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

Who Should Attend
This course is targeted at technical staff, team leaders and project managers who need to understand how to design a data warehouse using multi-dimensional data modeling techniques.

Course 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 Ralph Kimball’s book *The Data Warehouse Toolkit, Second Edition*, Wiley, ISBN: 0471200247, published in April, 2002 with additional topics and commentary from Daniel J Hall. The class uses either DB Designer V4 (free), MS Access, or ER/win as the data modeling software tool. Other data modeling tools may be used if you check with our instructor first.

Course 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
I. DIMENSIONAL MODELING PRIMER
   A. This chapter contrasts operational systems with data warehouses. It also compares OLTP with OLAP systems. It lists requirements that all data warehouses should fulfill. It lists the responsibilities of the data warehouse team. It discusses the four components of a data warehouse: the operational source systems, the data staging area, the data presentation area, and data access tools. It defines and describes ETL, star schemas, multi-dimensional data bases, and cubes. It defines metadata and the various categories of metadata. It defines fact and dimension tables and gives some examples. It discusses 5 dimensional modeling myths and lists 10 common dimensional modeling pitfalls to avoid.
   B. Each chapter in this course discusses different dimensional modeling situations and techniques using different industries to illustrate the techniques. The techniques, however, apply to all industries.

II. RETAIL SALES
   A. In this chapter the student learns about Ralph Kimball’s four-step dimensional modeling methodology. Transaction-level fact tables are discussed. The difference between additive and non-additive facts is illustrated. Causal dimensions like tracking a sales promotion are discussed. The reasons for having Degenerate Dimensions (DDs) like a purchase order number or sales transaction number in the fact tables are presented. The student is shown how relatively straight forward it is to add new dimension tables and facts to a star schema. The student is shown why snowflaking dimension attributes is not a good idea. The student is given several rules about how to avoid having too many or too few dimensions in their star schema. The student is shown why surrogate keys rather than natural keys should be used to link the fact table to the dimension tables.

III. INVENTORY
   A. In this chapter the student learns about 3 types of fact table: the periodic snapshot fact table, the inventory transaction fact, and the accumulating snapshot fact. How and when to use these different types of fact tables is discussed. Next, semi-additive facts are defined and illustrated. The student is introduced to the data warehouse bus architecture and shown why dimension tables should be shared by as many fact tables as possible. The student is shown why the data warehouse bus matrix is such an important design document. Examples of conformed dimension and fact tables are also shown to the students.
IV. PROCUREMENT
A. In this chapter, the student learns how to determine if there should be one fact table in a data mart fed by multiple source systems or whether there should be multiple fact tables fed by one source system per fact table. The strengths and weaknesses of each approach are discussed. Next, strategies are presented as to how to handle Slowly Changing Dimension (SCD) table attributes. The SCD type 1 approach overwrites the old values in the dimension table. The SCD type 2 approach saves the old dimension table row and adds a new row containing the updated values. The SCD type 3 approach adds an historical column to the dimension table. Two hybrid design approaches are also discussed in this chapter. The Dan Hall SCD/RCD type 4 approach to handling dimension table changes is discussed at the end of this chapter.

V. ORDER MANAGEMENT
A. An e-commerce order entry star schema is shown in this chapter. The student is shown how to accomplish date dimension role-playing where there is one date dimension table with different views built on the date dimension table for different date roles such as ship date and bill date. The student is shown how to transform the product master ODS into a product dimension table. The student is shown why the operational order header table should not become an order header dimension table. Where to place switches and indicators is discussed next. The strengths and weaknesses of placing indicators and switches in the fact table or in various types of “Junk” dimensions are discussed. How to handle multiple currencies and units of measure is discussed. How to allocate facts from a fact table at a higher granularity to a fact table at a lower granularity is discussed. The chapter finishes by discussing how to handle real-time data warehouse inquiries.

VI. CUSTOMER RELATIONSHIP MANAGEMENT
A. The CRM star schema is discussed in this chapter. The customer dimension table is one of the most challenging dimension tables in a star schema since it usually has millions of rows, many columns, and the data in it changes rapidly. Examples of well designed and not so well designed customer dimension tables are presented to the student. Dimension table outriggers are discussed next. This is an acceptable form of snowflaking. The student is shown how to handle Rapidly Changing Dimensions (RCDs) using the customer dimension as an example. RCDs can very often be handled through the use of mini-dimensions employing a technique known as banding. The chapter ends by showing the student how to handle customer hierarchies that are fixed-depth and variable-depth.

VII. ACCOUNTING
A. This chapter discusses general ledger periodic snapshot and transaction fact tables. Annual budget and budget variance schemas are discussed. The chapter ends with a discussion of the strengths and weaknesses of using data marts supplied with purchased financial software versus building your own financial data mart and various combinations of both.
VIII. HUMAN RESOURCES MANAGEMENT
A. This chapter discusses an employee transaction fact table and a human resource periodic snapshot fact table. This chapter introduces an audit dimension table to any star schema to track and evaluate the various ETL processes. This chapter also discusses various design techniques as to how to handle job skill categorization and searching using employee and job skill dimensions. This chapter ends by showing how to handle survey questionnaire data using a star schema.

IX. FINANCIAL SERVICES
A. This chapter introduces a dimension table checklist that can be used to verify that you have a minimum number of dimension tables. This chapter shows how to handle a many-to-many mapping between the customer and account dimension tables through the use of an associative bridge table. It also discusses how and why to use weighting factors in the bridge table. Next, the student is shown how to use a band definition table in order to make the banding more generalized in a mini-dimension. Finally, the student is shown how to code SQL that allows point-in-time balances to be displayed.

X. TELECOMMUNICATIONS AND UTILITIES
A. This chapter is very important since it shows the student how to properly do a star schema design review. It highlights common design mistakes to look for in the design review. The students conduct a design review of a badly designed star schema and then suggest design changes based on what they have learned so far. The students then do a workshop where they conduct design reviews of their own in-house star schemas. They present their findings and design improvement recommendations to the class.

XI. TRANSPORTATION
A. This chapter presents more design considerations when designing fact tables at different levels of granularity. Star schemas from the transportation and shipping industries are used as the examples in this chapter. The student is shown how to combine various role-playing dimensions into a super-dimension. The student is shown how to collapse smaller related dimensions into fewer dimensions. Next, the student is shown how to do country-specific date dimension outrigging. Finally, the student is shown how to properly handle a time dimension and multiple time zones.

XII. EDUCATION
A. This chapter shows how to handle admissions or hiring tracking by using an applicant accumulating snapshot fact table. It shows how to design factless fact tables by analyzing several different factless fact tables: a student registration event factless fact table, a building utilization factless fact table, a student attendance factless fact table, and how to analyze whether items that were on sale actually did sell by using a sales promotion factless fact table.
XIII. HEALTH CARE

A. In this chapter, the student learns about the health care value circle with a patient treatment fact table at the center of the star schema surrounded by various health care provider dimension tables. A health care billing line item accumulating snapshot fact table is studied. The student is shown how to handle various default values, not applicable, missing, and null data for the date dimension. The student is shown how to properly join multiple date dimension role-playing views to get the correct results. In fact, the students do a workshop where they are shown that database engines treat views as if they were separate tables. The student is shown how to handle multi-valued diagnoses through the use of diagnosis groups and a bridge table. This is generalized to show how to handle multi-valued data wherever it occurs. The student is also shown how to handle fixed versus variable length text data. The student is shown how to handle unique types of data such as: laboratory test results, graphs, charts, and photographs as well as how to design for sparse data. Finally, the student is shown how to handle late-arriving fact data rows that missed the ETL fact table load processes.

XIV. ELECTRONIC COMMERCE

A. This chapter is a high-level summary of the Data Webhouse Toolkit by Ralph Kimball and Richard Merz (Wiley, 2000). It explains at a high level how web browsers interact with web servers. This chapter defines the click stream and how the various data connected with it is captured for a data web-house. It describes how visitors to a web site are identified. It describes how data is stored in cookies and session variables. It identifies the usual click stream dimensions found in a data web-house. It looks at web page, session, and application fact table granularities and facts. The chapter finishes by studying a star schema that can analyze e-commerce profitability.

XV. INSURANCE

A. This chapter introduces the student to the insurance value chain and its star schema bus matrix. The bus matrix includes a policy transaction fact table, a policy premium accumulating snapshot fact table, a policy claims accumulating snapshot fact table, and a consolidated policy/claims fact table. The student sees how to handle multi-valued insured drivers on one policy premium snapshot fact table through the use of a bridge table with weighting factors. The student sees how to handle accident events by using a factless fact table. The chapter ends with a discussion of frequently made dimensional modeling mistakes.
XVI. BUILDING THE DATA WAREHOUSE

A. This chapter discusses Ralph Kimball’s methodology for building the data warehouse. It is a high-level summary of his detailed methodology discussed in his book: *The Data Warehouse Lifecycle Toolkit*, by Ralph Kimball, Laura Reeves, Margy Ross, and Warren Thornthwaite (Wiley, 1998). The chapter begins by showing a diagram of Kimball’s data warehouse life cycle road map. It then discusses data warehouse project planning and management issues, scoping the project, justification, staffing, requirements, interviewing, documentation, prioritization, and building consensus. It shows the student Kimball’s 8 step process for producing the data warehouse technical architecture.