ISIT312 Big Data Management

Logical Data Warehouse Design

Dr Fenghui Ren

School of Computing and Information Technology -
University of Wollongong

Logical Data Warehouse Design
Outline

OLAP Technologies

Relational OLAP (ROLAP): Stores data in relational databases, supports extensions to SQL and special access methods to efficiently implement the model and its operations

Multidimensional OLAP (MOLAP): Stores data in special data structures (e.g., arrays) and implement OLAP operations in these structures

- Better performance than ROLAP for query and aggregation, less storage capacity than ROLAP

Hybrid OLAP (HOLAP): Combines both technologies

- For example, detailed data stored in relational databases, aggregations kept in a separate MOLAP store, etc

Logical Data Warehouse Design

Outline

Relational Data Warehouse Design

In ROLAP systems, tables organized in specialized structures

Star schema: One fact table and a set of dimension tables

- Referential integrity constraints between fact table and dimension tables
- Dimension tables may contain redundancy in the presence of hierarchies
- Dimension tables denormalized, fact tables normalized

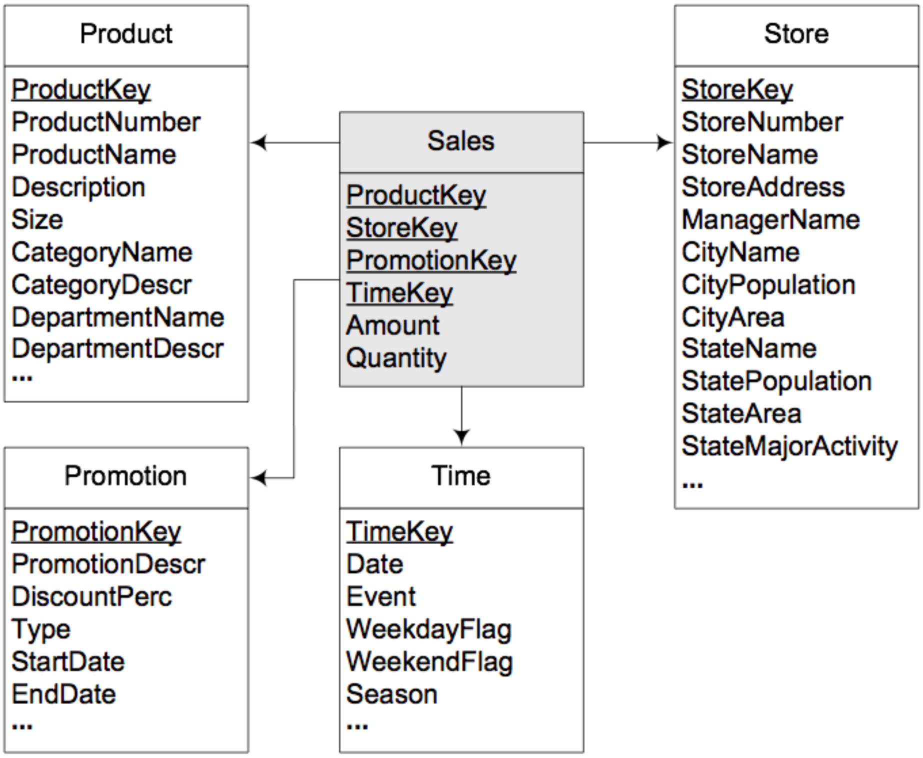
Snowflake schema: Avoids redundancy of star schemas by normalizing dimension tables

- Normalized tables optimize storage space, but decrease performance

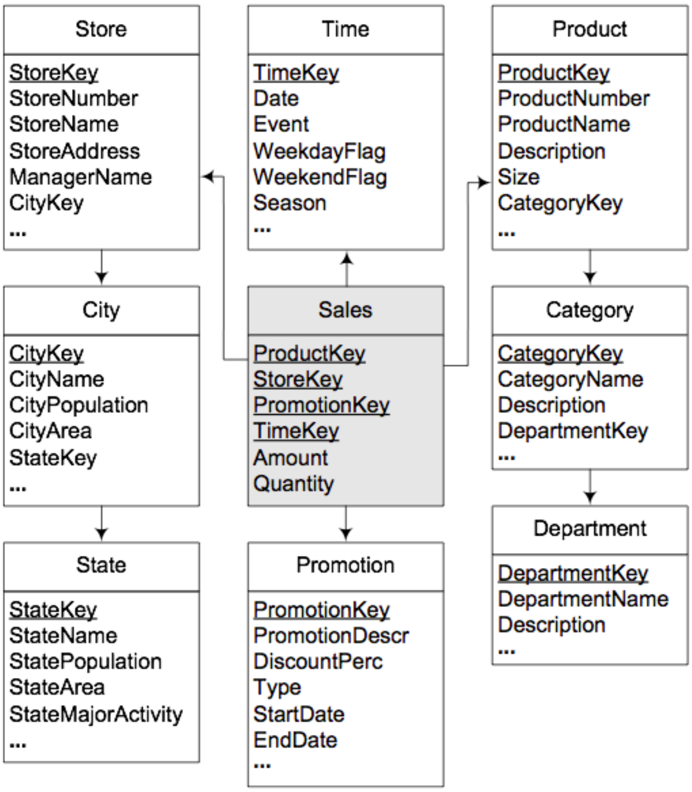
Starflake schema: Combination of the star and snowflake schemas, some dimensions normalized, other not

Constellation schema: Multiple fact tables that share dimension tables

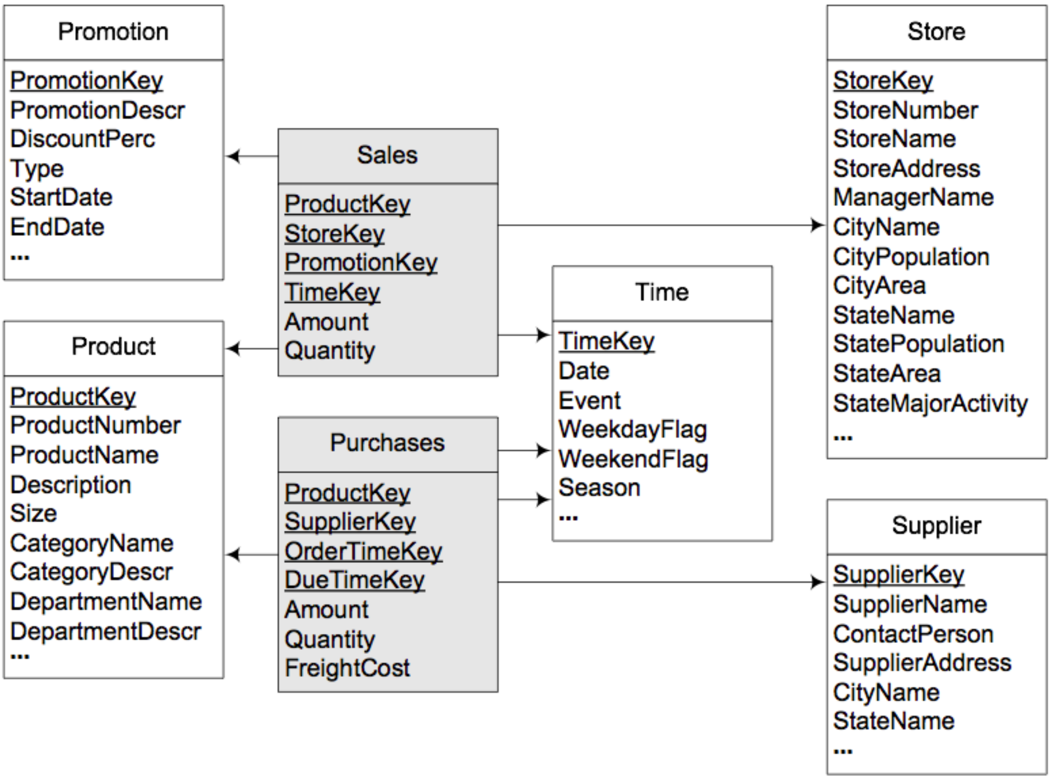
Example of a Star Schema



Example of a Snowflake Schema



Example of a Constellation Schema



Logical Data Warehouse Design
Outline

Relational Implementation of the Conceptual Model

A set of rules to translate the conceptual model (the MultiDim model) into the relational mode

Rule 1: A level L, provided it is not related to a fact with a one-to-one relationship, is mapped to a table TL that contains all attributes of the level

- A surrogate key may be added to the table, otherwise the identifier of the level will be the key of the table
- Additional attributes will be added to this table when mapping relationships using Rule 3 below

Rule 2: A fact F is mapped to a table TF that includes as attributes all measures of the fact

- A surrogate key may be added to the table
- Additional attributes will be added to this table when mapping relationships using Rule 3 below

Relational Implementation of the Conceptual Model

Rule 3: A relationship between either a fact F and a dimension level L, or between dimension levels LP and LC (standing for the parent and child levels, respectively), can be mapped in three different ways, depending on its cardinalities:

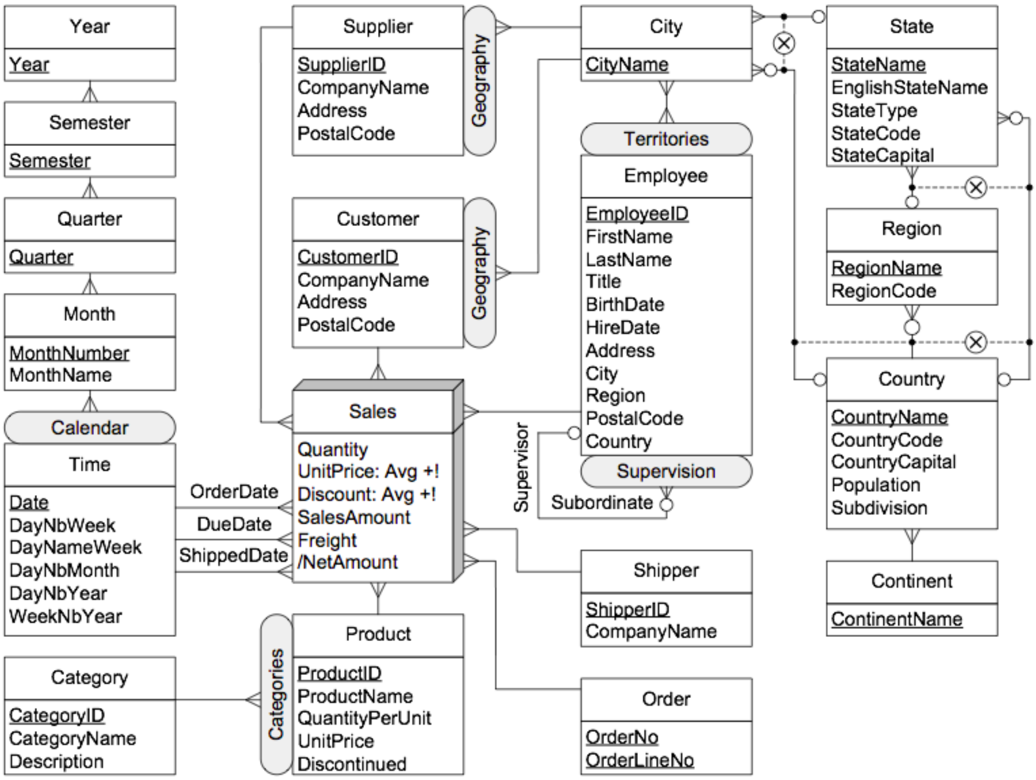
Rule 3a: If the relationship is one-to-one, the table corresponding to the fact (TF) or to the child level (TC) is extended with all the attributes of the dimension level or the parent level, respectively

Rule 3b: If the relationship is one-to-many, the table corresponding to the fact (TF) or to the child level (TC) is extended with the surrogate key of the table corresponding to the dimension level (TL) or the parent level (TP), respectively, that is, there is a foreign key in the fact or child table pointing to the other table

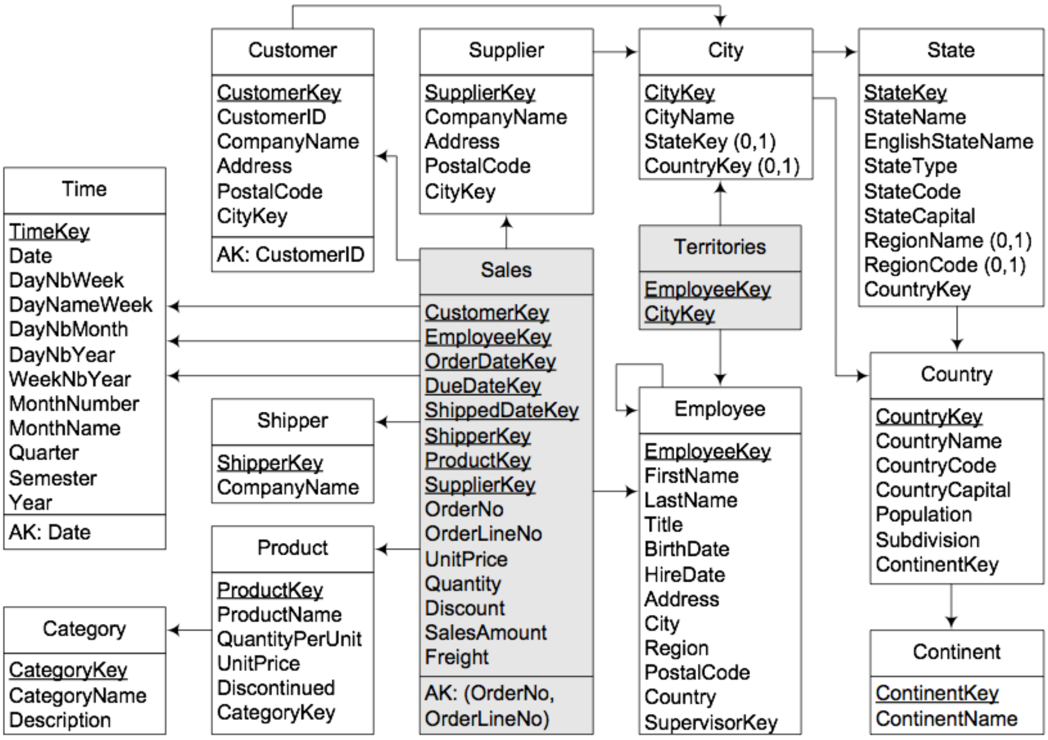
Relational Implementation of the Conceptual Model

Rule 3c: If the relationship is many-to-many, a new table TB (standing for bridge table) is created that contains as attributes the surrogate keys of the tables corresponding to the fact (TF) and the dimension level (TL), or the parent (TP) and child levels (TC), respectively. If the relationship has a distributing attribute, an additional attribute is added to the table to store this information

MultiDim Conceptual Schema of the Northwind Data Warehouse



Relational Representation of the Northwind Data Warehouse



Relational Representation of the Northwind Data Warehouse

- The `Sales` table includes one FK for each level related to the fact with a one-to-many relationship
- For `Time` , several roles: `OrderDate` , `DueDate` , and `ShippedDate`
- `Order` : related to the fact with a one-to-one relationship, called a degenerate, or a fact dimension
- Fact table contains five attributes representing the measures:
  - `UnitPrice` , `Quantity` , `Discount` , `SalesAmount` , and `Freight` .
- The many-to-many parent-child relationship between `Employee` and `Territory` is mapped to the table `Territories` , containing two foreign keys
- `Customer` has a surrogate key `CustomerKey` and a database key `CustomerAltKey`
- `SupplierKey` in `Supplier` is a database key

Logical Data Warehouse Design

Outline

The Time Dimension

Data warehouse: a historical database

Time dimension present in almost all data warehouses.

In a star or snowflake schema, time is included both as foreign key(s) in a fact table and as a time dimension containing the aggregation levels

OLTP databases: temporal information is usually derived from attributes of a `DATE` data type

- Example: A weekend is computed on-the-fly using appropriate functions

In a data warehouse time information is stored as explicit attributes in the time dimension

- Easy to compute: Total sales during weekends

SELECT statement filtering time dimension

```
SELECT SUM(SalesAmount)
FROM Time T, Sales S
WHERE T.TimeKey = S.TimeKey AND T.WeekendFlag
```

The Time Dimension

The granularity of the time dimension varies depending on their use

Time dimension with a granularity `month` spanning 5 years will have 5 * 12 = 60 tuples

References

A. VAISMAN, E. ZIMANYI, Data Warehouse Systems: Design and Implementation, Chapter 5 Logical Data Warehouse Design, Springer Verlag, 2014