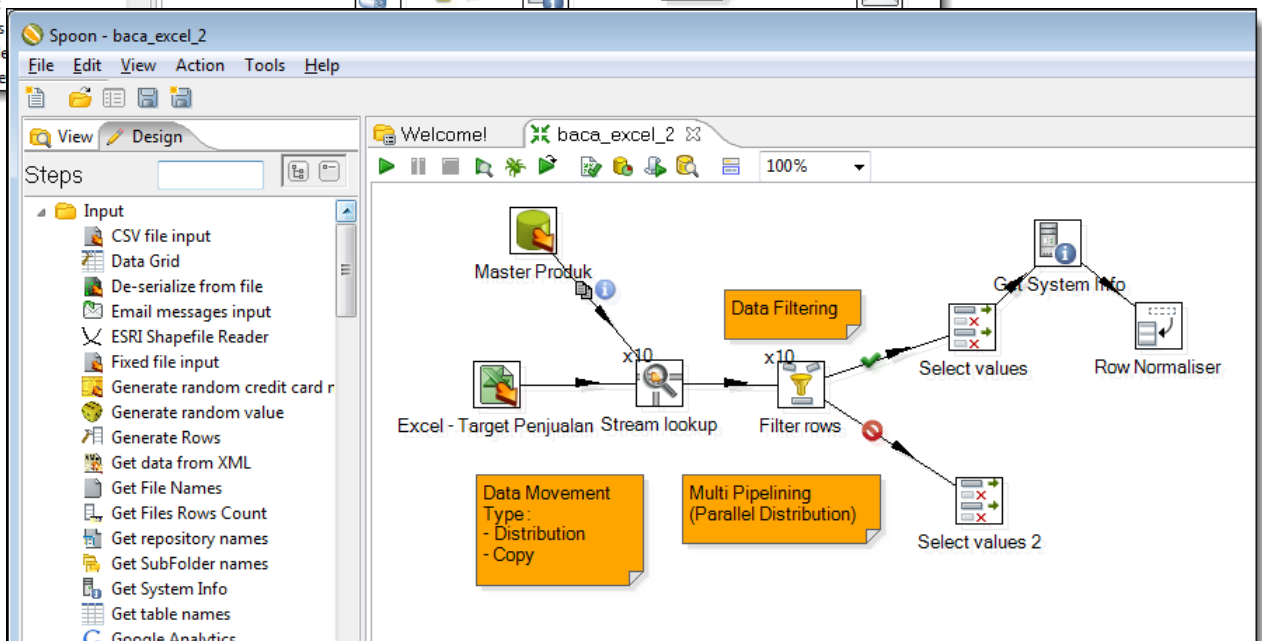
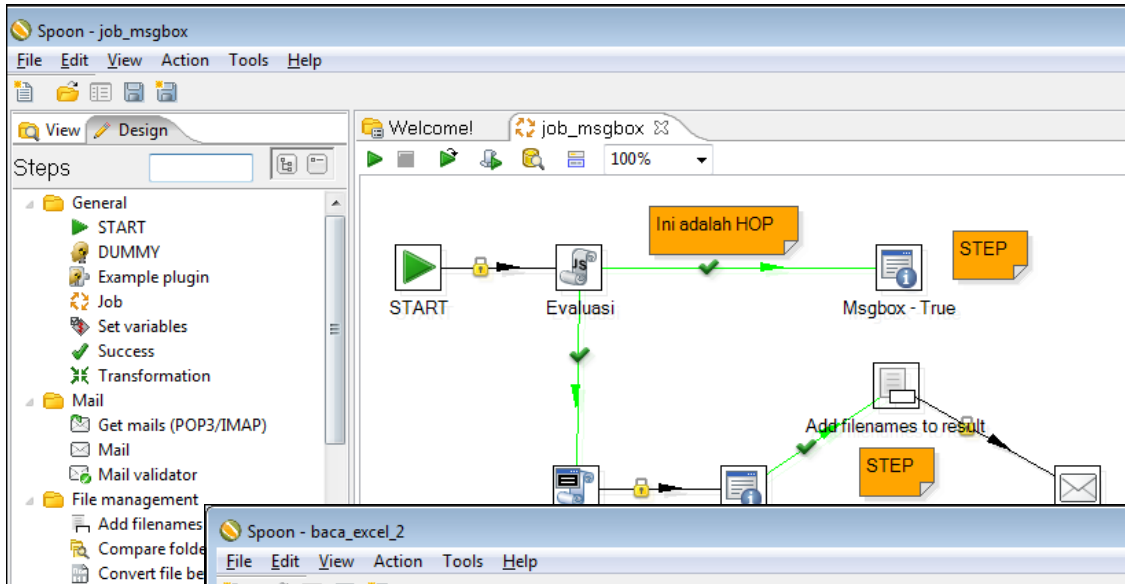


Data Integration with Kettle - Open Source ETL



Course ID : PHI-DW-2013-08

Platform : Java – Windows

Language : Indonesian

<http://www.phi-Integration.com>

I. Who Should Attend ?

This course is designed for those new to ETL (Extract, Transform, and Load) or need to understand usage of Kettle in daily data integration activities.

II. Objectives

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

- Understand the concepts of ETL and why it is needed.
- Create and manage simple jobs / transformations.
- Consume data from several data sources.
- Troubleshooting techniques.
- Scheduling Execution of Job / Transformation.
- Manage Complex Multi Database Environments using Kettle.

III. Duration

3 days / 18 hours

IV. Prerequisites

- Basic understanding one of several popular DBMS (Oracle, SQL Server, MySQL, etc) and of Structured Query Language (SQL).
- No understanding of other Pentaho tools is needed.

V. Requirements

- PC or Laptop with minimum of 2GHz CPU (Multi Cores), 2 GB of RAM, DVD Drive and 10 GB of available hard disk space.
- Softwares :
 - Microsoft Windows XP or Windows 7.
 - Java Runtime Environment (JRE) / Java Development
 - MySQL 5.0 Database Server.
 - PostgreSQL 9.2
 - Pentaho Data Integration (Kettle)
 - Pentaho BI Server Community Edition.

VI. Course Outline

1. Installation and Configuration

- Java Runtime Environment / Java Development Kit 1.6.
- Pentaho Data Integration 4.4.
- MySQL 5.0 Database Server Clients.
- PostgreSQL 9.2 Database Server and Clients.
- Data and Script Samples.

2. Introduction to Pentaho Data Integration a.k.a. Kettle

- Kettle as open source ETL suite.
- Kettle's Architecture.
- Kettle Application Stacks.
 - Spoon : Graphical UI Designer for job / transformation steps.
 - Pan : Command line batch script for transformation execution.
 - Kitchen : Command line batch script for transformation execution.
 - Carte : Cluster server for MPP Purpose.
- Kettle Concepts : Job / Transformation, Step, Hop, Row and Meta Data.

3. ETL Development with Spoon

- Spoon Development Environment.
- Creating and managing database connections.
- Data Explorer.
- Creating and Executing Job and Transformation.
- Basic Input / Output Steps : Text File, Excel, and Relational / JDBC based Table.
- Understanding Apache VFS (Virtual File System) in Kettle.

4. Short Introduction to PostgreSQL and MySQL.

- MySQL 5 database features introduction.
- Accessing MySQL's PHI-Minimart DB using SQLYog and PHPMysqlAdmin.
- PostgreSQL 9.2 database features introduction.
- Accessing and Administering PostgreSQL's PHI-Minimart DB using PGAdmin.

5. More Job and Transformation Steps

- Job Steps : Conditional, Email, SQL, and File Management.
- Data Normalisation and Denormalisation Steps.
- Lookups and Joins.
- Calculator and Modified Java Script Value.
- Selecting, Removing and Changing Meta Data.
- Controlling Job and Transformation Flows via Hop.

6. Variables

- Understanding Kettle.properties.
- Variable and Its Common Usage in Kettle.
- Ways to Configure and Accessing Variables.
- Advance Usage of Variable in Complex Environment.

7. Repository

- Filesystem and RDBMS based script repository.
- Setting up Kettle's RDBMS repository.
- Variable Behavior in Job and Transformation.
- Initialize variable in kettle.properties.

8. Logging

- Setting and Accessing Logging in Spoon.
- Logging Levels.
- Configure Log Outputs to RDBMS Tables.

9. Data Staging

- What is Data Staging?
- Physical I/O versus In-Memory Processing.
- Data Structure Reconstruction.

10. Advance Controls

- Shared Objects.

- Error Handling in Data Transformation.
- Email job results.

11. Automation and Scheduling

- Pan : Executing transformation from Command Line / Terminal.
- Kitchen : Executing transformation from .
- Using Windows Task Scheduler to Execute Job and Transformation.