DBMS schemas for decision support

T.R.Lekhaa

AP-IT

SNSCE

Introduction

- To solve modern business problems like market analysis, financial forecasting ->Multidimensional

How much revenue did the new product generate??

How much revenue did the new product generate by month, in the northeastern division, by sales office, compared with the plan??

Multidimensional data model -> cube

Many dimensions

Basics of DM

- The basic concepts of dimensional modeling are:

facts, dimensions and measures

- A fact is a collection of related data items, consisting of measures and context data. It typically represents business items or business transactions.
- e.g., sales revenue by month by product. Facts are also known as measurements or metrics.
- A dimension is a collection of data that describe one business dimension.
- A measure is a numeric attribute of a fact, representing the performance or behavior of the business relative to the dimensions.

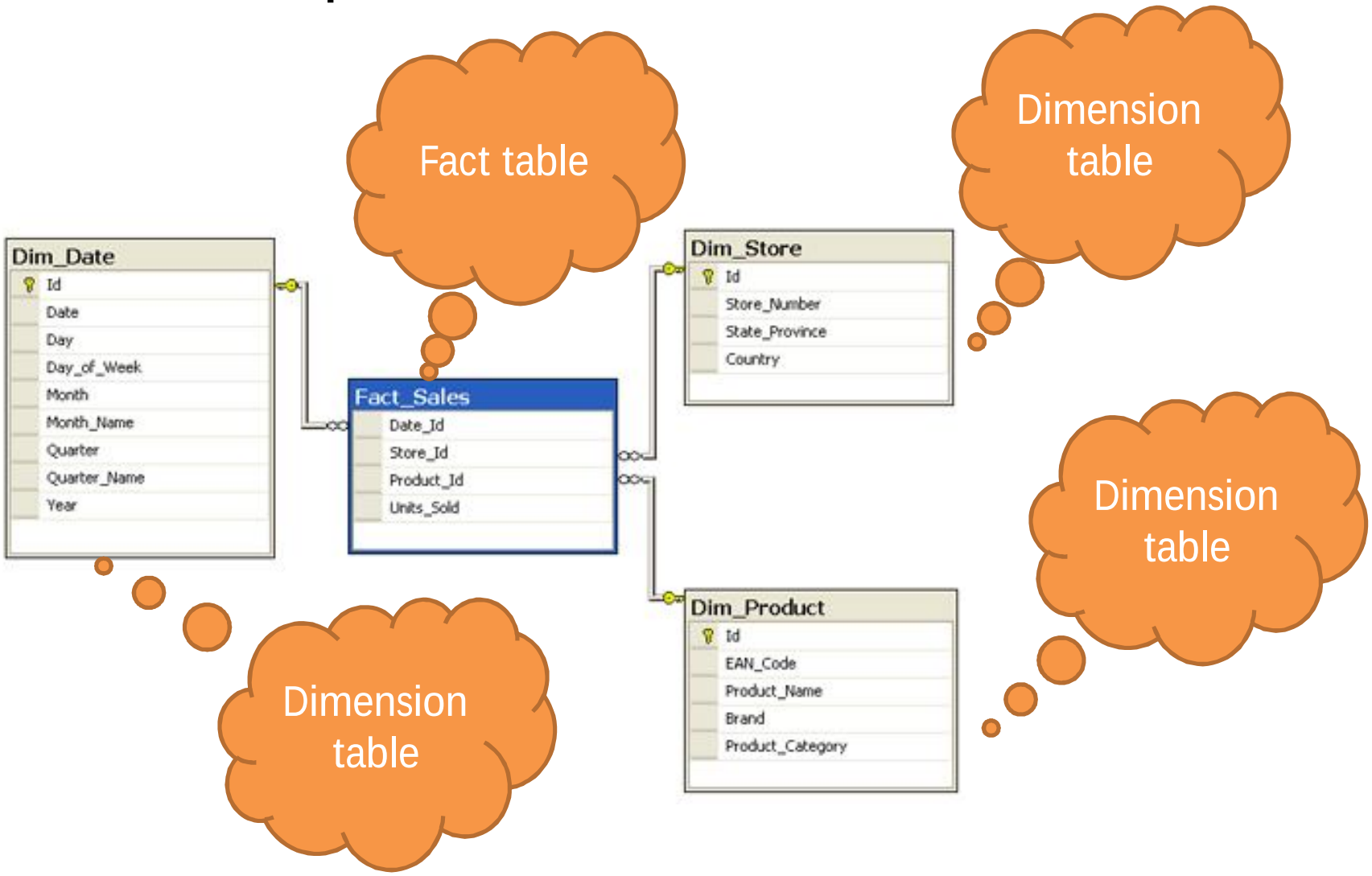
Considering Relational context, there are three basic schemas that are used in dimensional modeling

1. Star schema
2. Snowflake schema
3. Fact constellation schema

Star schema

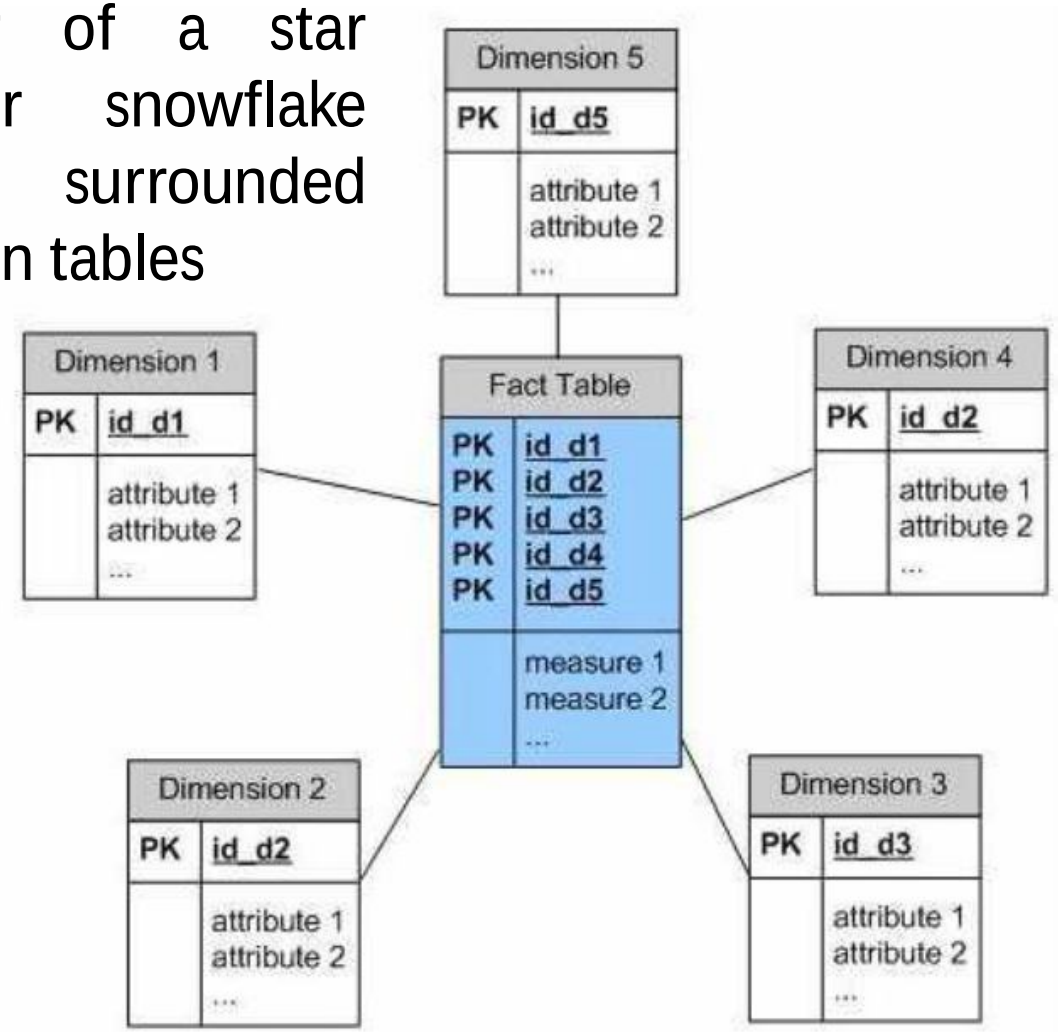
- The basic of star schema is that information can be classified into two groups:
  - Facts
  - Dimension
- Star schema has
  - one large central table (fact table)
  - set of smaller tables (dimensions)
  - arranged in a radial pattern around the central table.
- Facts are core data element being analyzed while dimensions are attributes about the facts.
- Which schema?
  - analysis of project requirements, accessible tools and project team preferences.

Example of fact table – STAR SCHEMA



Example of fact table – STAR SCHEMA

A fact table is found at the center of a star schema or snowflake schema surrounded by dimension tables



STAR SCHEMA

- simplest data warehouse schema.
- Why named as star schema??
  - the diagram resembles a star, with points radiating from a center.
- The center of the star consists of fact table and the points of the star are the dimension tables.
- most commonly used nowadays and is recommended by Oracle.

Fact Tables

- Contains summarized numerical and historical data (facts) and a multipart index composed of foreign keys from the primary keys of related dimension tables.
- A fact table typically has two types of columns:
  - foreign keys to dimension tables and
  - measures those that contain numeric facts.

Dimension Tables

- Dimensions are categories by which summarized data can be viewed.
- E.g.
- a profit summary in a fact table can be viewed
  - by a Time dimension (profit by month, quarter, year),
  - Region dimension (profit by country, state, city),
  - Product dimension (profit for product1, product2).
- A dimension is a structure usually composed of one or more hierarchies that categorizes data.
- If a dimension hasn't got a hierarchies and levels it is called flat dimension or list.
- The primary keys of each of the dimension tables are part of the composite primary key (set of more than one **key** that, together, uniquely identifies each record) of the fact table.
- Example:
- Fact tables store data about sales while dimension tables data about geographic region (markets, cities), clients, products, times, channels.

Measures

- Measures are numeric data based on columns in a fact table.
- They are the primary data which end users are interested in. E.g. a sales fact table may contain a profit measure which represents profit on each sale.

Snowflake schema and Fact constellation schema

- **Snowflake schema:**
  - decomposing one or more of the dimensions.
  - The decomposed snowflake structure visualizes the hierarchical structure of dimensions very well.
- **Fact constellation schema:**
  - splitting the original star schema into more star schemes each of them describes facts on another level of dimension hierarchies)
  - Multiple fact tables share dimension tables.