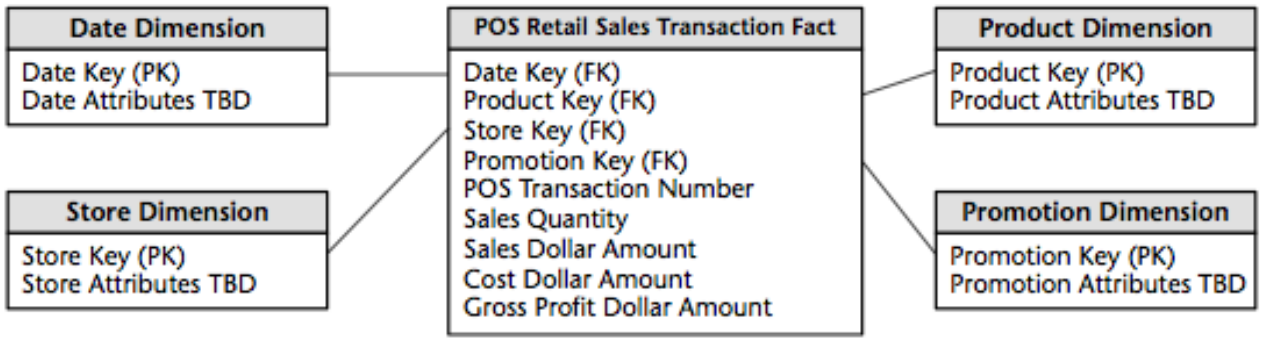
UNIVERSIDADE NOVA DE LISBOA
Faculdade de Ciências e Tecnologia
Departamento de Informática **MIEI**

Data Modeling
2017 / 2018
Test 2

18 December, 2017
1:45 hours

1 - Question [13]
Multidimensional modeling addresses the data model generation for OLAP Data Marts, namely when they are implemented using relational databases.

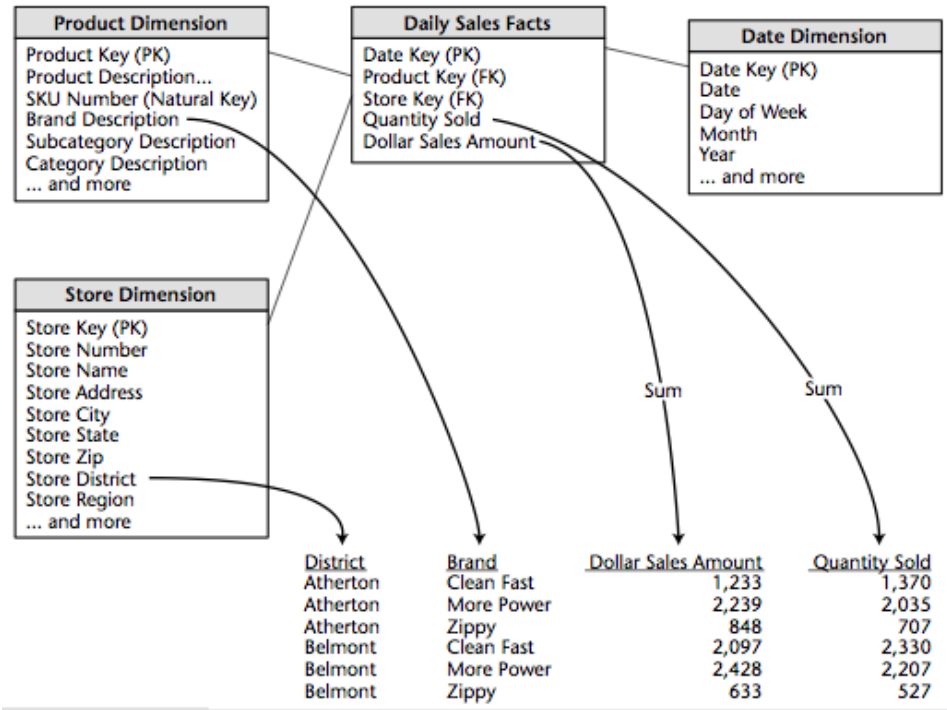
a) The OLAP paradigm is based on a metaphor of a multidimensional cube and has a set of basic operations that are used in interaction with the OLAP cube. Explain the following expressions/operations, their role in OLAP interaction and give an example: slice; drill-down; drill-up; drill-across; aggregation operator.
b) It is a well known recommendation that the primary keys of the dimension tables are surrogate keys. **Explain three** of the most important reasons for such rule.
c) Explain how the Data Warehouse's surrogate keys are managed and mapped to the primary keys from the source systems.
d) Kimball proposes three main techniques to deal with the slowly changing dimension issues. Explain the type 3 approach and indicate when its use is appropriate. Give an illustrative **example.**
e) Very large dimensions may present some challenges namely for browsing and managing changes. Kimball proposes 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 it is appropriate to use the approach.
f) Three types of fact tables can be used on multidimensional models: Transaction Fact Tables; Periodic Snapshot Fact Tables; Periodic Snapshot Fact Tables; Accumulating Snapshot Fact Tables. Describe each one based on the following characteristic: (*i*) Time Period Represented; (*ii*) Grain; (*iii*) Fact table loads and updates. Give an appropriate **example for each case**.
g) Explain Kimball's proposal for dealing with real-time requirements with a transaction fact table. How to model and how to answer the queries?
h) Consider the following star-schema for the POS retail Sales. Suppose that one intends to extend this model to cope with the following **additional dimensions**: frequent client, clerck and the time of day. What changes do you propose into the model? What are the implications in terms of data and in terms of previous reports?
i) The POS retail Sales cube includes a Promotion Dimension. How can we solve the requirement to know the products under promotion on one particular day and on a particular store? How can you get which of those products had no sales?



2 - Question [3]

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. Write the appropriate SQL query that returns the intend result.

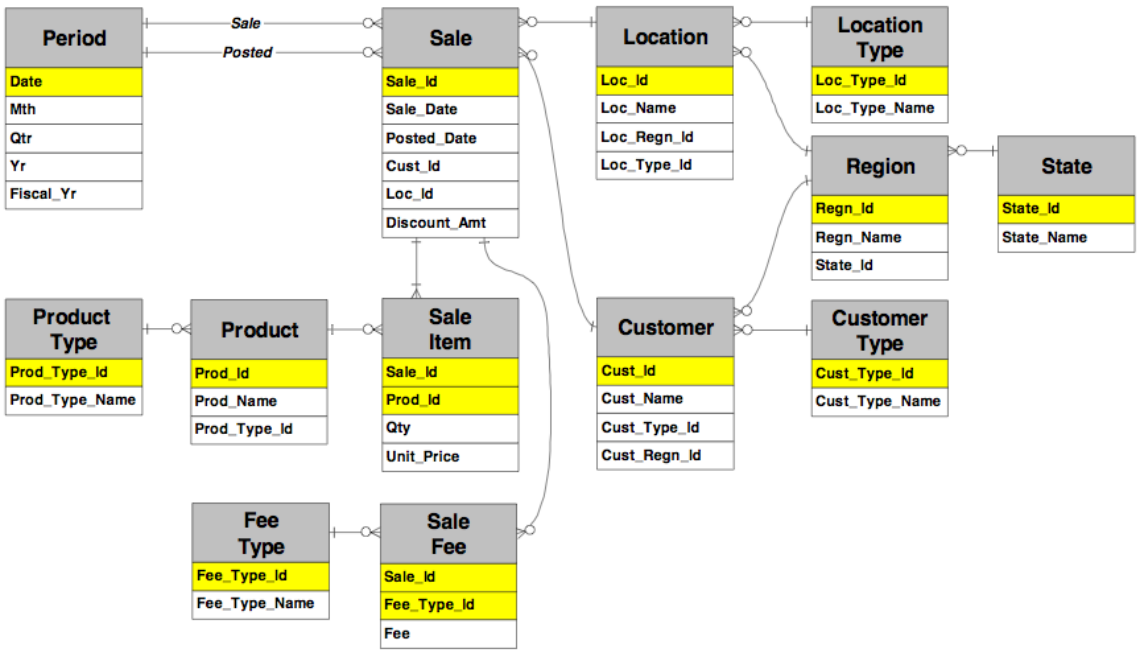


b) Show the results presented in the previous diagram in a matrix form view. The SQL query has to be a different one? **Justify** your answer.

c) Now modify your SQL query to get focused on the date concerning only the 2010 year but showing the results distributed by Month . What kind of OLAP operations are being considered here?

3 - Question [4]

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.



a) What are the minimal and the maximal entities? Justify your answer.
b) What are the transactional, the component and the classification entities? Explain your reasoning.
c) Describe, in detail, the proposed algorithm to derive a **star schema** based on the transaction table "Sale Item".
d) Indicate, which **flat schemas** is possible to derive from this OLTP data model. Explain the algorithm. Explain which issues are raised by the flat schemas for some metrics.