Q1. In this assignment, you will be developing a Web-based framework that allows one to create a fact star schema for a data warehouse that consists of only one data mart and can support some simple OLAP operations. Followings are the required functionalities in this assignment. Each function may require one or more (static or dynamically generated) Web pages, along with interactions with the database.

a) Creating dimensions and its attributes for the specified Data Mart. You should allow the user to specify the number of dimensions and the number of attributes in each dimension. Note that for each dimension, once the user submits the dimension name and its corresponding attributes from the Web, a table corresponding to this information will be created in the database in a programmatic way. For simplicity, apart from the primary key, you can assume that the other attributes are of string (or varchar) type. Users are also allowed to input some records into any created dimension table. (20 marks)

b) Defining a concept hierarchy for any of the dimensions. This information can also be stored in the database as a table. Once created, all dimensions could be reused (selected) by future data marts. (10 marks)

c) Creating a Data Mart. Here a user is allowed to enter the name of the Data Mart into a Web form. After submitting the Data Mart name, it can be recorded in the back-end database. Input a fact table name. User should also be allowed to specify the number of measures in the fact table. Upon submitting the fact table name and its measures, a database table corresponding to the submitted fact table should be created programatically. Then select the previously created dimensions for this data mart. All attributes other than foreign keys in the fact table can be of either int or real number type. When all tables of the data mart have been created, you should provide the option to display the created star schema of the Data Mart in a web page, including all the dimensions, any concept hierarchies, and the fact table. Users can also enter some records into any created fact tables. (10 marks)

d) Performing OLAP operation. Provide an interface for users to perform Roll Up and Drill Down operations by any dimensions on any measures. Note: The Drill Down operations are virtual, since you are required to materialize only the base cuboid. (20 marks)
Once again, note that all the above functionalities are provided from of a Web application. During the demo, we will be creating a new Data Mart and performing some simple OLAP operations online from your system.

To start with, you may also create the following Meta tables in the database:
DataMart (did (primary key), dm_name)

Dimensions (did references DataMart.did, dimid, dim_name). Here did and dimid together form a unique key.

FactTables (did references DataMart.did, factid, fact_name). Here did and factid together form a unique key.

Q2. Suppose that a data warehouse consists of the four dimensions date, spectator, location and game and the two measures, count and charge, where charge is the fare that a spectator pays when watching a game on a given date. Spectators may be students, adults, or seniors with each category having its own charge rate. Assume that the concept hierarchy for the spectator dimension has two levels, date dimension has 4 levels and the other two dimensions have 3 levels (excluding all) [20 marks]

a) Show sample concept hierarchies for each dimension according to the number of levels specified.
b) Draw a Snow-flake schema diagram for the data warehouse.
c) Draw a Star schema diagram for the data warehouse

d) Starting with the base cuboid [date, spectator, location, game] what specific OLAP operations (roll up, slice, select, etc.) that should performed in order to list the total charge paid by student spectators at GM_Place in 2004?

e) How many different cuboids will this cube contain (including the base and apex cuboids)?
Q3. Suppose we define a data cube with three dimensions: place, age, and salary.

The dimensional hierarchies used are “city < state < region” for place, “age_value < age_category [20-30, 30-40, 40-50]” for age and “salary_value < salary_class [50-60K, 60-65K, 65-70K]” for salary.

The measure monitored in the Fact Table is the Number of Subscribers.

What is the total number of cuboids possible from this data cube? [1 marks]

We have the following three materialized cuboids:

Cuboid 1: {state} where age_category = 20-30 AND salary_class = 60-65K
Cuboid 2: {city, age_value} where salary_class = 60-65K
Cuboid 3: {region, salary_class} where age_category = 30-40

Which of the above three cuboids can be selected (i.e. feasible) and, if there are more than one, which is the best among those to process each of the following queries? and why? NOTE: For some queries, the base cuboid may have to be selected.

(a) The query to be processed is {city, age_category} with the selection constant “salary_class = 60-65K”. [3 marks]
(b) The query to be processed is {salary_class} with the selection constant “age_value = 31”. [3 marks]
(c) The query to be processed is {age_value} where salary_class = 60-65K with the selection constant “city = Lafayette”. [3 marks]

Q4. Suppose that the following table is derived by attribute-oriented induction. [10 marks]

<table>
<thead>
<tr>
<th>class</th>
<th>Place</th>
<th>Year</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Programmer</td>
<td>USA</td>
<td>2000</td>
<td>180</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2001</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>2000</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2001</td>
<td>150</td>
</tr>
<tr>
<td>DBA</td>
<td>USA</td>
<td>2000</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2001</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>2000</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2001</td>
<td>85</td>
</tr>
</tbody>
</table>

a) Transform the table into a crosstab showing the associated t-weights and d-weights.

b) Map the target class, Programmer, into a (bidirectional) quantitative descriptive rule with attribute Place, for example, ∀ X, Programmer(X) ⇔ (place(X) = “USA”)[t : x%, d : y%]… ∨ (…)[t : w%, d : z%].

c) Map the target class, DBA, into a (bidirectional) quantitative descriptive rule with attributes Place and Year, for example, ∀ X, DBA(X) ⇔ (place(X) = “USA”∧ year(X)="2000")[t : x%, d : y%]… ∨ (…)[t : w%, d : z%].