

Designing the Data Warehouse

Course Summary

Description

This course provides students with the skills necessary to design a successful data warehouse. It is based on the following Ralph Kimball book: [The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, Third Edition](#), Wiley, ISBN: 1118530802, published on July 1, 2013. The students each get this book plus a set of printed PowerPoint slides. The course also discusses additional industry-wide best practices concerning Data Warehouse Design. The class uses the free CA Erwin Community edition data modeling software. The customer should talk with the instructor before the course concerning which data modeling software to use.

Topics

- Introduction to Data Warehousing
- Data Warehouse Design Methodology
- Data Warehouse Components, Aggregations and Measurements
- Designing For Slowly Changing Data
- Data Warehouse Design Solutions
- Designing For Rapidly Changing Data
- Designs For Finance and Accounting
- Designing Complex Associations
- Aggregating Data at the Right Level
- Data Warehouse Design Review
- Handling Multiple Granularities
- Designing Factless Fact Tables
- Handling Multi-Valued Measurements
- Designing For Electronic Commerce
- Data Warehousing Design Mistakes
- Building the Data Warehouse

Audience

This course is targeted at technical staff, team leaders, project managers, and users who need to understand how to design a data warehouse.

Prerequisites

Students should have at least some experience with any relational database management system.

Duration

Three days

Designing the Data Warehouse

Course Outline

- I. Introduction to Data Warehousing**
 - A. Operational Systems (OLTP)
 - B. Analytical Processing (OLAP)
 - C. Data Warehousing Requirements
 - D. Data Warehousing Team Responsibilities
 - E. Data Warehousing Components
 - F. Data Warehouse Terminology
 - G. What Is Metadata?
 - H. What Is a Fact?
 - I. What Is a Dimension?
 - J. Dimensional Modeling Myths
 - K. Avoiding Common Pitfalls
- II. Data Warehouse Design Methodology**
 - A. Retail Sales Case Study
 - B. The Kimball 4 Step Design Process
 - C. Non-Additive Facts
 - D. Date Dimension
 - E. Product Dimension
 - F. Sales Amount and Quantity By Department Report
 - G. What Is a Drill Down?
 - H. What Is a Roll Up?
 - I. Location (Geographical or Store) Dimension
 - J. Promotion Dimension
 - K. What Is a Factless Fact Table?
 - L. What Is a Degenerate Dimension?
 - M. Star Schema Extensibility
 - N. What Is Snow Flaking?
 - O. Too Few or Too Many Dimensions
 - P. Surrogate Keys Versus Natural Keys
- III. Data Warehouse Components, Aggregations and Measurements**
 - A. Inventory Periodic Snapshot Fact Table
 - B. Semi-Additive Facts
 - C. Enhanced Inventory Facts
 - D. Inventory Transaction Fact Table
 - E. Inventory Accumulating Snapshot Fact Table
 - F. Shared Common Dimensions
 - G. The Data Warehouse Bus Matrix
 - H. What Are Conformed Dimensions?
 - I. What Are Conforming Roll-Up Dimensions?
 - J. What Are Conforming Dimension Subsets?
 - K. What Are Conformed Facts?
- IV. Designing For Slowly Changing Data**
 - A. Procurement Case Study
 - B. Procurement Transaction-Grained Star Schema
 - C. Multiple Versus Single Procurement Transaction-Grained Fact Tables
 - D. One Fact Table Per Major Feeder System
 - E. What Are Slowly Changing Dimensions?
 - F. Type 1 Slowly Changing Dimension Strategy
 - G. Type 2 Slowly Changing Dimension Strategy
 - H. Type 3 Slowly Changing Dimension Strategy
 - I. Type 4 Slowly Changing Dimension Strategy
 - J. Predictable Changes with Multiple Version Overlay Strategy
 - K. Unpredictable Changes with Single Version Overlay Strategy
- V. Data Warehouse Design Solutions**
 - A. Order Management Bus Matrix
 - B. Order Product Transaction Fact
 - C. Role-Playing Dimensions (Using Views on Dimensions)
 - D. Common Product Dimensions
 - E. Product Master to Dimension Mapping
 - F. Customer Ship-To and Bill-To Dimension
 - G. An Outrigger Address Dimension
 - H. Deal Dimension
 - I. Order Number Degenerate Dimension
 - J. Junk Dimension
 - K. Designing for Multiple Currencies
 - L. Allocating Facts to Lower Granularities
 - M. Shipment Invoice Line Item Fact
 - N. Order Fulfillment Pipeline
 - O. Order Fulfillment Accumulating Snapshot Fact
 - P. Designing for Multiple Units of Measure
 - Q. Three Types of Fact Table Comparisons
 - R. Introduction to Real-Time Data Warehouse Design

Designing the Data Warehouse

Course Outline (cont'd)

- VI. Designing For Rapidly Changing Data**
 - A. Customer Relationship Management Overview
 - B. Operational and Analytic CRM
 - C. Buy or Build Your CRM System
 - D. Customer Dimension
 - E. Name and Address Quality Issues
 - F. Customer Dimension Columns
 - G. Date Dimension Outtrigger
 - H. Customer Segmentation Columns
 - I. Adding a County Outtrigger Dimension
 - J. Designing Dimension Outtriggers
 - K. Large and Rapidly Changing Customer Dimension
 - L. Tuning Rapidly Changing Dimensions
 - M. Mini-Dimension Characteristics
 - N. Customer Mini-Dimension ERD
 - O. Variable-Width Dimension Columns
 - P. Fixed-Depth Customer Hierarchies
 - Q. Variable-Depth Customer Hierarchies
 - R. Parent Subsidiary Company Hierarchies
 - S. Parent Subsidiary Bridge Table
 - T. Parent Subsidiary ERD
 - U. Parent Subsidiary Revenue Select Statements
- VII. Designs For Finance and Accounting**
 - A. Finance and Accounting Data Warehouses
 - B. General Ledger Periodic Snapshot
 - C. General Ledger Periodic Snapshot Schema
 - D. General Ledger Journal Entry Transaction Fact
 - E. Annual Budget Schema
 - F. Budget Variance Schema
 - G. OLAP Analytic Solutions
- VIII. Designing Complex Associations**
 - A. Employee Transaction Schema
 - B. Employee Transaction Dimension
 - C. Employee Periodic Snapshot Schema
 - D. Audit Dimension
 - E. Employee Skill Outtrigger Dimension
 - F. Employee Skill Outtrigger Select Statement
 - G. Another Employee Skill Outtrigger Solution
- H. Another Employee Skill Outtrigger Select Statement**
- I. A Third Employee Skill Outtrigger Solution**
- J. Analyzing Survey Data with Star Schemas**
- IX. Aggregating Data at the Right Level**
 - A. Banking Data Warehouse Requirements
 - B. Monthly Account Balance Periodic Snapshot
 - C. Dimension Table Check List
 - D. Account Household Periodic Snapshot
 - E. The Fact Table as an Associative Entity
 - F. Customer Account Associative Bridge
 - G. More On Mini-Dimension Outtriggers
 - H. Arbitrary Banding Range Technique
 - I. How to Design a Band Definition Table
 - J. How to Track Point-In-Time Balances
 - K. Account Balance as of a Given Date Select Statement
 - L. Context-Dependent Outtriggers or Dimension Super-Type and Sub-Types
- X. Data Warehouse Design Review**
 - A. Telecommunications Data Warehouse Bus Matrix
 - B. How to Perform a Star Schema Design Review
 - C. Version 1 Customer Billing Star Schema
 - D. Problems with the Version 1 Customer Billing Star Schema
 - E. Version 2 Customer Billing Star Schema
 - F. Problems with the Version 2 Customer Billing Star Schema
 - G. Version 3 Customer Billing Star Schema
 - H. Geographic Location Dimension
 - I. Geographic Information Systems
 - J. Analyzing Geo-Spatial Data
- XI. Handling Multiple Granularities**
 - A. Airline Frequent-Flyer Case Study
 - B. Multiple Fact Granularities
 - C. Segment-Level Flight Activity
 - D. Segment and Trip-Level Granularity in One Fact Table
 - E. Cargo Shipper Schema
 - F. Travel Services Hotel Schema
 - G. Combining Small Dimensions Into One
 - H. Country-Specific Calendar Outtriggers

Designing the Data Warehouse

Course Outline (cont'd)

- I. Time of Day as a Dimension or a Fact?
 - J. Designing for Multiple Time Zones
- XII. Designing Factless Fact Tables**
- A. Admissions Tracking Accumulating Snapshot
 - B. Accumulating Snapshot Design for Moving Forward Through a Series of Events
 - C. Accumulating Snapshot Design for Events that Happen Out of Sequence
 - D. Accumulating Snapshot Design for Tracking Events that Back Up
 - E. Horizontal Versus Vertical Event Tracking
 - F. Factless Fact Table for Student Registration
 - G. Factless Fact Table for Facilities Utilization
 - H. Factless Fact Table for Tracking Events that May Not Happen
 - I. Factless Fact Table for Tracking Weighted Events
 - J. Factless Promotion Coverage Fact Table
 - K. Promoted Products that Did Not Sell Select Statement
- XIII. Handling Multi-Valued Measurements**
- A. What Is a Value Circle and Why Do Designers Need to Know About It?
 - B. Discovering Dimensions in the Value Circle
 - C. Health Care Billing Accumulating Snapshot
 - D. Designing the Date Dimension to Track Events that Have Not Happened Yet
 - E. Designing the Date Dimension for Events that Have Been Skipped
 - F. Dropping the Granularity of Accumulating Snapshots to Avoid the Multi-Valued Problem
 - G. Three Techniques for Eliminating Multiple Values within a Fact Table Column
 - H. Turning a Billing Fact into a Profitability Fact
 - I. Handling Complex Events with an Accumulating Snapshot
- J. Mining Free-Form Comments into Dimension Constraints and Facts
 - K. Handling Facts with Sparse Data Values
 - L. Handling Late Arriving Historical Dimension Data
- XIV. Designing For Electronic Commerce**
- A. What Is a Uniform Resource Locator (URL)?
 - B. Client Server Interactions On the Internet
 - C. Click Stream Characteristics
 - D. Identifying Web-Site Visitor Origin
 - E. Identifying the Session
 - F. Identifying the Visitor
 - G. Proxy Web Server Challenges
 - H. Click Stream Dimensions
 - I. Click Stream Session Tracking
 - J. Click Stream Web Page Tracking
 - K. Aggregated Click Stream Facts
 - L. Tracking E-Commerce Profitability
- XV. Data Warehousing Design Mistakes**
- A. Insurance Value Chain
 - B. Insurance Bus Matrix
 - C. Insurance Policy Transaction Schema
 - D. Automobile Transaction Schema
 - E. Policy Premium Periodic Snapshot Schema
 - F. Multi-Valued Policy Holder Insured Driver Bridge
 - G. Keeping Your Bus Matrix Up To Date
 - H. Tracking Policy Claims
 - I. Policy Claims Transactions
 - J. Policy Claims Transaction Schema
 - K. Policy Claims Periodic Snapshot
 - L. Policy Claims Accumulating Snapshot
 - M. Consolidating Policy Premium and Claim Amounts
 - N. Factless Fact Table Accident Events
 - O. Dimensional Modeling Design Mistakes
- XVI. Building the Data Warehouse**
- A. Data Warehouse Life Cycle Road Map
 - B. DW Project Planning and Management
 - C. DW Project Scoping
 - D. DW Project Justification
 - E. DW Project Business Staffing

Designing the Data Warehouse

Course Outline (cont'd)

- F.** DW Project Business or IT Staffing
- G.** DW Project IT Staffing
- H.** DW Requirements Planning
- I.** Collecting DW Requirements
- J.** DW Effective Interviewing Techniques
- K.** DW Interview Wrap-up
- L.** DW Post Interview Documentation

- M.** DW Requirements Prioritization and Consensus
- N.** DW Business Impact and Feasibility Prioritization of Requirements in Quadrants
- O.** DW Technical Architecture in 8 Steps
- P.** DW Tool Suite Selection and Installation

- Q.** DW Star Schema Physical Design Considerations
- R.** DW Star Schema Aggregation Strategies
- S.** DW Star Schema Indexing Strategies
- T.** DW Data Staging Physical Design Considerations
- U.** DW Dimension Table Staging Considerations
- V.** DW Master Dimension Cross-Referencing Strategies
- W.** DW Fact Table Staging
- X.** DW Analytics Specification
- Y.** DW Deployment
- Z.** DW Maintenance and Growth
- AA.** Ten Common DW Design Mistakes to Avoid