

# Spectrum™ Technology Platform

Version 2018.2.0

Guide Data Quality



# Table des matières

## 1 - Démarrage

---

Introduction à la qualité des données	5
---------------------------------------	---

## 2 - Parsing

---

Présentation de l'analyse	8
Définition de grammaires d'analyse indépendantes du domaine dans des flux de données	9
Parsing spécifique à une culture	10
Analyse des résultats de parsing	37
Parser des noms de personne	41
Modèles de dataflow pour le parsing	42

## 3 - Standardisation

---

Standardisation des termes	59
Standardisation des noms de personne	60
Modèles de standardisation	62

## 4 - Correspondance

---

Mise en correspondance de la terminologie	66
Champs standard	68
Techniques de définition de match keys	69
Règles de correspondance	72
Mise en correspondance d'enregistrements d'une source unique	87
Mise en correspondance d'enregistrements d'une source avec ceux d'une autre source	93
Mise en correspondance d'enregistrements entre et dans des sources	100
Mise en correspondance d'enregistrements par rapport à une base de données	106

Rapprochement d'enregistrements à l'aide de plusieurs règles de rapprochement	109
Création d'un service de rapprochement universel	113
Utilisation d'une Match Key Express	117
Analyse des résultats de correspondance	121
Modèles de dataflows pour la correspondance	138

## 5 - Déduplication

---

Filtrage des enregistrements doublons	147
Création d'un enregistrement Best of Breed	152

## 6 - Enregistrements d'exception

---

Conception d'un flux de donnée afin de gérer les exceptions	160
Désignation d'un dataflow à des fins de revalidation en temps réel	162

## 7 - Tables de recherche

---

Présentation des tables de recherche	166
Tables du module Data Normalization	166
Tables du module Universal Name	172
Affichage du contenu d'une table de recherche	173
Ajout d'un terme dans une table de recherche	174
Suppression d'un terme d'une table de recherche	175
Modification de la forme standardisée d'un terme	175
Rétablissement de la Personnalisation de la table	175
Création d'une table de recherche	176
Importer des données	177

## 8 - Référence aux stages

---

Module Advanced Matching	181
Module Business Steward	250
Module Data Normalization	297
Module Universal Name	317

## 9 - Prise en charge du module et des codes ISO de pays

---

Prise en charge du module et des codes ISO de pays	353
---	-----

# 1 - Démarrage

## In this section

---

Introduction à la qualité des données

5

# Introduction à la qualité des données

La qualité des données implique la précision, la transmission en temps opportun, l'exhaustivité et la cohérence des données utilisées par une organisation afin que les données conviennent à l'utilisation. Spectrum™ Technology Platform prend en charge les initiatives de qualité des données en fournissant les capacités suivantes.

## *Parsing*

L'analyse est le processus consistant à analyser une séquence de caractères d'entrée dans un champ et à la diviser en plusieurs champs. Par exemple, vous pouvez disposer d'un champ appelé Nom qui contient la valeur « John A. Smith » et, grâce à l'analyse, vous pouvez le diviser afin d'obtenir un champ Prénom contenant « John », un champ Deuxième prénom contenant « A » et un champ Nom de famille contenant « Smith ».

## *Standardisation*

La standardisation prend des données de même type et leur applique le même format. Par exemple, les numéros de téléphone, les dates, les noms, les adresses et les numéros d'identification sont des types de données qui peuvent être standardisés. Par exemple, les numéros de téléphone peuvent être formatés pour éliminer les caractères non numériques, tels que les parenthèses, les points ou les tirets.

Vous devez standardiser vos données avant d'effectuer des activités de mise en correspondance ou de déduplication puisque les données standardisées seront mises en correspondance de manière plus précise que des données au format incohérent.

## *Correspondance*

La mise en correspondance est le processus d'identification des enregistrements liés les uns aux autres, d'une manière significative pour vos besoins. Par exemple, si vous essayez d'éliminer les informations redondantes de vos données client, il serait judicieux d'identifier les enregistrements doublons pour le même client ou, si vous essayez d'éliminer des courriers marketing doublons arrivant à la même adresse, vous souhaitez peut-être identifier les enregistrements des clients vivant dans le même foyer.

## *Déduplication*

La déduplication identifie les enregistrements qui représentent une entité mais qui, pour une raison quelconque, ont été saisis dans le système plusieurs fois, parfois avec des données légèrement différentes. Par exemple, votre système peut contenir des informations sur les fournisseurs, issues de différents services dans votre organisation qui, chacun, utilisent un ID de fournisseur différent pour le même fournisseur. À l'aide de Spectrum™ Technology Platform, vous pouvez consolider ces enregistrements dans un enregistrement unique pour chaque fournisseur.

### *Examen des enregistrements d'exception*

Dans certains cas, vous pouvez disposer de données qui ne peuvent pas être traitées automatiquement de manière confidentielle et qui doivent être examinées par un data steward fiable. Voici quelques exemples d'enregistrements qui peuvent exiger une vérification manuelle :

- Échecs de vérification d'adresse
- Échecs de géocodage
- Correspondances incertaines
- Décisions de fusion/consolidation

Le module Business Steward rassemble des fonctions qui vous permettent d'identifier et de résoudre des enregistrements d'exception.

# 2 - Parsing

## In this section

---

Présentation de l'analyse	8
Définition de grammaires d'analyse indépendantes du domaine dans des flux de données	9
Parsing spécifique à une culture	10
Analyse des résultats de parsing	37
Parser des noms de personne	41
Modèles de dataflow pour le parsing	42

## Présentation de l'analyse

L'analyse est le processus consistant à analyser une séquence de caractères d'entrée dans un champ et à la diviser en plusieurs champs. Par exemple, vous pouvez disposer d'un champ appelé Nom qui contient la valeur « John A. Smith » et, grâce à l'analyse, vous pouvez le diviser afin d'obtenir un champ Prénom contenant « John », un champ Deuxième prénom contenant « A » et un champ Nom de famille contenant « Smith ».

Pour créer un flux de données qui effectue l'analyse, utilisez le stage Open Parser. Open Parser vous permet d'écrire les règles d'analyse appelées *grammaires*. Une grammaire est un ensemble d'expressions qui mappent une séquence dans un ensemble d'entités nommées appelé *modèles de domaine*. Un modèle de domaine est une séquence d'un ou plusieurs jetons dans vos données d'entrée que vous souhaitez représenter par une structure de données, tels qu'un nom, une adresse ou des numéros de compte. Un modèle de domaine peut être composé de n'importe quel nombre de jetons pouvant être analysés depuis vos données d'entrée. Un modèle de domaine est représenté dans la grammaire de parsing comme l'expression <root>. Les données d'entrée contiennent souvent ce type de jetons sous un format difficile à utiliser ou un format mixte. Par exemple :

- Vos données d'entrée contiennent des noms dans un seul champ que vous souhaitez séparer en des prénoms et des noms de famille.
- Vos données d'entrée contiennent des adresses provenant de plusieurs cultures et vous souhaitez extraire les données d'adresse uniquement pour une culture spécifique.
- Vos données d'entrée comprennent du texte en forme libre contenant des adresses électroniques intégrées et vous souhaitez extraire les adresses électroniques et les mettre en correspondance avec des données personnelles et les stocker dans une base de données.

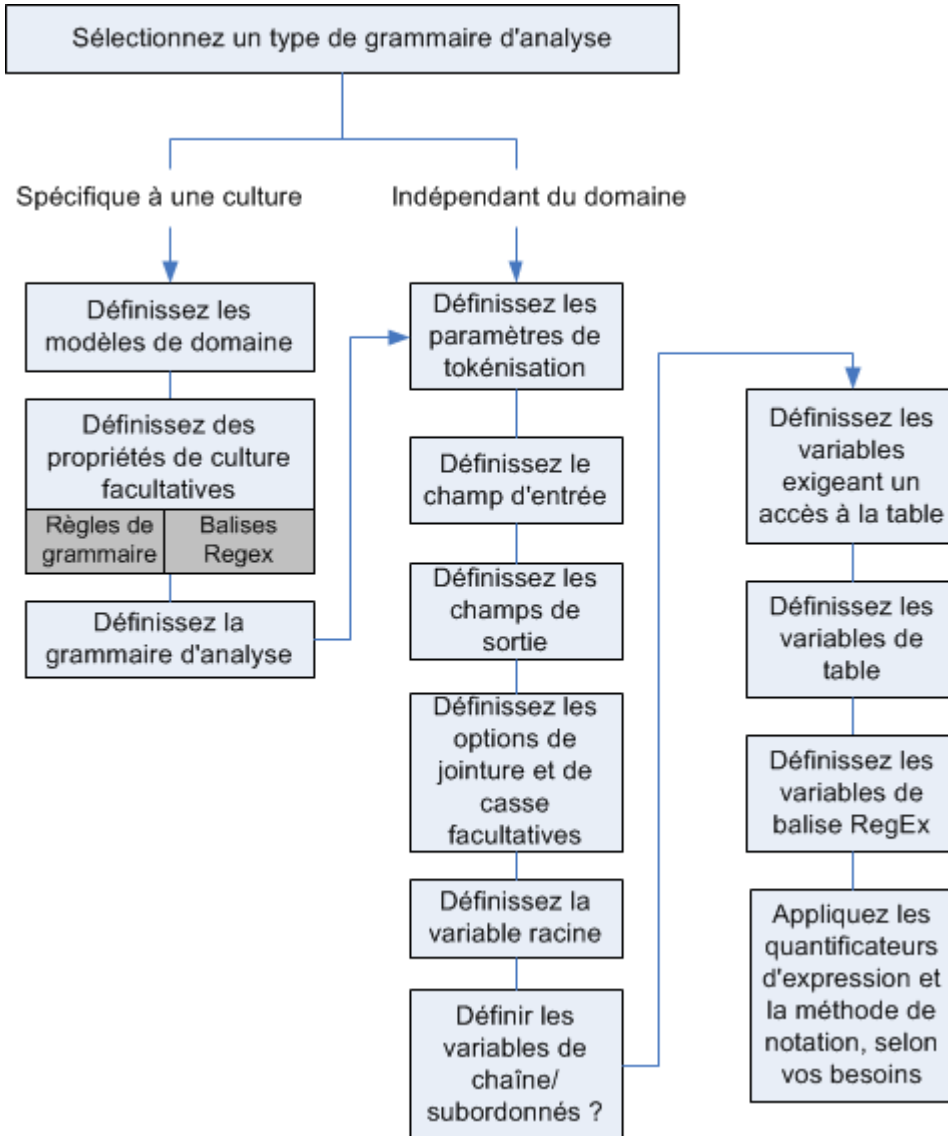
Il existe deux types de grammaires : *propre à la culture* et *indépendante au domaine*. Une grammaire d'analyse propre à la culture est associée à une culture et/ou à une langue (telle que l'anglais, l'anglais canadien, l'espagnol, l'espagnol mexicain, etc.) et à un type particulier de données (numéros de téléphone, noms personnels, etc.). Lorsqu'un stage Open Parser est configuré pour effectuer une analyse propre à la culture, la grammaire d'analyse de chaque culture s'applique à tous les enregistrements. La grammaire dotée du meilleur score d'analyse (ou la première atteignant un score de 100) est celle dont les résultats sont renvoyés. Sinon, les grammaires d'analyse propres à la culture peuvent utiliser la valeur dans le champ CultureCode de l'enregistrement d'entrée et traiter les données en fonction des paramètres de la culture contenus dans la grammaire d'analyse de la culture. Les grammaires de parsing spécifiques à une culture peuvent hériter des propriétés d'un parent. Une grammaire d'analyse indépendante du domaine n'est pas associée à une langue ou à un type particulier de données. Les grammaires de parsing indépendantes d'un domaine n'héritent pas des propriétés d'un parent et ignorent toute information de CultureCode des données d'entrée.

Open Parser analyse une séquence de caractères dans les champs d'entrée et les catégorise dans un séquence de jetons via un processus appelé *segmentation*. La segmentation consiste à délimiter et à classifier les sections d'une chaîne de caractères d'entrée en un ensemble de jetons par



marquage des caractères (également appelé caractères de segmentation), par exemple, un espace, un trait d'union, etc. Les jetons sont ensuite placés dans les champs de sortie que vous indiquez.

Le schéma suivant illustre la procédure de création d'une grammaire d'analyse :



## Définition de grammaires d'analyse indépendantes du domaine dans des flux de données

Pour définir des grammaires d'analyse indépendantes du domaine dans un flux de données, procédez comme suit :

1. Dans Enterprise Designer, ajoutez un stage Open Parser à votre flux de données.
2. Double-cliquez sur le stage Open parser sur le canevas.
3. Cliquez sur **Définir une grammaire indépendante du domaine** depuis l'onglet **Règles**.
4. Utilisez l'Éditeur de grammaire pour créer les règles de grammaire. Vous pouvez saisir des commandes et des variables dans la case de texte ou utiliser les commandes mises à votre disposition depuis l'onglet **Commandes**. Pour plus d'informations, reportez-vous à la section **Grammaires** à la page 27.
5. Faites un clic-droit dans l'Éditeur de grammaire et choisissez la commande adéquate pour couper, copier, coller et rechercher et remplacer des chaînes de texte dans votre grammaire de parsing.
6. Pour vérifier la grammaire de parsing que vous avez créée, cliquez sur **Valider**.  
 La fonctionnalité valider énumère toutes les erreurs dans la syntaxe de votre grammaire, y compris la rangée et la colonne où l'erreur s'est produite, une description de l'erreur et le nom ou la valeur de la commande ayant provoqué l'erreur.
7. Cliquez sur l'onglet **Aperçu** afin de tester la grammaire de parsing.
8. Une fois que vous avez fini de créer votre grammaire de parsing, cliquez sur **OK**.

## Parsing spécifique à une culture

### Définition d'une grammaire de parsing spécifique à une culture

Une grammaire d'analyse propre à la culture vous permet de spécifier différentes règles d'analyse pour différentes langues et cultures. Cela vous permet d'analyser les données de différents pays dans un stage Open Parser unique, par exemple des numéros de téléphone des Etats-Unis et du Royaume-Uni. Par défaut, chaque enregistrement d'entrée est analysé à l'aide de la grammaire d'analyse de chaque culture, dans l'ordre indiqué dans le stage Open Parser. Vous pouvez également ajouter un champ CultureCode aux enregistrements d'entrée si vous souhaitez utiliser une grammaire d'analyse de culture spécifique pour l'enregistrement. Pour plus d'informations, reportez-vous à la section **Affectation d'une culture d'analyse à un enregistrement** à la page 12.

**Remarque :** Si vous souhaitez créer une grammaire d'analyse indépendante du domaine, reportez-vous à la section **Définition de grammaires d'analyse indépendantes du domaine dans des flux de données** à la page 9.

1. Dans Enterprise Designer, accédez à **Outils > Open Parser Domain Editor**.
2. Cliquez sur l'onglet **Domaines**.
3. Cliquez sur **Ajouter**.

4. Saisissez un nom de domaine dans le champ **Nom**.
5. Saisissez la description du nom de domaine dans le champ **Description**.
6. Si vous souhaitez créer un nouveau domaine vierge, cliquez sur **OK**. Si vous souhaitez créer un domaine basé sur un autre domaine, procédez comme suit :
  - a) Sélectionnez **Utiliser un autre domaine comme modèle** pour créer un domaine en fonction d'un autre domaine.
  - b) Choisissez un domaine dans la liste. Le nouveau domaine sera créé une fois que vous aurez cliqué sur **OK** lors de l'étape suivante. Le nouveau domaine contiendra toutes les grammaires de parsing spécifiques à une culture définies dans le modèle de domaine sélectionné.
  - c) Cliquez sur **OK**.
7. Définissez la grammaire d'analyse pour la culture générale. La culture générale est la culture par défaut et permet d'analyser les enregistrements disposant d'une culture pour laquelle aucune grammaire d'analyse propre à la culture n'a été définie.
  - a) Dans l'onglet **Grammaires**, sélectionnez le domaine que vous avez créé.
  - b) Si vous avez créé un domaine à partir d'un modèle, des cultures peuvent déjà être répertoriées.
    - Si des cultures sont répertoriées, sélectionnez **Culture générale**, puis cliquez sur **Modifier**.
    - Si aucune culture n'est répertoriée, cliquez sur **Ajouter**, sélectionnez **Culture générale**, puis cliquez sur **OK**.
  - c) Dans l'onglet **Grammaire**, écrivez la grammaire d'analyse de la culture générale. Vous pouvez utiliser les onglets **Commandes**, **Règles de la grammaire** et **Balises RegEx** pour insérer des éléments de grammaire d'analyse prédéfinis. Pour entrer un élément prédéfini, placez le curseur là où vous souhaitez insérer l'élément, puis double-cliquez sur l'élément à ajouter.
 

L'onglet **Commandes** affiche les commandes d'analyse. Pour obtenir des informations sur les commandes disponibles, reportez-vous à la section **Grammaires** à la page 27.

L'onglet **Règles de grammaire** affiche les règles de grammaire que vous créez dans la boîte de dialogue Propriétés de la culture. Pour plus d'informations sur la création de règles de grammaire, voir **Définition des règles de grammaire d'une culture** à la page 32.

L'onglet **Balises RegEx** affiche les balises RegEx que vous créez dans la boîte de dialogue Propriétés de la culture. Pour plus d'informations sur la création de balises RegEx, voir **Définir des balises RegEx de culture** à la page 33.
  - d) Pour vérifier la syntaxe de grammaire que vous avez créée, cliquez sur **Valider**. La fonction de validation de la grammaire de parsing affiche toute erreur dans la syntaxe de votre grammaire et inclut l'erreur rencontrée, la ligne et la colonne dans laquelle se trouve l'erreur, et la commande, la règle de grammaire ou la balise RegEx dans laquelle se produit l'erreur.
  - e) Pour tester les résultats de votre grammaire avec les exemples de données, cliquez sur l'onglet **Aperçu**. Sous **Données d'entrée**, entrez les exemples de données à analyser. Entrez un enregistrement par ligne. Ensuite, cliquez sur le bouton **Aperçu**. Les champs

analysés s'afficheront dans la grille des **Résultats**. Pour obtenir des informations sur les champs de sortie, voir **Sortie** à la page 305. Pour obtenir des informations sur les traces, voir **Suivi des résultats finaux de parsing** à la page 37. Si vous n'obtenez pas les résultats escomptés, cliquez sur l'onglet **Grammaires** et continuez d'éditer la grammaire de parsing et de tester les données d'entrée représentatives jusqu'à ce que la grammaire de parsing produise les résultats escomptés.

- f) Cliquez sur **OK** lorsque vous avez terminé de définir la grammaire d'analyse pour la culture générale.
8. Définissez une grammaire spécifique à une culture pour chaque culture de votre choix. Pour ajouter des grammaires propres à une culture, cliquez sur **Ajouter** et définissez la grammaire à l'aide des mêmes étapes que pour la culture générale. Répétez l'opération autant de fois que nécessaire pour ajouter le nombre de cultures dont vous avez besoin.
9. Lorsque vous avez terminé d'ajouter des grammaires d'analyse spécifiques à la culture, cliquez sur **OK**.

Le domaine et les cultures que vous avez créés peuvent désormais être utilisés dans le stage Open Parser pour effectuer l'analyse.

## Affectation d'une culture d'analyse à un enregistrement

Lorsque vous configurez un stage Open Parser pour utiliser des grammaires d'analyse propres à une culture, les grammaires d'analyse de chaque culture sont appliquées à chaque enregistrement d'entrée dans l'ordre dans lequel les cultures sont répertoriées dans le stage Open Parser.

Cependant, si vous souhaitez appliquer une grammaire d'analyse d'une culture spécifique à un enregistrement, vous pouvez ajouter un champ nommé CultureCode. Le champ doit contenir un des codes de culture pris en charge répertoriés dans la table suivante.

### Codes de culture

Les codes de culture se composent d'un code de langue de deux lettres en minuscules et d'un code de pays ou de région de deux lettres en majuscules. Par exemple, « es-MX » pour l'espagnol (Mexique) et « fr-FR » pour l'anglais (États-Unis). Dans les cas où un code de langue à deux lettres n'est pas disponible, un code à trois lettres sera utilisé : par exemple, « uz-Cyrl-UZ » pour Ouzbek (Ouzbékistan, Cyrillique). Seul le code de langue à deux lettres minuscules désigne une langue. Par exemple, « fr » indique la culture neutre Français, et « de » désigne la culture neutre Allemand.

**Remarque :** Deux noms de culture suivent un modèle différent. Les cultures « zh-Hans » (Chinois simplifié) et « zh-Hant » (Chinois traditionnel) sont des cultures neutres. Les noms de culture représentent le standard actuel et devraient être employés à moins d'avoir une raison d'employer les anciens noms « zh-CHS » et « zh-CHT ».

Le tableau suivant présente les codes de culture pris en charge.

Langue (Culture/Région)	Code culture
Culture générale	Culture générale
Africains	af
Africains (Afrique du Sud)	af-ZA
Albanais	sq
Albanais (Albanie)	sq-AL
Arabe	ar
Arabe (Algérie)	ar-DZ
Arabe (Bahreïn)	ar-BH
Arabe (Égypte)	ar-EG
Arabe (Irak)	ar-IQ
Arabe (Jordanie)	ar-JO
Arabe (Koweït)	ar-KW
Arabe (Liban)	ar-LB
Arabe (Libye)	ar-LY

Langue (Culture/Région)	Code culture
-------------------------	--------------

Arabe (Maroc)	ar-MA
---------------	-------

Arabe (Oman)	ar-OM
--------------	-------

Arabe (Qatar)	ar-QA
---------------	-------

Arabe (Arabie saoudite)	ar-SA
-------------------------	-------

Arabe (Syrie)	ar-SY
---------------	-------

Arabe (Tunisie)	ar-TN
-----------------	-------

Arabe (E.A.U.)	ar-AE
----------------	-------

Arabe (Yémen)	ar-YE
---------------	-------

Arménien	hy
----------	----

Arménien (Arménie)	hy-AM
--------------------	-------

Azéris	az
--------	----

Azéris (Azerbaïdjan, Cyrillique)	az-Cyrl-AZ
----------------------------------	------------

Azéris (Azerbaïdjan, Latin)	az-Latn-AZ
-----------------------------	------------

Basque	eu
--------	----

Langue (Culture/Région)	Code culture
-------------------------	--------------

Basque (Basque)	eu-ES
-----------------	-------

Bielorusse	be
------------	----

Bielorusse (Biélorussie)	be-BY
--------------------------	-------

Bulgare	bg
---------	----

Bulgare (Bulgarie)	bg-BG
--------------------	-------

Catalan	ca
---------	----

Catalan (Catalogne)	ca-ES
---------------------	-------

Chinois	zh
---------	----

Chinois (Hong Kong RAS, RPC)	zh-HK
------------------------------	-------

Chinois (Macao RAS)	zh-MO
---------------------	-------

Chinois (PRC)	zh-CN
---------------	-------

Chinois (Simplifié)	zh-Hans
---------------------	---------

Chinois (Singapour)	zh-SG
---------------------	-------

Chinois (Taïwan)	zh-TW
------------------	-------

Langue (Culture/Région)	Code culture
-------------------------	--------------

Chinois (Traditionnel)	zh-Hant
------------------------	---------

Croate	hr
--------	----

Croate (Croatie)	hr-HR
------------------	-------

Tchèque	cs
---------	----

Tchèque (République tchèque)	cs-CZ
------------------------------	-------

Danois	da
--------	----

Danois (Danemark)	da-DK
-------------------	-------

Divehi	dv
--------	----

Divehi (Maldives)	dv-MV
-------------------	-------

Néerlandais	nl
-------------	----

Néerlandais (Belgique)	nl-BE
------------------------	-------

Néerlandais (Pays-Bas)	nl-NL
------------------------	-------

Anglais	en
---------	----

Anglais (Australie)	en-AU
---------------------	-------



Langue (Culture/Région)	Code culture
-------------------------	--------------

---

Anglais (Belize)	en-BZ
------------------	-------

---

Anglais (Canada)	en-CA
------------------	-------

---

Anglais (Caraïbe)	en-029
-------------------	--------

---

Anglais (Irlande)	en-IE
-------------------	-------

---

Anglais (Jamaïque)	en-JM
--------------------	-------

---

Anglais (Nouvelle-Zélande)	en-NZ
----------------------------	-------

---

Anglais (Philippines)	en-PH
-----------------------	-------

---

Anglais (Afrique du sud)	en-ZA
--------------------------	-------

---

Anglais (Trinité-et-Tobago)	en-TT
-----------------------------	-------

---

Anglais (Royaume-Uni)	en-GB
-----------------------	-------

---

Anglais (États-Unis)	fr-FR
----------------------	-------

---

Anglais (Zimbabwe)	en-ZW
--------------------	-------

---

Estonien	et
----------	----

---

Estonien (Estonie)	et-EE
--------------------	-------

---

Langue (Culture/Région)	Code culture
-------------------------	--------------

Féroé	fo
-------	----

Féroé (Îles Féroé)	fo-FO
--------------------	-------

Farsi	fa
-------	----

Farsi (Iran)	fa-IR
--------------	-------

Finnois	fi
---------	----

Finlandais (Finlande)	fi-FI
-----------------------	-------

Français	fr
----------	----

Français (Belgique)	fr-BE
---------------------	-------

Français (Canada)	fr-CA
-------------------	-------

Français (France)	fr-FR
-------------------	-------

Français (Luxembourg)	fr-LU
-----------------------	-------

Français (Monaco)	fr-MC
-------------------	-------

Français (Suisse)	fr-CH
-------------------	-------

Galicien	gl
----------	----

Langue (Culture/Région)	Code culture
-------------------------	--------------

Galicien (Espagne)	gl-ES
--------------------	-------

Géorgien	ka
----------	----

Géorgien (Géorgie)	ka-GE
--------------------	-------

Allemand	de
----------	----

Allemand (Autriche)	de-AT
---------------------	-------

Allemand (Allemagne)	de-DE
----------------------	-------

Allemand (Liechtenstein)	de-LI
--------------------------	-------

Allemand (Luxembourg)	de-LU
-----------------------	-------

Allemand (Suisse)	de-CH
-------------------	-------

Grec	el
------	----

Grec (Grèce)	el-GR
--------------	-------

Gujarâṭī	gu
----------	----

Gujarâṭī (Inde)	gu-IN
-----------------	-------

Hébreu	he
--------	----

Langue (Culture/Région)	Code culture
-------------------------	--------------

---

Hébreu (Israël)	he-IL
-----------------	-------

---

Hindi	hi
-------	----

---

Hindi (Inde)	hi-IN
--------------	-------

---

Hongrois	hu
----------	----

---

Hongrois (Hongrie)	hu-HU
--------------------	-------

---

Islandais	is
-----------	----

---

Islandais (Islande)	is-IS
---------------------	-------

---

Indonésien	id
------------	----

---

Indonésien (Indonésie)	id-ID
------------------------	-------

---

Italien	it
---------	----

---

Italien (Italie)	it-IT
------------------	-------

---

Italien (Suisse)	it-CH
------------------	-------

---

Japonais	ja
----------	----

---

Japonais (Japon)	ja-JP
------------------	-------

---

Langue (Culture/Région)	Code culture
-------------------------	--------------

Kannada	kn
---------	----

Kannada (Inde)	kn-IN
----------------	-------

Kazakh	kk
--------	----

Kazakh (Kazakhstan)	kk-KZ
---------------------	-------

Konkani	kok
---------	-----

Konkani (Inde)	kok-IN
----------------	--------

Coréen	ko
--------	----

Coréen (Corée)	ko-KR
----------------	-------

Kyrgyz	ky
--------	----

Kyrgyz (Kyrgyzstan)	ky-KG
---------------------	-------

Letton	lv
--------	----

Letton (Lettonie)	lv-LV
-------------------	-------

Lituanien	lt
-----------	----

Lituanien (Lituanie)	lt-LT
----------------------	-------

Langue (Culture/Région)	Code culture
-------------------------	--------------

Macédonien	mk
------------	----

Macédonien (Macédoine, ARYM)	mk-MK
------------------------------	-------

Malais	ms
--------	----

Malais (Brunei)	ms-BN
-----------------	-------

Malais (Malaisie)	ms-MY
-------------------	-------

Marathi	mr
---------	----

Marathi (Inde)	mr-IN
----------------	-------

Mongol	mn
--------	----

Mongol (Mongolie)	mn-MN
-------------------	-------

Norvégien	non
-----------	-----

Norvégien (Bokmål, Norvège)	nb-NO
-----------------------------	-------

Norvégien (Nynorsk, Norvège)	nn-NO
------------------------------	-------

Polonais	pl
----------	----

Polonais (Pologne)	pl-PL
--------------------	-------

Langue (Culture/Région)	Code culture
-------------------------	--------------

Portugais	pt
-----------	----

Portugais (Brésil)	pt-BR
--------------------	-------

Portugais (Portugal)	pt-PT
----------------------	-------

Punjabi	pa
---------	----

Punjabi (Inde)	pa-IN
----------------	-------

Roumain	ro
---------	----

Roumain (Roumanie)	ro-RO
--------------------	-------

Russe	ru
-------	----

Russe (Russie)	ru-RU
----------------	-------

Sanskrit	SA
----------	----

Sanskrit (Inde)	sa-IN
-----------------	-------

Serbe	sr
-------	----

Serbe (Serbie, Cyrillique)	sr-Cyrl-CS
----------------------------	------------

Serbe (Serbie, Latin)	sr-Latn-CS
-----------------------	------------

Langue (Culture/Région)	Code culture
-------------------------	--------------

---

Slovaque	sk
----------	----

---

Slovaque (Slovaquie)	sk-SK
----------------------	-------

---

---

Slovène	sl
---------	----

---

Slovène (Slovénie)	sl-SI
--------------------	-------

---

---

Espagnol	es
----------	----

---

Espagnol (Argentine)	es-AR
----------------------	-------

---

---

Espagnol (Bolivie)	es-BO
--------------------	-------

---

Espagnol (Chili)	es-CL
------------------	-------

---

---

Espagnol (Colombie)	es-CO
---------------------	-------

---

Espagnol (Costa Rica)	es-CR
-----------------------	-------

---

---

Espagnol (République dominicaine)	es-DO
-----------------------------------	-------

---

Espagnol (Équateur)	es-EC
---------------------	-------

---

---

Espagnol (El Salvador)	es-SV
------------------------	-------

---

Espagnol (Guatemala)	es-GT
----------------------	-------

---



Langue (Culture/Région)	Code culture
-------------------------	--------------

---

Espagnol (Honduras)	es-HN
---------------------	-------

---

Espagnol (Mexique)	es-MX
--------------------	-------

---

Espagnol (Nicaragua)	es-NI
----------------------	-------

---

Espagnol (Panama)	es-PA
-------------------	-------

---

Espagnol (Paraguay)	es-PY
---------------------	-------

---

Espagnol (Pérou)	es-PE
------------------	-------

---

Espagnol (Porto Rico)	es-PR
-----------------------	-------

---

Espagnol (Espagne)	es-ES
--------------------	-------

---

Espagnol (Espagne, traditionnel)	es-ES_tradnl
----------------------------------	--------------

---

Espagnol (Uruguay)	es-UY
--------------------	-------

---

Espagnol (Venezuela)	es-VE
----------------------	-------

---

Swahili	sw
---------	----

---

Swahili (Kenya)	sw-KE
-----------------	-------

---

Suédois	sv
---------	----

---

Langue (Culture/Région)	Code culture
-------------------------	--------------

---

Suédois (Finlande)	sv-FI
--------------------	-------

---

Suédois (Suède)	sv-SE
-----------------	-------

---

Syriaque	syr
----------	-----

---

Syriaque (Syrie)	syr-SY
------------------	--------

---

Tamoul	ta
--------	----

---

Tamoul (Inde)	ta-IN
---------------	-------

---

Tatar	tt
-------	----

---

Tatar (Russie)	tt-RU
----------------	-------

---

Telugu	te
--------	----

---

Telugu (Inde)	te-IN
---------------	-------

---

Thaï	th
------	----

---

Thai (Thaïlande)	th-TH
------------------	-------

---

Turque	tr
--------	----

---

Turque (Turquie)	tr-TR
------------------	-------

---

Langue (Culture/Région)	Code culture
Ukrainien	uk
Ukrainien (Ukraine)	uk-UA
Urdu	ur
Urdu (Pakistan)	ur-PK
Uzbek	uz
Uzbek (Ouzbékistan, Cyrillique)	uz-Cyrl-UZ
Uzbek (Ouzbékistan, Latin)	uz-Latn-UZ
Vietnamien	vi
Vietnamien (Vietnam)	vi-VN

## Grammaires

Une grammaire de parsing valide contient :

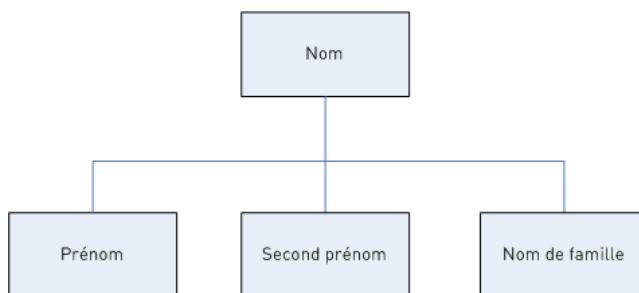
- Une variable racine définit la séquence de jetons, ou modèle de domaine, comme variables de règle.
- Les variables de règle définissant l'ensemble valide de caractère et la séquence selon laquelle ces caractères peuvent survenir afin d'être considéré comme composant d'un modèle de domaine. Pour plus d'informations, reportez-vous à la section [Commandes de la section de règles](#) à la page 31.
- Le champ d'entrée à analyser. Le champ d'entrée désigne le champ à analyser dans les enregistrements de données source.

- Les champs de sortie pour les données analysées produites. Les champs de sortie définissent l'emplacement où stocker chaque jeton produit analysé.

Une grammaire de parsing valide contient également d'autres commandes facultatives pour :

- Les caractères utilisés pour segmenter les données d'entrée que vous analysez. Les caractères de segmentation sont des caractères comme des espaces et des tirets définissant le début et la fin d'un jeton. Le caractère de segmentation par défaut est un espace. Les caractères de segmentation représentent la principale voie permettant de décomposer une séquence de caractères en un ensemble de jetons. Vous pouvez régler la commande Segmenter sur AUCUN pour empêcher la jetonisation du champ. Lorsque la commande Segmenter est réglée sur Aucun, tout espace des règles de grammaire doit être compris dans la définition des règles.
- L'option de sensibilité à la casse pour les jetons se trouve dans les données d'entrée.
- Un caractère attaché permettant de délimiter les jetons correspondants.
- Mettre en correspondance des jetons dans des tableaux
- Mettre en correspondance des jetons composés dans des tableaux
- Définir des balises RegEx
- Les chaînes littérales entre guillemets
- Les quantificateurs d'expressions (facultatif). Pour de plus amples informations sur les quantificateurs d'expressions, voir [Commandes de la section de règles](#) à la page 31 et [Quantificateurs d'expression : Comportement Avide, Réticent et Possessif](#).
- D'autres indicateurs divers servant au regroupement, aux commentaires et aux attributions (facultatif). Pour de plus amples informations sur les expressions regroupées, voir [Opérateur de regroupement \( \)](#).

Les variables de règle dans votre grammaire de parsing forment une structure d'arbre à plusieurs niveaux avec la séquence de caractères ou avec des jetons dans un modèle de domaine. Par exemple, vous pouvez créer une grammaire de parsing définissant un modèle de domaine basé sur des données d'entrée de nom contenant les jetons <FirstName>, <MiddleName> et <LastName>.



Utiliser les données d'entrée :

Joseph Arnold Cowers

Vous pouvez représenter cette chaîne de données comme trois jetons dans un modèle de domaine :

```
<root> = <FirstName><MiddleName><LastName>;
```

Les variables de règle pour ce modèle de domaine sont :

```
<FirstName> = <given>;
<MiddleName> = <given>;
<LastName> = @Table("Family Names");
<given> = @Regex("[A-Za-z]+");
```

En se basant sur cette exemple de grammaire simple, l'Open Parser segmente à la hauteur des espaces et interprète le jeton `Joseph` comme étant un prénom, car les caractères dans ce premier jeton correspondent à la définition `[A-Za-z]+` et que le jeton est dans la séquence définie.

Facultativement, toute expression peut être suivie par une autre.

### Exemple

```
<variable> = "some leading string" <variable2>;
<variable2> = @Table ("given") @Regex("[0-9]+");
```

Une règle de grammaire est une phrase grammaticale où une variable est égale à une ou plusieurs expressions. Chaque règle de grammaire suit la forme suivante :

```
<rule> = expression [| expression...];
```

Les règles de grammaire doivent suivre les règles suivantes :

- `<root>` est un nom de variable spécial et la première règle exécutée dans la grammaire, car il définit le modèle du domaine. `<root>` peut ne pas être référencé par une autre règle de la grammaire.
- Une variable `<rule>` ne peut pas se référer à elle-même de manière directe ou indirecte. Lorsque la règle A se réfère à la règle B, qui se réfère à la règle C, qui se réfère à la règle A, cela crée une référence circulaire. Les références circulaires ne sont pas autorisées.
- Une variable `<rule>` est égale à une ou plusieurs expressions.
- Chaque `expression` est séparée par un OR, qui est indiqué par le caractère de la barre verticale (`|`).
- Les expressions sont examinées l'une après l'autre. La première `expression` à mettre en correspondance est sélectionnée. Aucune autre expression ne sera examinée.
- Le nom de la variable peut être composé de caractères alphabétiques, numériques, de tiret bas (`_`) et de tirets (`-`). Le nom de la variable peut commencer par tout caractère valide. Si le nom du champ de sortie spécifié ne respecte pas cette forme, utilisez la fonctionnalité `alias` pour attribuer le nom de variable au champ de sortie.

Il existe plusieurs types d'expression :

- Autre variable

- Chaîne se composant d'un ou de plusieurs caractères entre guillemets ou entre apostrophes. Par exemple :

```
"McDonald" 'McDonald' "O'Hara" 'O\'Hara' 'D"har' "D\"har"
```

- Table
- CompoundTable
- Commandes RegEx

### Métacaractères de commande

Open Parser prend en charge l'ensemble standard de métacaractères de classe de caractère de Java RegEx sur les commandes %Tokenize et @RegEx. Un métacaractère est un caractère ayant une certaine signification dans la mise en correspondance de modèles. Les métacaractères pris en charge sont les suivants :

```
([{\^-$| ])?*+.
```

Pour obliger un métacaractère à être traité comme un caractère ordinaire, deux méthodes s'offrent à vous :

- Placez une barre oblique avant le métacaractère.
- Encadrez-le avec \Q (ouvrant le guillemet) et \E (le fermant).

%Tokenize suit la règle des classes de caractère des expressions régulières de Java—pas l'ensemble des expressions régulières de Java.

De manière générale, les caractères réservés pour un ensemble de caractères sont :

- « | » et « ] » indiquent un autre ensemble.
- « - » est un métacaractère s'il se trouve entre deux autres caractères.
- « ^ » est un métacaractère s'il est le premier caractère d'un ensemble.
- « && » sont des métacaractères s'ils se trouvent entre deux autres caractères.
- « \ » signifie que le caractère suivant est littéral.

Si vous avez des doutes sur si un caractère sera traité comme un métacaractère et souhaitez qu'il soit traité comme littéral, utilisez la barre oblique pour vous en assurer.

### Commandes de section d'en-tête

Cette section décrit les commandes de section d'en-tête. Certaines commandes sont facultatives. Si une commande est facultative, le comportement ou la valeur par défaut est répertorié.

- **Tokenize** (facultatif)
- **Tokenize (None)**
- **InputField** (requis si Input Fields n'est pas utilisé)
- **InputFields** (requis si Input Field n'est pas utilisé)
- **OutputFields** (requis)
- **IgnoreCase** (facultatif)

- **Join** (facultatif)

### Commandes de la section de règles

Les commandes de la section de règles sont :

- **RegEx**
- **Table**
- **CompoundTable**
- **Token**
- **Scoring**
- **RuleID**
- **Variable <root>**
- **rule|rule**
- **Opérateur de groupement ( )**
- **Opérateur d'occurrence Min/Max {min,max}**
- **Opérateur exact d'occurrence {exact}**
- **Opérateur d'assignation (=)**
- **Opérateur OR (|)**
- **Opérateur de fin de règle (;)**
- **Opérateur de commentaires (!)**
- **Quantificateur d'occurrence zéro ou un (?)**
- **Quantificateur d'occurrence zéro ou plus (?)**
- **One or More Occurrences Quantifier (+)**
- **Quantificateurs d'expression : Comportement Avide, Réticent et Possessif**

## Cultures

Une culture est la méthode principale permettant d'organiser des grammaires d'analyse propres à une culture. Vous pouvez utiliser des cultures pour créer différentes règles d'analyse pour les différentes cultures et langues. La culture suit une hiérarchie :

- **Culture générale** : la culture générale est indépendante d'une culture et linguistiquement agnostique. Utilisez la culture générale pour créer des règles de grammaire de parsing s'appliquant sur toutes les cultures et toutes les langues.
- **Langue** : une langue est associée à une langue mais pas avec une culture/région spécifique. Par exemple, l'anglais.
- **Culture/Région** : une culture/région est associée à une langue et un pays ou une région. Par exemple, l'anglais au Royaume-Uni ou l'anglais aux Etats-Unis.

Dans la hiérarchie de culture, le parent d'une culture/région est une langue et le parent d'une langue est la culture générale.

La culture/les régions héritent des propriétés de la langue parente. Les langues héritent des propriétés de la culture générale. Vous pouvez de cette manière définir des grammaires de parsing dans une langue pouvant être utilisée dans plusieurs pays partageant cette langue. Vous pouvez ainsi ignorer les règles de grammaire de langue avec des grammaires de parsing spécialisées à un pays ou une région particulière partageant la même langue que la culture et langue de base, mais ayant un adressage et une dénomination différents ou d'autres différences de pays ou de région.

Vous pouvez également utiliser l'héritage de culture pour analyser des enregistrements entrants ayant un code culture attribué mais pas de règle de grammaire définie pour ce code culture. Dans ce cas là, l'Open Parser cherche un code de langue ayant une règle de grammaire attribuée. Si celui-ci n'existe pas, l'Open Parser cherche une règle de grammaire attribuée dans la culture générale.

Domain Editor utilise une combinaison d'un code de langue et d'un code culture afin de représenter la langue et la culture/région respectivement.

### Définition des règles de grammaire d'une culture

Vous pouvez utiliser les règles de grammaire d'une culture pour remplacer une partie d'une grammaire d'analyse de la culture générale par des chaînes, des commandes ou des expressions propres à la culture et/ou au langage. En définissant une règle de grammaire, vous pouvez personnaliser les parties de la grammaire d'analyse de la culture générale en fonction de la culture et/ou du langage de l'enregistrement. Cela s'avère utile si vous ne souhaitez pas créer de grammaire d'analyse totalement distincte pour chaque culture et que vous voulez plutôt utiliser la grammaire de la culture générale, ne personnalisant que certaines parties de la grammaire de culture générale pour chaque culture.

Cette rubrique explique comment créer une règle de grammaire pour une culture.

1. Dans Enterprise Designer, accédez à **Outils > Open Parser Domain Editor**.
2. Cliquez sur l'onglet **Cultures**.

Afin d'obtenir la liste complète des cultures prises en charge, voir [Affectation d'une culture d'analyse à un enregistrement](#) à la page 12.

3. Sélectionnez la culture à laquelle vous souhaitez ajouter une règle de grammaire, puis cliquez sur **Propriétés**.
4. Cliquez sur l'onglet **Règles de grammaire**. Les informations affichées comprennent les noms des règles de grammaire définies pour la culture sélectionnée, la culture source associée, la valeur définie de la règle de grammaire et la description.
5. Cliquez sur **Ajouter**.
6. Tapez un nom pour la règle de grammaire dans le champ **Nom**.
7. Tapez une description pour la règle de grammaire dans le champ **Description**.
8. Saisissez la règle de grammaire dans le champ **Valeur**.



La règle de grammaire peut être n'importe quelle variable, chaîne, commande, ou expression groupée valide. Pour plus d'informations, reportez-vous à la section [Grammaires](#) à la page 27.

9. Sélectionnez **Activer le retour automatique à la ligne** pour afficher la valeur dans la zone de texte sans faire défiler.
10. Cliquez sur **OK**.

La valeur de la règle de grammaire que vous avez tapée est validée. Si la valeur contient des erreurs de syntaxe grammaticale, un message affiche une description des erreurs rencontrées, la ligne et colonne où l'erreur a eu lieu, et la commande, règle de grammaire, ou balise RegEx où l'erreur s'est produite.

#### Exemple de règle de grammaire

Vous disposez d'une grammaire qui analyse les noms occidentaux. La structure du modèle peut être la même pour toutes les cultures (<Prénom><DeuxièmePrénom><Nom>) et beaucoup de règles peuvent correspondre au même modèle ou à la même table. Cependant, vous disposez également de tables propres à la culture pour les noms de famille et vous souhaitez utiliser la table appropriée en fonction du code de culture de l'enregistrement.

Pour ce faire, vous pouvez définir une règle de grammaire pour chaque culture remplaçant l'élément <LastName> dans la culture générale avec une référence à la table propre à la culture. Par exemple, si vous disposez d'une table de noms de famille néerlandais, vous devez créer une règle de grammaire pour la culture néerlandaise (nl) comme suit :

**Nom** : nom de famille

**Description** : noms de famille néerlandais

**Valeur** : @Table("Dutch Last Names");

#### Définir des balises RegEx de culture

Cette rubrique explique comment définir des balises RegEx de culture lors de la définition d'une grammaire d'analyse propre à la culture.

1. Dans Enterprise Designer, accédez à **Outils > Open Parser Domain Editor**.
2. Cliquez sur l'onglet **Cultures**. L'onglet **Cultures** affiche une liste des cultures prises en charge. Afin d'obtenir la liste complète des cultures prises en charge, voir [Affectation d'une culture d'analyse à un enregistrement](#) à la page 12.
3. Choisissez une culture depuis la liste puis cliquez sur **Propriétés**. La boîte de dialogue **Propriétés de culture** s'affiche.
4. Cliquez sur l'onglet **Balises RegEx**. Les informations affichées comprennent les noms des balises RegEx définies pour la culture sélectionnée et la culture source associée, la valeur de la balise RegEx ainsi que la description.
5. Cliquez sur **Ajouter** ou **Modifier**.

6. Saisissez un nom dans la zone de texte **Nom** pour la balise RegEx.

Si vous saisissez un nom existant déjà pour la culture sélectionnée, une icône d'avertissement se met à clignoter. Saisissez un nom différent ou fermez la boîte de dialogue, supprimez la balise RegEx existante puis cliquez à nouveau sur **Ajouter**.

7. Saisissez la description de l'onglet RegEx dans la zone de texte **Description**.
8. Saisissez une valeur dans la zone de texte **Valeur** pour la balise RegEx.

Cette valeur peut être toute expression régulière valable mais ne peut pas être mise en correspondance avec une chaîne vide.

Domain Editor comprend plusieurs balises RegEx prédéfinies que vous pouvez utiliser pour définir des propriétés de culture. Vous pouvez également utiliser ces balises RegEx pour définir des caractères de jetonisation dans votre grammaire de parsing.

Vous pouvez modifier les balises RegEx prédéfinies ou les copier afin de créer vos propres variantes. Vous pouvez également désactiver les propriétés pour créer des balises RegEx spécialisées pour des langues spécifiques.

- Lettre : toute lettre de n'importe quelle langue. Cette balise RegEx comprend des neutralisateurs pour plusieurs langues à cause des différences dans les scripts utilisés, tels que les scripts cyrilliques, les scripts de langues asiatiques et le script thaïlandais.
- Minuscule : lettre minuscule ayant une variante majuscule.
- Numérique : tout caractère numérique dans n'importe quel script.
- Ponctuation : tout caractère de ponctuation.
- Majuscule : toute lettre majuscule ayant une variante minuscule.
- Espace : tout espace blanc ou séparateur invisible.

9. Cliquez sur **OK**.

### Importation et exportation de cultures

Outre la création des cultures, vous pouvez importer des cultures que vous avez créés ailleurs et exporter des cultures que vous créez dans Domain Editor.

1. Dans Enterprise Designer, accédez à **Outils > Open Parser Domain Editor**.
2. Cliquez sur l'onglet **Cultures**.
3. Cliquez sur **Importer** ou **Exporter**.
4. Réalisez l'une des actions suivantes :
  - Si vous importez une culture, accédez à une culture et sélectionnez-la. Cliquez sur **Ouvrir**. La culture importée apparaît dans Domain Editor.
  - Si vous exportez une culture, accédez à l'emplacement dans lequel enregistrer la culture exportée et sélectionnez-le. Cliquez sur **Enregistrer**. La culture exportée est enregistrée et Domain Editor réapparaît.

## Domaines

### Ajouter un domaine

Un domaine représente un type de données, tel que les données de nom, d'adresse et de numéro de téléphone. Il s'agit d'un modèle qui représente une séquence d'un ou plusieurs jetons dans vos données d'entrée que vous aurez couramment besoin d'analyser et que vous associez avec une ou plusieurs cultures.

Cette rubrique explique comment ajouter un domaine dans Domain Editor lors de la définition d'une grammaire d'analyse propre à la culture.

1. Dans Enterprise Designer, accédez à **Outils > Open Parser Domain Editor**.
2. Cliquez sur l'onglet **Domaines**.
3. Cliquez sur **Ajouter**.
4. Saisissez un nom de domaine dans le champ **Nom**.
5. Saisissez la description du nom de domaine dans le champ **Description**.
6. Si vous souhaitez créer un nouveau domaine vierge, cliquez sur **OK**. Si vous souhaitez créer un domaine basé sur un autre domaine, procédez comme suit :
  - a) Sélectionnez **Utiliser un autre domaine comme modèle** pour créer un domaine en fonction d'un autre domaine.
  - b) Choisissez un domaine dans la liste. Le nouveau domaine sera créé une fois que vous aurez cliqué sur **OK** lors de l'étape suivante. Le nouveau domaine contiendra toutes les grammaires de parsing spécifiques à une culture définies dans le modèle de domaine sélectionné.
  - c) Cliquez sur **OK**.

### Modification d'un domaine

Un domaine représente un type de données, tel que les données de nom, d'adresse et de numéro de téléphone. Il s'agit d'un modèle qui représente une séquence d'un ou plusieurs jetons dans vos données d'entrée que vous aurez couramment besoin d'analyser et que vous associez avec une ou plusieurs cultures.

Cette rubrique explique comment modifier un domaine.

1. Dans Enterprise Designer, accédez à **Outils > Open Parser Domain Editor**.
2. Cliquez sur l'onglet **Domaines**.
3. Sélectionnez un domaine dans la liste, puis cliquez sur **Modifier**. La boîte de dialogue **Modifier un domaine** apparaît.
4. Modifiez les informations contenues dans la description.

5. Si vous ne souhaitez que modifier la description du domaine, cliquez sur **OK**. Si vous avez effectué des mises à jour du modèle de domaine et que vous voulez à présent ajouter ces changements au domaine que vous êtes en train de modifier, passez à l'étape suivante.
6. Sélectionnez **Utiliser un autre domaine comme modèle** pour hériter des changements réalisés au modèle de domaine.
7. Choisissez une base de modèle de domaine depuis la liste. Quand vous cliquez sur **OK** à l'étape suivante, le modèle de domaine sera modifié. Le modèle de domaine modifié contiendra toutes les grammaires de parsing spécifique à une culture définies dans la base du modèle de domaine que vous avez sélectionné. Toute grammaire de parsing dans le modèle de domaine sélectionné sera remplacée par la grammaire de parsing de la base du modèle de domaine.
8. Cliquez sur **OK**.

Pour voir comment cela fonctionne, procédez comme suit :

1. Créez un modèle de domaine nommé **NameParsing** et définissez les grammaires de parsing pour Culture générale, en, et fr-FR.
2. Créez un modèle de domaine nommé **NameParsing2** et utilisez **NameParsing** comme base de modèle de domaine. **NameParsing2** est créé comme copie exacte et contient des grammaires de parsing pour Culture générale, en, et fr-FR
3. Modifier les grammaires de parsing spécifiques à une culture de **NameParsing** en changeant certaines des règles de grammaire dans la grammaire de Culture générale et ajoutez en-CA comme nouvelle culture.
4. Sélectionnez **NameParsing2** sur l'onglet Domaines, cliquez sur Modifier, et utilisez encore une fois **NameParsing** comme base de modèle de domaine.

Les résultats seront :

- La grammaire de parsing de Culture générale sera mise à jour (ce qui écrasera vos modifications s'il y en a).
- Les cultures en et fr-FR resteront identiques (sauf si elles ont été modifiées dans le domaine cible, auquel cas la version de Name Parsing serait rétablie).
- Une grammaire spécifique à une culture pour en-CA sera ajoutée.

## Supprimer un domaine

Un domaine représente un type de données, tel que les données de nom, d'adresse et de numéro de téléphone. Il s'agit d'un modèle qui représente une séquence d'un ou plusieurs jetons dans vos données d'entrée que vous aurez couramment besoin d'analyser et que vous associez avec une ou plusieurs cultures.

Cette rubrique explique comment supprimer un domaine.

1. Dans Enterprise Designer, accédez à **Outils > Open Parser Domain Editor**.
2. Cliquez sur l'onglet **Domaines**.
3. Sélectionnez un domaine dans la liste.

4. Cliquez sur **Supprimer**.

Si le domaine est associé à une ou plusieurs grammaires d'analyse propres à une culture, un message s'affiche vous invitant à confirmer que vous souhaitez supprimer le domaine. Si aucune grammaire de parsing spécifique à une culture n'est associée à ce domaine, un message s'affiche demandant confirmation de manière à supprimer le domaine sélectionné.

5. Cliquez sur **Oui**. Le domaine ainsi que toutes grammaires de parsing spécifiques à une culture associées à ce domaine sont supprimés.

### Importation et exportation de domaines

Outre la création de domaines, vous pouvez importer des domaines que vous avez créés ailleurs et exporter des domaines que vous créez dans Domain Editor.

1. Cliquez sur l'onglet **Domaines**. L'onglet **Domaines** apparaît.
2. Cliquez sur **Importer** ou **Exporter**.
3. Réalisez l'une des actions suivantes :
  - Si vous importez un domaine, accédez à un nom de domaine et sélectionnez-le. Cliquez sur **Ouvrir**. Le domaine importé apparaît dans Domain Editor.
  - Si vous exportez un domaine, accédez à l'emplacement dans lequel enregistrer le domaine exporté et sélectionnez-le. Cliquez sur **Enregistrer**. Le domaine exporté est enregistré et Domain Editor réapparaît.

## Analyse des résultats de parsing

### Suivi des résultats finaux de parsing

La fonction Suivre les détails d'Open Parser affiche une vue graphique de la manière dont le champ d'entrée a été analysé, jeton par jeton, dans les valeurs de champ de sortie. Le suivi affiche les résultats de correspondance, les résultats de non correspondance et les résultats intermédiaires.

Les résultats finaux de parsing affichent l'arbre de grammaire de parsing et les résultats sortants. Utilisez cet affichage lorsque vous ne voulez voir que les résultats du processus de mise en correspondance. Il s'agit de la vue par défaut.

1. Dans Enterprise Designer, ouvrez le flux de données qui contient le stage Open Parser dont vous souhaitez suivre les résultats d'analyse.
2. Double-cliquez sur le stage Open Parser sur le canevas.
3. Cliquez sur l'onglet **Aperçu**.

4. Entrez des exemples de données à analyser, puis cliquez sur le bouton **Aperçu**.
5. Dans la colonne Suivi, cliquez sur le lien **Cliquez ici...** pour afficher le schéma du suivi.

L'arbre d'exploration de la grammaire de parsing montre un ou plusieurs des composants suivants, en fonction de l'option choisie :

- La variable <root>. Le premier nœud de l'arbre est la variable <root>.
  - Les expressions définies dans la variable <root>. Les nœuds de deuxième niveau sont les expressions définies dans la variable <root>. Les expressions <root> définissent également les noms des champs de sortie.
  - Les définitions des variables des nœuds du deuxième niveau. Les nœuds du troisième niveau et chaque niveau se trouvant en dessous de ce nœud sont les définitions de chaque expression <root>. Les définitions d'expression peuvent être d'autres variables, des alias ou des définitions de règle.
  - Les valeurs et jetons de sortie. Le dernier nœud de l'arbre d'exploration montre les valeurs attribuées à chaque jeton séquentiel dans la grammaire de parsing.
  - La note d'analyse pour les composants pertinents de la grammaire de parsing. Les notes de parsing sont définies du bas d'une expression racine à son sommet. Par exemple, si un modèle d'expression a un coefficient de 80 et qu'une règle parente a un coefficient de 75, le score final de l'expression parente est le produit des notes enfant et des notes parentes, qui seraient de 60 pour cent dans cet exemple.
  - Le caractère espace affiche la zone de texte **Données d'entrée** comme un espace insécable (crochets vers le haut) afin de pouvoir mieux voir les caractères espace. Les délimiteurs ne peuvent pas être utilisés lorsque les jetons sont affichés en gris.
6. Dans le champ **Information**, sélectionnez **Résultat final de l'analyse**.

**Remarque :** Pour consulter les événements, consultez [Accomplissement des événements d'analyse pas à pas](#) à la page 39.

7. Dans la liste **Niveau de détail**, sélectionnez une des options.
  - **Cacher les expressions sans résultats.** Affiche les branches menant à un résultat de correspondance ou de non correspondance. Toute branche d'expression racine ne menant pas à une correspondance est affichée comme une ellipse. Si vous voulez examiner une branche ne menant pas à une correspondance, faites un double-clic sur l'ellipse.
  - **Cacher les expressions racines sans résultats.** Montre toute les branches des expressions racines contenant des résultats de correspondance ou de non correspondance. Toutes les autres expressions racines ne sont pas affichées.
  - **Montrer toutes les racines.** Montre toutes les expressions racines. Si une racine n'a pas de résultat de correspondance, l'affichage s'effondrera pour cette expression racine grâce au symbole d'ellipse.
  - **Afficher toutes les expressions.** Afficher les expressions racine et toutes les branches. Les expressions racine ne s'affichent plus comme une ellipse ; au lieu de cela, les règles pour chaque expression de la branche s'affichent.

Si vous avez un affichage de niveau du détail sélectionné cachant les expressions sans résultats et que vous sélectionnez une expression racine n'étant pas actuellement affichée, Suivre les détails change la sélection au niveau du détail et sélectionne un élément de la liste affichant le nombre minimum d'expressions racines tout en affichant toujours l'expression racine.

8. Cliquez sur **Afficher les scores** pour afficher des scores d'analyse pour les expressions racine, les expressions de variable, et les correspondances et non-correspondances en résultant.
9. Dans le champ **Zoom**, sélectionnez la taille de la vue d'arborescence.
10. Dans le champ **Clause racine**, choisissez l'une des options afin d'afficher les branches de cet arbre d'expression racine.

Lorsque vous cliquez sur une branche d'expression dans le schéma de trace, la liste **Clause racine** se met à jour pour afficher la clause sélectionnée. Double-cliquez sur une ellipse pour afficher une expression réduite.

11. Cliquez sur **OK** une fois que vous avez terminé. Les réglages du niveau du détail, de l'affichage des scores et du contrôle de zoom sont enregistrés lorsque vous cliquez sur **OK**.

## Accomplissement des événements d'analyse pas à pas

La vue Suivre les détails d'Open Parser vous permet d'afficher un graphique des étapes événement par événement dans le processus de mise en correspondance. Utilisez cet affichage lorsque vous dépannez le processus de mise en correspondance et souhaitez voir comment chaque jeton est évalué, explorer la jetonisation de grammaire de parsing et les résultats de correspondance jeton par jeton.

1. Dans Enterprise Designer, ouvrez le flux de données qui contient le stage Open Parser dont vous souhaitez suivre les résultats d'analyse.
2. Double-cliquez sur le stage Open Parser sur le canevas.
3. Cliquez sur l'onglet **Aperçu**.
4. Entrez des exemples de données à analyser, puis cliquez sur le bouton **Aperçu**.
5. Dans la colonne Suivi, cliquez sur le lien **Cliquez ici...** pour afficher le schéma du suivi.

L'arbre d'exploration de la grammaire de parsing montre un ou plusieurs des composants suivants, en fonction de l'option choisie :

- La variable <root>. Le premier nœud de l'arbre est la variable <root>.
- Les expressions définies dans la variable <root>. Les nœuds de deuxième niveau sont les expressions définies dans la variable <root>. Les expressions <root> définissent également les noms des champs de sortie.
- Les définitions des variables des nœuds du deuxième niveau. Les nœuds du troisième niveau et chaque niveau se trouvant en dessous de ce nœud sont les définitions de chaque expression <root>. Les définitions d'expression peuvent être d'autres variables, des alias ou des définitions de règle.



- Les valeurs et jetons de sortie. Le dernier nœud de l'arbre d'exploration montre les valeurs attribuées à chaque jeton séquentiel dans la grammaire de parsing.
  - La note d'analyse pour les composants pertinents de la grammaire de parsing. Les notes de parsing sont définies du bas d'une expression racine à son sommet. Par exemple, si un modèle d'expression a un coefficient de 80 et qu'une règle parente a un coefficient de 75, le score final de l'expression parente est le produit des notes enfant et des notes parentes, qui seraient de 60 pour cent dans cet exemple.
  - Le caractère espace affiche la zone de texte **Données d'entrée** comme un espace insécable (crochets vers le haut) afin de pouvoir mieux voir les caractères espace. Les délimiteurs ne peuvent pas être utilisés lorsque les jetons sont affichés en gris.
6. Les correspondances et non correspondances sont coloriées selon un code couleur dans le graphique de suivi :
- Les cases vertes désignent les correspondances faisant partie du résultat validé.
  - Les cases rouges désignent les non correspondances.
  - Les cases jaunes désignent les correspondances intermédiaires qui seront finalement abandonnées au fur et à mesure que les événements progressent. Les correspondances intermédiaires ne s'affichent que dans les Événements de parsing étape par étape.
  - Les cases grises désignent des correspondances intermédiaires abandonnées afin de libérer ce jeton pour l'utiliser avec une autre expression. Les correspondances intermédiaires ne s'affichent que dans les Événements de parsing étape par étape.
7. Dans la liste **Informations**, sélectionnez **Événements de parsing étape par étape**.
8. Dans la liste **Niveau de détail**, sélectionnez une des options.
- **Cacher les expressions sans résultats.** Affiche les branches menant à un résultat de correspondance ou de non correspondance. Toute branche d'expression racine ne menant pas à une correspondance est affichée comme une ellipse. Si vous voulez examiner une branche ne menant pas à une correspondance, faites un double-clic sur l'ellipse.
  - **Cacher les expressions racines sans résultats.** Montre toute les branches des expressions racines contenant des résultats de correspondance ou de non correspondance. Toutes les autres expressions racines ne sont pas affichées.
  - **Montrer toutes les racines.** Montre toutes les expressions racines. Si une racine n'a pas de résultat de correspondance, l'affichage s'effondrera pour cette expression racine grâce au symbole d'ellipse.
  - **Afficher toutes les expressions.** Afficher les expressions racine et toutes les branches. Les expressions racine ne s'affichent plus comme une ellipse ; au lieu de cela, les règles pour chaque expression de la branche s'affichent.
- Si vous avez un affichage de niveau du détail sélectionné cachant les expressions sans résultats et que vous sélectionnez une expression racine n'étant pas actuellement affichée, Suivre les détails change la sélection au niveau du détail et sélectionne un élément de la liste affichant le nombre minimum d'expressions racines tout en affichant toujours l'expression racine.
9. Cliquez sur **Afficher les scores** pour afficher des scores d'analyse pour les expressions racine, les expressions de variable, et les correspondances et non-correspondances en résultant.



10. Dans le champ **Zoom**, sélectionnez la taille de la vue d'arborescence.
11. Dans le champ **Clause racine**, choisissez l'une des options afin d'afficher les branches de cet arbre d'expression racine.  
  
Lorsque vous cliquez sur une branche d'expression dans le schéma de trace, la liste **Clause racine** se met à jour pour afficher la clause sélectionnée. Double-cliquez sur une ellipse pour afficher une expression réduite.
12. La case **Aller automatiquement au nœud sélectionné** est cochée par défaut. Lorsque celle-ci est sélectionnée et que vous cliquez sur le bouton **Lecture**, les événements s'exécutent du début et s'arrêtent au premier événement se produisant au niveau du nœud sélectionné ou de celui d'un de ses enfants. Pour lire tous les événements sans interruption, décochez cette case avant de cliquer sur le bouton **Lecture**.
13. Dans le champ **Délai avant lecture (en secondes)**, indiquez un délai pour contrôler la vitesse du débit de lecture.
14. Cliquez sur le bouton **Lecture** pour commencer à exécuter les événements de parsing.
15. Cliquez sur **OK** une fois que vous avez terminé.

## Parser des noms de personne

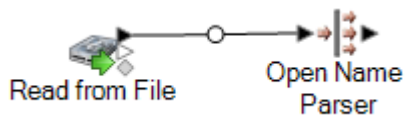
Si vous disposez de données de nom, figurant toutes dans un champ, il serait judicieux d'analyser le nom dans des champs distincts correspondant aux différentes parties du nom, telles que le prénom, le nom de famille, la civilité, etc. Ces éléments nom analysés peuvent ensuite être utilisés par d'autres opérations automatisées telles que le rapprochement par nom, la standardisation de noms ou la consolidation de noms enregistrés plusieurs fois.

1. Si ce n'est pas déjà fait, chargez les tables suivantes sur le serveur Spectrum™ Technology Platform :
  - Base Open Parser
  - Noms améliorés Open Parser

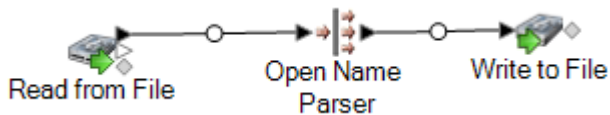
Utilisez l'utilitaire de chargement de la base de données du module Data Normalization pour charger ces tables. Pour en savoir plus sur le chargement de tables, voir le *Guide d'installation*.

2. Dans Enterprise Designer, créez un flux de données.
3. Faites glisser un stage source sur le canevas.
4. Double-cliquez sur le stage source et configurez-le. Pour obtenir les instructions sur la configuration des stages source, reportez-vous au *Guide du concepteur du flux de données*.
5. Faites glisser un stage Open Name Parser sur le canevas et connectez-le au stage source.

Par exemple, si vous utilisez un stage Read from File, votre flux de données se présente comme suit :



6. Faites glisser un stage de collecteur de données et connectez-le au stage Open Name Parser. Par exemple, si vous utilisez un collecteur de données Write to File, votre flux de données se présente comme suit :



7. Double-cliquez sur le stage de collecteur de données et configurez-le. Pour obtenir les instructions sur la configuration des stages source, reportez-vous au *Guide du concepteur du flux de données*.

Vous avez créé un flux de données capable d'analyser les noms personnels en des parties de composant, plaçant chaque partie du nom dans son propre champ.

## Modèles de dataflow pour le parsing

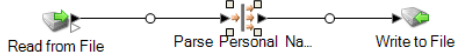
### Analyse de noms anglais

Ce modèle de flux de données montre comment prendre des données de nom de personne (par exemple, « John P. Smith »), les décomposer en prénom, deuxième prénom et nom de famille et leur ajouter des données de sexe.

#### *Scénario commercial*

Vous travaillez pour une compagnie d'assurance qui désire envoyer des devis personnalisés basés sur le sexe des clients à prospecter. Vos données d'entrée incluent des données de nom comme des noms entiers et vous désirez parser les données de nom en champs de prénom, second prénom et nom de famille. Vous souhaitez également déterminer le sexe des individus de vos données d'entrée.

Les flux de données suivants apportent une solution au scénario commercial :



Ce modèle de flux de données est disponible dans Enterprise Designer. Accédez à **Fichier > Nouveau > Flux de données > À partir d'un modèle** et sélectionnez **Analyser les noms personnels**.

Ce flux de données requiert les éléments suivants :

- Le module Universal Name
- Les tables de base Open Parser
- Les tables de noms améliorées Open Parser

Dans ce flux de données, les données sont lues dans un fichier et traitées par le stage Open Name Parser. (Open Name Parser fait partie du module Universal Naming.) Pour chaque nom, le flux de données procède comme suit :

#### *Read from File*

Ce stage identifie le nom du fichier, l'emplacement et la disposition du fichier contenant les noms que vous désirez parser. Le fichier contient à la fois des noms féminins et masculins.

#### *Open Name Parser*

Le stage Open Name Parser examine les champs de nom et les compare aux données de nom stockées dans les fichiers de la base de données de noms Spectrum™ Technology Platform. Suivant les résultats de la comparaison, il répartit les données de nom entre les champs Prénom, Deuxième prénom et Nom de famille.

#### *Write to File*

Le modèle comporte un stage Write to File. Outre les champs d'entrée, le fichier de sortie contient les champs FirstName, MiddleName, LastName, EntityType, GenderCode et GenderDeterminationSource.

## Parser des noms arabes

Ce modèle décrit comment parser des noms arabes européens en éléments de composants. La règle d'analyse sépare chaque jeton du champ **Nom** et copie chacun dans cinq champs : **Kunya**, **Ism**, **Laqab**, **Nasab** et **Nisba**. Ces champs de sortie représentent les cinq parties d'un nom arabe ; ils sont décrits dans le scénario commercial.

### Scénario commercial

Vous travaillez pour une banque qui désire mieux comprendre le système des noms arabes dans un effort pour améliorer son service client avec la clientèle de langue arabe. Vous avez reçu des plaintes de clients dont les informations de facturation ne faisaient pas figurer leur nom correctement. Dans un souci de connaissance plus personnelle de ses clients, le groupe marketing pour lequel vous travaillez souhaite adopter une meilleure approche de ses clients de langue arabe par le biais de campagnes marketing et d'une assistance téléphonique.

Pour pouvoir comprendre le système des noms arabes, vous trouverez les ressources expliquant comment les noms arabes sont formés sur Internet :

- [en.wikipedia.org/wiki/Arabic\\_names](https://en.wikipedia.org/wiki/Arabic_names)
- [heraldry.sca.org/laurel/names/arabic-naming2.htm](https://heraldry.sca.org/laurel/names/arabic-naming2.htm)

Les noms arabes sont basés sur un système de nommage qui inclut les cinq parties de nom suivantes : Ism, Kunya, Nasab, Laqab et Nisba.

- Ism est le nom principal, ou nom personnel d'une personne arabe.
- Souvent, kunya se rapportant au premier fils de la personne est utilisé à la place de ism.
- Nasab est un patronyme, ou une série de patronymes. Il indique le rang de naissance de la personne par le mot ibn ou bin, qui signifie fils de, et bint, qui signifie fille de.
- Laqab s'entend comme une description de la personne. Par exemple, al-Rashid signifie le vertueux ou celui qui est guidé vers le droit chemin, et al-Jamil signifie le magnifique.
- Nisba décrit l'activité d'une personne, sa zone géographique de résidence, ou son ascendance (tribu, famille, et ainsi de suite). Il suit une famille sur plusieurs générations. Nisba, parmi les composants d'un nom arabe, est peut-être celui qui s'approche le plus d'un nom de famille Européen. Par exemple, al-Filistin signifie le palestinien.

Les flux de données suivants apportent une solution au scénario commercial :



Ce modèle de flux de données est disponible dans Enterprise Designer. Accédez à **Fichier > Nouveau > Flux de données > À partir d'un modèle** et sélectionnez **ParseArabicNames**. Ce flux de données requiert le module Data Normalization.

Dans ce flux de données, les données sont lues dans un fichier et traitées par le stage Open Parser. Pour chaque ligne de données du fichier d'entrée, ce flux de données effectue les opérations suivantes :

#### Read from File

Ce stage identifie le nom du fichier, l'emplacement et la disposition du fichier contenant les noms que vous désirez parser. Le fichier contient à la fois des noms féminins et masculins.

## Open Parser

Ce stage définit s'il faut utiliser une grammaire spécifique à une culture créée dans Domain Editor ou définir une grammaire indépendante du domaine. Une grammaire de parsing spécifique à une culture, que vous pouvez créer dans Domain Editor, est une grammaire de parsing validée associée à une culture et à un domaine. Une grammaire de parsing indépendante d'un domaine, que vous pouvez créer dans l'Open Parser, est une grammaire de parsing validée associée à une culture et à un domaine.

Dans ce modèle, la grammaire de parsing est définie comme grammaire indépendante du domaine.

Le stage Open Parser contient une grammaire de parsing définissant les commandes et expressions suivantes :

- `%Tokenize` est défini sur le caractère d'espace (`\s`). Cela signifie que Open Parser utilisera le caractère espace pour séparer les champs d'entrée en jetons. Par exemple, Abu Mohammed al-Rahim ibn Salamah contient cinq jetons : Abu, Mohammed, al-Rahim, ibn et Salamah.
- `%InputField` est défini de sorte à analyser les données d'entrée du champ **Name**.
- `%OutputFields` est défini de sorte à copier les données analysées dans cinq champs : **Kunya**, **Ism**, **Laqab**, **Nasab** et **Nisba**.
- L'expression `<root>` définit le modèle des noms arabes :
  - Zéro ou une occurrence de **Kunya**
  - Exactement une ou deux occurrences de **Ism**
  - Zéro ou une occurrence de **Laqab**
  - Zéro ou une occurrence de **Nasab**
  - Zéro ou plusieurs occurrences de **Nisba**

Les variables de règle qui définissent le domaine doivent utiliser les mêmes noms que les champs de sortie définis dans la commande `OutputFields` requise.

La grammaire de parsing utilise une combinaison d'expressions régulières et d'expressions de quantificateurs pour construire un modèle pour les noms arabes. La grammaire de parsing utilise ces caractères spéciaux :

- Le caractère « ? » signifie qu'une expression régulière peut se produire zéro ou une fois.
- Le caractère « \* » signifie qu'une expression régulière peut se produire zéro fois ou plus.
- Le caractère « ; » signifie la fin d'une règle.

Utilisez l'onglet **Commandes** pour explorer la signification des autres symboles spéciaux que vous pouvez utiliser dans les grammaires d'analyse en survolant la description avec la souris.

Les quantificateurs sont avides par défaut. Avide signifie que l'expression accepte le plus grand nombre possible de jetons tout en autorisant une mise en correspondance valide. Vous pouvez outrepasser ce comportement en attachant un « ? » pour la mise en correspondance réticente ou un « + » pour la mise en correspondance possessive. Réticent signifie que l'expression accepte le moins de jetons possibles tout en autorisant une mise en correspondance valide. Possessif signifie que l'expression accepte le plus grand nombre de jetons possibles, même si cela empêche une mise en correspondance.

Pour tester la grammaire d'analyse, cliquez sur l'onglet **Aperçu**. Saisissez les noms affichés ci-dessous dans le champ **Nom**, puis cliquez sur **Aperçu**

Name	Kunya	Ism	Laqab	Nasab	Nisba
Abu Karim Muhammad al-Jamil ibn Nidal ibn Abdulaziz al-Filistini	Abu Karim	Muhammad	al-Jamil	ibn Nidal ibn Abdulaziz	al-Filistini
Layla bint Zuhayr ibn Yazid al-Nahdiyah		Layla		bint Zuhayr ibn Yazid	al-Nahdiyah
Yazid ibn Abi Hakim		Yazid		ibn Abi Hakim	
Abu Bishr al-Yaman ibn Abi al-Yaman al-Bandaniji	Abu Bishr	al-Yaman		ibn Abi	al-Yaman al-Bandaniji
Abu al-Tayyib 'Abd al-Rahim ibn Ahmad al-Harrani	Abu al-Tayyib	'Abd	al-Rahim	ibn Ahmad	al-Harrani
Ahmad ibn Sa'id al-Bahali		Ahmad		ibn Sa'id	al-Bahali
Abu al-Abbas Muhammad ibn Ya'qub ibn Yusuf al-Asamm al-Naysaburi	Abu al-Abbas	Muhammad		ibn Ya'qub ibn Yusuf	al-Asamm al-Naysaburi
Abu al-Qasim Mansur ibn al-Zabriqan ibn Salamah al-Namari	Abu al-Qasim	Mansur		ibn al-Zabriqan ibn Salamah	al-Namari
'Ubayd ibn Mu'awiyah ibn Zayd ibn Thabit ibn al-Dahhak		'Ubayd		ibn Mu'awiyah ibn Zayd ibn Thabit ibn al-Dahhak	
Umm Ja'far Zubaydah	Umm Ja'far	Zubaydah			

Vous pouvez aussi saisir d'autres noms valides et non valides pour observer comment les données d'entrée sont parsées.

Vous pouvez utiliser la fonctionnalité Suivi pour voir une représentation graphique soit du résultat de parsing final, soit étape par étape au cours des événements de parsing. Cliquez sur le lien de la colonne **Suivre** pour afficher Suivre les détails de la ligne de données.

### Write to File

Le modèle comporte un stage Write to File. Outre le champ d'entrée, le fichier de sortie contient les champs **Kunya**, **Ism**, **Laqab**, **Nasab** et **Nisba**.

## Parser des noms chinois

Ce modèle expose comment parser les noms chinois en éléments de composants. La règle d'analyse sépare chaque jeton du champ **Name** et copie chacun dans deux champs : **LastName** et **FirstName**.

### Scénario commercial

Vous travaillez pour une société de services financiers qui désire savoir s'il est possible d'inclure des caractères chinois pour ses clients de langue chinoise dans diverses correspondances.

Pour pouvoir comprendre le système des noms chinois, vous trouverez les ressources expliquant comment les noms chinois sont formés :

[en.wikipedia.org/wiki/Chinese\\_names](http://en.wikipedia.org/wiki/Chinese_names)

Les flux de données suivants apportent une solution au scénario commercial :



Ce modèle de flux de données est disponible dans Enterprise Designer. Accédez à **Fichier > Nouveau > Flux de données > À partir d'un modèle** et sélectionnez **ParseChineseNames**. Ce flux de données requiert le module Data Normalization.

Dans ce flux de données, les données sont lues dans un fichier et traitées par le stage Open Parser. Pour chaque ligne de données dans le fichier d'entrée, ce flux de données effectuera ce qui suit :

### *Read from File*

Ce stage identifie le nom du fichier, l'emplacement et la disposition du fichier contenant les noms que vous désirez parser. Le fichier contient à la fois des noms féminins et masculins.

### *Open Parser*

Ce stage définit s'il faut utiliser une grammaire spécifique à une culture créée dans Domain Editor ou définir une grammaire indépendante du domaine. Une grammaire de parsing spécifique à une culture, que vous pouvez créer dans Domain Editor, est une grammaire de parsing validée associée à une culture et à un domaine. Une grammaire de parsing indépendante d'un domaine, que vous pouvez créer dans l'Open Parser, est une grammaire de parsing validée associée à une culture et à un domaine.

Dans ce modèle, la grammaire de parsing est définie comme grammaire indépendante du domaine.

Le stage Open Parser contient une grammaire de parsing définissant les commandes et expressions suivantes :

- `%Tokenize` est défini sur `None`. Lorsque `Tokenize` est défini sur `None`, la règle de grammaire de parsing doit comprendre tout espace ou autre séparateur de jetons dans sa définition.
- `%InputField` est défini de sorte à analyser les données d'entrée du champ **Name**.
- `%OutputFields` est défini de sorte à copier les données analysées dans deux champs : **LastName** et **FirstName**.

L'expression `<root>` définit le modèle des noms chinois :

- Une occurrence de **LastName**
- Une à trois occurrences de **FirstName**

Les variables de règle qui définissent le domaine doivent utiliser les mêmes noms que les champs de sortie définis dans la commande `OutputFields` requise.

La variable de règle `CJKCharacter` définit le modèle des caractères pour le chinois/japonais/coréen (CJK). Le modèle du caractère est défini de façon à n'utiliser que des caractères qui sont des lettres. La règle est :

```
<CJKCharacter> = @RegEx("([\p{InCJKUnifiedIdeographs}&&\p{L}])");
```

- L'expression régulière `\p{InX}` sert à indiquer un bloc Unicode dans une culture donnée, où `X` est la culture. Dans cette instance, la culture est `CJKUnifiedIdeographs`.
- Dans les expressions régulières, une classe de caractères sont un ensemble de caractères que vous désirez faire correspondre. Par exemple, `[aeiou]` est la classe de caractères ne contenant que des voyelles. Les classes de caractères peuvent apparaître dans d'autres classes de caractères, et peuvent être composées par l'opérateur d'union (implicite) et l'opérateur d'intersection (`&&`). L'opérateur d'union dénote une classe contenant chaque caractère se trouvant dans au

moins une des ses classes d'opérande. L'opérateur d'intersection dénote une classe contenant chaque caractère recouvrant les blocs Unicode intersectés.

- L'expression régulière `\p{L}` sert à indiquer le bloc Unicode n'incluant que des lettres.

Pour tester la grammaire d'analyse, cliquez sur l'onglet **Aperçu**. Saisissez les noms affichés ci-dessous dans le champ **Nom**, puis cliquez sur **Aperçu**

Name	FirstName	LastName
王若琳	若琳	王
刘耕宏	耕宏	刘
许惠欣	惠欣	许
蔡依林	依林	蔡
水沫	沫	水
蕭亞軒	亞軒	蕭
郑元畅	元畅	郑
林依晨	依晨	林

Vous pouvez aussi saisir d'autres noms valides et non valides pour observer comment les données d'entrée sont parsées.

Vous pouvez utiliser la fonctionnalité **Suivi** pour voir une représentation graphique soit du résultat de parsing final, soit étape par étape au cours des événements de parsing. Cliquez sur le lien de la colonne **Suivre** pour afficher Suivre les détails de la ligne de données.

### *Write to File*

Le modèle comporte un stage **Write to File**. Outre le champ d'entrée, le fichier de sortie contient les champs **LastName** et **FirstName**. Choisissez les résultats d'une correspondance depuis la Liste des résultats de correspondance et cliquez sur **Supprimer**.

## Parsing de noms espagnols ou allemands

Ce modèle expose comment parser des noms de diverses cultures, comme les noms espagnols ou allemands, en éléments de composants. La règle d'analyse sépare chaque jeton du champ **Nom** et copie chacun dans les champs définis dans la grammaire d'analyse des noms de personnes et d'entreprises. Pour plus d'informations sur cette grammaire d'analyse, sélectionnez **Outils > Open Parser Domain Editor**, puis le domaine **Personal and Business Names** et soit la culture **Allemand (de)**, soit la culture **Espagnol (es)**.

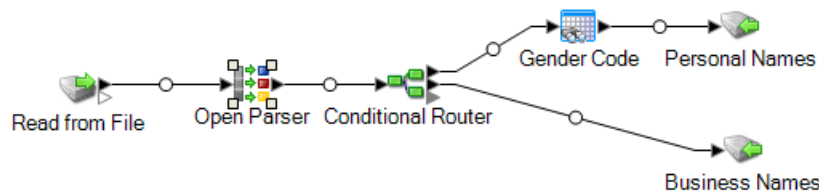
Ce modèle applique aussi des codes de genre aux noms de personnes en utilisant la table de données contenue dans **Table Management**. Pour plus d'informations sur **Table Management**, sélectionnez **Outils > Table Management**.

### *Scénario commercial*

Vous travaillez pour une société pharmaceutique basée à Bruxelles qui a consolidé ses opérations en Allemagne et en Espagne. Votre société désire déployer une base de données de cultures mixtes contenant des données de nom, et votre mission est d'analyser les variations de nom entre les deux cultures.

Les flux de données suivants apportent une solution au scénario commercial :





Ce modèle de flux de données est disponible dans Enterprise Designer. Accédez à **Fichier > Nouveau > Flux de données > À partir d'un modèle** et sélectionnez **ParseSpanish&GermanNames**. Ce flux de données requiert le module Data Normalization.

Dans ce flux de données, les données sont lues dans un fichier et traitées par le stage Open Parser. Pour chaque ligne de données dans le fichier d'entrée, ce flux de données effectuera ce qui suit :

#### *Read from File*

Ce stage identifie le nom du fichier, l'emplacement et la disposition du fichier contenant les noms que vous désirez parser. Le fichier contient à la fois des noms masculins et féminins et inclut des informations CultureCode pour chaque nom. Les informations CultureCode désignent les noms d'entrée soit allemands (de), soit espagnols (es).

#### *Open Name Parser*

Le stage Open Name Parser examine les champs de nom et les compare aux données de nom stockées dans les fichiers de la base de données de noms Spectrum™ Technology Platform. Suivant les résultats de la comparaison, il répartit les données de nom entre les champs Prénom, Deuxième prénom et Nom de famille.

#### *Conditional Router*

Ce stage route l'entrée pour que les noms de personnes soient routés vers le stage Gender Codes et que les noms commerciaux soient routés vers le stage Business Names.

#### *Code de genre*

Faites un double clic sur ce stage du canevas, puis cliquez sur **Modifier** pour afficher les Options de règle de Table Lookup.

L'option **Catégoriser** utilise la valeur Source comme clé et copie la valeur correspondante de l'entrée de table dans le champ sélectionné dans la liste Destination. Dans ce modèle, **Compléter le champ** est sélectionné et **Source** est défini de sorte à utiliser le champ **FirstName**. Table Lookup considère le champ entier comme une seule chaîne et marque l'enregistrement si la chaîne peut être catégorisée dans son entier.

**Destination** est défini sur le champ **GenderCode** et utilise les termes de recherche contenus dans la table **Code sexe** pour catégoriser les noms masculins et féminins. Si un terme ne se trouve pas dans les données d'entrée, Table Lookup assigne une valeur **U**, qui signifie inconnu. Pour mieux comprendre comment cela fonctionne, sélectionnez **Outils > Gestion des tables**, puis la table Code sexe.

### Write to File

Le modèle contient deux stages Write to File, l'un pour les noms de personnes et l'autre pour les noms d'entreprises. Outre le champ d'entrée, le fichier de sortie des noms de personnes contient les champs **Name**, **TitleOfRespect**, **FirstName**, **MiddleName**, **LastName**, **PaternalLastName**, **MaternalLastName**, **MaturitySuffix**, **GenderCode**, **CultureUsed** et **ParserScore**.

Le fichier de sortie des noms d'entreprises, quant à lui, contient les champs **Name**, **FirmName**, **FirmSuffix**, **CultureUsed** et **ParserScore**.

## Parsing d'adresses de courrier électronique

Ce modèle expose comment parser des adresses de courrier électronique en composants. La règle d'analyse sépare chaque jeton du champ **E-mail** et copie chacun dans trois champs : **Local-Part**, **DomainName** et **DomainExtension**. **Local-Part** représente la partie nom de domaine de l'adresse électronique, **DomainName** représente le nom de domaine de l'adresse électronique et **DomainExtension** représente l'extension de domaine de l'adresse électronique. Par exemple, dans `pb.com`, « pb » est le nom de domaine et « com » est l'extension de domaine.

Internet est une grande source d'informations du domaine public qui peut vous aider dans vos opérations d'open parsing. Dans cet exemple, les informations de formatage de courrier électronique ont été obtenues par diverses ressources Internet et ont été ensuite importées dans Table Management pour créer une table de valeurs de domaine. L'opération d'extension de domaine que vous réaliserez dans ce modèle expose l'utilité de cette méthode.

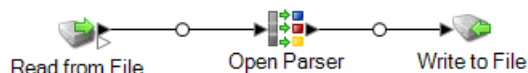
Ce modèle démontre également comment utiliser efficacement une table de données que vous chargez dans Table Management pour effectuer des recherches de table dans vos opérations de parsing.

### Scénario commercial

Vous travaillez pour une compagnie d'assurance qui désire faire sa première campagne commerciale par courrier électronique. Votre base de données contient les adresses électroniques de vos clients et on vous a demandé de trouver le moyen de vous assurer que ces adresses électroniques ont un format SMTP valide.

Avant de créer ce flux de données, vous devrez charger une table d'extensions de noms de domaine valides dans Table Management pour pouvoir y consulter les extensions de noms de domaine faisant partie du processus de validation.

Les flux de données suivants apportent une solution au scénario commercial :



Ce modèle de flux de données est disponible dans Enterprise Designer. Accédez à **Fichier > Nouveau > Flux de données > À partir d'un modèle** et sélectionnez **ParseEmail**. Ce flux de données requiert le module Data Normalization.

Dans ce flux de données, les données sont lues dans un fichier et traitées par le stage Open Parser. Pour chaque ligne de données du fichier d'entrée, ce flux de données effectue les opérations suivantes :

### *Créer une table d'extension de domaine*

La première opération est de créer une table Open Parser dans Table Management que vous pourrez utiliser pour vérifier si les extensions de domaine de vos adresses électronique sont valides.

1. Dans le menu **Outils**, sélectionnez **Gestion des tables**.
2. Dans la liste **Type**, sélectionnez **Open Parser**.
3. Cliquez sur **Nouveau**.
4. Dans la boîte de dialogue **Ajouter une table définie par l'utilisateur**, saisissez `EmailDomains` dans le champ **Table Name**, vérifiez que **None** est sélectionné dans la liste **Copier à partir de**, puis cliquez sur **OK**.
5. Lorsque **EmailDomains** apparaît dans la liste **Nom**, cliquez sur **Importer**.
6. Dans la boîte de dialogue **Importer**, cliquez sur **Parcourir** et recherchez le fichier source de la table. Son emplacement par défaut est le suivant : `<drive>:\Program Files\Pitney Bowes\Spectrum\server\modules\coretemplates\data\Email_Domains.txt`. Table Management affiche un aperçu des termes contenus dans le fichier d'importation.
7. Cliquez sur **OK**. Table Management importe les fichiers sources et affiche une liste d'extensions de domaine Internet.
8. Cliquez sur **Fermer**. La table `EmailDomains` est créée. Créez maintenant le flux de données à l'aide du modèle `ParseEmail`.

### *Read from File*

Ce stage identifie le nom de fichier, l'emplacement et la disposition du fichier contenant les adresses électroniques que vous souhaitez analyser.

### *Open Parser*

La grammaire de parsing du stage Open Parser définit les commandes et expressions suivantes :

- `%Tokenize` est défini sur `None`. Lorsque `Tokenize` est défini sur `None`, la règle de grammaire de parsing doit comprendre tout espace ou autre séparateur de jetons dans sa définition.
- `%InputField` est défini de sorte à analyser les données d'entrée du champ **Email\_Address**.
- `%OutputFields` est défini de sorte à copier les données analysées dans trois champs : **Local-Part**, **DomainName** et **DomainExtension**.
- L'expression racine définit le modèle des jetons ayant été parsés :

```
<root> = <Local-Part>"@"<DomainName>". "<DomainExtension>;
```

Les variables de règle qui définissent le domaine doivent utiliser les mêmes noms que les champs de sortie définis dans la commande `OutputFields` requise.

- Ce qui reste de la grammaire de parsing définit chacune des variables de règle comme des expressions.

```
<Local-Part> = (<alphanumeric> ".")* <alphanumeric> | (<alphanumeric> "_")*
<alphanumeric> ;
<DomainName> = (<alphanumeric> ".")? <alphanumeric>;
<DomainExtension> = @Table("EmailDomains")* "."? @Table("EmailDomains");
<alphanumeric>=@Regex("[A-Za-z0-9]+");
```

La variable `<Local-Part>` est définie comme une chaîne de texte contenant la variable `<alphanumeric>`, le caractère de point et une autre variable `<alphanumeric>`.

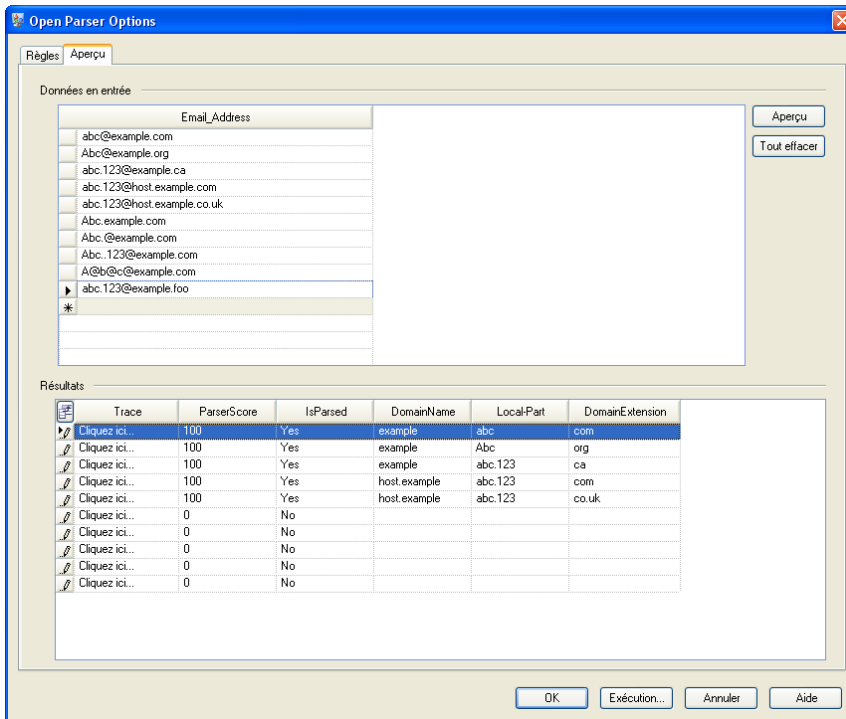
La définition de la variable `<alphanumeric>` est une expression régulière signifiant toute chaîne de caractères de A à Z, de a à a et de 0 à 9. La variable `<alphanumeric>` est utilisée tout au long de cette grammaire d'analyse et définie une seule fois sur la dernière ligne de la grammaire d'analyse.

La grammaire de parsing utilise une combinaison d'expressions régulières et de caractères libéraux pour construire un modèle pour les adresses électroniques. Tout caractère entre guillemets dans cette grammaire de parsing sont des caractères littéraux, le noms d'une table servant à la recherche, ou une expression régulière. La grammaire de parsing utilise ces caractères spéciaux :

- Le caractère « + » signifie qu'une expression régulière peut se produire une ou plusieurs fois.
- Le caractère « ? » signifie qu'une expression régulière peut se produire zéro ou une fois.
- Le caractère « | » signifie que la variable a une condition OR.
- Le caractère « ; » signifie la fin d'une règle.

Utilisez l'onglet **Commandes** pour explorer la signification des autres symboles spéciaux que vous pouvez utiliser dans les grammaires d'analyse en survolant la description avec la souris.

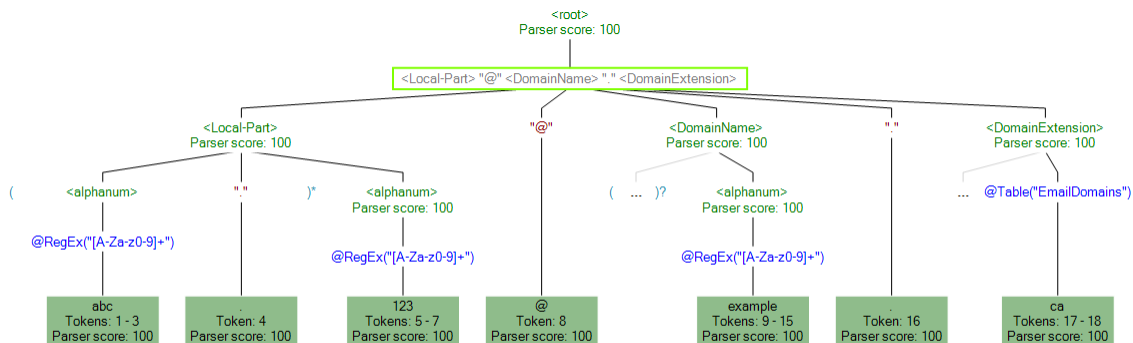
Pour tester la grammaire d'analyse, cliquez sur l'onglet **Aperçu**. Saisissez les adresses électroniques affichées ci-dessous dans le champ **Email Address**, puis cliquez sur **Aperçu**.



Vous pouvez aussi saisir d'autres adresses de courriel pour observer comment les données d'entrée sont parsées.

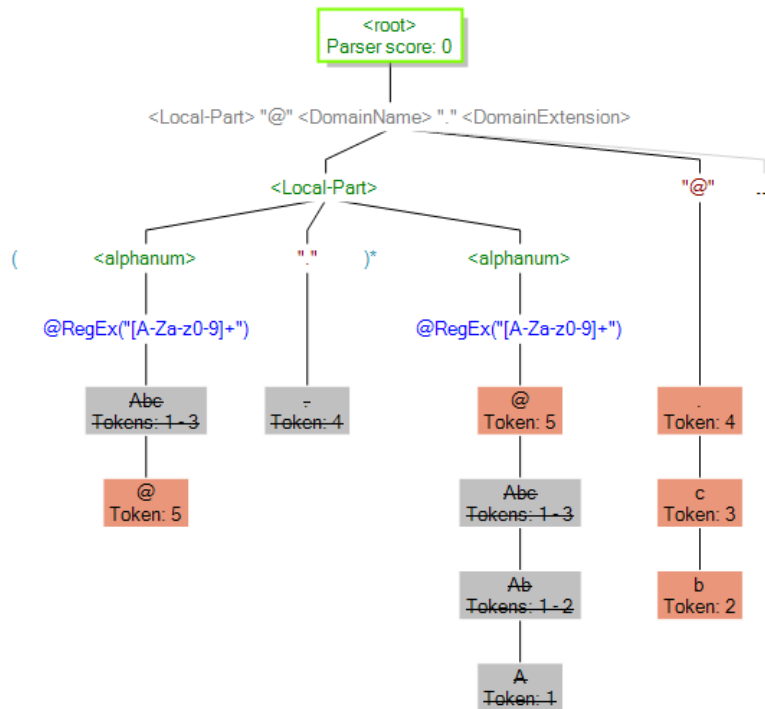
Vous pouvez aussi utiliser la fonctionnalité Suivi pour voir une représentation graphique soit du résultat de parsing final, soit étape par étape au cours des événements de parsing. Cliquez sur le lien de la colonne **Suivre** pour afficher Suivre les détails de la ligne de données.

Détails du suivi indique un résultat de correspondance. Comparez les jetons en correspondance pour chaque expression de la grammaire de parsing.



Vous pouvez aussi utiliser Aperçu pour afficher les résultats ne correspondant pas. Le graphique suivant montre un résultat non correspondant. Comparez les jetons en correspondance pour chaque expression de la grammaire de parsing. La raison pour laquelle ces données d'entrée

(Abc.example.com) ne correspondaient pas est qu'elles ne contenaient pas tous les jetons requis pour correspondre : il n'existe pas de caractère @ séparant le jeton Local- Part des jetons Domain.



*Write to File*

Le modèle comporte un stage Write to File. Outre le champ d'entrée, le fichier de sortie contient les champs **Local-Part**, **DomainName**, **DomainExtension**, **IsParsed** et **ParserScore**.

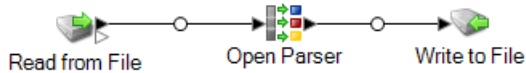
## Parsing pour les États Unis Numéros de téléphone

Ce modèle montre comment analyser les numéros de téléphone aux États-Unis et les diviser en parties de composant. La règle d'analyse sépare chaque jeton du champ **PhoneNumber** et copie chacun dans quatre champs : **CountryCode**, **AreaCode**, **Exchange** et **Number**.

*Scénario commercial*

Vous travaillez pour un fournisseur d'accès sans fil et un projet analysant les données de numéros de téléphone entrants vous a été assigné pour une région de votre entreprise en pleine expansion.

Les flux de données suivants apportent une solution au scénario commercial :



Ce modèle de flux de données est disponible dans Enterprise Designer. Accédez à **Fichier > Nouveau > Flux de données > À partir d'un modèle** et sélectionnez **ParseUSPhoneNumbers**. Ce flux de données requiert le module Data Normalization.

Dans ce flux de données, les données sont lues dans un fichier et traitées par le stage Open Parser. Pour chaque ligne de données dans le fichier d'entrée, ce flux de données effectuera ce qui suit :

### *Read from File*

Ce stage identifie le nom de fichier, l'emplacement et la disposition du fichier contenant les numéros de téléphone que vous désirez analyser.

### *Open Parser*

Ce stage définit s'il faut utiliser une grammaire spécifique à une culture créée dans Domain Editor ou définir une grammaire indépendante du domaine. Une grammaire de parsing spécifique à une culture, que vous pouvez créer dans Domain Editor, est une grammaire de parsing validée associée à une culture et à un domaine. Une grammaire de parsing indépendante d'un domaine, que vous pouvez créer dans l'Open Parser, est une grammaire de parsing validée associée à une culture et à un domaine.

Dans ce modèle, la grammaire de parsing est définie comme grammaire indépendante du domaine.

Le stage Open Parser contient une grammaire de parsing définissant les commandes et expressions suivantes :

- `%Tokenize` est défini sur `None`. Lorsque `Tokenize` est défini sur `None`, la règle de grammaire de parsing doit comprendre tout espace ou autre séparateur de jetons dans sa définition.
- `%InputField` est défini de sorte à analyser les données d'entrée du champ **PhoneNumber**.
- `%OutputFields` est défini de sorte à séparer les données analysées en quatre champs : **CountryCode**, **AreaCode**, **Exchange** et **Number**.
- L'expression `<root>` définit le modèle des jetons en cours d'analyse et inclut des instructions OR (`|`), de sorte qu'un numéro de téléphone corresponde à :
  - **CountryCode**, **AreaCode**, **Exchange** et **Number** OR
  - **AreaCode**, **Exchange** et **Number** OR
  - **Exchange** et **Number**

La grammaire de parsing utilise une combinaison d'expressions régulières et de caractères libéraux pour construire un modèle pour les numéros de téléphone. Tout caractère entre guillemets dans cette grammaire de parsing sont des caractères littéraux ou une expression régulière.

Le caractère plus (+) utilisé dans cette commande `<root>` est défini comme un caractère littéral, parce qu'il figure entre guillemets. Vous pouvez utiliser des guillemets simples ou doubles pour

indiquer un caractère littéral. Si le caractère plus est utilisé sans guillemets, il signifie que l'expression qu'il suit peut se produire une ou plusieurs fois.

Les règles du domaine du numéro de téléphone sont définies pour correspondre aux modèles de caractères suivants :

- Zéro ou une occurrence d'un caractère « + ».
- La règle **CountryCode**, qui est constituée d'un seul chiffre compris entre 0 et 9.
- Zéro ou une occurrence d'une parenthèse ouverte ou un caractère tiret ou espace. Deux de ces caractères ayant une occurrence dans les résultats de séquence d'une non correspondance, en d'autres termes, un numéro de téléphone non valide.
- La règle **AreaCode**, qui est une séquence d'exactly trois chiffres compris entre 0 et 9.
- Zéro ou une occurrence d'une parenthèse ouverte ou un caractère tiret ou espace. Deux de ces caractères ayant une occurrence dans les résultats de séquence d'une non correspondance, en d'autres termes, un numéro de téléphone non valide.
- La règle **Exchange**, qui est une séquence d'exactly trois chiffres compris entre 0 et 9.
- Zéro ou une occurrence d'une parenthèse ouverte ou un caractère tiret ou espace. Deux de ces caractères ayant une occurrence dans les résultats de séquence d'une non correspondance, en d'autres termes, un numéro de téléphone non valide.
- La règle **Number**, qui est une séquence d'exactly quatre chiffres compris entre 0 et 9.

Les variables de règle qui définissent le domaine doivent utiliser les mêmes noms que les champs de sortie définis dans la commande `OutputFields` requise.

### Expressions régulières et quantificateurs d'expression

La grammaire d'analyse utilise une combinaison d'expressions régulières et de quantificateurs d'expression pour construire un modèle pour les numéros de téléphone aux États-Unis. La grammaire de parsing utilise ces caractères spéciaux :

- Le caractère « ? » signifie qu'une expression régulière peut se produire zéro ou une fois.
- Le caractère (|) indique une condition OR.
- Le caractère « ; » signifie la fin d'une règle.

Utilisez l'onglet **Commandes** pour explorer la signification des autres symboles spéciaux que vous pouvez utiliser dans les grammaires d'analyse en survolant la description avec la souris.

### Utilisation de l'onglet Aperçu

Pour tester la grammaire d'analyse, cliquez sur l'onglet Aperçu. Saisissez les numéros de téléphone affichés ci-dessous dans le champ **PhoneNumber**, puis cliquez sur **Aperçu**.

PhoneNumber	CountryCode	AreaCode	Exchange	Number
1(410)286-7334	1	410	286	7334
14042867534	1	404	286	7534
(410)286-7256		410	286	7256
301-868-9999		301	868	9999
1-222-458-7799	1	222	458	7799
+1(410)286-7334	1	410	286	7334
901 888 9990		901	888	9990
1 410 888 2345	1	410	888	2345
234-4567			234	4567
234 6789			234	6789

Vous pouvez aussi saisir d'autres numéros de téléphone valides et non valides pour observer comment les données d'entrée sont parsées.



Vous pouvez aussi utiliser la fonctionnalité Suivi pour voir une représentation graphique soit du résultat de parsing final, soit étape par étape au cours des événements de parsing. Cliquez sur le lien de la colonne **Suivre** pour afficher Suivre les détails de la ligne de données.

#### *Write to File*

Le modèle comporte un stage Write to File. Outre le champ d'entrée, le fichier de sortie contient les champs **CountryCode**, **AreaCode**, **Exchange** et **Number**.

# 3 - Standardisation

## In this section

---

Standardisation des termes	59
Standardisation des noms de personne	60
Modèles de standardisation	62

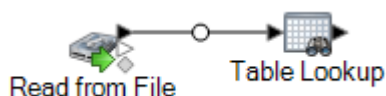
## Standardisation des termes

Une utilisation incohérente de la terminologie peut représenter un problème de qualité des données qui cause des difficultés d'analyse, de recherche, etc. Vous pouvez créer un flux de données qui recherche dans vos données des termes utilisés de manière incohérente et les standardise. Par exemple, si vos données incluent les termes « Incorporated », « Inc. » et « Inc » dans le nom de société, vous pouvez créer un flux de données permettant de standardiser le format (par exemple, « Inc. »).

**Remarque :** Avant d'effectuer cette procédure, votre administrateur doit installer la base de données du module Data Normalization contenant les termes standardisés que vous souhaitez appliquer à vos données. Vous trouverez les instructions d'installation des bases de données dans le *Guide d'installation*.

1. Dans Enterprise Designer, créez un flux de données.
2. Faites glisser un stage source sur le canevas.
3. Double-cliquez sur le stage source et configurez-le. Pour obtenir les instructions sur la configuration des stages source, reportez-vous au *Guide du concepteur du flux de données*.
4. Faites glisser un stage Table Lookup sur le canevas et connectez-le au stage source.

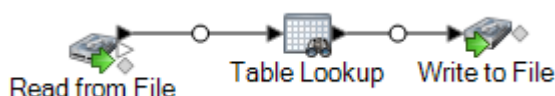
Par exemple, si vous utilisiez un stage Read from File, votre flux de données se présenterait comme suit :



5. Double-cliquez sur le stage Table Lookup sur le canevas.
6. Pour spécifier des options pour Table Lookup, vous créez une règle. Vous pouvez créer plusieurs règles puis spécifier l'ordre dans lequel ces règles seront appliquées. Cliquez sur **Ajouter** pour créer une règle.
7. Dans le champ **Action**, laissez l'option **Standardiser** par défaut sélectionnée.
8. Dans le champ **Le**, laissez l'option **Compléter le champ** sélectionnée si l'ensemble du champ correspond au terme que vous souhaitez standardiser. Vous pouvez également choisir **Termes individuels dans le champ** pour standardiser différents mots dans le champ.
9. Dans le champ **Source**, sélectionnez le champ à standardiser.
10. Dans le champ **Destination**, sélectionnez le champ qui doit contenir le terme standardisé. Si vous indiquez le même champ comme champ source, la valeur du champ source sera remplacée par le terme standardisé.
11. Dans le champ **Table**, sélectionnez la table contenant les termes standardisés.

**Remarque :** Si vous ne voyez pas la table dont vous avez besoin, contactez votre administrateur système. La base de données du module Data Normalization doit être chargée.

12. Dans le champ **Quand l'entrée de table est introuvable, définir la valeur de destination sur**, sélectionnez **Valeur de la source**.
13. Cliquez sur **OK**.
14. Définissez des règles supplémentaires pour standardiser les valeurs d'autres champs. Lorsque vous avez terminé de définir les règles, cliquez sur **OK**.
15. Faites glisser un stage de collecteur de données sur le canevas et connectez-le à Table Lookup. Par exemple, si vous utilisiez Write to File, votre flux de données se présenterait comme suit :



16. Double-cliquez sur le stage de collecteur de données et configurez-le. Pour obtenir des informations sur la configuration des stages de collecteur de données, reportez-vous au *Guide du concepteur de flux de données*.

Vous disposez désormais d'un flux de données qui standardise les termes.

## Standardisation des noms de personne

Cette procédure explique comment prendre les données de nom de personne (par exemple, « John P. Smith »), identifie les surnoms courants d'un même nom et crée une version standard du nom qui peut alors être utilisée pour consolider les enregistrements redondants.

**Remarque :** Avant de commencer, assurez-vous que vos données d'entrée disposent d'un champ appelé « Nom » qui contient le nom complet de la personne.

1. Si ce n'est pas déjà fait, chargez les tables suivantes sur le serveur Spectrum™ Technology Platform :

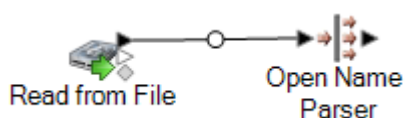
- Base Open Parser
- Noms améliorés Open Parser

Utilisez l'utilitaire de chargement de la base de données du module Data Normalization pour charger ces tables. Pour en savoir plus sur le chargement de tables, voir le *Guide d'installation*.

2. Dans Enterprise Designer, créez un flux de données.
3. Faites glisser un stage source sur le canevas.

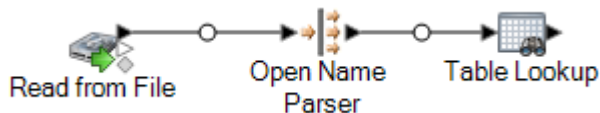
4. Double-cliquez sur le stage source et configurez-le. Pour obtenir les instructions sur la configuration des stages source, reportez-vous au *Guide du concepteur du flux de données*.
5. Faites glisser un stage Open Name Parser sur le canevas et connectez-le au stage source.

Par exemple, si vous utilisez un stage Read from File, votre flux de données se présente comme suit :



6. Faites glisser un stage Table Lookup sur le canevas et connectez-le au stage Open Name Parser.

Votre flux de données prend maintenant la forme suivante :

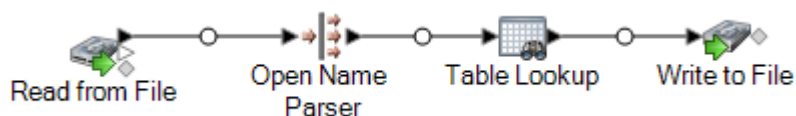


7. Double-cliquez sur le stage Table Lookup sur le canevas.
8. Dans le champ **Source**, sélectionnez **Prénom**.
9. Dans le champ **Destination**, sélectionnez **Prénom**.

En indiquant le même champ pour la source et la destination, le champ sera mis à jour avec la version standardisée du nom.

10. Dans le champ **Table**, sélectionnez **NickNames.xml**.
11. Cliquez sur **OK**.
12. Cliquez de nouveau sur **OK** pour fermer la fenêtre **Options Table Lookup**.
13. Faites glisser un stage de collecteur de données sur le canevas et connectez-le au stage Table Lookup.

Par exemple, si vous utilisiez un collecteur de données Write To File, votre flux de données se présenterait alors comme suit :



14. Double-cliquez sur le stage de collecteur de données et configurez-le. Pour obtenir les instructions sur la configuration des stages source, reportez-vous au *Guide du concepteur du flux de données*.

Vous disposez désormais d'un flux de données qui prend les noms personnels et standardise le prénom, remplaçant les surnoms par la forme standard du nom.

## Modèles de standardisation

### Formalisation des noms de personnes

Ce modèle de flux de données montre comment prendre des données de nom de personne (par exemple, « John P. Smith »), identifier les surnoms courants d'un même nom et créer une version standard du nom qui peut alors être utilisée pour consolider les enregistrements redondants. Il vous indique également comment ajouter les données de civilité en fonction des données de sexe.

#### *Scénario commercial*

Vous travaillez pour une organisation à but non lucratif qui désire envoyer des invitations pour un gala. Vos données d'entrée comprennent des données de nom comme des noms complets et vous désirez parser les données de nom en champs de prénom et de nom de famille, et ajouter un titre de respect pour rendre vos invitations plus formelles. Vous voulez aussi remplacer tous les surnoms de vos données de nom par une variante du nom plus formelle.

Les flux de données suivants apportent une solution au scénario commercial :



Ce modèle de flux de données est disponible dans Enterprise Designer. Accédez à **Fichier > Nouveau > Flux de données > À partir d'un modèle** et sélectionnez **StandardizePersonalNames**. Ce flux de données requiert les modules Data Normalization et Universal Name.

Pour chaque ligne de données dans le fichier d'entrée, ce flux de données effectuera ce qui suit :

#### *Read from File*

Ce stage identifie le nom du fichier, l'emplacement et la disposition du fichier contenant les noms que vous désirez parser. Le fichier contient à la fois des noms féminins et masculins.

#### *Name Parser*

Dans ce modèle, le stage Parser les noms de personne est nommée Parse Personal Name. Le stage Parse Personal Name examine les champs de noms et les compare aux données de noms stockées dans les fichiers de la base de données de noms Spectrum™ Technology Platform. En se basant sur la comparaison, il parse les données de noms en champs de prénom, second prénom

et nom de famille, et assigne à chaque nom un type d'entité et un genre. Il utilise aussi une reconnaissance de modèle en plus des données de nom.

Dans ce modèle, le stage Parser les noms de personne est configurée comme suit.

- Le parsing des noms de personnes est sélectionné et le parsing des noms commerciaux est effacé. Lorsque vous sélectionnez ces options, les prénoms sont évalués afin de déterminer leur sexe, leur ordre et leur ponctuation et aucune évaluation de noms d'entreprises n'est réalisée.
- Gender Determination Source (Source de détermination du sexe) est définie par défaut. Dans la plupart des cas, le paramètre par défaut est le plus adapté à la détermination par sexe parce qu'il couvre une grande variété de noms. Toutefois, si vous traitez des noms pour une culture spécifique, sélectionnez cette culture. La sélection d'une culture spécifique aide à garantir l'assignation d'un genre approprié aux noms. Par exemple, si vous laissez la sélection par défaut, le nom de Jean sera identifié comme un prénom féminin. Par contre, si vous sélectionnez Français, il sera identifié comme un prénom masculin.
- L'ordre est paramétré sur naturel. Les champs de nom sont classés par titre, prénom, deuxième prénom, nom de famille et suffixe.
- Conserver les points est effacé. Aucune ponctuation figurant dans le nom n'est conservée.

### Transformer

Dans ce modèle, le stage Transformer se nomme Assign Titles (Assigner des titres). Le stage Assign Titles utilise un script personnalisé pour effectuer une recherche sur chaque ligne du flux de données émis par le stage Parse Personal Name et assigne une valeur **TitleOfRespect** en fonction de la valeur **GenderCode**.

Le script personnalisé est :

```
if (row.get('TitleOfRespect') == '')
{
  if (row.get('GenderCode') == 'M')
    row.set('TitleOfRespect', 'Mr')
  if (row.get('GenderCode') == 'F')
    row.set('TitleOfRespect', 'Ms')
```

Chaque fois que le stage Assign Titles trouve **M** dans le champ **GenderCode**, il définit la valeur de **TitleOfRespect** sur **Mr**. Chaque fois que le stage Assign Titles trouve **F** dans le champ **GenderCode**, il définit la valeur de **TitleOfRespect** sur **Ms**.

### Standardisation

Dans ce modèle, le stage Standardisation se nomme Standardize Nicknames (Standardiser les surnoms). Le stage Standardize Nickname recherche les prénoms dans la base de données Nicknames.xml et remplace tous les surnoms par la forme plus réglementaire du nom. Par exemple, le nom Tommy est remplacé par Thomas.

### *Write to File*

Le modèle comporte un stage Write to File. En plus des champs d'entrée, le fichier de sortie contient les champs TitleOfRespect, FirstName, MiddleName, LastName, EntityType, GenderCode, et GenderDeterminationSource.



# 4 - Correspondance

## In this section

---

Mise en correspondance de la terminologie	66
Champs standard	68
Techniques de définition de match keys	69
Règles de correspondance	72
Mise en correspondance d'enregistrements d'une source unique	87
Mise en correspondance d'enregistrements d'une source avec ceux d'une autre source	93
Mise en correspondance d'enregistrements entre et dans des sources	100
Mise en correspondance d'enregistrements par rapport à une base de données	106
Rapprochement d'enregistrements à l'aide de plusieurs règles de rapprochement	109
Création d'un service de rapprochement universel	113
Utilisation d'une Match Key Express	117
Analyse des résultats de correspondance	121
Modèles de dataflows pour la correspondance	138

## Mise en correspondance de la terminologie

<b>Score moyen</b>	Score de correspondance moyen de tous les doublons. Les valeurs possibles sont 0-100, où 0 indique une faible correspondance et 100 indique une correspondance exacte.
<b>Ligne de base</b>	Le résultat de correspondance sélectionné qui sera comparé à un autre résultat de correspondance.
<b>Groupe de candidats</b>	Enregistrements suspects et candidats groupés les uns avec les autres via un identifiant qu'ils ont reçu de CandidateFinder. Le suspect (le premier enregistrement du groupe) est un enregistrement lu à partir de la source d'entrée alors que les enregistrements candidat sont habituellement extraits d'une base de données à l'aide d'une requête SQL.
<b>Enregistrements candidats</b>	Tous les enregistrements non suspects d'un groupe d'enregistrements ou d'un groupe de candidats.
<b>Drop</b>	Un réduction du nombre de doublons.
<b>Enregistrement détaillé de rapprochement</b>	Un enregistrement unique qui correspond à un enregistrement traité par un stage de rapprochement. Chaque enregistrement indique s'il s'agit d'un enregistrement de référence, unique ou doublon d'un autre, et fournit des informations sur son groupe d'appartenance. Les enregistrements candidats fournissent des informations sur les raisons pour lesquelles l'enregistrement donné correspond ou non à son enregistrement de référence.
<b>Groupes de doublons</b>	Un groupe de doublons comprend un enregistrement de référence et ses enregistrements doublons regroupés par numéro de groupe. Un groupe de doublons est également appelé une Collection dans Spectrum. Les enregistrements uniques appartiennent toujours au groupe 0.
<b>Enregistrements doublons</b>	Enregistrement correspondant à un autre enregistrement dans un groupe de doublons. Il peut s'agir d'un enregistrement de référence ou d'un candidat.
<b>Rapprochements express</b>	On parle de rapprochement express lorsque les contenus d'un champ donné d'un enregistrement de référence et d'un enregistrement candidat correspondent parfaitement. Il s'agit généralement d'une clé de rapprochement express fournie par le Match Key Generator. Lorsqu'un rapprochement express est trouvé, aucune autre opération de traitement n'est conduite et l'enregistrement de référence et l'enregistrement candidat sont considérés comme doublons.
<b>Enregistrements d'entrée</b>	Ordre des enregistrements dans le stage de correspondance avant que le tri de correspondance soit effectué
<b>Interflow Match</b>	Un stage de mise en correspondance localisant les correspondances entre des enregistrements de données semblables entre deux flux d'enregistrements d'entrée. Le premier flux d'enregistrement est une

	source d'enregistrements suspects et le second flux est une source d'enregistrements candidats.
<b>Intraflow Match</b>	Un stage de mise en correspondance localisant les correspondances entre des enregistrements de données semblables dans un seul flux d'entrée.
<b>Augmentation</b>	Une augmentation des doublons.
<b>Groupes de rapprochements</b>	(Regrouper par) Les enregistrements regroupés soit par une clé de correspondance ou une fenêtre coulissante.
<b>Résultats de correspondance</b>	(ou Paquet de ressources) Regroupement logique de fichiers produits par une étape. Ces données sont enregistrées à chaque étape exécutée et stockées sur disque. Toute exécution ultérieure n'efface ni ne modifie les résultats de l'exécution précédente. Dans MAT, les paquets sont utilisés pour offrir des informations sur le sommaire et le détail des résultats, ainsi que des informations de réglages.
<b>Liste des résultats de correspondance</b>	Liste des résultats de correspondance un seul type que MAT peut analyser dans la session d'analyse en cours.
<b>Type de résultats de correspondance</b>	Désigne le contenu des résultats de correspondance. MAT utilise le type des résultats de correspondance pour définir comment utiliser les données.
<b>Stage de correspondance</b>	Un stage sur le canevas effectuant les programmes de mise en correspondance. Les stages de mise en correspondance sont Interflow Match, Intraflow Match et Transactional Match.
<b>Rapprochements manqué</b>	Enregistrement ayant précédemment été identifié comme suspect ou doublon mais étant maintenant unique.
<b>Nouveau rapprochement</b>	Un enregistrement ayant précédemment été identifié comme unique mais étant maintenant suspect ou doublon.
<b>Fenêtre dynamique</b>	La méthode de correspondance de fenêtre coulissante remplit séquentiellement une taille tampon prédéterminée nommée fenêtre, avec le nombre correspondant de lignes de données. Chaque ligne ajoutée est comparée aux éléments déjà contenus dans la fenêtre.
<b>Enregistrements suspects</b>	Un enregistrement pilote mis en correspondance avec un groupe de correspondance ou un groupe candidat.
<b>Transactional Match</b>	Un stage de mise en correspondance qui met en correspondance les enregistrements suspects avec les enregistrements suspects renvoyé par Candidate Finder ou par une application externe.
<b>Enregistrements uniques</b>	Enregistrement de référence ou candidat qui ne correspond à aucun autre enregistrement dans un groupe de doublons. S'il est le seul enregistrement dans un groupe de doublons, un enregistrement de référence est automatiquement unique.

## Champs standard

Ce tableau répertorie certains des champs standard que vous rencontrez dans les tâches de mise en correspondance et en donne une description.

Champs	Descriptions
CandidateGroup	Ce champ identifie un groupe composé d'un enregistrement suspect et de ses candidats. Chaque enregistrement suspect reçoit un numéro CandidateGroup. Les candidats correspondant à ce suspect reçoivent le même numéro CandidateGroup. Par exemple, si Pierre Martin est un enregistrement suspect et que ses enregistrements candidats sont Pierre Martin et Perre Martn, ces trois enregistrements auraient la même valeur CandidateGroup.
CollectionNumber	Identifie une collection de doublons. Les valeurs possibles sont 1 ou des valeurs supérieures à 1. Les enregistrements marqués avec la même valeur sont des enregistrements en double.
CollectionRecordType	Identifie les enregistrements modèles et Best of Breed dans une collection d'enregistrements doublons. Les valeurs possibles sont : <ul style="list-style-type: none"> <li>• Primary : si l'enregistrement est l'enregistrement du modèle sélectionné dans une collection.</li> <li>• Secondary : si l'enregistrement n'est pas l'enregistrement du modèle sélectionné dans une collection.</li> <li>• BestOfBreed : si l'enregistrement est le nouvel enregistrement créé Best of Breed de la collection.</li> </ul>
ExpressMatchIdentified	Indique si un rapprochement a été obtenu à l'aide de la clé de correspondance express. Les valeurs possibles sont Yes ou No.
HasDuplicates	Identifie si l'enregistrement est un doublon d'un autre enregistrement. Les options sont : <ul style="list-style-type: none"> <li>• Y : l'enregistrement est un enregistrement suspect et possède des doublons.</li> <li>• N : l'enregistrement est un enregistrement suspect et ne possède pas de doublons.</li> <li>• D : l'enregistrement est un enregistrement candidat et est un doublon de l'enregistrement suspect.</li> <li>• U : l'enregistrement est un enregistrement candidat, mais n'est pas un doublon de l'enregistrement suspect.</li> </ul>

Champs	Descriptions
MatchRecordType	<p>Identifie le type d'enregistrement de correspondance dans une collection. Les valeurs possibles sont :</p> <ul style="list-style-type: none"> <li>• Suspect : un enregistrement auquel les autres enregistrements sont comparés afin de déterminer s'ils sont des doublons les uns des autres. Chaque collection possède un seul enregistrement suspect.</li> <li>• Duplicate : un enregistrement qui est un doublon de l'enregistrement suspect.</li> <li>• Unique : un enregistrement qui n'a pas de doublons.</li> </ul>
MatchScore	<p>Identifie le score global entre deux enregistrements. Les valeurs possibles sont 0-100, où 0 indique une faible correspondance et 100 indique une correspondance exacte. Les scores sont générés sur la base de la configuration de la règle de correspondance.</p>
TotalMatchCount	<p>Ce champ indique le nombre total de correspondances effectuées lors du traitement.</p>
TransactionRecordType	<p>Il peut s'agir de l'une des options suivantes :</p> <ul style="list-style-type: none"> <li>• Suspect : un enregistrement suspect est utilisé comme entrée dans une requête.</li> <li>• Candidate : un enregistrement candidat est un résultat renvoyé par une requête.</li> </ul>

## Techniques de définition de match keys

Un rapprochement efficace et efficient nécessite un bon équilibre entre précision et performances. En matière de rapprochement, l'approche la plus précise consisterait à analyser chaque enregistrement par rapport à tous les autres enregistrements, mais elle n'est pas pratique, car le nombre d'enregistrements à traiter entraînerait une lenteur inacceptable des performances. Il est préférable de limiter le nombre d'enregistrements impliqués dans le processus de rapprochement aux enregistrements les plus susceptibles de correspondre. Pour ce faire, vous pouvez utiliser des match keys. Une match key est une valeur créée pour chaque enregistrement à l'aide d'un algorithme que vous définissez vous-même. L'algorithme prend des valeurs de l'enregistrement et les utilise pour produire une valeur de match key, qui est ensuite stockée sous forme de nouveau champ dans l'enregistrement.

Par exemple, si l'enregistrement entrant est :

Prénom : Fred Nom de famille : Mertz Code postal : 21114-1687 Code de genre : M

Et que vous définissez une règle de match key qui génère une match key en combinant des données de l'enregistrement comme suit :

Champ de saisie	Position de début	Longueur
Code postal	1	5
Code postal	7	4
Nom	1	5
Prénom	1	5
Code de genre	1	1

Alors, la clé serait :

211141687MertzFredM

Tous les enregistrements présentant la même match key sont placés dans un groupe de doublons. Le processus de rapprochement compare ensuite les enregistrements du groupe les uns aux autres pour identifier les rapprochements.

Pour créer une match key, si vous rapprochez des enregistrements via Interflow Match ou Intraflow Match, utilisez un stage Match Key Generator. Si vous rapprochez des enregistrements via Transactional Match, utilisez le stage Candidate Finder pour créer des groupes de doublons.

**Remarque :** Les instructions qui suivent peuvent être appliquées au clés Match Key Generator et aux requêtes Candidate Finder. Dans Candidate Finder, ces instructions indiquent la manière dont vous définissez l'instruction SELECT.

### *Taille de groupe de doublons et performances*

La match key détermine la taille du groupe de doublons, et, par conséquent, les performances de votre dataflow. Si la taille du groupe de doublons double, le temps d'exécution double lui aussi. Par exemple, si vous définissez une match key qui produit un groupe de 20 enregistrements susceptibles de correspondre, le traitement prend deux fois plus de temps que si vous modifiez la match key pour que le groupe de doublons contienne seulement 10 enregistrements susceptibles de correspondre. L'inconvénient de réduire la règle de match key pour produire un groupe de doublons plus petit réside dans le fait que vous courez le risque d'exclure des enregistrements qui correspondent. Des règles de match key « lâches », en contrepartie, réduisent les chances d'exclusion d'enregistrements correspondants du groupe, mais augmentent la taille du groupe. Pour trouver le juste équilibre pour vos données, il est important de faire des tests avec différentes règles de match key à l'aide de données représentatives de celles que vous avez l'intention de traiter en production.

### *Densité*

Lors de la conception d'une match key, il convient de tenir compte de la densité des données. La densité fait référence au degré auquel les données peuvent être distribuées sur l'ensemble des

groupes doublons. Étant donné que les performances sont déterminées par le nombre de comparaisons à effectuer par le système, les match keys qui produisent un petit nombre de gros groupes de doublons ralentissent les performances par rapport aux match keys qui produisent un grand nombre de petits groupes de doublons.

Pour illustrer ce concept, prenons une situation dans laquelle vous avez un million d'enregistrements de noms et d'adresses à rapprocher. Vous pouvez définir une match key sur les trois premiers octets du code postal et la première lettre du nom de famille. Si les enregistrements proviennent de l'ensemble des États-Unis, la match key produit un nombre satisfaisant de groupes de doublons, avec des performances probablement acceptables. En revanche, si tous les enregistrements proviennent de New York, les codes postaux commencent tous par « 100 » et vous aboutissez à, au plus, seulement 26 groupes de doublons. Cela produit de gros groupes de doublons contenant, en moyenne, quelque 38 000 enregistrements.

Vous pouvez calculer le nombre maximal de comparaisons effectuées pour chaque groupe de doublons à l'aide de la formule suivante :

$$N * (N-1) / 2$$

où N est le nombre d'enregistrements du groupe de doublons.

Ainsi, si vous disposez de 26 groupes de doublons contenant chacun 38 000 enregistrements, le nombre maximal de comparaisons effectuées est d'environ 18,7 milliards. Voici comment ce nombre est calculé :

Tout d'abord, déterminez le nombre maximal de comparaisons par groupe de doublons :

$$38,000 * (38,000-1) / 2 = 721,981,000$$

Ensuite, multipliez ce nombre par le nombre de groupes de doublons :

$$721,981,000 * 26 = 18,771,506,000$$

Si, au lieu de cela, il existait 100 valeurs uniques pour les trois premiers octets du code postal, vous disposeriez de 2 600 groupes de doublons contenant en moyenne 380 enregistrements. Dans ce cas, le nombre maximal de comparaisons serait de 187 millions, à savoir, 100 fois moins. Ainsi, si les enregistrements proviennent uniquement de New York, vous pouvez envisager d'utiliser les quatre, voire les cinq premiers octets du code postal pour la match key, afin de produire davantage de groupes de doublons et de réduire le nombre de comparaisons. Il se peut que vous manquiez quelques rapprochements, mais, en revanche, le temps d'exécution sera considérablement réduit.

En réalité, une match key comme celle utilisée dans cet exemple ne donne pas des groupes de doublons de taille égale, à cause des variations des données. Par exemple, beaucoup plus de personnes portent un nom de famille qui commence par la lettre « S » que par la lettre « X ». C'est pourquoi vous devez vous efforcer de réduire la taille des plus gros groupes de doublons. Un groupe de doublons comptant 100 000 enregistrements est 10 fois plus gros qu'un groupe de doublons en comptant 10 000, mais il nécessite 100 fois plus de comparaisons et prend 100 fois plus de temps. Par exemple, imaginons que vous utilisiez cinq octets de code postal et six octets du champ AddressLine1 pour votre match key. En apparence, cette match key semble relativement satisfaisante. Le problème, ce sont les adresses de boîte postale. Même si la plupart des groupes de doublons sont d'une taille acceptable, il existera quelques très gros groupes de doublons avec

des clés comme 10002PO BOX contenant un très grand nombre d'enregistrements. Pour diviser les gros groupes de doublons, vous pouvez modifier votre match key pour qu'elle comprenne les deux premiers chiffres du numéro de boîte postale.

### *Alignement de la match key sur la règle de rapprochement*

Pour obtenir les résultats les plus précis, vous devez concevoir la match key de sorte qu'elle fonctionne bien avec la règle de rapprochement avec laquelle vous comptez l'utiliser. Pour ce faire, vous devez tenir compte de la définition de la règle de rapprochement.

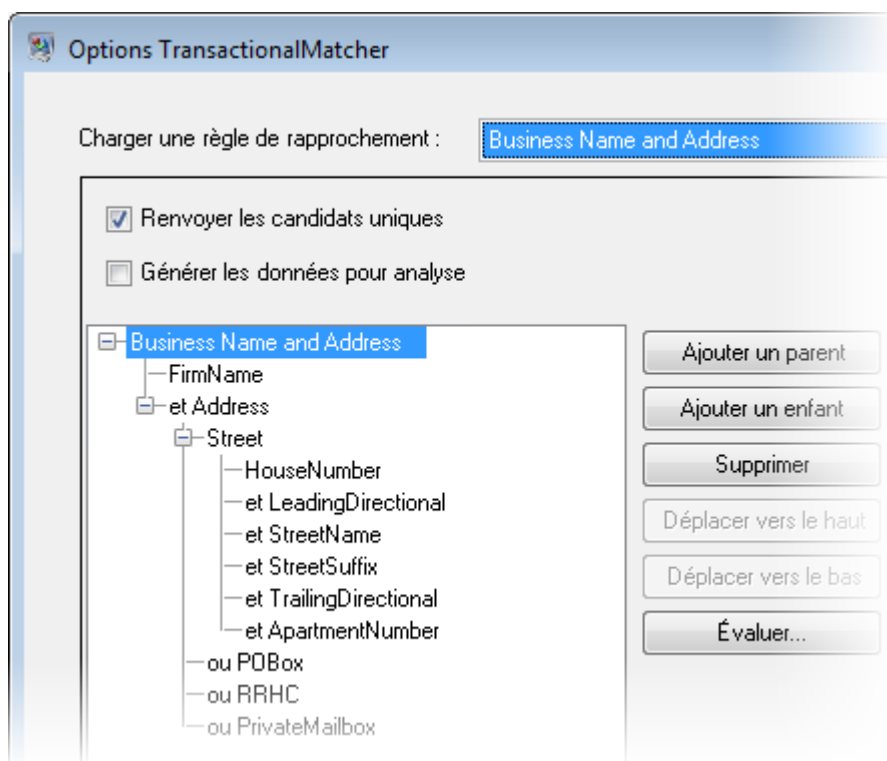
- La match key doit inclure tous les champs requis par la règle de rapprochement pour obtenir une correspondance exacte.
- La match key doit utiliser le même type d'algorithme que celui utilisé dans la règle de rapprochement. Par exemple, si vous concevez une match key à utiliser avec une règle de rapprochement qui emploie un algorithme phonétique, la match key doit elle aussi utiliser un algorithme phonétique.
- La match key doit être créée à l'aide des données de tous les champs utilisés dans la règle de rapprochement.
- Imaginez l'impact sur la match key s'il manque des données d'un ou de plusieurs des champs utilisés dans la règle de rapprochement. Par exemple, imaginons que vous utilisiez l'initiale du deuxième prénom dans la match key et que vous disposiez d'un enregistrement pour John A. Smith et d'un autre pour John Smith. Vous avez configuré la règle de rapprochement de sorte qu'elle ignore les valeurs vierges du champ du deuxième prénom. C'est pourquoi, suivant votre règle de rapprochement, ces deux enregistrements correspondent. Cependant, étant donné que la match key utilise l'initiale du deuxième prénom, les deux enregistrements aboutissent dans deux groupes de doublons différents et ne sont pas comparés l'un à l'autre, annulant ainsi l'effet souhaité de votre règle de rapprochement.

## Règles de correspondance

Chacun des stages de correspondance (Interflow Match, Intraflow Match et Transactional Match) nécessitent que vous configuriez une règle de correspondance. Une règle de correspondance définit les critères permettant de déterminer si un enregistrement correspond à un autre. Elle spécifie les champs à comparer, la manière dont comparer les champs et une hiérarchie de comparaisons pour les règles de correspondance complexes.

Créer un ensemble hiérarchique de comparaisons vous permet de former des règles de correspondance booléennes imbriquées. Par exemple, considérez la règle de correspondance suivante :





Dans cet exemple, la règle de correspondance tente de mettre en correspondance les enregistrements en fonction d'une raison sociale et d'une adresse. Le premier élément de la règle de correspondance est le champ FirmName. Cet élément signifie que la valeur dans le champ FirmName doit correspondre pour les enregistrements correspondent. Le second élément évalue l'adresse. Celui-ci est précédé par l'opérateur logique "and" qui signifie que les éléments FirmName et Address doivent correspondre pour que les enregistrements correspondent. La partie Address de la règle de correspondance se compose de règles enfant qui évaluent quatre types d'adresses : adresses de rue, adresses de boîtes postales, adresses Route rurale/Highway Contract (RRHC) et adresses de boîte postale privée. L'enfant Street examine les champs de flux de données HouseNumber, LeadingDirectional, StreetName, StreetSuffix, TrailingDirectional et ApartmentNumber. S'ils correspondent tous, la règle parent "Street" et sa règle parent "Address" sont évaluées sur "true". Si la règle Street n'est pas évaluée sur true, le champ POBox est évalué, puis RRHC et PrivateMailbox. Si l'un de ces trois champ est mis en correspondance, l'élément parent Address est également mis en correspondance.

## Création d'une règle de correspondance

Les règles de correspondance sont utilisées dans Interflow Match, Intraflow Match et Transactional Match pour définir les critères qui déterminent si un enregistrement correspond à un autre. Les règles de correspondance spécifient les champs à comparer, la manière dont comparer les champs et une hiérarchie de comparaisons pour les règles de correspondance complexes.

Vous pouvez créer des règles de correspondance dans Interflow Match, Intraflow Match et Transactional Match. Vous pouvez également créer des règles de correspondance dans l'outil Gestion des règles de correspondance d'Enterprise Designer. Créer une règle dans l'outil Gestion des règles de correspondance rend la règle disponible pour être utilisée dans tout flux de données et la rend également disponible aux autres utilisateurs. Créer une règle de correspondance dans un des stages de mise en correspondance rend la règle disponible uniquement pour ce stage, à moins d'enregistrer la règle en cliquant sur le bouton **Enregistrer**, ce qui la rend disponible aux autres stages et utilisateurs.

1. Ouvrez Enterprise Designer.
2. Procédez de l'une des façons suivantes :
  - Si vous souhaitez définir une règle de correspondance dans Interflow Match, Intraflow Match ou Transactional Match, double-cliquez sur le stage de mise en correspondance pour lequel vous souhaitez définir une règle de correspondance. Dans le champ **Charger une règle de correspondance**, choisissez une règle de correspondance prédéfinie comme point de départ. Si vous souhaitez démarrer par une règle de correspondance vide, cliquez sur **Nouveau**.
  - Si vous souhaitez définir une règle de correspondance dans l'outil Gestion des règles de correspondance, sélectionnez **Outils > Gestion des règles de correspondance**. Si vous souhaitez utiliser une règle existante comme point de départ pour votre règle, cochez la case **Copier à partir de** et sélectionnez la règle à utiliser comme point de départ.
3. Spécifiez les champs de flux de données à utiliser dans la règle de correspondance, ainsi que la hiérarchie de règle de correspondance.
  - a) Cliquez sur **Ajouter un parent**.
  - b) Saisissez un nom pour le parent. Le nom doit être unique et ne peut pas être un champ. Le premier parent dans la hiérarchie est utilisé comme le nom de la règle de correspondance dans le champ **Charger une règle de correspondance**. Toutes les règles de correspondance personnalisées que vous créez et les règles prédéfinies que vous modifiez sont sauvegardées avec le mot « Personnalisé » précédant le nom.
  - c) Cliquez sur **Ajouter un enfant**. Un menu déroulant apparaît dans la hiérarchie de règle. Sélectionnez un champ à ajouter au parent.
 

**Remarque** : Tous les enfants sous un parent doivent utiliser le même opérateur logique. Si vous souhaitez utiliser différents opérateurs logiques entre les champs, vous devez d'abord créer des parents intermédiaires.
  - d) Recommencez afin de compléter votre hiérarchie de correspondance.
4. Définir des options de parent. Les options de parent sont affichées à droite de la hiérarchie de règle lorsqu'un nœud parent est sélectionné.
  - a) Cliquez sur **Correspondance lorsque la valeur n'est pas True** pour modifier l'opérateur logique pour le parent de AND à AND NOT. Si vous sélectionnez cette option, les enregistrements correspondront uniquement s'ils ne correspondent pas non à la logique définie dans ce parent.

**Remarque :** Cocher l'option **Correspondance** lorsque la valeur n'est pas True a pour effet d'annuler les options de la **Méthode de correspondance**. Pour plus d'informations, reportez-vous à la section **Conditions de correspondance négatives** à la page 83.

- b) Dans le champ **Méthode de correspondance**, indiquez comment déterminer si un parent est une correspondance ou une non-correspondance. L'un des éléments suivants :

**Tous vrais** Un parent est considéré comme une correspondance si tous les enfants sont déterminés de manière à obtenir une correspondance. Cette méthode crée un connecteur « ET » entre les enfants.

**Au moins un vrai** Un parent est considéré comme étant une correspondance si au moins un enfant est déterminé de manière à obtenir une correspondance. Cette méthode crée un connecteur « OU » entre les enfants.

**Basé sur le seuil** Un parent est considéré comme étant une correspondance si le score du parent est supérieur ou égal au seuil du parent. Lorsque vous choisissez cette option, le curseur **Seuil** apparait. Utilisez ce curseur pour choisir un seuil. La méthode de score choisit quel connecteur logique utiliser. Les seuils au parent ne peuvent pas être plus élevés que le seuil des enfants.

**Remarque :** Le seuil définit ici peut être supplanté à tout moment dans la boîte de dialogue Options de flux de données. Allez dans **Édition > Options de flux de données**, puis cliquez sur **Ajouter**. Développez le stage, cliquez sur **Seuil de premier niveau**, et saisissez le seuil dans le champ **Valeur par défaut**.

- c) Dans le champ **Données manquantes**, indiquez comment noter des données vides dans un champ. L'un des éléments suivants :

**Ignorer les blancs** Ignorer le champ s'il contient des données vierges.

**Comptez comme 0** Noter le champ comme 0 s'il contient des données vierges.

**Comptez comme 100** Noter le champ comme 100 s'il contient des données vierges.

**Comparer les valeurs vides** Donne aux champs suspects et candidats une valeur 100 s'ils contiennent des données vierges ; sinon, donne aux champs suspects et candidats une valeur 0.

- d) Dans le champ **Méthode de notation**, sélectionnez la méthode utilisée pour déterminer le score de correspondance. L'un des éléments suivants :

**Moyenne pondérée** Utilise le poids de chaque enfant pour définir le score de correspondance moyen.

**Moyenne** Utilise le score moyen de chaque enfant pour définir le score d'un parent.

<b>Maximum</b>	Utilise le score enfant le plus élevé pour définir le score du parent.
<b>Minimum</b>	Utilise le score enfant le plus faible pour définir le score du parent.
<b>Sommation des vecteurs</b>	Utilise la sommation des vecteurs de chaque score enfant pour définir le score du parent. La formule pour le calcul est la suivante :
	$\text{sqrt}(a^2 + b^2 + c^2) / \text{sqrt}(n)$ , où : a, b et c sont les scores de trois enfants et n est le nombre d'enfants.

Le tableau suivant montre la relation logique entre les méthodes de correspondance et les méthodes de notation et comment chaque combinaison modifie la logique employée durant le traitement de correspondances.

**Tableau 1 : Matrice de méthode de correspondance de type « Méthode et Score »**

Méthode de notation	Méthode de correspondance			Commentaires
	Au moins un vrai	Tous vrais	Basé sur le seuil	
Moyenne pondérée	S/O	AND	AND	Disponible que lorsque Tous sont vrais ou Basés sur le seuil sont sélectionnés comme méthodes de correspondance.
Moyenne	S/O	AND	AND	
Sommation des vecteurs	S/O	AND	AND	
Maximum	OR	S/O	OR	Disponible que lorsque Au moins un vrai ou Basés sur le seuil sont sélectionnés comme méthodes de correspondance.
Minimum	OR	S/O	OR	

5. Définissez les options d'enfant. Les options d'enfant sont affichées à droite de la hiérarchie de règle lorsqu'un enfant est sélectionné.
  - a) Cochez l'option **Champ candidat** pour mapper le champ d'enregistrement enfant sélectionné vers un champ du fichier d'entrée.
  - b) Cochez l'option **Recouper avec** et sélectionnez un ou plusieurs éléments de la liste déroulante pour mettre en correspondance différents champs les uns avec les autres entre deux enregistrements. Si vous utilisez l'outil Gestion des règles de correspondance pour créer ou modifier une règle de correspondance, aucune liste déroulante n'apparaît et, à la place, vous devrez saisir chaque nom de champ en séparant les noms de champ par des virgules.

- c) Cliquez sur **Correspondance lorsque la valeur n'est pas True** pour modifier l'opérateur logique pour le parent de AND à NOT. Si vous sélectionnez cette option, la règle de correspondance sera uniquement évaluée comme True si les enregistrements ne correspondent pas à la logique définie dans cet enfant.

Par exemple, si vous souhaitez identifier les individus qui sont associés à des comptes multiples, vous pouvez créer une règle de correspondance qui met en correspondance le nom, sauf lorsque le numéro de compte ne correspond pas. Vous devez utiliser l'option de **Correspondance lorsque la valeur n'est pas True** pour l'enfant qui correspond au numéro de compte.

- d) Dans le champ **Données manquantes**, indiquez comment noter des données vides dans un champ. L'un des éléments suivants :

<b>Ignorer les blancs</b>	Ignorer le champ s'il contient des données vierges.
<b>Comptez comme 0</b>	Noter le champ comme 0 s'il contient des données vierges.
<b>Comptez comme 100</b>	Noter le champ comme 100 s'il contient des données vierges.
<b>Comparer les valeurs vides</b>	Donne aux champs suspects et candidats une valeur 100 s'ils contiennent des données vierges ; sinon, donne aux champs suspects et candidats une valeur 0.

- e) Dans le **champ Seuil**, indiquez le seuil devant être atteint au niveau du champ individuel afin que ce champ soit considéré comme une correspondance.
- f) Dans le champ **Méthode de notation**, sélectionnez la méthode utilisée pour déterminer le score de correspondance. L'un des éléments suivants :

<b>Moyenne pondérée</b>	Utilise le poids de chaque algorithme pour définir le score de correspondance moyen.
<b>Moyenne</b>	Utilise la moyenne de chaque algorithme pour définir le score de correspondance moyen.
<b>Maximum</b>	Utilise le score d'algorithme le plus élevé pour déterminer le score correspondant.
<b>Minimum</b>	Utilise le score d'algorithme le plus bas pour déterminer le score correspondant.
<b>Sommation des vecteurs</b>	Utilise la sommation des vecteurs du score de chaque algorithme pour définir le score de correspondance. Cette méthode de notation est utile si vous souhaitez qu'un score de correspondance supérieur dans un ou plusieurs algorithmes soit proportionnellement représenté dans le score de correspondance final. La formule utilisée pour calculer le score final est la suivante :

$\sqrt{a^2 + b^2 + c^2} / \sqrt{n}$ , où : a, b et c sont les scores de trois algorithmes différents et n est le nombre d'algorithmes utilisés.

- g) Choisissez au moins un algorithme à utiliser pour déterminer si les valeurs du champ correspondent. Un des éléments suivants.

**Acronyme** Détermine si un nom de société correspond à son acronyme en recherchant des données d'acronyme ; sinon, il crée un acronyme utilisant le premier caractère de chaque mot. Exemple : Internal Revenue Service et son acronyme IRS seraient considérés comme étant une correspondance et afficheraient un score de correspondance de 100.

**Fréquence de caractères** Détermine la fréquence des instances de chaque caractère dans une chaîne de caractère et compare la fréquence globale entre deux chaînes de caractères.

**Daitch-Mokotoff Soundex** Algorithme phonétique qui permet une plus grande exactitude dans la mise en correspondance de noms de famille slaves et yiddish avec une prononciation similaire mais des différences dans l'orthographe. Les noms codés sont composés de six chiffres et de multiples encodages possibles peuvent être renvoyés pour un seul nom. Cette option a été développée afin de respecter les limitations de Soundex dans le traitement des noms de famille allemands ou slaves.

**Date** Compare les champs de date sans prendre en compte le format de date des enregistrements d'entrée. Cliquez sur Éditer dans la colonne Options pour préciser ce qui suit :

- **Exige un mois** : empêche une date uniquement composée d'une année de correspondre
- **Exige un jour** : empêche une date uniquement composée d'un mois et d'une année de correspondre
- **Correspond au MM/LL transposé** : où le mois et le jour sont fournis sous forme numérique, compare le mois suspect au jour candidat et le jour suspect au mois candidat en plus d'effectuer la comparaison standard du mois suspect au mois candidat et du jour suspect au jour candidat
- **Préférer le format JJ/MM/AAAA au format MM/JJ/AAAA** : contribue à la décomposition analytique de la date dans les cas où le mois et le jour sont fournis sous un format numérique et que leur identification ne peut pas être déterminée par le contexte. Par exemple, en prenant les nombres 5 et 13, l'analyseur assignera automatiquement 5 au mois et 13 au jour car il n'y a que 12 mois par an. Toutefois, en prenant les nombres 5 et 12 (ou toute paire de nombres de 12 ou en dessous), l'analyseur prendra le premier nombre comme étant le mois. Cocher cette option vous permettra de vous assurer que l'analyseur lit le premier nombre comme étant le jour plutôt que le mois.
- **Options de durée—Globale** : vous permet de définir un nombre maximum de jours entre les dates correspondantes. Par exemple, si vous saisissez une durée globale de 35 jours et que votre date candidate est le 31 décembre 2000, une date suspecte du

5 février 2001 correspondrait mais pas une date suspecte du 6 février. Si vous saisissez une durée globale de 1 jour et que la date candidate est janvier 2000, une date suspecte de 1999 correspondrait (en comparant le 31 décembre 1999) mais pas une date suspecte de janvier 2001.

- **Options de durée—Année** : vous permet de définir le nombre d'années entre les dates correspondantes, indépendamment du mois et du jour. Par exemple, si vous saisissez une durée annuelle de 3 et que votre date candidate est le 31 janvier 2000, une date suspecte du 31 janvier 2003 correspondrait mais pas une date suspecte de février 2003. De même, si votre date candidate est 2000, une date suspecte de Mars 2003 correspondrait parce que les mois ne sont pas en conflit et qu'elle se trouve dans la durée des trois ans.
- **Options de durée—Mois** : vous permet de définir le nombre de mois entre les dates correspondantes, indépendamment de l'année et du jour. Par exemple, si vous saisissez une durée mensuelle de 4 et que votre date candidate est le 1er janvier 2000, une date suspecte de mai 2000 correspondrait car il n'y a pas de conflit de jour et qu'elle se trouve dans la durée des quatre mois, mais pas une date suspecte du 2 mai 2000 à cause du conflit de jours.
- **Options de durée—Jour** : vous permet de définir le nombre de jours entre les dates correspondantes, indépendamment de l'année et du mois. Par exemple, si vous saisissez une durée mensuelle de 5 et que votre date candidate est le 1er janvier 2000, une date suspecte de mai 2000 correspondrait car il n'y a pas de conflit de jour et qu'elle se trouve dans la durée des quatre mois, mais pas une date suspecte du 27 déc. 99 à cause du conflit de jours.

#### **Metaphone double**

Détermine la similitude entre deux chaînes basée sur la représentation phonétique de leurs caractères. Le double Metaphone est une version améliorée de l'algorithme Metaphone et tente de prendre en compte les nombreuses irrégularités de plusieurs langues.

#### **Distance d'édition**

Détermine la similitude entre deux chaînes en fonction du nombre de suppressions, d'insertions ou de substitutions requises pour transformer une chaîne en une autre.

#### **Distance euclidienne**

Fournit une mesure de la similitude entre deux chaînes via l'espace vectoriel de termes combinés sous forme de dimensions. Cela définit également le plus grand diviseur commun de deux entiers. Cela prend une paire d'entiers positifs et forme une nouvelle paire qui se compose du plus petit nombre et de la différence entre les nombres le plus élevé et le plus faible. Le processus se répète jusqu'à ce que les nombres soient égaux. Ce nombre est le plus grand diviseur commun de la paire d'origine. Par exemple, 21 est le plus grand diviseur commun de 252 et de 105 : ( $252 = 12 \times 21$  ;  $105 = 5 \times 21$ ) ; car



$252 - 105 = (12 - 5) \times 21 = 147$ , le plus grand diviseur commun de 147 et de 105 est également 21.

<b>Correspondance exacte</b>	Détermine si deux chaînes sont identiques.
<b>Initiales</b>	Utilisé pour mettre en correspondance les initiales de noms personnels décomposés.
<b>Distance Jaro-Winkler</b>	Détermine la similitude entre deux chaînes basée sur le nombre de remplacements de caractère requis pour transformer une chaîne en l'autre. Cette option a été développée pour les chaînes courtes, telles que les noms de personnes.
<b>Distance du clavier</b>	Détermine la similitude entre deux chaînes basée sur le nombre de suppressions, d'insertions ou de substitutions requises pour transformer un champ en l'autre, pondérée par la position des touches sur le clavier. Cliquez sur Éditer dans la colonne Options pour spécifier le type de clavier que vous utilisez : QWERTY (États-Unis), QWERTZ (Autriche et Allemagne) ou AZERTY (France).
<b>Koeln</b>	Noms d'index par son, tels qu'ils sont prononcés en allemand. Permet aux noms ayant la même prononciation d'être encodés avec la même représentation afin qu'ils puissent être mis en correspondance, en dépit de différences mineures au niveau de l'orthographe. Le résultat est toujours une séquence de nombres ; les caractères spéciaux et les espaces blancs sont ignorés. Cette option a été développée en réponse aux limites du Soundex.
<b>Distance Kullback-Liebler</b>	Détermine la similitude entre deux chaînes en fonction des différences entre la répartition des mots dans les deux chaînes.
<b>Metaphone</b>	Détermine la similitude entre deux chaînes en langue anglaise en fonction d'une représentation phonétique de leurs caractères. Cette option a été développée en réponse aux limites du Soundex.
<b>Metaphone (Espagnol)</b>	Détermine la similitude entre deux chaînes basée sur la représentation phonétique de leurs caractères. Cette option a été développée en réponse aux limites du Soundex.
<b>Metaphone3</b>	Procède à une amélioration en fonction des algorithmes Metaphone et Double Metaphone avec des paramètres de consonne et de voyelle interne exacts qui vous permet de produire des mots ou des noms mis en correspondance de manière plus ou moins proche pour rechercher des termes au niveau phonétique. Metaphone 3 augmente l'exactitude de l'encodage phonétique à 98 %. Cette option a été développée en réponse aux limites du Soundex.
<b>Variante de nom</b>	Détermine si deux noms sont des variantes l'un de l'autre. L'algorithme rapporte un score de correspondance de 100 si deux noms sont des variations l'un de l'autre, et qu'un score de correspondance de 0 si deux



noms ne sont pas des variations l'un de l'autre. Par exemple, JOHN est une variation de JAKE et renvoie un score de correspondance de 100. JOHN n'est pas une variante de HENRY et renvoie un score de correspondance de 0. Cliquez sur **Éditer** dans la colonne Options pour choisir les options de variante de nom. Pour plus d'informations, reportez-vous à la section [Name Variant Finder](#) à la page 338.

### Distance NGram

Calcule dans le texte ou l'expression la probabilité du terme suivant en fonction des n termes précédents, qui peuvent inclure des phonèmes, des syllabes, des lettres, des mots ou des paires de base et être constitués de toute combinaison de lettres. Cet algorithme inclut une option permettant de saisir la taille de NGram ; la valeur par défaut est 2.

### Similarité NGram

Détermine la similarité entre deux chaînes en fonction de la longueur de la sous-séquence commune la plus longue des phonèmes, syllabes, lettres, mots ou paires de base.

L'algorithme inclut les options suivantes :

- **Taille Ngram** : Saisissez la taille de NGram. La valeur par défaut est 2.
- **Ignorer les caractères de bruit** : Cochez la case pour remplacer la ponctuation par un espace.
- **Ignorer les espaces** : Cochez la case pour fusionner des mots.

### Chaîne numérique

Compare les lignes d'adresse en séparant les attributs numériques d'une ligne d'adresse des caractères. Par exemple, dans l'adresse de la chaîne 1234 Main Street Apt 567, l'attribut numérique de la chaîne (1234567) sont décomposés et gérés différemment des valeurs restantes de la chaîne (Main Street Apt). L'algorithme met d'abord en correspondance les données numériques dans la chaîne avec l'algorithme numérique. Si la correspondance des données numériques est de 100, les données alphabétiques sont mises en correspondance en utilisant la Distance d'édition et Fréquence de caractère. Le score de correspondance final est calculé ainsi :

$$(\text{numericScore} + (\text{EditDistanceScore} + \text{CharacterFrequencyScore}) / 2) / 2$$

Par exemple, le score de correspondance de ces deux adresses est de 95,5, calculé ainsi :

123 Main St Apt 567  
123 Maon St Apt 567

Numeric Score = 100  
Edit Distance = 91  
Character Frequency = 91

$$91 + 91 = 182$$

$$182/2 = 91$$

$$100 + 91 = 191$$

$$191/2 = 95,5$$

**Nysiis**

L'algorithme de code phonétique qui met en correspondance une prononciation approximative avec une orthographe exacte et indexe des mots prononcés de manière similaire. Fait partie du système New York State Identification and Intelligence System. Imaginons, par exemple, que vous recherchez des informations sur une personne dans une base de données de personnes. Vous pensez que le nom de la personne sonne comme « John Smith », mais il est en fait orthographié « Jon Smath ». Si vous procédez à une recherche de la correspondance exacte de « John Smith », aucun résultat n'est renvoyé. Cependant, si vous indexez la base de données à l'aide de l'algorithme NYSIIS et procédez à une recherche en utilisant de nouveau l'algorithme NYSIIS, la correspondance correcte est renvoyée car « John Smith » et « Jon Smath » sont indexés comme « JANSNATH » par l'algorithme. Cette option a été développée en réponse aux limites de Soundex ; elle gère certains n-grammes à caractères multiples et maintient un positionnement de voyelle relative, ce qui n'est pas le cas de Soundex.

**Remarque :** Cet algorithme ne traite pas les caractères non-alpha ; les enregistrements les contenant échoueront lors du traitement.

**Phonix**

Pré-traite les chaînes de nom en appliquant plus de 100 règles de transformation à des caractères uniques ou à des séquences de plusieurs caractères. 19 de ces règles s'appliquent uniquement si les caractères figurent au début de la chaîne, tandis que 12 des règles s'appliquent uniquement si les caractères figurent au milieu de la chaîne et 28 des règles s'appliquent uniquement si les caractères figurent à la fin de la chaîne. La chaîne de nom transformée est cryptée en un code composé d'une lettre au début, suivie de trois chiffres (en enlevant les zéros et les nombres en double). Cette option a été développée pour répondre aux limites de Soundex ; elle est plus complexe et donc plus lente que Soundex.

**Soundex**

Détermine la similitude entre deux chaînes basée sur la représentation phonétique de leurs caractères.

**Sous-chaîne**

Détermine si une chaîne figure dans une autre.

**Alignement de syllabes**

Combine des informations phonétiques avec des calculs basées sur la distance d'édition. Convertit les chaînes à comparer dans leurs séquences de syllabes correspondantes et calcule le nombre d'éditions requises pour convertir une séquence de syllabes en une autre.

Le tableau suivant décrit la relation logique entre le nombre d'algorithmes que vous pouvez utiliser suivant la méthode de notation du parent sélectionnée.

**Tableau 2 : Matrice de méthode de correspondance de type « Algorithme et Notation »**

Méthode de notation	Algorithmes	
	Seul	Multiple
Moyenne pondérée	S/O	Oui
Moyenne	S/O	Oui
Maximum	Oui	Oui
Minimum	S/O	Oui
Sommation des vecteurs	S/O	Oui

- Si vous définissez une règle dans Interflow Match, Intraflow Match ou Transactional Match, et que vous souhaitez partager la règle avec d'autres stages et/ou utilisateurs, cliquez sur le bouton **Enregistrer** en haut de la fenêtre.

### Conditions de correspondance négatives

Les conditions de correspondance sont des instructions qui indiquent les champs que vous souhaitez mettre en correspondance afin que deux enregistrements soient considérés comme une correspondance. Toutefois, dans certains cas, vous pouvez définir une condition indiquant que deux champs ne doivent *pas* correspondre afin que deux enregistrements soient considérés comme une correspondance. Cette technique, connue sous le nom de *négation*, inverse la logique d'une condition dans une règle de correspondance.

Par exemple, imaginons que vous des enregistrements d'assistance clientèle pour un centre d'appel et que vous souhaitez identifier les clients qui ont contacté le centre d'appel mais pour plusieurs comptes. En d'autres termes, vous souhaitez identifier les individus qui sont associés à plusieurs comptes. Afin d'identifier les clients qui ont plusieurs comptes, il serait judicieux de mettre en correspondance des enregistrements où les correspondances de nom, mais pas le numéro de compte, ne correspondent *pas*. Dans ce cas, vous utiliseriez la négation sur une condition de correspondance pour le numéro de compte.

Pour utiliser la négation, cochez la case **Correspondance lorsque la valeur n'est pas True** lors de la définition de votre règle de correspondance. Cette option est disponible pour les parents (groupes de conditions) et les enfants (conditions individuelles) dans la règle de correspondance. L'effet de cette option est légèrement différent lorsqu'elle est utilisée sur un parent plutôt que sur

un enfant. Quand elle est utilisée sur un parent, l'option de **Correspondance lorsque la valeur n'est pas True** inverse effectivement l'option de méthode de correspondance comme suit :

- La méthode de correspondance **Toutes vraies** devient alors « Toutes fausses ». La règle de correspondance peut uniquement mettre des enregistrements en correspondance si au moins un des enfants sous le parent a la valeur Fausse, ce qui rend donne au parent la valeur Fausse. Étant donné que l'option **Correspondance lorsque la valeur n'est pas True** est activée, cette évaluation Fausse entraînera une correspondance.
- La méthode de correspondance **Au moins une** devient alors « Aucune vraie ». La règle de correspondance peut uniquement mettre en correspondance les enregistrements où aucun des enfants n'est évalué sur True, car si les enfants sont évalués sur True (vrai), le parent sera True, mais avec l'option **Correspondance lorsque la valeur n'est pas True** est activée, cette évaluation sur True ne génèrera pas de correspondance. La règle trouvera une correspondance uniquement si aucun des enfants n'est True, ce qui génère un parent évalué sur Non true.
- La méthode de correspondance **Basé sur un seuil** passe efficacement d'enregistrements correspondants qui sont égaux ou supérieurs à un seuil déterminé à des enregistrements correspondants qui sont inférieurs à ce seuil. C'est parce que les enregistrements avec une valeur de seuil inférieure à celle indiquée seront évalués sur False, et parce que la **Correspondance lorsque la valeur n'est pas True** est activée, que cela entraînera une correspondance.

L'option **Correspondance lorsque la valeur n'est pas True** est plus facile à comprendre lorsqu'elle est appliquée aux éléments enfants dans une règle de correspondance. Cela indique simplement que deux enregistrements sont considérés comme une correspondance si l'algorithme n'indique *pas* de correspondance.

## Test d'une règle de correspondance

Après avoir défini une règle de correspondance, vous pouvez la tester pour en voir les résultats. Pour ce faire, vous pouvez utiliser Match Rule Evaluation pour examiner les effets d'une règle de correspondance sur un petit ensemble de données d'exemple.

1. Ouvrez le flux de données dans Enterprise Designer.
2. Double-cliquez sur le stage contenant la règle de correspondance que vous souhaitez tester.

Les règles de correspondance sont utilisées dans Interflow Match, Intraflow Match et Transactional Match.

3. Dans la hiérarchie de règle de correspondance, sélectionnez le noeud à tester et cliquez sur **Évaluer**.
4. Sous l'onglet **Importer**, entrez les données test (un suspect et jusqu'à 10 candidats). Il existe deux façons de saisir des données test.
  - Pour taper manuellement dans les données de test, saisissez un enregistrement de suspect sous **Suspect** et jusqu'à dix candidats sous **Candidat**. Après avoir saisi les enregistrements, vous pouvez cliquer sur **Exporter** pour enregistrer les enregistrements dans un fichier que vous pouvez importer plus tard au lieu de saisir de nouveau les données manuellement.

- Pour importer les données test à partir d'un fichier, cliquez sur **Importer...** et sélectionnez le fichier contenant les exemples d'enregistrements. Les fichiers délimités peuvent être délimités par une virgule, par une barre verticale ou par une tabulation et devrait avoir un enregistrement d'en-tête avec des champs d'en-tête correspondant aux noms de champs affichés dans **Candidats**. Un échantillon d'en-tête d'enregistrement pour l'entrée Ménage serait :

```
Name,AddressLine1,City,StateProvince
```

5. Évaluez la règle à l'aide de l'une de ces méthodes :

- Cliquez sur **Règle actuelle**. Cela exécute la règle définie dans l'onglet **Règle de correspondance**. Les résultats s'affichent pour une paire de suspect et de candidat à la fois. Pour naviguer parmi les résultats, cliquez sur les flèches. Les scores des champs et des algorithmes sont affichés dans un format d'arbre semblable aux contrôles de la règle de correspondance. Facultativement, les résultats peuvent être exportés sous un fichier XML.

**Remarque** : Si vous apportez des modifications à la règle de correspondance et que vous souhaitez appliquer les modifications à la règle de correspondance du stage, cliquez sur **Enregistrer**.

- Cliquez sur **Tous les algorithmes**. Cela ignore la règle de correspondance et au lieu de cela exécute tous les algorithmes sur chaque champ de paires de suspect et de candidat. Les résultats s'affichent pour une paire de suspect et de candidat à la fois, et il est possible de consulter les résultats en utilisant les flèches.

Pour automatiquement mettre à jour les résultats alors que vous effectuez des changements sur la règle de correspondance et l'entrée, cochez la case **Mises à jour automatiques**. Lorsque vous utilisez cette fonctionnalité dans l'option **Tous les algorithmes**, seuls les changements apportés à l'entrée mettront les résultats à jour.

Les résultats affichés dans Scores suivent le code couleur suivant :

- Vert—La règle a produit une correspondance.
- Rouge—La règle n'a pas produit de correspondance.
- Gris—La règle a été ignorée
- Bleu—Les résultats des algorithmes individuels respectant la règle.

Pour exporter les résultats d'évaluation dans un format XML, cliquez sur **Exporter**.

## Partage d'une règle de correspondance

Vous pouvez créer des règles de correspondance qui peuvent être partagées entre modules, stages, flux de données et utilisateurs. En partageant une règle de correspondance, vous pouvez développer plus facilement des flux de données en définissant une règle de correspondance, puis en la référençant là où cela est nécessaire. Cela permet également de garantir la cohérence des règles de correspondance destinées à exécuter les mêmes fonctions dans les différents flux de données.

Les règles de correspondance enregistrées dans le référentiel de règles de correspondance peuvent être exportées sous forme de fichiers .mr ou .json à l'aide de l'utilitaire Administration. Cela est particulièrement utile lorsque vous tentez de partager ces règles avec d'autres installations Spectrum ou lors de l'utilisation de ces éléments dans le SDK Spectrum Qualité des Big Data.

**Remarque :** Par défaut, les règles de correspondance sont exportées sous forme de fichier de .mr.

- Pour partager une règle de correspondance que vous avez créée dans Interflow Match, Intraflow Match ou Transactional Match, cliquez sur le bouton **Enregistrer** en haut de la fenêtre des options du stage.
- Si vous créez la règle dans l'outil Gestion des règles de correspondance, la règle est automatiquement mise à disposition de tous les utilisateurs pour l'utiliser dans les flux de données. Pour afficher l'outil Gestion des règles de correspondance, dans Enterprise Designer, sélectionnez **Outils > Gestion des règles de correspondance**.

## Affichage des règles de correspondance partagées

Dans Enterprise Designer, vous pouvez parcourir toutes les règles de correspondance partagées, disponibles sur votre système Spectrum™ Technology Platform. Ces règles de correspondance peuvent être utilisées par les stages Interflow Match, Intraflow Match et Transactional Match dans un flux de données pour effectuer un rapprochement.

Pour parcourir les règles de correspondance dans le référentiel de règle de correspondance, suivez la procédure suivante.

1. Ouvrez Enterprise Designer.
2. Sélectionnez **Outils > Gestion des règles de correspondance**.
3. Sélectionnez la règle à afficher et cliquez sur **Afficher**.

## Création d'un règle de rapprochement personnalisée sous forme d'objet JSON

Les règles de rapprochement peuvent être configurées et transmises lors de l'exécution si elles sont exposées sous forme d'options de flux de données. Cela vous permet de partager des règles de rapprochement entre différents ordinateurs et d'écraser les règles de rapprochement existantes par des chaînes de règles de rapprochement au format JSON. Vous pouvez également définir des options de stage lors de l'appel d'un job via un flux de processus ou l'outil de ligne de commande Exécuteur de job.

Vous pouvez rechercher des schémas MatchRule et MatchInfo dans le dossier suivant :

```
<Spectrum Location>\server\modules\jsonSchemas\matcher.
```

1. Enregistrez et exposez le flux de données contenant la règle de rapprochement.

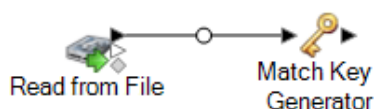
2. Ouvrez le flux de données utilisant la règle de rapprochement.
3. Allez dans `Edit > Dataflow Options`.
4. Dans la table **Map dataflow options to stages**; cliquez sur le stage de rapprochement utilisant la règle de rapprochement et cochez la case **Règle de rapprochement personnalisée**.
5. Facultatif : remplacez le nom de la règle de rapprochement « Règle de rapprochement personnalisée » dans le champ **Libellé d'option** par le nom de votre choix.
6. Cliquez deux fois sur **OK**.

## Mise en correspondance d'enregistrements d'une source unique

Cette procédure explique comment utiliser un stage Intraflow Match pour identifier les groupes d'enregistrements dans une source de données unique (telle qu'un fichier ou une table de base de données) reliés entre eux, en fonction des critères de correspondance que vous indiquez. Le flux de données regroupe les enregistrements dans des collections et écrit ces collections dans un fichier de sortie.

1. Dans Enterprise Designer, créez un flux de données.
2. Faites glisser un stage source sur le canevas.
3. Double-cliquez sur le stage source et configurez-le. Pour obtenir les instructions sur la configuration des stages source, reportez-vous au *Guide du concepteur du flux de données*.
4. Faites glisser un stage Match Key Generator sur le canevas et connectez-le au stage source.

Par exemple, si vous utilisez un stage source Read from File, votre flux de données se présenterait désormais comme suit :



Match Key Generator crée une clé non unique pour chaque enregistrement, qui peut ensuite être utilisée par les stages de rapprochement pour identifier les groupes d'enregistrements doublons potentiels. Les match keys facilitent la procédure de correspondance en vous permettant de regrouper les enregistrements par match key, puis de ne comparer les enregistrements que dans ces groupes.

5. Double-cliquez sur le Match Key Generator.
6. Cliquez sur **Ajouter**.
7. Définissez la règle à utiliser pour générer une match key pour chaque enregistrement.

**Tableau 3 : Options de Match Key Generator**

Nom de l'option	Description/Valeurs valides
-----------------	-----------------------------

---

Algorithme	
------------	--



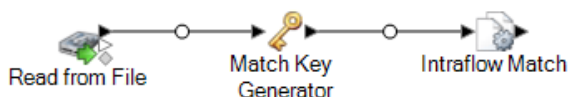
Nom de l'option	Description/Valeurs valides
	Définit l'algorithme à utiliser pour générer la clé de correspondance. L'un des éléments suivants :
<b>Consonne</b>	Renvoie les champs indiqués, les consonnes étant supprimées.
<b>Metaphone double</b>	Renvoie un code basé sur la représentation phonétique de leurs caractères. Le double Metaphone est une version améliorée de l'algorithme Metaphone et tente de prendre en compte les nombreuses irrégularités de plusieurs langues.
<b>Koeln</b>	Noms d'index par son, tels qu'ils sont prononcés en allemand. Permet aux noms ayant la même prononciation d'être encodés avec la même représentation afin qu'ils puissent être mis en correspondance, en dépit de différences mineures au niveau de l'orthographe. Le résultat est toujours une séquence de nombres ; les caractères spéciaux et les espaces blancs sont ignorés. Cette option a été développée en réponse aux limites du Soundex.
<b>MD5</b>	Algorithme qui produit une valeur hash de 128 bits. Cet algorithme est généralement utilisé pour vérifier l'intégrité des données.
<b>Metaphone</b>	Renvoie une clé codée Metaphone des champs sélectionnés. Metaphone est un algorithme qui code les mots à l'aide de leur sonorité lorsque prononcé en anglais.
<b>Metaphone (Espagnol)</b>	Renvoie une clé codée Metaphone des champs sélectionnés pour la langue espagnole. Cet algorithme Metaphone code les mots à l'aide de leur sonorité lorsque prononcé en espagnol.
<b>Metaphone3</b>	Procède à une amélioration en fonction des algorithmes Metaphone et Double Metaphone avec des paramètres de consonne et de voyelle interne exacts qui vous permet de produire des mots ou des noms mis en correspondance de manière plus ou moins proche pour rechercher des termes au niveau phonétique. Metaphone 3 augmente l'exactitude de l'encodage phonétique à 98 %. Cette option a été développée en réponse aux limites du Soundex.
<b>Nysiis</b>	L'algorithme de code phonétique qui met en correspondance une prononciation approximative avec une orthographe exacte et indexe des mots prononcés de manière similaire. Fait partie du système New York State Identification and Intelligence System. Imaginons, par exemple, que vous recherchez des informations sur une personne dans une base de données de personnes. Vous pensez que le nom de la personne sonne comme « John Smith », mais il est en fait orthographié « Jon Smyth ». Si vous procédez à une recherche de la correspondance exacte de « John Smith », aucun résultat n'est renvoyé. Cependant, si vous indexez la base de données à l'aide de l'algorithme NYSIIS et procédez à une recherche en utilisant de nouveau l'algorithme NYSIIS, la correspondance correcte est renvoyée car « John Smith » et « Jon Smyth » sont indexés comme « JAN SNATH » par l'algorithme.
<b>Phonix</b>	Pré-traite les chaînes de nom en appliquant plus de 100 règles

Nom de l'option	Description/Valeurs valides
	<p>de transformation à des caractères uniques ou à des séquences de plusieurs caractères. 19 de ces règles s'appliquent uniquement si les caractères figurent au début de la chaîne, tandis que 12 des règles s'appliquent uniquement si les caractères figurent au milieu de la chaîne et 28 des règles s'appliquent uniquement si les caractères figurent à la fin de la chaîne. La chaîne de nom transformée est cryptée en un code composé d'une lettre au début, suivie de trois chiffres (en enlevant les zéros et les nombres en double). Cette option a été développée pour répondre aux limites de Soundex ; elle est plus complexe et donc plus lente que Soundex.</p> <p><b>Soundex</b> Renvoie un code Soundex des champs sélectionnés. Soundex produit un code de longueur fixe en s'appuyant sur la sonorité du mot lorsque prononcé en anglais.</p> <p><b>Substring</b> Renvoie une partie spécifiée du champ sélectionné.</p>
Nom de champ	Indique le champ auquel vous souhaitez appliquer l'algorithme sélectionné pour générer la match key. Par exemple, si vous sélectionnez un champ appelé LastName et que vous choisissez l'algorithme Soundex, ce dernier est appliqué aux données dans le champ LastName pour produire une match key.
Position de début	Spécifie la position de départ dans le champ spécifié. Tous les algorithmes ne permettent pas de spécifier une position de départ.
Longueur	Spécifie la longueur de caractère à inclure à partir de la position de départ. Tous les algorithmes ne permettent pas de spécifier une longueur.
Supprimer les caractères parasites :	Supprime tout caractère non numérique et non alphanumérique, comme les traits d'union, les espaces blancs, et autres caractères spéciaux du champ d'entrée.
Trier les entrées :	<p>Trie tous les caractères dans un champ d'entrée ou tous les termes dans un champ d'entrée par ordre alphabétique.</p> <p><b>Caractères</b> Trie les valeurs de caractères d'un champ d'entrée avant de créer un identifiant unique.</p> <p><b>Termes</b> Trie les valeurs de termes d'un champ d'entrée avant de créer un identifiant unique.</p>

8. Lorsque vous avez terminé de définir la règle, cliquez sur **OK**.

9. Pour ajouter d'autres règles de correspondance, cliquez sur **Ajouter** et ajoutez-les, sinon cliquez sur **OK** lorsque vous avez terminé.
10. Faites glisser un stage Intraflow Match sur le canevas et connectez-le au stage Match Key Generator.

Par exemple, si vous utilisez un stage source Read from File, votre flux de données se présenterait désormais comme suit :



11. Double-cliquez sur Intraflow Match.
12. Dans le champ **Charger une règle de rapprochement**, sélectionnez une des règles de correspondance prédéfinies que vous pouvez utiliser tel quel ou modifier pour répondre à vos besoins. Si vous souhaitez créer une nouvelle règle de correspondance sans utiliser une des règles de correspondance prédéfinies comme point de départ, cliquez sur **Nouveau**. Vous ne pouvez disposer que d'une règle personnalisée dans un flux de données.

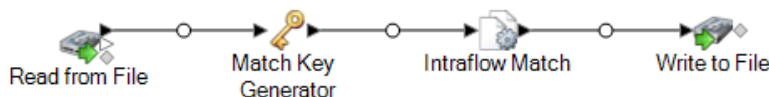
**Remarque :** La fonction Options de flux de données dans Enterprise Designer permet d'afficher la règle de correspondance pour la configuration au moment de l'exécution.

13. Dans le champ **Grouper par**, sélectionnez **Match Key**.

Cette opération aura pour effet de placer les enregistrements disposant de la même match key dans un groupe. La règle de correspondance s'applique aux enregistrements d'un groupe pour vérifier s'ils sont des doublons. La match key de chaque enregistrement est générée par le stage Generate Match Key que vous avez configuré précédemment dans cette procédure.

14. Pour plus d'informations sur la modification des autres options, voir [Création d'une règle de correspondance](#) à la page 73.
15. Cliquez sur **OK** pour enregistrer votre configuration Intraflow Match et revenir au canevas du flux de données.
16. Faites glisser un stage de collecteur de données sur le canevas et connectez-le au stage Generate Match key.

Par exemple, si vous utilisiez un stage de collecteur de données Write To File, votre flux de données se présenterait comme suit :



17. Double-cliquez sur le stage de collecteur de données et configurez-le.

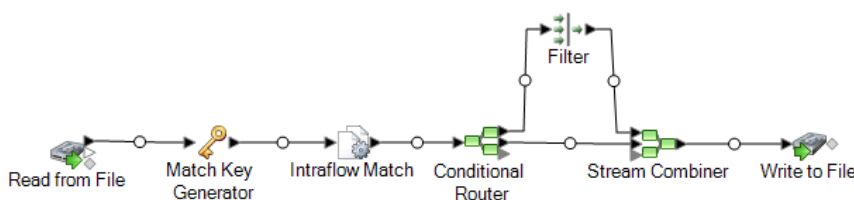
Pour obtenir des informations sur la configuration des stages de collecteur de données, reportez-vous au *Guide du concepteur de flux de données*.

Vous disposez désormais d'un flux de données qui met en correspondance les enregistrements d'une source unique.

### Exemple de mise en correspondance d'enregistrements dans une source de données unique

En tant que data steward pour une nouvelle société de carte de crédit, vous souhaitez analyser la base de données de vos clients et déterminer quelles adresses ont plusieurs occurrences et sous quel noms, afin que vous puissiez minimiser le nombre d'offres de carte de crédit dupliquées, envoyées au même foyer.

Cet exemple expose la façon d'identifier les membres d'un même foyer en comparant les informations d'un champ d'entrée unique et en créant un fichier de sortie contenant un enregistrement par foyer.



Le stage Read from File lit les données contenant des enregistrements uniques pour chaque foyer et des enregistrements éventuellement issus du même foyer. Le fichier d'entrée contient des noms et des adresses.

Match Key Generator crée une match key qui est une clé non unique partagée par des enregistrements semblables identifiant les enregistrements comme des doublons.

Le stage Intraflow Match compare les enregistrements présentant la même match key et marque chaque enregistrement comme un enregistrement unique ou comme l'un des différents enregistrements pour le même foyer.

Conditional Router envoie des enregistrements qui composent des collections d'enregistrements pour chaque foyer au stage Filter, qui filtre un seul enregistrement pour chaque foyer et l'envoie au stage Stream Combiner. Le stage Conditional Router envoie également des enregistrements uniques directement à Stream Combiner.

Enfin, le stage Write to File crée un fichier de sortie contenant un enregistrement pour chaque foyer.

## Mise en correspondance d'enregistrements d'une source avec ceux d'une autre source

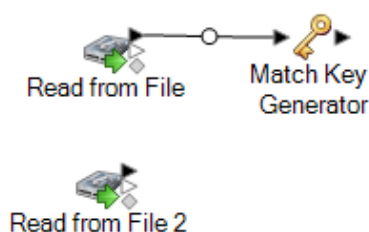
Cette procédure explique comment utiliser un stage Interflow Match pour identifier les enregistrements d'une source correspondant à ceux d'une autre source. La première source contient des enregistrements suspects et la deuxième source contient les enregistrements candidats. Le flux de données ne met en correspondance que les enregistrements d'une source avec ceux d'une autre source. Il n'essaie pas de mettre en correspondance des enregistrements issus de la même source. Le flux de données regroupe les enregistrements dans des collections d'enregistrements correspondants et écrit ces collections dans un fichier de sortie.

1. Dans Enterprise Designer, créez un flux de données.
2. Faites glisser deux stages source sur le canevas. Configurez l'un d'eux de telle sorte qu'il pointe vers la source des enregistrements suspects et configurez l'autre pour qu'il pointe vers la source des enregistrements candidats.

Pour obtenir les instructions sur la configuration des stages source, reportez-vous au *Guide du concepteur du flux de données*.

3. Faites glisser un stage Match Key Generator sur le canevas et connectez-le à l'un des stages source.

Par exemple, si vous utilisez un stage source Read from File, votre flux de données se présenterait désormais comme suit :



Match Key Generator crée une clé non unique pour chaque enregistrement, qui peut ensuite être utilisée par les stages de rapprochement pour identifier les groupes d'enregistrements doublons potentiels. Les match keys facilitent la procédure de correspondance en vous permettant de regrouper les enregistrements par match key, puis de ne comparer les enregistrements que dans ces groupes.

**Remarque :** Vous ajouterez un second stage Match Key Generator ultérieurement. Pour l'instant, vous n'en avez besoin que d'un sur le canevas.

4. Double-cliquez sur le stage Match Key Generator.
5. Cliquez sur **Ajouter**.

6. Définissez la règle à utiliser pour générer une match key pour chaque enregistrement.

**Tableau 4 : Options de Match Key Generator**

Nom de l'option	Description/Valeurs valides
-----------------	-----------------------------

---

Algorithme	
------------	--

Nom de l'option	Description/Valeurs valides
	Définit l'algorithme à utiliser pour générer la clé de correspondance. L'un des éléments suivants :
<b>Consonne</b>	Renvoie les champs indiqués, les consonnes étant supprimées.
<b>Metaphone double</b>	Renvoie un code basé sur la représentation phonétique de leurs caractères. Le double Metaphone est une version améliorée de l'algorithme Metaphone et tente de prendre en compte les nombreuses irrégularités de plusieurs langues.
<b>Koeln</b>	Noms d'index par son, tels qu'ils sont prononcés en allemand. Permet aux noms ayant la même prononciation d'être encodés avec la même représentation afin qu'ils puissent être mis en correspondance, en dépit de différences mineures au niveau de l'orthographe. Le résultat est toujours une séquence de nombres ; les caractères spéciaux et les espaces blancs sont ignorés. Cette option a été développée en réponse aux limites du Soundex.
<b>MD5</b>	Algorithme qui produit une valeur hash de 128 bits. Cet algorithme est généralement utilisé pour vérifier l'intégrité des données.
<b>Metaphone</b>	Renvoie une clé codée Metaphone des champs sélectionnés. Metaphone est un algorithme qui code les mots à l'aide de leur sonorité lorsque prononcé en anglais.
<b>Metaphone (Espagnol)</b>	Renvoie une clé codée Metaphone des champs sélectionnés pour la langue espagnole. Cet algorithme Metaphone code les mots à l'aide de leur sonorité lorsque prononcé en espagnol.
<b>Metaphone3</b>	Procède à une amélioration en fonction des algorithmes Metaphone et Double Metaphone avec des paramètres de consonne et de voyelle interne exacts qui vous permet de produire des mots ou des noms mis en correspondance de manière plus ou moins proche pour rechercher des termes au niveau phonétique. Metaphone 3 augmente l'exactitude de l'encodage phonétique à 98 %. Cette option a été développée en réponse aux limites du Soundex.
<b>Nysiis</b>	L'algorithme de code phonétique qui met en correspondance une prononciation approximative avec une orthographe exacte et indexe des mots prononcés de manière similaire. Fait partie du système New York State Identification and Intelligence System. Imaginons, par exemple, que vous recherchez des informations sur une personne dans une base de données de personnes. Vous pensez que le nom de la personne sonne comme « John Smith », mais il est en fait orthographié « Jon Smyth ». Si vous procédez à une recherche de la correspondance exacte de « John Smith », aucun résultat n'est renvoyé. Cependant, si vous indexez la base de données à l'aide de l'algorithme NYSIIS et procédez à une recherche en utilisant de nouveau l'algorithme NYSIIS, la correspondance correcte est renvoyée car « John Smith » et « Jon Smyth » sont indexés comme « JAN SNATH » par l'algorithme.
<b>Phonix</b>	Pré-traite les chaînes de nom en appliquant plus de 100 règles

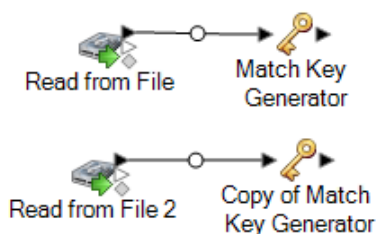


Nom de l'option	Description/Valeurs valides
	<p>de transformation à des caractères uniques ou à des séquences de plusieurs caractères. 19 de ces règles s'appliquent uniquement si les caractères figurent au début de la chaîne, tandis que 12 des règles s'appliquent uniquement si les caractères figurent au milieu de la chaîne et 28 des règles s'appliquent uniquement si les caractères figurent à la fin de la chaîne. La chaîne de nom transformée est cryptée en un code composé d'une lettre au début, suivie de trois chiffres (en enlevant les zéros et les nombres en double). Cette option a été développée pour répondre aux limites de Soundex ; elle est plus complexe et donc plus lente que Soundex.</p> <p><b>Soundex</b> Renvoie un code Soundex des champs sélectionnés. Soundex produit un code de longueur fixe en s'appuyant sur la sonorité du mot lorsque prononcé en anglais.</p> <p><b>Substring</b> Renvoie une partie spécifiée du champ sélectionné.</p>
Nom de champ	Indique le champ auquel vous souhaitez appliquer l'algorithme sélectionné pour générer la match key. Par exemple, si vous sélectionnez un champ appelé LastName et que vous choisissez l'algorithme Soundex, ce dernier est appliqué aux données dans le champ LastName pour produire une match key.
Position de début	Spécifie la position de départ dans le champ spécifié. Tous les algorithmes ne permettent pas de spécifier une position de départ.
Longueur	Spécifie la longueur de caractère à inclure à partir de la position de départ. Tous les algorithmes ne permettent pas de spécifier une longueur.
Supprimer les caractères parasites :	Supprime tout caractère non numérique et non alphanumérique, comme les traits d'union, les espaces blancs, et autres caractères spéciaux du champ d'entrée.
Trier les entrées :	<p>Trie tous les caractères dans un champ d'entrée ou tous les termes dans un champ d'entrée par ordre alphabétique.</p> <p><b>Caractères</b> Trie les valeurs de caractères d'un champ d'entrée avant de créer un identifiant unique.</p> <p><b>Termes</b> Trie les valeurs de termes d'un champ d'entrée avant de créer un identifiant unique.</p>

7. Lorsque vous avez terminé de définir la règle, cliquez sur **OK**.

8. Faites un clic droit sur le stage Match Key Generator sur le canevas et sélectionnez **Copier le stage**.
9. Faites un clic droit dans une zone vide du canevas et sélectionnez **Coller**.
10. Connectez la copie du Match Key Generator à l'autre stage source.

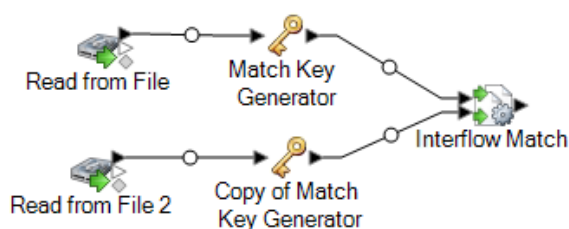
Par exemple, si vous utilisez des stages d'entrée Read from File, votre flux de données se présenterait désormais comme suit :



Le flux de données contient désormais deux stages Match Key Generator produisant des match keys pour toutes les sources utilisant exactement les mêmes règles. Il est essentiel que les stages Match Key Generator soient configurés de manière identique pour que ce flux de données fonctionne correctement.

11. Faites glisser un stage Interflow Match sur le canevas et connectez chacun des stages Match Key Generator à celui-ci.

Par exemple, si vous utilisez des stages d'entrée Read from File, votre flux de données se présenterait désormais comme suit :



12. Double-cliquez sur le stage Interflow Match.
13. Dans le champ **Charger une règle de rapprochement**, sélectionnez une des règles de correspondance prédéfinies que vous pouvez utiliser tel quel ou modifier pour répondre à vos besoins. Si vous souhaitez créer une nouvelle règle de correspondance sans utiliser une des règles de correspondance prédéfinies comme point de départ, cliquez sur **Nouveau**. Vous ne pouvez disposer que d'une règle personnalisée dans un flux de données.

**Remarque :** La fonction Options de flux de données dans Enterprise Designer permet d'afficher la règle de correspondance pour la configuration au moment de l'exécution.

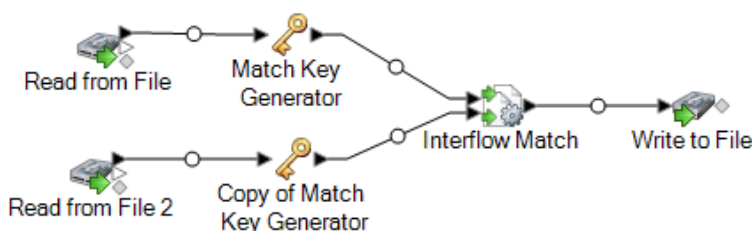
14. Dans le champ **Grouper par**, sélectionnez **Match Key**.

Cette opération aura pour effet de placer les enregistrements disposant de la même match key dans un groupe. La règle de correspondance s'applique aux enregistrements d'un groupe pour

vérifier s'ils sont des doublons. La match key de chaque enregistrement est générée par les stages Generate Match Key que vous avez configurés précédemment dans cette procédure.

15. Pour plus d'informations sur la modification des autres options, voir [Création d'une règle de correspondance](#) à la page 73.
16. Faites glisser un stage de collecteur de données sur le canevas et connectez-le au stage Interflow Match.

Par exemple, si vous utilisiez un stage de collecteur de données Write To File, votre flux de données se présenterait comme suit :



17. Double-cliquez sur le stage de collecteur de données et configurez-le.

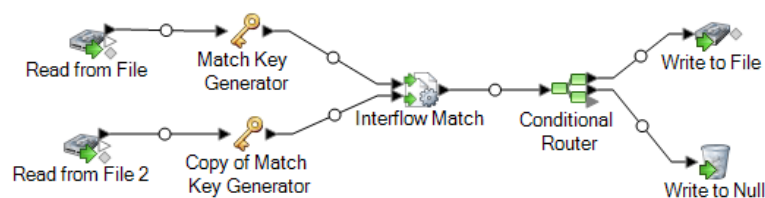
Pour obtenir des informations sur la configuration des stages de collecteur de données, reportez-vous au *Guide du concepteur de flux de données*.

Vous disposez désormais d'un flux de données qui met en correspondance les enregistrements de deux sources de données.

### Exemple de mise en correspondance d'enregistrements de plusieurs sources

En tant qu'entreprise de publipostage direct, vous souhaitez identifier les personnes figurant sur une liste d'expédition interdite, afin de ne pas leur envoyer de courrier direct. Vous disposez d'une liste de destinataires dans un fichier et d'une liste de personnes ne souhaitant pas recevoir de courrier marketing direct dans un autre fichier (fichier de suppression).

Le flux de données suivant offre une solution à ce scénario commercial :



Le stage Read from File lit les données de votre liste de publipostage et le stage Read from File 2 lit les données de la liste de suppression. Les deux stages Match Key Generator sont configurés de manière identique, afin de produire une match key qui peut être utilisée par Interflow Match pour former des groupes de correspondances potentielles. Interflow Match identifie les enregistrements de la liste de publipostage figurant également dans la liste de suppression et marque ces enregistrements

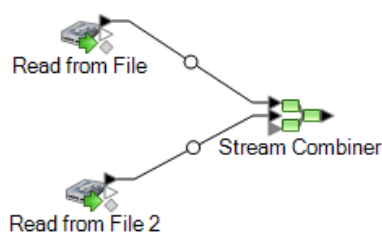
comme des doublons. Conditional Router envoie des enregistrements uniques, c'est-à-dire les enregistrements non trouvés dans la liste de suppression, au stage Write to File pour que celui-ci les écrive dans un fichier. Le stage Conditional Router envoie tous les autres enregistrements à Write to Null, où ils sont rejetés.

## Mise en correspondance d'enregistrements entre et dans des sources

Cette procédure explique comment utiliser un stage Intraflow Match pour identifier les enregistrements d'un fichier correspondant à ceux d'un autre fichier et à d'autres enregistrements du même fichier. Par exemple, vous disposez de deux fichiers (fichier A et fichier B) et vous souhaitez savoir si des enregistrements du fichier A correspondent à des enregistrements du fichier B, mais aussi si des enregistrements du fichier A correspondent à d'autres enregistrements du fichier A. Vous pouvez y parvenir en utilisant un stage Stream Combiner et un stage Intraflow Match.

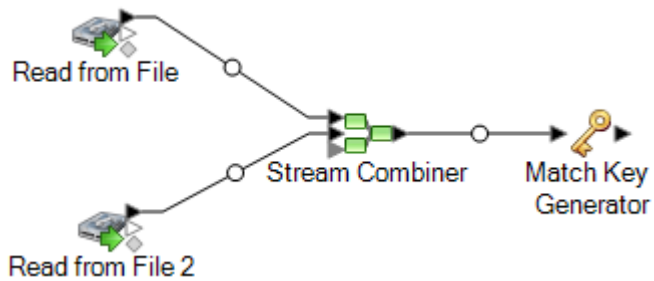
1. Dans Enterprise Designer, créez un flux de données.
2. Faites glisser un stage source sur le canevas.
3. Double-cliquez sur le stage source et configurez-le. Pour obtenir les instructions sur la configuration des stages source, reportez-vous au *Guide du concepteur du flux de données*.
4. Faites glisser un deuxième stage source sur le canevas et configurez-le pour lire la deuxième source de données dans le flux de données.
5. Faites glisser un stage Stream Combiner sur le canevas et reliez-le aux deux stages source.

Par exemple, si votre flux de données dispose de deux stages Read from File, il ressemblerait à ce qui suit après l'ajout de Stream Combiner :



6. Faites glisser un stage Match Key Generator sur le canevas et connectez-le au stage Stream Combiner.

Par exemple, votre flux de données peut se présenter comme suit :



Match Key Generator crée une clé non unique pour chaque enregistrement, qui peut ensuite être utilisée par les stages de rapprochement pour identifier les groupes d'enregistrements doublons potentiels. Les match keys facilitent la procédure de correspondance en vous permettant de regrouper les enregistrements par match key, puis de ne comparer les enregistrements que dans ces groupes.

7. Double-cliquez sur le Match Key Generator.
8. Cliquez sur **Ajouter**.
9. Définissez la règle à utiliser pour générer une match key pour chaque enregistrement.

**Tableau 5 : Options de Match Key Generator**

Nom de l'option	Description/Valeurs valides
-----------------	-----------------------------

---

Algorithme	
------------	--

Nom de l'option	Description/Valeurs valides
	Définit l'algorithme à utiliser pour générer la clé de correspondance. L'un des éléments suivants :
<b>Consonne</b>	Renvoie les champs indiqués, les consonnes étant supprimées.
<b>Metaphone double</b>	Renvoie un code basé sur la représentation phonétique de leurs caractères. Le double Metaphone est une version améliorée de l'algorithme Metaphone et tente de prendre en compte les nombreuses irrégularités de plusieurs langues.
<b>Koeln</b>	Noms d'index par son, tels qu'ils sont prononcés en allemand. Permet aux noms ayant la même prononciation d'être encodés avec la même représentation afin qu'ils puissent être mis en correspondance, en dépit de différences mineures au niveau de l'orthographe. Le résultat est toujours une séquence de nombres ; les caractères spéciaux et les espaces blancs sont ignorés. Cette option a été développée en réponse aux limites du Soundex.
<b>MD5</b>	Algorithme qui produit une valeur hash de 128 bits. Cet algorithme est généralement utilisé pour vérifier l'intégrité des données.
<b>Metaphone</b>	Renvoie une clé codée Metaphone des champs sélectionnés. Metaphone est un algorithme qui code les mots à l'aide de leur sonorité lorsque prononcé en anglais.
<b>Metaphone (Espagnol)</b>	Renvoie une clé codée Metaphone des champs sélectionnés pour la langue espagnole. Cet algorithme Metaphone code les mots à l'aide de leur sonorité lorsque prononcé en espagnol.
<b>Metaphone3</b>	Procède à une amélioration en fonction des algorithmes Metaphone et Double Metaphone avec des paramètres de consonne et de voyelle interne exacts qui vous permet de produire des mots ou des noms mis en correspondance de manière plus ou moins proche pour rechercher des termes au niveau phonétique. Metaphone 3 augmente l'exactitude de l'encodage phonétique à 98 %. Cette option a été développée en réponse aux limites du Soundex.
<b>Nysiis</b>	L'algorithme de code phonétique qui met en correspondance une prononciation approximative avec une orthographe exacte et indexe des mots prononcés de manière similaire. Fait partie du système New York State Identification and Intelligence System. Imaginons, par exemple, que vous recherchez des informations sur une personne dans une base de données de personnes. Vous pensez que le nom de la personne sonne comme « John Smith », mais il est en fait orthographié « Jon Smyth ». Si vous procédez à une recherche de la correspondance exacte de « John Smith », aucun résultat n'est renvoyé. Cependant, si vous indexez la base de données à l'aide de l'algorithme NYSIIS et procédez à une recherche en utilisant de nouveau l'algorithme NYSIIS, la correspondance correcte est renvoyée car « John Smith » et « Jon Smyth » sont indexés comme « JAN SNATH » par l'algorithme.
<b>Phonix</b>	Pré-traite les chaînes de nom en appliquant plus de 100 règles

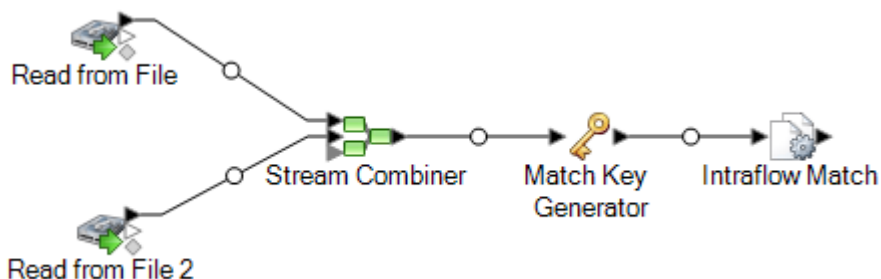
Nom de l'option	Description/Valeurs valides
	<p>de transformation à des caractères uniques ou à des séquences de plusieurs caractères. 19 de ces règles s'appliquent uniquement si les caractères figurent au début de la chaîne, tandis que 12 des règles s'appliquent uniquement si les caractères figurent au milieu de la chaîne et 28 des règles s'appliquent uniquement si les caractères figurent à la fin de la chaîne. La chaîne de nom transformée est cryptée en un code composé d'une lettre au début, suivie de trois chiffres (en enlevant les zéros et les nombres en double). Cette option a été développée pour répondre aux limites de Soundex ; elle est plus complexe et donc plus lente que Soundex.</p> <p><b>Soundex</b> Renvoie un code Soundex des champs sélectionnés. Soundex produit un code de longueur fixe en s'appuyant sur la sonorité du mot lorsque prononcé en anglais.</p> <p><b>Substring</b> Renvoie une partie spécifiée du champ sélectionné.</p>
Nom de champ	Indique le champ auquel vous souhaitez appliquer l'algorithme sélectionné pour générer la match key. Par exemple, si vous sélectionnez un champ appelé LastName et que vous choisissez l'algorithme Soundex, ce dernier est appliqué aux données dans le champ LastName pour produire une match key.
Position de début	Spécifie la position de départ dans le champ spécifié. Tous les algorithmes ne permettent pas de spécifier une position de départ.
Longueur	Spécifie la longueur de caractère à inclure à partir de la position de départ. Tous les algorithmes ne permettent pas de spécifier une longueur.
Supprimer les caractères parasites :	Supprime tout caractère non numérique et non alphanumérique, comme les traits d'union, les espaces blancs, et autres caractères spéciaux du champ d'entrée.
Trier les entrées :	<p>Trie tous les caractères dans un champ d'entrée ou tous les termes dans un champ d'entrée par ordre alphabétique.</p> <p><b>Caractères</b> Trie les valeurs de caractères d'un champ d'entrée avant de créer un identifiant unique.</p> <p><b>Termes</b> Trie les valeurs de termes d'un champ d'entrée avant de créer un identifiant unique.</p>

10. Lorsque vous avez terminé de définir la règle, cliquez sur **OK**.



11. Pour ajouter d'autres règles de correspondance, cliquez sur **Ajouter** et ajoutez-les, sinon cliquez sur **OK** lorsque vous avez terminé.
12. Faites glisser un stage Intraflow Match sur le canevas et connectez-le au stage Match Key Generator.

Par exemple, votre flux de données peut se présenter comme suit :



13. Double-cliquez sur Intraflow Match.
14. Dans le champ **Charger une règle de rapprochement**, sélectionnez une des règles de correspondance prédéfinies que vous pouvez utiliser tel quel ou modifier pour répondre à vos besoins. Si vous souhaitez créer une nouvelle règle de correspondance sans utiliser une des règles de correspondance prédéfinies comme point de départ, cliquez sur **Nouveau**. Vous ne pouvez disposer que d'une règle personnalisée dans un flux de données.

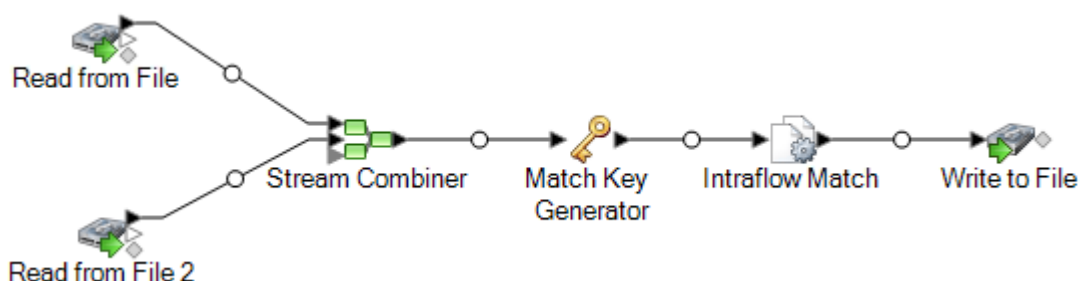
**Remarque :** La fonction Options de flux de données dans Enterprise Designer permet d'afficher la règle de correspondance pour la configuration au moment de l'exécution.

15. Dans le champ **Grouper par**, sélectionnez **Match Key**.

Cette opération aura pour effet de placer les enregistrements disposant de la même match key dans un groupe. La règle de correspondance s'applique aux enregistrements d'un groupe pour vérifier s'ils sont des doublons. La match key de chaque enregistrement est générée par le stage Generate Match Key que vous avez configuré précédemment dans cette procédure.

16. Pour plus d'informations sur la modification des autres options, voir [Création d'une règle de correspondance](#) à la page 73.
17. Cliquez sur **OK** pour enregistrer votre configuration Intraflow Match et revenir au canevas du flux de données.
18. Faites glisser un stage de collecteur de données sur le canevas et connectez-le au stage Generate Match key.

Par exemple, si vous utilisiez un stage de collecteur de données Write To File, votre flux de données se présenterait comme suit :



19. Double-cliquez sur le stage de collecteur de données et configurez-le.

Pour obtenir des informations sur la configuration des stages de collecteur de données, reportez-vous au *Guide du concepteur de flux de données*.

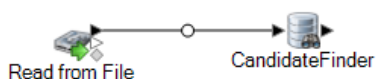
## Mise en correspondance d'enregistrements par rapport à une base de données

Cette procédure explique comment mettre en correspondance des enregistrements, dont les enregistrements suspects proviennent d'une source, telle qu'un fichier ou une base de données, et les enregistrements candidats figurent dans une base de données avec d'autres enregistrements non liés. Pour chaque enregistrement d'entrée, le flux de données interroge la base de données pour trouver les candidats de l'enregistrement, puis utilise un stage Transactional Match pour mettre en correspondance les enregistrements. Enfin, le flux de données écrit des collections d'enregistrements correspondants dans un fichier de sortie.

**Remarque :** Transactional Match ne met en correspondance les enregistrements suspects qu'avec les candidats. Il ne tente pas de mettre en correspondance les enregistrements suspects avec d'autres enregistrements suspects, comme cela est le cas dans Intraflow Match.

1. Dans Enterprise Designer, créez un flux de données.
2. Faites glisser un stage source sur le canevas.
3. Double-cliquez sur le stage source et configurez-le. Pour obtenir les instructions sur la configuration des stages source, reportez-vous au *Guide du concepteur du flux de données*.
4. Faites glisser un stage Candidate Finder sur le canevas et connectez-le au stage source.

Par exemple, si vous utilisiez le stage source Read from File, votre flux de données se présenterait comme suit :



Candidate Finder obtient les enregistrements candidats qui formeront l'ensemble des rapprochements potentiels pouvant être évalué ultérieurement par Transactional Match dans le flux de données.

5. Double-cliquez sur le stage Candidate Finder sur le canevas.
6. Dans le champ **Connexion**, sélectionnez la base de données à interroger pour trouver les enregistrements candidats. Si la base de données souhaitée n'est pas répertoriée, ouvrez Management Console et commencez par définir la connexion à la base de données.
7. Dans le champ SQL, entrez une instruction SQL `SELECT` qui recherche les enregistrements candidats, en fonction de la valeur contenue dans l'un des champs du flux de données. Pour référencer les champs de flux de données, utilisez le format `${FieldName}`, où `FieldName` correspond au nom du champ à référencer.

Par exemple, si vous souhaitez rechercher les enregistrements de la base de données dans lesquels la valeur de la colonne `LastName` est identique au champ `Customer_LastName` des enregistrements du flux de données, vous devez écrire une instruction SQL semblable à l'exemple suivant :

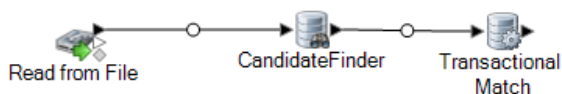
```
SELECT FirstName, LastName, Address, City, State, PostalCode
FROM Customer_Table
WHERE LastName = ${Customer_LastName};
```

8. Dans l'onglet **Mapping**, sélectionnez les champs du flux de données qui doivent contenir les données de chaque colonne de base de données.

La colonne **Champs sélectionnés** répertorie les colonnes de la base de données et **Champs de stage** répertorie les champs dans le flux de données.

9. Cliquez sur **OK**.
10. Faites glisser un stage Transactional Match sur le canevas et connectez-le au stage Candidate Finder.

Par exemple, si vous utilisez un stage d'entrée Read from File, votre flux de données se présenterait désormais comme suit :



Transactional Match met en correspondance les enregistrements suspects par rapport aux enregistrements candidats renvoyés par le stage Candidate Finder. Transactional Match utilise des règles de correspondance pour comparer l'enregistrement suspect à tous les enregistrements candidats possédant le même numéro de groupe de candidats (attribués dans Candidate Finder) pour identifier les doublons.

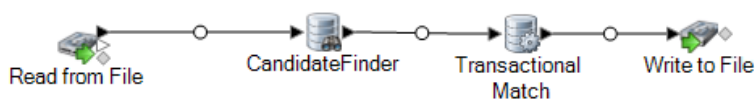
11. Double-cliquez sur le stage Transactional Match sur le canevas.
12. Dans le champ **Charger une règle de rapprochement**, sélectionnez une des règles de correspondance prédéfinies que vous pouvez utiliser tel quel ou modifier pour répondre à vos

besoins. Si vous souhaitez créer une nouvelle règle de correspondance sans utiliser une des règles de correspondance prédéfinies comme point de départ, cliquez sur **Nouveau**. Vous ne pouvez disposer que d'une règle personnalisée dans un flux de données.

**Remarque :** La fonction Options de flux de données dans Enterprise Designer permet d'afficher la règle de correspondance pour la configuration au moment de l'exécution.

13. Pour plus d'informations sur la modification des autres options, voir [Création d'une règle de correspondance](#) à la page 73.
14. Lorsque vous avez terminé de configurer le stage Transactional Match, cliquez sur **OK**.
15. Faites glisser un stage de collecteur de données sur le canevas et connectez-le au stage Transactional Match.

Par exemple, si vous utilisiez un stage de collecteur de données Write To File, votre flux de données se présenterait comme suit :



16. Double-cliquez sur le stage de collecteur de données et configurez-le.

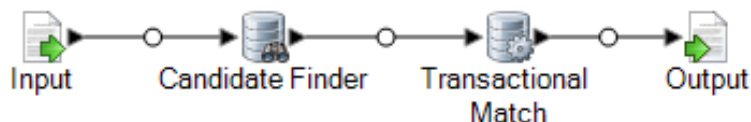
Pour obtenir des informations sur la configuration des stages de collecteur de données, reportez-vous au *Guide du concepteur de flux de données*.

Vous disposez désormais d'un flux de données qui met en correspondance les enregistrements de deux sources de données.

#### Exemple de mise en correspondance d'enregistrements par rapport à une base de données

En tant que cadre commercial d'une société de vente en ligne, vous désirez déterminer si un prospect en ligne est un client existant ou un nouveau client.

Le service de flux de données suivant apporte une solution au scénario commercial :



Ce flux de données est un service qui évalue les données de prospect qui lui sont envoyées par un appel API ou un appel de service Web. Il évalue les données par rapport aux données de client d'une base de données pour déterminer si un prospect est un client.

Le stage Input est configuré afin que le flux de données accepte les champs d'entrée suivants : AddressLine1, City, Name, PostalCode et StateProvince. Dans ce modèle, AddressLine1 et Name sont des champs clé pour le traitement du flux de données.

Le stage Candidate Finder obtient les enregistrements candidats qui formeront l'ensemble des correspondances potentielles qui seront ensuite évaluées par le stage Transactional Match.

Le stage Transactional Match vous permet de rapprocher des enregistrements suspects avec des enregistrements candidats potentiels renvoyés par le stage Candidate Finder. Transactional Match utilise des règles de correspondance pour comparer l'enregistrement suspect à tous les enregistrements candidats possédant le même numéro de groupe de candidats (attribués dans Candidate Finder) pour identifier les doublons. Dans cet exemple, Transactional Match compare LastName et AddressLine1.

Le stage Output renvoie les résultats du flux de données via une réponse API ou de service Web.

## Rapprochement d'enregistrements à l'aide de plusieurs règles de rapprochement

### ■ Télécharger l'échantillon de flux de données

Si vous disposez d'enregistrements que vous souhaitez rapprocher et que vous décidez d'utiliser plusieurs opérations de rapprochement, vous pouvez créer un flux de données utilisant plusieurs match keys, puis combiner les résultats pour effectuer un rapprochement efficace en fonction de plusieurs critères distincts. Par exemple, imaginons que vous souhaitiez créer un flux de données rapprochant des enregistrements où :

Le nom et l'adresse correspondent  
OR

La date de naissance et l'ID de gouvernement correspondent.

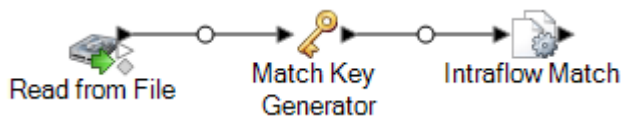
Pour procéder au rapprochement à l'aide de cette logique, vous créez un flux de données qui rapproche le nom et l'adresse dans un stage, et la date de naissance et l'ID de gouvernement dans un autre stage, puis vous combinez les enregistrements correspondants en une seule collection.

Cette rubrique explique la procédure générale pour configurer un flux de données dans lequel le rapprochement s'effectue sur deux stages de rapprochement. À des fins d'illustration, cette procédure utilise des stages Intraflow Match. Mais vous pouvez également utiliser cette technique avec des stages Interflow Match.

1. Dans Enterprise Designer, créez un flux de données.
2. Faites glisser un stage source sur le canevas.
3. Double-cliquez sur le stage source et configurez-le. Pour obtenir les instructions sur la configuration des stages source, reportez-vous au *Guide du concepteur du flux de données*.

4. Définissez le premier rapprochement. Les résultats de ce premier rapprochement est constitué de collections d'enregistrements rapprochés en fonction de votre premier jeu de critères de rapprochement, par exemple, des enregistrements rapprochés en fonction du nom et de l'adresse.

- a) Glissez un stage Match Key Generator et un stage Intraflow Match jusqu'au canevas et reliez-les de sorte à disposer d'un flux de données comme suit :



- a) Dans le stage Match Key Generator, définissez la match key à utiliser pour le premier rapprochement.

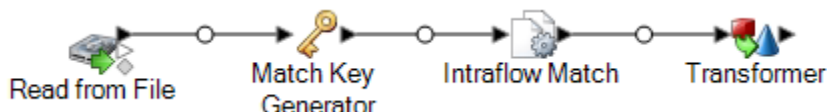
Par exemple, si vous souhaitez que le premier rapprochement s'effectue sur le nom et l'adresse, vous pouvez créer une match key basée sur les champs contenant le nom de famille et le code postal.

- b) Dans le stage Intraflow Match, définissez les règles de rapprochement à appliquer par le premier rapprochement.

Par exemple, vous pouvez configurer ce stage de rapprochement pour que le rapprochement s'effectue sur le nom et l'adresse.

5. Enregistrez les numéros de la collection du premier rapprochement dans un autre champ. Cette opération est nécessaire, car le champ CollectionNumber est écrasé lors du deuxième rapprochement. Il convient de renommer le champ CollectionNumber afin de préserver les résultats du premier rapprochement.

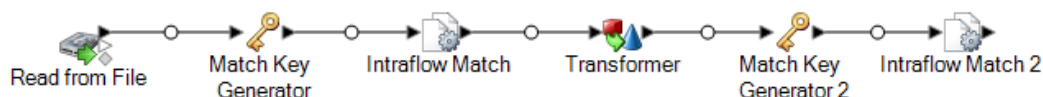
- a) Glissez un stage Transformer jusqu'au canevas et reliez-le au stage Intraflow Match de sorte à disposer d'un flux de données comme suit :



- b) Configurez le stage Transformer pour renommer le champ CollectionNumber en CollectionNumberPass1.

6. Définissez le deuxième rapprochement. Les résultats de ce deuxième rapprochement est constitué de collections d'enregistrements rapprochés en fonction de votre deuxième jeu de critères, par exemple, des enregistrements rapprochés en fonction de la date de naissance et de l'ID de gouvernement.

- a) Glissez un stage Match Key Generator et un stage Intraflow Match jusqu'au canevas et reliez-les de sorte à disposer d'un flux de données comme suit :



- b) Dans le deuxième stage Match Key Generator, définissez la match key à utiliser pour le deuxième rapprochement.

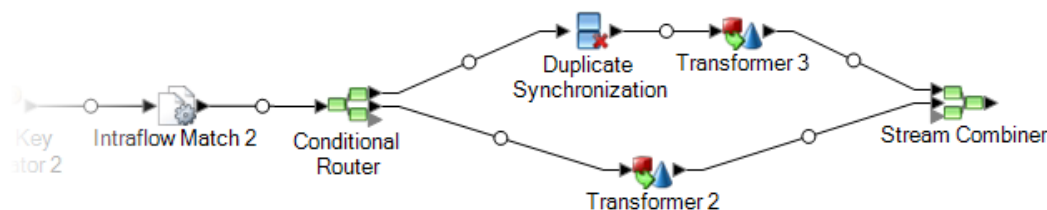
Par exemple, si vous souhaitez que le deuxième rapprochement s'effectue sur la date de naissance et l'ID de gouvernement, vous pouvez créer une match key basée sur les champs contenant la date de naissance et l'ID de gouvernement.

- c) Dans le deuxième stage Intraflow Match, définissez la règle de rapprochement à utiliser pour le deuxième rapprochement.

Par exemple, vous pouvez configurer ce stage de rapprochement pour que le rapprochement s'effectue sur la date de naissance et l'ID de gouvernement.

7. Déterminez si certains des enregistrements doublons identifiés par le deuxième rapprochement ont été également identifiés comme doublons lors du premier rapprochement.

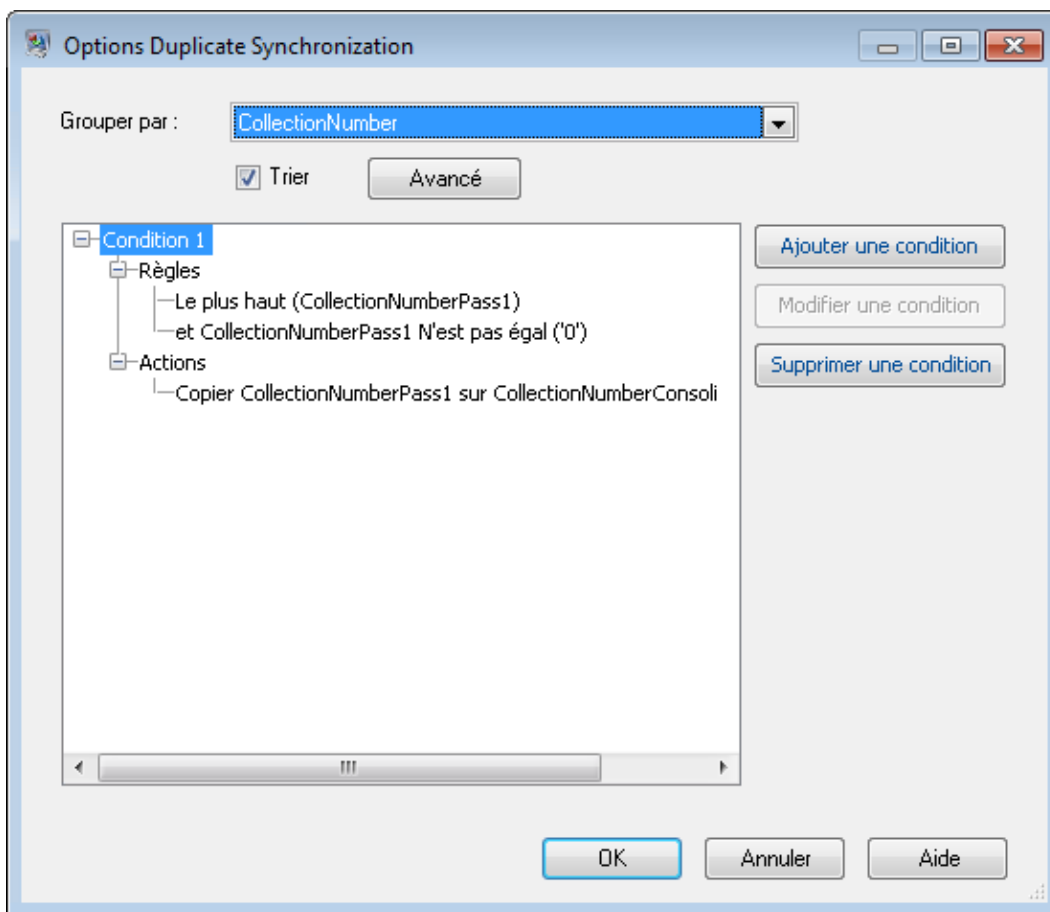
- a) Créez l'extrait de code de flux de données illustré ci-dessous après le deuxième stage Intraflow Match :



- b) Configurez le stage Conditional Router de sorte que les enregistrements dans lesquels le champ CollectionNumber n'est pas égal à 0 soient routés vers le stage Duplicate Synchronization.

Cette opération dirige les doublons du deuxième rapprochement vers le stage Duplicate Synchronization.

- c) Configurez le stage Duplicate Synchronization de sorte qu'il regroupe les enregistrements en fonction du champ CollectionNumber (il s'agit du numéro de collection) du deuxième rapprochement). Ensuite, au sein de chaque collection, déterminez si certains des enregistrements de la collection ont également été identifiés comme doublons lors du premier rapprochement. Si c'est le cas, copiez le numéro de collection du premier rapprochement dans un nouveau champ nommé CollectionNumberConsolidated. Pour ce faire, configurez le stage Duplicate Synchronization comme suit :



- d) Dans le stage Transformer qui suit le stage Duplicate Synchronization, créez une transformation personnalisée à l'aide du script suivant :

```
if (data['CollectionNumberConsolidated'] == null) {
  data['CollectionNumberConsolidated'] = data['CollectionNumber']
}
```

- e) Dans le Transformer qui suit immédiatement le Conditional Router, (Transformer 2 dans l'échantillon de flux de données), configurez une transformation copiant CollectionNumberPass1 dans CollectionNumberConsolidated.

Cette opération récupère les enregistrements uniques du deuxième rapprochement et copie CollectionNumberPass1 dans CollectionNumberConsolidated.

8. Après le Stream Combiner, vous disposez de collections d'enregistrements rapprochés lors de l'un ou l'autre des rapprochements. Le champ CollectionNumberConsolidated indique les enregistrements correspondants. Vous pouvez ajouter un récepteur ou tout autre traitement supplémentaire de votre choix après le stage Stream Combiner.

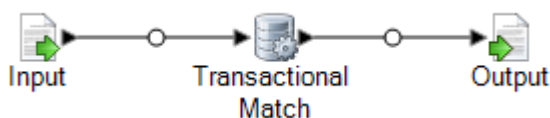


## Création d'un service de rapprochement universel

Un service de rapprochement universel est un service capable d'utiliser n'importe laquelle de vos règles de rapprochement pour effectuer un rapprochement et d'accepter tout champ d'entrée. Le service prend un nom de règle de rapprochement comme option d'entrée, vous permettant ainsi d'indiquer la règle de rapprochement que vous souhaitez utiliser dans l'appel de l'API ou la requête de service Web. Le service ne dispose pas de schéma d'entrée prédéfini, si bien que vous pouvez inclure tous les champs appropriés au type d'enregistrements que vous souhaitez rapprocher. En créant un service de rapprochement universel, vous pouvez éviter d'avoir des services distincts pour chaque règle de rapprochement, ce qui vous permet d'ajouter de nouvelles règles de rapprochement sans avoir à ajouter de service.

Cette procédure explique comment créer un service de rapprochement universel et inclut un exemple de requête de service Web envoyée au service de rapprochement universel.

1. Dans Enterprise Designer, créez un nouveau flux de données de service.
2. Glissez un stage Input, un stage Transactional Match et un stage Output jusqu'au canevas et reliez-les de sorte à obtenir un flux de données comme suit :



3. Double-cliquez sur le stage Transactional Match.
4. Dans le champ **Charger une règle de rapprochement**, sélectionnez une règle de rapprochement, quelle qu'elle soit. Par exemple, vous pouvez sélectionner la règle de rapprochement par défaut **Foyer**.

Même si vous indiquez la règle de rapprochement dans la requête de service, pour que le flux de données soit valide, vous devez configurer le stage Transactional Match avec une règle de rapprochement par défaut. Si vous ne sélectionnez pas de règle de rapprochement, le flux de données n'est pas validé et vous ne pouvez plus l'exposer.

5. Cliquez sur **OK**.
6. Double-cliquez sur le stage Output.
7. Exposez les champs `MatchRecordType` et `MatchScore`.
8. Cliquez sur **OK**.

**Remarque :** Il n'est pas nécessaire d'exposer de champs dans le stage Input, car les champs d'entrée sont indiqués comme des champs définis par l'utilisateur dans la requête de service.

9. Cliquez sur **Édition > Options de flux de données**.

10. Cliquez sur **Ajouter**.
11. Développez **Transactional Match** et cochez la case en regard de **Règle de rapprochement**.

Cette opération expose l'option de règle de rapprochement comme option d'exécution, permettant ainsi d'indiquer la règle de rapprochement dans la requête de service.

12. Cliquez **OK**, puis de nouveau sur **OK** pour fermer la fenêtre **Options de flux de données**.
13. Enregistrez et exposez le flux de données.

Vous disposez à présent d'un service de rapprochement universel que vous pouvez utiliser pour effectuer un rapprochement à l'aide de n'importe laquelle des règles de rapprochement définies dans l'outil Gestion des règles de correspondance d'Enterprise Designer. Si vous appelez le service, indiquez la règle de rapprochement dans l'option `MatchRule` et les champs d'entrée comme champs définis par l'utilisateur.

#### Exemple : Appel du service de rapprochement universel

Vous avez créé une règle de rapprochement nommée `AddressAndBirthday` dans l'outil Gestion des règles de correspondance. Cette règle de rapprochement rapproche les enregistrements en fonction des champs `Address` et `Birthday`. Vous souhaitez utiliser le service de rapprochement universel pour effectuer un rapprochement à l'aide de cette règle via une requête de service Web SOAP.

Pour ce faire, vous devez disposer d'une requête SOAP qui indique `AddressAndBirthday` dans l'élément `MatchRule` et les champs d'enregistrement dans l'élément `user_fields`.

```
<soapenv:Envelope
xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/"
xmlns:univ="http://www.pb.com/spectrum/services/UniversalMatchingService">

  <soapenv:Header/>
  <soapenv:Body>
    <univ:UniversalMatchingServiceRequest>
      <univ:options>

<univ:MatchRule>AddressAndBirthday</univ:MatchRule>
      </univ:options>
      <univ:Input>
        <univ:Row>
          <univ:user_fields>
            <univ:user_field>
              <univ:name>Name</univ:name>
              <univ:value>Bob Smith</univ:value>
            </univ:user_field>
            <univ:user_field>
              <univ:name>Address</univ:name>
              <univ:value>4200 Parliament
Pl</univ:value>
            </univ:user_field>
```

```

        <univ:user_field>
          <univ:name>Birthday</univ:name>
          <univ:value>1973-6-15</univ:value>
        </univ:user_field>
      </univ:user_fields>
    </univ:Row>
    <univ:Row>
      <univ:user_fields>
        <univ:user_field>
          <univ:name>Name</univ:name>
          <univ:value>Robert M. Smith</univ:value>

        </univ:user_field>
        <univ:user_field>
          <univ:name>Address</univ:name>
          <univ:value>4200 Parliament
Pl</univ:value>
        </univ:user_field>
        <univ:user_field>
          <univ:name>Birthday</univ:name>
          <univ:value>1973-6-15</univ:value>
        </univ:user_field>
      </univ:user_fields>
    </univ:Row>
    <univ:Row>
      <univ:user_fields>
        <univ:user_field>
          <univ:name>Name</univ:name>
          <univ:value>Bob Smith</univ:value>
        </univ:user_field>
        <univ:user_field>
          <univ:name>Address</univ:name>
          <univ:value>424 Washington
Blvd</univ:value>
        </univ:user_field>
        <univ:user_field>
          <univ:name>Birthday</univ:name>
          <univ:value>1959-2-19</univ:value>
        </univ:user_field>
      </univ:user_fields>
    </univ:Row>
  </univ:Input>
</univ:UniversalMatchingServiceRequest>
</soapenv:Body>
</soapenv:Envelope>

```

Cette requête renvoie la réponse suivante :

```

<soap:Envelope
xmlns:soap="http://schemas.xmlsoap.org/soap/envelope/">
  <soap:Body>

```

```

    <ns3:UniversalMatchingServiceResponse
xmlns:ns2="http://spectrum.pb.com/"

xmlns:ns3="http://www.pb.com/spectrum/services/UniversalMatchingService">

    <ns3:Output>
      <ns3:Row>
        <ns3:MatchScore/>

<ns3:MatchRecordType>Suspect</ns3:MatchRecordType>
      <ns3:user_fields>
        <ns3:user_field>
          <ns3:name>Name</ns3:name>
          <ns3:value>Bob Smith</ns3:value>
        </ns3:user_field>
        <ns3:user_field>
          <ns3:name>Birthday</ns3:name>
          <ns3:value>1973-6-15</ns3:value>
        </ns3:user_field>
        <ns3:user_field>
          <ns3:name>Address</ns3:name>
          <ns3:value>4200 Parliament Pl</ns3:value>

          </ns3:user_field>
        </ns3:user_fields>
      </ns3:Row>
      <ns3:Row>
        <ns3:MatchScore>100</ns3:MatchScore>

<ns3:MatchRecordType>Duplicate</ns3:MatchRecordType>
      <ns3:user_fields>
        <ns3:user_field>
          <ns3:name>Name</ns3:name>
          <ns3:value>Robert M. Smith</ns3:value>
        </ns3:user_field>
        <ns3:user_field>
          <ns3:name>Birthday</ns3:name>
          <ns3:value>1973-6-15</ns3:value>
        </ns3:user_field>
        <ns3:user_field>
          <ns3:name>Address</ns3:name>
          <ns3:value>4200 Parliament Pl</ns3:value>

          </ns3:user_field>
        </ns3:user_fields>
      </ns3:Row>
    </ns3:Output>
  </ns3:UniversalMatchingServiceResponse>
</soap:Body>
</soap:Envelope>

```

## Utilisation d'une Match Key Express

Le rapprochement par clé express peut s'avérer un outil utile pour réduire le nombre de comparaisons réalisé et ainsi améliorer la vitesse d'exécution dans les flux de données utilisant un stage Interflow Match ou Intraflow Match. Si deux enregistrements donnent une correspondance exacte sur la clé express, le candidat est considéré comme une correspondance à 100 % et aucune autre tentative de correspondance n'est effectuée. Si deux enregistrements ne correspondent pas sur une valeur de clé express, ceux-ci sont comparés via la méthode fondée sur les règles. Néanmoins, une clé express lâche donne lieu à de nombreuses correspondances de type faux positifs.

1. Ouvrez votre flux de données dans Enterprise Designer.
2. Double-cliquez sur le stage Match Key Generator.
3. Cochez la case **Créer une Match Key Express**.
4. Cliquez sur **Ajouter**.
5. Renseignez les champs suivants :

**Tableau 6 : Options de Match Key Generator**

Nom de l'option	Description/Valeurs valides
-----------------	-----------------------------

---

Algorithme	
------------	--

Nom de l'option	Description/Valeurs valides
	Définit l'algorithme à utiliser pour générer la clé de correspondance. L'un des éléments suivants :
<b>Consonne</b>	Renvoie les champs indiqués, les consonnes étant supprimées.
<b>Metaphone double</b>	Renvoie un code basé sur la représentation phonétique de leurs caractères. Le double Metaphone est une version améliorée de l'algorithme Metaphone et tente de prendre en compte les nombreuses irrégularités de plusieurs langues.
<b>Koeln</b>	Noms d'index par son, tels qu'ils sont prononcés en allemand. Permet aux noms ayant la même prononciation d'être encodés avec la même représentation afin qu'ils puissent être mis en correspondance, en dépit de différences mineures au niveau de l'orthographe. Le résultat est toujours une séquence de nombres ; les caractères spéciaux et les espaces blancs sont ignorés. Cette option a été développée en réponse aux limites du Soundex.
<b>MD5</b>	Algorithme qui produit une valeur hash de 128 bits. Cet algorithme est généralement utilisé pour vérifier l'intégrité des données.
<b>Metaphone</b>	Renvoie une clé codée Metaphone des champs sélectionnés. Metaphone est un algorithme qui code les mots à l'aide de leur sonorité lorsque prononcé en anglais.
<b>Metaphone (Espagnol)</b>	Renvoie une clé codée Metaphone des champs sélectionnés pour la langue espagnole. Cet algorithme Metaphone code les mots à l'aide de leur sonorité lorsque prononcé en espagnol.
<b>Metaphone3</b>	Procède à une amélioration en fonction des algorithmes Metaphone et Double Metaphone avec des paramètres de consonne et de voyelle interne exacts qui vous permet de produire des mots ou des noms mis en correspondance de manière plus ou moins proche pour rechercher des termes au niveau phonétique. Metaphone 3 augmente l'exactitude de l'encodage phonétique à 98 %. Cette option a été développée en réponse aux limites du Soundex.
<b>Nysiis</b>	L'algorithme de code phonétique qui met en correspondance une prononciation approximative avec une orthographe exacte et indexe des mots prononcés de manière similaire. Fait partie du système New York State Identification and Intelligence System. Imaginons, par exemple, que vous recherchez des informations sur une personne dans une base de données de personnes. Vous pensez que le nom de la personne sonne comme « John Smith », mais il est en fait orthographié « Jon Smyth ». Si vous procédez à une recherche de la correspondance exacte de « John Smith », aucun résultat n'est renvoyé. Cependant, si vous indexez la base de données à l'aide de l'algorithme NYSIIS et procédez à une recherche en utilisant de nouveau l'algorithme NYSIIS, la correspondance correcte est renvoyée car « John Smith » et « Jon Smyth » sont indexés comme « JAN SNATH » par l'algorithme.
<b>Phonix</b>	Pré-traite les chaînes de nom en appliquant plus de 100 règles

Nom de l'option	Description/Valeurs valides
	<p>de transformation à des caractères uniques ou à des séquences de plusieurs caractères. 19 de ces règles s'appliquent uniquement si les caractères figurent au début de la chaîne, tandis que 12 des règles s'appliquent uniquement si les caractères figurent au milieu de la chaîne et 28 des règles s'appliquent uniquement si les caractères figurent à la fin de la chaîne. La chaîne de nom transformée est cryptée en un code composé d'une lettre au début, suivie de trois chiffres (en enlevant les zéros et les nombres en double). Cette option a été développée pour répondre aux limites de Soundex ; elle est plus complexe et donc plus lente que Soundex.</p> <p><b>Soundex</b> Renvoie un code Soundex des champs sélectionnés. Soundex produit un code de longueur fixe en s'appuyant sur la sonorité du mot lorsque prononcé en anglais.</p> <p><b>Substring</b> Renvoie une partie spécifiée du champ sélectionné.</p>
Nom de champ	Indique le champ auquel vous souhaitez appliquer l'algorithme sélectionné pour générer la match key. Par exemple, si vous sélectionnez un champ appelé LastName et que vous choisissez l'algorithme Soundex, ce dernier est appliqué aux données dans le champ LastName pour produire une match key.
Position de début	Spécifie la position de départ dans le champ spécifié. Tous les algorithmes ne permettent pas de spécifier une position de départ.
Longueur	Spécifie la longueur de caractère à inclure à partir de la position de départ. Tous les algorithmes ne permettent pas de spécifier une longueur.
Supprimer les caractères parasites :	Supprime tout caractère non numérique et non alphanumérique, comme les traits d'union, les espaces blancs, et autres caractères spéciaux du champ d'entrée.
Trier les entrées :	<p>Trie tous les caractères dans un champ d'entrée ou tous les termes dans un champ d'entrée par ordre alphabétique.</p> <p><b>Caractères</b> Trie les valeurs de caractères d'un champ d'entrée avant de créer un identifiant unique.</p> <p><b>Termes</b> Trie les valeurs de termes d'un champ d'entrée avant de créer un identifiant unique.</p>

6. Cliquez sur **OK**.



7. Si vous souhaitez indiquer un champ et/ou algorithme supplémentaire pour créer une match key express, cliquez sur **Ajouter**, sinon cliquez sur **OK**.
8. Double-cliquez sur le stage Interflow Match ou Intraflow Match sur le canevas.
9. Sélectionnez l'option **Rapprochement express sur** et choisissez le champ **ExpressMatchKey**.  
Ce champ contient la match key express produite par Match Key Generator.
10. Cliquez sur **OK**.
11. Enregistrez et exécutez votre flux de données.

Pour savoir si un candidat a été mis en correspondance en utilisant une clé expresse, vérifiez la valeur du champ **ExpressKeyIdentified**, où Y signifie qu'il y a correspondance et N signifie qu'il n'y en a pas. Notez que les enregistrements suspects ont toujours une valeur **ExpressKeyIdentified** de N.

## Analyse des résultats de correspondance

L'outil Analyse des rapprochements dans Enterprise Designer affiche les résultats d'un ou de plusieurs stages de correspondance du même type. L'outil fournit un récapitulatif des résultats de correspondance d'un flux de données et vous permet d'afficher les résultats de correspondance, enregistrement par enregistrement. Vous pouvez utiliser ces informations pour dépanner ou affiner vos règles de correspondance, afin de produire les résultats de votre choix.

L'outil Analyse des rapprochements fournit les fonctions suivantes :

- Résultats de la synthèse de rapprochements : affiche les décomptes des enregistrements de synthèse pour un seul résultat de rapprochement ou des comparaisons entre deux résultats de rapprochement.
- Graphiques d'augmentation/réduction : utilise des histogrammes pour afficher une augmentation ou une réduction des rapprochements.
- Règles de rapprochement : affiche les règles de rapprochement utilisées pour un seul résultat de rapprochement ou les modifications apportées aux règles de rapprochement en comparant deux résultats de rapprochement.
- Résultats de détails de rapprochement : affiche les détails de traitement d'enregistrements pour un seul résultat de rapprochement ou la comparaison entre deux résultats de rapprochement.

## Affichage d'une synthèse des résultats de correspondance

L'outil Analyse des rapprochements peut afficher une synthèse d'informations sur les processus de mise en correspondance dans un dataflow, telles que le nombre de doublons d'enregistrements, le

score de correspondance moyen, etc. Vous pouvez afficher les résultats d'un job unique ou vous pouvez comparer les résultats entre plusieurs jobs.

1. Dans Enterprise Designer, ouvrez le dataflow à analyser.
2. Pour chaque stage Interflow Match, Intraflow Match ou Transactional Match, dont vous souhaitez analyser les correspondances, double-cliquez sur le stage et cochez la case **Générer les données pour analyse**.

**Important :** L'activation de l'option **Générer les données pour analyse** réduit les performances. Désactivez cette option lorsque vous avez terminé d'utiliser l'outil Analyse des rapprochements.

3. Sélectionnez **Exécuter > Exécuter le flux actif**.

**Remarque :** Afin d'obtenir des résultats optimaux, utilisez des données qui produiront au maximum 100 000 enregistrements. Plus vous obtenez de résultats de correspondance, plus les performances de l'outil Analyse des rapprochements sont lentes.

4. Lorsque l'exécution du dataflow se termine, sélectionnez **Outils > Analyse des rapprochements**.

La boîte de dialogue **Parcourir les résultats du rapprochement** affiche une liste des dataflow disposant de résultats de correspondance, que vous pouvez afficher dans l'outil Analyse des rapprochements. Si le job que vous souhaitez analyser n'est pas répertorié, ouvrez le dataflow et assurez-vous que le stage de correspondance présente la case **Générer les données pour analyse** cochée.

**Conseil :** Si vous disposez d'un grand nombre de dataflow et que vous souhaitez les filtrer, sélectionnez une option de filtre dans la liste déroulante **Afficher uniquement les jobs où**.

5. Cliquez sur l'icône "+" située en regard du dataflow à afficher pour le développer.
6. Sous le dataflow se trouve une entrée pour chaque stage de correspondance dans le dataflow. Sélectionnez le stage dont vous souhaitez afficher les résultats et cliquez sur **Ajouter**.

L'outil Analyse des rapprochements apparaît dans la partie inférieure de la fenêtre Enterprise Designer.

7. Pour comparer les résultats de correspondance, côte à côte, avec les résultats d'un autre outil de mise en correspondance, procédez comme suit :
  - a) Cliquez sur **Ajouter**.
  - b) Sélectionnez l'outil de mise en correspondance dont vous souhaitez comparer les résultats.
  - c) Cliquez sur **Ajouter**.
  - d) Dans la liste de dataflow, sélectionnez l'outil de mise en correspondance que vous venez d'ajouter et cliquez sur **Comparer**.

L'onglet **Résumé** répertorie les statistiques de correspondance du job. Selon le type de stage de correspondance utilisé dans le dataflow, vous obtiendrez des informations différentes.

Pour Intraflow Match, vous verrez les informations de synthèse suivantes :

**Enregistrements d'entrée**      Nombre total des enregistrements traités par le stage de correspondance.

<b>Enregistrements uniques</b>	Enregistrement de référence ou candidat qui ne correspond à aucun autre enregistrement dans un groupe de doublons. S'il est le seul enregistrement dans un groupe de doublons, un enregistrement de référence est automatiquement unique.
<b>Groupes de rapprochements</b>	(Regrouper par) Les enregistrements regroupés soit par une clé de correspondance ou une fenêtre coulissante.
<b>Groupes de doublons</b>	Un groupe de doublons comprend un enregistrement de référence et ses enregistrements doublons regroupés par numéro de groupe. Un groupe de doublons est également appelé une Collection dans Spectrum. Les enregistrements uniques appartiennent toujours au groupe 0.
<b>Rapprochements express</b>	On parle de rapprochement express lorsque les contenus d'un champ donné d'un enregistrement de référence et d'un enregistrement candidat correspondent parfaitement. Il s'agit généralement d'une clé de rapprochement express fournie par le Match Key Generator. Lorsqu'un rapprochement express est trouvé, aucune autre opération de traitement n'est conduite et l'enregistrement de référence et l'enregistrement candidat sont considérés comme doublons.
<b>Score moyen</b>	Score de correspondance moyen de tous les doublons. Les valeurs possibles sont 0-100, où 0 indique une faible correspondance et 100 indique une correspondance exacte.

Pour Interflow Match, vous verrez les informations de synthèse suivantes :

<b>Groupes de doublons</b>	Un groupe de doublons comprend un enregistrement de référence et ses enregistrements doublons regroupés par numéro de groupe. Un groupe de doublons est également appelé une Collection dans Spectrum. Les enregistrements uniques appartiennent toujours au groupe 0.
<b>Rapprochements express</b>	On parle de rapprochement express lorsque les contenus d'un champ donné d'un enregistrement de référence et d'un enregistrement candidat correspondent parfaitement. Il s'agit généralement d'une clé de rapprochement express fournie par le Match Key Generator. Lorsqu'un rapprochement express est trouvé, aucune autre opération de traitement n'est conduite et l'enregistrement de référence et l'enregistrement candidat sont considérés comme doublons.
<b>Score moyen</b>	Score de correspondance moyen de tous les doublons. Les valeurs possibles sont 0-100, où 0 indique une faible correspondance et 100 indique une correspondance exacte.
<b>Suspects d'entrée</b>	Nombre d'enregistrements dans le flux d'entrée que l'outil de mise en correspondance a tenté de rapprocher d'autres enregistrements
<b>Suspects contenant des doublons</b>	Nombre de suspects d'entrée mis en correspondance avec au moins un enregistrement candidat.
<b>Suspects uniques</b>	Nombre de suspects d'entrée qui n'ont été mis en correspondance avec aucun enregistrement candidat.
<b>Suspects contenant des candidats</b>	Nombre de suspects d'entrée qui disposaient d'au moins un enregistrement candidat dans leur groupe de correspondance et qui donc présentaient au moins une tentative de correspondance.

**Suspects sans candidats** Nombre de suspects d'entrée qui ne disposaient d'aucun enregistrement candidat dans leur groupe de correspondance et qui donc ne présentaient aucune tentative de correspondance.

Pour Transactional Match, vous verrez les informations de synthèse suivantes :

**Score moyen** Score de correspondance moyen de tous les doublons. Les valeurs possibles sont 0-100, où 0 indique une faible correspondance et 100 indique une correspondance exacte.

**Suspects d'entrée** Nombre d'enregistrements dans le flux d'entrée que l'outil de mise en correspondance a tenté de rapprocher d'autres enregistrements

**Suspects contenant des doublons** Nombre de suspects d'entrée mis en correspondance avec au moins un enregistrement candidat.

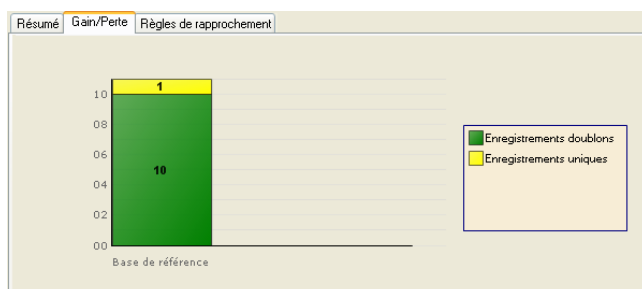
**Suspects uniques** Nombre de suspects d'entrée qui n'ont été mis en correspondance avec aucun enregistrement candidat.

**Suspects contenant des candidats** Nombre de suspects d'entrée qui disposaient d'au moins un enregistrement candidat dans leur groupe de correspondance et qui donc présentaient au moins une tentative de correspondance.

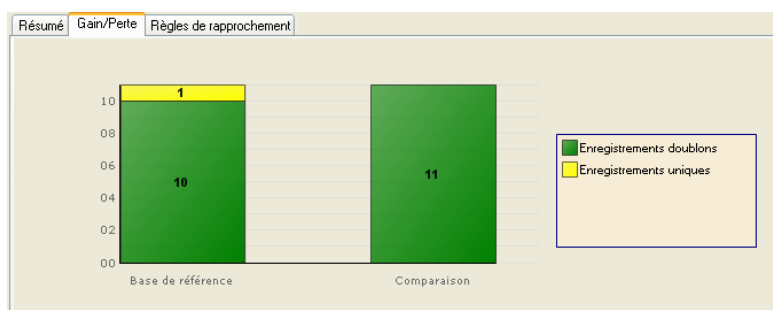
**Suspects sans candidats** Nombre de suspects d'entrée qui ne disposaient d'aucun enregistrement candidat dans leur groupe de correspondance et qui donc ne présentaient aucune tentative de correspondance.

L'onglet **Gain/Perte** de l'outil Analyse des rapprochements affiche le nombre de doublons et d'enregistrements uniques dans un histogramme pour la base de référence sélectionnée, ainsi que les résultats de comparaison éventuellement. L'*Augmentation* correspond à la augmentation du nombre d'enregistrements doublons. La *perte* correspond à la réduction du nombre d'enregistrements doublons. Les enregistrements uniques apparaissent en jaune et les enregistrements doublons en vert.

Si un seul job de base de référence est sélectionné, le graphique affichera les résultats de ce job unique :



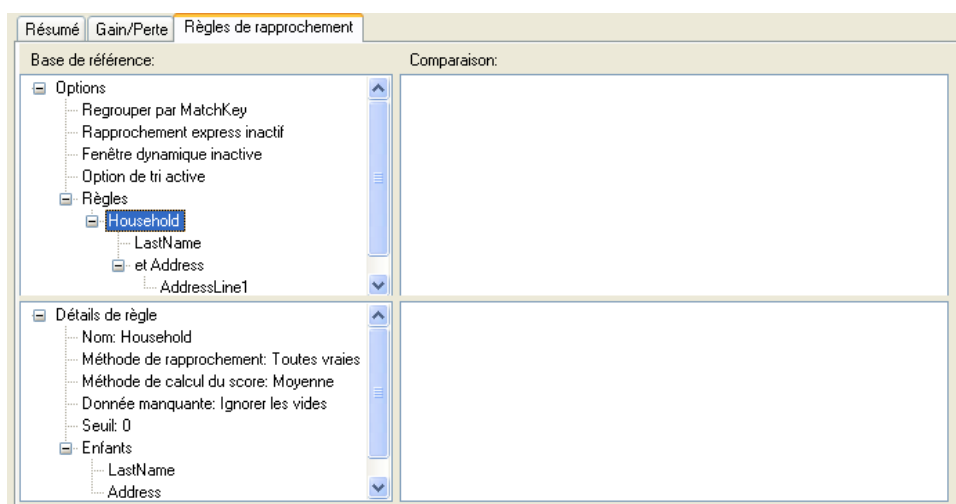
Si un job de base de référence et un job de comparaison sont sélectionnés, le graphique affiche les jobs de base de référence et de comparaison côte à côte :



L'onglet **Règles de correspondance** de l'outil Analyse des rapprochements affiche les règles de correspondance utilisées pour un seul résultat de correspondance ou pour les modifications apportées aux règles de correspondance lors de la comparaison de deux résultats de correspondance.

Les règles de rapprochement sont affichées sous la forme d'une structure hiérarchique similaire à la façon dont elles apparaissent dans le stage dans lequel elles sont créées. La hiérarchie de règle contient deux noeuds : Options et Règles. Le noeud Options montre les paramètres de stage pour le résultat de rapprochement sélectionné. Le noeud Règles montre les règles de rapprochement pour le résultat de rapprochement sélectionné.

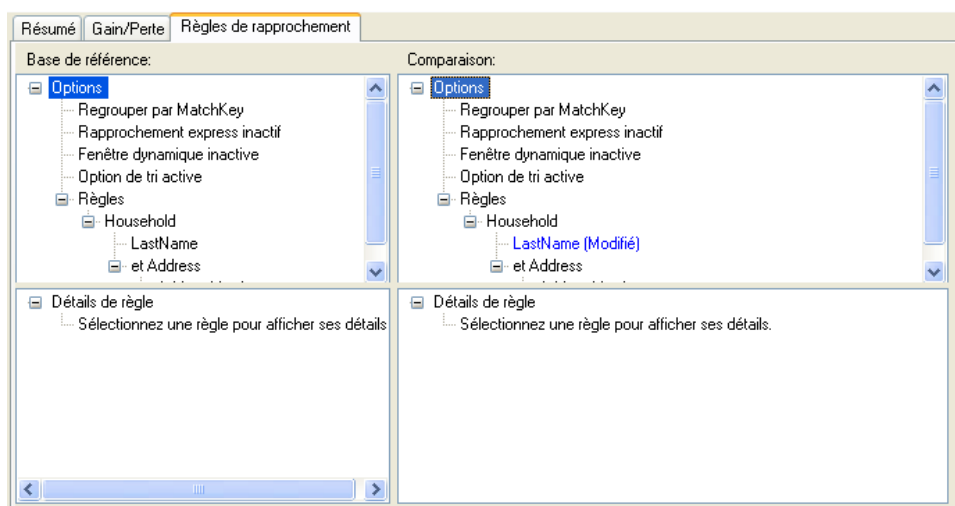
Pour visualiser une règle en détails, sélectionnez un noeud dans la hiérarchie.



Si vous comparez des règles de correspondance entre plusieurs jobs, les différences entre les résultats de correspondance de base de référence et de comparaison sont codées par couleur, comme suit :

- Bleu** Indique que la règle de rapprochement du résultat de rapprochement comparatif a été modifiée.
- Vert** Indique que la règle de rapprochement du résultat de rapprochement comparatif a été ajoutée.
- Rouge** Indique que la règle de rapprochement du résultat de rapprochement comparatif a été omise.

Par exemple :



## Affichage des résultats de correspondance de niveau d'enregistrement

Les détails d'un résultat sont un groupe de détails à propos des enregistrements de rapprochement pour un ensemble de résultats de rapprochements.

Pour afficher les résultats détaillés :

1. Dans l'outil Analyse des rapprochements, indiquez un job de ligne de référence et, éventuellement, un job de comparaison.
2. Cliquez sur **Détails**.

Les résultats de la correspondance de ligne de référence sont affichés en fonction de la vue sélectionnée dans la liste déroulante Afficher. La tableau suivant répertorie les colonnes affichées pour chaque type de stage de correspondance.

**Tableau 7 : Données de résultats détaillés affichées**

Résultats liés aux détails	Intraflow	Interflow	Transactional
Numéro d'enregistrement d'entrée	X	X	X
Groupe de correspondance	X	X	
Clé de rapprochement express	X	X	

Résultats liés aux détails	Intraflow	Interflow	Transactional
Enregistrement pilote de clé de rapprochement express	X	X	
Collection Number	X	X	X
Match Record Type	X	X	X
Champs utilisés par les règles	X	X	X
Score de règle global (niveau supérieur)	X		
Groupe de candidats		X	X
Score de rapprochement. Sélectionnez les résultats d'une correspondance dans la Liste des résultats de correspondance, puis cliquez sur <b>Supprimer</b> .		X	X

Pour plus d'informations sur le graphique de taux de réponse, voir [Graphique de taux de réponse](#) à la page 129.

3. Dans le champ **Analyser**, choisissez l'un des éléments suivants :

**Ligne de base** Affiche les résultats de rapprochement à partir de l'exécution de la base de référence.

**Comparaison** Affiche les résultats de rapprochement de l'exécution de comparaison.

4. Sélectionnez une des valeurs suivantes de la liste **afficher**, puis cliquez sur **Actualiser**. Si vous analysez les résultats de la base de référence, les options sont les suivantes :

- Suspects contenant des candidats : (Tous analyseurs) Affiche les enregistrements suspects et tous les enregistrements candidats qui ont tenté de se rapprocher de chaque suspect.
- Suspects contenant des doublons : (Tous analyseurs) Affiche tous les enregistrements suspects et les enregistrements candidats qui se sont rapprochés de chaque suspect.
- Suspects contenant des rapprochements express : (Interflow Match et Intraflow Match, quand Express Match Key est activé) Affiche les enregistrements suspects et candidats correspondant sur la base de la Match Key Express.

- Groupe de doublons : (Interflow Match et Intraflow Match) Affiche tous les groupes de doublons par numéro de groupe.
- Groupes de doublons : (Interflow Match et Intraflow Match) Affiche les enregistrements par groupes de doublons.
- Groupes de candidats : (Transactional Match) Affiche les enregistrements par groupes de candidats.
- Suspects uniques : (Interflow et Transactional Match) Affiche tous les enregistrements suspects qui n'ont correspondu à aucun enregistrement candidat.
- Enregistrements uniques : (Intraflow) Affiche les enregistrements non rapprochés.
- Suspects sans candidats : (Interflow et Transactional Match) Affiche tous les suspects qui n'ont contenu aucun candidat à rapprocher.
- Tous les enregistrements : Affiche tous les enregistrements traités par le stage de rapprochement.

Si vous analysez les résultats de comparaison, les options d'affichage sont les suivantes :

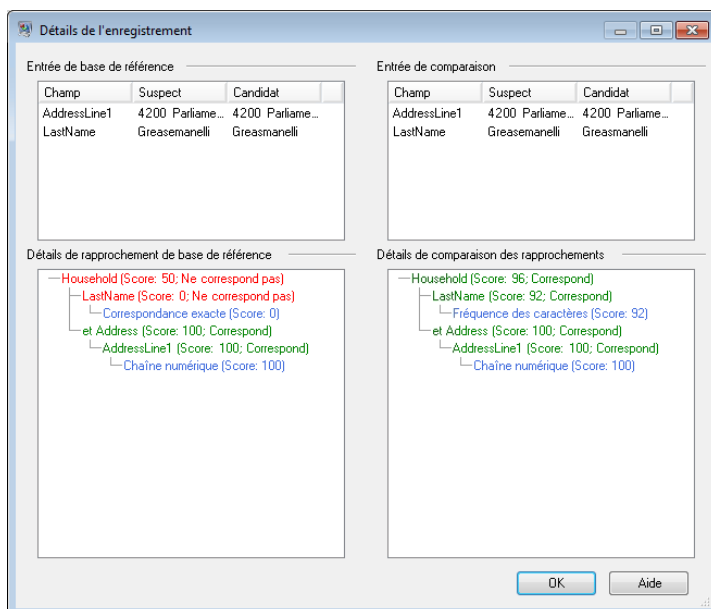
- Nouveaux rapprochements : (Intraflow) Affiche tous les nouveaux rapprochements et ses suspects associés. Cette vue combine les résultats des Suspects contenant de nouveaux doublons et des Nouveaux suspects en une seule vue.
  - Nouveaux suspects rapprochés : (Interflow et Transactional Match) Affiche les suspects qui n'avaient pas de doublons dans la base de référence mais au moins un doublon dans la comparaison.
  - Nouveaux suspects uniques : (Interflow et Transactional Match) Affiche les suspects qui avaient des doublons dans la base de référence mais n'en ont aucun dans la comparaison.
  - Rapprochements manquants : (Intraflow) Affiche tous les rapprochements manquants. Cette vue combine les résultats des Suspects contenant des doublons manquants et les Nouveaux suspects en une seule vue.
  - Suspects contenant de nouveaux doublons : (Tous analyseurs) Affiche des enregistrements qui sont de nouveaux doublons d'enregistrements qui étaient des suspects dans la base de référence et sont restés des suspects dans la comparaison.
  - Suspects contenant des doublons manquants : (Tous analyseurs) Affiche des enregistrements qui sont des doublons manquants d'enregistrements qui étaient des suspects dans la base de référence et sont restés des suspects dans la comparaison.
  - Nouveaux suspects : (Intraflow) Affiche les enregistrements qui sont des suspects dans le résultat de rapprochement comparatif, mais qui n'étaient pas suspects dans la base de référence.
  - Suspects manquants (Intraflow) Affiche les enregistrements qui ne sont pas des suspects dans les résultats de la comparaison, mais étaient des suspects dans la base de référence.
5. Développez un enregistrement suspect pour voir ses candidats.
  6. Sélectionnez un enregistrement candidat et cliquez sur **Détails**.

**Remarque :** Cette option n'est pas disponible lorsque la Fenêtre dynamique est activée dans les stages Intraflow Match.



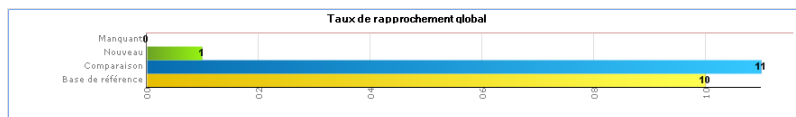
La fenêtre **Détails de l'enregistrement** affiche les données de niveau de champ, ainsi que le score de rapprochement de l'enregistrement pour chaque règle de correspondance. Si vous indiquez une base de référence et une exécution de job de comparaison, vous pouvez voir les résultats de l'enregistrement pour la base de référence et les exécutions de comparaison.

- Entrée de base de référence : affiche les données au niveau du champ, à partir à la fois du suspect et du candidat utilisés dans le rapprochement.
- Détails de rapprochement de base de référence : affiche les informations liées au score de chaque nœud dans les règles de rapprochement.
- Entrée de comparaison : affiche les données au niveau du champ, à partir à la fois du suspect et du candidat utilisés dans le rapprochement.
- Détails de comparaison des rapprochements : affiche les informations liées au score de chaque nœud dans les règles de rapprochement. Le texte en vert représente une correspondance pour un nœud dans les règles. Le texte en rouge représente une non-correspondance pour un nœud dans les règles.



## Graphique de taux de réponse

Les graphiques de taux de réponse affichent sous forme graphique les informations concernant un rapprochement en vues détaillées.




Pour les rapprochements de type dédoublement (Intraflow), un graphique apparaît, montrant les rapprochements globaux :

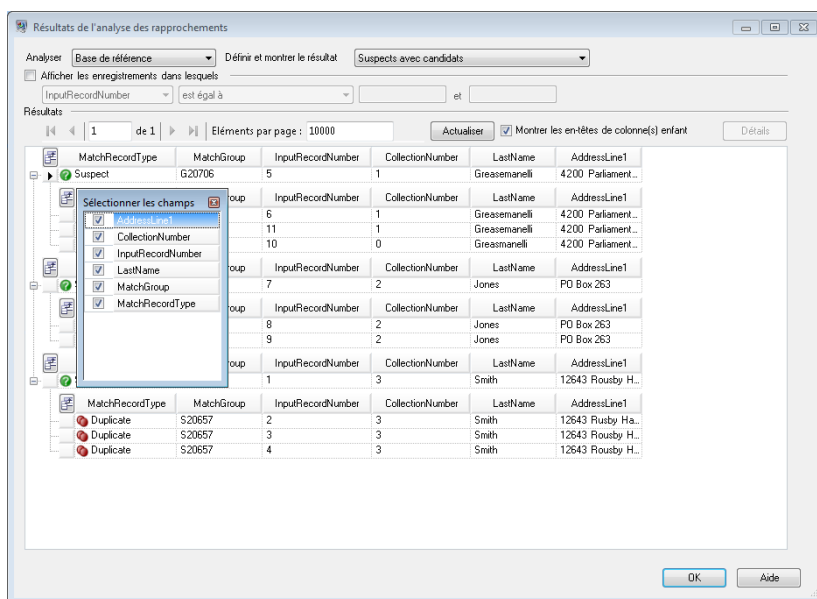
- Rapprochement de base de référence : Nombre total des rapprochements dans les résultats de base de référence.
- Comparaison des rapprochements : Nombre total des rapprochements dans les résultats de comparaison.
- Nouveaux rapprochements : Décompte de tous les enregistrements qui étaient uniques dans le résultat de base de référence, mais sont un suspect ou un doublon dans le résultat de la comparaison.
- Rapprochements manquants : Décompte de tous les enregistrements qui étaient suspects ou doublons dans le résultat de base de référence, mais sont uniques dans le résultat de la comparaison.

Pour les déduPLICATIONS (Interflow) et les rapprochements du composant Transactional Match, deux graphiques s'affichent :

- Taux de réponse global :
- Rapprochement de base de référence : Nombre total des rapprochements dans les résultats de base de référence.
- Comparaison des rapprochements : Nombre total des rapprochements dans les résultats de comparaison.
- Nouveaux rapprochements : Décompte de tous les enregistrements qui étaient uniques dans le résultat de base de référence, mais sont un suspect ou un doublon dans le résultat de la comparaison.
- Rapprochements manquants : Décompte de tous les enregistrements qui étaient suspects ou doublons dans le résultat de base de référence, mais sont uniques dans le résultat de la comparaison.
- Taux de réponse suspect
- Rapprochements de base de référence : Décompte de tous les suspects qui n'étaient pas uniques dans la base de référence.
- Comparaison des rapprochements : Décompte de tous les suspects qui n'étaient pas uniques dans la comparaison.
- Nouveaux rapprochements : Décompte de tous les suspects qui étaient uniques dans la base de référence, mais sont des rapprochements dans le résultat de la comparaison.
- Rapprochements manquants : Décompte de tous les suspects qui étaient des rapprochements dans la base de référence, mais sont uniques dans le résultat de la comparaison.

### Utilisation du Sélectionneur de champ

Cliquez sur l'icône du Sélectionneur de champ  pour afficher les colonnes sélectionnées dans les résultats de l'analyse des rapprochements. Le Sélectionneur de champ s'affiche au niveau parent et au niveau enfant. Vous pouvez choisir indépendamment des colonnes d'affichage pour les parents et les enfants.



## Filtrage des enregistrements

Utilisez la case à cocher **Afficher les enregistrements dans lesquels** pour filtrer les enregistrements détaillés de rapprochement affichés. Vous pouvez filtrer les enregistrements sur plusieurs opérateurs pour comparer les valeurs fournies par l'utilisateur avec les données dans un champ de chaque enregistrement détaillé de rapprochement.

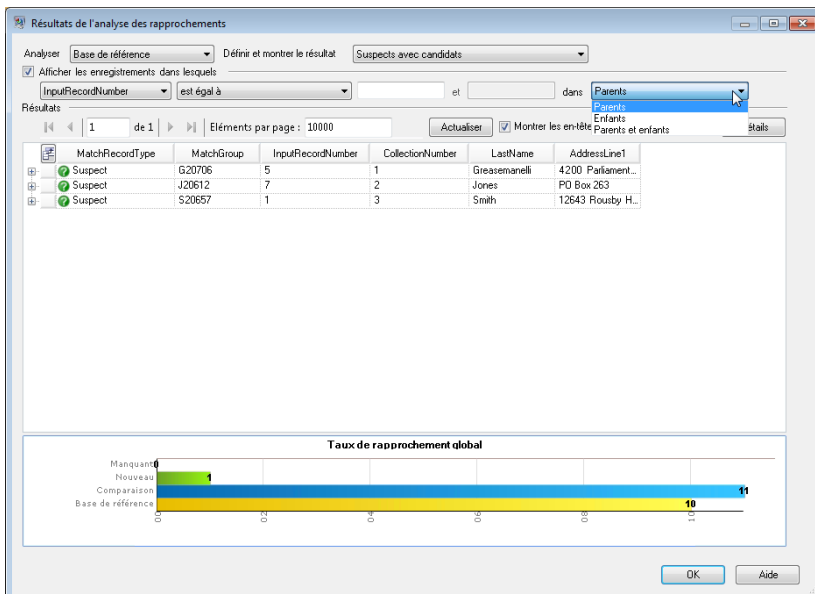
Les opérateurs que vous pouvez choisir sont :

- Les champs de type chaîne (GroupeCorrespondance, Typedenregistrementdecorrespondance, toute les données correspondantes)
- contient
- est compris entre
- est égal à
- n'est pas égal à
- commence par
- Les champs de type numérique (Numerodecollection, EntreeNumeroEnregistrement, MatchScore)
- est compris entre
- est égal à
- n'est pas égal à
- est supérieur à
- est supérieur ou égal à
- est inférieur à
- est inférieur ou égal à

Pour filtrer les enregistrements :

1. Sélectionnez la base de référence ou un résultat de rapprochement comparatif depuis la vue des Résultats de l'analyse des rapprochements et cliquez sur **Actualiser**.

## 2. Cochez la case **Afficher les enregistrements dans lesquels**.



- Sélectionnez un champ dans la Liste de champs.
- Sélectionnez un opérateur.
- Saisissez une valeur pour le type d'opérateur sélectionné. Si vous sélectionnez **est entre**, saisissez une plage de valeurs.
- Lorsque vous filtrez des vues suspectes, vous pouvez filtrer sur :
  - Parents : filtre sur les parents uniquement (Suspects), tous les enfants renvoyés.
  - Enfants : filtre sur tous les enfants qui n'entrent pas dans la plage de filtre. Nœuds parent (Suspect) renvoyés.
  - Parents et enfants : filtre sur les parents (Suspects), puis si les parents sont renvoyés, filtre sur leurs enfants
- Cliquez sur **Actualiser**. Les enregistrements qui se situent dans la plage des options et des valeurs sont affichés. Si aucun enregistrement ne se situe dans la plage des options choisies et des valeurs, un message affiche qu'aucun enregistrement n'a été renvoyé.

## Analyse des modifications apportées aux règles de correspondance

Vous pouvez utiliser l'outil Analyse des rapprochements dans Enterprise Designer pour afficher en détail l'effet qu'a une modification de règle de correspondance sur les résultats de correspondance du dataflow. Pour ce faire, vous pouvez exécuter le dataflow, apporter des modifications, réexécuter le dataflow, puis afficher les résultats dans l'outil Match Analysis. Cette procédure explique comment procéder.

**Important :** Lors de la comparaison des résultats de correspondance, les données d'entrée utilisées pour la base de référence et les exécutions de comparaison doivent être identiques. L'utilisation

des données d'entrée différentes peuvent entraîner des résultats trompeurs. Respectez les étapes suivantes pour garantir une comparaison exacte :

- Utiliser les mêmes fichiers ou tables d'entrée
  - Trier les données de la même façon qu'avant le stage de correspondance
  - Utiliser les mêmes requêtes Candidate Finder lors de l'utilisation de Transactional Match
1. Dans Enterprise Designer, ouvrez le dataflow à analyser.
  2. Pour chaque stage Interflow Match, Intraflow Match ou Transactional Match, dont vous souhaitez analyser les correspondances, double-cliquez sur le stage et cochez la case **Générer les données pour analyse**.

**Important :** L'activation de l'option **Générer les données pour analyse** réduit les performances. Désactivez cette option lorsque vous avez terminé d'utiliser l'outil Analyse des rapprochements.

3. Sélectionnez **Exécuter > Exécuter le flux actif**.

**Remarque :** Afin d'obtenir des résultats optimaux, utilisez des données qui produiront au maximum 100 000 enregistrements. Plus vous obtenez de résultats de correspondance, plus les performances de l'outil Analyse des rapprochements sont lentes.

4. Dans le ou les stages de correspondance du dataflow, apportez les modifications de votre choix à la règle de correspondance, puis exécutez à nouveau le dataflow.

Par exemple, si vous voulez tester l'effet de l'augmentation de la valeur seuil, modifiez la valeur seuil et réexécutez le dataflow.

5. Lorsque l'exécution du dataflow se termine, sélectionnez **Outils > Analyse des rapprochements**.

La boîte de dialogue **Parcourir les résultats du rapprochement** affiche une liste des dataflow disposant de résultats de correspondance, que vous pouvez afficher dans l'outil Analyse des rapprochements. Si le job que vous souhaitez analyser n'est pas répertorié, ouvrez le dataflow et assurez-vous que le stage de correspondance présente la case **Générer les données pour analyse** cochée.

**Conseil :** Si vous disposez d'un grand nombre de dataflow et que vous souhaitez les filtrer, sélectionnez une option de filtre dans la liste déroulante **Afficher uniquement les jobs où**.

6. Dans la partie gauche du volet Analyse des rapprochements, une liste affiche les stages de correspondance, un par exécution. Sélectionnez le stage de correspondance dans l'exécution à utiliser comme base de référence pour la comparaison, puis cliquez sur **Base de référence**. Ensuite, sélectionnez l'exécution avec laquelle vous souhaitez comparer la base de référence, puis cliquez sur **Comparer**.

Vous pouvez désormais comparer les résultats de correspondance de synthèse, tels que le nombre total de doublons d'enregistrements, ainsi que les informations de niveau d'enregistrement qui indiquent la manière dont chaque enregistrement a été évalué par rapport aux règles de correspondance.

### Exemple de comparaison de résultats de correspondance

Par exemple, imaginons que vous exécutiez un job nommé HouseholdRelationshipsAnalysis. Vous souhaitez tester l'effet d'une modification apportée au stage Household Match 2. Tout d'abord, vous exécutez le job à l'aide des paramètres d'origine, puis vous modifiez les règles de correspondance dans le stage Household Match 2 et réexécutez le job. Dans l'outil Analyse des rapprochements, l'exécution portant l'ID de job 10 correspond à l'exécution avec les paramètres d'origine ; vous la définissez donc comme base de référence. L'exécution portant l'ID de job 13 est l'exécution avec la règle de correspondance modifiée. Lorsque vous cliquez sur **Comparer**, vous pouvez voir que la règle de correspondance modifiée (ID de job 13) a produit un doublon de plus et un enregistrement unique de moins que la règle de correspondance d'origine.

	Baseline	Comparison	Changes
Input Records	11	11	0
Duplicate Records	10	11	1
Unique Records	1	0	-1
Match Groups	3	3	0
Duplicate Collections	3	3	0
Express Matches	0	0	0
Average Score	99.8	99.8	0.0

## Ajout de résultats de correspondance

Si vous exécutez un job alors que l'Outil d'analyse de correspondance est ouvert et que la Liste des résultats de correspondance est vide, les résultats de correspondance seront automatiquement ajoutés à la liste. Une fois qu'un résultats de correspondance a été ajouté, l'outil Analyse des rapprochements n'ajoute que les résultats de correspondance du même type de correspondance (Interflow Match, Intraflow Match ou Transactional Match).

Si vous souhaitez analyser les résultats de correspondance d'un autre type que celui actuellement sélectionné dans l'outil Analyse des rapprochements, procédez comme suit.

1. Choisissez tous les résultats de correspondance depuis la Liste des résultats de correspondance et cliquez sur **Supprimer**.
2. Ouvrez une tâche depuis l'Explorateur de serveur qui utilise une étape de correspondance différente ou cliquez sur l'onglet au dessus de la toile si la tâche est déjà ouverte.
3. Exécutez la tâche.

Lorsque la tâche a terminé son exécution, les résultats de correspondance de la dernière tâche seront ajoutés à la Liste des résultats de correspondance.

## Suppression des résultats de correspondance

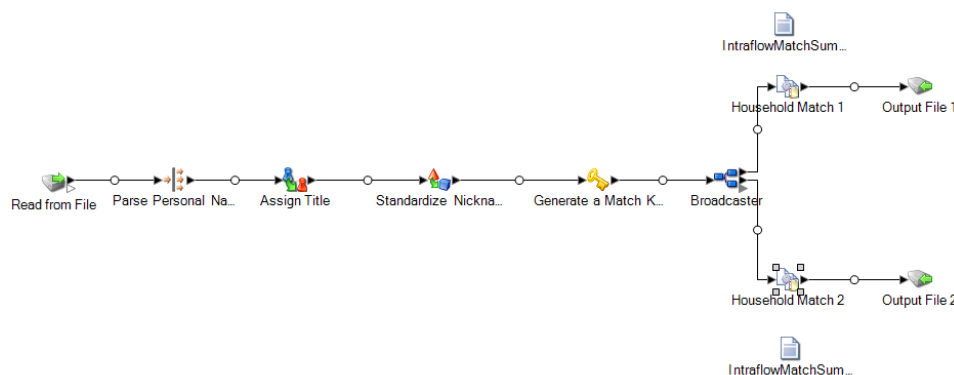
Pour supprimer les résultats d'une correspondance depuis la Liste des résultats de correspondance, choisissez les résultats d'une correspondance dans la Liste des résultats de correspondance et cliquez sur **Supprimer**.

Le système met à jour la Liste des résultats de correspondance et l'onglet Résumé comme suit :

- Si les résultats de correspondance supprimés n'étaient ni les résultats de correspondance de base ni de comparaison, les résultats de correspondance sont supprimés et aucun changement n'a lieu sur l'onglet Résumé.
- Si les résultats de correspondances supprimés étaient réglés comme étant la base, le système fixe les anciens résultats de correspondance les plus proches comme étant la nouvelle base et met à jour l'onglet Résumé pour afficher les nouvelles données de base uniquement.
- Si les résultats de correspondance supprimés étaient réglés comme étant les résultats de correspondance de comparaison, le système met l'onglet Résumé à jour pour afficher les données de base existante uniquement.
- Si les résultats de correspondance supprimés représentent l'un de deux résultats affichés dans la Liste de résultats de correspondance, les résultats de correspondance restants sont fixés comme étant la nouvelle base et le système met l'onglet Résumé à jour afin d'afficher les nouvelles données de base uniquement.

## Exemple : utilisation de Match Analysis

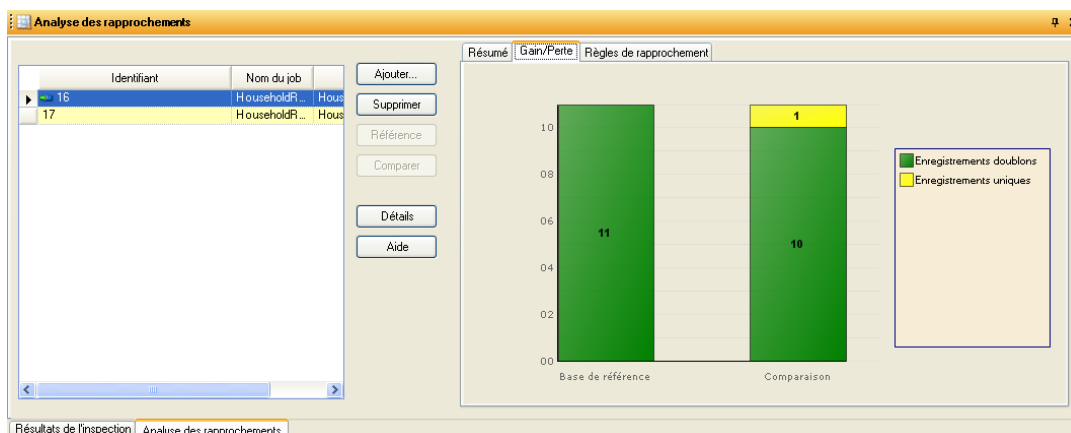
Cet exemple montre comment utiliser l'outil Match Analysis pour comparer les taux de Gain/Perte de deux correspondances différentes. Avant l'envoi des données à travers un comparateur, celles-ci sont divisées en deux flux à l'aide d'un diffuseur. Chaque flux est ensuite envoyé à travers un stage Intraflow Match. Chaque flux de données inclut des copies identiques des données traitées. Chaque stage Intraflow Match utilise un algorithme de correspondance différent et génère des données d'analyse de correspondance qui peuvent vous servir à comparer l'augmentation/diminution des diverses correspondances.



Cet exemple de flux de données est disponible dans Enterprise Designer. Accédez à **Fichier > Nouveau > Flux de données > À partir d'un modèle** et sélectionnez **HouseholdRelationshipsAnalysis**. Cet flux de données requiert les modules suivants : Advanced Matching, Data Normalization et Universal Name. Il requiert également que vous chargiez la base de données essentielle Table Lookup et les tables de base Open Parser.

Pour utiliser et afficher cet exemple :

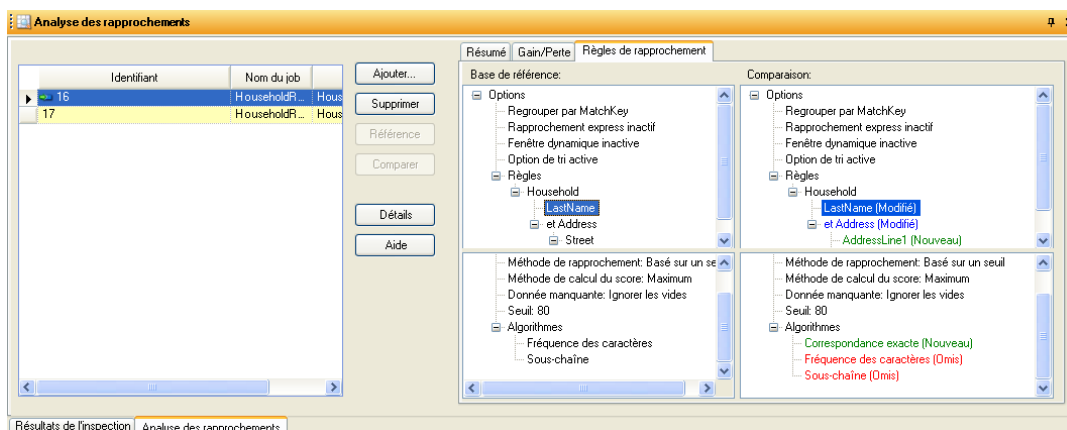
1. Exécutez le flux de données.
2. Sélectionnez **Outils > Analyse des rapprochements**.
3. Dans la fenêtre **Parcourir les résultats du rapprochement**, développez **HouseholdRelationshipAnalysis**, sélectionnez **Household Match 1** et **Household Match 2** dans la liste Source, puis cliquez sur **Ajouter**.
4. Sélectionnez **Household Match 1** dans la liste Résultats de correspondance et cliquez sur **Comparer**. La synthèse des résultats s'affiche.
5. Cliquez sur l'onglet **Gain/Perte**. Le graphique Gain/Perte apparaît.



Ce graphique indique les différences entre les enregistrements dupliqués et les enregistrements uniques générés par les différentes règles de correspondance utilisées.

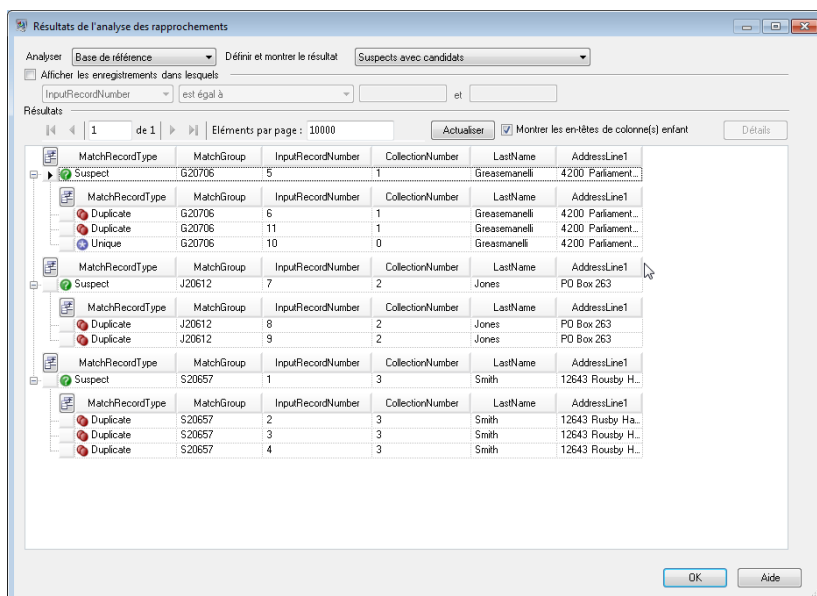
6. Cliquez sur l'onglet **Règles de rapprochement**. La comparaison des règles de correspondance s'affiche.





Depuis cet onglet, vous pouvez constater que l'algorithme a été modifié ; Fréquence de caractère est omis et Correspondance exacte a été ajouté.

7. Cliquez sur **Détails**.
8. Sélectionnez **Groupe de doublons** dans la liste affichée et cliquez sur **Actualiser**.
9. Développez chaque **CollectionNumber** pour visualiser les enregistrements Suspect et Doublet de chaque groupe de doublons.



10. Comparez les groupes dans la vue Détail par rapport au fichier de sortie créé.

## Modèles de dataflows pour la correspondance

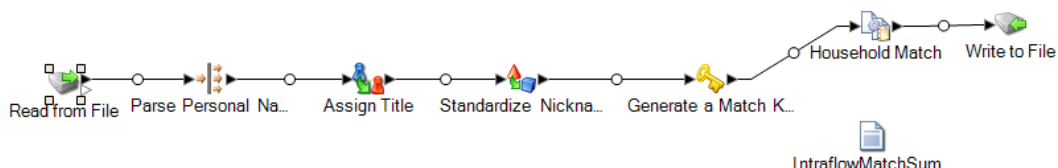
### Identification des membres d'un même foyer

Ce modèle de flux de données montre comment identifier les membres d'un même foyer en comparant les informations de chaque fichier d'entrée et en créant un fichier de sortie de groupes de foyers.

#### Scénario commercial

En tant que data steward pour une société de cartes de crédit, vous souhaitez analyser votre base de données client et déterminer les adresses doublons et leurs noms, afin de minimiser le nombre de courriers électroniques dupliqués et d'offres de carte de crédit envoyées à la même adresse.

Les flux de données suivants apportent une solution au scénario commercial :



Ce modèle de flux de données est disponible dans Enterprise Designer. Accédez à **Fichier > Nouveau > Flux de données > À partir d'un modèle** et sélectionnez **HouseholdRelationships**. Ce flux de données requiert les modules suivants : Advanced Matching, Data Normalization et Universal Name

Pour chaque enregistrement du fichier d'entrée, ce flux de données procède comme suit :

#### Read from File

Ce stage identifie le nom du fichier, l'emplacement et la disposition du fichier contenant les noms que vous désirez parser. Le fichier contient à la fois des noms féminins et masculins.

#### Open Name Parser

Le stage Open Name Parser examine les champs de nom et les compare aux données de nom stockées dans les fichiers de la base de données de noms Spectrum™ Technology Platform. En se basant sur la comparaison, il parse les données de noms en champs de prénom, second prénom et nom de famille, et assigne à chaque nom un type d'entité et un genre. Il utilise aussi une reconnaissance de modèle en plus des données de nom.

### Standardize Nicknames

Dans ce modèle, le stage Table Lookup se nomme Standardize Nicknames. Le stage Standardize Nickname recherche les prénoms dans la base de données Nicknames.xml et remplace tous les surnoms par la forme plus réglementaire du surnom. Par exemple, le nom Tommy est remplacé par Thomas.

### Transformer

Dans ce modèle, le stage Transformer se nomme Assign Titles (Assigner des titres). Le stage Assign Titles utilise un script personnalisé pour effectuer une recherche sur chaque ligne du flux de données émis par le stage Parse Personal Name et assigne une valeur **TitleOfRespect** en fonction de la valeur **GenderCode**.

Le script personnalisé est :

```
if (row.get('TitleOfRespect') == '')
{
  if (row.get('GenderCode') == 'M')
    row.set('TitleOfRespect', 'Mr')
  if (row.get('GenderCode') == 'F')
    row.set('TitleOfRespect', 'Ms')
```

Chaque fois que le stage Assign Titles trouve **M** dans le champ **GenderCode**, il définit la valeur de **TitleOfRespect** sur **Mr**. Chaque fois que le stage Assign Titles trouve **F** dans le champ **GenderCode**, il définit la valeur de **TitleOfRespect** sur **Ms**.

### Match Key Generator

Match Key Generator traite les règles définies par l'utilisateur composées d'algorithmes et de champs de source d'entrée pour générer le champ de clé de correspondance. Une clé de correspondance est un clé non-unique partagée par des enregistrements semblables identifiant les enregistrements comme des doublons. Les clés de correspondance servent à faciliter le processus de correspondance simplement en comparant des enregistrements contenant la même clé de correspondance. Une match key se compose de champs d'entrée. Un algorithme sélectionné s'applique sur chaque champ d'entrée indiqué. Le résultat de chaque champ est ensuite concaténé pour créer un seul champ de clé de correspondance.

Dans ce modèle, deux champs clés de correspondance sont définis : SubString (LastName (1:3)) et SubString (PostalCode (1:5)).

Par exemple, si l'adresse entrante était :

FirstName - Fred

LastName - Mertz

PostalCode - 21114-1687

Et les règles spécifiaient que :

Champ de saisie	Position de début	Longueur
LastName	1	3
PostalCode	1	5

Alors la clé, basée sur les règles et les données d'entrée indiquées plus haut, serait :

Mer21114

### *Household Match*

Dans ce modèle de flux de données, le stage Intraflow Match est appelé Household Match. Ce stage repère les correspondances entre des enregistrements de données similaires à l'intérieur d'un seul flux d'entrée. Des enregistrements correspondants peuvent aussi être qualifiés à l'aide d'informations ne se rapportant ni à un nom ni à une adresse. Le moteur de correspondance vous permet de créer des règles hiérarchiques basées sur n'importe quel champ défini ou créé dans d'autres stages.

Un flux d'enregistrements devant être comparés ainsi que les paramètres spécifiant quels champs doivent être comparés, comment les scores doivent être calculés, et d'une façon plus générale tout ce qui constitue une correspondance réussie.

Dans ce modèle, créez une règle de correspondance personnalisée qui compare LastName et AddressLine1. Cochez la case **Générer les données pour analyse** pour générer les données pour le rapport Interflow Summary.

Il y a quelques consignes à suivre lors de la création de votre hiérarchie de correspondance :

- Un nœud parent doit recevoir un nom unique. Cela ne peut pas être un champ.
- Le champ enfant doit être un champ de type de données Spectrum™ Technology Platform, il s'agit donc d'un champ disponible via un ou plusieurs composants.
- Tous les enfants sous un parent doivent utiliser les mêmes opérateurs logiques. Pour combiner des connecteurs, vous devez d'abord créer des nœuds parents intermédiaires.
- Les seuils au nœud parent peuvent être plus élevés que le seuil de l'enfant.
- Les nœuds parent n'ont pas besoin d'avoir un seuil.

### *Write to File*

Le modèle contient une étape Write to File qui crée un fichier texte indiquant les adresses sous forme d'un ensemble de seuils.

### *Rapport de synthèse intra-flux*

Le modèle contient le rapport de synthèse de Intraflow Match. Après avoir exécuté la tâche, développez **Rapports** dans la fenêtre Détails d'exécution, puis cliquez sur **IntraflowMatchSummary**.

Le Rapport de synthèse de Intraflow Match liste les statistiques des enregistrements traités et affiche un graphique à barres illustrant de façon graphique le décompte des enregistrements et le score général de correspondance.

## Déterminer si un prospecté est un client

Ce modèle de flux de données indique comment évaluer les données d'un prospect d'un fichier d'entrée par rapport aux données client d'une base de données client pour déterminer si un prospect est un client. Il s'agit d'un flux de données de type service, à savoir, accessible via l'API ou les services Web.

### Scénario commercial

En tant que cadre commercial d'une société de vente en ligne, vous désirez déterminer si un prospect en ligne est un client existant ou un nouveau client.

Le service de flux de données suivant apporte une solution au scénario commercial :



Ce modèle de flux de données est disponible dans Enterprise Designer. Accédez à **Fichier > Nouveau > Flux de données > À partir d'un modèle** et sélectionnez **ProspectMatching**. Ce flux de données requiert les modules Advanced Matching et Universal Name.

Pour chaque enregistrement du fichier d'entrée, le flux de données procède comme suit :

### Input

Les champs d'entrée sélectionnés dans ce modèle sont AddressLine1, City, Name, PostalCode, et StateProvince. Dans ce modèle, AddressLine1 et Name sont des champs clé pour le traitement du flux de données.

### Name Parser

Dans ce modèle, le stage Parser les noms de personne est nommée Parse Personal Name. Le stage Parse Personal Name examine les champs de noms et les compare aux données de noms stockées dans les fichiers de la base de données de noms Spectrum™ Technology Platform. En se basant sur la comparaison, il parse les données de noms en champs de prénom, second prénom et nom de famille, et assigne à chaque nom un type d'entité et un genre. Il utilise aussi une reconnaissance de modèle en plus des données de nom.

Dans ce modèle, le stage Parser les noms de personne est configurée comme suit.

- Le parsing des noms de personnes est sélectionné et le parsing des noms commerciaux est effacé. Lorsque vous sélectionnez ces options, les prénoms sont évalués afin de déterminer leur sexe, leur ordre et leur ponctuation et aucune évaluation de noms d'entreprises n'est réalisée.

- Gender Determination Source (Source de détermination du sexe) est définie par défaut. Dans la plupart des cas, le paramètre par défaut est le plus adapté à la détermination par sexe parce qu'il couvre une grande variété de noms. Toutefois, si vous traitez des noms pour une culture spécifique, sélectionnez cette culture. La sélection d'une culture spécifique aide à garantir l'assignation d'un genre approprié aux noms. Par exemple, si vous laissez la sélection par défaut, le nom de Jean sera identifié comme un prénom féminin. Par contre, si vous sélectionnez Français, il sera identifié comme un prénom masculin.
- L'ordre est paramétré sur naturel. Les champs de nom sont classés par titre, prénom, deuxième prénom, nom de famille et suffixe.
- Conserver les points est effacé. Aucune ponctuation figurant dans le nom n'est conservée.

### *Candidate Finder*

Le stage Candidate Finder s'utilise conjointement au stage Transactional Match.

Le stage Candidate Finder obtient les enregistrements candidats qui formeront l'ensemble des correspondances potentielles qui seront ensuite évaluées par le stage Transactional Match. De plus, en fonction du format de vos données, Candidate Finder peut aussi avoir besoin de parser le nom ou l'adresse de l'enregistrement suspect, des enregistrements candidats, ou des deux.

Comme cela fait partie de la configuration Candidate Finder, sélectionnez la connexion à la base de données par laquelle la requête spécifiée sera exécutée. Vous pouvez sélectionner n'importe quelle connexion configurée dans Management Console. Pour vous connecter à une base de données non répertoriée, configurez une connexion à cette base de données dans Management Console, puis fermez et rouvrez Candidate Finder pour actualiser la liste des connexions.

Pour définir la requête SQL, vous pouvez saisir n'importe quelle instruction dans la zone de texte de la vue Options de Candidate Finder. Par exemple, imaginez que vous avez une table dans votre base de données nommée Customer\_Table qui possède les colonnes suivantes :

Customer\_Table

---

Cust\_Name

---

Cust\_Address

---

Cust\_City

---

Cust\_State

---

Cust\_Zip

---

**Remarque :** Vous pouvez saisir n'importe quelle sélection SQL valide ; toutefois, `Select *` n'est pas valide dans cette commande.

Pour extraire toutes les rangées de la base de données, vous pourriez exécuter une requête ressemblant à ceci :

```
select Cust_Name, Cust_Address, Cust_City, Cust_State, Cust_Zip from
Customer_Table;
```

Toutefois, il est peu vraisemblable que vous souhaitiez comparer votre transaction à toutes les lignes de la base de données. Pour ne renvoyer que les enregistrements des candidats concernés, vous préférerez ajouter la clause `WHERE` en utilisant une substitution variable. Une substitution variable fait référence à une notation particulière dont vous vous servez pour que le moteur de sélection de candidats remplace la variable par les données actuelles de votre enregistrement suspect.

Pour utiliser la substitution de variable, encadrez le nom de champ entre des accolades précédées du signe dollar comme ceci : **`${FieldName}`**. Par exemple, la requête suivante ne renverra que les enregistrements qui ont une valeur dans `Cust_Zip` qui correspond à la valeur dans `PostalCode` sur l'enregistrement suspect.

```
select Cust_Name, Cust_Address, Cust_City, Cust_State, Cust_Zip
from Customer_Table
where Cust_Zip = ${PostalCode};
```

Vous devrez ensuite mapper les colonnes de votre base de données sur les champs d'étape si les noms de colonnes de votre base de données ne correspondent pas exactement aux noms du champ de composant. S'ils correspondent, ils seront automatiquement mappés sur les champs d'étape correspondants. Vous devrez utiliser les champs `Selected` (colonnes de la base de données) pour mapper les champs `Stage` (noms de champ définis dans le flux de données).

Considérez de nouveau `Customer_Table` à partir de l'exemple suivant :

Customer\_Table

---

Cust\_Name

---

Cust\_Address

---

Cust\_City

---

Cust\_State

---

Cust\_Zip

Quand vous extrayez ces enregistrements de la base de données, vous devez mapper les noms de colonne aux noms de champ employés par le stage Transactional Match et d'autres stages de votre flux de données. Par exemple, Cust\_Address pourrait être mappé sur AddressLine1, et Cust\_Zip serait mappé sur PostalCode.

1. Sélectionnez la liste déroulante sous **Champs sélectionnés** dans la vue Options Candidate Finder. Sélectionnez ensuite la colonne de base de données Cust\_Zip.
2. Sélectionnez la liste déroulante sous **Champs de stage**. Puis, sélectionnez le champ sur lequel vous voulez mapper.

Par exemple, si vous voulez mapper Cust\_Zip sur Postal Code, sélectionnez d'abord Cust\_Zip sous Champs sélectionnés, puis sélectionnez PostalCode sur la ligne Champ de stage correspondante.

En plus du mappage de champs décrit plus haut, vous devez utiliser une notation spéciale dans votre requête SQL pour réaliser le mappage. Pour ce faire, vous allez entrer le nom du champ de stage, entre accolades, après le nom de colonne de votre requête. Lorsque vous faites ceci, les champs sélectionnés seront automatiquement mappés sur les champs de stage correspondants.

En voici un exemple utilisant la requête des exemples précédents :

```
select Cust_Name {Name}, Cust_Address {AddressLine1},
       Cust_City {City}, Cust_State {StateProvince},
       Cust_Zip {PostalCode}
from Customer
where Cust_Zip = ${PostalCode};
```

### Transactional Match

Le stage Transactional Match s'utilise conjointement au stage Candidate Finder.

Le stage Transactional Match vous permet de comparer des enregistrements suspects à des enregistrements candidats potentiels renvoyés par le stage Candidate Finder.

Transactional Match utilise des règles de correspondance pour comparer l'enregistrement suspect à tous les enregistrements candidats possédant le même numéro de groupe de candidats (attribués dans Candidate Finder) pour identifier les doublons. Si l'enregistrement candidat est un doublon, un numéro de groupe lui est attribué, le type d'enregistrement de rapprochement est étiqueté Doublon, et l'enregistrement est ensuite écrasé. Tous candidats sans rapprochement dans le groupe reçoit un numéro de groupe de 0, est étiqueté Unique, puis écrasé lui-aussi.

Dans ce modèle, créez une règle de correspondance personnalisée qui compare LastName et AddressLine1.

Il y a quelques consignes à suivre lors de la création de votre hiérarchie de correspondance :

- Un nœud parent doit recevoir un nom unique. Cela ne peut pas être un champ.



- Le champ enfant doit être un champ de type de données Spectrum™ Technology Platform, à savoir, un champ disponible via un ou plusieurs stages.
- Tous les enfants sous un parent doivent utiliser les mêmes opérateurs logiques. Pour combiner des connecteurs, vous devez d'abord créer des nœuds parents intermédiaires.
- Les seuils au nœud parent peuvent être plus élevés que le seuil de l'enfant.
- Les nœuds parent n'ont pas besoin d'avoir un seuil.

### *Sortie*

En tant que service, ce modèle envoie tous les champs disponibles vers la sortie. Vous pouvez limiter la sortie en vous basant sur vos besoins.

# 5 - Déduplication

## In this section

---

Filtrage des enregistrements doublons	147
Création d'un enregistrement Best of Breed	152

## Filtrage des enregistrements doublons

La façon la plus simple de supprimer des enregistrements doublons est d'ajouter un stage Filter à votre flux de données après un stage de mise en correspondance. Le stage Filter supprime les enregistrements des collections d'enregistrements doublons en fonction des paramètres que vous indiquez.

1. Dans Enterprise Designer, créez un flux de données identifiant les enregistrements en double à travers une mise en correspondance.

La mise en correspondance est la première étape dans la déduplication, car vous devez identifier des enregistrements similaires, tels que des enregistrements disposant du même nom ou numéro de compte. Pour en savoir plus sur la création d'un flux de données qui met en correspondance les enregistrements, reportez-vous aux rubriques suivantes.

[Mise en correspondance d'enregistrements d'une source unique](#) à la page 87

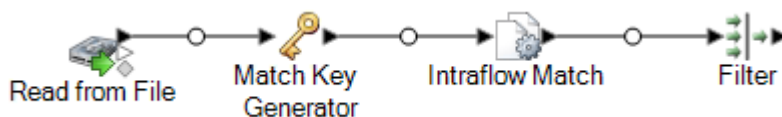
[Mise en correspondance d'enregistrements d'une source avec ceux d'une autre source](#) à la page 93

[Mise en correspondance d'enregistrements par rapport à une base de données](#) à la page 106

**Remarque :** Vous devez uniquement créer le flux de données au stade où il lit les données et effectue une mise en correspondance avec un stage Interflow Match, Intraflow Match ou Transactional Match. Une fois que vous avez créé un flux de données à ce stade, continuez avec les étapes suivantes.

2. Une fois que vous avez défini un flux de données qui lit les données et met les enregistrements en correspondance, faites glisser un stage Filter sur le canevas et connectez-le au stage qui effectue la mise en correspondance (Interflow Match, Intraflow Match ou Transactional Match).

Par exemple, si votre flux de données lit des données à partir d'un fichier et effectue une mise en correspondance avec Intraflow Match, votre flux de données ressemblera à ce qui suit, après avoir ajouté un stage Filter :



3. Double-cliquez sur le stage Filter sur le canevas.
4. Dans le champ **Grouper par**, sélectionnez **NuméroCollection**.
5. Laissez l'option **Limiter le nombre d'enregistrements doublons renvoyés** sélectionnée et la valeur définie sur 1. Il s'agit des paramètres par défaut.

6. Décidez si vous souhaitez conserver le premier enregistrement de chaque collection ou définir une règle permettant de déterminer l'enregistrement à conserver dans chaque collection. Si vous souhaitez conserver le premier enregistrement de chaque collection, ignorez cette étape. Si vous souhaitez définir une règle, dans l'arborescence des règles, sélectionnez **Règles** et procédez comme suit :

- a) Cliquez sur **Ajouter une règle**.

Les enregistrements de chaque groupe sont évalués pour déterminer s'ils respectent les règles que vous définissez ici. Si un enregistrement respecte la règle, il s'agit de l'enregistrement survivant et les autres enregistrements du groupe sont rejetés.

- b) Définissez une règle permettant d'identifier l'enregistrement de chaque groupe à retenir.

Utilisez les options suivantes pour définir une règle :

Option	Description
Nom du champ	Indique le nom du champ de flux de données dont vous souhaitez évaluer la valeur afin de déterminer si l'enregistrement doit être filtré.
Type de champ	Indique le type de données dans le champ. L'un des éléments suivants : <ul style="list-style-type: none"> <li><b>Non-numérique</b>      Choisissez cette option si le champ contient des données non numériques (par exemple, des données de chaîne).</li> <li><b>numériques</b>        Choisissez cette option si le champ contient des données numériques (par exemple, double, flottantes, etc.).</li> </ul>

Option	Description
--------	-------------

---

Opérateur	
-----------	--

Option	Description
	Indique le type de comparaison à utiliser pour évaluer le champ. L'un des éléments suivants :
<b>Contient</b>	Détermine si le champ contient la valeur indiquée. Par exemple, « bateau à voile » contient la valeur « bateau ».
<b>Est égal à</b>	Détermine si le champ contient la valeur exacte indiquée.
<b>Est supérieur à</b>	Détermine si la valeur du champ est supérieure à la valeur spécifiée. Cette opération ne fonctionne que sur les champs numériques.
<b>Supérieur ou Egal à</b>	Détermine si la valeur du champ est supérieure ou égale à la valeur spécifiée. Cette opération ne fonctionne que sur les champs numériques.
<b>Le plus haut</b>	Compare la valeur du champ pour tout le groupe d'enregistrements et identifie l'enregistrement qui possède la valeur la plus élevée dans le champ. Par exemple, si les champs dans ce groupe contiennent les valeurs 10, 20, 30, et 100, l'enregistrement qui dispose de la valeur 100 est sélectionné. Cette opération ne fonctionne que sur les champs numériques. Si plusieurs enregistrements sont à égalité pour la valeur la plus longue, un enregistrement est sélectionné.
<b>Est vide</b>	Détermine si le champ ne contient aucune valeur.
<b>N'est pas vide</b>	Détermine si le champ contient une valeur.
<b>Inférieur à</b>	Détermine si la valeur du champ est inférieure à la valeur spécifiée. Cette opération ne fonctionne que sur les champs numériques.
<b>Inférieur ou Egal à</b>	Détermine si la valeur du champ est inférieure ou égale à la valeur spécifiée. Cette opération ne fonctionne que sur les champs numériques.
<b>Le plus long</b>	Compare la valeur du champ pour tout le groupe d'enregistrements et identifie l'enregistrement qui possède la valeur la plus longue (en octets) dans le champ. Par exemple, si le groupe contient les valeurs « Mike » et « Michael », l'enregistrement possédant la valeur « Michael » est choisi. Si plusieurs enregistrements sont à égalité pour la valeur la plus longue, un enregistrement est sélectionné.
<b>Le plus bas</b>	Compare la valeur du champ pour tout le groupe d'enregistrements et identifie l'enregistrement qui possède la valeur la plus basse dans le champ. Par exemple, si les champs dans ce groupe contiennent les valeurs 10, 20, 30, et 100, l'enregistrement qui dispose de la valeur 10 est sélectionné. Cette opération ne fonctionne que sur les champs numériques. Si plusieurs enregistrements sont à égalité pour la valeur la plus longue, un enregistrement est sélectionné.
<b>Le plus commun</b>	Détermine si la valeur de champ contient la valeur qui se produit le plus fréquemment dans ce champ parmi les enregistrements dans ce groupe. Si deux valeurs, ou plus, sont les plus

Option	Description
	<p>courantes, aucun action ne sera réalisée.</p> <p><b>N'est pas égal</b> Détermine si la valeur du champ n'est pas égale à la valeur spécifiée.</p>
Type de valeur	<p>Indique le type de valeur à comparer avec la valeur du champ. L'un des éléments suivants :</p> <p><b>Remarque :</b> Cette option n'est pas disponible si vous sélectionnez l'opérateur Highest, Lowest ou Longest.</p> <p><b>Champ</b> Choisissez cette option si vous souhaitez comparer une autre valeur du champ de dataflow avec le champ.</p> <p><b>Chaîne</b> Choisissez cette option si vous souhaitez comparer le champ avec une valeur spécifique.</p>
Valeur	<p>Indique la valeur à comparer avec la valeur du champ. Si vous avez sélectionné Champ dans le champ <b>Type de champ</b>, sélectionnez un champ de dataflow. Si vous avez sélectionné Chaîne dans le champ <b>Type de valeur</b>, tapez la valeur à utiliser dans la comparaison.</p> <p><b>Remarque :</b> Cette option n'est pas disponible si vous sélectionnez l'opérateur Highest, Lowest ou Longest.</p>

c) Cliquez sur **OK**.

Vous avez désormais configuré le filtre avec une règle. Vous pouvez ajouter d'autres règles si nécessaire.

7. Cliquez sur **OK** pour fermer la fenêtre **Options de filtre**.

8. Faites glisser un stage de collecteur de données sur le canevas et connectez-le au stage Filter.

Par exemple, si vous utilisiez un stage de collecteur de données Write To File, votre flux de données se présenterait comme suit :



9. Double-cliquez sur le stage de collecteur de données et configurez-le.

Pour obtenir des informations sur la configuration des stages de collecteur de données, reportez-vous au *Guide du concepteur de flux de données*.

Vous disposez désormais d'un flux de données identifiant les enregistrements correspondants et supprimant tous les enregistrements, sauf un, de chaque groupe de doublons, ce qui produit un fichier de sortie contenant des données déduplicuées.

## Création d'un enregistrement Best of Breed

Pour éliminer des enregistrements en double de vos données, vous pouvez choisir de fusionner des données issues de groupes d'enregistrements en double en un enregistrement « best of breed » unique. Cette approche s'avère utile lorsque chaque enregistrement en double contient des données du même type (par exemple, des numéros de téléphone ou des noms) et que vous souhaitez préserver les meilleures données de chaque enregistrement dans l'enregistrement restant.

Cette procédure décrit la manière dont créer un flux de données qui fusionne des enregistrements en double en un enregistrement best of breed.

1. Dans Enterprise Designer, créez un flux de données identifiant les enregistrements en double à travers une mise en correspondance.

La mise en correspondance est la première étape dans la déduplication, car vous devez identifier des enregistrements similaires, tels que des enregistrements disposant du même nom ou numéro de compte. Pour en savoir plus sur la création d'un flux de données qui met en correspondance les enregistrements, reportez-vous aux rubriques suivantes.

[Mise en correspondance d'enregistrements d'une source unique](#) à la page 87

[Mise en correspondance d'enregistrements d'une source avec ceux d'une autre source](#) à la page 93

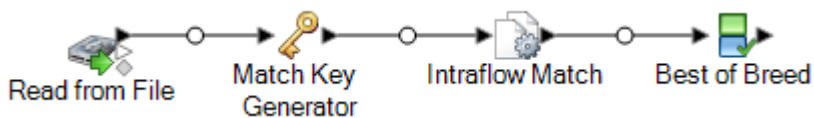
[Mise en correspondance d'enregistrements par rapport à une base de données](#) à la page 106

**Remarque :** Vous devez uniquement créer le flux de données au stade où il lit les données et effectue une mise en correspondance avec un stage Interflow Match, Intraflow Match ou Transactional Match. Une fois que vous avez créé un flux de données à ce stade, continuez avec les étapes suivantes.

2. Une fois que vous avez défini un flux de données qui lit les données et met en correspondance les enregistrements, faites glisser un stage Best of Breed sur le canevas et connectez-le au stage qui effectue la mise en correspondance (Interflow Match, Intraflow Match ou Transactional Match).

Par exemple, si votre flux de données lit des données à partir d'un fichier et effectue une mise en correspondance avec Intraflow Match, votre flux de données ressemblera à ce qui suit, après avoir ajouté un stage Best of Breed :





3. Double-cliquez sur le stage Best of Breed sur le canevas.
4. Dans le champ **Grouper par**, sélectionnez **NuméroCollection**.
5. Sous **Paramètres Best Of Breed**, sélectionnez **Règles** dans l'arborescence des conditions.
6. Cliquez sur **Ajouter une règle**.

Les enregistrements de chaque groupe sont évalués pour déterminer s'ils respectent les règles que vous définissez ici. Si un enregistrement correspond à une règle, ses données peuvent être copiées dans l'enregistrement best of breed, selon la manière dont vous configurez les actions associées à la règle. Vous définirez les actions ultérieurement.

7. Définissez une règle qu'un enregistrement en double doit respecter pour que ses données soient copiées dans l'enregistrement best of breed.

Utilisez les options suivantes pour définir une règle :

Option	Description
Nom du champ	Indique le nom du champ de flux de données dont vous souhaitez évaluer la valeur afin de déterminer si la condition est remplie, ainsi que les actions associées à effectuer.
Type de champ	Indique le type de données dans le champ. L'un des éléments suivants : <ul style="list-style-type: none"> <li><b>Non-numérique</b>      Choisissez cette option si le champ contient des données non numériques (par exemple, des données de chaîne).</li> <li><b>numériques</b>        Choisissez cette option si le champ contient des données numériques (par exemple, double, flottantes, etc.).</li> </ul>

Option	Description
--------	-------------

---

Opérateur	
-----------	--

Option	Description
	Indique le type de comparaison à utiliser pour évaluer le champ. L'un des éléments suivants :
<b>Contient</b>	Détermine si le champ contient la valeur indiquée. Par exemple, « bateau à voile » contient la valeur « bateau ».
<b>Est égal à</b>	Détermine si le champ contient la valeur exacte indiquée.
<b>Est supérieur à</b>	Détermine si la valeur du champ est supérieure à la valeur spécifiée. Cette opération ne fonctionne que sur les champs numériques.
<b>Supérieur ou Egal à</b>	Détermine si la valeur du champ est supérieure ou égale à la valeur spécifiée. Cette opération ne fonctionne que sur les champs numériques.
<b>Le plus haut</b>	Compare la valeur du champ pour tout le groupe d'enregistrements et identifie l'enregistrement qui possède la valeur la plus élevée dans le champ. Par exemple, si les champs dans ce groupe contiennent les valeurs 10, 20, 30, et 100, l'enregistrement qui dispose de la valeur 100 est sélectionné. Cette opération ne fonctionne que sur les champs numériques. Si plusieurs enregistrements sont à égalité pour la valeur la plus longue, un enregistrement est sélectionné.
<b>Est vide</b>	Détermine si le champ ne contient aucune valeur.
<b>N'est pas vide</b>	Détermine si le champ contient une valeur.
<b>Inférieur à</b>	Détermine si la valeur du champ est inférieure à la valeur spécifiée. Cette opération ne fonctionne que sur les champs numériques.
<b>Inférieur ou Egal à</b>	Détermine si la valeur du champ est inférieure ou égale à la valeur spécifiée. Cette opération ne fonctionne que sur les champs numériques.
<b>Le plus long</b>	Compare la valeur du champ pour tout le groupe d'enregistrements et identifie l'enregistrement qui possède la valeur la plus longue (en octets) dans le champ. Par exemple, si le groupe contient les valeurs « Mike » et « Michael », l'enregistrement possédant la valeur « Michael » est choisi. Si plusieurs enregistrements sont à égalité pour la valeur la plus longue, un enregistrement est sélectionné.
<b>Le plus bas</b>	Compare la valeur du champ pour tout le groupe d'enregistrements et identifie l'enregistrement qui possède la valeur la plus basse dans le champ. Par exemple, si les champs dans ce groupe contiennent les valeurs 10, 20, 30, et 100, l'enregistrement qui dispose de la valeur 10 est sélectionné. Cette opération ne fonctionne que sur les champs numériques. Si plusieurs enregistrements sont à égalité pour la valeur la plus longue, un enregistrement est sélectionné.
<b>Le plus commun</b>	Détermine si la valeur de champ contient la valeur qui se produit le plus fréquemment dans ce champ parmi les enregistrements dans ce groupe. Si deux valeurs, ou plus, sont les plus courantes, aucune action ne sera réalisée.

Option	Description
	<p><b>N'est pas égal</b> Détermine si la valeur du champ n'est pas égale à la valeur spécifiée.</p>
Type de valeur	<p>Indique le type de valeur à comparer avec la valeur du champ. L'un des éléments suivants :</p> <p><b>Remarque</b> : Cette option n'est pas disponible si vous sélectionnez l'opérateur Highest, Lowest ou Longest.</p> <p><b>Champ</b> Choisissez cette option si vous souhaitez comparer une autre valeur du champ de dataflow avec le champ.</p> <p><b>Chaîne</b> Choisissez cette option si vous souhaitez comparer le champ avec une valeur spécifique.</p>
Valeur	<p>Indique la valeur à comparer avec la valeur du champ. Si vous avez sélectionné Champ dans le champ <b>Type de champ</b>, sélectionnez un champ de dataflow. Si vous avez sélectionné Chaîne dans le champ <b>Type de valeur</b>, tapez la valeur à utiliser dans la comparaison.</p> <p><b>Remarque</b> : Cette option n'est pas disponible si vous sélectionnez l'opérateur Highest, Lowest ou Longest.</p>

8. Cliquez sur **OK**.
9. Cliquez sur le nœud **Actions** dans l'arborescence.
10. Cliquez sur **Ajouter une action**.
11. Indiquez les données à copier dans l'enregistrement best of breed si l'enregistrement répond aux critères que vous avez défini dans la règle.

Option	Description
Type de la source	<p>Indique le type de données à copier dans l'enregistrement Best Of Breed. Un des éléments suivants.</p> <p><b>Champ</b> Choisissez cette option pour copier une valeur d'un champ dans l'enregistrement Best Of Breed.</p> <p><b>Chaîne</b> Choisissez cette option pour copier une valeur constante dans l'enregistrement Best Of Breed.</p>

Option	Description
Données sources	Indique les données à copier dans l'enregistrement Best Of Breed. Si le type de source est Field, sélectionnez le champ dont vous souhaitez copier la valeur dans le champ de destination. Si le type de source est String, indiquez une valeur constante à copier dans le champ de destination.
Cible	Indique le champ dans l'enregistrement Best Of Breed dans lequel vous souhaitez copier les données spécifiées dans le champ <b>Données source</b> .
Accumuler les données sources	<p>Si les données dans le champ <b>Données source</b> sont des données numériques, vous pouvez activer cette option afin de combiner les données source pour tous les enregistrements doublons et insérer la valeur totale dans l'enregistrement Best Of Breed.</p> <p>Par exemple, s'il existe trois enregistrements doublons dans le groupe et que ceux-ci contenaient ces valeurs dans le champ Deposits :</p> <p>100.00 20.00 5.00</p> <p>Les trois valeurs serait combinées et la valeur totale (125,00) serait insérée dans le champ Deposits de l'enregistrement Best Of Breed.</p>

12. Cliquez sur **OK**.

Vous avez désormais configuré Best Of Breed avec une règle et une action. Vous pouvez ajouter d'autres règles et actions si nécessaire.

13. Cliquez sur **OK** pour fermer la fenêtre **Options Best of Breed**.

14. Faites glisser un stage de collecteur de données sur le canevas et connectez-le au stage Best of Breed.

Par exemple, si vous utilisiez un stage de collecteur de données Write To File, votre flux de données se présenterait comme suit :



15. Double-cliquez sur le stage de collecteur de données et configurez-le.

Pour obtenir des informations sur la configuration des stages de collecteur de données, reportez-vous au *Guide du concepteur de flux de données*.

Vous disposez désormais d'un flux de données identifiant les enregistrements correspondants et fusionnant les enregistrements en une collection dans un enregistrement Best Of Breed.

# 6 - Enregistrements d'exception

## In this section

---

Conception d'un flux de donnée afin de gérer les exceptions	160
Désignation d'un dataflow à des fins de revalidation en temps réel	162

## Conception d'un flux de donnée afin de gérer les exceptions

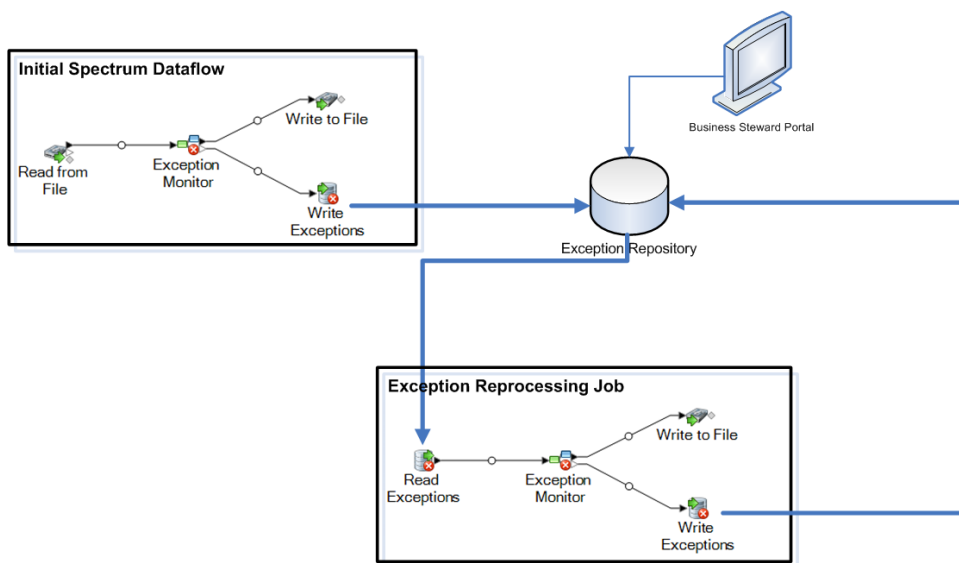
Si vous détenez une licence du module Business Steward, vous pouvez inclure un processus de gestion des exceptions dans vos flux de données. Les principales composantes d'une procédure de gestion d'exception sont :

- Un flux de donnée effectuant une procédure de qualité de donnée, tel que la déduplication d'enregistrement, la validation d'adresse ou le géocodage.
- Un stage Exception Monitor qui identifie les enregistrements qui n'ont pas pu être traités.
- Un stage Write Exceptions prend les enregistrements d'exceptions identifiés par le stage Exception Monitor et les écrit dans le référentiel d'exception afin qu'ils soient manuellement vérifiés.
- Le Business Steward Portal, outil basé sur un navigateur, vous permet de vérifier et de modifier les enregistrements d'exceptions. Une fois modifiés, les enregistrements sont marqués comme « approuvés », ce qui les rend prêts au retraitement.
- Un job de retraitement d'exceptions qui utilise le stage Read Exceptions pour lire les enregistrements approuvés du référentiel d'exception vers le job. Le job tente ensuite de retraiter les enregistrements corrigés, généralement à l'aide de la même logique que celle du flux de données initial. Le stage Exception Monitor recherche encore une fois les exceptions. Le stage Write Exceptions renvoie ensuite les exceptions au référentiel d'exception afin de procéder à une vérification supplémentaire.

**Remarque :** Ne placez pas d'autres stages entre les stages Exception Monitor et Write Exceptions dans un flux de données ; cela pourrait avoir un impact sur la configuration des exceptions dans Business Steward Portal.

Voici un scénario type permettant de mieux comprendre la mise en place de base de gestion d'exception :





Dans cet exemple, il existe deux flux de données : le flux de données initial, qui évalue les données de code postal des enregistrements d'entrée, et le job de retraitement des exceptions, qui prend les exceptions modifiées et vérifie que les enregistrements contiennent maintenant les données de code postal valides.

Dans les deux flux de données figure un stage Exception Monitor. Ce stage contient les conditions que vous souhaitez utiliser pour déterminer si un enregistrement doit être routé pour une révision manuelle. Ces conditions sont constituées d'une ou de plusieurs expressions, comme `PostalCode is empty`, qui signifie que tout enregistrement ne contenant pas de code postal doit être considéré comme une exception, routé vers le stage Write Exceptions et écrit dans le référentiel d'exception. Pour plus d'informations, reportez-vous à la section [Exception Monitor](#) à la page 251.

Tout enregistrement que l'Exception Monitor pourrait identifier comme étant une exception serait routé vers un dépositaire d'exception grâce à l'étape Écrire les exceptions. Les data stewards examinent les exceptions dans le dépositaire en utilisant le Business Steward Portal, un outil basé sur navigateur permettant de voir et de modifier des enregistrements d'exception. Dans notre exemple, le data steward peut utiliser Exception Editor dans le Business Steward Portal pour ajouter manuellement des codes postaux aux enregistrements d'exception et les marquer comme « approuvés ».

Une fois qu'un enregistrement est marqué comme « approuvé » dans le Business Steward Portal, l'enregistrement peut être relu dans un flux de données Spectrum™ Technology Platform. Le stage Read Exceptions réalise cette action. Si des enregistrements continuent à constituer des exceptions, ils seront à nouveau écrits dans le dépositaire d'exception afin d'être examinés par un data steward.

Pour définir la meilleure approche possible dans votre cas, considérez ces questions :

- **Comment voulez-vous identifier les enregistrements d'exception ?** Le stage Exception Monitor peut évaluer la valeur de n'importe quel champ ou n'importe quelle combinaison de champs afin de déterminer si un enregistrement est une exception. Vous pouvez analyser les résultats que vous obtenez actuellement avec votre flux de données afin de savoir comment vous souhaitez identifier les exceptions. Vous devriez peut être identifier les enregistrements dans la fourchette

moyenne du continuum de qualité de données et non pas ceux qui ont été clairement validés ou qui ont clairement échoué.

- **Vous voulez-vous traiter à nouveau les enregistrements d'exception édités et approuvés grâce à la même logique utilisée dans le flux de données initial.** Dans ce cas, vous devriez peut être utiliser un sous-flux afin de créer une logique métier réutilisable. Par exemple, le sous-flux peut être utilisé dans un flux de données initial effectuant une validation d'adresses et dans un job de retraitement des exceptions qui traite de nouveau les enregistrements corrigés afin de vérifier les corrections. Vous pouvez ensuite utiliser des sources et des étapes différentes entre les deux. Le flux de données initial peut contenir une étape Lire depuis la base de données qui récupère des données depuis votre base de données de clients afin qu'elles soient traitées. Le job de retraitement des exceptions contiendrait un stage Read Exceptions qui récupère les enregistrements d'exception modifiés et approuvés du référentiel d'exception.
- **Vous voulez-vous traiter à nouveau les exceptions corrigées et approuvées selon un planning pré-défini ?** Si c'est le cas, vous pouvez programmer votre job de retraitement via Planification dans Management Console.

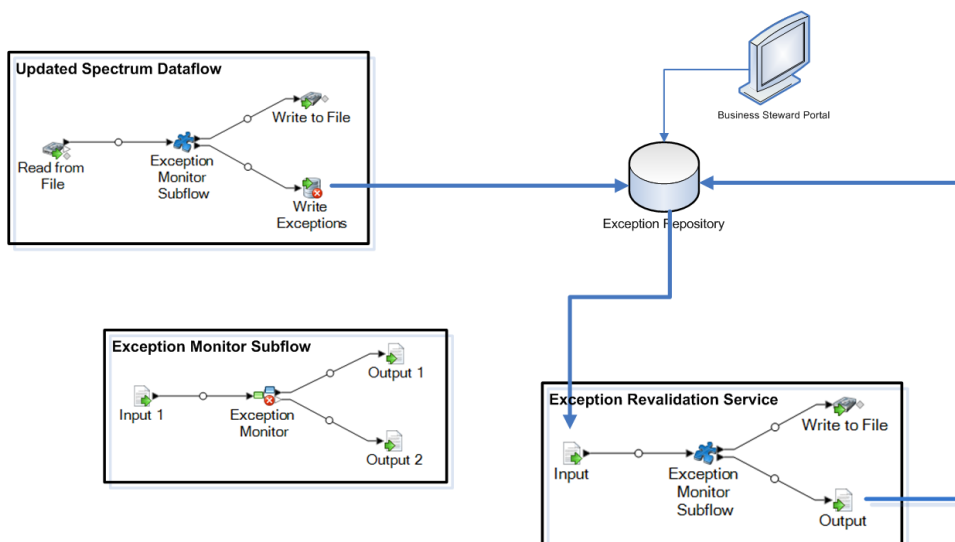
## Désignation d'un dataflow à des fins de revalidation en temps réel

Si vous utilisez la gestion des exceptions dans votre dataflow, vous pouvez utiliser la fonction de revalidation pour exécuter de nouveau des enregistrements d'exception via le processus de validation après les avoir corrigés dans le Business Steward Portal. Cette opération vous permet de déterminer si les modifications que vous avez apportées permettent le traitement correct en temps réel de l'enregistrement ; vous n'avez pas besoin d'attendre la réexécution du job par lot Read Exceptions pour voir le résultat.

Les blocs de construction de base d'un environnement de revalidation sont les suivants :

- Un job ou un service qui réutilise ou contient un sous-flux exposé. Il doit également contenir une source d'entrée, le stage Subflow qui traite l'entrée, un stage Write Exceptions et un récepteur de sortie des enregistrements correctement traités.
- Un sous-flux exposé contenant un stage Exception Monitor pointant vers un service de revalidation et configuré pour la revalidation, y compris la désignation des enregistrements comme des enregistrements à retraiter ou à approuver.
- Un service exposé qui réutilise ou contient également le sous-flux exposé. Il traite les enregistrements qui ont été modifiés, enregistrés et envoyés pour revalidation au Business Steward Portal.

Voici un scénario type permettant de mieux comprendre l'implémentation d'une revalidation :



Cet exemple présente dataflow : un job, un sous-flux et un service. Le job exécute les données d'entrée via le sous-flux. Le sous-flux contient un stage Exception Monitor, qui détermine si un enregistrement doit être routé à des fins de vérification manuelle. Si l'on poursuit avec notre exemple, cela signifie que tout enregistrement sans données dans le champ PostalCode est considéré comme une exception et dirigé vers le stage Write Exceptions ; ces exceptions apparaissent dans le Business Steward Portal. Les enregistrements contenant toute autre donnée dans ce champ sont dirigés vers le stage Write to File.

**Remarque :** Si votre dataflow est également configuré pour utiliser la fonctionnalité Best Of Breed, vous devrez ajouter et exposer manuellement le champ CollectionRecordType dans le stage/sous-flux de revalidation Exception Monitor et le service lui-même. Reportez-vous aux options [Write Exceptions](#) et à [Création d'un enregistrement Best of Breed](#) à la page 281 pour obtenir davantage d'informations sur la fonctionnalité Best Of Breed.

Le service de revalidation des exceptions que vous avez désigné lors de la configuration du stage Exception Monitor est appelé lorsque vous modifiez un ou plusieurs enregistrements d'exception dans l'Éditeur d'exceptions de Business Steward Portal et que vous cliquez sur **Enregistrer**. Comme le job, le service contient le sous-flux Exception Monitor, qui utilise la même logique commerciale pour retraiter le ou les enregistrements. Si les enregistrements ne remplissent pas une ou plusieurs conditions définies dans le stage Exception Monitor, les exceptions sont mises à jour dans le référentiel. Si les enregistrements remplissent les conditions définies dans le stage Exception Monitor, une ou deux actions se produisent, suivant la sélection effectuée dans le champ « Action après revalidation » :

- **Traiter à nouveau les enregistrements** : les enregistrements sont supprimés du référentiel et retraités.
- **Approuver les enregistrements** : les enregistrements sont marqués comme approuvés et renvoyés au référentiel.

Pour créer et utiliser un scénario de revalidation en temps réel, procédez comme suit :

1. Ouvrez ou créez un dataflow de job ou de service contenant un stage Exception Monitor, une source d'entrée (comme un stage Read from File ou Input), un récepteur de sortie (comme un stage Write to File ou Output) et un stage Write Exceptions.
2. Convertissez le stage Exception Monitor en un sous-flux et mappez les champs d'entrée et de sortie de sorte qu'ils correspondent à ceux du dataflow initial. Veillez à inclure le champ ExceptionMetadata pour la source d'entrée ainsi que le stage Output renseignant le stage Write Exceptions du job. Exposez le sous-flux de sorte qu'il puisse être utilisé par le job et le service.
3. Créez un service contenant un stage Input, le sous-flux que vous avez créé à l'étape 2, un stage Output et un récepteur de sortie (comme un stage Write to File ou Write to DB). Mappez les champs d'entrée et de sortie de sorte qu'ils correspondent à ceux du sous-flux initial ; veillez à inclure le champ ExceptionMetadata pour le stage Input et le stage Output. Exposez le service de sorte qu'il puisse être utilisé par le sous-flux.
4. Renvoyez le sous-flux et ouvrez l'onglet Configuration du stage Exception Monitor. Sélectionnez le service de revalidation que vous avez créé à l'étape 3 et indiquez l'action à effectuer après la revalidation. Enregistrez et exposez de nouveau le sous-flux.
5. Revenez au service, où un message apparaît, vous informant des modifications apportées au sous-flux et indiquant que le service va être actualisé. Cliquez sur **OK**, puis enregistrez et exposez de nouveau le service.
6. Revenez au job ou au service initial, où un message apparaît, vous informant des modifications apportées au sous-flux et indiquant que le dataflow va être actualisé. Cliquez sur **OK**, puis enregistrez le dataflow.
7. Exécutez le job.

**Remarque :** Même si vous avez précédemment exécuté le job ou le service initial, vous devez le réexécuter après avoir créé le scénario de revalidation pour enregistrer les enregistrements éligibles à la validation dans le référentiel. Vous pouvez identifier les enregistrements éligibles à la revalidation figurant dans l'Éditeur d'exceptions, car le bouton **Enregistrer** est actif pour ces enregistrements.

# 7 - Tables de recherche

## [In this section](#)

---

Présentation des tables de recherche	166
Tables du module Data Normalization	166
Tables du module Universal Name	172
Affichage du contenu d'une table de recherche	173
Ajout d'un terme dans une table de recherche	174
Suppression d'un terme d'une table de recherche	175
Modification de la forme standardisée d'un terme	175
Rétablissement de la Personnalisation de la table	175
Création d'une table de recherche	176
Importer des données	177

## Présentation des tables de recherche

Une table de recherche est une table de paires de valeurs clés utilisée par les stages de Spectrum™ Technology Platform pour normaliser les données en effectuant un remplacement de jeton. Pour modifier le contenu des tables de recherche utilisé dans Advanced Transformer, Open Parser et Table Lookup, utilisez l'outil Gestion de table dans Enterprise Designer.

## Tables du module Data Normalization

### Tables Advanced Transformer

Advanced Transformer utilise les tables suivantes pour identifier les termes. Utilisez la Gestion de Table pour créer de nouvelles tables ou pour modifier des tables existantes. Pour plus d'informations, reportez-vous à la section [Présentation des tables de recherche](#) à la page 166.

- Abréviations aéronautiques
- Sigles de tous les acronymes
- Abréviations de noms d'entreprise
- Abréviations du territoire canadien
- Abréviations informatiques
- Délimiteurs
- Entreprises allemandes
- Fortune 1000
- Abréviations de point cardinal
- Termes de bruit Global Sentry
- Pays sanctionnés par Global Sentry
- Abréviations d'institutions gouvernementales
- Code de compagnies aériennes IATA
- Pays du code de compagnies aériennes IATA
- Abréviations juridiques
- Abréviations médicales
- Acronymes d'associations médicales
- Abréviations militaires
- Surnoms

- Abréviations d'unité secondaire
- Unité secondaire inversée
- Abréviations de Singapour
- Abréviations espagnoles
- Abréviations de point cardinal espagnol
- Abréviations de suffixe de rue espagnol
- Abréviations de nom d'État
- Nom d'État inversé
- Abréviations de suffixe de rue
- Suffixe de rue inversé
- Filiale d'une société mère
- Acronymes de l'armée américaine
- Acronymes de la marine américaine

## Tables de Open Parser

Open Parser utilise les tables suivantes pour identifier des termes. Utilisez la Gestion de Table pour créer de nouvelles tables ou pour modifier des tables existantes. Pour plus d'informations, reportez-vous à la section [Présentation des tables de recherche](#) à la page 166.

### *Tables de base*

Les tables de base sont fournies avec le paquet d'installation Module Data Normalization.

- Descriptions du compte
- Entreprises
- Conjonctions d'entreprise
- Prépositions d'entreprise
- Suffixes d'entreprise
- Termes d'entreprise
- Conjonctions
- Préfixes de nom de famille
- Noms de famille
- Suffixes généraux
- Entreprises allemandes
- Prénoms
- Suffixes de maturité
- Prénoms espagnols
- Noms de famille espagnols
- Titres

### Tables Core Names

Les tables Core Names ne sont pas fournies avec le kit d'installation du module Data Normalization et requièrent donc une licence supplémentaire. Pour plus d'informations, contactez votre responsable de compte.

Les tables de Noms de noyau doivent être chargées à l'aide de l'utilitaire de chargement de base de données du Module Data Normalization. Pour obtenir des instructions, reportez-vous au *Guide d'installation Spectrum™ Technology Platform*.

- Noms de famille améliorés
- Prénoms améliorés

### Tables Company Names

Les tables Company Names ne sont pas fournies avec le kit d'installation du module Data Normalization et requièrent donc une licence supplémentaire. Pour plus d'informations, contactez votre responsable de compte.

Les tables de Noms de société doivent être chargées à l'aide de l'utilitaire de chargement de base de données du Module Data Normalization. Pour obtenir des instructions, reportez-vous au *Guide d'installation Spectrum™ Technology Platform*.

- Sociétés – Amériques
- Sociétés – Asie-Pacifique
- Sociétés – EMEA
- Articles d'entreprise
- Conjonctions d'entreprise

### Tables Arabic Plus Pack

Les tables Arabic Plus Pack ne sont pas fournies avec le kit d'installation du module Data Normalization et requièrent donc une licence supplémentaire. Pour plus d'informations, contactez votre responsable de compte.

Les tables du pack Arabic Plus Pack doivent être chargées à l'aide de l'utilitaire de chargement de base de données du Module Data Normalization. Pour obtenir des instructions, reportez-vous au *Guide d'installation Spectrum™ Technology Platform*.

- Noms de famille arabe (écriture arabe)
- Noms de famille arabes (écriture latine)
- Prénoms arabes (écriture arabe)
- Prénoms arabes (écriture latine)



### Tables Asian Plus Pack

Les tables Asian Plus Pack ne sont pas fournies avec le kit d'installation du module Data Normalization et requièrent donc une licence supplémentaire. Pour plus d'informations, contactez votre responsable de compte.

Les tables du pack Asian Plus doivent être chargées à l'aide de l'utilitaire de chargement de base de données du Module Data Normalization. Pour obtenir des instructions, reportez-vous au *Guide d'installation Spectrum™ Technology Platform*.

- Noms de famille chinois (natif)
- Noms de famille chinois (écriture latine)
- Prénoms chinois (natif)
- Prénoms chinois (écriture latine)
- Noms de famille coréens (natif)
- Noms de famille coréens (écriture latine)
- Prénoms coréens (natif)
- Prénoms coréens (écriture latine)
- Noms de famille japonais (kana)
- Noms de famille japonais (kanji)
- Noms de famille japonais (écriture latine)
- Prénoms japonais (kana)
- Prénoms japonais (kanji)
- Prénoms japonais (écriture latine)

## Tables Table Lookup

Table Lookup utilise les tables suivantes pour identifier des termes. Utilisez la Gestion de Table pour créer de nouvelles tables ou pour modifier des tables existantes. Pour plus d'informations, reportez-vous à la section [Présentation des tables de recherche](#) à la page 166.

### Tables de base

Les tables de base sont fournies avec le paquet d'installation Module Data Normalization.

- Abréviations aéronautiques
- Sigles de tous les acronymes
- Abréviations de noms d'entreprise
- Abréviations du territoire canadien
- Abréviations informatiques
- Acronymes de l'EU
- Fortune 1000
- Abréviations françaises
- Arrondissement à numéro de département français

- Communes à codes postaux français
- Départements à régions français
- Numéro de départements à départements français
- Codes de sexe
- Abréviations de point cardinal
- Acronymes allemands
- Ville à code d'État allemand
- Code de zone à ville allemand
- District à code d'État allemand
- Abréviations des États allemands
- Pays sanctionnés par Global Sentry
- Abréviations d'institutions gouvernementales
- Code de compagnies aériennes IATA
- Pays du code de compagnies aériennes IATA
- Abréviations juridiques
- Abréviations médicales
- Acronymes d'associations médicales
- Abréviations militaires
- Surnoms
- Abréviations d'unité secondaire
- Unité secondaire inversée
- Abréviations de Singapour
- Abréviations espagnoles
- Abréviations de point cardinal espagnol
- Abréviations de suffixe de rue espagnol
- Abréviations de nom d'État
- Nom d'État inversé
- Abréviations de suffixe de rue
- Suffixe de rue inversé
- Filiale d'une société mère
- Villes et zones de codes postaux britanniques
- Préfixes des codes téléphoniques britanniques
- Codes téléphoniques des villes britanniques
- Zones de codes postaux et villes britanniques
- Acronymes de l'armée américaine
- Acronymes de la marine américaine
- ZREPLACE (utilisé par le module SAP pour la validation d'adresses françaises)

### *Noms de noyau*

Les tables de Noms de noyau nécessitent une licence supplémentaire. Pour plus d'informations, contactez votre responsable de compte.

Les tables de Noms de noyau doivent être chargées à l'aide de l'utilitaire de chargement de base de données du Module Data Normalization. Pour obtenir des instructions, reportez-vous au *Guide d'installation Spectrum™ Technology Platform*.

- Amélioration des noms de famille ethniques
- Codes de sexe améliorés
- Amélioration de prénoms de famille ethniques

### *Pack Arabic Plus*

Les tables du pack Arabic Plus nécessitent une licence supplémentaire. Pour plus d'informations, contactez votre responsable de compte.

Les tables du pack Arabic Plus Pack doivent être chargées à l'aide de l'utilitaire de chargement de base de données du Module Data Normalization. Pour obtenir des instructions, reportez-vous au *Guide d'installation Spectrum™ Technology Platform*.

- Noms de famille arabes ethniques (écriture arabe)
- Noms de famille arabes ethniques (écriture latine)
- Codes de sexe arabes (écriture arabe)
- Codes de sexe arabes (écriture latine)
- Prénoms arabes ethniques (écriture arabe)
- Prénoms arabes ethniques (écriture latine)

### *Pack Asian Plus*

Les tables du pack Asian Plus Pack nécessitent une licence supplémentaire. Pour plus d'informations, contactez votre responsable de compte.

Les tables du pack Asian Plus doivent être chargées à l'aide de l'utilitaire de chargement de base de données du Module Data Normalization. Pour obtenir des instructions, reportez-vous au *Guide d'installation Spectrum™ Technology Platform*.

- Noms de famille CJK ethniques (natif)
- Noms de famille CJK ethniques (écriture latine)
- Prénoms CJK ethniques (natif)
- Prénoms CJK ethniques (écriture latine)
- Codes de sexe japonais (kana)
- Codes de sexe japonais (kanji)
- Codes de sexe japonais (écriture latine)

## Tables du module Universal Name

### Tables Name Variant Finder

Le stage Localisateur de variantes de nom utilise les tables suivantes. Chaque table nécessite une licence distincte.

- Arabic Plus Pack : `g1-cdq-cjki-arabic-<date>.jar`
- Pack Asian Plus - Chinois : `g1-cdq-cjki-chinese-<date>.jar`
- Pack Asian Plus - Japonais : `g1-cdq-cjki-japanese-<date>.jar`
- Asian Plus Pack - Coréen : `g1-cdq-cjki-korean-<date>.jar`
- Base de données Core Names : `g1-cdq-nomino-base-<date>.jar`

### Tables Open Name Parser

Open Name Parser utilise les tables suivantes pour identifier des termes. Utilisez la Gestion de Table pour créer de nouvelles tables ou pour modifier des tables existantes. Pour plus d'informations, reportez-vous à la section [Présentation des tables de recherche](#) à la page 166.

#### *Tables de base*

Les tables de base sont fournies avec le kit d'installation du module Universal Name.

- Descriptions du compte
- Conjonctions d'entreprise
- Conjonctions
- Préfixes de nom de famille
- Noms de famille
- Suffixes généraux
- Prénoms
- Suffixes de maturité
- Prénoms espagnols
- Noms de famille espagnols
- Titres

#### *Tables Core Names*

Les tables Core Names ne sont pas fournies avec le kit d'installation du module Universal Name et requièrent donc une licence supplémentaire.

- Noms de famille améliorés
- Prénoms améliorés

#### *Tables Company Names*

Les tables Company Names suivantes sont fournies avec le kit d'installation du module Universal Name.

- Descriptions du compte
- Entreprises
- Articles d'entreprise
- Conjonctions d'entreprise
- Prépositions d'entreprise
- Suffixes d'entreprise
- Termes d'entreprise
- Conjonctions

Les tables Company Names suivantes ne sont pas fournies avec le kit d'installation du module Universal Name et requièrent donc une licence supplémentaire.

- Sociétés – Amériques
- Sociétés – Asie-Pacifique
- Sociétés – EMEA

#### *Tables Asian Plus Pack*

Les tables Asian Plus Pack ne sont pas fournies avec le kit d'installation du module Universal Name et requièrent donc une licence supplémentaire.

- Noms de famille japonais (kana)
- Noms de famille japonais (kanji)
- Noms de famille japonais (écriture latine)
- Prénoms japonais (kana)
- Prénoms japonais (kanji)
- Prénoms japonais (écriture latine)
- Civilités japonaises

## Affichage du contenu d'une table de recherche

Vous pouvez afficher le contenu d'une table de recherche à l'aide de l'outil Gestion de table dans Enterprise Designer.

1. Dans Enterprise Designer, sélectionnez **Outils > Gestion de table**.
2. Dans le champ **Type**, sélectionnez le stage dont vous souhaitez afficher la table de recherche.
3. Dans le champ **Nom**, sélectionnez la table à afficher.

4. Vous pouvez utiliser les options suivantes pour modifier la manière dont la table est affichée :

Option	Description
<b>Trouvez un terme spécifique</b>	Dans le champ <b>Commence par</b> , tapez le terme à rechercher, puis cliquez sur <b>Actualiser</b> .
<b>Navigation dans la table</b>	Cliquez sur les icônes avant et arrière situées à droite du bouton <b>Actualiser</b> .
<b>Modifiez le nombre de termes affichés par page</b>	Modifiez la valeur dans le champ <b>Éléments par page</b> .
<b>Affichez tous les termes de recherche pour chaque terme standardisé dans une table Table Lookup.</b>	Dans le champ <b>Afficher par</b> , sélectionnez <b>Terme standardisé (regroupement)</b> . Cette option n'est disponible que pour les tables Table Lookup

## Ajout d'un terme dans une table de recherche

Si vous remarquez que vos données présentent des termes qui ne sont pas inclus dans la table de recherche et que vous souhaitez ajouter le terme à une table de recherche, suivez cette procédure.

1. Dans Enterprise Designer, sélectionnez **Outils > Gestion de table**.
2. Dans le champ **Type**, sélectionnez le stage dont vous souhaitez modifier la table de recherche.
3. Dans le champ **Nom**, sélectionnez la table dans laquelle vous souhaitez ajouter un terme.
4. Cliquez sur **Ajouter**.
5. Dans le champ **Rechercher un terme**, saisissez le terme qui existe dans vos données. Il s'agit de la clé de recherche qui sera utilisée.
6. Pour les tables Table Lookup, dans le champ **Terme standardisé**, entrez le terme qui doit remplacer le terme de recherche dans votre flux de données.

Par exemple, si vous souhaitez modifier le terme PB à Pitney Bowes, vous devez entrer PB comme terme de recherche et Pitney Bowes comme terme standardisé.

7. Pour les tables Table Lookup, cochez la case **Écraser le terme existant** si un terme existe déjà dans la table et que vous désirez le remplacer par la valeur que vous avez saisie à l'étape 5.
8. Cliquez sur **Ajouter**.

## Suppression d'un terme d'une table de recherche

Pour supprimer un terme d'une table de recherche, procédez comme suit :

1. Dans Enterprise Designer, sélectionnez **Outils > Gestion de table**.
2. Sélectionnez le terme et cliquez sur **Supprimer**.
3. Cliquez sur **Oui** pour supprimer le terme de la table.

## Modification de la forme standardisée d'un terme

Pour les tables utilisées par Table Lookup pour standardiser les termes, vous pouvez modifier la forme standardisée d'un terme. Par exemple, si vous disposez d'une table dotée des termes de recherche PB et PB Software, que le terme standardisé est Pitney Bowes, et que vous souhaitez modifier la forme standardisé pour Pitney Bowes Inc, vous pouvez procéder comme suit pour y parvenir.

1. Dans Enterprise Designer, sélectionnez **Outils > Gestion de table**.
2. Dans le champ **Type**, sélectionnez **Recherche dans la table**.
3. Dans le champ **Nom**, sélectionnez la table à modifier.
4. Sélectionnez le terme que vous voulez modifier et cliquez sur **Modifier**.

**Conseil :** S'il existe plusieurs termes de recherche pour un terme standardisé, vous pouvez aisément modifier tous les termes de recherche pour utiliser le nouveau terme standardisé en sélectionnant **Afficher par terme standardisé (regroupement)** dans le champ **Afficher par**, en sélectionnant le groupe et en cliquant sur **Modifier**.

5. Saisissez une nouvelle valeur dans le champ **Terme normalisé**.
6. Cliquez sur **OK**.

## Rétablissement de la Personnalisation de la table

Si vous apportez des modifications à une table, vous pouvez rétablir l'état d'origine de la table. Pour rétablir la personnalisation de la table :

1. Dans Enterprise Designer, sélectionnez **Outils > Gestion de table**.
2. Sélectionnez la table à rétablir.

3. Cliquez sur **Rétablir**.

La fenêtre **Rétablir** s'affiche. Elle répertorie tous les termes ajoutés, supprimés et modifiés.

4. Cochez la case **Rétablir** de chaque entrée que vous désirez rétablir. Vous pouvez aussi cliquer sur **Sélectionner tout** ou **Désélectionner tout** pour sélectionner ou effacer toutes les cases à cocher **Rétablir**.
5. Cliquez sur **OK**.

## Création d'une table de recherche

Le module Advanced Matching, le module Data Normalization et le module Universal Name présentent différentes tables que vous pouvez utiliser pour une large palette de processus de remplacement de terme et de standardisation. Cependant, si ces tables ne répondent pas à vos besoins, vous pouvez créer votre propre table de termes de recherche à utiliser avec Advanced Transformer, Open Parser ou Table Lookup. Pour créer une table, suivez cette procédure.

1. Dans Enterprise Designer, sélectionnez **Outils > Gestion de table**.
2. Dans le champ **Type**, sélectionnez le stage pour lequel vous souhaitez créer une table de recherche.
3. Cliquez sur **Nouveau**. La boîte de dialogue **Ajouter une table** s'affiche.
4. Dans le champ **Nom de la table**, entrez le nom de la nouvelle table.
5. Si vous désirez une nouvelle table vierge du type sélectionné, laissez **Copier à partir de** défini sur **Aucune**. Si vous souhaitez que la nouvelle table soit renseignée à partir d'une table existante, sélectionnez un nom de table dans la liste **Copier à partir de**.
6. Cliquez sur **OK**.

Pour des informations à propos de l'ajout d'éléments d'une table dans votre nouvelle table, voir [Ajout d'un terme dans une table de recherche](#) à la page 174.



# Importer des données

## Import de données dans une table de recherche

Vous pouvez importer des données à partir d'un fichier dans une table de recherche que peut utiliser Advanced Transformer, Open Parser ou Table Lookup. Pour être en mesure d'importer des données à partir d'un fichier dans une table de recherche, le fichier doit respecter les exigences suivantes :

- Doit être codé UTF-8.
- Doit être un fichier délimité. Les caractères délimiteurs pris en charge sont la virgule (,), le point virgule (;), la barre verticale (|), et tab (\t).
- Les champs ayant des délimiteurs imbriqués doivent commencer et finir par des doubles guillemets, par exemple "1,a","2,b","3,c".
- Une apostrophe dans un champ commençant et finissant par des guillemets doivent avoir deux guillemets, par exemple "2"" pieds".

Pour importer des données à partir d'un fichier dans une table de recherche, procédez comme suit :

1. Dans Enterprise Designer, sélectionnez **Outils > Gestion de table**.
2. Sélectionnez la table dans laquelle vous souhaitez importer les données. Vous pouvez également créer une table. Pour obtenir les instructions permettant de créer une table, voir [Création d'une table de recherche](#) à la page 176.
3. Cliquez sur **Importer**.
4. Cliquez sur **Parcourir** et sélectionnez le fichier contenant des données que vous désirez importer.
5. Cliquez sur **Ouvrir**. Un aperçu des données dans le fichier importé s'affiche dans Aperçu du fichier.
6. Vous pouvez sélectionner des colonnes d'une table définie par l'utilisateur et les mapper dans la table existante. Par exemple, supposez qu'il y a deux colonnes dans la table définie utilisateur que vous désirez importer. Il y a la colonne 1 et la colonne 2. La liste de colonnes devrait indiquer colonne 1 et colonne 2. Vous pourriez alors sélectionner la colonne 2 pour mapper un terme de recherche et sélectionner la colonne 1 pour mapper un terme normalisé.
7. Sélectionnez **Importer uniquement les nouveaux termes** pour n'importer que les nouveaux enregistrements de la table définie utilisateur, ou **Écraser les termes existants** pour importer tous les enregistrements des colonnes sélectionnées.
8. Cliquez sur **OK**.


## Utiliser Importation avancée



La fonction Import avancé vous permet d'importer des données sélectionnée dans des tables de recherche, utilisées par Advanced Transformer, Table Lookup et Open Parser. Utilisez Import avancé pour effectuer les opérations suivantes :

- Extraire des termes d'une colonne sélectionnée dans un fichier délimité défini utilisateur.
- Extraire les termes ne contenant qu'un seul mot (jeton) d'une colonne sélectionnée dans un fichier délimité défini utilisateur. Lorsque vous extrayez des jetons, vous pouvez identifier le nombre d'occurrences d'un terme dans une colonne donnée du fichier et créer des groupements pour les termes concernés et les ajouter à la table.

Le fichier contenant les données à importer doit respecter les exigences suivantes :

- Doit être codé UTF-8.
- Doit être un fichier délimité. Les caractères délimiteurs pris en charge sont la virgule (,), le point virgule (;), la barre verticale (|), et tab (\t).
- Les champs ayant des délimiteurs imbriqués doivent commencer et finir par des doubles guillemets, par exemple "1,a","2,b","3,c".
- Une apostrophe dans un champ commençant et finissant par des guillemets doivent avoir deux guillemets, par exemple "2"" pieds".

1. Dans Enterprise Designer, sélectionnez **Outils > Gestion de table**.
2. Sélectionnez la table dans laquelle vous souhaitez importer des données.
3. Cliquez sur **Import avancé**.
4. Cliquez sur **Parcourir** et sélectionnez le fichier que vous désirez importer.
5. Cliquez sur **Ouvrir**.
6. Sélectionnez la colonne d'une table dans la liste **Colonne**. Les données de l'exemple indiquent la fréquence d'occurrence de chaque terme répertorié dans la table définie utilisateur. La fréquence n'est affichée que pour les termes ne figurant pas déjà dans la table existante.
7. Pour visualiser les termes comme les mots simples, sélectionnez **Séparer en termes ne contenant qu'un seul mot**.
8. Pour les tables Advanced Transformer et Open Parser :
  - a) Sélectionnez un terme dans la liste sur la gauche.
  - b) Cliquez sur la flèche vers la droite pour ajouter le terme à la liste de droite. Cliquez sur la flèche vers la gauche pour supprimer un terme sélectionné dans la liste de tables.
  - c) Cliquez sur **OK** pour enregistrer les modifications apportées à la table.
9. Pour les tables Table Lookup :
  - a) Cliquez sur  pour ajouter un regroupement de tables.
  - b) Cliquez sur **Nouveau**.
  - c) Tapez un nouveau terme, puis cliquez sur **Ajouter**. Continuez d'ajouter des termes jusqu'à ce que vous ayez terminé, puis cliquez sur **Fermer**.

- d) Sélectionnez un terme dans la liste, puis cliquez sur Ajouter. Continuez d'ajouter des termes jusqu'à ce que vous ayez terminé, puis cliquez sur **Fermer**. Les nouveaux termes sont ajoutés à la liste de termes sur la droite.
- e) Sélectionnez un terme à gauche puis cliquez sur la flèche vers la droite pour ajouter le terme au groupement sélectionné. Cliquez sur la flèche vers la gauche pour supprimer un terme d'un des groupements.
- f) Pour modifier un terme, sélectionnez-le dans la liste de droite puis cliquez sur .
- g) Pour supprimer un terme, sélectionnez-le dans la liste de droite puis cliquez sur .
- h) Cliquez sur **OK** pour enregistrer les modifications apportées à la table.

# 8 - Référence aux stages

## In this section

---

Module Advanced Matching	181
Module Business Steward	250
Module Data Normalization	297
Module Universal Name	317

# Module Advanced Matching

## Module Advanced Matching

Le module Advanced Matching rapproche des enregistrements entre et/ou à l'intérieur de n'importe quel nombre de fichier d'entrée. Vous pouvez également utiliser le Module Advanced Matching pour rapprocher divers champs, dont nom, adresse, nom et adresse, ou des champs n'ayant pas trait aux noms ou aux adresses, comme le numéro de sécurité sociale ou la date de naissance.

### Composants

Le module Advanced Matching est constitué des éléments suivants :

- **Best-of-Breed** : ce stage sélectionne un enregistrement de type best-of-breed à partir du cluster de doublons en sélectionnant un enregistrement modèle, puis en l'utilisant pour générer le meilleur enregistrement composite. Ensuite, cet enregistrement est renvoyé en tant qu'enregistrement rémanent.
- **Candidate Finder** : ce stage obtient les enregistrements candidats qui formeront l'ensemble des rapprochements potentiels qui seront ensuite évalués par le stage Transactional Match. Candidate Finder s'utilise conjointement à Transactional Match.
- **Duplicate Synchronization** : ce stage vous permet d'indiquer quels champs d'un groupe d'enregistrements seront copiés (postés) dans les champs correspondants de tous les enregistrements du groupe.
- **Filter** : ce stage vous permet d'indiquer les critères auxquels les enregistrements doivent satisfaire afin d'être conservés ou supprimés d'un groupe de doublons, soit en vue pour un traitement ultérieur, soit pour être écrits vers un fichier de sortie.
- **Interflow Match** : ce stage identifie les rapprochements entre des enregistrements de données similaires en croisant plusieurs flux d'entrée.
- **Intraflow Match** : ce stage identifie des rapprochements entre des enregistrements de données similaires à l'intérieur d'un seul flux d'entrée.
- **Match Analysis** : utilisez Match Analysis pour analyser et comparer les résultats des stages de rapprochement afin d'obtenir une meilleure compréhension du résultat de la procédure de rapprochement et de vous assister dans la tâche consistant à améliorer les résultats de vos rapprochements.
- **Match Key Generator** : ce stage crée une clé non-unique partagée par tous les enregistrements. Les enregistrements qui partagent la même clé peuvent ensuite être groupés ensemble pour comparaison.
- **Private Match** : ce stage permet à deux entités de comparer des jeux de données et d'identifier des enregistrements, ou les clients, communs sans compromettre les informations sensibles. Grâce au recours à une fonction de chiffrement, la sécurité est conservée lors de l'exécution d'une

fonction de correspondance, puis une fonction de déchiffrement affiche la sortie des données en correspondance.

- **Transactional Match** : ce stage rapproche les transactions suspectes avec une base de données à l'aide du stage Candidate Finder pour interroger et renvoyer des enregistrements candidats potentiels.
- **Write to Search Index** : ce stage vous permet de créer un index plein texte basé sur les données arrivant dans le stage. Disposer de ces données dans un index de recherche dédié permet d'obtenir un temps de réponse plus rapide lorsque vous effectuez des recherches par rapport à l'index de Candidate Finder.

## Best of Breed

Best of Breed consolide les enregistrements doublons en sélectionnant les meilleures données parmi un groupe d'enregistrements doublons et en créant un nouvel enregistrement consolidé à l'aide des meilleures données. Ce « super » enregistrement s'appelle l'enregistrement Best of Breed. Vous définissez les règles à utiliser lors de la sélection des enregistrements à traiter. Une fois le traitement terminé, l'enregistrement Best of Breed est conservé par le système.

### Options

Le tableau suivant répertorie les options de Best Of Breed.

Nom de l'option	Description/Valeurs valides
Grouper par	Spécifie le champ à utiliser pour créer des groupes d'enregistrements à fusionner dans un enregistrement Best Of Breed unique, créant ainsi un enregistrement Best of Breed à partir de chaque groupe. Si vous avez utilisé un stage de correspondance précédemment dans le dataflow, vous devez sélectionner le champ CollectionNumber pour utiliser les collections créées par le stage de correspondance en tant que groupes. Cependant, si vous souhaitez regrouper les enregistrements par un autre champ, choisissez ce dernier ici. Par exemple, si vous souhaitez fusionner les enregistrements dotés de la même valeur dans le champ AccountNumber en un enregistrement Best Of Breed, vous devez sélectionner AccountNumber.
Trier	Si vous spécifiez un champ dans le champ <b>Grouper par</b> , cochez cette case pour trier les enregistrements par la valeur contenu dans le champ de votre choix. Cette option est activée par défaut.

Nom de l'option	Description/Valeurs valides
Avancé	<p>Cliquez sur ce bouton pour indiquer les options de performances du tri. Par défaut, les options de performances du tri spécifiées dans Management Console, qui représentent les options de performances par défaut pour votre système, sont appliquées. Si vous souhaitez remplacer les options de performances par défaut de votre système, cochez la case <b>Écraser les options de performance du tri</b>, puis spécifiez les valeurs que vous souhaitez obtenir dans ces champs :</p>
<b>Limite du nombre d'enregistrement en mémoire</b>	<p>Spécifie le nombre maximum de rangées de données qu'un trieur peut contenir en mémoire avant que celui-ci commence à pager sur le disque. Par défaut, un tri de 10 000 enregistrements ou moins sera effectué en mémoire et un tri de plus de 10 000 enregistrements sera effectué sur le disque. La limite maximale est de 100 000 enregistrements. En général, un tri en mémoire est beaucoup plus rapide qu'un tri sur le disque ; donc, il faut définir une valeur assez haute pour que la plupart des tris s'effectuent en mémoire et que seuls les groupes de grande taille soient écrits sur le disque.</p> <p><b>Remarque :</b> Soyez conscient du fait qu'au sein d'environnements où des jobs s'exécutent de manière simultanée, une augmentation du paramètre <b>Dans la limite d'enregistrement mémoire</b> augmente la probabilité de ne plus disposer de suffisamment de mémoire.</p>
<b>Nombre maximal de fichiers temporaires</b>	<p>Spécifie le nombre maximal de fichiers temporaires pouvant être employés par un processus de tri. L'utilisation d'un plus grand nombre de fichiers temporaires peut améliorer les performances. Cependant, le nombre optimal dépend très largement de la configuration du serveur qui exécute Spectrum™ Technology Platform. Nous vous conseillons d'essayer différents paramètres et d'observer l'effet de l'utilisation d'un plus ou moins grand nombre de fichiers temporaires sur les performances. Pour calculer le nombre approximatif de fichiers temporaires nécessaires, utilisez l'équation suivante :</p> $\frac{(NumberOfRecords \times 2)}{InMemoryRecordLimit} = NumberOfTempFiles$ <p>Notez que le nombre maximal de fichiers temporaires ne peut pas être supérieure à 1 000.</p>
<b>Activer la compression</b>	<p>Spécifie si les fichiers temporaires sont compressés lors de leur écriture sur le disque.</p> <p><b>Remarque :</b> Les paramètres de performances de tri optimal dépendent de la configuration matérielle de votre serveur. Néanmoins, l'équation suivante produit généralement de bonnes performances de tri :</p> $(InMemoryRecordLimit \times MaxNumberOfTempFiles \div 2) \geq TotalNumberOfRecords$

Nom de l'option	Description/Valeurs valides
Conserver les enregistrements originaux	Sélectionnez cette option pour conserver tous les enregistrements dans la collection avec l'enregistrement Best Of Breed. Désélectionnez l'option si vous voulez que seul l'enregistrement Best of Breed soit enregistré.
Utiliser le premier enregistrement	Sélectionnez cette option pour que Best of Breed sélectionne automatiquement le premier enregistrement de la collection comme enregistrement modèle. L'enregistrement modèle est l'enregistrement sur lequel se base l'enregistrement Best Of Breed.
Définissez l'enregistrement modèle	Sélectionnez cette option pour définir les règles de sélection de l'enregistrement modèle. Pour plus d'informations, reportez-vous à la section <a href="#">Définition des règles d'enregistrement modèle</a> à la page 184.

### **Définition des règles d'enregistrement modèle**

Dans le traitement Best of Breed, l'enregistrement modèle est l'enregistrement d'un groupe permettant de créer l'enregistrement Best Of Breed. L'enregistrement modèle sert de point de départ pour la construction de l'enregistrement Best Of Breed et est modifié en fonction des paramètres Best of Breed que vous définissez. Le stage Best of Breed peut sélectionner l'enregistrement modèle automatiquement, ou vous pouvez définir des règles de sélection de l'enregistrement modèle. Cette rubrique décrit la procédure de définition des règles afin de sélectionner l'enregistrement modèle.

Les règles concernant les modèles s'écrivent en spécifiant le nom de champ, un opérateur, un type de valeur, et une valeur. Voici un exemple d'options d'enregistrement modèle :

Nom du champ : MatchScore Type de champ : Numérique Opérateur : Est égal Type de valeur : Chaîne Valeur : 100

Cette règle de modèle sélectionne l'enregistrement de la collection où le Score de rapprochement est égal à la valeur 100.

La procédure suivante décrit la manière dont définir une règle d'enregistrement modèle dans le stage Best of Breed.

1. Dans le stage Best of Breed, sous **Paramètres de l'enregistrement modèle**, sélectionnez l'option **Définir l'enregistrement modèle**.
2. Dans l'arborescence, cliquez sur **Règles**.
3. Cliquez sur **Ajouter une règle**.
4. Renseignez les champs suivants.



Option	Description
Nom du champ	Indique le nom du champ de dataflow dont vous souhaitez évaluer la valeur afin de déterminer si l'enregistrement doit être l'enregistrement modèle.
Type de champ	Indique le type de données dans le champ. L'un des éléments suivants : <b>Non-numérique</b> Choisissez cette option si le champ contient des données non numériques (par exemple, des données de chaîne). <b>numériques</b> Choisissez cette option si le champ contient des données numériques (par exemple, double, flottantes, etc.).

Option	Description
--------	-------------

---

Opérateur	
-----------	--

Option	Description
	Indique le type de comparaison à utiliser pour évaluer le champ. L'un des éléments suivants :
<b>Contient</b>	Détermine si le champ contient la valeur indiquée. Par exemple, « bateau à voile » contient la valeur « bateau ».
<b>Est égal à</b>	Détermine si le champ contient la valeur exacte indiquée.
<b>Est supérieur à</b>	Détermine si la valeur du champ est supérieure à la valeur spécifiée. Cette opération ne fonctionne que sur les champs numériques.
<b>Supérieur ou Egal à</b>	Détermine si la valeur du champ est supérieure ou égale à la valeur spécifiée. Cette opération ne fonctionne que sur les champs numériques.
<b>Le plus haut</b>	Compare la valeur du champ pour tout le groupe d'enregistrements et identifie l'enregistrement qui possède la valeur la plus élevée dans le champ. Par exemple, si les champs dans ce groupe contiennent les valeurs 10, 20, 30, et 100, l'enregistrement qui dispose de la valeur 100 est sélectionné. Cette opération ne fonctionne que sur les champs numériques. Si plusieurs enregistrements sont à égalité pour la valeur la plus longue, un enregistrement est sélectionné.
<b>Est vide</b>	Détermine si le champ ne contient aucune valeur.
<b>N'est pas vide</b>	Détermine si le champ contient une valeur.
<b>Inférieur à</b>	Détermine si la valeur du champ est inférieure à la valeur spécifiée. Cette opération ne fonctionne que sur les champs numériques.
<b>Inférieur ou Egal à</b>	Détermine si la valeur du champ est inférieure ou égale à la valeur spécifiée. Cette opération ne fonctionne que sur les champs numériques.
<b>Le plus long</b>	Compare la valeur du champ pour tout le groupe d'enregistrements et identifie l'enregistrement qui possède la valeur la plus longue (en octets) dans le champ. Par exemple, si le groupe contient les valeurs « Mike » et « Michael », l'enregistrement possédant la valeur « Michael » est choisi. Si plusieurs enregistrements sont à égalité pour la valeur la plus longue, un enregistrement est sélectionné.
<b>Le plus bas</b>	Compare la valeur du champ pour tout le groupe d'enregistrements et identifie l'enregistrement qui possède la valeur la plus basse dans le champ. Par exemple, si les champs dans ce groupe contiennent les valeurs 10, 20, 30, et 100, l'enregistrement qui dispose de la valeur 10 est sélectionné. Cette opération ne fonctionne que sur les champs numériques. Si plusieurs enregistrements sont à égalité pour la valeur la plus longue, un enregistrement est sélectionné.
<b>Le plus commun</b>	Détermine si la valeur de champ contient la valeur qui se produit le plus fréquemment dans ce champ parmi les enregistrements dans ce groupe. Si deux valeurs, ou plus, sont les plus courantes, aucune action ne sera réalisée.

Option	Description
	<p><b>N'est pas égal</b> Détermine si la valeur du champ n'est pas égale à la valeur spécifiée.</p>
Type de valeur	<p>Indique le type de valeur à comparer avec la valeur du champ. L'un des éléments suivants :</p> <p><b>Remarque</b> : Cette option n'est pas disponible si vous sélectionnez l'opérateur Highest, Lowest ou Longest.</p> <p><b>Champ</b> Choisissez cette option si vous souhaitez comparer une autre valeur du champ de dataflow avec le champ.</p> <p><b>Chaîne</b> Choisissez cette option si vous souhaitez comparer le champ avec une valeur spécifique.</p>
Valeur	<p>Indique la valeur à comparer avec la valeur du champ. Si vous avez sélectionné Champ dans le champ <b>Type de champ</b>, sélectionnez un champ de dataflow. Si vous avez sélectionné Chaîne dans le champ <b>Type de valeur</b>, tapez la valeur à utiliser dans la comparaison.</p> <p><b>Remarque</b> : Cette option n'est pas disponible si vous sélectionnez l'opérateur Highest, Lowest ou Longest.</p>

5. Cliquez sur **OK**.
6. Si vous souhaitez spécifier d'autres règles, cliquez sur **Ajouter une règle**.

Si vous ajoutez d'autres règles, vous devrez sélectionner un opérateur logique à utiliser entre chaque règle. Choisissez **Et** si vous souhaitez que la nouvelle règle et la règle précédente soient toutes les deux respectées pour que l'enregistrement soit sélectionné comme enregistrement modèle. Sélectionnez **Ou** si vous souhaitez que la règle précédente ou la nouvelle règle soit respectée pour que l'enregistrement soit sélectionné comme enregistrement modèle.

Vous avez désormais configuré des règles à utiliser pour sélectionner l'enregistrement modèle. Configurez les paramètres Best Of Breed pour terminer la configuration du stage Best of Breed.

### **Définitions des règles et actions de type Best of Breed**

Les règles et actions de type Best of Breed permettent ensemble de déterminer les champs d'enregistrements doublons dans un groupe à copier dans l'enregistrement Best of Breed (le meilleur enregistrement). Les règles testent les valeurs d'un enregistrement et, si l'enregistrement satisfait les règles, les données sont copiées d'un enregistrement à l'enregistrement modèle. Les actions définissent les données à copier, ainsi que le champ de l'enregistrement modèle devant recevoir les données. Une fois les règles et les actions exécutées, l'enregistrement modèle sera l'enregistrement Best Of Breed.

Les règles et actions peuvent être regroupées en conditions et vous pouvez disposer de plusieurs conditions. Cela vous permet

1. Dans le stage Best of Breed, sous **Paramètres Best Of Breed**, cliquez sur le noeud **Règles** dans l'arborescence.
2. Cliquez sur **Ajouter une règle**.
3. Renseignez les champs suivants :

Option	Description
Nom du champ	Indique le nom du champ de flux de données dont vous souhaitez évaluer la valeur afin de déterminer si la condition est remplie, ainsi que les actions associées à effectuer.
Type de champ	Indique le type de données dans le champ. L'un des éléments suivants : <ul style="list-style-type: none"> <li><b>Non-numérique</b>      Choisissez cette option si le champ contient des données non numériques (par exemple, des données de chaîne).</li> <li><b>numériques</b>        Choisissez cette option si le champ contient des données numériques (par exemple, double, flottantes, etc.).</li> </ul>

Option	Description
--------	-------------

---

Opérateur	
-----------	--

Option	Description
	Indique le type de comparaison à utiliser pour évaluer le champ. L'un des éléments suivants :
<b>Contient</b>	Détermine si le champ contient la valeur indiquée. Par exemple, « bateau à voile » contient la valeur « bateau ».
<b>Est égal à</b>	Détermine si le champ contient la valeur exacte indiquée.
<b>Est supérieur à</b>	Détermine si la valeur du champ est supérieure à la valeur spécifiée. Cette opération ne fonctionne que sur les champs numériques.
<b>Supérieur ou Egal à</b>	Détermine si la valeur du champ est supérieure ou égale à la valeur spécifiée. Cette opération ne fonctionne que sur les champs numériques.
<b>Le plus haut</b>	Compare la valeur du champ pour tout le groupe d'enregistrements et identifie l'enregistrement qui possède la valeur la plus élevée dans le champ. Par exemple, si les champs dans ce groupe contiennent les valeurs 10, 20, 30, et 100, l'enregistrement qui dispose de la valeur 100 est sélectionné. Cette opération ne fonctionne que sur les champs numériques. Si plusieurs enregistrements sont à égalité pour la valeur la plus longue, un enregistrement est sélectionné.
<b>Est vide</b>	Détermine si le champ ne contient aucune valeur.
<b>N'est pas vide</b>	Détermine si le champ contient une valeur.
<b>Inférieur à</b>	Détermine si la valeur du champ est inférieure à la valeur spécifiée. Cette opération ne fonctionne que sur les champs numériques.
<b>Inférieur ou Egal à</b>	Détermine si la valeur du champ est inférieure ou égale à la valeur spécifiée. Cette opération ne fonctionne que sur les champs numériques.
<b>Le plus long</b>	Compare la valeur du champ pour tout le groupe d'enregistrements et identifie l'enregistrement qui possède la valeur la plus longue (en octets) dans le champ. Par exemple, si le groupe contient les valeurs « Mike » et « Michael », l'enregistrement possédant la valeur « Michael » est choisi. Si plusieurs enregistrements sont à égalité pour la valeur la plus longue, un enregistrement est sélectionné.
<b>Le plus bas</b>	Compare la valeur du champ pour tout le groupe d'enregistrements et identifie l'enregistrement qui possède la valeur la plus basse dans le champ. Par exemple, si les champs dans ce groupe contiennent les valeurs 10, 20, 30, et 100, l'enregistrement qui dispose de la valeur 10 est sélectionné. Cette opération ne fonctionne que sur les champs numériques. Si plusieurs enregistrements sont à égalité pour la valeur la plus longue, un enregistrement est sélectionné.
<b>Le plus commun</b>	Détermine si la valeur de champ contient la valeur qui se produit le plus fréquemment dans ce champ parmi les enregistrements dans ce groupe. Si deux valeurs, ou plus, sont les plus courantes, aucune action ne sera réalisée.

Option	Description
	<p><b>N'est pas égal</b> Détermine si la valeur du champ n'est pas égale à la valeur spécifiée.</p>
Type de valeur	<p>Indique le type de valeur à comparer avec la valeur du champ. L'un des éléments suivants :</p> <p><b>Remarque</b> : Cette option n'est pas disponible si vous sélectionnez l'opérateur Highest, Lowest ou Longest.</p> <p><b>Champ</b> Choisissez cette option si vous souhaitez comparer une autre valeur du champ de dataflow avec le champ.</p> <p><b>Chaîne</b> Choisissez cette option si vous souhaitez comparer le champ avec une valeur spécifique.</p>
Valeur	<p>Indique la valeur à comparer avec la valeur du champ. Si vous avez sélectionné Champ dans le champ <b>Type de champ</b>, sélectionnez un champ de dataflow. Si vous avez sélectionné Chaîne dans le champ <b>Type de valeur</b>, tapez la valeur à utiliser dans la comparaison.</p> <p><b>Remarque</b> : Cette option n'est pas disponible si vous sélectionnez l'opérateur Highest, Lowest ou Longest.</p>

4. Cliquez sur **OK**.
5. Si vous souhaitez spécifier d'autres règles pour la condition, cliquez sur **Ajouter une règle**.  
Si vous ajoutez d'autres règles, vous devrez sélectionner un opérateur logique à utiliser entre chaque règle. Choisissez **Et** si vous souhaitez que la nouvelle règle et la règle précédente soient toutes les deux respectées pour que la condition soit remplie et les actions associées effectuées. Sélectionnez **Ou** si vous souhaitez que la règle précédente ou la nouvelle règle soit respectée pour que la condition soit remplie.
6. Cliquez sur le nœud **Actions** dans l'arborescence.
7. Cliquez sur **Ajouter une action**.
8. Renseignez les champs suivants.



Option	Description
Type de la source	<p>Indique le type de données à copier dans l'enregistrement Best Of Breed. Un des éléments suivants.</p> <p><b>Champ</b> Choisissez cette option pour copier une valeur d'un champ dans l'enregistrement Best Of Breed.</p> <p><b>Chaîne</b> Choisissez cette option pour copier une valeur constante dans l'enregistrement Best Of Breed.</p>
Données sources	Indique les données à copier dans l'enregistrement Best Of Breed. Si le type de source est Field, sélectionnez le champ dont vous souhaitez copier la valeur dans le champ de destination. Si le type de source est String, indiquez une valeur constante à copier dans le champ de destination.
Cible	Indique le champ dans l'enregistrement Best Of Breed dans lequel vous souhaitez copier les données spécifiées dans le champ <b>Données source</b> .
Accumuler les données sources	<p>Si les données dans le champ <b>Données source</b> sont des données numériques, vous pouvez activer cette option afin de combiner les données source pour tous les enregistrements doublons et insérer la valeur totale dans l'enregistrement Best Of Breed.</p> <p>Par exemple, s'il existe trois enregistrements doublons dans le groupe et que ceux-ci contenaient ces valeurs dans le champ Deposits :</p> <p>100.00 20.00 5.00</p> <p>Les trois valeurs serait combinées et la valeur totale (125,00) serait insérée dans le champ Deposits de l'enregistrement Best Of Breed.</p>

9. Cliquez sur **OK**.
10. Si vous souhaitez indiquer d'autres actions à réaliser pour cette condition, cliquez sur **Ajouter une action** et répétez les étapes ci-dessus.
11. Pour ajoutez une autre condition, cliquez sur la condition racine dans l'arborescence, puis cliquez sur **Ajouter une condition**.

#### Exemple de règle et d'action Best of Breed

Cette règle Best of Breed sélectionne l'enregistrement où le Score de rapprochement est égal à la valeur 100. Les données de numéro de compte qui correspondent au champ sélectionné sont ensuite copiées dans le champ AccountNumber sur l'enregistrement Best of Breed.

**Règle**

Nom de champ : MatchScore

Type de champ : Numeric

Opérateur : Equal

Type de valeur : String

Valeur : 100

**Action**

Type de source : Field

Données source : AccountNumber

Destination : AccountNumber

**Sortie****Tableau 8 : Sortie Best of Breed**

Nom du champ	Format	Description/Valeurs valides						
CollectionRecordType	Chaîne	Identifie les enregistrements modèle et Best of Breed dans un groupe de doublons. Les valeurs possibles sont : <table border="0" style="margin-left: 20px;"> <tr> <td><b>Primary</b></td> <td>L'enregistrement est l'enregistrement modèle sélectionné dans un groupe.</td> </tr> <tr> <td><b>Secondary</b></td> <td>L'enregistrement n'est pas l'enregistrement modèle sélectionné dans un groupe.</td> </tr> <tr> <td><b>BestOfBreed</b></td> <td>L'enregistrement est l'enregistrement Best of Breed nouvellement créé dans le groupe.</td> </tr> </table>	<b>Primary</b>	L'enregistrement est l'enregistrement modèle sélectionné dans un groupe.	<b>Secondary</b>	L'enregistrement n'est pas l'enregistrement modèle sélectionné dans un groupe.	<b>BestOfBreed</b>	L'enregistrement est l'enregistrement Best of Breed nouvellement créé dans le groupe.
<b>Primary</b>	L'enregistrement est l'enregistrement modèle sélectionné dans un groupe.							
<b>Secondary</b>	L'enregistrement n'est pas l'enregistrement modèle sélectionné dans un groupe.							
<b>BestOfBreed</b>	L'enregistrement est l'enregistrement Best of Breed nouvellement créé dans le groupe.							

**Candidate Finder**

Candidate Finder obtient les enregistrements candidats qui formeront l'ensemble des rapprochements potentiels. Les recherches de base de données fonctionnent conjointement à Transactional Match, tandis que les recherches d'index de recherche fonctionnent indépendamment de Transactional Match. En fonction du format de vos données, Candidate Finder peut aussi avoir besoin de parser le nom ou l'adresse de l'enregistrement suspect, des enregistrements candidats, ou des deux.

Candidate Finder active également les index de recherche plein texte et permet de définir des critères de recherche simples et complexes par rapport à des caractères et du texte à l'aide de différents types de recherche (Tout mot commence par, Contient, Contient tout, Contient au moins un, Ne contient aucun, Correspondance approximative, Modèle, Proximité, Plage, Caractère générique) et de différentes conditions (Toutes vraies, Au moins une, Aucun vrai).

## Options de base de données

La boîte de dialogue de Candidate Finder vous permet de définir les instructions SQL qui extraient les candidats de rapprochement potentiels d'une base de données, ainsi que de mapper les colonnes que vous sélectionnez dans la base de données sur les noms de champs qui sont définis dans votre flux de données.

**Tableau 9 : Options de base de données de Candidate Finder**

Nom de l'option	Description/Valeurs valides
Type de Finder	Sélectionner la base de données.
Connexion	Sélectionnez la base de données qui contient les enregistrements candidats. Vous pouvez sélectionner n'importe quelle connexion configurée dans Management Console. Pour vous connecter à une base de données non répertoriée, configurez une connexion à cette base de données dans Management Console, puis fermez et rouvrez Candidate Finder pour actualiser la liste des connexions.  <b>Remarque :</b> La fonction Options de flux de données dans Enterprise Designer permet d'afficher le nom de la connexion pour la configuration au moment de l'exécution.
Instruction SQL	Tapez une instruction SQL dans la zone de texte comme décrit dans <a href="#">Définition de la requête SQL</a> à la page 195
Onglet Mapping	Choisissez les paramètres de mappage de champ comme décrit dans <a href="#">Mappage des colonnes de base de données sur les Champs de stage</a> à la page 196.
Onglet Aperçu	Cliquez sur cet onglet pour entrer un exemple de match key pour tester votre instruction SQL SELECT ou votre requête d'index.

### Définition de la requête SQL

Vous pouvez taper n'importe quelle instruction SQL SELECT valide dans la zone de texte de la boîte de dialogue **Options de Candidate Finder**.

**Remarque :** `Select *` n'est pas valide.

Par exemple, imaginez que vous avez une table dans votre base de données nommée Customer\_Table qui possède les colonnes suivantes :

- Customer\_Table
- Cust\_Name
- Cust\_Address
- Cust\_City
- Cust\_State
- Cust\_Zip

Pour extraire toutes les rangées de la base de données, vous pourriez exécuter une requête ressemblant à ceci :

```
SELECT Cust_Name, Cust_Address, Cust_City, Cust_State, Cust_Zip from
Customer_Table;
```

Il est rare que vous ayez besoin de faire correspondre votre transaction et toutes les rangées de la base de données. Pour ne renvoyer que les enregistrements candidats pertinents, ajouter une clause `WHERE` avec substitution de variable. La substitution de variable fait référence à une notation spéciale que vous utiliserez pour faire en sorte que le moteur de sélection de candidats (Candidate Selection) remplace la variable par les données réelles de votre enregistrement suspect.

Pour utiliser la substitution de variable, encadrez le nom de champ entre des accolades précédées du signe dollar comme ceci : `${FieldName}`. Par exemple, la requête suivante ne renverra que les enregistrements qui ont une valeur dans `Cust_Zip` qui correspond à la valeur dans `PostalCode` sur l'enregistrement suspect.

```
SELECT Cust_Name, Cust_Address, Cust_City, Cust_State, Cust_Zip
FROM Customer_Table
WHERE Cust_Zip = ${PostalCode};
```

Pour SQL 2000, le type de données doit être identique au type de données pour Candidate Finder. Le pilote JDBC définit la variable d'entrée Candidate Finder (par ex. : `${MatchKey}`) qui est utilisée dans la clause `WHERE` sur un type de données `nVarChar(4000)`. Si les données dans la base de données sont définies sur un type de données `VarChar`, SQL Server ignorera l'index sur la base de données. Si l'index est ignoré, la performance en souffrira. Par conséquent, utilisez la requête suivante pour SQL 2000 :

```
SELECT Cust_Name, Cust_Address, Cust_City, Cust_State, Cust_Zip
FROM Customer_Table
WHERE Cust_Zip = CAST(${PostalCode} AS VARCHAR(255));
```

### ***Mappage des colonnes de base de données sur les Champs de stage***

Si les noms de colonnes dans votre base de données correspondent aux noms de Champ de composant de manière exacte, ceux-ci sont automatiquement mappés sur les Champs de stage correspondants. Si ces noms ne sont pas exactement identiques, vous devrez utiliser les Champs sélectionnés (colonnes de la base de données) à mapper sur les Champs de stage (noms de champ définis dans le flux de données).

Prenons, par exemple, une table nommée `Customer_Table` contenant les colonnes suivantes :

- Cust\_Name
- Cust\_Address
- Cust\_City
- Cust\_State
- Cust\_Zip

Quand vous extrayez ces enregistrements de la base de données, vous devez mapper les noms des colonnes aux noms des champs qui sont employés par Transactional Match et d'autres composants de votre flux de données. Par exemple, Cust\_Address pourrait être mappé sur AddressLine1, et Cust\_Zip serait mappé sur PostalCode.

1. Sélectionnez la liste déroulante sous **Champs sélectionnés** dans la boîte de dialogue **Options de Candidate Finder**. Sélectionnez ensuite la colonne de base de données Cust\_Zip.
2. Sélectionnez la liste déroulante sous **Champs de stage**. Puis, sélectionnez le champ sur lequel vous voulez mapper.

Par exemple, si vous voulez mapper Cust\_Zip sur Postal Code, sélectionnez d'abord Cust\_Zip sous Champs sélectionnés, puis sélectionnez PostalCode sur la ligne Champ de stage correspondante.

### Méthode alternative pour le mappage des champs

Vous pouvez utiliser une notation spéciale dans votre requête SQL pour réaliser le mappage. Pour ce faire, encadrez le nom de champ sur lequel vous voulez mapper entre des accolades après le nom de colonne dans votre requête. Lorsque vous effectuez cela, les champs sélectionnés sont automatiquement mappés sur les champs de stages correspondants.

Par exemple,

```
select Cust_Name {Name}, Cust_Address {AddressLine1},
       Cust_City {City}, Cust_State {StateProvince},
       Cust_Zip {PostalCode}
from Customer
where Cust_Zip = ${PostalCode};
```

### Configuration du nom de connexion lors de l'exécution

Il est possible de configurer le nom de connexion et de le transmettre lors de l'exécution s'il est exposé sous forme d'option de flux de données. Cela vous permet d'exécuter votre flux de données tout en utilisant un nom de connexion différent.

1. Dans Enterprise Designer, ouvrez un flux de données utilisant le stage Candidate Finder.
2. Enregistrez et exposez ce flux de données.
3. Accédez à **Modifier > Options de dataflow**.
4. Dans la table **Map dataflow options to stages**, développez Candidate Finder et modifiez les options, le cas échéant. Cochez la case de l'option que vous souhaitez modifier, puis modifiez la valeur dans la liste déroulante **Valeur par défaut**.
5. Facultatif : modifiez le nom des options du champ **Option label**.
6. Cliquez deux fois sur **OK**.

## Options d'index de recherche

La boîte de dialogue Candidate Finder vous permet d'effectuer une recherche simple qui correspond aux valeurs de champ d'entrée dans l'index de recherche ou une recherche avancée pour créer des règles de correspondance qui extraient les candidats correspondants potentiels d'index de recherche.

### Options d'index de recherche simple

**Tableau 10 : Options de Candidate Finder**

Nom de l'option	Description/Valeurs valides
Type de Finder	Sélectionnez l'index de recherche.
Votre nom	Sélectionnez l'index approprié ayant été créé via le stage <a href="#">Write to Search Index</a> à la page 245 sous les stages déployés Advanced Matching dans Enterprise Designer.
Enregistrement de départ	Saisissez le numéro d'enregistrement à partir duquel doivent commencer les résultats de recherche. La valeur par défaut est 1.
Nombre maximal de résultats	Saisissez le nombre maximal de résultats que l'index de recherche doit renvoyer. La valeur par défaut est 10. <b>Remarque</b> : Si le nombre maximal de résultats est arbitrairement grand, traitez-les par lots, en utilisant le champ <b>Extraire la taille du lot</b> .
Extraire la taille du lot	Si le <b>Nombre maximal de résultats</b> est arbitrairement grand, saisissez la taille des lots dans lesquels vous voulez que les résultats soient traités. Cela optimise le traitement d'un grand nombre d'enregistrements. La valeur par défaut est 10 000. La valeur <b>Extraire la taille du lot</b> recommandée est une valeur inférieure au <b>Nombre maximal de résultats</b> et si la valeur <b>Extraire la taille du lot</b> est supérieure au <b>Nombre maximal de résultats</b> , les enregistrements sont traités dans un seul lot. <b>Remarque</b> : Ce champ s'applique uniquement au moteur de recherche pris en charge par le cluster et non au moteur de recherche patrimonial.

Nom de l'option	Description/Valeurs valides
Renvoyer le nombre total de correspondances	Renvoie le nombre total de correspondances effectuées. Par exemple, si vous utilisez la valeur par défaut « 10 » pour le champ Nombre maximal de résultats ci-dessus, seuls 10 résultats seront renvoyés. En revanche, si vous cochez cette case, le champ de sortie TotalMatchCount vous indique le nombre de correspondances effectuées lors du traitement.
Type de recherche d'index	Détermine le type de recherche d'index que vous souhaitez effectuer. Sélectionnez <b>recherche Simple</b> .
Champs d'index	Sélectionnez les champs d'index que vous souhaitez utiliser pour la comparaison de la recherche simple.
Champ de saisie	Sélectionnez le champ d'entrée que vous souhaitez utiliser pour la comparaison de la recherche simple.

Nom de l'option	Description/Valeurs valides
-----------------	-----------------------------

---

Analyseur d'entrée :



## Nom de l'option

## Description/Valeurs valides

Indiquez quel l'analyseur à utiliser pour segmenter la chaîne d'entrée. L'une de ces valeurs :

- Standard : fournit un segmenteur basé sur la grammaire qui contient un super ensemble d'analyseurs d'espace et de mot d'arrêt. Comprend la ponctuation anglaise pour segmenter les mots, connaît les mots à ignorer (via l'analyseur de mot d'arrêt), et effectue une recherche techniquement insensible à la casse en procédant à des comparaisons de minuscules. Par exemple, la chaîne « Pitney Bowes Software » serait renvoyée sous forme de trois jetons : « Pitney », « Bowes » et « Software ».
- Espace : sépare les jetons par des espaces. Sorte de sous-ensemble de l'analyseur standard dans la mesure où il comprend les segmentations de mots en anglais, en fonction des espaces et des retours à la ligne.
- Mot d'arrêt : supprime les articles, tels que "the," "and" et "a" pour réduire la taille de index et augmenter les performances.
- Mot clé : crée un jeton unique pour un dataflow. Par exemple, la chaîne « Pitney Bowes Software » serait renvoyée sous forme d'un jeton : « Pitney Bowes Software ».
- Russe : prend en charge les index et les services de saisie anticipée en langue russe. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « et », « je » et « vous » pour réduire la taille de l'index et améliorer les performances.
- Allemand : prend en charge les index et les services de saisie anticipée en langue allemande. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « le », « et » et « un » pour réduire la taille de l'index et améliorer les performances.
- Danois : prend en charge les index et les services de saisie anticipée en langue danoise. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « à », « et » et « un » pour réduire la taille de l'index et améliorer les performances.
- Néerlandais : prend en charge les index et les services de saisie anticipée en langue néerlandaise. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « le », « et » et « un » pour réduire la taille de l'index et améliorer les performances.
- Finnois : prend en charge les index et les services de saisie anticipée en langue finnoise. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « est », « et » et « de » pour réduire la taille de l'index et améliorer les performances.
- Français : prend en charge les index et les services de saisie anticipée en langue française. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « le », « et » et « un » pour réduire la taille de l'index et améliorer les performances.
- Hongrois : prend en charge les index et les services de saisie anticipée en langue hongroise. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « le », « et » et « un » pour réduire la taille de l'index et améliorer les performances.
- Italien : prend en charge les index et les services de saisie anticipée en langue italienne. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « le », « et » et « un » pour réduire la taille de l'index et améliorer les performances.

Nom de l'option	Description/Valeurs valides
	<p>les performances.</p> <ul style="list-style-type: none"> <li>• Norvégien : prend en charge les index et les services de saisie anticipée en langue norvégienne. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « le », « et » et « un » pour réduire la taille de l'index et améliorer les performances.</li> <li>• Portugais : prend en charge les index et les services de saisie anticipée en langue portugaise. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « le », « et » et « un » pour réduire la taille de l'index et améliorer les performances.</li> <li>• Espagnol : prend en charge les index et les services de saisie anticipée en langue espagnole. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « le », « et » et « un » pour réduire la taille de l'index et améliorer les performances.</li> <li>• Suédois : prend en charge les index et les services de saisie anticipée en langue suédoise. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « le », « et » et « un » pour réduire la taille de l'index et améliorer les performances.</li> <li>• Hindi : prend en charge les index et les services de saisie anticipée en langue hindi. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « par », « et » et « un » pour réduire la taille de l'index et améliorer les performances.</li> </ul>
Onglet Champs de sortie	<p>Cochez la case <b>Inclure</b> pour sélectionner les champs à inclure dans la sortie.</p> <p><b>Remarque</b> : Si le champ d'entrée provient d'un stage précédent dans le dataflow et qu'il porte le même nom que le le champ enregistré de l'index de recherche, les valeurs du champ d'entrée écraseront les valeurs dans le champ de sortie.</p>

### Options d'Index de recherche avancées

**Tableau 11 : Options de Candidate Finder**

Nom de l'option	Description/Valeurs valides
Type de Finder	Sélectionnez l'index de recherche.
Votre nom	Sélectionnez l'index approprié ayant été créé via le stage <a href="#">Write to Search Index</a> à la page 245 sous les stages déployés Advanced Matching dans Enterprise Designer.

Nom de l'option	Description/Valeurs valides
Enregistrement de départ	Saisissez le numéro d'enregistrement à partir duquel doivent commencer les résultats de recherche. La valeur par défaut est 1.
Nombre maximal de résultats	Saisissez le nombre maximal de réponses que l'index de recherche doit renvoyer. La valeur par défaut est 10.  <b>Remarque</b> : Si le nombre maximal de résultats est arbitrairement grand, traitez-les par lots, en utilisant le champ <b>Extraire la taille du lot</b> .
Extraire la taille du lot	Si le <b>Nombre maximal de résultats</b> est arbitrairement grand, saisissez la taille des lots dans lesquels vous voulez que les résultats soient traités. Cela optimise le traitement d'un grand nombre d'enregistrements. La valeur par défaut est 10 000.  La valeur <b>Extraire la taille du lot</b> recommandée est une valeur inférieure au <b>Nombre maximal de résultats</b> et si la valeur <b>Extraire la taille du lot</b> est supérieure au <b>Nombre maximal de résultats</b> , les enregistrements sont traités dans un seul lot.  <b>Remarque</b> : Ce champ s'applique uniquement au moteur de recherche pris en charge par le cluster et non au moteur de recherche patrimonial.
Trier	Trie les enregistrements candidats sur la base de champs indexés lors de l'exécution d'une requête de recherche.  Cochez la case <b>Trier</b> , le champ d'index souhaité dans la liste déroulante <b>Trier par</b> , et sélectionnez <b>Croissant</b> ou <b>Décroissant</b> dans la liste déroulante <b>Trier par</b> .  <b>Remarque</b> : Vous ne pouvez effectuer le tri que sur <b>Type de champ chaîne</b> et <b>Analyseur de mots clés</b> .
Renvoyer le nombre total de correspondances	Renvoie le nombre total de correspondances effectuées. Par exemple, si vous utilisez la valeur par défaut « 10 » pour le champ Nombre maximal de résultats ci-dessus, seuls 10 résultats seront renvoyés. En revanche, si vous cochez cette case, le champ de sortie TotalMatchCount vous indique le nombre de correspondances effectuées lors du traitement.
Pertinence	Contrôle la pertinence du <b>champ d'index</b> .
Type de recherche d'index	Détermine le type de recherche d'index que vous souhaitez effectuer. Sélectionnez <b>recherche avancée</b> .
Bouton Ajouter un parent	Accédez aux options parent.

Nom de l'option	Description/Valeurs valides
Options parents : nom	Saisissez le nom de le parent.
Options parents : Méthode de recherche	<p>Définissez comment déterminer si un parent est une correspondance ou non. L'une de ces valeurs :</p> <p>Toutes vraies : un parent est considéré comme une correspondance si tous les enfants sont déterminés correspondre. Cette méthode crée un connecteur « ET » entre les enfants.</p> <p>Au moins une : un parent est considéré comme une correspondance si au moins un enfant est déterminé correspondre. Cette méthode crée un connecteur « OU » entre les enfants.</p> <p>Aucun vrai : un parent est considéré comme une correspondance si aucun des enfants n'est déterminé correspondre. Cette méthode crée un connecteur "NOT" entre les enfants.</p>
Bouton Ajouter un enfant	Accédez aux options enfant.
Options enfant : champ Index	Sélectionnez le champ d'index que vous souhaitez utiliser pour la comparaison dans la recherche avancée.
Options enfant : type de recherche	Indique les critères de recherche/mise en correspondance qui déterminent si les données d'entrée sont recherchées/mises en correspondance avec les données indexées. Toutes les recherches sont sensibles à la casse.
Options enfant : champ d'entrée	Sélectionnez le champ d'entrée que vous souhaitez utiliser pour la comparaison dans la recherche avancée.
Tout terme/expression commence par	<p>Détermine si le texte contenu dans le champ d'index de recherche commence par le texte contenu dans le champ d'entrée.</p> <p>Par exemple, le texte dans le champ d'entrée "tech" serait considéré comme une correspondance pour les champs d'index de recherche contenant "Technical", "Technology", "Technologies", "Technician" ou même "National University of Technical Sciences". De même, une expression dans le champ d'entrée « DEF Sof » sera considérée comme une correspondance pour les champs d'index de recherche contenant « ABC DEF Software », « DEF Software » et « DEF Software India », mais ce ne serait pas une correspondance pour les champs d'index de recherche contenant « Software DEF » ou « DEF ABC Software ».</p>

Nom de l'option	Description/Valeurs valides
Contient	Détermine si le champ de l'index de recherche contient les données du champ d'entrée. Ce type de recherche prend en considération la séquence de mots dans le champ d'entrée tout en recherchant le champ d'index de recherche. Par exemple, les données de champ d'entrée « Pitney » et « Pitney Bowes » seraient contenues dans un champ d'index de recherche de « Pitney Bowes Software Inc ».
Contient tout	Détermine si tous les mots alphanumériques du champ d'entrée sont contenus dans le champ d'index de recherche. Ce type de recherche ne prend pas en considération la séquence de mots dans le champ d'entrée lors de la recherche du champ d'index de recherche.
Contient n'importe quel	Détermine si l'un des mots alphanumériques du champ d'entrée est contenu dans le champ d'index de recherche.
Ne contient aucun	Détermine si aucun des mots alphanumériques du champ d'entrée n'est contenu dans le champ Index de recherche.
Correspondance approximative	<p>Détermine la similitude entre deux mots alphanumériques en fonction du nombre de suppressions, d'insertions ou de substitutions requises pour transformer un mot en un autre.</p> <p>Utilisez le paramètre <b>Nombre maximal de modifications</b> pour définir une limite du nombre de modifications autorisées pour être considéré comme une correspondance réussie :</p> <ul style="list-style-type: none"> <li>• 0 : n'autorise aucune suppression, insertion, ni substitution. Les données du champ d'entrée et celles du champ d'index de recherche doivent être identiques.</li> <li>• 1 : n'autorise pas plus d'une suppression, insertion ou substitution. Par exemple, un champ d'entrée contenant "Barton" sera mis en correspondance avec un champ d'index de recherche contenant "Carton".</li> <li>• 2 : n'autorise pas plus de deux suppressions, insertions ou substitutions. Par exemple, un champ d'entrée contenant "Barton" sera mis en correspondance avec un champ d'index de recherche contenant "Martin".</li> </ul> <p>Le type de recherche Correspondance approximative n'est utilisé que pour les recherches de mots uniques. Cliquez sur <b>Ignorer les mots en trop</b> pour que Candidate Finder ne considère que le premier mot dans le champ lors de la comparaison du champ d'entrée avec le champ d'index. Par exemple, si le champ d'index indique "Pitney" et que le champ d'entrée indique "Pitney Bowes", ils ne sont pas considérés comme une correspondance à cause de "Bowes". Cependant, si vous cochez cette case, "Bowes" sera ignoré et dans la mesure où "Pitney" est le premier mot, les deux mots seraient considérés comme une correspondance.</p>

Nom de l'option	Description/Valeurs valides
numériques	<p>Détermine si tous les mots alphanumériques du champ d'entrée sont contenus dans le champ d'index de recherche.</p> <p>Le type de recherche Caractère générique n'est utilisé que pour les recherches de mots uniques.</p> <p>Cliquez sur <b>Ignorer les mots supplémentaires</b> pour que Candidate Finder ne prenne en compte que le premier mot du champ lors de la comparaison du champ d'entrée au champ d'index.</p>
Motif	<p>Détermine si le modèle de texte du champ d'entrée correspond au modèle de texte des critères de recherche. Vous pouvez encore affiner le modèle de texte dans le champ <b>Chaîne de modèles</b>. Par exemple, si le champ d'entrée contient "nlm" et que le modèle défini est "a*b?c", il sera mis en correspondance avec les mots suivants "Neelam", "nelam", "neelum", "nilam", etc.</p> <p>Le type de recherche Modèle n'est utilisé que pour les recherches de mots uniques. Cliquez sur <b>Ignorer les mots en trop</b> pour que Candidate Finder ne considère que le premier mot dans le champ lors de la comparaison du champ d'entrée avec le champ d'index.</p>
Proximité	<p>Détermine si les mots dans les champs d'entrée se situent à une certaine distance l'un de l'autre.</p> <ul style="list-style-type: none"> <li>• Définissez l'entrée <b>Champ de première entrée</b> et <b>Champ de deuxième entrée</b> à rechercher dans l'index.</li> <li>• Utilisez le paramètre <b>Distance</b> pour déterminer la distance maximale autorisée entre les mots indiqués dans le champ Premier et le champ Second afin d'être considéré comme une correspondance.</li> </ul> <p>Par exemple, vous pouvez utiliser ce type de recherche pour rechercher le premier champ "Spectrum" et le second champ "Pitney" à dix mots d'intervalle, dans un champ d'index de recherche contenant la phrase "Spectrum Technology Platform est un produit de Pitney Bowes Software Inc."</p> <p>Le type de recherche Proximité n'est utilisé que pour les recherches de mots uniques. Cliquez sur <b>Ignorer les mots en trop</b> pour que Candidate Finder ne considère que le premier mot dans le champ lors de la comparaison du champ d'entrée avec le champ d'index.</p>

Nom de l'option	Description/Valeurs valides
Plage	<p>Effectue une recherche inclusive de termes dans une plage spécifiée à l'aide d'un champ de limite inférieure (terme de départ) et un champ de limite supérieure (terme de fin). Tous les mots alphanumériques sont organisés de manière lexicographique dans le champ d'index de recherche.</p> <ul style="list-style-type: none"> <li>• Utilisez le paramètre <b>Champ de limite inférieure</b> pour sélectionner le champ à utiliser en tant que terme de départ.</li> <li>• Utilisez le paramètre <b>Champ de limite supérieure</b> pour sélectionner le champ à utiliser en tant que terme de fin.</li> </ul> <p>Par exemple, si vous avez recherché des codes postaux à partir de 20001 (défini dans le champ de limite inférieure) à 20009 (défini dans le champ de limite supérieure), la recherche renverra toutes les adresses contenant des codes postaux compris dans cette plage.</p> <p>Le type de recherche Plage n'est utilisé que pour les recherches de mots uniques. Cliquez sur <b>Ignorer les mots en trop</b> pour que Candidate Finder ne considère que le premier mot dans le champ lors de la comparaison du champ d'entrée avec le champ d'index.</p>
Caractère générique	<p>Recherches utilisant des <b>caractères génériques</b> uniques ou multiples :</p> <p>Sélectionnez la <b>position</b> dans votre fichier d'entrée dans lequel vous insérez le caractère générique.</p> <p>Le type de recherche Caractère générique n'est utilisé que pour les recherches de mots uniques. Cliquez sur <b>Ignorer les mots en trop</b> pour que Candidate Finder ne considère que le premier mot dans le champ lors de la comparaison du champ d'entrée avec le champ d'index.</p>
Options enfant : facteur de pertinence	<p>Contrôlez la pertinence d'un champ enfant en saisissant ici un nombre positif jusqu'à 100. Le nombre peut aussi être inférieur à « 1 » ; par exemple, « 0,05 » serait valide.</p> <p>Plus le facteur de pertinence sera élevé, plus le champ sera pertinent. Par exemple, si vous souhaitez que les résultats du champ Nom de la société soient plus pertinents que ceux des autres champs, sélectionnez Nom de la société dans <b>Nom du champ d'index</b> et saisissez "5".</p> <p><b>Remarque :</b> Par défaut, cette option est désactivée. Cochez la case pour l'activer.</p>
Ignorer les vides	<p>Décochez cette case si vous souhaitez que la requête prenne en compte les champs de fichier d'entrée vides.</p> <p><b>Remarque :</b> Par défaut, la requête ignore les champs vides.</p>

Nom de l'option	Description/Valeurs valides
Onglet Champs de sortie	<p>Cochez la case <b>Inclure</b> pour sélectionner les champs à inclure dans la sortie.</p> <p><b>Remarque</b> : Si le champ d'entrée provient d'un stage précédent dans le dataflow et qu'il porte le même nom que le le champ enregistré de l'index de recherche, les valeurs du champ d'entrée écraseront les valeurs dans le champ de sortie.</p>

### Configuration des options lors de l'exécution

Les options Some Candidate Finder peuvent être configurées et transmises lors de l'exécution si elles sont exposées sous forme d'options de flux de données. Cela vous permet d'exécuter votre flux de données tout en utilisant des configurations différentes. Voici les options de flux de données disponibles pour Candidate Finder :

- **ConnectionString** : nom de la base de données contenant les enregistrements candidats.
- **SearchIndexName** : nom de l'index de recherche utilisé dans le flux de données de Candidate Finder.
- **StartingRecord** : numéro d'enregistrement à partir duquel doivent commencer les résultats de recherche.
- **MaximumResults** : nombre maximal de réponses que l'index de recherche doit renvoyer.
- **ReturnMatchCount** : nombre total de correspondances effectuées. Ce champ est utile si vous saisissez un nombre inférieur dans le champ **MaximumResults**, mais que vous souhaitez connaître le nombre total de correspondances effectuées.

Pour définir des options Candidate Finder lors de l'exécution, procédez comme suit :

1. Dans Enterprise Designer, ouvrez un flux de données utilisant le stage Candidate Finder.
2. Enregistrez et exposez ce flux de données.
3. Accédez à `Edit > Dataflow Options`.
4. Dans la table **Map dataflow options to stages**, développez Candidate Finder et modifiez les options, le cas échéant. Cochez la case de l'option que vous souhaitez modifier, puis modifiez la valeur dans la liste déroulante **Valeur par défaut**.
5. Facultatif : modifiez le nom des options du champ **Option label**.
6. Cliquez deux fois sur **OK**.



## Sortie

Tableau 12 : Sorties de Candidate Finder

Nom du champ	Format	Description/Valeurs valides
CandidateCount	Chaîne	Ce champ indique le nombre total de candidats renvoyés pendant le traitement.
CandidateGroup	Chaîne	Ce champ identifie un groupe composé d'un enregistrement suspect et de ses candidats. Chaque enregistrement suspect reçoit un numéro CandidateGroup. Les candidats correspondant à ce suspect reçoivent le même numéro CandidateGroup. Par exemple, si Pierre Martin est un enregistrement suspect et que ses enregistrements candidats sont Pierre Martin et Perre Martn, ces trois enregistrements auraient la même valeur CandidateGroup.
HasDuplicates	Chaîne	Identifie si les candidats sont détectés ou non. Les valeurs possibles sont :  <b>Y</b> - L'enregistrement est suspect et a des candidats <b>N</b> - L'enregistrement est suspect et n'a pas de candidats <b>D</b> - L'enregistrement est un enregistrement candidat
TotalMatchCount	Chaîne	Ce champ indique le nombre total de correspondances effectuées lors du traitement.
TransactionRecordType	Chaîne	L'un des éléments suivants :  <b>Suspect</b> Un enregistrement suspect est utilisé comme entrée dans une requête.  <b>Candidate</b> C'est un enregistrement candidat qui sera un résultat renvoyé par une requête.

## Duplicate Synchronization

Duplicate Synchronization détermine quels champs d'un groupe d'enregistrements doivent être copiés dans les champs correspondants de tous les enregistrements du groupe. Vous pouvez indiquer les règles auxquelles les enregistrements doivent satisfaire afin de copier les données d'un

champ vers d'autres enregistrements du groupe. Une fois le traitement effectué, tous les enregistrements du groupe sont conservés.

## Options

Le tableau répertorie les options du stage **Duplicate Synchronization**.

Nom de l'option	Description/Valeurs valides
Grouper par	Indique le champ à utiliser pour créer des groupes d'enregistrements à synchroniser. Si vous avez utilisé un stage de correspondance précédemment dans le flux de données, tel que Interflow Match, Intraflow Match ou Transactional Match, vous devez sélectionner le champ CollectionNumber pour utiliser les collections créées par le stage de correspondance en tant que groupes. Cependant, si vous souhaitez regrouper les enregistrements par un autre champ, choisissez ce dernier ici. Par exemple, si vous souhaitez synchroniser les enregistrements dotés de la même valeur dans le champ AccountNumber, vous devez sélectionner AccountNumber.
Trier	Si vous spécifiez un champ dans le champ <b>Grouper par</b> , cochez cette case pour trier les enregistrements par la valeur contenu dans le champ de votre choix. Cette option est activée par défaut.

Nom de l'option	Description/Valeurs valides
-----------------	-----------------------------

---

Avancé	
--------	--

## Nom de l'option

## Description/Valeurs valides

Cliquez sur ce bouton pour indiquer les options de performances du tri. Par défaut, les options de performances du tri spécifiées dans Management Console, qui représentent les options de performances par défaut pour votre système, sont appliquées. Si vous souhaitez remplacer les options de performances par défaut de votre système, cochez la case **Écraser les options de performance du tri**, puis spécifiez les valeurs que vous souhaitez obtenir dans ces champs :

**Limite du nombre d'enregistrement en mémoire** Spécifie le nombre maximum de rangées de données qu'un trieur peut contenir en mémoire avant que celui-ci commence à pager sur le disque. Par défaut, un tri de 10 000 enregistrements ou moins sera effectué en mémoire et un tri de plus de 10 000 enregistrements sera effectué sur le disque. La limite maximale est de 100 000 enregistrements. En général, un tri en mémoire est beaucoup plus rapide qu'un tri sur le disque ; donc, il faut définir une valeur assez haute pour que la plupart des tris s'effectuent en mémoire et que seuls les groupes de grande taille soient écrits sur le disque.

**Remarque :** Soyez conscient du fait qu'au sein d'environnements où des jobs s'exécutent de manière simultanée, une augmentation du paramètre **Dans la limite d'enregistrement mémoire** augmente la probabilité de ne plus disposer de suffisamment de mémoire.

**Nombre maximal de fichiers temporaires** Spécifie le nombre maximal de fichiers temporaires pouvant être employés par un processus de tri. L'utilisation d'un plus grand nombre de fichiers temporaires peut améliorer les performances. Cependant, le nombre optimal dépend très largement de la configuration du serveur qui exécute Spectrum™ Technology Platform. Nous vous conseillons d'essayer différents paramètres et d'observer l'effet de l'utilisation d'un plus ou moins grand nombre de fichiers temporaires sur les performances. Pour calculer le nombre approximatif de fichiers temporaires nécessaires, utilisez l'équation suivante :

$$\frac{(NumberOfRecords \times 2)}{InMemoryRecordLimit} = NumberOfTempFiles$$

Notez que le nombre maximal de fichiers temporaires ne peut pas être supérieure à 1 000.

**Activer la compression** Spécifie si les fichiers temporaires sont compressés lors de leur écriture sur le disque.

**Remarque :** Les paramètres de performances de tri optimal dépendent de la configuration matérielle de votre serveur. Néanmoins, l'équation suivante produit généralement de bonnes performances de tri :

Nom de l'option	Description/Valeurs valides
-----------------	-----------------------------

$$(InMemoryRecordLimit \times MaxNumberOfTempFiles \div 2) \geq TotalNumberOfRecords$$

### Règles

Les règles Duplicate Synchronization déterminent les enregistrements dont les données doivent être copiées dans tous les autres enregistrements de la collection.

Pour ajouter une règle, sélectionnez Règles dans la hiérarchie des règles et cliquez sur **Ajouter une règle**

Si vous spécifiez plusieurs règles, vous devrez sélectionner un opérateur logique à utiliser entre chaque règle. Choisissez **Et** si vous souhaitez que la nouvelle règle et la règle précédente soient toutes les deux respectées pour que la condition soit remplie. Sélectionnez **Ou** si vous souhaitez que la règle précédente ou la nouvelle règle soit respectée pour que la condition soit remplie.

Option	Description				
Nom de champ	Indique le nom du champ de flux de données dont vous souhaitez évaluer la valeur afin de déterminer si l'enregistrement doit être filtré.				
Type de champ	Indique le type de données dans le champ. L'un des éléments suivants : <table border="0"> <tr> <td><b>Non-numérique</b></td> <td>Choisissez cette option si le champ contient des données non numériques (par exemple, des données de chaîne).</td> </tr> <tr> <td><b>numériques</b></td> <td>Choisissez cette option si le champ contient des données numériques (par exemple, double, flottantes, etc.).</td> </tr> </table>	<b>Non-numérique</b>	Choisissez cette option si le champ contient des données non numériques (par exemple, des données de chaîne).	<b>numériques</b>	Choisissez cette option si le champ contient des données numériques (par exemple, double, flottantes, etc.).
<b>Non-numérique</b>	Choisissez cette option si le champ contient des données non numériques (par exemple, des données de chaîne).				
<b>numériques</b>	Choisissez cette option si le champ contient des données numériques (par exemple, double, flottantes, etc.).				

Option	Description
Opérateur	Indique le type de comparaison à utiliser pour évaluer le champ. L'un des éléments suivants :
<b>Contient</b>	Détermine si le champ contient la valeur indiquée. Par exemple, « bateau à voile » contient la valeur « bateau ».
<b>Est égal à</b>	Détermine si le champ contient la valeur exacte indiquée.
<b>Est supérieur à</b>	Détermine si la valeur du champ est supérieure à la valeur spécifiée. Cette opération ne fonctionne que sur les champs numériques.
<b>Supérieur ou Egal à</b>	Détermine si la valeur du champ est supérieure ou égale à la valeur spécifiée. Cette opération ne fonctionne que sur les champs numériques.
<b>Le plus haut</b>	Compare la valeur du champ pour tout le groupe d'enregistrements et identifie l'enregistrement qui possède la valeur la plus élevée dans le champ. Par exemple, si les champs dans ce groupe contiennent les valeurs 10, 20, 30, et 100, l'enregistrement qui dispose de la valeur 100 est sélectionné. Cette opération ne fonctionne que sur les champs numériques. Si plusieurs enregistrements sont à égalité pour la valeur la plus longue, un enregistrement est sélectionné.
<b>Est vide</b>	Détermine si le champ ne contient aucune valeur.
<b>N'est pas vide</b>	Détermine si le champ contient une valeur.
<b>Inférieur à</b>	Détermine si la valeur du champ est inférieure à la valeur spécifiée. Cette opération ne fonctionne que sur les champs numériques.
<b>Inférieur ou Egal à</b>	Détermine si la valeur du champ est inférieure ou égale à la valeur spécifiée. Cette opération ne fonctionne que sur les champs numériques.
<b>Le plus long</b>	Compare la valeur du champ pour tout le groupe d'enregistrements et identifie l'enregistrement qui possède la valeur la plus longue (en octets) dans le champ. Par exemple, si le groupe contient les valeurs « Mike » et « Michael », l'enregistrement possédant la valeur « Michael » est choisi. Si plusieurs enregistrements sont à égalité pour la valeur la plus longue, un enregistrement est sélectionné.
<b>Le plus bas</b>	Compare la valeur du champ pour tout le groupe d'enregistrements et identifie l'enregistrement qui possède la valeur la plus basse dans le champ. Par exemple, si les champs dans ce groupe contiennent les valeurs 10, 20, 30, et 100, l'enregistrement qui dispose de la valeur 10 est sélectionné. Cette opération ne fonctionne que sur les champs numériques. Si plusieurs enregistrements sont à égalité pour la valeur la plus longue, un enregistrement est sélectionné.
<b>Le plus commun</b>	Détermine si la valeur de champ contient la valeur qui se produit le plus fréquemment dans ce champ parmi les enregistrements dans ce groupe. Si deux valeurs, ou plus, sont les plus courantes, aucune action ne sera réalisée.
<b>N'est pas égal</b>	Détermine si la valeur du champ n'est pas égale à la valeur spécifiée.

Option	Description
Type de valeur	<p>Indique le type de valeur à comparer avec la valeur du champ. L'un des éléments suivants :</p> <p><b>Remarque</b> : Cette option n'est pas disponible si vous sélectionnez l'opérateur Highest, Lowest ou Longest.</p> <p><b>Champ</b> Choisissez cette option si vous souhaitez comparer une autre valeur du champ de dataflow avec le champ.</p> <p><b>Chaîne</b> Choisissez cette option si vous souhaitez comparer le champ avec une valeur spécifique.</p>
Valeur	<p>Indique la valeur à comparer avec la valeur du champ. Si vous avez sélectionné Champ dans le champ <b>Type de champ</b>, sélectionnez un champ de dataflow. Si vous avez sélectionné Chaîne dans le champ <b>Type de valeur</b>, tapez la valeur à utiliser dans la comparaison.</p> <p><b>Remarque</b> : Cette option n'est pas disponible si vous sélectionnez l'opérateur Highest, Lowest ou Longest.</p>

### Actions

Les actions déterminent le champ à copier dans les autres enregistrements du groupe. Pour ajouter une action, sélectionnez Actions dans l'arborescence de la condition Duplicate Synchronization, puis cliquez sur **Ajouter une action**. Utilisez les options suivantes pour définir l'action.

Option	Description
Type de la source	<p>Spécifie le type de données à copier dans les autres enregistrements du groupe. Un des éléments suivants.</p> <p><b>Champ</b> Choisissez cette option si vous souhaitez copier une valeur d'un champ dans d'autres enregistrements du groupe.</p> <p><b>Chaîne</b> Choisissez cette option si vous souhaitez copier une valeur constante dans les autres enregistrements du groupe.</p>
Données sources	<p>Spécifie les données à copier dans les autres enregistrements du groupe. Si le type de source est Champ, sélectionnez le champ dont vous souhaitez copier la valeur dans les autres enregistrements du groupe. Si le type source est Chaîne, indiquez une valeur constante à copier dans les autres enregistrements du groupe.</p> <p><b>Remarque</b> : Dans le cas où les données sources ont une valeur null, elles ne sont pas copiées dans les autres enregistrements du groupe. Les autres enregistrements conservent leurs valeurs d'origine.</p>

Option	Description
Cible	Spécifie le champ des autres enregistrements dans lequel vous souhaitez copier les données indiquées dans le champ <b>Données source</b> . Par exemple, si vous souhaitez copier les données dans le champ AccountBalance de tous les autres enregistrements du groupe, vous devez indiquer AccountBalance.

### Exemple de règle et d'action Duplicate Synchronization

Cette règle et action Duplicate Synchronization sélectionne l'enregistrement doté d'un score de correspondance de 100 et copie le champ AccountNumber de numéro de compte dans tous les autres enregistrements du groupe.

#### Règle

Nom du champ : MatchScore  
 Type de champ : Numérique  
 Opérateur : Est égal  
 Type de valeur : Chaîne  
 Valeur : 100

#### Action

Type de source : Champ  
 Données source : AccountNumber  
 Destination : NewAccountNumber

## Filtre

Le stage Filter conserve ou enlève des enregistrements d'un groupe en fonction des règles que vous indiquez.

### Options

Le tableau suivant répertorie les options du stage Filter.



Nom de l'option	Description/Valeurs valides
Grouper par	<p>Indique le champ à utiliser pour créer des groupes d'enregistrements à filtrer. Le stage Filter conservera un ou plusieurs enregistrements de chaque groupe, selon la manière dont vous configurez le stage. Si vous avez utilisé un stage de correspondance précédemment dans le flux de données, tel que Interflow Match, Intraflow Match ou Transactional Match, vous devez sélectionner le champ CollectionNumber pour utiliser les collections créées par le stage de correspondance en tant que groupes. Cependant, si vous souhaitez regrouper les enregistrements par un autre champ, choisissez ce dernier ici. Par exemple, si vous souhaitez filtrer tous les enregistrements dotés de la même valeur dans le champ AccountNumber sauf un, vous devez sélectionner AccountNumber.</p>
Trier	<p>Si vous spécifiez un champ dans le champ <b>Grouper par</b>, cochez cette case pour trier les enregistrements par la valeur contenu dans le champ de votre choix. Cette option est activée par défaut.</p>

Nom de l'option	Description/Valeurs valides
-----------------	-----------------------------

---

Avancé	
--------	--

## Nom de l'option

## Description/Valeurs valides

Cliquez sur ce bouton pour indiquer les options de performances du tri. Par défaut, les options de performances du tri spécifiées dans Management Console, qui représentent les options de performances par défaut pour votre système, sont appliquées. Si vous souhaitez remplacer les options de performances par défaut de votre système, cochez la case **Écraser les options de performance du tri**, puis spécifiez les valeurs que vous souhaitez obtenir dans ces champs :

**Limite du nombre d'enregistrement en mémoire** Spécifie le nombre maximum de rangées de données qu'un trieur peut contenir en mémoire avant que celui-ci commence à pager sur le disque. Par défaut, un tri de 10 000 enregistrements ou moins sera effectué en mémoire et un tri de plus de 10 000 enregistrements sera effectué sur le disque. La limite maximale est de 100 000 enregistrements. En général, un tri en mémoire est beaucoup plus rapide qu'un tri sur le disque ; donc, il faut définir une valeur assez haute pour que la plupart des tris s'effectuent en mémoire et que seuls les groupes de grande taille soient écrits sur le disque.

**Remarque :** Soyez conscient du fait qu'au sein d'environnements où des jobs s'exécutent de manière simultanée, une augmentation du paramètre **Dans la limite d'enregistrement mémoire** augmente la probabilité de ne plus disposer de suffisamment de mémoire.

**Nombre maximal de fichiers temporaires** Spécifie le nombre maximal de fichiers temporaires pouvant être employés par un processus de tri. L'utilisation d'un plus grand nombre de fichiers temporaires peut améliorer les performances. Cependant, le nombre optimal dépend très largement de la configuration du serveur qui exécute Spectrum™ Technology Platform. Nous vous conseillons d'essayer différents paramètres et d'observer l'effet de l'utilisation d'un plus ou moins grand nombre de fichiers temporaires sur les performances. Pour calculer le nombre approximatif de fichiers temporaires nécessaires, utilisez l'équation suivante :

$$\frac{(NumberOfRecords \times 2)}{InMemoryRecordLimit} = NumberOfTempFiles$$

Notez que le nombre maximal de fichiers temporaires ne peut pas être supérieure à 1 000.

**Activer la compression** Spécifie si les fichiers temporaires sont compressés lors de leur écriture sur le disque.

**Remarque :** Les paramètres de performances de tri optimal dépendent de la configuration matérielle de votre serveur. Néanmoins, l'équation suivante produit généralement de bonnes performances de tri :

Nom de l'option	Description/Valeurs valides
-----------------	-----------------------------

$$(InMemoryRecordLimit \times MaxNumberOfTempFiles \div 2) \geq TotalNumberOfRecords$$

<p>Limiter le nombre d'enregistrements doublons renvoyés</p>	<p>Définit le nombre maximal d'enregistrements renvoyés de chaque groupe. Si vous définissez cette option sur 1, vous pouvez définir des règles de filtre pour déterminer l'enregistrement à renvoyer dans chaque groupe. Si aucune règle n'est définie, le premier enregistrement de chaque collection sera renvoyé et le reste ignoré. Dans ce mode, les règles de filtre définissent l'enregistrement qui sera conservé.</p> <p>Par exemple, si vous définissez une règle dans laquelle l'enregistrement doté du score de correspondance le plus élevé dans un groupe est conservé et que vous définissez cette option sur 1, l'enregistrement doté du score de correspondance le plus élevé dans chaque groupe survivra et les autres enregistrements du groupe seront abandonnés.</p> <p>Si vous définissez cette option sur une valeur supérieure à un, vous ne pouvez pas spécifier les règles de filtre.</p>
--	--

**Remarque :** Dans l'éventualité où aucun enregistrement de la collection ne répond aux critères de la règle définie, aucun enregistrement du groupe n'est renvoyé.

<p>Supprimer les doublons du groupe</p>	<p>Indique d'utiliser les règles de filtre pour déterminer les enregistrements qui sont enlevés de la collection. Les enregistrements restants dans la collection sont conservés. Quand cette option est sélectionnée, vous devez définir une règle.</p>
---	--

**Remarque :** Si un groupe ne contient qu'un enregistrement, les règles de filtre sont ignorées et l'enregistrement est conservé.

### Options de règle

Les règles de filtre déterminent les enregistrements d'un groupe à conserver ou à enlever. Si vous sélectionnez l'option **Limiter le nombre d'enregistrements doublons renvoyés**, les règles déterminent les enregistrements qui survivent au filtre. Si vous sélectionnez l'option **Supprimer les doublons du groupe**, les règles déterminent les enregistrements qui sont enlevés du flux de données.

Pour ajouter une règle, sélectionnez Règles dans la hiérarchie des règles et cliquez sur **Ajouter une règle**

Si vous spécifiez plusieurs règles, vous devrez sélectionner un opérateur logique à utiliser entre chaque règle. Choisissez **Et** si vous souhaitez que la nouvelle règle et la règle précédente soient toutes les deux respectées pour que la condition soit remplie. Sélectionnez **Ou** si vous souhaitez que la règle précédente ou la nouvelle règle soit respectée pour que la condition soit remplie.

**Remarque :** Vous ne pouvez disposer que d'une condition dans un stage Filter. Quand vous sélectionnez Condition dans la hiérarchie de règles, les boutons sont grisés.

Option	Description
Nom du champ	Indique le nom du champ de flux de données dont vous souhaitez évaluer la valeur afin de déterminer si l'enregistrement doit être filtré.
Type de champ	Indique le type de données dans le champ. L'un des éléments suivants : <b>Non-numérique</b> Choisissez cette option si le champ contient des données non numériques (par exemple, des données de chaîne). <b>numériques</b> Choisissez cette option si le champ contient des données numériques (par exemple, double, flottantes, etc.).

Option	Description
Opérateur	Indique le type de comparaison à utiliser pour évaluer le champ. L'un des éléments suivants :
<b>Contient</b>	Détermine si le champ contient la valeur indiquée. Par exemple, « bateau à voile » contient la valeur « bateau ».
<b>Est égal à</b>	Détermine si le champ contient la valeur exacte indiquée.
<b>Est supérieur à</b>	Détermine si la valeur du champ est supérieure à la valeur spécifiée. Cette opération ne fonctionne que sur les champs numériques.
<b>Supérieur ou Egal à</b>	Détermine si la valeur du champ est supérieure ou égale à la valeur spécifiée. Cette opération ne fonctionne que sur les champs numériques.
<b>Le plus haut</b>	Compare la valeur du champ pour tout le groupe d'enregistrements et identifie l'enregistrement qui possède la valeur la plus élevée dans le champ. Par exemple, si les champs dans ce groupe contiennent les valeurs 10, 20, 30, et 100, l'enregistrement qui dispose de la valeur 100 est sélectionné. Cette opération ne fonctionne que sur les champs numériques. Si plusieurs enregistrements sont à égalité pour la valeur la plus longue, un enregistrement est sélectionné.
<b>Est vide</b>	Détermine si le champ ne contient aucune valeur.
<b>N'est pas vide</b>	Détermine si le champ contient une valeur.
<b>Inférieur à</b>	Détermine si la valeur du champ est inférieure à la valeur spécifiée. Cette opération ne fonctionne que sur les champs numériques.
<b>Inférieur ou Egal à</b>	Détermine si la valeur du champ est inférieure ou égale à la valeur spécifiée. Cette opération ne fonctionne que sur les champs numériques.
<b>Le plus long</b>	Compare la valeur du champ pour tout le groupe d'enregistrements et identifie l'enregistrement qui possède la valeur la plus longue (en octets) dans le champ. Par exemple, si le groupe contient les valeurs « Mike » et « Michael », l'enregistrement possédant la valeur « Michael » est choisi. Si plusieurs enregistrements sont à égalité pour la valeur la plus longue, un enregistrement est sélectionné.
<b>Le plus bas</b>	Compare la valeur du champ pour tout le groupe d'enregistrements et identifie l'enregistrement qui possède la valeur la plus basse dans le champ. Par exemple, si les champs dans ce groupe contiennent les valeurs 10, 20, 30, et 100, l'enregistrement qui dispose de la valeur 10 est sélectionné. Cette opération ne fonctionne que sur les champs numériques. Si plusieurs enregistrements sont à égalité pour la valeur la plus longue, un enregistrement est sélectionné.
<b>Le plus commun</b>	Détermine si la valeur de champ contient la valeur qui se produit le plus fréquemment dans ce champ parmi les enregistrements dans ce groupe. Si deux valeurs, ou plus, sont les plus courantes, aucune action ne sera réalisée.
<b>N'est pas égal</b>	Détermine si la valeur du champ n'est pas égale à la valeur spécifiée.

Option	Description
Type de valeur	<p>Indique le type de valeur à comparer avec la valeur du champ. L'un des éléments suivants :</p> <p><b>Remarque</b> : Cette option n'est pas disponible si vous sélectionnez l'opérateur Highest, Lowest ou Longest.</p> <p><b>Champ</b> Choisissez cette option si vous souhaitez comparer une autre valeur du champ de dataflow avec le champ.</p> <p><b>Chaîne</b> Choisissez cette option si vous souhaitez comparer le champ avec une valeur spécifique.</p>
Valeur	<p>Indique la valeur à comparer avec la valeur du champ. Si vous avez sélectionné Champ dans le champ <b>Type de champ</b>, sélectionnez un champ de dataflow. Si vous avez sélectionné Chaîne dans le champ <b>Type de valeur</b>, tapez la valeur à utiliser dans la comparaison.</p> <p><b>Remarque</b> : Cette option n'est pas disponible si vous sélectionnez l'opérateur Highest, Lowest ou Longest.</p>

#### Exemple de règle de filtre

Cette règle conserve l'enregistrement dans chaque groupe doté de la valeur la plus élevée dans le champ MatchScore. Notez que les options **Valeur** et **Type de valeur** ne s'appliquent pas quand l'opérateur est le plus haut ou le plus bas.

Nom du champ = MatchScore  
 Type de champ = Numérique  
 Opérateur = Le plus haut

Cette règle conserve l'enregistrement pour lequel la valeur dans AccountNumber est "12345".

Nom du champ = AccountNumber  
 Type de champ = Numérique  
 Opérateur = Est égal à  
 Type de valeur = Chaîne  
 Valeur = 12345

## Interflow Match

Interflow Match localise des rapprochements entre des enregistrements de données similaires en croisant deux flux d'enregistrements d'entrée. Le premier flux d'enregistrement est une source d'enregistrements suspects et le second flux est une source d'enregistrements candidats.

À l'aide des critères d'un groupe de rapprochements (par exemple une match key), Interflow Match identifie un groupe d'enregistrements pouvant potentiellement être des doublons d'un enregistrement suspect particulier.

Chaque candidat est rapproché de manière séparée au Suspect, et un score lui est attribué en fonction des règles du rapprochement. Si le candidat est un doublon, un numéro de groupe lui est attribué, le type de l'enregistrement de rapprochement est étiqueté comme doublon, et écrasé ; les candidats uniques ne disposant d'aucun rapprochement peuvent être écrasés de manière facultative par l'utilisateur. Lorsque Interflow Match a épuisé tous les enregistrements candidats du groupe de doublons actuel, l'enregistrement suspect rapproché reçoit un numéro de groupe qui correspond à son enregistrement doublon. Ou bien, si aucun rapprochement n'a été identifié, le suspect reçoit un numéro de groupe « 0 » et il est étiqueté enregistrement unique.

**Remarque :** Interflow Match ne met en correspondance les enregistrements suspects qu'avec les enregistrements candidats. Il ne tente pas de mettre en correspondance les enregistrements suspects avec d'autres enregistrements suspects, comme cela est le cas dans Intraflow Match.

La procédure de rapprochement d'un suspect particulier peut se terminer avant le rapprochement de tous les candidats possibles si vous avez défini un limiteur sur les doublons et que la limite a été dépassée pour le suspect en cours.

Le type de rapprochement (Intraflow ou Interflow) détermine la manière selon laquelle les résultats de rapprochement d'une clé express se traduisent en scores de rapprochements candidats (Candidate Match Scores). Lors d'un rapprochement de type Interflow, un rapprochement à l'aide d'une clé de rapprochement express réussi donne toujours un score de rapprochement (MatchScore) de 100 au Candidat. En revanche, lors d'un rapprochement de type Intraflow, le score que gagne le Candidat à l'issue d'un rapprochement grâce à une clé de rapprochement express dépend de l'éventualité selon laquelle l'enregistrement avec lequel ce Candidat a été rapproché était le rapprochement d'un autre Suspect : les doublons d'une clé de rapprochement express d'un Suspect auront toujours des scores de rapprochements de 100, alors que les doublons d'une clé de rapprochement express d'un autre Candidat (qui était le doublon d'un Suspect) hériteront du score de rapprochement de ce Candidat (pas nécessairement de 100).

## Options

1. Dans le champ **Charger une règle de rapprochement**, sélectionnez une des règles de correspondance prédéfinies que vous pouvez utiliser tel quel ou modifier pour répondre à vos besoins. Si vous souhaitez créer une nouvelle règle de correspondance sans utiliser une des règles de correspondance prédéfinies comme point de départ, cliquez sur **Nouveau**. Vous ne pouvez disposer que d'une règle personnalisée dans un flux de données.

**Remarque :** La fonction Options de flux de données dans Enterprise Designer permet d'afficher la règle de correspondance pour la configuration au moment de l'exécution.

2. Cliquez sur **Grouper par** pour sélectionner un champ à utiliser pour grouper les champs dans la file de correspondance. La Correspondance intra-flux n'essaie de faire correspondre que les enregistrements de la même file de correspondance.



3. Cochez la case **Trier** pour réaliser un tri de pré-rapprochement de votre entrée en fonction du champ sélectionné dans le champ **Grouper par**.
4. Cliquez sur **Avancé** pour indiquer d'autres options de performances du tri.

**Limite du nombre d'enregistrements en mémoire** Spécifie le nombre maximum de rangées de données qu'un trieur peut contenir en mémoire avant que celui-ci commence à pager sur le disque. Par défaut, un tri de 10 000 enregistrements ou moins sera effectué en mémoire et un tri de plus de 10 000 enregistrements sera effectué sur le disque. La limite maximale est de 100 000 enregistrements. En général, un tri en mémoire est beaucoup plus rapide qu'un tri sur le disque ; donc, il faut définir une valeur assez haute pour que la plupart des tris s'effectuent en mémoire et que seuls les groupes de grande taille soient écrits sur le disque.

**Remarque :** Soyez conscient du fait qu'au sein d'environnements où des jobs s'exécutent de manière simultanée, une augmentation du paramètre **Dans la limite d'enregistrement mémoire** augmente la probabilité de ne plus disposer de suffisamment de mémoire.

**Nombre maximal de fichiers temporaires** Spécifie le nombre maximal de fichiers temporaires pouvant être employés par un processus de tri. L'utilisation d'un plus grand nombre de fichiers temporaires peut améliorer les performances. Cependant, le nombre optimal dépend très largement de la configuration du serveur qui exécute Spectrum™ Technology Platform. Nous vous conseillons d'essayer différents paramètres et d'observer l'effet de l'utilisation d'un plus ou moins grand nombre de fichiers temporaires sur les performances. Pour calculer le nombre approximatif de fichiers temporaires nécessaires, utilisez l'équation suivante :

$$\frac{(NumberOfRecords \times 2)}{InMemoryRecordLimit} = NumberOfTempFiles$$

Notez que le nombre maximal de fichiers temporaires ne peut pas être supérieure à 1 000.

**Activer la compression** Spécifie si les fichiers temporaires sont compressés lors de leur écriture sur le disque.

**Remarque :** Les paramètres de performances de tri optimal dépendent de la configuration matérielle de votre serveur. Néanmoins, l'équation suivante produit généralement de bonnes performances de tri :

$$(InMemoryRecordLimit \times MaxNumberOfTempFiles \div 2) \geq TotalNumberOfRecords$$

5. Cliquez sur **Rapprochement express sur** pour effectuer une comparaison initiale des valeurs de clé express pour déterminer si les deux enregistrements sont considérés comme correspondants.

Le rapprochement par clé express peut s'avérer un outil utile pour réduire le nombre de comparaisons réalisé et ainsi améliorer la vitesse d'exécution. Une clé express lâche donne lieu à de nombreuses correspondances de type faux positifs. Vous pouvez créer une clé express en créant une clé de correspondance dans le MatchKeyGenerator. Pour plus d'informations, reportez-vous à la section [Match Key Generator](#) à la page 234.

Si deux enregistrements donnent une correspondance exacte sur la clé express, le candidat est considéré comme un doublon à 100%. Si deux enregistrements ne correspondent pas sur une valeur de clé express, ceux-ci sont comparés via la méthode fondée sur les règles.

Pour savoir si un candidat a été mis en correspondance en utilisant une clé expresse, vérifiez la valeur du champ **ExpressKeyIdentified**, où Y signifie qu'il y a correspondance et N signifie qu'il n'y en a pas. Notez que les enregistrements suspects ont toujours une valeur **ExpressKeyIdentified** de N.

6. Dans la zone de texte **Numéro de collection initial**, indiquez le numéro de départ à assigner au champ du numéro de collection pour les enregistrements doublons.

Le numéro de collection identifie chaque enregistrement doublon dans une file de correspondance. Les enregistrements uniques se voient assigner un numéro de collection de 0. Chaque enregistrement doublon se voit assigner un numéro de collection débutant pas la valeur indiquée dans l'encadré **Numéro de collection initial**.

7. Sélectionnez l'un des choix suivants :

Option	Description
<b>Comparer le suspect avec tous les candidats</b>	<p>Cette option met en correspondance le suspect avec tous les candidats du même groupe de correspondance (groupe par option) même si un doublon a déjà été trouvé dans le groupe de correspondance. Par exemple :</p> <p>Suspect : John Smith            Candidat : Bill Jones            Candidat : John Smith            Candidat : John Smith</p> <p>Dans l'exemple, le suspect John Smith sera comparé aux deux candidats John Smith.</p> <p>Cochez la case <b>Rapporter les candidats uniques</b> afin de rapporter les enregistrements avec un groupe correspondant depuis le port de candidat identifié comme enregistrements uniques.</p>

Option	Description
<b>Arrêter la comparaison du suspect avec les candidats après avoir trouvé <math>n</math> doublons</b>	<p>Cette option met en correspondance le suspect avec tous les candidats du même groupe de correspondance (groupe par option), mais interrompt la comparaison lorsque le nombre de doublons défini par l'utilisateur a été identifié. Par exemple, si vous avez choisi d'arrêter la comparaison de candidats après avoir trouvé un doublon et que vous disposiez des données suivantes :</p> <p>Suspect : John Smith  Candidat : Bill Jones  Candidat : John Smith  Candidat : John Smith</p> <p>Dans l'exemple, l'enregistrement suspect John Smith ne sera plus comparé dans le groupe de correspondance une fois que le premier candidat John Smith est identifié comme doublon.</p>

8. Cliquez sur **Générer des données pour l'analyse** pour générer des résultats de correspondance. Pour plus d'informations, voir [Analyse des résultats de correspondance](#) à la page 121.
9. L'option **Affecter le numéro de collection 0 aux enregistrements uniques**, sélectionnée par défaut, affectera le numéro de collection 0 aux enregistrements uniques. Désélectionnez cette option pour générer des numéros de collection autre que zéro pour les enregistrements uniques. Les numéros de collection d'enregistrement unique suivront tout autre numéro de collection. Par exemple, si votre flux de données de correspondance trouve cinq enregistrements et que les trois premiers sont uniques, les numéros de collection seront assignés comme indiqué dans le premier groupe ci-dessous. Si votre flux de données de correspondance trouve cinq enregistrements et que les deux derniers sont uniques, les numéros de collection seront assignés comme indiqué dans le second groupe ci-dessous.

Option	Description
<b>Numéro du groupe</b>	<b>Type d'enregistrement</b>
1	Unique
2	Unique
3	Unique
4	Doublon/Suspect

Option	Description
4	Doublon/Suspect
<b>Numéro du groupe</b>	<b>Type d'enregistrement</b>
1	Doublon/Suspect
1	Doublon/Suspect
2	Unique
3	Unique
4	Unique

Si vous laissez cette case cochée, tout enregistrement unique trouvé dans votre flux de données sera affecté à un numéro de collection de zéro par défaut.

10. Si vous créer une nouvelle règle personnalisée de correspondance, voir [Création d'une règle de correspondance](#) à la page 73 pour plus d'informations.
11. Cliquez sur **Évaluer** pour évaluer le score d'un enregistrement suspect par rapport aux enregistrements du candidat. Pour plus d'informations, voir [Interflow Match](#) à la page 223.

## Sortie

**Tableau 13 : Champs de sortie de Interflow Match**

Nom du champ	Description/Valeurs valides
CollectionNumber	Identifie une collection de doublons. Les valeurs possibles sont 1 et plus.
ExpressMatchIdentified	Indique si un rapprochement a été obtenu à l'aide de la clé de correspondance express. Les valeurs possibles sont Oui ou Non.
InterflowSourceType	Les valeurs possibles sont input_port_0 ou input_port_1

Nom du champ	Description/Valeurs valides
MatchRecordType	<p>Identifie le type d'enregistrement de correspondance dans une collection. Les valeurs possibles sont :</p> <p><b>suspect</b> L'enregistrement d'entrée d'origine qui a été marqué comme ayant peut-être des enregistrements doublons.</p> <p><b>duplicate</b> Un enregistrement qui est un doublon de l'enregistrement d'entrée.</p> <p><b>unique</b> Un enregistrement qui n'a aucun doublon.</p>
MatchScore	<p>Identifie le score global entre deux enregistrements. Les valeurs possibles sont 0-100, où 0 indique une faible correspondance et 100 indique une correspondance exacte.</p>

**Remarque :** Les stages de Validate Address et du module Advanced Matching utilisent tous les deux le champ MatchScore. La valeur du champ MatchScore dans la sortie d'un flux de données est déterminée par le dernier stage pour modifier la valeur avant qu'elle soit envoyée sur le stage de sortie. Si vous avez un flux de données qui contient les stages Validate Address et le module Advanced Matching et que vous souhaitez voir la sortie de champ MatchScore pour chaque stage, utilisez un stage Transformer pour copier la valeur MatchScore sur un autre champ. Par exemple, Validate Address produit un fichier de sortie appelé MatchScore, puis un stage Transformer copie le champ MatchScore à partir de Validate Address sur un champ appelé AddressMatchScore. Lorsque le stage de mise en correspondance fonctionne, il remplit le champ MatchScore avec la valeur de la mise en correspondance et passe par la valeur AddressMatchScore à partir de Validate Address.

## Intraflow Match

Intraflow Match identifie des rapprochements entre des enregistrements de données similaires à l'intérieur d'un seul flux d'entrée. Vous pouvez créer des règles hiérarchiques basées sur des champs que vous avez définis ou créés dans d'autres stages du flux de données.

### Options

1. Dans le champ **Charger une règle de rapprochement**, sélectionnez une des règles de correspondance prédéfinies que vous pouvez utiliser tel quel ou modifier pour répondre à vos besoins. Si vous souhaitez créer une nouvelle règle de correspondance sans utiliser une des règles de correspondance prédéfinies comme point de départ, cliquez sur **Nouveau**. Vous ne pouvez disposer que d'une règle personnalisée dans un flux de données.

**Remarque :** La fonction Options de flux de données dans Enterprise Designer permet d'afficher la règle de correspondance pour la configuration au moment de l'exécution.

2. Cliquez sur **Grouper par** pour sélectionner un champ à utiliser pour grouper les champs dans la file de correspondance. La Correspondance intra-flux n'essaie de faire correspondre que les enregistrements de la même file de correspondance.
3. Cochez la case **Trier** pour réaliser un tri de pré-rapprochement de votre entrée en fonction du champ sélectionné dans le champ **Grouper par**.
4. Cliquez sur **Avancé** pour indiquer d'autres options de performances du tri.

**Limite du nombre d'enregistrement en mémoire** Spécifie le nombre maximum de rangées de données qu'un trieur peut contenir en mémoire avant que celui-ci commence à pager sur le disque. Par défaut, un tri de 10 000 enregistrements ou moins sera effectué en mémoire et un tri de plus de 10 000 enregistrements sera effectué sur le disque. La limite maximale est de 100 000 enregistrements. En général, un tri en mémoire est beaucoup plus rapide qu'un tri sur le disque ; donc, il faut définir une valeur assez haute pour que la plupart des tris s'effectuent en mémoire et que seuls les groupes de grande taille soient écrits sur le disque.

**Remarque :** Soyez conscient du fait qu'au sein d'environnements où des jobs s'exécutent de manière simultanée, une augmentation du paramètre **Dans la limite d'enregistrement mémoire** augmente la probabilité de ne plus disposer de suffisamment de mémoire.

**Nombre maximal de fichiers temporaires** Spécifie le nombre maximal de fichiers temporaires pouvant être employés par un processus de tri. L'utilisation d'un plus grand nombre de fichiers temporaires peut améliorer les performances. Cependant, le nombre optimal dépend très largement de la configuration du serveur qui exécute Spectrum™ Technology Platform. Nous vous conseillons d'essayer différents paramètres et d'observer l'effet de l'utilisation d'un plus ou moins grand nombre de fichiers temporaires sur les performances. Pour calculer le nombre approximatif de fichiers temporaires nécessaires, utilisez l'équation suivante :

$$\frac{(NumberOfRecords \times 2)}{InMemoryRecordLimit} = NumberOfTempFiles$$

Notez que le nombre maximal de fichiers temporaires ne peut pas être supérieure à 1 000.

**Activer la compression** Spécifie si les fichiers temporaires sont compressés lors de leur écriture sur le disque.

**Remarque :** Les paramètres de performances de tri optimal dépendent de la configuration matérielle de votre serveur. Néanmoins, l'équation suivante produit généralement de bonnes performances de tri :

$$(InMemoryRecordLimit \times MaxNumberOfTempFiles \div 2) \geq TotalNumberOfRecords$$

5. Cliquez sur **Rapprochement express sur** pour effectuer une comparaison initiale des valeurs de clé express pour déterminer si les deux enregistrements sont considérés comme correspondants.

Vous pouvez créer une clé express en créant une clé de correspondance dans le MatchKeyGenerator. Pour plus d'informations, reportez-vous à la section [Match Key Generator](#) à la page 234.

6. Dans la zone de texte **Numéro de collection initial**, indiquez le numéro de départ à assigner au champ du numéro de collection pour les enregistrements doublons.

Le numéro de collection identifie chaque enregistrement doublon dans une file de correspondance. Les enregistrements uniques se voient assigner un numéro de collection de 0. Chaque enregistrement doublon se voit assigner un numéro de collection débutant pas la valeur indiquée dans l'encadré **Numéro de collection initial**.

7. Cliquez sur **Fenêtre coulissante** pour activer cette méthode de correspondance. Pour plus d'informations sur la fenêtre coulissante, voir [Méthode de correspondance de fenêtre coulissante](#) à la page 233
8. Cliquez sur **Générer des données pour l'analyse** pour générer des résultats de correspondance. Pour plus d'informations, voir [Analyse des résultats de correspondance](#) à la page 121.
9. L'option **Affecter le numéro de collection 0 aux enregistrements uniques**, sélectionnée par défaut, affectera le numéro de collection 0 aux enregistrements uniques. Désélectionnez cette option pour générer des numéros de collection autre que zéro pour les enregistrements uniques. Les numéros de collection d'enregistrement unique suivront tout autre numéro de collection. Par exemple, si votre flux de données de correspondance trouve cinq enregistrements et que les trois premiers sont uniques, les numéros de collection seront assignés comme indiqué dans le premier groupe ci-dessous. Si votre flux de données de correspondance trouve cinq enregistrements et que les deux derniers sont uniques, les numéros de collection seront assignés comme indiqué dans le second groupe ci-dessous.

Option	Description
<b>Numéro du groupe</b>	<b>Type d'enregistrement</b>
1	Unique
2	Unique

Option	Description
3	Unique
4	Doublon/Suspect
4	Doublon/Suspect
Numéro du groupe	Type d'enregistrement
1	Doublon/Suspect
1	Doublon/Suspect
2	Unique
3	Unique
4	Unique

Si vous laissez cette case cochée, tout enregistrement unique trouvé dans votre flux de données sera affecté à un numéro de collection de zéro par défaut.

10. Pour plus d'informations sur la modification des autres options, voir [Création d'une règle de correspondance](#) à la page 73.
11. Cliquez sur **Évaluer** pour évaluer le score d'un enregistrement suspect par rapport aux enregistrements du candidat. Pour plus d'informations, voir [Interflow Match](#) à la page 223.

### Méthode de rapprochement par défaut

À l'aide de l'option grouper par (groupe de doublons) définie par l'utilisateur, l'analyseur de rapprochement identifie les groupes d'enregistrements pouvant potentiellement être des doublons les uns des autres. L'analyseur de rapprochement analyse ensuite chaque enregistrement du groupe ; si l'enregistrement correspond à un Suspect existant, l'enregistrement est considéré Doublon de ce suspect, reçoit un score, un CollectionNumber, et MatchRecordType (Doublon), et il est éliminé de la correspondance. Si, par contre, l'enregistrement ne correspond à aucun Suspect au sein du groupe, l'enregistrement devient un nouveau Suspect, et en tant que tel il est ajouté au Groupe d'enregistrements actuel de manière à être mis en correspondance avec les enregistrements suivants. Quand l'analyseur de correspondance a épuisé tous les enregistrements dans le groupe de doublons actuel, il élimine tous les Suspects de la correspondance, en marquant leur type d'enregistrement de correspondance comme Unique et attribuant un numéro de collection de 0.



Ces Suspects ayant au moins un doublon conservent un type d'enregistrement de correspondance de Suspect et reçoivent le même numéro de collection que leur enregistrement doublon correspondant. Enfin, quand tous les enregistrements d'un groupe de doublons ont été écrits en résultat. Un nouveau groupe de correspondance est comparé.

**Remarque :** La Méthode de rapprochement par défaut ne comparera que les enregistrements qui sont dans le même groupe de doublons.

Le type de rapprochement (Intraflow ou Interflow) détermine la manière selon laquelle les résultats de rapprochement d'une clé express se traduisent en scores de rapprochements candidats (Candidate Match Scores). Lors d'un rapprochement de type Interflow, un rapprochement à l'aide d'une clé de rapprochement express réussi donne toujours un score de rapprochement (MatchScore) de 100 au Candidat. En revanche, lors d'un rapprochement de type Intraflow, le score que gagne le Candidat à l'issue d'un rapprochement grâce à une clé de rapprochement express dépend de l'éventualité selon laquelle l'enregistrement avec lequel ce Candidat a été rapproché était le rapprochement d'un autre Suspect : les doublons d'une clé de rapprochement express d'un Suspect auront toujours des scores de rapprochements de 100, alors que les doublons d'une clé de rapprochement express d'un autre Candidat (qui était le doublon d'un Suspect) hériteront du score de rapprochement de ce Candidat (pas nécessairement de 100).

### Méthode de correspondance de fenêtre coulissante

L'algorithme fenêtre coulissante remplit de manière séquentielle une taille de mémoire tampon prédéfinie, qui s'appelle une fenêtre, à l'aide de la quantité de rangées de données correspondante. À mesure que chaque rangée s'ajoute à la fenêtre, celle-ci est comparée à chaque élément déjà contenu dans la fenêtre. Si un rapprochement avec un élément est déterminé, l'enregistrement pilote (le nouvel élément à ajouter à la fenêtre) et les candidats (éléments déjà dans la fenêtre) reçoivent tous le même identifiant de groupe. Cette comparaison continue jusqu'à ce que l'enregistrement pilote ait été comparé à tous les éléments contenus dans la fenêtre.

À mesure que de nouveaux pilotes sont ajoutés, la fenêtre atteindra sa capacité prédéterminée. À ce moment, la fenêtre coulissera, d'où le terme de fenêtre coulissante. Coulisser signifie tout simplement que la mémoire tampon de la fenêtre va supprimer et écrire l'élément le plus ancien de la fenêtre au fur et à mesure qu'elle ajoutera un nouvel enregistrement pilote dans la fenêtre.

### Sortie

**Tableau 14 : Sortie de Correspondance intra-flux**

Nom du champ	Description/Valeurs valides
CollectionNumber	Identifie une collection de doublons. Les valeurs possibles sont 1 et plus.

Nom du champ	Description/Valeurs valides
ExpressMatchIdentified	Indique si un rapprochement a été obtenu à l'aide de la clé de correspondance express. Les valeurs possibles sont Oui ou Non.
MatchRecordType	Identifie le type d'enregistrement de correspondance dans une collection. Les valeurs possibles sont : <ul style="list-style-type: none"> <li><b>suspect</b> Un enregistrement auquel les autres enregistrements sont comparés afin de déterminer s'ils sont des doublons les uns des autres. Chaque collection possède un seul enregistrement suspect.</li> <li><b>duplicate</b> Un enregistrement qui est un doublon de l'enregistrement suspect.</li> <li><b>unique</b> Un enregistrement qui n'a aucun doublon.</li> </ul>
MatchScore	Identifie le score global entre deux enregistrements. Les valeurs possibles sont 0-100, où 0 indique une faible correspondance et 100 indique une correspondance exacte.

**Remarque :** Les stages de Validate Address et du module Advanced Matching utilisent tous les deux le champ MatchScore. La valeur du champ MatchScore dans la sortie d'un flux de données est déterminée par le dernier stage pour modifier la valeur avant qu'elle soit envoyée sur le stage de sortie. Si vous avez un flux de données qui contient les stages Validate Address et le module Advanced Matching et que vous souhaitez voir la sortie de champ MatchScore pour chaque stage, utilisez un stage Transformer pour copier la valeur MatchScore sur un autre champ. Par exemple, Validate Address produit un fichier de sortie appelé MatchScore, puis un stage Transformer copie le champ MatchScore à partir de Validate Address sur un champ appelé AddressMatchScore. Lorsque le stage de mise en correspondance fonctionne, il remplit le champ MatchScore avec la valeur de la mise en correspondance et passe par la valeur AddressMatchScore à partir de Validate Address.

## Match Key Generator

Match Key Generator crée une clé non unique pour chaque enregistrement, qui peut ensuite être utilisée par les stages de rapprochement pour identifier les groupes d'enregistrements doublons potentiels. Les match keys facilitent la procédure de correspondance en vous permettant de regrouper les enregistrements par match key, puis de ne comparer les enregistrements que dans ces groupes.

La match key est créée à l'aide de règles que vous définissez et est constituée de champs d'entrée. Un algorithme sélectionné s'applique sur chaque champ d'entrée indiqué. Le résultat de chaque algorithme est ensuite concaténé pour créer un seul champ de match key.

Outre la création de match keys, vous pouvez également créer des match keys express, à utiliser ultérieurement dans le dataflow par un stage Intraflow Match ou un stage Interflow Match.

Vous pouvez créer plusieurs match keys et match keys express.

Par exemple, si l'enregistrement entrant est :

Prénom : Fred Nom de famille : Mertz Code postal : 21114-1687 Code de genre : M

Et que vous définissez une règle de match key qui génère une match key en combinant des données de l'enregistrement comme suit :

Champ de saisie	Position de début	Longueur
Code postal	1	5
Code postal	7	4
Nom	1	5
Prénom	1	5
Code de genre	1	1

Alors, la clé serait :

211141687MertzFredM

## Entrée

L'entrée représente tout champ dans les données source.

## Options

Pour définir les options de Match Key Generator, cliquez sur le bouton **Ajouter**. La boîte de dialogue **Match Key Field** s'affiche.

**Remarque :** La fonction Options de flux de données dans Enterprise Designer permet d'afficher Match Key Generator pour la configuration au moment de l'exécution.

**Tableau 15 : Options de Match Key Generator**

Nom de l'option	Description/Valeurs valides
-----------------	-----------------------------

---

Algorithme	
------------	--

## Nom de l'option

## Description/Valeurs valides

Définit l'algorithme à utiliser pour générer la clé de correspondance. L'un des éléments suivants :

- Consonne** Renvoie les champs indiqués, les consonnes étant supprimées.
- Metaphone double** Renvoie un code basé sur la représentation phonétique de leurs caractères. Le double Metaphone est une version améliorée de l'algorithme Metaphone et tente de prendre en compte les nombreuses irrégularités de plusieurs langues.
- Koeln** Noms d'index par son, tels qu'ils sont prononcés en allemand. Permet aux noms ayant la même prononciation d'être encodés avec la même représentation afin qu'ils puissent être mis en correspondance, en dépit de différences mineures au niveau de l'orthographe. Le résultat est toujours une séquence de nombres ; les caractères spéciaux et les espaces blancs sont ignorés. Cette option a été développée en réponse aux limites du Soundex.
- MD5** Algorithme qui produit une valeur hash de 128 bits. Cet algorithme est généralement utilisé pour vérifier l'intégrité des données.
- Metaphone** Renvoie une clé codée Metaphone des champs sélectionnés. Metaphone est un algorithme qui code les mots à l'aide de leur sonorité lorsque prononcé en anglais.
- Metaphone (Espagnol)** Renvoie une clé codée Metaphone des champs sélectionnés pour la langue espagnole. Cet algorithme Metaphone code les mots à l'aide de leur sonorité lorsque prononcé en espagnol.
- Metaphone3** Procède à une amélioration en fonction des algorithmes Metaphone et Double Metaphone avec des paramètres de consonne et de voyelle interne exacts qui vous permet de produire des mots ou des noms mis en correspondance de manière plus ou moins proche pour rechercher des termes au niveau phonétique. Metaphone 3 augmente l'exactitude de l'encodage phonétique à 98 %. Cette option a été développée en réponse aux limites du Soundex.
- Nysiis** L'algorithme de code phonétique qui met en correspondance une prononciation approximative avec une orthographe exacte et indexe des mots prononcés de manière similaire. Fait partie du système New York State Identification and Intelligence System. Imaginons, par exemple, que vous recherchez des informations sur une personne dans une base de données de personnes. Vous pensez que le nom de la personne sonne comme « John Smith », mais il est en fait orthographié « Jon Smyth ». Si vous procédez à une recherche de la correspondance exacte de « John Smith », aucun résultat n'est renvoyé. Cependant, si vous indexez la base de données à l'aide de l'algorithme NYSIIS et procédez à une recherche en utilisant de nouveau l'algorithme NYSIIS, la correspondance correcte est renvoyée car « John Smith » et « Jon Smyth » sont indexés comme « JAN SNATH » par l'algorithme.

Nom de l'option	Description/Valeurs valides
	<p><b>Phonix</b> Pré-traite les chaînes de nom en appliquant plus de 100 règles de transformation à des caractères uniques ou à des séquences de plusieurs caractères. 19 de ces règles s'appliquent uniquement si les caractères figurent au début de la chaîne, tandis que 12 des règles s'appliquent uniquement si les caractères figurent au milieu de la chaîne et 28 des règles s'appliquent uniquement si les caractères figurent à la fin de la chaîne. La chaîne de nom transformée est cryptée en un code composé d'une lettre au début, suivie de trois chiffres (en enlevant les zéros et les nombres en double). Cette option a été développée pour répondre aux limites de Soundex ; elle est plus complexe et donc plus lente que Soundex.</p> <p><b>Soundex</b> Renvoie un code Soundex des champs sélectionnés. Soundex produit un code de longueur fixe en s'appuyant sur la sonorité du mot lorsque prononcé en anglais.</p> <p><b>Substring</b> Renvoie une partie spécifiée du champ sélectionné.</p>
Nom de champ	Indique le champ auquel vous souhaitez appliquer l'algorithme sélectionné pour générer la match key. Par exemple, si vous sélectionnez un champ appelé LastName et que vous choisissez l'algorithme Soundex, ce dernier est appliqué aux données dans le champ LastName pour produire une match key.
Position de début	Spécifie la position de départ dans le champ spécifié. Tous les algorithmes ne permettent pas de spécifier une position de départ.
Longueur	Spécifie la longueur de caractère à inclure à partir de la position de départ. Tous les algorithmes ne permettent pas de spécifier une longueur.
Supprimer les caractères parasites :	Supprime tout caractère non numérique et non alphanumérique, comme les traits d'union, les espaces blancs, et autres caractères spéciaux du champ d'entrée.
Trier les entrées :	Trie tous les caractères dans un champ d'entrée ou tous les termes dans un champ d'entrée par ordre alphabétique. <p><b>Caractères</b> Trie les valeurs de caractères d'un champ d'entrée avant de créer un identifiant unique.</p> <p><b>Termes</b> Trie les valeurs de termes d'un champ d'entrée avant de créer un identifiant unique.</p>

Si vous ajoutez plusieurs algorithmes de génération de clé de correspondance, vous pouvez utiliser les boutons **Monter** et **Abaisser** afin de changer l'ordre selon lequel les algorithmes sont appliqués.

### Génération d'une Match Key Express

Activez l'option **Créer une Match Key Express** et cliquez sur **Ajouter** pour définir une match key express à utiliser ultérieurement dans le flux de données par un stage Intraflow Match ou un stage Interflow Match.

Si l'option **Créer une Match Key Express** est activée et que l'option **Match Key Express sur** est sélectionnée dans un stage Interflow Match en aval ou dans un stage Intraflow Match, une première tentative de rapprochement est effectuée à l'aide de la match key express créée ici. Si les match keys express de deux enregistrements sont rapprochées, l'enregistrement est considéré comme un rapprochement et aucun autre traitement n'est effectué. Si les match keys express des enregistrements ne sont pas rapprochés, les règles de correspondance définies dans Interflow Match ou Intraflow Match sont utilisées pour déterminer si les enregistrements correspondent.

## Sortie

**Tableau 16 : Sortie de Match Key Generator**

Nom du champ	Description/Valeurs valides
ExpressMatchKey	Une valeur indiquant le niveau de correspondance. Si la match key express est une correspondance, le score est 100. Si la match key express n'est pas une correspondance, un score de 0 sera renvoyé.
MatchKey	La clé générée pour identifier les enregistrements.

## Private Match

Private Match permet à deux entités de comparer des jeux de données et d'identifier les enregistrements communs sans compromettre les informations sensibles. Par exemple, deux sociétés peuvent souhaiter lancer une campagne marketing conjointe. Chaque société comporte sa propre base de données contenant des informations client, et les deux sociétés souhaitent déterminer les clients qui font des achats dans les deux sociétés pour adopter une approche plus ciblée au sein de la campagne. En revanche, pour garantir la sécurité des données et respecter la réglementation en matière de confidentialité, elles ne souhaitent pas partager ces bases de données l'une avec l'autre ni les fournir à un tiers pour réaliser une correspondance. La fonctionnalité Private Match permet d'effectuer une correspondance entre les deux bases de données sans violation de la sécurité ni infraction des lois en matière de confidentialité.

Private Match est utilisé dans l'un des trois modes suivants :

- **Mode Encrypt** : le premier utilisateur saisit ses données et un champ d'index ainsi qu'un champ de correspondance sont extraits et chiffrés. Une clé publique et une table de déplacement contenant les données du premier utilisateur sont générées pour le deuxième utilisateur, et une clé privée est générée pour le premier utilisateur, pour un usage ultérieur.
- **Mode Private Match** : le deuxième utilisateur saisit ses données et les données chiffrées du premier utilisateur, fournit la clé publique et la table de déplacement, et effectue une correspondance. Un fichier contenant les données mises en correspondance est généré et envoyé au premier utilisateur.
- **Mode Decrypt** : le premier utilisateur saisit les données chiffrées du deuxième utilisateur, fournit la clé privée et génère la sortie contenant un index mis en correspondance des données des deux utilisateurs.

Grâce à la fonction de chiffrement (mode Encrypt), la sécurité est maintenue pendant que la fonction de correspondance est exécutée (mode Private Match), puis une fonction de déchiffrement affiche la sortie des données mises en correspondance (mode Decrypt). L'ensemble des fichiers générés et partagés entre les utilisateurs sont chiffrés et illisibles.

## Entrée

Les exigences en entrée du stage Private Match varient suivant la tâche que vous effectuez :

- **Mode Encrypt** : un fichier contenant les données du premier utilisateur doit être joint au port d'entrée. Il peut s'agir d'un fichier texte, d'une base de données ou de presque tout type de fichier source.
- **Mode Private Match** : un fichier contenant les données du deuxième utilisateur doit être joint au premier port d'entrée ; il peut également s'agir d'un fichier texte, d'une base de données ou de presque tout type de fichier source. Les données chiffrées générées par le premier utilisateur doivent être jointes au deuxième port d'entrée.
- **Mode Decrypt** : le fichier de sortie généré par le deuxième utilisateur.

## Options

Les options du stage Private Match varient suivant la tâche que vous effectuez.

### *Mode Encrypt*

1. Sélectionnez l'opération **Encrypt**.
2. Sélectionnez le **champ d'index** fournissant un ID unique à chaque enregistrement du fichier.
3. Sélectionnez le **champ de correspondance** à utiliser pour effectuer une correspondance avec les données du deuxième utilisateur.
4. Indiquez le chemin d'accès et le nom du fichier de **clé publique** à créer lors de l'exécution du job.
5. Indiquez le chemin d'accès et le nom du fichier de **clé privée** à créer lors de l'exécution du job.
6. Indiquez le chemin d'accès et le nom du fichier de **table de déplacement** à créer lors de l'exécution du job.



7. Saisissez un nom pour la **colonne de sortie** qui contiendra les données chiffrées du fichier de sortie envoyé au deuxième utilisateur.
8. Appuyez sur **OK**.

#### *Mode Private Match*

1. Sélectionnez l'opération **Private Match**.
2. Sélectionnez le **champ d'index** fournissant un ID unique à chaque enregistrement du fichier.
3. Sélectionnez le **champ de correspondance** à utiliser pour effectuer une correspondance avec les données du premier utilisateur.
4. Saisissez le nom du **champ de données chiffrées** que le premier utilisateur a saisi à l'étape 4 des instructions du mode Encrypt.
5. Accédez au fichier de **clé publique**.
6. Accédez au fichier de **table de déplacement**.
7. Saisissez un nom pour la **colonne de sortie** qui contiendra les données chiffrées du fichier de sortie envoyé au premier utilisateur.

#### *Mode Decrypt*

1. Sélectionnez l'opération **Decrypt**.
2. Sélectionnez le **champ d'index** fournissant un ID unique à chaque enregistrement du fichier.
3. Accédez au fichier de **clé privée**.
4. Sélectionnez la **colonne de sortie** qui contiendra les données déchiffrées du fichier de sortie. Le format des données de ce champ est l'index mis en correspondance des données du premier utilisateur et l'index mis en correspondance des données du deuxième utilisateur, séparés par une barre verticale (|), comme suit :

User1Data|User2Data

### **Sortie**

Les exigences en sortie du stage Private Match varient suivant la tâche que vous effectuez :

- Mode Encrypt : un fichier généré par le stage Write to File contenant les données chiffrées du premier utilisateur.
- Mode Private Match : un fichier généré par le stage Write to File contenant les informations chiffrées sur les résultats de correspondance.
- Mode Decrypt : un fichier généré par le stage Write to File contenant l'index mis en correspondance des données des deux utilisateurs.

## Transactional Match

Transactional Match met en correspondance les enregistrements suspects par rapport aux enregistrements candidats renvoyés par le stage Candidate Finder. Transactional Match utilise des

règles de correspondance pour comparer l'enregistrement suspect à tous les enregistrements candidats possédant le même numéro de groupe de candidats (attribués dans Candidate Finder) pour identifier les doublons. Si l'enregistrement candidat est un doublon, un numéro de groupe lui est attribué, le type d'enregistrement de rapprochement est étiqueté Doublon, et l'enregistrement est ensuite écrasé. Tous candidats sans rapprochement dans le groupe reçoit un numéro de groupe de 0, est étiqueté Unique, puis écrasé lui-aussi.

**Remarque :** Transactional Match ne met en correspondance les enregistrements suspects qu'avec les candidats. Il ne tente pas de mettre en correspondance les enregistrements suspects avec d'autres enregistrements suspects, comme cela est le cas dans Intraflow Match.

## Options

1. Dans le champ **Charger une règle de rapprochement**, sélectionnez une des règles de correspondance prédéfinies que vous pouvez utiliser tel quel ou modifier pour répondre à vos besoins. Si vous souhaitez créer une nouvelle règle de correspondance sans utiliser une des règles de correspondance prédéfinies comme point de départ, cliquez sur **Nouveau**. Vous ne pouvez disposer que d'une règle personnalisée dans un flux de données.

**Remarque :** La fonction Options de flux de données dans Enterprise Designer permet d'afficher la règle de correspondance pour la configuration au moment de l'exécution.

2. Sélectionnez **Renvoyer les candidats uniques** si vous souhaitez que les enregistrements candidats uniques soient inclus dans la sortie du stage.
3. Sélectionnez **Générer les données pour analyse** si vous souhaitez utiliser l'outil Analyse des rapprochements pour analyser les résultats du flux de données. Pour plus d'informations, voir [Analyse des résultats de correspondance](#) à la page 121.
4. Pour plus d'informations sur la modification des autres options, voir [Création d'une règle de correspondance](#) à la page 73.
5. Cliquez sur **Évaluer** pour évaluer le score d'un enregistrement suspect par rapport aux enregistrements du candidat. Pour plus d'informations, voir [Interflow Match](#) à la page 223.

## Sortie

Tableau 17 : Transactional Match sortie

Nom du champ	Description/Valeurs valides
HasDuplicates	<p>Identifie si l'enregistrement est un doublon d'un autre enregistrement. L'un des éléments suivants :</p> <p><b>Y</b> L'enregistrement est un enregistrement suspect et possède des doublons.</p> <p><b>N</b> L'enregistrement est un enregistrement suspect et ne possède pas de doublons.</p> <p><b>D</b> L'enregistrement est un enregistrement candidat et est un doublon de l'enregistrement suspect.</p> <p><b>U</b> L'enregistrement est un enregistrement candidat mais n'est pas un doublon de l'enregistrement suspect.</p>
MatchRecordType	<p>Identifie le type d'enregistrement de correspondance dans une collection. Les valeurs possibles sont :</p> <p><b>Suspect</b> L'enregistrement d'entrée d'origine qui a été marqué comme ayant peut-être des enregistrements doublons.</p> <p><b>Duplicate</b> Un enregistrement qui est un doublon de l'enregistrement d'entrée.</p> <p><b>Unique</b> Un enregistrement qui n'a aucun doublon.</p>
MatchScore	<p>Identifie le score global entre deux enregistrements. Les valeurs possibles sont 0-100, où 0 indique une faible correspondance et 100 indique une correspondance exacte.</p>
MatchInfo.<RuleName>.IsMatch	<p>Affiche l'état de correspondance de la règle sous forme de valeur <code>True</code> ou <code>False</code>, où <code>True</code> représente une correspondance et <code>False</code> indique aucune correspondance. Ces informations s'affichent si vous sélectionnez l'option <b>Envoyer les informations de correspondance détaillées</b> sur l'écran <b>Transactional Match Options</b>.</p> <p><b>Remarque</b> : La valeur de ce champ est dérivée de l'état de correspondance des nœuds enfants en fonction de la configuration de règle parente définie.</p>

Nom du champ	Description/Valeurs valides
MatchInfo.<RuleName>.Score	<p>Affiche le score de correspondance de la règle. Les valeurs possibles sont 0-100, où 0 indique une faible correspondance et 100 une correspondance parfaite. Ces informations s'affichent si vous sélectionnez l'option <b>Renvoyer les informations de correspondance détaillées</b> sur l'écran <b>Transactional Match Options</b>.</p> <p><b>Remarque :</b> Le score de ce champ est dérivé des scores des nœuds enfants en fonction de la configuration de règle parente définie.</p>
MatchInfo.<RuleName>.<RuleNodeName>.IsMatch	<p>Affiche l'état de correspondance de chaque nœud de la hiérarchie de règles sous forme de valeur <code>True</code> ou <code>False</code>, où <code>True</code> représente une correspondance et <code>False</code> indique aucune correspondance. <code>RuleNodeName</code> est une variable, remplacée par les noms de nœud de hiérarchie dans vos règles de rapprochement.</p> <p>Chaque nœud de la hiérarchie de règles émet ce champ, si vous avez sélectionné l'option <b>Renvoyer les informations de correspondance détaillées</b> sur l'écran <b>Transactional Match Options</b>.</p>
MatchInfo.<RuleName>.<RuleNodeName>.Score	<p>Affiche le score de rapprochement de chaque nœud de la hiérarchie de règles. <code>RuleNodeName</code> est une variable, remplacée par les noms de nœud de hiérarchie dans vos règles de rapprochement.</p> <p>Chaque nœud de la hiérarchie de règles émet ce champ, si vous avez sélectionné l'option <b>Renvoyer les informations de correspondance détaillées</b> sur l'écran <b>Transactional Match Options</b>.</p> <p>Les valeurs possibles sont 0-100, où 0 indique une faible correspondance et 100 une correspondance parfaite.</p>

**Remarque :** Les stages de Validate Address et du module Advanced Matching utilisent tous les deux le champ MatchScore. La valeur du champ MatchScore dans la sortie d'un flux de données est déterminée par le dernier stage pour modifier la valeur avant qu'elle soit envoyée sur le stage de sortie. Si vous avez un flux de données qui contient les stages Validate Address et le module Advanced Matching et que vous souhaitez voir la sortie de champ MatchScore pour chaque stage, utilisez un stage Transformer pour copier la valeur MatchScore sur un autre champ. Par exemple, Validate Address produit un fichier de sortie appelé MatchScore, puis un stage Transformer copie le champ MatchScore à partir de Validate Address sur un champ appelé AddressMatchScore. Lorsque le stage de mise en

correspondance fonctionne, il remplit le champ MatchScore avec la valeur de la mise en correspondance et passe par la valeur AddressMatchScore à partir de Validate Address.

## Write to Search Index

Write to Search Index vous permet de créer un index plein texte basé sur les données arrivant dans le stage. Disposer de ces données dans un index de recherche dédié permet d'obtenir un temps de réponse plus rapide lorsque vous effectuez des recherches par rapport à l'index des autres stages Spectrum™ Technology Platform. Les index de recherche plein texte sont préférables aux bases de données relationnelles si vous disposez d'un grand nombre de données textuelles de format libre qui doivent faire l'objet de recherches ou être catégorisées ou si vous prenez en charge un gros volume de requêtes interactives basées sur du texte.

Write to Search Index utilise un outil d'analyse pour diviser le texte d'entrée en de petits éléments d'indexation, appelés jetons. Ensuite, il extrait les termes d'index de recherche de ces jetons. Le type d'outil d'analyse utilisé (la manière dont le texte d'entrée est divisé en jetons) détermine la façon dont vous pourrez ensuite rechercher ce texte. Par exemple, l'analyseur de *mots clés* effectue toujours une correspondance exacte et transforme la chaîne toute entière en un seul jeton, tandis que l'analyseur *standard* rompt la chaîne pour créer des jetons. L'analyseur *Mot d'arrêt* est beaucoup plus compliqué. Il supprime les articles tels que « a » ou « le » de la chaîne avant la transformation en jetons, réduisant ainsi la taille d'index.

Les index de recherche prennent en charge la fonction presque en temps réel, ce qui permet de mettre à jour les index presque immédiatement, sans besoin de fermer et de recréer le stage avec l'index de recherche.

Pour plus d'informations sur les tâches d'index de recherche que vous pouvez effectuer via la ligne de commande, reportez-vous à Index de recherche de la section [Utilitaire d'administration](#) du Guide d'administration.

### Options

1. Dans Enterprise Designer, double-cliquez sur le stage Write to Search Index sur le canevas.
2. Saisissez le **Nom** de l'index.
3. Sélectionnez un **Mode d'écriture**. Lorsque vous régénérez un index, vous disposez d'options liées à la manière dont les nouvelles données doivent affecter les données existantes.
  - **Create or Overwrite** : les nouvelles données remplaceront les données existantes et celles-ci ne figureront plus dans l'index.
  - **Update or Append** : les nouvelles données remplaceront les données existantes et toute nouvelle donnée qui n'existait pas auparavant sera ajoutée à l'index.
  - **Append** : les nouvelles données seront ajoutées aux données existantes et celles-ci resteront intactes.
  - **Delete** : les données du champ sélectionné seront supprimées de l'index de recherche.

4. Sélectionnez le **Champ clé** en fonction duquel vous souhaitez effectuer des opérations **Update or Append** ou **Delete** sur les enregistrements.

- En mode **Create ou Overwrite**, le **Champ clé** doit être unique pour les index de recherche Elastic (utilisés dans un environnement distribué). Si vous laissez le champ vide, tous les enregistrements sont stockés dans l'index, quels que soient les doublons. Cependant, vous ne pouvez effectuer aucune opération d'écriture, telle qu'une mise à jour, une modification ou une suppression, sur cet index. Le tableau suivant explique le comportement d'indexation si le **Champ clé** n'est pas unique pour les index de recherche Lucene et Elastic.

Mode d'écriture	Champ clé	Index de recherche Lucene	Index de recherche Elastic
Create or Overwrite	Enregistrements doublons avec le même <b>Champ clé</b>	Tous les enregistrements sont stockés.  <b>Remarque :</b> Les enregistrements doublons avec le même champ clé sont écrasés dès que vous exécutez l'opération de mise à jour.	Tous les enregistrements doublons avec le même champ clé sont <b>écrasés</b> .
Mettre à jour ou ajouter	Enregistrements doublons avec le même <b>Champ clé</b>	Les doublons sont remplacés.	Les doublons sont remplacés.

5. Cochez la case **Validation par lot** si vous souhaitez indiquer le nombre d'enregistrements à valider dans un lot lors de la création de l'index de recherche. Saisissez ensuite ce nombre dans le champ **Taille de lot**. La valeur par défaut est 5000.

6. Sélectionnez un **Outil d'analyse** pour créer les éléments suivants :

- **Standard** : fournit un segmenteur basé sur la grammaire qui contient un super ensemble d'analyseurs d'espace et de mot d'arrêt. Comprend la ponctuation anglaise pour segmenter les mots, connaît les mots à ignorer (via l'analyseur de mot d'arrêt), et effectue une recherche techniquement insensible à la casse en procédant à des comparaisons de minuscules. Par exemple, la chaîne « Pitney Bowes Software » serait renvoyée sous forme de trois jetons : « pitney », « bowes » et « software ». Pour une comparaison entre les analyseurs standard et de mots clés, reportez-vous à la section [Analyseurs standard et de mots clés](#) à la page 249.
- **Espace** : sépare les jetons par des espaces. Sorte de sous-ensemble de l'analyseur standard dans la mesure où il comprend les segmentations de mots en anglais, en fonction des espaces et des retours à la ligne.

- Mot d'arrêt : supprime les articles, tels que "the," "and" et "a" pour réduire la taille de l'index et augmenter les performances.
- Mot clé : crée un jeton unique à partir d'un dataflow et le conserve tel quel. Par exemple, la chaîne « Pitney Bowes Software » serait renvoyée sous forme d'un jeton : « Pitney Bowes Software ». Pour une comparaison entre les analyseurs standard et de mots clés, reportez-vous à la section [Analyseurs standard et de mots clés](#) à la page 249.
- Russe : prend en charge les index et les services de saisie anticipée en langue russe. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « et », « je » et « vous » pour réduire la taille de l'index et améliorer les performances.
- Allemand : prend en charge les index et les services de saisie anticipée en langue allemande. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « le », « et » et « un » pour réduire la taille de l'index et améliorer les performances.
- Danois : prend en charge les index et les services de saisie anticipée en langue danoise. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « à », « et » et « un » pour réduire la taille de l'index et améliorer les performances.
- Néerlandais : prend en charge les index et les services de saisie anticipée en langue néerlandaise. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « le », « et » et « un » pour réduire la taille de l'index et améliorer les performances.
- Finnois : prend en charge les index et les services de saisie anticipée en langue finnoise. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « est », « et » et « de » pour réduire la taille de l'index et améliorer les performances.
- Français : prend en charge les index et les services de saisie anticipée en langue française. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « le », « et » et « un » pour réduire la taille de l'index et améliorer les performances.
- Hongrois : prend en charge les index et les services de saisie anticipée en langue hongroise. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « le », « et » et « un » pour réduire la taille de l'index et améliorer les performances.
- Italien : prend en charge les index et les services de saisie anticipée en langue italienne. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « le », « et » et « un » pour réduire la taille de l'index et améliorer les performances.
- Norvégien : prend en charge les index et les services de saisie anticipée en langue norvégienne. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « le », « et » et « un » pour réduire la taille de l'index et améliorer les performances.
- Portugais : prend en charge les index et les services de saisie anticipée en langue portugaise. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « le », « et » et « un » pour réduire la taille de l'index et améliorer les performances.
- Espagnol : prend en charge les index et les services de saisie anticipée en langue espagnole. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « le », « et » et « un » pour réduire la taille de l'index et améliorer les performances.
- Suédois : prend en charge les index et les services de saisie anticipée en langue suédoise. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « le », « et » et « un » pour réduire la taille de l'index et améliorer les performances.

- Hindi : prend en charge les index et les services de saisie anticipée en langue hindi. Prend également en charge de nombreux mots d'arrêt et supprime les articles tels que « par », « et » et « un » pour réduire la taille de l'index et améliorer les performances.
7. Pour mettre à jour l'analyseur de tous les champs présents dans la liste, sélectionnez un analyseur dans la liste déroulante **Mettre les analyseurs à jour à...**
  8. Pour recharger le schéma à partir du serveur, cliquez sur **Recharger le schéma**.
 

**Remarque** : Vous pouvez modifier le nom du champ en saisissant le nouveau nom directement dans la colonne **Champs**. Cependant, vous ne pouvez pas modifier le nom **Champs de stage**, ni le **Type**.
  9. Pour ajouter/supprimer des champs de votre source d'entrée de manière sélective, cliquez sur **Ajout rapide**. La fenêtre contextuelle **Ajout rapide** affiche une liste de tous les champs de la source d'entrée. Sélectionnez les champs que vous souhaitez ajouter et cliquez sur **OK**.
  10. Sélectionnez le(s) champ(s) dont vous souhaitez stocker les données. Par exemple, à l'aide d'un fichier d'entrée d'adresses, vous pouvez indexer uniquement le code postal et choisir de stocker les champs restants (tels que Address Line 1, City, State), de telle sorte que l'ensemble de l'adresse soit renvoyé lorsqu'une correspondance est détectée via la recherche d'index.
  11. Sélectionnez le(s) champ(s) dont vous souhaitez ajouter les données à l'index pour une requête de recherche.
 

**Remarque** : Si vous voulez supprimer certains champs, sélectionnez-les et cliquez sur **Supprimer**.
  12. Si nécessaire, modifiez l'analyseur pour tout champ censé utiliser autre chose que ce que vous avez sélectionné dans le champ Analyzer.
  13. Cliquez sur **OK**.

## Sortie

Write to Search Index comporte un port de sortie, appelé **Port d'erreur**, qui collecte les données des enregistrements qui n'ont pas pu être traités ni ajoutés à l'index de recherche. Les enregistrements qui traversent le port d'erreur jusqu'au collecteur de données sont considérés comme non conformes.

Les enregistrements sont transmis au port d'erreur si la valeur de **KeyField** de l'enregistrement d'entrée est vide ou null. Pour les moteurs d'index de recherche qui prennent en charge les recherches autonomes (le moteur d'index patrimonial), le port d'erreur fonctionne uniquement pour les opérations Update et Delete. Cependant, pour le moteur d'index de prise en charge distribuée, il collecte les enregistrements non conformes pour l'ensemble des quatre opérations : Create, Update, Delete et Append.

## Gestion d'index de recherche

L'outil **Gestion d'index de recherche** vous permet d'effectuer les tâches suivantes :



- Ajouter/supprimer des champs dans un index de recherche existant
- Supprimer un ou plusieurs index de recherche

Pour ajouter/supprimer des champs dans un index de recherche existant :

1. Sélectionnez **Outils > Gestion d'index de recherche**. La fenêtre contextuelle **Gestion d'index de recherche** s'affiche, indiquant les index de recherche existants.
2. Sélectionnez l'index que vous voulez modifier et cliquez sur **Modifier**. La page **Modifier index** affiche une liste de tous les champs de l'index de recherche avec leurs détails.
3. Pour supprimer un champ, sélectionnez-le dans la liste, puis cliquez sur **Supprimer le champ**.
4. Pour ajouter un champ à la liste, cliquez sur **Ajouter un champ**. La fenêtre contextuelle **Ajouter un champ** s'affiche. Elle vous permet d'ajouter les détails suivants :
  - **Nom de champ** : Saisissez le nom du champ.
  - **Type** : Sélectionnez le type de champ. Les options sont : chaîne, entier, long, flottant et double.
  - **Index** : Cochez la case pour l'ajouter à l'index pour la demande de recherche.
  - **Stocker** : Cochez la case pour stocker le champ.
  - **Analyseur** : Dans la liste des options, sélectionnez l'analyseur que vous souhaitez utiliser pour indexer le champ.

**Remarque** : Cette option est activée uniquement si vous avez sélectionné l'indexation de ce champ.

Pour supprimer un index de recherche existant :

1. Sélectionnez **Outils > Gestion d'index de recherche**.
2. Sélectionnez les index de recherche à supprimer.
3. Cliquez sur **Supprimer**.
4. Cliquez sur **Fermer**.

**Remarque** : Vous pouvez également supprimer un index de recherche à l'aide de l'utilitaire Administration. La commande est `index delete --d IndexName`, où « IndexName » est le nom de l'index à supprimer.

## Analyseurs standard et de mots clés

Avant de choisir entre les analyseurs standard et de mots clés, vous devez connaître la différence suivante dans le comportement de ces deux analyseurs :

- **Analyseur standard** : Il rompt la chaîne pour la transformer en jetons et convertit également l'ensemble de ses jetons en minuscules.
- **Analyseur de mots clés** : Il convertit la chaîne toute entière en jetons et la conserve telle quelle (sans modifier la casse).

**Exemple** :

Prenons un exemple d'algorithme *contains any* et supposons que l'entrée est P O. Dans ce cas, l'analyseur **standard** renvoie à la fois les enregistrements P O et P, tandis que l'analyseur **de mots clés** renvoie uniquement les enregistrements P.O. (comme indiqué ci-dessous).

Résultats de la recherche avec l'analyseur standard :

Point		Point 2					
SearchTerm	CandidateGroup	HasDuplicates	Index	IndexField	Term	TransactionRecordType	
P O	1	Y				Suspect	
P O	1	D	P O	4	Test4	Candidate	
P O	1	D	P	1	Test 1	Candidate	

Résultats de la recherche avec l'analyseur de mots clés :

Point		Point 2					
SearchTerm	CandidateGroup	HasDuplicates	Index	IndexField	Term	TransactionRecordType	
P O	1	Y				Suspect	
P O	1	D	P O	4	Test4	Candidate	

## Module Business Steward

### Module Business Steward

Le module Business Steward rassemble des fonctions qui vous permettent d'identifier et de résoudre des enregistrements d'exception. Les enregistrements d'exception sont ceux que Spectrum™ Technology Platform n'a pas pu traiter de façon fiable et qui nécessitent une révision manuelle. Exemples d'exceptions :

- Échecs de vérification d'adresse
- Échecs de géocodage
- Correspondances incertaines
- Décisions de fusion/consolidation

Le module Business Steward fournit un outil basé sur un navigateur pour la suppression manuelle des enregistrements d'exception. Une fois que les enregistrements d'exception sont manuellement corrigés et approuvés, ils peuvent être réintégrés dans votre processus de qualité de données de Spectrum™ Technology Platform.

La plupart des actions qui ont lieu dans le module Business Steward sont reflétées dans le journal d'audit, un outil de Management Console qui enregistre l'activité de l'utilisateur. Les actions suivantes sont incluses dans le journal :

- Ajout d'exceptions dans le stage Write Exceptions

- Suppression d'exceptions dans le stage Read Exceptions et sur la page Gérer les exceptions du Business Steward Portal
- Affectation d'exceptions dans le Gestionnaire d'exceptions du Business Steward Portal
- Extraction d'exceptions dans l'Éditeur d'exceptions du Business Steward Portal
- Mise à jour d'exceptions dans l'Éditeur d'exceptions du Business Steward Portal
- Revalidation des exceptions dans l'Éditeur d'exceptions de Business Steward Portal

Pour en savoir plus sur le journal d'audit, reportez-vous au Guide d'administration Spectrum™ Technology Platform pour WebUI.

### Composants

Le module Business Steward est constitué des éléments suivants :

- **Exception Monitor**—Un stage qui examine des enregistrements suivant un ensemble de conditions afin de définir si l'enregistrement exige une révision manuelle par un data steward. Lorsque les enregistrements remplissent ces conditions, ce stage peut envoyer un courrier électronique indiquant les exceptions aux destinataires.
- **Write Exceptions** : stage écrivant les enregistrements d'exception sur le référentiel d'exception. Une fois que les enregistrements d'exception sont dans le référentiel d'exception, ils peuvent être révisés par un data steward.
- **Tableau de bord et éditeur d'exception** : cet outil basé sur un navigateur affiche un tableau de bord du résumé des statistiques et des graphiques pour vous aider à comprendre le genre d'exceptions ayant lieu dans vos données. Il s'agit également d'un éditeur vous permettant de modifier des enregistrements d'exception et d'approuver leur retraitement.
- **Gérer les exceptions** : outil basé sur un navigateur qui vous permet de vérifier et de gérer l'activité d'enregistrement des exceptions pour tous les utilisateurs.
- **Performances qualitatives des données** : outil basé sur un navigateur qui fournit des informations sur les tendances de vos enregistrements d'exceptions, et vous permet d'identifier les indicateurs clés de performance et d'envoyer des notifications lorsque certaines conditions ont été remplies.
- **Read Exceptions** : stage lisant les enregistrements depuis le référentiel d'exception jusqu'à un dataflow. Cette étape vous permet de retraiter les enregistrements d'exception qui ont été corrigés par un data steward.

## Exception Monitor

Le stage Exception Monitor examine des enregistrements suivant un ensemble de conditions afin de définir si l'enregistrement exige une révision manuelle par un data steward. Exception Monitor vous permet de router les enregistrements que Spectrum™ Technology Platform ne pouvait pas correctement traiter jusqu'à un outil de révision manuelle (Business Steward Portal)

Outre la définition de conditions qui déterminent si des enregistrements nécessitent une révision manuelle, vous pouvez également configurer Exception Monitor pour envoyer une notification à une ou plusieurs adresses électroniques lorsque ces conditions ont été remplies un certain nombre de fois.

Pour de plus amples informations sur le traitement d'exception, voir [Module Business Steward](#) à la page 250.

## Entrée

Exception Monitor prend tous les enregistrements comme entrée. Si les données d'entrée ne contiennent pas de champ appelé « CollectionNumber », l'option **Renvoyer tous les enregistrements dans le groupe de l'exception** est désactivée.

**Remarque :** L'Exception Monitor ne peut pas surveiller les champs contenant des données complexes telles que des listes ou des objets géométriques.

## Options

### Onglet Conditions

**Tableau 18 : Options d'Exception Monitor**

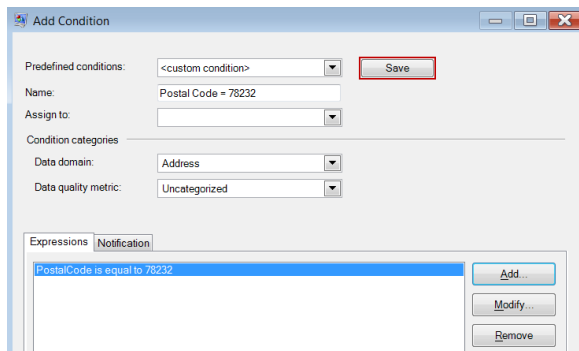
Nom de l'option	Description
Interrompt l'évaluation lorsqu'une condition est remplie	Définit si l'évaluation d'un enregistrement devrait continuer selon les conditions restantes une fois qu'une condition est remplie. L'activation de cette option pourrait améliorer les performances en réduisant potentiellement le nombre d'évaluations que le système doit effectuer. Toutefois, si toutes les conditions ne sont pas évaluées, vous perdrez dans l'exhaustivité des rapports d'exception affichés dans le Business Steward Portal. Par exemple, si vous définissez trois conditions (Intégralité de l'adresse, Certitude du nom et Certitude du géocode) et qu'un enregistrement remplisse les critères définis dans Intégralité de l'adresse, et que cette option est activée, l'enregistrement ne serait pas évalué selon Certitude du nom et Certitude du géocode. Si l'enregistrement serait également qualifié comme étant une exception car il remplirait la condition Certitude du nom, cette information ne serait pas recueillie. Au lieu de cela, l'enregistrement serait rapporté comme n'ayant qu'un problème d'Intégralité de l'adresse, au lieu de montrer un problème d'Intégralité de l'adresse et de Certitude du nom.

### *Ajout ou modification des conditions et des expressions*

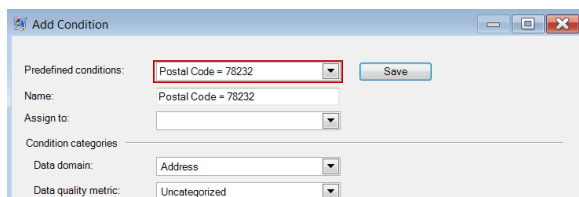
Une condition définit les critères employés afin de déterminer si un enregistrement constitue une « exception » et doit être routée de manière à être examinée manuellement. Cela signifie habituellement que vous devriez définir des conditions pouvant systématiquement identifier les enregistrements qui ont soit échoué le traitement automatique plus tôt dans le flux de données ou qui ont un faible niveau de certitude et devraient ainsi être examinés manuellement.

Le stage Exception Monitor vous permet de créer des conditions prédéfinies et des conditions personnalisées via la boîte de dialogue Ajouter une condition. Les conditions prédéfinies sont disponibles pour tous les flux de données, tandis que les conditions personnalisées sont disponibles uniquement pour les flux de données pour lesquels elles ont été créées. Le processus de

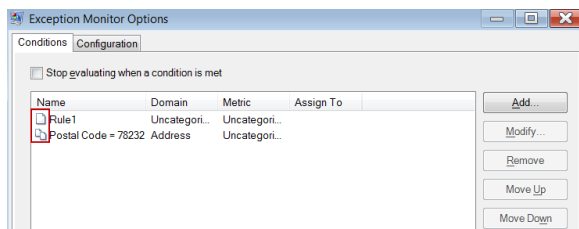
configuration est presque identique pour les deux types ; cependant, pour créer une condition prédéfinie, vous devez enregistrer la condition en renseignant les champs, puis en cliquant sur **Enregistrer**, indiqué dans la case rouge dessous.



Une fois que vous avez enregistré une condition personnalisée, le champ **Conditions prédéfinies** change pour afficher le nom de la condition au lieu de « <condition personnalisée> ».



Une fois que vous avez créé des conditions prédéfinies ou personnalisées, elles apparaissent dans l'onglet Conditions de la boîte de dialogue Options d'Exception Monitor. Comme indiqué sur l'image suivante, l'icône en regard du nom de la condition l'identifie comme condition prédéfinie ou condition personnalisée. Une icône de document double désigne une condition prédéfinie et une icône de document simple désigne une condition personnalisée.



1. Dans l'onglet Conditions de la fenêtre **Options d'Exception Monitor**, cliquez sur **Ajouter** afin de créer une nouvelle condition, ou sur **Modifier** pour modifier une condition existante. Remplissez ces champs :

- **Conditions prédéfinies** : sélectionnez une condition prédéfinie ou conservez « <condition personnalisée> » dans le menu déroulant pour créer une nouvelle condition.
- **Nom** : nom pour la condition. Vous pouvez choisir n'importe quel nom. Vous pouvez utiliser un nom descriptif car le nom de la condition s'affiche dans le Business Steward Portal. Par exemple, « MatchScore<80 » ou « FailedDPV ». Si vous tentez de donner à une nouvelle condition un nom identique à celui d'une condition existante, mais avec d'autres caractères ajoutés à la fin (par exemple, « FailedDPV » et « FailedDPV2 »), il vous est demandé si vous

souhaitez écraser la condition existante dès que vous saisissez le dernier caractère correspondant à son nom (dans notre exemple, « V »). Répondez « Oui » à l'invite, terminez de nommer la condition, puis appuyez sur **OK** ou sur **Enregistrer**. Les deux conditions sont visibles dans la boîte de dialogue Options d'Exception Monitor. La nouvelle condition n'écrase pas la condition existante, sauf si le nom est 100 % identique.

- **Affecter à** : sélectionne un utilisateur auquel les enregistrements d'exception remplissant cette condition doivent être affectés. Si vous n'effectuez pas de sélection dans ce champ, les enregistrements d'exception seront automatiquement affectés à l'utilisateur ayant exécuté le job.
- **Domaine de données**—(Facultatif) Indique le type de données évaluées par la condition. Ceci n'est utilisé qu'aux fins de rapports du Business Steward Portal pour montrer quels types d'exceptions surviennent dans vos données. Par exemple, si la condition évalue la réussite ou l'échec de validation d'adresse, le domaine de données pourrait être « Adresse » ; si la condition évalue la réussite ou l'échec d'une opération de géocodage, le domaine de données pourrait être « Spatial » etc. Vous pouvez choisir votre propre domaine de données ou choisir l'un des domaines prédéfinis :
  - Account—La condition vérifie la raison sociale ou le nom d'entreprise associé à un compte commercial.
  - Address—La condition vérifie les données d'adresse, telles que l'adresse postale complète ou un code postal.
  - Asset—La condition vérifie les données sur les propriétés d'une entreprise, telles que les propriétés physiques, l'immobilier, les ressources humaines ou d'autres actifs.
  - Date—La condition vérifie les données de date.
  - Email—La condition vérifie les données de courrier électronique.
  - Financial—La condition vérifie les données liées aux devises, titres, etc.
  - Name—La condition vérifie les données de nom personnelles, telles que le prénom ou le nom de famille.
  - Phone—La condition vérifie les données de numéro de téléphone.
  - Product—La condition vérifie les données de matériaux, de pièces, de marchandises, etc.
  - Spatial—La condition vérifie les données de point, de polygone ou de ligne représentant une entité géographique, telle que les plaines inondables, les zones littorales, les maisons, les régions de vente, etc.
  - SSN—La condition vérifie les données de numéros américains de sécurité sociale.
  - Uncategorized—Sélectionnez cette option si vous ne souhaitez pas catégoriser cette condition.
- **Métrique de qualité de données**—(Facultatif) Indique la métrique que cette condition mesurera. Ceci n'est utilisé qu'aux fins de rapports du Business Steward Portal pour montrer quels types d'exceptions surviennent dans vos données. Par exemple, si la condition est conçue pour évaluer l'intégralité d'un enregistrement (que toutes les adresses comprennent des codes postaux par exemple), vous pourrez alors utiliser « Intégralité » comme métrique de qualité de données. Vous pouvez choisir votre propre métrique ou choisir l'une des métriques prédéfinies :

- **Accuracy**—La condition vérifie si les données peuvent être comparées à une source de confiance pour vérification. Par exemple, si une adresse n'a pas pu être vérifiée en utilisant les données provenant d'une administration postale, elle pourrait être considérée comme étant une exception car elle manque de précision.
  - **Completeness**—La condition vérifie s'il manque des attributs clés aux données. Par exemple, une adresse n'ayant pas de code postal, ou un compte n'ayant pas de nom de contact.
  - **Consistency**—La condition vérifie si les données du système sont cohérentes entre différents systèmes. Par exemple, si votre système de données de clients utilisent les codes de sexe H et F, mais que les données que vous traitez ont les codes de sexe 0 et 1, les données pourraient être considérées comme ayant des problèmes de consistance.
  - **Interpretability**—La condition vérifie si les données sont correctement décomposées selon une structure de données pouvant être interprétée par un autre système. Par exemple, les numéros de sécurité sociale ne devraient contenir que des données numériques. Si les données comportent des lettres, tels que xxx-xx-xxxx, les données pourraient être considérées comme ayant des problèmes d'intelligibilité.
  - **Recency**—La condition vérifie si les données sont à jour. Par exemple, si une personne déménage mais que l'adresse que vous possédez dans votre système contient l'ancienne adresse de cette personne, les données pourraient être considérées comme ayant un problème concernant leur caractère peu récent.
  - **Uncategorized**—Sélectionnez cette option si vous ne souhaitez pas catégoriser cette condition.
  - **Uniqueness**—La condition vérifie s'il existe des données doublons. Si le flux de données n'a pas pu consolider de données doublons, l'enregistrement serait considéré être une exception.
2. Vous devez ajouter au moins une expression à une condition. Une expression est une phrase logique vérifiant la valeur d'un champ. Pour ajouter une expression, cliquez sur **Ajouter**. Pour modifier une expression existante, cliquez sur **Modifier**. Remplissez ces champs :
- **Expression créée à l'aide du générateur d'expressions** : sélectionnez cette option pour créer une expression de base.
  - **Expression personnalisée** : sélectionnez cette expression pour écrire une expression en utilisant les scripts Groovy. Si vous avez besoin d'utiliser une logique plus complexe, comme les évaluations imbriquées, utilisez une expression personnalisée. Pour plus d'informations, reportez-vous à la section [Utilisation d'expressions personnalisées dans Exception Monitor](#) à la page 257.
  - Si d'autres expressions sont déjà définies pour cette condition, vous pouvez sélectionner un opérateur dans le champ **Logical operator**. L'un des éléments suivants :
    - **And** : cette expression doit avoir une valeur true, tout comme l'expression précédente, pour que la condition ait une valeur true.
    - **Or** : si cette expression a une valeur true, la condition a une valeur true, même si l'expression précédente n'a pas cette valeur.
  - Si vous choisissez de créer une expression avec le générateur d'expression, les champs suivants sont disponibles :



- **Nom du champ** : sélectionnez le champ qui doit être évalué par cette expression. La liste des champs disponibles est peuplée en fonction des stages en amont du stage Exception Monitor.
  - **Opérateur** : sélectionnez l'opérateur qui doit être utilisé dans l'évaluation.
  - **Valeur** : indiquez la valeur que l'expression doit rechercher à l'aide de l'opérateur choisi dans le champ Operator.
3. Cliquez sur **Ajouter** pour ajouter l'expression. Cliquez sur **Fermer** une fois que vous avez fini d'ajouter les expressions.
  4. Utilisez les boutons **Monter** et **Descendre** pour changer l'ordre suivi lors de l'évaluation des expressions.
  5. Cliquez sur l'onglet **Notification** si vous souhaitez qu'Exception Monitor envoie un message à une ou plusieurs adresses électroniques lorsque cette condition est remplie un certain nombre de fois. Cet e-mail inclut un lien vers les enregistrements en échec d'Exception Editor du Business Steward Portal, où vous pouvez saisir manuellement les données correctes. Si vous ne souhaitez pas configurer les notifications, passez à l'étape 11. Pour arrêter de recevoir des notifications à une adresse électronique particulière, supprimez cette adresse de la liste des destinataires dans la ligne **Envoyer une notification à** de l'onglet Notification dans la boîte de dialogue Modifier une condition.

**Remarque** : Les notifications doivent être configurées dans Management Console avant de pouvoir utiliser une notification avec succès, à partir d'Exception Monitor. Pour plus d'informations sur la configuration des notifications, reportez-vous au *Guide d'administration*.

6. Entrez les adresses électroniques auxquelles la notification doit être envoyée. Séparez les adresses par des virgules, des espaces ou des point virgules.
7. Désignez le point auquel une notification doit être envoyée. Elle peut être envoyée à la première occurrence de la condition ou lorsque la condition a été remplie un nombre de fois spécifique. La valeur maximale est de 1 000 000 occurrences.
8. Cochez la case **Envoyer un rappel après** si les messages de rappel doivent être envoyés aux adresses électroniques désignées après l'e-mail initial.
9. Entrez le nombre de jours suivant l'e-mail initial au moment duquel l'e-mail de rappel doit être envoyé.
10. Cliquez sur **Rappel quotidien** si vous souhaitez que les messages de rappel soient envoyés tous les jours suivant le premier e-mail de rappel.
11. Saisissez l'**Objet** de la notification par courrier électronique.
12. Si vous souhaitez enregistrer cette condition pour la réutiliser comme condition prédéfinie, cliquez sur **Enregistrer**. Si vous modifiez une condition existante et que vous cliquez sur Enregistrer, il vous est demandé si vous souhaitez écraser la condition existante ; notez que si vous écrasez une condition prédéfinie, ces modifications s'appliquent à tous les flux de données utilisant la condition.
13. Lorsque vous avez fini de travailler avec les expressions, cliquez sur **OK**.



14. Ajoutez ou modifiez les conditions supplémentaires selon vos besoins.
15. Utilisez les boutons **Monter** et **Descendre** pour changer l'ordre suivi lors de l'évaluation des conditions. L'ordre des conditions n'est important que si vous avez activé l'option **Arrêter l'évaluation lorsqu'une condition est remplie**. Pour plus d'informations, reportez-vous à la section **Onglet Configuration** à la page 259.
16. Cliquez sur **OK** quand vous aurez terminé.

#### *Suppression d'une condition ou d'une expression*

- Pour supprimer une condition, ouvrez Exception Monitor, sélectionnez la condition qui doit être supprimée, puis cliquez sur **Supprimer**. Notez bien que lorsque vous supprimez une condition, toutes les expressions de la condition sont supprimées.
- Pour supprimer une expression, ouvrez la condition qui contient l'expression, sélectionnez l'expression, puis cliquez sur **Supprimer**.

#### *Utilisation d'expressions personnalisées dans Exception Monitor*

Vous pouvez écrire vos propres expressions personnalisées pour contrôler la manière dont Exception Monitor achemine les enregistrements en utilisant le langage de script Groovy pour créer une expression.

#### *Utilisation de la méthode de script Groovy*

Pour obtenir des informations sur Groovy, consultez l'adresse suivante : [groovy-lang.org](http://groovy-lang.org).

Les expressions Groovy utilisées dans le stage Exception Monitor doivent être évaluées par rapport à une valeur booléenne (true ou false) qui indique si l'enregistrement est considéré comme une exception et doit être acheminé à des fins de vérification manuelle. Les enregistrements d'exception sont routés au port d'exception.

Par exemple, si vous devez vérifier des enregistrements d'un niveau de confiance de validation <85, votre script prendra la forme suivante :

```
data['Confidence']<85
```

Le moniteur évalue alors la valeur du champ Confidence en fonction de vos critères pour déterminer sur quel port de sortie il doit l'envoyer.

#### **Vérification qu'un champ contient une seule valeur**

Cet exemple est évalué sur true si le champ Status contient la lettre « F ». Il faut que ce soit une correspondance exacte, par conséquent « f » ne serait pas évalué sur true.

```
return data['Status'] == 'F';
```

**Vérification qu'un champ contient plusieurs valeurs**

Cet exemple est évalué sur true si le champ Status contient la lettre « F » ou « f ».

```
boolean returnValue = false;
if (data['Status'] == 'F' || data['Status'] == 'f')
{
    returnValue = true;
}
return returnValue;
```

**Évaluation de la longueur d'un champ**

Cet exemple est évalué sur true si le champ PostalCode contient plus de 5 caractères.

```
return data['PostalCode'].length() > 5;
```

**Vérification qu'un caractère se trouve dans une valeur de champ**

Cet exemple est évalué sur true si le champ PostalCode contient un tiret.

```
boolean returnValue = false;
if (data['PostalCode'].indexOf('-') != -1)
{
    returnValue = true;
}
return returnValue;
```

**Erreurs courantes**

Ce qui suit illustre les erreurs courantes concernant l'usage de scripts.

Ce qui suit est incorrect car PostalCode (le nom de la colonne) doit être écrit entre guillemets simples ou doubles.

```
return data[PostalCode];
```

Ce qui suit est incorrect car aucune colonne n'est spécifiée

```
return data[];
```

Ce qui suit est incorrect, car row.set() ne renvoie pas de valeur booléenne. Il sera toujours évalué sur false et changera toujours la valeur du champ PostalCode sur 88989.

```
return row.set('PostalCode', '88989');
```

Employez un seul signe égal pour définir la valeur d'un champ, et deux signes égal pour vérifier la valeur d'un champ.

### Onglet Configuration

**Tableau 19 : Options d'Exception Monitor**

Nom de l'option	Description
Désactiver Exception Monitor	Éteint ou allume Exception Monitor. Si vous désactivez Exception Monitor, les enregistrements passeront simplement par l'étape sans qu'aucune action n'ait lieu. Cela reviendrait à supprimer Exception Monitor du dataflow.
Interrompre la tâche après avoir atteint la limite d'exception.	Indique si l'exécution de la tâche devrait être interrompue lorsque le nombre défini d'enregistrements respecte les conditions d'exception.
Nombre maximum d'enregistrement d'exception	Si <b>Interrompre la tâche après avoir atteint la limite d'exception.</b> est sélectionné, utilisez ce champ pour indiquer le nombre maximum d'enregistrements d'exception à autoriser avant que l'exécution de la tâche soit interrompue. Par exemple, si vous indiquez 100, la tâche s'arrêterait une fois que le 101ème enregistrement d'exception est rencontré.
Rapport uniquement (ne pas créer d'exceptions)	Vous permet de suivre les enregistrements remplissant les conditions d'exception et signale ces statistiques sur la page Performances qualitatives des données du Business Steward Portal, mais ne crée pas d'exceptions pour ces enregistrements.
Renvoyer tous les enregistrements du groupe de l'exception	<p>Indique si les enregistrements appartenant à un groupe d'un enregistrement d'exception devraient être renvoyés au lieu de juste renvoyer l'enregistrement d'exception. Par exemple, un groupe de correspondance (basé sur une MatchKey) contient quatre enregistrements. L'un d'eux est l'enregistrement Suspect, un autre est un doublon ayant une note de 90 et deux sont des enregistrements uniques ayant pour note 80 et 83. Si vous avez une condition indiquant que tout enregistrement avec un MatchScore entre 80 et 89 constitue une exception, par défaut, seuls les enregistrements avec un score de correspondance de 80 et 83 seraient envoyés au port d'exception. Si toutefois vous activez cette option, les quatre enregistrements seraient envoyés au port d'exception.</p> <p>Activez cette option si vous voulez que les data stewards soient en mesure de comparer l'enregistrement d'exception avec les autres enregistrements du groupe. En comparant tous les enregistrements du groupe, les data stewards pourraient être capable des décisions plus éclairées sur quoi faire avec un enregistrement d'exception. Par exemple, dans une situation de mise en correspondance, un data steward pourrait voir tous les candidats afin de déterminer si l'exception est une réplique des autres.</p> <p><b>Remarque :</b></p> <p>Si les données d'entrée ne contiennent pas d'un champ appelé « CollectionNumber », cette option est désactivée.</p>

Nom de l'option	Description
Grouper par	<p>Si vous avez sélectionné <b>Renvoyer tous les enregistrements dans le groupe de l'exception</b>, sélectionnez le champ suivant lequel regrouper les enregistrements.</p> <p><b>Remarque :</b> Le champ d'entrée « CollectionNumber » n'apparaît pas dans cette liste, car il ne s'agit pas d'une sélection valide pour la fonctionnalité Grouper par.</p>
Service de revalidation	Sélectionnez le service que vous souhaitez exécuter lorsque vous revalidez les enregistrements de ce dataflow.
Action après revalidation	Indique si vous souhaitez retraiter les enregistrements ou approuver les enregistrements qui ont été revalidés.
Rapprocher les enregistrements d'exception à l'aide du champ de correspondance	<p>Utilise des champs de rapprochement pour rapprocher les enregistrements d'entrée des enregistrements d'exception du référentiel. Activez cette option si votre entrée contient des enregistrements qui généraient auparavant des exceptions, mais qui sont maintenant corrigés dans l'entrée.</p> <p>Les enregistrements d'entrée sont validés par rapport à la ou aux conditions, puis rapprochés des enregistrements d'exception existants du référentiel. Si un enregistrement d'entrée remplit la ou les conditions et correspond à un enregistrement d'exception, cet enregistrement d'exception est supprimé du référentiel. Si un enregistrement d'entrée ne remplit pas la ou les conditions et correspond à un enregistrement d'exception, cet enregistrement d'exception est mis à jour et conservé dans le référentiel. En outre, s'il existe des doublons dans le référentiel, une seule exception rapprochée par dataflow est mise à jour ; toutes les autres de ce dataflow sont supprimées.</p>
Optimisé pour les enregistrements uniques ou les petits lots	<p>Cette option est activée lorsque vous sélectionnez <b>Champs de correspondance</b>. Lorsque cette option est décochée (par défaut), le serveur charge en mémoire tous les enregistrements d'exception existants du dataflow et du stage actuels avant de traiter les enregistrements d'exception entrants. Ceci est recommandé lorsque le référentiel comporte un petit nombre d'enregistrements d'exception existants et un grand nombre de nouveaux enregistrements d'exception ou de nouvelles mises à jour. Ce scénario implique généralement un temps de chargement initial plus long et un besoin en mémoire accru. Il est plus rapide lors du traitement de lots plus volumineux, tels que les mises à jour quotidiennes, hebdomadaires ou mensuelles.</p> <p>Il est recommandé de cocher cette option lorsque le référentiel contient un grand nombre d'enregistrements d'exception existants et un nombre relativement petit de nouveaux enregistrements d'exception ou de nouvelles mises à jour, car le serveur interroge le référentiel pour savoir s'il contient des enregistrements d'exception existants lors de la lecture de chaque enregistrement d'entrée. Ce scénario implique généralement un temps de chargement initial plus court et des besoins en mémoire moins importants. Il est plus rapide lors du traitement de quelques enregistrements en temps réel.</p>

Nom de l'option	Description
Champs de correspondance	Fournit une liste de tous les champs d'entrée utilisés pour créer une clé permettant de rapprocher un enregistrement d'exception du référentiel. Vous devez définir au moins un champ de rapprochement si vous avez coché la case <b>Rapprocher les enregistrements d'exception à l'aide du champ de correspondance</b> .

## Sortie

L'Exception Monitor renvoie des enregistrements sur deux ports. L'un contient les enregistrements qui ne satisfont à aucune des conditions définies dans le stage Exception Monitor. L'autre, le port d'exception, contient tous les enregistrements qui correspondent à une ou plusieurs conditions d'exception. Le port d'exception peut aussi inclure des enregistrements qui ne sont pas des enregistrements d'exception si vous activez l'option **Renvoyer tous les enregistrements du groupe de l'exception**. L'Exception Monitor n'ajoute ou ne modifie pas les champs au sein d'un enregistrement.

## Read Exceptions

Lire les exceptions est une étape qui lit les enregistrements depuis le dépositaire d'exception en tant qu'entrée d'un flux de données. (Pour plus d'informations sur le dépositaire d'exception, voir [Module Business Steward](#) à la page 250.)

**Remarque :** Une fois qu'un enregistrement est lu dans un flux de données par Lire les exceptions, celui-ci est supprimé du dépositaire.

## Entrée

Lire les exceptions lit les données à partir d'un dépositaire d'exception. Il n'accepte pas d'entrée d'un autre stage dans un flux de données.

**Remarque :** Seuls les enregistrements marqués comme « approuvés » dans le Business Steward Portal sont lus dans le flux de données.

## Options

Le stage Lire les exceptions inclut les options suivantes.

### Onglet Général

Les options de l'onglet **Général** indique quels enregistrements d'exception vous devriez lire dans le flux de données.

Les options **Filter** vous permettent de sélectionner un sous-ensemble d'enregistrements depuis le référentiel d'exceptions en utilisant les critères suivants :

- **Utilisateur** : utilisateur ayant exécuté le flux de données qui a généré les exceptions que vous souhaitez lire dans le flux de données.
- **Nom du flux de données** : le nom du flux de données ayant généré les exceptions que vous devriez lire dans le flux de données.
- **Nom du stage** : l'étiquette d'étape de l'Exception Monitor tel qu'il apparaît dans le flux de données dans Enterprise Designer. Ce critère s'avère utile si le flux de données qui a généré les exceptions contient plusieurs stages Exception Monitor et que vous ne voulez lire que les exceptions d'une de ces stages Exception Monitor.
- **Date de départ** : la date et l'heure du plus ancien enregistrement que vous voulez lire dans le flux de données. La date d'un enregistrement d'exception est la date de sa dernière modification.
- **Date d'expiration** : la date et l'heure du plus récent enregistrement que vous voulez lire dans le flux de données. La date d'un enregistrement d'exception est la date de sa dernière modification.

La liste des **Champs** affiche les champs qui seront lus dans le flux de données. Les champs sont par défaut tous inclus, mais vous pouvez exclure des champs en décochant la case depuis la colonne **Inclure**.

La liste **Aperçu** affiche les enregistrements remplissant les critères indiqués dans **Filter**.

**Remarque** : L'aperçu n'affiche que les enregistrements marqués « Approuvés » dans le Business Steward Portal et remplissant le critère de filtre.

### Onglet Tri

Utiliser l'onglet Tri pour trier les enregistrements d'entrée basés sur les valeurs de champ.

- **Ajouter** : ajoute un champ à trier.
- Colonne **Nom du champ** : affiche le nom du champ servant au tri. Vous pouvez choisir un champ en cliquant sur le bouton déroulant.
- Colonne **Classer** : indique si le tri s'effectue de manière croissante ou décroissante.
- **Haut** et **Bas** : modifie l'ordre du tri. Les enregistrements sont d'abord triés par le champ en haut de la liste, puis par le second, etc.
- **Supprimer** : supprime un champ de tri.

### L'onglet Exécution

- **Enregistrement de départ** : indiquez la position dans le référentiel du premier enregistrement que vous voulez lire vers le flux de données. Par exemple, si vous voulez sauter les premier 99 enregistrements dans le répertoire, vous devez spécifier 100. Le 100ème enregistrement serait le premier enregistrement lu vers le référentiel s'il correspond aux critères indiqués dans l'onglet **Général**. La position d'un enregistrement est déterminée par l'ordre des enregistrements dans le Business Steward Portal.
- **Tous les enregistrements** : choisissez cette option si vous voulez lire tous les enregistrements correspondant au critère de recherche indiqué dans l'onglet **Général**.

- **Nombre d'enregistrements max** : choisissez cette option si vous voulez limiter le nombre d'enregistrements lus dans le flux de données. Par exemple, si vous ne voulez lire que les 1000 premiers enregistrements qui correspondent au critère de sélection, choisissez cette option et spécifiez 1000.

## Sortie

L'étape Lire les exceptions renvoie les enregistrements du dépositaire d'exception qui ont été approuvés et qui correspondent au critère de sélection indiqué dans les options de Lire les exceptions. En plus des champs des enregistrements, Lire les exceptions renvoie les champs décrivant les dernières modifications apportées à l'enregistrement dans le Business Steward Portal.

**Tableau 20 : Sortie de Lire les exceptions**

Nom du champ	Description
Exception.Comment	Tout commentaire saisi par la personne ayant résolu l'exception. Par exemple, les commentaires peuvent décrire les modifications que le business steward a apportées à l'enregistrement.
Exception.LastModifiedBy	Le dernier utilisateur à avoir modifié l'enregistrement dans le Business Steward Portal
Exception.LastModifiedMilliseconds	L'instant où l'enregistrement a été modifié la dernière fois dans le Business Steward Portal. L'instant est exprimé en millisecondes depuis le 1er janvier 1970 à 0h00 GMT. Il s'agit de la manière standard de calculer l'heure dans le langage de programmation Java. Vous pouvez utiliser cette valeur pour effectuer des comparaisons de date ou pour créer une transformation pour convertir cette valeur dans le format de date voulu.
Exception.LastModifiedString	L'instant où l'enregistrement a été modifié la dernière fois dans le Business Steward Portal. Ce champ offre une représentation plus compréhensible de la date que le champ Exception.LastModifiedMilliseconds. L'instant est exprimé sous ce format :  Thu Feb 17 13:34:32 CST 2011

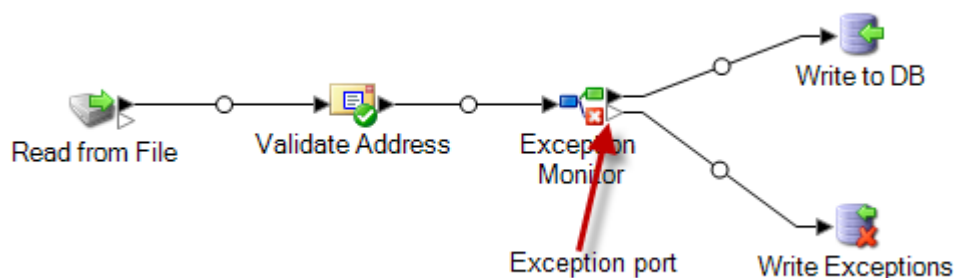
## Write Exceptions

Écrire les exceptions est une étape qui prend les enregistrements que le stage Exception Monitor a identifiés comme étant des exceptions et les écrits sur le dépositaire d'exception. Une fois dans le dépositaire d'exception, les enregistrements peuvent être révisés et édités via le Business Steward Portal.

**Remarque :** Les enregistrements d'exception ne sont pas écrits sur le Business Steward Portal lorsque les jobs ou services sont exécutés en mode d'inspection dans Enterprise Designer ou en mode d'aperçu dans Management Console.

### Entrée

L'étape Écrire les exceptions récupère les enregistrements depuis le port d'exception de le stage Exception Monitor et les écrit vers le dépositaire d'exception. L'étape Écrire les exceptions devrait être placé en aval du port d'exception de le stage Exception Monitor. Le port d'exception est le port de sortie de dessous de le stage Exception Monitor



### Options

Le stage Écrire les exceptions vous permet de sélectionner les données de champ qui doivent être renvoyées vers le dépositaire d'exception. Les champs qui apparaissent dépendent des stages qui se produisent en amont dans le flux de données. Si, par exemple, vous disposez d'un stage Validate Address dans le flux de données, vous voyez des champs tels que AddressLine1, AddressLine2, City, PostalCode et ainsi de suite dans le stage Write Exceptions. Par défaut, tous ces champs sont sélectionnés ; décochez les cases des champs que le référentiel d'exceptions ne doit pas renvoyer. L'ordre des champs est déterminé par la manière dont ils sont classés lorsqu'ils parviennent dans le stage Write Exceptions. Vous pouvez réorganiser les champs en sélectionnant une ligne et en utilisant les flèches à droite de l'écran pour déplacer la ligne vers le haut ou vers le bas. L'ordre que vous sélectionné ici est conservé pour tous les utilisateurs dans Business Steward Portal ; en revanche, chaque utilisateur peut réorganiser les champs comme il le souhaite dans le portail.

Sélectionnez **Autoriser l'utilisateur à créer des enregistrements Best Of Breed dans Portal** pour effectuer une consolidation manuelle des enregistrements lorsque des correspondances de



données ne peuvent pas être effectuées à l'aide des règles standard. Cette fonction copie un enregistrement sélectionné dans un groupe et l'utilise à la place des doublons pour effectuer le traitement. Cette option est disponible uniquement pour les enregistrements d'exception groupés générés à partir d'un job de mise en correspondance, qui est identifié par la présence du champ CandidateGroup ou du champ CollectionNumber. Lorsque vous utilisez cette option, un champ en lecture seule appelé « CollectionRecordType » est ajouté à l'enregistrement d'exception. Vous pouvez voir ce champ au bas de la liste. Notez que toutes les options de ce champ sont désactivées.

**Remarque :** Si votre flux de données est également configuré pour utiliser la [revalidation](#), vous devrez ajouter et exposer manuellement le champ CollectionRecordType dans le stage/sous-flux Exception Monitor et le service lui-même.

Une fois que vous avez ajouté un enregistrement **Best Of Breed** dans la vue Résoudre les doublons de Exception Editor de Business Steward Portal, ce champ est défini sur « BestOfBreed ». Si vous choisissez de créer des enregistrements Best Of Breed, vous ne pourrez pas utiliser l'option **Tout approuver** pour ces enregistrements dans Business Steward Portal. Découvrez-en plus sur les enregistrements Best Of Breed dans Business Steward Portal [ici](#).

Il se peut que vous souhaitiez disposer de certains champs d'entrée dans le référentiel, sans qu'ils soient visibles dans le Business Steward Portal. En effet, ce champ peut contenir des données sensibles ou, tout simplement, vous souhaitez simplifier l'affichage du portail. Cochez la case **Autoriser l'affichage** pour indiquer les champs sélectionnés à afficher une fois qu'ils ont été transmis au référentiel d'exceptions. Par défaut, tous les champs sont affichables. Décochez la case de tout champ que vous souhaitez masquer dans le portail.

Vous pouvez également désigner, parmi les champs sélectionnés, ceux qui doivent être modifiables dans le portail une fois qu'ils sont transmis au référentiel d'exceptions. Par défaut, la colonne **Autoriser la modification** est cochée pour tous les champs entrant dans le stage Write Exceptions. Décochez la case des champs que le référentiel d'exceptions doit renvoyer dans un état de lecture seule.

Enfin, vous pouvez utiliser la fonction **Recherche** pour affecter une recherche à un champ contenant les données problématiques. Vous pouvez sélectionner dans la liste des recherches qui ont été définies dans l'outil de configuration Business Steward, ou vous pouvez saisir manuellement le nom de la recherche. Pour plus d'informations sur les recherches, reportez-vous à la section [Recherches](#) à la page 290.

**Remarque :** Les recherches peuvent être affectées uniquement aux champs dont le type est chaîne.

## Sortie

Écrire les exceptions ne renvoie pas de sorties dans le flux de données. Il écrit les enregistrements d'exception vers le dépositaire d'exception.

## Business Steward Portal

### Introduction au Business Steward Portal

#### *Qu'est-ce que le Business Steward Portal ?*

Le Business Steward Portal est un outil qui permet de réviser, de modifier et d'approuver les enregistrements dont le traitement automatique a échoué ou dont le traitement n'a pas obtenu un niveau de confiance suffisant. Utilisez le Business Steward Portal pour saisir manuellement les données correctes ou des données supplémentaires dans un enregistrement. Par exemple, si l'enregistrement d'un client échoue lors d'un processus de validation d'adresse, vous pouvez utiliser les outils de recherche pour rechercher et déterminer l'adresse du client, puis modifier l'enregistrement afin qu'il contienne l'adresse correcte. L'enregistrement modifié peut ensuite être approuvé et retraité par Spectrum™ Technology Platform, envoyé à un autre processus de validation ou d'enrichissement, ou écrit dans une base de données, selon votre configuration. Vous pouvez également utiliser le portail pour ajouter des informations qui n'étaient pas dans l'enregistrement d'origine.

Le Business Steward Portal offre également des graphiques récapitulatifs offrant un aperçu des différents types de données déclenchant le traitement d'exceptions, y compris le domaine des données (nom, adresse, spatial, etc.), ainsi que les valeurs de qualité des données engendrant l'échec des données (exhaustivité, précision, cohérence, etc.).

En outre, la page Gérer les exceptions du Business Steward Portal vous permet de consulter et de gérer l'activité des enregistrements d'exception, notamment de réaffecter des enregistrements d'un utilisateur à un autre. Enfin, la page Qualité des données du portail Business Steward fournit des informations sur les tendances au sein des dataflows et des stages.

Pour de plus amples informations sur le traitement d'exception, voir [Module Business Steward](#) à la page 250.

#### *Accès au Business Steward Portal*

Pour ouvrir le Business Steward Portal, allez dans **Démarrer > Tous les programmes > Pitney Bowes > Spectrum Technology Platform > Serveur > Page d'accueil** et sélectionnez **Qualité des données Spectrum**, puis **Business Steward Portal** et cliquez sur **Ouvrir le Business Steward Portal**.

Vous pouvez également suivre les étapes suivantes :

1. Ouvrez un navigateur Web et accédez à `http://<servername>:8080/bsmportal`.

Par exemple,

`http://myserver:8080/bsmportal`

Contactez l'administrateur de Spectrum™ Technology Platform si vous ne connaissez pas le nom ni le port du serveur.

2. Connectez-vous à Spectrum™ Technology Platform. Contactez l'administrateur de Spectrum™ Technology Platform si vous avez des difficultés pour vous connecter.

### **Menu Business Steward Portal**

Le menu du Business Steward Portal se compose de quatre options et accède au système d'aide, comme indiqué ci-dessous :

- **Tableau de bord** : affichez l'état des exceptions qui vous sont affectées. Si vous disposez de droits d'affichage, vous pouvez également afficher l'état des exceptions affectées aux autres utilisateurs.
- **Éditeur** : passez en revue, modifiez et approuvez des enregistrements d'exception pour le retraitement.
- **Gérer** : si vous disposez de droits d'affichage, vous pouvez affecter des exceptions, à vous ou aux autres. Si vous disposez de droits de suppression, vous pouvez supprimer des enregistrements d'exception du référentiel.
- **Qualité des données** : si vous disposez de droits d'affichage, vous pouvez accéder à cette page, afficher des informations statistiques et configurer des indicateurs clés de performance pour les enregistrements d'exception. Si vous ne disposez pas de droits d'affichage, vous ne pouvez pas accéder à cette page.
- Menu déroulant utilisateur : accès au système d'aide Business Steward Portal, à toute la documentation Spectrum Technology Platform, à la page de profil, sur laquelle vous pouvez définir une préférence de langue et spécifier un pays pour indiquer les préférences culturelles, fournir une adresse électronique pour les notifications et modifier votre mot de passe.

### **Page Tableau de bord**

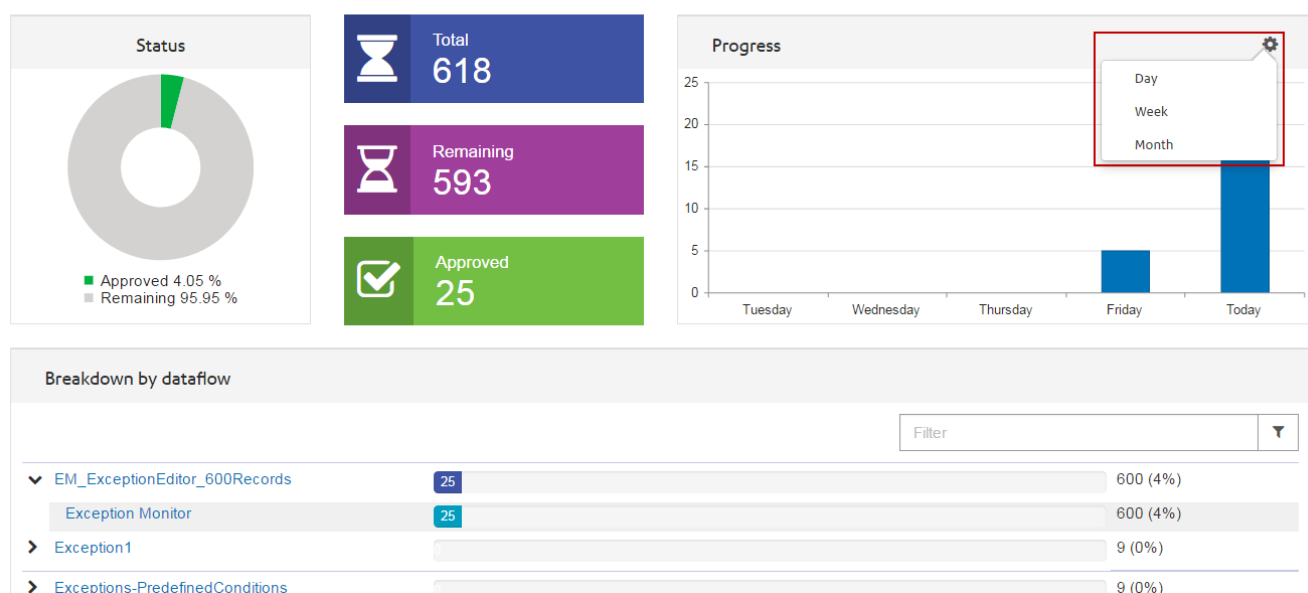
Le tableau de bord des exceptions affiche les données qui récapitulent l'état des enregistrements d'exception qui vous appartiennent et de ceux des autres utilisateurs. (Notez que vous ne pouvez voir les données des autres utilisateurs que si vous disposez d'autorisations de modification). Ceci inclut les éléments suivants :

- Nombre total d'exceptions
- Le nombre et le pourcentage d'exceptions qui ont été approuvées par rapport à celles qui restent non approuvées
- La progression de l'utilisateur sélectionné, chaque jour, semaine ou mois
- Progression de l'approbation des enregistrements d'exception par flux de données et stage dans le flux de données

**Remarque** : Les graphiques de progression de l'utilisateur et du dataflow n'apparaîtront pas si vous avez désactivé le suivi sur la page Paramètres de Business Steward de Management Console.

1. Dans l'onglet Tableau de bord, sélectionnez l'utilisateur dont vous souhaitez afficher les activités d'exceptions dans la première case à liste déroulante mise en surbrillance ci-dessous. Notez que seuls les utilisateurs qui ont actuellement des exceptions affectées apparaîtront dans cette liste.

## My Status



2. Cliquez sur la roue crantée **Paramètres** (affichée dans la deuxième case à liste déroulante mise en surbrillance ci-dessous) pour sélectionner une durée en jours, semaines ou mois.
3. Saisissez des critères de recherche dans le champ **Filtre** pour réduire la liste de dataflows ou d'un sous-ensemble de dataflows. Le terme de recherche doit être inclus dans le nom du flux de données.

## Page Éditeur

L'Éditeur d'exceptions vous permet de réviser, de modifier et d'approuver manuellement des enregistrements d'exception. L'objectif d'une révision manuelle est de déterminer les données incorrectes ou manquantes pour les actualiser et les approuver, en particulier si Spectrum™ Technology Platform n'a pas été en mesure de les corriger dans le cadre d'un processus de flux de données automatique. Vous pouvez ensuite revalider les enregistrements d'exception pour les approuver ou les retraiter.

Vous pouvez également utiliser l'Éditeur d'exceptions pour résoudre les enregistrements d'exception en double. Vous pouvez aussi utiliser des outils de recherche pour rechercher des informations qui vous permettent de modifier, d'approuver et de réexécuter des enregistrements.

### Personnalisation du contenu de l'Éditeur d'exceptions

Il existe plusieurs manières de personnaliser l'affichage dans l'Éditeur d'exceptions. Vous pouvez utiliser **Options de sélection** pour renvoyer les enregistrements d'un utilisateur, d'un flux de données ou d'un ID du job, etc. Vous pouvez utiliser l'outil **Filtre de champ** pour afficher les enregistrements répondant à certains critères, tandis que les enregistrements qui ne remplissent pas les critères sont masqués.

Si vous utilisez l'outil **Éléments par page** et si vous appliquez ensuite des options de sélection ou un filtre de champs pour affiner la liste d'enregistrements d'exception affichés, n'oubliez pas qu'il peut exister plusieurs pages de résultats et non pas seulement celle affichée sur l'écran initial. Par exemple, imaginons que vous ayez défini Éléments par page sur la valeur 10, puis que vous appliquez un filtre de champs pour renvoyer uniquement les enregistrements qui ont un code postal spécifique. Au départ, il peut sembler que seulement 10 enregistrements ont été renvoyés, mais il se peut qu'il existe plusieurs pages de résultats, étant donné que vous avez défini la limite du nombre d'enregistrements à afficher simultanément sur 10.

#### *Utilisation des options de sélection*

Sélectionnez les critères des champs de liste déroulante du volet des options de sélection pour affiner les enregistrements affichés dans la grille Exception.

<b>Utilisateur</b>	Obligatoire. ID de l'utilisateur affecté au flux de données. Ces informations sont visibles uniquement si vous disposez des droits de modification.
<b>Nom de flux de données</b>	Obligatoire. Nom du flux de données qui a généré les enregistrements d'exception.
<b>Libellé de stage</b>	Obligatoire. Le nom donné à le stage Exception Monitor dans le flux de données. Cette information s'avère particulièrement utile lorsque le flux de données contient plusieurs stages Exception Monitor. Si la personne qui a créé le flux de données a étiqueté chaque stage Exception Monitor de manière descriptive, vous pouvez identifier quel Exception Monitor a produit l'enregistrement d'exception. L'intitulé par défaut est « Exception Monitor ».
<b>ID de job</b>	Obligatoire. Un numéro d'identification assigné au job par le système. Chaque fois qu'un job est exécuté, il se voit attribuer un nouveau numéro de job.
<b>Status</b>	Facultatif. État d'approbation de l'enregistrement.
<b>Date</b>	Facultatif. La date (et facultativement, l'heure) à laquelle le flux de données s'est exécuté. Pour entrer l'heure, saisissez l'heure après la date.
<b>Domaine de données</b>	Facultatif. Le type de données ayant produit une exception. Nom, Adresse et Numéro de téléphone sont des exemples de domaines de données. Ces informations vous aident à identifier quels champs de l'enregistrement demandent à être édités.
<b>Mesures de la qualité</b>	Facultatif. Les mesures de qualité que l'enregistrement a ratées. La Précision, l'Intégralité et l'Unicité sont des exemples de métriques de qualité. Ces informations vous aident à définir pourquoi l'enregistrement a été identifié comme constituant une exception.

#### *Utilisation du filtre de champs*

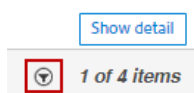
Une fois que vous avez appliqué des options de sélection et que les enregistrements d'exception sont chargés, vous pouvez utiliser le filtrage de champs pour afficher uniquement les enregistrements qui vous intéressent. Le Business Steward Portal n'affiche par défaut que les enregistrements d'un seul flux de Spectrum™ Technology Platform à la fois. Vous pouvez encore filtrer la liste d'enregistrements pour afficher uniquement les enregistrements répondant à certains critères dans un champ donné. Une fois qu'un filtre est créé, il est automatiquement enregistré et la prochaine fois que vous ouvrez le flux de données dans l'Éditeur d'exceptions, le filtre est appliqué.

**Remarque :** Vous devez appliquer des options de sélection avant de pouvoir utiliser l'outil de filtrage de champs.

Vous pouvez créer des filtres pour plusieurs champs, mais vous ne pouvez créer qu'un seul filtre pour chaque champ. Si un champ comporte déjà un filtre, l'arrière-plan de la flèche est bleu :

City	FirstName
KEENE	LAKSHMI
KEENE	LAKSHMI

Les filtres ne peuvent être créés que dans la vue sous forme de tableau. Cependant, les filtres définis dans la vue sous forme de tableau sont également reflétés dans la vue sous forme de formulaire. À l'instar de la vue sous forme de tableau, un indicateur est présent près du bas de l'écran pour indiquer qu'un enregistrement a été filtré :



Pour filtrer la liste des enregistrements :

1. Cliquez sur le bouton **Filtrer**. Une icône de filtre figure en regard de chaque en-tête de colonne.
2. Cliquez sur l'icône de filtre du champ dont vous souhaitez filtrer les données.
3. Sélectionnez un opérateur approprié au type de données du champ, suivi d'une valeur.

**is equal to** Cherche les enregistrements possédant la valeur exacte que vous avez spécifiée. Il peut s'agir d'une valeur numérique ou d'une valeur textuelle. Par exemple, vous pouvez rechercher des enregistrements d'une valeur MatchScore égale à exactement 82, ou des enregistrements d'une valeur LastName « Smith ».

**is not equal to** Cherche les enregistrements ayant n'importe quelle valeur autre que celle que vous avez indiquée. Il peut s'agir d'une valeur numérique ou d'une valeur textuelle. Par exemple, vous pouvez rechercher les enregistrements de toute valeur MatchScore sauf 100, ou les enregistrements de toute valeur LastName sauf « Smith ».

**starts with** Cherche les enregistrements commençant par une valeur particulière dans le champ sélectionné. Par exemple, si vous appliquez un filtre pour la valeur « Van » dans le champ Nomdefamille, vous verrez des enregistrements avec « Van Buren », « Vandenburg » ou « Van Dyck ».

**contains** Cherche les enregistrements contenant la valeur que vous avez indiquée selon n'importe quelle position dans le champ sélectionné. Par exemple, si vous appliquez un filtre pour la valeur « South » dans le champ Ligne d'adresse 1, vous verrez des enregistrements avec « 12 South Ave. », « 9889 Southport St. », « 600 South Shore Dr. » et « 4089 5th St. South ».

**does not contain** Recherche les enregistrements qui ne contiennent pas la valeur que vous avez indiquée dans n'importe quelle position au sein du champ sélectionné. Par exemple, si vous appliquez un filtre pour la valeur « South » dans le

champ AddressLine1, vous ne voyez pas apparaître d'enregistrement avec « 12 South Ave. », « 9889 Southport St. », « 600 South Shore Dr. » ni « 4089 5th St. South ».

<b>ends with</b>	Cherche les enregistrements finissant par une valeur particulière dans le champ sélectionné. Si par exemple vous appliquez un filtre pour les enregistrements finissant par « burg » dans le champ Ville, vous verrez des enregistrements avec « Gettysburg », « Fredricksburg » et « Blacksburg ».
<b>is greater than</b>	Cherche les enregistrements ayant une valeur numérique supérieure à celle que vous avez indiquée.
<b>is greater than or equal to</b>	Cherche les enregistrements ayant une valeur numérique supérieure ou égale à celle que vous avez indiquée. Par exemple, si vous utilisez 50, vous verrez les enregistrements ayant une valeur de 50 ou une valeur supérieure dans le champ sélectionné.
<b>is less than</b>	Cherche les enregistrements ayant une valeur numérique inférieure à celle que vous avez indiquée
<b>is less than or equal to</b>	Cherche les enregistrements ayant une valeur numérique inférieure ou égale à celle que vous avez indiquée. Par exemple, si vous utilisez 50, vous verrez les enregistrements ayant une valeur de 50 ou une valeur inférieure dans le champ sélectionné.
<b>is after</b>	Recherche les enregistrements ayant une valeur de date ou d'heure ultérieure à celle que vous avez indiquée.
<b>is after or equal to than</b>	Recherche les enregistrements ayant une valeur de date ou d'heure égale ou ultérieure à celle que vous avez indiquée.
<b>is before</b>	Recherche les enregistrements ayant une valeur de date ou d'heure antérieure à celle que vous avez indiquée.
<b>is before or equal to than</b>	Recherche les enregistrements ayant une valeur de date ou d'heure égale ou antérieure à celle que vous avez indiquée.

4. Cliquez sur le bouton **Filtrer** pour appliquer les critères. Seuls les enregistrements dont les données correspondent aux critères de ce champ apparaissent.
5. Cliquez de nouveau sur l'icône de filtre, puis sur **Effacer** pour supprimer le filtre. Sinon, cliquez sur le bouton **Filtre** pour supprimer tous les filtres ; cette action peut être réalisée en mode tabulaire ou en mode formulaire.

### **Affichage des enregistrements**

Vous pouvez afficher les enregistrements d'exception dans deux formats différents. L'affichage par défaut est une vue sous forme de tableau, dans laquelle vous pouvez charger jusqu'à 100 enregistrements d'exception par page. Vous pouvez faire défiler la liste et modifier les enregistrements dans n'importe quel ordre. Si vous modifiez plusieurs enregistrements, vous pouvez enregistrer toutes les modifications en une seule fois ; il est inutile d'enregistrer les modifications



individuellement, pour chaque enregistrement. Si vous utilisez cette vue, vous pouvez spécifier le nombre d'enregistrements que vous souhaitez afficher par page dans la liste déroulante au bas de l'écran.

L'autre méthode d'affichage des enregistrements d'exception est la vue sous forme de formulaire, dans laquelle vous pouvez afficher et modifier un seul enregistrement à la fois ; dans cette vue, vous ne pouvez pas modifier plusieurs enregistrements en même temps. De même, vous devez enregistrer les modifications individuellement, pour chaque enregistrement ; vous ne pouvez pas enregistrer les modifications pour plusieurs enregistrements à la fois.

#### *Affichage des détails d'un enregistrement*

Quelle que soit la vue que vous utilisiez, la grille d'exceptions affiche tous les champs d'un enregistrement, ainsi que son état d'approbation, son type d'exception et tout commentaire ajouté à l'enregistrement. Dans la vue sous forme de tableau, vous pouvez afficher des détails supplémentaires sur un enregistrement en cliquant sur la flèche à l'extrémité gauche d'un enregistrement. Dans la vue sous forme de formulaire, cliquez sur **Afficher les détails**. Ces actions ouvrent l'onglet **Détail**, qui affiche les informations suivantes.

<b>ID de job</b>	Un numéro d'identification assigné au job par le système. Chaque fois qu'un job est exécuté, il se voit attribuer un nouveau numéro de job.
<b>Grouper par</b>	Si la base de données est configurée pour afficher tous les enregistrements du groupe d'enregistrements d'exception, le champ selon lequel les enregistrements sont regroupés s'affiche. Cela ne s'applique qu'aux flux de données ayant effectué de mise en correspondance, tels que les flux de données identifiant les enregistrements doublons ou les flux de données regroupant les enregistrements par ménages.
<b>Heure de l'exception</b>	La date et l'heure correspondant à l'Exception Monitor ayant identifié l'enregistrement comme une exception.
<b>Type d'enregistrement</b>	Le type d'enregistrement que vous avez sélectionné. L'un des éléments suivants : <ul style="list-style-type: none"> <li>• <b>E</b> : exception</li> <li>• <b>GE</b> : exception de groupe</li> <li>• <b>N</b>—Best Of Breed</li> </ul>
<b>État</b>	L'état de l'enregistrement que vous avez sélectionné. L'un des éléments suivants : <ul style="list-style-type: none"> <li>• Nouvelle</li> <li>• Résolu (possible uniquement si l'enregistrement a été modifié au moins une fois depuis sa création)</li> </ul>

Cliquez sur l'onglet **Conditions** pour afficher les informations suivantes :

<b>Condition</b>	Le nom de la condition qui a identifié l'enregistrement comme une exception. Les noms de condition sont définis par la personne ayant installé le flux de données.
<b>Domaine</b>	Le type de données ayant produit une exception. Nom, Adresse et Numéro de téléphone sont des exemples de domaines de données. Ces informations vous aident à identifier quels champs de l'enregistrement demandent à être édités.



**Mesure** Les mesures de qualité que l'enregistrement a ratées. La Précision, l'Intégralité et l'Unicité sont des exemples de métriques de qualité. Ces informations vous aident à définir pourquoi l'enregistrement a été identifié comme constituant une exception.

Chaque fois qu'un enregistrement est modifié, certaines informations sont conservées par le système, indiquant qui a modifié l'enregistrement, lorsqu'il a été modifié, le nom de l'utilisateur auquel l'enregistrement a été attribué et toutes les données qui ont été saisies dans le champ de commentaires de l'enregistrement. L'enregistrement qui s'affiche dans la grille d'exception reflète les toutes dernières modifications et le commentaire le plus récent (le cas échéant) ; en revanche, les informations de l'onglet **Historique** indiquent les informations suivantes pour toute la durée de vie de l'enregistrement, de son ajout initial comme exception dans le référentiel au point auquel vous l'affichez :

<b>Version</b>	Le numéro de révision du changement.
<b>Auteur de la dernière modification</b>	L'utilisateur ayant effectué le changement.
<b>Affecté à</b>	L'utilisateur auquel l'enregistrement d'exception a été attribué au moment de sa révision.
<b>Quand</b>	La date et l'heure à laquelle le changement a été sauvegardé.
<b>Commentaires</b>	Les commentaires (le cas échéant) saisis par la personne ayant effectué les changements.

### *Tri des champs*

Si vous utilisez la vue sous forme de tableau, vous pouvez trier l'ordre des enregistrements affichés en fonction d'un champ spécifique en cliquant n'importe où dans un en-tête de colonne. Par exemple, si vous souhaitez trier les enregistrements dans l'ordre alphabétique par état, il vous suffit de cliquer sur l'en-tête de colonne des états. Un premier clic permet de trier dans l'ordre croissant, tandis qu'un deuxième clic permet de trier dans l'ordre décroissant ; un troisième clic efface l'ordre de tri et rétablit l'ordre de tri des enregistrements avant le tri.

### *Configuration de champs*

Vous pouvez sélectionner les champs qui apparaissent et modifier l'ordre dans lequel ils apparaissent en cliquant sur le bouton **Configurer la vue** (roue dentée sur le côté droit de l'écran sous le menu déroulant utilisateur) et en apportant les modifications souhaitées. Ces modifications sont enregistrées sur le serveur en fonction du nom d'utilisateur et du nom du dataflow ; par conséquent, lorsque vous ouvrez ultérieurement le dataflow, la configuration reste appliquée. De même, les modifications que vous apportez ici affectent également l'affichage lorsque vous modifiez des enregistrements d'exception via la vue Formulaire. Utilisez ces fonctions de Configurer la vue pour personnaliser les champs affichés dans l'Éditeur d'exceptions :

### **Activation de la navigation clavier**

Cochez **Optimiser l'éditeur tabulaire pour la navigation clavier** pour utiliser la touche Tab pour parcourir les enregistrements d'exception dans la vue tabulaire de l'Éditeur d'exceptions.

### **Recherche de champs dans la vue Configurer**

Saisissez tout ou partie d'un nom de champ dans la zone de recherche et la liste des champs disponibles est dynamiquement mise à jour. La recherche n'est pas sensible à la casse.

### Masquage de champs de la vue

Si vous ne souhaitez pas afficher tous les champs d'un enregistrement d'exception, cliquez sur **Configurer la vue** et décochez les champs que vous souhaitez masquer. La liste affichée figure dans le même ordre que celui affiché dans la grille Exceptions.

### Modification de l'ordre des champs

Vous pouvez modifier l'ordre dans lequel les champs sont affichés en faisant glisser des champs et en les déposant de sorte à les placer dans l'ordre de votre choix. Cependant, vous ne pouvez pas réorganiser les champs lors de l'affichage des résultats de recherche. Vous devez effacer la fenêtre de recherche et sélectionner « Tout » pour avoir de nouveau la possibilité de réorganiser les champs.


**Remarque :** Si vous utilisez la vue tabulaire, vous pouvez glisser-déposer les en-têtes de colonne directement à partir de l'Éditeur d'exceptions pour modifier l'ordre dans lequel les champs sont affichés ; vous n'avez pas besoin de le faire depuis Configurer la vue.

### Configurer la disposition Vue sous forme de formulaire

Vous pouvez remplacer la disposition par défaut des champs dans la vue sous forme de formulaire en sélectionnant les colonnes devant apparaître, mais également en spécifiant leur largeur. La seule règle de cette procédure est que les largeurs de colonnes pour tous les champs visibles doivent s'ajouter jusqu'à un total de 12. Par exemple, si vous souhaitez placer quatre champs côte à côte sur une seule ligne, vous pouvez donner à chaque champ une largeur d'une colonne « 3 ». De même, si vous souhaitez inclure seulement deux champs, vous pouvez saisir « 6 » comme la largeur de colonne pour chaque champ. Sinon, si vous souhaitez inclure cinq champs d'adresse sur une ligne, vous devez spécifier un nombre de colonnes pour chacun de ces champs, de telle sorte qu'ils puissent s'ajouter jusqu'à 12. La première image ci-dessous décrit la manière dont vous pouvez configurer cinq champs ; la deuxième image montre un exemple du rendu de cette configuration.


**Remarque :** Lorsque vous remplacez la disposition par défaut pour un champ ou une combinaison de champs, ces champs ne seront plus sensibles à la taille de la fenêtre. Cependant, les champs avec les valeurs par défaut seront toujours réactifs en fonction de la taille de la fenêtre du navigateur.

## Configure view








Filter   All Visible Hidden

Field name	Form columns
<input checked="" type="checkbox"/> Approved	1
<input type="checkbox"/> Type	Default
<input checked="" type="checkbox"/> Address1	4
<input type="checkbox"/> FirmName	Default
<input checked="" type="checkbox"/> City	4
<input checked="" type="checkbox"/> StateProvince	1
<input checked="" type="checkbox"/> PostalCode	2

Optimize tabular editor for keyboard navigation


[Reset to default](#) 

OK Cancel

Approved

Address1	City	StateProvince	PostalCode
1594 Spring St.	Smyrna	<input type="text"/>	30080

[Show detail](#) 

« < > » 1 of 11 items

**Modification des enregistrements d'exception**

L'édition d'un enregistrement d'exception a pour objectif de corriger, d'augmenter ou d'approuver l'enregistrement afin qu'il puisse être traité correctement. L'édition d'un enregistrement d'exception peut demander à faire appel à d'autres services Spectrum Technology Platform ou à des ressources de conseils externes, telles que des cartes, Internet, ou d'autres systèmes d'informations dont votre entreprise dispose.

Après avoir passé en revue les enregistrements, vous pouvez les modifier et les approuver directement dans l'Éditeur d'exceptions. Vous pouvez modifier un ou plusieurs enregistrements à la fois dans la vue sous forme de tableau, et vous pouvez modifier un enregistrement à la fois dans la vue sous forme de formulaire.

Les champs en lecture seule ne peuvent pas être modifiés. Si vous souhaitez qu'un champ en lecture seule puisse être modifié, vous devez supprimer tous les enregistrements d'exceptions de ce flux de données et de cet ID de job et réexécuter le flux de données après avoir configuré les

champs en conséquence dans le stage Write Exceptions. Cette opération produit de nouveaux enregistrements d'exceptions avec des champs modifiables. En outre, vous ne pouvez pas modifier un champ pour qu'il contienne une valeur qui ne correspond pas au type de données. Par exemple, vous ne pouvez pas modifier un champ avec un type de données numérique pour qu'il contienne des lettres.

#### *Vue sous forme de tableau*

Pour modifier un champ d'un seul enregistrement dans la vue sous forme de tableau :

1. Cliquez sur le champ que vous souhaitez modifier et saisissez la valeur de champ de manière appropriée. Les champs en lecture seule sont grisés. Cliquez avec le bouton droit de la souris sur le champ pour accéder aux options, les couper, les copier et les coller. Lorsque vous avez modifié un champ, notez qu'un triangle vert apparaît dans le coin supérieur gauche de ce champ. Il s'agit d'un indicateur visuel vous rappelant que la valeur du champ a été modifiée, mais qu'elle n'a pas encore été enregistrée.
2. Cochez la case **Approuvé** de l'enregistrement modifié. Cela marque l'enregistrement comme étant prêt au traitement par Spectrum™ Technology Platform.
3. Si vous devez annuler une modification, sélectionnez l'enregistrement en question et cliquez sur **Annuler les modifications**.
4. Cliquez sur le bouton **Enregistrer** lorsque vous avez terminé de modifier les enregistrements.

Pour modifier un champ de plusieurs enregistrements dans la vue sous forme de tableau :

1. Tout en appuyant sur la touche Ctrl ou Maj, cliquez sur le champ que vous souhaitez modifier pour tous les enregistrements sélectionnés. Par exemple, si vous souhaitez remplacer toutes les instances de « L.A. » par « Los Angeles », appuyez sur la touche Ctrl et cliquez dans le champ City pour l'un des enregistrements. Ensuite, tout en maintenant la touche Ctrl ou Maj enfoncée, cliquez dans ce même champ sur les autres enregistrements que vous souhaitez modifier.
2. Modifiez les valeurs de champ de manière appropriée. Vous pouvez modifier ces champs, mais sachez que les modifications que vous apportez ici s'appliqueront à tous les enregistrements sélectionnés, même si auparavant les valeurs de ces champs variaient. De même, si vous effacez les données d'un champ lors de la modification de plusieurs enregistrements, celles-ci seront effacées pour tous les enregistrements sélectionnés.
3. Cochez la case **Approuvé** des enregistrements modifiés. Cela marque l'enregistrement comme étant prêt au traitement par Spectrum™ Technology Platform. Sinon, vous pouvez cliquer sur le bouton **Tout approuver** ; ce qui entraînera l'affichage de chaque enregistrement d'exception simple dans l'éditeur à approuver. (La fonction Tout approuver ne s'applique pas à la mise en correspondance des enregistrements d'exception).
4. Si vous devez annuler une modification, sélectionnez les enregistrements en question et cliquez sur **Annuler les modifications**.
5. Cliquez sur le bouton **Enregistrer** lorsque vous avez terminé de modifier les enregistrements. Les modifications apportées aux enregistrements sont enregistrées dans le référentiel d'exceptions et la vue est actualisée. Si vous n'avez pas défini de workflow de revalidation, la liste des exceptions est rechargée. Cependant, il se peut qu'un ou que plusieurs des enregistrements modifiés ne figurent pas dans la liste actualisée, car ils ne correspondent plus

aux critères de recherche/filtre. Si vous avez défini un workflow de revalidation, il se peut que les enregistrements modifiés ne figurent pas dans la liste actualisée s'ils sont désormais valides et ont été supprimés du référentiel.

- Utilisez les boutons de navigation au bas de l'écran pour accéder à la page d'enregistrements d'exception précédente ou suivante ; vous pouvez également utiliser ces boutons pour accéder directement au premier ou au dernier enregistrement d'exception.

#### *Vue sous forme de formulaire*

Pour modifier des enregistrements à l'aide de la vue sous forme de Formulaire, procédez comme suit :

- Cliquez sur **Vue sous forme de formulaire**. Le premier enregistrement du jeu apparaît.
- Cliquez sur le champ que vous souhaitez modifier. Modifiez la valeur du champ de manière appropriée. Les champs en lecture seule sont grisés. Cliquez avec le bouton droit de la souris sur le champ pour accéder aux options, les couper, les copier et les coller. Lorsque vous avez modifié un champ, notez que le contour du champ devient vert. Il s'agit d'un indicateur visuel vous rappelant que la valeur du champ a été modifiée, mais qu'elle n'a pas encore été enregistrée.
- Vous pouvez ajouter des commentaires sur vos changements depuis la colonne **Commentaires**. Les commentaires sont visibles aux autres utilisateurs et peuvent être utilisés pour mesurer et surveiller les changements appliqués à l'enregistrement.
- Lorsque vous jugez avoir appliqué les changements nécessaires afin que l'enregistrement devienne valide, cochez la case **Approuvé**. Cela marque l'enregistrement comme étant prêt au traitement par Spectrum™ Technology Platform.
- Si vous avez besoin d'annuler une modification que vous avez apportée, cliquez sur le bouton **Annuler les modifications**.
- Cliquez sur **Enregistrer**. Les modifications apportées à l'enregistrement sont enregistrées dans le référentiel d'exceptions et la vue est actualisée. Soit vous voyez le même enregistrement contenant vos modifications, soit vous voyez l'enregistrement suivant dans la liste, parce que l'enregistrement que vous avez modifié n'est plus disponible ou qu'il ne correspond plus aux critères de recherche/filtre.
- Utilisez les boutons de navigation au bas de l'écran pour accéder à l'enregistrement précédent ou suivant ; vous pouvez également utiliser ces boutons pour accéder directement au premier ou au dernier enregistrement d'exception.

#### **Résolution des enregistrements doublons**

Les exceptions de résolution doublons surviennent lorsque Spectrum™ Technology Platform ne parvient pas à définir avec certitude si un enregistrement est un doublon d'un autre. Il existe deux manières de résoudre les enregistrements doublons.

**Remarque :** Les enregistrements en double peuvent uniquement être résolus avec la fonction Résoudre les doublons de la vue sous forme de tableau. En revanche, vous pouvez toujours modifier ces enregistrements dans la vue sous forme de formulaire.

La première consiste à regrouper les enregistrements doublons dans des collections. Une fois que vous avez approuvé les enregistrements, ils peuvent être traités via une procédure de consolidation afin d'éliminer les enregistrements doublons dans chaque collection de vos données.

Une autre approche consiste à modifier les enregistrements afin qu'ils deviennent plus facilement reconnaissables comme étant des doublons, par exemple, en corrigeant l'orthographe d'un nom de rue. Lorsque vous approuvez les enregistrements, Spectrum™ Technology Platform traite à nouveau les enregistrements en suivant un processus de mise en correspondance et de consolidation. Si vous réussissez à corriger les enregistrements, Spectrum™ Technology Platform peut identifier l'enregistrement comme étant un doublon.

Une autre approche pour résoudre les enregistrements doublons est de créer un enregistrement Best Of Breed. Cela combine les deux autres approches en gérant les collections d'enregistrements, puis en modifiant l'un des enregistrements de la collection pour inclure les champs des deux enregistrements original et doublon. Cet enregistrement s'appelle alors enregistrement Best Of Breed.

#### *Faire d'un enregistrement un doublon d'un autre*

Les enregistrements doublons s'affichent comme des groupes d'enregistrements dans le Business Steward Portal. Vous pouvez faire d'un enregistrement un doublon d'un autre en le plaçant dans le même groupe que l'enregistrement doublon.

Pour faire d'un enregistrement un doublon :

1. Choisissez l'enregistrement sur lequel vous souhaitez travailler, puis cliquez sur **Résoudre les doublons**.

L'affichage **Résolution de doublon** montre les enregistrements doublons. Les enregistrements sont regroupés dans des collections ou des groupes candidats contenant ces types d'enregistrement de correspondance :

<b>suspect</b>	Un enregistrement auquel les autres enregistrements sont comparés afin de déterminer s'ils sont des doublons les uns des autres. Chaque collection possède un seul enregistrement suspect.
<b>duplicate</b>	Un enregistrement qui est un doublon de l'enregistrement suspect.
<b>unique</b>	Un enregistrement qui n'a aucun doublon.

Vous pouvez définir le type d'un enregistrement en lisant la colonne MatchRecordType.

2. Si besoin est, corrigez les enregistrements individuels de la manière que vous voulez. Pour plus d'informations, reportez-vous à la section [Modification des enregistrements d'exception](#) à la page 275. Sinon, vous pouvez glisser- déposer des enregistrements entre des groupes.
3. Dans le champ Numerodecollection, ou Groupecandidat, saisissez le numéro du groupe vers lequel vous souhaitez déplacer l'enregistrement. L'enregistrement devient un doublon des autres enregistrements du groupe.

Dans certains cas, vous ne pouvez pas déplacer un enregistrement avec une valeur Typedenregistrementdecorrespondance « suspect » vers une autre collection de doublons.

**Remarque :** Les enregistrements sont regroupés soit par le champ Numerodecollection ou le champ Groupecandidat, en fonction du type de logique de mise en correspondance employée dans le dataflow produisant les exceptions. Contactez votre administrateur Spectrum™ Technology Platform si vous voulez obtenir plus d'informations sur la mise en correspondance.

4. Lorsque vous avez terminé de modifier des enregistrements, cochez la case **Approuvé**. Cela signifie que l'enregistrement est prêt à être re-traité par Spectrum™ Technology Platform.
5. Pour enregistrer vos modifications, cliquez sur le bouton **Enregistrer**.

#### *Création d'un nouveau groupe d'enregistrements doublons*

Dans certaines situations, vous pouvez créer un nouveau groupe d'enregistrements pour les enregistrements que vous voulez dupliquer. Dans d'autres situations, créer de nouveaux groupes vous est impossible. Votre capacité à créer de nouveaux groupes est déterminée par le type de traitement de Spectrum™ Technology Platform qui a créé les enregistrements d'exception.

1. Choisissez l'enregistrement sur lequel vous souhaitez travailler, puis cliquez sur **Résoudre les doublons**.

L'affichage **Résolution de doublon** montre les enregistrements doublons. Les enregistrements sont regroupés dans des collections ou des groupes candidats contenant ces types d'enregistrement de correspondance :

<b>suspect</b>	Un enregistrement auquel les autres enregistrements sont comparés afin de déterminer s'ils sont des doublons les uns des autres. Chaque collection possède un seul enregistrement suspect.
<b>duplicate</b>	Un enregistrement qui est un doublon de l'enregistrement suspect.
<b>unique</b>	Un enregistrement qui n'a aucun doublon.

Vous pouvez définir le type d'un enregistrement en lisant la colonne MatchRecordType.

2. Si besoin est, corrigez les enregistrements individuels de la manière que vous voulez. Pour plus d'informations, reportez-vous à la section **Modification des enregistrements d'exception** à la page 275.
3. Sélectionnez un enregistrement que vous voulez ajouter à la nouvelle collection, puis cliquez sur **Nouvelle collection**. Un numéro unique de collection est automatiquement attribué à la nouvelle collection, et l'enregistrement que vous avez sélectionné devient un enregistrement suspect.

**Remarque :** Si vous ne voyez pas le bouton **Nouvelle collection**, vous ne pouvez pas créer une nouvelle collection pour les enregistrements sur lesquels vous travaillez. Vous ne pouvez créer de nouvelles collections que si le flux de données qui a produit les exceptions contenait le stage Interflow Match ou Intraflow Match, mais pas si celui-ci contenait le stage Transactional Match. Contactez l'administrateur de Spectrum™ Technology Platform si vous voulez obtenir plus d'informations sur ces stages de correspondance.



4. Ajoutez des enregistrements à la collection en entrant le numéro de la nouvelle collection dans le champ CollectionNumber de l'enregistrement.
5. Lorsque vous avez terminé de modifier des enregistrements, cochez la case **Approuvé**. Cela signifie que l'enregistrement est prêt à être re-traité par Spectrum™ Technology Platform.
6. Pour enregistrer vos modifications, cliquez sur le bouton **Enregistrer**.

#### *Rendre un enregistrement unique*

Pour faire d'un doublon un enregistrement unique.

1. Dans le champ MatchRecordType, entrez « Unique ».
2. Lorsque vous avez terminé de modifier des enregistrements, cochez la case **Approuvé**. Cela signifie que l'enregistrement est prêt à être re-traité par Spectrum™ Technology Platform.
3. Pour enregistrer vos modifications, cliquez sur le bouton **Enregistrer**.

#### *Champs automatiquement ajustés pendant la résolution des doublons*

Quand vous modifiez des enregistrements dans l'affichage Résolution de doublon du Business Steward Portal, certains champs sont automatiquement ajustés pour prendre en compte la nouvelle disposition de l'enregistrement.

**Tableau 21 : Enregistrements traités par Interflow Match et intra-flux**

Action	Valeurs automatiquement appliquées aux champs
Déplacement d'un enregistrement d'une collection à une autre	<p><b>Si vous introduisez un enregistrement dans une collection de doublons :</b></p> <ul style="list-style-type: none"> <li>• MatchRecordType : Doublon</li> <li>• MatchScore : 100</li> <li>• HasDuplicates : D (ce champ n'est présent que si le flux de données contenait un stage Interflow Match.)</li> </ul> <p><b>Si vous introduisez un enregistrement doublon dans une collection d'enregistrements uniques (collection 0) :</b></p> <ul style="list-style-type: none"> <li>• MatchRecordType : Unique</li> <li>• MatchScore : Aucune modification</li> <li>• HasDuplicates : U (ce champ n'est présent que si le flux de données contenait un stage Interflow Match.)</li> </ul> <p><b>Si vous introduisez un enregistrement suspect dans une collection d'enregistrements uniques (collection 0) :</b></p> <ul style="list-style-type: none"> <li>• MatchRecordType : Unique</li> <li>• MatchScore : 0</li> <li>• HasDuplicates : N (ce champ n'est présent que si le flux de données contenait un stage Interflow Match.)</li> </ul>



Action	Valeurs automatiquement appliquées aux champs
Création d'une nouvelle collection	<ul style="list-style-type: none"> <li>• MatchRecordType : Suspect</li> <li>• MatchScore : Aucune valeur</li> <li>• HasDuplicates : Y (ce champ n'est présent que si le flux de données contenait un stage Interflow Match.)</li> </ul> <p><b>Remarque :</b> Si l'enregistrement provient d'un flux de données qui contenait un stage Interflow Match, seuls les enregistrements qui ont une valeur de « input_port_0 » dans le champ InterflowSourceType peuvent être un enregistrement suspect.</p>

**Tableau 22 : Enregistrements traités par Transactional Match.**


Action	Valeurs automatiquement appliquées aux champs
Modifier MatchRecordType en valeur Duplicate	<ul style="list-style-type: none"> <li>• HasDuplicates : D</li> <li>• MatchScore : 100</li> </ul>
Réglez MatchRecordType sur Unique	<ul style="list-style-type: none"> <li>• HasDuplicates : U</li> <li>• MatchScore : inchangé</li> </ul>
Réglez HasDuplicates sur D	<ul style="list-style-type: none"> <li>• MatchRecordType : Doublon</li> <li>• MatchScore : 100</li> </ul>
Réglez HasDuplicates sur U	<ul style="list-style-type: none"> <li>• MatchRecordType : Unique</li> <li>• MatchScore : inchangé</li> </ul>
Réglez HasDuplicates sur Y	<ul style="list-style-type: none"> <li>• MatchRecordType : Suspect</li> <li>• MatchScore : vide</li> </ul>
Réglez HasDuplicates sur N	<ul style="list-style-type: none"> <li>• MatchRecordType : Suspect</li> <li>• MatchScore : vide</li> </ul>

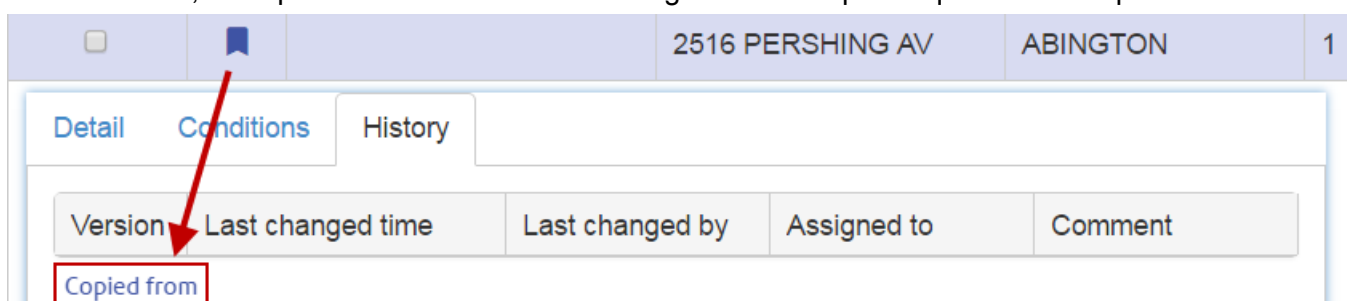
***Création d'un enregistrement Best of Breed***


Pour créer un enregistrement Best Of Breed, procédez comme suit.

**Remarque :** Les enregistrements Best Of Breed peuvent uniquement être ajoutées à des collections non zéro. De plus, un enregistrement Best Of Breed doit être créé pour chaque collection d'un groupe avant que vous puissiez améliorer l'ensemble des enregistrements de ce groupe.

1. Pour créer un enregistrement Best Of Breed dans un groupe, sélectionnez un enregistrement

de la collection, puis cliquez sur le bouton **Ajouter un enregistrement Best Of Breed** (). Cela copie l'enregistrement que vous avez sélectionné et le place juste après cet enregistrement dans Exception Editor. Lorsque vous enregistrez l'enregistrement Best Of Breed, il est déplacé au bas de la collection d'enregistrements. Si vous examinez l'historique des enregistrements Best Of Breed, vous pouvez voir un lien vers l'enregistrement depuis lequel il a été copié :



**Remarque :** Vous pouvez annuler cette opération en sélectionnant l'enregistrement Best Of Breed, puis en cliquant sur le bouton **Supprimer un enregistrement Best Of Breed** (.

2. Pour combiner des champs dans l'enregistrement Best Of Breed, glissez et déposez les champs des enregistrements doublons dans l'enregistrement Best Of Breed. Notez qu'il existe certaines limites quant aux opérations glisser-déposer dans cet environnement :

- La cellule depuis laquelle vous glissez un enregistrement vers une autre ne peut pas être à l'état « edit » (modifier) (même si la cellule dans laquelle vous déposez l'enregistrement peut l'être).
- Lorsque vous déposez une cellule sur une autre cellule (que ce soit sur la même ligne ou sur une autre ligne), la deuxième cellule prend la valeur de la première cellule.
- Vous ne pouvez pas glisser-déposer des cellules de type booléen.
- Vous pouvez glisser des cellules de type non chaîne sur des cellules de type chaîne (telles que des champs de date), mais vous ne pouvez pas déposer des cellules de type chaîne sur des cellules de type non chaîne.
- Vous ne pouvez pas déposer une cellule sur une autre cellule qui utilise des valeurs de recherche (comme des éditeurs de cases de combinaison), mais vous pouvez déposer une valeur de ces cellules sur d'autres cellules de type chaîne.
- Vous ne pouvez pas déplacer une cellule sur une cellule en lecture seule.

Reportez-vous aux options **Write Exceptions** pour obtenir davantage d'informations sur la fonctionnalité Best Of Breed.

### Utilisation des outils de recherche

L'Éditeur d'exceptions du Business Steward Portal fournit des outils de recherche qui vous aident à rechercher des informations susceptibles de vous permettre de modifier et d'approuver des enregistrements d'exception et de les réexécuter correctement. Les outils incluent Bing Maps et les services pour lesquels vous disposez d'une licence dans Spectrum™ Technology Platform. Les outils avec des icônes qui ont un signe plus au centre sont des outils premium et entraînent des frais lorsqu'il sont utilisés.

**Remarque :** Ces outils de recherche doivent d'abord être configurés sur la page Paramètres de Business Steward de Management Console. Si aucun outil n'est sélectionné sur cette page, aucun n'y apparaît. En outre, quels que soient les outils sélectionnés dans Management Console, les services qui apparaissent à l'utilisateur ici sont limités à ceux pour lesquels il dispose de droits d'affichage.

### Utilisation de Bing Maps

Bing Maps affiche l'emplacement d'une adresse sur une carte et fournit des commandes qui vous permettent de zoomer sur la carte et de la déplacer. De plus, vous pouvez cliquer sur la carte pour obtenir des adresses.

**Remarque :** L'outil de recherche Bing Maps est fourni par Microsoft ; pour utiliser ce service, vous devez être connecté à Internet.

1. Cliquez sur l'enregistrement que vous souhaitez rechercher.
2. Sous le tableau des enregistrements, cliquez sur l'onglet **Outils de recherche** pour développer la vue.
3. Dans le champ **Service**, sélectionnez **Bing Maps**.
4. Sélectionnez les champs à utiliser dans votre recherche dans la colonne **Nom de champ**. Par exemple, si vous souhaitez rechercher une adresse sur une carte, vous pouvez sélectionner Ligne d'adresse 1 et Ville. Si vous souhaitez afficher uniquement la ville sur une carte, sélectionnez uniquement City et StateProvince.
5. Sélectionnez « Road » pour une vue de carte traditionnelle, « Bird's Eye » pour une vue aérienne ou « Automatic » pour laisser Bing Maps décider de la vue la plus appropriée.
6. Cliquez sur **Activer**. Les résultats, y compris la latitude et la longitude, sont affichés dans la case **Résultats** et sur la carte sous forme de punaise. Cliquez sur les boutons de rotation pour déplacer le point de vue de 90 degrés. Cliquez sur les flèches de la boussole pour déplacer la mise au point par incréments dans le sens sélectionné. Utilisez les boutons Zoom avant et Zoom arrière pour effectuer une mise au point plus proche ou plus lointaine. Vous pouvez également déplacer la punaise en la faisant glisser vers un nouvel emplacement ; les informations de carte se mettent à jour dynamiquement.
7. Pour obtenir l'adresse d'autres bâtiments, cliquez sur la carte. Il peut être utile de passer à la vue Bird's Eye lorsque vous recherchez des bâtiments.

Après avoir effectué une recherche initiale sur la carte, vous pouvez cliquer sur un autre enregistrement d'exception et sur le bouton Accéder, et la carte est mise à jour en conséquence.

### Utilisation des outils de recherche du service Spectrum

Les outils de recherche de services Pitney Bowes incluent tous les services dont vous possédez une licence, comme `ValidateAddress`, `GetPostalCodes`, etc. Vous pouvez utiliser ces services dans l'éditeur d'exceptions pour rechercher et valider des données d'exception que vous tentez de corriger ou d'augmenter.

Notez que lorsque vous utilisez cette fonctionnalité, vous ne voyez les services que si vous disposez d'autorisations d'affichage sur les services sous le groupe de plate-forme de sécurité des rôles. De même, afin d'exécuter des services, vous devez disposer d'autorisation d'exécution sur les services. Cependant, il est possible de modifier ces autorisations à l'aide de substitutions d'entité sécurisée. Grâce à une combinaison d'autorisations de niveau supérieur et de substitutions, l'administrateur peut gérer la liste des services auquel un utilisateur ou un rôle donné peut accéder dans la liste déroulante de services du portail BSM.

1. Sélectionnez l'enregistrement contenant les données que vous souhaitez rechercher.
2. Sous l'Éditeur d'exceptions, cliquez sur **Outils de recherche**.
3. Dans le champ **Service**, sélectionnez le service à utiliser, tel que `ValidateAddress` ou `GetCandidateAddresses`.
4. Si l'enregistrement d'exception contient des champs utilisés dans ce service, mais sous des noms différents, mappez les champs d'entrée du service vers les champs d'exception dans l'onglet **Champ de service**. Par exemple, si vous utilisez `ValidateAddress` et si votre enregistrement d'exception n'inclut pas de champ `AddressLine1`, mais qu'il inclut en revanche un champ `AddrLine1`, sélectionnez « `AddrLine1` » dans la colonne Champ d'exception de la ligne `AddressLine1`. Vous devez avoir au moins un champ d'entrée mappé avant d'exécuter le service.

**Remarque :** Business Steward Portal mémorise les mappages que vous créez depuis des champs de service vers des champs d'exception, à condition que vous mappiez les enregistrements d'exception avec les mêmes noms de champ, de stage et de flux de données. Par exemple, si votre enregistrement d'exception comporte un champ nommé « `AddrLine1` » et si vous le mappez vers « `AddressLine1` », il mémorise ce mappage à condition que vous travailliez avec des enregistrements contenant « `AddrLine1` » et qui ont été créés dans le même stage par le même flux de données.

5. Répétez l'étape 4 pour chaque champ de sortie de service. Cette étape est facultative, mais vous devez disposer d'au moins un champ de sortie mappé avant de pouvoir appliquer les données de service. Notez que vous ne pouvez pas mapper un champ d'exception en lecture seule avec les champs de sortie de service.
6. Cliquez sur l'onglet **Options** pour sélectionner les ressources de base de données et pour afficher et modifier les options de service qui ont été définies dans Management Console. Si vous ne connaissez pas le but d'une option spécifique, survolez l'option à l'aide de la souris pour en afficher la description. Les modifications apportées ici persistent lorsqu'elles sont utilisées par le même utilisateur, le même flux de données et le même stage de l'enregistrement d'exception.

**Remarque :** Si le service que vous utilisez nécessite une base de données, vous devez avoir configuré la ressource de base de données dans Management Console. Par exemple, si vous consultez des enregistrements américains à l'aide de Validate Address, vous devez avoir configuré une base de données aux États-Unis dans Management Console.

7. Cliquez sur **Exécuter le service**. L'enregistrement actualisé apparaît dans l'onglet **Résultat** avec le code d'état indiquant que l'enregistrement a été correctement effectué. Les champs qui sont mappés vers les enregistrements d'exception seront désignés par un astérisque.
8. Sélectionnez l'enregistrement du résultat et cliquez sur **Appliquer les données de service** pour transférer les données aux champs mappés de l'enregistrement d'exception.
9. Si vous souhaitez retraiter l'enregistrement actualisé, cochez la case **Approuvé** de cet enregistrement, puis cliquez sur **Enregistrer**.

## Page Gérer

La page Gérer permet à un utilisateur disposant des droits d'affichage et d'édition de réviser et de gérer l'activité d'enregistrement des exceptions pour tous les bénéficiaires. Elle permet également de réaffecter les enregistrements d'exception d'un utilisateur à un autre. Si vous disposez de droits de suppression, vous pouvez supprimer le groupe d'enregistrements d'exception tout entier du système en fonction du nom du flux de données et de l'ID du job.

### Vérification de l'activité des enregistrements d'exception

La section **Statut** de la page Gérer les exceptions indique l'activité d'enregistrement des exceptions par affectation ou nom de flux de données. Vous pouvez spécifier le type d'affichage en cliquant sur le bouton « Affectations » ou « Flux de données » dans le coin supérieur droit de l'écran. La vue « Affectations » fournit le nombre d'enregistrements d'exception affectés à chaque utilisateur ainsi que le nombre d'enregistrements approuvés et le nombre d'enregistrements restant. Elle affiche également le nombre d'enregistrements hors exception existant dans les groupes qui contiennent également des enregistrements d'exception. Notez que les données hors exception apparaissent uniquement s'il existe des enregistrements hors exception dans un groupe contenant également des enregistrements d'exception et s'il n'existe aucune substitution de sécurité empêchant l'affichage de ces données.

La vue « Flux de données » fournit le nombre et le pourcentage d'enregistrements qui ont été approuvés pour chaque flux de données. Pour les flux de données contenant plus d'un stage Exception Monitor, vous pouvez décomposer davantage ces informations pour chacun de ces stages. Par exemple, si un flux de données contient deux stages Exception Monitor, « NameExceptions » peut être chargé de 37 % de toutes les exceptions du flux de données, et « DataExceptions » des 63 % restants. Dans ce cas, l'état de chaque **flux de données** est toujours calculé comme un tout, car il fait référence à tous les flux de données du référentiel, tandis que l'état de chaque **stage** est calculé en fonction du nombre total d'exceptions du flux de données sélectionné.

De plus, vous pouvez filtrer les informations qui s'affichent en saisissant des critères de recherche dans la ligne Filtre. La liste est automatiquement et dynamiquement renseignée avec les flux de données ou les bénéficiaires dont le nom correspond aux lettres que vous saisissez.

La section **Progression** de cette page indique la progression cumulée pour tous les utilisateurs ou tous les dataflows dont les états individuels sont affichés sur le tableau de bord. Sélectionnez « Affectations » en haut de l'écran pour afficher la progression de l'utilisateur ; sélectionnez « Dataflows » pour afficher la progression du flux de données. L'échelle permet de choisir la mesure de l'échelle (jour, semaine, mois) et le nombre d'unités à afficher.

### Affectation des enregistrements d'exception

La section **Affectation** de la page Gérer les exceptions (sous l'onglet **Maintenance**), vous permet de réaffecter les enregistrements d'exception d'un utilisateur à l'autre.

1. Sélectionnez un utilisateur dont vous souhaitez affecter les exceptions à un autre utilisateur dans le champ **Utilisateur**.
2. Pour réaffecter tous les enregistrements d'exception appartenant à l'utilisateur, passez à l'étape 4. Pour réaffecter une partie des enregistrements d'exception d'un utilisateur, renseignez un ou plusieurs de ces champs :
  - Nom du flux de données : nom du flux de données produisant les enregistrements d'exception.
  - Nom du stage : nom du stage produisant les enregistrements d'exception.
  - ID de job : identifiant affecté au job contenant les enregistrements d'exception.
  - Domaine de données : genre de **données** affectées dans Exception Monitor.
  - Métriques de qualité : genre de **métrique** affectée dans Exception Monitor.
  - Date/heure de début : date de début dans une plage de dates au cours de laquelle les enregistrements d'exception ont été créés.
  - Date/l'heure de fin : date de fin dans une plage de dates au cours de laquelle les enregistrements d'exception ont été créés.
  - Statut d'approbation : indique si les enregistrements d'exception ont été approuvés ou non.
3. Après avoir effectué des sélections dans les champs **Utilisateur**, **Nom de flux de données** et **Libellé de stage** (au minimum), vous pouvez affiner encore davantage le filtre en fonction de valeurs de champ d'exception.
  - a) Cliquez sur l'icône Ajouter un filtre de champ :
  - b) Dans la colonne **Nom du champ**, choisissez le champ que vous voulez utiliser pour le filtre.
  - c) Dans la colonne **Opérateur**, sélectionnez l'une des options suivantes :
 

<b>is equal to</b>	Cherche les enregistrements possédant la valeur exacte que vous avez spécifiée. Il peut s'agir d'une valeur numérique ou d'une valeur textuelle. Par exemple, vous pouvez rechercher des enregistrements avec une valeur MatchScore de 82 exactement, ou des enregistrements avec une valeur LastName « Smith ».
<b>is not equal to</b>	Cherche les enregistrements ayant n'importe quelle valeur autre que celle que vous avez indiquée. Il peut s'agir d'une valeur numérique ou

d'une valeur textuelle. Par exemple, vous pouvez rechercher les enregistrements avec toute valeur MatchScore sauf 100, ou les enregistrements avec toute valeur LastName sauf « Smith ».

- is greater than** Cherche les enregistrements ayant une valeur numérique supérieure à celle que vous avez indiquée.
- is greater than or equal to** Cherche les enregistrements ayant une valeur numérique supérieure ou égale à celle que vous avez indiquée. Par exemple, si vous utilisez 50, vous verrez les enregistrements ayant une valeur de 50 ou une valeur supérieure dans le champ sélectionné.
- is less than** Cherche les enregistrements ayant une valeur numérique inférieure à celle que vous avez indiquée
- is less than or equal to** Cherche les enregistrements ayant une valeur numérique inférieure ou égale à celle que vous avez indiquée. Par exemple, si vous utilisez 50, vous verrez les enregistrements ayant une valeur de 50 ou une valeur inférieure dans le champ sélectionné.
- contains** Cherche les enregistrements contenant la valeur que vous avez indiquée selon n'importe quelle position dans le champ sélectionné. Par exemple, si vous appliquez un filtre pour la valeur « South » dans le champ Ligne d'adresse 1, vous verrez des enregistrements avec « 12 South Ave. », « 9889 Southport St. », « 600 South Shore Dr. » et « 4089 5th St. South ».
- does not contain** Recherche les enregistrements qui ne contiennent pas la valeur que vous avez indiquée dans n'importe quelle position au sein du champ sélectionné. Par exemple, si vous appliquez un filtre pour la valeur « South » dans le champ Ligne d'adresse 1, mais que vous sélectionnez « Ne contient pas », vous ne verrez pas apparaître les enregistrements contenant « 12 South Ave. », « 9889 Southport St. », « 600 South Shore Dr. » ni « 4089 5th St. South ».
- starts with** Cherche les enregistrements commençant par une valeur particulière dans le champ sélectionné. Par exemple, si vous appliquez un filtre pour la valeur « Van » dans le champ Nomdefamille, vous verrez des enregistrements avec « Van Buren », « Vandenburg » ou « Van Dyck ».
- ends with** Cherche les enregistrements finissant par une valeur particulière dans le champ sélectionné. Si par exemple vous appliquez un filtre pour les enregistrements finissant par « burg » dans le champ Ville, vous verrez des enregistrements avec « Gettysburg », « Fredricksburg » et « Blacksburg ».

- d) Dans la colonne **Valeur du champ**, saisissez la valeur que vous voulez employer comme critère de filtrage.

**Remarque :** La valeur de recherche est sensible à la casse. Cela signifie que les recherches de la valeur SMITH ne renvoient que les enregistrements avec « SMITH » tout en majuscule, mais pas « smith » ni « Smith ».



- e) Pour filtrer en utilisant plus d'un filtre, ajoutez plusieurs filtres en cliquant de nouveau sur l'icône d'ajout de filtre de champ. Par exemple, si vous souhaitez que tous les enregistrements ayant une valeur LastName de « Smith » et une valeur État de « NY », vous pouvez utiliser deux filtres, un pour le champ LastName et l'autre pour le champ État.

Cet exemple affichera tous les enregistrements avec la valeur « FL » saisie dans le champ État

:

Field name	Operator	Value
State	Is equal to	FL

Cet exemple affichera tous les enregistrements dont la valeur Codepostal n'est pas de 60510

:

Field name	Operator	Value
StateProvince	Is equal to	NY
PostalCode	Is not equal to	14226

Cet exemple renverra tous les enregistrements dotés d'un élément StateProvince "NY" avec tous les codes postaux, excepté 14226.

:

Field name	Operator	Value
PostalCode	Is not equal to	60510

4. Cliquez sur **Réaffecter**.
5. Facultatif. Sélectionnez le nombre d'exceptions à réaffecter. Vous pouvez affecter tout ou partie des exceptions obtenues au nouvel utilisateur. Par exemple, si vous saisissez « 10 » comme limite, seuls les 10 premiers enregistrements répondant aux critères seront réaffectés et les enregistrements restants répondant aux critères ne seront pas réaffectés.
6. Sélectionnez un autre utilisateur dans la liste déroulante **Réaffecter**.
7. Cliquez sur **Confirmer**.

### Suppression des enregistrements d'exception

Il peut arriver que vous souhaitiez supprimer des enregistrements d'exception du référentiel. Par exemple, vous pouvez disposer d'enregistrements résiduels suite à un test du système ou d'enregistrements qui ont été considérés à tort comme des exceptions après traitement ; ou encore, vous pouvez souhaiter traiter et supprimer des enregistrements approuvés en premier, puis réexécuter le même job. La section **Purger** de la page Gérer (sous l'onglet **Maintenance**) vous permet de procéder à cette opération.

Vous devez effectuer des sélections à partir des champs **Dataflow name** et **Job ID** avant de cliquer sur **Remove**. Cependant, vous pouvez sélectionner Tout dans le champ Job ID pour supprimer les enregistrements d'exception de chaque job exécuté par le flux de données sélectionné. Cliquez sur **Supprimer les données de rapport qualité de données** pour supprimer toutes les données de performances. Si cette option n'est pas sélectionnée, les enregistrements d'exception du job sont



supprimés du référentiel, mais les données de performances continuent à apparaître sur la page Qualité des données.

### Page de qualité des données

La page Qualité des données du portail Business Steward fournit des informations sur les tendances de vos enregistrements d'exception.

**Remarque :** Cette page est désactivée si vous désactivez le rapport sur la qualité des données sur la page Paramètres de Business Steward de Management Console.

### Identification des tendances

La page Tendances de Qualité des données décrit les informations statistiques suivantes par flux de données et par domaine :

- Nombre total de conditions traitées
- Nombre total de conditions d'exception
- Pourcentage de conditions ayant été correctement traitées
- Tendances de vos données à intervalles de 30 jours

Ces informations peuvent être ventilées encore davantage par stage si vous sélectionnez un flux de données ou par métrique si vous sélectionnez un domaine. Les valeurs qui apparaissent ici sont déterminées par les flux de données contenant des stages Exception Monitor qui ont été exécutés au cours des 30 derniers jours, ainsi que par leurs domaines et métriques respectifs.

- Si vous affichez des informations par domaine, sélectionnez un **Nom de dataflow** si vous souhaitez afficher les informations d'un flux de données spécifique, et un **Libellé de stage** si vous souhaitez afficher les informations d'un stage spécifique. Notez que vous devez sélectionner un seul flux de données si vous souhaitez également filtrer les résultats en fonction d'un stage, sinon, vous verrez s'afficher les données de tous les flux de données.
- Si vous affichez des informations par flux de données, sélectionnez un **Domaine** si vous souhaitez afficher les informations d'un domaine spécifique, et une **Métrique** si vous souhaitez afficher les informations d'une métrique spécifique. Sinon, vous verrez les données de tous les domaines.
- Sélectionnez une durée pour **Échelle** pour indiquer jusqu'à quelle date vous souhaitez remonter. Vous pouvez sélectionner n'importe quel nombre de jours, semaines ou mois.
- Sélectionnez un flux de données si vous souhaitez ventiler les résultats par stage.
- Sélectionnez un domaine si vous souhaitez ventiler les résultats par métrique de qualité des données.

À mesure que vous effectuez des sélections pour afficher différentes parties des données, vous constaterez que des graphiques apparaissent, reflétant l'état des données en question. Par exemple, si vous examinez les données des domaines, vous pouvez affiner ces données par domaine spécifique (tel que « Address »), puis par métrique spécifique (telle que « Completeness »). Les graphiques sont mis à jour en conséquence, à mesure que vous effectuez différentes sélections.

## Business Steward Settings

### Introduction

#### Introduction aux paramètres de Business Steward

Les paramètres Business Steward fournissent les outils suivants pour les utilisateurs ayant des autorisations d'écriture :

- **Recherches** : fournit une méthode de correction des enregistrements d'exceptions à l'aide d'un jeu de données spécifique pour ces corrections.
- **Domaines** : vous permet de spécifier le type de données évaluées.
- **Métriques** : vous permet de spécifier la manière dont les données sont mesurées.
- **Notifications** : permet au système d'envoyer un message à une ou plusieurs adresses électroniques lorsqu'un nombre désigné d'exceptions est lié à un domaine ou à une mesure spécifique.
- **Rapport sur la qualité des données** : définit les préférences de suivi des conditions de réussite / d'échec et des IPC.
- **Services Outils de recherche** : définissent les préférences pour les outils de recherche dans Exception Editor du portail Business Steward.
- **Options** : définit les préférences de suivi des journaux d'audit et de la progression.

#### Accès aux paramètres Business Steward

Pour accéder aux paramètres Business Steward :

1. Dans un navigateur Web, accédez à l'URL suivante :

`http://server.port/managementconsole`

Où : *server* est l'adresse IP ou le nom de serveur de votre serveur Spectrum™ Technology Platform et *port* le port HTTP utilisé par Spectrum™ Technology Platform. Par défaut, le port HTTP est 8080.

2. Saisissez un nom d'utilisateur et un mot de passe valides.
3. Cliquez sur le bouton **Ressources**.
4. Sélectionnez **Paramètres Business Steward**.

### Recherches

L'outil Recherches fournit un moyen de choisir parmi une liste de valeurs, un champ spécifique lors de la mise à jour des enregistrements dans l'éditeur d'exceptions. Cette fonctionnalité est particulièrement utile lorsque vous souhaitez modifier plusieurs enregistrements comportant des données dans le même champ. Par exemple, vous pouvez disposer d'un ensemble d'enregistrements d'exception qui contiennent des données bancaires. Un des champs de ces données peut se

composer des codes qui représentent le type de compte lié à l'enregistrement (1 = vérification, 2 = épargne, 3 = marché monétaire et ainsi de suite). dont les adresses comprennent des codes ISO au lieu de noms dans le champ Pays, empêchant ainsi ces adresses d'être validées. Pour résoudre ce problème, vous pourriez créer une recherche qui fournit des codes ISO et leurs noms de pays respectifs et effectuer les corrections dans Exception Editor, où vous sélectionneriez le code ISO à partir d'une liste, pour remplir ensuite le champ avec le nom de pays lié à ce code ISO.

Un autre avantage de cet outil est qu'il limite les options disponibles pour les corrections, ce qui réduit le risque d'erreur supplémentaire. Prenons le même exemple avec des noms de pays incorrects ou manquants, en créant une recherche qui fournit une liste de noms de pays au lieu de demander une saisie manuelle de ces noms pour chaque enregistrement d'exception, vous vous assurez ainsi que ces noms sont correctement orthographiés et sont plus susceptibles d'être validés lorsqu'ils sont retraités.

### **Qu'est-ce que le processus de recherche ?**

Le processus de recherche comporte trois étapes après avoir consulté les exceptions et identifié un problème périodique parmi ces exceptions (tels que des données non valides dans un champ de pays) :

- Créez la recherche à l'aide des valeurs et/ou des libellés de données exactes qui écraseront les données incorrectes.
- À l'aide du stage Write Exceptions dans le flux de données qui produit les enregistrements d'exceptions, pointez le champ problématique de la recherche que vous avez créée, et réexécutez le flux de données.
- Corrigez les enregistrements d'exceptions dans l'éditeur d'exceptions du portail Business Steward, en écrasant les données erronées du champ problématique avec les données correctes à partir de la recherche.

### **Création de recherches**

Une recherche est constituée de valeurs ou de paires valeur/libellé qui contiennent les données destinées à remplacer les données existantes, problématiques dans le flux de données produisant les enregistrements d'exceptions. La valeur est ce qui remplacera les données problématiques et le libellé est ce qui est affiché dans une liste que vous sélectionnez lors de l'utilisation de la table de recherche pour corriger les enregistrements dans Exception Editor.

**Remarque :** Si vous incluez uniquement les valeurs dans votre table de recherche, les valeurs seront également utilisées comme des libellés.

Vous pouvez renseigner une recherche en entrant manuellement les informations ou en les copiant à partir d'une source externe et en les collant dans la boîte de dialogue **Ajouter plusieurs**. La source externe peut être une feuille de calcul, un fichier texte ou n'importe quel autre fichier tant que les informations sont présentées dans une ou deux colonnes séparées par une virgule, un onglet ou un point-virgule.

**Remarque :** Lorsque vous utilisez la fonction **Ajouter plusieurs**, puis cliquez sur **Enregistrer**, toute valeur ou paire valeur/libellé précédemment existante pour cette recherche est

supprimée. Cependant, une fois que vous avez utilisé la fonction **Ajouter plusieurs**, vous pouvez manuellement ajouter des valeurs supplémentaires ou paires valeur/libellé.

1. Cliquez sur le bouton **Ajouter une recherche**.
2. Saisissez un nom pour la nouvelle recherche dans la zone de texte.
3. Pour ajouter manuellement une paire valeur/libellé :
  - Cliquez sur le bouton **Ajouter une valeur de recherche**.
  - Entrez une valeur et/ou un libellé pour la recherche.

Pour utiliser la fonction **Ajouter plusieurs** :

- Cliquez sur le bouton **Ajouter plusieurs** pour ouvrir la boîte de dialogue.
- Sélectionnez les colonnes et les séparateurs en conséquence. Si vous collez des données à partir de Microsoft Excel, utilisez le délimiteur onglet. Si vous définissez un délimiteur incorrect, l'outil importe toute la ligne comme valeur ou libellé (selon ce qui est désigné comme étant la première colonne).
- Saisissez les valeurs, séparateurs et/ou libellés pour tout le contenu ou collez le contenu d'une autre application.

Après avoir ajouté toutes les valeurs ou paires valeur/libellés, vous pouvez les trier dans l'ordre croissant ou décroissant sur la colonne Valeur ou Libellé. Notez qu'une fois que vous avez trié la liste, vous ne pouvez pas annuler le tri (sauf pour effectuer le tri dans l'ordre inverse ou effectuer un tri sur l'autre colonne).

4. Répétez l'étape 3, le cas échéant.
5. Cliquez sur **Enregistrer**.

### **Affectation de recherches**

Après la création d'une recherche, vous devez attribuer cette recherche dans le champ avec des données problématiques dans le stage Write Exceptions du flux de données.

1. Dans Enterprise Designer, ouvrez le flux de données qui produit les enregistrements d'exceptions.
2. Ouvrez le stage Write Exceptions.
3. Dans la colonne **Nom de la recherche** pour le champ contenant des données problématiques, sélectionnez la recherche qui contient les nouvelles données précises à partir de la liste déroulante, puis cliquez sur **OK**.
4. Enregistrez et renvoyez le flux de données.

### **Correction des enregistrements**

Après la création d'une recherche et l'affectation de cette recherche à un champ dans le flux de données, vous devez corriger les enregistrements d'exceptions dans le portail Business Steward.

1. Dans Exception Editor, sélectionnez le flux de données qui produit les enregistrements d'exceptions.
2. Pour le premier enregistrement problématique, cliquez sur le champ que vous avez affecté à la recherche.

3. Cliquez sur le bouton de menu déroulant dans ce champ et sélectionnez le libellé de l'enregistrement correct. N'oubliez pas que ce libellé n'est pas nécessairement identique à la valeur. Par exemple, si vous souhaitez que votre champ ait la valeur "California" vous pourriez cliquer sur un libellé disant "CA".
4. Répétez l'étape 3 pour chaque enregistrement problématique.
5. Enregistrer les exceptions modifiées

#### **Modification ou suppression des recherches**

1. Ouvrez la page Recherches.
2. Cochez la case en regard de la recherche appropriée.
3. Cliquez sur le bouton **Modifier recherche**, modifier la recherche autant que nécessaire, puis cliquez sur **Enregistrer**.

ou

Cliquez sur le bouton **Supprimer une recherche**.

#### **Domaines**

Domaines spécifie le type de données évaluées. Ceci est utilisé pour créer un rapport illustrant les types d'exceptions survenant dans vos données. Par exemple, si la condition évalue la réussite ou l'échec de validation d'adresse, le domaine de données pourrait être « Adresse » ; si la condition évalue la réussite ou l'échec d'une opération de géocodage, le domaine de données pourrait être « Spatial » etc.

**Remarque :** Les domaines que vous définissez ici serviront d'options par défaut à la fois pour la configuration de Business Steward et le stage Exception Monitor.

Vous pouvez sélectionner un des domaines prédéfinis ci-dessous ou choisir votre propre domaine en cliquant sur le bouton **Ajouter un élément** et en remplissant les champs, le cas échéant. Vous pouvez également modifier des domaines en sélectionnant un domaine, en cliquant sur le bouton **Modification d'un élément** et en y apportant les modifications nécessaires. Vous pouvez également filtrer la liste des domaines affichés en entrant des données de recherche dans le champ **Filtre**. Les résultats seront mis à jour dynamiquement.

- Account—La condition vérifie la raison sociale ou le nom d'entreprise associé à un compte commercial.
- Address—La condition vérifie les données d'adresse, telles que l'adresse postale complète ou un code postal.
- Asset—La condition vérifie les données sur les propriétés d'une entreprise, telles que les propriétés physiques, l'immobilier, les ressources humaines ou d'autres actifs.
- Date—La condition vérifie les données de date.
- Email—La condition vérifie les données de courrier électronique.
- Financial—La condition vérifie les données liées aux devises, titres, etc.
- Name—La condition vérifie les données de nom personnelles, telles que le prénom ou le nom de famille.
- Phone—La condition vérifie les données de numéro de téléphone.

- **Product**—La condition vérifie les données de matériaux, de pièces, de marchandises, etc.
- **Spatial**—La condition vérifie les données de point, de polygone ou de ligne représentant une entité géographique, telle que les plaines inondables, les zones littorales, les maisons, les régions de vente, etc.
- **SSN/SSN**—La condition vérifie les données de numéros américains de sécurité sociale.
- **Uncategorized**—Sélectionnez cette option si vous ne souhaitez pas catégoriser cette condition.

## Mesures

Mesures vous permet de spécifier la manière dont les données sont mesurées. Ceci est utilisé pour créer un rapport illustrant les types d'exceptions survenant dans vos données. Par exemple, si la condition est conçue pour évaluer l'intégralité d'un enregistrement (que toutes les adresses comprennent des codes postaux par exemple), vous pourrez alors utiliser « Intégralité » comme métrique de qualité de données.

**Remarque :** Les mesures que vous définissez ici serviront d'options par défaut à la fois pour la configuration de Business Steward et le stage Exception Monitor.

Vous pouvez sélectionner une des mesures prédéfinies ci-dessous ou définir votre propre mesure en cliquant sur le bouton **Ajouter un élément** et en remplissant les champs, le cas échéant. Vous pouvez également modifier des mesures en sélectionnant un domaine, en cliquant sur le bouton **Modification d'un élément** et en y apportant les modifications nécessaires. Vous pouvez également filtrer la liste des mesures affichées en entrant des données de recherche dans le champ **Filtre**. Les résultats seront mis à jour dynamiquement.

- **Accuracy**—La condition vérifie si les données peuvent être comparées à une source de confiance pour vérification. Par exemple, si une adresse n'a pas pu être vérifiée en utilisant les données provenant d'une administration postale, elle pourrait être considérée comme étant une exception car elle manque de précision.
- **Completeness**—La condition vérifie s'il manque des attributs clés aux données. Par exemple, une adresse n'ayant pas de code postal, ou un compte n'ayant pas de nom de contact.
- **Consistency**—La condition vérifie si les données du système sont cohérentes entre différents systèmes. Par exemple, si votre système de données de clients utilisent les codes de sexe H et F, mais que les données que vous traitez ont les codes de sexe 0 et 1, les données pourraient être considérées comme ayant des problèmes de consistance.
- **Interpretability**—La condition vérifie si les données sont correctement décomposées selon une structure de données pouvant être interprétée par un autre système. Par exemple, les numéros de sécurité sociale ne devraient contenir que des données numériques. Si les données comportent des lettres, tels que xxx-xx-xxxx, les données pourrait être considérées comme ayant des problèmes d'intelligibilité.
- **Recency**—La condition vérifie si les données sont à jour. Par exemple, si une personne déménage mais que l'adresse que vous possédez dans votre système contient l'ancienne adresse de cette personne, les données pourraient être considérées comme ayant un problème concernant leur caractère peu récent.
- **Uncategorized**—Sélectionnez cette option si vous ne souhaitez pas catégoriser cette condition.
- **Uniqueness**—La condition vérifie s'il existe des données doublons. Si le flux de données n'a pas pu consolider de données doublons, l'enregistrement serait considéré être une exception.

## Avis

La fonction Notifications permet au système d'envoyer un message à une ou plusieurs adresses électroniques lorsqu'un nombre désigné d'exceptions est lié à un domaine ou à une mesure spécifique. Cet e-mail inclut un lien vers les enregistrements en échec d'Exception Editor du Business Steward Portal, où vous pouvez saisir manuellement les données correctes. Pour arrêter de recevoir des notifications à une adresse électronique particulière, supprimez cette adresse de la liste des destinataires dans la ligne Envoyer une notification de la page Modifier un domaine.

**Remarque :** Les notifications doivent être configurées dans Management Console avant de pouvoir utiliser une notification avec succès, au sein de la configuration de Business Steward. Pour plus d'informations sur la configuration des notifications, reportez-vous au Guide d'administration.

1. Dans la configuration Business Steward, ouvrez la page **Domaines** page ou la page **Mesures**.
2. Sélectionnez le domaine ou la mesure pour laquelle vous souhaitez ajouter une notification et cliquez sur le bouton **Modification d'un élément**.
3. Sélectionnez le ou les noms d'utilisateur dans la liste déroulante (telle que configurée dans Management Console) ou saisissez les nouvelles adresses électroniques auxquelles les notifications doivent être envoyées.
4. Sélectionnez le nombre d'enregistrements d'exceptions devant déclencher une notification.
5. Saisissez le texte qui doit être envoyé sous l'objet de la notification.
6. Entrez le message qui doit apparaître dans le corps de la notification. Vous pouvez utiliser des variables dans le message pour transmettre des informations importantes sur l'(les)exception(s), notamment les informations suivantes :
  - `#{jobID}`: le numéro d'identifiant du job qui a produit le(s) enregistrement(s) d'exceptions.
  - `#{jobName}` : le nom du job qui a produit le(s) enregistrement(s) d'exceptions.
  - `#{userName}` : le nom de l'utilisateur dont le job a produit le(s) enregistrement(s) d'exceptions.
  - `#{stageLabel}` : le nom du stage du dataflow qui a produit le(s) enregistrement(s) d'exceptions.
  - `#{link}` : un lien vers la page éditeur dans le portail Business Steward, affichant les enregistrements pour un dataflow particulier.
7. Cochez la case **Envoyer un rappel** si vous souhaitez envoyer une notification de rappel, puis sélectionnez le nombre de jours qui doivent s'écouler avant que le rappel soit envoyé.
8. Sélectionnez le ou les noms d'utilisateur dans la liste déroulante (telle que configurée dans Management Console) ou saisissez une adresse électronique dans le champ **Envoyer un rappel à** pour envoyer des notifications à des adresses électroniques autres que celles spécifiées à l'étape 3.
9. Saisissez le texte qui doit être envoyé sous l'objet de la notification de rappel.
10. Entrez le message qui doit apparaître dans le corps de la notification de rappel. Le rappel utilise une variable supplémentaire :
  - `#{Count}` : le nombre d'exceptions pour le dataflow ou le stage spécifié qui doivent encore être résolues.



11. Cochez la case **Rappel quotidien** si vous souhaitez que la notification de rappel soit envoyée chaque jour jusqu'à ce que les exceptions soient résolues.

### Rapports de qualité des données

1. Cliquez sur **Rapport sur la qualité des données** pour suivre les conditions de réussite ou d'échec dans le stage Exception Monitor. Si vous désactivez cette option, la page Qualité des données du portail Business Steward ne contiendra aucune donnée. De même, le champ « Rapport uniquement » dans le stage Exception Monitor pour tous les dataflows est désactivé.
2. Dans la liste déroulante **Rétention**, sélectionnez la durée de conservation, en mois, des données.

### Configuration des indicateurs clés de performance

La section **Configuration ICP** de l'onglet Rapport sur la qualité des données vous permet d'indiquer des indicateurs clés de performance (ICP) pour vos données et d'assigner des notifications lorsque ces ICP remplissent certaines conditions.

1. Cliquez sur **Ajouter un ICP**.
2. Saisissez un **nom** pour l'indicateur clé de performance. Ce nom doit être unique sur votre serveur Spectrum™ Technology Platform.
3. Sélectionnez l'une des qualités de données **Mesures** pour l'indicateur clé de performance ; si vous n'effectuez pas de sélection, cet indicateur clé de performance est lié à toutes les **mesures**.
4. Sélectionnez un nom de **flux de données** pour l'indicateur clé de performance ; si vous n'effectuez pas de sélection, cet indicateur clé de performance est lié à tous les flux de données du module Business Steward.
5. Sélectionnez un **nom de stage** pour l'indicateur clé de performance ; si vous n'effectuez pas de sélection, cet indicateur clé de performance est lié à tous les stages du module Business Steward dans vos flux de données.
6. Sélectionnez un **domaine** de données pour l'indicateur clé de performance ; si vous n'effectuez pas de sélection, cet indicateur clé de performance est lié à tous les **domaines**. Si vous sélectionnez un domaine ici, le champ Condition sera désactivé.
7. Sélectionnez une **condition** pour l'indicateur clé de performance. Si vous n'effectuez pas de sélection, cet indicateur clé de performance présente la valeur par défaut « All ». Pour sélectionner une condition, vous devez au préalable avoir sélectionné « All » dans le champ Domain. Une fois qu'une condition a été sélectionnée, le champ Domain est désactivé.
8. Sélectionnez une **période de KPI** pour désigner les intervalles auxquels le module Business Steward doit surveiller vos données et envoyer des notifications. Par exemple, si vous sélectionnez « 1 » et « Monthly », une notification de KPI sera envoyée lorsque le pourcentage d'exceptions dépasse le seuil ou que la variance dépasse une période de mois à mois.
9. Fournissez un pourcentage de **Variance** ou de **Seuil**. Les valeurs de variance représentent le pourcentage augmenté d'échec dans les enregistrements d'exception depuis la dernière période. Les valeurs de seuil représentent le pourcentage d'échec à partir duquel des notifications doivent être envoyées. Cette valeur doit être 1 ou plus.



10. Sélectionnez des adresses électroniques de la liste ou saisissez les adresses électroniques des **Destinataires** qui doivent être notifiés lorsque ces conditions sont remplies. Si possible, ce champ sera renseigné automatiquement lorsque vous entrez les adresses électroniques. Vous ne devez pas séparer les adresses par des virgules, des point virgules ou toute autre ponctuation.
11. Entrez l'**objet** que l'e-mail de notification doit utiliser.
12. Entrez le **message** que doit contenir la notification lorsque ces conditions sont remplies.
13. Cliquez sur **OK**. Le nouveau KPI apparaîtra parmi tout autre KPI existant. Vous pouvez trier des KPI en fonction d'une des colonnes contenant des données.
14. Cliquez sur **Enregistrer**.

Vous pouvez modifier et supprimer des ICP en cliquant sur **Modifier l'ICP sélectionné** ou **Supprimer l'ICP sélectionné**.

### Services Outils de recherche

1. Sélectionnez les services d'outils de recherche que vous souhaitez rendre disponibles dans l'éditeur d'exceptions de Business Steward Portal. La liste des services disponibles dépend des droits d'utilisateur et est renseignée par les modules sous licence et les services dans Spectrum Technology Platform. Utilisez le **filtre** pour réduire la liste des services en fonction des critères de filtre.
2. Cliquez sur **Premium** pour indiquer aux utilisateurs qu'ils doivent escompter des frais supplémentaires lorsqu'ils utilisent les services (tels que les services Dun & Bradstreet).

### Options

1. Cliquez sur **Faire l'audit des événements d'exception** pour que le module Business Steward tienne à jour un journal lorsque des enregistrements d'exceptions sont créés, lus, mis à jour ou supprimés.
2. Cliquez sur **Suivre la progression** pour suivre l'approbation des enregistrements d'exceptions dans le portail Business Steward. Si vous désactivez cette option, les graphiques de progression du tableau de bord du portail Business Steward n'apparaîtront pas.

## Module Data Normalization

### Module Data Normalization

Le module Data Normalization examine les termes dans un enregistrement et détermine si le terme est au format préféré.

## Composants

Le module Data Normalization est constitué des éléments suivants :

- **Advanced Transformer** : ce stage scanne et divise les chaînes de données en plusieurs champs, en plaçant les données extraites et non extraites dans un champ existant ou un nouveau champ.
- **Open Parser** : ce stage décompose et analyse vos données d'entrée provenant de plusieurs cultures du monde en utilisant une grammaire simple mais puissante. En utilisant cette grammaire, vous pourrez définir une séquence d'expressions représentant des modèles de domaine utiles au parsing (décomposition analytique) de vos données d'entrée. Open Parser recueille également des statistiques et note les correspondances de parsing afin de vous aider à décider de l'efficacité de vos grammaires de parsing.
- **Table Lookup** : ce stage évalue un terme et le compare avec une forme de ce terme ayant été préalablement validée. Si le terme n'est pas dans la forme requise, alors la version standard le remplace. Table Lookup peut aussi remplacer un mot complet par son abréviation, une abréviation par un mot entier, changer un surnom en nom entier ou corriger une faute d'orthographe
- **Transliterator** : convertit une chaîne en écriture latine vers d'autres scripts.

## Advanced Transformer

Le job Advanced Transformer scanne et divise des chaînes de données en différents champs à l'aide de tables ou d'expressions régulières. Extraction d'un terme ou nombre de mots spécifique en partant de la gauche ou de la droite de ce terme. Les données extraites et non extraites peuvent être placées dans un champ existant ou un nouveau champ.

Par exemple, si vous voulez extraire les informations liées à la Suite (STE) de ce champ d'adresse et la placer dans un champ séparé.

2300 BIRCH RD STE 100

Pour accomplir ceci, vous pouvez créer un Advanced Transformer qui extraira le terme STE et tous les mots à la droite du terme STE, ce qui laissera le champ sous la forme :

2300 BIRCH RD

### Entrée

Advanced Transformer utilise tout champ d'entrée défini dans le flux de données.

### Options

Il est possible de configurer les options Advanced Transformer au niveau du stage, via l'un des clients Spectrum™ Technology Platform, ou lors de l'exécution, à l'aide des options de flux de données.

### Configuration des options

Vous devez créer une règle pour définir les options d'Advanced Transformer. Vous pouvez créer plusieurs règles puis spécifier l'ordre dans lequel ces règles seront appliquées. Pour créer une règle, procédez comme suit :

1. Faites un double-clic sur l'instance Advanced Transformer du canevas. La boîte de dialogue Options d'Advanced Transformer s'affiche.
2. Sélectionnez le nombre d'instances d'exécution, puis cliquez sur **OK**. Utilisez l'option Instances d'exécution pour paramétrer un dataflow afin de mener plusieurs exécutions parallèles d'un stage afin d'augmenter potentiellement les performances.
3. Cliquez sur le bouton **Ajouter**. Le dialogue Options d'Advanced Transformer s'affiche.

**Remarque** : Si vous ajoutez plusieurs règles de transformateur, vous pouvez utiliser les boutons **Monter** et **Descendre** pour modifier l'ordre dans lequel les règles s'appliquent.

4. Sélectionnez le type d'action de transformation que vous souhaitez effectuer, puis cliquez sur **OK**. Les options sont répertoriées dans le tableau ci-dessous.

**Tableau 23 : Options d'Advanced Transformer**

Option	Description
Source	Définit le champ de source d'entrée à évaluer pour être scanné et divisé.
Extraire en utilisant	<p>Choisir <b>Données de tableau</b> ou <b>Expressions régulières</b>.</p> <p>Sélectionnez <b>Données de table</b> si vous souhaitez effectuer une analyse et une scission à l'aide des tables XML situées à l'emplacement &lt;Drive&gt;:\Program Files\Pitney Bowes\Spectrum\server\modules\advancedtransformer\data. Voir les options de données de table ci-dessous pour plus d'informations sur chaque option.</p> <p>Choisir <b>Expressions régulières</b> si vous souhaitez scanner et diviser en utilisant des expressions régulières. Les expressions régulières offrent plusieurs options supplémentaires permettant de diviser les données. Vous pouvez utiliser les expressions régulières préemballées en choisissant une depuis la liste. Vous pouvez également en construire une en utilisant la syntaxe RegEx.</p> <p>Vous pouvez par exemple diviser les données lorsque la première valeur numérique est trouvée, comme dans « John Smith 123 Main St. » où le texte « John Smith » peut être assigné à un champ et « 123 Main St. » à un autre champ. Consulter les options d'Expressions régulières ci-dessous pour en savoir plus sur chaque option.</p>
<b>Options Données de table</b>	

Option	Description
Non-extracted Data	<p>Indique le champ de sortie qui devrait contenir les données transformées. Si vous souhaitez remplacer la valeur initiale, indiquer le même champ dans le dossier de Destination que dans la liste déroulante Source.</p> <p>Vous pouvez également saisir un nouveau nom de champ dans le champ Destination. Si vous saisissez un nouveau nom de champ, ce nom de champ sera disponible dans les stages de votre dataflow découlant d'Advanced Transformer.</p>
Extracted Data	<p>Indique le champ de sortie où vous voulez placer les données extraites.</p> <p>Vous pouvez également saisir un nouveau nom de champ dans le champ Données extraites. Si vous saisissez un nouveau nom de champ, ce nom de champ sera disponible dans les stages de votre dataflow découlant d'Advanced Transformer.</p>
Caractère de séparation	<p>Indique tout caractère spécial que vous souhaitez tokeniser. La tokenisation est un processus qui consiste à séparer des termes. Par exemple, s'il y a un champ avec les données « Smith, John », vous devriez insérer une virgule. Vous obtiendrez les termes :</p> <ul style="list-style-type: none"> <li>• Smith</li> <li>• ,</li> <li>• John</li> </ul> <p>Maintenant que les termes sont séparés, les données peuvent être divisées en scannant et en extrayant à la hauteur de la virgule afin que « Smith » et « John » soient identifiés de manière propre comme étant les données à standardiser.</p>
table	<p>Définit le tableau contenant les termes sur lesquels la division du champ se base. Pour une liste des tableaux, voir <a href="#">Tables Advanced Transformer</a> à la page 166. Pour obtenir des informations sur la création ou la modification des tableaux, voir <a href="#">Présentation des tables de recherche</a> à la page 166.</p>
Recherche de termes contenant plusieurs mots	<p>Cochez cette case pour activer les recherches de plusieurs mots dans une chaîne donnée. Par exemple :</p> <p>Chaîne d'entrée = « Cedar Rapids 52401 » Règle d'entreprise = Identifier « Cedar Rapids » dans les chaînes basées sur un tableau contenant l'entrée ; Cedar Rapids = Sortie Etats-Unis = Identifie la présence de « Cedar Rapids » et place les termes dans un nouveau champ, par exemple, <b>Ville</b>.</p> <p>Pour des recherches avec plusieurs mots, la recherche est interrompue après qu'une première mise en correspondance soit établie.</p> <p><b>Remarque</b> : En choisissant cette option, il est possible que les performances en soient affectées de manière négative.</p>

Option	Description
Extraire	<p>Indique le type d'extraction à effectuer. Sélectionnez l'une des options suivantes :</p> <p><b>Extraire le terme</b>                      Extrait le terme identifié par le tableau sélectionné</p> <p><b>Extraire N mots à droite du terme</b>      Extrait des mots à droite du terme. Vous indiquez le nombre de mots à extraire. Par exemple, si vous souhaitez extraire les deux mots à droite du terme identifié, choisissez 2.</p> <p><b>Extraire N mots à gauche du terme</b>      Extrait des mots à gauche du terme. Vous indiquez le nombre de mots à extraire. Par exemple, si vous souhaitez extraire les deux mots à gauche du terme identifié, choisissez 2.</p> <p>Si vous choisissez d'extraire des mots à droite ou à gauche du terme, vous pouvez indiquer si vous souhaitez également inclure le terme en question dans les données de destination ou les données extraites. Par exemple, dans le champ :</p> <p>2300 BIRCH RD STE 100</p> <p>vous désirez extraire « STE 100 » et le placer dans le champ indiqué dans les données extraites, vous choisiriez d'inclure le terme dans le champ de données extrait, comprenant ainsi l'abréviation « STE » et le mot « 100 ».</p> <p>Si vous ne choisissez ni Données de destination ni Données extraites, le terme ne sera pas inclus et sera éliminé.</p>

### Options Expressions régulières

Expressions régulières	<p>Choisissez une expression régulière préemballée depuis la liste, ou assemblez-en une depuis la zone de texte. L'Advanced Transformer prend en charge la syntaxe RegEx standard.</p> <p>La plateforme Java 2 contient un package nommé <b>java.util.regex</b> qui permet l'utilisation d'expressions régulières. Pour plus d'informations, rendez-vous sur : <a href="http://java.sun.com/docs/books/tutorial/essential/regex/index.html">java.sun.com/docs/books/tutorial/essential/regex/index.html</a>.</p>
Bouton Ellipse	<p>Cliquez sur ce bouton pour ajouter ou supprimer une nouvelle expression régulière.</p>
Remplissage des groupes	<p>Après avoir choisi une nouvelle expression RegEx prédéfinie ou après en avoir saisie une, cliquez sur <b>Peupler le groupe</b> afin d'extraire tous groupes RegEx et de placer l'expression complète ainsi que tous groupes RegEx trouvé dans la liste de Groupes.</p>

Option	Description
Groupes	<p>Cette colonne montre les expressions régulières du groupe d'Expressions régulières sélectionné.</p> <p>Par exemple, si vous sélectionnez l'expression Date Regex, l'expression suivante s'affiche sur la zone de texte :  <code>(1[012]{1,2} 0?[1-9])[-/.]([12][0-9] 3[01]{1,2} 0?[1-9])[-/.](((0-9){4}))</code>. Cette expression Regex est divisée en trois et l'ensemble de l'expression et chacune des trois parties peuvent être envoyées à des champs de sortie différents. L'ensemble de l'expression est recherchée dans le champ texte, et si une mise en correspondance à lieu dans le champ de source, les parties associées seront déplacées jusqu'au champ de sortie assigné. Si le champ de source est « Le 12/14/2006 » et que vous lui appliquez l'expression Date et assignez toute la date (par exemple, «12/14/2006 ») devant être placée dans le champ DATE, le « 12 » devra être placé dans le champ MOIS, le « 14 » dans le champ JOUR et « 2006 » dans le champ Année La recherche de date sera donc effectuée et les informations adéquates trouvées (le cas échéant) seront déplacée jusqu'au champ de sortie adéquat.</p> <p>Champ source : "On 12/14/2006" DATE: "12/14/2006" MONTH: "12" DAY: "14" YEAR: "2006"</p>
Champ de sortie	Faites dérouler la liste du menu pour choisir un champ de sortie.

### Configuration des options lors de l'exécution

Les règles Advanced Transformer peuvent être configurées et transmises lors de l'exécution si elles sont exposées sous forme d'options de flux de données. Cela vous permet de remplacer la configuration existante par des chaînes au format JSON. Vous pouvez également définir des options de stage lors de l'appel d'un job via un flux de processus ou l'outil de ligne de commande Exécuteur de job.

Vous pouvez rechercher des schémas AdvancedTransformerRules dans le dossier suivant :

```
<Spectrum Location>\server\modules\jsonSchemas\advancedTransformer
```

Pour définir des règles Advanced Transformer lors de l'exécution :

1. Dans Enterprise Designer, ouvrez un flux de données qui utilise le stage Advanced Transformer.
2. Enregistrez et exposez ce flux de données.
3. Accédez à **Modifier > Options de dataflow**.
4. Dans la table **Map dataflow options to stages**, développez Advanced Transformer. Cochez la case AdvancedTransformerRules.
5. Facultatif : modifiez le nom des options du champ **Option label**.
6. Cliquez deux fois sur **OK**.

## Sortie

L'Advanced Transformer ne crée pas de nouveaux champs de sortie. Seuls les champs que vous définissez sont écrits à la sortie.

## Open Parser

L'Open Parser décompose et analyse vos données d'entrée provenant de plusieurs cultures du monde différentes en utilisant une grammaire simple mais puissante. En utilisant cette grammaire, vous pourrez définir une séquence d'expressions représentant des modèles de domaine utiles au parsing (décomposition analytique) de vos données d'entrée. Open Parser recueille également des statistiques et note les correspondances de parsing afin de vous aider à décider de l'efficacité de vos grammaires de parsing.

Utilisez Open Name Parser pour :

- Décomposer et analyser des données d'entrée en utilisant des grammaires de culture spécifiques à un domaine/à une culture que vous pourrez définir dans Domain Editor.
- Décomposer et analyser des données en utilisant des grammaires de parsing indépendant du domaine que vous pourrez définir dans l'Open Parser en utilisant la même grammaire de parsing simple et puissante à votre disposition dans Domain Editor.
- Analyser les données d'entrée via des grammaires d'analyse indépendantes du domaine lors de l'exécution, que vous définissez dans les options de dataflow.
- Avoir un aperçu des grammaires de parsing permettant de tester le parsing d'un échantillon de données d'entrée avant d'exécuter le job en utilisant le fichier de données d'entrée cible.
- Surveiller les résultats de la grammaire de parsing afin de savoir si les jetons ont correspondu ou non aux expressions que vous aviez défini et afin de mieux comprendre le processus de mise en correspondance.

## Entrée

Open Parser accepte les champs d'entrée définis dans votre grammaire d'analyse. Pour plus d'informations, reportez-vous à la section [Commandes de section d'en-tête](#) à la page 30.

Si vous effectuez une analyse propre à la culture, vous pouvez éventuellement inclure un champ CultureCode dans les données d'entrée pour utiliser une grammaire d'analyse d'une culture spécifique pour un enregistrement. Si vous omettez le champ CultureCode ou s'il est vide, chaque culture répertoriée dans le stage Open Parser est appliquée, dans l'ordre indiqué. Le résultat de la culture présentant le score d'analyse le plus élevé ou la première culture ayant un score de 100 est renvoyé. Pour plus d'informations sur le champ CultureCode, voir [Affectation d'une culture d'analyse à un enregistrement](#) à la page 12.

## Options

Les tableaux suivants répertorient les options du stage Open Parser.

## Onglet Règles

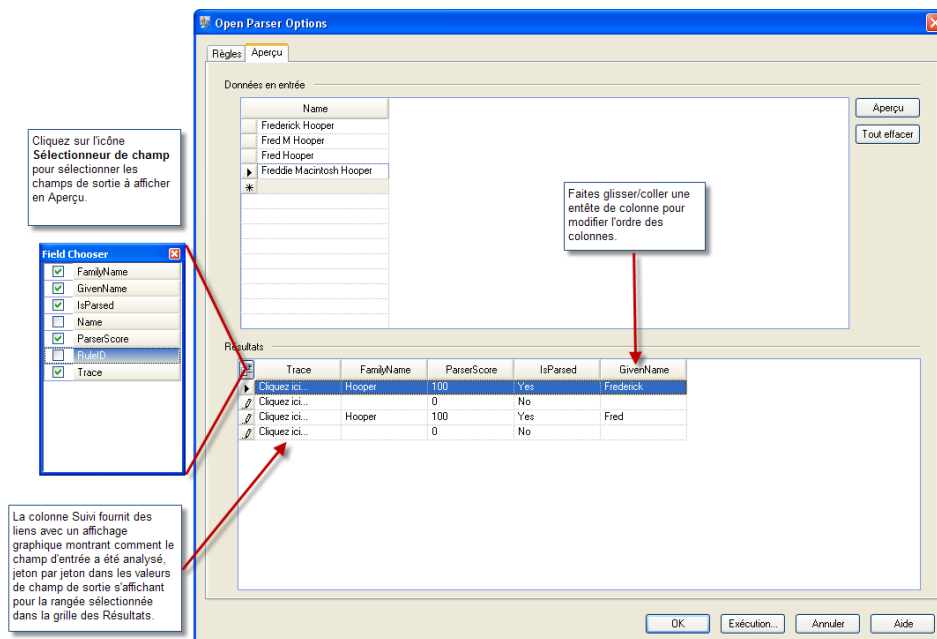
Option	Description
Utiliser une grammaire de domaine de culture spécifique	<p>Indique d'utiliser une grammaire d'analyse propre à une langue et à un domaine qui a déjà été définie dans l'outil Open Parser Domain Editor dans Enterprise Designer. Pour plus d'informations sur la définition de domaines, voir <a href="#">Définition d'une grammaire de parsing spécifique à une culture</a> à la page 10.</p> <p>Si vous choisissez cette option, vous verrez également les options suivantes :</p> <p><b>Domaine</b> Indique la grammaire d'analyse à utiliser.</p> <p><b>Cultures</b> Indique la langue ou la culture des données à analyser. Cliquez sur le bouton <b>Ajouter</b> pour ajouter une culture. Vous pouvez modifier l'ordre des tentatives d'Open Parser d'analyser les données avec chaque culture à l'aide des boutons <b>Déplacer vers le haut</b> et <b>Déplacer vers le bas</b>. Pour plus d'informations sur les cultures, voir <a href="#">Définition d'une grammaire de parsing spécifique à une culture</a> à la page 10.</p> <p><b>Renvoyer plusieurs enregistrements parsés</b> Activez cette option pour qu'Open Parser affiche les enregistrements pour chaque culture parvenant à l'entrée. Si vous ne cochez pas cette case, l'Open Parser affichera les résultats pour le premier enregistrement atteignant 100 comme score d'analyseur, qu'elle que soit la culture. Si toutes les cultures s'exécutent sans trouver d'enregistrement avec un score d'analyseur de 100, l'Open Parser affichera les enregistrements dont le score est le plus proche de 100. Si plusieurs cultures trouvent des enregistrements avec le même meilleur score en dessous de 100, l'ordre définit dans l'étape 4 déterminera quel enregistrement de culture sera affiché.</p>
Définir une grammaire indépendante du domaine	<p>Choisissez cette option si vous voulez définir une grammaire d'analyse qui doit être appliquée sans considération de la langue ou du domaine des données d'entrée. Si vous choisissez cette option, l'éditeur de grammaire apparaît et vous pouvez définir la grammaire d'analyse directement dans le stage Open Parser plutôt que d'utiliser l'outil Open Parser Domain Editor dans Enterprise Designer.</p> <p><b>Remarque :</b> Vous pouvez également définir une grammaire indépendante du domaine lors de l'exécution. Pour de plus amples informations, cliquez <a href="#">ici</a>.</p>

## Onglet Aperçu

La création d'une grammaire de parsing fonctionnant correctement est un processus itératif. L'aperçu est utile lors de test de variations dans votre entrée afin de vous assurer que la grammaire de parsing produise bien les résultats escomptés.



Saisissez les valeurs de test dans le champ d'entrée puis cliquez sur **Aperçu**.



Les champs analysés s'afficheront dans la grille des **Résultats**. Pour obtenir des informations sur les champs de sortie, voir [Sortie](#) à la page 305. Pour obtenir des informations sur les traces, voir [Suivi des résultats finaux de parsing](#) à la page 37. Si vous n'obtenez pas les résultats escomptés, cliquez sur l'onglet **Règles** et continuez d'éditer la grammaire de parsing et de tester les données d'entrée jusqu'à obtenir les résultats escomptés.

## Sortie

**Tableau 24 : Sortie Open Parser**

Nom du champ	Description/Valeurs valides
<Champ de saisie>	Le champ de saisie définit dans la grammaire de parsing.
<Champs de sortie...>	Les champs de sortie définis dans la grammaire de parsing.
CultureCode	Les codes de culture contenus dans les données d'entrée. Pour obtenir la liste complète des codes de culture pris en charge, voir <a href="#">Affectation d'une culture d'analyse à un enregistrement</a> à la page 12.

Nom du champ	Description/Valeurs valides
CultureUsedtoParseSelect a match results in the Match Results List and then click <b>Remove</b> .	La valeur de code de culture utilisé pour analyser chaque enregistrement de sortie. Cette valeur se base sur les correspondances à une grammaire de parsing spécifique à une culture.
IsParsed	Indique si un enregistrement de sortie a été analysé ou non. Les valeurs possibles sont Oui ou Non.
ParserScoreSelect a match results in the Match Results List and then click <b>Remove</b> .	Indique le score total moyen. La valeur du ParserScore sera comprise entre 0 et 100, tel que définit dans la grammaire de parsing. 0 signifie qu'aucun résultat n'a été trouvé.  Pour plus d'informations, reportez-vous à la section <a href="#">Scoring</a> .
Trace	Cliquez sur cette commande pour voir un affichage graphique permettant de savoir comment chaque jeton dans la grammaire de parsing a été analysé puis déplacé jusqu'à un champ de sortie pour la ligne correspondante dans la grille des Résultats.

## Recherche dans la table

Le stage Table Lookup standardise les termes en s'appuyant sur une forme de ce terme ayant été préalablement validée et en appliquant la version standard. Cette évaluation s'effectue en cherchant dans une table afin d'y trouver le terme à normaliser.

Par exemple :

	Prénom	Nom de famille
Entrée source :	Fréd	Smith
Sortie normalisée :	Frédéric	Smith

Trois types d'actions peuvent être effectués : standardiser, identifier et catégoriser.

Si le terme est trouvé lors de l'action de normalisation, Table Lookup remplace soit le champ entier, soit les termes individuels de ce champ par le terme normalisé, même si le champ contient plusieurs

mots. Table Lookup peut aussi remplacer un mot complet par son abréviation, une abréviation par un mot entier, changer un surnom en nom entier ou corriger une faute d'orthographe.

Si le terme est trouvé lors de l'action d'identification, Table Lookup marque l'enregistrement comme contenant un terme qui peut être normalisé, mais n'effectue aucune action.

Si le terme est trouvé lors de l'action de catégorisation, Table Lookup utilise la valeur source comme clé et copie la valeur correspondante de l'entrée de table dans le champ sélectionné. Si aucun des termes source ne correspond, **Catégoriser** utilise la valeur spécifiée par défaut.

## Entrée

**Tableau 25 : Champs d'entrée de Table Lookup**

Nom du champ	Description/Valeurs valides
Source	Définit le champ de source d'entrée à évaluer pour être scanné et divisé.
StandardizationTable	L'une des tables énumérées dans <a href="#">Tables Table Lookup</a> à la page 169.

## Options

Il est possible de configurer les options Table Lookup au niveau du stage, via l'un des clients Spectrum™ Technology Platform, ou lors de l'exécution, à l'aide des options de flux de données.

### Configuration des options

Pour spécifier des options pour Table Lookup, vous créez une règle. Vous pouvez créer plusieurs règles puis spécifier l'ordre dans lequel ces règles seront appliquées. Pour créer une règle, ouvrez le stage Table Lookup et cliquez sur **Ajouter**, puis renseignez les champs suivants.

**Remarque :** Si vous ajoutez plusieurs règles Table Lookup, vous pouvez vous servir des boutons **Monter** et **Descendre** pour modifier l'ordre dans lequel les règles sont appliquées.

Option	Description
Action	<p data-bbox="553 373 1429 401">Indique le type d'action à effectuer sur le champ source. L'un des éléments suivants :</p> <ul data-bbox="553 422 1429 894" style="list-style-type: none"><li data-bbox="553 422 1429 579"><b>Normaliser</b> Modifie les données dans un champ pour mettre en correspondance le terme standardisé trouvé dans la table de recherche. Si le champ contient plusieurs termes, seuls les termes trouvés dans la table de recherche sont remplacés par le terme standardisé. Les autres données dans le champ ne sont pas modifiées.</li><li data-bbox="553 600 1429 758"><b>Identifier</b> Marque l'enregistrement comme contenant un terme qui peut être standardisé, mais n'effectue aucune action sur les données du champ. Le champ de sortie <b>StandardizedTermIdentified</b> est ajouté à l'enregistrement avec une valeur Oui si le champ peut être normalisé, et Non si ce n'est pas possible.</li><li data-bbox="553 779 1429 894"><b>Catégoriser</b> Utilise la valeur source comme clé et copie la valeur correspondante de la table dans le champ sélectionné dans la liste Destination. Cela crée un champ dans vos données qui permet de catégoriser les enregistrements.</li></ul>

Option	Description
Le	<p>Détermine si le système utilise l'ensemble du champ comme terme de recherche ou s'il recherche chaque termes du champ dans la table de recherche. L'un des éléments suivants :</p> <p><b>Compléter le champ</b> Traite l'ensemble du champ comme un terme, ce qui produit ce qui suit :</p> <ul style="list-style-type: none"> <li>• Si vous avez sélectionné l'action <b>Standardiser</b>, Table Lookup traite l'ensemble du champ comme une chaîne et tente de standardiser le champ en utilisant la chaîne dans son ensemble. Par exemple, « International Business Machines » deviendrait « IBM ».</li> <li>• Si vous avez sélectionné l'action <b>Identifier</b>, Table Lookup traite le champ dans son intégralité comme une chaîne et marque l'enregistrement si la chaîne comme un tout peut être standardisée.</li> <li>• Si vous avez sélectionné l'action <b>Catégoriser</b>, Table Lookup traite le champ dans son intégralité comme une chaîne et marque l'enregistrement si la chaîne comme un tout peut être catégorisée.</li> </ul> <p><b>Termes individuels dans le champ</b> Traite chaque mot du champ comme un terme propre, ce qui produit ce qui suit :</p> <ul style="list-style-type: none"> <li>• Si vous avez sélectionné l'action <b>Standardiser</b>, Table Lookup analyse le champ et essaie de standardiser les termes individuels dans ce champ. Par exemple, « Bill Mike Smith » deviendrait « William Michael Smith ».</li> <li>• Si vous avez sélectionné l'action <b>Identifier</b>, Table Lookup analyse le champ et marque l'enregistrement si n'importe quel terme simple dans ce champ peut être standardisé.</li> <li>• Si vous sélectionnez l'action <b>Catégoriser</b>, Au contraire de <b>Normaliser</b>, <b>Catégoriser</b> ne copie pas le terme source s'il n'y a pas de correspondance avec une table. Si aucun des termes source ne correspond, <b>Catégoriser</b> utilise la valeur spécifiée par défaut. Contrairement à <b>Normaliser</b>, <b>Catégoriser</b> ne renvoie que cette valeur de table et aucune valeur provenant de <b>Source</b>. Si aucun des termes source ne correspond, <b>Catégoriser</b> utilise la valeur spécifiée par défaut.</li> </ul>
Source	Indique le champ qui doit contenir le terme à rechercher.
Cible	<p>Indique le champ dans lequel les termes renvoyés par la table de recherche doivent être écrits.</p> <p>Si vous désirez remplacer la valeur, spécifiez le même champ dans le champ <b>Destination</b> comme vous avez fait dans le champ <b>Source</b>. Vous pouvez aussi créer un nouveau champ en saisissant le nom du champ que vous désirez créer.</p> <p>Le champ <b>Destination</b> n'est pas disponible si vous sélectionnez l'action <b>Identifier</b>.</p>

Option	Description				
table	<p>Indique la table à utiliser pour rechercher les termes correspondant aux données dans votre flux de données.</p> <p>Pour une liste de tables que vous pouvez éditer, voir <a href="#">Tables Table Lookup</a> à la page 169. Pour obtenir des informations sur la création ou la modification des tableaux, voir <a href="#">Présentation des tables de recherche</a> à la page 166.</p>				
Recherche de termes contenant plusieurs mots	<p>Active les recherches de plusieurs mots dans une chaîne donnée. Par exemple :</p> <p>Chaîne d'entrée : « Major General John Smith »  Règle d'entreprise : identifier « Major General » dans une chaîne basée sur un tableau contenant l'entrée  Sortie : remplacer « Major General » par « Maj. Gen ».</p> <p>Pour des recherches avec plusieurs mots, la recherche est interrompue après qu'une première mise en correspondance soit établie.</p> <p>Cette option est désactivée lorsque <b>On</b> est défini sur <b>Champ entier</b>.</p> <p><b>Remarque</b> : En choisissant cette option, il est possible que les performances en soient affectées de manière négative.</p>				
Quand l'entrée de table est introuvable, définissez la valeur de destination sur	<p>Indique la valeur à placer dans le champ de destination si un terme correspondant est introuvable dans la table de recherche. L'un des éléments suivants :</p> <table border="0"> <tr> <td style="padding-right: 20px;"><b>Valeur de la source</b></td> <td>Placez la valeur du champ source dans le champ de destination.</td> </tr> <tr> <td><b>Autre</b></td> <td>Placez une valeur spécifique dans le champ de destination.</td> </tr> </table>	<b>Valeur de la source</b>	Placez la valeur du champ source dans le champ de destination.	<b>Autre</b>	Placez une valeur spécifique dans le champ de destination.
<b>Valeur de la source</b>	Placez la valeur du champ source dans le champ de destination.				
<b>Autre</b>	Placez une valeur spécifique dans le champ de destination.				

### Configuration des options lors de l'exécution

Les options Table Lookup peuvent être configurées et transmises lors de l'exécution si elles sont exposées sous forme d'options de flux de données. Cela vous permet de remplacer la configuration existante par des chaînes au format JSON. Vous pouvez également définir des options de stage lors de l'appel d'un job via un flux de processus ou l'outil de ligne de commande Exécuteur de job.

Vous pouvez rechercher un schéma LookupRule dans le dossier suivant :

```
<Spectrum Location>\server\modules\jsonSchemas\tableLookup
```

Pour définir des règles Table Lookup lors de l'exécution :

1. Dans Enterprise Designer, ouvrez un flux de données utilisant le stage Table Lookup.
2. Enregistrez et exposez ce flux de données.
3. Accédez à `Edit > Dataflow Options`.

4. Dans la table **Map dataflow options to stages**, développez Table Lookup. Cochez la case LookupRule.
5. Facultatif : modifiez le nom des options du champ **Option label**.
6. Cliquez deux fois sur **OK**.

## Sortie

**Tableau 26 : Sorties Table Lookup**

Nom du champ	Description/Valeurs valides
StandardizedTermIdentified	Indique si le champ contient ou non un terme pouvant être normalisé. Sortie uniquement si vous sélectionnez l'option <b>Champ entier</b> ou <b>Termes individuels dans un champ</b> . <b>Yes</b> L'enregistrement contient un terme pouvant être normalisé. <b>No</b> L'enregistrement ne contient pas de terme pouvant être normalisé.

## Transliterateur

Transliterateur convertit une chaîne en écriture latine vers d'autres scripts. Par exemple :

Source	Transliteration
キャンパス	kyanpasu
Αλφαβητικός Κατάλογος	
биологическом	biologichyeskom

Il est important de noter que translittérer n'est pas traduire. La translittération est plutôt la conversion de lettres d'un script à un autre sans traduire les mots sous-jacents.

**Remarque :** Souvent, les méthodes de translittération standard ne suivent pas les règles de prononciation d'une langue particulière dans le texte cible.

Le stage Transliterator prend en charge les scripts suivants. En général, le stage Transliterator suit les lignes directrices de translittération en écriture latine du groupe de travail GENUNG (Groupe d'experts des Nations unies pour les noms géographiques). Pour plus d'informations, voir [www.eki.ee/wgrs](http://www.eki.ee/wgrs).

<b>Arabe</b>	Script utilisé par plusieurs langues asiatiques et africaines, y compris l'arabe, le persan et l'ourdou.
<b>Cyrillique</b>	Script utilisé par l'Europe orientale et les langues asiatiques, y compris les langues slaves comme le russe. Le stage Transliterator suit généralement la norme ISO 9 pour l'ensemble cyrillique de base.
<b>Devanagari</b>	Le script utilisé par plusieurs langues indiennes, notamment l'hindi et le sanskrit. Ce script est un descendant de l'écriture brahmi qui est l'un des plus anciens systèmes d'écriture utilisés dans l'Inde ancienne et dans l'Asie du sud et centrale actuelle.
<b>Grec</b>	Script utilisé pour la langue grecque. Ce script appartient à la branche hellénique de la famille des langues indo-européennes.
<b>Gujarâti</b>	Le script utilisé par l'état du Gujarat dans l'ouest de l'Inde. C'est l'un des scripts modernes de l'Inde qui a été adapté de l'écriture devanagari.
<b>Gurmukhi</b>	Le script utilisé par la langue indienne punjabi. Ce script a reçu une influence considérable du script nagari, qui est une forme plus ancienne du script devanagari.
<b>Hangul</b>	Script utilisé pour le coréen. Le stage Transliterator suit les règles de translittération du ministère coréen du Tourisme et de la Culture. Pour plus d'informations, consultez le site Web de <a href="#">L'Institut National de la langue coréenne</a> .
<b>Han</b>	Script utilisé par la langue chinoise. Il s'agit d'une branche de la famille des langues tibéto-birmanes et a été écrite avec des scripts basés sur le thaï et le chinois.
<b>Chinois traditionnel/simplifié</b>	Le stage Transliterator prend en charge à la fois le chinois traditionnel et le chinois simplifié. Voici un exemple de chinois traditionnel : 人人生而自由. Voici un exemple de chinois simplifié : 人人生而自由.
<b>Kannada</b>	Le script utilisé par plusieurs langues indiennes du sud, telles que le konkani. Ce script est un descendant de l'écriture brahmi de l'Inde ancienne.
<b>Katakana et hiragana</b>	L'un des scripts qui peuvent être utilisés pour écrire le japonais. Le stage Transliterator utilise une légère variante du système Hepburn. Avec le système Hepburn, ZI (ジ) et DI (ヂ) sont représentés par « ji » et ZU (ズ) et DU (ヅ) sont représentés par « zu ». La réversibilité effectue une légère modification en utilisant « dji » pour DI et « Dzou » pour DU. La translittération Katakana est réversible. La translittération Hiragana-Katakana n'est pas complètement réversible puisqu'il y a plusieurs lettres de Katakana qui n'ont pas d'équivalents correspondants en Hiragana. Aussi, la marque de longueur



n'est pas utilisée avec Hiragana. La translittération Hiragana-écriture latine n'est pas non plus réversible puisqu'en interne, il s'agit d'une combinaison de Hiragana-Katakana et d'Hiragana-écriture latine.

<b>Demi-chasse/pleine chasse.</b>	Le stage Transliterator peut convertir des scripts demi-chasse étroits et des scripts pleine chasse plus larges. Voici un exemple en demi-chasse : アノリウ. Voici un exemple en pleine chasse : アルアノリウ.
<b>Latin</b>	Script utilisé par la plupart des langues européennes, tel que l'anglais. Il était à l'origine utilisé par les anciens Romains pour écrire en langue latine.
<b>Malayalam</b>	Le script utilisé par la langue malayalam, la langue officielle de l'état indien du Kerala. Ce script a d'abord été écrit avec l'alphabet vatteluttu, qui signifie « écriture ronde », et développé à partir de l'écriture brahmi de l'Inde ancienne.
<b>Oriya</b>	Le script utilisé par la langue oriya, la langue officielle de l'état indien d'Odisha. Le script oriya a été développé à partir de l'écriture kalinga, l'un des nombreux descendants de l'écriture brahmi de l'Inde ancienne.
<b>Tamoul</b>	Le script utilisé par la langue tamoule dans plusieurs états de l'Inde, du Sri Lanka et de la Malaisie. Ce script a été écrit à l'origine avec une version du script brahmi connue sous le nom de tamil brahmi.
<b>Telugu</b>	Le script utilisé par plusieurs langues du sud de l'Inde. Ce script est un descendant de l'écriture brahmi de l'Inde ancienne.
<b>Thaï</b>	Le script utilisé par la langue thaï. Ce script est influencé par l'écriture brahmi de l'Inde ancienne et les alphabets khmers.

Transliterator fait partie du Module Data Normalization. Pour obtenir une liste des autres stages, reportez-vous à la section [Module Data Normalization](#) à la page 297.

## Concepts de translittération

Les translittérations de script nécessitent un certain nombre de qualités. Une bonne translittération doit être :

- Terminés
- Prévisible
- Prononçable
- Sans équivoque

Ces qualités sont rarement satisfaites simultanément, de sorte que le stage Transliterator tente d'équilibrer ces exigences.

### Terminés

Chaque séquence bien formée de caractères dans le script source doit translittérer en une séquence de caractères à partir du script cible.

### Prévisible

Les lettres elles-mêmes (sans aucune connaissance des langues écrites dans ce script) doivent être suffisantes pour la translittération, basée sur un nombre relativement restreint de règles. Cela permet à la translittération d'être effectuée mécaniquement.

### Prononçable

La translittération offre moins d'utilité si le processus fait correspondre simplement les caractères sans tenir compte de leur prononciation. Mapper simplement « αβγδεζηθ... » en « abcdefgh... » donne des chaînes qui peuvent être complètes et sans ambiguïté, mais impossible à prononcer.

Souvent, les méthodes de translittération standard ne suivent pas les règles de prononciation d'une langue particulière dans le texte cible. Par exemple, le système Hepburn japonais utilise un « j » qui a la valeur phonétique anglaise (par opposition aux français, allemand ou espagnol), mais utilise des voyelles qui n'ont pas les sons de l'anglais standard. Une méthode de translittération peut également exiger des connaissances particulières pour avoir la bonne prononciation. Par exemple, dans le système japonais kunrei-siki, « tu » se prononce « tsu ». Cela ressemble à des situations où il y a différentes langues dans le même script. Par exemple, savoir que le mot Gewalt vient de l'allemand permet à un lecteur compétent de prononcer le « w » comme un « v ».

Dans certains cas, la translittération peut être fortement influencée par la tradition. Par exemple, la lettre bêta (β) en grec moderne sonne comme un « v », mais une transformation peut continuer à utiliser un b (comme dans biologie). Dans ce cas, l'utilisateur doit savoir qu'un « b » dans le mot translittéré correspond à bêta (β) et doit être prononcé comme un « v » en grec moderne. Les lettres peuvent aussi être translittérées différemment selon leur contexte pour rendre la prononciation plus prévisible. Par exemple, depuis la séquence grecque GAMMA GAMMA (γγ) est prononcé comme « ng », le premier GAMMA peut être transcrit comme un « n ».

**Remarque :** En général, afin de produire des résultats prévisibles lors de la translittération de scripts latins en scripts différents, le texte anglais ne produit pas de résultats phonétiques. En effet, la prononciation de l'anglais ne peut pas être prédite facilement à partir des lettres d'un mot. Par exemple, grove, move et love se terminent avec « ove », mais sont prononcés différemment.

### Sans équivoque

Il doit toujours être possible de récupérer le texte du script source à partir de la translittération du script cible. Par exemple, le mot Elláda doit pouvoir revenir vers l'original Ελλάδα. Cependant, en translittération, des caractères multiples peuvent produire des ambiguïtés. Par exemple, le caractère grec PSI (ψ) mappe vers ps, mais ps peut aussi découler de la séquence PI, SIGMA (πσ) comme PI (π) mappe vers p et SIGMA (σ) mappe vers s.

Pour gérer le problème de l'ambiguïté, Transliterator utilise une apostrophe pour désambiguïser les séquences de caractères. Avec cette procédure, le caractère grec PI SIGMA (πσ) mappe vers p's. En japonais, chaque fois qu'une séquence ambiguë dans le script cible ne découle pas d'une seule lettre, la transformation utilise une apostrophe pour lever l'ambiguïté sur elle. Par exemple, il utilise cette procédure pour faire la distinction entre man'ichi et manichi.

**Remarque :** Certains caractères dans un script cible ne se trouvent pas normalement à l'extérieur de certains contextes. Par exemple, le petit caractère japonais « ya », comme dans « kya » (キヤ), n'est pas isolé généralement. Pour gérer ces caractères, Transliterator utilise un tilde. Par exemple, l'entrée « ~ya » peut produire un petit « ya » isolé. Lors de la translittération du grec, l'entrée « a~s » peut produire un sigma grec non-final (ασ) à la fin d'un mot. De même, l'entrée « ~sa » peut produire un sigma final dans une position non-finale (σα).


Pour les transformations de scripts généraux, une technique courante pour la réversibilité est d'utiliser des accents supplémentaires pour distinguer les lettres qui ne peuvent être distinguées autrement. L'exemple suivant représente le texte grec qui est mappé en latin entièrement réversible :

## Entrée

Nom du champ	Description										
Tout champ de chaîne	Le stage Transliterator peut translittérer un champ de type chaîne. Vous pouvez spécifier les champs à translittérer dans les options du stage Transliterator.										
TransliteratorID	<p>Substitue la translittération par défaut spécifiée dans les options du stage Transliterator. Utilisez ce champ si vous souhaitez spécifier une translittération différente pour chaque enregistrement.</p> <p>Par exemple :</p> <table border="0"> <tr> <td><b>Arabic-Latin</b></td> <td>De l'arabe au latin.</td> </tr> <tr> <td><b>Greek-Latin</b></td> <td>Du grec au latin.</td> </tr> <tr> <td><b>Latin-Hangul</b></td> <td>Du latin à l'hangûl.</td> </tr> <tr> <td><b>Latin-Katakana</b></td> <td>Du latin au katakana.</td> </tr> <tr> <td><b>Fullwidth-Halfwidth</b></td> <td>De pleine chasse à demi-chasse.</td> </tr> </table>	<b>Arabic-Latin</b>	De l'arabe au latin.	<b>Greek-Latin</b>	Du grec au latin.	<b>Latin-Hangul</b>	Du latin à l'hangûl.	<b>Latin-Katakana</b>	Du latin au katakana.	<b>Fullwidth-Halfwidth</b>	De pleine chasse à demi-chasse.
<b>Arabic-Latin</b>	De l'arabe au latin.										
<b>Greek-Latin</b>	Du grec au latin.										
<b>Latin-Hangul</b>	Du latin à l'hangûl.										
<b>Latin-Katakana</b>	Du latin au katakana.										
<b>Fullwidth-Halfwidth</b>	De pleine chasse à demi-chasse.										

## Options

**Tableau 27 : Options du Transliterator**

Option	Description/Valeurs valides
De	<p>Le script utilisé par les champs que vous souhaitez translittérer. Pour obtenir une description complète des scripts pris en charge, voir <a href="#">Transliterator</a> à la page 311.</p> <p><b>Remarque :</b> Le stage Transliterator ne prend pas en charge la translittération entre tous les scripts. Les champs <b>De</b> et <b>À</b> reflètent automatiquement les valeurs valides basées sur votre sélection.</p>
To	<p>Le script souhaité pour convertir le champ. Pour obtenir une description complète des scripts pris en charge, voir <a href="#">Transliterator</a> à la page 311.</p> <p><b>Remarque :</b> Le stage Transliterator ne prend pas en charge la translittération entre tous les scripts. Les champs <b>De</b> et <b>À</b> reflètent automatiquement les valeurs valides basées sur votre sélection.</p>
Bouton Échanger	<p>Cliquez sur le bouton d'échange pour échanger des langues dans les champs <b>De</b> et <b>À</b>.</p>
	
Champs à Translittérer	Indique les champs que vous souhaitez translittérer.
Supprimer les marques d'accent	Cochez la case pour supprimer les accents du champ. Cette option reste désactivée par défaut.

## Sortie

Le stage Transliterator translittère les champs spécifiés. Il ne produit pas de d'autre sortie.

# Module Universal Name

## Module Universal Name

Pour obtenir la plus grande précision de standardisation possible, vous devez décomposer les chaînes de données en plusieurs champs. Spectrum™ Technology Platform fournit des fonctionnalités d'analyse avancées vous permettant d'analyser les noms de personnes, les noms d'entreprises et de nombreux autres termes et abréviations. De plus, vous pouvez créer votre propre liste de termes personnalisés à utiliser comme base pour vos opérations d'analyse et d'extraction.

### Composants

Le module Universal Name est constitué des éléments suivants :

- **Name Parser (déprécié)** : ce composant décompose les noms de personnes, les noms d'entreprises et les autres termes du champ de données de nom en leurs éléments de composant. Ces éléments de nom analysés sont ensuite disponibles pour d'autres opérations automatisées telles que le rapprochement de noms, la standardisation de noms ou la consolidation de noms enregistrés plusieurs fois.

**Avertissement** : Le stage Name Parser est déprécié et il se peut qu'il ne soit pas pris en charge dans les versions à venir. Pour l'analyse de noms, utilisez Open Name Parser.

- **Name Variant Finder** : ce composant fonctionne en mode prénom ou nom de famille pour interroger la base de données des noms et renvoyer des variations de noms. Name Variant Finder peut limiter le nombre de résultats par sexe et culture. Spectrum™ Technology Platform inclut un fichier de noms de base. Les fichiers de noms d'extension des autres cultures sont disponibles et peuvent être déployés en copiant le fichier JAR approprié dans le dossier `modules/tables/ext`. Les noms arabes sont disponibles en tant qu'extension. Les données de ce fichier sont de Nomino, Inc.
- **Open Name Parser** : Open Name Parser décompose les noms de personnes et d'entreprises et d'autres termes du champ de données de nom en leurs éléments de composant. Ces éléments de nom analysés sont ensuite disponibles pour d'autres opérations automatisées telles que le rapprochement de noms, la standardisation de noms ou la consolidation de noms enregistrés plusieurs fois.

## Name Parser (DÉPRÉCIÉ)

**Avertissement** : Le stage Name Parser est déprécié et il se peut qu'il ne soit pas pris en charge dans les versions à venir. Pour l'analyse de noms, utilisez Open Name Parser.

Name Parser décompose les noms de personnes et d'entreprises et d'autres termes du champ de données de nom en leurs éléments de composant. Le processus de parsing comprend une explication de la fonction, la forme et la relation syntaxique de chaque partie de l'ensemble. Ces éléments de nom analysés sont ensuite disponibles pour d'autres opérations automatisées telles que le rapprochement de noms, la standardisation de noms ou la consolidation de noms enregistrés plusieurs fois.

L'option de parsing de nom effectue l'opération suivante :

- Détermine le type d'entité d'un nom afin de décrire la fonction que le nom effectue. Les types d'entité de nom sont divisés en deux principaux groupes : les noms de personnes et les noms d'entreprises. Ces groupes sont divisés en sous-groupes.
- Détermine la forme d'un nom afin de comprendre la syntaxe que le parseur doit suivre pour le parsing. Les noms personnels prennent habituellement un ordre naturel (signature) ou un ordre inverse. Les noms d'entreprises sont généralement ordonnés hiérarchiquement.
- Détermine et classe les éléments constitutifs d'un nom afin que la relation syntaxique de chaque partie du nom complet soit identifiée. La syntaxe de nom de personne comprend des préfixes, des prénoms, des deuxièmes prénoms et des noms de famille, des suffixes et des termes de description du compte, entre autres parties de noms de personnes. La syntaxe de nom d'entreprise inclut le texte principal, les termes insignifiants, les prépositions, les objets de la préposition et les termes de suffixe parmi les autres parties du nom d'entreprise.
- Détermine le sexe du nom. Le sexe est déterminé à partir d'hypothèses culturelles que vous spécifiez. Par exemple, Jean est un prénom masculin en France, mais un prénom féminin aux États-Unis. Si vous savez que les noms que vous traitez viennent de France, vous pouvez spécifier Français comme culture pour la détermination du sexe. Name Parser utilise les données des tables Prénom et Prénoms composés pour déterminer le sexe. Si un nom n'est pas trouvé dans l'une des tables et qu'un titre est présent dans le nom, le parseur vérifie la table Titre pour déterminer le sexe. Sinon, le genre est marqué comme inconnu.

**Remarque** : Si un champ dans votre enregistrement d'entrée contient déjà l'une des cultures prises en charge, vous pouvez prédéfinir le champ GenderDeterminationSource dans votre entrée pour remplacer le Gender Determination Source (Source de détermination du sexe) dans l'interface utilisateur graphique.

- Attribue un score de parsing qui indique le degré de confiance dont le parseur dispose pour vérifier si son parsing est correct.

## Entrée

**Avertissement** : Le stage Name Parser est déprécié et il se peut qu'il ne soit pas pris en charge dans les versions à venir. Pour l'analyse de noms, utilisez Open Name Parser.

**Tableau 28 : Entrée Name Parser**

Nom du champ	Description/Valeurs valides																										
GenderDeterminationSource	<p>Culture des données de nom à utiliser pour déterminer le genre. Utilise par défaut des règles interculturelles. Par exemple, Jean est généralement un nom féminin et la valeur par défaut l'identifie comme tel mais, si vous sélectionnez Français, le système l'identifie comme un nom masculin. Les options sont répertoriées ci-dessous avec des exemples de pays pour chaque culture. La liste de pays située sous chaque culture n'est pas exhaustive.</p> <table border="0"> <tr> <td><b>SLAVIC</b></td> <td>Bosnie, Pologne, Albanie.</td> </tr> <tr> <td><b>ARMENIAN</b></td> <td>Arménie.</td> </tr> <tr> <td><b>DEFAULT</b></td> <td>Bulgarie, îles Caïmans, Irlande, États-Unis, Royaume-Uni.</td> </tr> <tr> <td><b>FRENCH</b></td> <td>France.</td> </tr> <tr> <td><b>SCANDINAVIAN</b></td> <td>Danemark, Finlande, Islande, Norvège, Suède.</td> </tr> <tr> <td><b>GERMANIC</b></td> <td>Autriche, Allemagne, Luxembourg, Suisse, Pays-Bas.</td> </tr> <tr> <td><b>GREEK</b></td> <td>Grèce.</td> </tr> <tr> <td><b>HUNGARIAN</b></td> <td>Hongrie.</td> </tr> <tr> <td><b>ITALIAN</b></td> <td>Italie.</td> </tr> <tr> <td><b>PORTUGUESE</b></td> <td>Portugal.</td> </tr> <tr> <td><b>ROMANIA</b></td> <td>Roumanie.</td> </tr> <tr> <td><b>HISPANIC</b></td> <td>Espagne.</td> </tr> <tr> <td><b>ARABIC</b></td> <td>Tunisie.</td> </tr> </table> <p>GenderDeterminationSource est également utilisé par Name Variant Finder pour limiter les variations de nom renvoyées, basées sur la culture. Pour plus d'informations, reportez-vous à la section <a href="#">Name Variant Finder</a> à la page 338.</p>	<b>SLAVIC</b>	Bosnie, Pologne, Albanie.	<b>ARMENIAN</b>	Arménie.	<b>DEFAULT</b>	Bulgarie, îles Caïmans, Irlande, États-Unis, Royaume-Uni.	<b>FRENCH</b>	France.	<b>SCANDINAVIAN</b>	Danemark, Finlande, Islande, Norvège, Suède.	<b>GERMANIC</b>	Autriche, Allemagne, Luxembourg, Suisse, Pays-Bas.	<b>GREEK</b>	Grèce.	<b>HUNGARIAN</b>	Hongrie.	<b>ITALIAN</b>	Italie.	<b>PORTUGUESE</b>	Portugal.	<b>ROMANIA</b>	Roumanie.	<b>HISPANIC</b>	Espagne.	<b>ARABIC</b>	Tunisie.
<b>SLAVIC</b>	Bosnie, Pologne, Albanie.																										
<b>ARMENIAN</b>	Arménie.																										
<b>DEFAULT</b>	Bulgarie, îles Caïmans, Irlande, États-Unis, Royaume-Uni.																										
<b>FRENCH</b>	France.																										
<b>SCANDINAVIAN</b>	Danemark, Finlande, Islande, Norvège, Suède.																										
<b>GERMANIC</b>	Autriche, Allemagne, Luxembourg, Suisse, Pays-Bas.																										
<b>GREEK</b>	Grèce.																										
<b>HUNGARIAN</b>	Hongrie.																										
<b>ITALIAN</b>	Italie.																										
<b>PORTUGUESE</b>	Portugal.																										
<b>ROMANIA</b>	Roumanie.																										
<b>HISPANIC</b>	Espagne.																										
<b>ARABIC</b>	Tunisie.																										
Name	Nom à analyser. Ce champ est obligatoire.																										

## Options

**Avertissement** : Le stage Name Parser est déprécié et il se peut qu'il ne soit pas pris en charge dans les versions à venir. Pour l'analyse de noms, utilisez Open Name Parser.

Pour indiquer les options de Name Parser, double-cliquez sur l'instance de Name Parser sur le canevas. La boîte de dialogue Options de Name Parser apparaît.

**Tableau 29 : Options de Name Parser**

Option	Description
Analyser les noms de personnes	Cochez cette case pour analyser les noms des personnes.
Séparer les noms conjoints en plusieurs enregistrements. Sélectionnez les résultats d'une correspondance dans la Liste des résultats de correspondance, puis cliquez sur <b>Supprimer</b> .	<p>Cochez cette case pour séparer les noms contenant plus d'un individu en plusieurs enregistrements, par exemple, <i>Bill &amp; Sally Smith</i>.</p> <p>Lorsqu'un enregistrement conjoint engendre deux enregistrements de nom distincts, un champ de sortie ID d'enregistrement de l'analyseur est généré. Chaque paire d'enregistrements de nom distincts est identifiée avec le même ID d'enregistrement de l'analyseur.</p>
Source de détermination du genre. Sélectionnez les résultats d'une correspondance dans la Liste des résultats de correspondance, puis cliquez sur <b>Supprimer</b> .	<p>Détermine la manière dont le Name Parser attribue un genre au nom. Dans la plupart des cas, le paramètre par défaut est le plus adapté parce qu'il couvre une grande variété de noms. Si vous traitez des noms pour une culture spécifique, sélectionnez cette culture. La sélection d'une culture spécifique aide à garantir l'assignation d'un genre approprié aux noms. Par exemple, si vous laissez la sélection par défaut, le prénom Jean est identifié comme un prénom féminin. Si vous sélectionnez Français, il est identifié comme un prénom masculin.</p> <p><b>Remarque</b> : Si vous sélectionnez une culture mais que le nom est introuvable dans celle-ci, le genre est déterminé en fonction de la culture par défaut, qui inclut des données de différentes cultures.</p>
Order	<p>Indique la manière dont les champs de nom sont classés dans vos enregistrements d'entrée. L'un des éléments suivants :</p> <p><b>Naturel</b> Les champs de nom sont classés par titre, prénom, deuxième prénom, nom de famille et suffixe.</p> <p><b>Inverse</b> Les champs de nom sont classés d'abord par le nom de famille.</p> <p><b>Mélangé</b> Les champs de nom sont classés à l'aide d'une combinaison des classements naturel et inverse.</p>



Option	Description
Conserver les points	Conserve la ponctuation dans le champ de nom de personne analysé.
Analyser les noms professionnels	Cochez cette case pour analyser les noms commerciaux.
Conserver les points	Cochez cette case pour rétablir la ponctuation dans le champ du nom de personne analysé.
Table définie par l'utilisateur	Cliquez sur une des tables définies par l'utilisateur pour ajouter des valeurs aux valeurs existantes dans les différentes tables de l'analyseur. Cette fonction vous permet de personnaliser les tables pour obtenir un environnement commercial unique. Cliquez sur Configurer pour sélectionner un fichier XML qui contient les valeurs à ajouter. Pour plus d'informations sur les tables définies par l'utilisateur, reportez-vous à <a href="#">Modification des tables Name Parser définies par l'utilisateur</a> à la page 321.

### **Modification des tables Name Parser définies par l'utilisateur**

**Avertissement** : Le stage Name Parser est déprécié et il se peut qu'il ne soit pas pris en charge dans les versions à venir. Pour l'analyse de noms, utilisez Open Name Parser.

Vous pouvez ajouter, modifier et supprimer des valeurs des tables Name Parser pour les personnaliser en fonction de votre environnement d'entreprise unique.

Les tables définies par l'utilisateur Name Parser sont des fichiers XML qui figurent par défaut dans le dossier `<Drive>:\Program Files\Pitney Bowes\Spectrum\server\modules\parser\data`. Spectrum™ Technology Platform inclut les tables définies par l'utilisateur suivantes :

[\*UserAccountDescriptions.xml\*](#)

**Tableau 30 : Colonnes UserAccountDescriptions.xml**

Nom de colonne	Description/Valeurs valides
LookupValue	Terme de recherche généralement trouvé dans une description de compte. Tout texte d'un seul mot. Non-respect de la casse.

Exemple d'entrée :

```
<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

      LookupValue
      ART
      AND

    </deleted-entry-group>
  </deleted-entries>
  <added-entries delimiter-character="|">

    LookupValue
    A/C
    ACCOUNT
    EXP

  </added-entries>
</table-data>
```

[UserCompanyPrepositions.xml](#)

**Tableau 31 : Colonnes UserCompanyPrepositions.xml**

Nom de colonne	Description/Valeurs valides
LookupValue	Toute préposition (par exemple, « de » ou « sur ») généralement trouvée dans les noms de société. Tout texte d'un seul mot. Non-respect de la casse.

Exemple d'entrée :

```
<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

      LookupValue
      AROUND
      NEAR

    </deleted-entry-group>
  </deleted-entries>
  <added-entries delimiter-character="|">
```

```

        LookupValue
        ABOUT
        AFTER
        ACROSS

    </added-entries>
</table-data>

```

### *UserCompanySuffixes.xml*

**Tableau 32 : Colonnes UserCompanySuffixes.xml**

Nom de colonne	Description/Valeurs valides
LookupValue	Tout suffixe généralement trouvé dans les noms de société. Exemples : « Inc. » et « Co. ». Tout texte constitué d'un seul mot. Non-respect de la casse.

#### Exemple d'entrée :

```

<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

        LookupValue
        SANDY
        CLUE

    </deleted-entry-group>
  </deleted-entries>
  <added-entries delimiter-character="|">

        LookupValue
        LTD
        LLC
        CO
        INC

    </added-entries>
</table-data>

```

[UserCompanyTerms.xml](#)**Tableau 33 : Colonnes UserCompanyTerms.xml**

Nom de colonne	Description/Valeurs valides
LookupValue	Tout terme généralement trouvé dans un nom de société. Tout texte d'un seul mot. Non-respect de la casse.

Exemple d'entrée :

```

<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

      LookupValue
      MARY
      BLUE

    </deleted-entry-group>
  </deleted-entries>
  <added-entries delimiter-character="|">

    LookupValue
    ARC
    ARCADE
    ASSEMBLY
    ARIZONA

  </added-entries>
</table-data>

```

[UserCompoundFirstNames.xml](#)

Cette table contient les prénoms composés définis par l'utilisateur. Les noms composés se composent de deux mots.

**Tableau 34 : Colonnes UserCompoundFirstNames.xml**

Nom de colonne	Description/Valeurs valides
FirstName	Prénom composé. Maximum de deux mots. Non-respect de la casse.

Nom de colonne	Description/Valeurs valides
Culture	Culture dans laquelle la combinaison Prénom/Genre s'applique. Vous pouvez utiliser une des valeurs valides dans le champ d'entrée GenderDeterminationSource. Pour plus d'informations, reportez-vous à la section <a href="#">Entrée</a> à la page 319.
Genre	Genre le plus souvent associé à cette combinaison Prénom/Culture. L'un des éléments suivants : <ul style="list-style-type: none"> <li><b>M</b> Le nom est masculin.</li> <li><b>F</b> Le nom est féminin.</li> <li><b>A</b> Ambigu. Le nom peut être masculin ou féminin.</li> <li><b>U</b> Inconnu. Le genre du nom n'est pas connu. Si ce champ est laissé vide, le genre Inconnu est déduit par défaut.</li> </ul>
Fréquence	Non utilisé dans cette version. Vous pouvez laisser cette colonne vide.

## Exemple d'entrée :

```

<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

      FirstName
      ANN MARIE
      BILLY JOE

    </deleted-entry-group>
    <deleted-entry-group>

      FirstName|Frequency
      KAREN SUE|0.126
      BILLY JOE|0.421

    </deleted-entry-group>
    <deleted-entry-group>

      FirstName|Gender|Culture
      JEAN ANN|M|DEFAULT
      JEAN CLUADE|F|FRENCH

    </deleted-entry-group>
  </deleted-entries>
  <added-entries delimiter-character="|">

```

```

        FirstName|Gender|Culture
        JOHN Henry|M|DEFAULT
        A'SHA A'MAR|F|ARABIC
        BILLY JO|A|DEFAULT

    </added-entries>
</table-data>

```

### *UserConjunctions.xml*

Cette table contient une liste de conjonctions définies par l'utilisateur, telles que « et », « ou » ou « & ».

#### **Tableau 35 : Colonnes UserConjunctions.xml**

Nom de colonne	Description/Valeurs valides
LookupValue	Toute conjonction. Doit être un mot unique. Non-respect de la casse.

Exemple d'entrées :

```

<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

        LookupValue
        FIND
        CARE
        %

    </deleted-entry-group>
  </deleted-entries>
  <added-entries delimiter-character="|">

        LookupValue
        &
        AND
        OR

    </added-entries>
</table-data>

```

*UserFirstNames.xml***Tableau 36 : Colonnes UserFirstNames.xml**

Nom de colonne	Description/Valeurs valides
FirstName	Prénom décrit par la ligne de la table. Non-respect de la casse.
Genre	Genre le plus souvent associé à cette combinaison Prénom/Culture. L'un des éléments suivants : <ul style="list-style-type: none"> <li><b>M</b> Le nom est masculin.</li> <li><b>F</b> Le nom est féminin.</li> <li><b>A</b> Ambigu. Le nom peut être masculin ou féminin.</li> <li><b>U</b> Inconnu. Le genre du nom n'est pas connu. Si ce champ est laissé vide, le genre Inconnu est déduit par défaut.</li> </ul>
Culture	Culture dans laquelle la combinaison Prénom/Genre s'applique. Vous pouvez utiliser une des valeurs valides dans le champ d'entrée GenderDeterminationSource. Pour plus d'informations, reportez-vous à la section <a href="#">Entrée</a> à la page 319.

## Exemple d'entrée :

```

<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

      FirstName
      AADEL
      AADIL

    </deleted-entry-group>
    <deleted-entry-group>

      FirstName
      A'SACE
      A'BOCKETT

    </deleted-entry-group>
    <deleted-entry-group>

      FirstName | Gender | Culture
      ALII | M | DEFAULT
  </deleted-entries>

```

```

        AISHA|F|ARABIC

    </deleted-entry-group>
</deleted-entry-group>

        FirstName|Gender
        JOHE|M

    </deleted-entry-group>
</deleted-entries>
<added-entries delimiter-character="|">

        FirstName|Gender|Culture
        JOHE|M|DEFAULT
        A'SHAN|F|ARABIC

    </added-entries>
</table-data>

```

### [UserGeneralSuffixes.xml](#)

Cette table contient une liste de suffixes définis par l'utilisateur, utilisés dans les noms de personnes et qui ne sont pas des suffixes de maturité, tels que « MD » ou « PhD ».

**Tableau 37 : Colonnes UserGeneralSuffixes.xml**

Nom de colonne	Description/Valeurs valides
LookupValue	Tout suffixe qui n'est pas fréquemment appliqué aux noms de personnes et qui n'est pas un suffixe de maturité. Doit être un mot unique. Non-respect de la casse.

Exemple d'entrée :

```

<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

        LookupValue
        AND
        WILL
        TUNA

    </deleted-entry-group>
  </deleted-entries>
</table-data>

```



```

        LookupValue
        ACCOUNTANT
        ATTORNEY
        ANALYST
        ASSISTANT

    </added-entries>
</table-data>

```

### *UserLastNamePrefixes.xml*

Cette table contient une liste de préfixes définis par l'utilisateur qui apparaissent dans le nom de famille d'une personne, tels que « Van », « De » ou « La ».

#### **Tableau 38 : Colonnes UserLastNamePrefixes.xml**

Nom de colonne	Description/Valeurs valides
LookupValue	Tout préfixe qui fait partie du nom de famille d'un individu. Tout texte d'un seul mot. Non-respect de la casse.

#### Exemple d'entrée :

```

<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

        LookupValue
        DO
        RUN
        ANIMAL

    </deleted-entry-group>
  </deleted-entries>
  <added-entries delimiter-character="|">

        LookupValue
        D'
        DA
        DEN
        DEL

    </added-entries>

```

```
</table-data>
```

### *UserLastNames.xml*

**Tableau 39 : Colonnes UserLastNames.xml**

Nom de colonne	Description/Valeurs valides
LastName	Nom de famille décrit par la ligne de la table. Non-respect de la casse.
Genre	Genre le plus souvent associé à cette combinaison Prénom/Culture. L'un des éléments suivants : <ul style="list-style-type: none"> <li><b>M</b> Le nom est masculin.</li> <li><b>F</b> Le nom est féminin.</li> <li><b>A</b> Ambigu. Le nom peut être masculin ou féminin.</li> <li><b>U</b> Inconnu. Le genre du nom n'est pas connu. Si ce champ est laissé vide, le genre Inconnu est déduit par défaut.</li> </ul>
Culture	Culture dans laquelle la combinaison Prénom/Genre s'applique. Vous pouvez utiliser une des valeurs valides dans le champ d'entrée GenderDeterminationSource. Pour plus d'informations, reportez-vous à la section <a href="#">Entrée</a> à la page 319.

#### Exemple d'entrée :

```
<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

      LastName
      Rusod
      AADIL

    </deleted-entry-group>
  </deleted-entry-group>

  <deleted-entry-group>

    LastName
    KAASEEY
    JOIEN

  </deleted-entry-group>
```

```

</deleted-entries>
<added-entries delimiter-character="|">
    LastName|Culture|Gender
    SMITH|ENGLISH|A
    WILSON|ENGLISH|A
    JONES|ENGLISH|A
</added-entries>
</table-data>

```

### *UserMaturitySuffixes.xml*

Cette table contient des suffixes générationnels définis par l'utilisateur, utilisés dans le nom d'une personne, tels que « Jr. » ou « Sr. ».

#### **Tableau 40 : Colonnes UserMaturitySuffixes.xml**

Nom de colonne	Description/Valeurs valides
LookupValue	Suffixe générationnel utilisé dans les noms de personnes. Tout texte d'un seul mot. Non-respect de la casse.

#### Exemple d'entrée :

```

<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>
      LookupValue
      I
      V
      18
      VI
    </deleted-entry-group>
  </deleted-entries>
  <added-entries delimiter-character="|">
    LookupValue
    I
    II
    III
  </added-entries>

```

```
</table-data>
```

### UserTitles.xml

Cette table contient des titres définis par l'utilisateur, utilisés dans le nom d'une personne, tels que « Mr. » ou « Ms. ».

**Tableau 41 : Colonnes UserTitles.xml**

Nom de colonne	Description/Valeurs valides
LookupValue	Titre utilisé dans les noms de personnes. Tout texte d'un seul mot. Non-respect de la casse.
Genre	Genre le plus souvent associé à ce titre. L'un des éléments suivants : <ul style="list-style-type: none"> <li><b>M</b> Le nom est masculin.</li> <li><b>F</b> Le nom est féminin.</li> <li><b>A</b> Ambigu. Le nom peut être masculin ou féminin.</li> <li><b>U</b> Inconnu. Le genre du nom n'est pas connu. Si ce champ est laissé vide, le genre Inconnu est déduit par défaut.</li> </ul>

### Exemple d'entrée :

```
<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>
      LookupValue
      Belt
      Friend
      Thursday
      Red
    </deleted-entry-group>
  </deleted-entries>
  <added-entries delimiter-character="|">
    LookupValue | Gender
    Mrs | F
    Mr | M
    Most | F
```

```

    </added-entries>
</table-data>

```

### Exemple de table définie par l'utilisateur

La figure ci-dessous présente l'exemple de table UserFirstNames.xml et la syntaxe à utiliser lors de la modification de tables définies par l'utilisateur.

```

<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

      FirstName
      AADEL
      AADIL

    </deleted-entry-group>
    <deleted-entry-group>

      FirstName|Frequency
      A'SACE|0.126
      A'BECKETT|0.421

    </deleted-entry-group>
    <deleted-entry-group>

      FirstName|Gender|Culture|VariantGroup
      ALI|M|DEFAULT|GROUP88
      AISHA|F|ARABIC|GROUP43

    </deleted-entry-group>
    <deleted-entry-group>

      FirstName|Gender
      JOHN|M

    </deleted-entry-group>
  </deleted-entries>
  <added-entries delimiter-character="|">

    FirstName|Gender|Culture
    JOHN|M|DEFAULT
    A'SHA|F|ARABIC
    JAMES|M|DEFAULT

  </added-entries>
</table-data>

```

## Sortie

**Avertissement** : Le stage Name Parser est déprécié et il se peut qu'il ne soit pas pris en charge dans les versions à venir. Pour l'analyse de noms, utilisez Open Name Parser.

**Tableau 42 : Sortie de Name Parser**

Nom du champ	Format	Description/Valeurs valides
AccountDescription	Chaîne	Description de compte qui fait partie du nom. Par exemple, dans « Mary Jones Account # 12345 », la description de compte est « Account#12345 ».
EntityType	Chaîne	Indique le type de nom. L'un des éléments suivants : <b>Firm</b> Le nom est un nom de société. <b>Personal</b> Le nom est un nom de personne individuelle.
<b>Champs liés aux noms de sociétés</b>		
FirmModifier.1.Object	Chaîne	Premier objet d'une préposition apparaissant dans le nom de société. Par exemple, dans le nom de société « Pratt & Whitney Division of United Technologies », le premier objet d'une préposition est « United Technologies ».
FirmModifier.1.Preposition	Chaîne	Première préposition apparaissant dans le nom de société. Par exemple, dans le nom de société « Pratt & Whitney Division of United Technologies », « of » est la première préposition.
FirmModifier.2.Object	Chaîne	Deuxième objet d'une préposition apparaissant dans le nom de société. Par exemple, dans le nom de société « Church of Our Lady of Lourdes », le deuxième objet d'une préposition est « Lourdes ».
FirmModifier.2.Preposition	Chaîne	Deuxième préposition apparaissant dans le nom de société. Par exemple, dans le nom de société « Church of Our Lady of Lourdes », la deuxième préposition est le deuxième « of ».
FirmName	Chaîne	Le nom d'une société. Par exemple, « Pitney Bowes, Inc. ».

Nom du champ	Format	Description/Valeurs valides
FirmPrimary	Chaîne	base du nom d'une société. Par exemple, « Pitney Bowes ».
FirmSuffix	Chaîne	Suffixe d'entreprise. Par exemple, « Co. » et « Inc. »
<b>Champs liés aux noms des personnes individuelles</b>		
FirstName	Chaîne	Prénom d'une personne.
FirstNameVariantGroup	Chaîne	ID numérique indiquant le groupe de noms similaires auquel le prénom appartient. Par exemple, Muhammad, Mohammed et Mehmet appartiennent tous au même groupe de variante de nom. L'ID de groupe réel est attribué lorsque les données complémentaires sont chargées.  Ce champ n'est renseigné que si vous avez acheté la fonction Groupe de variante de nom.
GenderCode	Chaîne	Genre d'une personne tel que déterminé par l'analyse du prénom. L'un des éléments suivants :  <b>A</b> Ambigu. Le nom est masculin et féminin. Par exemple, Pat. <b>F</b> Féminin. Le nom est féminin. <b>M</b> Masculin. Le nom est masculin. <b>U</b> Inconnu. Le nom est introuvable dans la table des genres.
GenderDeterminationSource	Chaîne	Culture utilisée pour déterminer le genre d'un nom. Si le nom est introuvable dans la table des genres, ce champ est vide.
GeneralSuffix	Chaîne	Suffixe général/professionnel d'une personne. Par exemple, MD ou PhD.
LastName	Chaîne	Nom de famille d'une personne.
MaturitySuffix	Chaîne	Suffixe de maturité/générationnel d'une personne. Par exemple, Jr. ou Sr.

Nom du champ	Format	Description/Valeurs valides
MiddleName	Chaîne	Deuxième prénom d'une personne.
NameScore	Chaîne	Score représentant la qualité de l'opération d'analyse, compris entre 0 et 100. 0 indique une mauvaise qualité et 100 une excellente qualité.
ParserRecordID	Chaîne	ID unique attribué à chaque enregistrement d'entrée.
TitleOfRespect	Chaîne	Titre d'une personne, tel que Mr., Mrs., Dr. ou Rev.
<b>Champs liés aux noms conjoints</b>		
PersonalName.2.FirstName	Chaîne	Prénom de la deuxième personne dans un nom conjoint. Exemple de nom conjoint : « John and Jane Smith ».
PersonalName.2.FirstNameVariantGroup	Chaîne	ID numérique indiquant le groupe de noms similaires auquel le prénom de la deuxième personne d'un nom conjoint appartient. Par exemple, Muhammad, Mohammed et Mehmet appartiennent tous au même groupe de variante de nom. L'ID de groupe réel est attribué lorsque les données complémentaires sont chargées.  Ce champ n'est renseigné que si vous avez acheté la fonction Groupe de variante de nom.
PersonalName.2.GenderCode	Chaîne	Genre de la deuxième personne dans un nom conjoint, tel que déterminé par l'analyse du prénom effectuée par Name Parser. Exemple de nom conjoint : « John and Jane Smith ». L'un des éléments suivants :  <b>A</b> Ambigu. Le nom est masculin et féminin. Par exemple, Pat. <b>F</b> Féminin. Le nom est féminin. <b>M</b> Masculin. Le nom est masculin. <b>U</b> Inconnu. Le nom est introuvable dans la table des genres.
PersonalName.2.GenderDeterminationSource	Chaîne	Culture utilisée pour déterminer le genre de la deuxième personne dans un nom conjoint. Exemple de nom conjoint : « John and Jane Smith ».



Nom du champ	Format	Description/Valeurs valides
PersonalName.2.GeneralSuffix	Chaîne	Suffixe général/professionnel de la deuxième personne dans un nom conjoint. Exemple de nom conjoint : « John and Jane Smith ». Exemples de suffixes généraux : MD et PhD.
PersonalName.2.LastName	Chaîne	Nom de famille de la deuxième personne dans un nom conjoint. Exemple de nom conjoint : « John and Jane Smith ».
PersonalName.2.MaturitySuffix	Chaîne	Suffixe de maturité/générationnel de la deuxième personne dans un nom conjoint. Exemple de nom conjoint : « John and Jane Smith ». Exemple de suffixe de maturité : Jr. et Sr.
PersonalName.2.MiddleName	Chaîne	Deuxième prénom de la deuxième personne dans un nom conjoint. Exemple de nom conjoint : « John and Jane Smith ».
PersonalName.2.TitleOfRespect	Chaîne	Civilité de la deuxième personne dans un nom conjoint. Par exemple, « Mr. and Mrs. Smith » est un nom lié. Exemples de civilités : Mr., Mrs. et Dr.
PersonalName.3.FirstName	Chaîne	Prénom de la troisième personne dans un nom conjoint. Par exemple, « Mr. & Mrs. John Smith & Dr. Mary Jones » est un nom lié.
PersonalName.3.FirstNameVariantGroup	Chaîne	ID numérique indiquant le groupe de noms similaires auquel le prénom de la deuxième personne d'un nom conjoint appartient. Par exemple, Muhammad, Mohammed et Mehmet appartiennent tous au même groupe de variante de nom. L'ID de groupe réel est attribué lorsque les données complémentaires sont chargées.  Ce champ n'est renseigné que si vous avez acheté la fonction Groupe de variante de nom.

Nom du champ	Format	Description/Valeurs valides
PersonalName.3.GenderCode	Chaîne	Genre de la troisième personne dans un nom conjoint, tel que déterminé par l'analyse du prénom effectuée par Name Parser. Exemple de nom lié : « Mr. & Mrs. John Smith & Adam Jones ». L'un des éléments suivants :  <b>A</b> Ambigu. Le nom est masculin et féminin. Par exemple, Pat. <b>F</b> Féminin. Le nom est féminin. <b>M</b> Masculin. Le nom est masculin. <b>U</b> Inconnu. Le nom est introuvable dans la table des genres.
PersonalName.3.GenderDeterminationSource	Chaîne	Culture utilisée pour déterminer le genre de la troisième personne dans un nom conjoint. « Mr. & Mrs. John Smith & Adam Jones ».
PersonalName.3.GeneralSuffix	Chaîne	Suffixe général/professionnel de la troisième personne dans un nom conjoint. Exemple de nom lié : « Mr. & Mrs. John Smith & Adam Jones PhD ». Exemples de suffixes généraux : MD et PhD.
PersonalName.3.LastName	Chaîne	Nom de famille de la troisième personne dans un nom conjoint. Par exemple, « Mr. & Mrs. John Smith & Dr. Mary Jones » est un nom lié.
PersonalName.3.MaturitySuffix	Chaîne	Suffixe de maturité/générationnel de la troisième personne dans un nom conjoint. Exemple de nom lié : « Mr. & Mrs. John Smith & Adam Jones Sr. ». Exemples de suffixes de maturité : Jr. et Sr.
PersonalName.3.MiddleName	Chaîne	Deuxième prénom de la troisième personne dans un nom conjoint. Par exemple, « Mr. & Mrs. John Smith & Dr. Mary Jones » est un nom lié.
PersonalName.3.TitleOfRespect	Chaîne	Civilité de la troisième personne dans un nom conjoint. Par exemple, « Mr. & Mrs. John Smith & Dr. Mary Jones » est un nom lié. Exemples de civilités : Mr., Mrs. et Dr.

## Name Variant Finder

Name Variant Finder fonctionne en mode prénom ou nom de famille pour interroger une base de données afin de renvoyer différentes versions d'un nom. Par exemple, « John » et « Jon » sont des

variantes du prénom « Jonathan ». Name Variant Finder exige des dictionnaires complémentaires qui peuvent être installés à l'aide de l'utilitaire de chargement de base de données du module Universal Name, du module Data Normalization, et du module Advanced Matching. Contactez votre représentant commercial pour plus d'informations sur la façon d'obtenir ces dictionnaires facultatifs spécifiques à la culture.

## Entrée

**Tableau 43 : Champs d'entrée Name Variant Finder**

Nom du champ	Description/Valeurs valides
FirstName	Nom pour lequel rechercher des variantes, s'il s'agit d'un nom donné.
LastName	Nom pour lequel rechercher des variantes, s'il s'agit d'un nom de famille.
GenderCode	<p>Genre du nom dans le champ Prénom. L'un des éléments suivants :</p> <p><b>Remarque :</b> Les codes de genre ne s'appliquent qu'aux prénoms, par aux noms de famille.</p> <p><b>M</b> Le nom est masculin.</p> <p><b>F</b> Le nom est féminin.</p> <p><b>A</b> Ambigu. Le nom peut être masculin ou féminin.</p> <p><b>U</b> Inconnu. Le genre du nom n'est pas connu.</p>
Ethnicity	<p>Culture la plus souvent associée au nom figurant dans le champ Prénom ou Nomdefamille. Vous pouvez utiliser les étapes de Name Parser ou Open Parser pour renseigner ce champ si vous ne connaissez pas l'ethnicité d'un nom.</p> <p><b>Remarque :</b> Ce champ était autrefois nommé GenderDeterminationSource.</p>

## Options

Tableau 44 : Options de Name Variant Finder

Option	Description
Prénom	Recherche des variations de nom en fonction du prénom.
Nom de famille	Recherche des variations de nom en fonction du nom de famille.
Code de genre	Renvoie les variations de nom uniquement pour le genre spécifié dans le champ GenderCode de l'enregistrement. Pour plus d'informations sur le champ GenderCode, reportez-vous à <a href="#">Entrée</a> à la page 339.
Ethnicité	Renvoie les variations de nom uniquement pour la culture spécifiée dans le champ Ethnicité de l'enregistrement. Pour plus d'informations sur le champ Ethnicité, reportez-vous à <a href="#">Entrée</a> à la page 339.
Romanisé	Renvoie la version romanisée anglaise du nom. Un nom romanisé est un nom converti d'un script non latin à un script latin. Par exemple, Achin est la version romanisée du nom coréen 아진.
Natif	Renvoie le nom dans le script natif de la culture du nom. Par exemple, un nom coréen serait renvoyé en Hangul.
Kana	Si vous sélectionnez l'option <b>Natif</b> , vous pouvez choisir de renvoyer les noms japonais en Kana, en sélectionnant l'option correspondante. Le Kana se compose des scripts hiragana et katakana.  <b>Remarque</b> : Votre licence doit inclure la base de données Pack Asian Plus pour rechercher des variantes de nom japonaises. Pour plus d'informations, contactez votre cadre commercial.

Option	Description
Kanji	<p>Si vous sélectionnez l'option <b>Natif</b>, vous pouvez choisir de renvoyer les noms japonais en Kanji, en sélectionnant l'option correspondante. Le Kanji est l'un des scripts utilisés dans la langue japonaise.</p> <p><b>Remarque</b> : Votre licence doit inclure la base de données Pack Asian Plus pour rechercher des variantes de nom japonaises. Pour plus d'informations, contactez votre cadre commercial.</p>

## Sortie

**Tableau 45 : Sorties de Name Variant Finder**

Nom du champ	Format	Description/Valeurs valides
CandidateGroup	Chaîne	Identifie le regroupement d'un nom d'entrée et de ses variations de nom. Chaque nom d'entrée reçoit un numéro CandidateGroup. Les variations de ce nom d'entrée reçoivent le même numéro CandidateGroup.
Ethnicity	Chaîne	<p>Culture d'un nom déterminée par les dictionnaires Core Name et complémentaires.</p> <p><b>Remarque</b> : Ce champ était autrefois nommé GenderDeterminationSource.</p>
FirstName	Chaîne	Nom donné d'une personne.
GenderCode	Chaîne	<p>Genre d'un nom déterminé par les dictionnaires Core Name et complémentaires. L'un des éléments suivants :</p> <p><b>M</b> Le nom est masculin.</p> <p><b>F</b> Le nom est féminin.</p> <p><b>A</b> Ambigu. Le nom peut être masculin ou féminin.</p> <p><b>U</b> Inconnu. Le genre du nom n'est pas connu.</p>
LastName	Chaîne	Nom de famille d'une personne.

Nom du champ	Format	Description/Valeurs valides
TransactionalRecordType	Chaîne	Indique la manière dont le nom a été utilisé dans le processus de correspondance. L'un des éléments suivants : <ul style="list-style-type: none"> <li><b>Suspect</b> Un enregistrement suspect est utilisé comme entrée dans une requête.</li> <li><b>Candidate</b> C'est un enregistrement candidat qui sera un résultat renvoyé par une requête.</li> </ul>

## OpenNameParser

L'option OpenNameParser décompose les noms personnels et d'entreprises et d'autres termes dans le champ de données nom en composantes. Ces éléments de nom analysés sont ensuite disponibles pour d'autres opérations automatisées telles que le rapprochement de noms, la standardisation de noms ou la consolidation de noms enregistrés plusieurs fois.

OpenNameParser effectue l'opération suivante :

- Détermine le type d'entité d'un nom afin de décrire la fonction que le nom remplit. Les types d'entité de nom sont divisés en deux principaux groupes : les noms de personnes et les noms d'entreprises. Au sein de chacun de ces principaux groupes figurent des sous-groupes.
- Détermine la forme d'un nom afin de comprendre la syntaxe que le parseur doit suivre pour le parsing. Les noms personnels prennent habituellement un ordre naturel (signature) ou un ordre inverse. Les noms d'entreprises sont généralement ordonnés hiérarchiquement.
- Détermine et classe les éléments constitutifs d'un nom afin que la relation syntaxique de chaque partie du nom complet soit identifiée. La syntaxe de nom de personne comprend des préfixes, des prénoms, des deuxièmes prénoms et des noms de famille, des suffixes et des termes de description du compte, entre autres parties de noms de personnes. La syntaxe de nom d'entreprise comprend des noms d'entreprises et des termes de suffixe.
- Les analyses lient les noms de personnes et d'entreprises et soit elles les conservent sous forme d'enregistrement unique, soit elles les divisent en plusieurs enregistrements. Exemple de noms liés : « Mr. and Mrs. John Smith » et « Baltimore Gas & Electric dba Constellation Energy ».
- Les résultats des analyses se présentent sous forme d'enregistrements ou de liste.
- Vous permet d'utiliser Open Parser Domain Editor pour créer de nouveaux domaines susceptibles d'être utilisés dans les options avancées Open Name Parser.
- Affecte un score d'analyse qui reflète le niveau de confiance vis-à-vis de l'exactitude de l'analyse.

## Entrée

**Tableau 46 : Entrée Open Name Parser**

Nom du champ	Description								
CultureCode	<p>Culture des données de nom en entrée. Une liste des options figure ci-dessous.</p> <table border="0"> <tr> <td><b>Null (empty)</b></td> <td>Culture mondiale (par défaut)</td> </tr> <tr> <td><b>de</b></td> <td>Allemand</td> </tr> <tr> <td><b>es</b></td> <td>Espagnol</td> </tr> <tr> <td><b>ja</b></td> <td>Japonais</td> </tr> </table> <p><b>Remarque :</b> Si vous avez ajouté votre propre domaine via Open Parser Domain Editor, les cultures et les codes de culture de ce domaine sont également valides.</p>	<b>Null (empty)</b>	Culture mondiale (par défaut)	<b>de</b>	Allemand	<b>es</b>	Espagnol	<b>ja</b>	Japonais
<b>Null (empty)</b>	Culture mondiale (par défaut)								
<b>de</b>	Allemand								
<b>es</b>	Espagnol								
<b>ja</b>	Japonais								
Name	Nom à analyser. Ce champ est obligatoire.								

## Options

Il est possible de configurer les options Open Name Parser au niveau du stage, via l'un des clients Spectrum™ Technology Platform, ou lors de l'exécution, à l'aide des options de flux de données.

### Options de parsing

Le tableau suivant répertorie les options qui contrôlent l'analyse des noms.

Tableau 47 : Options de parsing Open Name Parser

Nom de l'option	Description
Analyser les noms de personnes	Indique s'il convient ou non d'analyser les noms de personnes.  <b>Naturel</b> Les champs de nom sont classés par titre, prénom, deuxième prénom, nom de famille et suffixe.  <b>Inverse</b> Les champs de nom sont classés d'abord par le nom de famille.  <b>Les deux</b> Les champs de nom sont classés à l'aide d'une combinaison des classements naturel et inverse.
Noms liés	Indique s'il convient d'analyser les noms liés.
Scinder les noms liés en plusieurs enregistrements	Indique s'il convient de séparer les noms contenant plus d'un individu en plusieurs enregistrements, par exemple, Bill & Sally Smith.  Utilisez un stage Unique ID Generator pour créer un ID pour chacun des enregistrements scindés.
Analyser les noms professionnels	Indique s'il convient d'analyser les noms d'entreprises.
Résultats de sortie sous forme de liste	Indique s'il convient de renvoyer les éléments de nom analysés sous forme de liste.
Seuil de raccourci	Indique le mode d'équilibrage des performances par rapport à la qualité. Des performances plus rapides entraînent une sortie de moindre qualité ; en revanche, une qualité supérieure entraîne des performances plus lentes. Lorsque ce seuil est atteint, aucun autre traitement n'est réalisé sur l'enregistrement.  La valeur par défaut est 100.

### Options de cultures

Le tableau suivant répertorie les options qui contrôlent les cultures des noms.



**Tableau 48 : Options de cultures Open Name Parser**

Nom de l'option	Description
Cultures	<p>Indique la ou les cultures à inclure dans la grammaire d'analyse. Culture mondiale est la sélection par défaut.</p> <p><b>Remarque</b> : Si vous avez ajouté votre propre domaine via Open Parser Domain Editor, les cultures et les codes de culture de ce domaine apparaissent ici eux aussi.</p> <p>Cliquez sur les boutons <b>Monter</b> et <b>Descendre</b> pour définir l'ordre d'exécution des cultures.</p>

### Options avancées

Le tableau suivant répertorie les options avancées de l'analyse de noms.

**Tableau 49 : Options avancées Open Name Parser**

Option	Description
Options avancées	<p>Utilisez la liste déroulante Domaine pour sélectionner le domaine approprié pour chaque Nom.</p> <p>Cliquez sur les boutons <b>Monter</b> et <b>Descendre</b> pour définir l'ordre d'exécution des analyseurs. Les résultats sont renvoyés pour le premier domaine dont le score est supérieur à la valeur définie dans le champ <b>Seuil de raccourci</b>. Si aucun domaine n'atteint ce seuil, les résultats du domaine dont le score est le plus élevé sont renvoyés. Si plusieurs domaines atteignent le seuil en même temps, la priorité est donnée au domaine exécuté en premier (déterminé par l'ordre défini ici) et ses résultats sont renvoyés.</p> <p><b>Remarque</b> : Si vous avez ajouté votre propre domaine via Open Parser Domain Editor, ce domaine apparaît ici lui aussi.</p>

### Configuration des options lors de l'exécution

Les options Open Name Parser peuvent être configurées et transmises lors de l'exécution si elles sont exposées sous forme d'options de flux de données. Cela vous permet de remplacer la configuration existante par des chaînes d'analyse de noms au format JSON. Vous pouvez également

définir des options de stage lors de l'appel d'un job via un flux de processus ou l'outil de ligne de commande Exécuteur de job.

Pour définir des options Open Name Parser lors de l'exécution :

1. Dans Enterprise Designer, ouvrez un flux de données qui utilise le stage Open Name Parser.
2. Enregistrez et exposez ce flux de données.
3. Accédez à `Edit > Dataflow Options`.
4. Dans la table **Map dataflow options to stages**, développez Open Name Parser et modifiez les options, le cas échéant. Cochez la case de l'option que vous souhaitez modifier, puis modifiez la valeur dans la liste déroulante **Valeur par défaut**.
5. Facultatif : modifiez le nom des options du champ **Option label**.
6. Cliquez deux fois sur **OK**.

## Réponse

### Tableau 50 : Sortie Open Name Parser

Nom du champ	Format	Description
AccountDescription	Chaîne	Description de compte qui fait partie du nom. Par exemple, dans « Mary Jones Account # 12345 », la description de compte est « Account#12345 ».
Names	Chaîne	Champ hiérarchique contenant une liste d'éléments analysés. Ce champ est renvoyé lorsque vous cochez la case <b>Résultats de sortie sous forme de liste</b> sous Options de parsing.
<b>Champs liés aux noms de sociétés</b>		
FirmConjunction	Chaîne	Indique que le nom d'une société contient une conjonction comme « d/b/a » (doing business as), « o/a » (operating as) ou « t/a » (trading as).
FirmName	Chaîne	Le nom d'une société. Par exemple, « Pitney Bowes ».
FirmSuffix	Chaîne	Suffixe d'entreprise. Par exemple, « Co. » et « Inc. »

Nom du champ	Format	Description
IsFirm	Chaîne	Indique que le nom est celui d'une entreprise et non d'un individu.
<b>Champs liés aux noms des personnes individuelles</b>		
Conjunction	Chaîne	Indique que le nom contient une conjonction comme « and », « or » ou « & ».
CultureCode	Chaîne	Les codes de culture contenus dans les données d'entrée.
CultureCodeUsedToParse	Chaîne	Identifie la grammaire propre à une culture utilisée pour analyser les données. <b>Null (empty)</b> Culture mondiale (par défaut) <b>de</b> Allemand <b>es</b> Espagnol <b>ja</b> Japonais  <b>Remarque :</b> Si vous avez ajouté votre propre domaine via Open Parser Domain Editor, les cultures et les codes de culture de ce domaine apparaissent eux aussi dans ce champ.
FirstName	Chaîne	Prénom d'une personne.
GeneralSuffix	Chaîne	Suffixe général/professionnel d'une personne. Par exemple, MD ou PhD.
IsParsed	Chaîne	Indique si un enregistrement de sortie a été analysé ou non. Les valeurs valides sont true ou false.
IsPersonal	Chaîne	Indique si le nom est celui d'un individu et non d'une société. Les valeurs valides sont true ou false.
IsReverseOrder	Chaîne	Indique si le nom d'entrée est dans l'ordre inverse ou non. Les valeurs valides sont true ou false.

Nom du champ	Format	Description
LastName	Chaîne	Nom de famille d'une personne. Inclut le nom de famille paternel.
LeadingData	Chaîne	Informations autres que le nom qui apparaissent avant un nom.
MaturitySuffix	Chaîne	Suffixe de maturité/générationnel d'une personne. Par exemple, Jr. ou Sr.
MiddleName	Chaîne	Deuxième prénom d'une personne.
Name.	Chaîne	Nom de personne ou d'entreprise fourni en entrée.
NameScore	Chaîne	Indique le score moyen des jetons connus et inconnus pour chaque nom. La valeur de NameScore est comprise entre 0 et 100, tel que défini dans la grammaire d'analyse. 0 signifie qu'aucun résultat n'a été trouvé.
SecondaryLastName	Chaîne	Dans la grammaire d'analyse espagnole, nom de famille de la mère d'une personne.
TitleOfRespect	Chaîne	Informations qui apparaissent avant un nom, comme « Mr. », « Mrs. » ou « Dr. ».
TrailingData	Chaîne	Informations autres que le nom qui apparaissent après un nom.
<b>Champs liés aux noms conjoints</b>		
Conjunction2	Chaîne	Indique qu'un deuxième nom lié contient une conjonction comme « and », « or » ou « & ».
Conjunction3	Chaîne	Indique qu'un troisième nom lié contient une conjonction comme « and », « or » ou « & ».

Nom du champ	Format	Description
FirmName2	Chaîne	Nom d'une deuxième entreprise liée. Par exemple : « Baltimore Gas & Electric dba Constellation Energy ».
FirmSuffix2	Chaîne	Suffixe d'une deuxième entreprise liée.
FirstName2	Chaîne	Prénom d'un deuxième nom lié.
FirstName3	Chaîne	Prénom d'un troisième nom lié.
GeneralSuffix2	Chaîne	Suffixe général/professionnel d'un deuxième nom lié. Par exemple, MD ou PhD.
GeneralSuffix3	Chaîne	Suffixe général/professionnel d'un troisième nom lié. Par exemple, MD ou PhD.
IsConjoined	Chaîne	Indique que le nom d'entrée est lié. Exemple de nom conjoint : « John and Jane Smith ».
LastName2	Chaîne	Nom de famille d'un deuxième nom lié.
LastName3	Chaîne	Nom de famille d'un troisième nom lié.
MaturitySuffix2	Chaîne	Suffixe de maturité/générationnel d'un deuxième nom lié. Par exemple, Jr. ou Sr.
MaturitySuffix3	Chaîne	Suffixe de maturité/générationnel d'un troisième nom lié. Par exemple, Jr. ou Sr.
MiddleName2	Chaîne	Deuxième prénom d'un deuxième nom lié.
MiddleName3	Chaîne	Deuxième prénom d'un troisième nom lié.

Nom du champ	Format	Description
TitleOfRespect2	Chaîne	Informations qui apparaissent avant un deuxième nom lié, comme « Mr. », « Mrs. » ou « Dr. ».
TitleOfRespect3	Chaîne	Informations qui apparaissent avant un troisième nom lié, comme « Mr. », « Mrs. » ou « Dr. ».

### Rapport de synthèse Open Name Parser

Le rapport de synthèse Open Name Parser répertorie les statistiques de synthèse concernant le job, comme le nombre total d'enregistrements d'entrée et le nombre total d'enregistrements ne contenant aucune donnée de nom, ainsi que différentes statistiques d'analyse. Pour obtenir des instructions d'utilisation des rapports, reportez-vous au *Guide Dataflow Designer Spectrum™ Technology Platform*.

#### Résultats généraux

- **Total number of input records** : nombre d'enregistrements contenus dans le fichier d'entrée.
- **Total number of records that contained no name data** : nombre d'enregistrements contenus dans le fichier d'entrée qui ne contenaient aucune donnée de nom à analyser.
- **Total number of names parsed out** : nombre de noms contenus dans le fichier d'entrée qui ont été analysés.
- **Total Records** : nombre total d'enregistrements traités.
- **Lowest name parsing score** : score d'analyse le plus faible attribué à un nom du fichier d'entrée.
- **Highest name parsing score** : score d'analyse le plus élevé attribué à un nom du fichier d'entrée.
- **Average name parsing score** : score d'analyse moyen de l'ensemble des noms analysés contenus dans le fichier d'entrée.

#### Résultats d'analyse de noms de personnes

- **Number of personal name records written** : nombre de noms de personnes contenus dans le fichier d'entrée.
- **Number of names parsed from conjoined names** : nombre de noms analysés provenant d'enregistrements qui contenaient des noms liés. Par exemple, si votre fichier d'entrée compte cinq enregistrements avec deux noms liés et sept enregistrements avec trois noms liés, la valeur de ce champ est 31, conformément à l'équation suivante :  $(5 \times 2) + (7 \times 3)$ .
- **Records with 2 conjoined names** : nombre d'enregistrements en entrée contenant deux noms liés.
- **Records with 3 conjoined names** : nombre d'enregistrements en entrée contenant trois noms liés.

- **Number of names with title of respect present** : nombre de noms analysés contenant une civilité.
- **Number of names with maturity suffix present** : nombre de noms analysés contenant un suffixe de maturité.
- **Number of names with general suffix present** : nombre de noms analysés contenant un suffixe général.
- **Number of names that contained account descriptions** : nombre de noms analysés contenant une description de compte.
- **Total Reverse Order Names** : nombre de noms analysés figurant dans l'ordre inverse et produisant une valeur « True » dans le champ de sortie isReversed.

#### *Résultats d'analyse de noms d'entreprises*

- **Number of business name records written** : nombre de noms d'entreprises contenus dans le fichier d'entrée.
- **Number of names with firm suffix present** : nombre de noms analysés contenant un suffixe d'entreprise.
- **Number of names that contained account descriptions** : nombre d'enregistrements en entrée contenant une description de compte.
- **Total DBA Records** : nombre d'enregistrements en entrée contenant des conjonctions Doing Business As (DBA), produisant une valeur « True » dans les deux champs de sortie isPersonal et isFirm.

# 9 - Prise en charge du module et des codes ISO de pays

## In this section

---

Prise en charge du module et des codes ISO de pays

353



## Prise en charge du module et des codes ISO de pays

Ce tableau répertorie les codes ISO pour chaque pays, ainsi que les modules qui prennent en charge l'adressage, le géocodage et le routage pour chaque pays.

Notez que le module Enterprise Geocoding inclut des bases de données pour l'Afrique (30 pays), le Moyen-Orient (8 pays) et l'Amérique latine (20 pays). Ces bases de données couvrent les pays plus petits dans les régions qui ne disposent pas de leurs propres bases de données de géocodage. La colonne Modules pris en charge indique les pays couverts par ces bases de données pour l'Afrique, le Moyen-Orient et l'Amérique latine.

En outre, la base de données Geocode Address World fournit un géocodage géographique et un géocodage postal limité (mais pas de géocodage au niveau des rues) pour tous les pays.

Nom de pays ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Modules pris en charge
Afghanistan	AF	AFG	Module Universal Addressing
Aland, îles	AX	ALA	Module Universal Addressing
Albanie	AL ou SQ (Routage)	ALB	Module Universal Addressing Module Enterprise Geocoding Module Enterprise Routing
Algérie	DZ	DZA	Module Enterprise Geocoding (Afrique) Module Universal Addressing
Samoa américaines	AS	ASM	Module Universal Addressing
Andorre	AD	AND	Module Enterprise Geocoding (Andorre est couverte par le géocodeur de l'Espagne) Module Universal Addressing

Nom de pays ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Modules pris en charge
Angola	AO	AGO	Module Enterprise Geocoding (Afrique)Module Universal Addressing
Anguilla	AI	AIA	Module Universal Addressing
Antarctique	AQ	ATA	Module Universal Addressing
Antigua et Barbuda	AG	ATG	Module Universal Addressing
Argentine	AR	ARG	Module Enterprise GeocodingModule Universal AddressingModule Enterprise Routing
Arménie	AM	ARM	Module Universal Addressing
Aruba	AW	ABW	Module Enterprise Geocoding (Amérique latine)Module Universal Addressing
Australie	AU	AUS	Module Enterprise GeocodingModule Enterprise RoutingModule Universal Addressing
Autriche	AT	AUT	Module Enterprise GeocodingModule Enterprise RoutingModule Universal Addressing
Azerbaïdjan	AZ	AZE	Module Universal Addressing
Bahamas	BS	BHS	Module Enterprise GeocodingModule Universal AddressingModule Enterprise Routing

Nom de pays ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Modules pris en charge
Bahreïn	BH	BHR	Module Enterprise Geocoding (Moyen-Orient) Module Universal Addressing
Bangladesh	BD	BGD	Module Universal Addressing
Barbade	BB	BRB	Module Enterprise Geocoding (Amérique latine) Module Universal Addressing
Biélorussie	BY	BLR	Module Universal Addressing Module Enterprise Routing
Belgique	BE	BEL	Module Enterprise Geocoding Module Enterprise Routing Module Universal Addressing
Belize	BZ	BLZ	Module Enterprise Geocoding (Amérique latine) Module Universal Addressing
Bénin	BJ	BEN	Module Enterprise Geocoding (Afrique) Module Universal Addressing
Bermudes	BM	BMU	Module Universal Addressing Module Enterprise Routing
Bhoutan	BT	BTN	Module Universal Addressing
Bolivie	BO	BOL	Module Enterprise Geocoding (Amérique latine) Module Universal Addressing

Nom de pays ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Modules pris en charge
Bonaire, Saint-Eustache et Saba	BQ	BES	Module Universal Addressing
Bosnie-Herzégovine	BA	BIH	Module Enterprise Geocoding Module Universal Addressing Module Enterprise Routing Module Enterprise Geocoding
Botswana	BW	BWA	Module Enterprise Geocoding (Afrique) Module Universal Addressing
Île Bouvet	BV	BVT	Module Universal Addressing
Brésil	BR	BRA	Module Enterprise Geocoding Module Enterprise Routing Module Universal Addressing
Territoire britannique de l'Océan indien	IO	IOT	Module Universal Addressing
Brunéi Darussalam	BN	BRN	Module Enterprise Geocoding Module Universal Addressing
Bulgarie	BG	BGR	Module Enterprise Geocoding Module Universal Addressing
Burkina Faso	BF	BFA	Module Enterprise Geocoding (Afrique) Module Universal Addressing
Burundi	BI	BDI	Module Enterprise Geocoding (Afrique) Module Universal Addressing
Cambodge	KH	KHM	Module Universal Addressing

Nom de pays ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Modules pris en charge
Cameroun	CM	CMR	Module Enterprise Geocoding (Afrique)Module Universal Addressing
Canada	CA	CAN	Module Enterprise GeocodingModule Enterprise RoutingModule Universal Addressing
Cap-Vert	CV	CPV	Module Universal Addressing
Îles Caïmans	KY	CYM	Module Universal Addressing
République centrafricaine	CF	CAF	Module Universal Addressing
Tchad	TD	TCD	Module Universal Addressing
Chili	CL	CHL	Module Enterprise GeocodingModule Universal AddressingModule Enterprise Routing
Chine	CN ou zh_CN (Routage)	CHN	Module Enterprise GeocodingModule Universal AddressingModule Enterprise Routing
Île Christmas	CX	CXR	Module Universal Addressing
Îles Cocos (Keeling)	CC	CCK	Module Universal Addressing
Colombie	CO	COL	Module Enterprise GeocodingModule Universal Addressing
Comores	KM	COM	Module Universal Addressing

Nom de pays ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Modules pris en charge
Congo, République du	CG	COG	Module Enterprise Geocoding (Afrique) Module Universal Addressing
Congo, République démocratique du	CD	COD	Module Enterprise Geocoding (Afrique) Module Universal Addressing Module Enterprise Routing
Îles Cook	CK	COK	Module Universal Addressing
Costa Rica	RETOUR CHARIOT (CR)	CRI	Module Enterprise Geocoding (Amérique latine) Module Universal Addressing
Côte d'Ivoire	CI	CIV	Module Universal Addressing
Croatie	HR	HRV	Module Enterprise Geocoding Module Enterprise Routing Module Universal Addressing
Cuba	CU	CUB	Module Enterprise Geocoding (Amérique latine) Module Enterprise Routing Module Universal Addressing
Curaçao	CW	CUW	Module Universal Addressing
Chypre	CY	CYP	Module Enterprise Geocoding Module Universal Addressing
République tchèque	CZ ou CS (Routage)	CZE	Module Enterprise Geocoding Module Universal Addressing Module Enterprise Routing
Danemark	DK	DNK	Module Enterprise Geocoding Module Enterprise Routing Module Universal Addressing

Nom de pays ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Modules pris en charge
Djibouti	DJ	DJI	Module Universal Addressing
Dominique	DM	DMA	Module Universal Addressing
République dominicaine	DO	DOM	Module Enterprise Geocoding (Amérique latine) Module Universal Addressing
Équateur	EC	ECU	Module Enterprise Geocoding (Amérique latine) Module Universal Addressing
Égypte	EG	EGY	Module Enterprise Geocoding (Moyen-Orient) Module Universal Addressing
El Salvador	SV	SLV	Module Enterprise Geocoding (Amérique latine) Module Universal Addressing
Guinée équatoriale	GQ	GNQ	Module Universal Addressing
Erythrée	ER	ERI	Module Universal Addressing
Estonie	EE	EST	Module Enterprise Geocoding Module Enterprise Routing Module Universal Addressing
Éthiopie	ET	ETH	Module Universal Addressing
Falkland, îles (Malvinas)	FK	FLK	Module Universal Addressing
Îles Féroé	FO	FRO	Module Universal Addressing

Nom de pays ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Modules pris en charge
Fiji	FJ	FJI	Module Universal Addressing
Finlande	FI	FIN	Module Entreprise Geocoding Module Entreprise Routing Module Universal Addressing
France	FR	FRA	Module Entreprise Geocoding Module Entreprise Routing Module Universal Addressing
Guyane française	GF	GUF	Module Entreprise Geocoding ( <i>La Guyane française est couverte par le géocodeur de la France</i> ) Module Universal Addressing
Polynésie française	PF	PYF	Module Universal Addressing
Terres australes françaises	TF	ATF	Module Universal Addressing
Gabon	GA	GAB	Module Entreprise Geocoding (Afrique) Module Universal Addressing
Gambie	GM	GMB	Module Universal Addressing
Géorgie	GE	GEO	Module Universal Addressing
Allemagne	DE	DEU	Module Entreprise Geocoding Module Entreprise Routing Module Universal Addressing
Ghana	GH	GHA	Module Entreprise Geocoding (Afrique) Module Universal Addressing Module Entreprise Routing



Nom de pays ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Modules pris en charge
Gibraltar	GI	GIB	Module Entreprise Geocoding ( <i>Gibraltar est couverte par le géocodeur de l'Espagne</i> )Module Universal Addressing
Grèce	GR	GRC	Module Entreprise GeocodingModule Universal Addressing
Groenland	GL	GRL	Module Universal Addressing
Grenade	GD	GRD	Module Universal Addressing
Guadeloupe	GP	GLP	Module Entreprise Geocoding ( <i>La Guadeloupe est couverte par le géocodeur de la France.</i> )Module Universal Addressing
Guam	GU	GUM	Module Universal Addressing
Guatemala	GT	GTM	Module Entreprise Geocoding (Amérique latine)Module Universal Addressing
Guernesey	GG	GGY	Module Universal Addressing
Guinée	GN	GIN	Module Universal Addressing
Guinée-Bissau	GW	GNB	Module Universal Addressing
Guyane	GY	GUY	Module Entreprise Geocoding (Amérique latine)Module Universal Addressing

Nom de pays ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Modules pris en charge
Haïti	HT	HTI	Module Universal Addressing
Îles Heard-et-MacDonald	HM	HMD	Module Universal Addressing
Saint-Siège (État de la cité du Vatican)	VA	TVA	Module Entreprise Geocoding ( <i>Le Vatican est couvert par le géocodeur de l'Italie</i> )Module Universal Addressing
Honduras	HN	HND	Module Entreprise Geocoding (Amérique latine)Module Universal Addressing
Hong Kong	HK	HKG	Module Entreprise GeocodingModule Universal Addressing
Hongrie	HU	HUN	Module Entreprise GeocodingModule Entreprise RoutingModule Universal Addressing
Islande	IS	ISL	Module Entreprise GeocodingModule Universal Addressing
Inde	IN	IND	Module Entreprise GeocodingModule Universal Addressing
Indonésie	ID	IDN	Module Entreprise GeocodingModule Universal Addressing
Iran, République islamique d'	IR	IRN	Module Universal Addressing
Irak	IQ	IRQ	Module Entreprise Geocoding (Moyen-Orient)Module Universal Addressing

Nom de pays ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Modules pris en charge
Irlande	IE	IRL	Module Enterprise Geocoding Module Enterprise Routing Module Universal Addressing
île de Man	IM	IMN	Module Universal Addressing
Israël	IL	ISR	Module Enterprise Geocoding Module Universal Addressing Module Enterprise Routing
Italie	IT	ITA	Module Enterprise Geocoding Module Enterprise Routing Module Universal Addressing
Jamaïque	JM	JAM	Module Enterprise Geocoding (Amérique latine) Module Universal Addressing
Japon	JP	JPN	Module Enterprise Geocoding Module Enterprise Routing Module Universal Addressing
Jersey	JE	JEY	Module Universal Addressing
Jordanie	JO	JOR	Module Universal Addressing Module Enterprise Geocoding (Moyen-Orient) Module Enterprise Routing
Kazakhstan	KZ	KAZ	Module Universal Addressing
Kenya	KE	KEN	Module Enterprise Geocoding (Afrique) Module Universal Addressing Module Enterprise Routing
Kiribati	KI	KIR	Module Universal Addressing

Nom de pays ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Modules pris en charge
Corée, République populaire démocratique de	KP	PRK	Module Universal Addressing
Corée, République de	KR	KOR	Module Enterprise Geocoding Module Universal Addressing
Kosovo	Xk	XKX	Module Enterprise Geocoding Module Universal Addressing
Koweït	KW	KWT	Module Enterprise Geocoding (Moyen-Orient) Module Universal Addressing
Kirghizistan	KG	KGZ	Module Universal Addressing
République démocratique populaire Lao	LA	LAO	Module Universal Addressing
Lettonie	LV	LVA	Module Enterprise Geocoding Module Enterprise Routing Module Universal Addressing
Liban	LB	LBN	Module Enterprise Geocoding (Moyen-Orient) Module Universal Addressing
Lesotho	LS	LSO	Module Enterprise Geocoding (Afrique) Module Universal Addressing Module Enterprise Routing
Libéria	LR	LBR	Module Universal Addressing
Libyenne, Jamahiriya arabe	LY	LBY	Module Universal Addressing

Nom de pays ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Modules pris en charge
Liechtenstein	LI	LIE	Module Entreprise Geocoding ( <i>Le Liechtenstein est couvert par le géocodeur de la Suisse</i> ).Module Entreprise RoutingModule Universal Addressing
Lituanie	LT	LTU	Module Entreprise GeocodingModule Entreprise RoutingModule Universal Addressing
Luxembourg	LU	LUX	Module Entreprise Geocoding ( <i>Le Luxembourg est couvert par le géocodeur de la Belgique</i> ).Module Entreprise RoutingModule Universal Addressing
Macao	MO	MAC	Module Entreprise GeocodingModule Universal Addressing
Macédoine, ancienne République yougoslave de	MK	MKD	Module Entreprise GeocodingModule Universal Addressing
Madagascar	MG	MDG	Module Universal Addressing
Malawi	MW	MWI	Module Entreprise Geocoding (Afrique)Module Universal Addressing
Malaisie	MY	MYS	Module Entreprise GeocodingModule Entreprise RoutingModule Universal Addressing
Maldives	MV	MDV	Module Universal Addressing
Mali	ML	MLI	Module Entreprise Geocoding (Afrique)Module Universal Addressing

Nom de pays ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Modules pris en charge
Malte	ML	MLT	Module Entreprise Geocoding Module Universal Addressing
Îles Marshall	MH	MHL	Module Universal Addressing
Martinique	MQ	MTQ	Module Entreprise Geocoding ( <i>La Martinique est couverte par le géocodeur de la France</i> ).Module Universal Addressing
Mauritanie	MR	MRT	Module Entreprise Geocoding (Afrique)Module Universal Addressing
Maurice	MU	MUS	Module Entreprise Geocoding (Afrique)Module Universal Addressing
Mayotte	YT	MYT	Module Entreprise Geocoding ( <i>Mayotte est couverte par le géocodeur de la France</i> )Module Universal Addressing
Mexique	MX	MEX	Module Entreprise GeocodingModule Universal Addressing
Micronésie, États fédérés de	FM	FSM	Module Universal Addressing
Moldova, République de	MD	MDA	Module Universal AddressingModule Entreprise Routing
Monaco	MC	MCO	Module Entreprise Geocoding ( <i>Monaco est couverte par le géocodeur de la France</i> ).Module Universal Addressing

Nom de pays ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Modules pris en charge
Mongolie	MN	MNG	Module Universal Addressing
Monténégro	ME	MNE	Module Enterprise Geocoding Module Universal Addressing
Montserrat	MS	MSR	Module Universal Addressing
Maroc	MA	MAR	Module Enterprise Geocoding (Afrique) Module Universal Addressing
Mozambique	MZ	MOZ	Module Enterprise Geocoding (Afrique) Module Universal Addressing Module Enterprise Routing
Myanmar	MM	MMR	Module Universal Addressing
Namibie	NA	NAM	Module Enterprise Geocoding (Afrique) Module Universal Addressing
Nauru	NR	NRU	Module Universal Addressing
Népal	NP	NPL	Module Universal Addressing
Pays-Bas	NL	NLD	Module Enterprise Geocoding Module Enterprise Routing Module Universal Addressing
Nouvelle-Calédonie	NC	NCL	Module Universal Addressing

Nom de pays ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Modules pris en charge
Nouvelle-Zélande	NZ	NZL	Module Enterprise Geocoding Module Enterprise Routing Module Universal Addressing
Nicaragua	NI	NIC	Module Enterprise Geocoding (Amérique latine) Module Universal Addressing
Niger	NE	NER	Module Enterprise Geocoding (Afrique) Module Universal Addressing
Nigéria	NG	NGA	Module Enterprise Geocoding (Afrique) Module Universal Addressing
Niué	NU	NIU	Module Universal Addressing
Île Norfolk	NF	NFK	Module Universal Addressing
Îles Mariannes du Nord	MP	MNP	Module Universal Addressing
Norvège	NO	NOR	Module Enterprise Geocoding Module Enterprise Routing Module Universal Addressing
Oman	OM	OMN	Module Enterprise Geocoding (Moyen-Orient) Module Universal Addressing
Pakistan	PK	PAK	Module Universal Addressing
Palaos	PW	PLW	Module Universal Addressing



Nom de pays ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Modules pris en charge
Palestine, territoire occupé	PS	PSE	Module Universal Addressing
Panama	PA	PAN	Module Enterprise Geocoding (Amérique latine) Module Universal Addressing
Papouasie - Nouvelle-Guinée	PG	PNG	Module Universal Addressing
Paraguay	PY	PRY	Module Enterprise Geocoding (Amérique latine) Module Universal Addressing
Pérou	PE	PER	Module Enterprise Geocoding (Amérique latine) Module Universal Addressing
Philippines	PH	PHL	Module Enterprise Geocoding Module Universal Addressing Module Enterprise Routing
Pitcairn	PN	PCN	Module Universal Addressing
Pologne	PL	POL	Module Enterprise Geocoding Module Enterprise Routing Module Universal Addressing
Portugal	PT	PRT	Module Enterprise Geocoding Module Enterprise Routing Module Universal Addressing
Porto Rico	PR	PRI	Module Universal Addressing
Qatar	QA	QAT	Module Enterprise Geocoding (Moyen-Orient) Module Universal Addressing

Nom de pays ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Modules pris en charge
Réunion	RE	REU	Module Entreprise Geocoding ( <i>La Réunion est couverte par le géocodeur de la France</i> ).Module Universal Addressing
Roumanie	RO	ROU	Module Entreprise GeocodingModule Entreprise RoutingModule Universal Addressing
Russie, Fédération de	RU	RUS	Module Entreprise GeocodingModule Universal Addressing
Rwanda	RW	RWA	Module Entreprise Geocoding (Afrique)Module Universal Addressing
Saint-Barthélemy	BL	BLM	Module Universal Addressing
Sainte-Hélène, Ascension et Tristan da Cunha	SH	SHE	Module Universal Addressing
Saint-Christophe-et-Niévès	KN	KNA	Module Entreprise Geocoding (Amérique latine)Module Universal Addressing
Sainte Lucie	LC	LCA	Module Universal Addressing
Saint-Martin (partie française)	MF	MAF	Module Universal Addressing
Saint-Pierre-et-Miquelon	PM	SPM	Module Universal Addressing
Saint-Vincent-et-les-Grenadines	VC	VCT	Module Universal Addressing

Nom de pays ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Modules pris en charge
Samoa	WS	WSM	Module Universal Addressing
Saint-Marin	SM	SMR	Module Entreprise Geocoding ( <i>Saint-Marin est couvert par le géocodeur de l'Italie</i> ).Module Universal Addressing
Sao Tomé-et-Principe	ST	STP	Module Universal Addressing
Arabie saoudite	SA	SAU	Module Entreprise Geocoding (Moyen-Orient)Module Universal Addressing
Sénégal	SN	SEN	Module Entreprise Geocoding (Afrique)Module Universal Addressing
Serbie	RS	SRB	Module Entreprise GeocodingModule Universal Addressing
Seychelles	SC	SYC	Module Universal Addressing
Sierra Leone	SL	SLE	Module Universal Addressing
Singapour	SG	SGP	Module Entreprise GeocodingModule Entreprise RoutingModule Universal Addressing
Saint-Martin (partie hollandaise)	SX	SXM	Module Universal Addressing
Slovaquie	SK	SVK	Module Entreprise GeocodingModule Entreprise RoutingModule Universal Addressing

Nom de pays ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Modules pris en charge
Slovénie	SI	SVN	Module Enterprise Geocoding Module Enterprise Routing Module Universal Addressing
Îles Salomon	SB	SLB	Module Universal Addressing
Somalie	SO	SOM	Module Universal Addressing
Afrique du Sud	ZA	ZAF	Module Enterprise Geocoding Module Universal Addressing
Géorgie du Sud et les îles Sandwich du Sud	GS	SGS	Module Enterprise Geocoding Module Universal Addressing
Soudan du Sud	SS	SSD	Module Universal Addressing
Espagne	ES	ESP	Module Enterprise Geocoding Module Enterprise Routing Module Universal Addressing
Sri Lanka	LK	LKA	Module Universal Addressing
Soudan	SD	SDN	Module Universal Addressing
Surinam	SR	SUR	Module Enterprise Geocoding (Amérique latine) Module Universal Addressing
Svalbard et Jan Mayen	SJ	SJM	Module Universal Addressing
Swaziland	SZ	SWZ	Module Enterprise Geocoding (Afrique) Module Universal Addressing

Nom de pays ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Modules pris en charge
Suède	SE	SWE	Module Enterprise Geocoding Module Enterprise Routing Module Universal Addressing
Suisse	CH	CHE	Module Enterprise Geocoding Module Enterprise Routing Module Universal Addressing
République arabe syrienne	SY	SYR	Module Universal Addressing
Taiwan, province de Chine	TW ou zh_TW (Routage)	TWN	Module Universal Addressing Module Enterprise Routing
Tadjikistan	TJ	TJK	Module Universal Addressing
Tanzanie, République unie de	TZ	TZA	Module Enterprise Geocoding (Afrique) Module Universal Addressing Module Enterprise Routing
Thaïlande	th	THA	Module Enterprise Geocoding Module Enterprise Routing Module Universal Addressing
Timor-Leste	TL	TLS	Module Universal Addressing
Togo	TG	TGO	Module Enterprise Geocoding (Afrique) Module Universal Addressing
Tokelau	TK	TKL	Module Universal Addressing
Tonga	TO	TON	Module Universal Addressing

Nom de pays ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Modules pris en charge
Trinité-et-Tobago	TT	TTO	Module Enterprise Geocoding (Amérique latine)Module Universal Addressing
Tunisie	TN	TUN	Module Enterprise Geocoding (Afrique)Module Universal Addressing
Turquie	TR	TUR	Module Enterprise GeocodingModule Universal Addressing
Turkménistan	TM	TKM	Module Universal Addressing
Îles Turques-et-Caïques	TC	TCA	Module Universal Addressing
Tuvalu	TV	TUV	Module Universal Addressing
Ouganda	UG	UGA	Module Enterprise Geocoding (Afrique)Module Universal Addressing
Ukraine	UA	UKR	Module Enterprise GeocodingModule Universal Addressing
Émirats Arabes Unis	AE	ARE	Module Enterprise Geocoding (Moyen-Orient)Module Universal Addressing
Royaume-Uni	GB	GBR	Module Enterprise GeocodingModule Enterprise RoutingModule Universal Addressing
États-Unis	US	États-Unis	Module Enterprise GeocodingModule Enterprise RoutingModule Universal Addressing

Nom de pays ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Modules pris en charge
Îles mineures éloignées des États-Unis	UM	UMI	Module Universal Addressing
Uruguay	UY	URY	Module Enterprise Geocoding Module Universal Addressing
Ouzbékistan	UZ	UZB	Module Universal Addressing
Vanuatu	UV	VUT	Module Universal Addressing
Venezuela (République bolivarienne du)	VE	VEN	Module Enterprise Geocoding Module Universal Addressing
Vietnam	VN	VNM	Module Enterprise Geocoding Module Universal Addressing
Îles Vierges britanniques	VG	VGB	Module Universal Addressing
Îles Vierges des États-Unis	VI	VIR	Module Universal Addressing
Wallis-et-Futuna	WF	WLF	Module Universal Addressing
Sahara occidental	EH	ESH	Module Universal Addressing
Yémen	YE	YEM	Module Enterprise Geocoding (Moyen-Orient) Module Universal Addressing
Zambie	ZM	ZMB	Module Enterprise Geocoding (Afrique) Module Universal Addressing

Nom de pays ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Modules pris en charge
Zimbabwe	ZW	ZWE	Module Enterprise Geocoding (Afrique)Module Universal Addressing



# Notices

© 2018 Pitney Bowes. Tous droits réservés. MapInfo et Group 1 Software sont des marques commerciales de Pitney Bowes Software Inc. Toutes les autres marques et marques commerciales sont la propriété de leurs détenteurs respectifs.

### *Avis USPS®*

Pitney Bowes Inc. détient une licence non exclusive pour la publication et la vente de bases de données ZIP + 4® sur des supports optiques et magnétiques. Les marques de commerce suivantes appartiennent à United States Postal Service : CASS, CASS Certified, DPV, eLOT, FASTforward, First-Class Mail, Intelligent Mail, LACS<sup>Link</sup>, NCOA<sup>Link</sup>, PAVE, PLANET Code, Postal Service, POSTNET, Post Office, RDI, Suite<sup>Link</sup>, United States Postal Service, Standard Mail, United States Post Office, USPS, ZIP Code et ZIP + 4. Cette liste de marques de commerce appartenant à U.S. Postal Service n'est pas exhaustive.

Pitney Bowes Inc. détient une licence non exclusive de USPS® pour le traitement NCOA<sup>Link®</sup>.

Les prix des produits, des options et des services de Pitney Bowes Software ne sont pas établis, contrôlés ni approuvés par USPS® ni par le gouvernement des États-Unis. Lors de l'utilisation de données RDI™ pour déterminer les frais d'expédition de colis, le choix commercial de l'entreprise de distribution de colis à utiliser n'est pas fait par USPS® ni par le gouvernement des États-Unis.

### *Fournisseur de données et avis associés*

Les produits de données contenus sur ce support et utilisés au sein des applications Pitney Bowes Software sont protégés par différentes marques de commerce et par un ou plusieurs des copyrights suivants :

© Copyright United States Postal Service. Tous droits réservés.

© 2014 TomTom. Tous droits réservés. TomTom et le logo TomTom logo sont des marques déposées de TomTom N.V.

© 2016 HERE

Source : INEGI (Instituto Nacional de Estadística y Geografía)

Basées sur les données électroniques © National Land Survey Sweden.

© Copyright United States Census Bureau

© Copyright Nova Marketing Group, Inc.

Des portions de ce programme sont sous © Copyright 1993-2007 de Nova Marketing Group Inc. Tous droits réservés.

© Copyright Second Decimal, LLC

© Copyright Canada Post Corporation

Ce CD-ROM contient des données provenant d'une compilation dont Canada Post Corporation possède le copyright.

© 2007 Claritas, Inc.

Le jeu de données Geocode Address World contient des données distribuées sous licence de GeoNames Project ([geonames.org](http://geonames.org)) fournies sous la licence Creative Commons Attribution License (« Attribution License ») à l'adresse : <http://creativecommons.org/licenses/by/3.0/legalcode>. Votre utilisation des données GeoNames (décrites dans le Manuel de l'utilisateur Spectrum™ Technology Platform) est régie par les conditions de la licence Attribution License et tout conflit entre votre accord avec Pitney Bowes Software, Inc. et la licence Attribution License sera résolu en faveur de la licence Attribution License uniquement s'il concerne votre utilisation des données GeoNames.



3001 Summer Street  
Stamford CT 06926-0700  
USA

[www.pitneybowes.com](http://www.pitneybowes.com)