Database Systems
Homework 9 & 10

Jonathan Henrique Maia de Moraes (ID: 1620855)

12/02/2015

Chapter 13

1 Questions (Pages 362 & 363)

1.1 Question (3.)

A data warehouse is a broad-based, shared database for management decision making that contains data gathered over time. A subject oriented, integrated, non-volatile, and time variant collection of data in support of management's decisions.

1.2 Question (4.)

- **Subject Oriented**: DW are organized around subjects, the major entities of concern in the business environment;

- **Integrated**: All of the data about a subject must be organized or integrated in such a way as to provide a unified overall picture of all the important details about the subject over time;

- **Non-Volatile**: Once a data is added to the DW, it doesn't change;

- **Time Variant**: DW always includes some kind of timestamp to its data;

- **High Quality**: To support a management decision, the data must be high quality;

- **Aggregated**: The type of data that management requires for decision making is generally summarized data. The sheer volume of all of the historical detail data would often make the DW unacceptably huge. The detail data were stored in the DW, the amount of time needed

to summarize the data for management every time a query was posed would often be unacceptable;

- **Denormalized**: Since the data is non-volatile, the existing data in the DW never has to be updated. Being denormalized improves the performance of read-only queries;

- **Not Necessarily Absolutely Current**: A consequence of the kind of typical time schedule for loading new data into the DW.

1.3 Question (5.)

- **EDW**: A large-scale DW that incorporates the data of an entire company or of a major division, site, or activity of a company;

- **DM**: A small-scale DW that is designed to support a small part of an organization, say a department or a related group of departments.

1.4 Question (10.)

- Data Extraction;

- Data Cleaning;

- Data Transformation;

- Data Loading.

1.5 Question (13.)

OLAP is a decision support methodology based on viewing data in multiple dimensions. The OLAP environment's multidimensional data is very well suited for querying and for multi-time period trend analyses. Several other data search concepts are commonly associated with OLAP.

1.6 Question (15.)

Data mining is the searching out of hidden knowledge in a company's data that can give the company a competitive advantage in its marketplace.