

Dimensional Modeling for the Data Warehouse

Course Summary

Description

This course provides students with the skills necessary to design a successful data warehouse using multi-dimensional data modeling techniques. 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 Dimensional Modeling and Star Schemas. 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

- Dimensional Modeling Primer
- Retail Sales
- Inventory
- Procurement
- Order Management
- Customer Relationship Management
- Accounting
- Human Resources Management
- Financial Services
- Telecommunications And Utilities
- Transportation
- Education
- Health Care
- Electronic Commerce
- Insurance
- 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 using multi-dimensional data modeling techniques.

Prerequisites

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

Duration

Four days

Dimensional Modeling for the Data Warehouse

Course Outline

- I. Dimensional Modeling Primer**
 - A. Operational Systems (OLTP)
 - B. Analytical Processing (OLAP)
 - C. Data Warehousing Requirements
 - D. Data Warehousing Team Responsibilities
 - E. Data Warehousing Components
 - 1. Operational Source Systems
 - 2. Data Staging Area
 - a. Extraction
 - b. Transformation
 - i. Cleaning
 - ii. Conforming
 - c. Loading
 - 3. Data Presentation Area
 - 4. Data Access Tooling
 - F. Data Warehouse Terminology
 - 1. Star Schema
 - 2. On-Line Analytical Processing (OLAP)
 - 3. Cubes
 - G. What Is Metadata?
 - 1. Staging Meta Data
 - 2. DBMS Meta Data
 - 3. Data Access Tooling Meta Data
 - H. What Is a Fact?
 - I. What Is a Dimension?
 - J. Dimensional Modeling Myths
 - K. Avoiding Common Pitfalls
- II. Retail Sales**
 - A. Retail Sales Case Study
 - B. The Kimball 4 Step Design Process
 - 1. Select the Business Process
 - 2. Declare the Granularity of the Fact Table
 - 3. Choose the Dimensions
 - 4. Identify the Facts
 - 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. Inventory**
 - 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. Procurement**
 - 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. Order Management**
 - A. Order Management Bus Matrix
 - B. Order Product Transaction Fact

Dimensional Modeling for the Data Warehouse

Course Outline (cont'd)

- 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
- VI. Customer Relationship Management**
- 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 Outrigger
 - H. Customer Segmentation Columns
 - I. Adding a County Outrigger Dimension
 - J. Designing Dimension Outriggers
 - 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. 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. Human Resources Management**
- A. Employee Transaction Schema
 - B. Employee Transaction Dimension
 - C. Employee Periodic Snapshot Schema
 - D. Audit Dimension
 - E. Employee Skill Outrigger Dimension
 - F. Employee Skill Outrigger Select Statement
 - G. Another Employee Skill Outrigger Solution
 - H. Another Employee Skill Outrigger Select Statement
 - I. A Third Employee Skill Outrigger Solution
 - J. Analyzing Survey Data with Star Schemas
- IX. Financial Services**
- 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

Dimensional Modeling for the Data Warehouse

Course Outline (cont'd)

- G. More On Mini-Dimension Outriggers
 - 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 Outriggers or Dimension Super-Type and Sub-Types
- X. Telecommunications And Utilities**
- 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. Transportation**
- 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 Outriggers
 - I. Time of Day as a Dimension or a Fact?
 - J. Designing for Multiple Time Zones
- XII. Education**
- 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. Health Care**
- 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

Dimensional Modeling for the Data Warehouse

Course Outline (cont'd)

- XIV. 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. Insurance**
- 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
 - 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
- XVII. WORKSHOPS [The Students Do Various Workshops Depending On How Many Days The Course Is Taught.]**
- A. Use the instructor supplied order and shipment star schema tables to do various drill downs and roll ups. [We use MS Access since it is readily available in most IT environments, not that you would ever implement a star schema with Access.]

Dimensional Modeling for the Data Warehouse

Course Outline (cont'd)

- B. Use Ralph Kimball's 4 step methodology to begin to develop a star schema for a business process identified at each student team. Each team identifies a fact table, defines the granularity of the fact table, chooses dimensions, and defines additive, semi-additive, or non-additive numeric facts. Each team may do the same or different star schemas as other teams. Each team writes up their star schema requirements with MS Word and implements it with the data modeling software chosen for the class. Each team presents their star schema to the class.
- C. Each team modifies their star schemas based upon the comments that the class has given them.
- D. The class is shown how to do a drill across using the instructor supplied order and shipment star schemas developed in Microsoft Access. The students do their own drill across in Access.
- E. If the students bring real star schemas to class, the class may do a design review on the selected star schemas.