

CIPFP www.fpmislata.com							
Actividad:	Pentaho Data Integration (PDI). Ejemplo Instituto						
Ciclo:	IABD	Modulo:	BDA	Curso:		Agrupación:	1
Alumno/a:						Grupo:	

Actividad

En un centro de estudios necesitan analizar los datos de las notas de los alumnos, la información la tienen repartida en dos ficheros uno en formato CSV con los datos de las Actividades y otro en formato XLS con las calificaciones de los distintos grupos uno en cada hoja.

Esta es la información de la que se dispone y la estructura final de los datos

Ficheros:

- Notas (Fichero de Excel con 3 Hojas)
- Actividades (Fichero CSV)

Tablas Finales

- Actividades
 - ID_Actividad (Varchar(6))
 - Descripción (Varchar(35))
 - Trimestre (Integer)
 - Peso (Integer)
 - Fecha_Ini (Date)
 - Fecha_Fin (Date)
- Alumnos
 - ID_Alumno (Varchar(5))
 - Nombre (Varchar((25)) → En mayúsculas
 - Apellido1 (Varchar((25)) → En mayúsculas
 - Apellido2 (Varchar((25)) → En mayúsculas
 - Grupo (Varchar((10))
 - Email (Varchar((40))
- Notas
 - ID_Alumno (Varchar((5))
 - ID_Actividad (Varchar((6))
 - Nota (Double)

CIPFP www.fpmislata.com							
Actividad:	Pentaho Data Integration (PDI). Ejemplo Instituto						
Ciclo:	IABD	Modulo:	BDA	Curso:		Agrupación:	1
Alumno/a:						Grupo:	

Los pasos y acciones que debes realizar son:

1. Revisión y análisis de los datos de entrada:
2. Elabora un plan de trabajo definiendo jobs y transformaciones. Dibújalo si te facilita el trabajo.
3. Analiza las comprobaciones previas antes de empezar. Conexiones a BBDD, existencia de ficheros, etc...
4. Ingesta de los datos y transformación
5. Carga de los datos
 - a. Carga los datos en tablas con la estructura definida al inicio del enunciado de la actividad

Ten en cuenta que:

- Deberás crear tantas transformaciones y jobs como sean necesarios
- Crea etiquetas para facilitar la lectura de la ETL
- Definir una estructura de jobs y transformaciones clasificada y organizada
- Deberás realizar comprobaciones para evitar fallos en el proceso.

AYUDA:

Necesitarás una de estas dos transformaciones para poder obtener la estructura de tablas de salida

- ROW DENORMALISER
- ROW NORMALISER