Module 3
Data Warehouse Design Practices and Methodologies

Lesson 1: Relational Database Concepts for Multidimensional Data

Lesson Objectives

- Discuss motivation for relational database representation of multidimensional data
- Explain importance of grain determination
- Provide examples of types of fact tables

Motivation for Table Design

- Lack of scalability and integration of data cube storage engines

- Dominance of relational model and products

- Large amounts of research and development on relational database features for data warehouses

- Predominant usage of relational databases for large data warehouses

Multidimensional Data Representations

Data cubes
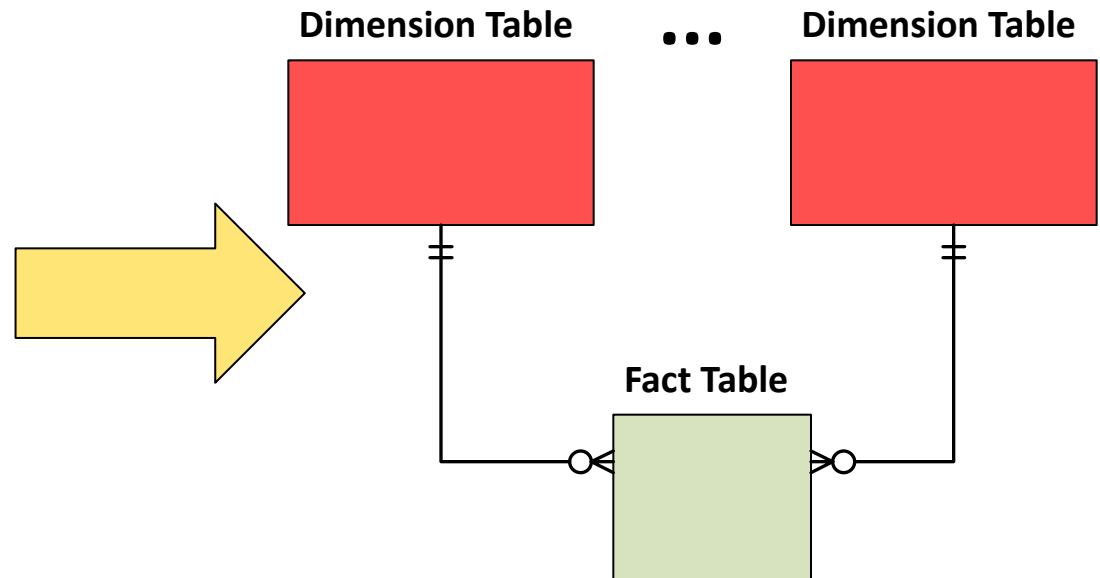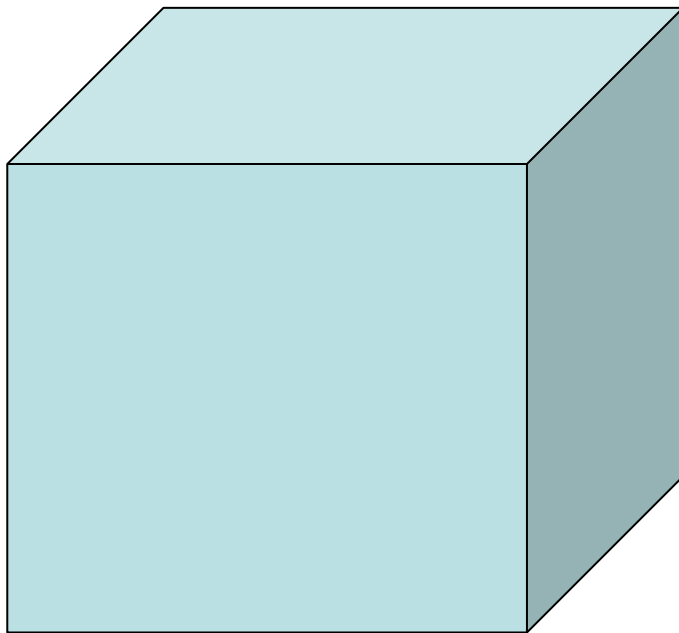
Grain

- Finest level of detail for a fact table
- Determined by the finest level of each dimension
- Completely specify related dimensions
- Determine size of fact tables using dimension cardinalities and sparsity
- Tradeoff
  - Flexibility and size
  - Trend towards finer grains

Grain Example

- Sales fact table grain
  - Coarse: customer postal codes (1,000), product type (100), store (200), week (52)
  - Fine: individual customer (200,000), individual product (2,000), store (200), day (365)
  - Sparsity: coarse (5%), fine (75%)
- Impact
  - Higher storage requirements for fine grain
  - More reporting flexibility for fine grain

Types of Fact Tables

- Transaction
  - Most common
  - Usually additive measures
- Snapshot
  - Periodic or accumulating view of asset level
  - Usually semi-additive measures
- Factless
  - Event occurrence
  - No measures, just FKs

Fact Table Examples

Transaction	Periodic	Factless
Store	Account	Student
Product	Account Type	Semester
Customer	Balance Date	Course
Associate	Dividend Date	Faculty
Date	Balance	Date
Quantity	Transaction Count	Period
Extended Price	Dividend Cumulative	
	Dividend Current Year	

Summary

- Importance of relational DBMS usage
- Choose grain carefully
- Understand types of fact tables and measure aggregation properties