(BI) Designing a Data Warehouse

Course Description:

This course is aimed at data warehouse and business intelligence designers, implementers and managers, but will also serve as a good basis for business and data analysts who will be involved in working with data warehouses and business intelligence deployments. All individuals who will be involved in some element of working with a data warehouse will benefit from gaining an understanding of the dimension modelling concepts that influence the final design of a data warehouse.

Delegates will be introduced to the basic terminology and concepts of a Kimball based Data Warehouse architecture. The course covers Dimension Modelling Techniques introducing the basics of Fact Tables, Dimension Tables, Dimension Hierarchies and more.

Case studies are used throughout to demonstrate the concepts taught and to introduce the need for some of the more advanced modelling techniques such as Factless Tables, Enterprise Data Warehouse Bus Architecture, handling Slowly Changing Dimensions and Snapshots, Bridge Tables, Recursive Hierarchies.

As well as introducing the concepts and terminology of the data warehouse design delegates will learn about the design cycle and processes, the ETL requirements and processes and ensuring that the model is fit for both current and future business requirements.

This course is generic, but use of SQL Server and its business intelligence service and tools (DBMS, SSIS, SSAS and SSRS) will allow for some hands on elements of data warehouse design to be carried out to re-enforce concepts.

This course can be followed by the SQL Server Business Intelligence courses to address the implementation of a data warehouse, multidimensional cube and reporting deployment in a SQL Server environment: (SSAS) SQL Server Analysis Services, (SSIS) SQL Server Integration Services, (SSRS) SQL Server Reporting Services.

Pre-requisites:

Delegates should have an understanding of relational database design and be familiar with a database management system such as SQL Server or Oracle.

Course Content:
Course Outline

• Introduction to Data Warehousing, Business Intelligence & Dimension Modelling
  • Data Capture & Data Analysis
  • Goals of Data Warehousing & BI
  • Data Warehouses & Data Marts
  • Kimball’s Data Warehouse/BI Architecture
  • Alternative Data Warehouse/BI Architecture
  • Star Schemas Versus OLAP Cubes
  • OLAP Deployment Considerations
  • Extract Transform Load (ETL)
  • BI Applications

• Kimball Dimensional Modelling Techniques
  • Fundamental Concepts
  • Four Step Design Process
  • Basic Fact Tables
  • Dimension Tables
  • Calendar Dimensions
  • Role Playing Dimensions
  • Snowflake Dimensions
  • Conformed Dimensions
  • Slowly Changing Dimensions
  • Dimension Hierarchies
  • Advanced Fact Table Techniques
  • Advanced Dimension Table Techniques
  • Special Purpose Schemas
Course Outline

- **Dimensional Design Process**
  - **Part 1: Transactional Data**
  - Retail Case Study
  - Four Step Process for Designing Dimensional Models
  - Nulls in dimensional model
  - Snowflake Dimension Attributes
  - Degenerate Dimensions
  - Fact Table Granularity
  - Transaction Fact Tables
  - Factless Fact Tables
  - Additive, Non- Additive and Derived Facts
  - Dimension & Fact Table Keys
  - Resisting Normalisation

- **Dimensional Design Process**
  - **Part 2: Inventory Data**
  - Inventory Case Study
  - Chain of Key Business Processes
  - Semi- additive Facts
  - Periodic Snapshots, Transaction, and Accumulating Snapshots
  - Enterprise Data Warehouse Bus Architecture
  - Enterprise Data Warehouse Bus Matrix

- **Dimensional Design Process**
  - **Part 3: Changing Data**
  - Procurement Case Study
  - Slowly Changing Dimensions
  - Types 0 to 7 SCD Techniques
  - Blended Versus Separate Transaction Schemas
Course Outline

• **Dimensional Design Process**  
  Part 4: Order Management  
  • Order Management Case Study  
  • Bus Matrix for Order Management Processes  
  • Fact Table Normalisation Considerations  
  • Role Playing Dimensions  
  • Factors to Determine Single or Multiple Dimensions  
  • Junk Dimensions  
  • More on Degenerate Dimensions  
  • Multiple Currencies  
  • Handling Facts with Different Granularities  
  • Audit Dimensions  
  • Accumulating Snapshot Schema  

• **Dimensional Design Process**  
  Part 5: Financial Data  
  • Accounting Case Study  
  • Chart of Accounts  
  • Year to Date Facts  
  • Fiscal Calendars  
  • Drilling Down through Multi-ledger Hierarchy  
  • Consolidated Fact Tables  
  • Role of OLAP and Analytic Financial Solutions  

• **Dimensional Design Process**  
  Part 6: HR Data  
  • Human Resources Case Study  
  • Employee Profile Tracking  
  • Headcount Periodic Snapshot  
  • Recursive Employee Hierarchies
Course Outline

- Kimball Data Warehouse/BI Lifecycle
  - Project Planning & Management
  - Needs Analysis: Collecting Business Requirements
  - Developing a Technical Architecture
  - BI Application Design & Development: Dimensional Modelling, Physical Design
  - Physical Storage
  - ETL Design & Development
  - Prototype
  - Deployment, Maintenance & Growth

- Dimensional Modelling Process & Tasks
  - Preparation
  - High Level Dimensional Model
  - Detailed Dimensional Model Development
  - Model Review & Validation
  - Final Design Documentation: Description, Data Model Diagram, Detailed Dimension Design Worksheet

- ETL Subsystems & Techniques
  - Requirements & Constraints to be Considered: Business Needs, Compliance, Data Quality, Security, Integration, Latency, Archiving
  - Data Profiling
  - Extracting Data From Source Systems
  - Cleansing and Conforming Source Data: Quality Screens, Error Events, Auditing, Deduplication
  - Dimensional Structures to Monitor Quality Errors
  - Deliver Data to Dimensional Structures
  - Implement Slowly Changing Dimensions
  - Manage Production ETL Environment: Reliability, Scheduling, Backup, Recovery & Restart, Version Control, Workflow Monitor, Lineage & Dependency, Problem Escalation, Security, Compliance
ETL System Design & Development Process

- Planning & Design: High Level Plan, Choose a Tool, Default Strategies, Documentation
- One Time Historic Data Load: Populate Dimension Tables, Load Fact Tables
- Incremental Load Processing
- Real Time Data Warehousing Considerations

Course Duration: 3 Days