

## AUDIENCIA

Diseñado para analistas e integradores de datos, desarrolladores y administradores de bases de datos.

De gran valor para cualquier profesional de la informática con una experiencia mínima de 2 años con lenguaje SQL y diseño e implementación de bases de datos relacionales.

## DURACIÓN

3 Días.

## Pentaho Data Integration Fundamentals - DI1000

Al finalizar el curso el participante debe estar en capacidad de:

- Instalar Pentaho Data Integration
- Crear, pre-visualizar, y ejecutar transformaciones
- Gestión del Repositorio de PDI
- Analizar resultados en «*Step Metrics*» y «*Log View*»
- Crear conexiones a bases de datos con el *Database Explorer* para explorarlas
- Crear múltiples transformaciones de datos con diversos tipos de “*steps*”
- Mapear estructuras de datos transaccionales a bases de datos analíticas
- Crear transformaciones que utilicen parámetros
- Crear dimensiones que cambian lentamente (*Slowly-Changing Dimensions*)
- Cargar *data warehouse* con patrones de diseño ETL
- Crear Trabajos (*Jobs*) de PDI : ejecutar múltiples transformaciones, uso de variables y sub-trabajos.
- Auditar los pasos de las transformaciones y trabajos, examinar los registros
- Identificar y resolver errores en transformaciones y tareas (*Jobs*)
- Administrar el Repositorio Empresarial.



Contáctenos:  
[cursos@matrixcpmsolutions.com](mailto:cursos@matrixcpmsolutions.com)

## Día 1

### Módulo 1: Introducción a *Pentaho Data Integration*

- Lección 1: Objetivos
- Lección 2: ¿Qué es *Pentaho Data Integration* (PDI)?

### Módulo 2: Básicos de las transformaciones

- Lección 1: Ejecución y personalización del PDI
- Lección 2: Crear una transformación “Hola Mundo”
- Lección 3: Manejo de errores y registro de actividades
- Ejercicio 1: Su primera transformación

### Módulo 3: Repositorio

- Lección 1: Creación del repositorio
- Ejercicio 1: Crear y cargar el repositorio

## Día 2

### Módulo 4: Lectura y escritura de archivos

- Lección 1: Pasos de Entrada y Salida
- Lección 2: Limpieza de datos
- Lección 3: Crear parámetros de rutas de archivos
- Ejercicio 1: Entradas con múltiples salidas usando parámetros
- Ejercicio 2: Crear archivos serializables desde múltiples archivos
- Ejercicio 3: Des-serializar archivos

### Módulo 5: Trabajando con bases de datos

- Lección 1: Conectar y explorar bases de datos orígenes
- Ejercicio 1: Lectura y escritura de tablas de bases de datos
- Ejercicio 2: Lectura desde tablas y escritura hacia archivos
- Lección 2: Limpieza de datos
- Lección 3: Parámetros y variables

- Ejercicio 3: Insertar datos usando pasos de inserción/actualización
- Ejercicio 4: *Table Input* y parámetros
- Ejercicio 5: Asistente para el copiado de tablas

### Módulo 6: Flujos de datos y *Lookups*

- Lección 1: Copia y distribución de datos
- Lección 2: *Lookups* (búsquedas)
- Lección 3: Unión y fusión de datos
- Ejercicio 1: Trabajando con procesamiento en paralelo
- Ejercicio 2: Uso de paso *Stream Lookup*

## Día 3

### Módulo 7: Cálculos

- Lección 1: Paso *Group By*
- Lección 2: Paso *Calculator*
- Lección 3: Expresiones regulares
- Lección 4: Expresiones Java definidas por el usuario
- Lección 5: *Javascript*
- Ejercicio 1: Cálculos

### Módulo 8: Trabajos (*Jobs*)

- Lección 1: Introducción a los *Jobs*
- Lección 2: Envío de alertas
- Lección 3: Uso del *Kitchen* para ejecutar *Job* vía línea de comando
- Chequeo de JVM
- Ciclos e iteraciones
- Manejo de errores a través de correos electrónicos

### Módulo 9: Programación de tareas

- Configurando el *Scheduler*
- Monitoreando trabajos programados
- Programación y monitoreo de trabajos.



Contáctenos:  
cursos@matrixcpmsolutions.com

