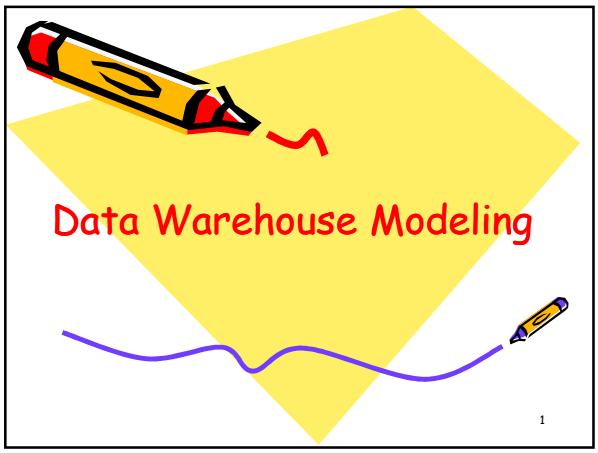


1

Data warehouse modeling

- Data warehouse modeling is the process of designing the schemas of the detailed and summarized information of the data warehouse.
- The goal of data warehouse modeling is to develop a schema describing the reality, or at least a part of the fact, which the data warehouse is needed to support.

2

Data warehouse modeling

- Data warehouse modeling is an essential stage of building a data warehouse for two main reasons.
  - Firstly, through the schema, data warehouse clients can visualize the relationships among the warehouse data, to use them with greater ease.
  - Secondly, a well-designed schema allows an effective data warehouse structure to emerge, to help decrease the cost of implementing the warehouse and improve the efficiency of using it.

3

Data warehouse modeling

- Data modeling in data warehouses is different from data modeling in operational database systems.
- The primary function of data warehouses is to support DSS processes.
- Thus, the objective of data warehouse modeling is to make the data warehouse efficiently support complex queries on long term information.

4

Data warehouse modeling

- data modeling in operational database targets supporting simple transactions in the database such as retrieving, inserting, deleting, and changing data.
- data warehouses are designed for the customer with general information knowledge about the enterprise, whereas operational database systems are more oriented toward use by software specialists for creating distinct applications..

5

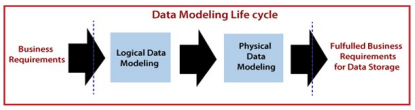
Data Modeling Life Cycle

- a data modeling life cycle is a process of transforming the business requirements to fulfill the goals for storing, maintaining, and accessing the data within IT systems.
- The result is a logical and physical data model for an enterprise data warehouse.

6

Data Modeling Life Cycle

- The objective of the data modeling life cycle is primarily the creation of a storage area for business information.
- That area comes from the logical and physical data modeling stages, as shown in Figure:



A generic data modeling life cycle

7

Conceptual Data Model

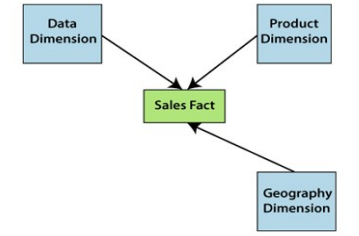
- A conceptual data model recognizes the highest-level relationships between the different entities.
- Characteristics of the conceptual data model
  - It contains the essential entities and the relationships among them.
  - No attribute is specified.
  - No primary key is specified.

8

Conceptual Data Model

- It can be seen that the only data shown via the conceptual data model is the entities that define the data and the relationships between those entities.
- No other data, as shown through the conceptual data model.

9



Example of Conceptual Data Model

10

Logical Data Model

- A logical data model defines the information in as much structure as possible, without observing how they will be physically achieved in the database.
- The primary objective of logical data modeling is to document the business data structures, processes, rules, and relationships by a single view - the logical data model.

11

Logical Data Model

- **Features of a logical data model**
  - It involves all entities and relationships among them.
  - All attributes for each entity are specified.
  - The primary key for each entity is stated.
  - Referential Integrity is specified (FK Relation).

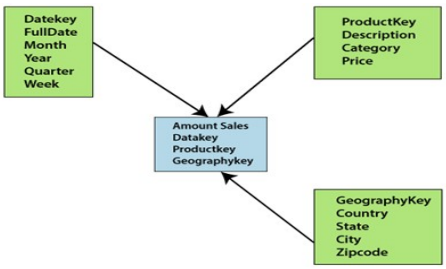
12

Logical Data Model

- The phase for designing the logical data model which are as follows:
  - Specify primary keys for all entities.
  - List the relationships between different entities.
  - List all attributes for each entity.
  - Normalization.
  - No data types are listed

13

Logical Data Model



Example of Logical Data Model

14

Physical Data Model

- Physical data model describes how the model will be presented in the database.
- A physical database model demonstrates all table structures, column names, data types, constraints, primary key, foreign key, and relationships between tables.

15

Physical Data Model

- The purpose of physical data modeling is the mapping of the logical data model to the physical structures of the RDBMS system hosting the data warehouse.
- This contains defining physical RDBMS structures, such as tables and data types to use when storing the information.
- It may also include the definition of new data structures for enhancing query performance.

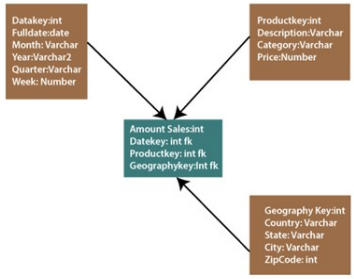
16

Physical Data Model

- Characteristics of a physical data model:
  - Specification all tables and columns.
  - Foreign keys are used to recognize relationships between tables.
- The steps for physical data model design which are as follows:
  - Convert entities to tables.
  - Convert relationships to foreign keys.
  - Convert attributes to columns.

17

Physical Data Model



Example of Physical Data Model

18

Data warehouse Schemas

- The following schemas will be discussed in this course:
  - Star schema: One fact table and a set of dimension tables
  - Snowflake schema: Avoids redundancy of star schemas by normalizing dimension tables
  - Starflake schema: Combination of the star and snowflake schemas, some dimensions normalized, other not
  - Constellation schema: Multiple fact tables that share dimension tables

19

Star schema

- **Star Schema** in data warehouse, is a schema in which the center of the star can have one fact table and a number of associated dimension tables.
  - It is known as star schema as its structure resembles a star.
  - The Star Schema data model is the simplest type of Data Warehouse schema.
  - It is also known as Star Join Schema and is optimized for querying large data sets.

20

Fact Tables

- A Fact table in a star schema contains facts and is connected to dimensions.
- A fact table has two types of columns:
  - A column that includes Facts
  - Foreign Key to Dimensions Table
- Generally, the primary key of a fact table is a composite key that is made up of all the foreign keys that make up the table.

21

Fact Tables

- Fact tables can contain detail-level facts or aggregated facts. Fact tables that include aggregated facts are often called summary tables.
- Fact tables usually contain facts that have been aggregated to some level.

22

Dimension Tables

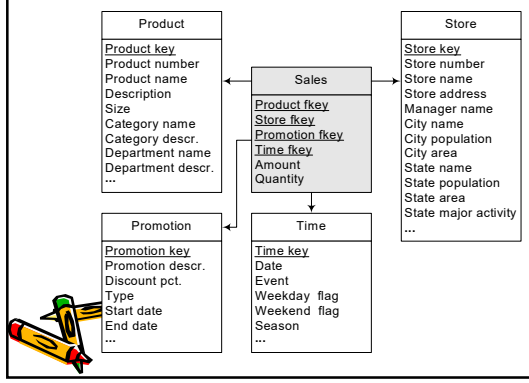
- A dimension is an architecture that categorizes data in a hierarchy.
  - A dimension without hierarchies and levels is called a flat dimension or list.
  - Each dimension table's primary key is part of the composite primary key of the fact table.
  - A dimension attribute is a descriptive, textual attribute that helps describe a dimensional value.
  - Fact tables are usually larger than dimension tables.

23

- Example of Star Schema Data Modelling
  - In the following Star Schema example, the fact table is at the center which contains keys to every dimension table such as product key, store key, promotion key, time key & other attributes like amount and quantity.

24

Star Schema Example



25

Characteristics of Star Schema

- Every dimension in a star schema is represented with the only one-dimension table.
- The dimension table should contain the set of attributes.
- The dimension table is joined to the fact table using a foreign key
- The dimension table are not joined to each other

26

Characteristics of Star Schema

- Fact table would contain key and measur
- The Star schema is easy to understand and provides optimal disk usage.
- The dimension tables are not normalized.
- The schema is widely supported by BI Tools

27

Advantages of Star Schema

- Star schemas have a more straightforward join logic compared to other schemas for fetching data from highly normalized transactional schemas.
- As opposed to highly normalized transactional schemas, the star schema simplifies common business reporting logic, such as reporting and period-over-period.

28

Advantages of Star Schema

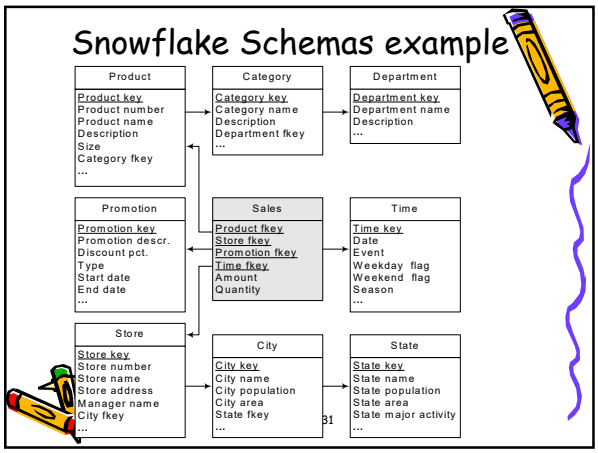
- Star schemas are widely used by OLAP systems to design cubes efficiently.
- A star schema can be used as a source without designing a cube structure in most major OLAP systems.
- By enabling specific performance schemes that can be applied to queries, the query processor software in Star Schema can offer better execution plans.

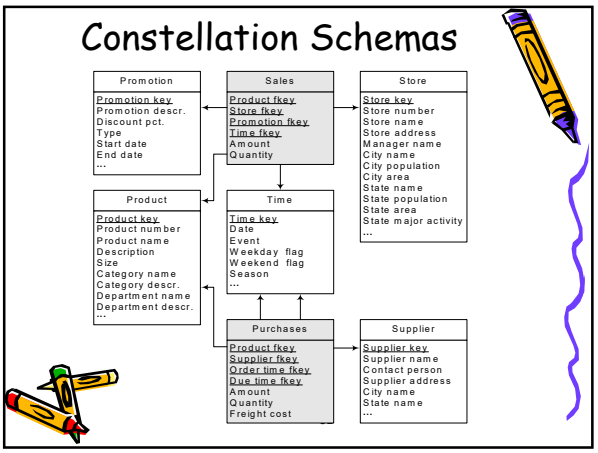
29

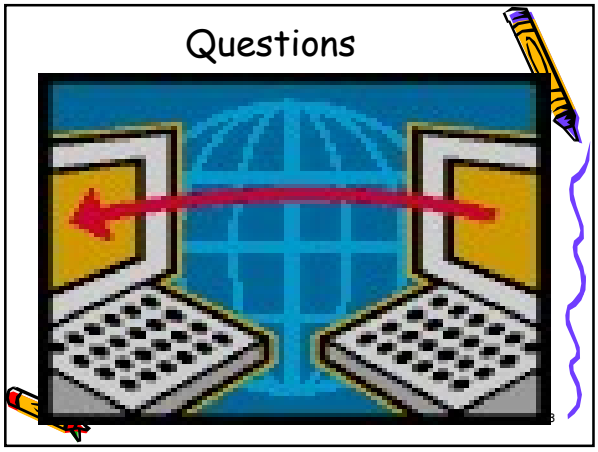
Disadvantage of Star Schema

- Since the schema is highly de-normalized, data integrity is not enforced well.
- Not flexible in terms of analytical needs.
- Star schemas do not reinforce many-to-many relationships within business entities.

30


Snowflake Schemas example
Product
Product key
Product number
Product name
Description
Size
Category fkey
...
Category
Category key
Category name
Description
Department fkey
...
Department
Department key
Department name
Description
...
Promotion
Promotion key
Promotion descr.
Discount pct.
Type
Start date
End date
...
Sales
Product fkey
Store fkey
Promotion fkey
Time fkey
Amount
Quantity
Time
Time key
Date
Event
Weekday flag
Weekend flag
Season
...
Store
Store key
Store number
Store name
Store address
Manager name
City fkey
...
City
City key
City name
City population
City area
State fkey
...
State
State key
State name
State population
State area
State major activity
...


Constellation Schemas
Promotion
Promotion key
Promotion descr.
Discount pct.
Type
Start date
End date
...
Sales
Product fkey
Store fkey
Promotion fkey
Time fkey
Amount
Quantity
Store
Store key
Store number
Store name
Store address
Manager name
City name
City population
City area
State name
State population
State area
State major activity
...
Product
Product key
Product number
Product name
Description
Size
Category name
Category descr.
Department name
Department descr.
...
Time
Time key
Date
Event
Weekday flag
Weekend flag
Season
...
Purchases
Product fkey
Supplier fkey
Order time fkey
Due time fkey
Amount
Quantity
Freight cost
Supplier
Supplier key
Supplier name
Contact person
Supplier address
City name
State name
...


Questions