



Dimensional Modeling in Depth

Duration: 4 days

Why attend: Excellence in dimensional modeling remains the keystone of a well designed data warehouse. This course gives you the opportunity to learn directly from the industry's dimensional modeling leaders, Ralph Kimball and Margy Ross.

You will learn practical dimensional modeling techniques covering basic to advanced issues. Following the tradition of the *Data Warehouse Toolkit*, all the techniques in this class are motivated by specific industry scenarios. The goal is for students to gain an in-depth understanding of dimensional modeling so they can confidently apply the concepts in their workplace following the training. Every attendee will receive a copy of *The Data Warehouse Toolkit, 2nd Edition*, co-authored by Ralph Kimball and Margy Ross.

Who should attend: The course is designed for data warehouse architects, data modelers, database administrators, application developers and system designers. It's appropriate for anyone interested in an A to Z coverage of dimensional modeling.

Course Outline:

DAY 1

Dimensional Modeling Fundamentals

- *Publishing responsibilities of DW/BI professionals*
- *Role of dimensional modeling in Kimball versus Corporate Information Factory architectures*
- *Fact and dimension table characteristics*
- *Surrogate key recommendations*
- *Fact table granularity*
- *Dimensional modeling fables and myths*

Retail Sales Case Study as Class Design

- *4-step design process*
- *Denormalized dimension table hierarchies*
- *Degenerate dimensions*
- *Dimension role-playing*
- *Date and time-of-day dimension considerations*
- *Centipede fact tables with too many dimensions*
- *Gracefully extending an existing dimensional model*
- *Star versus snowflake schemas*
- *Factless fact tables*

Order Management Design Workshop as Small Group Exercise

- *Complications with operational header/line data*
- *Allocated facts*
- *Abstract, generic dimensions*
- *Freeform text comments*
- *Junk dimensions for miscellaneous transaction indicators*
- *Multiple currencies*

DAY 2

Inventory Case Study as Class Design

- *Value chain implications*
- *Semi-additive facts*
- *Three fundamental types of fact tables (transaction, periodic snapshot and accumulating snapshot)*
- *Conformed dimensions*
- *Enterprise Data Warehouse Bus Architecture and matrix to integrate dimensional models*
- *Drilling across fact tables*
- *Individual exercise: Translate requirements into bus matrix*
- *Consolidated cross-process fact tables*

Billing Design Review as Individual Exercise

- *Common design flaws*
- *Checklist for conducting design reviews*

Slowly Changing Dimensions

- *Basic Type 1, 2 and 3*
- *Advanced hybrid techniques for dealing with a series of predictable and unpredictable changes*
- *Mini-dimensions for rapidly changing large dimensions*

Credit Card Design Workshop as Small Group Exercise

- *Complementary transaction and periodic snapshot schemas*
- *Design considerations for one dimension versus two dimensions*
- *Fact table normalization*

Insurance Case Study as Class Design

- *Review of earlier design patterns and techniques*
- *Development of bus matrix from extended case study*
- *Communicating dimensional models to users*
- *Further recommendations regarding modeling process activities*
- *Detailed implementation bus matrix*

DAY 3 – FINANCIAL APPLICATIONS

Automobile Options Case Study as Class Design

- *Trading off columns versus rows*
- *Impact on user interface design and application scalability*

Profit Equation

- *Starting with revenue, then bringing costs to same grain*
- *What to do when your business refuses to allocate*

- *Tracking allocation metadata*
- *Profit margin point analysis*
- *Profit margin value banding*

General Ledger

- *Cleanest schema in your data warehouse*
- *Non-conforming dimensions from the general ledger*
- *Tracking instantaneous balances across all time*
- *Why not to store year-to-date, what to do instead*
- *Drilling down in the general ledger all the way to a document*

Budgeting Value Chain

- *Budgets, commitments and expenditures*
- *Ragged hierarchies for financial reporting*
- *Bridge tables for ragged hierarchies*
- *Shared ownership in financial rollups*
- *Time varying ragged hierarchies*
- *Techniques for modifying ragged hierarchies*
- *Rolling up the value chain through a ragged hierarchy*

Specific Financial Application Challenges

- *Tracking the “age of the book”*
- *Calculating the “policy loss triangle” in insurance*

Retail Bank Account Tracking as Small Group Exercise

- *Serving the need for householding all possible account types and full account detail with 100's of facts*
- *Many-to-many account to customer map*
- *General many valued dimensions*
- *Very rapidly changing account demographics and status*
- *Correctly weighted and “impact” reports by individual customer*
- *Tagging an account as “about to go bankrupt”*
- *Super-types and sub-types in financial applications*

Compliance Enabled Data Warehouses

- *Eliminating Type 1 and Type 3 updates*
- *Accessing all prior versions of a database at points in time*
- *Protecting the custody of your data*
- *Showing why and when changes to data occurred*

Dimensional Designs in the ETL Back Room

- *Tracking data quality with error event fact table (brief overview)*
- *Column, structure, and business rule tests for data quality*
- *Reporting data quality with audit dimension*

DAY 4 – CUSTOMER BEHAVIOR APPLICATIONS

Customer Relationship Management Payoffs Class Discussion

- *What do our end users expect from a CRM system?*
- *How do CRM results impact the bottom line?*
- *What data sources are needed to support CRM?*
- *What data quality and integration problems are common?*
- *Where are real-time CRM solutions required? What is real-time?*

Capturing Complex Customer Behavior

- *Building study groups from existing reports*
- *Attaching study group tables to all customer facing applications*
- *Combining study groups with union, intersection, set difference*
- *Sequential time dependent study groups*
- *Applying study groups to marketing panel studies*
- *Applying study groups to medical outcomes analysis*

Building Visual Basic (or similar) Custom User Interfaces

- *Car option selection, value band definition, study group creation*

Typical Customer Dimension Modeling Challenges

- *Hundreds or even thousands of demographic attributes*
- *Sparse but wide demographics coverage*
- *Implicit time spans defined by all types of transactions*
- *Finding detailed customer status at random times in the past*
- *Tricky time span queries made simple*
- *Multiple and growing lists of names in complex customer profile*
- *Customer satisfaction dimensions: causal dimensions*
- *When is something both a dimension and a fact?*
- *Relationship between a prospect and a customer*
- *Maintaining customer identity after aggressive de-duplication*

Real Time Customer Tracking (brief overview)

- *Hot partition*
- *How to handle unresolved customer identities in real time*

Modeling Sequential Behavior

- *Step dimension for describing sequential behavior*
- *RFID and web page challenges (brief discussion)*
- *Link exposure data sets: 10 terabytes per day*
- *Modeling and querying product purchase sequences*

Text Facts to Describe Cluster Evolution

- *Building text facts with cluster identification data mining tool*

Final Customer Topics

- *Modeling very complex events involving many parties such as automobile accidents and complex surgical procedures*
- *Structured questionnaire designs*