Politecnico di Torino
Database Management Systems

June 21st 2012

1. (7 Points) The following relations are given (primary keys are underlined):

   > CONFERENCE(<u>IDC</u>, Title, City, State, Sponsor, Start_Date, End_Date)
   > PAPER(<u>IDP</u>, Title, #Pages, Type, IDA)
   > AUTHOR(<u>IDA</u>, Full_Name, Birth_Date, Affiliation, Nation, Email)
   > REGISTRATION(<u>IDC</u>, <u>IDP</u>, IDA, Date, Amount)

   Assume the following cardinalities:

   - card(CONFERENCE)= 10^4 tuples,
     distinct values of City $\simeq 10^2$,
     distinct values of State $\simeq 10$,
     distinct values of Sponsor $\simeq 10$,
     MIN(Start_Date) = 1/1/2010, MAX(End_Date) = 31/12/2011,
   - card(PAPER)= $5*10^5$ tuples,
     MIN(#Pages) = 2, MAX(#Pages) = 12,
   - card(AUTHOR)= 10^6 tuples,
     MIN(Birth_Date) = 1/1/1960, MAX(Birth_Date) = 31/12/1989.
     distinct values of Affiliation $\simeq 10^3$,
     distinct values of Nation $\simeq 10^2$,
   - card(REGISTRATION)= $5*10^9$ tuples,
     MIN(Date) = 1/1/2010, MAX(Date) = 31/12/2011,
     MIN(Amount) = 100 \$, MAX(Amount) = 600 \$.

   Furthermore, assume the following reduction factor for the group by condition:
   - having count(*)>1 $\simeq \frac{1}{4}$.

   Consider the following SQL query:

   ```
   select Title, Type
   from PAPER
   where #Pages ≥ 8 and
         IDA not in (select IDA from AUTHOR A, CONFERENCE C, REGISTRATION R
                        where R.IDC=C.IDC and R.IDA=A.IDA
                            and Date ≥ 1/11/2011
                            and C.State = 'USA' and A.Nation <> 'USA'
                            and Birth_Date ≥ 1/1/1987
                        group by IDA
                        having count(*)>1);
   ```

   For the SQL query:

   (a) Report the corresponding algebraic expression and specify the cardinality of each node (representing an intermediate result or a leaf). If necessary, assume a data distribution. Also analyze the group by anticipation.

   (b) Select one or more secondary physical structures to increase query performance. Justify your choice and report the corresponding execution plan (join orders, access methods, etc.).

2. (8 Points) The following relations are given (primary keys are underlined, optional attributes are denoted with *):

```
DRUG(DCode, DName, DrugCompany, Price, PerCentDiscount*)
SHOPPING_CART(CCode, CustomerID, Date, State)
CONTENTS_OF_SHOPPING_CART(CCode, DCode, NumberOfPackages, Amount,
AvailabilityState)
LEVEL_OF_WAREHOUSE(DCode, NumberOfAvailablePackages)
SENDING_REQUEST(CCode, DCode, NumberOfPackages)
DRUG_SELECTION(SCode, CCode, DCode, NumberOfPackages)
```

Write the triggers managing the following activities in a on-line drug store. The CONTENTS_OF_SHOPPING_CART table lists the drugs included in the electronic shopping cart described in the SHOPPING_CART table.

(1) *Selection of a drug for the purchase.* Write the trigger to update the contents of the shopping cart when a new product to be entered in the shopping cart is selected (insert in the DRUG_SELECTION table). You should examine the availability in stock of the drug (table LEVEL_OF_WAREHOUSE). Assume that the drug is always present in the warehouse, also with zero packaging available. If the number of available packages is greater than the number of requested packages, you must enter the requested drug in the shopping cart (table CONTENTS_OF_SHOPPING_CART). The availability state should be set to 'A' (i.e., it is available). You should also calculate the amount, according to the price of the drug and the possible applied discount (attribute PerCentDiscount, value expressed as a percentage). Finally, the availability in stock must be updated. If the number of available packages is less than the number of requested packages, the drug should be inserted in the shopping cart for the available quantity. The availability state should be set to 'P' (i.e., it is partial). The available quantity in stock must be updated accordingly. Finally, if there is no availability, the drug should not be included in the shopping cart.

(2) *Integrity constraint on the maximum number of drugs sold with high discount.* The same shopping cart (identified by the CodC attribute) can not include more than 20 items sold at a discount greater than 40%. Write the trigger that manages this integrity constraint.

3. Data Warehouse design

The digital music market is booming and Internet platforms offering legal music downloads grow day by day. Each platform collects data on all the latest song downloads requested by customers. To evaluate the profitability of this new market and predict future developments, a data warehouse providing an overview of the current trends and usage insights is required. In particular, the data warehouse must be designed to analyze:

- the average download time per song,
- the average daily downloads (number of songs),
- the average daily income per song.

Analysis must be performed according to:

- date, month, quarter, 4-month period, 6-month period, year of the download,
- time slot (8:00-12:00, 12:00-16:00, 16:00-20:00, etc.) and time of the day (a.m., or p.m.) of the download,
- song, record label (i.e., the company publishing the song), album where the song was published, year of recording of the album,
- song author,
- musical genre of the song,
- singer or band, and their nationality,
- city, province, region, and country of the customer who downloaded the song,
- payment method (credit card, pre-paid, etc.).

The data warehouse contains data related to the years 2007, 2008, 2009, 2010, and 2011. The data warehouse must allow efficient execution of all the following queries:

(a) Considering only downloads in the second quarter of 2010, of customers from the Abruzzo region, select the total number of downloaded songs and the average download time per song, separately for each album.

(b) For each author and for each time slot, select the total number of downloaded songs, separately for each record label. Sort and rank the results for increasing total download number.

(c) For each song authored by Italian singers/bands and considering only downloads in 2011, for each month select the total income, the average daily income, the average daily number of downloads, and the cumulative number of downloads since the beginning of the year, separately for each musical genre.

Design

(a) (6 Points) Design the data warehouse to address the described issues. In particular, the designed data warehouse must allow efficient execution of all the above queries.

(b) (8 Points) Write the frequent queries (a) and (c) using the extended SQL language.