**Produced by:**
**Dr. Brenda Mullally** bmullally@wit.ie
**Ruth Barry** rbarry@wit.ie

**Department Computing Maths and Physics**
**Waterford Institute of Technology**
www.wit.ie
moodle.wit.ie

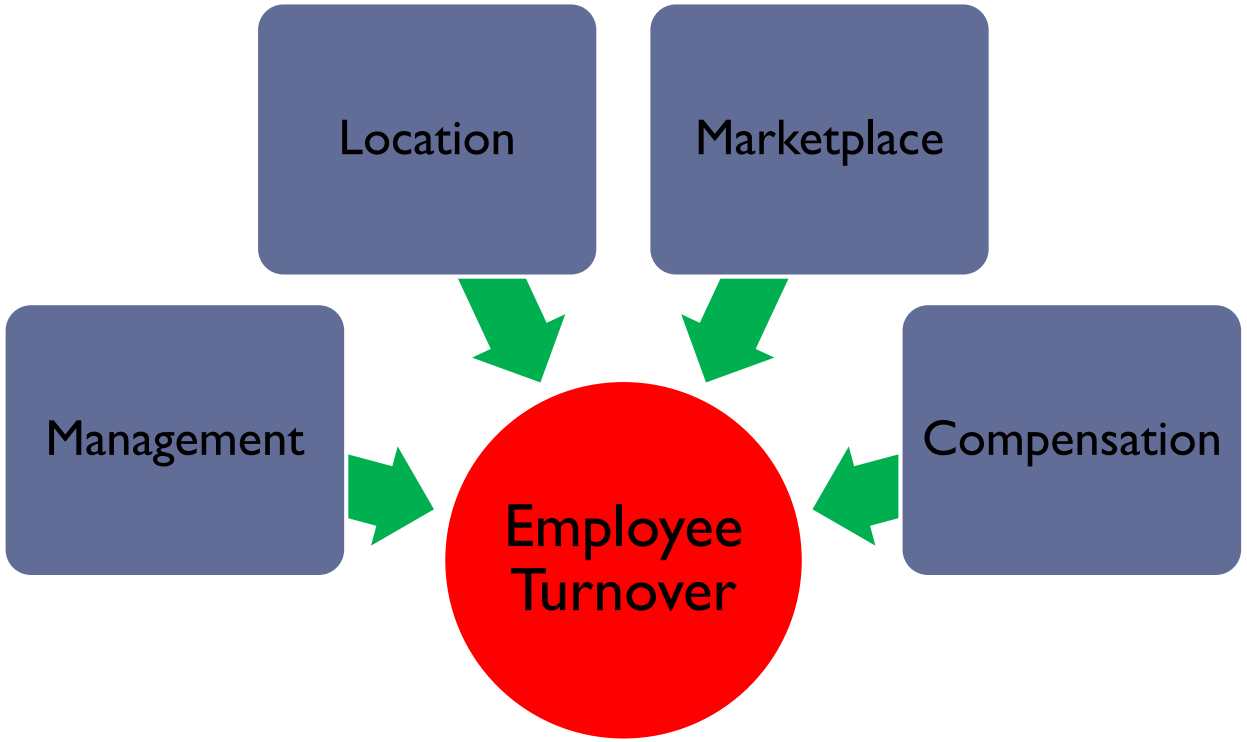
# MSc Enterprise Software Systems

# Business Intelligence

# Objectives

- Business analyst perspective
- Explain reasons for sparsity
- Provide examples of measure aggregation properties
- Explain use of each operator
- Discuss motivation for relational database representation of multidimensional data
- Explain importance of grain determination
- Provide examples of types of fact tables

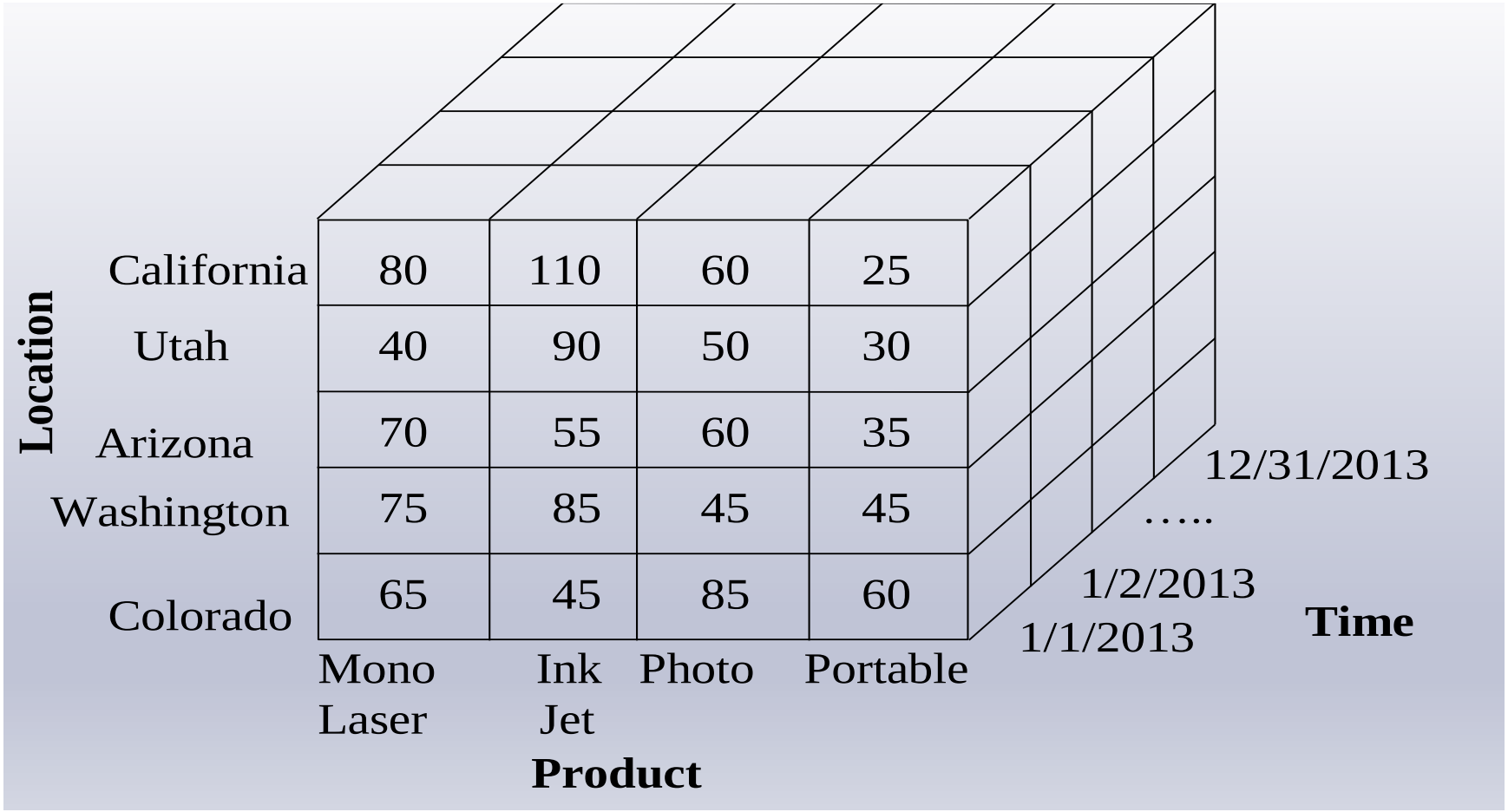
# Business Analyst Perspective



# Data Cube Basics

- Business analyst model
  - Factors or influencing variables of interest
  - Quantitative variables
  - Multidimensional arrangement
- Terminology
  - Dimension: subject label for a row or column
  - Member: value of dimension
  - Measure: quantitative variables stored in cells

# Sales Data Cube Example



# Notes on Dimensions and Measures

- Hierarchical dimensions with sub members
- Sparsity
  - Many cells do not have values
  - Increases with dimension detail and number of dimensions
- Measures
  - Derived measures
  - Multiple measures in cells

# Measure Aggregation Properties

- Additive
  - Summarized by addition across all dimensions
  - Common measures such as sales, cost, and profit
- Semi-Additive
  - Summarized by addition in some but not all dimensions such as time
  - Periodic measurements such as account balances and inventory levels
- Non-Additive
  - Cannot be summarized by addition through any dimension
  - Historical facts such as unit price for a sale

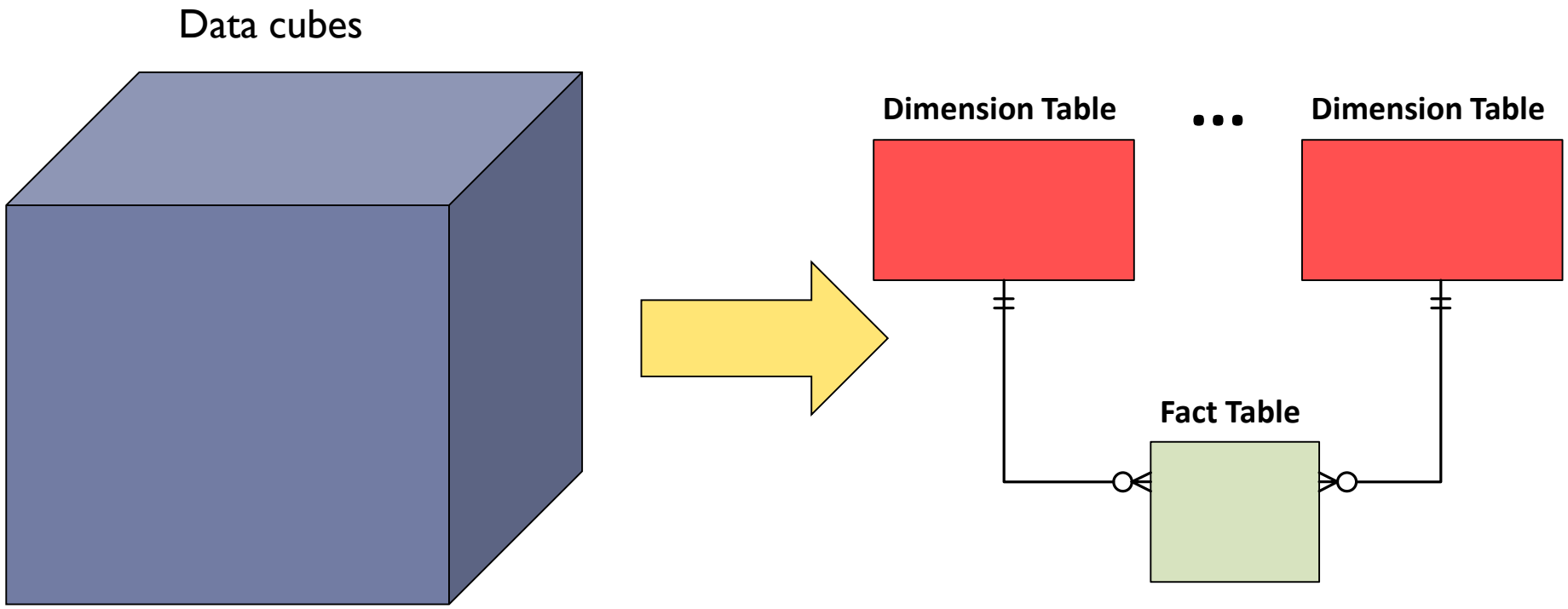
# Measure Aggregation Example

- Dimensions
  - Course: course id, degree, department, and college
  - Student: student id, major, department, and college
  - Time: semester, academic year, academic decade
- Measures:
  - Credit hours
  - Grade
  - Unit tuition
  - Tuition
- Aggregation properties for measures: ?

# 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



# 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

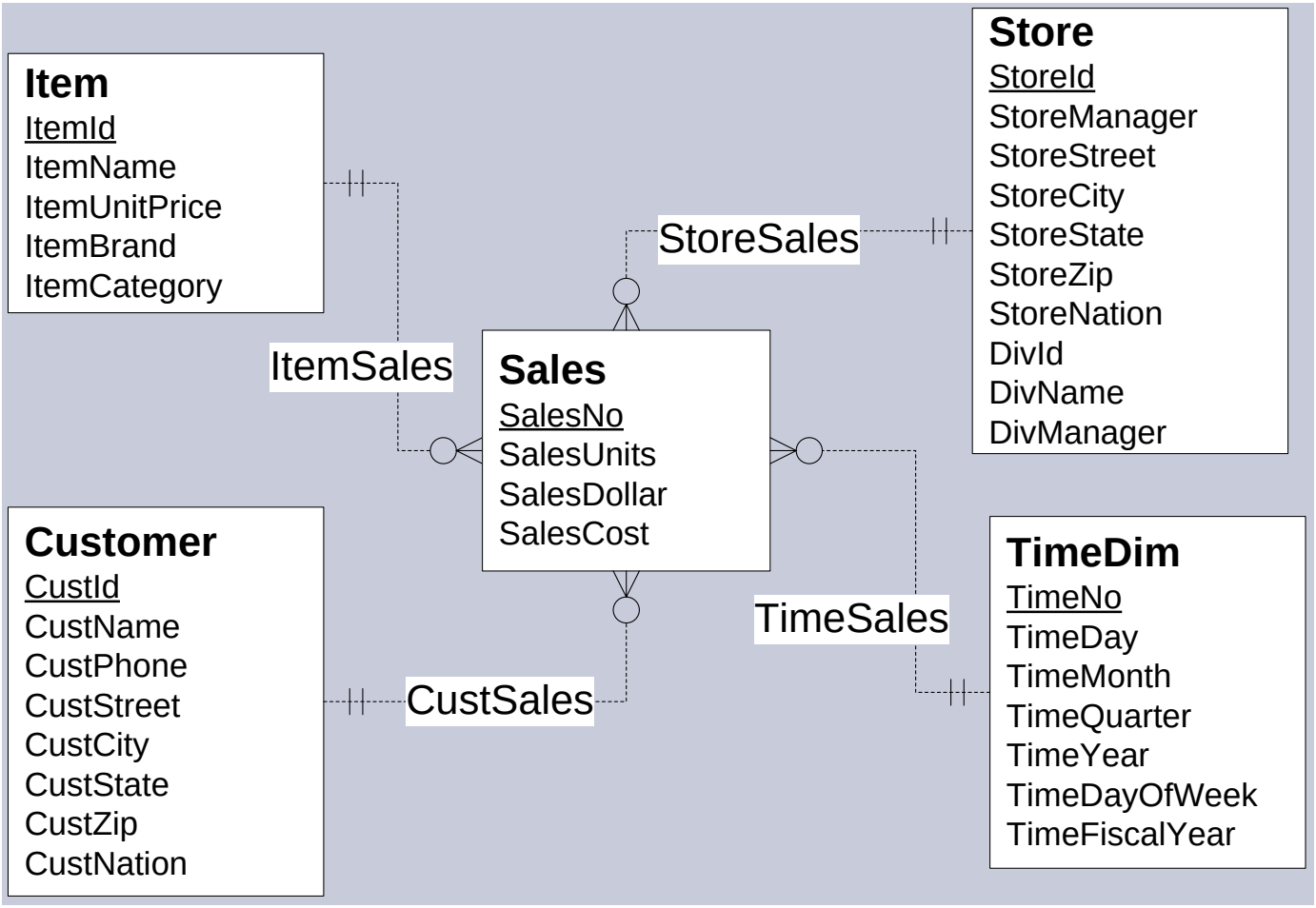
# Fact Table Examples

| 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 | |

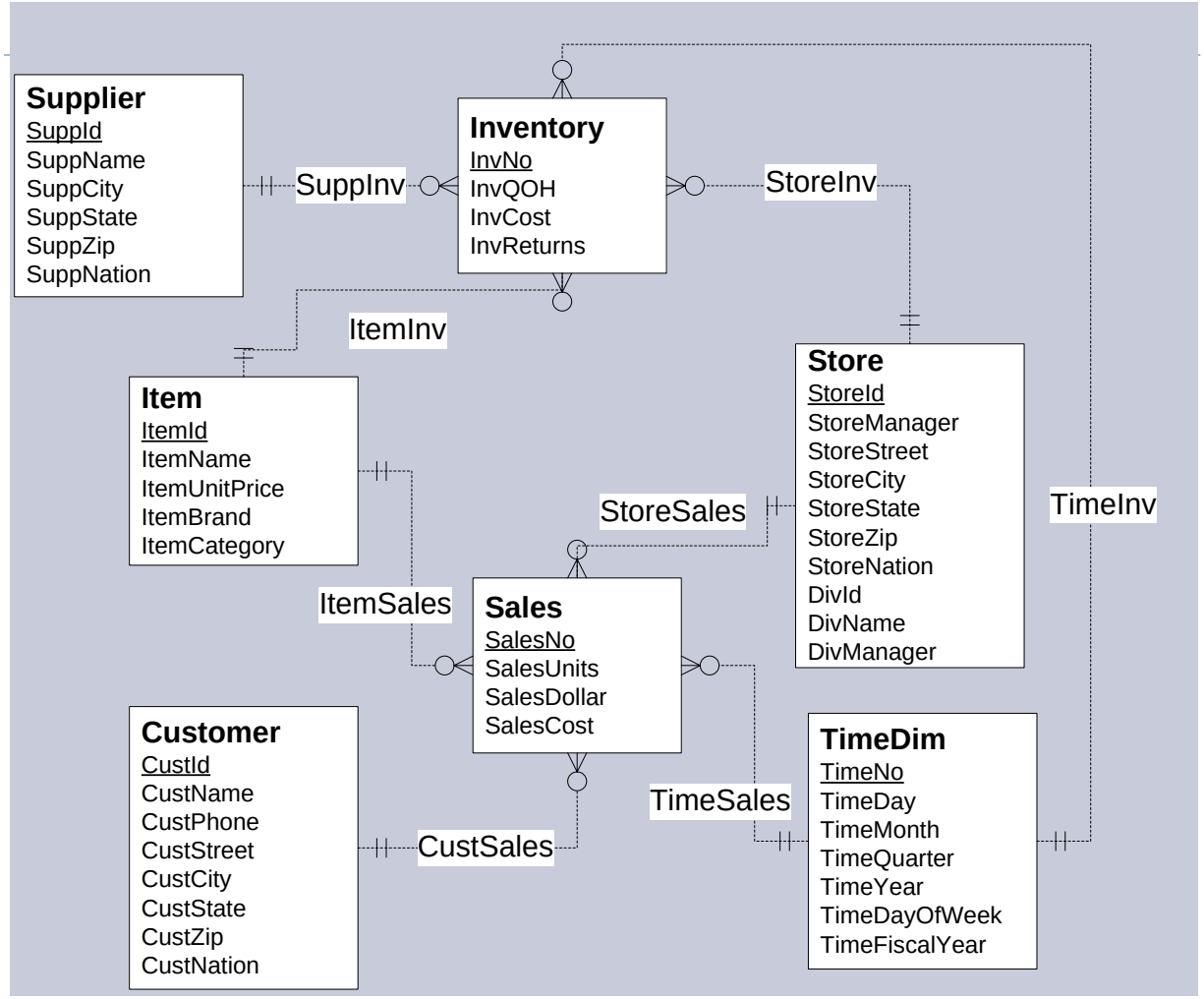
# DW Design

- Explain schema patterns
- Explain patterns for historical integrity
- Reflect on the importance of historical integrity

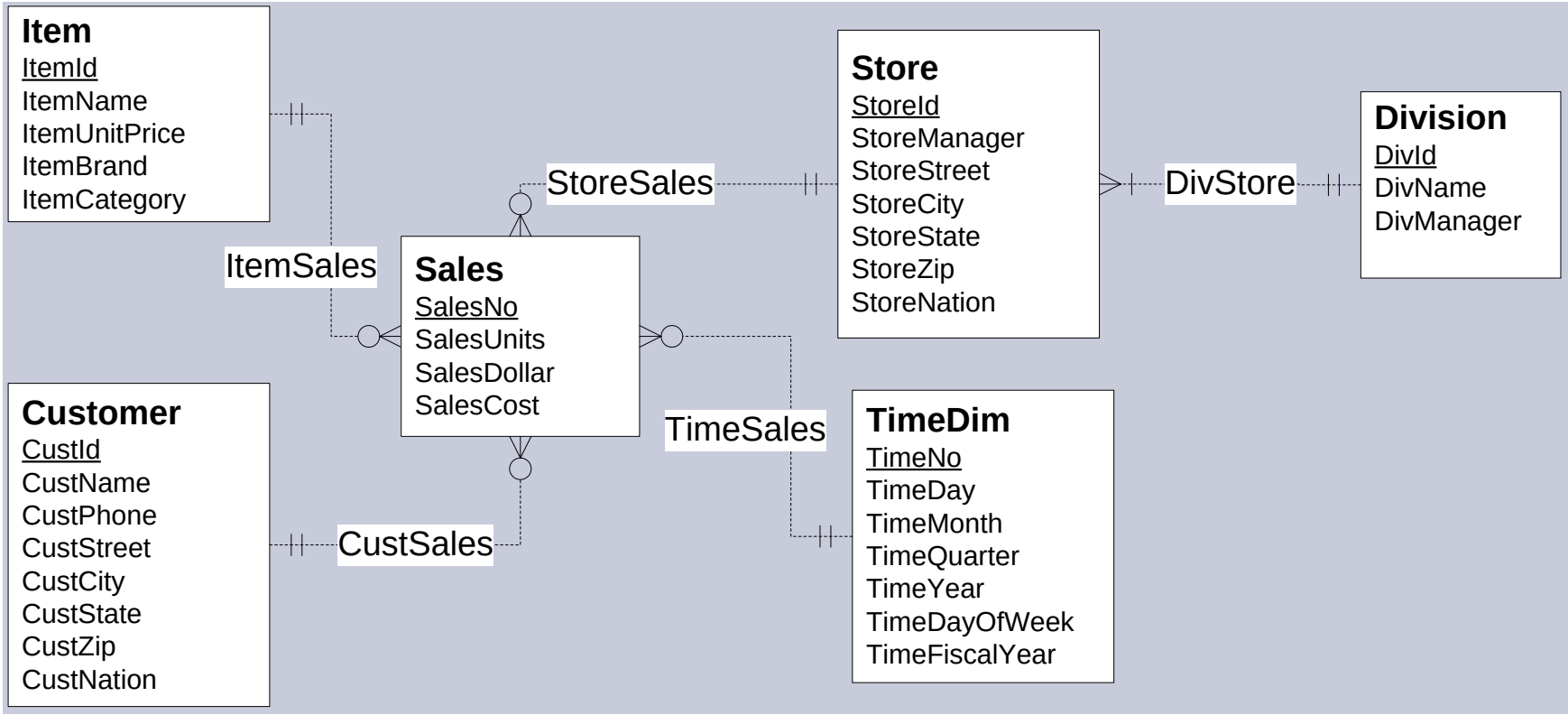
# Star Schema Example



# Constellation Schema Example




# Snowflake Schema Example



# Time Representation for Fact Tables

- Alternatives
  - Timestamp
  - Time dimension table for organization specific calendar features
- Variations
  - Time of day columns
  - Accumulating fact table for representation of multiple events

# Historical Integrity for Dimensions

- Primarily an issue for dimension changes
- Fact rows no longer historically accurate after a dimension update
- Determine importance of history preservation for dimension columns
- Some inaccuracy tolerated with summary query results

# Historical Integrity Examples

## Type I Representation

**Customer**
CustId
CustName
CustPhone
CustStreet
CustCity
CustState
CustZip
CustNation

## Type II Representation

**Customer**
CustId
VersionNo
CustName
CustPhone
CustStreet
CustCity
CustCityBegEffDate
CustCityEndEffDate
CustState
CustZip
CustNation

## Type III Representation

**Customer**
CustId
CustName
CustPhone
CustStreet
CustCityCurr
CustCityCurrBegEffDate
CustCityCurrEndEffDate
CustCityPrev
CustCityPrevBegEffDate
CustCityPrevEndEffDate
CustCityPast
CustCityPastBegEffDate
CustCityPastEndEffDate
CustState
CustZip
CustNation