



MASTERING DATA WAREHOUSE AGGREGATES HDT807 Three Days

Prerequisites

Students should have at least some experience with any relational database management system and some experience using star schemas.

Who Should Attend

This course is targeted at software developers, data modelers, data base analysts, and any other technical personnel who need to understand how to plan, design, and build star schema aggregates.

Course Description

This course provides students with the skills necessary to plan, design, and build star schema aggregates. It is based on the Christopher Adamson book Mastering Data Warehouse Aggregates published in 2006 by Wiley Publishing, Inc, ISBN: 0-471-77709-9. The book contains examples of Materialized Query Tables (MQTs) from DB2 and Materialized Views from Oracle.

Course Topics

- Fundamentals Of Aggregates
- Choosing Aggregates
- Designing Aggregates
- Using Aggregates
- ETL Part 1: Incorporating Aggregates
- ETL Part 2: Loading Aggregates
- Aggregates And Your Projects
- Advanced Aggregate Design
- Related Topics



MASTERING DATA WAREHOUSE AGGREGATES HDT807

- I. Fundamentals of Aggregates**
 - A. Star Schema Basics**
 - 1. Operational Systems and the Data Warehouse
 - 2. Facts and Dimensions
 - 3. The Star Schema
 - 4. Using the Star Schema
 - 5. Multiple Stars and Conformance
 - 6. Data Warehouse Architecture
 - B. Invisible Aggregates**
 - 1. Improving Performance
 - 2. The Base Schema and the Aggregate Schema
 - 3. The Aggregate Navigator
 - 4. Principles of Aggregation
 - 5. Other Types of Summarization
 - C. Summary**
- II. Choosing Aggregates**
 - A. What Is a Potential Aggregate**
 - 1. Aggregate Fact Tables: A Question of Grain
 - 2. Aggregate Dimensions Must Conform
 - 3. Pre-Joined Aggregates Have Grain Too
 - 4. Enumerating Potential Aggregates
 - B. Identifying Potentially Useful Aggregates**
 - 1. Drawing on Initial Design
 - 2. Where Subject Areas Meet
 - 3. Query Patterns of an Existing System
 - C. Assessing the Value of Potential Aggregates**
 - 1. Number of Aggregates
 - 2. How Many Rows Are Summarized
 - 3. Who Will Benefit From the Aggregate
 - D. Summary**



MASTERING DATA WAREHOUSE AGGREGATES HDT807

III. Designing Aggregates

A. The Base Schema

1. Identification of Grain
2. Conformance Bus
3. Rollup Dimensions
4. Hierarchies
5. Housekeeping Columns

B. Design Principles for the Aggregate Schema

1. A Separate Star for Each Aggregation
2. Naming Conventions
3. Aggregate Dimension Design
4. Aggregate Fact Table Design
5. Pre-Joined Aggregate Design

C. Documenting the Aggregate Schema

1. Identify Schema Families
2. Identify Dimensional Conformance
3. Documenting Aggregate Dimension Tables
4. Documenting Aggregate Fact Tables
5. Pre-Joined Aggregates
6. Materialized Views and Materialized Query Tables

D. Summary



MASTERING DATA WAREHOUSE AGGREGATES HDT807

- IV. Using Aggregates**
 - A. Which Tables to Use**
 - 1. The Schema Design
 - 2. Relative Size
 - 3. Aggregate Portfolio and Availability
 - B. Requirements for the Aggregate Navigator**
 - 1. Why an Aggregate Navigator?
 - 2. Two Views and Query Rewrite
 - 3. Dynamic Availability
 - 4. Multiple Front Ends
 - 5. Multiple Back Ends
 - C. Evaluating Aggregate Navigators**
 - 1. Front-End Aggregate Navigators
 - 2. Back-End Aggregate Navigation
 - 3. Performance Add-On Technologies and OLAP
 - D. Specific Solutions**
 - 1. Living With Materialized Views
 - 2. Living With Materialized Query Tables
 - 3. Working Without an Aggregate Navigator
 - E. Summary**



MASTERING DATA WAREHOUSE AGGREGATES HDT807

- V. ETL Part 1: Incorporating Aggregates**
 - A. The Load Process**
 - 1. The Importance of the Load
 - 2. Tools of the Load
 - 3. Incremental Loads and Changed Data Identification
 - 4. The Top-Level Process
 - B. Loading the Base Star Schema**
 - 1. Loading Dimension Tables
 - 2. Loading Fact Tables
 - 3. Putting It All Together
 - C. Loading the Aggregate Schema**
 - 1. Loading Aggregates Separately from Base Schema Tables
 - 2. Invalid Aggregates
 - 3. Materialized Views and Materialized Query Tables
 - 4. Drop and Rebuild Versus Incremental Load
 - D. Real-Time Loads**
 - 1. Real-Time Load of the Base Schema
 - 2. Real-Time Load and Aggregate Tables
 - 3. Partitioning the Schema
 - E. Summary**



MASTERING DATA WAREHOUSE AGGREGATES HDT807

- VI. ETL Part 2: Loading Aggregates**
 - A. The Source Data For Aggregate Tables**
 - 1. Changed Data Identification
 - 2. Elimination of Redundant Processing
 - 3. Ensuring Conformance
 - 4. Loading the Base Schema and Aggregates Simultaneously
 - B. Loading Aggregate Dimensions**
 - 1. Requirements for the Aggregate Dimension Load Process
 - 2. Extracting and Preparing the Records
 - 3. Identifying and Processing New Records
 - 4. Identifying and Processing Type 1 Changes
 - 5. Processing Type 2 Changes
 - 6. Key Mapping
 - C. Loading Aggregate Fact Tables**
 - 1. Requirements for Loading Aggregate Fact Tables
 - 2. Acquire Data and Assemble Facts
 - 3. Identification of Surrogate Keys
 - 4. Aggregating Over Time
 - D. Dropping and Rebuilding Aggregates**
 - 1. Dropping and Rebuilding Aggregate Dimension Tables
 - 2. Dropping and Rebuilding Aggregate Fact Tables
 - E. Pre-Joined Aggregates**
 - 1. Dropping and Rebuilding a Pre-Joined Aggregate
 - 2. Incrementally Loading a Pre-Joined Aggregate
 - F. Materialized Views and Materialized Query Tables**
 - 1. Defining Attributes for Aggregate Dimensions
 - 2. Optimizing the Hierarchy
 - G. Summary**



MASTERING DATA WAREHOUSE AGGREGATES HDT807

VII. Aggregates and Your Project

- A. Data Warehouse Implementation**
 - 1. Incremental Implementation of the Data Warehouse
 - 2. Incorporating Aggregates Into the Project
- B. The Aggregate Project**
 - 1. Strategy Stage
 - 2. Design Stage
 - 3. Build Stage
 - 4. Deployment
- C. Management of Aggregates**
 - 1. Maintenance Responsibilities
 - 2. Ad Hoc Changes to Aggregate Portfolio
 - 3. An Ongoing Process
- D. Summary**

VIII. Advanced Aggregate Design

- A. Aggregating Facts**
 - 1. Periodic Snapshots Design
 - 2. Accumulating Snapshots
 - 3. Factless Fact Tables
- B. Aggregating Dimensions**
 - 1. Transaction Dimensions
 - 2. Bridge Tables
 - 3. Core and Custom Stars
- C. Other Schema Types**
 - 1. Snowflakes and Aggregates
 - 2. Third Normal Form Schemas and Aggregates
- D. Summary**



MASTERING DATA WAREHOUSE AGGREGATES HDT807

IX. Related Topics

- A. Aggregates and the Archive Strategy**
 - 1. The Data Warehouse Archive Strategy
 - 2. Aggregates and Archives
- B. Aggregates and Security**
 - 1. Dimensionally Driven Security and Aggregates
 - 2. Unrestricted Access to Summary Data
- C. Derived Tables**
 - 1. The Merged Fact Table
 - 2. The Pivoted Fact Table
 - 3. The Sliced Fact Table
- D. When Rollups Are Deployed Before Detail**
 - 1. Building the Base Table First
 - 2. Building the Rollup First
- E. Summary**