Politecnico di Torino
Database Management Systems

September 7^{th} 2012

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

   USER (UID, Nickname, Name, Surname, State, Birth_Date, E-Mail)
   BLOG_ARTICLE(BID, Title, Text, UID, Publication_Date, Language)
   COMMENT(BID, UID, Timestamp, Text, Vote)
   DAILY_VISIT(BID, Date, #Visitors)

   Assume the following cardinalities:

   - card(USER)= 10^4 tuples,
     distinct values of State $\simeq 10$,
     MIN(Birth_Date) = 1/1/1945, MAX(Birth_Date) = 31/12/1994,
   - card(BLOG_ARTICLE)= 10^7 tuples,
     MIN(Publication_Date) = 1/1/2009, MAX(Publication_Date) = 31/12/2011,
     distict values of Language $\simeq 10$,
   - card(COMMENT)= 10^{10} tuples,
     MIN(TO_DATE(Timestamp)) = 1/1/2009, MAX(TO_DATE(Timestamp)) = 31/12/2011,
     MIN(Vote) = 1, MAX(Vote) = 10,
   - card(DAILY_VISIT)= 10^9 tuples,
     MIN(Date) = 1/1/2009, MAX(Date) = 31/12/2011,
     MIN(#Visitors) = 1 , MAX(#Visitors) = 1000.

   Furthermore, assume the following reduction factor for the group by condition:

   - having AVG(Vote)>7 $\simeq \frac{1}{10}$.

   Consider the following SQL query:

```
select Title
from BLOG_ARTICLE B, DAILY_VISITS V
where B.BID=V.BID and
      V.Date >= 1/1/2011 and Language='Italian' and
      exist (select * from COMMENT C, USER U
                     where C.UID=U.UID
                           and TO_DATE(Timestamp) ≥ 1/1/2011
                           and U.State = 'Italy' and Birth_Date ≥ 1/1/1985
                           and U.UID=B.UID
                     group by UID
                     having AVG(Vote)>7);
```

   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 *):

```
ACTIVITY (ACode, AName, TypeOfActivity, Date, TimeSlot, AgeRange, RCode, NumberOfTeachers)
ROOM (RCode, RName, Capacity)
VALUES_OF_AGE_RANGE (AgeRange, MinimumAge, MaximumAge)
VALUES_OF_TIME_SLOT (TimeSlot, StartHour, EndHour)
CHILD (CCode, CName, Age)
ENROLLMENT (CCode, ACode)
ENROLLMENT_REQUEST(RequestCode, CCode, ACode)
```

Write the triggers to manage the following activities relating to a baby parking. The `ACTIVITY` table lists the planned activities. The activities are held every day in different time slots (attribute `TimeSlot`) and vary according to the age of children (attribute `AgeRange`).

(1) *Enrollment in an activity.* It is requested to register a child for an activity (insert in the `ENROLLMENT_REQUEST` table). The registration is possible only if all the following conditions are satisfied. (a) The child meets the constraint on the minimum and maximum age required for the activity. (b) The number of enrollments made for that activity (table `ENROLLMENT`) is less than the capacity of the room in which the activity will take place (table `ROOM`). (c) The child is not already registered to other activities on the same date and time slot. If at least one of the above conditions is not met, the application for enrollment should not be executed.

(2) *Integrity constraint on planning activities.* Each type of activity can be proposed, for each age range, at most three times in the same day. All modification operations on the `ACTIVITY` table violating the integrity constraint must not be executed. Write the trigger enforcing this integrity constraint.

3. Data Warehouse design

The Italian Civil Aviation Authority (ENAC) performs monitoring functions related to the air transport system. The company would like to evaluate the efficiency of the air companies by analyzing their income and the travel time for each route. To evaluate these features and predict possible inconveniences, a data warehouse providing an overview of the current flight distribution is required. In particular, the data warehouse must be designed to analyze:

- the average number of passengers per flight,
- the average travel time per flight,
- the daily average income per flight.

Analysis must be performed according to:

- date, day of the week, day of the year, month, 2-month period, 3-month period, 6-month period, year,
- the air company
- the airplane model, the building firm,
- the comfort configuration of each airplane model (i.e., TV, business class, etc.
- airport, city, province, region, country and continent both of the departure and arrival.

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) For each air company, select the monthly number of flights, the monthly number of passengers, and the cumulative monthly income since the beginning of the year, separately for each departure state. Consider only the year 2011 and BOEING 747 airplanes.

(b) Considering only the airports located in Europe, for each city of departure located in Italy, for each city of arrival and for each month, select the daily average total income, the average number of passengers per flight, the average travel time per flight, and the total number of passengers for the departure city, separately for each air company.

(c) For each airplane model with the business class, select the total number of flights and the total number of passengers, separately for each air company. Sort and rank the results for increasing total number of flights.

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 (b) using the extended SQL language.