Data Warehouse and Data Mining

ASSIGNMENT 1

Q1. How is data ware house different from a database? Also Identify the similarity.

Q2. Suppose that a data warehouse consists of the three dimensions time, doctor, and patient, and the two measures count and charge, where charge is the fee that a doctor charges a patient for a visit.

a. Enumerate three classes of schemas that are popularly used for modelling data warehouses.
b. Draw a schema diagram for the above data warehouse using one of the schema classes.
c. Starting with the base cuboid [day,doctor,patient], what specific OLAP operations should be performed in order to list the total fee collected by each doctor in 2010?

Q3. A data warehouse can be modelled by either a star schema or a snowflake schema or fact constellation. Briefly describe the similarities and the differences of the two models, and then analyze their advantages and disadvantages with regard to one another. Give your opinion of which might be more empirically useful and state the reasons behind your answer.

Q4. In data warehouse technology, a multiple dimensional view can be implemented by a relational database technique (ROLAP), by a multidimensional database technique (MOLAP), or by a hybrid database technique (HOLAP).

a. Briefly describe each implementation technique.

b. Which implementation techniques do you prefer, and why?
c. For each technique, explain how each of the following functions may be implemented:
    i. The generation of a data warehouse (including aggregation)
    ii. Roll-up
    iii. Drill-down

Q5. Briefly compare the following concepts. You may use an example to explain your point(s).

a. Data cleaning, data transformation, refresh

Q6. Suppose that a data warehouse for Big University consists of the four dimensions student, course, semester, and instructor, and two measures count and avg grade. At the lowest conceptual level (e.g., for a given student, course, semester, and instructor combination), the avg grade measure stores the actual course grade of the student. At higher conceptual levels, avg grade stores the average grade for the given combination.

a. Starting with the base cuboid [student,course,semester,instructor], what specific OLAP operations (e.g., roll-up from semester to year) should you perform in order to list the average grade of CS courses for each Big University student.