Duration: 3 Days

Course Content

The volume, variety and velocity of data are increasing rapidly. Organizations need fast and easy-to-use tools to harness data for actionable insight. One of the biggest challenges facing organizations today is the requirement to provide a consistent, single version of the truth across all sources of information in an analytics-ready format.

With powerful data extract, transform and load (ETL) capabilities, an intuitive and rich graphical design environment, and an open and standards-based architecture, Pentaho Data Integration is increasingly the choice over proprietary and homegrown data integration tools.

Pentaho Data Integration provides a full ETL solution, including:

- Rich graphical designer to empower ETL developers
- Broad connectivity to any type of data, including diverse and big data
- Enterprise scalability and performance, including in-memory caching
- Big data integration, analytics and reporting, including Hadoop, NoSQL, traditional OLTP & analytic databases
- Modern, open, standards-based architecture

COURSE BENEFITS

- Improve productivity by giving your data integration team the skills they need to succeed with Pentaho Data Integration
- Learn to deliver data to a wide variety of applications using Pentaho's out-of-the-box data standardization, enrichment and quality capabilities
- Interactive, hands-on training materials significantly improve skill development and maximize retention

SKILLS ACHIEVED

At the completion of this course, you should be able to:

- Create, preview, and run basic transformations containing steps and hops
- View transformation results in the Step Metrics view and the Log view
- Configure the Pentaho Enterprise Repository, including basic security
- Use the Pentaho Enterprise Repository to: create folders, store transformations and jobs, move, lock, revise, delete, and restore artifacts.
- Configure error handling for transformation steps
- Create a database connection and use Database Explorer to interact with data sources
- Create transformations that involve configuring the following steps: Table input, Table output, Text file output, CSV file input, Insert/Update, add constants, Filter, Value Mapper, Stream lookup, Join rows, Merge join, Sort rows, JavaScript, Database Lookup, Set Environment Variables
- Learn how to use transformation steps to perform complex calculations on the data stream
- Create reusable transformations using parameterized values and environment variables
- Use Pentaho Data Integration to cleanse and correct data

- Load data from and write data to different data sources
- Create Pentaho Data Integration jobs that: run multiple transformations, use variables, contain sub-jobs, provide built-in error notification, load and process multiple text files, and convert files into Microsoft Excel format
- Configure logging for transformation steps and for job entries and examine the logged data
- Schedule and monitor the execution of a transformation in Pentaho Data Integration and in the Pentaho Enterprise Console

Who Should Attend

This course is the third course in the Data Analyst learning path. Students with prior database development or administration experience who are new to Pentaho Data Integration should take this course.

Prerequisites

There are no prerequisites for this course but some ETL experience is preferred.

Course Outline

Day 1

Module 1: Introduction to Pentaho Data Integration

- Lesson 1: Objectives & Class Logistics
- Lesson 2: What is Pentaho Data Integration (PDI)?

Module 2: Transformation Basics

- Lesson 1: Learning the PDI User Interface
- Lesson 2: Creating Transformations
  Exercise 1: Generate Rows, Sequence, Select Values
- Lesson 3: Error Handling & Logging Introduction
- Lesson 4: Introduction to Repositories

Module 3: Reading & Writing Files

- Lesson 1: Input & Output Steps
- Lesson 2: Parameters & kettle. Properties
  Exercise 2: CSV Input to Multiple Text Output Using Switch/Case
  Exercise 3: Serializing Multiple Text Files
  Exercise 4: De-serialize a File

inixindo

Day 2

Module 4: Working with Databases

- Lesson 1: Connecting to & Exploring a Database
- Lesson 2: Table Input & Output
  Exercise 5: Reading & Writing to Database Tables
- Lesson 3: Insert, Update, & Delete Steps
- Lesson 4: Data Cleansing
- Lesson 5: Using Parameters & Arguments in SQL
  Exercise 6: Input with Parameters & Table Copy Wizard

Module 5: Data Flows & Lookups

- Lesson 1: Copying and Distributing Data
  Exercise 7: Parallel Processing
- Lesson 2: Lookups
  Exercise 8: Lookups & Data Formatting
- Lesson 3: Merging Data

Day 3

Module 6: Calculations

- Lesson 1: Using the Group by Step
- Lesson 2: Calculator
  Exercise 9: Calculating & Aggregating Order Quantity
- Lesson 3: Regular Expression
- Lesson 4: User Defined Java Expression
- Lesson 5: JavaScript

Module 7: Job Orchestration

- Lesson 1: Introduction to Jobs
  Exercise 10: Loading JVM Data into a Table
- Lesson 2: Sending Alerts
- Lesson 3: Looping & Conditions
  Exercise 11: Creating a Job with a Loop
- Lesson 4: Executing Jobs from a Terminal Window (Kitchen)

Module 8: Scheduling

- Lesson 1: Setting up the Scheduler
- Lesson 2: Monitoring Scheduled Tasks

Module 9: Exploring Data Integration Repositories
- Lesson 1: The Pentaho Data Integration Repository
  Exercise 12: Using the Pentaho Enterprise Repository

Module 10: Detailed Logging
- Lesson 1: Detailed Logging throughout Execution