# DS-306 Data Warehousing and Business Intelligence
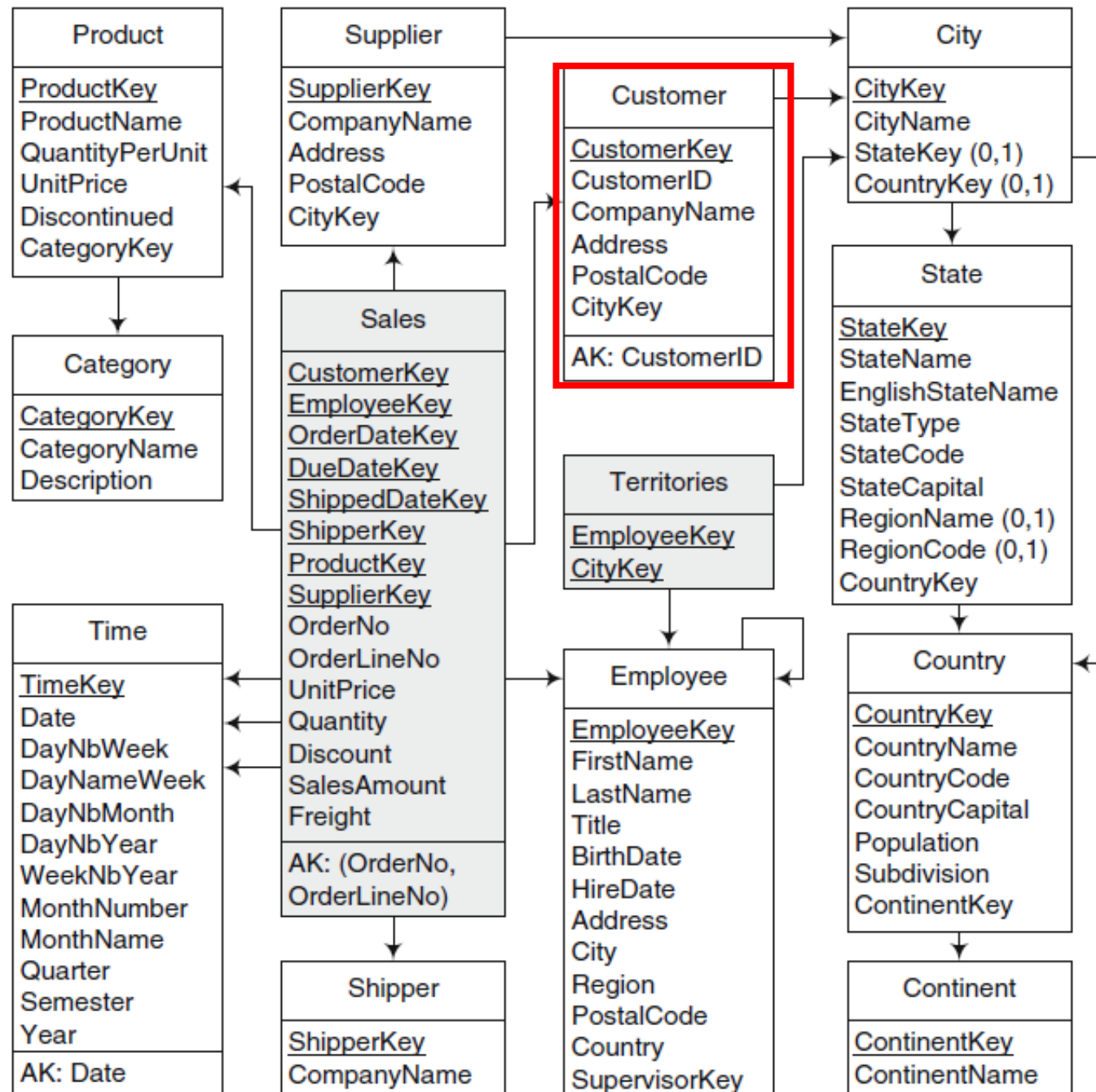
**Topic 4: Logical DW Design**

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**

**Product**

ProductKey
ProductName
QuantityPerUnit
UnitPrice
Discontinued
CategoryKey

**Category**

CategoryKey
CategoryName
Description

**Supplier**

SupplierKey
CompanyName
Address
PostalCode
CityKey

**Sales**

CustomerKey
EmployeeKey
OrderDateKey
DueDateKey
ShippedDateKey
ShipperKey
ProductKey
SupplierKey
OrderNo
OrderLineNo
UnitPrice
Quantity
Discount
SalesAmount
Freight

AK: (OrderNo, OrderLineNo)

**Customer**

CustomerKey
CustomerID
CompanyName
Address
PostalCode
CityKey

AK: CustomerID

**Territories**

EmployeeKey
CityKey

**Time**

TimeKey
Date
DayNbWeek
DayNameWeek
DayNbMonth
DayNbYear
WeekNbYear
MonthNumber
MonthName
Quarter
Semester
Year

AK: Date

**Shipper**

ShipperKey
CompanyName

**Employee**

EmployeeKey
FirstName
LastName
Title
BirthDate
HireDate
Address
City
Region
PostalCode
Country
SupervisorKey

**City**

CityKey
CityName
StateKey (0,1)
CountryKey (0,1)

**State**

StateKey
StateName
EnglishStateName
StateType
StateCode
StateCapital
RegionName (0,1)
RegionCode (0,1)
CountryKey

**Country**

CountryKey
CountryName
CountryCode
CountryCapital
Population
Subdivision
ContinentKey

**Continent**

ContinentKey
ContinentName

# 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

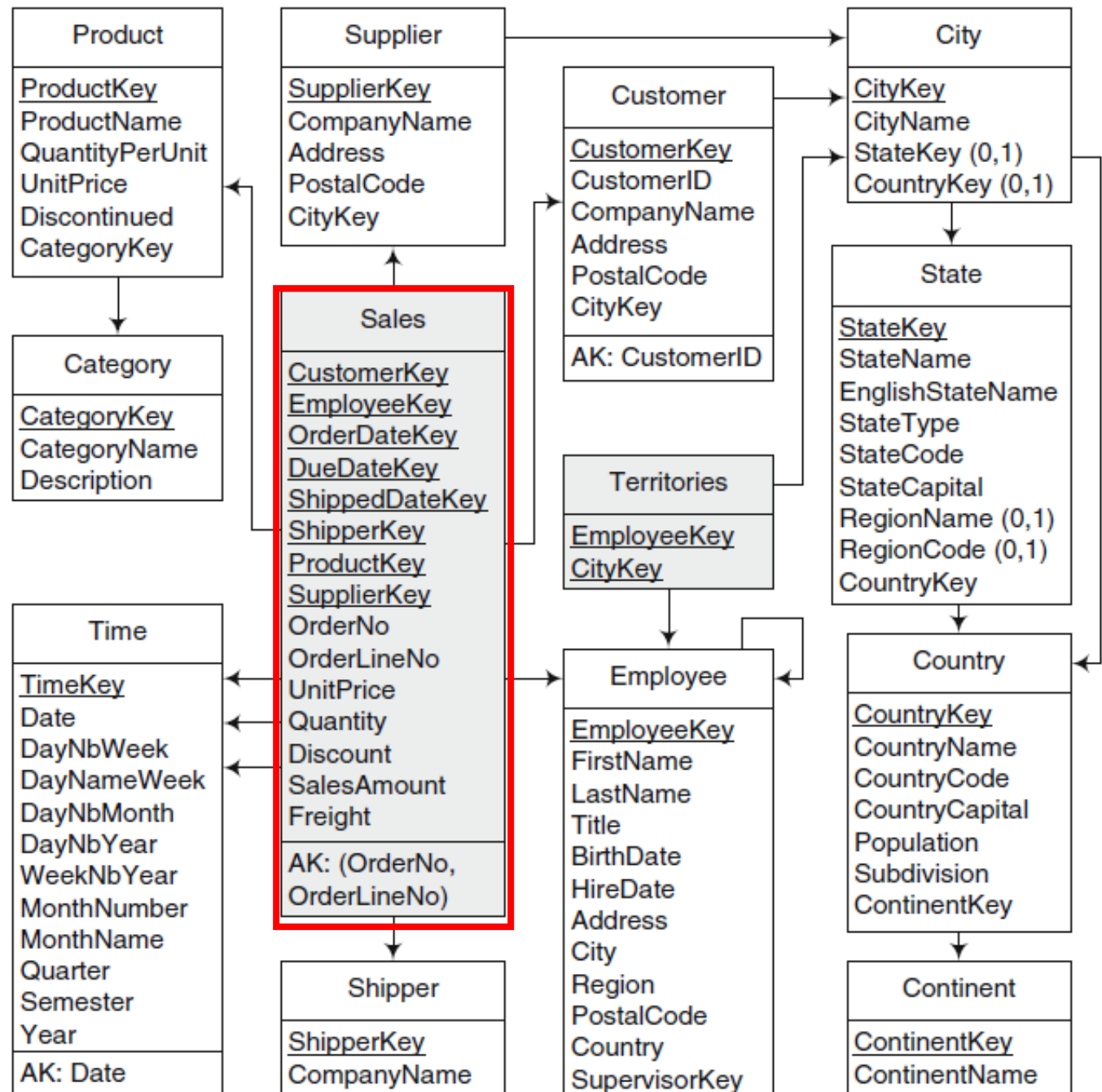- Further, a **surrogate key** may be added to the table

# Rule 3

- A relationship **between** either a **fact F** and a **dimension level L**, or between dimension level $L_P$ and $L_C$ (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
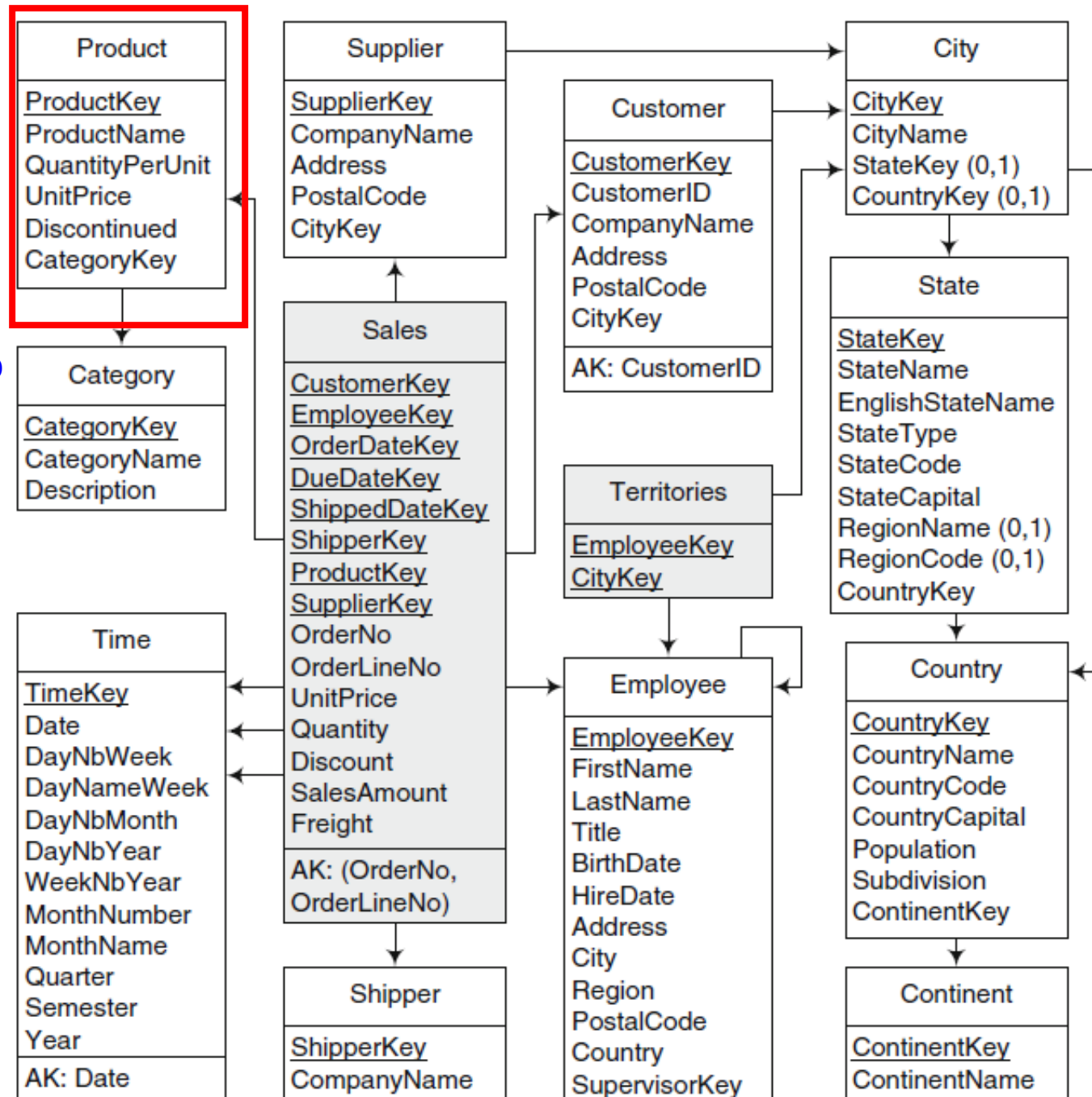
**OrderNo**
**OrderlineNo**
**are dimensions but**
**included in fact table**

**Product**

ProductKey
ProductName
QuantityPerUnit
UnitPrice
Discontinued
CategoryKey

**Category**

CategoryKey
CategoryName
Description

**Time**

TimeKey
Date
DayNbWeek
DayNameWeek
DayNbMonth
DayNbYear
WeekNbYear
MonthNumber
MonthName
Quarter
Semester
Year

AK: Date

**Supplier**

SupplierKey
CompanyName
Address
PostalCode
CityKey

**Sales**

CustomerKey
EmployeeKey
OrderDateKey
DueDateKey
ShippedDateKey
ShipperKey
ProductKey
SupplierKey
OrderNo
OrderLineNo
UnitPrice
Quantity
Discount
SalesAmount
Freight

AK: (OrderNo,
OrderLineNo)

**Shipper**

ShipperKey
CompanyName

**Customer**

CustomerKey
CustomerID
CompanyName
Address
PostalCode
CityKey

AK: CustomerID

**Territories**

EmployeeKey
CityKey

**Employee**

EmployeeKey
FirstName
LastName
Title
BirthDate
HireDate
Address
City
Region
PostalCode
Country
SupervisorKey

**City**

CityKey
CityName
StateKey (0,1)
CountryKey (0,1)

**State**

StateKey
StateName
EnglishStateName
StateType
StateCode
StateCapital
RegionName (0,1)
RegionCode (0,1)
CountryKey

**Country**

CountryKey
CountryName
CountryCode
CountryCapital
Population
Subdivision
ContinentKey

**Continent**

ContinentKey
ContinentName

# 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
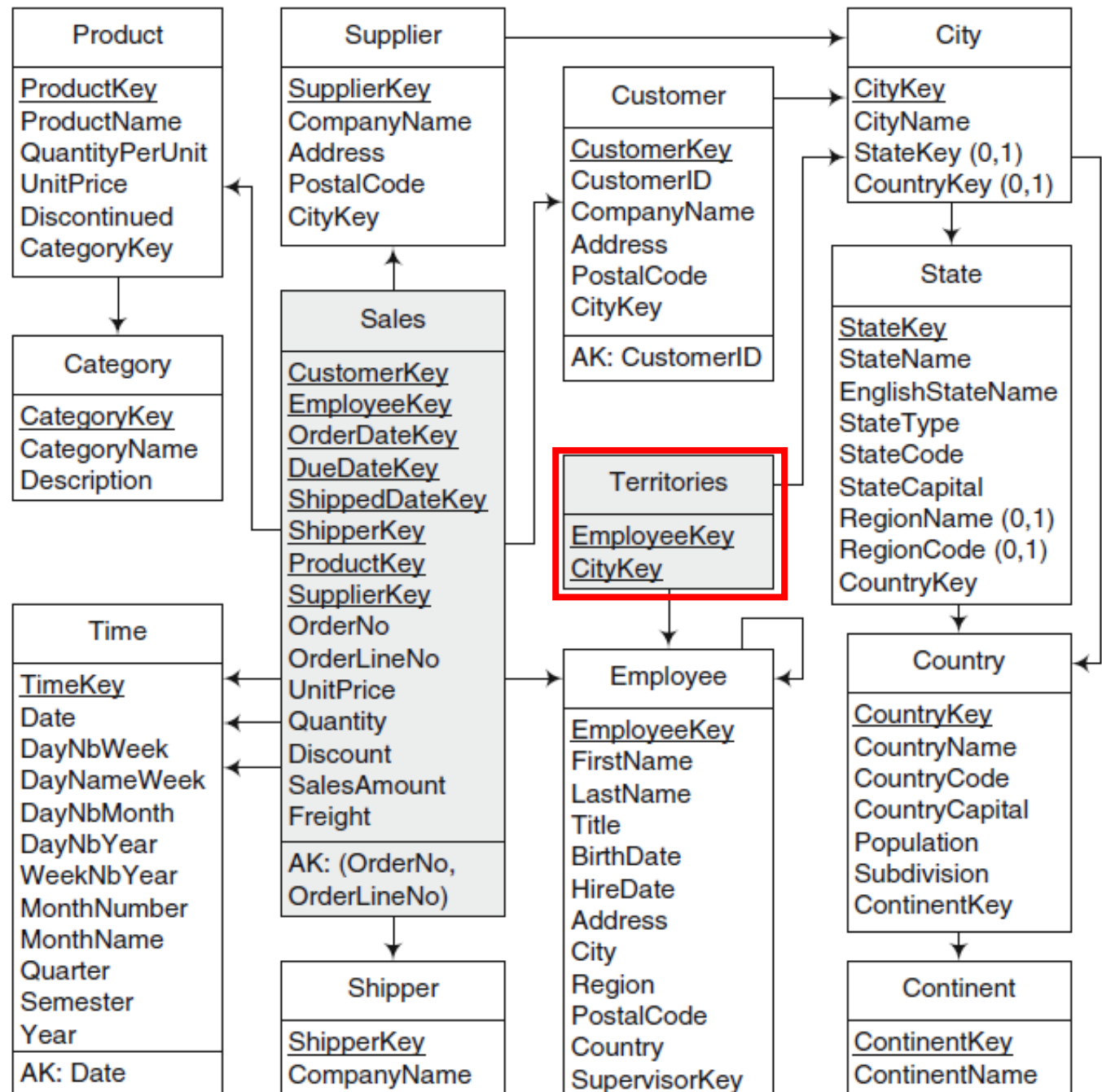
**Product, Supplier Have 1:M relationship**

**Product**
- ProductKey
- ProductName
- QuantityPerUnit
- UnitPrice
- Discontinued
- CategoryKey

**Category**
- CategoryKey
- CategoryName
- Description

**Supplier**
- SupplierKey
- CompanyName
- Address
- PostalCode
- CityKey

**Sales**
- CustomerKey
- EmployeeKey
- OrderDateKey
- DueDateKey
- ShippedDateKey
- ShipperKey
- ProductKey
- SupplierKey
- OrderNo
- OrderLineNo
- UnitPrice
- Quantity
- Discount
- SalesAmount
- Freight

AK: (OrderNo, OrderLineNo)

**Customer**
- CustomerKey
- CustomerID
- CompanyName
- Address
- PostalCode
- CityKey

AK: CustomerID

**Territories**
- EmployeeKey
- CityKey

**City**
- CityKey
- CityName
- StateKey (0,1)
- CountryKey (0,1)

**State**
- StateKey
- StateName
- EnglishStateName
- StateType
- StateCode
- StateCapital
- RegionName (0,1)
- RegionCode (0,1)
- CountryKey

**Time**
- TimeKey
- Date
- DayNbWeek
- DayNameWeek
- DayNbMonth
- DayNbYear
- WeekNbYear
- MonthNumber
- MonthName
- Quarter
- Semester
- Year

AK: Date

**Shipper**
- ShipperKey
- CompanyName

**Employee**
- EmployeeKey
- FirstName
- LastName
- Title
- BirthDate
- HireDate
- Address
- City
- Region
- PostalCode
- Country
- SupervisorKey

**Country**
- CountryKey
- CountryName
- CountryCode
- CountryCapital
- Population
- Subdivision
- ContinentKey

**Continent**
- ContinentKey
- ContinentName

# 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**

**Product**

ProductKey
ProductName
QuantityPerUnit
UnitPrice
Discontinued
CategoryKey

**Category**

CategoryKey
CategoryName
Description

**Supplier**

SupplierKey
CompanyName
Address
PostalCode
CityKey

**Sales**

CustomerKey
EmployeeKey
OrderDateKey
DueDateKey
ShippedDateKey
ShipperKey
ProductKey
SupplierKey
OrderNo
OrderLineNo
UnitPrice
Quantity
Discount
SalesAmount
Freight

AK: (OrderNo,
OrderLineNo)

**Customer**

CustomerKey
CustomerID
CompanyName
Address
PostalCode
CityKey

AK: CustomerID

**Territories**

EmployeeKey
CityKey

**City**

CityKey
CityName
StateKey (0,1)
CountryKey (0,1)

**State**

StateKey
StateName
EnglishStateName
StateType
StateCode
StateCapital
RegionName (0,1)
RegionCode (0,1)
CountryKey

**Time**

TimeKey
Date
DayNbWeek
DayNameWeek
DayNbMonth
DayNbYear
WeekNbYear
MonthNumber
MonthName
Quarter
Semester
Year

AK: Date

**Shipper**

ShipperKey
CompanyName

**Employee**

EmployeeKey
FirstName
LastName
Title
BirthDate
HireDate
Address
City
Region
PostalCode
Country
SupervisorKey

**Country**

CountryKey
CountryName
CountryCode
CountryCapital
Population
Subdivision
ContinentKey

**Continent**

ContinentKey
ContinentName

# 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)**

| All | all |
|-----|-----|
| Category | Beverages ... Seafood |
| Product | Chai Chang ... Ikura Konbu ... |
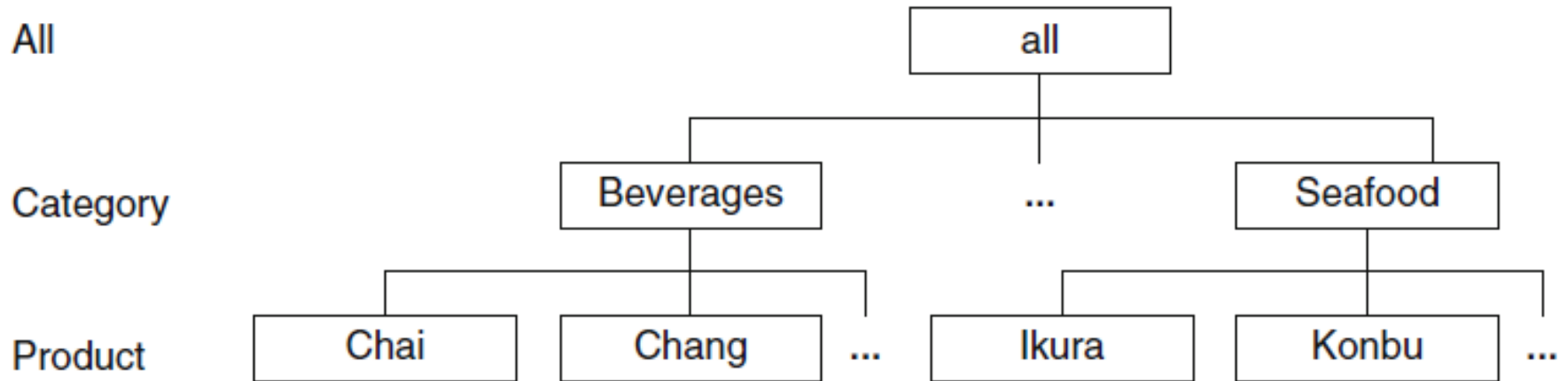
# Balanced Hierarchies: **Logical**

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



| Product | Category | Time |
|---|---|---|
| ProductKey | CategoryKey | TimeKey |
| ProductName | CategoryName | Date |
| QuantityPerUnit | Description | ... |
| UnitPrice | | MonthNumber |
| Discontinued | | MonthName |
| CategoryKey | | Quarter |
| | | Semester |
| | | Year |
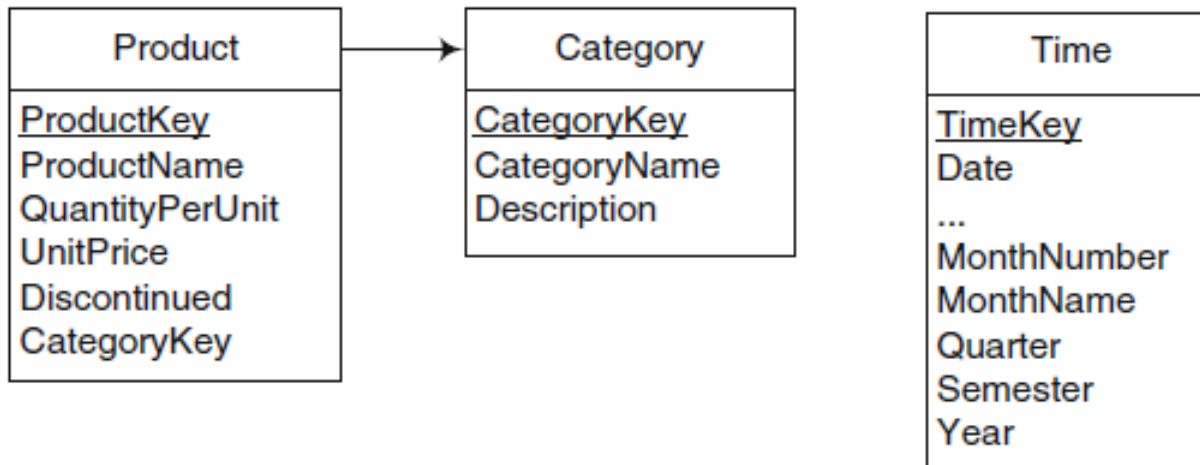
Fig. 5.5 Relations for a balanced hierarchy. (a) Snowflake structure. (b) Flat table

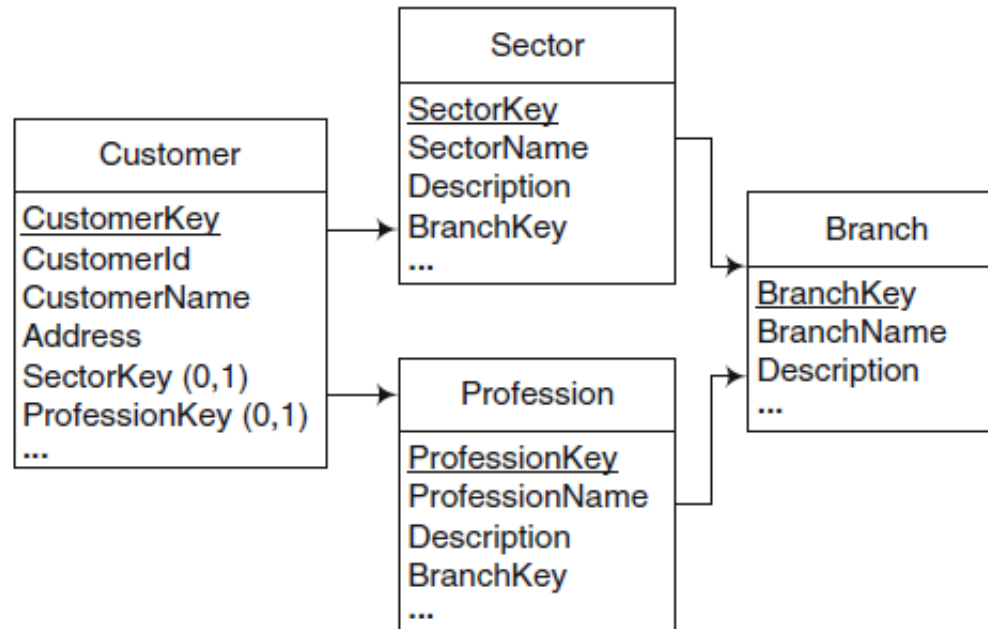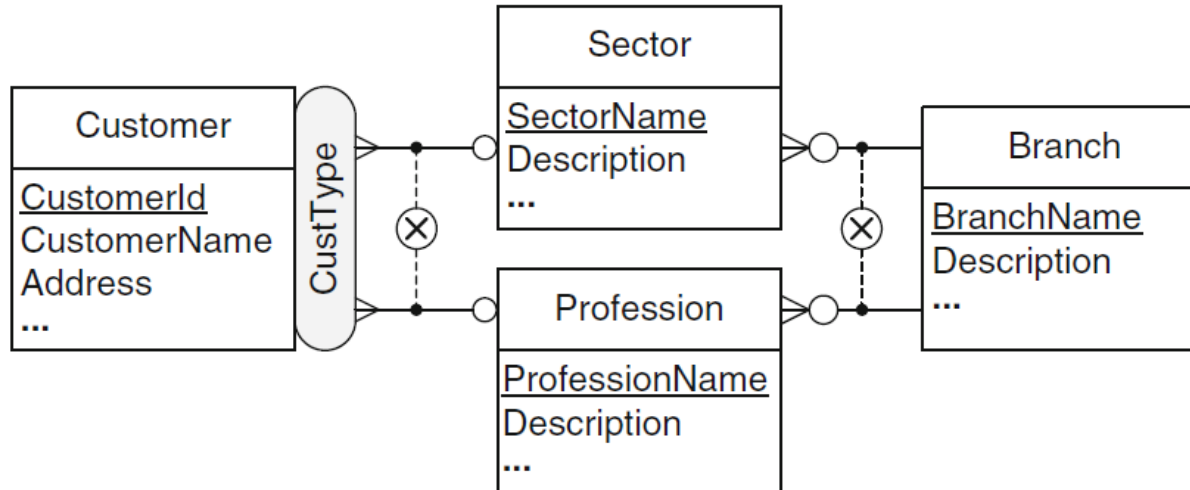# Unbalanced Hierarchies: **Conceptual**
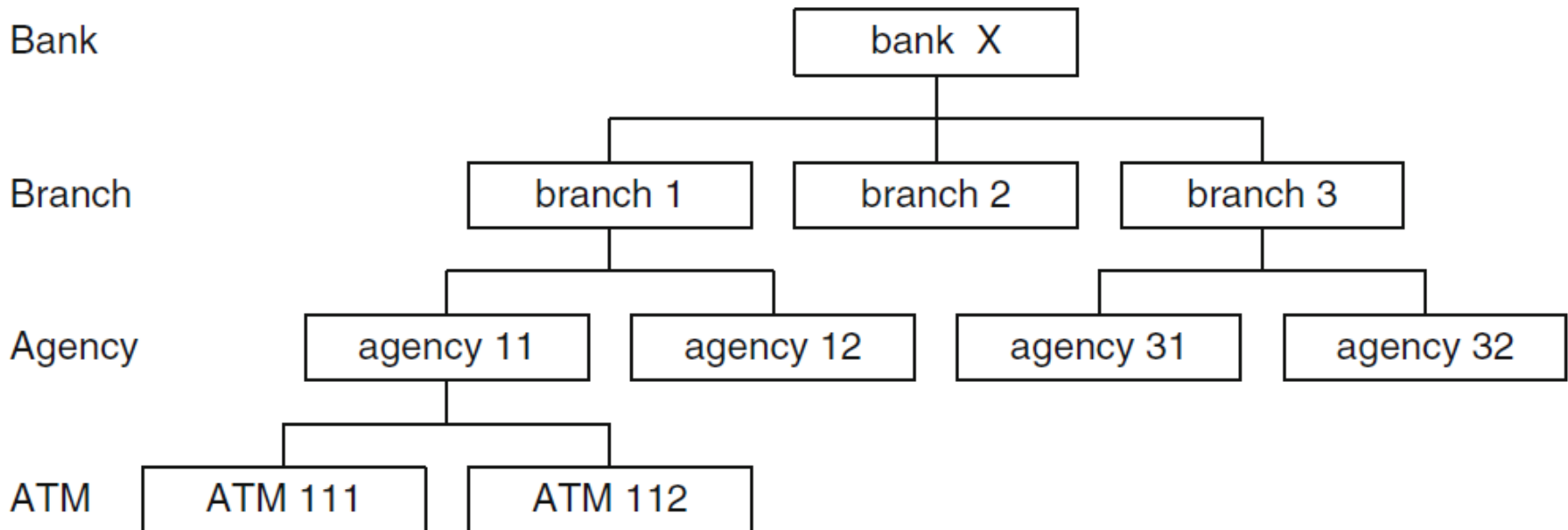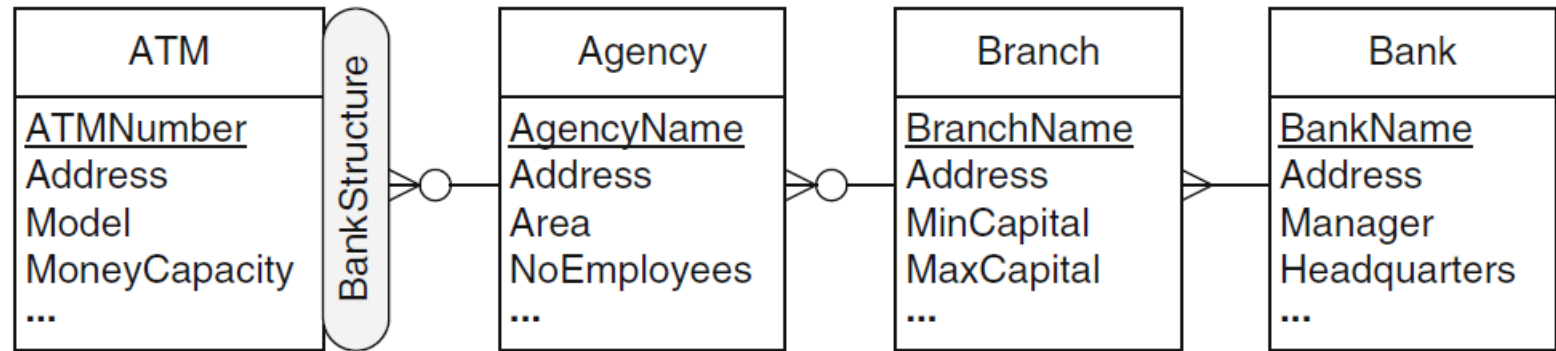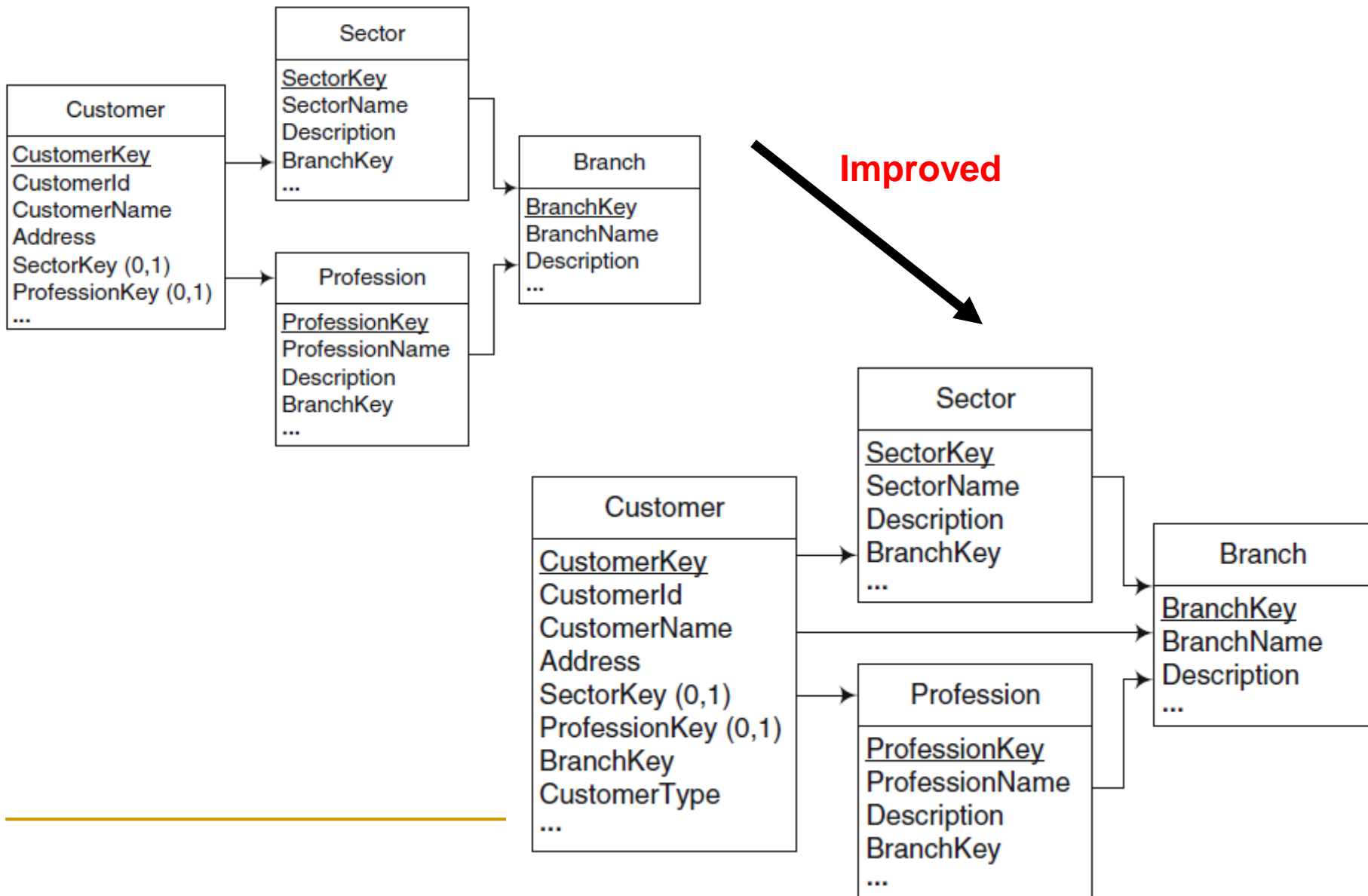
# Unbalanced Hierarchies: **Logical**

- **Problem:** Measures are associated with ATMs
  - Measures will aggregate only for those agencies that have ATMs

- **Solution:** Transform unbalanced hierarchy to balanced using placeholders
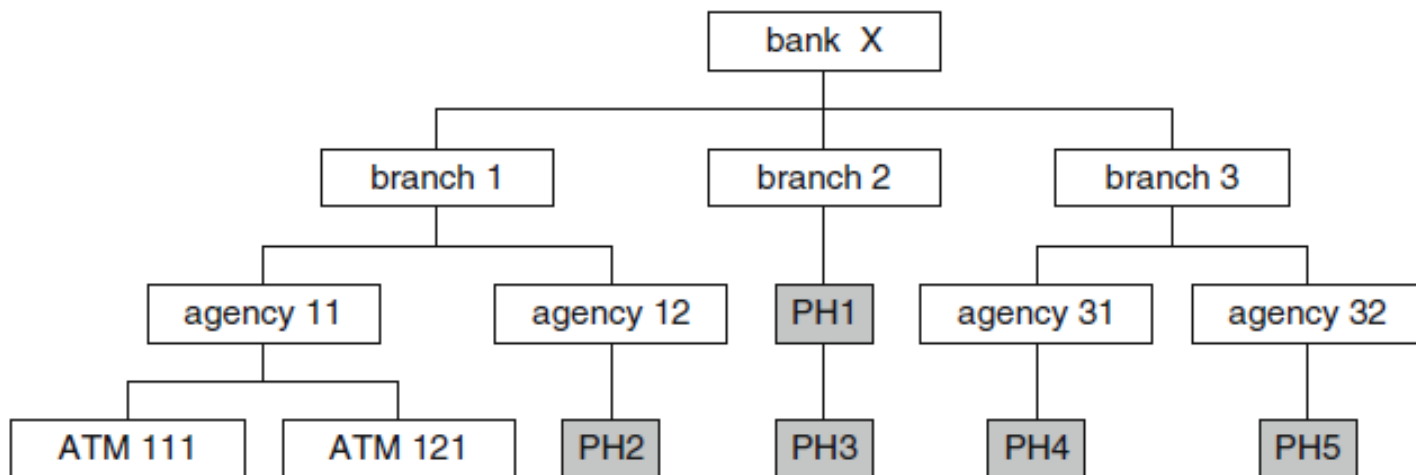
# Generalized Hierarchies: **Logical**

# Generalized Hierarchies: **Logical**

# Alternative Hierarchy: **Logical**
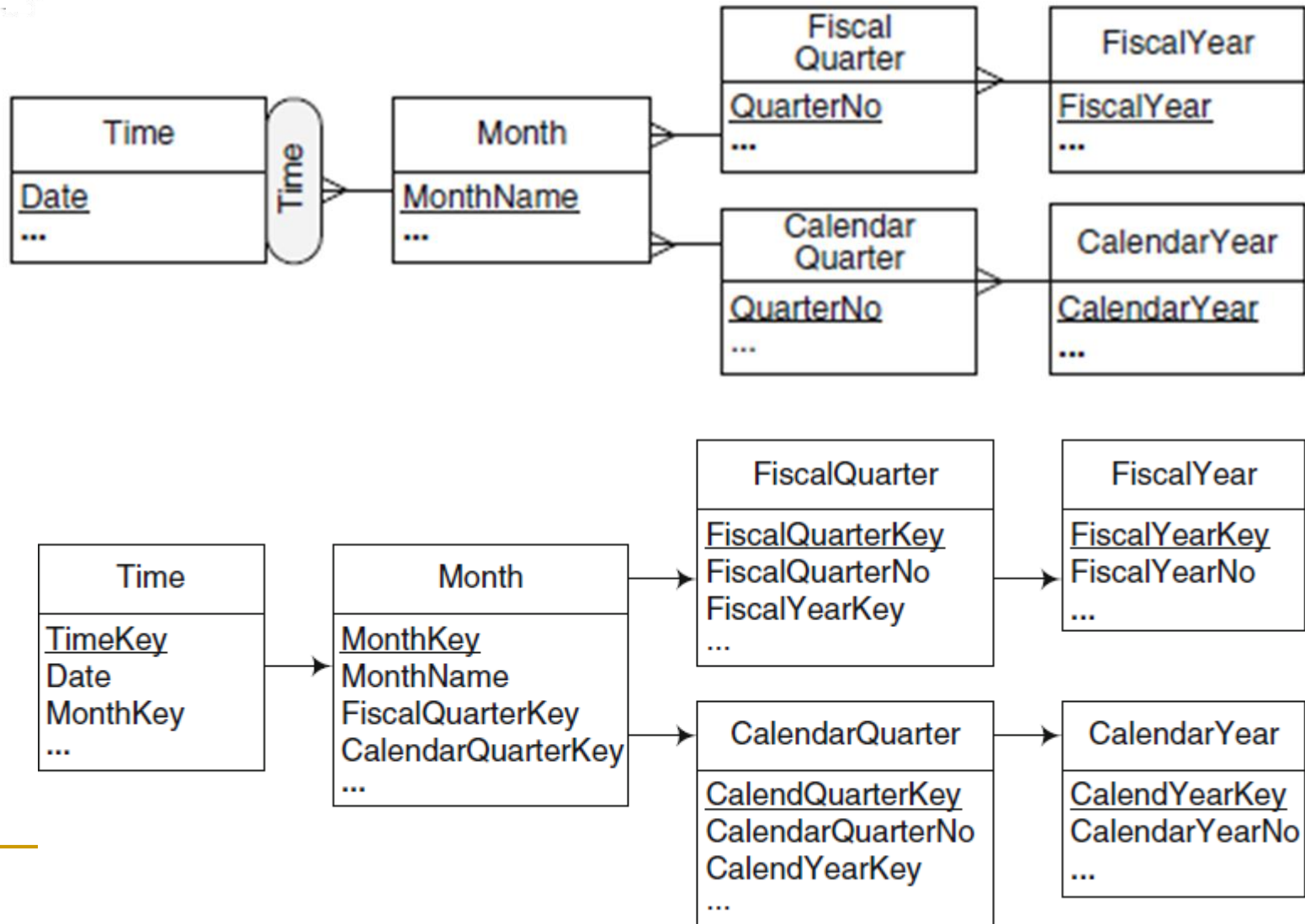
# Parallel Dependent hierarchies
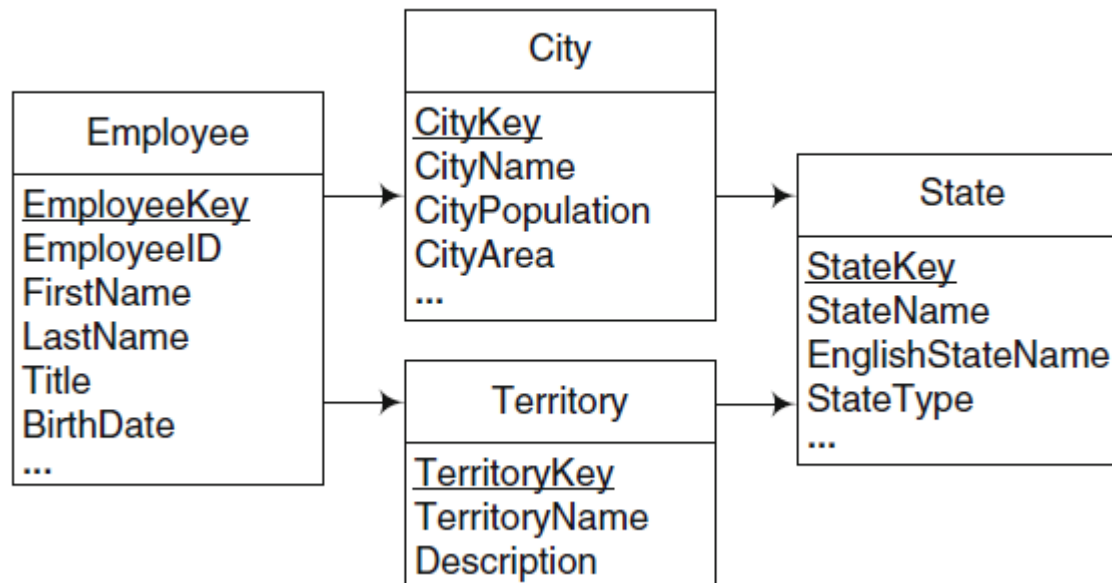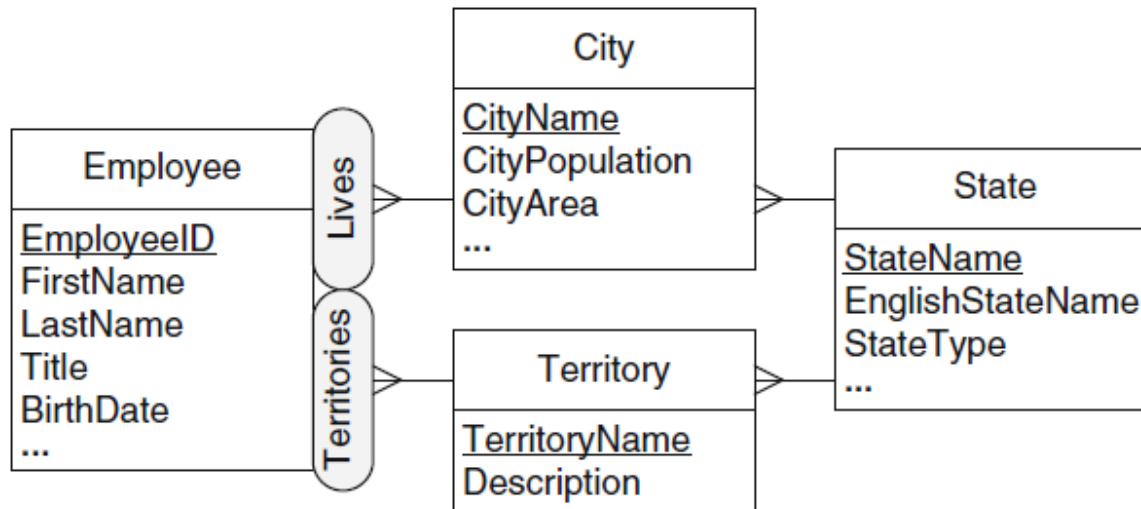
# Parallel Nonstrict hierarchies

# Multiple granularities for the sales fact



**Time**
Date
DayNbWeek
DayNameWeek
DayNbMonth
DayNbYear
WeekNbYear

**Sales**
Quantity
UnitPrice: Avg +!
Discount: Avg +!
SalesAmount
Freight
/NetAmount

**Product**
ProductNo
ProductName
QuantityPerUnit
UnitPrice
Discontinued

**City**
CityName
Population
Altitude

Geography

**State**
StateName
EnglishStateName
StateType
StateCode
StateCapital

**Country**
CountryName
CountryCode
CountryCapital
Population
Subdivision

# Multiple granularities for the sales fact

**Product**

ProductKey
ProductNo
ProductName
QuantityPerUnit
UnitPrice
Discontinued

**City**

CityKey
CityName
Population
Altitude
StateKey

**Sales**

TimeKey
ProductKey
CityKey (0,1)
StateKey (0,1)
Quantity
UnitPrice
Discount
SalesAmount
Freight

**Time**

TimeKey
Date
DayNbWeek
DayNameWeek
DayNbMonth
DayNbYear
WeekNbYear

**State**

StateKey
StateName
EnglishStateName
StateType
StateCode
StateCapital
CountryKey

**Country**

CountryKey
CountryName
CountryCode
CountryCapital
Population
Subdivision

# Many-to-Many Dimensions

# Many-to-Many Dimensions
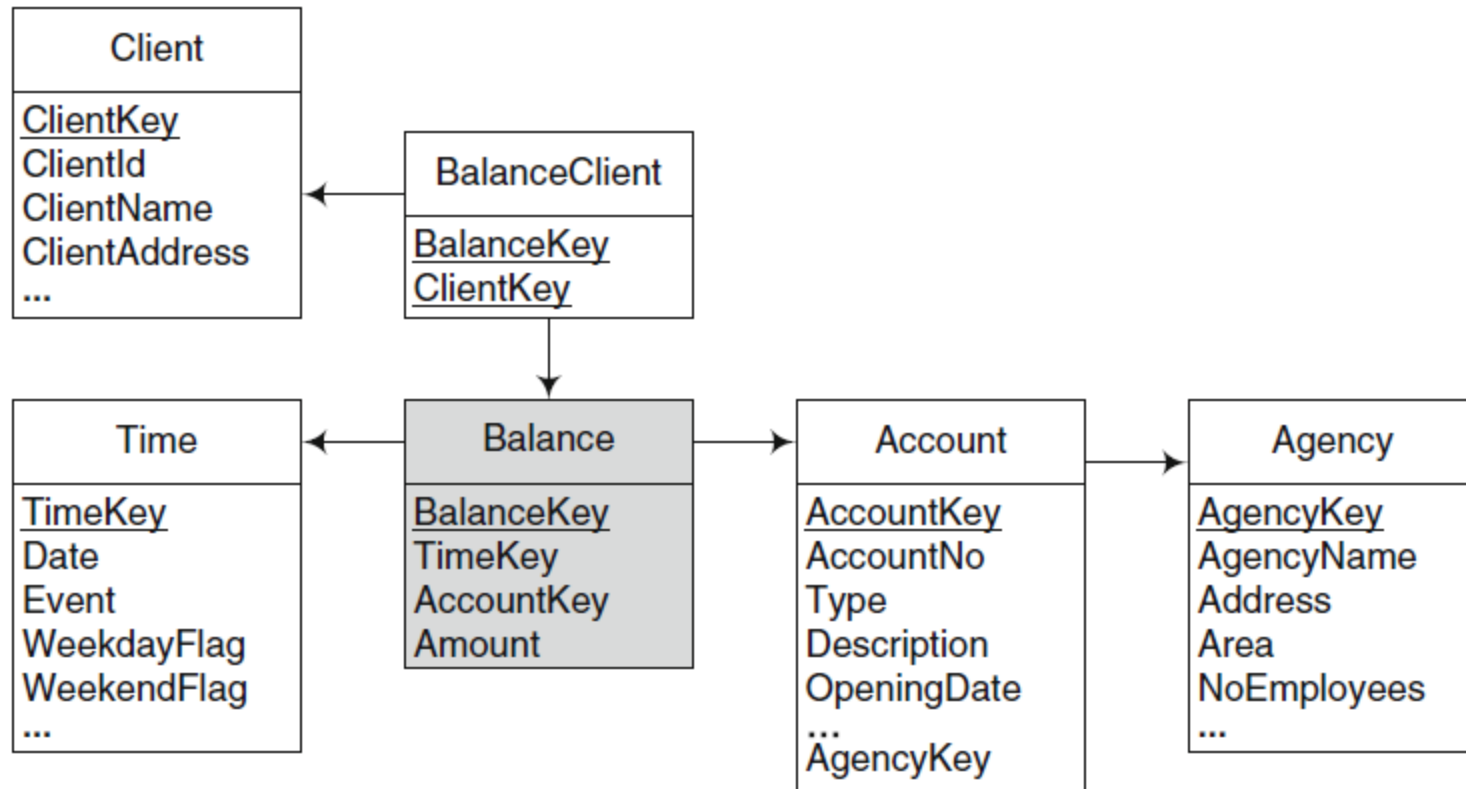
# Lab Case Study

- Lab  Case Study

**Conceptual Model**

**Employee**
- _EmployeeID_
- FirstName
- LastName
- Title
- BirthDate
- HireDate
- Address
- City
- Region
- PostalCode
- Country

Territories

Supervision / Supervisor / Subordinate

**Supplier**
- _SupplierID_
- CompanyName
- Address
- PostalCode

Geography

**Customer**
- _CustomerID_
- CompanyName
- Address
- PostalCode

Geography

**City**
- _CityName_

**State**
- _StateName_
- EnglishStateName
- StateType
- StateCode
- StateCapital

**Region**
- _RegionName_
- RegionCode

**Country**
- _CountryName_
- CountryCode
- CountryCapital
- Population
- Subdivision

**Continent**
- _ContinentName_

**Time**
- _Date_
- DayNbWeek
- DayNameWeek
- DayNbMonth
- DayNbYear
- WeekNbYear

Calendar

OrderDate
DueDate
ShippedDate

**Sales**
- Quantity
- UnitPrice: Avg +!
- Discount: Avg +!
- SalesAmount
- Freight
- /NetAmount

**Month**
- _MonthNumber_
- MonthName

**Quarter**
- _Quarter_

**Semester**
- _Semester_

**Year**
- _Year_

**Product**
- _ProductID_
- ProductName
- QuantityPerUnit
- UnitPrice
- Discontinued

Categories

**Category**
- _CategoryID_
- CategoryName
- Description

**Shipper**
- _ShipperID_
- CompanyName

**Order**
- _OrderNo_
- _OrderLineNo_

35

**Logical Model**

**Product**
- ProductKey
- ProductName
- QuantityPerUnit
- UnitPrice
- Discontinued
- CategoryKey

**Category**
- CategoryKey
- CategoryName
- Description

**Supplier**
- SupplierKey
- CompanyName
- Address
- PostalCode
- CityKey

**Sales**
- CustomerKey
- EmployeeKey
- OrderDateKey
- DueDateKey
- ShippedDateKey
- ShipperKey
- ProductKey
- SupplierKey
- OrderNo
- OrderLineNo
- UnitPrice
- Quantity
- Discount
- SalesAmount
- Freight

AK: (OrderNo, OrderLineNo)

**Customer**
- CustomerKey
- CustomerID
- CompanyName
- Address
- PostalCode
- CityKey

AK: CustomerID

**Territories**
- EmployeeKey
- CityKey

**City**
- CityKey
- CityName
- StateKey (0,1)
- CountryKey (0,1)

**State**
- StateKey
- StateName
- EnglishStateName
- StateType
- StateCode
- StateCapital
- RegionName (0,1)
- RegionCode (0,1)
- CountryKey

**Time**
- TimeKey
- Date
- DayNbWeek
- DayNameWeek
- DayNbMonth
- DayNbYear
- WeekNbYear
- MonthNumber
- MonthName
- Quarter
- Semester
- Year

AK: Date

**Shipper**
- ShipperKey
- CompanyName

**Employee**
- EmployeeKey
- FirstName
- LastName
- Title
- BirthDate
- HireDate
- Address
- City
- Region
- PostalCode
- Country
- SupervisorKey

**Country**
- CountryKey
- CountryName
- CountryCode
- CountryCapital
- Population
- Subdivision
- ContinentKey

**Continent**
- ContinentKey
- ContinentName