

IBM InfoSphere QualityStage  
Versión 11 Release 3

*Guía de aprendizaje*





IBM InfoSphere QualityStage  
Versión 11 Release 3

*Guía de aprendizaje*



**Nota**

Antes de utilizar esta información y el producto al que da soporte, lea la información del apartado "Avisos y marcas registradas" en la página 51.

---

# Contenido

## Guía de aprendizaje de trabajos paralelos de InfoSphere QualityStage . . . 1

Acerca de IBM InfoSphere QualityStage . . . . .	1
Proyectos en IBM InfoSphere QualityStage . . . . .	1
Acerca de los trabajos de InfoSphere QualityStage . . . . .	2
Etapas de IBM InfoSphere DataStage and QualityStage . . . . .	2
Componentes de servidor y cliente . . . . .	3
Objetivos del proyecto de la guía de aprendizaje . . . . .	3
Configurar la guía de aprendizaje . . . . .	4
Crear una carpeta para los archivos de la guía de aprendizaje . . . . .	4
Crear el proyecto de guía de aprendizaje . . . . .	5
Copiar datos de la guía de aprendizaje . . . . .	5
Iniciar un proyecto . . . . .	6
Módulo 1: Investigar los de datos de origen . . . . .	8
Lección 1.1: Configurar y enlazar un trabajo Investigate . . . . .	8
Lección 1.2: Renombrar enlaces y etapas en un trabajo Investigate . . . . .	10
Lección 1.3: Configurar el archivo de origen . . . . .	11
Lección 1.4: Configurar la etapa Copy . . . . .	12
Lección 1.5: Etapa Investigate, configurar para revisar nombres . . . . .	13
Lección 1.6: Etapa Investigate, configurar para revisar regiones geográficas . . . . .	14
Lección 1.7: Configurar informes de destino . . . . .	16
Lección 1.8: Compilar y ejecutar trabajos . . . . .	17
Módulo 1: Resumen . . . . .	18
Módulo 2: Estandarizar datos . . . . .	19
Lección 2.1: Configurar un trabajo Standardize . . . . .	19
Lección 2.2: Configurar las propiedades de etapa del trabajo Standardize . . . . .	21

Lección 2.3: Configurar los conjuntos de datos de destino . . . . .	28
Módulo 2: Resumen . . . . .	29
Módulo 3: Agrupar registros con atributos comunes . . . . .	31
Lección 3.1: Configurar un trabajo One-source Match . . . . .	31
Lección 3.2: Configurar las propiedades de etapa del trabajo One-source Match . . . . .	33
Lección 3.3: Configurar los archivos de destino del trabajo One-source Match . . . . .	36
Módulo 3: Resumen . . . . .	38
Módulo 4: Crear un registro único. . . . .	39
Lección 4.1: Configurar un trabajo Survive . . . . .	39
Lección 4.2: Configurar las propiedades de etapa del trabajo Survive . . . . .	40
Módulo 4: Resumen . . . . .	43
Guía de aprendizaje de IBM InfoSphere QualityStage: resumen. . . . .	44

## Apéndice A. Accesibilidad de los productos . . . . . 45

## Apéndice B. Cómo ponerse en contacto con IBM . . . . . 47

## Apéndice C. Acceso a la documentación del producto . . . . . 49

## Avisos y marcas registradas . . . . . 51

## Índice . . . . . 57



---

# Guía de aprendizaje de trabajos paralelos de InfoSphere QualityStage

Utilice la guía de aprendizaje de trabajos paralelos para conocer las habilidades básicas necesarias para desarrollar este tipo de trabajos.

Los “Objetivos del proyecto de la guía de aprendizaje” en la página 3 son: utilizar las etapas del cliente del Diseñador para limpiar datos de clientes, eliminar todas las direcciones de clientes duplicadas y facilitar el mejor formato para la dirección correcta.

---

## Acerca de IBM InfoSphere QualityStage

IBM® InfoSphere QualityStage es un componente de limpieza de datos que forma parte del cliente de IBM InfoSphere DataStage and QualityStage Designer.

El cliente del Diseñador proporciona una interfaz de usuario común en la que puede diseñar sus trabajos de calidad de datos. Además, cuenta con la potencia de un motor de procesamiento en paralelo para procesar grandes almacenes de datos de origen.

Las etapas integradas disponibles en el Repositorio proporcionan la base para llevar a cabo los siguientes conceptos de limpieza de datos:

- Resolver ambigüedades y conflictos de datos
- Descubrir atributos ocultos o nuevos de columnas de origen con formato libre o poco controladas
- Ajustar los datos transformando los tipos de datos en un formato estándar
- Crear un único resultado

### Objetivos de aprendizaje

Los puntos clave que debe tener en cuenta al completar esta guía de aprendizaje incluyen los siguientes temas:

- Cómo mejoran la calidad de los datos los procesos de estandarización y coincidencia
- La facilidad de combinar las dos etapas de cliente de InfoSphere DataStage and QualityStage Designer en un mismo trabajo
- Cómo fluyen los datos en un proceso interactivo entre dos trabajos
- Los datos supervivientes tienen como resultado el mejor registro disponible

## Proyectos en IBM InfoSphere QualityStage

El cliente de IBM InfoSphere DataStage and QualityStage Designer proporciona una vista para los proyectos. Los proyectos son un método para organizar los datos remodelados. Usted puede definir etapas y archivos de datos, y crear trabajos dentro de un proyecto concreto. IBM InfoSphere QualityStage utiliza estos proyectos para crear y almacenar archivos en el cliente y el servidor.

Cada proyecto de InfoSphere QualityStage contiene los siguientes componentes:

- Trabajos de InfoSphere QualityStage

- Etapas utilizadas para crear cada trabajo
- Especificaciones de coincidencia
- Reglas de estandarización
- Definiciones de tabla

En esta guía de aprendizaje, creará un proyecto y utilizará los datos facilitados para crear trabajos en el proyecto.

## Acerca de los trabajos de InfoSphere QualityStage

IBM InfoSphere QualityStage utiliza trabajos para procesar datos.

Para iniciar un trabajo de InfoSphere QualityStage, abra el cliente del Diseñador y cree un nuevo trabajo paralelo. Puede crear un trabajo de InfoSphere QualityStage añadiendo etapas, archivos de origen y de destino, y enlaces del Repositorio, y colocándolos en el lienzo del Diseñador. El cliente del Diseñador compila el trabajo paralelo y crea un archivo ejecutable. Al ejecutarse el trabajo, las etapas procesan los datos a través de las propiedades de datos que usted ha definido. El resultado es un conjunto de datos que se puede utilizar como entrada para el trabajo siguiente.

En esta guía de aprendizaje puede crear cuatro trabajos de InfoSphere QualityStage. Cada trabajo se crea alrededor de una de las etapas de Data Quality y de etapas adicionales de IBM InfoSphere DataStage.

## Etapas de IBM InfoSphere DataStage and QualityStage

Una etapa de IBM InfoSphere DataStage and QualityStage lleva a cabo una acción sobre los datos. El tipo de acción depende de la etapa en la que se encuentre.

Las etapas del cliente de InfoSphere DataStage and QualityStage Designer se almacenan en la paleta de herramientas del Diseñador. Puede acceder a todas las etapas de IBM InfoSphere QualityStage del grupo de calidad de datos en la paleta. Usted mismo configura cada etapa para realizar en los datos el tipo de acciones que obtienen los resultados necesarios. Esos resultados se utilizan más tarde como datos de entrada en la etapasiguiente. Las siguientes etapas están incluidas en InfoSphere QualityStage:

- etapa Investigate
- etapa Standardize
- etapa Match Frequency
- etapa One-source Match
- etapa Two-source Match
- etapa Survive
- etapa Standardization Quality Assessment (SQA)

En esta guía de aprendizaje utilizará la mayor parte de etapas de InfoSphere QualityStage.

Además puede añadir a su trabajo cualquier etapa de IBM InfoSphere DataStage. En algunas de las lecciones puede añadir etapas de InfoSphere DataStage para ampliar los tipos de herramientas para procesar los datos.



## Componentes de servidor y cliente

Debe cargar los componentes de cliente y servidor que se utilizan para crear trabajos para limpiar datos.

Los siguientes componentes de servidor se instalan en el servidor:

### Repositorio

Un almacén central que contiene toda la información necesaria para crear un trabajo de IBM InfoSphere QualityStage.

### Motor de InfoSphere Information Server

Ejecuta los trabajos de InfoSphere QualityStage.

Los siguientes componentes de cliente de InfoSphere DataStage se instalan en un sistema personal:

- IBM InfoSphere DataStage and QualityStage Designer
- IBM InfoSphere DataStage and QualityStage Director
- IBM InfoSphere DataStage and QualityStage Administrator

En esta guía de aprendizaje utilizará todos estos componentes al crear y ejecutar su proyecto de InfoSphere QualityStage.

---

## Objetivos del proyecto de la guía de aprendizaje

El objetivo de esta guía de aprendizaje es utilizar las etapas del IBM InfoSphere DataStage and QualityStage Designer para limpiar datos de clientes eliminando todas las direcciones de clientes duplicadas y facilitando el mejor formato para la dirección correcta.

En esta guía de aprendizaje su rol es el de un analista de bases de datos de un banco que proporciona varios servicios financieros. El banco dispone de una extensa base de datos de clientes. Sin embargo, hay problemas con la lista de clientes porque contiene muchos registros de nombres y direcciones para una única familia. El departamento de marketing quiere vender servicios adicionales a sus clientes, por lo que su tarea es encontrar y eliminar las direcciones duplicadas.

Por ejemplo, un matrimonio tiene cuatro cuentas, todas ellas a su nombre. Las cuentas incluyen dos cuentas corrientes, un plan de pensiones y un fondo de inversión.

En el sistema del banco el seguimiento de la información de los clientes se realiza por número de cuenta más que por nombre, número o dirección del cliente. Para este único cliente el banco dispone de cuatro entradas de dirección.

Con el fin de ahorrar dinero en los envíos, el banco quiere unificar la información de las familias, de manera que cada familia reciba solamente un envío. En esta guía de aprendizaje va a utilizar InfoSphere QualityStage para estandarizar todas las direcciones de los clientes. Además, debe ubicar y unificar todos los registros de clientes que viven en la misma dirección.

### Objetivos de aprendizaje

La finalidad de esta guía de aprendizaje es proporcionar un conocimiento del sistema de trabajo que se utiliza en el flujo de procesos de InfoSphere QualityStage a través de los distintos trabajos. Además, aprenderá a llevar a cabo las siguientes tareas:

- Configurar los trabajos del proyecto
- Configurar las etapas del proyecto
- Valorar los resultados de cada trabajo
- Aplicar los resultados a sus prácticas empresariales

Una vez completadas estas tareas, debería entender cómo las etapas de InfoSphere QualityStage reestructuran y limpian los datos a través de reglas empresariales aplicadas.

El tiempo para completar esta guía de aprendizaje es de 2'5 horas aproximadamente.

### **Nivel de habilidad**

Para entender esta guía de aprendizaje necesita tener un nivel de comprensión de análisis de datos intermedio-avanzado.

### **Audiencia**

Esta guía de aprendizaje está destinada a analistas empresariales y analistas de sistemas que tengan interés en entender InfoSphere QualityStage.

### **Requisitos del sistema**

- IBM InfoSphere Information Server
- Sistemas operativos: Microsoft Windows XP o Linux

### **Requisitos previos**

Para completar esta guía de aprendizaje debe saber utilizar

- IBM InfoSphere DataStage and QualityStage Designer
- Sistemas personales

### **Resultados esperados**

Una vez completada esta guía de aprendizaje, debería poder crear sus propios proyectos de InfoSphere QualityStage a través del cliente del Diseñador para cumplir los requisitos empresariales y los estándares de calidad de datos de su empresa.

---

## **Configurar la guía de aprendizaje**

El proceso de configuración de la guía de aprendizaje incluye crear una carpeta, crear un proyecto, copiar el trabajo y el archivo de datos de entrada al proyecto y luego iniciar el proyecto. Debe completar las tareas de configuración antes de empezar con el Módulo 1 de esta guía.

### **Crear una carpeta para los archivos de la guía de aprendizaje**

Copie los archivos de la guía de aprendizaje en una carpeta que cree en el sistema cliente de IBM InfoSphere QualityStage.

#### **Procedimiento**

1. Cree una carpeta en su sistema (por ejemplo, C:\TutorialData).

2. Busque el archivo `Parallel_job_tutorial.zip`, que se encuentra en el soporte de instalación. En el directorio que contiene el soporte de instalación, el archivo `Parallel_job_tutorial.zip` se encuentra en el directorio `directorio_padre\TutorialData\QualityStage`. Por ejemplo, el archivo `Parallel_job_tutorial.zip` podría estar en el directorio `is-client\TutorialData\QualityStage`.
3. Extraiga los archivos del archivo `Parallel_job_tutorial.zip` en la carpeta que ha creado en el paso 1 en la página 4.

## Crear el proyecto de guía de aprendizaje

Cree un nuevo proyecto para la guía de aprendizaje para mantener sus ejercicios de la guía de aprendizaje separados del resto de trabajos en InfoSphere QualityStage.

### Acerca de esta tarea

Para hacerlo, debe disponer de privilegios de administrador de IBM InfoSphere QualityStage.

Para crear el proyecto de guía de aprendizaje:

### Procedimiento

1. Seleccione **Iniciar > Todos los programas > IBM InfoSphere Information Server > Administrador de IBM InfoSphere DataStage and QualityStage**.
2. En la ventana Adjuntar a DataStage, escriba su nombre de usuario y contraseña y pulse **Aceptar**.
3. En el separador Proyectos, pulse **Añadir** para abrir la ventana Añadir nuevo proyecto.
4. En el campo **Nombre** especifique el nombre del nuevo proyecto (por ejemplo, Guía de aprendizaje).
5. Pulse **Aceptar** para crear el nuevo proyecto.
6. Pulse **Cerrar** para cerrar el cliente del Administrador.

## Copiar datos de la guía de aprendizaje

Copie los archivos de datos de la guía de aprendizaje de la carpeta de la guía de aprendizaje que usted mismo ha creado en el sistema cliente en la carpeta o el directorio del proyecto del IBM InfoSphere QualityStage sistema donde está instalada la capa del motor.

### Acerca de esta tarea

Al crear el proyecto de la guía de aprendizaje ha creado automáticamente una carpeta o directorio para el proyecto en el sistema donde está instalada la capa del motor. La capa del motor de InfoSphere Information Server puede estar instalada en el mismo sistema Windows que los clientes, o bien en un sistema Windows, UNIX o Linux separado. A veces se conoce a la capa del motor como el servidor de IBM InfoSphere DataStage and QualityStage.

### Procedimiento

1. Abra la carpeta de la guía de aprendizaje `TutorialData\QualityStage` que usted mismo ha creado en el sistema cliente y localice el archivo `input.csv`.

2. Abra la carpeta del proyecto en el sistema donde está instalada la capa del motor para el proyecto de la guía de aprendizaje que ha creado. Donde *tutorial\_project* corresponde al nombre del proyecto que ha creado. Algunos ejemplos de vía de acceso son:
  - Para un servidor de Windows: C:\IBM\InformationServer\Server\Projects\*tutorial\_project*
  - Para un servidor de UNIX o Linux: opt/IBM/InformationServer/Server/Projects/*tutorial\_project*
3. Copie el archivo .csv a la carpeta del proyecto en el servidor.

## Iniciar un proyecto

Utilice un proyecto del cliente del IBM InfoSphere DataStage and QualityStage Designer como contenedor para sus trabajos de IBM InfoSphere QualityStage.

### Acerca de esta tarea

Abra el cliente del Diseñador para iniciar la guía de aprendizaje. El trabajo paralelo del Diseñador proporciona un archivo ejecutable que ejecuta sus trabajos de InfoSphere QualityStage.

### Procedimiento

1. Pulse **Iniciar > Todos los programas > IBM InfoSphere Information Server > Diseñador de IBM InfoSphere DataStage and QualityStage**. Se abrirá la ventana Adjuntar a Proyecto.
2. En el campo **Dominio** introduzca el nombre del servidor al que está conectado.
3. En el campo **Nombre de usuario** introduzca su nombre de usuario.
4. En el campo **Contraseña** introduzca su contraseña.
5. En el campo **Proyecto** seleccione el proyecto que ha creado (por ejemplo, Guía de aprendizaje).
6. Pulse **Aceptar**. Se abrirá la ventana Nuevo en el cliente del Diseñador.

### Crear un trabajo

El cliente de IBM InfoSphere DataStage and QualityStage Designer proporciona una interfaz al motor paralelo que procesa los trabajos. En este paso usted guardará un trabajo en una carpeta en el repositorio de metadatos.

### Antes de empezar

Si no lo tiene ya abierto, abra el cliente.

### Procedimiento

1. En la ventana Nuevo seleccione la carpeta **Trabajos** en el panel de la izquierda y, a continuación, seleccione el icono **Trabajo paralelo** en el panel de la derecha.
2. Pulse **Aceptar**. En el área Diseño de trabajo aparecerá una nueva ventana de diseño de trabajo vacía.
3. Pulse **Guardar > Archivo**.
4. En la ventana Guardar trabajo paralelo como, pulse el botón derecho del ratón en la carpeta Trabajos y seleccione **Nuevo > Carpeta** en el menú de atajo.
5. Introduzca un nombre para la carpeta (por ejemplo, MyTutorial).
6. Pulse la nueva carpeta (MyTutorial) e introduzca Investigate1 en el campo **Nombre del elemento**.
7. Pulse **Guardar** para guardar el trabajo.

## Resultados

Ha creado un nuevo trabajo paralelo llamado Investigate y lo ha guardado en la carpeta Jobs\MyTutorial del repositorio. Utilice los procedimientos anteriores para crear 3 trabajos paralelos más en esta carpeta y denomínelos Standardize1, Unduplicate1 y Survive1.

## Qué hacer a continuación

Importe los datos de la guía de aprendizaje a su proyecto.

## Importar componentes de la guía de aprendizaje

Utilice el cliente de IBM InfoSphere DataStage and QualityStage Designer para importar los componentes de la guía de aprendizaje, que incluyen trabajos y definiciones de tabla de ejemplo, al proyecto de guía de aprendizaje.

## Acerca de esta tarea

Importe los componentes de la guía de aprendizaje para empezar las lecciones de la guía de aprendizaje.

Para importar componentes de la guía de aprendizaje:

## Procedimiento

1. Seleccione **Inicio > Todos los programas > IBM InfoSphere Information Server > IBM InfoSphere DataStage and QualityStage Designer**.
2. En la ventana Conectar con DataStage, escriba su nombre de usuario y contraseña.
3. Seleccione el proyecto **Guía de aprendizaje** en la lista **Proyectos** y pulse **Aceptar**. Se abrirá la ventana Nuevo del cliente del Diseñador.
4. Pulse **Cancelar** para cerrar la ventana Nuevo si está abriendo un trabajo existente en lugar de crear un trabajo u objeto nuevo.
5. Seleccione **Importar > ComponentesDataStage**.
6. En el campo **Importar desde archivo**, vaya al directorio en el cliente en el que copió los datos de la guía de aprendizaje, por ejemplo: C:\TutorialData\QualityStage. Seleccione el archivo **QualityStage\_Tutorial.dsx**.
7. Asegúrese de que se ha seleccionado **Importar todo**. También puede seleccionar **Realizar análisis de impacto**.
8. Pulse **Aceptar** para importar los trabajos y definiciones de tabla de ejemplo a una carpeta del repositorio llamada Guía de aprendizaje de QualityStage.

## Resultados

Los componentes aparecen en el repositorio bajo la carpeta **Guía de aprendizaje de QualityStage**. Puede abrir cada trabajo y ver cómo ha sido diseñado en el lienzo. Utilice estos trabajos como referencia para crear sus propios trabajos.

## Qué hacer a continuación

Puede empezar con el Módulo 1.

---

## Módulo 1: Investigar los de datos de origen

Este módulo explica cómo configurar y procesar un trabajo Investigate para facilitar datos a partir de los cuales pueda crear informes en la consola Web de IBM Information Server.

Puede utilizar la información de los informes para hacer supuestos básicos sobre los datos y los pasos que debe llevar a cabo para lograr el objetivo de facilitar una dirección legítima para cada usuario incluido en la base de datos.

### Objetivos de aprendizaje

Una vez completadas las lecciones de este módulo, deberá saber cómo llevar a cabo las siguientes tareas:

- Añadir etapas y enlaces a un trabajo IBM InfoSphere DataStage and QualityStage
- Configurar propiedades de etapa para especificar qué acción realizan al ejecutar el trabajo
- Cargar y procesar datos y metadatos de clientes
- Compilar y ejecutar un trabajo
- Generar datos para informes

El tiempo para completar este módulo es de 30 minutos aproximadamente.

### Lección 1.1: Configurar y enlazar un trabajo Investigate

Cree cada trabajo añadiendo etapas de Data Quality y archivos secuenciales y etapas de IBM InfoSphere DataStage al lienzo del IBM InfoSphere DataStage and QualityStage Designer. Todos los iconos del lienzo están enlazados para permitir que los datos fluyan del archivo de origen a cada etapa.

Si todavía no lo ha hecho, abra el cliente del Diseñador.

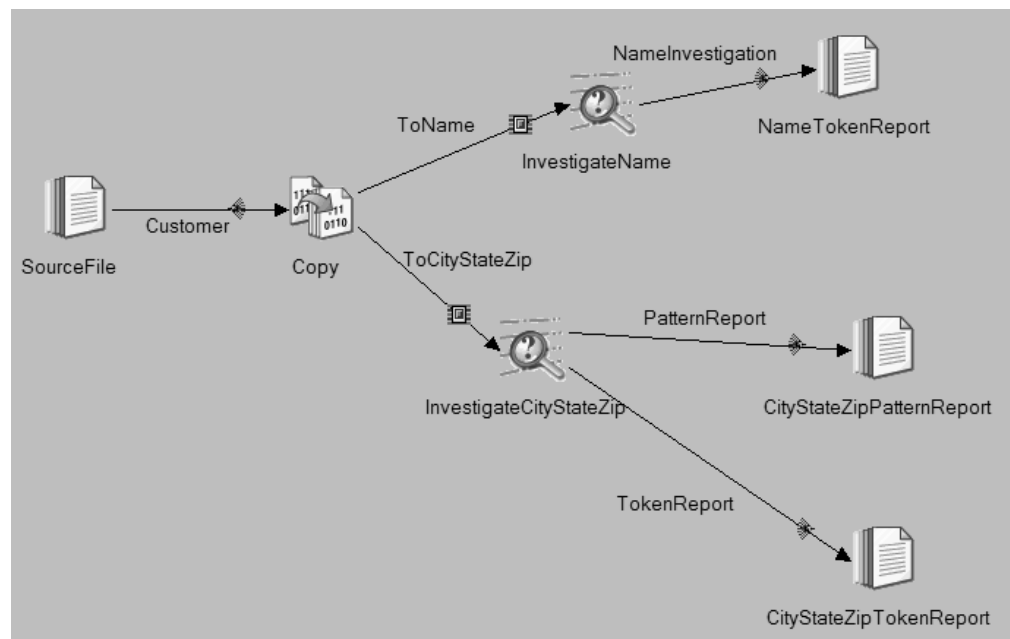
1. En el panel izquierdo del Diseñador vaya a la carpeta MyTutorial que usted mismo ha creado para esta guía de aprendizaje y efectúe una doble pulsación en Investigate1 para abrir el trabajo.
2. Pulse **Paleta > Calidad de datos** para seleccionar la etapa Investigate. Si no aparece la paleta, pulse **Ver > Paleta**.
3. Arrastre la etapa Investigate al lienzo del Diseñador y suéltela en medio del lienzo.
4. Arrastre una segunda etapa Investigate y suéltela bajo la primera etapa **Investigate**. Debe utilizar dos etapas Investigate para crear los datos de los informes.
5. Pulse **Paleta > Archivo** y seleccione **Archivo secuencial**.
6. Arrastre el **Archivo secuencial** al lienzo del Diseñador y suéltelo a la izquierda de la primera etapa Investigate. Este archivo secuencial es el archivo de origen.
7. Pulse **Paleta > Proceso** y seleccione la etapa Copy. Esta etapa duplica los datos del archivo de origen y los copia en las dos etapas Investigate.
8. Arrastre la etapa Copy al lienzo del Diseñador y suéltela entre el **Archivo secuencial** y la primera etapa Investigate.
9. Pulse **Paleta > Archivo**, y arrastre un segundo **Archivo secuencial** al lienzo del Diseñador y suéltelo a la derecha de la primera etapa Investigate.

Los datos de la etapa Investigate se envían al segundo **Archivo secuencial**, que es el archivo de destino.

10. Arrastre un tercer **Archivo secuencial** al lienzo del Diseñador y suéltelo a la derecha de la etapa Investigate y bajo el segundo **Archivo secuencial**. Ahora tiene un archivo de origen, una etapa Copy, dos etapas Investigate y dos archivos de destino.
11. Arrastre un cuarto **Archivo secuencial** al lienzo del Diseñador y suéltelo bajo el tercer **Archivo secuencial** como último archivo de destino. En el paso siguiente, enlace todas las etapas.
12. Pulse **Paleta > General > Enlazar**.
  - a. Pulse el botón derecho del ratón y arrastre un enlace del archivo de origen a la etapa Copy.

Si el enlace es rojo, púselo para activarlo y arrástrelo hasta que llegue a la etapa. Tiene que volverse de color negro.

Una vez enlazados todos los iconos del lienzo, puede pulsar una etapa y arrastrarla para cambiar su posición.
  - b. Siga enlazando el resto de etapas. La siguiente imagen muestra el trabajo Investigate completado con los nombres que deberá asignar a las etapas y a los enlaces en la siguiente lección.



### Punto de comprobación de la lección

Al configurar un trabajo Investigate, usted conecta el archivo de origen con sus datos y metadatos de origen a todas las etapas y enlaza las etapas a los archivos de destino.

Una vez completada esta lección, usted habrá aprendido acerca del Diseñador lo siguiente:

- Cómo añadir etapas al lienzo del Diseñador
- Cómo combinar etapas de calidad de datos y de proceso en el lienzo del Diseñador
- Cómo enlazar todas las etapas entre ellas

## Lección 1.2: Renombrar enlaces y etapas en un trabajo Investigate

Al crear un trabajo grande en el cliente del IBM InfoSphere DataStage and QualityStage Designer, es importante renombrar todas las etapas, archivos y enlaces con nombres con significado para evitar así cualquier confusión al seleccionar las vías de acceso durante la configuración de las etapas.

No utilice espacios al renombrar los enlaces y las etapas. El cliente del Diseñador restablece el nombre a su valor genérico si introduce algún espacio. El objetivo de esta lección es sustituir los nombres genéricos de los iconos del lienzo con nombres más adecuados.

Para renombrar los iconos del lienzo:

1. Para renombrar una etapa, siga los siguientes pasos:
  - a. Pulse el nombre del ArchivoSecuencial de origen hasta que aparezca un recuadro resaltado alrededor del nombre.
  - b. Introduzca ArchivoOrigen en el recuadro.
  - c. Pulse fuera del recuadro para deseleccionarlo.
2. Para renombrar un enlace, siga los siguientes pasos:
  - a. Pulse con el botón derecho del ratón en el nombre genérico del enlace DSLinkXX que conecta el ArchivoOrigen a la etapa Copy y seleccione **Renombrar** en el menú de atajo. Aparecerá un recuadro resaltado alrededor del nombre predeterminado.
  - b. Introduzca Cliente y pulse fuera del recuadro. El nombre del enlace predeterminado cambiará a Cliente.
3. Pulse con el botón derecho del ratón el nombre genérico del enlace que conecta la etapa Copy a la primera etapa Investigate.
4. Repita el paso 2, pero introduzca ToName en el recuadro.
5. Pulse con el botón derecho del ratón el nombre genérico del enlace que conecta la etapa Copy a la segunda etapa Investigate.
6. Repita el paso 2, pero introduzca ACityStateZip en el recuadro.
7. Pulse en los nombres de las siguientes etapas y escriba el nuevo nombre de etapa en el recuadro resaltado:

Etapas	Cambiar a
Copy	Copy
Investigate (la primera)	InvestigateName
Investigate (la segunda)	InvestigateCityStateZip

8. Renombre los tres archivos de destino de la parte superior en el orden siguiente:
  - a. NameTokenReport
  - b. CityStateZipPatternReport
  - c. CityStateZipTokenReport
9. En los nombres de los siguientes enlaces, seleccione **Renombrar** e introduzca el nuevo nombre del enlace en el recuadro resaltado:

Enlace	Cambiar a
De InvestigateName a NameTokenReport	NameInvestigation



Enlace	Cambiar a
De InvestigateCityStateZip a CityStateZipPatternReport	PatternReport
De InvestigateCityStateZip a CityStateZipTokenReport	TokenReport

Renombrar los elementos del lienzo del Diseñador ayuda a una mejor organización del trabajo Investigate.

### Punto de comprobación de la lección

En esta lección usted ha cambiado los nombres genéricos de las etapas y enlaces a nombres más adecuados para el trabajo.

Ha aprendido a realizar las siguientes tareas:

- Cómo seleccionar el campo de nombre predeterminado para editarlo
- El método correcto para cambiar el nombre


## Lección 1.3: Configurar el archivo de origen

Los datos y metadatos de origen van adjuntos al ArchivoOrigen como los datos de origen del trabajo.

El objetivo de esta lección es adjuntar los datos de entrada de nombres y direcciones de clientes y cargar los metadatos.

Para añadir datos y metadatos al trabajo Investigate debe configurar el archivo de origen para ubicar el archivo de datos de entrada **input.csv** almacenado en su sistema y cargar las columnas de metadatos.

Para configurar el archivo de origen:

1. Efectúe una doble pulsación en el icono **ArchivoOrigen** para abrir el separador **Propiedades** en la ventana ArchivoOrigen - Archivo secuencial.
2. Seleccione el archivo de datos de la guía de aprendizaje:
  - a. Pulse **Origen > Archivo** para activar el campo **Archivo**.
  - b. Pulse el  en el campo **Archivo** y seleccione **Localizar Archivo**.
  - c. Ubique el directorio en el servidor donde ha copiado el archivo input.csv del DVD (por ejemplo, C:\IBM\InformationServer\Server\Projects\tutorial).
  - d. Pulse **input.csv** para seleccionar el archivo y, a continuación, pulse **Aceptar**.
3. Pulse **Opciones > La primera línea contiene los nombres de las columnas** y luego seleccione **True** en la lista **La primera línea contiene los nombres de las columnas**.
4. Pulse el separador **Columnas**.
5. Pulse **Cargar**.
6. En la ventana Definiciones de tabla pulse la carpeta **Guía de aprendizaje de QualityStage > Definiciones de tabla**. Esta carpeta ha sido creada al importar los metadatos de ejemplo de la guía de aprendizaje.
7. Pulse **Entrada** en la carpeta Definiciones de tabla y pulse **Aceptar**.

8. Pulse **Aceptar** en la ventana Seleccionar columnas para cargar los metadatos de ejemplo.
9. Pulse **Ver datos** para visualizar la calidad de los datos de entrada.
10. En la primera ventana Navegador de datos para el enlace de salida, seleccione el número filas a visualizar y pulse **Aceptar**. Puede dejar el número de filas como 100.
11. En la segunda ventana Navegador de datos para el enlace de salida, verá los nombres y direcciones de los clientes del banco. Las direcciones aparecen desordenadas, complicando así la tarea del banco de analizar los datos.
12. Pulse **Cerrar** para cerrar la ventana Navegador de datos.
13. Pulse **Aceptar** para actualizar la etapa Sequential File con los cambios que ha realizado.

### Punto de comprobación de la lección

En esta lección usted ha adjuntado los datos de entrada (nombres y direcciones de clientes) y ha cargado los metadatos.

Ha aprendido a realizar las siguientes tareas:

- Adjuntar datos de origen al archivo de origen
- Añadir metadatos de columna al archivo de origen

## Lección 1.4: Configurar la etapa Copy

La etapa Copy duplica los datos de origen y los envía a las dos etapas Investigate.

Esta lección explica cómo configurar una etapa Processing, la etapa Copy, para duplicar el origen y enviar los metadatos de salida a las dos etapas Investigate.

Para configurar una etapa Copy:

1. Efectúe una doble pulsación en el icono de la etapa Copy para abrir el separador **Propiedades** en la ventana Copy - Copy.
2. Pulse el separador **Entrada > Columnas**. Los metadatos que ha cargado en el ArchivoOrigen se han propagado a la etapa Copy.
3. Pulse el separador **Salida > Correlacionar** para correlacionar las columnas del panel **Columnas** de la izquierda con el panel **ToName**.
4. En el campo **Nombre de salida** encima del panel **Columnas** de la pantalla, seleccione **ToName** en caso de que no esté ya seleccionado. Seleccionar el nombre de salida correcto asegura que los datos van a la etapa Investigate, InvestigateName, o InvestigateCityStateZip correcta.
5. Copie los datos del panel **Columnas** en el panel **ToName**:
  - a. Coloque el cursor en el panel **Columnas**, pulse el botón derecho del ratón y seleccione **Seleccionar todo** en el menú de atajo.
  - b. Pulse el botón derecho del ratón y seleccione **Copiar** en el menú de atajo.
  - c. Coloque el cursor en el panel **ToName**, pulse el botón derecho del ratón y seleccione **Pegar columna** en el menú de atajo. Los metadatos de columna se copian en el panel **ToName** y las líneas muestran el enlace del panel **Columnas** al panel **ToName**.
6. En el campo **Nombre de salida** sobre el panel **Columnas**, seleccione **ToCityStateZip** en el menú desplegable.
7. Repita el paso 5 para correlacionar el panel **Columnas** con el panel **ToCityStateZip**.
8. Pulse **Aceptar** para guardar la etapa Copy actualizada.

Este proceso le muestra cómo correlacionar columnas a dos salidas diferentes.

### Punto de comprobación de la lección

En esta lección usted ha correlacionado los metadatos de entrada a los dos enlaces de salida para seguir propagando los metadatos a las dos etapas posteriores.

Ha aprendido a realizar las siguientes tareas:

- Añadir una etapa de IBM InfoSphere DataStage a un trabajo de IBM InfoSphere QualityStage
- Propagar metadatos a la etapa posterior
- Correlacionar metadatos a dos enlaces de salida

## Lección 1.5: Etapa Investigate, configurar para revisar nombres

La opción Investigación de palabras de la etapa Investigate analiza los datos de nombres y direcciones en patrones reconocibles utilizando conjuntos de reglas que clasifican los nombres y las direcciones personales.


La etapa Investigate analiza todos los registros del archivo de origen. En esta lección debe seleccionar el conjunto de reglas NAME para aplicar los estándares USPS.

Para configurar la etapa Investigate:

1. Efectúe una doble pulsación en el icono **InvestigateName**.
2. Pulse el separador **Investigación de palabras** para abrir la ventana Investigación de palabras.
3. Seleccione **Nombre** en la sección **Columnas de datos disponibles** y pulse

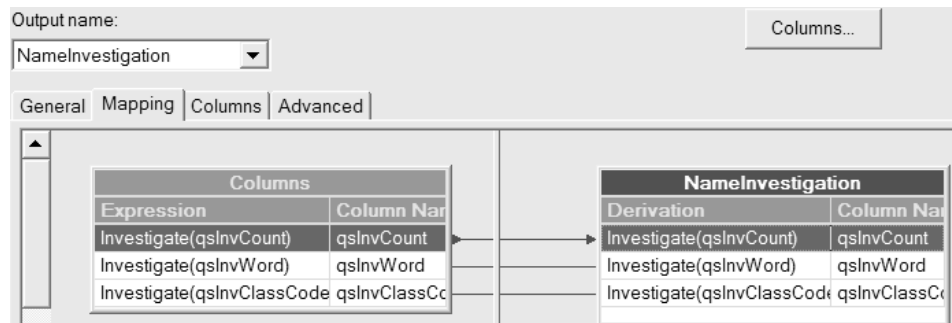


para mover la columna **Nombre** al panel **Columnas estándar**. La etapa InvestigateName analiza la columna **Nombre** utilizando el conjunto de reglas que usted mismo ha seleccionado en el paso 4.

4. En el campo **Conjunto de reglas**, pulse  para seleccionar un conjunto de reglas para la etapa InvestigateName.
  - a. En la ventana Conjuntos de reglas efectúe una doble pulsación en la carpeta **Reglas de estandarización** para abrir el árbol Reglas de estandarización.
  - b. Efectúe una doble pulsación en la carpeta **USA**, una doble pulsación en la carpeta **USNAME** y, a continuación, seleccione **USNAME**. El conjunto de reglas USNAME analiza la columna **Nombre** de conformidad con la normativa postal de los Estados Unidos para los nombres.
  - c. Pulse **Aceptar** para salir de la ventana Conjuntos de reglas.
5. Pulse el recuadro de selección **Informe de señales** en la sección **Conjunto de datos de salida** de la ventana.
6. Pulse el separador **Propiedades de etapa > Salida > Correlación**.
7. Correlacione las columnas de salida:
  - a. Pulse el panel **Columnas**.
  - b. Pulse el botón derecho del ratón y seleccione **Seleccionar todo** en el menú de atajo.
  - c. Pulse el botón derecho del ratón y seleccione **Copiar** en el menú de atajo.

- d. Pulse en el panel **NameInvestigation**.
- e. Pulse el botón derecho del ratón y seleccione **Pegar columna** en el menú de atajo. Las columnas de la izquierda están correlacionadas con las columnas de la derecha.

La correlación de **NameInvestigation** debería tener el siguiente aspecto:



8. Pulse el separador **Columnas**. Fíjese en que las columnas de Salida se llenan al correlacionar las columnas en el separador **Correlación**.
9. Amplíe el tipo de datos para las columnas `qsInvWord` y `qsInvClassCode`.
  - a. En la fila de la cuadrícula **Columnas** para la columna `qsInvWord`, seleccione **Unicode** en la lista **Ampliado**.
  - b. Repita el paso 9a para la columna `qsInvClassCode`.
  - c. Pulse **Aceptar**.
10. Pulse **Aceptar** y luego pulse **Archivo > Guardar** para guardar el trabajo de investigación actualizado.

## Resumen de la lección

Esta lección ha explicado cómo configurar la etapa **Investigate** utilizando el conjunto de reglas **USNAME**.

Usted ha aprendido a configurar la etapa **Investigate** del trabajo **Investigate** realizando las siguientes tareas:

- Seleccionar las columnas a investigar
- Seleccionar un conjunto de reglas para aplicarlo a los datos
- Correlacionar las columnas de salida

## Lección 1.6: Etapa **Investigate**, configurar para revisar regiones geográficas

La opción **Investigación de palabras** de la etapa **Investigate** analiza los datos de nombres y direcciones en patrones reconocibles utilizando conjuntos de reglas que clasifican los nombres y las direcciones personales.

La etapa **Investigate** analiza todos los registros del archivo de origen. En esta lección debe aplicar el conjunto de reglas **USAREA** para aplicar los estándares **USPS**.


Para configurar el icono **InvestigateCityStateZip**:

1. Efectúe una doble pulsación en el icono **InvestigateCityStateZip**.
2. Pulse el separador **Investigación de palabras** para abrir la ventana **Investigación de palabras**.

3. Seleccione las siguientes columnas en el panel **Columnas de datos disponibles** para moverlas al panel **Columnas estándar**. La segunda etapa Investigate analiza las columnas de direcciones mediante el conjunto de reglas que usted mismo ha seleccionado en el paso 5.

- Ciudad
- Región
- Zip5
- Zip4

4. Pulse  para mover las columnas seleccionadas al panel **Columnas estándar**.


5. En el campo **Conjunto de reglas**, pulse  para ubicar un conjunto de reglas para InvestigateCityStateZip.


- a. En la ventana Conjuntos de reglas efectúe una doble pulsación en la carpeta **Reglas de estandarización** para abrir el árbol Reglas de estandarización.
- b. Efectúe una doble pulsación en la carpeta **USA** y efectúe otra doble pulsación en la carpeta **USAREA** y seleccione el archivo **USAREA**. El conjunto de reglas USAREA analiza las columnas Ciudad, Estado, Zip5 y Zip4 de conformidad con la normativa postal de los Estados Unidos.
- c. Pulse **Aceptar** para salir de la ventana Conjuntos de reglas. Aparecerá USAREA.SET en el campo **Conjunto de reglas**.

6. Pulse los recuadros de selección **Informe de señales** y **Informe de patrones** en la sección **Conjunto de datos de salida** de la ventana. Al asignar datos a 2 salidas debe verificar que la ordenación de enlaces es correcta. La ordenación de enlaces garantiza que los datos son enviados a los informes correctos a través de los enlaces asignados que usted mismo ha nombrado en la Lección 1.2. El separador **Ordenación de enlaces** no se muestra si solamente hay un enlace.

7. Si necesita cambiar el orden de visualización de los enlaces, pulse el separador **Propiedades de etapa > Ordenación de enlaces** y seleccione el enlace de salida que desea mover.

8. Desplace los enlaces hacia arriba o hacia abajo tal y como se describe a continuación:

• Pulse  para subir el nombre del enlace un nivel.

• Pulse  para bajar el nombre del enlace un nivel.

La siguiente imagen muestra el orden correcto de los enlaces.



9. Pulse el separador **Salida > Correlación**. Puesto que existen dos enlaces de salida desde la segunda etapa Investigate, debe correlacionar las columnas con cada enlace:
  - a. En la lista **Nombre de salida** encima del panel **Columnas**, seleccione **PatternReport**.
  - b. Seleccione el panel **Columnas**.
  - c. Pulse el botón derecho del ratón y seleccione **Seleccionar todo** en el menú de atajo.
  - d. Pulse el botón derecho del ratón y seleccione **Copiar** en el menú de atajo.
  - e. Seleccione el panel **PatternReport**, pulse el botón derecho del ratón y seleccione **Pegar columna** en el menú de atajo. Las columnas están correlacionadas con el enlace de salida **PatternReport**.
  - f. En la lista **Nombre de salida** encima del panel **Columnas**, seleccione **TokenReport**.
  - g. Repita los pasos de b a e, excepto el paso e, donde debe seleccionar el panel **TokenReport**.
10. Amplíe el tipo de datos para las columnas de ambos enlaces de salida.
  - a. Pulse el separador **Salida > Columnas**.
  - b. En la lista **Nombre de salida** encima de la página **Columnas**, seleccione **PatternReport**.
  - c. En la fila de la cuadrícula **Columnas** para cada columna que tenga un tipo de SQL de VarChar o Char, seleccione **Unicode** en la lista **Ampliado**.
  - d. En la lista **Nombre de salida** encima del panel **Columnas**, seleccione **TokenReport**.
  - e. Repita el paso 10c para las columnas de este enlace de salida.
  - f. Pulse **Aceptar**.
11. Pulse **Aceptar** para cerrar la ventana InvestigateCityStateZip.

## Resumen de la lección

Esta lección le ha explicado cómo configurar la segunda etapa Investigate para el conjunto de reglas AREA.

Ha aprendido a configurar la segunda etapa Investigate del trabajo Investigate mediante los siguientes temas:

- Seleccionar las columnas a investigar
- Seleccionar un conjunto de reglas para aplicarlo a los datos
- Verificar la ordenación de enlaces de los informes de salida
- Correlacionar las columnas de salida con dos enlaces de salida


## Lección 1.7: Configurar informes de destino

La información de los datos de origen y los metadatos de columna se propagan a los archivos de datos de destino para utilizarlos posteriormente en la creación de informes de Investigación.

El trabajo Investigate convierte los datos de origen sin formato en datos legibles que se configuran más tarde como informes de Investigación.

Para configurar los archivos de datos:

1. Efectúe una doble pulsación en el icono **NameTokenReport** en el lienzo del cliente del Diseñador.

2. En **Entrada > Propiedades**, pulse **Destino > Archivo**.
3. En el campo **Archivo**, pulse  y vaya hasta la vía de acceso de la carpeta en el sistema servidor donde reside el archivo de datos de entrada. En los pasos siguientes, va a especificar nombres de archivo de destino en separadores de entrada de etapa.
4. En el campo **Nombre de archivo**, escriba `tokrpt.csv` para visualizar la vía de acceso y el nombre del archivo en el campo **Archivo**, (por ejemplo, `C:\IBM\InformationServer\Server\Projects\tutorial\tokrpt.csv`).
5. Especifique la forma en que la etapa recopila los datos antes de grabarlos en el archivo secuencial. Debe especificar detalles de recopilación porque la etapa Investigate se ejecuta en modalidad paralela y la etapa Sequential File se ejecuta en modalidad secuencial.
  - a. Pulse el separador **Particionamiento**.
  - b. En la lista **Tipo de recopilador**, seleccione **Ordenado**. Este método lee todas las filas de la primera partición, luego todas las filas de la segunda partición, y así sucesivamente.
  - c. En la sección **Ordenación**, seleccione **Ordenar**.
  - d. En las columnas disponibles, pulse **qsInvCount**.
  - e. Pulse **Aceptar**.
6. Efectúe una doble pulsación en el icono **CityStateZipPatternReport**.
7. Repita los pasos 2 a 5, pero escriba `areapatrpt.csv` para el nombre de archivo.
8. Efectúe una doble pulsación en el icono **CityStateZipTokenReport**.
9. Repita los pasos 2 a 5, pero escriba `areatokrpt.csv` para el nombre de archivo.

### Punto de comprobación de la lección

Esta lección le ha explicado cómo configurar los archivos de destino para utilizarlos como informes.



Usted ha configurado los tres archivos de datos de destino enlazando los datos a cada archivo de informe.

## Lección 1.8: Compilar y ejecutar trabajos

Puede probar el trabajo Investigate ejecutando el compilador y ejecutando después el trabajo para procesar los datos de los informes.

Compile el trabajo Investigate en el cliente del Diseñador. Una vez el trabajo está compilado correctamente, abra el cliente del Director y ejecute el trabajo.

Para compilar y ejecutar el trabajo:

1. Pulse **Archivo > Guardar** para guardar el trabajo Investigate en el lienzo del Diseñador.
2. Pulse el  para compilar el trabajo. Se abrirá la ventana Compilar trabajo y el trabajo se empezará a compilar. Una vez finaliza el compilador, se mostrará el siguiente mensaje Trabajo compilado satisfactoriamente sin errores.
3. Pulse **Herramientas > Ejecutar Director**. Se abrirá la aplicación del Director mostrando el trabajo en la vista de estado.
4. Pulse el  para abrir la ventana Opciones de ejecución del trabajo.

## 5. Pulse Ejecutar.

Tras ejecutar el trabajo, aparecerá **Finalizado** en la columna **Estado**.

### **Punto de comprobación de la lección**

En esta lección usted ha aprendido a compilar y procesar un trabajo Investigate.

Usted ha procesado los datos en tres archivos de salida mediante las siguientes tareas:

- Compilar el trabajo Investigate
- Ejecutar el trabajo Investigate en el Director

## **Módulo 1: Resumen**

En el Módulo 1 ha establecido, configurado y procesado un trabajo Investigate en IBM InfoSphere DataStage and QualityStage Designer.

Un trabajo Investigate busca cada registro columna por columna y analiza el contenido de los datos de las columnas que usted ha seleccionado. El trabajo Investigate carga los datos de origen de nombres y direcciones almacenados en la base de datos del banco, analiza las columnas en un formato que pueda ser analizado y organiza los datos en tres archivos de datos.

El trabajo Investigate convierte los datos de origen sin formato en datos legibles que puede configurar como informes Investigation mediante Consola web de IBM InfoSphere Information Server. Debe seleccionar los **Informes de QualityStage** para acceder a la interfaz de informes de la consola Web.

El siguiente módulo organiza los datos sin formato en datos estandarizados que proporcionan datos utilizables para la coincidencia y la supervivencia.

### **Lecciones aprendidas**

Una vez completado este módulo, usted habrá aprendido los siguientes conceptos y tareas:

- Cómo establecer y enlazar etapas en un trabajo de manera que los datos se propaguen de una etapa a la etapa posterior
- Cómo configurar las propiedades de la etapa para aplicar el conjunto de reglas correcto al analizar los datos
- Cómo compilar y ejecutar un trabajo
- Cómo generar datos para análisis



---

## Módulo 2: Estandarizar datos

Este módulo explica cómo configurar y procesar un trabajo de Estandarización para estandarizar la información de nombres y direcciones procedente de la base de datos del banco.

Cuando ha trabajado sobre los datos en el Módulo 1, el formato de algunas direcciones era libre y no estaba estandarizado. Eliminar las direcciones de clientes duplicadas y garantizar que una dirección corresponde a un cliente en concreto sería muy difícil si no se estandarizaran antes los datos.

La estandarización o el condicionamiento garantizan que los datos de origen son coherentes internamente, es decir, cada tipo de datos incluye el mismo tipo de contenido y está en el mismo formato. Cuando se utilizan datos coherentes, el sistema puede buscar coincidencias en los datos de direcciones con mayor precisión utilizando una de las etapas de coincidencia.

### Objetivos de aprendizaje

Una vez completadas las lecciones de este módulo, deberá saber cómo llevar a cabo las siguientes tareas:

1. Añadir etapas y enlaces a un trabajo Standardize
2. Configurar las propiedades de las distintas etapas para procesar correctamente los datos al ejecutar el trabajo
3. Trabajar con nulos de manejo mediante derivaciones
4. Generar la distribución de frecuencia para datos estandarizados

El tiempo para completar este módulo es de 60 minutos aproximadamente.

### Lección 2.1: Configurar un trabajo Standardize

Estandarizar los datos es el primer paso de la limpieza de datos. En la Lección 2.1, ha añadido varias etapas al lienzo del IBM InfoSphere DataStage and QualityStage Designer. Estas etapas incluyen la etapa Transformer, que aplica derivaciones para manejar nulos y la etapa Match Frequency, que añade datos de frecuencia.

Si todavía no lo ha hecho, abra el cliente del Diseñador.

Tal y como ha aprendido en la Lección 1.1, debe añadir etapas y enlaces al lienzo del Diseñador para crear un trabajo Standardize. El trabajo Investigate que ha completado le ha ayudado a determinar cómo formular una estrategia empresarial a través de informes de Investigación. El trabajo Standardize aplica conjuntos de reglas a los datos de origen, condicionándolos así para que coincidan.

Para configurar un trabajo Standardize:

1. En el panel izquierdo del Diseñador vaya a la carpeta MyTutorial que ha creado para esta guía de aprendizaje y efectúe una doble pulsación en Standardize1 para abrir el trabajo.
2. Arrastre los siguientes iconos al lienzo del Diseñador desde la paleta.
  - El icono **Calidad de datos > Standardize** al medio del lienzo
  - El icono **Archivo > Archivo secuencial** a la izquierda de la etapa Standardize
  - El icono **Archivo > Conjunto de datos** a la derecha de la etapa Standardize
  - El icono **Proceso > Transformer** entre la etapa Standardize y el archivo **Conjunto de datos**

- El icono **Proceso** > **Copy** entre la etapa Transformer y el archivo **Conjunto de datos**
- El icono **Calidad de datos** > **Match Frequency** debajo de la etapa Copy
- El segundo icono **Archivo** > **Conjunto de datos** a la derecha de la etapa Match Frequency

Una vez enlazados los archivos y las etapas, puede adjuntar su ubicación en el lienzo.

3. Pulse con el botón derecho del ratón el icono **Archivo secuencial** y arrástrelo para crear un enlace del icono **Archivo secuencial** al icono de la etapa Standardize.

4. Arrastre los enlaces a las etapas restantes, del mismo modo que en el paso 3. Si el enlace es rojo, púlselo para activarlo y arrástrelo hasta que llegue a la etapa. Tiene que volverse de color negro.

Una vez enlazados todos los iconos del lienzo, puede pulsar sobre una etapa y arrastrarla para cambiar su posición.

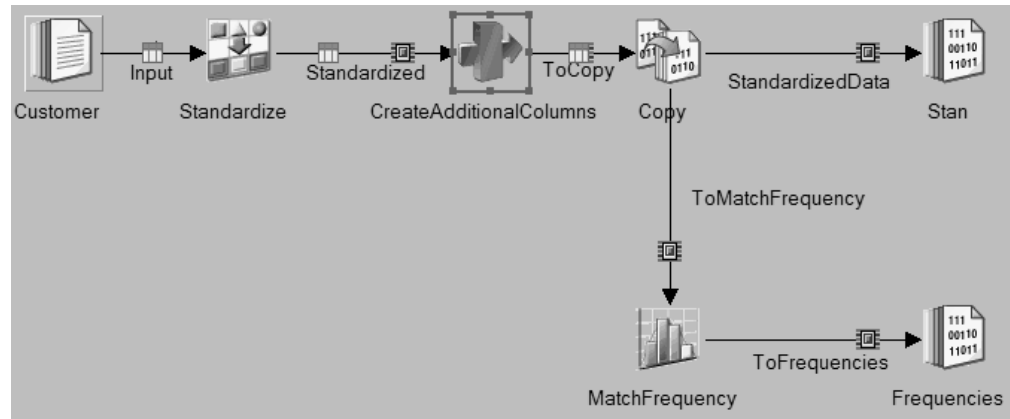
5. Pulse en los nombres de las siguientes etapas y escriba el nuevo nombre de etapa en el recuadro resaltado:

Etapa	Cambiar a
SequentialFile	Customer
etapa Standardize	Standardize
Etapa Transformer	CreateAdditionalColumns
Etapa Copy	Copy
Archivo Conjunto_Datos (la salida de la etapa Copy)	Stan
etapa Match Frequency	MatchFrequency
Archivo Conjunto_Datos (la salida de la etapa Match Frequency)	Frequencies

6. Pulse con el botón derecho del ratón los nombres de los siguientes enlaces, seleccione **Renombrar** e introduzca el nuevo nombre del enlace en el recuadro resaltado:

Enlace	Cambiar a
De Customer a Standardize	Input
De Standardize a CreateAdditionalColumns	Standardized
De CreateAdditionalColumns a Copy	ToCopy
De Copy a Stan	StandardizedData
De Copy a MatchFrequency	ToMatchFrequency
De MatchFrequency a Frequencies	ToFrequencies

La siguiente imagen muestra las etapas y enlaces del trabajo Standardize.



### Punto de comprobación de la lección

En esta lección ha aprendido a configurar un trabajo Standardize. La importancia de la etapa Standardize reside en generar el tipo de datos que puedan utilizarse luego en un trabajo de coincidencia.

Usted ha configurado y enlazado un trabajo Standardize mediante las siguientes tareas:

- Añadir etapas de Calidad de datos y de Proceso al lienzo del Diseñador
- Enlazar todas las etapas
- Renombrar los enlaces y etapas

## Lección 2.2: Configurar las propiedades de etapa del trabajo Standardize

Las propiedades de cada una de las etapas del trabajo Standardize deben configurarse en el lienzo de IBM InfoSphere DataStage and QualityStage Designer.


Complete las siguientes tareas para configurar el trabajo Standardize:

- Cargue los datos y metadatos de origen
- Añada conjuntos de reglas compatibles para nombres y direcciones de Estados Unidos
- Aplique derivaciones a los conjuntos de nulos
- Copie los datos en los dos enlaces de salida
- Cree datos de frecuencia

### Configurar las propiedades del archivo de Cliente

Para configurar las propiedades de etapa del Cliente (archivo de origen):

1. Efectúe una doble pulsación en el icono del archivo de origen **Cliente** para abrir el separador **Propiedades** en la ventana Cliente - Archivo secuencial.
2. Pulse **Origen > Archivo**.

3. En el campo **Archivo**, pulse  y vaya hasta la vía de acceso de la carpeta en el sistema servidor donde reside el archivo de datos de entrada.
4. Seleccione **input.csv** y luego pulse **Aceptar**. Este es el archivo de origen que lee la etapa Standardize al ejecutarse el trabajo.


5. Pulse **Opciones > La primera línea contiene los nombres de las columnas** y luego seleccione **True** en la lista **La primera línea contiene los nombres de las columnas**.
6. Pulse el separador **Columnas** y pulse **Cargar**.
7. En la ventana Definiciones de tabla pulse la carpeta **Guía de aprendizaje de QualityStage**. Esta carpeta ha sido creada al importar los metadatos de ejemplo de la guía de aprendizaje.
8. Pulse **Definiciones de tabla > Entrada**. Las definiciones de tabla se cargan en el separador **Columnas** del archivo de origen **Ciente**.
9. Pulse **Aceptar** para cerrar la ventana Definiciones de tabla.
10. Vuelva a pulsar **Aceptar** en la ventana Seleccionar columnas para cargar los metadatos de ejemplo.
11. Pulse **Aceptar** para cerrar el archivo de origen **Ciente**.

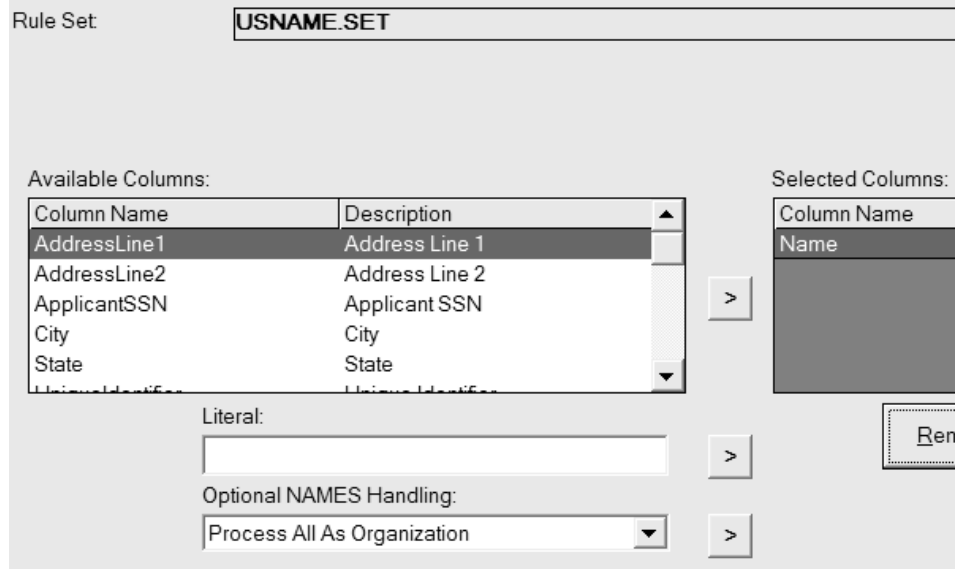
Los datos de origen se adjuntan al archivo de origen **Ciente** y las definiciones de tabla se cargan para organizar los datos en columnas de direcciones estándar.

### Configurar la etapa Standardize

La etapa Standardize aplica reglas a los datos de nombres y direcciones para analizar los datos en un formato de columna estándar.

Para configurar la etapa Standardize:

1. Efectúe una doble pulsación sobre el icono de la etapa Standardize para abrir la ventana Etapa Standardize.
2. Pulse el separador **Nuevo Proceso** para abrir la ventana Proceso de reglas Standardize.
3. En el campo **Conjunto de reglas** pulse **Reglas de estandarización > USA**. Los conjuntos de reglas de la carpeta **Reglas de estandarización** son específicos del dominio para trabajos de estandarización. Puede seleccionar conjuntos de reglas de esta carpeta para crear estructuras de datos estándar de la industria coherentes y estructuras coincidentes.
4. Abra la carpeta **USNAME**.
  - a. Seleccione el conjunto de reglas **USNAME** y pulse **Aceptar**. Aparecerá **USNAME.SET** en el campo **Conjunto de reglas**. Debe seleccionar este conjunto de reglas porque los datos de nombres y direcciones pertenecen a los Estados Unidos.
  - b. En el panel **Columnas disponibles** seleccione **Nombre**.
  - c. Pulse  para mover la columna **Nombre** al panel **Columnas seleccionadas**. El campo **Manejo de NOMBRES opcional** está activado.
  - d. Pulse **Aceptar**.



5. Pulse el separador **Nuevo Proceso** para abrir la ventana Proceso de reglas Standardize.
  6. En el campo **Conjunto de reglas** pulse **Reglas Standardize > EEUU** y seleccione el conjunto de reglas **USADDR**.
  7. Seleccione los siguientes nombres de columnas en el panel **Columnas Disponibles** y muévalas al panel **Columnas Seleccionadas**:
    - LíneaDirección1
    - LíneaDirección2
  8. Pulse **Aceptar**.
  9. Pulse el separador **Nuevo Proceso** para abrir la ventana Proceso de reglas Standardize.
  10. En el campo **Conjunto de reglas** pulse **Reglas Standardize > USA** y seleccione el conjunto de reglas **USAREA**.
  11. Seleccione los siguientes nombres de columnas en el panel **Columnas Disponibles** y muévalas al panel **Columnas Seleccionadas**:
    - Ciudad
    - Región
    - Zip5
    - Zip4
- Nota:** Mantenga el orden de las columnas. Zip5 debe preceder a Zip4.
12. Pulse **Aceptar**.
  13. Pulse el separador **Nuevo Proceso** para abrir la ventana Proceso de reglas Standardize.
  14. En el campo **Conjunto de reglas**, pulse **Reglas de estandarización > EEUU** y seleccione el conjunto de reglas **USTAXID**.
  15. Seleccione el siguiente nombre de columna en el panel **Columnas disponibles** y muévalas al panel **Columnas seleccionadas**:
    - ApplicantSSN
  16. Pulse **Aceptar**.
  17. Correlacione las columnas de salida de la etapa Standardize.
    - a. Pulse el separador **Propiedades de etapa**.

- b. Pulse el separador **Salida > Correlación**.
  - c. En el panel **Columns** pulse el botón derecho del ratón y seleccione **Seleccionar todo** en el menú de atajo.
  - d. Pulse el botón derecho del ratón y seleccione **Copiar** en el menú de atajo.
  - e. Vaya al panel **Standardized**, pulse el botón derecho del ratón y seleccione **Pegar columna** en el menú de atajo.
18. Guarde las definiciones de tabla en la carpeta **Definiciones de tabla**.
- a. Pulse el separador **Columns**.
  - b. Pulse **Guardar**. Se abre la ventana Guardar definición de tabla.
  - c. En el campo **Tipo de origen de datos**, introduzca Definiciones de tabla.
  - d. En el campo **Nombre del origen de datos**, escriba Standardized.
  - e. En el campo **Nombre de archivo/tabla** escriba Standardized.
  - f. Pulse **Aceptar** para abrir la ventana Guardar definición de tabla como.
  - g. Guarde las definiciones de tabla de Standardization en la carpeta Definición de tabla que está en un nivel inferior a la carpeta del proyecto, por ejemplo, **Guía de aprendizaje de QualityStage > Definiciones de tabla**.
  - h. Confirme los cambios y salga de las ventanas.

Ha configurado la etapa Standardize para aplicar los conjuntos de reglas USNAME, USADDR, USAREA y USTAXID a los datos de cliente y ha guardado las definiciones de tabla.

### Configurar la etapa Transformer

La etapa Transformer aumenta el número de columnas que la etapa de coincidencia utiliza para seleccionar las coincidencias. La etapa Transformer también aplica derivaciones para manejar valores nulos.

Para configurar las propiedades de la etapa Transformer:

1. Efectúe una doble pulsación sobre el icono de la etapa CreateAdditionalColumns para abrir la ventana Etapa Transformer.
2. En la sección superior de la ventana, pulse el botón derecho del ratón en cualquier columna del recuadro de enlace **Standardized** y seleccione **Seleccionar todo** en el menú de atajo. Todas las columnas del recuadro de enlace **Standardized** quedan resaltadas.
3. Pulse el botón derecho del ratón y seleccione **Copiar** en el menú de atajo.
4. Vaya al panel **ToCopy** en la sección superior de la ventana, pulse el botón derecho del ratón y seleccione **Pegar columna** en el menú de atajo. La correlación de columnas de entrada a derivaciones especificadas debería tener este aspecto:

The screenshot shows a data integration tool interface. On the left, a table named 'Standardized' has columns: UniquelIdentifier, ApplicantSSN, Name, AddressLine1, AddressLine2, City, State, and Zip5. On the right, a 'Stage Variables' table is partially visible. Below it, a 'ToCopy' table is shown with columns: Constraint, Derivation, and Column Name. The 'ToCopy' table contains the following rows:

Constraint	Derivation	Column Name
Standardized.Uni		UniquelIdentifier
Standardized.App		ApplicantSSN
Standardized.Nar		Name
Standardized Adr		Addressl ine1

At the bottom, there are two tables. The left one is titled 'Standardized' and the right one is titled 'ToCopy'. Both have columns: Column name, Key, and SQL type.

Column name	Key	SQL type
UniquelIdentifier	<input checked="" type="checkbox"/>	VarChar
ApplicantSSN	<input type="checkbox"/>	VarChar
Name	<input type="checkbox"/>	VarChar

	Column name	Key	SQL type
1	UniquelIdentifier	<input checked="" type="checkbox"/>	VarChar
2	ApplicantSSN	<input type="checkbox"/>	VarChar
3	Name	<input type="checkbox"/>	VarChar

5. En la sección inferior derecha de la ventana, seleccione la fila superior, fila 1, del panel **ToCopy** y añada tres derivaciones y columnas a la etapa CreateAdditionalColumns:
  - a. Pulse el botón derecho del ratón en la fila y seleccione **Insertar fila** en el menú de atajo.
  - b. Añada dos filas más mediante el procedimiento descrito en el paso 5a.
  - c. Pulse el botón derecho del ratón en la fila insertada superior y seleccione **Editar fila** en el menú de atajo para abrir la ventana Editar metadatos de columna.
  - d. En el campo **Nombre de columna**, escriba MatchFirst1.
  - e. En la lista **Tipo de SQL**, seleccione **VarChar**.
  - f. En el campo **Longitud**, escriba 1.
  - g. En la lista **Con posibilidad de nullos**, seleccione **Sí**.
  - h. Pulse **Aplicar** y luego pulse **Cerrar** para cerrar la ventana.
  - i. En la página Paralelo, en la parte inferior de la ventana, seleccione **Ampliado (Unicode)**.
  - j. Pulse el botón derecho del ratón en la siguiente fila y seleccione **Editar fila** en el menú de atajo.
  - k. En el campo **Nombre de columna**, escriba HouseNumberFirstChar.
  - l. Repita los subpasos 5e a 5h.
  - m. Pulse el botón derecho del ratón en la última fila nueva y seleccione **Editar fila** en el menú de atajo.
  - n. En el campo **Nombre de columna**, escriba ZipCode3.
  - o. Repita los subpasos 5e a 5h, pero en el campo **Longitud**, seleccione 3.

La correlación de columnas de entrada a derivaciones especificadas debería tener este aspecto:

ToCopy					
Constraint:					
Derivation		Column Name			
		MatchFirst1			
		HouseNumberFirstCha			
		ZipCode3			

ToCopy					
	Column name	Key	SQL type	Extended	Length
1	MatchFirst1	<input type="checkbox"/>	VarChar	Unicode	1
2	HouseNumberFirs	<input type="checkbox"/>	VarChar	Unicode	1
3	ZipCode3	<input type="checkbox"/>	VarChar	Unicode	3

6. Añada derivaciones a las columnas:
  - a. Efectúe una doble pulsación en la celda que se encuentra en la columna **Derivación** y en la misma fila que la columna **MatchFirst1**, en la ventana ToCopy. Escriba la derivación: `if IsNull(Standardized.MatchFirstName_USNAME) then Setnull() Else Standardized.MatchFirstName_USNAME[1,1]`. Esta expresión detecta si la columna MatchFirstName contiene un nulo. Si la columna contiene un nulo, lo maneja. Si la columna contiene una serie, extrae el primer carácter y lo graba en la columna MatchFirst1.

ToCopy					
Constraint:					
Derivation		Column Name			
if isNull(Standardized.MatchFirstName_USNAME) then Setnull() Else Standardized.MatchFirstName_USNAME[1,1]		MatchFirst1			
		HouseNumberFirstChar			
		ZipCode3			
Standardize.UniqueIdentifier		<b>UniqueIdentifier</b>			

- b. Repita el subpaso a en la columna HouseNumberFirstChar y escriba el derivativo: `if IsNull(Standardized.HouseNumber_USADDR) then Setnull() Else Standardized.HouseNumber_USADDR[1,1]`.
  - c. Repita el subpaso a en la columna ZipCode3 y escriba el derivativo: `if IsNull(Standardized.ZipCode_USAREA) then Setnull() Else Standardized.ZipCode_USAREA[1,3]`.
7. Correlacione las tres derivaciones y las columnas a las columnas de entrada.
  - a. Vaya al panel superior izquierdo y desplace el panel **Standardized** hasta que localice MatchFirstName\_USNAME.
  - b. Pulse y arrastre la celda al panel **ToCopy** y a la celda que contiene Standardized.MatchFirstName\_USNAME.
  - c. Cuando se le solicite que sustituya los datos existentes, pulse **Sí**.
  - d. Repita los pasos del a al c para HouseNumber\_USADDR y ZipCode\_USAREA, emparejando los nombres de columna del panel **Standardized** con las columnas de nombre similar en el panel **ToCopy**.



- e. Pulse **Aceptar** para cerrar la ventana Etapa Transformer.

### Configurar la etapa Copy

La etapa Copy duplica los datos y los graba en más de un enlace de salida. En esta lección, la etapa Copy duplica los metadatos de la etapa Transformer y los graba en la etapa Match Frequency y en el archivo de destino.

Los metadatos de las etapas Standardize y Transformer se duplican y se graban en dos enlaces de salida.

Para configurar la etapa Copy:

1. Efectúe una doble pulsación en el icono de la etapa Copy para abrir la ventana Etapa Copy.
2. Pulse el separador **Salida > Correlación**.
3. Copie los datos del enlace de salida **StandardizedData**:
  - a. En el campo **Nombre salida** sobre el panel **Columnas** seleccione **StandardizedData**.
  - b. Pulse el botón derecho del ratón en el panel **Columnas** y seleccione **Seleccionar todo** en el menú de atajo.
  - c. Pulse el botón derecho del ratón y seleccione **Copiar** en el menú de atajo.
  - d. Vaya al panel **StandardizedData**, pulse el botón derecho del ratón y seleccione **Pegar columna** en el menú de atajo.
4. Para copiar los datos en el enlace de salida **ToMatchFrequency**, seleccione **ToMatchFrequency** en el campo **Nombre de salida** sobre el panel **Columnas** y repita los pasos del b al d, copiando los datos al panel **ToMatchFrequency**.
5. Pulse **Aceptar** para copiar los datos y cerrar la etapa Copy.

### Configurar la etapa Match Frequency

La etapa Match Frequency genera información de distribución de frecuencia analizando los datos que se utilizan para realizar los emparejamientos.

La etapa Match Frequency procesa datos de frecuencia independientemente de ejecutar una coincidencia. El enlace de salida de esta etapa contiene cuatro columnas:

- qsFreqVal
- qsFreqCounts
- qsFreqColumnID
- qsFreqHeaderFlag

Para configurar la etapa Match Frequency:

1. Efectúe una doble pulsación en el icono de etapa Match Frequency para abrir la ventana Etapa Match Frequency.
2. Seleccione el recuadro de selección **No utilizar una especificación de coincidencia**. En este momento usted desconoce qué columnas se utilizan en la especificación de coincidencia.
3. Pulse el separador **Propiedades de etapa**.
4. Pulse el separador **Salida > Correlación**.
  - a. En el campo **Nombre de salida** seleccione **Frecuencias**.
  - b. Pulse el botón derecho del ratón en el panel **Columnas** y seleccione **Seleccionar todo** en el menú de atajo.
  - c. Pulse el botón derecho del ratón y seleccione **Copiar** en el menú de atajo.

- d. Vaya al panel **Frecuencias**, pulse el botón derecho del ratón y seleccione **Pegar column** en el menú de atajo.
5. Guarde las definiciones de tabla en la carpeta **Definiciones de tabla**.
  - a. Pulse el separador **Columnas**.
  - b. Pulse **Guardar**. Se abre la ventana Guardar definición de tabla.
  - c. En el campo **Tipo de origen de datos**, introduzca Definiciones de tabla.
  - d. En el campo **Nombre del origen de datos**, escriba ToFrequencies.
  - e. En el campo **Nombre de tabla/archivo**, escriba ToFrequencies.
  - f. Pulse **Aceptar** para abrir la ventana Guardar definición de tabla como.
  - g. Guarde las definiciones de tabla de Standardization en la carpeta Definición de tabla que está en un nivel inferior a la carpeta del proyecto, por ejemplo, **Guía de aprendizaje de QualityStage > Definiciones de tabla**.
  - h. Pulse **Guardar**.
  - i. Confirme los cambios y salga de las ventanas.
6. Pulse **Aceptar** para cerrar el separador **Salida > Columna** y la etapa Match Frequency.
7. Pulse **Aceptar** para cerrar la etapa.

### Punto de comprobación de la lección

Esta lección le ha explicado cómo configurar el archivo de origen y todas las etapas del trabajo Standardize.

Usted ha aplicado los valores a cada etapa y ha correlacionado los archivos de salida con la siguiente etapa del trabajo Standardize. Ha aprendido a realizar las siguientes tareas:

- Configurar el archivo de origen para cargar los datos y metadatos de clientes
- Aplicar conjuntos de reglas conformes al servicio postal de los Estados Unidos a los datos de nombres y direcciones de clientes
- Añadir columnas adicionales para seleccionar coincidencias y crear derivaciones para manejar nulos
- Grabar datos en dos enlaces de salida y asociar los datos a los enlaces correctos
- Cree datos de frecuencia

## Lección 2.3: Configurar los conjuntos de datos de destino


Los dos conjuntos de datos de destino del trabajo Standardize almacenan los datos estandarizados y de frecuencia que puede utilizar como datos de origen en el trabajo One-source Match.



Complete las siguientes tareas para configurar los conjuntos de datos de destino:

- Adjunte el archivo al conjunto de datos de destino Stan
- Adjunte el archivo al conjunto de datos Frecuencias

Para configurar los conjuntos de datos de destino:

1. Efectúe una doble pulsación en el icono de conjunto de datos de salida **Stan** para abrir la ventana Conjunto de datos.
2. Pulse **Entrada > Propiedades** y seleccione **Destino > Archivo**.

3. Pulse  y vaya a la carpeta en el sistema servidor donde reside el archivo de datos de entrada (por ejemplo, input.csv).

4. En el campo **Nombre de archivo**, escriba Stan y luego pulse **Aceptar** para visualizar la vía de acceso y el nombre del archivo en el campo **Archivo**, (por ejemplo, C:\IBM\InformationServer\Server\Projects\tutorial\Stan).
5. Guarde las definiciones de tabla en la carpeta **Definiciones de tabla**.
  - a. Pulse el separador **Columnas**.
  - b. Pulse **Guardar**. Se abre la ventana Guardar definición de tabla.
  - c. En el campo **Tipo de origen de datos**, introduzca Definiciones de tabla.
  - d. En el campo **Nombre del origen de datos**, escriba StandardizedData1.
  - e. En el campo **Nombre de archivo/tabla**, escriba StandardizedData1.
  - f. Pulse **Aceptar** para abrir la ventana Guardar definición de tabla como.
  - g. Guarde las definiciones de tabla en la carpeta Definición de tabla que está en un nivel inferior a la carpeta del proyecto, por ejemplo, **Guía de aprendizaje de QualityStage > Definiciones de tabla**.
  - h. Confirme los cambios y salga de las ventanas.
6. Efectúe una doble pulsación en el icono de conjunto de datos de destino **Frequencies**.
7. Repita los pasos del 2 al 5 con el archivo Frequencies, pero sustituya Stan por Frequencies y sustituya StandardizedData1 por ToFrequencies1 en los campos correspondientes. El archivo Stan y el archivo Frequencies son los conjuntos de datos de origen para el trabajo One-source Match.
8. Pulse **Archivo > Guardar** para guardar el trabajo Standardize.
9. Pulse  para compilar el trabajo en el cliente del Diseñador.
10. Pulse  para ejecutar el trabajo.

El trabajo estandariza los datos de acuerdo con las reglas aplicadas y añade columnas de coincidencia adicionales a los metadatos. Los datos se graban en dos conjuntos de datos de destino, que sirven como archivos de origen para un trabajo posterior.

### **Punto de comprobación de la lección**

Esta lección le ha explicado cómo adjuntar archivos a conjuntos de datos de destino para almacenar los datos de nombres y direcciones de clientes estandarizados procesados y los datos de frecuencia.

Usted ha configurado los archivos de conjuntos de datos de destino Stan y Frequencies para aceptar los datos cuando se procesan.

## **Módulo 2: Resumen**

En el Módulo 2 usted ha establecido y configurado un trabajo Standardize.

La ejecución de un trabajo de Standardize condiciona los datos a fin de garantizar que todos los datos de nombres y direcciones de clientes incluyen el mismo contenido y están en el mismo formato. El trabajo de Standardize carga los datos de origen de nombres y direcciones almacenados en la base de datos del banco y añade definiciones de tabla a fin de organizar los datos en un formato que pueda ser analizado por el conjunto de reglas. El proceso de la etapa Transformer aumenta el número de columnas y se generan datos de frecuencia para su entrada en el trabajo de coincidencia.

## Lecciones aprendidas

Una vez completado este módulo, usted habrá aprendido los siguientes conceptos y tareas:

- Cómo generar datos estandarizados de manera que los registros coincidan correctamente
- Cómo ejecutar las etapas de IBM InfoSphere DataStage y Data Quality conjuntamente en un trabajo
- Cómo aplicar conjuntos de reglas de país o región para analizar los datos de direcciones
- Cómo utilizar derivaciones para manejar nulos
- Cómo generar datos que puedan utilizarse luego como datos de origen en un trabajo posterior

---

## Módulo 3: Agrupar registros con atributos comunes

En este módulo se explica cómo configurar y ejecutar un trabajo One-source Match utilizando datos estandarizados y datos de frecuencia generados para buscar registros coincidentes y eliminar registros duplicados.

La etapa One-source Match es una de las dos etapas que busca registros coincidentes al tiempo que elimina registros duplicados y no coincidentes. La otra etapa de coincidencia es la etapa Two-source Match.

La etapa One-source Match agrupa los registros que comparten atributos comunes. La especificación de coincidencia que debe aplicar ha sido configurada para separar todos los registros con un peso superior a un límite de coincidencia concreto como duplicados. Para identificar el registro maestro debe seleccionar el registro del conjunto que coincida consigo mismo con el peso más elevado.

Todos los registros que no forman parte de un conjunto de duplicados son registros no coincidentes. Estos registros y los registros maestros se utilizan en el siguiente pase. No debe incluir duplicados, ya que es importante que pertenezcan a un único conjunto.

La utilización de una etapa de coincidencia asegura la integridad de los datos, ya que se aplica tecnología de coincidencia probabilística. Esta tecnología se aplica a los atributos relevantes para evaluar columnas, componentes de columnas o caracteres individuales definidos por usted. Además, puede aplicar pesos de acuerdo o desacuerdo a elementos de datos clave.

### Objetivos de aprendizaje

Una vez completadas las lecciones de este módulo, deberá saber cómo llevar a cabo las siguientes tareas:

- Añadir enlaces y etapas de IBM InfoSphere DataStage a un trabajo
- Utilizar datos estandarizados y datos de frecuencia como los archivos de origen
- Configurar propiedades de etapa para especificar qué acción realizan las etapas cuando se ejecuta el trabajo
- Eliminar las direcciones duplicadas después del primer pase
- Aplicar una especificación de coincidencia para determinar cómo se seleccionan las coincidencias
- Canalizar los datos de atributos comunes a un archivo de destino independiente

El tiempo para completar este módulo es de 30 minutos aproximadamente.

### Lección 3.1: Configurar un trabajo One-source Match

La recopilación de registros en grupos con atributos relacionados es el próximo paso de la limpieza de datos. En esta lección va a añadir la etapa One-source Match de Calidad de Datos y una etapa Funnel para buscar registros coincidentes y eliminar los duplicados.

Si todavía no lo ha hecho, abra el cliente de IBM InfoSphere DataStage and QualityStage Designer.

Tal como ha aprendido en el módulo anterior, debe añadir etapas y enlaces al lienzo del Diseñador para crear un trabajo One-source Match. El trabajo Standardize que acaba de completar ha creado un conjunto de datos **Stan** y un

conjunto de datos **Frecuencias**. La información de estos conjuntos de datos se utiliza como los datos de entrada al diseñar el trabajo One-source Match.

Para configurar un trabajo One-source Match:

1. En el panel izquierdo del Diseñador, vaya a la carpeta MyTutorial que ha creado para esta guía de aprendizaje y efectúe una doble pulsación en OneSource1 para abrir el trabajo.
2. Arrastre los siguientes iconos al lienzo del Diseñador desde la paleta.
  - El icono **Calidad de datos > One-source Match** al centro del lienzo.
  - El icono **Archivo > Conjunto de datos** a la parte superior izquierda del icono **One-source Match**.
  - Un segundo icono **Archivo > Conjunto de datos** a la parte inferior izquierda del icono **One-source Match**.
  - El icono **Proceso > Funnel** a la parte superior derecha del icono **One-source Match**.
  - Tres iconos **Archivo > Archivo secuencial**, uno a la derecha de la etapa Funnel y los otros dos a la derecha de la etapa One-source Match.
3. Pulse el botón derecho del ratón en el icono **Conjunto de datos** superior y arrástrelo para crear un enlace desde este conjunto de datos a la etapa One-source Match.

**Nota:** El orden en que cree los enlaces afectará a la ejecución satisfactoria del trabajo. Más adelante en esta guía de aprendizaje, modificará propiedades de etapas para cambiar el orden de algunos de los enlaces.

4. Arrastre los enlaces a las etapas restantes. Arrastre dos enlaces desde la etapa One-source Match a la etapa Funnel.
5. Pulse en los nombres de las siguientes etapas y escriba el nuevo nombre de etapa en el recuadro resaltado:

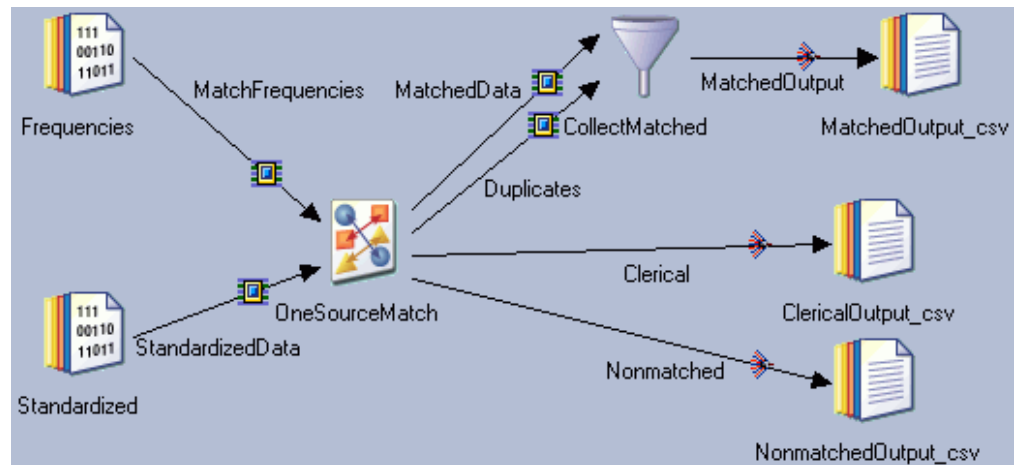
Etapa	Cambiar a
Data Set de la parte superior izquierda	Frecuencias
Data Set de la parte inferior izquierda	Standardized
One-source Match	OneSourceMatch
Funnel	CollectMatched
Sequential File de la parte superior derecha	MatchedOutput_csv
Sequential File de la parte central derecha	ClericalOutput_csv
Sequential File de la parte inferior derecha	NonMatchedOutput_csv

6. Pulse el botón derecho del ratón en los nombres de los siguientes enlaces, seleccione **Renombrar** en el menú de atajo y escriba el nuevo nombre de enlace en el recuadro resaltado:

Enlaces	Cambiar a
De Frecuencias a OneSourceMatch	MatchFrecuencias
De Standardized a OneSourceMatch	StandardizedData
De OneSourceMatch a CollectMatched	MatchedData
De OneSourceMatch a CollectMatched	Duplicates
De CollectMatched a MatchOutput_csv	MatchedOutput
De OneSourceMatch a ClericalOutput_csv	Clerical

Enlaces	Cambiar a
De OneSourceMatch a NonMatchedOutput_csv	NonMatched

7. Pulse **Archivo > Guardar** para guardar el trabajo.



### Punto de comprobación de la lección para el trabajo One-source Match

En esta lección ha aprendido a configurar un trabajo One-source Match. Durante el proceso de este trabajo, los registros coincidentes se buscan utilizando la especificación de coincidencia creada para esta guía de aprendizaje. Posteriormente, los registros se ordenan de acuerdo con sus atributos y se graban en varios enlaces de salida.

Ha configurado y enlazado un trabajo One-source Match mediante las siguientes tareas:

- Añadir etapas de Calidad de datos y de Proceso al lienzo del Diseñador
- Enlazar todas las etapas
- Renombrar los enlaces y las etapas con los nombres adecuados

## Lección 3.2: Configurar las propiedades de etapa del trabajo One-source Match


Configure las propiedades para cada etapa del trabajo One-source Match en el lienzo del Diseñador.

Realice las siguientes tareas para configurar el trabajo One-source Match:

- Cargue los datos y metadatos de los dos archivos de origen
- Aplique una especificación de coincidencia al trabajo One-source Match y seleccione enlaces de salida
- Combine los registros sin ordenar

Para configurar los conjuntos de datos Frequencies y Standardized:

1. Efectúe una doble pulsación en el icono del conjunto de datos **Frequencies** para abrir el separador **Propiedades** en la ventana Frequencies - Conjunto de datos.
2. Pulse **Archivo > Origen**.


3. En el campo **Archivo**, pulse  y vaya hasta la vía de acceso de la carpeta en el sistema servidor donde reside el archivo de datos de entrada.
4. En el campo **Nombre de archivo**, escriba *Frecuencias*. (Por ejemplo, C:\IBM\InformationServer\Server\Projects\tutorial\Frecuencias).
5. Pulse **Aceptar** para cerrar la ventana.
6. Pulse el separador **Columnas** y pulse **Cargar**. Se abrirá la ventana Definiciones de tabla.
7. Seleccione el archivo *Carpeta\_proyecto* > **Definiciones de tabla** > **ToFrecuencias1** y pulse **Aceptar**.
8. Confirme los cambios y salga de las ventanas. Las definiciones de tabla se cargan en el separador **Columnas** del archivo de origen.
9. Efectúe una doble pulsación en el icono del conjunto de datos **Standardized**.
10. Repita los pasos del 2 al 9, pero escriba *Stan* en el paso 4 y seleccione el archivo **StandardizedData1** en el paso 7.

Los datos del trabajo Standardize se cargan en los archivos de origen del trabajo One-source Match.

### Configurar la etapa One-source Match

La etapa One-source Match agrupa los registros con atributos comunes.

Para configurar la etapa One-source Match:

1. Efectúe una doble pulsación en el icono de la etapa One-source Match.
2. Pulse el botón  de Especificación de coincidencia.
3. En la ventana Repositorio, expanda carpetas hasta localizar la carpeta **NameandAddress**.
4. Pulse el botón derecho del ratón en la especificación de coincidencia **NameAndAddress** y seleccione **Suministrar todo** en el menú de atajo.
5. Pulse **Aceptar** para adjuntar la especificación de coincidencia de un origen **NameAndAddress** para la guía de aprendizaje.
6. Pulse los recuadros de selección para las siguientes opciones de **Salida de coincidencia**:
  - Coincidencia - Envía los registros coincidentes como datos de salida.
  - Administrativo - Separa los registros que requieren una revisión administrativa.
  - Duplicado - Incluye los registros duplicados que están por encima del límite de coincidencia.
  - No coincidente - separa los registros que no sean coincidentes, duplicados ni administrativos como registros no coincidentes.
7. Mantenga la configuración predeterminada **Dependiente** en el panel **Tipo de coincidencia**. Después de ejecutar este primer pase se eliminan los duplicados con cada pase adicional.
8. Pulse el separador **Propiedades de etapa** > **Ordenación de enlaces**. Asegúrese de que los enlaces de entrada y salida aparecen en el siguiente orden. Si es necesario, utilice los botones de flecha arriba y abajo para poner los enlaces en el orden correcto.



Etiqueta del enlace de entrada	Nombre de enlace
Datos	StandardizedData
FrecDatos	MatchFrequencies

Etiqueta del enlace de salida	Nombre de enlace
Coincidencia	MatchedData
Administrativo	Administrativo
Duplicado	Duplicates
No coincidente	NonMatched

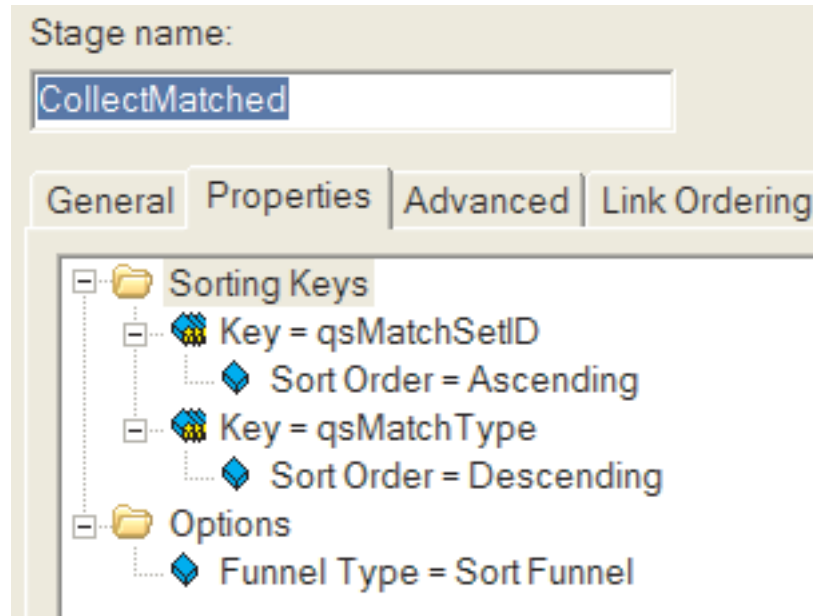
9. Pulse el separador **Salida > Correlación** y correlacione las siguientes columnas con los enlaces correctos:
  - a. En la lista **Nombre de salida** encima del panel **Columnas**, seleccione **MatchedData**.
  - b. Pulse el botón derecho del ratón en el panel **Columnas** y seleccione **Seleccionar todo** en el menú de atajo.
  - c. Pulse el botón derecho del ratón y seleccione **Copiar** en el menú de atajo.
  - d. Vaya al panel **MatchedData**, pulse el botón derecho del ratón y seleccione **Pegar columna** en el menú de atajo.
  - e. Seleccione **Duplicates** en la lista **Nombre de salida** encima del panel **Columnas**.
  - f. Repita los pasos 9b a 9d para los datos de Duplicates.
  - g. Seleccione **Clerical** en la lista **Nombre de salida** encima del panel **Columnas**.
  - h. Repita los pasos 9b a 9d para los datos de Clerical.
  - i. Seleccione **NonMatched** en la lista **Nombre de salida** encima del panel **Columnas**.
  - j. Repita los pasos 9b a 9d para los datos de Nonmatched.
10. Pulse **Aceptar** para cerrar la ventana Propiedades de etapa.
11. Pulse **Aceptar** para cerrar la etapa.

## Configurar la etapa Funnel

La etapa Funnel combina registros cuando estos se reciben en un formato sin ordenar.

Para configurar una etapa Funnel:

1. Efectúe una doble pulsación en el icono de la etapa CollectMatched y pulse el separador **Etapa > Propiedades**.
2. En el árbol **Opciones** seleccione **Tipo de canalización**.
3. En la lista **Tipo de canalización**, seleccione **Canalización ordenada**.
4. Pulse **Claves de ordenación > Clave**, y luego seleccione **qsMatchSetID** en la lista **Clave**. El orden de clasificación predeterminado es **Ascendente**.
5. Vuelva a pulsar **Claves de ordenación**.
6. En el campo **Propiedades disponibles para añadir**, pulse **Clave**.
7. En la lista **Clave**, seleccione **qsMatchType**.
8. Pulse **Orden de clasificación** y luego seleccione **Descendente** en la lista **Orden de clasificación**.



9. Pulse el separador **Salida > Correlación**.
10. Pulse el botón derecho del ratón en el panel **Columns** y seleccione **Seleccionar todo** en el menú de atajo.
11. Pulse el botón derecho del ratón y seleccione **Copiar** en el menú de atajo.
12. Vaya a la columna **MatchedOutput**, pulse el botón derecho del ratón y seleccione **Pegar columna** en el menú de atajo.
13. Pulse **Aceptar** para cerrar la ventana de la etapa.

### Puntos clave de la lección 3.2

En la Lección 3.2, ha configurado los archivos de origen y las etapas del trabajo One-source Match.

Ha aprendido a realizar las siguientes tareas:

- Cargar los datos y metadatos generados en un trabajo anterior
- Aplicar una especificación de coincidencia para procesar los datos e identificar coincidencias y duplicados
- Combinar registros en un único archivo

## Lección 3.3: Configurar los archivos de destino del trabajo One-source Match

Para configurar los archivos de destino para el trabajo One-source Match, debe adjuntar archivos a los cuatro registros de salida. Los registros del archivo MatchedOutput se convierten en los registros de origen para el siguiente trabajo.

Para configurar los archivos de destino:

1. Efectúe una doble pulsación en el icono **MatchedOutput\_csv** para abrir el separador **Propiedades** en la ventana MatchedOutput\_csv - Archivo secuencial. Va a adjuntar un nombre de archivo a los registros coincidentes.
2. Pulse **Destino > Archivo**.

3. Junto al campo **Archivo**, pulse  y vaya a la carpeta en el sistema servidor donde reside en el archivo de datos de entrada.

4. En el campo **Nombre de archivo**, escriba MatchedOutput.csv para visualizar la vía de acceso y el nombre de archivo en el campo **Archivo**, (por ejemplo, C:\IBM\InformationServer\Server\Projects\tutorial\MatchedOutput.csv).
5. Pulse **Opciones > La primera línea contiene los nombres de las columnas** y cambie el valor a **True**.
6. Pulse el separador **Formato**.
7. Pulse el botón derecho del ratón en **Valores predeterminados de los campos**, y luego pulse **Añadir subpropiedad > Valor de campo nulo**.
8. Escriba "" en el campo **Valor de campo nulo**. El valor de campo nulo es un par de comillas dobles sin ningún espacio entre ellas.
9. Especifique la forma en que la etapa recopila los datos antes de grabarlos en el archivo secuencial. Debe especificar detalles de recopilación porque la etapa anterior se ejecuta en modalidad paralela y la etapa Sequential File se ejecuta en modalidad secuencial.
  - a. Pulse el separador **Particionamiento**.
  - b. En la lista **Tipo de recopilador**, seleccione **Ordenado**. Este método lee todas las filas de la primera partición, luego todas las filas de la segunda partición, y así sucesivamente.
  - c. En la sección **Ordenación**, seleccione **Ordenar**.
  - d. En las columnas disponibles, pulse **qsMatchDataID**.
  - e. Pulse **Aceptar**.
10. Guarde las definiciones de tabla en la carpeta **Definiciones de tabla**.
  - a. Pulse el separador **Columnas**.
  - b. Pulse **Guardar**. Se abre la ventana Guardar definición de tabla.
  - c. En el campo **Tipo de origen de datos**, introduzca Definiciones de tabla.
  - d. En el campo **Nombre del origen de datos**, escriba MatchedOutput1.
  - e. En el campo **Nombre de tabla/archivo**, escriba MatchedOutput1.
  - f. Pulse **Aceptar** para abrir la ventana Guardar definición de tabla como.
  - g. Guarde las definiciones de tabla de Standardization en la carpeta Definición de tabla que está en un nivel inferior a la carpeta del proyecto, por ejemplo, **Guía de aprendizaje de QualityStage > Definiciones de tabla**.
  - h. Confirme los cambios y salga de las ventanas.
11. Repita los pasos 1 en la página 36 a 10 para cada una de las siguientes etapas:
  - Para la etapa ClericalOutput\_csv, escriba ClericalOutput.csv y Clerical1 en los campos correspondientes.
  - Para la etapa NonMatchedOutput\_csv, escriba NonMatchedOutput.csv y NonMatched1 en los campos correspondientes.
12. Pulse **Archivo > Guardar** para guardar el trabajo.



13. Pulse  para compilar el trabajo en el cliente de IBM InfoSphere DataStage and QualityStage Designer.
14. Pulse **Herramientas > Ejecutar Director** para abrir el Director de IBM InfoSphere DataStage and QualityStage Director. El Director se abre con el trabajo One-source Match visible en la ventana del Director con el estado Compilado.



15. Pulse  .

Ha configurado los archivos de destino y luego ha compilado y ejecutado el trabajo.

### **Punto de comprobación de la lección**

En esta lección usted ha combinado los registros de direcciones coincidentes y duplicados en un archivo. Los registros de salida no coincidentes y administrativos se han separado en archivos individuales. Los registros de salida administrativos se pueden revisar manualmente para encontrar registros coincidentes. Los registros no coincidentes se utilizan en el siguiente pase. Los registros de direcciones coincidentes y duplicados se utilizan en el trabajo Survive.

Ha aprendido a separar los registros de salida de la etapa One-source Match en varios archivos de destino.

## **Módulo 3: Resumen**

En el Módulo 3 ha establecido y configurado un trabajo utilizando la etapa One-source Match para consolidar los datos de nombres y direcciones coincidentes y duplicados en un archivo.

Al crear un trabajo de etapa One-source Match, usted ha añadido una especificación de coincidencia para aplicar los criterios de bloqueo y coincidencia a los datos estandarizados y de frecuencia creados en el trabajo Standardize. Después de aplicar la especificación de coincidencia, los registros resultantes se enviaron a través de cuatro enlaces de salida, uno para cada tipo de registro. Las coincidencias y duplicados se enviaron a una etapa Funnel que combinó los registros en una salida, que se grabó en un archivo. Los registros no coincidentes se enviaron a un archivo, al igual que los registros de salida administrativos.

### **Lecciones aprendidas**

Una vez completado el Módulo 3, usted habrá aprendido los siguientes conceptos y tareas:

- Cómo aplicar una especificación de coincidencia a la etapa One-source Match
- Cómo agrupa la etapa One-source Match los registros con atributos similares
- Cómo asegurar la integridad de los datos aplicando la tecnología de coincidencia probabilística

---

## Módulo 4: Crear un registro único

Este módulo diseña un trabajo Survive para aislar el mejor registro de nombre y dirección de cada cliente.

El trabajo One-source Match identifica un grupo de registros con atributos similares. En el trabajo Survive usted debe especificar qué columnas y valores de columna de cada grupo crean el registro de salida del grupo. El registro de salida puede incluir la siguiente información:

- Un registro de entrada completo
- Las columnas seleccionadas del registro
- Las columnas seleccionadas de distintos registros del grupo

Seleccione los valores de columna en función de las reglas para probar las columnas. Una regla incluye un conjunto de condiciones y una lista de destinos. Si una columna resulta verdadera frente a las condiciones, el valor de la columna de ese registro se convierte en el mejor candidato para el destino. Tras probar cada registro del grupo, se combinan las columnas declaradas como mejores candidatas para convertirse en el registro de salida del grupo. La supervivencia de la columna está determinada por el destino. La supervivencia del valor de la columna está determinada por las reglas.

### Objetivos de aprendizaje

Una vez completadas las lecciones de este módulo, deberá saber cómo llevar a cabo las siguientes tareas:

- Añadir etapas y enlaces a un trabajo Survive
- Elegir la columna seleccionada
- Añadir las reglas
- Correlacionar las columnas de salida

El tiempo para completar este módulo es de 20 minutos aproximadamente.

## Lección 4.1: Configurar un trabajo Survive

Crear el mejor registro de resultados de la etapa Survive es el último trabajo del proceso de limpieza de datos. El mejor registro de resultados es el nombre y la dirección con mayor probabilidad de exactitud para cada cliente del banco.

En esta lección añadirá la etapa Survive de Calidad de datos, el archivo de origen de datos combinados del trabajo One-source Match y el archivo de destino para los mejores registros.

Para configurar un trabajo Survive:

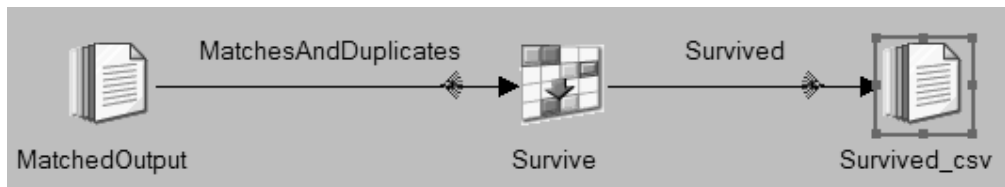
1. En el panel izquierdo del cliente de IBM InfoSphere DataStage and QualityStage Designer, vaya a la carpeta MyTutorial que ha creado para esta guía de aprendizaje y efectúe una doble pulsación en Survive1 para abrir el trabajo.
2. Arrastre los siguientes iconos al lienzo del Diseñador desde la paleta:
  - El icono **Calidad de datos** > **Survive** al centro del lienzo
  - El icono **Archivo** > **Archivo secuencial** a la izquierda de la etapa Survive
  - El segundo icono **Archivo** > **Archivo secuencial** a la derecha de la etapa Survive

- Pulse con el botón derecho del ratón el icono **Archivo secuencial** de la izquierda y arrastre un enlace a la etapa Survive.
- Arrastre un segundo enlace desde la etapa Survive al icono de salida **Archivo secuencial**.
- Pulse en los nombres de las siguientes etapas y escriba el nuevo nombre de etapa en el recuadro resaltado:

Etapa	Cambiar a
Archivo secuencial de la izquierda	MatchedOutput
etapa Survive	Survive
Archivo secuencial de la derecha	Survived_csv

- Pulse el botón derecho del ratón en los nombres de los siguientes enlaces, seleccione **Renombrar** en el menú de atajo y escriba el nuevo nombre de enlace en el recuadro resaltado:

Enlaces	Cambiar a
De MatchedOutput a Survive	MatchesAndDuplicates
De Survive a Survived_csv	Survived



### Punto de comprobación de la lección

En esta lección ha aprendido a configurar un trabajo Survive añadiendo como datos de origen los resultados del trabajo One-source Match, la etapa Survive y el archivo de destino como el registro de salida para el grupo.

Además, ha aprendido que la etapa Survive tiene un enlace de entrada y un enlace de salida.


## Lección 4.2: Configurar las propiedades de etapa del trabajo Survive

Para configurar las propiedades de etapa del trabajo Survive, debe cargar los datos coincidentes y duplicados del trabajo One-source Match, configurar la etapa Survive con reglas que prueben columnas para un conjunto de condiciones y configurar el archivo de destino.

En el trabajo Survive probará los valores de columnas para determinar qué columnas son las mejores candidatas para el registro en cuestión. Estas columnas se combinan para convertirse en el registro de salida del grupo. Al seleccionar un mejor candidato puede especificar que se prueben estos valores de columnas:

- Datos de creación del registro
- Origen de datos
- Longitud de los datos de una columna
- Frecuencia de datos en un grupo

Para configurar el archivo de origen:

1. Efectúe una doble pulsación en el icono **Archivo MatchedOutput** para acceder a la página **Propiedades**.
2. Pulse **Archivo > Origen**.
3. En el campo **Archivo**, pulse  y vaya hasta la vía de acceso de la carpeta en el sistema servidor donde reside el archivo de datos de entrada.
4. En el campo **Nombre de archivo**, escriba MatchedOutput.csv para visualizar la vía de acceso y el nombre de archivo en el campo **Archivo**, (por ejemplo, C:\IBM\InformationServer\Server\Projects\tutorial\MatchedOutput.csv).
5. Pulse **Opciones > La primera línea contiene los nombres de las columnas** y cambie el valor a **True**.
6. Pulse el separador **Formato**.
7. Pulse el botón derecho del ratón en **Valores predeterminados de los campos**, y luego pulse **Añadir subpropiedad > Valor de campo nulo**.
8. Escriba "" en el campo **Valor de campo nulo**. El valor de campo nulo es un par de comillas dobles sin ningún espacio entre ellas.
9. Pulse el separador **Columnas** y pulse **Cargar**. Se abrirá la ventana Definiciones de tabla.
10. Pulse el archivo **Carpeta\_proyecto > Definiciones de tabla > MatchedOutput1**. Las definiciones de tabla se cargan en el separador **Columnas** del archivo de origen.
11. Confirme los cambios y salga de las ventanas.

Usted ha adjuntado el archivo MatchedOutput.csv y ha cargado las Definiciones de tabla al archivo MatchedOutput.

## Configurar la etapa Survive

Configure la etapa Survive con reglas para comparar las columnas en el mejor formato.

Para configurar la etapa Survive:

1. Efectúe una doble pulsación en el icono de la etapa Survive.
2. Pulse **Nueva regla** para abrir la ventana Definición de reglas Survive. La etapa Survive requiere una regla que contenga uno o más destinos y una expresión de condición TRUE.

Las reglas se definen especificando los siguientes elementos:

- Columna o columnas de destino
- Columna a analizar
- Técnica a aplicar a la columna que se está analizando

3. En el panel **Columnas disponibles**, seleccione **Todas las columnas** y pulse



para mover **Todas las columnas** al panel **Destino**. Al seleccionar Todas las columnas está asignando el primer registro del grupo como el mejor registro.

4. En la sección **Regla de supervivencia** de la ventana, seleccione **qsMatchType** en la lista **Analizar columna**. Está seleccionando qsMatchType como destino con el que comparar el resto de columnas.
5. En la lista **Técnica**, seleccione **Igual a**.

6. En el campo **Datos**, escriba MP. MP significa Par de coincidencia para la etapa One-source Match.
7. Pulse **Aceptar** para cerrar la ventana Definición de reglas de supervivencia.
8. Repita los pasos del 2 en la página 41 al 5 en la página 41 para añadir las siguientes columnas y reglas. No especifique valores en el campo **Datos**.

Especificar columnas de salida > Destinos	Analizar Columna	Técnica
GenderCode_USNAME	GenderCode_USNAME	La más frecuente (no en blanco)
FirstName_USNAME	FirstName_USNAME	La más frecuente (no en blanco)
MiddleName_USNAME	MiddleName_USNAME	La más larga
PrimaryName_USNAME	PrimaryName_USNAME	La más frecuente (no en blanco)

Puede ver las reglas que ha añadido en la cuadrícula Survive.

9. En la sección **Seleccionar la columna de datos de identificación de grupo**, seleccione la columna qsMatchSetID.
10. Pulse **Propiedades de etapa**, y luego pulse el separador **Salida > Correlación**.
11. Pulse el botón derecho del ratón en el panel **Columnas** y seleccione **Seleccionar todo** en el menú de atajo.
12. Seleccione **Copiar** en el menú de atajo.
13. Vaya al panel **Survived**, pulse el botón derecho del ratón y seleccione **Pegar columna** en el menú de atajo.
14. Confirme los cambios y salga de las ventanas.

Target(s):	Analyze Column:	Technique:	Data
<AllColumns>	qsMatchType	Equals	"MP"
GenderCode	GenderCode	Most Frequent	
FirstName	FirstName	Most Frequent	
MiddleName	MiddleName	Longest	
PrimaryName	PrimaryName	Most Frequent	



Column Name	Description
qsMatchLRFlag	Match output: left / right flag (set by eg double interval c
qsMatchPassNumber	Match output: pass number
qsMatchPattern	Match output: pattern
qsMatchSetID	Match output: set id (ie rec id of group's master)
qsMatchType	Match output: type code, eg XA, MP etc.

Selected Column      **qsMatchSetID**

## Configurar el archivo de destino

Va a configurar el archivo de destino de la etapa Survive.



1. Efectúe una doble pulsación en el icono del archivo de destino **Survived\_csv** y pulse **Destino > Archivo** para activar el campo **Archivo**.
2. En el campo **Archivo** pulse  y vaya a la carpeta en el sistema servidor donde reside el archivo de datos de entrada.
3. En el campo **Nombre de archivo**, escriba `record.csv` para visualizar la vía de acceso y el nombre del archivo en el campo **Archivo**, (por ejemplo, `C:\IBM\InformationServer\Server\Projects\tutorial\record.csv`).
4. Pulse **Opciones > La primera línea contiene los nombres de las columnas** y cambie el valor a **True**.
5. Pulse el separador **Formato**.
6. Pulse el botón derecho del ratón en **Valores predeterminados de los campos**, y luego pulse **Añadir subpropiedad > Valor de campo nulo**.
7. Escriba `" "` en el campo **Valor de campo nulo**. El valor de campo nulo es un par de comillas dobles sin ningún espacio entre ellas.
8. Confirme los cambios y salga de las ventanas.
9. Pulse **Archivo > Guardar** para guardar el trabajo.
10. Pulse  para compilar el trabajo en el cliente del Diseñador.
11. Pulse **Herramientas > Ejecutar Director** para abrir el Director de DataStage. Se abrirá el Director mostrando el trabajo Standardize en la ventana del Director con el estado Compilado.
12. Pulse **Ejecutar**.

### Punto de comprobación de la lección

Usted ha configurado el trabajo Survive, renombrado los enlaces y las etapas, y configurado los archivos de origen y destino, y la etapa Survive.

En la Lección 4.2, ha aprendido a especificar reglas simples que luego se aplican a una columna seleccionada. Esta combinación se compara más tarde con todas las columnas para encontrar el mejor registro.

## Módulo 4: Resumen

En el Módulo 4 ha completado el último trabajo del flujo de trabajo de IBM InfoSphere QualityStage. En este módulo usted ha establecido y configurado el trabajo Survive para seleccionar el mejor registro de los datos de nombres y direcciones coincidentes y duplicados que usted mismo ha creado en la etapa One-source Match.

Al configurar la etapa Survive, usted ha seleccionado una regla, ha incluido columnas del archivo de origen, ha añadido una regla a cada columna y ha aplicado los datos. Una vez la etapa Survive ha procesado los registros para seleccionar el mejor registro, la información se envía al archivo de salida.

### Lecciones aprendidas

Al completar el Módulo 4, habrá aprendido los siguientes conceptos y tareas:

- Cómo utilizar la etapa Survive para crear el mejor candidato de un registro
- Cómo aplicar reglas simples a los valores de columna

---

## Guía de aprendizaje de IBM InfoSphere QualityStage: resumen

Con las lecciones de esta guía de aprendizaje se aprende cómo se puede utilizarse InfoSphere QualityStage para ayudar a una organización a gestionar y mantener la calidad de sus datos. Es fundamental que las empresas mantengan la máxima calidad en los datos de sus clientes. Por ello, la información debe estar actualizada y ser completa, precisa y fácil de utilizar.

La guía de aprendizaje presentaba un problema empresarial común: verificar los nombres y las direcciones de clientes y mostrar los pasos a seguir en los trabajos de InfoSphere QualityStage para unificar los distintos nombres asignados a una única familia. Esta guía de aprendizaje presentaba cuatro módulos que cubrían los cuatro trabajos del flujo de trabajo de InfoSphere QualityStage. Estos trabajos ofrecen a los clientes las siguientes garantías:

- Investigación de datos para identificar errores y validar los contenidos de los campos de un archivo de datos
- Condicionamiento de datos para asegurar que los datos de origen son coherentes internamente
- Coincidencia de datos para identificar todos los registros de un archivo que corresponden a registros similares de otro archivo
- Identificación de los registros de los datos de coincidencia que sobreviven para crear un mejor registro de candidatos

### Lecciones aprendidas

Una vez completada esta guía de aprendizaje, usted habrá aprendido los siguientes conceptos y tareas:

- Acerca del flujo de trabajo de InfoSphere QualityStage
- Cómo configurar un trabajo de InfoSphere QualityStage
- Cómo sirven los datos creados en un trabajo como datos de origen para el trabajo siguiente
- Cómo crear datos de calidad mediante InfoSphere QualityStage

---

## Apéndice A. Accesibilidad de los productos

Puede obtener información sobre el estado de accesibilidad de los productos de IBM.

Los módulos de producto y las interfaces de usuario de IBM InfoSphere Information Server no son totalmente accesibles.

Para obtener información sobre el estado de accesibilidad de los productos de IBM, consulte la información de accesibilidad de productos de IBM en [http://www.ibm.com/able/product\\_accessibility/index.html](http://www.ibm.com/able/product_accessibility/index.html).

### Documentación sobre accesibilidad

Se proporciona documentación accesible para los productos en IBM Knowledge Center. IBM Knowledge Center presenta la documentación en formato XHTML 1.0, que se puede ver en la mayoría de navegadores web. Dado que IBM Knowledge Center utiliza XHTML, puede establecer preferencias de visualización en el navegador. Esto también le permite utilizar lectores de pantalla y otras tecnologías de asistencia para acceder a la documentación.

La documentación que está en IBM Knowledge Center se proporciona en archivos PDF, que no son totalmente accesibles.

### IBM y la accesibilidad

Consulte el sitio web IBM Human Ability and Accessibility Center para obtener más información sobre el compromiso de IBM con la accesibilidad.



---

## Apéndice B. Cómo ponerse en contacto con IBM

Puede ponerse en contacto con IBM para obtener soporte al cliente, servicios de software, información sobre productos e información general. También puede facilitar comentarios a IBM sobre los productos y la documentación.

En la tabla siguiente se listan los recursos para soporte al cliente, servicios de software, formación e información sobre productos y soluciones.

Tabla 1. Recursos de IBM

Recurso	Descripción y ubicación
Portal de soporte de IBM	Puede personalizar la información de soporte eligiendo los productos y los temas que le interesen en <a href="http://www.ibm.com/support/entry/portal/Software/Information_Management/InfoSphere_Information_Server">www.ibm.com/support/entry/portal/Software/Information_Management/InfoSphere_Information_Server</a>
Servicios de software	Puede encontrar información sobre servicios de software, de tecnologías de la información y de consultoría de negocio en el sitio de soluciones, en <a href="http://www.ibm.com/businesssolutions/">www.ibm.com/businesssolutions/</a>
Mi IBM	Puede gestionar enlaces a sitios web de IBM y a información que satisfaga sus necesidades específicas de soporte técnico creando una cuenta en el sitio Mi IBM en <a href="http://www.ibm.com/account/">www.ibm.com/account/</a>
Formación y certificación	Puede obtener información sobre formación técnica y servicios de educación diseñados para personas, empresas y organizaciones públicas, a fin de adquirir, mantener y optimizar sus habilidades de TI en <a href="http://www.ibm.com/training">http://www.ibm.com/training</a>
Representantes de IBM	Puede contactar con un representante de IBM para obtener información sobre soluciones en <a href="http://www.ibm.com/connect/ibm/us/en/">www.ibm.com/connect/ibm/us/en/</a>



---

## Apéndice C. Acceso a la documentación del producto

La documentación se proporciona en diversos formatos: en el IBM Knowledge Center en línea, en un centro de información opcional instalado localmente y como manuales PDF. Puede acceder a la ayuda en línea o instalada localmente directamente desde las interfaces de cliente del producto.

IBM Knowledge Center es el mejor lugar para encontrar la información más actualizada de InfoSphere Information Server. IBM Knowledge Center contiene ayuda para la mayoría de las interfaces del producto, así como documentación completa para todos los módulos de producto de la suite. Puede abrir IBM Knowledge Center desde el producto instalado o desde un navegador web.

### Cómo acceder a IBM Knowledge Center

Existen varias maneras de acceder a la documentación en línea:

- Pulse el enlace **Ayuda** en la parte superior derecha de la interfaz de cliente.
- Pulse la tecla F1. Normalmente, la tecla F1 abre el tema que describe el contexto actual de la interfaz de cliente.

**Nota:** La tecla F1 no funciona en clientes web.

- Escriba la dirección en un navegador web, por ejemplo, cuando no tenga iniciada una sesión en el producto.

Escriba la siguiente dirección para acceder a todas las versiones de la documentación de InfoSphere Information Server:

<http://www.ibm.com/support/knowledgecenter/SSZJPZ/>

Si desea acceder a un tema concreto, especifique el número de versión con el identificador de producto, el nombre del plug-in de documentación y la vía de acceso al tema en el URL. Por ejemplo, el URL para la versión 11.3 de este tema es el siguiente. (El símbolo  $\Rightarrow$  indica una continuación de línea):

[http://www.ibm.com/support/knowledgecenter/SSZJPZ\\_11.3.0/=>com.ibm.swg.im.iis.common.doc/common/accessingiidoc.html](http://www.ibm.com/support/knowledgecenter/SSZJPZ_11.3.0/=>com.ibm.swg.im.iis.common.doc/common/accessingiidoc.html)

### Consejo:

El Knowledge Center tiene también un URL corto:

<http://ibm.biz/knowctr>

Para especificar un URL corto a una página de producto, versión o tema específico, utilice un carácter de almohadilla (#) entre el URL corto y el identificador de producto. Por ejemplo, el URL corto a toda la documentación de InfoSphere Information Server es el siguiente URL:

<http://ibm.biz/knowctr#SSZJPZ/>

Y el URL corto al tema anterior para crear un URL ligeramente más corto es el siguiente URL (El símbolo  $\Rightarrow$  indica una continuación de línea):

[http://ibm.biz/knowctr#SSZJPZ\\_11.3.0/com.ibm.swg.im.iis.common.doc/=>common/accessingiidoc.html](http://ibm.biz/knowctr#SSZJPZ_11.3.0/com.ibm.swg.im.iis.common.doc/=>common/accessingiidoc.html)

## Cambiar los enlaces de ayuda para que hagan referencia a la documentación instalada localmente

IBM Knowledge Center contiene la versión más actualizada de la documentación. Sin embargo, puede instalar una versión local de la documentación como un centro de información y configurar los enlaces de ayuda para que apunten a él. Un centro de información local es útil si su empresa no proporciona acceso a Internet.

Siga las instrucciones de instalación que vienen con el paquete de instalación del centro de información para instalarlo en el sistema que elija. Después de instalar e iniciar el centro de información, puede utilizar el mandato **iisAdmin** en el sistema de la capa de servicios para cambiar la ubicación de la documentación a la que hacen referencia la tecla F1 y los enlaces de ayuda del producto. (El símbolo ⇒ indica una continuación de línea):

### Windows

```
vía_instalación_IS\ASBServer\bin\iisAdmin.bat -set -key ⇒  
com.ibm.iis.infocenter.url -value http://<host>:<puerto>/help/topic/
```

### AIX Linux

```
vía_instalación_IS/ASBServer/bin/iisAdmin.sh -set -key ⇒  
com.ibm.iis.infocenter.url -value http://<host>:<puerto>/help/topic/
```

Donde <host> es el nombre del sistema donde está instalado el centro de información y <puerto> es el número de puerto para el centro de información. El número de puerto predeterminado es 8888. Por ejemplo, en un sistema llamado server1.example.com que utilice el puerto predeterminado, el valor del URL sería <http://server1.example.com:8888/help/topic/>.

## Obtener la documentación en PDF y en copia impresa

- Los manuales en archivos PDF están disponibles en línea y puede accederse a ellos desde este documento de soporte: <https://www.ibm.com/support/docview.wss?uid=swg27008803&wv=1>.
- También puede solicitar publicaciones de IBM en formato impreso en línea o a través de su representante local de IBM. Para solicitar publicaciones en línea, vaya al Centro de Publicaciones de IBM en <http://www.ibm.com/e-business/linkweb/publications/servlet/pbi.wss>.



---

## Avisos y marcas registradas

Esta información ha sido desarrollada para productos y servicios ofrecidos en los Estados Unidos. Este material puede estar disponible en IBM en otros idiomas. Sin embargo, es posible que deba tener una copia del producto o de la versión del producto en ese idioma para poder acceder al mismo.

### Avisos

Es posible que IBM no ofrezca en otros países los productos, servicios o características que se describen en este documento. Póngase en contacto con el representante local de IBM para obtener información acerca de los productos y servicios que actualmente están disponibles en su localidad. Cualquier referencia a un producto, programa o servicio de IBM no implica ni establece que sólo se pueda utilizar dicho producto, programa o servicio de IBM. En su lugar, se puede utilizar cualquier producto, programa o servicio funcionalmente equivalente que no infrinja ninguno de los derechos de propiedad intelectual de IBM. Sin embargo, es responsabilidad del usuario evaluar y verificar el funcionamiento de cualquier producto, programa o servicio que no sea de IBM.

IBM puede tener patentes o solicitudes de patente pendientes de aprobación que cubran temas tratados en este documento. La posesión de este documento no confiere ninguna licencia sobre dichas patentes. Puede enviar consultas sobre licencias, por escrito, a la siguiente dirección:

IBM Director of Licensing  
IBM Corporation  
North Castle Drive  
Armonk, NY 10504-1785 EE.UU.

Para realizar consultas relativas a la información de juego de caracteres de doble byte (DBCS), póngase en contacto con el departamento de propiedad intelectual de IBM de su país o envíe las consultas, por escrito, a:

Intellectual Property Licensing  
Legal and Intellectual Property Law  
IBM Japan Ltd.  
19-21, Nihonbashi-Hakozakicho, Chuo-ku  
Tokio 103-8510, Japón

**El párrafo siguiente no se aplica en el Reino Unido ni en ningún otro país en el que las disposiciones en él expuestas sean incompatibles con la legislación local:** INTERNATIONAL BUSINESS MACHINES CORPORATION PROPORCIONA ESTA PUBLICACIÓN "TAL CUAL", SIN GARANTÍAS DE NINGÚN TIPO, NI EXPLÍCITAS NI IMPLÍCITAS, INCLUYENDO, PERO SIN LIMITARSE A ELLAS, LAS GARANTÍAS IMPLÍCITAS DE NO INFRACCIÓN, COMERCIALIZACIÓN E IDONEIDAD PARA UN PROPÓSITO DETERMINADO. Algunas legislaciones no contemplan la declaración de limitación de garantías, ni implícitas ni explícitas, en determinadas transacciones, por lo que cabe la posibilidad de que esta declaración no se aplique en su caso.

Esta información puede contener imprecisiones técnicas o errores tipográficos. La información incluida en este documento está sujeta a cambios periódicos, que se

incorporarán en nuevas ediciones de la publicación. IBM puede realizar en cualquier momento mejoras o cambios en los productos o programas descritos en esta publicación sin previo aviso.

Cualquier referencia hecha en esta información a sitios web que no sean de IBM se proporciona únicamente para su comodidad y no debe considerarse en modo alguno como una aprobación de dichos sitios web. Los materiales de estos sitios web no forman parte de los materiales de este producto de IBM y el uso que haga de estos sitios web es de la entera responsabilidad del usuario.

IBM puede utilizar o distribuir cualquier información que se le facilite de la manera que considere adecuada, sin contraer por ello ninguna obligación con el remitente.

Los licenciatarios de este programa que deseen obtener información acerca del mismo con el fin de: (i) intercambiar la información entre los programas creados independientemente y otros programas (incluido éste) y (ii) utilizar mutuamente la información que se ha intercambiado, deben ponerse en contacto con:

IBM Corporation  
J46A/G4  
555 Bailey Avenue  
San Jose, CA 95141-1003 EE.UU.

Esta información puede estar disponible, sujeta a los términos y condiciones pertinentes, incluido en algunos casos el pago de una cantidad determinada.

IBM proporciona el programa bajo licencia descrito en este documento, y todo el material bajo licencia disponible para el mismo, bajo los términos del Acuerdo de cliente de IBM, el Acuerdo acuerdo internacional de licencia de programa de IBM o cualquier otro acuerdo equivalente entre las partes.

Los datos de rendimiento incluidos en este documento se determinaron en un entorno controlado. Por tanto, los resultados obtenidos en otros entornos operativos pueden variar significativamente. Algunas mediciones pueden haberse efectuado en sistemas a nivel de desarrollo, y no existe ninguna garantía de que dichas mediciones sean las mismas en sistemas de disponibilidad general. Además, es posible que algunas mediciones se hayan estimado mediante extrapolación. Los resultados reales pueden variar. Los usuarios de este documento deben verificar los datos aplicables a su entorno específico.

La información relacionada con productos no de IBM se ha obtenido de los suministradores de dichos productos, de sus anuncios publicados o de otras fuentes de información pública disponibles. IBM no ha probado dichos productos y no puede confirmar la precisión del rendimiento, la compatibilidad ni ninguna otra afirmación relacionada con productos que no son de IBM. Las consultas acerca de las prestaciones de los productos que no son de IBM deben dirigirse a los suministradores de tales productos.

Todas las declaraciones relativas a la dirección o intención futura de IBM están sujetas a cambios o anulación sin previo aviso y representan únicamente metas y objetivos.

Esta información se suministra sólo con fines de planificación. La presente información esta sujeta a cambios antes de que los productos que en ella se describen estén disponibles.

Esta información contiene ejemplos de datos e informes utilizados en las operaciones de negocios diarias. Para ilustrarlos de la forma más completa posible, los ejemplos incluyen nombres de personas, empresas, marcas y productos. Todos estos nombres son ficticios y cualquier similitud con los nombres y direcciones utilizados por una empresa real es totalmente casual.

#### LICENCIA DE COPYRIGHT:

Esta información contiene programas de aplicación de ejemplo en idioma de origen, que ilustra las técnicas de programación en diversas plataformas operativas. Puede copiar, modificar y distribuir estos programas de ejemplo de cualquier forma, sin pagar a IBM, con la finalidad de desarrollar, utilizar, comercializar o distribuir programas de aplicación que se ajusten a la interfaz de programación de aplicaciones para la plataforma operativa para la que se han escrito los programas de ejemplo. Estos ejemplos no se han probado bajo todas las condiciones posibles. Por lo tanto, IBM no puede garantizar ni dar por sentada la fiabilidad, capacidad de servicio o funcionamiento de esos programas. Los programas de ejemplo se suministran "TAL CUAL", sin garantía de ninguna clase. IBM no se hará responsable de los daños que puedan derivarse del uso de los programas de ejemplo.

Cada copia, parcial o completa, de estos programas de ejemplo o cualquier trabajo derivado, debe incluir un aviso de copyright como el siguiente:

© (el nombre de su empresa) (año). Partes de este código provienen de programas de ejemplo de IBM Corp. © Copyright IBM Corp. \_escriba el año o años\_. Reservados todos los derechos.

Si está viendo esta información en copia software, es posible que las fotografías y las ilustraciones en color no aparezcan.

### Consideraciones sobre la política de privacidad

Los productos de software de IBM, incluidas las soluciones de software como servicio, ("Ofertas de software"), pueden utilizar cookies u otras tecnologías para recopilar información sobre el uso de productos, para ayudar a mejorar la experiencia del usuario final, para personalizar las interacciones con el usuario final o para otros fines. En muchos casos, las Ofertas de software no recopilan información de identificación personal. Algunas de nuestras Ofertas de software pueden ayudarle a recopilar información de identificación personal. Si esta Oferta de software utiliza cookies para recopilar información de identificación personal, la información específica sobre el uso de cookies por parte de esta oferta se expone más abajo.

Dependiendo de las configuraciones desplegadas, esta Oferta de software puede utilizar cookies de sesión o persistentes. Si un producto o componente no está en la lista, ese producto o componente no utiliza cookies.

Tabla 2. Uso de cookies de los productos y componentes de InfoSphere Information Server

Módulo de producto	Componente o característica	Tipo de cookie que se utiliza	Recopilar estos datos	Finalidad de los datos	Inhabilitación de las cookies
Cualquiera (parte de la instalación de InfoSphere Information Server)	Consola web de InfoSphere Information Server	<ul style="list-style-type: none"> <li>Sesión</li> <li>Persistente</li> </ul>	Nombre de usuario	<ul style="list-style-type: none"> <li>Gestión de sesiones</li> <li>Autenticación</li> </ul>	No se pueden inhabilitar

Tabla 2. Uso de cookies de los productos y componentes de InfoSphere Information Server (continuación)

Módulo de producto	Componente o característica	Tipo de cookie que se utiliza	Recopilar estos datos	Finalidad de los datos	Inhabilitación de las cookies
Cualquiera (parte de la instalación de InfoSphere Information Server)	InfoSphere Metadata Asset Manager	<ul style="list-style-type: none"> <li>Sesión</li> <li>Persistente</li> </ul>	Ninguna información de identificación personal	<ul style="list-style-type: none"> <li>Gestión de sesiones</li> <li>Autenticación</li> <li>Usabilidad de usuario mejorada</li> <li>Configuración de inicio de sesión único</li> </ul>	No se pueden inhabilitar
InfoSphere DataStage	Etapas Big Data File	<ul style="list-style-type: none"> <li>Sesión</li> <li>Persistente</li> </ul>	<ul style="list-style-type: none"> <li>Nombre de usuario</li> <li>Firma digital</li> <li>ID de sesión</li> </ul>	<ul style="list-style-type: none"> <li>Gestión de sesiones</li> <li>Autenticación</li> <li>Configuración de inicio de sesión único</li> </ul>	No se pueden inhabilitar
InfoSphere DataStage	Etapas XML	Sesión	Identificadores internos	<ul style="list-style-type: none"> <li>Gestión de sesiones</li> <li>Autenticación</li> </ul>	No se pueden inhabilitar
InfoSphere DataStage	Consola de operaciones de IBM InfoSphere DataStage and QualityStage	Sesión	Ninguna información de identificación personal	<ul style="list-style-type: none"> <li>Gestión de sesiones</li> <li>Autenticación</li> </ul>	No se pueden inhabilitar
InfoSphere Data Click	Consola web de InfoSphere Information Server	<ul style="list-style-type: none"> <li>Sesión</li> <li>Persistente</li> </ul>	Nombre de usuario	<ul style="list-style-type: none"> <li>Gestión de sesiones</li> <li>Autenticación</li> </ul>	No se pueden inhabilitar
InfoSphere Data Quality Console		Sesión	Ninguna información de identificación personal	<ul style="list-style-type: none"> <li>Gestión de sesiones</li> <li>Autenticación</li> <li>Configuración de inicio de sesión único</li> </ul>	No se pueden inhabilitar
InfoSphere QualityStage Standardization Rules Designer	Consola web de InfoSphere Information Server	<ul style="list-style-type: none"> <li>Sesión</li> <li>Persistente</li> </ul>	Nombre de usuario	<ul style="list-style-type: none"> <li>Gestión de sesiones</li> <li>Autenticación</li> </ul>	No se pueden inhabilitar
InfoSphere Information Governance Catalog		<ul style="list-style-type: none"> <li>Sesión</li> <li>Persistente</li> </ul>	<ul style="list-style-type: none"> <li>Nombre de usuario</li> <li>Identificadores internos</li> <li>Estado del árbol</li> </ul>	<ul style="list-style-type: none"> <li>Gestión de sesiones</li> <li>Autenticación</li> <li>Configuración de inicio de sesión único</li> </ul>	No se pueden inhabilitar
InfoSphere Information Analyzer	Etapas Reglas de datos en el cliente del Diseñador de InfoSphere DataStage and QualityStage	Sesión	ID de sesión	Gestión de sesiones	No se pueden inhabilitar

Si las configuraciones desplegadas para esta Oferta de software le ofrecen como cliente la posibilidad de recopilar información de identificación personal de los usuarios finales mediante cookies y otras tecnologías, debe buscar asesoramiento jurídico sobre la legislación aplicable a dicha recopilación de datos, incluidos los requisitos de notificación y consentimiento.

Para obtener más información sobre el uso de diversas tecnologías, incluidas las cookies, para estos fines, consulte la Política de privacidad de IBM en <http://www.ibm.com/privacy>, la sección “Cookies, balizas web y otras tecnologías” de la Declaración de privacidad en línea de IBM en <http://www.ibm.com/privacy/details> y la “Declaración de privacidad de productos de software y software como servicio de IBM” (en inglés) en <http://www.ibm.com/software/info/product-privacy>.

## **Marcas registradas**

IBM, el logotipo de IBM e [ibm.com](http://www.ibm.com) son marcas comerciales o marcas registradas de International Business Machines Corp., registradas en muchas jurisdicciones en todo el mundo. Otros nombres de productos y servicios pueden ser marcas registradas de IBM u otras empresas. Encontrará una lista actual de las marcas registradas de IBM en el sitio web [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

Los términos siguientes son marcas comerciales o marcas registradas de otras empresas:

Adobe es una marca registrada de Adobe Systems Incorporated en los Estados Unidos y/o en otros países.

Intel e Itanium son marcas comerciales o marcas registradas de Intel Corporation o sus filiales en los Estados Unidos y otros países.

Linux es una marca registrada de Linus Torvalds en los Estados Unidos y/ en otros países.

Microsoft, Windows y Windows NT son marcas registradas de Microsoft Corporation en los Estados Unidos y/o en otros países.

UNIX es una marca registrada de The Open Group en los Estados Unidos y en otros países.

Java<sup>™</sup> y todas las marcas registradas y logotipos basados en Java son marcas comerciales o marcas registradas de Oracle y/o sus filiales.

El Servicio de correos de Estados Unidos (United States Postal Service) es propietario de las siguientes marcas registradas: CASS, CASS Certified, DPV, LACS<sup>Link</sup>, ZIP, ZIP + 4, ZIP Code, Post Office, Postal Service, USPS y United States Postal Service. IBM Corporation es un licenciataria no exclusivo de DPV y LACS<sup>Link</sup> del Servicio de correos de Estados Unidos.

Otros nombres de empresas, productos y servicios pueden ser marcas comerciales o marcas de servicio de terceros.



---

# Índice

## A

- accesibilidad de los productos
  - accesibilidad 45
- análisis de direcciones 14
- analizar datos que están en formato libre 8
- analizar direcciones 14
- archivo
  - origen 11
  - Secuencial 11
- archivo de origen
  - configurar 11
  - renombrar 11
- Archivo secuencial 16, 19
- archivos de destino del trabajo
  - One-source Match
    - configurar 36
- avisos legales 51

## C

- caso de ejemplo del proyecto de la guía de aprendizaje 3
- columnas, correlación 12
- componentes de cliente 3
- componentes de la guía de aprendizaje
  - importar 7
- componentes de servidor 3
- configurar
  - etapa Match Frequency 27
  - trabajo de etapa Investigate 8
  - trabajo Standardize 19
- configurar guía de aprendizaje 4
- configurar la etapa Copy 27
- copiando metadatos 27
- copiar datos de la guía de aprendizaje 5
- correlación de columnas 27
- crear proyecto de la guía de aprendizaje 5

## D

- data
  - analizar formato libre 8
  - standardize 19
- datos de la guía de aprendizaje
  - copiar 5
- Diseñador de InfoSphere DataStage 6
- documentación del producto
  - acceder 49

## E

- elementos del proyecto 1
- enlaces, renombrar 10
- Etapa Copy
  - configurar 12, 27
- etapa Investigate 8
  - configurar 13, 14

- etapa Match Frequency
  - columnas 27
  - configurar 27
- etapa One-source Match
  - configurar archivos de destino 36
- etapa Standardize
  - configurar 21
  - datos de condicionamiento 19
  - Estandarizar conjuntos de reglas 21
- etapa Survive
  - configurar 39
  - renombrar enlaces y etapas 39
- etapas
  - Copy 12, 19, 27
  - Investigate 8
  - Match Frequency 19, 27
  - Standardize 19, 21
  - Transformer 19
- etapas, renombrar 10

## G

- guía de aprendizaje
  - configuración 4

## I

- importar componentes de la guía de aprendizaje 7
- informe de patrones 8, 16
- Informe de patrones de palabra 8
- informe de señales 8, 16
- informes
  - configurar salida 16
  - patrón 8, 16
  - Patrón de palabra 8
  - señal 8, 16
  - Señal de palabra 8
- informes de salida, configurar 16
- InfoSphere DataStage
  - Cliente del Diseñador 1
  - crear un trabajo 6
  - etapa Copy 27
  - Etapa Copy 12
- InfoSphere QualityStage
  - etapa One-source Match 31
  - etapa Survive 39, 40
- etapas 2
- proyectos 1
- Trabajo de etapas Survive 39
  - resumen 43
- trabajos 2
- valor 1
- investigación de columnas de dominio único 8

## L

- Lección 1.1 8
- limpiar datos 1

- limpieza de datos 1

## M

- marcas registradas
  - lista de 51
- metadatos 12
  - carga 11
- Módulo 2, acerca de 19

## O

- objetivos del proyecto de la guía de aprendizaje 3

## P

- Palabra 8
- Paleta de herramientas del Diseñador
  - grupo de calidad de datos 2
- proyecto de la guía de aprendizaje
  - crear 5
- proyectos 1
  - apertura 6

## R

- registros
  - agrupar 31

## S

- servicios de software
  - contactar 47
- soporte
  - cliente 47
- soporte al cliente
  - contactar 47

## T

- trabajo de etapa Investigate
  - configurar 8
  - renombrar enlaces y etapas 10
- trabajo de etapa Standardize
  - configurar 19
- trabajo One-source Match
  - configurar 31
- Trabajo paralelo
  - guardar 6
- trabajo Survive
  - configurar 40
- trabajo Survive, configurar 39
- trabajos
  - visión general 2
- trabajos de etapa One-source Match
  - agrupar registros con atributos comunes 31

trabajos de etapa One-source Match (continuación)	
configurar archivos de origen	33
trabajos de la etapa One-source Matchconfigurar la etapa	34







Impreso en España

SC43-1188-00

