Recurso complementario

Ejemplo 1: extracción y carga de datos desde archivos

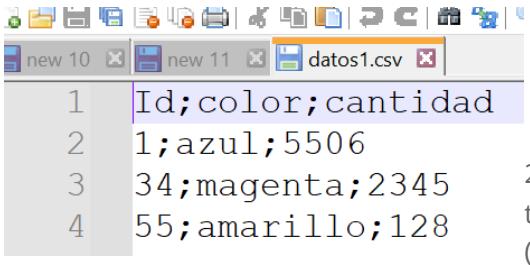
El primer ejemplo de aplicación consiste en desarrollar un proceso *ETL* para leer un archivo de tipo CSV y volcar los datos dentro de un archivo de texto plano (.TXT). Un archivo CSV, o *comma-separated values*, contiene datos tabulados en filas y columnas; los datos en una fila estarán delimitados por algún separador, como puede ser una coma, un punto, un guion, un tabulador, o cualquier otro símbolo adecuado. Por lo tanto, se creará una transformación de Pentaho Data Integration (*kettle*), a través de la interfaz *Spoon*, usando los métodos antes descritos.

Para construir la transformación (que llamaremos **transformación01**), se deben cumplir los siguientes pasos:

1. Crea un archivo de texto en algún directorio vacío, con nombre **datos1.csv**, que contenga los datos mostrados en la figura 1. Observa que los datos están separados por **puntos y comas**.

Figura 1.

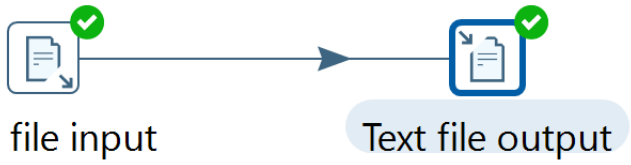
Contenido del archivo datos1.csv a leerse en la transformación



2. Crea la siguiente transformación en *Spoon* (figura 2).

Figura 2.

Transformación del ejercicio a crearse dentro de PDI

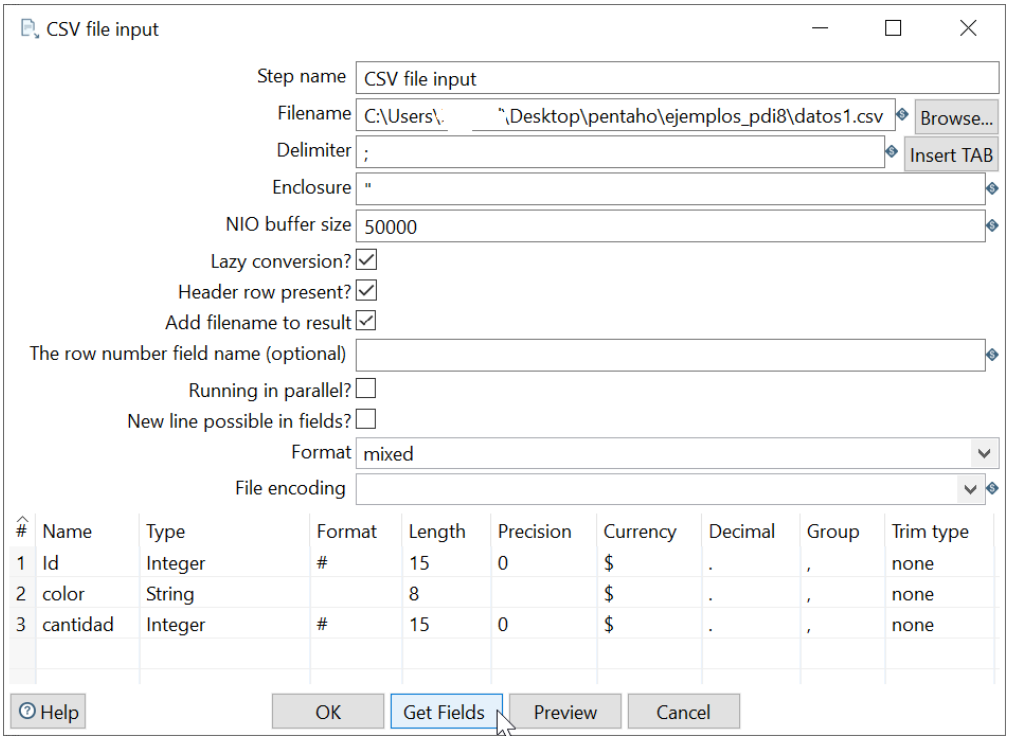


3. Haz doble clic en el primer paso (**CSV file input**) y conecta el paso con el archivo CSV ya creado. Además, completa la configuración interna (son varios pasos), según lo indicado en la figura 3. Una vez que seleccionas el archivo puedes presionar el botón **Get Fields** para leer automáticamente los nombres de columnas desde el archivo; también con el botón

Preview podrás leer previamente los datos desde el archivo CSV y fijarte si la configuración es correcta o no.

Figura 3.

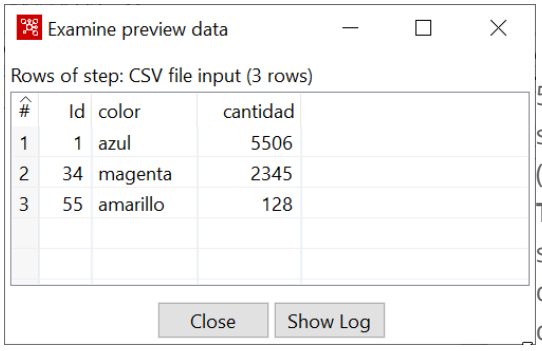
Datos de configuración para el primer paso (paso de entrada) de la transformación ETL



4. Si previsualiza los datos con el botón **Preview**, verá algo como lo que se muestra en la figura 4

Figura 4.

Previsualización de datos desde el archivo CSV

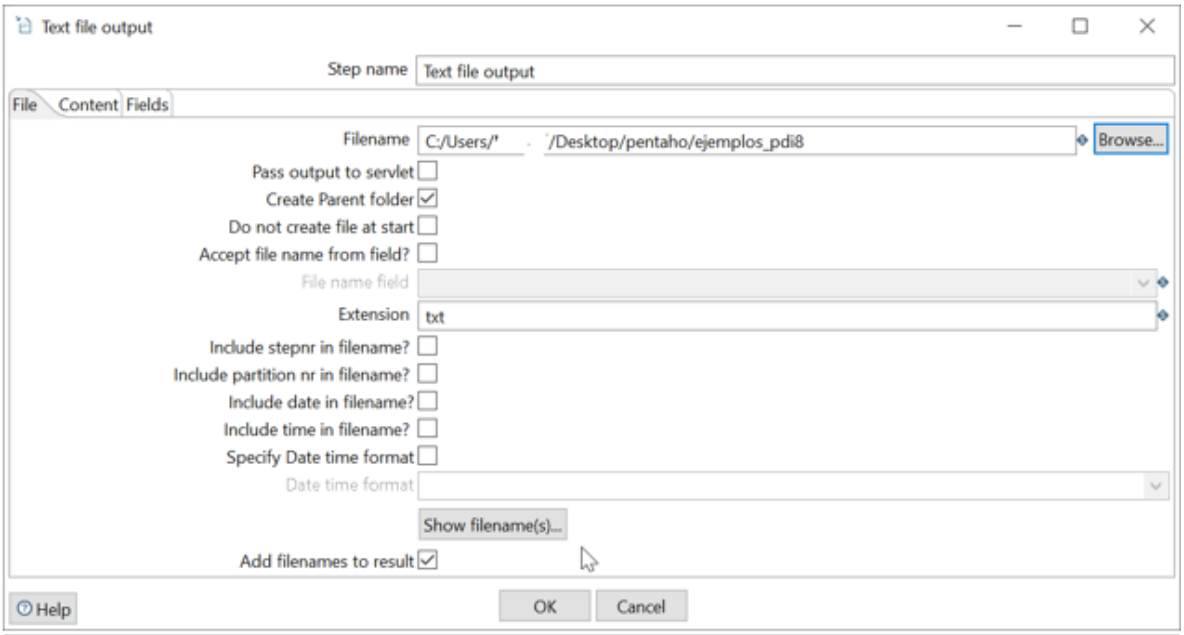


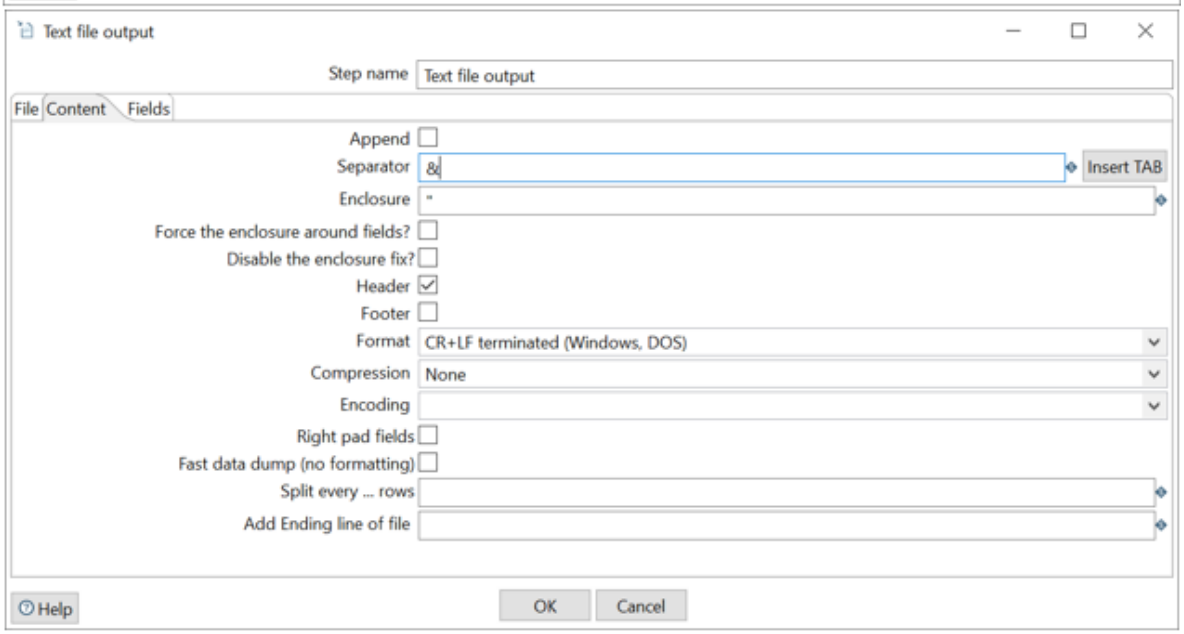
5. Ahora, configura el segundo paso (correspondiente al paso **Text file output**), siguiendo la configuración de la figura 5. Considera que acá este paso creará

un archivo de texto plano en el directorio, por lo que tendrás que inventar un nombre de salida para el nuevo archivo (que recibirá los datos desde el archivo CSV).

Figura 5.

*Configuración del paso **Text file output**, dividido en dos pestañas de configuración - nombre de archivo (arriba) y formato del contenido (abajo)*



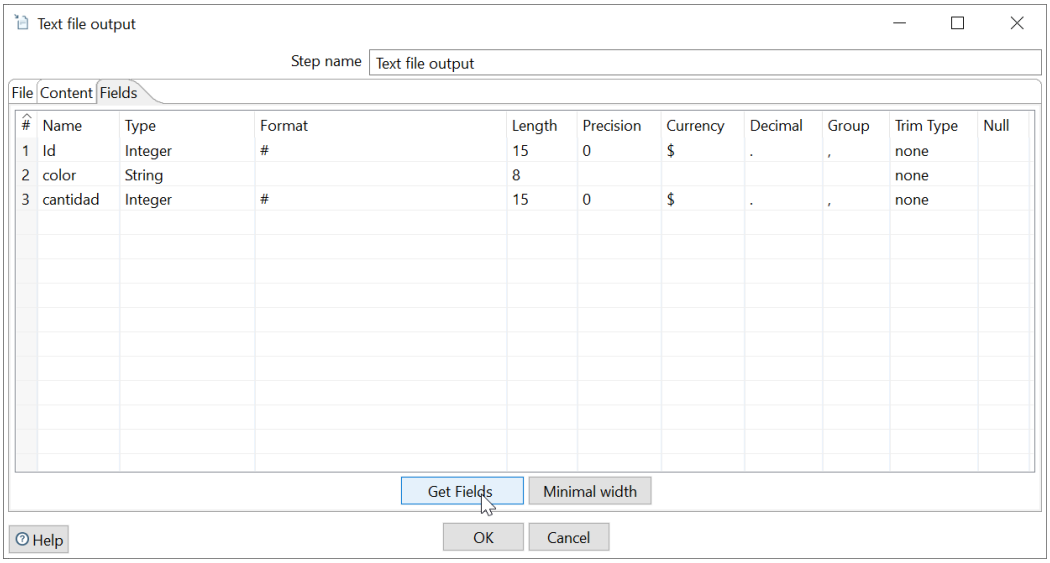


Procura configurar ambas partes dentro del componente.

6. En la tercera pestaña de configuración (*Fields*), podrá presionar el botón **Get Fields** para leer automáticamente los campos desde el paso CSV anterior (figura 6).

Figura 6.

Obtención automática de campos desde el paso de lectura del archivo CSV



7. Finalmente, presiona el botón de ejecución **Run** en la ventana de transformación para ejecutar el proceso *ETL*. Luego, mira que en el panel inferior no haya mensajes de error (en color rojo) y revisa el directorio de salida del archivo de texto para encontrar el archivo .TXT con los datos.

Figura 7.

*Para ejecutar la transformación, se debe presionar el botón **Run** en la ventana de PDI*

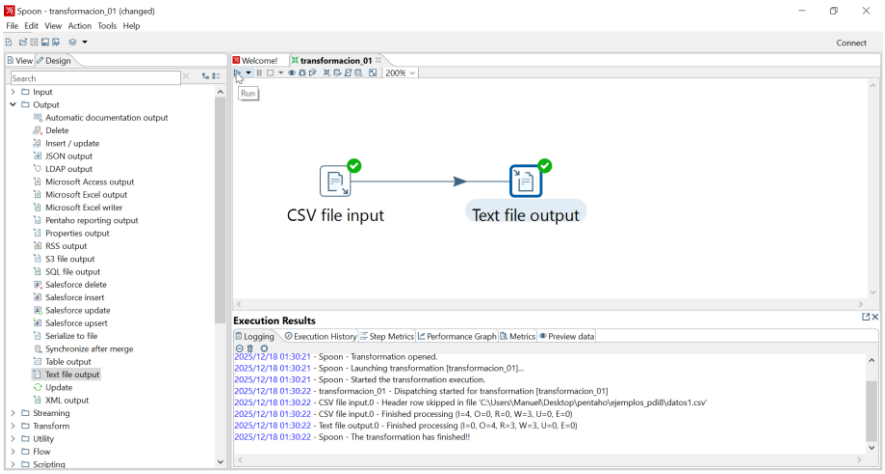
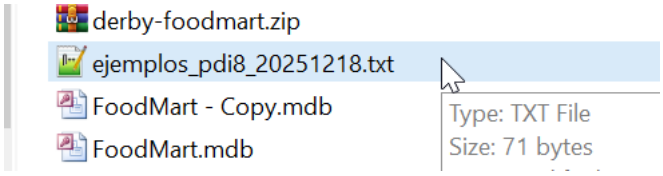
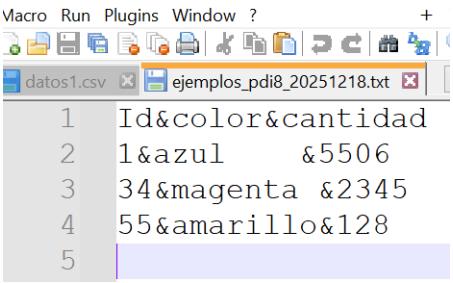


Figura 8.

Al terminar la ejecución de la transformación, se verá el archivo de salida con su contenido (el mismo contenido del archivo CSV original, pero con delimitadores diferentes, ya que se configuró el uso del símbolo '&' para ello, en el paso 5)





¡Felicitaciones! Ya creaste tu primera transformación en Pentaho Data Integration (*Kettle*), utilizando la interfaz gráfica *Spoon*. Ahora, se entregarán dos nuevos ejemplos de nuevas transformaciones con otros componentes para que amplíe su conocimiento acerca del sistema.