CS561: Advanced Topics In Database Systems
Spring-2012

Homework 3

Total Points: 70
Release Date: 03/22/2012
Due Date: 04/05/2012

References: >>> Slides for Data Mining & OLAP, plus the slides for Active Databases.

Question 1 [20 Points]—Data Mining

Transaction ID	Items
1	A, B, C, D
2	A, C, D, F
3	C, D, E, G, A
4	A, D, F, B
5	B, C, G
6	D, F, G
7	A, B, G
8	C, D, F, G

Given the table above where each row represents a transaction and the items sold in this transaction, answer the following questions:

Q1) Find all frequent itemsets using the Apriori technique (given in class) with support higher than or equal to 30% (support is the percentage of transactions containing the itemset).

Hint: As given in lecture (slide 35), create a table divided into scans, and for each scan (say scan number i) identify what are the candidates itemsets of size i considered in this scan (second column in slide 35), and then report the frequent itemsets of size i along with the support of each one (third column in slide 35, but add the support of each itemset).

Q2) What is the support and confidence of the following association rules (check the slides on how to compute the confidence):

A → BD

BD → AC

A → CD

Question 2 [50 Points]—Active Databases & OLAP

Accessing Oracle

You can access your Oracle account by following these instructions:

To Access Oracle account from any CCC machine, follow these steps:
(If you faced problems connecting to Oracle, send email to Taylor, Mark J. <mtaylor@WPI.EDU>)

1) Login to the CCC machine using your WPI account. Or "> ssh ccc.wpi.edu" to remotely login to CCC machine.

1) In the command line terminal, execute the following command that will set the Oracle environment variables:
> **source /usr/local/bin/oraenv**

Otherwise, you can set the Oracle variables manually as follow:
ORACLE_BASE=/usr/local/oracle11gr2
ORACLE_HOME=/usr/local/oracle11gr2/product/11.2.0/db_1
ORACLE_SID=WPI11GR2

2) To connect to Oracle, use command:
> **sqlplus <username>@WPI11GR2**

3) Your password is your username capitalized. After your first login change the password using the following command.
Sql > alter user *<username>* identified by *<new_password>*;

1) **[10 Points]** Create the following relational tables:

Doctor**(SSN, FirstName, LastName, Specialty, YearsOfExperience, city)**
Patient**(SSN, FirstName, LastName, Address, DOB, PrimaryDoctor_SSN)**
Medicine**(TradeName, UnitPrice, GenericFlag)**
Prescription**(Id, Date, Doctor_SSN, Patient_SSN, TotalCost)**
Prescription_Medicine**(Prescription Id, TradeName, NumOfUnits)**

- The **Doctor** relation has attributes Social Security Number (SSN), first and last names, specialty, the number of experience years, and the city (s)he lives in.
- The **Patient** relation has attributes SSN, first and last names, address, date of birth (DOB), and the SSN of the patient's primary doctor.
- The **Medicine** relation has attributes trade name, unit price, and whether or not the medicine is generic (True or False).
- The **Prescription** relation has attributes the prescription id, the date in which the prescription is written, the SSN of the doctor who wrote the prescription, and the SSN of the patient to whom the prescription is written.
- The **Prescription_Medicine** relation stores the medicines (TradeName) written in each prescription (prescriptionId) along with their quantities (number of units).

Hint: Assign to the columns the appropriate data types of your choice and create the correct foreign keys/ primary keys relationships. The primary keys are underlined in the tables above.

2) **[20 Points]** Database triggers convert a passive database to an active one. Your task is to create triggers that perform the following:

- Create database triggers to ensure that no doctor is allowed to serve as a primary doctor for more than 5 patients (You may need multiple triggers for that).

- Create database triggers to ensure that Prescription.TotalCost is always up-to-date and computed as SUM of (Prescription_Medicine.NumOfUnits * Medicine.UnitPrice) for all medicine in that prescription (You may need multiple triggers for that).

3) **[20 Points]** OLAP queries are queries that involve aggregation over the data to compute some statistics. Your task is to write the following aggregation queries:

- For each city, report how many units are sold in that city of each medicine (TradeName) and the total cost of that medicine.

- Group the doctors by their specialty and years of experience, and report for each group, how many patients they serve and how many prescriptions they write.

Hint: You may want to create and insert sample data sets to test your code in the previous questions.

What to Submit
You will submit a single zip file containing the following:
- A text, word doc, or pdf file containing the answer to Question 1
- A single text file containing the SQL commands to answer Question 2
    - The first set of commands are for creating the tables
    - The second set of commands are for creating the required triggers
    - The third set of commands are SQL queries for answering the OLAP queries

How to Submit
Use blackboard system to submit your file zip file.

Late Policy:

We follow the late policy stated on the course website.