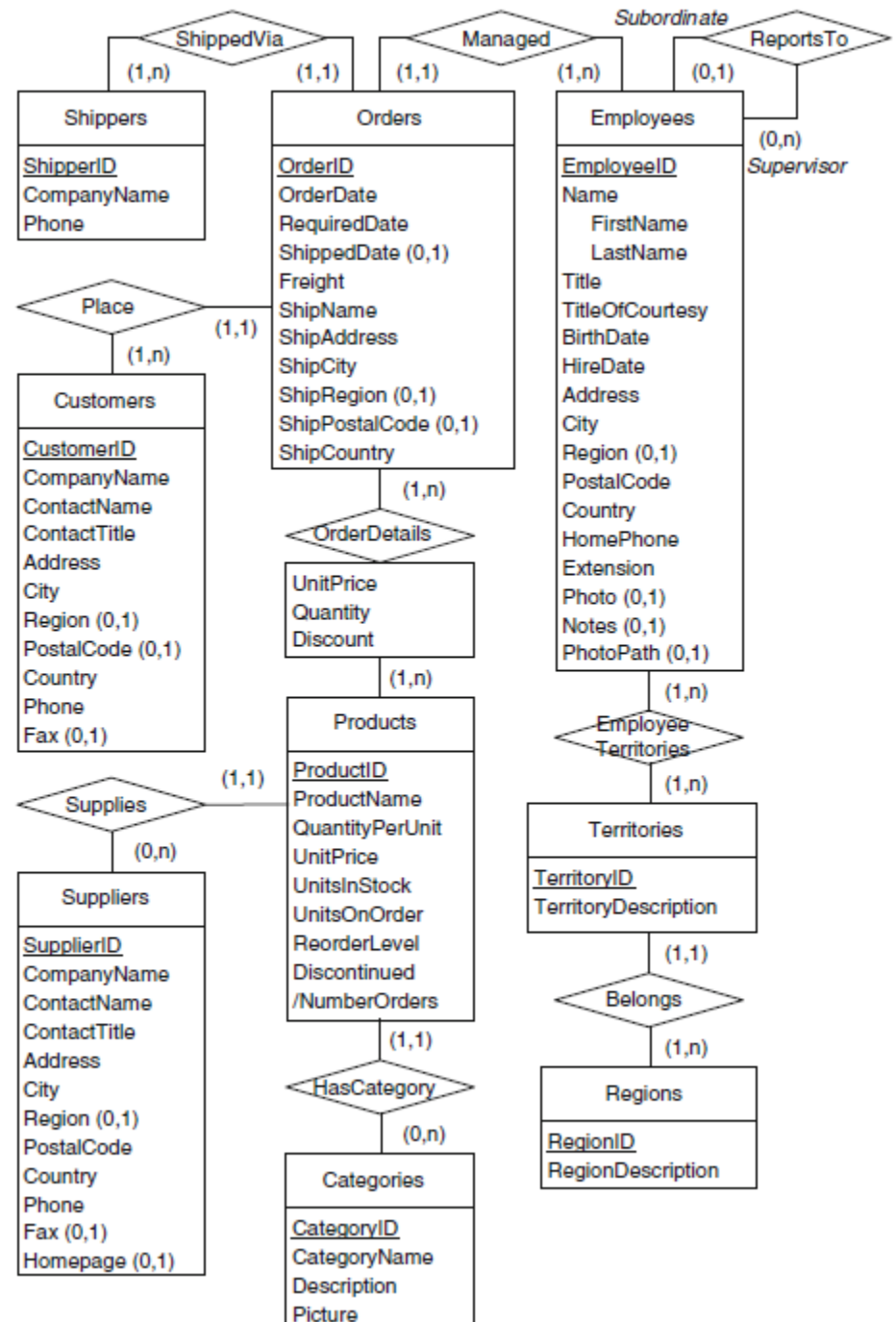

DS-306 Data Warehousing and Business Intelligence

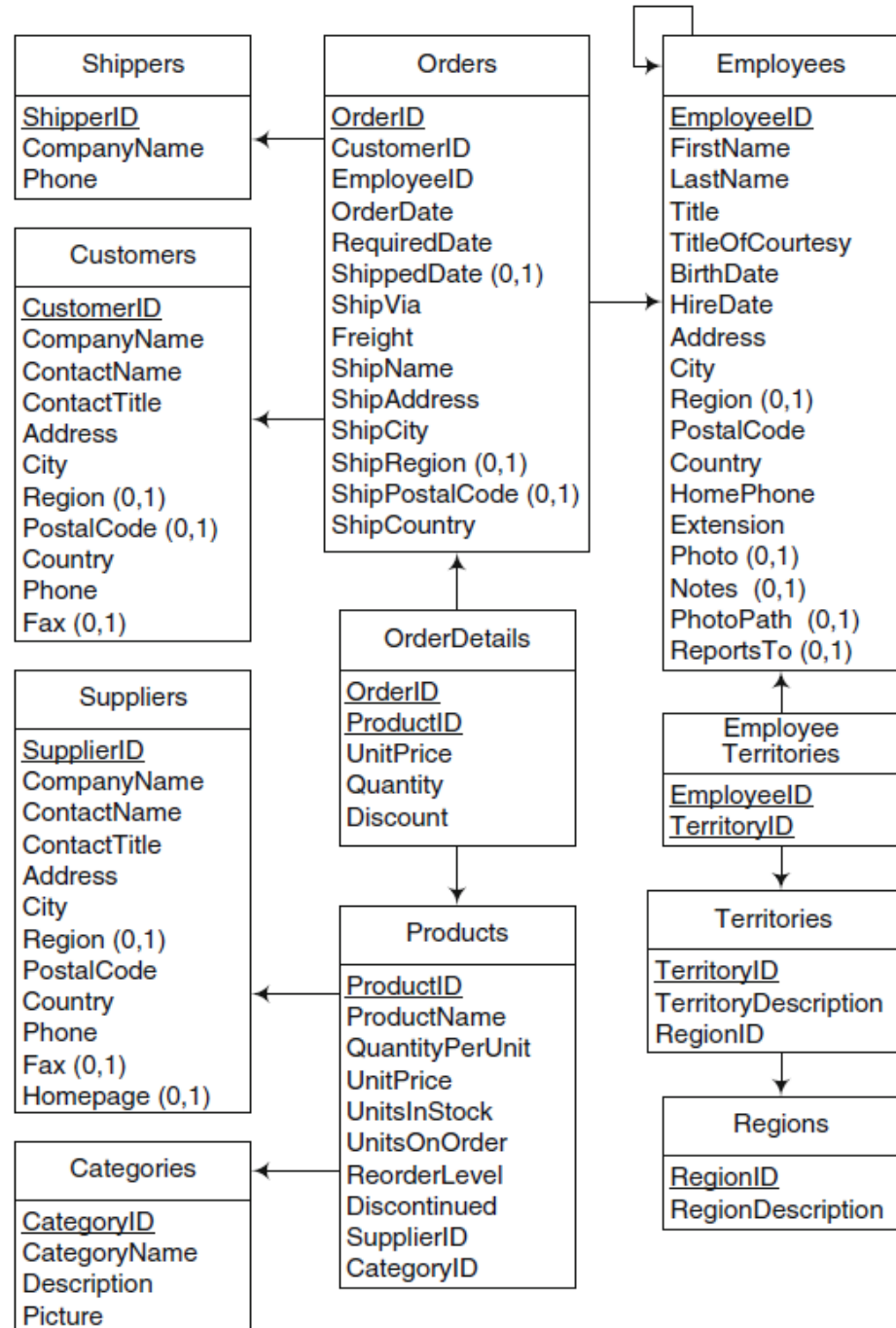
Topic 3: Conceptual Data Warehouse Design

Dr. Khurram Shahzad

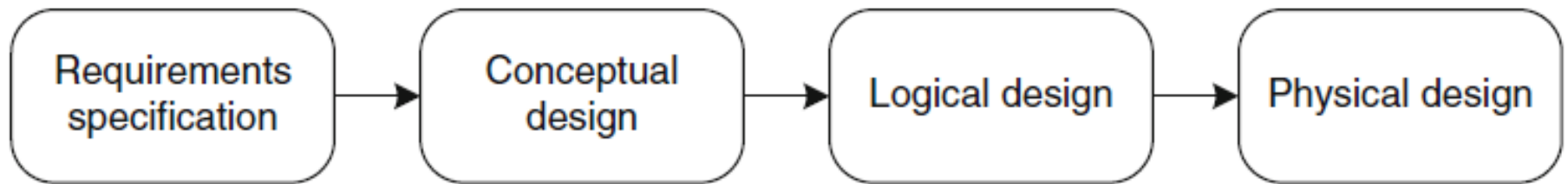
Northwind ERD



Northwind Logical



Data Warehouse Design Process



Dimensional Modeling

- Dimensional Modeling focuses subject-orientation, critical factors of business
- Critical factors are stored in **facts**
- Redundancy is no problem, achieve efficiency

Dimensional Modeling

- Two important concepts

- **Fact**

- Numeric measurements, represent business activity/event
 - Are pre-computed, redundant
 - Example: Profit, quantity sold

Dimensional Modeling

- Two important concepts

- **Dimension**

- Qualifying characteristics, perspective to a fact
 - Example: date (Date, month, quarter, year), product(type, category)

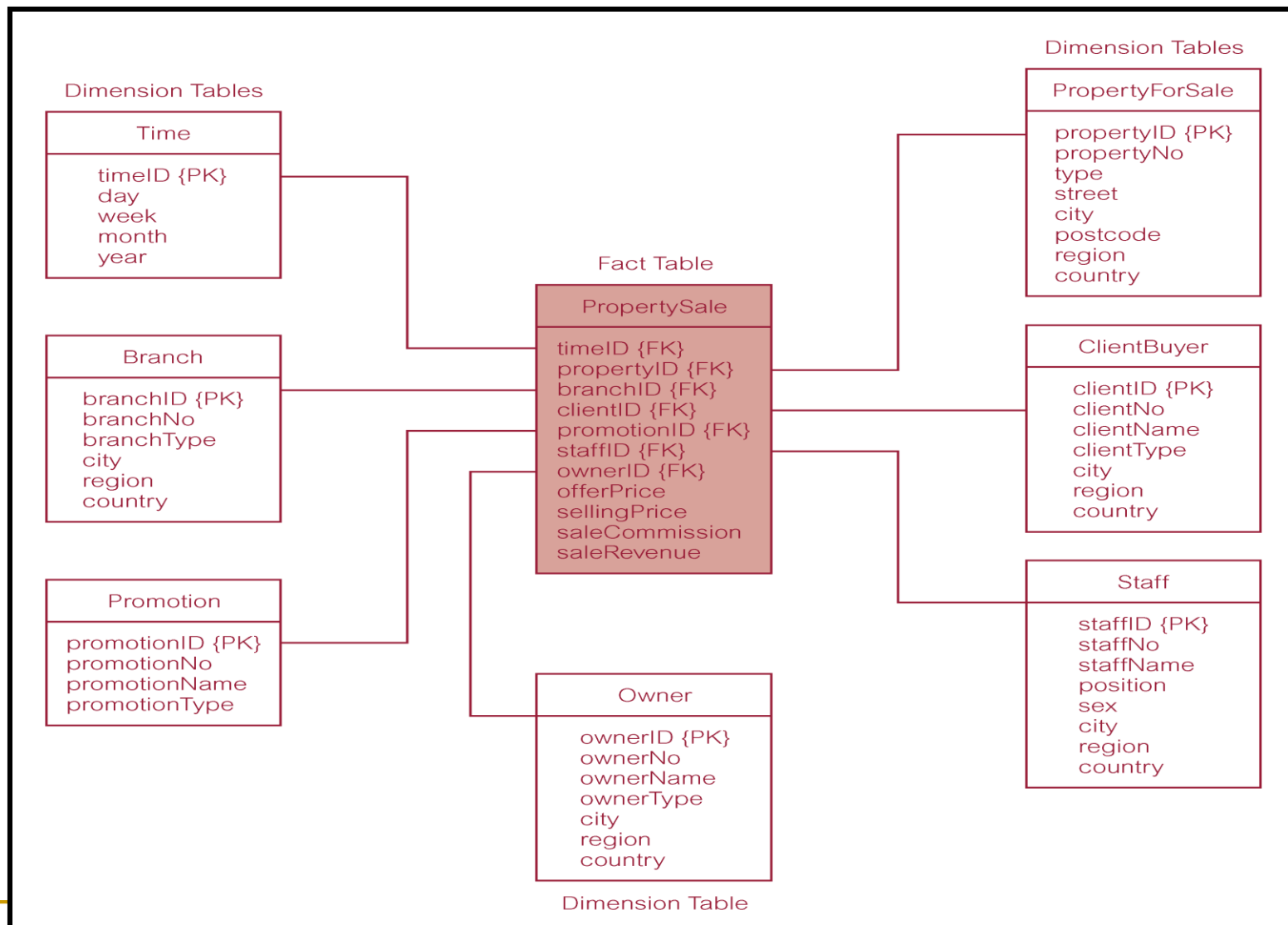
Dimensional Modeling

- Every dimensional model (DM) is composed **of one (or more) facts**, and **a set of dimensions**.
- Look on facts through one (or more) dimensions.
 - What is the **sale amount** in Consumer **Product category**, for **elderly customers** in the second quarter of 2004?

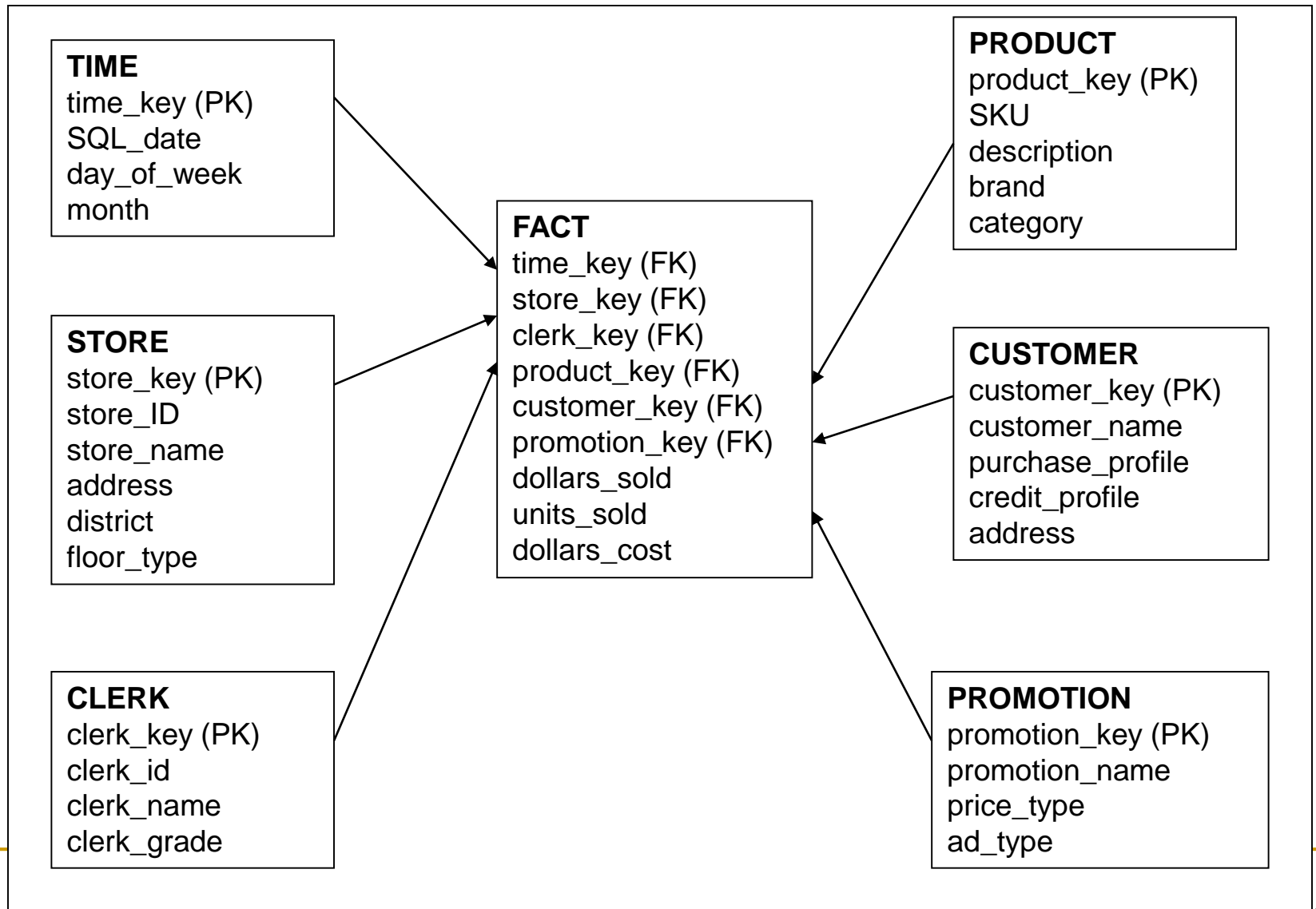
Dimensional Modeling (Contd.)

- Forms '**star-like**' structure, which is called a **star schema** or **star join**
- Dimensional Model has three types
 - Star Schema
 - Snowflake Schema
 - Star-flake Schema

A Typical Dimensional Model



Example



Conceptual Model

Conceptual modeling: MultiDim Model

- Remember, A **star schema** is composed of a set of dimensions and a set of facts
- The conceptual model of DW is called MultiDim Model

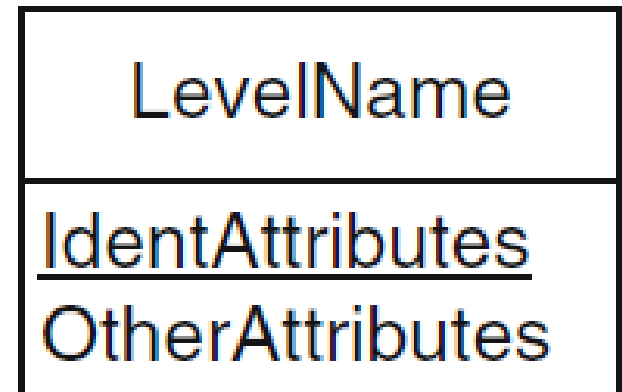
Notation of MultiDim Model

1. Level
2. Hierarchy
3. Cardinalities
4. Fact with measures and associate levels
5. Types of measures
6. Hierarchy name
7. Exclusive relationships

Notation of MultiDim Model

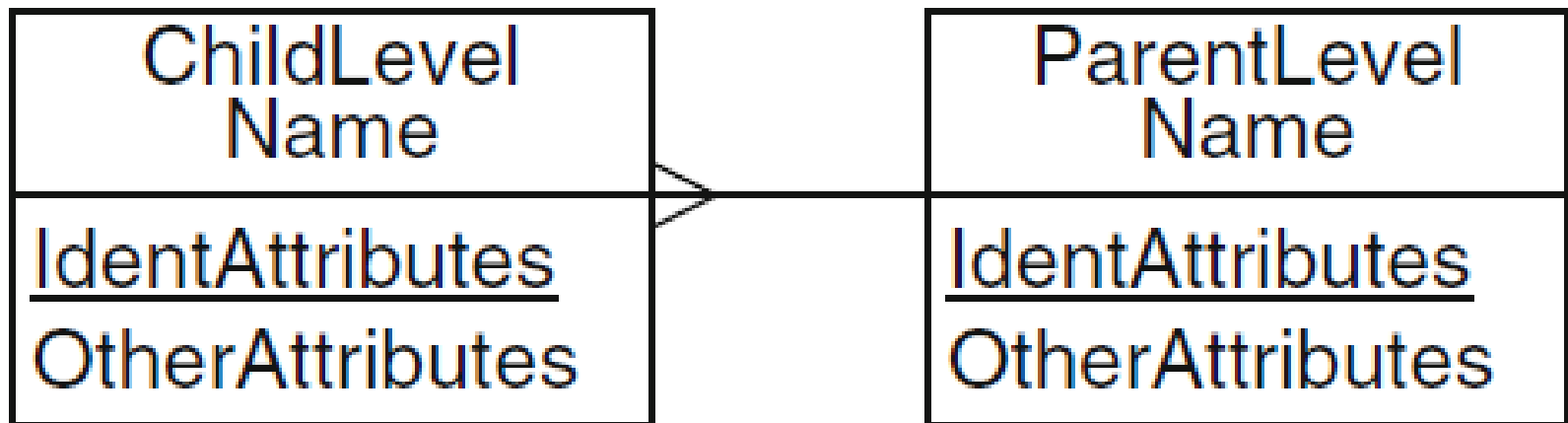
■ Level

- ❑ Level is analogous to an entity type in the ER model
- ❑ Level has a set of **attributes**
- ❑ Level has **identifiers** that uniquely identifies the members of a level
- ❑ Instances of a level are called **members**

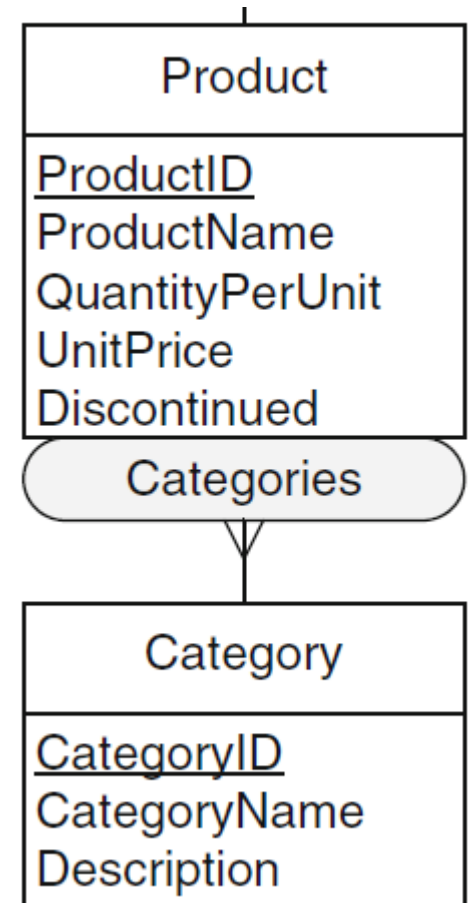
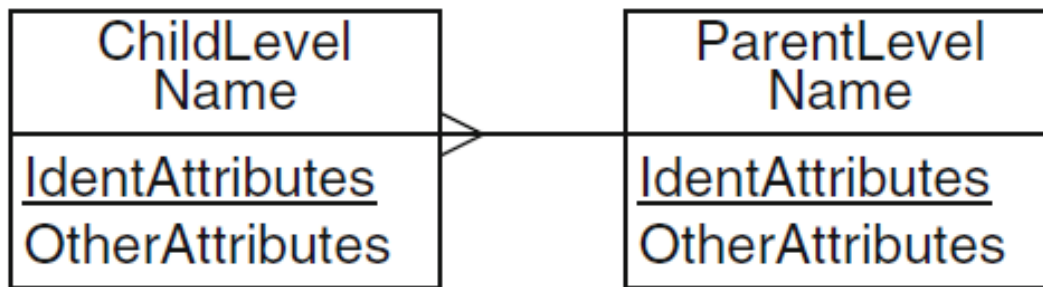


Notation of MultiDim Model

- Hierarchy
 - Product and Category



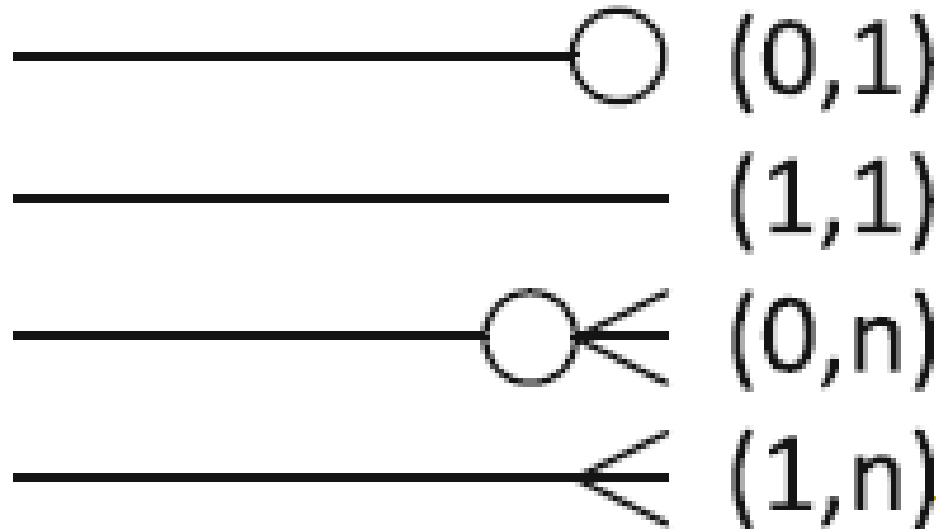
Example



Notation of MultiDim Model

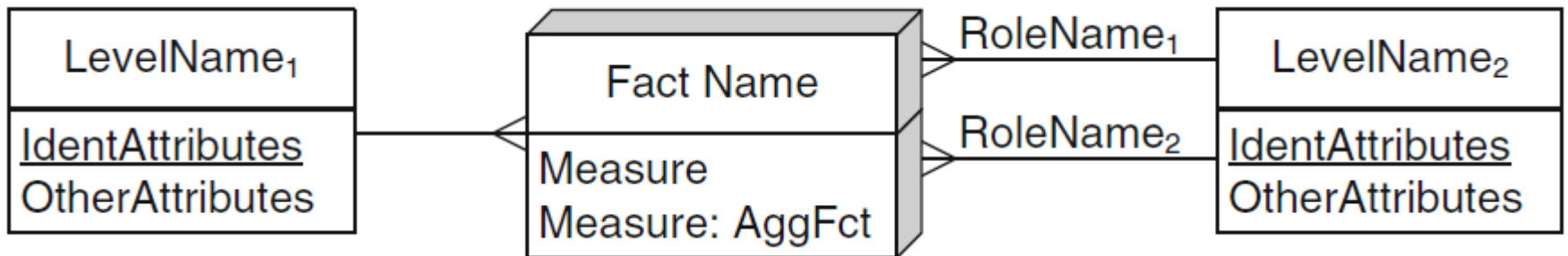
■ Cardinalities

- Indicates the minimum and maximum number of fact members that can be related to level member
- Sales fact is related to Product with one-to-many



Notation of MultiDim Model

- A fact may contain attributes commonly called **measures**
- Fact with measures and associated levels
 - Sales fact can include, measures, quantity, unit process, discount, sales amount, net-amount



Types of Measures

- Additive measure
 - Meaningfully summarized along all the dimensions, using addition e.g. quantity
- Semiadditive measures
 - Can be meanfully summarized using addition along *some*, but not all, dimensions e.g. inventory quantity

Types of Measures

- Nonadditive measures

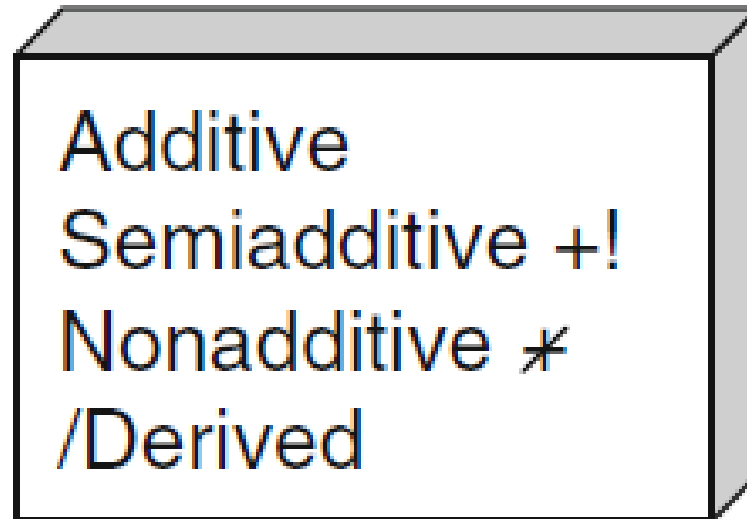
- That cannot be meaningfully summarized using addition across dimensions e.g. price, cost per unit, exchange rate

- Derived

- Measures can be derived. That is, they are calculated on the basis of other measures

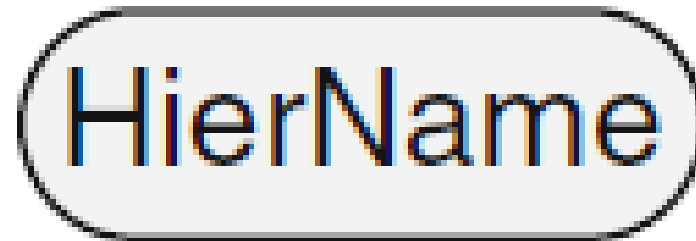
Notation of MultiDim Model

- Types of measures



Notation of MultiDim Model

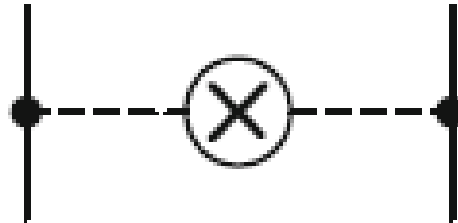
- Hierarchy name
 - Hierarchy comprises of several related levels
 - Lower level is called **child**, and higher is called **parent**



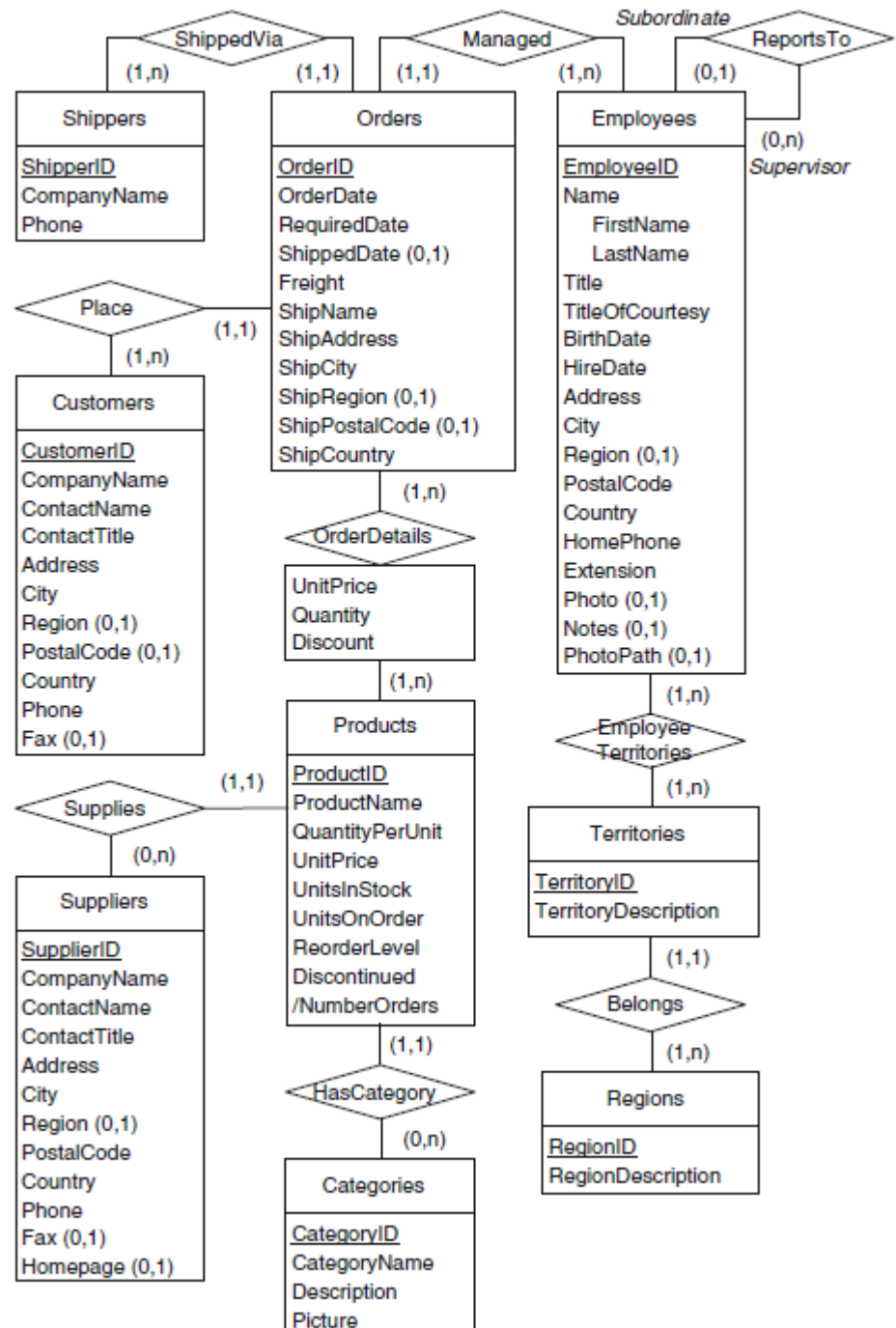
HierName

Notation of MultiDim Model

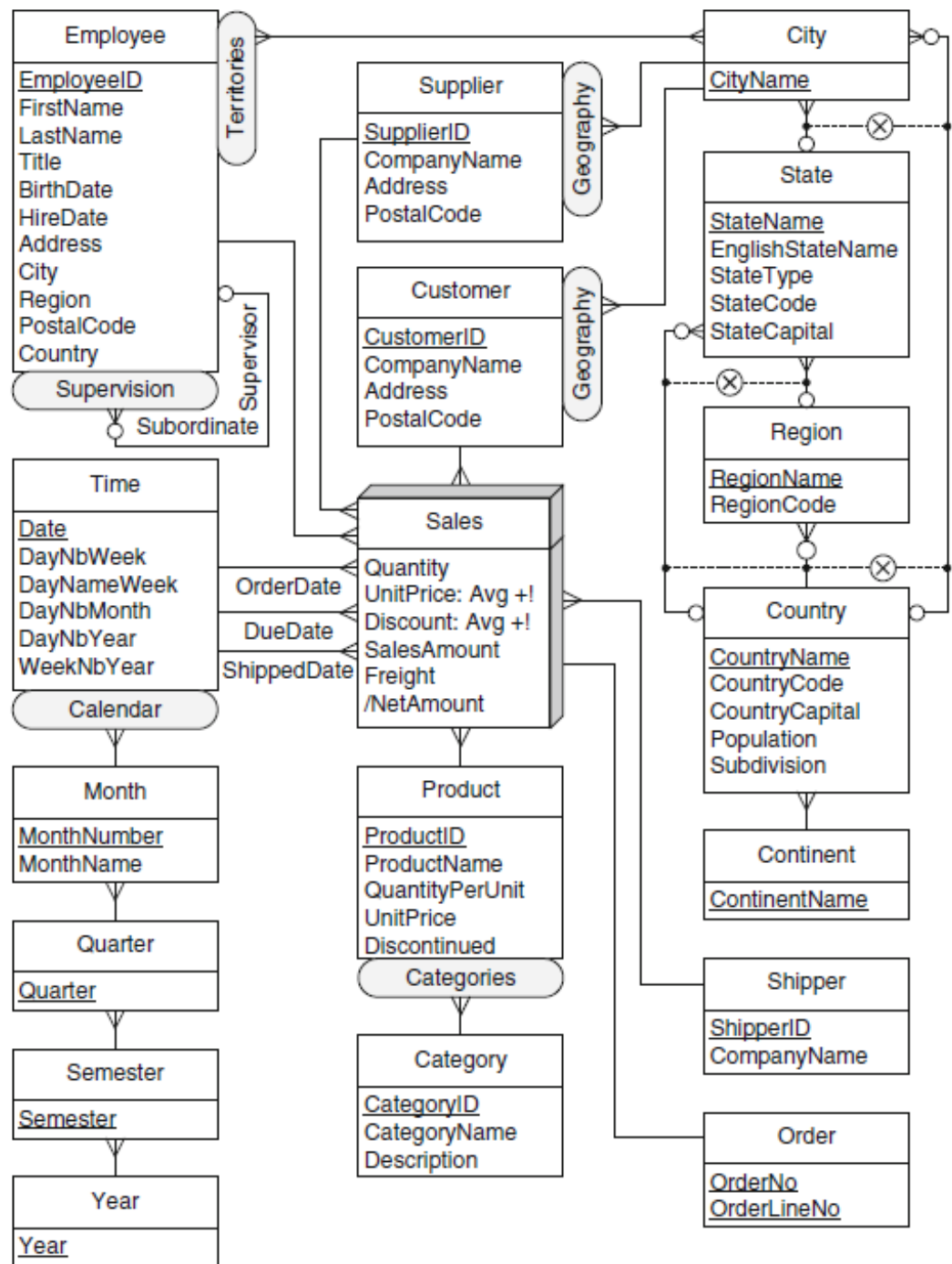
- Exclusive relationships
 - States can be aggregated into regions
 - States can be aggregated into countries



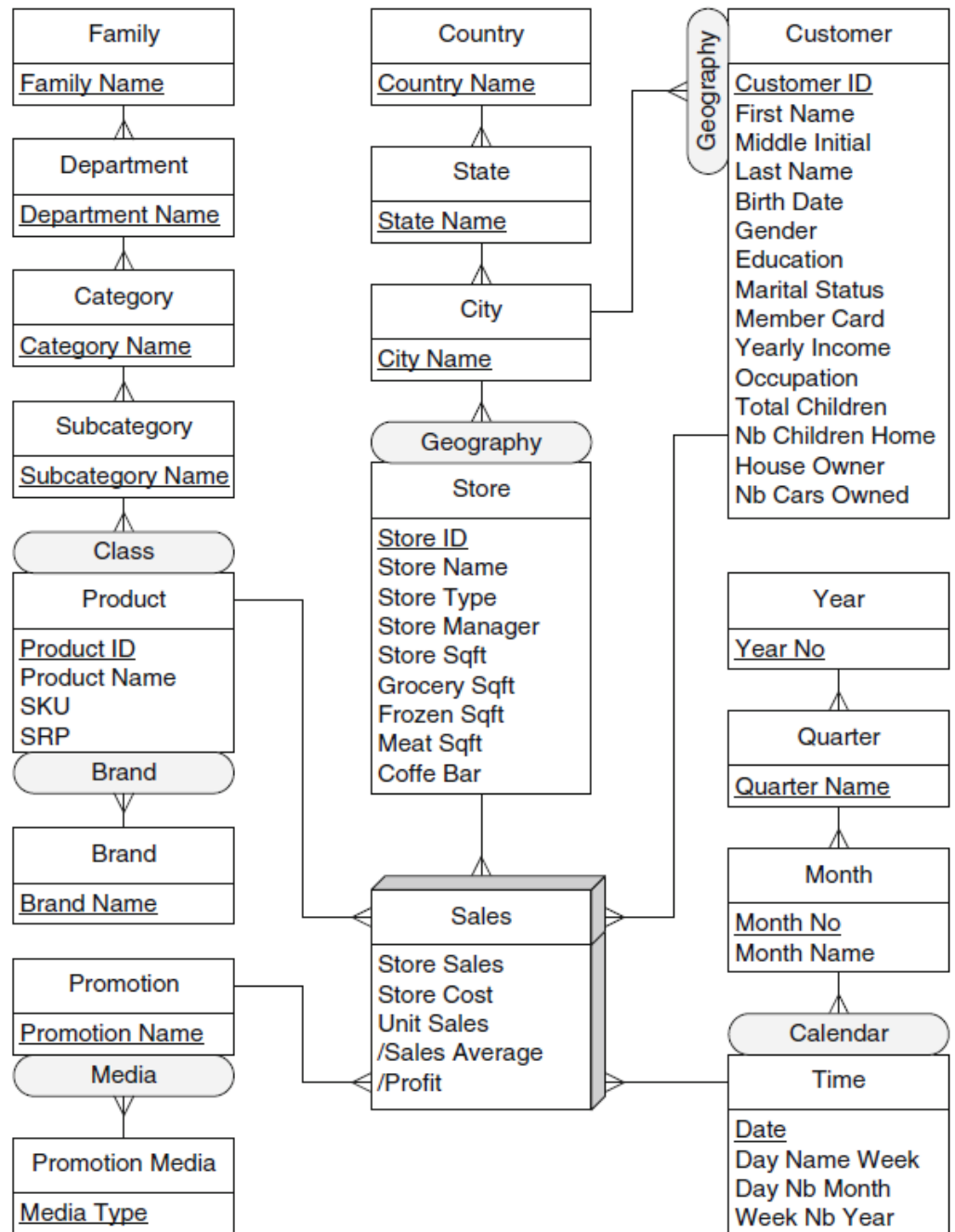
Northwind ERD



Northwind DM



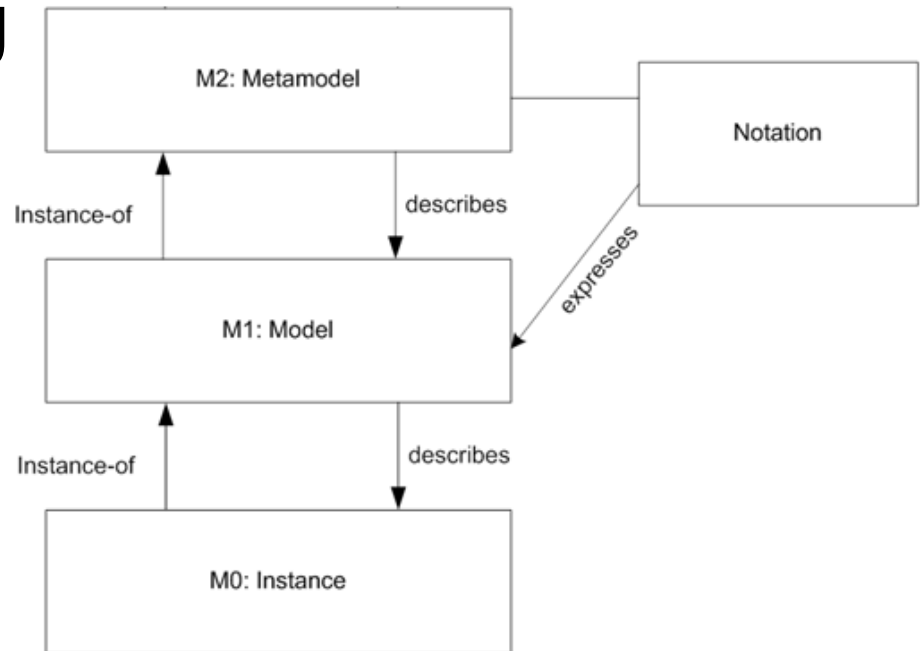
Foodmart DW



Foundations of Conceptual Modeling

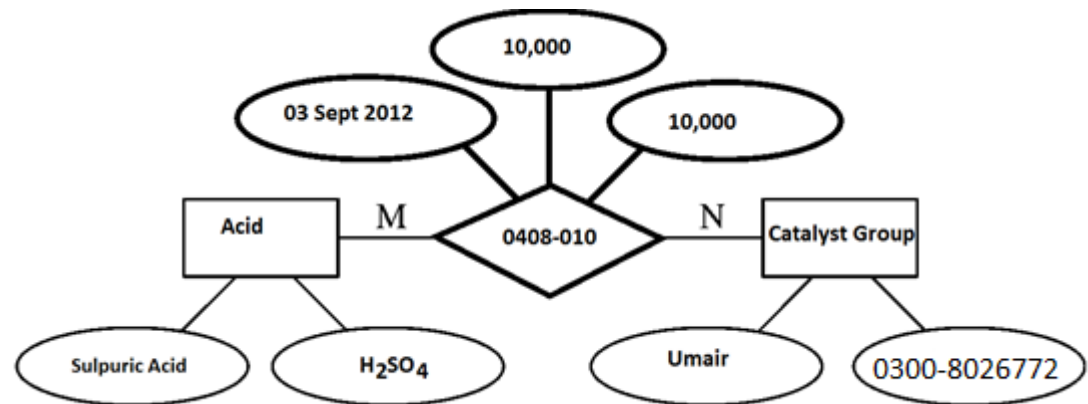
What is Metamodel?

- A model that contains
 - All the modeling concepts
 - All possible relationships between the concepts
- Levels in modelling



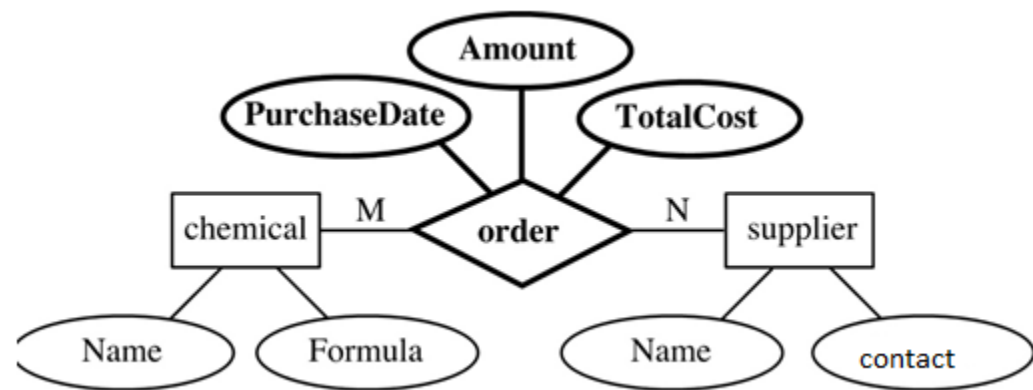
Abstraction concepts ERD

- *Instance level* reflects the concrete entities that are involved in business processes
 - At this level, executed activities, concrete data values, and resources & person are represented
 - Examples of instances

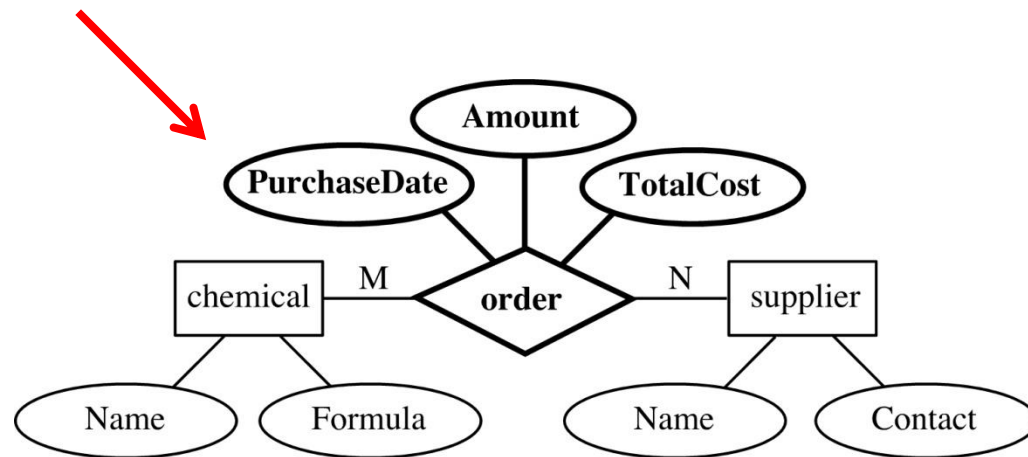
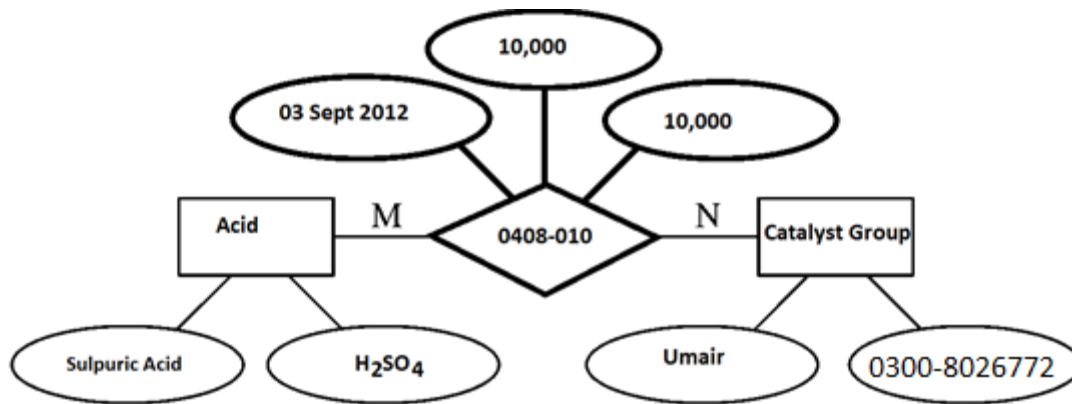


Abstraction concepts

- To organize complexity of process scenarios, a set of similar entities at instance level are identified and classified at *model level*
 - ❑ A set of similar process instances are classified and represented by a process
 - ❑ Examples of model level

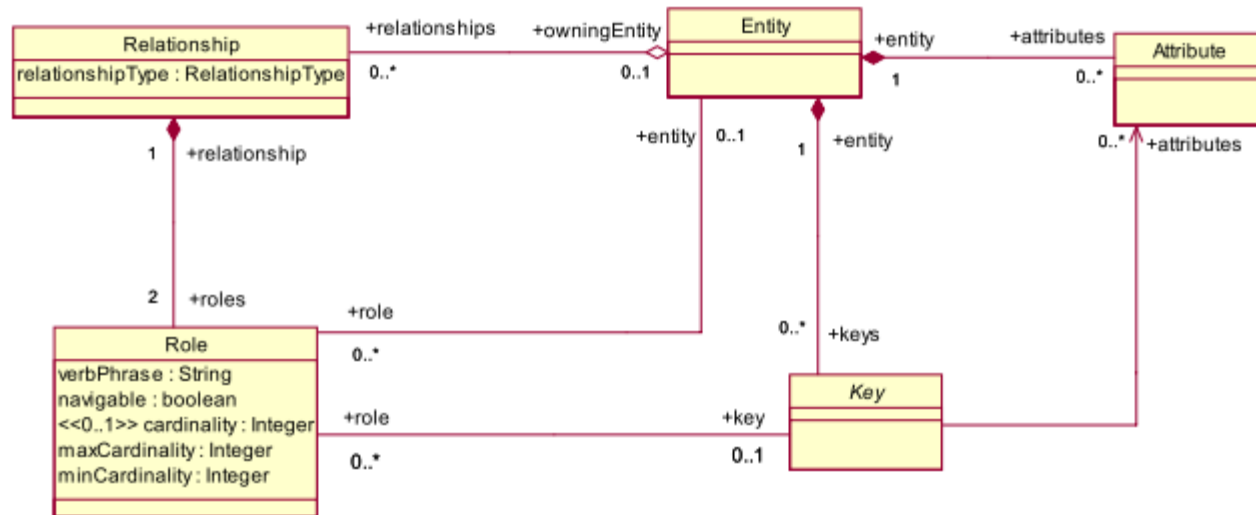


Abstraction concepts

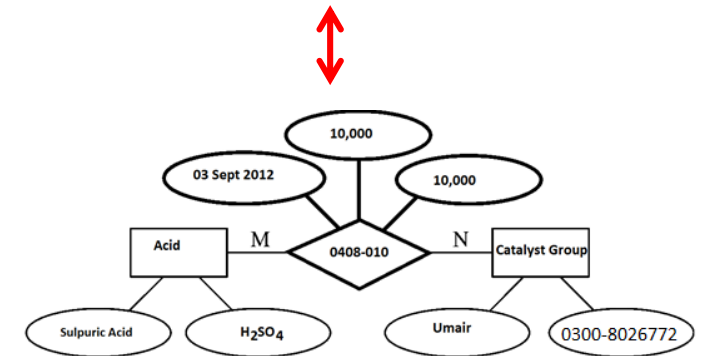
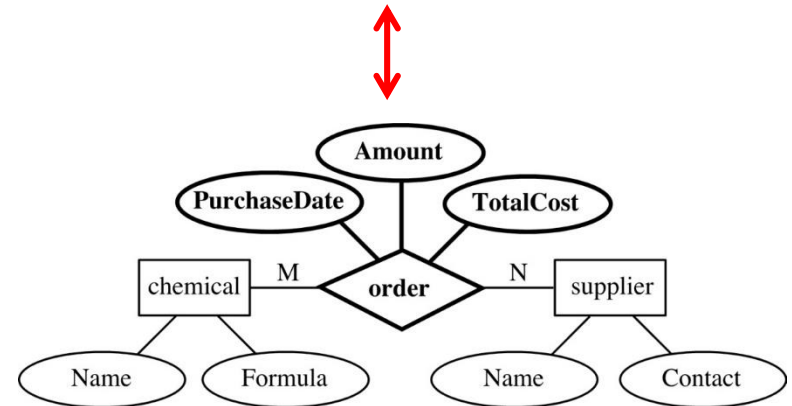
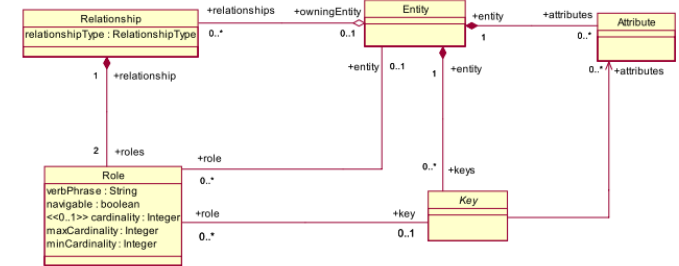
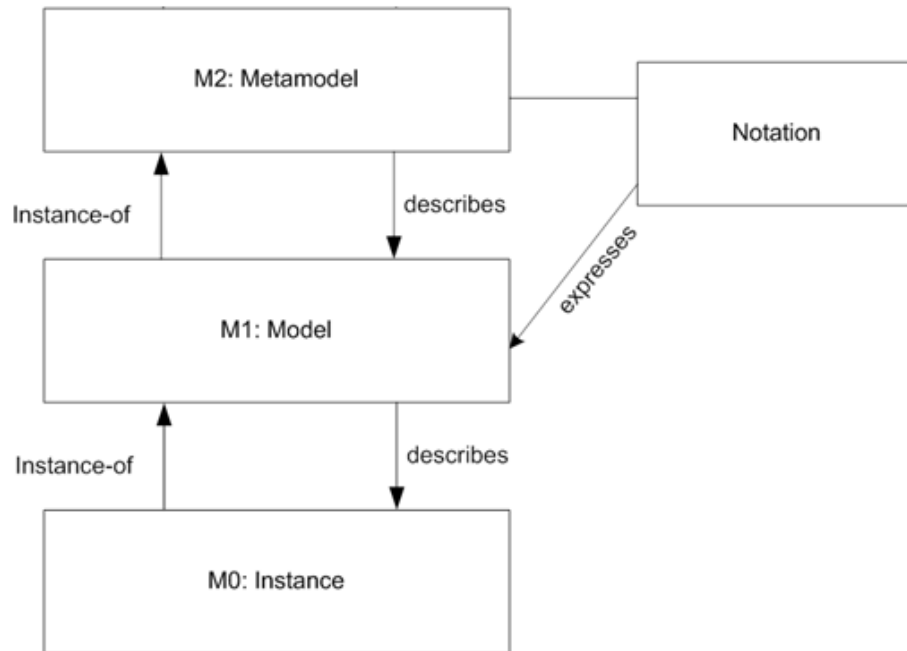


Abstraction concepts

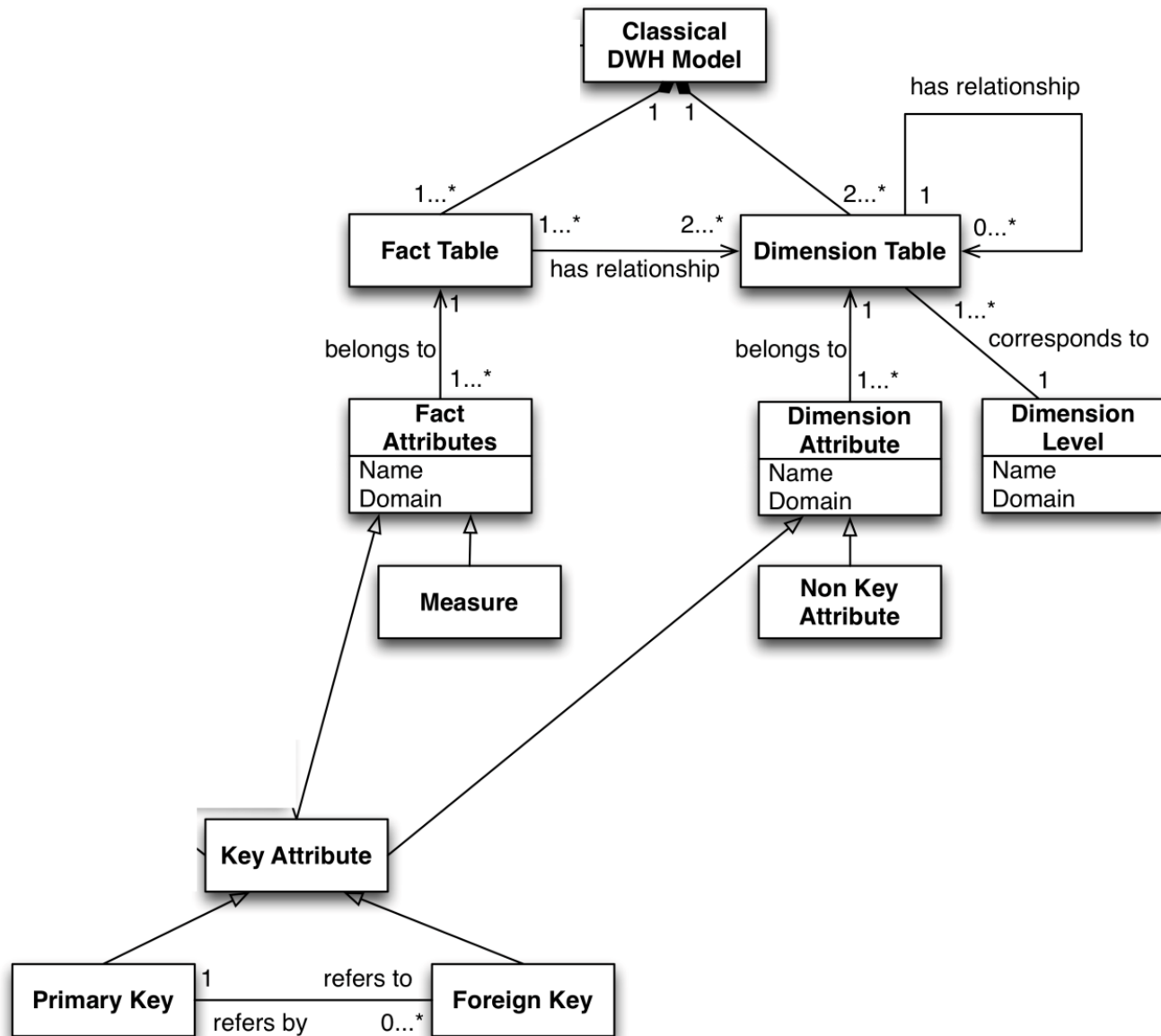
- The complete set of concepts and associations between concepts is called *metamodel*
 - Example



(Horizontal) Abstraction concepts



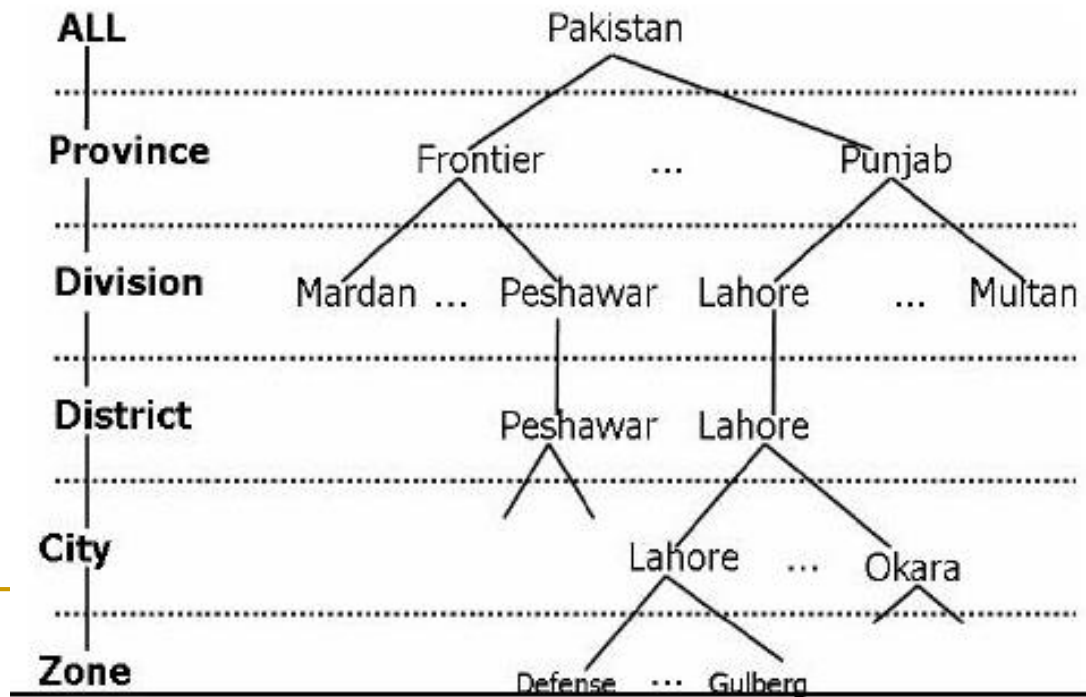
Meta model for DW



Dimensional Hierarchies

Dimensional Hierarchies

- Dimensions have hierarchies
 - Model level representation of dimension
 - Instance level representation of dimension



Dimensional Hierarchies

- Types of dimensional hierarchies
 - Balanced Hierarchies
 - Unbalanced Hierarchies
 - Generalized Hierarchies
 - Alternative Hierarchies
 - Parallel Hierarchies
 - Nonstrict Hierarchies

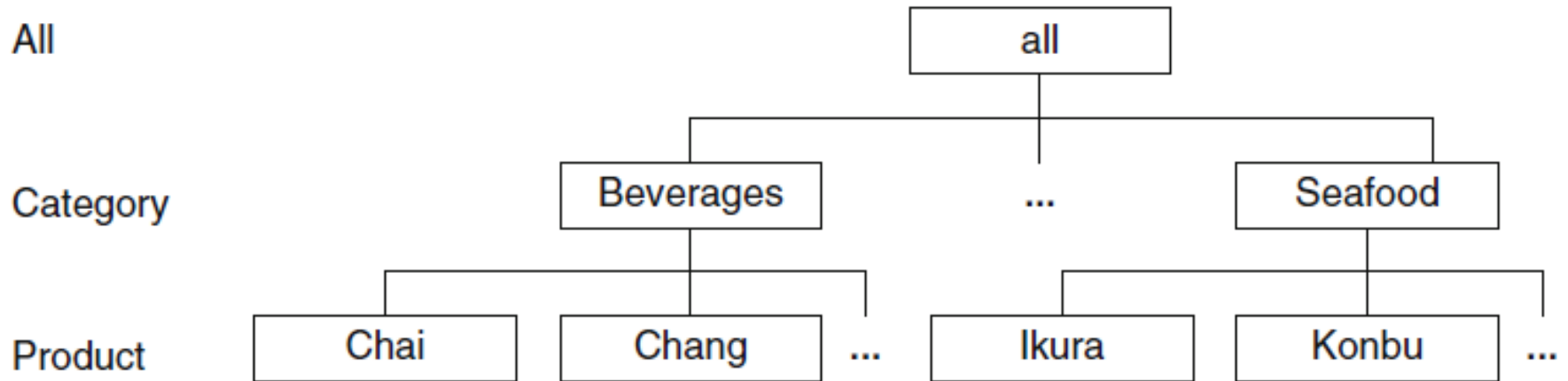
Balanced Hierarchies

■ Balanced Hierarchy

- A balanced hierarchy has only **one path** where all the levels are **mandatory**
- All the branches have the **same length**
- All parent members have **at least one** child members
- A child member belongs **exactly to one** parent member

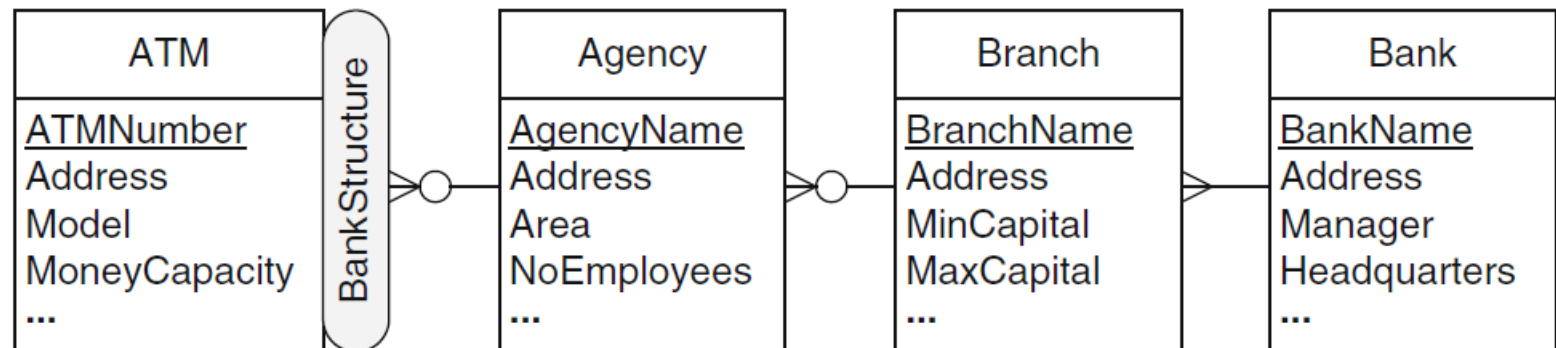
Balanced Hierarchies

■ Balanced Hierarchy (**instance level**)

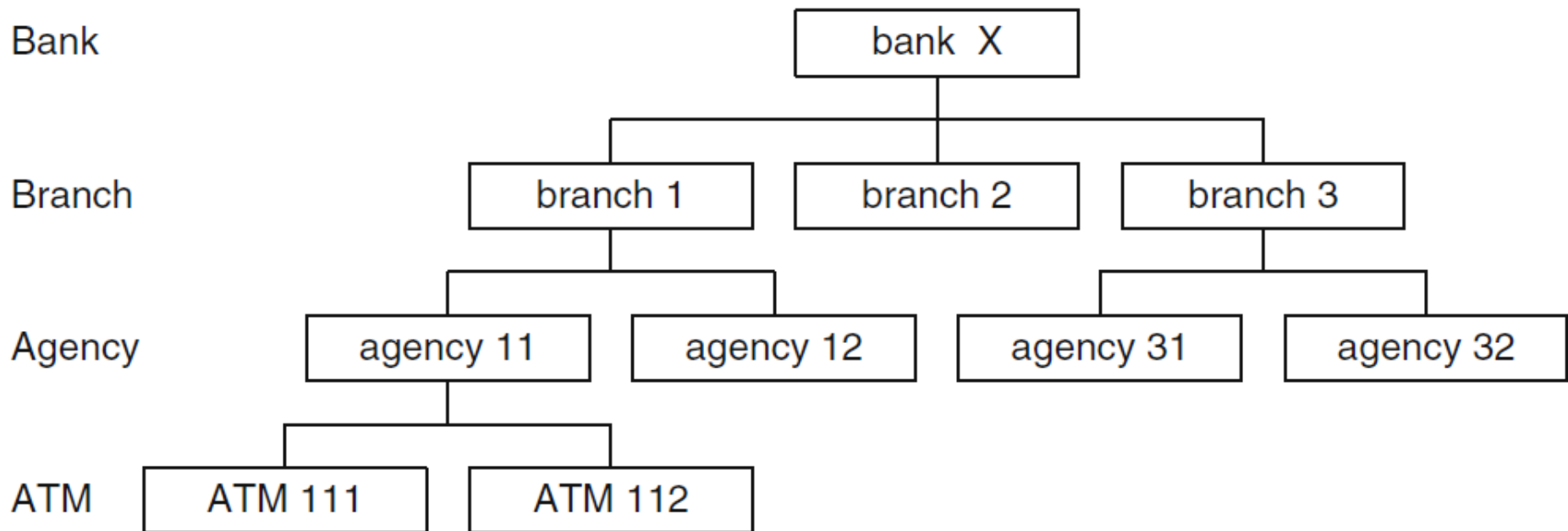


Unbalanced Hierarchies

- Unbalanced Hierarchy (**model level**)
 - Unbalanced hierarchy has only one path where **at least one level is not mandatory**
 - Therefore, at the instance level, there can be parent members **without** associated child members



Unbalanced Hierarchies (Instance level)



Unbalanced Hierarchies

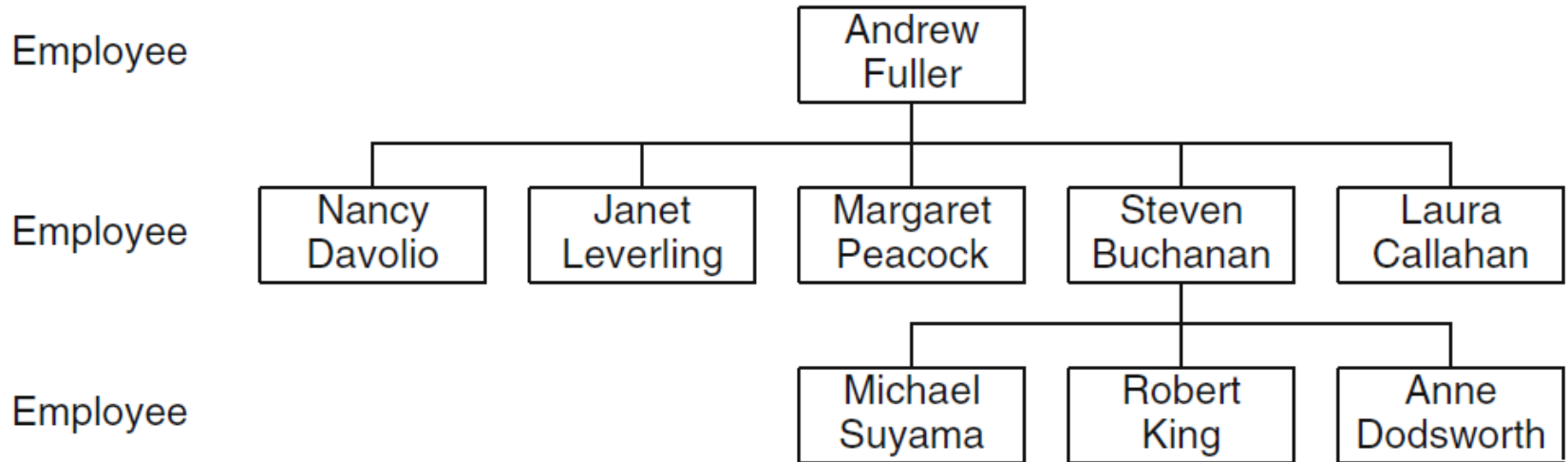
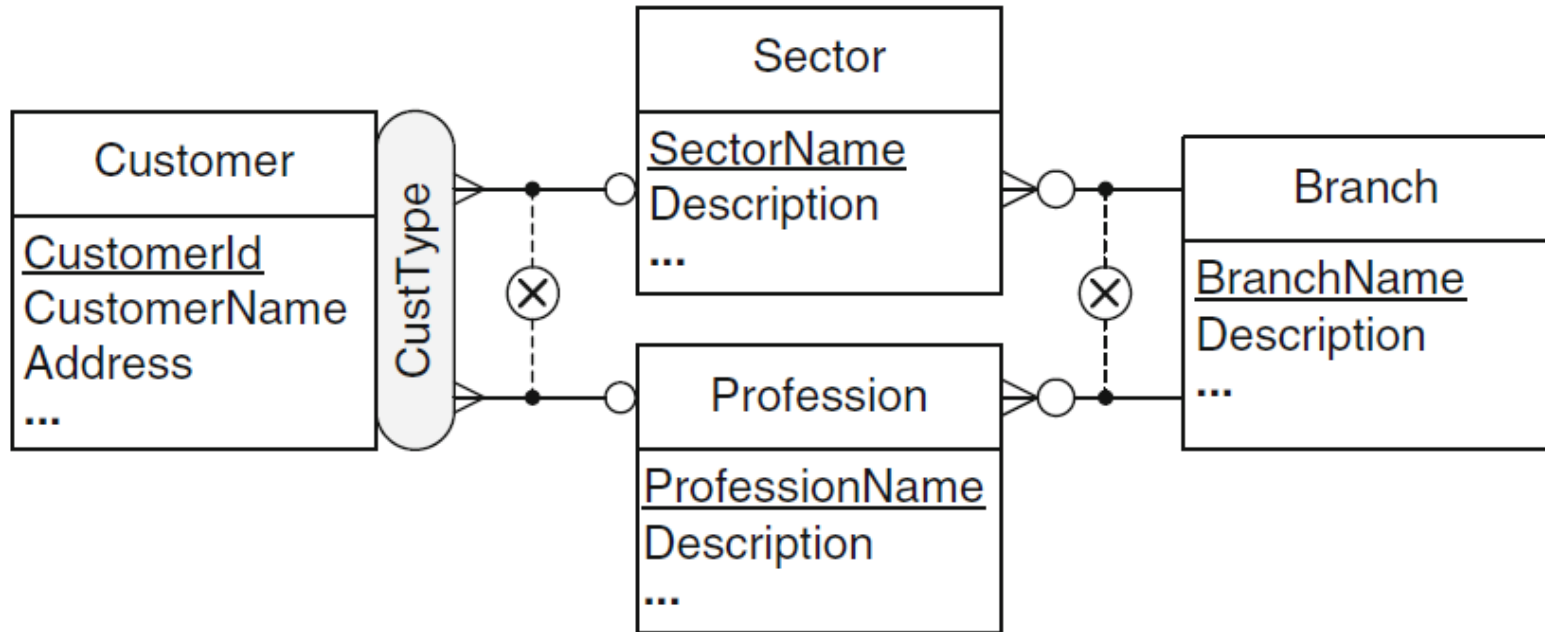


Fig. 4.5 Instances of the parent-child hierarchy in the Northwind data warehouse

Generalized Hierarchies

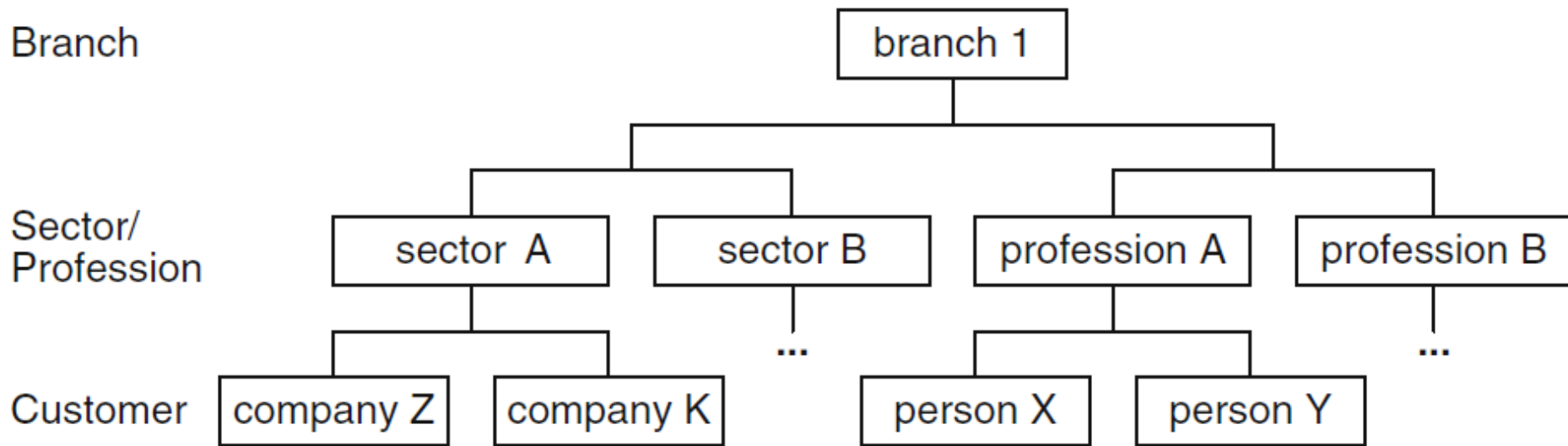
- Hierarchies can be of **different types**
 - Customer can be a **company** or a **person**
- Suppose measures pertaining to customer must be aggregated differently
 - Customer → **Sector** → Branch
 - Customer → **Profession** → Branch
- Such hierarchies are called **generalized hierarchies**

Generalized Hierarchies



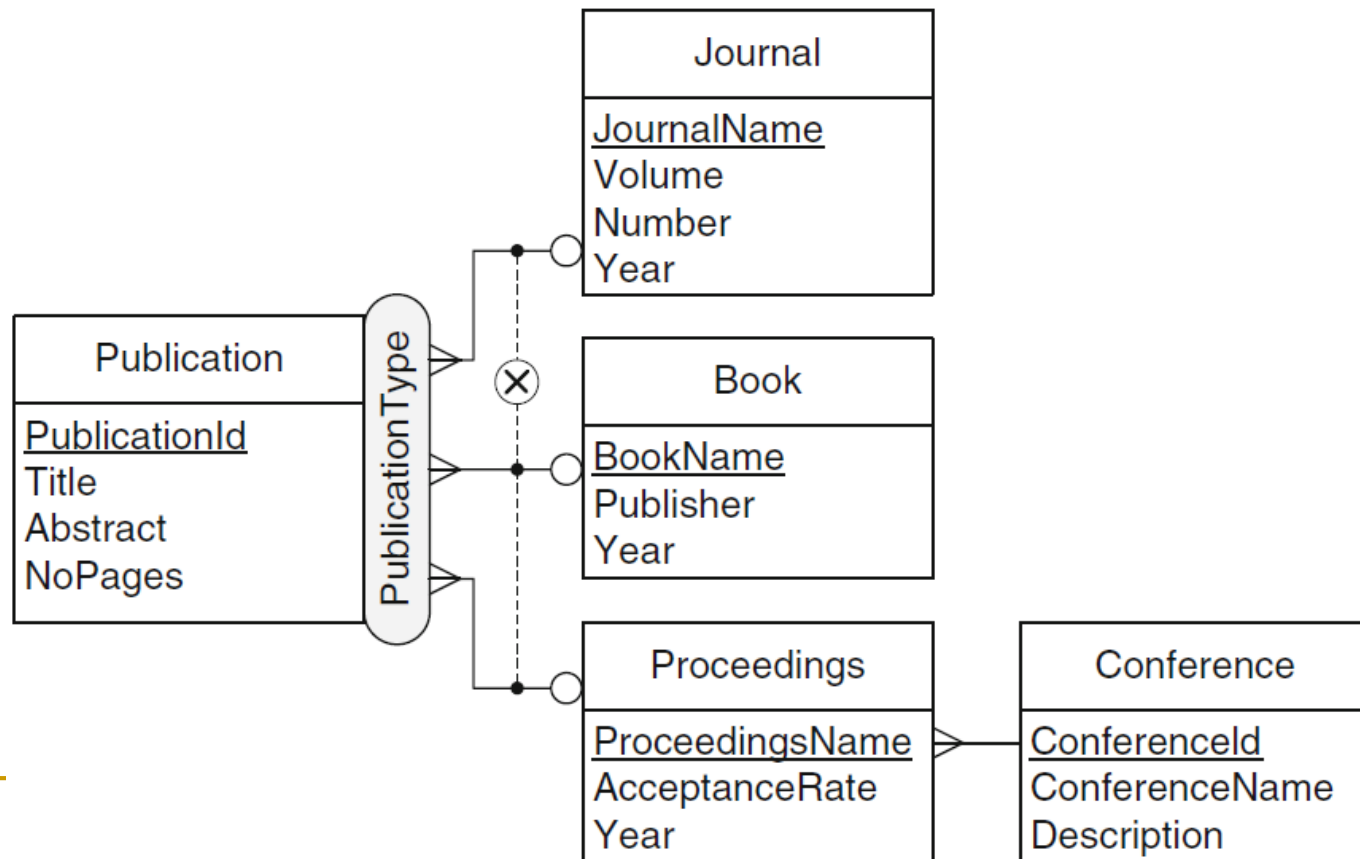
- The rounded symbol of exclusivity indicates that **paths are exclusive** for **every** member
 - The levels at which alternative paths split and join are called, the **splitting** and **joining** levels

Generalized Hierarchies



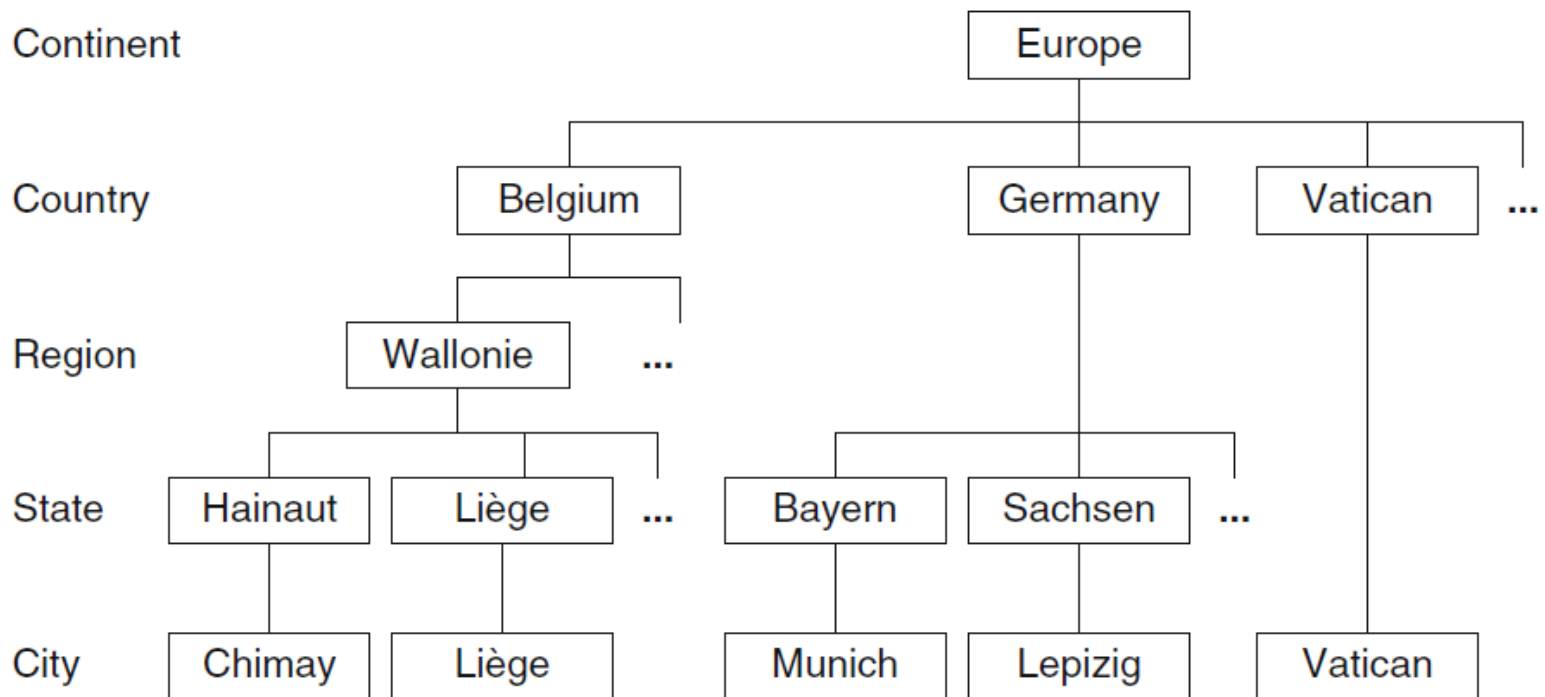
Generalized Hierarchies

- Example generalized hierarchy **without joining level**



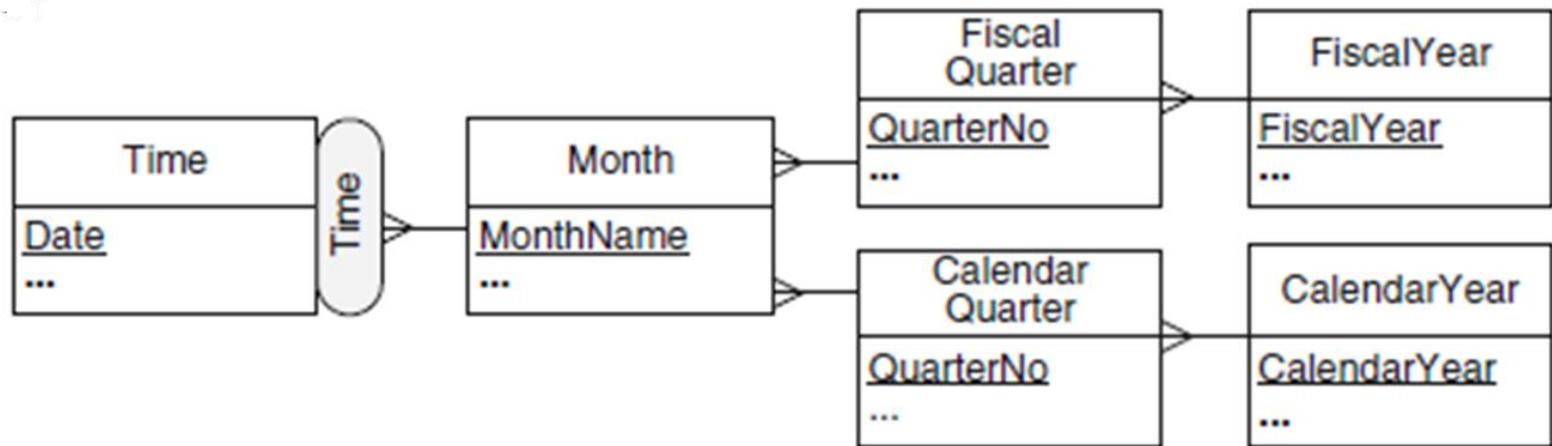
Generalized Hierarchies

- Generalized hierarchies include a special case, **ragged hierarchies**
 - Some countries, such as Belgium, are divided into regions, such as Germany, are not

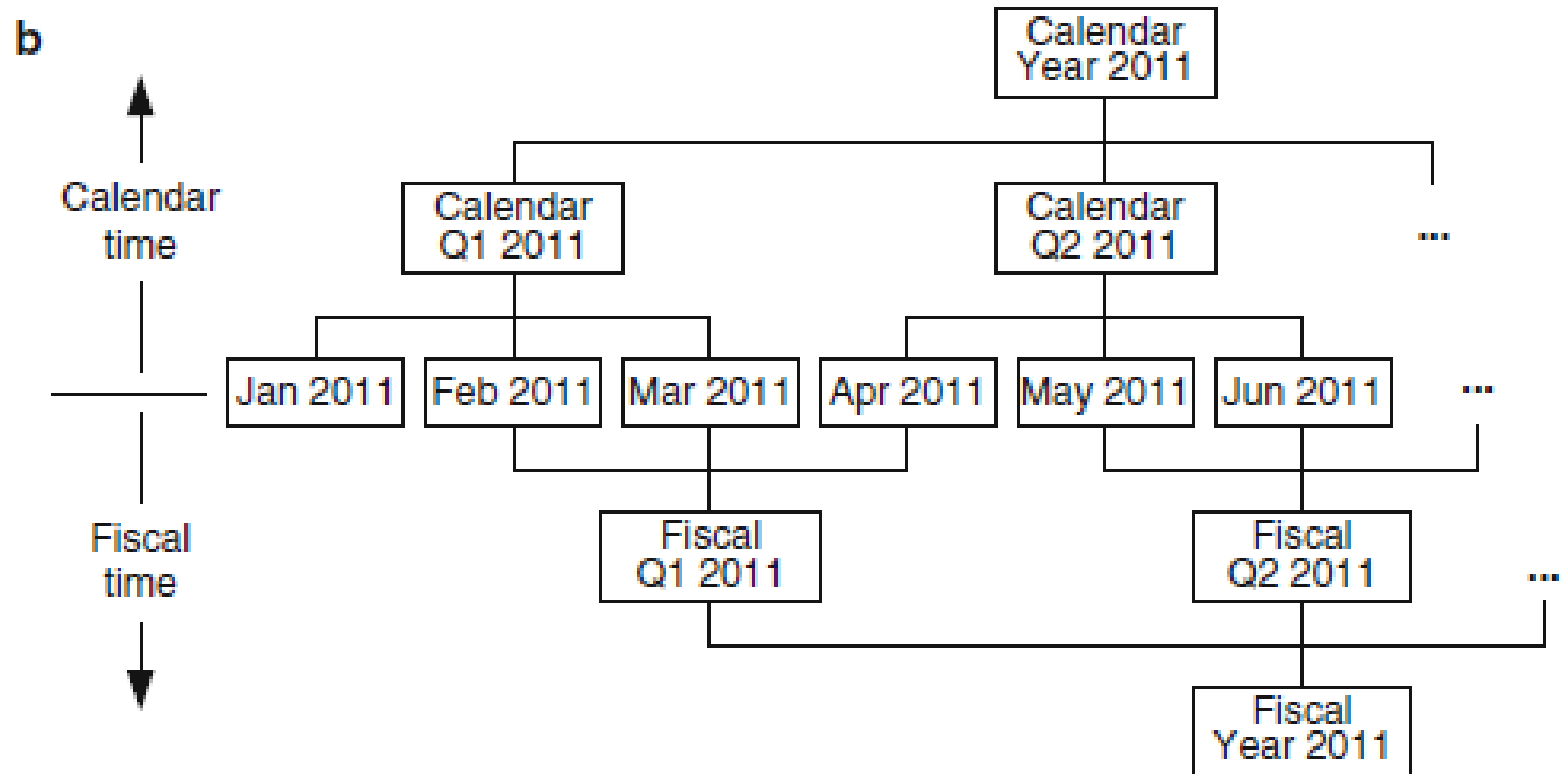


Alternative Hierarchy

- Represents the situation where there are several nonexclusive **hierarchies that share at least the leaf level**



Alternative Hierarchy

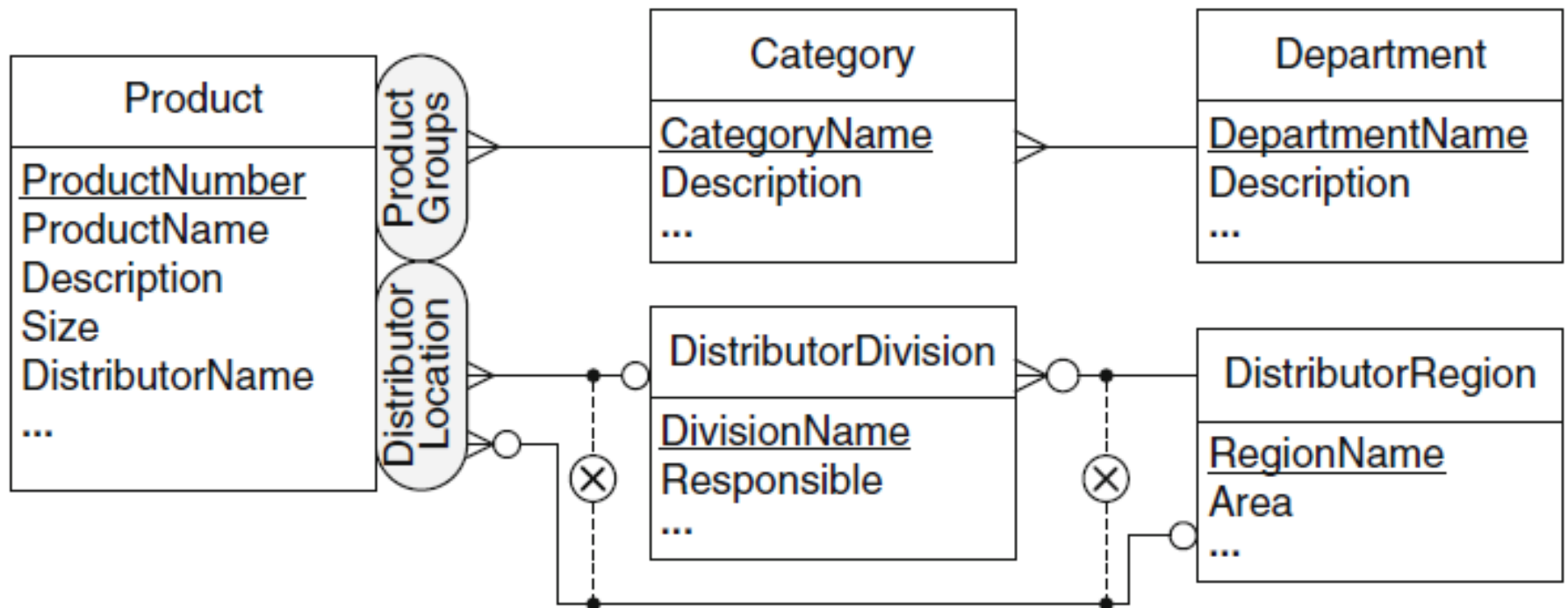


Parallel hierarchies

- Two types of parallel hierarchies depending upon the whether they **share level**
 - Dependent parallel hierarchies
 - Independent parallel hierarchies

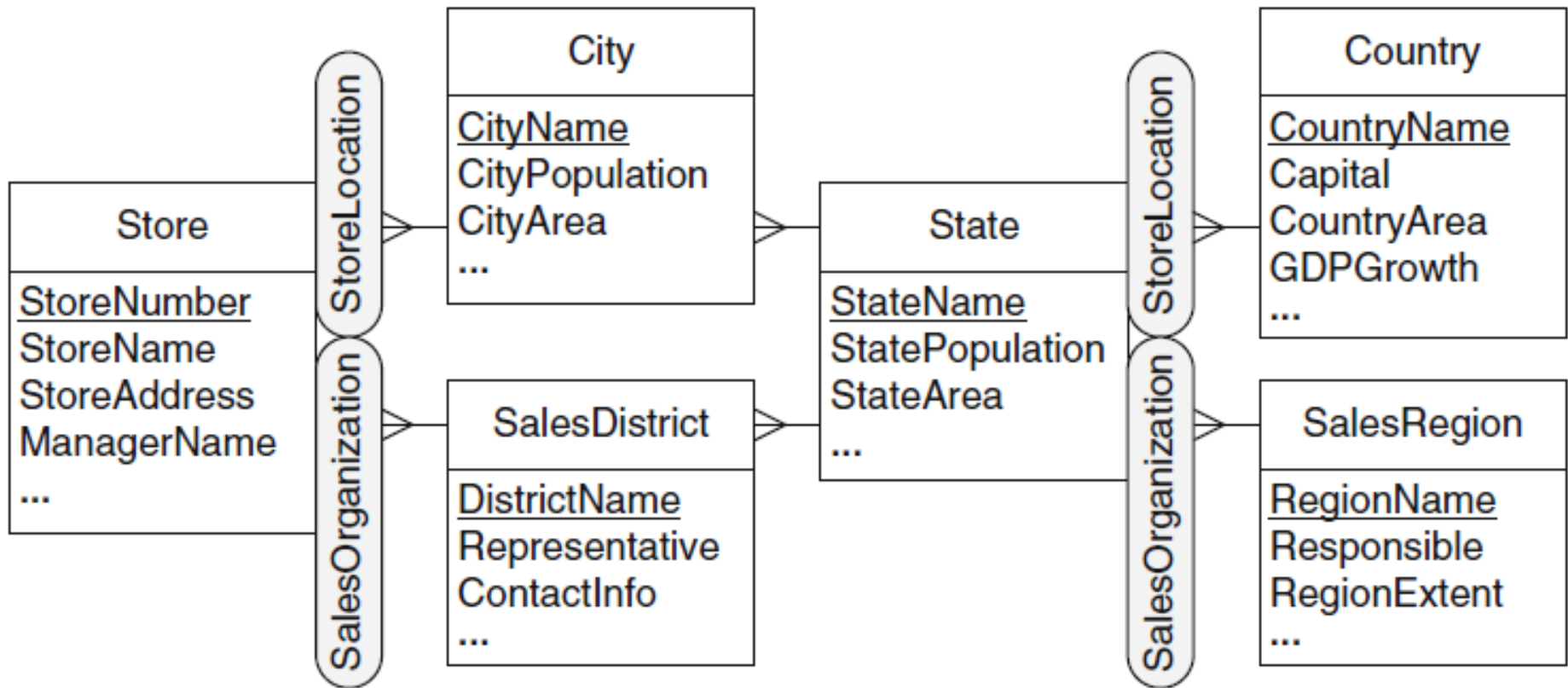
Parallel Independent hierarchies

- Dimensions have two independent hierarchies



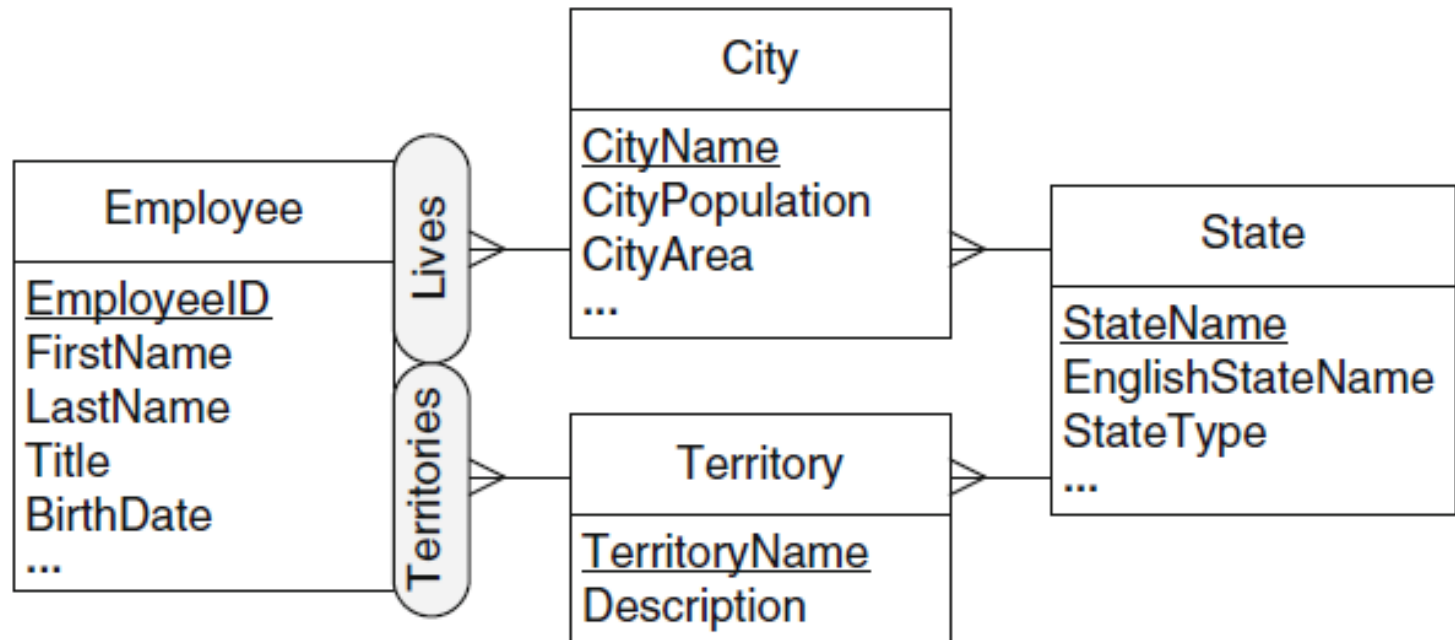
Parallel Dependent hierarchies

- Hierarchy levels are dependent on each other



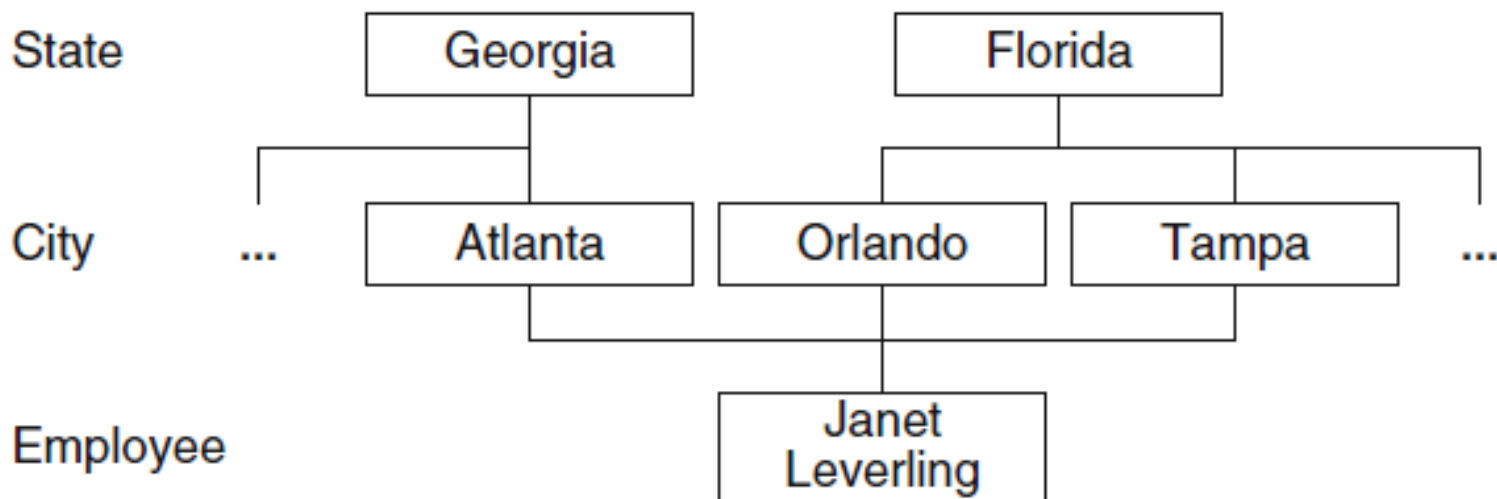
Parallel Dependent hierarchies

- Parallel dependent hierarchies leading to different parent members of shared level



Nonstrict Hierarchies

- Many-to-many relationship between parent and child level are common in real life
- A hierarchy that has **at least one many-to-many** relationship is called **nonstrict**
 - Otherwise, it is called **strict** hierarchy



Strict vs Nonstrict Hierarchies

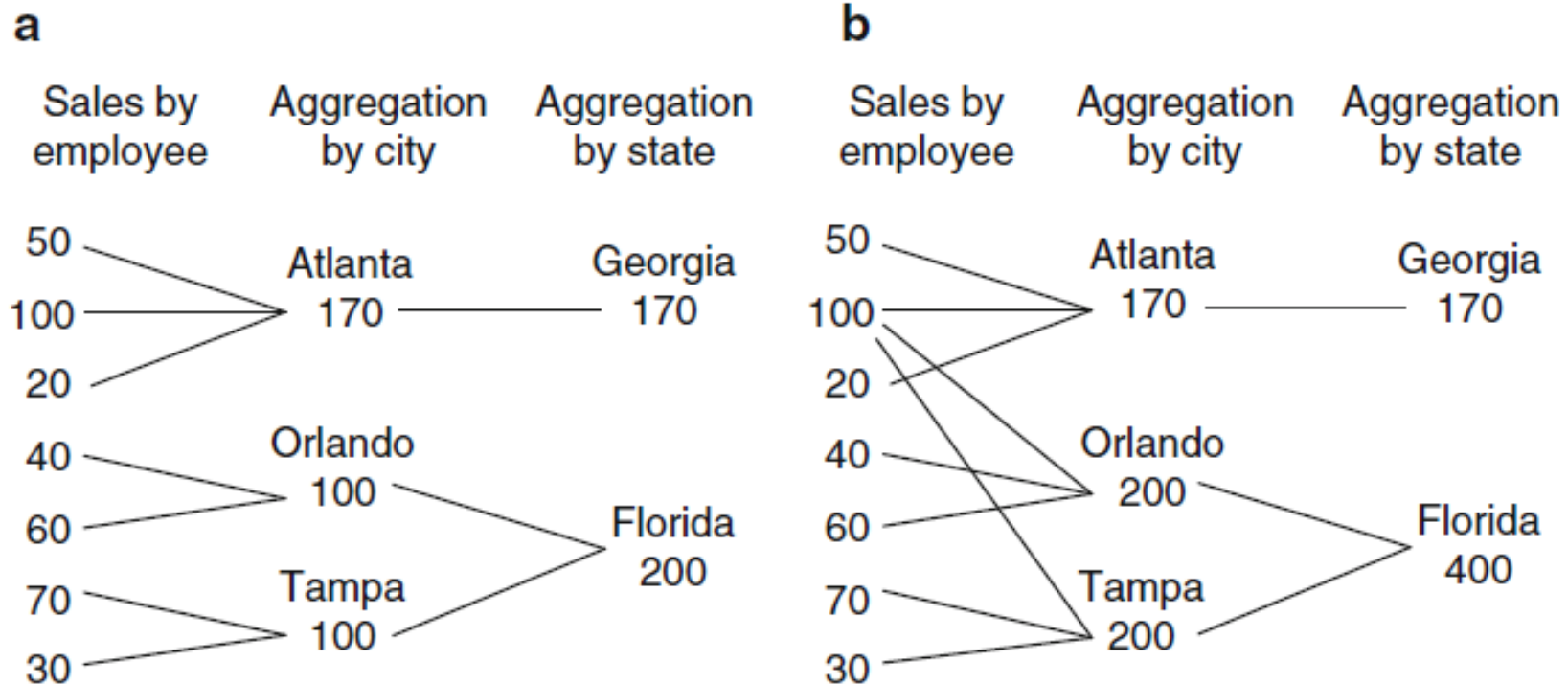
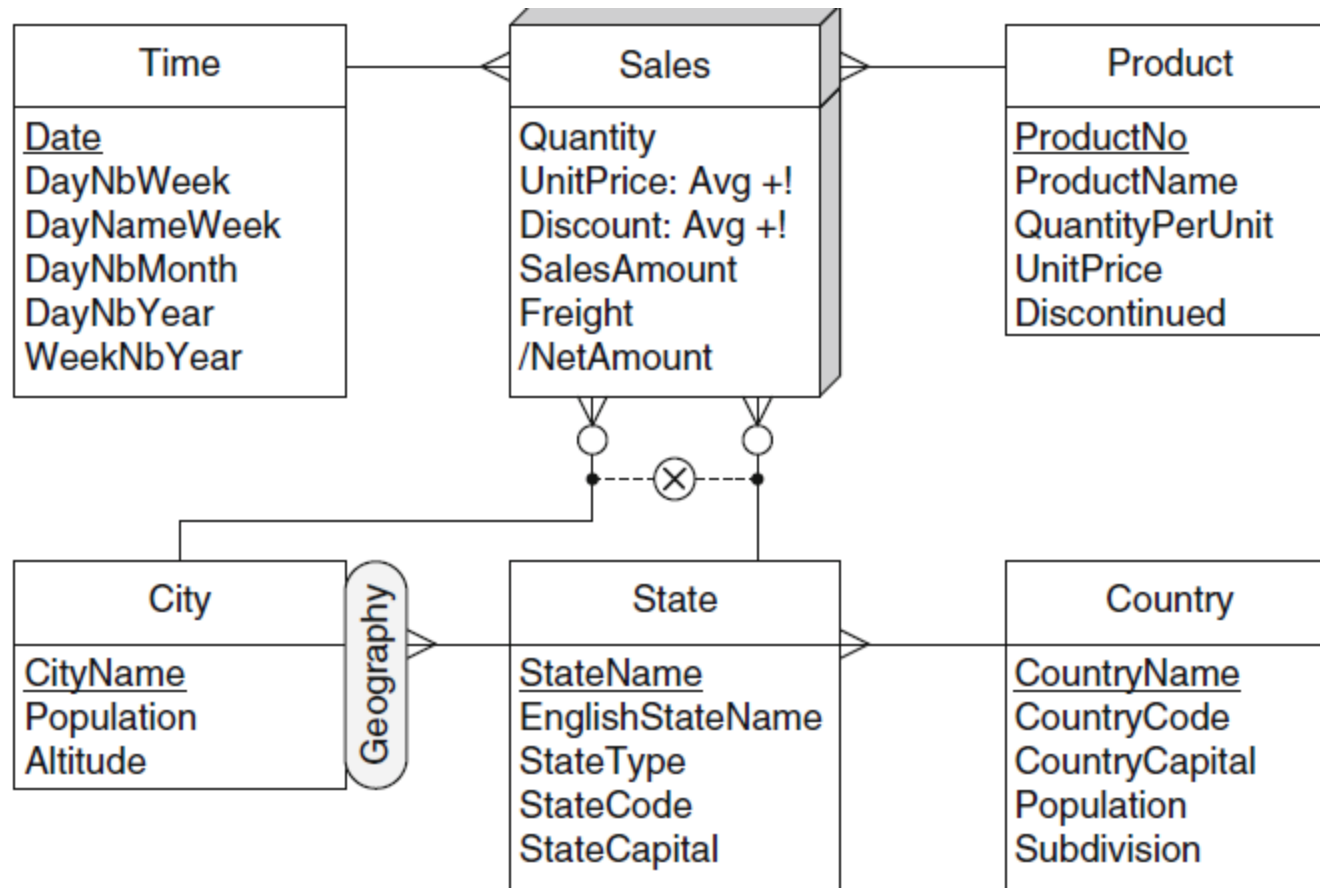
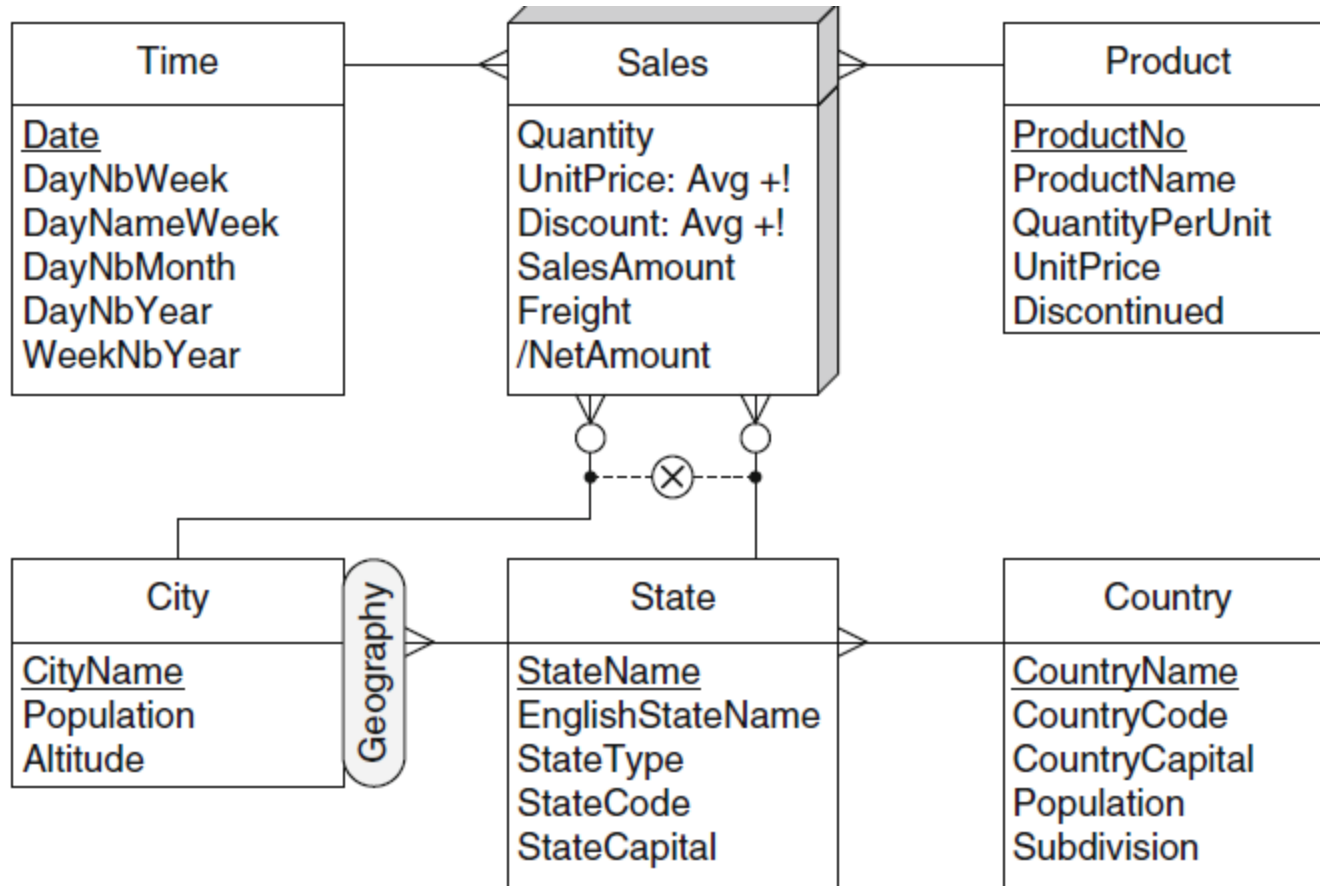


Fig. 4.14 Double-counting problem when aggregating a sales amount measure in Fig. 4.13. (a) Strict hierarchy. (b) Nonstrict hierarchy

Multiple granularities for the sales fact



Multiple granularities for the sales fact



Many-to-Many Dimensions

