

IBM InfoSphere DataStage
Versión 11 Release 3

*Introducción a IBM InfoSphere
DataStage*



IBM InfoSphere DataStage
Versión 11 Release 3

*Introducción a IBM InfoSphere
DataStage*



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 35.

Contenido

Introducción a InfoSphere DataStage . . . 1

Ciclo de vida de trabajo	1
Desarrollar	3
Probar	5
Desplegar	7
Operar	8
Diseños de trabajos	10
Extraer y cargar datos	10
Transformar datos	14
Enriquecer los datos	17
Limpiar datos	18
Proceso en tiempo real.	21
Proceso de datos grandes.	23
Combinar trabajos en un trabajo de secuencia	24
Integración de InfoSphere DataStage en InfoSphere Information Server	25

Apéndice A. Accesibilidad de los productos 27

Apéndice B. Cómo ponerse en contacto con IBM 29

Apéndice C. Acceso a la documentación del producto 31

Apéndice D. Cómo proporcionar comentarios sobre la documentación del producto 33

Avisos y marcas registradas 35

Índice 41

Introducción a InfoSphere DataStage

IBM® InfoSphere DataStage es una herramienta de integración de datos para diseñar, desarrollar y ejecutar trabajos que mueven y transforman datos.

InfoSphere DataStage es el componente de integración de datos de IBM InfoSphere Information Server. Proporciona una infraestructura gráfica para desarrollar los trabajos que mueven datos los sistemas de origen a los sistemas de destino. Los datos transformados se pueden entregar a depósitos de datos, despensas de datos y almacenes de datos operativos, servicios web y sistemas de mensajería en tiempo real, y otras aplicaciones empresariales. InfoSphere DataStage da soporte a la extracción, transformación y carga (ETL) y la extracción, carga y transformación (ELT) de patrones. InfoSphere DataStage utiliza el proceso en paralelo y la conectividad de la empresa para proporcionar una plataforma realmente escalable.

Con InfoSphere DataStage, su empresa puede conseguir estos objetivos:

- Diseñar flujos de datos que extraen información de varios sistemas de origen, transformar los datos según convenga, y entregar los datos a bases de datos o aplicaciones de destino.
- Conectar directamente a aplicaciones empresariales como orígenes o destinos para garantizar que los datos sean relevantes, completos y precisos.
- Reducir el tiempo de desarrollo y mejorar la coherencia del diseño y el despliegue, utilizando las funciones precreadas.
- Minimizar el ciclo de desarrollo de proyectos trabajando con un conjunto común de herramientas a nivel de InfoSphere Information Server.

Ciclo de vida de trabajo

Puede desarrollar, probar, desplegar y ejecutar trabajos que muevan datos de sistemas de origen a sistemas de destino, mediante IBM InfoSphere DataStage.

InfoSphere DataStage funciona con los demás componentes de IBM InfoSphere Information Server para proporcionar una solución de integración de información unificada. Cada componente proporciona funciones clave a través de las fases de un proyecto de integración de información:

- Planificar
- Descubrir y analizar
- Diseñar
- Desarrollar
- Desplegar
- Entregar

InfoSphere DataStage proporciona la posibilidad de desarrollar, probar, desplegar y ejecutar los trabajos que proporcionan los datos. Los usuarios que completan las tareas del ciclo de vida de trabajo son el desarrollador, el coordinador de release y el operador. Su empresa puede tener distintos roles de trabajo, o puede llamar a estos roles mediante otros nombres. Un usuario puede cubrir todos los roles, o varios usuarios pueden compartirlos.

En la figura siguiente se muestra el ciclo de vida de un trabajo de InfoSphere DataStage, desde su desarrollo hasta la entrega de los datos.

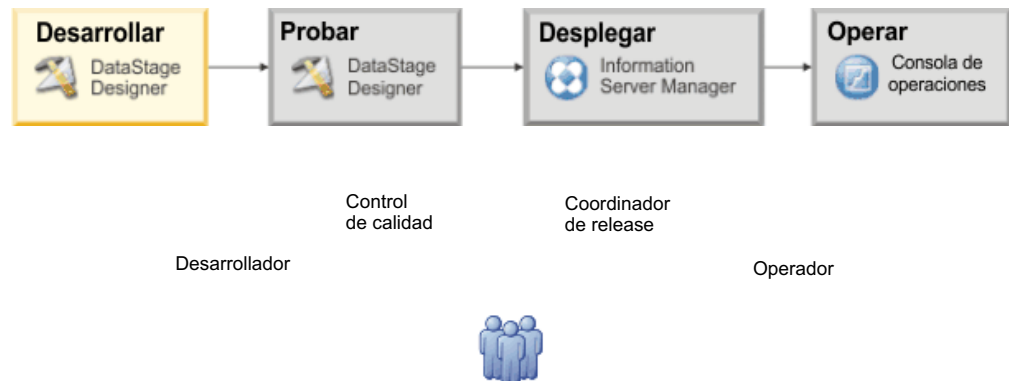


Figura 1. Ciclo de vida de trabajo

Desarrollar

El desarrollador crea los trabajos que mueven los datos del origen al destino, utilizando el IBM InfoSphere DataStage and QualityStage Cliente del Diseñador. El desarrollador diseña el flujo de datos organizando las etapas que sean necesarios para los orígenes de datos, las transformaciones y los destinos. El desarrollador define los enlaces que conectan las etapas y configura éstas.

Una etapa describe un origen de datos, un paso de proceso o un sistema de destino. La etapa también define la lógica de proceso que mueve los datos de los enlaces de entrada a los enlaces de salida. Si la instalación de InfoSphere Information Server incluye IBM InfoSphere QualityStage, también se proporcionan las etapas que analizan y limpian los datos. Si la instalación incluye IBM InfoSphere Information Analyzer, se proporciona la etapa que contiene reglas de datos. Las reglas de datos validan los datos y comprueban la calidad de los mismos.

Cuando el desarrollador guarde el trabajo, los detalles se grabarán en el repositorio de metadatos. Los demás usuarios que estén autorizados en el proyecto, podrán ver y editar los diseños de trabajo.

Un proyecto es un contenedor que organiza y proporciona seguridad para los trabajos. El administrador crea un proyecto en el cual trabajen los desarrolladores, mediante IBM InfoSphere DataStage and QualityStage Administrator. El administrador define la seguridad a nivel de proyecto, de modo que sólo los usuarios que están autorizados para el proyecto puedan acceder a los trabajos.

Probar

El desarrollador prueba el trabajo hasta que esté listo para la producción mediante el Cliente del Diseñador. El Cliente del Diseñador incluye herramientas de depuración visual y etapas de prueba específicas que dan soporte al proceso que los desarrolladores siguen, a medida que diseñan los trabajos de integración de datos.

Desplegar

El coordinador de release verifica y aprueba el trabajo, y mueve el trabajo al entorno de prueba o al entorno de producción, utilizando IBM InfoSphere Information Server Manager. Las organizaciones que despliegan

artefactos a través de su sistema de control de códigos fuente de la empresa pueden comprobar los artefactos directamente, y gestionar el trabajo de promoción y creación de versiones.

Operar

El operador ejecuta y gestiona los trabajos en producción mediante la Consola de operaciones de IBM InfoSphere DataStage and QualityStage. El operador supervisa el proyecto para evaluar cómo está trabajando con otras partes del sistema. Para dar soporte al análisis, la consola proporciona gráficos sobre el rendimiento del trabajo y el rendimiento del sistema, que incluye el uso de la CPU, el uso de la memoria y del disco.

Conceptos relacionados:

- ➡ Fases de integración de la información
- ➡ Diseño de trabajos de DataStage y QualityStage

Desarrollar

Desarrollar los trabajos que mueven los datos del origen al destino.

Puede crear y desarrollar trabajos mediante el Cliente del Diseñador. Para desarrollar un trabajo, puede definir el flujo de trabajo, y configurar el trabajo tal como se muestra en la figura siguiente. A continuación, puede guardar y compilar el trabajo, y revisar los errores de compilación.

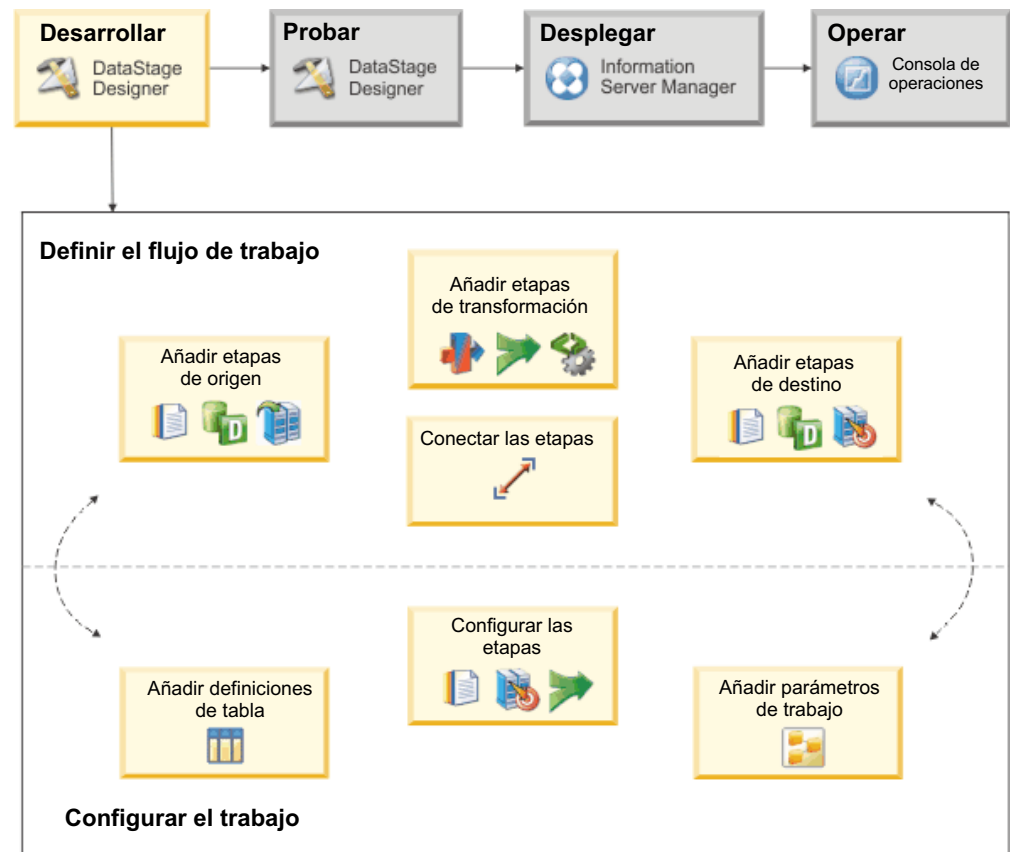


Figura 2. Desarrollo de un trabajo

Definir el flujo de trabajo

Puede diseñar el flujo de datos para un trabajo organizando las etapas en el lienzo del Diseñador, y conectando las etapas con enlaces. Las etapas representan los orígenes de datos, las transformaciones y los destinos. Las etapas contienen la lógica de proceso del trabajo. Los enlaces contienen metadatos sobre los datos que se mueve entre etapas.

Configurar el trabajo

Puede configurar el trabajo añadiendo definiciones de tabla, configurando las etapas y añadiendo parámetros de trabajo.

- Puede añadir definiciones de tabla a los enlaces. Las definiciones de tabla contienen información sobre la estructura de los datos y sobre la ubicación de las tablas o los archivos que contienen los datos. Puede importar una definición de tabla de la base de datos de origen o de destino, o puede crear la definición de tabla utilizando el Cliente del Diseñador.
- Puede configurar las etapas especificando las propiedades que definen el modo en que la etapa procesa los datos.
- Puede utilizar parámetros en los trabajos para especificar valores en tiempo de ejecución, en lugar de grabar los valores en el código.

Revisar los errores de compilación

Puede revisar las etapas que tengan errores al compilar un trabajo. Si un trabajo tiene un error de compilación, puede seleccionar una opción para resaltar la etapa que contenga el primer error en el diseño. Puede revisar la información sobre el error y, a continuación, editar la etapa para cambiar los valores incorrectos, y volver a compilar el trabajo.

En este trabajo de ejemplo, a una columna de la etapa Transformer, Trim_Prep, le falta una propiedad necesaria, por lo tanto, la etapa Transformer aparece resaltada.



Figura 3. Revisión de errores de compilación

Conceptos relacionados:

- ➡ Diseño de trabajos de DataStage y QualityStage
- ➡ Esbozo de los diseños de trabajo
- ➡ Definición de los datos
- ➡ Configuración de los diseños
- ➡ Lista alfabética de las etapas
- ➡ Consejos para el diseño de trabajos

Probar

Depurar los trabajos utilizando las etapas de depuración y la herramienta de depuración interactiva.

Puede utilizar las etapas de depuración al principio del ciclo de desarrollo para generar datos de prueba y datos de ejemplo durante el proceso. Tras compilar un trabajo, puede ejecutar éste en modalidad de depuración y examinar los datos de la columna de los puntos de interrupción, para comprender el flujo de trabajo.

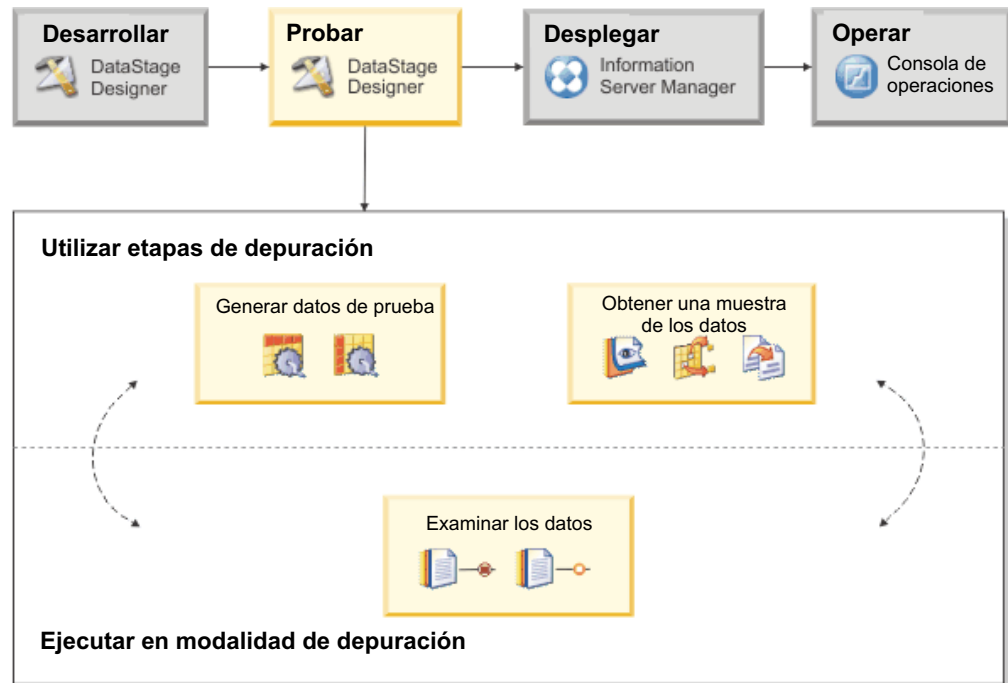


Figura 4. Probar un trabajo durante el desarrollo

Generar datos de prueba

Puede utilizar las etapas de depuración para generar datos de prueba. Por ejemplo, puede generar datos basados en la definición de tabla para el sistema de origen y, a continuación, puede crear y probar la lógica en sentido descendente, en base a los datos generados. Puede utilizar los datos para probar la lógica de trabajo al principio del ciclo de vida de trabajo, antes de los datos del sistema de origen estén disponibles.

Este trabajo de ejemplo genera datos de clientes utilizando la etapa Row Generator. El trabajo enriquece los datos generados con datos de cliente reales, procedentes de una base de datos Oracle, utilizando la etapa Lookup.

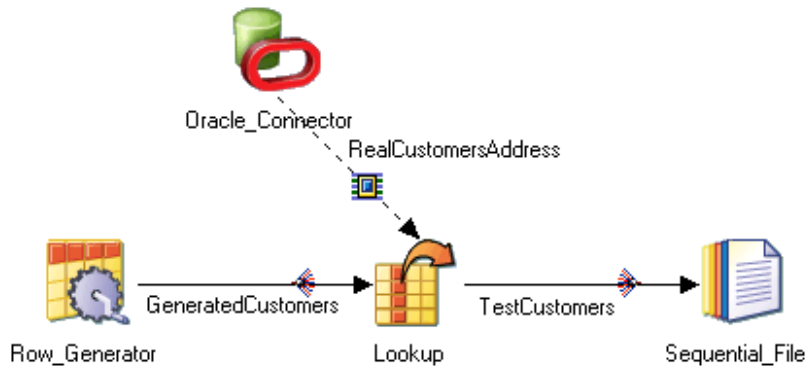


Figura 5. Generación de datos de prueba

Obtener una muestra de los datos

Puede utilizar las etapas de depuración para obtener una muestra de los datos de salida, mientras desarrolla un diseño de trabajo. Por ejemplo, puede obtener una muestra de los datos en un conjunto de datos que sea la salida de un trabajo, utilizando la etapa Peek. La etapa Peek graba los datos en el registro de trabajos. En este trabajo, los datos originales se copian a otro conjunto de datos, y las columnas de muestreo se graban en el archivo de registro para su revisión.



Figura 6. Datos de ejemplo

Examinar los datos de la columna en los puntos de interrupción

Puede depurar un trabajo estableciendo puntos de interrupción en los enlaces donde desee que el trabajo se detenga cuando se ejecute el trabajo en la modalidad de depuración. Cuando el trabajo se detenga en un punto de interrupción, podrá examinar los datos de la columna que se esté procesando. También podrá cambiar los puntos de interrupción para otras condiciones de datos y examinar las propiedades de la etapa y los valores de parámetro del trabajo. En este trabajo de ejemplo, se establecen dos puntos de interrupción para detener el trabajo después de cada cinco filas. El trabajo se detiene en los puntos de interrupción, y en la Ventana de depuración se muestran los datos de columna de cada punto de interrupción.

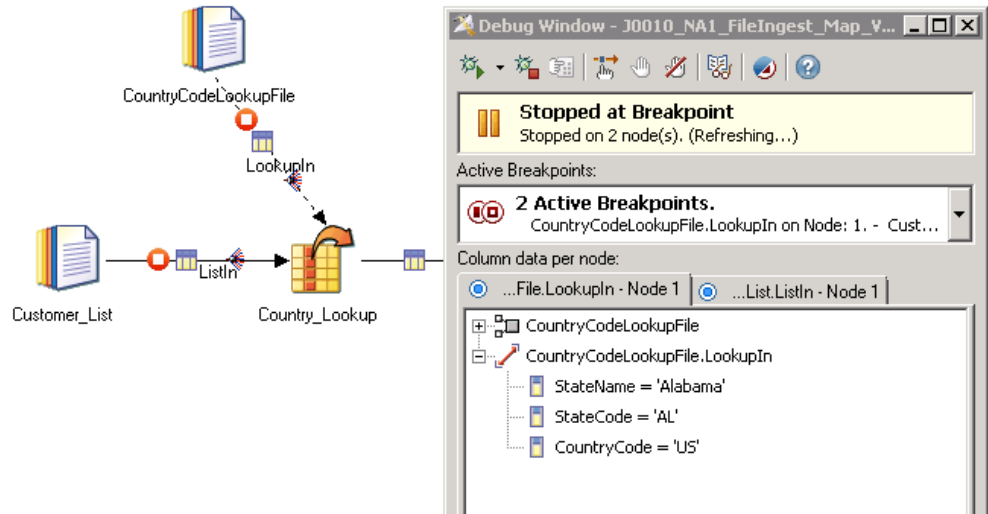


Figura 7. Establecer puntos de interrupción

Conceptos relacionados:

➡ Depuración de trabajos paralelos

Tareas relacionadas:

➡ Depuración de trabajos paralelos con las etapas de depuración

➡ Ejecución de trabajos paralelos en modalidad de depuración

Desplegar

Despliegue trabajos entre los entornos que den soporte a su ciclo de vida de desarrollo de software: entornos de desarrollo, prueba, producción y otros.

En un escenario típico, puede promocionar sus trabajos de un entorno a otro, a medida que los trabajos pasen las pruebas o demás requisitos de preparación para su organización. Puede mover un único trabajo, o varios trabajos, y sus activos relacionados entre dichos entornos, utilizando IBM InfoSphere Information Server Manager.

Puede completar otras tareas que den soporte al proceso de despliegue mediante InfoSphere Information Server Manager. Por ejemplo, puede buscar activos que hayan cambiado desde una fecha concreta, o identificar objetos dependientes. Puede añadir activos que se encuentren fuera de InfoSphere DataStage, que también deban desplegarse con sus trabajos. Puede gestionar los trabajos de InfoSphere DataStage y los activos asociados al sistema de control de origen.

El flujo de trabajo básico para desplegar un trabajo utilizando InfoSphere Information Server Manager incluye las tareas siguientes:

1. Definir un paquete para especificar los activos que deben desplegarse.
2. Crear el paquete para producir un archivo de paquete.
3. Copiar el archivo de paquete al sistema de destino, si es necesario.
4. Desplegar el paquete en el sistema de destino.

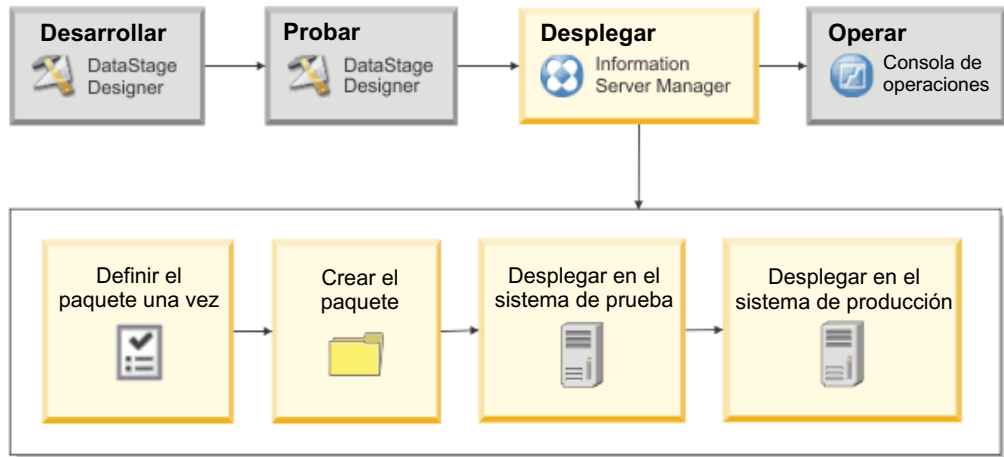


Figura 8. Despliegue de un trabajo

Definir el paquete

Al definir el paquete de despliegue, puede seleccionar los activos de los motores de InfoSphere Information Server y los proyectos que compartan un repositorio de metadatos. Cuando los activos del sistema de origen cambien, puede volver a crear, volver a copiar y volver a desplegar el paquete. No es necesario que vuelva a definir el paquete.

Crear el paquete

Al crear el paquete, las copias de los activos que se hayan definido para el paquete, se añaden al archivo de paquete. Los elementos se agregan a la compilación en su estado, en el momento de la compilación. Si vuelve a crear un paquete, puede sustituir el paquete completo o sólo actualizar los elementos del paquete que han cambiado desde la última compilación. El paquete ya está preparado para el despliegue.

Desplegar el paquete

Puede desplegar el paquete en proyectos distintos en el mismo sistema de InfoSphere DataStage and QualityStage, por ejemplo, cuando el entorno de desarrollo y el de prueba se encuentran en el mismo servidor. O puede desplegar el paquete en diferentes sistemas de InfoSphere DataStage and QualityStage.

Gestionar trabajos en el control de origen

Puede comprobar artefactos directamente dentro y fuera de los sistemas de control de códigos fuente, entre los que se incluyen IBM Rational Team Concert, IBM ClearCase, Subversion, Microsoft Team Foundation Server y otros. Puede utilizar el sistema de control de códigos fuente de su empresa para gestionar promoción de trabajos y creación de versiones, si éste es el estándar de su empresa.

Conceptos relacionados:

☞ Despliegue de trabajos y acceso al control de versión

Operar

Supervisar y gestionar trabajos en producción.

Puede acceder a la información sobre trabajos, actividad de trabajo, recursos del sistema y colas de gestión de carga de trabajo para los motores de InfoSphere Information Server utilizando la Consola de operaciones de IBM InfoSphere DataStage and QualityStage. Puede ejecutar trabajos bajo demanda y gestionar los recursos del sistema.

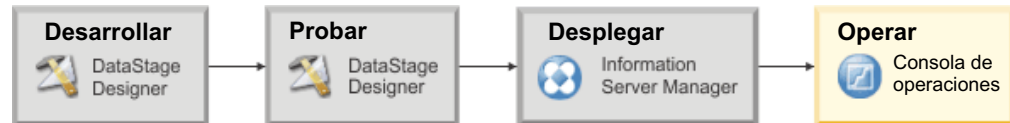


Figura 9. Operar

Buscar respuestas a las preguntas más comunes

El panel de instrumentos de la consola proporciona respuestas rápidas al ver y analizar el entorno de ejecución. En las gráficas del panel de instrumentos puede revisar la carga de trabajo de ejecución del trabajo, los trabajos que hayan concluido recientemente, el estado de los servicios del motor y el uso de los recursos del sistema. El panel de instrumentos también proporciona el recuento de ejecuciones de trabajo satisfactorias y fallidas, y el estado de los servicios clave que se estén ejecutando en el motor. En el panel de instrumentos se muestran alertas cuando los trabajos de ejecución fallan y cuando se pasan los umbrales de métrica.

Acceder a los detalles y el historial de los trabajos

En el panel de instrumentos puede profundizar en la información para ver detalles de trabajo y proyecto específicos. Por ejemplo, puede obtener más detalles sobre el recuento de trabajos de ejecución fallidos, para ver una lista de los trabajos fallidos. A continuación, puede obtener más detalles sobre un trabajo específico, para ver más información sobre las propiedades, los parámetros, el rendimiento y los mensajes. En la página Proyectos puede revisar la información de resumen de alto nivel, y la información detallada sobre la actividad actual e histórica. También puede recuperar y ver el registro completo directamente desde el servidor.

Optimizar los recursos del sistema

Mediante las capacidades de gestión de la carga de trabajo de la Consola de operaciones, puede gestionar los recursos del sistema de forma dinámica, donde varios equipos comparten una infraestructura de hardware común. Puede establecer políticas del sistema, supervisar trabajos en cola y gestionar las colas de gestión de carga de trabajo.

Priorizar las tareas de misión crítica

Puede regular una actividad de trabajo donde los recursos del sistema excedan los umbrales que haya especificado y asignar la prioridad de cualquier trabajo sometido. Si es un usuario con privilegios, puede utilizar las alteraciones temporales manuales para promover trabajos específicos al principio de la cola.

Conceptos relacionados:

➤ Supervisión de trabajos y ejecuciones de trabajos utilizando la ejecución del trabajo

➤ Administración de la gestión de carga de trabajo


Diseños de trabajos


Diseñar trabajos que extraen datos de varios orígenes, los transforman y los entregan a los sistemas de destino, en los formatos pertinentes.

En los ejemplos siguientes se muestran algunas de las formas en que puede diseñar trabajos mediante IBM InfoSphere DataStage. Los elementos clave de un diseño de trabajo son las etapas y los enlaces entre ellas. Una etapa describe un origen de datos, un paso de proceso o un sistema de destino. La etapa también define la lógica de proceso que mueve los datos de los enlaces de entrada a los enlaces de salida. Puesto que las etapas son flexibles y configurables, puede enlazar varias etapas juntas para satisfacer requisitos empresariales complejos.

Estos diseños de trabajo incluyen sólo un ejemplo de las etapas disponibles. Las etapas que puede utilizar dependen de sus requisitos y de su entorno.

Información relacionada:

 [IBM InfoSphere DataStage Data Flow and Job Design](#) (una publicación de IBM Redbooks)

 [InfoSphere DataStage Parallel Framework Standard Practices](#) (una publicación de IBM Redbooks)

 [Consejos para el diseño de trabajos](#)

Extraer y cargar datos

Puede desarrollar trabajos para mover datos entre varios orígenes, y en varios formatos. En estos ejemplos se muestra cómo puede extraer datos, identificar los cambios entre archivos, y cargar datos en una base de datos o un almacén de datos.

- “Extraer datos de una base de datos a un archivo o conjunto de datos”
- “Cargar datos con conectividad nativa de alta velocidad” en la página 11
- “Cargar datos en tiempo real en un almacén de datos” en la página 11
- “Extraer datos de SAP” en la página 12
- “Captar datos de un servicio web para enriquecer datos” en la página 12
- “Identificar cambios de la base de datos” en la página 13

Extraer datos de una base de datos a un archivo o conjunto de datos

Puede extraer datos masivos de una base de datos. Por ejemplo, la etapa Oracle Connector puede utilizar SQL personalizado en la base de datos Oracle para extraer la dirección del cliente, número de teléfono y el saldo de su cuenta, en paralelo. Puede desarrollar un trabajo similar para extraer datos de cualquier origen de base de datos relacional.

En el primer ejemplo, los datos extraídos se graban en un archivo secuencial para compartir con otros departamentos de la empresa. En el segundo ejemplo, los datos extraídos se graban en un conjunto de datos para su uso en sentido descendente en otro trabajo. Un conjunto de datos almacena datos en un formato constante que utiliza el almacenamiento en paralelo; es decir, el almacenamiento de datos particionados. Los conjuntos de datos optimizan el rendimiento de E/S al mantener el grado de particionamiento.



Figura 10. Extraer datos de una base de datos a un archivo secuencial



Figura 11. Extraer datos de una base de datos a un conjunto de datos

Cargar datos con conectividad nativa de alta velocidad

Puede leer datos de un archivo, ordenar los datos numéricamente para optimizar el almacenamiento, y cargar de forma masiva los datos en una base de datos en formato nativo. Por ejemplo, puede cargar datos de terceros que su organización recibe en una tabla de Netezza para su análisis, utilizando la etapa Netezza Connector. Al igual que cada una de las etapas de conector de base de datos, la etapa Netezza Connector de este ejemplo utiliza la API nativa de la base de datos para maximizar el rendimiento.



Figura 12. Cargar en una tabla de Netezza

Cargar datos en tiempo real en un almacén de datos

Puede desarrollar un trabajo para mover datos en tiempo real a un almacén de datos. Por ejemplo, puede capturar información en tiempo real sobre las ventas, el inventario y los envíos, de sistemas transaccionales de su empresa, en varias plataformas. A continuación, puede estandarizar e integrar la información, y entregarla a un almacén de datos que dé soporte a una única infraestructura de creación de informes.

En este ejemplo, el trabajo carga los datos de pedidos en tiempo real, utilizando la etapa WebSphere MQ Connector. El trabajo analiza los datos utilizando la etapa XML, y transforma los datos utilizando la etapa Transformer. A continuación, el trabajo la tabla de referencia y carga los datos en una base de datos DB2, utilizando la etapa Slowly Changing Dimension. La etapa Slowly Changing Dimension es una etapa de proceso que funciona dentro del contexto de una base de datos con esquema en estrella.

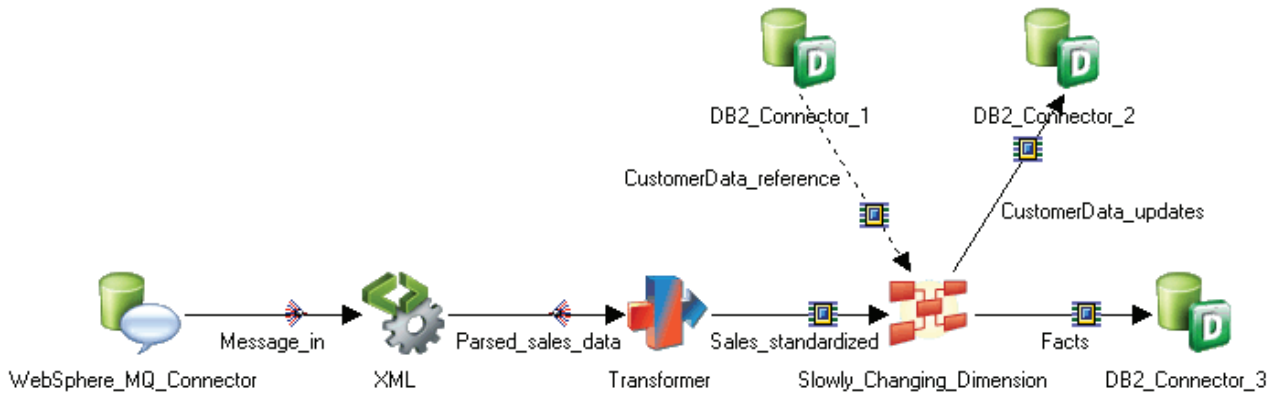


Figura 13. Cargar datos en un almacén de datos

Extraer datos de SAP

Puede extraer datos de SAP y cargarlos en SAP Business Warehouse (SAP BW), a un almacén de datos empresariales existente, a un archivo, o a otro destino. Por ejemplo, puede extraer datos utilizando la etapa ABAP (Advanced Business Application Programming) Extract. La etapa ABAP Extract genera un programa ABAP y lo sube al sistema SAP. El programa ABAP ejecuta una consulta SQL, y extrae los datos.

En este trabajo de ejemplo, se da formato a los datos extraídos para la etapa de destino utilizando una etapa Transformer. Los datos formateados se cargan en un almacén comercial SAP, utilizando la etapa SAP BW. Las etapas SAP las proporcionan IBM InfoSphere Information Server Pack for SAP Applications e InfoSphere Information Server Pack for SAP BW.



Figura 14. Extraer datos de SAP

Captar datos de un servicio web para enriquecer datos

Puede enriquecer los datos utilizando un servicio web para recuperar datos de una aplicación. Por ejemplo, puede recuperar datos de una aplicación de recursos humanos, por ejemplo Día laborable, utilizando la etapa Web Services Transformer. Este trabajo de ejemplo recupera los datos y, a continuación, los analiza utilizando la etapa XML, para generar filas y columnas.

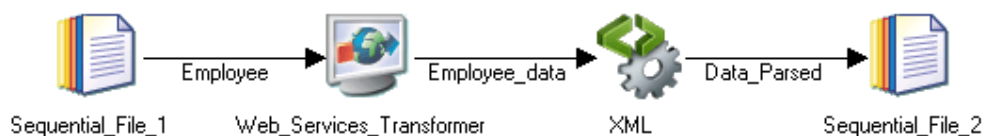


Figura 15. Captar datos de un servicio web

Identificar cambios de la base de datos

Puede identificar los cambios de la base de datos y aplicarlos a una base de datos de destino.

Este trabajo de ejemplo lee los datos que captura InfoSphere Change Data Capture (CDC) y aplica los datos cambiados a una base de datos de destino, utilizando la etapa CDC Transaction.

InfoSphere CDC forma parte de IBM InfoSphere Data Replication, que es una tecnología de réplica que captura los cambios de la base de datos a medida que se producen, y los entrega a InfoSphere DataStage, bases de datos de destino o colas de mensajes. InfoSphere Data Replication recopila los cambios de los datos utilizando el análisis de registro de base de datos de bajo impacto y de baja latencia.

Esta solución proporciona varias ventajas. La sobrecarga en el sistema de base de datos es menor porque el proceso de los datos se basa en el registro. A través de la red sólo se mueven los datos cambiados. El proceso total en el servidor ETL es menor, ya no es necesario que el servidor encuentre los cambios.

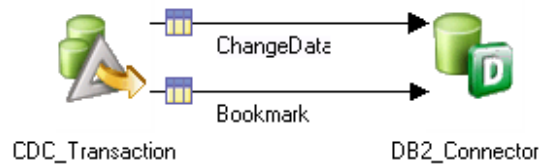


Figura 16. Identificar los cambios con la etapa CDC Transaction

Conceptos relacionados:

- ➡ Etapa Sequential file
- ➡ Etapa Data set
- ➡ Conector Oracle
- ➡ Conector Netezza
- ➡ Conector de IBM WebSphere MQ
- ➡ Transformación XML
- ➡ Etapa Slowly Changing Dimension
- ➡ Etapa Web Services Transformer
- ➡ Visión general de la etapa Transacción CDC
- ➡ InfoSphere Data Replication
- ➡ IBM InfoSphere Information Server Pack for SAP Applications

Transformar datos

Puede transformar datos para satisfacer tanto tareas de integración de datos simples como complejas. En estos ejemplos se muestra cómo puede unir datos, resumir datos, utilizar una tabla dinámica para reordenar los datos e identificar los cambios respecto a un archivo anterior.

- “Transformar datos con expresiones”
- “Unir datos de varios archivos sin formato”
- “Unir orígenes heterogéneos” en la página 15
- “Resumir datos por características comunes” en la página 15
- “Datos dinámicos de una tabla” en la página 16
- “Buscar los deltas del archivo de ayer” en la página 16

Transformar datos con expresiones

Puede crear transformaciones para aplicar a los datos utilizando la etapa Transformer. Estas transformaciones pueden ser simples o complejas, y pueden aplicarse a columnas individuales de los datos. Por ejemplo, puede recortar datos eliminando los espacios iniciales y finales, concatenar datos, realizar operaciones en fechas y horas, realizar operaciones matemáticas, y aplicar la lógica condicional. Las transformaciones se especifican mediante un conjunto de funciones. Puede definir las transformaciones utilizando el editor de expresiones.

Este trabajo de ejemplo incluye varias transformaciones: el número de cliente se recorta; el código de país se sustituye por Estados Unidos, si es EE.UU.; el proveedor y los códigos de tipo se concatenan para crear un código de tipo de proveedor; y se inserta la fecha actual.



Figura 17. Transformar datos con expresiones

Constraint	Column Name
TrimLeadingTrailing(Customers_in.CUSTOMER_NUMBER)	CUSTOMER_NUMBER
Customers_in.CUST_NAME	NAME
If Customers_in.COUNTRY_CODE = "US" Then "United States" Else ""	COUNTRY
Customers_in.CUST_VEND : Customers_in.CUST_TYPE_CODE	VENDER_TYPE
Customers_in.LAST_SALES_DATE	LAST_SALES_DATE
CurrentDate()	DATE

Figura 18. Definición de transformaciones en el editor de expresiones

Unir datos de varios archivos sin formato

Puede leer datos de varios archivos sin formato, y grabar los datos en un archivo de destino único. Puede combinar los registros de los archivos utilizando la etapa Join y usar técnicas como, por ejemplo, el particionamiento basado en claves y la

clasificación en memoria, para acelerar el proceso. Por ejemplo, puede combinar los registros de cliente con los detalles del pedido y grabar el pedido del cliente en el archivo de destino.

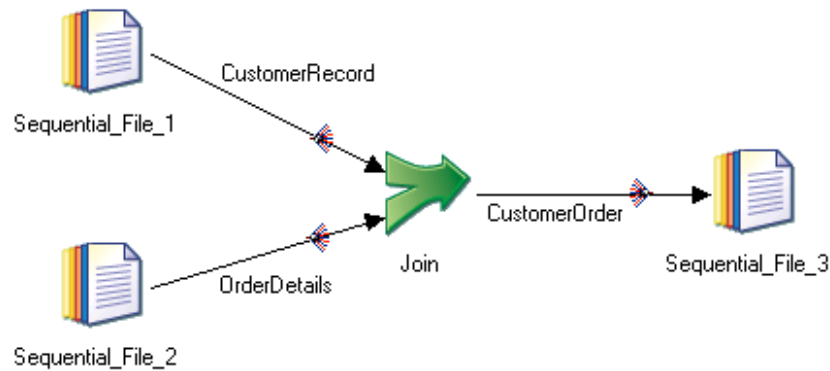


Figura 19. Unir datos de archivos sin formato

Unir orígenes heterogéneos

Puede recopilar, estandarizar y consolidar los datos de una amplia gama de orígenes de datos y estructuras de datos, y crear una fuente de información única, precisa y fiable. En este ejemplo, el trabajo extrae datos de dos tipos distintos de bases de datos, y carga los resultados en otro tipo de base de datos, en el formato adecuado. El trabajo utiliza una etapa Join para combinar los datos, las etapas Transformer para recortar los datos y darles formato.

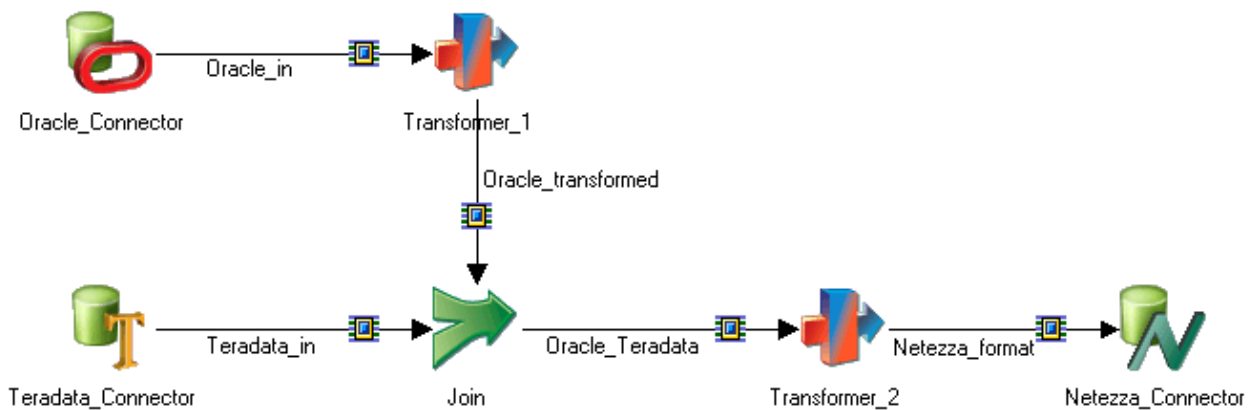


Figura 20. Unir orígenes heterogéneos

Resumir datos por características comunes

Puede resumir los datos utilizando la etapa Aggregator. Los registros se pueden agrupar por una o más características. Por ejemplo, puede agrupar los datos de ventas por día de la semana y mes, y calcular los totales. Estas agrupaciones pueden mostrar que el día más ocupado de la semana varía según la temporada.



Figura 21. Resumir datos por día

Datos dinámicos de una tabla

Puede correlacionar un conjunto de columnas de una fila de entrada en una sola columna, en varias filas de salida, mediante un giro horizontal, utilizando la etapa Pivot Enterprise. También puede realizar giros verticales, que combinan datos relacionados de varias filas en una sola fila con tipos de campo de repetición.



Figura 22. Girar datos

En este ejemplo, el esquema cambia, tal como se muestra en las dos tablas siguientes. Los datos se pueden girar en cualquier dirección, horizontal de la Tabla 1 a la Tabla 2, o vertical, de la Tabla 2 a la Tabla 1.

Tabla 1. Original

ID pedido	Fecha pedido	ID cliente	Nombre	Elemento1	Elemento2	Elemento3
123456	6-6-2013	ABCX	Juan López	portátil	monitor	teclado

Tabla 2. Después del giro horizontal

ID giro	ID pedido	Fecha pedido	ID cliente	Nombre	Elemento
1	123456	6-6-2013	ABCX	Juan López	portátil
2	123456	6-6-2013	ABCX	Juan López	monitor
3	123456	6-6-2013	ABCX	Juan López	teclado

Buscar los deltas del archivo de ayer

Puede identificar los cambios de un archivo anterior. Por ejemplo, puede comparar el archivo de ayer con el archivo de hoy, y realizar un registro de los cambios utilizando la etapa Change Capture. La etapa Change Capture toma dos conjuntos de datos de entrada, indicados antes y después, y genera un conjunto de datos único. Los registros de salida representan los cambios que se hayan realizado en los datos antes de obtener el conjunto de datos posterior. En este trabajo de ejemplo, se cargan los cambios entre dos archivos secuenciales en una base de datos Teradata, para su análisis.

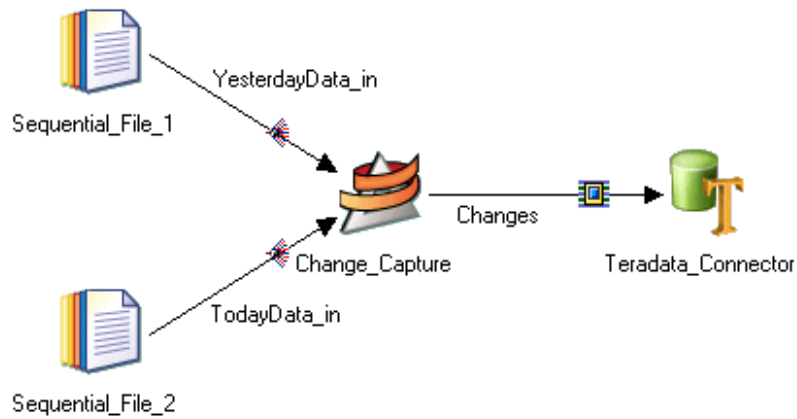


Figura 23. Identificar los cambios con la etapa Change Capture

Conceptos relacionados:

- ☞ Etapa Transformer
- ☞ Etapa Join
- ☞ Etapa Agregator
- ☞ Etapa Pivot Enterprise
- ☞ Etapa Change Capture

Enriquecer los datos

Estos ejemplos muestran cómo puede enriquecer los datos de cliente con datos de referencia, y cómo se puede aplicar las reglas empresariales para determinar un curso apropiado de la acción.

- “Buscar datos de referencia para el enriquecimiento de los datos”
- “Aplicar reglas empresariales” en la página 18

Buscar datos de referencia para el enriquecimiento de los datos

Puede enriquecer los datos añadiendo datos de un origen de referencia. Por ejemplo, puede añadir el código del país a una lista de direcciones de cliente, utilizando la etapa Lookup. En este trabajo de ejemplo se extraen las direcciones de cliente de una base de datos DB2, se hace referencia a los códigos de país de una base de datos Oracle, y se combinan los datos utilizando la etapa Lookup.

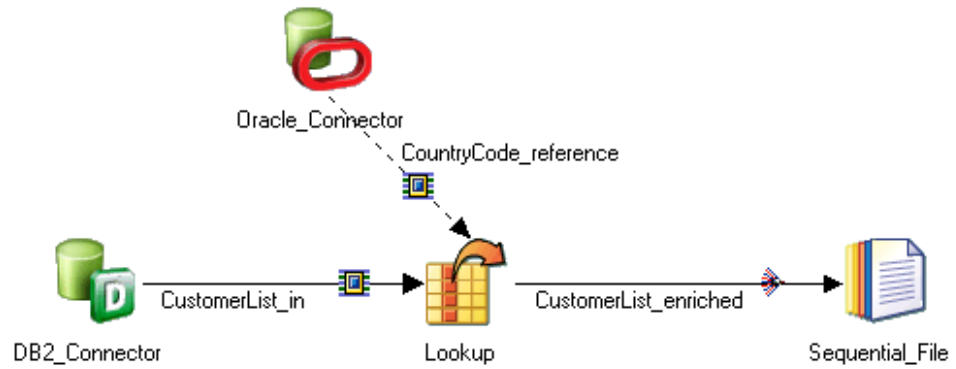


Figura 24. Enriquecer los datos buscando datos de referencia

Aplicar reglas empresariales

Además de las reglas de transformación y las reglas de datos que están incluidas en InfoSphere Information Server, puede invocar reglas empresariales complejas utilizando la etapa ILOG JRules Connector. Esta etapa accede a los conjuntos de reglas que proporciona IBM Operational Decision Manager. Las organizaciones que desarrollen una infraestructura de reglas empresariales podrán utilizar dichas reglas empresariales compartidas en sus flujos de trabajo de integración de datos.

Por ejemplo, puede determinar el curso de la acción apropiado que deba tomarse para cada uno de sus clientes, en base a las reglas empresariales aplicadas.



Figura 25. Aplicar reglas empresariales

Conceptos relacionados:

- ➡ Etapa Lookup
- ➡ Etapa ILOG JRules Connector

Limpiar datos

Puede desarrollar trabajos para estandarizar, verificar y validar los datos. En estos ejemplos se muestra cómo puede estandarizar los datos de cliente, verifique direcciones internacionales, identificar registros duplicados, y aplicar reglas de datos integrando las etapas de IBM InfoSphere QualityStage e IBM InfoSphere Information Analyzer.

- “Estandarizar datos” en la página 19
- “Verificar direcciones internacionales” en la página 19
- “Hacer coincidir los registros para identificar los duplicados” en la página 19
- “Validar datos con reglas de datos” en la página 20

Estandarizar datos

Puede analizar y estandarizar datos para alinearlos según sus necesidades, utilizando la etapa Standardize. Por ejemplo, puede estandarizar los nombres de organización para que la organización y los nombres de contacto estén claramente identificados, y que el nombre de la organización tenga una representación. Por ejemplo, las variaciones del nombre de la empresa ficticia Sample Outdoor, que pueden incluir SOC, S. O. C y Sample Outdoor Co, se representan mediante Sample Outdoor Company. También puede estandarizar otros tipos de datos, entre los que se incluye el correo electrónico, las descripciones de productos o las direcciones, de forma que 100 W. Main St y 100 West Main Street se convierten, tanto una dirección como la otra, en 100 W Main St.

La etapa Standardize utiliza conjuntos de reglas que se han diseñado para estandarizar los datos de forma que cumplan los estándares de la industria, para mejorar la coincidencia de datos, o facilitar el análisis en sentido descendente. La etapa Standardize la proporciona IBM InfoSphere QualityStage. Puede desarrollar los conjuntos de reglas utilizando IBM InfoSphere QualityStage Standardization Rules Designer, que es una herramienta de clasificación orientada a datos, basada en web.



Figura 26. Estandarizar datos

Verificar direcciones internacionales

Puede validar y estandarizar datos de dirección para garantizar que cada dirección se ajuste a los estándares postales y determinar si la dirección resulta adecuada para entregar la información, utilizando la etapa Address Verification, que proporciona IBM InfoSphere QualityStage Address Verification Interface. Por ejemplo, puede extraer datos de dirección de un sistema de gestión de relaciones con los clientes, verificar las direcciones internacionales y cargarlos en un almacén de datos de Netezza. La Address Verification proporciona procesos que pueden organizar, verificar y transformar datos de dirección para idiomas y postales, en varios países y regiones.



Figura 27. Verificar direcciones internacionales

Hacer coincidir los registros para identificar los duplicados

Puede identificar los registros duplicados para entidades como, por ejemplo, individuos, empresas, proveedores, productos o sucesos, haciendo coincidir los datos. La coincidencia es un sistema probabilístico de enlace de registros, que

proporciona IBM InfoSphere QualityStage, que identifica los registros que probablemente representen la misma entidad. El proceso de coincidencia mejora la integridad de los datos. Por ejemplo, puede identificar los registros de cliente duplicados en el sistema de gestión de relaciones con los clientes.

Para este trabajo de ejemplo, los datos de frecuencia de coincidencia se han generado en un trabajo anterior, utilizando una etapa Match Frequency. Los datos de clientes se han extraído de una base de datos DB2. La etapa One Source Match identifica las coincidencias, los duplicados, y las no coincidencias, combina los datos utilizando una etapa Funnel, y carga los datos en una base de datos Netezza. Los registros administrativos, que requieren una revisión adicional para poder determinar si coinciden con los registros maestros, se cargan en una base de datos DB2.

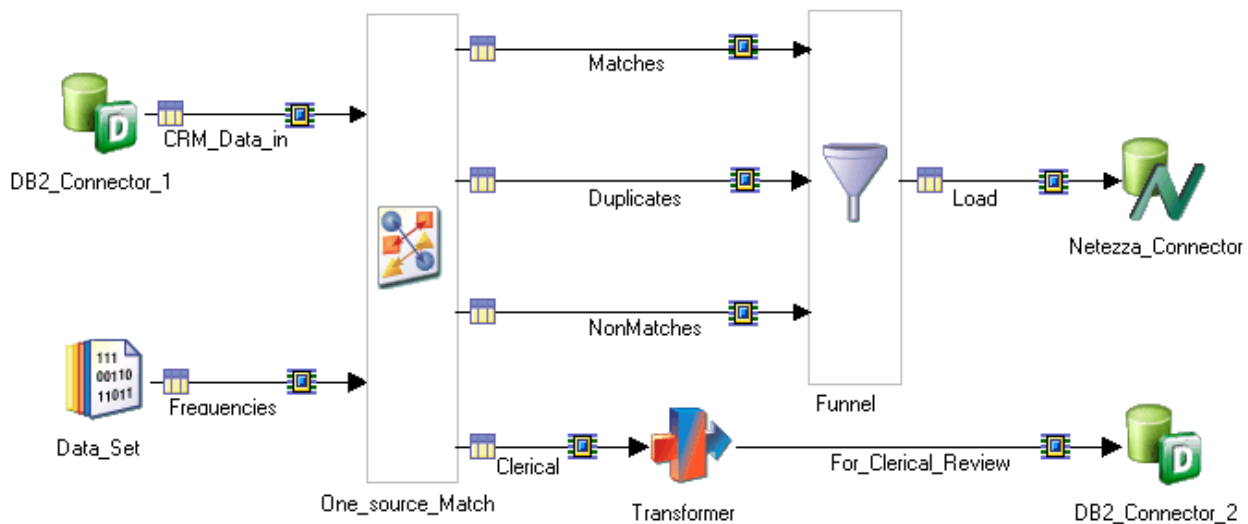


Figura 28. Hacer coincidir los registros para identificar los duplicados

Validar datos con reglas de datos

Puede validar los datos y asegurarse de que la calidad de los datos se ajuste a las expectativas empresariales relativas a la limpieza de datos, utilizando la etapa Data Rules, que proporciona IBM InfoSphere Information Analyzer. Por ejemplo, puede comprobar que un proveedor tiene un ID de proveedor en el formato correcto, un nombre de proveedor y un ID de impuesto que tenga nueve caracteres numéricos. Las reglas de datos que se utilizan en la etapa Data Rules suele crearlas un analista de datos utilizando InfoSphere Information Analyzer.

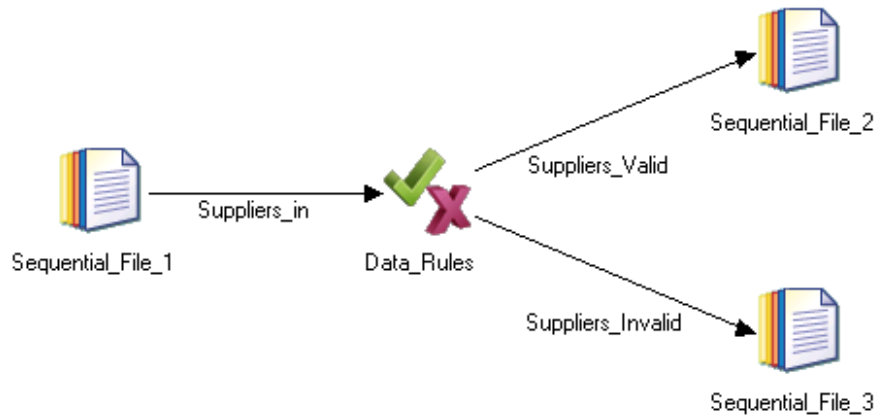


Figura 29. Validar datos con reglas de datos

Conceptos relacionados:

- Establecer datos coherentes mediante estandarización
- Ampliación de conjuntos de reglas de estandarización utilizando Standardization Rules Designer
- IBM InfoSphere QualityStage Address Verification Interface
- Coincidencia de datos
- Etapa Funnel
- Etapa Data Rules
- InfoSphere QualityStage
- InfoSphere Information Analyzer

Proceso en tiempo real

Puede desarrollar trabajos para la integración de datos en tiempo real. En estos ejemplos se muestra cómo puede responder a una solicitud de servicio web y verificar los datos entre las colas de mensajes.

- “Responder a una solicitud de servicio web”
- “Verificar los datos entre colas de mensajes” en la página 22

Responder a una solicitud de servicio web

Puede responder a una solicitud de servicio web en tiempo real. Por ejemplo, puede devolver un número de cuenta después de que un cliente envíe la información de cuenta necesaria utilizando las etapas que enlacen a IBM InfoSphere Information Services Director. En este trabajo de ejemplo, los datos de la cuenta llegan en tiempo real a través de una etapa ISD Input. Los datos se enriquecen con el número de cuenta, utilizando una etapa Lookup. Los datos enriquecidos se devuelven en tiempo real a través de una etapa ISD Output. Al utilizar InfoSphere Information Services Director, puede desplegar los trabajos que funcionen como servicios de información para aplicaciones orientadas a servicios.

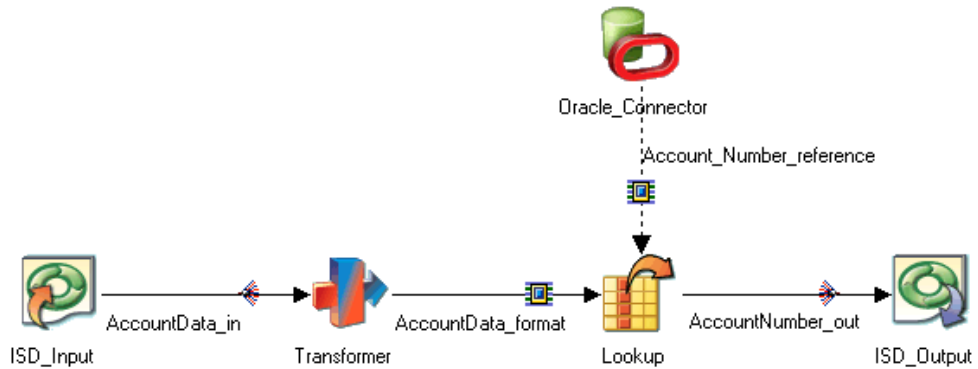


Figura 30. Extraer y cargar datos con servicios web

Verificar los datos entre colas de mensajes

Puede desarrollar un trabajo para recibir datos de una cola de mensajería, comprobar los datos y cargarlos en otra cola de mensajes. En este ejemplo, el trabajo utiliza la etapa IBM WebSphere MQ Connector para conectarse a las colas de mensajes. El trabajo estandariza las direcciones y, a continuación, las valida utilizando una etapa Data Rules. Las direcciones no válidas se envían a una tabla para los registros rechazados, y las válidas se cargan en la cola de mensajes.

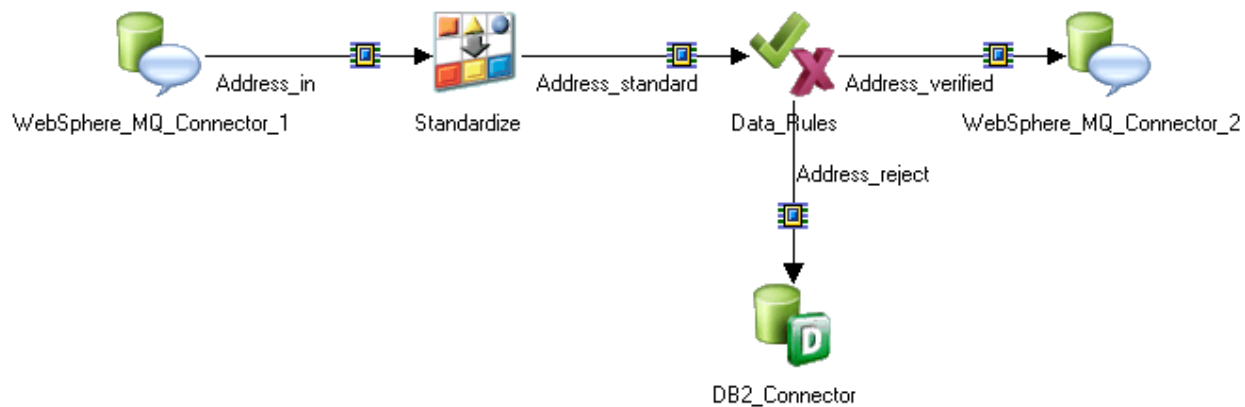


Figura 31. Verificar los datos entre colas de mensajes

Conceptos relacionados:

- ➡ Etapa Lookup
- ➡ Conector de IBM WebSphere MQ
- ➡ Establecer datos coherentes mediante estandarización
- ➡ Etapa Data Rules
- ➡ Diseño de trabajos de InfoSphere DataStage y QualityStage como servicios
- ➡ InfoSphere Information Services Director

Proceso de datos grandes

Puede desarrollar trabajos que intercambien datos con orígenes de datos grandes. En estos ejemplos se muestra cómo puede acceder al sistema de archivos distribuido Hadoop (HDFS) y aumentar la analítica basada en Hadoop.

- “Acceder a datos en HDFS”
- “Aumentar los datos con la analítica basadas en Hadoop”

Acceder a datos en HDFS

Puede acceder a los archivos que se encuentren en HDFS. Este trabajo de ejemplo accede a los pedidos de los archivos HDFS utilizando la etapa Big Data File. El trabajo utiliza una etapa Transformer para seleccionar un subconjunto de los pedidos, combina éstos con los detalles de pedido, y graba los artículos solicitados a los archivos HDFS posteriores. Puede desplegar este trabajo directamente en InfoSphere DataStage, que proporciona una escalabilidad masiva mediante la ejecución de trabajos en el motor paralelo de InfoSphere Information Server. Como alternativa, puede utilizar IBM InfoSphere DataStage Balanced Optimization para procesar esta lógica dentro del clúster Hadoop. La lógica del trabajo se representa como scripts MapReduce.

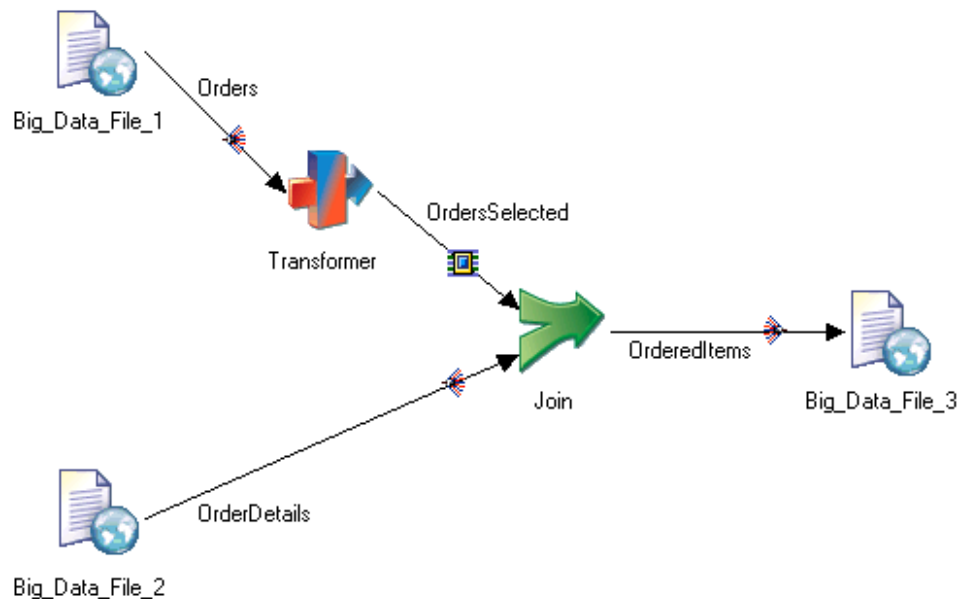


Figura 32. Acceder a datos en HDFS

Aumentar los datos con la analítica basadas en Hadoop

Puede aumentar los datos en un almacén de datos con los resultados analíticos basados en Hadoop. En este trabajo de ejemplo se mueven los datos analíticos de un sistema de almacén de datos Hive a un almacén de datos Netezza.

La etapa Hive se ejecuta al principio la etapa Java™ Integration, y proporciona un conector Hive para InfoSphere DataStage. La etapa Hive forma parte del código de ejemplo Hive para la etapa Java Integration, que está disponible desde InfoSphere Information Server y InfoSphere Discovery Exchange en IBM developerWorks. Otros aceleradores de conector que están disponibles en developerWorks amplían la etapa Java Integration con la lógica personalizada para estos orígenes de datos:

HBase, Java Message Service (JMS), MongoDB, y Cassandra.



Figura 33. Aumento de datos con la analítica basada en Hadoop

Conceptos relacionados:

- ☞ Etapa Big Data File
- ☞ InfoSphere DataStage Balanced Optimization

Información relacionada:

- ☞ InfoSphere Information Server and InfoSphere Discovery Exchange
- ☞ IBM developerWorks
- ☞ ¿Qué es Hadoop?

Combinar trabajos en un trabajo de secuencia

Puede combinar los trabajos en un trabajo de secuencia, para desarrollar patrones de proceso mayores.

Puede desarrollar trabajos de secuencia para que ejecuten varios trabajos en secuencia, y para que integren controles de programación como, por ejemplo, ramificaciones y bucles en el flujo de trabajo. Puede especificar la información de control, como los distintos cursos de acción que deben realizarse, dependiendo de si un trabajo de la secuencia resulta satisfactorio, o falla. Cuando los trabajos se combinan en un trabajo de secuencia, resulta más sencillo resolver los problemas que surjan. Los trabajos de secuencia también proporcionan las posibilidades de reinicio de comprobación, de modo que tras el reinicio, el proceso continúa después del último paso satisfactorio en el flujo de trabajo.

En este trabajo de ejemplo se ejecutan tres trabajos en secuencia, utilizando la etapa Job Activity. El primer trabajo extrae los datos. El segundo trabajo los transforma. El tercer trabajo los carga. La última etapa Job Activity también envía mensajes de aviso para que se distribuyan utilizando una actividad de notificación. Las excepciones de las actividades del trabajo se procesan utilizando una etapa Exception Handler, que envía los errores que deben procesarse utilizando una etapa Routine Activity.

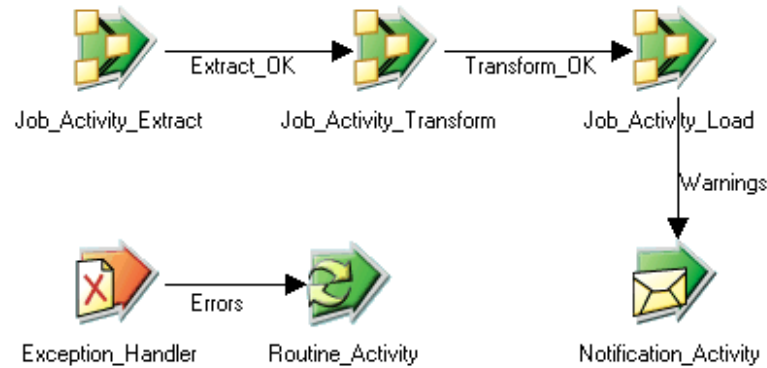


Figura 34. Combinar trabajos en un trabajo de secuencia

Conceptos relacionados:

- ➡ Creación de trabajos de secuencia
- ➡ Actividades de trabajo de secuencia

Integración de InfoSphere DataStage en InfoSphere Information Server

Puede utilizar las posibilidades de otros componentes de IBM InfoSphere Information Server a medida que desarrolla y despliega trabajos de IBM InfoSphere DataStage.

InfoSphere DataStage está integrado con los demás componentes de InfoSphere Information Server en los dominios de integración de datos, la calidad de los datos, y la gobernabilidad de los datos. A continuación figuran algunos ejemplos de cómo puede utilizar los demás componentes.

Colaborar a través de proyectos utilizando el repositorio de metadatos

El repositorio de metadatos sirve como punto de colaboración para los usuarios de los componentes de InfoSphere Information Server. Puede investigar los metadatos en el repositorio de varios proyectos para dar soporte a los análisis de impacto y el linaje de datos, para que la organización comprenda cómo se utiliza la información.

Por ejemplo, un analista de datos de su empresa puede perfilar los datos y descubrir que algunos datos de un recurso específico se han dañado. Como desarrollador, puede revisar el perfil de datos, y desarrollar un trabajo para transforme los datos y corrija el problema. Un encargado de datos de su empresa puede ejecutar un informe para revisar el linaje de datos y evaluar cómo se está utilizando los datos. Cualquier persona de su empresa que tenga autorización para trabajar en el proyecto puede revisar quién ha perfilado los datos y quién ha desarrollado el trabajo.

Cuando esté desarrollando un trabajo en InfoSphere DataStage, puede utilizar los metadatos que se hayan compartido de varios proyectos. Puede importar metadatos de tablas, vistas y procedimientos almacenados para utilizarlos en sus diseños de trabajo. Los trabajos y metadatos se incluyen automáticamente en el repositorio de metadatos.

Determinar los requisitos de desarrollo para un proyecto mediante IBM InfoSphere Blueprint Director

Un arquitecto de empresa puede crear un plan o panorama de la información para un proyecto mediante InfoSphere Blueprint Director. Como desarrollador del proyecto, puede revisar el plan, incluyendo las actividades y tareas contenidas en cada fase, para identificar los requisitos de desarrollo. Cuando el plan esté vinculado a metadatos reales, podrá identificar los problemas potenciales en la terminología, lagunas en el proceso de información, o áreas para mejorar la reutilización y la coherencia.

Colaborar con analistas empresariales para impulsar los diseños de trabajo mediante IBM InfoSphere FastTrack

Un analista empresarial puede crear especificaciones para las correlaciones de origen a destino mediante InfoSphere FastTrack. Las correlaciones en las especificaciones pueden contener transformaciones de valor de datos que definen cómo crear aplicaciones. El analista empresarial puede generar trabajos a partir de las especificaciones. Como desarrollador, puede utilizar esos trabajos como punto de partida para los trabajos de InfoSphere DataStage.



Desarrollar reglas de datos mediante IBM InfoSphere Information Analyzer

Puede utilizar las reglas de datos de InfoSphere Information Analyzer para mejorar la calidad de los datos en sus trabajos de InfoSphere DataStage. Por ejemplo, un analista de datos puede utilizar InfoSphere Information Analyzer para perfilar orígenes de datos para desarrollar reglas de calidad de datos. A continuación, podrá utilizar dichas reglas de datos en sus trabajos de InfoSphere DataStage. Como desarrollador, también puede crear reglas directamente en el diseñador de trabajo y, a continuación, compartir dichas reglas para que los analistas empresariales también puede utilizarlas en InfoSphere Information Analyzer.

Desplegar trabajos como servicios mediante IBM InfoSphere Information Services Director

Puede desplegar trabajos que funcionen como servicios para aplicaciones orientadas a servicios mediante InfoSphere Information Services Director. Por ejemplo, puede desarrollar un trabajo que utilice las etapas de entrada y salida de InfoSphere Information Services Director. A continuación, puede publicar la lógica de integración de datos como servicios compartidos, que se pueden invocar mediante REST, SOAP u otros enlaces, y reutilizarse en toda la empresa. InfoSphere Information Services Director procesa las solicitudes de servicio de sus trabajos de InfoSphere DataStage.

Conceptos relacionados:

-  [InfoSphere Information Governance Catalog](#)
-  [InfoSphere Blueprint Director](#)
-  [InfoSphere FastTrack](#)
-  [InfoSphere Information Analyzer](#)
-  [InfoSphere Information Services Director](#)
-  [Arquitectura y conceptos de IBM InfoSphere Information Server](#)

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.

Documentación sobre accesibilidad

Se proporciona documentación accesible para los productos InfoSphere Information Server en un centro de información. El centro de información presenta la documentación en formato XHTML 1.0, que se puede ver en la mayoría de navegadores web. Dado que el centro de información 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 el centro de información 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 3. 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 www.ibm.com/support/entry/portal/Software/Information_Management/InfoSphere_Information_Server
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 www.ibm.com/businesssolutions/
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 www.ibm.com/account/
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 http://www.ibm.com/training
Representantes de IBM	Puede contactar con un representante de IBM para obtener información sobre soluciones en www.ibm.com/connect/ibm/us/en/

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

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

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>.

Apéndice D. Cómo proporcionar comentarios sobre la documentación del producto

Puede proporcionar comentarios y observaciones útiles sobre la documentación de IBM.

Sus comentarios ayudarán a IBM a ofrecer información de calidad. Puede utilizar cualquiera de los métodos siguientes para proporcionar comentarios:

- Para hacer un comentario sobre un tema de IBM Knowledge Center que esté alojado en el sitio web de IBM, inicie la sesión y añada un comentario pulsando el botón **Añadir comentario** en la parte inferior del tema. Los comentarios enviados de esta forma son visibles para el público.
- Para enviar un comentario sobre el tema de IBM Knowledge Center a IBM que no pueda ver nadie más, inicie la sesión y pulse el enlace **Comentarios** en la parte inferior de la página de IBM Knowledge Center.
- Envíe sus comentarios utilizando el formulario de comentarios del lector que encontrará en www.ibm.com/software/awdtools/rcf/.
- Envíe sus comentarios por correo electrónico a hojacom@es.ibm.com. Incluya el nombre y el número de versión del producto, así como el nombre y el número de pieza de la información (si es pertinente). Si su comentario es sobre un texto específico, incluya la ubicación del texto (por ejemplo, un título, un número de tabla o un número de página).

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. Consulte 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 4. 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"> Sesión Persistente 	Nombre de usuario	<ul style="list-style-type: none"> Gestión de sesiones Autenticación 	No se pueden inhabilitar

Tabla 4. 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"> Sesión Persistente 	Ninguna información de identificación personal	<ul style="list-style-type: none"> Gestión de sesiones Autenticación Usabilidad de usuario mejorada Configuración de inicio de sesión único 	No se pueden inhabilitar
InfoSphere DataStage	Etapas Big Data File	<ul style="list-style-type: none"> Sesión Persistente 	<ul style="list-style-type: none"> Nombre de usuario Firma digital ID de sesión 	<ul style="list-style-type: none"> Gestión de sesiones Autenticación Configuración de inicio de sesión único 	No se pueden inhabilitar
InfoSphere DataStage	Etapas XML	Sesión	Identificadores internos	<ul style="list-style-type: none"> Gestión de sesiones Autenticación 	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"> Gestión de sesiones Autenticación 	No se pueden inhabilitar
InfoSphere Data Click	Consola web de InfoSphere Information Server	<ul style="list-style-type: none"> Sesión Persistente 	Nombre de usuario	<ul style="list-style-type: none"> Gestión de sesiones Autenticación 	No se pueden inhabilitar
InfoSphere Data Quality Console		Sesión	Ninguna información de identificación personal	<ul style="list-style-type: none"> Gestión de sesiones Autenticación Configuración de inicio de sesión único 	No se pueden inhabilitar
InfoSphere QualityStage Standardization Rules Designer	Consola web de InfoSphere Information Server	<ul style="list-style-type: none"> Sesión Persistente 	Nombre de usuario	<ul style="list-style-type: none"> Gestión de sesiones Autenticación 	No se pueden inhabilitar
InfoSphere Information Governance Catalog		<ul style="list-style-type: none"> Sesión Persistente 	<ul style="list-style-type: none"> Nombre de usuario Identificadores internos Estado del árbol 	<ul style="list-style-type: none"> Gestión de sesiones Autenticación Configuración de inicio de sesión único 	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.

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 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^{Link}, 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^{Link} 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

- ABAP Extract, etapa 10
- accesibilidad de los productos
 - accesibilidad 27
- Address Verification, etapa 18
- Aggregator, etapa 14
- avisos legales 35

B

- Balanced Optimization 23
- Big Data File, etapa 23
- Blueprint Director 25

C

- cargar
 - trabajos 10
- Cassandra 23
- CDC Transaction, etapa 10
- Change Capture, etapa 14
- colas de mensajes
 - verificar datos entre 21
- consola de operaciones 9
- Consola de operaciones de DataStage 9
- coordinador de release 1

D

- Data Rules, etapa 18
- Data Set, etapa 10, 14, 18
- datos
 - enriquecimiento 17
 - estandarizar 18
 - limpiar 18
- datos de referencia 17
- datos grandes
 - trabajos 23
- DB2 Connector, etapa 10, 17, 18, 21
- depurar 5
- desarrollador 1
- desarrollar
 - trabajos 3
- despliegue
 - paquete 7
 - trabajos 7
- direcciones
 - verificar 18
- documentación del producto
 - acceder 31

E

- enriquecer datos
 - trabajos 17
- estandarizar
 - datos 18
- etapas
 - ABAP Extract 10

etapas (continuación)

- Address Verification 18
- Aggregator 14
- Big Data File 23
- CDC Transaction 10
- Change Capture 14
- Data Rules 18
- Data Set 10, 14, 18
- DB2 Connector 10, 17, 18, 21
- depurar 5
- Exception Handler 24
- Hive 23
- ILOG JRules 17
- ISD Input 21
- ISD Output 21
- Java Integration 23
- Job Activity 24
- Join 14, 23
- JRules 17
- Lookup 17, 21
- Match Frequency 18
- Netezza Connector 10, 14, 18, 23
- Notification Activity 24
- One Source Match 18
- Oracle Connector 10, 14, 17, 21
- Peak 5
- Pivot Enterprise 14
- Routine Activity 24
- Row Generator 5
- SAP BW 10
- Sequential File 10, 14, 17, 18
- Slowly Changing Dimension 10
- Sort 10
- Standardize 18, 21
- Teradata Connector 14
- Transformer 10, 14, 18, 21, 23
- Web Services Transformer 10
- WebSphere MQ Connector 10, 21
- XML 10
- Exception Handler, etapa 24
- extraer
 - trabajos 10

F

- FastTrack 25

G

- generar datos de prueba 5

H

- Hadoop 23
- HBase 23
- HDFS 23
- herramienta de integración de datos 1
- Hive, etapa 23

I

- ILOG JRules, etapa 17
- Information Analyzer 25
- Information Services Director 25
- InfoSphere Blueprint Director 25
- InfoSphere DataStage
 - ciclo de vida de trabajo 1
 - visión general 1
- InfoSphere DataStage Balanced Optimization 23
- InfoSphere FastTrack 25
- InfoSphere Information Analyzer 25
- InfoSphere Information Server
 - integración 25
- InfoSphere Information Services Director 21, 25
- integración de la información 1
- ISD Input, etapa 21
- ISD Output, etapa 21

J

- Java Integration, etapa 23
- Java Message Service (JMS) 23
- Job Activity, etapa 24
- Join, etapa 14, 23
- JRules, etapa 17

L

- limpiar datos
 - trabajos 18
- Lookup, etapa 17, 21

M

- marcas registradas
 - lista de 35
- Match Frequency, etapa 18
- MongoDB 23

N

- Netezza Connector, etapa 10, 14, 18, 23
- Notification Activity, etapa 24

O

- One Source Match, etapa 18
- operador 1
- Oracle Connector, etapa 10, 14, 17, 21

P

- Peak, etapa 5
- Pivot Enterprise, etapa 14
- probar
 - trabajos 5

- proceso en tiempo real
 - trabajos 21
- puntos de interrupción 5

R

- reglas de datos 25
- reglas empresariales
 - aplicar 17
- repositorio de metadatos 25
- Routine Activity, etapa 24
- Row Generator, etapa 5

S

- SAP 10
- SAP Business Warehouse 10
- SAP BW, etapa 10
- secuencia
 - trabajos 24
- Sequential File, etapa 10, 14, 17, 18
- servicios de software
 - contactar 29
- servicios web 21
- sistema de archivos distribuido
 - Hadoop 23
- Slowly Changing Dimension, etapa 10
- soporte
 - cliente 29
- soporte al cliente
 - contactar 29
- Sort, etapa 10
- Standardize, etapa 18, 21

T

- Teradata Connector, etapa 14
- trabajos
 - ciclo de vida 1
 - combinar 24
 - desarrollar 3
 - desplegar 7
 - diseños 10
 - ejemplos 10
 - probar 5
- transformación
 - trabajos 14
- Transformer, etapa 10, 14, 18, 21, 23

V

- visión general 1

W

- Web Services Transformer, etapa 10
- WebSphere MQ Connector, etapa 10, 21

X

- XML, etapa 10



Impreso en España

GC43-1245-00

