

Pentaho Data Integration - Kettle (CE)

I. Objetivos

Este curso está enfocado en el uso del componente de ETL de Pentaho: Pentaho Data Integration.

Al finalizar el curso, el alumno tendrá suficientes conocimientos para:

- Instalar el producto y la versión de Java correspondiente.
- Realizar transformación de datos (ETL) desde distintas fuentes de datos a diferentes destinos.

II. Requisitos

- Conocimientos de SQL.
- Opcional:
 - Conocimiento de otras herramientas de ETL.

III. Duración

- 20 horas.

IV. Metodología

El curso se desarrolla mediante la exposición teórica acompañada de demostraciones prácticas y explicaciones de los resultados obtenidos.

El alumno lleva a cabo prácticas con el producto para cada concepto explicado. Se utilizarán distintas fuentes de datos, tanto ficheros de texto, como hojas de cálculo y bases de datos, principalmente relacionales (MySQL).

Resolución de dudas de los conceptos expuestos.

V. Contenido

- Introducción a Pentaho Community Edition y sus componentes.
 - Componentes de PCE.
 - Requisitos de Java.
 - Instalación de Java y PDI.

Pentaho Data Integration - Kettle (CE)

- PDI. Pentaho Data Integrator.
 - Bases de datos:
 - Conexiones a bases de datos.
 - Uso de conexiones compartidas.
 - Componentes básicos de transformaciones:
 - Creación de transformaciones.
 - Importación y exportación usando tablas.
 - Importación y exportación desde ficheros de texto plano, csv, Excel, xml, etc.
 - Uso de calculadora y fórmulas.
 - Selección de columnas y filtros de datos.
 - Uso de lookup, group by, split, pivot.
 - Merge join.
 - Mapping.
 - Trabajos, variables y propiedades:
 - Modificación de propiedades de Kettle.
 - Creación de trabajos.
 - Uso de variables.
 - Uso de parámetros.
 - Flujo de trabajos y gestión de errores.
 - Elementos adicionales:
 - Uso de PDI desde terminal.
 - Ejecutar código en base de datos.
 - Procesamiento dinámico de ficheros.
 - Mover archivos en el sistema operativo.
 - Escritura en el "log" de PDI.
 - Espera de ficheros (filewatcher).
 - Comprobación de existencia de archivos.
 - Comprobación de existencia de tabla (BBDD).
 - Prácticas con transformaciones y trabajos.