

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
- () A Data Warehouse purpose is mainly for decision support
- () A (Enterprise) Data Warehouse contains enterprise wide data
- () In a Data Warehouse, data is mostly structured and analytics-ready
- () In a Data Lake, data can be structured, semi-structured, and unstructured
- () 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
- () 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 updated 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.
- () 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
- () 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
- () Independent Data Marts are possible, which means they can be set up without a centralised enterprise data warehouse metadata are data about data. Metadata describe the structure of and some meaning about data, thereby contributing to their effective or ineffective use.

Metadata is used to show how the data was used, how it was changed and how it is organised

1.2 Explain the meaning and significance of Metadata in Data Warehousing

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

- (X) Sharing functionality enterprise-wide

- () Sharing data enterprise-wide

provides a vehicle for pushing data from source systems into the data warehouse. It involves integrating application functionality and is focused on sharing functionality (rather than data) across systems,

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

Extract, transform and load

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

- () The EDW approach is top-down, the DM approach bottom-up
- () The EDW approach is better suited for large enterprises and large-scale Data Warehousing projects
- () The EDW approach requires longer time for implementation than the DM approach
- () 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.

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

1.6 Conceptual logical and physical.

Conceptual: overall view of the structure of data in business context. It is independent of any database or physical storage structure and it may contain objects that are never being implemented in physical DB but needed for understanding.

Logical: Still independent of any database or physical storage structure it represents the first step in designing architecture of the application and specifies the entities & attributes to be implemented. It also further identifies relationships between these entities & attributes, and business rules. It finally defines primary keys, foreign keys, alternate keys.

Physical: Represents the logical data model in a DB schema, DBMS specific Defines physical objects such as tables and columns Specifies referential integrity rules, including foreign keys, constraints, event triggers Contains DBMS-specific performance and optimisation entities

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.

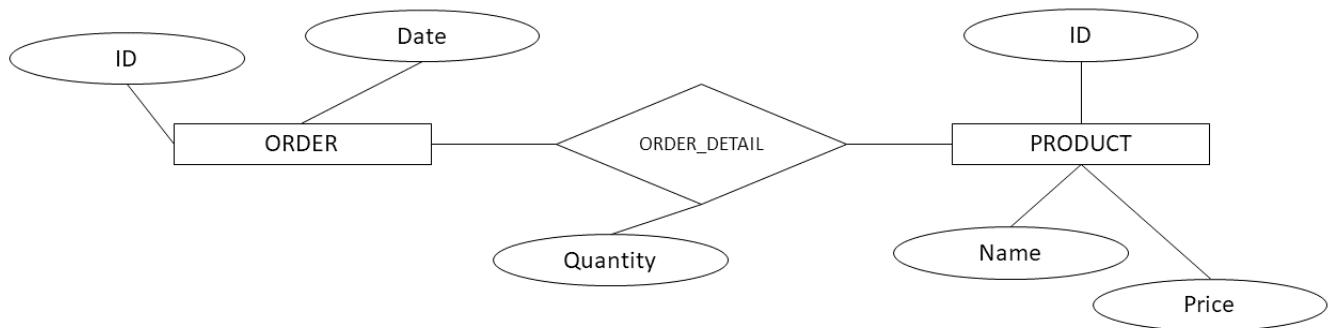


Figure 1: ER Schema

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
2. How many dimension tables do you need? 2: quantity and time
3. When do you compute the average quantity? Justify your answer.
 - () In the operational database
 - () During ETL
 - () When facts and dimensions are created in the data warehouse (CREATE TABLE)
 - () At query time The most efficient and can change in a later stage
4. Draw a (relational) star schema and indicate if properties are primary keys (PK) or foreign keys(FK)

1.9 Indicate which of the following statements about OLAP are correct

- () A Data Cube is a multidimensional array
- () A Data Cube is optimised for fast analysis of data
- () A Data Cube is optimised for efficient storage of data
- () 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
- () 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 activities that have an immediate impact (thus "drive" it)