



## Data Warehouse Lifecycle in Depth

**Duration:** 4 days

**Why attend:** The data warehouse/business intelligence (DW/BI) system continues to be one of the most organizationally complex and technically interesting IT projects. This course prepares you to successfully implement your DW/BI environment.

Taught by Margy Ross and Warren Thornthwaite, co-authors of the best selling *Data Warehouse Lifecycle Toolkit, Second Edition*, this course is packed with specific techniques, guidance and advice from initial project planning through deployment and maintenance.

**Who should attend:** We've designed this course to appeal to all major roles on a DW/BI project, from project managers to DBAs to data modelers to application developers. It's for anyone new to data warehousing who wants to learn how to do it, or who has been through a project or two and wants to learn how to do it right.

**Course Outline:** **DAY 1**

### **Introduction to the Kimball Lifecycle Approach**

- *Major task sequencing and roadmap*

### **Program/Project Planning and Management**

- *Readiness factors*
- *Risk assessment and mitigation plans*
- *Scoping and business justification*
- *Team roles and responsibilities*
- *Project plan development and maintenance*
- *Program management*

### **Business Requirements Definition**

- *Program versus project requirements preparation*
- *Requirements gathering participants*
- *Techniques for gathering requirements and handling obstacles*
- *Program/project requirements deliverables*
- *Requirements prioritization*

### **Dimensional Modeling**

- *Role of dimensional modeling in Kimball versus Corporate Information Factory architectures*
- *Fact and dimension table characteristics*
- *4-step process for designing dimensional models*
- *Denormalizing dimension table hierarchies*
- *Degenerate dimensions*

- *Surrogate key recommendations for dimension tables*
- *Star versus snowflake schemas*
- *Centipede fact tables with too many dimensions*
- *Factless fact tables*
- *Additive, semi-additive, and non-additive facts*
- *Value chain implications on data architecture*
- *Data Warehouse Bus Architecture and matrix for integration*
- *Conformed dimensions – identical versus shrunken roll-ups*
- *Exercise: Translate requirements into bus matrix*

## **DAY 2**

### **Dimensional Modeling continued**

- *Slowly changing dimensions – type 1, 2, 3 and advanced hybrid techniques*
- *Mini-dimensions for large, rapidly changing dimensions*
- *Dimension table role-playing*
- *Exercise: Small group design case study workshop #1*
- *Dealing with facts at different header/line levels of detail*
- *Multiple currencies*
- *Junk dimensions for miscellaneous attributes*
- *Exercise: Design review to identify common flaws*
- *Exercise: Design enhancements to embellish existing design*
- *Dimensional modeling process flow, tasks, and deliverables*
- *Exercise: Small group design case study workshop #2*
- *Transaction versus periodic snapshot versus accumulating snapshot fact tables*
- *Exercise: Convert E-R model into dimensional model*

### **Mature DW/BI System Check-ups**

- *Symptoms of sponsorship, data, infrastructure, acceptance, and organizational/cultural disorders*
- *Prescribed treatment plans for common maturity problems*

## **DAY 3**

### **Technical Architecture Design**

- *Architecture concepts*
- *Topology options: independent data marts, enterprise data warehouse, and the conformed data warehouse*
- *Common components and functionality*
  - *ETL system*
  - *Presentation servers (RDBMS/OLAP)*
  - *Real time options: direct to source, ODS, real time layer*
  - *BI application types and services*
- *Creating the architecture plan*
- *Exercise: Translating requirements into architecture implications*

### **Product Selection and Installation**

- *Architecture-based evaluation approach and matrices*
- *Infrastructure considerations*
- *Metadata management*
- *Securing the system*

### **Physical Database Design**

- *Standards and naming conventions*
- *Physical model development*
- *Aggregation, indexing and storage plans*
- *Usage monitoring*

### **Extract, Transformation and Load**

- *ETL planning*
  - *Data profiling*
  - *High-level and detailed ETL schematics*
  - *Quality assurance and data validation system*
  - *Warehouse operations system*
- *ETL workflow*
  - *Extract to create, filter and transfer source data*
  - *Cleaning and conforming dimensions and facts*
  - *Exercise: Processing slowing changing dimensions type 2*

## **DAY 4**

### **Extract, Transformation and Load continued**

- *ETL workflow continued*
  - *Preparing and delivering dimensions and facts*
  - *Data integration and master data management*
  - *Dealing with data quality issues*
- *Aggregate management*
- *Load cycle management*
- *Exercise: “High-level ETL schematic” case study*

### **BI Applications**

- *BI application types (ad hoc, standard reporting, analytic applications, dashboards) and audiences*
- *Specification of templates, applications and navigation framework*
- *Development of applications and BI portal*

### **DW/BI System Deployment and Support**

- *System deployment*
- *Communication and documentation*
- *Training and support*
- *On-going user, data and system maintenance*

### **DW/BI System Growth**

- *Planning for growth*