

Spectrum™ Technology Platform

Version 2018.2.0

Guía de calidad de datos



Contents

1 - Introducción

Introducción a Data Quality	5
-----------------------------	---

2 - Análisis

Introducción al análisis	8
Cómo definir gramáticas de análisis independientes de dominio en los flujos de datos	10
Análisis específico de la cultura	10
Inspección de resultados de análisis	37
Análisis de nombres personales	41
Plantillas de flujo de datos para análisis	42

3 - Plantillas de flujo de datos para comparación

Normalización de la terminología	59
Estandarización de nombres personales	60
Plantillas para estandarización	62

4 - Comparación

Coincidencia de terminología	65
Campos estándares	66
Técnicas para la definición de claves de cruce	68
Reglas de cruce	71
Cruce de registros de un origen único	85
Comparación de registros de un origen a otro origen	91
Coincidencia de registros entre y dentro orígenes	98

Comparación de registros con una base de datos	104
Comparación de registros con varias reglas de cruce	107
Creación de un servicio de comparación universal	111
Uso de una clave de cruce inmediato	115
Análisis de resultados de cruce	119
Plantillas de flujo de datos para comparación	135

5 - Desduplicación

Filtrado de registros duplicados	144
Creación de un registro Best of Breed.	148

6 - Registros de excepción

Diseño de un flujo de datos para procesar excepciones	154
Diseño de un flujo de datos para revalidación en tiempo real	156

7 - Tablas de búsqueda

Introducción a las tablas de búsqueda	160
Referencia de etapas	160
Tablas del módulo Universal Name	166
Cómo ver el contenido de Table Lookup	167
Cómo agregar un término a Table Lookup	168
Eliminación de un término en una tabla de búsqueda	169
Cómo modificar la forma estandarizada de un término	169
Cómo revertir personalización de tabla	169
Creación de una tabla de búsqueda	170

Importar datos	170
----------------	-----

8 - Referencia de etapas

Módulo Advanced Matching	174
Módulo Business Steward	240
Módulo Data Normalization	285
Módulo Universal Name	304

9 - Códigos de país ISO y compatibilidad de módulos

Códigos de país ISO y compatibilidad de módulos	340
---	-----

1 - Introducción

In this section

Introducción a Data Quality

5

Introducción a Data Quality

Data Quality implica garantizar la exactitud, puntualidad, integridad y consistencia de los datos utilizados por una organización, de modo tal que dichos datos son aptos para ser utilizados. Spectrum™ Technology Platform respalda las iniciativas de calidad de datos al proporcionar las siguientes capacidades.

Análisis

El análisis es el proceso a través del cual se evalúa una secuencia de caracteres de entrada en un campo y se disemina en varios campos. Por ejemplo, podría existir un campo llamado Nombre que contenga el valor "John A. Smith". Mediante el análisis, se puede dividir para tener un campo FirstName que contenga "John", un campo MiddleName que contenga "A" y un campo LastName que contenga "Smith".

Plantillas de flujo de datos para comparación

Normalización toma datos del mismo tipo y los pone en el mismo formato. Algunos tipos de datos que pueden normalizarse incluyen números telefónicos, fechas, nombres, direcciones y números identificatorios. Por ejemplo, los números telefónicos pueden formatearse de manera tal que se eliminen los caracteres no numéricos tales como paréntesis, puntos o guiones.

Es recomendable que realice la normalización de sus datos antes de realizar actividades de cruce o de deduplicación, puesto que los datos normalizados experimentarán cruces más exactos que los datos con inconsistencias en su formato.

Comparación

El cruce es el proceso de identificar registros relacionados entre sí en cierta forma que tenga significado para sus propósitos. Por ejemplo, si intenta eliminar información redundante de los datos de su cliente, puede que desee identificar los registros duplicados de dicho cliente; o, si intenta evitar que el mismo material de marketing se envíe dos veces a la misma dirección, es posible que desee identificar los registros de clientes que habitan en el mismo hogar.

Desduplicación

La desduplicación identifica registros que representan a una entidad, pero que por alguna razón fueron ingresados varias veces en el sistema, en ocasiones con datos con leves diferencias. Por ejemplo, su sistema puede contener información de proveedores de distintos departamentos de su organización, usando cada departamento, para el mismo proveedor, una Id. de proveedor diferente. Al usar Spectrum™ Technology Platform puede consolidar esos registros en un único registro de cada proveedor.

Revisión de registros de excepción

En ciertos casos es posible que cuente con datos que no pueden procesarse automáticamente de manera confidencial y que deben ser revisados por un administrador de datos que esté informado. Algunos ejemplos de registros que pueden requerir de revisión manual incluyen los siguientes:

- Fallas en la verificación de direcciones
- Fallas de geocodificación
- Cruces con bajo nivel de confianza
- Decisiones de combinación/consolidación

El módulo Business Steward es un conjunto de funciones que permiten identificar y resolver registros de excepciones.

2 - Análisis

In this section

Introducción al análisis	8
Cómo definir gramáticas de análisis independientes de dominio en los flujos de datos	10
Análisis específico de la cultura	10
Inspección de resultados de análisis	37
Análisis de nombres personales	41
Plantillas de flujo de datos para análisis	42

Introducción al análisis

El análisis es el proceso a través del cual se evalúa una secuencia de caracteres de entrada en un campo y se disemina en varios campos. Por ejemplo, podría existir un campo llamado Nombre que contenga el valor "John A. Smith". Mediante el análisis, se puede dividir para tener un campo FirstName que contenga "John", un campo MiddleName que contenga "A" y un campo LastName que contenga "Smith".

Para crear un flujo de datos que analice, use la etapa Open Parser. Open Parser permite escribir reglas de análisis llamadas *gramáticas*. Una gramática es un conjunto de expresiones que mapean una secuencia de caracteres en un conjunto de entidades llamadas *patrones de dominio*. Un patrón de dominio es una secuencia de una o más muestras en sus datos de entrada que desea representar como una estructura de datos, como nombre, dirección o números de cuenta. Un patrón de dominio puede consistir en cualquier cantidad de muestras que pueden analizarse a partir de sus datos de entrada. Un patrón de dominio está representado en la gramática de análisis como la expresión <root>. Los datos de entrada generalmente contienen esas muestras en formatos mixtos o difíciles de usar. Por ejemplo:

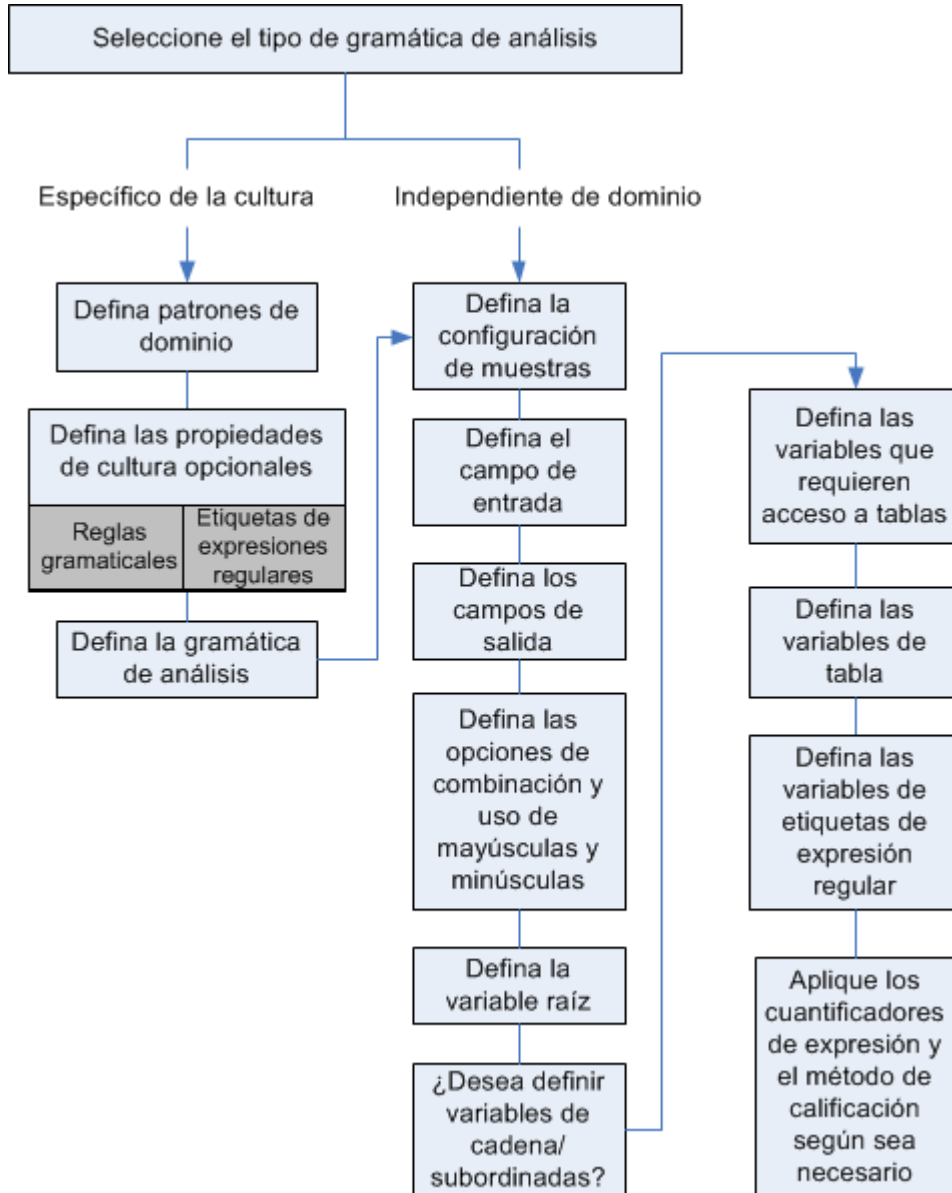
- Sus datos de entrada contienen nombres en un solo campo que desea separar en nombres de pila y apellidos.
- Sus datos de entrada contienen direcciones de varias culturas y usted quiere extraer datos de dirección de una sola cultura específica.
- Sus datos de entrada incluyen texto de forma libre que contiene direcciones de correo electrónico incrustadas y usted quiere extraer direcciones de correo electrónico, cruzarlas con datos personales y almacenarlas en una base de datos.

Hay dos tipos de gramáticas: *específicas de cada cultura e independientes de dominio*. Una gramática de análisis específica de cada cultura se asocia con una cultura o un idioma (por ejemplo, inglés, inglés de Canadá, español, español de México, etc.) y un determinado tipo de datos (números de teléfono, nombres de personas, etc.) Cuando Open Parse se configura para realizar análisis específico de cultura, la gramática de análisis de cada cultura se aplica a cada registro. La gramática con el mejor puntaje de análisis (o la primera en obtener un puntaje de 100) es la que mostrará sus resultados. Por otra parte, las gramáticas de análisis específicas de cada cultura pueden utilizar el valor del campo CultureCode del registro de entrada y procesar los datos según la configuración cultural que se haya incluido en la gramática de análisis de la cultura. Las gramáticas de análisis específicas de la cultura pueden heredar propiedades de un elemento principal. Una gramática de análisis independiente de dominio no se asocia ni a un lenguaje ni a un tipo de datos en particular. Las gramáticas de análisis independientes de dominio no heredan las propiedades de un elemento principal e ignoran cualquier información de CultureCode en los datos de entrada.

Open Parser analiza una secuencia de caracteres en los campos de entrada y los clasifica en una secuencia de muestras a través de un proceso llamado *formación de muestras*. La formación de muestras es el proceso de delimitar y clasificar secciones de una cadena de caracteres de entrada en un conjunto de muestras basado en caracteres separadores (también llamados caracteres

formadores de muestras), como los espacios, los guiones, etc. Las muestras se colocan en los campos de salida que especifique.

El siguiente diagrama ilustra el proceso de creación de una gramática de análisis:



Cómo definir gramáticas de análisis independientes de dominio en los flujos de datos

Para definir gramáticas de análisis independientes de dominio en un flujo de datos:

1. En Enterprise Designer, agregue una etapa Open Parser al flujo de datos.
2. Haga doble clic en la etapa Open parser en el lienzo.
3. Haga clic en **Definir gramática independiente de dominio** en la ficha **Reglas**.
4. Utilice Grammar Editor para crear las reglas gramaticales. Puede ingresar comandos y variables en el cuadro de texto o utilizar los comandos provistos en la ficha **Comandos**. Para obtener más información, consulte [Gramáticas](#) en la página 27.
5. Para cortar, copiar, pegar, y buscar y reemplazar cadenas de texto en su gramática de análisis, haga clic con el botón derecho del mouse en Grammar Editor y seleccione el comando apropiado.
6. Para comprobar la gramática de análisis que ha creado, haga clic en **Validar**.

La función de validación enumera todos los errores de su sintaxis gramatical, incluidas la línea y la columna en las que ocurre el error, una descripción del error y el nombre o valor del comando que está causando el error.

7. Haga clic en la ficha **Vista previa** para probar la gramática de análisis.
8. Cuando haya finalizado de crear su gramática de análisis, haga clic en **Aceptar**.

Análisis específico de la cultura

Definición de gramáticas de análisis específicas de culturas

Una gramática de análisis de culturas específicas permite especificar diferentes reglas de análisis para los diferentes idiomas y culturas. Esto permite analizar datos de diferentes países en una sola etapa Open Parser, como por ejemplo, números de teléfono de los Estados Unidos y números de teléfono del Reino Unido. De forma predeterminada, cada registro de entrada se analiza mediante la gramática de análisis de cada cultura, en el orden especificado en la etapa Open Parser. También puede agregar un campo CultureCode a los registros de entrada en caso de que desee usar para ese registro una gramática de análisis de una cultura determinada. Para obtener más información, consulte [Asignación de un análisis de cultura a un registro](#) en la página 12.

Nota: Si desea crear una gramática de análisis independiente de dominio, consulte [Cómo definir gramáticas de análisis independientes de dominio en los flujos de datos](#) en la página 10.

1. En Enterprise Designer, vaya a **Herramientas > Abrir el editor de Parser Domain**.
2. Haga clic en la ficha **Dominios**.
3. Haga clic en **Agregar**.
4. Ingrese un nombre de dominio en el campo **Nombre**.
5. Ingrese una descripción del nombre de dominio en el campo **Descripción**.
6. Si desea crear un dominio nuevo y vacío, haga clic en **Aceptar**. Si desea crear un nuevo dominio basado en otro dominio, haga lo siguiente.
 - a) Seleccione **Usar otro dominio como plantilla** si desea crear un nuevo dominio basado en otro dominio.
 - b) Seleccione un dominio de la lista. Cuando haga clic en **Aceptar** en el siguiente paso, se creará el nuevo dominio. El nuevo dominio contendrá todas las gramáticas de análisis específicas de la cultura definidas en la plantilla de dominio seleccionada.
 - c) Haga clic en **Aceptar**.
7. Defina la gramática de análisis de la cultura global. La cultura global es la cultura predeterminada, y se utiliza para analizar registros que tienen una cultura para la cual no se ha definido una gramática de análisis específica.
 - a) En la ficha **Gramáticas**, seleccione el nuevo dominio que ha creado.
 - b) Si ha creado un dominio a partir de una plantilla, puede que haya culturas ya incluidas en la lista.
 - Si hay culturas en la lista, seleccione **Cultura global**, y luego haga clic en **Editar**.
 - Si no hay culturas en la lista, haga clic en **Agregar**, seleccione **Cultura global** y luego haga clic en **Aceptar**.
 - c) En la ficha **Gramática**, escriba la gramática de análisis para la cultura global. Puede utilizar las fichas **Comandos**, **Reglas gramaticales** y **Rótulos de RegEx** para insertar elementos de gramática de análisis predefinidos. Para introducir un elemento predefinido, coloque el cursor donde desea insertar el elemento y luego haga doble clic en el elemento que desea agregar.

La ficha **Comandos** muestra los comandos de análisis. Para obtener información sobre los comandos disponibles, consulte [Gramáticas](#) en la página 27.

La ficha **Reglas gramaticales** muestra las reglas gramaticales que ha creado en el cuadro de diálogo Propiedades de cultura. Para obtener más información sobre la creación de reglas gramaticales, consulte [Definición de las reglas gramaticales de una cultura](#) en la página 32.

La ficha **Rótulos de RegEx** muestra los rótulos de expresiones regulares que haya creado en el cuadro de diálogo Propiedades de cultura. Para obtener más información sobre cómo crear etiquetas de expresiones regulares, consulte [Definición de etiquetas de expresiones regulares de culturas](#) en la página 33.

- d) Para comprobar la sintaxis gramatical que ha creado, haga clic en **Validar**. La función de validación de la gramática de análisis muestra cualquier error en su sintaxis gramatical, e incluye el error que se haya encontrado, la línea y la columna donde se produjo el error, y el comando, la regla gramatical o la etiqueta de expresión regular se produjo ocurre el error.
 - e) Para probar los resultados de su gramática con datos de ejemplo, haga clic en la ficha **Vista previa**. En **Datos de entrada**, ingrese los datos de ejemplo que desea analizar. Ingrese un registro por fila. Luego haga clic en el botón **Vista previa**. Los campos de salida analizados se muestran en la cuadrícula **Resultados**. Para obtener información sobre los campos de salida, consulte **Salida** en la página 292. Para obtener información sobre el seguimiento, consulte **Seguimiento de resultados de análisis finales** en la página 37. Si sus resultados no son los esperados, haga clic en la ficha **Gramáticas** y continúe editando la gramática de análisis y probando datos de entrada representativos hasta que la gramática de análisis produzca los resultados esperados.
 - f) Haga clic en **Aceptar** cuando haya terminado de definir la gramática de análisis de la cultura global.
8. Defina una gramática de análisis específica por cada cultura que desee. Para agregar gramáticas específicas de las culturas, haga clic en **Agregar** y defina las reglas de la gramática siguiendo los mismos pasos que para la cultura global. Repita cuantas veces que sea necesario para añadir tantas culturas como necesite.
 9. Cuando termine de agregar gramáticas de análisis de culturas específicas, haga clic en **Aceptar**.

El dominio y las culturas que ha creado se pueden usar ahora en la etapa Open Parser para realizar análisis.

Asignación de un análisis de cultura a un registro

Cuando se configura una etapa Open Parser para el uso de análisis gramaticales específicos de una cultura, los análisis gramaticales de cada cultura se aplican a cada registro de entrada en el orden en que las culturas se enumeran en la etapa de Open Parser. Sin embargo, si se desea aplicar un análisis gramatical específico de una cultura a un registro, puede agregar un campo denominado CultureCode. El campo debe contener uno de los códigos culturales enumerados en la siguiente tabla.

Código cultural

Los códigos culturales se componen de un código de idioma de dos minúsculas y un código de país o región de dos mayúsculas. Por ejemplo, "es-MX" para español (México) y "en-US" para inglés (Estados Unidos). En los casos en que un código de idioma de dos letras no está disponible, se utiliza un código de tres letras, por ejemplo, "uz-Cyrl-UZ" para Uzbek (Uzbekistán, cirílico). El idioma se especifica únicamente mediante el código de idioma de dos letras minúsculas. Por ejemplo, "fr" especifica la cultura neutral del francés y "de" especifica la cultura neutral del alemán.

Nota: Hay dos nombres culturales que siguen una pauta diferente. Las cultura "zh-Hans" (chino simplificado) y "zh-Hant" (chino tradicional) son culturas neutrales. Los nombres de

cultura representan el estándar actual y deben ser usados a menos que tenga una razón para utilizar los nombres más antiguos "zh-CHS" y "zh-CHT".

La siguiente tabla muestra los códigos culturales. que se admiten.

Idioma (Cultura/Región)	Código de cultura
Cultura global	Cultura global
Afrikaans	af
Afrikaans (Sudáfrica)	af-ZA
Albanés	sq
Albanés (Albania)	sq-AL
Árabe	ar
Árabe (Argelia)	ar-DZ
Árabe (Bahréin)	ar-BH
Árabe (Egipto)	ar-EG
Árabe (Irak)	ar-IQ
Árabe (Jordania)	ar-JO
Árabe (Kuwait)	ar-KW

Idioma (Cultura/Región)	Código de cultura
-------------------------	-------------------

Árabe (Líbano)	ar-LB
----------------	-------

Árabe (Libia)	ar-LY
---------------	-------

Árabe (Marruecos)	ar-MA
-------------------	-------

Árabe (Omán)	ar-OM
--------------	-------

Árabe (Qatar)	ar-QA
---------------	-------

Árabe (Arabia Saudita)	ar-SA
------------------------	-------

Árabe (Siria)	ar-SY
---------------	-------

Árabe (Túnez)	ar-TN
---------------	-------

Árabe (Emiratos Árabes Unidos)	ar-AE
--------------------------------	-------

Árabe (Yemen)	ar-YE
---------------	-------

Armenio	hy
---------	----

Armenio (Armenia)	hy-AM
-------------------	-------

Azerí	az
-------	----

Azerí (Azerbaiyán, cirílico)	az-Cyrl-AZ
------------------------------	------------

Idioma (Cultura/Región)	Código de cultura
-------------------------	-------------------

Azerí (Azerbaiyán, latino)

az-Latn-AZ

Vasco

eu

Vasco (Vasco)

eu-ES

Bielorruso

be

Bielorruso (Bielorrusia)

be-BY

Búlgaro

bg

Búlgaro (Bulgaria)

bg-BG

Catalán

ca

Catalán (Cataluña)

ca-ES

Chino

zh

Chino (Hong Kong ZAE, RPC)

zh-HK

Chino (Macao ZAE)

zh-MO

Chino (RPC)

zh-CN

Chino (simplificado)

zh-Hans

Idioma (Cultura/Región)	Código de cultura
-------------------------	-------------------

Chino (Singapur)	zh-SG
------------------	-------

Chino (Taiwán)	zh-TW
----------------	-------

Chino (tradicional)	zh-Hant
---------------------	---------

Croata	hr
--------	----

Croata (Croacia)	hr-HR
------------------	-------

Checo	cs
-------	----

Checo (República Checa)	cs-CZ
-------------------------	-------

Danés	da
-------	----

Danés (Dinamarca)	da-DK
-------------------	-------

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

Divehi (Maldivas)	dv-MV
-------------------	-------

Holandés	nl
----------	----

Holandés (Bélgica)	nl-BE
--------------------	-------

Holandés (Países Bajos)	nl-NL
-------------------------	-------

Idioma (Cultura/Región)	Código de cultura
Inglés	in
Inglés (Australia)	en-AU
Inglés (Belice)	en-BZ
Inglés (Canadá)	en-CA
Inglés (Caribe)	en-029
Inglés (Irlanda)	en-IE
Inglés (Jamaica)	en-JM
Inglés (Nueva Zelanda)	en-NZ
Inglés (Filipinas)	en-PH
Inglés (Sudáfrica)	en-ZA
Inglés (Trinidad y Tobago)	en-TT
Inglés (Reino Unido)	en-GB
Inglés (Estados Unidos)	en-US
Inglés (Zimbabwe)	en-ZW

Idioma (Cultura/Región)	Código de cultura
-------------------------	-------------------

Estonio	et
---------	----

Estonio (Estonia)	et-EE
-------------------	-------

Feroés	fo
--------	----

Feroés (Islas Feroe)	fo-FO
----------------------	-------

Persa	fa
-------	----

Persa (Irán)	fa-IR
--------------	-------

Finlandés	fi
-----------	----

Finlandés (Finlandia)	fi-FI
-----------------------	-------

Francés	fr
---------	----

Francés (Bélgica)	fr-BE
-------------------	-------

Francés (Canadá)	fr-CA
------------------	-------

Francés (Francia)	fr-FR
-------------------	-------

Francés (Luxemburgo)	fr-LU
----------------------	-------

Francés (Mónaco)	fr-MC
------------------	-------

Idioma (Cultura/Región)	Código de cultura
-------------------------	-------------------

Francés (Suiza)	fr-CH
-----------------	-------

Gallego	gl
---------	----

Gallego (España)	gl-ES
------------------	-------

Georgiano	ka
-----------	----

Georgiano (Georgia)	ka-GE
---------------------	-------

Germano	de
---------	----

Alemán (Austria)	de-AT
------------------	-------

Alemán (Alemania)	de-DE
-------------------	-------

Alemán (Liechtenstein)	de-LI
------------------------	-------

Alemán (Luxemburgo)	de-LU
---------------------	-------

Alemán (Suiza)	de-CH
----------------	-------

Griego	el
--------	----

Griego (Grecia)	el-GR
-----------------	-------

Gujaratí	gu
----------	----

Idioma (Cultura/Región)	Código de cultura
-------------------------	-------------------

Gujaratí (India)	gu-IN
------------------	-------

Hebreo	he
--------	----

Hebreo (Israel)	he-IL
-----------------	-------

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

Hindi (India)	hi-IN
---------------	-------

Húngaro	hu
---------	----

Húngaro (Hungría)	hu-HU
-------------------	-------

Islandés	es
----------	----

Islandés (Islandia)	is-IS
---------------------	-------

Indonesio	id
-----------	----

Indonesio (Indonesia)	id-ID
-----------------------	-------

Italiano	it
----------	----

Italiano (Italia)	it-IT
-------------------	-------

Italiano (Suiza)	it-CH
------------------	-------

Idioma (Cultura/Región)	Código de cultura
-------------------------	-------------------

Japonés	ja
---------	----

Japonés (Japón)	ja-JP
-----------------	-------

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

Kannada (India)	kn-IN
-----------------	-------

Kazajo	kk
--------	----

Kazajo (Kazajistán)	kk-KZ
---------------------	-------

Konkaní	kok
---------	-----

Konkaní (India)	kok-IN
-----------------	--------

Coreano	ko
---------	----

Coreano (Corea)	ko-KR
-----------------	-------

Kirguís	ky
---------	----

Kirguís (Kirguistán)	ky-KG
----------------------	-------

Letón	lv
-------	----

Letón (Letonia)	lv-LV
-----------------	-------

Idioma (Cultura/Región)	Código de cultura
-------------------------	-------------------

Lituano	lt
---------	----

Lituano (Lituania)	lt-LT
--------------------	-------

Macedonio	mk
-----------	----

Macedonio (Macedonia, ARYM)	mk-MK
-----------------------------	-------

Malayalam	ms
-----------	----

Malayo (Brunéi)	ms-BN
-----------------	-------

Malayo (Malasia)	ms-MY
------------------	-------

Maratí	mr
--------	----

Maratí (India)	mr-IN
----------------	-------

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

Mongol (Mongolia)	mn-MN
-------------------	-------

Noruego	no
---------	----

Noruego (Bokmål, Noruega)	nb-NO
---------------------------	-------

Noruego (Nynorsk, Noruega)	nn-NO
----------------------------	-------

Idioma (Cultura/Región)	Código de cultura
-------------------------	-------------------

Polaco	pl
--------	----

Polaco (Polonia)	pl-PL
------------------	-------

Portugués	pt
-----------	----

Portugués (Brasil)	pt-BR
--------------------	-------

Portugués (Portugal)	pt-PT
----------------------	-------

Punyabí	pa
---------	----

Punyabí (India)	pa-IN
-----------------	-------

Rumano	ro
--------	----

Rumano (Rumania)	ro-RO
------------------	-------

Ruso	ru
------	----

Ruso (Rusia)	ru-RU
--------------	-------

Sánscrito	SA
-----------	----

Sánscrito (India)	sa-IN
-------------------	-------

Serbio	sr
--------	----

Idioma (Cultura/Región)	Código de cultura
-------------------------	-------------------

Serbio (Serbia, cirílico)	sr-Cyrl-CS
---------------------------	------------

Serbio (Serbia, latino)	sr-Latn-CS
-------------------------	------------

Eslovaco	sk
----------	----

Eslovaco (Eslovaquia)	sk-SK
-----------------------	-------

Esloveno	sl
----------	----

Esloveno (Eslovenia)	sl-SI
----------------------	-------

Español	es
---------	----

Español (Argentina)	es-AR
---------------------	-------

Español (Bolivia)	es-BO
-------------------	-------

Español (Chile)	es-CL
-----------------	-------

Español (Colombia)	es-CO
--------------------	-------

Español (Costa Rica)	es-CR
----------------------	-------

Español (República Dominicana)	es-DO
--------------------------------	-------

Español (Ecuador)	es-EC
-------------------	-------

Idioma (Cultura/Región)	Código de cultura
-------------------------	-------------------

Español (El Salvador)	es-SV
-----------------------	-------

Español (Guatemala)	es-GT
---------------------	-------

Español (Honduras)	es-HN
--------------------	-------

Español (México)	es-MX
------------------	-------

Español (Nicaragua)	es-NI
---------------------	-------

Español (Panamá)	es-PA
------------------	-------

Español (Paraguay)	es-PY
--------------------	-------

Español (Perú)	es-PE
----------------	-------

Español (Puerto Rico)	es-PR
-----------------------	-------

Español (España)	es-ES
------------------	-------

Español (España, variedad tradicional)	es-ES_tradnl
--	--------------

Español (Uruguay)	es-UY
-------------------	-------

Español (Venezuela)	es-VE
---------------------	-------

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

Idioma (Cultura/Región)	Código de cultura
-------------------------	-------------------

Swahili (Kenia)	sw-KE
-----------------	-------

Sueco	sv
-------	----

Sueco (Finlandia)	sv-FI
-------------------	-------

Sueco (Suecia)	sv-SE
----------------	-------

Sirio	syr
-------	-----

Sirio (Siria)	syr-SY
---------------	--------

Tamil	ta
-------	----

Tamil (India)	ta-IN
---------------	-------

Tártaro	tt
---------	----

Tártaro (Rusia)	tt-RU
-----------------	-------

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

Telugu (India)	te-IN
----------------	-------

Tailandés	th
-----------	----

Tailandés (Tailandia)	th-TH
-----------------------	-------

Idioma (Cultura/Región)	Código de cultura
Turco	tr
Turco (Turquía)	tr-TR
Ucraniano	uk
Ucraniano (Ucrania)	uk-UA
Urdu	ur
Urdu (Pakistán)	ur-PK
Uzbeko	uz
Uzbeko (Uzbekistán, cirílico)	uz-Cyrl-UZ
Uzbeko (Uzbekistán, latino)	uz-Latn-UZ
Vietnamita	vi
Vietnamita (Vietnam)	vi-VN

Gramáticas

Una gramática de análisis válida contiene:

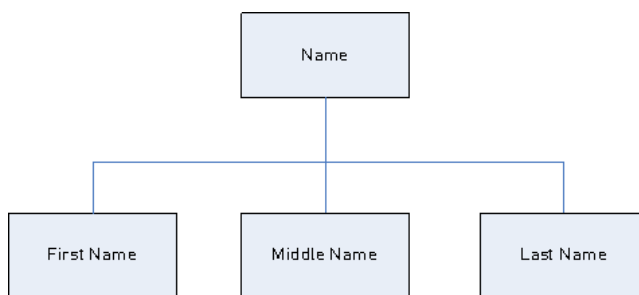
- Una variable raíz que define la secuencia de muestras o el patrón de dominio como variables de raíz.

- Las variables de regla que definen el conjunto válido de caracteres y la secuencia en que pueden aparecer para ser considerados como miembro de un patrón de dominio. Para obtener más información, consulte [Comandos de la sección Regla](#) en la página 30.
- El campo de entrada por analizar. El campo de entrada designa el campo por analizar en los registros de datos de origen.
- Los campos de salida de los datos analizados resultantes. Los campos de salida definen dónde almacenar cada muestra resultante que es analizada.

Una gramática de análisis válida también contiene otros comandos opcionales para lo siguiente:

- Los caracteres usados para formar muestras de los datos de entrada que está analizando: Los caracteres que forman muestras son caracteres, como el espacio y el guión, que determinan el comienzo y el final de una muestra. El carácter de muestra predeterminado es un espacio. Los caracteres que forman muestras son la forma principal de dividir una secuencia de caracteres en un conjunto de muestras. Es posible configurar el comando de formación de muestras en NINGUNO para evitar que se formen muestras de un campo. Cuando la función de formación de muestras está configurada en Ninguno, las reglas de gramática deben incluir todos los espacios dentro de la definición de regla.
- Opciones de distinción de mayúsculas y minúsculas para las muestras en los datos de entrada.
- Carácter de unión para delimitar muestras coincidentes.
- Comparación de muestras en tablas
- Comparación de muestras compuestas en tablas
- Definición de etiquetas RegEx
- Cadena de caracteres literales entre comillas
- Cuantificadores de expresión (opcional). Para obtener más información sobre cuantificadores de expresión, consulte [Comandos de la sección Regla](#) en la página 30 y [Cuantificadores de expresión: comportamiento expansivo, reactivo y posesivo](#).
- Otros indicadores varios para agrupar, comentar y realizar asignaciones (opcional). Para obtener más información sobre expresiones agrupadas, consulte [Operador de agrupación \(\)](#).

Las variables de regla en su gramática de análisis forman una estructura de árbol con capas de la secuencia de caracteres o muestras en un patrón de dominio. Por ejemplo, puede crear una gramática de análisis que defina un patrón de dominio basado en datos de entrada de nombre que incluyan las muestras <FirstName>, <MiddleName> y <LastName>.



Uso de los datos de entrada:

```
Joseph Arnold Cowers
```

Es posible representar esa cadena de datos como tres muestras de un patrón de dominio:

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

Las variables de regla para este patrón de dominio son:

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

Según este ejemplo de gramática simple, Open Parser forma muestras con espacios e interpreta la muestra `Joseph` como nombre porque los caracteres de la primera muestra coinciden con la definición `[A-Za-z]+` y la muestra está en la secuencia definida. De forma opcional, cualquier expresión puede estar seguida de otra expresión.

Ejemplo

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

Una regla gramatical es una instrucción gramatical en la que una variable es equivalente a una o más expresiones. Cada regla gramatical sigue esta forma:

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

Las reglas gramaticales deben seguir estas reglas:

- `<root>` es un nombre de variable especial y es la primera regla ejecutada en la gramática porque define el patrón de dominio. Ninguna otra regla en la gramática puede hacer referencia a `<root>`.
- Una variable `<rule>` no puede referirse a sí misma directa o indirectamente. Cuando la regla A hace referencia a la regla B, que hace referencia a la regla C, que a su vez hace referencia a A, se crea una referencia circular. No se permiten referencias circulares.
- Una variable `<rule>` es igual a una o más expresiones.
- Cada `expression` está separada por "O", que se indica utilizando el carácter de barra vertical "|".
- Las expresiones pueden examinarse una por una. Se selecciona la primera `expression` a cruzar. No se examinan nuevas expresiones.
- El nombre de variable puede estar compuesto por letras, números, guión bajo (`_`) y guión (`-`). El nombre de la variable puede comenzar con cualquier carácter válido. Si el nombre del campo de salida especificado no cumple con esa forma, use la función de alias para asignar el nombre de la variable al campo de salida.

Una expresión puede ser cualquiera de los siguientes tipos:

- Otra variable
- Una cadena consistente en uno o más caracteres entre comillas simples o dobles. Por ejemplo:

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

- Tabla
- CompoundTable
- Comandos RegEx

Metacaracteres de comando

Open Parser admite el conjunto estándar de metacaracteres de clase de carácter RegEx de Java en los comandos %Tokenize y @RegEx. Un metacarácter es un carácter que contiene un significado especial en la comparación de patrones. Los metacaracteres compatibles son:

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

Existen dos maneras de lograr que un metacarácter sea tratado como un carácter ordinario:

- Coloque una barra invertida delante del metacarácter.
- Enciérrelo dentro de \Q (que indica el comienzo de las comillas) y \E (que indica el final de las comillas).

%Tokenize sigue la regla de las clases de carácter de expresiones regulares de Java, no de las expresiones regulares de Java en su conjunto.

En general, los caracteres reservados de un conjunto de caracteres son:

- '[' and ']' indica otro conjunto.
- '-' es un metacarácter si se encuentra entre otros dos caracteres.
- '^' es un metacarácter si es el primer carácter de un conjunto.
- '&&' son metacaracteres si se encuentran entre otros dos caracteres.
- '\' significa que el carácter contiguo es un literal.

Si tiene alguna duda sobre cómo será tratado un carácter pero desea que sea tratado como un literal, utilice como carácter de escape la barra invertida.

Comandos de la sección de encabezado

Esta sección describe los comandos de la sección de encabezado. Algunos comandos son opcionales. Si un comando es opcional, el valor o comportamiento predeterminado está en la lista.

- **Tokenize** (opcional)
- **Tokenize (None)**
- **InputField** (obligatorio si no se usan los campos de entrada)
- **InputFields** (obligatorio si no se usa el campo de entrada)
- **OutputFields** (obligatorio)
- **IgnoreCase** (opcional)
- **Join** (opcional)

Comandos de la sección Regla

Los comandos de la sección Regla son:

- RegEx
- Tabla
- CompoundTable
- Muestra
- Calificación
- RuleID
- Variable <root>
- rule|rule
- Operador de agrupación ()
- Operador de ocurrencias mínimas y máximas {min,max}
- Operador de ocurrencias exactas {exact}
- Operador de asignación (=)
- Operador OR (|)
- Operador End-of-Rule (;)
- Operador de comentario (!)
- Cuantificador cero o una ocurrencia (?)
- Cuantificador cero o más ocurrencias (*)
- Cuantificador una o más ocurrencias (+)
- Cuantificadores de expresión: comportamiento expansivo, reacio y posesivo

Culturas

Una cultura es el concepto primario para organizar las gramáticas de análisis específicas de la cultura. Puede utilizar las culturas para crear diferentes reglas de análisis para distintas culturas e idiomas. La cultura sigue una jerarquía:

- **Cultura global:** la cultura global es independiente de las culturas específicas y no posee un idioma. Utilice la cultura global para crear reglas gramaticales de análisis que abarquen todas las culturas y los idiomas.
 - **Idioma:** un idioma está asociado con un idioma, pero no con una cultura/región específica. Por ejemplo, inglés.
 - **Cultura/región:** una cultura/región está asociada con un idioma y un país o región. Por ejemplo, inglés en el Reino Unido o inglés en Estados Unidos.

En la jerarquía de cultura, el elemento principal de una cultura/región es un idioma y el elemento principal de un idioma es la cultura global.

Las culturas/regiones heredan las propiedades del idioma principal. Los idiomas heredan las propiedades de la cultura global. Por lo tanto, se pueden definir las gramáticas de análisis en un idioma para ser utilizadas en muchos países que comparten el mismo idioma. Luego, puede invalidar las reglas gramaticales de idioma con gramáticas de análisis especializadas para un país o región

particular que comparte el mismo idioma que la cultura de idioma base, pero tiene direcciones, nombres u otras diferencias regionales o de país específicos.

También puede usar una herencia cultural para analizar registros entrantes que tienen un código de cultura asignado, pero no tienen una regla gramatical definida para ese código de cultura. En este caso, Open Parser busca un código de idioma que tenga una regla gramatical asignada. Si no existe, Open Parser busca una regla gramatical asignada en la cultura global.

Domain Editor utiliza una combinación de un código de idioma y un código de cultura para representar el idioma y la cultura/región, respectivamente.

Definición de las reglas gramaticales de una cultura

Puede usar las reglas gramaticales de una cultura para reemplazar una parte de la gramática de análisis de una cultura global con las cadenas, comandos, o expresiones específicas de una cultura y/o idioma. Al definir una regla gramatical, puede personalizar partes de la gramática de análisis de la cultura global con base en la cultura y/o idioma de los registros. Esto resulta conveniente si no desea crear una gramática de análisis completamente diferente para cada cultura, y en cambio desea usar la gramática de la cultura global y personalizar solamente las partes específicas de la cultura global para cada una de las culturas.

Esta sección describe la forma de crear una regla gramatical para una cultura.

1. En Enterprise Designer, vaya a **Herramientas > Abrir el editor de Parser Domain**.
2. Haga clic en la ficha **Culturas**.

Para acceder a una lista completa de las culturas admitidas, consulte [Asignación de un análisis de cultura a un registro](#) en la página 12.

3. Seleccione la cultura a la cual desea añadir una regla gramatical y luego haga clic en **Propiedades**.
4. Haga clic en la ficha **Reglas gramaticales**. La información que se muestra incluye los nombres de las reglas gramaticales definidas para la cultura seleccionada, la cultura de origen asociada, el valor definido de la regla gramatical y la descripción.
5. Haga clic en **Agregar**.
6. Ingrese un nombre para la regla gramatical en el cuadro de texto **Nombre**.
7. Ingrese una descripción de la regla gramatical en el campo **Descripción**.
8. Ingrese la regla gramatical en el campo **Valor**.

La regla gramatical puede ser cualquier variable, cadena, comando o expresión agrupada válida. Para obtener más información, consulte [Gramáticas](#) en la página 27.

9. Seleccione **Activar ajuste automático de línea** para visualizar el valor en el cuadro de texto sin necesidad de desplazarse en pantalla.
10. Haga clic en **Aceptar**.

Se validará el valor de la regla gramatical que ingresó. Si el valor contiene errores de sintaxis gramatical, un mensaje mostrará una descripción de los errores encontrados, la línea y la

columna donde se ubica el error, y el comando, la regla gramatical o la etiqueta de expresión regular donde se ubica el error.

Ejemplo de regla gramatical

Usted cuenta con una gramática que analiza nombres occidentales. La estructura del patrón puede ser la misma para todas las culturas (<FirstName><MiddleName><LastName>) y muchas de las reglas pueden coincidir con el mismo patrón o tabla. Sin embargo, también cuenta con tablas específicas de una cultura para los apellidos, y desea usar la tabla adecuada basándose en el código cultural del registro.

Para lograrlo, puede definir una regla gramatical para cada cultura que sustituya el elemento <LastName> en la cultura global con una referencia a la tabla específica de la cultura. Por ejemplo, si tiene una tabla con apellidos holandeses, creará una regla gramatical para la cultura holandesa (nl) como se indica a continuación:

Nombre: LastName

Descripción: apellidos holandeses

Valor: @Table("Dutch Last Names");

Definición de etiquetas de expresiones regulares de culturas

Esta sección describe la forma de definir etiquetas de expresiones regulares de culturas cuando se define una gramática de análisis de una cultura específica.

1. En Enterprise Designer, vaya a **Herramientas > Abrir el editor de Parser Domain**.
2. Haga clic en la ficha **Culturas**. La ficha **Culturas** muestra una lista de culturas admitidas. Para acceder a una lista completa de las culturas admitidas, consulte [Asignación de un análisis de cultura a un registro](#) en la página 12.
3. Seleccione una cultura de la lista y luego haga clic en **Propiedades**. Aparecerá el cuadro de diálogo **Propiedades de cultura**.
4. Haga clic en la ficha **Etiquetas de expresiones regulares**. La información que se muestra incluye los nombres de etiquetas de expresiones regulares definidas para la cultura seleccionada y la cultura de origen asociada, el valor de la etiqueta de expresión regular y la descripción.
5. Haga clic en **Agregar** o **Modificar**.
6. Ingrese un nombre para la etiqueta de expresión regular en el cuadro de texto **Nombre**.

Si ingresa un nombre que ya existe en la cultura seleccionada, aparecerá un icono de advertencia. Ingrese un nombre diferente o cierre el cuadro de diálogo, elimine la etiqueta de expresión regular existente y haga clic en **Agregar** nuevamente.

7. Ingrese una descripción de la etiqueta de expresión regular en el cuadro de texto **Descripción**.
8. Ingrese un valor para la etiqueta de expresión regular en el cuadro de texto **Valor**.

El valor puede ser cualquier expresión regular válida pero no puede coincidir con una cadena vacía.

Domain Editor incluye varias etiquetas de expresiones regulares predefinidas que pueden usarse para definir propiedades culturales. También puede utilizar estas etiquetas de expresiones regulares para definir los caracteres de formación de muestras en su gramática de análisis.

Puede modificar las etiquetas de expresiones regulares predefinidas o copiarlas y crear sus propias variantes. También es posible utilizar propiedades de invalidación para crear etiquetas de expresiones regulares especializadas para idiomas específicos.

- Letra: cualquier letra de cualquier idioma. Esta etiqueta de expresión regular incluye invalidaciones para varios idiomas debido a los diferentes alfabetos utilizados, por ejemplo, alfabetos cirílicos, de idiomas asiáticos y tailandeses.
- Minúscula: una letra minúscula que tiene una variante en mayúscula.
- Número: cualquier carácter numérico de cualquier alfabeto.
- Puntuación: cualquier carácter de puntuación.
- Mayúscula: una letra mayúscula que tiene una variante en minúscula.
- Espacio en blanco: cualquier espacio en blanco o separador invisible.

9. Haga clic en **Aceptar**.

Importación y exportación de culturas

Además de crear dominios, también puede importar dominios creados en otro lugar y exportar dominios creados en Domain Editor.

1. En Enterprise Designer, vaya a **Herramientas > Abrir el editor de Parser Domain**.
2. Haga clic en la ficha **Culturas**.
3. Haga clic en **Importar** o **Exportar**.
4. Realice una de estas acciones:
 - Si está importando esta cultura, explore hasta la ubicación y seleccione una cultura. Haga clic en **Abrir**. La cultura importada aparece en Domain Editor.
 - Si está exportando una cultura, explore hasta la ubicación donde desee guardar la cultura exportada y selecciónela. Haga clic en **Guardar**. La cultura exportada se guarda y se regresa a Domain Editor.

Dominios

Cómo agregar un dominio

Un dominio representa un tipo de datos, como por ejemplo datos de nombre, dirección, y número telefónico. Consiste en un patrón que representa una secuencia de una o más muestras en sus datos de entrada que comúnmente se necesitan para el análisis y que se asocian con una o más culturas.

Esta sección describe la forma de agregar un dominio en Domain Editor cuando se define una gramática de análisis específica de una cultura.

1. En Enterprise Designer, vaya a **Herramientas > Abrir el editor de Parser Domain**.
2. Haga clic en la ficha **Dominios**.
3. Haga clic en **Agregar**.
4. Ingrese un nombre de dominio en el campo **Nombre**.
5. Ingrese una descripción del nombre de dominio en el campo **Descripción**.
6. Si desea crear un dominio nuevo y vacío, haga clic en **Aceptar**. Si desea crear un nuevo dominio basado en otro dominio, haga lo siguiente.
 - a) Seleccione **Usar otro dominio como plantilla** si desea crear un nuevo dominio basado en otro dominio.
 - b) Seleccione un dominio de la lista. Cuando haga clic en **Aceptar** en el siguiente paso, se creará el nuevo dominio. El nuevo dominio contendrá todas las gramáticas de análisis específicas de la cultura definidas en la plantilla de dominio seleccionada.
 - c) Haga clic en **Aceptar**.

Modificación de un dominio

Un dominio representa un tipo de datos, como por ejemplo datos de nombre, dirección, y número telefónico. Consiste en un patrón que representa una secuencia de una o más muestras en sus datos de entrada que comúnmente se necesitan para el análisis y que se asocian con una o más culturas.

Esta sección describe la forma de modificar un dominio.

1. En Enterprise Designer, vaya a **Herramientas > Abrir el editor de Parser Domain**.
2. Haga clic en la ficha **Dominios**.
3. Seleccione un dominio de la lista y haga clic en **Modificar**. Aparecerá el cuadro de diálogo **Modificar dominio**.
4. Cambie la información de la descripción.
5. Si solo desea modificar la descripción del dominio, haga clic en **Aceptar**. Si realizó actualizaciones al dominio de plantilla y desea agregar esos cambios al dominio que está modificando, siga adelante con el próximo paso.
6. Seleccione **Usar otro dominio como plantilla** para heredar los cambios aplicados a la plantilla de dominio.
7. Seleccione una plantilla de patrón de dominio de la lista. Cuando haga clic en **Aceptar** en el siguiente paso, se modificará el patrón de dominio. El patrón de dominio modificado contendrá todas las gramáticas de análisis específicas de la cultura definidas en la plantilla de patrón de dominio seleccionada. Todas las gramáticas de análisis del patrón de dominio se sobrescribirán con la gramática de análisis de la plantilla de patrón de dominio.
8. Haga clic en **Aceptar**.

Para ver cómo funciona, haga lo siguiente:

1. Cree un patrón de dominio denominado **NameParsing** y defina gramáticas de análisis para Cultura global, en (inglés) y en-US (inglés de Estados Unidos).
2. Cree un patrón de dominio denominado **NameParsing2** y utilice **NameParsing** como plantilla de patrón de dominio. **NameParsing2** se crea como una copia exacta y contiene las gramáticas de análisis para Cultura global, en (inglés) y en-US (inglés de Estados Unidos).
3. Modifique las gramáticas de análisis específicas de la cultura para **NameParsing** cambiando algunas de las reglas gramaticales de la gramática de Cultura global y agregue en-CA (inglés de Canadá) como una nueva cultura.
4. Seleccione **NameParsing2** en la ficha Dominios, haga clic en Modificar y vuelva a utilizar **NameParsing** como plantilla de patrón de dominio.

Los resultados serán:

- Se actualizará la gramática de análisis de Cultura global (se sobrescribirá cualquier cambio que se haya realizado).
- Las culturas en (inglés) y en-US (inglés de Estados Unidos) no sufrirán modificaciones (a menos que se hayan modificado en el dominio de destino, en cuyo caso se revertirán a la versión de Name Parsing).
- Se agregará una gramática específica de la cultura para en-CA (inglés de Canadá).

Eliminación de un dominio

Un dominio representa un tipo de datos, como por ejemplo datos de nombre, dirección, y número telefónico. Consiste en un patrón que representa una secuencia de una o más muestras en sus datos de entrada que comúnmente se necesitan para el análisis y que se asocian con una o más culturas.

Esta sección describe la forma de eliminar un dominio.

1. En Enterprise Designer, vaya a **Herramientas > Abrir el editor de Parser Domain**.
2. Haga clic en la ficha **Dominios**.
3. Seleccione un dominio de la lista.
4. Haga clic en **Quitar**.

Si el dominio está asociado con una o más gramáticas específicas de la cultura, se mostrará un mensaje similar que le solicitará confirmar que desea eliminar el dominio. Si no hay ninguna gramática de análisis específica de la cultura asociada a este dominio, aparecerá un mensaje que confirma si desea eliminar el dominio seleccionado.

5. Haga clic en **Sí**. Se eliminarán el dominio y todas las gramáticas de análisis específicas de una cultura asociadas con ese dominio.

Importación y exportación de dominios

Además de crear dominios, también puede importar dominios creados en otro lugar y exportar dominios creados en Domain Editor.

1. Haga clic en la ficha **Dominios**. Se muestra la ficha **Dominios**.
2. Haga clic en **Importar** o **Exportar**.
3. Realice una de estas acciones:
 - Si está importando un dominio, explore hasta la ubicación y seleccione un nombre de dominio. Haga clic en **Abrir**. El dominio importado aparece en Domain Editor.
 - Si está exportando un dominio, explore hasta la ubicación donde desee guardar el dominio exportado y selecciónela. Haga clic en **Guardar**. El dominio exportado se guarda y se regresa a Domain Editor.

Inspección de resultados de análisis

Seguimiento de resultados de análisis finales

La función Detalles de seguimiento de Open Parser muestra una vista gráfica de cómo el campo de entrada se analizó, ejemplo-por-ejemplo, en el campo de salida. Seguimiento muestra los resultados coincidentes, los resultados no coincidentes, y los resultados provisionales.

Resultados de análisis finales muestra el árbol de gramática de análisis y la salida resultante. Use esta vista cuando desee ver solo los resultados del proceso de cruce. Esta es la vista predeterminada.

1. En Enterprise Designer, abra el flujo de datos que contiene la etapa Open Parser cuyos resultados de análisis desea rastrear.
2. Haga doble clic en el lienzo en la etapa Open Parser.
3. Haga clic en la ficha **Preview**.
4. Ingrese los datos de muestra que se desea analizar y luego haga clic en el botón **Vista previa**.
5. En la columna Seguimiento, haga clic en **Haga clic aquí...** para mostrar el diagrama de seguimiento.

La vista de árbol de la gramática de análisis muestra uno o más de los siguientes elementos, según las opciones seleccionadas.

- La variable <root>. El nodo superior del árbol es la variable <root>.

- Las expresiones definidas en la variable <root>. Los nodos de segundo nivel son las expresiones definidas en la variable <root>. Las expresiones <root> también definen los nombres de los campos de salida.
 - Las definiciones de variable de los nodos de segundo nivel. Los nodos de tercer nivel y de cada nivel inferior son las definiciones de cada una de las expresiones <root>. Las definiciones de expresión pueden ser otras variables, alias o definiciones de regla.
 - Los valores y las muestras que son salida. El nodo inferior del árbol muestra los valores asignados a cada muestra secuencial en la gramática de análisis.
 - La calificación del analizador para elementos relevantes de la gramática de análisis. Las calificaciones del analizador se determinan, respecto de una expresión de raíz, desde abajo hacia arriba. Por ejemplo, si un patrón de expresión tiene un peso de 80 y una regla antecesora tiene un peso de 75, la calificación final para la expresión antecesora es el producto de las calificaciones secundarias y las calificaciones antecesoras, que en este ejemplo serían de 60 por ciento.
 - El carácter de espacio aparece en el cuadro de texto **Datos de entrada** como un carácter de espacio de no separación (paréntesis hacia arriba) para ver mejor los caracteres de espacio. Los delimitadores no se usan ya que las fichas se muestran atenuadas.
6. En el campo **Información** seleccione **Resultados de análisis finales**.
- Nota:** Para revisar los eventos de análisis, consulte [Recorrer eventos de análisis](#) en la página 39.
7. En la lista **Nivel de detalle**, seleccione una de las opciones.
- **Ocultar expresiones sin resultados.** Muestra esas ramas que derivan en un resultado de coincidencia o de no coincidencia. Cualquier rama de expresión de raíz que no deriva en una coincidencia se muestra con puntos suspensivos. Si desea ver una rama que no deriva en un cruce, haga doble clic en los puntos suspensivos.
 - **Ocultar expresiones radicales sin resultados** Muestra todas las ramas de las expresiones de raíz que contienen resultados de coincidencia o de no coincidencia. No se muestra ninguna otra expresión raíz.
 - **Mostrar todas las raíces.** Muestra cada una de las expresiones radicales. Si una raíz no tiene resultados de coincidencia, la visualización de esa expresión de raíz se contrae utilizando el símbolo de puntos suspensivos.
 - **Mostrar todas las expresiones.** Muestra las expresiones de raíz y todas las ramas. Las expresiones de raíz ya no se muestran como puntos suspensivos; en su lugar, se muestran las reglas de cada expresión en la rama.
- Si ha seleccionado la vista de nivel de detalle que oculta expresiones sin resultados y selecciona una expresión de raíz que actualmente no se muestra, Detalles de seguimiento cambia la selección de nivel de detalle a un elemento de lista que muestra la cantidad mínima de expresiones radicales y, al mismo tiempo, también muestra esa expresión radical seleccionada.
8. Haga clic en **Mostrar puntajes** para ver los puntajes del análisis para las expresiones radicales y variables, y las coincidencias y no coincidencias.

9. En el campo **Zoom** seleccione el tamaño de la vista de árbol.
10. En el campo **Cláusula radical**, seleccione una de las opciones para mostrar esa rama del árbol de expresiones de raíz.

Cuando se hace clic en una rama de expresión en el diagrama de seguimiento, la lista **Cláusula raíz** se actualiza para mostrar la cláusula seleccionada. Haga doble clic en los puntos suspensivos para mostrar una expresión contraída.

11. Haga clic en **Aceptar** cuando haya terminado. Las configuraciones de nivel de detalle, mostrar calificaciones y el control de acercamiento y alejamiento se guardan cuando hace clic en **Aceptar**.

Recorrer eventos de análisis

La vista de Detalles de seguimiento de Open Parser le permite ver un diagrama de los pasos evento por evento en el proceso de comparación. Use esta vista cuando soluciona un error en el proceso de comparación y desea ver cómo se evalúa cada muestra, la formación de muestras de la gramática de análisis y los resultados de comparación muestra por muestra.

1. En Enterprise Designer, abra el flujo de datos que contiene la etapa Open Parser cuyos resultados de análisis desea rastrear.
2. Haga doble clic en el lienzo en la etapa Open Parser.
3. Haga clic en la ficha **Preview**.
4. Ingrese los datos de muestra que se desea analizar y luego haga clic en el botón **Vista previa**.
5. En la columna Seguimiento, haga clic en **Haga clic aquí...** para mostrar el diagrama de seguimiento.

La vista de árbol de la gramática de análisis muestra uno o más de los siguientes elementos, según las opciones seleccionadas.

- La variable <root>. El nodo superior del árbol es la variable <root>.
- Las expresiones definidas en la variable <root>. Los nodos de segundo nivel son las expresiones definidas en la variable <root>. Las expresiones <root> también definen los nombres de los campos de salida.
- Las definiciones de variable de los nodos de segundo nivel. Los nodos de tercer nivel y de cada nivel inferior son las definiciones de cada una de las expresiones <root>. Las definiciones de expresión pueden ser otras variables, alias o definiciones de regla.
- Los valores y las muestras que son salida. El nodo inferior del árbol muestra los valores asignados a cada muestra secuencial en la gramática de análisis.
- La calificación del analizador para elementos relevantes de la gramática de análisis. Las calificaciones del analizador se determinan, respecto de una expresión de raíz, desde abajo hacia arriba. Por ejemplo, si un patrón de expresión tiene un peso de 80 y una regla antecesora tiene un peso de 75, la calificación final para la expresión antecesora es el producto de las

calificaciones secundarias y las calificaciones antecesoras, que en este ejemplo serían de 60 por ciento.

- El carácter de espacio aparece en el cuadro de texto **Datos de entrada** como un carácter de espacio de no separación (paréntesis hacia arriba) para ver mejor los caracteres de espacio. Los delimitadores no se usan ya que las fichas se muestran atenuadas.
6. Las coincidencias y las no coincidencias están identificadas con un código de colores en el diagrama de seguimiento:
 - Las casillas de color verde indican coincidencias que son parte del resultado correcto final.
 - Las casillas de color rojo indican falta de coincidencia.
 - Las casillas de color amarillo indican coincidencias provisionales que finalmente se revertirán a medida que se recorran los eventos. Las coincidencias provisionales se muestran solo en Recorrer eventos de análisis.
 - Las casillas de color gris indican coincidencias provisionales que se han revertido para dejar libre esa unidad de texto para otra expresión. Las coincidencias provisionales se muestran solo en Recorrer eventos de análisis.
 7. En la lista **Información**, seleccione **Recorrer eventos de análisis**.
 8. En la lista **Nivel de detalle**, seleccione una de las opciones.
 - **Ocultar expresiones sin resultados**. Muestra esas ramas que derivan en un resultado de coincidencia o de no coincidencia. Cualquier rama de expresión de raíz que no deriva en una coincidencia se muestra con puntos suspensivos. Si desea ver una rama que no deriva en un cruce, haga doble clic en los puntos suspensivos.
 - **Ocultar expresiones radicales sin resultados** Muestra todas las ramas de las expresiones de raíz que contienen resultados de coincidencia o de no coincidencia. No se muestra ninguna otra expresión raíz.
 - **Mostrar todas las raíces**. Muestra cada una de las expresiones radicales. Si una raíz no tiene resultados de coincidencia, la visualización de esa expresión de raíz se contrae utilizando el símbolo de puntos suspensivos.
 - **Mostrar todas las expresiones**. Muestra las expresiones de raíz y todas las ramas. Las expresiones de raíz ya no se muestran como puntos suspensivos; en su lugar, se muestran las reglas de cada expresión en la rama.

Si ha seleccionado la vista de nivel de detalle que oculta expresiones sin resultados y selecciona una expresión de raíz que actualmente no se muestra, Detalles de seguimiento cambia la selección de nivel de detalle a un elemento de lista que muestra la cantidad mínima de expresiones radicales y, al mismo tiempo, también muestra esa expresión radical seleccionada.
 9. Haga clic en **Mostrar puntajes** para ver los puntajes del análisis para las expresiones radicales y variables, y las coincidencias y no coincidencias.
 10. En el campo **Zoom** seleccione el tamaño de la vista de árbol.
 11. En el campo **Cláusula radical**, seleccione una de las opciones para mostrar esa rama del árbol de expresiones de raíz.

Cuando se hace clic en una rama de expresión en el diagrama de seguimiento, la lista **Cláusula raíz** se actualiza para mostrar la cláusula seleccionada. Haga doble clic en los puntos suspensivos para mostrar una expresión contraída.

12. La casilla de verificación **Ir automáticamente al nodo seleccionado** está marcada de forma predeterminada. Cuando esta opción está seleccionada y hace clic en el botón **Reproducir**, los eventos se ejecutan desde el comienzo y se detienen en el primer evento que ocurre con el nodo seleccionado o cualquiera de sus elementos secundarios. Para reproducir todos los eventos sin pausas, desmarque esta casilla antes de hacer clic en el botón **Reproducir**.
13. En el campo **Retardo de reproducción (segundos)**, especifique un retardo para controlar la velocidad de la reproducción.
14. Haga clic en el botón **Reproducir** para comenzar a ejecutar los eventos de análisis.
15. Haga clic en **Aceptar** cuando haya terminado.

Análisis de nombres personales

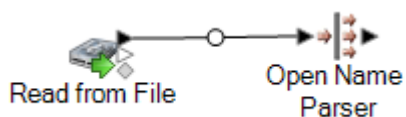
Si tiene datos de nombres que están en un solo campo, es posible que desee analizar los nombres en campos separados por cada parte de dichos nombres, es decir, nombre de pila, apellido, título de respeto, entre otros. De este modo, estos elementos de nombre analizados pueden utilizarse en otras operaciones automatizadas tales como comparación y normalización de nombres o consolidación de nombres provenientes de múltiples registros.

1. Si aún no lo ha hecho, cargue las siguientes tablas en el servidor Spectrum™ Technology Platform:
 - Base de Open Parser
 - Nombres mejorados de Open Parser

Use la utilidad de carga de bases de datos del módulo Data Normalization para cargar estas tablas. Para obtener instrucciones sobre la carga de tablas, consulte la *Guía de instalación*.

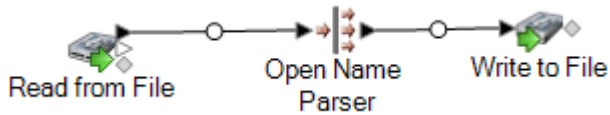
2. Genere un nuevo flujo de datos en Enterprise Designer.
3. Arrastre hacia el lienzo una etapa de origen.
4. Haga doble clic en la etapa de origen y configúrela. Consulte la *Guía de Dataflow Designer* para obtener instrucciones acerca de cómo configurar etapas de origen.
5. Arrastre hacia el lienzo una etapa Open Name Parser y conéctela a la etapa de origen.

Por ejemplo, si está utilizando la etapa Read from File, su flujo de datos tendrá el siguiente aspecto:



- Arrastre hacia el lienzo la etapa receptora y conecte a esta Open Name Parser.

Por ejemplo, si está utilizando una etapa receptora Write to File, su flujo de datos puede tener el siguiente aspecto:



- Haga doble clic en la etapa receptora y configúrela. Consulte la *Guía de Dataflow Designer* para obtener instrucciones acerca de cómo configurar etapas de origen.

Ha creado un flujo de datos que puede analizar nombres de personas en sus partes componentes, colocando en su propio campo cada parte del nombre.

Plantillas de flujo de datos para análisis

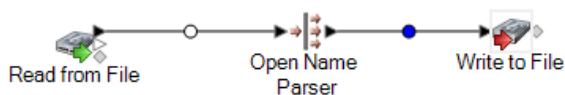
Análisis de nombres en inglés

Esta plantilla de flujo de datos muestra cómo tomar datos de nombres personales (por ejemplo "Juan P. Smith"), lo analiza en primer nombre, segundo nombre y apellido, y añade datos sobre género.

Situación empresarial posible

Usted trabaja para una empresa de seguros que quiere enviar presupuestos personalizados según el género a potenciales clientes. Sus datos de entrada incluyen datos de nombres completos y quiere dividir esos datos en los campos de nombre, segundo nombre y apellido. También deberá determinar el género de las personas en sus datos de entrada.

El siguiente flujo de datos ofrece una solución ante una posible situación empresarial:



Esta plantilla de flujo de datos está disponible en Enterprise Designer. Vaya a **Archivo > Nuevo > Flujo de datos > Desde plantilla** y seleccione **Parse Personal Name**

Este flujo de datos requiere lo siguiente:

- Módulo Universal Name
- Bases de datos de Open Parser

- Tablas de nombres mejoradas de Open Parser

En este flujo de datos, los datos se leen desde un archivo y se procesan mediante la etapa Open Name Parser. (La etapa Open Name Parser es parte del módulo Universal Naming). Para cada nombre, el flujo de datos hace lo siguiente:

Read from File

La etapa Read from File (Lectura desde archivo) identifica el nombre, la ubicación y el diseño del archivo que contiene los nombres que desea analizar. El archivo contiene tanto nombres masculinos como femeninos.

Open Name Parser

Open Name Parser examina los campos de nombre y los compara con los datos almacenados en los archivos de bases de datos de nombres de Spectrum™ Technology Platform. Basado en la comparación, analiza los datos de nombre en los campos Primer nombre, Segundo nombre y Apellido.

Write to File

La plantilla contiene una etapa Write to File (Escritura en archivo). Además de los campos de entrada, el archivo de salida contiene los campos FirstName, MiddleName, LastName, EntityType, GenderCode y GenderDeterminationSource.

Análisis de nombres árabes

Esta plantilla muestra cómo dividir los nombres árabes occidentalizados en las partes que los componen. La regla de análisis separa cada muestra en el campo **Nombre** y copia cada muestra en cinco campos: **Kunya**, **Ism**, **Laqab**, **Nasab**, **Nisba**. Estos campos de salida representan las cinco partes de un nombre en árabe, que se describen en el escenario de negocios.

Situación empresarial posible

Usted trabaja para un banco que quiere entender mejor el sistema de nombres árabes a fin de mejorar el servicio de atención a los clientes que hablan árabe. Ha recibido quejas de los clientes porque la información de facturación no contiene el nombre correcto del cliente. Para mejorar la relación con el cliente, el equipo de marketing en el que trabaja quiere apelar a los clientes que hablan árabe mediante campañas de marketing y asistencia telefónica.

Para entender el sistema de nombres árabes, hace una búsqueda y encuentra los siguientes recursos en Internet que explican tal sistema:

- en.wikipedia.org/wiki/Arabic_names
- heraldry.sca.org/laurel/names/arabic-naming2.htm

Los nombres árabes se basan en un sistema de asignación de nombres que incluye estas piezas: Ism, Kunya, Nasab, Laqab y Nisba.

- Ism es el nombre principal, o nombre personal, de una persona árabe.
- Por lo general, kunya hace referencia al hijo primogénito y sirve de sustituto de ism.
- Nasab es un patronímico o una serie de patronímicos. Indica el legado de una persona mediante la palabra ibn o bin, que significa hijo, y bint, que significa hija.
- Laqab es una descripción de la persona. Por ejemplo, al-Rashid significa "el honrado" o "el bien guiado" y al-Jamil significa "el bello".
- Nisba describe la ocupación de la persona, el área geográfica natal o la ascendencia (tribu, familia, etc.). Acompaña a una familia a lo largo de varias generaciones. El elemento nisba, entre todos los componentes del nombre árabe, es quizá el que más se asemeja al apellido occidental. Por ejemplo, al-Filistin significa "el palestino".

El siguiente flujo de datos ofrece una solución ante una posible situación empresarial:



Esta plantilla de flujo de datos está disponible en Enterprise Designer. Vaya a **Archivo > Nuevo > Flujo de datos > Desde plantilla** y seleccione **ParseArabicNames**. Este flujo de datos requiere el módulo Data Normalization.

En este flujo de datos, los datos se leen desde un archivo y se procesan mediante la etapa Open Parser (Analizador abierto). Para cada fila de datos del archivo de entrada, este flujo de datos hará lo siguiente:

Read from File

La etapa Read from File (Lectura desde archivo) identifica el nombre, la ubicación y el diseño del archivo que contiene los nombres que desea analizar. El archivo contiene tanto nombres masculinos como femeninos.

Open Parser

La etapa Open Parser (Analizador abierto) define si se utiliza una gramática de dominio específica de una cultura creada en Domain Editor o si se define una gramática independiente de dominio. Una gramática de análisis específica de una cultura creada en Domain Editor es una gramática de análisis validada que está asociada a una cultura y un dominio. Una gramática de análisis independiente de dominio creada en Open Parser es una gramática de análisis validada que no está asociada a una cultura ni un dominio.

En esta plantilla, la gramática de análisis se define como una gramática independiente de dominio.

La etapa Open Parser contiene una gramática de análisis que define los siguientes comandos y expresiones:

- `%Tokenize` está configurado con el carácter de espacio (`\s`). Esto significa que Open Parser utilizará el carácter de espacio para separar el campo de entrada en muestras. Por ejemplo, Abu Mohammed al-Rahim ibn Salamah contiene cinco fichas: Abu, Mohammed, al-Rahim, ibn y Salamah.

- %InputField está configurado para analizar los datos de entrada provenientes del campo **Nombre**.
- %OutputFields está configurado para copiar datos analizados en cinco campos: **Kunga, Ism, Laqab, Nasab y Nisba**.
- La expresión <root> define el patrón de los nombres árabes:
- Cero o una ocurrencia de **Kunya**
- Exactamente una o dos ocurrencias de **Ism**
- Cero o una ocurrencia de **Laqab**
- Cero o una ocurrencia de **Nasab**
- Cero o más ocurrencias de **Nisba**

Las variables de regla que definen el dominio deben utilizar los mismos nombres que los campos de salida definidos en el comando OutputFields requerido.

La gramática de análisis utiliza una combinación de expresiones regulares y cuantificadores de expresión para construir un patrón de nombres árabes. La gramática de análisis utiliza estos caracteres especiales:

- El carácter "?" significa que una expresión regular puede ocurrir cero veces o una vez.
- El carácter "*" significa que una expresión regular puede ocurrir cero o más veces.
- El carácter ";" indica el final de una regla.

Utilice la ficha **Comandos** para explorar el significado del resto de los símbolos especiales que puede utilizar en las gramáticas de análisis pasando el mouse sobre la descripción.

De forma predeterminada, los cuantificadores son expansivos. Se dice que son expansivos porque la expresión acepta todas las muestras posibles, pero a la vez, permite un cruce exitoso. Es posible invalidar este comportamiento agregando '?' en los cruces reacios o '+' en los cruces posesivos. Se los llama cruces reacios porque la expresión acepta la menor cantidad de muestras posibles, pero a la vez, permite un cruce exitoso. En el caso de los cruces posesivos, la expresión acepta todas las muestras posibles aunque ello impida un cruce.

Para probar la gramática de análisis, haga clic en la ficha Vista previa. Ingrese los nombres que se muestran más abajo en el campo **Nombre** y haga clic en **Vista previa**.

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

También es posible ingresar otros nombres válidos y no válidos para ver cómo se analizan los datos de entrada.

Es posible utilizar la función de seguimiento para ver una representación gráfica de los resultados finales de análisis o para recorrer los eventos de análisis. Haga clic en el enlace de la columna **Seguimiento** para ver los Detalles de seguimiento de la fila de datos.

Write to File

La plantilla contiene una etapa Write to File (Escritura en archivo). Además del campo de entrada, el campo de salida contiene los campos **Kunya, lsm, Laqab, Nasab, y Nisba**.

Análisis de nombres chinos

Esta plantilla muestra cómo dividir los nombres chinos en las partes que los componen. La regla de análisis separa cada muestra del campo **Nombre** y copia cada muestra a dos campos: **Apellido** y **Primer nombre**.

Situación empresarial posible

Usted trabaja para una empresa de servicios financieros que quiere explorar si es posible incluir los caracteres chinos en varias comunicaciones con los clientes que hablan chino.

Para entender el sistema de nombres chinos, hace una búsqueda y encuentra este recurso en Internet, que explica cómo están conformados los nombres chinos:

en.wikipedia.org/wiki/Chinese_names

El siguiente flujo de datos ofrece una solución ante una posible situación empresarial:



Esta plantilla de flujo de datos está disponible en Enterprise Designer. Vaya a **Archivo > Nuevo > Flujo de datos > A partir de plantilla** y seleccione **Analizar nombre chinos**. Este flujo de datos requiere el módulo Data Normalization.

En este flujo de datos, los datos se leen desde un archivo y se procesan mediante la etapa Open Parser (Analizador abierto). Para cada fila de datos del archivo de entrada, este flujo de datos hará lo siguiente:

Read from File

La etapa Read from File (Lectura desde archivo) identifica el nombre, la ubicación y el diseño del archivo que contiene los nombres que desea analizar. El archivo contiene tanto nombres masculinos como femeninos.

Open Parser

La etapa Open Parser (Analizador abierto) define si se utiliza una gramática de dominio específica de una cultura creada en Domain Editor o si se define una gramática independiente de dominio. Una gramática de análisis específica de una cultura creada en Domain Editor es una gramática de análisis validada que está asociada a una cultura y un dominio. Una gramática de análisis

independiente de dominio creada en Open Parser es una gramática de análisis validada que no está asociada a una cultura ni un dominio.

En esta plantilla, la gramática de análisis se define como una gramática independiente de dominio.

La etapa Open Parser contiene una gramática de análisis que define los siguientes comandos y expresiones:

- %Tokenize está configurado en Ninguno. Cuando se define Tokenize en None, la regla gramatical de análisis debe incluir todos los espacios u otros separadores de muestra al interior de su definición de regla.
- %InputField está configurado para analizar los datos de entrada provenientes del campo **Nombre**.
- %OutputFields está configurado para copiar los datos analizados a dos campos: **LastName** y **FirstName**.

La expresión <root> define el patrón de los nombres chinos:

- Una ocurrencia de **Apellido**
- Una a tres ocurrencias de **Primer nombre**

Las variables de regla que definen el dominio deben utilizar los mismos nombres que los campos de salida definidos en el comando OutputFields requerido.

La variable de la regla CJKCharacter define el patrón de caracteres para chino, japonés y coreano (CJK, según la inicial en inglés de cada idioma). El patrón de caracteres está definido para utilizar solo caracteres que son letras. La regla es:

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

- La expresión regular \p{InX} se utiliza para indicar un bloque Unicode para una determinada cultura, en la que X es la cultura. En este ejemplo, la cultura es CJKUnifiedIdeographs.
- En las expresiones regulares, una clase de carácter es un conjunto de caracteres que se desea cruzar. Por ejemplo, [aeiou] es la clase de carácter que contiene solo vocales. Las clases de caracteres pueden aparecer dentro de otras clases de caracteres y pueden estar compuestas por la operación de unión (implícitas) y el operador de intersección (&&). El operador de unión indica una clase que contiene cada carácter que se encuentra en al menos una de las clases de operandos. El operador de intersección indica una clase que contiene cada carácter que se superpone con los bloques Unicode intersecados.
- La expresión regular \p{L} se utiliza para indicar el bloque Unicode que incluye solo letras.

Para probar la gramática de análisis, haga clic en la ficha Vista previa. Ingrese los nombres que se muestran más abajo en el campo **Nombre** y haga clic en **Vista previa**.

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

También es posible ingresar otros nombres válidos y no válidos para ver cómo se analizan los datos de entrada.

Es posible utilizar la función de seguimiento para ver una representación gráfica de los resultados finales de análisis o para recorrer los eventos de análisis. Haga clic en el enlace de la columna **Seguimiento** para ver los Detalles de seguimiento de la fila de datos.

Write to File

La plantilla contiene una etapa Write to File (Escritura en archivo). Además del campo de entrada, el archivo de salida contiene los campos **Apellido** y **Primer nombre**. Seleccione todos los resultados de la Lista Resultados de cruce y haga clic en **Eliminar**.

Análisis de nombres españoles y alemanes

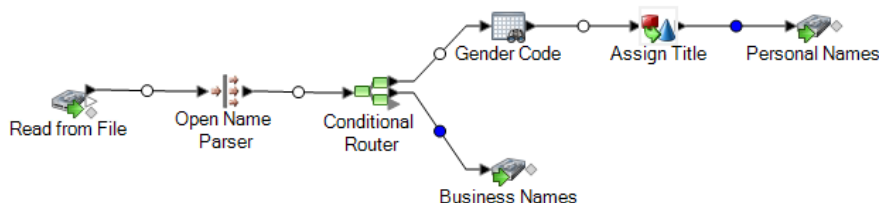
Esta plantilla muestra cómo dividir nombres de varias culturas, por ejemplo, nombres españoles y alemanes, en las partes que los componen. La regla de análisis separa cada muestra en el campo **Nombre** y copia cada muestra en los campos definidos en la gramática de análisis de nombres personales y nombres de empresas. Para obtener más información sobre esta gramática de análisis, seleccione **Herramientas > Editor del dominio Open Parser** y luego seleccione el dominio de **Nombres personales y de empresas** y cualquiera de las culturas en **Alemán (de)** o **Español (es)**.

Esta plantilla también aplica los códigos de género a los nombres personales cuando se utilizan datos de tablas contenidos en Administración de tablas. Para obtener más información sobre Administración de tablas, seleccione **Herramientas > Administración de tablas**.

Situación empresarial posible

Usted trabaja para una empresa farmacéutica ubicada en Bruselas que ha consolidado sus operaciones en Alemania y España. Su empresa quiere implementar una base de datos de varias culturas que contenga datos de nombre, y es su trabajo analizar las variantes de nombre que existen entre las dos culturas.

El siguiente flujo de datos ofrece una solución ante una posible situación empresarial:



Esta plantilla de flujo de datos está disponible en Enterprise Designer. Vaya a **Archivo > Nuevo > Flujo de datos > Desde plantilla** y seleccione **ParseSpanish&GermanNames**. Este flujo de datos requiere el módulo Data Normalization.

En este flujo de datos, los datos se leen desde un archivo y se procesan mediante la etapa Open Parser (Analizador abierto). Para cada fila de datos del archivo de entrada, este flujo de datos hará lo siguiente:

Read from File

La etapa Read from File (Lectura desde archivo) identifica el nombre, la ubicación y el diseño del archivo que contiene los nombres que desea analizar. El archivo contiene tanto nombres masculinos como femeninos e incluye información de CultureCode para cada nombre. La información de CultureCode designa los nombres de entrada como alemanes (de) o españoles (es).

Open Name Parser

Open Name Parser examina los campos de nombre y los compara con los datos almacenados en los archivos de bases de datos de nombres de Spectrum™ Technology Platform. Basado en la comparación, analiza los datos de nombre en los campos Primer nombre, Segundo nombre y Apellido.

Conditional Router

La etapa Conditional Router (Enrutador condicional) envía la entrada para que los nombres personales se dirijan a la etapa Gender Codes (Códigos de género) y los nombres de empresas se dirijan a la etapa Business Names (Nombres de empresas).

Código de género

Haga doble clic en el lienzo de esta etapa y, luego, haga clic en **Modificar** para mostrar las opciones de regla de búsqueda en tablas.

La opción **Categorizar** utiliza el valor Fuente como clave y copia el valor correspondiente de la entrada de la tabla en el campo seleccionado en la lista Destino. En esta plantilla, **Completar campo** se selecciona y **Fuente** se configura para utilizar el campo **FirstName**. Table Lookup trata todo el campo como una cadena y marca el registro si es posible categorizar la cadena en su totalidad.

La opción **Destino** está configurada según el campo **Códigos de género** y utiliza los términos de búsqueda de la tabla **Códigos de género** para realizar la categorización de nombres masculinos y femeninos. Si no se encuentra un término en los datos de entrada, Table Lookup le asigna un valor **U**, en referencia a la palabra "unknown" (desconocido). Para entender mejor cómo funciona esta opción, seleccione **Herramientas > Administración de tablas** y seleccione la tabla Códigos de género.

Write to File

La plantilla contiene dos etapas Write to File, una para nombres personales y una para nombres de las empresas. Además del campo de entrada, el archivo de salida de nombres personales contiene los campos **Name**, **TitleOfRespect**, **FirstName**, **MiddleName**, **LastName**, **PaternalLastName**, **MaternalLastName**, **MaturitySuffix**, **GenderCode**, **CultureUsed**, y **ParserScore**

El archivo de salida de nombres de empresas contiene los campos **Name**, **FirmName**, **FirmSuffix**, **CultureUsed**, y **ParserScore**

Análisis de direcciones de correo electrónico

Esta plantilla muestra cómo dividir las direcciones de correo electrónico en las partes que la componen. La regla de análisis separa cada muestra del campo **Correo electrónico** y copia las distintas muestras a tres campos: **Parte local**, **Nombre de dominio** y **extensión del dominio**. **Parte local** representa la parte del nombre del dominio de la dirección de correo electrónico, **Nombre de dominio** representa el nombre del dominio de la dirección de correo electrónico y **Extensión del dominio** representa la extensión del dominio de la dirección de correo electrónico. Por ejemplo, en `pb.com`, "pb" es el nombre de dominio y ".com" es la extensión de dominio.

Internet es una gran fuente de información de dominio público que puede ayudarlo en sus tareas de análisis abierto. En este ejemplo, la información de formato de correo electrónico se obtuvo a través de varios recursos de Internet y luego se importó en Administración de tablas para crear una tabla de valores de dominio. La tarea de extensión de dominio que se realiza en esta actividad de plantilla muestra la utilidad de este método.

Esta plantilla también muestra cómo utilizar de forma efectiva los datos de las tablas que se cargan en Administración de tablas para realizar búsquedas en ellas como parte de las tareas de análisis.

Situación empresarial posible

Usted trabaja para una compañía de seguros que quiere realizar su primera campaña de marketing por correo electrónico. Su base de datos contiene direcciones de correo electrónico de sus clientes y se le ha solicitado buscar la forma de asegurarse de que esas direcciones de correo electrónico tengan un formato SMTP válido.

Antes de crear este flujo de datos, es necesario cargar una tabla con extensiones de nombres de dominio válidos en Administración de tablas para poder buscar las extensiones del nombre de dominio como parte del proceso de validación.

El siguiente flujo de datos ofrece una solución ante una posible situación empresarial:



Esta plantilla de flujo de datos está disponible en Enterprise Designer. Vaya a **Archivo > Nuevo > Flujo de datos > A partir de plantilla** y seleccione **Analizar correo electrónico**. Este flujo de datos requiere el módulo Data Normalization.

En este flujo de datos, los datos se leen desde un archivo y se procesan mediante la etapa Open Parser (Analizador abierto). Para cada fila de datos del archivo de entrada, este flujo de datos hará lo siguiente:

Creación de una tabla de extensiones de dominio

La primera tarea es crear una tabla Open Parser en Administración de tareas que pueda usarse para controlar si las extensiones de dominio en sus direcciones de correo electrónico son válidas.

1. En el menú **Herramientas**, seleccione **Administración de tablas**.
2. En la lista **Tipo**, seleccione **Open Parser**.
3. Haga clic en **Nuevo**.
4. En el cuadro de diálogo **Agregar tabla definida por el usuario**, ingrese `EmailDomains` en el campo **Nombre de tabla**; asegúrese de que esté seleccionado **Ninguna** esté seleccionada en la lista **Copiar desde** y después haga clic en **Aceptar**.
5. Estando mostrado **Dominios de correo electrónico** en la lista **Nombre**, haga clic en **Importar**.
6. En el cuadro de diálogo **Importar**, haga clic en **Examinar** y ubique el archivo fuente de la tabla. La ubicación predeterminada es: `<drive>:\Program Files\Pitney Bowes\Spectrum\server\modules\coretemplates\data\Email_Domains.txt`. Administración de tablas muestra una vista previa de los términos que se encuentran en el archivo de importación.
7. Haga clic en **Aceptar**. Administración de tablas importa los archivos fuente y muestra una lista de extensiones de dominio de Internet.
8. Haga clic en **Cerrar**. Se crea la tabla `EmailDomains`. Luego, cree el flujo de datos utilizando la plantilla `ParseEmail`.

Read from File

Esta etapa Read from File identifica el nombre, la ubicación y el diseño del archivo que contiene las direcciones de correo electrónico que desea analizar.

Open Parser

La gramática de análisis de la etapa Open Parser (Analizador abierto) define los siguientes comandos y expresiones:

- `%Tokenize` está configurado en Ninguno. Cuando se define `Tokenize` en `None`, la regla gramatical de análisis debe incluir todos los espacios u otros separadores de muestra al interior de su definición de regla.
- `%InputField` está configurado para analizar los datos de entrada provenientes del campo **Email_Address**.
- `%OutputFields` está configurado para copiar los datos analizados a tres campos: **Local-Part**, **DomainName** y **DomainExtension**.
- La expresión de raíz define el patrón de muestras que se analiza:

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

Las variables de regla que definen el dominio deben utilizar los mismos nombres que los campos de salida definidos en el comando `OutputFields` requerido.

- El resto de la gramática de análisis define cada una de las variables de regla como expresiones.

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

La variable `<Local-Part>` está definida como una cadena de texto que contiene la variable `<alphanum>`, el carácter de punto y otra variable `<alphanum>`.

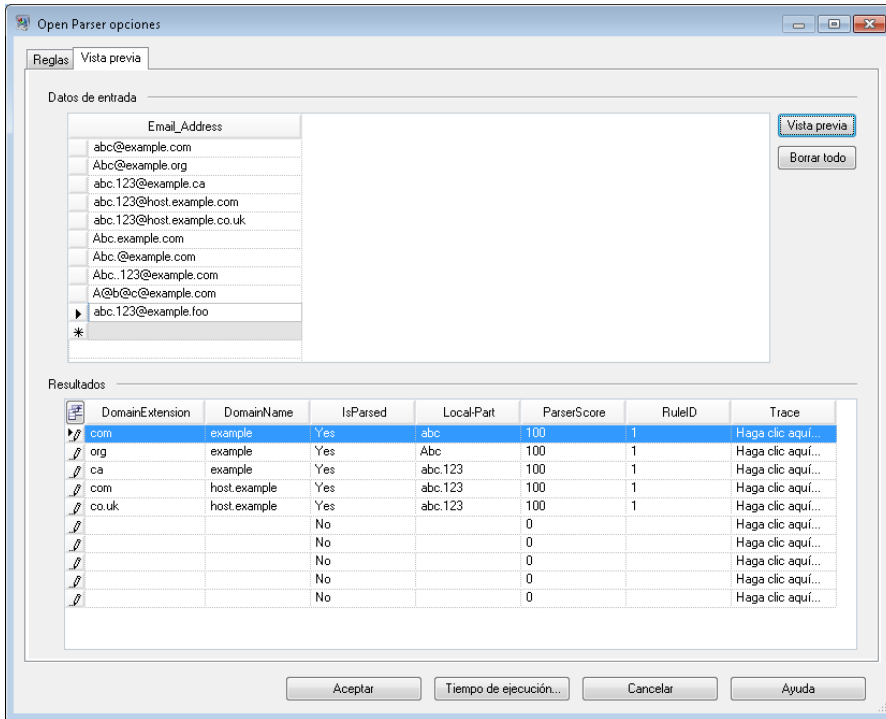
La definición de la variable `<alphanum>` es una expresión regular que coincide con cualquier cadena de caracteres de la A a la Z, de la "a" a la "z" y del 0 al 9. La variable `<alphanum>` se utiliza a lo largo de esta gramática de análisis y está definida una vez en la última línea de la gramática de análisis.

La gramática de análisis utiliza una combinación de expresiones regulares y caracteres literales para construir un patrón de direcciones de correo electrónico. Los caracteres entre comillas dobles en esta gramática de análisis son caracteres literales, el nombre de una tabla utilizada para búsquedas o una expresión regular. La gramática de análisis utiliza estos caracteres especiales:

- El carácter "+" significa que una expresión regular puede ocurrir una o más veces.
- El carácter "?" significa que una expresión regular puede ocurrir cero veces o una vez.
- El carácter "|" significa que la variable tiene una condición OR.
- El carácter ";" indica el final de una regla.

Utilice la ficha **Comandos** para explorar el significado del resto de los símbolos especiales que puede utilizar en las gramáticas de análisis pasando el mouse sobre la descripción.

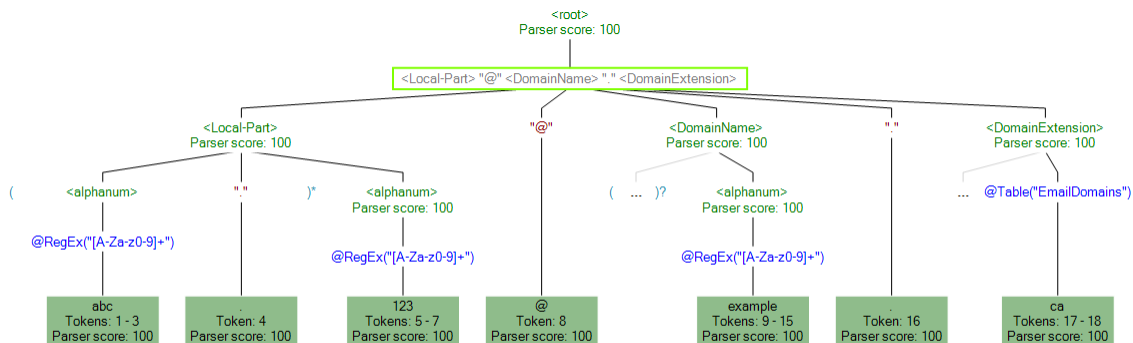
Para probar la gramática de análisis, haga clic en la ficha Vista previa. Ingrese en el campo **Dirección de correo electrónico** las direcciones de correo electrónico que se muestran más adelante, y después haga clic en **Vista previa**.



También puede ingresar otras direcciones de correo electrónico para ver cómo se analizan los datos de entrada.

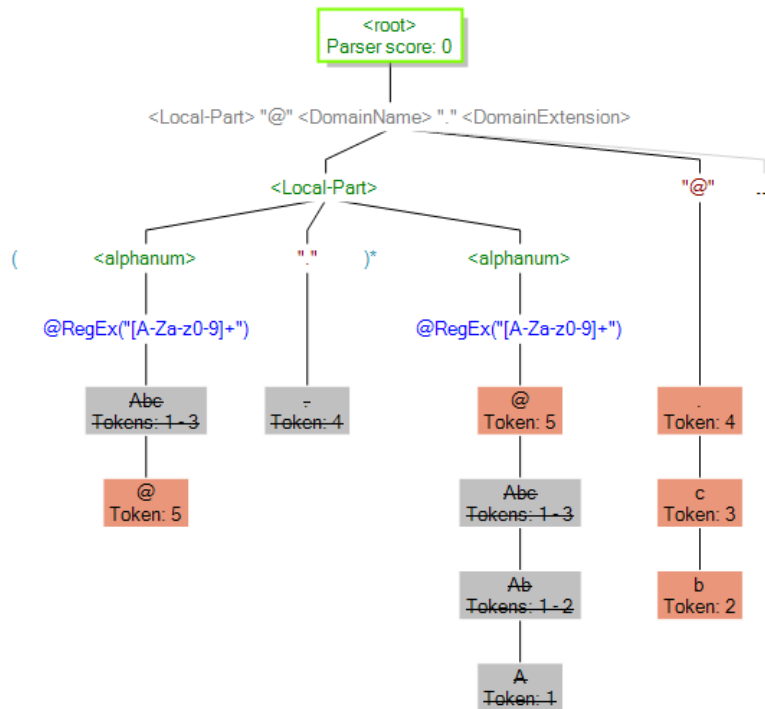
También es posible utilizar la función de seguimiento para ver una representación gráfica de los resultados finales de análisis o para recorrer los eventos de análisis. Haga clic en el enlace de la columna **Seguimiento** para ver los Detalles de seguimiento de la fila de datos.

En Detalles de seguimiento se muestra un resultado de comparación. Compare las muestras cruzadas de cada expresión de la gramática de análisis.



También puede utilizar la función de seguimiento para ver resultados sin coincidencias. El siguiente gráfico muestra un resultado sin coincidencias. Compare las muestras cruzadas de cada expresión de la gramática de análisis. La razón por la que estos datos de entrada (Abc . example . com) no

coincidieron es que no contienen todas las muestras requeridas para el cruce: no hay un carácter @ que separe la muestra Local- Part y las muestras de dominio..



Write to File

La plantilla contiene una etapa Write to File (Escritura en archivo). Además del campo de entrada, el archivo de salida contiene los campos **Parte local**, **Nombre de dominio**, **Extensión del dominio**, **IsParsed** y **ParserScore**.

Análisis de números telefónicos de Estados Unidos

Esta plantilla muestra la manera de analizar números telefónicos de EE. UU. en sus segmentos componentes. La regla de análisis separa cada muestra en el campo **Número de teléfono** y copia las distintas muestras en cuatro campos: **Código de país**, **Código de área**, **Central** y **Número**.

Situación empresarial posible

Usted trabaja para un proveedor de servicios inalámbricos y le han asignado la tarea de analizar los datos de números telefónicos entrantes de una región en crecimiento de su sector empresarial.

El siguiente flujo de datos ofrece una solución ante una posible situación empresarial:



Esta plantilla de flujo de datos está disponible en Enterprise Designer. Vaya a **Archivo > Nuevo > Flujo de datos > A partir de plantilla** y seleccione **Analizar número telefónico de EE.UU.**. Este flujo de datos requiere el módulo Data Normalization.

En este flujo de datos, los datos se leen desde un archivo y se procesan mediante la etapa Open Parser (Analizador abierto). Para cada fila de datos del archivo de entrada, este flujo de datos hará lo siguiente:

Read from File

Esta etapa identifica el nombre de archivo, la ubicación y el diseño del archivo que contiene los números de teléfono que desea analizar.

Open Parser

La etapa Open Parser (Analizador abierto) define si se utiliza una gramática de dominio específica de una cultura creada en Domain Editor o si se define una gramática independiente de dominio. Una gramática de análisis específica de una cultura creada en Domain Editor es una gramática de análisis validada que está asociada a una cultura y un dominio. Una gramática de análisis independiente de dominio creada en Open Parser es una gramática de análisis validada que no está asociada a una cultura ni un dominio.

En esta plantilla, la gramática de análisis se define como una gramática independiente de dominio.

La etapa Open Parser contiene una gramática de análisis que define los siguientes comandos y expresiones:

- `%Tokenize` está configurado en Ninguno. Cuando se define `Tokenize` en None, la regla gramatical de análisis debe incluir todos los espacios u otros separadores de muestra al interior de su definición de regla.
- `%InputField` está configurado para analizar los datos de entrada provenientes del campo **PhoneNumber**.
- `%OutputFields` está configurado para segmentar en cuatro campos los datos analizados: **CountryCode, AreaCode, Intercambiar y Número**.
- La expresión `<root>` define el patrón de las muestras que se analizan, e incluye instrucciones OR (|), de modo que un número telefónico válido sea:
- **Código de país, Código de área, Central y Número O BIEN**
- **Código de área, Central y Número O BIEN**
- **Central y Número**

La gramática de análisis utiliza una combinación de expresiones regulares y caracteres literales para construir un patrón de números telefónicos. Los caracteres entre comillas dobles en esta gramática de análisis son caracteres literales o una expresión regular.

El carácter más (+) que se utiliza en este comando `<root>` se define como un carácter literal porque está encerrado entre comillas. Es posible utilizar comillas simples o dobles para indicar que un carácter es literal. Si el signo más se utiliza sin comillas, significa que la expresión que precede al signo puede ocurrir una o más veces.

Las reglas de dominio del número telefónico están definidas para coincidir con los siguientes patrones de caracteres:

- Cero o una ocurrencia de un carácter "+".
- La regla **Código de país**, que es un dígito único entre 0 y 9.
- Cero o una ocurrencia de un carácter de paréntesis abierto, un guion o un espacio. Si dos de estos caracteres aparecen en secuencia, se produce un resultado sin coincidencias o, en otras palabras, un número telefónico no válido.
- La regla **Código de área**, que es una secuencia de exactamente tres dígitos entre 0 y 9.
- Cero o una ocurrencia de un carácter de paréntesis abierto, un guion o un espacio. Si dos de estos caracteres aparecen en secuencia, se produce un resultado sin coincidencias o, en otras palabras, un número telefónico no válido.
- La regla **Central**, que es una secuencia de exactamente tres dígitos entre 0 y 9.
- Cero o una ocurrencia de un carácter de paréntesis abierto, un guion o un espacio. Si dos de estos caracteres aparecen en secuencia, se produce un resultado sin coincidencias o, en otras palabras, un número telefónico no válido.
- La regla **Número**, que es una secuencia de exactamente cuatro dígitos entre 0 y 9.

Las variables de regla que definen el dominio deben utilizar los mismos nombres que los campos de salida definidos en el comando `OutputFields` requerido.

Expresiones regulares y cuantificadores de expresión

La gramática de análisis utiliza una combinación de expresiones regulares y de cuantificadores de expresión a fin de construir un patrón de números telefónicos de Estados Unidos. La gramática de análisis utiliza estos caracteres especiales:

- El carácter "?" significa que una expresión regular puede ocurrir cero veces o una vez.
- El carácter (|) indica una condición OR.
- El carácter ";" indica el final de una regla.

Utilice la ficha **Comandos** para explorar el significado del resto de los símbolos especiales que puede utilizar en las gramáticas de análisis pasando el mouse sobre la descripción.

Utilización de la ficha Vista previa

Para probar la gramática de análisis, haga clic en la ficha Vista previa. Ingrese los números telefónicos que se muestran más abajo en el campo **Número telefónico** y después haga clic en **Vista previa**.

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

También es posible ingresar otros números telefónicos válidos y no válidos para ver cómo se analizan los datos de entrada.

También es posible utilizar la función de seguimiento para ver una representación gráfica de los resultados finales de análisis o para recorrer los eventos de análisis. Haga clic en el enlace de la columna **Seguimiento** para ver los Detalles de seguimiento de la fila de datos.

Write to File

La plantilla contiene una etapa Write to File (Escritura en archivo). Además del campo de entrada, el archivo de salida incluye los campos **Código de país**, **Código de área**, **Central** y **Número**.

3 - Plantillas de flujo de datos para comparación

In this section

Normalización de la terminología	59
Estandarización de nombres personales	60
Plantillas para estandarización	62

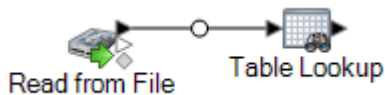
Normalización de la terminología

El uso poco consistente de la terminología puede ser un problema para la calidad de los datos, y puede ocasionar problemas en el análisis y en las búsquedas, entre otros. Puede generar un flujo de datos que busque entre sus datos los términos que se usan en forma poco consistente y los normalice. Por ejemplo, si sus datos incluyen los términos "Incorporated", "Inc.", e "Inc" en los nombres de la empresa, puede generar un flujo de datos para normalizar todo en una forma única (por ejemplo, "Inc.").

Nota: Antes de realizar este procedimiento, su administrador debe instalar la base de datos del módulo Data Normalization, la que contendrá los términos normalizados que desea aplicar a sus datos. Las instrucciones para la instalación de bases de datos pueden encontrarse en la *Guía de instalación*.

1. Genere un nuevo flujo de datos en Enterprise Designer.
2. Arrastre hacia el lienzo una etapa de origen.
3. Haga doble clic en la etapa de origen y configúrela. Consulte la *Guía de Dataflow Designer* para obtener instrucciones acerca de cómo configurar etapas de origen.
4. Arrastre hacia el lienzo una etapa Table Lookup y conéctela a la etapa de origen.

Por ejemplo, si estuviera utilizando una etapa de origen Read from File, su flujo de datos tendría el siguiente aspecto:

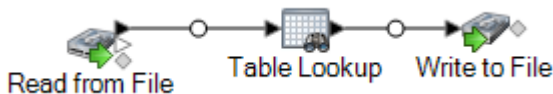


5. Haga doble clic en la etapa Table Lookup en el lienzo.
6. Para especificar las opciones de Table Lookup, debe crear una regla. Puede crear múltiples reglas y luego especificar el orden en que quiere aplicarlas. Haga clic en **Agregar** para crear una regla.
7. En el campo **Acción**, deje seleccionada la opción predeterminada **Normalizar**.
8. En el campo **Activo**, deje seleccionado **Completar campo** si el campo completo es el término que desea normalizar. O bien, seleccione **Términos individuales dentro de un campo** para normalizar palabras individuales en dicho campo.
9. En el campo **Origen**, seleccione el campo que desea normalizar.
10. En el campo **Destino**, seleccione el campo que desea que contenga el término normalizado. Si especifica como campo de origen el mismo campo, el valor del campo de origen será reemplazado por el término normalizado.
11. En el campo **Tabla**, seleccione la tabla que contenga los términos normalizados.

Nota: Si no ve la tabla que necesita, comuníquese con su administrador de sistema. Debe estar cargada la base de datos del módulo Data Normalization.

12. En el campo **Cuando no se encuentre la entrada de tabla, defina el valor de destino en,** seleccione **Valor del origen**.
13. Haga clic en **Aceptar**.
14. Defina reglas adicionales si desea normalizar valores en más campos. Cuando haya terminado de definir reglas, haga clic en **Aceptar**.
15. Arrastre hacia el lienzo una etapa receptora y conéctela a Table Lookup.

Por ejemplo, si estuviese usando Write to File, su flujo de datos tendrá el siguiente aspecto:



16. Haga doble clic en la etapa receptora y configúrela.

Para obtener información acerca de cómo configurar etapas receptoras, consulte la *Guía de Dataflow Designer*.

Ahora tiene un flujo de datos que normaliza términos.

Estandarización de nombres personales

Este procedimiento muestra la forma de generar un flujo de datos que toma datos de nombres de personas (por ejemplo, "John P. Smith"), identifica apodos comunes para el mismo nombre, y genera una versión normalizada del nombre que puede utilizarse entonces para consolidar registros redundantes.

Nota: Antes de comenzar, cerciórese de que sus datos de entrada tengan un campo denominado "Nombre" que contenga el nombre completo de la persona.

1. Si aún no lo ha hecho, cargue las siguientes tablas en el servidor Spectrum™ Technology Platform:

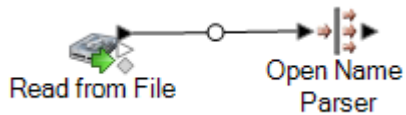
- Base de Open Parser
- Nombres mejorados de Open Parser

Use la utilidad de carga de bases de datos del módulo Data Normalization para cargar estas tablas. Para obtener instrucciones sobre la carga de tablas, consulte la *Guía de instalación*.

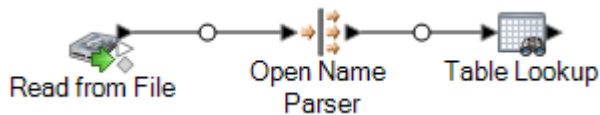
2. Genere un nuevo flujo de datos en Enterprise Designer.
3. Arrastre hacia el lienzo una etapa de origen.

4. Haga doble clic en la etapa de origen y configúrela. Consulte la *Guía de Dataflow Designer* para obtener instrucciones acerca de cómo configurar etapas de origen.
5. Arrastre hacia el lienzo una etapa Open Name Parser y conéctela a la etapa de origen.

Por ejemplo, si está utilizando la etapa Read from File, su flujo de datos tendrá el siguiente aspecto:



6. Arrastre hacia el lienzo la etapa Table Lookup y conéctela a la etapa Open Name Parser.
- Su flujo de datos debería tener ahora el siguiente aspecto:

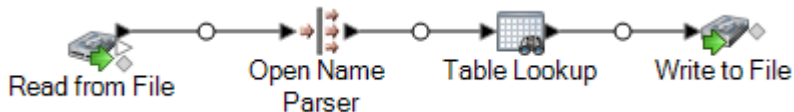


7. Haga doble clic en la etapa Table Lookup en el lienzo.
8. En el campo **Origen**, seleccione **FirstName**.
9. En el campo **Destino**, seleccione **FirstName**.

Al especificar el mismo campo tanto como origen y como destino, el campo se actualizará con la versión normalizada del nombre.

10. En el campo **Tabla**, seleccione **NickNames.xml**.
11. Haga clic en **Aceptar**.
12. Haga clic otra vez en **Aceptar** para cerrar la ventana de **Opciones de Table Lookup**.
13. Arrastre hacia el lienzo una etapa receptora y conéctela a la etapa Table Lookup.

Por ejemplo, si estuviera utilizando una etapa receptora Write to File, su flujo de datos tendría ahora el siguiente aspecto:



14. Haga doble clic en la etapa receptora y configúrela. Consulte la *Guía de Dataflow Designer* para obtener instrucciones acerca de cómo configurar etapas de origen.

Ahora tiene un flujo de datos que toma nombres de personas y normaliza los nombres de pila, reemplazando los apodos con la forma normalizada del nombre.

Plantillas para estandarización

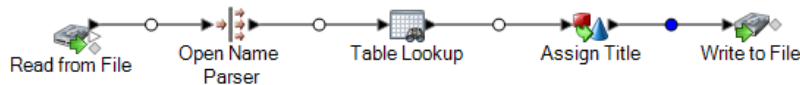
Formalización de análisis de nombres personales

Esta plantilla de flujo de datos muestra cómo tomar datos de nombres personales (por ejemplo "Juan P. Smith"), identificar apodos comunes del mismo nombre, y crear una versión estándar de los nombres que luego puede ser utilizada para consolidar registros repetidos. También muestra cómo se puede agregar datos de título de respeto basándose en datos sobre género.

Situación empresarial posible

Usted trabaja para una organización sin fines de lucro que quiere enviar invitaciones para una fiesta. Sus datos de entrada incluyen nombres completos y quiere dividirlos en los campos de nombre, segundo nombre y apellido, y agregar un campo de tratamiento para que sus invitaciones sean más formales. También quiere reemplazar cualquier apodo de los datos de nombre para usar una variante más formal del nombre.

El siguiente flujo de datos ofrece una solución ante una posible situación empresarial:



Esta plantilla de flujo de datos está disponible en Enterprise Designer. Vaya a **Archivo > Nuevo > Flujo de datos > Desde plantilla** y seleccione **StandardizePersonalNames**. Este flujo de datos requiere de los módulos Data Normalization y Universal Name.

Para cada fila de datos del archivo de entrada, este flujo de datos hará lo siguiente:

Read from File

La etapa Read from File (Lectura desde archivo) identifica el nombre, la ubicación y el diseño del archivo que contiene los nombres que desea analizar. El archivo contiene tanto nombres masculinos como femeninos.

Name Parser

En esta plantilla, la etapa Name Parser (Analizador de nombre) se denomina Parse Personal Name (Análisis de nombres personales). La etapa Parse Personal Name examina los campos de nombre y los compara con los datos almacenados en los archivos de bases de datos de nombres de Spectrum™ Technology Platform. Según la comparación, divide los datos de nombre en los campos de nombre, segundo nombre y apellido, y asigna un tipo de entidad y un género a cada nombre. También utiliza el reconocimiento de patrones además de los datos de nombre.

En esta plantilla, la etapa Parse Personal Name se configura de la siguiente manera.

- Se selecciona la opción Analizar nombres personales y se desmarca la opción Analizar nombres de empresas. Cuando seleccionan estas opciones, los nombres de pila se evalúan en cuanto a género, orden, puntuación y no se realiza una evaluación de nombres de empresas.
- El Origen de determinación de género está configurado de manera predeterminada. En la mayoría de los casos, la configuración predeterminada es la mejor opción para determinar el género porque cubre una amplia variedad de nombres. De todas maneras, si está procesando nombres de una cultura específica, seleccione esa cultura. La selección de una cultura específica contribuye a garantizar la asignación del género correcto a los nombres. Por ejemplo, si se deja la opción predeterminada seleccionada, el nombre Jean será identificado como un nombre femenino. Sin embargo, si selecciona francés, será identificado como un nombre masculino.
- El orden configurado es el natural. Los campos de nombre están ordenados por tratamiento, nombre, segundo nombre, apellido y sufijo.
- La opción Retener puntos está desmarcada. No se mantiene ninguna puntuación en los datos de nombres.

Transformer

En esta plantilla, la etapa Transformer (Transformador) se denomina Assign Titles (Asignación de tratamientos). La etapa Assign Titles utiliza una secuencia de comandos personalizada para examinar cada fila en la transmisión de datos enviados por la etapa Parse Personal Name y les asigna un valor **TitleOfRespect** según el valor **GenderCode**.

La secuencia de comandos personalizada:

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

Cada vez que la etapa Assign Titles encuentra **M** en el campo **GenderCode**, define el valor para **TitleOfRespect** como **Mr**. Cada vez que las etapas Assign Titles encuentran **F** en el campo **GenderCode**, define el valor para **TitleOfRespect** como **Ms**.

Plantillas de flujo de datos para comparación

En esta plantilla, la etapa Standardization (Estandarización) se denomina Standardize Nicknames (Estandarización de apodos). La etapa Standardize Nickname busca nombres en la base de datos Nicknames.xml y reemplaza cualquier apodo con la forma más común del nombre. Por ejemplo, el nombre Tommy es reemplazado por Thomas.

Write to File

La plantilla contiene una etapa Write to File (Escritura en archivo). Además de los campos de entrada, el archivo de salida contiene los campos TitleOfRespect, FirstName, MiddleName, LastName, EntityType, GenderCode y GenderDeterminationSource.

4 - Comparación

In this section

Coincidencia de terminología	65
Campos estándares	66
Técnicas para la definición de claves de cruce	68
Reglas de cruce	71
Cruce de registros de un origen único	85
Comparación de registros de un origen a otro origen	91
Coincidencia de registros entre y dentro orígenes	98
Comparación de registros con una base de datos	104
Comparación de registros con varias reglas de cruce	107
Creación de un servicio de comparación universal	111
Uso de una clave de cruce inmediato	115
Análisis de resultados de cruce	119
Plantillas de flujo de datos para comparación	135

Coincidencia de terminología

Calificación promedio	La media coincide con puntuación de todos los duplicados. Los valores posibles son 0-100; 0 indica una coincidencia poco satisfactoria y 100, una coincidencia exacta.
Punto de referencia	Resultado de cruce seleccionado que se comparará con otro resultado de cruce.
Grupo de candidatos	Registros sospechosos y candidatos agrupados en función de una ID asignada por Candidate Finder (Localizador de candidatos). El registro sospechoso (el primer registro del grupo) es aquel leído de una fuente de entrada mientras que los registros candidatos suelen encontrarse en una base de datos mediante una consulta SQL.
Registros de candidatos	Todos los registros no sospechosos de un grupo de cruce o un grupo candidato.
Caída	Reducción de los duplicados.
Registro de cruce detallado	Registro único que corresponde a un registro procesado por una etapa de cruce. Cada registro brinda información sobre su tipo de registro (si es sospechoso, único o duplicado), como también información sobre el grupo de cruce, el grupo de candidatos y la colección de salida. Los registros candidatos presentan información sobre la razón por la que el registro de entrada coincide o no con su sospechoso.
Colecciones duplicadas	Una colección duplicada consiste en un registro sospechoso y sus registros duplicados agrupados mediante CollectionNumber. Los registros únicos siempre corresponden a CollectionNumber 0.
Registros duplicados	Un registro duplicado es aquel que coincide con otro dentro de un grupo de cruce. Puede ser un sospechoso o un candidato.
Cruces inmediatos	Un cruce inmediato se realiza cuando un sospechoso y un candidato tienen una coincidencia exacta en los contenidos de un campo designado; por lo general, es un campo ExpressMatchKey provisto por Match Key Generator (Generador de clave de cruce). Si se realiza un cruce inmediato, no se realizan más procesamientos para determinar si el sospechoso y el candidato son duplicados.
Registros de entrada	Orden de los registros en la etapa de comparación antes de que se realice la ordenación de comparación.
Interflow Match	Una etapa de cruce que ubica las coincidencias entre registros de datos similares entre dos flujos de registros de entrada. El primer flujo de registros es una fuente de origen de registros sospechosos, mientras que el segundo es una fuente de origen de registros candidatos.
Intraflow Match	Etapa de cruce que localiza las coincidencias entre registros de datos similares en un mismo flujo de entrada.
Elevación	Aumento de duplicados.

Grupos de cruce	(Agrupar por) Los registros están agrupados por una clave de cruce o una ventana dinámica.
Resultado de coincidencia	(o Paquete de recursos) Agrupación lógica de archivos producida por una etapa. Estos datos se guardan por cada ejecución de una etapa y se almacenan en el disco. Las ejecuciones subsiguientes no sobrescriben ni cambian los resultados de las ejecuciones anteriores. En la herramienta de análisis de cruce, los paquetes se utilizan para proveer información acerca de los resultados resumidos y detallados, así como configurar información.
Lista Resultados de cruce	Lista de resultados de cruce de un solo tipo que la herramienta de análisis de cruce puede analizar en la sesión actual.
Tipo de resultados de cruce	Hace referencia a los contenidos de los resultados de cruce. La herramienta de análisis de cruce utiliza los tipos de resultados de cruce para determinar cómo se utilizarán los datos.
Etapas de cruce	Etapas del lienzo que realiza rutinas de cruce. Las etapas de cruce son Interflow Match (Cruce de interflujo), Intraflow Match (Cruce de intraflujo) y Transaccional Match (Cruce transaccional).
Cruce perdido	Registro que anteriormente era sospechoso o duplicado pero que ahora es único.
Nuevo cruce	Registro que anteriormente era único, pero en la actualidad es sospechoso o duplicado.
Ventana dinámica	El método de comparación de ventana dinámica llena de forma secuencial el tamaño de un área de influencia predeterminada que se denomina ventana con la correspondiente cantidad de filas de datos. Puesto que cada fila se agrega a la ventana, se la compara con cada elemento que se encuentra en la ventana.
Registro sospechoso	Registro de controlador que se cruza con candidatos dentro de un grupo de cruce o un grupo de candidatos.
Transaccional Match	Etapas de cruce que cruza los registros sospechosos con los registros candidatos que devuelve Candidate Finder o una aplicación externa.
Registros únicos	Registro sospechoso o candidato que no coincide con ningún otro registro en un grupo de cruce. Si es el único registro de un grupo de coincidencia, el sospechoso automáticamente es único.

Campos estándares

Esta tabla contiene una lista de algunos de los campos estándar que encontrará en las tareas de comparación y muestra una descripción de estos.

Campos	Descripciones
CandidateGroup	Este campo identifica una agrupación de un registro sospechoso y sus candidatos. Cada registro sospechoso recibe un número de CandidateGroup. Los candidatos para ese registro sospechoso reciben el mismo número de CandidateGroup. Por ejemplo, si John Smith es un registro sospechoso y sus registros candidatos son John Smith y Jon Smth, entonces los tres registros tendrán el mismo valor de CandidateGroup.
CollectionNumber	Identifica una colección de registros duplicados. Los valores posibles son 1 o valores mayores a 1. Los registros con un marcador de mismo valor son registros duplicados.
CollectionRecordType	Identifica la plantilla y los registros Best of Breed en una colección de registros duplicados. Los valores posibles son: <ul style="list-style-type: none"> • Principal: el registro es el registro de plantilla seleccionado en una colección. • Secundario: el registro no es el registro de plantilla seleccionado en una colección. • BestOfBreed: el registro es el registro Best of Breed que se acaba de crear en la colección.
ExpressMatchIdentified	Indica si la coincidencia se obtuvo utilizando la clave de cruce inmediato. Los valores posibles son Sí o No.
HasDuplicates	Identifica si el registro es un duplicado de otro registro. Las opciones son: <ul style="list-style-type: none"> • Y: el registro es un registro sospechoso y tiene duplicados. • N: el registro es un registro sospechoso y no tiene duplicados. • D: el registro es un registro candidato y es un duplicado del registro sospechoso. • U: el registro es un candidato, pero no es un duplicado del registro sospechoso.
MatchRecordType	Identifica el tipo de registro de cruce de una colección. Los valores posibles son: <ul style="list-style-type: none"> • Sospechoso: registro con el que se comparan otros registros para determinar si son duplicados entre sí. Cada colección tiene un solo registro sospechoso. • Duplicado: registro que es un duplicado del registro sospechoso. • Único: registro que no tiene duplicados.
MatchScore	Identifica la calificación general entre dos registros. Los valores posibles son 0-100; 0 indica una coincidencia poco satisfactoria y 100, una coincidencia exacta. Las calificaciones se generan según la regla de cruce configurada.
TotalMatchCount	Este campo indica el número total de cruces realizados durante el procesamiento.

Campos	Descripciones
TransactionRecordType	<p>Puede ser una de estas opciones:</p> <ul style="list-style-type: none"> • Sospechoso: se utiliza un registro sospechoso como entrada a una consulta. • Candidato: un registro candidato es un resultado devuelto de una consulta.

Técnicas para la definición de claves de cruce

La comparación eficaz y eficiente requiere el equilibrio correcto entre precisión y rendimiento. El enfoque más preciso para la comparación sería analizar cada registro en relación con el resto de los registros, pero esto no es factible debido a que la cantidad de registros que se debería procesar produciría un rendimiento inaceptablemente lento. Un mejor enfoque es limitar la cantidad de registros involucrados en el proceso de comparación, a aquellos con mayor probabilidad de cruce. Puede hacer esto mediante el uso de las claves de cruce. Una clave de cruce es un valor que se crea por cada registro, mediante el uso de un algoritmo definido por el usuario. El algoritmo toma los valores del registro y los utiliza para generar un valor de clave de cruce, que se almacena como un nuevo campo en el registro.

Por ejemplo, si el registro entrante es:

Primer nombre: Fred

Apellido: Mertz

Código postal: 21114-1687

Código de género: M

Y define una regla de clave de cruce que genera una clave de cruce al combinar los datos del registro, de la siguiente manera:

Campo de entrada	Posición de inicio	Longitud
Código postal	1	5
Código postal	7	4
Apellido	1	5
Nombre	1	5
Código de género	1	1

Entonces la clave será:

211141687MertzFredM

Todo registro que contenga la misma clave de cruce se coloca en un grupo de cruce. El proceso de comparación posteriormente compara los registros en el grupo entre sí para identificar las coincidencias.

Para crear una clave de cruce, utilice la etapa Match Key Generator si compara registros con Interflow Match o Intraflow Match. Si compara registros con Transactional Match, use la etapa Candidate Finder para crear los grupos de cruce.

Nota: Las pautas que siguen se pueden aplicar a las consultas de Candidate Finder y claves de Match Key Generator. En Candidate Finder, estas pautas se aplican a la manera en que se define la instrucción SELECT.

Tamaño y rendimiento del grupo de cruce

La clave de cruce determina el tamaño de un grupo de cruce y, por lo tanto, el rendimiento de su flujo de datos. A medida que se duplica el tamaño del grupo de cruce, también se duplica el tiempo de ejecución. Por ejemplo, si define una clave de búsqueda que genera un grupo de 20 registros potencialmente coincidentes, tardará el doble de tiempo procesar como si modificara la clave de cruce, de modo que el grupo de cruce solo contenga 10 registros potencialmente coincidentes. La desventaja de "restringir" la regla de la clave de búsqueda para generar un grupo de cruce más pequeño, es que se corre el riesgo de excluir registros que sí coinciden. "Aflojar" la regla de la clave de cruce reduce la posibilidad de excluir del grupo a un registro coincidente, pero se incrementa el tamaño del grupo. A fin de encontrar el equilibrio correcto para sus datos, es importante que pruebe una variedad de reglas de clave de cruce, mediante el uso de los datos que sean representativos de los datos que intenta procesar en la producción.

Densidad

Al diseñar una clave de cruce, es importante considerar la densidad del área. La densidad se refiere al grado en el que los datos se pueden distribuir en todos los grupos de cruce. Dado que el rendimiento se determina mediante el número de comparaciones que el sistema debe realizar, las claves de cruce que generan una pequeña cantidad de grupos de cruce de gran tamaño producen un rendimiento más lento, en comparación con las claves de cruce que generan una mayor cantidad de grupos de cruce de tamaño pequeño.

Para ilustrar este concepto, considere una situación donde posee un conjunto de un millón de registros de nombres y direcciones que desea comparar. Debe definir una clave de cruce como los tres primeros bytes del código postal y la primera letra del apellido. Si los registros son de EE.UU., la clave de cruce producirá una gran cantidad de grupos de cruce y es probable que tenga un rendimiento aceptable. Sin embargo, si todos los registros son de Nueva York, todos los códigos postales comenzarán con "100" y terminará al menos con solo 26 grupos de cruce. Esto generará grandes grupos de cruce con aproximadamente 38.000 registros en promedio.

Puede calcular la cantidad máxima de comparaciones realizadas por cada grupo de cruce, al utilizar la siguiente fórmula:

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

Donde N es la cantidad de registros en el grupo de cruce.

Entonces, si tiene 26 grupos de cruce con 38.000 registros cada uno, la cantidad máxima de comparaciones realizadas sería de 18,7 mil millones aproximadamente. A continuación se muestra cómo se calcula la cantidad:

Primero, determine la cantidad máxima de comparaciones por grupo de cruce:

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

Luego, multiplique esta cantidad por el número de grupos de cruce:

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

Si en su lugar hubieran 100 valores únicos para los primeros 3 bytes del código postal, tendría 2.600 grupos de cruce con un promedio de 380 registros. En este caso, la cantidad máxima de comparaciones sería de 187 millones, lo que es 100 veces menos. Entonces, si los registros son de Nueva York, debe considerar el uso de los primeros cuatro o cinco bytes del código postal, para que la clave de cruce genere más grupos de cruce y reduzca la cantidad de comparaciones. Puede que pierda unas pocas coincidencias, pero el equilibrio reducirá en gran medida el tiempo de ejecución.

En la realidad, una clave de búsqueda como la que se usa en este ejemplo, no generará grupos de cruce del mismo tamaño debido a las variaciones de los datos. Por ejemplo, habrá una mayor cantidad de personas cuyo apellido comience con "S" que con "X". Debido a esto, debe enfocar sus esfuerzos en reducir el tamaño de los grupos de cruce más grandes. Un grupo de cruce de 100.000 registros es 10 veces más grande que un grupo de 10.000, pero requerirá 100 veces más comparaciones y tardará 100 veces más. Por ejemplo, considere que usa cinco bytes del código postal y seis bytes del campo AddressLine1 para la clave de cruce. A primera vista parece una clave de cruce bastante buena. El problema es con las direcciones de casilla postal. Aunque la mayoría de los grupos de cruce pueden tener un tamaño aceptable, habrá unos pocos grupos de cruce muy grandes con claves similares a 10002PO BOX, con una cantidad muy grande de registros. Para dividir los grupos de cruce grandes, puede modificar la clave de cruce para incluir el primer par de dígitos del número de la casilla postal.

Alineación de la clave de cruce con la regla de cruce

Para obtener los resultados más precisos, debe diseñar una clave de cruce que funcione bien con la regla de cruce que usará en ella. Esto requerirá que considere la manera en que se define la regla de cruce.

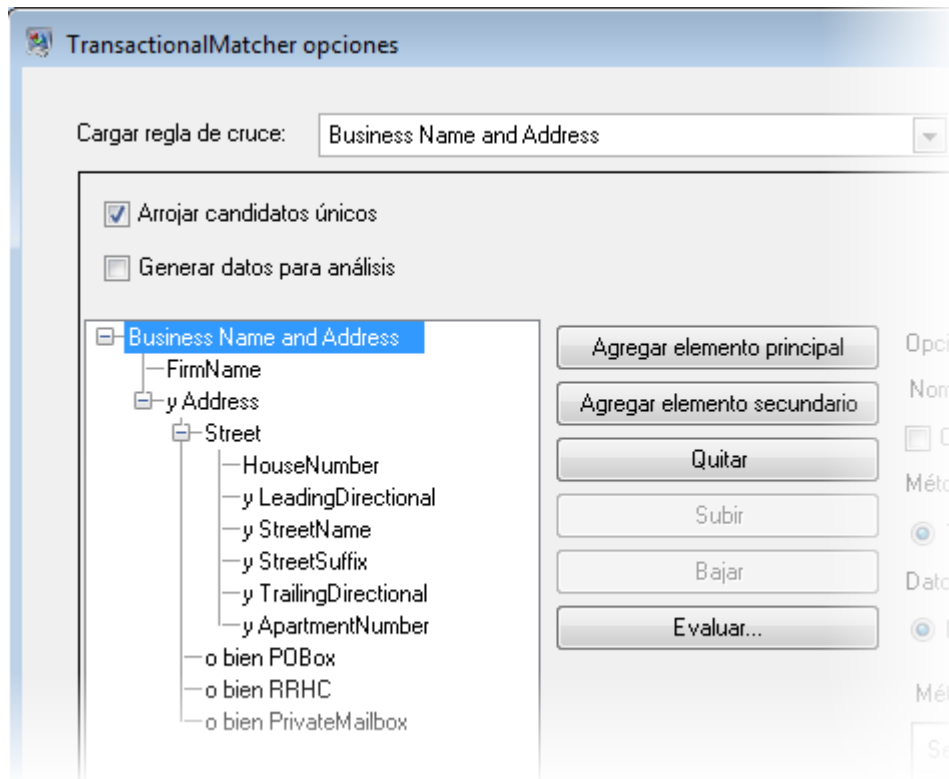
- La clave de cruce debe incluir cualquier campo que la regla de cruce exija que sea una coincidencia exacta.
- La clave de cruce debe utilizar el mismo tipo de algoritmo que se usa en la regla de cruce. Por ejemplo, si diseña una clave de cruce para su uso con una regla de cruce que utiliza un algoritmo fonético, entonces la clave de cruce también debe usar un algoritmo fonético.
- La clave de cruce se debe desarrollar mediante el uso de los datos de todos los campos que se utilizan en la regla de cruce.

- Considere cómo se verá afectada la clave de cruce si faltan datos de uno o más de los campos que se utilizan para la clave de cruce. Por ejemplo, considere que utiliza la inicial del segundo nombre como parte de la clave de búsqueda y tiene el registro de John A. Smith y otro de John Smith. Configuró la regla de cruce para que ignore los valores en blanco en el campo de inicial del segundo nombre, de modo que estos dos registros coincidirán según su regla de cruce. Sin embargo, debido a que la regla de cruce utiliza la inicial del segundo nombre, los dos registros terminarían en grupos de cruce distintos y no se compararían entre sí, con lo que se anularía su regla de cruce.

Reglas de cruce

Cada una de las etapas de búsqueda de cruce (Interflow Match, Intraflow Match y Transactional Match) requiere que se configure una regla de cruce. Una regla de cruce define los criterios que se utilizan para determinar si un registro coincide con otro. Esta regla especifica los campos que van a compararse, cómo van a compararse y establece una jerarquía de comparaciones para las reglas de cruce complejas.

Crear un conjunto jerárquico de comparaciones le permite formar reglas de cruce booleanas anidadas. Por ejemplo, considere la siguiente regla de cruce:



En este ejemplo, la regla de cruce está intentando hacer coincidir los registros en base a un nombre y dirección de empresa. El primer elemento de la regla de cruce es el campo FirmName. Este elemento implica que el valor del campo FirmName debe coincidir para que coincidan los registros. El segundo elemento evalúa la dirección. Tenga en cuenta que es precedido por el operador lógico "y", lo que significa que tanto FirmName como Dirección deben coincidir para que coincidan los registros. La parte de dirección de la regla de cruce consta de reglas secundarias que evalúan cuatro tipos de direcciones: direcciones de calle, direcciones de casilla postal, direcciones de ruta rural o contrato de autopista (RRHC) y direcciones de casilla de correo privado. La regla secundaria Calle analiza los campos de flujo de datos HouseNumber, LeadingDirectional, StreetName, StreetSuffix, TrailingDirectional y ApartmentNumber. Si todos coinciden, entonces la regla principal "Calle" y su regla principal "Dirección" se evalúan como "verdaderas". Si la regla Calle no se evalúa como verdadera, se evalúa el campo POBox, el RRHC y, a continuación el PrivateMailbox. Si alguno de estos tres coincide, entonces coincidirá el elemento principal Dirección.

Generación de reglas de cruce

Las reglas de cruce se usan en Interflow Match, Intraflow Match y Transactional Match para definir los criterios que se utilizan para determinar si un registro coincide con otro. Las reglas de cruce especifican los campos que se van a comparar y cómo van a compararse, así como una jerarquía de comparaciones para las reglas de comparación complejas.

Puede generar reglas de cruce en Interflow Match, Intraflow Match y Transactional Match. También puede generar reglas de cruce en la herramienta Match Rule Management de Enterprise Designer. Si genera una regla en la herramienta Match Rule Management, esa regla podrá usarse en cualquier flujo de datos y también estará disponible para otros usuarios. Si genera una regla en alguna de las etapas de cruce, esa regla estará disponible solo para esa etapa, a menos que la guarde haciendo clic en el botón **Guardar**, con lo cual quedará disponible para otras etapas y usuarios.

1. Abra Enterprise Designer.
2. Realice una de estas acciones:
 - Si desea definir una regla de cruce en Interflow Match, Intraflow Match o Transactional Match, haga doble clic en la etapa de cruce para la que desea definir la regla. En el campo **Cargar regla de cruce**, seleccione una regla de cruce predefinida como punto de partida. Si desea empezar con una regla de cruce en blanco, haga clic en **Nuevo**.
 - Si desea definir una regla de cruce en la herramienta Administración de reglas de coincidencia, seleccione **Herramientas > Administración de reglas de coincidencia**. Si desea usar una regla ya existente como punto de partida para su regla, marque la casilla **Copiar desde** y seleccione la regla que va a usar como punto de partida.
3. Especifique los campos de flujo de datos que desea usar en la regla de cruce, así como la jerarquía de regla de cruce.
 - a) Haga clic en **Agregar elemento principal**.
 - b) Ingrese un nombre para el elemento principal. El nombre debe ser único y no puede ser un campo. El primer elemento principal de la jerarquía se usa como nombre de la regla de

cruce en el campo **Cargar regla de cruce**. Todas las reglas de coincidencia personalizadas que cree y las reglas predefinidas que modifique se guardan con la palabra "Custom" anexada al principio del nombre.

- c) Haga clic en **Agregar elemento secundario**. Aparecerá un menú desplegable en la jerarquía de regla. Seleccione un campo para agregarle al elemento principal.

Nota: Todos los nodos secundarios de un nodo principal deben usar el mismo operador lógico. Si desea usar diferentes operadores lógicos para cada campo, primero deberá crear elementos principales intermedios.

- d) Repita el proceso hasta completar la jerarquía de comparación.

4. Defina las opciones de los elementos principales. Las opciones para los elementos principales se muestran a la derecha de la jerarquía de regla cuando hay un nodo principal seleccionado.

- a) Haga clic en **Coincidir cuando no sea verdadero** para cambiar el operador lógico del elemento principal de Y a Y NO. Si selecciona esta opción, los registros solo coincidirán si no coinciden con la lógica definida en este elemento principal.

Nota: Comprobar la opción **Coincidir cuando no sea verdadero** tiene el efecto de negar las opciones del **Método de comparación**. Para obtener más información, consulte [Condiciones de cruce negativas](#) en la página 82.

- b) En el campo **Método de comparación** especifique cómo determinar si un elemento principal presenta un cruce o no. Una de las siguientes:

Todo verdadero	Se considera que un elemento principal es un cruce si todos los elementos secundarios están determinados para el cruce. Este método crea un conector "AND" entre los elementos secundarios.
Cualquier elemento verdadero	Se considera que un elemento principal es un cruce si al menos uno de los elementos secundarios está determinado para el cruce. Este método crea un conector "OR" entre los elementos secundarios.
Según el umbral	Se considera que un elemento principal es un cruce si la calificación del elemento principal es mayor o igual que el umbral del elemento principal. Cuando selecciona esta opción, aparece la barra deslizante Umbral . Utilice esta barra deslizante para especificar un umbral. El método de calificación determina qué conector lógico utilizar. Los umbrales del elemento principal no pueden ser mayores que el umbral de los elementos secundarios.

Nota: El umbral que se configura aquí puede invalidarse durante la ejecución en el cuadro de diálogo Opciones de flujo de datos. Vaya a **Editar > Opciones de flujo de datos** y haga clic en **Agregar**. Expanda la etapa, haga clic en **Umbral de nivel superior** e ingrese el umbral en el campo **Valor predeterminado**.

- c) En el campo **Datos faltantes** especifique cómo calificar los campos vacíos. Una de las siguientes:

Ignorar espacios en blanco	Ignora el campo si está vacío.
Contar como 0	Asigna al campo una calificación de 0 si está vacío.
Contar como 100	Asigna al campo una calificación de 100 si está vacío.
Comparar espacios en blanco	Otorga campos sospechosos y candidatos como 100 si ambos contienen datos en blanco; de otro modo, otorga campos sospechosos y candidatos como 0.

- d) En el campo **Método de calificación** seleccione el método utilizado para determinar la calificación de comparación. Una de las siguientes:

Promedio ponderado	Utiliza el peso de cada elemento secundario para determinar la calificación de cruce promedio.
Promedio	Utiliza la calificación promedio de cada elemento secundario para determinar la calificación del elemento principal.
Máximo	Utiliza la calificación más alta de los elementos secundarios para determinar la calificación del elemento principal.
Mínimo	Utiliza la calificación más baja de los elementos secundarios para determinar la calificación del elemento principal.
Suma total del vector	Uses the vector summation of each child score to determine the score of the parent. The formula for calculation is: $\sqrt{a^2 + b^2 + c^2} / \sqrt{n}$, where: a, b, and c are the scores of three children and n is the number of children.

La siguiente tabla muestra la relación lógica existente entre los métodos de comparación y los métodos de calificación, y cómo cada combinación cambia la lógica utilizada durante el proceso de cruce.

Tabla 1: Matriz del método de comparación y el método de calificación

Método de calificación	Método de comparación			Comentarios
	Cualquier elemento verdadero	Todo verdadero	Según el umbral	
Promedio ponderado	N/A	AND	AND	Disponible solamente cuando se seleccionan como método de comparación Todo verdadero o Según el umbral.
Promedio	N/A	AND	AND	
Suma total del vector	N/A	AND	AND	
Máximo	OR	N/A	OR	Disponible solamente cuando se seleccionan como método de comparación Cualquier elemento verdadero o Según el umbral.
Mínimo	OR	N/A	OR	

5. Defina las opciones de los elementos secundarios. Las opciones para los elementos secundarios se muestran a la derecha de la jerarquía de regla cuando hay un elemento secundario seleccionado.

- Marque la opción **Campo del candidato** para asignar el campo del registro secundario seleccionado al campo en el archivo de entrada.
- Verifique la opción **Comparar de forma cruzada con** y seleccione uno o más elementos de la lista desplegable para comparar diferentes entre dos registros. Si están utilizando la herramienta Administración de reglas de coincidencia para crear o editar una regla de coincidencia, no habrá lista desplegable sino que deberá ingresar cada nombre de campo, separado por comas.
- Haga clic en **Coincidir cuando no sea verdadero** para cambiar el operador lógico de Y a NO. Si selecciona esta opción, la regla de cruce solo evaluará como verdadero si los registros no coinciden con la lógica definida en este elemento secundario.

Por ejemplo, si desea identificar individuos que están asociados a múltiples cuentas, podría crear una regla de cruce que coincida con el nombre pero que no coincida con el número de cuenta. Utilizaría la opción **Coincidir cuando no sea verdadero** para el elemento secundario que no coincide con el número de cuenta.

- En el campo **Datos faltantes** especifique cómo calificar los campos vacíos. Una de las siguientes:

Ignorar espacios en blanco Ignora el campo si está vacío.

- Contar como 0** Asigna al campo una calificación de 0 si está vacío.
- Contar como 100** Asigna al campo una calificación de 100 si está vacío.
- Comparar espacios en blanco** Otorga campos sospechosos y candidatos como 100 si ambos contienen datos en blanco; de otro modo, otorga campos sospechosos y candidatos como 0.
- e) En el **campo Umbral**, especifique el umbral cuyo criterio debe cumplirse a nivel del campo individual para que se determine que ese campo es un cruce.
- f) En el campo **Método de calificación** seleccione el método utilizado para determinar la calificación de comparación. Una de las siguientes:
- Promedio ponderado** Utiliza el peso de cada algoritmo para determinar la calificación de cruce promedio.
- Promedio** Utiliza la calificación promedio de cada algoritmo para determinar la calificación de cruce.
- Máximo** Utiliza la calificación más alta de los algoritmos para determinar la calificación de cruce.
- Mínimo** Utiliza la calificación más baja de los algoritmos para determinar la calificación de cruce.
- Suma total del vector** Uses vector summation of the score of each algorithm to determine the match score. This scoring method is useful if you want a higher match score in one or more algorithms to get proportionately represented in the final match score. The formula used for calculating the final score is:
- $$\sqrt{a^2 + b^2 + c^2} / \sqrt{n}, \text{ where: } a, b, \text{ and } c \text{ are the scores of three different algorithms and } n \text{ is the number of algorithms used.}$$
- g) Elija uno o más algoritmos que se utilizarán para determinar si los valores del campo constituyen un cruce. Las opciones son:
- Sigla** Determina si un nombre comercial coincide con su acrónimo buscando datos de acrónimos; de lo contrario, crea un acrónimo con la primera letra de cada palabra. Ejemplo: Internal Revenue Service y su sigla IRS serían considerados un cruce y arrojarían una calificación de cruce de 100.
- Frecuencia de caracteres** Determina la frecuencia de aparición de cada carácter en una cadena y compara las frecuencias generales entre dos cadenas.
- Daitch-Mokotoff Soundex** Algoritmo fonético que identifica con mayor precisión los apellidos eslavos e yiddish que tienen una pronunciación similar pero se escriben diferente. Los nombres codificados tienen seis dígitos y es posible obtener diversas codificaciones para un único nombre. Esta opción se desarrolló para

responder a las limitaciones de Soundex en el procesamiento de apellidos germanos o eslavos.

Fecha

Compara los campos de fecha independientemente del formato de fecha de los registros de entrada. Haga clic en Editar en la columna Opciones para especificar lo siguiente:

- **Requerir mes:** evita que las fechas que consisten solamente en un año presenten un cruce.
- **Requerir día:** evita que las fechas que consisten solamente en un mes y un año presenten un cruce
- **Comparar MM/DD transpuestos:** en los casos en que el mes y el día están en formato numérico, compara el mes sospechoso con el día candidato y el día sospechoso con el mes candidato, además de realizar la comparación estándar entre mes sospechoso y mes candidato y día sospechoso y día candidato
- **Preferir el formato DD/MM/AAAA al formato MM/DD/AAAA:** contribuye al análisis de la fecha en aquellos casos en que tanto el mes como el día están en formato numérico y su identificación no se puede determinar por el contexto. Por ejemplo, dados los números 5 y 13, el analizador asignará automáticamente 5 al mes y 13 al día porque solo hay 12 meses en el año. Sin embargo, dados los números 5 y 12 (o dos números cualquiera que sean iguales o menores que 12), el analizador presupondrá que el número que aparece primero es el mes. Seleccionar esta opción asegura que el analizador considere que el primer número es el día en vez del mes.
- **Opciones de rango: General:** permite configurar la cantidad máxima de días entre las fechas de comparación. Por ejemplo, si ingresa un rango general de 35 días y su fecha candidata es el 31 de diciembre de 2000, una fecha sospechosa del 5 de febrero de 2001 sería un cruce, pero no así una del 6 de febrero. Si ingresa un rango general de 1 día y su fecha candidata es enero de 2000, una fecha sospechosa de 1999 sería un cruce (al comparar el 31 de diciembre de 1999) pero no así una fecha sospechosa de enero de 2001.
- **Opciones de rango: Año:** permite configurar la cantidad de años entre las fechas de comparación, independientemente del mes y el día. Por ejemplo, si ingresa 3 como rango de año y su fecha candidata es el 31 de enero de 2000, una fecha sospechosa del 31 de enero de 2003 sería un cruce, pero no así una fecha sospechosa de febrero de 2003. De la misma manera, si su fecha candidata es el año 2000, una fecha sospechosa de marzo de 2003 sería un cruce porque no se especifican los meses y está dentro del rango de tres años.
- **Opciones de rango: Mes:** permite configurar la cantidad de meses entre las fechas de comparación, independientemente del año y el día. Por ejemplo, si ingresa 4 como rango de mes y su fecha candidata es el 1 de enero de 2000, una fecha sospechosa de mayo de 2000 es

un cruce porque no se especifica un día y está dentro del rango de cuatro meses, pero una fecha sospechosa del 2 de mayo de 2000 no es un cruce porque los días no coinciden.

- **Opciones de rango: día:** permite configurar la cantidad de días entre las fechas de comparación, independientemente del año y el mes. Por ejemplo, si ingresa 5 como rango de día y su fecha candidata es el 1 de enero de 2000, una fecha sospechosa de enero de 2000 es un cruce porque no se especifica un día, pero una fecha sospechosa del 27 de diciembre de 1999 no es un cruce, porque los meses no coinciden.

Doble metaphone	Determina la similitud entre dos cadenas según la representación fonética de sus caracteres. Double Metaphone es una versión mejorada del algoritmo Metaphone e intenta explicar la gran cantidad de irregularidades encontradas en los distintos idiomas.
Distancia de edición	Determina la similitud entre dos cadenas según el número de eliminaciones, inserciones o sustituciones que se requieren para transformar una cadena en otra.
Distancia euclidiana	Ofrece una medida de semejanza entre dos cadenas mediante el uso del espacio de vector de términos combinados como si fuera las dimensiones. También determina el máximo factor común de dos números enteros. Toma un par de números enteros positivos y forma un nuevo par compuesto por el número más pequeño y la diferencia entre los números más grande y más pequeño. El proceso se repite hasta que los números se igualan. El número resultante constituye el máximo factor común del par original. Por ejemplo, 21 es el máximo factor común de 252 y 105: ($252 = 12 \times 21$; $105 = 5 \times 21$); ya que $252 - 105 = (12 - 5) \times 21 = 147$, el máximo factor común de 147 y 105 también es 21.
Cruce exacto	Determina si dos cadenas son iguales.
Iniciales	Se utiliza para cruzar las iniciales en nombres personales analizados.
Distancia de Jaro-Winkler	Determina la similitud entre dos cadenas según el número de caracteres que hay que reemplazar para transformar una cadena en otra. Esta opción se desarrolló para cadenas cortas, por ejemplo, nombres personales.
Distancia de teclado	Determina la similitud entre dos cadenas según el número de eliminaciones, inserciones o sustituciones que se requieren para transformar una cadena en otra, ponderada según la posición de las teclas en el teclado. Haga clic en Editar en la columna Opciones para especificar el tipo de teclado que utiliza: QWERTY (Estados Unidos), QWERTZ (Austria y Alemania) o AZERTY (Francia).
Koeln	Indexa los nombres por sonido, tal como se los pronuncia en alemán. Permite que los nombres que tienen la misma pronunciación se

	<p>codifiquen con la misma representación, de modo tal que puedan cruzarse a pesar de las pequeñas diferencias en el modo en que se escriben. El resultado es siempre una secuencia de números; los caracteres especiales y los espacios en blanco se ignoran. Esta opción se desarrolló para responder a las limitaciones de Soundex.</p>
Distancia de Kullback-Liebler	<p>Determina la similitud entre dos cadenas según las diferencias que presentan en su distribución de palabras.</p>
Metaphone	<p>Determina la similitud entre dos cadenas en inglés según la representación fonética de sus caracteres. Esta opción se desarrolló para responder a las limitaciones de Soundex.</p>
Metaphone (español)	<p>Determina la similitud entre dos cadenas según la representación fonética de sus caracteres. Esta opción se desarrolló para responder a las limitaciones de Soundex.</p>
Metaphone 3	<p>Mejora los algoritmos de Metaphone y Double Metaphone con parámetros más exactos de consonantes y vocales internas que permiten pronunciar palabras y nombres que, fonéticamente, coinciden en mayor o menor medida con los términos de búsqueda. Metaphone 3 aumenta la precisión de la codificación fonética en un 98%. Esta opción se desarrolló para responder a las limitaciones de Soundex.</p>
Variante de nombre	<p>Determina si dos nombres son variantes entre sí. El algoritmo arroja una calificación de cruce de 100 si dos nombres son variaciones entre sí, y una calificación de cruce de 0 si los dos nombres no son variaciones entre sí. Por ejemplo, JOHN es una variación de JAKE y arroja una calificación de cruce de 100. JOHN no es una variación de HENRY y arroja una calificación de cruce de 0. Haga clic en Editar en la columna Opciones para seleccionar las opciones de Variante de nombre. Para obtener más información, consulte Name Variant Finder en la página 326.</p>
Distancia NGram	<p>Calcula en texto o habla la probabilidad del siguiente término basado en los términos n anteriores, que pueden incluir fonemas, sílabas, letras, palabras o pares de base y que pueden consistir en cualquier combinación de letras. Este algoritmo incluye una opción para ingresar el tamaño de NGram; el valor predeterminado es 2.</p>
Similitud de NGram	<p>Determines similarity between two strings based on the length of the longest common subsequence of phonemes, syllables, letters, words or base pairs.</p> <p>The algorithm includes the following options:</p> <ul style="list-style-type: none"> • NGram size: Enter the size of the NGram. El valor predeterminado es 2. • Drop Noise Characters: Select the check-box to replace punctuation with space.

- **Drop Spaces:** Select the check-box to merge words.

Cadena numérica

Compara las líneas de dirección separando los atributos numéricos de los caracteres en una línea de dirección. Por ejemplo, si la cadena de la dirección es 1234 Main Street Apt 567, los atributos numéricos de la cadena (1234567) se analizan y procesan de forma diferente al resto del valor de la cadena (Main Street Apt). Primero, el algoritmo cruza los datos numéricos de la cadena con el algoritmo numérico. Si el cruce de los datos numéricos es de 100, se cruzan los datos alfabéticos utilizando Distancia de edición y Frecuencia de caracteres. La calificación de cruce final se calcula de la siguiente manera:

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

Por ejemplo, la calificación de cruce de las siguientes dos direcciones es de 95,5 y se calcula de la siguiente manera:

123 Main St Apt 567
123 Maon St Apt 567

Calificación numérica = 100
Distancia de edición = 91
Frecuencia de caracteres = 91

$91 + 91 = 182$
 $182/2 = 91$
 $100 + 91 = 191$
 $191/2 = 95.5$

Nysiis

Algoritmo de codificación fonética que cruza una pronunciación aproximada con una palabra escrita exacta e indexa palabras que se pronuncian de manera similar. Parte del sistema de inteligencia e identificación del estado de Nueva York (New York State Identification and Intelligence System). Supongamos que está buscando información sobre alguien en una base de datos de personas. Usted cree que el nombre de esa persona suena algo así como "John Smith", pero en realidad se escribe "Jon Smath". Si busca una coincidencia exacta para "John Smith" no obtendrá resultados. Sin embargo, si indexa la base de datos con el algoritmo de NYSIIS y realiza la búsqueda usando ese mismo algoritmo, obtendrá una coincidencia correcta porque el algoritmo indexa "John Smith" y "Jon Smath" como "JANSNATH". Esta opción se desarrolló como respuesta a las limitaciones de Soundex; a diferencia de este último, Nyssis utiliza algunos n-gramas multicarácter y mantiene la posición relativa de las vocales.

Nota: Este algoritmo no procesa caracteres no alfabéticos; por lo tanto, los registros que contengan caracteres de este tipo se procesarán con errores.

Phonix	Preprocesa cadenas de nombres aplicando más de 100 reglas de transformación a caracteres únicos o secuencias de varios caracteres. Diecinueve de esas reglas se aplican solo si los caracteres se encuentran al comienzo de la cadena, mientras que 12 de ellas se aplican solo cuando los caracteres se encuentran en medio de la cadena, y 28 se aplican únicamente si están al final de la cadena. La cadena de nombre transformada se codifica en un código compuesto por la primera letra seguida de tres dígitos (sin ceros ni números duplicados). Esta opción se desarrolló como respuesta a las limitaciones de Soundex; es más compleja y, por lo tanto, más lenta que esta última.
Soundex	Determina la similitud entre dos cadenas según la representación fonética de sus caracteres.
Subcadena de caracteres	Determina si una cadena aparece dentro de otra.
Alineación silábica	Combina la información fonética con cálculos basados en la distancia de edición. Convierte las cadenas que se van a comparar en su correspondiente secuencia de sílabas y calcula la cantidad de ediciones que se necesitan para convertir una secuencia de sílabas a otra.

La siguiente tabla describe la relación lógica que existe entre el número de algoritmos que usted puede utilizar según el método de calificación del elemento principal que haya seleccionado.

Tabla 2: Matriz del algoritmo de comparación y el método de calificación

Método de calificación	Algoritmos	
	Simple	Múltiple
Promedio ponderado	N/A	Sí
Promedio	N/A	Sí
Máximo	Sí	Sí
Mínimo	N/A	Sí
Suma total del vector	N/A	Sí

- Si define una regla en Interflow Match, Intraflow Match o Transactional Match y desea compartirla con otras etapas o usuarios, haga clic en el botón **Guardar** ubicado en la parte superior de la pantalla.

Condiciones de cruce negativas

Las condiciones de cruce son enunciados que indican qué campos desea hacer coincidir para que dos registros se consideren un cruce. Sin embargo, es posible que en algunas situaciones desee definir una condición que indique que dos campos *no* deben coincidir a fin de que dos registros se consideren un cruce. Esta técnica, conocida como *negación*, invierte la lógica de una condición dentro de una regla de cruce.

Por ejemplo, supongamos que tiene registros de soporte de un cliente para un centro de atención telefónica y desea identificar qué clientes se han contactado con el centro de atención telefónica con relación a varias cuentas. En otras palabras, desea identificar a las personas que están asociadas con varias cuentas. Para identificar a los clientes que tienen varias cuentas, debe hacer coincidir registros en los cuales el nombre sea el mismo, pero el número de cuenta *no* coincida. En este caso, deberá usar la negación en una condición de cruce para el número de cuenta.

Para usar la negación, active la casilla **Coincidir cuando no sea verdadero** cuando defina la regla de cruce. Esta opción está disponible tanto para elementos principales (grupos de condiciones) como para elementos secundarios (condiciones individuales) en la regla de cruce. El efecto de esta opción es levemente diferente cuando se usa con elementos principales y con elementos secundarios. Cuando se usa con un elemento principal, la opción **Coincidir cuando no sea verdadero** efectivamente invierte la opción del método de cruce de la siguiente manera:

- El método de coincidencia **Todo verdadero** se convierte efectivamente en "algún falso". La regla de cruce solo puede coincidir con registros si, al menos uno de los elementos secundarios debajo del elemento principal resulta falso y, por tanto, el elemento principal resultará falso. Como la opción **Coincidir cuando no sea verdadero** está activada, esta determinación de falsedad originará una coincidencia.
- El método de coincidencia **Algún verdadero** se convierte efectivamente en "ningún verdadero". La regla de cruce solo puede coincidir con registros en los cuales ninguno de los elementos secundarios sea considerado verdadero, ya que si alguno de los elementos secundarios se considera verdadero, el elemento principal será verdadero y, con la opción **Coincidir cuando no sea verdadero** activada, esta determinación de verdadero no generará ninguna coincidencia. Solo si ninguno de los elementos secundarios son verdaderos, por lo que el elemento principal será "no verdadero", la regla encontrará una coincidencia.
- El método de coincidencia **Según el umbral** efectivamente cambia de los registros coincidentes que son iguales o mayores que el umbral especificado por los registros coincidentes que son menores que el umbral. Esto se debe a que los registros con un valor de umbral menor que el especificado se considerarán falsos y, como la opción **Coincidir cuando no sea verdadero** está activada, se generará una coincidencia.

La opción **Coincidir cuando no sea verdadero** es fácil de comprender cuando se la aplica a elementos secundarios en una regla de cruce. Esta simplemente indica que dos registros se consideran coincidentes si el algoritmo *no* indica una coincidencia.

Prueba de reglas de cruce

Después de definir una regla de cruce, puede que desee probarla para ver sus resultados. Para hacerlo, puede utilizar Evaluación de regla de cruce para examinar los efectos de una regla de cruce en un pequeño conjunto de datos de muestra.

1. Abra el flujo de datos en Enterprise Designer.
2. Haga doble clic en la etapa que contiene la regla de cruce que desea probar.

Las reglas de cruce se utilizan en Interflow Match, Intraflow Match y Transactional Match.

3. En la jerarquía de la regla de cruce, seleccione el nodo que desea probar y haga clic en **Evaluar**.
4. En la ficha **Importar**, introduzca los datos de la prueba (un sospechoso y hasta 10 candidatos). Hay dos maneras de introducir los datos de prueba.

- Para introducir los datos de prueba manualmente, ingrese un registro sospechoso en **Sospechoso** y hasta diez candidatos en **Candidatos**. Después de ingresar los registros, puede hacer clic en **Exportar** para guardar los registros en un archivo que se pueda importar posteriormente, en lugar de volver a ingresar los datos de forma manual.
- Para importar los datos de prueba de un archivo, haga clic en **Importar...** y seleccione el archivo que contiene los registros de muestras. Los archivos pueden estar delimitados por coma, barra vertical o tabulación y deben tener un registro de encabezado cuyos campos de encabezado coincidan con los nombres de campo exhibidos en **Candidatos**. Un registro de encabezado de muestra para la entrada de Hogar sería:

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

5. Evalúe la regla mediante uno de estos métodos:

- Haga clic en **Regla actual**. Así, se ejecuta la regla definida en la ficha **Regla de cruce**. Se muestran los resultados de un par compuesto por un sospechoso y un candidato por vez. Para recorrer los resultados, haga clic en los botones con flechas. Las calificaciones de los campos y los algoritmos se muestran en un formato de árbol similar al del control de regla de cruce. Opcionalmente, los resultados pueden exportarse a un archivo XML.

Nota: Si se hacen cambios a la regla de cruce y desea aplicarlos a la regla de cruce de la etapa, haga clic en **Guardar**.

- Haga clic en **Todos los algoritmos**. Así, se ignora la regla de cruce y se ejecutan todos los algoritmos para cada campo de pares de sospechosos y candidatos. Se muestran los resultados de un par compuesto por un sospechoso y un candidato por vez. Es posible recorrer esos resultados utilizando los botones con flechas.

Para actualizar automáticamente los resultados a medida que se realizan cambios en la regla de cruce o la entrada, seleccione la casilla de verificación **Actualizar automáticamente**. Al

utilizar esta función con la opción **Todos los algoritmos**, solo los cambios realizados en la entrada actualizarán los resultados.

Los resultados que se muestran en Calificaciones tienen el siguiente código de colores:

- Verde: la regla dio como resultado una coincidencia.
- Rojo: la regla no dio como resultado ninguna coincidencia.
- Gris: la regla fue ignorada.
- Azul: se trata de los resultados de los algoritmos individuales dentro de la regla.

Para exportar los resultados de la evaluación en formato XML, haga clic en **Exportar**.

Compartir reglas de cruce

You can create match rules that can be shared among modules, stages, dataflows, and users. Si comparte una regla de cruce, puede facilitar el desarrollo de flujos de datos mediante la definición de una regla de cruce única, a la cual podrá referirse cuando sea necesario. Esto también ayuda a garantizar que las reglas de cruce que tengan por objeto realizar la misma función sean uniformes en todo el flujo de datos.

The match rules saved in the Match Rule Repository can be exported as .mr or .json files using the Administration utility. This is especially useful when trying to share these rules with other Spectrum installations or while consuming these in the Spectrum Big Data Quality SDK.

Nota: By default the match rules are exported as .mr file.

- Para compartir una regla de cruce que haya creado en Interflow Match, Intraflow Match o Transactional Match, haga clic en el botón **Guardar** que se encuentra en la parte superior de la ventana de opciones de la etapa.
- Cuando crea una regla con la herramienta Match Rules Management, la regla queda disponible automáticamente para que todos los usuarios la utilicen con los flujos de datos. Para ver la herramienta Match Rules Management de Enterprise Designer, seleccione **Herramientas > Match Rules Management**.

Visualización de reglas de cruce compartidas

En Enterprise Designer, puede buscar todas las reglas de cruce compartidas disponibles en su sistema Spectrum™ Technology Platform. Estas reglas de cruce se pueden usar en las etapas Interflow Match, Intraflow Match, y Transactional Match en un flujo de datos para realizar la comparación.

Para buscar esas reglas de cruce en el repositorio de reglas de cruce, siga este procedimiento.

1. Abra Enterprise Designer.

2. Seleccione **Herramientas > Administración de reglas de coincidencia**.
3. Seleccione la regla que desea ver y haga clic en **Ver**.

Creación de una regla de cruce personalizada como un objeto JSON

Las reglas de cruce se pueden configurar y pasar en el tiempo de ejecución, si estas se exponen como opciones del flujo de datos. Esto le permite compartir las reglas de cruce entre los equipos y anular las reglas de cruce existentes con cadenas de reglas de cruce con formato JSON. También puede establecer las opciones de tiempo de ejecución cuando se invoque al trabajo, mediante un flujo de proceso o mediante la herramienta de línea de comandos del ejecutor del trabajo.

Puede encontrar esquemas para MatchRule y MatchInfo en la siguiente carpeta:

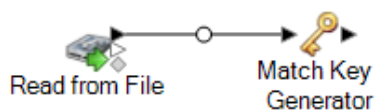
1. Guarde y exponga el flujo de datos que contiene la regla de cruce.
2. Abra el flujo de datos que usa la regla de cruce.
3. Vaya a **Edit > Dataflow Options**.
4. En la tabla **Aplicar opciones de flujo de datos a etapas** haga clic en la etapa de comparación que usa la regla de cruce y marque la casilla **Regla de cruce personalizada**.
5. Opcional: cambie el nombre de la regla de cruce en el campo **Rótulo de opción** de "Regla de cruce personalizada" al nombre de su preferencia.
6. Haga clic en **Aceptar** dos veces.

Cruce de registros de un origen único

Este procedimiento describe la forma de utilizar una etapa Intraflow Match para identificar grupos de registros dentro de un origen de datos único (como por ejemplo un archivo o tabla de base de datos) que se relacionan entre sí con base en los criterios de cruce especificados por usted. El flujo de datos agrupa los registros en colecciones y escribe las colecciones en un archivo de salida.

1. Genere un nuevo flujo de datos en Enterprise Designer.
2. Arrastre hacia el lienzo una etapa de origen.
3. Haga doble clic en la etapa de origen y configúrela. Consulte la *Guía de Dataflow Designer* para obtener instrucciones acerca de cómo configurar etapas de origen.
4. Arrastre hacia el lienzo una etapa Match Key Generator y conéctela a la etapa de origen.

Por ejemplo, si está utilizando una etapa de origen Read from File, su flujo de datos tendrá ahora el siguiente aspecto:



Match Key Generator crea una clave no exclusiva para cada registro, que posteriormente se puede usar en las etapas de comparación a fin de identificar grupos de registros potencialmente duplicados. Las claves de cruce facilitan el proceso de comparación al permitir la agrupación de registros por clave de cruce y posteriormente solo comparando los registros al interior de estos grupos.

5. Haga doble clic en Match Key Generator.
6. Haga clic en **Agregar**.
7. Defina la regla que se usará para generar una clave de cruce para cada registro.

Tabla 3: Opciones de Match Key Generator

Nombre de la opción	Descripción / Valores válidos
---------------------	-------------------------------

Algoritmo	
-----------	--

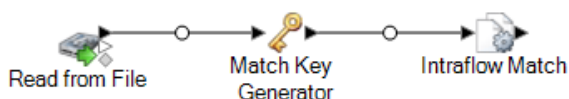
Nombre de la opción	Descripción / Valores válidos
	Especifica el algoritmo que se utilizará para generar la clave de cruce. Las opciones son:
Consonante	Devuelve campo especificados con consonantes eliminadas.
Doble metaphone	Devuelve un código basado en una representación fonética de sus caracteres. Double Metaphone es una versión mejorada del algoritmo Metaphone e intenta explicar la gran cantidad de irregularidades encontradas en los distintos idiomas.
Koeln	Indexa los nombres por sonido, tal como se los pronuncia en alemán. Permite que los nombres que tienen la misma pronunciación se codifiquen con la misma representación, de modo tal que puedan cruzarse a pesar de las pequeñas diferencias en el modo en que se escriben. El resultado es siempre una secuencia de números; los caracteres especiales y los espacios en blanco se ignoran. Esta opción se desarrolló para responder a las limitaciones de Soundex.
MD5	Un algoritmo de síntesis del mensaje que produce un valor hash de 128 bits. Este algoritmo suele usarse para comprobar la integridad de los datos.
Metaphone	Devuelve una clave codificada Metaphone de campos seleccionados. Metaphone es un algoritmo para codificar palabras utilizando su pronunciación del inglés.
Metaphone (español)	Devuelve una clave codificada Metaphone de campos seleccionados para el idioma español. El algoritmo Metaphone codifica palabras con base en su pronunciación en español.
Metaphone 3	Mejora los algoritmos de Metaphone y Double Metaphone con parámetros más exactos de consonantes y vocales internas que permiten pronunciar palabras y nombres que, fonéticamente, coinciden en mayor o menor medida con los términos de búsqueda. Metaphone 3 aumenta la precisión de la codificación fonética en un 98%. Esta opción se desarrolló para responder a las limitaciones de Soundex.
Nysiis	Algoritmo de codificación fonética que cruza una pronunciación aproximada con una palabra escrita exacta e indexa palabras que se pronuncian de manera similar. Parte del sistema de inteligencia e identificación del estado de Nueva York (New York State Identification and Intelligence System). Supongamos que está buscando información sobre alguien en una base de datos de personas. Cree que el nombre de esa persona suena algo así como "John Smith", pero en realidad se escribe "Jon Smyth". Si busca una coincidencia exacta para "John Smith" no obtendrá resultados. Sin embargo, si indexa la base de datos con el algoritmo de NYSIIS y realiza la búsqueda usando ese mismo algoritmo, obtendrá una coincidencia correcta porque el algoritmo indexa "John Smith" y "Jon Smyth" como "JAN SNATH".
Phonix	Preprocesa cadenas de nombres aplicando más de 100 reglas de transformación a caracteres únicos o secuencias de varios

Nombre de la opción	Descripción / Valores válidos
	<p>caracteres. Diecinueve de esas reglas se aplican solo si los caracteres se encuentran al comienzo de la cadena, mientras que 12 de ellas se aplican solo cuando los caracteres se encuentran en medio de la cadena, y 28 se aplican únicamente si están al final de la cadena. La cadena de nombre transformada se codifica en un código compuesto por la primera letra seguida de tres dígitos (sin ceros ni números duplicados). Esta opción se desarrolló como respuesta a las limitaciones de Soundex; es más compleja y, por lo tanto, más lenta que esta última.</p> <p>Soundex Devuelve un código Soundex de campos seleccionados. El algoritmo Soundex produce un código de longitud fija basado en la pronunciación en inglés de las palabras.</p> <p>Subcadena de caracteres Devuelve una porción específica del campo seleccionado.</p>
Nombre de campo	Especifica el campo en el que desea aplicar el algoritmo seleccionado para generar la clave de cruce. Por ejemplo, si selecciona un campo denominado LastName y elige el algoritmo Soundex, se aplicará ese algoritmo a los datos del campo LastName para generar una clave de cruce.
Posición de inicio	Especifica la posición de inicio dentro del campo especificado. No todos los algoritmos permiten especificar una posición inicial.
Longitud	Especifica la longitud, en caracteres, por incluirse desde la posición inicial. No todos los algoritmos permiten especificar una longitud.
Quitar caracteres irrelevantes	Elimina los caracteres que no son numéricos ni alfabéticos, como los guiones, los espacios en blanco y otros caracteres especiales de un campo de entrada.
Ordenar entrada	<p>Ordena alfabéticamente todos los caracteres de un campo de entrada o todos los términos de un campo de entrada.</p> <p>Caracteres Ordena los valores de los caracteres de un campo de entrada antes de crear una ID única:</p> <p>Términos Ordena el valor de cada término de un campo de entrada antes de crear una ID única.</p>

8. Cuando haya terminado de definir la regla, haga clic en **Aceptar**.

9. Si desea añadir reglas de cruce adicionales, haga clic en **Agregar** y añádalas, o bien haga clic en **Aceptar** cuando haya terminado.
10. Arrastre hacia el lienzo una etapa Interflow Match y conéctela a la etapa Match Key Generator.

Por ejemplo, si está utilizando una etapa de origen Read from File, su flujo de datos tendrá ahora el siguiente aspecto:



11. Haga doble clic en Intraflow Match.
12. En el campo **Cargar regla de cruce** seleccione una de las reglas de cruce predefinidas, que puede usar como está o bien puede modificarla para que se ajuste a sus necesidades. Si desea crear una nueva regla de cruce sin usar una de las reglas de cruce predefinidas como punto de partida, haga clic en **Nueva**. Solamente puede tener una regla personalizada en un flujo de datos.

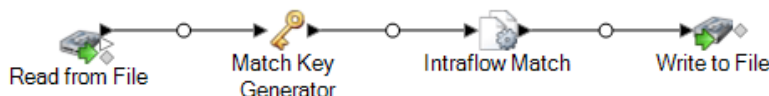
Nota: La función Opciones del flujo de datos en Enterprise Designer permite exponer la regla de cruce para la configuración durante la ejecución.

13. En el campo **Group by**, seleccione **MatchKey**.

Esto colocará en un grupo los registros que tengan la misma clave de cruce. La regla de cruce se aplica a los registros en un grupo para ver si hay duplicados. La clave de cruce de cada registro será generada por la etapa Generate Match Key que configuró antes en este procedimiento.

14. Para obtener más información acerca de cómo modificar las otras opciones, consulte [Generación de reglas de cruce](#) en la página 72.
15. Haga clic en **Aceptar** para guardar su configuración de Intraflow Match y vuelva al lienzo del flujo de datos.
16. Arrastre hacia el lienzo la etapa receptora y conéctela a la etapa Generate Match Key.

Por ejemplo, si estuviese utilizando la etapa receptora Write to File, su flujo de datos tendría el siguiente aspecto:



17. Haga doble clic en la etapa receptora y configúrela.

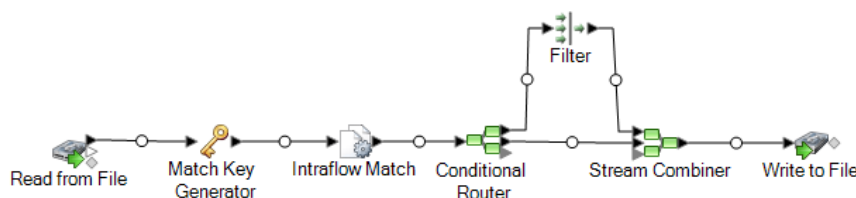
Para obtener información acerca de cómo configurar etapas receptoras, consulte la *Guía de Dataflow Designer*.

Ahora tiene un flujo de datos que cruzará registros de un origen único.

Ejemplo de Cruce de registros en un origen único de datos

En su calidad de administrador de datos de una empresa de tarjetas de crédito, usted desea analizar su base de datos de clientes y averiguar cuáles direcciones figuran varias veces y con qué nombres, a fin de poder reducir a un mínimo la cantidad de ofertas repetidas de tarjetas de crédito que se envían al mismo hogar.

Este ejemplo demuestra la forma de identificar a los miembros de un mismo hogar por medio de la comparación de información al interior de un único archivo de entrada y la generación de un archivo de salida que contenga solamente un registro por hogar.



La etapa Read from File lee datos que contienen tanto registros únicos por cada hogar como registros que posiblemente correspondan al mismo hogar. El archivo de entrada contiene nombres y direcciones.

Match Key Generator genera una clave de cruce, la cual se trata de una clave no exclusiva compartida por registros afines que identifica registros como posibles duplicados.

La etapa Intraflow Match compara registros que tienen la misma clave de cruce y marca cada registro ya sea como registro único o como uno de varios registros correspondientes al mismo hogar.

Conditional Router envía registros que son colecciones de registros de cada hogar a la etapa Filter, que filtra cada uno de los archivos correspondientes a cada hogar, salvo uno, el cual envía a a la etapa Stream Combiner. La etapa Conditional Router envía además registros únicos a Stream Combiner en forma directa.

Por último, la etapa Write to File genera un archivo de salida que contiene un registro por cada hogar.

Comparación de registros de un origen a otro origen

Este procedimiento describe cómo usar una etapa de Interflow Match para identificar registros en un origen que coincidan con registros en otro origen. El primer origen contiene registros sospechosos y el segundo contiene registros candidatos. El flujo de datos solo hace coincidir registros de un origen con registros en otro origen. No intenta hacer coincidir registros desde el interior del mismo

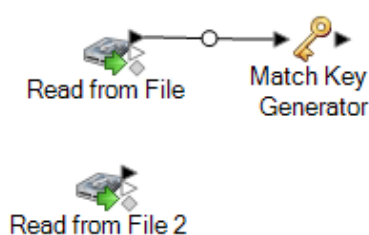
origen. El flujo de datos agrupa los registros en colecciones de registros coincidentes y escribe dichas colecciones en un archivo de salida.

1. Genere un nuevo flujo de datos en Enterprise Designer.
2. Arrastre hacia el lienzo dos etapas de origen. Configure una de ellas apuntando a los registros sospechosos y la otra a origen de registros candidatos.

Consulte la *Guía de Dataflow Designer* para obtener instrucciones acerca de cómo configurar etapas de origen.

3. Arrastre hacia el lienzo una etapa Match Key Generator y conéctela a una de las etapas de origen.

Por ejemplo, si está utilizando una etapa de origen Read from File, su flujo de datos tendrá ahora el siguiente aspecto:



Match Key Generator crea una clave no exclusiva para cada registro, que posteriormente se puede usar en las etapas de comparación a fin de identificar grupos de registros potencialmente duplicados. Las claves de cruce facilitan el proceso de comparación al permitir la agrupación de registros por clave de cruce y posteriormente solo comparando los registros al interior de estos grupos.

Nota: Más adelante añadirá una segunda etapa Match Key Generator. Por el momento necesitará en el lienzo solamente una.

4. Haga doble clic en la etapa Match Key Generator.
5. Haga clic en **Agregar**.
6. Defina la regla que se usará para generar una clave de cruce para cada registro.

Tabla 4: Opciones de Match Key Generator

Nombre de la opción	Descripción / Valores válidos
---------------------	-------------------------------

Algoritmo	
-----------	--

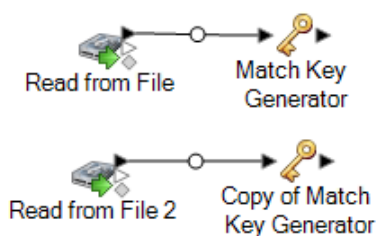
Nombre de la opción	Descripción / Valores válidos
	Especifica el algoritmo que se utilizará para generar la clave de cruce. Las opciones son:
Consonante	Devuelve campo especificados con consonantes eliminadas.
Doble metaphone	Devuelve un código basado en una representación fonética de sus caracteres. Double Metaphone es una versión mejorada del algoritmo Metaphone e intenta explicar la gran cantidad de irregularidades encontradas en los distintos idiomas.
Koeln	Indexa los nombres por sonido, tal como se los pronuncia en alemán. Permite que los nombres que tienen la misma pronunciación se codifiquen con la misma representación, de modo tal que puedan cruzarse a pesar de las pequeñas diferencias en el modo en que se escriben. El resultado es siempre una secuencia de números; los caracteres especiales y los espacios en blanco se ignoran. Esta opción se desarrolló para responder a las limitaciones de Soundex.
MD5	Un algoritmo de síntesis del mensaje que produce un valor hash de 128 bits. Este algoritmo suele usarse para comprobar la integridad de los datos.
Metaphone	Devuelve una clave codificada Metaphone de campos seleccionados. Metaphone es un algoritmo para codificar palabras utilizando su pronunciación del inglés.
Metaphone (español)	Devuelve una clave codificada Metaphone de campos seleccionados para el idioma español. El algoritmo Metaphone codifica palabras con base en su pronunciación en español.
Metaphone 3	Mejora los algoritmos de Metaphone y Double Metaphone con parámetros más exactos de consonantes y vocales internas que permiten pronunciar palabras y nombres que, fonéticamente, coinciden en mayor o menor medida con los términos de búsqueda. Metaphone 3 aumenta la precisión de la codificación fonética en un 98%. Esta opción se desarrolló para responder a las limitaciones de Soundex.
Nysiis	Algoritmo de codificación fonética que cruza una pronunciación aproximada con una palabra escrita exacta e indexa palabras que se pronuncian de manera similar. Parte del sistema de inteligencia e identificación del estado de Nueva York (New York State Identification and Intelligence System). Supongamos que está buscando información sobre alguien en una base de datos de personas. Cree que el nombre de esa persona suena algo así como "John Smith", pero en realidad se escribe "Jon Smyth". Si busca una coincidencia exacta para "John Smith" no obtendrá resultados. Sin embargo, si indexa la base de datos con el algoritmo de NYSIIS y realiza la búsqueda usando ese mismo algoritmo, obtendrá una coincidencia correcta porque el algoritmo indexa "John Smith" y "Jon Smyth" como "JAN SNATH".
Phonix	Preprocesa cadenas de nombres aplicando más de 100 reglas de transformación a caracteres únicos o secuencias de varios

Nombre de la opción	Descripción / Valores válidos
	<p>caracteres. Diecinueve de esas reglas se aplican solo si los caracteres se encuentran al comienzo de la cadena, mientras que 12 de ellas se aplican solo cuando los caracteres se encuentran en medio de la cadena, y 28 se aplican únicamente si están al final de la cadena. La cadena de nombre transformada se codifica en un código compuesto por la primera letra seguida de tres dígitos (sin ceros ni números duplicados). Esta opción se desarrolló como respuesta a las limitaciones de Soundex; es más compleja y, por lo tanto, más lenta que esta última.</p> <p>Soundex Devuelve un código Soundex de campos seleccionados. El algoritmo Soundex produce un código de longitud fija basado en la pronunciación en inglés de las palabras.</p> <p>Subcadena de caracteres Devuelve una porción específica del campo seleccionado.</p>
Nombre de campo	Especifica el campo en el que desea aplicar el algoritmo seleccionado para generar la clave de cruce. Por ejemplo, si selecciona un campo denominado LastName y elige el algoritmo Soundex, se aplicará ese algoritmo a los datos del campo LastName para generar una clave de cruce.
Posición de inicio	Especifica la posición de inicio dentro del campo especificado. No todos los algoritmos permiten especificar una posición inicial.
Longitud	Especifica la longitud, en caracteres, por incluirse desde la posición inicial. No todos los algoritmos permiten especificar una longitud.
Quitar caracteres irrelevantes	Elimina los caracteres que no son numéricos ni alfabéticos, como los guiones, los espacios en blanco y otros caracteres especiales de un campo de entrada.
Ordenar entrada	<p>Ordena alfabéticamente todos los caracteres de un campo de entrada o todos los términos de un campo de entrada.</p> <p>Caracteres Ordena los valores de los caracteres de un campo de entrada antes de crear una ID única:</p> <p>Términos Ordena el valor de cada término de un campo de entrada antes de crear una ID única.</p>

7. Cuando haya terminado de definir la regla, haga clic en **Aceptar**.
8. Haga clic derecho en la etapa Match Key Generator en el lienzo y seleccione **Copiar etapa**.

9. Haga clic derecho en un área vacía del lienzo y seleccione **Pegar**.
10. Conecte la copia de Match Key Generator a la otra etapa de origen.

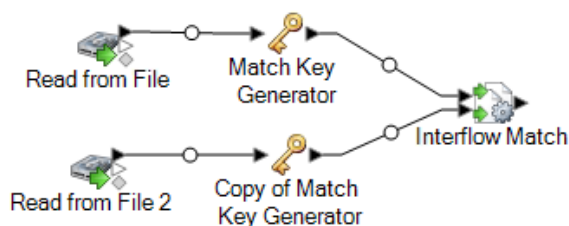
Por ejemplo, si está usando las etapas de entrada Read from File, su flujo de datos tendrá ahora el siguiente aspecto:



El flujo de datos contiene ahora dos etapas Match Key Generator que producen claves de cruce para cada origen, y que usan exactamente las mismas reglas. Configurar de forma idéntica las etapas Match Key Generator es esencial para el apropiado funcionamiento de este flujo de datos.

11. Arrastre hacia el lienzo una etapa Interflow Match y conecte a esta cada una de las etapas Match Key Generator.

Por ejemplo, si está usando las etapas de entrada Read from File, su flujo de datos tendrá ahora el siguiente aspecto:



12. Haga doble clic en la etapa Interflow Match.
13. En el campo **Cargar regla de cruce** seleccione una de las reglas de cruce predefinidas, que puede usar como está o bien puede modificarla para que se ajuste a sus necesidades. Si desea crear una nueva regla de cruce sin usar una de las reglas de cruce predefinidas como punto de partida, haga clic en **Nueva**. Solamente puede tener una regla personalizada en un flujo de datos.

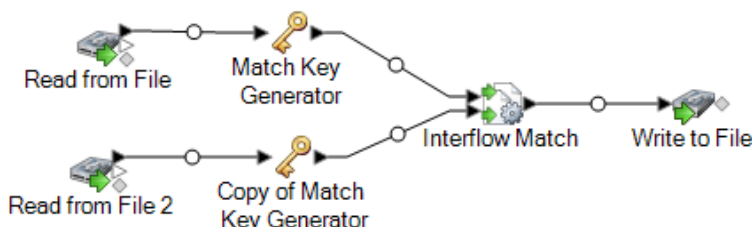
Nota: La función Opciones del flujo de datos en Enterprise Designer permite exponer la regla de cruce para la configuración durante la ejecución.

14. En el campo **Group by**, seleccione **MatchKey**.

Esto colocará en un grupo los registros que tengan la misma clave de cruce. La regla de cruce se aplica a los registros en un grupo para ver si hay duplicados. La clave de cruce para cada registro será generada por las etapas Generate Match Key que configuró antes en este procedimiento.

15. Para obtener más información acerca de cómo modificar las otras opciones, consulte [Generación de reglas de cruce](#) en la página 72.
16. Arrastre hacia el lienzo una etapa receptora y conéctela a la etapa Interflow Match.

Por ejemplo, si estuviese utilizando la etapa receptora Write to File, su flujo de datos tendría el siguiente aspecto:



17. Haga doble clic en la etapa receptora y configúrela.

Para obtener información acerca de cómo configurar etapas receptoras, consulte la *Guía de Dataflow Designer*.

Ahora cuenta con un flujo de datos que se corresponderá con los registros de dos orígenes de datos.

Ejemplo de Cruce de registros de múltiples orígenes

Como empresa de marketing directo, usted desea identificar a las personas que están en una lista de "no enviar correos" de modo que no se les envíen correos. Tiene una lista de destinatarios en un archivo, y en otro archivo tiene una lista de personas que no desean recibir publicidad por correo (es decir, un archivo para eliminaciones).

El siguiente flujo de datos entrega una solución para esta situación empresarial:

La etapa Read from File lee los datos de su lista de correos y la etapa Read from File 2 lee los datos de la lista de eliminaciones. Las dos etapas Match Key Generator están configuradas en forma idéntica, por lo que generan una clave de cruce que puede ser utilizada en Interflow Match para conformar grupos de posibles cruces. Interflow Match identifica registros en la lista de correos que están además en el archivo de eliminaciones y marca dichos registros como duplicados. Conditional Router envía registros únicos, lo que quiere decir que aquellos registros que no se encontraron en la lista de eliminaciones, son enviados a Write to File para que sean

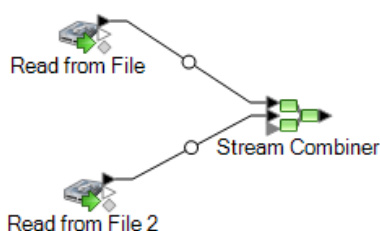
escritos en un archivo. La etapa Conditional Router envía todos los demás registros a Write to Null, en donde son descartados.

Coincidencia de registros entre y dentro orígenes

Este procedimiento describe cómo usar una etapa de Intraflow Match para identificar registros en un archivo que coincidan con otro archivo y en el mismo archivo. Por ejemplo, tiene dos archivos (archivos A y B) y desea ver si existen registros en el archivo A que coincidan con registros en el archivo B, pero también desea ver si existen registros en el archivo A que coincidan con otros registros en el archivo A. Puede lograr esto al usar Stream Combiner y una etapa de Intraflow Match.

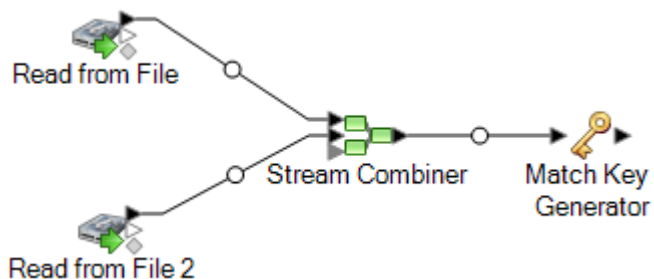
1. Genere un nuevo flujo de datos en Enterprise Designer.
2. Arrastre hacia el lienzo una etapa de origen.
3. Haga doble clic en la etapa de origen y configúrela. Consulte la *Guía de Dataflow Designer* para obtener instrucciones acerca de cómo configurar etapas de origen.
4. Arrastre una segunda etapa de origen al lienzo y configúrelo para que lea el segundo origen de datos en el flujo de datos.
5. Arrastre una etapa Stream Combiner sobre el lienzo y conecte las dos etapas de origen a esta última.

Por ejemplo, si su flujo de datos tuvo dos etapas Read From File, luciría así después de agregar Stream Combiner:



6. Arrastre una etapa de Match Key Generator hacia el lienzo y conéctela a la etapa Stream Combiner.

Por ejemplo, su flujo de datos podría tener ahora el siguiente aspecto:



Match Key Generator crea una clave no exclusiva para cada registro, que posteriormente se puede usar en las etapas de comparación a fin de identificar grupos de registros potencialmente duplicados. Las claves de cruce facilitan el proceso de comparación al permitir la agrupación de registros por clave de cruce y posteriormente solo comparando los registros al interior de estos grupos.

7. Haga doble clic en Match Key Generator.
8. Haga clic en **Agregar**.
9. Defina la regla que se usará para generar una clave de cruce para cada registro.

Tabla 5: Opciones de Match Key Generator

Nombre de la opción	Descripción / Valores válidos
---------------------	-------------------------------

Algoritmo	
-----------	--

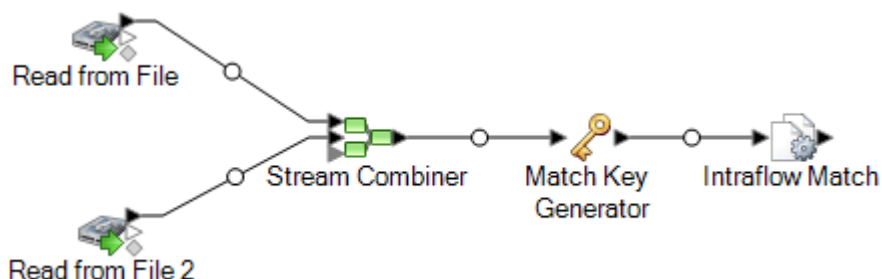
Nombre de la opción	Descripción / Valores válidos
	Especifica el algoritmo que se utilizará para generar la clave de cruce. Las opciones son:
Consonante	Devuelve campo especificados con consonantes eliminadas.
Doble metaphone	Devuelve un código basado en una representación fonética de sus caracteres. Double Metaphone es una versión mejorada del algoritmo Metaphone e intenta explicar la gran cantidad de irregularidades encontradas en los distintos idiomas.
Koeln	Indexa los nombres por sonido, tal como se los pronuncia en alemán. Permite que los nombres que tienen la misma pronunciación se codifiquen con la misma representación, de modo tal que puedan cruzarse a pesar de las pequeñas diferencias en el modo en que se escriben. El resultado es siempre una secuencia de números; los caracteres especiales y los espacios en blanco se ignoran. Esta opción se desarrolló para responder a las limitaciones de Soundex.
MD5	Un algoritmo de síntesis del mensaje que produce un valor hash de 128 bits. Este algoritmo suele usarse para comprobar la integridad de los datos.
Metaphone	Devuelve una clave codificada Metaphone de campos seleccionados. Metaphone es un algoritmo para codificar palabras utilizando su pronunciación del inglés.
Metaphone (español)	Devuelve una clave codificada Metaphone de campos seleccionados para el idioma español. El algoritmo Metaphone codifica palabras con base en su pronunciación en español.
Metaphone 3	Mejora los algoritmos de Metaphone y Double Metaphone con parámetros más exactos de consonantes y vocales internas que permiten pronunciar palabras y nombres que, fonéticamente, coinciden en mayor o menor medida con los términos de búsqueda. Metaphone 3 aumenta la precisión de la codificación fonética en un 98%. Esta opción se desarrolló para responder a las limitaciones de Soundex.
Nysiis	Algoritmo de codificación fonética que cruza una pronunciación aproximada con una palabra escrita exacta e indexa palabras que se pronuncian de manera similar. Parte del sistema de inteligencia e identificación del estado de Nueva York (New York State Identification and Intelligence System). Supongamos que está buscando información sobre alguien en una base de datos de personas. Cree que el nombre de esa persona suena algo así como "John Smith", pero en realidad se escribe "Jon Smyth". Si busca una coincidencia exacta para "John Smith" no obtendrá resultados. Sin embargo, si indexa la base de datos con el algoritmo de NYSIIS y realiza la búsqueda usando ese mismo algoritmo, obtendrá una coincidencia correcta porque el algoritmo indexa "John Smith" y "Jon Smyth" como "JAN SNATH".
Phonix	Preprocesa cadenas de nombres aplicando más de 100 reglas de transformación a caracteres únicos o secuencias de varios

Nombre de la opción	Descripción / Valores válidos
	<p>caracteres. Diecinueve de esas reglas se aplican solo si los caracteres se encuentran al comienzo de la cadena, mientras que 12 de ellas se aplican solo cuando los caracteres se encuentran en medio de la cadena, y 28 se aplican únicamente si están al final de la cadena. La cadena de nombre transformada se codifica en un código compuesto por la primera letra seguida de tres dígitos (sin ceros ni números duplicados). Esta opción se desarrolló como respuesta a las limitaciones de Soundex; es más compleja y, por lo tanto, más lenta que esta última.</p> <p>Soundex Devuelve un código Soundex de campos seleccionados. El algoritmo Soundex produce un código de longitud fija basado en la pronunciación en inglés de las palabras.</p> <p>Subcadena de caracteres Devuelve una porción específica del campo seleccionado.</p>
Nombre de campo	Especifica el campo en el que desea aplicar el algoritmo seleccionado para generar la clave de cruce. Por ejemplo, si selecciona un campo denominado LastName y elige el algoritmo Soundex, se aplicará ese algoritmo a los datos del campo LastName para generar una clave de cruce.
Posición de inicio	Especifica la posición de inicio dentro del campo especificado. No todos los algoritmos permiten especificar una posición inicial.
Longitud	Especifica la longitud, en caracteres, por incluirse desde la posición inicial. No todos los algoritmos permiten especificar una longitud.
Quitar caracteres irrelevantes	Elimina los caracteres que no son numéricos ni alfabéticos, como los guiones, los espacios en blanco y otros caracteres especiales de un campo de entrada.
Ordenar entrada	<p>Ordena alfabéticamente todos los caracteres de un campo de entrada o todos los términos de un campo de entrada.</p> <p>Caracteres Ordena los valores de los caracteres de un campo de entrada antes de crear una ID única:</p> <p>Términos Ordena el valor de cada término de un campo de entrada antes de crear una ID única.</p>

10. Cuando haya terminado de definir la regla, haga clic en **Aceptar**.

11. Si desea añadir reglas de cruce adicionales, haga clic en **Agregar** y añádalas, o bien haga clic en **Aceptar** cuando haya terminado.
12. Arrastre hacia el lienzo una etapa Interflow Match y conéctela a la etapa Match Key Generator.

Por ejemplo, su flujo de datos podría tener ahora el siguiente aspecto:



13. Haga doble clic en Intraflow Match.
14. En el campo **Cargar regla de cruce** seleccione una de las reglas de cruce predefinidas, que puede usar como está o bien puede modificarla para que se ajuste a sus necesidades. Si desea crear una nueva regla de cruce sin usar una de las reglas de cruce predefinidas como punto de partida, haga clic en **Nueva**. Solamente puede tener una regla personalizada en un flujo de datos.

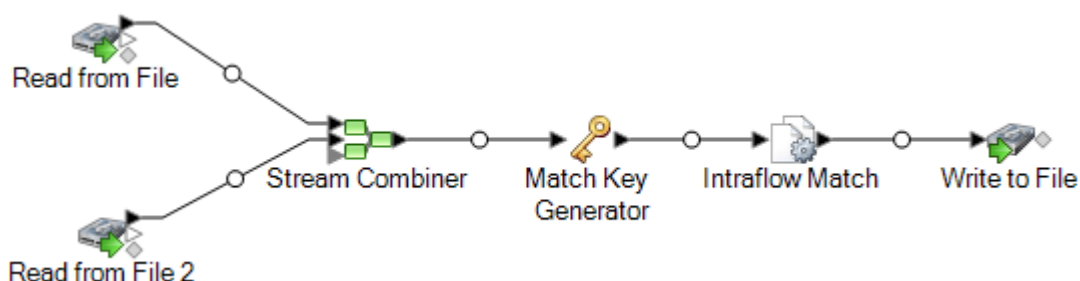
Nota: La función Opciones del flujo de datos en Enterprise Designer permite exponer la regla de cruce para la configuración durante la ejecución.

15. En el campo **Group by**, seleccione **MatchKey**.

Esto colocará en un grupo los registros que tengan la misma clave de cruce. La regla de cruce se aplica a los registros en un grupo para ver si hay duplicados. La clave de cruce de cada registro será generada por la etapa Generate Match Key que configuró antes en este procedimiento.

16. Para obtener más información acerca de cómo modificar las otras opciones, consulte [Generación de reglas de cruce](#) en la página 72.
17. Haga clic en **Aceptar** para guardar su configuración de Intraflow Match y vuelva al lienzo del flujo de datos.
18. Arrastre hacia el lienzo la etapa receptora y conéctela a la etapa Generate Match Key.

Por ejemplo, si estuviese utilizando la etapa receptora Write to File, su flujo de datos tendría el siguiente aspecto:



- Haga doble clic en la etapa receptora y configúrela.

Para obtener información acerca de cómo configurar etapas receptoras, consulte la *Guía de Dataflow Designer*.

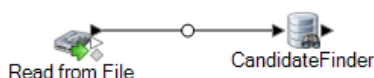
Comparación de registros con una base de datos

Este procedimiento describe la forma de cruzar registros en los cuales los registros sospechosos provienen de un origen que podría ser un archivo o base de datos, y los registros candidatos se encuentran en una base de datos con otros registros no relacionados. Para cada registro de entrada, el flujo de datos consulta la base de datos en busca de candidatos para dicho registro, y utiliza luego una etapa Transactional Match para comparar registros. Por último, el flujo de datos escribe las colecciones de registros coincidentes en un archivo de salida.

Nota: Transactional Match solo cruza registros sospechosos con candidatos. No intenta cruzar registros sospechosos con otros registros sospechosos, como se hace en Intraflow Match.

- Genere un nuevo flujo de datos en Enterprise Designer.
- Arrastre hacia el lienzo una etapa de origen.
- Haga doble clic en la etapa de origen y configúrela. Consulte la *Guía de Dataflow Designer* para obtener instrucciones acerca de cómo configurar etapas de origen.
- Arrastre hacia el lienzo una etapa Candidate Finder y conecte a esta la etapa de origen.

Por ejemplo, si estuviese usando la etapa de origen Read from File, su flujo de datos tendrá el siguiente aspecto:



Candidate Finder obtiene los registros candidatos que conformarán el conjunto de posibles cruces que Transactional Match evaluará más adelante en el flujo de datos.

- Haga doble clic en el lienzo en la etapa Candidate Finder.

6. En el campo **Conexión**, seleccione la base de datos que desea consultar para buscar registros candidatos. Si la base de datos que desea no figura en la lista, abra Management Console y defina antes la conexión de la base de datos.
7. En el campo SQL, ingrese una instrucción SQL `SELECT` que busca registros candidatos en función del valor en uno de los campos del flujo de datos. Para referenciar los campos del flujo de datos, use el formato `${FieldName}`, donde `FieldName` es el nombre del campo que desea referenciar.

Por ejemplo, si desea buscar registros en la base de datos, donde el valor en la columna `LastName` es el mismo que el campo de registros de flujo de datos, `Customer_LastName`, escribiría una instrucción SQL como la siguiente:

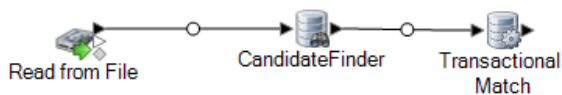
```
SELECT FirstName, LastName, Address, City, State, PostalCode
FROM Customer_Table
WHERE LastName = ${Customer_LastName};
```

8. En la ficha **Mapa de campo**, seleccione los campos del flujo de datos que deben contener los datos de cada columna de base de datos.

La columna **Campos seleccionados** lista las columnas de base de datos y **Campos de etapa** lista los campos del flujo de datos.

9. Haga clic en **Aceptar**.
10. Arrastre hacia el lienzo una etapa Transactional Match y conecte Candidate Finder a esta.

Por ejemplo, si está utilizando una etapa de entrada Read from File, su flujo de datos tendrá ahora el siguiente aspecto:



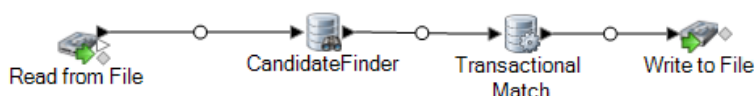
Transactional Match compara los registros sospechosos con los registros candidatos que se obtienen en la etapa Candidate Finder. Transactional Match utiliza reglas de cruce para comparar el registro sospechoso con todos los registros candidatos que tienen el mismo número de grupo de candidatos (asignado en Candidate Finder) a fin de identificar duplicados.

11. Haga doble clic en el lienzo en la etapa Transactional Match.
12. En el campo **Cargar regla de cruce** seleccione una de las reglas de cruce predefinidas, que puede usar como está o bien puede modificarla para que se ajuste a sus necesidades. Si desea crear una nueva regla de cruce sin usar una de las reglas de cruce predefinidas como punto de partida, haga clic en **Nueva**. Solamente puede tener una regla personalizada en un flujo de datos.

Nota: La función Opciones del flujo de datos en Enterprise Designer permite exponer la regla de cruce para la configuración durante la ejecución.

13. Para obtener más información acerca de cómo modificar las otras opciones, consulte [Generación de reglas de cruce](#) en la página 72.
14. Cuando haya terminado de configurar la etapa Transactional Match, haga clic en **Aceptar**.
15. Arrastre hacia el lienzo una etapa receptora y conéctela a la etapa Transactional Match.

Por ejemplo, si estuviese utilizando la etapa receptora Write to File, su flujo de datos tendría el siguiente aspecto:



16. Haga doble clic en la etapa receptora y configúrela.

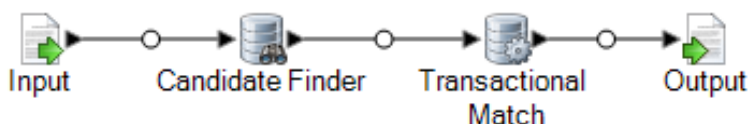
Para obtener información acerca de cómo configurar etapas receptoras, consulte la *Guía de Dataflow Designer*.

Ahora cuenta con un flujo de datos que se corresponderá con los registros de dos orígenes de datos.

Ejemplo de Comparación de registros con una base de datos

Como ejecutivo de cuenta de una empresa de ventas por Internet, quiere determinar si un posible cliente ya ha comprado o si es un cliente nuevo.

El siguiente servicio de flujo de datos ofrece una solución a la posible situación empresarial:



Este flujo de datos es un servicio que evalúa datos candidatos que se envían mediante una invocación a API o mediante una invocación de un servicio web. Evalúa los datos contra datos de clientes en una base de datos del cliente para determinar si un candidato es un cliente.

La etapa de entrada está configurada de modo que el flujo de datos acepta los siguientes campos de entrada: AddressLine1, City, Name, PostalCode y StateProvince. AddressLine1 y Name son los campos clave para el procesamiento del flujo de datos de esta plantilla.

La etapa Candidate Finder obtiene los registros candidatos que conformarán los posibles cruces que serán evaluados por la etapa Transactional Match.

La etapa Transactional Match compara registros sospechosos con los posibles registros candidatos que son devueltos en la etapa Candidate Finder. Transactional Match utiliza reglas de cruce para comparar el registro sospechoso con todos los registros candidatos que tienen el mismo número de grupo de candidatos (asignado

en Candidate Finder) a fin de identificar duplicados. En este ejemplo, Transactional Match compara LastName con AddressLine1.

La etapa Output devuelve los resultados del flujo de datos a través de una respuesta de API o de un servicio web.

Comparación de registros con varias reglas de cruce

■ Descargue el flujo de datos de muestra

Si posee registros que desea comparar y requiere usar más de una operación de comparación, puede crear un flujo de datos que utilice más de una clave de cruce y luego combine los resultados para realizar la comparación eficazmente a partir de varios criterios independientes. Por ejemplo, imagine que desea crear un flujo de datos que compare registros donde:

Coincida el nombre y la dirección

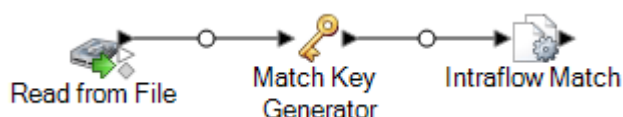
OR

Coincida la fecha de nacimiento y la identificación emitida por el gobierno

Para realizar la comparación mediante el uso de esta lógica, debe crear un flujo de datos que realice la comparación de nombre y dirección en una etapa y la de la fecha de nacimiento y la identificación emitida por el gobierno en otra etapa y luego combinar los registros coincidentes en una sola colección.

Este tema ofrece un procedimiento general para la configuración de un flujo de datos, donde la comparación se realiza durante el transcurso de dos etapas de comparación. Este procedimiento utiliza etapas Intraflow Match para propósitos de ilustración. Sin embargo, también puede utilizar esta técnica con Interflow Match.

1. Genere un nuevo flujo de datos en Enterprise Designer.
2. Arrastre hacia el lienzo una etapa de origen.
3. Haga doble clic en la etapa de origen y configúrela. Consulte la *Guía de Dataflow Designer* para obtener instrucciones acerca de cómo configurar etapas de origen.
4. Defina la primera sesión de comparación. Los resultados de esta primera sesión de comparación serán las colecciones de registros que coinciden con el primer conjunto de criterios de comparación, por ejemplo los registros que coinciden en el nombre y la dirección.
 - a) Arrastre una etapa Match Key Generator y una etapa Intraflow Match al lienzo y conéctelas de modo que el flujo de datos tenga la siguiente apariencia:



- a) En la etapa Match Key Generator, defina la clave de cruce que se va a utilizar para la primera sesión de comparación.

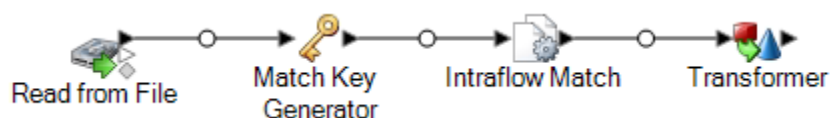
Por ejemplo, si desea que la primera sesión de comparación haga coincidir los nombres y las direcciones, puede crear una clave de cruce a partir de los campos que contienen el apellido y código postal.

- b) En la etapa Intraflow Match, defina las reglas de cruce que desea que ejecute la primera sesión de comparación.

Por ejemplo, puede configurar esta etapa de comparación para que haga coincidir el nombre y la dirección.

5. Guarde los números de colección de la primera etapa de comparación en otro campo. Esto es necesario ya que el campo CollectionNumber se sobrescribirá durante la segunda sesión de comparación. Se requiere cambiar el nombre del campo CollectionNumber para conservar los resultados de la primera sesión de comparación.

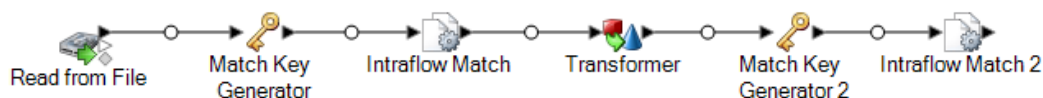
- a) Arrastre una etapa Transformer al lienzo y conéctela a la etapa Intraflow Match, de modo que el flujo de datos con la siguiente apariencia:



- b) Configure la etapa Transformer para cambiar el nombre del campo CollectionNumber a CollectionNumberPass1.

6. Defina la segunda sesión de comparación. Los resultados de esta segunda sesión de comparación serán las colecciones de registros que coinciden con el segundo conjunto de criterios de comparación, por ejemplo los registros de la fecha de nacimiento y la identificación emitida por el gobierno.

- a) Arrastre una etapa Match Key Generator y una etapa Intraflow Match al lienzo y conéctelas de modo que el flujo de datos tenga la siguiente apariencia:



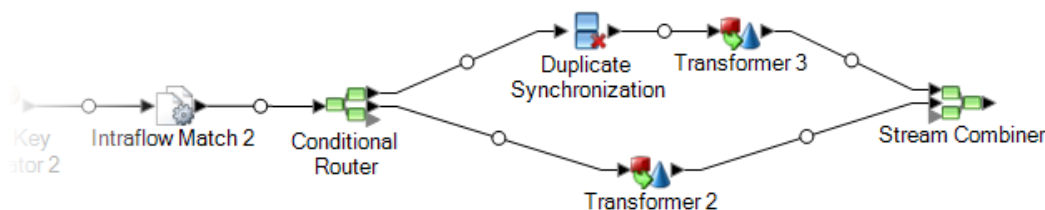
- b) En la segunda etapa Match Key Generator, defina la clave de cruce que se va a utilizar para la segunda sesión de comparación.

Por ejemplo, si desea que la segunda sesión de comparación haga coincidir la fecha de nacimiento y la identificación emitida por el gobierno, puede crear una clave de cruce a partir de los campos que contienen la fecha de nacimiento y la identificación emitida por el gobierno.

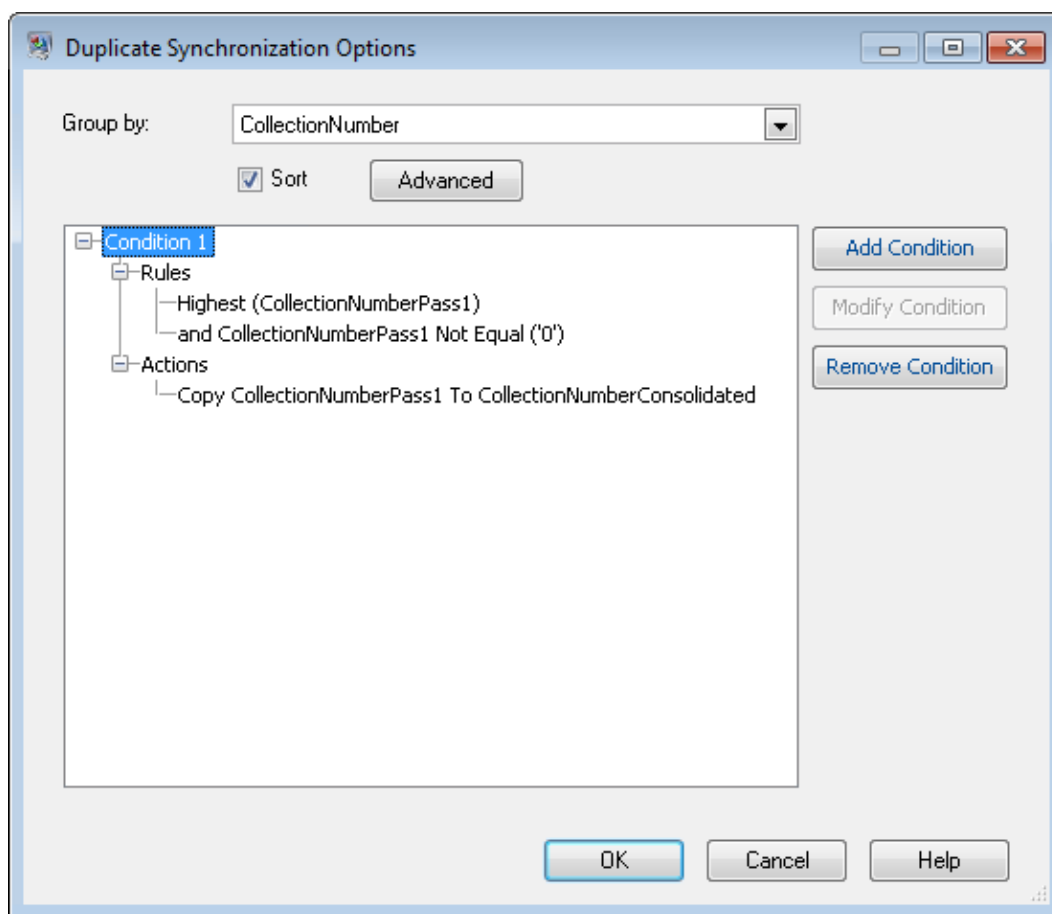
- c) En la segunda etapa Intraflow Match, defina la regla de cruce que se va a utilizar para la segunda sesión de comparación.

Por ejemplo, puede configurar esta etapa de comparación para que haga coincidir la fecha de nacimiento y la identificación emitida por el gobierno.

7. Determine si alguno de los registros duplicados identificados por la segunda sesión de comparación, también se identificaron como duplicados en la primera sesión de comparación.
 - a) Cree el fragmento de código del flujo de datos que se muestra a continuación de la segunda etapa Intraflow Match:



- b) Configure la etapa Conditional Router de modo que los registros donde el campo CollectionNumber no es igual a 0 se envíen a la etapa Duplicate Synchronization. Esto enviará los duplicados desde la segunda sesión de comparación a la etapa Duplicate Synchronization.
- c) Configure la etapa Duplicate Synchronization para agrupar los registros según el campo CollectionNumer (este es el número de colección desde la segunda etapa de comparación) Luego, dentro de cada colección, identifique si alguno de los registros en la colección también se identificó como duplicado en la primera sesión de comparación. Si los hubo, copie el número de colección de la primera sesión al nuevo campo denominado CollectionNumberConsolidated. Para hacer esto, configure Duplicate Synchronization como se muestra aquí:



- d) En la etapa Transformer que viene después de la etapa Duplicate Synchronization, cree una transformación personalizada mediante el uso de esta secuencia de comandos:

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

- e) En la etapa Transformer que viene inmediatamente después de Conditional Router (Transformer 2 en el flujo de datos de muestra) configure una transformación para copiar CollectionNumberPass1 en CollectionNumberConsolidated.

Esta acción toma los registros únicos desde la segunda sesión de comparación y copia CollectionNumberPass1 en CollectionNumberConsolidated.

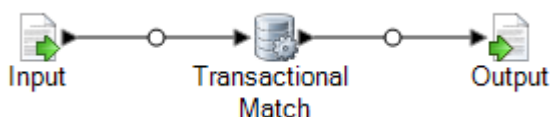
8. Después de Stream Combiner tendrá colecciones de registros que coinciden en cualquiera de las sesiones de comparación. El campo CollectionNumberConsolidated indica los registros coincidentes. Puede agregar un receptor o algún procesamiento adicional que desee realizar después de la etapa Stream Combiner.

Creación de un servicio de comparación universal

Un servicio de comparación universal es un servicio que puede usar cualquiera de sus reglas de cruce para realizar la comparación y puede admitir cualquier campo de entrada. El servicio toma un nombre de regla de cruce como una opción y le permite especificar la regla de cruce que desee usar en la invocación de API o la solicitud de servicio web. Este servicio no tiene un esquema de entrada predefinido, de modo que puede incluir cualquier campo que sea apropiado para el tipo de registro que desee comparar. Al crear un servicio de comparación universal se evita tener servicios separados para cada regla de cruce, lo que le permite agregar nuevas reglas de cruce sin tener que agregar un servicio.

Este procedimiento muestra cómo crear un servicio de comparación universal e incluye un ejemplo de una solicitud de servicio web para un servicio de comparación universal.

1. Genere un nuevo flujo de datos del servicio en Enterprise Designer.
2. Arrastre una etapa Input, una etapa Transactional Match y una etapa Output al lienzo y conéctelas de modo que el flujo de datos tenga la siguiente apariencia:



3. Haga doble clic en la etapa Transactional Match.
4. En el campo **Cargar regla de cruce** seleccione cualquier regla de cruce. Por ejemplo, puede seleccionar la regla de cruce predeterminada **Hogar**.

Aunque especificará la regla de cruce en la solicitud de servicio, debe configurar la etapa Transactional Match con una regla de cruce predeterminada para validar el flujo de datos. Si no selecciona una regla de cruce, el flujo de datos no se validará y no podrá exponerlo.

5. Haga clic en **Aceptar**.
6. Haga doble clic en la etapa Output.
7. Elija exponer los campos `MatchRecordType` y `MatchScore`.
8. Haga clic en **Aceptar**.

Nota: No se requiere exponer campos en la etapa Input, ya que los campos de entrada se especificarán como campos definidos por el usuario en la solicitud de servicio.

9. Haga clic en **Editar > Opciones del flujo de datos**.
10. Haga clic en **Agregar**.
11. Expanda **Transactional Match** y marque la casilla junto a **Regla de cruce**.

Esto expone la opción de regla de cruce como una opción de tiempo de ejecución y, así, permite especificar la regla de cruce en la solicitud de servicio.

12. Haga clic en **Aceptar** luego haga clic en **Aceptar** nuevamente para cerrar la ventana **Opciones del flujo de datos**.
13. Guarde y exponga el flujo de datos.

Ahora dispone de un servicio de comparación universal, que puede utilizar para realizar una comparación mediante el uso de cualquier regla de cruce definida en la herramienta Match Rules Management en Enterprise Designer. Cuando invoque el servicio, especifique una regla de cruce en la opción `MatchRule` y especifique los campos de entrada como campos definidos por el usuario.

Ejemplo: invocación del servicio de cruce universal

Creo una regla de cruce denominada `AddressAndBirthday` en la herramienta Match Rules Management. Esta regla de cruce compara los registros mediante el uso de los campos Dirección y Cumpleaños. Desea utilizar el servicio de comparación universal para realizar la comparación mediante el uso de esta regla, a través de una solicitud de servicio web de SOAP.

Para lograrlo, tendría una solicitud SOAP que especifique `AddressAndBirthday` en el elemento `MatchRule` y los campos del registro en el elemento `user_fields`.

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

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

<univ:MatchRule>AddressAndBirthday</univ:MatchRule>
      </univ:options>
      <univ:Input>
        <univ:Row>
          <univ:user_fields>
            <univ:user_field>
              <univ:name>Name</univ:name>
              <univ:value>Bob Smith</univ:value>
            </univ:user_field>
            <univ:user_field>
              <univ:name>Address</univ:name>
              <univ:value>4200 Parliament
Pl</univ:value>
            </univ:user_field>
            <univ:user_field>
              <univ:name>Birthday</univ:name>
              <univ:value>1973-6-15</univ:value>
            </univ:user_field>
          </univ:user_fields>
        </univ:Row>
        <univ:Row>
          <univ:user_fields>
```



```

        <univ:user_field>
          <univ:name>Name</univ:name>
          <univ:value>Robert M. Smith</univ:value>

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

```

Esta solicitud produciría la siguiente respuesta:

```

<soap:Envelope
xmlns:soap="http://schemas.xmlsoap.org/soap/envelope/">
  <soap:Body>
    <ns3:UniversalMatchingServiceResponse
xmlns:ns2="http://spectrum.pb.com/"

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

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

```

```

<ns3:MatchRecordType>Suspect</ns3:MatchRecordType>
  <ns3:user_fields>
    <ns3:user_field>
      <ns3:name>Name</ns3:name>
      <ns3:value>Bob Smith</ns3:value>
    </ns3:user_field>
    <ns3:user_field>
      <ns3:name>Birthday</ns3:name>
      <ns3:value>1973-6-15</ns3:value>
    </ns3:user_field>
    <ns3:user_field>
      <ns3:name>Address</ns3:name>
      <ns3:value>4200 Parliament Pl</ns3:value>
    </ns3:user_field>
  </ns3:user_fields>
</ns3:Row>
<ns3:Row>
  <ns3:MatchScore>100</ns3:MatchScore>
</ns3:MatchRecordType>Duplicate</ns3:MatchRecordType>
  <ns3:user_fields>
    <ns3:user_field>
      <ns3:name>Name</ns3:name>
      <ns3:value>Robert M. Smith</ns3:value>
    </ns3:user_field>
    <ns3:user_field>
      <ns3:name>Birthday</ns3:name>
      <ns3:value>1973-6-15</ns3:value>
    </ns3:user_field>
    <ns3:user_field>
      <ns3:name>Address</ns3:name>
      <ns3:value>4200 Parliament Pl</ns3:value>
    </ns3:user_field>
  </ns3:user_fields>
</ns3:Row>
</ns3:Output>
</ns3:UniversalMatchingServiceResponse>
</soap:Body>
</soap:Envelope>

```

Uso de una clave de cruce inmediato

El cruce de clave inmediato puede ser una herramienta útil para reducir el número de comparaciones realizadas y, de ese modo, mejorar la velocidad de ejecución de flujos de datos que utilizan una etapa de Interflow Match o Intraflow Match. Si dos registros presentan un cruce exacto en la clave inmediata, el candidato se considera un duplicado 100% y no se realizan más intentos. Si dos registros no presentan un cruce en el valor de clave inmediata, se los compara utilizando el método basado en reglas. Sin embargo, una clave inmediata imprecisa genera muchos falsos positivos en el cruce.

1. Abra el flujo de datos en Enterprise Designer.
2. Haga doble clic en la etapa Match Key Generator.
3. Active la casilla **Generar clave de cruce inmediato**.
4. Haga clic en **Agregar**.
5. Complete los siguientes campos:

Tabla 6: Opciones de Match Key Generator

Nombre de la opción	Descripción / Valores válidos
---------------------	-------------------------------

Algoritmo	
-----------	--

Nombre de la opción	Descripción / Valores válidos
	Especifica el algoritmo que se utilizará para generar la clave de cruce. Las opciones son:
Consonante	Devuelve campo especificados con consonantes eliminadas.
Doble metaphone	Devuelve un código basado en una representación fonética de sus caracteres. Double Metaphone es una versión mejorada del algoritmo Metaphone e intenta explicar la gran cantidad de irregularidades encontradas en los distintos idiomas.
Koeln	Indexa los nombres por sonido, tal como se los pronuncia en alemán. Permite que los nombres que tienen la misma pronunciación se codifiquen con la misma representación, de modo tal que puedan cruzarse a pesar de las pequeñas diferencias en el modo en que se escriben. El resultado es siempre una secuencia de números; los caracteres especiales y los espacios en blanco se ignoran. Esta opción se desarrolló para responder a las limitaciones de Soundex.
MD5	Un algoritmo de síntesis del mensaje que produce un valor hash de 128 bits. Este algoritmo suele usarse para comprobar la integridad de los datos.
Metaphone	Devuelve una clave codificada Metaphone de campos seleccionados. Metaphone es un algoritmo para codificar palabras utilizando su pronunciación del inglés.
Metaphone (español)	Devuelve una clave codificada Metaphone de campos seleccionados para el idioma español. El algoritmo Metaphone codifica palabras con base en su pronunciación en español.
Metaphone 3	Mejora los algoritmos de Metaphone y Double Metaphone con parámetros más exactos de consonantes y vocales internas que permiten pronunciar palabras y nombres que, fonéticamente, coinciden en mayor o menor medida con los términos de búsqueda. Metaphone 3 aumenta la precisión de la codificación fonética en un 98%. Esta opción se desarrolló para responder a las limitaciones de Soundex.
Nysiis	Algoritmo de codificación fonética que cruza una pronunciación aproximada con una palabra escrita exacta e indexa palabras que se pronuncian de manera similar. Parte del sistema de inteligencia e identificación del estado de Nueva York (New York State Identification and Intelligence System). Supongamos que está buscando información sobre alguien en una base de datos de personas. Cree que el nombre de esa persona suena algo así como "John Smith", pero en realidad se escribe "Jon Smyth". Si busca una coincidencia exacta para "John Smith" no obtendrá resultados. Sin embargo, si indexa la base de datos con el algoritmo de NYSIIS y realiza la búsqueda usando ese mismo algoritmo, obtendrá una coincidencia correcta porque el algoritmo indexa "John Smith" y "Jon Smyth" como "JAN SNATH".
Phonix	Preprocesa cadenas de nombres aplicando más de 100 reglas de transformación a caracteres únicos o secuencias de varios

Nombre de la opción	Descripción / Valores válidos
	<p>caracteres. Diecinueve de esas reglas se aplican solo si los caracteres se encuentran al comienzo de la cadena, mientras que 12 de ellas se aplican solo cuando los caracteres se encuentran en medio de la cadena, y 28 se aplican únicamente si están al final de la cadena. La cadena de nombre transformada se codifica en un código compuesto por la primera letra seguida de tres dígitos (sin ceros ni números duplicados). Esta opción se desarrolló como respuesta a las limitaciones de Soundex; es más compleja y, por lo tanto, más lenta que esta última.</p> <p>Soundex Devuelve un código Soundex de campos seleccionados. El algoritmo Soundex produce un código de longitud fija basado en la pronunciación en inglés de las palabras.</p> <p>Subcadena de caracteres Devuelve una porción específica del campo seleccionado.</p>
Nombre de campo	Especifica el campo en el que desea aplicar el algoritmo seleccionado para generar la clave de cruce. Por ejemplo, si selecciona un campo denominado LastName y elige el algoritmo Soundex, se aplicará ese algoritmo a los datos del campo LastName para generar una clave de cruce.
Posición de inicio	Especifica la posición de inicio dentro del campo especificado. No todos los algoritmos permiten especificar una posición inicial.
Longitud	Especifica la longitud, en caracteres, por incluirse desde la posición inicial. No todos los algoritmos permiten especificar una longitud.
Quitar caracteres irrelevantes	Elimina los caracteres que no son numéricos ni alfabéticos, como los guiones, los espacios en blanco y otros caracteres especiales de un campo de entrada.
Ordenar entrada	<p>Ordena alfabéticamente todos los caracteres de un campo de entrada o todos los términos de un campo de entrada.</p> <p>Caracteres Ordena los valores de los caracteres de un campo de entrada antes de crear una ID única:</p> <p>Términos Ordena el valor de cada término de un campo de entrada antes de crear una ID única.</p>

6. Haga clic en **Aceptar**.

7. Si desea especificar un campo y/o algoritmo adicional que se utiliza para generar una clave de cruce inmediato, haga clic en **Agregar** , de lo contrario, haga clic en **Aceptar**.
8. Haga doble clic en la etapa de Interflow Match o Intraflow Match en el lienzo.
9. Seleccione la opción **Cruce inmediato en** y seleccione el campo **ExpressMatchKey**.
Este campo contiene la clave de cruce inmediato producida por Match Key Generator.
10. Haga clic en **Aceptar**.
11. Guarde y ejecute su flujo de datos.

Para determinar si un candidato presenta un cruce utilizando una clave inmediata, observe el valor del campo **ExpressKeyIdentified** que será Y si hay cruce y N si no. Cabe destacar que los registros sospechosos siempre tienen un valor de N en **ExpressKeyIdentified**.

Análisis de resultados de cruce

La herramienta de análisis de cruce en Enterprise Designer muestra los resultados de una o más etapas de comparación del mismo tipo. La herramienta proporciona resultados de cruce de resumen para un flujo de datos y también le permite ver los resultados de cruce registro por registro. Puede usar esta información para solucionar problemas o ajustar las reglas de cruce para que produzcan los resultados que desea.

La herramienta de análisis de cruce ofrece las siguientes características:

- Resultados resumidos de cruce: muestra recuentos de registros resumidos para un solo resultado de cruce o bien comparaciones entre dos resultados de cruce.
- Gráficos de elevación/caída: usa gráficos de barras para mostrar un aumento o disminución en los cruces.
- Reglas de cruce: muestra las reglas de cruce utilizadas para un solo resultado de cruce o los cambios realizados en las reglas de cruce al comparar dos resultados.
- Resultados detallados de cruce: muestra información detallada sobre el procesamiento de registros para un solo resultado de cruce o bien la comparación entre dos resultados.

Ver un resumen de resultados de comparación

La herramienta de Análisis de cruce puede mostrar información de resumen sobre los procesos de correspondencia en un flujo de datos, como el número de registros duplicados, la calificación de cruce promedio, y así sucesivamente. Puede ver los resultados de un solo trabajo o puede comparar los resultados entre varios trabajos.

1. En Enterprise Designer, abra el flujo de datos que desea analizar.

- Para etapa de correspondencia de Interflow Match, Intraflow Match, o Transactional, cuyas correspondencias desee analizar, haga doble clic en la etapa y seleccione la casilla de verificación **Generar datos para análisis**.

Importante: Si se activa la opción **Generar datos para análisis** se reduce el rendimiento. Debe desactivar esta opción cuando haya terminado de utilizar la herramienta Análisis de cruce.

- Seleccione **Ejecutar > Ejecutar flujo actual**

Nota: Para obtener resultados óptimos, utilice datos que producirán 100 000 o menos registros. Mientras más resultados de cruce haya, más lento será el rendimiento de la herramienta Análisis de cruce.

- Cuando finalice la ejecución del flujo de datos, seleccione **Herramientas > Análisis de cruce**.

En el cuadro de diálogo **Examinar resultados de cruce** se muestra una lista de flujos de datos que tienen resultados de cruce que se ven en la herramienta Análisis de cruce. Si el trabajo que desea analizar no se encuentra en la lista, abra el flujo de datos y asegúrese de que la etapa tenga seleccionada la casilla de verificación **Generar datos para análisis**.

Consejo: Si hay un gran número de flujos de datos y si desea filtrarlos, seleccione una opción de filtro de la lista desplegable **Mostrar solo trabajos donde**.

- Haga clic con el botón secundario en el ícono " +" al lado del flujo de datos que desee ver para ampliarlo.
- En el flujo de datos hay una entrada para cada etapa de cruce en el flujo de datos. Seleccione la etapa que desea ver y haga clic en **Agregar**.

La herramienta Análisis de cruce aparece en la parte inferior de la ventana del Enterprise Designer.

- Si desea comparar los resultados de cruce al lado de los resultados de otra función de cruce:
 - Haga clic en **Agregar**.
 - Seleccione la función de cruce cuyos resultados desea comparar.
 - Haga clic en **Agregar**.
 - En la lista de flujo de datos, seleccione la función de cruce que acaba de agregar y haga clic en **Comparar**.

En la pestaña **Resumen** se muestra una lista con las estadísticas de cruce del trabajo. Dependiendo del tipo de etapa de cruce utilizado en el flujo de datos, podrá ver información diferente.

Para Intraflow Match, verá la siguiente información de resumen:

Registros de entrada El número total de registros procesados por la función de cruce.

Registros únicos Registro sospechoso o candidato que no coincide con ningún otro registro en un grupo de cruce. Si es el único registro de un grupo de coincidencia, el sospechoso automáticamente es único.

Grupos de cruce (Agrupar por) Los registros están agrupados por una clave de cruce o una ventana dinámica.

Colecciones duplicadas	Una colección duplicada consiste en un registro sospechoso y sus registros duplicados agrupados mediante CollectionNumber. Los registros únicos siempre corresponden a CollectionNumber 0.
Cruces inmediatos	Un cruce inmediato se realiza cuando un sospechoso y un candidato tienen una coincidencia exacta en los contenidos de un campo designado; por lo general, es un campo ExpressMatchKey provisto por Match Key Generator (Generador de clave de cruce). Si se realiza un cruce inmediato, no se realizan más procesamientos para determinar si el sospechoso y el candidato son duplicados.
Calificación promedio	La media coincide con puntuación de todos los duplicados. Los valores posibles son 0-100; 0 indica una coincidencia poco satisfactoria y 100, una coincidencia exacta.

Para Interflow Match, verá la siguiente información de resumen:

Colecciones duplicadas	Una colección duplicada consiste en un registro sospechoso y sus registros duplicados agrupados mediante CollectionNumber. Los registros únicos siempre corresponden a CollectionNumber 0.
Cruces inmediatos	Un cruce inmediato se realiza cuando un sospechoso y un candidato tienen una coincidencia exacta en los contenidos de un campo designado; por lo general, es un campo ExpressMatchKey provisto por Match Key Generator (Generador de clave de cruce). Si se realiza un cruce inmediato, no se realizan más procesamientos para determinar si el sospechoso y el candidato son duplicados.
Calificación promedio	La media coincide con puntuación de todos los duplicados. Los valores posibles son 0-100; 0 indica una coincidencia poco satisfactoria y 100, una coincidencia exacta.
Sospechosos de entrada	El número de registros en el flujo de entrada que el comprobador intentó hacer coincidir con otros registros.
Sospechosos con duplicados	El número de sospechosos de entrada que se cruzan con al menos uno de los registros de candidatos.
Sospechosos únicos	El número de sospechosos de entrada que no se cruzan con ninguno de los registros de candidatos.
Sospechosos con candidatos	El número de sospechosos de entrada que tenían por lo menos a un candidato de su grupo de coincidencia y que, por lo tanto, tenían al menos un intento de cruce.
Sospechosos sin candidatos	El número de sospechosos de entrada que no tenían ningún candidato en su grupo de cruce y que, por lo tanto, no tenían intentos de cruce.

Para Transactional Match, verá la siguiente información de resumen:

Calificación promedio	La media coincide con puntuación de todos los duplicados. Los valores posibles son 0-100; 0 indica una coincidencia poco satisfactoria y 100, una coincidencia exacta.
Sospechosos de entrada	El número de registros en el flujo de entrada que el comprobador intentó hacer coincidir con otros registros.

Sospechosos con duplicados

El número de sospechosos de entrada que se cruzan con al menos uno de los registros de candidatos.

Sospechosos únicos

El número de sospechosos de entrada que no se cruzan con ninguno de los registros de candidatos.

Sospechosos con candidatos

El número de sospechosos de entrada que tenían por lo menos a un candidato de su grupo de coincidencia y que, por lo tanto, tenían al menos un intento de cruce.

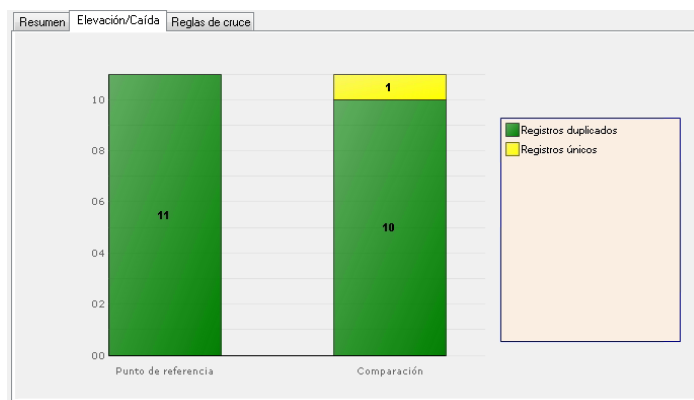
Sospechosos sin candidatos El número de sospechosos de entrada que no tenían ningún candidato en su grupo de cruce y que, por lo tanto, no tenían intentos de cruce.

Los gráficos de la pestaña **elevación/caída** de la herramienta de Análisis de cruce muestran los recuentos de registros duplicados y únicos en un gráfico de barras para los resultados de punto de referencia seleccionados y, opcionalmente, para los de comparación. La *elevación* se define como el incremento en la cantidad de registros duplicados. La *caída* se define como la reducción en la cantidad de registros duplicados. Los registros únicos se muestran en amarillo y los registros duplicados, en verde.

Si sólo un trabajo de punto de referencia está seleccionado, el gráfico mostrará los resultados para ese trabajo:



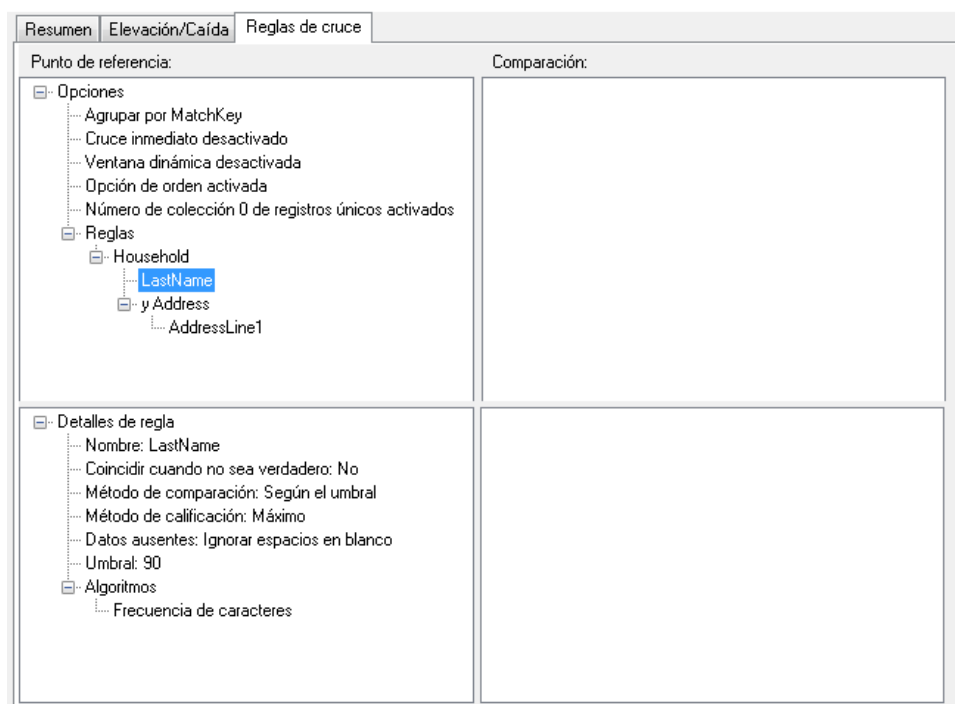
Si se seleccionan un trabajo de punto de referencia y un trabajo de comparación, se muestra un gráfico con los trabajos de punto de referencia al lado de los trabajos de comparación:



En la pestaña **Reglas de cruce** se muestran las reglas de cruce utilizadas para un solo resultado de cruce o los cambios realizados en las reglas de cruce al comparar dos resultados.

Las reglas de cruce se muestran en una estructura jerárquica, similar a la forma en que se visualizan en el cuadro de la etapa al crearse. La jerarquía de regla contiene dos nodos: opciones y reglas. El nodo Opciones muestra la configuración de la etapa correspondiente al resultado de cruce seleccionado. El nodo Reglas muestra las reglas de cruce correspondientes al resultado de cruce seleccionado.

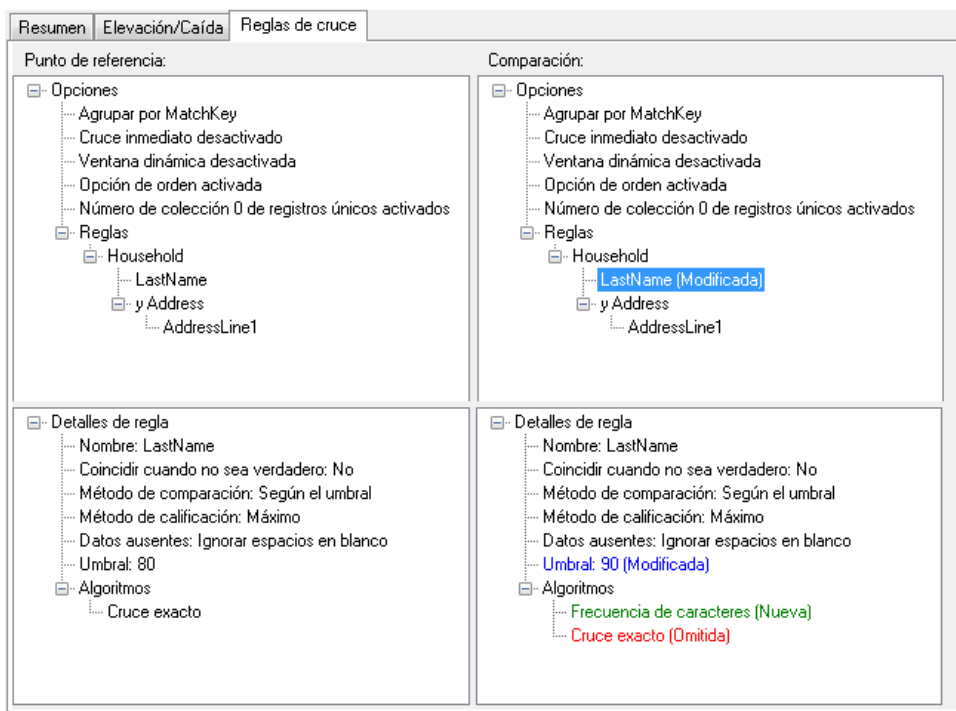
Para ver detalles de reglas, seleccione un nodo en la jerarquía.



Si compara las reglas de cruce entre varios trabajos, las diferencias entre los resultados de punto referencia y de cruce de comparación se codifican con colores de la siguiente forma:

- Azul** Indica que la regla de cruce en el resultado de comparación fue modificada.
- Verde** Indica que la regla de cruce en el resultado de comparación fue agregada.
- Rojo** Indica que la regla de cruce en el resultado de comparación fue omitida.

Por ejemplo:



Visualización de resultados de cruce de nivel de registro

Resultados detallados muestra una colección de los detalles sobre los registros de cruce para el conjunto de resultados de cruce.

Para mostrar los resultados detallados:

1. En la herramienta Análisis de cruce, especifique un trabajo de punto de referencia y opcionalmente, un trabajo de comparación.
2. Haga clic en **Detalles**.

Los resultados de cruce de punto de referencia se muestran conforme a la vista seleccionada en la lista desplegable Mostrar. La siguiente tabla enumera las columnas que se muestran para cada tipo de etapa de cruce.

Tabla 7: Datos de resultados detallados visualizados

Resultados relacionados con la información detallada	Intraflujo	Interflujo	Transaccional
Número de registro de entrada	X	X	X

Resultados relacionados con la información detallada	Intraflujo	Interflujo	Transaccional
Grupo de cruce	X	X	
Clave inmediata	X	X	
Registro de controlador de clave inmediata	X	X	
Número de colección	X	X	X
Tipo de registro de cruce	X	X	X
Campos usados por las reglas	X	X	X
Calificación de regla general (nivel superior)	X		
Grupo de candidatos		X	X
Calificación de cruce Seleccione un resultado de cruce de la lista Resultados de cruce y haga clic en Eliminar .		X	X

Para obtener más información acerca del gráfico de índice de cruce, consulte [Gráfico de índice de cruce](#) en la página 127.

3. En el campo **Analizar** elija una de las siguientes opciones:

Punto de referencia	Muestra los resultados de cruce desde la ejecución del punto de referencia.
Comparación	Muestra los resultados de cruce desde la ejecución de la comparación.

4. Seleccione uno de los siguientes valores en la lista **mostrar** y luego haga clic en **Actualizar**. Si analiza resultados del punto de referencia, las opciones son:

- Sospechosos con candidatos: (todos los cruces) muestra los registros sospechosos y todos los registros candidatos que se intentaron cruzar con cada sospechoso.

- Sospechosos con duplicados: (todos los cruces) muestra todos los registros sospechosos y los registros candidatos cruzados con cada sospechoso.
- Sospechosos con cruces inmediatos: (Interflow Match [Cruce de interflujo] e Intraflow Match [Cruce de intraflujo], cuando la clave de cruce inmediato está activada) muestra los registros sospechosos y candidatos que coinciden sobre la base de la clave de cruce inmediato.
- Colecciones duplicadas: (Intraflow e Interflow) muestra todas las colecciones duplicadas por número de colección.
- Grupos de cruce: (Intraflow e Interflow) muestra los registros por grupos de cruce.
- Grupo de candidatos: (Transactional Match) muestra los registros por grupos de candidatos.
- Sospechosos únicos: (Interflow y Transactional Match) muestra todos los registros sospechosos que no coinciden con ningún registro candidato.
- Registros únicos: (Intraflow) muestra todos los registros sin cruce coincidente.
- Sospechosos sin candidatos: (Interflow y Transactional Match) muestra todos los sospechosos que no contienen candidatos para cruzar.
- Todos los registros: muestra todos los registros procesados por la etapa de comparación.

Si analiza resultados de comparación, las opciones de muestra son:

- Nuevos cruces: (Intraflow) muestra todos los nuevos cruces y sus sospechosos relacionados. Esta vista combina los resultados de las vistas Sospechosos con nuevos duplicados y Nuevos sospechosos en una sola vista.
 - Nuevos sospechosos cruzados: (Interflow y Transactional Match) muestra los sospechosos que no tienen duplicados en el punto de referencia pero tienen al menos un duplicado en la comparación.
 - Nuevos sospechosos únicos: (Interflow y Transactional Match) muestra los sospechosos que tienen duplicados en el punto de referencia y no tienen ninguno en la comparación.
 - Cruces perdidos: (Intraflow) muestra todas las coincidencias. Esta vista combina los resultados de las vistas Sospechosos con duplicados perdidos y Sospechosos perdidos en una sola vista.
 - Sospechosos con nuevos duplicados: (todos los cruces) muestra los registros que son nuevos duplicados para los registros que figuraron como sospechosos en el punto de referencia y permanecieron como tales en la comparación.
 - Sospechosos con duplicados perdidos: (todos los cruces) muestra los registros que son duplicados perdidos para los registros que figuraron como sospechosos en el punto de referencia y permanecieron como tales en la comparación.
 - Nuevos sospechosos: (Intraflow) muestra los registros que son sospechosos en los resultados de cruce de comparación pero que no eran sospechosos en el punto de referencia.
 - Sospechosos perdidos: (Intraflow) muestra los registros que no son sospechosos en los resultados de la comparación pero eran sospechosos en el punto de referencia.
5. Expanda un registro sospechoso para ver sus candidatos.
 6. Seleccione un registro candidato y haga clic en **Detalles**.

Nota: Esta opción no se encuentra disponible si la opción Ventana dinámica está activada en las etapas Intraflow Match (Cruce de intraflujo).

La ventana **Detalles del registro** muestra los datos de nivel de campo, así como también la calificación de cruce del registro para cada regla de cruce. Si especifica una ejecución de trabajo de punto de referencia y comparación, puede visualizar los resultados del registro tanto en las ejecuciones de punto de referencia como de comparación.

- Entrada de punto de referencia: muestra los datos de nivel de campo del sospechoso y el candidato que se utilizaron en el cruce.
- Detalles de cruce de punto de referencia: muestra información de calificación para cada nodo en las reglas de cruce.
- Entrada de comparación: muestra los datos de nivel de campo del sospechoso y el candidato que se utilizaron en el cruce.
- Detalles de cruce de comparación: muestra la información de calificación para cada nodo en las reglas de cruce. El texto de color verde representa un cruce (coincidencia) para un nodo en las reglas. El texto de color rojo representa una falta de coincidencia para un nodo en las reglas.

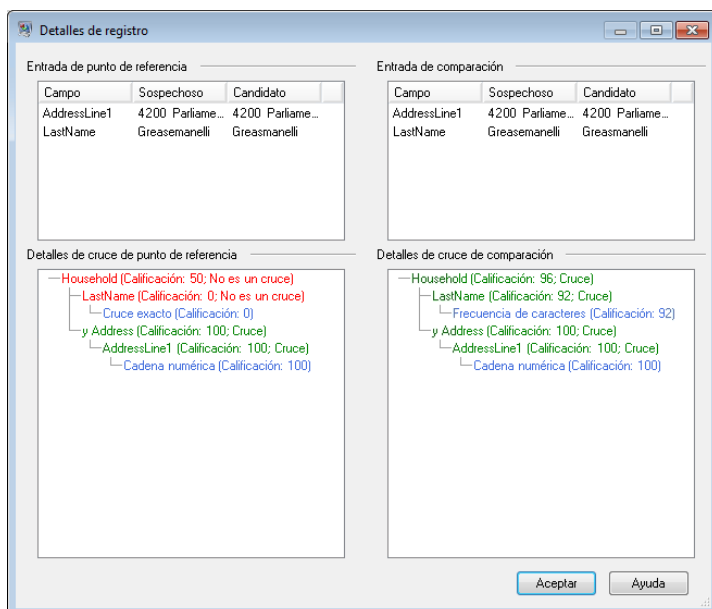
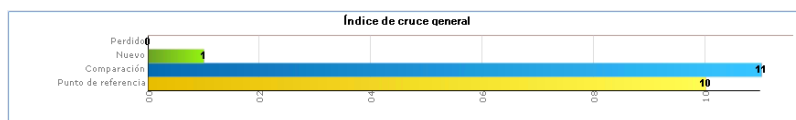


Gráfico de índice de cruce

Los gráficos de índice de cruce muestran la información de cruce en vistas detalladas.



Para los cruces de intraflujo, se muestra un gráfico de cruces generales:


- Cruces de punto de referencia: cantidad total de cruces en los resultados de punto de referencia. Cruces de punto de referencia.
- Cruces de comparación: cantidad total de cruces en los resultados de comparación.

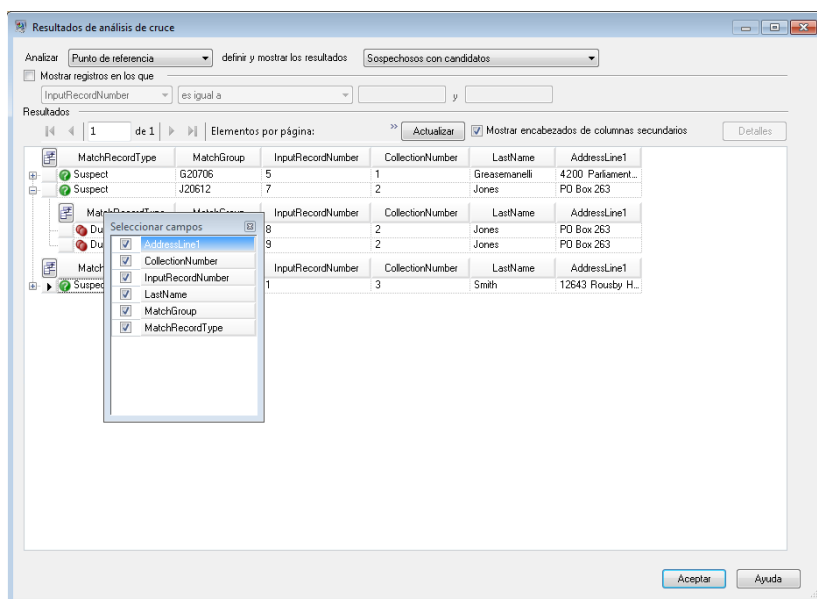
- Nuevos cruces: un recuento de todos los registros que fueron únicos en los resultados de punto de referencia pero aparecen como sospechosos o duplicados en los resultados de comparación.
- Cruces perdidos: un recuento de todos los registros que fueron sospechosos o duplicados en los resultados de punto de referencia pero aparecen como únicos en los resultados de comparación.

Para los cruces de interflujo y transaccionales, se muestran dos gráficos:

- Índice de cruce general
- Cruces de punto de referencia: cantidad total de cruces en los resultados de punto de referencia. Cruces de punto de referencia.
- Cruces de comparación: cantidad total de cruces en los resultados de comparación.
- Nuevos cruces: un recuento de todos los registros que fueron únicos en los resultados de punto de referencia pero aparecen como sospechosos o duplicados en los resultados de comparación.
- Cruces perdidos: un recuento de todos los registros que fueron sospechosos o duplicados en los resultados de punto de referencia pero aparecen como únicos en los resultados de comparación.
- Índice de cruce de registros sospechosos
- Cruces de punto de referencia: un recuento de todos los registros sospechosos que no fueron únicos en el punto de referencia.
- Cruces de comparación: un recuento de todos los registros sospechosos que no fueron únicos en la comparación.
- Nuevos cruces: un recuento de todos los registros sospechosos que fueron únicos en el punto de referencia pero aparecen como cruces con coincidencia en los resultados de comparación.
- Cruces perdidos: un recuento de todos los registros sospechosos que fueron cruces con coincidencia en el punto de referencia pero aparecen como únicos en los resultados de comparación.

Uso del selector de campos

Haga clic en el icono del selector de campos  para mostrar las columnas seleccionadas en los resultados de análisis de cruce. El selector de campos se muestra en el nivel principal y secundario. Además, usted puede seleccionar de forma independiente columnas de visualización para los elementos principales y secundarios.



Filtro de registros

Utilice la casilla **Mostrar registros en los que** para filtrar los registros de cruce detallados que se mostrarán. Puede filtrar los registros de acuerdo con diferentes operadores para comparar los valores proporcionados por el usuario con los datos de un campo de cada registro de cruce detallado.

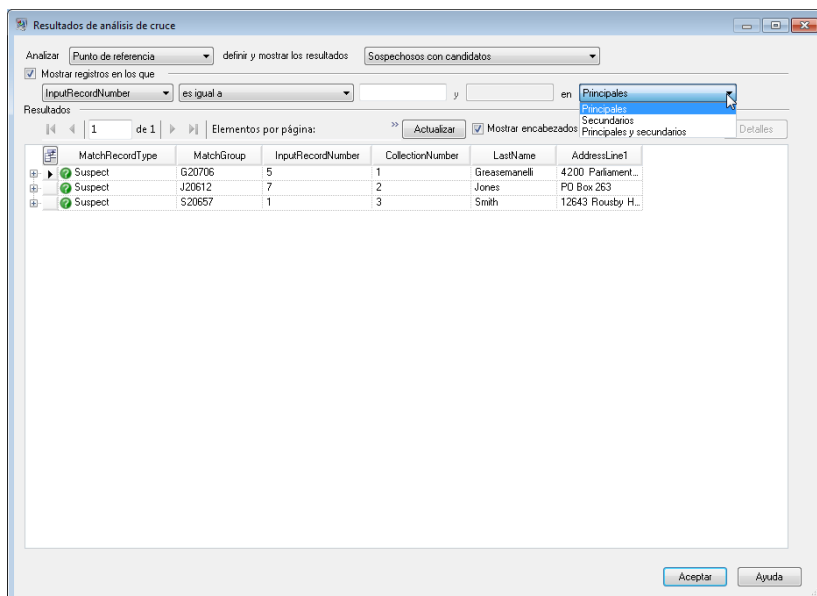
Los operadores que puede seleccionar son:

- Campos de cadena de caracteres (MatchGroup, MatchRecordType, todos los datos que pueden coincidir)
 - contiene
 - está entre
 - es igual a
 - no es igual a
 - comienza con
- Campos numéricos (CollectionNumber, InputRecordNumber, MatchScore)
 - está entre
 - es igual a
 - no es igual a
 - es mayor que
 - es mayor o igual que
 - es menor que
 - es menor o igual que

Para filtrar registros:

1. Seleccione un resultado de punto de referencia o comparación de la vista Resultados de análisis de cruce y haga clic en **Actualizar**.

2. Seleccione la casilla **Mostrar registros en los que**.



3. Seleccione un campo en la lista desplegable **Campo**.
4. Seleccione un operador.
5. Escriba un valor para el tipo de operador seleccionado. Si selecciona **está entre**, introduzca un rango de valores.
6. Al aplicar filtros en vistas de registros sospechosos, puede filtrar por:
 - Principales: el filtro se aplica solo a los elementos principales (sospechosos) y se muestran los elementos secundarios.
 - Secundarios: el filtro se aplica a los elementos secundarios que no estén incluidos en el rango de filtro. Se muestran los nodos principales (sospechosos).
 - Principales y secundarios: el filtro se aplica a los elementos principales (sospechosos) y, si los resultados muestran algún elemento principal, se filtran sus elementos secundarios
7. Haga clic en **Actualizar**. Se mostrarán los registros que estén incluidos en el rango de las opciones y los valores seleccionados. Si no hay registros que estén incluidos en ese rango, aparecerá un mensaje para indicar que no se obtuvieron registros.

Análisis de los cambios de las reglas de cruce

Puede utilizar la herramienta Análisis de cruce en Enterprise Designer para ver detalladamente el efecto que tiene un cambio en una regla de cruce en los resultados de cruce de los flujos de datos. Esto lo puede realizar al ejecutar el flujo de datos, implementar los cambios, volver a ejecutar el flujo de datos y luego visualizar los resultados en la herramienta Análisis de cruce. Este procedimiento describe cómo llevarlo a cabo.

Importante: Al comparar los resultados de cruce, los datos de entrada que se usan para las ejecuciones de comparación y punto de referencia deben ser idénticos. Utilizar diferentes datos de entrada puede causar resultados incorrectos. Siga las siguientes indicaciones para ayudar a garantizar una comparación exacta:

- Utilice los mismos archivos o tablas de entrada
- Ordene los datos de la misma forma antes de la etapa de comparación
- Utilice las mismas consultas de Candidate Finder cuando se use Transactional Match

1. En Enterprise Designer, abra el flujo de datos que desea analizar.
2. Para etapa de correspondencia de Interflow Match, Intraflow Match, o Transactional, cuyas correspondencias desee analizar, haga doble clic en la etapa y seleccione la casilla de verificación **Generar datos para análisis**.

Importante: Si se activa la opción **Generar datos para análisis** se reduce el rendimiento. Debe desactivar esta opción cuando haya terminado de utilizar la herramienta Análisis de cruce.

3. Seleccione **Ejecutar > Ejecutar flujo actual**

Nota: Para obtener resultados óptimos, utilice datos que producirán 100 000 o menos registros. Mientras más resultados de cruce haya, más lento será el rendimiento de la herramienta Análisis de cruce.

4. En las etapas de cruce de los flujos de datos, realice los cambios de regla de cruce que desee y luego ejecute nuevamente el flujo de datos.

Por ejemplo, si desea probar el efecto de aumentar el valor del umbral, cambie dicho valor y ejecute nuevamente el flujo de datos.

5. Cuando finalice la ejecución del flujo de datos, seleccione **Herramientas > Análisis de cruce**.

En el cuadro de diálogo **Examinar resultados de cruce** se muestra una lista de flujos de datos que tienen resultados de cruce que se ven en la herramienta Análisis de cruce. Si el trabajo que desea analizar no se encuentra en la lista, abra el flujo de datos y asegúrese de que la etapa tenga seleccionada la casilla de verificación **Generar datos para análisis**.

Consejo: Si hay un gran número de flujos de datos y si desea filtrarlos, seleccione una opción de filtro de la lista desplegable **Mostrar solo trabajos donde**.

6. En el lado izquierdo del panel Análisis de cruce, existe una lista de las etapas de cruce, una por ejecución. Seleccione la etapa de cruce en la ejecución que desee usar como punto de referencia y haga clic en **Punto de referencia**. Luego, seleccione la ejecución con la que desee comparar el punto de referencia y haga clic en **Comparar**.

Ahora puede comparar los resultados de cruce de resumen, como por ejemplo el número total de registros duplicados, además de información detallada de nivel de registro que muestra cómo se evaluó cada registro con respecto a las reglas de cruce.

Ejemplo de comparación de resultados de cruce

Por ejemplo, supongamos que ejecuta un trabajo denominado HouseholdRelationshipsAnalysis. Desea probar el efecto de un cambio en la etapa Cruce de hogares 2. Primero ejecuta el trabajo con la configuración original, luego modifica las reglas de cruce en la etapa Cruce de hogares 2 y ejecuta el trabajo nuevamente. En la herramienta Análisis de cruce, la ejecución con una ID de trabajo de 10 es la ejecución con la configuración original, de modo que debe establecerla como el punto de referencia. La ejecución con una ID de trabajo de 13 es la ejecución con la regla de cruce modificada. Al hacer clic en **Comparar**, puede ver que la regla de cruce modificada (ID de trabajo 113) generó más de un registro duplicado y un registro único menos que la regla de cruce original.

The screenshot shows the 'Análisis de cruce' tool interface. On the left, there is a list of jobs with 'HouseholdRelations...' selected. In the center, there are buttons for 'Agregar...', 'Quitar', 'Punto de referencia', 'Comparar', 'Detalles', and 'Ayuda'. On the right, a table displays the comparison results for 'Reglas de cruce'.

	Punto de referen	Comparación	Cambios
Registros de entrada	11	11	0
Registros duplicados	10	11	1
Registros únicos	1	0	-1
Grupos de cruce	3	3	0
Colecciones duplicadas	3	3	0
Cruces inmediatos	0	0	0
Calificación promedio	99.8	99.3	-0.5

Adición de resultados de cruce

Si ejecuta un trabajo cuando la herramienta de análisis de cruce está abierta y la Lista Resultados de cruce está vacía, los resultados de cruce se agregan automáticamente a la lista. Después de añadido un resultado de un cruce, la herramienta de Análisis de cruces sólo agrega los resultados coincidentes del mismo tipo (Interflow Match, Intraflow Match, o Transactional Match).

Si desea analizar resultados de cruce de tipo diferente a lo que actualmente se ha seleccionado en la herramienta de Análisis de cruce, siga los siguientes pasos.

1. Seleccione todos los resultados de la Lista Resultados de cruce y haga clic en **Eliminar**.
2. Desde el Explorador del servidor, abra un trabajo que utilice una etapa de comparación diferente o haga clic en la ficha que se encuentra sobre el lienzo si el trabajo ya está abierto.
3. Ejecute el trabajo.

Cuando finalice la ejecución del trabajo, los resultados de cruce de la última instancia de trabajo se agregan a la Lista Resultados de cruce.

Eliminación de resultados de cruce

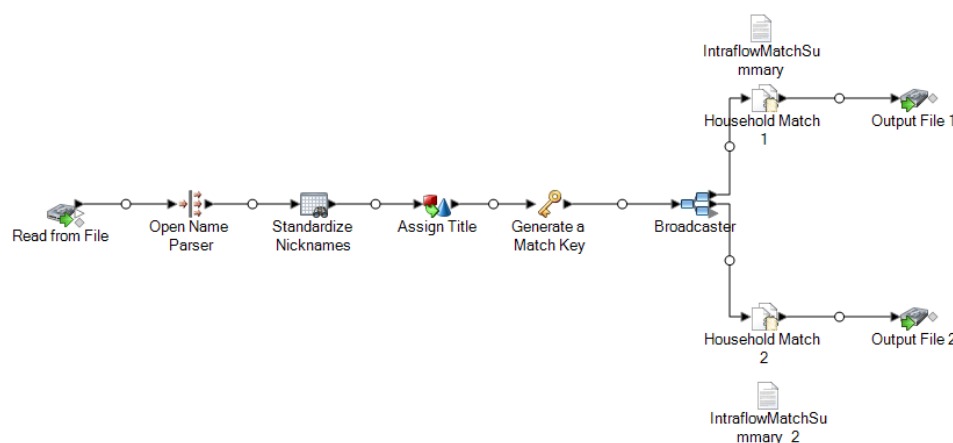
Para eliminar resultados de cruce desde la lista Resultados de cruce, seleccione un resultado de cruce en la lista Resultados de cruce y luego haga clic en **Eliminar**.

El sistema actualiza la lista de resultados de cruce y la ficha Resumen de la siguiente forma:

- Si los resultados de cruce eliminados no eran resultados de Punto de referencia ni de Comparación, los resultados de cruce se eliminan y no se registran cambios en la ficha Resumen.
- Si los resultados de cruce eliminados estaban configurados como Punto de referencia, el sistema configura el resultado de cruce que le sucede en antigüedad como el nuevo Punto de referencia y actualiza la ficha Resumen para mostrar solo los datos del nuevo Punto de referencia.
- Si los resultados de cruce eliminados estaban configurados como los resultados de cruce de una Comparación, el sistema actualiza la ficha Resumen para mostrar solo los datos del Punto de referencia existente.
- Si los resultados de cruce eliminados conforman uno de los dos que se muestran en la Lista Resultados de cruce, los resultados restantes son configurados como el nuevo Punto de referencia y el sistema actualiza la ficha Resumen para mostrar solo los datos del nuevo Punto de referencia.

Ejemplo: uso de análisis de cruce

Este ejemplo demuestra cómo usar la herramienta de Análisis de flujo para comparar la tasa de elevación y caída de dos cruces diferentes. Antes de que los datos se envíen mediante un cruce, se dividen en dos flujos utilizando Broadcaster. Luego, cada flujo se envía mediante una etapa Intraflow Match (Cruce de intraflujo). Cada flujo de datos incluye copias idénticas de los datos procesados. Cada etapa Intraflow Match utiliza diferentes algoritmos de comparación y genera datos de análisis de cruce que pueden usarse para comparar la elevación/caída de varios cruces.

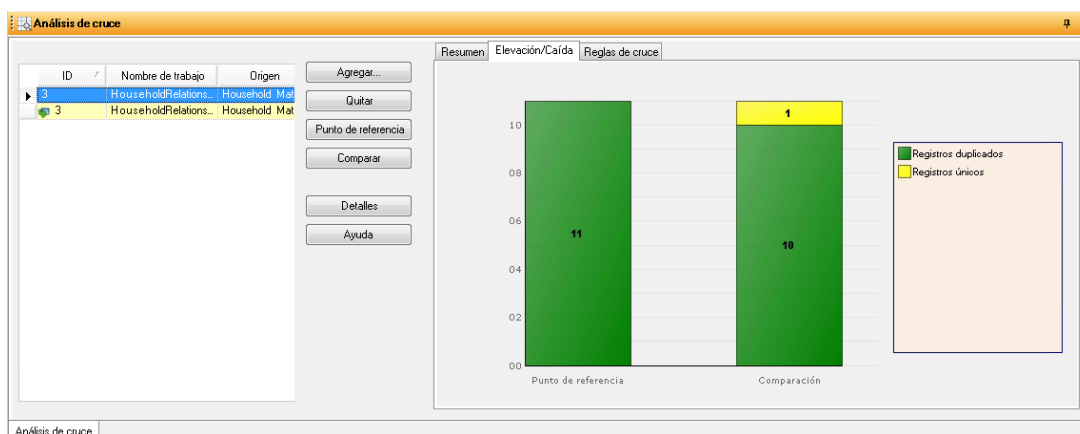


Este ejemplo de flujo de datos está disponible en Enterprise Designer. Vaya a **Archivo > Nuevo > Flujo de datos > Desde plantilla** y seleccione **HouseholdRelationshipsAnalysis**. Este flujo de

datos requiere de los siguientes módulos: Advanced Matching Module, Data Normalization Module, y Universal Name Module. Es preciso que además cargue la base de datos principal de Table Lookup y las tablas básicas de Open Parser.

Para usarlo, vea el siguiente ejemplo:

1. Ejecute el flujo de datos.
2. Seleccione **Herramientas > Análisis de cruce**.
3. Desde la ventana **Examinar resultados de cruce**, expanda **HouseholdRelationshipAnalysis**, seleccione **Cruce de hogares 1** y **Cruce de hogares 2** de la lista Origen y, luego, haga clic en **Agregar**.
4. Seleccione **Cruce de hogares 1** en la Lista de resultados de cruce y haga clic en **Comparar**. Se visualizarán los resultados resumidos.
5. Haga clic en la ficha **Elevación/Caída**. Se muestra el gráfico Elevación/Caída.



Este gráfico muestra las diferencias entre los registros duplicados y los registros únicos generados para las diferentes reglas de cruce utilizadas.

6. Haga clic en la ficha **Reglas de cruce**. Se visualizará la comparación de las reglas de cruce.

Reglas de cruce

Punto de referencia:

- Opciones
 - Agrupar por MatchKey
 - Cruce inmediato desactivado
 - Ventana dinámica desactivada
 - Opción de orden activada
 - Número de colección 0 de registros únicos activados
- Reglas
 - Household
 - LastName
 - y Address
 - AddressLine1

Comparación:

- Opciones
 - Agrupar por MatchKey
 - Cruce inmediato desactivado
 - Ventana dinámica desactivada
 - Opción de orden activada
 - Número de colección 0 de registros únicos activados
- Reglas
 - Household
 - LastName (Modificada)
 - y Address
 - AddressLine1

Método de comparación: Según el umbral

Método de calificación: Máximo

Datos ausentes: Ignorar espacios en blanco

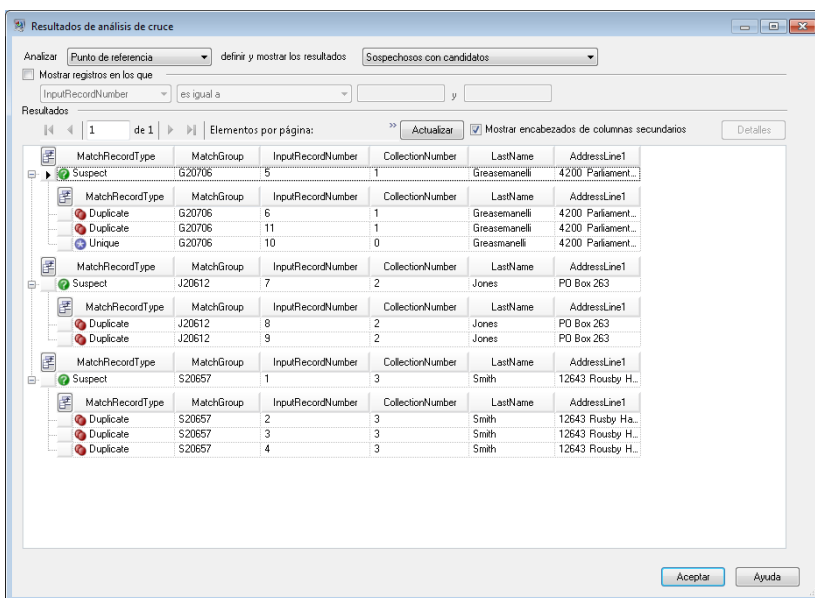
Umbral: 80

Algoritmos

- Frecuencia de caracteres
- Cruce exacto (Nueva)
- Frecuencia de caracteres (Omitida)

Desde esta ficha, es posible ver que el algoritmo ha cambiado: se omitió la frecuencia de caracteres y se agregó el cruce exacto.

7. Haga clic en **Detalles**.
8. Seleccione **Colecciones duplicadas** desde la lista Mostrar y luego haga clic en **Actualizar**.
9. Expanda cada **CollectionNumber** para ver los registros sospechosos y los duplicados de cada colección duplicada.



10. Compare las colecciones en la vista de detalles del archivo de salida creado.

Plantillas de flujo de datos para comparación

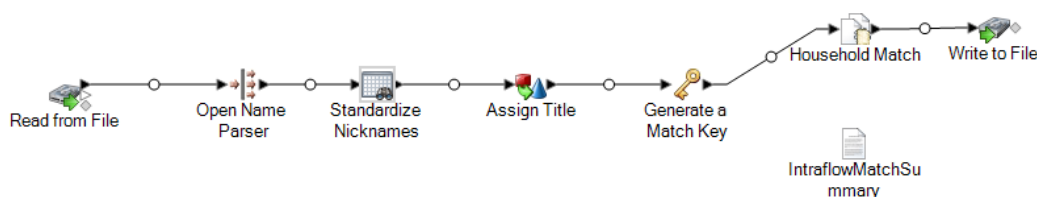
Identificación de miembros de un hogar

Esta plantilla de flujo de datos demuestra la manera de identificar a los miembros de un mismo hogar, por medio de la comparación de información contenida en un único archivo de entrada y la generación de un archivo de salida que contenga colecciones de hogar.

Situación empresarial posible

En su calidad de administrador de datos de una empresa de tarjetas de crédito, usted desea analizar su base de datos de clientes y averiguar cuáles direcciones figuran varias veces y con qué nombres, a fin de poder reducir a un mínimo la cantidad de correspondencia y de ofertas repetidas de tarjetas de crédito que se envían a una misma dirección.

El siguiente flujo de datos ofrece una solución ante una posible situación empresarial:



Esta plantilla de flujo de datos está disponible en Enterprise Designer. Vaya a **Archivo > Nuevo > Flujo de datos > A partir de plantilla** y seleccione **Vínculos de hogar**. Este flujo de datos requiere de los siguientes módulos: Advanced Matching, Data Normalization y Universal Name.

Para cada registro del archivo de entrada, este flujo de datos hará lo siguiente:

Read from File

La etapa Read from File (Lectura desde archivo) identifica el nombre, la ubicación y el diseño del archivo que contiene los nombres que desea analizar. El archivo contiene tanto nombres masculinos como femeninos.

Open Name Parser

Open Name Parser examina los campos de nombre y los compara con los datos almacenados en los archivos de bases de datos de nombres de Spectrum™ Technology Platform. Según la comparación, divide los datos de nombre en los campos de nombre, segundo nombre y apellido, y asigna un tipo de entidad y un género a cada nombre. También utiliza el reconocimiento de patrones además de los datos de nombre.

Estandarizar los apodos

En esta plantilla, la etapa Table Lookup (búsqueda de tabla) se denomina Standardize Nicknames (estandarización de apodos). La etapa Standardize Nickname busca nombres en la base de datos Nicknames.xml y reemplaza cualquier apodo con la forma más común del apodo. Por ejemplo, el nombre Tommy es reemplazado por Thomas.

Transformer

En esta plantilla, la etapa Transformer (Transformador) se denomina Assign Titles (Asignación de tratamientos). La etapa Assign Titles utiliza una secuencia de comandos personalizada para examinar cada fila en la transmisión de datos enviados por la etapa Parse Personal Name y les asigna un valor **TitleOfRespect** según el valor **GenderCode**.

La secuencia de comandos personalizada:

```

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


Cada vez que la etapa Assign Titles encuentra **M** en el campo **GenderCode**, define el valor para **TitleOfRespect** como **M_r**. Cada vez que las etapas Assign Titles encuentran **F** en el campo **GenderCode**, define el valor para **TitleOfRespect** como **M_s**.

Match Key Generator

Match Key Generator (Generador de claves de cruce) procesa reglas definidas por el usuario que consisten en algoritmos y campos de origen de entrada para generar el campo de clave de cruce. Una clave de cruce es una clave no exclusiva compartida por registros similares que identifican registros como posibles duplicados. La clave de cruce se utiliza para facilitar el proceso de comparación porque limita la comparación únicamente a los registros que contienen la misma clave de cruce. La clave de cruce está compuesta por campos de entrada. Cada campo de entrada especificado cuenta con un algoritmo seleccionado que se ejecuta para el campo. A continuación, el resultado de cada campo es concatenado para crear un único campo de clave de cruce.

En esta plantilla se definen dos campos de clave de cruce: SubString (apellido (1:3) y SubString (código postal (1:5)).

Por ejemplo, si la dirección entrante fuera:

FirstName: Fred

LastName: Mertz

PostalCode: 21114-1687

Y las reglas especificaran lo siguiente:

Campo de entrada	Posición de inicio	Longitud
LastName	1	3
PostalCode	1	5

Entonces la clave, basada en las reglas y los datos de entrada que se muestran arriba, sería:

Mer21114

Coincidencia de hogar

En esta plantilla de flujo de datos la etapa de cruce de intraflujo (Intraflow Match) se denomina Household Match (coincidencia de hogar). Esta etapa ubica los cruces (coincidencias) entre registros de datos similares, de un mismo flujo de entrada. Los registros cruzados también pueden calificarse mediante los datos sin nombre ni dirección. El motor de comparaciones permite crear reglas jerárquicas basadas en cualquier campo que se haya definido o creado en otras etapas.

Un flujo de registros por ser cruzados así como parámetros que especifican qué campos se deben comparar, cómo se deben computar las calificaciones y, por lo general, qué constituye un cruce exitoso.

En esta plantilla, se crea una regla de cruce personalizada que compara LastName y AddressLine1. Seleccione la casilla **Generar datos para análisis** a fin de generar datos para el informe resumido de intraflujo.

Estas son algunas pautas por seguir durante la creación de una jerarquía de comparación:

- Los nodos principales deben tener un nombre único. No pueden ser campos.
- El campo secundario debe ser un campo de tipo de dato de Spectrum™ Technology Platform, es decir, un tipo de dato disponible a través de uno o más componentes.
- Todos los nodos secundarios de un nodo principal deben usar los mismos operadores lógicos. Para combinar conectores, primero deben crearse nodos principales intermedios.
- Los umbrales del nodo principal pueden ser más altos que los umbrales de los nodos secundarios.
- No es necesario que los nodos principales tengan umbral.

Write to File

La plantilla contiene una etapa Write to File (Escritura en archivo) que crea un archivo de texto que muestra las direcciones como una colección de hogares.

Informe resumido de intraflujo

La plantilla contiene el informe resumido Intraflow Match (Cruce de intraflujo). Después de ejecutar el trabajo, expanda **Informas** en la ventana Detalles de ejecución y después haga clic en **Informe resumido de intraflujo**.

El informe resumido Intraflow Match muestra las estadísticas de los registros procesados y un gráfico de barras que ilustra el recuento de registros y la calificación general de cruce.

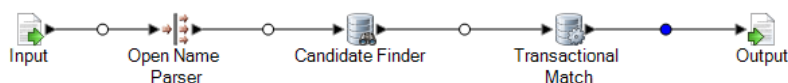
DetECCIÓN DE POSIBLES CLIENTES QUE YA SON CLIENTES

Esta plantilla de flujo de datos muestra cómo comparar los datos de posibles clientes en un archivo de entrada con los datos de clientes en una base de datos de clientes para determinar si un posible cliente ya es cliente. Se trata de un flujo de datos de servicios, lo que quiere decir que se puede tener acceso al flujo de datos por medio de API o de servicios web.

Situación empresarial posible

Como ejecutivo de cuenta de una empresa de ventas por Internet, quiere determinar si un posible cliente ya ha comprado o si es un cliente nuevo.

El siguiente servicio de flujo de datos ofrece una solución a la posible situación empresarial:



Esta plantilla de flujo de datos está disponible en Enterprise Designer. Vaya a **Archivo > Nuevo > Flujo de datos > Desde plantilla** y seleccione **servicios**. Este flujo de datos requiere de los módulos Advanced Matching Module y Universal Name Module.

Para cada registro en el archivo de entrada, este flujo de datos realiza lo siguiente:

Input

Los campos de entrada seleccionados para esta plantilla son AddressLine1, City, Name, PostalCode y StateProvince. AddressLine1 y Name son los campos clave para el procesamiento del flujo de datos de esta plantilla.

Name Parser

En esta plantilla, la etapa Name Parser (Analizador de nombre) se denomina Parse Personal Name (Análisis de nombres personales). La etapa Parse Personal Name examina los campos de nombre y los compara con los datos almacenados en los archivos de bases de datos de nombres de Spectrum™ Technology Platform. Según la comparación, divide los datos de nombre en los campos de nombre, segundo nombre y apellido, y asigna un tipo de entidad y un género a cada nombre. También utiliza el reconocimiento de patrones además de los datos de nombre.

En esta plantilla, la etapa Parse Personal Name se configura de la siguiente manera.

- Se selecciona la opción Analizar nombres personales y se desmarca la opción Analizar nombres de empresas. Cuando seleccionan estas opciones, los nombres de pila se evalúan en cuanto a género, orden, puntuación y no se realiza una evaluación de nombres de empresas.
- El Origen de determinación de género está configurado de manera predeterminada. En la mayoría de los casos, la configuración predeterminada es la mejor opción para determinar el género porque cubre una amplia variedad de nombres. De todas maneras, si está procesando nombres de una cultura específica, seleccione esa cultura. La selección de una cultura específica contribuye a garantizar la asignación del género correcto a los nombres. Por ejemplo, si se deja la opción predeterminada seleccionada, el nombre Jean será identificado como un nombre femenino. Sin embargo, si selecciona francés, será identificado como un nombre masculino.
- El orden configurado es el natural. Los campos de nombre están ordenados por tratamiento, nombre, segundo nombre, apellido y sufijo.
- La opción Retener puntos está desmarcada. No se mantiene ninguna puntuación en los datos de nombres.

Candidate Finder

La etapa Candidate Finder (Localizador de candidatos) se utiliza en combinación con la etapa Transactional Match (Cruce transaccional).

La etapa Candidate Finder obtiene los registros candidatos que conformarán los posibles cruces que serán evaluados por la etapa Transactional Match. Además, según el formato de sus datos,

es posible que Candidate Finder también necesite analizar el nombre o la dirección del registro sospechoso, los registros candidatos, o ambos.

Durante la configuración de Candidate Finder, se selecciona la conexión de base de datos mediante la cual se ejecutará la consulta especificada. Es posible seleccionar cualquier conexión configurada en Management Console. Para establecer una conexión con una base de datos que no está en la lista, configure una conexión a esa base de datos en Management Console y, luego, cierre y vuelva a abrir Candidate Finder para actualizar la lista de conexiones.

Para definir la consulta SQL, puede ingresar cualquier instrucción SQL Select válida en el cuadro de texto de la vista Opciones de Candidate Finder. Por ejemplo, suponga que, en su base de datos, tiene una tabla denominada Customer_Table que tiene las siguientes columnas:

Customer_Table

Cust_Name

Cust_Address

Cust_City

Cust_State

Cust_Zip

Nota: Puede ingresar cualquier instrucción SQL Select válida, pero la opción `Select *` no es válida en este control.

Para recuperar todas las filas de la base de datos, puede construir una consulta similar a la siguiente:

```
select Cust_Name, Cust_Address, Cust_City, Cust_State, Cust_Zip from Customer_Table;
```

Sin embargo, es poco probable que quiera cruzar su transacción con todas las filas de la base de datos. Para obtener solo los registros candidatos relevantes, es recomendable agregar una cláusula WHERE utilizando la sustitución de variable. La sustitución de variable es una notación especial que se utiliza para que el motor de selección de candidatos reemplace la variable con los datos reales de su registro sospechoso.

Para utilizar la sustitución de variable, encierre el nombre del campo entre llaves y coloque adelante un signo de dólar **`\${FieldName}**. Por ejemplo, la siguiente consulta arrojará solo aquellos registros con un valor en Cust_Zip que coincida con el valor en PostalCode del registro sospechoso.

```
select Cust_Name, Cust_Address, Cust_City, Cust_State,Cust_Zip
from Customer_Table
where Cust_Zip = ${PostalCode};
```

Luego, debe asignar las columnas de la base de datos a los campos de la etapa si los nombres de las columnas de su base de datos no coinciden de manera exacta con los nombres de los campos del componente. Si los nombres coinciden, se asignarán automáticamente a los campos correspondientes de la etapa. Es necesario asignar los campos seleccionados (columnas de la base de datos) a los campos de la etapa (nombres de campos definidos en el flujo de datos).

Considere nuevamente la tabla Customer_Table del ejemplo anterior:

Customer_Table

Cust_Name

Cust_Address

Cust_City

Cust_State

Cust_Zip

Cuando se recuperan estos registros de la base de datos, es necesario asignar los nombres de las columnas a los nombres de los campos que serán utilizados por Transactional Match y otras etapas de su flujo de datos. Por ejemplo, Cust_Address podría asignarse a AddressLine1 y Cust_Zip se asignaría a PostalCode.

1. Seleccione la lista desplegable de **Campos seleccionados** en la vista Opciones de Candidate Finder. Luego, seleccione la columna Cust_Zip de la base de datos.
2. Seleccione la lista desplegable en **Campos de la etapa**. Luego, seleccione el campo al que desea asignarle la columna.

Por ejemplo, si desea asignar Cust_Zip a PostalCode, primero seleccione Cust_Zip en Campos seleccionados y luego seleccione PostalCode en la fila correspondiente del Campo de la etapa.

Otra alternativa para asignar campos (además de la que se describió anteriormente) es utilizar una notación especial en su consulta SQL. Para realizar la asignación, ingrese el nombre del campo de la etapa, encerrado entre llaves, después del nombre de la columna en su consulta. Al hacerlo, los campos seleccionados se asignarán automáticamente a los campos correspondientes de la etapa.

A continuación, se muestra un ejemplo utilizando la consulta mencionada anteriormente:

```
select Cust_Name {Name}, Cust_Address {AddressLine1},
       Cust_City {City}, Cust_State {StateProvince},
       Cust_Zip {PostalCode}
from Customer
where Cust_Zip = ${PostalCode};
```

Transactional Match

La etapa Transactional Match se utiliza en combinación con la etapa Candidate Finder.

La etapa Transactional Match le permite comparar los registros sospechosos con posibles registros candidatos que se obtienen en la etapa Candidate Finder.

Transactional Match utiliza reglas de cruce para comparar el registro sospechoso con todos los registros candidatos que tienen el mismo número de grupo de candidatos (asignado en Candidate Finder) a fin de identificar duplicados. Si el registro candidato es un duplicado, se le asigna un número de colección, el tipo de registro de cruce se etiqueta como duplicado y luego se copia. A los candidatos sin coincidencias en el grupo se les asigna el número de colección 0, se etiquetan como únicos y luego también se copian.

En esta plantilla, se crea una regla de cruce personalizada que compara LastName y AddressLine1.

Estas son algunas pautas por seguir durante la creación de una jerarquía de comparación:

- Los nodos principales deben tener un nombre único. No pueden ser campos.
- El campo secundario debe ser un campo de tipo de dato de Spectrum™ Technology Platform, es decir, un tipo de dato disponible a través de una o más etapas.
- Todos los nodos secundarios de un nodo principal deben usar los mismos operadores lógicos. Para combinar conectores, primero deben crearse nodos principales intermedios.
- Los umbrales del nodo principal pueden ser más altos que los umbrales de los nodos secundarios.
- No es necesario que los nodos principales tengan umbral.

Salida

Como servicio, esta plantilla envía todos los campos disponibles a la salida. Es posible limitar la salida según sus necesidades.

5 - Desduplicación

In this section

Filtrado de registros duplicados	144
Creación de un registro Best of Breed.	148

Filtrado de registros duplicados

La manera más sencilla de eliminar registros duplicados es agregar a su flujo de datos una etapa Filter luego de una etapa de cruce. La etapa Filter elimina registros de colecciones de registros duplicados basándose en los ajustes que haya especificado.

1. En Enterprise Designer, cree un flujo de datos que identifique los registros duplicados mediante comparación.

El cruce es el primer paso en la desduplicación, ya que se necesita identificar registros que son similares, como por ejemplo registros que tienen el mismo número de cuenta o nombre. Consulte las siguientes secciones para obtener instrucciones acerca de la creación de flujo de datos que realicen cruce de registros.

Cruce de registros de un origen único en la página 85

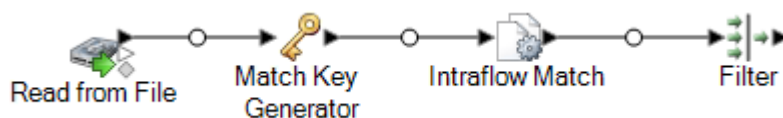
Comparación de registros de un origen a otro origen en la página 91

Comparación de registros con una base de datos en la página 104

Nota: Solo necesita generar el flujo de datos en el punto donde lee datos y realiza el cruce con una etapa Interflow Match, Intraflow Match o Transactional Match. Cuando haya creado el flujo de datos hasta este punto, siga adelante con los pasos siguientes.

2. Cuando haya definido un flujo de datos que lea datos y cruce registros, arrastre hacia el lienzo una etapa Filter y conéctela a la etapa que realiza el cruce (Interflow Match, Intraflow Match o Transactional Match).

Por ejemplo, si su flujo de datos lee datos de un archivo y realiza cruces con Intraflow Match, tendría el siguiente aspecto luego de añadir una etapa Filter:



3. Haga doble clic en el lienzo en la etapa Filter.
4. En el campo **Agrupar por**, seleccione **CollectionNumber**.
5. Deje seleccionada la opción **Limitar la cantidad de registros duplicados obtenidos** y defina el valor en 1. Esta es la configuración predeterminada.
6. Decida si desea conservar el primer registro de cada colección, o si desea definir una regla para decidir cuál registro se mantendrá de cada una de las colecciones. Si desea conservar el primer registro de cada colección, omita este paso. Si desea definir una regla, en el árbol de regla, seleccione **Reglas** y luego siga los pasos siguientes:
 - a) Haga clic en **Agregar regla**.

Los registros de cada grupo son evaluados para ver si cumplen con las reglas que defina aquí. Si un registro cumple una regla, será este el registro sobreviviente y los demás registros del grupo serán descartados.

- b) Defina una regla para identificar el registro que se conservará de cada grupo.

Use las siguientes opciones para definir una regla:

Opción	Descripción
Nombre de campo	Especifica el nombre del campo de flujo de datos cuyo valor desea evaluar, para determinar si se va a filtrar el registro.
Tipo de campo	Especifica el tipo de datos del campo. Una de las siguientes: No numérico Seleccione esta opción si el campo contiene datos no numéricos (por ejemplo, datos de cadena). Numérico Seleccione esta opción si el campo contiene datos numéricos (por ejemplo, doble, flotante, entre otros).

Opción	Descripción
Operador	Especifica el tipo de comparación que desea utilizar en la evaluación del campo. Una de las siguientes:
Contiene	Determina si el campo contiene el valor especificado. Por ejemplo, "sailboat" contiene el valor "boat".
IGUAL	Determina si el campo contiene el valor exacto especificado.
Mayor que	Determina si el valor de campo es mayor que el valor especificado. Esta operación solamente funciona con campos numéricos.
Mayor o igual que	Determina si el valor de campo es mayor o igual que el valor especificado. Esta operación solamente funciona con campos numéricos.
El más alto	Compara el valor del campo en todo el grupo de registros y determina qué registro tiene el valor más alto en el campo. Por ejemplo, si los campos del grupo contienen valores de 10, 20, 30 y 100, quedaría seleccionado el registro con el valor 100. Esta operación solamente funciona con campos numéricos. Si hay múltiples registros que poseen el valor más largo, se selecciona uno de ellos.
Está vacío	Determina si el campo no contiene ningún valor.
No está vacío	Determina si el campo contiene algún valor.
Menor que	Determina si el valor de campo es menor que el valor especificado. Esta operación solamente funciona con campos numéricos.
Menor o igual que	Determina si el valor de campo es menor o igual que el valor especificado. Esta operación solamente funciona con campos numéricos.
El más largo	Compara el valor del campo en todo el grupo de registros y determina qué registro tiene el valor más largo (en bytes) en el campo. Por ejemplo, si el grupo contiene los valores "Mike" y "Michael", quedaría seleccionado el registro con el valor "Michael". Si hay múltiples registros que poseen el valor más largo, se selecciona uno de ellos.
El más bajo	Compara el valor del campo en todo el grupo de registros y determina qué registro tiene el valor más bajo en el campo. Por ejemplo, si los campos del grupo contienen valores de 10, 20, 30 y 100, quedaría seleccionado el registro con el valor 10. Esta operación solamente funciona con campos numéricos. Si hay múltiples registros que poseen el valor más largo, se selecciona uno de ellos.
Más común	Determina si el valor de campo contiene el valor que aparece más frecuentemente en este campo entre los registros del grupo. Si hay dos o más valores que son los más comunes, no se realiza acción alguna.
Diferente	Determina si el valor de campo no es el mismo que el valor especificado.

Opción	Descripción
Tipo de valor	<p>Especifica el tipo de valor que desea comparar al valor del campo. Una de las siguientes:</p> <p>Nota: Esta opción no está disponible si selecciona el operador El más alto, El más bajo, o El más largo.</p> <p>Campo Seleccione esta opción si desea comparar con este campo el valor del campo de otro flujo de datos.</p> <p>Cadena Seleccione esta opción si desea comparar el campo con un valor específico.</p>
Valor	<p>Especifica el valor que se va a comparar con el valor del campo. Si seleccionó Campo en el campo Tipo de campo seleccione un campo de flujo de datos. Si seleccionó Cadena en el campo de Tipo de valor, ingrese el valor que desea usar para la comparación.</p> <p>Nota: Esta opción no está disponible si selecciona el operador El más alto, El más bajo, o El más largo.</p>

c) Haga clic en **Aceptar**.

Ahora ha configurado Filter con una regla. Puede añadir reglas adicionales si lo requiere.

- Haga clic en **Aceptar** para cerrar la ventana de **Opciones de Filter**.
- Arrastre hacia el lienzo una etapa receptora y conéctela a la etapa Filter.

Por ejemplo, si estuviese utilizando la etapa receptora Write to File, su flujo de datos tendría el siguiente aspecto:



- Haga doble clic en la etapa receptora y configúrela.

Para obtener información acerca de cómo configurar etapas receptoras, consulte la *Guía de Dataflow Designer*.

Ahora tiene un flujo de datos que identifica registros coincidentes y elimina todos los registros de cada grupo de duplicados, salvo uno, lo cual da como resultado un archivo de salida que contiene datos desduplicados.

Creación de un registro Best of Breed.

Para eliminar de sus datos registros duplicados, puede decidir combinar dichos datos a partir de grupos de registros duplicados en un único registro "best of breed". Esta metodología es conveniente cuando cada registro duplicado contiene datos del mismo tipo (por ejemplo, números telefónicos o nombres) y desea conservar los mejores datos de cada registro en el registro sobreviviente.

Este procedimiento describe la forma de crear un flujo de datos que combine registros duplicados en un registro best of breed.

1. En Enterprise Designer, cree un flujo de datos que identifique los registros duplicados mediante comparación.

El cruce es el primer paso en la desduplicación, ya que se necesita identificar registros que son similares, como por ejemplo registros que tienen el mismo número de cuenta o nombre. Consulte las siguientes secciones para obtener instrucciones acerca de la creación de flujo de datos que realicen cruce de registros.

[Cruce de registros de un origen único](#) en la página 85

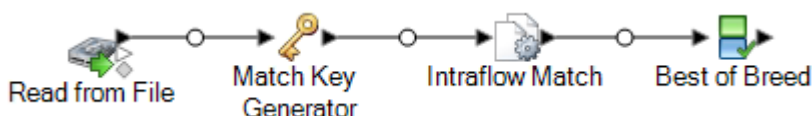
[Comparación de registros de un origen a otro origen](#) en la página 91

[Comparación de registros con una base de datos](#) en la página 104

Nota: Solo necesita generar el flujo de datos en el punto donde lee datos y realiza el cruce con una etapa Interflow Match, Intraflow Match o Transactional Match. Cuando haya creado el flujo de datos hasta este punto, siga adelante con los pasos siguientes.

2. Cuando haya definido un flujo de datos que lee datos y cruza registros, arrastre hacia el lienzo una etapa Best of Breed y conéctela a la etapa que realiza el cruce (Interflow Match, Intraflow Match o Transactional Match).

Por ejemplo, si su flujo de datos lee datos de un archivo y realiza un cruce con Intraflow Match, su flujo de datos tendrá el siguiente aspecto luego de añadir una etapa Best of Breed:



3. Haga doble clic en el lienzo en la etapa Best of Breed.
4. En el campo **Agrupar por**, seleccione **CollectionNumber**.
5. En **Configuración de Best of Breed**, seleccione **Reglas** en el árbol de condiciones.
6. Haga clic en **Agregar regla**.

Los registros de cada grupo son evaluados para ver si cumplen con las reglas que defina aquí. Si un registro coincide con una regla, sus datos pueden copiarse en el registro best of breed,

dependiendo de cómo se hayan configurado las acciones asociadas a la regla. Definirá las acciones más adelante.

- Defina una regla que un registro duplicado deba cumplir a fin de que sus datos sean copiados en el registro best of breed.

Use las siguientes opciones para definir una regla:

Opción	Descripción
Nombre de campo	Especifica el nombre del campo de flujo de datos cuyo valor usted desea determinar en caso de que se cumpla la condición y se deban realizar las siguientes acciones.
Tipo de campo	Especifica el tipo de datos del campo. Una de las siguientes: <ul style="list-style-type: none"> No numérico Seleccione esta opción si el campo contiene datos no numéricos (por ejemplo, datos de cadena). Numérico Seleccione esta opción si el campo contiene datos numéricos (por ejemplo, doble, flotante, entre otros).

Opción	Descripción
Operador	Especifica el tipo de comparación que desea utilizar en la evaluación del campo. Una de las siguientes:
Contiene	Determina si el campo contiene el valor especificado. Por ejemplo, "sailboat" contiene el valor "boat".
IGUAL	Determina si el campo contiene el valor exacto especificado.
Mayor que	Determina si el valor de campo es mayor que el valor especificado. Esta operación solamente funciona con campos numéricos.
Mayor o igual que	Determina si el valor de campo es mayor o igual que el valor especificado. Esta operación solamente funciona con campos numéricos.
El más alto	Compara el valor del campo en todo el grupo de registros y determina qué registro tiene el valor más alto en el campo. Por ejemplo, si los campos del grupo contienen valores de 10, 20, 30 y 100, quedaría seleccionado el registro con el valor 100. Esta operación solamente funciona con campos numéricos. Si hay múltiples registros que poseen el valor más largo, se selecciona uno de ellos.
Está vacío	Determina si el campo no contiene ningún valor.
No está vacío	Determina si el campo contiene algún valor.
Menor que	Determina si el valor de campo es menor que el valor especificado. Esta operación solamente funciona con campos numéricos.
Menor o igual que	Determina si el valor de campo es menor o igual que el valor especificado. Esta operación solamente funciona con campos numéricos.
El más largo	Compara el valor del campo en todo el grupo de registros y determina qué registro tiene el valor más largo (en bytes) en el campo. Por ejemplo, si el grupo contiene los valores "Mike" y "Michael", quedaría seleccionado el registro con el valor "Michael". Si hay múltiples registros que poseen el valor más largo, se selecciona uno de ellos.
El más bajo	Compara el valor del campo en todo el grupo de registros y determina qué registro tiene el valor más bajo en el campo. Por ejemplo, si los campos del grupo contienen valores de 10, 20, 30 y 100, quedaría seleccionado el registro con el valor 10. Esta operación solamente funciona con campos numéricos. Si hay múltiples registros que poseen el valor más largo, se selecciona uno de ellos.
Más común	Determina si el valor de campo contiene el valor que aparece más frecuentemente en este campo entre los registros del grupo. Si hay dos o más valores que son los más comunes, no se realiza acción alguna.
Diferente	Determina si el valor de campo no es el mismo que el valor especificado.

Opción	Descripción
Tipo de valor	<p>Especifica el tipo de valor que desea comparar al valor del campo. Una de las siguientes:</p> <p>Nota: Esta opción no está disponible si selecciona el operador El más alto, El más bajo, o El más largo.</p> <p>Campo Seleccione esta opción si desea comparar con este campo el valor del campo de otro flujo de datos.</p> <p>Cadena Seleccione esta opción si desea comparar el campo con un valor específico.</p>
Valor	<p>Especifica el valor que se va a comparar con el valor del campo. Si seleccionó Campo en el campo Tipo de campo seleccione un campo de flujo de datos. Si seleccionó Cadena en el campo de Tipo de valor, ingrese el valor que desea usar para la comparación.</p> <p>Nota: Esta opción no está disponible si selecciona el operador El más alto, El más bajo, o El más largo.</p>

8. Haga clic en **Aceptar**.
9. Haga clic en el nodo **Acciones** en el árbol.
10. Haga clic en **Agregar acción**.
11. Especifique los datos que se copiarán en el registro best of breed si es que dicho registro cumple con los criterios definidos en la regla.

Opción	Descripción
Tipo de origen	<p>Especifica el tipo de datos que se copiará en el registro Best of breed. Las opciones son:</p> <p>Campo Seleccione esta opción si desea copiar un valor de un campo en el registro Best of breed.</p> <p>String Seleccione esta opción si desea copiar un valor constante en el registro Best of breed.</p>
Datos de origen	<p>Especifica los datos que se copiarán en el registro Best of breed. Si el tipo de origen es Campo, seleccione el campo cuyos valores desea copiar en el campo de destino. Si el tipo de origen es Cadena, especifique un valor constante para copiar en el campo de destino.</p>

Opción	Descripción
Destino	Especifica el destino en el registro Best of breed al cual desea copiar los datos especificados en el campo Datos de origen .
Acumular datos de origen	<p>Si los datos en el campo Datos de origen son datos numéricos, puede habilitar esta opción para combinar los datos de origen en todos los registros duplicados y colocar el valor total en el registro Best of breed.</p> <p>Por ejemplo, si hubiese tres registros duplicados en el grupo, y estos contuviesen estos valores en el campo de Depósitos:</p> <p>100.00 20.00 5.00</p> <p>Estos tres valores serían combinados y el valor total, de 125,00, se colocaría en el campo Depósitos del registro Best of breed.</p>

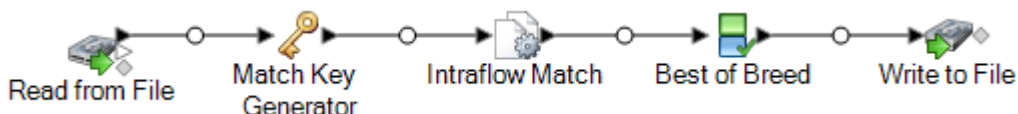
12. Haga clic en **Aceptar**.

Ahora ha configurado Best of Breed con una regla y una acción. Puede añadir reglas y acciones adicionales en caso necesario.

13. Haga clic en **Aceptar** para cerrar la ventana de **Opciones de Best of Breed**.

14. Arrastre hacia el lienzo una etapa receptora y conéctela a la etapa Best of Breed.

Por ejemplo, si estuviese utilizando la etapa receptora Write to File, su flujo de datos tendría el siguiente aspecto:



15. Haga doble clic en la etapa receptora y configúrela.

Para obtener información acerca de cómo configurar etapas receptoras, consulte la *Guía de Dataflow Designer*.

Ahora tiene un flujo de datos que identifica registros coincidentes y combina los registros dentro de una colección en un único registro best of breed.

6 - Registros de excepción

In this section

Diseño de un flujo de datos para procesar excepciones	154
Diseño de un flujo de datos para revalidación en tiempo real	156

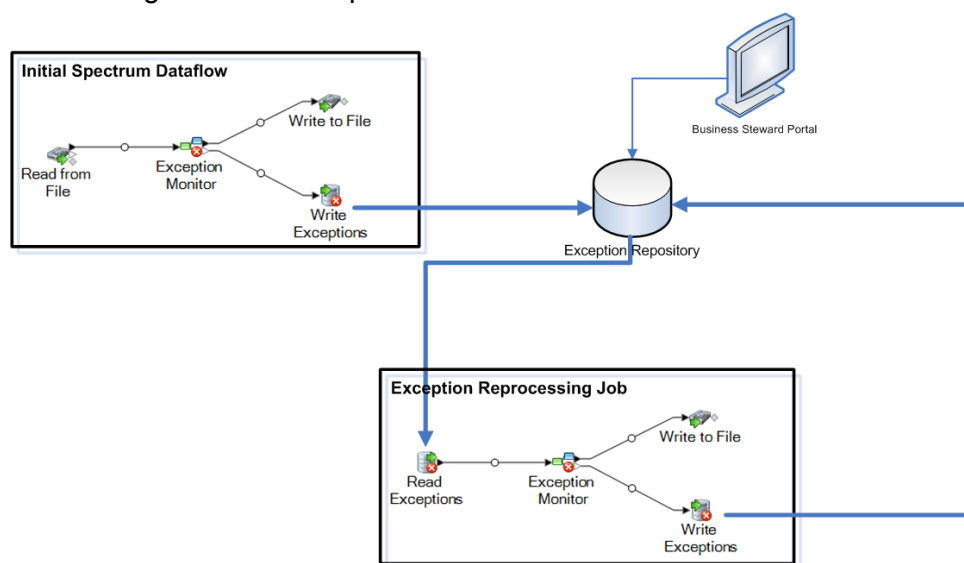
Diseño de un flujo de datos para procesar excepciones

Si posee la licencia para el módulo Business Steward, puede incluir el proceso de gestión de excepciones en sus flujos de datos. Los factores básicos del proceso de gestión de una excepción son los siguientes:

- Un flujo de datos inicial que lleva a cabo un proceso de calidad de datos, por ejemplo, deduplicación de registros, validación de direcciones o geocodificación.
- Una etapa Exception Monitor que identifica registros que no se pudieron procesar.
- La etapa Write Exceptions toma los registros de excepción identificados por la etapa Exception Monitor y los escribe en el repositorio de excepciones para su revisión manual.
- Business Steward Portal, una herramienta basada en navegador, permite revisar y editar los registros de excepciones. Cuando está editado, los registros se marcan como "Aprobados", lo que permite que estén disponibles para volver a ser procesados.
- Un trabajo de reprocesamiento de excepciones que usa la etapa Read Exceptions para leer registros aprobados desde el repositorio de excepciones en trabajo. Entonces el trabajo intenta volver a procesar los registros corregidos, normalmente mediante la misma lógica que el flujo de datos original. La etapa Exception Monitor una vez más verifica la presencia de excepciones. La etapa Write Exceptions envía excepciones de vuelta al repositorio de excepciones para una revisión adicional.

Nota: Do not place other stages between Exception Monitor and Write Exceptions stages in a dataflow; doing so could impact the configuration of exceptions in the Business Steward Portal.

A continuación, se muestra un caso hipotético de ejemplo que ayuda a ilustrar una implementación básica de gestión de excepciones:



En este ejemplo, hay dos flujos de datos: el flujo de datos inicial que evalúa los datos de código postal de los registros de entrada y el trabajo de reprocesamiento de excepciones, el que toma las excepciones editadas y verifica que los registros ahora contengan datos de código postal válidos.

En ambos flujos de datos, hay una etapa Exception Monitor. Esta etapa contiene las condiciones que desea utilizar para determinar si un registro se debe enrutar para revisión manual. Estas condiciones constan de una o más expresiones, tales como `PostalCode is empty`, lo que significa que cualquier registro que no contenga un código postal se considerará como excepción y se enrutará a la etapa Write Exceptions y se escribirá en el repositorio de excepciones. Para obtener más información, consulte [Exception Monitor](#) en la página 241.

Todos los registros que Exception Monitor identifica como excepciones son enviados a un repositorio de excepciones utilizando la etapa Write Exceptions. Los administradores de datos revisan las excepciones del repositorio utilizando Business Steward Portal, una herramienta de exploración para ver y modificar registros de excepción. Si utilizamos nuestro ejemplo, el administrador de datos podría usar el editor de excepciones del portal Business Steward Portal para agregar manualmente códigos postales a los registros de excepción y marcarlos como "Aprobados".

Cuando se marca un registro como "Aprobado" en Business Steward Portal, ese registro está disponible para que se lo vuelva a leer en un flujo de datos de Spectrum™ Technology Platform. Este proceso se realiza mediante el uso de una etapa Read Exceptions. Si todavía algún registro genera una excepción, se lo escribe nuevamente en el repositorio de excepciones para que un administrador de datos la revise.

Para determinar cuál es la mejor opción para su situación, tenga en cuenta estas preguntas:

- **¿Cómo quiere identificar los registros de excepción?** La etapa Exception Monitor puede evaluar cualquier valor de un campo o cualquier combinación de campos para determinar si un registro es una excepción. Es recomendable analizar los resultados que está obteniendo actualmente con su flujo de datos para determinar cómo quiere identificar las excepciones. También se recomienda identificar registros en el rango medio de la secuencia de calidad de datos y no entre aquellos que claramente están validados o han fallado.
- **¿Quiere que los registros de excepción editados y aprobados vuelvan a ser procesados utilizando la misma lógica que se utilizó en el flujo de datos original?** Si es así, es recomendable utilizar un subflujo para crear una lógica de negocio reutilizable. Por ejemplo, el subflujo puede utilizarse en un flujo de datos inicial que lleva a cabo la validación de direcciones y en un trabajo de reprocesamiento de excepciones que reprocesa los registros corregidos para verificar las correcciones. Luego, puede utilizar diferentes etapas de origen y recepción entre los dos. El flujo de datos inicial puede contener una etapa Read from DB que toma datos desde la base de datos de sus clientes para ser procesados. El trabajo de reprocesamiento de excepciones contiene una etapa Read Exceptions que toma los registros de excepción editados y aprobados desde el repositorio de excepciones.
- **¿Quiere volver a procesar correctamente y aprobar las excepciones siguiendo una programación predeterminada?** Si es así, puede programar su trabajo de reprocesamiento utilizando Programación en la Consola de administración.

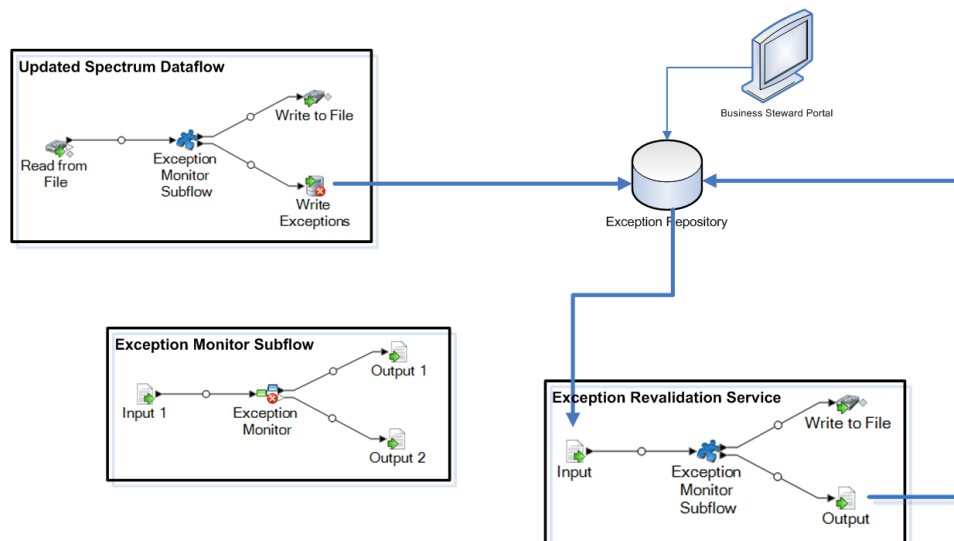
Diseño de un flujo de datos para revalidación en tiempo real

Si utiliza la gestión de excepciones en su flujo de datos, puede usar la función de revalidación para volver a ejecutar los registros de excepción a través del proceso de validación, después de su corrección en el portal Business Steward. Esto le permite determinar si el cambio realizado hace que el registro se procese correctamente en tiempo real; no es necesario que espere hasta que se vuelva a ejecutar el trabajo en lote Read Exceptions para ver los resultados.

Los factores básicos del entorno de revalidación son:

- Un trabajo o servicio que reutiliza o contiene un subflujo expuesto. También debe contener una fuente de entrada, la etapa del subflujo que procesa la entrada, una etapa Write Exceptions y un receptor de salida para los registros procesados correctamente.
- Un subflujo expuesto con una etapa Exception Monitor que apunte a un servicio de revalidación y que esté configurado para la revalidación, incluida la designación en cuanto a si los registros revalidados se deben volver a procesar o aprobar.
- Un servicio expuesto que también se reutiliza o contiene el subflujo expuesto. Procesa los registros que se editaron, guardaron y enviaron para la revalidación en el portal Business Steward.

A continuación se muestra un caso hipotético de ejemplo, que ayuda a ilustrar una implementación de revalidación:



En este ejemplo, existen tres flujos de datos: un trabajo, un subflujo y un servicio. El trabajo ejecuta los datos de entrada a través del subflujo. El subflujo contiene una etapa Exception Monitor, que determina si un registro debe ser asignado para revisión manual. Continuando con el ejemplo, eso quiere decir que cualquier registro sin datos en el campo PostalCode se considerará una excepción y será asignado para la etapa Write Exceptions; estas excepciones son las que aparecen en el

portal Business Steward. Los registros que contengan cualquier otra cosa en dicho campo se asignarán para la etapa Write to File.

Nota: Si el flujo de datos también se configura para utilizar la funcionalidad Best Of Breed, deberá agregar y exponer manualmente el campo CollectionRecordType en la etapa/subflujo de revalidación Exception Monitor y el servicio mismo. Consulte las opciones de **Write Exceptions** y **Creación de un registro Best of Breed**, en la página 270 para obtener más información sobre la funcionalidad Best Of Breed.

El servicio de revalidación de excepción que designe cuando configure la etapa Exception Monitor se invoca cuando edita uno o más registros de excepción en el editor de excepciones de Business Steward Portal y hace clic en **Guardar**. Al igual que el trabajo, el servicio contiene un subflujo de Exception Monitor, que utiliza la misma lógica comercial para volver a procesar los registros. Si los registros no cumplen una o más condiciones establecidas en la etapa Exception Monitor, las excepciones se actualizarán en el repositorio. Si los registros cumplen las establecidas en la etapa Exception Monitor, ocurrirá una de dos acciones, según la selección realizada en el campo "Acción después de revalidación":

- **Volver a procesar registros:** los registros se borrarán del repositorio y se volverán a procesar.
- **Aprobar registros:** los registros se marcarán como aprobados y se enviarán de vuelta al repositorio.

Siga estos pasos para crear y utilizar un caso hipotético de revalidación en tiempo real:

1. Abra o cree un flujo de datos de trabajo o servicio, que contenga una etapa Exception Monitor, una fuente de entrada (como por ejemplo una etapa Read From File o Input), un receptor de salida (como una etapa Write To File o Output) y una etapa Write Exceptions.
2. Convierta la etapa Exception Monitor a un subflujo y asigne los campos de entrada y salida para que coincidan con los del flujo de datos inicial. Asegúrese de incluir el campo ExceptionMetadata para la fuente de entrada, así como también, la etapa de salida que completa la etapa Write Exceptions en el trabajo. Exponga el subflujo de modo que el trabajo y el servicio lo puedan usar.
3. Cree un servicio que contenga una etapa Input, el subflujo que creó en el paso 2, una etapa Output y un receptor de salida (como una etapa Write to File o Write to DB). Asigne los campos de entrada y salida para que coincidan con los del flujo de datos inicial; asegúrese incluir el campo ExceptionMetadata para la etapa Input y la etapa Output. Exponga el servicio de modo que el subflujo lo pueda utilizar.
4. Vuelva al subflujo y abra la ficha Configuración de la etapa Exception Monitor. Seleccione el servicio de revalidación que creó en el paso 3 y especifique la acción que se realizará después de la revalidación. Guarde y exponga nuevamente el subflujo.
5. Vuelva al servicio, donde aparecerá un mensaje, que le notifica de los cambios en el subflujo e indica que el servicio se actualizará. Haga clic en **Aceptar**, luego guarde y exponga nuevamente el servicio.
6. Vuelva al servicio o trabajo inicial, donde aparecerá un mensaje, que le notifica de los cambios en el subflujo e indica que el flujo de datos se actualizará. Haga clic en **Aceptar**, luego guarde el flujo de datos.
7. Ejecute el trabajo.

Nota: Aunque haya ejecutado antes el servicio o trabajo inicial, debe volver a ejecutarlo nuevamente después de crear el caso hipotético de revalidación, para completar el repositorio con los registros que pueden ser seleccionados para la revalidación. Puede identificar si los registros en el editor de excepciones son adecuados para la revalidación, ya que el botón **Guardar** estará activo para dichos registros.

7 - Tablas de búsqueda

In this section

Introducción a las tablas de búsqueda	160
Referencia de etapas	160
Tablas del módulo Universal Name	166
Cómo ver el contenido de Table Lookup	167
Cómo agregar un término a Table Lookup	168
Eliminación de un término en una tabla de búsqueda	169
Cómo modificar la forma estandarizada de un término	169
Cómo revertir personalización de tabla	169
Creación de una tabla de búsqueda	170
Importar datos	170

Introducción a las tablas de búsqueda

Una tabla de búsqueda es una tabla de pares de claves o valores utilizada por las etapas de Spectrum™ Technology Platform para estandarizar datos realizando reemplazo de muestras. Para modificar el contenido de las tablas de búsqueda que se usan en Advanced Transformer, Open Parser y Table Lookup, utilice la herramienta Table Management de Enterprise Designer.

Referencia de etapas

Tablas de Advanced Transformer

Advanced Transformer utiliza las siguientes tablas para identificar términos. Utilice Administración de tablas para crear tablas nuevas o modificar las ya existentes. Para obtener más información, consulte [Introducción a las tablas de búsqueda](#) en la página 160.

- Abreviaturas aeronáuticas
- Iniciales de todas las siglas
- Abreviaturas de nombres de empresas
- Abreviaturas de territorios canadienses
- Abreviaturas de informática/TI
- Delimitadores
- Empresas alemanas
- Fortune 1000
- Abreviaturas de direcciones geográficas
- Términos irrelevantes de Global Sentry
- Países sancionados de Global Sentry
- Abreviaturas de agencias gubernamentales
- Designadores de aerolíneas de IATA (Asociación de Transporte Aéreo Internacional)
- País designador de aerolíneas de IATA
- Abreviaturas legales
- Abreviaturas médicas
- Siglas de organizaciones médicas
- Abreviaturas militares
- Apodos

- Abreviaturas de unidades secundarias
- Inverso de unidades secundarias
- Abreviaturas de Singapur
- Abreviaturas en español
- Abreviaturas de direcciones en español
- Abreviaturas de sufijos de calles en español
- Abreviaturas de nombres de estados
- Inverso de nombre de estado
- Abreviaturas de sufijos de calle
- Inverso de sufijo de calle
- Filial a matriz
- Bases de datos de Estados Unidos
- Bases de datos de Estados Unidos

Tablas de Open Parser

Open Parser utiliza las siguientes tablas para identificar términos. Utilice Administración de tablas para crear tablas nuevas o modificar las ya existentes. Para obtener más información, consulte [Introducción a las tablas de búsqueda](#) en la página 160.

Tablas base

Las tablas base están incluidas en el paquete de instalación del módulo Data Normalization.

- Descripciones de cuentas
- Compañías
- Conjunciones de empresa
- Preposiciones de compañía
- Sufijos de compañía
- Términos de compañía
- Conjunciones
- Prefijos de apellido
- Apellidos
- Sufijos generales
- Empresas alemanas
- Nombres
- Sufijos generacionales
- Nombres de pila en español
- Apellidos españoles
- Títulos

Tablas de nombres básicos

Las tablas de nombres básicos no se incluyen en el paquete de instalación del módulo Data Normalization y, por lo tanto, requieren una licencia adicional. Para obtener más información, contacte a su ejecutivo de cuenta.

Las tablas de nombres básicos deben cargarse usando la utilidad de carga de bases de datos del módulo Data Normalization. Para obtener instrucciones, consulte la *Guía de instalación de Spectrum™ Technology Platform*.

- Apellidos mejorados
- Nombres de pila mejorados

Tablas de nombres de empresas

Las tablas de nombres de empresas no se incluyen en el paquete de instalación del módulo Data Normalization y, por lo tanto, requieren una licencia adicional. Para obtener más información, contacte a su ejecutivo de cuenta.

Las tablas de nombres de empresas deben cargarse usando la utilidad de carga de bases de datos del módulo Data Normalization. Para obtener instrucciones, consulte la *Guía de instalación de Spectrum™ Technology Platform*.

- Empresas - América
- Empresas - Pacífico Asiático
- Empresas - EMEA
- Artículos de la empresa
- Conjunciones de empresa

Tablas de paquete Plus para árabe

Las tablas de paquete Plus para árabe no se incluyen en el paquete de instalación del módulo Data Normalization y, por lo tanto, requieren una licencia adicional. Para obtener más información, contacte a su ejecutivo de cuenta.

Las tablas de Paquete Plus para árabe deben cargarse usando la utilidad de carga de bases de datos del módulo Data Normalization. Para obtener instrucciones, consulte la *Guía de instalación de Spectrum™ Technology Platform*.

- Apellidos árabes (árabe)
- Apellidos árabes (romanizados)
- Nombres de pila árabes (árabe)
- Nombres de pila árabes (romanizados)

Tablas de paquete Plus para idiomas asiáticos

Las tablas de paquete Plus para idiomas asiáticos no se incluyen en el paquete de instalación del módulo Data Normalization y, por lo tanto, requieren una licencia adicional. Para obtener más información, contacte a su ejecutivo de cuenta.

Las tablas de Paquete Plus para idiomas asiáticos deben cargarse usando la utilidad de carga de bases de datos del módulo Data Nominalization. Para obtener instrucciones, consulte la *Guía de instalación de Spectrum™ Technology Platform*.

- Apellidos chinos (nativos)
- Apellidos chinos (romanizados)
- Nombres de pila chinos (nativos)
- Nombres de pila chinos (romanizados)
- Apellidos en coreano (nativos)
- Apellidos en coreano (romanizados)
- Nombres de pila en coreano (nativos)
- Nombres de pila en coreano (romanizados)
- Apellidos japoneses (Kana)
- Apellidos japoneses (Kanji)
- Apellidos japoneses (romanizados)
- Nombres de pila japoneses (Kana)
- Nombres de pila japoneses (Kanji)
- Nombres de pila japoneses (romanizados)

Tablas de Table Lookup

Table Lookup utiliza las siguientes tablas para identificar términos. Utilice Administración de tablas para crear tablas nuevas o modificar las ya existentes. Para obtener más información, consulte [Introducción a las tablas de búsqueda](#) en la página 160.

Tablas base

Las tablas base están incluidas en el paquete de instalación del módulo Data Normalization.

- Abreviaturas aeronáuticas
- Iniciales de todas las siglas
- Abreviaturas de nombres de empresas
- Abreviaturas de territorios canadienses
- Abreviaturas de informática/TI
- Siglas de la Unión Europea
- Fortune 1000
- Abreviaturas en francés
- Distrito (arrondissement) a número de departamento de Francia

- Comuna a código postal de Francia
- Departamento a región de Francia
- Número de departamento a departamento de Francia
- Códigos de género
- Abreviaturas de direcciones geográficas
- Siglas alemanas
- Ciudad a código de estado de Alemania
- Código de área a ciudad de Alemania
- Distrito a código de estado de Alemania
- Abreviaturas de estados alemanes
- Países sancionados de Global Sentry
- Abreviaturas de agencias gubernamentales
- Designadores de aerolíneas de IATA (Asociación de Transporte Aéreo Internacional)
- País designador de aerolíneas de IATA
- Abreviaturas legales
- Abreviaturas médicas
- Siglas de organizaciones médicas
- Abreviaturas militares
- Apodos
- Abreviaturas de unidades secundarias
- Inverso de unidades secundarias
- Abreviaturas de Singapur
- Abreviaturas en español
- Abreviaturas de direcciones en español
- Abreviaturas de sufijos de calles en español
- Abreviaturas de nombres de estados
- Inverso de nombre de estado
- Abreviaturas de sufijos de calle
- Inverso de sufijo de calle
- Filial a matriz
- Reino Unido Municipio a área de código postal
- Reino Unido Prefijos telefónicos
- Reino Unido Códigos telefónicos a municipios
- Reino Unido Área de código postal a municipio
- Bases de datos de Estados Unidos
- Bases de datos de Estados Unidos
- ZREPLACE (utilizado por el módulo SAP para la validación de direcciones de Francia)

Nombres básicos

Las tablas de nombres básicos requieren una licencia adicional. Para obtener más información, contacte a su ejecutivo de cuenta.

Las tablas de nombres básicos deben cargarse usando la utilidad de carga de bases de datos del módulo Data Nominalization. Para obtener instrucciones, consulte la *Guía de instalación de Spectrum™ Technology Platform*.

- Origen étnico mejorado de apellidos
- Códigos mejorados de género
- Origen étnico mejorado de nombres de pila

Paquete Plus para árabe

Las tablas de Paquete Plus para árabe requieren una licencia adicional. Para obtener más información, contacte a su ejecutivo de cuenta.

Las tablas de Paquete Plus para árabe deben cargarse usando la utilidad de carga de bases de datos del módulo Data Nominalization. Para obtener instrucciones, consulte la *Guía de instalación de Spectrum™ Technology Platform*.

- Origen étnico de apellidos árabes (arábigos)
- Origen étnico de apellidos árabes (romanizados)
- Códigos árabes de género (árabes)
- Códigos árabes de género (romanizados)
- Origen étnico de nombres de pila árabes (arábigos)
- Origen étnico de nombres de pila árabes (romanizados)

Paquete Plus para idiomas asiáticos

Las tablas de Paquete Plus para idiomas asiáticos requieren una licencia adicional. Para obtener más información, contacte a su ejecutivo de cuenta.

Las tablas de Paquete Plus para idiomas asiáticos deben cargarse usando la utilidad de carga de bases de datos del módulo Data Nominalization. Para obtener instrucciones, consulte la *Guía de instalación de Spectrum™ Technology Platform*.

- Origen étnico de apellidos chinos, japoneses y coreanos (nativos)
- Origen étnico de apellidos chinos, japoneses y coreanos (romanizados)
- Origen étnico de nombres de pila chinos, japoneses y coreanos (nativos)
- Origen étnico de nombres de pila chinos, japoneses y coreanos (romanizados)
- Códigos japoneses de género (Kana)
- Códigos japoneses de género (Kanji)
- Códigos japoneses de género (romanizados)

Tablas del módulo Universal Name

Tablas de Name Variant Finder

La etapa Name Variant Finder utiliza las siguientes tablas. Cada tabla requiere una licencia por separado.

- Paquete Plus para idioma árabe: `g1-cdq-cjki-arabic-<date>.jar`
- Paquete Plus para idiomas asiáticos (chino): `g1-cdq-cjki-chinese-<date>.jar`
- Paquete Plus para idiomas asiáticos (japonés): `g1-cdq-cjki-japanese-<date>.jar`
- Paquete Plus para idioma asiático-coreano: `g1-cdq-cjki-korean-<date>.jar`
- Base de datos de nombres básicos: `g1-cdq-nomino-base-<date>.jar`

Tablas de Open Name Parser

Open Name Parser utiliza las tablas a continuación para identificar términos. Utilice Administración de tablas para crear tablas nuevas o modificar las ya existentes. Para obtener más información, consulte [Introducción a las tablas de búsqueda](#) en la página 160.

Tablas base

Las tablas base están incluidas en el paquete de instalación del módulo Universal Name.

- Descripciones de cuentas
- Conjunciones de empresa
- Conjunciones
- Prefijos de apellido
- Apellidos
- Sufijos generales
- Nombres
- Sufijos generacionales
- Nombres de pila en español
- Apellidos españoles
- Títulos

Tablas de nombres básicos

Las tablas de nombres básicos no se incluyen en el paquete de instalación del módulo Universal Name y, por lo tanto, requieren una licencia adicional.

- Apellidos mejorados
- Nombres de pila mejorados

Tablas de nombres de empresas

Las siguientes tablas de nombres de empresas se incluyen en el paquete de instalación del módulo Universal Name.

- Descripciones de cuentas
- Compañías
- Artículos de la empresa
- Conjunciones de empresa
- Preposiciones de compañía
- Sufijos de compañía
- Términos de compañía
- Conjunciones

Las siguientes tablas de nombres de empresas no se incluyen en el paquete de instalación del módulo Universal Name y, por lo tanto, requieren una licencia adicional.

- Empresas - América
- Empresas - Pacífico Asiático
- Empresas - EMEA

Tablas de paquete Plus para idiomas asiáticos

Las tablas del paquete Plus para Asia no se incluyen en el paquete de instalación del módulo Universal Name y, por lo tanto, requieren una licencia adicional.

- Apellidos japoneses (Kana)
- Apellidos japoneses (Kanji)
- Apellidos japoneses (romanizados)
- Nombres de pila japoneses (Kana)
- Nombres de pila japoneses (Kanji)
- Nombres de pila japoneses (romanizados)
- Títulos japoneses

Cómo ver el contenido de Table Lookup

Puede ver el contenido de Table Lookup mediante la administración de tablas en Enterprise Designer.

1. En Enterprise Designer, seleccione **Herramientas > Administración de tablas**.
2. En el campo **Tipo**, seleccione la etapa de la tabla de búsqueda que desea ver.
3. En el campo **Nombre**, elija la tabla que desea ver.
4. Puede utilizar las siguientes opciones para cambiar la forma en que se muestra el cuadro:

Opción	Descripción
Encontrar un término específico	En el campo Comienza con , escriba el término que desea buscar y haga clic en Actualizar .
Página a través de la tabla	Haga clic en los íconos hacia adelante y hacia atrás a la derecha del botón Actualizar .
Cambie el número de términos que se muestra por página.	Cambie el valor del campo Elementos por página .
Vea todos los términos de búsqueda para cada término estandarizado en una tabla de Table Lookup.	En el campo Ver por , seleccione Tpermino estandarizado (agrupación) . Esta opción solo está disponible para las tabla de Table Lookup.

Cómo agregar un término a Table Lookup

Si descubre que sus datos tienen términos que no se incluyen en la tabla de búsqueda y desea agregarlos a una tabla de búsqueda, siga este procedimiento.

1. En Enterprise Designer, seleccione **Herramientas > Administración de tablas**.
2. En el campo **Tipo**, seleccione la etapa cuya tabla de búsqueda desea modificar.
3. En el campo **Nombre**, seleccione la tabla a la que desea agregar un término.
4. Haga clic en **Agregar**.
5. En el campo **Término de búsqueda**, escriba el término que existe en los datos. Esta es la clave de búsqueda que se va a utilizar.
6. Para Table Lookup, ingrese el término que desea que reemplace el término de búsqueda en su flujo en **Término estandarizado**.

Por ejemplo, si desea cambiar el término PB a Pitney Bowes, debe ingresar PB como el término de búsqueda y de Pitney Bowes como el término estandarizado.

7. Para las tablas de Table Lookup, seleccione la casilla de verificación **Invalidar término existente** si el término ya existe en la tabla y desea reemplazarlo por el valor que ingresó en el paso 5.
8. Haga clic en **Agregar**.

Eliminación de un término en una tabla de búsqueda

Para eliminar un término de tabla de búsqueda:

1. En Enterprise Designer, seleccione **Herramientas > Administración de tablas**.
2. Seleccione el término y haga clic en **Eliminar**.
3. Haga clic en **Sí** para eliminar el término de la tabla.

Cómo modificar la forma estandarizada de un término

Para las tablas que se utilizan en Table Lookup para estandarizar términos, puede cambiar la forma estandarizada de un término. Por ejemplo, si tiene una tabla con los términos de búsqueda PB y PB Software, el término normalizado es Pitney Bowes y desea cambiar la forma estandarizada a Pitney Bowes Inc, puede hacerlo siguiendo este procedimiento.

1. En Enterprise Designer, seleccione **Herramientas > Administración de tablas**.
2. En el campo **Tipo**, seleccione **Table Lookup**.
3. En el campo **Nombre**, seleccione la tabla que desea modificar.
4. Seleccione el término que desea modificar y haga clic en **Modificar**.

Consejo: Si hay varios términos de búsqueda estandarizados para un término, puede modificar fácilmente todos los términos de búsqueda para utilizar el nuevo término estandarizado, seleccionando **Ver por término estandarizado (agrupación)** en el campo **Ver por**. Seleccione el grupo y haga clic en **Modificar**.

5. Ingrese un nuevo valor en el campo **Término estandarizado**.
6. Haga clic en **Aceptar**.

Cómo revertir personalización de tabla

Si realiza modificaciones en una tabla, puede revertir la tabla a su estado original. Para revertir personalizaciones de tabla:

1. En Enterprise Designer, seleccione **Herramientas > Administración de tablas**.
2. Seleccione el informe que desea revertir.
3. Haga clic en **Revertir**.

Aparecerá la ventana **Restaurar**. Se muestra una lista de todos los términos agregados, eliminados y modificados.

4. Seleccione la casilla de verificación **Revertir** para cada entrada de tabla que desee revertir. También puede hacer clic en **Seleccionar todo** o **Deseleccionar todo** para seleccionar o deseleccionar todas las casillas de verificación **Revertir**.
5. Haga clic en **Aceptar**.

Creación de una tabla de búsqueda

El módulo Advanced Matching, Data Normalization y Universal Name vienen con una variedad de tablas que se pueden utilizar para una amplia gama de sustitución de términos o procesos de estandarización. Sin embargo, si estas tablas no satisfacen sus necesidades, puede crear su propia tabla de términos de búsqueda con Advanced Transformer, Open Parser o Table Lookup. Para crear una tabla, siga este procedimiento.

1. En Enterprise Designer, seleccione **Herramientas > Administración de tablas**.
2. En el campo **Tipo**, seleccione la etapa para la que desea crear una tabla de búsqueda.
3. Haga clic en **Nuevo**. Aparecerá el cuadro de diálogo **Agregar tabla**.
4. En el campo **Nombre de la tabla**, ingrese un nombre para la nueva tabla.
5. Si desea obtener una tabla en blanco del tipo seleccionado, deje el campo **Copiar desde** establecido en **Ninguno**. Si desea que la nueva tabla se complete a partir de una ya existente, seleccione un nombre de la lista **Copiar desde**.
6. Haga clic en **Aceptar**.

Para obtener más información sobre cómo agregar elementos de tabla a una nueva tabla, consulte [Cómo agregar un término a Table Lookup](#) en la página 168.

Importar datos

Importar datos en una tabla de búsqueda

Puede importar datos desde un archivo en una tabla de búsqueda para usar con Advanced Transformer, Open Parser, o Table Lookup. A fin de poder importar datos desde un archivo en una tabla de búsqueda, el archivo debe cumplir los siguientes requisitos:

- Debe tener codificación UTF-8.
- Debe ser un archivo delimitado. Los caracteres delimitadores admitidos son la coma (,), el punto y coma (;), la barra vertical (|) y la tabulación (\t).
- Los campos con delimitadores incrustados deben comenzar y finalizar con comillas dobles, por ejemplo "1,a","2,b","3,c".
- Un carácter entre comillas en un campo que comienza y finaliza con comillas dobles debe tener dos comillas, por ejemplo "2"" pies".

Para importar datos de un archivo en una tabla de búsqueda:

1. En Enterprise Designer, seleccione **Herramientas > Administración de tablas**.
2. Seleccione la tabla a la que desea importar los datos. O bien, cree una nueva tabla. Para obtener instrucciones sobre cómo crear una tabla, consulte [Creación de una tabla de búsqueda](#) en la página 170.
3. Haga clic en **Importar**.
4. Haga clic en **Examinar** y seleccione el archivo que desea importar.
5. Haga clic en **Abrir**. En Ver el archivo en vista previa, se muestra una vista previa de los datos del archivo importado.
6. Puede seleccionar las columnas de una tabla definida por el usuario y asignarlas a la tabla existente. Por ejemplo, suponga que hay dos columnas en la tabla definida por el usuario que desea importar. Contiene column1 y column2. La lista de columnas mostraría la column1 y column2. Puede seleccionar column2 para asignarla a un término de búsqueda y seleccionar column1 para asignarla a un término estandarizado.
7. Seleccione **Importar solo términos nuevos** para importar solo los registros nuevos desde la tabla definida por el usuario o **Sobrescribir términos ya existentes** para importar todos los registros de las columnas seleccionadas.
8. Haga clic en **Aceptar**.




Uso de Importación avanzada

La función de Importación avanzada permite importar selectivamente los datos en las tablas de búsqueda que usa Advanced Transformer, Table Lookup, y Open Parser. Use Importación avanzada para:

- Extraiga términos de una columna seleccionada en un archivo delimitado y definido por el usuario.
- Extraiga términos de una sola palabra (muestras) de una columna seleccionada en un archivo definido por el usuario y delimitado. Cuando extrae muestras, puede identificar la cantidad de veces que los términos ocurren para una columna dada en el archivo y crear agrupamientos para los términos relacionados y agregarlos a la tabla.

El archivo que contiene los datos que desea importar debe cumplir los siguientes requisitos:

- Debe tener codificación UTF-8.

- Debe ser un archivo delimitado. Los caracteres delimitadores admitidos son la coma (,), el punto y coma (;), la barra vertical (|) y la tabulación (\t).
 - Los campos con delimitadores incrustados deben comenzar y finalizar con comillas dobles, por ejemplo "1,a","2,b","3,c".
 - Un carácter entre comillas en un campo que comienza y finaliza con comillas dobles debe tener dos comillas, por ejemplo "2"" pies".
1. En Enterprise Designer, seleccione **Herramientas > Administración de tablas**.
 2. Seleccione la tabla a la que desee importar los datos.
 3. Haga clic en **Importación avanzada**.
 4. Haga clic en **Examinar** y seleccione el archivo que desea importar.
 5. Haga clic en **Abrir**.
 6. Seleccione una columna de tabla de la lista **Columna**. Los datos de muestra exhiben la frecuencia con que ocurre cada término de la lista en la tabla definida por el usuario. Solo se muestra la frecuencia de los términos que todavía no están en la tabla existente.
 7. Para ver los términos como palabras únicas, seleccione **Separar en términos de una palabra**.
 8. Para las tablas de Advanced Transformer y Open Parser:
 - a) Seleccione un término de la lista de la izquierda.
 - b) Haga clic en la flecha derecha para agregar el término a la lista de la derecha. Haga clic en la flecha izquierda para eliminar un término seleccionado de la lista de tablas.
 - c) Haga clic en **Aceptar** para guardar los cambios realizados en la tabla.
 9. Para tablas de Table Lookup:
 - a) Haga clic en  para agregar una agrupación de tablas.
 - b) Haga clic en **Nuevo**.
 - c) Ingrese un término nuevo y luego haga clic en **Agregar**. Siga agregando términos hasta que finalice y luego haga clic en **Cerrar**.
 - d) Seleccione un término de la lista y luego haga clic en Agregar. Siga agregando términos hasta que finalice y luego haga clic en **Cerrar**. Los términos nuevos se agregarán a la lista de términos de la derecha.
 - e) Seleccione un término de la izquierda y luego haga clic en la flecha derecha para agregar el término a la agrupación seleccionada. Haga clic en la flecha izquierda para eliminar un término de una de las agrupaciones.
 - f) Para modificar un término, selecciónelo en la lista de la derecha y luego haga clic en .
 - g) Para eliminar un término, selecciónelo en la lista de la derecha y luego haga clic en .
 - h) Haga clic en **Aceptar** para guardar los cambios realizados en la tabla.

8 - Referencia de etapas

In this section

Módulo Advanced Matching	174
Módulo Business Steward	240
Módulo Data Normalization	285
Módulo Universal Name	304

Módulo Advanced Matching

Módulo Advanced Matching

El módulo Advanced Matching (Comparación avanzada) establece cruces de registros entre una cantidad indefinida de archivos de entrada y/ o dentro de estos archivos. También es posible utilizar el módulo Advanced Matching para establecer cruces en una amplia variedad de campos, lo que incluye los campos de nombre, dirección, nombre y dirección, o bien campos sin nombre ni dirección, como los de número de seguro social o fecha de nacimiento.

Componentes

El módulo Advanced Matching consiste en:

- **Best-of-Breed:** esta etapa selecciona un registro "best-of-breed" del clúster de duplicados seleccionando un registro de plantilla que luego usa para elaborar el mejor registro compuesto. A continuación, este registro se devuelve como registro sobreviviente.
- **Candidate Finder:** esta etapa obtiene los registros candidatos que conformarán los posibles cruces que serán evaluados por la etapa Transactional Match. Candidate Finder (Localizador de candidatos) se utiliza junto con Transactional Match (Cruce transaccional).
- **Duplicate Synchronization:** esta etapa le permite especificar cuáles son los campos de una colección de registros que se copiarán (publicarán) en los campos correspondientes de todos los registros de la colección.
- **Filter:** esta etapa le permite especificar los criterios con los que deben cumplir los registros para ser conservados o eliminados de una colección de registros duplicados, ya sea para procesarlos posteriormente o copiarlos en el archivo de salida.
- **Interflow Match:** identifica los cruces (coincidencias) entre registros de datos similares en múltiples flujos de entrada.
- **Intraflow Match:** esta etapa identifica los cruces (coincidencias) entre registros de datos similares en un mismo flujo de entrada.
- **Análisis de cruce:** utilice el análisis de cruce para analizar y comparar los resultados de las etapas de comparación a fin de tener una mejor comprensión de los resultados del proceso de comparación y mejorarlos.
- **Match Key Generator:** esta etapa crea una clave no exclusiva que es compartida por todos los registros similares. Los registros que comparten la misma clave pueden agruparse a los fines de la comparación.
- **Private Match:** esta etapa permite que dos entidades comparen conjuntos de datos e identifiquen registros o clientes comunes, sin comprometer información confidencial. La seguridad se mantiene mientras se realiza la función de cruce mediante el uso de la función de cifrado y, luego, la función de descifrado muestra el resultado de los datos cruzados.

- **Transactional Match:** esta etapa compara las transacciones sospechosas con una base de datos en la etapa Candidate Finder para consultar y obtener posibles registros candidatos.
- **Write to Search Index:** esta etapa permite crear un índice de texto completo según los datos que entran en la etapa. Al tener estos datos en un índice de búsqueda exclusivo, se obtiene un tiempo de respuesta más rápido cuando se realizan búsquedas en el índice desde Candidate Finder.

Best of Breed

Best of Breed consolida los registros duplicados. Para ello, selecciona los mejores datos de una colección de registros duplicados y crea un nuevo registro consolidado con los mejores datos. Este "súper" registro se conoce como el registro "best of breed" (el mejor de la especie). Usted debe definir las reglas a usar para seleccionar los registros que habrán de procesarse. Cuando el proceso finaliza, el registro "best of breed" se conserva en el sistema.

Opciones

La siguiente tabla muestra las opciones para Best of Breed.

Nombre de la opción	Descripción / Valores válidos
Agrupar por	Especifica el campo que se utiliza para crear grupos de registros para fusionar en un solo registro best of breed que se crea a partir de cada grupo. En casos en que haya usado una etapa de comparación anteriormente en el flujo de datos, debe seleccionar el campo CollectionNumber para usar las colecciones que la etapa de comparación creó como los grupos. Sin embargo, si desea agrupar los recursos por otros campos, elija el campo aquí. Por ejemplo, si desea fusionar en un registro best of breed todos los registros que tienen el mismo valor en el campo AccountNumber, debe seleccionar AccountNumber.
Ordenar	Si especifica un campo en el campo Agrupar por , marque esta casilla para ordenar los registros por valor en el campo que escoja. Esta opción está habilitada de forma predeterminada.

Nombre de la opción	Descripción / Valores válidos
Avanzada	<p>Haga clic en este botón para especificar las opciones de rendimiento de orden. De manera predeterminada, las opciones de rendimiento de orden que se especifican en Management Console, que son las opciones de rendimiento predeterminadas para su sistema, están activas. Si desea reemplazar las opciones de rendimiento predeterminadas del sistema, marque la casilla Invaldar opciones de rendimiento de orden y luego especifique en estos campos los valores que desea:</p> <p>Límite de registros en memoria Especifica la cantidad máxima de filas de datos que un ordenador guardará en la memoria antes de empezar la paginación al disco. De forma predeterminada, la ordenación de 10.000 registros o menos se realizará en la memoria y la de más de 10.000 registros se realizará como una ordenación en disco. El límite máximo es de 100.000 registros. Normalmente, la ordenación que se realiza en la memoria es más rápida que la que se realiza en disco; por lo tanto, este valor se debe establecer lo suficientemente alto para que la mayor parte de la ordenación sea en la memoria y solo los conjuntos de datos grandes se escriban en el disco.</p> <p>Nota: Tenga cuidado en entornos en los que se ejecuten trabajos de forma concurrente, porque aumentar la configuración del Límite de registros en memoria aumenta las posibilidades de agotar la memoria.</p> <p>Número máximo de archivos temporales Especifica el número máximo de archivos temporales que puede utilizar el proceso de orden. El uso de una gran cantidad de archivos temporales puede producir un mejor rendimiento. Sin embargo, la cantidad óptima depende en gran medida de la configuración en el servidor que ejecuta Spectrum™ Technology Platform. Debe experimentar distintas configuraciones y observar el efecto en el rendimiento de usar una mayor o menor cantidad de archivos temporales. Para calcular la cantidad aproximada de archivos temporales que se pueda necesitar, use la siguiente ecuación:</p> $\frac{(NumberOfRecords \times 2)}{InMemoryRecordLimit} = NumberOfTempFiles$ <p>Tenga en cuenta que la cantidad máxima de archivos temporales no puede ser superior a 1.000.</p> <p>Activar compresión Especifica que los archivos temporales se comprimen cuando se los escribe en el disco.</p> <p>Nota: La configuración del rendimiento óptimo de orden depende de la configuración del hardware de su servidor. No obstante, la siguiente ecuación generalmente produce un buen rendimiento de orden:</p> $(InMemoryRecordLimit \times MaxNumberOfTempFiles \div 2) \geq TotalNumberOfRecords$

Nombre de la opción	Descripción / Valores válidos
Mantener registros originales	Marque esta opción para mantener todos los registros en la colección, junto con el registro best of breed. Desactive la opción si solamente desea tener el registro Best of Breed.
Usar primer registro	Marque esta opción si desea que Best of Breed seleccione automáticamente el primer registro de la colección como el registro de plantilla. El registro de plantilla es el registro en el que se basa el registro best of breed.
Definir registro de plantilla	Marque esta opción para definir reglas de selección del registro de plantilla. Para obtener más información, consulte Definición de Reglas de registro de plantilla en la página 177.

Definición de Reglas de registro de plantilla

En el procesamiento Best of Breed, el registro de plantilla es el registro en una colección que se utiliza para crear el mejor registro. El registro de plantilla se usa como punto de partida para generar el mejor registro, y se modifica en función de la mejor configuración definida por usted. La etapa Best Of Breed puede seleccionar automáticamente el registro de plantilla, o usted puede definir reglas para la selección del registro de plantilla. Esta sección describe la forma de definir las reglas de selección de registro de plantilla.

Las reglas de plantilla se escriben especificando el nombre de campo, un operador, un tipo de valor y un valor. A continuación se muestra un ejemplo de opciones de registro de plantilla:

Nombre del campo: MatchScore
 Tipo de campo: Numérico
 Operador: Igual
 Tipo de valor: cadena
 Valor: 100

Esta regla de plantilla selecciona el registro en la colección que presenta una Calificación de cruce igual al valor de 100.

El siguiente procedimiento describe la forma de definir una regla de registro de plantilla en la etapa Best of Breed.

1. En la etapa Best of Breed, en **Configuración de registro de plantilla**, seleccione la opción **Definir registro de plantilla**.
2. En el árbol, haga clic en **Reglas**.
3. Haga clic en **Agregar regla**.
4. Complete los siguientes campos.

Opción	Descripción
Nombre de campo	Especifica el nombre del campo de flujo de datos cuyo valor usted desea evaluar para determinar si el registro debe ser el registro de plantilla.
Tipo de campo	Especifica el tipo de datos del campo. Una de las siguientes: No numérico Seleccione esta opción si el campo contiene datos no numéricos (por ejemplo, datos de cadena). Numérico Seleccione esta opción si el campo contiene datos numéricos (por ejemplo, doble, flotante, entre otros).

Opción	Descripción
Operador	Especifica el tipo de comparación que desea utilizar en la evaluación del campo. Una de las siguientes:
Contiene	Determina si el campo contiene el valor especificado. Por ejemplo, "sailboat" contiene el valor "boat".
IGUAL	Determina si el campo contiene el valor exacto especificado.
Mayor que	Determina si el valor de campo es mayor que el valor especificado. Esta operación solamente funciona con campos numéricos.
Mayor o igual que	Determina si el valor de campo es mayor o igual que el valor especificado. Esta operación solamente funciona con campos numéricos.
El más alto	Compara el valor del campo en todo el grupo de registros y determina qué registro tiene el valor más alto en el campo. Por ejemplo, si los campos del grupo contienen valores de 10, 20, 30 y 100, quedaría seleccionado el registro con el valor 100. Esta operación solamente funciona con campos numéricos. Si hay múltiples registros que poseen el valor más largo, se selecciona uno de ellos.
Está vacío	Determina si el campo no contiene ningún valor.
No está vacío	Determina si el campo contiene algún valor.
Menor que	Determina si el valor de campo es menor que el valor especificado. Esta operación solamente funciona con campos numéricos.
Menor o igual que	Determina si el valor de campo es menor o igual que el valor especificado. Esta operación solamente funciona con campos numéricos.
El más largo	Compara el valor del campo en todo el grupo de registros y determina qué registro tiene el valor más largo (en bytes) en el campo. Por ejemplo, si el grupo contiene los valores "Mike" y "Michael", quedaría seleccionado el registro con el valor "Michael". Si hay múltiples registros que poseen el valor más largo, se selecciona uno de ellos.
El más bajo	Compara el valor del campo en todo el grupo de registros y determina qué registro tiene el valor más bajo en el campo. Por ejemplo, si los campos del grupo contienen valores de 10, 20, 30 y 100, quedaría seleccionado el registro con el valor 10. Esta operación solamente funciona con campos numéricos. Si hay múltiples registros que poseen el valor más largo, se selecciona uno de ellos.
Más común	Determina si el valor de campo contiene el valor que aparece más frecuentemente en este campo entre los registros del grupo. Si hay dos o más valores que son los más comunes, no se realiza acción alguna.
Diferente	Determina si el valor de campo no es el mismo que el valor especificado.

Opción	Descripción
Tipo de valor	<p>Especifica el tipo de valor que desea comparar al valor del campo. Una de las siguientes:</p> <p>Nota: Esta opción no está disponible si selecciona el operador El más alto, El más bajo, o El más largo.</p> <p>Campo Seleccione esta opción si desea comparar con este campo el valor del campo de otro flujo de datos.</p> <p>Cadena Seleccione esta opción si desea comparar el campo con un valor específico.</p>
Valor	<p>Especifica el valor que se va a comparar con el valor del campo. Si seleccionó Campo en el campo Tipo de campo seleccione un campo de flujo de datos. Si seleccionó Cadena en el campo de Tipo de valor, ingrese el valor que desea usar para la comparación.</p> <p>Nota: Esta opción no está disponible si selecciona el operador El más alto, El más bajo, o El más largo.</p>

- Haga clic en **Aceptar**.
- Si desea especificar reglas adicionales, haga clic en **Agregar regla**.

Si agrega reglas adicionales, deberá seleccionar un operador lógico para usar entre cada regla. Seleccione **Y** si desea que ambas reglas, la nueva y la anterior, pasen con el fin de ser seleccionadas como registro de plantilla. Seleccione **O** si desea que la regla anterior o la nueva regla pasen con el fin de ser seleccionadas como registro de plantilla.

Ahora ha configurado las reglas que se usarán para seleccionar el registro de plantilla. Configure los mejores ajustes para completar la configuración de la etapa Best of Breed.

Definición de Reglas y acciones Best of Breed

Las reglas y acciones Best of Breed trabajan en conjunto para determinar los campos de los registros duplicados de una colección que se copiarán en el registro Best of breed. Las reglas someten a prueba los valores de un registro, y si el registro aprueba dichas reglas, los datos se copian desde el registro en el registro de plantilla. Las acciones definen cuáles datos se copiarán, y el campo en el registro de plantilla que debería recibir los datos. Luego de que se hayan ejecutado todas las reglas y acciones, el registro de plantilla será el registro Best of breed.

Las reglas y acciones pueden agruparse según condiciones, y se puede tener una variedad de condiciones. Esto le permite lo siguiente

- En la etapa Best of Breed, en **Configuración de Best of Breed**, haga clic en el nodo **Reglas** en el árbol.
- Haga clic en **Agregar regla**.

3. Complete los siguientes campos:

Opción	Descripción
Nombre de campo	Especifica el nombre del campo de flujo de datos cuyo valor usted desea determinar en caso de que se cumpla la condición y se deban realizar las siguientes acciones.
Tipo de campo	Especifica el tipo de datos del campo. Una de las siguientes: No numérico Seleccione esta opción si el campo contiene datos no numéricos (por ejemplo, datos de cadena). Numérico Seleccione esta opción si el campo contiene datos numéricos (por ejemplo, doble, flotante, entre otros).

Opción	Descripción
Operador	Especifica el tipo de comparación que desea utilizar en la evaluación del campo. Una de las siguientes:
Contiene	Determina si el campo contiene el valor especificado. Por ejemplo, "sailboat" contiene el valor "boat".
IGUAL	Determina si el campo contiene el valor exacto especificado.
Mayor que	Determina si el valor de campo es mayor que el valor especificado. Esta operación solamente funciona con campos numéricos.
Mayor o igual que	Determina si el valor de campo es mayor o igual que el valor especificado. Esta operación solamente funciona con campos numéricos.
El más alto	Compara el valor del campo en todo el grupo de registros y determina qué registro tiene el valor más alto en el campo. Por ejemplo, si los campos del grupo contienen valores de 10, 20, 30 y 100, quedaría seleccionado el registro con el valor 100. Esta operación solamente funciona con campos numéricos. Si hay múltiples registros que poseen el valor más largo, se selecciona uno de ellos.
Está vacío	Determina si el campo no contiene ningún valor.
No está vacío	Determina si el campo contiene algún valor.
Menor que	Determina si el valor de campo es menor que el valor especificado. Esta operación solamente funciona con campos numéricos.
Menor o igual que	Determina si el valor de campo es menor o igual que el valor especificado. Esta operación solamente funciona con campos numéricos.
El más largo	Compara el valor del campo en todo el grupo de registros y determina qué registro tiene el valor más largo (en bytes) en el campo. Por ejemplo, si el grupo contiene los valores "Mike" y "Michael", quedaría seleccionado el registro con el valor "Michael". Si hay múltiples registros que poseen el valor más largo, se selecciona uno de ellos.
El más bajo	Compara el valor del campo en todo el grupo de registros y determina qué registro tiene el valor más bajo en el campo. Por ejemplo, si los campos del grupo contienen valores de 10, 20, 30 y 100, quedaría seleccionado el registro con el valor 10. Esta operación solamente funciona con campos numéricos. Si hay múltiples registros que poseen el valor más largo, se selecciona uno de ellos.
Más común	Determina si el valor de campo contiene el valor que aparece más frecuentemente en este campo entre los registros del grupo. Si hay dos o más valores que son los más comunes, no se realiza acción alguna.
Diferente	Determina si el valor de campo no es el mismo que el valor especificado.

Opción	Descripción
Tipo de valor	<p>Especifica el tipo de valor que desea comparar al valor del campo. Una de las siguientes:</p> <p>Nota: Esta opción no está disponible si selecciona el operador El más alto, El más bajo, o El más largo.</p> <p>Campo Seleccione esta opción si desea comparar con este campo el valor del campo de otro flujo de datos.</p> <p>Cadena Seleccione esta opción si desea comparar el campo con un valor específico.</p>

Valor	<p>Especifica el valor que se va a comparar con el valor del campo. Si seleccionó Campo en el campo Tipo de campo seleccione un campo de flujo de datos. Si seleccionó Cadena en el campo de Tipo de valor, ingrese el valor que desea usar para la comparación.</p> <p>Nota: Esta opción no está disponible si selecciona el operador El más alto, El más bajo, o El más largo.</p>
-------	---

- Haga clic en **Aceptar**.
- Si desea especificar especificar reglas adicionales a esta condición, haga clic en **Agregar regla**.
Si agrega reglas adicionales, deberá seleccionar un operador lógico para usar entre cada regla. Seleccione **Y** si desea que pasen ambas reglas, la nueva y la anterior, con el fin de cumplir la condición y que se realicen las acciones asociadas. Seleccione **O** si desea que pase la regla anterior o la regla siguiente, para que se cumpla la condición.
- Haga clic en el nodo **Acciones** en el árbol.
- Haga clic en **Agregar acción**.
- Complete los siguientes campos.

Opción	Descripción
Tipo de origen	<p>Especifica el tipo de datos que se copiará en el registro Best of breed. Las opciones son:</p> <p>Campo Seleccione esta opción si desea copiar un valor de un campo en el registro Best of breed.</p> <p>String Seleccione esta opción si desea copiar un valor constante en el registro Best of breed.</p>

Opción	Descripción
Datos de origen	Especifica los datos que se copiarán en el registro Best of breed. Si el tipo de origen es Campo, seleccione el campo cuyos valores desea copiar en el campo de destino. Si el tipo de origen es Cadena, especifique un valor constante para copiar en el campo de destino.
Destino	Especifica el destino en el registro Best of breed al cual desea copiar los datos especificados en el campo Datos de origen .
Acumular datos de origen	<p>Si los datos en el campo Datos de origen son datos numéricos, puede habilitar esta opción para combinar los datos de origen en todos los registros duplicados y colocar el valor total en el registro Best of breed.</p> <p>Por ejemplo, si hubiese tres registros duplicados en el grupo, y estos contuviesen estos valores en el campo de Depósitos:</p> <p>100.00 20.00 5.00</p> <p>Estos tres valores serían combinados y el valor total, de 125,00, se colocaría en el campo Depósitos del registro Best of breed.</p>

9. Haga clic en **Aceptar**.
10. Se desea especificar acciones adicionales para aplicar a esta condición, haga clic en **Agregar acción** y repita los pasos anteriores.
11. Para agregar otra condición, haga clic en la condición raíz del árbol y luego haga clic en **Agregar condición**.

Ejemplo de regla y acción Best of Breed

Esta regla selecciona el registro que presenta una calificación de cruce igual al valor de 100. Luego, los datos Número de cuenta que corresponden al campo seleccionado se copian al campo AccountNumber en el registro Best of Breed.

Regla

Nombre de archivo: MatchScore
 Tipo de campo: numérico
 Operador: igual
 Tipo de valor: cadena
 Valor: 100

Acción

Tipo de origen: campo
 Datos de origen: AccountNumber
 Destino: AccountNumber

Salida

Tabla 8: Datos de salida de Best of Breed

Nombre de campo Formato Descripción / Valores válidos

Nombre de campo	Formato	Descripción / Valores válidos
CollectionRecordType	String	Identifica la plantilla y los registros Best of Breed (Mejor de la especie) en una colección de registros duplicados. Los valores posibles son:
	Primary	El registro es el registro de plantilla seleccionado en una colección.
	Secondary	El registro no es el registro de plantilla seleccionado en una colección.
	BestOfBreed	El registro es el registro Best of Breed que se acaba de crear en la colección.

Candidate Finder

Candidate Finder (Localizador de candidatos) obtiene los registros candidatos que formarán el conjunto de cruces potenciales. Las búsquedas de base de datos trabajan en conjunto con Transactional Match (Cruce transaccional), y las búsquedas de índice de búsqueda trabajan independientemente de Transactional Match. De acuerdo con el formato de sus datos, es posible que Candidate Finder también necesite analizar el nombre o la dirección del registro sospechoso, los registros candidatos o ambos.

Candidate Finder también permite las búsquedas de índice de texto completo y ayuda a definir los criterios de búsqueda simples y complejos en comparación con caracteres y texto, mediante el uso de distintos tipos de búsqueda (Cualquier palabra que comience con, Contiene, Contiene todo, Contiene cualquiera, No contiene, Aproximada, Patrón, Proximidad, Rango, Comodín) y condiciones (Todas verdaderas, Cualquiera verdadera, Ninguna verdadera).

Opciones de base de datos

El cuadro de diálogo de Candidate Finder no solo permite definir instrucciones SQL que recuperan posibles candidatos de cruce de una base de datos, sino que también le permite asignar las columnas que seleccione de la base de datos a los nombres de campo definidos en su flujo de datos.

Tabla 9: Opciones de base de datos de Candidate Finder

Nombre de la opción	Descripción / Valores válidos
Tipo de buscador	Seleccione Base de datos.
Conexión	<p>Seleccione la base de datos que contenga los registros candidatos. Es posible seleccionar cualquier conexión configurada en Management Console. Para establecer una conexión con una base de datos que no está en la lista, configure una conexión a esa base de datos en Management Console y, luego, cierre y vuelva a abrir Candidate Finder para actualizar la lista de conexiones.</p> <p>Nota: La función Opciones de flujo de datos en Enterprise Designer permite exponer el nombre de conexión para la configuración durante la ejecución.</p>
Instrucción SQL	Ingrese una instrucción SQL en el cuadro de texto como se describe en Definición de la consulta SQL en la página 186
Ficha Asignación de campo	Seleccione la configuración de asignación de campos como se describe en Asignación de columnas de la base de datos a los campos de etapa en la página 187.
Ficha Vista previa	Haga clic en esta ficha para ingresar una clave de cruce de muestra para probar la instrucción SQL SELECCIONAR o la consulta de índice. <apiname>SELECT</apiname>SELECT

Definición de la consulta SQL

Puede ingresar cualquier instrucción SQL select válida en el cuadro de texto del cuadro de diálogo **Opciones de Candidate Finder**.

Nota: `Select *` no es válido.

Por ejemplo, suponga que, en su base de datos, tiene una tabla denominada Customer_Table que tiene las siguientes columnas:

- Customer_Table
- Cust_Name
- Cust_Address
- Cust_City
- Cust_State
- Cust_Zip

Para recuperar todas las filas de la base de datos, puede construir una consulta similar a la siguiente:

```
SELECT Cust_Name, Cust_Address, Cust_City, Cust_State, Cust_Zip from
Customer_Table;
```

Es poco probable que quiera cruzar su transacción con todas las filas de la base de datos. Para obtener solo los registros candidatos relevantes, agregue una cláusula WHERE utilizando la sustitución de la variable. `<apiname>WHERE</apiname>WHERE` La sustitución de variable es una notación especial que se utiliza para que el motor de selección de candidatos reemplace la variable con los datos reales de su registro sospechoso.

Para utilizar la sustitución de variable, encierre el nombre del campo entre llaves y coloque adelante un signo de dólar `${FieldName}`. `<codeph>${FieldName}</codeph>${FieldName}` Por ejemplo, la siguiente consulta arrojará solo aquellos registros con un valor en Cust_Zip que coincida con el valor en PostalCode del registro sospechoso.

```
SELECT Cust_Name, Cust_Address, Cust_City, Cust_State,Cust_Zip
FROM Customer_Table
WHERE Cust_Zip = ${PostalCode};
```

Para SQL 2000, el tipo de datos debe ser idéntico al tipo de datos de Candidate Finder. El controlador JDBC define la variable de entrada de Candidate Finder (Ex: `${MatchKey}`) que se utiliza en la cláusula DONDE para un tipo de datos de nVarChar(4000). `<apiname>WHERE</apiname>WHERE` Si los datos de la base de datos están configurados como un tipo de datos de VarChar, SQL Server omitirá el índice en la base de datos. Si se omite el índice, el rendimiento se deteriorará. Por lo tanto, utilice la siguiente consulta para SQL 2000:

```
SELECT Cust_Name, Cust_Address, Cust_City, Cust_State,Cust_Zip
FROM Customer_Table
WHERE Cust_Zip = CAST(${PostalCode} AS VARCHAR(255));
```

Asignación de columnas de la base de datos a los campos de etapa

Si los nombres de las columnas de su base de datos coinciden de manera exacta con los nombres de los campos de componente, se los asigna automáticamente a los campos correspondientes de la etapa. Si no tienen nombres idénticos, deberá asignar los campos seleccionados (columnas de la base de datos) a los campos de etapa (nombres de campos definidos en el flujo de datos).

Por ejemplo, considere una tabla denominada Customer_Table que tenga las siguientes columnas:

- Cust_Name
- Cust_Address
- Cust_City
- Cust_State
- Cust_Zip

Cuando se recuperan estos registros de la base de datos, es necesario asignar los nombres de las columnas a los nombres de los campos utilizados por Transactional Match (Cruce transaccional)

y otros componentes del flujo de datos. Por ejemplo, Cust_Address podría asignarse a AddressLine1 y Cust_Zip se asignaría a PostalCode.

1. Seleccione la lista desplegable de **Campos seleccionados** en la vista **Opciones de Candidate Finder**. Luego, seleccione la columna Cust_Zip de la base de datos.
2. Seleccione la lista desplegable en **Campos de la etapa**. Luego, seleccione el campo al que desea asignarle la columna.

Por ejemplo, si desea asignar Cust_Zip a PostalCode, primero seleccione Cust_Zip en Campos seleccionados y luego seleccione PostalCode en la fila correspondiente del Campo de la etapa.

Método alternativo para la asignación de campos

Puede utilizar una notación especial en su consulta SQL para asignar campos. Para realizar la asignación, encierre entre llaves el nombre del campo al que desea asignar después del nombre de la columna en su consulta. Al hacer esto, los campos seleccionados se asignan automáticamente a los campos correspondientes de la etapa.

Por ejemplo:

```
select Cust_Name {Name}, Cust_Address {AddressLine1},
       Cust_City {City}, Cust_State {StateProvince},
       Cust_Zip {PostalCode}
from Customer
where Cust_Zip = ${PostalCode};
```

Opciones de índice de búsqueda

El cuadro de diálogo Candidate Finder le permite realizar una búsqueda simple que hace coincidir los valores de campo de entrada de los índices de búsqueda, o una búsqueda avanzada para construir reglas de coincidencias que recuperen posibles candidatos de coincidencia de los índices de búsqueda.

Opciones de índice de búsqueda simple

Tabla 10: Opciones de Candidate Finder

Nombre de la opción	Descripción/Valores válidos
Tipo de buscador	Seleccione Índice de búsqueda.
Nombre	Seleccione el índice correspondiente que se creó con la etapa Índice Write to Search en la página 235 en las etapas implementadas de Advanced Matching en Enterprise Designer.

Nombre de la opción	Descripción/Valores válidos
Starting record (Registro de inicio)	Ingrese el número de registro donde deben comenzar los resultados de la búsqueda. El valor predeterminado es 1.
Resultados máximos	Ingrese la cantidad máxima de respuestas que desea que arroje el índice de búsqueda. El valor predeterminado es 10. Nota: Si la cantidad de resultados máximos es demasiado alta, procese los resultados en lotes mediante el campo Tamaño de lotes para la recuperación .
Recuperar tamaño de lote	Si el valor de Resultados máximos es demasiado alto, especifique el tamaño de lote que usará para procesar los resultados. Esto optimiza el procesamiento de cantidades altas de registros. El valor predeterminado es 10000. El valor Recuperar tamaño de lote recomendado es menor que Resultados máximos y si Recuperar tamaño de lote es mayor que Resultados máximos , los registros se procesan en un solo lote. Nota: Este campo solo se aplica a motores de búsqueda compatibles con clúster y no a motores de búsqueda de datos heredados.
Obtener recuento de cruces	Devuelve el número total de cruces realizados. Por ejemplo, si usa el valor predeterminado "10" para el campo Resultados máximos, solo se mostrarán 10 resultados. Sin embargo, si marca esta casilla, el campo de salida TotalMatchCount le indicará cuántos cruces se realizaron durante el procesamiento.
Tipo de búsqueda de índice	Determina el tipo de búsqueda de índice que desea realizar. Seleccione Búsqueda simple .
Campos del índice	Seleccione los campos de índice que desea usar para comparación en la búsqueda simple.
Campo de entrada	Seleccione el campo de entrada que desea usar para comparación en la búsqueda simple.

Nombre de la opción	Descripción/Valores válidos
---------------------	-----------------------------

Analizador de entrada	
-----------------------	--

Nombre de la opción	Descripción/Valores válidos
---------------------	-----------------------------

Especifique el analizador que se usará para la función Tokenize en la cadena de entrada. Puede tomar uno de los siguientes valores:

- Standard: Ofrece formación de muestras basada en gramática que contiene un superconjunto de analizadores de Whitespace y Stop Word. Comprende la puntuación del español para separar las palabras, sabe qué palabras ignorar (a través de Stop Word Analyzer) y realiza búsquedas que técnicamente no distinguen entre mayúsculas y minúsculas, con comparaciones de minúsculas. Por ejemplo, la cadena de caracteres "Pitney Bowes Software" se devolvería como tres muestras: "Pitney", "Bowes" y "Software".
- Whitespace: Separa las muestras con espacios en blanco. Es un subconjunto de Standard Analyzer en cuanto comprende los espacios entre palabras en español en textos basados en espacios y saltos de línea.
- StopWord: quita los artículos, como "la", "y" y "un/una" para reducir el tamaño del índice y aumentar el rendimiento.
- Palabra clave: crea una sola muestra desde un flujo de datos. Por ejemplo, la cadena de caracteres "Pitney Bowes Software" se devolvería como una muestra: "Pitney Bowes Software".
- Ruso: admite índices de idioma ruso y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "y", "yo" y "tú" para reducir el tamaño del índice y aumentar el rendimiento.
- Alemán: admite índices de idioma alemán y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "el", "y" y "un" para reducir el tamaño del índice y aumentar el rendimiento.
- Danés: admite índices de idioma danés y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "en", "y" y "un" para reducir el tamaño del índice y aumentar el rendimiento.
- Holandés: admite índices de idioma holandés y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "el", "y" y "un" para reducir el tamaño del índice y aumentar el rendimiento.
- Finlandés: admite índices de idioma finlandés y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "es", "y" y "de" para reducir el tamaño del índice y aumentar el rendimiento.
- Francés: admite índices de idioma francés y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "el", "y" y "un" para reducir el tamaño del índice y aumentar el rendimiento.
- Húngaro: admite índices de idioma húngaro y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "el", "y" y "un" para reducir el tamaño del índice y aumentar el rendimiento.
- Italiano: admite índices de idioma italiano y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "el", "y" y "un" para reducir el tamaño del índice y aumentar el rendimiento.
- Noruego: admite índices de idioma noruego y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "el", "y" y "un" para reducir el tamaño del índice y aumentar el rendimiento.
- Portugués: admite índices de idioma portugués y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "el", "y" y "un" para reducir el tamaño del índice y aumentar el rendimiento.

Nombre de la opción	Descripción/Valores válidos
	<ul style="list-style-type: none"> • Español: admite índices de idioma español y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "el", "y" y "un" para reducir el tamaño del índice y aumentar el rendimiento. • Sueco: admite índices de idioma sueco y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "el", "y" y "un" para reducir el tamaño del índice y aumentar el rendimiento. • Hindi: admite índices de idioma hindi y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "por", "y" y "un" para reducir el tamaño del índice y aumentar el rendimiento.
Pestaña Campos de salida	<p>Marque la casilla Incluir para seleccionar los campos almacenados que se deben incluir en el resultado.</p> <p>Nota: Si el campo de salida es de una etapa anterior en el flujo de datos y posee el mismo nombre que el del campo almacenado del índice de búsqueda, los valores del campo de entrada sobrescribirán los valores en el campo de salida.</p>

Opciones de índice de búsqueda avanzada

Tabla 11: Opciones de Candidate Finder

Nombre de la opción	Descripción/Valores válidos
Tipo de buscador	Seleccione Índice de búsqueda.
Nombre	Seleccione el índice correspondiente que se creó con la etapa Índice Write to Search en la página 235 en las etapas implementadas de Advanced Matching en Enterprise Designer.
Starting record (Registro de inicio)	Ingrese el número de registro donde deben comenzar los resultados de la búsqueda. El valor predeterminado es 1.
Resultados máximos	<p>Ingrese la cantidad máxima de respuestas que desea que arroje la búsqueda de índice. El valor predeterminado es 10.</p> <p>Nota: Si la cantidad de resultados máximos es demasiado alta, procese los resultados en lotes mediante el campo Tamaño de lotes para la recuperación.</p>

Nombre de la opción	Descripción/Valores válidos
Recuperar tamaño de lote	<p>Si el valor de Resultados máximos es demasiado alto, especifique el tamaño de lote que usará para procesar los resultados. Esto optimiza el procesamiento de cantidades altas de registros. El valor predeterminado es 10000.</p> <p>El valor Recuperar tamaño de lote recomendado es menor que Resultados máximos y si Recuperar tamaño de lote es mayor que Resultados máximos, los registros se procesan en un solo lote.</p> <p>Nota: Este campo solo se aplica a motores de búsqueda compatibles con clúster y no a motores de búsqueda de datos heredados.</p>
Ordenar	<p>Ordena los registros candidatos según los campos indexados mientras realiza una consulta de búsqueda.</p> <p>Seleccione la casilla Ordenar, haga clic en el campo de índice deseado en la lista desplegable Ordenar por y seleccione la opción Ascendente o Descendente desde la lista desplegable Organizar por.</p> <p>Nota: Solo puede ordenar en Tipo de campo de cadena de caracteres y Analizador de palabra clave.</p>
Obtener recuento de cruces	<p>Devuelve el número total de cruces realizados. Por ejemplo, si usa el valor predeterminado "10" para el campo Resultados máximos, solo se mostrarán 10 resultados. Sin embargo, si marca esta casilla, el campo de salida TotalMatchCount le indicará cuántos cruces se realizaron durante el procesamiento.</p>
Relevancia	<p>Controla la relevancia del Campo de índice.</p>
Tipo de búsqueda de índice	<p>Determina el tipo de búsqueda de índice que desea realizar. Seleccione Búsqueda avanzada.</p>
Botón Agregar elemento principal	<p>Opciones de Elemento principal de acceso.</p>
Opciones del elemento principal: Nombre	<p>Ingrese un nombre para el elemento principal.</p>

Nombre de la opción	Descripción/Valores válidos
Opciones del elemento principal: Método de búsqueda	<p>Especifique cómo determinar si un elemento principal presenta un cruce o no. Puede tomar uno de los siguientes valores:</p> <p>Todo verdadero: se considera que un elemento principal es un cruce si todos los elementos secundarios están determinados para el cruce. Este método crea un conector "AND" entre los elementos secundarios.</p> <p>Cualquier elemento verdadero: se considera que un elemento principal es un cruce si al menos uno de los elementos secundarios está determinado para el cruce. Este método crea un conector "OR" entre los elementos secundarios.</p> <p>Ningún elemento verdadero: se considera que un elemento principal es un cruce si ninguno de los elementos secundarios está determinado para el cruce. Este método crea un conector "NOT" entre los elementos secundarios.</p>
Botón Agregar elemento secundario	Opciones de Elemento secundario de acceso.
Opciones del elemento secundario: campo Índice	Seleccione el campo de índice que desea usar para comparación en la búsqueda avanzada.
Opciones del elemento secundario: tipo de búsqueda	Especifica los criterios de coincidencia/búsqueda que determinan si los datos de entrada se comparan/buscan con los datos indexados. Ninguna búsqueda distingue mayúsculas de minúsculas.
Opciones del elemento secundario: campo de entrada	Seleccione el campo de entrada que desea usar para comparación en la búsqueda avanzada.
Cualquier palabra o frase que comience con	<p>Determina si el texto incluido en el campo de índice de búsqueda comienza con el texto que se incluye en el campo de entrada.</p> <p>Por ejemplo, el texto en el campo de entrada "tech" se consideraría un cruce para los campos de índice de búsqueda que contengan "Technical", "Technology", "Technologies", "Technician" o incluso "National University of Technical Sciences". De igual forma, una frase en el campo de datos de entrada "DEF Sof" se consideraría una coincidencia para los campos de índice de búsqueda que contengan "ABC DEF Software", "DEF Software" y "DEF Software India", pero no sería una coincidencia para los campos de índice de búsqueda que contengan "Software DEF" o "DEF ABC Software".</p>

Nombre de la opción	Descripción/Valores válidos
Contiene	<p>Determina si el campo de índice de búsqueda contiene los datos del campo de entrada. Este tipo de búsqueda considera la secuencia de palabras en el campo de entrada durante la búsqueda en el campo de índice de búsqueda. Por ejemplo, los datos del campo de entrada "Pitney" y "Pitney Bowes" se incluirían en un campo de índice de búsqueda de "Pitney Bowes Software Inc."</p>
Contiene todo	<p>Determina si todas las palabras alfanuméricas del campo de entrada se incluyen en el campo de índice de búsqueda. Este tipo de búsqueda no considera la secuencia de palabras en el campo de entrada durante la búsqueda en el campo de índice de búsqueda.</p>
Contiene cualquiera	<p>Determina si cualquiera de las palabras alfanuméricas del campo de entrada se incluyen en el campo de índice de búsqueda.</p>
Contiene ninguno	<p>Determina si ninguna de las palabras alfanuméricas del campo de entrada se incluye en el campo de índice de búsqueda.</p>
Parcial	<p>Determina la similitud entre dos palabras alfanuméricas según el número de eliminaciones, inserciones o sustituciones que se requieren para transformar una palabra en otra.</p> <p>Use el parámetro Ediciones máximas para establecer un límite en la cantidad permitida de ediciones para que sean consideradas como un cruce exitoso:</p> <ul style="list-style-type: none"> • 0: no permite eliminaciones, inserciones o sustituciones. Los datos del campo de entrada y los del campo de índice deben ser idénticos. • 1: permite no más de una eliminación, inserción o sustitución. Por ejemplo, un campo de entrada que contenga "Barton" coincidirá con un campo de índice de búsqueda que contenga "Carton". • 2: permite no más de dos eliminaciones, inserciones o sustituciones. Por ejemplo, un campo de entrada que contenga "Barton" coincidirá con un campo de índice de búsqueda que contenga "Martin". <p>El tipo de búsqueda Aproximada se usa solamente para búsquedas de una sola palabra. Haga clic en Ignorar palabras adicionales para disponer que Candidate Finder solo considere la primera palabra en el campo, al comparar el campo de entrada con el campo de índice. Por ejemplo, si el campo de índice dice "Pitney" y el campo de entrada dice "Pitney Bowes", no se considerarían como un cruce debido a "Bowes". Sin embargo, si marca esta casilla, se ignoraría "Bowes" y con "Pitney" como la primera palabra, las dos palabras se considerarían como un cruce.</p>

Nombre de la opción	Descripción/Valores válidos
Numérico	<p>Determina si los números del campo de entrada se incluyen en el campo de índice de búsqueda.</p> <p>El tipo de búsqueda Numérico se usa solamente para búsquedas de una sola palabra.</p> <p>Haga clic en Ignorar palabras adicionales para disponer que Candidate Finder solo considere la primera palabra en el campo, al comparar el campo de entrada con el campo de índice.</p>
Patrón	<p>Determina si el patrón del texto del campo de entrada coincide con el patrón del texto de los criterios de búsqueda. Puede refinar adicionalmente el patrón del texto en el campo Cadena de patrón. Por ejemplo, si el campo de entrada contiene "nlm" y el patrón definido es "a*b?c", entonces coincidirá con las siguiente palabras "Neelam", "nelam", "neelum", "nilam", etc.</p> <p>El tipo de búsqueda Patrón se usa solamente para búsquedas de una sola palabra. Haga clic en Ignorar palabras adicionales para disponer que Candidate Finder solo considere la primera palabra en el campo, al comparar el campo de entrada con el campo de índice.</p>
Proximidad	<p>Determina si las palabras en los campos de entrada están dentro de una cierta distancia entre sí.</p> <ul style="list-style-type: none"> • Defina la entrada del campo Primera entrada y campo Segunda entrada que desee buscar en el índice. • Use el parámetro Distancia para determinar la distancia máxima permitida entre las palabras especificadas en el campo Primera y el campo Segunda, para que sean consideradas como un cruce <p>Por ejemplo, puede usar satisfactoriamente este tipo de búsqueda para buscar en el campo Primera "Spectrum" y campo Segunda "Pitney" dentro de diez palabras de cada uno, en un campo de índice de búsqueda que contenga la oración "Spectrum Technology Platform es un producto de Pitney Bowes Software Inc."</p> <p>El tipo de búsqueda Proximidad se usa solamente para búsquedas de una sola palabra. Haga clic en Ignorar palabras adicionales para disponer que Candidate Finder solo considere la primera palabra en el campo, al comparar el campo de entrada con el campo de índice.</p>

Nombre de la opción	Descripción/Valores válidos
Rango	<p>Realiza búsquedas inclusivas de términos dentro de un rango, que se especifica mediante el uso de un campo asociado Inferior (término de inicio) y un campo asociado Superior (término final). Todas las palabras alfanuméricas se organizan lexicográficamente en el campo de índice de búsqueda.</p> <ul style="list-style-type: none"> • Use el parámetro campo asociado Inferior para seleccionar el campo que se va a usar como término de inicio. • Use el parámetro campo asociado Superior para seleccionar el campo que se va a usar como término final. <p>Por ejemplo, si busca códigos postales desde 20001 (definido en el campo asociado Inferior) hasta 20009 (definido en el campo asociado Superior), la búsqueda arrojará todas las direcciones con códigos postales dentro de dicho rango.</p> <p>El tipo de búsqueda Rango se usa solamente para búsquedas de una sola palabra. Haga clic en Ignorar palabras adicionales para disponer que Candidate Finder solo considere la primera palabra en el campo, al comparar el campo de entrada con el campo de índice.</p>
Carácter comodín	<p>Realiza búsquedas usando uno o varios caracteres comodín.</p> <p>Seleccione la posición en el archivo de entrada donde va a insertar el carácter comodín.</p> <p>El tipo de búsqueda Comodín se usa solamente para búsquedas de una sola palabra. Haga clic en Ignorar palabras adicionales para disponer que Candidate Finder solo considere la primera palabra en el campo, al comparar el campo de entrada con el campo de índice.</p>
Opciones del elemento secundario: factor de relevancia	<p>Controle la relevancia de un campo secundario ingresando aquí cualquier número positivo hasta 100. El número puede ser incluso menor que "1". Por ejemplo, ".05" sería un número válido.</p> <p>Mientras más alto sea el factor de aumento, más relevante será el campo. Por ejemplo, si desea que los resultados del campo Nombre de la firma sean más relevantes que los resultados de los otros campos, seleccione "Nombre de la firma" en nombre del campo Índice e ingrese "5" aquí.</p> <p>Nota: De forma predeterminada, esta opción está desactivada. Seleccione la casilla de verificación para habilitar la opción.</p>
Ignorar espacios en blanco	<p>Desmarque esta casilla de verificación si desea que la consulta considere los campos de archivo de entrada en blanco.</p> <p>Nota: De forma predeterminada, la consulta ignora los campos en blanco.</p>

Nombre de la opción	Descripción/Valores válidos
Pestaña Campos de salida	<p>Marque la casilla Incluir para seleccionar los campos almacenados que se deben incluir en el resultado.</p> <p>Nota: Si el campo de salida es de una etapa anterior en el flujo de datos y posee el mismo nombre que el del campo almacenado del índice de búsqueda, los valores del campo de entrada sobrescribirán los valores en el campo de salida.</p>

Cómo configurar opciones en el tiempo de ejecución

Algunas opciones de Candidate Finder se pueden configurar y pasar en el tiempo de ejecución, si estas se exponen como opciones del flujo de datos. Esto le permite ejecutar su flujo de datos mientras usa diferentes configuraciones. Estas son las opciones de flujo de datos disponibles para Candidate Finder:

- **ConnectionString:** el nombre de la base de datos que contiene los registros de candidatos.
- **SearchIndexName:** el nombre del índice de búsqueda usado en el flujo de datos Candidate Finder.
- **StartingRecord:** el número de registro donde deben comenzar los resultados de la búsqueda.
- **MaximumResults:** la cantidad máxima de respuestas que desea que arroje la búsqueda de índice.
- **ReturnMatchCount:** el número total de coincidencias realizadas. Este campo es útil si ingresa un número más bajo en el campo MaximumResults, pero desea conocer la cantidad total de coincidencias realizadas.

Para definir las opciones de Candidate Finder en el tiempo de ejecución:

1. En Enterprise Designer, abra un flujo de datos que use la etapa Candidate Finder.
2. Guarde y exponga dicho flujo de datos.
3. Vaya a **Edit > Dataflow Options**.
4. En la tabla **Aplicar opciones de flujo de datos a etapas**, amplíe Candidate Finder y edite las opciones según sea necesario. Marque la casilla de la opción que desee editar y luego cambie el valor en la lista desplegable **Valor predeterminado**.
5. Opcional: cambie el nombre de las opciones en el campo **Rótulo de opción**.
6. Haga clic en **Aceptar** dos veces.

Salida

Tabla 12: Datos de salida de Candidate Finder

Nombre de campo	Formato	Descripción/Valores válidos
CandidateCount	Cadena	En este campo se indica el número total de candidatos devueltos durante el procesamiento.
CandidateGroup	Cadena	Este campo identifica una agrupación de un registro sospechoso y sus candidatos. Cada registro sospechoso recibe un número de CandidateGroup. Los candidatos para ese registro sospechoso reciben el mismo número de CandidateGroup. Por ejemplo, si John Smith es un registro sospechoso y sus registros candidatos son John Smith y Jon Smth, entonces los tres registros tendrán el mismo valor de CandidateGroup.
HasDuplicates	Cadena	Permite identificar si se detectan los candidatos o no. Los valores posibles son: Y: el registro es sospechoso y tiene candidatos N: el registro es sospechoso, pero no tiene candidatos D: el registro es un registro candidato
TotalMatchCount	Cadena	En este campo se indica el número total de cruces establecidos durante el procesamiento.
TransactionRecordType	Cadena	Una de las siguientes: Suspect Se utiliza un registro sospechoso como entrada a una consulta. Candidate Un registro candidato es un resultado devuelto de una consulta.

Duplicate Synchronization

Duplicate Synchronization (Sincronización duplicada) determina cuáles son los campos de una colección de registros que deben copiarse en los campos correspondientes de todos los registros de la colección. Usted puede especificar las reglas que deben cumplir los registros para copiar los

datos del campo en los demás registros de la colección Una vez que el proceso finaliza, todos los registros de la colección se conservan.

Opciones

En la siguiente tabla se enumeran las opciones para la etapa **Duplicate Synchronization**.

Nombre de la opción	Descripción / Valores válidos
Agrupar por	Especifica el campo que se usará para crear grupos de registros para sincronizar. En los casos donde haya usado una etapa de comparación tempranamente en el flujo de datos, como por ejemplo, Interflow Match, Intraflow Match o Transactional Match, debe seleccionar el campo CollectionNumber para usar las colecciones creadas por la etapa de comparación como los grupos. Sin embargo, si desea agrupar los recursos por otros campos, elija el campo aquí. Por ejemplo, si desea sincronizar registros que tienen el mismo valor en el campo AccountNumber, tendrá que seleccionar AccountNumber.
Ordenar	Si especifica un campo en el campo Agrupar por , marque esta casilla para ordenar los registros por valor en el campo que escoja. Esta opción está habilitada de forma predeterminada.

Nombre de la opción	Descripción / Valores válidos
---------------------	-------------------------------

Avanzada	
----------	--

Nombre de la opción	Descripción / Valores válidos
---------------------	-------------------------------

Haga clic en este botón para especificar las opciones de rendimiento de orden. De manera predeterminada, las opciones de rendimiento de orden que se especifican en Management Console, que son las opciones de rendimiento predeterminadas para su sistema, están activas. Si desea reemplazar las opciones de rendimiento predeterminadas del sistema, marque la casilla **Invaldar opciones de rendimiento de orden** y luego especifique en estos campos los valores que desea:

Límite de registros en memoria	Especifica la cantidad máxima de filas de datos que un ordenador guardará en la memoria antes de empezar la paginación al disco. De forma predeterminada, la ordenación de 10.000 registros o menos se realizará en la memoria y la de más de 10.000 registros se realizará como una ordenación en disco. El límite máximo es de 100.000 registros. Normalmente, la ordenación que se realiza en la memoria es más rápida que la que se realiza en disco; por lo tanto, este valor se debe establecer lo suficientemente alto para que la mayor parte de la ordenación sea en la memoria y solo los conjuntos de datos grandes se escriban en el disco.
---------------------------------------	---

Nota: Tenga cuidado en entornos en los que se ejecuten trabajos de forma concurrente, porque aumentar la configuración del **Límite de registros en memoria** aumenta las posibilidades de agotar la memoria.

Número máximo de archivos temporales	Especifica el número máximo de archivos temporales que puede utilizar el proceso de orden. El uso de una gran cantidad de archivos temporales puede producir un mejor rendimiento. Sin embargo, la cantidad óptima depende en gran medida de la configuración en el servidor que ejecuta Spectrum™ Technology Platform. Debe experimentar distintas configuraciones y observar el efecto en el rendimiento de usar una mayor o menor cantidad de archivos temporales. Para calcular la cantidad aproximada de archivos temporales que se pueda necesitar, use la siguiente ecuación:
---	--

$$InMemoryRecordLimit = \frac{(NumberOfRecords \times 2)}{NumberOfTempFiles}$$

Tenga en cuenta que la cantidad máxima de archivos temporales no puede ser superior a 1.000.

Activar compresión	Especifica que los archivos temporales se comprimen cuando se los escribe en el disco.
---------------------------	--

Nota: La configuración del rendimiento óptimo de orden depende de la configuración del hardware de su servidor. No obstante, la siguiente ecuación generalmente produce un buen rendimiento de orden:

$$(InMemoryRecordLimit \times MaxNumberOfTempFiles)$$

Nombre de la opción	Descripción / Valores válidos
---------------------	-------------------------------

	$\div 2) \geq TotalNumberOfRecords$
--	-------------------------------------

Reglas

Las reglas de Duplicate Synchronization determinan los registros cuyos datos deberán copiarse en todos los demás registros de la colección.

Para agregar una regla, seleccione Reglas en la jerarquía de reglas y haga clic en **Agregar regla**

Si especifica varias reglas, deberá seleccionar un operador lógico para usar entre cada regla. Seleccione **Y** si desea que pase tanto la regla nueva como la regla anterior, para que se cumpla la condición. Seleccione **O** si desea que pase la regla anterior o la regla siguiente, para que se cumpla la condición.

Opción	Descripción
Nombre de campo	Especifica el nombre del campo de flujo de datos cuyo valor desea evaluar, para determinar si se va a filtrar el registro.
Tipo de campo	Especifica el tipo de datos del campo. Una de las siguientes: <ul style="list-style-type: none"> No numérico Seleccione esta opción si el campo contiene datos no numéricos (por ejemplo, datos de cadena). Numérico Seleccione esta opción si el campo contiene datos numéricos (por ejemplo, doble, flotante, entre otros).

Opción	Descripción
Operador	Especifica el tipo de comparación que desea utilizar en la evaluación del campo. Una de las siguientes:
Contiene	Determina si el campo contiene el valor especificado. Por ejemplo, "sailboat" contiene el valor "boat".
IGUAL	Determina si el campo contiene el valor exacto especificado.
Mayor que	Determina si el valor de campo es mayor que el valor especificado. Esta operación solamente funciona con campos numéricos.
Mayor o igual que	Determina si el valor de campo es mayor o igual que el valor especificado. Esta operación solamente funciona con campos numéricos.
El más alto	Compara el valor del campo en todo el grupo de registros y determina qué registro tiene el valor más alto en el campo. Por ejemplo, si los campos del grupo contienen valores de 10, 20, 30 y 100, quedaría seleccionado el registro con el valor 100. Esta operación solamente funciona con campos numéricos. Si hay múltiples registros que poseen el valor más largo, se selecciona uno de ellos.
Está vacío	Determina si el campo no contiene ningún valor.
No está vacío	Determina si el campo contiene algún valor.
Menor que	Determina si el valor de campo es menor que el valor especificado. Esta operación solamente funciona con campos numéricos.
Menor o igual que	Determina si el valor de campo es menor o igual que el valor especificado. Esta operación solamente funciona con campos numéricos.
El más largo	Compara el valor del campo en todo el grupo de registros y determina qué registro tiene el valor más largo (en bytes) en el campo. Por ejemplo, si el grupo contiene los valores "Mike" y "Michael", quedaría seleccionado el registro con el valor "Michael". Si hay múltiples registros que poseen el valor más largo, se selecciona uno de ellos.
El más bajo	Compara el valor del campo en todo el grupo de registros y determina qué registro tiene el valor más bajo en el campo. Por ejemplo, si los campos del grupo contienen valores de 10, 20, 30 y 100, quedaría seleccionado el registro con el valor 10. Esta operación solamente funciona con campos numéricos. Si hay múltiples registros que poseen el valor más largo, se selecciona uno de ellos.
Más común	Determina si el valor de campo contiene el valor que aparece más frecuentemente en este campo entre los registros del grupo. Si hay dos o más valores que son los más comunes, no se realiza acción alguna.
Diferente	Determina si el valor de campo no es el mismo que el valor especificado.

Opción	Descripción
Tipo de valor	<p>Especifica el tipo de valor que desea comparar al valor del campo. Una de las siguientes:</p> <p>Nota: Esta opción no está disponible si selecciona el operador El más alto, El más bajo, o El más largo.</p> <p>Campo Seleccione esta opción si desea comparar con este campo el valor del campo de otro flujo de datos.</p> <p>Cadena Seleccione esta opción si desea comparar el campo con un valor específico.</p>
Valor	<p>Especifica el valor que se va a comparar con el valor del campo. Si seleccionó Campo en el campo Tipo de campo seleccione un campo de flujo de datos. Si seleccionó Cadena en el campo de Tipo de valor, ingrese el valor que desea usar para la comparación.</p> <p>Nota: Esta opción no está disponible si selecciona el operador El más alto, El más bajo, o El más largo.</p>

Acciones

Las acciones determinan qué campo se copiará en otros registros del grupo. Para agregar una acción, seleccione Acciones en el árbol de condición Duplicate Synchronization y luego haga clic en **Agregar acción**. Use las siguientes opciones para definir la acción.

Opción	Descripción
Tipo de origen	<p>Especifica el tipo de los datos que serán copiados a otros registros del grupo. Las opciones son:</p> <p>Campo Seleccione esta opción si desea copiar un valor de un campo en los demás registros del grupo.</p> <p>Cadena Seleccione esta opción si desea copiar un valor constante en los demás registros del grupo.</p>
Datos de origen	<p>Especifica los datos que se deben copiar en los demás registros del grupo. Si el tipo de origen es campo, seleccione el campo cuyos valores desea copiar en los demás registros del grupo. Si el tipo de origen es Cadena, especifique un valor constante para copiar en los demás registros del grupo.</p> <p>Nota: En caso de que los datos de origen tengan un valor nulo, este no se copiará a otros registros del grupo. Los demás registros conservarán sus valores originales.</p>

Opción	Descripción
Destino	Especifica el campo en los demás registros en los cuales desea copiar los datos especificados en el campo Datos de origen . Por ejemplo, si desea copiar los datos en el campo AccountBalance en todos los demás registros del grupo, deberá especificar AccountBalance.

Ejemplo de una regla y acción de Duplicate Synchronization

Esta regla y acción de Duplicate Synchronization selecciona el registro en el cual la calificación de cruce es 100 y copia el campo de número de cuenta AccountNumber en todos los demás registros del grupo.

Regla

Nombre del campo: MatchScore
 Tipo de campo: Numérico
 Operador: Igual
 Tipo de valor: cadena
 Valor: 100

Acción

Tipo de origen: Campo
 Datos de origen: AccountNumber
 Destino: NewAccountNumber

Filtro

La etapa Filter (Filtro) conserva o elimina registros de un grupo de registros según las reglas que usted especifique.

Opciones

La siguiente tabla enumera las opciones en la etapa Filter.

Nombre de la opción	Descripción / Valores válidos
Agrupar por	Especifica el campo que se usará para crear grupos de registros para filtrar. La etapa Filter conservará uno o más registros de cada grupo, dependiendo de la configuración de la etapa. En los casos donde haya usado una etapa de comparación tempranamente en el flujo de datos, como por ejemplo, Interflow Match, Intraflow Match o Transactional Match, debe seleccionar el campo CollectionNumber para usar las colecciones creadas por la etapa de comparación como los grupos. Sin embargo, si desea agrupar los recursos por otros campos, elija el campo aquí. Por ejemplo, si desea filtrar todos los registros que tengan el mismo valor en el campo AccountNumber, excepto uno, tendrá que seleccionar AccountNumber.
Ordenar	Si especifica un campo en el campo Agrupar por , marque esta casilla para ordenar los registros por valor en el campo que escoja. Esta opción está habilitada de forma predeterminada.

Nombre de la opción	Descripción / Valores válidos
---------------------	-------------------------------

Avanzada	
----------	--

Nombre de la opción	Descripción / Valores válidos
---------------------	-------------------------------

Haga clic en este botón para especificar las opciones de rendimiento de orden. De manera predeterminada, las opciones de rendimiento de orden que se especifican en Management Console, que son las opciones de rendimiento predeterminadas para su sistema, están activas. Si desea reemplazar las opciones de rendimiento predeterminadas del sistema, marque la casilla **Invaldar opciones de rendimiento de orden** y luego especifique en estos campos los valores que desea:

Límite de registros en memoria	<p>Especifica la cantidad máxima de filas de datos que un ordenador guardará en la memoria antes de empezar la paginación al disco. De forma predeterminada, la ordenación de 10.000 registros o menos se realizará en la memoria y la de más de 10.000 registros se realizará como una ordenación en disco. El límite máximo es de 100.000 registros. Normalmente, la ordenación que se realiza en la memoria es más rápida que la que se realiza en disco; por lo tanto, este valor se debe establecer lo suficientemente alto para que la mayor parte de la ordenación sea en la memoria y solo los conjuntos de datos grandes se escriban en el disco.</p>
---------------------------------------	--

Nota: Tenga cuidado en entornos en los que se ejecuten trabajos de forma concurrente, porque aumentar la configuración del **Límite de registros en memoria** aumenta las posibilidades de agotar la memoria.

Número máximo de archivos temporales	<p>Especifica el número máximo de archivos temporales que puede utilizar el proceso de orden. El uso de una gran cantidad de archivos temporales puede producir un mejor rendimiento. Sin embargo, la cantidad óptima depende en gran medida de la configuración en el servidor que ejecuta Spectrum™ Technology Platform. Debe experimentar distintas configuraciones y observar el efecto en el rendimiento de usar una mayor o menor cantidad de archivos temporales. Para calcular la cantidad aproximada de archivos temporales que se pueda necesitar, use la siguiente ecuación:</p>
---	---

$$\text{InMemoryRecordLimit} = \frac{(\text{NumberOfRecords} \times 2)}{\text{NumberOfTempFiles}}$$

Tenga en cuenta que la cantidad máxima de archivos temporales no puede ser superior a 1.000.

Activar compresión	<p>Especifica que los archivos temporales se comprimen cuando se los escribe en el disco.</p>
---------------------------	---

Nota: La configuración del rendimiento óptimo de orden depende de la configuración del hardware de su servidor. No obstante, la siguiente ecuación generalmente produce un buen rendimiento de orden:

$$\text{InMemoryRecordLimit} \times \text{MaxNumberOfTempFiles}$$

Nombre de la opción	Descripción / Valores válidos
---------------------	-------------------------------

$\div 2) \geq TotalNumberOfRecords$

<p>Limitar la cantidad de registros duplicados obtenidos</p>	<p>Especifica la cantidad máxima de registros que se devuelven de cada grupo. Si la opción está establecida en 1, se pueden definir reglas de filtrado para determinar qué registro de cada grupo se debe devolver. Si no se define regla alguna, el primer registro de cada colección se devuelve y los demás se descartan. En este modo, las reglas de filtrado definen los registros que se conservarán.</p> <p>Por ejemplo, si se define una regla en la cual se conserva el registro con el puntaje más alto de coincidencia de un grupo y la opción está establecida en 1, el registro con el puntaje más alto de coincidencia de cada grupo sobrevivirá y los demás registros del grupo se descartarán.</p> <p>Si la opción está establecida en un valor superior a uno, no se puede especificar reglas de filtrado.</p> <p>Nota: En caso de que ningún registro de la colección cumpla con los criterios definidos, no se devuelve ningún registro del grupo.</p>
<p>Eliminar duplicados de la colección</p>	<p>Especifica que se deben utilizar reglas de filtrado para determinar los registros se eliminarán de la colección. Los registros restantes de la colección se conservarán. Al seleccionar esta opción, debe definir una regla.</p> <p>Nota: Si el grupo contiene solo un registro, las reglas de filtrado se ignoran y el registro se conserva.</p>

Opciones de regla

Las reglas de filtrado determinan los registros de un grupo que se conservarán o eliminarán. Si selecciona la opción **Limitar la cantidad de registros duplicados**, las reglas determinan qué registros sobreviven el filtro. Si selecciona la opción **Eliminar los duplicados de colección**, las reglas determinan qué registros se eliminan del flujo de datos.

Para agregar una regla, seleccione Reglas en la jerarquía de reglas y haga clic en **Agregar regla**

Si especifica varias reglas, deberá seleccionar un operador lógico para usar entre cada regla. Seleccione **Y** si desea que pase tanto la regla nueva como la regla anterior, para que se cumpla la condición. Seleccione **O** si desea que pase la regla anterior o la regla siguiente, para que se cumpla la condición.

Nota: Solo puede existir una condición en una etapa Filter. Cuando selecciona Condición en la jerarquía de reglas, los botones aparecen atenuados.

Opción	Descripción
Nombre de campo	Especifica el nombre del campo de flujo de datos cuyo valor desea evaluar, para determinar si se va a filtrar el registro.
Tipo de campo	Especifica el tipo de datos del campo. Una de las siguientes: No numérico Seleccione esta opción si el campo contiene datos no numéricos (por ejemplo, datos de cadena). Númeroico Seleccione esta opción si el campo contiene datos numéricos (por ejemplo, doble, flotante, entre otros).

Opción	Descripción
Operador	Especifica el tipo de comparación que desea utilizar en la evaluación del campo. Una de las siguientes:
Contiene	Determina si el campo contiene el valor especificado. Por ejemplo, "sailboat" contiene el valor "boat".
IGUAL	Determina si el campo contiene el valor exacto especificado.
Mayor que	Determina si el valor de campo es mayor que el valor especificado. Esta operación solamente funciona con campos numéricos.
Mayor o igual que	Determina si el valor de campo es mayor o igual que el valor especificado. Esta operación solamente funciona con campos numéricos.
El más alto	Compara el valor del campo en todo el grupo de registros y determina qué registro tiene el valor más alto en el campo. Por ejemplo, si los campos del grupo contienen valores de 10, 20, 30 y 100, quedaría seleccionado el registro con el valor 100. Esta operación solamente funciona con campos numéricos. Si hay múltiples registros que poseen el valor más largo, se selecciona uno de ellos.
Está vacío	Determina si el campo no contiene ningún valor.
No está vacío	Determina si el campo contiene algún valor.
Menor que	Determina si el valor de campo es menor que el valor especificado. Esta operación solamente funciona con campos numéricos.
Menor o igual que	Determina si el valor de campo es menor o igual que el valor especificado. Esta operación solamente funciona con campos numéricos.
El más largo	Compara el valor del campo en todo el grupo de registros y determina qué registro tiene el valor más largo (en bytes) en el campo. Por ejemplo, si el grupo contiene los valores "Mike" y "Michael", quedaría seleccionado el registro con el valor "Michael". Si hay múltiples registros que poseen el valor más largo, se selecciona uno de ellos.
El más bajo	Compara el valor del campo en todo el grupo de registros y determina qué registro tiene el valor más bajo en el campo. Por ejemplo, si los campos del grupo contienen valores de 10, 20, 30 y 100, quedaría seleccionado el registro con el valor 10. Esta operación solamente funciona con campos numéricos. Si hay múltiples registros que poseen el valor más largo, se selecciona uno de ellos.
Más común	Determina si el valor de campo contiene el valor que aparece más frecuentemente en este campo entre los registros del grupo. Si hay dos o más valores que son los más comunes, no se realiza acción alguna.
Diferente	Determina si el valor de campo no es el mismo que el valor especificado.

Opción	Descripción
Tipo de valor	<p>Especifica el tipo de valor que desea comparar al valor del campo. Una de las siguientes:</p> <p>Nota: Esta opción no está disponible si selecciona el operador El más alto, El más bajo, o El más largo.</p> <p>Campo Seleccione esta opción si desea comparar con este campo el valor del campo de otro flujo de datos.</p> <p>Cadena Seleccione esta opción si desea comparar el campo con un valor específico.</p>
Valor	<p>Especifica el valor que se va a comparar con el valor del campo. Si seleccionó Campo en el campo Tipo de campo seleccione un campo de flujo de datos. Si seleccionó Cadena en el campo de Tipo de valor, ingrese el valor que desea usar para la comparación.</p> <p>Nota: Esta opción no está disponible si selecciona el operador El más alto, El más bajo, o El más largo.</p>

Ejemplo de una regla de filtrado

Esta regla conserva el registro de cada grupo que posee el mayor valor en el campo MatchScore. Cabe destacar que las opciones **Valor** y **Tipo de valor** no se aplican cuando el operador es el más alto o el más bajo.

Nombre del campo = MatchScore
 Tipo de campo = Numérico
 Operador = Mayor

Esta regla conserva el registro en los que el valor en AccountNumber es "12345".

Nombre del campo = AccountNumber
 Tipo de campo = Numérico
 Operador = Igual
 Tipo de valor = Cadena
 Valor = 12345

Interflow Match

Interflow Match (Cruce de interflujo) localiza los cruces (coincidencias) entre registros de datos similares de dos flujos de registros de entrada. El primer flujo de registros es una fuente de origen de registros sospechosos, mientras que el segundo es una fuente de origen de registros candidatos.

Mediante el criterio de grupo de cruce (por ejemplo, una clave de cruce), Interflow Match identifica un grupo de registros que son posibles duplicados de un registro sospechoso en particular.

Cada candidato se cruza por separado con el registro sospechoso y se califica de acuerdo con las reglas de cruce definidas. Si el candidato es un duplicado, se le asigna un número de colección y el tipo de registro de cruce se etiqueta como duplicado y se copia. Los candidatos únicos sin coincidencias pueden copiarse o no según la preferencia del usuario. Una vez que Interflow Match finaliza con todos los registros candidatos en el actual grupo de cruce, al registro sospechoso cruzado (con coincidencia) se le asigna un número de colección que corresponde a su registro duplicado. O bien, si no se identifican coincidencias en el cruce, al registro sospechoso se le asigna el número de colección 0 y se lo etiqueta como registro único.

Nota: Interflow Match solo cruza registros sospechosos con registros candidatos. No intenta cruzar registros sospechosos con otros registros sospechosos, como se hace en Intraflow Match.

El proceso de comparación para un registro sospechoso en particular puede cancelarse antes de cruzar todos los posibles candidatos si se ha definido un limitador para los duplicados y se supera el límite para el registro sospechoso actual.

El tipo de comparación (Intraflow or Interflow, intraflujo o interflujo) determina la forma en que los resultados del cruce de clave inmediato se reflejan en las calificaciones de cruce de candidatos. En la comparación de interflujo, un cruce de clave inmediato que se realiza con éxito siempre confiere una calificación de cruce (MatchScore) de 100 al candidato. Por otra parte, en la comparación de intraflujo, la calificación obtenida por un registro candidato como resultado de un cruce de clave inmediato dependerá de la posibilidad de que el registro cruzado con el candidato sea un cruce (coincidencia) de algún otro registro sospechoso: los duplicados de un registro sospechoso detectados por medio del cruce de clave inmediato siempre tendrán una calificación de cruce (MatchScore) de 100, en tanto los duplicados de cruce de clave inmediato de otro candidato (que era un duplicado de un registro sospechoso) heredarán la calificación de cruce (no necesariamente 100) de ese candidato.

Opciones

1. En el campo **Cargar regla de cruce** seleccione una de las reglas de cruce predefinidas, que puede usar como está o bien puede modificarla para que se ajuste a sus necesidades. Si desea crear una nueva regla de cruce sin usar una de las reglas de cruce predefinidas como punto de partida, haga clic en **Nueva**. Solamente puede tener una regla personalizada en un flujo de datos.

Nota: La función Opciones del flujo de datos en Enterprise Designer permite exponer la regla de cruce para la configuración durante la ejecución.

2. Haga clic en **Agrupar por** para seleccionar el campo que quiere utilizar para agrupar registros en la cola de cruce. Intraflow Match solamente intenta establecer cruces entre registros de la misma cola de cruce.
3. Seleccione la casilla **Ordenar** para ordenar sus datos de entrada antes del cruce en el campo seleccionado en el campo **Agrupar por**.
4. Haga clic en **Opciones avanzadas** para especificar opciones de rendimiento de orden adicionales.

Límite de registros en memoria Especifica la cantidad máxima de filas de datos que un ordenador guardará en la memoria antes de empezar la paginación al disco. De forma predeterminada, la ordenación de 10.000 registros o menos se realizará en la memoria y la de más de 10.000 registros se realizará como una ordenación en disco. El límite máximo es de 100.000 registros. Normalmente, la ordenación que se realiza en la memoria es más rápida que la que se realiza en disco; por lo tanto, este valor se debe establecer lo suficientemente alto para que la mayor parte de la ordenación sea en la memoria y solo los conjuntos de datos grandes se escriban en el disco.

Nota: Tenga cuidado en entornos en los que se ejecuten trabajos de forma concurrente, porque aumentar la configuración del **Límite de registros en memoria** aumenta las posibilidades de agotar la memoria.

Número máximo de archivos temporales Especifica el número máximo de archivos temporales que puede utilizar el proceso de orden. El uso de una gran cantidad de archivos temporales puede producir un mejor rendimiento. Sin embargo, la cantidad óptima depende en gran medida de la configuración en el servidor que ejecuta Spectrum™ Technology Platform. Debe experimentar distintas configuraciones y observar el efecto en el rendimiento de usar una mayor o menor cantidad de archivos temporales. Para calcular la cantidad aproximada de archivos temporales que se pueda necesitar, use la siguiente ecuación:

$$(NumberOfRecords \times 2) \div InMemoryRecordLimit = NumberOfTempFiles$$

Tenga en cuenta que la cantidad máxima de archivos temporales no puede ser superior a 1.000.

Activar compresión Especifica que los archivos temporales se comprimen cuando se los escribe en el disco.

Nota: La configuración del rendimiento óptimo de orden depende de la configuración del hardware de su servidor. No obstante, la siguiente ecuación generalmente produce un buen rendimiento de orden:

$$(InMemoryRecordLimit \times MaxNumberOfTempFiles \div 2) \geq TotalNumberOfRecords$$

- Haga clic en **Cruce inmediato en** para realizar una comparación inicial de los valores de clave inmediata y determinar si se considerará que dos registros presentan un cruce.

El cruce de clave inmediato puede ser una herramienta útil para reducir el número de comparaciones realizadas y, de ese modo, mejorar la velocidad de ejecución. Una clave inmediata imprecisa genera muchos falsos positivos en el cruce. Usted puede generar una

clave inmediata como parte de la creación de una clave de cruce por medio de MatchKeyGenerator. Consulte [Match Key Generator](#) en la página 224 para obtener más información.

Si dos registros presentan un cruce exacto en la clave inmediata, el candidato se considera un duplicado 100%. Si dos registros no presentan un cruce en el valor de clave inmediata, se los compara utilizando el método basado en reglas.

Para determinar si un candidato presenta un cruce utilizando una clave inmediata, observe el valor del campo **ExpressKeyIdentified** que será Y si hay cruce y N si no. Cabe destacar que los registros sospechosos siempre tienen un valor de N en **ExpressKeyIdentified**.

- En el cuadro de texto **Número de colección inicial** especifique el número inicial que debe asignarse al campo de número de colección para los registros duplicados.

El número de colección identifica cada registro duplicado en una cola de cruce. Los registros únicos reciben un número de colección de 0. Cada registro duplicado recibe un número de colección a partir del valor especificado en el cuadro de texto **Número de colección inicial**.

- Seleccione una de las siguientes opciones:

Opción	Descripción
Comparar el registro sospechoso con todos los candidatos	<p>Esta opción compara el registro sospechoso con todos los candidatos en el mismo grupo de cruce (agrupar según opción) aún cuando ya se haya encontrado un duplicado en el grupo de cruce. Por ejemplo:</p> <p>Sospechoso- John Smith Candidato - Bill Jones Candidato - John Smith Candidato - John Smith</p> <p>En el ejemplo, el registro sospechoso John Smith se comparará con ambos candidatos John Smith.</p> <p>Marque la casilla Devolver candidatos únicos para devolver los registros dentro de un grupo de cruce del puerto candidato que hayan sido identificados como registros únicos.</p>
Detener cruce de sospechosos con candidatos después de encontrar <i>n</i> duplicados	<p>Esta opción compara el registro sospechoso con todos los candidatos del mismo grupo de cruce (opción agrupar por), pero detiene el cruce cuando se identifica la cantidad de duplicados definida por el usuario. Por ejemplo, si decide detener la comparación de candidatos luego de encontrar un duplicado y cuenta con los siguientes datos:</p> <p>Sospechoso- John Smith Candidato - Bill Jones</p>

Opción	Descripción
	Candidato - John Smith Candidato - John Smith
	En el ejemplo, el registro sospechoso John Smith detendría la comparación dentro del grupo de cruce al identificarse como duplicado el primer candidato John Smith.

8. Haga clic en **Generar datos para análisis** para generar resultados de cruce. Para obtener más información, consulte [Análisis de resultados de cruce](#) en la página 119.
9. **Al asignar el número de colección 0 a los registros únicos**, marcados de forma predeterminada, se asignarán ceros como números de colección a los registros únicos. Desmarque esta opción para generar números de colección distintos a cero para los registros únicos. Los números de colección de los registros únicos estarán en secuencia con cualquiera de los otros números de colección. Por ejemplo, si su flujo de datos de cruce encuentra cinco registros y los tres primeros registros son únicos, los números de colección se asignarán como se muestra en el primer grupo a continuación. Si su flujo de datos de cruce encuentra cinco registros y los dos últimos son únicos, los números de colección se asignarán como se muestra en el segundo grupo a continuación.

Opción	Descripción
Número de colección	Tipo de registro
1	Único
2	Único
3	Único
4	Duplicado/sospechoso
4	Duplicado/sospechoso
Número de colección	Tipo de registro
1	Duplicado/sospechoso
1	Duplicado/sospechoso

Opción	Descripción
2	Único
3	Único
4	Único

Si deja marcada esta opción, a todos los registros únicos que se encuentren en su flujo de datos se les asignará predeterminadamente un número de colección de cero.

10. Si crea una nueva regla de comparación personalizada, consulte [Generación de reglas de cruce](#) en la página 72 para obtener más información.
11. Haga clic en **Evaluar** para evaluar qué calificación obtuvo el registro sospechoso en comparación con los registros candidatos. Para obtener más información, consulte [Interflow Match](#) en la página 213.

Salida

Tabla 13: Campos de salida de Interflow Match

Nombre de campo	Descripción / Valores válidos
CollectionNumber	Identifica una colección de registros duplicados. Los valores posibles son 1 o mayor que 1.
ExpressMatchIdentified	Indica si la coincidencia se obtuvo utilizando la clave de cruce inmediato. Los valores posibles son Sí o No.
HasDuplicates	Identifica si el registro es un duplicado de otro registro. Las opciones son: Y El registro es un registro sospechoso y tiene duplicados. N El registro es un registro sospechoso y no tiene duplicados. D El registro es un candidato y es un duplicado del registro sospechoso. U El registro es un candidato, pero no es un duplicado del registro sospechoso.
InterflowSourceType	Los valores posibles son input_port_0 o input_port_1

Nombre de campo	Descripción / Valores válidos
MatchRecordType	<p>Identifica el tipo de registro de cruce de una colección. Los valores posibles son:</p> <p>suspect El registro de entrada original que fue marcado como un registro que posiblemente tuviese registros duplicados.</p> <p>duplicate Registro que es un duplicado del registro de entrada.</p> <p>unique Registro que no tiene duplicados.</p>
MatchScore	<p>Identifica la calificación general entre dos registros. Los valores posibles son 0-100; 0 indica una coincidencia poco satisfactoria y 100, una coincidencia exacta.</p>

Nota: Tanto las etapas de Validate Address como las del módulo Advanced Matching utilizan el campo MatchScore. El valor del campo MatchScore en la salida de un flujo de datos es determinado por la última etapa que modifica el valor antes de enviarlo a la etapa de salida. Si el flujo de datos contiene etapas de Validate Address y del módulo Advanced Matching y desea ver los datos de salida del campo MatchScore para cada etapa, utilice una etapa Transformer para copiar el valor de MatchScore en otro campo. Por ejemplo, Validate Address genera un campo de salida denominado MatchScore y luego una etapa Transformer puede copiar el campo MatchScore desde Validate Address a un campo denominado AddressMatchScore. Al ejecutarse, la etapa de cruce completa el campo MatchScore con el valor proveniente de la etapa y transmite el valor del campo AddressMatchScore a través de la etapa Validate Address.

Intraflow Match

Intraflow Match (Cruce de intraflujo) localiza los cruces (coincidencias) entre registros de datos similares en un mismo flujo de entrada. Puede crear reglas jerárquicas basadas en cualquier campo que se haya definido o creado en otras etapas en el flujo de datos.

Opciones

1. En el campo **Cargar regla de cruce** seleccione una de las reglas de cruce predefinidas, que puede usar como está o bien puede modificarla para que se ajuste a sus necesidades. Si desea crear una nueva regla de cruce sin usar una de las reglas de cruce predefinidas como punto de partida, haga clic en **Nueva**. Solamente puede tener una regla personalizada en un flujo de datos.

Nota: La función Opciones del flujo de datos en Enterprise Designer permite exponer la regla de cruce para la configuración durante la ejecución.

2. Haga clic en **Agrupar por** para seleccionar el campo que quiere utilizar para agrupar registros en la cola de cruce. Intraflow Match solamente intenta establecer cruces entre registros de la misma cola de cruce.
3. Seleccione la casilla **Ordenar** para ordenar sus datos de entrada antes del cruce en el campo seleccionado en el campo **Agrupar por**.
4. Haga clic en **Opciones avanzadas** para especificar opciones de rendimiento de orden adicionales.

Límite de registros en memoria

Especifica la cantidad máxima de filas de datos que un ordenador guardará en la memoria antes de empezar la paginación al disco. De forma predeterminada, la ordenación de 10.000 registros o menos se realizará en la memoria y la de más de 10.000 registros se realizará como una ordenación en disco. El límite máximo es de 100.000 registros. Normalmente, la ordenación que se realiza en la memoria es más rápida que la que se realiza en disco; por lo tanto, este valor se debe establecer lo suficientemente alto para que la mayor parte de la ordenación sea en la memoria y solo los conjuntos de datos grandes se escriban en el disco.

Nota: Tenga cuidado en entornos en los que se ejecuten trabajos de forma concurrente, porque aumentar la configuración del **Límite de registros en memoria** aumenta las posibilidades de agotar la memoria.

Número máximo de archivos temporales

Especifica el número máximo de archivos temporales que puede utilizar el proceso de orden. El uso de una gran cantidad de archivos temporales puede producir un mejor rendimiento. Sin embargo, la cantidad óptima depende en gran medida de la configuración en el servidor que ejecuta Spectrum™ Technology Platform. Debe experimentar distintas configuraciones y observar el efecto en el rendimiento de usar una mayor o menor cantidad de archivos temporales. Para calcular la cantidad aproximada de archivos temporales que se pueda necesitar, use la siguiente ecuación:

$$\frac{(NumberOfRecords \times 2)}{InMemoryRecordLimit} = NumberOfTempFiles$$

Tenga en cuenta que la cantidad máxima de archivos temporales no puede ser superior a 1.000.

Activar compresión

Especifica que los archivos temporales se comprimen cuando se los escribe en el disco.

Nota: La configuración del rendimiento óptimo de orden depende de la configuración del hardware de su servidor. No obstante, la siguiente ecuación generalmente produce un buen rendimiento de orden:

$$(InMemoryRecordLimit \times MaxNumberOfTempFiles \div 2) \geq TotalNumberOfRecords$$

5. Haga clic en **Cruce inmediato en** para realizar una comparación inicial de los valores de clave inmediata y determinar si se considerará que dos registros presentan un cruce.
Usted puede generar una clave inmediata como parte de la creación de una clave de cruce por medio de MatchKeyGenerator. Consulte [Match Key Generator](#) en la página 224 para obtener más información.
6. En el cuadro de texto **Número de colección inicial** especifique el número inicial que debe asignarse al campo de número de colección para los registros duplicados.
El número de colección identifica cada registro duplicado en una cola de cruce. Los registros únicos reciben un número de colección de 0. Cada registro duplicado recibe un número de colección a partir del valor especificado en el cuadro de texto **Número de colección inicial**.
7. Haga clic en **Ventana dinámica** para activar este método de cruce. Para obtener más información acerca de la Ventana dinámica, consulte [Método de comparación de ventana dinámica](#) en la página 223
8. Haga clic en **Generar datos para análisis** para generar resultados de cruce. Para obtener más información, consulte [Análisis de resultados de cruce](#) en la página 119.
9. **Al asignar el número de colección 0 a los registros únicos**, marcados de forma predeterminada, se asignarán ceros como números de colección a los registros únicos. Desmarque esta opción para generar números de colección distintos a cero para los registros únicos. Los números de colección de los registros únicos estarán en secuencia con cualquiera de los otros números de colección. Por ejemplo, si su flujo de datos de cruce encuentra cinco registros y los tres primeros registros son únicos, los números de colección se asignarán como se muestra en el primer grupo a continuación. Si su flujo de datos de cruce encuentra cinco registros y los dos últimos son únicos, los números de colección se asignarán como se muestra en el segundo grupo a continuación.

Opción	Descripción
Número de colección	Tipo de registro
1	Único
2	Único
3	Único

Opción	Descripción
4	Duplicado/sospechoso
4	Duplicado/sospechoso
Número de colección	Tipo de registro
1	Duplicado/sospechoso
1	Duplicado/sospechoso
2	Único
3	Único
4	Único

Si deja marcada esta opción, a todos los registros únicos que se encuentren en su flujo de datos se les asignará predeterminadamente un número de colección de cero.

- Para obtener más información acerca de cómo modificar las otras opciones, consulte [Generación de reglas de cruce](#) en la página 72.
- Haga clic en **Evaluar** para evaluar qué calificación obtuvo el registro sospechoso en comparación con los registros candidatos. Para obtener más información, consulte [Interflow Match](#) en la página 213.

Método de comparación predeterminado

Al utilizar la función Agrupar por (grupo de cruce) configurada por el usuario, la función de cruce identifica grupos de registro que podrían ser duplicados entre sí. Luego, la función de cruce analiza cada registro del grupo; si un registro coincide con un registro Sospechoso existente, se lo considera un Duplicado de ese sospechoso, se le asigna una Calificación, un CollectionNumber (Número de colección) y un MatchRecordType (Tipo de registro de cruce) (Duplicado), y se lo elimina del cruce. Por el contrario, si el registro no coincide con un Sospechoso existente dentro del grupo de cruce, el registro se convierte en un nuevo Sospechoso, porque se lo agrega al actual Grupo de cruce para que pueda ser comparado con los próximos registros. Cuando la función de cruce ha analizado todos los registros del actual Grupo de cruce, elimina todos los Sospechosos del cruce, rotula el tipo de Registro de cruce como Único y asigna el número de colección 0. Aquel Sospechoso que tiene, al menos, un duplicado mantiene el tipo de Registro de cruce Sospechoso y recibe el mismo número de colección que el registro duplicado que coincide con él. Finalmente, cuando todos los

registros dentro de un grupo de cruce han sido escritos en la salida, Un nuevo grupo de cruce se compara.

Nota: El Método de comparación predeterminado solo compara registros que están dentro del mismo grupo de cruce.

El tipo de comparación (Intraflow or Interflow, intraflujo o interflujo) determina la forma en que los resultados del cruce de clave inmediato se reflejan en las calificaciones de cruce de candidatos. En la comparación de interflujo, un cruce de clave inmediato que se realiza con éxito siempre confiere una calificación de cruce (MatchScore) de 100 al candidato. Por otra parte, en la comparación de intraflujo, la calificación obtenida por un registro candidato como resultado de un cruce de clave inmediato dependerá de la posibilidad de que el registro cruzado con el candidato sea un cruce (coincidencia) de algún otro registro sospechoso: los duplicados de un registro sospechoso detectados por medio del cruce de clave inmediato siempre tendrán una calificación de cruce (MatchScore) de 100, en tanto los duplicados de cruce de clave inmediato de otro candidato (que era un duplicado de un registro sospechoso) heredarán la calificación de cruce (no necesariamente 100) de ese candidato.

Método de comparación de ventana dinámica

El algoritmo de ventana dinámica es un algoritmo que llena de forma secuencial el tamaño de un área de influencia predeterminada denominada ventana con la correspondiente cantidad de filas de datos. A medida que cada fila se agrega a la ventana, se la compara con cada elemento que se encuentra en la ventana. Si se determina un cruce con un elemento, tanto el registro de controlador (el nuevo elemento por agregar a la ventana) como los candidatos (elementos que ya están en la ventana) reciben la misma ID de grupo. Esta comparación continúa hasta que el registro de controlador haya sido comparado con todos los elementos que se encuentran en la ventana.

A medida que se agregan los controladores nuevos, la ventana estará más cerca de alcanzar su capacidad predeterminada. En este punto, la ventana se deslizará, por ello se la denomina Ventana dinámica. Se dice que la ventana es dinámica simplemente porque el área de influencia de la ventana eliminará y escribirá el elemento más viejo en la ventana a medida que agregue el registro de controlador más nuevo.

Salida

Tabla 14: Datos de salida de Intraflow Match

Nombre de campo	Descripción / Valores válidos
CollectionNumber	Identifica una colección de registros duplicados. Los valores posibles son 1 o mayor que 1.

Nombre de campo	Descripción / Valores válidos
ExpressMatchIdentified	Indica si la coincidencia se obtuvo utilizando la clave de cruce inmediato. Los valores posibles son Sí o No.
MatchRecordType	Identifica el tipo de registro de cruce de una colección. Los valores posibles son: <ul style="list-style-type: none"> suspect Registro con el que se comparan otros registros para determinar si son duplicados entre sí. Cada colección tiene un solo registro sospechoso. duplicate Registro que es un duplicado del registro sospechoso. unique Registro que no tiene duplicados.
MatchScore	Identifica la calificación general entre dos registros. Los valores posibles son 0-100; 0 indica una coincidencia poco satisfactoria y 100, una coincidencia exacta.

Nota: Tanto las etapas de Validate Address como las del módulo Advanced Matching utilizan el campo MatchScore. El valor del campo MatchScore en la salida de un flujo de datos es determinado por la última etapa que modifica el valor antes de enviarlo a la etapa de salida. Si el flujo de datos contiene etapas de Validate Address y del módulo Advanced Matching y desea ver los datos de salida del campo MatchScore para cada etapa, utilice una etapa Transformer para copiar el valor de MatchScore en otro campo. Por ejemplo, Validate Address genera un campo de salida denominado MatchScore y luego una etapa Transformer puede copiar el campo MatchScore desde Validate Address a un campo denominado AddressMatchScore. Al ejecutarse, la etapa de cruce completa el campo MatchScore con el valor proveniente de la etapa y transmite el valor del campo AddressMatchScore a través de la etapa Validate Address.

Match Key Generator

Match Key Generator crea una clave no exclusiva para cada registro, que posteriormente se puede usar en las etapas de comparación a fin de identificar grupos de registros potencialmente duplicados. Las claves de cruce facilitan el proceso de comparación al permitir la agrupación de registros por clave de cruce y posteriormente solo comparando los registros al interior de estos grupos.

La clave de cruce se crea por medio de reglas definidas por el usuario y se compone de los campos de entrada. Cada campo de entrada especificado cuenta con un algoritmo seleccionado que se ejecuta para el campo. A continuación, el resultado de cada algoritmo es concatenado para crear un único campo de clave de cruce.

Además de crear claves de cruce, también puede crear claves de cruce inmediato, las que posteriormente usará una etapa Intraflow Match o Interflow Match en el flujo de datos.

Puede crear varias claves de cruce y claves de cruce inmediato.

Por ejemplo, si el registro entrante es:

Primer nombre: Fred

Apellido: Mertz

Código postal: 21114-1687

Código de género: M

Y define una regla de clave de cruce que genera una clave de cruce al combinar los datos del registro, de la siguiente manera:

Campo de entrada	Posición de inicio	Longitud
Código postal	1	5
Código postal	7	4
Apellido	1	5
Nombre	1	5
Código de género	1	1

Entonces la clave será:

211141687MertzFredM

Input

La entrada es cualquier campo en los datos de origen.

Opciones

Para definir las opciones de Match Key Generator (Generador de clave de cruce), haga clic en el botón **Agregar**. Se muestra el cuadro de diálogo **Campo de clave de cruce**.

Nota: La función Opciones de flujo de datos de Enterprise Designer permite que el generador de clave de cruce (Match Key Generator) esté expuesto para ser establecido al modo de tiempo de ejecución.

Tabla 15: Opciones de Match Key Generator

Nombre de la opción	Descripción / Valores válidos
---------------------	-------------------------------

Algoritmo	
-----------	--

Nombre de la opción

Descripción / Valores válidos

Especifica el algoritmo que se utilizará para generar la clave de cruce. Las opciones son:

- Consonante** Devuelve campo especificados con consonantes eliminadas.
- Doble metaphone** Devuelve un código basado en una representación fonética de sus caracteres. Double Metaphone es una versión mejorada del algoritmo Metaphone e intenta explicar la gran cantidad de irregularidades encontradas en los distintos idiomas.
- Koeln** Indexa los nombres por sonido, tal como se los pronuncia en alemán. Permite que los nombres que tienen la misma pronunciación se codifiquen con la misma representación, de modo tal que puedan cruzarse a pesar de las pequeñas diferencias en el modo en que se escriben. El resultado es siempre una secuencia de números; los caracteres especiales y los espacios en blanco se ignoran. Esta opción se desarrolló para responder a las limitaciones de Soundex.
- MD5** Un algoritmo de síntesis del mensaje que produce un valor hash de 128 bits. Este algoritmo suele usarse para comprobar la integridad de los datos.
- Metaphone** Devuelve una clave codificada Metaphone de campos seleccionados. Metaphone es un algoritmo para codificar palabras utilizando su pronunciación del inglés.
- Metaphone (español)** Devuelve una clave codificada Metaphone de campos seleccionados para el idioma español. El algoritmo Metaphone codifica palabras con base en su pronunciación en español.
- Metaphone 3** Mejora los algoritmos de Metaphone y Double Metaphone con parámetros más exactos de consonantes y vocales internas que permiten pronunciar palabras y nombres que, fonéticamente, coinciden en mayor o menor medida con los términos de búsqueda. Metaphone 3 aumenta la precisión de la codificación fonética en un 98%. Esta opción se desarrolló para responder a las limitaciones de Soundex.
- Nysiis** Algoritmo de codificación fonética que cruza una pronunciación aproximada con una palabra escrita exacta e indexa palabras que se pronuncian de manera similar. Parte del sistema de inteligencia e identificación del estado de Nueva York (New York State Identification and Intelligence System). Supongamos que está buscando información sobre alguien en una base de datos de personas. Cree que el nombre de esa persona suena algo así como "John Smith", pero en realidad se escribe "Jon Smyth". Si busca una coincidencia exacta para "John Smith" no obtendrá resultados. Sin embargo, si indexa la base de datos con el algoritmo de NYSIIS y realiza la búsqueda usando ese mismo algoritmo, obtendrá una coincidencia correcta porque el algoritmo indexa "John Smith" y "Jon Smyth" como "JAN SNATH".

Nombre de la opción	Descripción / Valores válidos
	<p>Phonix Preprocesa cadenas de nombres aplicando más de 100 reglas de transformación a caracteres únicos o secuencias de varios caracteres. Diecinueve de esas reglas se aplican solo si los caracteres se encuentran al comienzo de la cadena, mientras que 12 de ellas se aplican solo cuando los caracteres se encuentran en medio de la cadena, y 28 se aplican únicamente si están al final de la cadena. La cadena de nombre transformada se codifica en un código compuesto por la primera letra seguida de tres dígitos (sin ceros ni números duplicados). Esta opción se desarrolló como respuesta a las limitaciones de Soundex; es más compleja y, por lo tanto, más lenta que esta última.</p> <p>Soundex Devuelve un código Soundex de campos seleccionados. El algoritmo Soundex produce un código de longitud fija basado en la pronunciación en inglés de las palabras.</p> <p>Subcadena de caracteres Devuelve una porción específica del campo seleccionado.</p>
Nombre de campo	Especifica el campo en el que desea aplicar el algoritmo seleccionado para generar la clave de cruce. Por ejemplo, si selecciona un campo denominado LastName y elige el algoritmo Soundex, se aplicará ese algoritmo a los datos del campo LastName para generar una clave de cruce.
Posición de inicio	Especifica la posición de inicio dentro del campo especificado. No todos los algoritmos permiten especificar una posición inicial.
Longitud	Especifica la longitud, en caracteres, por incluirse desde la posición inicial. No todos los algoritmos permiten especificar una longitud.
Quitar caracteres irrelevantes	Elimina los caracteres que no son numéricos ni alfabéticos, como los guiones, los espacios en blanco y otros caracteres especiales de un campo de entrada.
Ordenar entrada	<p>Ordena alfabéticamente todos los caracteres de un campo de entrada o todos los términos de un campo de entrada.</p> <p>Caracteres Ordena los valores de los caracteres de un campo de entrada antes de crear una ID única:</p> <p>Términos Ordena el valor de cada término de un campo de entrada antes de crear una ID única.</p>

Si agrega muchos algoritmos de generación de claves de cruce, puede utilizar los botones **Subir** y **Bajar** para cambiar el orden en el que se aplican los algoritmos.

Generación de claves de cruce inmediato

Habilite la opción **Generar clave de cruce inmediato** y haga clic en **Agregar** para definir una clave de cruce inmediato que se utilizará en una etapa de cruce de intraflujo (Intraflow Match) o en una etapa de cruce de interflujo (Interflow Match), más adelante en el flujo de datos.

Si en una etapa de cruce de intraflujo (Intraflow Match) o en una etapa de cruce de interflujo (Interflow Match) están habilitadas las opciones **Generar clave de cruce inmediato** y **Cruce inmediato en**, el intento de cruce se hace primero con la clave de cruce inmediato creada aquí. Si coinciden las claves de cruce inmediato de dos registros, entonces se considera que el registro es un cruce y no se intenta realizar ningún procesamiento adicional. Si las claves de cruce inmediato de dos registros no coinciden, entonces se utilizan las reglas de cruce definidas en el cruce de intraflujo (Intraflow Match) o en el cruce de interflujo (Interflow Match) para determinar si los registros coinciden.

Salida

Tabla 16: Datos de salida de Match Key Generator

Nombre de campo	Descripción / Valores válidos
ExpressMatchKey	Valor que indica el nivel de cruce. Si la clave de cruce inmediato es una coincidencia, la calificación es 100. Si la clave de cruce inmediato no es una coincidencia, se obtiene una calificación de 0.
MatchKey	La clave que se genera para identificar registros.

Private Match

Private Match permite que dos entidades comparen conjuntos de datos e identifiquen registros comunes, sin comprometer información confidencial. Por ejemplo, dos empresas podrían estar interesadas en lanzar una campaña de marketing conjunta. Cada empresa posee sus propias bases de datos con información de clientes y desean determinar qué clientes compran en ambas empresas, para utilizar un enfoque más específico en la campaña. Sin embargo, para garantizar la seguridad de los datos y cumplir las regulaciones de privacidad, las empresas no desean compartir estas bases de datos entre sí o dárselas a terceros para que realicen los cruces. La función de cruce privado permite que ambas bases de datos se comparen entre sí, sin vulnerar la seguridad o infringen las leyes de privacidad.

Private Match se usa en uno de los tres modos siguientes:

- **Modo Cifrado:** el primer usuario introduce sus datos y los campos de índice y cruce se extraen y cifran. Se genera una clave pública y una tabla de desplazamiento con los datos del primer usuario para el segundo usuario y además se genera una clave privada para que el primer usuario utilice posteriormente.
- **Modo de coincidencia privada:** el segundo usuario ingresa sus datos y los datos cifrados del primer usuario, proporciona la clave pública y la tabla de desplazamiento y realiza una coincidencia. Se genera un archivo con los datos cruzados, para su envío al primer usuario.
- **Modo Descifrado:** el primer usuario introduce los datos cifrados del segundo usuario, proporciona la clave privada y genera los resultados con un índice coincidente de los datos de ambos usuarios.

Al usar la función de cifrado (modo de cifrado), la seguridad se conserva mientras se realiza una función de cruce (modo de cruce privado) y luego, una función de descifrado muestra el resultado de los datos cruzados (modo de descifrado). Todos los datos generados y compartidos entre los usuarios están cifrados y no se pueden leer.

Input

Los requisitos de entrada de la etapa Private Match varían según la tarea que se realiza:

- **Modo Cifrado:** se debe adjuntar un archivo con los datos del primer usuario al puerto de entrada. Este puede ser un archivo de texto o prácticamente cualquier tipo de archivo de origen.
- **Modo Cruce privado:** se debe adjuntar un archivo con los datos del segundo usuario al primer puerto de entrada; este también puede ser un archivo de texto, una base de datos o prácticamente cualquier tipo de archivo de origen. Los datos cifrados generados por el primer usuario se deben adjuntar al segundo puerto de entrada.
- **Modo Descifrado:** el archivo de salida generado por el segundo usuario.

Opciones

Las opciones de la etapa Private Match varían según la tarea que se realiza.

Nota: Si realiza una actualización a la versión 11.0 o posterior desde la versión 10.x y generó claves privadas, tendrá que volver a generar esas claves debido a que las claves que se generaron en la versión 10.x no son compatibles con las versiones 11.0 y posteriores.

Modo Cifrado

1. Seleccione la operación **Cifrar**.
2. Seleccione el **campo de índice** que proporcione una ID única para cada registro en el archivo. La ID única debe ser un valor numérico.
3. Seleccione el **campo de cruce** que se debe usar para comparar los datos del segundo usuario.
4. Especifique la ruta de acceso y el nombre del archivo **Clave pública** que se creará cuando ejecute el trabajo.
5. Especifique la ruta de acceso y el nombre del archivo **Almacén de claves** que se creará cuando ejecute el trabajo.

6. Especifique la ruta de acceso y el nombre del archivo **Tabla de desplazamiento** que se creará cuando ejecute el trabajo.
7. Escriba un nombre para la **columna de salida** que incluirá los datos cifrados en el archivo de salida que se envía al segundo usuario.
8. Presione **Aceptar**.

Modo Cruce privado

1. Seleccione la operación **Cruce privado**.
2. Seleccione el **campo de índice** que proporcione una ID única para cada registro en el archivo. La ID única debe ser un valor numérico.
3. Seleccione el **campo de cruce** que se debe usar para comparar los datos del primer usuario.
4. Escriba el nombre del **campo de datos cifrados** que el primer usuario ingresó en el paso 7 de las instrucciones del modo Cifrado.
5. Vaya al archivo **Clave pública**.
6. Vaya al archivo **Tabla de desplazamiento**.
7. Escriba un nombre para la **columna de salida** que incluirá los datos cifrados en el archivo de salida que se envía al primer usuario.

Modo Descifrado

1. Seleccione la operación **Descifrar**.
2. Seleccione el **campo de datos cifrados** que se ingreso en el paso 7 de las instrucciones Modo Private Match.
3. Vaya al archivo **Almacén de claves**.
4. Seleccione la **columna de salida** que incluirá los datos descifrados en el archivo de salida. El formato de los datos en este campo es el índice coincidente de los datos del primer usuario y el índice coincidente de los datos del segundo usuario, separados por un carácter de coma (,), como se en el siguiente ejemplo:

```
User1Data,User2Data
```

Salida

Los requisitos de salida de la etapa Private Match varían según la tarea que se realiza:

- Modo Cifrado: un archivo generado por la etapa Write to File, que contiene los datos cifrados del primer usuario.
- Modo Cruce privado: un archivo generado por la etapa Write to File, que contiene la información cifrada de los resultados del cruce.
- Modo Descifrado: un archivo generado por la etapa Write to File, que contiene el índice coincidente de los datos de ambos usuarios.

Transactional Match

Transactional Match compara los registros sospechosos con los registros candidatos que se obtienen en la etapa Candidate Finder. Transactional Match utiliza reglas de cruce para comparar el registro sospechoso con todos los registros candidatos que tienen el mismo número de grupo de candidatos (asignado en Candidate Finder) a fin de identificar duplicados. Si el registro candidato es un duplicado, se le asigna un número de colección, el tipo de registro de cruce se etiqueta como duplicado y luego se copia. A los candidatos sin coincidencias en el grupo se les asigna el número de colección 0, se etiquetan como únicos y luego también se copian.

Nota: Transactional Match solo cruza registros sospechosos con candidatos. No intenta cruzar registros sospechosos con otros registros sospechosos, como se hace en Intraflow Match.

Opciones

1. En el campo **Cargar regla de cruce** seleccione una de las reglas de cruce predefinidas, que puede usar como está o bien puede modificarla para que se ajuste a sus necesidades. Si desea crear una nueva regla de cruce sin usar una de las reglas de cruce predefinidas como punto de partida, haga clic en **Nueva**. Solamente puede tener una regla personalizada en un flujo de datos.

Nota: La función Opciones del flujo de datos en Enterprise Designer permite exponer la regla de cruce para la configuración durante la ejecución.

2. Seleccione **Arrojar candidatos únicos** si desea que los registros de un candidato único sean incluidos en el resultado de la etapa.
3. Seleccione **Generar datos para análisis** si desea usar la herramienta Análisis de cruce para analizar los resultados del flujo de datos. Para obtener más información, consulte [Análisis de resultados de cruce](#) en la página 119.
4. Para obtener más información acerca de cómo modificar las otras opciones, consulte [Generación de reglas de cruce](#) en la página 72.
5. Haga clic en **Evaluar** para evaluar qué calificación obtuvo el registro sospechoso en comparación con los registros candidatos. Para obtener más información, consulte [Interflow Match](#) en la página 213.

Salida

Tabla 17: Datos de salida de Transactional Match

Nombre de campo	Descripción / Valores válidos
HasDuplicates	<p>Identifica si el registro es un duplicado de otro registro. Una de las siguientes:</p> <p>Y El registro es un registro sospechoso y tiene duplicados.</p> <p>N El registro es un registro sospechoso y no tiene duplicados.</p> <p>D El registro es un candidato y es un duplicado del registro sospechoso.</p> <p>U El registro es un candidato, pero no es un duplicado del registro sospechoso.</p>
MatchRecordType	<p>Identifica el tipo de registro de cruce de una colección. Los valores posibles son:</p> <p>Suspect El registro de entrada original que fue marcado como un registro que posiblemente tuviese registros duplicados.</p> <p>Duplicate Registro que es un duplicado del registro de entrada.</p> <p>Unique Registro que no tiene duplicados.</p>
MatchScore	<p>Identifica la calificación general entre dos registros. Los valores posibles son 0-100; 0 indica una coincidencia poco satisfactoria y 100, una coincidencia exacta.</p>
MatchInfo.<RuleName>.IsMatch	<p>Muestra el estado de cruce de la regla como Verdadero o Falso, donde Verdadero representa un cruce y Falso indica que no hay cruces. Esta información aparece si selecciona la opción Arrojar información detallada del cruce en la pantalla Opciones de cruce transaccional.</p> <p>Nota: El valor de este campo se deriva del estado de cruce de los nodos secundarios según la configuración de reglas principal definida.</p>

Nombre de campo	Descripción / Valores válidos
MatchInfo.<RuleName>.Score	<p>Muestra la puntuación de cruce para la regla. Los valores posibles son 0-100, donde 0 indica una coincidencia poco satisfactoria y 100 una coincidencia exacta. Esta información aparece si selecciona la opción Arrojar información detallada del cruce en la pantalla Opciones de cruce transaccional.</p> <p>Nota: La puntuación de este campo se deriva de las puntuaciones de los nodos secundarios según la configuración de reglas principal definida.</p>
MatchInfo.<RuleName>.<RuleNodeName>.IsMatch	<p>Muestra el estado de cruce de cada nodo en la jerarquía de reglas como <code>Verdadero</code> o <code>Falso</code>, donde <code>Verdadero</code> representa un cruce y <code>Falso</code> indica que no hay cruces. <code>RuleNodeName</code> es una variable que es reemplazada por los nombres de nodos jerárquicos en sus reglas de cruce.</p> <p>Cada nodo de la jerarquía de reglas produce este campo, si se selecciona la opción Arrojar información detallada del cruce en la pantalla Opciones de cruce transaccional.</p>
MatchInfo.<RuleName>.<RuleNodeName>.Score	<p>Muestra la puntuación de cruce de cada nodo en la jerarquía de reglas. <code>RuleNodeName</code> es una variable que es reemplazada por los nombres de nodos jerárquicos en sus reglas de cruce.</p> <p>Cada nodo de la jerarquía de reglas produce este campo, si se selecciona la opción Arrojar información detallada del cruce en la pantalla Opciones de cruce transaccional.</p> <p>Los valores posibles son 0-100, donde 0 indica una coincidencia poco satisfactoria y 100 una coincidencia exacta.</p>

Nota: Tanto las etapas de `Validate Address` como las del módulo `Advanced Matching` utilizan el campo `MatchScore`. El valor del campo `MatchScore` en la salida de un flujo de datos es determinado por la última etapa que modifica el valor antes de enviarlo a la etapa de salida. Si el flujo de datos contiene etapas de `Validate Address` y del módulo `Advanced Matching` y desea ver los datos de salida del campo `MatchScore` para cada etapa, utilice una etapa `Transformer` para copiar el valor de `MatchScore` en otro campo. Por ejemplo, `Validate Address` genera un campo de salida denominado `MatchScore` y luego una etapa `Transformer` puede copiar el campo `MatchScore` desde `Validate Address` a un campo denominado `AddressMatchScore`. Al ejecutarse, la etapa de cruce completa el campo `MatchScore` con el valor proveniente de la etapa y transmite el valor del campo `AddressMatchScore` a través de la etapa `Validate Address`.

Índice Write to Search

Write to Search Index (Escritura en índice de búsqueda) le permite crear un índice de texto completo, a partir de los datos que ingresan a la etapa. Al tener estos datos en un índice de búsqueda exclusivo se obtiene un tiempo de respuesta más rápido cuando se realizan búsquedas en el índice desde otras etapas de Spectrum™ Technology Platform. Se prefieren los índices de búsqueda de texto completo antes que las bases de datos relacionales, cuando se tiene una gran cantidad de datos de texto de forma libre que se debe buscar o categorizar o bien si tiene compatibilidad para un alto volumen de consultas interactivas basadas en texto.

Write to Search Index utiliza un analizador para dividir el texto de entrada en pequeños elementos de índice denominados muestras. Posteriormente, extrae los términos del índice de búsqueda de dichas muestras. El tipo de analizador que se use, la manera en la que el texto de entrada se divide en muestras, determina cómo se podrá buscar dicho texto posteriormente. For example, the *Keyword* analyzer always does an exact match and tokenizes the whole string as single token, while the *Standard* analyzer breaks the string to create tokens. The *Stop Word* analyzer is all the more sophisticated and removes articles, such as "a" or "the" from the string before tokenizing, thus shrinking the index size.

Los índices de búsqueda admiten la función de tiempo real cercana, o que permite que los índices se actualicen casi inmediatamente, sin tener que cerrar y reconstruir la etapa por medio del índice de búsqueda.

Para obtener información sobre las tareas de índice de búsqueda que se pueden realizar a través de la línea de comandos, consulte Índice de búsqueda en la sección [Utilidad de administración](#) de la Guía de administración.

Opciones

1. En Enterprise Designer, haga doble clic en el lienzo en la etapa de índice de Write to Search.
2. Ingrese un **nombre** para la conexión.
3. Seleccione un **modo de escritura**. Cuando se regenera un índice, se tienen opciones relacionadas con cómo deben afectar los nuevos datos a los datos ya existentes.
 - **Crear o sobrescribir**: los nuevos datos sobrescribirán los datos existentes y estos últimos ya no estarán en el índice.
 - **Actualizar o Adjuntar**: los nuevos datos sobrescriben los datos existentes y todos los datos nuevos que antes no existían, se agregarán al índice.
 - **Adjuntar**: los nuevos datos se adjuntarán a los datos existentes y estos últimos permanecerán intactos.
 - **Eliminar**: los datos del campo seleccionado se eliminarán del índice de búsqueda.
4. Select the **Key field** on the basis of which you want to **Update or Append** or **Delete** the records.

- In case of **Create or Overwrite** mode, the **Key field** needs to be unique for Elastic search indexes (used in a distributed environment). Si deja el campo en blanco, todos los registros se almacenan en el índice, independientemente de cualquier duplicación. Sin embargo, no podrá realizar ninguna operación de escritura, como actualizar, anexar y eliminar, en este índice. The following table explains the indexing behavior if the **Key field** is non-unique for Lucene and Elastic search indexes.

Modo de escritura:	Campo de clave	Índice de búsqueda Lucene	Índice de búsqueda Elastic
Create or Overwrite	Duplicate records with same Key field	Se almacenan todos los registros. Nota: Los registros duplicados con el mismo campo de clave se sobrescriben en el momento en que ejecuta la operación de actualización.	All duplicate records with the same Key field are overwritten .
Actualizar o adjuntar	Duplicate records with same Key field	Los duplicados se sobrescriben.	Los duplicados se sobrescriben.

5. Marque la casilla **Confirmación de lote** si desea especificar el número de registros para confirmar en un lote, mientras se crea el índice de búsqueda. Luego, ingrese el número en el campo **Tamaño del lote**. El valor predeterminado es 5000.
6. Seleccione un **analizador** para construir:
 - Standard: Ofrece formación de muestras basada en gramática que contiene un superconjunto de analizadores de Whitespace y Stop Word. Comprende la puntuación del español para separar las palabras, sabe qué palabras ignorar (a través de Stop Word Analyzer) y realiza búsquedas que técnicamente no distinguen entre mayúsculas y minúsculas, con comparaciones de minúsculas. Por ejemplo, la cadena de caracteres "Pitney Bowes Software" se devolvería como tres muestras: "Pitney", "Bowes" y "Software". For a comparison of Standard and Keyword analyzers, see [Standard and Keyword Analyzer](#) en la página 239.
 - Whitespace: Separa las muestras con espacios en blanco. Es un subconjunto de Standard Analyzer en cuanto comprende los espacios entre palabras en español en textos basados en espacios y saltos de línea.
 - Stop Word: Elimina los artículos, como "la", "y" y "a" para reducir el tamaño del índice y aumentar el rendimiento.
 - Palabras claves: crea una sola muestra desde un flujo de datos y la mantiene como está. Por ejemplo, la cadena de caracteres "Pitney Bowes Software" se devolvería como una muestra: "Pitney Bowes Software". For a comparison of Standard and Keyword analyzers, see [Standard and Keyword Analyzer](#) en la página 239.

- Ruso: admite índices de idioma ruso y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "y", "yo" y "tú" para reducir el tamaño del índice y aumentar el rendimiento.
 - Alemán: admite índices de idioma alemán y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "el", "y" y "un" para reducir el tamaño del índice y aumentar el rendimiento.
 - Danés: admite índices de idioma danés y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "en", "y" y "un" para reducir el tamaño del índice y aumentar el rendimiento.
 - Holandés: admite índices de idioma holandés y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "el", "y" y "un" para reducir el tamaño del índice y aumentar el rendimiento.
 - Finlandés: admite índices de idioma finlandés y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "es", "y" y "de" para reducir el tamaño del índice y aumentar el rendimiento.
 - Francés: admite índices de idioma francés y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "el", "y" y "un" para reducir el tamaño del índice y aumentar el rendimiento.
 - Húngaro: admite índices de idioma húngaro y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "el", "y" y "un" para reducir el tamaño del índice y aumentar el rendimiento.
 - Italiano: admite índices de idioma italiano y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "el", "y" y "un" para reducir el tamaño del índice y aumentar el rendimiento.
 - Noruego: admite índices de idioma noruego y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "el", "y" y "un" para reducir el tamaño del índice y aumentar el rendimiento.
 - Portugués: admite índices de idioma portugués y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "el", "y" y "un" para reducir el tamaño del índice y aumentar el rendimiento.
 - Español: admite índices de idioma español y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "el", "y" y "un" para reducir el tamaño del índice y aumentar el rendimiento.
 - Sueco: admite índices de idioma sueco y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "el", "y" y "un" para reducir el tamaño del índice y aumentar el rendimiento.
 - Hindi: admite índices de idioma hindi y servicios de avance de caracteres. También es compatible con diversas palabras no significativas y elimina artículos como "por", "y" y "un" para reducir el tamaño del índice y aumentar el rendimiento.
7. Para actualizar el analizador de todos los campos presentes en la lista, seleccione un analizador en el menú desplegable **actualizar Analizadores para....**
 8. To reload the schema from the server, click **Reload Schema**.

Nota: Puede cambiar el nombre del campo, escribiendo el nuevo nombre directamente en la columna de **campos**. Sin embargo, no puede cambiar el nombre de los **Campos de etapa** o el **Tipo**.

9. To selectively add/remove fields from your input source, click **Quick Add**. The **Quick Add** pop-up window displays a list of all the fields from the input source. Select the fields that you want to add and click **OK**.
10. Seleccione los campos de los que desea almacenar datos. Por ejemplo, si se utiliza un archivo de entrada de direcciones, puede indexar solo el campo de código postal y almacenar el resto de los campos (como la línea de dirección 1, la ciudad, el estado) para que se entregue toda la dirección cuando se encuentre una coincidencia con la búsqueda de índice.
11. Seleccione los campos de los que desea agregar datos al índice para consultas de búsqueda.

Nota: If you want to delete certain fields, select those and click **Delete**.

12. Si fuera necesario, cambie el analizador por cualquier campo que use algo distinto a lo que haya seleccionado en el campo Analizador.
13. Haga clic en **Aceptar**.

Salida

Write to Search Index tiene un puerto de salida, llamado **Puerto de error**, que recopila datos de los registros que no se pudieron procesar y agregar al índice de búsqueda. Los registros que pasan a través del puerto de error hacia el receptor se consideran malformados.

Los registros se reenvían al puerto de error si el valor de **KeyField** en el registro de entrada está en blanco o es nulo. Para motores de índice de búsqueda que admiten búsquedas independientes (el motor de índice heredado), el puerto de error funciona solo para las operaciones de Actualización y Eliminación. Sin embargo, para el motor de índice de soporte distribuido, recopila registros malformados para las cuatro operaciones: crear, actualizar, eliminar y adjuntar.

Administración de índice de búsqueda

The **Search Index Management** tool enables you to perform the following tasks:

- Agregar o quitar campos en un índice de búsqueda existente
- Eliminar uno o más índices de búsqueda

Para agregar o quitar campos en un índice de búsqueda existente:

1. Seleccione **Herramientas > Administración de índice de búsqueda**. The **Search Index Management** pop-up window is displayed showing the existing search indexes.
2. Select the index to be modified, and click **Modify**. The **Modify Index** page is displayed showing a list of all the fields in the search index with their details.
3. To delete a field, select it from the list, and click **Remove field**.

4. Para agregar u nuevo campo a la lista de campos, haga clic en **Agregar**. The **Add Field** pop-up window is displayed, which allows you to add the following details:

- **Field name:** Enter the name of the field.
- **Type:** Select the field type. Las opciones son: cadena, entero, largo, flotante y doble.
- **Index:** Select the check-box to add it to the index for search query.
- **Store:** Select the check-box to store the field.
- **Analyzer:** From the list of options, select the analyzer you want to use for indexing the field.

Nota: Esta opción se activará solo si seleccionó indexar este campo.

Para eliminar un índice de búsqueda existente:

1. Seleccione **Herramientas > Administración de índice de búsqueda**.
2. Seleccione los índices de búsqueda que desee eliminar.
3. Haga clic en **Eliminar**.
4. Haga clic en **Cerrar**.

Nota: También puede eliminar un índice de búsqueda mediante el uso de la Utilidad de administración. El comando es `index delete --d IndexName`, en el cual "IndexName" es el nombre del índice que desea borrar.

Standard and Keyword Analyzer

Before choosing between the Standard and Keyword analyzers you should be aware of the following difference in the behavior of these two analyzers:

- **Standard analyzer:** It breaks the string to tokenize it and also converts all its tokens to lower case.
- **Keyword analyzer:** It tokenizes the whole string and keeps it as is (does not change the case).

Ejemplo:

Let us take an example of *contains any* algorithm and assume the input is P O. In this case, the **Standard** analyzer will return both P O and P records, while the **Keyword** analyzer will return only the P O records (shown below).

Search result with Standard analyzer:

Point		Point 2					
SearchTerm	CandidateGroup	HasDuplicates	Index	IndexField	Term	TransactionRecordType	
PO	1	Y				Suspect	
PO	1	D	PO	4	Test4	Candidate	
PO	1	D	P	1	Test 1	Candidate	

Search result with Keyword analyzer:

Point		Point 2					
SearchTerm	CandidateGroup	HasDuplicates	Index	IndexField	Term	TransactionRecordType	
PO	1	Y				Suspect	
PO	1	D	PO	4	Test4	Candidate	

Módulo Business Steward

Módulo Business Steward

El módulo Business Steward es un conjunto de funciones que permiten identificar y resolver registros de excepciones. Los registros de excepciones son registros que Spectrum™ Technology Platform no pudo procesar adecuadamente y deben ser revisados de forma manual por un administrador de datos. Algunos ejemplos de excepciones son:

- Fallas en la verificación de direcciones
- Fallas de geocodificación
- Cruces con bajo nivel de confianza
- Decisiones de combinación/consolidación

El módulo Business Steward incluye una herramienta de explorador para revisar los registros de excepción de forma manual. Cuando se hayan corregido y aprobado los registros de excepción de forma manual, se los puede reincorporar al proceso de calidad de datos de su Spectrum™ Technology Platform.

Muchas de las acciones que se llevan a cabo al interior del módulo Business Steward Module se reflejan en el registro de auditoría, una herramienta de Management Console que registra la actividad del usuario. Las siguientes acciones se incluyen en el registro:

- Adición de excepciones en la etapa Write Exceptions
- Eliminación de excepciones en la etapa Read Exceptions y la página Administrar excepciones de Business Steward Portal
- Asignación de excepciones en el Administrador de excepciones de Business Steward Portal
- Recuperación de excepciones en el Editor de excepciones de Business Steward Portal
- Actualización de excepciones en el Editor de excepciones de Business Steward Portal
- Revalidación de excepciones en el Editor de excepciones de Business Steward Portal

Lea más acerca del registro de auditoría en la Spectrum™ Technology Platform Guía de administración para la IU web.

Componentes

El módulo Business Steward consiste en lo siguiente:

- **Exception Monitor:** es una etapa que evalúa los registros con respecto a una serie de condiciones para determinar si el registro requiere la revisión manual por parte de un administrador de datos. Cuando los registros cumple esas condiciones, esta etapa puede enviar una notificación por correo electrónico los destinatarios de las excepciones.

- **Write Exceptions:** una etapa que escribe los registros de excepción en el repositorio de excepciones. Cuando los registros de excepción están en el repositorio de excepciones, están disponibles para su revisión por parte de un administrador de datos.
- **Panel y editor de excepciones:** una herramienta basada en navegador que muestra un panel de estadísticas en resumen y gráficos para poder entender los tipos de excepciones que ocurren en sus datos. También es un editor que permite modificar los registros de excepción y aprobarlos para su reprocesamiento.
- **Gestionar excepciones:** una herramienta basada en navegador que permite revisar y gestionar actividades de registro de excepción para todos los usuarios.
- **Rendimiento de calidad de los datos:** una herramienta basada en navegador que proporciona información sobre las tendencias en los registros de excepción y permite identificar los indicadores clave de rendimiento y enviar notificaciones cuando se cumplen ciertas condiciones.
- **Read Exceptions:** una etapa que lee registros desde el repositorio de excepciones en un flujo de datos. Esta etapa permite reprocesar registros de excepción que han sido corregidos por un administrador de datos.

Exception Monitor

La etapa Exception Monitor evalúa los registros con respecto a una serie de condiciones para determinar si el registro requiere la revisión manual por parte de un administrador de datos. Exception Monitor le permite destinar registros que Spectrum™ Technology Platform no pudo procesar correctamente hacia una herramienta de revisión manual (Business Steward Portal).

Además de establecer las condiciones que determinan si los registros requieren revisión manual, también puede configurar Exception Monitor para enviar una notificación a una o más direcciones de correo electrónico cuando se hayan cumplido esas condiciones un número determinado de veces.

Para obtener más información acerca del procesamiento de excepciones, consulte [Módulo Business Steward](#) en la página 240.

Input

Exception Monitor toma cualquier registro como entrada. Si los datos de entrada no incluyen un campo llamado "CollectionNumber", la opción **Obtener todos los registros en el grupo de excepción** estará deshabilitada.

Nota: Exception Monitor no puede monitorear campos que contienen datos complejos, como listas u objetos geométricos.

Opciones

Pestaña Condiciones

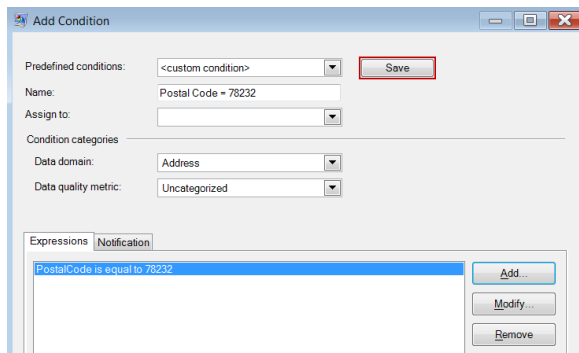
Tabla 18: Opciones de Exception Monitor

Nombre de la opción	Descripción
Detener la evaluación cuando se cumple una condición	Especifica si se continúa la evaluación de un registro según el resto de las condiciones una vez que se cumple con una condición. La activación de esta opción puede mejorar el rendimiento porque reduce potencialmente la cantidad de evaluaciones que el sistema debe realizar. Sin embargo, si no se evalúan todas las condiciones, perderá cierto grado de finalización en los informes de excepciones que se muestran en Business Steward Portal. Por ejemplo, si define tres condiciones (Finalización de dirección, Confianza de nombre y Confianza de geocodificación) y un registro cumple los criterios definidos en Finalización de dirección y usted activa esta opción, el registro no será evaluado con respecto a las condiciones Confianza de nombre y Confianza de geocodificación. Si el registro también califica como una excepción porque coincide con la condición Confianza de nombre, esta información no será capturada. Por el contrario, el registro sería reportado solo como un problema de Finalización de dirección, en lugar de un problema de Finalización de dirección y de Confianza de nombre.

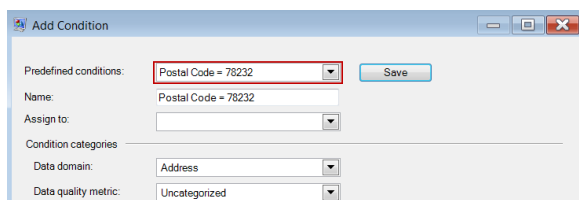
Adición o modificación de condiciones y expresiones

Una condición define los criterios utilizados para determinar si un registro es una "excepción" y debe ser asignado para revisión manual. Normalmente, esto significa que desea definir las condiciones que pueden identificar de forma homogénea los registros que han fallado durante el procesamiento automático al principio del flujo de datos o que tienen un bajo nivel de confianza y, por lo tanto, deben ser revisados en forma manual.

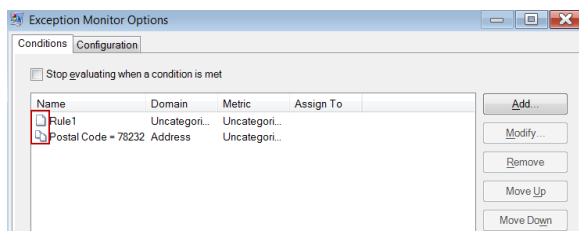
La etapa Exception Monitor le permite crear condiciones predefinidas y personalizadas utilizando el cuadro de diálogo Agregar condición. Hay condiciones predefinidas para todos los flujos de datos, mientras que hay condiciones personalizadas disponibles solo para los flujos de datos para los que se crearon. El proceso de configuración es casi idéntico en ambos tipos; sin embargo, para crear una condición predefinida, debe guardar la condición completando los campos y haciendo clic en **Guardar**, en el cuadro rojo que se muestra abajo.



Después de guardar una condición personalizada, el campo **Condiciones predefinidas** cambia para mostrar el nombre de la condición en lugar de "<custom condition>".



Después de crear condiciones predefinidas o personalizadas, estas aparecerán en la ficha Condiciones del cuadro de diálogo Opciones de Exception Monitor. Como se muestra en la siguiente imagen, el icono que está junto al nombre de la condición identifica como condición predefinida o condición personalizada. Un icono de documento doble designa una condición predefinida y un icono de documento sencillo designa una condición personalizada.



1. En la ficha Condiciones de la ventana **Opciones de Exception Monitor**, haga clic en **Agregar** para crear una condición nueva o **Modificar** para editar una condición existente. Complete estos campos:

- **Condiciones predefinidas:** seleccione una condición predefinida o mantenga "<custom condition>" en el menú desplegable para crear una nueva condición.
- **Name (Nombre):** un nombre para la condición. Puede elegir cualquier nombre que desee. Dado que el nombre de la condición se muestra en Business Steward Portal, es recomendable que utilice un nombre descriptivo. Por ejemplo, "MatchScore<80" o "FailedDPV". Si intenta entregar una nueva condición a un nombre que sea idéntica a una condición existente, pero con otros caracteres anexados al final (por ejemplo, "FailedDPV" y "FailedDPV2"), se le preguntará si desea sobrescribir la condición existente apenas escriba el último carácter que coincida con su nombre (mediante nuestro ejemplo, "V"). Diga "Sí" a la indicación, termine de nombrar la condición y cuando presione **OK** o **Guardar**, ambas condiciones se podrán

ver en el cuadro de diálogo Opciones de Exception Monitor. La nueva condición no sobrescribirá la condición existente a menos que el nombre sea 100% idéntico.

- **Assign to** (Asignar a): seleccione un usuario a quien se deben asignar los registros de excepción que cumplan esta condición. Si no hace una selección en este campo, los registros de excepción se asignarán automáticamente al usuario que ejecutó el trabajo.
- **Data domain** (Dominio de datos): (opcional) especifica el tipo de datos que la condición evalúa. Se utiliza exclusivamente con una finalidad informativa para mostrar qué tipos de excepciones ocurren en sus datos. Por ejemplo, si la condición evalúa el éxito o la falla de una validación de direcciones, el dominio de datos sería "Dirección"; si la condición evalúa el éxito o la falla de una operación de geocodificación, el dominio de datos sería "Espacial", etc. Puede especificar su propio dominio de datos o seleccionar uno de los dominios predefinidos:
 - Account: la condición comprueba el nombre de una empresa u organización asociada con una cuenta de ventas.
 - Address: la condición comprueba los datos de dirección, como una dirección de correo completa o un código postal.
 - Asset: la condición comprueba datos sobre la propiedad de una empresa, por ejemplo, bienes materiales, inmuebles, recursos humanos u otros activos.
 - Date: la condición comprueba los datos de fecha.
 - Email: la condición comprueba los datos de correo electrónico.
 - Financial: la condición comprueba datos relacionados con moneda, títulos valores, etc.
 - Name: la condición comprueba los datos de nombre personal, como el nombre o el apellido.
 - Phone: la condición comprueba los datos de número telefónico.
 - Product: la condición comprueba datos sobre materiales, piezas, mercancías, etc.
 - Spatial: la condición comprueba los datos de punto, polígono o línea que representan un accidente geográfico, por ejemplo, llanuras aluviales, líneas costeras, casas, zonas de ventas, etc.
 - SSN: la condición comprueba los números de seguro social de Estados Unidos.
 - Uncategorized: elija esta opción si no desea categorizar esta condición.
- **Data quality metric** (Métrica de calidad de datos): (opcional) especifica la métrica que mide esta condición. Se utiliza exclusivamente con una finalidad informativa para mostrar qué tipos de excepciones ocurren en sus datos. Por ejemplo, si la condición es diseñada para evaluar la finalización del registro (por ejemplo, que todas las direcciones contienen códigos postales), podría especificar "Finalización" como métrica de calidad de datos. Puede especificar su propia métrica o seleccionar una de las métricas predefinidas:
 - Accuracy: la condición mide si los datos pueden ser cotejados con una fuente de confianza. Por ejemplo, si no es posible verificar una dirección utilizando datos de la autoridad postal, se la puede considerar una excepción porque no es precisa.
 - Completeness: la condición mide si a los datos les falta algún atributo esencial. Por ejemplo, una dirección a la que le falta el código postal o una cuenta a la que le falta un nombre de contacto.

- **Consistency:** la condición mide si los datos son uniformes entre múltiples sistemas. Por ejemplo, si su sistema de datos de clientes utiliza los códigos de género M y F, pero los datos que está procesando tienen el código de género 0 y 1, posiblemente se considere que los datos presentan problemas de uniformidad.
 - **Interpretability:** la condición mide si los datos son correctamente analizados e ingresados en una estructura de datos que otro sistema puede interpretar. Por ejemplo, los números de seguro social deben contener solo datos numéricos. Si los datos contienen letras, por ejemplo xxx-xx-xxxx, posiblemente se considere que presentan problemas de interpretación.
 - **Recency:** la condición mide si los datos están actualizados. Por ejemplo, si una persona se muda, pero la información que usted tiene en su sistema es la dirección anterior, posiblemente se considere que los datos tienen un problema de actualización.
 - **Uncategorized:** elija esta opción si no desea categorizar esta condición.
 - **Uniqueness:** la condición mide si hay datos duplicados. Si el flujo de datos no pudo consolidar datos duplicados, los registros pueden ser considerados una excepción.
2. Debe agregar al menos una expresión a una condición. Una expresión es una instrucción lógica que comprueba el valor de un campo. Para agregar una expresión, haga clic en **Agregar**. Para modificar una expresión existente, haga clic en **Modificar**. Complete estos campos:
- **Expresión creada con el generador de expresiones:** seleccione esta opción para crear una expresión básica.
 - **Expresión personalizada:** seleccione esta opción para escribir una expresión usando comandos Groovy. Si necesita utilizar una lógica más compleja, como evaluaciones anidadas, utilice una expresión personalizada. Para obtener más información, consulte [Uso de expresiones personalizadas de Exception Monitor](#) en la página 247.
 - Si hay otras expresiones que ya están definidas para esta condición, puede seleccionar un operador en el campo **Operador lógico**. Una de las siguientes:
 - **And:** esta expresión y la que la precede deben ser verdaderas para que la condición sea verdadera.
 - **Or:** si la expresión es verdadera, la condición es verdadera aunque la expresión que la preceda no lo sea.
 - Si elige crear una expresión mediante un generador de expresiones, los siguientes campos están disponibles:
 - **Nombre de campo:** seleccione el campo que desea que sea evaluado por esta expresión. La lista de campos disponibles se completa según las etapas anteriores a la etapa Exception Monitor.
 - **Operador:** seleccione el operador que quiere utilizar en la evaluación.
 - **Valor:** especifique el valor que quiere que la expresión compruebe utilizando el operador elegido en el campo Operador.
3. Haga clic en **Agregar** para agregar la expresión. Haga clic en **Cerrar** cuando haya terminado de agregar expresiones.
4. Utilice los botones **Subir** y **Bajar** para cambiar el orden en que se evalúan las expresiones.

5. Haga clic en la ficha **Notificación** si desea que Exception Monitor envíe un mensaje a una o más direcciones de correo electrónico cuando esta condición se haya cumplido una cantidad especificada de veces. Dicho correo electrónico incluirá un enlace a los registros fallidos en el Editor de excepciones de Business Steward Portal, en donde podrá ingresar manualmente los datos correctos. Si no desea configurar notificaciones, siga al paso 11. Para que una dirección de correo electrónico en particular no siga recibiendo notificaciones, elimine dicha dirección de la lista de destinatarios en la línea **Enviar notificación a** de la ficha Notificación en el cuadro de diálogo Modificar condición.

Nota: Las notificaciones deben configurarse en la consola de administración antes de que puedan usar en forma correcta desde dentro de Exception Monitor. Consulte *Guía de administración* para obtener información acerca de la configuración de notificaciones.

6. Ingrese la dirección o direcciones de correo electrónico a las cuales debe enviarse la notificación. Separe las diferentes direcciones por comas, espacios o punto y coma.
7. Designe el punto en el cual usted desea que se envíe una notificación. Puede hacer que sea enviado la primera vez que se cumpla la condición, o que se envíe cuando la condición se haya cumplido una en cantidad específica de oportunidades. El valor máximo es 1.000.000 instancias.
8. Marque la casilla **Enviar recordatorio después de** si desea que se envíen mensajes recordatorios a las direcciones de correo electrónico definidas luego del correo electrónico inicial.
9. Ingrese la cantidad de días durante los cuales usted desea que el recordatorio sea enviado.
10. Haga clic en **Recordar a diario** si desea que los mensajes recordatorios sean enviados todos los días a partir del primer mensaje recordatorio.
11. Enter the **Subject** for the email notification.
12. Si desea guardar esta condición para volver a utilizarla como condición predefinida, haga clic en **Guardar**. Si modifica una condición existente y hace clic en Guardar, se le preguntará si desea sobrescribir la condición existente; tenga en cuenta que si sobrescribe una condición predefinida, dichos cambios se aplicarán en todos los flujos de datos que utilicen la condición.
13. Cuando haya terminado de trabajar con las expresiones, haga clic en **Aceptar**.
14. Agregue o modifique condiciones adicionales según sea necesario.
15. Utilice los botones **Subir** y **Bajar** para cambiar el orden en que se evalúan las condiciones. El orden de las condiciones es importante solo si tiene activada la opción **Detener la evaluación cuando se cumple una condición**. Para obtener más información sobre [Pestaña de configuración](#) en la página 248, consulte .
16. Cuando haya terminado, haga clic en **Aceptar**.

Eliminación de una condición o expresión

- Para eliminar una condición, abra Exception Monitor, seleccione la condición que desea eliminar y, luego, haga clic en **Eliminar**. Cabe destacar que cuando elimina una condición, se eliminan todas las expresiones de esa condición.
- Para eliminar una expresión, abra la condición que contiene esa expresión, selecciónela y haga clic en **Eliminar**.

Uso de expresiones personalizadas de Exception Monitor

Puede escribir sus propias expresiones personalizadas para controlar cómo Exception Monitor redirige los registros, mediante el lenguaje de scripting Groovy para crear una expresión.

Uso de comandos Groovy

Para obtener más información acerca de Groovy, consulte groovy-lang.org.

Las expresiones Groovy que se utilizan en la etapa Exception Monitor, se deben evaluar como un valor booleano (verdadero o falso), que indica si un registro se considera una excepción y se debe redirigir para su revisión manual. Los registros de excepción son enviados al puerto de excepción.

Por ejemplo, si necesita revisar registros con un nivel de confianza de validación menor que 85, su secuencia de comandos tendría la siguiente apariencia:

```
data['Confidence']<85
```

El monitor evaluaría el valor del campo Confianza según sus criterios para determinar a qué puerto de salida enviarlo.

Verificación de un valor único en un campo

Este ejemplo se evalúa como verdadero si el campo Estado contiene 'F'. Tiene que ser un cruce exacto, por lo que 'f' no se evaluaría como verdadero.

```
return data['Status'] == 'F';
```

Verificación de valores múltiples en un campo

Este ejemplo se evalúa como verdadero si el campo Estado contiene 'F' o 'f'.

```
boolean returnValue = false;
if (data['Status'] == 'F' || data['Status'] == 'f')
{
    returnValue = true;
}
return returnValue;
```

Evaluación de la longitud del campo

Este ejemplo se evalúa como verdadero si el campo PostalCode tiene más de 5 caracteres.

```
return data['PostalCode'].length() > 5;
```

Verificación de un carácter dentro de un valor de campo

Este ejemplo se evalúa como verdadero si el campo PostalCode contiene un guión.

```
boolean returnValue = false;
if (data['PostalCode'].indexOf('-') != -1)
{
    returnValue = true;
}
return returnValue;
```

Errores habituales

A continuación se muestran errores habituales que se cometen al utilizar secuencias de comandos.

La siguiente secuencia es incorrecta porque PostalCode (el nombre de la columna) debe estar dentro de comillas simples o dobles

```
return data[PostalCode];
```

La siguiente secuencia es incorrecta porque no se especifica ninguna columna

```
return data[];
```

La siguiente secuencia es incorrecta porque row.set() no devuelve un valor booleano. Siempre se evaluará como falso y se cambiará el campo PostalCode a 88989.

```
return row.set('PostalCode', '88989');
```

Use un solo signo igual para configurar el valor de un campo y un signo igual doble para verificar el valor de un campo.

Pestaña de configuración**Tabla 19: Opciones de Exception Monitor**

Nombre de la opción	Descripción
Desactivar Exception Monitor	Activa y desactiva Exception Monitor. Si desactiva Exception Monitor, los registros simplemente pasarán por la etapa y no se realizará acción alguna. El efecto es similar al que se produce cuando se elimina Exception Monitor del flujo de datos.
Detener el trabajo después de alcanzar el límite de excepciones	Especifica si se detiene la ejecución de un trabajo cuando la cantidad especificada de registros cumple con las condiciones de excepción.

Nombre de la opción	Descripción
Cantidad máxima de registros de excepciones	Si selecciona Detener el trabajo después de alcanzar el límite de excepciones, utilice este campo para especificar la cantidad máxima de registros de excepción que se permiten antes de detener la ejecución del trabajo. Por ejemplo, si especifica 100, el trabajo se detendrá cuando se encuentre el registro de excepción número 101.
Solo informar (no crear excepciones)	Le permite hacer un seguimiento de registros que cumplen con las condiciones de excepción e informar dichas estadísticas en la página Rendimiento de calidad de datos en el portal Business Steward Portal, pero no crea excepciones para dichos registros.
Obtener todos los registros en el grupo de excepción	<p>Especifica si se obtienen todos los registros que pertenecen a un grupo de registro de excepción en lugar de solo un registro de excepción. Por ejemplo, un grupo de cruce (basado en MatchKey) contiene cuatro registros. Uno es un registro sospechoso, otro es un duplicado cuya calificación es 90 y dos son registros únicos que tienen las calificaciones 80 y 83. Si tiene una condición que indica que cualquier registro con MatchScore entre 80 y 89 es una excepción, los registros con una calificación de cruce de 80 y 83 se envían al puerto de excepción de forma predeterminada. Sin embargo, si activa esta opción, los cuatro registros serán enviados al puerto de excepción.</p> <p>Active esta opción si desea que los administradores de datos puedan comparar el registro de excepción con los otros registros del grupo. Al comparar todos los registros del grupo, los administradores de datos pueden tomar decisiones más informadas sobre qué hacer con un registro de excepción. Por ejemplo, en un cruce, un administrador de datos podría ver todos los candidatos para determinar si la excepción es un duplicado de otros.</p> <p>Nota:</p> <p>Si los datos de entrada no incluyen un campo llamado "CollectionNumber", esta opción estará deshabilitada.</p>
Agrupar por	<p>Si seleccionó Obtener todos los registros en el grupo de excepción, elija el campo que determinará la agrupación de los registros.</p> <p>Nota: El campo de entrada "CollectionNumber" no aparecerá en esta lista ya que no es una selección válida la opción Agrupar por función.</p>
Servicio de revalidación	Seleccione el servicio que desea ejecutar cuando revalida los registros de este flujo de datos.
Acción tras la revalidación	Especifica si desea volver a procesar registro o aprobar registros que se hayan revalidado.

Nombre de la opción	Descripción
Comparar los registros de excepción mediante un campo de cruce	<p>Utiliza campos de cruce para comparar registros de entrada con registros de excepción en el repositorio. Active esta opción si su entrada contiene registros que generaron excepciones anteriormente, pero ahora están corregidos en la entrada.</p> <p>Los registros de entrada se evaluarán con comparación con las condiciones y, a continuación, contra los registros de excepción existentes en el repositorio. Si un registro de entrada pasa las condiciones y se compara con un registro de excepción, este se eliminará del repositorio. Si un registro de entrada no pasa las condiciones y se comparara con un registro de excepción, dicho registro de excepción se actualizará y guardará en el repositorio. Además, si existen duplicados en el repositorio, solo se actualizará una excepción comparada por flujo de datos, todas las demás se eliminarán para ese flujo de datos.</p>
Optimizado para registros únicos o lotes pequeños	<p>Esta opción está activada cuando selecciona Campos de coincidencia. Cuando esta opción no está marcada (opción predeterminada), el servidor cargará en la memoria todos los registros de excepciones existentes en la etapa y el flujo de datos actuales antes de procesar los registros de excepciones entrantes. Esto se recomienda cuando el repositorio tiene un bajo número de registros de excepciones existentes y un alto número de registros de excepciones nuevos o de actualizaciones. Este caso hipotético generalmente implica un tiempo de carga inicial más largo y un mayor requerimiento de memoria; es más rápido cuando se procesan lotes más grandes, como las actualizaciones diarias, semanales o mensuales.</p> <p>Se recomienda marcar esta opción cuando el repositorio tiene un alto número de registros de excepciones existentes y un número relativamente bajo de registros de excepciones nuevas o actualizaciones, ya que el servidor consulta al repositorio por registros de excepciones existentes a medida que se lee cada registro de entrada. Este caso hipotético generalmente implica un tiempo de carga inicial más breve y un requerimiento de memoria menor; es más rápido cuando se procesan unos pocos registros en tiempo real.</p>
Campos de cruce	<p>Proporciona una lista de todos los campos de entrada utilizados para generar una clave para comparar un registro de excepción en el repositorio. Debe definir al menos un campo de cruce si marcó la casilla Comparar registros de excepción utilizando campos de cruce.</p>

Salida

Exception Monitor arroja registros en dos puertos. Un puerto contiene registros que no cumplen ninguna de las condiciones definidas en la etapa Exception Monitor. El otro puerto, el puerto de excepción, contiene todos los registros que cumplen con una o más condiciones de excepción. El puerto de excepciones también puede incluir registros que no sean de excepciones si está activa la opción **Obtener todos los registros en el grupo de excepción**. Exception Monitor no agrega ni modifica los campos que están dentro de un registro.

Leer excepciones

Read Exceptions es una etapa que lee registros desde el repositorio de excepciones como entrada en un flujo de datos. (Para obtener más información acerca del repositorio de excepciones, consulte [Módulo Business Steward](#) en la página 240).

Nota: Cuando Read Exceptions ya ha leído un registro en un flujo de datos, este es eliminado del repositorio.

Input

Read Exceptions lee datos desde el repositorio de excepciones. No toma entradas desde otra etapa del flujo de datos.

Nota: Solo los registros marcados como "aprobados" en Business Steward Portal se leen en el flujo de datos.

Opciones

La etapa Read Exceptions tiene las siguientes opciones.

Ficha General

Las opciones de la ficha **General** especifican qué registros de excepción quiere que se lean en el flujo de datos.

Las opciones de **Filter** permiten seleccionar un subconjunto de registros desde el repositorio de excepciones utilizando estos criterios:

- **Usuario:** el usuario que ejecuta el flujo de datos que generó las excepciones que usted quiere leer en el flujo de datos.
- **Nombre del flujo de datos:** el nombre del flujo de datos que generó las excepciones que usted quiere leer en el flujo de datos.
- **Rótulo de etapa:** el rótulo de la etapa Exception Monitor que se muestra en el flujo de datos en Enterprise Designer. Estos criterios son útiles si el flujo de datos que generó las excepciones contiene múltiples etapas Exception Monitor y usted solo quiere realizar la lectura en las excepciones de una de esas etapas Exception Monitor.
- **Fecha desde:** la fecha y la hora de los registros más antiguos que desea leer en el flujo de datos. La fecha de un registro de excepción es la fecha de la última vez que fue modificado.
- **Fecha hasta:** la fecha y la hora de los registros más nuevos que desea leer en el flujo de datos. La fecha de un registro de excepción es la fecha de la última vez que fue modificado.

La lista **Campos** muestra los campos que se leerán en el flujo de datos. De forma predeterminada, todos los campos están incluidos, pero puede excluir campos desmarcando la casilla de la columna **Incluir**.

La lista **Vista previa** muestra los registros que cumplen con los criterios especificados en **Filter**.

Nota: La vista previa muestra solo los registros que han sido "Aprobados" en Business Steward Portal y cumplen con los criterios de filtro.

Ficha Ordenar

Utilice la ficha Ordenar para ordenar los registros de entrada según los valores de campo.

- **Agregar:** agrega un campo para ordenar.
- Columna **Nombre de campo:** muestra el nombre del campo por ordenar. Puede seleccionar un campo haciendo clic en el botón desplegable.
- Columna **Orden:** especifica si el orden será descendente o ascendente.
- **Ascendente y Descendente:** cambian la disposición del orden. Los registros están ordenados, en primer lugar, por el campo en el extremo superior de la lista, luego por el segundo, etc.
- **Eliminar:** elimina un campo de orden.

Ficha Tiempo de ejecución

- **Registro de inicio:** especifica la posición en el repositorio del primer registro que desea leer en el flujo de datos. Por ejemplo, si desea omitir los primeros 99 registros de un repositorio, especifique 100. El registro número 100 será el primero que se leerá en el repositorio si cumple con los criterios especificados en la ficha **General**. La posición de un registro está determinada por el orden de los registros en Business Steward Portal.
- **Todos los registros:** seleccione esta opción si quiere leer todos los registros que cumplen con los criterios de búsqueda especificados en la ficha **General**.
- **Máximo de registros:** seleccione esta opción si desea limitar el número de registros que se leen en el flujo de datos. Por ejemplo, si solo quiere leer los primeros 1000 registros que cumplen con los criterios de selección, seleccione esta opción y especifique 1000.

Salida

La etapa Read Exceptions arroja registros desde el repositorio de excepciones que han sido aprobados y cumplen con los criterios de selección especificados en las opciones de Read Exception. Además de los campos de registros, Read Exceptions arroja aquellos campos que describen las últimas modificaciones que se le hicieron al registro en Business Steward Portal.

Tabla 20: Datos de salida de Read Exceptions

Nombre de campo	Descripción
Exception.Comment	Todo comentario que ingresó la persona que resolvió la excepción. Por ejemplo, los comentarios pueden describir las modificaciones que Business Steward le hizo al registro.

Nombre de campo	Descripción
Exception.LastModifiedBy	El último usuario que modificó el registro en Business Steward Portal.
Exception.LastModifiedMilliseconds	La hora de la última vez que se modificó el registro en Business Steward Portal. La hora está expresada en milisegundos desde el 1 de enero de 1970 0:00 GMT. Esta es la forma estándar de calcular el tiempo en el lenguaje de programación Java. Puede utilizar este valor para realizar comparaciones de datos o crear una transformación para convertir este valor al formato de datos que desee.
Exception.LastModifiedString	La hora de la última vez que se modificó el registro en Business Steward Portal. Este campo ofrece una representación más comprensible de la fecha que el campo Exception.LastModifiedMilliseconds. La hora está expresada en el siguiente formato: Thu Feb 17 13:34:32 CST 2011

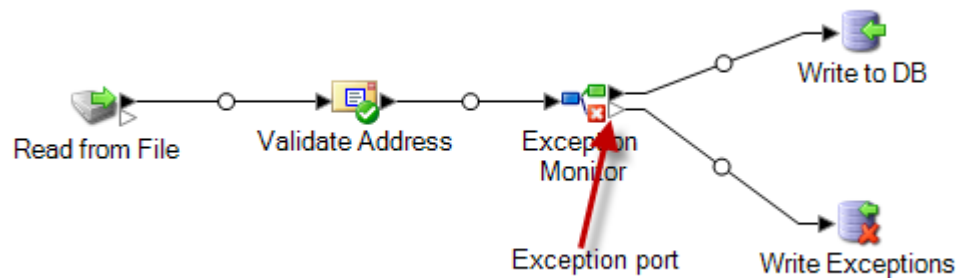
Write Exceptions

Write Exceptions es una etapa que toma registros identificados por la etapa Exception Monitor como excepciones y los escribe en el repositorio de excepciones. Cuando están en el repositorio de excepciones, los registros pueden ser revisados y editados utilizando Business Steward Portal (Portal de Business Steward).

Nota: No se escriben registros de excepción en el Business Steward Portal cuando se ejecutan trabajos o servicios en el modo de inspección de Enterprise Designer o el modo de vista previa de Management Console.

Input

La etapa Write Exceptions toma registros desde el puerto de excepción en la etapa Exception Monitor y, luego, los escribe en el repositorio de excepciones. La etapa Write Exceptions debe ser posterior al puerto de excepción de la etapa Exception Monitor. El puerto de excepción es el último puerto de salida de la etapa Exception Monitor:



Opciones

La etapa Write Exceptions permite seleccionar qué datos de campo deben arrojarse al repositorio de excepciones. Los campos que aparecen dependen de las etapas que ocurren antes en el flujo de datos. Si, por ejemplo, tiene una etapa Validate Address en el flujo de datos, verá los campos AddressLine1, AddressLine2, City (Ciudad) y PostalCode en la etapa Write Exceptions. De forma predeterminada, todos esos campos están seleccionados; desactive las casillas de los campos que no desea que se devuelvan al repositorio de excepciones. El orden de los campos está definido por su orden al llegar a etapa Write Exceptions. Puede reorganizar los campos si selecciona una fila y usa las flechas de la derecha de la pantalla para mover la fila hacia arriba o hacia abajo. El orden que seleccione aquí, se mantendrá para todos los usuarios en Business Steward Portal. Sin embargo, cada usuario puede reorganizar los campos a su gusto en el portal.

Seleccione **Permitir al usuario crear registros de Best Of Breed en el portal** para realizar la consolidación manual de registros cuando los cruces de datos no se pueden realizar con las reglas estándar. Esta función copia un registro seleccionado dentro de un grupo y lo utiliza en lugar de los duplicados para el procesamiento. Esta opción solo está disponible para los registros de excepción agrupados que se generan a partir de un trabajo coincidente, que se identifica mediante la presencia de los campos CandidateGroup o CollectionNumber. Cuando utiliza esta opción, se agregará al registro de excepción un campo de solo lectura denominado "CollectionRecordType". Puede ver este campo en la parte inferior de la lista; tenga en cuenta que todas las opciones del campo están desactivadas.

Nota: Si el flujo de datos también se configura para **revalidación**, deberá agregar y exponer manualmente el campo CollectionRecordType en la etapa/subflujo Exception Monitor y el servicio mismo.

Después de agregar un registro **Best of Breed** en la vista Resolver duplicados del editor de excepciones Business Steward Portal, este campo se establecerá en "BestOfBreed." Si opta por crear registros Best Of Breed, no podrá utilizar la opción **Aprobar todos** para esos registros en el Business Steward Portal. Lea más sobre los registros Best Of Breed en el Business Steward Portal [aquí](#).

Quizás quiere que ciertos campos de entrada estén en el repositorio, pero que no se vean en Business Steward Portal. Puede ser el caso de campos que contienen datos privados o simplemente porque desea ordenar lo que aparece en el Portal. Marque la casilla **Permitir visualización** para

designar los campos seleccionados que deben ser visibles, una vez que se transfieran al repositorio de excepciones. De forma predeterminada, todos los campos son visibles. Desmarque la casilla de los campos que no desea que se vean en el portal.

También puede designar los campos seleccionados que deben ser editables en el portal, una vez que se transfieran al repositorio de excepciones. De forma predeterminada, la columna **Permitir edición** está marcada para todos los campos que pasan a la etapa Write Exceptions. Desactive la casilla de todos los campos que desee devolver al repositorio de excepciones en un estado de solo lectura.

Por último, puede utilizar la función **Búsqueda** para asignar una búsqueda a un campo que contenga datos problemáticos. Puede seleccionar de la lista de búsquedas que se han definido en la herramienta de configuración de Business Steward o puede ingresar manualmente el nombre de la búsqueda. Para obtener más información acerca de las búsquedas, consulte [Búsquedas](#) en la página 278.

Nota: Las búsquedas pueden asignarse únicamente a los campos de tipo cadena.

Salida

Write Exceptions no arroja ninguna salida en el flujo de datos. Escribe registros de excepción en el repositorio de excepciones.

Business Steward Portal

Introducción a Business Steward Portal

¿Qué es Business Steward Portal?

Business Steward Portal es una herramienta para revisar, modificar y aprobar registros que fallaron durante el procesamiento automático o que no fueron procesados con el suficiente nivel de confianza. Utilice Business Steward Portal para ingresar de forma manual los datos correctos u otros adicionales en un registro. Por ejemplo, si el registro de un cliente falla durante el proceso de validación de direcciones, puede usar las herramientas de búsqueda para realizar una investigación, determinar cuál es la dirección del cliente y modificar el registro para que contenga la dirección correcta. Luego, Spectrum™ Technology Platform puede aprobar y volver a procesar el archivo modificado, enviarlo a otra validación de datos o proceso de enriquecimiento, o escribirlo en una base de datos, según la configuración. También puede utilizar el portal para agregar información que no estaba incluida en el registro original.

Business Steward Portal además ofrece gráficos de resumen que permiten ver mejor los tipos de datos que provocan el procesamiento de excepciones, entre ellos, el dominio de los datos (nombre, direcciones, datos espaciales, etc.), así como la métrica de calidad en la que fallan los datos (finalización, precisión, coherencia, etc.).

Además, la página Administrar excepciones de Business Steward Portal permite revisar y administrar la actividad de los registros de excepción, incluso reasignar registros de un usuario a otro. Por último, la página de calidad de datos de Business Steward Portal entrega información sobre tendencias en etapas y flujos de datos.

Para obtener más información acerca del procesamiento de excepciones, consulte **Módulo Business Steward** en la página 240.

Acceso a Business Steward Portal

Para abrir Business Steward Portal, vaya a **Inicio > Todos los programas > Pitney Bowes > Spectrum Technology Platform > Servidor > Página de bienvenida** y escoja **Spectrum Data Quality**, luego **Business Steward Portal** y después haga clic en **Abrir Business Steward Portal**.

De forma alternativa, podría seguir estos pasos:

1. Abra un navegador web y vaya a `http://<servername>:8080/bsmportal`.

Por ejemplo:

`http://myserver:8080/bsmportal`

Comuníquese con su administrador de Spectrum™ Technology Platform si no conoce el nombre de servidor ni el puerto.

2. Inicie sesión en Spectrum™ Technology Platform. Comuníquese con su administrador de Spectrum™ Technology Platform si tiene problemas para iniciar sesión.

El menú de Business Steward Portal

El menú de Business Steward Portal consta de cuatro opciones y un acceso al sistema de ayuda, tal como se muestra a continuación:

- **Tablero:** vea el estado de las excepciones asignadas a usted. Si cuenta con permisos de visualización, también puede ver el estado de las excepciones asignadas a otros usuarios.
- **Editor:** revise, edite y apruebe registros de excepción para su reprocesamiento.
- **Administración:** si tiene permisos de visualización, puede asignar excepciones a sí mismo o a otras personas. Si tiene permisos de eliminación, puede purgar los registros de excepción del repositorio.
- **Calidad de datos:** si tiene permisos de visualización, puede acceder a esta página para ver información estadística y configurar indicadores clave de rendimiento para registros de excepción. Si no tiene permisos de visualización, no puede acceder a la página.
- Menú desplegable del usuario: acceda al sistema de ayuda de Business Steward Portal, a toda la documentación de la plataforma tecnológica de Spectrum y a la página de perfil, donde se puede definir preferencia de idioma y especificar un país para indicar la preferencia cultural, ingresar un correo electrónico para recibir notificaciones y cambiar su contraseña.

La página del tablero

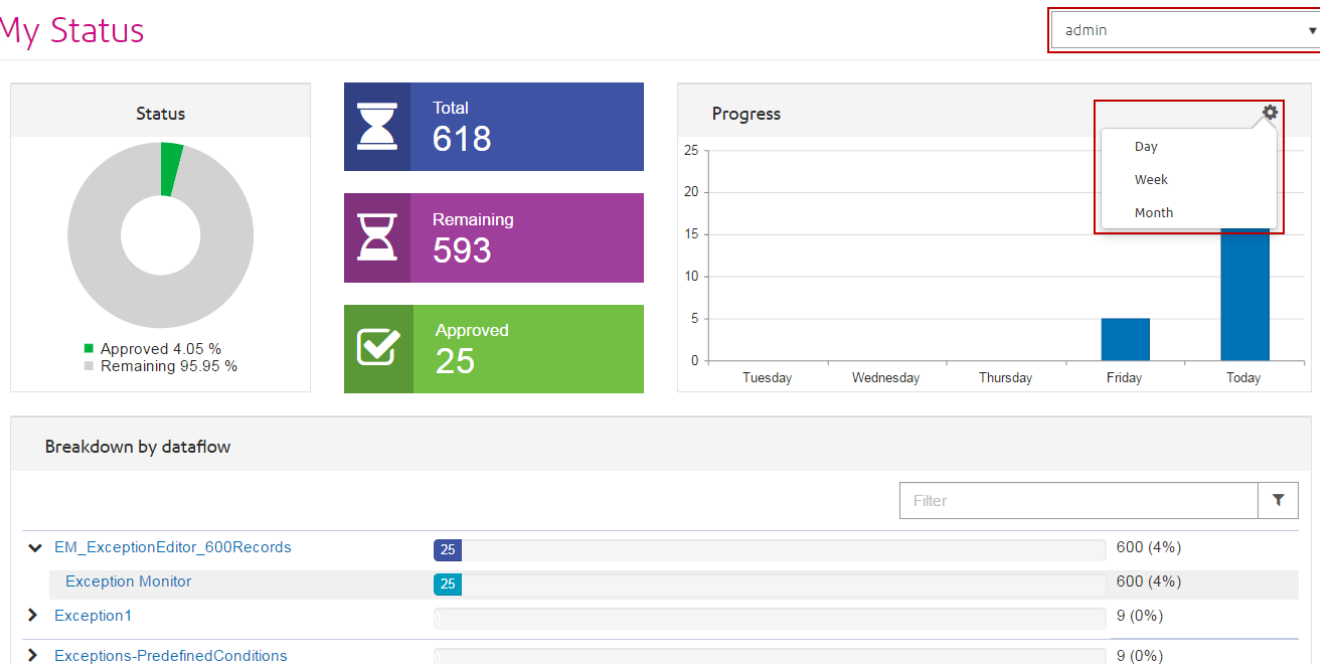
El tablero de excepciones presenta datos que resumen el estado de los registros de excepción que le pertenecen a usted y a otros usuarios. (Tenga en cuenta que solo puede ver los datos de otras personas si tiene permisos de modificación). Esto incluye lo siguiente:

- The total number of exceptions
- El número y el porcentaje de excepciones que se han aprobado en comparación con los que no se han aprobado
- El progreso del usuario seleccionado por día, semana o mes
- Exception record approval progress by dataflow and stage(s) within the dataflow

Nota: Los gráficos de progreso de usuario y flujo de datos no aparecerán si desactivó el seguimiento del progreso en la página de configuración de Business Steward de Management Console.

1. En la ficha Tablero, seleccione el usuario cuya actividad de excepciones desea ver en la primera casilla desplegable que aparece destacada a continuación. Tenga en cuenta que solo los usuarios que actualmente tengan excepciones asignadas aparecerán en esta lista.

My Status



2. Haga clic en la rueda de engranajes **Configuración** (que aparece en el segundo cuadro desplegable destacado arriba) para seleccionar la duración en días, semanas o meses.
3. Ingrese los criterios de búsqueda en el campo **Filter** para reducir la lista de flujos de datos o subconjunto de flujo de datos. El término de búsqueda se debe incluir en el nombre del flujo de datos.

La página del editor

El Editor de excepciones proporciona un medio para revisar, modificar y aprobar manualmente los registros de excepciones. El objetivo de las revisiones manuales es determinar qué datos son incorrectos o faltan y luego actualizarlos y aprobarlos, especialmente si Spectrum™ Technology Platform no pudo corregirlos como parte de un proceso automatizado de flujo de datos. Puede revalidar los registros de excepción para que sean aprobados o reprocesados.

Además puede utilizar el Editor de excepciones para resolver los registros de excepción duplicados y utilizar las herramientas de búsqueda para buscar la información que lo ayude a editar, aprobar o volver a ejecutar los registros.

Personalización del contenido del Editor de excepciones

Existen muchas maneras de personalizar lo que aparece en el Editor de excepciones. Puede usar **Opciones de selección** para devolver registros de un usuario, flujo de datos, ID de trabajo, etc. específicos. Puede usar la herramienta **Filtro de campos** para visualizar los registros que cumplen ciertos criterios, mientras que los registros que no los cumplen quedan ocultos.

Si emplea la herramienta **Elementos por página** y luego aplica opciones de selección o un filtro de campos para limitar la lista de registros de excepción que aparece, no olvide que puede haber varias páginas de resultados y no solo lo que se ve en la pantalla inicial. Por ejemplo, digamos que estableció los elementos por página en 10 y luego aplica un filtro de campos para obtener solo los registros que tienen un código postal determinado. Al principio, podría parecer que solo se devolvieron 10 registros, pero puede haber varias páginas de resultados, dado que limitó a 10 los registros.

Uso de las opciones de selección

Seleccione criterios en los campos desplegable del panel de opciones de selección para limitar los registros que ve en la cuadrícula de excepciones.

Usuario	Obligatorio. La ID del usuario asignado al flujo de datos. Solo podrá ver esta información si tiene permisos de modificación.
Nombre de flujo de datos	Obligatorio. El nombre del flujo de datos que generó los registros de excepciones.
Rótulo de etapa	Obligatorio. Nombre definido por el usuario que se le asigna a la etapa Exception Monitor en el flujo de datos. Esta información es particularmente útil cuando el flujo de datos contiene múltiples etapas Exception Monitor. Si la persona que creó el flujo de datos le asignó a cada etapa Exception Monitor un rótulo significativo puede identificar qué Exception Monitor produjo el registro de excepción. El rótulo predeterminado es "Exception Monitor".
ID de trabajo	Obligatorio. Identificador numérico que el sistema le asignó a un trabajo. Cada vez que un trabajo se ejecuta, recibe una nueva ID de trabajo.
Estado	Opcional. El estado de aprobación del registro.
Fecha	Opcional. La fecha (y, opcionalmente, la hora) en que el flujo de datos se ejecutó. Para ingresar la hora, escriba la hora después de la fecha.

Dominio de datos Opcional. El tipo de datos que generó una excepción. Ejemplo de dominios de datos son Nombre, Dirección y Número telefónico. Esta información ayuda a identificar los campos del registro que necesitan ser editados.

Métrica de calidad Opcional. Medida de calidad de la falla del registro. Ejemplos de métricas de calidad son Precisión, Finalización y Singularidad. Esta información ayuda a determinar por qué el registro fue identificado como una excepción.

Uso del filtro de campos

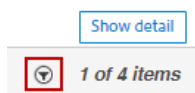
Después de realizar opciones de selección y cargar los registros de excepción, puede usar el filtro de campos para mostrar solo los registros que le interesan. De forma predeterminada, Business Steward Portal solo muestra registros de un flujo de datos de Spectrum™ Technology Platform por vez. Puede filtrar más la lista de registros para ver solo los registros que cumplen ciertos criterios dentro de un campo específico. Después de crear un filtro, este se guarda automáticamente y la siguiente vez que abre el flujo de datos en el Editor de excepciones, se aplica el filtro.

Nota: Debe aplicar las opciones de selección antes de usar la herramienta de filtro de campos.

Puede crear filtros para varios campos, pero solo puede crear un filtro para cada campo. Si un campo ya tiene aplicado un filtro, el fondo de la flecha será azul:

City	FirstName
KEENE	LAKSHMI
KEENE	LAKSHMI

Solo es posible crear filtros en la vista de tabla. Sin embargo, los filtros definidos en la vista de tabla también se reflejan en la vista de formulario. Al igual que en la vista de tabla, en la parte inferior del formulario hay un indicador que señala que se filtró un registro:



Para filtrar la lista de registros:

1. Haga clic en el botón **Filtro**. Verá los íconos de filtro junto a cada encabezado de columna.
2. Haga clic en el ícono de filtro correspondiente al campo cuyos datos desea filtrar.
3. Seleccione un operador que sea adecuado para el tipo de dato del campo, seguido por un valor.

es igual a Busca registros que tengan exactamente el valor especificado. Puede ser un valor numérico o uno de texto. Por ejemplo, puede buscar registros con un valor MatchScore de exactamente 82 o registros con el valor LastName "Smith".

no iguala Busca registros que tienen cualquier valor distinto del especificado. Puede ser un valor numérico o uno de texto. Por ejemplo, puede buscar registros con cualquier valor MatchScore excepto 100 o registros con cualquier valor LastName excepto "Smith".

comienza con	Busca los registros que comienzan con un valor particular en el campo seleccionado. Por ejemplo, si filtra la palabra "Van" en el campo LastName, verá registros con los datos "Van Buren", "Vandenburg" o "Van Dyck".
contiene	Busca registros que contienen el valor especificado en cualquier posición dentro del campo seleccionado. Por ejemplo, si filtra la palabra "South" en el campo AddressLine1, verá los registros con los datos "12 South Ave.", "9889 Southport St.", "600 South Shore Dr." y "4089 5th St. South".
no contiene	Busca registros que no contienen el valor especificado en cualquier posición dentro del campo seleccionado. Por ejemplo, si filtra la palabra "South" en el campo AddressLine1, no verá los registros con los datos "12 South Ave.", "9889 Southport St.", "600 South Shore Dr." y "4089 5th St. South".
finaliza con	Busca los registros que finalizan con un valor particular en el campo seleccionado. Por ejemplo, si filtra los registros que finalizan con "burg" en el campo Ciudad, verá registros con los datos "Gettysburg", "Fredricksburg" y "Blacksburg".
es mayor que	Busca registros que tengan un valor numérico mayor que el valor especificado.
es mayor o igual que	Busca registros que tengan un valor numérico mayor o igual que el valor especificado. Por ejemplo, si especificó 50, verá registros con un valor igual o mayor que 50 en el campo seleccionado.
es menor que	Busca registros que tengan un valor numérico menor que el valor especificado.
es menor o igual que	Busca registros que tengan un valor numérico menor o igual que el valor especificado. Por ejemplo, si especificó 50, verá registros con un valor igual o menor que 50 en el campo seleccionado.
está después de	Busca registros que tengan un valor de fecha u hora que sea posterior al valor especificado.
está después de o es equivalente a	Busca registros que tengan un valor de fecha u hora que sea equivalente o posterior al valor especificado.
está antes de	Busca registros que tengan un valor de fecha u hora que sea anterior al valor especificado.
está antes de o es equivalente a	Busca registros que tengan un valor de fecha u hora que sea equivalente o anterior al valor especificado.

- Haga clic en el botón **Filtro** para aplicar los criterios. Solo aparecerán los registros cuyos datos coincidan con los criterios para ese campo.

- Haga clic nuevamente en el ícono de filtro y luego en **Borrar** para eliminar el filtro. O bien, haga clic en el botón **Filtro** para eliminar todos los filtros. Esta acción se puede llevar a cabo en vista de tabla o vista de formulario.

Visualización de registros

Puede ver los registros de excepción en dos formatos. La vista predeterminada es la vista de tabla, donde puede cargar hasta 100 registros de excepción por página. Puede desplazarse a través de la lista y editar los registros en cualquier orden. Si edita varios registros, puede guardar todas las modificaciones simultáneamente, no es necesario guardar cada registro en forma individual. Con esta vista, puede especificar cuántos registros desea ver por página en el menú desplegable que se encuentra en la parte inferior de la pantalla.

El otro método para visualizar registros de excepción es la vista de formulario, donde puede ver y editar un registro a la vez; en esta vista no es posible editar varios registros al mismo tiempo. Asimismo, debe guardar las modificaciones para cada registro de manera individual, no puede guardar varios registros simultáneamente.

Visualización de detalles de registro

No importa qué vista use, la cuadrícula de excepciones muestra todos los campos para un registro, así como su estado de aprobación, tipo de excepción y cualquier comentario que se haya agregado al registro. En la vista de tabla, puede ver más detalles sobre un registro si hace clic en la flecha que se encuentra en el extremo izquierdo del registro. En la vista de formulario, haga clic en **Mostrar detalles**. Estas acciones abrirán la ficha **Detalle**, que muestra la siguiente información:

ID de trabajo	Identificador numérico que el sistema le asignó a un trabajo. Cada vez que un trabajo se ejecuta, recibe una nueva ID de trabajo.
Agrupar por	Si el flujo de datos fue configurado para arrojar todos los registros en el grupo de registros de excepción, esta función muestra el campo según el cual están agrupados los registros. Esta función solo se aplica a flujos de datos que realizan comparaciones, como los flujos de datos que identifican registros duplicados o flujos de datos que agrupan registros en hogares.
Tiempo de excepción	Fecha y hora en que Exception Monitor identificó el registro como excepción.
Tipo de registro	El tipo de registro que seleccionó. Una de las siguientes: <ul style="list-style-type: none"> • E: excepción • GE: excepción grupal • N—Best of Breed
Estado	El estado del registro que seleccionó. Una de las siguientes: <ul style="list-style-type: none"> • Nuevo • Resuelto (solo posible si el registro se ha editado al menos una vez desde su creación)

Haga clic en la ficha **Condiciones** para ver la siguiente información:

Condición	Nombre de la condición que identificó el registro como una excepción. Los nombres de condición son definidos por la persona que configuró el flujo de datos.
Dominio	El tipo de datos que generó una excepción. Ejemplo de dominios de datos son Nombre, Dirección y Número telefónico. Esta información ayuda a identificar los campos del registro que necesitan ser editados.
Métrica	Medida de calidad de la falla del registro. Ejemplos de métricas de calidad son Precisión, Finalización y Singularidad. Esta información ayuda a determinar por qué el registro fue identificado como una excepción.

Cada vez que se modifica un registro, el sistema conserva cierta información que indica quién cambió el registro, cuándo se modificó, el nombre del usuario al que se asignó y cualquier dato que se haya ingresado en el campo de comentario para dicho registro. El registro que aparece en la cuadrícula de excepción refleja los cambios y comentarios más recientes (si los hubiera). Sin embargo, la ficha **Historial** muestra la siguiente información para la duración completa del registro, desde que se agregó al repositorio como una excepción hasta el momento en que lo está viendo:

Versión	Número de revisión del cambio.
Última modificación por	Usuario que realizó el cambio.
Asignado a	El usuario al que se asignó el registro de excepción al momento de su revisión.
Cuándo	La fecha y la hora en que se guardó el cambio.
Comentarios	Los comentarios (si hay) que ingresó la persona que realizó el cambio.

Orden de campos

Si está usando la vista de tabla, puede ordenar los registros visibles según un campo específico. Para ello, debe hacer clic en cualquier parte de un encabezado de encabezado de columna. Por ejemplo, si desea clasificar los registros en orden alfabético por estado, simplemente haga clic en el encabezado de la columna Estado. El primer clic los ordenará en forma ascendente y el segundo clic los ordenará en forma descendente. Un tercer clic eliminará el orden y los registros regresarán al orden que tenían al principio.

Configuración de campos

Puede seleccionar qué campos aparecen y cambiar el orden en que aparecen si hace clic en el botón **Configurar vista** (la rueda dentada en la parte derecha de la pantalla, bajo el menú desplegable del usuario) y realiza los cambios según corresponda. Estos cambios se guardan en el servidor según el nombre de usuario y del flujo de datos, por lo tanto, cuando abre el flujo de datos posteriormente aún se aplicará la configuración. Del mismo modo, los cambios que aquí realice también afectarán lo que se muestra cuando edita los registros de excepción en la vista de formulario. Use estas funciones de Configurar vista para personalizar los campos que aparecen en el Editor de excepciones:

Activar navegación por teclado

Marque la opción **Optimizar editor de tabla para la navegación por teclado** para usar la tecla Tabulación para navegar por los registros de excepción en la Vista de tabla del Editor de excepciones.

Búsqueda de campos en Configurar vista

Ingrese total o parcialmente un nombre de campo en la casilla de búsqueda y la lista de campos disponibles se actualizará dinámicamente. La búsqueda no distingue mayúsculas de minúsculas.

Cómo ocultar campos de la visualización

Si no desea ver todos los campos de un registro de excepción, haga clic en **Configurar vista** y desmarque los campos que desea ocultar. La lista se mostrará en el mismo orden que se visualiza en la cuadrícula Excepciones.

Cambio de orden de campos

Puede cambiar el orden en que se muestran los campos si arrastra y suelta los campos para organizarlos. Sin embargo, no puede reordenar los campos mientras ve los resultados de la búsqueda. Debe borrar la ventana de búsqueda y seleccionar "Todo" para recuperar la capacidad de reordenar campos.


Nota: Si utiliza la vista de tabla, puede arrastrar y soltar los encabezados de columna directamente desde el Editor de excepciones para cambiar el orden en que se muestran los campos; no necesita hacerlo desde Configurar vista.

Configurar el diseño Vista de formulario

Puede reemplazar el diseño predeterminado de los campos de la Vista de formulario seleccionando las columnas que deben aparecer, pero también especificando el ancho de esas columnas. La única regla para hacerlo así es que el ancho de columna de todos los campos visibles deben sumar un total de 12. Por ejemplo, si quisiera ubicar cuatro campos uno al lado del otro en una sola fila, podría asignar a cada campo un ancho de columna de "3". Asimismo, si solo quisiera incluir dos campos, podría ingresar "6" como el ancho de columna para cada campo. O bien, si quisiera incluir cinco campos de dirección en una fila, deberá especificar un número de columnas para cada uno de esos campos de modo que sumen 12. La primera imagen que se ve a continuación, muestra cómo podría configurar cinco campos; la segunda imagen, muestra un ejemplo de cómo queda dicha configuración.


Nota: Cuando se reemplaza el diseño predeterminado de cualquier campo o combinación de campos, esos campos ya no responderán al tamaño de la ventana; sin embargo, los campos con valores predeterminados seguirán siendo sensibles de acuerdo al tamaño de ventana del navegador.

Configure view




Filter  All Visible Hidden

Field name	Form columns
<input checked="" type="checkbox"/> Approved	1
<input type="checkbox"/> Type	Default
<input checked="" type="checkbox"/> Address1	4
<input type="checkbox"/> FirmName	Default
<input checked="" type="checkbox"/> City	4
<input checked="" type="checkbox"/> StateProvince	1
<input checked="" type="checkbox"/> PostalCode	2

Optimize tabular editor for keyboard navigation

[Reset to default](#) 

OK Cancel

   Filter Calendar Table Settings

Approved

Address1	City	StateProvince	PostalCode
1594 Spring St.	Smyrna	<input type="text"/>	30080

[Show detail](#)

« < > » 1 of 11 items

Edición de un registro de excepción

La finalidad de editar un registro de excepción es corregir o ampliar y aprobar el registro para que pueda ser procesado con éxito. Editar un registro de excepción puede implicar la consulta de otros servicios de Spectrum Technology Platform, o bien recursos externos, tales como mapas, Internet u otros sistemas de información en su empresa.

Después de revisar los registros, puede editarlos y aprobarlos directamente en el Editor de excepciones. Puede editar uno o más registros a la vez en la vista de tabla y editar los registros uno a uno en la vista de formulario.

Tome en cuenta que los campos de solo lectura no pueden editarse. Si desea hacer que un campo de solo lectura sea editable, necesita eliminar todos los registros de excepciones para ese flujo de datos e ID de trabajo, y ejecutar el flujo de datos nuevamente luego de configurar los campos en conformidad en la etapa Write Exceptions. Esto produciría nuevos registros de excepción con campos editables. Además, no puede editar un campo para que contenga un valor que no coincide

con el tipo de dato. Por ejemplo, no puede editar un campo con un tipo de dato numérico para que contenga letras.

Vista de tabla

Para editar un campo de un solo registro en la vista de tabla:

1. Haga clic en el campo que desea editar y cambie el valor. Los campos de solo lectura se verán atenuados. Haga clic derecho en el campo para tener acceso a las opciones de cortar, copiar, y pegar. Cuando haya editado un campo, notará que aparece un triángulo verde en la esquina superior izquierda del campo. Esta es una indicación visual para recordarle que se cambió el valor del campo, pero aún no se ha guardado.
2. Marque la casilla **Aprobado** para el registro modificado. Así, se marca el registro como listo para ser procesado por Spectrum™ Technology Platform.
3. Si necesita deshacer un cambio, seleccione el registro que desea y haga clic en el botón **Deshacer cambios**.
4. Haga clic en el botón **Guardar** cuando haya terminado de editar los registros.

Para editar un campo de varios registros en la vista de tabla:

1. Mientras presiona la tecla Ctrl o Mayús, haga clic en el campo que desea cambiar para todos los registros seleccionados. Por ejemplo, si desea cambiar todas las apariciones de "L.A." y decir "Los Ángeles", presione la tecla Ctrl y haga clic dentro del campo Ciudad para uno de los registros. Entonces, mientras está presionando Ctrl o Mayús, haga clic dentro de ese mismo campo para los otros registros que desea cambiar.
2. Cambie, conforme a esto, los valores del campo. Puede editar estos campos, pero hágalo con cuidado, ya que los cambios que realice aquí se aplicarán a todos los registros seleccionados, aún cuando hayan variado previamente los valores de dichos campos. De igual manera, si borra los datos de múltiples registros, se borrarán los datos de todos los registros seleccionados.
3. Marque la casilla **Aprobado** para los registros modificados. Así, se marca el registro como listo para ser procesado por Spectrum™ Technology Platform. O bien, puede hacer clic en el botón **Aprobar todos**; esto causará que todos los registros de excepción simple aparezca en el editor para aprobación. (La función Aprobar todos no se aplica a registros excepción que coincidan).
4. Si necesita deshacer un cambio, seleccione los registros que desea y haga clic en el botón **Deshacer cambios**.
5. Haga clic en el botón **Guardar** cuando haya terminado de editar los registros. Los cambios en los registros se guardan en el repositorio de excepciones y se actualiza la vista. Si no ha definido un flujo de trabajo de revalidación, se vuelve a cargar la lista de excepciones. Sin embargo, uno o más de los registros editados podría no aparecer en la lista actualizada, debido a que ya no coincide con los criterios de búsqueda o filtro. Si *definió* un flujo de trabajo de revalidación, los registros editados podrían no aparecer en la lista actualizada si ahora son válidos y se han purgado del repositorio.
6. Use los botones de navegación que se encuentran en la parte inferior de la pantalla para ir a la página anterior o siguiente de registros de excepción. También puede usar estos botones para ir directamente al primer o último registro de excepción.

Vista de formulario

Para editar registros con la vista de formulario, siga estos pasos:

1. Haga clic en **Vista de formulario**. Aparecerá el primer registro del conjunto.
2. Haga clic en el campo que desea editar. y cambie, conforme a esto, el valor del campo. Los campos de solo lectura se verán atenuados. Haga clic derecho en el campo para tener acceso a las opciones de cortar, copiar, y pegar. Cuando haya editado un campo, notará que el contorno de ese campo cambia a verde. Esta es una indicación visual para recordarle que se cambió el valor del campo, pero aún no se ha guardado.
3. Puede agregar comentarios sobre sus cambios en la columna **Comentarios**. Los comentarios son visibles para otros usuarios y pueden ser usados para ayudar a mantener el seguimiento de los cambios que se le hacen al registro.
4. Cuando esté seguro de haber realizado todos los cambios para que el registro sea válido, marque la casilla **Aprobado**. Así, se marca el registro como listo para ser procesado por Spectrum™ Technology Platform.
5. Si necesita deshacer un cambio, haga clic en el botón **Deshacer cambios**.
6. Haga clic en **Guardar**. Los cambios en el registro se guardan en el repositorio de excepciones y se actualiza la vista. Verá el mismo registro que contiene sus cambios o bien el siguiente registro de la lista, en caso de que el registro que cambió ya no esté disponible o ya no coincida con los criterios de búsqueda o filtro.
7. Use los botones de navegación que se encuentran en la parte inferior de la pantalla para ir al registro de excepción anterior o siguiente. También puede usar estos botones para ir directamente al primer o último registro de excepción.

Resolución de registros duplicados

Las excepciones de resolución de duplicados ocurren cuando Spectrum™ Technology Platform no puede determinar con seguridad si un registro es un duplicado de otro. Hay dos formas de resolver registros duplicados.

Nota: Los registros duplicados solo se pueden resolver con la función Resolver duplicados en la vista de tabla. Sin embargo, puede editar esos registros en la vista de formulario.

Una forma es agrupar los registros duplicados en colecciones. Cuando ya se aprobaron los registros, se los puede procesar mediante un proceso de consolidación para eliminar los registros duplicados en cada colección de sus datos.

Otra forma es editar los registros para que sean más fáciles de reconocer como duplicados, por ejemplo, corrigiendo la ortografía del nombre de la calle. Cuando aprueba los registros, Spectrum™ Technology Platform vuelve a procesar los registros mediante una comparación y un proceso de consolidación. Si corrigió los registros con éxito, Spectrum™ Technology Platform podrá identificar el registro como un duplicado.

Yet another approach to resolving duplicate records is to create a best of breed record. This combines the other two approaches by managing record collections and then editing one of the records in the

collection to include fields from both the original and duplicate records. Este "súper" registro se conoce como el registro "best of breed" (el mejor de la especie).

Conversión de un registro en el duplicado de otro

Los registros duplicados se muestran como grupos de registros en Business Steward Portal. Puede convertir un registro en un duplicado de otro moviéndolo al mismo grupo donde está el registro duplicado.

Para convertir un registro en un duplicado:

1. Seleccione el registro con el que quiere trabajar y haga clic en **Resolver duplicados**.

La vista **Resolver duplicados** muestra registros duplicados. Los registros están agrupados en colecciones o grupos candidatos que contienen estos tipos de registros de cruce:

suspect	Registro con el que se comparan otros registros para determinar si son duplicados entre sí. Cada colección tiene un solo registro sospechoso.
duplicate	Registro que es un duplicado del registro sospechoso.
unique	Registro que no tiene duplicados.

Puede determinar el tipo de un registro mirando la columna MatchRecordType.

2. Si es necesario, corrija los registros individuales. Para obtener más información, consulte [Edición de un registro de excepción](#) en la página 264. También puede arrastrar y soltar registros entre los grupos.
3. En el campo CollectionNumber o CandidateGroup, ingrese el número del grupo al que quiere mover el registro. El registro se convierte en un duplicado de otros registros del grupo.

En algunos casos no es posible mover un registro con un valor MatchRecordType del tipo "sospechoso" a otra colección de duplicados.

Nota: Los registros son agrupados de acuerdo con los campos CollectionNumber o CandidateGroup, según la lógica de comparación utilizada en el flujo de datos que produjo las excepciones. Comuníquese con su administrador de Spectrum™ Technology Platform si desea obtener más información sobre comparaciones.

4. Cuando termine de modificar los registros, marque la casilla **Aprobado**. lo que indica que el registro está listo para ser nuevamente procesado por Spectrum™ Technology Platform.
5. Para guardar los cambios, haga clic en el botón **Guardar**.

Creación de un nuevo grupo de registros duplicados

En algunas situaciones, puede crear un nuevo grupo de registros que desea convertir en duplicados entre sí. En otras situaciones no puede crear grupos nuevos. Su capacidad para crear grupos nuevos está determinada por el tipo de procesamiento de Spectrum™ Technology Platform que generó los registros de excepción.

1. Seleccione el registro con el que quiere trabajar y haga clic en **Resolver duplicados**.

La vista **Resolver duplicados** muestra registros duplicados. Los registros están agrupados en colecciones o grupos candidatos que contienen estos tipos de registros de cruce:

suspect	Registro con el que se comparan otros registros para determinar si son duplicados entre sí. Cada colección tiene un solo registro sospechoso.
duplicate	Registro que es un duplicado del registro sospechoso.
unique	Registro que no tiene duplicados.

Puede determinar el tipo de un registro mirando la columna MatchRecordType.

- Si es necesario, corrija los registros individuales. Para obtener más información, consulte [Edición de un registro de excepción](#) en la página 264.
- Seleccione un registro que desea colocar en la nueva colección, y luego haga clic en **Nueva colección**. La nueva colección recibe automáticamente un número de colección único y el registro que usted seleccionó se convierte en sospechoso.

Nota: Si no ve el botón **Nueva colección**, no puede crear una nueva colección para los registros con los que está trabajando. Solo puede crear nuevas colecciones si el flujo de datos que produjo las excepciones incluía una etapa Interflow Match o Intraflow Match, pero no si incluía una etapa Transactional Match. Comuníquese con su administrador de Spectrum™ Technology Platform si desea obtener más información sobre estas etapas de comparación.

- Coloque registros adicionales en la colección ingresando el número de colección nuevo en el campo CollectionNumber del registro.
- Cuando termine de modificar los registros, marque la casilla **Aprobado**. lo que indica que el registro está listo para ser nuevamente procesado por Spectrum™ Technology Platform.
- Para guardar los cambios, haga clic en el botón **Guardar**.

Conversión de un registro en único

Para convertir un registro en único:

- En el campo MatchRecordType, ingrese "Único".
- Cuando termine de modificar los registros, marque la casilla **Aprobado**. lo que indica que el registro está listo para ser nuevamente procesado por Spectrum™ Technology Platform.
- Para guardar los cambios, haga clic en el botón **Guardar**.

Campos ajustados automáticamente durante la resolución duplicada

Cuando modifica registros en la vista de resolución duplicada de Business Steward Portal, algunos campos se ajustan automáticamente para reflejar la nueva disposición del registro.

Tabla 21: Registros procesados por Interflow o Intraflow Match

Acción	Valores automáticamente aplicados a los campos
Movimiento de un registro de una colección a otra.	<p>Si mueve un registro a una colección de duplicados:</p> <ul style="list-style-type: none"> • MatchRecordType: Duplicado • MatchScore: 100 • HasDuplicates: D (Este campo solo está presente si el flujo de datos contenía una etapa Interflow Match). <p>Si mueve un registro duplicado a una colección de registros únicos (colección 0):</p> <ul style="list-style-type: none"> • MatchRecordType: Único • MatchScore: Sin cambios • HasDuplicates: U (Este campo solo está presente si el flujo de datos contenía una etapa Interflow Match). <p>Si mueve un registro sospechoso a una colección de registros únicos (colección 0):</p> <ul style="list-style-type: none"> • MatchRecordType: Único • MatchScore: 0 • HasDuplicates: N (Este campo solo está presente si el flujo de datos contenía una etapa Interflow Match).
Creación de una nueva colección	<ul style="list-style-type: none"> • MatchRecordType: Sospechoso • MatchScore: Sin valor • HasDuplicates: Y (Este campo solo está presente si el flujo de datos contenía una etapa Interflow Match). <p>Nota: Si el registro provino de un flujo de datos que contenía una etapa Interflow Match, solo los registros con un valor "input_port_0" en el campo InterflowSourceType pueden ser sospechosos.</p>

Tabla 22: Registros procesados por Transactional Match

Acción	Valores automáticamente aplicados a los campos
Cambio de MatchRecordType a duplicado	<ul style="list-style-type: none"> • HasDuplicates: D • MatchScore: 100
Cambio de MatchRecordType a único	<ul style="list-style-type: none"> • HasDuplicates: U • MatchScore: no cambiado

Acción Valores automáticamente aplicados a los campos


Cambio de HasDuplicates a D	<ul style="list-style-type: none"> • MatchRecordType: Duplicado • MatchScore: 100
Cambio de HasDuplicates a U	<ul style="list-style-type: none"> • MatchRecordType: Único • MatchScore: no cambiado
Cambio de HasDuplicates a Y	<ul style="list-style-type: none"> • MatchRecordType: Sospechoso • MatchScore: En blanco
Cambio de HasDuplicates a N	<ul style="list-style-type: none"> • MatchRecordType: Sospechoso • MatchScore: En blanco

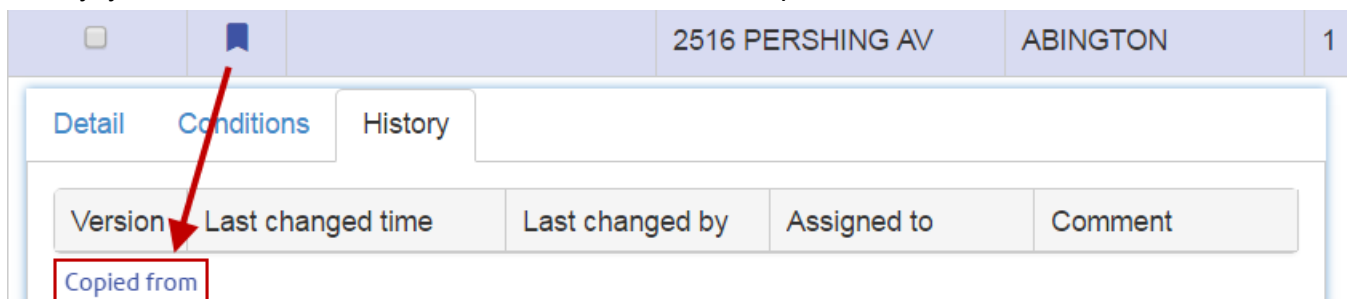
Creación de un registro Best of Breed.

Follow these steps to create a best of breed record.

Nota: Best of breed records can only be added to non-zero collections. Also, a best of breed record must be created for each collection within a group before you can approve all the records within that group.

1. To create a best of breed record within a collection, select a record in that collection and click

the **Add Best of Breed Record** button (). This will make a copy of the record you selected and place it just after that record in the Exception Editor. When you save the best of breed record, it will move to the bottom of the record collection. If you look at the best of breed record's history, you will see a link to the record from which it was copied:



Nota: You can undo this by selecting the best of breed record and then clicking the

Remove Best of Breed Record button ().

2. To combine fields into the best of breed record, drag and drop fields from the duplicate records into the best of breed record. Note that there are some limitations to drag-and-drop in this environment:

- The cell you are dragging from one record to another cannot be in an "edit" state (though the cell where you are dropping it can be).
- When you drop one cell onto another cell (whether in the same row or a different row) the second cell will take on the value of the first cell.
- You cannot drag and drop Boolean-type cells.
- You can drag non-string-type cells to string-type cells (such as date fields), but you cannot drop string-type cells to non-string-type cells.
- You cannot drop a cell onto another cell that uses lookup values (such as combination box editors), but you can drag a value from these cells to other string-type cells.
- You cannot drop a cell onto a read-only cell.

See [Write Exceptions](#) options for more information on best of breed functionality.

Uso de las herramientas de búsqueda

El Editor de excepciones de Business Steward Portal proporciona herramientas de búsqueda que facilitan la búsqueda de información que podrían ayudarlo a editar y aprobar registros de excepción y volver a ejecutarlos correctamente. Las herramientas incluyen mapas Bing y los servicios para los cuales tiene licencia en Spectrum™ Technology Platform. Las herramientas con íconos que tienen un signo más en el medio son herramientas avanzadas que implican gastos cuando se usan.

Nota: En primero lugar, estas herramientas de búsqueda se deben configurar en la página de configuración de Business Steward de Management Console. Si no se selecciona ninguno en esa página, no aparecerá en esta página. Además, independientemente de las herramientas seleccionadas en Management Console, los servicios que le aparecen al usuario en esta página, se limitan a los que tiene permisos de visualización.

Uso de Mapas Bing

Bing Maps muestra la ubicación de una dirección en un mapa y ofrece controles que permiten acercar o hacer una vista panorámica del mapa. Además, puede hacer clic en el mapa para obtener direcciones.

Nota: La herramienta de búsqueda Bing Maps es proporcionada por Microsoft; debe estar conectado a Internet para usar este servicio.

1. Haga clic en el registro que desea investigar.
2. Bajo la tabla de registros, haga clic en **Herramientas de búsqueda** para ampliar la vista.
3. En el campo **Herramientas**, seleccione **Mapas Bing**.
4. Seleccione los campos que desea utilizar en la búsqueda en la columna **Nombre de campo**. Por ejemplo, si desea buscar una dirección en el mapa, podría elegir AddressLine1 y City. Si desea ver solo la ciudad en un mapa, puede seleccionar solo City y StateProvince.
5. Seleccione "Camino" para una vista de mapa tradicional, "Vista aérea" para una vista desde arriba y "Automático" para dejar que Bing Maps decida cuál es más apropiada.

6. Haga clic en **Ir**. Los resultados, incluidas latitud y longitud, se muestran en la casilla **Resultados** y en el mapa como un marcador. Haga clic en los botones Girar para cambiar la perspectiva en 90 grados. Haga clic en las flechas de la brújula para cambiar el enfoque de manera gradual en la dirección seleccionada. Use los botones Acercar y Alejar para enfocar más o menos cerca. También puede mover el marcador arrastrándolo a una nueva ubicación; la información del mapa se actualizará dinámicamente.
7. Para obtener la dirección de otros edificios, haga clic en el mapa. Cambiar a la vista aérea puede ser útil para encontrar edificios.

Después de finalizar una búsqueda inicial en el mapa, si hace clic en otro registro de excepción y en el botón **Ir a**, el mapa se actualizará según corresponda.

Uso de la herramienta de búsqueda del servicio Spectrum

Las herramientas de búsqueda del servicio Pitney Bowes incluyen todos los servicios para los cuales tiene licencia, como ValidateAddress, GetPostalCodes, etc. Puede utilizar estos servicios en Exception Editor para buscar y validar datos de excepción que está intentando corregir o expandir.

Tenga en cuenta que cuando utilice esta función, solo verá los servicios si tiene los permisos de vista para los servicios en el grupo de la plataforma para la seguridad del rol. Del mismo modo, a fin de ejecutar los servicios, deberá ejecutar los permisos para los servicios. Sin embargo, estos permisos pueden ser modificados al utilizar anulaciones de entidad protegida. Mediante una combinación de permisos y anulaciones de nivel superior, el administrador puede administrar la lista de servicios a la cual tiene acceso un usuario o rol particular en la lista desplegable de servicios del portal BSM.

1. Seleccione el registro que contiene los datos que desea buscar.
2. Bajo el Editor de excepciones, haga clic en **Herramientas de búsqueda**.
3. En el campo **Servicio**, seleccione el servicio que desea utilizar, como ValidateAddress o GetCandidateAddresses.
4. Si el registro de excepciones contiene campos utilizados en ese servicio, pero con nombres distintos, asigne los campos de entrada del servicio a los campos de excepción en la ficha **Campos de servicio**. Por ejemplo, si utiliza ValidateAddress y su registro de excepción no incluye un campo AddressLine1 pero sí un campo AddrLine1, seleccione "AddrLine1" en la columna del campo de excepción de la fila AddressLine1. Debe tener al menos un campo de entrada asignado antes de ejecutar el servicio.

Nota: El Business Steward Portal recuerda los mapas que creó de los campos de servicio a los campos de excepción, siempre y cuando asigne los registros de excepción con el mismo nombre de campo, nombre de etapa y nombre de flujo de datos. Por ejemplo, si el registro de excepción tiene un campo llamado "AddrLine1" y usted lo asigna como "AddressLine1", recordará esta asignación, siempre y cuando trabaje con registros que contengan "AddrLine1" y que fueron creados en la misma etapa por el mismo flujo de datos.

5. Repita el paso 4 para los campos de salida del servicio. Este paso es opcional, pero debe tener al menos un campo de salida asignado antes de aplicar los datos del servicio. Tenga en cuenta

que no se puede asignar un campo de excepción de solo lectura a los campos de salida del servicio.

- Haga clic en la ficha **Opciones** para seleccionar los recursos de bases de datos y ver y cambiar las opciones de servicio que se definieron en Management Console. Si no sabe cuál es la finalidad de una determinada opción, suspenda el mouse sobre esa opción para ver su descripción. Los cambios que haga aquí se mantendrán cuando los utilice el mismo usuario, flujo de datos y etapa del registro de excepción.

Nota: Si el servicio que está utilizando requiere una base de datos, en primer lugar debe configurar el recurso de la base de datos en Management Console. Por ejemplo, si está revisando registros de los Estados Unidos por medio de Validate Address, debe tener configurada una base de datos de Estados Unidos en Management Console

- Haga clic en **Ejecutar servicio**. El registro actualizado aparecerá en la ficha **Resultado** con un código de estado que indica el éxito del registro. Los campos asignados a registros de excepción se indicarán con un asterisco.
- Seleccione el registro de resultado y haga clic en **Aplicar datos de servicio** para transferir los datos a los campos asignados del registro de excepción.
- Si desea procesar el registro actualizado, haga clic en la casilla de verificación **Aprobado** del registro y luego en **Guardar**.

La página Administrar

La página Administrar permite que un usuario con permisos de visualización y modificación pueda revisar y administrar actividades de registro de excepción para todos los usuarios asignados. También proporciona la capacidad de reasignar registros de excepción de un usuario a otro. Si tiene permisos de eliminación, puede eliminar el grupo completo de registros de excepción desde el sistema, en base al nombre del flujo de datos y la ID del trabajo.

Revisión de la actividad de registros de excepción

La sección **Estado** de la página Administrar excepciones muestra la actividad de los registros de excepción por asignación o nombre de flujo de datos. Puede especificar cuál mostrar si hace clic en el botón "Asignaciones" o "Flujos de datos", que se encuentra cerca de la esquina superior derecha de la pantalla. "Asignaciones" proporciona la cantidad de registros de excepción asignados a cada usuario, así como también cuántos de dichos registros se han aprobado y cuántos quedan. It also shows the number of non-exception records that exist within groups that also contain exception records. Note that non-exception data will appear only if there are non-exception records within a group also containing exception records and if there are no security overrides preventing this data from being shown.

"Flujos de datos" proporciona el porcentaje de registros que han sido aprobados para cada flujo de datos. For dataflows containing more than one Exception Monitor stage, you can further break down this information for each of those stages. For example, if a dataflow contains two Exception Monitor stages, "NameExceptions" might be responsible for 37% of all exceptions in the dataflow and "DataExceptions" might be responsible for the other 63%. In this situation, the status of each **dataflow**

is still calculated as a whole as it relates to all dataflows in the repository, while the status of each **stage** is calculated based on the total number of exceptions for the selected dataflow.

Puede filtrar la información que se muestra si ingresa criterios de búsqueda en la fila Filtro. La lista se completa automáticamente en forma dinámica con los flujos de datos o los usuarios que coincidan con las letras que se escriben.

La sección **Progreso** de esta página muestra el progreso acumulativo de todos los usuarios o todos los flujos de datos cuyos estados individuales se muestran en el Tablero. Seleccione “Asignaciones” en la parte superior de la pantalla para ver el progreso del usuario; seleccione “Flujos de datos” para ver progreso de flujo de datos. Utilice la escala para seleccionar las métricas para la escala (día, semana, mes) y el número de unidades para mostrar.

Asignación de registros de excepciones

La sección **Asignación** de la página Administrar excepciones (en **Mantenimiento**) le permite reasignar registros de excepciones de un usuario a otro.

1. Seleccione el usuario cuyas excepciones desea asignar a otro usuario en el campo **Usuario**.
2. Para reasignar todos los registros de excepciones que pertenecen a un usuario, salte al Paso 4. Para reasignar una parte de los registros de excepciones de un usuario, complete uno o más de estos campos:
 - Nombre de flujo de datos: el nombre del flujo de datos que produce los registros de excepciones.
 - Rótulo de etapa: el nombre de la etapa que produce los registros de excepciones.
 - ID de trabajo: la ID asignada al trabajo que contiene los registros de excepciones.
 - Dominio de datos: el tipo de **datos** asignados en Exception Monitor.
 - Métricas de calidad: el tipo de **métrica** asignada a Exception Monitor.
 - Fecha/hora desde: la fecha de inicio en un intervalo de fechas en el que se crearon los registros de excepciones.
 - Fecha/hora hasta: la fecha de finalización en un intervalo de fechas en el que se crearon los registros de excepciones.
 - Estado de aprobación: establece si se han aprobado o no los registros de excepciones.
3. Después de realizar las selecciones en los campos **Usuario**, **Nombre de flujo de datos** y **Rótulo de etapa** (como mínimo), puede refinar más el filtro según los valores del campo de excepción.
 - a) Haga clic en el icono Agregar filtro de campo.
 - b) En la columna **Nombre de campo** seleccione el campo que quiere filtrar.
 - c) En la columna **Operador** seleccione una de las siguientes opciones:

es igual a	Busca registros que tengan exactamente el valor especificado. Puede ser un valor numérico o uno de texto. Por ejemplo, puede buscar registros con un valor MatchScore de exactamente 82 o registros con el valor LastName "Smith".
-------------------	--

- | | |
|-----------------------------|---|
| no iguala | Busca registros que tienen cualquier valor distinto del especificado. Puede ser un valor numérico o uno de texto. Por ejemplo, puede buscar registros con cualquier valor MatchScore excepto 100 o registros con cualquier valor LastName excepto "Smith". |
| es mayor que | Busca registros que tengan un valor numérico mayor que el valor especificado. |
| es mayor o igual que | Busca registros que tengan un valor numérico mayor o igual que el valor especificado. Por ejemplo, si especificó 50, verá registros con un valor igual o mayor que 50 en el campo seleccionado. |
| es menor que | Busca registros que tengan un valor numérico menor que el valor especificado. |
| es menor o igual que | Busca registros que tengan un valor numérico menor o igual que el valor especificado. Por ejemplo, si especificó 50, verá registros con un valor igual o menor que 50 en el campo seleccionado. |
| contiene | Busca registros que contienen el valor especificado en cualquier posición dentro del campo seleccionado. Por ejemplo, si filtra la palabra "South" en el campo AddressLine1, verá los registros con los datos "12 South Ave.", "9889 Southport St.", "600 South Shore Dr." y "4089 5th St. South". |
| no contiene | Busca registros que no contienen el valor especificado en cualquier posición dentro del campo seleccionado. Por ejemplo, si filtra la palabra "South" en el campo AddressLine1, pero selecciona "no contiene", no verá los registros con los datos "12 South Ave.", "9889 Southport St.", "600 South Shore Dr." y "4089 5th St. South". |
| comienza con | Busca los registros que comienzan con un valor particular en el campo seleccionado. Por ejemplo, si filtra la palabra "Van" en el campo LastName, verá registros con los datos "Van Buren", "Vandenburg" o "Van Dyck". |
| finaliza con | Busca los registros que finalizan con un valor particular en el campo seleccionado. Por ejemplo, si filtra los registros que finalizan con "burg" en el campo Ciudad, verá registros con los datos "Gettysburg", "Fredricksburg" y "Blacksburg". |
- d) En la columna **Valor de campo** ingrese el valor que usará como criterio de filtro.
- Nota:** El valor de búsqueda distingue mayúsculas de minúsculas. Esto significa que la búsqueda de SMITH arrojará solo los registros con la palabra "SMITH" en mayúsculas, pero no "smith" ni "Smith".
- e) Para filtrar más de un campo, agregue varios filtros haciendo clic nuevamente en el ícono de agregar filtro de campo. Por ejemplo, si quiere todos los registros que contienen el valor de LastName "SMITH" y el valor de Estado "NY", puede utilizar dos filtros, uno para el campo LastName y uno para el campo Estado.

Este ejemplo arrojaría todos los registros con un valor "FL" en el campo Estado:

Field name	Operator	Value
State	Is equal to	FL

Este ejemplo arrojaría todos los registros que no tienen el valor 60510 de PostalCode:

Field name	Operator	Value
StateProvince	Is equal to	NY
PostalCode	Is not equal to	14226

Este ejemplo arrojaría todos los registros con StateProvince "NY" con todos los códigos postales, excepto 14226.

Field name	Operator	Value
PostalCode	Is not equal to	60510

- Haga clic en **Reasignar**.
- Opcional. Seleccione el número de excepciones que desea reasignar. Puede asignar todas o algunas de las excepciones resultantes a un nuevo usuario. Por ejemplo, si ingresa "10" como el límite, solo se reasignarán los 10 primeros registros que cumplan los criterios.
- Seleccione otro usuario en la lista desplegable **Reasignar**.
- Haga clic en **Confirmar**.

Eliminación de registros de excepción

Cada cierto tiempo, elimine los registros de excepción del repositorio. Por ejemplo, puede tener registros residuales de pruebas realizadas en el sistema o registros que se consideraron erróneamente excepciones después del procesamiento, o bien, quizás desee procesar y eliminar primero los registros aprobados y luego volver a ejecutar el mismo trabajo. Puede hacer esto en la sección **Purgar** de la página Administrar (bajo **Mantenimiento**).

Debe realizar una selección en los campos **Nombre del flujo de datos** y **Id. de trabajo** antes de hacer clic en **Eliminar**. Sin embargo, puede seleccionar "Todos" en el campo Id. de trabajo para eliminar los registros de excepción de todos los trabajos que ejecute el flujo de datos seleccionado. Haga clic en **Eliminar datos de informe de calidad de datos** para eliminar todos los datos de rendimiento. Si esta opción no está seleccionada, los registros de excepción del trabajo se quitarán del repositorio, pero los datos de rendimiento seguirán apareciendo en la página de calidad de datos.

Página de calidad de datos

La página de calidad de datos de Business Steward Portal entrega información acerca de las tendencias en los registros de excepción.

Nota: Esta página se desactivará si apaga la generación de informes de calidad de datos en la página de Business Steward Settings de Management Console.

Identificar las tendencias

The Data Quality Trends page depicts the following statistical information by dataflow and domain:

- Número total de registros procesados
- Cantidad total de registros de excepciones
- Porcentaje de registros que se procesan correctamente
- La tendencia de los datos en intervalo de 30 días

This information can be further broken down by stage if you select a dataflow or by metric if you select a domain. The values that appear here are determined by dataflows with Exception Monitor stages that have been run over the last 30 days as well as their respective domains and metrics.

- If you are viewing information by domain, select a **Dataflow name** if you want to view information for a specific dataflow and a **Stage label** if you want to view information for a specific stage. Note that you must select a single dataflow if you want to also filter the results based on a stage; otherwise, you will see data for all dataflows.
- If you are viewing information by dataflow, select a **Domain** if you want to view information for a specific domain and a **Metric** if you want to view information for a specific metric. De lo contrario, verá todos los datos de flujos.
- Seleccione una duración para que la **Escala** especifique cuán atrás desea que lleguen los datos. Puede seleccionar cualquier número de días, semanas o meses.
- Select a dataflow if you want to break out results by stage.
- Select a domain if you want to break out results by data quality metric.

As you make selections to view different parts of the data, you will notice that charts appear reflecting the status of that particular data. For instance, if you look at domain data, you could narrow that data down by a specific domain (such as "Address") and then by a particular metric (such as "Completeness"). The charts will update accordingly as you make different selections.

Business Steward Settings

Introducción

Introducción a Business Steward Settings

Business Steward Settings brinda las siguientes herramientas para usuarios con permisos de escritura:

- **Búsquedas**: brinda un método para corregir registros de excepción con un conjunto específico de datos para esas correcciones.
- **Dominios**: permiten especificar el tipo de datos que se está evaluando.
- **Métricas**: permiten especificar la forma en que se miden los datos.

- **Notificaciones:** permite que el sistema le envíe un mensaje a una o más direcciones de correo electrónico cuando una cantidad designada de excepciones están ligadas a un dominio o métrica específica.
- **Informe de calidad de datos:** establece las preferencias para hacer un seguimiento de las condiciones de aprobado/desaprobado y los KPI.
- **Servicios de herramientas de búsqueda:** establece las preferencias para herramientas de búsqueda en el Editor de excepciones de Business Steward Portal.
- **Opciones:** establece las preferencias para registros de auditorías y seguimiento de progreso.

Acceso a Business Steward Settings

Para acceder a Business Steward Settings:

1. En un navegador web, vaya a este URL:

`http://server:port/managementconsole`

Donde *servidor* es la dirección IP o el nombre de servidor de su servidor Spectrum™ Technology Platform y *puerto* es el puerto HTTP que usa Spectrum™ Technology Platform. De manera predeterminada, el puerto HTTP es 8080.

2. Ingrese un nombre de usuario y contraseña válidos.
3. Haga clic en el botón **Recursos**.
4. Seleccione **Business Steward Settings**.

Búsquedas

La herramienta de búsqueda ofrece una manera para seleccionar un campo específico de una lista de valores durante la actualización de registros en el editor de excepciones. Esta función es particularmente útil cuando hay varios registros con datos en el mismo campo que desea cambiar. Por ejemplo, podría tener un conjunto de registros de excepción que contienen datos bancarios. Uno de los campos de dichos datos podría estar compuesto por códigos que representan qué tipo de cuenta está vinculada con el registro (1 = cheque, 2 = ahorro, 3 = mercado monetario, etc.). cuyas direcciones incluyan códigos ISO en vez de nombres en el campo Country, lo que causa que no se puedan validar esas direcciones. Para corregir esto, puede crear una búsqueda que brinde códigos ISO con sus nombres de país correspondientes y hacer las correcciones en el Editor de excepciones, donde debe seleccionar el código ISO de una lista que luego completa el campo con el nombre del país ligado a ese código ISO.

Otro beneficio de utilizar esta herramienta es que limita las opciones disponibles para hacer correcciones, lo que reduce la posibilidad de nuevos errores. Para continuar con el ejemplo de los nombres de país incorrectos o faltantes, si crea una búsqueda que brinde una lista de nombres de país en vez de requerir que se ingresen esos nombres manualmente para cada registro de excepción, se asegura de que esos nombres estén escritos correctamente, lo cual incrementa las posibilidades de validación al volver a procesarse.

¿Qué es el proceso de búsqueda?

El proceso de búsqueda consiste en tres pasos luego de haber revisado excepciones e identificado un problema recurrente entre esas excepciones (como datos no válidos en un campo de país):

- Cree la búsqueda utilizando valores y rótulos de datos correctos que van a sobrescribir los datos incorrectos.
- Con la etapa Write Exceptions en el flujo de datos que está produciendo los registros de excepción, indique el campo problemático en la búsqueda que creó y vuelva a ejecutar el flujo de datos.
- Corrija los registros de excepción en el Editor de excepciones de Business Steward Portal; para eso, sobrescriba los datos incorrectos en el campo problemático con datos correctos de la búsqueda.

Creación de búsquedas

Una búsqueda está compuesta de valores o pares de valor/rótulo que contienen datos para remplazar datos problemáticos existentes en un flujo de datos que está produciendo registros de excepción. El valor es lo que remplazará los datos problemáticos, y el rótulo es lo que se muestra en una lista desde la que selecciona cuando utiliza la tabla de búsqueda para corregir registros en el Editor de excepciones.

Nota: Si incluye únicamente valores en su tabla de búsqueda, los valores también se utilizarán como rótulos.

Puede completar una búsqueda ingresando manualmente la información o copiándola de una fuente externa y pegándola en el cuadro de diálogo **Añadir varios**. La fuente externa puede ser una hoja de cálculo, un archivo de texto o casi cualquier otro archivo siempre que la información se presente en una o dos columnas con una coma, tabulación o punto y coma.

Nota: Cuando utilice la función **Añadir varios** y luego haga clic en **Guardar**, se borrarán todos los valores o pares valor/rótulo existentes previos para esa búsqueda. Sin embargo, una vez que utilizó la función **Añadir varios**, puede añadir manualmente valores o pares valor/rótulo adicionales.

1. Haga clic en el botón **Añadir búsqueda**.
2. Ingrese un nombre para la nueva búsqueda en el cuadro de texto.
3. Para añadir manualmente un par de valor/rótulo:
 - Haga clic en el botón **Añadir valor de búsqueda**.
 - Ingrese un valor o rótulo para la búsqueda.

Para usar la función **Añadir varios**:

- Haga clic en el botón **Añadir varios** para abrir el cuadro de diálogo.
- Seleccione las columnas y separador según corresponda. Si está pegando datos de Microsoft Excel, use el delimitador por tabulaciones. Si define el delimitador incorrecto, la herramienta importará la línea completa como el valor o rótulo (el que se haya designado en la primera columna).

- Escriba los valores, separadores y r tulos para todos los contenidos o pegue los contenidos de otra aplicaci n.

Luego de que se hayan a nadido todos los valores o pares valor/r tulo, puede ordenarlos en orden ascendente o descendente en la columna Valor o la columna R tulo. Tenga en cuenta que una vez que orden  la lista, no puede desordenarla (excepto para ordenarla en orden inverso o para ordenarla en la otra columna).

4. Repita el paso 3 seg n sea necesario.
5. Haga clic en **Guardar**

Asignaci n de b squedas

Luego de crear una b squeda, debe asignarla al campo con los datos problem ticos en la etapa Write Exceptions del flujo de datos.

1. En Enterprise Designer, abra el flujo de datos que est  produciendo los registros de excepci n.
2. Abra la etapa Write Exceptions.
3. En la columna **Nombre de b squeda** para el campo con datos problem ticos, seleccione la b squeda que contenga datos correctos nuevos de la lista desplegable y haga clic en **Aceptar**.
4. Guarde y ejecute nuevamente el flujo de datos.

Correcci n de registros

Luego de crear una b squeda y asignarla a un campo en el flujo de datos, debe corregir los registros de excepci n en Business Steward Portal.

1. En el Editor de excepciones, seleccione el flujo de datos que est  produciendo los registros de excepci n.
2. Para el primer registro problem tico, haga clic en el campo al que asign  la b squeda.
3. Haga clic en el bot n desplegable en ese campo y seleccione el r tulo correcto para ese registro. Recuerde que este r tulo no es necesariamente igual que el valor. Por ejemplo, si quiere que su campo tenga como valor "California" puede hacer clic en un r tulo que diga "CA".
4. Repita el paso 3 para cada registro problem tico.
5. Guarde las excepciones cambiadas.

Modificar o eliminar b squedas

1. Abra la p gina B squedas.
2. Marque la casilla junto a la b squeda apropiada.
3. Haga clic en el bot n **Editar b squeda**, modifique la b squeda seg n sea necesario y haga clic en **Guardar**.

o bien

Haga clic en el bot n **Eliminar b squeda**.

Dominios

Los dominios especifican el tipo de datos que se está evaluando. Se utiliza con una finalidad informativa para mostrar qué tipos de excepciones ocurren en sus datos. Por ejemplo, si la condición evalúa el éxito o la falla de una validación de direcciones, el dominio de datos sería "Dirección"; si la condición evalúa el éxito o la falla de una operación de geocodificación, el dominio de datos sería "Espacial", etc.

Nota: Los dominios que establece aquí servirán como opciones predeterminadas tanto para la configuración de Business Steward como para la etapa Exception Monitor.

Puede seleccionar uno de los dominios predefinidos detallados a continuación o especificar su propio dominio haciendo clic en el botón **Añadir elemento** y completar los campos según sea necesario. También puede editar dominios si selecciona un dominio, hace clic en el botón **Editar elemento** y realiza los cambios necesarios. A su vez, puede filtrar la lista de dominios que aparecen si ingresa datos de búsqueda en el campo **Filter**. Los resultados se actualizarán de forma dinámica.

- Account: la condición comprueba el nombre de una empresa u organización asociada con una cuenta de ventas.
- Address: la condición comprueba los datos de dirección, como una dirección de correo completa o un código postal.
- Asset: la condición comprueba datos sobre la propiedad de una empresa, por ejemplo, bienes materiales, inmuebles, recursos humanos u otros activos.
- Date: la condición comprueba los datos de fecha.
- Email: la condición comprueba los datos de correo electrónico.
- Financial: la condición comprueba datos relacionados con moneda, títulos valores, etc.
- Name: la condición comprueba los datos de nombre personal, como el nombre o el apellido.
- Phone: la condición comprueba los datos de número telefónico.
- Product: la condición comprueba datos sobre materiales, piezas, mercancías, etc.
- Spatial: la condición comprueba los datos de punto, polígono o línea que representan un accidente geográfico, por ejemplo, llanuras aluviales, líneas costeras, casas, zonas de ventas, etc.
- SSN: la condición comprueba los números de seguro social de Estados Unidos.
- Uncategorized: elija esta opción si no desea categorizar esta condición.

Métricas

Las métricas especifican la forma en que se miden los datos. Se utiliza con una finalidad informativa para mostrar qué tipos de excepciones ocurren en sus datos. Por ejemplo, si la condición es diseñada para evaluar la finalización del registro (por ejemplo, que todas las direcciones contienen códigos postales), podría especificar "Finalización" como métrica de calidad de datos.

Nota: Las métricas que establece aquí servirán como opciones predeterminadas tanto para la configuración de Business Steward como para la etapa Exception Monitor.

Puede seleccionar una de las métricas predefinidas detalladas a continuación o especificar su propia métrica haciendo clic en el botón **Añadir elemento** y completar los campos según sea necesario. También puede editar métricas si selecciona una métrica, hace clic en el botón **Editar**

elemento y realiza los cambios necesarios. A su vez, puede filtrar la lista de métricas que aparecen si ingresa datos de búsqueda en el campo **Filter**. Los resultados se actualizarán de forma dinámica.

- **Accuracy**: la condición mide si los datos pueden ser cotejados con una fuente de confianza. Por ejemplo, si no es posible verificar una dirección utilizando datos de la autoridad postal, se la puede considerar una excepción porque no es precisa.
- **Completeness**: la condición mide si a los datos les falta algún atributo esencial. Por ejemplo, una dirección a la que le falta el código postal o una cuenta a la que le falta un nombre de contacto.
- **Consistency**: la condición mide si los datos son uniformes entre múltiples sistemas. Por ejemplo, si su sistema de datos de clientes utiliza los códigos de género M y F, pero los datos que está procesando tienen el código de género 0 y 1, posiblemente se considere que los datos presentan problemas de uniformidad.
- **Interpretability**: la condición mide si los datos son correctamente analizados e ingresados en una estructura de datos que otro sistema puede interpretar. Por ejemplo, los números de seguro social deben contener solo datos numéricos. Si los datos contienen letras, por ejemplo xxx-xx-xxxx, posiblemente se considere que presentan problemas de interpretación.
- **Recency**: la condición mide si los datos están actualizados. Por ejemplo, si una persona se muda, pero la información que usted tiene en su sistema es la dirección anterior, posiblemente se considere que los datos tienen un problema de actualización.
- **Uncategorized**: elija esta opción si no desea categorizar esta condición.
- **Uniqueness**: la condición mide si hay datos duplicados. Si el flujo de datos no pudo consolidar datos duplicados, los registros pueden ser considerados una excepción.

Notificaciones

La función Notificaciones permite que el sistema le envíe un mensaje a una o más direcciones de correo electrónico cuando una cantidad designada de excepciones están ligadas a un dominio o métrica específica. Dicho correo electrónico incluirá un enlace a los registros fallidos en el Editor de excepciones de Business Steward Portal, en donde podrá ingresar manualmente los datos correctos. Para que una dirección de correo electrónico en particular no siga recibiendo notificaciones, elimine dicha dirección de la lista de destinatarios en la línea Enviar notificación a de la página Editar dominio.

Nota: Las notificaciones deben configurarse en Management Console antes de que se puedan usar en forma correcta desde dentro de la configuración de Business Steward. Consulte la Guía de administración para obtener información acerca de la configuración de notificaciones.

1. En la configuración de Business Steward, abra la página **Dominios** o la página **Métricas**.
2. Seleccione el dominio o métrica al que desee agregar una notificación y haga clic en el botón **Editar elemento**.
3. Seleccione uno o más nombres de usuario de la lista desplegable (según se configuró en Management Console) o ingrese nuevas direcciones de correo electrónico a las que se enviarán las notificaciones.
4. Seleccione la cantidad de registros de excepción que deberían activar una notificación.
5. Ingrese el texto que debería enviarse como asunto de la notificación.

6. Ingrese el mensaje que debería aparecer en el cuerpo de la notificación. Puede utilizar variables en el mensaje para comunicar información importante sobre la excepción, por ejemplo:
 - `#{jobID}`—El número de ID del trabajo que produjo los registros de excepción.
 - `#{jobName}`—El nombre del trabajo que produjo los registros de excepción.
 - `#{userName}`—El nombre del usuario cuyo trabajo produjo los registros de excepción.
 - `#{stageLabel}`—El nombre de la etapa del flujo de datos que produjo los registros de excepción.
 - `#{link}`—Un enlace a la página Editor en Business Steward Portal, donde se muestran los registros para un flujo de datos determinado.
7. Marque la casilla **Enviar recordatorio** si desea enviar una notificación de recordatorio y seleccione la cantidad de días que deberían pasar para enviar el recordatorio.
8. Seleccione uno o más nombres de usuario de la lista desplegable (según se configuró en Management Console) o ingrese una dirección de correo electrónico en el campo **Enviar recordatorio a** para enviar notificaciones a direcciones de correo electrónico distintas de las que especificó en el paso 3.
9. Ingrese el texto que debería enviarse como asunto de la notificación de recordatorio.
10. Ingrese el mensaje que debería aparecer en el cuerpo de la notificación de recordatorio. El recordatorio utiliza una variable adicional:
 - `#{Count}`—La cantidad de excepciones para el flujo de datos o etapa especificado que todavía deben resolverse.
11. Marque la casilla **Recordar a diario** si desea que se envíe un recordatorio todos los días hasta que se resuelva la excepción.

Informe de calidad de datos

1. Haga clic en **Informe de calidad de datos** para registrar las condiciones de éxito o error de la etapa Exception Monitor. Si desactiva esta opción, la página de calidad de datos de Business Steward Portal no tendrá datos. Del mismo modo, se deshabilitará el campo "Report only" de la etapa Exception Monitor para todos los flujos de datos.
2. En el menú desplegable **Retención** seleccione cuánto tiempo, en meses, se deben conservar los datos.

Configuración de indicadores clave de rendimiento

La sección **Cálculo de KPI** de la ficha Informe de calidad de datos le permite designar indicadores clave de rendimiento (KPI) para sus datos, y asignar notificaciones para cuando dichos KPI cumplan ciertas condiciones.

1. Haga clic en **Añadir un KPI**.
2. Ingrese un **Nombre** para el indicador clave de rendimiento. Este nombre debe ser único en su servidor Spectrum™ Technology Platform.
3. Seleccione una de las **Métricas** de calidad de datos para el indicador clave de rendimiento; si no hace una selección, este indicador clave de rendimiento se ligará a todas las **métricas**.

4. Seleccione un nombre de **Flujo de datos** para el indicador clave de rendimiento; si no hace una selección, este indicador clave de rendimiento se ligará a todos los flujos de datos del módulo Business Steward.
5. Seleccione un **Rótulo de etapa** para el indicador clave de rendimiento; si no hace una selección, este indicador clave de rendimiento se ligará a todas las etapas del módulo Business Steward en sus flujos de datos.
6. Seleccione un **Dominio** de datos para el indicador clave de rendimiento; si no hace una selección, este indicador clave se ligará a todos los **dominios**. Tenga presente que al seleccionar un dominio aquí se deshabilitará el campo Condición.
7. Seleccione una **Condición** para el indicador clave de rendimiento. Si no hace una selección, este indicador clave de rendimiento se establece automáticamente en "Todo". Tenga presente que para seleccionar una condición, primero tiene que haber seleccionado "Todo" en el campo Dominio. Después de seleccionar una condición, el campo Dominio se deshabilitará.
8. Seleccione un **Período de KPI** para designar los intervalos para los que desee que el módulo Business Steward Module supervise los datos y envíe las notificaciones. Por ejemplo, si selecciona "1" y "Mensualmente", se enviará una notificación de KPI cuando el porcentaje de excepciones haya aumentado conforme al umbral o la variación en un período de mes a mes.
9. Proporcione un porcentaje para una **Variación** o un **Umbral**. Los valores de variación representan el porcentaje incrementado de fallas en los registros de excepción desde el último período. Los valores de umbral representan el porcentaje de fallas en los cuales usted desea que se envíen las notificaciones. Su valor debe ser 1 o superior.
10. Ingrese las direcciones de correo electrónico, o selecciónelas en la lista, de los **Destinatarios** que deben ser notificados cuando se cumplan estas condiciones. Cuando sea posible, este campo se autocompletará a medida que ingrese las direcciones de correo electrónico. No es necesario que separe las direcciones con comas, punto y coma u otro signo de puntuación.
11. Ingrese el **Asunto** que desee que se use en el correo electrónico de notificación.
12. Ingrese el **Mensaje** que desee que la notificación comunique cuando se cumplan estas condiciones.
13. Haga clic en **Aceptar**. El nuevo KPI aparecerá entre cualquier otro KPI existente. Puede clasificar los KPI en cualquiera de las columnas con datos.
14. Haga clic en **Guardar**

Puede modificar y eliminar los KPI al seleccionarlos y hacer clic en **Editar KPI seleccionado** o **Eliminar KPI seleccionado**.

Servicios de herramientas de búsqueda

1. Seleccione qué servicios de herramientas de búsqueda desea que estén disponibles en el Editor de excepciones de Business Steward Portal. La lista de servicios disponibles se basa en permisos de usuario y se completa a partir de sus módulos y servicios con licencia dentro de Spectrum Technology Platform. Utilice **Filtro** para acortar la lista de servicios a partir de criterios de filtro.
2. Haga clic en **Premium** para indicarle a los usuarios que acumularán tarifas adicionales cuando utilicen estos servicios (como los servicios Dun & Bradstreet).

Opciones

1. Haga clic en **Auditar eventos de excepción** para que el módulo Business Steward mantenga un registro de cuándo se crean, leen, actualizan o eliminan los registros de excepción.
2. Haga clic en **Hacer seguimiento del progreso** para hacer el seguimiento de cuándo se aprueban los registros de excepción en Business Steward Portal. Si desactiva esta opción, los gráficos de progreso del Tablero Business Steward Portal no aparecerán.

Módulo Data Normalization

Módulo Data Normalization

El Módulo de normalización de datos examina los términos en un registro y determina si el término está en el formato preferido.

Componentes

El módulo Data Normalization consta de:

- **Advanced Transformer:** en esta etapa se exploran y dividen las cadenas de datos en múltiples campos, y se colocan los datos extraídos y sin extraer en un campo ya existente o nuevo.
- **Open Parser:** esta etapa analiza los datos de entrada de múltiples culturas de todo el mundo por medio de una gramática de análisis simple pero eficaz. Esta gramática permite definir una secuencia de expresiones que representan patrones de dominio para analizar los datos de entrada. Open Parser también recopila datos estadísticos y otorga calificaciones a los cruces de análisis para ayudar a determinar la efectividad de las gramáticas de análisis utilizadas.
- **Table Lookup:** esta etapa evalúa un término y lo compara con un formato previamente validado de ese término. Si el término no está en el formato adecuado, es reemplazado por su versión estándar. Las funciones de Table Lookup incluyen el cambio de palabras completas a abreviaturas y viceversa, el cambio de apodos a nombres completos o la corrección de palabras mal escritas.
- **Transliterator:** Transliterator convierte una cadena de caracteres entre el alfabeto latino y otros alfabetos.

Advanced Transformer

El trabajo Advanced Transformer explora y divide las cadenas de datos en múltiples campos por medio de tablas o expresiones regulares. Esta herramienta extrae un término específico o una

cantidad determinada de palabras situadas a la derecha o la izquierda de un término. Los datos extraídos y sin extraer pueden colocarse en un campo nuevo o ya existente.

Por ejemplo, supongamos que se desea extraer la información de habitación (suite) de este campo de dirección para colocarla en un campo separado.

2300 BIRCH RD STE 100

Para lograrlo, se crea una instancia de Advanced Transformer que extrae el término STE (suite) y todas las palabras a la derecha de ese término, de forma tal que el campo queda así:

2300 BIRCH RD

Input

Advanced Transformer utiliza cualquier campo de entrada definido en el flujo de datos.

Opciones

Las opciones de Advanced Transformer se pueden configurar a nivel de etapa, por medio de cualquier cliente Spectrum™ Technology Platform o bien en el tiempo de ejecución, mediante las opciones de flujo de datos.

Opciones de configuración

Para especificar las opciones de Advanced Transformer, debe crear una regla. Puede crear múltiples reglas y luego especificar el orden en que quiere aplicarlas. Para crear una regla:

1. Haga doble clic en la instancia de Advanced Transformer en el lienzo. Se muestra el cuadro de diálogo Opciones de Advanced Transformer.
2. Seleccione la cantidad de instancias de ejecución y haga clic en **Aceptar**. Utilice la opción Instancias de ejecución a fin de configurar un flujo de datos que ejecute instancias múltiples y paralelas de una etapa para así aumentar potencialmente el rendimiento.
3. Haga clic en el botón **Agregar**. Aparece el cuadro de diálogo Opciones de regla de Advanced Transformer.

Nota: Si agrega muchas reglas de transformación, puede utilizar los botones **Subir** y **Bajar** para cambiar el orden en el que se aplican las reglas.

4. Seleccione el tipo de acción de transformación que desea realizar y haga clic en **Aceptar**. Las opciones se indican en la siguiente tabla.

Tabla 23: Opciones de Advanced Transformer

Opción	Descripción
Origen	Especifica el campo de entrada de origen que se evalúa para explorar y dividir.

Opción	Descripción
Extraer por medio de	<p>Seleccione Datos de tabla o Expresiones regulares.</p> <p>Seleccione Datos de tabla si desea explorar y dividir mediante las tablas XML, ubicadas en <Drive>:\Program Files\Pitney Bowes\Spectrum\server\modules\advancedtransformer\data. Consulte opciones de tabla de datos a continuación para obtener más información sobre cada opción.</p> <p>Seleccione Expresiones regulares si desea explorar y dividir utilizando expresiones regulares. Las expresiones regulares brindan muchas opciones adicionales para dividir datos. Puede utilizar expresiones regulares previamente empaquetadas seleccionando una de la lista o puede construirlas utilizando sintaxis de RegEx (expresiones regulares).</p> <p>Por ejemplo, puede dividir datos cuando se encuentra el primer valor numérico. Así, por ejemplo, en "John Smith 123 Main St.", "John Smith" quedaría en un campo y "123 Main St.", en otro. Vea las Opciones de expresiones regulares a continuación para obtener más información sobre cada opción.</p>
Opciones de tabla de datos	
Datos sin extraer	<p>Especifica el campo de salida donde desea que se incluyan los datos transformados. Si desea reemplazar el valor original, especifique en el campo Destino el mismo campo que en el cuadro desplegable Origen.</p> <p>También puede ingresar un nuevo nombre de campo en el campo Destino. Si ingresa un nuevo nombre de campo, ese nombre de campo estará disponible para etapas de su flujo de datos posteriores a Advanced Transformer.</p>
Datos extraídos	<p>Especifica el campo de salida donde desea ubicar los datos extraídos.</p> <p>También puede ingresar un nuevo nombre de campo en el campo Datos extraídos. Si ingresa un nuevo nombre de campo, ese nombre de campo estará disponible para etapas de su flujo de datos posteriores a Advanced Transformer.</p>
Caracteres de muestra	<p>Especifica cualquier carácter especial que desee usar de muestra. La formación de muestras es el proceso de separar términos. Por ejemplo, si tiene un campo con los datos "Smith, John", sería conveniente que usara la coma como muestra. Esto daría como resultado los siguientes términos:</p> <ul style="list-style-type: none"> • Smith • , • John <p>Ahora que todos los términos están separados, los datos pueden dividirse explorando y extrayendo en función de la coma para que "Smith" y "John" sean identificados claramente como los datos que deben estandarizarse.</p>

Opción	Descripción						
tabla	<p>Especifica la tabla que contiene los términos sobre los que se basará la división del campo. Para obtener una lista de tablas, consulte Tablas de Advanced Transformer en la página 160. Para obtener información sobre la creación y modificación de tablas, consulte Introducción a las tablas de búsqueda en la página 160.</p>						
Buscar términos de múltiples palabras	<p>Seleccione esta casilla de verificación para activar las búsquedas de diversas palabras dentro de una cadena en particular. Por ejemplo:</p> <p>Cadena de entrada = "Cedar Rapids 52401" Reglas de negocio = Identificar "Cedar Rapids" en una cadena sobre la base de una tabla que contiene la entrada; Cedar Rapids = Estados Unidos Salida = Identifica la presencia de "Cedar Rapids" y ubica los términos en un nuevo campo, como por ejemplo Ciudad.</p> <p>En las búsquedas de varias palabras, la búsqueda se detiene cuando encuentra el primer cruce.</p> <p>Nota: Seleccionar esta opción puede afectar negativamente el rendimiento.</p>						
Extraer	<p>Especifica el tipo de extracción que se realizará. Seleccione una de estas opciones:</p> <table border="0"> <tr> <td>Extraer término</td> <td>Extrae el término identificado por la tabla seleccionada.</td> </tr> <tr> <td>Extraer N palabras a la derecha del término</td> <td>Extrae palabras a la derecha del término. Usted especifica la cantidad de palabras que se extraen. Por ejemplo, si desea extraer dos palabras a la derecha del término identificado, especifique 2.</td> </tr> <tr> <td>Extraer N palabras a la izquierda del término</td> <td>Extrae palabras a la izquierda del término. Usted especifica la cantidad de palabras que se extraen. Por ejemplo, si desea extraer dos palabras a la izquierda del término identificado, especifique 2.</td> </tr> </table> <p>Si elige extraer palabras a la derecha o a la izquierda del término, puede especificar si desea incluir el término en sí en los datos de destino o en los datos extraídos. Por ejemplo, si tiene el siguiente campo:</p> <p>2300 BIRCH RD STE 100</p> <p>y desea extraer "STE 100" y ubicarlo en un campo especificado en los datos extraídos, debe elegir incluir el término en el campo de datos extraído. De esta forma, incluirá la abreviatura "STE" y la palabra "100".</p> <p>Si no selecciona Destino ni Datos extraídos, el término no se incluirá y será descartado.</p>	Extraer término	Extrae el término identificado por la tabla seleccionada.	Extraer N palabras a la derecha del término	Extrae palabras a la derecha del término. Usted especifica la cantidad de palabras que se extraen. Por ejemplo, si desea extraer dos palabras a la derecha del término identificado, especifique 2.	Extraer N palabras a la izquierda del término	Extrae palabras a la izquierda del término. Usted especifica la cantidad de palabras que se extraen. Por ejemplo, si desea extraer dos palabras a la izquierda del término identificado, especifique 2.
Extraer término	Extrae el término identificado por la tabla seleccionada.						
Extraer N palabras a la derecha del término	Extrae palabras a la derecha del término. Usted especifica la cantidad de palabras que se extraen. Por ejemplo, si desea extraer dos palabras a la derecha del término identificado, especifique 2.						
Extraer N palabras a la izquierda del término	Extrae palabras a la izquierda del término. Usted especifica la cantidad de palabras que se extraen. Por ejemplo, si desea extraer dos palabras a la izquierda del término identificado, especifique 2.						

Opciones de expresiones regulares

Opción	Descripción
Expresiones regulares	<p>Seleccione una expresión regular previamente empaquetada de la lista o construya una en el cuadro de texto. Advanced Transformer admite sintaxis estándar de RegEx (expresiones regulares).</p> <p>La plataforma Java 2 contiene un paquete denominado java.util.regex, que habilita el uso de expresiones regulares. Para obtener más información, visite: java.sun.com/docs/books/tutorial/essential/regex/index.html.</p>
Botón de puntos suspensivos	Haga clic en este botón para agregar o eliminar una expresión regular nueva.
Populate Group	Después de seleccionar una expresión Regex predefinida o de ingresar una nueva, haga clic en Populate Group para extraer cualquier grupo de expresiones Regex y ubicar en la lista Grupos tanto la expresión completa como cualquier otro grupo de expresiones Regex encontrado.
Grupos	<p>Esta columna muestra las expresiones regulares del grupo de expresiones regulares seleccionado.</p> <p>Por ejemplo, si selecciona la expresión regular Fecha, en el cuadro de texto se mostrará la siguiente expresión: <code>(1[012]{1,2} 0?[1-9])[-./]([12][0-9] 3[01]{1,2} 0?[1-9])[-./]([0-9]{4})</code>. Esta expresión Regex tiene tres partes y se puede enviar la expresión entera y cada una de las partes a un campo de salida diferente. Se busca la expresión entera en el campo de origen y, si se encuentra un cruce, se mueven las partes asociadas al campo de salida asignado. Si el campo de origen es "En 12/14/2006" y se le aplica la expresión Fecha, se asignará la fecha completa (por ejemplo, "12/14/2006") al campo FECHA, el "12" al campo MES, el "14" al campo DÍA y "2006" al campo AÑO. Buscará los datos y, si los encuentra, moverá la información correspondiente a los campos de salida pertinentes.</p> <p>Campo de origen: "En 12/14/2006" FECHA: "12/14/2006" MES: "12" DÍA: "14" AÑO: "2006"</p>
Campo de salida	Despliegue el menú para seleccionar un campo de salida.

Salida

Advanced Transformer no crea campos de salida nuevos. Solo los campos que defina se escriben en la salida.

Open Parser

Open Parser analiza los datos de entrada de múltiples culturas de todo el mundo por medio de una gramática de análisis simple pero eficaz. Esta gramática permite definir una secuencia de expresiones que representan patrones de dominio para analizar los datos de entrada. Open Parser también recopila datos estadísticos y otorga calificaciones a los cruces de análisis para ayudar a determinar la efectividad de las gramáticas de análisis utilizadas.

Utilice Open Name Parser para:

- Analizar datos de entrada utilizando gramáticas de análisis específicas de una cultura o específicas de un dominio, las cuales puede definir en Domain Editor.
- Analizar datos de entrada utilizando gramáticas de análisis independientes de dominio, las cuales puede definir en Open Parser utilizando la misma gramática de análisis simple pero eficaz disponible en Domain Editor.
- Analice los datos de entrada mediante las gramáticas de análisis independientes de dominio en el tiempo de ejecución, que defina en Opciones del flujo de datos.
- Previsualizar las gramáticas de análisis para probar cómo se analizan los datos de entrada de ejemplo antes de ejecutar el trabajo usando el archivo de datos de entrada de destino.
- Hacer un seguimiento de los resultados de las gramáticas de análisis para ver si las muestras presentaron o no un cruce con las expresiones definidas y para comprender mejor el proceso de comparación.

Input

Open Parser acepta los campos de entrada que usted define en su gramática de análisis. Para obtener más información, consulte [Comandos de la sección de encabezado](#) en la página 30.

Si se realiza el análisis específico de una cultura, puede incluir de forma opcional un campo CultureCode en los datos de entrada para utilizar una gramática de análisis de una cultura específica con un registro. Si se omite el campo CultureCode o si está vacío, entonces cada cultura enumerada en la etapa Open Parser se aplica, en el orden especificado. Se devuelve el resultado de la cultura con la puntuación más alta del analizador, o la primera cultura que alcance una puntuación de 100. Para obtener más información sobre el campo CultureCode, consulte [Asignación de un análisis de cultura a un registro](#) en la página 12.

Opciones

Las siguientes tablas enumeran las opciones para la etapa Open Parser.

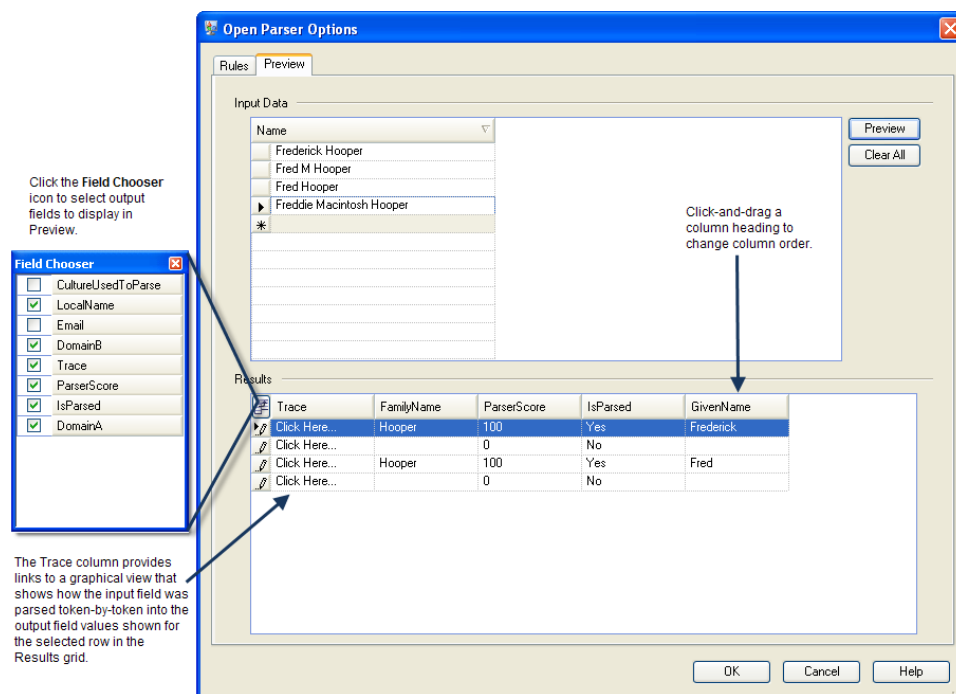
Ficha Reglas

Opción	Descripción
Usar gramática de dominio específica de la cultura	<p>Especifica que se debe utilizar un idioma y una gramática de análisis específica de dominio que ya se ha definido en la herramienta Open Parser Domain Editor en Enterprise Designer. Para obtener más información sobre la definición de dominios, consulte Definición de gramáticas de análisis específicas de culturas en la página 10.</p> <p>Si elige esta opción, también verá estas opciones:</p> <p>Dominio Especifica la gramática de análisis que se utilizará.</p> <p>Culturas Especifica el idioma o la cultura de los datos que desea analizar. Haga clic en el botón Agregar para agregar una cultura. Puede cambiar el orden en el que Open Parser intenta analizar los datos con cada cultura mediante los botones Subir y Bajar. Para obtener más información sobre las culturas, consulte Definición de gramáticas de análisis específicas de culturas en la página 10.</p> <p>Arrojar múltiples registros analizados Habilite esta opción para que Open Parser arroje registros para cada cultura que analice los datos de entrada correctamente. Si no selecciona esta casilla, Open Parser arrojará los resultados del primer registro que consiga una calificación del analizador de 100, independientemente de la cultura. Si se ejecutan todas las culturas sin que ningún registro consiga una calificación del analizador de 100, Open Parser arrojará el registro con la calificación más cercana a 100. Si hay múltiples culturas que arrojan registros con la misma calificación alta inferior a 100, el orden elegido en el Paso 4 determinará qué registro de cultura se arroja.</p>
Definir gramática independiente de dominio	<p>Elija esta opción si desea definir una gramática de análisis que deba aplicarse sin considerar el idioma o el dominio de los datos de entrada. Si elige esta opción, aparecerá el editor de gramática y se puede definir la gramática de análisis directamente en la etapa Open Parser en lugar de usar la herramienta Open Parser Domain Editor en Enterprise Designer.</p> <p>Nota: También puede definir la gramática independiente del dominio en el tiempo de ejecución. Haga clic aquí para obtener más información.</p>

Ficha Vista previa

La creación de una gramática de análisis de trabajo es un proceso iterativo. La vista previa es útil para probar variaciones en su entrada a fin de asegurarse de que la gramática de análisis produzca los resultados deseados.

Ingrese valores de prueba en el campo de entrada y haga clic en **Vista previa**.



Los campos de salida analizados se muestran en la cuadrícula **Resultados**. Para obtener información sobre los campos de salida, consulte [Salida](#) en la página 292. Para obtener información sobre el seguimiento, consulte [Seguimiento de resultados de análisis finales](#) en la página 37. Si los resultados no son los esperados, haga clic en la ficha **Reglas** y continúe editando la gramática de análisis y probando los datos de entrada hasta que se produzcan los resultados deseados.

Salida

Tabla 24: Datos de salida de Open Parser

Nombre de campo	Descripción / Valores válidos
<Campo de entrada>	El campo de entrada original definido en la gramática de análisis.
<Campos de salida...>	Los campos de salida definidos en la gramática de análisis.
CultureCode	Los códigos de cultura contenidos en los datos de entrada. Para acceder a una lista completa de los códigos de cultura admitidos, consulte Asignación de un análisis de cultura a un registro en la página 12.

Nombre de campo	Descripción / Valores válidos
CultureUsedtoParseSelect a match results in the Match Results List and then click Remove .	El valor del código de cultura utilizado para analizar cada registro de salida. Este valor se basa en cruces con una gramática de análisis específica de la cultura.
IsParsed	Indica si se analizó un registro de salida. Los valores posibles son Sí o No.
ParserScoreSelect a match results in the Match Results List and then click Remove .	Indica la calificación promedio total. El valor de ParserScore estará entre 0 y 100, como se define en la gramática de análisis. Se arroja 0 cuando no se encuentran cruces. Para obtener más información, consulte Calificación .
Trace	Haga clic en este control para obtener una vista gráfica de cómo se analizó cada muestra en la gramática de análisis respecto del campo de salida para la fila seleccionada en la cuadrícula Resultados.

Table Lookup

La etapa Table Lookup estandariza los términos en función de un formato validado anteriormente para el término en cuestión y aplica la versión estándar. Para realizar esta evaluación, se realiza una búsqueda en una tabla para encontrar el término a estandarizar.

Por ejemplo:

	Nombre	Apellido
Datos de entrada de origen:	Bill	Smith
Datos de salida estandarizados:	William	Smith

Hay tres tipos de acciones que puede realizar: estandarizar, identificar y clasificar.

Si el término se encuentra al realizar la acción de estandarizar, Table Lookup reemplaza todo el campo o los términos individuales dentro del campo por el término estandarizado, incluso aunque el campo contenga varias palabras. Las funciones de Table Lookup incluyen el cambio de palabras

completas a abreviaturas y viceversa, el cambio de apodos a nombres completos o la corrección de palabras mal escritas.

Si el término se encuentra al realizar la acción de identificar, Table Lookup marca el registro para señalar que contiene un término que puede estandarizarse, pero no ejecuta ninguna otra acción.

Si el términos e encuentra al realizar la acción de categorizar, Table Lookup utiliza el valor de origen como clave y copia el valor correspondiente de la entrada de la tabla en el campo seleccionado. Si no coincide ninguno de los términos, **Categorizar** usa el valor predeterminado especificado.

Input

Tabla 25: Campos de entrada de Table Lookup

Nombre de campo	Descripción / Valores válidos
Source	Especifica el campo de entrada de origen que se evalúa para explorar y dividir.
StandardizationTable	Una de las tablas enumeradas en Tablas de Table Lookup en la página 163.

Opciones

Las opciones de Table Lookup se pueden configurar a nivel de etapa, por medio de cualquiera de los clientes Spectrum™ Technology Platform o en el tiempo de ejecución, mediante las opciones de flujo de datos.

Opciones de configuración

Para especificar las opciones de Table Lookup, debe crear una regla. Puede crear múltiples reglas y luego especificar el orden en que quiere aplicarlas. Para crear una regla, abra la etapa de tabla de búsqueda y haga clic en **Agregar** y, luego, complete los campos siguientes.

Nota: Si agrega muchas reglas de Table Lookup, puede utilizar los botones **Subir** y **Bajar** para cambiar el orden en el que se aplican las reglas.

Opción	Descripción
Acción	<p>Especifica el tipo de acción a tomar en el campo de origen. Una de las siguientes:</p> <p>Estandarizar Cambia los datos en un campo para que corresponda con el término estandarizado que se encuentre en la tabla de búsqueda. Si el campo contiene varios términos, solo los términos que se encuentran en la tabla de búsqueda se sustituyen por el término estandarizado. El resto de datos en el campo no cambia.</p> <p>Identificar Marca el registro para señalar que contiene un término que puede estandarizarse, pero no ejecuta ninguna otra acción en el campo. El campo de salida StandardizedTermIdentified se agrega al registro con el valor "Sí" si el campo se puede estandarizar y "No" en caso contrario.</p> <p>Categorizar Utiliza el valor Origen como clave y copia el valor correspondiente de la entrada de tabla en el campo seleccionado en la lista Destino. Esto crea un nuevo campo en sus datos que se pueden utilizar para categorizar registros.</p>

Opción	Descripción
Activar	<p>Especifica si utilizar todo el campo como término de búsqueda o buscar en la tabla de búsqueda de cada término en el campo. Una de las siguientes:</p> <p>Completar campo Trata todo el campo como un término, lo que se traduce en lo siguiente:</p> <ul style="list-style-type: none"> • Si selecciona la acción Estandarizar, Table Lookup considera el campo entero como una sola cadena e intenta estandarizar el campo utilizando la cadena en su totalidad. Por ejemplo, "International Business Machines" se cambiaría a "IBM". • Si selecciona la acción Identificar, Table Lookup trata todo el campo como una cadena y marca el registro si es posible estandarizar la cadena en su totalidad. • Si selecciona la acción Categorizar, Table Lookup tratará todo el campo como una cadena y marcará el registro si es posible categorizar la cadena en su totalidad. <p>Términos individuales en el campo Trata cada palabra en el campo como su propio término, lo que se traduce en lo siguiente:</p> <ul style="list-style-type: none"> • Si selecciona la acción Estandarizar, Table Lookup analizará el campo e intentará estandarizar los términos individuales dentro del campo. Por ejemplo, "Bill Mike Smith" se cambiaría a "William Michael Smith". • Si selecciona la acción Identificar, Table Lookup analiza el campo y marca el registro si se puede estandarizar algún término individual dentro del campo. • Si ha seleccionado la acción Categorizar, A diferencia de Estandarizar, Categorizar no copia el término de origen si no hay una correspondencia de tabla. Si no coincide ninguno de los términos, Categorizar usa el valor predeterminado especificado. A diferencia de Estandarizar, Categorizar solo arroja ese valor de tabla y nada desde Origen. Si no coincide ninguno de los términos, Categorizar usa el valor predeterminado especificado.
Origen	Especifica el campo en el que desea que se incluya el término que quiere buscar.
Destino	<p>Especifica el campo en el que deben escribirse los términos obtenidos de la búsqueda de la tabla.</p> <p>Si desea reemplazar el valor, especifique en el campo Destino el mismo campo que en el cuadro Origen. También puede crear un nuevo campo ingresando el nombre del campo que desea crear.</p> <p>El campo Destino no está disponible si se selecciona la acción Identificar.</p>

Opción	Descripción				
Tabla	<p>Especifica la tabla que desea utilizar para encontrar términos que correspondan con los datos en su flujo de datos.</p> <p>Para obtener una lista de tablas que puede editar, consulte Tablas de Table Lookup en la página 163. Para obtener información sobre la creación y modificación de tablas, consulte Introducción a las tablas de búsqueda en la página 160.</p>				
Buscar términos de múltiples palabras	<p>Permite múltiples búsquedas de palabras dentro de una cadena determinada. Por ejemplo:</p> <p>Cadena de entrada = "Major General John Smith" Regla de negocio: Identificar "Major General" en una cadena basada en una tabla que contiene la entrada Salida: Sustituir "Major General" con "Maj. Gen.".</p> <p>En las búsquedas de varias palabras, la búsqueda se detiene cuando encuentra el primer cruce.</p> <p>Esta opción está deshabilitada cuando se selecciona Activado en Campo completo.</p> <p>Nota: Seleccionar esta opción puede afectar negativamente el rendimiento.</p>				
Si no se encuentra la entrada de la tabla, definir el valor de destino en	<p>Especifica el valor que se debe colocar en el campo destino en un término correspondiente no puede encontrarse en la tabla de búsqueda. Una de las siguientes:</p> <table border="0"> <tr> <td>Valor del origen</td> <td>Poner el valor del campo de origen en el campo destino.</td> </tr> <tr> <td>Otros</td> <td>Poner un valor específico en el campo destino.</td> </tr> </table>	Valor del origen	Poner el valor del campo de origen en el campo destino.	Otros	Poner un valor específico en el campo destino.
Valor del origen	Poner el valor del campo de origen en el campo destino.				
Otros	Poner un valor específico en el campo destino.				

Cómo configurar opciones en el tiempo de ejecución

Las opciones de Table Lookup se pueden configurar y pasar en el tiempo de ejecución, si estas se exponen como opciones del flujo de datos. Esto le permite anular la configuración existente con cadenas con formato JSON. También puede establecer las opciones de tiempo de ejecución cuando se invoque al trabajo, mediante un flujo de proceso o mediante la herramienta de línea de comandos del ejecutor del trabajo.

Puede encontrar un esquema para LookupRule en la siguiente carpeta:

```
<Spectrum Location>\server\modules\jsonSchemas\tableLookup
```

Para definir las reglas de Table Lookup en el tiempo de ejecución:

1. En Enterprise Designer, abra un flujo de datos que use una etapa Table Lookup.
2. Guarde y exponga dicho flujo de datos.

3. Vaya a `Edit > Dataflow Options`.
4. En la tabla **Aplicar opciones de flujo de datos a etapas** expanda Table Lookup. Marque la casilla de LookupRule.
5. Opcional: cambie el nombre de las opciones en el campo **Rótulo de opción**.
6. Haga clic en **Aceptar** dos veces.

Salida

Tabla 26: Datos de salida de Table Lookup

Nombre de campo	Descripción / Valores válidos				
StandardizedTermIdentified	Indica si el campo contiene o no contiene un término que se pueda estandarizar. Solo salida si selecciona las opciones Campo completo o Términos individuales en el campo . <table border="0" style="margin-left: 20px;"> <tr> <td>Yes</td> <td>El registro contiene un término que se puede estandarizar.</td> </tr> <tr> <td>No</td> <td>El registro no contiene un término que se puede estandarizar.</td> </tr> </table>	Yes	El registro contiene un término que se puede estandarizar.	No	El registro no contiene un término que se puede estandarizar.
Yes	El registro contiene un término que se puede estandarizar.				
No	El registro no contiene un término que se puede estandarizar.				

Transliterator

El componente Transliterator (Transcriptor) convierte una cadena de caracteres entre el alfabeto latino y otros alfabetos. Por ejemplo:

Origen	Transcripción
キャンパス	kyanpasu
Αλφαβητικός Κατάλογος	
биологическом	biologichyeskom

Cabe señalar que la transcripción o transliteración no es una traducción, sino que consiste en la conversión de letras de un alfabeto a otro sin traducir las palabras.

Nota: Con frecuencia, los métodos de transcripción estándar no siguen las reglas de pronunciación de ningún idioma en particular en el sistema de escritura de destino.

La etapa Transliterator admite estas cadenas de caracteres. Como regla general, en esta etapa se siguen las pautas del Grupo de Trabajo sobre Sistemas de Romanización de UNGEGN (UNEGN Working Group on Romanization Systems). Para obtener más información, consulte www.eki.ee/wgrs.

Árabe	El alfabeto utilizado por varios idiomas de Asia y África, lo que incluye el árabe, el persa y el urdu.
Cirílico	El alfabeto utilizado por diversos idiomas de Europa del Este y Asia, lo que incluye las lenguas eslavas como el ruso. La etapa Transliterator por lo general sigue la norma ISO 9 para el conjunto de caracteres cirílicos básicos.
Devanagari	El alfabeto que utilizan los idiomas indios, incluidos el hindi y el sánscrito. Este alfabeto proviene del alfabeto brahmi, que es uno de los sistemas de escritura más antiguos utilizados en la India antigua, presente en el sur y el centro de Asia.
Griego	El alfabeto utilizado por el idioma griego. Este alfabeto pertenece a la rama helénica de la familia del idioma indoeuropeo.
Gujarati	El alfabeto se utiliza en el estado de Gujarat en India occidental. Es uno de los alfabetos modernos de la India, que se adaptó a partir del alfabeto del devanagari.
Gurmukhi	El alfabeto que se utiliza en el idioma indio panyabí. Este alfabeto tiene una influencia considerable del alfabeto del nagari, que es una forma anterior del alfabeto del devanagari.
Hangul	El alfabeto utilizado por el idioma coreano. La etapa Transliterator sigue las normas de transcripción del Ministerio de Cultura y Turismo de Corea. Para obtener más información, visite el sitio web The National Institute of the Korean Language .
Han	El alfabeto que se utiliza en el idioma chino. Es una rama de la familia del idioma tibetano-birmano, y se ha escrito con cadenas de caracteres que se basan en el tailandés y el chino.
Chino tradicional/simplificado	La etapa Transliterator admite el chino tradicional y el simplificado. Por ejemplo, este es chino tradicional: 人人生而自由。 Este es chino simplificado: 人人生而自由
Kannada	El alfabeto que utilizan los idiomas indios del sur, como el konkani. Este alfabeto proviene del alfabeto brahmi de la India antigua.
Katakana y hiragana	Uno de los diversos alfabetos que pueden utilizarse para escribir en japonés. La etapa Transliterator utiliza una leve variante del sistema Hepburn. Con el sistema Hepburn, los caracteres ZI (ジ) y DI (ヂ) se

representan como "ji", y los caracteres ZU (ズ) y DU (ヅ) se representan como "zu". A los fines de una posible inversión, esto se modificó mínimamente para usar "dji" para DI y "dzu" para DU. La transcripción de Katakana puede invertirse. La transcripción de Hiragana - Katakana no puede invertirse por completo porque hay varias letras en Katakana que no cuentan con un equivalente en Hiragana. Además, la marca de extensión no se usa con Hiragana. La transcripción Hiragana - Latino tampoco puede invertirse porque internamente es una combinación de Katakana - Hiragana e Hiragana - Latino.

Ancho medio/ Ancho completo	La etapa Transliterador puede hacer conversiones entre caracteres de medio ancho (con menos espacio) y caracteres de ancho completo (con más espacio). Por ejemplo, este es el medio ancho: アルアノリウ. Este es el ancho completo: アルアノリウ.
Latino	El alfabeto utilizado por la mayoría de los idiomas europeos, como el inglés o el español. Originalmente, lo usaban los romanos de la antigüedad para escribir el latín.
Malabar	El alfabeto que utiliza el idioma malabar, el idioma oficial del estado indio de Kerala. En sus inicios, esta cadena de caracteres se escribía con el alfabeto vatteluttu, que significa "escritura redondeada", y que se desarrolló a partir del alfabeto brahmi de la India antigua.
Oriya	El alfabeto que utiliza el idioma oriya, el idioma oficial del estado indio de Odisha. El alfabeto oriya se desarrolló a partir del alfabeto kalinga, uno de los numerosos descendientes del alfabeto brahmi de la India antigua.
Tamil	El alfabeto que se utiliza en el idioma tamil en varios estados de la India, Sri Lanka y Malasia. En sus inicios, este alfabeto se escribía con una versión del alfabeto brahmi conocida como tamil brahmi.
Telugu	El alfabeto que se utiliza en varios idiomas del sur de la India. Este alfabeto proviene del alfabeto brahmi de la India antigua.
Tailandés	El alfabeto que utiliza el idioma tailandés. La cadena de caracteres tiene influencias del alfabeto brahmi de la India antigua y de los alfabetos del camboyano.

El componente Transliterador forma parte del módulo Data Normalization (Normalización de datos). Para acceder a una lista de otras etapas, consulte [Módulo Data Normalization](#) en la página 285.

Conceptos de transcripción

Existe una serie de características que, por lo general, son recomendables para las transcripciones de alfabetos. Una buena transcripción debe ser:

- Finalizado
- Predecible

- Pronunciable
- Inequívoca

Rara vez se satisfacen simultáneamente todas estas características, por eso la etapa Transliterador intenta equilibrar estos requerimientos.

Finalizado

Toda secuencia de caracteres bien formada en el sistema de escritura de origen debe transcribirse a una secuencia de caracteres del sistema de escritura de destino.

Predecible

Las letras por sí mismas (sin conocimiento alguno de los idiomas escritos en ese alfabeto) deben ser suficientes para la transcripción, de acuerdo con una cantidad relativamente pequeño de reglas. De esa forma, es posible que la transcripción se realice mecánicamente.

Pronunciable

La transcripción no es tan útil si el proceso simplemente asigna los caracteres independientemente de su pronunciación. Asignar simplemente "αβγδεζηθ..." a "abcdefgh..." produciría cadenas que pueden ser completas e inequívocas, pero imposibles de pronunciar.

Con frecuencia, los métodos de transcripción estándar no siguen las reglas de pronunciación de ningún idioma en particular en el sistema de escritura de destino. Por ejemplo, el sistema Hepburn japonés utiliza una "j" que tiene el valor fonético del inglés (en oposición al del francés, el alemán o el español), pero utiliza vocales que no tienen los sonidos estándares del inglés. Un método de transcripción también puede requerir cierto conocimiento especial para lograr la pronunciación correcta. Por ejemplo, en el sistema japonés kunrei-siki, "tu" se pronuncia "tsu". Es similar a las situaciones en las que existen diferentes idiomas dentro del mismo alfabeto. Por ejemplo, si se sabe que la palabra Gewalt proviene del alemán, un lector entendido pronunciará la "w" como una "v".

En algunos casos, la transcripción puede ser fuertemente influenciada por la tradición. Por ejemplo, la letra griega moderna beta (β) suena como una "v", pero una transformación posiblemente continúe utilizándola como una b (como en "biología"). En ese caso, el usuario necesitaría saber que una "b" en la palabra transcrita corresponde a beta (β) y se pronuncia como una "v" en griego moderno. Las letras también pueden ser transcritas de forma diferente de acuerdo con el contexto para hacer que la pronunciación sea más predecible. Por ejemplo, puesto que la secuencia del griego GAMMA GAMMA ($\gamma\gamma$) se pronuncia como "ng", la primera GAMMA puede ser transcrita como una "n".

Nota: En general, para producir resultados predecibles en la transcripción de alfabetos latinos a otros alfabetos, el texto en inglés no producirá resultados fonéticos. Ello se debe a que la pronunciación del inglés no puede ser predicha fácilmente a partir de las letras que conforman una palabra. Por ejemplo, las palabras "grove", "move" y "love" terminan con las letras "ove", pero se pronuncian de maneras muy diferentes.

Inequívoca

Siempre debería ser posible recuperar el texto en el tipo de escritura de origen a partir de la transcripción en el tipo de escritura de destino. Por ejemplo, debería ser posible volver de Elláda al original Ελλάδα. Sin embargo, en la transcripción muchos caracteres pueden producir ambigüedades. Por ejemplo, el carácter griego PSI (ψ) se asigna a ps, pero ps también podría ser el resultado de la secuencia PI, SIGMA($\pi\sigma$) ya que PI (π) se asigna a p y SIGMA (σ) se asigna a s.

Para solucionar el problema de la ambigüedad, Transliterator utiliza un apóstrofo para desambiguar las secuencias de caracteres. Al utilizar este procedimiento, el carácter griego PI SIGMA ($\pi\sigma$) se asigna a p's. En japonés, cuando una secuencia ambigua en el sistema de escritura de destino no proviene de una única letra, la transcripción utiliza un apóstrofo para desambiguarla. Por ejemplo, utiliza este procedimiento para distinguir entre man'ichi y manichi.

Nota: Algunos caracteres del sistema de escritura de destino, por lo general, no se encuentran fuera de ciertos contextos. Por ejemplo, el carácter japonés "ya" pequeño, que puede encontrarse en "kya" (キヤ), normalmente no aparece solo. Para marcar una distinción entre esos caracteres, Transliterator utiliza una tilde. Por ejemplo, la entrada "~ya" produciría una "ya" pequeña aislada. Cuando se transcribe al griego, la entrada "a~s" produciría una sigma ($\alpha\sigma$) no final del griego al final de una palabra. De la misma manera, la entrada "~sa" produciría una sigma final en una posición no final ($\varsigma\alpha$).

Para las transformaciones generales de alfabetos, una técnica frecuente a fin de lograr la reversibilidad es utilizar acentos que distingan letras que, de lo contrario, no podrían distinguirse. Por ejemplo, el siguiente ejemplo muestra texto en griego que se asigna al latín y es completamente reversible:

Entrada

Nombre de campo	Descripción
Any string field	La etapa Transliterator puede transcribir cualquier campo de cadena. Puede especificar los campos a transcribir en las opciones de la etapa Transliterator.

Nombre de campo	Descripción										
TransliteratorID	<p>Se anula la transcripción predeterminada especificada en las opciones de la etapa Transliterador. Utilice este campo si desea especificar una transcripción diferente para cada registro.</p> <p>Por ejemplo:</p> <table border="0"> <tr> <td>Arabic-Latin</td> <td>De árabe a latino.</td> </tr> <tr> <td>Greek-Latin</td> <td>De griego a latino.</td> </tr> <tr> <td>Latin-Hangul</td> <td>De latino a hangul.</td> </tr> <tr> <td>Latin-Katakana</td> <td>De latino a katakana.</td> </tr> <tr> <td>Fullwidth-Halfwidth</td> <td>Del ancho completo al ancho medio.</td> </tr> </table>	Arabic-Latin	De árabe a latino.	Greek-Latin	De griego a latino.	Latin-Hangul	De latino a hangul.	Latin-Katakana	De latino a katakana.	Fullwidth-Halfwidth	Del ancho completo al ancho medio.
Arabic-Latin	De árabe a latino.										
Greek-Latin	De griego a latino.										
Latin-Hangul	De latino a hangul.										
Latin-Katakana	De latino a katakana.										
Fullwidth-Halfwidth	Del ancho completo al ancho medio.										

Opciones

Tabla 27: Opciones de Transliterador

Opción	Descripción / Valores válidos
Desde	<p>Alfabeto utilizado por los campos que desea transcribir. Para obtener una descripción de los alfabetos admitidos, consulte Transliterador en la página 298.</p> <p>Nota: La etapa Transliterador no admite la transcripción entre todos los alfabetos. Los campos De y A reflejan automáticamente los valores válidos según su selección.</p>
Para	<p>El alfabeto al que quiere convertir el campo. Para obtener una descripción de los alfabetos admitidos, consulte Transliterador en la página 298.</p> <p>Nota: La etapa Transliterador no admite la transcripción entre todos los alfabetos. Los campos De y A reflejan automáticamente los valores válidos según su selección.</p>
Botón Intercambiar	<p>Haga clic en el botón Intercambiar para intercambiar los idiomas de los campos De y A.</p>



Opción	Descripción / Valores válidos
Campos a transcribir	Especifica los campos que desea transcribir.
Eliminar marcas de acento	Select the check box to remove accent marks from the field. This option remains turned off by default.

Salida

La etapa Transliterador transcribe los campos que usted especifica. No produce ninguna otra salida.

Módulo Universal Name

Módulo Universal Name

Para llevar a cabo la normalización más precisa, puede ser necesario dividir las cadenas de datos en varios campos. Spectrum™ Technology Platform proporciona características de análisis avanzado que le permiten analizar nombres personales, nombres de empresas y muchos otros términos y abreviaturas. Además, puede crear su propia lista de términos personalizados para usar como base para las operaciones de exploración y extracción.

Componentes

El Módulo Universal Name consta de:

- **Name Parser (obsoleto):** el componente Name Parser (Analizador de nombres) divide los nombres personales, nombres de empresas y otros términos del campo de datos de nombres en las partes que los conforman. De este modo, estos elementos de nombre analizados se encuentran disponibles para otras operaciones automatizadas tales como comparación y estandarización de nombres o consolidación de nombres de múltiples registros.

Atención: La etapa Name Parser se ha vuelto obsoleta y es posible que no se admita en futuras versiones. Utilice Open Name Parser para analizar nombres.

- **Name Variant Finder:** este componente funciona en modo de nombre o apellido para realizar consultas en la base de datos de nombres y obtener variaciones de nombre. Name Variant Finder puede limitar la cantidad de resultados por género y cultura. Spectrum™ Technology Platform incluye un archivo base de nombres. Hay archivos de nombres agregados de otras culturas disponibles y pueden implementarse al copiar el archivo JAR correspondiente en la carpeta

`modules/tables/ext`. Los nombres en formato árabe se encuentran disponibles como complementos. Los datos de este archivo provienen de Nomino, Inc.

- **Open Name Parser:** divide los nombres personales, nombres de empresas y otros términos del campo de datos de nombres en las partes que los conforman. De este modo, estos elementos de nombre analizados se encuentran disponibles para otras operaciones automatizadas tales como comparación y estandarización de nombres o consolidación de nombres de múltiples registros.

Name Parser (OBSOLETO)

Atención: La etapa Name Parser se ha vuelto obsoleta y es posible que no se admita en futuras versiones. Utilice Open Name Parser para analizar nombres.

Name Parser (Analizador de nombres) divide los nombres personales, nombres de empresas y otros términos del campo de datos de nombres en las partes que los conforman. El proceso de análisis incluye una explicación de la función, la forma y la relación sintáctica de cada parte con el todo. De este modo, estos elementos de nombre analizados se encuentran disponibles para otras operaciones automatizadas tales como comparación y estandarización de nombres o consolidación de nombres de múltiples registros.

El proceso de análisis de nombres ejecuta las siguientes funciones:

- Determina el tipo de entidad de un nombre para describir la función que cumple ese nombre. Los tipos de entidades de nombre se dividen en dos grupos principales: nombres personales y nombres de empresas, ambos con subgrupos.
- Determina la forma de un nombre para comprender la sintaxis que debe seguir la herramienta de análisis. Por lo general, los nombres personales siguen un orden natural (firma) o un orden inverso, en tanto los nombres de empresas se ordenan jerárquicamente.
- Determina y etiqueta los componentes de un nombre de forma tal de identificar la relación sintáctica de cada parte del nombre con el nombre completo. La sintaxis de nombres personales incluye prefijos, partes del nombre, segundo nombre y apellido, sufijos y términos de descripciones de cuentas, entre otras partes del nombre. La sintaxis de nombres de empresas incluye el texto principal, términos insignificantes, preposiciones, objetos de la preposición y términos de sufijos, entre otras partes del nombre.
- Determina el género del nombre. El género se establece sobre la base de las premisas culturales que se especifican. Por ejemplo, Jean es un nombre masculino en Francia, pero es un nombre femenino en Estados Unidos. Si usted ya sabe que los nombres que se procesan son franceses, puede especificar Francia como cultura de determinación de género. Name Parser utiliza datos de las tablas de nombres y nombres compuestos para determinar el género. Si un nombre no aparece en ninguna tabla e incluye un título, el analizador consulta la tabla de títulos para determinar el género. Si no se obtienen resultados, el género se marca como desconocido.

Nota: Si un campo en un registro de entrada contiene una de las culturas admitidas, puede definir previamente el campo GenderDeterminationSource en la entrada para invalidar el origen de determinación de género en la interfaz gráfica del usuario (GUI).

- Asigna una calificación de análisis que indica el grado de confianza del analizador respecto de la exactitud del análisis.

Input

Atención: La etapa Name Parser se ha vuelto obsoleta y es posible que no se admita en futuras versiones. Utilice Open Name Parser para analizar nombres.

Tabla 28: Datos de entrada de Name Parser

Nombre de campo	Descripción / Valores válidos																										
GenderDeterminationSource	<p>La cultura de los datos de nombre que se utilizará para determinar el género. La opción predeterminada utiliza reglas de diferentes culturas. Por ejemplo, "Jean" generalmente es un nombre femenino y la opción predeterminada lo identifica como tal, pero se identifica como un nombre masculino si se selecciona la cultura francesa. Las opciones se enumeran a continuación junto con países de ejemplo para cada cultura. Cabe destacar de la lista de países de cada cultura no es exhaustiva.</p> <table border="0"> <tr> <td>SLAVIC</td> <td>Bosnia, Polonia, Albania.</td> </tr> <tr> <td>ARMENIAN</td> <td>Armenia.</td> </tr> <tr> <td>DEFAULT</td> <td>Bulgaria, Islas Caimán, Irlanda, Estados Unidos, Reino Unido.</td> </tr> <tr> <td>FRENCH</td> <td>Francia.</td> </tr> <tr> <td>SCANDINAVIAN</td> <td>Dinamarca, Finlandia, Islandia, Noruega, Suecia.</td> </tr> <tr> <td>GERMANIC</td> <td>Austria, Alemania, Luxemburgo, Suiza, Países Bajos.</td> </tr> <tr> <td>GREEK</td> <td>Grecia.</td> </tr> <tr> <td>HUNGARIAN</td> <td>Hungría.</td> </tr> <tr> <td>ITALIAN</td> <td>Italia.</td> </tr> <tr> <td>PORTUGUESE</td> <td>Portugal.</td> </tr> <tr> <td>ROMANIA</td> <td>Rumania.</td> </tr> <tr> <td>HISPANIC</td> <td>España.</td> </tr> <tr> <td>ARABIC</td> <td>Túnez.</td> </tr> </table> <p>Name Variant Finder también utiliza GenderDeterminationSource para limitar las variaciones de nombre obtenidas según la cultura. Para obtener más información, consulte Name Variant Finder en la página 326.</p>	SLAVIC	Bosnia, Polonia, Albania.	ARMENIAN	Armenia.	DEFAULT	Bulgaria, Islas Caimán, Irlanda, Estados Unidos, Reino Unido.	FRENCH	Francia.	SCANDINAVIAN	Dinamarca, Finlandia, Islandia, Noruega, Suecia.	GERMANIC	Austria, Alemania, Luxemburgo, Suiza, Países Bajos.	GREEK	Grecia.	HUNGARIAN	Hungría.	ITALIAN	Italia.	PORTUGUESE	Portugal.	ROMANIA	Rumania.	HISPANIC	España.	ARABIC	Túnez.
SLAVIC	Bosnia, Polonia, Albania.																										
ARMENIAN	Armenia.																										
DEFAULT	Bulgaria, Islas Caimán, Irlanda, Estados Unidos, Reino Unido.																										
FRENCH	Francia.																										
SCANDINAVIAN	Dinamarca, Finlandia, Islandia, Noruega, Suecia.																										
GERMANIC	Austria, Alemania, Luxemburgo, Suiza, Países Bajos.																										
GREEK	Grecia.																										
HUNGARIAN	Hungría.																										
ITALIAN	Italia.																										
PORTUGUESE	Portugal.																										
ROMANIA	Rumania.																										
HISPANIC	España.																										
ARABIC	Túnez.																										
Name	El nombre que desea analizar. Este campo es obligatorio.																										

Opciones

Atención: La etapa Name Parser se ha vuelto obsoleta y es posible que no se admita en futuras versiones. Utilice Open Name Parser para analizar nombres.

Para especificar las opciones de Name Parser, haga doble clic en la instancia de Name Parser en el lienzo. Aparece el diálogo Name Parser Options.

Tabla 29: Opciones de Name Parser

Opción	Descripción
Analizar nombres personales	Marque esta casilla para analizar nombres personales.
<p>Separe nombres conjuntos en varios registros Seleccione un resultado de cruce en la lista Resultados de cruce y haga clic en Eliminar.</p>	<p>Haga clic en esta casilla para separar los nombres que pertenecen a más de una persona en múltiples registros, por ejemplo, Bill & Sally Smith.</p> <p>Cuando se obtienen dos registros de nombre separados a partir de un registro conjunto, se genera un campo de salida ID de análisis de registro. Cada par de registros de nombre separados se identifican con la misma ID de análisis de registro.</p>
<p>Fuente de determinación de género Seleccione todos los resultados de la Lista Resultados de cruce y haga clic en Eliminar.</p>	<p>Determina cómo Name Parser asigna un género al nombre. En la mayoría de los casos, la configuración predeterminada es la mejor opción porque cubre una amplia variedad de nombres. Si está procesando nombres de una cultura específica, seleccione esa cultura. La selección de una cultura específica contribuye a garantizar la asignación del género correcto a los nombres. Por ejemplo, si se deja la opción predeterminada seleccionada, el nombre Jean será identificado como un nombre femenino. Si selecciona que la cultura es francesa, el nombre será identificado como un nombre masculino.</p> <p>Nota: Si selecciona una cultura, pero el nombre no se encuentra en esa cultura, el género se determina utilizando la cultura predeterminada, que incluye datos de varias culturas.</p>
Orden	<p>Especifica cómo los campos de nombre se ordenan en sus registros de entrada. Una de las siguientes:</p> <p>Natural Los campos de nombre están ordenados por tratamiento, nombre, segundo nombre, apellido y sufijo.</p> <p>Revertir Los campos de nombre están ordenados según el apellido en primer lugar.</p> <p>Combinado Los campos de nombre se ordenan utilizando una combinación del criterio Natural y Revertir.</p>

Opción	Descripción
Retener períodos	Conserva la puntuación en el campo de nombre personal analizado.
Analizar nombres de empresas	Marque esta casilla para analizar nombres de empresas.
Retener períodos	Marque esta casilla para que la puntuación vuelva al campo de nombre de empresa analizado.
Tabla definida por el usuario	Haga clic en cualquier tabla definida por el usuario para agregar valores a los valores existentes en las distintas tablas de análisis. Esta opción permite personalizar tablas para su entorno de negocios específico. Haga clic en Configurar para seleccionar un archivo XML que contenga los valores que desea agregar. Para obtener más información sobre tablas definidas por el usuario, consulte Modificación de tablas de Name Parser definidas por el usuario en la página 308.

Modificación de tablas de Name Parser definidas por el usuario

Atención: La etapa Name Parser se ha vuelto obsoleta y es posible que no se admita en futuras versiones. Utilice Open Name Parser para analizar nombres.

Se puede agregar, modificar y eliminar valores en las tablas de Name Parser a fin de adaptarlas a su entorno de negocios específico

Las tablas de Name Parser definidas por el usuario son archivos XML ubicados de forma predeterminada en la carpeta <Drive>:\Program Files\Pitney Bowes\Spectrum\server\modules\parser\data. Spectrum™ Technology Platform incluye las siguientes tablas definidas por el usuario:

UserAccountDescriptions.xml

Tabla 30: Columnas UserAccountDescriptions.xml

Nombre de columna	Descripción / Valores válidos
LookupValue	Término de búsqueda que comúnmente se encuentra en una Descripción de cuenta. Cualquier texto de una sola palabra. Sin distinción de mayúsculas y minúsculas

Entrada de ejemplo:

```
<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

      LookupValue
      ART
      AND

    </deleted-entry-group>
  </deleted-entries>
  <added-entries delimiter-character="|">

    LookupValue
    A/C
    ACCOUNT
    EXP

  </added-entries>
</table-data>
```

[UserCompanyPrepositions.xml](#)

Tabla 31: Columnas UserCompanyPrepositions.xml

Nombre de columna	Descripción / Valores válidos
LookupValue	Cualquier preposición (por ejemplo, "de" o "en") que comúnmente se encuentra en los nombres de empresas. Cualquier texto de una sola palabra. Sin distinción de mayúsculas y minúsculas

Entrada de ejemplo:

```
<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

      LookupValue
      AROUND
      NEAR

    </deleted-entry-group>
  </deleted-entries>
```

```

<added-entries delimiter-character="|">
    LookupValue
    ABOUT
    AFTER
    ACROSS
</added-entries>
</table-data>

```

UserCompanySuffixes.xml

Tabla 32: Columnas UserCompanySuffixes.xml

Nombre de columna	Descripción / Valores válidos
LookupValue	Cualquier sufijo que comúnmente se encuentra en los nombres de las empresas. Los ejemplos incluyen "SRL" y "SA". Cualquier texto de una sola palabra. Sin distinción de mayúsculas y minúsculas

Entrada de ejemplo:

```

<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>
      LookupValue
      SANDY
      CLUE
    </deleted-entry-group>
  </deleted-entries>
  <added-entries delimiter-character="|">
    LookupValue
    LTD
    LLC
    CO
    INC
  </added-entries>
</table-data>

```

[UserCompanyTerms.xml](#)**Tabla 33: Columnas UserCompanyTerms.xml**

Nombre de columna	Descripción / Valores válidos
LookupValue	Cualquier término que comúnmente se encuentra en el nombre de una empresa. Cualquier texto de una sola palabra. Sin distinción de mayúsculas y minúsculas

Entrada de ejemplo:

```
<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

      LookupValue
      MARY
      BLUE

    </deleted-entry-group>
  </deleted-entries>
  <added-entries delimiter-character="|">

    LookupValue
    ARC
    ARCADE
    ASSEMBLY
    ARIZONA

  </added-entries>
</table-data>
```

[UserCompoundFirstNames.xml](#)

Esta tabla contiene nombres compuestos definidos por el usuario. Los nombres compuestos son nombres que tienen dos palabras.

Tabla 34: Columnas UserCompoundFirstNames.xml

Nombre de columna	Descripción / Valores válidos
FirstName	El primer nombre compuesto. Dos palabras como máximo. Sin distinción de mayúsculas y minúsculas
Cultura	La cultura en que se aplica esta combinación FirstName/Gender (Nombre/Género). Puede utilizar cualquiera de los valores que son válidos en el campo de entrada GenderDeterminationSource. Para obtener más información, consulte Input en la página 306.
Género	El género que comúnmente se asocia con esta combinación FirstName/Culture (Nombre/Cultura). Una de las siguientes: <ul style="list-style-type: none"> M El nombre es masculino. F El nombre es femenino. A Ambiguo. El nombre puede ser masculino o femenino. U Desconocido. El género correspondiente a este nombre es desconocido. Esta opción se presume si el campo está vacío.
Frecuencia	No se utiliza en este lanzamiento. Puede dejar esta columna vacía.

Entrada de ejemplo:

```
<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

      FirstName
      ANN MARIE
      BILLY JOE

    </deleted-entry-group>
    <deleted-entry-group>

      FirstName|Frequency
      KAREN SUE|0.126
      BILLY JOE|0.421

    </deleted-entry-group>
  </deleted-entries>
</table-data>
```



```

        FirstName|Gender|Culture
        JEAN ANN|M|DEFAULT
        JEAN CLUADE|F|FRENCH

    </deleted-entry-group>
</deleted-entries>
<added-entries delimiter-character="|">

        FirstName|Gender|Culture
        JOHN Henry|M|DEFAULT
        A'SHA A'MAR|F|ARABIC
        BILLY JO|A|DEFAULT

    </added-entries>
</table-data>

```

UserConjunctions.xml

Esta tabla contiene una lista de conjunciones definidas por el usuario, por ejemplo "y", "o" o "&".

Tabla 35: Columnas UserConjunctions.xml

Nombre de columna	Descripción / Valores válidos
LookupValue	Cualquier conjunción. Debe ser una sola palabra. Sin distinción de mayúsculas y minúsculas

Entradas de ejemplo:

```

<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

        LookupValue
        FIND
        CARE
        %

    </deleted-entry-group>
  </deleted-entries>
  <added-entries delimiter-character="|">

        LookupValue
        &
        AND

```

OR

```
</added-entries>
</table-data>
```

UserFirstNames.xml

Tabla 36: Columnas UserFirstNames.xml

Nombre de columna	Descripción / Valores válidos
FirstName	El primer nombre descrito en esta fila de la tabla. Sin distinción de mayúsculas y minúsculas
Género	<p>El género que comúnmente se asocia con esta combinación FirstName/Culture (Nombre/Cultura). Una de las siguientes:</p> <p>M El nombre es masculino.</p> <p>F El nombre es femenino.</p> <p>A Ambiguo. El nombre puede ser masculino o femenino.</p> <p>U Desconocido. El género correspondiente a este nombre es desconocido. Esta opción se presume si el campo está vacío.</p>
Cultura	La cultura en que se aplica esta combinación FirstName/Gender (Nombre/Género). Puede utilizar cualquiera de los valores que son válidos en el campo de entrada GenderDeterminationSource. Para obtener más información, consulte Input en la página 306.

Entrada de ejemplo:

```
<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

      FirstName
      AADEL
      AADIL

    </deleted-entry-group>
  </deleted-entry-group>
```

```

        FirstName
        A 'SACE
        A 'BOCKETT

</deleted-entry-group>
<deleted-entry-group>

        FirstName|Gender|Culture
        ALII|M|DEFAULT
        AISHA|F|ARABIC

</deleted-entry-group>
<deleted-entry-group>

        FirstName|Gender
        JOHE|M

</deleted-entry-group>
</deleted-entries>
<added-entries delimiter-character="|">

        FirstName|Gender|Culture
        JOHE|M|DEFAULT
        A 'SHAN|F|ARABIC

</added-entries>
</table-data>

```

[UserGeneralSuffixes.xml](#)

Esta tabla contiene una lista de sufijos definidos por el usuario en nombres personales que no son sufijos generacionales, como "MD" o "PhD".

Tabla 37: Columnas UserGeneralSuffixes.xml

Nombre de columna	Descripción / Valores válidos
LookupValue	Cualquier sufijo que se aplique frecuentemente a nombres personales y no sea un sufijo generacional. Debe ser una sola palabra. Sin distinción de mayúsculas y minúsculas

Entrada de ejemplo:

```

<table-data>
  <deleted-entries delimiter-character="|">

```

```

    <deleted-entry-group>
        LookupValue
        AND
        WILL
        TUNA
    </deleted-entry-group>
</deleted-entries>
<added-entries delimiter-character="|">
    LookupValue
    ACCOUNTANT
    ATTORNEY
    ANALYST
    ASSISTANT
</added-entries>
</table-data>

```

UserLastNamePrefixes.xml

Esta tabla contiene una lista de los prefijos definidos por el usuario que forman parte del apellido de una persona, por ejemplo, "Van", "De" o "La".

Tabla 38: Columnas UserLastNamePrefixes.xml

Nombre de columna	Descripción / Valores válidos
LookupValue	Cualquier prefijo que forme parte del apellido de una persona. Cualquier texto de una sola palabra. Sin distinción de mayúsculas y minúsculas

Entrada de ejemplo:

```

<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>
        LookupValue
        DO
        RUN
        ANIMAL
    </deleted-entry-group>
  </deleted-entries>

```

```

<added-entries delimiter-character="|">
    LookupValue
    D'
    DA
    DEN
    DEL
</added-entries>
</table-data>

```

UserLastNames.xml

Tabla 39: Columnas UserLastNames.xml

Nombre de columna	Descripción / Valores válidos
LastName	El apellido descrito en esta fila de la tabla. Sin distinción de mayúsculas y minúsculas
Género	<p>El género que comúnmente se asocia con esta combinación FirstName/Culture (Nombre/Cultura). Una de las siguientes:</p> <p>M El nombre es masculino.</p> <p>F El nombre es femenino.</p> <p>A Ambiguo. El nombre puede ser masculino o femenino.</p> <p>U Desconocido. El género correspondiente a este nombre es desconocido. Esta opción se presume si el campo está vacío.</p>
Cultura	La cultura en que se aplica esta combinación FirstName/Gender (Nombre/Género). Puede utilizar cualquiera de los valores que son válidos en el campo de entrada GenderDeterminationSource. Para obtener más información, consulte Input en la página 306.

Entrada de ejemplo:

```

<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>
      LastName
      Rusod
    </deleted-entry-group>
  </deleted-entries>
</table-data>

```

```

AADIL

</deleted-entry-group>
<deleted-entry-group>

    LastName
    KAASEEY
    JOIEN

</deleted-entry-group>
</deleted-entries>
<added-entries delimiter-character="|">

    LastName|Culture|Gender
    SMITH|ENGLISH|A
    WILSON|ENGLISH|A
    JONES|ENGLISH|A

</added-entries>
</table-data>

```

[UserMaturitySuffixes.xml](#)

Esta tabla contiene sufijos generacionales definidos por el usuario utilizados en los nombres de personas, como "Jr." o "Sr."

Tabla 40: Columnas UserMaturitySuffixes.xml

Nombre de columna	Descripción / Valores válidos
LookupValue	Sufijo generacional utilizado en los nombres de personas. Cualquier texto de una sola palabra. Sin distinción de mayúsculas y minúsculas

Entrada de ejemplo:

```

<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

        LookupValue
        I
        V
        18
        VI
    
```

```

        </deleted-entry-group>
    </deleted-entries>
    <added-entries delimiter-character="|">
        LookupValue
        I
        II
        III
    </added-entries>
</table-data>

```

UserTitles.xml

Esta tabla contiene títulos de tratamiento definidos por el usuario que se utilizan con los nombres de personas, por ejemplo "Sr." o "Sra."

Tabla 41: Columnas UserTitles.xml

Nombre de columna	Descripción / Valores válidos
LookupValue	Título de tratamiento que se utiliza con los nombres de personas. Cualquier texto de una sola palabra. Sin distinción de mayúsculas y minúsculas
Género	El género más frecuentemente asociado con ese título. Una de las siguientes: <ul style="list-style-type: none"> M El nombre es masculino. F El nombre es femenino. A Ambiguo. El nombre puede ser masculino o femenino. U Desconocido. El género correspondiente a este nombre es desconocido. Esta opción se presume si el campo está vacío.

Entrada de ejemplo:

```

<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>
        LookupValue
        Belt
        Friend
        Thursday
    </deleted-entry-group>
  </deleted-entries>
</table-data>

```

```

                Red

        </deleted-entry-group>
</deleted-entries>
<added-entries delimiter-character="|">

        LookupValue|Gender
        Mrs|F
        Mr|M
        Most|F

</added-entries>
</table-data>

```

Tabla definida por el usuario de ejemplo

La siguiente figura muestra una tabla UserFirstNames.xml de ejemplo y la sintaxis que se utiliza cuando se modifican las tablas definidas por el usuario.

```

<table-data>
  <deleted-entries delimiter-character="|">
    <deleted-entry-group>

        FirstName
        AADEL
        AADIL

    </deleted-entry-group>
  </deleted-entry-group>

        FirstName|Frequency
        A'SACE|0.126
        A'BECKETT|0.421

    </deleted-entry-group>
  </deleted-entry-group>

        FirstName|Gender|Culture|VariantGroup
        ALI|M|DEFAULT|GROUP88
        AISHA|F|ARABIC|GROUP43

    </deleted-entry-group>
  </deleted-entry-group>

        FirstName|Gender
        JOHN|M

    </deleted-entry-group>
  </deleted-entries>
<added-entries delimiter-character="|">

```



```

      FirstName | Gender | Culture
      JOHN | M | DEFAULT
      A' SHA | F | ARABIC
      JAMES | M | DEFAULT

    </added-entries>
  </table-data>

```

Salida

Atención: La etapa Name Parser se ha vuelto obsoleta y es posible que no se admita en futuras versiones. Utilice Open Name Parser para analizar nombres.

Tabla 42: Datos de salida de Name Parser

Nombre de campo	Formato	Descripción / Valores válidos
AccountDescription	Cadena	Descripción de cuenta que es parte del nombre. Por ejemplo, en "Mary Jones Account # 12345", la descripción de cuenta es "Account#12345".
EntityType	Cadena	Indica el tipo de nombre. Una de las siguientes: <ul style="list-style-type: none"> Firm El nombre es un nombre de empresa. Personal El nombre es un nombre de persona.
Campos relacionados con nombres de empresas		
FirmModifier.1.Object	Cadena	El primer objeto de una preposición que aparece en el nombre de una empresa. Por ejemplo, en el nombre de la empresa "División Pratt & Whitney de United Technologies", el primer objeto de la preposición es "United Technologies".
FirmModifier.1.Preposition	Cadena	La primera preposición que aparece en el nombre de una empresa. Por ejemplo, en el nombre de la empresa "División Pratt & Whitney de United Technologies", "de" sería la primera preposición.

Nombre de campo	Formato	Descripción / Valores válidos
FirmModifier.2.Object	Cadena	El segundo objeto de una preposición que aparece en el nombre de una empresa. Por ejemplo, en el nombre de la empresa "Iglesia de Nuestra Señora de Lourdes", el segundo objeto de una preposición es "Lourdes".
FirmModifier.2.Preposition	Cadena	La segunda preposición que aparece en el nombre de una empresa. Por ejemplo, en el nombre de la empresa "Iglesia de Nuestra Señora de Lourdes", la segunda preposición es "de".
FirmName	Cadena	El nombre de una empresa. Por ejemplo, "Pitney Bowes, Inc."
FirmPrimary	Cadena	La parte principal del nombre de una empresa. Por ejemplo, "Pitney Bowes".
FirmSuffix	Cadena	El sufijo corporativo. Por ejemplo, "SRL" y "SA".
Campos relacionados con nombres de personas		
FirstName	Cadena	El primer nombre de la persona.
FirstNameVariantGroup	Cadena	Una ID numérica que indica el grupo de nombres similares al que pertenece este nombre. Por ejemplo, Muhammad, Mohammed y Mehmet corresponden al mismo Grupo de variante de nombre. La ID de grupo real se asigna cuando se cargan los datos complementarios. Este campo solo se completa si ha comprado la función Grupo de variante de nombre.

Nombre de campo	Formato	Descripción / Valores válidos
GenderCode	Cadena	Género de una persona según lo determinado a partir del análisis del primer nombre. Una de las siguientes: A Ambiguo. El nombre es tanto masculino como femenino. Por ejemplo, Pat. F Femenino. El nombre es femenino. M Masculino. El nombre es masculino. U Desconocido. No se encontró el nombre en la tabla de género.
GenderDeterminationSource	Cadena	Cultura utilizada para determinar el género del nombre. Si no se pudo encontrar el nombre en la tabla de género, este campo está vacío.
GeneralSuffix	Cadena	El sufijo general o profesional de una persona. Por ejemplo, MD o PhD.
LastName	Cadena	El apellido de la persona.
MaturitySuffix	Cadena	Un sufijo generacional de una persona. Por ejemplo, Jr. o Sr.
MiddleName	Cadena	El segundo nombre de una persona.
NameScore	Cadena	Puntuación que representa la calidad de la operación de análisis, de 0 a 100. 0 indica baja calidad y 100 indica la calidad más alta.
ParserRecordID	Cadena	ID única asignada a cada registro de entrada.
TitleOfRespect	Cadena	Título de una persona, por ejemplo Sr., Sra., Dr. o Rdo.
Campos relacionados con nombres conjuntos		
PersonalName.2.FirstName	Cadena	El primer nombre de la segunda persona de un nombre conjunto. Un ejemplo de nombre conjunto es "John y Jane Smith".

Nombre de campo	Formato	Descripción / Valores válidos
PersonalName.2.FirstNameVariantGroup	Cadena	<p>ID numérica que indica el grupo de nombres similares al que pertenece el primer nombre de la segunda persona de un nombre conjunto. Por ejemplo, Muhammad, Mohammed y Mehmet corresponden al mismo Grupo de variante de nombre. La ID de grupo real se asigna cuando se cargan los datos complementarios.</p> <p>Este campo solo se completa si ha comprado la función Grupo de variante de nombre.</p>
PersonalName.2.GenderCode	Cadena	<p>El género de la segunda persona de un nombre conjunto determinado a partir del análisis del primer nombre mediante Name Parser. Un ejemplo de nombre conjunto es "John y Jane Smith". Uno de los siguientes:</p> <p>A Ambiguo. El nombre es tanto masculino como femenino. Por ejemplo, Pat.</p> <p>F Femenino. El nombre es femenino.</p> <p>M Masculino. El nombre es masculino.</p> <p>U Desconocido. No se encontró el nombre en la tabla de género.</p>
PersonalName.2.GenderDeterminationSource	Cadena	<p>La cultura utilizada para determinar el género de la segunda persona de un nombre conjunto. Un ejemplo de nombre conjunto es "John y Jane Smith".</p>
PersonalName.2.GeneralSuffix	Cadena	<p>El sufijo general o profesional de la segunda persona de un nombre conjunto. Un ejemplo de nombre conjunto es "John y Jane Smith". Ejemplos de sufijos generales son Médico y Doctorado.</p>
PersonalName.2.LastName	Cadena	<p>El apellido de la segunda persona de un nombre conjunto. Un ejemplo de nombre conjunto es "John y Jane Smith".</p>
PersonalName.2.MaturitySuffix	Cadena	<p>El sufijo generacional de la segunda persona de un nombre conjunto. Un ejemplo de nombre conjunto es "John y Jane Smith". Ejemplos de sufijos generacionales son Jr. y Sr.</p>
PersonalName.2.MiddleName	Cadena	<p>El segundo nombre de la segunda persona de un nombre conjunto. Un ejemplo de nombre conjunto es "John y Jane Smith".</p>

Nombre de campo	Formato	Descripción / Valores válidos
PersonalName.2.TitleOfRespect	Cadena	El título de tratamiento para el segundo nombre en un nombre conjunto. Por ejemplo, "Sr. and Mrs. Smith" es un nombre conjunto. Ejemplos de título de tratamiento son Sr., Sra. y Dr.
PersonalName.3.FirstName	Cadena	El primer nombre de la tercera persona de un nombre conjunto. Por ejemplo, "Sr. y Sra. Smith y Dr. Mary Jones" es un nombre conjunto.
PersonalName.3.FirstNameVariantGroup	Cadena	ID numérica que indica el grupo de nombres similares al que pertenece el primer nombre de la segunda persona de un nombre conjunto. Por ejemplo, Muhammad, Mohammed y Mehmet corresponden al mismo Grupo de variante de nombre. La ID de grupo real se asigna cuando se cargan los datos complementarios. Este campo solo se completa si ha comprado la función Grupo de variante de nombre.
PersonalName.3.GenderCode	Cadena	El género de la tercera persona de un nombre conjunto determinado a partir del análisis del primer nombre mediante Name Parser. Un ejemplo de un nombre conjunto es "Sr. y Sra. Smith y Adam Jones". Una de las siguientes: <ul style="list-style-type: none"> A Ambiguo. El nombre es tanto masculino como femenino. Por ejemplo, Pat. F Femenino. El nombre es femenino. M Masculino. El nombre es masculino. U Desconocido. No se encontró el nombre en la tabla de género.
PersonalName.3.GenderDeterminationSource	Cadena	La cultura utilizada para determinar el género de la tercera persona de un nombre conjunto. "Sr. y Sra. Smith y Adam Jones".
PersonalName.3.GeneralSuffix	Cadena	El sufijo general o profesional de la tercera persona de un nombre conjunto. Un ejemplo de un nombre conjunto es "Sr. y Sra. Smith y Adam Jones, PhD." Ejemplos de sufijos son MD y PhD.
PersonalName.3.LastName	Cadena	El apellido de la tercera persona de un nombre conjunto. Por ejemplo, "Sr. y Sra. Smith y Dr. Mary Jones" es un nombre conjunto.

Nombre de campo	Formato	Descripción / Valores válidos
PersonalName.3.MaturitySuffix	Cadena	El sufijo generacional de la tercera persona de un nombre conjunto. Un ejemplo de un nombre conjunto es "Sr. y Sra. Smith y Adam Jones Sr." Ejemplos de sufijos de madurez son Jr. y Sr.
PersonalName.3.MiddleName	Cadena	El segundo nombre de la tercera persona de un nombre conjunto. Por ejemplo, "Sr. y Sra. Smith y Dr. Mary Jones" es un nombre conjunto.
PersonalName.3.TitleOfRespect	Cadena	El título de tratamiento de la tercera persona de un nombre conjunto. Por ejemplo, "Sr. y Sra. Smith y Dr. Mary Jones" es un nombre conjunto. Ejemplos de título de tratamiento son Sr., Sra. y Dr.

Name Variant Finder

Name Variant Finder (Localizador de variantes de nombres) funciona en modo de nombre o apellido para realizar consultas en una base de datos a fin de obtener versiones alternativas de un nombre. Por ejemplo, "John" y "Jon" son variantes del nombre "Johnathan". Name Variant Finder requiere diccionarios complementarios que pueden instalarse mediante la utilidad de carga de bases de datos de los módulos Universal Name, Data Normalization y Advanced Matching. Comuníquese con el representante de ventas para obtener información sobre cómo obtener estos diccionarios opcionales específicos por culturas.

Input

Tabla 43: Campos de entrada de Name Variant Finder

Nombre de campo	Descripción / Valores válidos
FirstName	El nombre respecto del cual quiere encontrar variantes, si se trata de un nombre dado.
LastName	El nombre respecto del cual quiere encontrar variantes, si se trata de un apellido.

Nombre de campo	Descripción / Valores válidos
GenderCode	<p>El género del nombre que se encuentra en el campo FirstName. Una de las siguientes:</p> <p>Nota: Los códigos de género solo se aplican a los primeros nombres, no a los apellidos.</p> <p>M El nombre es masculino.</p> <p>F El nombre es femenino.</p> <p>A Ambiguo. El nombre puede ser masculino o femenino.</p> <p>U Desconocido. El género correspondiente a este nombre es desconocido.</p>
Ethnicity	<p>La cultura más comúnmente asociada con el nombre del campo FirstName o LastName. Puede utilizar las etapas Name Parser u Open Parser para completar este campo si no conoce el origen étnico de un nombre.</p> <p>Nota: Anteriormente este campo se llamaba GenderDeterminationSource.</p>

Opciones

Tabla 44: Opciones de Name Variant Finder

Opción	Descripción
Nombre	Busca variaciones de nombre según el primer nombre.
Apellido	Busca variaciones de nombre según el apellido.
Código de género	Arroja variaciones de nombre solo para los géneros especificados en el campo GenderCode del registro. Para obtener más información sobre el campo GenderCode, consulte Input en la página 326.
Origen étnico	Arroja variaciones de nombre solo para las culturas especificadas en el campo Origen étnico del registro. Para obtener más información sobre el campo Origen étnico, consulte Input en la página 326.

Opción	Descripción
Romanizado	Arroja la versión romanizada en inglés de un nombre. Un nombre romanizado es aquel que se ha convertido de un tipo de escritura no latina a una latina. Por ejemplo, "Achin" es la versión romanizada del nombre coreano 아진.
Nativo	arroja el nombre en el tipo de escritura nativo de esa cultura. Por ejemplo, un nombre coreano puede devolverse en Hangul.
Kana	Si seleccionó Nativo , puede elegir obtener nombres japoneses en Kana al marcar esta opción. Kana está compuesto por los tipos de escritura hiragana y katakana. Nota: Debe obtener la licencia de la base de datos Paquete Plus para idiomas asiáticos para buscar variantes de nombres japoneses. Para obtener más información, consulte a su ejecutivo de ventas.
Kanji	Si seleccionó Nativo , puede elegir obtener nombres japoneses en Kanji al marcar esta opción. Kanji es uno de los tipos de escritura que se utilizan en el idioma japonés. Nota: Debe obtener la licencia de la base de datos Paquete Plus para idiomas asiáticos para buscar variantes de nombres japoneses. Para obtener más información, consulte a su ejecutivo de ventas.

Salida

Tabla 45: Datos de salida de Name Variant Finder

Nombre de campo	Formato	Descripción / Valores válidos
CandidateGroup	Cadena	Identifica un grupo de nombre de entrada y sus variaciones de nombre. Cada nombre de entrada recibe un número de CandidateGroup. Las variaciones de nombre de entrada obtienen el mismo número de CandidateGroup.
Ethnicity	Cadena	La cultura de un nombre determinado por el Nombre básico y los diccionarios adicionales. Nota: Anteriormente este campo se llamaba GenderDeterminationSource.

Nombre de campo	Formato	Descripción / Valores válidos
FirstName	Cadena	El nombre de pila de una persona.
GenderCode	Cadena	El género de un nombre determinado por el Nombre básico y los diccionarios adicionales. Una de las siguientes: M El nombre es masculino. F El nombre es femenino. A Ambiguo. El nombre puede ser masculino o femenino. U Desconocido. El género correspondiente a este nombre es desconocido.
LastName	Cadena	El apellido de la persona.
TransactionalRecordType	Cadena	Especifica la forma en que el nombre se utilizó en el proceso de comparación. Una de las siguientes: Suspect Se utiliza un registro sospechoso como entrada a una consulta. Candidate Un registro candidato es un resultado devuelto de una consulta.

OpenNameParser

OpenNameParser divide los nombres personales, nombres de empresas y otros términos del campo de datos de nombres en las partes que los conforman. De este modo, estos elementos de nombre analizados se encuentran disponibles para otras operaciones automatizadas tales como comparación y estandarización de nombres o consolidación de nombres de múltiples registros.

OpenNameParser hace lo siguiente:

- Determina el tipo de un nombre para describir la función que cumple ese nombre. Los tipos de entidades de nombre se dividen en dos grupos principales: nombres personales y nombres de empresas. Estos grupos principales también se dividen en subgrupos.
- Determina la forma de un nombre para comprender la sintaxis que debe seguir la herramienta de análisis. Por lo general, los nombres personales siguen un orden natural (firma) o un orden inverso, en tanto los nombres de empresas se ordenan jerárquicamente.
- Determina y etiqueta los componentes de un nombre de forma tal de identificar la relación sintáctica de cada parte del nombre con el nombre completo. La sintaxis de nombres personales incluye

prefijos, las secciones de primer y segundo nombre y apellido, sufijos y términos de descripciones de cuentas, entre otras partes del nombre. La sintaxis de nombres de empresas incluye el nombre de la firma y los términos de sufijos.

- Analiza los nombres conjuntos de empresas y personales, y los retiene como un registro o los divide en varios registros. Ejemplos de nombres conjuntos son "Mr. and Mrs. John Smith" y "Baltimore Gas & Electric dba Constellation Energy".
- Analiza el resultado como registros o como una lista.
- Le permite usar Domain Editor de Open Parser para crear nuevos dominios que se pueden usar en las opciones avanzadas de Open Name Parser.
- Asigna una calificación de análisis que indica el grado de confianza de la exactitud del análisis.

Parámetros d

Tabla 46: Entrada de Open Name Parser

Nombre de campo	Descripción								
CultureCode	<p>La cultura de los datos de nombres de entrada. Las opciones se muestran a continuación.</p> <table border="0"> <tr> <td>Null (empty)</td> <td>Cultura global (predeterminado).</td> </tr> <tr> <td>de</td> <td>Alemán</td> </tr> <tr> <td>es</td> <td>Español</td> </tr> <tr> <td>ja</td> <td>Japonés</td> </tr> </table> <p>Nota: Si ha agregado un dominio propio con el Open Parser Domain Editor, las culturas y códigos culturales para ese dominio también son válidos.</p>	Null (empty)	Cultura global (predeterminado).	de	Alemán	es	Español	ja	Japonés
Null (empty)	Cultura global (predeterminado).								
de	Alemán								
es	Español								
ja	Japonés								
Name	El nombre que desea analizar. Este campo es obligatorio.								

Opciones

Las opciones de Open Name Parser se pueden configurar a nivel de etapa, por medio de cualquiera de los clientes Spectrum™ Technology Platform o en el tiempo de ejecución, mediante las opciones de flujo de datos.

Opciones de análisis

La siguiente tabla muestra las opciones que controlan el análisis de nombres.

Tabla 47: Abrir Name Parser Opciones de análisis

Nombre de la opción	Descripción
Analizar nombres personales	<p>Especifica si se van a analizar nombres de personas.</p> <p>Natural Los campos de nombre están ordenados por tratamiento, nombre, segundo nombre, apellido y sufijo.</p> <p>Revertir Los campos de nombre están ordenados según el apellido en primer lugar.</p> <p>Ambos Los campos de nombre se ordenan utilizando una combinación del criterio Natural y Revertir.</p>
Nombres conjuntos	Especifica si se van a analizar nombres conjuntos.
Dividir nombres conjuntos en múltiples registros	<p>Haga clic en esta casilla para separar los nombres que pertenecen a más de una persona en múltiples registros, por ejemplo, Bill & Sally Smith.</p> <p>Utilizar una etapa Unique ID Generator para crear una ID. para cada uno de los registros divididos.</p>
Analizar nombres de empresas	Especifica si se van a analizar nombres de empresas.
Resultados de salida como lista	Especifica si se van a arrojar los elementos de nombre analizados en un formulario de lista.
Umbral de acceso directo	<p>Especifica cómo mantener un equilibrio en el rendimiento en comparación con la calidad. Una velocidad de rendimiento mayor generará una calidad de los datos de salida menor; del mismo modo, una mayor calidad generará una velocidad de rendimiento menor. Cuando se cumple este umbral, no se realizará ningún otro procesamiento en el registro.</p> <p>El valor predeterminado es 100.</p>

Opciones de culturas

La siguiente tabla muestra las opciones que controlan las culturas de nombre.

Tabla 48: Abra Opciones de culturas de Name Parser

Nombre de la opción	Descripción
Culturas	<p>Especifique qué cultura o culturas desea incluir en la gramática de análisis. Cultura global es la selección predeterminada.</p> <p>Nota: Si se ha agregado un dominio propio usando Open Parser Domain Editor, las culturas y códigos culturales para dicho dominio se mostrarán aquí.</p> <p>Utilice los botones Subir y Bajar para configurar el orden en que desea ejecutar las culturas.</p>

Opciones avanzadas

La siguiente tabla enumera las opciones avanzadas de análisis de nombres.

Tabla 49: Opciones avanzadas de Open Name Parser

Opción	Descripción
Opciones avanzadas	<p>Use el menú desplegable Dominio para seleccionar el dominio apropiado para cada nombre.</p> <p>Utilice los botones Subir y Bajar para configurar el orden en que desea ejecutar los analizadores. Los resultados se devuelven en el primer dominio que obtenga una calificación más alta que el número establecido en el campo Umbral de acceso directo. Si ningún dominio alcanza ese umbral, se devuelven los resultados para el dominio con la puntuación más alta. Si varios dominios alcanzan el umbral al mismo tiempo, se da prioridad al dominio que se ejecuta primero (determinado por el orden establecido aquí) y se devuelven sus resultados.</p> <p>Nota: Si agregó un dominio propio con el Open Parser Domain Editor, este dominio también aparecerá aquí.</p>

Cómo configurar opciones en el tiempo de ejecución

Las opciones de Open Name Parser se pueden configurar y pasar en el tiempo de ejecución, si estas se exponen como opciones del flujo de datos. Esto le permite anular la configuración existente

con cadenas de análisis de nombre con formato JSON. También puede establecer las opciones de tiempo de ejecución cuando se invoque al trabajo, mediante un flujo de proceso o mediante la herramienta de línea de comandos del ejecutor del trabajo.

Para definir las opciones de Open Name Parser en el tiempo de ejecución:

1. En Enterprise Designer, abra un flujo de datos que use una etapa Open Name Parser.
2. Guarde y exponga dicho flujo de datos.
3. Vaya a `Edit > Dataflow Options`.
4. En la tabla **Aplicar opciones de flujo de datos a etapas** expanda Open Name Parser y edite las opciones según sea necesario. Marque la casilla de la opción que desee editar y luego cambie el valor en la lista desplegable **Valor predeterminado**.
5. Opcional: cambie el nombre de las opciones en el campo **Rótulo de opción**.
6. Haga clic en **Aceptar** dos veces.

Output

Tabla 50: Datos de salida de Open Name Parser

Nombre de campo	Formato	Descripción
AccountDescription	Cadena	Descripción de cuenta que es parte del nombre. Por ejemplo, en "Mary Jones Account # 12345", la descripción de cuenta es "Account#12345".
Names	Cadena	Una estructura jerárquica que contiene una lista de elementos analizados. Este campo se devuelve al comprobar la casilla Resultados de salida como lista en las opciones de análisis.

Campos relacionados con nombres de empresas

FirmConjunction	Cadena	Indica que el nombre de una empresa contiene una conjunción como "d/b/a" (nombre con que hace negocios), "o/a" (nombre con que opera), y "t/a" (nombre con que comercializa).
FirmName	Cadena	El nombre de una empresa. Por ejemplo, "Pitney Bowes".
FirmSuffix	Cadena	El sufijo corporativo. Por ejemplo, "SRL" y "SA".

Nombre de campo	Formato	Descripción								
IsFirm	Cadena	Indica que el nombre es una empresa en vez de a una persona.								
Campos relacionados con nombres de personas										
Conjunction	Cadena	Indica que el nombre contiene una conjunción "y", "o" o "&".								
CultureCode	Cadena	Los códigos de cultura contenidos en los datos de entrada.								
CultureCodeUsedToParse	Cadena	<p>Identifica la gramática de cultura específica que se utilizó para analizar los datos.</p> <table border="0"> <tr> <td>Null (empty)</td> <td>Cultura global (predeterminado).</td> </tr> <tr> <td>de</td> <td>Alemán</td> </tr> <tr> <td>es</td> <td>Español</td> </tr> <tr> <td>ja</td> <td>Japonés</td> </tr> </table> <p>Nota: Si agregó un dominio propio con Open Parser Domain Editor, las culturas y los códigos de cultura para ese dominio también aparecerán en el campo.</p>	Null (empty)	Cultura global (predeterminado).	de	Alemán	es	Español	ja	Japonés
Null (empty)	Cultura global (predeterminado).									
de	Alemán									
es	Español									
ja	Japonés									
FirstName	Cadena	El primer nombre de la persona.								
GeneralSuffix	Cadena	El sufijo general o profesional de una persona. Por ejemplo, MD o PhD.								
IsParsed	Cadena	Indica si un registro de salida fue analizado. Los valores son "true" (verdadero) o "false" (falso).								
IsPersonal	Cadena	Indica si el nombre es una persona en lugar de una empresa. Los valores son "true" (verdadero) o "false" (falso).								
IsReverseOrder	Cadena	Indica si el nombre de la entrada está en el orden inverso. Los valores son "true" (verdadero) o "false" (falso).								

Nombre de campo	Formato	Descripción
LastName	Cadena	El apellido de la persona. Incluye el apellido paterno.
LeadingData	Cadena	Información que no es el nombre y que aparece antes que un nombre.
MaturitySuffix	Cadena	Un sufijo generacional de una persona. Por ejemplo, Jr. o Sr.
MiddleName	Cadena	El segundo nombre de una persona.
Name.	Cadena	El nombre personal o de empresa que se proporcionó en la entrada.
NameScore	Cadena	Indica la puntuación media de muestras conocidas y desconocidas para cada nombre. El valor de NameScore estará entre 0 y 100, como se define en la gramática de análisis. Se arroja 0 cuando no se encuentran cruces.
SecondaryLastName	Cadena	Gramática de análisis en español, el apellido de la madre de una persona.
TitleOfRespect	Cadena	Información que aparece antes que un nombre, como, por ejemplo, "Sr.", "Sra." o "Dr."
TrailingData	Cadena	Información que no es el nombre y que aparece después de un nombre.
Campos relacionados con nombres conjuntos		
Conjunction2	Cadena	Indica que un segundo nombre conjunto contiene una conjunción "y", "o" o "&".
Conjunction3	Cadena	Indica que un tercer nombre conjunto contiene una conjunción "y", "o" o "&".

Nombre de campo	Formato	Descripción
FirmName2	Cadena	El nombre de una segunda empresa conjunta. Por ejemplo, Baltimore Gas & Electric dba Constellation Energy.
FirmSuffix2	Cadena	El sufijo de una segunda empresa conjunta.
FirstName2	Cadena	El primer nombre de un segundo nombre conjunto.
FirstName3	Cadena	El primer nombre de un tercer nombre conjunto.
GeneralSuffix2	Cadena	El sufijo general/profesional para un segundo nombre conjunto. Por ejemplo, MD o PhD.
GeneralSuffix3	Cadena	El sufijo general/profesional para un tercer nombre conjunto. Por ejemplo, MD o PhD.
IsConjoined	Cadena	Indica que el nombre de la entrada es conjunto. Un ejemplo de nombre conjunto es "John y Jane Smith".
LastName2	Cadena	El apellido de un segundo nombre conjunto.
LastName3	Cadena	El apellido de un tercer nombre conjunto.
MaturitySuffix2	Cadena	El sufijo de madurez/generacional de un segundo nombre conjunto. Por ejemplo, Jr. o Sr.
MaturitySuffix3	Cadena	El sufijo de madurez/generacional de un tercer nombre conjunto. Por ejemplo, Jr. o Sr.
MiddleName2	Cadena	El segundo nombre de un segundo nombre conjunto.
MiddleName3	Cadena	El segundo nombre de un tercer nombre conjunto.

Nombre de campo	Formato	Descripción
TitleOfRespect2	Cadena	Información que aparece antes de que un segundo nombre conjunto, como "Mr.", "Sra." o "Dr. "
TitleOfRespect3	Cadena	Información que aparece antes de que un tercer nombre conjunto, como "Mr.", "Sra." o "Dr. "

Informe de resumen de Open Name Parser

Un informe de resumen Open Name Parser muestra estadísticas resumidas sobre el trabajo, como el número total de registros de entrada y el número total de registros que no contienen datos de nombre, además de varias estadísticas de análisis sintáctico. Para obtener instrucciones sobre cómo usar los informes, consulte la *Guía de Dataflow Designer de Spectrum™ Technology Platform*.

Resultados generales

- **Número total de registros de entrada:**-el número de registros en el archivo de entrada.
- **Número total de registros que no contienen datos de nombre:** el número de registros para analizar en el archivo de entrada que no contienen datos de nombre.
- **Número total de nombres analizados:** el número de nombres en el archivo de entrada que se analizaron.
- **Total de registros:**-el número total de registros procesados.
- **Calificación más baja del análisis del nombre:** la calificación más baja obtenida por un nombre en el archivo de entrada.
- **Calificación más alta del análisis del nombre:** la calificación más alta obtenida por un nombre en el archivo de entrada.
- **Calificación promedio del análisis del nombre:** la calificación promedio obtenida por los nombres analizados en el archivo de entrada.

Resultados del análisis de los nombres personales

- **Número de registros de nombres personales escritos:** el número de nombres personales en el archivo de entrada.
- **Número de nombres conjuntos analizados:** el número de nombres de los registros analizados que contienen nombres conjuntos. Por ejemplo, si el archivo de entrada tenía cinco registros con dos nombres conjuntos y siete registros con tres nombres conjuntos, el valor de este campo sería 31, como se expresa en la siguiente ecuación: $(5 \times 2) + (7 \times 3)$.
- **Registros con 2 nombres conjuntos:** el número de registros de entrada que contienen dos nombres conjuntos.
- **Registros con 3 nombres conjuntos:** el número de registros de entrada que contienen tres nombres conjuntos.

- **Número de nombres con título de tratamiento:** el número de nombres analizados que contienen un título de tratamiento.
- **Número de nombres con sufijo generacional:** el número de nombres analizados que contienen un sufijo generacional.
- **Número de nombres con sufijo general:** el número de nombres analizados que contienen un sufijo general.
- **Número de nombres que contienen descripciones de cuenta:** el número de nombres analizados que contienen una descripción de cuenta.
- **Total de nombres con orden inverso:** el número de nombres analizados que siguen un orden inverso, lo que se traduce en el campo de salida isReversed como "verdadero".

Resultados del análisis de los nombres de empresas

- **Número de registros de nombres de empresas escritos:** el número de nombres de empresas en el archivo de entrada.
- **Número de nombres con sufijo de empresa:** el número de nombres analizados que contienen un sufijo de empresarial.
- **Número de nombres que contienen descripciones de cuenta:** el número de nombres en registros de entrada que contienen una descripción de cuenta.
- **Total de registros DBA:** el número de registros de entrada que contienen conjunciones Doing Business As (DBA), lo que se traduce en campos de salida isFirm y isPersonal como "verdadero".

9 - Códigos de país ISO y compatibilidad de módulos

In this section

Códigos de país ISO y compatibilidad de módulos

340

Códigos de país ISO y compatibilidad de módulos

La siguiente tabla contiene los códigos ISO de cada país y los módulos compatibles con las funciones de procesamiento de direcciones, geocodificación y rutas según el país.

tenga en cuenta que el módulo Enterprise Geocoding incluye bases de datos para África (30 países), Medio Oriente (8 países) y Latinoamérica (20 países). Estas bases de datos abarcan los países más pequeños en dichas regiones, que no poseen sus propias bases de datos de geocodificación específica del país. La columna Módulos compatibles indica qué países abarcan estas bases de datos de África, Medio Oriente y Latinoamérica.

Asimismo, la base de datos Geocode Address World proporciona geocodificación postal geográfica y limitada para todos los países (pero no geocodificación de nivel de calle).

Nombre de país ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Módulos compatibles
Afganistán	AF	AFG	Módulo Universal Addressing
Islas Aland	AX	ALA	Módulo Universal Addressing
Albania	AL o SQ (enrutamiento)	ALB	Módulo Universal Addressing Módulo Enterprise Geocoding Módulo Enterprise Routing
Argelia	DZ	DZA	Módulo Enterprise Geocoding (África) Módulo Universal Addressing
Samoa Americana	AS	ASM	Módulo Universal Addressing
Andorra	AD	AND	Módulo Enterprise Geocoding. (Andorra está cubierta por el código geográfico de España) Módulo Universal Addressing

Nombre de país ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Módulos compatibles
Angola	AO	AGO	Módulo Enterprise Geocoding (África) Módulo Universal Addressing
Anguila	AI	AIA	Módulo Universal Addressing
Antártida	AQ	ATA	Módulo Universal Addressing
Antigua y Barbuda	AG	ATG	Módulo Universal Addressing
Argentina	AR	ARG	Módulo Enterprise Geocoding Módulo Universal Addressing Módulo Enterprise Routing
Armenia	AM	ARM	Módulo Universal Addressing
Aruba	AW	ABW	Módulo Enterprise Geocoding (Latinoamérica) Módulo Universal Addressing
Australia	AU	AUS	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Austria	AT	AUT	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Azerbaiyán	AZ	AZE	Módulo Universal Addressing

Nombre de país ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Módulos compatibles
Bahamas	BS	BHS	Módulo Enterprise Geocoding Módulo Universal Addressing Módulo Enterprise Routing
Bahréin	BH	BHR	Módulo Enterprise Geocoding (Medio Oriente) Módulo Universal Addressing
Bangladesh	BD	BGD	Módulo Universal Addressing
Barbados	BB	BRB	Módulo Enterprise Geocoding (Latinoamérica) Módulo Universal Addressing
Bielorrusia	BY	BLR	Módulo Universal Addressing Módulo Enterprise Routing
Bélgica	BE	BEL	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Belice	BZ	BLZ	Módulo Enterprise Geocoding (Latinoamérica) Módulo Universal Addressing
Benín	BJ	BEN	Módulo Enterprise Geocoding (África) Módulo Universal Addressing
Bermudas	BM	BMU	Módulo Universal Addressing Módulo Enterprise Routing
Bután	BT	BTN	Módulo Universal Addressing

Nombre de país ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Módulos compatibles
Bolivia	BO	BOL	Módulo Enterprise Geocoding (Latinoamérica) Módulo Universal Addressing
Bonaire, Saba y San Eustaquio	BQ	BES	Módulo Universal Addressing
Bosnia-Herzegovina	BA	BIH	Módulo Enterprise Geocoding Módulo Universal Addressing Módulo Enterprise Routing Módulo Enterprise Geocoding
Botsuana	BW	BWA	Módulo Enterprise Geocoding (África) Módulo Universal Addressing
Isla Bouvet	BV	BVT	Módulo Universal Addressing
Brasil	BR	BRA	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Territorio Oceánico Indio-Británico	IO	IOT	Módulo Universal Addressing
Brunéi	BN	BRN	Módulo Enterprise Geocoding Módulo Universal Addressing
Bulgaria	BG	BGR	Módulo Enterprise Geocoding Módulo Universal Addressing
Burkina Faso	BF	BFA	Módulo Enterprise Geocoding (África) Módulo Universal Addressing

Nombre de país ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Módulos compatibles
Burundi	BI	BDI	Módulo Enterprise Geocoding (África) Módulo Universal Addressing
Camboya	KH	KHM	Módulo Universal Addressing
Camerún	CM	CMR	Módulo Enterprise Geocoding (África) Módulo Universal Addressing
Canadá	CA	CAN	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Cabo Verde	CV	CPV	Módulo Universal Addressing
Islas Caimán	KY	CYM	Módulo Universal Addressing
República Centroafricana	CF	CAF	Módulo Universal Addressing
Chad	TD	TCD	Módulo Universal Addressing
Chile	CL	CHL	Módulo Enterprise Geocoding Módulo Universal Addressing Módulo Enterprise Routing
China	CN o zh_CN (enrutamiento)	CHN	Módulo Enterprise Geocoding Módulo Universal Addressing Módulo Enterprise Routing
Isla de Navidad	CX	CXR	Módulo Universal Addressing

Nombre de país ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Módulos compatibles
Islas Cocos (Keeling)	CC	CCK	Módulo Universal Addressing
Colombia	CO	COL	Módulo Enterprise Geocoding Módulo Universal Addressing
Comoras	KM	WITH	Módulo Universal Addressing
Congo, República del	CG	COG	Módulo Enterprise Geocoding (África) Módulo Universal Addressing
Congo, República Democrática del	CD	COD	Módulo Enterprise Geocoding (África) Módulo Universal Addressing Módulo Enterprise Routing
Islas Cook	CK	COK	Módulo Universal Addressing
Costa Rica	CR	CRI	Módulo Enterprise Geocoding (Latinoamérica) Módulo Universal Addressing
Costa de Marfil	CI	CIV	Módulo Universal Addressing
Croacia	HR	HRV	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Cuba	CU	CUB	Módulo Enterprise Geocoding (Latinoamérica) Módulo Enterprise Routing Módulo Universal Addressing
Curaçao	CW	CUW	Módulo Universal Addressing

Nombre de país ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Módulos compatibles
Chipre	CY	CYP	Módulo Enterprise Geocoding Módulo Universal Addressing
República Checa	CZ o CS (enrutamiento)	CZE	Módulo Enterprise Geocoding Módulo Universal Addressing Módulo Enterprise Routing
Dinamarca	DK	DNK	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Yibuti	DJ	DJI	Módulo Universal Addressing
Dominica	DM	DMA	Módulo Universal Addressing
República Dominicana	DO	DOM	Módulo Enterprise Geocoding (Latinoamérica) Módulo Universal Addressing
Ecuador	EC	ECU	Módulo Enterprise Geocoding (Latinoamérica) Módulo Universal Addressing
Egipto	EG	EGY	Módulo Enterprise Geocoding (Medio Oriente) Módulo Universal Addressing
El Salvador	SV	SLV	Módulo Enterprise Geocoding (Latinoamérica) Módulo Universal Addressing
Guinea Ecuatorial	GQ	GNQ	Módulo Universal Addressing
Eritrea	ER	ERI	Módulo Universal Addressing

Nombre de país ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Módulos compatibles
Estonia	EE	EST	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Etiopía	ET	ETH	Módulo Universal Addressing
Islas Malvinas (Falkland)	FK	FLK	Módulo Universal Addressing
Islas Feroe	FO	FRO	Módulo Universal Addressing
Islas Fiji	FJ	FJI	Módulo Universal Addressing
Finlandia	FI	FIN	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Francia	FR	FRA	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Guayana Francesa	GF	GUF	Módulo Enterprise Geocoding (<i>Guayana Francesa está cubierta por el geocodificador de Francia</i>). Módulo Universal Addressing
Polinesia Francesa	PF	PYF	Módulo Universal Addressing
Territorios Australes Franceses	TF	ATF	Módulo Universal Addressing
Gabón	GA	GAB	Módulo Enterprise Geocoding (África) Módulo Universal Addressing

Nombre de país ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Módulos compatibles
Gambia	GM	GMB	Módulo Universal Addressing
Georgia	GE	GEO	Módulo Universal Addressing
Alemania	OF	DEU	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Ghana	GH	GHA	Módulo Enterprise Geocoding (África) Módulo Universal Addressing Módulo Enterprise Routing
Gibraltar	GI	GIB	Módulo Enterprise Geocoding (<i>Gibraltar está cubierto por el geocodificador de España</i>). Módulo Universal Addressing
Grecia	GR	GRC	Módulo Enterprise Geocoding Módulo Universal Addressing
Groenlandia	GL	GRL	Módulo Universal Addressing
Granada	GD	GRD	Módulo Universal Addressing
Guadalupe	GP	GLP	Módulo Enterprise Geocoding (<i>Guadalupe está cubierta por el geocodificador de Francia</i>). Módulo Universal Addressing
Guam	GU	GUM	Módulo Universal Addressing

Nombre de país ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Módulos compatibles
Guatemala	GT	GTM	Módulo Enterprise Geocoding (Latinoamérica) Módulo Universal Addressing
Guernsey	GG	GGY	Módulo Universal Addressing
Guinea	GN	GIN	Módulo Universal Addressing
Guinea-Bissau	GW	GNB	Módulo Universal Addressing
Guyana	GY	GUY	Módulo Enterprise Geocoding (Latinoamérica) Módulo Universal Addressing
Haití	HT	HTI	Módulo Universal Addressing
Islas Heard y McDonald	HM	HMD	Módulo Universal Addressing
Santa Sede (Estado de la Ciudad del Vaticano)	VA	VAT	Módulo Enterprise Geocoding (<i>El Vaticano está cubierto por el geocodificador de Italia</i>). Módulo Universal Addressing
Honduras	HN	HND	Módulo Enterprise Geocoding (Latinoamérica) Módulo Universal Addressing
Hong Kong	HK	HKG	Módulo Enterprise Geocoding Módulo Universal Addressing
Hungría	HU	HUN	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing

Nombre de país ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Módulos compatibles
Islandia	IS	ISL	Módulo Enterprise Geocoding Módulo Universal Addressing
India	IN	IND	Módulo Enterprise Geocoding Módulo Universal Addressing
Indonesia	ID	IDN	Módulo Enterprise Geocoding Módulo Universal Addressing
República Islámica de Irán	IR	IRN	Módulo Universal Addressing
Irak	IQ	IRQ	Módulo Enterprise Geocoding (Medio Oriente) Módulo Universal Addressing
Irlanda	IE	IRL	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Isla de Man	IM	IMN	Módulo Universal Addressing
Israel	IL	ISR	Módulo Enterprise Geocoding Módulo Universal Addressing Módulo Enterprise Routing
Italia	IT	ITA	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Jamaica	JM	JAM	Módulo Enterprise Geocoding (Latinoamérica) Módulo Universal Addressing

Nombre de país ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Módulos compatibles
Japón	JP	JPN	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Jersey	JE	JEY	Módulo Universal Addressing
Jordania	JO	JOR	Módulo Universal Addressing Módulo Enterprise Geocoding (Medio Oriente) Módulo Enterprise Routing
Kazajistán	KZ	KAZ	Módulo Universal Addressing
Kenia	KE	KEN	Módulo Enterprise Geocoding (África) Módulo Universal Addressing Módulo Enterprise Routing
Kiribati	KI	KIR	Módulo Universal Addressing
Corea, República Popular Democrática de	KP	PRK	Módulo Universal Addressing
Corea, República de	KR	KOR	Módulo Enterprise Geocoding Módulo Universal Addressing
Kosovo	Xk	XKX	Módulo Enterprise Geocoding Módulo Universal Addressing
Kuwait	KW	KWT	Módulo Enterprise Geocoding (Medio Oriente) Módulo Universal Addressing

Nombre de país ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Módulos compatibles
Kirguistán	KG	KGZ	Módulo Universal Addressing
República Democrática Popular Lao	LA	LAO	Módulo Universal Addressing
Letonia	LV	LVA	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Líbano	LB	LBN	Módulo Enterprise Geocoding (Medio Oriente) Módulo Universal Addressing
Lesoto	LS	LSO	Módulo Enterprise Geocoding (África) Módulo Universal Addressing Módulo Enterprise Routing
Liberia	LR	LBR	Módulo Universal Addressing
Libia	LY	LBY	Módulo Universal Addressing
Liechtenstein	LI	LIE	Módulo Enterprise Geocoding (<i>Liechtenstein está cubierto por el geocodificador de Suiza</i>). Módulo Enterprise Routing Módulo Universal Addressing
Lituania	LT	LTU	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing

Nombre de país ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Módulos compatibles
Luxemburgo	LU	LUX	Módulo Enterprise Geocoding (<i>Luxemburgo está cubierto por el geocodificador de Bélgica</i>). Módulo Enterprise Routing Módulo Universal Addressing
Macao	MO	MAC	Módulo Enterprise Geocoding Módulo Universal Addressing
Macedonia, Antigua República Yugoslava de	MK	MKD	Módulo Enterprise Geocoding Módulo Universal Addressing
Madagascar	MG	MDG	Módulo Universal Addressing
Malawi	MW	MWI	Módulo Enterprise Geocoding (África) Módulo Universal Addressing
Malasia	MY	MYS	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Maldivas	MV	MDV	Módulo Universal Addressing
Malí	ML	MLI	Módulo Enterprise Geocoding (África) Módulo Universal Addressing
Malta	ML	MLT	Módulo Enterprise Geocoding Módulo Universal Addressing
Islas Marshall	MH	MHL	Módulo Universal Addressing

Nombre de país ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Módulos compatibles
Martinica	MQ	MTQ	Módulo Enterprise Geocoding (<i>Martinica está cubierta por el geocodificador de Francia</i>). Módulo Universal Addressing
Mauritania	MR	MRT	Módulo Enterprise Geocoding (África) Módulo Universal Addressing
Mauricio	MU	MUS	Módulo Enterprise Geocoding (África) Módulo Universal Addressing
Mayotte	YT	MYT	Módulo Enterprise Geocoding (<i>Mayotte está cubierta por el geocodificador de Francia</i>). Módulo Universal Addressing
México	MX	MEX	Módulo Enterprise Geocoding Módulo Universal Addressing
Micronesia, Estados Federados de	FM	FSM	Módulo Universal Addressing
Moldavia, República de	MD	MDA	Módulo Universal Addressing Módulo Enterprise Routing
Mónaco	MC	MCO	Módulo Enterprise Geocoding (<i>Mónaco está cubierta por el geocodificador de Francia</i>). Módulo Universal Addressing
Mongolia	MN	MNG	Módulo Universal Addressing
Montenegro	ME	MNE	Módulo Enterprise Geocoding Módulo Universal Addressing

Nombre de país ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Módulos compatibles
Montserrat	MS	MSR	Módulo Universal Addressing
Marruecos	MA	MAR	Módulo Enterprise Geocoding (África) Módulo Universal Addressing
Mozambique	MZ	MOZ	Módulo Enterprise Geocoding (África) Módulo Universal Addressing Módulo Enterprise Routing
Myanmar	MM	MMR	Módulo Universal Addressing
Namibia	NA	NAM	Módulo Enterprise Geocoding (África) Módulo Universal Addressing
Nauru	NR	NRU	Módulo Universal Addressing
Nepal	NP	NPL	Módulo Universal Addressing
Países Bajos	NL	NLD	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Nueva Caledonia	NC	NCL	Módulo Universal Addressing
Nueva Zelanda	NZ	NZL	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Nicaragua	NI	NIC	Módulo Enterprise Geocoding (Latinoamérica) Módulo Universal Addressing

Nombre de país ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Módulos compatibles
Níger	NE	NER	Módulo Enterprise Geocoding (África) Módulo Universal Addressing
Nigeria	NG	NGA	Módulo Enterprise Geocoding (África) Módulo Universal Addressing
Niue	NU	NIU	Módulo Universal Addressing
Isla Norfolk	NF	NFK	Módulo Universal Addressing
Islas Marianas del Norte	MP	MNP	Módulo Universal Addressing
Noruega	NO	NOR	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Omán	OM	OMN	Módulo Enterprise Geocoding (Medio Oriente) Módulo Universal Addressing
Pakistán	PK	PAK	Módulo Universal Addressing
Palaos	PW	PLW	Módulo Universal Addressing
Territorio Palestino, Ocupado	PS	PSE	Módulo Universal Addressing
Panamá	PA	OFFSET	Módulo Enterprise Geocoding (Latinoamérica) Módulo Universal Addressing

Nombre de país ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Módulos compatibles
Papúa Nueva Guinea	PG	PNG	Módulo Universal Addressing
Paraguay	PY	PRY	Módulo Enterprise Geocoding (Latinoamérica) Módulo Universal Addressing
Perú	PE	PER	Módulo Enterprise Geocoding (Latinoamérica) Módulo Universal Addressing
Las Filipinas	PH	PHL	Módulo Enterprise Geocoding Módulo Universal Addressing Módulo Enterprise Routing
Pitcairn	PN	PCN	Módulo Universal Addressing
Polonia	PL	POL	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Portugal	PT	PRT	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Puerto Rico	PR	PRI	Módulo Universal Addressing
Qatar	QA	QAT	Módulo Enterprise Geocoding (Medio Oriente) Módulo Universal Addressing
Reunión	RE	REU	Módulo Enterprise Geocoding (<i>Reunión está cubierta por el geocodificador de Francia</i>). Módulo Universal Addressing

Nombre de país ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Módulos compatibles
Rumania	RO	ROU	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Federación Rusa	RU	RUS	Módulo Enterprise Geocoding Módulo Universal Addressing
Ruanda	RW	RWA	Módulo Enterprise Geocoding (África) Módulo Universal Addressing
Saint Barthélemy	BL	BLM	Módulo Universal Addressing
Saint Helena, Ascension y Tristan Da Cunha	SH	SHE	Módulo Universal Addressing
San Cristóbal y Nieves	KN	KNA	Módulo Enterprise Geocoding (Latinoamérica) Módulo Universal Addressing
Santa Lucía	LC	LCA	Módulo Universal Addressing
Saint Martin (parte francesa)	MF	MAF	Módulo Universal Addressing
San Pedro y Miquelón	PM	SPM	Módulo Universal Addressing
San Vicente y las Granadinas	VC	VCT	Módulo Universal Addressing
Samoa	WS	WSM	Módulo Universal Addressing

Nombre de país ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Módulos compatibles
San Marino	SM	SMR	Módulo Enterprise Geocoding (<i>San Marino está cubierto por el geocodificador de Italia</i>). Módulo Universal Addressing
Santo Tomé y Príncipe	ST	STP	Módulo Universal Addressing
Arabia Saudita	SA	SAU	Módulo Enterprise Geocoding (Medio Oriente) Módulo Universal Addressing
Senegal	SN	SEN	Módulo Enterprise Geocoding (África) Módulo Universal Addressing
Serbia	RS	SRB	Módulo Enterprise Geocoding Módulo Universal Addressing
Seychelles	SC	SYC	Módulo Universal Addressing
Sierra Leona	SL	SLE	Módulo Universal Addressing
República de Singapur	SG	SGP	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Sint Maarten (parte holandesa)	SX	SXM	Módulo Universal Addressing
Eslovaquia	SK	SVK	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing

Nombre de país ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Módulos compatibles
Eslovenia	SI	SVN	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Islas Salomón	SB	SLB	Módulo Universal Addressing
Somalia	SO	SOM	Módulo Universal Addressing
Sudáfrica	ZA	ZAF	Módulo Enterprise Geocoding Módulo Universal Addressing
Islas Georgias del Sur y Sandwich del Sur	GS	SGS	Módulo Enterprise Geocoding Módulo Universal Addressing
Sudán del Sur	SS	SSD	Módulo Universal Addressing
España	ES	ESP	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Sri Lanka	LK	LKA	Módulo Universal Addressing
Sudán	SD	SDN	Módulo Universal Addressing
Surinam	SR	SUR	Módulo Enterprise Geocoding (Latinoamérica) Módulo Universal Addressing
Svalbard y Jan Mayen	SJ	SJM	Módulo Universal Addressing

Nombre de país ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Módulos compatibles
Suazilandia	SZ	SWZ	Módulo Enterprise Geocoding (África) Módulo Universal Addressing
Suecia	IF	SWE	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Suiza	CH	CHE	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
República Árabe Siria	SY	SYR	Módulo Universal Addressing
Taiwán, Provincia de China	TW o zh_TW (enrutamiento)	TWN	Módulo Universal Addressing Módulo Enterprise Routing
Tayikistán	TJ	TJK	Módulo Universal Addressing
Tanzania	TZ	TZA	Módulo Enterprise Geocoding (África) Módulo Universal Addressing Módulo Enterprise Routing
Tailandia	TH	THA	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Timor Oriental	TL	TLS	Módulo Universal Addressing
Togo	TG	TGO	Módulo Enterprise Geocoding (África) Módulo Universal Addressing

Nombre de país ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Módulos compatibles
Tokelau	TK	TKL	Módulo Universal Addressing
Tonga	TO	TON	Módulo Universal Addressing
Trinidad y Tobago	TT	TTO	Módulo Enterprise Geocoding (Latinoamérica) Módulo Universal Addressing
Túnez	TN	TUN	Módulo Enterprise Geocoding (África) Módulo Universal Addressing
Turquía	TR	TUR	Módulo Enterprise Geocoding Módulo Universal Addressing
Turkmenistán	TM	TKM	Módulo Universal Addressing
Islas Turcas y Caicos	TC	TCA	Módulo Universal Addressing
Tuvalu	TV	TUV	Módulo Universal Addressing
Uganda	UG	UGA	Módulo Enterprise Geocoding (África) Módulo Universal Addressing
Ucrania	UA	UKR	Módulo Enterprise Geocoding Módulo Universal Addressing
Emiratos Árabes Unidos	AE	ARE	Módulo Enterprise Geocoding (Medio Oriente) Módulo Universal Addressing

Nombre de país ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Módulos compatibles
Reino Unido	GB	GBR	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Estados Unidos	Estados Unidos	Estados Unidos	Módulo Enterprise Geocoding Módulo Enterprise Routing Módulo Universal Addressing
Islas Ultramarinas Menores de los Estados Unidos	UM	UMI	Módulo Universal Addressing
Uruguay	UY	URY	Módulo Enterprise Geocoding Módulo Universal Addressing
Uzbekistán	UZ	UZB	Módulo Universal Addressing
Vanuatu	VU	VUT	Módulo Universal Addressing
Venezuela, República Bolivariana de	VE	VEN	Módulo Enterprise Geocoding Módulo Universal Addressing
Vietnam	VN	VNM	Módulo Enterprise Geocoding Módulo Universal Addressing
Islas Vírgenes Británicas	VG	VGB	Módulo Universal Addressing
Islas Vírgenes de los Estados Unidos	VI	VIR	Módulo Universal Addressing
Wallis y Futuna	WF	WLF	Módulo Universal Addressing

Nombre de país ISO	ISO 3116-1 Alpha-2	ISO 3116-1 Alpha-3	Módulos compatibles
Sahara Occidental	EH	ESH	Módulo Universal Addressing
Yemen	YE	YEM	Módulo Enterprise Geocoding (Medio Oriente) Módulo Universal Addressing
Zambia	ZM	ZMB	Módulo Enterprise Geocoding (África) Módulo Universal Addressing
Zimbabwe	ZW	ZWE	Módulo Enterprise Geocoding (África) Módulo Universal Addressing

Notices

© 2018 Pitney Bowes. Todos los derechos reservados. MapInfo y Group 1 Software son marcas comerciales de Pitney Bowes Software Inc. El resto de marcas comerciales son propiedad de sus respectivos propietarios.

Avisos de USPS®

Pitney Bowes Inc. posee una licencia no exclusiva para publicar y vender bases de datos ZIP + 4® en medios magnéticos y ópticos. Las siguientes marcas comerciales son propiedad del Servicio Postal de los Estados Unidos: CASS, CASS Certified, DPV, eLOT, FASTforward, First-Class Mail, Intelligent Mail, LACS^{Link}, NCOA^{Link}, PAVE, PLANET Code, Postal Service, POSTNET, Post Office, RDI, Suite^{Link}, United States Postal Service, Standard Mail, United States Post Office, USPS, ZIP Code, y ZIP + 4. Esta lista no es exhaustiva de todas las marcas comerciales que pertenecen al servicio postal.

Pitney Bowes Inc. es titular de una licencia no exclusiva de USPS® para el procesamiento NCOA^{Link}®.

Los precios de los productos, las opciones y los servicios del software de Pitney Bowes no los establece, controla ni aprueba USPS® o el gobierno de Estados Unidos. Al utilizar los datos RDI™ para determinar los costos del envío de paquetes, la decisión comercial sobre qué empresa de entrega de paquetes se va a usar, no la toma USPS® ni el gobierno de Estados Unidos.

Proveedor de datos y avisos relacionados

Los productos de datos que se incluyen en este medio y que se usan en las aplicaciones del software de Pitney Bowes Software, están protegidas mediante distintas marcas comerciales, además de un o más de los siguientes derechos de autor:

© Derechos de autor, Servicio Postal de los Estados Unidos. Todos los derechos reservados.

© 2014 TomTom. Todos los derechos reservados. TomTom y el logotipo de TomTom son marcas comerciales registradas de TomTom N.V.

© 2016 HERE

Fuente: INEGI (Instituto Nacional de Estadística y Geografía)

Basado en los datos electrónicos de © National Land Survey Sweden.

© Derechos de autor Oficina del Censo de los Estados Unidos

© Derechos de autor Nova Marketing Group, Inc.

Algunas partes de este programa tienen © Derechos de autor 1993-2007 de Nova Marketing Group Inc. Todos los derechos reservados

© Copyright Second Decimal, LLC

© Derechos de autor Servicio de correo de Canadá

Este CD-ROM contiene datos de una compilación cuyos derechos de autor son propiedad del servicio de correo de Canadá.

© 2007 Claritas, Inc.

El conjunto de datos Geocode Address World contiene datos con licencia de GeoNames Project (www.geonames.org) suministrados en virtud de la licencia de atribución de Creative Commons (la “Licencia de atribución”) que se encuentra en <http://creativecommons.org/licenses/by/3.0/legalcode>. El uso de los datos de GeoNames (según se describe en el manual de usuario de Spectrum™ Technology Platform) se rige por los términos de la Licencia de atribución. Todo conflicto entre el acuerdo establecido con Pitney Bowes Software, Inc. y la Licencia de atribución se resolverá a favor de la Licencia de atribución exclusivamente en cuanto a lo relacionado con el uso de los datos de GeoNames.



3001 Summer Street
Stamford CT 06926-0700
USA

www.pitneybowes.com