

Training Content

Pentaho Data Integration DI1000



Overview

Pentaho Data Integration DI 1000 training provides an overview of the Data Integration. This training will develop the skill to create easy ETL Solutions using the UI based designer. The Pentaho Data Integration can connect to various sources including Hadoop, NoSQL and relational databases. This training will also help you install the Pentaho Data Integration.

After the course

Once the course is complete, you will be able to achieve the following:

- Create Jobs and Transformation.
- Create, preview, and run basic transformations containing steps and hops
- View transformation results in the Step Metrics view and the Log view
- Create a database connection and use Database Explorer to interact with a data source
- Create more complex 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, Row normalizer, JavaScript, Dimension lookup/update, Database Lookup, Get Data from XML, Set Environment Variables, and Analytic query
- Map the structure of an online transaction processing database to the structure of an online analytical processing database
- Use ETL design patterns to populate a data warehouse
- 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
- Create a transformation that uses a partition schema to partition data to slave servers in the cluster

Pre-requisites

The participants should have the basic knowledge on the Linux/Windows system as the training may require them to do some basic operations. They need not have the understanding of Business Intelligence or Data Warehouse. The understanding of any of the Relational Databases and dimensional modeling will be an added advantage.

You may also need to bring a Windows/Linux Laptop, minimum with the following configuration:

1. At least 2 GB RAM on the laptop.
2. At least 1 GB of space on Hard Drive.
3. 1.4 GHz CPU or above.
4. DVD drive
5. USB port
6. SUN JAVA 1.6 or above.
7. MySql 5.x/PostgreSQL 8.x or above

Topic	Hours
Day 1	
MODULE 1: PENTAHO DATA INTEGRATION OVERVIEW Exercise 1: Introducing Pentaho Data Integration MODULE 2: INPUTS AND OUTPUTS Exercise: Inputs and Outputs MODULE 3: INTRODUCTION TO THE TRAINING DATA (LECTURE AND DEMO) MODULE 4: DATA WAREHOUSE STEPS Exercise 3: Data Warehouse Steps	8
Day 2	
MODULE 5: LOOKUPS (LECTURE AND DEMO) MODULE 6: FIELD TRANSFORMATIONS, PART 1 Exercise 4: Lookups and Field Transformations MODULE 7: SET TRANSFORMATIONS Exercise 5: Set Transformations MODULE 9: FIELD TRANSFORMATIONS, PART 2 Exercise: Field Transformations, Part 2 MODULE 10: LOADING THE TIME DIMENSION AND THE FACT TABLE Exercise 7: Loading a Fact Table	8
Day 3	
MODULE 11: INTRODUCTION TO JOBS Exercise 8: Creating a Job MODULE 12: ADVANCED JOB CONCEPTS Exercise 9: Advanced Job Concepts MODULE 13: COMMON SCRIPTING USES Exercise 10: Using JavaScript MODULE 14: DYNAMIC TRANSFORMATIONS (LECTURE AND DEMO) MODULE 15: USING XML IN PENTAHO DATA INTEGRATION Exercise 11: Using XML MODULE 16: PORTABLE TRANSFORMATIONS AND JOBS Exercise 12: Portable Transformations and Jobs	8
Day 4	
MODULE 17: LOGGING Exercise 13: Configuring Logging MODULE 18: ERROR HANDLING IN TRANSFORMATIONS Exercise 14: Error Handling in Transformations MODULE 19: PENTAHO ENTERPRISE REPOSITORY Exercise 15: Pentaho Enterprise Repository MODULE 20: SCHEDULING AND MONITORING Exercise 16: Scheduling and Monitoring MODULE 21: PRE AND POST-PROCESSING Exercise 17: Constraint and Index Management	8

<p>MODULE 22: INTERPRETING RUNTIME DATA (LECTURE AND DEMO) (OPTIONAL) MODULE 23: CLUSTERING AND PARTITIONING Exercise 18: Clustering and Partitioning</p>	
--	--