Answer ALL the Questions

1. Fill in the Blanks [1 x 10 =10]
    a. The type of relationship in star schema is _____
    b. _____describes the data contained in the data warehouse.
    c. _____predicts future trends & behaviors, allowing business managers to make proactive, knowledge-driven decisions.
    d. The granularity of the fact is the _____of detail at which it is recorded
    e. With respect to normalization, Fact tables are _____.
    f. In horizontal splitting, we split a relation into multiple tables on the basis of _____.
    g. a process to change data from a detailed level to a summary level is called _____
    h. Data warehouse architecture is based on _____.
    i. Removing duplicate records is a process called _____.
    j. The core of the multidimensional model is the _____. , which consists of a large set of facts and a number of dimensions.

2. Write at least 2 differences between [3 x 2=6]
    a. Star and Snowflake Schema
    b. View and Materialized View
    c. OLAP and Data Warehouse

3. Short Answers [2 x 2= 4]
    a. What is factless fact tables?
    b. What are non-additive measures?

4. Long Answers
    a. How Normalization is different in DWH compared to Database. Discuss different normalization requirements in DWH. [2 + 4 =6]
    b. Discuss the issues that needs to be addressed for designing of data Warehouse that does not follow RDBMS architecture for their day to day operations. Ex: Facebook [4]

c. All Electronics company have sales dept. Sales consider four dimensions namely time, items, branch and location. The schema contains a central fact tables sales with two measures dollars-sold and unit sold. [5 + 3= 8]
    i. Define Snowflake schema for above case
    ii. List the security issues that needs to be managed with the above schema.

d. Consider the following business scenario. A telecom company plans to maintain a CRM data warehouse. There are 10 million customers of the company. Besides the usual attributes, the company wants to maintain additional demographic information like literacy percentage, male/female ratio, average life expectancy and average income of the people belonging to the state to which each customer belongs. The company also wants to maintain information about the age group, income level and marital status of its customers. They also need to run queries like the number of married and unmarried customers they have at any point in time.

[10+4=14]

    i. Design an efficient data warehouse schema that satisfies the above business scenario. Clearly identify the fact table(s), dimension table(s), primary key(s) and foreign key(s).
    ii. Write an SQL statement that generates the number of married and unmarried customers that the company has today.

e. Suppose the fact data is calculated to be 36GB of data per year, and 4 years worth of data are to be kept online. [8]
    i. Calculate the space required
    ii. Now suppose the data is to be partitioned by month and there are four concurrent queries to be allowed , estimate the size of temporary space
    iii. Calculate the total database size
    iv. Calculate the initial space for 6 months worth of data