DS-306 Data Warehousing and Business Intelligence

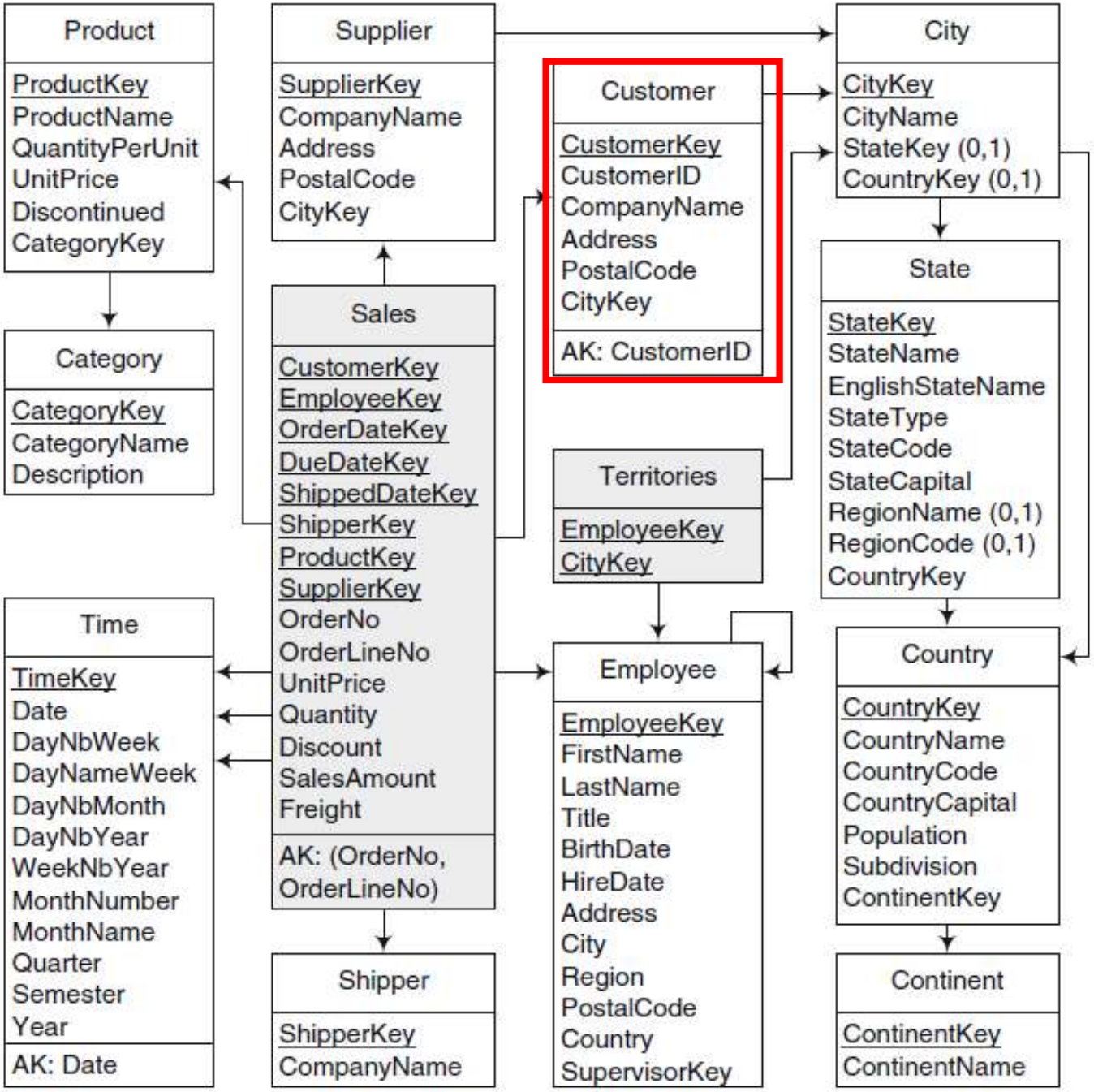
Topic 4: Logical DW Design

Dr. Khurram Shahzad

Two possibilities of keys

- There are two possibilities of defining the keys of dimension levels
  - Generating Surrogate keys
  - Keeping DB key as DW key

Surrogate key is added to Customer



Implications of rules

- In all rules, surrogate keys can be generated for each dimension level
- Reasons: Provide independence from keys of the underlying sources
- That is, underlying keys can change across time
- Also, Surrogate keys are integers that increases efficiency compared to strings

Conceptual to Logical

Relational Implementation

- Set of rules are applied to translate conceptual model (MultiDim model) to relational model

Rule 1

- A **level L**, provided that it is **not related to a fact** with **one-to-one relationship**, is mapped to table T_L that contains all attributes of the level
- A **surrogate key** may be added
  - Depending upon the type of changes
- Or, the **identifier** of the level will be the key of the table

Rule 2

- A **fact F** is mapped to a table T_F that includes as attributes all measures of the fact
- Further, a **surrogate key** may be added to the table

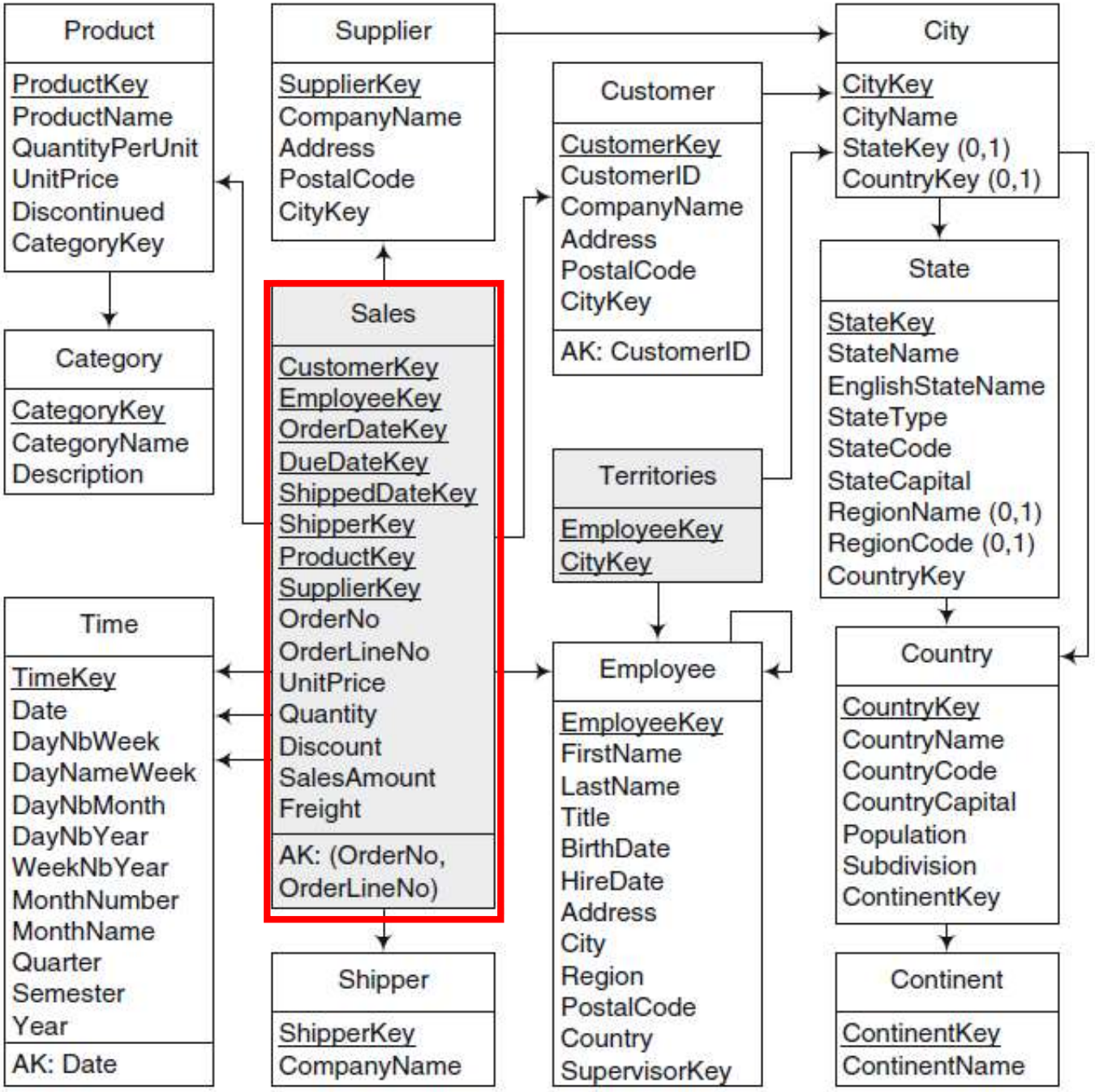
Rule 3

- A relationship **between** either a **fact F** and a **dimension level L**, or between dimension level $\mathbf{L_P}$ and $\mathbf{L_C}$ (parent child) can be mapped in three different ways
  - **One-to-one relation**
  - **One-to-many relation**
  - **Many-to-many**

Rule 3a

- **Rule 3a:** If the relationship is one-to-one, the table corresponding to the fact T_F or to the child T_C is extended with all the attributes of dimension level or the parent level

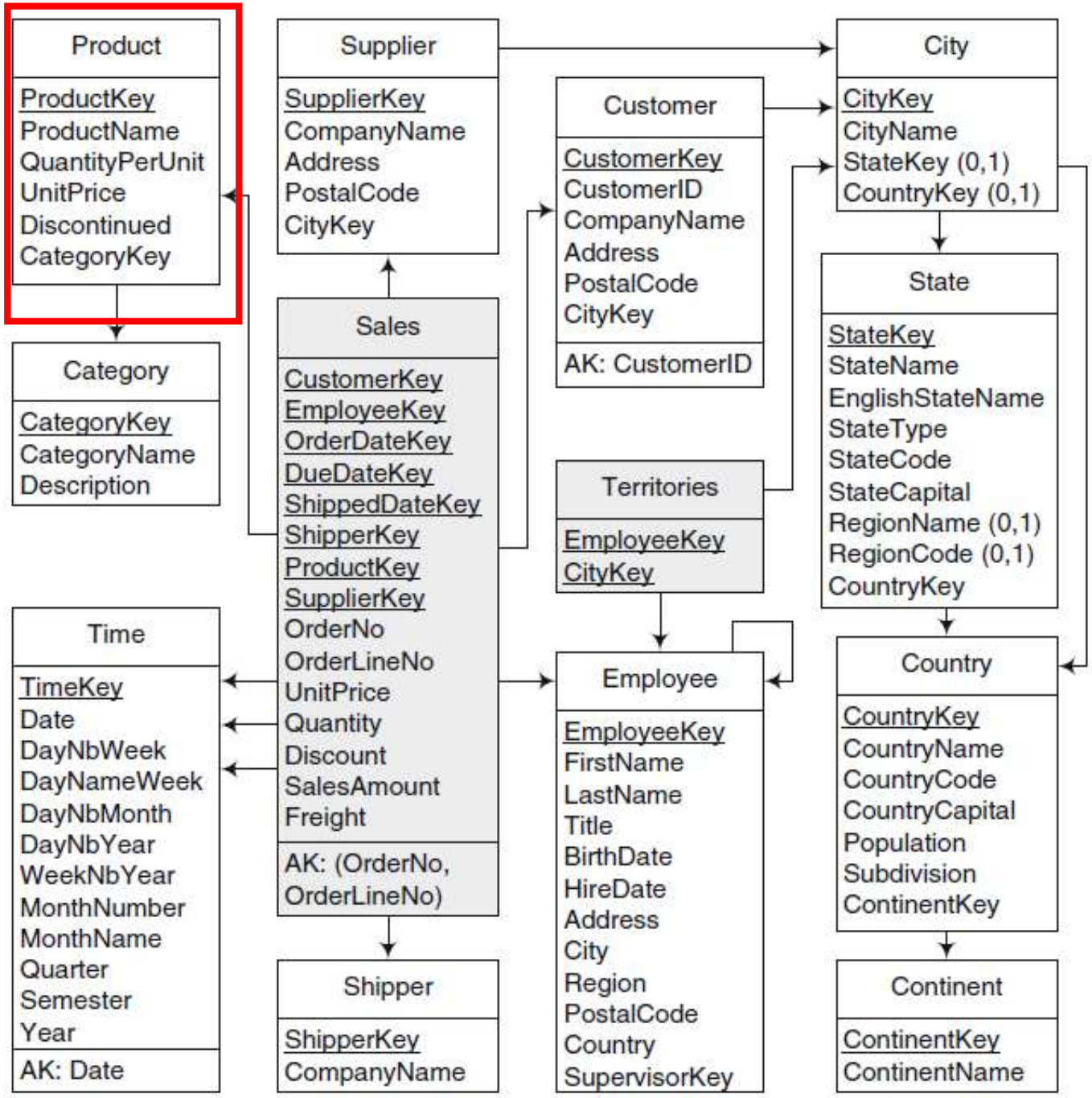
OrderNo
OrderlineNo
are dimensions but included in fact table



Rule 3b

- **Rule 3b:** If the relationship is one-to-many, the **table corresponding to the fact** T_F or to the child level T_C **is extended with the surrogate** key of the table corresponding to the dimension level T_L or the parent level
- That is, a foreign key in the fact or child table pointing to the other table

Product, Supplier
Have 1:M relationship



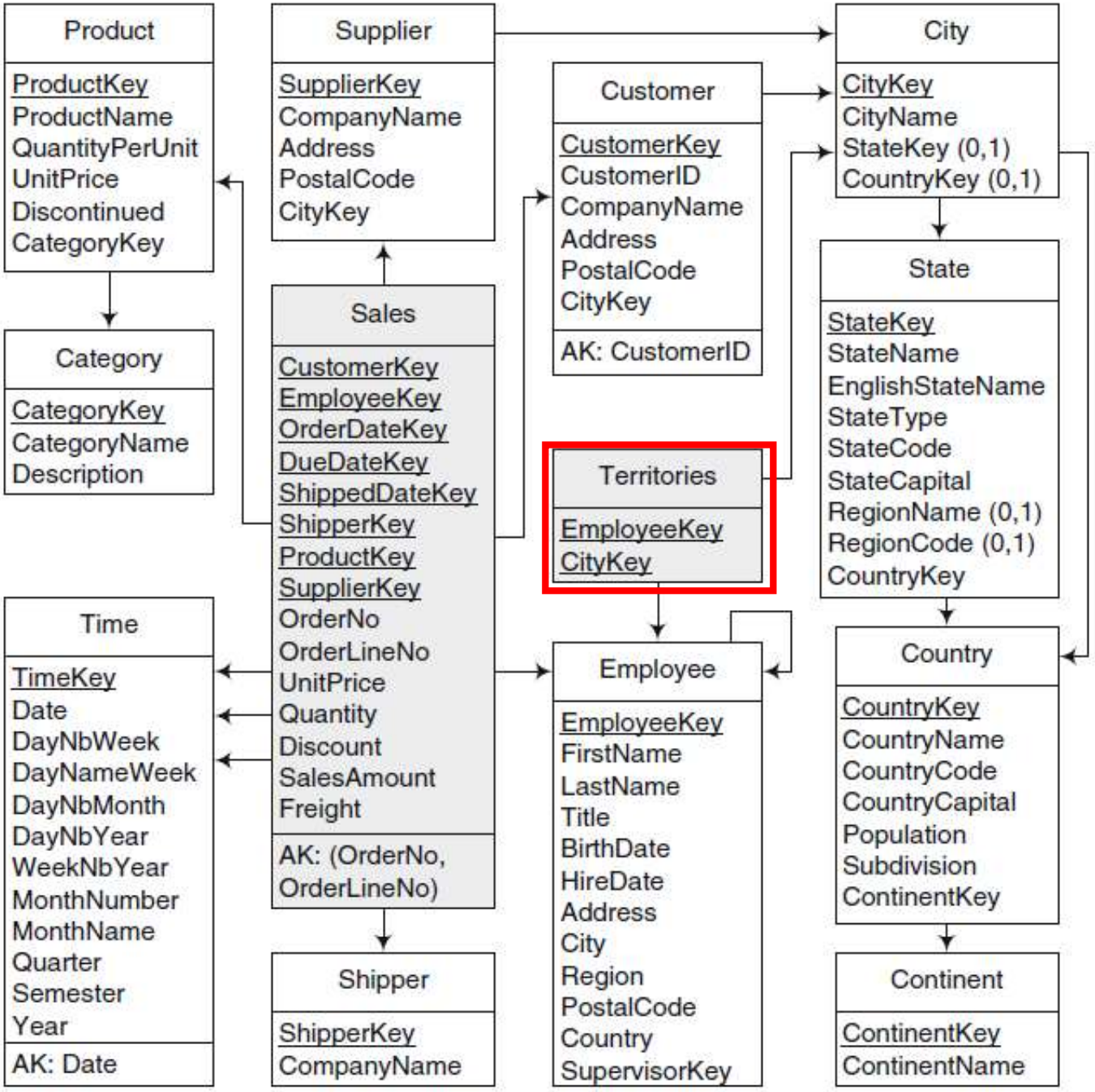
Rule 3c

- **Rule 3c:** If the relationship is many-to-many, a new table T_B (bridge table) is created that contains as attributes the surrogate keys to the corresponding to the fact T_F and the dimension level T_L, or the parent T_P and child level T_C
- The key of the table is the combination of both surrogate keys

Many-to-Many Relationship

- Many-to-many relationship parent-child relationship between Employee and territory is mapped to the table territories containing two foreign keys

Employee, City have M:M relation
New table is added



Time Dimension

- DW is historical DB. So, time dimension is present in every DW
- Time information is included both as foreign keys in a fact table
- Contains aggregation level in which facts can be aggregated across time
- This information is defined from DATE

Logical representation of hierarchies

Recall, Dimensional Hierarchies

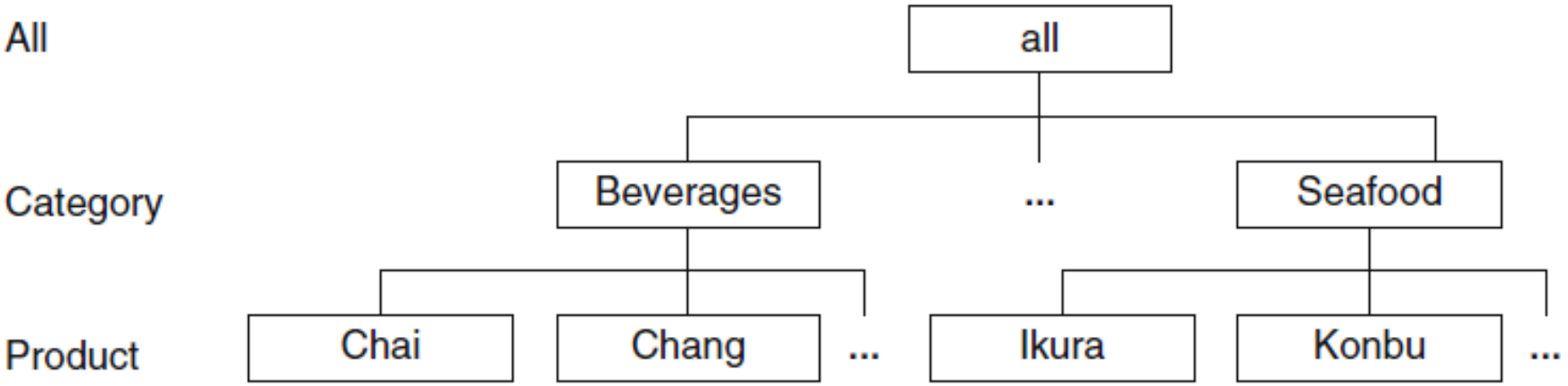
- Types of dimensional hierarchies
  - Balanced Hierarchies
  - Unbalanced Hierarchies
  - Generalized Hierarchies
  - Alternative Hierarchies
  - Parallel Hierarchies
  - Nonstrict Hierarchies

Balanced Hierarchies

- **Recall,**
  - A balanced hierarchy has only **one path** where all the levels are **mandatory**
  - All the branches have the **same length**
- Levels of dimension hierarchies are represented indepdentently
- Are linked by parent-child relationships

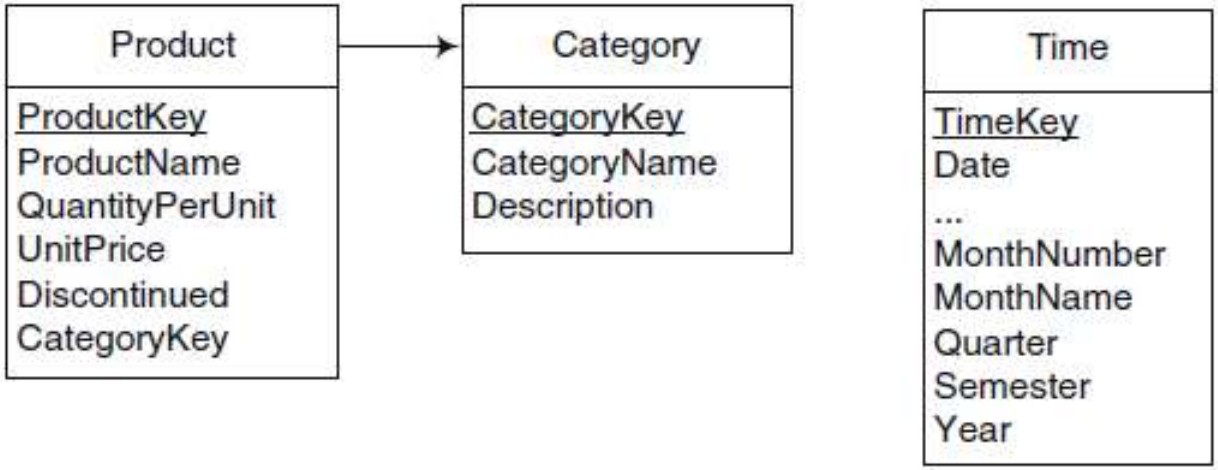
Balanced Hierarchies

- Balanced Hierarchy (**instance level**)



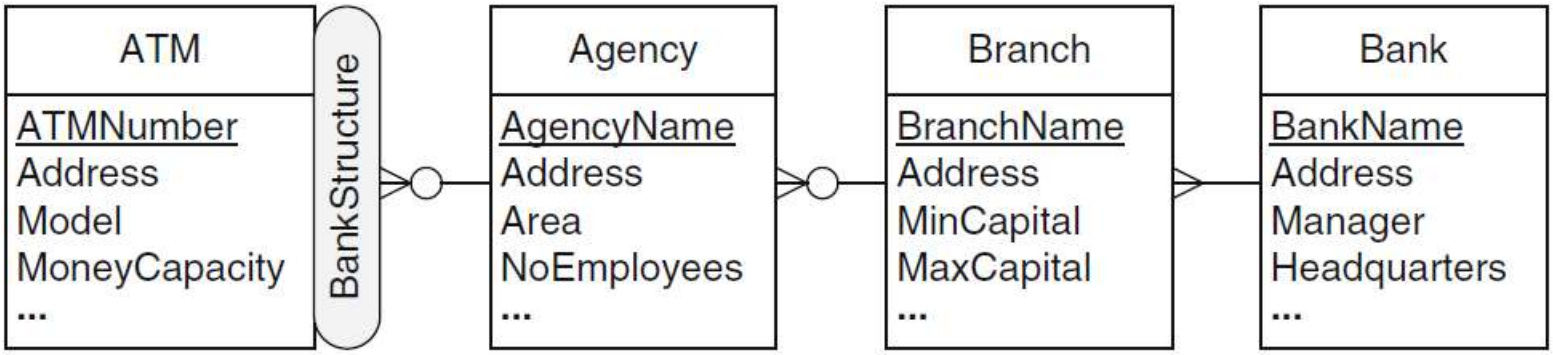
Balanced Hierarchies: **Logical**

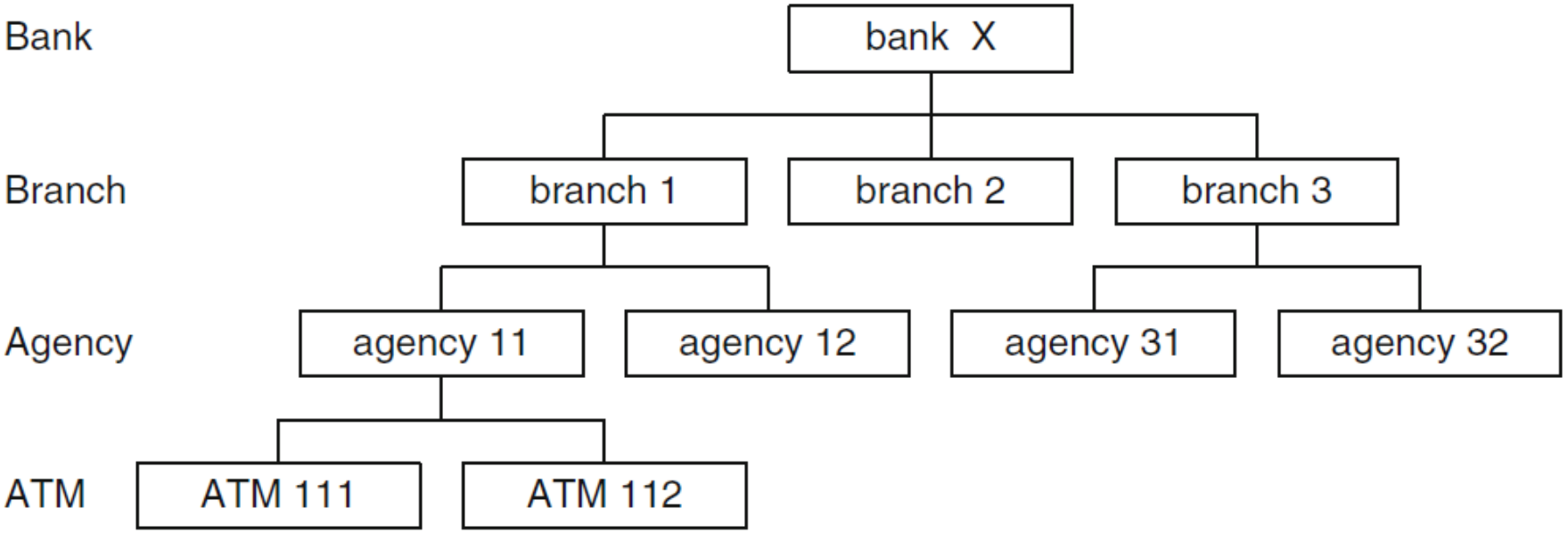
- Example snowflake structure on left
- Flat table on right (denormalized table), if star is required
  - May contain several levels



g. 5.5 Relations for a balanced hierarchy. (a) Snowflake structure. (b) Flat table

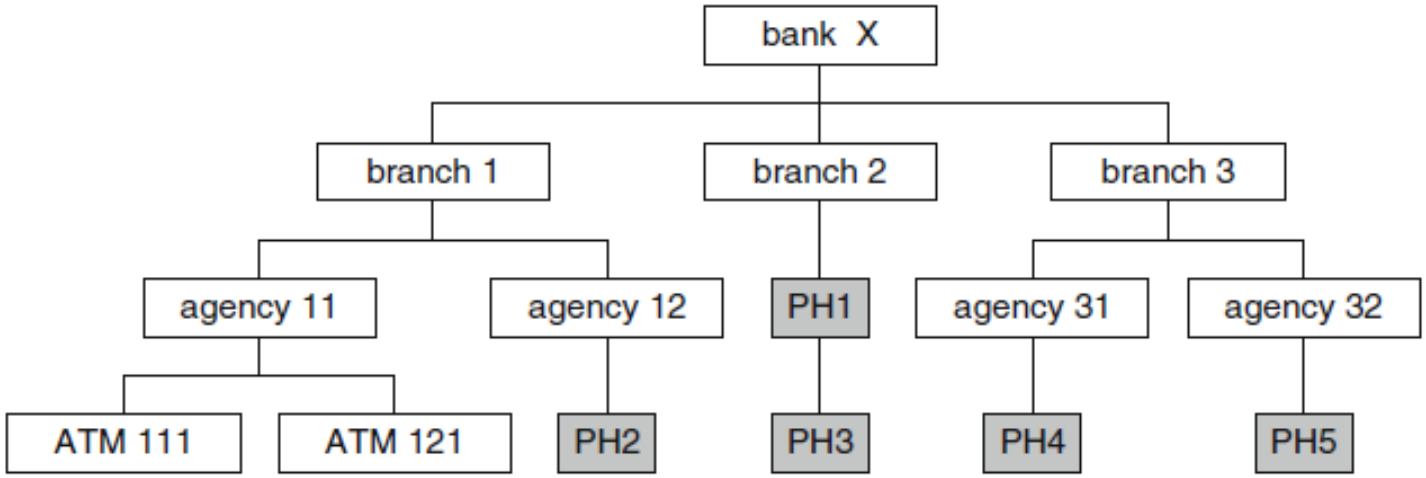
Unbalanced Hierarchies: **Conceptual**



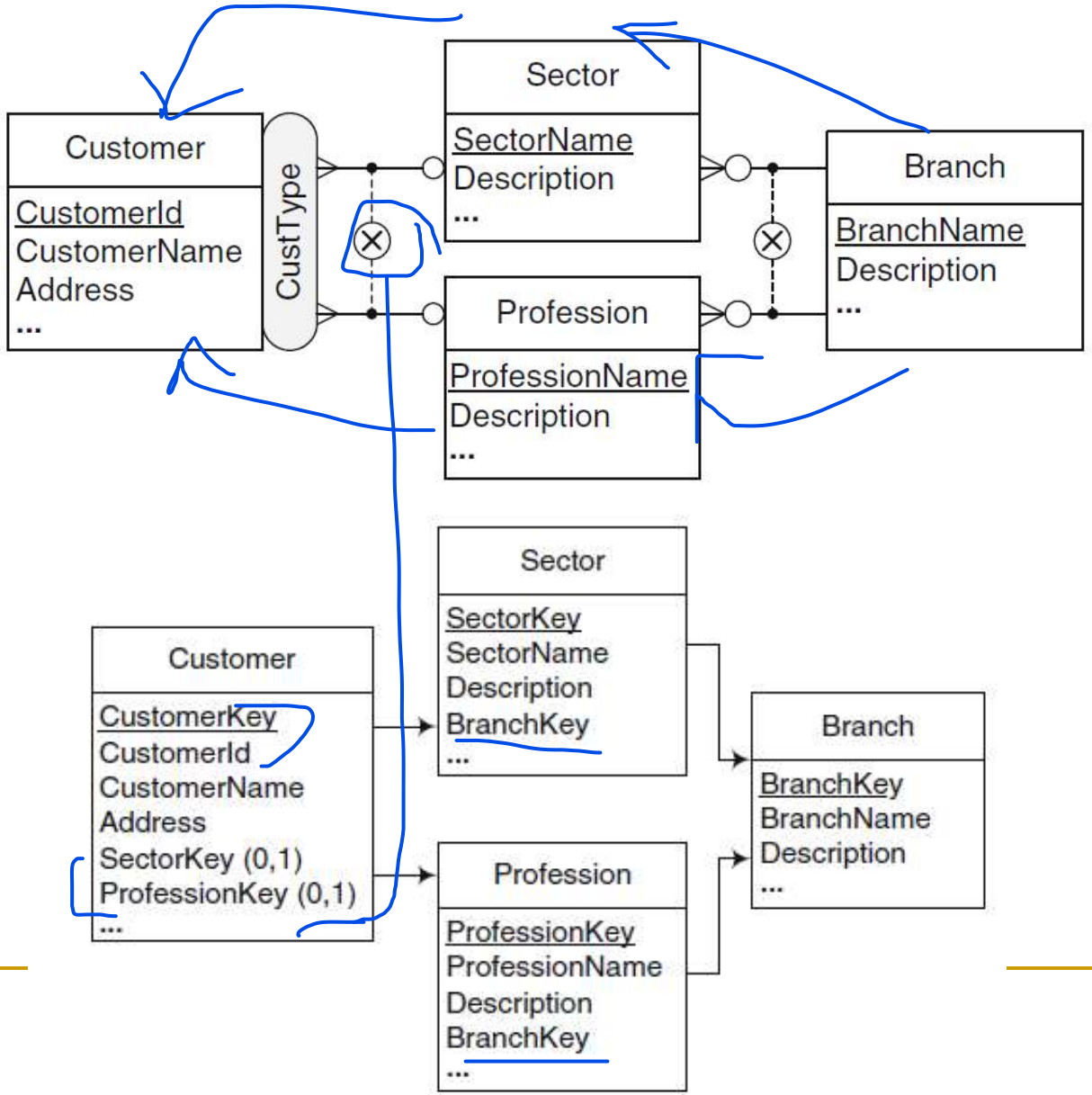


Unbalanced Hierarchies: **Logical**

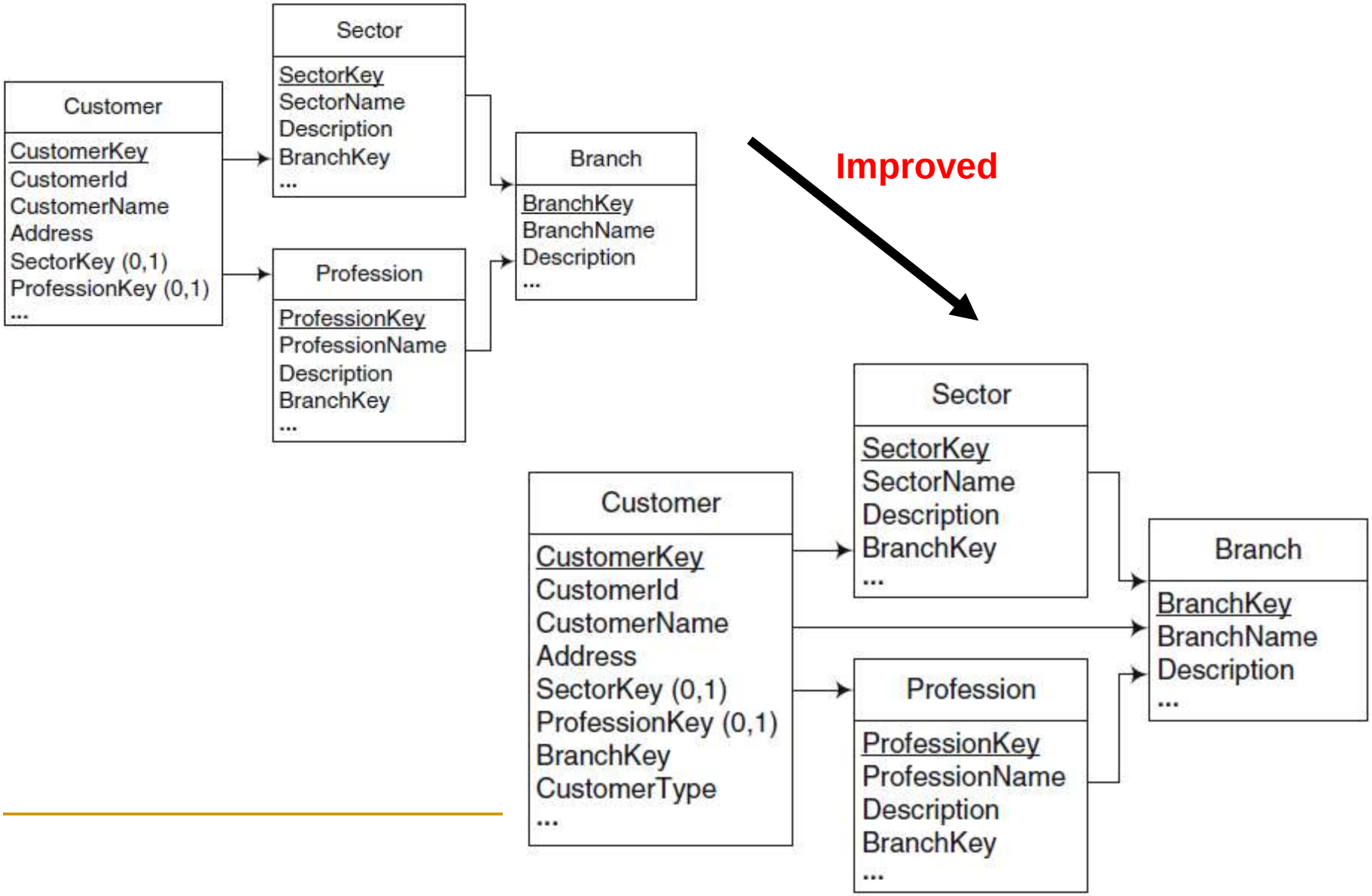
- **Problem:** Measures are associated with ATMs
  - Measures will aggregate only for those agencies that have ATMs
- **Solution:** Transform unbalanced hierarchy to balanced using placeholders



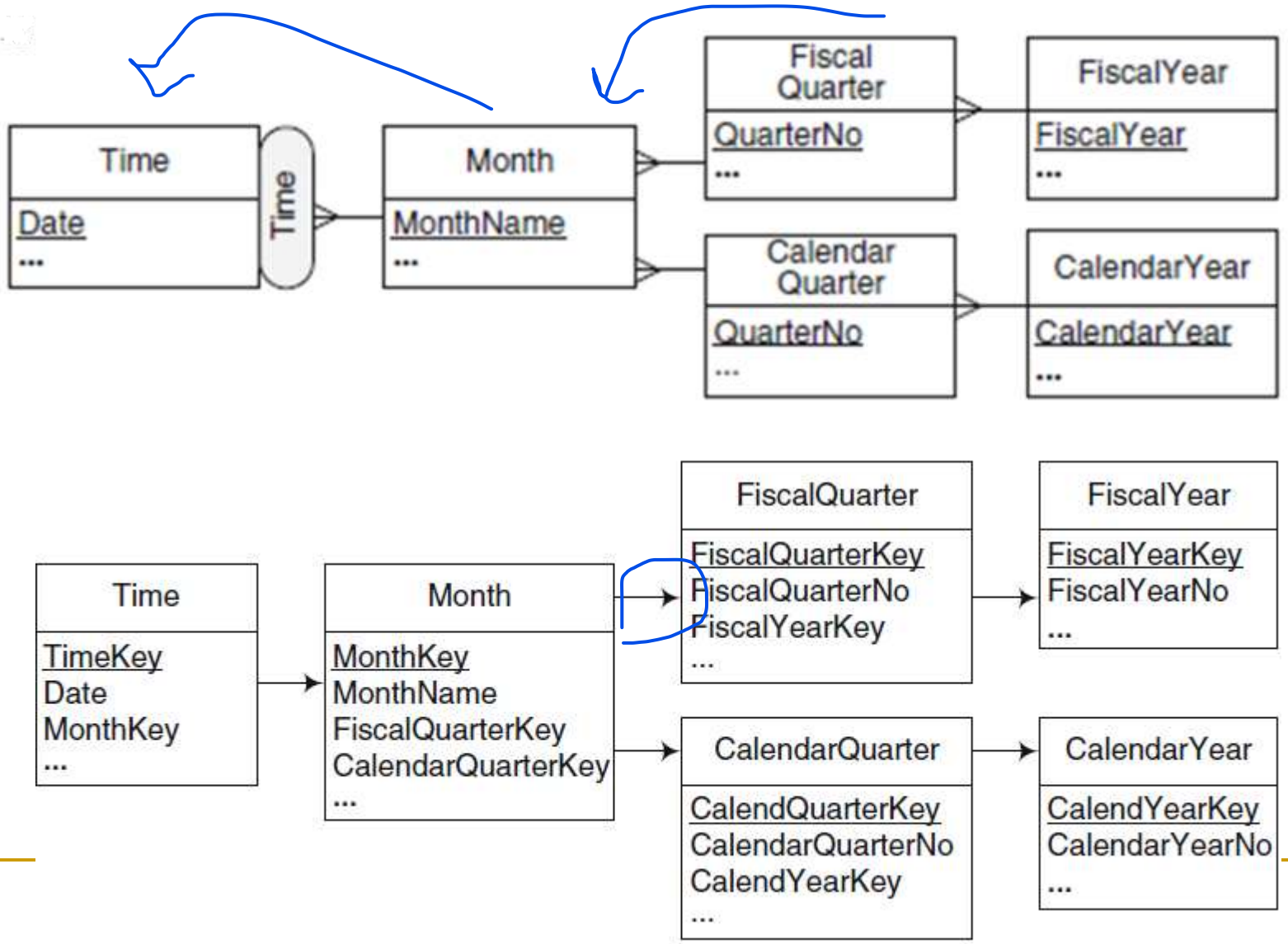
Generalized Hierarchies: **Logical**



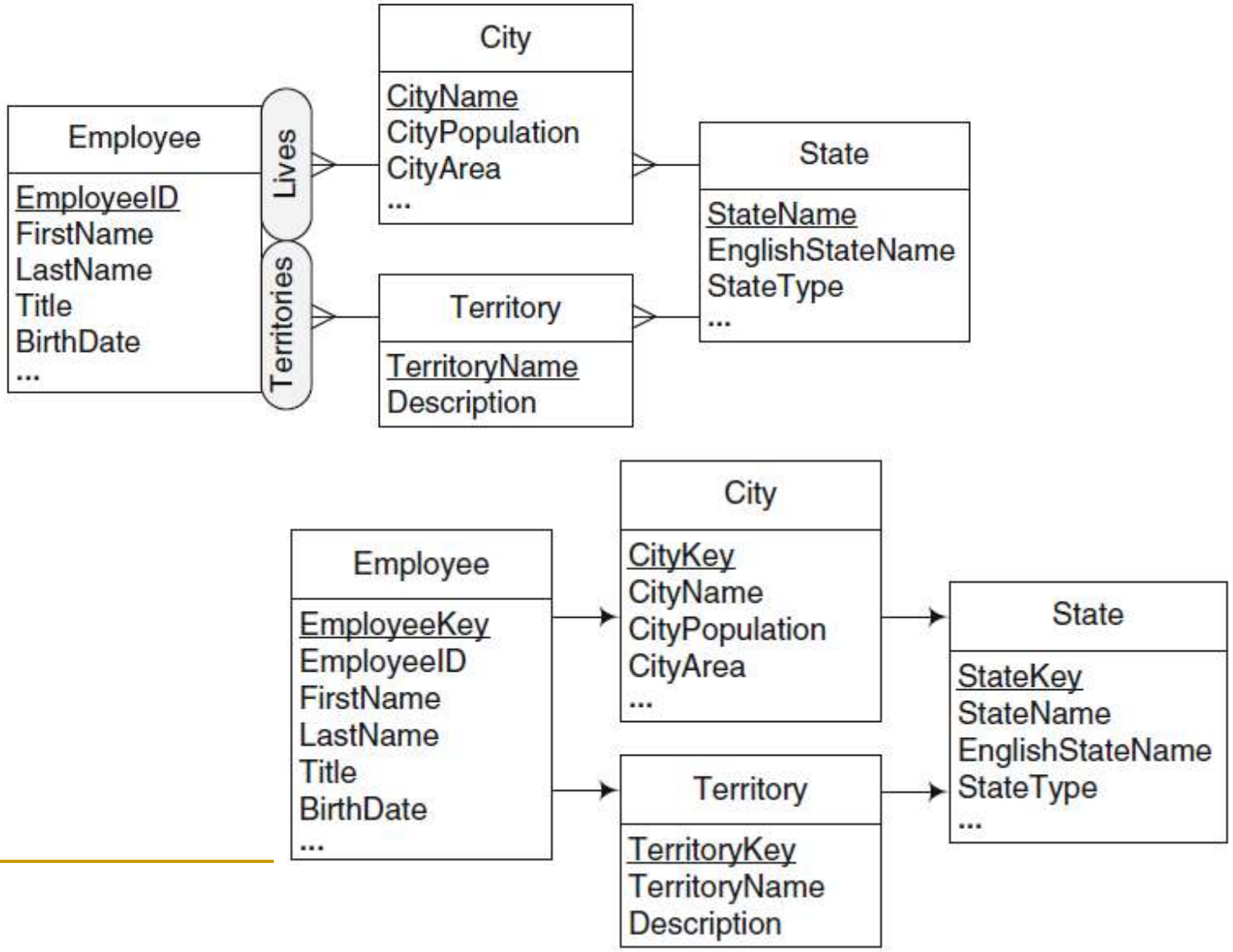
Generalized Hierarchies: **Logical**



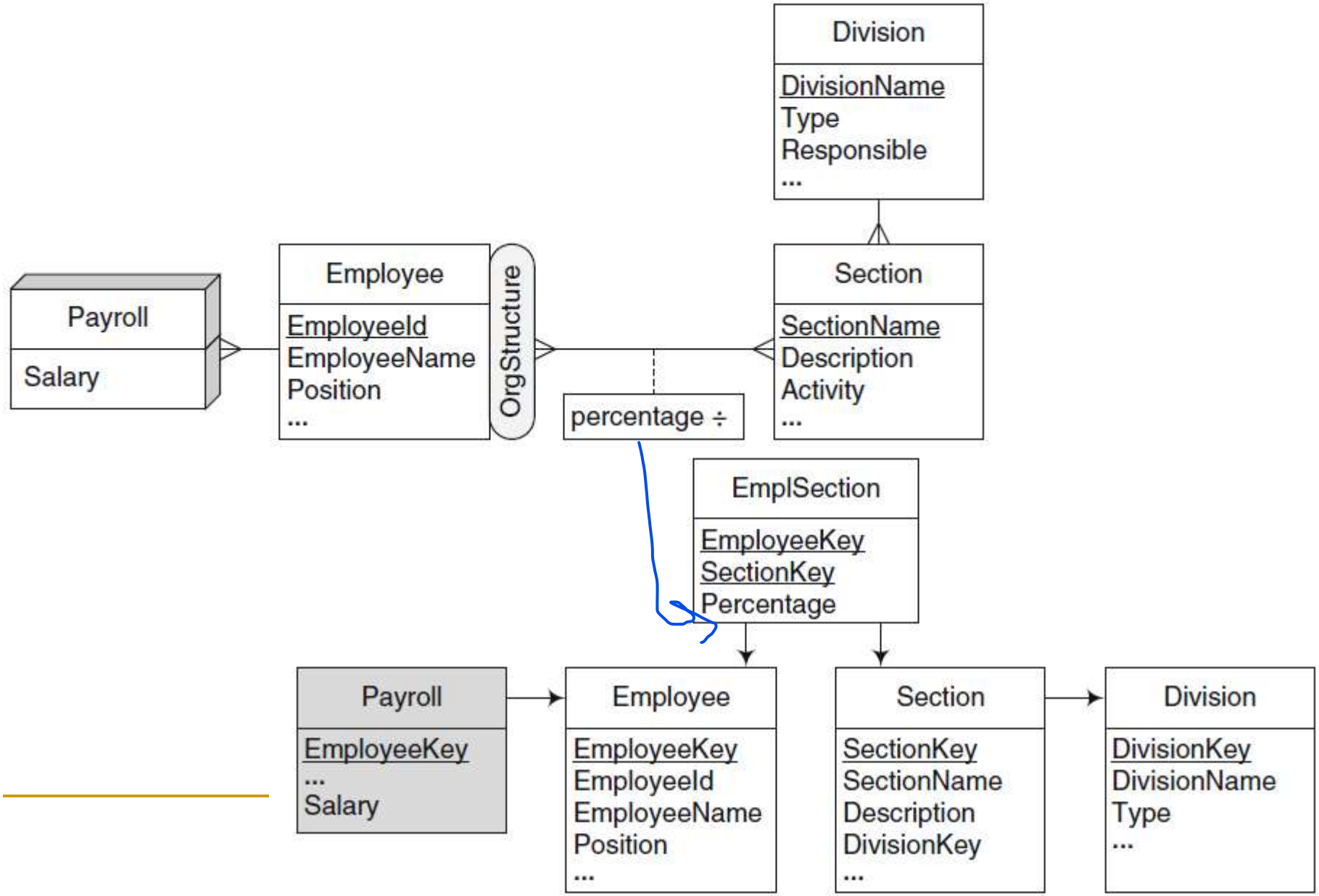
Alternative Hierarchy: **Logical**



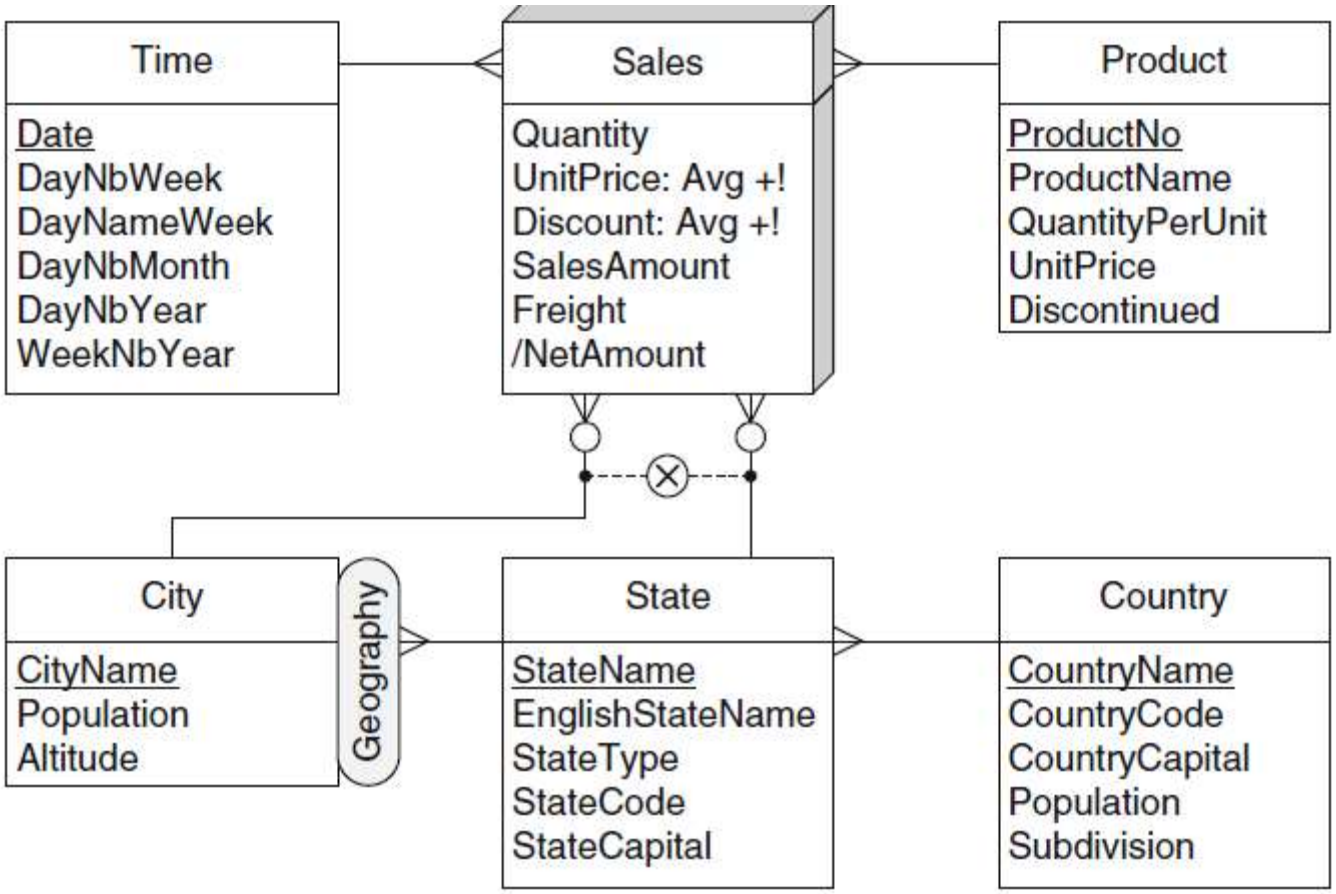
Parallel Dependent hierarchies



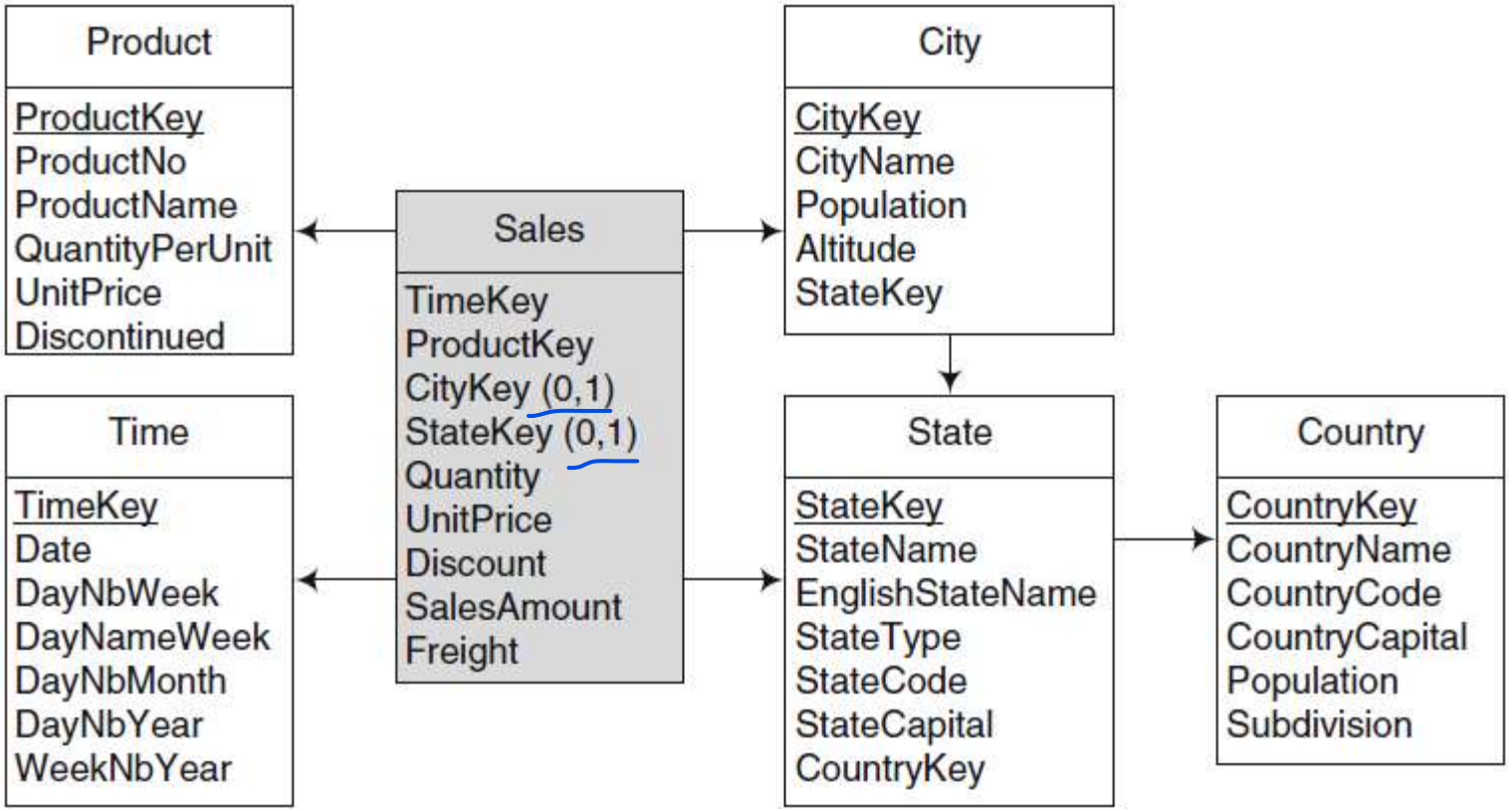
Parallel Nonstrict hierarchies



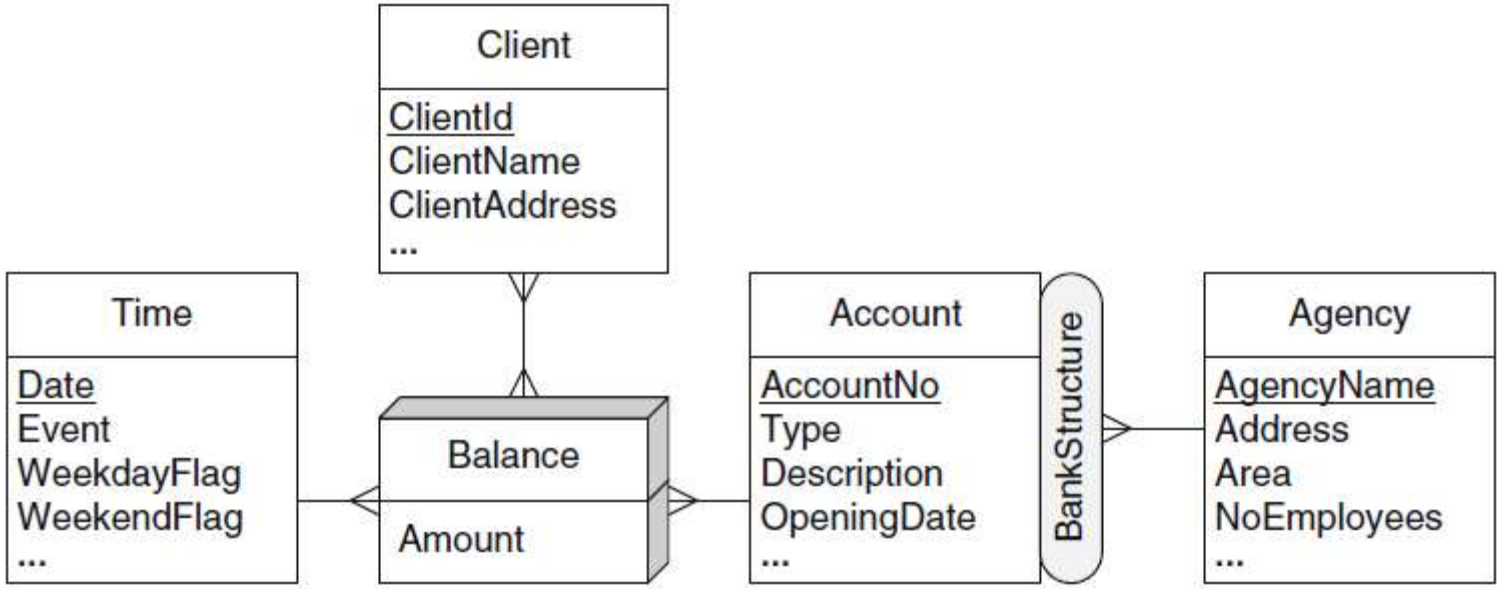
Multiple granularities for the sales fact



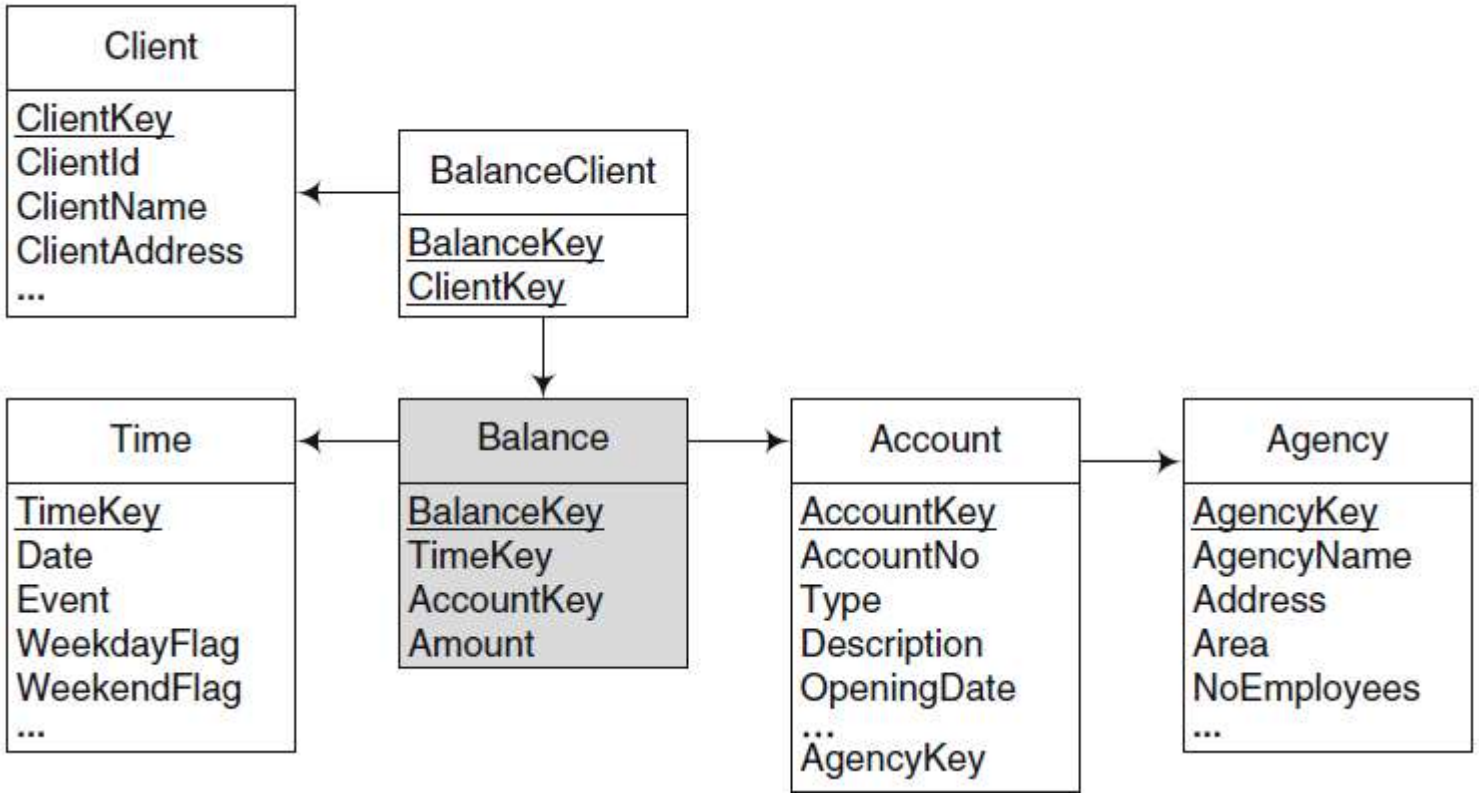
Multiple granularities for the sales fact



Many-to-Many Dimensions



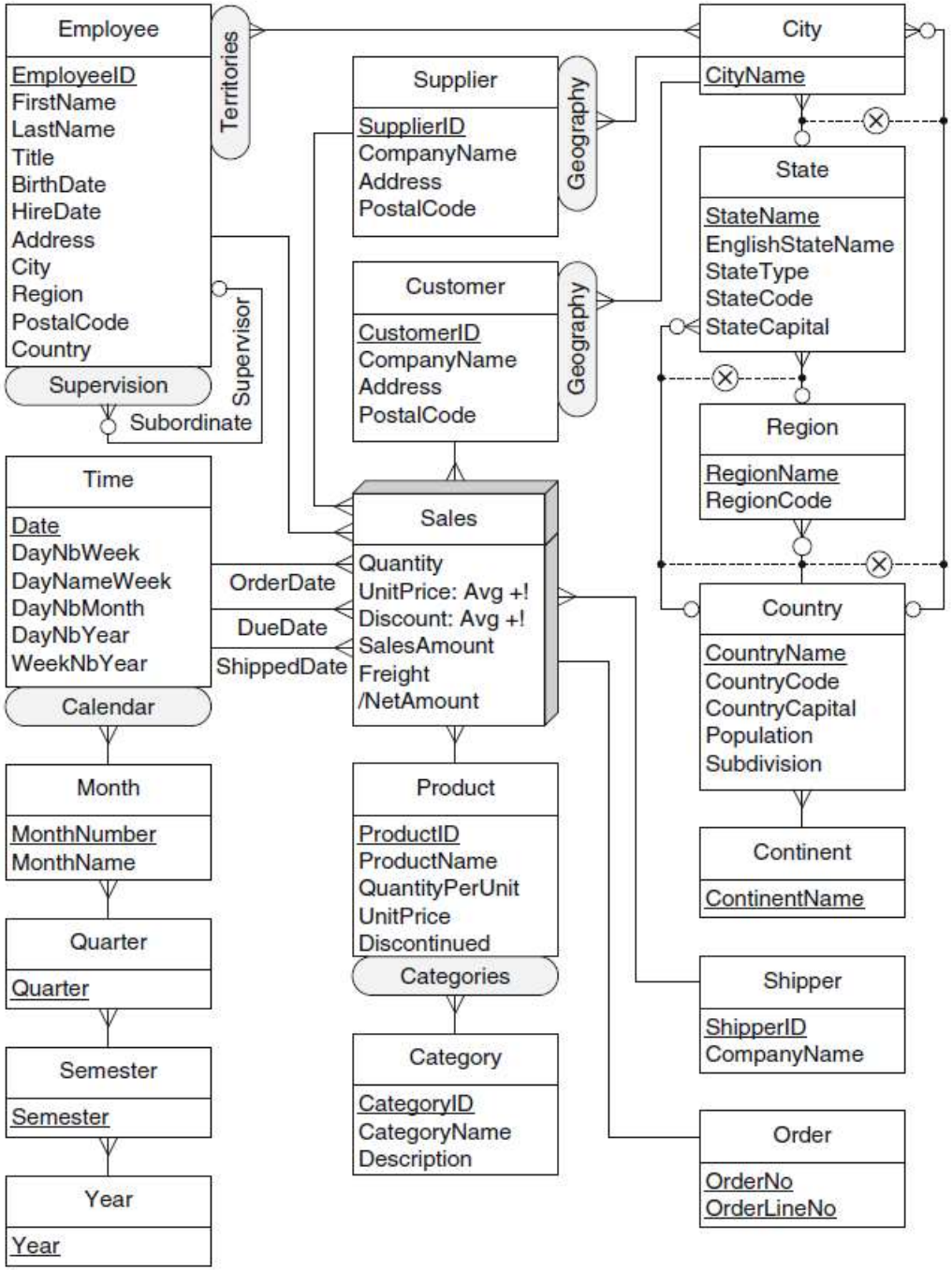
Many-to-Many Dimensions



Lab Case Study

- Lab Case Study

Conceptual Model



Logical Model