

Caso práctico: Almacén de datos para el análisis del impacto ambiental y consumo energético derivados de la actividad económica

PRA2- Carga de datos

Presentación

A partir de la solución oficial de la primera práctica (PRA1), el estudiante debe diseñar, implementar y ejecutar los procesos de extracción, transformación y carga de los datos de las fuentes de datos proporcionadas.

Esta actividad tiene como objetivo identificar y desarrollar los procesos de carga del almacén de datos y su ejecución efectiva. Para su realización dispone de una serie de recursos de aprendizaje asociados que están disponibles en el aula, junto al enunciado de esta actividad. Se recomienda consultar el documento *Guía Buenas Prácticas ETL.pdf*, así como los vídeos proporcionados.

Se facilitarán algunos archivos de código fuente de Spoon para que los incorporéis en el repositorio de archivos que crearéis al realizar esta actividad con PDI. Al inicio de la guía de muestra que encontraréis en este documento, se indican los ficheros que os vienen dados y cómo utilizarlos.

Descripción

El alcance de esta segunda parte del caso práctico se concreta en los siguientes apartados:

- Crear la matriz de dimensiones y métricas.
- Identificar los procesos de extracción, transformación y carga de datos (ETL) hacia el almacén de datos.
- Diseñar y desarrollar los procesos ETL mediante las herramientas de diseño proporcionadas (*Pentaho Data Integration - Spoon*)
- Implementar con trabajos (*jobs*) los procesos ETL, para realizar la carga efectiva del almacén de datos.

Por tanto, el documento de la solución de la PRA2 debe incluir los siguientes aspectos:

- Descripción de todas las acciones realizadas.
- Captura de pantalla con el flujo completo del ETL / JOB.
- Capturas de pantalla que demuestren: 1) las partes más significativas del ETL, sus características y su correspondiente explicación, 2) la correcta ejecución del ETL y el tiempo de ejecución y 3) la correcta carga de los datos en la base de datos.
- Captura de pantalla del resultado de todas las ejecuciones (pantalla «Execution Results»). Se debe incluir la pestaña «Step Metrics» o «Job Metrics», correspondiente a cada ETL/JOB.

Además de entregar el documento con la solución de la PRA2, también se tendrá en consideración la implementación sobre la máquina virtual proporcionada para realizar las actividades de la asignatura.

Guía de muestra

Con el fin de ayudar a alcanzar los objetivos planteados de la PRA2, se proporciona esta guía de muestra que servirá de ejemplo de cómo realizar alguna de las tareas anteriormente descritas. Es decir, el diseño y el desarrollo de los procesos ETL y la carga efectiva al almacén de datos.

Esta guía va a seguir las pautas recomendadas en el documento *Guía Buenas Prácticas ETL.pdf* que tenéis disponible en el aula junto a esta actividad. Es importante no tomar esta guía como un tutorial o guía a seguir paso a paso para la resolución de la PRA. Únicamente se facilita a modo de ejemplo de resolución de un caso similar, por tanto no idéntico.

Antes de entrar en detalle, se indican los archivos de código fuente de Spoon (.ktr) que tenéis a vuestra disposición por bloque:

Bloque IN: IN_INVESTMENTS.ktr

Bloque TR: TR_DIM_Country.ktr

TR_DIM_EconomicActivitySector.ktr

TR_DIM_TypeEquipmentInstallation.ktr

Así como el siguiente *job* (.kjb): JOB_TR_FACT.kjb

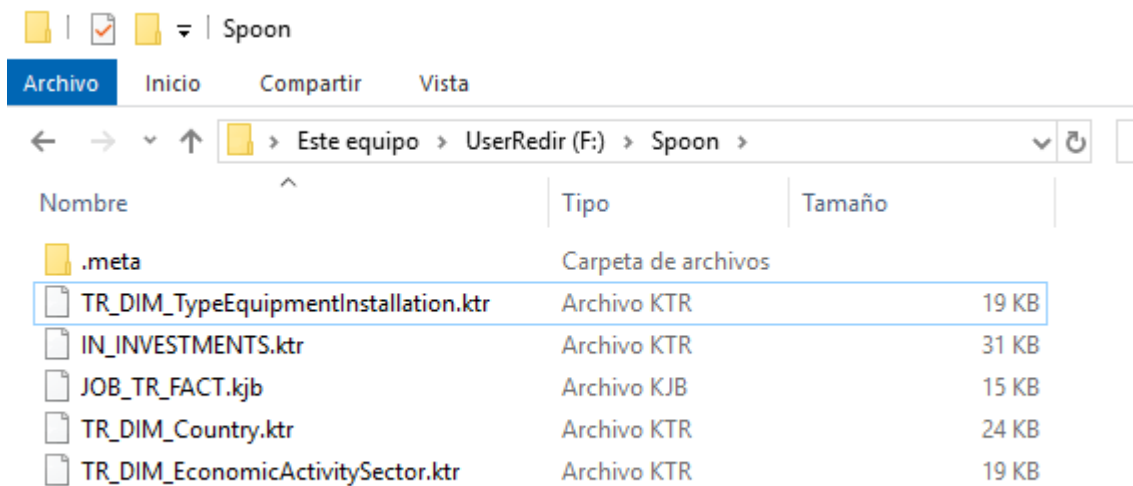
Estos archivos pueden ser incorporados al repositorio de Spoon que deberéis crear para esta PRA2.

Para poder utilizar estos archivos en vuestras prácticas, será necesario adaptar las cadenas de conexión, sustituyendo *loginuoc* por vuestro propio *login* del campus de la UOC.

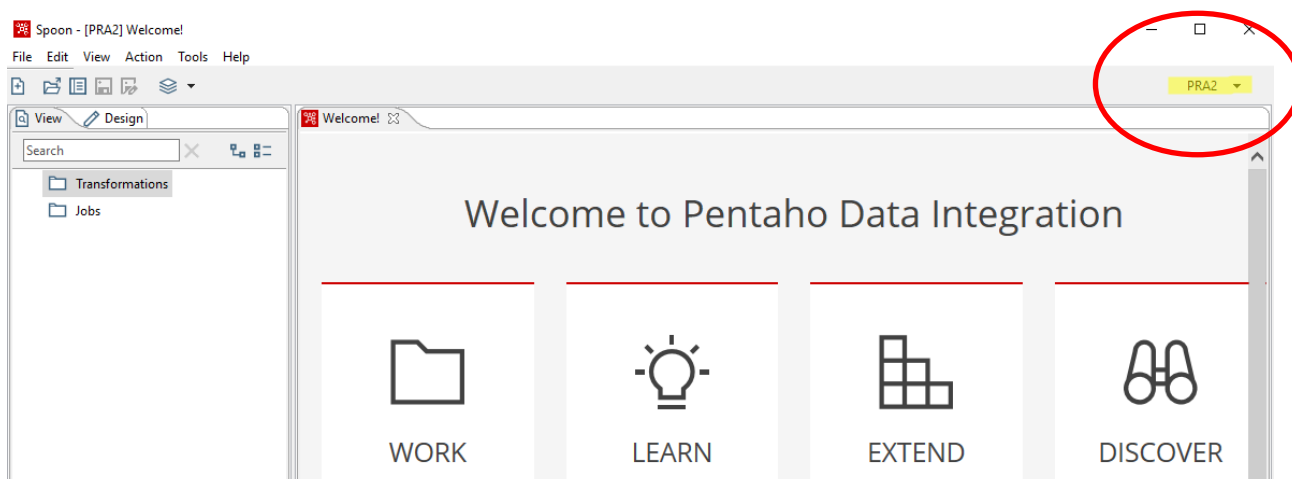
Además, dado que se dispone de varios servidores, también será necesario modificar el nombre del servidor: UCS1R1UOCSQLXX, sustituyendo XX por el número de servidor asignado a vuestra aula.

Indicaciones para incorporar los archivos facilitados al proyecto de Spoon:

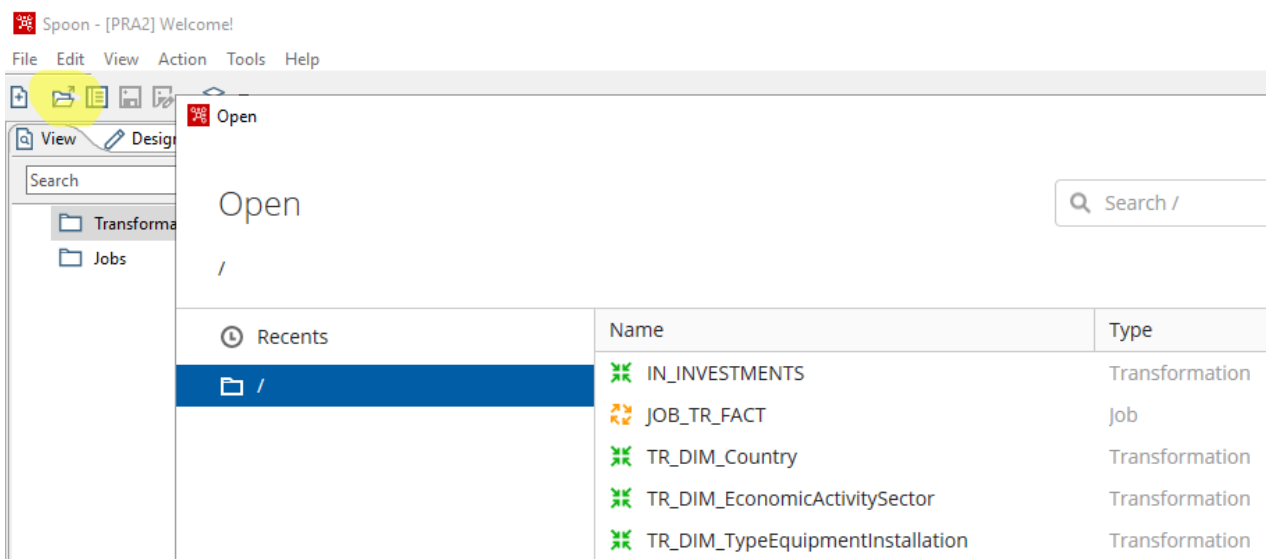
1. Crear un repositorio de tipo «File». En este ejemplo se ha creado en la ruta F:\Spoon con el nombre PRA2
2. Copiar los archivos facilitados en la carpeta del repositorio.



3. Abrir Spoon y conectar con el repositorio.

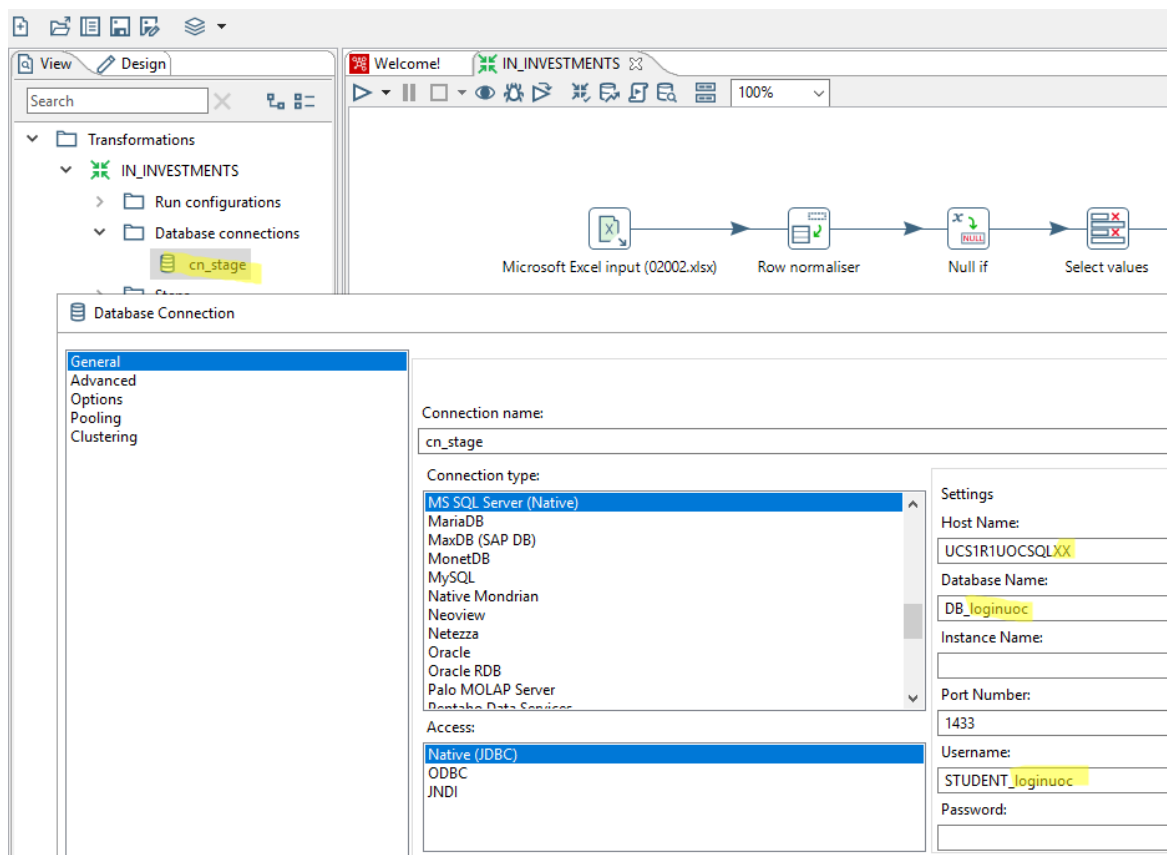


4. Abrir un nuevo fichero. Al habernos conectado previamente al repositorio aparecerán los ficheros copiados.



5. Abrir cada uno de los archivos para modificar los parámetros indicados de la conexión:

- Nombre del servidor. Sustituir UCS1R1UOCSQLXX por el número de servidor asignado a cada aula.
- Nombre de la base de datos. Sustituir DB_loginuoc por el login del campus de la UOC de cada estudiante
- Nombre de usuario. Sustituir STUDENT_loginuoc por el login del campus de la UOC de cada estudiante. Recordad indicar también la contraseña correctamente.



1. Identificación de los procesos ETL

A la hora de diseñar los procesos de carga de una base de datos analítica no hay una única estrategia. Es habitual estructurar los procesos ETL sobre la base de las entidades de datos que se deben actualizar, ya que existen diferencias conceptuales en la actualización de una tabla de dimensión con respecto a la de una tabla de hechos. La división del proceso de carga inicial en diferentes bloques de actualización facilitará el diseño, un orden de ejecución y la gestión de las dependencias. Cada uno de estos bloques de actualización se dividirá en las correspondientes etapas de extracción, transformación y carga.

Se identifican los dos bloques siguientes:

- **Bloque IN:** procesos de carga de los datos desde las fuentes a las tablas intermedias en el área de maniobras (*staging area*). Estos procesos se distinguen por el prefijo «IN» en el nombre.
- **Bloque TR:** procesos de transformación para cargar los datos desde las tablas intermedias hasta nuestro almacén de datos, según el modelo multidimensional diseñado. Así pues, los procesos ETL de transformación para cargar las dimensiones (TR_DIM_) son diferentes de aquellos que se realizan para cargar las tablas de hechos (TR_FACT_). Estos procesos se distinguen con el prefijo «TR_» en el nombre.

A continuación, se muestra un ejemplo de cómo identificar **algunos de los procesos** que forman parte de cada uno de los bloques de actualización:

Bloque IN (de las fuentes a las tablas intermedias)

Nombre del ETL	Descripción	Orígenes de los datos	Tabla de destino (<i>stage</i>)
IN_OBJECTIVES	Carga de los datos correspondientes a los Objetivos de Desarrollo Sostenible y su relación con los ámbitos medioambientales	ODS.xlsx (dos pestañas)	STG_Objectives STG_ObjectivesAreas
IN_INVESTMENTS	Carga de datos correspondientes a la evolución de la inversión en protección ambiental por tipo de equipo e instalación, ámbito medioambiental y sector de actividad económica.	02002.xlsx	STG_Investments

Bloque TR (de las tablas intermedias a nuestro almacén)

El bloque TR de procesos ETL para poblar el modelo multidimensional del almacén tiene dos partes diferenciadas. Por un lado, los procesos de carga y transformación de las dimensiones y, por otro, los de las tablas de hechos. El orden de ejecución es importante para que la carga de datos sea la correcta. Las dimensiones se cargarán primero y, después, las tablas de hechos para que no haya errores durante la carga.

A continuación, se muestran algunos de los procesos del bloque TR de carga y transformación de las dimensiones:

Nombre del ETL	Descripción	Tabla de origen	Tabla de destino (dimensión)
TR_DIM_Date	Carga y transformación de la dimensión temporal (fecha en la que se realiza la medición).	SQL	DIM_Date
TR_DIM_SDG	Carga y transformación de la dimensión de Objetivos de Desarrollo Sostenible.	STG_Objectives	DIM_SDG
TR_DIM_Country	Carga de la dimensión de países donde se localiza la medición del balance energético.	STG_Countries	Dim_Country
TR_DIM_EconomicActivitySector	Carga y transformación de la dimensión con datos del sector de la actividad económica al que se clasifica la medición.	STG_Investments	Dim_EconomicActivitySector
TR_DIM_TypeEquipmentInstallation	Carga de la dimensión con los tipos de equipamientos o instalación de la medición.	STG_Investments	Dim_TypeEquipmentInstallation

Por otra, el proceso del bloque de carga y transformación de la tabla de hechos es el siguiente:

Nombre del ETL	Descripción	Tabla de origen
TR_FACT_EnvironmentalMeasurements	Carga y transformación de la tabla de hechos «FACT_EnvironmentalMeasurements».	STG_Investments STG_ProtectedAreas STG_UrbanWastes

En este punto, el estudiante deberá completar la identificación de todos los procesos de cada uno de los bloques (IN y TR) que desarrollará para cargar las dimensiones y las tablas de hechos del modelo multidimensional del almacén de datos.

2. Diseño y desarrollo de los procesos ETL

En este apartado, se deben diseñar los procesos de carga identificados en el punto anterior con la herramienta de diseño proporcionada, *Pentaho Data Integration* (PDI).

Creación de tablas

El primer paso para la implementación de los procesos ETL consiste en la creación de las tablas. Esto se llevará a cabo una única vez, mediante *scripts*, sobre la base de datos proporcionada (en nuestro caso: SQL Server). Se deberán crear las tablas intermedias y las tablas del modelo dimensional de la solución oficial, es decir, las dimensiones y las tablas de hechos. Para hacerlo, se utilizarán los *scripts* facilitados junto a la solución de la PRA1 (*Script_Diseño_Fisico.sql*). En caso de modificar los *scripts* facilitados, se deberá explicar y argumentar la estrategia de uso y definición de restricciones, claves primarias, claves foráneas, etc. en el modelo multidimensional.

Una vez implementado el modelo físico del almacén, el siguiente paso es el diseño de los procesos ETL de cada uno de los bloques (IN y TR). Estos procesos permitirán poblar las tablas del área intermedia (*staging area*), las de dimensiones y las de hechos del almacén de datos diseñado.

Bloque IN

Transformación de «IN_OBJECTIVES»

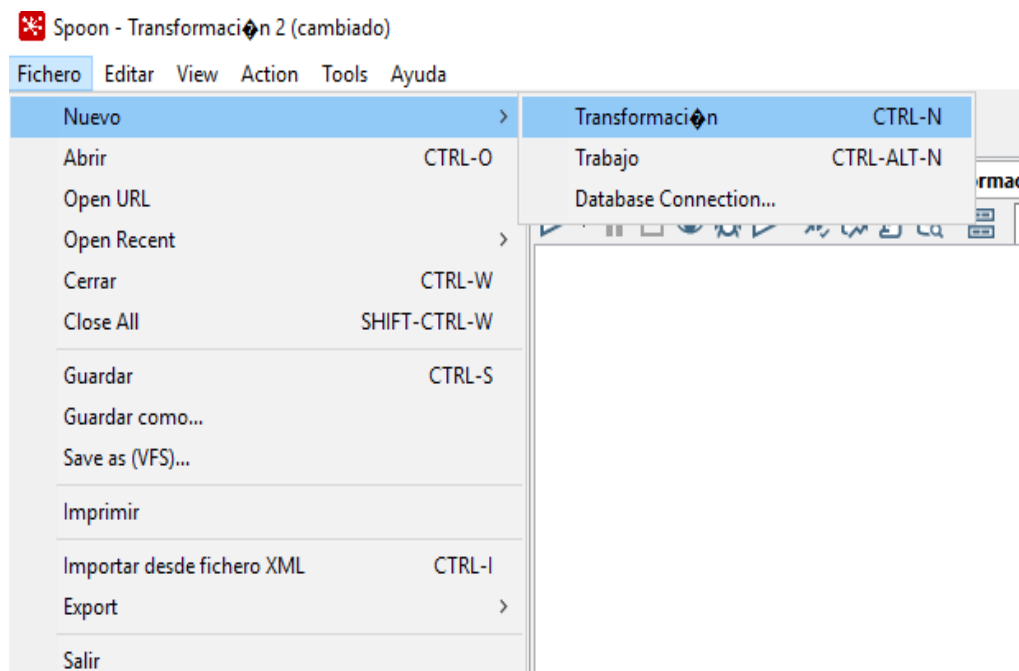
A continuación, se describe parte del desarrollo de la transformación de «IN_OBJECTIVES» (identificada en el primer punto de la guía) mediante Spoon. El objetivo es cargar uno de los orígenes de los datos identificados, «ODS.xlsx», en las tablas «STG_Objectives» y «STG_ObjectivesAreas» del área intermedia (*staging area*). La tabla intermedia tendrá que haber sido creada con anterioridad en la base de datos analítica. En esta guía de muestra únicamente se cargará la tabla «STG_OBJECTIVES».

```
USE [DB_XXX]
GO
```

```
CREATE TABLE [dbo].[STG_Objectives](
    [Objetivo] [int] NOT NULL,
    [Nombre] [varchar](100) NULL,
    [Descripcion] [varchar](512) NULL
) ON [PRIMARY]
GO
) ON [PRIMARY]
GO
```

Para este caso práctico se han utilizado fuentes externas (no operacionales) que se emplean para descubrir el conocimiento mediante el análisis de los datos. Es muy habitual manipular los ficheros realizando manualmente una serie de acciones de preparación antes de su procesamiento (preprocesado). En el ejemplo tratado en esta Guía de Muestra no será necesario realizar ningún preprocesado de los datos.

La transformación «IN_OBJECTIVES» contiene cuatro pasos: lectura del fichero XLSX, operaciones con cadenas, ordenación de las filas y carga a la tabla intermedia «STG_Objectives».



Este es el primer paso de la transformación, en el que se realizará la extracción o entrada del fichero XLSX. Para ello se utilizará un componente de tipo «Entrada Excel».



Microsoft Excel input (ODS.xlsx)

Aquí se indicará el fichero desde donde se extraerán los datos. Para ello se utilizará la variable de entorno «DIR_IN» y se indicará el tipo de motor utilizado, si es para ficheros de tipo XLS o XLSX.

Entrada Excel

Nombre paso: Objetivos Desarrollo Sostenible (ODS.xlsx)

Ficheros | Hojas | Contenido | Manejador de Error | Campos | Additional output fields

Spread sheet type (engine): Excel 2007 XLSX (Apache POI)

Fichero o directorio: [Añadir] [Examinar...]

Expresión Regular:

Exclude Regular Expression:

Password:

Ficheros seleccionados:

#	Fichero/Directorio	Comodón	Exclude wildcard	Requerido	Include subfolde
1	\$(DIR_IN)\ODS.xlsx			N	N

[Eliminar] [Editar]

Aceptar nombres de fichero de pasos anteriores

Aceptar nombres de fichero de pasos: ☐

Paso desde el que se obtienen los:

Campo de entrada a utilizar como:

[Mostrar Fichero(s)...]

[?] Help [Vale] [Previsualizar filas] [Cancelar]

A continuación, se indicarán las hojas del fichero de origen que se deberán tener en cuenta y, en caso de ser necesario, la fila y columna donde empezar y finalizar la lectura de datos:

Entrada Excel

Nombre paso: Objetivos Desarrollo Sostenible (ODS.xlsx)

Ficheros | Hojas | Contenido | Manejador de Error | Campos | Additional output fields

Lista de hojas a leer

#	Nombre hoja	Fila inicial	Columna inicial
1	ODS		

[Obtener hoja(s)...]

[?] Help [Vale] [Previsualizar filas] [Cancelar]

El siguiente paso consiste en recuperar los campos a tratar con el botón «Traer campos» y completar la definición de estos. Hay que especificar, donde se considere necesario, la precisión y la longitud de los campos o indicar el formato de la fecha para los campos «date».

Si en los campos del tipo «Number» se indica una longitud y una precisión con el valor -1, estos se tratarán como un dato numérico de coma flotante (*float*).

Entrada Excel

Nombre paso:

Ficheros | Hojas | Contenido | Manejador de Error | Campos | Additional output fields

#	Nombre	Tipo	Longitud	Precisión	Tipo de poda	Repetir	Formato	Moneda	Decimal	Ag
1	Objetivo	Number			none	N				
2	Nombre	String			none	N				
3	Descripción	String			none	N				

Obtener campos de cabecera...

Help Vale Previsualizar filas Cancelar

Para realizar una visualización previa de los datos que se cargarán, se utiliza el botón «Previsualizar filas».

Examine preview data

Rows of step: Objetivos Desarrollo Sostenible (ODS.xlsx) (17 rows)

#	Objetivo	Nombre	Descripción
1	1,0	Fin de la pobreza	Para lograr este Objetivo de acabar con la pobreza, el crecimiento económico debe ser inclusivo, con el fin de crear empleo.
2	2,0	Hambre cero	El sector alimentario y el sector agrícola ofrecen soluciones claves para el desarrollo y son vitales para la eliminación del hambre.
3	3,0	Salud y bienestar	Para lograr los Objetivos de Desarrollo Sostenible, es fundamental garantizar una vida saludable y promover el bienestar universal.
4	4,0	Educación de calidad	La educación es la base para mejorar nuestra vida y el desarrollo sostenible.
5	5,0	Igualdad de género	La igualdad entre los géneros no es solo un derecho humano fundamental, sino la base necesaria para conseguir un mundo más justo.
6	6,0	Agua limpia y saneamiento	El agua libre de impurezas y accesible para todos es parte esencial del mundo en que queremos vivir.
7	7,0	Energía asequible y no contaminante	La energía es central para casi todos los grandes desafíos y oportunidades a los que se enfrenta el mundo en la actualidad.
8	8,0	Trabajo decente y crecimiento económico	Debemos reflexionar sobre este progreso lento y desigual, y revisar nuestras políticas económicas y sociales destinadas a crear empleos.
9	9,0	Industria, innovación e infraestructuras	Las inversiones en infraestructura son fundamentales para lograr un desarrollo sostenible.
10	10,0	Reducción de las desigualdades	Reducir la desigualdad en y entre los países.
11	11,0	Ciudades y comunidades sostenibles	Las inversiones en infraestructura son cruciales para lograr el desarrollo sostenible.
12	12,0	Producción y consumo responsables	El objetivo del consumo y la producción sostenibles es hacer más y mejores cosas con menos recursos.
13	13,0	Acción por el clima	El cambio climático es un reto global que no respeta las fronteras nacionales.
14	14,0	Vida submarina	Conservar y utilizar en forma sostenible los océanos, los mares y los recursos marinos para el desarrollo sostenible.
15	15,0	Vida de ecosistemas terrestres	Gestionar sosteniblemente los bosques, luchar contra la desertificación, detener e invertir la degradación de las tierras y del suelo.
16	16,0	Paz, justicia e instituciones sólidas	Acceso universal a la justicia y la construcción de instituciones responsables y eficaces a todos los niveles.
17	17,0	Alianzas para lograr los objetivos	Revitalizar la Alianza Mundial para el Desarrollo Sostenible.

Cerrar Show Log

El siguiente paso de la transformación es asegurar la calidad de los datos mediante la normalización de los valores de los campos de tipo «String» que se consideren necesarios, en este ejemplo se aplicará al campo «Nombre». Para ello, con el componente «String operations», se convertirán los datos de las fuentes de origen en mayúsculas y se eliminarán los espacios en blanco que pudiera haber al inicio y al final de la cadena de caracteres.



String operations

Step name: String operations

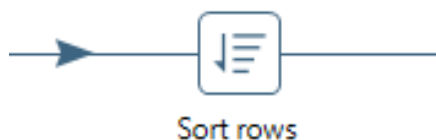
The fields to process:

#	In stream field	Out stream field	Trim type	Lower/Upper	Padding	Pad char	Pad Length	InitCap	Escape	Digits	Remove Special character
1	Nombre		none	upper	none			N	None	none	none
2	Descripción		none	none	none						
3											
4											

< >

Help Vale Get fields Cancelar

A continuación, hay que ordenar de manera ascendente los campos. Para ello, se utilizará el componente «Ordenar filas» de las transformaciones que tenemos disponibles.



Ordenar filas

Nombre paso: Ordenar filas

Directorio ordenación: %%java.io.tmpdir%% Examinar...

Prefijo para ficheros temporales: out

Tamaño de ordenación (filas en memoria): 1000000

Free memory threshold (in %):

Comprimir ficheros temporales? ☐

Only pass unique rows? (verifies keys only) ☐

Campos:

#	Nombre Campo	Ascendente	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presorted?
1	Objetivo	S	N	N	0	N

Help Vale Cancelar Traer Campos

Por último, se cargarán los datos en la tabla intermedia del *stage*, utilizando el paso «Salida Tabla» de la carpeta «Salida». Este paso necesita especificar la conexión de la base de datos correspondiente.



Table output (STG_Objectives)

El paso de cargar los datos a la tabla intermedia del *stage* se configurará como se indica en el menú principal del componente, que veis a continuación:

Para dejar la transformación preparada para posibles reprocesos, es necesario realizar un borrado previo de las tablas. Para esto, una posible opción es activar el *check* «Vaciar tabla».

A continuación, será necesario cargar los campos y establecer la correspondencia o mapeo, entre los campos del flujo de datos y los campos de la tabla física.

El proceso de la transformación completa es el siguiente:



Y el resultado de la ejecución:

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data											
#	Nombre paso	Numero Copia	Leído	Escrito	Entrada	Salida	Actualizado	Rejected	Errores	Activo	Tiempo
1	Objetivos Desarrollo Sostenible (ODS.xlsx)	0	0	17	17	0	0	0	0	Finalizado	0.1s
2	String operations	0	17	17	0	0	0	0	0	Finalizado	0.1s
3	Ordenar filas	0	17	17	0	0	0	0	0	Finalizado	0.2s
4	Salida Tabla (STG_OBJETIVES)	0	17	17	0	17	0	0	0	Finalizado	0.2s

Como se puede observar en las métricas, se cargan los 17 registros del fichero de entrada.

Bloque TR_DIM

Transformación de «TR_DIM_SDG»

Como en el caso anterior, el primer paso será la creación de la tabla física mediante el *script* facilitado.

```
USE [DB_XXX]
GO

CREATE TABLE [dbo].[DIM_SDG](
    [pk_sdg] [int] NOT NULL,
    [sdg_name] [varchar](50) NULL,
    [sdg_description] [varchar](500) NULL,
    CONSTRAINT [PK_DIM_SDG] PRIMARY KEY CLUSTERED
(
    [pk_sdg] ASC
)) ON [PRIMARY]
GO
```

A continuación, se trataría de implementar el proceso ETL de carga para esta dimensión. En este caso se mostrará una sencilla carga de los datos a la tabla del modelo dimensional. **Recordad que este es un ejemplo parcial que debe ser completado.**

Primero se extraen los datos de la tabla STG_Objectives desde el *staging area*

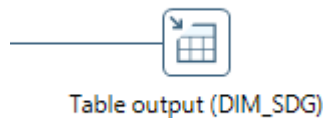


Table input (STG_Objectives)

En este paso se utiliza la conexión al *staging area* (cn_stage) para acceder a la tabla STG_Objectives. En este punto se podría utilizar una consulta SQL más compleja, en caso de ser necesario para acceder a los datos.

Table input	
Step name	Table input (STG_Objectives)
Connection	cn_stage
SQL <pre>SELECT Objetivo , Nombre , Descripcion FROM dbo.STG_Objectives</pre>	

Y posteriormente se cargan los datos a la tabla de dimensión DIM_SDG del modelo dimensional en el DataMart.



Se utiliza la conexión al DataMart (cn_dw) para acceder a la tabla DIM_SDG

Table output	
Step name	Table output (DIM_SDG)
Connection	cn_dw Edit... New... Wizard...
Target schema	dbo Browse...
Target table	DIM_SDG Browse...
Commit size	1000
Truncate table	<input type="checkbox"/>
Ignore insert errors	<input type="checkbox"/>
Specify database fields	<input checked="" type="checkbox"/>

Finalmente se mapean los campos físicos de la tabla con los datos del flujo de datos del ETL.

Main options		Database fields	
Fields to insert:			
#	Table field	Stream field	
1	pk_sdg	Objetivo	
2	sdg_name	Nombre	
3	sdg_description	Descripcion	
			Get fields Enter field mapping

La transformación de ejemplo (**solución parcial**) quedaría de la siguiente manera:

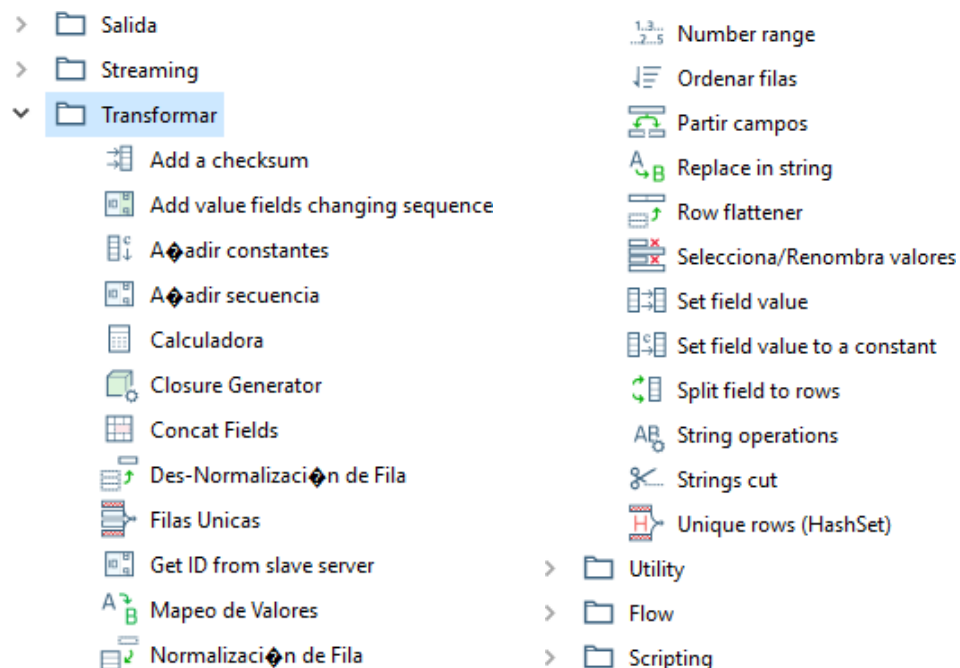


Para completar este ejemplo, sería necesario, por ejemplo, tratar el diseño de los registros NA (*Not Available*)

En la pestaña «Step Metrics» se puede ver la carga de los 17 registros.

Execution Results											
Logging Execution History Step Metrics Performance Graph Metrics Preview data											
#	Nombre paso	Numero Copia	Leído	Escrito	Entrada	Salida	Actualizado	Rejected	Errores	Activo	Tiempo
1	Entrada Tabla (STG_OBJETIVOS)	0	0	17	17	0	0	0	0	Finalizado	0.0s
2	Salida Tabla (DIM_SDG)	0	17	17	0	17	0	0	0	Finalizado	0.1s

Para la entrega de la PRA2, el estudiante deberá diseñar todos los procesos ETL de cada uno de los bloques (IN y TR). En este ejemplo se ha mostrado un caso básico de carga de datos, pero, según el formato de origen de los datos y de la calidad de estos, tal vez sea necesario utilizar otras transformaciones. Spoon dispone de una gran cantidad de componentes a los que se puede acceder desde el menú lateral. Están organizados por categorías, como se puede observar en el siguiente ejemplo:



3. Implementación de los trabajos con procesos ETL

Los bloques de procesos ETL implementados que hay que tener en cuenta son los siguientes:

- **Bloque IN_:** procesos ETL de transformación y carga al área intermedia.
- **Bloque TR_DIM_:** procesos ETL de transformación y carga de dimensiones.
- **Bloque TR_FACT_:** procesos ETL de transformación y carga de hechos.

En este punto, para realizar la carga efectiva de los datos, el estudiante debe diseñar mediante PDI los trabajos (*jobs*) que permitan la ejecución de todos los procesos ETL incluidos en cada bloque. En este apartado se deben incluir también las volumetrías obtenidas (número de registros cargados en cada tabla).

En la siguiente imagen se puede observar la carga efectiva de la tabla del modelo dimensional (17 registros cargados) mediante una consulta realizada en la consola SSMS (SQL Server Management Studio).

```
select * from [dbo].[DIM_SDG]
```

pk_sdg	sdg_name	sdg_description
1	FIN DE LA POBREZA	Para lograr este Objetivo de acabar con la pobreza, el crecimiento económico debe ser inclusivo, con...
2	HAMBRE CERO	El sector alimentario y el sector agrícola ofrecen soluciones claves para el desarrollo y son vitales par...
3	SALUD Y BIENESTAR	Para lograr los Objetivos de Desarrollo Sostenible, es fundamental garantizar una vida saludable y pro...
4	EDUCACIÓN DE CALIDAD	La educación es la base para mejorar nuestra vida y el desarrollo sostenible.
5	IGUALDAD DE GÉNERO	La igualdad entre los géneros no es solo un derecho humano fundamental, sino la base necesaria par...
6	AGUA LIMPIA Y SANEAMIENTO	El agua libre de impurezas y accesible para todos es parte esencial del mundo en que queremos vivir.
7	ENERGÍA ASEQUIBLE Y NO CONTAMINANTE	La energía es central para casi todos los grandes desafíos y oportunidades a los que se enfrenta el m...
8	TRABAJO DECENTE Y CRECIMIENTO ECONÓMICO	Debemos reflexionar sobre este progreso lento y desigual, y revisar nuestras políticas económicas y s...
9	INDUSTRIA, INNOVACIÓN E INFRAESTRUCTURAS	Las inversiones en infraestructura son fundamentales para lograr un desarrollo sostenible.
10	REDUCCIÓN DE LAS DESIGUALDADES	Reducir la desigualdad en y entre los países.
11	CIUDADES Y COMUNIDADES SOSTENIBLES	Las inversiones en infraestructura son cruciales para lograr el desarrollo sostenible.
12	PRODUCCIÓN Y CONSUMO RESPONSABLES	El objetivo del consumo y la producción sostenibles es hacer más y mejores cosas con menos recursos.
13	ACCIÓN POR EL CLIMA	El cambio climático es un reto global que no respeta las fronteras nacionales.
14	VIDA SUBMARINA	Conservar y utilizar en forma sostenible los océanos, los mares y los recursos marinos para el desarroll...
15	VIDA DE ECOSISTEMAS TERRESTRES	Gestionar sosteniblemente los bosques, luchar contra la desertificación, detener e invertir la degradaci...
16	PAZ, JUSTICIA E INSTITUCIONES SÓLIDAS	Acceso universal a la justicia y la construcción de instituciones responsables y eficaces a todos los ni...
17	ALIANZAS PARA LOGRAR LOS OBJETIVOS	Revitalizar la Alianza Mundial para el Desarrollo Sostenible.

Recursos externos de apoyo

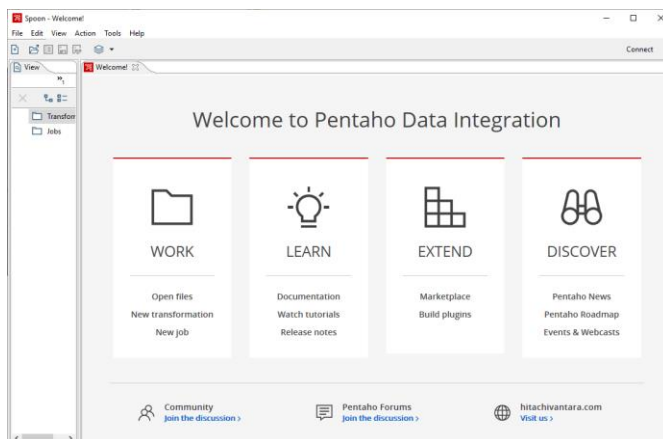
- **Pentaho Data Integration (también conocido como PDI, Kettle o Spoon).**

Enlace: https://www.youtube.com/playlist?list=PLPgjON4ZM0JBdxxDUAfCS84X79e_2CLNQ

Objetivo: Curso completo en castellano sobre PDI

Comentarios: canal youtube, videos detallados de funciones y componentes PDI, compuesto por 96 minivideos (2-10min) sobre temas muy concretos, ligados a la documentación de la herramienta.

- **Pentaho Data Integration LEARN (oficial).**



Enlaces:

- Documentación :
https://help.pentaho.com/Documentation/9.0/Products/Pentaho_Data_Integration?ecid=ms_glo_bd_en_sscepen01
- Watch tutorials: https://www.hitachivantara.com/en-us/products/data-management-analytics/pentaho/tutorials.html?ecid=ms_glo_bd_en_sscepen02#getting-started
- Samples: (desde VDI) C:\Program Files\pdi-ce-9.0.0.0-423\data-integration\samples

Objetivo: Documentación, tutoriales y ejemplos oficiales PDI

Comentarios: Consulta rápida de funciones y componentes PDI, ejemplos funcionales de los jobs y transformaciones más frecuentes.

Formato y fecha de entrega

La entrega final de esta actividad debe realizarse enviando un único mensaje al buzón de «Registro de AC» del apartado «Evaluación» del aula. Junto con el mensaje se enviará un único archivo en formato Word o PDF con la solución de la PRA2. El nombre del archivo debe ser la composición del nombre de usuario y «_BDA_PRA2». Por ejemplo, si el nombre de usuario es «bantich», el nombre del archivo debe ser «bantich_BDA_PRA2.pdf».

La fecha máxima de entrega es el 14/12/2021 a las 23:59 horas.