

Curso de Entrenamiento

Este curso está diseñado para introducirle a los diversos conceptos de Big Data y el entorno de trabajo *Hadoop*, con tecnologías y productos *Pentaho*.

Sobre la base del curso *Pentaho Data Integration Fundamentals*, usted aprenderá cómo funciona *Pentaho* y los entornos de trabajo *Hadoop*, con las siguientes tecnologías: *HDFS*, *Sqoop*, *Pig*, *Oozie*, *MapReduce*, *YARN*, *Hive*, *Impala*, *Hbase*, *Flume*

DURACIÓN

2 Días.



Contáctenos:

pentaho@matrixcpmsolutions.com

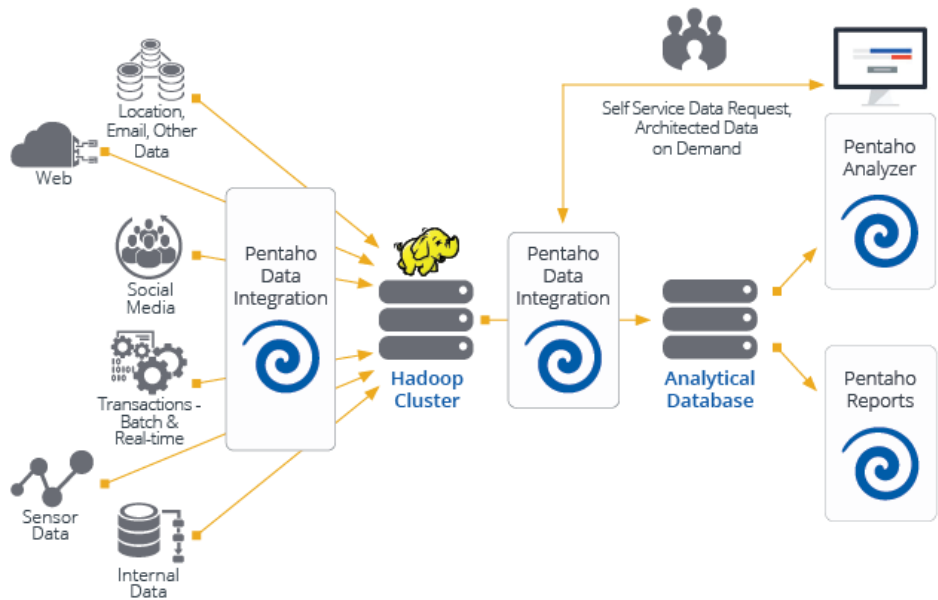
Pentaho and Hadoop Framework Fundamentals – DI2000

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

- Utilizar las tecnologías de *Hadoop* desde la línea de comandos nativa y con *Pentaho Data Integration*
- Emplear las mejores prácticas en la inyección y procesamiento de datos
- Utilizar *Pentaho Interactive Reporting and Analyzer* para hacer reportes desde *Impala*

Beneficios del curso:

- Mejora la productividad brindándole a su equipo de integración de datos las habilidades que necesita para utilizar *Pentaho Data Integration* con fuentes de datos *Hadoop*
- Ayuda a implementar el modelo de solución *Streamlined Data Refinery* (Refinería de Datos Optimizada) para Big Data usando *PDI* y *Hadoop*.
- Interactivo, con un práctico material de capacitación que mejora significativamente el desarrollo de habilidades, maximizando la retención del contenido explicado.



Día 1

Módulo 1: Introducción a Pentaho y Big Data

- Lección 1: Arquitecturas Big Data
- Lección 2: Casos de uso de *Streamlined Data Refinery* (Refinería de Datos Optimizada)
- Lección 3: Descripción general de las herramientas de Pentaho
 - Ejercicio 1: Explorando el entorno

Módulo 2: HDFS

- Lección 1: Fundamentos de HDFS
 - Descripción de lectura y escritura de datos en HDFS.
 - Replicación de datos y Tolerancia a fallas
- Lección 2: Las mejores prácticas HDFS.
 - Tamaños de archivos
 - Tipos de archivos
 - Compresión
- Lección 3: *Pentaho Data Integration* con HDFS
 - Importar / Exportar datos entre el sistema de archivos local y HDFS
 - Pasos para el manejo de Archivos
 - Las mejores prácticas con PDI / HDFS

Módulo 3: Inyección de los datos

- Lección 1: *Sqoop*
 - Fundamentos de *Sqoop*
 - *PDI* y *Sqoop*
 - Ejercicio 3: Importación / exportación de datos entre una BD y *Hadoop* Usando *Sqoop*
- Lección 2: *Flume*
 - Fundamentos de *Flume*
 - *PDI* y *Flume*
- Lección 3: Las mejores prácticas de Ingestión de Datos
 - *PDI* vs. *Sqoop* vs. *Flume*
 - Agregación de archivos de menor tamaño en archivos más grandes
 - Las mejores formas de almacenar archivos no divisibles como XML y JSON

Módulo 4: Procesamiento de Datos

- Lección 1: Conceptos de MapReduce
 - Mapeador
 - Reductor
 - Combinador
- Lección 2: Arquitectura MR1
 - Controlador
 - Seguimiento de procesos
 - Seguimiento de tareas
 - Con orden / aleatorio
 - Particionador
- Lección 3: YARN / MR2
 - Fundamentos YARN
 - Arquitectura MR2
- Lección 4: Utilizando PDI para escribir en MapReduce
 - Desarrollando MR Usando PDI
 - MR1 vs. HILO / MR2 en PDI
- Lección 5: Las mejores prácticas de PDI / MR
 - Qué hacer y qué no al escribir aplicaciones PDI y MapReduce
 - El uso de caché distribuida de Hadoop
 - Compresión
 - Ejercicio 4: Uso de PDI para desarrollar PMR en MR2 / YARN
- Lección 6: *PDI / Carte* sobre YARN
 - Fundamentos de *Carte* sobre YARN
 - Ejercicio 5: Construir una transformación que se ejecute en YARN
- Lección 7: *Pig*
 - Fundamentos de *Pig*
 - *PDI* y *Pig*
 - Ejercicio 6: Ejecutar la aplicación *Pig* y ejecutar un Script de *Pig* utilizando PDI.



Día 2

Módulo 5: Arreglo de Procesos

- Lección 1: Conceptos básicos Oozie
- Lección 2: Características de Arreglo de Procesos PDI
- Lección 3: PDI y Oozie
- Ejercicio 7: Ejecutar la aplicación Oozie y ejecutar un script Oozie desde PDI

Módulo 6: Hadoop y SQL

- Lección 1: Hive tradicional
- Lección 2: Hive / TEZ
- Lección 3: Impala
- Lección 4: El uso de PDI con Hive e Impala
- Lección 5: Las mejores prácticas con SQL / Hadoop / PDI
- Ejercicio 8: Trabajo con Impala utilizando el Command Line, HUE y PDI

Módulo 7: HBase

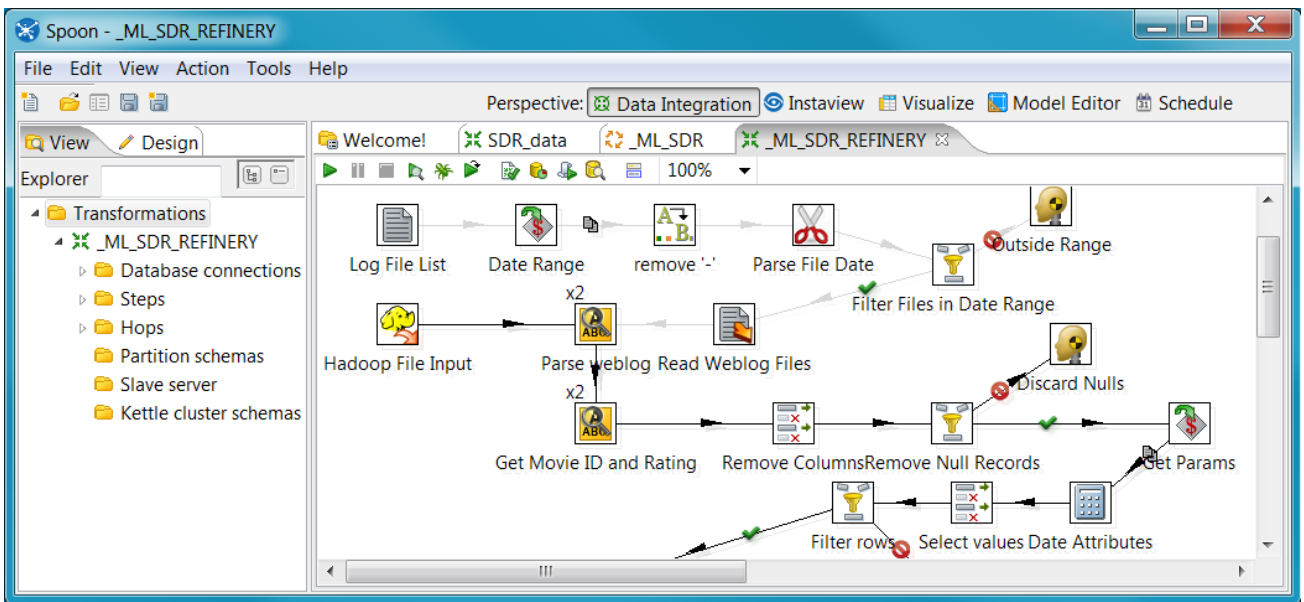
- Lección 1: Conceptos básicos Hbase
- Lección 2: PDI y Hbase
- Lección 3: Las mejores prácticas con PDI / HBase
- Ejercicio 9: Trabajar con HBase y PDI

Módulo 8: Elaborando reportes en Big Data

- Lección 1: Uso de Pentaho Report Designer con Hadoop
- Lección 2: El uso de Pentaho Analyzer con Hadoop
- Lección 3: Mejores prácticas para la presentación de informes sobre datos de Hadoop
- Ejercicio 9: Crear un PRD y Analyzer Report utilizando datos de Hadoop

Módulo 9: Tecnologías Adicionales de Pentaho y Big Data

- Lección 1: Flume y PDI
- Lección 2: Storm y PDI
- Lección 3: Kafka y PDI



Contáctenos:
pentaho@matrixcpmsolutions.com