

Data Modeling

2022 / 2023

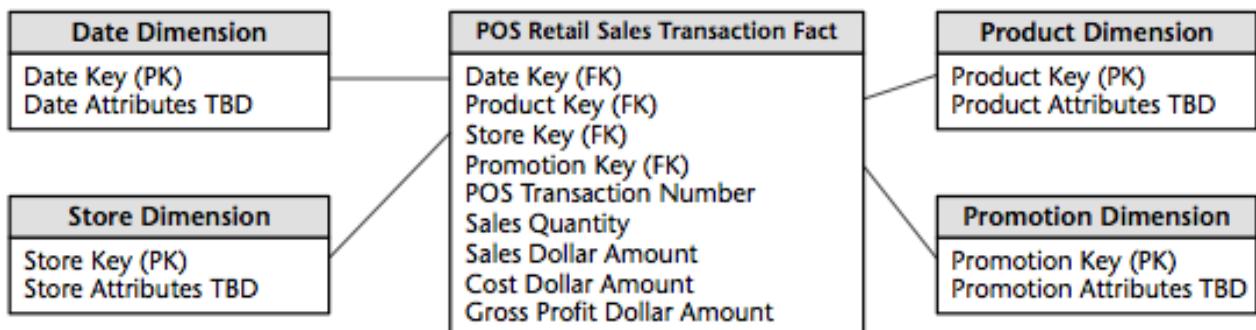
Test 2

16 December, 2022

1:45 hours

1 - Question [10]

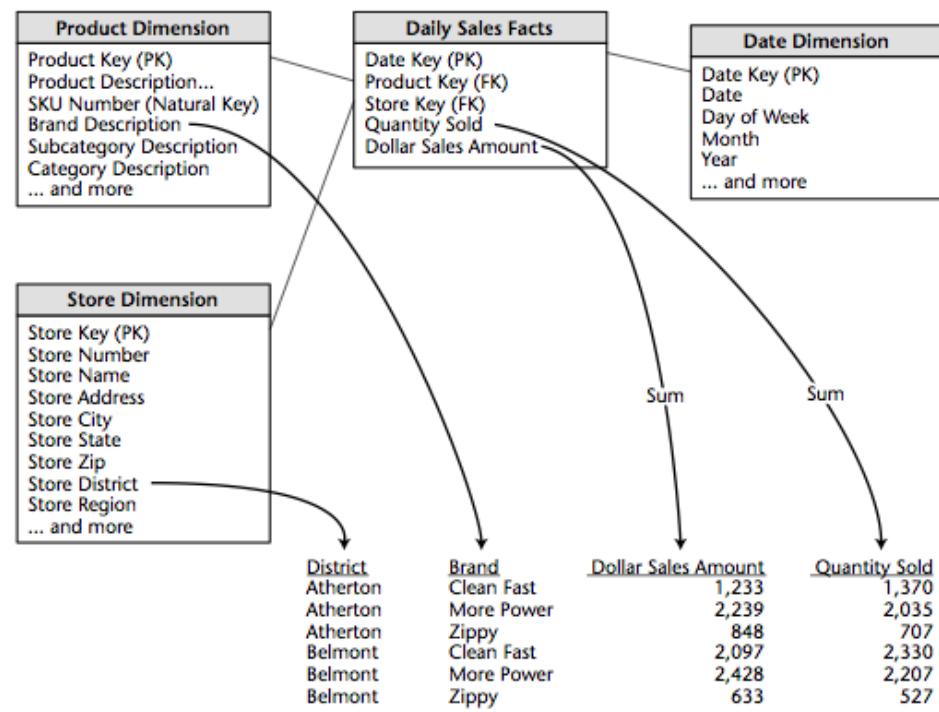
- It is well known that OLTP systems have many differences from DW / OLAP systems. For the following aspects compare OLTP versus DW / OLAP systems: (i) the users; (ii) the pattern of hardware utilization; (iii) the granularity and the amount of information processed in each query.
- Explain the following terms/concepts: (i) degenerate dimension; (ii) fact-less table; (iii) “junk” dimension. For each one provide an appropriate example.
- Kimball proposes three main techniques to deal with the slowly changing dimensions issues. Explain the **type 2** approach and indicate when its use is appropriate. Give an illustrative **example**. **Explain its advantages and its limitations.**
- Very large dimensions may present some challenges namely for browsing and to manage changes. Kimball propose an approach both to cope with the most visited attributes and the attribute's changes. Explain the proposed approach and give an **example** of dimension in which is appropriate to use the approach.
- Consider the next star-schema for the POS retail Sales. Suppose that one intends to extend this model to cope with the following **additional dimensions**: (i) frequent client, (ii) clerk and (iii) the time of day (in minutes). What changes do you propose into the model? Explicitly indicate the modifications on the fact table schema and on the existing records of the fact table. How do you populate any new dimension? What are the implications in terms of previous reports?
- Consider the next star-schema for the POS retail Sales. Explain how to build: (i) a pre-aggregate that summarizes the sales at month, store, product and promotion levels. (ii) a pre-aggregate that summarizes the sales at day and product levels. For each pre-aggregate write the corresponding SQL statement.



2 - Question [4]

The most common schema adopted to implement multidimensional cubes using relational technology is the Star Schema.

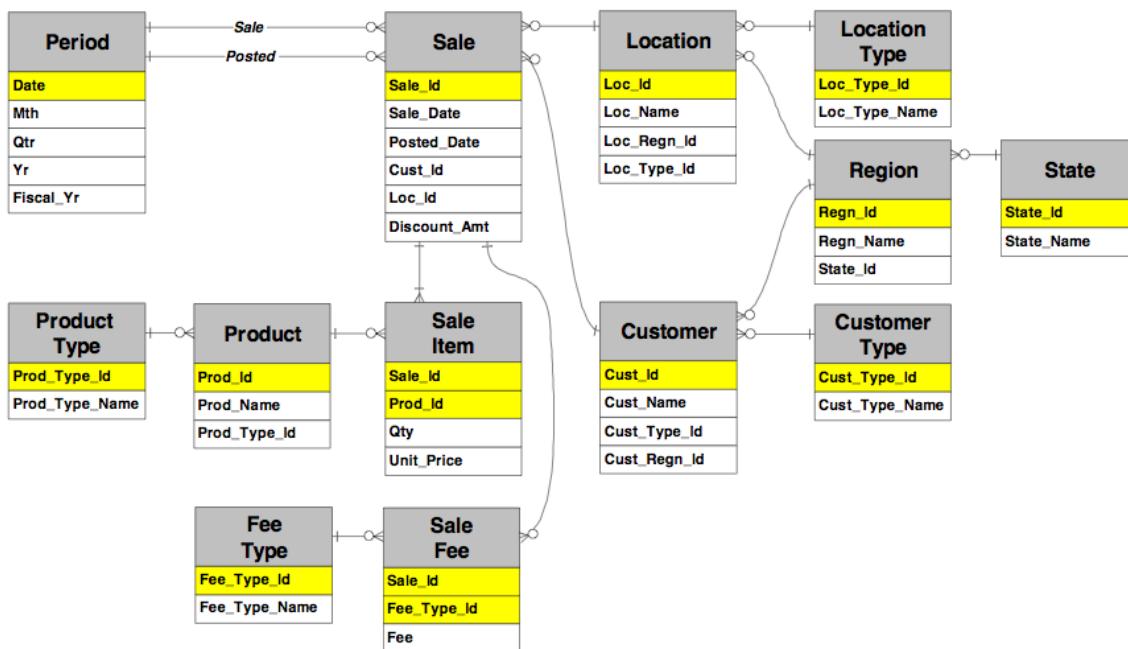
- a) Consider the example shown in the next diagram and the associated result of a typical OLAP query. Assume that: (i) a region named “South Region” that has only two districts “Alberton” and “Belmont”; we are interested on the quarter Q1, 2016 sales for the products from the Category “Home Appliances”. Write the appropriate SQL query that returns the intend result.



- b) Now **modify** your SQL query to consider only the sales of “Belmont”, not detailed by Brand, but showing the results distributed by Month. Explain what kind of OLAP operations are being considered here and justify your answer. Explain how many rows we get as the result of the operation?.

3 - Question [6]

In the paper “From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design” from Daniel L. Moody e Mark A. R. Kortink, a Methodology is proposed to derive multidimensional models from OLTP models. For the next questions consider the following OLTP data model.



- What are the minimal and the maximal entities. **Justify your answer.**
- Describe, in detail, the proposed algorithm to derive a **snowflake schema** based on the transaction table “**Sale Item**”, and consider that we want facts about the number of sold products, the average unity price and the total amount of dollars. What will be the **primary key of the fact table**? Include in your explanation **the sql aggregation** to compute the fact table.
- Indicate, which **terraced schemas** is possible to derive from this OLTP data model. Explain the algorithm.