

# Spectrum™ Technology Platform

Version 12.0 SP2

Guide Enterprise Data Integration



# Table des matières

## 1 - Introduction

---

Architecture de gestion des données d'entreprise	5
Modèle d'entrepôt de données basé sur un schéma en étoile	8

## 2 - Connexion aux sources de données et aux entrepôts de données

---

Data Source Connections	14
Définition de connexions	14
Prise en charge de la compression pour les serveurs de fichiers Cloud	68
Suppression d'une connexion	69

## 3 - Renseignement de l'entrepôt de données

---

Préparation de vos données	71
Renseignement d'une table de dimensions temporelles	72
Renseignement d'une table de dimensions	73
Renseignement d'une table de faits	75
Ajout d'un horodateur aux enregistrements d'un entrepôt de données	80

## 4 - Mise à jour de l'entrepôt de données

---

Définition d'un calendrier de mise à jour d'un entrepôt de données	83
Mise à jour d'une table de faits	84
Utilisation d'un cache global pour les requêtes	89
Utilisation d'un cache local pour les requêtes	90

## 5 - Référence aux stages

---

Call Stored Procedure	94
DB Change Data Reader	96
DB Loader	100
Field Parser	109
Field Combiner	113
Field Selector	114
Generate Time Dimension	115
Query Cache	123
Query DB	132
Requête NoSQL DB	136
Read From DB	139
Read from file	148
Read from Hadoop Sequence File	165
Read from Hive File	169
Read from HL7 File	173
Read from NoSQL DB	185
Read from SAP	189
Read from Spreadsheet	194
Read from Variable Format File	197
Read From XML	212
Commande SQL	219
Transpose	230
Unique ID Generator	234
Write To Cache	243

Write to DB	245
Write to File	251
Write to Hadoop Sequence File	269
Write to Hive File	272
Write to NoSQL DB	279
Write to Spreadsheet	283
Write to Variable Format File	286
Write to XML	296
Modèles de date et de nombre	306

## 6 - Configurations

---

Configurations de LogMiner Oracle	313
-----------------------------------	-----

## 7 - Optimisation des performances

---

Définition d'une taille d'extraction optimale	316
---	-----

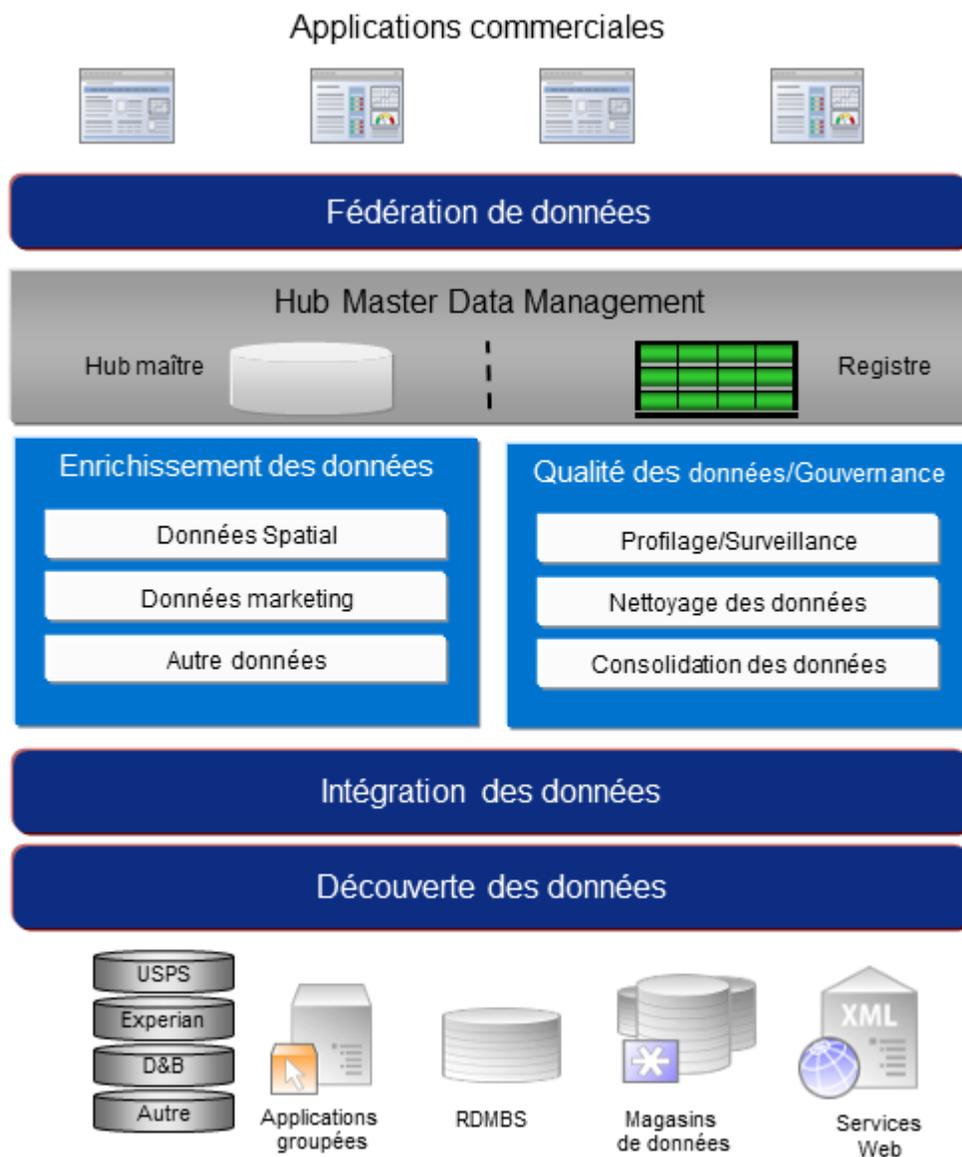
# 1 - Introduction

## In this section

Architecture de gestion des données d'entreprise	5
Modèle d'entrepôt de données basé sur un schéma en étoile	8

# Architecture de gestion des données d'entreprise

Avec Spectrum™ Technology Platform, vous pouvez générer une procédure complète de gestion des données d'entreprise, ou vous pouvez l'utiliser comme une solution plus ciblée. Le schéma suivant illustre une solution complète qui extrait les données d'une source, les fiabilise et les enrichit, et qui alimente un hub de gestion de données de référence qui offre une vue unique des données pour différentes applications métier.



## Data Discovery

Data Discovery est le processus de balayage de vos ressources de données visant à obtenir un inventaire complet de votre paysage de données. Spectrum™ Technology Platform peut balayer les données structurées, non structurées et semi-structurées à l'aide d'une grande variété de techniques de profilage de données. Les résultats du balayage sont utilisés pour générer automatiquement une bibliothèque de documentation décrivant les actifs de données de votre entreprise et pour créer un référentiel de métadonnées. Cette documentation et le référentiel de métadonnées associé vous fournissent les informations dont vous avez besoin avant de commencer des projets d'intégration de données, de gestion de la qualité des données, de gouvernance de données ou de gestion de données de référence.

Pour plus d'informations sur le module Data Discovery Spectrum™ Technology Platform, contactez votre responsable de compte.

## Data Integration

Une fois que vous disposez d'un inventaire de votre paysage de données, vous devez envisager le mode d'accès aux données que vous devez gérer. Spectrum™ Technology Platform peut se connecter à vos données via plusieurs sources, soit directement, soit par l'intermédiaire d'une intégration à vos technologies d'accès aux données existantes. Il prend en charge des fonctionnalités d'intégration de données par lots et en temps réel pour une grande variété de besoins d'entreprise, notamment l'entreposage de données, la gestion de la qualité des données, l'intégration système et la migration. Spectrum™ Technology Platform peut accéder aux données de bases de données RDBMS, de magasins de données, de fichiers XML, de fichiers plats, etc. Spectrum™ Technology Platform prend en charge les requêtes SQL avec des agrégations et des jointures complexes et constitue un outil de développement de requêtes visuel. De plus, Spectrum™ Technology Platform peut accéder aux données via les services Web REST et SOAP.

Spectrum™ Technology Platform peut déclencher un traitement par lots en fonction de l'apparence d'un ou de plusieurs fichiers sources dans un dossier spécifié. Ce déclenchement de « dossier à chaud » permet de surveiller les chargements FTP et de les traiter à mesure qu'ils se produisent.

Certaines de ces fonctions d'intégration de données nécessitent une licence pour le module Enterprise Data Integration. Pour plus d'informations, contactez votre responsable de compte.

Enfin, Spectrum™ Technology Platform peut s'intégrer à des applications groupées, telles que SAP.

## Qualité des données/gouvernance

Les procédures de gestion de la qualité et de gouvernance des données recherchent les enregistrements en double, les informations incohérentes et les informations inexacts dans vos données.

Le rapprochement de doublons identifie les enregistrements doublons potentiels ou les relations entre les enregistrements, si les données sont de type nom et adresse, ou appartiennent à un autre type d'informations clients. Spectrum™ Technology Platform vous permet d'indiquer un ensemble cohérent de règles de correspondance pour votre entreprise en utilisant des méthodes de correspondance booléennes, des méthodes de calcul du score, des seuils, des algorithmes et des pondérations afin de déterminer si un groupe d'enregistrements contient des doublons. Spectrum™

Technology Platform prend en charge une personnalisation étendue pour vous permettre d'ajuster ces règles aux besoins uniques de votre entreprise.

Une fois les enregistrements doublons identifiés, vous pouvez souhaiter consolider les enregistrements. Spectrum™ Technology Platform vous permet d'indiquer la manière de lier ou de fusionner les enregistrements doublons pour pouvoir créer l'enregistrement le plus complet et précis à partir de toute collecte d'informations client. Par exemple, un seul enregistrement de type best-of-breed (premier de groupe) peut être généré à partir de tous les enregistrements d'un même foyer. Le module Advanced Matching permet d'identifier les doublons et de les éliminer.

Les procédures de gestion de qualité des données normalisent également vos données. La normalisation est une procédure critique car les éléments de données normalisées sont requis pour atteindre les meilleurs résultats possibles pour mettre en correspondance et identifier les relations entre les enregistrements. Tandis que plusieurs modules réalisent une normalisation d'un type ou d'un autre, c'est le module Data Normalization de la Spectrum™ Technology Platform qui fournit l'ensemble des fonctions de normalisation le plus complet. En outre, le module Universal Name fournit des fonctions de fiabilisation des données de noms personnel et de noms d'entreprises.

Des données normalisées ne sont pas forcément des données précises. Spectrum™ Technology Platform peut comparer vos données aux données de référence connues et à jour à des fins de correction. Les sources utilisées pour cette procédure peuvent inclure des organismes de réglementation (par exemple, l'U.S. Postal Service), des fournisseurs de données tiers (par exemple, Experian ou D&B) ou les sources de références internes d'une entreprise (par exemple, des données comptables). Spectrum™ Technology Platform est particulièrement efficace pour la validation de données d'adresse. Il peut valider ou normaliser les adresses de 250 pays et territoires dans le monde. Il existe deux modules de validation des adresses : le module Address Now et le module Universal Addressing.

Pour déterminer celui qui vous convient, parlez de vos besoins avec votre responsable de compte.

Alors que Spectrum™ Technology Platform peut gérer automatiquement un large éventail de problèmes de qualité des données, dans certains cas, une révision manuelle par un data steward est souhaitable. Pour cela, le module Business Steward permet d'indiquer les règles qui déclencheront une révision manuelle et fournit un outil Web pour la révision des enregistrements d'exception. Cela inclut l'accès intégré aux outils tiers, tels que les cartes Bing et les données Experian qui aident les data stewards dans la procédure de révision et de résolution.

### *Enrichissement des données*

Les procédures d'enrichissement des données améliorent vos données avec des informations supplémentaires. L'enrichissement peut être basé sur des données spatiales, des données marketing ou des données d'autres sources à utiliser pour ajouter des détails supplémentaires à vos données. Par exemple, si vous disposez d'une base de données d'adresses de clients, vous pouvez géocoder l'adresse pour déterminer les coordonnées de latitude/longitude de l'adresse et stocker ces coordonnées dans l'enregistrement. Vos données client peuvent ensuite être utilisées pour réaliser divers calculs spatiaux, par exemple pour rechercher l'agence bancaire la plus proche du client. Spectrum™ Technology Platform vous permet d'enrichir vos données de différentes informations, comme le géocodage (à l'aide du module Enterprise Geocoding), l'affectation de juridictions fiscales (à l'aide du module Enterprise Tax), les calculs géospatiaux (à l'aide du module Location Intelligence)

et les directions d'itinéraire à pied et à bord d'un véhicule entre des points (à l'aide du module Enterprise Routing).

### *Master Data Management Hub*

Le hub de gestion de données maître (MDM) permet une modélisation rapide des entités et de leurs relations complexes entre les rôles, les procédures et les interactions. Il fournit des fonctions d'analyse de réseau social intégrées qui vous permettent de comprendre les influenceurs, de prévoir les désabonnements, de détecter les relations non évidentes et les modèles frauduleux, et de fournir des recommandations.

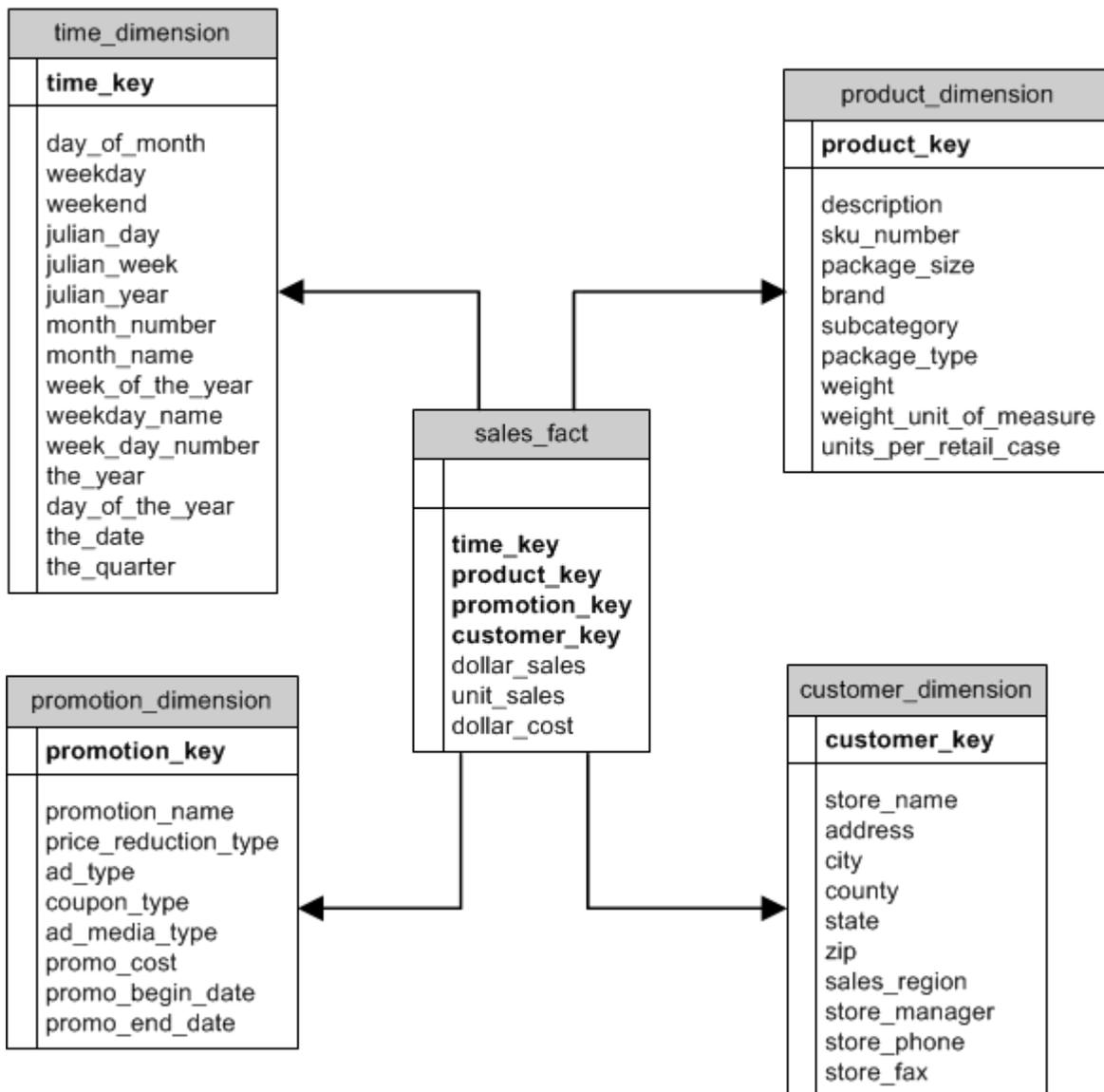
Spectrum™ Technology Platform prend en charge deux approches pour le hub MDM. Dans l'approche de hub maître, les données sont gérées dans une base de données MDM unique et les applications accèdent aux données à partir de la base de données MDM. Dans l'approche registre, les données sont gérées dans chaque application métier et le registre du hub MDM contient les clés permettant de trouver les enregistrements associés. Par exemple, un enregistrement de client peut exister dans une base de données de saisie de commande et une base de données de support clientèle. Le registre MDM contient une clé unique qui permettrait d'accéder aux données client dans les deux emplacements.

Le module Data Hub fournit des fonctions MDM.

## Modèle d'entrepôt de données basé sur un schéma en étoile

Spectrum™ Technology Platform prend en charge la création et la maintenance d'entrepôts de données utilisant un modèle de schéma en étoile. Dans un schéma en étoile, les données sont stockées soit sous forme de faits, à savoir, des descriptions spécifiques d'un événement, soit sous forme d'attributs dimensionnels, à savoir, des descriptions des faits de la table de faits. Les faits changent régulièrement, tandis que les dimensions changent lentement, voire ne changent jamais.

L'illustration suivante montre le modèle d'un schéma en étoile :



Cette illustration indique les principales caractéristiques d'un schéma en étoile : une table de faits, des tables de dimensions et des jointures.

### Table de faits

Les tables de faits sont les tables centrales du schéma en étoile de votre entrepôt de données. Les tables de faits contiennent généralement des informations numériques ou quantitatives (appelées mesures) qui décrivent un événement spécifique. Par exemple, si vous disposez d'un entrepôt de données que vous utilisez pour générer un rapport sur le chiffre d'affaires de l'entreprise, votre table de faits contient les colonnes **dollar\_sales** et **dollar\_cost**, comme illustré ci-dessus. En règle générale, les faits sont continuellement évalués et additifs. Un fait « continuellement évalué » signifie qu'il s'agit d'une mesure numérique qui comporte une valeur chaque fois qu'elle est mesurée. Un fait « additif » indique qu'il est résumé par addition.

Les tables de faits contiennent également un ensemble de colonnes qui forment une clé concaténée, ou composite. Chaque colonne de la clé concaténée est constituée d'une clé étrangère tirée de la clé primaire d'une table de dimensions. Par exemple, dans l'illustration ci-dessus, la table de faits contient un colonne `product_key` qui associe le fait à un produit donné de la table `product_dimension`.

Le niveau de détail d'une table de faits est dit grain. Chaque ligne de la table de faits doit être enregistrée au même niveau de détail. Dans le diagramme ci-dessus, les mesures de la table de faits sont des totaux quotidiens de ventes en dollars, de ventes en unités et de coûts en dollars pour chaque produit vendu. Le grain est quotidien. Chaque enregistrement de la table de faits représente les ventes totales d'un produit donné dans un magasin en un jour. Chaque nouvelle combinaison de produit, de magasin ou de jour génère un enregistrement différent dans la table de faits.

Les tables de faits sont renseignées à l'aide de données provenant d'une source de données. La source de données peut être un système OLTP ou un entrepôt de données. Spectrum™ Technology Platform réalise une capture d'écran des données sources suivant un calendrier régulier et transfère les données à l'entrepôt de données, généralement à la même heure chaque jour, chaque semaine ou chaque mois.

Un schéma en étoile peut comporter différentes tables de faits. Utilisez un schéma comportant différentes tables de faits pour séparer les ensembles de mesures qui partagent un sous-ensemble commun de tables de dimensions, ou pour suivre des mesures présentant des grains différents.

### Table de dimensions

Les tables de dimensions stockent des données qui décrivent les informations de la table de faits. Par exemple, si `sales_total` varie d'un mois à l'autre, vous pouvez analyser les dimensions pour savoir pourquoi. Une même table de dimensions peut être utilisée avec différentes tables de faits.

Les tables de dimensions comportent des attributs et une clé primaire en une seule partie qui joint la table de dimensions à la table de faits. Les attributs constituent les colonnes de la table de dimensions. La clé primaire en une seule partie vous permet de parcourir rapidement une seule table de dimensions. Le parcours d'une table de dimensions peut vous aider à déterminer la meilleure façon de lancer une requête sur la table de faits.

Les tables de dimensions temporelles sont nécessaires pour réaliser des calculs temporels précis, car il s'avère parfois difficile d'extraire les données de date requises des enregistrements. Par exemple, les enregistrements suivants figurent dans une base de données de ventes. Notez qu'il existe des laps de temps entre les enregistrements. Par exemple, il n'existe aucun enregistrement pour la journée du 04/01/2012.

Date	Produit	Montant
03/01/2012	Chemise rouge	10,00 €
05/01/2012	Chemise rouge	5,00 €
07/01/2012	Chemise rouge	15,00 €

Si vous lancez une requête sur ces enregistrements et calculez les ventes moyennes par jour, la réponse est 10,00 € (30 €/3 enregistrements). Cependant, ce résultat est incorrect, car les trois enregistrements se sont en fait déroulés sur une période de cinq jours. Si vous disposez d'une table de dimensions temporelles comportant un enregistrement pour chaque jour, vous pouvez joindre cette table à la table ci-dessus pour obtenir le résultat suivant :

Date	Produit	Montant
03/01/2012	Chemise rouge	10,00 €
04/01/2012		
05/01/2012	Chemise rouge	5,00 €
06/01/2012		
07/01/2012	Chemise rouge	15,00 €

Si vous calculez les ventes moyennes par jour en fonction de ces enregistrements, vous obtenez la réponse correcte suivante : 6,00 € (30 €/5 jours).

Vous pouvez également prendre en compte dans vos calculs les attributs temporels arbitraires tels que les jours fériés, les weekends et les trimestres. Par exemple, si la date du 06/01/2012 est un jour férié et que seules les ventes moyennes par jour ouvré vous intéressent, la réponse est 7,50 €.

### Jointures

Les jointures définissent les relations entre une table de faits et des tables de dimensions dans le schéma en étoile. La clé primaire de la table de dimensions constitue la clé étrangère de la table de faits. La table de faits doit contenir une valeur de clé primaire de chaque table de dimensions. La référence de la clé étrangère à la clé primaire constitue le mécanisme de vérification des valeurs entre les deux tables. Les relations de jointure de ce type garantissent l'intégrité référentielle d'un entrepôt de données. Pour assurer la validité des résultats de requête, il convient de maintenir l'intégrité référentielle.

Chaque enregistrement d'une table de dimensions peut décrire de nombreux enregistrements de la table de faits, faisant ainsi de la cardinalité de jointure des tables de dimensions aux tables de faits une cardinalité de type un-à-plusieurs.

Dans l'illustration ci-dessus, `product_key` est la clé primaire de la table `product_dimension` et la clé étrangère de la table `sales_fact`. Cette jointure représente la relation entre les produits de l'entreprise et ses ventes.

## Avantages d'un schéma en étoile

Un schéma intelligemment conçu vous permet de comprendre, de parcourir et d'analyser rapidement des jeux de données multidimensionnels volumineux. Les principaux avantages des schémas en étoile d'un environnement d'aide à la prise de décisions sont les suivants :

### *Performances des requêtes*

Les requêtes lancées sur une base de données basée sur un schéma en étoile sont plus rapides que celles lancées sur un système OLTP, car le schéma en étoile comporte moins de tables et des chemins d'accès joints clairs. Dans un modèle de schéma en étoile, les dimensions sont liées via la table de faits centrale. Les dimensions sont liées les unes aux autres via un chemin d'accès joint intersectant la table de faits. Cette fonctionnalité permet d'obtenir des résultats de requête précis et cohérents.

### *Performances et administration du chargement*

La structure de schéma en étoile réduit le temps nécessaire au chargement de lots de données volumineux dans une base de données. En définissant les faits et les dimensions et en les séparant dans des tables différentes, il est possible de réduire l'impact d'une opération de chargement. Les tables de dimensions sont renseignées une seule fois et actualisées de temps en temps. Les nouveaux faits peuvent être régulièrement ajoutés, de manière sélective, via l'ajout d'enregistrements à une table de faits.

### *Intégrité référentielle intégrée*

Un schéma en étoile est conçu pour garantir l'intégrité référentielle des données chargées. L'intégrité référentielle est assurée grâce à l'utilisation de clés primaires et étrangères. Les clés primaires des tables de dimensions deviennent des clés étrangères dans les tables de faits pour associer chaque enregistrement des tables de dimensions et de faits.

### *Parcours efficace des données*

Le parcours des données s'avère efficace, car les dimensions sont jointes via des tables de faits. Ces jointures sont importantes, car elles représentent les relations fondamentales des processus métier réels. Vous pouvez parcourir une seule table de dimensions, afin de sélectionner des valeurs d'attribut pour générer une requête efficace.

# 2 - Connexion aux sources de données et aux entrepôts de données

## In this section

---

Data Source Connections	14
Définition de connexions	14
Prise en charge de la compression pour les serveurs de fichiers Cloud	68
Suppression d'une connexion	69

## Data Source Connections

Une source de données est une base de données, un serveur de fichiers, un service Cloud ou une autre source de données que vous souhaitez traiter via Spectrum™ Technology Platform. Spectrum™ Technology Platform peut se connecter à plus de 20 types de sources de données.

Pour connecter Spectrum™ Technology Platform à une source de données, vous devez commencer par définir la connexion. Par exemple, si vous souhaitez lire les données d'un fichier XML dans un dataflow, et que le fichier XML est situé sur un serveur de fichiers distant, vous devez définir une connexion au serveur de fichiers avant de pouvoir définir le fichier XML d'entrée dans un dataflow. De même, pour écrire une sortie de dataflow dans une base de données, vous devez définir au préalable la base de données en tant que ressource externe.

## Définition de connexions

Pour définir une nouvelle connexion dans Spectrum™ Technology Platform, utilisez l'un de ces modules :

- Management Console
- Onglet Sources de données de Metadata Insights

**Remarque :** Pour lire ou écrire des données situées dans un fichier sur le serveur Spectrum™ Technology Platform lui-même, il n'est pas nécessaire de définir une connexion.

## Connexion à Amazon

### Connexion à Amazon DynamoDB

Vous pouvez utiliser cette connexion :

- Dans Enterprise Designer pour la lecture et l'écriture de données
- Dans Metadata Insights pour créer des Physical Models

1. Accédez à la page **Sources de données** en utilisant l'un de ces modules :

**Management Console :** Ouvrez Management Console à l'aide de l'URL : URL: `http://serveur:port/managementconsole`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Ressources > Sources de données**.

**Metadata Insights :** Ouvrez Metadata Insights à l'aide de l'URL : URL: `http://serveur.port/metadata-insights`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Sources de données**.

2. Cliquez sur le bouton **Ajouter** .
3. Dans le champ **Nom**, entrez un nom pour la nouvelle connexion. Ce nom est entièrement de votre choix.

**Remarque :** Une fois que vous enregistrez une connexion, vous ne pouvez plus modifier le nom.

4. Dans le champ **Type**, sélectionnez **Amazon DynamoDB**.
5. Dans le champ **ID de clé d'accès**, saisissez la séquence alphanumérique de 20 caractères qui vous a été fournie pour accéder à votre compte Amazon AWS.
6. Dans le champ **Clé d'accès secrète**, saisissez la clé de 40 caractères nécessaire pour authentifier la connexion.
7. Dans le champ **Région** sélectionnez la région du compte Amazon AWS.
8. Pour tester la connexion, cliquez sur **Tester**.
9. Cliquez sur **Enregistrer**.

#### Limitations Amazon DynamoDB

1. Les types de données hiérarchiques comme les listes, les cartes et les ensembles sont interprétés comme des types de données Chaîne. Cela est dû au fait que ces types de données ne sont pas pris en charge.
2. Les valeurs Null d'une source de données DynamoDB sont interprétées comme des valeurs de colonne vides.
3. La fonction d'agrégation `count` n'est pas prise en charge dans une requête sur un Model Store.

#### Connexion à Amazon S3

Vous pouvez utiliser cette connexion dans les job de qualité des données Hadoop MapReduce et Spark sur les Big Data.

1. Accédez à la page **Sources de données** en utilisant l'un de ces modules :

**Management Console :** Ouvrez Management Console à l'aide de l'URL : URL: `http://serveur.port/managementconsole`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Ressources > Sources de données**.

**Metadata Insights :** Ouvrez Metadata Insights à l'aide de l'URL : URL: `http://serveur.port/metadata-insights`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Sources de données**.

2. Cliquez sur le bouton **Ajouter** .
3. Dans le champ **Nom**, entrez un nom pour la nouvelle connexion. Ce nom est entièrement de votre choix.

**Remarque :** Une fois que vous enregistrez une connexion, vous ne pouvez plus modifier le nom.

4. Dans le champ **Type**, sélectionnez **Cloud**.
5. Dans le champ **Service Cloud**, sélectionnez **AmazonS3**.
6. Dans le champ **Nom de compartiment**, saisissez le nom de compartiment tel que défini dans votre service Cloud Amazon S3. Il s'agit du compartiment où Spectrum™ Technology Platform lira et écrira des fichiers.
7. Saisissez votre clé d'accès et la clé secrète qui vous sont affectées par Amazon.
8. Dans le champ **Type de stockage**, sélectionnez le niveau de redondance que vous souhaitez autoriser pour le stockage des données.

**Standard** Niveau de redondance par défaut fourni par Amazon S3.

**Redondance réduite** Stocke les données non critiques et facilement reproduisibles à des niveaux de redondance inférieurs. Cela fournit un stockage assez fiable à moindre coût.

9. Dans la section **Chiffrement**, sélectionnez la méthode de chiffrement des données. Vous pouvez sélectionner le chiffrement côté serveur et/ou le chiffrement côté client.

**Clé côté serveur** Les données sont chiffrées et déchiffrées côté serveur. Vos données sont transmises sous forme de texte en clair au service Cloud Amazon, où elles sont chiffrées et stockées. Lors de leur récupération, les données sont déchiffrées

par le service Cloud Amazon, puis transmises sous forme de texte en clair à votre système.

Vous disposez de deux options permettant de spécifier la clé :

- **Géré par AWS** : la clé est automatiquement générée par le service Cloud Amazon S3.
- **Fourni par le client** : saisissez la clé à utiliser par le service Cloud Amazon S3 pour chiffrer et déchiffrer les données côté serveur.

#### **Clé côté client**

Les données sont chiffrées et déchiffrées côté client. Les données sont chiffrées localement sur votre système client, puis transmises au stockage Cloud Amazon S3. Lors de leur récupération, les données sont retransmises à votre système au format chiffré et déchiffrées sur le système client.

**Clé côté client** : saisissez la clé à utiliser par votre système client pour chiffrer et déchiffrer les données.

Si vous sélectionnez à la fois **Clé côté serveur** et **Clé côté client**, le chiffrement et le déchiffrement sont effectués côté serveur et côté client. Les données sont tout d'abord chiffrées à l'aide de votre clé côté client, puis transmises au format chiffré à Amazon, où elles sont de nouveau chiffrées à l'aide de la clé côté serveur et stockées. Lors de leur récupération, Amazon commence par déchiffrer les données à l'aide de la clé côté serveur, puis transmet les données au format chiffré à votre système, où elles sont enfin déchiffrées à l'aide de la clé côté client.

**Remarque** : Pour utiliser la fonctionnalité de chiffrement du Cloud Amazon S3, vous devez installer les fichiers JAR Amazon S3 Security. Pour plus d'informations, reportez-vous à la section [Utilisation du chiffrement Cloud Amazon S3](#) à la page 18.

Pour plus d'informations sur les fonctionnalités de chiffrement d'Amazon S3, reportez-vous à l'adresse suivante :

[docs.aws.amazon.com/AmazonS3/latest/dev/UsingEncryption.html](https://docs.aws.amazon.com/AmazonS3/latest/dev/UsingEncryption.html)

10. Si vous souhaitez définir des droits d'accès; dans la section **Droits d'accès**, cliquez sur .

Les trois types de bénéficiaires sont les suivants :

<b>Everyone</b>	Tous les utilisateurs, à l'exception des AuthenticatedUsers et du groupe LogDelivery.
<b>AuthenticatedUsers</b>	Pour les utilisateurs qui sont connectés à Amazon.
<b>LogDelivery</b>	Pour les utilisateurs qui écrivent des journaux d'activités dans un compartiment spécifié par l'utilisateur, si la journalisation de compartiment est activée.

Pour chaque bénéficiaire, sélectionnez les autorisations souhaitées :

<b>Open/Download</b>	L'utilisateur peut télécharger le fichier.
<b>Vue</b>	L'utilisateur peut afficher les autorisations existantes sur le fichier.

**Modifier**

L'utilisateur peut modifier et définir les autorisations sur le fichier.

11. Pour tester la connexion, cliquez sur **Tester**.
12. Cliquez sur **Enregistrer**.

**Utilisation du chiffrement Cloud Amazon S3**

Pour utiliser la fonctionnalité de sécurité par chiffrement du service Cloud Amazon S3, vous devez télécharger les fichiers JAR de sécurité et les placer sur le serveur Spectrum™ Technology Platform. L'utilisation du chiffrement est facultatif.

1. Accédez au site de téléchargement.

Pour les plates-formes Windows et Linux utilisant Java7, les fichiers JAR peuvent être téléchargés depuis :

<http://www.oracle.com/technetwork/java/javase/downloads/jce-7-download-432124.html>

Pour les plates-formes AIX utilisant Java7, les fichiers JAR peuvent être téléchargés depuis :

<https://www14.software.ibm.com/webapp/iwm/web/preLogin.do?source=jcesdk>

2. Téléchargez ces deux fichiers JAR :

- local\_policy.jar
- US\_export\_policy.jar

3. Placez ces fichiers JAR à l'emplacement suivant :

```
SpectrumFolder\Pitney Bowes\Spectrum\java64\jre\lib\security
```

4. Redémarrez le serveur.

**Connexion à Amazon SimpleDB**

Vous pouvez utiliser cette connexion :

- Dans Enterprise Designer pour la lecture et l'écriture de données
- Dans Metadata Insights pour créer des Physical Models

1. Accédez à la page **Sources de données** en utilisant l'un de ces modules :

**Management Console :**

Ouvrez Management Console à l'aide de l'URL : URL:

**Console :**

`http://serveur.port/managementconsole`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Ressources > Sources de données**.

**Metadata Insights :** Ouvrez Metadata Insights à l'aide de l'URL : URL: `http://serveur.port/metadata-insights`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Sources de données**.

2. Cliquez sur le bouton **Ajouter** .
3. Dans le champ **Nom**, entrez un nom pour la nouvelle connexion. Ce nom est entièrement de votre choix.

**Remarque :** Une fois que vous enregistrez une connexion, vous ne pouvez plus modifier le nom.

4. Dans le champ **Type**, sélectionnez **Amazon SimpleDB**.
5. Dans le champ **ID de clé d'accès**, saisissez la séquence alphanumérique de 20 caractères qui vous a été fournie pour accéder à votre compte Amazon AWS.
6. Dans le champ **Clé d'accès secrète**, saisissez la clé de 40 caractères nécessaire pour authentifier la connexion.
7. Pour tester la connexion, cliquez sur **Tester**.
8. Cliquez sur **Enregistrer**.

### Restrictions d'Amazon SimpleDB

#### Restriction d'écriture

Dans le stage Write to DB, le mode d'écriture **Mettre à jour** n'est pas disponible lors de l'écriture dans une table Amazon SimpleDB. L'option **Insérer** gère à la fois les opérations d'insertion et de mise à jour. Elle fait la distinction entre une insertion et une mise à jour à l'aide de la valeur unique de la colonne `ItemName` qui est présente dans toutes les tables Amazon SimpleDB.

Motif : une requête de mise à jour nécessite une clé principale pour chaque enregistrement de la table à mettre à jour, ce qui n'est pas pris en charge par les bases de données Amazon SimpleDB.

#### Restriction de lecture

Les fonctions d'agrégation `SUM` et `AVG` ne sont pas prises en charge lors de l'exécution des requêtes sur un Model Store.

## Connexion à Apache Cassandra

Vous pouvez utiliser cette connexion :

- Dans Enterprise Designer pour la lecture et l'écriture de données
- Dans Metadata Insights pour créer des Physical Models

1. Accédez à la page **Sources de données** en utilisant l'un de ces modules :

**Management Console :** Ouvrez Management Console à l'aide de l'URL : URL: `http://serveur.port/managementconsole`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Ressources > Sources de données**.

**Metadata Insights :** Ouvrez Metadata Insights à l'aide de l'URL : URL: `http://serveur.port/metadata-insights`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Sources de données**.

2. Cliquez sur le bouton **Ajouter** .
3. Dans le champ **Nom**, entrez un nom pour la nouvelle connexion. Ce nom est entièrement de votre choix.

**Remarque :** Une fois que vous enregistrez une connexion, vous ne pouvez plus modifier le nom.

4. Dans le champ **Type**, sélectionnez **Apache Cassandra**.
5. Dans le champ **Hôte**, saisissez le nom de l'ordinateur ou l'adresse IP d'installation de la base de données Apache Cassandra.
6. Dans le champ **Espace de clé**, saisissez le nom de l'espace de clé du centre de données auquel vous souhaitez accéder.
7. Dans le champ **Port**, saisissez le port de configuration de la base de données Apache Cassandra.
8. Saisissez le nom d'utilisateur et le mot de passe à utiliser pour l'authentification auprès de la base de données Cassandra.
9. Dans le champ **Niveau de cohérence**, sélectionnez le niveau de cohérence des lignes de données sur les nœuds répliqués pour permettre une transaction de données réussie. Il peut s'agir d'un nœud, de tous les nœuds ou d'une combinaison de nœuds disponibles.
10. Dans **Taille de l'extraction**, saisissez le nombre de lignes de jeux de résultats que vous souhaitez extraire à chaque opération de lecture.
11. Pour tester la connexion, cliquez sur **Tester**.
12. Cliquez sur **Enregistrer**.

## Restriction Apache Cassandra

La fonction d'agrégation `count` n'est pas prise en charge dans une requête sur un Model Store.

## Connexion au Cloud Azure

1. Accédez à la page **Sources de données** en utilisant l'un de ces modules :

**Management Console :** Ouvrez Management Console à l'aide de l'URL : URL: `http://serveur.port/managementconsole`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Ressources > Sources de données**.

**Metadata Insights :** Ouvrez Metadata Insights à l'aide de l'URL : URL: `http://serveur.port/metadata-insights`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Sources de données**.

2. Cliquez sur le bouton **Ajouter** .
3. Dans le champ **Nom**, entrez un nom pour la nouvelle connexion. Ce nom est entièrement de votre choix.

**Remarque :** Une fois que vous enregistrez une connexion, vous ne pouvez plus modifier le nom.

4. Dans le champ **Type**, sélectionnez **Cloud**.
5. Dans le champ **Service Cloud**, sélectionnez **AzureBlobStorage**.
6. Dans le champ **Protocole**, sélectionnez si vous souhaitez que la connexion entre Azure et Spectrum™ Technology Platform utilise HTTP ou HTTPS.
7. Dans le champ **Nom de compte**, saisissez le nom de votre compte de stockage Azure.
8. Dans le champ **Clé d'accès** saisissez la clé d'accès à votre compte Azure.
9. Pour tester la connexion Cloud, cliquez sur **Tester**.
10. Cliquez sur **Enregistrer**.

## Connexion à Flat File

### Connexion à Delimited Flat File

Vous pouvez utiliser cette connexion :

- Dans Enterprise Designer pour la lecture et l'écriture de données
- Dans Metadata Insights pour créer des Physical Models

1. Accédez à **Connexions > Flat File**.
2. Par défaut, l'écran s'ouvre en mode Création. Sinon, cliquez sur  pour ajouter une nouvelle connexion Flat File.
3. Saisissez un **Nom de connexion** pour la connexion aux données Flat File.
4. Saisissez le **Chemin d'accès** en cliquant sur **Parcourir**, puis en sélectionnant le répertoire du fichier.
5. Sélectionnez le **Chiffrement de caractères** du fichier plat dans la liste déroulante.
6. Définissez le **Type d'enregistrement** sur **Delimited**.
7. Dans **Délimiteur de champ**, sélectionnez le séparateur prévu entre deux champs d'un enregistrement de fichier.
8. Sélectionnez le **Qualificatif de texte (facultatif)**, le cas échéant, qui englobe les valeurs de champ d'un enregistrement de fichier.
9. Dans **Séparateur de ligne**, la valeur *Par défaut* est sélectionnée, indiquant que le séparateur de ligne prévu dépend de l'exécution de Spectrum™ Technology Platform sous un système Unix ou Windows.
10. Pour indiquer si la première ligne du fichier est ou non une ligne d'en-tête, déplacez le curseur **La première ligne est l'enregistrement d'en-tête** sur **Oui** ou sur **Non**.
11. Pour spécifier si le type de données des différents champs d'un enregistrement du fichier doit être automatiquement détecté ou non, déplacez le curseur **Détecter le type de données du fichier** sur **Oui** ou sur **Non**.
12. Pour ignorer les enregistrements non conformes lors de l'analyse de fichier, déplacez le curseur **Ignorer les enregistrements incorrects** sur **Activé**.
13. Cliquez sur **Test**.  
Un message confirme que le test de la connexion a réussi.
14. Cliquez sur **Enregistrer**.  
Un message confirme la création de la connexion.

Pour afficher un échantillon d'enregistrement extrait à l'aide de la connexion Delimited Flat File, cliquez sur **Prévisualisation** dans la barre d'en-tête. Les enregistrements de fichier sont extraits et les champs triés en fonction des détails que vous avez fournis.

## Connexion à Fixed Width Flat File

Vous pouvez utiliser cette connexion :

- Dans Enterprise Designer pour la lecture et l'écriture de données
- Dans Metadata Insights pour créer des Physical Models

1. Accédez à **Connexions > Flat File**.
2. Par défaut, l'écran s'ouvre en mode Création. Sinon, cliquez sur  pour ajouter une nouvelle connexion Flat File.
3. Saisissez un **Nom de connexion** pour la connexion aux données Flat File.
4. Saisissez le **Chemin d'accès** en cliquant sur **Parcourir**, puis en sélectionnant le répertoire du fichier.
5. Sélectionnez le **Chiffrement de caractères** du fichier plat dans la liste déroulante.
6. Définissez **Type d'enregistrement** sur **Fixed Width**.
7. Dans le champ **Longueur d'enregistrement**, saisissez le nombre total de caractères d'un enregistrement de fichier.

Répétez l'étape 8 à l'étape 13 pour saisir les détails de tous les champs prévus dans un enregistrement de fichier.

8. Cliquez sur **Ajouter un champ** pour ajouter une ligne pour un champ dans un enregistrement de fichier.
9. Dans le champ **Nom**, saisissez le nom de la valeur de champ.
10. Dans le champ **Type**, sélectionnez le type de données de la valeur de champ.
11. Dans la colonne **Position de début**, saisissez la position de l'enregistrement de fichier à partir de laquelle commence la valeur de champ.  
Pour le premier champ d'un enregistrement de fichier, le comptage de la **Position de début** commence à partir de 1.
12. Dans le champ **Longueur**, saisissez le nombre total de caractères couverts par le champ, y compris le caractère à la **Position de début**.

La somme des valeurs **Position de début** et **Longueur** de tout champ doit être inférieure ou égale à la **Longueur d'enregistrement**

Si l'enregistrement de fichier est :

```
01234Rob Smith29PitneyBowes
```

**Longueur d'enregistrement = 27**

Pour le champ « Nom » :

**Position de début = 6**

**Longueur = 9**

```
Name = Rob Smith
```

13. Cochez la case **Raccourcir** si vous souhaitez supprimer tous les espaces blancs au début et/ou à la fin d'une valeur de champ.
14. Cliquez sur **Test**.  
Un message confirme que le test de la connexion a réussi.
15. Cliquez sur **Enregistrer**.  
Un message confirme la création de la connexion.

Pour afficher un échantillon d'enregistrement extrait à l'aide de la connexion Fixed Width Flat File, cliquez sur **Prévisualisation** dans la barre d'en-tête. Les enregistrements de fichier sont extraits et les champs triés en fonction des détails que vous avez fournis.

### Formats de date et d'heure d'une connexion à un fichier

Lors de la lecture des valeurs de date et d'heure des fichiers à l'aide d'une connexion à un fichier dans Spectrum™ Technology Platform, les valeurs doivent respecter certains formats de date et d'heure spécifiques.

#### Formats de date et d'heure acceptés

- Date : « aaaa-mm-jj »
- Date et heure : « aaaa-mm-jj HH:mm:ss »
- Heure : « HH:mm:ss »

#### Fichiers délimités

Si la fonctionnalité **Détecter le type** est activée lors de la configuration de la connexion Delimited File, les valeurs de date et d'heure des enregistrements de fichier qui respectent les formats susmentionnés sont automatiquement détectées comme de type Date.

Si une valeur date-heure n'est pas conforme à l'un des formats acceptés, la valeur est lue comme une valeur de type Chaîne au lieu d'une valeur de type Date.

#### Fichiers de largeur fixe

Pour les fichiers de largeur fixe, les valeurs de type Date sont configurées lors de la création de la connexion au fichier de largeur fixe. Par conséquent, ces valeurs sont lues comme des valeurs de type Date, qu'elles respectent ou non les formats acceptés.

Si la valeur date-heure d'un fichier de type Fixed Width ne respecte pas les formats acceptés, elle doit être traitée à l'aide de **Transformations** dans le stage Logical Model Creation en appliquant cette fonction de catégorie *Conversion* à la valeur :

```
parsedate(String date, String format)
```

Dans ce cas, la *date* est la valeur reçue du fichier, tandis que le *format* est le format de date-heure dans lequel la valeur est reçue du fichier. Cela permet d'analyser correctement la valeur date-heure.

Par exemple, si la *date* = 23-fév-2008, le *format* = jj-*MMM*-aaaa.

### Formats de valeur obtenus

Lors de l'aperçu des données dans un Model Store :

- Si la valeur a été lue comme une valeur date/heure, elle apparaît dans l'aperçu sous l'un des formats de date/heure acceptés.
- Si la valeur a été lue comme une valeur Chaîne, elle apparaît comme telle dans l'aperçu.

## Connexion à un serveur FTP

Afin que Spectrum™ Technology Platform puisse accéder aux fichiers d'un serveur FTP, vous devez définir une connexion au serveur FTP à l'aide de Management Console. Une fois que cela est fait, vous pouvez créer des flux de données dans Enterprise Designer qui peuvent lire ou écrire des données dans les fichiers sur le serveur FTP.

Avant de vous connecter à un serveur FTP, vérifiez que les paramètres de délai d'attente de votre serveur FTP sont appropriés pour les jobs qui vont utiliser cette connexion. Suivant la conception d'un job, il peut se produire des périodes d'inactivité de la connexion, qui peuvent entraîner une temporisation de la connexion. Par exemple, vous pouvez disposer d'un flux de données avec deux stages Read From File connectés à un stage Import To Hub. Lorsque le stage Import To Hub lit les enregistrements d'un stage Read From File, l'autre est inactif, entraînant éventuellement une temporisation de la connexion au serveur FTP. Pour éviter la temporisation des connexions, vous pouvez envisager de définir la valeur de délai d'attente de votre serveur FTP sur 0.

**Remarque :** Le serveur FTP doit être exécuté en mode de connexion active. Le mode de connexion passive n'est pas pris en charge.

1. Accédez à la page **Sources de données** en utilisant l'un de ces modules :

**Management Console :** Ouvrez Management Console à l'aide de l'URL : URL: `http://serveur.port/managementconsole`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Ressources > Sources de données**.

**Metadata Insights :** Ouvrez Metadata Insights à l'aide de l'URL : URL: `http://serveur.port/metadata-insights`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Sources de données**.

2. Cliquez sur le bouton **Ajouter** .
3. Dans le champ **Nom**, entrez un nom pour la nouvelle connexion. Ce nom est entièrement de votre choix.
 

**Remarque** : Une fois que vous enregistrez une connexion, vous ne pouvez plus modifier le nom.
4. Dans le champ **Type**, sélectionnez **FTP**.
5. Dans les champs **Nom d'utilisateur** et **Mot de passe** saisissez les informations d'identification à utiliser pour authentifier le serveur FTP. Cela n'est nécessaire que si le serveur FTP l'exige.
6. Dans le champ **Hôte**, saisissez le nom d'hôte ou l'adresse IP du serveur FTP.
7. Dans le champ **Port** saisissez le numéro de port réseau utilisé par le serveur FTP.
8. Cliquez sur **Test** pour vérifier que le serveur Spectrum™ Technology Platform peut se connecter au serveur FTP.
9. Cliquez sur **Enregistrer**.

## Connexion à Google Cloud Storage

1. Accédez à la page **Sources de données** en utilisant l'un de ces modules :

**Management Console** : Ouvrez Management Console à l'aide de l'URL : URL: `http://serveur.port/managementconsole`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque** : Par défaut, le port HTTP est 8080.

Accédez à **Ressources > Sources de données**.

**Metadata Insights** : Ouvrez Metadata Insights à l'aide de l'URL : URL: `http://serveur.port/metadata-insights`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque** : Par défaut, le port HTTP est 8080.

Accédez à **Sources de données**.

2. Cliquez sur le bouton **Ajouter** .
3. Dans le champ **Nom**, entrez un nom pour la nouvelle connexion. Ce nom est entièrement de votre choix.

**Remarque** : Une fois que vous enregistrez une connexion, vous ne pouvez plus modifier le nom.

4. Dans le champ **Type**, sélectionnez **Cloud**.
5. Dans le champ **Service Cloud**, sélectionnez **GoogleCloudStorage**.
6. Dans le champ **Nom de compartiment**, saisissez le nom de compartiment tel que défini dans votre service Google Cloud. Il s'agit du compartiment où Spectrum™ Technology Platform lira et écrira des fichiers.
7. Saisissez le nom d'application, le compte de service et le fichier de clé privée fournis par Google.

**Remarque :** Assurez-vous que le fichier de clé privée est présent sur le serveur Spectrum™ Technology Platform.

8. Vous pouvez définir des droits d'accès dans la section **Droits d'accès**.
 

<b>Gérer vos données et autorisations</b>	Permet à l'utilisateur de gérer les données et les autorisations.
<b>Afficher vos données</b>	Permet à l'utilisateur d'afficher les données.
<b>Gérer vos données</b>	Permet à l'utilisateur de gérer les données.
9. Pour tester la connexion, cliquez sur **Tester**.
10. Cliquez sur **Enregistrer**.

Pour plus d'informations, reportez-vous à la documentation [Authentification du compte de service](#) de Google.

## Connexion à Hadoop

Vous pouvez créer des flux dans Enterprise Designer en utilisant ces stages du module Enterprise Big Data pour lire des données et écrire des données dans cette base de données.

- **Read from Hadoop Sequence File**
- **Write to Hadoop Sequence File**

**Avertissement :** Spectrum™ Technology Platform ne prend pas en charge *Hadoop 2.x* pour Kerberos sur les plates-formes Windows.

1. Accédez à la page **Sources de données** en utilisant l'un de ces modules :

**Management Console :** Ouvrez Management Console à l'aide de l'URL : URL: `http://serveur.port/managementconsole`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Ressources > Sources de données**.

**Metadata Insights :** Ouvrez Metadata Insights à l'aide de l'URL : URL: `http://serveur.port/metadata-insights`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Sources de données**.

2. Cliquez sur le bouton **Ajouter** .
3. Dans le champ **Nom**, entrez un nom pour la nouvelle connexion. Ce nom est entièrement de votre choix.

**Remarque :** Une fois que vous enregistrez une connexion, vous ne pouvez plus modifier le nom.

4. Dans le champ **Type**, sélectionnez **HDFS**.
5. Dans le champ **Hôte**, saisissez le nom d'hôte ou l'adresse IP de NameNode dans le cluster HDFS.
6. Dans le champ **Port**, entrez le numéro du port réseau.
7. Dans **Utilisateur**, sélectionnez l'une de ces options :

**Utilisateur serveur** Sélectionnez cette option si l'authentification est activée dans votre cluster HDFS. Cette option utilisera les noms d'utilisateur sous lequel le serveur Spectrum™ Technology Platform est exécuté pour s'authentifier dans HDFS.

**Nom utilisateur** Sélectionnez cette option si l'authentification est désactivée dans votre cluster HDFS.

8. Cochez la case **Kerberos** si vous souhaitez activer la fonctionnalité d'authentification Kerberos pour cette connexion au serveur de fichiers HDFS.
9. Si vous avez choisi d'activer l'authentification **Kerberos**, saisissez le chemin d'accès au fichier keytab dans le champ **Chemin d'accès au fichier Keytab**.

**Remarque :** Assurez-vous que le fichier de clé TAB est présent sur le serveur Spectrum™ Technology Platform.

10. Dans le champ **Protocole**, sélectionnez l'une des options suivantes :

**WEBHDFS** Sélectionnez cette option si le cluster HDFS exécute HDFS 1.0 ou version ultérieure. Ce protocole prend en charge les opérations de lecture et d'écriture.

**HFTP** Sélectionnez cette option si le cluster HDFS exécute une version antérieure à HDFS 1.0 ou si votre organisation n'autorise pas le protocole WEBHDFS. Ce protocole ne prend en charge que l'opération de lecture.

**HAR** Sélectionnez cette option pour accéder aux fichiers d'archive Hadoop. Si vous choisissez cette option, indiquez le chemin au fichier d'archive dans le champ **Path**. Ce protocole ne prend en charge que l'opération de lecture.

11. Développez **Options avancées**.

12. Si vous avez sélectionné le protocole WEBHDFS, vous pouvez spécifier les options avancées suivantes, le cas échéant :

**Facteur de réplication** Indique le nombre de nœuds de données dans lesquels répliquer chaque bloc. Par exemple, le paramètre par défaut de 3 réplique chaque bloc en trois nœuds différents dans le cluster. Le facteur de réplication maximal est de 1 024.

**Taille du bloc** Indique la taille de chaque bloc. HDFS décompose un fichier en blocs de la taille que vous indiquez ici. Par exemple, si vous indiquez la valeur par défaut 64 Mo, chaque fichier est décomposé en blocs de 64 Mo. Chaque bloc est ensuite répliqué en un nombre de nœuds dans le cluster spécifié dans le champ **Replication factor**.

**Permissions de fichier** Indique le niveau d'accès aux fichiers écrits dans le cluster HDFS par Spectrum™ Technology Platform. Vous pouvez spécifier des droits de lecture et d'écriture pour chacune des options suivantes :

**Remarque :** L'autorisation *Exécuter* n'est pas applicable à Spectrum™ Technology Platform.

**Utilisateur** Il s'agit de l'utilisateur indiqué ci-dessus, **Server user** ou l'utilisateur spécifié dans le champ **User Name**.

**Groupe** Cela fait référence à un groupe dont l'utilisateur est membre. Par exemple, si l'utilisateur est john123, les permissions Groupe s'appliquent à tout groupe dont john123 est membre.

**Autre** Cela fait référence à tout autre utilisateur, ainsi qu'aux groupes dont l'utilisateur spécifié n'est pas membre.

13. Utilisez les descriptions **Permissions de fichier** ci-dessous pour définir les propriétés du serveur pour Hadoop pour garantir que les fonctions de tri et de filtrage fonctionnent comme vous le souhaitez lorsque la connexion est utilisée dans un stage ou une activité. Pour ajouter des propriétés, effectuez l'une des étapes suivantes :

- Cliquez sur  et ajoutez les propriétés et leurs valeurs respectives dans les champs **Propriété** et **Valeur**.
- Cliquez sur  et chargez votre fichier de configuration XML. Le fichier XML doit être semblable à hdfs-site.xml, yarn-site.xml ou core-site.xml.

**Remarque :** Placez le fichier de configuration sur le serveur.

**Permissions de fichier et paramètres - Hadoop 1.x**

Cette section s'applique à ce stage et à cette activité :

- Stage **Read from Sequence File**
- Activité **Run Hadoop Pig**

**fs.default.name**

Indique le nœud et le port d'exécution de Hadoop. Par exemple,  
`hdfs://152.144.226.224:9000`

**mapred.job.tracker**

Indique le nom d'hôte ou l'adresse IP et le port d'exécution du suiveur de job MapReduce. Si le nom d'hôte est saisi comme local, les jobs sont exécutés sous la forme d'une seule carte et réduisent la tâche. Par exemple, `152.144.226.224:9001`

**dfs.namenode.name.dir**

Indique l'endroit où un nœud de nom DFS doit stocker la table nommée dans le système de fichiers local. S'il s'agit d'une liste de répertoires délimitée par des virgules, la table nommée est copiée dans tous les répertoires, à des fins de redondance. Par exemple, `file:/home/hduser/Data/namenode`

**hadoop.tmp.dir**

Indique l'emplacement de base des autres répertoires temporaires. Par exemple,  
`/home/hduser/Data/tmp`

**Permissions de fichier et paramètres - Hadoop 2.x**

Cette section s'applique à ce stage et à cette activité :

- Stage **Read from Sequence File**
- Activité **Run Hadoop Pig**

**fs.defaultFS**

Indique le nœud et le port d'exécution de Hadoop. Par exemple,  
`hdfs://152.144.226.224:9000`

**REMARQUE** : Pour les versions 11.0 et antérieures de Spectrum, le nom de paramètre `fs.defaultfs` doit être utilisé. Notez la différence de casse. Pour les versions 11 SP1 et ultérieures, les deux noms `fs.defaultfs` et `fs.defaultFS` sont valides. Nous vous recommandons d'utiliser le nom de paramètre `fs.defaultFS` pour les versions 11.0 SP1 et ultérieures.

**yarn.resourcemanager.resource-tracker.address**

Indique le nom d'hôte ou l'adresse IP de Resource Manager. Par exemple,  
`152.144.226.224:8025`

**yarn.resourcemanager.scheduler.address**

Indique l'adresse de Scheduler Interface. Par exemple, `152.144.226.224:8030`

**yarn.resourcemanager.address**

Indique l'adresse de l'interface Applications Manager contenue dans Resource Manager. Par exemple, 152.144.226.224:8041

### **mapreduce.jobhistory.address**

Indique le nom d'hôte ou l'adresse IP et le port d'exécution de MapReduce Job History Server. Par exemple, 152.144.226.224:10020

### **mapreduce.application.classpath**

Indique le CLASSPATH des applications MapReduce. Ce CLASSPATH indique l'emplacement des classes associées aux applications MapReduce. Les entrées doivent être séparées par des virgules.

Par exemple :

```
$HADOOP_CONF_DIR, $HADOOP_COMMON_HOME/share/hadoop/common/*,
$HADOOP_COMMON_HOME/share/hadoop/common/lib/*,
$HADOOP_HDFS_HOME/share/hadoop/hdfs/*,
$HADOOP_HDFS_HOME/share/hadoop/hdfs/lib/*,
$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/*,
$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/lib/*,
$HADOOP_YARN_HOME/share/hadoop/yarn/*,
$HADOOP_YARN_HOME/share/hadoop/yarn/lib/*
```

### **mapreduce.app-submission.cross-platform**

Gère plusieurs problèmes de plate-forme qui se produisent si votre serveur Spectrum est exécuté sur un ordinateur Windows et que vous installez Cloudera dessus. Si votre serveur Spectrum et Cloudera sont exécutés sur des systèmes d'exploitation différents, définissez la valeur de ce paramètre sur `true`. Sinon, définissez-la sur `false`.

**Remarque :** Cloudera ne prend pas en charge les clients Windows. La configuration de ce paramètre contourne le problème et ne constitue pas une solution à tous les problèmes de plate-forme qui se produisent.

### **Permissions de fichier et paramètres - Kerberos**

Cette section s'applique à ce stage et à cette activité :

- Stage **Read from Sequence File**
- Activité **Run Hadoop Pig**

Si vous avez coché la case **Kerberos**, ajoutez ces propriétés de configuration Kerberos : **hadoop.security.authentication**

Type de sécurité d'authentification utilisé. Saisissez la valeur `kerberos`.

### **yarn.resourcemanager.principal**

Principal Kerberos utilisé pour le Gestionnaire de ressources pour votre négociateur de ressources Hadoop YARN. Par exemple : `yarn/_HOST@HADOOP.COM`

### **dfs.namenode.kerberos.principal**

Principal Kerberos utilisé pour le namenode de votre Hadoop Distributed File System (HDFS). Par exemple, `hdfs/_HOST@HADOOP.COM`

#### **dfs.datanode.kerberos.principal**

Principal Kerberos utilisé pour le datanode de votre Hadoop Distributed File System (HDFS). Par exemple, `hdfs/_HOST@HADOOP.COM`

#### **Permissions de fichier et paramètres - Hadoop 1.x**

Cette section s'applique à ces stages :

- Stage **Read from File**
- Stage **Write to File**
- Stage **Read from Hive ORC File**
- Stage **Read from Hive ORC File**

#### **fs.default.name**

Indique le nœud et le port d'exécution de Hadoop. Par exemple, `hdfs://152.144.226.224:9000`

#### **Permissions de fichier et paramètres - Hadoop 2.x**

Cette section s'applique à ces stages :

- Stage **Read or write from File**
- Stage **Read or write from Hive ORC File**

#### **fs.defaultFS**

Indique le nœud et le port d'exécution de Hadoop. Par exemple, `hdfs://152.144.226.224:9000`

**REMARQUE** : Pour les versions 11.0 et antérieures de Spectrum, le nom de paramètre `fs.defaultfs` doit être utilisé. Notez la différence de casse. Pour les versions 11 SP1 et ultérieures, les deux noms `fs.defaultfs` et `fs.defaultFS` sont valides. Nous vous recommandons d'utiliser le nom de paramètre `fs.defaultFS` pour les versions 11.0 SP1 et ultérieures.

14. Pour tester la connexion, cliquez sur **Tester**.

15. Cliquez sur **Enregistrer**.

Après avoir défini une connexion à un cluster HDFS, celle-ci devient disponible dans les stages source et de collecteur de données dans Enterprise Designer, tels que **Read from File** et **Write to File**. Vous pouvez sélectionner le cluster HDFS lorsque vous cliquez sur **Ordinateur distant** lors de la définition d'un fichier dans un stage source ou de collecteur de données.

### **Prise en charge de la compression pour Hadoop**

Spectrum™ Technology Platform prend en charge les formats de compression `gzip` (`.gz`) et `bzip2` (`.bz2`) sur Hadoop. Lors de l'utilisation des stages **Read from File** et **Write to File** avec une connexion HDFS, incluez l'extension correspondant au format de compression requis (`.gz` ou `.bz2`) dans le champ **Nom de fichier**. Le fichier est décompressé ou compressé en fonction de

l'extension de compression spécifiée. Spectrum™ Technology Platform gère la compression et la décompression des fichiers.

## Connexion à une base de données JDBC

Définissez une connexion à l'aide de la page **Sources de données**. Vous pouvez vous rendre sur cette page via le module **Management Console** ou **Metadata Insights**.

1. Accédez à la page **Sources de données** en utilisant l'un de ces modules :

**Management Console :** Ouvrez Management Console à l'aide de l'URL : URL: `http://serveur.port/managementconsole`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Ressources > Sources de données**.

**Metadata Insights :** Ouvrez Metadata Insights à l'aide de l'URL : URL: `http://serveur.port/metadata-insights`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Sources de données**.

2. Cliquez sur le bouton **Ajouter** .
3. Dans le champ **Nom**, entrez un nom pour la nouvelle connexion. Ce nom est entièrement de votre choix.

**Remarque :** Une fois que vous enregistrez une connexion, vous ne pouvez plus modifier le nom.

4. Dans le champ **Type**, sélectionnez le type de base de données auquel vous souhaitez vous connecter.

Le module Spectrum™ Technology Platform Data Integration comprend les pilotes JDBC pour les bases de données SQL Server, Oracle et PostgreSQL. Pour vous connecter à un autre type de base de données, vous devez ajouter le pilote JDBC avant de définir une connexion.

5. Dans le champ **URL**, saisissez l'URL de connexion JDBC. Votre administrateur de base de données peut fournir cette URL.

Par exemple, pour vous connecter à une base de données MySQL nommée « SampleDatabase », hébergée sur un serveur nommé « MyServer », vous devez saisir :

```
jdbc:mysql://MyServer/SampleDatabase
```

6. Il se peut que vous deviez renseigner des champs supplémentaires, suivant le pilote JDBC. Les champs représentent les propriétés de la chaîne de connexion du pilote JDBC que vous avez sélectionné dans le champ **Type**. Reportez-vous à la documentation du fournisseur de pilote JDBC ou contactez votre administrateur de base de données pour avoir plus d'informations sur les propriétés de connexion spécifiques et les valeurs requises par le type de connexion.
7. Cliquez sur **Enregistrer**.
8. Testez la connexion en cochant la case en regard de la nouvelle connexion et en cliquant sur le bouton Test .

### Importation d'un pilote JDBC

Spectrum™ Technology Platform peut accéder aux données à partir de toute base de données via un pilote JDBC. Les pilotes pour SQL, Oracle et PostgreSQL sont fournis avec le module Spectrum™ Technology Platform Data Integration, qui inclut également des pilotes pour d'autres types de bases de données. Si Spectrum™ Technology Platform n'est pas fourni avec le pilote du type de base de données dont vous avez besoin, vous pouvez ajouter un pilote JDBC.

Dans cette procédure, vous allez importer un pilote JDBC en copiant les fichiers du pilote sur le serveur Spectrum™ Technology Platform. Une fois la procédure terminée, le pilote pourra être utilisé lors de la définition d'une connexion à une base de données JDBC dans Management Console.

**Remarque :** Cette procédure fonctionne pour les pilotes *JDBC 4.x*. Si le pilote que vous souhaitez ajouter utilise une version antérieure de JDBC, vous devez ajouter le pilote manuellement dans Management Console. (Pour plus d'informations, reportez-vous à la section [Ajout manuel d'un pilote JDBC](#) à la page 35.)

1. Placez tous les fichiers de pilote JDBC pour la base de données dans un dossier nommé :

```
Name.jdbc
```

Où *Name* est le nom de votre choix. Le nom du dossier doit se terminer par `.jdbc`.

2. Connectez-vous au serveur exécutant Spectrum™ Technology Platform.
3. Copiez le dossier contenant le pilote dans le dossier suivant :

```
Spectrum Location\server\app\drivers
```

Le pilote est automatiquement importé.

4. Pour vérifier que le pilote a été correctement importé, connectez-vous à Management Console et accédez à **Système > Pilotes**. Le pilote doit être répertorié.

Si le pilote n'est pas répertorié, ouvrez le journal système dans Management Console et recherchez les erreurs liées au déploiement de pilotes JDBC.

## Ajout manuel d'un pilote JDBC

Spectrum™ Technology Platform peut accéder aux données à partir de toute base de données via un pilote JDBC. Les pilotes pour SQL, Oracle et PostgreSQL sont fournis avec le module Spectrum™ Technology Platform Data Integration, qui inclut également des pilotes pour d'autres types de bases de données. Si Spectrum™ Technology Platform n'est pas fourni avec le pilote du type de base de données dont vous avez besoin, vous pouvez ajouter un pilote JDBC.

Dans cette procédure, vous allez ajouter des fichiers de pilote JDBC sur le serveur, puis définir manuellement la chaîne de connexion et les propriétés de connexion. Avant de commencer, assurez-vous que vous comprenez le format et les propriétés de la chaîne de connexion requis par le pilote. Pour que le pilote puisse fonctionner, vous devez les définir correctement dans l'ordre. Vous pouvez généralement trouver les informations sur la chaîne de connexion et les propriétés d'un pilote sur le site Web du fournisseur du pilote.

**Remarque :** Nous vous recommandons d'utiliser cette procédure uniquement lors de l'ajout d'un pilote JDBC qui utilise *JDBC 1.x*, *2.x* ou *3.x*. Si le pilote utilise *JDBC 4.x*, nous vous recommandons d'utiliser la méthode d'importation pour ajouter le pilote. Pour plus d'informations, reportez-vous à la section [Importation d'un pilote JDBC](#) à la page 34.

1. Ouvrez Management Console.
2. Accédez à **Système > Pilotes**.
3. Cliquez sur le bouton Ajouter .
4. Dans le champ **Nom**, saisissez un nom pour le pilote. Ce nom est entièrement de votre choix.
5. Dans le champ **JDBC driver class name**, saisissez le nom de la classe Java du pilote. En général, vous pouvez trouver le nom de la classe dans la documentation de votre pilote JDBC.

Par exemple, pour utiliser le pilote JDBC Microsoft, vous pouvez saisir les éléments suivants :

```
com.microsoft.sqlserver.jdbc.SQLServerDriver
```

6. Dans le champ **Modèle de chaîne de connexion**, saisissez l'URL de connexion JDBC à utiliser pour vous connecter à la base de données, y compris toutes les propriétés que vous souhaitez définir dans la chaîne de connexion. Les différents fournisseurs de bases de données utilisent des chaînes de connexion différentes ; c'est pourquoi vous devez consulter la documentation de votre base de données pour obtenir plus d'informations sur la chaîne de connexion.

Si le pilote doit être utilisé par plusieurs connexions de base de données, vous pouvez utiliser des jetons de propriété dans la chaîne de connexion au lieu des valeurs de propriété codées en dur qui peuvent varier pour chaque connexion. Par exemple, si vous souhaitez que certaines connexions utilisent le chiffrement et d'autres non, vous pouvez définir un jeton de propriété pour la propriété de chiffrement.

Pour utiliser un jeton de propriété dans la chaîne de connexion, appliquez la syntaxe suivante :

```
${PropertyToken}
```

Tout jeton de propriété que vous incluez dans le modèle de chaîne de connexion constituera un champ obligatoire lors de la définition d'une connexion à une base de données.

**Remarque :** Utilisez le nom de jeton de propriété `${mot de passe}` pour la propriété qui contiendra le mot de passe de base de données. Grâce à ce nom de jeton, le mot de passe est masqué dans le champ dans Management Console et est chiffré dans la base de données.

Par exemple, cette chaîne de connexion SQL contient des jetons de propriété pour l'hôte, le port, l'instance et le chiffrement :

```
jdbc:sqlserver://${host}:${port};databaseName=${instance};encrypt=${encryption};  
TrustServerCertificate=true
```

Ces jetons constituent des champs obligatoires lors de la définition d'une connexion à une base de données utilisant ce pilote :

Accueil > Ressources : Sources de données > Ajouter une source de données

## Ajouter une source de données

\*Nom

### Connexion

\*Type

\*Host

\*Port

\*Instance

\*encryption

7. S'il existe des propriétés que vous souhaitez rendre facultatives pour les connexions aux bases de données, définissez-les dans la section **Propriétés de connexion**.
  - a) Dans la section **Propriétés de connexion**, cliquez sur le bouton Ajouter .
  - b) Dans le champ **Libellé**, saisissez une description conviviale de la propriété. Le libellé que vous saisissez ici est utilisé comme libellé de champ dans la fenêtre Connexions lors de la création d'une connexion à l'aide de ce pilote.
  - c) Dans le champ **Jeton de propriété**, saisissez le jeton de la propriété facultative. Pour connaître les propriétés prises en charge par le pilote, reportez-vous à la documentation du pilote de base de données.

**Remarque :** Utilisez le nom de jeton de propriété `mot de passe` pour la propriété qui contiendra le mot de passe de base de données. Grâce à ce nom de jeton, le mot de passe est masqué dans le champ dans Management Console et est chiffré dans la base de données.

Par exemple, si vous souhaitez rendre le chiffrement facultatif pour les connexions aux bases de données utilisant ce pilote, vous pouvez définir la propriété de chiffrement comme suit :

Accueil > Système : pilotes > Modifier un pilote

## Modifier le pilote déployé

\*Nom

com.mysql.jdbc.Driver.5.1

\*Nom de classe du pilote JDBC ?

com.mysql.jdbc.Driver

\*Modèle de chaîne de connexion ?

jdbc:mysql://\${host}/\${instance}

### Propriétés et pilotes

Propriétés de la connexion ?



Libellé	Code de propriété
<input type="checkbox"/> username	user
<input type="checkbox"/> password	password
<input type="checkbox"/> Use SSL	useSSL

Lorsqu'une connexion à une base de données utilise ce pilote, la propriété de chiffrement est affichée sous forme de propriété facultative dans la connexion à la base de données :

Accueil > Ressources : Sources de données > Ajouter une source de données

## Ajouter une source de données

\*Nom

MyConnection

### Connexion

\*Type

com.mysql.jdbc.Driver.5.1

\*Host

\*Instance

User Name

Password

Use SSL

- Connectez-vous au serveur exécutant Spectrum™ Technology Platform et placez le fichier de pilote de base de données dans un dossier sur le serveur. L'emplacement n'a pas d'importance.
- Dans la section **Fichiers du pilote**, cliquez sur le bouton Ajouter .
- Dans le champ **File path**, saisissez le chemin d'accès au fichier de pilote de base de données sur le serveur.
- Cliquez sur **Enregistrer**.

## Suppression d'un pilote JDBC importé

Il n'est pas possible de supprimer des pilotes JDBC via Management Console si le pilote JDBC a été importé dans Spectrum™ Technology Platform au lieu d'avoir été ajouté manuellement dans Management Console. Au lieu de cela, suivez cette procédure pour supprimer le pilote.

**Important :** Avant de supprimer un pilote, assurez-vous qu'il n'existe aucune connexion à une base de données via le pilote.

1. Arrêtez le serveur Spectrum™ Technology Platform.
2. Accédez à ce dossier :

```
Spectrum Location\server\app\drivers
```

3. Dans le dossier `drivers`, supprimez le dossier contenant le pilote.
4. Démarrez le serveur Spectrum™ Technology Platform.
5. Pour vérifier que le pilote a été supprimé, connectez-vous à Management Console, accédez à **System > Drivers**, et vérifiez que le pilote n'est plus répertorié.

## Types de données de base de données pris en charge

Spectrum™ Technology Platform prend en charge les types de données suivants, couramment utilisés dans les bases de données :

<b>bigdecimal</b>	Type de données numérique prenant en charge 38 points de précision décimaux. Utilisez ce type de données pour les données utilisées dans les calculs mathématiques requérant un haut niveau de précision, en particulier ceux impliquant des données financières. Le type de données bigdecimal prend en charge des calculs plus précis que le type de données double.
<b>booléen</b>	Type logique doté de deux valeurs : True et False.
<b>date</b>	Type de données contenant un mois, un jour et une année. Par exemple, 2012-01-30 ou January 30, 2012. Vous pouvez indiquer un format de date par défaut dans Management Console.
<b>datetime</b>	Type de données contenant un mois, un jour, une année et des heures, des minutes et des secondes. Par exemple, 2012/01/30 6:15 PM.
<b>double</b>	Un type de données numérique qui contient à la fois des nombres en double précision négatifs et positifs entre $2^{-1074}$ et $(2-2^{-52}) \times 2^{1023}$ . En notation E, la plage de valeurs est comprise entre -1.79769313486232E+308 et 1.79769313486232E+308.
<b>flottant</b>	Un type de données numérique contenant des nombres de précision seuls négatifs et positifs entre $2^{-149}$ et $(2-2^{23}) \times 2^{127}$ . En notation E, la plage de valeurs est comprise entre -3.402823E+38 et 3.402823E+38.
<b>integer</b>	Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre $-2^{31}$ (-2,147,483,648) et $2^{31}-1$ (2,147,483,647).
<b>long</b>	Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre $-2^{63}$ (-9,223,372,036,854,775,808) et $2^{63}-1$ (9,223,372,036,854,775,807).

<b>chaîne</b>	Séquence de caractères.
<b>time</b>	Type de données contenant l'heure du jour. Par exemple, 21:15:59 ou 9:15:59 PM.
<b>Raw</b>	Type de données Oracle pour le stockage de données binaires ) longueur variable. La taille maximale est de 2 000 octets (la longueur maximale dans Oracle 7 était de 255 octets).

D'autres types de données de base de données sont automatiquement mappés vers l'un des types de données pris en charge comme suit :

Type de données de base de données	Type de données prises en charge
<b>Types Date/Time</b>	
TIMESTAMP	datetime
<b>Types chaînes</b>	
CHAR	string
CLOB	string
LONGVARCHAR	string
NCHAR	string
NVARCHAR	string
VARCHAR	string
<b>Types numériques</b>	
BIGINT	long
DECIMAL	double
FLOAT	double
NUMERIC	bigdecimal
REAL	flottant

Type de données de base de données	Type de données prises en charge
SMALLINT	integer
TINYINT	integer
<b>Types Booléen</b>	
BIT	booléen

### Prise en charge des types de données de base de données pour le module Location Intelligence

Ces types de données de base de données sont automatiquement mappés vers l'un des types de données pris en charge du module Location Intelligence.

Type de données de base de données	Type de données prises en charge
<b>SQL Server</b>	
tinyint	SHORT_INTEGER
smallint	SHORT_INTEGER
int	INTEGER
bigint	LONG_INTEGER
flottant	DOUBLE
real	DOUBLE
decimal(10, 5)	DOUBLE
numeric(10, 5)	DOUBLE
date	DATE
time	TIME
datetime	DATE_TIME

Type de données de base de données	Type de données prises en charge
smalldatetime	DATE_TIME
char (10)	STRING
varchar(10)	STRING
nchar(10)	STRING
nvarchar(10)G	STRING
binary(10)	BINARY
varbinary(10)	BINARY
<b>PostGIS</b>	
smallint	SHORT_INTEGER
integer	INTEGER
bigint	LONG_INTEGER
numeric(10, 5)	DOUBLE
real	DOUBLE
double precision	DOUBLE
serial	INTEGER
bigserial	LONG_INTEGER
bytea	BINARY
date	DATE
time	TIME

Type de données de base de données	Type de données prises en charge
timestamp	DATE_TIME
character(10)	STRING
character varying(10)	STRING
nchar(10)	STRING
<b>Oracle</b>	
NUMBER	DOUBLE
CHAR(10)	STRING
VARCHAR(10)	STRING
VARCHAR2(10)	STRING
NCHAR(10)	STRING
NVARCHAR2(10)	STRING
DATE	DATE_TIME
TIMESTAMP	DATE_TIME
BLOB	BINARY
<b>SAP HANA</b>	
tinyint	SHORT_INTEGER
smallint	SHORT_INTEGER
integer	INTEGER

Type de données de base de données	Type de données prises en charge
bigint	LONG_INTEGER
smalldecimal	DOUBLE
decimal(10, 5)	DOUBLE
real	DOUBLE
double	DOUBLE
float(30)	DOUBLE
varchar(30)	STRING
nchar(10)	STRING
nvarchar(30)	STRING
alphanum(30)	STRING
date	DATE
time	TIME
seconddate	DATE_TIM
timestamp	DATE_TIM
varbinary(30)	BINARY

### Restrictions

- Les connecteurs MongoDB/Cassandra ne sont pas pris en charge via PrestoDB dans Metadata Insights. Il existe des connecteurs distincts pour MongoDB et Cassandra.
- %Write to Any DB; via Presto n'est pas recommandé par Presto DB et n'est donc pas pris en charge par le connecteur Presto JDBC.

## Connexion à Knox

Une passerelle Apache Knox vous permet d'accéder à un service Hadoop via une couche de sécurité Knox.

Avec cette connexion, vous pouvez créer des flux dans Enterprise Designer à l'aide de stages du module Enterprise Big Data pour lire des données et écrire des données dans Hadoop via Knox.

1. Accédez à la page **Sources de données** en utilisant l'un de ces modules :

**Management Console :** Ouvrez Management Console à l'aide de l'URL : URL: `http://serveur.port/managementconsole`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Ressources > Sources de données**.

**Metadata Insights :** Ouvrez Metadata Insights à l'aide de l'URL : URL: `http://serveur.port/metadata-insights`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Sources de données**.

2. Cliquez sur le bouton **Ajouter** .
3. Dans le champ **Nom**, entrez un nom pour la nouvelle connexion. Ce nom est entièrement de votre choix.

**Remarque :** Une fois que vous enregistrez une connexion, vous ne pouvez plus modifier le nom.

4. Dans le champ **Type**, sélectionnez *Passerelle*.
5. Dans le champ **Type de passerelle**, sélectionnez **Knox**.
6. Dans le champ **Hôte**, saisissez le nom d'hôte ou l'adresse IP du nœud du cluster HDFS exécutant la passerelle.
7. Dans le champ **Port**, saisissez le numéro de port de la passerelle Knox.
8. Dans le champ **Nom d'utilisateur**, saisissez le nom d'utilisateur de la passerelle Knox.
9. Dans le champ **Mot de passe**, saisissez le mot de passe vous autorisant à accéder à la passerelle Knox.
10. Dans le champ **Nom de passerelle**, saisissez le nom de la passerelle Knox à laquelle vous souhaitez accéder.

11. Dans le champ **Nom de cluster**, saisissez le nom du cluster Hadoop auquel accéder.
12. Dans le champ **Protocole**, sélectionnez `webhdfs`.
13. Dans le champ **Nom de service**, saisissez le nom du service Hadoop auquel accéder.
14. Pour tester la connexion, cliquez sur **Tester**.
15. Cliquez sur **Enregistrer**.

Après avoir défini une connexion Knox à un cluster HDFS, vous pouvez utiliser la connexion dans Enterprise Designer, dans les stages **Read from File** et **Write to File**. Vous pouvez sélectionner le cluster HDFS lorsque vous cliquez sur **Ordinateur distant** lors de la définition d'un fichier dans un stage source ou de collecteur de données.

## Connexion à Marketo

1. Accédez à la page **Sources de données** en utilisant l'un de ces modules :

**Management Console :** Ouvrez Management Console à l'aide de l'URL : URL: `http://serveur.port/managementconsole`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Ressources > Sources de données**.

**Metadata Insights :** Ouvrez Metadata Insights à l'aide de l'URL : URL: `http://serveur.port/metadata-insights`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Sources de données**.

2. Cliquez sur le bouton **Ajouter** .
3. Dans le champ **Nom**, entrez un nom pour la nouvelle connexion. Ce nom est entièrement de votre choix.

**Remarque :** Une fois que vous enregistrez une connexion, vous ne pouvez plus modifier le nom.

4. Dans le champ **Type**, sélectionnez **Marketo**.
5. Dans le champ **URL d'accès**, saisissez l'URL de votre compte Marketo.

Pour trouver votre URL d'accès, connectez-vous à votre compte Marketo et accédez à **Admin > Intégration > Services Web**. L'URL d'accès se trouve sous l'en-tête **API REST** et se présente au format suivant :

```
https://AccountID.mktorest.com/rest
```

Copiez la partie de l'URL avant /rest. Il peut s'agir par exemple de  
https://AccountID.mktorest.com.

6. Saisissez l'ID client et la clé secrète client de votre compte Marketo.

Pour trouver votre ID client et votre clé secrète, connectez-vous à votre compte Marketo et accédez à **Admin > Intégration > LaunchPoint > API Rest > Afficher les détails**. La fenêtre contextuelle affiche les détails.

7. Pour tester la connexion, cliquez sur **Tester**.
8. Cliquez sur **Enregistrer**.

### Restrictions de Marketo

1. Cette requête s'applique uniquement aux entités `List` et `Activity_type`. Pour les autres, fournissez le type de filtre.

```
Select * from Marketo_Table
```

2. Ne prend pas en charge les jointures, sauf entre les entités `Lead` et `Lead_List`. La requête de jointure entre `Lead` et `Lead_List` d'un élément `List_Id` se présente comme suit :

```
Select Lead.* from Lead Inner Join Lead_List
On Lead.ID = Lead_List.Lead_ID
And Lead_List.List_ID = <List ID>
```

### Entités et opérations prises en charge

Les entités sont de ces types :

1. Entité
2. Entity Update : il s'agit d'une table virtuelle utilisée pour la mise à jour de l'entité `Lead`. Par exemple, **Merge\_Leads** doit être utilisé pour la fusion des différents prospects Marketo.

## Connexion à Microsoft Dynamics 365

### Connexion à Microsoft Dynamics 365 Online

Spectrum™ Technology Platform prend en charge la connexion à Microsoft Dynamics 365 (version 9) uniquement.

1. Accédez à la page **Sources de données** en utilisant l'un de ces modules :

**Management Console :** Ouvrez Management Console à l'aide de l'URL : URL: `http://serveur.port/managementconsole`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Ressources > Sources de données**.

**Metadata Insights :** Ouvrez Metadata Insights à l'aide de l'URL : URL: `http://serveur.port/metadata-insights`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Sources de données**.

2. Cliquez sur le bouton **Ajouter** .
3. Dans le champ **Nom**, entrez un nom pour la nouvelle connexion. Ce nom est entièrement de votre choix.

**Remarque :** Une fois que vous enregistrez une connexion, vous ne pouvez plus modifier le nom.

4. Dans le champ **Type**, sélectionnez **Microsoft Dynamics 365**.
5. Dans le champ **Type de déploiement**, sélectionnez **En ligne**.
6. Dans le champ **Nom d'utilisateur**, saisissez votre nom d'utilisateur Microsoft Dynamics.
7. Dans le champ **Mot de passe**, saisissez votre mot de passe Microsoft Dynamics.
8. Dans le champ **Nom d'organisation** saisissez votre nom unique d'organisation, qui identifie votre instance CRM.

Pour trouver votre nom unique d'organisation, connectez-vous à Microsoft Dynamics et accédez à **Paramètres > Personnalisation > Personnalisations > Ressources destinées aux développeurs**. Votre nom unique d'organisation s'affiche.

9. Dans le champ **Région** sélectionnez la région géographique de votre compte Microsoft Dynamics.
10. Pour tester la connexion, cliquez sur **Tester**.
11. Cliquez sur **Enregistrer**.

### Connexion à Microsoft Dynamics 365 On Premise

Ce connecteur de Spectrum™ Technology Platform prend en charge les demandes basées sur les authentifications pour Microsoft Dynamics 365 On Premises.

**Pré-requis :**

**Importer le certificat dans le fichier de magasin de clés :** pour importer les certificats du serveur Dynamics CRM dans le magasin de clés de distribution Spectrum Java, procédez comme suit :

1. Copiez les certificats de serveur dans un dossier local.
2. Accédez à ce chemin d'accès à la distribution Spectrum JAVA :  
<SPECTRUM\_HOME>\java\jre\lib\security.
3. Exécutez cette commande pour importer les certificats : `keytool -importcert -alias <certificate alias name> -file " <certificate path>\<certificate name>" -keystore keystore.jks` sous Windows et `keytool -import -alias <certificate alias name> -file "<certificate path>/<certificate name>" -keystore keystore.jks` sous Unix.

**Définition d'une connexion à Microsoft Dynamics 365 On Premise**

Procédez comme suit pour définir une connexion à **Microsoft Dynamics 365 On Premises** :

1. Accédez à la page **Sources de données** en utilisant l'un de ces modules :

**Management Console :** Ouvrez Management Console à l'aide de l'URL : URL: `http://serveur:port/managementconsole`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Ressources > Sources de données**.

**Metadata Insights :** Ouvrez Metadata Insights à l'aide de l'URL : URL: `http://serveur:port/metadata-insights`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Sources de données**.

2. Cliquez sur le bouton **Ajouter** .
3. Dans le champ **Nom**, entrez un nom pour la nouvelle connexion. Ce nom est entièrement de votre choix.

**Remarque :** Une fois que vous enregistrez une connexion, vous ne pouvez plus modifier le nom.

4. Cliquez sur **Microsoft Dynamics 365** dans le **Type**.
5. Cliquez sur **On Premise** dans le **Type de déploiement**.

6. Saisissez votre nom d'utilisateur Microsoft Dynamics dans le **Nom d'utilisateur**.
7. Saisissez votre mot de passe Microsoft Dynamics dans le **Mot de passe**.
8. Saisissez le nom de l'hôte dans le **Nom d'hôte**.
9. Saisissez le nom de port dans le **Nom de port**.
10. Saisissez l'URL de STS dans **URL STS**.
11. Cliquez sur **Test** pour tester la connexion.
12. Cliquez sur **Enregistrer**.

## Restrictions

**Créer/Mettre à jour** :Créer/Mettre à jour peut échouer si une colonne d'une entité est mappée vers plusieurs entités de référence. Par exemple, « ParentCustomerId » de Customer peut être associé à Account, Lead, etc. Pour résoudre ce problème, les données de cette colonne doivent se présenter au format suivant : « ReferenceEntityName:GUID » au lieu de « GUID ».

## Entités et opérations prises en charge

Les entités sont de ces types :

1. Appartenant à l'utilisateur
2. Appartenant à l'organisation
3. Appartenant à l'entreprise
4. Aucun

## Connexion à un Model Store

Connectez-vous à un Model Store pour utiliser les données fédérées de différentes sources telles que des bases de données, des serveurs de fichiers et des services Cloud. Une fois qu'une connexion est définie, vous pouvez utiliser les données des Logical Models et des Physical Models d'un Model Store (créés et déployés dans Metadata Insights) dans les stages **Read from DB** et **Write to DB** d'Enterprise Designer.

1. Accédez à la page **Sources de données** en utilisant l'un de ces modules :

**Management Console :** Ouvrez Management Console à l'aide de l'URL : URL: `http://serveur.port/managementconsole`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Ressources > Sources de données**.

**Metadata Insights :** Ouvrez Metadata Insights à l'aide de l'URL : URL: `http://serveur.port/metadata-insights`, où *serveur* est le nom du serveur ou

l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Sources de données**.

2. Cliquez sur le bouton **Ajouter** .
3. Dans le champ **Nom**, entrez un nom pour la nouvelle connexion. Ce nom est entièrement de votre choix.

**Remarque :** Une fois que vous enregistrez une connexion, vous ne pouvez plus modifier le nom.

4. Dans le champ **Type**, sélectionnez **Model Store**.
5. Dans le champ **Model Store**, saisissez le nom du Model Store auquel vous souhaitez vous connecter.

Pour trouver les noms des Model Stores disponibles, ouvrez Metadata Insights sur Modeling, puis cliquez sur l'onglet **Model Store**.

6. Pour tester la connexion, cliquez sur **Tester**.
7. Cliquez sur **Enregistrer**.

## Connexion à NetSuite

Lors de la lecture et de l'écriture d'une connexion NetSuite, les deux modes, interactif et par lots, sont pris en charge. Spectrum™ Technology Platform prend en charge les types d'entité NetSuite suivants :

- Enregistrements standard
- Enregistrements personnalisés
- Recherches enregistrées
- Jointures entre des enregistrements standard

Pour vous connecter à NetSuite :

1. Accédez à la page **Sources de données** en utilisant l'un de ces modules :

**Management Console :** Ouvrez Management Console à l'aide de l'URL : URL: `http://serveur.port/managementconsole`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Ressources > Sources de données**.

**Metadata Insights :**

Ouvrez Metadata Insights à l'aide de l'URL : URL: `http://serveur.port/metadata-insights`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Sources de données**.

2. Cliquez sur le bouton **Ajouter** .
3. Dans le champ **Nom**, entrez un nom pour la nouvelle connexion. Ce nom est entièrement de votre choix.

**Remarque :** Une fois que vous enregistrez une connexion, vous ne pouvez plus modifier le nom.

4. Dans le champ **Type**, sélectionnez **NetSuite**.
5. Dans le champ **Email**, saisissez l'adresse électronique associée au compte NetSuite à utiliser pour la connexion.
6. Dans le champ **Mot de passe**, saisissez le mot de passe du compte NetSuite.
7. Dans le champ **Compte**, saisissez le nom d'utilisateur du compte NetSuite.
8. Dans le champ **Rôle**, sélectionnez le rôle approprié pour cette connexion parmi les rôles mappés vers le compte d'utilisateur NetSuite donné.

Le champ **Rôle** est facultatif. Si vous laissez le champ **Rôle** vide, le rôle par défaut est utilisé pour se connecter via la connexion.

**Avertissement :** Seuls les rôles standard sont pris en charge. Les rôles personnalisés ne sont pas pris en charge.

9. Pour tester la connexion, cliquez sur **Tester**.
10. Cliquez sur **Enregistrer**.

**Remarque :** Pour effectuer une opération `INSERT` sur un enregistrement à l'aide d'une connexion NetSuite, utilisez une requête `UPSERT` avec la clé primaire (`internalId`) vide.

### Restrictions de NetSuite

1. Lors de l'interrogation à l'aide de jointures, vous devez indiquer des colonnes spécifiques. Par exemple, la requête suivante n'est pas prise en charge :

```
select * from CUSTOMER_M
```

2. Les connexions simultanées à NetSuite ne sont pas prises en charge, car NetSuite autorise une seule connexion par compte.
3. Vous pouvez écrire uniquement des enregistrements standard et personnalisés.
4. Pour les deux requêtes `UPDATE` et `UPSERT`, une opération `UPSERT` est réalisée.
5. Dans le stage Write to DB, la taille de lot maximale autorisée pour une opération `insert` est de 200 et la taille de lot maximale pour une opération `update` est de 100.

### Entités et opérations prises en charge

Les entités sont de ces types :

1. Enregistrements standard
2. Enregistrements personnalisés
3. Jointures
4. Recherches enregistrées

**Remarque :** Dans une table de connexion NetSuite, la colonne de clé primaire est `internalId`.

## Connexion à NoSQL

Les types de base de données NoSQL pris en charge sont les suivants :

- Couchbase
- MongoDB

Vous pouvez créer des flux dans Enterprise Designer en utilisant ces stages du module Enterprise Big Data pour lire des données et écrire des données dans cette base de données.

- Requête NoSQL DB
- Read from NoSQL DB
- Write to NoSQL DB

1. Accédez à la page **Sources de données** en utilisant l'un de ces modules :

**Management Console :** Ouvrez Management Console à l'aide de l'URL : URL: `http://serveur.port/managementconsole`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Ressources > Sources de données**.

**Metadata Insights :** Ouvrez Metadata Insights à l'aide de l'URL : URL: `http://serveur.port/metadata-insights`, où *serveur* est le nom du serveur ou

l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Sources de données**.

2. Cliquez sur le bouton **Ajouter** .
3. Dans le champ **Nom**, entrez un nom pour la nouvelle connexion. Ce nom est entièrement de votre choix.

**Remarque :** Une fois que vous enregistrez une connexion, vous ne pouvez plus modifier le nom.

4. Dans le champ **Type**, sélectionnez l'une des options suivantes :
  - Couchbase
  - MongoDB
5. Spécifiez l'**Hôte**, le **Port**, la **Base de données**, le **Nom d'utilisateur** et le **Mot de passe** de la base de données NoSQL spécifique à laquelle vous souhaitez accéder.
6. Cliquez sur **Test** pour vérifier que la connexion à la base de données fonctionne.
7. Cliquez sur **OK**.

## Connexion à Salesforce

Vous pouvez utiliser cette connexion :

- Dans Enterprise Designer pour la lecture et l'écriture de données
- Dans Metadata Insights pour créer des Physical Models

1. Accédez à la page **Sources de données** en utilisant l'un de ces modules :

**Management Console :** Ouvrez Management Console à l'aide de l'URL : URL: `http://serveur.port/managementconsole`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Ressources > Sources de données**.

**Metadata Insights :** Ouvrez Metadata Insights à l'aide de l'URL : URL: `http://serveur.port/metadata-insights`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Sources de données**.

2. Cliquez sur le bouton **Ajouter** .
3. Dans le champ **Nom**, entrez un nom pour la nouvelle connexion. Ce nom est entièrement de votre choix.

**Remarque :** Une fois que vous enregistrez une connexion, vous ne pouvez plus modifier le nom.

4. Dans le champ **Type**, sélectionnez **Salesforce**.
5. Dans le champ **Nom d'utilisateur**, saisissez l'ID de courrier électronique enregistré dans le magasin de données Salesforce.
6. Dans le champ **Mot de passe**, saisissez une combinaison du mot de passe du portail Salesforce et du jeton de sécurité généré via le portail Salesforce.

Par exemple, si votre mot de passe est Sales@Test et si le jeton de sécurité fourni par Salesforce est 56709367, le mot de passe d'authentification de cette connexion Salesforce est Sales@Test56709367.

7. Définissez **Utiliser le point d'extrémité par défaut** sur **Non** si vous souhaitez utiliser une URL de point d'extrémité spécifique pour accéder aux données Salesforce. Saisissez l'URL requise dans le champ **URL Salesforce** affiché juste en dessous de **Utiliser le point d'extrémité par défaut**.

**Remarque :** L'**URL Salesforce** est obligatoire.

8. Définissez **Utiliser la lecture en masse** sur **Oui** si vous souhaitez extraire des données en masse à partir de Salesforce. La valeur par défaut est **Non**.

**Remarque :** La requête d'extraction de données en masse ne fonctionne pas pour les champs composés d'adresse et de géolocalisation.

. Pour plus d'informations sur les considérations et les limitations d'une requête en masse, reportez-vous aux sections [Considérations et restrictions relatives aux champs composés](#) et [Limites de l'API Bulk](#).

9. Pour tester la connexion, cliquez sur **Tester**.
10. Cliquez sur **Enregistrer**.

**Remarque :** Les champs d'audit sont activés sur toutes les tables par défaut. Les champs d'audit Salesforce sont les suivants :

- Date de création
- Date de la dernière modification
- Créé par

- Dernière modification effectuée par

**Avertissement** : Les Physical Models créés dans Spectrum™ Technology Platform, versions 10 et antérieures, à l'aide de connexions Salesforce, doivent être ouvertes et enregistrées de nouveau pour que les champs d'audit de leurs tables soient activés.

### Restriction Salesforce

Les fonctions d'agrégation ne sont pas prises en charge lors de l'exécution des requêtes sur un Model Store.

## Connexion à SAP NetWeaver

La création d'une connexion SAP NetWeaver dans Management Console via OData Services vous permet de lire, d'écrire et de synchroniser vos données CRM et ERP. Lors de la lecture et de l'écriture d'une connexion SAP, les deux modes, interactif et par lots, sont pris en charge.

Pour définir une connexion SAP NetWeaver, procédez comme suit :

1. Accédez à la page **Sources de données** en utilisant l'un de ces modules :

**Management Console :** Ouvrez Management Console à l'aide de l'URL : URL: `http://serveur.port/managementconsole`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque** : Par défaut, le port HTTP est 8080.

Accédez à **Ressources > Sources de données**.

**Metadata Insights :** Ouvrez Metadata Insights à l'aide de l'URL : URL: `http://serveur.port/metadata-insights`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque** : Par défaut, le port HTTP est 8080.

Accédez à **Sources de données**.

2. Cliquez sur le bouton **Ajouter** .
3. Dans le champ **Nom**, entrez un nom pour la nouvelle connexion. Ce nom est entièrement de votre choix.

**Remarque** : Une fois que vous enregistrez une connexion, vous ne pouvez plus modifier le nom.

4. Dans le champ **Type**, sélectionnez **SAP**.
5. Dans le champ **Nom d'utilisateur**, saisissez le nom d'utilisateur permettant d'accéder au service Web SAP.
6. Dans le champ **Mot de passe**, saisissez le mot de passe du service Web SAP.
7. Dans le champ **OdataURL**, saisissez l'adresse du service Web Odata à utiliser pour cette connexion.
8. Cliquez sur **Test**.  
Un message confirme que le test de la connexion a réussi.
9. Cliquez sur **Enregistrer**.  
Un message confirme la création de la connexion.

**Remarque :** Pour effectuer des opérations d'extraction, un service OData doit prendre en charge les opérations \$skip et \$top. Si le service ne prend pas en charge ces opérations, les enregistrements extraits affichent des incohérences dans l'aperçu Model Store.

### Restrictions de SAP NetWeaver

Pour les deux opérations UPDATE et UPSERT, une opération UPDATE est réalisée.

### Entités et opérations prises en charge

Les colonnes d'entités sont de deux types :

- Natif : les colonnes avec des types de données natifs sont affichées avec leurs types de données respectifs.
- Personnalisé : les colonnes avec des types de données personnalisés sont affichées avec un type de données vide.

Pour déployer un Model Store dérivé d'une connexion SAP, assurez-vous que ses Physical Models et Logical Models incluent uniquement les entités dont les colonnes sont de types de données natifs. Si les modèles possèdent des entités avec des types de données personnalisés, il est impossible de déployer le Model Store.

## Connexion à SharePoint

1. Accédez à la page **Sources de données** en utilisant l'un de ces modules :

**Management Console :** Ouvrez Management Console à l'aide de l'URL : URL: `http://serveur.port/managementconsole`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Ressources > Sources de données**.

**Metadata Insights :** Ouvrez Metadata Insights à l'aide de l'URL : URL: `http://serveur.port/metadata-insights`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Sources de données**.

2. Cliquez sur le bouton **Ajouter** .
3. Dans le champ **Nom**, entrez un nom pour la nouvelle connexion. Ce nom est entièrement de votre choix.

**Remarque :** Une fois que vous enregistrez une connexion, vous ne pouvez plus modifier le nom.

4. Dans le champ **Type**, sélectionnez **Cloud**.
5. Dans le champ **Service Cloud**, sélectionnez **Sharepoint**.
6. Dans le champ **Versión**, sélectionnez **v2010**. Spectrum™ Technology Platform prend actuellement en charge SharePoint version 2010.
7. Dans le champ **Protocole**, sélectionnez le protocole requis pour se connecter Sharepoint.
8. Dans le champ **Adresse du serveur**, saisissez le nom d'hôte ou l'adresse IP du serveur SharePoint auquel vous souhaitez vous connecter.
9. Saisissez le nom d'utilisateur et le mot de passe à utiliser pour l'authentification dans SharePoint.
10. Dans le champ **Projet**, saisissez le projet spécifique correspondant à l'emplacement SharePoint auquel vous souhaitez accéder.
11. Pour tester la connexion, cliquez sur **Tester**.
12. Cliquez sur **Enregistrer**.

#### Exemple

Par exemple, imaginons que vous souhaitiez créer une connexion à cette URL SharePoint :

```
https://sharepoint.example.com/sites/myportal
```

Vous devez renseigner les champs **Protocole**, **Adresse du serveur** et **Projet** comme suit :

- **Protocole** : https
- **Adresse du serveur** : sharepoint.example.com
- **Projet** : myportal

## Connexion à Splunk

1. Accédez à la page **Sources de données** en utilisant l'un de ces modules :

**Management Console :** Ouvrez Management Console à l'aide de l'URL : URL: `http://serveur.port/managementconsole`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Ressources > Sources de données**.

**Metadata Insights :** Ouvrez Metadata Insights à l'aide de l'URL : URL: `http://serveur.port/metadata-insights`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Sources de données**.

2. Cliquez sur le bouton **Ajouter** .
3. Dans le champ **Nom**, entrez un nom pour la nouvelle connexion. Ce nom est entièrement de votre choix.

**Remarque :** Une fois que vous enregistrez une connexion, vous ne pouvez plus modifier le nom.

4. Dans le champ **Type**, sélectionnez **Splunk**.
5. Dans le champ **Nom d'utilisateur**, saisissez le nom d'utilisateur du compte Splunk d'authentification de l'instance Splunk.
6. Dans le champ **Mot de passe**, saisissez le mot de passe du compte Splunk.
7. Dans le champ **Hôte**, saisissez l'adresse ou le nom d'hôte du serveur d'hébergement de la source de données Splunk.
8. Dans le champ **Port**, saisissez le numéro de port de la source de données Splunk.
9. Pour tester la connexion, cliquez sur **Tester**.
10. Cliquez sur **Enregistrer**.

### Restrictions de Splunk

Cette requête n'est pas prise en charge :

```
select count(*) from SplunkTable
```

## Entités et opérations prises en charge

### Opérations prises en charge

LIKE, ORDER BY, LIMIT, IN, BETWEEN, !=, <=, >=, <,>, différents opérateurs AND/OR.

### Fonctions prises en charge

- Fonctions de chaîne : upper, lower, length, len, ltrim, rtrim, substring, max, min
- Fonctions mathématiques : abs, ceil, exp, floor, sqrt, round

**Remarque :** Pour toutes les autres opérations de requête, utilisez la colonne Splunk `search` comme indiqué ci-dessous.

Spectrum™ Technology Platform fournit une colonne `search` dans la table `Splunk`, à l'aide de laquelle vous pouvez rechercher les données requises dans la connexion Splunk.

Lors de l'exécution d'une requête `select` sur `SplunkTable`, utilisez la colonne `search` dans la clause `where` dans l'un des scénarios ci-dessous :

1. Pour inclure les critères de recherche qui ne peuvent pas être spécifiés à l'aide de la syntaxe SQL ANSI.
2. Pour inclure les critères de recherche propres à Splunk qui ne peuvent pas être inclus dans le cadre de la requête SQL principale.

Par exemple, la requête ci-dessous recherche une valeur `_raw` qui contient la clé `opp` avec la valeur `ACC`.

```
select "_raw" from SplunkTable where "search"='search opp=ACC'
```

## Connexion à SuccessFactors

1. Accédez à la page **Sources de données** en utilisant l'un de ces modules :

**Management Console :** Ouvrez Management Console à l'aide de l'URL : URL: `http://serveur.port/managementconsole`, où `serveur` est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et `port` est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Ressources > Sources de données**.

**Metadata Insights :** Ouvrez Metadata Insights à l'aide de l'URL : URL: `http://serveur.port/metadata-insights`, où `serveur` est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et `port` est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Sources de données**.

2. Cliquez sur le bouton **Ajouter** .
3. Dans le champ **Nom**, entrez un nom pour la nouvelle connexion. Ce nom est entièrement de votre choix.

**Remarque :** Une fois que vous enregistrez une connexion, vous ne pouvez plus modifier le nom.

4. Dans le champ **Type**, sélectionnez **SuccessFactors**.
5. Dans le **ID de société**, saisissez votre ID de société, qui identifie l'instance unique de votre société dans un centre de données SuccessFactors spécifique.
6. Dans le champ **URL du service**, saisissez l'URL du serveur SuccessFactors auquel vous souhaitez vous connecter. Cette URL est spécifique au centre de données international auquel l'ID de société est mappé.
7. Saisissez le nom d'utilisateur et le mot de passe de votre instance client SuccessFactors.
8. Pour tester la connexion, cliquez sur **Tester**.
9. Cliquez sur **Enregistrer**.

### Restrictions de SuccessFactors

1. Les opérations de traitement par lots peuvent être effectuées uniquement à l'aide des requêtes `upsert`. Par conséquent, les requêtes `insert` et `update` sont également effectuées sous forme de requêtes `upsert` dans les opérations de traitement par lots.
2. Il se peut que les propriétés de la colonne/table, telles qu'indiquées dans le schéma Physical Model d'une connexion SuccessFactors, ne fonctionnent pas comme prévu lors de l'utilisation de l'opération correspondante. Par exemple, une colonne qui a été marquée comme pouvant être mise à jour, peut lever une exception lorsque vous essayez de mettre à jour cette colonne.

### Entités et opérations prises en charge

Les entités sont de deux types :

- **Entité** : table représentant une entité métier.
- **Jointure** : mappage entre deux tables de type *Entité* : une table parent et n'importe laquelle de ses tables enfant.

**Remarque :** Les liens ne sont pas présents entre les tables dans le schéma Physical Model calculé à partir d'une connexion SuccessFactors. C'est parce que les clés étrangères ne figurent pas dans les tables SuccessFactors et que les jointures entre les tables sont indiquées par des tables de type *Jointure* dans la Spectrum™ Technology Platform.

Les fonctions de tables de jointure sont les suivantes :

- Le nom d'une table de *jointure* indique que les deux tables *Entité* ont été mappées entre elles.
- Chaque enregistrement d'une table de *jointure* contient la clé principale de l'entité parente et les colonnes de l'entité enfant respective. Par conséquent, la clé principale de l'entité parente particulière est mappée sur les détails de l'entité enfant. Par exemple, `User#HR` est une table de *jointure* dans laquelle `User` est l'entité parente et `Hr` est l'entité enfant. Cette jointure représente tous les utilisateurs et leurs représentants HR respectifs. La table de jointure `User#HR` utilise par conséquent la table parent `User` dont la clé principale `UserId` est mappée vers les colonnes de la table enfant `HR`, comme `hr_userId`, `hr_username` et `hr_email`.
- En cas de tables de *Jointure*, les fonctions `insert` et `update` s'utilisent de la même façon que la fonction `upsert`. C'est parce que les tables de *jointure* ne sont pas des entités réelles dans `SuccessFactors`, mais des mappages entre les entités et leurs propriétés de navigation ou leurs tables enfant.

Lors d'une opération `insert` ou `update` sur toute table de *jointure*, l'entité parente est mise à jour, tandis qu'un nouvel enregistrement est inséré dans la table enfant ou l'enregistrement existant est mis à jour d'après l'enregistrement parent.

**Remarque :** Lors de la mise à jour, le mappage entre un parent et un enfant est modifié. Il est également possible de modifier les attributs individuels de l'enfant selon ce qui est requis.

## Connexion à SugarCRM

1. Accédez à la page **Sources de données** en utilisant l'un de ces modules :

**Management Console :** Ouvrez Management Console à l'aide de l'URL : URL: `http://serveur.port/managementconsole`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Ressources > Sources de données**.

**Metadata Insights :** Ouvrez Metadata Insights à l'aide de l'URL : URL: `http://serveur.port/metadata-insights`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Sources de données**.

2. Cliquez sur le bouton **Ajouter** .

3. Dans le champ **Nom**, entrez un nom pour la nouvelle connexion. Ce nom est entièrement de votre choix.

**Remarque :** Une fois que vous enregistrez une connexion, vous ne pouvez plus modifier le nom.

4. Dans le champ **Type**, sélectionnez **SugarCRM**.
5. Saisissez vos nom d'utilisateur et votre mot de passe SugarCRM.
6. Dans le champ **URL**, saisissez l'URL du compte SugarCRM à utiliser pour cette connexion.
7. Saisissez l'**Id Client** et le **Secret Client** de votre compte SugarCRM.
8. Pour tester la connexion, cliquez sur **Tester**.
9. Cliquez sur **Enregistrer**.

### Restrictions de SugarCRM

1. Pour les deux requêtes `UPDATE` et `UPSERT`, une opération `UPSERT` est réalisée.
2. Il est possible que les colonnes **Annulable** et **Actualisable** des propriétés de la table, telles qu'affichées dans le **Schéma Physical Model** de la connexion, ne représentent pas le fonctionnement correct. Par exemple, une colonne non marquée comme actualisable peut lever une exception système lorsque vous essayez de la mettre à jour et, inversement, une colonne marquée comme annulable peut ne pas lever d'exception lors de sa mise à jour.
3. Lors d'une requête utilisant des jointures, vous devez utiliser un alias.

### Entités et opérations prises en charge

#### Opérations prises en charge

`LIKE` (son opération est limitée à la sélection des options commençant par la valeur spécifiée, comme dans l'instruction `WHERE name LIKE 's%'`, qui sélectionne tous les noms commençant par la lettre S de l'alphabet), `ISNULL`, `IS NOT NULL`, `IN`, `NOT IN`, `>`, `>=`, `<`, `<=`, `=`, `<>`, `AND`, `OR`

## Connexion à Oracle Eloqua

1. Accédez à la page **Sources de données** en utilisant l'un de ces modules :

**Management Console :** Ouvrez Management Console à l'aide de l'URL : URL: `http://serveur.port/managementconsole`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Ressources > Sources de données**.

**Metadata Insights :** Ouvrez Metadata Insights à l'aide de l'URL : URL: `http://serveur.port/metadata-insights`, où *serveur* est le nom du serveur ou l'adresse IP de votre serveur Spectrum™ Technology Platform et *port* est le port HTTP utilisé par Spectrum™ Technology Platform.

**Remarque :** Par défaut, le port HTTP est 8080.

Accédez à **Sources de données**.

2. Cliquez sur le bouton **Ajouter** .
3. Dans le champ **Nom**, entrez un nom pour la nouvelle connexion. Ce nom est entièrement de votre choix.

**Remarque :** Une fois que vous enregistrez une connexion, vous ne pouvez plus modifier le nom.

4. Dans le champ **Type**, cliquez sur **Eloqua**.
5. Saisissez le **Nom du site** qui est identique au nom de la société.
6. Saisissez le nom d'utilisateur dans le champ **Nom d'utilisateur**.
7. Saisissez le mot de passe dans le champ **Mot de passe**.
8. Cliquez sur **Test** pour tester la connexion.
9. Cliquez sur **Enregistrer**.

## Opérations spéciales

1. Utilisez cette requête de jointure pour récupérer les contacts d'une Liste de contacts :

```
select * from Contacts inner join ContactListMembers on
Contacts.Eloqua_Contact_ID = ContactListMembers.Contact_Id where
ContactListMembers.ContactList_Id = '<id>'
```

Utilisez cette requête de jointure pour récupérer les contacts d'un Segment de contacts :

```
select * from Contacts inner join ContactSegmentMembers on
Contacts.Eloqua_Contact_ID = ContactSegmentMembers.Contact_Id where
ContactSegmentMembers.Contactlist_Id = '<id>'
```

2. Utilisez cette instruction pour insérer des contacts dans une Liste de contacts :

```
insert into ContactListMembers (ContactList_ID,Contact_ID) values
('<contactlist_id>','<contact_id>')
```

3. Utilisez cette instruction pour supprimer des contacts d'une Liste de contacts :

```
delete from ContactListMembers where ContactList_ID =
'<contactlist_id>' and Contact_ID = '<contact_id>'
```

## Restrictions

Voici les restrictions :

### 1. Créer/Mettre à jour :

- a. Insérer/Upsert échoue si des colonnes Non nul sont vides ou n'existent pas.
- b. Insérer/Upsert échoue si les valeurs des colonnes Unique ne sont pas uniques pour un lot donné.
- c. Afin d'éviter une exception de retour, conservez la valeur de **Nombre de lots à valider** sur 1.

### 2. Read : Pour les entités personnalisées, l'opération Select s'applique uniquement aux jointures avec l'entité Contacts.

### 3. Filtrer :

- a. Les filtres pris en charge sont =, ! =, >, <, > =, < =
- b. Les opérateurs de condition IN et NOT IN ne sont pas pris en charge lors de la fourniture de plus d'une valeur.
- c. Jointures n'est pas pris en charge entre les entités.
- d. Les opérateurs conditionnels d'opérations AND et OR sont pris en charge uniquement pour les entités Comptes et Contacts.
- e. Seul l'opérateur conditionnel AND est pris en charge. Il ne fonctionne pas pour le reste des entités de la colonne.
- f. Le filtre = ne fonctionne pas toujours sur les champs présentant un type de données Horodatage.

## Entités et opérations prises en charge

Le tableau suivant décrit les entités et les opérations prises en charge pour Oracle Eloqua.

Ces entités sont prises en charge :

- **Entité** : indique une table représentant une entité métier.
- **Activité** : indique une table représentant une entité métier dont les données sont générées en fonction d'une activité.
- **Entité personnalisée** : indique les entités qui sont utilisées dans le cadre d'opérations spéciales fournies avec le Connecteur.

Ce tableau répertorie les entités et les opérations prises en charge pour celles-ci.

Nom de l'entité	Créer	Lire	Mettre à jour	Supprimer	Prise en charge par lots	Taille de lot maximale
-----------------	-------	------	---------------	-----------	--------------------------	------------------------

Comptes	X	X	X	X	Insérer/Mettre à jour*	1000
Groupes de comptes		X				
Campagne		X				
Contacts	X	X	X	X	Insérer/Mettre à jour*	1000
Liste de contacts	X	X	X	X		
Segment de contacts	X	X	X	X		
E-mails		X				
Dossiers d'e-mails		X				
Groupes d'e-mails		X				
Microsites		X				
Utilisateurs		X				
Visiteurs		X				
<b>Activité</b>						
Ouverture d'e-mails		X				
EmailClickthrough		X				
Envoi d'e-mails		X				
S'abonner		X				
Se désinscrire		X				

Nom de l'entité	Créer	Lire	Mettre à jour	Supprimer	Prise en charge par lots	Taille de lot maximale
Rebond		X				
WebVisit		X				
PageView		X				
FormSubmit		X				
<b>Entités personnalisées</b>						
ContactListMembers	X	X		X	Insérer/Supprimer	1000
ContactSegmentMembers		X				

L'opération Mettre à jour fonctionne comme Upsert.

## Prise en charge de la compression pour les serveurs de fichiers Cloud

Les serveurs de fichiers Amazon S3, Google cloud storage et MS Azure Blobstore prennent en charge les formats compressés gzip (.gz) et zip (.zip).

Spectrum™ Technology Platform gère la compression et la décompression des fichiers écrits et lus vers/depuis les serveurs de fichiers.

**Remarque :** Vous pouvez utiliser le même serveur de fichiers pour gérer la lecture et l'écriture de fichiers normales et la compression et la décompression de fichiers.

### Lecture d'un fichier au format compressé

Lors de la lecture d'un fichier à partir du serveur, son format de compression est calculé à partir de la propriété de clé de métadonnées `Content-Encoding` provenant du serveur.

### Écriture d'un fichier au format compressé

Lors de l'écriture d'un fichier dans un serveur, indiquez le format de compression requis : `.gz` ou `.zip`. Le fichier est compressé en fonction de l'extension de compression spécifiée.

La propriété de clé de métadonnées `Content-Encoding` est également définie en fonction du format de compression sélectionné. Cette valeur de propriété est transmise au serveur de fichiers Cloud lors de l'écriture du fichier.

## Suppression d'une connexion

Vous pouvez supprimer la connexion en utilisant l'un de ces modules :

- Management Console
- Metadata Insights

1. Accédez à la page **Sources de données** du module requis.
  - Dans Management Console, cliquez sur **Ressources** > **Sources de données**.
  - Dans Metadata Insights, cliquez sur **Sources de données**.
2. Cochez la case en regard de la connexion que vous souhaitez supprimer, puis cliquez sur le bouton **Supprimer** .

# 3 - Renseignement de l'entrepôt de données

## In this section

---

Préparation de vos données	71
Renseignement d'une table de dimensions temporelles	72
Renseignement d'une table de dimensions	73
Renseignement d'une table de faits	75
Ajout d'un horodateur aux enregistrements d'un entrepôt de données	80

## Préparation de vos données

Avant de renseigner votre entrepôt de données, vous devez vous assurer que vos données sont de bonne qualité et qu'elles présentent tous les attributs nécessaires pour fournir des informations significatives aux utilisateurs commerciaux. Pour ce faire, une approche commune consiste à utiliser un magasin de données opérationnelles (Operational Data Store ou ODS). Un magasin ODS est une base de données dans laquelle vous pouvez réaliser des opérations relatives à la qualité des données afin de préparer le chargement de vos données dans l'entrepôt de données. Spectrum™ Technology Platform offre différentes fonctionnalités déployables dans un ODS pour améliorer la qualité de vos données ainsi que pour y ajouter des données spatiales, démographiques ou autres. Si vous ne disposez actuellement pas des fonctionnalités suivantes, fournies sous licence, contactez Pitney Bowes pour plus d'informations.

### *Parsing, normalisation des noms, et validation des noms.*

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/extraction. Le module Universal Name fournit ces fonctionnalités.

### *Déduplication et consolidation*

L'identification d'entité unique vous permet de consolider des enregistrements, d'éliminer des doublons et de développer des enregistrements de type Best of Breed (Best of Breed). Un enregistrement de type "Best of Breed" est un enregistrement composite élaboré à l'aide de données provenant d'autres enregistrements. Le module Advanced Matching et le module Data Normalization fournissent ces fonctionnalités.

### *Validation d'adresse*

La validation d'adresse applique des règles en provenance des autorités postales appropriées afin d'écrire une adresse sous sa forme normalisée et même la valider comme adresse livrable. La validation d'adresse peut vous aider à être éligible à certaines réductions sur les tarifs postaux et améliorer la délivrabilité de vos courriers. Le module Universal Addressing et le module Address Now fournissent ces fonctionnalités.

### *Géocodage*

Le géocodage est la procédure qui consiste à déterminer des coordonnées géographiques (latitude et longitude) à partir d'une adresse complète ou partielle. Le géocodage peut être utilisé pour la création de cartes, mais ce n'est pas sa seule application. Les données de localisation sous-jacentes peuvent vous servir dans vos décisions professionnelles. En renversant le processus, vous pouvez

saisir un géocode (un point représenté par ses coordonnées de longitude et de latitude) et recevoir des informations d'adresse se rapportant à ce géocode. Le module Enterprise Geocoding fournit ces fonctionnalités.

### *Location Intelligence*

La Location Intelligence (localisation avancée) génère de nouvelles informations sur vos données en évaluant, analysant et modélisant des liens entre territoires. En utilisant le traitement de Location Intelligence, vous pouvez vérifier des lieux et transformer des informations en renseignements commerciaux précieux. Le module Location Intelligence fournit ces fonctionnalités.

### *Assignment de juridiction fiscale*

L'assignation de juridiction fiscale prend une adresse et en détermine les juridictions fiscales qui s'appliquent à l'emplacement correspondant à l'adresse en question. Une assignation la plus précise possible des juridictions fiscales peut réduire vos risques financiers et vos responsabilités en matière de réglementation.

Le logiciel Spectrum™ Technology Platform de Pitney Bowes intègre les toutes dernières circonscriptions juridiques avec l'adresse postale exacte des enregistrements de votre client, vous permettant d'ajouter à vos enregistrements les informations sur l'État, le pays, le canton, la municipalité et des informations particulières à la juridiction fiscale. Quelques exemples d'utilisation de l'assignation de juridiction fiscale :

- Taxe sur les ventes et la consommation
- Impôt mobilier
- Taxe sur les primes d'assurance

Le module Enterprise Tax Module fournit ces fonctionnalités.

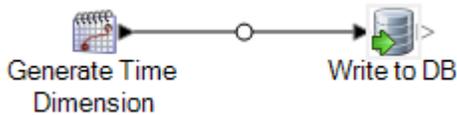
## Renseignement d'une table de dimensions temporelles

Une table de dimensions temporelles est une table d'une base de données qui permet d'analyser des données historiques sans avoir recours à des calculs SQL complexes. Par exemple, vous pouvez analyser vos données par jour ouvré par rapport aux jours fériés, par jour de la semaine par rapport aux weekends, par exercice fiscal ou par événement spécial.

La procédure suivante explique comment utiliser Spectrum™ Technology Platform pour renseigner une table de dimensions temporelles de votre entrepôt de données.

**Remarque :** Avant de commencer la procédure, vous devez avoir défini les connexions à l'entrepôt de données dans lequel vous souhaitez créer une table de dimensions temporelles. Si vous n'avez pas défini la connexion nécessaire, reportez-vous à la section [Data Source Connections](#) à la page 14.

1. Dans Enterprise Designer, sélectionnez **Fichier > Nouveau > Flux de données > Job**.
2. Glissez le stage Generate Time Dimension sur le canevas.
3. Glissez un stage Write to DB sur le canevas et connectez-lui le stage Generate Time Dimension. Le flux de données prend maintenant la forme suivante :



4. Double-cliquez sur le stage Generate Time Dimension et configurez-le pour qu'il produise les dimensions temporelles de votre choix. Pour plus d'informations, reportez-vous à la section [Generate Time Dimension](#) à la page 115.
 

**Remarque :** Si le grain est un jour ou plus, le jour julien est généralement utilisé comme valeur de clé d'une table de dimensions temporelles. Si le grain est inférieur à un jour, vous pouvez générer une clé distincte en ajoutant un stage Unique ID Generator au flux de données. Si vous utilisez le jour julien comme clé, configurez Generate Time Dimension pour qu'il produise une colonne de nombres entiers pour les valeurs de jour julien et une colonne de type de données date ou datetime pour les valeurs de date.
5. Double-cliquez sur le stage Write to DB du canevas et configurez-le pour qu'il pointe vers la base de données et la table où vous souhaitez créer la table de dimensions temporelles. Pour plus d'informations sur la configuration de Write to DB, reportez-vous à la section [Write to DB](#) à la page 245.
6. Pour afficher un aperçu des valeurs des dimensions temporelles avant de les écrire dans la table de dimensions temporelles, procédez comme suit :
  - a) Cliquez avec le bouton droit de la souris sur le canal connectant le stage Generate Time Dimension au stage Write to DB, puis sélectionnez **Ajouter un point d'inspection**.
  - b) Sélectionnez **Exécuter > Inspecter le flux actif**.  
Le volet d'inspection apparaît au bas de la fenêtre Enterprise Designer et affiche les données qui vont être écrites dans la table de dimensions temporelles. Si nécessaire, vous pouvez modifier le stage Generate Time Dimension et réexécuter le processus d'inspection pour afficher les résultats de vos modifications.
7. Lorsque vous êtes satisfait(e) du flux de données, sélectionnez **Exécuter > Exécuter le flux actif** pour exécuter le flux de données et renseigner la table de dimensions temporelles.

## Renseignement d'une table de dimensions

Les tables de dimensions font partie d'un schéma en étoile. Elles contiennent des informations détaillées sur les colonnes de la table de faits. Les tables de dimensions comportent des attributs et une clé primaire en une seule partie qui joint la table de dimensions à la table de faits. La clé

primaire en une seule partie vous permet de parcourir rapidement une seule table de dimensions. Le parcours d'une table de dimensions peut vous aider à déterminer la meilleure façon de lancer une requête sur la table de faits.

La procédure suivante explique comment utiliser Spectrum™ Technology Platform pour renseigner pour la première fois une table de dimensions de votre entrepôt de données.

**Remarque :** Si vous utilisez une base de données, un serveur de fichiers ou un Service Web comme source de données, avant de commencer cette procédure, vous devez avoir défini les connexions aux ressources externes que vous souhaitez utiliser comme source pour la table de dimensions. Vous devez également définir une connexion à l'entrepôt de données dans lequel vous souhaitez créer une table de dimensions. Si vous n'avez pas défini les connexions nécessaires, reportez-vous à la section [Data Source Connections](#) à la page 14.

1. Dans votre entrepôt de données, créez la table que vous souhaitez utiliser comme table de dimensions.
2. Dans Management Console, créez des connexions à votre source de données et à l'entrepôt de données.
  - Pour vous connecter à une base de données, reportez-vous à la section [Connexion à une base de données JDBC](#) à la page 33.
  - Pour vous connecter à un serveur de fichiers, reportez-vous à la section [Connexion à un serveur FTP](#) à la page 25.
3. Dans Enterprise Designer, sélectionnez **Fichier > Nouveau > Flux de données > Job**.
4. Glissez le stage source sur le canevas.
  - Pour utiliser des données d'une base de données pour renseigner la table, glissez le stage **Read from DB** sur le canevas.
  - Pour utiliser des données d'un fichier plat pour renseigner la table, glissez le stage **Read from File** sur le canevas.
  - Pour utiliser des données d'un fichier de format variable pour renseigner la table, glissez le stage **Read from Variable Format File** sur le canevas.
  - Pour utiliser des données d'un fichier XML pour renseigner la table, glissez le stage **Read from XML** sur le canevas.
5. Double-cliquez sur le stage source que venez de placer sur le canevas et configurez-le pour qu'il pointe vers la source des données avec lesquelles vous souhaitez renseigner la table de dimensions.
  - Pour plus d'informations sur la configuration de Read from DB, reportez-vous à la section [Read From DB](#) à la page 139.
  - Pour plus d'informations sur la configuration de Read from File, reportez-vous à la section [Read from file](#) à la page 148.
  - Pour plus d'informations sur la configuration de Read from Variable Format File, reportez-vous à la section [Read from Variable Format File](#) à la page 197.

- Pour plus d'informations sur la configuration de Read from XML, reportez-vous à la section [Read From XML](#) à la page 212.
6. Glissez un stage Unique ID Generator sur le canevas et connectez-lui le stage source. Par exemple, si vous utilisez Read from DB comme stage source, connectez Read from DB à Unique ID Generator.
  7. Double-cliquez sur le stage Unique ID Generator du canevas et configurez-le de sorte à créer une clé de substitution.
 

**Remarque** : En règle générale, la clé du système opérationnel n'est pas utilisée comme clé primaire pour une dimension de l'entrepôt. Cela permet de maintenir la cohérence historique, car une valeur de clé peut changer dans le système d'exploitation.
  8. Glissez un stage Write to DB sur le canevas et connectez-lui Unique ID Generator.
  9. Double-cliquez sur le stage Write to DB du canevas et configurez-le pour qu'il pointe vers la base de données et la table de dimensions que vous souhaitez renseigner. Pour plus d'informations sur la configuration de Write to DB, reportez-vous à la section [Write to DB](#) à la page 245.
  10. Sélectionnez **Fichier > Enregistrer** et enregistrez le flux de données.
  11. Pour exécuter le flux de données et renseigner la table de dimensions, sélectionnez **Exécuter > Run Current Flow**.

## Renseignement d'une table de faits

Une fois que vous avez renseigné les tables de dimensions de votre entrepôt de données, vous pouvez renseigner la table de faits. Renseignez une table de faits à l'aide de mesures numériques provenant de tables de la base de données OLTP.

**Important** : Avant de renseigner la table de faits, vous devez renseigner les tables de dimensions.

La procédure suivante explique comment utiliser Spectrum™ Technology Platform pour renseigner une table de faits de votre entrepôt de données. Dans cette procédure, vous allez créer un flux de données lisant des données sources d'une table de votre base de données source, remplaçant les clés naturelles des tables sources par des clés de substitution des tables de dimensions, puis chargeant l'enregistrement mis à jour contenant les clés de substitution et les données de faits des tables sources dans la table de faits.

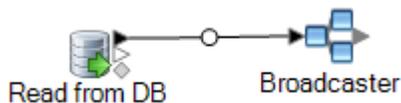
1. Dans Management Console, créez des connexions à votre source de données et à l'entrepôt de données.
  - Pour vous connecter à une base de données, reportez-vous à la section [Connexion à une base de données JDBC](#) à la page 33.
  - Pour vous connecter à un serveur de fichiers, reportez-vous à la section [Connexion à un serveur FTP](#) à la page 25.

2. Dans Enterprise Designer, sélectionnez **Fichier > Nouveau > Flux de données > Job**.
3. Suivant la source des données que vous souhaitez écrire dans la table de faits, glissez le stage approprié sur le canevas.
  - Pour utiliser des données d'une base de données pour renseigner la table, glissez le stage **Read from DB** sur le canevas.
  - Pour utiliser des données d'un fichier plat pour renseigner la table, glissez le stage **Read from File** sur le canevas.
  - Pour utiliser des données d'un fichier de format variable pour renseigner la table, glissez le stage **Read from Variable Format File** sur le canevas.
  - Pour utiliser des données d'un fichier XML pour renseigner la table, glissez le stage **Read from XML** sur le canevas.
4. Double-cliquez sur le stage source que venez de placer sur le canevas et configurez-le pour qu'il pointe vers la source des données avec lesquelles vous souhaitez renseigner la table de faits.
  - Pour plus d'informations sur la configuration de Read from DB, reportez-vous à la section [Read From DB](#) à la page 139.
  - Pour plus d'informations sur la configuration de Read from File, reportez-vous à la section [Read from file](#) à la page 148.
  - Pour plus d'informations sur la configuration de Read from Variable Format File, reportez-vous à la section [Read from Variable Format File](#) à la page 197.
  - Pour plus d'informations sur la configuration de Read from XML, reportez-vous à la section [Read From XML](#) à la page 212.

**Remarque :** En règle générale, un flux de données qui renseigne une table de fait lit les données d'une base de données et non d'un fichier. Étant donné qu'il s'agit du scénario le plus courant, les exemples du reste de cette procédure utilisent Read from DB.

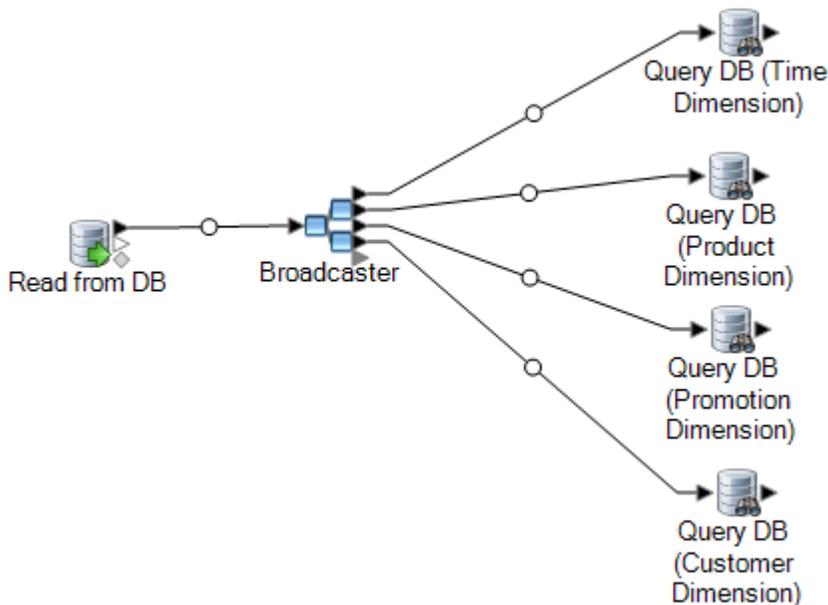
5. Glissez un stage Broadcaster sur le canevas et connectez-lui le stage source.

Votre flux de données prend maintenant la forme suivante :



6. Glissez un stage Query DB sur le canevas pour chaque table de dimensions de votre entrepôt de données et connectez-les au stage Broadcaster.

Par exemple, si votre entrepôt de données comporte quatre tables de dimensions, glissez quatre stages Query DB sur le canevas. Votre flux de données prend maintenant la forme suivante :



Les stages Query DB sont utilisés pour rechercher la clé de substitution pour chaque dimension utilisant la clé naturelle de la source de données. La clé de substitution vient ensuite remplacer la clé naturelle dans chaque enregistrement chargé dans la table de faits.

**Conseil :** Pour connaître facilement la table interrogée par chaque stage, vous pouvez modifier le nom du stage.

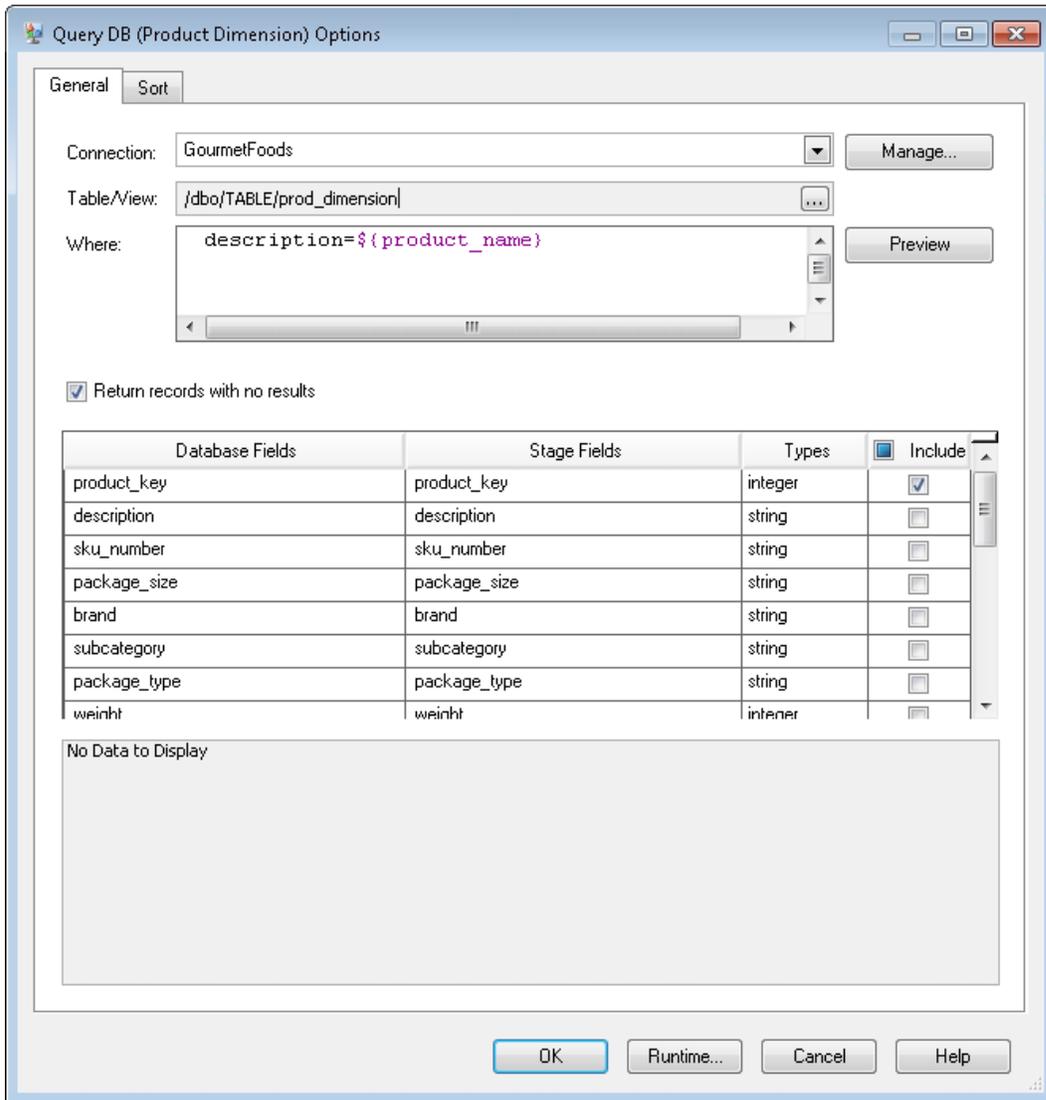
7. Configurez chaque stage Query DB de sorte qu'il recherche la clé de substitution pour chaque clé naturelle de la source de données. Pour ce faire, procédez comme suit :
  - a) Dans le champ **Connexion**, indiquez la connexion à l'entrepôt de données.
  - b) Dans le champ **Table/Vue**, sélectionnez la table de dimensions que vous souhaitez que ce stage interroge.
  - c) Dans le champ **Où**, écrivez une instruction `WHERE` qui recherche la clé de substitution en fonction de la valeur champ de flux de données approprié.

L'exemple suivant recherche la clé de substitution d'un produit en trouvant l'enregistrement de la table de dimensions dont la valeur de la colonne `description` correspond à la valeur du champ `product_name` de la source de données.

```
description=${product_name}
```

- d) Dans la colonne **Inclure**, sélectionnez la colonne de base de données contenant la clé de substitution.

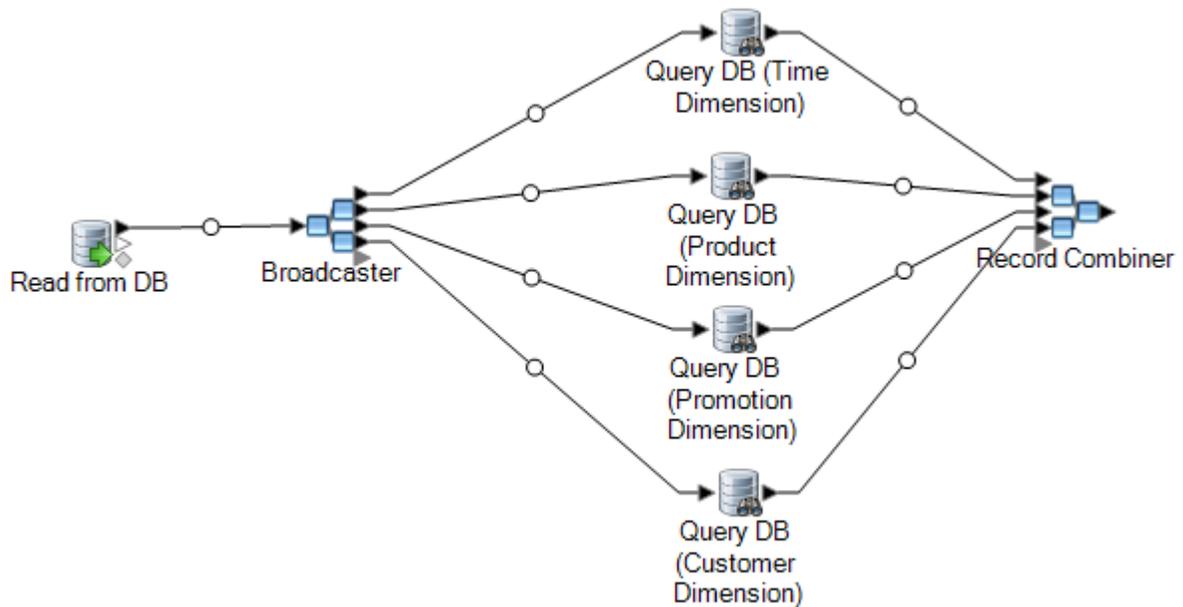
Par exemple, un stage Query DB qui recherche la clé de substitution d'un nom de produit ressemble à l'exemple suivant :



Dans cet exemple, la requête recherche la clé de produit en trouvant l'enregistrement de la table `prod_dimension` dont la valeur de la colonne `description` correspond à celle du champ de flux de données `product_name`. Le stage renvoie le champ `product_key` de la table et l'ajoute au flux de données, comme indiqué par la case cochée de la colonne **Include**.

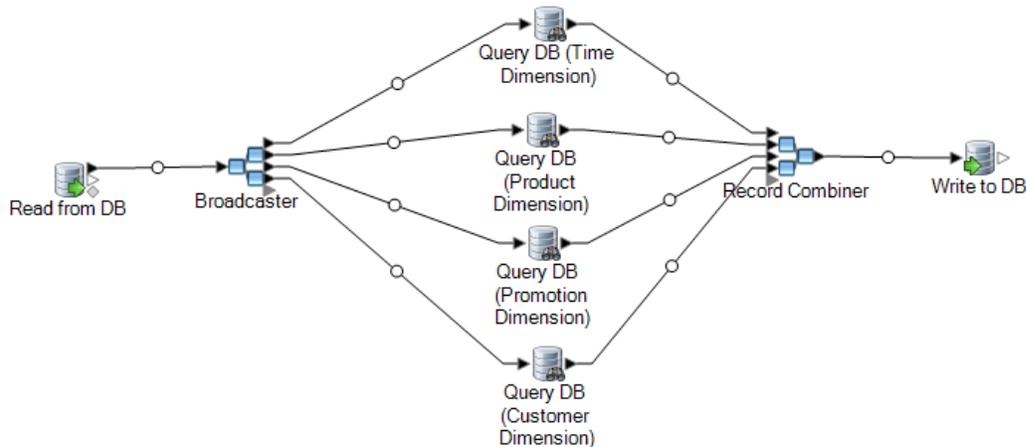
- Glissez un stage Record Combiner sur le canevas et connectez-lui l'ensemble des stages Query DB.

Votre flux de données prend maintenant la forme suivante :



9. Glissez un stage Write to DB sur le canevas et connectez-le au stage Record Combiner.

Votre flux de données prend maintenant la forme suivante :



10. Configurez le stage Write to DB pour qu'il écrive les enregistrements dans la table de faits. Pour ce faire, procédez comme suit :
- Dans le champ **Connexion**, indiquez la connexion à l'entrepôt de données.
  - Dans le champ **Table/Vue**, sélectionnez la table de faits que vous souhaitez que ce stage interroge. Si la table de faits n'existe pas déjà dans l'entrepôt de données, cliquez sur **Créer une table** pour la créer.
  - Pour chaque champ que vous souhaitez écrire dans la table de faits, cochez la case de la colonne **Inclure**.
  - Dans l'onglet **Exécution**, notez que, par défaut, l'option **Insérer** est sélectionnée comme mode d'écriture. En règle générale, le renseignement de la table de faits est exécuté en mode d'insertion ; vous pouvez donc laisser cette option sélectionnée.

11. Enregistrez et exécutez votre flux de données.

### Exemple de remplacement de données sources par des clés de la table de dimensions

Prenons l'enregistrement suivant :

```
March 28 2013,Parsley Garlic Pasta,Mile High Gourmet  
Market,78.35
```

Cet exemple contient un champ de date, suivi d'un nom de produit (Parsley Garlic Pasta), d'un client (Mile High Gourmet Market) et d'un montant (78.25). L'entrepôt de données comporte des tables de dimensions pour la date, le nom de produit et le client. Les clés naturelles de l'enregistrement doivent donc être remplacées par les clés de substitution des tables de dimensions. Pour ce faire, le flux de données doit comporter trois stages Query DB : un premier stage qui recherche la clé de substitution pour la date, un deuxième pour le nom de produit et un troisième pour le client.

Chaque stage Query DB comporte une instruction `WHERE` qui recherche la clé de substitution respective.

Suite à ces recherches, l'enregistrement, une fois écrit dans la table de faits, peut prendre la forme suivante :

```
711,1,15,78.35
```

Notez que les clés naturelles de la date, du nom de produit et du client ont été remplacées par les clés de substitution.

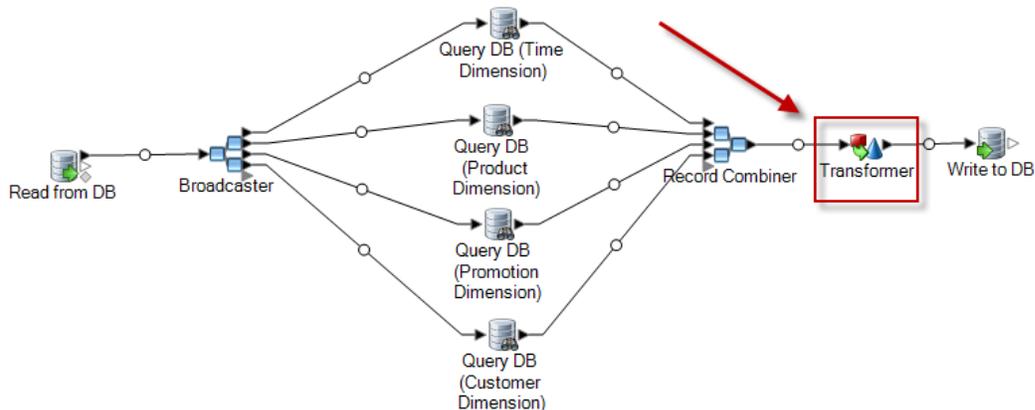
## Ajout d'un horodateur aux enregistrements d'un entrepôt de données

Une manière pratique de garantir la qualité des données est de repérer les enregistrements de l'entrepôt de données en fonction de leur date de chargement. Si la procédure de chargement ne se termine pas correctement ou si vous remarquez des problèmes une fois les données chargées, une colonne d'horodateurs facilite l'identification des enregistrements affectés. Vous pouvez ensuite supprimer tous les enregistrements traités lors d'une phase donnée, revenir à l'état précédant le chargement et résoudre tout problème avant de tenter de charger de nouveau les données. Vous pouvez horodater l'opération de chargement en ajoutant une colonne supplémentaire, comme `load_date`, à votre table de faits via le stage SQL Command.

Lors du renseignement ou de la mise à jour d'un entrepôt de données, pour ajouter un horodateur au dataflow, procédez comme suit :

1. Dans Enterprise Designer, ouvrez le dataflow renseignant ou mettant à jour l'entrepôt de données :
2. Glissez un stage Transformer jusqu'au canevas et connectez-le au dataflow juste avant le stage Write to DB.

Par exemple :



3. Double-cliquez sur le stage Transformer.
4. Cliquez sur **Ajouter**.
5. Sous **Général**, sélectionnez **Personnalisé**.
6. Dans le champ **Nom de la transformation personnalisée**, saisissez un nom pour cette transformation. Vous pouvez choisir n'importe quel nom. Par exemple, Ajouter horodateur.
7. Dans le champ **Script personnalisé**, saisissez ceci :

```
data['<timestamp field>']=currentDateTime()
```

Où <timestamp field> est le nom du dataflow de données où vous souhaitez placer l'horodateur.

Par exemple, si vous souhaitez placer l'horodateur dans un champ de dataflow nommé Timestamp, votre script personnalisé prend la forme suivante :

```
data['Timestamp']=currentDateTime()
```

8. Cliquez sur le bouton **Ajouter** au bas de la fenêtre.
9. Cliquez sur **Fermer**.
10. Cliquez sur **OK** pour fermer la fenêtre **Options Transformer**.

Le dataflow ajoute l'heure actuelle à un champ de chaque enregistrement, à l'aide d'un horodateur figurant dans l'entrepôt de données qui affiche l'heure de chargement de chaque enregistrement.

# 4 - Mise à jour de l'entrepôt de données

## In this section

---

Définition d'un calendrier de mise à jour d'un entrepôt de données	83
Mise à jour d'une table de faits	84
Utilisation d'un cache global pour les requêtes	89
Utilisation d'un cache local pour les requêtes	90

# Définition d'un calendrier de mise à jour d'un entrepôt de données

Vous pouvez programmer les flux de données Spectrum™ Technology Platform de sorte à extraire et à transformer les données de la structure normalisée de la source de données dans la structure de schéma en étoile de l'entrepôt de données. La programmation de flux de données est utile, car la plupart des opérations de chargement requièrent des ressources système qui ne sont pas disponibles au cours de la journée.

Au moment de décider du calendrier de mise à jour, prenez en compte les points suivants :

- Fréquence
- Séquence
- Dépendances

## Fréquence

Vous devez programmer les flux de données à exécuter en fonction du grain de la table de faits la plus détaillée. Par exemple :

- Si le grain de la table de faits est quotidien, programmez son flux de renseignement une fois par jour.
- Si son grain est mensuel, programmez son flux de renseignement une fois par mois, pas plus souvent, car les utilisateurs utilisent uniquement les données de mois passés complets.

La plupart des flux de données de renseignement traitent de grandes quantités de données. Par conséquent, programmez-les de sorte qu'ils soient exécutés lorsque l'utilisation du serveur Spectrum™ Technology Platform, des bases de données source et de l'entrepôt de données et du réseau est minimale.

Lors du chargement initial, renseignez l'ensemble des tables de dimensions et de faits. Après le chargement initial, actualisez les tables en fonction des ajouts ou des modifications. En règle générale, les tables de faits sont actualisées plus fréquemment que les tables de dimensions. Plusieurs raisons à cela :

- Les tables de dimensions sont généralement statiques, sauf en cas de modification ou d'ajout d'un attribut de la source.
- Les données des tables de faits d'une base de données d'aide à la prise de décisions sont généralement historiques et requièrent des mises à jour et des ajouts réguliers pour rester d'actualité. Le chargement initial et la plupart des chargements incrémentiels affectent les tables de faits.

## Séquence

Il existe des dépendances entre les données des bases de données de l'entrepôt de données, c'est pourquoi, avant de définir le calendrier d'exécution, vous devez déterminer la séquence d'exécution des flux de données de renseignement.

Renseignez les tables de dimensions avant les tables de faits, car chaque enregistrement et chaque clé de dimension doivent exister avant de pouvoir renseigner la table de faits associée. Cette restriction constitue une fonction de la relation clé primaire-clé étrangère entre les tables de dimensions et de faits d'un schéma en étoile.

Actualisez les tables de niveau de base avant de renseigner les tables agrégées de votre base de données d'aide à la prise de décisions. Cette séquence garantit la synchronisation des tables de niveau de base et des tables agrégées.

L'ordre correct d'exécution des flux de données de renseignement est le suivant :

1. Plans de tables de dimensions de niveau de base
2. Plans de tables de faits de niveau de base
3. Plans de tables de dimensions agrégées
4. Plans de tables de faits agrégées

## Dépendances

Si plusieurs flux de données de renseignement doivent être exécutés dans un ordre précis, ou si le temps d'exécution des flux de données est difficile à prévoir, vous pouvez créer des dépendances de flux de données. Un flux de données est exécuté uniquement si certaines conditions sont remplies, par exemple, si le flux de données précédent est terminé, ou s'il a échoué.

Pour créer des dépendances de flux de données, créez un flux de processus dans Enterprise Designer. Pour plus d'informations sur les flux de processus, reportez-vous au *Guide de Dataflow Designer*.

# Mise à jour d'une table de faits

Cette procédure explique comment créer un flux de données qui lit les données d'une base de données ou d'un fichier source et utilise ces données pour mettre à jour une table de faits de votre entrepôt de données.

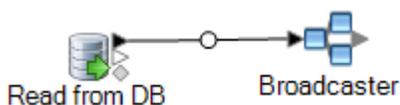
1. Dans Enterprise Designer, sélectionnez **Fichier > Nouveau > Flux de données > Job**.
2. Suivant la source des données que vous souhaitez écrire dans la table de faits, glissez le stage approprié sur le canevas.
  - Pour utiliser des données d'une base de données pour renseigner la table, glissez le stage **Read from DB** sur le canevas.

- Pour utiliser des données d'un fichier plat pour renseigner la table, glissez le stage **Read from File** sur le canevas.
- Pour utiliser des données d'un fichier de format variable pour renseigner la table, glissez le stage **Read from Variable Format File** sur le canevas.
- Pour utiliser des données d'un fichier XML pour renseigner la table, glissez le stage **Read from XML** sur le canevas.

**Remarque :** Si vous souhaitez lire les données d'un fichier et non d'une base de données, assurez-vous que le fichier contient uniquement les nouveaux enregistrements que vous souhaitez ajouter à la table de faits et qu'il ne contient pas d'enregistrements existant déjà dans la table de faits. Si vous souhaitez lire les données d'une table de données, vous devez définir une requête pour filtrer ultérieurement les enregistrements au cours de cette procédure

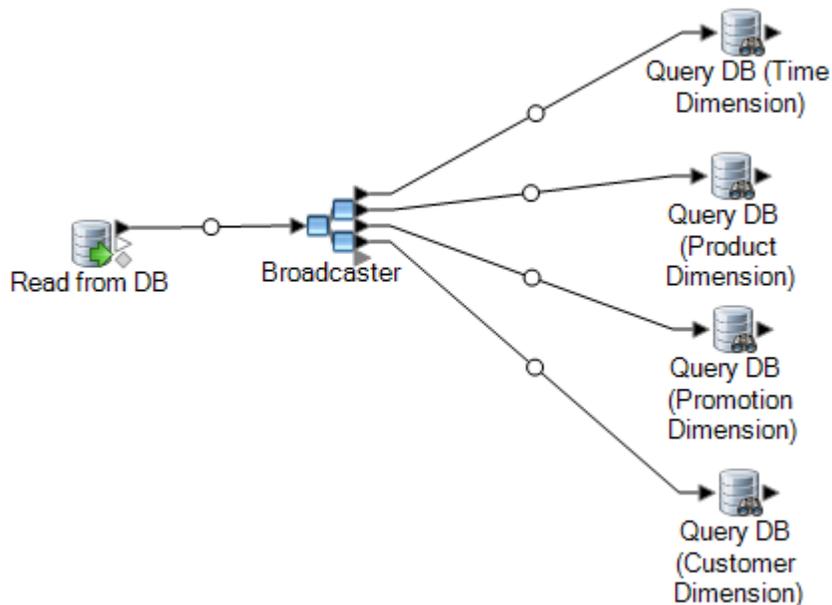
3. Double-cliquez sur le stage source que venez de placer sur le canevas et configurez-le pour qu'il pointe vers la source des données avec lesquelles vous souhaitez renseigner la table de faits.
  - Pour plus d'informations sur la configuration de Read from DB, reportez-vous à la section [Read From DB](#) à la page 139.
  - Pour plus d'informations sur la configuration de Read from File, reportez-vous à la section [Read from file](#) à la page 148.
  - Pour plus d'informations sur la configuration de Read from Variable Format File, reportez-vous à la section [Read from Variable Format File](#) à la page 197.
  - Pour plus d'informations sur la configuration de Read from XML, reportez-vous à la section [Read From XML](#) à la page 212.
4. Si vous souhaitez lire les données d'une base de données, filtrez les enregistrements de sorte que seuls les enregistrements nouveaux soient ajoutés à la table de faits. Pour ce faire, définissez l'instruction SQL SELECT de sorte qu'elle lise uniquement les enregistrements qui ont été modifiés depuis la dernière mise à jour de la table de faits.
5. Glissez un stage Broadcaster sur le canevas et connectez-lui le stage source.

Votre flux de données prend maintenant la forme suivante :



6. Glissez un stage Query DB sur le canevas pour chaque table de dimensions de votre entrepôt de données et connectez-les au stage Broadcaster.

Par exemple, si votre entrepôt de données comporte quatre tables de dimensions, glissez quatre stages Query DB sur le canevas. Votre flux de données prend maintenant la forme suivante :



Les stages Query DB sont utilisés pour rechercher la clé de substitution pour chaque dimension utilisant la clé naturelle de la source de données. La clé de substitution vient ensuite remplacer la clé naturelle dans chaque enregistrement chargé dans la table de faits.

**Conseil :** Pour connaître facilement la table interrogée par chaque stage, vous pouvez modifier le nom du stage.

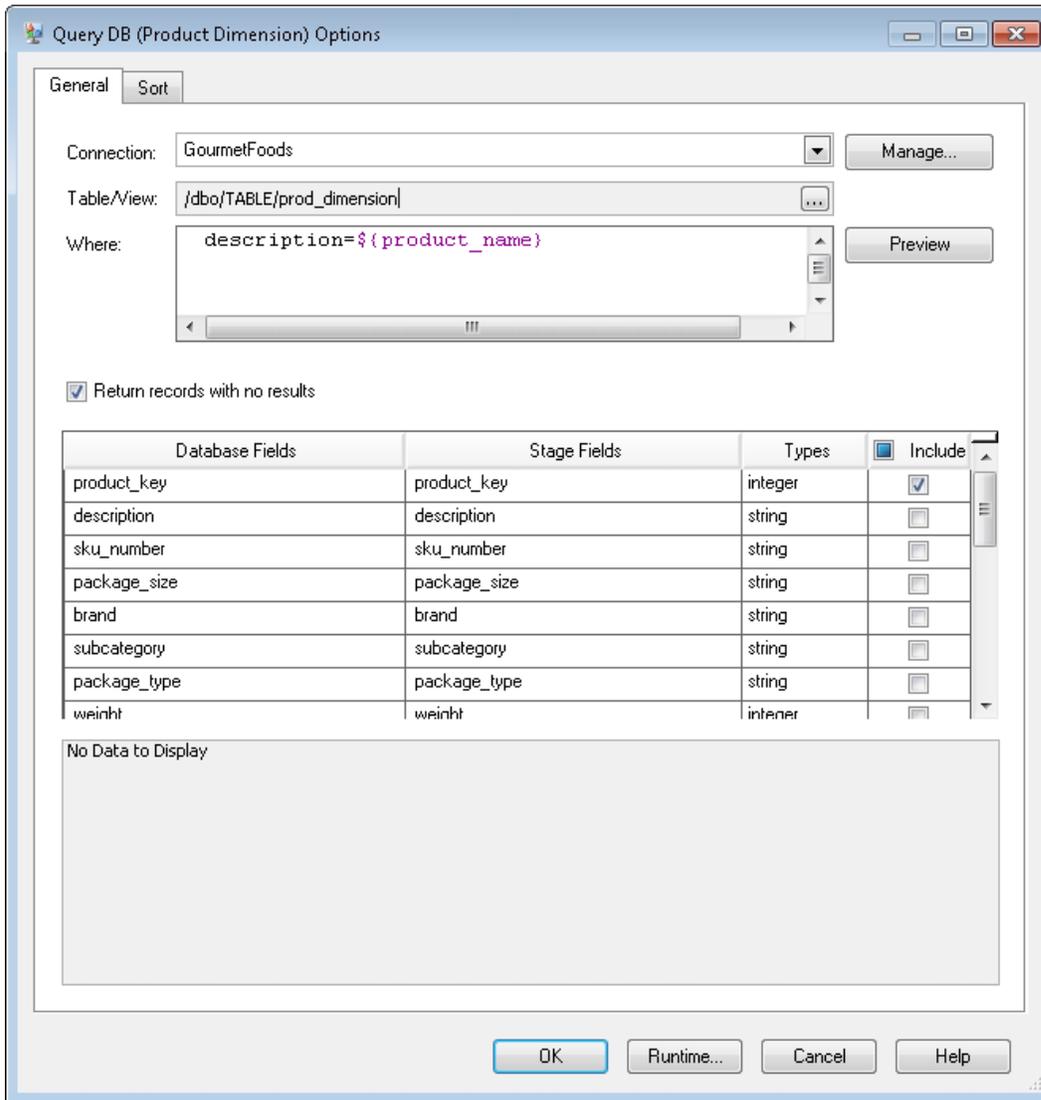
7. Configurez chaque stage Query DB de sorte qu'il recherche la clé de substitution pour chaque clé naturelle de la source de données. Pour ce faire, procédez comme suit :
  - a) Dans le champ **Connexion**, indiquez la connexion à l'entrepôt de données.
  - b) Dans le champ **Table/Vue**, sélectionnez la table de dimensions que vous souhaitez que ce stage interroge.
  - c) Dans le champ **Où**, écrivez une instruction `WHERE` qui recherche la clé de substitution en fonction de la valeur champ de flux de données approprié.

L'exemple suivant recherche la clé de substitution d'un produit en trouvant l'enregistrement de la table de dimensions dont la valeur de la colonne `description` correspond à la valeur du champ `product_name` de la source de données.

```
description=${product_name}
```

- d) Dans la colonne **Inclure**, sélectionnez la colonne de base de données contenant la clé de substitution.

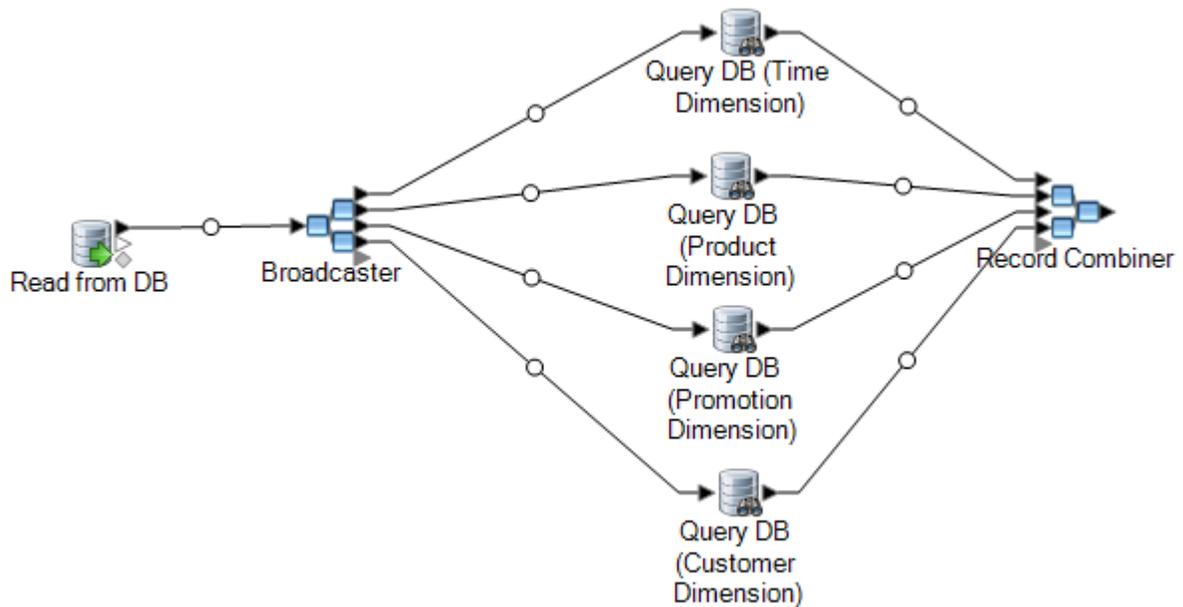
Par exemple, un stage Query DB qui recherche la clé de substitution d'un nom de produit ressemble à l'exemple suivant :



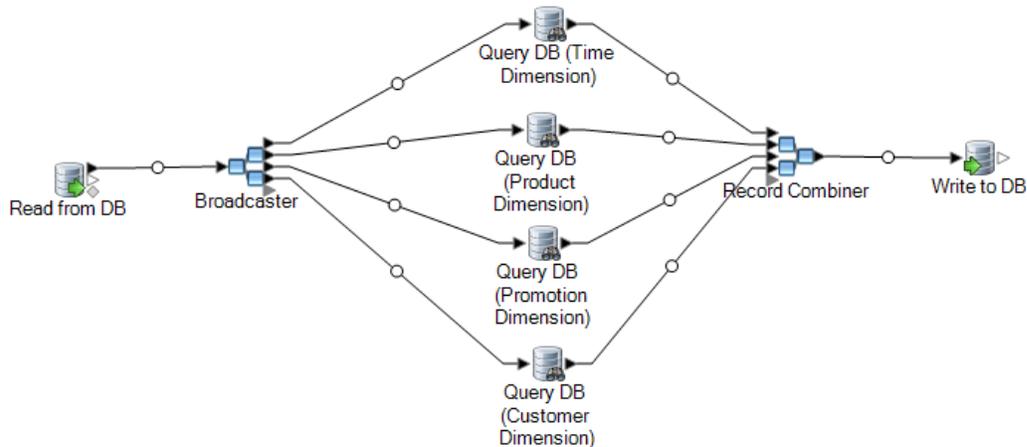
Dans cet exemple, la requête recherche la clé de produit en trouvant l'enregistrement de la table `prod_dimension` dont la valeur de la colonne `description` correspond à celle du champ de flux de données `product_name`. Le stage renvoie le champ `product_key` de la table et l'ajoute au flux de données, comme indiqué par la case cochée de la colonne **Include**.

8. Glissez un stage Record Combiner sur le canevas et connectez-lui l'ensemble des stages Query DB.

Votre flux de données prend maintenant la forme suivante :



9. Glissez un stage Write to DB sur le canevas et connectez-le au stage Record Combiner. Votre flux de données prend maintenant la forme suivante :

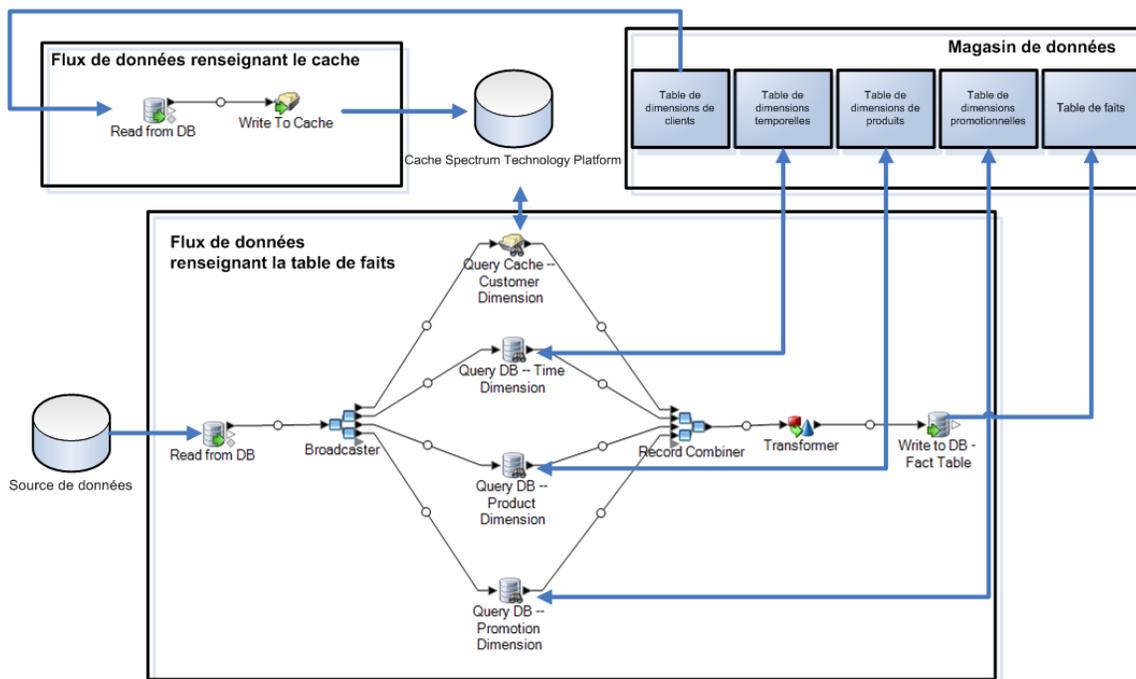


10. Configurez le stage Write to DB pour qu'il écrive les enregistrements dans la table de faits. Pour ce faire, procédez comme suit :
- Dans le champ **Connexion**, indiquez la connexion à l'entrepôt de données.
  - Dans le champ **Table/Vue**, sélectionnez la table de faits que vous souhaitez que ce stage interroge. Si la table de faits n'existe pas déjà dans l'entrepôt de données, cliquez sur **Créer une table** pour la créer.
  - Pour chaque champ que vous souhaitez écrire dans la table de faits, cochez la case de la colonne **Inclure**.
  - Dans l'onglet **Exécution**, notez que, par défaut, l'option **Insérer** est sélectionnée comme mode d'écriture. En règle générale, le renseignement de la table de faits est exécuté en mode d'insertion ; vous pouvez donc laisser cette option sélectionnée.

## Utilisation d'un cache global pour les requêtes

Si vous disposez d'une table de dimensions volumineuse, vous pouvez charger les données de cette table dans un cache et utiliser ce dernier pour rechercher des clés de substitution. L'utilisation d'un cache améliore les performances par rapport aux recherches directes dans la table de dimensions à l'aide de Query DB.

Pour utiliser un cache, vous devez créer deux flux de données : l'un pour renseigner le cache à l'aide de données provenant de la table de dimensions, et l'autre pour utiliser le cache lors de la mise à jour de la table de faits. Le diagramme suivant montre comment les deux flux de travail fonctionnent ensemble :



1. Créez un flux de données renseignant le cache à l'aide des données de la table de dimensions volumineuse.

Ce flux de données doit comprendre les deux stages suivants :

- un stage Read from DB lisant les données de la table de dimensions que vous souhaitez charger dans le cache ;
- un stage Write to Cache renseignant le cache à l'aide des données de la table de dimensions.

2. Exécutez ce flux de données pour renseigner le cache.
3. Dans le flux de données renseignant la table de faits, ajoutez un stage Query Cache.
4. Configurez le stage Query Cache pour qu'il interroge le cache créé par le stage Write to Cache.

5. Exécutez ce flux de données pour renseigner la table de faits.

Si vous souhaitez vérifier que le cache est renseigné à l'aide des données les plus récentes de la table de dimensions chaque fois que vous mettez à jour votre table de faits, vous pouvez créer un flux de processus qui exécute d'abord le job de renseignement de la table de dimensions, puis le job de mise à jour de la table de faits. Cela vous permet d'exécuter le flux de processus de sorte que les deux flux de données se suivent. Pour plus d'informations sur les flux de processus, reportez-vous au *Guide de Dataflow Designer*.

## Suppression d'un cache

Le stage Query Cache utilise un cache global pour les recherches de données. Si le cache n'est plus nécessaire, vous pouvez le supprimer comme suit. La suppression d'un cache libère de la mémoire.

**Remarque :** La **suppression d'un cache** est utilisée uniquement pour supprimer un cache global de Management console. Avant de supprimer un cache, assurez-vous qu'il n'est pas utilisé par un flux de données contenant un stage Query Cache. Si vous supprimez un cache utilisé par un flux de données, ce dernier échoue.

1. Ouvrez Management Console.
2. Développez **Modules > Enterprise Data Integration > Outils > Gestion de la mémoire cache**.
3. Sélectionnez le cache que vous souhaitez supprimer et cliquez sur **Supprimer**.

## Utilisation d'un cache local pour les requêtes

Si vous disposez d'une table de dimensions volumineuse, vous pouvez charger les données de cette table dans un cache et utiliser ce dernier pour rechercher des clés de substitution. L'utilisation d'un cache améliore les performances par rapport aux recherches directes dans la table de dimensions à l'aide de Query DB.

Un cache local est un cache temporaire utilisé uniquement lors de l'exécution du stage Query Cache. Il recherche ensuite les données du cache en fonction des champs de clé et des conditions de recherche et renvoie les données des enregistrements correspondants figurant dans le cache, en les ajoutant à l'enregistrement du flux de données. Sur un seul flux de données ou si la table objet de la recherche change fréquemment, utilisez un cache local plutôt qu'un cache global.

Pour utiliser un cache local pour les demandes, procédez comme suit :

1. Dans Enterprise Designer, ouvrez le flux de données dans lequel vous souhaitez exécuter une requête à l'aide d'un cache.
2. Glissez un stage Query Cache sur le canevas et connectez-le au flux de données.

3. Double-cliquez sur le stage Query Cache.
4. Sélectionnez **Cache local**.
5. Sélectionnez la connexion de base de données à utiliser. Vos choix varient en fonction des connexions qui sont définies dans le Gestionnaire de connexion de Management Console. Si vous devez établir une nouvelle connexion à la base de données, ou en modifier ou supprimer une existante, cliquez sur **Gérer**.

Si vous ajoutez ou modifiez une connexion de base de données, remplissez ces champs :

<b>Nom de la connexion</b>	Saisissez le nom de la connexion.: Ce nom est entièrement de votre choix.
<b>Pilote de la base de données</b>	Sélectionnez le type de base de données approprié.
<b>Options de connexion</b>	Indiquez l'hôte, le port, l'instance, le nom d'utilisateur et le mot de passe à utiliser pour vous connecter à la base de données.

6. Indiquez la table ou la vue de la base de données que vous souhaitez interroger.
7. Sélectionnez une clé sous l'option **Champ clé**.
8. Sélectionnez un champ d'entrée sous l'option **Champ d'entrée**. Le champ **Champ d'entrée** contient les champs provenant du stage précédent. Si la valeur de ce champ correspond à la clé du champ **Champ clé** de la base de données, la requête renvoie les données de cet enregistrement de la base de données.
9. Cliquez sur **OK**.
10. Exécutez le flux de données.

# 5 - Référence aux stages

## In this section

---

Call Stored Procedure	94
DB Change Data Reader	96
DB Loader	100
Field Parser	109
Field Combiner	113
Field Selector	114
Generate Time Dimension	115
Query Cache	123
Query DB	132
Requête NoSQL DB	136
Read From DB	139
Read from file	148
Read from Hadoop Sequence File	165
Read from Hive File	169
Read from HL7 File	173
Read from NoSQL DB	185
Read from SAP	189
Read from Spreadsheet	194
Read from Variable Format File	197
Read From XML	212
Commande SQL	219
Transposer	230
Unique ID Generator	234
Write To Cache	243
Write to DB	245
Write to File	251
Write to Hadoop Sequence File	269
Write to Hive File	272
Write to NoSQL DB	279
Write to Spreadsheet	283

Write to Variable Format File	286
Write to XML	296
Modèles de date et de nombre	306

## Call Stored Procedure

Le stage source Call Stored Procedure exécute une procédure stockée dans une base de données et renvoie les résultats de l'appel de la procédure stockée sous forme d'entrée du flux de données. Utilisez Call Stored procédure lorsque vous souhaitez obtenir des données d'une base de données à l'aide d'une procédure stockée d'une base de données plutôt que d'une requête sur une table ou une vue.

**Remarque :** Si vous souhaitez lire des données dans un flux de données directement depuis une table ou une vue, utilisez le stage Read from DB.

Si vous disposez d'une logique métier intégrée à la procédure stockée et si vous souhaitez utiliser cette logique dans votre environnement Spectrum™ Technology Platform, vous pouvez utiliser le stage Call Stored Procedure pour lire des données dans un flux de données. Par exemple, de nombreux systèmes opérationnels n'utilisent pas de contrôles d'intégrité référentielle dans la base de données pour les tables volumineuses constamment mises à jour, à cause de la réduction des performances que lesdits contrôles entraîneraient. C'est pourquoi, pour maintenir l'intégrité référentielle, vous pouvez créer des procédures stockées et les utiliser pour toutes les mises à jour du système.

Vous pouvez également utiliser des procédures stockées pour simplifier la gestion de l'environnement Spectrum™ Technology Platform. Par exemple, si vous disposez de centaines de processus ETL qui lisent tous les mêmes données, vous pouvez placer la requête dans une procédure stockée, de sorte qu'elle se trouve à un seul et même endroit. Cette procédure facilite la maintenance, car il vous suffit de modifier cette seule procédure stockée au lieu de centaines de processus différents.

Nom de l'option	Description
Connexion	<p>Sélectionnez la connexion de base de données à utiliser. Vos choix varient en fonction des connexions qui sont définies dans le Gestionnaire de connexion de Management Console. Si vous devez établir une nouvelle connexion à la base de données, ou en modifier ou supprimer une existante, cliquez sur <b>Gérer</b>.</p> <p>Si vous ajoutez ou modifiez une connexion de base de données, remplissez ces champs :</p> <p><b>Nom de la connexion</b> Saisissez le nom de la connexion.: Ce nom est entièrement de votre choix.</p> <p><b>Pilote de la base de données</b> Sélectionnez le type de base de données approprié.</p> <p><b>Options de connexion</b> Indiquez l'hôte, le port, l'instance, le nom d'utilisateur et le mot de passe à utiliser pour vous connecter à la base de données.</p>

Nom de l'option	Description
Schéma	Indique le schéma contenant la procédure stockée que vous souhaitez appeler.
Procédure	Indique la procédure stockée que vous souhaitez appeler.
Paramètres de la procédure stockée	<p>Cette table indique les valeurs des paramètres de la procédure stockée.</p> <p><b>Paramètres</b> Cette colonne indique les paramètres définis dans la procédure stockée.</p> <p><b>Champs de stage</b> Pour les paramètres OUT, INOUT et RETURN, cette colonne indique le nom de champ de flux de données qui va contenir les données renvoyées par le paramètre. Initialement, le nom de champ est identique au nom de paramètre. Vous pouvez modifier le nom de champ de stage en cliquant sur le nom de champ, puis en saisissant un nouveau nom pour les paramètres. Cette colonne n'est pas utilisée pour les paramètres IN.</p> <p><b>Sens</b> L'un des éléments suivants :</p> <p><b>IN</b> Il s'agit d'un paramètre d'entrée. La valeur que vous définissez pour ce paramètre est transmise à la procédure stockée sous forme d'entrée.</p> <p><b>OUT</b> Il s'agit d'un paramètre de sortie. La procédure stockée renvoie des données au stage dans ce paramètre.</p> <p><b>INOUT</b> Il peut s'agir d'un paramètre d'entrée, qui transmet une valeur à la procédure stockée, ou d'un paramètre de sortie, qui reçoit des données renvoyées par la procédure stockée.</p> <p><b>RETURN</b> Ce paramètre contient un code de renvoi provenant de la procédure stockée.</p> <p><b>Types</b> Cette colonne affiche le type de données de la valeur de paramètre. Si le type de données n'est pas pris en charge par Spectrum™ Technology Platform, il est déclaré « Non pris en charge » et la procédure stockée échoue.</p> <p><b>Valeur</b> Dans cette colonne, saisissez la valeur que vous souhaitez définir pour le paramètre. Pour les paramètres OUT, cette colonne est désactivée.</p>

Nom de l'option	Description
Champs de jeu de résultats	<p>Cette table indique les champs de flux de données à utiliser pour les données renvoyées par la procédure stockée.</p> <p><b>Tables de la base de données</b> Cette colonne indique les tables dont la procédure stockée a renvoyé des données.</p> <p><b>Champs de la base de données</b> Cette colonne indique le champ dont la procédure stockée a renvoyé des données.</p> <p><b>Champs de stage</b> Cette colonne indique le nom de champ de flux de données qui va contenir les données du champ de base de données.</p> <p><b>Types</b> Cette colonne indique le type de données du champ. Si le type de données n'est pas pris en charge par Spectrum™ Technology Platform, le type est « Non pris en charge ».</p> <p><b>Include</b> Pour inclure le champ au flux de données, cochez la case de cette colonne. Si la case est décochée, le champ n'est pas utilisé dans le flux de données.</p>
Obtenir les champs	Cliquez sur ce bouton pour renseigner la table Champs de jeu de résultats à l'aide du schéma de jeu de résultats renvoyé par la procédure stockée. Cette opération exécute la procédure stockée et permet d'obtenir le schéma de jeu de résultats.
Ajouter	Cliquez sur ce bouton pour ajouter manuellement un champ de jeu de résultats.
Supprimer	Cliquez sur ce bouton pour supprimer un champ de jeu de résultats de la liste de champs disponibles.

## DB Change Data Reader

Le stage **DB Change Data Reader** vous permet de sélectionner les colonnes à inclure dans le jobflow en cours, où les colonnes disposent de la fonction Change Data Capture activée.

Dans le stage, vous pouvez créer une ressource Change Data Capture (CDC), qui est la table de source de données requise. Les colonnes de la ressource CDC pour lesquelles la fonction Change Data Capture est activée sont indiquées à l'aide des cases à cocher.

### Modifier la gestion de la capture de données

La fonction Change Data Capture permet la capture de toutes les modifications apportées dans une colonne. Pour chaque colonne sélectionnée, toutes les insertions, les mises à jour et les suppressions sont capturées.

### Bases de données prises en charge

Actuellement, Spectrum™ Technology Platform prend en charge la fonction Change Data Capture (CDC) pour les bases de données MS SQL et Oracle uniquement.

**MS SQL** Pour les sources de données MS SQL, la fonction Change Data Capture peut être activée ou désactivée pour les colonnes de tables en arrière-plan. Cliquez [ici](#) pour connaître les étapes nécessaires.

**Remarque :** La fonction Change Data Capture de Spectrum™ n'est pas prise en charge dans l'édition Express de SQL Server.

**Oracle** Pour les sources de données Oracle, Spectrum™ suit les modifications apportées aux données des colonnes de table à l'aide de l'utilitaire LogMiner d'Oracle. CDC ne peut pas être activé ni désactivé sur les colonnes des tables d'une source de données Oracle via Spectrum™ Technology Platform.

Les changements de données résultant des requêtes `insert`, `update` et `delete` sont suivis et capturés, de la date et de l'heure de début saisies à la date et l'heure en cours. Cette heure de début saisie s'applique lors de l'exécution de la première requête avec CDC activé.

Lors des exécutions suivantes sur la même connexion Oracle, les modifications apportées aux données sont capturées à partir de l'heure de la dernière exécution jusqu'à l'heure en cours.

**Remarque :** Il s'agit d'un processus incrémentiel. Lors de la première capture, les modifications sont capturées depuis la date et l'heure de début saisies jusqu'à la date et l'heure en cours. Lors des captures suivantes, les modifications sont capturées depuis la date et l'heure de fin de la précédente capture jusqu'à la date et l'heure en cours.

## Ajout d'une ressource CDC

**Remarque :** Pour utiliser la fonction Change Data Capture de Spectrum™, assurez-vous que l'Agent SQL Server est exécuté sur le serveur MS SQL.

- Ouvrez le menu contextuel **Modifier la gestion de la capture des données** à l'aide de l'une des deux méthodes suivantes :
  - Accédez à **Outils > Modifier la gestion de la capture des données**.
  - Ajoutez le stage **DB Change Data Reader** à un job, ouvrez les paramètres de stage et cliquez sur **Gérer**.

2. Cliquez sur **Ajouter**.
3. Saisissez le **Nom** de la ressource CDC.
4. Dans le champ **Connexion**, sélectionnez la connexion à la base de données SQL que vous souhaitez utiliser. Pour effectuer une nouvelle connexion de base de données, cliquez sur **Gérer**. Pour plus d'informations sur la création des connexions de bases de données, consultez le [Gestionnaire de connexion à la base de données](#).
5. Dans le champ **Table/Vue**, indiquez la table dont les colonnes doivent être incluses dans le jobflow. Cliquez sur le bouton Parcourir ([...]) pour naviguer jusqu'à la table ou la vue que vous souhaitez utiliser.

La grille ci-dessous affiche toutes les colonnes de la table sélectionnée, ainsi que les types de données de chaque colonne.

6. Si vous sélectionnez une connexion Oracle dans le champ **Connexion**, le champ **Date de début** devient disponible.  
Le champ est renseigné avec la valeur par défaut de la date du jour et avec l'heure 12:00 AM. Vous pouvez saisir la date et l'heure de début de votre choix.

La date et l'heure de fin sont la date du jour et l'heure en cours.

**Avertissement :** Vous devez disposer de droits d'*exécution* sur l'utilitaire Oracle LogMiner pour pouvoir utiliser la fonction CDC sur une connexion Oracle. Pour plus d'informations, reportez-vous à la section [Configurations de LogMiner Oracle](#) à la page 313.

**Remarque :** Le champ **Date de début** n'est pas disponible lors de la sélection d'une connexion MS SQL dans le champ **Connexion**.

7. Cliquez sur **OK**.

La table de ressource CDC créée est maintenant prête à être utilisée dans le stage **DB Change Data Reader**, dont les colonnes peuvent être incluses ou exclues dans le/du jobflow.

Le stage affiche les colonnes de la table pour lesquelles la fonction CDC a été activée.

## Modification d'une ressource CDC

1. Ouvrez le menu contextuel **Modifier la gestion de la capture des données** à l'aide de l'une des deux méthodes suivantes :
  - Accédez à **Outils > Modifier la gestion de la capture des données**.
  - Ajoutez le stage **DB Change Data Reader** à un job, ouvrez les paramètres de stage et cliquez sur **Gérer**.
2. Sélectionnez la ressource CDC que vous devez modifier.
3. Cliquez sur **Modifier**.
4. Modifiez les détails de la ressource CDC ajoutée, comme requis.
5. Cliquez sur **OK**.

## Suppression d'une ressource CDC

- Ouvrez le menu contextuel **Modifier la gestion de la capture des données** à l'aide de l'une des deux méthodes suivantes :
  - Accédez à **Outils > Modifier la gestion de la capture des données**.
  - Ajoutez le stage **DB Change Data Reader** à un job, ouvrez les paramètres de stage et cliquez sur **Gérer**.
- Sélectionnez la ressource CDC que vous devez supprimer.
- Cliquez sur **Supprimer**.

## Sélection de Change Data Reader Options

**DB Change Data Reader Options** reflète les colonnes de table de la ressource CDC sélectionnée pour lesquelles la fonction CDC est activée.

Vous pouvez sélectionner les colonnes à inclure ou à exclure dans le/du jobflow.

- Dans un job, ajoutez le stage **DB Change Data Reader**.
- Ouvrez le stage **DB Change Data Reader Options** en double-cliquant sur l'icône de stage.
- Sélectionnez la ressource CDC de votre choix dans la liste déroulante **Sélectionner une ressource....**

Vous pouvez ajouter ou modifier une ressource CDC en cliquant sur **Gérer**.

La grille ci-dessous affiche toutes les colonnes de table avec leurs types de données. Elle indique également si une colonne donnée est incluse dans le flux de jobs et si la colonne est sélectionnée pour la fonction Change Data Capture.

- À l'aide des cases à cocher figurant sous la colonne **Inclure** de la grille, sélectionnez les colonnes de table à inclure dans le flux de jobs.
- Les cases à cocher figurant sous la colonne **CDC activé** de la grille reflètent les colonnes de table pour lesquelles la fonction CDC est activée.

Les cases à cocher **CDC activé** en lecture seule sont cochées pour les colonnes pour lesquelles la fonction CDC est activée.

**Remarque :** Pour les sources de données MS SQL, cliquez [ici](#) pour connaître les étapes d'activation ou de désactivation de la fonction Change Data Capture sur des colonnes de table données de l'application dorsale.

- Cliquez sur **OK**.

Les données des colonnes de table pour lesquelles la fonction Change Data Capture a été sélectionnée sont recueillies et enregistrées.

## DB Loader

Le stage **DB Loader** vous permet d'accéder aux données et de les charger depuis/vers les bases de données configurées dans le module Spectrum™ Data Integration. Ce stage fournit une interface avec un utilitaire de chargement de données haut débit. Actuellement, la plate-forme Spectrum™ Data Integration prend en charge **Oracle Loader**, **DB2 Loader**, **PostgreSQL Loader** et **Teradata Loader**.

## Oracle Loader

Oracle Loader vous permet de charger des données dans toute base de données Oracle configurée sur la plate-forme Spectrum™ Data Integration.

**Remarque :** Avant de pouvoir utiliser Oracle Loader, vous devez installer le client Oracle avec une configuration d'administrateur.

Nom de l'option	Description
Connexion	<p>Sélectionnez la connexion de base de données à utiliser. Vos choix varient en fonction des connexions qui sont définies dans le Gestionnaire de connexion de Management Console. Si vous devez établir une nouvelle connexion à la base de données, ou en modifier ou supprimer une existante, cliquez sur <b>Gérer</b>.</p> <p>Si vous ajoutez ou modifiez une connexion de base de données, remplissez ces champs :</p> <p><b>Nom de la connexion</b> Saisissez le nom de la connexion.: Ce nom est entièrement de votre choix.</p> <p><b>Pilote de la base de données</b> Sélectionnez le type de base de données approprié.</p> <p><b>Options de connexion</b> Indiquez l'hôte, le port, l'instance, le nom d'utilisateur et le mot de passe à utiliser pour vous connecter à la base de données.</p>
Table/Vue	<p>Après avoir sélectionné une connexion, spécifiez la vue ou la table pour l'écriture. Cliquez sur le bouton Parcourir ([...]) pour accéder à la table ou vue à utiliser, ou cliquez sur <b>Créer une table</b> pour créer une table dans la base de données.</p> <p><b>Remarque :</b> Si vous écrivez sur une base de données SQL, vous ne pouvez pas écrire dans les vues qui font référence à plus d'une table. Cela est dû à une limitation dans SQL Server.</p>

Nom de l'option	Description
Écouteur	Indique un nom de variable contenant l'adresse et les détails de connexion requis pour établir une connexion à la base de données Oracle. Par exemple, « XE ». Cette variable figure dans <code>tnsnames.ora</code> , fichier qui contient les paramètres de configuration réseau côté client.
Champs de stage	Cette colonne contient une liste des noms de champ utilisés dans le flux de données. Vous ne pouvez pas modifier les noms de ces champs.
Types	Cette colonne contient une liste des types de données de chaque champ.

### L'onglet Exécution

Nom de l'option	Description
Méthode de chargement	Indique comment vous souhaitez écrire des données dans la base de données Oracle. <ul style="list-style-type: none"> <li><b>Ajouter</b> Ajoute les données à la table cible sans effacer les données existantes de la table.</li> <li><b>Insérer</b> Charge les données dans la base de données. Avant d'être chargée, la table doit être vide. Cette opération ne fonctionne pas sur plusieurs instances d'exécution.</li> <li><b>Tronquer et insérer</b> Supprime les lignes existantes, le cas échéant, puis charge les données dans la table. Cette opération ne fonctionne pas sur plusieurs instances d'exécution.</li> </ul>
Utiliser l'utilitaire de chargement du chemin direct	Sélectionnez cette option pour charger directement les données dans la base de données Oracle en contournant une grande partie du traitement de données normalement effectué.
Irrécupérable	Cette case est activée lorsque vous sélectionnez <b>Utiliser l'utilitaire de chargement du chemin direct</b> . Sélectionnez cette option si vous ne souhaitez pas écrire de journaux Rétablir dans la base de données. Pour obtenir davantage d'informations sur les journaux Rétablir, consultez l'adresse suivante : <a href="http://docs.oracle.com/cd/B28359_01/server.111/b28310/onlineredo001.htm#ADMIN11302">http://docs.oracle.com/cd/B28359_01/server.111/b28310/onlineredo001.htm#ADMIN11302</a> .
Dossier des fichiers journaux	Indique le chemin d'accès au dossier. Pour accéder au dossier de votre choix, cliquez sur le bouton de sélection (...). Le fichier journal contient un enregistrement des activités de ce stage au cours d'une session de chargement.
Dossier de fichiers incorrects	Indique le chemin d'accès au dossier. Pour accéder au dossier de votre choix, cliquez sur le bouton de sélection (...). Le fichier incorrect contient une liste d'enregistrements que le stage ne parvient pas à charger dans la base de données.

Nom de l'option	Description
Nombre max. d'erreurs autorisées	Indique le nombre maximal d'erreurs à autoriser avant d'interrompre l'opération de chargement. Pour interrompre l'opération de chargement dès la première erreur, définissez cette valeur sur 0. Le nombre maximal d'erreurs autorisé est 32 767.

**Remarque :** Vous pouvez obtenir des améliorations significatives des performances en utilisant plusieurs instances d'exécution de cette opération. Pour ce faire, cliquez sur le bouton **Exécution** et saisissez la valeur requise dans le champ **Instances d'exécution**.

## DB2 Loader

DB2 Loader vous permet de charger des données dans une base de données DB2 configurée sur la plate-forme Spectrum™ Data Integration.

**Remarque :** Avant de pouvoir utiliser DB2 Loader, vous devez installer le client d'exécution DB2 avec une configuration administrateur.

Nom de l'option	Description
Connexion	<p>Sélectionnez la connexion de base de données à utiliser. Vos choix varient en fonction des connexions qui sont définies dans le Gestionnaire de connexion de Management Console. Si vous devez établir une nouvelle connexion à la base de données, ou en modifier ou supprimer une existante, cliquez sur <b>Gérer</b>.</p> <p>Si vous ajoutez ou modifiez une connexion de base de données, remplissez ces champs :</p> <p><b>Nom de la connexion</b> Saisissez le nom de la connexion.: Ce nom est entièrement de votre choix.</p> <p><b>Pilote de la base de données</b> Sélectionnez le type de base de données approprié.</p> <p><b>Options de connexion</b> Indiquez l'hôte, le port, l'instance, le nom d'utilisateur et le mot de passe à utiliser pour vous connecter à la base de données.</p>
Table/View	Après avoir sélectionné une connexion, spécifiez la vue ou la table pour l'écriture. Cliquez sur le bouton Parcourir ([...]) pour accéder à la table ou vue à utiliser, ou cliquez sur <b>Créer une table</b> pour créer une table dans la base de données.

Nom de l'option	Description
Base de données/Alias	<p>Il s'agit d'une variable qui place la base de données et le serveur DB2 dans un catalogue.</p> <p><b>Pour placer le serveur DB2 dans un catalogue</b> Utilisez le processeur de ligne de commande DB2 sur la machine du serveur Spectrum et saisissez la commande suivante :</p> <pre>CATALOG TCPIP NODE &lt;nodename&gt; REMOTE &lt;hostname&gt; SERVER &lt;port&gt;</pre> <p>où :</p> <p><b>nodename</b> : nom de la connexion</p> <p><b>hostname</b> : nom TCP/IP de l'ordinateur du serveur DB2</p> <p><b>port</b> : port du serveur</p> <p><b>Pour placer la base de données dans un catalogue</b> Utilisez la commande :</p> <pre>CATALOG DATABASE &lt;databasename&gt; AS &lt;local_database_alias&gt; AT NODE &lt;nodename&gt;</pre> <p>où :</p> <p><b>databasename</b> : nom de la base de données sur le serveur DB2</p> <p><b>local_database_alias</b> : nom local donné à la base de données lors de la connexion depuis l'ordinateur du serveur</p> <p><b>nodename</b> : nom utilisé dans la commande CATALOG TCP/IP précédente</p>
Champs de stage	Cette colonne contient une liste des noms de champ utilisés dans le flux de données. Vous ne pouvez pas modifier les noms de ces champs.
Types	Cette colonne contient une liste des types de données de chaque champ.

## L'onglet Exécution

Nom de l'option	Description
Méthode de chargement	<p>Indique le mode d'écriture des données dans une table DB2.</p> <p><b>Insérer</b>      Insère les données chargées dans la table, tandis que les données existantes de la table restent inchangées.</p> <p><b>Remplacer</b>      Insère les données chargées dans la table après avoir supprimé toutes les données existantes qu'elle contient.</p> <p>Les définitions de schéma et d'index de la table restent intactes.</p> <p><b>Redémarrer</b>      Redémarre le chargement de données, dans le cas où la tentative de chargement précédente a été interrompue.</p>
Non récupérable	<p>Indique que cette transaction de chargement n'est pas récupérable.</p> <p>Si vous sélectionnez cette option, la transaction de chargement est marquée comme non récupérable. Les espaces de la table ne sont pas définis sur l'état Backup Pending après le chargement, et les données chargées ne sont pas copiées lors du chargement. Par conséquent, une transaction non récupérable ne peut pas être récupérée en cas d'échec de chargement des données, même si une opération <code>rollforward</code> est tentée ultérieurement.</p> <p>Si vous sélectionnez cette option, vous ne pouvez pas la transaction, même si vous utilisez l'utilitaire DB2 <code>rollforward</code>, parce que l'utilitaire ignore une telle transaction non récupérable et que la table est marquée comme « non valide ». En outre, les transactions suivantes sur la table sont elles aussi ignorées par <code>rollforward</code>.</p> <p>Pour restaurer une table contenant des transactions non-récupérables, vous devez utiliser une sauvegarde au niveau des espaces de la table ou une sauvegarde complète effectuée à un point de validation suivant le chargement non récupérable.</p> <p><b>Remarque :</b> Ne sélectionnez pas cette option si les données contiennent des colonnes Datalink avec l'attribut File Link Control.</p>
CPU	<p>Nombre de threads parallèles que l'utilitaire de chargement peut générer et gérer pour charger, analyser et formater les enregistrements, tout en créant des objets de table dans chaque partition de base de données.</p>
Disque	<p>Nombre de threads parallèles que l'utilitaire de chargement peut générer et gérer pour écrire des données dans les conteneurs d'espaces de table.</p>

Nom de l'option	Description
Mode d'indexation	<p>Indique le mode de gestion des index par l'utilitaire de chargement.</p> <p><b>Autoselect</b> L'utilitaire de chargement décide s'il faut appliquer le mode Régénérer ou Incrémentiel, en fonction de la quantité de données et de la profondeur de l'arborescence des index.</p> <p><b>Régénérer</b> Tous les index sont recréés.</p> <p><b>Incrémentiel</b> Les nouvelles données sont ajoutées aux index existants.</p> <p>Ce mode peut être appliqué uniquement si l'objet indexé est valide et accessible au début d'une opération de chargement.</p> <p><b>Remarque :</b> L'indexation incrémentielle n'est pas pris en charge lorsque TOUTES les conditions suivantes sont réunies :</p> <ol style="list-style-type: none"> <li>1. L'option Load Copy est indiquée <code>logretain</code> ou <code>userexit</code> est activé).</li> <li>2. La table se trouve dans un espace de table DMS.</li> <li>3. L'objet index se trouve dans un espace de table partagé par d'autres objets table appartenant à la table en cours de chargement.</li> </ol> <p>Pour contourner cette restriction, placez les index dans des espaces de table distincts.</p> <p><b>Deferred</b> L'utilitaire de chargement ne tente pas de créer un index. Les index existants sont marqués comme devant être actualisés.</p> <p><b>Remarque :</b> La construction d'index nécessite plus de temps en mode Deferred qu'en mode Régénérer. Par conséquent, lorsque vous effectuez plusieurs opérations de chargement, laissez la dernière opération de chargement régénérer tous les index au lieu de les régénérer lors du premier accès par une opération n'effectuant pas de chargement.</p> <p><b>Remarque :</b> Ce mode est pris en charge uniquement pour les tables comportant des index qui ne sont pas uniques.</p>
Analyse rapide	<p>Indique si la validation syntaxique des valeurs des colonnes doit être ignorée, pour améliorer les performances.</p> <p>Si cette option est cochée, toutes les erreurs syntaxiques des données sont ignorées afin d'optimiser les performances.</p> <p>Par exemple, si une valeur de chaîne <code>12wxvg56</code> est détectée dans un champ mappé vers une colonne d'entiers dans un fichier ASCII, l'utilitaire de chargement doit normalement générer une erreur de syntaxe. Mais, si Analyse rapide est sélectionné, l'erreur de syntaxe est ignorée et un nombre arbitraire est chargé dans le champ d'entier.</p> <p><b>Remarque :</b> Assurez-vous que vous utilisez cette option uniquement avec des données correctes et propres.</p>

Nom de l'option	Description
Nom du schéma	Schéma dans lequel les tables d'exceptions sont stockées.
Nom de la table	Table d'exceptions dans laquelle sont copiées les lignes contenant des erreurs lors du chargement.
Dossier des fichiers journaux	<p>Chemin d'accès au répertoire dans lequel les fichiers journaux doivent être stockés.</p> <p>Un fichier journal contient une liste des transactions de chargement de base de données exécutées par un stage <b>DB Loader</b> au cours d'une session de chargement.</p> <p>Cliquez sur le bouton d'ellipse (...) pour spécifier le répertoire de votre choix pour le stockage des fichiers journaux.</p>
Dossier de fichiers incorrects	<p>Chemin d'accès au répertoire du serveur DB2 dans lequel les fichiers incorrects doivent être stockés.</p> <p>Un fichier incorrect contient une liste des enregistrements qu'un stage <b>DB Loader</b> ne parvient pas à charger dans la base de données.</p> <p>Cliquez sur le bouton d'ellipse (...) pour spécifier le répertoire de votre choix pour le stockage des fichiers incorrects.</p>
Nombre max. d'erreurs autorisées	<p>Nombre maximal d'erreurs autorisées avant qu'une opération de chargement soit abandonnée.</p> <p>Pour abandonner une opération dès que la première erreur est rencontrée, définissez la valeur de ce champ sur 0.</p> <p style="text-align: center;"><b>Remarque :</b> Le nombre maximal d'erreurs autorisé est 32 767.</p>
<b>Parallélisme</b>	<p>Une base de données DB2 peut être divisée en plusieurs partitions via le clonage de l'environnement sur différents nœuds physiques.</p> <p>Les requêtes de base de données distinctes pour l'extraction et la mise à jour des données sont automatiquement réparties sur les différentes partitions et exécutées en parallèle pour optimiser les performances.</p>
<b>Gestion des exceptions</b>	<p>Une base de données DB2 vous permet d'enregistrer les erreurs et les exceptions rencontrées lors de l'exécution des requêtes et des procédures, ainsi que de les gérer de manière appropriée.</p> <p>Pour ce faire, une base de données DB2 fournit des tables d'exceptions et un schéma qui stockent la source ainsi que les traces de consignation de chaque exception de la base de données.</p>

## PostgreSQL Loader

PostgreSQL Loader vous permet de charger des données dans une base de données PostgreSQL configurée sur la plate-forme Spectrum™ Data Integration.

Nom de l'option	Description
Connexion	<p>Sélectionnez la connexion de base de données à utiliser. Vos choix varient en fonction des connexions qui sont définies dans le Gestionnaire de connexion de Management Console. Si vous devez établir une nouvelle connexion à la base de données, ou en modifier ou supprimer une existante, cliquez sur <b>Gérer</b>.</p> <p>Si vous ajoutez ou modifiez une connexion de base de données, remplissez ces champs :</p> <p><b>Nom de la connexion</b> Saisissez le nom de la connexion.: Ce nom est entièrement de votre choix.</p> <p><b>Pilote de la base de données</b> Sélectionnez le type de base de données approprié.</p> <p><b>Options de connexion</b> Indiquez l'hôte, le port, l'instance, le nom d'utilisateur et le mot de passe à utiliser pour vous connecter à la base de données.</p>
Table/Vue	Cliquez sur le bouton Parcourir ([...]) pour accéder à la table ou la vue que vous souhaitez utiliser.
Champs de la base de données	Cette colonne contient une liste des noms de champ de la base de données. Vous ne pouvez pas modifier les noms de ces champs.
Champs de stage	Cette colonne contient une liste des noms de champ utilisés dans le flux de données. Vous ne pouvez pas modifier les noms de ces champs.
Types	Cette colonne contient une liste des types de données de chaque champ.

### L'onglet Exécution

Nom de l'option	Description
Méthode de chargement	<p>Indique la méthode d'écriture des données dans les tables de base de données PostgreSQL.</p> <ul style="list-style-type: none"> <li>• Sélectionnez <b>Insérer</b> pour écrire des données dans une table vide ou pour les ajouter à une table de données existante.</li> <li>• Sélectionnez <b>Tronquer et Insérer</b> pour tronquer les données avant de les charger dans la table de base de données.</li> </ul>

**Remarque :** Vous pouvez obtenir des améliorations significatives des performances en utilisant plusieurs instances d'exécution de cette opération. Pour ce faire, cliquez sur le bouton **Exécution** et saisissez la valeur requise dans le champ **Instances d'exécution**.

## Teradata Loader

Teradata Loader vous permet de charger des données dans une base de données Teradata configurée sur la plate-forme Spectrum™ Data Integration.

**Remarque :** Le chargeur n'est pris en charge que sur les systèmes Windows.

Nom de l'option	Description
Connexion	<p>Sélectionnez la connexion de base de données à utiliser. Vos choix varient en fonction des connexions qui sont définies dans le Gestionnaire de connexion de Management Console. Si vous devez établir une nouvelle connexion à la base de données, ou en modifier ou supprimer une existante, cliquez sur <b>Gérer</b>.</p> <p>Si vous ajoutez ou modifiez une connexion de base de données, remplissez ces champs :</p> <p><b>Nom de la connexion</b> Saisissez le nom de la connexion.: Ce nom est entièrement de votre choix.</p> <p><b>Pilote de la base de données</b> Sélectionnez le type de base de données approprié.</p> <p><b>Options de connexion</b> Indiquez l'hôte, le port, l'instance, le nom d'utilisateur et le mot de passe à utiliser pour vous connecter à la base de données.</p>
Table/Vue	Cliquez sur le bouton Parcourir ([...]) pour accéder à la table ou la vue que vous souhaitez utiliser.
Champs de la base de données	Cette colonne contient une liste noms de champ de la base de données. Vous ne pouvez pas modifier les noms de ces champs.
Champs de stage	Cette colonne contient une liste des noms de champ utilisés dans le flux de données. Vous ne pouvez pas modifier les noms de ces champs.
Types	Cette colonne contient une liste des types de données de chaque champ.

## L'onglet Exécution

Nom de l'option	Description
Méthode de chargement	Indique la méthode d'écriture des données dans les tables de base de données Teradata. <ul style="list-style-type: none"> <li>• Sélectionnez <b>Insérer</b> pour écrire des données dans une table vide.</li> <li>• Sélectionnez <b>Ajouter</b> pour écrire des données dans une table de données existante.</li> <li>• Sélectionnez <b>Tronquer et Insérer</b> pour tronquer les données avant de les écrire dans la table de base de données.</li> </ul>
Dossier des fichiers journaux	Sélectionnez l'emplacement de dossier pour enregistrer les fichiers journaux de la procédure de chargement.
Générer un mauvais fichier	Cochez cette case pour générer un fichier journal consignait les enregistrements ayant échoué au chargement.
Dossier de fichiers incorrects	Sélectionnez l'emplacement de dossier pour l'enregistrement du journal d'enregistrements ayant échoué. Le journal donne plus d'informations, telles que les codes d'erreur et les noms de champ d'erreur des enregistrements ayant échoué.
Nombre max. d'erreurs autorisées	Spécifiez la limite d'erreurs autorisée pour la session de chargement. Si le nombre d'erreurs dépasse cette valeur, le processus de chargement s'arrête.

**Remarque :** Teradata Loader ne prend en charge qu'une seule instance d'exécution à la fois.

## Field Parser

Le stage **Field Parser** extrait les champs des données XML et Delimited dans la colonne d'entrée spécifiée. Pour configurer les options de **Field Parser**, procédez comme suit.

1. Dans le champ **Source**, sélectionnez la colonne contenant les données XML ou Delimited à analyser.

**Remarque :** La liste déroulante affiche toutes les colonnes d'entrée de chaîne.

2. Sélectionnez le **Format** XML ou Delimited suivant le type de données que vous souhaitez analyser et, en conséquence, sélectionnez les options décrites ci-dessous.

## Options Field Parser pour les données XML

Nom de l'option	Description
Server name	Indique si le fichier que vous avez sélectionné pour inférer le schéma se trouve sur l'ordinateur exécutant Enterprise Designer ou sur le serveur. Si vous sélectionnez un fichier sur l'ordinateur local, le nom du serveur sera Mon ordinateur. Si vous sélectionnez un fichier sur le serveur, le nom du serveur sera Spectrum™ Technology Platform.
Fichier de Schéma	<p>Indique le chemin du fichier de schéma XSD. Cliquez sur le bouton d'ellipse (...) pour accéder à l'emplacement du fichier. Le fichier de schéma peut résider sur le serveur ou sur votre système local.</p> <p>Sinon, vous pouvez également spécifier un fichier XML au lieu d'un fichier XSD. Si vous spécifiez un fichier XML, le schéma est inféré en fonction de la structure du fichier XML. Utiliser un fichier XML au lieu d'un fichier XSD présente les limites suivantes :</p> <ul style="list-style-type: none"> <li>• Le fichier XML ne peut pas être supérieur à 1 Mo. Si la taille du fichier XML est supérieure à 1 Mo, essayez d'enlever quelques données en maintenant la structure du fichier XML.</li> <li>• Le fichier de données ne sera pas validé en fonction du schéma déduit.</li> </ul> <p><b>Remarque :</b> Si le Spectrum™ Technology Platform est en cours d'exécution sur les systèmes Unix ou Linux, rappelez-vous que les noms de fichiers et les chemins sur ces plates-formes sont sensibles à la casse.</p>
Champs de sortie	<p>Cette section affiche les détails du schéma sélectionné. Elle inclut l'élément racine, suivi des éléments enfants ainsi que de leurs attributs.</p> <p>Par défaut, tous les nœuds du schéma restent sélectionnés. Cependant, vous pouvez décocher la case des nœuds que vous ne souhaitez pas transmettre au stage suivant.</p> <ul style="list-style-type: none"> <li>• <b>Rechercher un nœud :</b> saisissez le nom du nœud auquel vous souhaitez accéder dans l'arborescence de schémas. Le nœud saisi est mis en surbrillance dans le volet d'aperçu sous le champ.</li> <li>• <b>XPath :</b> cliquez n'importe où dans ce champ pour afficher le chemin d'accès XML (XPath) des éléments et attributs du nœud en surbrillance dans l'arborescence de schémas. Pour afficher tous les XPath que vous avez précédemment affichés, cliquez sur la flèche en bas à droite du champ.</li> </ul> <p><b>Remarque :</b> XPath est un langage qui permet de trouver des informations dans un document XML. Pour plus d'informations, consultez <a href="http://www.w3schools.com/xpath/">www.w3schools.com/xpath/</a>.</p>

### Options Field Parser pour les données au format Delimited

Nom de l'option	Description
Séparateur de champs	<p>Dans la liste déroulante, sélectionnez le séparateur de champs utilisé dans la colonne Delimited à analyser.</p> <p>Si la colonne Delimited utilise un caractère différent comme séparateur de champ, cliquez sur le bouton d'ellipse pour sélectionner un autre caractère comme séparateur de champ.</p>
Qualificateur de texte	<p>Dans la liste déroulante, sélectionnez le qualificateur de texte utilisé dans la colonne Delimited à analyser.</p> <p><b>Remarque :</b> Les qualificateurs de texte sont le caractère utilisé pour entourer les valeurs de texte dans des données au format Delimited.</p> <p>Si la colonne Delimited utilise un qualificateur de texte différent, cliquez sur le bouton d'ellipse pour sélectionner un autre caractère comme qualificateur de texte.</p>
Type de sortie	<p>Sélectionnez si vous souhaitez que la sortie analysée se présente sous la forme d'une <b>Liste</b> (affichage hiérarchique des valeurs) ou de <b>Champs</b>.</p> <p><b>Remarque :</b> Pour obtenir une liste comme type de sortie, vous pouvez ajouter un seul champ de sortie, tandis que l'option <b>Champs</b> vous permet d'ajouter plusieurs champs dans lesquels vous pouvez obtenir les valeurs distinctes pendant l'analyse.</p>

Nom de l'option	Description
Champs de sortie	<p>Cette section vous permet d'ajouter/de modifier les différents champs dans lesquels vous souhaitez que les détails de la colonne au format Delimited soient séparés. Vous pouvez également supprimer n'importe quel champ de sortie ajouté.</p> <p>Pour ajouter un nouveau champ à afficher dans la sortie analysée, cliquez sur le bouton <b>Ajouter</b> et effectuez ces étapes dans la fenêtre contextuelle <b>Paramètre de Champ</b> qui s'affiche :</p> <ol style="list-style-type: none"> <li>Saisissez le <b>Nom</b> du champ.</li> <li>Dans la liste déroulante <b>Type</b>, sélectionnez le type de données du champ à ajouter. Suivant le type sélectionné, quelques autres champs peuvent être définis. Par exemple, en cas de <i>date</i>, vous pouvez définir son format sur <i>M/j/aa, MMM j .aaaa</i> ou <i>MMMM j .aaaa</i>. Pour des informations plus détaillées sur les types de données et la définition de leurs détails, reportez-vous à la section <a href="#">Définition de champs dans un fichier d'entrée délimité</a> à la page 154.</li> </ol> <p style="padding-left: 40px;"><b>Remarque</b> : Si vous sélectionnez <i>Chaîne</i> comme type de données, n'importe quel type de données au format Delimited est analysé. Cependant, vous pouvez également utiliser le type spécifique, en fonction des données que vous souhaitez analyser dans le champ.</p> <ol style="list-style-type: none"> <li>Dans le champ <b>Position</b>, saisissez la position du type de données (dans le fichier d'entrée) à analyser pour ce champ. Par exemple, dans l'extrait de fichier suivant, si vous souhaitez analyser les valeurs de date et d'heure dans le champ en cours d'ajout, définissez la <b>Position</b> sur 3.</li> </ol> <pre style="background-color: #f0f0f0; padding: 10px;">true;"02/02/2022";"10/2/92 5:05 AM";598985994665542.25634;1;"Arjun";74785.155;5:05PM,1,Deepak,65152 false;"15/03/1923";"3/23/90 11:55 AM";3425699466554.2563;2;"sharma";5.1;5:45AM,2,Arjun,365273</pre> <ol style="list-style-type: none"> <li>Cliquez sur <b>Ajouter champ et Fermer</b>.</li> </ol> <p>Le champ ajouté et ses détails sont affichés dans la case.</p> <p style="padding-left: 40px;"><b>Remarque</b> : Si vous souhaitez que tout caractère d'espacement superflu soit supprimé au début et à la fin d'une chaîne de valeur dans un champ, cochez la case <b>Trim</b>.</p> <p><b>Modifier</b> : cliquez sur ce bouton pour modifier les détails d'un des champs de sortie ajoutés.</p> <p><b>Modifier</b> : cliquez sur ce bouton pour supprimer l'un des champs de sortie ajoutés.</p>

**Exécution** : ce bouton permet de spécifier plusieurs instances d'exécution de Field Parser. Cela se traduit par une amélioration significative des performances.

**OK** : cliquez sur ce bouton pour enregistrer toutes les informations saisies dans ce stage.

**Annuler** : cliquez sur ce bouton pour annuler toutes les mises à jour que vous avez apportées.

**Aide** : cliquez sur ce bouton pour consulter le fichier d'aide de ce stage.

## Field Combiner

Le stage **Field Combiner** combine les champs provenant du stage précédent dans un flux de données pour créer une chaîne XML.

### Options du stage Field Combiner

Nom de l'option	Description
Nom de colonne de sortie	Spécifiez le nom de la colonne contenant les champs combinés sous forme de chaîne XML.
Champs de sortie	<p>Cette section vous permet de sélectionner les champs à combiner et d'effectuer différentes actions sur ces champs.</p> <p><b>Remarque</b> : La liste déroulante <b>Appliquer l'espace de nommage</b> et les boutons <b>Élément</b>, <b>Attribut</b> et <b>Supprimer</b> sont activés quand vous sélectionnez un ou plusieurs éléments/attributs dans le volet de schéma.</p> <ul style="list-style-type: none"> <li>• <b>Ajout rapide</b> : cliquez sur ce bouton pour ouvrir la fenêtre contextuelle <b>Ajouter/Supprimer des champs</b>. Cette fenêtre affiche une liste de tous les champs provenant du stage précédent. Sélectionnez les champs à combiner.</li> </ul> <p><b>Remarque</b> : Les champs sélectionnés sont affichés dans le volet de schéma sous le champ <b>Rechercher un nœud</b>.</p> <ul style="list-style-type: none"> <li>• <b>Rechercher un nœud</b> : saisissez le nom du nœud auquel vous souhaitez accéder dans le volet de schéma. Le nœud saisi est mis en surbrillance et son chemin d'accès XML est affiché dans le champ <b>XPath</b> sous le volet.</li> <li>• <b>Élément</b> : cliquez sur ce bouton pour convertir un attribut en un élément de la chaîne XML.</li> <li>• <b>Élément</b> : cliquez sur ce bouton pour convertir un élément en un attribut de la chaîne XML.</li> <li>• <b>Appliquer l'espace de nommage</b> : si vous souhaitez spécifier un espace de nommage XML pour un élément ou un attribut, sélectionnez-le ici.</li> </ul> <p><b>Remarque</b> : Vous pouvez créer des espaces de nommage en cliquant sur le bouton <b>Espace de nommage</b> au-dessus. Pour plus d'informations, reportez-vous à la section <a href="#">Définition des espaces de nommage</a>.</p> <ul style="list-style-type: none"> <li>• <b>Supprimer</b> : cliquez sur ce bouton pour supprimer les éléments/attributs que vous ne souhaitez pas dans la chaîne XML.</li> </ul>

Nom de l'option	Description
XPath	<p>Cliquez n'importe où dans ce champ pour afficher le chemin d'accès XML (XPath) du nœud mis en surbrillance dans le volet de schéma. Pour afficher tous les XPath que vous avez précédemment affichés, cliquez sur la flèche en bas à droite du champ.</p> <p><b>Remarque :</b> XPath est un langage qui permet de trouver des informations dans un document XML. Pour plus d'informations, consultez <a href="http://www.w3schools.com/xpath/">www.w3schools.com/xpath/</a>.</p>

- **Exécution** : ce bouton permet de spécifier plusieurs instances d'exécution de Field Combiner. Cela se traduit par une amélioration significative des performances.
- **OK** : cliquez sur ce bouton pour enregistrer toutes les informations saisies dans ce stage.
- **Annuler** : cliquez sur ce bouton pour annuler toutes les mises à jour que vous avez apportées.
- **Aide** : cliquez sur ce bouton pour consulter le fichier d'aide de ce stage.

### Définition d'un espace de nommage

Les espaces de nommage vous permettent de disposer de noms d'élément et d'attribut en double dans votre sortie en assignant chaque élément ou attribut à un espace de nommage XML.

Pour définir un ou plusieurs espaces de nommage :

1. Sur l'écran **Options du stage Field Combiner**, cliquez sur le bouton **Espace de nommage**. La fenêtre contextuelle **Détails de l'espace de nommage** s'affiche.
2. Dans la colonne **Préfixe**, saisissez le préfixe à associer à un élément ou un attribut.
3. Dans la colonne **Espace de nommage**, indiquez l'URL de l'espace de nommage.
4. Répétez cette procédure pour définir autant d'espaces de nommage que vous le souhaitez.

## Field Selector

Field Selector vous permet de sélectionner les champs à transmettre au stage suivant du flux de données. Vous pouvez utiliser Field Selector pour supprimer les champs non désirés d'un flux de données. Par exemple, si vous avez créé un nouveau champ en combinant les données de deux champs et si vous n'avez plus besoin des deux champs sources, vous pouvez utiliser Field Selector pour conserver uniquement le nouveau champ et supprimer les deux champs sources du flux de données.

## Options

Option	Description
Sélectionnez les champs à envoyer à la phase suivante	Cochez la case en regard de chaque champ que vous souhaitez envoyer au stage suivant du flux de données. Pour ne pas envoyer un champ au stage suivant, décochez la case.
Sélectionner tout	Pour sélectionner tous les champs du flux de données, cochez cette case. Pour désélectionner tous les champs, décochez cette case.

## Generate Time Dimension

Generate Time Dimension crée des enregistrements de date, un pour chaque jour de la plage de dates que vous définissez. Vous pouvez ensuite écrire ces enregistrements dans une table de dimensions temporelles figurant dans une base de données via le stage Write to DB. La table de dimensions temporelles peut ensuite être utilisée pour réaliser des calculs précis en fonction d'une période temporelle. Par exemple, les ventes par trimestre, le budget dépensé par trimestre et le revenu par jour constituent tous des exemples d'analyse nécessitant une dimension temporelle. Les tables de dimensions temporelles vous permettent également de prendre en compte dans l'analyse les exercices financiers ou les trimestres non standard.

### Exemple d'utilisation d'une table de dimensions temporelles

Les tables de dimensions temporelles sont nécessaires pour réaliser des calculs temporels précis, car il s'avère parfois difficile d'extraire les données de date requises des enregistrements. Par exemple, les enregistrements suivants figurent dans une base de données de ventes. Notez qu'il existe des laps de temps entre les enregistrements. Par exemple, il n'existe aucun enregistrement pour la journée du 04/01/2012.

Date	Produit	Montant
03/01/2012	Chemise rouge	10,00 €
05/01/2012	Chemise rouge	5,00 €
07/01/2012	Chemise rouge	15,00 €

Si vous lancez une requête sur ces enregistrements et calculez les ventes moyennes par jour, la réponse est 10,00 € (30 €/3 enregistrements). Cependant, ce résultat est incorrect, car les trois enregistrements se sont en fait déroulés sur une période de cinq jours. Si vous disposez d'une table de dimensions temporelles comportant un enregistrement pour chaque jour, vous pouvez joindre cette table à la table ci-dessus pour obtenir le résultat suivant :

Date	Produit	Montant
03/01/2012	Chemise rouge	10,00 €
04/01/2012		
05/01/2012	Chemise rouge	5,00 €
06/01/2012		
07/01/2012	Chemise rouge	15,00 €

Si vous calculez les ventes moyennes par jour en fonction de ces enregistrements, vous obtenez la réponse correcte suivante : 6,00 € (30 €/5 jours).

Vous pouvez également prendre en compte dans vos calculs les attributs temporels arbitraires tels que les jours fériés, les weekends et les trimestres. Par exemple, si la date du 06/01/2012 est un jour férié et que seules les ventes moyennes par jour ouvré vous intéressent, la réponse est 7,50 €.

## Options

Generate Time Dimension comporte les options suivantes.

Option	Description
Date de début	Premier jour de la plage de dates dont vous souhaitez générer des enregistrements de dimensions temporelles.

Option	Description
Date de fin	Sélectionnez cette option si vous souhaitez définir une date de fin spécifique pour la dimension temporelle. Un enregistrement est généré pour chaque jour compris entre la date de début et la date que vous définissez ici.
Durée	Sélectionnez cette option si vous souhaitez que les dimensions temporelles couvrent un certain nombre de jours, de mois ou d'années. Un enregistrement est généré pour chaque jour de la durée. Par exemple, si vous indiquez une semaine, sept enregistrements sont générés, en commençant par le jour que vous avez défini dans le champ <b>Date de début</b> .

Option	Description
Attribut temps	

## Option

## Description

Option	Description
	Indique le type d'informations temporelles que vous souhaitez inclure dans la dimension temporelle. Chaque attribut est un champ de l'enregistrement de chaque jour. L'un des éléments suivants :
<b>Date</b>	Date du jour au format <Jour> <Mois> <Année>, avec le nom du mois dans la langue des paramètres régionaux du serveur. Par exemple, si le serveur est configuré en anglais, la date doit être lue comme suit : 30 October 2012.
<b>Jour du mois</b>	Nombre représentant le jour du mois. Par exemple, 10 indique qu'il s'agit du 10 <sup>ème</sup> jour du mois.
<b>Jour de l'année</b>	Nombre représentant le jour de l'année. Par exemple, 304 indique qu'il s'agit du 304ème jour de l'année.
<b>Est une année bissextile</b>	Valeur booléenne indiquant si le jour fait partie d'une année bissextile, auquel cas la valeur serait <code>true</code> , ou non, auquel cas la valeur serait <code>false</code> .
<b>Est un jour de la semaine</b>	Valeur booléenne indiquant si le jour est un jour de la semaine (du lundi au vendredi), auquel cas la valeur serait <code>true</code> , ou non, auquel cas la valeur serait <code>false</code> .
<b>Est un jour de week-end</b>	Valeur booléenne indiquant si le jour est un jour de week-end (samedi ou dimanche), auquel cas la valeur serait <code>true</code> , ou non, auquel cas la valeur serait <code>false</code> .
<b>Jour julien</b>	Nombre unique pour le jour. Le système de jours juliens ou calendaires considère le 1er janvier 4713 av. J.-C. comme le jour n° 1 et numérote tous les jours suivants dans l'ordre séquentiel. Par exemple, 2 456 231 indique qu'il s'agit du 2 456 231ème jour depuis le 1 <sup>er</sup> janvier 4713 av. J.-C.
<b>Semaine julienne</b>	Nombre unique pour la semaine de sept jours. Le système julien considère la première semaine de l'année 4713 av. J.-C. comme la semaine n° 1 et numérote toutes les semaines suivantes dans l'ordre séquentiel. Par exemple, 350 890 indique qu'il s'agit de la 350 890ème semaine depuis la première semaine de l'année 4713 av. J.-C.
<b>Année julienne</b>	Nombre unique pour l'année. Le système julien considère l'année 4713 av. J.-C. comme l'année n° 1 et numérote les années suivantes dans l'ordre séquentiel. Par exemple, 6 725 signifie que le jour fait partie de l'année 2012 ap. J.-C.
<b>Nom du mois</b>	Nom du mois du jour en anglais.
<b>Numéro du mois</b>	Numéro du mois de l'année. Par exemple, 3 signifie que le jour fait partie du troisième mois de l'année.
<b>Trimestre</b>	Nombre représentant le trimestre de l'année. Par exemple, 1 signifie que le jour fait partie du premier trimestre de l'année.
<b>Semaine de l'année</b>	Nombre représentant la semaine de l'année. Par exemple, 43 signifie que le jour fait partie de la 43ème semaine de

Option	Description
	<p>l'année.</p> <p><b>Nom du jour de la semaine</b> Nom du jour en anglais. Par exemple, Monday (lundi).</p> <p><b>Numéro du jour de la semaine</b> Nombre représentant le jour de la semaine, le jour n° 1 étant le lundi. Par exemple, mardi est le jour n° 2, mercredi le jour n°3, etc.</p> <p><b>Year</b> L'année du jour suivant le calendrier grégorien. Par exemple, 2012 signifie que le jour fait partie de l'année 2012.</p>
Champ	Nom du champ à créer dans le dataflow pour contenir l'attribut temporel. Un nom de champ par défaut est fourni, mais vous pouvez le modifier.
Type	Type de données de ce champ. Generate Time Dimension sélectionne automatiquement le type de données approprié pour chaque attribut temporel.
Calendrier	Indique s'il convient d'utiliser un calendrier personnalisé ou le calendrier grégorien par défaut pour le calcul de l'attribut temporel. Pour définir un calendrier personnalisé, cliquez sur le bouton <b>Calendrier</b> .

### Création d'un calendrier

Un calendrier définit les caractéristiques importantes de l'année en fonction du type d'analyse que vous souhaitez réaliser. Par défaut, le calendrier est le calendrier grégorien standard. Cependant, le calendrier grégorien standard n'est pas toujours approprié. Par exemple, si l'exercice financier de votre entreprise commence le 1er juin au lieu du 1er janvier, vous pouvez définir le mois de juin comme le premier mois de l'année, et l'attribut temporel « mois de l'année » indique juin comme le mois n° 1, juillet comme le mois n° 2, etc., et non pas janvier comme le mois n° 1, février comme le mois n° 2, etc.

- Réalisez l'une des actions suivantes :
  - Si vous configurez le stage Generate Time Dimension dans Enterprise Designer, cliquez sur **Calendriers**.
  - Dans Management Console, accédez à **Modules > Enterprise Data Integration > Outils > Calendriers**.
- Cliquez sur **Ajouter**.
- Dans le champ **Nom du calendrier**, donnez au calendrier un nom qui signifie quelque chose pour vous.



## Nom du champ

## Description

Nom du champ	Description
IsLeapYear	Valeur booléenne indiquant si le jour fait partie d'une année bissextile, auquel cas la valeur serait <code>true</code> , ou non, auquel cas la valeur serait <code>false</code> .
IsWeekday	Valeur booléenne indiquant si le jour est un jour de la semaine (du lundi au vendredi), auquel cas la valeur serait <code>true</code> , ou non, auquel cas la valeur serait <code>false</code> .
IsWeekend	Valeur booléenne indiquant si le jour est un jour de week-end (samedi ou dimanche), auquel cas la valeur serait <code>true</code> , ou non, auquel cas la valeur serait <code>false</code> .
JulianDay	Nombre unique pour le jour. Le système de jours juliens ou calendaires considère le 1er janvier 4713 av. J.-C. comme le jour n° 1 et numérote tous les jours suivants dans l'ordre séquentiel. Par exemple, 2 456 231 indique qu'il s'agit du 2 456 231ème jour depuis le 1 <sup>er</sup> janvier 4713 av. J.-C.
JulianWeek	Nombre unique pour la semaine de sept jours. Le système julien considère la première semaine de l'année 4713 av. J.-C. comme la semaine n° 1 et numérote toutes les semaines suivantes dans l'ordre séquentiel. Par exemple, 350 890 indique qu'il s'agit de la 350 890ème semaine depuis la première semaine de l'année 4713 av. J.-C.
JulianYear	Nombre unique pour l'année. Le système julien considère l'année 4713 av. J.-C. comme l'année n° 1 et numérote les années suivantes dans l'ordre séquentiel. Par exemple, 6 725 signifie que le jour fait partie de l'année 2012 ap. J.-C.
MonthName	Nom du mois du jour en anglais.
MonthNumber	Numéro du mois de l'année. Par exemple, 3 signifie que le jour fait partie du troisième mois de l'année.
Quarter	Nombre représentant le trimestre de l'année. Par exemple, 1 signifie que le jour fait partie du premier trimestre de l'année.
WeekOfYear	Nombre représentant la semaine de l'année. Par exemple, 43 signifie que le jour fait partie de la 43ème semaine de l'année.

Nom du champ	Description
WeekdayName	Nom du jour en anglais. Par exemple, Monday (lundi).
WeekdayNumber	Nombre représentant le jour de la semaine, le jour n° 1 étant le lundi. Par exemple, mardi est le jour n° 2, mercredi le jour n°3, etc.
Year	L'année du jour suivant le calendrier grégorien. Par exemple, 2012 signifie que le jour fait partie de l'année 2012.

## Query Cache

Query Cache effectue une recherche sur les données d'un cache en fonction des valeurs d'un ou de plusieurs champs de données et renvoie les données des enregistrements correspondants du cache, en les ajoutant à l'enregistrement du flux de données. La recherche de données dans un cache peut s'avérer plus performante que celle de données dans une base de données.

Il existe deux types de cache : les caches globaux et les caches locaux.

### *Options de cache global*

Un cache global est un cache partagé à l'échelle du système qui réside en mémoire. Si vous souhaitez que le cache soit disponible pour différents flux de données, lorsque les données ne changent pas souvent ou restent relativement statiques ou encore lorsque l'espace de stockage n'est pas limité, optez pour un cache global. Un cache global est statique, car vous ne pouvez y écrire qu'une seule fois. Une fois créé, le cache ne peut pas être mis à jour.

Un cache global est créé par le stage Write to Cache. Avant d'utiliser un cache global, vous devez le renseigner à l'aide des données sur lesquelles vous souhaitez effectuer des recherches. Pour ce faire, créez un flux de données contenant le stage **Write to Cache**.

Nom de l'option	Description
Cache type	Sélectionnez l'option Cache global.

Nom de l'option	Description
Nom du cache	Indique le cache que vous souhaitez interroger. Pour créer un cache, utilisez le stage Write to Cache.
Champs de cache	Cette colonne contient une liste des champs du cache. Vous ne pouvez pas modifier les noms de ces champs.
Champs de stage	Cette colonne contient une liste des noms de champ utilisés dans le flux de données. Pour modifier le nom d'un champ, cliquez dessus et saisissez un nouveau nom.
Type	Cette colonne contient une liste des types de données de chaque champ de flux de données.
Include	Pour que la requête renvoie la valeur du champ de cache, cochez la case de cette colonne. Pour que la requête ne renvoie pas le champ de cache, décochez la case.

Nom de l'option Description

Valeur d'erreur par défaut Indique la valeur à afficher dans le champ de flux de données en cas d'échec de la requête. La liste déroulante affiche les valeurs valides correspondant au type de données du champ interrogé. Par exemple, en cas d'**entier**, l'option affichée est **-1**. Vous pouvez également saisir une valeur dans ce champ. Voir le tableau ci-dessous pour obtenir une liste des valeurs d'erreur par défaut valides pour différents types de données.

Type de données Valeur d'erreur par défaut valide ainsi que le type de données (entre parenthèses)

Type de données	Null	-1	1899-12-30	1899-12-30	12:00:00	False	Vide
Date							
Entier		<input type="checkbox"/>					
Long		<input type="checkbox"/>					
Flottant		<input type="checkbox"/>					
Grand Decimal		<input type="checkbox"/>					
Double		<input type="checkbox"/>					
Chaîne	<input type="checkbox"/>						
Durée					<input type="checkbox"/>		
Date/Heure			<input type="checkbox"/>				
Booléen						<input type="checkbox"/>	

Nom de l'option	Description
Champ Clé	Indique le champ du cache à utiliser comme clé de recherche. Si la valeur du champ de la colonne <b>Champ d'entrée</b> correspond à la valeur du champ de clé du cache, la requête renvoie les données de cet enregistrement dans le cache.
Champ de saisie	Indique le champ de flux de données dont la valeur va être utilisée comme clé. Si la valeur de ce champ correspond à la valeur du champ de clé du cache, la requête renvoie les données de cet enregistrement dans le cache.

### Options de cache local

Un cache local est un cache temporaire uniquement utilisé lors de l'exécution du stage Query Cache. Query Cache construit le cache à partir de la table de base de données de votre choix. Il recherche ensuite les données du cache en fonction des champs de clé et des conditions de recherche et renvoie les données des enregistrements correspondants figurant dans le cache, en les ajoutant à l'enregistrement du flux de données.

Un cache local est dynamique, car il est créé lors de l'exécution d'un job de Query Cache. Une fois que Query Cache a terminé la lecture des données, le cache est automatiquement supprimé de la mémoire. Un cache local est recréé à chaque exécution du stage Query Cache. Si le cache va être utilisé dans un seul flux de travail ou si la table objet de la recherche change fréquemment, optez pour un cache local.

Nom de l'option	Description
Cache type	Indique l'option Cache local.
Connexion	<p>Sélectionnez la connexion de base de données à utiliser. Vos choix varient en fonction des connexions qui sont définies dans le Gestionnaire de connexion de Management Console. Si vous devez établir une nouvelle connexion à la base de données, ou en modifier ou supprimer une existante, cliquez sur <b>Gérer</b>.</p> <p>Si vous ajoutez ou modifiez une connexion de base de données, remplissez ces champs :</p> <p><b>Nom de la connexion</b> Saisissez le nom de la connexion.: Ce nom est entièrement de votre choix.</p> <p><b>Pilote de la base de données</b> Sélectionnez le type de base de données approprié.</p> <p><b>Options de connexion</b> Indiquez l'hôte, le port, l'instance, le nom d'utilisateur et le mot de passe à utiliser pour vous connecter à la base de données.</p>
Table/Vue	Indiquez la table ou la vue de la base de données que vous souhaitez interroger.

Nom de l'option	Description
Champs de la base de données	Cette colonne contient une liste des champs de la base de données. Vous ne pouvez pas modifier les noms de ces champs.
Champs de stage	Cette colonne contient une liste des noms de champ utilisés dans le flux de données. Pour modifier le nom d'un champ, cliquez dessus et saisissez le nouveau nom.
Type	Cette colonne contient une liste des types de données de chaque champ de flux de données.
Include	Pour que la requête renvoie la valeur du champ de cache, cochez la case de cette colonne. Pour que la requête ne renvoie pas le champ de cache, décochez la case.

Nom de l'option      Description

Valeur d'erreur par défaut      Indique la valeur à afficher dans le champ de flux de données en cas d'échec de la requête. La liste déroulante affiche les valeurs valides correspondant au type de données du champ interrogé. Par exemple, en cas d'**entier**, l'option affichée est **-1**.  
  
Vous pouvez également saisir une valeur dans ce champ. Voir le tableau ci-dessous pour obtenir une liste des valeurs d'erreur par défaut valides pour différents types de données.

Type de données      Valeur d'erreur par défaut valide ainsi que le type de données (entre parenthèses)

Type de données	Valeur d'erreur par défaut valide ainsi que le type de données (entre parenthèses)
Null	
Entier	-1 (Entier)
Long	1899-12-30 1899-12-30 12:00:00 (Date/Heure)
Flottant	
Grand Decimal	
Double	
Chaîne	
Durée	
Date/Heure	
Booléen	

Nom de l'option	Description
Champ Clé	Indique le champ de la base de données à utiliser comme clé de recherche. Si la valeur du champ de la colonne <b>Champ d'entrée</b> correspond à la valeur de <b>Champ clé</b> dans la base de données, la requête renvoie les données de cet enregistrement de la base de données.
Champ de saisie	Indique le champ de flux de données dont la valeur va être utilisée comme clé. Si la valeur de ce champ correspond à la valeur de <b>Champ clé</b> dans la base de données, la requête renvoie les données de cet enregistrement de la base de données.

### Options de cache avancé

Un cache avancé est un cache temporaire semblable au cache local. Il est utilisé lors de l'exécution du stage Query Cache. Il construit le cache en fonction de la requête SQL qui lit les données des tables mentionnées dans la requête. Il recherche ensuite les données du cache en fonction des clés de recherche mentionnées dans la clause where et renvoie les données des enregistrements correspondants figurant dans le cache, en ajoutant les données de l'enregistrement en cache à l'enregistrement du flux de données.

Un cache avancé est dynamique, car il est créé lors de l'exécution d'un job de Query Cache. Une fois que Query Cache a terminé la lecture des données, le cache est automatiquement supprimé de la mémoire. Un cache avancé est recréé à chaque exécution du stage Query Cache. Sélectionnez une option de cache avancé en cas de lecture des données de plusieurs tables et de besoins en requête complexes à exécuter pour la création de cache.

Nom de l'option	Description
Cache type	Indique l'option Cache avancé.
Connexion	<p>Sélectionnez la connexion de base de données à utiliser. Vos choix varient en fonction des connexions qui sont définies dans le Gestionnaire de connexion de Management Console. Si vous devez établir une nouvelle connexion à la base de données, ou en modifier ou supprimer une existante, cliquez sur <b>Gérer</b>.</p> <p>Si vous ajoutez ou modifiez une connexion de base de données, remplissez ces champs :</p> <p><b>Connection Name</b> Saisissez le nom de la connexion.: Ce nom est entièrement de votre choix.</p> <p><b>Database Driver</b> Sélectionnez le type de base de données approprié.</p> <p><b>Connection Options</b> Indiquez l'hôte, le port, l'instance, le nom d'utilisateur et le mot de passe à utiliser pour vous connecter à la base de données.</p>

Nom de l'option	Description
Requête	Fournit la requête SQL permettant de lire les données de la base de données. La requête peut lire les données de plusieurs tables.
Où	Ce texte est utilisé comme clause where pour rechercher le cache créé en fonction de la requête. L'utilisateur peut spécifier le champ d'entrée dans la requête à l'aide de l'opérateur \$ comme préfixe. Par exemple, <code>_id = \${_inputId}</code> , où <code>_inputId</code> est le champ d'entrée et <code>_id</code> la colonne de recherche du cache.
Obtenir les champs	Cette opération renseigne la grille avec les champs qui sont sélectionnés pour être mis en cache à l'aide de la requête SQL.
Champs de la base de données	Cette colonne contient une liste des champs récupérés dans la base de données. Vous ne pouvez pas modifier les noms de ces champs.
Champs de stage	Cette colonne contient une liste des noms de champ utilisés dans le flux de données. Pour modifier le nom d'un champ, cliquez dessus et saisissez le nouveau nom.
Type	Cette colonne contient une liste des types de données de chaque champ de flux de données.

Nom de l'option      Description

Valeur d'erreur par défaut      Indique la valeur à afficher dans le champ de flux de données en cas d'échec de la requête. La liste déroulante affiche les valeurs valides correspondant au type de données du champ interrogé. Par exemple, en cas d'**entier**, l'option affichée est **-1**.  
  
Vous pouvez également saisir une valeur dans ce champ. Voir le tableau ci-dessous pour obtenir une liste des valeurs d'erreur par défaut valides pour différents types de données.

Type de données      Valeur d'erreur par défaut valide ainsi que le type de données (entre parenthèses)

Type de données	Valeur d'erreur par défaut valide ainsi que le type de données (entre parenthèses)
Null	-1 (Entier)
1899-12-30	1899-12-30 (Date)
1899-12-30 12:00:00	12:00:00 (Heure)
False	
Vide	
Date	
Entier	<input type="text" value="-1"/>
Long	<input type="text" value=""/>
Flottant	<input type="text" value=""/>
Grand Decimal	<input type="text" value=""/>
Double	<input type="text" value=""/>
Chaîne	<input type="text" value=""/>
Durée	<input type="text" value=""/>
Date/Heure	<input type="text" value=""/>
Booléen	<input type="text" value=""/>

## L'onglet Exécution

Les options disponibles dans l'onglet Exécution sont communes aux caches globaux, locaux et avancés.

Nom de l'option	Description
Match options	<p>Indique comment procéder s'il existe plusieurs enregistrements du cache correspondant à la requête. L'un des éléments suivants :</p> <p><b>Renvoyer toutes les correspondances</b> Renvoie les données de tous les enregistrements du cache dont la valeur correspond à celle du ou des champs de clé.</p> <p><b>Renvoyer le premier enregistrement correspondant</b> Renvoie uniquement les données du premier enregistrement du cache dont la valeur correspond à celle du ou des champs de clé.</p> <p><b>Renvoyer le dernier enregistrement correspondant</b> Renvoie uniquement les données du dernier enregistrement du cache dont la valeur correspond à celle du ou des champs de clé.</p>
Options de stage	<p>Cette section répertorie les options de dataflow utilisées dans la requête SQL de ce stage et vous permet de fournir une valeur par défaut pour toutes ces options. La colonne <b>Nom</b> répertorie les options, tandis que vous pouvez saisir les valeurs par défaut dans la colonne <b>Valeur</b> correspondante.</p> <p><b>Remarque :</b> La valeur par défaut fournie ici est également affichée dans la section <b>Map dataflow options to stages</b> de la boîte de dialogue <b>Options de dataflow</b>. La boîte de dialogue vous permet également de modifier la valeur par défaut. En cas de conflit des valeurs par défaut fournies pour une option via <b>Options de stage</b>, <b>Options de dataflow</b> et <b>Job Executor</b>, l'ordre de priorité est le suivant : valeur fournie via <b>Job Executor</b> &gt; valeur définie via la boîte de dialogue <b>Options de dataflow</b> &gt; valeur saisie via <b>Options de stage</b>.</p>

## Query DB

Le stage Query DB vous permet de vous servir des champs comme paramètres dans votre requête de base de données et de renvoyer les résultats de la requête comme des nouveaux champs dans le flux de données.

**Remarque :** Si vous voulez interroger une base de données spatiales, utilisez Query Spatial Data au lieu de Query DB.

## Onglet Général

Option	Description
Connexion	<p>Sélectionnez la connexion de base de données à utiliser. Vos choix varient en fonction des connexions qui sont définies dans le Gestionnaire de connexion de Management Console. Si vous devez établir une nouvelle connexion à la base de données, ou en modifier ou supprimer une existante, cliquez sur <b>Gérer</b>.</p> <p>Si vous ajoutez ou modifiez une connexion de base de données, remplissez ces champs :</p> <p><b>Nom de la connexion</b>      Saisissez le nom de la connexion.: Ce nom est entièrement de votre choix.</p> <p><b>Pilote de la base de données</b>      Sélectionnez le type de base de données approprié.</p> <p><b>Options de connexion</b>      Indiquez l'hôte, le port, l'instance, le nom d'utilisateur et le mot de passe à utiliser pour vous connecter à la base de données.</p>
Table/Vue	Indiquez la table ou la vue de la base de données que vous souhaitez interroger.

Option	Description
Où	<p>Si vous souhaitez utiliser une instruction <code>WHERE</code>, saisissez-la ici. Notez que vous ne devez pas, en réalité, inclure le terme <code>WHERE</code> dans l'instruction. L'objectif d'une instruction <code>WHERE</code> est de ne renvoyer que les données des enregistrements qui correspondent à la condition que vous indiquez.</p> <p>Pour spécifier une valeur d'un champ de flux de données, utilisez la syntaxe suivante :</p> <pre data-bbox="557 569 769 594">\${&lt;field name&gt;}</pre> <p>Où <code>&lt;field name&gt;</code> est le nom d'un champ du flux de données.</p> <p>Par exemple :</p> <pre data-bbox="557 716 1094 747">account_number=\${customer_key}</pre> <p>Dans cet exemple, la requête renvoie des données des enregistrements dont la valeur de la colonne de la table <code>account_number</code> correspond à la valeur du champ de flux de données <code>customer_key</code>.</p> <p><b>Remarque</b> : Si vous interrogez une base de données sensible à la casse, assurez-vous de saisir le nom de champ au même format que celui utilisé dans la table de base de données. En d'autres termes, placez le nom de champ entre guillemets doubles (") s'il l'était lors de la création de la table.</p> <p>Cliquez sur <b>Aperçu</b> pour voir un aperçu des données (les 50 premiers enregistrements) selon les critères que vous avez définis.</p> <p><b>Remarque</b> : La fonction d'aperçu dans Query DB ne fonctionne pas si vous utilisez un champ de flux de données dans l'instruction <code>WHERE</code>. Au lieu de cela, vous pouvez prévisualiser le résultat à l'aide de l'outil d'inspection du flux de données dans Enterprise Designer.</p>
Renvoyer les enregistrements sans résultat	<p>Cochez cette case pour que les enregistrements dont les requêtes ne renvoient aucun résultat soient toujours renvoyés par Query DB. Si vous décochez cette case, l'enregistrement ne sera pas renvoyé. Nous vous recommandons de laisser cette option cochée.</p>
Include	<p>Dans la table de champs, sélectionnez les champs que vous désirez inclure en cliquant sur la case <b>Inclure</b> en regard du champ.</p>

### Onglet Tri

Si vous voulez trier les enregistrements selon la valeur d'un champ, indiquez les champs que vous voulez utiliser pour le tri.

## Configuration du stage Query DB lors de l'exécution

Vous pouvez configurer le stage Query DB de sorte que les valeurs de la clause `WHERE` soient définies lors de l'exécution. Cette opération s'avère utile dans les cas où vous souhaitez permettre la configuration du nom de colonne de la clause `WHERE` via les Options de flux de données.

1. Ouvrez le flux de données dans Enterprise Designer.
2. Configurez les champs de nom **Connexion** et **Table/Vue** pour qu'ils pointent vers la base de données devant faire l'objet de la requête.
3. Dans le champ **Where**, saisissez une instruction `WHERE` au format suivant pour les valeurs que vous souhaitez configurer : `${paramètre}`.

Par exemple :

```
${COL}=${EmployeeID}
```

Ici, COL représente une option Flux de données qui sera renseignée par le nom de colonne de la table lors de l'exécution.

4. Fermez la fenêtre d'options Query DB.
5. Cliquez sur l'icône Options de flux de données de la barre d'outils ou sur **Édition > Options de flux de données**. La fenêtre **Options de flux de données** apparaît.
6. Cliquez sur **Ajouter**. La fenêtre **Définir les options de flux de données** apparaît.
7. Sélectionnez le stage Query DB.
8. Précisez les valeurs de **Nom d'option** et de **Libellé d'option**.

La valeur du champ **Nom d'option** doit être identique à la valeur saisie au format `${paramètre}` dans la clause `WHERE`. Dans le champ **Libellé d'option**, vous pouvez indiquer un autre libellé ou conserver la même valeur que celle du champ **Nom d'option**.

Par exemple : COL.

9. Indiquez la **Valeur par défaut**. Par exemple : EmpID.
10. Cliquez sur **OK**.

Cette procédure mappe le nom de colonne réel de la base de données, à savoir, EmpID, au nom d'option à l'exécution, COL. Le nom de colonne de la base de données doit être correctement placé entre guillemets avec l'identifiant de propre à la base de données spécifique.

## Requête NoSQL DB

Le stage **Requête NoSQL DB** permet des recherches de données requises à partir d'une base de données NoSQL. Le stage prend en charge les bases de données MongoDB.

### Onglet Général

Nom du champ	Description
Connexion	<p>Sélectionnez la connexion à la base de données requise dans la liste déroulante. Les options affichées sont basées sur les connexions définies dans Management Console.</p> <p>Pour ajouter une nouvelle connexion, voir <a href="#">Connexion à NoSQL</a> à la page 54.</p> <p>Pour modifier une connexion existante, sélectionnez-la et ouvrez-la dans la liste des connexions de la page Sources de données de Management Console, effectuez les mises à jour requises, puis cliquez sur le bouton <b>Enregistrer</b>.</p>
Table/Vue	<p>Indique la collection ou la vue de la base de données que vous souhaitez interroger.</p> <p><b>Remarque</b> : Dans une base de données MongoDB, une table/vue est appelée une <i>collection</i>.</p>
Fichier de schéma	<p>Cliquez sur le bouton Parcourir (...) pour sélectionner un fichier de schéma JSON. Ce fichier est facultatif. Les champs de l'onglet Champs peuvent être régénérés via le fichier de schéma ou la table/vue de base de données.</p> <p>Pour effacer le chemin d'accès au fichier sélectionné, cliquez sur <b>Effacer</b>.</p> <p><b>Remarque</b> : Si un fichier de schéma est sélectionné, les champs sont toujours générés via ce fichier de schéma.</p>
Où	<p>Saisissez la clause <code>where</code> pour définir les critères de la recherche.</p> <p>Pour obtenir la liste de tous les opérateurs inclus dans la clause <code>WHERE</code>, consultez la section <a href="http://docs.mongodb.org/manual/reference/operator/query/">http://docs.mongodb.org/manual/reference/operator/query/</a>.</p>
Aperçu	<p>Affiche les enregistrements de la table ou de la vue sélectionnée.</p> <p><b>Remarque</b> : Cliquer sur <b>Aperçu</b> extrait les 50 premiers enregistrements de la base de données sélectionnée, sans appliquer les critères de filtre spécifiés dans le champ <b>Où</b>.</p>

Nom du champ	Description
Développer tout	Développe les éléments de l'arborescence d'aperçu.
Réduire tout	Réduit les éléments de l'arborescence d'aperçu.

### Onglet Champs

L'onglet **Champs** vous permet de sélectionner les données que vous souhaitez transmettre au stage suivant. Pour plus d'informations, reportez-vous à la section [Définition de champs - Requête NoSQL DB](#) à la page 137.

## Définition de champs - Requête NoSQL DB

L'onglet **Champs** affiche le champ et son type, tels que définis dans le fichier de schéma/NoSQL DB.

1. Dans l'onglet **Champs**, cliquez sur **Régénérer**.

Cette opération génère les données agrégées en fonction des 50 premiers enregistrements. Les données sont affichées au format suivant : `Fieldname (datatype)`.

**Remarque** : Si le fichier de schéma est localisé, les champs sont générés via le fichier de schéma. La table/vue est ignorée. Pour réinitialiser le fichier de schéma, cliquez sur **Effacer**.

2. Pour modifier le nom et le type d'un champ, mettez le champ en surbrillance, puis cliquez sur **Modifier**.
3. Dans le champ **Name**, choisissez le champ à ajouter ou saisissez le nom du champ.
4. Dans le champ **Type**, vous pouvez laisser le type de données défini sur chaîne si vous ne prévoyez pas d'effectuer des opérations mathématiques sur les données. Cependant, si vous avez l'intention d'effectuer ce genre d'opérations, sélectionnez un type de données approprié. Cela convertira les données de chaîne à partir du fichier en un type de données qui activera la manipulation appropriée des données dans le flux de données.

Le stage prend en charge les types de données suivants :

<b>double</b>	Un type de données numérique qui contient à la fois des nombres en double précision négatifs et positifs entre $2^{-1074}$ et $(2-2^{-52}) \times 2^{1023}$ . En notation E, la plage de valeurs est comprise entre -1.79769313486232E+308 et 1.79769313486232E+308.
<b>flottant</b>	Un type de données numérique contenant des nombres de précision seuls négatifs et positifs entre $2^{-149}$ et $(2-2^{-23}) \times 2^{127}$ . En notation E, la plage de valeurs est comprise entre -3.402823E+38 et 3.402823E+38.

<b>integer</b>	Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre $-2^{31}$ (-2,147,483,648) et $2^{31}-1$ (2,147,483,647).
<b>long</b>	Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre $-2^{63}$ (-9,223,372,036,854,775,808) et $2^{63}-1$ (9,223,372,036,854,775,807).
<b>chaîne</b>	Séquence de caractères.

- Vous pouvez également ajouter des champs supplémentaires non présents dans le fichier de schéma ou la table. Cliquez sur **Ajouter** pour ajouter un nouveau champ. Pour supprimer un champ, cliquez sur **Supprimer**.

**Remarque :** Vous ne pouvez ajouter un nouveau champ que sous le type de liste.

- Cliquez sur **OK**.

## Configuration des options de flux de données - Requête NoSQL DB

Cette procédure explique comment configurer un flux de données pour qu'il prenne en charge les options d'exécution **Requête NoSQL DB**.

- Ouvrez le flux de données dans Enterprise Designer.
- Pour configurer les options d'exécution d'un stage dans un flux de données intégré, ouvrez ce dernier.
- Cliquez sur l'icône Options de flux de données de la barre d'outils ou sur **Édition > Options de flux de données**. La boîte de dialogue **Options de flux de données** apparaît.
- Cliquez sur **Ajouter**. La boîte de dialogue **Définir l'option de flux des données** apparaît.
- Développez le stage **Requête NoSQL DB**.
- Les options de flux de données sont exposées pour interroger une base de données Mongo DB :
  - Connexion
  - Table

Le nom d'option **Query NoSQL DB** sélectionné s'affiche dans les champs **Nom de l'option** et **Libellé d'option**. Il s'agit du nom de l'option qui devra être spécifié lors de l'exécution afin de définir cette option.

- Saisissez une description de l'option dans le champ **Description**.
- Dans le champ **Target**, sélectionnez l'option **Selected stage(s)**.
- Si vous voulez limiter les valeurs qu'un utilisateur peut sélectionner, modifiez les options disponibles dans le champ **Legal values** en cliquant sur l'icône située directement à droite de ce champ.

10. Pour modifier la valeur par défaut, indiquez une valeur différente dans le champ **Valeur par défaut**.

**Remarque :** Pour un service, vous ne pouvez modifier que les valeurs par défaut avant d'exposer le service pour la première fois. Une fois que vous exposez le service, vous ne pouvez plus modifier les valeurs par défaut à l'aide d'Enterprise Designer. Au lieu de cela, vous devez utiliser Management Console. Pour plus d'informations, reportez-vous à la section [Spécification des Options de services par défaut](#).

11. Cliquez sur **OK**.
12. Continuez d'ajouter les options souhaitées.
13. Cliquez sur **OK** dans la boîte de dialogue des **Options du flux de données** lorsque vous avez terminé d'ajouter vos options.
14. Enregistrez et exposez le flux de données.

## Read From DB

Le stage **Read From DB** lit des données d'une table ou d'une vue de base de données sous forme d'entrée d'un dataflow. Le stage est disponible pour les jobs, les services et les sous-flux, mais pas pour les processflow.

### Onglet Général

Nom du champ	Description						
Connexion	<p>Sélectionnez la connexion de base de données à utiliser. Vos choix varient en fonction des connexions qui sont définies dans le Gestionnaire de connexion de Management Console. Si vous devez établir une nouvelle connexion à la base de données, ou en modifier ou supprimer une existante, cliquez sur <b>Gérer</b>.</p> <p>Si vous ajoutez ou modifiez une connexion de base de données, remplissez ces champs :</p> <table border="0"> <tr> <td><b>Nom de la connexion</b></td> <td>Saisissez le nom de la connexion.: Ce nom est entièrement de votre choix.</td> </tr> <tr> <td><b>Pilote de la base de données</b></td> <td>Sélectionnez le type de base de données approprié.</td> </tr> <tr> <td><b>Options de connexion</b></td> <td>Indiquez l'hôte, le port, l'instance, le nom d'utilisateur et le mot de passe à utiliser pour vous connecter à la base de données.</td> </tr> </table>	<b>Nom de la connexion</b>	Saisissez le nom de la connexion.: Ce nom est entièrement de votre choix.	<b>Pilote de la base de données</b>	Sélectionnez le type de base de données approprié.	<b>Options de connexion</b>	Indiquez l'hôte, le port, l'instance, le nom d'utilisateur et le mot de passe à utiliser pour vous connecter à la base de données.
<b>Nom de la connexion</b>	Saisissez le nom de la connexion.: Ce nom est entièrement de votre choix.						
<b>Pilote de la base de données</b>	Sélectionnez le type de base de données approprié.						
<b>Options de connexion</b>	Indiquez l'hôte, le port, l'instance, le nom d'utilisateur et le mot de passe à utiliser pour vous connecter à la base de données.						

Nom du champ	Description
SQL	<p>Saisissez la requête SQL spécifiant les enregistrements à lire à partir de la source de données lors de l'exécution du dataflow. Vous pouvez saisir manuellement la requête SQL dans ce champ. Sinon, utiliser Visual Query Builder pour construire la requête en cliquant sur <b>Créer SQL....</b></p> <p>La requête SQL peut inclure des variables au lieu de noms de colonne réels. L'utilisation de variables vous permet de personnaliser la requête lors de l'exécution. Pour plus d'informations, reportez-vous à la section <a href="#">Variables de requête</a> à la page 145.</p>
Créer SQL...	<p>Créez une requête complexe en sélectionnant plusieurs colonnes et en créant des jointures et des requêtes imbriquées en cliquant sur <b>Créer SQL. Visual Query Builder</b> s'ouvre. Pour plus d'informations, reportez-vous à la section <a href="#">Visual Query Builder</a> à la page 141.</p> <p><b>Remarque :</b> Une requête créée à l'aide de Visual Query Builder s'affiche avec les noms complets des colonnes et des tables dans le champ <b>SQL</b>.</p>
Régénérer les champs	<p>Pour afficher le schéma des données à extraire par la requête, cliquez sur <b>Régénérer les champs</b>.</p> <p>Si vous modifiez une requête existante, cliquez sur <b>Régénérer les champs</b> pour récupérer le schéma modifié.</p> <p><b>Remarque :</b> Si vous cliquez sur <b>Régénérer les champs</b>, les noms d'entité de la requête SQL sont conservés et non remplacés par leurs noms complets.</p>
Aperçu	<p>Pour voir un échantillon des enregistrements récupérés par la requête SQL, cliquez sur <b>Prévisualisation</b>.</p>

**Remarque :** Le stage **Read From DB** vous permet de modifier le type d'un champ d'entrée.

**Remarque :** Le stage **Read from DB** lit toutes les valeurs du type de données `date` sous forme de valeurs `String`. Il s'agit du comportement du pilote *jTDS*, qui est le pilote par défaut utilisé par Spectrum. Pour gérer toutes les valeurs du type de données `date` en l'état, utilisez le pilote JDBC de Microsoft.

## L'onglet Exécution

Nom du champ	Description
Taille de la récupération :	<p>Sélectionnez cette option pour indiquer le nombre d'enregistrements à lire en une seule fois depuis la table de base de données. Par exemple, si la valeur de <b>Taille de l'extraction</b> est 100 et que le nombre total d'enregistrements à lire est 1 000, la base de données lit l'ensemble des enregistrements en 10 fois.</p> <p>La définition d'une valeur Taille de l'extraction optimale peut considérablement améliorer les performances.</p> <p><b>Remarque :</b> Vous pouvez calculer une taille d'extraction optimale pour votre environnement en testant les temps d'exécution entre un stage Read from DB et un stage Write to Null. Pour plus d'informations, reportez-vous à la section <a href="#">Définition d'une taille d'extraction optimale</a> à la page 316.</p>
Options de stage	<p>Cette section répertorie les options de dataflow utilisées dans la requête SQL de ce stage et vous permet de fournir une valeur par défaut pour toutes ces options. La colonne <b>Nom</b> répertorie les options, tandis que vous pouvez saisir les valeurs par défaut dans la colonne <b>Valeur</b> correspondante.</p> <p><b>Remarque :</b> La valeur par défaut fournie ici est également affichée dans la section <b>Map dataflow options to stages</b> de la boîte de dialogue <b>Options de dataflow</b>. La boîte de dialogue vous permet également de modifier la valeur par défaut. En cas de conflit des valeurs par défaut fournies pour une option via <b>Options de stage</b>, <b>Options de dataflow</b> et <b>Job Executor</b>, l'ordre de priorité est le suivant : valeur fournie via <b>Job Executor</b> &gt; valeur définie via la boîte de dialogue <b>Options de dataflow</b> &gt; valeur saisie via <b>Options de stage</b>.</p>

## Visual Query Builder

Visual Query Builder fournit une interface visuelle pour la création de requêtes SQL complexes dans le stage Read from DB. Pour utiliser Visual Query Builder, vous avez besoin de connaissances élémentaires à propos des concepts SQL.

Pour accéder à Visual Query Builder, cliquez sur le bouton **Créer SQL** dans Read from DB.

La fenêtre principale du générateur de requêtes est divisée en les parties suivantes :

- La **Zone de création de requête** est la zone principale dans laquelle la représentation visuelle de la requête est affichée. Cette zone vous permet de définir les objets de base de données source, de définir les liens entre eux et de configurer les propriétés des tables et des liens.

- Le **Volet Colonnes** figure sous la zone de création de requête. Il permet d'effectuer toutes les opérations nécessaires avec les colonnes et expressions de sortie de requête. Vous pouvez y définir les alias, le tri et le regroupement des champs, ainsi que des critères.
- La commande de page au-dessus de la zone de création de requête vous permet de basculer entre la requête principale et les sous-requêtes.

### Ajout d'objets à une requête

Pour ajouter un objet à une requête, utilisez l'arborescence **Objets de base de données**. Les objets dans l'arborescence sont regroupés par base de données, schéma et type. Glissez l'objet que vous souhaitez ajouter à la requête et déposez-le dans la zone de création de requête. Sinon, vous pouvez également double-cliquer sur l'objet pour l'ajouter à la requête.

### Définition des alias d'objet

Pour définir un alias pour un objet ou une table dérivée dans la requête, double-cliquez l'objet et sélectionnez **Propriétés**. La boîte de dialogue **Propriétés de la source de données** apparaît. Elle peut contenir d'autres options de source de données propres au serveur, mais la propriété Alias est la même pour tous les serveurs de base de données.

### Jointure de tables

Lorsque deux objets référencés par une relation de clé étrangère sont ajoutés à la requête, ils sont joints automatiquement avec la commande INNER JOIN. Pour les serveurs qui ne prennent pas en charge la clause JOIN, le générateur de requête ajoute une condition à la partie WHERE de la requête.

Pour joindre deux objets manuellement, sélectionnez le champ qui doit lier l'objet à un autre et, faites-le glisser dans le champ correspondant de l'autre objet. Une fois le glissement terminé, une ligne associant les champs liés apparaît. Des symboles de cardinalité clés sont placés aux extrémités du lien lorsqu'une relation correspondante existe dans la base de données.



Pour enlever un lien entre des objets, double-cliquez sur le lien et sélectionnez **Enlever**.

Pour modifier le type de jointure, double-cliquez sur la ligne du lien.

### Sélection de champs de sortie

Pour ajouter un champ à la liste des champs de sortie de requête, cochez la case située à gauche du nom de champ dans la liste des champs de la source de données, dans la zone **Création de requête**. Pour inclure tous les champs de l'objet, cochez la case située à gauche de l'astérisque dans la liste de champs de la source de données. Vous pouvez également y faire glisser des champs, de la zone **Création de requête** au volet **Colonnes** pour obtenir le même résultat.

Si vous ne sélectionnez aucun champ dans les sources de données de requête, un astérisque est ajouté à la liste de sélection de la requête obtenue (« Sélectionner \* dans ... »). Cela est dû au fait

qu'une requête SELECT sans aucune colonne engendre une erreur pour la plupart des serveurs de base de données. L'astérisque est enlevé de la requête si vous sélectionnez un champ ou que vous ajoutez une expression de sortie à la requête.

**Conseil** : Un autre moyen d'ajouter un champ consiste à sélectionner un nom de champ dans la liste déroulante de la colonne Expression dans le volet **Colonnes**. Vous pouvez également taper une expression valide dans la colonne **Expression** du volet **Colonnes**. Pour insérer une liste vide au volet **Colonnes**, appuyez sur les touches Alt+Inser.

Pour enlever un champ du volet **Colonnes**, désélectionnez la case située à gauche du nom de champ dans la zone **Création de requête** ou appuyez sur les touches Alt+Suppr dans le volet **Colonnes**.

Pour déplacer une ligne vers le haut, appuyez sur les touches Alt+Haut. Pour déplacer une ligne vers le bas, appuyez sur les touches Alt+Bas.

Pour enlever une expression de la liste SELECT de la requête, désélectionnez la case dans la colonne **Sortie**.

Pour définir un alias pour une expression, saisissez l'alias dans la colonne **Alias**. Les alias deviennent les en-têtes des colonnes dans le jeu de données obtenu.

### Tri d'un jeu de données

Pour trier le jeu de données obtenu, utilisez les colonnes Type de tri et Ordre de tri du volet **Colonnes**. La colonne Type de tri vous permet de procéder au tri par ordre croissant ou décroissant. La colonne Ordre de tri vous permet de configurer l'ordre dans lequel les champs sont triés, si plusieurs champs sont triés.

Pour désactiver le tri par un champ, effacez la colonne Type de tri du champ correspondant.

### Définition de critères

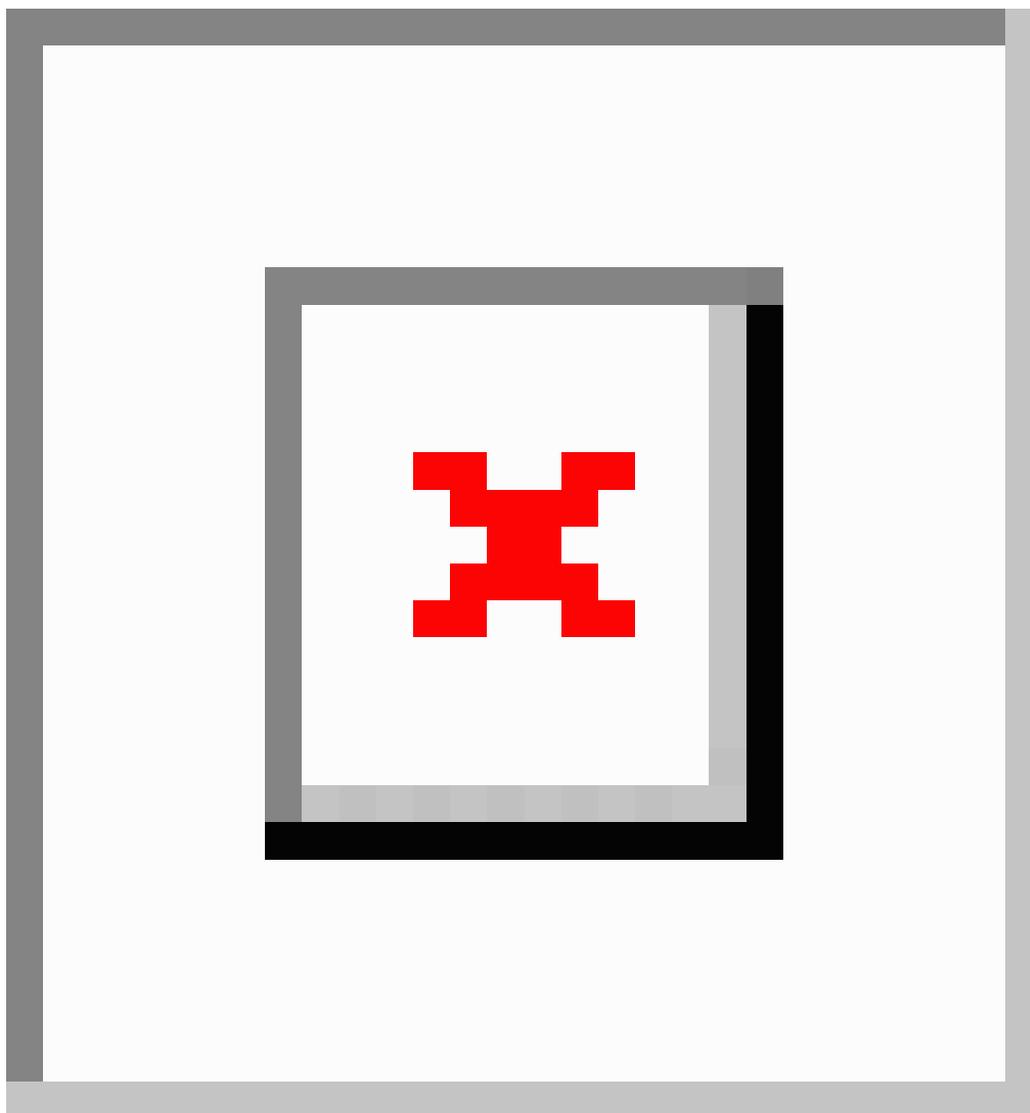
Pour définir des critères, utilisez la colonne **Critères** et les colonnes **Ou** du volet **Colonnes**. Lorsque vous écrivez des conditions dans ces colonnes, omettez l'expression elle-même. Par exemple, pour obtenir les critères suivants dans votre requête :

```
WHERE (Field1 >= 10) AND (Field1 <= 20)
```

Tapez ce qui suit dans la cellule Critères de l'expression Field1 :

```
>= 10 AND <= 20
```

Les critères placés dans les colonnes **Or** sont regroupés par colonnes à l'aide de l'opérateur AND, puis concaténés dans la clause WHERE (ou HAVING) à l'aide l'opérateur OR. Par exemple, les critères indiquées ci-dessous produiront l'instruction SQL suivante. Les critères de Field1 sont placés à la fois dans les colonnes Critères et Ou.



```
WHERE (Field1= 10) AND ((Field2 < 0) OR (Field2 > 10))
```

Certaines expressions peuvent être de type booléen, par exemple, la clause EXISTS. Dans ce cas, vous devez taper « = True » dans la colonne Critères de ces expressions ou « = False » si vous souhaitez placer un opérateur NOT avant l'expression.

### Regroupement de champs de sortie

Pour créer une requête avec le regroupement, les expressions de marque pour le regroupement avec la case **Regroupement**.

Une requête avec regroupement peut ne disposer que d'expressions de regroupement ou d'agrégat dans la liste SELECT. Par conséquent, le générateur de requêtes vous permet de définir la case Sortie pour les expressions de regroupement et d'agrégat. Si vous tentez de définir cette case pour

une colonne sans le jeu de fonctions de regroupements ou d'agrégat, une case à cocher **Regroupement** est définie automatiquement pour maintenir la validité de la requête SQL obtenue.

Lorsque le volet **Colonnes** contient des colonnes marquées par la case à cocher **Regroupement**, une nouvelle colonne appelée **Critères de** apparaît dans la grille. Cette colonne applique des critères aux groupes d'expression ou à leurs valeurs.

Par exemple, vous avez une colonne « Quantité » avec une fonction d'agrégat « Moy » dans votre requête et vous saisissez  $> 10$  dans la colonne Critères. Lorsque la valeur « pour les groupes » est définie dans la colonne Critères de, la requête obtenue contient uniquement des groupes avec une quantité moyenne supérieure à 10 et votre requête présente la condition « Avg(Quantity)  $> 10$  » dans une clause HAVING. Lorsque la valeur « pour les valeurs » est définie dans la colonne Critères de, la requête de résultats calculera la fonction d'agrégat Moyenne uniquement pour les enregistrements doté d'une valeur Quantité supérieure à 10, et votre requête disposera de la condition « Quantity  $> 10$  » dans la clause WHERE.

### Définition de propriétés de requête SQL

Vous pouvez définir des options propres à votre serveur de base de données à l'aide du menu contextuel de la zone **Création de requête**.

## Variables de requête

Dans le stage **Read From DB**, lors de la définition de la requête à exécuter, vous pouvez inclure des variables au lieu de noms de colonne réels. L'utilisation de variables dans la requête vous permet de personnaliser les conditions de requête lors de l'exécution (à l'aide de **Options de flux de données** ou via **Job Executor**).

Cependant, vous pouvez également fournir la valeur Options de Stage dans l'onglet **Exécution** et afficher le schéma et des exemples d'enregistrements à extraire via la requête à l'aide des boutons **Régénérer les champs** et **Aperçu**, respectivement.

Une variable est définie à l'aide du format `#{variable}`, et insérée dans la clause `select` ou `where` d'une requête SQL.

**Remarque :** Vous pouvez modifier une requête générée via **Visual Query Builder** pour y inclure des variables. En revanche, la requête modifiée ne pourra plus être lue par Visual Query Builder. Le bouton **Créer SQL** est désactivé lorsque vous incluez une variable dans une requête SQL manuellement écrite ou générée.

### Insertion d'une variable de requête

1. Ouvrez le job requis, qui inclut un stage **Read From DB**. Sinon, ajoutez un stage **Read from DB** au job.
2. Ouvrez la boîte de dialogue **Options de Read from DB** du stage **Read from DB**.

3. Créez la requête SQL dans le champ **SQL**, soit manuellement, soit à l'aide de Visual Query Builder. Pour plus d'informations, reportez-vous à la section [Visual Query Builder](#) à la page 141.
4. Ajouter les conditions de votre choix à la clause `where` de la requête à l'aide de variables sous la syntaxe `#{variable}`.  
Par exemple, dans une table `CUSTOMERS`, qui contient la colonne `AGE` avec des valeurs telles que 28, 32, 30, etc., et une colonne `SALARY` avec des valeurs telles que 1 000, 1 500, 2 200, etc., créez une requête SQL comme suit :

```
select * from CUSTOMERS where #{condition1} > 28 and #{condition2} > 1200
```

**Remarque :** Pour l'insertion d'une variable dans la clause `where` de la requête SQL, le bouton **Créer SQL...** est désactivé.

5. Pour afficher le schéma et les exemples d'enregistrements à extraire via la requête, saisissez la valeur Options de stage de l'onglet **Exécution**, puis cliquez sur les boutons **Régénérer les champs** et **Aperçu**, respectivement.
6. Cliquez sur **OK**.

La clause `where` de la requête SQL peut désormais être personnalisée lors de l'exécution à l'aide des **Options de flux de données** ou lors de l'exécution du job via JobExecutor.

**Remarque :** Une variable peut également être placée dans la clause `select` d'une requête SQL. Cependant, le nom de la variable doit correspondre au nom de l'une des colonnes de la table faisant l'objet de la requête.

### Configuration d'une variable de requête comme option de flux de données

1. Ouvrez le job requis pour lequel la requête contenant la ou les variables a été définie dans un stage **Read From DB**.
2. Ouvrez **Édition > Options de flux de données....**
3. Cliquez sur **Ajouter**.
4. Dans la section **Map dataflow options to stages**, développez l'entrée **Read From DB**. Les variables définies dans la requête SQL du stage **Read From DB** sont répertoriées avec les autres attributs du stage.
5. Sélectionnez la variable que vous souhaitez personnaliser à l'aide de la case à cocher correspondante.
6. Saisissez un nom pertinent pour la variable dans le champ **Libellé d'option**.
7. Dans le champ **Valeur par défaut**, saisissez le nom de colonne qui doit être utilisé au lieu de la variable sélectionnée dans la clause `where` de la requête SQL. Vous pouvez également saisir une valeur de constante à utiliser au lieu de la variable de la clause `where`.

Par exemple, pour la requête SQL ci-dessous définie dans le stage **Read From DB** :

```
select * from CUSTOMERS where #{condition1} > 28 and #{condition2} > 1200
```

Vous pouvez sélectionner la colonne `AGE` de la table `CUSTOMERS` comme **Valeur par défaut** pour la variable `condition1`, et la colonne `SALARY` comme **Valeur par défaut** pour la variable `condition2`.

Lors de l'exécution, la requête est interprétée comme suit :

```
select * from CUSTOMERS where AGE > 28 and SALARY > 1200
```

8. Répétez les étapes 5 à 7 pour toutes les variables placées dans la requête SQL du stage **Read From DB**.
9. Cliquez sur **OK**.

Lors de l'exécution du flux de données, la requête personnalisée permet de récupérer les données requises.

### Configuration d'une variable de requête pour Job Executor

**Remarque :** Vérifiez que vous avez téléchargé Spectrum™ Job Executor sur votre serveur.

1. Créez un fichier texte définissant les valeurs par défaut des variables incluses dans la requête SQL du stage **Read From DB** du job.

Pour attribuer un nom de colonne `AGE` comme valeur par défaut à une variable `condition1`, créez un fichier texte, par exemple, `variables.txt`, et incluez la ligne ci-dessous dans le fichier :

```
condition1=AGE
```

Pour attribuer une valeur de constante, par exemple, 20, comme valeur par défaut à une variable `condition1`, incluez la ligne ci-dessous dans le fichier :

```
condition1=20
```

2. Lors de l'exécution du job à l'aide de Job Executor à l'invite de commande, utilisez l'argument `-o` suivi du chemin d'accès au fichier texte créé.  
Par exemple, pour exécuter le job `ReadCustomerDataJob`, dont les valeurs par défaut de ses variables ont été définies dans le fichier texte `variables.txt`, exécutez la commande ci-dessous à l'invite de commande :

```
java -jar jobexecutor.jar -h "localhost" -u "admin" -p "admin" -s "8080" -j "ReadCustomerDataJob" -o "variables.txt"
```

Lors de l'exécution du job à l'aide de Job Executor, la requête personnalisée permet de récupérer les données requises.

**Remarque :** Pour obtenir des instructions et la syntaxe de ligne de commande, reportez-vous à la section *Exécution d'un job* du Guide Dataflow Designer.

## Read from file

Le stage Read from File spécifie un fichier d'entrée pour un job ou un sous-flux. Il n'est pas disponible pour les services.

**Remarque :** Si vous souhaitez utiliser un fichier XML en entrée pour votre dataflow, utilisez le stage Read from XML au lieu de Read from File. Si vous souhaitez utiliser un fichier au format variable en entrée, utilisez Read from Variable Format File.

### Onglet de propriétés du fichier

Nom du champ	Description
Server name	Indique si le fichier que vous sélectionnez comme entrée figure sur l'ordinateur exécutant Enterprise Designer ou sur le serveur Spectrum™ Technology Platform. Si vous sélectionnez un fichier sur l'ordinateur local, le nom du serveur sera Mon ordinateur. Si vous sélectionnez un fichier sur le serveur, le nom du serveur sera Spectrum™ Technology Platform.

Nom du champ	Description
Nom de fichier	<p>Indique le chemin du fichier. Cliquez sur le bouton d'ellipse (...) pour accéder au fichier souhaité.</p> <p>Vous pouvez lire plusieurs fichiers en utilisant un caractère générique pour lire les données de plusieurs fichiers du répertoire. Les caractères génériques * et ? sont pris en charge. Par exemple, vous pouvez utiliser *.CSV pour lire tous les fichiers portant une extension .CSV placés dans le répertoire. Pour que plusieurs fichiers puissent être lus, tous doivent présenter la même mise en page (les mêmes champs aux mêmes emplacements). Tout enregistrement ne correspondant pas à la mise en page indiquée dans l'onglet <b>Champs</b> est traité comme un enregistrement non conforme.</p> <p>Lors de la lecture d'un fichier à partir d'un serveur de fichiers HDFS, les formats de compression pris en charge sont les suivants :</p> <ol style="list-style-type: none"> <li>1. GZIP (.gz)</li> <li>2. BZIP2 (.bz2)</li> </ol> <p><b>Remarque :</b> L'extension du fichier indique le format de compression à utiliser pour décompresser le fichier.</p> <p><b>Avertissement :</b> Si le Spectrum™ Technology Platform est en cours d'exécution sur les systèmes Unix ou Linux, rappelez-vous que les noms de fichiers et les chemins sur ces plates-formes sont sensibles à la casse.</p>
Type d'enregistrement	<p>Le format des enregistrements dans le fichier. Sélectionnez l'un des éléments suivants :</p> <p><b>Séquence de lignes</b> Un fichier texte dans lequel les enregistrements sont séparés par un caractère de fin de ligne (EOL) comme un retour chariot ou un retour à la ligne (CR ou LF) et chaque champ dispose d'une position de caractère fixe de départ et de fin.</p> <p><b>Largeur fixe</b> Un fichier texte dans lequel chaque enregistrement est un nombre déterminé de caractères de longueur, et chaque champ dispose d'une position de caractère de départ et de fin fixe .</p> <p><b>Délimité</b> Un fichier texte dans lequel les enregistrements sont séparés par un caractère de fin de ligne (EOL) comme un retour chariot ou un retour à la ligne (CR ou LF) et chaque champ est séparé par un caractère désigné comme une virgule.</p>

Nom du champ	Description
Encodage des caractères	L'encodage du fichier texte. Sélectionnez l'un des éléments suivants :
<b>UTF-8</b>	Prend en charge tous les caractères Unicode et est rétrocompatible avec le format ASCII. Pour plus d'informations sur UTF, reportez-vous à <a href="https://unicode.org/faq/utf_bom.html">unicode.org/faq/utf_bom.html</a> .
<b>UTF-16</b>	Prend en charge tous les caractères Unicode, mais n'est pas rétrocompatible avec le format ASCII. Pour plus d'informations sur UTF, reportez-vous à <a href="https://unicode.org/faq/utf_bom.html">unicode.org/faq/utf_bom.html</a> .
<b>US-ASCII</b>	L'encodage des caractères en fonction de l'ordre alphabétique anglais.
<b>UTF-16BE</b>	Encodage UTF-16 avec sérialisation d'octets gros-boutiste (octet le plus significatif en premier).
<b>UTF-16LE</b>	Encodage UTF-16 avec sérialisation d'octets petit-boutiste (octet le moins significatif en premier).
<b>ISO-8859-1</b>	Chiffrement de caractères ASCII utilisé habituellement pour les langues européennes occidentales. Également désigné par l'acronyme Latin-1.
<b>ISO-8859-3</b>	Chiffrement de caractères ASCII utilisé habituellement pour les langues européennes méridionales. Également connu sous le nom Latin-3.
<b>ISO-8859-9</b>	Chiffrement de caractères ASCII utilisé habituellement pour la langue turque. Également connu sous le nom Latin-5.
<b>CP850</b>	Une page de code ASCII utilisée pour écrire les langues européennes occidentales.
<b>CP500</b>	Une page de code EBCDIC utilisée pour écrire les langues européennes occidentales.
<b>Shift_JIS</b>	Un encodage de caractères pour la langue japonaise.
<b>MS932</b>	Extension Microsoft de Shift_JIS permettant d'inclure les caractères spéciaux NEC, la sélection NEC des extensions IBM et les extensions IBM.
<b>CP1047</b>	Une page de code EBCDIC avec le jeu de caractères Latin-1 complet.

Nom du champ	Description
Séparateur de champs	<p>Indique le caractère utilisé pour séparer des champs dans un fichier délimité.</p> <p>Par exemple, cet enregistrement utilise une barre verticale ( ) comme séparateur de champ :</p> <pre>7200 13TH ST   MIAMI   FL   33144</pre> <p>Ces caractères disponibles à définir en tant que séparateurs de champs sont :</p> <ul style="list-style-type: none"><li>• Espace</li><li>• Tabulation</li><li>• Virgule</li><li>• Point (.)</li><li>• Point-virgule</li><li>• Conduite</li></ul> <p>Si le fichier utilise un caractère différent comme séparateur de champ, cliquez sur le bouton de sélection pour sélectionner un autre caractère délimiteur.</p>
Qualificateur de texte	<p>Le caractère utilisé pour entourer les valeurs de texte dans un fichier délimité.</p> <p>Par exemple, cet enregistrement utilise des guillemets doubles (") en tant que qualificateur de texte.</p> <pre>"7200 13TH ST"   "MIAMI"   "FL"   "33144"</pre> <p>Ces caractères disponibles à définir en tant que qualificateurs de texte sont :</p> <ul style="list-style-type: none"><li>• Guillemets simples (')</li><li>• Guillemets doubles (")</li></ul> <p>Si le fichier utilise un délimiteur de texte, cliquez sur le bouton de sélection pour sélectionner un autre caractère délimiteur.</p>

Nom du champ	Description
Séparateur d'enregistrements	<p>Indique le caractère utilisé pour séparer des enregistrements dans un fichier séquentiel ligne par ligne ou délimité. Ce champ n'est pas disponible si vous activez la case <b>Utiliser EOL par défaut</b>.</p> <p>Les paramètres de séparateur d'enregistrements disponibles sont :</p> <p><b>Unix (U+000A)</b> Un caractère de saut de ligne sépare les enregistrements. Il s'agit du séparateur d'enregistrement standard pour les systèmes Unix.</p> <p><b>Macintosh (U+000D)</b> Un caractère de retour chariot sépare les enregistrements. Il s'agit du séparateur d'enregistrement standard pour les systèmes Macintosh.</p> <p><b>Windows (U+000D U+000A)</b> Un retour chariot suivi d'un saut de ligne sépare les enregistrements. Il s'agit du séparateur d'enregistrement standard pour les systèmes Windows.</p> <p>Si votre fichier utilise un séparateur d'enregistrement, cliquez sur le bouton de sélection pour sélectionner un autre caractère séparateur.</p>
Utiliser EOL par défaut	<p>Indique que le séparateur d'enregistrements du fichier est le caractère de fin de ligne (EOL) par défaut utilisé sur le système d'exploitation sur lequel le serveur Spectrum™ Technology Platform est exécuté.</p> <p>Ne sélectionnez pas cette option si le fichier utilise un caractère EOL qui est différent du caractère EOL par défaut utilisé sur le système d'exploitation du serveur. Par exemple, si le fichier utilise un EOL Windows, mais que le serveur fonctionne sous Linux, ne cochez pas cette option. Au lieu de cela, sélectionnez l'option Windows dans le champ <b>Record separator</b>.</p>
Longueur d'enregistrement	<p>Pour les fichiers à largeur fixe, précisez le nombre exact de caractères dans chaque enregistrement.</p> <p>Pour les fichiers de séquence de lignes, indiquez la longueur, en caractères, de l'enregistrement le plus long dans le fichier.</p>
La première ligne est l'enregistrement d'en-tête.	<p>Indiquez si le premier enregistrement dans un fichier délimité contient des informations d'en-tête, mais aucune donnée.</p> <p>Par exemple, cet extrait de fichier illustre une ligne d'en-tête dans le premier enregistrement.</p> <pre>"AddressLine1" "City" "StateProvince" "PostalCode" "7200 13TH ST" "MIAMI" "FL" "33144" "One Global View" "Troy" "NY" "12180"</pre>

Nom du champ	Description
Traiter les enregistrements avec moins de champs que ce qui est défini comme incorrect	Les enregistrements de fichiers délimités contenant moins de champs que le nombre défini dans l'onglet <b>Champs</b> seront traités comme des enregistrements incorrects.
Importer	Importe la définition du dessin d'enregistrement du fichier, le paramètre d'encodage et les options de tri à partir d'un fichier de paramètres. Le fichier de paramètres est créé en exportant les paramètres à partir d'un autre stage Read from File ou Write to File qui utilisait le même fichier d'entrée ou un fichier qui dispose du même dessin d'enregistrement que le fichier avec lequel vous joblez.
Exporter	<p>Enregistre la définition du dessin d'enregistrement du fichier, le paramètre d'encodage et les options de tri dans un fichier de paramètres. Vous pouvez ensuite importer ces paramètres dans d'autres stages Read from File ou Write to File qui utilisent le même fichier d'entrée ou un fichier semble à celui que vous utilisez actuellement. Vous pouvez utiliser le fichier de paramètres avec Job Executor pour indiquer les paramètres de fichier lors de l'exécution.</p> <p>Pour plus d'informations sur le fichier de paramètres, voir <a href="#">Fichier de paramètres de définition du fichier</a> à la page 161.</p>

### Onglet Champs

L'onglet Champs définit les noms, les positions, et, pour les fichiers à longueur fixe et à séquence de lignes, les longueurs de champs dans le fichier. Pour plus d'informations, reportez-vous à ces rubriques :

[Définition de champs dans un fichier d'entrée délimité](#) à la page 154

[Définition de champs dans un fichier à séquence de ligne ou à longueur fixe](#) à la page 157

### Onglet Trier les champs

L'onglet Trier les champs définit les champs permettant de trier les enregistrements d'entrée avant qu'ils ne soient envoyés dans le dataflow. Le tri est facultatif. Pour plus d'informations, reportez-vous à la section [Tri des enregistrements d'entrée](#) à la page 159.

### L'onglet Exécution

Nom du champ	Description
Nom de fichier	Affiche le nom de fichier sélectionné dans le premier onglet.

Nom du champ	Description
Enregistrement de départ	Si vous voulez ignorer les enregistrements au début du fichier lors de la lecture des enregistrements dans le dataflow, spécifiez le premier enregistrement que vous souhaitez lire. Par exemple, si vous voulez sauter les 50 premiers enregistrements dans un fichier, spécifiez 51. Le 51ème enregistrement sera le premier enregistrement lu dans le dataflow.
Tous les enregistrements	Sélectionnez cette option si vous souhaitez lire tous les enregistrements à partir de l'enregistrement spécifié dans le champ <b>Starting record</b> jusqu'à la fin du fichier.
Nombre max d'enregistrements	Sélectionnez cette option si vous souhaitez uniquement lire un certain nombre d'enregistrements à partir de l'enregistrement spécifié dans le champ <b>Starting record</b> . Par exemple, si vous voulez lire les 100 premiers enregistrements, sélectionnez cette option et entrez 100.

## Définition de champs dans un fichier d'entrée délimité

L'onglet **Champs** définit les noms, la position et, pour certains types de fichier, les longueurs des champs du fichier. Après avoir défini un fichier d'entrée sur l'onglet **Propriétés de fichier**, vous pouvez définir les champs.

Si le fichier d'entrée ne contient pas d'enregistrement d'en-tête, ou si vous souhaitez définir manuellement les champs, suivez ces étapes dans l'onglet **Champs** :

1. Pour définir les champs déjà présents dans le fichier d'entrée, cliquez sur **Régénérer**. Ensuite, cliquez sur **Détecter le type**. Cela définira automatiquement le type de données pour chaque champ en fonction des 50 premiers enregistrements dans le fichier.
2. Pour ajouter des champs supplémentaires dans la sortie, cliquez sur **Ajouter**.
3. Dans le champ **Name**, choisissez le champ à ajouter ou saisissez le nom du champ.
4. Dans le champ **Type**, vous pouvez laisser le type de données défini sur `string` si vous ne prévoyez pas d'effectuer des opérations mathématiques ou d'horodatage sur les données. Cependant, si vous avez l'intention d'effectuer ce genre d'opérations, sélectionnez un type de données approprié. Cela convertira les données de chaîne à partir du fichier en un type de données qui activera la manipulation appropriée des données dans le flux de données.

Spectrum™ Technology Platform prend en charge ces types de données :

- bigdecimal** Type de données numérique prenant en charge 38 points de précision décimaux. Utilisez ce type de données pour les données utilisées dans les calculs mathématiques requérant un haut niveau de précision, en particulier ceux impliquant des données financières. Le type de données `bigdecimal` prend en charge des calculs plus précis que le type de données `double`.

<b>booléen</b>	Type logique doté de deux valeurs : True et False.
<b>bytearray</b>	Tableau (liste) d'octets.  <b>Remarque :</b> ByteArray n'est pas pris en charge en entrée pour un service REST.
<b>date</b>	Type de données contenant un mois, un jour et une année. Par exemple, 2012-01-30 ou January 30, 2012. Vous pouvez indiquer un format de date par défaut dans Management Console.
<b>datetime</b>	Type de données contenant un mois, un jour, une année et des heures, des minutes et des secondes. Par exemple, 2012/01/30 6:15 PM.
<b>double</b>	Un type de données numérique qui contient à la fois des nombres en double précision négatifs et positifs entre $2^{-1074}$ et $(2-2^{-52}) \times 2^{1023}$ . En notation E, la plage de valeurs est comprise entre -1.79769313486232E+308 et 1.79769313486232E+308.
<b>flottant</b>	Un type de données numérique contenant des nombres de précision seuls négatifs et positifs entre $2^{-149}$ et $(2-2^{-23}) \times 2^{127}$ . En notation E, la plage de valeurs est comprise entre -3.402823E+38 et 3.402823E+38.
<b>integer</b>	Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre $-2^{31}$ (-2,147,483,648) et $2^{31}-1$ (2,147,483,647).
<b>liste</b>	Au sens strict du terme, une liste n'est pas un type de données. Cependant, lorsqu'un champ contient des données hiérarchiques, il est traité comme un champ de « liste ». Dans Spectrum™ Technology Platform, une liste représente un ensemble de données composées de plusieurs valeurs. Par exemple, un champ Noms peut contenir une liste de valeurs de noms. Cela peut être représenté sous forme d'une structure XML, comme suit :  <pre>&lt;Names&gt;   &lt;Name&gt;John Smith&lt;/Name&gt;   &lt;Name&gt;Ann Fowler&lt;/Name&gt; &lt;/Names&gt;</pre> <p>Il est important de souligner que le type de données de liste Spectrum™ Technology Platform est différent du type de données de liste de schéma XML dans le sens où ce dernier est un type de données simple composé de plusieurs valeurs, tandis que le type de données de liste Spectrum™ Technology Platform est semblable à un type de données complexe XML.</p>
<b>long</b>	Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre $-2^{63}$ (-9,223,372,036,854,775,808) et $2^{63}-1$ (9,223,372,036,854,775,807).
<b>chaîne</b>	Séquence de caractères.

**time** Type de données contenant l'heure du jour. Par exemple, 21:15:59 ou 9:15:59 PM.

5. Si vous avez sélectionné une date, une heure ou un type de données numérique, vous pouvez utiliser le format date/heure ou numérique par défaut, ou vous pouvez spécifier un autre format pour ce champ spécifique. Le format par défaut est le format par défaut du système qui a été défini dans les options de conversion du type dans Management Console, ou le format par défaut du flux de données spécifié dans les options de conversion du type dans Enterprise Designer. Le format qui s'applique apparaît. Pour utiliser le format par défaut, laissez l'option **Valeur par défaut** sélectionnée. Pour indiquer un autre format, choisissez **Personnalisé** et suivez les étapes suivantes :

**Remarque** : Il est important que vous choisissiez un format de date et heure qui reflète avec précision les données que vous lisez à partir du fichier. Par exemple, si le fichier contient des données de date au format Mois/Jour/Année, mais que vous choisissiez Jour/Mois/Année, tout calcul de date que vous effectuez dans le flux de données, tel que le tri par date, ne reflètera pas la date correcte. En outre, la conversion de type peut échouer pour les enregistrements, auquel cas le comportement d'échec spécifié dans les options de conversion de type dans Management Console ou Enterprise Designer prendra effet.

- a) Dans le champ **Locale**, sélectionnez le pays dont vous souhaitez utiliser la convention de formatage. Votre sélection déterminera les valeurs par défaut dans le champ **Format**. Pour les données de date, votre sélection déterminera également la langue utilisée lorsqu'un mois est épelé. Par exemple, si vous indiquez Anglais, le premier mois de l'année sera « January », alors que si vous indiquez Français, ce sera « Janvier ».
- b) Dans le champ **Format**, sélectionnez le format des données. Le format dépend du type de données du champ. Une liste des formats les plus couramment utilisés pour les paramètres régionaux sélectionnés est fournie.

Un exemple du format sélectionné s'affiche à droite du champ **Format**.

Vous pouvez également spécifier vos propres formats de date, heure et nombre si ceux disponibles ne répondent pas à vos besoins. Pour indiquer votre propre format de date ou heure, tapez le format dans le champ à l'aide de la notation décrite dans [Modèles de date et d'heure](#) à la page 306. Pour indiquer votre propre format numérique, tapez le format dans le fichier à l'aide de la notation décrite dans [Modèles numériques](#) à la page 309.

6. Dans le champ **Position**, entrez la position de ce champ dans l'enregistrement.

Par exemple, dans ce fichier d'entrée, AddressLine1 est en position 1, City est en position 2, StateProvince est en position 3, et PostalCode est en position 4.

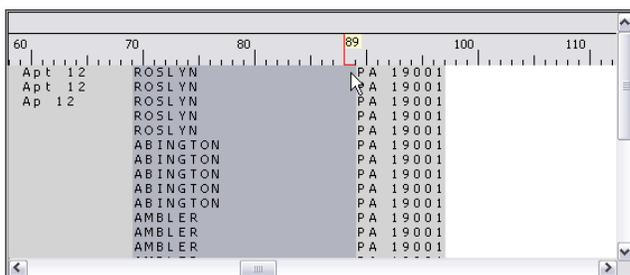
```
"AddressLine1"|"City"|"StateProvince"|"PostalCode"
"7200 13TH ST"|"MIAMI"|"FL"|"33144"
"One Global View"|"Troy"|"NY"|"12180"
```

- Si vous souhaitez que tout caractère d'espace superflu soit supprimé au début et à la fin d'une chaîne de valeur dans un champ, cochez la case **Trim**.

## Définition de champs dans un fichier à séquence de ligne ou à longueur fixe

Dans le stage Read from File, l'onglet **Champs** définit les noms, la position et pour certains types de fichiers, les longueurs des champs dans le fichier. Après avoir défini un fichier d'entrée sur l'onglet **Propriétés de fichier**, vous pouvez définir les champs.

- Dans l'onglet **Champs**, sous **Aperçu**, cliquez au début d'un champ et faites glisser vers la gauche pour sélectionner le champ, comme indiqué ici :



- Dans le champ **Name**, entrez le champ que vous souhaitez ajouter.
- Dans le champ **Type**, vous pouvez laisser le type de données défini sur chaîne si vous ne prévoyez pas d'effectuer d'opérations mathématiques ou d'horodatage sur les données. Cependant, si vous avez l'intention d'effectuer ce genre d'opérations, sélectionnez un type de données approprié. Cela convertira les données de chaîne à partir du fichier en un type de données qui activera la manipulation appropriée des données dans le flux de données.

Spectrum™ Technology Platform prend en charge les types de données suivants :

**bigdecimal** Type de données numérique prenant en charge 38 points de précision décimale. Utilisez ce type de données pour les données utilisées dans les calculs mathématiques requérant un haut niveau de précision, en particulier ceux impliquant des données financières. Le type de données bigdecimal prend en charge des calculs plus précis que le type de données double.

**booléen** Type logique doté de deux valeurs : True et False.

**bytearray** Tableau (liste) d'octets.

**Remarque :** ByteArray n'est pas pris en charge en entrée pour un service REST.

**date** Type de données contenant un mois, un jour et une année. Par exemple, 2012-01-30 ou January 30, 2012. Vous pouvez indiquer un format de date par défaut dans Management Console.

**datetime** Type de données contenant un mois, un jour, une année et des heures, des minutes et des secondes. Par exemple, 2012/01/30 6:15 PM.

<b>double</b>	Un type de données numérique qui contient à la fois des nombres en double précision négatifs et positifs entre $2^{-1074}$ et $(2-2^{-52}) \times 2^{1023}$ . En notation E, la plage de valeurs est comprise entre -1.79769313486232E+308 et 1.79769313486232E+308.
<b>flottant</b>	Un type de données numérique contenant des nombres de précision seuls négatifs et positifs entre $2^{-149}$ et $(2-2^{-23}) \times 2^{127}$ . En notation E, la plage de valeurs est comprise entre -3.402823E+38 et 3.402823E+38.
<b>integer</b>	Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre $-2^{31}$ (-2,147,483,648) et $2^{31}-1$ (2,147,483,647).
<b>liste</b>	Au sens strict du terme, une liste n'est pas un type de données. Cependant, lorsqu'un champ contient des données hiérarchiques, il est traité comme un champ de « liste ». Dans Spectrum™ Technology Platform, une liste représente un ensemble de données composées de plusieurs valeurs. Par exemple, un champ Noms peut contenir une liste de valeurs de noms. Cela peut être représenté sous forme d'une structure XML, comme suit :
	<pre>&lt;Names&gt;   &lt;Name&gt;John Smith&lt;/Name&gt;   &lt;Name&gt;Ann Fowler&lt;/Name&gt; &lt;/Names&gt;</pre>
	Il est important de souligner que le type de données de liste Spectrum™ Technology Platform est différent du type de données de liste de schéma XML dans le sens où ce dernier est un type de données simple composé de plusieurs valeurs, tandis que le type de données de liste Spectrum™ Technology Platform est semblable à un type de données complexe XML.
<b>long</b>	Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre $-2^{63}$ (-9,223,372,036,854,775,808) et $2^{63}-1$ (9,223,372,036,854,775,807).
<b>chaîne</b>	Séquence de caractères.
<b>time</b>	Type de données contenant l'heure du jour. Par exemple, 21:15:59 ou 9:15:59 PM.

- Pour appliquer le format de stockage `packed decimal` au champ, cochez la case **Décimal condensé**. Le type `packed decimal` utilise 4 bits, contrairement au type `integer`, qui utilise 8 bits.

**Remarque :** Ce format de stockage est disponible uniquement lors de la sélection des types de données `double`, `integer` et `long` lors de la lecture de fichiers à largeur fixe et à séquence de lignes.

- Si vous avez sélectionné une date, une heure ou un type de données numérique, vous pouvez utiliser le format date/heure ou numérique par défaut, ou vous pouvez spécifier un autre format pour ce champ spécifique. Le format par défaut est le format par défaut du système qui a été

défini dans les options de conversion du type dans Management Console, ou le format par défaut du flux de données spécifié dans les options de conversion du type dans Enterprise Designer. Le format qui s'applique apparaît. Pour utiliser le format par défaut, laissez l'option **Valeur par défaut** sélectionnée. Pour indiquer un autre format, choisissez **Personnalisé** et suivez les étapes suivantes :

**Remarque** : Il est important que vous choisissiez un format de date et heure qui reflète avec précision les données que vous lisez à partir du fichier. Par exemple, si le fichier contient des données de date au format Mois/Jour/Année, mais que vous choisissiez Jour/Mois/Année, tout calcul de date que vous effectuez dans le flux de données, tel que le tri par date, ne reflètera pas la date correcte. En outre, la conversion de type peut échouer pour les enregistrements, auquel cas le comportement d'échec spécifié dans les options de conversion de type dans Management Console ou Enterprise Designer prendra effet.

- a) Dans le champ **Locale**, sélectionnez le pays dont vous souhaitez utiliser la convention de formatage. Votre sélection déterminera les valeurs par défaut dans le champ **Format**. Pour les données de date, votre sélection déterminera également la langue utilisée lorsqu'un mois est épilé. Par exemple, si vous indiquez Anglais, le premier mois de l'année sera « January », alors que si vous indiquez Français, ce sera « Janvier ».
- b) Dans le champ **Format**, sélectionnez le format des données. Le format dépend du type de données du champ. Une liste des formats les plus couramment utilisés pour les paramètres régionaux sélectionnés est fournie.

Un exemple du format sélectionné s'affiche à droite du champ **Format**.

Vous pouvez également spécifier vos propres formats de date, heure et nombre si ceux disponibles ne répondent pas à vos besoins. Pour indiquer votre propre format de date ou heure, tapez le format dans le champ à l'aide de la notation décrite dans [Modèles de date et d'heure](#) à la page 306. Pour indiquer votre propre format numérique, tapez le format dans le fichier à l'aide de la notation décrite dans [Modèles numériques](#) à la page 309.

6. Les champs **Position de début** et **Longueur** sont remplis automatiquement en fonction du choix effectué dans l'aperçu du fichier.
7. Si vous souhaitez que tout caractère d'espacement superflu soit supprimé au début et à la fin d'une chaîne de caractères dans un champ, cochez la case **Espaces courts**.
8. Cliquez sur **OK**.

## Tri des enregistrements d'entrée

Dans le stage Read from File, l'onglet **Trier les champs** définit les champs permettant de trier les enregistrements d'entrée avant qu'ils ne soient envoyés dans le flux de données. Le tri est facultatif.

1. Dans l'onglet **Trier des champs**, cliquez sur **Ajouter**.

2. Cliquez sur la flèche de déroulement dans la colonne **Nom du champ** et sélectionnez le champ que vous voulez trier. Les champs disponibles à la sélection dépendent des champs définis dans ce fichier d'entrée.
3. Dans la colonne **Ordre**, sélectionnez Ascending ou Descending.
4. Répétez cette opération jusqu'à l'ajout de tous les champs d'entrée que vous souhaitez utiliser pour le tri. Changez l'ordre du tri en surlignant la ligne pour le champ que vous voulez déplacer et cliquez sur **Vers le haut** ou **Vers le bas**.
5. Les options de performance de tri par défaut de votre système sont définies dans Management Console. Si vous voulez remplacer les options de performance de tri par défaut, cliquez sur l'onglet **Avancé**. La boîte de dialogue **Options avancées** contient les options de performance de tri suivantes :

**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 :

$$(InMemoryRecordLimit \times MaxNumberOfTempFiles \div 2) \geq TotalNumberOfRecords$$

## Fichier de paramètres de définition du fichier

Un fichier de paramètres de définition du fichier contient le dessin d'enregistrement du fichier, l'encodage et les options de tri ayant été exportés à partir d'un stage Read from File ou Write to File. Le fichier de paramètres de définition du fichier peut être importé dans Read from File ou Write to File pour définir rapidement les options du stage au lieu d'indiquer manuellement les options.

La manière la plus simple de créer un fichier de paramètres de définition du fichier est d'indiquer les paramètres de fichier à l'aide de Read from File ou de Write to File, puis de cliquer sur le bouton **Exporter** pour générer le fichier de paramètres de définition du fichier.

Cependant, à titre informatif, le schéma du fichier de paramètres de définition du fichier s'affiche ci-dessous. Chaque élément dans le fichier XML possède un type et si ce type est autre qu'une chaîne ou un nombre entier, les valeurs autorisées s'affichent. Ces valeurs correspondent directement aux options contenues dans la boîte de dialogue du stage. Par exemple, l'élément FileTypeEnum correspond au champ Type d'enregistrement sous l'onglet Propriétés de fichier, et les trois valeurs suivantes apparaissent dans le schéma : linesequential, fixedwidth et delimited.

**Remarque :** Si vous saisissez « personnalisé » pour les champs LineSeparator, FieldSeparator ou TextQualifier, un élément personnalisé correspondant doit être inclus (par ex., « CustomLineSeparator », « CustomFieldSeparator » ou « CustomTextQualifier ») avec un nombre hexadécimal représentant le caractère ou la séquence de caractères à utiliser.

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema elementFormDefault="qualified"
  xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="FileSchema" nillable="true" type="FileSchema"/>
  <xs:complexType name="FileSchema">
    <xs:sequence>
      <xs:element
        minOccurs="0"
        maxOccurs="1"
        default="linesequential"
        name="Type"
        type="FileTypeEnum"/>
      <xs:element
        minOccurs="0"
        maxOccurs="1"
        default="UTF-8" name="Encoding" type="xs:string"/>
    </xs:sequence>
  </xs:complexType>
</xs:schema>
```

```
<xs:element
  minOccurs="0"
  maxOccurs="1"
  name="RecordLength"
  type="xs:int"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  default="default"
  name="LineSeparator"
  type="LineSeparatorEnum"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  name="CustomLineSeparator"
  type="xs:string"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  default="comma"
  name="FieldSeparator"
  type="FieldSeparatorEnum"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  name="CustomFieldSeparator"
  type="xs:string"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  default="none"
  name="TextQualifier"
  type="TextQualifierEnum"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  name="CustomTextQualifier"
  type="xs:string"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  default="false"
  name="HasHeader"
  type="xs:boolean"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  default="true"
  name="EnforceColumnCount"
  type="xs:boolean"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
```

```

        name="Fields"
        type="ArrayOfFieldSchema"/>
    </xs:sequence>
</xs:complexType>
<xs:simpleType name="FileTypeEnum">
    <xs:restriction base="xs:string">
        <xs:enumeration value="linesequential"/>
        <xs:enumeration value="fixedwidth"/>
        <xs:enumeration value="delimited"/>
    </xs:restriction>
</xs:simpleType>
<xs:simpleType name="LineSeparatorEnum">
    <xs:restriction base="xs:string">
        <xs:enumeration value="default"/>
        <xs:enumeration value="windows"/>
        <xs:enumeration value="unix"/>
        <xs:enumeration value="mac"/>
        <xs:enumeration value="custom"/>
    </xs:restriction>
</xs:simpleType>
<xs:simpleType name="FieldSeparatorEnum">
    <xs:restriction base="xs:string">
        <xs:enumeration value="comma"/>
        <xs:enumeration value="tab"/>
        <xs:enumeration value="space"/>
        <xs:enumeration value="semicolon"/>
        <xs:enumeration value="period"/>
        <xs:enumeration value="pipe"/>
        <xs:enumeration value="custom"/>
    </xs:restriction>
</xs:simpleType>
<xs:simpleType name="TextQualifierEnum">
    <xs:restriction base="xs:string">
        <xs:enumeration value="none"/>
        <xs:enumeration value="single"/>
        <xs:enumeration value="double"/>
        <xs:enumeration value="custom"/>
    </xs:restriction>
</xs:simpleType>
<xs:complexType name="ArrayOfFieldSchema">
    <xs:sequence>
        <xs:element
            minOccurs="0"
            maxOccurs="unbounded"
            name="Field"
            nillable="true"
            type="FieldSchema"/>
    </xs:sequence>
</xs:complexType>
<xs:complexType name="FieldSchema">
    <xs:sequence>
        <xs:element
            minOccurs="0"

```

```

        maxOccurs="1"
        name="Name"
        type="xs:string"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  default="string"
  name="Type"
  type="xs:string"/>
<xs:element
  minOccurs="1"
  maxOccurs="1"
  name="Position"
  type="xs:int"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  name="Length"
  type="xs:int"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  default="false"
  name="Trim"
  type="xs:boolean"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  name="Locale"
  type="Locale"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  name="Pattern"
  type="xs:string"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  default="none"
  name="Order"
  type="SortOrderEnum"/>
</xs:sequence>
</xs:complexType>
<xs:complexType name="Locale">
  <xs:sequence>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      name="Country"
      type="xs:string"/>
    <xs:element
      minOccurs="0"
      maxOccurs="1"

```

```

        name="Language"
        type="xs:string"/>
    <xs:element
        minOccurs="0"
        maxOccurs="1"
        name="Variant"
        type="xs:string"/>
    </xs:sequence>
</xs:complexType>
<xs:simpleType name="SortOrderEnum">
    <xs:restriction base="xs:string">
        <xs:enumeration value="none"/>
        <xs:enumeration value="ascending"/>
        <xs:enumeration value="descending"/>
    </xs:restriction>
</xs:simpleType>
</xs:schema>

```

## Read from Hadoop Sequence File

Le stage Read from Hadoop Sequence File lit les données d'un fichier de séquence comme l'entrée d'un flux de données. Un fichier de séquence est un fichier plat constitué de paires de valeurs clés binaires. Pour obtenir davantage d'informations, consultez l'adresse suivante :

[wiki.apache.org/hadoop/SequenceFile](http://wiki.apache.org/hadoop/SequenceFile).

**Remarque :** Le stage Read from Hadoop Sequence File prend uniquement en charge les fichiers de séquence délimités non compressés figurant sur Hadoop Distributed File System (HDFS).

### Onglet de propriétés du fichier

Champs	Description
Serveur	Indique que le fichier que vous sélectionnez dans le champ <b>Nom de fichier</b> se trouve sur le système Hadoop. Avant de pouvoir l'utiliser dans le stage, vous devez créer une connexion au serveur de fichiers Hadoop dans Management Console. Si vous sélectionnez un fichier sur le système Hadoop, le nom de serveur est le nom que vous indiquez dans Management Console lors de la création d'un serveur de fichiers.
Nom du fichier	Indique le chemin du fichier. Cliquez sur le bouton de sélection (...) pour trouver le fichier souhaité.

Champs	Description
Séparateur de champs	<p>Indique le caractère utilisé pour séparer des champs dans un fichier délimité.</p> <p>Par exemple, cet enregistrement utilise une barre verticale ( ) comme séparateur de champ :</p> <pre>7200 13TH ST   MIAMI   FL   33144</pre> <p>Ces caractères disponibles à définir en tant que séparateurs de champs sont :</p> <ul style="list-style-type: none"><li>• Espace</li><li>• Tabulation</li><li>• Virgule</li><li>• Point (.)</li><li>• Point-virgule</li><li>• Conduite</li></ul> <p>Si le fichier utilise un caractère différent comme séparateur de champ, cliquez sur le bouton de sélection pour sélectionner un autre caractère délimiteur.</p>
Qualificateur de texte	<p>Le caractère utilisé pour entourer les valeurs de texte dans un fichier délimité.</p> <p>Par exemple, cet enregistrement utilise des guillemets doubles (") en tant que qualificateur de texte.</p> <pre>"7200 13TH ST"   "MIAMI"   "FL"   "33144"</pre> <p>Ces caractères disponibles à définir en tant que qualificateurs de texte sont :</p> <ul style="list-style-type: none"><li>• Guillemets simples (')</li><li>• Guillemets doubles (")</li></ul> <p>Si le fichier utilise un délimiteur de texte, cliquez sur le bouton de sélection pour sélectionner un autre caractère délimiteur.</p>

### Onglet Champs

L'onglet Champs définit les noms, positions et types des champs du fichier. Pour plus d'informations, reportez-vous à la section [Définition de champs dans un fichier de séquence d'entrée](#) à la page 167.

### Onglet Trier les champs

L'onglet Trier les champs définit les champs permettant de trier les enregistrements d'entrée avant qu'ils ne soient envoyés dans le flux de données. Le tri est facultatif. Pour plus d'informations, reportez-vous à la section [Tri des enregistrements d'entrée](#) à la page 168.

### Onglet Filtre

L'onglet Champ définit les champs permettant de filtrer les enregistrements d'entrée avant leur envoi au flux de données. Pour plus d'informations, reportez-vous à la section **Filtrage des enregistrements d'entrée** à la page 168.

## Définition de champs dans un fichier de séquence d'entrée

Dans le stage Read from Hadoop Sequence File, l'onglet **Champs** définit les noms, les positions et les types des champs du fichier. Après avoir défini un fichier d'entrée sur l'onglet **Propriétés de fichier**, vous pouvez définir les champs.

Si le fichier d'entrée ne contient pas d'enregistrement d'en-tête, ou si vous souhaitez définir manuellement les champs, suivez ces étapes dans l'onglet **Champs** :

1. Pour définir les champs déjà présents dans le fichier d'entrée, cliquez sur **Régénérer**. Ensuite, cliquez sur **Détecter le type**. Cela définira automatiquement le type de données pour chaque champ en fonction des 50 premiers enregistrements dans le fichier.
2. Pour ajouter des champs supplémentaires dans la sortie, cliquez sur **Ajouter**.
3. Dans le champ **Name**, choisissez le champ à ajouter ou saisissez le nom du champ.
4. Dans le champ **Type**, vous pouvez laisser le type de données défini sur chaîne si vous ne prévoyez pas d'effectuer des opérations mathématiques sur les données. Cependant, si vous avez l'intention d'effectuer ce genre d'opérations, sélectionnez un type de données approprié. Cela convertira les données de chaîne à partir du fichier en un type de données qui activera la manipulation appropriée des données dans le flux de données.

Le stage prend en charge les types de données suivants :

<b>double</b>	Un type de données numérique qui contient à la fois des nombres en double précision négatifs et positifs entre $2^{-1074}$ et $(2 \cdot 2^{-52}) \times 2^{1023}$ . En notation E, la plage de valeurs est comprise entre -1.79769313486232E+308 et 1.79769313486232E+308.
<b>flottant</b>	Un type de données numérique contenant des nombres de précision seuls négatifs et positifs entre $2^{-149}$ et $(2 \cdot 2^{23}) \times 2^{127}$ . En notation E, la plage de valeurs est comprise entre -3.402823E+38 et 3.402823E+38.
<b>integer</b>	Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre $-2^{31}$ (-2,147,483,648) et $2^{31}-1$ (2,147,483,647).
<b>long</b>	Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre $-2^{63}$ (-9,223,372,036,854,775,808) et $2^{63}-1$ (9,223,372,036,854,775,807).
<b>chaîne</b>	Séquence de caractères.

5. Dans le champ **Position**, entrez la position de ce champ dans l'enregistrement.

Par exemple, dans ce fichier d'entrée, AddressLine1 est en position 1, City est en position 2, StateProvince est en position 3, et PostalCode est en position 4.

```
"AddressLine1"|"City"|"StateProvince"|"PostalCode"
"7200 13TH ST"|"MIAMI"|"FL"|"33144"
"One Global View"|"Troy"|"NY"|"12180"
```

- Si vous souhaitez que tout caractère d'espace superflu soit supprimé au début et à la fin d'une chaîne de valeur dans un champ, cochez la case **Trim**.

## Tri des enregistrements d'entrée

Dans le stage Read from Hadoop Sequence File, l'onglet **Trier les champs** définit les champs en fonction desquels trier les enregistrements d'entrée avant leur envoi au flux de données. Le tri est facultatif.

- Dans Read from Hadoop Sequence File, cliquez sur l'onglet **Trier les champs**.
- Dans l'onglet **Trier des champs**, cliquez sur **Ajouter**.
- Cliquez sur la flèche de déroulement dans la colonne **Nom du champ** et sélectionnez le champ que vous voulez trier. Les champs disponibles à la sélection dépendent des champs définis dans ce fichier d'entrée.
- Dans la colonne **Ordre**, sélectionnez Ascending ou Descending.
- Répétez cette opération jusqu'à l'ajout de tous les champs d'entrée que vous souhaitez utiliser pour le tri. Changez l'ordre du tri en surlignant la ligne pour le champ que vous voulez déplacer et cliquez sur **Vers le haut** ou **Vers le bas**.

## Filtrage des enregistrements d'entrée

Dans le stage Read from Hadoop Sequence File, l'onglet **Filtre** définit les champs en fonction desquels filtrer les enregistrements d'entrée avant leur envoi au dataflow. Le filtrage est facultatif.

- Dans le stage Read from Hadoop Sequence File, cliquez sur l'onglet **Filtre**.
- Dans le champ **Combiner méthode d'expression**, sélectionnez **Tous** si vous souhaitez que l'évaluation de toutes les expressions soit définie sur true pour que l'enregistrement soit dirigé vers ce port; sélectionnez **N'importe lequel** si vous souhaitez que les enregistrements soient dirigés vers ce port si l'une ou plusieurs des expressions sont définies sur true.
- Cliquez sur **Ajouter** et précisez le champ à tester, l'opérateur et une valeur. Les opérateurs figurent dans le tableau suivant.

Opérateur	Description
Egal à	Vérifie si la valeur du champ correspond à la valeur indiquée.
n'est pas égal à	Vérifie si la valeur du champ ne correspond pas à la valeur indiquée.
Est supérieur à	Vérifie si le champ a une valeur numérique supérieure à la valeur indiquée. Cet opérateur fonctionne sur des types de données numériques ainsi que sur des champs de chaînes contenant des nombres.
est supérieur ou égal à	Vérifie si le champ a une valeur numérique supérieure ou égale à la valeur indiquée. Cet opérateur fonctionne sur des types de données numériques ainsi que sur des champs de chaînes contenant des nombres.
Est inférieur à	Vérifie si le champ a une valeur numérique inférieure à la valeur indiquée. Cet opérateur fonctionne sur des types de données numériques ainsi que sur des champs de chaînes contenant des nombres.
est inférieur ou égal à	Vérifie si le champ a une valeur numérique inférieure ou égale à la valeur indiquée. Cet opérateur fonctionne sur des types de données numériques ainsi que sur des champs de chaînes contenant des nombres.
Est nul	Vérifie si le champ est une valeur nulle.
N'est pas nul	Vérifie si le champ n'est pas une valeur nulle.

- Sélectionnez l'option **Raccourcir**, le cas échéant. Cette option commence par couper tous les espaces blancs susceptibles de figurer avant et après la valeur du champ, afin le filtrage des données du champ.
- Répétez cette opération jusqu'à l'ajout de tous les champs d'entrée que vous souhaitez utiliser pour le filtrage.

## Read from Hive File

Le stage **Read from Hive File** lit les données du fichier sélectionné, qui peut se présenter sous l'un des formats suivants :

- ORC
- RC

- PARQUET
- Avro

### Onglet *propriétés du fichier*

Champs	Description
Serveur	Indique que le fichier que vous sélectionnez dans le champ <b>Nom de fichier</b> se trouve sur le système Hadoop. Avant de pouvoir l'utiliser dans le stage, vous devez créer une connexion au serveur de fichiers Hadoop dans Management Console. Si vous sélectionnez un fichier sur le système Hadoop, le nom de serveur est le nom que vous indiquez dans Management Console lors de la création d'un serveur de fichiers.
Nom du fichier	Indique le chemin du fichier. Cliquez sur le bouton de sélection (...) pour trouver le fichier souhaité.  <p><b>Remarque :</b> Le schéma d'un fichier d'entrée est importé dès que vous naviguez vers un emplacement correct et que vous sélectionnez le fichier. Ce schéma importé ne peut pas être modifié.</p> <p>Vous pouvez, par contre, renommer les colonnes du schéma, le cas échéant.</p> <p>Les 50 premiers enregistrements du fichier sont extraits dans la grille <b>Aperçu</b> lors de la sélection du fichier.</p>
File type	Sélectionnez le type de fichier lu : <ul style="list-style-type: none"> <li>• ORC</li> <li>• RC</li> <li>• PARQUET</li> <li>• Avro</li> </ul>

**Remarque :** S'il s'agit de fichiers RC, pour générer l'Aperçu, définissez le schéma dans l'onglet **Champs**, puis cliquez sur **Aperçu** dans l'onglet **Propriétés de fichier**.

### Onglet *Champs*

L'onglet **Champs** définit les noms, les types et les positions des champs tels qu'ils figurent dans le fichier d'entrée, ainsi que les noms donnés par les utilisateurs pour les champs. Pour plus d'informations, reportez-vous à la section [Définition des champs pour Read from Hive File](#) à la page 171.

## Définition des champs pour Read from Hive File

Dans l'onglet **Champs** du stage **Read from Hive File**, les noms de schéma, les types de données, les positions et les noms donnés des champs du fichier sont répertoriés.

1. Cliquez sur **Régénérer**.

S'il s'agit de fichiers ORC, Avro et Parquet, cette opération génère le schéma en fonction des métadonnées du fichier existant. S'il s'agit de fichiers RC, tous les champs ajoutés avant de cliquer sur **Aperçu** sont effacés.

La grille affiche les colonnes **Nom**, **Type**, **Champ de stage** et **Inclure**.

La colonne **Nom** affiche le nom de champ dérivé de l'enregistrement d'en-tête du fichier.

La colonne **Type** répertorie les types de données de chaque champ respectif du fichier.

Le stage prend en charge les types de données suivants :

**booléen** Type logique doté de deux valeurs : True et False.

**date** Type de données contenant un mois, un jour et une année. Par exemple, 2012-01-30 ou January 30, 2012. Vous pouvez indiquer un format de date par défaut dans Management Console.

**datetime** Type de données contenant un mois, un jour, une année et des heures, des minutes et des secondes.

Par exemple, 2012/01/30 6:15 PM.

**Remarque :** Le type de données `datetime` de Spectrum correspond au type de données `timestamp` des fichiers Hive.

**double** Un type de données numérique qui contient à la fois des nombres en double précision négatifs et positifs entre  $2^{-1074}$  et  $(2-2^{-52}) \times 2^{1023}$ . En notation E, la plage de valeurs est comprise entre -1.79769313486232E+308 et 1.79769313486232E+308.

**bigdecimal** Type de données numérique prenant en charge 38 points de précision décimaux. Utilisez ce type de données pour les données utilisées dans les calculs mathématiques requérant un haut niveau de précision, en particulier ceux impliquant des données financières. Le type de données `bigdecimal` prend en charge des calculs plus précis que le type de données `double`.

**Remarque :** Pour les fichiers RC, Avro et Parquet Hive, les champs du type de données `decimal` du fichier d'entrée sont convertis en type de données `bigdecimal`.

**long** Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre  $-2^{63}$  (-9,223,372,036,854,775,808) et  $2^{63}-1$  (9,223,372,036,854,775,807).

**Remarque :** Le type de données `long` de Spectrum correspond au type de données `bigint` des fichiers Hive.

**integer** Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre  $-2^{31}$  (-2,147,483,648) et  $2^{31}-1$  (2,147,483,647).

**flottant** Un type de données numérique contenant des nombres de précision seuls négatifs et positifs entre  $2^{-149}$  et  $(2-2^{23})\times 2^{127}$ . En notation E, la plage de valeurs est comprise entre -3.402823E+38 et 3.402823E+38.

**chaîne** Séquence de caractères.

**Remarque :** S'il s'agit de fichiers RC, les types de données `smallint` et complexes ne sont pas pris en charge.

La colonne **Position** affiche la position de départ du champ respectif au sein d'un enregistrement.

2. Dans la colonne **Champ de stage**, modifiez le nom de champ existant de votre choix pour chaque champ.

Par défaut, cette colonne affiche les noms de champ lus à partir du fichier.

3. Dans la colonne **Inclure**, cochez les cases en regard des champs que vous souhaitez inclure dans la sortie du stage.

Par défaut, tous les champs sont sélectionnés dans cette colonne.

4. Pour les fichiers RC, vous pouvez ajouter et supprimer des champs et modifier l'ordre des colonnes sélectionnées dans la sortie en utilisant les boutons ci-dessous :

Nom de l'option	Description
Ajouter	Ajoute un champ dans la sortie.
Modifier	Modifie le nom et le type de données du champ sélectionné.
Supprimer	Enlève le champ sélectionné de la sortie.
Déplacer vers le haut/Déplacer vers le bas	Réorganise la position du champ sélectionné dans la sortie.

**Remarque :** Cette fonctionnalité est disponible uniquement pour les fichiers RC.

5. Cliquez sur **OK**.

## Read from HL7 File

Le stage **Read from HL7 File** lit les données Health Level Seven (HL7) d'un fichier texte sous forme d'entrée d'un dataflow. HL7 est une norme de messagerie utilisée dans l'industrie des soins de santé pour échanger des données entre systèmes. Pour plus d'informations sur HL7, consultez l'adresse [www.hl7.org](http://www.hl7.org).

### *Format de message HL7*

Les données d'un message HL7 sont organisées de manière hiérarchique, comme suit :

- message
  - segment
    - champ
      - composant
        - Sous-composant

Chaque ligne d'un message HL7 est un segment. Un *segment* est un regroupement logique de champs. Les trois premiers caractères d'un segment identifient le type de segment. Dans le message ci-dessus, il existe cinq segments : MSH (en-tête de message), PID (identification du patient), deux segments NK1 (proche parent) et IN1 (assurance).

Chaque segment se compose de champs. Un *champ* contient des informations liées à l'objectif du segment, telles que le nom de la compagnie d'assurance dans le segment IN1 (assurance). En général (mais pas systématiquement), les champs sont délimités par un caractère |.

Les champs peuvent être divisés en *composants*. Les composants sont généralement indiqués par un caractère ^. Dans l'exemple ci-dessus, le segment PID (identification du patient) contient un champ de nom du patient contenant LEVERKUHN^ADRIAN^C, divisé en trois parties : nom de famille (LEVERKUHN), prénom (ADRIAN), et initiale du deuxième prénom (C). Les composants peuvent être divisés en *sous-composants*. En règle générale, les sous-composants sont indiqués par un caractère &.

Voici un exemple de message HL7 :

```
MSH|^~\&||.||||199908180016||ADT^A04|ADT.1.1698593|P|2.7
PID|1||000395122||LEVERKUHN^ADRIAN^C||19880517180606|M|||6 66TH AVE
NE^^WEIMAR^DL^98052|| (157) 983-3296|||S||12354768|87654321
NK1|1|TALLIS^THOMAS^C|GRANDFATHER|12914 SPEM
ST^^ALIUM^IN^98052| (157) 883-6176
NK1|2|WEBERN^ANTON|SON|12 STRASSE MUSIK^^VIENNA^AUS^11212| (123) 456-7890
IN1|1|PRE2||LIFE PRUDENT BUYER|PO BOX
23523^WELLINGTON^ON^98111|||19601||||||THOMAS^JAMES^M|F||||||||||||||ZKA535529776
```

**Remarque :** Pour créer un fichier HL7 à l'aide de l'exemple de texte donné :

1. Copiez et collez l'exemple de texte dans un nouveau document à l'aide d'un logiciel d'édition de texte, comme Notepad++.
2. Apportez les modifications de contenu nécessaires.
3. Configurez les paramètres pour afficher EOL (End of Line – fin de ligne) sur le texte. Dans Notepad++, accédez à **Vue > Afficher les symboles > Afficher End of Line**.
4. Remplacez le format de conversion EOL (End of Line – fin de ligne) par CR (Carriage Return – retour chariot). Dans Notepad++, accédez à **Édition > Conversion EOL > Ancien format Mac**.
5. Enregistrez le fichier HL7 après avoir apporté cette modification au format.

### Onglet de propriétés du fichier

Nom du champ	Description
Server name	Indique si le fichier que vous sélectionnez comme entrée figure sur l'ordinateur exécutant Enterprise Designer ou sur le serveur Spectrum™ Technology Platform. Si vous sélectionnez un fichier sur l'ordinateur local, le nom du serveur sera Mon ordinateur. Si vous sélectionnez un fichier sur le serveur, le nom du serveur sera Spectrum™ Technology Platform.
Nom de fichier	Indique le chemin du fichier. Cliquez sur le bouton de sélection (...) pour trouver le fichier souhaité.  <b>Remarque :</b> Si le Spectrum™ Technology Platform est en cours d'exécution sur les systèmes Unix ou Linux, rappelez-vous que les noms de fichiers et les chemins sur ces plates-formes sont sensibles à la casse.
Version HL7	Version de la norme HL7 utilisée dans le fichier indiqué. Par exemple, « 2.7 » signifie que le fichier utilise la norme HL7 version 2.7. La version HL7 est indiquée par le 12ème champ du segment MSH.

Nom du champ	Description
Encodage des caractères	L'encodage du fichier texte. Sélectionnez l'un des éléments suivants :
<b>UTF-8</b>	Prend en charge tous les caractères Unicode et est rétrocompatible avec le format ASCII. Pour plus d'informations sur UTF, reportez-vous à <a href="https://unicode.org/faq/utf_bom.html">unicode.org/faq/utf_bom.html</a> .
<b>UTF-16</b>	Prend en charge tous les caractères Unicode, mais n'est pas rétrocompatible avec le format ASCII. Pour plus d'informations sur UTF, reportez-vous à <a href="https://unicode.org/faq/utf_bom.html">unicode.org/faq/utf_bom.html</a> .
<b>US-ASCII</b>	L'encodage des caractères en fonction de l'ordre alphabétique anglais.
<b>UTF-16BE</b>	Encodage UTF-16 avec sérialisation d'octets gros-boutiste (octet le plus significatif en premier).
<b>UTF-16LE</b>	Encodage UTF-16 avec sérialisation d'octets petit-boutiste (octet le moins significatif en premier).
<b>ISO-8859-1</b>	Chiffrement de caractères ASCII utilisé habituellement pour les langues européennes occidentales. Également désigné par l'acronyme Latin-1.
<b>ISO-8859-3</b>	Chiffrement de caractères ASCII utilisé habituellement pour les langues européennes méridionales. Également connu sous le nom Latin-3.
<b>ISO-8859-9</b>	Chiffrement de caractères ASCII utilisé habituellement pour la langue turque. Également connu sous le nom Latin-5.
<b>CP850</b>	Une page de code ASCII utilisée pour écrire les langues européennes occidentales.
<b>CP500</b>	Une page de code EBCDIC utilisée pour écrire les langues européennes occidentales.
<b>Shift_JIS</b>	Un encodage de caractères pour la langue japonaise.
<b>MS932</b>	Extension Microsoft de Shift_JIS permettant d'inclure les caractères spéciaux NEC, la sélection NEC des extensions IBM et les extensions IBM.
<b>CP1047</b>	Une page de code EBCDIC avec le jeu de caractères Latin-1 complet.

Nom du champ	Description
Validate	<p>Ces options déterminent si le fichier est contrôlé pour s'assurer qu'il est conforme à la norme HL7 2.7. Si un message du fichier échoue à la validation, il est traité comme un enregistrement non conforme et les options d'enregistrement non conforme spécifiées pour le job (dans Enterprise Designer sous <b>Édition &gt; Options de job</b>) ou pour le système (dans Management Console) prennent effet.</p> <p><b>Champs requis</b> Cochez cette case pour vous assurer que chaque segment, champ, composant et sous-composant contient les éléments requis par la norme HL7 2.7.</p> <p><b>Longueur</b> Cochez cette case pour vous assurer que chaque élément répond aux exigences minimales et maximales de longueur pour l'élément, tel que défini dans la norme HL7 2.7.</p>

Nom du champ	Description
--------------	-------------

---

Ignorer les éléments imprévus	
-------------------------------	--

## Nom du champ

## Description

Sélectionnez ces options si vous souhaitez autoriser les messages à contenir des segments, des champs, des composants et des sous-composants ne figurant pas à l'emplacement prévu. Les emplacements prévus sont définis dans la norme HL7 ou, dans le cas de types de message personnalisés, dans l'outil HL7 Schema Management d'Enterprise Designer.

Par exemple, notez le schéma de message personnalisé suivant :

```
MSH
[PID]
{ZSS}
PV1
NK1
{[DG1]}
```

Et les données suivantes :

```
MSH|^~\&|Pharm|GenHosp|CIS|GenHosp|198807050000||RAS^O17|RAS1234|P|2.7
ZSS|100|abc
PID|1234||PATID1234^5^M11^ADT1^MR^GOOD HEALTH
HOSPITAL~123456789^^^USSSA^SS|
PV1||O|O/R|||0148^ADDISON, JAMES|0148^ADDISON, JAMES
NK1|Son|John
```

Dans ce cas, le segment `PID` est imprévu, car il figure avant le segment `ZSS`.

Les messages contenant des éléments figurant à des emplacements imprévus sont traités comme des enregistrements non conformes et les options d'enregistrement non conforme spécifiées pour le job (dans Enterprise Designer, sous **Édition > Options de job**) ou pour le système (dans Management Console) prennent effet.

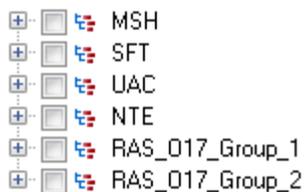
Par défaut, toutes les options **Ignorer les éléments imprévus** sont activées pour permettre le traitement d'autant d'enregistrements que possible.

- |                        |  |
|------------------------|--|
| <b>Segments</b>        | Cochez cette case pour que les messages puissent contenir des segments non définis dans la norme HL7 2.7. Des segments imprévus sont ignorés et d'autres segments du message sont traités.                                 |
| <b>Champs</b>          | Cochez cette case pour que les segments puissent contenir des champs non définis dans la norme HL7 2.7. Les fichiers imprévus sont ignorés et les autres champs du segment sont traités.                                   |
| <b>Composants</b>      | Cochez cette case pour que les champs puissent contenir des composants non définis dans la norme HL7 2.7. Les composants imprévus sont ignorés et les autres composants du champ sont traités.                             |
| <b>Sous-composants</b> | Cochez cette case pour que les composants puissent contenir des sous-composants non définis dans la norme HL7 2.7. Les sous-composants imprévus dans le composant sont ignorés et les autres sous-composants sont traités. |

## Onglet Champs

L'onglet **Champs** s'affiche dans les segments, les champs, les composants et les sous-composants. Utilisez l'onglet **Champs** pour sélectionner les données à lire dans le dataflow.

Les groupes de segments, qui sont des collections de segments utilisés ensemble pour contenir une catégorie de données, sont affichés à l'aide d'un système de numérotation qui indique l'emplacement du groupe dans le schéma de message. À chaque groupe de segments est attribué un libellé « Group\_n », où « n » est un numéro qui correspond à l'emplacement du groupe dans le schéma de message. Pour illustrer le fonctionnement du système de numérotation, prenez l'exemple suivant :

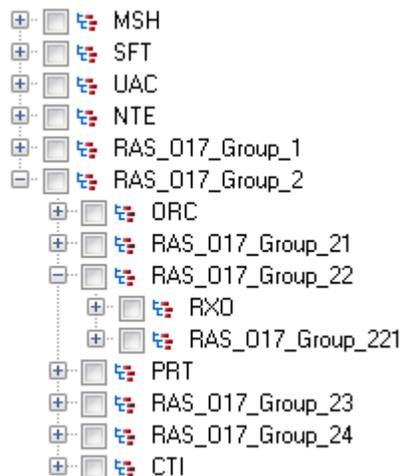


Cet exemple affiche la liste de champs du message RAS^017. Ce message comprend deux groupes de segments : RAS\_017\_Group\_1 et RAS\_017\_Group\_2. Le groupe de segments « Group\_1 » fait référence au premier groupe de segments du schéma RAS^017, tandis que le deuxième groupe, « Group\_2 », fait référence au deuxième groupe répertorié dans le schéma RAS^017.

Pour déterminer le groupe de segments représenté par « Group\_1 » et « Group\_2 », recherchez la description du message RAS^017 dans le document *Norme HL7 Version 2.7*. Vous pouvez télécharger une copie de ce document depuis [www.hl7.org](http://www.hl7.org).

Dans la description du message, recherchez le premier groupe, qui, dans le cas de RAS^017, est le groupe PATIENT. Le deuxième groupe du schéma est le groupe ORDER.

Les groupes de segments nichés sous un groupe de segments porte un numéro supplémentaire ajouté à leur numéro de groupe. Par exemple, Group\_21 représente le premier groupe niché sous le deuxième groupe. D'autres sous-groupes portent des numéros supplémentaires qui leur sont ajoutés, comme Group\_221, qui, pour le message RAS^017, représente le groupe de segments ORDER\_DETAIL\_SUPPLEMENT. Voici un exemple de groupes nichés :



Les contrôles de l'onglet **Champs** sont décrits dans le tableau suivant.

**Tableau 2 : Onglet Champs**

Nom de l'option	Description
Régénérer	<p>Cliquez sur ce bouton pour renseigner l'onglet <b>Champs</b> avec une liste de tous les segments, champs, composants et sous-composants pour le type de message contenu dans le fichier d'entrée. Tous les éléments du type de message sont affichés en fonction du schéma HL7, que le fichier d'entrée contienne tous les éléments ou non. Par exemple, si le fichier contient un message RAS, le schéma du type de message RAS entier sera affiché que le fichier contienne ou non les données de tous les segments, champs, composants et sous-composants.</p> <p>Si vous avez défini des éléments personnalisés à l'aide de l'outil HL7 Schema Management d'Enterprise Designer, ces éléments sont également répertoriés.</p>
Développer tout	Développe tous les éléments dans l'onglet Champs pour afficher tous les segments, champs, composants et sous-composants des types de messages contenus dans le fichier.
Réduire tout	Ferme tous les nœuds dans la vue pour que seuls les segments soient affichés. Utilisez cette option pour afficher aisément les segments des types de messages dans le fichier. Vous pouvez ensuite développer des segments individuels pour afficher les champs, les composants et les sous-composants dans un segment.
Sélectionner tout	Cochez cette case pour créer des champs de dataflow pour tous les segments, champs, composants et sous-composants pour tous les types de messages inclus dans le fichier.

## Simplification des données HL7

Les données HL7 sont organisées de manière hiérarchique. Cependant, de nombreux stages requièrent des données au format plat, vous devrez donc convertir les données afin de les rendre utilisables par les stages en aval pour des éléments tels que la validation d'adresse ou le géocodage.

La procédure suivantes explique comment utiliser un stage Splitter pour simplifier des données HL7.

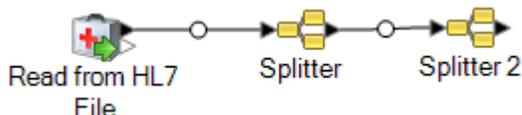
1. Ajoutez un stage Read from HL7 File à votre flux de données et configurez le stage.
2. Ajoutez un stage Splitter et reliez-le à Read from HL7 File.

- Ajoutez d'autres stages Splitter en fonction de vos besoins, afin de disposer d'un stage Splitter pour chaque segment, champ ou composant à simplifier.

**Remarque :** Vous ne devez que simplifier les données à traiter dans un stage en aval. Les autres données peuvent rester au format hiérarchique. Par exemple, si vous souhaitez traiter uniquement les données d'adresse, il vous suffit de les simplifier.

- Connectez tous les stages Splitter.

Votre flux de données doit désormais ressembler à ce qui suit :



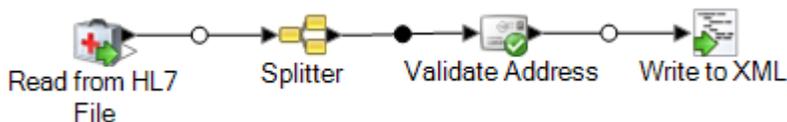
- Double-cliquez sur le premier stage Splitter pour ouvrir les options correspondantes.
- Dans le champ **Split at**, sélectionnez le segment, le champ ou le composant à simplifier.
- Cliquez sur **OK**.
- Configurez chaque stage Splitter supplémentaire, en sélectionnant un autre segment, champ ou composant dans chaque champ **Split at** de Splitter.
- Ajoutez d'autres stages en fonction de vos besoins après le dernier stage Splitter afin de compléter votre flux de données.

### Exemple

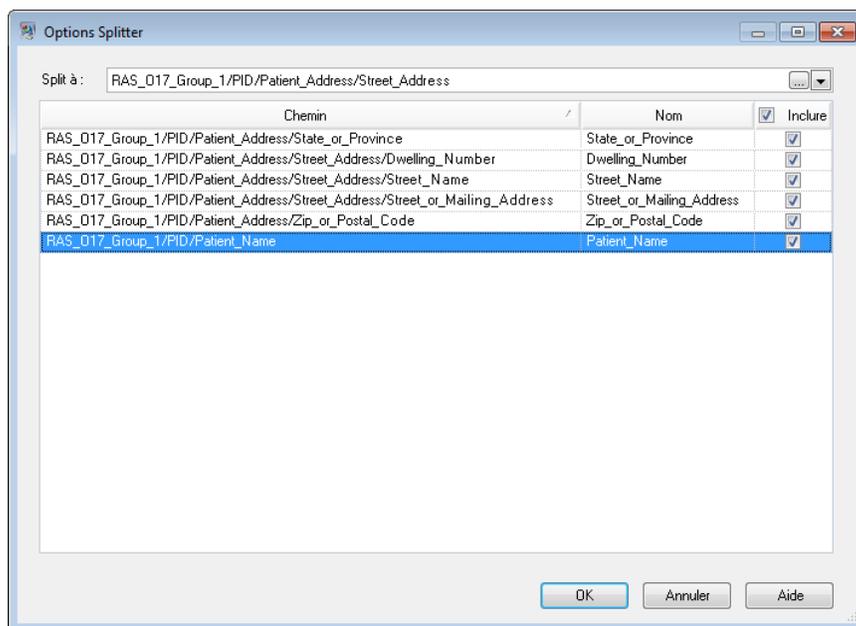
Vous disposez des données HL7 suivantes et vous souhaitez valider l'adresse dans le segment PID.

```
MSH|^~\&||.||||199908180016||RAS^O17|ADT.1.1698594|P|2.7
PID|1||000395122||SMITH^JOHN^D||19880517180606|M|||One Global
View^^Troy^NY^12180||(630)123-4567|||S||12354768|87654321
```

Pour ce faire, vous devez convertir ces données d'adresse en données plates afin que le stage Validate Address puisse les traiter. Ainsi, vous créez un flux de données contenant un stage Splitter, suivi d'un stage Validate Address, comme suit :



Le stage Splitter est configuré pour effectuer une division au composant PID/Patient\_Address/Street\_Address, qui convertit ces données en données plates.



Le canal qui connecte le stage Splitter au stage Validate Address renomme les champs pour utiliser les noms de champ requis par Validate Address :  
 Street\_or\_Mailing\_Addres est renommé en AddressLine1, State\_or\_Province est renommé en StateProvince et Zip\_or\_Postal\_Code est renommé en PostalCode.

Dans cet exemple, la sortie est écrite dans un fichier XML contenant ces données.

```
<?xml version='1.0' encoding='UTF-8'?>
<XmlRoot xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <PatientInformation>
    <Confidence>95</Confidence>
    <RecordType>Normal</RecordType>
    <CountryLevel>A</CountryLevel>
    <ProcessedBy>USA</ProcessedBy>
    <MatchScore>0</MatchScore>
    <AddressLine1>1 Global Vw</AddressLine1>
    <City>Troy</City>
    <StateProvince>NY</StateProvince>
    <PostalCode>12180-8371</PostalCode>
    <PostalCode.Base>12180</PostalCode.Base>
    <PostalCode.AddOn>8371</PostalCode.AddOn>
    <Country>United States Of America</Country>
    <Patient_Name>
      <Family_Name>
        <Surname>SMITH</Surname>
      </Family_Name>
      <Given_Name>JOHN</Given_Name>
    <Second_and_Further_Given_Names_or_Initials_Thereof>
```

```

D
</Second_and_Further_Given_Names_or_Initials_Thereof>
  </Patient_Name>
</PatientInformation>
</XmlRoot>

```

## Ajout d'un message HL7 personnalisé

Le stage Read from HL7 File valide les messages à l'aide du schéma HL7 2.7. Cependant, vos données HL7 peuvent contenir des messages ne faisant pas partie de la norme HL7. Si vous souhaitez que le stage Read from HL7 File valide vos données HL7 personnalisées, vous devez créer un schéma HL7 personnalisé. Cette rubrique explique comment créer un schéma HL7 personnalisé à l'aide de l'outil HL7 Schema Management. Pour plus d'informations sur HL7, consultez l'adresse [www.hl7.org](http://www.hl7.org).

1. Dans Enterprise Designer, accédez à **Outils > HL7 Schema Management**.

Cette opération ouvre la fenêtre HL7 Schema Management contenant une liste de messages pris en charge. Ces messages sont prédéfinis par HL7.

2. Dans la fenêtre **HL7 Schema Management**, cliquez sur **Ajouter**.
3. Dans le champ **Message type**, indiquez un type de message HL7 personnalisé.

Le type de message indique les informations associées à la santé fournies dans le message. Par exemple, un type de message ADT (Admin Discharge Transfer) permet d'échanger l'état d'un patient au sein d'un organisme de soins de santé et un type de message ORU (Observation Result) permet de transmettre les observations et les résultats du LIS (Lab Information System) avec le HIS (Hospital Information System).

4. Dans le champ **Trigger event**, indiquez un code d'événement.

L'événement déclencheur est un événement réel qui initie une communication et l'envoi d'un message. Le type de message et l'événement déclencheur figurent dans le champ MSH-9 du message. Par exemple, le champ MSH-9 peut contenir la valeur ADT^A01. Cela signifie qu'ADT est le type de message HL7 et A01 l'événement déclencheur.

5. Dans le champ **Description**, saisissez la description d'un message HL7 personnalisé.

Ce champ vous permet de mieux comprendre un type de message. Par exemple, si vous ajoutez le type de message XYZ, vous pouvez fournir une description utilisée pour échanger l'état du patient au sein de l'organisme de soins de santé.

Vous voyez maintenant apparaître un nouveau message créé sous **Definition**. Cliquez sur le signe plus pour développer le message. Vous pouvez voir que le segment MSH est automatiquement ajouté.

6. Pour ajouter un segment existant à un message
  - a) Cliquez sur **Select Segment**.
  - b) Sélectionnez les segments que vous souhaitez ajouter au message, puis cliquez sur **OK**.  
Un schéma du segment sélectionné s'affiche dans la grille **Schéma de segment** et les messages sélectionnés sont ajoutés au schéma du message.
  
7. Pour ajouter un segment personnalisé à un message
  - a) Cliquez sur **Select Segment**.
  - b) Cliquez sur **Add Segment**.
  - c) Dans le champ **Name**, indiquez le nom du segment, puis cliquez sur **OK**.  
Les nouveaux segments ajoutés apparaissent au bas de la liste **Segments pris en charge**.
  - d) Sélectionnez le segment personnalisé ajouté, puis cliquez sur le bouton **Add field**.
  - e) Dans le champ **Name**, indiquez un nom de champ pour le segment sélectionné.  
Par exemple, un segment PID (Patient information) contient des noms de champ Patient ID, Patient Name, Patient Address, County Code, etc.
  - f) Dans le champ **Type**, sélectionnez un type de données approprié.  
Les types de données HL7 définissent le type de données susceptible d'être inclus dans un champ. Ils sont utilisés tout au long de la structure de messages HL7. Par exemple, ST pour string (chaîne), TX pour données textuelles et FT pour données formatées.
  - g) Dans le champ **Normative length**, indiquez la longueur minimale et la longueur maximale du champ au format : m..n. Vous pouvez également indiquer une liste de valeurs possibles pour une longueur du champ au format suivant : x,y,z.  
Par exemple, la longueur 1..3 signifie que la longueur de l'élément peut être 1, 2 ou 3 et la longueur 1, 3, 4 signifie que la longueur de l'élément peut être 1, 3 ou 4, mais pas 2. Une valeur différente de 1, 3 et 4 sera traitée comme non valide.
  - h) Dans le champ **Optionality**, indiquez si un champ est facultatif ou obligatoire.
 

**O**  
Le champ est facultatif.

**R**  
Le champ est obligatoire.
  - i) Dans le champ **Repetition**, si vous souhaitez autoriser le champ à apparaître plusieurs fois dans le segment, cochez la case **Repetition** et indiquez le nombre de fois que le champ peut être utilisé.  
Par exemple, une valeur 3 signifie que le champ peut avoir trois occurrences. Si cette valeur n'est pas définie, il ne peut exister qu'une occurrence, et, par conséquent, le champ ne sera pas répété.
  
8. Cliquez sur **OK**.

Vous pouvez également sélectionner les options **Optional** et **Repeating** des propriétés du segment.

9. Sélectionnez **Optional** pour que le champ sélectionné soit facultatif, et **Repeating** pour autoriser la répétition d'un segment sélectionné dans un message.
10. Cliquez sur **OK**.

Le nouveau message HL7 ajouté apparaît au bas de la liste.

## Read from NoSQL DB

Le stage Read from NoSQL DB lit des données à partir d'une table de base de données comme entrée d'un dataflow. Le stage prend en charge les types de base de données MongoDB et Couchbase.

### Onglet Général

Nom du champ	Description
Connexion	<p>Sélectionnez la connexion à la base de données requise dans la liste déroulante. Les options affichées sont basées sur les connexions définies dans Management Console.</p> <p>Pour ajouter une nouvelle connexion, voir <a href="#">Connexion à NoSQL</a> à la page 54.</p> <p>Pour modifier une connexion existante, sélectionnez-la et ouvrez-la dans la liste des connexions de la page Sources de données de Management Console, effectuez les mises à jour requises, puis cliquez sur le bouton <b>Enregistrer</b>.</p>
Table/Vue	<p>Indique la collection ou la vue de la base de données que vous souhaitez interroger.</p> <p><b>Remarque</b> : Même si le terme <b>Table/Vue</b> est utilisé dans l'interface utilisateur, MongoDB l'appelle <i>collection</i> et Couchbase l'appelle <i>vue</i>.</p>
Fichier de schéma	<p>Cliquez sur le bouton Parcourir (...) pour sélectionner un fichier de schéma JSON. Ce fichier est facultatif. Les champs de l'onglet Champs peuvent être régénérés via le fichier de schéma ou la table/vue de base de données.</p> <p>Pour effacer le chemin d'accès au fichier sélectionné, cliquez sur <b>Effacer</b>.</p> <p><b>Remarque</b> : Si un fichier de schéma est sélectionné, les champs sont toujours générés via ce fichier de schéma.</p>

Nom du champ	Description
Où	<p>Entrez les critères de filtre requis, le cas échéant, à l'aide de la syntaxe de MongoDB, pour récupérer des enregistrements spécifiques. Laissez le champ vide si aucun critère de filtre n'est requis.</p> <p>La syntaxe suivante est pour une clause avec un opérateur <i>égal à</i> :</p> <pre>{ "&lt;column name&gt;" : "&lt;filter value&gt;" }</pre> <p>Vous pouvez lier plusieurs clauses via les opérateurs requis. Pour obtenir la liste de tous les opérateurs inclus dans la clause <i>where</i>, consultez la section <a href="http://docs.mongodb.org/manual/reference/operator/query/">http://docs.mongodb.org/manual/reference/operator/query/</a>.</p> <p>Par exemple, pour récupérer des enregistrements où la valeur de la colonne <i>customer_name</i> correspond à la valeur <i>John</i>, et la valeur de la colonne <i>customer_age</i> est supérieure ou égale à 45, procédez comme suit :</p> <pre>{ \$and: [ { "customer_name": "John" }, { \$gte: [ "customer_age", "45" ] } ] }</pre> <p><b>Avertissement</b> : Assurez-vous que vous n'incluez pas le mot clé <i>where</i> dans ce champ.</p> <p><b>Remarque</b> : Actuellement, ce champ est visible uniquement lors de la sélection d'une connexion MongoDB.</p>
Ignorer les champs absents	<p>Si cette option est sélectionnée, les champs définis dans le schéma, s'ils ne figurent pas dans l'enregistrement réel, ne sont pas transmis au stage suivant.</p> <p><b>Remarque</b> : Si vous n'activez pas cette option, les champs absents de la table de base de données ou de la vue sont ajoutés et traités avec la valeur NULL.</p>
Aperçu	<p>Affiche les enregistrements de la table sélectionnée.</p> <p><b>Remarque</b> : Pour les sources de données MongoDB, cliquer sur <b>Aperçu</b> affiche les enregistrements filtrés, si une ou plusieurs clauses <i>where</i> ont été saisies dans le champ <b>Où</b>. En l'absence de clause <i>where</i>, l'aperçu n'affiche tous les enregistrements.</p> <p><b>Remarque</b> : Pour les sources de données Couchbase, cliquer sur <b>Aperçu</b> affiche également le champ <i>_id</i> ajouté contenant la clé. Si l'enregistrement comporte déjà un champ <i>_id</i>, il est écrasé par le champ <i>_id</i> ajouté lors de l'aperçu des champs.</p>
Développer tout	Développe les éléments de l'arborescence d'aperçu.
Réduire tout	Réduit les éléments de l'arborescence d'aperçu.

### Onglet Champs

L'onglet Champs vous permet de sélectionner les données que vous souhaitez transmettre au stage suivant. (Pour plus d'informations, reportez-vous à la section [Définition de champs dans une base de données NoSQL](#) à la page 187.)

## Définition de champs dans une base de données NoSQL

L'onglet **Champs** affiche le champ et son type, tels que définis dans le fichier de schéma/NoSQL DB.

1. Dans l'onglet **Champs**, cliquez sur **Régénérer**.

Cette opération génère les données agrégées en fonction des 50 premiers enregistrements.

Les données sont affichées au format suivant : `Fieldname (datatype)`.

**Remarque :** Si le fichier de schéma est localisé, les champs sont générés via le fichier de schéma et la table ou la vue est ignorée. Pour réinitialiser le fichier de schéma, cliquez sur **Effacer**.

**Remarque :** Lors de la lecture de données depuis Couchbase DB, la clé de chaque enregistrement est également lue. Cette clé est stockée avec l'enregistrement à l'aide d'un champ `_id` ajouté lors de la régénération des champs et elle est également incluse dans les données envoyées au stage suivant. Si l'enregistrement comporte déjà un champ `_id`, il est écrasé par le champ `_id` ajouté lors de la régénération des champs.

2. Pour modifier le nom et le type d'un champ, mettez le champ en surbrillance, puis cliquez sur **Modifier**.
3. Dans le champ **Name**, choisissez le champ à ajouter ou saisissez le nom du champ.
4. Dans le champ **Type**, vous pouvez laisser le type de données défini sur chaîne si vous ne prévoyez pas d'effectuer des opérations mathématiques sur les données. Cependant, si vous avez l'intention d'effectuer ce genre d'opérations, sélectionnez un type de données approprié. Cela convertira les données de chaîne à partir du fichier en un type de données qui activera la manipulation appropriée des données dans le flux de données.

Le stage prend en charge les types de données suivants :

<b>double</b>	Un type de données numérique qui contient à la fois des nombres en double précision négatifs et positifs entre $2^{-1074}$ et $(2-2^{-52}) \times 2^{1023}$ . En notation E, la plage de valeurs est comprise entre $-1.79769313486232E+308$ et $1.79769313486232E+308$ .
<b>flottant</b>	Un type de données numérique contenant des nombres de précision seuls négatifs et positifs entre $2^{-149}$ et $(2-2^{-23}) \times 2^{127}$ . En notation E, la plage de valeurs est comprise entre $-3.402823E+38$ et $3.402823E+38$ .
<b>integer</b>	Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre $-2^{31}$ ( $-2,147,483,648$ ) et $2^{31}-1$ ( $2,147,483,647$ ).

**long** Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre  $-2^{63}$  (-9,223,372,036,854,775,808) et  $2^{63}-1$  (9,223,372,036,854,775,807).

**chaîne** Séquence de caractères.

- Vous pouvez également ajouter des champs supplémentaires non présents dans le fichier de schéma ou la table. Cliquez sur **Ajouter** pour ajouter un nouveau champ. Pour supprimer un champ, cliquez sur **Supprimer**.

**Remarque :** Vous ne pouvez ajouter un nouveau champ que sous le type de liste.

- Cliquez sur **OK**.

## Options de flux de données NoSQL DB

Cette procédure explique comment configurer un flux de données pour qu'il prenne en charge les options d'exécution NoSQL DB.

- Ouvrez le flux de données dans Enterprise Designer.
- Pour configurer les options d'exécution d'un stage dans un flux de données intégré, ouvrez ce dernier.
- Cliquez sur l'icône Options de flux de données de la barre d'outils ou sur **Édition > Options de flux de données**. La boîte de dialogue **Options de flux de données** apparaît.
- Cliquez sur **Ajouter**. La boîte de dialogue **Définir l'option de flux des données** apparaît.
- Développez le stage NoSQLDB.
- Les options de flux de données sont exposées comme décrit dans le tableau suivant :

Base de données	Lire	Écriture
Mongo DB	Connexion	Connexion
	table	table
Couchbase DB	Connexion	Connexion
	Vue	
	Nom de document de conception	

Le nom d'option NoSQL DB sélectionné s'affiche dans les champs **Nom de l'option** et **Libellé d'option**. Il s'agit du nom de l'option qui devra être spécifié lors de l'exécution afin de définir cette option.

- Saisissez une description de l'option dans le champ **Description**.

8. Dans le champ **Target**, sélectionnez l'option **Selected stage(s)**.
9. Si vous voulez limiter les valeurs qu'un utilisateur peut sélectionner, modifiez les options disponibles dans le champ **Legal values** en cliquant sur l'icône située directement à droite de ce champ.
10. Pour modifier la valeur par défaut, indiquez une valeur différente dans le champ **Valeur par défaut**.

**Remarque :** Pour un service, vous ne pouvez modifier que les valeurs par défaut avant d'exposer le service pour la première fois. Une fois que vous exposez le service, vous ne pouvez plus modifier les valeurs par défaut à l'aide d'Enterprise Designer. Au lieu de cela, vous devez utiliser Management Console. Pour plus d'informations, reportez-vous à la section [Spécification des Options de services par défaut](#).

11. Cliquez sur **OK**.
12. Continuez d'ajouter les options souhaitées.
13. Cliquez sur **OK** dans la boîte de dialogue des **Options du flux de données** lorsque vous avez terminé d'ajouter vos options.
14. Enregistrez et exposez le flux de données.

## Read from SAP

Le stage Read from SAP lit des données d'une base de données comme entrée d'un flux de données. Il peut lire les données d'une ou de plusieurs tables. Lors de la lecture de données de plusieurs tables, le stage effectue une opération de jointure pour déterminer les enregistrements à lire dans le flux de données.

## Connexion à SAP

Pour lire des données de SAP dans un flux de données à l'aide de Read from SAP, vous devez créer une connexion entre Spectrum™ Technology Platform et votre système SAP.

1. Ouvrez le Gestionnaire de connexions SAP. Vous pouvez le faire dans Enterprise Designer, sous **Outils > Gestion des connexions SAP**, ou dans le stage Read from SAP, en cliquant sur le bouton **Gérer** en regard du champ **Connexion**.
2. Cliquez sur **Ajouter**.
3. Dans le champ **Nom de connexion**, attribuez un nom à cette connexion.
4. Renseignez les autres champs à l'aide des informations du serveur SAP auquel vous souhaitez vous connecter. Pour obtenir les informations nécessaires, contactez votre administrateur SAP Basis.

**Important :** L'ID d'utilisateur et le mot de passe doivent être ceux d'un compte SAP avec des droits d'accès d'administrateur.

5. Cliquez sur **Test** pour vérifier la connexion.
6. Cliquez sur **OK**.

Vous n'avez pas créé de connexion pouvant être utilisée par le stage Read from SAP pour lire des données de SAP dans un flux de données.

## Lecture des données d'une seule table SAP

Le stage Read from SAP peut être configuré pour lire les données d'une seule table ou de plusieurs tables de la base de données SAP. Cette procédure explique comment configurer Read from SAP pour lire les données d'une seule table.

1. Dans Enterprise Designer, glissez Read from SAP sur le canevas.
2. Double-cliquez sur le stage Read from SAP du canevas.
3. Dans le champ **Connexion**, sélectionnez le serveur SAP contenant les données que vous souhaitez lire dans le flux de données. S'il n'existe aucune connexion définie pour le serveur SAP, vous devez créer la connexion en cliquant sur **Gérer**.
4. Dans le champ **Type de source**, sélectionnez **Seule**.
5. Cliquez sur **Sélectionner**.
6. Sélectionnez la table que vous souhaitez lire dans le flux de données, puis cliquez sur **OK**.

**Remarque :** Seules les 200 premières tables sont répertoriées. Utilisez la fonction de recherche pour trouver des tables non répertoriées dans les 200 premières. Le champ de recherche recherche uniquement les valeurs des colonnes **Nom** et **Intitulé**.

7. Pour afficher les noms de champ à utiliser dans le flux de données, cochez la case **Afficher le nom technique**.

Dans SAP, les champs portent un nom convivial utilisé à des fins d'affichage et un nom unique qui peut être moins lisible. Par exemple, un champ peut porter un nom convivial « Canal de distribution » et un nom technique « DIS\_CHANNEL ». Pour s'assurer que le nom de champ est valide dans le flux de données, il convient d'utiliser le nom technique comme nom de champ.

8. Cochez la case de la colonne **Inclure** pour chaque champ que vous souhaitez lire dans le flux de données.
9. Cliquez sur **OK**.
10. Si vous souhaitez lire uniquement certains enregistrements, vous pouvez indiquer des conditions de filtre dans l'onglet **Filtre**. Pour qu'un enregistrement soit lu dans le flux de données, il doit remplir l'ensemble des conditions que vous définissez.
11. Vous pouvez améliorer les performances en indiquant une taille d'extraction appropriée dans l'onglet **Exécution**.

Sélectionnez cette option pour indiquer le nombre d'enregistrements à lire en une seule fois depuis la table de base de données. Par exemple, si la valeur de **Taille de l'extraction** est 100 et que le nombre total d'enregistrements à lire est 1 000, la base de données lit l'ensemble des enregistrements en 10 fois.

La définition d'une valeur Taille de l'extraction optimale peut considérablement améliorer les performances.

**Remarque :** Vous pouvez calculer une taille d'extraction optimale pour votre environnement en testant les temps d'exécution entre un stage Read from DB et un stage Write to Null. Pour plus d'informations, reportez-vous à la section **Définition d'une taille d'extraction optimale** à la page 316.

La taille d'extraction par défaut de Read from SAP est 10 000.

12. Cliquez sur **OK**.

Le stage Read from SAP est désormais configuré pour lire les données d'une seule table de la base de données SAP dans le flux de données.

## Lecture des données de plusieurs tables SAP

Le stage Read from SAP peut être configuré pour lire les données d'une seule table ou de plusieurs tables de la base de données SAP. Cette procédure explique comment configurer Read from SAP pour lire les données de plusieurs tables. Pour lire les données de plusieurs tables, vous définissez une instruction JOIN pour combiner les données en un flux unique.

1. Dans Enterprise Designer, glissez Read from SAP sur le canevas.
2. Double-cliquez sur le stage Read from SAP du canevas.
3. Dans le champ **Connexion**, sélectionnez le serveur SAP contenant les données que vous souhaitez lire dans le flux de données. S'il n'existe aucune connexion définie pour le serveur SAP, vous devez créer la connexion en cliquant sur **Gérer**.
4. Dans le champ **Type de source**, sélectionnez **Multiple**.
5. Cliquez sur **Ajouter**.
6. Sélectionnez les tables que vous souhaitez lire dans le flux de données, puis cliquez sur **OK**.

**Remarque :** Seules les 200 premières tables sont répertoriées. Utilisez la fonction de recherche pour trouver des tables non répertoriées dans les 200 premières. Le champ de recherche recherche uniquement les valeurs des colonnes **Nom** et **Intitulé**.

7. Sélectionnez la première table de la liste, puis cliquez sur **Créer une relation**. Il s'agit de la table source.
8. Dans le champ **Clé source**, sélectionnez la colonne de la table source dont la valeur sera utilisée pour rapprocher les enregistrements des enregistrements de l'autre table.
9. Dans le champ **Type de jointure**, sélectionnez l'un des éléments suivants :

<b>INNER JOIN</b>	Renvoie uniquement les enregistrements contenant une correspondance entre les tables source et cible.
<b>LEFT JOIN</b>	Renvoie tous les enregistrements de la table source, même s'il n'existe aucune correspondance entre les tables source et cible. Cette option renvoie l'ensemble des enregistrements de la table source plus tout enregistrement correspondant à la table cible.

10. Dans le champ **Table**, sélectionnez la table cible.
11. Dans le champ **Clé de table**, sélectionnez la colonne de la table cible contenant les données que vous souhaitez comparer aux données du champ **Clé source** pour déterminer si l'enregistrement remplit la condition de jointure.
12. Cliquez sur **OK**.
13. Cliquez sur **Sélectionner un schéma**.
14. Sélectionnez les champs que vous souhaitez lire dans le flux de données. Pour afficher les noms de champ à utiliser dans le flux de données, cochez la case **Afficher le nom technique**.

Dans SAP, les champs portent un nom convivial utilisé à des fins d'affichage et un nom unique qui peut être moins lisible. Par exemple, un champ peut porter un nom convivial « Canal de distribution » et un nom technique « DIS\_CHANNEL ». Pour s'assurer que le nom de champ est valide dans le flux de données, il convient d'utiliser le nom technique comme nom de champ.

15. Cliquez sur **OK**.
16. Si vous souhaitez lire uniquement certains enregistrements, vous pouvez indiquer des conditions de filtre dans l'onglet **Filtre**. Pour qu'un enregistrement soit lu dans le flux de données, il doit remplir l'ensemble des conditions que vous définissez.
17. Vous pouvez améliorer les performances en indiquant une taille d'extraction appropriée dans l'onglet **Exécution**.

Sélectionnez cette option pour indiquer le nombre d'enregistrements à lire en une seule fois depuis la table de base de données. Par exemple, si la valeur de **Taille de l'extraction** est 100 et que le nombre total d'enregistrements à lire est 1 000, la base de données lit l'ensemble des enregistrements en 10 fois.

La définition d'une valeur Taille de l'extraction optimale peut considérablement améliorer les performances.

**Remarque :** Vous pouvez calculer une taille d'extraction optimale pour votre environnement en testant les temps d'exécution entre un stage Read from DB et un stage Write to Null. Pour plus d'informations, reportez-vous à la section [Définition d'une taille d'extraction optimale](#) à la page 316.

La taille d'extraction par défaut de Read from SAP est 10 000.

Le stage Read from SAP est désormais configuré pour lire les données de plusieurs tables de la base de données SAP dans le flux de données.

## Filtrage d'enregistrements dans Read from SAP

Les paramètres de filtre de Read from SAP vous permettent de lire un sous-ensemble d'enregistrements d'une table SAP plutôt que l'ensemble des enregistrements de la table. Pour filtrer les enregistrements, vous spécifiez les valeurs que doit contenir un enregistrement pour qu'il soit lu dans le flux de données. Si vous ne spécifiez aucune condition de filtre, l'ensemble des enregistrements de la table sont lus dans le flux de données. L'utilisation de conditions de filtre est facultative.

**Remarque :** Si le stage Read from SAP est configuré pour lire les données de plusieurs tables SAP, le filtre est appliqué après l'exécution de l'opération JOIN.

1. Dans le stage Read from SAP, cliquez sur l'onglet **Filtre**.
2. Cliquez sur **Ajouter**.
3. Dans le champ **Nom de table**, sélectionnez la table contenant les enregistrements que vous souhaitez filtrer.
4. Dans le champ **Filtrer par**, sélectionnez le champ contenant les données que vous souhaitez utiliser comme base de filtrage.
5. Sélectionnez l'un des opérateurs suivants :

**Remarque :** Les opérateurs disponibles varient suivant le type de données du champ utilisé pour le filtrage.

Opérateur	Description
Contient	Vérifie si la chaîne contient la valeur indiquée.
Est égal à	Vérifie si la valeur du champ correspond à la valeur indiquée.
N'est pas égal à	Vérifie si la valeur du champ ne correspond pas à la valeur indiquée.
Est supérieur à	Vérifie si le champ a une valeur numérique supérieure à la valeur indiquée. Cet opérateur fonctionne sur des types de données numériques ainsi que sur des champs de chaînes contenant des nombres.
Est supérieur ou égal à	Vérifie si le champ a une valeur numérique supérieure ou égale à la valeur indiquée. Cet opérateur fonctionne sur des types de données numériques ainsi que sur des champs de chaînes contenant des nombres.

Opérateur	Description
Inférieur à	Vérifie si le champ a une valeur numérique inférieure à la valeur indiquée. Cet opérateur fonctionne sur des types de données numériques ainsi que sur des champs de chaînes contenant des nombres.
Est inférieur ou égal à	Vérifie si le champ a une valeur numérique inférieure ou égale à la valeur indiquée. Cet opérateur fonctionne sur des types de données numériques ainsi que sur des champs de chaînes contenant des nombres.
Est nul	Vérifie si le champ est une valeur nulle.
N'est pas nul	Vérifie si le champ n'est pas une valeur nulle.
Commence par	Vérifie si le champ commence par la valeur indiquée.
Se termine par	Vérifie si le champ se termine par la valeur indiquée.

- Saisissez la valeur à laquelle vous souhaitez comparer la valeur du champ sélectionné.
- Cliquez sur **OK**.
- Ajoutez des conditions de filtre supplémentaires, le cas échéant.

**Remarque :** Si vous indiquez plusieurs conditions de filtre, l'ensemble des conditions de filtre doivent être définies sur true pour que l'enregistrement soit lu dans le flux de données. Si l'une des conditions n'est pas définie sur true, l'enregistrement n'est pas lu dans le flux de données.

## Read from Spreadsheet

Read from Spreadsheet ajoute des données d'une feuille de calcul Excel comme entrées d'un dataflow. Ce stage prend en charge les formats de fichier .xls et .xlsx.

### *Onglet de propriétés du fichier*

L'onglet **Propriétés du fichier** contient des options permettant de spécifier la feuille de calcul et les données à lire dans le dataflow.

Nom du champ	Description
Server name	Indique si le fichier que vous sélectionnez comme entrée figure sur l'ordinateur exécutant Enterprise Designer ou sur le serveur Spectrum™ Technology Platform. Si vous sélectionnez un fichier sur l'ordinateur local, le nom du serveur sera Mon ordinateur. Si vous sélectionnez un fichier sur le serveur, le nom du serveur sera Spectrum™ Technology Platform.
Nom de fichier	Indique le chemin du fichier. Cliquez sur le bouton de sélection (...) pour trouver le fichier souhaité.  <b>Remarque :</b> Si le Spectrum™ Technology Platform est en cours d'exécution sur les systèmes Unix ou Linux, rappelez-vous que les noms de fichiers et les chemins sur ces plates-formes sont sensibles à la casse.
Sélection de données	Indique comment vous souhaitez sélectionner les données de la feuille de calcul pour les lire dans le dataflow. L'un des éléments suivants :  <b>Données de la feuille</b> Sélectionnez cette option pour lire toutes les données d'une feuille de la feuille de calcul.  <b>Données de la plage</b> Sélectionnez cette option pour lire un sous-ensemble de données d'une feuille en indiquant une plage de cellules à lire.  <b>Plage nommée</b> Sélectionnez cette option pour lire un sous-ensemble de données d'une feuille en indiquant une plage nommée de la feuille de calcul.
Sélection de feuille	Si vous sélectionnez <b>Données de la feuille</b> ou <b>Données de la plage</b> dans le champ <b>Sélection de données</b> , utilisez cette option pour sélectionner la feuille dont vous souhaitez lire les données dans le dataflow.
Plage	Si vous sélectionnez <b>Données de la plage</b> dans le champ <b>Sélection de données</b> , utilisez cette option pour indiquer la cellule de début de la plage et la cellule de fin de la plage.
Plage nommée	Si vous sélectionnez <b>Plage nommée</b> dans le champ <b>Sélection de données</b> , utilisez cette option pour indiquer le nom de la plage que vous souhaitez lire dans le dataflow. Les plages sont définies dans la feuille de calcul. Si aucune plage n'apparaît, aucune plage n'est définie dans la feuille de calcul.

Nom du champ	Description
Ligne d'en-têtes	<p>Cochez cette case si vous souhaitez indiquer une ligne contenant des en-têtes de colonne. Les en-têtes de colonne deviennent les noms de champ du dataflow, même si vous pouvez modifier les noms de champ dans l'onglet <b>Champs</b>. Si vous ne cochez pas cette case, les champs du dataflow portent les noms génériques par défaut Column1, Column2, etc.</p> <p>La ligne d'en-tête que vous indiquez est fonction de la sélection de données. Par exemple, si vous sélectionnez <b>Données de la plage</b> dans le champ <b>Sélection de données</b> et que la plage commence à la cinquième ligne, et si vous indiquez 1 comme ligne d'en-tête, la cinquième ligne de la feuille de calcul est utilisée comme en-tête, car elle correspond à la première ligne de la plage.</p>
Décalage des données par rapport à l'en-tête	<p>Si vous indiquez une ligne d'en-tête, ce champ indique la première ligne contenant des données, par rapport à l'en-tête. Par exemple, si vous indiquez 1, la première ligne sous l'en-tête correspond à la première ligne de données à lire dans le dataflow. Si vous indiquez 2, la deuxième ligne sous l'en-tête correspond à la première ligne lue dans le dataflow, etc.</p>
Première ligne de données	<p>Si vous n'indiquez aucune ligne d'en-tête, ce champ indique la ligne de la sélection de données contenant la première ligne de données à lire dans le dataflow. La ligne que vous indiquez est fonction de la sélection de données. Par exemple, si vous sélectionnez <b>Données de la plage</b> dans le champ <b>Sélection de données</b> et que la plage commence à la cinquième ligne, et si vous indiquez 1 comme première ligne de données, la première ligne de données à lire dans le dataflow est la cinquième ligne.</p>
Ignorer les lignes vierges	<p>Sélectionnez cette option si vous souhaitez exclure les lignes vierges du dataflow. Si vous ne sélectionnez pas cette option, les lignes vierges de la feuille de calcul renvoient des enregistrements vierges dans le dataflow.</p> <p><b>Remarque :</b> Cette option n'affecte pas les données affichées dans l'aperçu. Les lignes vierges sont toujours affichées dans l'aperçu, même si cette option est sélectionnée.</p>

### Onglet Champs

L'onglet **Champs** contient des options de mappage des données de la feuille de calcul vers les champs du dataflow.

Option	Description
Régénérer	Cliquez sur ce bouton pour renseigner l'onglet Champs à l'aide des champs du fichier d'entrée défini dans l'onglet <b>Propriétés de fichier</b> .
Détecter le type	Cliquez sur ce bouton pour déterminer automatiquement le type de données pour l'ensemble des champs. Vous pouvez manuellement modifier le type de données d'un champ en sélectionnant le champ, puis en cliquant sur <b>Modifier</b> .
Modifier	Sélectionnez un champ, puis cliquez sur ce bouton pour modifier le nom de champ ou le type de données.

## Read from Variable Format File

Read from Variable Format File lit des données à partir d'un fichier contenant des enregistrements dont la mise en page varie. Chaque enregistrement est lu en tant que champ de liste. Vous pouvez spécifier la balise qui indique le type d'enregistrement parent et tous les autres types d'enregistrement deviendront des champs de liste sous le parent.

Les fichiers de format variable présentent les caractéristiques suivantes :

- Les enregistrements dans le fichier contiennent des champs différents et des nombres de champs différents.
- Chaque enregistrement doit contenir une balise (généralement un nombre) identifiant le type d'enregistrement.
- Les relations hiérarchiques sont prises en charge.

### Exemple de fichier au format variable

Cet exemple présente un fichier de format variable contenant des informations sur la vérification de l'activité de compte de deux clients, Joe Smith et Anne Johnson. Dans cet exemple, le fichier est un fichier délimité qui utilise la virgule comme délimiteur de champ.

```
001   Joe,Smith,M,100 Main St,555-234-1290
100   CHK12904567,12/2/2007,6/1/2012,CHK
200   1000567,1/5/2012,Fashion Shoes,323.12
001   Anne,Johnson,F,1202 Lake St,555-222-4932
100   CHK238193875,1/21/2001,4/12/2012,CHK
```

```
200 1000232,3/5/2012,Blue Goose Grocery,132.11
200 1000232,3/8/2012,Trailway Bikes,540.00
```

Le premier champ dans chaque enregistrement contient la balise qui identifie le type d'enregistrement et, par conséquent, le format de l'enregistrement :

- 001 : enregistrement de client
- 100 : enregistrement de compte
- 200 : enregistrement de transaction de compte

Pour les fichiers délimités, il est courant que la valeur de balise (001, 100, 200) consiste en un nombre d'octets fixe au début de l'enregistrement, comme indiqué dans l'exemple ci-dessus.

Chaque enregistrement est doté de son propre format :

- 001 : Prénom,Nomdefamille,Sexe,Adresse,NuméroTéléphone
- 100 : IDdeCompte,DateD'ouverture,DateD'expiration,TypeDeCompte
- 200 : IDdeTransaction,DateDeTransaction,Fournisseur,Montant

Le format d'enregistrement 100 (enregistrement de compte) est un enfant de l'enregistrement 001 précédent et le format d'enregistrement 200 (enregistrement de transaction de compte) est un enfant de l'enregistrement 100 précédent (enregistrement de compte). Ainsi, dans le fichier d'exemple, le compte de Joe Smith, CHK12904567, présentait une transaction datant du 05/01/12 d'un montant de 323,12 chez Fashion Shoes. De même, le compte d'Anne Johnson, CHK238193875, présentait deux transactions, l'une du 05/03/12 chez Blue Goose Grocery et l'autre du 08/03/12 chez Trailway Bikes.

### Onglet de propriétés du fichier

Nom de l'option	Description
Server name	Indique si le fichier que vous sélectionnez comme entrée figure sur l'ordinateur exécutant Enterprise Designer ou sur le serveur Spectrum™ Technology Platform. Si vous sélectionnez un fichier sur l'ordinateur local, le nom du serveur sera Mon ordinateur. Si vous sélectionnez un fichier sur le serveur, le nom du serveur sera Spectrum™ Technology Platform.

Nom de l'option	Description
Nom de fichier	<p>Indique le chemin du fichier. Cliquez sur le bouton d'ellipse (...) pour accéder au fichier souhaité.</p> <p>Vous pouvez lire plusieurs fichiers en utilisant un caractère générique pour lire les données de plusieurs fichiers du répertoire. Les caractères génériques * et ? sont pris en charge. Par exemple, vous pouvez utiliser *.CSV pour lire tous les fichiers portant une extension .CSV placés dans le répertoire. Pour que plusieurs fichiers puissent être lus, tous doivent présenter la même mise en page (les mêmes champs aux mêmes emplacements). Tout enregistrement ne correspondant pas à la mise en page indiquée dans l'onglet <b>Champs</b> est traité comme un enregistrement non conforme.</p> <p>Lors de la lecture d'un fichier à partir d'un serveur de fichiers HDFS, les formats de compression pris en charge sont les suivants :</p> <ol style="list-style-type: none"> <li>1. GZIP (.gz)</li> <li>2. BZIP2 (.bz2)</li> </ol> <p><b>Remarque :</b> L'extension du fichier indique le format de compression à utiliser pour décompresser le fichier.</p> <p><b>Avertissement :</b> Si le Spectrum™ Technology Platform est en cours d'exécution sur les systèmes Unix ou Linux, rappelez-vous que les noms de fichiers et les chemins sur ces plates-formes sont sensibles à la casse.</p>
Type d'enregistrement	<p>Le format des enregistrements dans le fichier. Sélectionnez l'un des éléments suivants :</p> <p><b>Séquence de lignes</b> Un fichier texte dans lequel les enregistrements sont séparés par un caractère de fin de ligne (EOL) comme un retour chariot ou un retour à la ligne (CR ou LF) et chaque champ dispose d'une position de caractère fixe de départ et de fin.</p> <p><b>Largeur fixe</b> Un fichier texte dans lequel chaque enregistrement est un nombre déterminé de caractères de longueur, et chaque champ dispose d'une position de caractère de départ et de fin fixe .</p> <p><b>Délimité</b> Un fichier texte dans lequel les enregistrements sont séparés par un caractère de fin de ligne (EOL) comme un retour chariot ou un retour à la ligne (CR ou LF) et chaque champ est séparé par un caractère désigné comme une virgule.</p>

Nom de l'option	Description
Encodage des caractères	L'encodage du fichier texte. Sélectionnez l'un des éléments suivants : <ul style="list-style-type: none"> <li><b>UTF-8</b> Prend en charge tous les caractères Unicode et est rétrocompatible avec le format ASCII. Pour plus d'informations sur UTF, reportez-vous à <a href="http://unicode.org/faq/utf_bom.html">unicode.org/faq/utf_bom.html</a>.</li> <li><b>UTF-16</b> Prend en charge tous les caractères Unicode, mais n'est pas rétrocompatible avec le format ASCII. Pour plus d'informations sur UTF, reportez-vous à <a href="http://unicode.org/faq/utf_bom.html">unicode.org/faq/utf_bom.html</a>.</li> <li><b>US-ASCII</b> L'encodage des caractères en fonction de l'ordre alphabétique anglais.</li> <li><b>UTF-16BE</b> Encodage UTF-16 avec sérialisation d'octets gros-boutiste (octet le plus significatif en premier).</li> <li><b>UTF-16LE</b> Encodage UTF-16 avec sérialisation d'octets petit-boutiste (octet le moins significatif en premier).</li> <li><b>ISO-8859-1</b> Chiffrement de caractères ASCII utilisé habituellement pour les langues européennes occidentales. Également désigné par l'acronyme Latin-1.</li> <li><b>ISO-8859-3</b> Chiffrement de caractères ASCII utilisé habituellement pour les langues européennes méridionales. Également connu sous le nom Latin-3.</li> <li><b>ISO-8859-9</b> Chiffrement de caractères ASCII utilisé habituellement pour la langue turque. Également connu sous le nom Latin-5.</li> <li><b>CP850</b> Une page de code ASCII utilisée pour écrire les langues européennes occidentales.</li> <li><b>CP500</b> Une page de code EBCDIC utilisée pour écrire les langues européennes occidentales.</li> <li><b>Shift_JIS</b> Un encodage de caractères pour la langue japonaise.</li> <li><b>MS932</b> Extension Microsoft de Shift_JIS permettant d'inclure les caractères spéciaux NEC, la sélection NEC des extensions IBM et les extensions IBM.</li> <li><b>CP1047</b> Une page de code EBCDIC avec le jeu de caractères Latin-1 complet.</li> </ul>
Longueur d'enregistrement	Pour les fichiers à largeur fixe, précisez le nombre exact de caractères dans chaque enregistrement.

Nom de l'option	Description
Séparateur de champs	<p>Indique le caractère utilisé pour séparer des champs dans un fichier délimité.</p> <p>Par exemple, cet enregistrement utilise une barre verticale ( ) comme séparateur de champ :</p> <pre>7200 13TH ST   MIAMI   FL   33144</pre> <p>Ces caractères disponibles à définir en tant que séparateurs de champs sont :</p> <ul style="list-style-type: none"> <li>• Espace</li> <li>• Tabulation</li> <li>• Virgule</li> <li>• Point (.)</li> <li>• Point-virgule</li> <li>• Conduite</li> </ul> <p>Si le fichier utilise un caractère différent comme séparateur de champ, cliquez sur le bouton de sélection pour sélectionner un autre caractère délimiteur.</p>
Séparateur de balises	<p>Indique le caractère placé après le champ de balise pour démarquer le champ d'identification pour chaque enregistrement dans un fichier délimité. Un séparateur de balises doit être un caractère unique.</p> <p>Par défaut, les caractères suivants sont disponibles comme séparateurs de balises :</p> <ul style="list-style-type: none"> <li>• Espace</li> <li>• Tabulation</li> <li>• Virgule</li> <li>• Point (.)</li> <li>• Point-virgule</li> <li>• Conduite</li> </ul> <p>Si le fichier utilise un caractère différent comme séparateur de balises, cliquez sur le bouton d'ellipse pour ajouter et sélectionner un séparateur de balises personnalisé.</p> <p><b>Remarque :</b> Par défaut, le caractère <b>Séparateur d'enregistrements</b> est le même que le caractère <b>Séparateur de champs</b> sélectionné. Pour activer ce champ et sélectionner un caractère différent, décochez la case <b>Identique au séparateur de champs</b>.</p>
Identique au séparateur de champs	<p>Indique si le séparateur de balises est le même que le séparateur de champs. Décochez cette case pour sélectionner un caractère différent comme séparateur de balises.</p> <p><b>Remarque :</b> Par défaut, cette case est cochée et le champ <b>Séparateur de balises</b> est désactivé.</p>

Nom de l'option	Description
Qualificateur de texte	<p>Le caractère utilisé pour entourer les valeurs de texte dans un fichier délimité. Par exemple, cet enregistrement utilise des guillemets doubles (") en tant que qualificateur de texte.</p> <pre>"7200 13TH ST" "MIAMI" "FL" "33144"</pre> <p>Ces caractères disponibles à définir en tant que qualificateurs de texte sont :</p> <ul style="list-style-type: none"> <li>• Guillemets simples (')</li> <li>• Guillemets doubles (")</li> </ul> <p>Si le fichier utilise un délimiteur de texte, cliquez sur le bouton de sélection pour sélectionner un autre caractère délimiteur.</p>
Séparateur d'enregistrements	<p>Indique le caractère utilisé pour séparer des enregistrements dans un fichier séquentiel ligne par ligne ou délimité. Ce champ n'est pas disponible si vous activez la case <b>Utiliser EOL par défaut</b>.</p> <p>Les paramètres de séparateur d'enregistrements disponibles sont :</p> <p><b>Unix (U+000A)</b> Un caractère de saut de ligne sépare les enregistrements. Il s'agit du séparateur d'enregistrement standard pour les systèmes Unix.</p> <p><b>Macintosh (U+000D)</b> Un caractère de retour chariot sépare les enregistrements. Il s'agit du séparateur d'enregistrement standard pour les systèmes Macintosh.</p> <p><b>Windows (U+000D U+000A)</b> Un retour chariot suivi d'un saut de ligne sépare les enregistrements. Il s'agit du séparateur d'enregistrement standard pour les systèmes Windows.</p> <p>Si votre fichier utilise un séparateur d'enregistrement, cliquez sur le bouton de sélection pour sélectionner un autre caractère séparateur.</p>
Nom de la balise racine	<p>Balise à utiliser pour les enregistrements qui représentent la valeur parent d'autres types d'enregistrement. Par exemple, si vous disposez de trois types d'enregistrement 001, 100 et 200, et que les types d'enregistrement 100 et 200 sont les enfants du type d'enregistrement 001, 001 est la balise racine.</p>
Utiliser des balises de largeur fixe	<p>Détermine si une quantité fixe d'espace doit être allouée au début de chaque enregistrement afin d'y placer la balise d'enregistrement. Par exemple, ce qui suit présente un fichier avec les balises 001, 100 et 200 dans un champ de largeur fixe :</p> <pre>001 Joe, Smith, M, 100 Main St, 555-234-1290 100 CHK12904567, 12/2/2007, 6/1/2012, CHK 200 1000567, 1/5/2012, Mike's Shoes, 323.12</pre>

Nom de l'option	Description
Position de départ de la balise	Si vous cochez la case <b>Utiliser des balises de largeur fixe</b> , cette option spécifie la position dans chaque enregistrement dans lequel la balise commence. Par exemple, si la balise commence au quatrième caractère de l'enregistrement, vous indiquerez 4.
Largeur de balise	Si vous cochez la case <b>Utiliser des balises de largeur fixe</b> , cette option indique le nombre d'espaces à allouer pour les balises commençant à la position indiquée dans le champ <b>Tag start position</b> . Par exemple, si vous indiquez 3 dans le champ <b>Tag start position</b> et 7 dans le champ <b>Tag width</b> , les positions 4 à 10 seront prises en considération dans la balise de l'enregistrement. La valeur que vous indiquez doit être suffisamment élevée pour inclure tous les caractères du nom de balise le plus long.  La valeur dans le champ <b>Tag width</b> est automatiquement augmentée si vous allongez le nom dans le champ <b>Root tag name</b> .  La largeur de balise maximale est de 1 024.
Utiliser EOL par défaut	Indique que le séparateur d'enregistrements du fichier est le caractère de fin de ligne (EOL) par défaut utilisé sur le système d'exploitation sur lequel le serveur Spectrum™ Technology Platform est exécuté.  Ne sélectionnez pas cette option si le fichier utilise un caractère EOL qui est différent du caractère EOL par défaut utilisé sur le système d'exploitation du serveur. Par exemple, si le fichier utilise un EOL Windows, mais que le serveur fonctionne sous Linux, ne cochez pas cette option. Au lieu de cela, sélectionnez l'option Windows dans le champ <b>Record separator</b> .
Traiter les enregistrements avec moins de champs que ce qui est défini comme incorrect	Si vous activez cette option, le format des enregistrements enfant contenant moins de champs qu'un enregistrement complet est considéré incorrect. Lorsqu'un enregistrement au format incorrect est détecté, le traitement passe à la balise racine suivante, en ne tenant pas compte de toutes les balises enfant intermédiaires. Une exception est écrite dans le journal contenant des informations sur les enregistrements enfant au format incorrect accompagnés d'un numéro de ligne.  <b>Remarque</b> : Le format des enregistrements est toujours considéré incorrect dans les situations suivantes, que vous activiez cette option ou non. <ul style="list-style-type: none"> <li>• La balise est inconnue</li> <li>• La ligne est vide</li> <li>• Une balise ne contient aucune donnée</li> <li>• Un enregistrement avec une balise qui est l'enfant d'une autre balise apparaît immédiatement après un enregistrement avec une balise racine</li> </ul>

### Onglet Champs

L'onglet **Champs** spécifie les caractéristiques de chaque champ lu à partir du fichier.

[Définition des champs dans des fichiers de format variable délimités](#) à la page 204

[Définition de champs dans un fichier de format variable à séquence de ligne ou à largeur fixe](#) à la page 208

### L'onglet Exécution

Nom du champ	Description
Nom de fichier	Affiche le nom de fichier sélectionné dans le premier onglet.
Enregistrement de départ	Si vous voulez ignorer les enregistrements au début du fichier lors de la lecture des enregistrements dans le dataflow, spécifiez le premier enregistrement que vous souhaitez lire. Par exemple, si vous voulez sauter les 50 premiers enregistrements dans un fichier, spécifiez 51. Le 51ème enregistrement sera le premier enregistrement lu dans le dataflow.
Tous les enregistrements	Sélectionnez cette option si vous souhaitez lire tous les enregistrements à partir de l'enregistrement spécifié dans le champ <b>Starting record</b> jusqu'à la fin du fichier.
Nombre max d'enregistrements	Sélectionnez cette option si vous souhaitez uniquement lire un certain nombre d'enregistrements à partir de l'enregistrement spécifié dans le champ <b>Starting record</b> . Par exemple, si vous voulez lire les 100 premiers enregistrements, sélectionnez cette option et entrez 100.

## Définition des champs dans des fichiers de format variable délimités

Cette procédure décrit la manière dont définir les champs dans le stage Read from Variable Format File pour les fichiers délimités.

1. Dans le stage Read from Variable Format File, cliquez sur l'onglet **Champs**.
2. Cliquez sur **Régénérer**.

Une liste de tous les champs pour chaque enregistrement apparaît. Pour chaque champ, les informations suivantes apparaissent :

**Parent** Balise du fichier d'entrée indiquant le type d'enregistrement dans lequel le champ apparaît. Si la balise commence par un nombre, la balise présente le préfixe « NumericTag\_ ». Par exemple, une balise nommée 100 deviendrait NumericTag\_100. Le préfixe est nécessaire car les noms de champ de dataflow ne peuvent pas commencer par un nombre.

**Champ** Le nom sera utilisé dans le dataflow pour le champ. Par défaut, les champs portent des noms au format <Tag Name>\_<Column n>. Par exemple, le premier champ de type d'enregistrement Owner serait Owner\_Column1, le deuxième serait Owner\_Column2, etc.

**Type** Type de données du champ.

**Remarque :** Les 50 premiers enregistrements permettent de générer la liste des champs. Le fichier d'entrée doit contenir au moins deux balises racine afin de générer une liste de champs.

3. Si vous souhaitez modifier les relations parent/enfant entre les balises :
  - a) Cliquez sur **Modifier la hiérarchie de balise**.
  - b) Cliquez et faites glisser les balises pour définir la hiérarchie de balise que vous souhaitez.
  - c) Cliquez sur **OK**.
4. Si vous souhaitez modifier le nom ou type de données d'un champ, sélectionnez le champ et cliquez sur **Modifier**.
5. Dans le champ **Name**, choisissez le champ à ajouter ou saisissez le nom du champ.

En général, il sera judicieux de remplacer les noms par défaut par des noms significatifs afin de représenter les données contenues dans le champ. Par exemple, envisagez les données d'entrée suivantes :

```
001 Joe,Smith,M,100 Main St,555-234-1290
```

Cet enregistrement possède une balise parent de 001 et présenterait ces champs créés par défaut :

```
NumericTag_001_Column1 : Joe
NumericTag_001_Column2 : Smith
NumericTag_001_Column3 : M
NumericTag_001_Column4 : 100 Main St
NumericTag_001_Column5 : 555-234-1290
```

Il serait judicieux de renommer les champs de telle sorte que les noms décrivent les données. Par exemple :

```
FirstName : Joe
LastName : Smith
Gender : M
AddressLine1 : 100 Main St
PhoneNumber : 555-234-1290
```

**Remarque :** Vous ne pouvez pas renommer les champs de liste. Les champs de liste contenant tous les champs d'un type d'enregistrement donné utilisent toujours le nom de balise du fichier d'entrée comme nom de champ.

6. Pour modifier le type de données d'un champ, sélectionnez le type de données de votre choix dans le champ **Type**.

Les types de données suivants sont disponibles :

**bigdecimal** Type de données numérique prenant en charge 38 points de précision décimaux. Utilisez ce type de données pour les données utilisées dans les calculs mathématiques requérant un haut niveau de précision, en particulier ceux impliquant des données financières. Le type de données bigdecimal prend en charge des calculs plus précis que le type de données double.

**booléen** Type logique doté de deux valeurs : True et False.

**bytearray** Tableau (liste) d'octets.

**Remarque :** ByteArray n'est pas pris en charge en entrée pour un service REST.

**date** Type de données contenant un mois, un jour et une année. Par exemple, 2012-01-30 ou January 30, 2012. Vous pouvez indiquer un format de date par défaut dans Management Console.

**datetime** Type de données contenant un mois, un jour, une année et des heures, des minutes et des secondes. Par exemple, 2012/01/30 6:15 PM.

**double** Un type de données numérique qui contient à la fois des nombres en double précision négatifs et positifs entre  $2^{-1074}$  et  $(2-2^{-52}) \times 2^{1023}$ . En notation E, la plage de valeurs est comprise entre -1.79769313486232E+308 et 1.79769313486232E+308.

**flottant** Un type de données numérique contenant des nombres de précision seuls négatifs et positifs entre  $2^{-149}$  et  $(2-2^{23}) \times 2^{127}$ . En notation E, la plage de valeurs est comprise entre -3.402823E+38 et 3.402823E+38.

**integer** Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre  $-2^{31}$  (-2,147,483,648) et  $2^{31}-1$  (2,147,483,647).

**liste** Au sens strict du terme, une liste n'est pas un type de données. Cependant, lorsqu'un champ contient des données hiérarchiques, il est traité comme un champ de « liste ». Dans Spectrum™ Technology Platform, une liste représente un ensemble de données composées de plusieurs valeurs. Par exemple, un champ Noms peut contenir une liste de valeurs de noms. Cela peut être représenté sous forme d'une structure XML, comme suit :

```
<Names>
  <Name>John Smith</Name>
  <Name>Ann Fowler</Name>
</Names>
```

Il est important de souligner que le type de données de liste Spectrum™ Technology Platform est différent du type de données de liste de schéma XML dans le sens où ce dernier est un type de données simple composé de plusieurs

valeurs, tandis que le type de données de liste Spectrum™ Technology Platform est semblable à un type de données complexe XML.

<b>long</b>	Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre $-2^{63}$ (-9,223,372,036,854,775,808) et $2^{63}-1$ (9,223,372,036,854,775,807).
<b>chaîne</b>	Séquence de caractères.
<b>time</b>	Type de données contenant l'heure du jour. Par exemple, 21:15:59 ou 9:15:59 PM.

7. Si vous avez sélectionné une date, une heure ou un type de données numérique, vous pouvez utiliser le format date/heure ou numérique par défaut, ou vous pouvez spécifier un autre format pour ce champ spécifique. Le format par défaut est le format par défaut du système qui a été défini dans les options de conversion du type dans Management Console, ou le format par défaut du flux de données spécifié dans les options de conversion du type dans Enterprise Designer. Le format qui s'applique apparaît. Pour utiliser le format par défaut, laissez l'option **Valeur par défaut** sélectionnée. Pour indiquer un autre format, choisissez **Personnalisé** et suivez les étapes suivantes :

**Remarque :** Il est important que vous choisissiez un format de date et heure qui reflète avec précision les données que vous lisez à partir du fichier. Par exemple, si le fichier contient des données de date au format Mois/Jour/Année, mais que vous choisissiez Jour/Mois/Année, tout calcul de date que vous effectuez dans le flux de données, tel que le tri par date, ne reflètera pas la date correcte. En outre, la conversion de type peut échouer pour les enregistrements, auquel cas le comportement d'échec spécifié dans les options de conversion de type dans Management Console ou Enterprise Designer prendra effet.

- a) Dans le champ **Locale**, sélectionnez le pays dont vous souhaitez utiliser la convention de formatage. Votre sélection déterminera les valeurs par défaut dans le champ **Format**. Pour les données de date, votre sélection déterminera également la langue utilisée lorsqu'un mois est épelé. Par exemple, si vous indiquez Anglais, le premier mois de l'année sera « January », alors que si vous indiquez Français, ce sera « Janvier ».
- b) Dans le champ **Format**, sélectionnez le format des données. Le format dépend du type de données du champ. Une liste des formats les plus couramment utilisés pour les paramètres régionaux sélectionnés est fournie.

Un exemple du format sélectionné s'affiche à droite du champ **Format**.

Vous pouvez également spécifier vos propres formats de date, heure et nombre si ceux disponibles ne répondent pas à vos besoins. Pour indiquer votre propre format de date ou heure, tapez le format dans le champ à l'aide de la notation décrite dans **Modèles de date et d'heure** à la page 306. Pour indiquer votre propre format numérique, tapez le format dans le fichier à l'aide de la notation décrite dans **Modèles numériques** à la page 309.

8. Cliquez sur **OK**.

## Définition de champs dans un fichier de format variable à séquence de ligne ou à largeur fixe

Cette procédure décrit la manière dont définir les champs dans le stage Read from Variable Format File pour des fichiers à séquence de ligne ou à largeur fixe.

1. Dans le stage Read from Variable Format File, cliquez sur l'onglet **Champs**.
2. Cliquez sur **Obtenir les balises**.

Une liste de tous les champs pour chaque enregistrement apparaît. Pour chaque champ, les informations suivantes apparaissent :

<b>Parent</b>	Balise du fichier d'entrée indiquant le type d'enregistrement dans lequel le champ apparaît. Si la balise commence par un nombre, la balise présente le préfixe « NumericTag_ ». Par exemple, une balise nommée 100 deviendrait NumericTag_100. Le préfixe est nécessaire car les noms de champ de dataflow ne peuvent pas commencer par un nombre.
<b>Champ</b>	Le nom sera utilisé dans le dataflow pour le champ. Par défaut, les champs portent des noms au format <Tag Name>_<Column n>. Par exemple, le premier champ de type d'enregistrement Owner serait Owner_Column1, le deuxième serait Owner_Column2, etc.
<b>Type</b>	Type de données du champ.

**Remarque :** Les 50 premiers enregistrements permettent de générer la liste des champs. Le fichier d'entrée doit contenir au moins deux balises racine afin de générer une liste de champs.

3. Dans le champ **Filter**, sélectionnez la balise pour le type d'enregistrement dont vous souhaitez définir les champs, puis cliquez sur **Ajouter**.

**Remarque :** Le filtre n'a aucun impact sur les champs qui sont lus dans le flux de données. Il ne filtre que la liste des champs pour faciliter la navigation.

4. Dans le champ **Name**, choisissez le champ à ajouter ou saisissez le nom du champ.
5. Dans le champ **Type**, vous pouvez laisser le type de données défini sur `string` si vous ne prévoyez pas d'effectuer des opérations mathématiques ou d'horodatage sur les données. Cependant, si vous avez l'intention d'effectuer ce genre d'opérations, sélectionnez un type de données approprié. Cela convertira les données de chaîne à partir du fichier en un type de données qui activera la manipulation appropriée des données dans le flux de données.

Spectrum™ Technology Platform prend en charge ces types de données :

<b>bigdecimal</b>	Type de données numérique prenant en charge 38 points de précision décimaux. Utilisez ce type de données pour les données utilisées dans les calculs mathématiques requérant un haut niveau de précision, en particulier ceux
-------------------	---

impliquant des données financières. Le type de données bigdecimal prend en charge des calculs plus précis que le type de données double.

**booléen** Type logique doté de deux valeurs : True et False.

**bytearray** Tableau (liste) d'octets.

**Remarque :** ByteArray n'est pas pris en charge en entrée pour un service REST.

**date** Type de données contenant un mois, un jour et une année. Par exemple, 2012-01-30 ou January 30, 2012. Vous pouvez indiquer un format de date par défaut dans Management Console.

**datetime** Type de données contenant un mois, un jour, une année et des heures, des minutes et des secondes. Par exemple, 2012/01/30 6:15 PM.

**double** Un type de données numérique qui contient à la fois des nombres en double précision négatifs et positifs entre  $2^{-1074}$  et  $(2-2^{-52}) \times 2^{1023}$ . En notation E, la plage de valeurs est comprise entre -1.79769313486232E+308 et 1.79769313486232E+308.

**flottant** Un type de données numérique contenant des nombres de précision seuls négatifs et positifs entre  $2^{-149}$  et  $(2-2^{23}) \times 2^{127}$ . En notation E, la plage de valeurs est comprise entre -3.402823E+38 et 3.402823E+38.

**integer** Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre  $-2^{31}$  (-2,147,483,648) et  $2^{31}-1$  (2,147,483,647).

**liste** Au sens strict du terme, une liste n'est pas un type de données. Cependant, lorsqu'un champ contient des données hiérarchiques, il est traité comme un champ de « liste ». Dans Spectrum™ Technology Platform, une liste représente un ensemble de données composées de plusieurs valeurs. Par exemple, un champ Noms peut contenir une liste de valeurs de noms. Cela peut être représenté sous forme d'une structure XML, comme suit :

```
<Names>
  <Name>John Smith</Name>
  <Name>Ann Fowler</Name>
</Names>
```

Il est important de souligner que le type de données de liste Spectrum™ Technology Platform est différent du type de données de liste de schéma XML dans le sens où ce dernier est un type de données simple composé de plusieurs valeurs, tandis que le type de données de liste Spectrum™ Technology Platform est semblable à un type de données complexe XML.

**long** Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre  $-2^{63}$  (-9,223,372,036,854,775,808) et  $2^{63}-1$  (9,223,372,036,854,775,807).

**chaîne** Séquence de caractères.

**time** Type de données contenant l'heure du jour. Par exemple, 21:15:59 ou 9:15:59 PM.

6. Si vous avez sélectionné une date, une heure ou un type de données numérique, vous pouvez utiliser le format date/heure ou numérique par défaut, ou vous pouvez spécifier un autre format pour ce champ spécifique. Le format par défaut est le format par défaut du système qui a été défini dans les options de conversion du type dans Management Console, ou le format par défaut du flux de données spécifié dans les options de conversion du type dans Enterprise Designer. Le format qui s'applique apparaît. Pour utiliser le format par défaut, laissez l'option **Valeur par défaut** sélectionnée. Pour indiquer un autre format, choisissez **Personnalisé** et suivez les étapes suivantes :

**Remarque** : Il est important que vous choisissiez un format de date et heure qui reflète avec précision les données que vous lisez à partir du fichier. Par exemple, si le fichier contient des données de date au format Mois/Jour/Année, mais que vous choisissiez Jour/Mois/Année, tout calcul de date que vous effectuez dans le flux de données, tel que le tri par date, ne reflètera pas la date correcte. En outre, la conversion de type peut échouer pour les enregistrements, auquel cas le comportement d'échec spécifié dans les options de conversion de type dans Management Console ou Enterprise Designer prendra effet.

- a) Dans le champ **Locale**, sélectionnez le pays dont vous souhaitez utiliser la convention de formatage. Votre sélection déterminera les valeurs par défaut dans le champ **Format**. Pour les données de date, votre sélection déterminera également la langue utilisée lorsqu'un mois est épelé. Par exemple, si vous indiquez Anglais, le premier mois de l'année sera « January », alors que si vous indiquez Français, ce sera « Janvier ».
- b) Dans le champ **Format**, sélectionnez le format des données. Le format dépend du type de données du champ. Une liste des formats les plus couramment utilisés pour les paramètres régionaux sélectionnés est fournie.

Un exemple du format sélectionné s'affiche à droite du champ **Format**.

Vous pouvez également spécifier vos propres formats de date, heure et nombre si ceux disponibles ne répondent pas à vos besoins. Pour indiquer votre propre format de date ou heure, tapez le format dans le champ à l'aide de la notation décrite dans [Modèles de date et d'heure](#) à la page 306. Pour indiquer votre propre format numérique, tapez le format dans le fichier à l'aide de la notation décrite dans [Modèles numériques](#) à la page 309.

7. Dans le champ **Start position**, entrez la position du premier caractère du champ et dans le champ **Length** entrez le nombre de caractères présents dans le champ.  
Par exemple, si le champ commence au dixième caractère de l'enregistrement et dispose de cinq caractères, vous spécifiez une position de début de 10 et une longueur de 5.
8. Cliquez sur **Ajouter**.
9. Répétez ce processus pour ajouter d'autres champs au type d'enregistrement ou cliquez sur **Fermer** si vous avez terminé d'ajouter des champs.

## Simplification de données de format variable

Les données de fichier de format variable contiennent souvent des enregistrements dotés d'une relation hiérarchique, dont l'un des types d'enregistrement est le parent des autres types d'enregistrement. Cependant, de nombreux stages requièrent des données au format plat, vous devrez donc convertir les données afin de les rendre utilisables par les stages en aval. Par exemple, envisagez les données d'entrée suivantes :

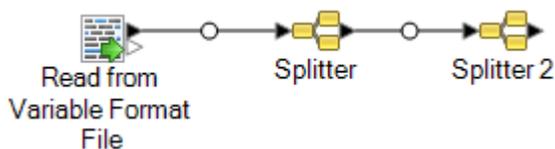
```
001 Joe,Smith,M,100 Main St,555-234-1290
100 CHK12904567,12/2/2007,6/1/2012,CHK
200 1000567,1/5/2012,Fashion Shoes,323.12
001 Anne,Johnson,F,1202 Lake St,555-222-4932
100 CHK238193875,1/21/2001,4/12/2012,CHK
200 1000232,3/5/2012,Blue Goose Grocery,132.11
200 1000232,3/8/2012,Trailway Bikes,540.00
```

Il serait judicieux de simplifier les enregistrements afin de disposer d'un enregistrement par transaction. Dans l'exemple précédent, cela revient à prendre les enregistrements de transaction (enregistrements avec la balise 200) et à les simplifier pour inclure les informations du titulaire de compte (enregistrements avec la balise 001) et les détails de compte (enregistrements avec la balise 100).

La procédure suivante indique comment utiliser les stages Splitter pour simplifier les enregistrements.

1. Ajoutez un stage Read from Variable Format File à votre flux de données et configurez le stage. Pour plus d'informations, reportez-vous à la section [Read from Variable Format File](#) à la page 197.
2. Ajoutez un stage Splitter et connectez-le à Read from Variable Format File.
3. Ajoutez d'autres stages Splitter en fonction de vos besoins, afin de disposer d'un stage Splitter pour chaque type d'enregistrement enfant dans vos données d'entrée.
4. Connectez tous les stages Splitter.

Votre flux de données doit désormais ressembler à ce qui suit :



5. Double-cliquez sur le premier stage Splitter pour ouvrir les options correspondantes.
6. Dans le champ **Split at**, sélectionnez un des types d'enregistrement enfant.
7. Cliquez sur **OK**.
8. Configurez chaque stage Splitter supplémentaire en sélectionnant un autre type d'enregistrement enfant dans le champ **Split at** de chaque stage Splitter.

## Read From XML

Le stage Read from XML lit un fichier XML dans un job ou un sous-flux. Il définit le chemin du fichier et le format de données, y compris un schéma XML et les détails des éléments de données.

Les éléments XML simples sont convertis en champs texte et transmis au stage suivant. Les données XML simples sont des enregistrements composés d'éléments XML qui contiennent uniquement des données et pas d'élément enfant. Voici un exemple de fichier de données XML simple :

```
<customers>
  <customer>
    <name>Sam</name>
    <gender>M</gender>
    <age>43</age>
    <country>United States</country>
  </customer>
  <customer>
    <name>Jeff</name>
    <gender>M</gender>
    <age>32</age>
    <country>Canada</country>
  </customer>
  <customer>
    <name>Mary</name>
    <gender>F</gender>
    <age>61</age>
    <country>Australia</country>
  </customer>
</customers>
```

Dans cet exemple, chaque enregistrement contient des éléments XML simples, tels que `<name>`, `<gender>`, `<age>` et `<country>`. Aucun des éléments ne contient d'éléments enfant.

Le stage Read from XML simplifie automatiquement les données simples de ce genre car la plupart des stages exigent que les données soient au format texte. Si vous souhaitez préserver la structure hiérarchique, utilisez un stage Aggregator après Read from XML pour convertir les données en données hiérarchiques.

Les éléments XML complexes restent au format hiérarchique et sont transmis en tant que champ de liste. Cependant, de nombreux stages requièrent des données au format plat, vous devrez donc simplifier des éléments XML complexes afin de les rendre utilisables par les stages en aval. Pour plus d'informations, reportez-vous à la section [Simplification des éléments XML complexes](#) à la page 217.

**Remarque :** Read From XML ne prend pas en charge les types XML `xs:anyType` ni `xs:anySimpleType`.

## Onglet de propriétés du fichier

Tableau 3 : Onglet de propriétés du fichier

Nom de l'option	Description
Fichier de Schéma	<p>Indique le chemin du fichier de schéma XSD. Cliquez sur le bouton de sélection (...) pour trouver le fichier souhaité. Notez que le fichier de schéma doit être sur le serveur pour que le fichier de données soit validé par rapport au schéma. Si le fichier de schéma n'est pas sur le serveur, la validation est désactivée.</p> <p>Sinon, vous pouvez spécifier un fichier XML au lieu d'un fichier XSD. Si vous spécifiez un fichier XML, le schéma est inféré en fonction de la structure du fichier XML. Utiliser un fichier XML au lieu d'un fichier XSD présente les limites suivantes :</p> <ul style="list-style-type: none"> <li>• Le fichier XML ne peut pas être supérieur à 1 Mo. Si la taille du fichier XML est supérieure à 1 Mo, essayez d'enlever quelques données en maintenant la structure du fichier XML.</li> <li>• Le fichier de données ne sera pas validé en fonction du schéma déduit.</li> </ul> <p><b>Remarque :</b> Si le Spectrum™ Technology Platform est en cours d'exécution sur les systèmes Unix ou Linux, rappelez-vous que les noms de fichiers et les chemins sur ces plates-formes sont sensibles à la casse.</p>
Fichier de données	<p>Indique le chemin du fichier de données XML. Cliquez sur le bouton de sélection (...) pour trouver le fichier souhaité.</p> <p><b>Remarque :</b> Si le Spectrum™ Technology Platform est en cours d'exécution sur les systèmes Unix ou Linux, rappelez-vous que les noms de fichiers et les chemins sur ces plates-formes sont sensibles à la casse.</p>
Aperçu	<p>Affiche un aperçu du fichier de schéma ou XML. Lorsque vous spécifiez un fichier XSD, la structure d'arborescence reflète le XSD sélectionné. Une fois que vous indiquez un fichier de schéma et un fichier de données, vous pouvez cliquer sur les éléments de schéma en gras pour voir un aperçu des données que l'élément contient.</p>

## Onglet Champs

**Tableau 4 : Onglet Champs**

Nom de l'option	Description
Filtrer	Filtre la liste des éléments et des attributs pour faciliter la navigation. Le filtre n'a aucun impact sur les champs qui sont inclus dans la sortie. Il filtre uniquement la liste des éléments et des attributs pour faciliter la navigation.
XPath	La colonne XPath affiche l'expression XPath pour l'élément ou l'attribut. Elle est affichée à titre indicatif. Pour plus d'informations sur XPath, consultez <a href="http://www.w3schools.com/xpath/">www.w3schools.com/xpath/</a> .
Champ	Le nom sera utilisé dans le flux de données pour l'élément ou l'attribut. Pour modifier le nom du champ, double-cliquez dessus et entrez le nom du champ souhaité.

Nom de l'option

Description

---

Type

## Nom de l'option

## Description

Le type de données à utiliser pour le champ.

**bigdecimal** Type de données numérique prenant en charge 38 points de précision décimaux. Utilisez ce type de données pour les données utilisées dans les calculs mathématiques requérant un haut niveau de précision, en particulier ceux impliquant des données financières. Le type de données bigdecimal prend en charge des calculs plus précis que le type de données double.

**booléen** Type logique doté de deux valeurs : True et False.

**date** Type de données contenant un mois, un jour et une année. Le format des dates doit être *aaaa-MM-jj*. Par exemple, 2012-01-30.

**datetime** Type de données contenant un mois, un jour, une année et des heures, des minutes et des secondes. Le format de datetime doit être *aaaa-MM-jj'T'HH:mm:ss*. Par exemple, 2012-01-30T06:15:30

**double** Un type de données numérique qui contient à la fois des nombres en double précision négatifs et positifs entre  $2^{-1074}$  et  $(2 \cdot 2^{-52}) \times 2^{1023}$ . En notation E, la plage de valeurs est comprise entre -1.79769313486232E+308 et 1.79769313486232E+308.

**flottant** Un type de données numérique contenant des nombres de précision seuls négatifs et positifs entre  $2^{-149}$  et  $(2 \cdot 2^{23}) \times 2^{127}$ . En notation E, la plage de valeurs est comprise entre -3.402823E+38 et 3.402823E+38.

**integer** Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre  $-2^{31}$  (-2,147,483,648) et  $2^{31}-1$  (2,147,483,647).

**liste** Au sens strict du terme, une liste n'est pas un type de données. Cependant, lorsqu'un champ contient des données hiérarchiques, il est traité comme un champ de « liste ». Dans Spectrum™ Technology Platform, une liste représente un ensemble de données composées de plusieurs valeurs. Par exemple, un champ Noms peut contenir une liste de valeurs de noms. Cela peut être représenté sous forme d'une structure XML, comme suit :

```
<Names>
  <Name>John Smith</Name>
  <Name>Ann Fowler</Name>
</Names>
```

Il est important de souligner que le type de données de liste Spectrum™ Technology Platform est différent du type de données de liste de schéma XML dans le sens où ce dernier est un type de données simple composé de plusieurs valeurs, tandis que le type de données de liste Spectrum™ Technology Platform est semblable à un type de données complexe XML.

**long** Un type de données numériques qui contient à la fois des nombres

Nom de l'option	Description
	entiers négatifs et positifs entre $-2^{63}$ (-9,223,372,036,854,775,808) et $2^{63}-1$ (9,223,372,036,854,775,807).
<b>chaîne</b>	Séquence de caractères.
<b>time</b>	Type de données contenant l'heure du jour. Le format de temps doit être <i>HH:mm:ss</i> . Par exemple, 21:15:59.
<b>Include</b>	Indique si ce champ doit être disponible dans le flux de données ou s'il faut l'exclure.

### Exemple : fichier XML simple

Dans cet exemple, vous voulez lire le fichier suivant dans un flux de données :

```
<addresses>
  <address>
    <addressline1>One Global View</addressline1>
    <city>Troy</city>
    <state>NY</state>
    <postalcode>12128</postalcode>
  </address>
  <address>
    <addressline1>1825B Kramer Lane</addressline1>
    <city>Austin</city>
    <state>TX</state>
    <postalcode>78758</postalcode>
  </address>
</addresses>
```

Dans cet exemple, vous pouvez choisir d'inclure `<addressline1>`, `<city>`, `<state>` et `<postalcode>`. Un enregistrement est alors créé pour chaque élément `<address>`, car `<address>` est l'élément parent commun de `<addressline1>`, `<city>`, `<state>` et `<postalcode>`.

## Simplification des éléments XML complexes

La plupart des stages dans un flux de données exigent que les données soient au format plat. Cela signifie que lorsque vous lisez des données hiérarchiques à partir d'un fichier XML dans un flux de données, vous devrez les aplatir si les données contiennent des éléments XML complexes. Un élément XML complexe est un élément qui contient d'autres éléments ou attributs. Par exemple,

dans le fichier de données suivant, les éléments `<address>` et `<account>` sont des éléments XML complexes :

```
<customers>
  <customer>
    <name>Sam</name>
    <gender>M</gender>
    <age>43</age>
    <country>United States</country>
    <address>
      <addressline1>1253 Summer St.</addressline1>
      <city>Boston</city>
      <stateprovince>MA</stateprovince>
      <postalcode>02110</postalcode>
    </address>
    <account>
      <type>Savings</type>
      <number>019922</number>
    </account>
  </customer>
  <customer>
    <name>Jeff</name>
    <gender>M</gender>
    <age>32</age>
    <country>Canada</country>
    <address>
      <addressline1>26 Wellington St.</addressline1>
      <city>Toronto</city>
      <stateprovince>ON</stateprovince>
      <postalcode>M5E 1S2</postalcode>
    </address>
    <account>
      <type>Checking</type>
      <number>238832</number>
    </account>
  </customer>
  <customer>
    <name>Mary</name>
    <gender>F</gender>
    <age>61</age>
    <country>Australia</country>
    <address>
      <addressline1>Level 7, 1 Elizabeth Plaza</addressline1>
      <city>North Sydney</city>
      <stateprovince>NSW</stateprovince>
      <postalcode>2060</postalcode>
    </address>
    <account>
      <type>Savings</type>
      <number>839938</number>
    </account>
  </customer>
</customers>
```

La procédure suivante indique comment utiliser les stages Splitter pour simplifier des données XML contenant plusieurs éléments XML complexes.

**Remarque :** Si vos données contiennent un seul élément XML complexe, vous pouvez utiliser un stage Splitter unique pour simplifier les données en associant simplement le stage Read from XML au stage Splitter. Vous ne devez pas utiliser les stages Broadcaster et Record Combiner, décrits dans cette procédure pour les fichiers de données contenant un élément XML complexe unique.

1. Ajoutez un stage Read from XML à votre flux de données et configurez le stage. Pour plus d'informations, reportez-vous à la section [Read From XML](#) à la page 212.
2. Ajoutez un stage Broadcaster et associez-lui le stage Read from XML.
3. Ajoutez un stage Splitter pour chaque élément XML complexe dans vos données.
4. Associez le stage Broadcaster à chaque stage Splitter.
5. Ajoutez un stage Record Combiner et associez-lui chaque stage Splitter.

Votre flux de données doit désormais ressembler à ce qui suit :



6. Double-cliquez sur le premier stage Splitter pour ouvrir les options correspondantes.
7. Dans le champ **Split at**, sélectionnez un des champs complexes. Dans le fichier de données d'exemple ci-dessus, il pourrait s'agir du champ d'adresse.
8. Cliquez sur **OK**.
9. Configurez chaque stage Splitter supplémentaire en sélectionnant un autre élément XML complexe dans le champ **Split at** de chaque stage Splitter.

Le flux de données est alors configuré afin de prendre l'entrée XML contenant les enregistrements dotés d'éléments XML complexes et de simplifier les données. Les enregistrements résultant du stage Record Combiner peuvent être envoyés à tout stage exigeant des données texte. Par exemple, vous pouvez joindre le stage Record Combiner à un stage Validate Address pour la validation d'adresse.

## Commande SQL

SQL Command exécute une ou plusieurs commandes SQL pour chaque enregistrement dans le dataflow. Vous pouvez utiliser SQL Command pour effectuer les tâches suivantes :

- Exécuter des instructions INSERT/UPDATE complexes, comme des instructions comportant des sous-requêtes/jointures avec d'autres tables
- Mettre des tables à jour après y avoir inséré/mis à jour des données pour conserver l'intégrité du référentiel

- Mettre à jour ou supprimer un enregistrement dans une base de données avant le chargement d'un enregistrement de remplacement
- Mettre plusieurs tables à jour en une seule transaction

Vous pouvez exécuter des commandes SQL supplémentaires avant et après l'exécution des commandes SQL principales, ainsi qu'appeler des procédures stockées.

**Remarque :** Les procédures stockées appelées à partir de SQL Command ne doivent pas utiliser de paramètres OUT.

**Remarque :** Une amélioration significative des performances peut être obtenue en utilisant plusieurs instances d'exécution de la commande SQL. Pour indiquer plusieurs instances d'exécution, cliquez sur le bouton **Exécution**.

### Général

L'onglet **Général** permet d'indiquer les instructions SQL dynamiques à exécuter une fois pour chaque enregistrement. La table suivante répertorie les options disponibles dans l'onglet **Général**.

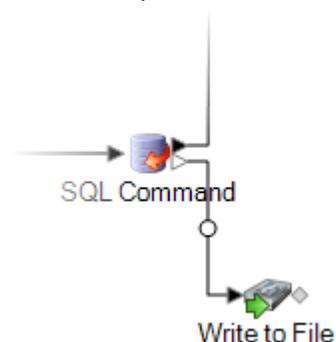
Option	Description
Connexion	<p>Sélectionnez la connexion de base de données à utiliser. Vos choix varient en fonction des connexions qui sont définies dans le Gestionnaire de connexion de Management Console. Si vous devez établir une nouvelle connexion à la base de données, ou en modifier ou supprimer une existante, cliquez sur <b>Gérer</b>.</p> <p>Si vous ajoutez ou modifiez une connexion de base de données, remplissez ces champs :</p> <p><b>Nom de la connexion</b> Saisissez le nom de la connexion.: Ce nom est entièrement de votre choix.</p> <p><b>Pilote de la base de données</b> Sélectionnez le type de base de données approprié.</p> <p><b>Options de connexion</b> Indiquez l'hôte, le port, l'instance, le nom d'utilisateur et le mot de passe à utiliser pour vous connecter à la base de données.</p>

Option	Description				
Instructions SQL	<p>Entrez les instructions SQL à exécuter pour chaque enregistrement dans le dataflow. Lorsque vous commencez à taper, une fenêtre contextuelle renseignée automatiquement affichera les commandes SQL valides. Séparez les instructions SQL par un point virgule (;).</p> <p>Pour spécifier une valeur d'un champ de dataflow, utilisez la syntaxe suivante :</p> <pre data-bbox="555 520 769 543">\${&lt;field name&gt;}</pre> <p>Où <code>&lt;field name&gt;</code> est le nom d'un champ du dataflow.</p> <p>Par exemple :</p> <pre data-bbox="555 653 1425 779">UPDATE MyDatabase.dbo.customer SET name=\${Name} WHERE id=\${ID};</pre> <p>Dans cet exemple, <code>\${Name}</code> est remplacé par la valeur du champ Nom du dataflow et <code>\${ID}</code> est remplacé par la valeur du champ ID du dataflow.</p> <p><b>Remarque :</b> Les requêtes doivent utiliser le nom entièrement qualifié. Par exemple, <code>MyDatabase.dbo.customer</code>.</p>				
Traitement de la transaction	<p>Indique si les enregistrements doivent être traités par lots ou tous à la fois. L'un des éléments suivants :</p> <table data-bbox="555 1098 1425 1234"> <tr> <td data-bbox="555 1098 721 1123"><b>Taille des lots</b></td> <td data-bbox="837 1098 1425 1157">Regroupe les enregistrements dans des lots de la taille que vous spécifiez et traite un lot à la fois.</td> </tr> <tr> <td data-bbox="555 1178 721 1234"><b>Ensemble de l'exécution</b></td> <td data-bbox="837 1178 1425 1234">Crée un grand lot pour tous les enregistrements et traite toutes les transactions simultanément.</td> </tr> </table>	<b>Taille des lots</b>	Regroupe les enregistrements dans des lots de la taille que vous spécifiez et traite un lot à la fois.	<b>Ensemble de l'exécution</b>	Crée un grand lot pour tous les enregistrements et traite toutes les transactions simultanément.
<b>Taille des lots</b>	Regroupe les enregistrements dans des lots de la taille que vous spécifiez et traite un lot à la fois.				
<b>Ensemble de l'exécution</b>	Crée un grand lot pour tous les enregistrements et traite toutes les transactions simultanément.				

Option	Description				
Traitement de l'erreur	Indique ce qu'il faut faire si une erreur est détectée lors de l'exécution des commandes SQL. L'un des éléments suivants : <table border="0" style="margin-left: 20px;"> <tr> <td style="vertical-align: top;"><b>Impossible d'arrêter le dataflow lors de l'erreur</b></td> <td style="vertical-align: top;">Le dataflow continue de s'exécuter si la base de données renvoie une erreur lors de l'exécution des commandes SQL.</td> </tr> <tr> <td style="vertical-align: top;"><b>Arrêter le dataflow après ce nombre d'erreurs</b></td> <td style="vertical-align: top;">Le dataflow arrêtera de s'exécuter après que la base de données a renvoyé le nombre d'erreurs indiqué.</td> </tr> </table>	<b>Impossible d'arrêter le dataflow lors de l'erreur</b>	Le dataflow continue de s'exécuter si la base de données renvoie une erreur lors de l'exécution des commandes SQL.	<b>Arrêter le dataflow après ce nombre d'erreurs</b>	Le dataflow arrêtera de s'exécuter après que la base de données a renvoyé le nombre d'erreurs indiqué.
<b>Impossible d'arrêter le dataflow lors de l'erreur</b>	Le dataflow continue de s'exécuter si la base de données renvoie une erreur lors de l'exécution des commandes SQL.				
<b>Arrêter le dataflow après ce nombre d'erreurs</b>	Le dataflow arrêtera de s'exécuter après que la base de données a renvoyé le nombre d'erreurs indiqué.				

**Remarque :** S'il existe une erreur de syntaxe dans le SQL, le dataflow se terminera toujours quel que soit le paramètre que vous choisissez ici.

En outre, vous pouvez éventuellement écrire des enregistrements d'erreur dans un récepteur en connectant le port d'erreur SQL Command au type de récepteur de votre choix. Le port d'erreur est le triangle blanc à droite de l'icône du stage dans le dataflow. Par exemple, pour écrire des enregistrements d'erreur dans un fichier plat, vous devez connecter le port d'erreur SQL Command à un stage Write to File, comme indiqué ici :



### SQL avant/après traitement

L'onglet **SQL avant/après traitement** permet d'indiquer les instructions SQL à exécuter une fois par exécution de dataflow, et non une fois par enregistrement comme c'est le cas avec le SQL que vous indiquez dans l'onglet **Général**. La table suivante répertorie les options disponibles dans l'onglet **SQL avant/après traitement**.

Option	Description
Pre-SQL	<p>Saisissez au moins une instruction SQL à exécuter avant de traiter les enregistrements entrant dans le stage. Les instructions SQL que vous saisissez ici sont exécutées une fois par exécution après que le dataflow démarre son exécution et avant que le stage SQL Command ne traite les premiers enregistrements.</p> <p>Un exemple d'utilisation du SQL avant traitement serait de créer une table pour les enregistrements qui seront traités.</p>
Autocommit pre-SQL	<p>Cochez cette case pour valider les instructions SQL avant traitement, avant d'exécuter les instructions SQL dans l'onglet <b>Général</b>.</p> <p>SI vous ne cochez pas cette case, les instructions SQL avant traitement seront validées dans la même transaction que les instructions SQL dans l'onglet <b>Général</b>.</p> <p><b>Remarque :</b> SI vous ne cochez ni la case <b>Validation automatique des instructions SQL avant traitement</b> ni <b>Validation automatique des instructions SQL après traitement</b>, toutes les instructions SQL du stage sont validées dans une transaction.</p>
Post-SQL	<p>Saisissez au moins une instruction SQL à exécuter une fois tous les enregistrements traités. Les instructions SQL que vous saisissez ici sont exécutées une fois par exécution après que le stage SQL Command est terminé mais avant que le dataflow ne se termine.</p> <p>Un exemple d'utilisation du SQL avant traitement serait de créer un index après le traitement des enregistrements.</p>
Autocommit post-SQL	<p>Cochez cette case pour valider les instructions SQL après traitement dans leur propre transaction après que les commandes SQL dans l'onglet <b>Général</b> sont validées.</p> <p>SI vous ne cochez pas cette case, les instructions SQL après traitement seront validées dans la même transaction que les instructions SQL dans l'onglet <b>Général</b>.</p> <p><b>Remarque :</b> SI vous ne cochez ni la case <b>Validation automatique des instructions SQL avant traitement</b> ni <b>Validation automatique des instructions SQL après traitement</b>, toutes les instructions SQL du stage sont validées dans une transaction.</p>

### L'onglet Exécution

L'onglet **Exécution** affiche **Options de stage** et vous offre la possibilité de définir les valeurs par défaut des options de stage.

Nom du champ	Description
Options de stage	<p>Cette section répertorie les options de dataflow utilisées dans la requête SQL de ce stage et vous permet de fournir une valeur par défaut pour toutes ces options. La colonne <b>Nom</b> répertorie les options, tandis que vous pouvez saisir les valeurs par défaut dans la colonne <b>Valeur</b> correspondante.</p> <p><b>Remarque</b> : La valeur par défaut fournie ici est également affichée dans la section <b>Map dataflow options to stages</b> de la boîte de dialogue <b>Options de dataflow</b>. La boîte de dialogue vous permet également de modifier la valeur par défaut. En cas de conflit des valeurs par défaut fournies pour une option via <b>Options de stage</b>, <b>Options de dataflow</b> et <b>Job Executor</b>, l'ordre de priorité est le suivant : valeur fournie via <b>Job Executor</b> &gt; valeur définie via la boîte de dialogue <b>Options de dataflow</b> &gt; valeur saisie via <b>Options de stage</b>.</p>

## Spécification d'un stage SQL Command lors de l'exécution

Cette procédure explique comment configurer un flux de données pour qu'il prenne en charge les options d'exécution d'un stage SQL Command ainsi que d'indiquer les arguments Job Executor pour le faire.

1. Ouvrez le flux de données dans Enterprise Designer.
2. Pour configurer les options d'exécution d'un stage dans un flux de données intégré, ouvrez ce dernier.
3. Cliquez sur l'icône Options de flux de données de la barre d'outils ou sur **Édition** > **Options de flux de données**. La boîte de dialogue **Options de flux de données** apparaît.
4. Cliquez sur **Ajouter**. La boîte de dialogue **Définir l'option de flux des données** apparaît.
5. Développez le stage SQL Command.
6. Sélectionnez une option SQL Command. Il peut s'agir de **PreSqlCommand**, de **SqlCommand** ou de **PostSqlCommand**.

**PreSqlCommand** Instructions SQL que vous souhaitez exécuter avant le traitement des enregistrements arrivant dans le stage. Ces instructions SQL sont exécutées une fois par exécution après le début de l'exécution du flux de données, mais avant le traitement du premier enregistrement par le stage SQL Command.

Un exemple d'utilisation du SQL avant traitement serait de créer une table pour les enregistrements qui seront traités.

**SqlCommand** Instructions SQL à exécuter pour chaque enregistrement du flux de données.

**PostSqlCommand** Instructions SQL à exécuter après le traitement de tous les enregistrements. Ces instructions SQL sont exécutées une fois par exécution après la fin du stage SQL Command, mais avant la fin du flux de données.

Par exemple, vous pouvez utiliser post-SQL pour créer un index après le traitement des enregistrements.

Le nom d'option SQL Command s'affiche dans les champs **Option name** et **Option label**. Il s'agit du nom de l'option qui devra être spécifié lors de l'exécution afin de définir cette option.

7. Saisissez une description de l'option dans le champ **Description**.
8. Dans le champ **Target**, sélectionnez l'option **Selected stage(s)**.
9. Si vous voulez limiter les valeurs qu'un utilisateur peut sélectionner, modifiez les options disponibles dans le champ **Legal values** en cliquant sur l'icône située directement à droite de ce champ.
10. Pour modifier la valeur par défaut, indiquez une valeur différente dans le champ **Valeur par défaut**.

**Remarque :** Pour un service, vous ne pouvez modifier que les valeurs par défaut avant d'exposer le service pour la première fois. Une fois que vous exposez le service, vous ne pouvez plus modifier les valeurs par défaut à l'aide d'Enterprise Designer. Au lieu de cela, vous devez utiliser Management Console. Pour plus d'informations, reportez-vous à la section [Spécification des Options de services par défaut](#).

11. Cliquez sur **OK**.
12. Continuez d'ajouter les options souhaitées.
13. Cliquez sur **OK** dans la boîte de dialogue des **Options du flux de données** lorsque vous avez terminé d'ajouter vos options.
14. Enregistrez et exposez le flux de données.
15. Créez un fichier texte contenant l'instruction SQL que vous souhaitez utiliser lors de l'exécution. Le fichier texte peut prendre la forme suivante :

```
SqlCommand = UPDATE CustomersSET
ContactName='Alfred Schmidt'
City='Hamburg'
WHERE CustomerName='Alfreds Futterkiste';
```

Dans cet exemple, SqlCommand est l'un des noms des options du stage SQL Command.

16. Utilisez l'argument **-o** lors de l'exécution d'un exécuteur de job depuis la ligne de commande.

```
java -jar jobexecutor.jar -h "noipa019sh-l1" -u "admin" -p "admin" -s
"8080" -o "options.txt" -j "FetchOracleData" -w
```

Le nom de fichier (options.txt) indique le nom du fichier texte que vous avez créé à l'étape 14. (Pour plus d'informations, reportez-vous à la section [Exécution d'un job à partir de la ligne de commande](#) à la page 226.)

## Exécution d'un job à partir de la ligne de commande

Avant de pouvoir exécuter un job à partir de la ligne de commande, il doit être exposé. Pour exposer un job, ouvrez-le dans Enterprise Designer et sélectionnez **Fichier > Exposer/Ne plus exposer et Enregistrer**.

Pour exécuter un job à partir de la ligne de commande, vous devez installer l'utilitaire Job Executor sur le système dans lequel exécuter le job. L'Exécuteur de job est disponible depuis la page d'accueil de Spectrum™ Technology Platform sur le serveur Spectrum™ Technology Platform (par exemple, <http://monserver:8080>).

### Utilisation

```
java -jar jobexecutor.jar -u UserID -p Password -j Job [Optional Arguments]
```

Obligatoire	Argument	Description
Non	-?	Imprime les informations d'utilisation.
Non	-d <i>delimiter</i>	Définit un délimiteur d'instance/état. Ceci apparaît uniquement dans les sorties synchrones.
Non	-e	Utilisez une connexion HTTPS sécurisée pour communiquer avec le serveur Spectrum™ Technology Platform.
Non	-f <i>property file</i>	Indique un chemin vers un fichier de propriétés d'un job. Un fichier de propriété de job contient des arguments de Job Executor. Pour plus d'informations sur les fichiers de propriétés d'un job, voir <a href="#">Utilisation d'un fichier de propriétés de job</a> .
Non	-h <i>host name</i>	Indique le nom ou l'adresse IP du serveur Spectrum™ Technology Platform.
Non	-i <i>poll interval</i>	Indique la fréquence de vérification des jobs achevés, en secondes. Ceci ne s'applique qu'en mode synchrone.
Oui	-j <i>job name</i>	Une liste de jobs séparés par des virgules à exécuter. Les noms de jobs sont sensibles à la casse. Les jobs sont lancés selon l'ordre répertorié.
Non	-n <i>email list</i>	Indique une liste d'adresses email supplémentaires séparées par des virgules pour les notifications de job configurées.

Obligatoire	Argument	Description
Non	<code>-o</code> <i>property file</i>	<p>Indique un chemin vers un fichier de propriétés des options de dataflow. Utilisez un fichier de propriétés d'options de dataflow pour définir des options pour les stages du dataflow. Pour définir des options de dataflow à l'aide d'un fichier de propriétés, vous devez configurer le dataflow pour qu'il expose les options de stage lors de l'exécution. Pour plus d'informations, reportez-vous à la section <a href="#">Ajout d'options d'exécution de flux de données</a>.</p> <p>Par exemple, un fichier de propriétés d'options de dataflow pour un dataflow contenant un stage Assign GeoTAX Info peut ressembler à ceci :</p> <pre>OutputCasing=U UseStreetLevelMatching=N TaxKey=T Database.GTX=gsl</pre>
Oui	<code>-p</code> <i>password</i>	Mot de passe de l'utilisateur.
Non	<code>-r</code>	<p>Spécifiez cet argument pour renvoyer un rapport détaillé sur le job. Cette option ne fonctionne que si vous spécifiez également <code>-w</code>. Le rapport contient les informations suivantes :</p> <ul style="list-style-type: none"> <li>• <b>Position 1</b>—Nom du job</li> <li>• <b>Position 2</b>—ID du processus du job</li> <li>• <b>Position 3</b>—État</li> <li>• <b>Position 4</b>—Date/Heure de début (JJ/MM/AAAAA HH:MM:SS)</li> <li>• <b>Position 5</b>—Date/Heure de fin (JJ/MM/AAAAA HH:MM:SS)</li> <li>• <b>Position 6</b>—Nombre d'enregistrements ayant réussi</li> <li>• <b>Position 7</b>—Nombre d'enregistrements ayant échoué</li> <li>• <b>Position 8</b>—Nombre d'enregistrements non conformes</li> <li>• <b>Position 9</b>—Actuellement non employé</li> </ul> <p>Par exemple :</p>

Obligatoire	Argument	Description
		<pre>MySimpleJob 4 succeeded 04/09/2010 14:50:47 04/09/2010 14:50:47 100 0 0 </pre> <p>Les informations sont délimitées à l'aide du délimiteur spécifié dans l'argument <code>-d</code>.</p>
Non	<code>-s port</code>	Le socket (port) sur lequel le serveur Spectrum™ Technology Platform s'exécute. La valeur par défaut est 8080.
Non	<code>-t timeout</code>	Configure le délai d'attente (en secondes) pour le mode synchrone. La valeur par défaut est 3 600. La valeur maximum est 2147483. Il s'agit d'un délai d'attente global agrégé qui représente la durée d'attente maximale que tous les jobs générés se terminent.
Oui	<code>-u user name</code>	Le nom de connexion de l'utilisateur.
Non	<code>-v</code>	Renvoie une sortie détaillée.
Non	<code>-w</code>	<p>Exécute Job Executor en mode synchrone. Cela signifie que l'exécution de Job Executor se poursuit jusqu'à ce que le job soit terminé.</p> <p>Si vous ne spécifiez pas <code>-w</code>, Job Executor s'arrête après le démarrage du job, sauf si le job lit ou écrit dans des fichiers sur le serveur. Dans ce cas, Job Executor s'exécutera jusqu'à ce que tous les fichiers locaux soient traités, puis il s'arrêtera.</p>
Non	<code>StageName=Protocol:FileName</code>	Remplace le fichier d'entrée ou de sortie indiqué dans Read from File ou Write to File. Pour plus d'informations, reportez-vous à la section <a href="#">Remplacement des emplacement des fichiers Job</a> .
Non	<code>StageName:schema=Protocol:SchemaFile</code>	Remplace la définition de disposition de fichier spécifiée dans Read from File ou Write to File par une définition définie dans un fichier de schéma. Pour plus d'informations, reportez-vous à la section <a href="#">Remplacement du format de fichier dans la ligne de commande</a> .

**Exemple d'utilisation de Job Executor**

L'exemple suivant affiche la ligne de commande et sa sortie :

```
D:\spectrum\job-executor>java -jar jobexecutor.jar -u user123
-p "mypassword" -j validateAddressJob1 -h spectrum.example.com
-s 8888 -w -d "%" -i 1 -t 9999

validateAddressJob1%105%succeeded
```

Dans cet exemple, la sortie indique que le job nommé « validateAddressJob1 » a été exécuté (avec l'identifiant 105) sans erreur. Le résultat aurait aussi pu être « failed » (échec) ou « running » (en cours d'exécution).

## Exécution de commandes SQL avant et après un flux de données

L'activité **Exécute SQL** effectue des opérations sur la base de données à tout moment au cours du processflow. Cette activité vous permet d'exécuter les instructions SQL avant et après l'exécution d'un flux de données Spectrum™ Technology Platform ou d'un programme externe. Par exemple, l'activité **Exécute SQL** peut être utilisée pour supprimer des index avant l'exécution d'un flux de données Spectrum™ Technology Platform et pour créer de nouveau des index après l'exécution du flux de données. Pour exécuter des instructions SQL à l'aide de l'activité **Exécute SQL**, vous devez créer un processflow.

**Remarque :** Pour obtenir des instructions sur la procédure de création et de planification d'un processflow, reportez-vous au *Guide Dataflow Designer*.

1. Glissez l'activité **Exécute SQL** jusqu'au canevas.
2. Double-cliquez sur l'activité **Exécute SQL**.
3. Sélectionnez la connexion à la base de données à utiliser.

Si vous devez établir une nouvelle connexion à la base de données, ou en modifier ou supprimer une existante, cliquez sur **Gérer**.

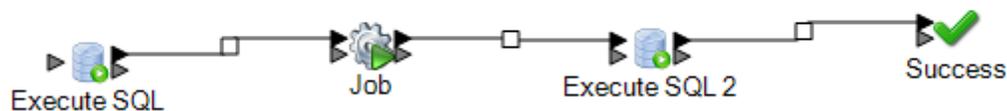
Si vous ajoutez ou modifiez une connexion de base de données, remplissez ces champs :

<b>Nom de la connexion</b>	Saisissez le nom de la connexion.: Ce nom est entièrement de votre choix.
<b>Pilote de la base de données</b>	Sélectionnez le type de base de données approprié.
<b>Options de connexion</b>	Indiquez l'hôte, le port, l'instance, le nom d'utilisateur et le mot de passe à utiliser pour vous connecter à la base de données.

4. Écrivez l'instruction SQL dans la zone **Instruction(s) SQL**.

Par défaut, l'option **Arrêter le processflow s'il y a une erreur** est cochée, ce qui signifie que le processflow est arrêté en cas d'exception. Si l'option **Arrêter le processflow s'il y a une erreur** n'est pas cochée et qu'une exception se produit, le processflow ne s'arrête pas et l'exception est consignée dans les journaux du serveur.

5. Ajoutez l'action que vous souhaitez que le processflow réalise.  
Vous pouvez ajouter un job en faisant glisser l'icône du job jusqu'au canevas, ou ajouter un programme externe en faisant glisser l'icône Exécuter le programme jusqu'au canevas.
6. Connectez les deux activités.
7. Ajoutez autant d'activités **Execute SQL** que nécessaire.  
Pour réaliser des actions sur **Execute SQL**, reportez-vous aux étapes 2 à 5.
8. Une fois tous les jobs et toutes les activités Exécuter le programme et Execute SQL ajoutés, pour les exécuter dans le processflow, glissez une activité Succès jusqu'au canevas et reliez-la à la dernière activité du



processflow

9. Exécutez le processflow.

## Transposer

Transposer convertit des colonnes en lignes. La transposition de données s'effectue à l'inverse du pivotement de données via le stage Group Statistics, qui transforme les données de ligne en colonnes.

Pour comprendre Transposer, prenons l'exemple suivant. Une table contient quatre trimestres de données de ventes et vous souhaitez ajouter l'ensemble du chiffre d'affaires généré et analyser la croissance sur les trois premiers trimestres. Pour ce faire, utilisez Transposer pour créer une colonne contenant l'ensemble du chiffre d'affaires des trois trimestres transposés. L'utilisation de Transposer pour ajouter l'ensemble du chiffre d'affaires généré figurant dans différentes colonnes à une seule colonne plutôt qu'à plusieurs peut améliorer les performances.

Le tableau suivant explique les options de la boîte de dialogue Transposer.

Option	Description
En-tête des champs transposés	Saisissez un nom d'en-tête pour la colonne devant contenir les colonnes à transposer. Cette nouvelle colonne est automatiquement ajoutée au flux de données.
En-tête des valeurs transposées	Saisissez un nom d'en-tête pour la colonne devant contenir les valeurs de colonne transposées. Cette nouvelle colonne est automatiquement ajoutée au flux de données.
Conserver les champs transposés	Cochez cette case pour conserver l'ensemble des champs transposés sous forme de colonnes dans la sortie.
Nom du champ	Affiche l'ensemble des en-têtes de colonne du fichier d'entrée.

Type

Affiche le type de données des champs respectifs (en-têtes de colonne).

Les colonnes à transposer doivent comporter un type de données compatible avec celui du fichier source d'entrée. Voici la matrice de compatibilité. Les grilles cochées correspondent aux types de données compatibles.

	Entier	Long	Chaîne	Date/Heure	Double	Big Decimal	Durée	Date
Entier	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
Long	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
Chaîne	<input checked="" type="checkbox"/>							
Datetime			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>				
Double	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>			
bigdecimal	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>		
Durée			<input checked="" type="checkbox"/>				<input checked="" type="checkbox"/>	
Date			<input checked="" type="checkbox"/>					<input checked="" type="checkbox"/>

Transposé

Cochez la case en regard de chaque champ que vous souhaitez convertir en colonne. Pour empêcher la transposition d'une colonne et la conserver dans la sortie, décochez la case.

**Exemple d'utilisation de Transposer**

Les données d'entrée suivantes contiennent quatre trimestres de ventes par magasin. Notez que T1, T2, T3 et T4 représentent quatre trimestres de ventes (en millions).

Magasin (États-Unis)	Q1	Q2	Q3	Q4
New York	100,00	200,10	300,00	400,00
Californie	250,10	450,00	550,00	650,00
Illinois	150,00	250,10	350,00	450,00

Les cas mentionnés ci-dessous illustrent le comportement de Transposer en fonction des options fournies dans le stage. Notez que Trimestre est le nom de colonne de l'en-tête des champs transposés et Chiffre d'affaires celui des valeurs des champs transposés.

**Cas 1**

Supposons que nous souhaitions transposer les colonnes T1, T2 et T3 et conserver T4 dans la sortie. Pour ce faire, cochez la case sous l'en-tête **Transposé** en regard de chaque colonne à transposer. Les colonnes T1, T2 et T3 sont maintenant affichées sous forme de lignes, tandis que la colonne T4 est conservée sous forme de colonne dans la sortie.

Magasin (États-Unis)	Trimestre	Chiffre d'affaires	Q4
New York	Q1	100,00	400,00
New York	Q2	200,10	400,00
New York	Q3	300,00	400,00
Californie	Q1	250,10	650,00
Californie	Q2	450,00	650,00
Californie	Q3	550,00	650,00

Magasin (États-Unis)	Trimestre	Chiffre d'affaires	Q4
Illinois	Q1	150,00	450,00
Illinois	Q2	250,10	450,00
Illinois	Q3	350,00	450,00

### Cas 2

Supposons que nous souhaitons transposer les colonnes T1 et T2 et conserver T3 et T4 dans la sortie. Nous souhaitons également conserver tous les champs transposés (T1 et T2) sous forme de colonnes dans la sortie. Pour ce faire, cochez la case **Conserver les champs transposés** et la case sous l'en-tête **Transposé** en regard de chaque colonne à transposer. Les colonnes T1 et T2 sont maintenant affichées sous forme de lignes, tandis que les colonnes T3 et T4 sont conservées sous forme de colonnes dans la sortie avec T1 et T2.

Magasin (États-Unis)	Trimestre	Chiffre d'affaires	Q1	Q2	Q3	Q4
New York	Q1	100,00	100,00	200,10	300,00	400,00
New York	Q2	200,10	100,00	200,10	300,00	400,00
Californie	Q1	250,10	250,10	450,00	550,00	650,00
Californie	Q2	450,00	250,10	450,00	550,00	650,00
Illinois	Q1	150,00	150,00	250,10	350,00	450,00
Illinois	Q2	250,10	150,00	250,10	350,00	450,00

## Unique ID Generator

Le stage Unique ID Generator crée une clé unique qui identifie un enregistrement spécifique. L'affectation d'un identifiant unique s'avère cruciale pour les initiatives d'entrepôt de données pour lesquelles les transactions ne peuvent pas porter toutes les données d'adresse et de nom, mais doivent néanmoins être attribuées au même enregistrement/contact. Un identifiant unique peut être appliqué au niveau de l'individu, du foyer, de la société et/ou de l'établissement. Unique ID Generator (générateur unique d'identifiant) propose plusieurs algorithmes pour créer des ID uniques.

L'identifiant unique s'appuie soit sur un numéro séquentiel ou un tampon Date/Heure. Par ailleurs, vous pouvez éventuellement utiliser une variété d'algorithmes pour générer les données à ajouter à l'ID, augmentant ainsi la possibilité pour l'ID d'être unique. Les identifiants de type numéro séquentiel ou tampon Date/Heure sont obligatoires et ne peuvent pas être supprimés de l'identifiant généré.

Unique ID Generator permet de générer une clé non unique à l'aide d'un des algorithmes de génération de clé. Dans le mode non unique, vous pouvez créer des clés à utiliser pour la correspondance. Cela peut être utile dans un entrepôt de données dans lequel vous avez déjà ajouté des clés à une dimension et si vous souhaitez générer une clé pour les nouveaux enregistrements, afin de voir si les nouveaux enregistrements correspondent à un enregistrement existant.

L'exemple suivant montre que chaque enregistrement de l'entrée reçoit un identifiant d'enregistrement séquentiel en sortie.

Enregistrement	RecordID
John Smith	0
Mary Smith	1
Jane Doe	2
John Doe	3

Le stage Unique ID produit un champ nommé RecordID qui contient l'ID unique. Vous pouvez renommer le champ RecordID, le cas échéant.

## Définition d'un identifiant unique

Par défaut, le stage Unique ID Generator crée un ID séquentiel, avec le premier enregistrement doté d'un ID de 0, le second enregistrement doté d'un ID de 1, le troisième enregistrement doté d'un ID de 2, etc. Pour modifier la manière dont l'ID unique est généré, suivez cette procédure.

1. Dans le stage Unique ID Generator, sous l'onglet **Règles**, cliquez sur **Modifier**.
2. Choisissez la méthode à utiliser pour générer l'ID unique.

Options	Description
<b>Le tag numérique séquentiel débute à</b>	Affecte une valeur numérique incrémentielle à chaque enregistrement en commençant par le nombre que vous indiquez. Si vous indiquez 0, le premier enregistrement aura un ID de 0, le deuxième aura un ID de 1, etc.

Options

Description

---

**Balise numérique séquentielle commençant à la valeur dans un champ de base de données**

## Options

## Description

Affecte une valeur numérique incrémentielle à chaque enregistrement en commençant par le nombre maximal de lectures à partir du champ de base de données. Ce nombre est ensuite incrémenté de 1 et affecté au premier enregistrement. Par exemple, si le nombre de lectures du champ de base de données est 30, le premier enregistrement aura un ID de 31, le deuxième enregistrement aura un ID de 32, etc.

**Connexion** Sélectionnez la connexion de base de données à utiliser. Vos choix varient en fonction des connexions qui sont définies dans le Gestionnaire de connexion de Management Console. Si vous devez établir une nouvelle base de données, ou modifier ou supprimer une connexion existante, cliquez sur **Gérer**.

Si vous ajoutez ou modifiez une connexion de base de données, remplissez ces champs :

**Nom de la connexion**

Saisissez le nom de la connexion.: Ce nom est entièrement de votre choix.

**Pilote de la base de données**

Sélectionnez le type de base de données approprié.

**Options de connexion**

Indiquez l'hôte, le port, l'instance, le nom d'utilisateur et le mot de passe à utiliser pour vous connecter à la base de données.

**Vue de la table** Indiquez la table ou la vue de la base de données que vous souhaitez interroger.

**Champ Database** Sélectionnez une colonne à partir de la liste pour générer une clé unique.

Les types de données pris en charge pour la génération d'ID unique sont les suivants :

**long** Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre  $-2^{63}$  (-9,223,372,036,854,775,808) et  $2^{63}-1$  (9,223,372,036,854,775,807).

**integer** Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre  $-2^{31}$  (-2,147,483,648) et  $2^{31}-1$  (2,147,483,647).

**bigdecimal** Type de données numérique prenant en charge 38 points de précision décimaux. Utilisez ce type de données pour les données utilisées dans les calculs mathématiques requérant un haut niveau de précision, en particulier ceux impliquant des données financières. Le type de données bigdecimal prend en charge des calculs plus précis que le type de données double.

**double** Un type de données numérique qui contient à la fois des nombres en double précision négatifs et positifs entre  $2^{-1074}$  et  $(2-2^{-52}) \times 2^{1023}$ . En

Options	Description
	notation E, la plage de valeurs est comprise entre -1.79769313486232E+308 et 1.79769313486232E+308.
<b>flottant</b>	Un type de données numérique contenant des nombres de précision seuls négatifs et positifs entre $2^{-149}$ et $(2-2^{-23}) \times 2^{127}$ . En notation E, la plage de valeurs est comprise entre -3.402823E+38 et 3.402823E+38.
<b>Horodatage</b>	Crée une clé unique en fonction de l'horodatage au lieu de la numérotation séquentielle.
<b>UUID</b>	Crée une clé d'identifiant de 32 chiffres universellement unique pour chaque enregistrement. Les chiffres de la clé sont affichés en cinq groupes séparés par des traits d'union, sous la forme 8-4-4-4-12, pour un total de 36 caractères (32 caractères alphanumériques et quatre traits d'union). Exemple : 123e4567-e89b-12d3-a456-432255330000
<b>Désactivé</b>	Sélectionnez cette option uniquement si vous souhaitez générer une clé non unique à l'aide d'un algorithme.

3. Cliquez sur **OK**.

## Utilisation d'algorithmes pour augmenter un ID unique

Unique ID Generator génère un ID unique pour chaque enregistrement en numérotant chaque enregistrement de manière séquentielle ou en générant un tampon date/heure pour chaque enregistrement. Vous pouvez éventuellement utiliser des algorithmes pour ajouter des informations supplémentaires à l'ID unique séquentiel ou d'horodatage, créant ainsi un ID unique plus complexe et qui est plus susceptible d'être réellement unique.

1. Dans le stage Unique ID Generator, cliquez sur **Ajouter**.
2. Dans le champ **Algorithm**, sélectionnez l'algorithme à utiliser pour générer d'autres informations dans l'ID. 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 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** 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.

**Substring** Renvoie une partie spécifiée du champ sélectionné.

3. Dans le champ **Field name**, choisissez le champ auquel appliquer l'algorithme. Par exemple, si vous choisissez l'algorithme soundex et un champ nommé City, l'ID sera généré en appliquant l'algorithme soundex aux données contenues dans le champ City.
4. Si vous avez sélectionné l'algorithme de sous-chaîne, indiquez la partie du champ à utiliser dans la sous-chaîne :
  - a) Dans le champ **Start position**, indiquez la position dans le champ où la sous-chaîne doit commencer.
  - b) Dans le champ **Length**, sélectionnez le nombre de caractères à partir de la position de départ que vous souhaitez inclure dans la sous-chaîne.

Par exemple, imaginons que vous disposez des données suivantes dans un champ nommé Nom de famille :

Augustine

Si vous avez indiqué 3 comme position de départ et 6 comme position de fin, la sous-chaîne produirait ce qui suit :

gustin

5. Cochez la case **Supprimer les caractères de bruit** pour supprimer tous les caractères non numériques et non alpha, tels que les traits d'union, les espaces blancs et autres caractères spéciaux du champ avant d'appliquer l'algorithme.
6. Pour les algorithmes de consonne et de sous-chaîne, vous pouvez trier les données dans le champ avant d'appliquer l'algorithme en cochant la case **Sort input**. Vous pouvez alors choisir de trier les caractères dans le champ ou les termes dans le champ par ordre alphabétique.
7. Cliquez sur **OK** pour enregistrer vos paramètres.
8. Répétez les étapes autant de fois que nécessaire, si vous souhaitez ajouter d'autres algorithmes pour produire un ID plus complexe.

**Remarque :** La définition de clé unique est toujours affichée d'une couleur différente et ne peut pas être supprimée.

## Définition d'un ID non unique

Unique ID Generator permet de générer une clé non unique à l'aide d'un des algorithmes de génération de clé. Dans le mode non unique, vous pouvez créer des clés à utiliser pour la correspondance. Cela peut être utile dans un entrepôt de données dans lequel vous avez déjà

ajouté des clés à une dimension et si vous souhaitez générer une clé pour les nouveaux enregistrements, afin de voir si les nouveaux enregistrements correspondent à un enregistrement existant.

1. Dans le stage Unique ID Generator, sous l'onglet **Règles**, cliquez sur **Modifier**.
2. Sélectionnez **Désactivé**.

Cela désactive la partie ID unique des règles de génération d'ID. Avec cette option désactivée, seul l'algorithme que vous choisissez dans les étapes suivantes sera utilisé pour créer l'ID. Cela signifie que tout enregistrement doté des mêmes données dans les champs que vous utilisez pour générer l'ID disposera du même ID. Vous pouvez ensuite utiliser l'ID pour la mise en correspondance.

3. Cliquez sur **OK**.
4. À l'invite d'avertissement, cliquez sur **Oui**.
5. Dans le stage Unique ID Generator, cliquez sur **Ajouter**.
6. Dans le champ **Algorithm**, sélectionnez l'algorithme à utiliser pour générer d'autres informations dans l'ID. 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.

**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.

**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** 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.

**Substring** Renvoie une partie spécifiée du champ sélectionné.

7. Dans le champ **Field name**, choisissez le champ auquel appliquer l'algorithme. Par exemple, si vous choisissez l'algorithme soundex et un champ nommé City, l'ID sera généré en appliquant l'algorithme soundex aux données contenues dans le champ City.
8. Si vous avez sélectionné l'algorithme de sous-chaîne, indiquez la partie du champ à utiliser dans la sous-chaîne :
  - a) Dans le champ **Start position**, indiquez la position dans le champ où la sous-chaîne doit commencer.
  - b) Dans le champ **Length**, sélectionnez le nombre de caractères à partir de la position de départ que vous souhaitez inclure dans la sous-chaîne.

Par exemple, imaginons que vous disposez des données suivantes dans un champ nommé Nom de famille :

Augustine

Si vous avez indiqué 3 comme position de départ et 6 comme position de fin, la sous-chaîne produirait ce qui suit :

gustin

9. Cochez la case **Supprimer les caractères de bruit** pour supprimer tous les caractères non numériques et non alpha, tels que les traits d'union, les espaces blancs et autres caractères spéciaux du champ avant d'appliquer l'algorithme.
10. Pour les algorithmes de consonne et de sous-chaîne, vous pouvez trier les données dans le champ avant d'appliquer l'algorithme en cochant la case **Sort input**. Vous pouvez alors choisir de trier les caractères dans le champ ou les termes dans le champ par ordre alphabétique.
11. Cliquez sur **OK** pour enregistrer vos paramètres.
12. Répétez les étapes autant de fois que nécessaire, si vous souhaitez ajouter d'autres algorithmes pour produire un ID plus complexe.

**Remarque :** La définition de clé unique est toujours affichée d'une couleur différente et ne peut pas être supprimée.

## Write To Cache

Write to Cache charge la sortie d'un flux de données dans un cache global, permettant ainsi le lancement d'une recherche sur les données via le stage Query Cache. L'utilisation d'un cache global pour les recherches sur les données améliore les performances par rapport aux recherches sur les bases de données.

Un cache global est un cache partagé à l'échelle du système qui réside en mémoire. Si vous souhaitez que le cache soit disponible pour différents flux de données, lorsque les données ne changent pas souvent ou restent relativement statiques ou encore lorsque l'espace de stockage n'est pas limité, optez pour un cache global. Un cache global est statique, car vous ne pouvez y écrire qu'une seule fois. Une fois créé, le cache ne peut pas être mis à jour.

**Remarque :** Write to Cache écrase le cache à chaque exécution du flux de données.

### Général

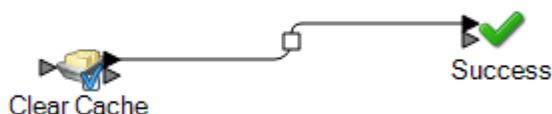
Nom de l'option	Description
Nom du cache	Indique le nom que vous souhaitez donner au cache. S'il existe déjà des caches sur le système, ils sont répertoriés dans une liste et vous pouvez en sélectionner un si vous souhaitez renseigner un cache existant à l'aide de nouvelles données. Pour créer un nouveau cache, saisissez le nom de votre choix pour ce cache. Le nom doit commencer par une lettre. Il peut contenir un trait de soulignement, mais aucun autre caractère spécial. Le nom peut contenir des valeurs numériques.

Nom de l'option	Description
Champs de cache	Cette colonne contient une liste des noms de champ à utiliser dans le cache. Pour modifier le nom d'un champ, cliquez dessus et saisissez un nouveau nom.
Champs de stage	Cette colonne contient une liste des noms de champ utilisés dans le flux de données. Vous ne pouvez pas modifier les noms de ces champs.
Type	Cette colonne contient une liste des types de données de chaque champ.
Include	Cochez la case de cette colonne pour que le champ soit écrit dans le cache. Décochez la case si vous ne souhaitez pas que le champ soit écrit dans le cache.
Champ Clé	<p>Cochez la case de cette colonne si vous souhaitez que le champ soit utilisé comme clé dans le stage Query Cache. Par exemple, si vous disposez d'un champ de flux de données nommé AccountNumber et si vous souhaitez que le stage Query Cache lance une recherche sur les données en demandant une valeur correspondante dans le champ AccountNumber, vous devez cocher la case de la colonne Champ clé du champ AccountNumber.</p> <p>Les champs que vous indiquez comme champs de clé peuvent être sélectionnés dans le stage Query Cache comme champs de clé.</p>

## Effacement d'un cache global

Pour effacer un cache global, vous devez créer et exécuter un processflow. Le processflow doit contenir une activité Effacer le cache. L'activité Effacer le cache permet d'effacer le cache global sans le supprimer. Vous pouvez également effacer le cache automatiquement en planifiant un processflow.

**Remarque :** Pour obtenir des instructions sur la procédure de création et de planification d'un processflow, reportez-vous au *Guide Dataflow Designer*.



Pour effacer manuellement les données du cache global, procédez comme suit :

1. Faites glisser l'activité **Clear cache** jusqu'au canevas.
2. Faites glisser l'activité **Success** jusqu'au canevas.
3. Connectez les deux activités.
4. Double-cliquez sur l'activité **Effacer le cache**.
5. Sélectionnez le cache. Vous pouvez également sélectionner plusieurs caches dont vous souhaitez effacer les données.

Les caches que vous créez dans le stage Write to Cache sont répertoriés dans l'activité Effacer le cache.

6. Exécutez le processflow.

## Write to DB

Le stage Write to DB écrit la sortie d'un flux de données vers une base de données.

**Remarque :** Une amélioration significative des performances peut être obtenue en utilisant de multiples instances d'exécution de Write to DB. Pour indiquer plusieurs instances d'exécution, cliquez sur le bouton **Exécution**.

### Onglet Général

Nom de l'option	Description
Connexion	Sélectionnez la connexion de la base de données que vous souhaitez utiliser dans le champ Connexion. Pour effectuer une nouvelle connexion de base de données, cliquez sur <b>Gérer</b> . Pour plus d'informations sur la création des connexions de bases de données, consultez le <a href="#">Gestionnaire de connexion à la base de données</a> .

Nom de l'option	Description
Table/Vue	<p>Après avoir sélectionné une connexion, spécifiez la vue ou la table pour l'écriture. Cliquez sur le bouton Parcourir ([...]) pour accéder à la table ou vue à utiliser, ou cliquez sur <b>Créer une table</b> pour créer une table dans la base de données.</p> <p><b>Remarque</b> : Si vous écrivez sur une base de données SQL, vous ne pouvez pas écrire dans les vues qui font référence à plus d'une table. Cela est dû à une limitation dans SQL Server.</p>
Créer une table	<p>Crée une table dans la base de données sélectionnée. Sélectionnez le propriétaire de la table dans le champ <b>Table owner</b> et indiquez le nom de la nouvelle table dans le champ <b>Table name</b>. Les noms de table sont sensibles à la casse. Indiquez une clé principale en cochant une case de la colonne <b>Clé principale</b>. Par ailleurs, indiquez les champs à écrire dans la nouvelle table en cochant la case correspondante dans la colonne <b>Inclure</b>. La colonne <b>Largeur</b> indique la largeur du champ pour le type de données chaîne. Par défaut, cette valeur est 512. Si la colonne <b>Autoriser null</b> est cochée et que <b>Champs d'entrée</b> contient une valeur null, le flux de données écrit la valeur null dans la base de données. Vous pouvez modifier le nom de la colonne en modifiant la valeur dans la colonne Champs de sortie.</p> <p>Le bouton <b>Créer une table</b> prend en charge la création de table dans les bases de données suivantes :</p> <ul style="list-style-type: none"> <li>• Axion</li> <li>• DB2</li> <li>• Derby/Cloudscape</li> <li>• Firebird</li> <li>• HSQLDB</li> <li>• Interbase</li> <li>• MaxDB/SapDB</li> <li>• McKoi</li> <li>• MySQL</li> <li>• Oracle</li> <li>• PostgreSQL</li> <li>• SQL Server</li> <li>• Sybase</li> </ul> <p><b>Remarque</b> : Pour les bases de données DB2, si vous essayez de créer une table et que la taille de page est inférieure à la longueur totale de toutes les colonnes de chaîne, vous obtiendrez une erreur indiquant « Échec de la création du corps à partir du contenu. La classe sérialisable n'est pas disponible pour le broker. »</p>
Champs de stage	<p>Dans la colonne Champs de stage, vous pouvez indiquer le champ à écrire dans le champ de base de données présenté dans la colonne Champ de base de données.</p>

Nom de l'option	Description
Include	La colonne Inclure vous permet de sélectionner les champs pour écriture. <b>Remarque</b> : Pour éviter des performances faibles, vous devez disposer d'un index ou d'une clé triée dans la table de base de données.

**Remarque** : Le stage **Write to DB** écrit toutes les valeurs du type de données `date` sous forme de valeurs `String`. Il s'agit du comportement du *pilote JTDS*, qui est le pilote par défaut utilisé par Spectrum. Pour gérer toutes les valeurs du type de données `date` en l'état, utilisez le pilote JDBC de Microsoft.

### Gestionnaire de connexion à la base de données

Le Gestionnaire de connexion à la base de données vous permet de gérer les connexions aux bases de données enregistrées. Pour ajouter, modifier, supprimer et tester des connexions, procédez comme suit :

1. Dans la boîte de dialogue **Options de Write To DB**, cliquez sur **Gérer**.
2. Cliquez sur **Ajouter**, **Modifier**, ou **Supprimer**.
3. Si vous ajoutez ou modifiez une connexion de base de données, remplissez ces champs :
  - Nom de la connexion : saisissez le nom de la nouvelle connexion.
  - Pilote de base de données : sélectionnez le type de base de données approprié.
  - Options de connexion : indiquez toutes les options, généralement, l'hôte, le port, l'instance, le nom utilisateur et le mot de passe.

**Remarque** : Vous pouvez tester la connexion en cliquant sur **Test**.

4. Si vous supprimez une connexion de base de données, sélectionnez la connexion qui doit être supprimée et cliquez sur **Supprimer**.

## L'onglet Exécution

Nom de l'option	Description
Mode Écriture	<p>Indique le type de mesures à prendre lors de l'écriture sur la base de données. L'un des éléments suivants :</p> <p><b>Insérer</b> Insérer les nouveaux enregistrements dans la base de données mais ne pas mettre à jour les enregistrements existants. Il s'agit du paramètre par défaut.</p> <p><b>Mettre à jour</b> Mettre à jour des enregistrements existants dans la base de données, mais ne pas insérer de nouveaux enregistrements</p> <p><b>Remarque :</b> Si vous sélectionnez <b>Mettre à jour</b>, le nom de la colonne de clé primaire utilisé dans la table d'entrée doit correspondre au nom de la colonne de clé primaire dans la table de sortie. Si vous essayez de mettre à jour une table où le nom de la colonne de clé primaire ne correspond pas à l'entrée, ou bien où la colonne de clé primaire n'est pas définie, la mise à jour ne fonctionnera pas.</p> <p><b>Insérer si impossible de mettre à jour</b> Insérer de nouveaux enregistrements dans la base de données si l'enregistrement n'existe pas, sinon mettre à jour l'enregistrement existant.</p>
Validation en lot	<p>Sélectionnez cette option pour valider les modifications sur la base de données après avoir indiqué un nombre d'enregistrements à traiter. Par défaut, cette option n'est pas sélectionnée, ce qui signifie que les modifications sont validées après le traitement de chaque enregistrement. Sélectionner cette option peut considérablement améliorer la performance du stage Write to DB.</p>

Nom de l'option	Description
Taille des lots	<p>L'activation de l'option <b>Validation en lot</b> spécifie le nombre d'enregistrements à valider sur la base de données dans chaque lot. La valeur par défaut est 1 000. Pour vos flux de données créés dans Spectrum™ Technology Platform 7.0 et versions antérieures, la valeur par défaut est 100.</p> <p>Une taille de lot plus grande n'offre pas toujours de meilleures performances de chargement. Tenez compte des facteurs suivants au moment de choisir une taille de lot :</p> <ul style="list-style-type: none"> <li>• <b>Taux d'arrivée des données au stage Write To DB</b> : si les données arrivent à un taux plus lent que ce que la base de données peut traiter, la modification de la taille de lot ne permettra pas d'améliorer les performances générales du flux de données. Par exemple, les flux de données avec la validation d'adresse ou le géocodage ne peuvent pas bénéficier d'un accroissement de taille.</li> <li>• <b>Trafic réseau</b> : pour les réseaux lents, l'augmentation de la taille des lots à une taille de lot moyenne (1 000 à 10 000) se traduit par de meilleures performances.</li> <li>• <b>Vitesse de chargement et/ou de traitement de la base de données</b> : pour les bases de données dotées d'une importante puissance de traitement, l'augmentation de la taille de lot permettra d'améliorer les performances.</li> <li>• <b>Instances d'exécution multiples</b> : si vous utilisez plusieurs instances d'exécution du stage Write to DB, une taille de lot volumineuse va consommer beaucoup de mémoire ; il est donc préférable d'utiliser une taille de lot petite ou moyenne (100 à 10 000).</li> <li>• <b>Restauration de base de données</b> : chaque fois qu'une instruction échoue, l'ensemble du lot est restauré. Plus la taille du lot est volumineuse, plus l'opération de restauration est longue.</li> </ul>
Valider à la fin	<p>Sélectionnez cette option pour être sûr que l'opération de validation de la base de données se produise après le transfert de tous les enregistrements à la base de données.</p>
Nombre de lots à valider	<p>Indiquez une valeur suite à laquelle les enregistrements doivent être validés. Les enregistrements sont validés dans la base de données après chaque nombre (Nombre de lots à valider * taille de lot) d'enregistrements transférés à la base de données. Par exemple, si <b>Taille de lot</b> est défini sur 1 000 et <b>Nombre de lots à valider</b> sur 3, la validation s'effectue à chaque transfert de 3 000 enregistrements vers la base de données.</p>
Tronquer la table avant d'insérer les données	<p>Sélectionnez cette option si vous voulez effacer toutes les données de la table avant d'écrire sur la base de données.</p>

Nom de l'option	Description
Ignorer la table si elle existe déjà et la recréer.	<p>Sélectionnez cette option pour supprimer et recréer la table avant d'écrire la sortie du flux de données dans la table. Cette option est utile si vous souhaitez que le schéma de la table corresponde aux champs du flux de données et ne contienne pas d'informations de schéma superflues.</p> <p>La table qui sera supprimée et recréée est celle indiquée dans le champ <b>Table/View</b> dans l'onglet <b>General</b>. Par exemple, si vous indiquez la table Clients dans le champ <b>Table/View</b> et que vous sélectionnez <b>Drop and recreate the table if it already exists</b>, la table Clients sera supprimée de la base de données et une nouvelle table nommée Clients sera créée avec un schéma qui correspond aux champs réels écrits dans la table.</p>

## Configuration de la gestion des erreurs dans Write to DB

Le stage Write to DB dispose d'un port d'erreur qui vous permet de filtrer les enregistrements qui sont à l'origine d'erreurs de base de données lors de l'écriture d'un enregistrement dans une base de données, tel qu'une violation de contrainte de clé primaire ou une violation de contrainte unique. Ces enregistrements peuvent ensuite être acheminés vers un autre chemin dans le flux de données, tandis que d'autres enregistrements sont validés. Par exemple, si vous traitez 100 enregistrements et que les enregistrements 4, 23 et 56 entraînent une erreur de base de données, ces trois enregistrements seraient routés via le port d'erreur, tandis que les 97 autres enregistrements seraient validés dans la base de données.

**Remarque :** L'utilisation du port d'erreur est facultative. Si vous n'utilisez pas le port d'erreur, le job échouera si un enregistrement entraîne une erreur.

1. Dans la palette, choisissez le stage du type qui doit gérer les enregistrements d'erreur (par exemple, Write to File) et faites-le glisser sur le canevas. Vous disposez de quelques options pour sélectionner un stage :
  - Pour écrire des enregistrements ayant échoué dans un fichier, faites glisser un des stages suivants sur le canevas : Write to File, Write to XML ou Write to Variable Format File.
  - Pour annuler simplement les enregistrements ayant échoué, faites glisser Write to Null sur le canevas.
2. Reliez le port d'erreur sur Write to DB au stage qui doit gérer les enregistrements ayant échoué. L'exemple suivant présente le port d'erreur sur Write to DB connecté au stage Write to File. Dans cet exemple, les enregistrements entraînant une erreur lorsqu'ils sont écrits dans la base de données sont écrits dans le fichier indiqué dans le stage Write to File.



Lorsque vous exécutez le flux de données, les enregistrements entraînant une erreur sont acheminés vers le port d'erreur. Les enregistrements du port d'erreur contiennent les champs indiqués dans Write to DB plus les champs suivants :

<b>Error.code</b>	Ce champ contient le code d'erreur numérique renvoyé depuis la base de données. Par exemple, dans le cas de l'erreur <code>ORA-00001: unique constraint ANKUSH.SYS_C0010018) violated</code> , la valeur du champ Error.code serait 1. Pour obtenir la liste des codes d'erreur, reportez-vous à la documentation du logiciel de base de données.
<b>Error.Message</b>	Ce champ contient le message d'erreur renvoyé depuis la base de données. Par exemple : <code>ORA-01034 ORACLE not available</code> . Dans ce cas, <code>ORACLE not available</code> serait la valeur du champ Error.Message. Pour obtenir la liste des messages d'erreur, reportez-vous à la documentation du logiciel de base de données.
<b>Error.SQLState</b>	Ce champ contient le code SQLSTATE qui fournit des informations détaillées sur la cause de l'erreur. Pour obtenir une liste des codes SQLSTATE, reportez-vous à la documentation du logiciel de base de données.
<b>Timestamp</b>	Date et heure sur le serveur Spectrum™ Technology Platform auxquelles l'erreur s'est produite.
<b>Nom d'utilisateur</b>	Nom de l'utilisateur Spectrum™ Technology Platform ayant exécuté le flux de données.

## Write to File

Sortie de dataflow Write to File dans un fichier texte. Les enregistrements contiennent tous les mêmes champs. Si vous souhaitez écrire des enregistrements de format variable, reportez-vous à [Write to Variable Format File](#) à la page 286. Si vous souhaitez écrire des enregistrements dans un fichier XML, reportez-vous à [Write to XML](#) à la page 296.

**Conseil :** Vous pouvez copier votre source et la coller comme récepteur dans votre dataflow pour configurer le fichier rapidement et utiliser les mêmes champs définis dans votre source.

## Onglet de propriétés du fichier

Nom du champ	Description
Server name	Indique si le fichier que vous sélectionnez comme entrée figure sur l'ordinateur exécutant Enterprise Designer ou sur le serveur Spectrum™ Technology Platform. Si vous sélectionnez un fichier sur l'ordinateur local, le nom du serveur sera Mon ordinateur. Si vous sélectionnez un fichier sur le serveur, le nom du serveur sera Spectrum™ Technology Platform.
Nom de fichier	<p>Indique le chemin du fichier. Cliquez sur le bouton de sélection (...) pour trouver le fichier souhaité.</p> <p>Lors de l'écriture d'un fichier à partir d'un serveur de fichiers HDFS, les formats de compression pris en charge sont les suivants :</p> <ol style="list-style-type: none"> <li>1. GZIP (.gz)</li> <li>2. BZIP2 (.bz2)</li> </ol> <p><b>Remarque :</b> Incluez l'extension appropriée dans le nom de fichier, pour indiquer le format de compression à utiliser pour écrire le fichier.</p> <p><b>Avertissement :</b> Si le Spectrum™ Technology Platform est en cours d'exécution sur les systèmes Unix ou Linux, rappelez-vous que les noms de fichiers et les chemins sur ces plates-formes sont sensibles à la casse.</p>
Type d'enregistrement	<p>Le format des enregistrements dans le fichier. Sélectionnez l'un des éléments suivants :</p> <p><b>Séquence de lignes</b> Un fichier texte dans lequel les enregistrements sont séparés par un caractère de fin de ligne (EOL) comme un retour chariot ou un retour à la ligne (CR ou LF) et chaque champ dispose d'une position de caractère fixe de départ et de fin.</p> <p><b>Largeur fixe</b> Un fichier texte dans lequel chaque enregistrement est un nombre déterminé de caractères de longueur, et chaque champ dispose d'une position de caractère de départ et de fin fixe .</p> <p><b>Délimité</b> Un fichier texte dans lequel les enregistrements sont séparés par un caractère de fin de ligne (EOL) comme un retour chariot ou un retour à la ligne (CR ou LF) et chaque champ est séparé par un caractère désigné comme une virgule.</p>

Nom du champ	Description
Encodage des caractères	L'encodage du fichier texte. Sélectionnez l'un des éléments suivants :
<b>UTF-8</b>	Prend en charge tous les caractères Unicode et est rétrocompatible avec le format ASCII. Pour plus d'informations sur UTF, reportez-vous à <a href="https://unicode.org/faq/utf_bom.html">unicode.org/faq/utf_bom.html</a> .
<b>UTF-16</b>	Prend en charge tous les caractères Unicode, mais n'est pas rétrocompatible avec le format ASCII. Pour plus d'informations sur UTF, reportez-vous à <a href="https://unicode.org/faq/utf_bom.html">unicode.org/faq/utf_bom.html</a> .
<b>US-ASCII</b>	L'encodage des caractères en fonction de l'ordre alphabétique anglais.
<b>UTF-16BE</b>	Encodage UTF-16 avec sérialisation d'octets gros-boutiste (octet le plus significatif en premier).
<b>UTF-16LE</b>	Encodage UTF-16 avec sérialisation d'octets petit-boutiste (octet le moins significatif en premier).
<b>ISO-8859-1</b>	Chiffrement de caractères ASCII utilisé habituellement pour les langues européennes occidentales. Également désigné par l'acronyme Latin-1.
<b>ISO-8859-3</b>	Chiffrement de caractères ASCII utilisé habituellement pour les langues européennes méridionales. Également connu sous le nom Latin-3.
<b>ISO-8859-9</b>	Chiffrement de caractères ASCII utilisé habituellement pour la langue turque. Également connu sous le nom Latin-5.
<b>CP850</b>	Une page de code ASCII utilisée pour écrire les langues européennes occidentales.
<b>CP500</b>	Une page de code EBCDIC utilisée pour écrire les langues européennes occidentales.
<b>Shift_JIS</b>	Un encodage de caractères pour la langue japonaise.
<b>MS932</b>	Extension Microsoft de Shift_JIS permettant d'inclure les caractères spéciaux NEC, la sélection NEC des extensions IBM et les extensions IBM.
<b>CP1047</b>	Une page de code EBCDIC avec le jeu de caractères Latin-1 complet.

Nom du champ	Description
Séparateur de champs	<p>Indique le caractère utilisé pour séparer des champs dans un fichier délimité.</p> <p>Par exemple, cet enregistrement utilise une barre verticale ( ) comme séparateur de champ :</p> <pre>7200 13TH ST MIAMI FL 33144</pre> <p>Ces caractères disponibles à définir en tant que séparateurs de champs sont :</p> <ul style="list-style-type: none"><li>• Espace</li><li>• Tabulation</li><li>• Virgule</li><li>• Point (.)</li><li>• Point-virgule</li><li>• Conduite</li></ul> <p>Si le fichier utilise un caractère différent comme séparateur de champ, cliquez sur le bouton de sélection pour sélectionner un autre caractère délimiteur.</p>
Qualificateur de texte	<p>Le caractère utilisé pour entourer les valeurs de texte dans un fichier délimité.</p> <p>Par exemple, cet enregistrement utilise des guillemets doubles (") en tant que qualificateur de texte.</p> <pre>"7200 13TH ST" "MIAMI" "FL" "33144"</pre> <p>Ces caractères disponibles à définir en tant que qualificateurs de texte sont :</p> <ul style="list-style-type: none"><li>• Guillemets simples (')</li><li>• Guillemets doubles (")</li></ul> <p>Si le fichier utilise un délimiteur de texte, cliquez sur le bouton de sélection pour sélectionner un autre caractère délimiteur.</p>

Nom du champ	Description
Séparateur d'enregistrements	<p>Indique le caractère utilisé pour séparer des enregistrements dans un fichier séquentiel ligne par ligne ou délimité. Ce champ n'est pas disponible si vous activez la case <b>Utiliser EOL par défaut</b>.</p> <p>Les paramètres de séparateur d'enregistrements disponibles sont :</p> <p><b>Unix (U+000A)</b> Un caractère de saut de ligne sépare les enregistrements. Il s'agit du séparateur d'enregistrement standard pour les systèmes Unix.</p> <p><b>Macintosh (U+000D)</b> Un caractère de retour chariot sépare les enregistrements. Il s'agit du séparateur d'enregistrement standard pour les systèmes Macintosh.</p> <p><b>Windows (U+000D U+000A)</b> Un retour chariot suivi d'un saut de ligne sépare les enregistrements. Il s'agit du séparateur d'enregistrement standard pour les systèmes Windows.</p> <p>Si votre fichier utilise un séparateur d'enregistrement, cliquez sur le bouton de sélection pour sélectionner un autre caractère séparateur.</p>
Utiliser EOL par défaut	<p>Indique que le séparateur d'enregistrements du fichier est le caractère de fin de ligne (EOL) par défaut utilisé sur le système d'exploitation sur lequel le serveur Spectrum™ Technology Platform est exécuté.</p> <p>Ne sélectionnez pas cette option si le fichier utilise un caractère EOL qui est différent du caractère EOL par défaut utilisé sur le système d'exploitation du serveur. Par exemple, si le fichier utilise un EOL Windows, mais que le serveur fonctionne sous Linux, ne cochez pas cette option. Au lieu de cela, sélectionnez l'option Windows dans le champ <b>Record separator</b>.</p>
Longueur d'enregistrement	<p>Pour les fichiers à largeur fixe, précisez le nombre exact de caractères dans chaque enregistrement.</p> <p>Pour les fichiers de séquence de lignes, indiquez la longueur, en caractères, de l'enregistrement le plus long dans le fichier.</p>
La première ligne est l'enregistrement d'en-tête.	<p>Indique si le premier enregistrement dans un fichier délimité contient des informations d'en-tête, mais aucune donnée.</p> <p>Par exemple, cet extrait de fichier illustre une ligne d'en-tête dans le premier enregistrement.</p> <pre>"AddressLine1" "City" "StateProvince" "PostalCode" "7200 13TH ST" "MIAMI" "FL" "33144" "One Global View" "Troy" "NY" 12180</pre>

Nom du champ	Description
Traiter les enregistrements avec moins de champs que ce qui est défini comme incorrect	Les enregistrements de fichiers délimités contenant moins de champs que le nombre défini dans l'onglet <b>Champs</b> seront traités comme des enregistrements incorrects.
Importer	Importe la définition du dessin d'enregistrement du fichier, le paramètre d'encodage et les options de tri à partir d'un fichier de paramètres. Le fichier de paramètres est créé en exportant les paramètres à partir d'un autre stage Read from File ou Write to File qui utilisait le même fichier d'entrée ou un fichier qui dispose du même dessin d'enregistrement que le fichier avec lequel vous joblez.
Exporter	<p>Enregistre la définition du dessin d'enregistrement du fichier, le paramètre d'encodage et les options de tri dans un fichier de paramètres. Vous pouvez ensuite importer ces paramètres dans d'autres stages Read from File ou Write to File qui utilisent le même fichier d'entrée ou un fichier semble à celui que vous utilisez actuellement. Vous pouvez utiliser le fichier de paramètres avec Job Executor pour indiquer les paramètres de fichier lors de l'exécution.</p> <p>Pour plus d'informations sur le fichier de paramètres, voir <a href="#">Fichier de paramètres de définition du fichier</a> à la page 161.</p>

### Onglet Champs

L'onglet Champs définit les noms, les positions, et, pour les fichiers à longueur fixe et à séquence de lignes, les longueurs de champs dans le fichier. Pour plus d'informations, reportez-vous aux rubriques suivantes :

[Définition de champs dans un fichier de sortie délimité](#) à la page 257

[Définition de champs dans un fichier à séquence de ligne ou à longueur fixe](#) à la page 260

### Onglet Trier les champs

L'onglet Trier les champs définit les champs permettant de trier les enregistrements de sortie avant leur écriture dans le fichier de sortie. Le tri est facultatif. Pour plus d'informations, reportez-vous à la section [Tri des enregistrements de sortie](#) à la page 263.

### L'onglet Exécution

Nom de l'option	Description
Nom de fichier	Cela affiche le fichier défini dans l'onglet <b>Propriétés de fichier</b> .

Nom de l'option	Description
Générer plusieurs fichiers	<p>Sélectionnez cette option pour écrire des enregistrements dans des fichiers différents, plutôt que d'écrire tous les enregistrements en un fichier. Le fichier dans lequel chaque enregistrement est écrit est indiqué dans l'enregistrement lui-même. Chaque enregistrement doit contenir un champ indiquant soit un nom de fichier, soit le chemin d'accès complet au fichier dans lequel vous souhaitez écrire l'enregistrement. Par exemple, pour envoyer les prix des actions de différentes sociétés (de groupes divers) à tous les clients séparément, cette fonctionnalité écrit les prix des actions des différentes sociétés dans des fichiers distincts qui peuvent être envoyés à chacun des clients, si vous le désirez. Si vous activez l'option <b>Générer plusieurs fichiers</b>, vous devez spécifier un fichier de sortie sur le serveur Spectrum™ Technology Platform ou sur un serveur FTP. Pour écrire les données dans un fichier sur un serveur FTP, vous devez définir une connexion au serveur de fichiers via Management Console.</p> <p><b>Remarque :</b> Les enregistrements de la colonne que vous sélectionnez dans <b>Champ de chemin d'accès au fichier</b> doivent se présenter dans l'ordre trié. Utilisez cette fonctionnalité lorsque l'enregistrement contient un nom de fichier ou le fichier de fichier complet.</p>
Champ de chemin d'accès au fichier	<p>Sélectionne le champ contenant le chemin (un nom de fichier ou le chemin de fichier complet) du fichier dans lequel vous souhaitez écrire l'enregistrement. Ce champ n'est activé que si vous sélectionnez <b>Générer plusieurs fichiers</b>.</p>
Mode Écriture	<p>Indique s'il faut ajouter la sortie du dataflow à la fin du fichier ou supprimer les données existantes dans le fichier avant écriture de la sortie. L'un des éléments suivants :</p> <p><b>Écraser</b> Remplace les données existantes dans le fichier de sortie chaque fois que le dataflow est exécuté.</p> <p><b>Ajouter</b> Ajoute le dataflow de sortie à la fin du fichier, sans effacer les données existantes du fichier</p>

## Définition de champs dans un fichier de sortie délimité

Dans le stage Write to File, l'onglet **Champs** définit les noms, la position et, pour certains types de fichier, les longueurs des champs du fichier. Après avoir défini un fichier de sortie sur l'onglet **Propriétés de fichier**, vous pouvez définir les champs.

Si le fichier de sortie contient un enregistrement d'en-tête, vous pouvez rapidement définir les champs en cliquant sur **Régénérer**.

Pour définir les champs avec des valeurs par défaut pour la position, la longueur et le type de données, cliquez sur **Ajout rapide** et sélectionnez les champs à ajouter.

Si le fichier d'entrée ne contient pas d'enregistrement d'en-tête, ou si vous souhaitez définir manuellement les champs, suivez les étapes suivantes :

1. Cliquez sur **Ajouter**.
2. Dans le champ **Name**, choisissez le champ que vous souhaitez ajouter.
3. Dans le champ **Type**, sélectionnez le type de données du champ provenant du flux de données.

Spectrum™ Technology Platform prend en charge les types de données suivants :

**bigdecimal** Type de données numérique prenant en charge 38 points de précision décimaux. Utilisez ce type de données pour les données utilisées dans les calculs mathématiques requérant un haut niveau de précision, en particulier ceux impliquant des données financières. Le type de données bigdecimal prend en charge des calculs plus précis que le type de données double.

**booléen** Type logique doté de deux valeurs : True et False.

**bytearray** Tableau (liste) d'octets.

**Remarque** : ByteArray n'est pas pris en charge en entrée pour un service REST.

**date** Type de données contenant un mois, un jour et une année. Par exemple, 2012-01-30 ou January 30, 2012. Vous pouvez indiquer un format de date par défaut dans Management Console.

**datetime** Type de données contenant un mois, un jour, une année et des heures, des minutes et des secondes. Par exemple, 2012/01/30 6:15 PM.

**double** Un type de données numérique qui contient à la fois des nombres en double précision négatifs et positifs entre  $2^{-1074}$  et  $(2-2^{-52}) \times 2^{1023}$ . En notation E, la plage de valeurs est comprise entre -1.79769313486232E+308 et 1.79769313486232E+308.

**flottant** Un type de données numérique contenant des nombres de précision seuls négatifs et positifs entre  $2^{-149}$  et  $(2-2^{23}) \times 2^{127}$ . En notation E, la plage de valeurs est comprise entre -3.402823E+38 et 3.402823E+38.

**integer** Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre  $-2^{31}$  (-2,147,483,648) et  $2^{31}-1$  (2,147,483,647).

**liste** Au sens strict du terme, une liste n'est pas un type de données. Cependant, lorsqu'un champ contient des données hiérarchiques, il est traité comme un champ de « liste ». Dans Spectrum™ Technology Platform, une liste représente un ensemble de données composées de plusieurs valeurs. Par exemple, un champ Noms peut contenir une liste de valeurs de noms. Cela peut être représenté sous forme d'une structure XML, comme suit :

```
<Names>
  <Name>John Smith</Name>
```

```
<Name>Ann Fowler</Name>
</Names>
```

Il est important de souligner que le type de données de liste Spectrum™ Technology Platform est différent du type de données de liste de schéma XML dans le sens où ce dernier est un type de données simple composé de plusieurs valeurs, tandis que le type de données de liste Spectrum™ Technology Platform est semblable à un type de données complexe XML.

<b>long</b>	Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre $-2^{63}$ (-9,223,372,036,854,775,808) et $2^{63}-1$ (9,223,372,036,854,775,807).
<b>chaîne</b>	Séquence de caractères.
<b>time</b>	Type de données contenant l'heure du jour. Par exemple, 21:15:59 ou 9:15:59 PM.

- Si vous avez sélectionné une date, une heure ou un type de données numérique, vous pouvez utiliser le format date/heure ou numérique par défaut, ou vous pouvez spécifier un autre format pour ce champ spécifique. Le format par défaut est le format par défaut du système qui a été défini dans les options de conversion du type dans Management Console, ou le format par défaut du flux de données spécifié dans les options de conversion du type dans Enterprise Designer. Le format qui s'applique apparaît. Pour utiliser le format par défaut, laissez l'option **Valeur par défaut** sélectionnée. Pour indiquer un autre format, choisissez **Personnalisé** et suivez les étapes suivantes :

- Dans le champ **Locale**, sélectionnez le pays dont vous souhaitez utiliser la convention de formatage. Votre sélection déterminera les valeurs par défaut dans le champ **Format**. Pour les données de date, votre sélection déterminera également la langue utilisée lorsqu'un mois est épilé. Par exemple, si vous indiquez Anglais, le premier mois de l'année sera « January », alors que si vous indiquez Français, ce sera « Janvier ».
- Dans le champ **Format**, sélectionnez le format des données. Le format dépend du type de données du champ. Une liste des formats les plus couramment utilisés pour les paramètres régionaux sélectionnés est fournie.

Un exemple du format sélectionné s'affiche à droite du champ **Format**.

Vous pouvez également spécifier vos propres formats de date, heure et nombre si ceux disponibles ne répondent pas à vos besoins. Pour indiquer votre propre format de date ou heure, tapez le format dans le champ à l'aide de la notation décrite dans [Modèles de date et d'heure](#) à la page 306. Pour indiquer votre propre format numérique, tapez le format dans le fichier à l'aide de la notation décrite dans [Modèles numériques](#) à la page 309.

- Cliquez sur **Ajouter**.

Après avoir défini les champs dans votre fichier de sortie, vous pouvez en modifier le contenu et la disposition.

Nom de l'option	Description
Ajouter	Ajoute un champ dans la sortie. Vous pouvez ajouter un champ à la fin de la disposition existante ou insérer un champ à une position existante et la position des champs restants est ajustée en conséquence.
Modifier	Modifie le nom et le type du champ.
Enlever	Enlève le champ sélectionné de la sortie.
Déplacer vers le haut/Déplacer vers le bas	Réorganise le champ sélectionné.

## Définition de champs dans un fichier à séquence de ligne ou à longueur fixe

Dans le stage Write to File, l'onglet **Champs** définit les noms, la position et, pour certains types de fichier, les longueurs des champs du fichier. Après avoir défini un fichier de sortie sur l'onglet **Propriétés de fichier**, vous pouvez définir les champs.

Pour définir les champs avec des valeurs par défaut pour la position, la longueur et le type de données, cliquez sur **Ajout rapide** et sélectionnez les champs à ajouter.

Pour ajouter des champs manuellement à partir d'une liste de champs utilisés dans le flux de données, procédez comme suit :

1. Cliquez sur **Ajouter**.
2. Dans le champ **Name**, choisissez le champ que vous souhaitez ajouter.
3. Dans le champ **Type**, sélectionnez le type de données du champ provenant du flux de données.

Spectrum™ Technology Platform prend en charge les types de données suivants :

- bigdecimal** Type de données numérique prenant en charge 38 points de précision décimaux. Utilisez ce type de données pour les données utilisées dans les calculs mathématiques requérant un haut niveau de précision, en particulier ceux impliquant des données financières. Le type de données bigdecimal prend en charge des calculs plus précis que le type de données double.
- booléen** Type logique doté de deux valeurs : True et False.

<b>bytearray</b>	Tableau (liste) d'octets.  <b>Remarque :</b> ByteArray n'est pas pris en charge en entrée pour un service REST.
<b>date</b>	Type de données contenant un mois, un jour et une année. Par exemple, 2012-01-30 ou January 30, 2012. Vous pouvez indiquer un format de date par défaut dans Management Console.
<b>datetime</b>	Type de données contenant un mois, un jour, une année et des heures, des minutes et des secondes. Par exemple, 2012/01/30 6:15 PM.
<b>double</b>	Un type de données numérique qui contient à la fois des nombres en double précision négatifs et positifs entre $2^{-1074}$ et $(2-2^{-52}) \times 2^{1023}$ . En notation E, la plage de valeurs est comprise entre -1.79769313486232E+308 et 1.79769313486232E+308.
<b>flottant</b>	Un type de données numérique contenant des nombres de précision seuls négatifs et positifs entre $2^{-149}$ et $(2-2^{23}) \times 2^{127}$ . En notation E, la plage de valeurs est comprise entre -3.402823E+38 et 3.402823E+38.
<b>integer</b>	Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre $-2^{31}$ (-2,147,483,648) et $2^{31}-1$ (2,147,483,647).
<b>liste</b>	Au sens strict du terme, une liste n'est pas un type de données. Cependant, lorsqu'un champ contient des données hiérarchiques, il est traité comme un champ de « liste ». Dans Spectrum™ Technology Platform, une liste représente un ensemble de données composées de plusieurs valeurs. Par exemple, un champ Noms peut contenir une liste de valeurs de noms. Cela peut être représenté sous forme d'une structure XML, comme suit :  <pre>&lt;Names&gt;   &lt;Name&gt;John Smith&lt;/Name&gt;   &lt;Name&gt;Ann Fowler&lt;/Name&gt; &lt;/Names&gt;</pre> <p>Il est important de souligner que le type de données de liste Spectrum™ Technology Platform est différent du type de données de liste de schéma XML dans le sens où ce dernier est un type de données simple composé de plusieurs valeurs, tandis que le type de données de liste Spectrum™ Technology Platform est semblable à un type de données complexe XML.</p>
<b>long</b>	Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre $-2^{63}$ (-9,223,372,036,854,775,808) et $2^{63}-1$ (9,223,372,036,854,775,807).
<b>chaîne</b>	Séquence de caractères.
<b>time</b>	Type de données contenant l'heure du jour. Par exemple, 21:15:59 ou 9:15:59 PM.

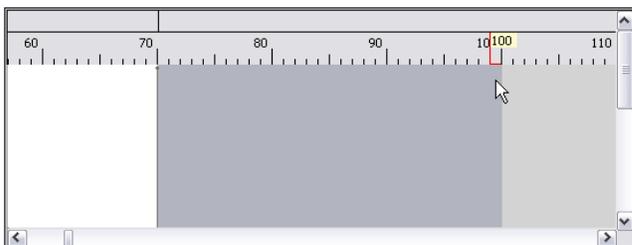
4. Si vous avez sélectionné une date, une heure ou un type de données numérique, vous pouvez utiliser le format date/heure ou numérique par défaut, ou vous pouvez spécifier un autre format pour ce champ spécifique. Le format par défaut est le format par défaut du système qui a été défini dans les options de conversion du type dans Management Console, ou le format par défaut du flux de données spécifié dans les options de conversion du type dans Enterprise Designer. Le format qui s'applique apparaît. Pour utiliser le format par défaut, laissez l'option **Valeur par défaut** sélectionnée. Pour indiquer un autre format, choisissez **Personnalisé** et suivez les étapes suivantes :
  - a) Dans le champ **Locale**, sélectionnez le pays dont vous souhaitez utiliser la convention de formatage. Votre sélection déterminera les valeurs par défaut dans le champ **Format**. Pour les données de date, votre sélection déterminera également la langue utilisée lorsqu'un mois est épilé. Par exemple, si vous indiquez Anglais, le premier mois de l'année sera « January », alors que si vous indiquez Français, ce sera « Janvier ».
  - b) Dans le champ **Format**, sélectionnez le format des données. Le format dépend du type de données du champ. Une liste des formats les plus couramment utilisés pour les paramètres régionaux sélectionnés est fournie.

Un exemple du format sélectionné s'affiche à droite du champ **Format**.

Vous pouvez également spécifier vos propres formats de date, heure et nombre si ceux disponibles ne répondent pas à vos besoins. Pour indiquer votre propre format de date ou heure, tapez le format dans le champ à l'aide de la notation décrite dans [Modèles de date et d'heure](#) à la page 306. Pour indiquer votre propre format numérique, tapez le format dans le fichier à l'aide de la notation décrite dans [Modèles numériques](#) à la page 309.

5. Les champs **Position de début** et **Longueur** sont remplis automatiquement en fonction des données du flux de données et le nombre de champs ajoutés.
6. Cliquez sur **Ajouter**.

Sinon, vous pouvez également ajouter un champ en définissant d'abord la position de départ et la longueur du champ. Pour ce faire, sous **Fichier modèle** cliquez à l'endroit où vous voulez ajouter un champ et faites glisser vers la gauche afin de sélectionner le champ souhaité, comme indiqué ici :



Après avoir défini les champs dans votre fichier de sortie, vous pouvez en modifier le contenu et la disposition. L'option **Recalculer la position de départ** indique au stage Write to File de recalculer les positions des champs lorsque vous modifiez, déplacez ou enlevez un champ dans le fichier de sortie. Décochez cette case si vous ne souhaitez pas que les positions soient recalculées et qu'à la place vous souhaitez que les champs restent dans leur position existante une fois le fichier de sortie modifié.

Nom de l'option	Description
Ajouter	Ajoute un champ dans la sortie.
Modifier	Modifie le nom, le type, la position de départ et la longueur du champ.
Supprimer	Enlève le champ sélectionné de la sortie.
Déplacer vers le haut/Déplacer vers le bas	Réorganise le champ sélectionné.

## Tri des enregistrements de sortie

Dans le stage Write to File, l'onglet **Trier les champs** définit les champs permettant de trier les enregistrements de sortie avant leur écriture dans le fichier de sortie. Le tri est facultatif.

1. Dans Write to File, cliquez sur l'onglet **Trier les champs**.
2. Cliquez sur **Ajouter**.
3. Cliquez sur la flèche de déroulement dans la colonne **Nom du champ** et sélectionnez le champ que vous voulez trier. Les champs disponibles à la sélection dépendent des champs dans le flux de données.
4. Dans la colonne **Ordre**, sélectionnez Ascending ou Descending.
5. Répétez cette opération jusqu'à l'ajout de tous les champs d'entrée que vous souhaitez utiliser pour le tri. Changez l'ordre du tri en surlignant la ligne pour le champ que vous voulez déplacer et cliquez sur **Vers le haut** ou **Vers le bas**.
6. Les options de performance de tri par défaut pour votre système sont définies dans Management Console Si vous voulez remplacer les options de performance de tri par défaut, cliquez sur l'onglet **Avancé**. La boîte de dialogue **Options avancées** contient les options de performance de tri suivantes :

**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 :

$$(NumberOfRecords \times 2) \div InMemoryRecordLimit = NumberOfTempFiles$$

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 :

$$(InMemoryRecordLimit \times MaxNumberOfTempFiles \div 2) \geq TotalNumberOfRecords$$

## Fichier de paramètres de définition du fichier

Un fichier de paramètres de définition du fichier contient le dessin d'enregistrement du fichier, l'encodage et les options de tri ayant été exportés à partir d'un stage Read from File ou Write to File. Le fichier de paramètres de définition du fichier peut être importé dans Read from File ou Write to File pour définir rapidement les options du stage au lieu d'indiquer manuellement les options.

La manière la plus simple de créer un fichier de paramètres de définition du fichier est d'indiquer les paramètres de fichier à l'aide de Read from File ou de Write to File, puis de cliquer sur le bouton **Exporter** pour générer le fichier de paramètres de définition du fichier.

Cependant, à titre informatif, le schéma du fichier de paramètres de définition du fichier s'affiche ci-dessous. Chaque élément dans le fichier XML possède un type et si ce type est autre qu'une chaîne ou un nombre entier, les valeurs autorisées s'affichent. Ces valeurs correspondent directement aux options contenues dans la boîte de dialogue du stage. Par exemple, l'élément `FileTypeEnum` correspond au champ `Type` d'enregistrement sous l'onglet `Propriétés de fichier`, et les trois valeurs suivantes apparaissent dans le schéma : `linesequential`, `fixedwidth` et `delimited`.

**Remarque :** Si vous saisissez « personnalisé » pour les champs `LineSeparator`, `FieldSeparator` ou `TextQualifier`, un élément personnalisé correspondant doit être inclus (par ex., « `CustomLineSeparator` », « `CustomFieldSeparator` » ou « `CustomTextQualifier` ») avec un nombre hexadécimal représentant le caractère ou la séquence de caractères à utiliser.

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema elementFormDefault="qualified"
  xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="FileSchema" nillable="true" type="FileSchema"/>
  <xs:complexType name="FileSchema">
    <xs:sequence>
      <xs:element
        minOccurs="0"
        maxOccurs="1"
        default="linesequential"
        name="Type"
        type="FileTypeEnum"/>
      <xs:element
        minOccurs="0"
        maxOccurs="1"
        default="UTF-8" name="Encoding" type="xs:string"/>
      <xs:element
        minOccurs="0"
        maxOccurs="1"
        name="RecordLength"
        type="xs:int"/>
      <xs:element
        minOccurs="0"
        maxOccurs="1"
        default="default"
        name="LineSeparator"
        type="LineSeparatorEnum"/>
      <xs:element
        minOccurs="0"
        maxOccurs="1"
        name="CustomLineSeparator"
        type="xs:string"/>
      <xs:element
        minOccurs="0"
        maxOccurs="1"
        default="comma"
        name="FieldSeparator"
        type="FieldSeparatorEnum"/>
    <xs:element
```

```

        minOccurs="0"
        maxOccurs="1"
        name="CustomFieldSeparator"
        type="xs:string"/>
<xs:element
    minOccurs="0"
    maxOccurs="1"
    default="none"
    name="TextQualifier"
    type="TextQualifierEnum"/>
<xs:element
    minOccurs="0"
    maxOccurs="1"
    name="CustomTextQualifier"
    type="xs:string"/>
<xs:element
    minOccurs="0"
    maxOccurs="1"
    default="false"
    name="HasHeader"
    type="xs:boolean"/>
<xs:element
    minOccurs="0"
    maxOccurs="1"
    default="true"
    name="EnforceColumnCount"
    type="xs:boolean"/>
<xs:element
    minOccurs="0"
    maxOccurs="1"
    name="Fields"
    type="ArrayOfFieldSchema"/>
</xs:sequence>
</xs:complexType>
<xs:simpleType name="FileTypeEnum">
    <xs:restriction base="xs:string">
        <xs:enumeration value="linesequential"/>
        <xs:enumeration value="fixedwidth"/>
        <xs:enumeration value="delimited"/>
    </xs:restriction>
</xs:simpleType>
<xs:simpleType name="LineSeparatorEnum">
    <xs:restriction base="xs:string">
        <xs:enumeration value="default"/>
        <xs:enumeration value="windows"/>
        <xs:enumeration value="unix"/>
        <xs:enumeration value="mac"/>
        <xs:enumeration value="custom"/>
    </xs:restriction>
</xs:simpleType>
<xs:simpleType name="FieldSeparatorEnum">
    <xs:restriction base="xs:string">
        <xs:enumeration value="comma"/>

```

```

    <xs:enumeration value="tab"/>
    <xs:enumeration value="space"/>
    <xs:enumeration value="semicolon"/>
    <xs:enumeration value="period"/>
    <xs:enumeration value="pipe"/>
    <xs:enumeration value="custom"/>
  </xs:restriction>
</xs:simpleType>
<xs:simpleType name="TextQualifierEnum">
  <xs:restriction base="xs:string">
    <xs:enumeration value="none"/>
    <xs:enumeration value="single"/>
    <xs:enumeration value="double"/>
    <xs:enumeration value="custom"/>
  </xs:restriction>
</xs:simpleType>
<xs:complexType name="ArrayOfFieldSchema">
  <xs:sequence>
    <xs:element
      minOccurs="0"
      maxOccurs="unbounded"
      name="Field"
      nillable="true"
      type="FieldSchema"/>
  </xs:sequence>
</xs:complexType>
<xs:complexType name="FieldSchema">
  <xs:sequence>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      name="Name"
      type="xs:string"/>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      default="string"
      name="Type"
      type="xs:string"/>
    <xs:element
      minOccurs="1"
      maxOccurs="1"
      name="Position"
      type="xs:int"/>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      name="Length"
      type="xs:int"/>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      default="false"

```

```

        name="Trim"
        type="xs:boolean"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  name="Locale"
  type="Locale"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  name="Pattern"
  type="xs:string"/>
<xs:element
  minOccurs="0"
  maxOccurs="1"
  default="none"
  name="Order"
  type="SortOrderEnum"/>
</xs:sequence>
</xs:complexType>
<xs:complexType name="Locale">
  <xs:sequence>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      name="Country"
      type="xs:string"/>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      name="Language"
      type="xs:string"/>
    <xs:element
      minOccurs="0"
      maxOccurs="1"
      name="Variant"
      type="xs:string"/>
  </xs:sequence>
</xs:complexType>
<xs:simpleType name="SortOrderEnum">
  <xs:restriction base="xs:string">
    <xs:enumeration value="none"/>
    <xs:enumeration value="ascending"/>
    <xs:enumeration value="descending"/>
  </xs:restriction>
</xs:simpleType>
</xs:schema>

```

## Write to Hadoop Sequence File

Le stage Write to Hadoop Sequence File écrit les données dans un fichier de séquence comme sortie d'un flux de données. Un fichier de séquence est un fichier plat constitué de paires de valeurs clés binaires. Pour obtenir davantage d'informations, consultez l'adresse suivante :

[wiki.apache.org/hadoop/SequenceFile](http://wiki.apache.org/hadoop/SequenceFile).

**Remarque :** Le stage Write to Hadoop Sequence File prend uniquement en charge les fichiers de séquence délimités non compressés figurant sur Hadoop Distributed File System (HDFS).

### Onglet de propriétés du fichier

Champs	Description
Serveur	Indique que le fichier que vous sélectionnez dans le champ <b>Nom de fichier</b> se trouve sur le système Hadoop. Avant de pouvoir l'utiliser dans le stage, vous devez créer une connexion au serveur de fichiers Hadoop dans Management Console. Si vous sélectionnez un fichier sur le système Hadoop, le nom de serveur est le nom que vous indiquez dans Management Console lors de la création d'un serveur de fichiers.
Nom du fichier	Indique le chemin du fichier. Cliquez sur le bouton de sélection (...) pour trouver le fichier souhaité.
Séparateur de champs	<p>Indique le caractère utilisé pour séparer des champs dans un fichier délimité.</p> <p>Par exemple, cet enregistrement utilise une barre verticale ( ) comme séparateur de champ :</p> <pre>7200 13TH ST MIAMI FL 33144</pre> <p>Ces caractères disponibles à définir en tant que séparateurs de champs sont :</p> <ul style="list-style-type: none"> <li>• Espace</li> <li>• Tabulation</li> <li>• Virgule</li> <li>• Point (.)</li> <li>• Point-virgule</li> <li>• Conduite</li> </ul> <p>Si le fichier utilise un caractère différent comme séparateur de champ, cliquez sur le bouton de sélection pour sélectionner un autre caractère délimiteur.</p>

Champs	Description
Qualificateur de texte	<p>Le caractère utilisé pour entourer les valeurs de texte dans un fichier délimité. Par exemple, cet enregistrement utilise des guillemets doubles (") en tant que qualificateur de texte.</p> <pre>"7200 13TH ST" "MIAMI" "FL" "33144"</pre> <p>Ces caractères disponibles à définir en tant que qualificateurs de texte sont :</p> <ul style="list-style-type: none"> <li>• Guillemets simples (')</li> <li>• Guillemets doubles (")</li> </ul> <p>Si le fichier utilise un délimiteur de texte, cliquez sur le bouton de sélection pour sélectionner un autre caractère délimiteur.</p>

### Onglet Champs

L'onglet Champs définit les noms, positions et types des champs du fichier. (Pour plus d'informations, reportez-vous à la section [Définition de champs dans un fichier de séquence de sortie](#) à la page 270.)

## Définition de champs dans un fichier de séquence de sortie

Dans le stage Write to Hadoop Sequence File, l'onglet **Champs** définit les noms, les positions et les types des champs du fichier. Après avoir défini un fichier d'entrée sur l'onglet **Propriétés de fichier**, vous pouvez définir les champs.

1. Pour sélectionner les champs de votre choix dans les données d'entrée ou un fichier existant, cliquez sur **Ajout rapide**.
  - a) Sélectionnez les champs spécifiques à partir des données d'entrée.
  - b) Cliquez sur **OK**.
2. Pour ajouter de nouveaux champs, cliquez sur **Ajouter**.
  - a) Saisissez le **Nom** du champ.
  - b) Sélectionnez le **Type** du champ. Le stage prend en charge les types de données suivants :
    - booléen** Type logique doté de deux valeurs : True et False.
    - date** Type de données contenant un mois, un jour et une année. Par exemple, 2012-01-30 ou January 30, 2012. Vous pouvez indiquer un format de date par défaut dans Management Console.

**datetime** Type de données contenant un mois, un jour, une année et des heures, des minutes et des secondes. Par exemple, 2012/01/30 6:15 PM.

**Remarque :** Dans les fichiers Parquet, les types de données `datetime` et `time` sont mappés sous forme de `String`. Dans les fichiers RC, le type de données `datetime` est mappé sous forme de `timestamp`.

**double** Un type de données numérique qui contient à la fois des nombres en double précision négatifs et positifs entre  $2^{-1074}$  et  $(2-2^{-52}) \times 2^{1023}$ . En notation E, la plage de valeurs est comprise entre -1.79769313486232E+308 et 1.79769313486232E+308.

**flottant** Un type de données numérique contenant des nombres de précision seuls négatifs et positifs entre  $2^{-149}$  et  $(2-2^{23}) \times 2^{127}$ . En notation E, la plage de valeurs est comprise entre -3.402823E+38 et 3.402823E+38.

**integer** Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre  $-2^{31}$  (-2,147,483,648) et  $2^{31}-1$  (2,147,483,647).

**bigdecimal** Type de données numérique prenant en charge 38 points de précision décimaux. Utilisez ce type de données pour les données utilisées dans les calculs mathématiques requérant un haut niveau de précision, en particulier ceux impliquant des données financières. Le type de données `bigdecimal` prend en charge des calculs plus précis que le type de données `double`.

**Remarque :** Pour les fichiers RC, Avro et Parquet Hive, le type de données `bigdecimal` est converti en type de données `decimal` à une précision 38 et une échelle 10.

**long** Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre  $-2^{63}$  (-9,223,372,036,854,775,808) et  $2^{63}-1$  (9,223,372,036,854,775,807).

**Remarque :** Dans les fichiers RC, le type de données `long` est mappé sous forme de type de données `bigint`.

**chaîne** Séquence de caractères.

- c) Dans le champ **Position**, entrez la position de ce champ dans l'enregistrement.

Par exemple, dans ce fichier d'entrée, `AddressLine1` est en position 1, `City` est en position 2, `StateProvince` est en position 3, et `PostalCode` est en position 4.

```
"AddressLine1"|"City"|"StateProvince"|"PostalCode"
"7200 13TH ST"|"MIAMI"|"FL"|"33144"
"One Global View"|"Troy"|"NY"|"12180"
```

3. Si vous écrasez un fichier existant, cliquez sur **Régénérer** pour récupérer le schéma du fichier existant, puis modifiez-le.  
Cette opération génère le schéma en fonction des 50 premiers enregistrements dans les données d'entrée de ce stage.
4. Si vous souhaitez que tout caractère d'espacement superflu soit supprimé au début et à la fin d'une chaîne de caractères dans un champ, cochez la case **Espaces courts**.
5. Indiquez l'une des options suivantes pour générer la clé :

**Générer automatiquement** Le cluster Hadoop génère automatiquement la clé. Pour la génération automatique, tous les champs de la grille sont considérés comme des champs de valeur. Le type de données de la clé est long.

**Défini par l'utilisateur** Par défaut, le premier champ de la grille est sélectionné comme champ de clé. Une icône est affichée pour indiquer que le champ est le champ de clé. Vous pouvez sélectionner tout autre champ comme champ de clé.

Après avoir défini les champs dans votre fichier de sortie, vous pouvez en modifier le contenu et la disposition.

Nom de l'option	Description
Ajouter	Ajoute un champ dans la sortie. Vous pouvez ajouter un champ à la fin de la disposition existante ou insérer un champ à une position existante et la position des champs restants est ajustée en conséquence.
Modifier	Modifie le nom et le type du champ.
Enlever	Enlève le champ sélectionné de la sortie.
Déplacer vers le haut/Déplacer vers le bas	Réorganise le champ sélectionné.

## Write to Hive File

Le stage **Write to Hive File** écrit l'entrée de flux de données dans le fichier Hive de sortie spécifié.

Pour le fichier de sortie, vous pouvez sélectionner l'un des formats de fichier Hive pris en charge ci-dessous :

- ORC
- RC
- PARQUET
- Avro

### Onglet propriétés du fichier

**Tableau 5 : Propriétés de fichier communes**

Champs	Description
Server name	Indique que le fichier que vous sélectionnez dans le champ <b>Nom de fichier</b> se trouve sur le système Hadoop. Une fois que vous sélectionnez un fichier qui se trouve sur un système Hadoop, le <b>nom du serveur</b> reflète le nom du serveur de fichiers respectif, comme spécifié dans Management Console.
Nom du fichier	<p>Cliquez sur le bouton d'ellipse (...) pour accéder au fichier de sortie Hive qui doit être créé dans le serveur de fichiers Hadoop défini. Les données de sortie de ce stage sont écrites dans le fichier sélectionné.</p> <p><b>Remarque :</b> Avant de pouvoir l'utiliser dans le stage, vous devez créer une connexion au serveur de fichiers Hadoop dans Management Console.</p>
File type	<p>Sélectionnez l'un des quatre formats de fichier Hive pris en charge :</p> <ul style="list-style-type: none"> <li>• ORC</li> <li>• RC</li> <li>• PARQUET</li> <li>• Avro</li> </ul>

**Tableau 6 : Propriétés de fichier ORC**

Champs	Description
Taille du tampon	<p>Définit la taille du tampon à allouer lors de l'écriture dans un fichier ORC. Cela est spécifié en kilooctets.</p> <p><b>Remarque :</b> La taille du tampon par défaut est 256 Ko.</p>

Champs	Description
Taille de bande	Définit la taille de bande à créer lors de l'écriture dans un fichier ORC. Cela est spécifié en mégaoctets. <b>Remarque</b> : La taille de bande par défaut est 64 Mo.
Row index stride	Définit le nombre de lignes à écrire entre deux entrées d'index de ligne consécutives. <b>Remarque</b> : La valeur par défaut de Row Index Stride est 10 000 lignes.
Type de compression	Définit le type de compression à utiliser lors de l'écriture dans un fichier ORC. Les types de compression disponibles sont <i>ZLIB</i> et <i>SNAPPY</i> . <b>Remarque</b> : Le type de compression par défaut est <i>ZLIB</i> .
Padding	Indique si les bandes sont complétées pour minimiser les bandes qui croisent les limites de bloc HDFS lors de l'écriture dans un fichier ORC. <b>Remarque</b> : Par défaut, la case <b>Pad</b> est cochée.
Aperçu	Les 50 premiers enregistrements du fichier écrit sont extraits et affichés dans la grille <b>Aperçu</b> , une fois que le flux de données a été exécuté au moins une fois, et que les données ont été écrites dans le fichier sélectionné.

**Tableau 7 : Propriétés de fichier RC**

Champs	Description
Taille du tampon	Définit la taille du tampon à allouer lors de l'écriture dans un fichier RC. Cela est spécifié en kilooctets. <b>Remarque</b> : La taille du tampon par défaut est 256 Ko.
Taille du bloc	Définit la taille de bloc à créer lors de l'écriture dans un fichier RC. Cela est spécifié en mégaoctets. <b>Remarque</b> : La taille de bloc par défaut est de 64 Mo.

Champs	Description
Type de compression	Définit le type de compression à utiliser lors de l'écriture dans un fichier RC. Les types de compression disponibles sont NONE et DEFLATE.  <b>Remarque</b> : Le type de compression par défaut est NONE.
Aperçu	Les 50 premiers enregistrements du fichier écrit sont extraits et affichés dans la grille <b>Aperçu</b> , une fois que le flux de données a été exécuté au moins une fois, et que les données ont été écrites dans le fichier sélectionné.  L'onglet <b>Champs</b> est utilisé pour définir la séquence et le type de données des champs obligatoires.  <b>Remarque</b> : Pour le type de fichier RC, vous devez définir les métadonnées du fichier de sortie avant de cliquer sur <b>Aperçu</b> pour charger la grille <b>Aperçu</b> .

### Tableau 8 : Propriétés de fichier Parquet

Champs	Description
Type de compression	Définit le type de compression à utiliser lors de l'écriture dans un fichier PARQUET. Les types de compression disponibles sont UNCOMPRESSED, GZIP et SNAPPY.  <b>Remarque</b> : Le type de compression par défaut est UNCOMPRESSED.
Taille du bloc	Définit la taille de bloc à créer lors de l'écriture dans un fichier PARQUET. Cela est spécifié en mégaoctets.  <b>Remarque</b> : La taille de bloc par défaut est de 128 Mo.
Taille de la page	La taille de page est pour la compression. Lors de la lecture, chaque page peut être décompressée indépendamment. Cela est spécifié en kilooctets.  <b>Remarque</b> : La taille de page par défaut est de 1 024 Ko.
Enable dictionary	Activer/désactiver le codage du dictionnaire.  <b>Avertissement</b> : Le dictionnaire doit être activé pour que Dictionary Page Size soit activé.  <b>Remarque</b> : La valeur par défaut est <code>true</code> .

Champs	Description
Dictionary Page size	<p>Il existe une page de dictionnaire par colonne par groupe de lignes lorsque le codage du dictionnaire est utilisé. La taille de page de dictionnaire fonctionne comme la taille de page. Cela est spécifié en kilooctets.</p> <p><b>Remarque :</b> La taille de page de dictionnaire par défaut est de 1 024 Ko.</p>
Writer version	<p>Parquet prend en charge deux versions de l'API Writer : PARQUET_1_0 et PARQUET_2_0.</p> <p><b>Remarque :</b> La valeur par défaut est PARQUET_1_0.</p>
Aperçu	<p>Les 50 premiers enregistrements du fichier écrit sont extraits et affichés dans la grille <b>Aperçu</b>, une fois que le flux de données a été exécuté au moins une fois, et que les données ont été écrites dans le fichier sélectionné.</p>

### Tableau 9 : Propriétés de fichier Avro

Champs	Description
Intervalle de sync. (en octets)	<p>Spécifie le nombre approximatif d'octets non compressés à écrire dans chaque bloc. La plage de valeurs valide est comprise entre 32 et 2<sup>30</sup>. Cependant, il est conseillé de conserver l'intervalle de synchronisation entre 2 K et 2 M.</p> <p><b>Remarque :</b> L'intervalle de sync. par défaut est 16000.</p>
Compression	<p>Définit le type de compression à utiliser lors de l'écriture dans un fichier Avro. Les types de compression disponibles sont <b>NONE</b>, <b>SNAPPY</b> et <b>DEFLATE</b>. La sélection de la compression <b>DEFLATE</b> vous permet également de sélectionner le niveau de compression (décrit ci-dessous).</p> <p><b>Remarque :</b> Le type de compression par défaut est NONE.</p>
Niveau de compression	<p>Ce champ s'affiche si vous sélectionnez l'option <b>DEFLATE</b> dans le champ <b>Compression</b> ci-dessus.</p> <p>Il peut prendre des valeurs comprises entre 0 et 9, où 0 indique aucune compression. Le niveau de compression augmente de 1 à 9, avec une augmentation simultanée du temps nécessaire pour compresser les données.</p> <p><b>Remarque :</b> Le type de compression par défaut est 1.</p>

Champs	Description
Aperçu	Les 50 premiers enregistrements du fichier écrit sont extraits et affichés dans cette grille, une fois que le flux de données a été exécuté au moins une fois, et que les données ont été écrites dans le fichier sélectionné.

### Onglet Champs

L'onglet **Champs** définit les noms et les types des champs tels qu'ils figurent dans le fichier source de ce stage, et qui doivent être sélectionnés pour être écrits dans le fichier de sortie.

Pour plus d'informations, reportez-vous à la section [Définition des champs du stage Writing to Hive File](#) à la page 277.

### Onglet Exécution

L'onglet **Exécution** offre la possibilité d'**écraser** un fichier existant portant le même nom dans le serveur de fichiers Hadoop configuré. Si vous cochez la case **Écraser**, lors de l'exécution du flux de données, le nouveau fichier de sortie Hive remplace tout fichier existant portant le même nom dans le même serveur de fichiers Hadoop.

Par défaut, la case **Écraser** est décochée.

**Remarque :** Si vous ne sélectionnez pas **Écraser**, une exception est générée lors de l'exécution du flux de données, si le fichier à écrire porte le même nom qu'un fichier existant dans le même serveur de fichiers Hadoop.

## Définition des champs du stage Writing to Hive File

L'onglet **Champs** du stage **Write to Hive File** répertorie les noms de schéma et les types de données des champs des données d'entrées du stage.

- Pour sélectionner les champs de votre choix dans les données d'entrée ou un fichier existant, cliquez sur **Ajout rapide**.
  - Sélectionnez les champs spécifiques à partir des données d'entrée.
  - Cliquez sur **OK**.
- Pour ajouter de nouveaux champs, cliquez sur **Ajouter**.
  - Saisissez le **Nom** du champ.
  - Sélectionnez le **Type** du champ. Le stage prend en charge les types de données suivants :

<b>booléen</b>	Type logique doté de deux valeurs : True et False.
<b>date</b>	Type de données contenant un mois, un jour et une année. Par exemple, 2012-01-30 ou January 30, 2012. Vous pouvez indiquer un format de date par défaut dans Management Console.

**datetime** Type de données contenant un mois, un jour, une année et des heures, des minutes et des secondes. Par exemple, 2012/01/30 6:15 PM.

**Remarque :** Dans les fichiers Parquet, les types de données `datetime` et `time` sont mappés sous forme de `String`. Dans les fichiers RC, le type de données `datetime` est mappé sous forme de `timestamp`.

**double** Un type de données numérique qui contient à la fois des nombres en double précision négatifs et positifs entre  $2^{-1074}$  et  $(2-2^{-52}) \times 2^{1023}$ . En notation E, la plage de valeurs est comprise entre -1.79769313486232E+308 et 1.79769313486232E+308.

**flottant** Un type de données numérique contenant des nombres de précision seuls négatifs et positifs entre  $2^{-149}$  et  $(2-2^{23}) \times 2^{127}$ . En notation E, la plage de valeurs est comprise entre -3.402823E+38 et 3.402823E+38.

**integer** Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre  $-2^{31}$  (-2,147,483,648) et  $2^{31}-1$  (2,147,483,647).

**bigdecimal** Type de données numérique prenant en charge 38 points de précision décimaux. Utilisez ce type de données pour les données utilisées dans les calculs mathématiques requérant un haut niveau de précision, en particulier ceux impliquant des données financières. Le type de données `bigdecimal` prend en charge des calculs plus précis que le type de données `double`.

**Remarque :** Pour les fichiers RC, Avro et Parquet Hive, le type de données `bigdecimal` est converti en type de données `decimal` à une précision 38 et une échelle 10.

**long** Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre  $-2^{63}$  (-9,223,372,036,854,775,808) et  $2^{63}-1$  (9,223,372,036,854,775,807).

**Remarque :** Dans les fichiers RC, le type de données `long` est mappé sous forme de type de données `bigint`.

**chaîne** Séquence de caractères.

- c) Dans le champ **Position**, entrez la position de ce champ dans l'enregistrement.

Par exemple, dans ce fichier d'entrée, `AddressLine1` est en position 1, `City` est en position 2, `StateProvince` est en position 3, et `PostalCode` est en position 4.

```
"AddressLine1"|"City"|"StateProvince"|"PostalCode"
"7200 13TH ST"|"MIAMI"|"FL"|"33144"
"One Global View"|"Troy"|"NY"|"12180"
```

- Si vous écrasez un fichier existant, cliquez sur **Régénérer** pour récupérer le schéma du fichier existant, puis modifiez-le.

Cette opération génère le schéma en fonction des métadonnées du fichier existant, s'il s'agit de fichiers de sortie ORC et Parquet. Pour les fichiers de sortie RC, vous devez explicitement ajouter les champs pour remplacer les champs existants.

La colonne **Nom** répertorie les noms des différentes colonnes des données d'entrée. La colonne **Type** répertorie les types de données de chaque champ respectif des données d'entrée.

**Remarque** : En cas de type de fichier *Parquet*, une autre colonne **Annulable** indique si le champ est annulable ou non. Vous pouvez cocher cette case pour un champ donné rendre le champ annulable, ou la décochez dans le cas contraire.

- Vous pouvez modifier les noms, les types de données et les séquence des colonnes sélectionnées dans la sortie en utilisant les boutons ci-dessous :

Nom de l'option	Description
Ajouter	Ajoute un champ dans la sortie.
Modifier	Modifie le nom et le type de données du champ sélectionné.
Supprimer	Enlève le champ sélectionné de la sortie.
Déplacer vers le haut/Déplacer vers le bas	Réorganise la position du champ sélectionné dans la sortie.

- Cliquez sur **OK**.

## Write to NoSQL DB

Write to NoSQL DB écrit la sortie d'un dataflow dans une base de données NoSQL. Le stage prend en charge les types de base de données MongoDB et Couchbase.

**Remarque** : Une amélioration significative des performances peut être obtenue en utilisant plusieurs instances d'exécution de Write to NoSQLDB. Pour indiquer plusieurs instances d'exécution, cliquez sur le bouton **Exécution**.

## Onglet Général

Champs	Description
Connexion	<p>Sélectionnez la connexion à la base de données requise dans la liste déroulante. Les options affichées sont basées sur les connexions définies dans Management Console.</p> <p>Pour ajouter une nouvelle connexion, voir <a href="#">Connexion à NoSQL</a> à la page 54.</p> <p>Pour modifier une connexion existante, sélectionnez-la et ouvrez-la dans la liste des connexions de la page Sources de données de Management Console, effectuez les mises à jour requises, puis cliquez sur le bouton <b>Enregistrer</b>.</p>
Table/Vue	<p>Indique le nom de la collection dans laquelle vous souhaitez écrire.</p> <p>Vous pouvez créer une nouvelle collection dans la base de données NoSQL en saisissant un nom de collection dans la liste déroulante Table/Vue, puis en cliquant sur <b>Créer une table</b>.</p> <p><b>Remarque :</b> Pour Couchbase la liste déroulante Table/Vue et le bouton Créer une table sont désactivés, car nous écrivons dans un compartiment et non dans une vue. En outre, le bouton Aperçu est lui aussi désactivé.</p>
Ignorer les valeurs nulles	<p>Si cette option est activée, tout champ comportant une valeur nulle est ignoré.</p> <p><b>Remarque :</b> Si vous n'activez pas cette option, tout champ comportant une valeur null est également écrit dans la base de données.</p>
Aperçu	<p>Affiche les enregistrements de la table sélectionnée.</p> <p><b>Remarque :</b> Pour les sources de données MongoDB, cliquer sur <b>Aperçu</b> affiche les enregistrements filtrés, si une ou plusieurs clauses <i>where</i> ont été saisies dans le champ <b>Où</b>. En l'absence de clause where, l'aperçu n'affiche tous les enregistrements.</p> <p><b>Remarque :</b> Pour les sources de données Couchbase, cliquer sur <b>Aperçu</b> affiche également le champ <code>_id</code> ajouté contenant la clé. Si l'enregistrement comporte déjà un champ <code>_id</code>, il est écrasé par le champ <code>_id</code> ajouté lors de l'aperçu des champs.</p>
Développer tout	Développe les éléments de l'arborescence d'aperçu.
Réduire tout	Réduit les éléments de l'arborescence d'aperçu.

### Onglet Champs

L'onglet Champs vous permet de sélectionner les données que vous souhaitez écrire dans la base de données. Pour plus d'informations, reportez-vous à la section [Définition de champs dans une base de données NoSQL](#) à la page 281.

## Définition de champs dans une base de données NoSQL

Dans le stage Write to NoSQL DB, l'onglet **Champs** définit les noms et les types des champs récupérés du stage précédent.

1. Dans l'onglet **Champs**, cliquez sur **Régénérer**.

Cela affiche les champs provenant du stage précédent.

Les données sont affichées au format suivant : `Fieldname (datatype)`.

**Remarque** : Pour Couchbase, si l'enregistrement dispose d'un champ `_id`, ce champ est utilisé comme clé pour l'écriture de l'enregistrement dans la base de données. Sinon, la clé de l'enregistrement est automatiquement générée et écrite dans la base de données.

2. Pour modifier le nom et le type d'un champ, mettez le champ en surbrillance, puis cliquez sur **Modifier**.
3. Dans le champ **Name**, choisissez le champ à ajouter ou saisissez le nom du champ.
4. Dans le champ **Type**, vous pouvez laisser le type de données défini sur chaîne si vous ne prévoyez pas d'effectuer des opérations mathématiques sur les données. Cependant, si vous avez l'intention d'effectuer ce genre d'opérations, sélectionnez un type de données approprié. Cela convertira les données de chaîne à partir du fichier en un type de données qui activera la manipulation appropriée des données dans le flux de données.

Le stage prend en charge les types de données suivants :

<b>double</b>	Un type de données numérique qui contient à la fois des nombres en double précision négatifs et positifs entre $2^{-1074}$ et $(2-2^{-52}) \times 2^{1023}$ . En notation E, la plage de valeurs est comprise entre $-1.79769313486232E+308$ et $1.79769313486232E+308$ .
<b>flottant</b>	Un type de données numérique contenant des nombres de précision seuls négatifs et positifs entre $2^{-149}$ et $(2-2^{-23}) \times 2^{127}$ . En notation E, la plage de valeurs est comprise entre $-3.402823E+38$ et $3.402823E+38$ .
<b>integer</b>	Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre $-2^{31}$ ( $-2,147,483,648$ ) et $2^{31}-1$ ( $2,147,483,647$ ).
<b>long</b>	Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre $-2^{63}$ ( $-9,223,372,036,854,775,808$ ) et $2^{63}-1$ ( $9,223,372,036,854,775,807$ ).
<b>chaîne</b>	Séquence de caractères.

5. Sous l'onglet **Champs**, vous pouvez sélectionner individuellement chaque champ à écrire dans la base de données ou cliquer sur **Tout sélectionner** pour sélectionner tous les champs.
6. Éventuellement, dans l'onglet **Exécution**, vous pouvez définir la taille de lot. La taille de lot indique le nombre d'enregistrements à écrire en une fois dans la base de données.
7. Cliquez sur **OK**.

## Options de flux de données NoSQL DB

Cette procédure explique comment configurer un flux de données pour qu'il prenne en charge les options d'exécution NoSQL DB.

1. Ouvrez le flux de données dans Enterprise Designer.
2. Pour configurer les options d'exécution d'un stage dans un flux de données intégré, ouvrez ce dernier.
3. Cliquez sur l'icône Options de flux de données de la barre d'outils ou sur **Édition > Options de flux de données**. La boîte de dialogue **Options de flux de données** apparaît.
4. Cliquez sur **Ajouter**. La boîte de dialogue **Définir l'option de flux des données** apparaît.
5. Développez le stage NoSQLDB.
6. Les options de flux de données sont exposées comme décrit dans le tableau suivant :

Base de données	Lire	Écriture
Mongo DB	Connexion	Connexion
	table	table
Couchbase DB	Connexion	Connexion
	Vue	
	Nom de document de conception	

Le nom d'option NoSQL DB sélectionné s'affiche dans les champs **Nom de l'option** et **Libellé d'option**. Il s'agit du nom de l'option qui devra être spécifié lors de l'exécution afin de définir cette option.

7. Saisissez une description de l'option dans le champ **Description**.
8. Dans le champ **Target**, sélectionnez l'option **Selected stage(s)**.
9. Si vous voulez limiter les valeurs qu'un utilisateur peut sélectionner, modifiez les options disponibles dans le champ **Legal values** en cliquant sur l'icône située directement à droite de ce champ.
10. Pour modifier la valeur par défaut, indiquez une valeur différente dans le champ **Valeur par défaut**.

**Remarque :** Pour un service, vous ne pouvez modifier que les valeurs par défaut avant d'exposer le service pour la première fois. Une fois que vous exposez le service, vous ne pouvez plus modifier les valeurs par défaut à l'aide d'Enterprise Designer. Au lieu de cela, vous devez utiliser Management Console. Pour plus d'informations, reportez-vous à la section [Spécification des Options de services par défaut](#).

11. Cliquez sur **OK**.
12. Continuez d'ajouter les options souhaitées.
13. Cliquez sur **OK** dans la boîte de dialogue des **Options du flux de données** lorsque vous avez terminé d'ajouter vos options.
14. Enregistrez et exposez le flux de données.

## Write to Spreadsheet

Write to Spreadsheet écrit les données dans une feuille de calcul Excel comme sortie d'un dataflow.

### Onglet de propriétés du fichier

L'onglet **Propriétés du fichier** contient des options permettant de spécifier la feuille de calcul et les données à écrire depuis un dataflow.

Nom du champ	Description
Server name	Indique que le fichier que vous sélectionnez dans le champ <b>Nom de fichier</b> se trouve sur le serveur Spectrum™ Technology Platform.
Nom de fichier	<p>Indique le chemin du fichier. Cliquez sur le bouton de sélection (...) pour trouver le fichier souhaité.</p> <p>Lors de l'écriture d'un fichier à partir d'un serveur de fichiers HDFS, les formats de compression pris en charge sont les suivants :</p> <ol style="list-style-type: none"> <li>1. GZIP (.gz)</li> <li>2. BZIP2 (.bz2)</li> </ol> <p><b>Remarque :</b> Incluez l'extension appropriée dans le nom de fichier, pour indiquer le format de compression à utiliser pour écrire le fichier.</p> <p><b>Avertissement :</b> Si le Spectrum™ Technology Platform est en cours d'exécution sur les systèmes Unix ou Linux, rappelez-vous que les noms de fichiers et les chemins sur ces plates-formes sont sensibles à la casse.</p>

Nom du champ	Description
Mode Écriture	Indique la manière dont vous souhaitez écrire les données dans la feuille de calcul pour écrire dans le dataflow. Vous pouvez créer une feuille de calcul lors de l'exécution à l'aide de l'une des options suivantes. Les options sont les suivantes : <ul style="list-style-type: none"> <li><b>Créer ou écraser</b> Crée un nouveau fichier et remplace les données existantes du fichier de sortie à chaque exécution du dataflow.</li> <li><b>Insérer</b> Ajoute la sortie du dataflow à la zone cartographiée et déplace les données déjà présentes, le cas échéant, vers le bas.</li> <li><b>Ajouter</b> Ajoute le dataflow de sortie à la fin du fichier, sans effacer les données existantes du fichier</li> </ul>
Sheet name	Indique le nom de la feuille de la feuille de calcul dans laquelle vous souhaitez écrire les données dans le dataflow.
Start writing from	Indique soit une combinaison de ligne-colonne (A1 ou B2), soit une colonne depuis laquelle vous souhaitez que les données soient écrites. Pour l'option <b>Insérer</b> , vous devez fournir une ligne et une colonne. Pour l'option <b>Ajouter</b> , vous ne pouvez fournir qu'une colonne, car cette option ignore la valeur de ligne.
First row as column header	Indique que la première ligne d'un fichier contient des informations d'en-tête, mais aucune donnée.

### Onglet Champs

L'onglet Champs définit les colonnes, les emplacements, les types et les valeurs annulables des champs du fichier. Pour plus d'informations, reportez-vous à la rubrique suivante :

**Définition des champs d'un fichier de sortie** à la page 284

## Définition des champs d'un fichier de sortie

Dans Write to Spreadsheet, l'onglet **Champs** définit les noms, l'emplacement et les types de données des champs du fichier. Après avoir défini un fichier de sortie sur l'onglet **Propriétés de fichier**, vous pouvez définir les champs. Si l'option **Nullable** est cochée et que le champ **Nom** contient une valeur nulle, le flux de données écrit la valeur nulle dans la feuille de calcul.

Si le fichier de sortie contient un enregistrement d'en-tête, vous pouvez rapidement définir les champs en cliquant sur **Régénérer**.

Pour définir les champs avec des valeurs par défaut pour la position, la longueur et le type de données, cliquez sur **Ajout rapide** et sélectionnez les champs à ajouter.

Si le fichier d'entrée ne contient pas d'enregistrement d'en-tête, ou si vous souhaitez définir manuellement les champs, suivez les étapes suivantes :

1. Cliquez sur **Ajouter**.
2. Dans le champ **Name**, choisissez le champ que vous souhaitez ajouter.
3. Dans le champ **Type**, sélectionnez le type de données du champ provenant du flux de données.

Spectrum™ Technology Platform prend en charge les types de données suivants :

**bigdecimal** Type de données numérique prenant en charge 38 points de précision décimaux. Utilisez ce type de données pour les données utilisées dans les calculs mathématiques requérant un haut niveau de précision, en particulier ceux impliquant des données financières. Le type de données bigdecimal prend en charge des calculs plus précis que le type de données double.

**booléen** Type logique doté de deux valeurs : True et False.

**bytearray** Tableau (liste) d'octets.

**Remarque** : ByteArray n'est pas pris en charge en entrée pour un service REST.

**date** Type de données contenant un mois, un jour et une année. Par exemple, 2012-01-30 ou January 30, 2012. Vous pouvez indiquer un format de date par défaut dans Management Console.

**datetime** Type de données contenant un mois, un jour, une année et des heures, des minutes et des secondes. Par exemple, 2012/01/30 6:15 PM.

**double** Un type de données numérique qui contient à la fois des nombres en double précision négatifs et positifs entre  $2^{-1074}$  et  $(2-2^{-52}) \times 2^{1023}$ . En notation E, la plage de valeurs est comprise entre -1.79769313486232E+308 et 1.79769313486232E+308.

**flottant** Un type de données numérique contenant des nombres de précision seuls négatifs et positifs entre  $2^{-149}$  et  $(2-2^{23}) \times 2^{127}$ . En notation E, la plage de valeurs est comprise entre -3.402823E+38 et 3.402823E+38.

**integer** Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre  $-2^{31}$  (-2,147,483,648) et  $2^{31}-1$  (2,147,483,647).

**long** Un type de données numériques qui contient à la fois des nombres entiers négatifs et positifs entre  $-2^{63}$  (-9,223,372,036,854,775,808) et  $2^{63}-1$  (9,223,372,036,854,775,807).

**chaîne** Séquence de caractères.

**time** Type de données contenant l'heure du jour. Par exemple, 21:15:59 ou 9:15:59 PM.

4. Cliquez sur **Ajouter**.

Après avoir défini les champs dans votre fichier de sortie, vous pouvez en modifier le contenu et la disposition.

Nom de l'option	Description
Ajouter	Ajoute un champ dans la sortie. Vous pouvez ajouter un champ à la fin de la disposition existante ou insérer un champ à un emplacement existant et Write to Spreadsheet ajustera les champs restants en conséquence.
Modifier	Modifie le nom et le type du champ.
Enlever	Enlève le champ sélectionné de la sortie.
Déplacer vers le haut/Déplacer vers le bas	Réorganise le champ sélectionné.

## Write to Variable Format File

Write to Variable Format File écrit des enregistrements dont la mise en page varie dans un fichier.

Les fichiers de format variable présentent les caractéristiques suivantes :

- Les enregistrements dans le fichier contiennent des champs différents et des nombres de champs différents.
- Chaque enregistrement doit contenir une balise (généralement un nombre) identifiant le type d'enregistrement.
- Les relations hiérarchiques sont prises en charge.

### Exemple de fichier au format variable

Cet exemple présente un fichier de format variable contenant des informations sur la vérification de l'activité de compte de deux clients, Joe Smith et Anne Johnson. Dans cet exemple, le fichier est un fichier délimité qui utilise la virgule comme délimiteur de champ.

```
001   Joe,Smith,M,100 Main St,555-234-1290
100   CHK12904567,12/2/2007,6/1/2012,CHK
```

```

200 1000567,1/5/2012,Fashion Shoes,323.12
001 Anne,Johnson,F,1202 Lake St,555-222-4932
100 CHK238193875,1/21/2001,4/12/2012,CHK
200 1000232,3/5/2012,Blue Goose Grocery,132.11
200 1000232,3/8/2012,Trailway Bikes,540.00

```

Le premier champ dans chaque enregistrement contient la balise qui identifie le type d'enregistrement et, par conséquent, le format de l'enregistrement :

- 001 : enregistrement de client
- 100 : enregistrement de compte
- 200 : enregistrement de transaction de compte

Pour les fichiers délimités, il est courant que la valeur de balise (001, 100, 200) consiste en un nombre d'octets fixe au début de l'enregistrement, comme indiqué dans l'exemple ci-dessus.

Chaque enregistrement est doté de son propre format :

- 001 : Prénom,Nomdefamille,Sexe,Adresse,NuméroTéléphone
- 100 : IDdeCompte,DateD'ouverture,DateD'expiration,TypeDeCompte
- 200 : IDdeTransaction,DateDeTransaction,Fournisseur,Montant

Le format d'enregistrement 100 (enregistrement de compte) est un enfant de l'enregistrement 001 précédent et le format d'enregistrement 200 (enregistrement de transaction de compte) est un enfant de l'enregistrement 100 précédent (enregistrement de compte). Ainsi, dans le fichier d'exemple, le compte de Joe Smith, CHK12904567, présentait une transaction datant du 05/01/12 d'un montant de 323,12 chez Fashion Shoes. De même, le compte d'Anne Johnson, CHK238193875, présentait deux transactions, l'une du 05/03/12 chez Blue Goose Grocery et l'autre du 08/03/12 chez Trailway Bikes.

### Onglet de propriétés du fichier

Nom de l'option	Description
Server name	Indique si le fichier que vous sélectionnez comme entrée figure sur l'ordinateur exécutant Enterprise Designer ou sur le serveur Spectrum™ Technology Platform. Si vous sélectionnez un fichier sur l'ordinateur local, le nom du serveur sera Mon ordinateur. Si vous sélectionnez un fichier sur le serveur, le nom du serveur sera Spectrum™ Technology Platform.

Nom de l'option	Description
Nom de fichier	<p>Indique le chemin du fichier. Cliquez sur le bouton de sélection (...) pour trouver le fichier souhaité.</p> <p><b>Remarque :</b> Si le Spectrum™ Technology Platform est en cours d'exécution sur les systèmes Unix ou Linux, rappelez-vous que les noms de fichiers et les chemins sur ces plates-formes sont sensibles à la casse.</p>
Nom de la balise racine	Balise à utiliser pour les enregistrements qui représentent la valeur parent d'autres types d'enregistrement. Par exemple, si vous disposez de trois types d'enregistrement 001, 100 et 200, et que les types d'enregistrement 100 et 200 sont les enfants du type d'enregistrement 001, 001 est la balise racine.
Utiliser des balises de largeur fixe	<p>Détermine si une quantité fixe d'espace doit être allouée au début de chaque enregistrement afin d'y placer la balise d'enregistrement. Par exemple, ce qui suit présente un fichier avec les balises 001, 100 et 200 dans un champ de largeur fixe :</p> <pre>001   Joe, Smith, M, 100 Main St, 555-234-1290 100   CHK12904567, 12/2/2007, 6/1/2012, CHK 200   1000567, 1/5/2012, Mike 's Shoes, 323.12</pre>
Largeur de balise	<p>Si vous cochez la case <b>Utiliser des balises de largeur fixe</b>, cette option indique le nombre d'espaces à allouer pour les balises au début de chaque enregistrement. Par exemple, si vous indiquez 7, les sept premières positions dans chaque enregistrement seront réservées pour la balise. La valeur que vous indiquez doit être supérieure ou égale à la taille dans les caractères du nom de balise le plus long. Pour plus d'informations sur les noms de balise, reportez-vous à <a href="#">Noms de balise dans des fichiers de format variable</a> à la page 295.</p> <p>La valeur dans le champ <b>Tag width</b> est automatiquement augmentée si vous ajoutez des champs sous l'onglet <b>Fields</b> dont le nom est plus long que la valeur indiquée.</p> <p>La largeur de balise maximale est de 1 024.</p>
Enlever le préfixe de balise numérique	<p>Enlève la portion « NumericTag_ » du nom de champ avant d'écrire la balise dans le fichier. Le préfixe « NumericTag_ » est ajouté aux noms de balise par le stage Read from Variable Format File pour tout nom de balise commençant par un nombre. En effet, le nom de balise est utilisé comme nom d'un champ de dataflow de liste contenant les données dans l'enregistrement, et les noms de champ de dataflow ne peuvent pas commencer par un nombre. Par exemple, une balise 100 deviendrait un champ de liste nommé « NumericTag_100 ». Si vous activez cette option, ce champ est écrit dans le fichier de sortie en tant qu'enregistrement avec une balise de « 100 » au lieu de « NumericTag_100 ».</p>

Nom de l'option	Description
Encodage des caractères	L'encodage du fichier texte. Sélectionnez l'un des éléments suivants :
<b>UTF-8</b>	Prend en charge tous les caractères Unicode et est rétrocompatible avec le format ASCII. Pour plus d'informations sur UTF, reportez-vous à <a href="https://unicode.org/faq/utf_bom.html">unicode.org/faq/utf_bom.html</a> .
<b>UTF-16</b>	Prend en charge tous les caractères Unicode, mais n'est pas rétrocompatible avec le format ASCII. Pour plus d'informations sur UTF, reportez-vous à <a href="https://unicode.org/faq/utf_bom.html">unicode.org/faq/utf_bom.html</a> .
<b>US-ASCII</b>	L'encodage des caractères en fonction de l'ordre alphabétique anglais.
<b>UTF-16BE</b>	Encodage UTF-16 avec sérialisation d'octets gros-boutiste (octet le plus significatif en premier).
<b>UTF-16LE</b>	Encodage UTF-16 avec sérialisation d'octets petit-boutiste (octet le moins significatif en premier).
<b>ISO-8859-1</b>	Chiffrement de caractères ASCII utilisé habituellement pour les langues européennes occidentales. Également désigné par l'acronyme Latin-1.
<b>ISO-8859-3</b>	Chiffrement de caractères ASCII utilisé habituellement pour les langues européennes méridionales. Également connu sous le nom Latin-3.
<b>ISO-8859-9</b>	Chiffrement de caractères ASCII utilisé habituellement pour la langue turque. Également connu sous le nom Latin-5.
<b>CP850</b>	Une page de code ASCII utilisée pour écrire les langues européennes occidentales.
<b>CP500</b>	Une page de code EBCDIC utilisée pour écrire les langues européennes occidentales.
<b>Shift_JIS</b>	Un encodage de caractères pour la langue japonaise.
<b>MS932</b>	Extension Microsoft de Shift_JIS permettant d'inclure les caractères spéciaux NEC, la sélection NEC des extensions IBM et les extensions IBM.
<b>CP1047</b>	Une page de code EBCDIC avec le jeu de caractères Latin-1 complet.

Nom de l'option	Description
Séparateur de champs	<p>Indique le caractère utilisé pour séparer des champs dans un fichier délimité.</p> <p>Par exemple, cet enregistrement utilise une barre verticale ( ) comme séparateur de champ :</p> <pre>7200 13TH ST   MIAMI   FL   33144</pre> <p>Ces caractères disponibles à définir en tant que séparateurs de champs sont :</p> <ul style="list-style-type: none"> <li>• Espace</li> <li>• Tabulation</li> <li>• Virgule</li> <li>• Point (.)</li> <li>• Point-virgule</li> <li>• Conduite</li> </ul> <p>Si le fichier utilise un caractère différent comme séparateur de champ, cliquez sur le bouton de sélection pour sélectionner un autre caractère délimiteur.</p>
Séparateur de balises	<p>Indique le caractère placé après le champ de balise pour démarquer le champ d'identification pour chaque enregistrement dans un fichier délimité. Un séparateur de balises doit être un caractère unique.</p> <p>Par défaut, les caractères suivants sont disponibles comme séparateurs de balises :</p> <ul style="list-style-type: none"> <li>• Espace</li> <li>• Tabulation</li> <li>• Virgule</li> <li>• Point (.)</li> <li>• Point-virgule</li> <li>• Conduite</li> </ul> <p>Si le fichier utilise un caractère différent comme séparateur de balises, cliquez sur le bouton d'ellipse pour ajouter et sélectionner un séparateur de balises personnalisé.</p> <p><b>Remarque :</b> Par défaut, le caractère <b>Séparateur d'enregistrements</b> est le même que le caractère <b>Séparateur de champs</b> sélectionné. Pour activer ce champ et sélectionner un caractère différent, décochez la case <b>Identique au séparateur de champs</b>.</p>
Identique au séparateur de champs	<p>Indique si le séparateur de balises est le même que le séparateur de champs. Décochez cette case pour sélectionner un caractère différent comme séparateur de balises.</p> <p><b>Remarque :</b> Par défaut, cette case est cochée et le champ <b>Séparateur de balises</b> est désactivé.</p>

Nom de l'option	Description
Qualificateur de texte	<p>Le caractère utilisé pour entourer les valeurs de texte dans un fichier délimité. Par exemple, cet enregistrement utilise des guillemets doubles (") en tant que qualificateur de texte.</p> <pre>"7200 13TH ST"   "MIAMI"   "FL"   "33144"</pre> <p>Ces caractères disponibles à définir en tant que qualificateurs de texte sont :</p> <ul style="list-style-type: none"> <li>• Guillemets simples (')</li> <li>• Guillemets doubles (")</li> </ul> <p>Si le fichier utilise un délimiteur de texte, cliquez sur le bouton de sélection pour sélectionner un autre caractère délimiteur.</p>
Séparateur d'enregistrements	<p>Indique le caractère utilisé pour séparer des enregistrements dans un fichier séquentiel ligne par ligne ou délimité. Ce champ n'est pas disponible si vous activez la case <b>Utiliser EOL par défaut</b>.</p> <p>Les paramètres de séparateur d'enregistrements disponibles sont :</p> <p><b>Unix (U+000A)</b> Un caractère de saut de ligne sépare les enregistrements. Il s'agit du séparateur d'enregistrement standard pour les systèmes Unix.</p> <p><b>Macintosh (U+000D)</b> Un caractère de retour chariot sépare les enregistrements. Il s'agit du séparateur d'enregistrement standard pour les systèmes Macintosh.</p> <p><b>Windows (U+000D U+000A)</b> Un retour chariot suivi d'un saut de ligne sépare les enregistrements. Il s'agit du séparateur d'enregistrement standard pour les systèmes Windows.</p> <p>Si votre fichier utilise un séparateur d'enregistrement, cliquez sur le bouton de sélection pour sélectionner un autre caractère séparateur.</p>
Utiliser EOL par défaut	<p>Indique que le séparateur d'enregistrements du fichier est le caractère de fin de ligne (EOL) par défaut utilisé sur le système d'exploitation sur lequel le serveur Spectrum™ Technology Platform est exécuté.</p> <p>Ne sélectionnez pas cette option si le fichier utilise un caractère EOL qui est différent du caractère EOL par défaut utilisé sur le système d'exploitation du serveur. Par exemple, si le fichier utilise un EOL Windows, mais que le serveur fonctionne sous Linux, ne cochez pas cette option. Au lieu de cela, sélectionnez l'option Windows dans le champ <b>Record separator</b>.</p>

### Onglet Champs

L'onglet **Champs** détermine les champs du dataflow qui sont inclus dans le fichier de sortie.

Nom de l'option	Description
Ajouter	<p>Cliquez sur cette option pour ajouter un champ à la sortie.</p> <p>Pour plus d'informations sur la création de champs de dataflow à utiliser avec Write to Variable Format File, reportez-vous à <a href="#">Writing Flat Data to a Variable Format File</a> à la page 293.</p>
Modifier	<p>Cliquez sur cette option pour modifier le nom de la balise. Ce bouton est activé uniquement lorsqu'une balise est sélectionnée. Si l'option <b>Utiliser des balises de largeur fixe</b> est activée sous l'onglet <b>Propriétés de fichier</b>, la largeur de balise est automatiquement ajustée si vous saisissez un nom de balise plus long.</p> <p><b>Remarque :</b> Modifier le nom de balise racine à l'aide de ce bouton a le même effet que modifier la valeur du champ <b>Nom de la balise racine</b> dans l'onglet <b>Propriétés de fichier</b>.</p>
Supprimer	<p>Enlève le champ sélectionné de la sortie. Si vous enlevez un champ de liste, tous les champs enfant sont également enlevés. Si vous enlevez un champ enfant, seul l'enfant sélectionné est enlevé du champ de liste.</p>
xx Tout supprimer	<p>Enlève tous les champs de la sortie.</p>
Déplacer vers le haut/Déplacer vers le bas	<p>Réorganise le champ sélectionné.</p>

### *L'onglet Exécution*

Nom de l'option	Description
Nom de fichier	<p>Cela affiche le fichier défini dans l'onglet <b>Propriétés de fichier</b>.</p>

Nom de l'option	Description
Générer plusieurs fichiers	<p>Sélectionnez cette option pour écrire des enregistrements dans des fichiers différents, plutôt que d'écrire tous les enregistrements en un fichier. Le fichier dans lequel chaque enregistrement est écrit est indiqué dans l'enregistrement lui-même. Chaque enregistrement doit contenir un champ indiquant soit un nom de fichier, soit le chemin d'accès complet au fichier dans lequel vous souhaitez écrire l'enregistrement. Par exemple, pour envoyer les prix des actions de différentes sociétés (de groupes divers) à tous les clients séparément, cette fonctionnalité écrit les prix des actions des différentes sociétés dans des fichiers distincts qui peuvent être envoyés à chacun des clients, si vous le désirez. Si vous activez l'option <b>Générer plusieurs fichiers</b>, vous devez spécifier un fichier de sortie sur le serveur Spectrum™ Technology Platform ou sur un serveur FTP. Pour écrire les données dans un fichier sur un serveur FTP, vous devez définir une connexion au serveur de fichiers via Management Console.</p> <p><b>Remarque :</b> Les enregistrements de la colonne que vous sélectionnez dans <b>Champ de chemin d'accès au fichier</b> doivent se présenter dans l'ordre trié. Utilisez cette fonctionnalité lorsque l'enregistrement contient un nom de fichier ou le fichier de fichier complet.</p>
Champ de chemin d'accès au fichier	<p>Sélectionne le champ contenant le chemin (un nom de fichier ou le chemin de fichier complet) du fichier dans lequel vous souhaitez écrire l'enregistrement. Seuls les éléments de type simple mappés directement sur une balise racine seront répertoriés dans le champs <b>File path</b>. Ce champ n'est activé que si vous sélectionnez l'option <b>Générer plusieurs fichiers</b>.</p>
Mode Écriture	<p>Indique s'il faut ajouter la sortie du dataflow à la fin du fichier ou supprimer les données existantes dans le fichier avant écriture de la sortie. L'un des éléments suivants :</p> <p><b>Écraser</b>                      Remplace les données existantes dans le fichier de sortie chaque fois que le dataflow est exécuté.</p> <p><b>Ajouter</b>                        Ajoute le dataflow de sortie à la fin du fichier, sans effacer les données existantes du fichier</p>

## Writing Flat Data to a Variable Format File

Dans un flux de données Spectrum™ Technology Platform, chaque enregistrement possède les mêmes champs. Toutefois, dans un fichier de format variable, les enregistrements ne contiennent pas tous les mêmes champs. Pour écrire des données texte issues d'un flux de données dans un fichier de format variable, vous devez diviser chaque enregistrement dans le flux de données, en regroupant les champs de chaque enregistrement dans les champs de liste correspondant aux types d'enregistrement à utiliser pour le fichier de format variable. Un champ de liste consiste en un ensemble de champs. Par exemple, les champs FirstName, LastName, Gender, Address et Phone peuvent être regroupés dans un champ de liste appelé AccountOwner.

Pour écrire des données plates dans un fichier de format variable, utilisez un stage Aggregator pour regrouper les champs dans les champs de liste correspondant aux types d'enregistrement à écrire dans le fichier de format variable. Pour ce faire, procédez comme suit :

1. Placez un stage Aggregator dans votre flux de données quelque part avant le stage Write to Variable Format File.
2. Double-cliquez sur le stage Aggregator pour ouvrir sa fenêtre d'options.
3. Sélectionnez **Grouper par**, puis cliquez sur **Ajouter**.
4. Dans le champ **Group By**, sélectionnez le champ qui contient un identifiant unique permettant d'identifier les données associées. La valeur de ce champ doit être unique dans les enregistrements des données texte. Par exemple, un numéro de compte, un numéro de sécurité sociale ou un numéro de téléphone.

**Remarque :** Le champ que vous sélectionnez doit être trié. Si ce n'est pas le cas, utilisez un stage Sorter pour trier les enregistrements par champ.

5. Cliquez sur **OK**.
6. Sélectionnez **Listes de sortie**, puis cliquez sur **Ajouter**.  
Chaque liste de sortie représentera un type d'enregistrement dans le fichier de format variable.
7. Sélectionnez **New data type** et dans le champ **Type name**, indiquez le type d'informations que contiendra le type de données. Cette valeur deviendra un type d'enregistrement dans le fichier de format variable. Par exemple, ce type de données contiendra des enregistrements associés aux transactions de compte ; vous pouvez donc nommer le type « AccountTransaction ».
8. Dans le champ **Name**, saisissez le nom à attribuer au champ. Il peut s'agir du même nom que celui que vous avez indiqué dans le champ **Nom du type**.
9. Cliquez sur **OK**.
10. Sélectionnez le type de données que vous venez de créer et cliquez sur **Ajouter**.
11. Laissez l'option **Champ existant** sélectionnée et sélectionnez un des champs à inclure dans ce type de données, puis cliquez sur **OK**. N'oubliez pas que cette valeur deviendra un type d'enregistrement dans le fichier de format variable. Répétez l'opération pour ajouter des champs supplémentaires au type d'enregistrement.
12. Créez des listes de sortie supplémentaires pour chaque type d'enregistrement dont vous souhaitez disposer dans le fichier de format variable. Lorsque vous avez terminé, cliquez sur **OK** pour fermer les options Aggregator.

Les champs renvoyés par le stage Aggregator sont désormais regroupés en champs de liste qui correspondent aux types d'enregistrement à inclure dans la sortie de fichier de format variable.

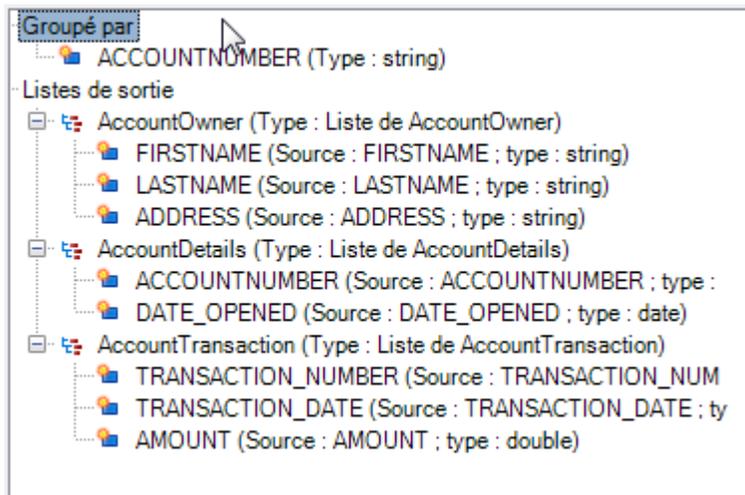
Par exemple, si l'on considère les données texte suivantes :

```
FIRSTNAME, LASTNAME, ADDRESS, ACCOUNTNUMBER, DATE_OPENED, TRANSACTION_NUMBER, TRANSACTION_DATE, AMOUNT
Joe, Smith, 100 Main St, CHK12904567, 12/2/2007, 1000567, 1/5/2012, 323.12
```

Il serait judicieux de les convertir en quelque chose de similaire dans le fichier de format variable :

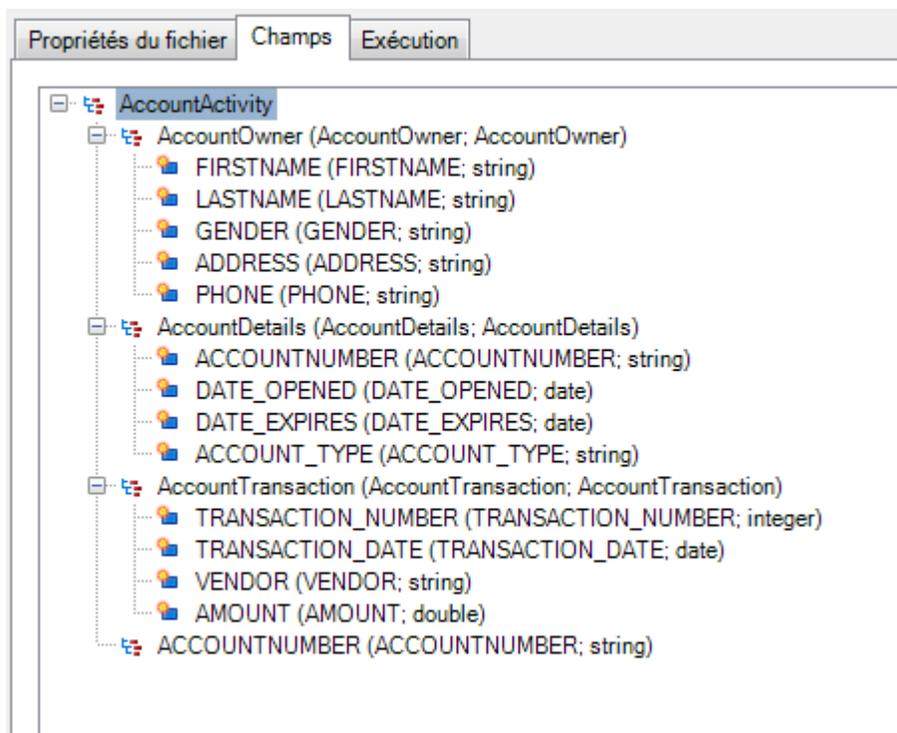
```
AccountOwner      Joe,Smith,100 Main St
AccountInformation CHK12904567,12/2/2007
Transaction       1000567,1/5/2012,323.12
```

Pour ce faire, vous devez créer un stage Aggregator configuré comme suit :



## Noms de balise dans des fichiers de format variable

Dans un fichier de format variable, chaque enregistrement dans le fichier de sortie possède une balise qui indique le type d'enregistrement. Dans Write To Variable Format File, le nom de champ est utilisé comme nom de balise dans le fichier de sortie. Par exemple, envisagez les champs suivants :



Ces champs doivent être écrits dans le fichier comme suit. Dans cet exemple, le compte possède deux enregistrements AccountTransaction.

```
AccountOwner      Anne,Johnson,F,1202 Lake St,555-222-4932
AccountDetails    CHK238193875,1/21/2001,4/12/2012,CHK
AccountTransaction 1000232,3/5/2012,Blue Goose Grocery,132.11
AccountTransaction 1000232,3/8/2012,Trailway Bikes,540.00
```

**Remarque :** Seuls les champs de liste contenant des champs simples, tels que des chaînes, sont écrits dans le fichier de sortie. Si un champ de liste se compose uniquement d'autres champs de liste, il n'est pas écrit dans le fichier de sortie. Dans l'exemple ci-dessus, aucun enregistrement doté d'une balise AccountActivity ne serait écrit dans le fichier de sortie car la balise AccountActivity n'est composée que d'autres champs de liste (AccountOwner, AccountDetails, and AccountTransaction).

## Write to XML

Le stage Write to XML écrit la sortie d'un travail ou d'un sous-flux sur un fichier XML.

### Onglet de propriétés du fichier

Nom du champ	Description
Fichier de données	Indique le chemin du fichier XML de sortie. Cliquez sur le bouton de sélection (...) pour trouver le fichier souhaité.  <b>Remarque</b> : Si le Spectrum™ Technology Platform est en cours d'exécution sur les systèmes Unix ou Linux, rappelez-vous que les noms de fichiers et les chemins sur ces plates-formes sont sensibles à la casse.
Fichier réel	Affiche la structure indiquée dans l'onglet <b>Champs</b> . Si vous cliquez sur un élément et que le fichier indiqué dans le champ <b>Data file</b> contient l'élément, un aperçu des données sera affiché. Seules les données d'éléments simples peuvent être affichées dans l'aperçu.
Exporter un schéma	Cliquez sur ce bouton pour enregistrer un fichier XSD qui représente le schéma affiché dans la vue Fichier Actuel. Le fichier de schéma est immédiatement enregistré dans l'emplacement que vous indiquez.

### Onglet Champs

L'onglet **Champs** définit les champs à inclure dans le fichier XML de sortie. Lorsque vous ajoutez des champs, ils sont affichés dans une structure d'arborescence. L'arborescence affiche le nom de l'élément ou de l'attribut qui sera écrit dans le fichier XML. Le nom du champ de flux de données, suivi par le type de données, figure entre parenthèses à la suite du nom d'élément/attribut. Par exemple :



Cela indique que quatre éléments et un attribut doivent être écrits dans le fichier XML. L'attribut est indiqué par le signe « @ » en rouge.

L'élément État contient les données du champ StateProvince de type chaîne. De même, l'élément ZIP contient les données du champ PostalCode de type chaîne. Le fichier XML peut être semblable à ce qui suit :

```

<XmlRoot>
  <Customer Status="0">
    <AddressLine1>7713 Mullen Dr</AddressLine1>

```

```
<City>Austin</City>
<State>TX</State>
<ZIP>78757-1346</ZIP>
</Customer>
<Customer Status="0">
  <AddressLine1>1825B Kramer Ln</AddressLine1>
  <City>Austin</City>
  <State>TX</State>
  <ZIP>78758-4260</ZIP>
</Customer>
</XmlRoot>
```

**Remarque :** Le nom de l'élément racine (dans cet exemple, <XmlRoot>) est indiqué sous l'onglet **Propriétés du fichier**.

La table suivante décrit les options sous l'onglet **Champs**.

Nom de l'option	Description
Ajouter	Ajoute un champ dans la sortie.

Nom de l'option

Description

---

Modifier

Nom de l'option	Description
	Modifie la manière dont le champ est écrit dans XML. Vous pouvez indiquer les options suivantes :
<b>Type de sortie</b>	Cette option est disponible si vous modifiez un champ simple. Cela indique si le champ du flux de données doit être écrit dans un élément ou attribut XML.
<b>Élément</b>	Sélectionnez cette option pour écrire les données du champ dans un élément XML. Indiquez le nom de l'élément à utiliser dans le champ <b>Element name</b> .
<b>Attribut</b>	Écrit les données du champ sur un attribut de l'élément parent. Indiquez le nom de l'attribut à utiliser dans le champ <b>Attribute name</b> .
<b>Nom de l'élément/Nom de l'attribut</b>	Indique le nom de l'élément ou de l'attribut à écrire dans le fichier XML. Le nom par défaut est le nom de champ du flux de données.
<b>Modifier tous les enfants en</b>	Cette option est disponible si vous modifiez un élément complexe. Elle spécifie le type de données XML que l'élément complexe doit contenir. L'un des éléments suivants :
<b>Aucune modification</b>	Les types enfant restent tels qu'ils sont actuellement définis, à savoir élément ou attribut. Vous pouvez indiquer le type pour chaque champ individuellement en sélectionnant le champ et en cliquant sur <b>Modifier</b> .
<b>Éléments</b>	Tous les champs simples sous l'élément sont écrits en tant qu'éléments XML.
<b>Attributs</b>	Tous les champs simples sous l'élément sont écrits en tant qu'attributs XML.
<b>Espace de nomns;</b>	Si vous souhaitez spécifier un espace de nommage XML à utiliser pour l'élément ou l'attribut, sélectionnez-le ici. Vous pouvez créer des espaces de nommage dans l'onglet <b>Champs</b> du stage Write to XML.
<b>Inclure les champs vides</b>	Cochez cette case pour inclure dans le fichier de sortie des éléments XML ayant une valeur nulle ou aucune donnée. Si vous ne cochez pas cette case, les éléments vides ne seront pas inclus dans la sortie.
	Par exemple, si vous définissez un élément nommé <City>, mais qu'un enregistrement ne dispose d'aucune donnée dans le champ City, la sortie XML contiendra les éléments suivants si vous cochez <b>Inclure les champs vides</b> :
	<pre>&lt;City xs:nil="true"&gt;&lt;/City&gt;</pre>

Nom de l'option	Description
	<p data-bbox="743 338 1429 411">Si vous ne cochez pas cette case, l'élément &lt;City&gt; ne sera pas écrit dans le fichier de sortie.</p> <p data-bbox="639 443 1429 562"><b>Remarque :</b> Le <b>champ Flux de données</b> affiche le champ dont les données seront écrites dans l'élément ou l'attribut. Ceci apparaît pour que, lorsque vous modifiez le nom d'élément ou d'attribut, vous puissiez toujours voir quelles données de champs sont contenues dans l'élément ou l'attribut.</p>
Supprimer	Enlève le champ sélectionné de la sortie. Si vous enlevez un champ de liste, tous les champs enfant sont également enlevés. Si vous enlevez un champ enfant, seul l'enfant sélectionné est enlevé du champ de liste.
Tout supprimer	Enlève tous les champs de la sortie.
Déplacer vers le haut/Déplacer vers le bas	<p data-bbox="553 926 915 953">Réorganise le champ sélectionné.</p> <p data-bbox="553 972 1429 1150">Vous ne pouvez pas déplacer des éléments simples dans des éléments complexes. Si vous voulez modifier les éléments dans un élément complexe, vous devez modifier le stage Aggregator de votre flux de données de façon à inclure les champs du flux de données dont vous souhaitez disposer dans l'élément complexe. Pour plus d'informations, reportez-vous à la section <a href="#">Création de fichier XML complexe à partir de données texte</a> à la page 303.</p>
Régénérer	Remplace les champs actuellement définis par les champs arrivant dans Write to XML depuis le canal en amont.

## L'onglet Exécution

Nom de l'option	Description
Générer plusieurs fichiers	<p>Sélectionnez cette option pour écrire des enregistrements dans des fichiers différents, plutôt que d'écrire tous les enregistrements en un fichier. Le fichier dans lequel chaque enregistrement est écrit est indiqué dans l'enregistrement lui-même. Chaque enregistrement doit contenir un champ indiquant soit un nom de fichier, soit le chemin d'accès complet au fichier dans lequel vous souhaitez écrire l'enregistrement. Par exemple, pour envoyer les prix des actions de différentes sociétés (de groupes divers) à tous les clients séparément, cette fonctionnalité écrit les prix des actions des différentes sociétés dans des fichiers distincts qui peuvent être envoyés à chacun des clients, si vous le désirez. Si vous activez l'option <b>Générer plusieurs fichiers</b>, vous devez spécifier un fichier de sortie sur le serveur Spectrum™ Technology Platform ou sur un serveur FTP. Pour écrire les données dans un fichier sur un serveur FTP, vous devez définir une connexion au serveur de fichiers via Management Console.</p> <p><b>Remarque :</b> Les enregistrements de la colonne que vous sélectionnez dans <b>Champ de chemin d'accès au fichier</b> doivent se présenter dans l'ordre trié. Utilisez cette fonctionnalité lorsque l'enregistrement contient un nom de fichier ou le chemin de fichier complet.</p>
Champ de chemin d'accès au fichier	<p>Sélectionne le champ contenant le chemin (un nom de fichier ou le chemin de fichier complet) du fichier dans lequel vous souhaitez écrire l'enregistrement. Seuls les éléments de type simple mappés directement sur une racine seront répertoriés dans le champ <b>File path</b>. Ce champ n'est activé que si vous sélectionnez l'option <b>Générer plusieurs fichiers</b>.</p>
Générer un schéma à l'exécution	<p>Sélectionnez cette option pour générer un XSD lors de l'exécution et insérez une référence <code>noNamespaceSchemaLocation</code> au schéma dans le fichier XML. La valeur d'attribut <code>noNamespaceSchemaLocation</code> est le nom du fichier XSD du fichier XML. Si vous exportez le schéma lors de l'édition d'un flux de données, il n'existe aucune référence au fichier XSD dans le fichier de sortie XML et l'utilisateur doit ajouter manuellement la référence au fichier XSD.</p>
Chemin d'accès au schéma	<p>Indique le chemin d'accès pour l'enregistrement du fichier XSD contenant le schéma du fichier XML de sortie. Cliquez sur le bouton de sélection (...) pour trouver le fichier souhaité. Le fichier de sortie est enregistré à l'endroit que vous indiquez lorsque vous exécutez le flux de données.</p>

## Utilisation d'espaces de nommage dans un fichier de sortie XML

Les espaces de nommage vous permettent de disposer de noms d'élément et d'attribut en double dans votre sortie en assignant chaque élément ou attribut à un espace de nommage XML.

1. Dans Enterprise Designer, ouvrez le flux de données.
2. Double-cliquez sur le stage Write to XML sur le canevas.

3. Cliquez sur l'onglet **Champs**.
4. Définissez au moins un espace de nommage :
  - a) Dans la colonne **Préfixe**, entrez le préfixe à utiliser pour associer un élément ou attribut à l'espace de nommage.
  - b) Dans la colonne **Espace de nommage**, indiquez l'URL de l'espace de nommage.
  - c) Répétez cette procédure pour définir autant d'espaces de nommage que vous souhaitez utiliser pour le fichier XML de sortie.
5. Associez au moins un élément ou attribut à l'espace de nommage.
  - a) Dans l'onglet **Champs**, sélectionnez l'élément ou l'attribut à associer à l'espace de nommage, puis cliquez sur **Modifier**, ou créez un élément ou attribut en cliquant sur **Ajouter**.
  - b) Dans le champ **Namespace**, choisissez le préfixe de l'espace de nommage à associer à l'élément ou à l'attribut.
  - c) Cliquez sur **OK**.

## Création de fichier XML complexe à partir de données texte

Les flux de données produisent souvent des enregistrements contenant des champs texte écrits dans le fichier XML en tant qu'éléments XML simples. Si vous souhaitez organiser des champs texte dans les éléments XML complexes pour produire des données hiérarchiques, vous pouvez utiliser un ou plusieurs stages Aggregator.

Par exemple, si l'on considère ces données texte où la première ligne est un enregistrement d'en-tête :

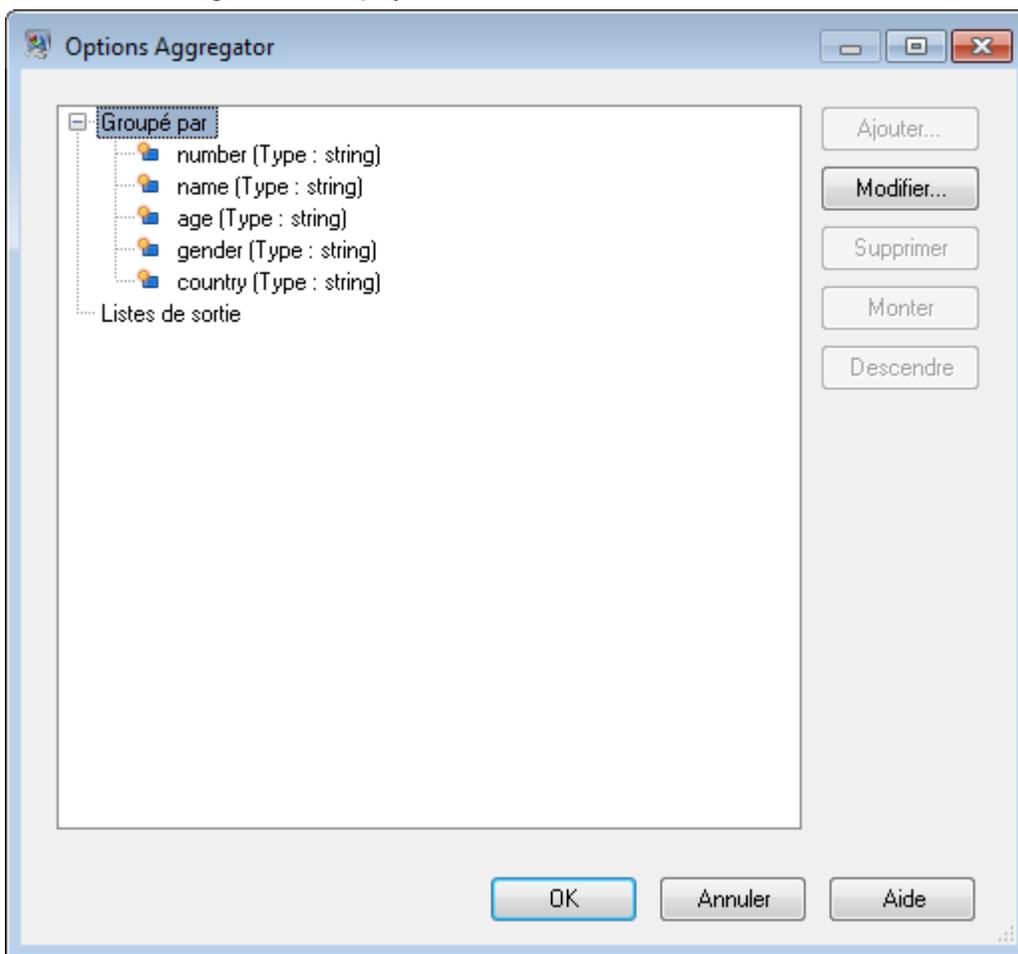
```
addressline1,age,city,country,gender,name,number,postalcode,stateprovince,type
1253 Summer St.,43,Boston,United States,M,Sam,019922,02110,MA,Savings
```

Il serait judicieux de regrouper les champs de données associés à l'adresse et les champs associés au compte dans des éléments XML complexes nommés `<Address>` et `<Account>` comme suit :

```
<CustomerRecord>
  <name>Sam</name>
  <age>43</age>
  <gender>M</gender>
  <country>United States</country>
  <Address>
    <addressline1>1253 Summer St.</addressline1>
    <city>Boston</city>
    <stateprovince>MA</stateprovince>
    <postalcode>02110</postalcode>
  </Address>
  <Account>
    <number>019922</number>
    <type>Savings</type>
  </Account>
</CustomerRecord>
```

1. Ajoutez un stage Aggregator au point dans le flux de données dans lequel vous souhaitez créer des éléments complexes.
2. Double-cliquez sur le stage Aggregator pour ouvrir les options correspondantes.
3. Sélectionnez **Grouper par** et cliquez sur **Ajouter**.
4. Sélectionnez le champ contenant une valeur unique pour chaque enregistrement, telle qu'un numéro de compte, et cliquez sur **OK**.
5. Si vous souhaitez transmettre d'autres champs simples, sélectionnez **Grouper par** et cliquez de nouveau sur **Ajouter** pour ajouter tous les champs simples à inclure.

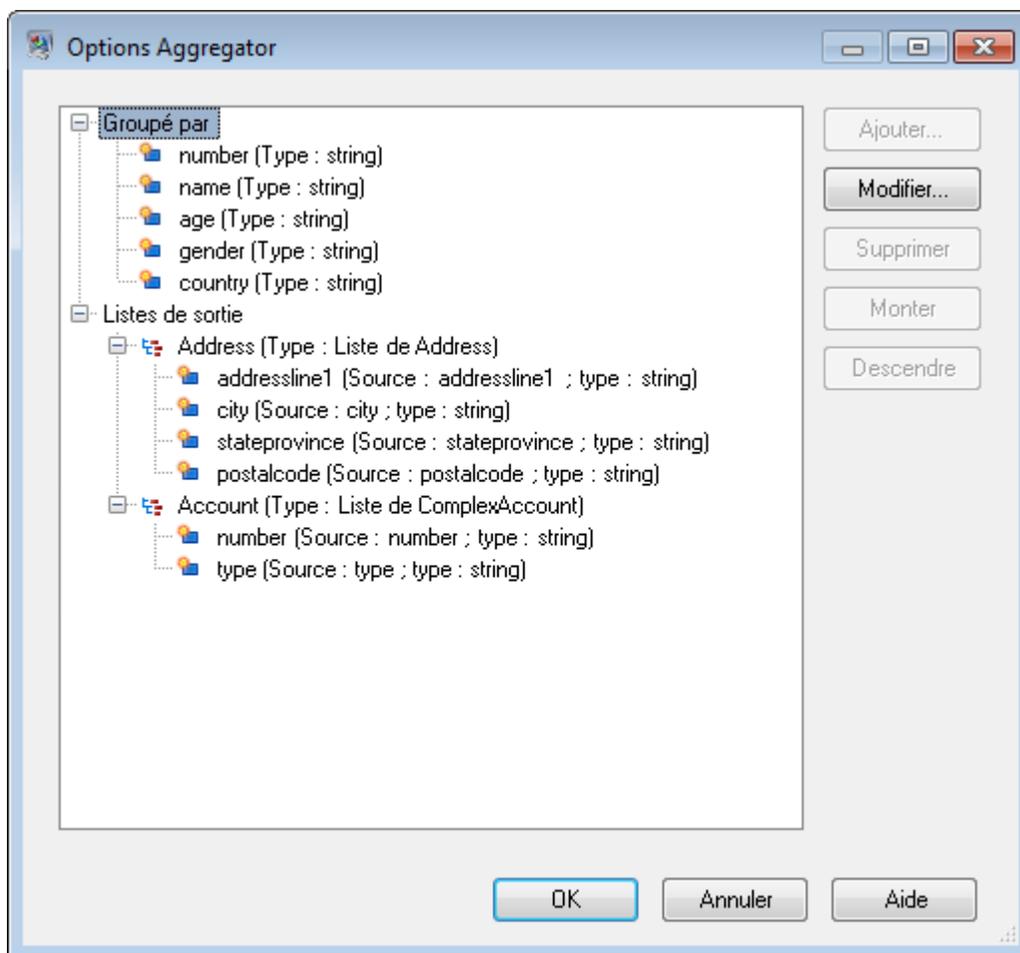
Par exemple, dans ce cas, cinq champs simples seront inclus dans chaque enregistrement : nombre, nom, âge, sexe et pays.



6. Sélectionnez **Listes de sortie** et cliquez sur **Ajouter**.
7. Sélectionnez **Nouveau type de données**. Ceci aura pour effet de définir un nouvel élément complexe. Saisissez une description du genre de données que contiendra cet élément complexe. Par exemple, vous pouvez saisir « Complexe », puisque vous créez un élément XML complexe. Vous pouvez définir librement le nom du type de données.

8. Dans le champ **Name**, saisissez le nom à utiliser pour le champ. Il s'agira également du nom de l'élément XML.
9. Cliquez sur **OK**.
10. Sélectionnez le champ que vous venez de créer et cliquez sur **Ajouter**.
11. Avec **Champ existant** sélectionné, choisissez un champ à ajouter comme champ enfant à l'élément complexe, puis cliquez sur **OK**.
12. Répétez les deux étapes précédentes pour ajouter des champs supplémentaires à l'élément complexe.
13. Ajoutez d'autres champs complexes si nécessaire.

Lorsque vous avez terminé, le stage Aggregator doit répertorier chaque champ simple et complexe à inclure dans chaque enregistrement. Par exemple :



14. Cliquez sur **OK**.

# Modèles de date et de nombre

## Modèles de date et d'heure

Lors de la définition des options de type de données pour les données de date et d'heure, vous pouvez créer votre propre modèle de date ou d'heure personnalisé si les modèles prédéfinis ne répondent pas à vos besoins. Pour créer un modèle de date ou d'heure, utilisez la notation décrite dans la table suivante. Par exemple, ce modèle :

dd MMMM yyyy

Produirait une date comme suit :

14 décembre 2012

Lettre	Description	Exemple
G	Code de l'ère	AD
yy	Année à deux chiffres	96
yyyy	Année à quatre chiffres	1996
M	Mois numérique de l'année.	7
MM	Mois numérique de l'année. Si le nombre est inférieur à 10, un zéro est ajouté pour qu'il s'agisse d'un nombre à deux chiffres.	07
MMM	Nom abrégé du mois	Jul
MMMM	Version longue du nom du mois	Juillet
w	Semaine de l'année	27
ww	Semaine de l'année à deux chiffres. Si la semaine est inférieure à 10, un zéro supplémentaire est ajouté.	06

Lettre	Description	Exemple
W	Semaine du mois	2
D	Jour de l'année	189
DDD	Jour de l'année à trois chiffres. Si le numéro contient moins de trois chiffres, des zéros sont ajoutés.	006
d	Jour du mois	10
dd	Jour du mois à deux chiffres. Les nombres inférieurs à 10 ont un zéro ajouté.	09
F	Jours de la semaine dans le mois	2
E	Nom abrégé du jour de la semaine	Ma
EEEE	Version longue du nom du jour de la semaine	Mardi
a	Marque AM/PM	PM
H	Heure du jour, avec la première heure étant 0 et la dernière heure étant 23.	0
HH	Heure du jour à deux chiffres, la première heure étant 0 et la dernière heure étant 23. Les nombres inférieurs à 10 ont un zéro ajouté.	08
k	Heure du jour, avec la première heure étant 1 et la dernière heure étant 24.	24
kk	Heure du jour à deux chiffres, la première heure étant 1 et la dernière heure étant 24. Les nombres inférieurs à 10 ont un zéro ajouté.	02
k	Heure du matin (AM) ou de l'après-midi (PM), avec 0 étant la première heure et 11 étant la dernière heure.	0
KK	Heure du jour à deux chiffres, la première heure étant 1 et la dernière heure étant 24. Les nombres inférieurs à 10 ont un zéro ajouté.	02

Lettre	Description	Exemple
h	Heure du matin (AM) ou de l'après-midi (PM), 1 étant la première heure et 12 étant la dernière heure.	12
hh	Heure à deux chiffres du matin (AM) ou de l'après-midi (PM), 1 étant la première heure et 12 étant la dernière heure. Les nombres inférieurs à 10 ont un zéro ajouté.	09
m	Minute de l'heure	30
mm	Minutes de l'heure à deux chiffres. Les nombres inférieurs à 10 ont un zéro ajouté.	05
s	Seconde de la minute	55
ss	Secondes de la minute à deux chiffres. Les nombres inférieurs à 10 ont un zéro ajouté.	02
S	Milliseconde de la seconde	978
SSS	Millisecondes de la seconde à trois chiffres. Les nombres contenant moins de trois chiffres ont un ou deux zéros ajoutés pour qu'il s'agisse de nombres à trois chiffres.	978 078 008
z	Abréviation de l'heure du nom de fuseau horaire. Si le fuseau horaire ne porte aucun nom, décalage GMT.	PST GMT-08:00
zzzz	Nom complet du fuseau horaire. Si le fuseau horaire ne porte aucun nom, décalage GMT.	Pacific Standard Time GMT-08:00
Z	Fuseau horaire RFC 822.	-0800
X	Fuseau horaire ISO 8601.	-08Z
XX	Fuseau horaire ISO 8601 avec les minutes.	-0800Z
XXX	Fuseau horaire ISO 8601 avec les minutes et deux-points comme séparateur entre les heures et les minutes.	-08:00Z

## Modèles numériques

Lors de la définition des options de type de données pour les données numériques, vous pouvez créer votre propre modèle numérique personnalisé si les modèles prédéfinis ne répondent pas à vos besoins. Un modèle numérique de base se compose des éléments suivants :

- Un préfixe tel qu'un symbole de devise (facultatif)
- Un modèle de nombres contenant un caractère de regroupement facultatif (par exemple, une virgule comme séparateur de milliers)
- Un suffixe (facultatif)

Par exemple, ce modèle :

\$ ###,###.00

Produirait un nombre formaté comme suit (remarquez l'utilisation d'un séparateur de milliers après les trois premiers chiffres) :

232 998,60 \$

### *Modèles pour les nombres négatifs*

Par défaut, les nombres négatifs sont formatés de la même manière que les nombres positifs mais présentent le signe négatif en préfixe. Le caractère utilisé pour le signe numérique dépend des paramètres régionaux. Dans la plupart des paramètres régionaux, le signe négatif est « - ». Par exemple, si vous spécifiez ce modèle numérique :

0.00

Le nombre négatif dix serait formaté comme suit dans la plupart des paramètres régionaux :

-10.00

Néanmoins, si vous souhaitez définir un autre préfixe ou suffixe à utiliser pour les nombres négatifs, indiquez un second modèle, en le séparant du premier modèle à l'aide d'un point virgule (« ; »). Par exemple :

0.00; (0.00)

Dans ce modèle, les nombres négatifs seraient contenus entre parenthèses :

(10.00)

### *Notation scientifique*

Pour formater un nombre dans la notation scientifique, utilisez le caractère `E` suivi d'un nombre minimal de chiffres à inclure dans l'exposant. Par exemple, dans le modèle suivant :

0.###E0

Le nombre 1234 serait formaté comme suit :

1.234E3

En d'autres termes,  $1.234 \times 10^3$ .

Remarques :

- Le nombre de chiffres après le caractère d'exposant indique le nombre minimal de chiffres exponentiels. Il n'existe pas de valeur maximale.
- Les exposants négatifs sont formatés à l'aide du signe moins localisé, non du préfixe ou du suffixe du modèle.
- Les modèles de notation scientifique ne peuvent pas contenir de séparateurs de regroupement (par exemple, un séparateur de milliers).

### Caractères de modèle numérique spéciaux

Les caractères suivants permettent de produire d'autres caractères, plutôt que d'être reproduits littéralement dans le nombre résultant. Si vous souhaitez utiliser un de ces caractères spéciaux comme caractères littéraux dans le préfixe ou suffixe de votre modèle numérique, entourez le caractère spécial de guillemets.

Symbole	Description
0	Représente un chiffre dans le modèle incluant des zéros où cela est nécessaire pour remplir le modèle. Par exemple, lorsqu'il est appliqué au modèle, le nombre vingt-sept 0000 serait : 0027
#	Représente un chiffre mais les zéros sont omis. Par exemple, lorsqu'il est appliqué au modèle, le nombre vingt-sept #### serait : 27
.	Le séparateur décimal ou le séparateur décimal monétaire utilisé dans les paramètres régionaux sélectionnés. Par exemple, aux États Unis, le point (.) est utilisé comme séparateur décimal mais, en France, la virgule (,) est utilisée comme séparateur décimal.
-	Le signe négatif utilisé dans les paramètres régionaux sélectionnés. Pour la plupart des paramètres régionaux, il s'agit du signe moins (-).

Symbole	Description
,	<p>Le caractère de regroupement utilisé dans les paramètres régionaux sélectionnés. Le caractère approprié pour les paramètres régionaux sélectionnés sera utilisé. Par exemple, aux États-Unis, la virgule (,) est utilisée comme séparateur.</p> <p>Le séparateur de regroupement est généralement utilisé pour les milliers mais, dans certains pays, il sépare les dizaines de milliers. La taille de regroupement est un nombre constant de chiffres entre les caractères de regroupement, tel que 3 pour 100,000,000 ou 4 pour 1,0000,0000. Si vous fournissez un modèle avec plusieurs caractères de regroupement, l'intervalle entre le dernier et la fin de l'entier est celui utilisé. Par exemple, tous les modèles suivants produisent le même résultat :</p> <pre data-bbox="553 638 743 751"># , ## , ### , #### ##### , ##### ## , #### , #####</pre>
E	<p>Sépare la mantisse et l'exposant dans la notation scientifique. Vous ne devez pas entourer le E de guillemets dans votre modèle. Reportez-vous à la section <a href="#">Notation scientifique</a> à la page 309.</p>
;	<p>Sépare les sous-modèles positifs et négatifs. Reportez-vous à la section <a href="#">Modèles pour les nombres négatifs</a> à la page 309.</p>
%	<p>Multipliez le nombre par 100 et affichez le nombre sous forme de pourcentage. Par exemple, le nombre .35, lorsqu'il est appliqué à ce modèle,</p> <pre data-bbox="553 1213 602 1234">##%</pre> <p>produirait le résultat suivant :</p> <pre data-bbox="553 1304 610 1325">35%</pre>
¤	<p>Le symbole de devise pour les paramètres régionaux sélectionnés. S'il est doublé, le symbole de devise international est utilisé. S'il est présent dans un modèle, le séparateur décimal monétaire est utilisé au lieu du séparateur décimal.</p>
'	<p>Utilisé pour mettre des caractères spéciaux entre guillemets dans un préfixe ou un suffixe. Par exemple :</p> <pre data-bbox="553 1654 643 1675">" '# '# "</pre> <p>Formate 123 en :</p> <pre data-bbox="553 1745 643 1766">"#123"</pre> <p>Pour créer un guillemet simple, utilisez-en deux à la suite :</p> <pre data-bbox="553 1835 727 1856">"# o' 'clock"</pre>

# 6 - Configurations

## In this section

---

Configurations de LogMiner Oracle

313

## Configurations de LogMiner Oracle

Oracle LogMiner est un utilitaire d'arrière-plan qui permet à Spectrum d'interroger des journaux créés pour une base de données Oracle et d'y accéder.

L'utilitaire permet à Spectrum™ de lire les journaux de la source de données Oracle pour suivre les modifications apportées aux colonnes de ses tables, dans le cadre du stage **DB Change Data Reader**.

### *Types de données et attributs de stockage de table non pris en charge*

Oracle LogMiner ne prend pas en charge les attributs de stockage de table et les types de données suivants :

- Type de données **BFILE**
- Types de données abstraits simples et imbriqués (**ADT**)
- Collections (tables et VARRAY imbriqués)
- Références d'objet
- Tables sur lesquelles la compression a été activée
- SecureFiles

### *Bases de données et versions de fichier journal Rétablir prises en charge*

LogMiner est exécuté sur des bases de données Oracle de version 8.1 ou ultérieure.

Vous pouvez également utiliser LogMiner pour analyser les fichiers journaux Rétablir des bases de données Oracle 8.0. Cependant, les informations extraites dépendent de la version du journal et non de la version de la base de données utilisée.

Par exemple, pour utiliser LogMiner de manière optimale, il est possible d'agrandir les fichiers journaux Rétablir d'Oracle9i pour qu'ils capturent des informations supplémentaires lorsque la journalisation complémentaire est activée. Les fichiers journaux Rétablir créés avec des versions antérieures d'Oracle n'ont pas de données supplémentaires et peuvent donc présenter des limites en ce qui concerne les opérations et les types de données pris en charge par LogMiner.

### *Restrictions de SQL\*Loader*

Spectrum CDC peut capturer des données qui ont été chargées dans des tables Oracle par l'utilitaire SQL\*Loader. Cependant, les restrictions suivantes s'appliquent :

1. Le chargement de données doit s'effectuer via le *chemin d'accès conventionnel*. Spectrum CDC ne peut pas capturer des données chargées via un *chargement par chemin d'accès direct*, car Oracle LogMiner ne prend pas en charge les chargements par chemin d'accès direct.
2. La méthode de chargement doit être INSERT, APPEND ou REPLACE.

La méthode TRUNCATE n'est pas prise en charge, car la commande TRUNCATE pousse SQL\*Loader à émettre la commande TRUNCATE TABLE DDL. Étant donné que la fonction

Spectrum CDC ne capture pas le DDL mentionné, les suppressions de lignes entraînées par l'utilisation de la commande TRUNCATE TABLE DDL ne sont pas capturées.

### Privilèges utilisateur requis

Le tableau suivant identifie les privilèges système minimaux dont les utilisateurs d'Oracle CDC doivent disposer :

Privilège système	Version d'Oracle
ALTER ANY TABLE	ALL
CONNECT	ALL
LOCK ANY TABLE	ALL
SELECT ANY TRANSACTION	10 g ou version ultérieure

Le tableau suivant identifie les privilèges objet minimaux dont les utilisateurs d'Oracle CDC doivent disposer :

Nom d'objet	Privilège
Tables sources	LOCK ANY TABLE OR SELECT
PUBLIC.V\$DATABASE	SELECT
PUBLIC.V\$LOGMNR_CONTENTS	SELECT
SYS.DBMS_LOGMNR	EXECUTE
SYS.DBMS_LOGMNR_D	EXECUTE

Pour plus d'informations sur Oracle LogMiner, cliquez [ici](#).

# 7 - Optimisation des performances

## In this section

---

Définition d'une taille d'extraction optimale

316

## Définition d'une taille d'extraction optimale

Dans le stage **Read from DB**, la taille d'extraction optimale est calculée en prenant les relevés de temps d'exécution entre un stage **Read from DB** et un stage **Write to Null**.

Assurez-vous de tester les temps d'exécution du job de test avec différentes valeurs de taille d'extraction en utilisant votre propre application avec le module Spectrum™ Data Integration.

1. Créez un job dans Enterprise Designer.
2. Glissez un stage **Read from DB** sur le canevas.
3. Glissez un stage **Write to Null** sur le canevas.
4. Créez un canal entre les deux stages.
5. Double-cliquez sur le stage **Read from DB** pour le configurer de sorte qu'il lise les données d'une table contenant les données de test.
  - a) Dans l'onglet **Général**, dans le champ **Connexion**, sélectionnez la base de données contenant les données de test.
  - b) Cliquez sur **Créer SQL...** pour créer la requête SQL à l'aide du schéma et des tables sélectionnés à partir desquels les données doivent être lues.  
Assurez-vous que la table sélectionnée comporte au moins 1 000 enregistrements pour permettre des valeurs de test optimales.
  - c) Dans l'onglet **Exécution**, cochez la case **Taille de l'extraction**.
  - d) Dans le champ associé, saisissez le nombre d'enregistrements que vous souhaitez lire dans une instance.  
Spectrum™ Technology Platform a été testé pour fonctionner de façon optimale avec une taille d'extraction pouvant atteindre 1 000 enregistrements.
  - e) Cliquez sur **OK**.
6. Enregistrez le job.
7. Exécutez le job.  
La fenêtre **Détails d'exécution** s'ouvre.
8. Cliquez sur **Actualiser**.
9. Notez les heures **Démarré** et **Terminé**.
10. Répétez les étapes 7 à 9 en augmentant progressivement la taille d'extraction pour identifier le paramètre optimal pour votre serveur.

Vous avez désormais identifié le paramètre de taille d'extraction offrant des performances optimales pour votre environnement.

# Notices

© 2018 Pitney Bowes Software Inc. 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 ([www.geonames.org](http://www.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)