

图书基本信息

书名：<<掌握数据仓库数据聚合 Mastering Data Warehouse Aggregates>>

13位ISBN编号：9780471777090

10位ISBN编号：0471777099

出版时间：2006-12

出版时间：Wiley

作者：Adamson, Christopher

页数：345

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

<<掌握数据仓库数据聚合 Mastering>>

内容概要

This is the first book to provide in-depth coverage of star schema aggregates used in dimensional modeling—from selection and design, to loading and usage, to specific tasks and deliverables for implementation projects

Covers the principles of aggregate schema design and the pros and cons of various types of commercial solutions for navigating and building aggregates

Discusses how to include aggregates in data warehouse development projects that focus on incremental development, iterative builds, and early data loads

书籍目录

Foreword. Acknowledgments. Introduction. Chapter 1 Fundamentals of Aggregates. Star Schema Basics. Operational Systems and the Data Warehouse. Operational Systems. Data Warehouse Systems. Facts and Dimensions. The Star Schema. Dimension Tables and Surrogate Keys. Fact Tables and Grain. Using the Star Schema. Multiple Stars and Conformance. Data Warehouse Architecture. Invisible Aggregates. Improving Performance. The Base Schema and the Aggregate Schema. The Aggregate Navigator. Principles of Aggregation. Providing the Same Results. The Same Facts and Dimension Attributes as the Base Schema. Other Types of Summarization. Pre-Joined Aggregates. Derived Tables. Tables with New Facts. Summary. Chapter 2 Choosing Aggregates. What Is a Potential Aggregate? Aggregate Fact Tables: A Question of Grain. Aggregate Dimensions Must Conform. Pre-Joined Aggregates Have Grain Too. Enumerating Potential Aggregates. Identifying Potentially Useful Aggregates. Drawing on Initial Design. Design Decisions. Listening to Users. Where Subject Areas Meet. The Conformance Bus. Aggregates for Drilling Across. Query Patterns of an Existing System. Analyzing Reports for Potential Aggregates. Choosing Which Reports to Analyze. Assessing the Value of Potential Aggregates. Number of Aggregates. Presence of an Aggregate Navigator. Space Consumed by Aggregate Tables. How Many Rows Are Summarized. Examining the Number of Rows Summarized. The Cardinality Trap and Sparsity. Who Will Benefit from the Aggregate. Summary. Chapter 3 Designing Aggregates. The Base Schema. Identification of Grain. When Grain Is Forgotten. Grain and Aggregates. Conformance Bus. Rollup Dimensions. Aggregation Points. Natural Keys. Source Mapping. Slow Change Processing. Hierarchies. Housekeeping Columns. Design Principles for the Aggregate Schema. A Separate Star for Each Aggregation. Single Schema and the Level Field. Drawbacks to the Single Schema Approach. Advantages of Separate Tables. Pre-Joined Aggregates. Naming Conventions. Naming the Attributes. Naming Aggregate Tables. Aggregate Dimension Design. Attributes of Aggregate Dimensions. Sourcing Aggregate Dimensions. Shared Dimensions. Aggregate Fact Table Design. Aggregate Facts: Names and Data Types. No New Facts, Including Counts. Degenerate Dimensions. Audit Dimension. Sourcing Aggregate Fact Tables. Pre-Joined Aggregate Design. Documenting the Aggregate Schema. Identify Schema Families. Identify Dimensional Conformance. Documenting Aggregate Dimension Tables. Documenting Aggregate Fact Tables. Pre-Joined Aggregates. Materialized Views and Materialized Query Tables. Summary. Chapter 4 Using Aggregates. Which Tables to Use? The Schema Design. Relative Size. Aggregate Portfolio and Availability. Requirements for the Aggregate Navigator. Why an Aggregate Navigator? Two Views and Query Rewrite. Dynamic Availability. Multiple Front Ends. Multiple Back Ends. Evaluating Aggregate Navigators. Front-End Aggregate Navigators. Approach. Pros and Cons. Back-End Aggregate Navigation. Approach. Pros and Cons. Performance Add-On Technologies and OLAP. Approach. Pros and Cons. Specific Solutions. Living with Materialized Views. Using Materialized Views. Materialized Views as Pre-Joined Aggregates. Materialized Views as Aggregate Fact Tables (Without Aggregate Dimensions). Materialized Views and Aggregate Dimension Tables. Additional Considerations. Living with Materialized Query Tables. Using Materialized Query Tables. Materialized Query Tables as Pre-Joined Aggregates. Materialized Query Tables as Aggregate Fact Tables (Without Aggregate Dimensions). Materialized Query Tables and Aggregate Dimension Tables. Additional Considerations. Working Without an Aggregate Navigator. Human Decisions. Maintaining the Aggregate Portfolio. Impact on the ETL Process. Summary. Chapter 5 ETL Part 1: Incorporating Aggregates. The Load Process. The Importance of the Load. Tools of the Load. Incremental Loads and Changed Data Identification. The Top-Level Process. Loading the Base Star Schema. Loading Dimension Tables. Attributes of the Dimension Table. Requirements for the Dimension Load Process. Extracting and Preparing the Record. Process New Records. Process Type 1 Changes. Process Type 2 Changes. Loading Fact Tables. Requirements for the Fact Table Load Process. Acquire Data and Assemble Facts. Identification of Surrogate Keys. Putting It All

Together. Loading the Aggregate Schema. Loading Aggregates Separately from Base Schema Tables. Invalid Aggregates. Load Frequency. Taking Aggregates Off-Line. Off-Line Load Processes. Materialized Views and Materialized Query Tables. Drop and Rebuild Versus Incremental Load. Drop and Rebuild. Incremental Loading of Aggregates. Real-Time Loads. Real-Time Load of the Base Schema. Real-Time Load and Aggregate Tables. Partitioning the Schema. Summary. Chapter 6 ETL Part 2: Loading Aggregates. The Source Data for Aggregate Tables. Changed Data Identification. Elimination of Redundant Processing. Ensuring Conformance. Loading the Base Schema and Aggregates Simultaneously. Loading Aggregate Dimensions. Requirements for the Aggregate Dimension Load Process. Extracting and Preparing the Records. Identifying and Processing New Records. Identifying and Processing Type 1 Changes. Processing Type 2 Changes. Key Mapping. Loading Aggregate Fact Tables. Requirements for Loading Aggregate Fact Tables. Acquire Data and Assemble Facts. Selecting Source Columns. Processing New Facts Only. Calculating and Aggregating Facts. One Query Does It All. Identification of Surrogate Keys. Aggregating Over Time. Dropping and Rebuilding Aggregates. Dropping and Rebuilding Aggregate Dimension Tables. Dropping and Rebuilding Aggregate Fact Tables. Pre-Joined Aggregates. Dropping and Rebuilding a Pre-Joined Aggregate. Incrementally Loading a Pre-Joined Aggregate. Materialized Views and Materialized Query Tables. Defining Attributes for Aggregate Dimensions. Optimizing the Hierarchy. Summary. Chapter 7 Aggregates and Your Project. Data Warehouse Implementation. Incremental Implementation of the Data Warehouse. Planning Data Marts Around Conformed Dimensions. Other Approaches. Incorporating Aggregates into the Project. Aggregates and the First Data Mart. Subsequent Subject Areas. The Aggregate Project. Strategy Stage. Technology Selection: Choosing an Aggregate Navigator. Additional Strategic Tasks and Deliverables. Design Stage. Design of the Aggregate Schema and Load Specification. Design Documentation. Developing Test Plans for Aggregates. Build Stage. Iterative Build and Aggregates. Build Tasks and Aggregates. Deployment. Transitioning to Production, Final Testing, and Documentation. End User Education. Management of Aggregates. Maintenance Responsibilities. Ad Hoc Changes to Aggregate Portfolio. An Ongoing Process. Summary. Chapter 8 Advanced Aggregate Design. Aggregating Facts. Periodic Snapshots Designs. Transactions. Snapshots. Semi-Additivity. Invisible Aggregates for Periodic Snapshots. Averaging Semi-Additive Facts Produces a Derived Schema. Taking Less Frequent Snapshots Does Not Produce an Invisible Aggregate. Accumulating Snapshots. The Accumulating Snapshot. Aggregating the Accumulating Snapshot. Factless Fact Tables. Factless Events and Aggregates. Coverage Tables and Aggregates. Aggregating Dimensions. Transaction Dimensions. Timestamping a Dimension. Aggregating a Timestamped Dimension. Bridge Tables. Dealing with Multi-Valued Attributes. Aggregates and Bridge Tables. Core and Custom Stars. Other Schema Types. Snowflakes and Aggregates. The Snowflake Schema. Aggregating Snowflakes. Third Normal Form Schemas and Aggregates. Summary. Chapter 9 Related Topics. Aggregates and the Archive Strategy. The Data Warehouse Archive Strategy. Aggregates and Archives. Maintaining Aggregates. Archive Versus Purge. Summarizing Off-Line Data. Aggregates and Security. Dimensionally Driven Security and Aggregates. Unrestricted Access to Summary Data. Derived Tables. The Merged Fact Table. The Pivoted Fact Table. The Sliced Fact Table. When Rollups Are Deployed Before Detail. Building the Base Table First. Building the Rollup First. Parallel Load Processes. Redeveloping the Load. Historic Detail. Summary. Glossary. Index.

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>