Bases de Dados e Armazéns de Dados

Departamento de Engenharia Informática (DEI/ISEP)
Paulo Oliveira
pjo@isep.ipp.pt

1

Introduction to Dimensional Modeling

2

Bibliography

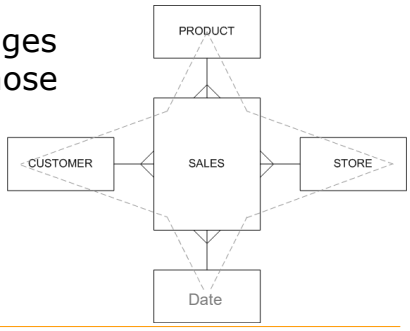
- Mastering the Data Warehouse Design
  Relational and Dimensional Techniques
  Claudia Imhoff, Nicholas Galemno, Jonathan G. Geiger
  Wiley, 2003, First Edition
  Chapters 2, 3, 4, 5 (pages 29-144)

- "From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design"
  Daniel L. Moody, Mark A. R. Kortink

- The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling
  Ralph Kimball, Margy Ross
  Wiley, 2002, Second Edition
  Chapters 1,…,6

Dimensional Modeling

- Data modeling for DW is also known as **star schema creation**
- Based on **fact tables** and **dimension tables**
- Dimensional model is very **asymmetric**
  - one large dominant table in the center of the schema: **fact table**, with multiple joins connecting to **dimension tables**
- **Dimensional model** packages the data in a format whose design goals are:
  - **User understandability**
  - **Query performance**
  - **Flexibility to change**



Dimensional Modeling

- **Model simplicity** achieved by the reduced number of tables
  - Business users benefit because **data is easier to understand** and **navigate**
- Has also **performance benefits**
- Dimensional models are gracefully extensible to **accommodate change**
  - Can accommodate **new dimensions** if a single value of that dimension is defined for each existing fact row
  - Can accommodate **new facts** to the fact table, when the level of detail is consistent with the existing fact table
  - Can supplement preexisting dimension tables with **new attributes**

5

5

Fact Table

- **Primary table** in a dimensional model
- Holds
  - **Primary key made up by the set or subset of foreign keys** (composite primary key) that connect to dimension tables
  - **Numerical measurements of the business** (the facts), which may be analyzed using statistical functions
    - units_sold, value_sold, order_cost, units_ordered, …

6

6

Dimension Tables

- Provide the **basis for analyzing data** in the fact table
- Used to answer ***"WHO"***, ***"WHAT"***, ***"WHEN"***, ***"WHERE"*** and ***"WHY"*** questions about the business events stored in the fact table
- Each dimension table has a **non-composite primary key** that corresponds exactly to one foreign key in the fact table

Fact and Dimension Tables

- There is a **1:N** relationship between each **dimension table** and the **fact table**
- **Fact tables** typically have large volumes of rows, while **dimension tables** have a smaller number
  - Key advantage: **JOIN performance is improved** when one large table can be joined with some small tables
  - In many cases the **dimension tables** are small enough to **be fully cached in memory**

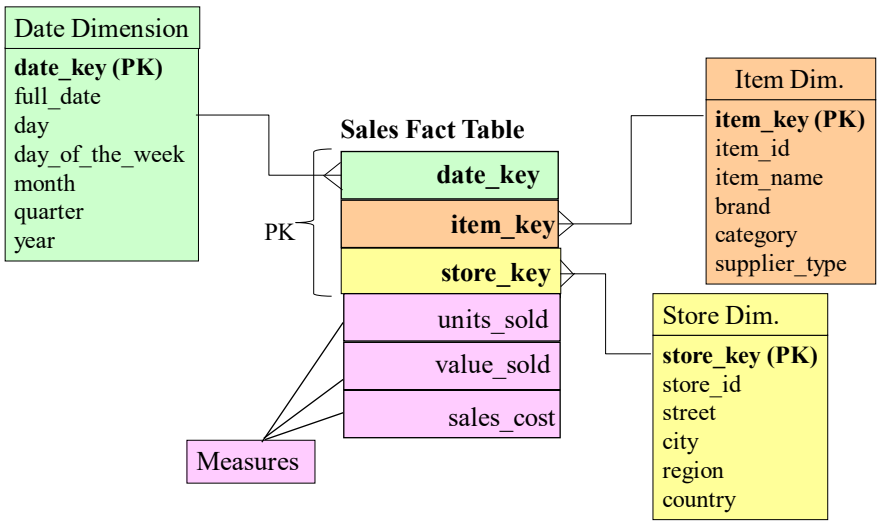
Schema Types

- **Star schema**
  - Fact table in the middle connected to a set of dimension tables
- **Snowflake schema**
  - Refinement of star schema where some dimensional table is normalized into a set of smaller dimension tables, forming a snowflake
- **Galaxy schema/Fact constellation**
  - Multiple fact tables share dimension tables, viewed as a collection of star schemas

9

9

Example of a Star Schema



10

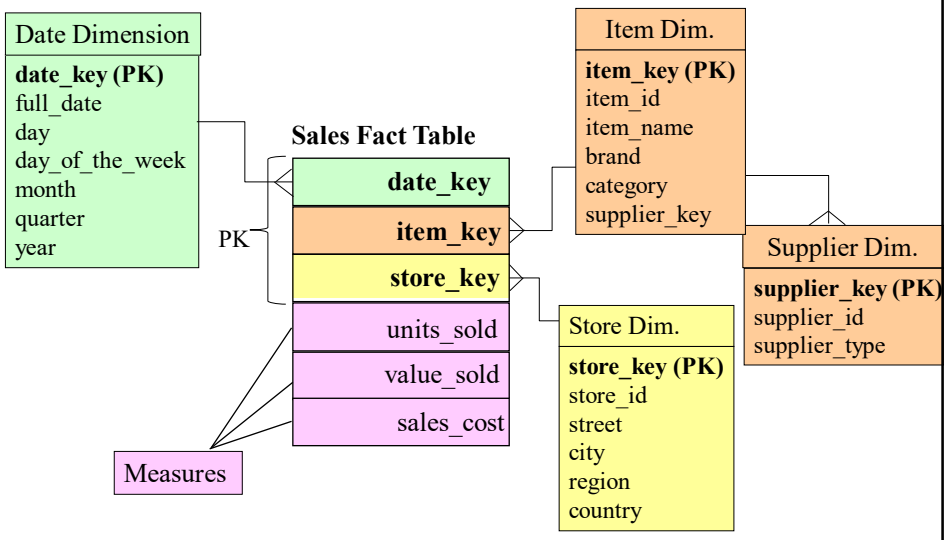
10

Date Dimension Example

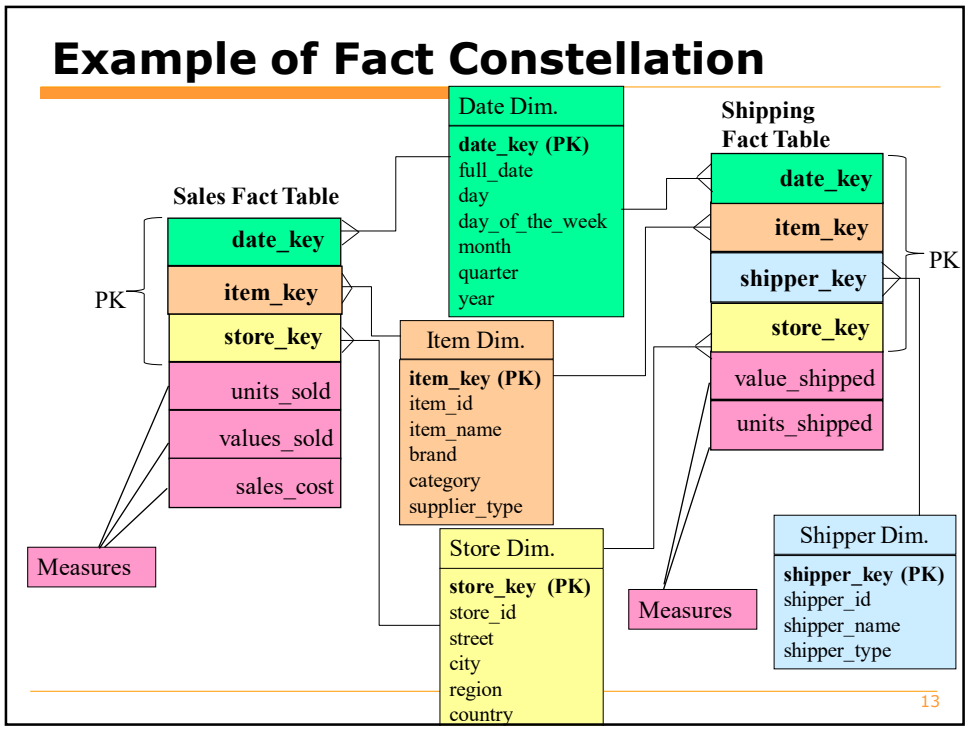
date_key	fulldate	day	day_of_week	month	quarter	year	...
...	...	...	...	...	...	...	...
727	28/12/2022	28	wednesday	12	4	2022	...
728	29/12/2022	29	thursday	12	4	2022	...
729	30/12/2022	30	friday	12	4	2022	...
730	31/12/2022	31	saturday	12	4	2022	...
731	01/01/2023	1	sunday	1	1	2023	...
...	...	...	...	...	...	...	...

11

Example of a Snowflake Schema



12


Example of Fact Constellation
Date Dim.
date_key (PK)
full_date
day
day_of_the_week
month
quarter
year
Shipping
Fact Table
date_key
item_key
shipper_key
store_key
value_shipped
units_shipped
PK
Sales Fact Table
date_key
item_key
store_key
units_sold
values_sold
sales_cost
PK
Item Dim.
item_key (PK)
item_id
item_name
brand
category
supplier_type
Measures
Store Dim.
store_key (PK)
store_id
street
city
region
country
Measures
Shipper Dim.
shipper_key (PK)
shipper_id
shipper_name
shipper_type
13

13