1 Descriptive Analytics Part 2: Data Warehousing & BPM

1.1 Indicate which of the following statements about Data Warehousing are correct

( ) A Data Warehouse is a decentralised storage facility

( ) A Data Warehouse purpose is mainly for operational use

(X) A Data Warehouse purpose is mainly for decision support

(X) A (Enterprise) Data Warehouse contains enter-prise wide data

(X) In a Data Warehouse, data is mostly structured and analytics-ready

(X) In a Data Lake, data can be structured, semi-structured, and unstructured

(X) Data Warehousing is subject-oriented

( ) Data Warehousing is product-oriented

( ) Metadata is the term used to define for raw data that has been cleansed and is ready for analytics

(X) Nonvolatility of a Data Warehouse means that old data is not updated, but rather discarded

( ) Nonvolatility of a Data Warehouse means that the amount of data is kept stable. That is, records are updating when necessary, but a good DW design means that no new records have to be generated because all relevant records are available right from the beginning.

(X) A Data Mart is a smaller "Data Warehouse" focused on a particular subject and business unit

( ) A Data Mart is always dependent on a centralised enterprise data warehouse

(X) In the Hub-and-Spoke architecture, data marts are dependent on a centralised normalised relational data warehouse

( ) In the Data Mart Bus architecture, is always dependent on a centralised enterprise data warehouse

(X) Independent Data Marts are possible, which means they can be set up without a centralised enterprise data warehouse

1.2 Explain the meaning and significance of Metadata in Data Warehousing

- "Data about Data", explains how data is organised and how it is (or should be) used
- Metadata should be handled automatically
- Important for understanding the business or technical context of data, its origin and transformations applied to it

1.3 In Enterprise Application Integration (EAI), the focus is on

( ) Sharing functionality enterprise-wide

( ) Sharing data enterprise-wide

1.4 Name the three main components in ETL and explain them.

extract Reading from one/many Databases

transform Converting in the form required for the DW (or another DB)

load Writing the data (to the DW or another DB)

1.5 Indicate which of the following statements about Enterprise Data Warehouses (EDW) and Data Marts (DM) are correct

(X) The EDW approach is top-down, the DM approach bottom-up

(X) The EDW approach is better suited for large enterprises and large-scale Data Warehousing projects

(X) The EDW approach requires longer time for implementation then the DM approach

(X) The EDW approach can handle larger volumes of data and numbers of users

1.6 Name the three levels of Data Models and explain them.

Conceptual Business view, structure of data in business context, independent of DBMS

Logical Architect view, still indendent of DBMS, specifies entities & attributes, and their relationships and primary/foreign/alternate keys

Physical Developer view, DB schema, DBMS specific; includes referential integrity rules, constraints, event triggers

1.7 Name the two types of tables in a Dimensional Model and explain them.

Fact table contains descriptive decision analysis attributes

Dimension tables contain classification and aggregation information, describing the data in the fact table and views for analysing & summarising it

1.8 Convert ER model to star schema

In this model, you have two entity-types (ORDER and PRODUCT) and one association-type (ORDER_DETAIL). The properties of each element are shown in an ellipse (for instance, product has the properties ID, Name and Price). Your task is to convert this entity-relationship model to a star schema. The data warehouse (for which you will create a star schema) does not store the data the same way as the operational database of the company. The ER model of the operational database is shown in Fig. 1 below.
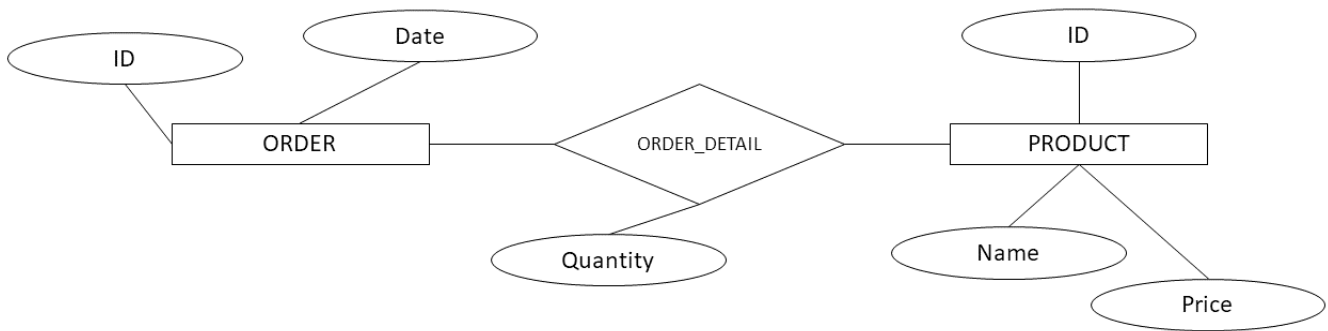

Figure 1: ER Schema

In its data warehouse, the company wants to store only the average quantity of products sold per quarter between January 2010 and December 2017.

1. How many fact tables do you need?   1 (SALES)

2. How many dimension tables do you need?   1 (DIM_QUARTER)

3. When do you compute the average quantity? Justify your answer.

   ( ) In the operational database

   (X) During ETL The company tells explicitly that they want to store only averages by quarter (max. level of granularity). So, you can/must get rid of lower level data and compute average during the transform step in ETL.

   ( ) When facts and dimensions are created in the data warehouse (CREATE TABLE)

   ( ) At query time

4. Draw a (relational) star schema and indicate if properties are primary keys (PK) or foreign keys(FK)
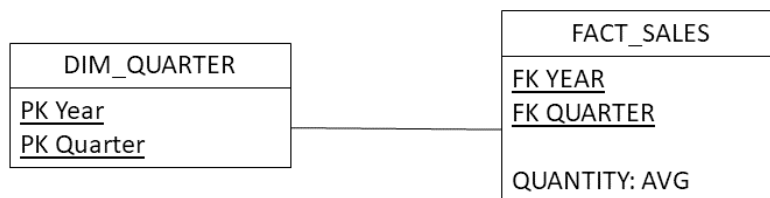

Figure 2: Star Schema

1.9   Indicate which of the following statements about OLAP are correct

(X) A Data Cube is a multidimensional array

(X) A Data Cube is optimised for fast analysis of data

( ) A Data Cube is optimised for efficient storage of data

(X) Slice and dice correspond to subsets of Data Cubes, by selecting selecting 2 or more dimensions for inclusion

1.10   Indicate which of the following statements about KPIs are correct

( ) A KPI is a measure that is independent of the company's strategy

(X) A KPI is a measure that is aligned to the company's strategy

( ) An Outcome KPI is measuring activities that have no immediate impact but might impact future outcomes

( ) A Driver KPI is measuring actvities that have an immediate impact (thus "drive" it)