

CIPFP www.fpmislata.com							
Actividad:	Iniciación en Pentaho Data Integración (PDI)						
Ciclo:	IABD	Modulo:	BDA	Curso:		Agrupación:	1
Alumno/a:						Grupo:	

Actividad

Una empresa de ámbito internacional nos proporciona datos de sus ventas en distintos formatos y nos pide que los reunamos todos y construyamos un único fichero con toda la información reunida.

Los ficheros que nos proporciona son:

- Datos de Ventas.
 - Ventas de MX, DE y FR en formato CSV
 - Ventas de CA en un fichero de Excel
- Datos de Productos en un fichero CSV
- Datos de Proveedores en una tabla de MySQL

Antes de empezar a reunir los datos hay que hacer un análisis del contenido de los ficheros para identificar posibles problemas en la unificación de los datos.

Realizaremos el trabajo en 3 fases:

Primera fase

Hay que realizar la extracción de la información en dos transformaciones (en la nomenclatura de PDI)

- Una primera transformación en la que reunimos en un fichero de texto todos los datos de venta
- Una segunda en la que unimos los datos de venta, primero con los productos y luego con los proveedores

Segunda fase

Modificamos la segunda transformación siguiendo estos pasos:

1. Añadir a la última salida de la segunda transformación (el segundo join) un componente que nos permita quedarnos con aquella información que realmente consideramos importante. Nos quedamos con los siguientes campos:
 - Units
 - Revenue
 - ManufacturerID
 - Manufacturer

CIPFP www.fpmislata.com							
Actividad:	Iniciación en Pentaho Data Integración (PDI)						
Ciclo:	IABD	Modulo:	BDA	Curso:		Agrupación:	1
Alumno/a:						Grupo:	

NOTA: el paso es **Transform**→**Select Values**

2. A continuación del componente Select Values, añadimos un nuevo componente del tipo group by que nos permite agrupar la información por los campos ManufacturerID y Manufacturer
3. Después del Group By añadimos un componente **Calculator** que nos va a permitir realizar operaciones sobre nuestros datos, a través de la columna Calculation la cual nos ofrece un amplio conjunto de operaciones que escoger.

En este caso seleccionamos: A/B donde A será el importe Total y B el número de unidades vendidas, simplemente queremos calcular el precio medio de cada unidad vendida para cada fabricante.

4. Finalmente añadimos un componente de salida Text file Output.

Tercera fase

Finalmente, crearemos un job que agrupe las dos transformaciones