

IBM InfoSphere DataStage and Quality Stage
Versión 11 Release 3

*Guía para el acceso a datos no
estructurados*



IBM InfoSphere DataStage and Quality Stage
Versión 11 Release 3

*Guía para el acceso a datos no
estructurados*



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 67.

Contenido

Capítulo 1. Etapa Unstructured Data . . . 1

Diseño de trabajos con etapas Unstructured Data . . .	1
Definición de un trabajo que incluya una etapa Unstructured Data	1
Extracción de datos de Microsoft Excel	1
Grabación de datos en un archivo de Microsoft Excel nuevo	18
Grabación de datos en archivos existentes de Microsoft Excel	32

Capítulo 2. Referencia 39

Conversiones de tipos de datos desde Microsoft Excel a InfoSphere DataStage	39
Conversiones de tipos de datos de InfoSphere DataStage a Microsoft Excel	45
Condiciones de terminación anómala de trabajo en Microsoft Excel	46

Capítulo 3. Resolución de problemas 49

Capítulo 4. Variables de entorno: Etapa Unstructured Data 53

CC_JNI_EXT_DIRS	53
CC_JVM_OPTIONS	53
CC_JVM_OVERRIDE_OPTIONS	53
CC_IGNORE_TIME_LENGTH_AND_SCALE	53

CC_MSG_LEVEL	54
CC_UNST_JAVA_HEAP	54

Apéndice A. Accesibilidad de los productos 55

Apéndice B. Lectura de la sintaxis de la línea de mandatos. 57

Apéndice C. Cómo leer diagramas de sintaxis 59

Apéndice D. Cómo ponerse en contacto con IBM 61

Apéndice E. Acceso a la documentación del producto 63

Apéndice F. Cómo aportar comentarios sobre la documentación del producto . 65

Avisos y marcas registradas 67

Índice 73

Capítulo 1. Etapa Unstructured Data

Los datos no estructurados son información que no tiene un modelo de datos predefinido o que no encaja bien en tablas relacionales. Son ejemplos de datos no estructurados el texto procedente de libros, periódicos, metadatos, audio, archivos de vídeo, el cuerpo de documentos de un procesador de textos, páginas web y diagramas de presentación. En este release, la etapa Unstructured Data solamente puede utilizar archivos de Microsoft Excel como orígenes de datos.

Utilice la etapa Unstructured Data para realizar las operaciones siguientes:

- Extraer información de orígenes de datos no estructurados e integrar la información en sus trabajos. De una sola una etapa Unstructured Data se pueden extraer datos de las hojas de Microsoft Excel que tengan definiciones de columna diferentes.
- Grabar datos en hojas de Microsoft Excel.

Diseño de trabajos con etapas Unstructured Data

Puede utilizar las etapas Unstructured Data en los trabajos para leer los orígenes de Unstructured Data o grabar datos en los orígenes de Unstructured Data en los contextos de dichos trabajos.

Definición de un trabajo que incluya una etapa Unstructured Data

Para poder leer datos desde, o grabarlos a, archivos de Microsoft Excel, debe crear un trabajo que incluya la etapa Unstructured Data, añadir las etapas adicionales necesarias y crear los enlaces necesarios.

Procedimiento

1. En el Cliente del Diseñador, pulse **Archivo > Nuevo**.
2. En la ventana Nuevo, pulse el icono **Trabajo paralelo** y luego pulse **Aceptar**.
3. En la Paleta, pulse **Archivo**.
4. Arrastre el icono **Etapa Unstructured Data** al lienzo.
5. Crear etapas para el trabajo.
6. En el lado izquierdo del cliente del Diseñador, en el menú Paleta, seleccione la categoría **General** y, a continuación, cree los enlaces necesarios para el trabajo.
7. (Opcional) Efectúe una doble pulsación en el icono **Etapa Unstructured Data** para especificar o modificar los atributos siguientes:
 - **Etapa:** Modifique el nombre predeterminado de la **Etapa**. Puede entrar hasta 255 caracteres. Como alternativa, puede modificar el nombre de la etapa en el lienzo de diseño del trabajo.
 - **Descripción:** Escriba una descripción de la etapa.
8. Pulse **Guardar**.

Extracción de datos de Microsoft Excel

Puede utilizar la etapa Unstructured Data para extraer varios tipos de datos de un archivo de Microsoft Excel.

Requisitos previos

Antes de iniciar la instalación y configuración de la etapa Unstructured Data, debe cumplir los requisitos del sistema y haber instalado todo el software previo necesario.

Antes de empezar

Procedimiento

1. Instale Information Server perteneciente al idioma que coincida con el idioma del archivo de Microsoft Excel que desee extraer.
2. Asegúrese de que el programa Visor de Microsoft Excel que muestra el contenido de las hojas de cálculo de Microsoft Excel (por ejemplo, Microsoft Excel, Microsoft Excel Viewer o IBM Lotus Symphony) está instalado en la máquina cliente.
3. Asegúrese de que las extensiones de archivo .xls y .xlsx están asociadas debidamente al programa Visor de Microsoft Excel utilizado.

Orígenes de datos soportados

La etapa Unstructured Data solamente soporta archivos de Microsoft Excel como archivo de origen.

La etapa Unstructured Data soporta los formatos de archivo siguientes:

- Microsoft Excel 97-2003 OLE2 (.xls), incluido el soporte para archivos cifrados por contraseña.
- Microsoft Excel 2007-2010 OOXML (.xlsx), incluido el soporte para archivos cifrados por contraseña.

La etapa Unstructured Data no soporta archivos de Microsoft Excel que se han creado mediante Microsoft Excel para Mac.

Rangos de datos

Cuando utiliza la etapa Unstructured Data, puede extraer datos de una hoja de cálculo Microsoft Excel pertenecientes a un rango de datos especificado.

El rango de datos representa una celda, una fila, una columna, o una selección de celdas que contienen uno o más bloques continuos de celdas. El rango de datos se especifica mediante la expresión de rango. En la etapa Unstructured Data, puede utilizar una expresión de rango para especificar el rango de datos que se debe extraer.

Por ejemplo, Employee_Salary!A1:G8 describe un rango de datos en el que la primera celda es A1 y la última celda es G8 en la hoja de cálculo Employee_Salary.

Tabla 1. Ejemplo de archivo de Microsoft Excel: hoja de cálculo Employee_Salary

	A	B	C	D	E	F	G
1	EMPNO	FIRSTNAME	LASTNAME	DEPT	JOB	SALARY	BONUS
2	20	MICHAEL	THOMPSON	B01	MANAGER	94250	800
3	30	SALLY	KWAN	C01	MANAGER	98250	800
4	60	IRVING	STERN	D11	MANAGER	72250	500
5	70	EVA	PULASKI	D21	MANAGER	96170	700
6	50	JOHN	GEYER	E01	MANAGER	80175	800
7	90	ELEEN	HENDERSON	E11	MANAGER	89750	600

Tabla 1. Ejemplo de archivo de Microsoft Excel: hoja de cálculo Employee_Salary (continuación)

	A	B	C	D	E	F	G
1	EMPNO	FIRSTNAME	LASTNAME	DEPT	JOB	SALARY	BONUS
8	100	THEODORE	SPENSER	E21	MANAGER	86150	500

La etapa Unstructured Data correlaciona la fila y columna de Microsoft Excel perteneciente al rango de datos especificado con la fila y columna de InfoSphere DataStage y extrae los registros.

La tabla siguiente describe los registros extraídos por la etapa Unstructured Data cuando la expresión de rango es Employee_Salary!A2:G8.

Tabla 2. Ejemplo de fila y columna de DataStage

20	MICHAEL	THOMPSON	B01	MANAGER	94250	800
30	SALLY	KWAN	C01	MANAGER	98250	800
60	IRVING	STERN	D11	MANAGER	72250	500
70	EVA	PULASKI	D21	MANAGER	96170	700
50	JOHN	GEYER	E01	MANAGER	80175	800
90	ELEEN	HENDERSON	E11	MANAGER	89750	600
100	THEODORE	SPENSER	E21	MANAGER	86150	500

La propiedad **Opción de rango** de la etapa Unstructured Data le permite especificar el rango de datos seleccionando la opción **Especificar la fila inicial** o la opción **Especificar todo el rango de datos**. Si selecciona la opción **Especificar la fila inicial**, identifique la fila inicial. La etapa Unstructured Data identificará entonces la fila final del rango de datos. Si selecciona la opción **Especificar todo el rango de datos**, debe especificar las filas inicial y final del rango de datos que se debe extraer.

Si desea utilizar el valor de las celdas de la primera fila como nombre de columna de IBM® InfoSphere DataStage, puede utilizar la propiedad **Cabecera de columna**. Si la propiedad **Cabecera de columna** se establece en **Primera fila de rangos de datos** y especifica Employee_Salary!A1:G8 como expresión de rango, la primera fila se trata como cabecera, y el valor de las celdas de la primera fila se utiliza como nombre de columna predeterminado de DataStage para el trabajo. Puede crear la expresión de rango durante la fase de diseño mediante la etapa Unstructured Data.

Tipos de datos que se pueden extraer de Microsoft Excel

Puede utilizar la etapa Unstructured Data para extraer varios tipos de datos de un archivo de Microsoft Excel.

Propiedades de archivo

La tabla siguiente muestra la información que se puede extraer como propiedades de archivo:

Tabla 3. Datos que se pueden extraer como propiedades de archivo

Datos	Descripción
Nombre de archivo	Nombre del archivo. Por ejemplo: Workbook1.xls
Vía de acceso de archivo	Vía de acceso del archivo. Por ejemplo: C:\excel\Workbook1.xls

Tabla 3. Datos que se pueden extraer como propiedades de archivo (continuación)

Datos	Descripción
Tamaño de archivo	Tamaño del archivo en bytes.
Fecha de última modificación	Fecha y hora en que se modificó el archivo por última vez.

Propiedades de documento

La tabla siguiente muestra la información que se puede extraer como propiedades de documento:

Tabla 4. Datos que se pueden extraer como propiedades de documento

Datos	Descripción
Autores	Autores del documento.
Comentarios de documento	Comentarios del documento.
Fecha de creación del contenido	Fecha y hora en que se creó el documento.
Palabras clave	Palabras clave del documento.
Número de revisión	Número de revisión del documento.
Asunto	Asunto del documento.
Título	Título del documento.
Empresa	Valor de la propiedad Empresa del documento.
Categoría	Categoría del documento.
Gestor	Gestor del documento.
Propiedades personalizadas	Propiedades personalizadas del documento. Debe especificar el nombre de la propiedad personalizada que se debe extraer.

Información de hoja

La tabla siguiente muestra la información que se puede extraer como información de hoja:

Tabla 5. Datos que se pueden extraer como información de hoja

Datos	Descripción
Nombre de hoja	Nombre de la hoja de Microsoft Excel.
Cabecera (izquierda, centro, derecha)	Cabecera de la posición especificada.
Pie de página (izquierda, centro, derecha)	Pie de página de la posición especificada.

Información de fila

La tabla siguiente muestra la información que se puede extraer como información de fila:

Tabla 6. Datos que se pueden extraer como información de fila

Datos	Descripción
Número de fila	Número de fila de Microsoft Excel dentro de la hoja. El número de la primera fila es 1.
Está oculta	Indicación de si la fila está oculta o no. Escribe True si la fila u hoja a la que pertenece la fila está oculta.

Información de celda

Puede extraer la información de celda de acuerdo con la columna de Microsoft Excel o posición de la celda. Puede especificar la columna de origen de Microsoft Excel de acuerdo con la posición relativa dentro del rango de datos al extraer la información de celda de acuerdo con la columna de Microsoft Excel.

La tabla siguiente muestra la información que se puede extraer como información de celda:

Tabla 7. Datos que se pueden extraer como información de celda

Datos	Descripción
Valor	Valor de una celda. Si la celda tiene una fórmula, la etapa extrae el valor de la memoria caché.
Comentario	Comentario de una celda.
Autor de comentario	Autor del comentario de una celda.
Fórmula	Fórmula de una celda en el texto.
Tipo de hipervínculo	Tipo de hipervínculo de una celda.
Dirección de hipervínculo	Dirección a la que apunta el hipervínculo. El formato depende del tipo de este hipervínculo.
Etiqueta de hipervínculo	Etiqueta de texto del hipervínculo.

Diseño de trabajos que extraen datos de archivos de Microsoft Excel

Puede utilizar la etapa Unstructured Data para diseñar trabajos que lean datos no estructurados de archivos de Microsoft Excel.

Acerca de esta tarea

La etapa Unstructured Data lee datos de los archivos Microsoft Excel y pasa las filas a una etapa Transformer. La etapa Transformer transforma los datos y los carga en el conector ODBC. Cuando configura la etapa Unstructured Data para leer datos de archivos de Microsoft Excel, crea un solo enlace de salida.

Procedimiento

1. Defina un trabajo que incluya una etapa Unstructured Data.
2. Para configurar la etapa Unstructured Data como una etapa de origen para leer los datos no estructurados de archivos de Microsoft Excel, complete los pasos siguientes:
 - a. Configure la etapa Unstructured Data como origen.
 - b. Modifique la definición de columna en el enlace.
3. Compile y ejecute el trabajo.

Configuración de la etapa Unstructured Data como origen:

Cuando crea un trabajo de la etapa Unstructured Data, debe configurar la etapa Unstructured Data para que extraiga los datos y genere datos de salida del tipo solicitado por el usuario.

Procedimiento

1. En el lienzo paralelo, efectúe una doble pulsación en la etapa **Unstructured Data**.
2. En la lista **Tipo de documento**, seleccione **Excel**.
3. Pulse **Configurar** para configurar propiedades adicionales, y definir la correlación entre elementos de Microsoft Excel y columnas de DataStage.
4. Especifique los detalles del nombre de archivo en el panel Origen de datos:
 - a. En el campo **Nombre de archivo**, especifique el nombre del archivo del que desee leer los datos. La compilación del trabajo falla si este campo está vacío. Si el archivo está protegido por contraseña, especifique la contraseña en el campo **Contraseña**.
 - b. Opcional: Si especifica caracteres comodín en el nombre de archivo, seleccione **Utilizar archivo de plantilla para tiempo de diseño** y especifique un nombre de archivo de plantilla. El archivo de plantilla se utiliza para pasos de configuración posteriores, y no se utiliza en tiempo de ejecución. Especifique un valor para **Contraseña de plantilla** si el archivo de plantilla especificado está protegido por contraseña.
 - c. Opcional: Pulse **Ver** para iniciar el programa visor externo de Microsoft Excel. Puede verificar el contenido del archivo de Microsoft Excel con el que está trabajando.
 - d. Pulse **Cargar**.
5. Especifique los detalles del rango de datos que deben leerse del archivo de Microsoft Excel, en el panel Opciones de lectura.
 - a. Opcional: Especifique un valor para **Opción de rango**. Si selecciona **Especificar la fila inicial**, sólo necesita especificar la primera fila. La etapa Unstructured Data encuentra la última fila en tiempo de ejecución. Si selecciona **Especificar todo el rango de datos**, debe especificar la fila inicial y la fila final.
 - b. Opcional: Especifique **Expresión de rango**. **Expresión de rango** es una propiedad necesaria en el momento de la ejecución, pero puede estar vacía cuando se pulsa el botón **Cargar**. La etapa Unstructured Data examina el documento completo y lista los rangos de datos posibles en el recuadro de lista **Rango de datos de plantilla** en el panel **Importar**. La propiedad **Expresión de rango** toma el valor adecuado al pulsar **Importar** en el panel **Importar**.
 - c. Opcional: Si desea omitir nombres de hoja de cálculo de la expresión de rango, especifique el nombre en el campo **Especificar nombres de hoja para omitir**. Utilice este campo cuando los nombres de hojas se omiten de la expresión de rango.
 - d. Opcional: Especifique Primera fila de rangos de datos. En tiempo de diseño, si selecciona **Ninguno**, los nombres de columna de Microsoft Excel se expresan en el formato: "Column#*número de columna*(*Columna*etiqueta de *columna Excel*)" en el panel Correlacionar. Si selecciona **La primera fila es la cabecera**, el valor de la primera fila se visualiza en el panel Correlacionar. En el tiempo de ejecución, si selecciona **Ninguna**, se extrae la primera fila. Si selecciona **La primera fila es la cabecera**, la primera fila se pasa por alto.
6. Especifique los detalles del rango de datos que deban importarse en el panel Importar.
 - a. Seleccione un rango de datos en **Rango de datos de plantilla**.
 - b. Opcional: Si desea extraer elementos adicionales de Microsoft Excel tales como propiedades de documento, seleccione el separador **Propiedad** y seleccione los elementos que deban extraerse.

- c. Pulse **Importar**.
7. Defina la correlación entre las columnas de InfoSphere DataStage y los elementos de Microsoft Excel importados, en el panel Correlacionar.
 - a. Defina la correlación entre las columnas de InfoSphere DataStage y los elementos de Microsoft Excel importados. Puede añadir correlaciones de columnas de InfoSphere DataStage o cambiar el orden de las columnas pulsando los botones **Subir**, **Bajar**, **Insertar** o **Suprimir**. En **Columna de InfoSphere DataStage**, especifique el nombre de la columna de InfoSphere DataStage correspondiente a cada elemento de Microsoft Excel. En **Elemento de Microsoft Excel**, puede seleccionar el elemento que desee correlacionar con la columna de InfoSphere DataStage. Todos los elementos que se pueden seleccionar en el panel Importar están listados en cada célula. En **Opción de importación**, puede seleccionar el elemento de Microsoft Excel si hay alguna opción de importación disponible. Por ejemplo, si selecciona Columna de Microsoft Excel en Excel, están disponibles las opciones Elemento, Valor, Comentario, Autor de comentario, Fórmula, Tipo de hipervínculo y Dirección de hipervínculo.
 - b. Pulse **Aceptar**.
8. Especifique los datos necesarios en los separadores **Propiedades** y **Avanzado**.
9. Pulse **Aceptar** para guardar los valores que ha especificado.

Modificación de la definición de columna en el enlace:

Puede modificar la definición de columna, tal como Tipo de SQL, Longitud, Escala, Admite nulos, en el enlace. Si desea cambiar el nombre de columna que se importó en la ventana **Configuración**, abra de nuevo la ventana **Configuración** para especificar el nombre.

Procedimiento

1. En el lienzo paralelo, efectúe una doble pulsación en la etapa **Unstructured Data**.
2. Seleccione el separador **Salida** y luego seleccione el enlace de salida en **Nombre de salida (etapa en sentido descendente)**.
3. Edite el Tipo de SQL, Longitud y Escala de cada columna.
4. Pulse **Aceptar** para guardar los cambios.

Utilización de parámetros de trabajo

La etapa Unstructured Data no puede crear parámetros de trabajo nuevos en la ventana Configuración. Pero el usuario puede utilizar los parámetros de trabajo en la ventana Configuración. Debe crear parámetros de trabajo en la ventana Propiedades de trabajo antes o después de trabajar en la ventana Configuración. Para ello seleccione **Editar > Propiedades de trabajo** en el cliente de IBM InfoSphere DataStage and QualityStage Designer. Para obtener más información sobre la creación de parámetros de trabajo, consulte la Lección 2.4: Adición de parámetros, en la guía de aprendizaje de trabajos paralelos de IBM InfoSphere DataStage.

Los parámetros de trabajo se especifican en la ventana Configuración entre caracteres #. Por ejemplo, el parámetro de trabajo *FileName*, se especifica como **#FileName#** en la ventana Configuración. Para un campo de tipo serie tal como la propiedad **Nombre de archivo**, puede escribir directamente el nombre del parámetro de trabajo entre los caracteres #.

Si desea utilizar un parámetro de trabajo para una propiedad de tipo List tal como **Opción de rango**, debe crear un parámetro de tipo List que contenga una lista de variables de tipo serie. Las variables de tipo serie deben coincidir con el texto de la etiqueta de la propiedad correspondiente contenida en la ventana Configuración. Por ejemplo, si desea utilizar el parámetro de trabajo para la propiedad **Opción de rango**, debe crear un parámetro de trabajo de tipo List que contenga la variable de tipo serie **Especificar la fila inicial** y **Especificar todo el rango de datos**. Después de crear un parámetro de trabajo, seleccione <Parametrizar...> en la ventana Configuración y especifique el nombre del parámetro de trabajo entre los caracteres # en el cuadro de diálogo **Parámetro de entrada**. Pulse **Cargar** para editar o seleccionar variables en el panel Resolver parámetros de trabajo.

Opciones para leer datos de archivos de Microsoft Excel

Utilice las opciones siguientes para modificar la forma en que la etapa Unstructured Data lee los datos.

Manejo de errores:

Puede especificar si se debe registrar un mensaje de error y detener el trabajo cuando se produce un error al extraer datos del archivo.

Puede establecer la propiedad **Acción para error** en **Concluir** u **Omitir**. El valor predeterminado es **Concluir**.

- Si selecciona **Concluir**, la etapa Unstructured Data registra un error no recuperable y detiene el trabajo cuando se produce un error al extraer datos.
- Si selecciona **Omitir**, la etapa Unstructured Data registra un mensaje de aviso y continúa el proceso de los campos de entrada y registros restantes cuando se produce un error al extraer datos.

Manejo de filas nulas:

Puede configurar la etapa Unstructured Data para pasar por alto las filas con valores nulos en sus celdas durante la extracción.

Puede establecer la propiedad **Omitir filas nulas** en **Sí** o **No**. El valor predeterminado es **No**.

- Para pasar por alto las filas con valores nulos en sus celdas, seleccione **Sí**.
- Para continuar con las filas con valores nulos en sus celdas, seleccione **No**.

Extracción del valor de una celda determinada o propiedades personalizadas:

Puede especificar el valor de una celda determinada o las propiedades personalizadas que se deben extraer.

Procedimiento

1. En el lienzo de diseño de trabajos, efectúe una doble pulsación en el icono **Etapa Unstructured Data**.
2. Pulse **Configurar**.
3. En el panel **Importar**, seleccione el separador **Avanzado**.
4. Para importar el valor de una celda determinada, seleccione **Celda determinada** en la columna **Tipo** y especifique la posición de la celda en la columna **Valor**. Por ejemplo, si desea importar el valor de la celda A1, especifique A1 en la columna **Valor**.

5. Para importar una propiedad personalizada, seleccione **Propiedad personalizada** en la columna **Tipo** y especifique el valor de la propiedad en el campo **Valor**. Por ejemplo, si desea importar la propiedad personalizada denominada Prop1, especifique Prop1 en el campo **Valor**.
6. Pulse **Importar**.
7. En el panel Correlacionar, defina una correlación entre columnas de InfoSphere DataStage y elementos de Microsoft Excel.
8. Pulse **Aceptar**.

Propagación de columnas ocultas:

Puede especificar la acción que debe emprender para las columnas ocultas durante la propagación de columnas en tiempo de ejecución.

Puede definir la propiedad **Columnas ocultas** en la categoría **Propagación de columnas en tiempo de ejecución** para extraer o pasar por alto las columnas ocultas durante la propagación de columnas en tiempo de ejecución.

- Si selecciona **Extraer**, las columnas ocultas se extraen.
- Si selecciona **Omitir**, las columnas ocultas se pasan por alto.

El valor predeterminado es **Extraer**.

Propagación de columnas en tiempo de ejecución

En InfoSphere DataStage, puede configurar un trabajo para propagar a través del resto del trabajo columnas adicionales que no están definidas en los metadatos. Este proceso se conoce como propagación de columnas en tiempo de ejecución (RCP).

Cuando la propagación de columnas en tiempo de ejecución está habilitada, la etapa Unstructured Data propaga columnas de Microsoft Excel basándose en el primer rango de datos. Si se utilizan caracteres comodín en el nombre de archivo, se utiliza el primer archivo que coincide con la expresión. El valor de la propiedad Columnas ocultas determina si se propaga una columna oculta. Para cada columna de Microsoft Excel propagada, sólo se extraen valores de celda. Para extraer información tal como el nombre de archivo, nombre de hoja o número de fila, puede definir columnas adicionales en la ventana Configuración.

Reglas de denominación de columnas

Los nombres de columna de InfoSphere DataStage están basados en la letra de columna de Microsoft Excel del primer rango de datos. El nombre de la columna tiene el prefijo "**Column_**" seguido de la letra de columna de Microsoft Excel. Por ejemplo, Column_A, Column_B, Column_C, etc.

Si el trabajo ya tiene una columna con el nombre, el trabajo se interrumpe.

Tipos de datos

Todas las columnas añadidas por la etapa Unstructured Data son de tipo **Unicode Varchar** y tienen una longitud no definida.

Ejemplos de extracción de datos de archivos de Microsoft Excel

Puede crear trabajos de ejemplo para extraer datos de archivos de Microsoft Excel.

Para obtener los archivos de los ejemplos, extraiga el archivo `vía_acceso_instalación_IS\Clients\Samples\Connectors\UnstructuredData_Samples.zip`.

Ejemplo 1: Extracción de datos de un rango en un archivo de Microsoft Excel:

Cree un trabajo que utilice la etapa Unstructured Data para recuperar datos de un rango en una hoja de cálculo de Microsoft Excel.

Acerca de esta tarea

En este ejemplo se utiliza el archivo de Microsoft Excel de ejemplo, `Employee1.xls`, que contiene datos de los empleados que trabajan en una empresa. Este archivo de ejemplo tiene tres hojas de cálculo: Hoja1, Hoja2 y Hoja3. Hoja1 contiene información sobre los empleados de cada departamento de la empresa. Hoja2 y Hoja3 están en blanco. En este ejemplo, extraerá información empresarial únicamente sobre los empleados que trabajan en el departamento B01.

Paso 1: Creación del trabajo:

Cree un trabajo de ejemplo que incluya una etapa Unstructured Data y una etapa Sequential File.

Procedimiento

1. Inicie el cliente de IBM InfoSphere DataStage and QualityStage Designer.
2. En el panel **Repositorio**, pulse con el botón derecho del ratón en la carpeta **Trabajos** y seleccione **Nuevo > Trabajo paralelo**.
3. En la sección **Archivo** de la paleta, arrastre la etapa Unstructured Data al lienzo.
4. Arrastre una etapa Sequential File al lienzo y coloque la etapa a la derecha de la etapa Unstructured Data.
5. Cree un enlace desde la etapa Unstructured Data a la etapa Sequential File.
6. Cambie el nombre de la etapa y del enlace.
7. Seleccione **Archivo > Guardar** y asigne un nombre al trabajo.

Paso 2: Configuración de la etapa Unstructured Data:

Configure la etapa Unstructured Data para extraer datos de un rango en un archivo de Microsoft Excel.

Procedimiento

1. Efectúe una doble pulsación en la etapa Unstructured Data.
2. Pulse **Configurar**.

Nota: No configure ninguna propiedad de la etapa en este paso, pues puede crear todas las configuraciones necesarias en la ventana Configuración.

3. En la ventana Configuración, especifique la vía de acceso completa del archivo de Microsoft Excel de entrada `Employee1.xls`.
4. En la lista **Opción de rango**, seleccione **Especificar todo el rango de datos** para extraer los datos de un rango específico.
5. En el campo **Expresión de rango**, especifique `Sheet1!A16:K28`.

6. En el campo **Cabecera de columna**, seleccione **Primera fila de rangos de datos**. Cuando se selecciona **Primera fila de rangos de datos**, se considera que la primera fila es la cabecera y la etapa Unstructured Data inicia la extracción a partir de la segunda fila.
7. Pulse **Cargar** y compruebe que estén seleccionadas las casillas situadas junto a las columnas de Microsoft Excel que se deben extraer mediante el trabajo. La etapa Unstructured Data accede al archivo especificado y el panel **Importar** muestra las columnas de Microsoft Excel pertenecientes al rango de datos especificado. De forma predeterminada, se seleccionan todas las columnas de Microsoft Excel.
8. Quite la marca de la casilla situada junto a las columnas de Microsoft Excel denominadas E (PHONE NO) e I (BIRTH DATE).
9. Pulse **Importar**. Cuando pulsa **Importar**, se actualiza el panel **Correlacionar** situado en la parte inferior derecha de la ventana **Configuración**.
10. Pulse **Aceptar**.
11. Compruebe que los valores que ha especificado en la ventana **Configuración** se han guardado en el separador **Propiedad** del editor de etapas.
12. En la página **Salida > Columna**, cambie el tipo de la columna **EMP_NO** a **Integer** y pulse **Aceptar**.

Paso 3: Configuración de la etapa Sequential File:

Configure la etapa Sequential File.

Acerca de esta tarea

En este ejemplo se utiliza la etapa Sequential File como etapa de salida. Puede utilizar cualquier otra etapa de salida para crear la salida.

Procedimiento

1. Efectúe una doble pulsación en la etapa Sequential File.
2. En la página **Propiedades**, especifique la vía de acceso donde desee crear el archivo de salida, seguido por el nombre de archivo `OutputOfExample1.txt`.
3. Pulse **Aceptar**.

Paso 4: Visualización del resultado del trabajo:

Después de ejecutar el trabajo, abra el archivo y verifique el resultado.

Acerca de esta tarea

Por ejemplo, si un archivo de entrada de Microsoft Excel contiene información sobre los empleados de los diferentes departamentos de una empresa, el trabajo puede extraer datos del departamento especificado.

Procedimiento

1. Guarde el trabajo.
2. Compile y ejecute el trabajo.

La tabla siguiente muestra la información de un archivo de entrada de Microsoft Excel que contiene la información sobre empleados de distintos departamentos.

Tabla 8. Ejemplo de archivo de Microsoft Excel con datos de empleados

EMP NO	FIRST NAME	MID INIT	LAST NAME	PHONE NO	HIRE DATE	JOB	SEX	BIRTH DATE
Empleados de DEPT_A00								
10	CHRISTINE	I	HAAS	3978	1/1/1995	PRES	F	8/24/1963
20	MICHAEL	L	THOMSON	3476	10/10/2003	MANAGER	M	2/2/1976
30	SALLY	A	KWAN	4738	4/5/2005	MANAGER	F	5/11/1971
50	JOHN	B	GEYER	6789	8/17/1979	MANAGER	M	9/15/1955
Empleados de DEPT_B01								
60	IRVING	F	STERN	6423	9/14/2003	MANAGER	M	7/7/1975
70	EVA	D	PULASKI	7831	9/30/2003	MANAGER	F	5/26/2003
90	EILEEN	W	HENDERSON	5498	8/15/2000	MANAGER	F	5/15/1971
100	THEODORE	Q	SPENSER	742	6/19/2000	MANAGER	M	12/18/1980
110	VINCENZO	G	LUCCHESI	3490	5/16/1988	SALESREP	M	11/5/1958
120	SEAN		O'CONNELL	2167	12/5/1993	CLERK	M	10/18/1972
130	DELORES	M	QUINTANA	4578	7/28/2001	ANALYST	F	9/15/1955
140	HEATHER	A	NICHOLLS	1793	12/15/2006	ANALYST	F	1/19/1976
150	BRUCE		ADAMSON	4510	2/12/2002	DESIGNER	M	5/17/1972
160	ELIZABETH	R	PIANKA	3782	10/11/2006	DESIGNER	F	4/12/1980
1770	MASATOSHI	J	YOSHIMURA	2890	9/15/1999	DESIGNER	M	1/5/1981
180	MARILYN	S	SCOUTTEN	1682	7/7/2003	DESIGNER	F	2/21/1978

- Después de ejecutarse el trabajo, abra el archivo OutputOfExample1.txt para ver el resultado.

El archivo OutputOfExample1.txt muestra datos correspondientes a DEPT_B01, excepto las columnas NO PHONE y BIRTH DATE:

```
"60","IRVING","F","STERN","2003-09-14","MANAGER","M","72250","500"
"70","EVA","D","PULASKI","2005-09-30","MANAGER","F","96170","700"
"90","EILEEN","W","HENDERSON","2000-08-15","MANAGER","F","89750","600"
"100","THEODORE","Q","SPENSER","2000-06-19","MANAGER","M","86150","500"
"110","VINCENZO","G","LUCCHESI","1988-05-16","SALESREP","M","66500","900"
"120","SEAN","","O'CONNELL","1993-12-05","CLERK","M","49250","600"
"130","DELORES","M","QUINTANA","2001-07-28","ANALYST","F","73800","500"
"140","HEATHER","A","NICHOLLS","2006-12-15","ANALYST","F","68420","600"
"150","BRUCE","","ADAMSON","2002-02-12","DESIGNER","M","55280","500"
"160","ELIZABETH","R","PIANKA","2006-10-11","DESIGNER","F","62250","400"
"170","MASATOSHI","J","YOSHIMURA","1999-09-15","DESIGNER","M","44680","500"
"180","MARILYN","S","SCOUTTEN","2003-07-07","DESIGNER","F","51340","500"
```

Ejemplo 2: Extracción de datos de varias hojas de cálculo de Microsoft Excel:

Cree un trabajo que utilice la etapa Unstructured Data para extraer datos de varias hojas de cálculo de Microsoft Excel.

Acercas de esta tarea

En este ejemplo se utiliza el archivo de Microsoft Excel de ejemplo, Employee2.xls. Este archivo de ejemplo tiene las hojas de cálculo siguientes: DEPT A00, DEPT B01, DEPT C01 y DEPT D01. Cada hoja contiene información sobre los empleados del departamento.

La estructura de datos de cada hoja es similar. Cada hoja tiene las columnas EMP NO, FIRST NAME, MID INIT, LAST NAME, PHONE NO, HIRE DATE, JOB y ADDRESS, y la tercera fila es la cabecera. Pero cada hoja tiene un número diferente de filas.

Paso 1: Creación del trabajo:

Cree un trabajo de ejemplo que incluya una etapa Unstructured Data y una etapa Sequential File.

Procedimiento

1. Inicie el cliente de IBM InfoSphere DataStage and QualityStage Designer.
2. En el panel Repositorio, pulse con el botón derecho del ratón en la carpeta **Trabajos** y seleccione **Nuevo > Trabajo paralelo**.
3. En la sección **Archivo** de la paleta, arrastre una etapa Unstructured Data al lienzo.
4. En la sección **Archivo** de la paleta, arrastre una etapa Sequential File al lienzo. Coloque la etapa a la derecha de la etapa Unstructured Data.
5. Cree un enlace desde la etapa Unstructured Data a la etapa Sequential File.
6. Renombre las etapas y los enlaces.
7. Seleccione **Archivo > Guardar** y asigne un nombre al trabajo.

Paso 2: Configuración de la etapa Unstructured Data:

Configure la etapa Unstructured Data para extraer datos de varias hojas de cálculo Microsoft Excel.

Procedimiento

1. Efectúe una doble pulsación en la etapa Unstructured Data para abrir las propiedades de la etapa.
2. Pulse **Configurar**.

Nota: No configure ninguna propiedad de la etapa en este paso, pues puede crear todas las configuraciones necesarias en la ventana Configuración.

3. En la ventana Configuración, especifique la vía de acceso completa del archivo de Microsoft Excel de entrada Employee2.xls.
4. En **Opción de rango**, seleccione **Especificar la fila inicial**.
5. En el campo **Expresión de rango**, especifique **A3:H3**. Cuando la etapa se ejecuta con la opción Especificar la primera fila y no se especifica ninguna hoja determinada en **Expresión de rango**, el trabajo encuentra dinámicamente la última fila y extrae filas hasta la última fila durante la ejecución.
6. En el campo **Cabecera de columna**, seleccione **Primera fila de rangos de datos**.
7. Pulse **Cargar**. Las columnas de Excel pertenecientes al rango de datos especificado aparecerán en el panel Importar.
8. En el separador **Propiedad**, seleccione el recuadro de selección situado junto a la propiedad para extraer el valor de la propiedad. En este ejemplo, seleccione **Sheetname** como propiedad.
9. Pulse **Importar**. La etapa creará las correlaciones de columnas.
10. Para hacer que la columna SheetName sea la primera columna de la lista:
 - a. Seleccione la columna SheetName.
 - b. Pulse **Subir** hasta que la columna SheetName sea la primera columna de la lista.
11. En la tabla de correlación, inserte una fila para la columna ADDRESS en el archivo de entrada que tiene hiperenlace.
 - a. Pulse **Insertar**.
 - b. En la opción **Elemento de Excel**, seleccione **Columna ADDRESS**.
 - c. En la celda de la opción **Importar** de la nueva fila, seleccione **Dirección de hiperenlace**.

- d. Especifique el nombre de la columna EMAIL_ADDRESS de DataStage para la fila nueva.
12. Pulse **Aceptar**.
13. Compruebe que los valores que ha especificado en la ventana Configuración se han guardado en el separador **Propiedad** del editor de etapas.
14. Pulse el separador **Salida > Columna** para cambiar el tipo de datos u otros atributos. Cambie el tipo de la columna **EMP_NO** por Integer.
15. Pulse **Aceptar**.

Paso 3: Configuración de la etapa Sequential File:

Configure la etapa Sequential File.

Acerca de esta tarea

En este ejemplo se utiliza la etapa Sequential File como etapa de salida. Puede utilizar cualquier otra etapa de salida para crear la salida.

Procedimiento

1. Efectúe una doble pulsación en la etapa Sequential File.
2. En la página **Propiedades**, especifique la vía de acceso donde desee crear el archivo de salida, seguido por el nombre de archivo OutputOfExample2.txt.
3. Pulse **Aceptar**.

Paso 4: Visualización del resultado del trabajo:

Después de ejecutar el trabajo, abra el archivo y verifique el resultado.

Procedimiento

1. Guarde el trabajo.
2. Compile y ejecute el trabajo.

Considere por ejemplo archivos de entrada de Microsoft Excel cuyas diferentes hojas contienen información sobre los empleados de cada departamento. El trabajo extrae datos sobre empleados a partir de todas las hojas y los datos se muestran en la forma indicada en las tablas siguientes:

Tabla 9. Información de empleados de DEPT_A00

EMP NO	FIRST NAME	MIDI INIT	LAST NAME	PHONE NO	HIRE DATE	JOB	SEX	BIRTH DATE
10	CHRISTINE	I	HAAS	3978	1/1/1995	PRES	F	8/24/1963
20	MICHAEL	L	THOMSON	3476	10/10/2003	MANAGER	M	2/2/1976
30	SALLY	A	KWAN	4738	4/5/2005	MANAGER	F	5/11/1971
50	JOHN	B	GEYER	6789	8/17/1979	MANAGER	M	9/15/1955

Tabla 10. Datos de los empleados de DEPT_B01

EMP NO	FIRST NAME	MIDI NIT	LAST NAME	PHONE NO	HIRE DATE	JOB	SEX	BIRTH DATE
60	IRVING	F	STERN	6423	9/14/2003	MANAGER	M	7/7/1975
70	EVA	D	PULASKI	7831	9/30/2003	MANAGER	F	5/26/2003
90	EILEEN	W	HENDERSON	5498	8/15/2000	MANAGER	F	5/15/1971
100	THEODORE	Q	SPENSER	742	6/19/2000	MANAGER	M	12/18/1980
110	VINCENZO	G	LUCCHESSI	3490	5/16/1988	SALESREP	M	11/5/1958
120	SEAN		O'CONNELL	2167	12/5/1993	CLERK	M	10/18/1972
130	DELORES	M	QUINTANA	4578	7/28/2001	ANALYST	F	9/15/1955
140	HEATHER	A	NICHOLLS	1793	12/15/2006	ANALYST	F	1/19/1976
150	BRUCE		ADAMSON	4510	2/12/2002	DESIGNER	M	5/17/1972

Tabla 10. Datos de los empleados de DEPT_B01 (continuación)

EMP NO	FIRST NAME	MIDI NIT	LAST NAME	PHONE NO	HIRE DATE	JOB	SEX	BIRTH DATE
160	ELIZABETH	R	PIANKA	3782	10/11/2006	DESIGNER	F	4/12/1980
1770	MASATOSHI	J	YOSHIMURA	2890	9/15/1999	DESIGNER	M	1/5/1981
180	MARILYN	S	SCOUTTEN	1682	7/7/2003	DESIGNER	F	2/21/1978

3. Después de ejecutarse el trabajo, abra el archivo OutputOfExample2.txt que contiene el resultado siguiente.

```
"DEPT A00", "10", "CHRISTINE", "I", "HAAS", "3978", "1995-01-01", "PRES    ",
"CHRISTINE HAAS", "mailto:CHRISTINE%20HAAS@abc.com"
"DEPT A00", "20", "MICHAEL", "L", "THOMPSON", "3476", "2003-10-10", "MANAGER ",
"MICHAEL THOMPSON", "mailto:MICHAEL%20THOMPSON@abc.com"
"DEPT A00", "30", "SALLY", "A", "KWAN", "4738", "2005-04-05", "MANAGER ",
"SALLY KWAN", "mailto:SALLY%20KWAN@abc.com"
"DEPT A00", "50", "JOHN", "B", "GEYER", "6789", "1979-08-17", "MANAGER ",
"JOHN GEYER", "mailto:JOHN%20GEYER@abc.com"
"DEPT B01", "60", "IRVING", "F", "STERN", "6423", "2003-09-14", "MANAGER ",
"IRVING STERN", "mailto:IRVING%20STERN@abc.com"
"DEPT B01", "70", "EVA", "D", "PULASKI", "7831", "2005-09-30", "MANAGER ",
"EVA PULASKI", "mailto:EVA%20PULASKI@abc.com"
"DEPT B01", "90", "EILEEN", "W", "ENDERSON", "5498", "2000-08-15", "MANAGER ",
"EILEEN HENDERSON", "mailto:EILEEN%20ENDERSON@abc.com"
"DEPT B01", "100", "THEODORE", "Q", "SPENSER", "972", "2000-06-19", "MANAGER ",
"THEODORE SPENSER", "mailto:THEODORE%20SPENSER@abc.com"
"DEPT B01", "110", "VINCENZO", "G", "LUCCHESSI", "3490", "1988-05-16", "SALESREP",
"VINCENZO LUCCHESSI", "mailto:VINCENZO%20LUCCHESSI@abc.com"
"DEPT B01", "120", "SEAN", " ", "O'CONNELL", "2167", "1993-12-05", "CLERK   ",
"SEAN O'CONNELL", "mailto:SEAN%20O'CONNELL@abc.com"
"DEPT B01", "130", "DELORES", "M", "QUINTANA", "4578", "2001-07-28", "ANALYST ",
"DELORES QUINTANA", "mailto:DELORES%20QUINTANA@abc.com"
"DEPT B01", "140", "HEATHER", "A", "NICHOLLS", "1793", "2006-12-15", "ANALYST ",
"HEATHER NICHOLLS", "mailto:HEATHER%20NICHOLLS@abc.com"
"DEPT B01", "150", "BRUCE", " ", "ADAMSON", "4510", "2002-02-12", "DESIGNER",
"BRUCE ADAMSON", "mailto:BRUCE%20ADAMSON@abc.com"
"DEPT B01", "160", "ELIZABETH", "R", "PIANKA", "3782", "2006-10-11", "DESIGNER",
"ELIZABETH PIANKA", "mailto:ELIZABETH%20PIANKA@abc.com"
"DEPT B01", "170", "MASATOSHI", "J", "YOSHIMURA", "2890", "1999-09-15", "DESIGNER",
"MASATOSHI YOSHIMURA", "mailto:MASATOSHI%20YOSHIMURA@abc.com"
"DEPT B01", "180", "MARILYN", "S", "SCOUTTEN", "1682", "2003-07-07", "DESIGNER",
"MARILYN SCOUTTEN", "mailto:MARILYN%20SCOUTTEN@abc.com"
"DEPT C01", "190", "JAMES", "H", "WALKER", "2986", "2004-07-26", "DESIGNER",
"JAMES WALKER", "mailto:JAMES%20WALKER@abc.com"
"DEPT C01", "200", "DAVID", " ", "BROWN", "4501", "2002-03-03", "DESIGNER",
"DAVID BROWN", "mailto:DAVID%20BROWN@abc.com"
"DEPT C01", "210", "WILLIAM", "T", "JONES", "942", "1998-04-11", "DESIGNER",
"WILLIAM JONES", "mailto:WILLIAM%20JONES@abc.com"
"DEPT C01", "220", "JENNIFER", "K", "LUTZ", "672", "1998-08-29", "DESIGNER",
"JENNIFER LUTZ", "mailto:JENNIFER%20LUTZ@abc.com"
"DEPT C01", "230", "JAMES", "J", "JEFFERSON", "2094", "1996-11-21", "CLERK   ",
"JAMES JEFFERSON", "mailto:JAMES%20JEFFERSON@abc.com"
"DEPT C01", "240", "SALVATORE", "M", "MARINO", "3780", "2004-12-05", "CLERK   ",
"SALVATORE MARINO", "mailto:SALVATORE%20MARINO@abc.com"
"DEPT C01", "250", "DANIEL", "S", "SMITH", "961", "1999-10-30", "CLERK   ",
"DANIEL SMITH", "mailto:DANIEL%20SMITH@abc.com"
"DEPT C01", "260", "SYBIL", "P", "JOHNSON", "8953", "2005-09-11", "CLERK   ",
"SYBIL JOHNSON", "mailto:SYBIL%20JOHNSON@abc.com"
"DEPT D01", "270", "MARIA", "L", "PEREZ", "9001", "2006-09-30", "CLERK   ",
"MARIA PEREZ", "mailto:MARIA%20PEREZ@abc.com"
"DEPT D01", "280", "ETHEL", "R", "SCHNEIDER", "8997", "1997-03-24", "OPERATOR",
"ETHEL SCHNEIDER", "mailto:ETHEL%20SCHNEIDER@abc.com"
"DEPT D01", "290", "JOHN", "R", "PARKER", "4502", "2006-05-30", "OPERATOR",
"JOHN PARKER", "mailto:JOHN%20PARKER@abc.com"
"DEPT D01", "300", "PHILIP", "X", "SMITH", "2095", "2002-06-19", "OPERATOR",
"PHILIP SMITH", "mailto:PHILIP%20SMITH@abc.com"
"DEPT D01", "310", "MAUDE", "F", "SETRIGHT", "3332", "1994-09-12", "OPERATOR",
```

```
"MAUDE SETRIGHT", "mailto:MAUDE%20SETRIGHT@abc.com"  
"DEPT D01", "320", "RAMLAL", "V", "MEHTA", "9990", "1995-07-07", "FIELDREP",  
"RAMLAL MEHTA", "mailto:RAMLAL%20MEHTA@abc.com"  
"DEPT D01", "330", "WING", " ", "LEE", "2103", "2006-02-23", "FIELDREP",  
"WING LEE", "mailto:WING%20LEE@abc.com"  
"DEPT D01", "340", "JASON", "R", "GOUNOT", "5698", "1977-05-05", "FIELDREP",  
"JASON GOUNOT", "mailto:JASON%20GOUNOT@abc.com"
```

Ejemplo 3: Extracción de datos de varios rangos que tienen estructuras de datos diferentes en un archivo de Microsoft Excel:

Cree un trabajo que utilice la etapa Unstructured Data para extraer datos de varios rangos que tienen estructuras de datos diferentes en un archivo de Microsoft Excel.

Acerca de esta tarea

En este ejemplo se utiliza el archivo de Microsoft Excel de ejemplo, Employee3.xls. Este archivo de ejemplo tiene dos hojas, Departments y Employees, que tienen estructuras de datos diferentes.

En este ejemplo, la etapa Unstructured Data tiene dos enlaces de salida. Puede extraer datos de la hoja Departments al primer enlace, y de la hoja Employees al segundo enlace.

Paso 1: Creación del trabajo:

Cree un trabajo de ejemplo que incluya una etapa Unstructured Data y dos etapas Sequential File.

Procedimiento

1. Inicie el cliente de IBM InfoSphere DataStage and QualityStage Designer.
2. En el panel Repositorio, pulse con el botón derecho del ratón en la carpeta **Trabajos** y seleccione **Nuevo > Trabajo paralelo**.
3. En la sección **Archivo** de la paleta, arrastre una etapa Unstructured Data al lienzo.
4. En la sección **Archivo** de la paleta, arrastre dos etapas Sequential File al lienzo. Coloque las etapas a la derecha de la etapa Unstructured Data.
5. Cambie el nombre de las nuevas etapas Sequential File por Output_1 y Output_2.
6. Cree un enlace desde la etapa Unstructured Data a las etapas Sequential File.
7. Cambie el nombre de los enlaces por Departments y Employees.
8. Guarde el trabajo.

Paso 2: Configuración de la etapa Unstructured Data:

Configure la etapa Unstructured Data para que extraiga datos de varias hojas de cálculo de Microsoft Excel.

Procedimiento

1. Efectúe una doble pulsación en la etapa Unstructured Data para abrir las propiedades de la etapa.
2. Pulse **Configurar**.
3. En la ventana Configuración, especifique la vía de acceso completa del archivo de Microsoft Excel de entrada Employee3.xls.

4. Especifique los datos que deban extraerse de la hoja de cálculo **Departments** y complete los subpasos siguientes para generar las correlaciones de columnas.
 - a. En el recuadro de lista **Enlace**, seleccione **Departments**.
 - b. En la lista **Opción de rango**, seleccione **Especificar todo el rango de datos**.
 - c. En el campo **Expresión de rango**, especifique **Departments!A2:C6**.
 - d. En **Cabecera de columna**, seleccione **Primera fila de rangos de datos**.
 - e. Pulse **Cargar**. Las columnas de Excel pertenecientes al rango de datos especificado aparecerán en el panel Importar.
5. Especifique los datos que deban extraerse de la hoja de cálculo **Employees** y complete los subpasos siguientes para generar las correlaciones de columnas.
 - a. En el recuadro de lista **Enlace**, seleccione **Employees**.
 - b. En la lista **Opción de rango**, seleccione **Especificar todo el rango de datos**.
 - c. En el campo **Expresión de rango**, especifique **Employees!A2:L34**.
 - d. En el campo **Cabecera de columna**, seleccione **Primera fila de rangos de datos**.
 - e. Pulse **Cargar**. Las columnas de Excel pertenecientes al rango de datos especificado aparecerán en el panel Importar.
 - f. Pulse **Importar** y, a continuación, pulse **Aceptar**. La etapa correlaciona las columnas.
6. Compruebe que los valores que ha especificado en la ventana Configuración se han guardado en el separador **Propiedad** del editor de etapas.
7. En la página Salida, seleccione el enlace **Employees** como el nombre de salida.
8. En la página Columnas, cambie el tipo de datos de la columna EMP_NO a Integer y, a continuación, pulse **Aceptar**.

Paso 3: Configuración de etapas Sequential File:

Configurar las etapas Sequential File. En este ejemplo se utiliza la etapa Sequential File como etapa de salida. Puede utilizar cualquier otra etapa para crear la salida.

Acerca de esta tarea

En este ejemplo se utiliza la etapa Sequential File como etapa de salida. Puede utilizar cualquier otra etapa de salida para crear la salida.

Procedimiento

1. Efectúe una doble pulsación en la etapa Sequential File Output_1.
2. En la página **Propiedades**, especifique la vía de acceso para crear el archivo de salida, seguido por el nombre de archivo OutputOfExample3_1.txt.
3. Pulse **Aceptar**.
4. Repita los pasos 1-3 para la segunda etapa Sequential File Output_2 y el nombre de archivo como OutputOfExample3_2.txt.

Paso 4: Visualización del resultado del trabajo:

Después de ejecutar el trabajo, abra el archivo y verifique el resultado.

Procedimiento

1. Guarde el trabajo.
2. Compile y ejecute el trabajo.

Un archivo de entrada de Microsoft Excel de ejemplo, Employee3.xls, contiene información de departamento en la hoja **Departments**, e información de los empleados en la hoja **Employees**. El trabajo extrae los datos del departamento al archivo OutputOfExample3_1.txt, y los datos de los empleados al archivo OutputOfExample3_2.txt. Los datos de cada hoja son los siguientes:

3. Después de ejecutarse el trabajo, abra los archivos OutputOfExample3.txt y OutputOfExample3_2.txt. El archivo OutputOfExample3_1.txt debería coincidir con la hoja **Departments** y el archivo OutputOfExample3_2.txt debería coincidir con la hoja **Employees** del archivo Employee.xls.

Grabación de datos en un archivo de Microsoft Excel nuevo

Puede utilizar la etapa Unstructured Data en los trabajos que graben datos en un archivo de Microsoft Excel nuevo, especificando el nombre de vía de acceso completo. También puede crear varios archivos de Microsoft Excel, especificando la ubicación desde la que se hayan creado los archivos, y un prefijo para los nombres de archivo.

Cada enlace de entrada de la etapa Unstructured Data está asociado a una hoja de Microsoft Excel distinta. Las hojas de Microsoft Excel se nombran asociadas a los enlaces de entrada, por ejemplo, **Sheet1**, **Sheet2**, **Sheet3**, etc. Cada columna de InfoSphere DataStage del enlace de entrada se correlaciona con una columna de Microsoft Excel.

Diseño de trabajos que tienen la etapa Unstructured Data

Puede utilizar una etapa Unstructured Data para grabar datos en orígenes de datos no estructurados.

Procedimiento

1. Defina un trabajo que incluya una etapa Unstructured Data.
2. Para configurar la etapa Unstructured Data como una etapa de destino para grabar datos en orígenes de datos no estructurado, complete los pasos siguientes:
 - a. Configure la etapa Unstructured Data como destino.
 - b. Especifique la definición de columna en el enlace.
3. Compile y ejecute el trabajo.

Configuración de la etapa Unstructured Data como destino:

Puede configurar la etapa Unstructured Data para un archivo de Microsoft Excel.

Acerca de esta tarea

La etapa Unstructured Data sólo da soporte al formato OOXML (.xlsx) de archivos de Microsoft Excel como archivo de destino.

La etapa Unstructured Data da soporte a la propagación de columnas en tiempo de ejecución. Cuando se habilita la propagación de columnas en tiempo de ejecución en un enlace de salida de una etapa en sentido ascendente, las columnas adicionales propagadas se añaden después de las columnas que se haya definido en el cliente del Diseñador de InfoSphere.

La etapa Unstructured Data no da soporte a la generación de archivos .xls o archivos cifrados por contraseña.

Procedimiento

1. En el lienzo paralelo, efectúe una doble pulsación en la etapa **Unstructured Data**.
2. En el separador **Etapa**, seleccione el tipo de documento como **Excel**, en la lista **Tipo de documento**.
3. En la lista **Modalidad de grabación**, seleccione **Crear un archivo**.
4. Pulse **Configurar** para configurar propiedades para grabar datos en un archivo de Microsoft Excel.
5. En el panel Archivo de salida, especifique lo siguiente:
 - a. En el campo Nombre de archivo, especifique el nombre de vía de acceso completo del archivo de Microsoft Excel en que deban grabarse los datos. Por ejemplo, puede especificar C:\tmp\employee.xlsx.
 - b. Opcional: Especifique **Modalidad de actualización de archivos**. Si selecciona **Crear (Generar error si existe)**, la ejecución del trabajo falla si ya existe el archivo de Microsoft Excel de destino. Si selecciona **Sobrescribir**, que es el valor predeterminado, la etapa Unstructured Data sobrescribe el archivo existente.
 - c. Opcional: Especifique el **Método de grabación**. Si selecciona **Generar varios archivos**, la etapa Unstructured Data crea varios archivos de Microsoft Excel, en base a los valores de propiedades adicionales. Si selecciona **Archivo específico**, que es el valor predeterminado, la etapa Unstructured Data crea un archivo de Microsoft Excel con el nombre que se especifique en la propiedad **Nombre de archivo**.
6. En el panel Propiedades, especifique lo siguiente:
 - a. Opcional: Desmarque el recuadro de selección **Establecer para todos los enlaces** para especificar las propiedades de cada enlace de entrada que se seleccione en la lista **Enlace**. Si selecciona el recuadro de selección **Establecer para todos los enlaces**, que es el valor predeterminado, los valores de propiedades se aplican para todos los enlaces de entrada.
 - b. Opcional: Especifique la **Columna de cabecera**. Si selecciona **Nombres de columna**, la etapa Unstructured Data graba los nombres de columna de InfoSphere DataStage en la primera fila de la hoja de Microsoft Excel. Si selecciona **Ninguna**, que es el valor predeterminado, la etapa Unstructured Data graba los datos en la primera fila de la hoja de Microsoft Excel.
 - c. Opcional: Especifique **Tamaño automático de columnas**. Si selecciona **Sí**, la etapa Unstructured Data ajusta cada ancho de columna en la hoja de Microsoft Excel generada, para ajustarse al contenido de la columna. Si selecciona **No**, que es el valor predeterminado, la etapa Unstructured Data no ajusta el ancho de columna.
7. Opcional: En el panel Orden de hojas, cambie el orden de hojas, y modifique los nombres de hoja.
8. Pulse **Aceptar** para guardar la configuración que haya especificado.

Tareas relacionadas:

“Grabación de datos en varios archivos de Microsoft Excel” en la página 20
Puede utilizar la etapa Unstructured Data para grabar datos en varios archivos de Microsoft Excel cuando tenga una gran cantidad de datos.

Especificación de la definición de columna en el enlace:

Puede especificar la definición de columna, tal como Tipo de SQL, Longitud, Escala, Admite nulos, en el enlace.

Procedimiento

1. En el lienzo paralelo, efectúe una doble pulsación en la etapa **Unstructured Data**.
2. Seleccione el separador **Entrada** y, a continuación, seleccione el enlace de entrada de **Nombre de entrada (etapa en sentido ascendente)**.
3. Edite el Tipo de SQL, Longitud y Escala de cada columna.
4. Pulse **Aceptar** para guardar los cambios.

Utilización de parámetros de trabajo al modificar un archivo existente de Microsoft Excel:

La etapa Unstructured Data no puede crear parámetros de trabajo nuevos en la ventana Configuración. Pero el usuario puede utilizar los parámetros de trabajo en la ventana Configuración. Debe crear parámetros de trabajo en la ventana Propiedades de trabajo antes o después de trabajar en la ventana Configuración. Para ello seleccione **Editar > Propiedades de trabajo** en el cliente de IBM InfoSphere DataStage and QualityStage Designer. Para obtener más información sobre la creación de parámetros de trabajo, consulte la Lección 2.4: Adición de parámetros, en la guía de aprendizaje de trabajos paralelos de IBM InfoSphere DataStage.

Los parámetros de trabajo se especifican en la ventana Configuración entre caracteres #. Por ejemplo, el parámetro de trabajo *FileName*, se especifica como **#FileName#** en la ventana Configuración. Para un campo de tipo serie tal como la propiedad **Nombre de archivo**, puede escribir directamente el nombre del parámetro de trabajo entre los caracteres #.

Si desea utilizar un parámetro de trabajo para una propiedad de tipo List tal como **Acción si el archivo ya existe**, debe crear un parámetro de tipo List que contenga una lista de variables de tipo serie. Las variables de tipo serie deben coincidir con el texto de la etiqueta de la propiedad correspondiente contenida en la ventana Configuración. Por ejemplo, si desea utilizar el parámetro de trabajo para la propiedad **Acción si el archivo ya existe**, debe crear un parámetro de trabajo de tipo List que contenga la variable de tipo serie **Error** y **Sobrescribir**. Después de crear un parámetro de trabajo, seleccione **<Parametrizar...>** en la ventana Configuración y especifique el nombre del parámetro de trabajo entre los caracteres # en el cuadro de diálogo **Parámetro de entrada**. Pulse **Cargar** para editar o seleccionar variables en el panel Resolver parámetros de trabajo.

Grabación de datos en varios archivos de Microsoft Excel

Puede utilizar la etapa Unstructured Data para grabar datos en varios archivos de Microsoft Excel cuando tenga una gran cantidad de datos.

Acerca de esta tarea

Puede utilizar la etapa Unstructured Data para diseñar trabajos que graben datos en varios archivos de Microsoft Excel. El número máximo de registros al que da soporte el formato OOXML de Microsoft Excel (.xlsx) es de 1.048.576. Cuando los enlaces de entrada tengan más de 1.048.576 registros, deberá dividir éstos últimos en varios archivos de Microsoft Excel. Aunque los enlaces no tengan más de 1.048.576 registros, es posible que desee grabarlos en varios archivos de Microsoft Excel porque abrir un archivo de Microsoft Excel grande requiere una gran cantidad de memoria.

Procedimiento

1. Defina un trabajo que incluya una etapa Unstructured Data.
2. Configure la etapa Unstructured Data.
3. Compile y ejecute el trabajo.

Tareas relacionadas:

“Configuración de la etapa Unstructured Data como destino” en la página 18
Puede configurar la etapa Unstructured Data para un archivo de Microsoft Excel.

Configuración de la etapa Unstructured Data:

Puede configurar la etapa Unstructured Data para generar varios archivos de Microsoft Excel.

Procedimiento

1. En el lienzo paralelo, efectúe una doble pulsación en la etapa **Unstructured Data**.
2. En el separador **Etapa**, seleccione el tipo de documento como **Excel**, en la lista **Tipo de documento**.
3. En la lista **Modalidad de grabación**, seleccione **Crear un archivo**.
4. Pulse **Configurar** para configurar propiedades para grabar datos en un archivo de Microsoft Excel.
5. En el panel Archivo de salida, especifique lo siguiente:
 - a. En el campo **Nombre de archivo**, especifique la ubicación donde desee crear varios archivos de Microsoft Excel. También puede especificar el prefijo de los nombres de archivo, a continuación de la ubicación del archivo. Por ejemplo, si desea crear archivos de Microsoft Excel en C:\tmp y utilizar Ejemplo como el prefijo de los nombres de archivo, especifique C:\tmp\Ejemplo en el campo **Nombre de archivo**. La etapa Unstructured Data añade, al prefijo, un número secuencial de tres dígitos y la extensión de archivo (.xlsx). Si el número de archivos supera los 999, el nombre de archivo contendrá el número necesario de dígitos.
 - b. Opcional: Especifique **Modalidad de actualización de archivos**. Si selecciona **Crear (Generar error si existe)**, la ejecución del trabajo falla si ya existe el archivo de Microsoft Excel de destino. Si selecciona **Sobrescribir**, que es el valor predeterminado, la etapa Unstructured Data sobrescribe el archivo existente.
 - c. Opcional: En la lista **Método de grabación**, seleccione **Generar varios archivos**. La etapa Unstructured Data crea varios archivos de Microsoft Excel en la ubicación que se especifique en el campo **Nombre de archivo**.
6. En el panel Propiedades, especifique lo siguiente:
 - a. Opcional: Seleccione el recuadro de selección **Establecer para todos los enlaces**, que es el valor predeterminado, los valores de propiedades se aplican para todos los enlaces de entrada.
 - b. Opcional: Especifique la **Columna de cabecera**. Si selecciona **Nombres de columna**, la etapa Unstructured Data graba los nombres de columna de InfoSphere DataStage en la primera fila de la hoja de Microsoft Excel. Si selecciona **Ninguna**, que es el valor predeterminado, la etapa Unstructured Data graba los datos en la primera fila de la hoja de Microsoft Excel.
 - c. Opcional: Especifique **Tamaño automático de columnas**. Si selecciona **Sí**, la etapa Unstructured Data ajusta cada ancho de columna en la hoja de

Microsoft Excel generada, para ajustarse al contenido de la columna. Si selecciona **No**, que es el valor predeterminado, la etapa Unstructured Data no ajusta el ancho de columna.

- d. Especifique el **Número máximo de filas en una hoja**. El valor predeterminado es 65536. Este número incluye una fila de nombre de columna cuando se selecciona **Nombres de columna** en la lista **Columna de cabecera**.
- 7. En el panel Orden de hojas, cambie el orden de hojas, y modifique los nombres de hoja.
- 8. Pulse **Aceptar** para guardar la configuración que haya especificado.

Consideración sobre el fin de oleada:

En los trabajos paralelos de InfoSphere DataStage, algunas etapas pueden enviar un marcador de fin de oleada (EOW), que indica el final de una unidad de trabajo o transacción. Cuando todos los registros que se extraen del enlace de entrada se incluyen en una unidad de trabajo (denominada oleada única), la etapa Unstructured Data genera hojas de Microsoft Excel que contienen el número máximo de registros hasta que se graben todos los registros.

Por ejemplo, supongamos que la etapa Unstructured Data tiene dos enlaces de entrada, DSLink1 y DSLink2. DSLink1 está asociado a Sheet1, y DSLink2 está asociado a Sheet2. El número máximo de registros de una hoja es 65.536, y DSLink1 tiene 100.000 registros; DSLink2 tiene 150.000 registros. Ninguna hoja tiene nombres de columna en la primera fila. En este caso, en cada hoja de cada archivo se incluye el número de registros siguiente.

Nombre de archivo	Sheet1	Sheet2
Workbook001.xlsx	65.536	65.536
Workbook002.xlsx	34.464	65.536
Workbook003.xlsx	0	18.928

Cuando los registros que se extraen del enlace de entrada se dividen en dos o más unidades de trabajo (denominadas varias oleadas), la etapa Unstructured Data deja de grabar registros en la hoja de Microsoft Excel, y crea un nuevo archivo de Microsoft Excel, si el número de registros de una oleada excede el número máximo en, al menos, una hoja. La etapa Unstructured Data no graba ningún registro de la oleada siguiente en el archivo anterior, aunque una hoja pueda contener más registros. Por ejemplo, supongamos que cada enlace contiene el número de registros siguiente en cada oleada.

Nº de oleada	DSLink1	DSLink2
1	90.000	50.000
2	5.000	90.000
3	5.000	10.000

En la primera oleada, la etapa Unstructured Data crea un archivo de Microsoft Excel denominado Workbook001.xlsx que tiene dos hojas, Sheet1 y Sheet2. La etapa Unstructured Data graba los registros que se extraen de DSLink1 a Sheet1, hasta que lee el número máximo de registros (65.536) y graba todos los registros (50.000) de DSLink2 a Sheet2. A continuación, la etapa Unstructured Data crea un archivo de Microsoft Excel denominado Workbook002.xlsx, y graba el resto de los

registros (24.464) de la primera oleada y todos los registros (5.000) de la segunda oleada, de DSLink1 a Sheet1. A pesar de que la hoja Sheet2 del archivo Workbook001.xlsx no supere el número máximo de registros, la etapa Unstructured Data graba los registros (65.536) de la segunda oleada extraída de DSLink2 en la hoja Sheet2 del archivo Workbook002.xlsx, no del archivo Workbook001.xlsx. Cuando el número de registros de la segunda oleada de DSLink2 excede el número máximo, la etapa Unstructured Data crea un archivo de Microsoft Excel denominado Workbook003.xlsx, y graba el resto de los registros (24.464) de la segunda oleada de DSLink2 en la hoja Sheet2 del archivo Workbook003.xlsx. En el caso de tercera oleada, puesto que ambas hojas tienen suficiente espacio, la etapa Unstructured Data graba todos los registros extraídos (5.000) y DSLink2 (10.000) a la hoja Sheet1 y la hoja Sheet2 del archivo Workbook003.xlsx.

Como resultado, en cada hoja de cada archivo se graba el número de registros siguiente:

Nombre de archivo	Sheet1	Sheet2
Workbook001.xlsx	65.536 de la 1ª oleada	50.000 procedentes de la primera oleada
Workbook002.xlsx	29.464 (= 24.464 de la 1ª oleada + 5.000 de la 2ª oleada)	65.536 de la 2ª oleada
Workbook003.xlsx	5.000 de la 3ª oleada	34.464 (= 24.464 de la 2ª oleada + 10.000 de la 3ª oleada)

Ejemplos de grabación de datos a archivos de Microsoft Excel

Puede crear trabajos de ejemplo que graben datos en archivos Microsoft Excel.

Para obtener los archivos de los ejemplos, extraiga el archivo vía acceso instalación IS\Clients\Samples\Connectors\UnstructuredData_Samples.zip.

Ejemplo 1: Grabación de datos en una hoja de cálculo de Microsoft Excel:

Cree un trabajo que utilice la etapa Unstructured Data para grabar datos en una hoja de cálculo de Microsoft Excel.

Acerca de esta tarea

En este ejemplo se utiliza un archivo de texto, Employee.txt, como datos de origen. El archivo de origen contiene información de los empleados en formato CSV. Puede grabar esta información en una hoja de cálculo de Microsoft Excel.

Paso 1: Creación del trabajo:

Cree un trabajo de ejemplo que incluya una etapa Unstructured Data y una etapa Sequential File.

Procedimiento

1. Inicie el cliente de IBM InfoSphere DataStage and QualityStage Designer.
2. En el panel Repositorio, pulse con el botón derecho del ratón en la carpeta **Trabajos** y seleccione **Nuevo > Trabajo paralelo**.
3. En la sección **Archivo** de la paleta, arrastre una etapa Sequential File al lienzo.

4. En la sección **Archivo** de la paleta, arrastre una etapa Unstructured Data al lienzo. Coloque la etapa a la derecha de la etapa Sequential File.
5. Cree un enlace desde la etapa Sequential File a la etapa Unstructured Data.
6. Cambie el nombre de la etapa y del enlace.
7. Guarde el trabajo.

Paso 2: Configuración de la etapa Unstructured Data:

Configure la etapa Unstructured Data para extraer datos de un rango en un archivo de Microsoft Excel.

Procedimiento

1. Efectúe una doble pulsación en la etapa Unstructured Data para abrir las propiedades de la etapa.
2. En el separador **Etapa**, seleccione **Excel** en la lista **Tipo de documento**.
3. En la lista **Modalidad de grabación**, seleccione **Crear un archivo**.
4. Pulse **Configurar**.
5. En la ventana Configuración, especifique la vía de acceso completa donde desee crear el archivo de salida, seguido del nombre de archivo OutputOfSample4.txt.
6. En **Modalidad de actualización de archivos**, seleccione **Sobrescribir**.
7. En **Método de grabación**, seleccione **Archivo específico**.
8. En el campo **Cabecera de columna**, seleccione **Nombres de columna**.
9. En el campo **Tamaño automático de columnas** en el separador **Propiedad**, seleccione **Sí**.
10. En el panel **Orden de hojas**, especifique **Employee** como **Nombre de hoja**.
11. Pulse **Aceptar**.
12. Compruebe que los valores que ha especificado en la ventana Configuración se han guardado en el separador **Propiedad** del editor de etapas.
13. Pulse **Aceptar**.

Paso 3: Configuración de la etapa Sequential File:

Configure la etapa Sequential File.

Acerca de esta tarea

En este ejemplo se utiliza la etapa Sequential File como etapa de salida. Puede utilizar cualquier otra etapa de salida para crear la salida.

Procedimiento

1. Efectúe una doble pulsación en la etapa Sequential File.
2. En la página **Propiedades**, especifique la vía de acceso del archivo de entrada Employee.txt.
3. En la página **Columnas**, defina las columnas, tal como se muestra en la figura siguiente.

	Column name	Key	SQL type	Extended	Length	Scale	Nullable	Data element
1	EMPNO	<input checked="" type="checkbox"/>	Char		6		No	
2	FIRSTNME	<input type="checkbox"/>	VarChar		12		No	
3	MIDINIT	<input type="checkbox"/>	Char		1		Yes	
4	LASTNAME	<input type="checkbox"/>	VarChar		15		No	
5	WORKDEPT	<input type="checkbox"/>	Char		3		Yes	
6	PHONENO	<input type="checkbox"/>	Char		4		Yes	
7	HIREDATE	<input type="checkbox"/>	Date		10		Yes	
8	JOB	<input type="checkbox"/>	Char		8		Yes	
9	EDLEVEL	<input type="checkbox"/>	SmallInt		2		No	
10	SEX	<input type="checkbox"/>	Char		1		Yes	
11	BIRTHDATE	<input type="checkbox"/>	Date		10		Yes	
12	SALARY	<input type="checkbox"/>	Decimal		9	2	Yes	
13	BONUS	<input type="checkbox"/>	Decimal		9	2	Yes	
14	COMM	<input type="checkbox"/>	Decimal		9	2	Yes	

4. Pulse **Aceptar**.

Paso 4: Visualización del resultado del trabajo:

Después de ejecutar el trabajo, abra el archivo y verifique el resultado.

Procedimiento

1. Guarde el trabajo.
2. Compile y ejecute el trabajo.
3. Vea los datos de origen del archivo **Employee.txt**.

En este ejemplo se muestra el principio de los datos del archivo:

```
"000010","CHRISTINE","I","HAAS","A00","3978","1995-01-01","PRES","18","F","1963-08-24","0152750.00",
"0001000.00","0004220.00"
"000020","MICHAEL","L","THOMPSON","B01","3476","2003-10-10","MANAGER","18","M","1978-02-02",
"0094250.00","0000800.00","0003300.00"
"000030","SALLY","A","KWAN","C01","4738","2005-04-05","MANAGER","20","F","1971-05-11","0098250.00",
"0000800.00","0003060.00"
"000050","JOHN","B","GEYER","E01","6789","1979-08-17","MANAGER","16","M","1955-09-15","0080175.00",
"0000800.00","0003214.00"
"000060","IRVING","F","STERN","D11","6423","2003-09-14","MANAGER","16","M","1975-07-07","0072250.00",
"0000500.00","0002580.00"
"000070","EVA","D","PULASKI","D21","7831","2005-09-30","MANAGER","16","F","2003-05-26","0096170.00",
"0000700.00","0002893.00"
```

4. Después de ejecutarse el trabajo, abra el archivo **OutputOfExample4.xlsx**, para ver el resultado. En la tabla siguiente se muestran los datos de salida del archivo **OutputOfExample4.xlsx**.

Tabla 11. Detalles de los datos de salida de un archivo de Microsoft Excel

EMP NO	FIRST NAME	MID INIT	LAST NAME	PHONE NO	WORK DEPT	HIREDATE	JOB	SEX	BIRTH DATE	SALARY	BONUS
10	CHRISTINE	I	HAAS	3978	A00	1/1/1995	PRES	F	8/24/1963	152750	1000
20	MICHAEL	L	THOMPSON	3476	A00	10/10/2003	MANAGER	M	2/2/1978	94250	800
30	SALLY	A	KWAN	4738	A00	4/5/2005	MANAGER	F	5/11/1971	98250	800

Tabla 11. Detalles de los datos de salida de un archivo de Microsoft Excel (continuación)

EMP NO	FIRST NAME	MID INIT	LAST NAME	PHONE NO	WORK DEPT	HIRE DATE	JOB	SEX	BIRTH DATE	SALARY	BONUS
50	JOHN	B	GEYER	6789	A00	8/17/1979	MANAGER	M	9/15/1955	80175	800
60	IRVING	F	STERN	6423	B01	9/14/2003	MANAGER	M	7/7/1975	72250	500
70	EVA	D	PULASKI	7831	B01	9/30/2005	MANAGER	F	5/26/2003	96170	700
90	EILEEN	W	HENDERSON	5498	B01	8/15/2000	MANAGER	F	5/15/1971	89750	600
100	THEODORE	Q	SPENSER	972	B01	6/19/2000	MANAGER	M	12/18/1980	86150	500
110	VINCENZO	G	LUCCHESSI	3490	B01	5/16/1988	SALES REP	M	11/5/1959	66500	900
120	SEAN		O'CONNELL	2167	B01	12/5/1993	CLERK	M	10/18/1972	9250	600
130	DELORES	M	QUINTANA	4578	B01	7/28/2001	ANALYST	F	9/15/1955	73800	500
140	HEATHER	A	NICHOLLS	1793	B01	12/15/2006	ANALYST	F	1/19/1976	68420	600
150	BRUCE		ADAMSON	4510	B01	2/12/2002	DESIGNER	M	5/17/1977	55280	500
160	ELIZABETH	R	PIANKA	3782	B01	10/11/2006	DESIGNER	F	4/12/1980	62250	400
170	MASATOSHI	J	YOSHIMURA	2890	B01	9/15/1999	DESIGNER	M	1/5/1981	44680	500
180	MARILYN	S	SCOUTTEN	1682	B01	7/7/2003	DESIGNER	F	2/21/1979	51340	500
190	JAMES	H	WALKER	2986	C01	7/26/2004	DESIGNER	M	6/25/1982	50450	400
200	DAVID		BROWN	4501	C01	3/3/2002	DESIGNER	M	5/29/1971	57740	600
210	WILLIAM	T	JONES	942	C01	4/11/1998	DESIGNER	M	2/23/2003	68270	400
220	JENNIFER	K	LUTZ	672	C01	8/29/1998	DESIGNER	F	3/19/1978	49840	600
230	JAMES	J	JEFFERSON	2094	C01	11/21/1996	CLERK	M	5/30/1980	42180	400
240	SALVATORE	M	MARINO	3780	C01	12/5/2004	CLERK	M	3/31/2002	48760	600
250	DANIEL	S	SMITH	961	C01	10/30/1999	CLERK	M	11/12/1969	49180	400
260	SYBIL	P	JOHNSON	8953	C01	9/11/2005	CLERK	F	10/5/1976	47250	300
270	MARIA	L	PEREZ	9001	D01	9/30/2006	CLERK	F	5/26/2003	37380	500
280	ETHEL	R	SCHNEIDER	8997	D01	3/24/1997	OPERATOR	F	3/28/1976	36250	500
290	JOHN	R	PARKER	4502	D01	5/30/2006	OPERATOR	M	7/9/1985	35340	300
300	PHILIP	X	SMITH	2095	D01	6/19/2002	OPERATOR	M	10/27/1976	37750	400
310	MAUDE	F	SETRIGHT	3332	D01	9/12/1994	OPERATOR	F	4/21/1961	35900	300

Tabla 11. Detalles de los datos de salida de un archivo de Microsoft Excel (continuación)

EMP NO	FIRST NAME	MID INIT	LAST NAME	PHONE NO	WORK DEPT	HIRE DATE	JOB	SEX	BIRTH DATE	SALARY	BONUS
320	RAMLAL	V	MEHTA	9990	D01	7/7/1995	FIELD REP	M	8/11/1962	39950	400
330	WING		LEE	2103	D01	2/23/2006	FIELD REP	M	7/18/1971	45370	500
340	JASON	R	GOUNOT	5698	D01	5/5/1977	FIELD REP	M	5/17/1956	43840	500

Ejemplo 2: Grabar datos a varias hojas de cálculo de un archivo de Microsoft Excel:

Cree un trabajo que utilice la etapa Unstructured Data para grabar datos en varias hojas de cálculo de Microsoft Excel.

Acerca de esta tarea

En este ejemplo se utilizan 7 archivos de texto, DEPT_A00.txt, DEPT_B01.txt, DEPT_C01.txt, DEPT_D11.txt, DEPT_D21.txt, DEPT_E11.txt y DEPT_E21.txt, como datos de origen. Cada archivo de origen contiene información de los empleados del departamento correspondiente, en formato CSV. Puede escribir información de cada archivo de origen en cada hoja de cálculo. El archivo de Microsoft Excel que se ha creado tiene varias hojas de cálculo.

Paso 1: Creación del trabajo:

Cree un trabajo de ejemplo que incluya siete etapas Sequential File y una etapa Unstructured Data.

Procedimiento

1. Inicie el cliente de IBM InfoSphere DataStage and QualityStage Designer.
2. En el panel Repositorio, pulse con el botón derecho del ratón en la carpeta **Trabajos** y seleccione **Nuevo > Trabajo paralelo**.
3. En la sección **Archivo** de la paleta, arrastre una etapa Sequential File al lienzo.
4. Repita el paso 3 seis veces más. Colóquelos en una línea vertical.
5. En la sección **Archivo** de la paleta, arrastre una etapa Unstructured Data al lienzo. Coloque la etapa a la derecha de las etapas Sequential File.
6. Cree un enlace desde una etapa Sequential File a la etapa Unstructured Data.
7. Renombre las etapas y los enlaces. Ponga nombre a los enlaces de forma que coincidan con el nombre del departamento correspondiente (A00, B01, C01, D11, D21, E11, E21).
8. Guarde el trabajo.

Paso 2: Configuración de la etapa Unstructured Data:

Configure la etapa Unstructured Data para extraer datos de un rango en un archivo de Microsoft Excel.

Procedimiento

1. Efectúe una doble pulsación en la etapa Unstructured Data para abrir las propiedades de la etapa.
2. En el separador **Etapa**, seleccione **Excel** en la lista **Tipo de documento**.

3. En la lista **Modalidad de grabación**, seleccione **Crear un archivo**.
4. Pulse **Configurar**.
5. En la ventana Configuración, especifique la vía de acceso completa donde desee crear el archivo de salida, seguido del nombre de archivo OutputOfSample5.txt.
 - a. En **Modalidad de actualización de archivos**, seleccione **Sobrescribir**.
 - b. En **Método de grabación**, seleccione **Archivo específico**. Asegúrese de que la opción **Establecer para todos los enlaces** esté seleccionada.
 - c. En el campo **Cabecera de columna**, seleccione **Nombres de columna**.
 - d. En el campo **Tamaño automático de columnas** en el separador Propiedad, seleccione **Sí**.
 - e. En el panel Orden de hojas, ordene los enlaces A00, B01, C01, D11, D21, E11 y E21 mediante los botones **Subir** y **Bajar**.
 - f. En el panel Orden de hojas, especifique los nombres de hoja, de forma que coincidan con los nombres de enlace correspondientes.
 - g. Pulse **Aceptar**.
6. Pulse **Aceptar**.

Paso 3: Configuración de etapas Sequential File:

Configurar las etapas Sequential File.

Acerca de esta tarea

En este ejemplo se utiliza la etapa Sequential File como etapa de salida. Puede utilizar cualquier otra etapa de salida para crear la salida.

Procedimiento

1. Efectúe una doble pulsación en la etapa Sequential File.
2. En la página **Propiedades**, especifique la vía de acceso del archivo DEPT_A00.txt.
3. En la página Columnas, defina las columnas, tal como se muestra en la figura siguiente.

	Column name	Key	SQL type	Extended	Length	Scale	Nullable	Data element
1	EMPNO	<input checked="" type="checkbox"/>	Char		6		No	
2	FIRSTNME	<input type="checkbox"/>	VarChar		12		No	
3	MIDINIT	<input type="checkbox"/>	Char		1		Yes	
4	LASTNAME	<input type="checkbox"/>	VarChar		15		No	
5	WORKDEPT	<input type="checkbox"/>	Char		3		Yes	
6	PHONENO	<input type="checkbox"/>	Char		4		Yes	
7	HIREDATE	<input type="checkbox"/>	Date		10		Yes	
8	JOB	<input type="checkbox"/>	Char		8		Yes	
9	EDLEVEL	<input type="checkbox"/>	SmallInt		2		No	
10	SEX	<input type="checkbox"/>	Char		1		Yes	
11	BIRTHDATE	<input type="checkbox"/>	Date		10		Yes	
12	SALARY	<input type="checkbox"/>	Decimal		9	2	Yes	
13	BONUS	<input type="checkbox"/>	Decimal		9	2	Yes	
14	COMM	<input type="checkbox"/>	Decimal		9	2	Yes	

4. Pulse **Aceptar**.
5. Repita el Paso 1 a 4 seis veces más para el resto de los archivos de origen, DEPT_B01.txt, DEPT_C01.txt, DEPT_D11.txt, DEPT_D21.txt, DEPT_E11.txt y DEPT_E21.txt. Todas las etapas Sequential File tienen la misma definición de columna en su enlace de salida.

Paso 4: Visualización del resultado del trabajo:

Después de ejecutar el trabajo, abra el archivo y verifique el resultado.

Procedimiento

1. Guarde el trabajo.
2. Compile y ejecute el trabajo.
3. Después de ejecutarse el trabajo, abra el archivo Output0fExample5.xlsx, para ver el resultado. Los datos de salida de una hoja de Microsoft Excel deben coincidir con la información que aparece en el archivo de texto de origen. El archivo Sheet A00 de Microsoft Excel debería coincidir con los datos de origen del archivo DEPT_A00.txt. El archivo Sheet B01 de Microsoft Excel debería coincidir con los datos de origen del archivo DEPT_B01.txt. El archivo Sheet C01 de Microsoft Excel debería coincidir con los datos de origen del archivo DEPT_C01.txt. El archivo Sheet D11 de Microsoft Excel debería coincidir con los datos de origen del archivo DEPT_D11.txt. El archivo Sheet D21 de Microsoft Excel debería coincidir con los datos de origen del archivo DEPT_D21.txt. El archivo Sheet E11 de Microsoft Excel debería coincidir con los datos de origen del archivo DEPT_E11.txt. El archivo Sheet E21 de Microsoft Excel debería coincidir con los datos de origen del archivo DEPT_E21.txt.

Ejemplo 3: Grabación de datos en varios archivos de Microsoft Excel:

Cree un trabajo que utilice la etapa Unstructured Data para grabar datos en varios archivos de Microsoft Excel.

Acerca de esta tarea

En este ejemplo se utiliza un archivo de texto, `Employee.txt`, como datos de origen. El archivo de origen contiene información de 42 empleados en formato CSV. Puede grabar esta información en varios archivos de Microsoft Excel divididos por el número máximo de filas especificado en una opción de hoja.

Paso 1: Creación del trabajo:

Cree un trabajo de ejemplo que incluya una etapa Sequential File y una etapa Unstructured Data.

Procedimiento

1. Inicie el cliente de IBM InfoSphere DataStage and QualityStage Designer.
2. En el panel Repositorio, pulse con el botón derecho del ratón en la carpeta **Trabajos** y seleccione **Nuevo > Trabajo paralelo**.
3. En la sección **Archivo** de la paleta, arrastre una etapa Sequential File al lienzo.
4. En la sección **Archivo** de la paleta, arrastre una etapa Unstructured Data al lienzo. Coloque la etapa a la derecha de la etapa Sequential File.
5. Cree un enlace desde la etapa Sequential File a la etapa Unstructured Data.
6. Cambie el nombre de la etapa y del enlace.
7. Guarde el trabajo.

Paso 2: Configuración de la etapa Unstructured Data:

Configure la etapa Unstructured Data para grabar los datos en una hoja de cálculo de Microsoft Excel.

Procedimiento

1. Efectúe una doble pulsación en la etapa Unstructured Data para abrir las propiedades de la etapa.
2. En el separador **Etapa**, seleccione el tipo de documento como **Excel**, en la lista **Tipo de documento**.
3. En la lista **Modalidad de grabación**, seleccione **Crear un archivo**.
4. Pulse **Configurar**.
5. En la ventana Configuración, especifique la vía de acceso donde desee crear el archivo de salida, seguido del prefijo de archivo `OutputOfSample6.txt`. Cuando se generan los archivos, a este prefijo se le añaden un número secuencial de tres dígitos y una extensión `.xlsx` (p.ej. `001.xlsx`, `002.xlsx`). Por ejemplo, `001.xlsx`, `002.xlsx`.
6. En **Modalidad de actualización de archivos**, seleccione **Sobrescribir**.
7. En **Método de grabación**, seleccione **Generar varios archivos**.
8. En el campo **Cabecera de columna**, seleccione **Nombres de columna**.
9. En el campo **Tamaño automático de columnas** en el separador Propiedad, seleccione **Sí**.
10. En **Número máximo de filas en una hoja**, especifique **10**.
11. En el panel Orden de hojas, especifique **Employee** como Nombre de hoja.
12. Pulse **Aceptar**.

Paso 3: Configuración de etapas Sequential File:

Configurar las etapas Sequential File.

Acerca de esta tarea

En este ejemplo se utiliza la etapa Sequential File como etapa de salida. Puede utilizar cualquier otra etapa de salida para crear la salida.

Procedimiento

1. Efectúe una doble pulsación en la etapa Sequential File.
2. En la página **Propiedades**, especifique la vía de acceso del archivo Employee.txt.
3. En la página Columnas, defina las columnas, tal como se muestra en la figura siguiente.

	Column name	Key	SQL type	Extended	Length	Scale	Nullable	Data element
1	EMPNO	<input checked="" type="checkbox"/>	Char		6		No	
2	FIRSTNME	<input type="checkbox"/>	VarChar		12		No	
3	MIDINIT	<input type="checkbox"/>	Char		1		Yes	
4	LASTNAME	<input type="checkbox"/>	VarChar		15		No	
5	WORKDEPT	<input type="checkbox"/>	Char		3		Yes	
6	PHONENO	<input type="checkbox"/>	Char		4		Yes	
7	HIREDATE	<input type="checkbox"/>	Date		10		Yes	
8	JOB	<input type="checkbox"/>	Char		8		Yes	
9	EDLEVEL	<input type="checkbox"/>	SmallInt		2		No	
10	SEX	<input type="checkbox"/>	Char		1		Yes	
11	BIRTHDATE	<input type="checkbox"/>	Date		10		Yes	
12	SALARY	<input type="checkbox"/>	Decimal		9	2	Yes	
13	BONUS	<input type="checkbox"/>	Decimal		9	2	Yes	
14	COMM	<input type="checkbox"/>	Decimal		9	2	Yes	

4. Pulse **Aceptar**.

Paso 4: Visualización del resultado del trabajo:

Después de ejecutar el trabajo, abra el archivo y verifique el resultado.

Procedimiento

1. Guarde el trabajo.
2. Compile y ejecute el trabajo.
3. Después de ejecutarse el trabajo, abra el archivo de salida para ver el resultado. Los datos de salida de una hoja de Microsoft Excel deben coincidir con la información que aparece en el archivo de texto de origen. Puesto que el número máximo de filas especificado en una hoja es 10, e incluye la fila 1 para la cabecera de columna, cada archivo de salida incluye un máximo de 9 registros del archivo de entrada. El archivo de entrada tiene 42 registros y, por tanto, se generan archivos 5. Se generan los archivos OutputOfExample6_001.xlsx, OutputOfExample6_002.xlsx, OutputOfExample6_003.xlsx, OutputOfExample6_004.xlsx y OutputOfExample6_005.xlsx.

Grabación de datos en archivos existentes de Microsoft Excel

Puede utilizar la etapa Unstructured Data en los trabajos que graben datos a un archivo existente de Microsoft Excel. También puede copiar un archivo de Microsoft Excel, y los datos de grabación, en la copia.

Cuando la etapa Unstructured Data graba datos en un archivo existente de Microsoft Excel, la etapa graba sólo los datos de celda. Si una celda de Microsoft Excel en la que la etapa grabe datos tiene algún formato concreto, la etapa mantendrá el formato existente. Si el archivo de Microsoft Excel tiene una fórmula o un gráfico que haga referencia a celdas que grabe la etapa Unstructured Data, la fórmula o el gráfico se volverán a calcular cuando Microsoft Excel los abra.

Puede grabar en columnas de Microsoft Excel desde cualquier columna de InfoSphere DataStage. No es necesario que los nombres y el orden de las columnas de InfoSphere DataStage y las columnas de Microsoft Excel coincidan. Si la hoja de Microsoft Excel tiene una cabecera en la primera fila, puede configurar la etapa Unstructured Data, de modo que los valores de la primera fila se utilicen para determinar la columna en que se graben los registros. Puede grabar hasta 1.048.576 filas de una hoja de Microsoft Excel. Las columnas de InfoSphere DataStage de origen deben definirse en el tiempo de diseño.

Referencia relacionada:

“Conversiones de tipos de datos de InfoSphere DataStage a Microsoft Excel” en la página 45

Para que etapa Unstructured Data pueda grabar los datos que se extraen de los enlaces de entrada en archivos de Microsoft Excel, los datos se convierten a tipos de datos de Microsoft Excel.

Diseño de trabajos que graben datos en una hoja existente de Microsoft Excel

Puede utilizar una etapa Unstructured Data para diseñar trabajos que graben datos en un archivo existente de Microsoft Excel. Una etapa Unstructured Data puede actualizar un archivo de Microsoft Excel cada vez. Sólo puede tener una etapa Unstructured Data que actualice el archivo de Microsoft Excel en un trabajo.

Antes de empezar

- Instale InfoSphere Information Server con el idioma que coincida con el idioma del archivo de Microsoft Excel que desee extraer.
- Asegúrese de que la aplicación que muestra el contenido de las hojas de cálculo de Microsoft Excel (por ejemplo, Microsoft Excel, Microsoft Excel Viewer o IBM Lotus Symphony) está instalado en el sistema cliente.
- Asegúrese de que la extensión de archivo .xlsx esté asociada a la aplicación que se utilice para ver hojas de cálculo de Microsoft Excel.

Procedimiento

1. Defina un trabajo que incluya una etapa Unstructured Data.
2. Configure la etapa Unstructured Data como destino.
3. Compile y ejecute el trabajo.

Definición de un trabajo que incluya una etapa Unstructured Data:

Para poder leer datos desde, o grabarlos a, archivos de Microsoft Excel, debe crear un trabajo que incluya la etapa Unstructured Data, añadir las etapas adicionales necesarias y crear los enlaces necesarios.

Procedimiento

1. En el Cliente del Diseñador, pulse **Archivo > Nuevo**.
2. En la ventana Nuevo, pulse el icono **Trabajo paralelo** y luego pulse **Aceptar**.
3. En la Paleta, pulse **Archivo**.
4. Arrastre el icono **Etapa Unstructured Data** al lienzo.
5. Creer etapas para el trabajo.
6. En el lado izquierdo del cliente del Diseñador, en el menú Paleta, seleccione la categoría **General** y, a continuación, cree los enlaces necesarios para el trabajo.
7. (Opcional) Efectúe una doble pulsación en el icono **Etapa Unstructured Data** para especificar o modificar los atributos siguientes:
 - **Etapa:** Modifique el nombre predeterminado de la **Etapa**. Puede entrar hasta 255 caracteres. Como alternativa, puede modificar el nombre de la etapa en el lienzo de diseño del trabajo.
 - **Descripción:** Escriba una descripción de la etapa.
8. Pulse **Guardar**.

Configuración de la etapa Unstructured Data como destino:

Puede configurar la etapa Unstructured Data para modificar un archivo existente de Microsoft Excel. No se da soporte a la ejecución en modalidad paralela cuando se modifica un archivo existente de Microsoft Excel. Debe configurar la etapa Unstructured Data en una modalidad secuencial.

Acerca de esta tarea

La etapa Unstructured Data sólo da soporte al formato OOXML (.xlsx) de archivos de Microsoft Excel como archivo de destino y archivo de plantilla.

La etapa Unstructured Data no da soporte a la modificación de archivos .xls o archivos cifrados por contraseña.

La etapa Unstructured Data no soporta archivos de Microsoft Excel que se han creado mediante Microsoft Excel para Mac.

Los nombres de archivo, nombres de hoja y nombres de cabecera deben ser capaces de expresarse en la página de códigos predeterminada de Microsoft Windows donde el cliente del Diseñador de InfoSphere DataStage está instalado. Si el archivo de Microsoft Excel que desea procesar incluye otros caracteres, deberá editarlo para que no incluya dichos caracteres.

Procedimiento

1. En el lienzo paralelo, efectúe una doble pulsación en la etapa **Unstructured Data**.
2. En el separador **Entrada**, seleccione el enlace de entrada del campo **Nombre de entrada (etapa en sentido ascendente)**.
3. En la página Columnas, asegúrese de que las columnas se hayan definido correctamente.
4. En el separador **Etapa**, seleccione el tipo de documento como **Excel**, en la página Propiedades.
5. En la lista **Modalidad de grabación**, seleccione **Modificar archivo existente**.
6. Para configurar las propiedades para grabar datos en un archivo existente de Microsoft Excel, pulse **Configurar**.

7. Especifique los detalles del archivo de destino en el que deban grabarse los datos.
 - a. En el campo **Nombre de archivo**, especifique el nombre del archivo en que deban grabarse los datos.
 - b. Si deben actualizarse más de 32 columnas en la hoja de Microsoft Excel, especifique el número de columnas en el campo **Número de columnas a cargar**.
 - c. Para cargar las columnas en el panel Importar, pulse **Cargar**.
8. En la lista **Enlace**, seleccione un enlace de entrada que se deba configurar.
9. Especifique los detalles de Microsoft Excel que deban importarse en el panel Importar.
 - a. En la lista de hojas, seleccione la hoja de Microsoft Excel que deba actualizarse.
 - b. En la lista de columnas, seleccione las columnas de Microsoft Excel que deban actualizarse.
 - c. Pulse **Importar**.
10. Correlacione las columnas de Microsoft Excel importadas con las columnas de InfoSphere DataStage que se hayan definido en el enlace de entrada. En la columna **InfoSphere DataStage**, seleccione una columna de InfoSphere DataStage en la que grabar los datos.
11. Repita los pasos 7 al 9 para todos los enlaces de entrada.
12. Pulse **Aceptar**.

Utilización de parámetros de trabajo al modificar un archivo existente de Microsoft Excel:

La etapa Unstructured Data no puede crear parámetros de trabajo nuevos en la ventana Configuración. Pero el usuario puede utilizar los parámetros de trabajo en la ventana Configuración. Debe crear parámetros de trabajo en la ventana Propiedades de trabajo antes o después de trabajar en la ventana Configuración. Para ello seleccione **Editar > Propiedades de trabajo** en el cliente de IBM InfoSphere DataStage and QualityStage Designer. Para obtener más información sobre la creación de parámetros de trabajo, consulte la Lección 2.4: Adición de parámetros, en la guía de aprendizaje de trabajos paralelos de IBM InfoSphere DataStage.

Los parámetros de trabajo se especifican en la ventana Configuración entre caracteres #. Por ejemplo, el parámetro de trabajo *FileName*, se especifica como **#FileName#** en la ventana Configuración. Para un campo de tipo serie tal como la propiedad **Nombre de archivo**, puede escribir directamente el nombre del parámetro de trabajo entre los caracteres #.

Si desea utilizar un parámetro de trabajo para una propiedad de tipo List tal como **Acción si el archivo ya existe**, debe crear un parámetro de tipo List que contenga una lista de variables de tipo serie. Las variables de tipo serie deben coincidir con el texto de la etiqueta de la propiedad correspondiente contenida en la ventana Configuración. Por ejemplo, si desea utilizar el parámetro de trabajo para la propiedad **Acción si el archivo ya existe**, debe crear un parámetro de trabajo de tipo List que contenga la variable de tipo serie **Error** y **Sobrescribir**. Después de crear un parámetro de trabajo, seleccione **<Parametrizar...>** en la ventana Configuración y especifique el nombre del parámetro de trabajo entre los caracteres # en el cuadro de diálogo **Parámetro de entrada**. Pulse **Cargar** para editar o seleccionar variables en el panel Resolver parámetros de trabajo.

Ejemplo: Grabación de datos en archivos existentes de Microsoft Excel

Cree un trabajo que utilice la etapa Unstructured Data para grabar datos en los archivos existentes de Microsoft Excel.

Acerca de esta tarea

En este ejemplo se utiliza un archivo de texto, Employee.txt, como datos de origen. El archivo de origen contiene información de los empleados en formato CSV. Puede grabar esta información en el archivo de Microsoft Excel ExcelModifySample1.xlsx. El archivo ExcelModifySample1.xlsx tiene una hoja denominada Employee que contiene las columnas **EMP NO**, **FIRST NAME**, **MIDINIT**, **LAST NAME**, **HIRE DATE**, **JOB**, **SEX**, **SALARY**, **BONUS** y **TOTAL PAY**. En este trabajo, puede grabar en estas columnas, excepto **TOTAL PAY**. La columna **TOTAL PAY** tiene una fórmula para calcular el salario y la bonificación.

Puede crear trabajos de ejemplo para grabar datos a un archivo existente de Microsoft Excel.

Para obtener los archivos de los ejemplos, extraiga el archivo vía acceso instalación_IS\Clients\Samples\Connectors\UnstructuredData_Samples.zip.

Paso 1: Creación del trabajo:

Cree un trabajo de ejemplo que incluya una etapa Sequential File y una etapa Unstructured Data. En este ejemplo, la etapa Sequential File lee datos de un archivo de Microsoft Excel y, a continuación, la etapa Unstructured Data graba datos en el origen de Unstructured Data.

Procedimiento

1. Inicie el cliente de IBM InfoSphere DataStage and QualityStage Designer.
2. En el panel Repositorio, pulse con el botón derecho del ratón en la carpeta **Trabajos** y seleccione **Nuevo > Trabajo paralelo**.
3. En la sección **Archivo** de la paleta, arrastre una etapa Sequential File al lienzo.
4. En la sección **Archivo** de la paleta, arrastre una etapa Unstructured Data al lienzo. Coloque la etapa a la derecha de la etapa Sequential File.
5. Cree un enlace desde la etapa Sequential File a la etapa Unstructured Data.
6. Cambie el nombre de la etapa y del enlace.
7. Guarde el trabajo.

Paso 2: Configuración de la etapa Sequential File:

Configure la etapa Sequential File para leer datos del archivo de origen. Debe especificar el nombre del archivo de origen, definir los nombres de columna y las propiedades SQL.

Antes de empezar

En este ejemplo se utiliza la etapa Sequential File como etapa de salida. Puede utilizar cualquier otra etapa de salida para crear la salida.

Procedimiento

1. Efectúe una doble pulsación en la etapa Sequential File.

2. En la página **Propiedades**, especifique la vía de acceso del archivo `Employee.txt`.
3. En la página **Columnas**, defina las columnas, tal como se muestra en la tabla siguiente.

Nombre de columna	Clave	Tipo de SQL	Ampliado	Longitud	Esca-la	Con posibilidad de nulos	Descripción
EMP_NO		Integer				Sí	
FIRST_NAME		VarChar				Sí	
MIDINIT		VarChar				Sí	
LAST_NAME		VarChar				Sí	
HIRE_DATE		Date				Sí	
JOB		VarChar				Sí	
SEX		VarChar				Sí	
SALARY		Integer				Sí	
BONUS		Integer				Sí	

4. Pulse **Aceptar**.

Paso 3: Configuración de la etapa Unstructured Data:

Configure la etapa Unstructured Data para grabar los datos en un archivo existente de Microsoft Excel.

Procedimiento

1. Efectúe una doble pulsación en la etapa Unstructured Data para abrir las propiedades de la etapa.
2. En el separador **Etapa**, seleccione **Excel** en la lista **Tipo de documento**.
3. En la lista **Modalidad de grabación**, seleccione **Crear un archivo**.
4. Pulse **Configurar**.
5. En la ventana Configuración, especifique la vía de acceso donde desee crear el archivo de salida, seguido del prefijo de archivo `ExcelModifySample1.txt`.
6. Pulse **Cargar**.
7. En el campo **Cabecera de columna**, seleccione **Primera fila**.
8. En el campo **Empezar a grabar desde esta fila**, especifique **2**.
9. En la lista **Hoja**, seleccione **Employee**. Asegúrese de que se hayan seleccionado las columnas de Microsoft Excel A a I.
10. Pulse **Importar**.
11. En el panel Correlacionar, defina la correlación entre la columna de Microsoft Excel y la columna de InfoSphere DataStage.
12. Pulse **Aceptar**.

Paso 4: Visualización del resultado del trabajo:

Después de ejecutar el trabajo, abra el archivo y verifique el resultado.

Procedimiento

1. Guarde el trabajo.
2. Compile y ejecute el trabajo.
3. Después de ejecutarse el trabajo, abra el archivo de salida para ver el resultado. Los datos de salida de una hoja de Microsoft Excel deben coincidir con la información que aparece en el archivo de texto de origen.

Capítulo 2. Referencia

Estos temas describen los tipos soportados de Microsoft Excel y las correlaciones de tipos entre Microsoft Excel y DataStage. También describen las condiciones de terminación anómala de trabajos de IBM Software Support.

Conversiones de tipos de datos desde Microsoft Excel a InfoSphere DataStage

Antes de que los datos extraídos de Microsoft Excel se escriban en el enlace de salida mediante la etapa Unstructured Data, los datos se convierten a tipos de datos de InfoSphere DataStage.

La tabla siguiente muestra la correlación entre los tipos de datos de Microsoft Excel e InfoSphere DataStage.

Nota: La etapa Unstructured Data correlaciona las conversiones de tipos de datos de Microsoft Excel a InfoSphere DataStage sólo cuando la etapa Unstructured Data lee los registros del origen de datos de Microsoft Excel.

Tabla 12. Correlación entre los tipos de datos de Microsoft Excel e InfoSphere DataStage

Tipo de datos de la celda Microsoft Excel	Tipo de datos de DataStage
En blanco	Tipos de datos enteros BigInt Integer SmallInt TinyInt Tipos de datos de texto Char VarChar LongVarChar Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar Tipos de datos fraccionarios Double Float Real Tipos de datos decimales Decimal Numeric Tipo de datos de fecha y hora Date Time Timestamp
Booleano	Tipos de datos enteros BigInt Integer SmallInt TinyInt Nota: Correlaciona TRUE: 1, FALSE: 0 Tipos de datos de texto Char VarChar LongVarChar Nota: Correlaciona TRUE: "true", FALSE: "false" Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar Nota: Correlaciona TRUE: "true", FALSE: "false"

Tabla 12. Correlación entre los tipos de datos de Microsoft Excel e InfoSphere DataStage (continuación)

Tipo de datos de la celda Microsoft Excel	Tipo de datos de DataStage
Error	<p>Tipos de datos de texto Char VarChar LongVarChar Nota: Expresión de tipo serie del error. Por ejemplo, #NAME?</p> <p>Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar Nota: Expresión de tipo serie del error. Por ejemplo, #NAME?</p>
Numérico	<p>Tipos de datos enteros BigInt Integer SmallInt TinyInt</p> <p>Tipos de datos de texto Char VarChar LongVarChar</p> <p>Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar</p> <p>Tipos de datos fraccionarios Double Float Real</p> <p>Tipos de datos decimales Decimal Numeric</p> <p>Tipo de datos de fecha y hora Date Time Timestamp</p>
Serie	<p>Tipos de datos de texto Char VarChar LongVarChar</p> <p>Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar</p> <p>Tipo de datos de fecha y hora Date Time Timestamp</p>

Tabla 13. Otros tipos de datos de Microsoft Excel e InfoSphere DataStage

Otros tipos de datos de Microsoft Excel	Tipo de datos de InfoSphere DataStage
Fórmula	<p>Tipos de datos de texto Char VarChar LongVarChar</p> <p>Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar</p>
Comentario	<p>Tipos de datos de texto Char VarChar LongVarChar</p> <p>Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar</p>
Autor de comentario	<p>Tipos de datos de texto Char VarChar LongVarChar</p> <p>Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar</p>

Tabla 13. Otros tipos de datos de Microsoft Excel e InfoSphere DataStage (continuación)

Otros tipos de datos de Microsoft Excel	Tipos de datos de InfoSphere DataStage
Tipo de hipervínculo	Tipos de datos enteros BigInt Integer SmallInt TinyInt Tipos de datos de texto Char VarChar LongVarChar Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar
Dirección de hipervínculo	Tipos de datos de texto Char VarChar LongVarChar Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar
Etiqueta de hipervínculo	Tipos de datos de texto Char VarChar LongVarChar Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar

Tabla 14. Correlación entre tipos de datos de Microsoft Excel e InfoSphere DataStage

Tipo de datos de la celda Microsoft Excel	Tipos de datos de InfoSphere DataStage
Nombre de archivo	Tipos de datos de texto Char VarChar LongVarChar Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar
Vía de acceso de archivo	Tipos de datos de texto Char VarChar LongVarChar Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar
Tamaño de archivo	Tipos de datos enteros BigInt Integer SmallInt TinyInt Tipos de datos de texto Char VarChar LongVarChar Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar
Fecha de última modificación	Tipos de datos de texto Char VarChar LongVarChar Nota: Expresión de tipo serie con el formato aaaa-mm-dd Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar Nota: Expresión de tipo serie con el formato aaaa-mm-dd Tipo de datos de fecha y hora Date Time Timestamp

Tabla 15. Correlación entre propiedades de documento de Microsoft Excel y tipos de datos de InfoSphere DataStage

Propiedades de documento de Microsoft Excel	Tipos de datos de InfoSphere DataStage
Autores	<p>Tipos de datos de texto Char VarChar LongVarChar</p> <p>Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar</p>
Comentarios de documento	<p>Tipos de datos de texto Char VarChar LongVarChar</p> <p>Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar</p>
Fecha de creación del contenido	<p>Tipos de datos de texto Char VarChar LongVarChar Nota: Expresión de tipo serie con el formato aaaa-mm-dd</p> <p>Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar Nota: Expresión de tipo serie con el formato aaaa-mm-dd</p> <p>Tipo de datos de fecha y hora Date Time Timestamp</p>
Palabras clave	<p>Tipos de datos de texto Char VarChar LongVarChar</p> <p>Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar</p>
Número de revisión	<p>Tipos de datos de texto Char VarChar LongVarChar</p> <p>Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar</p>
Asunto	<p>Tipos de datos de texto Char VarChar LongVarChar</p> <p>Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar</p>
Cargo	<p>Tipos de datos de texto Char VarChar LongVarChar</p> <p>Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar</p>
Empresa	<p>Tipos de datos de texto Char VarChar LongVarChar</p> <p>Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar</p>

Tabla 15. Correlación entre propiedades de documento de Microsoft Excel y tipos de datos de InfoSphere DataStage (continuación)

Propiedades de documento de Microsoft Excel	Tipos de datos de InfoSphere DataStage
Categoría	Tipos de datos de texto Char VarChar LongVarChar Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar
Gestor	Tipos de datos de texto Char VarChar LongVarChar Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar

Tabla 16. Correlación entre propiedades personalizadas de Microsoft Excel y tipos de datos de InfoSphere DataStage

Propiedad personalizada de Microsoft Excel	Tipos de datos de InfoSphere DataStage
Texto	Tipos de datos de texto Char VarChar LongVarChar Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar
Fecha	Tipos de datos de texto Char VarChar LongVarChar Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar Tipo de datos de fecha y hora Date Time Timestamp
Número	Tipos de datos enteros BigInt Integer SmallInt TinyInt Nota: Si el valor es un entero. Tipos de datos de texto Char VarChar LongVarChar Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar Tipos de datos fraccionarios Double Float Real Tipos de datos decimales Decimal Numeric

Tabla 16. Correlación entre propiedades personalizadas de Microsoft Excel y tipos de datos de InfoSphere DataStage (continuación)

Propiedad personalizada de Microsoft Excel	Tipos de datos de InfoSphere DataStage
Booleano	<p>Tipos de datos enteros BigInt Integer SmallInt TinyInt Nota: Correlaciona TRUE: 1, FALSE: 0</p> <p>Tipos de datos de texto Char VarChar LongVarChar Nota: Correlaciona TRUE: "true", FALSE: "false"</p> <p>Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar Nota: Correlaciona TRUE: "true", FALSE: "false"</p>

Tabla 17. Correlación entre la información de hojas de cálculo Microsoft Excel y tipos de datos de InfoSphere DataStage

Información de hoja de cálculo de Microsoft Excel	Tipos de datos de InfoSphere DataStage
Nombre de hoja	<p>Tipos de datos de texto Char VarChar LongVarChar</p> <p>Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar</p>
Cabecera	<p>Tipos de datos de texto Char VarChar LongVarChar</p> <p>Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar</p> <p>Nota: Para los tipos de datos de texto y tipos de datos de texto de idioma nacional, Microsoft Excel soporta mandatos especiales representados por una sola letra precedida de un carácter "&" en la cabecera y pie de página de Microsoft Excel. La etapa Unstructured Data no convierte esas letras y simplemente las conserva en el texto extraído. Consulte http://msdn.microsoft.com/en-us/library/dd773041%28v=office.12%29.aspx para obtener más información sobre los mandatos especiales.</p>

Tabla 17. Correlación entre la información de hojas de cálculo Microsoft Excel y tipos de datos de InfoSphere DataStage (continuación)

Información de hoja de cálculo de Microsoft Excel	Tipos de datos de InfoSphere DataStage
Pie de página	<p>Tipos de datos de texto Char VarChar LongVarChar</p> <p>Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar</p> <p>Nota: Para los tipos de datos de texto y tipos de datos de texto de idioma nacional, Microsoft Excel soporta mandatos especiales representados por una sola letra precedida de un carácter "&" en la cabecera y pie de página de Microsoft Excel. La etapa Unstructured Data no convierte esas letras y simplemente las conserva en el texto extraído. Consulte http://msdn.microsoft.com/en-us/library/dd773041%28v=office.12%29.aspx para obtener más información sobre los mandatos especiales.</p>

Tabla 18. Correlación entre la información de filas de Microsoft Excel y los tipos de datos equivalentes de InfoSphere DataStage

Información de filas de Microsoft Excel	Tipos de datos de InfoSphere DataStage
Número de fila	<p>Tipos de datos enteros BigInt Integer SmallInt TinyInt</p> <p>Tipos de datos de texto Char VarChar LongVarChar</p> <p>Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar</p>
Está oculta	<p>Tipos de datos enteros BigInt Integer SmallInt TinyInt Nota: Correlaciona TRUE: 1, FALSE: 0</p> <p>Tipos de datos de texto Char VarChar LongVarChar Nota: Correlaciona TRUE: "true", FALSE: "false"</p> <p>Tipos de datos de texto de idioma nacional NChar NVarChar LongNVarChar Nota: Correlaciona TRUE: "true", FALSE: "false"</p>

Conversiones de tipos de datos de InfoSphere DataStage a Microsoft Excel

Para que etapa Unstructured Data pueda grabar los datos que se extraen de los enlaces de entrada en archivos de Microsoft Excel, los datos se convierten a tipos de datos de Microsoft Excel.

En la tabla siguiente se muestra la correlación entre los tipos de datos de InfoSphere DataStage y los tipos de datos de Microsoft Excel.

Nota: La etapa Unstructured Data correlaciona las conversiones de tipos de datos de tipos de datos InfoSphere DataStage a archivos de Microsoft Excel, sólo cuando la etapa Unstructured Data graba registros en el origen de datos de Microsoft Excel.

Tabla 19. Correlación entre tipos de datos de Microsoft Excel e InfoSphere DataStage

Tipo de datos de InfoSphere DataStage	Tipo de datos de la celda Microsoft Excel
Tipos de datos de texto Char, VarChar, LongVarChar Tipos de datos de texto de idioma nacional NChar, NVarChar, LongNVarChar	Serie
Tipos de datos enteros BigInt, Integer, SmallInt, TinyInt Tipos de datos fraccionarios Double, Float, Real Tipos de datos decimales Decimal, Numeric	Numeric
Tipos de datos de fecha y hora Date, Time, Timestamp	Numeric Nota: Esta nota sólo es aplicable cuando se graban datos en los archivos existentes de Microsoft Excel. <ul style="list-style-type: none"> La etapa Unstructured Data no establece el formato para los tipos de datos de fecha y hora. Se recomienda establecer el formato adecuado para expresar la fecha, la hora o la indicación de fecha y hora del archivo de Microsoft Excel.
Tipos de datos binarios Binary, VarBinary, LongVarBinary, Bit	No están soportados.

Conceptos relacionados:

“Grabación de datos en archivos existentes de Microsoft Excel” en la página 32
 Puede utilizar la etapa Unstructured Data en los trabajos que graben datos a un archivo existente de Microsoft Excel. También puede copiar un archivo de Microsoft Excel, y los datos de grabación, en la copia.

Condiciones de terminación anómala de trabajo en Microsoft Excel

Las tablas describen las diferentes condiciones de terminación anómala de trabajo en archivos de Microsoft Excel.

Nota: La condición siguiente sólo se aplica cuando la etapa Unstructured Data lee los registros de los archivos de Microsoft Excel.

Nombre de archivo (no se utilizan caracteres comodín)

Cuando el nombre de archivo se utiliza sin caracteres comodín, se pueden producir las siguientes condiciones de terminación anómala de trabajo:

Tabla 20. Nombre de archivo (no se utilizan caracteres comodín)

Condición	Resultado
El archivo especificado no existe.	El trabajo termina de forma anómala.
El usuario no tiene permiso para leer el archivo especificado.	El trabajo termina de forma anómala.
El archivo especificado no es un archivo de Microsoft Excel válido.	El trabajo termina de forma anómala.
El archivo no se puede abrir con la contraseña especificada.	El trabajo termina de forma anómala.

Nombre de archivo (se utilizan caracteres comodín)

Cuando el nombre de archivo se utiliza con caracteres comodín, se pueden producir las siguientes condiciones de terminación anómala de trabajo:

Tabla 21. Nombre de archivo (se utilizan caracteres comodín)

Condición	Resultado
No existe ningún archivo con el nombre especificado.	El trabajo continúa con un mensaje de aviso (ninguna fila de salida).
El usuario no tiene permiso para leer el archivo especificado.	El trabajo continúa con un mensaje de aviso si la propiedad Acción para error está establecida en Omitir. En otro caso, el trabajo termina de forma anómala.
El archivo especificado no es un archivo de Microsoft Excel válido.	El trabajo continúa con un mensaje de aviso si la propiedad Acción para error está establecida en Omitir. En otro caso, el trabajo termina de forma anómala.
El archivo no se puede abrir con la contraseña especificada.	El trabajo continúa con un mensaje de aviso si la propiedad Acción para error está establecida en Omitir. En otro caso, el trabajo termina de forma anómala.

Nombre de hoja

Se pueden producir las siguientes condiciones de terminación anómala de trabajo para el nombre de hoja:

Tabla 22. Nombre de hoja

Condición	Resultado
El nombre de hoja está especificado en el rango de datos y la hoja no existe.	El trabajo continúa con un mensaje de aviso si la propiedad Acción para error está establecida en Omitir. En otro caso, el trabajo termina de forma anómala.

Cabecera de columna

Se puede producir la siguiente condición de terminación anómala de trabajo para la cabecera de columna:

Tabla 23. Cabecera de columna

Condición	Resultado
La primera fila es cabecera de columna y el valor de la primera fila no coincide con el valor de la primera fila del rango de datos de plantilla.	El trabajo continúa y se registra un mensaje informativo.

Tipo de datos

Se pueden producir las siguientes condiciones de terminación anómala de trabajo para el tipo de datos:

Tabla 24. Tipo de datos

Condición	Resultado
El tipo de datos no está soportado para extraer el tipo de objeto de Microsoft Excel que está correlacionado con la columna de DataStage.	El trabajo termina de forma anómala.
El tipo de datos está soportado para extraer el tipo de objeto de Microsoft Excel que está correlacionado con la columna de DataStage, pero no coincide con la instancia (Esto ocurre cuando el objeto de Microsoft Excel es una celda o propiedad personalizada).	El trabajo continúa con un mensaje de aviso si la propiedad Acción para error está establecida en <code>Omitir</code> . En otro caso, el trabajo termina de forma anómala.

Propiedad personalizada

Se puede producir la siguiente condición de terminación anómala de trabajo para la propiedad personalizada:

Tabla 25. Propiedad personalizada

Condición	Resultado
La propiedad especificada no existe.	El trabajo continúa con un mensaje de aviso si la propiedad Acción para error está establecida en <code>Omitir</code> . En otro caso, el trabajo termina de forma anómala.

Capítulo 3. Resolución de problemas

Utilice la información de esta sección como ayuda para comprender, identificar y resolver problemas de la etapa Unstructured Data de InfoSphere DataStage.

Los mensajes que se muestran en la parte inferior de la ventana de configuración están truncados

Si observa que los mensajes mostrados en la parte inferior de la ventana de configuración están truncados, coloque el ratón sobre el área de mensajes para ver el mensaje completo.

No se puede obtener el área de datos de plantilla esperada

Si pulsa **Cargar** sin especificar la expresión de rango o sólo especifica el nombre de la hoja en la expresión de rango, se muestra una lista de áreas de datos de plantilla para la correlación de columnas. Sin embargo, con algunos archivos de Microsoft Excel, puede que no aparezca el área de datos de plantilla de su interés.

Para obtener una lista completa de la expresión de rango para el área de datos de plantilla, especifique la expresión de rango completa para el área de datos de plantilla prevista y luego pulse **Cargar**. Si especifica la célula inicial para el área de datos de plantilla esperada, puede obtener el área de datos de plantilla que empieza en la celda especificada.

Se muestra un mensaje de aviso al modificar la configuración de la expresión de rango

El campo **Expresión de rango** se actualiza cuando selecciona una de las áreas de datos de plantilla en el recuadro de lista y pulsa **Correlacionar**. Puede ver una expresión de rango pertinente que está asociada con el área de datos de plantilla seleccionada. Pero puede que desee cambiar la expresión de rango. Por ejemplo, de forma predeterminada, la etapa Unstructured Data devuelve la información de la expresión de rango incluido el nombre de hoja. Pero puede modificar la configuración para visualizar la expresión de rango sin el nombre de hoja. Cuando pulsa **Aceptar** para guardar la configuración modificada, obtiene el mensaje de aviso siguiente:

```
El origen de datos ha cambiado desde  
que se creó la correlación de columnas.  
Los cambios podrían generar  
un error de tiempo de ejecución.  
¿Desea guardar los cambios?
```

Si los cambios no son necesarios para la correlación de columnas, pulse **Aceptar** para completar la configuración y confirmar que la expresión de rango actualizada es coherente con la correlación.

Error de tiempo de espera excedido

Se puede producir el error siguiente cuando intenta cargar un archivo grande de Microsoft Excel en la ventana Configuración y la operación no finaliza dentro del tiempo especificado para el servicio CAS:

No se ha podido procesar la solicitud:
No se ha podido recibir la respuesta del manejador: la solicitud ha excedido el tiempo de espera.

Cuando aparezca este error, modifique el valor de tiempo de espera especificado mediante el mandato **PropertyAdmin**.

Por ejemplo, para cambiar el valor de tiempo de espera a 180 segundos, especifique el mandato siguiente en la capa de servicios:

```
InformationServer/ASBServer/bin/PropertyAdmin -set -key cas.agent.timeout  
-value 180
```

Donde *InformationServer* es el directorio de instalación de InfoSphere Information Server.

Error de falta de memoria al cargar un archivo grande de Microsoft Excel

Puede encontrar el error siguiente al intentar cargar un archivo grande de Microsoft Excel en la ventana Configuración:

El archivo {0} no se puede cargar porque no hay memoria suficiente. Puede que el archivo sea demasiado grande (su tamaño es {1} bytes). Especifique un archivo más pequeño como plantilla.

Donde {0} indica el nombre de archivo especificado por el usuario y {1} indica el tamaño del archivo especificado.

Cuando carga un archivo grande de Microsoft Excel, se utiliza una gran cantidad de memoria de almacenamiento dinámico Java. Como resultado, el servicio CAS (Connection Access Service), que es un proceso Java utilizado por la etapa Unstructured Data, puede quedar detenido. Si el proceso se detiene, se cancela la solicitud y se muestra el mensaje de error.

Para evitar el error, siga uno de estos métodos:

- Especifique un archivo de Microsoft Excel más pequeño.
- Cree un archivo menor copiando el archivo de Microsoft Excel original. Suprima las filas y hojas que no se utilizan para la correlación de columnas.

Error relacionado con el cambio de tamaño del almacenamiento dinámico de ASBAgent

Puede encontrar el error siguiente al intentar cargar un archivo grande de Microsoft Excel:

No se ha encontrado cómo cambiar el tamaño del almacenamiento dinámico de ASBAgent

Si habilita la vista de registro para la categoría **ISF-CAS-NATIVE**, puede ver anotado el mensaje siguiente:

Se ha recibido un mensaje de aviso de la capa nativa (C++):
El archivo {0} no se puede cargar porque no hay memoria suficiente.

Puede que el archivo sea demasiado grande (su tamaño es {1} bytes).
 El tamaño máximo del almacenamiento dinámico de la JVM es {2}. El tamaño del almacenamiento dinámico consumido es {3}. {4}

Donde,

- {0} indica el nombre de archivo especificado por el usuario
- {1} indica el tamaño del archivo especificado
- {2} indica el tamaño máximo del almacenamiento dinámico de la JVM
- {3} indica el tamaño del almacenamiento dinámico consumido actualmente

Para corregir el problema, aumente el tamaño de la memoria de almacenamiento dinámico Java del ASBAgent.

Error de excepción de Java en tiempo de ejecución

Puede encontrar el error grave siguiente cuando se procesa un archivo grande de Microsoft Excel:

```
Unstructured_Data_0,0: Se ha producido una excepción Java en tiempo de ejecución:
java.lang.OutOfMemoryError (java.util.Arrays::copyOfRange, file Arrays.java,
line 4,138)
```

El error se produce cuando el tamaño de la memoria de almacenamiento dinámico Java no es suficiente.

Para corregir el problema, aumente el tamaño del almacenamiento dinámico de Java disponible estableciendo la variable de entorno `CC_UNST_JAVA_HEAP`. El valor de la variable de entorno es el valor entero del tamaño de almacenamiento dinámico Java expresado en MB. Por ejemplo, para establecer el tamaño del almacenamiento dinámico Java en 512 MB, especifique `CC_UNST_JAVA_HEAP=512`. El tamaño predeterminado del almacenamiento dinámico es 256 MB.

Extraer archivo de Microsoft Excel		
	Archivo de Microsoft Excel 97-2003 (.xls)	6 veces del tamaño de archivo
	Archivo de Microsoft Excel 2007-2010 (.xlsx) protegido por contraseña	30 veces del tamaño de archivo
	Archivo de Microsoft Excel 2007-2010 (.xlsx) desprotegido	Tamaño de almacenamiento dinámico predeterminado
Modificar archivo de Microsoft Excel		300 veces el resultado del tamaño de archivo de Microsoft Excel
Crear archivo de Microsoft Excel		Tamaño de almacenamiento dinámico predeterminado

Nota: El tamaño de almacenamiento dinámico necesario real varía en función del tipo de datos que contiene el archivo de Microsoft Excel y la plataforma, y puede exceder el tamaño que figura en la tabla anterior.

No se puede encontrar un rango apropiado para especificar

- Es posible que no pueda encontrar una expresión de rango apropiada para especificar cuando desee extraer rangos de datos de varias hojas o varios archivos en las situaciones siguientes:
 - Cuando cada rango de datos comienza en posiciones diferentes
 - Cuando el final de los rangos de datos no es la última fila de las hojas

En estos casos, defina el nombre de los rangos de datos en los archivos de origen de Microsoft Excel y especifique el nombre como expresión de rango.

No se ha podido compilar un trabajo creado en la versión anterior

Puede encontrar el error siguiente al compilar un trabajo que se cree en InfoSphere DataStage versión 9.1

Nota: No se ha especificado un valor para la propiedad Expresión de rango. En el editor de etapas, pulse **Configurar** y luego especifique un valor para la propiedad. Si este trabajo se ha creado en InfoSphere DataStage version 9.1, deberá abrir la ventana de configuración y guardar el trabajo.

Este error se produce si realiza una de las acciones siguientes.

1. Abra el editor de etapas.
2. Pulse **Aceptar** en Editor de etapas sin pulsar **Configurar**.

o bien

1. Abra el editor de etapas.
2. Pulse **Configurar**.
3. Pulse **Cancelar** en la ventana Configuración.
4. Pulse **Aceptar** en el editor de etapas.

Nota: Si no abre el editor de etapas o no pulsar **Aceptar** en el editor de etapas, puede ejecutar un trabajo creado en InfoSphere DataStage versión 9.1 sin tener que realizar ningún tipo de acción.

Para arreglar este error, efectúe los pasos siguientes.

1. Abra el editor de etapas.
2. Pulse **Configurar**.
3. Pulse **Aceptar** en la ventana Configuración.
4. Pulse **Aceptar** en el editor de etapas.

Capítulo 4. Variables de entorno: Etapa Unstructured Data

La etapa Unstructured Data utiliza estas variables de entorno.

CC_JNI_EXT_DIRS

Establezca esta variable de entorno para añadir un prefijo a la vía de acceso de clases de la propiedad del sistema `java.ext.dirs`.

Cuando se establece el valor de esta variable de entorno, se añade un prefijo a la vía de acceso de clases de la propiedad del sistema `java.ext.dirs`.

CC_JVM_OPTIONS

Establezca esta variable de entorno para especificar los argumentos JVM que se utilizan cuando se ejecuta un trabajo.

Cuando se especifica esta variable, ésta tiene prioridad sobre el valor de los argumentos JVM predeterminados para los conectores basados en Java. Por ejemplo, si establece **CC_JVM_OPTIONS** como `-Xmx512M`, el tamaño máximo de almacenamiento dinámico se establece en 512 MB cuando se crean instancias de JVM para el conector.

CC_JVM_OVERRIDE_OPTIONS

Establezca esta variable de entorno para alterar temporalmente las opciones de JVM del nodo conductor para poder evitar o solucionar un posible conflicto.

En el nodo conductor en un trabajo paralelo, los conectores Java se inicializan para la reconciliación de esquema. Por tanto, todos los conectores Java de un trabajo se inicializan en la misma JVM. En un trabajo individual, varias etapas podrían estar desarrolladas en Java. Cada una de estas etapas puede definir opciones de JVM, como vía de acceso de clases, propiedad del sistema, tamaño de almacenamiento dinámico, etc. Si dos etapas se ejecutan en la misma JVM física, las opciones de JVM pueden entrar en conflicto entre sí.

CC_IGNORE_TIME_LENGTH_AND_SCALE

Establezca esta variable de entorno para cambiar el comportamiento del conector en el lienzo paralelo.

Cuando esta variable de entorno se establece en 1, el conector que se ejecuta con el motor paralelo hace caso omiso de la longitud y escala especificadas para la columna de indicación de fecha y hora. Por ejemplo, cuando el valor de esta variable de entorno no está establecido y si la longitud de la columna de indicación de fecha y hora es 26 y la escala es 6, el conector en el lienzo paralelo considera que la indicación de fecha y hora tiene una resolución de microsegundos. Cuando el valor de esta variable de entorno se establece en 1, el conector en el lienzo paralelo no considera que la indicación de fecha y hora tiene una resolución de microsegundos a menos que la propiedad ampliada de microsegundos esté establecida, aunque la longitud de la columna de indicación de fecha y hora sea 26 y la escala sea 6.

CC_MSG_LEVEL

Establezca esta variable de entorno para especificar la gravedad mínima de los mensajes que el conector notifica en el archivo de registro.

Con el valor predeterminado de 3, los mensajes informativos y los mensajes de una gravedad superior se notifican en el archivo de registro.

La lista siguiente contiene los valores válidos:

- 1 - Rastreo
- 2 - Depuración
- 3 - Informativo
- 4 - Aviso
- 5 - Error
- 6 - Muy grave

CC_UNST_JAVA_HEAP

Establezca esta variable de entorno para controlar el tamaño del almacenamiento dinámico de Java que la etapa Unstructured Data puede utilizar.

Establezca la variable en un valor entero que represente el tamaño de almacenamiento dinámico de Java en MB. Por ejemplo, para establecer el tamaño del almacenamiento dinámico de Java en 512 MB, establezca **CC_UNST_JAVA_HEAP** en 512. El tamaño de almacenamiento dinámico de Java predeterminado es 256 MB.

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 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. Lectura de la sintaxis de la línea de mandatos

Esta documentación utiliza caracteres especiales para definir la sintaxis de la línea de mandatos.

Los siguientes caracteres especiales definen la sintaxis de la línea de mandatos:

- [] Identifica un argumento opcional. Se necesitan los argumentos que no están entre delimitadores.
- ... Indica que puede especificar varios valores para el argumento anterior.
- | Indica información que se excluye mutuamente. Puede utilizar el argumento a la izquierda del separador o el argumento a la derecha del separador. No puede utilizar los dos argumentos en un único uso del mandato.
- { } Delimita un conjunto de argumentos que se excluyen mutuamente cuando se necesita uno de los argumentos. Si los argumentos son opcionales, se escriben entre delimitadores ([]).

Nota:

- El número máximo de caracteres de un argumento es de 256.
- Escriba los valores de argumentos que tengan espacios incrustados entre comillas simples o dobles.

Por ejemplo:

```
wsetsrc[-S server] [-l label] [-n name] origen
```

El argumento *origen* es el único argumento necesario para el mandato **wsetsrc**. Los delimitadores de los otros argumentos indican que dichos argumentos son opcionales.

```
wlsac [formato -l | -f] [clave...] perfil
```

En este ejemplo, los argumentos de formato *-l* y *-f* se excluyen mutuamente y son opcionales. El argumento *perfil* es necesario. El argumento *clave* es opcional. La elipsis (...) que sigue al argumento *clave* indica que puede especificar varios nombres de clave.

```
wrb -import {rule_pack | rule_set}...
```

En este ejemplo, los argumentos *rule_pack* y *rule_set* se excluyen mutuamente, pero debe especificarse uno de ellos. Además, los puntos suspensivos (...) indican que puede especificar varios paquetes de reglas y conjuntos de reglas.

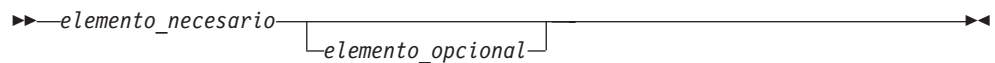
Apéndice C. Cómo leer diagramas de sintaxis

Las reglas siguientes se aplican a los diagramas de sintaxis que se utilizan en esta documentación:

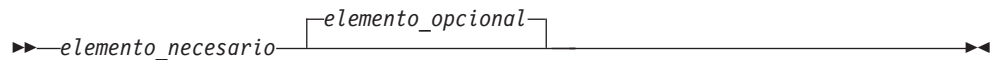
- Lea los diagramas de sintaxis de izquierda a derecha y de arriba abajo, siguiendo el recorrido de la línea. Se utilizan los convenios siguientes:
 - El símbolo >>--- indica el inicio de un diagrama de sintaxis.
 - El símbolo ---> indica que el diagrama de sintaxis continúa en la línea siguiente.
 - El símbolo >--- indica que el diagrama de sintaxis viene de la línea anterior.
 - El símbolo --->< indica el final de un diagrama de sintaxis.
- Los elementos necesarios aparecen en la línea horizontal (la línea principal).



- Los elementos opcionales aparecen debajo de la línea principal.

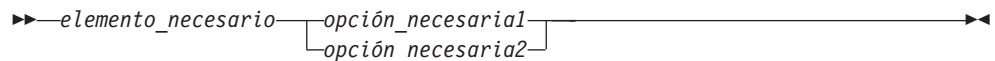


Si aparece un elemento opcional sobre la línea principal, dicho elemento no tendrá efecto sobre el elemento de sintaxis y sólo se utilizará para facilitar la lectura.

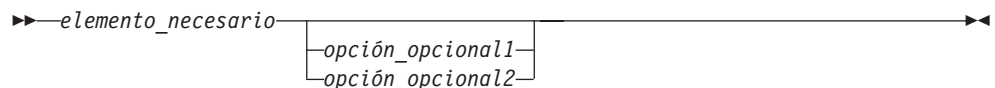


- Si se puede elegir entre dos o más elementos, éstos aparecerán apilados verticalmente.

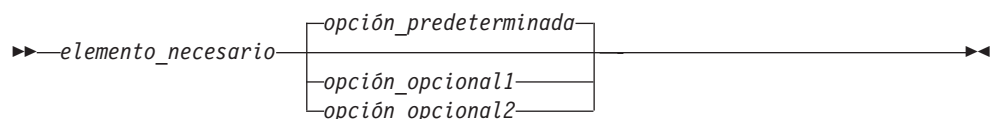
Si se debe elegir uno de los elementos, un elemento de la pila aparece en la línea principal.



Si la elección de uno de los elementos es opcional, toda la pila aparecerá por debajo de la línea principal.



Si uno de los elementos es el predeterminado, aparecerá por encima de la línea principal y las opciones restantes se mostrarán por debajo.



- Una flecha que vuelve hacia la izquierda, sobre la línea principal, indica un elemento que se puede repetir.

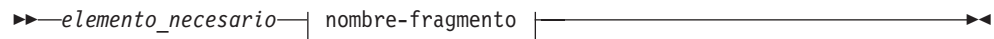


Si la flecha de repetición contiene una coma, los elementos repetidos se deben separar mediante una coma.

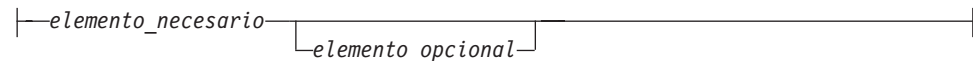


Una flecha de repetición sobre una pila indica que los elementos de la pila se pueden repetir.

- A veces, un diagrama se debe dividir en fragmentos. El fragmento de sintaxis se muestra por separado del diagrama de sintaxis principal, pero el contenido del fragmento se debe leer como si formara parte de la línea principal del diagrama.



Nombre-fragmento:



- Las palabras clave, y sus abreviaturas mínimas si las hay, aparecen en mayúsculas. Se deben escribir exactamente tal como se muestran.
- Las variables aparecen en letras minúsculas en cursiva (por ejemplo, **nombre-columna**). Representan nombres o valores proporcionados por el usuario.
- Separe las palabras clave y los parámetros con un espacio como mínimo si no se muestra ningún signo de puntuación en el diagrama.
- Entre los signos de puntuación, paréntesis, operadores aritméticos y otros símbolos exactamente como se muestran en el diagrama.
- Las notas a pie de página se muestran mediante un número entre paréntesis, por ejemplo (1).

Apéndice D. 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 26. 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 E. 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 F. Cómo aportar comentarios sobre la documentación del producto

Puede aportar valiosos comentarios en relación a la documentación de IBM.

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

- Para proporcionar un comentario acerca de un tema del IBM Knowledge Center que está alojado en el sitio web de IBM, inicie la sesión y pulse el botón **Añadir comentario** en la parte inferior del tema. Los comentarios enviados de esta manera serán visibles para todos los usuarios.
- Para enviar un comentario acerca de un tema del IBM Knowledge Center a IBM y que ningún otro usuario pueda ver, inicie la sesión y pulse en el enlace **Comentarios** en la parte inferior del 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 comments@us.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. 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 licenciarios 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 27. 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 27. 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

accesibilidad de los productos
 accesibilidad 55
avisos legales 67

C

caracteres especiales
 sintaxis de la línea de mandatos 57
CC_UNST_JAVA_HEAP, variable de entorno 53
correlación
 tipos de datos 39, 46

D

documentación del producto
 acceder 63

E

Ejemplo 2: Creación del trabajo 13
Ejemplo 3: Creación del trabajo 16
Ejemplo 4: Creación del trabajo 23
Ejemplo 5: Creación del trabajo 27, 30
Ejemplo 7: Creación del trabajo
 grabar datos en archivos existentes de Microsoft Excel 35
extracción de datos
 DataStage 2, 3
Extracción de datos
 rango de datos 2
Extracción de datos de un rango en un archivo de Microsoft Excel
 Creación del trabajo 10

M

mandatos
 sintaxis 57
marcas registradas
 lista de 67

O

Orígenes de datos
 DataStage 2

R

resolución de problemas
 Unstructured Data, etapa 49

S

servicios de software
 contactar 61

sintaxis
 línea de mandatos 57
sintaxis de línea de mandatos
 convenios 57
sitios web
 no IBM 59
soporte
 cliente 61
soporte al cliente
 contactar 61

T

tipos de datos
 cargar datos 39, 46
 DataStage 39, 46
 grabación de datos 39, 46
 Unstructured Data, etapa 39, 46

U

Unstructured Data, etapa 13, 16, 23, 27, 30, 35
 condiciones de terminación anómala de trabajo 46
 Configuración de la etapa Sequential File
 grabar datos en archivos existentes de Microsoft Excel 35
 configuración de la etapa Unstructured Data 18, 21
 configurar
 modificar un archivo existente de Microsoft Excel 33
 configurar el origen de datos 6
 Consideración sobre el fin de oleada 22
 consulta
 correlaciones soportadas 39
 definir trabajos 1, 33
 diseñar trabajos 1
 Diseño de trabajos 5, 32
 ejemplo 1
 extracción de datos de un rango 10
 grabación de datos 23
 grabar datos en archivos existentes de Microsoft Excel 35
 Ejemplo 1: Configuración de la etapa Sequential File 11, 14, 17
 Ejemplo 1: Configuración de la etapa Unstructured Data 10
 ejemplo 2 27
 extracción de datos de varias hojas de cálculo de Microsoft Excel 12
 Ejemplo 2: Configuración de la etapa Unstructured Data 13
 ejemplo 3
 extracción de datos de varios rangos 16

Unstructured Data, etapa (*continuación*)
 ejemplo 3 (*continuación*)
 grabar datos en varios archivos de Microsoft Excel 30
 Ejemplo 3: Configuración de la etapa Unstructured Data 16
 Ejemplo 4: Configuración de la etapa Sequential File 24
 Ejemplo 4: Configuración de la etapa Unstructured Data 24, 27
 Ejemplo 5: Configuración de etapas Sequential File 28
 Ejemplo 6: Configuración de etapas Sequential File 31
 Ejemplo 6: Configuración de la etapa Unstructured Data 30
 Ejemplo 7: Configuración de la etapa Unstructured Data
 grabar datos en un archivo existente de Microsoft Excel 36
 Ejemplos: Extracción de datos de archivos de Microsoft Excel 10
 Ejemplos: Grabación de datos en archivos de Microsoft Excel 23
 especificar definición de columna 20
 extraer valor de una celda o propiedades personalizadas 8
 grabación de datos 27
 grabar datos en archivos existentes de Microsoft Excel 32
 grabar datos en un archivo de Microsoft Excel 18, 20
 Diseño de trabajos 18
 hoja existente de Microsoft Excel 32
 instalar y configurar 2
 manejo de errores 8, 9
 manejo de filas nulas 8
 modificar definición de columna 7
 opciones para leer datos de archivos de Microsoft Excel 8
 parámetros de trabajo en la ventana Configuración 7, 20, 34
 propagación de columnas en tiempo de ejecución 9
 varias hojas de cálculo 27
 visión general 1
 visualización del resultado del trabajo 11, 14, 17, 25, 29, 31
 grabar datos en un archivo existente de Microsoft Excel 37

V

variables de entorno
 Unstructured Data, etapa 53



Impreso en España

SC43-1246-00

