Chapter 8 Assignment

Review Questions 2, 3, 4, 6, 11, 14, 15, 16, 17, 22, 24, 25, 27, 33, 38, 40, 42, 44, 46 and 47 on pages 485 – 487

2. How do BI systems differ from transaction processing systems?

* **They do not support operational activities, such as recording and processing orders. Instead, BI systems are used to support management assessment, analysis, planning, control, and, ultimately, decision making.**

3. Name and describe the two main categories of BI systems.

* **Reporting applications, they make elementary calculations on data.**
* **Data mining applications, they use sophisticated mathematical and statistical techniques.**

4. What are the three sources of data for BI systems?

* **Operational databases, extracts of operational databases, and purchased data.**

6. What is an ETL system, and what functions does it perform?

* **Extract, transform, and load systems are used to extract data from operational systems; transform the data and load them into the data warehouses; and maintain metadata that describes the source, format, assumptions, and constraints about the data.**

11. Explain the difference between a data warehouse and a data mart. Give an example other than the ones used in this book.

* **A data mart is a collection of data that is smaller than that held in a data warehouse. SOME EXAMPLE**

14. What is a star schema?

* **It visually represents a star, with a fact table at the center of the star and dimension tables radiating out from the center. The fact table is always fully normalized, but dimension tables may be non-normalized.**

15. What is a fact table? What types of data are stored in fact tables?

* **The central table in a dimensional database. Its attributes are data values that are summed, averaged, or processed in some simple arithmetic manner.**

16. What is a measure?

* **In OLAP, a data value that is summed, averaged, or processed in some simple arithmetic manner.**

17. What is a dimension table? What types of data are stored in dimension tables?

* **In a star schema dimensional database, the tables that connect to the central fact table. Dimensional tables hold attributes used in the organizing queries in analysis such as those of OLAP cubes.**

22. What is the distinguishing characteristic of OLAP reports?

* **They are Dynamic.**

24. Give an example, other than ones in this text, of a measure, two dimensions related to your measure, and a cube.

* **I’m still confused by measures**

25. What is drill down?

* **User-directed disaggregation of data used to break higher-level totals into components.**

27. Define *distributed database*.

* **A database that is stored and processed on more than one computer.**

33. Explain what problems can occur in a distributed database that is partitioned but not replicated.

* **If any transaction updates data that span two or more distributed partitions.**

38. According to this chapter, why were OODBMSs not successful?

40. What is *Big Data?*

* **The current term for the enormous datasets generated by Web applications.**

42. What is the NoSQL *movement*?

* **A Not Only SQL DBMS is often a distributed replicated database that is used for large databases.**

44. What is *Cassandra*, and what is the history of the development of Cassandra to its current state?

* **A nonrelational unstructured data store from the Apache Software Foundation.**

46. Explain MapReduce processing.

* **To break a large analytical task into smaller tasks, assign each task to a different cluster of computers, gather the results of each, and combine them into a final product of the original task.**

47. What is *Hadoop*, and what is the history of the development of Hadoop to its current state? What are *HBase* and *Pig*?

* **Provides standard file services to clustered servers so that their file systems can function as one distributed file system. Was originally part of Cassandra, but it has spun off a nonrelational data store of its own called HBase and a query language named Pig.**