

Designing Effective BI Systems

4 Days (BI-DSNEFCT-201-EN)

Description

Building a great BI system is not just about being able to use the software; it's also about getting the overall design right. This course is essentially vendor neutral; it specifically doesn't cover how to drive any particular software. Instead it looks at the complete design process for a BI system, focusing on the architectural challenges and design issues that arise when building business intelligence solutions. For example it examines how we collect the user's analytical requirements and convert them first into a logical model and then on to a star schema.

Dimensional modeling is examined in detail and compared and contrasted with relational modeling. This enables attendees, for example, to evaluate the pros and cons of the Inmon and Kimball models of warehouse design.

The course then looks at the challenges of data cleansing and effective ETL design. Finally it takes a look at how data mining can be used effectively.

Target Audience

This course is intended for people who want to understand how BI systems are designed as a whole. They may be people who already have some experience in one aspect of BI (perhaps of working in an area like ETL) and who want to move upwards into BI system design and ultimately into the role of BI architect. The course is also highly suitable for people who have recently been appointed to the role of BI technical project manager.

Prerequisites

Before attending this course, it is recommended that students have the following skills:

- A good understanding of relational databases and relational database design
- A good understanding of SQL possibly from attending classes such as

Introduction to Transact-SQL for SQL Server 2000 & 2005 or T-SQL Fundamentals

Course Objectives

Upon completion of this course, the student will be able to:

- Evaluate the costs and benefits of a BI system for the business
- Understand the difference between relational and multi-dimensional systems and how each should be used
- Turn analytical requirements into a model for the system
- Understand dimensional modeling, design and implementation
- Select the appropriate data structure for the warehouse
- Understand the appropriate use of data mining
- Design effective BI systems

Course Summary Outline

Section A: Introduction to Data Warehousing and BI

Module 01: Introduction to Data Warehousing & BI

- Decision makers need information that is intuitively easy to understand. There are two main problems to delivering this. The information is buried somewhere in a mass of data and, worse, that data is dispersed across multiple OLTP systems. The solution is (ETL). Extract the data from each system, Transform it into a consistent format, and Load it into a central repository (the data warehouse). Once there we can restructure the data to deliver intuitive analysis.

Module 02: Introduction to the Process of Building a BI System

- Overview of creating a BI system -- requirements gathering, modeling, defining extraction logic, data transformation. Overview of an operational BI system -- data processing, data loading, data maintenance, data delivery.

Module 03: Starting and Structuring the Project

- Prerequisites, resourcing, identifying the skills you need on the BI team, roles and responsibilities, funding techniques, project principles.

Section B: Dimensional Modeling, Design and Implementation

Module 04: Overview of Modeling a BI System

- Modeling for OLTP. The user model, how users think about business processes. The logical model, ER Modeling. The physical model. Modeling for OLAP. The user model, how users think about information. The logical model – Sun modeling. The physical model, star schema.

Module 05: Requirement Analysis – User model to Logical model

- The meaning of data, design approach, user involvement, interviews,

requirements definition workshops, turning the user view (graphs, grids, reports) into a logical model (Sun model), questions to ask, defining scope, data audit.

LAB 05: Turning requirements into a logical model

Module 06: Physical Modeling – Star Schemas

- Dimensions, measures, star schemas. The fact table, granularity, fields in the fact table, additive and semi-additive fields. Dimension tables, time dimensions, hierarchical and non-hierarchical attributes, conforming and the nonconforming dimensions, the UDM.

LAB 06: Turning the logical model into a physical model

Module 07: Further Dimensional Design

- Surrogate keys, slowly changing dimensions (types 0 to 4), mini dimensions, snowflaking, degenerate dimensions, aggregations, Bridge tables, KPIs. Relationships - one to one, one to many, many to many. Ragged and unbalanced hierarchies, step dimensions.

Section C: BI Architecture

Module 08: OLAP and the EDW

- The difference between ROLAP, MOLAP and HOLAP. Indexing ROLAP structures. Compare and contrast OnLine Analytical Processing with the Enterprise Data Warehouse.

Module 09: Data Warehouse Architectures

- The pros and cons of various BI architectures and outlines the business constraints (rather than the dogma) which should be used in making a choice. This will include discussion of Inmon vs. Kimball which should not be seen as a binary choice -- relational and dimensional structures should be seen more as opposite ends of a spectrum.

Section D: ETL

Module 10: Extraction

- Prerequisites, common sources of data, the importance of the food chain, data quality critical success factors, data quality assessments, the data quality map, extraction tools, other considerations. Avoiding common mistakes.

Module 11: Transformation

- Types of transformations, merging, the duplication and survivorship, new key overlay, time date stamping, recoding, data type conversion, normalization, transformation and derivation, householding, cleansing. Transformation and

conditioning tools, when to transform.

Module 12: Data Loading

- Optimizing data loading. Snapshots versus transactions, preloaded administration, pre-sorting, load initialization, know your RDBMS, post integrity checks, database reorganization.

Section E: Data Mining and Presentation

Module 13: Data Mining and Exploration

- Finding hidden information in data. Data mining versus querying. Data mining methodologies. The data mining process -- data, information, decision. Data mining application areas, data mining techniques -- clustering, classification, value prediction, association discovery, sequential pattern discovery, time sequence discovery. Result validation, solution deployment.

Module 14: Delivering Information

- Reporting, OLAP browsing. Tool types -- raw SQL, canned query tools, report generators, GIS systems, end user training, MDX.

Module 15: Summary

- We take a look at some of the common mistakes