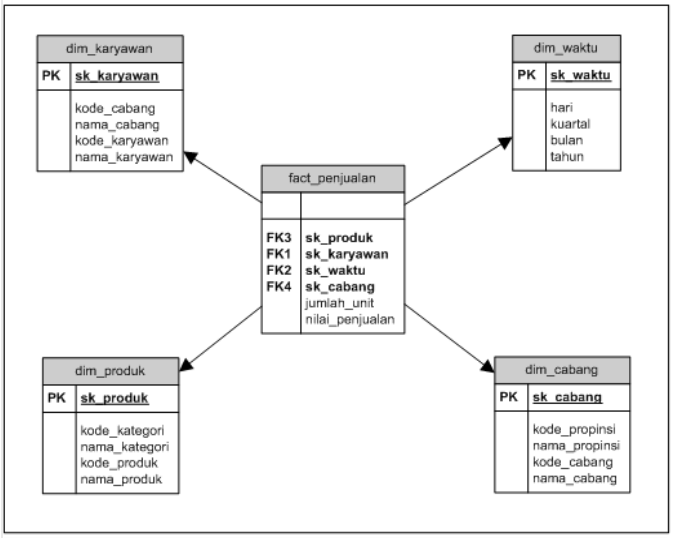
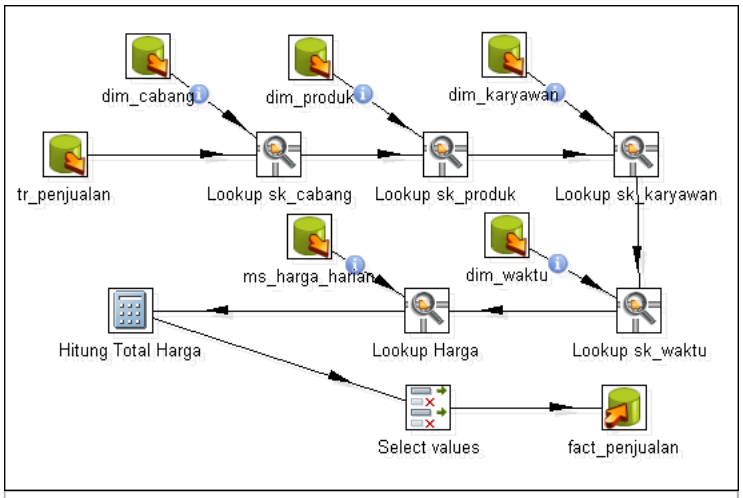
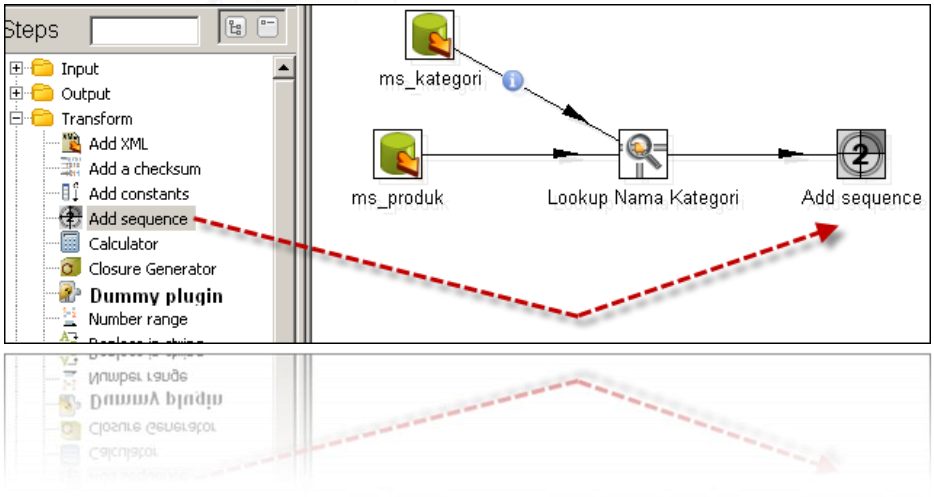
PHI-Integration.com
Aligning Business and IT !

Data Warehouse Series

Dimensional Modelling with Kettle







Course ID : PHI-DW-2013-09

Platform : Java – Windows - PostgreSQL

Language : Indonesian

http://www.phi-Integration.com

I. Who Should Attend ?

This course is designed for those new to Ralph Kimball's Dimensional Modelling but has gained proficiencies in Kettle (Pentaho Data Integration).

II. Objectives

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

- Understand the concepts of Data Warehouse with Dimensional Modelling.
- Building and Populating Dimension and Fact Tables using Kettle.
- Understand Several Subsystems of ETL Data Integration : Data Profiling, Several Slowly Changing Dimension (SCD), and Change Data Capture (CDC).
- Building and Populating Aggregate Tables and "Helper" Tables.

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).
- Basic understanding of Pentaho Data Integration.

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
  - PostgreSQL 9.2 Database Server.
  - 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.
- PostgreSQL 9.2 Database Server and Clients.
- Data and Script Samples.

2. Dimensional Modelling

- Fact and Dimension Tables.
- Star and Snowflake Schema.

3. Short Introduction to PostgreSQL

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

4. Dimension Table

- Structure of Dimension Table.
- Surrogate Keys.
- Dimension Hierarchies, Grain Levels, and Ragged Hierarchies.
- Bridge Tables.
- Late Arrival of Dimension Data.

5. Date Dimension

- Introduction to Date Dimension as Calendar Table.
- Build and Populate Date Dimension using SQL Queries.
- Build and Populate Date Dimension using ETL.

6. Fact Table

- Structure of Fact Table.
- Atomic and Aggregate Fact Tables.
- Additive and Non-Additive Measures.
- Late Arrival of Fact Data.

- Factless Fact Table.

7. ETL Subsystem : Slowly Changing Dimension (SCD)

- What is Slowly Changing Dimension ?
- Types of SCD
- Using “Dimensional Insert / Update” Step to build and populate SCD tables :
  - Insert / Update scenario.
  - Punch Through scenario.

8. ETL Subsystem : Change Data Capture (CDC)

- Whole Population versus Change Data Capture.
- Challenges on CDC.
- Supports of CDC on several RDBMS.
- SQL Server 2008’s CDC feature demonstration.

9. Automation

- Creating Kettle’s job and transformation to do whole or incremental population of dimension and fact tables.
- Creating and schedule batch script to execute Kettle’s job.

10. Multidimensional Data and OLAP using Mondrian / JPivot

- Mondrian Installation and Configuration.
- Creating schema with Pentaho Schema Workbench.
- View and Navigate our Dimensional Modelling Data as Cube in JPivot.