Lab 5: Dimension Modelling Types

Overview

Lab 5 delves into advanced dimension modelling concepts, focusing on various dimension types used in data warehouse design. It provides a comprehensive exploration of specialized dimension structures that address specific business requirements and data modelling challenges.

Key Dimension Types

Conformed Dimension

- **Definition**: A dimension that can be shared across different facts (business processes)
- **Characteristics**: Returns a distinct list of values mapped together from source systems
- **Application**: Used when the same dimension has the same meaning across multiple fact tables

Degenerate Dimension

- **Definition**: A dimension that has no content except for its primary key
- **Application**: Often used for transaction identifiers or reference numbers that don't require additional attributes

Junk Dimension (Garbage Dimension)

- **Definition**: Combines two or more related low-cardinality flags into a single dimension
- **Example**: Car color (red, black, blue) and body style (sedan, van, SUV)
- **Implementation**: Creates combinations of all possible values of individual indicator fields
- **Benefit**: Reduces the number of foreign keys in fact tables

Role-Playing (Re-usable) Dimension

- **Definition**: A single physical dimension referenced multiple times in a fact table

- **Example**: Date dimension used for order date, ship date, and delivery date
- **Implementation**: Creates separate views of the dimension with unique attribute column names
- **Benefit**: Reduces redundancy while maintaining logical separation

Outrigger Dimension

- **Definition**: A dimension containing a reference to another dimension table
- **Characteristics**: Secondary dimension entity attached to a main conformed dimension
- **Caution**: Should be used sparingly to avoid excessive snowflaking

Shrunken Rollup Dimension

- **Definition**: Used for developing aggregate (higher level of summary) fact tables
- **Application**: Supports summarized reporting at higher levels of granularity

Swappable Dimension

- **Definition**: A dimension with multiple alternate versions that can be swapped at query time
- **Characteristics**: Different meaning, structure, and data compared to primary dimension
- **Implementation Options**: Direct join with filters, logical views, or physical tables
- **Application**: Provides flexibility for different user needs or security requirements

Heterogeneous Dimension

- **Definition**: Used when different products have different attributes but share the same customer base
- **Implementation Options**:
- Separate dimensions for each product type
- Merged attributes with null values for unrelated attributes
- Generic design with only common attributes

Slowly Changing Dimension (SCD)

- **Definition**: Dimensions that change over time, requiring history tracking
- **Types**:
- Type 0: No changes made even if source changes
- Type 1: No history maintained, only latest value kept
- Type 2: Full history maintained with multiple records

- Type 3: Only current and previous values maintained
- Type 4: Split into current and historical tables

Fast Changing Dimension

- **Definition**: Dimensions with attributes that change very frequently
- **Solution**: Mini-dimensions that separate fast-changing attributes
- **Implementation**: Identify fast-changing columns and split them into separate junk dimensions

Multi-Valued Dimension

- **Definition**: Dimensions with many-to-many relationships to facts
- **Example**: Patients with multiple diagnoses, students with multiple majors
- **Implementation**: Uses bridge tables to relate multiple dimension values to a single fact record

Technical Implementation

The lab provides detailed explanations and examples for: 1. Identifying appropriate dimension types for specific business scenarios 2. Implementing various dimension structures in data warehouse design 3. Understanding the trade-offs between different dimension modelling approaches 4. Handling complex relationships between facts and dimensions

Practical Applications

This lab addresses advanced data modelling challenges: - Tracking historical changes in dimension attributes - Handling many-to-many relationships in dimensional models - Optimizing performance for dimensions with frequently changing attributes - Creating flexible dimension structures for diverse reporting needs