

Database Management Systems

February 7th, 2011

1. (6 Points) The following relations are given (primary keys are underlined):

```
SINGER(SCode, SName, City, DateofBirth)
RECORD_LABEL(RLCode, RLName, Address, City)
DISC(DCode, Title, SCode, RLCode, Type, Price)
SALE(DCode, Date, SoldCopyNumber)
```

Assume the following cardinalities:

- $\text{card}(\text{SINGER}) = 10^4$ tuples,
 $\text{MIN}(\text{DateofBirth}) = 1-1-1969$, $\text{MAX}(\text{DateofBirth}) = 31-12-1998$,
- $\text{card}(\text{RECORD_LABEL}) = 10^4$ tuples,
number of City $\simeq 100$,
- $\text{card}(\text{DISC}) = 10^6$ tuples,
 $\text{MIN}(\text{Price}) = 10$, $\text{MAX}(\text{Price}) = 29$,
- $\text{card}(\text{SALE}) = 10^8$ tuples for year 2010.

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

- $\text{having sum}(\text{SoldCopyNumber}) \geq 10.000 \simeq \frac{1}{100}$.

Consider the following SQL query:

```
select RLName, City
from RECORD_LABEL RL
where City='Rome' and RLCode IN (select D.RLCode
                                from SALE S, DISC D, SINGER SI
                                where S.DCode=D.DCode and SI.SCode=D.SCode
                                D.Price  $\leq$  11 and SI.DateofBirth  $\leq$  31/12/1983
                                and S.Date  $\leq$  31/01/2010
                                group by S.DCode, D.RLCode
                                having sum(SoldCopyNumber)  $\geq$  10.000)
```

For the SQL query:

- 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.
- 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. (7 Points) The following relations are given (primary keys are underlined, optional attributes are denoted with *).

```
EVENT(ECode, EventName, EventCategory, EventCost, EventDuration)
CALENDAR_OF_EVENTS(ECode, Date, StartHour, Place)
CATEGORY_SUMMARY(EventCategory, Date, TotalNumberOfEvents, TotalCostOfEvents)
```

The trigger application deals with scheduling the events for the 150th anniversary of Italy's unification (Italia 150) in the city of Torino. Each event belongs to a category (attribute `EventCategory`), e.g., exhibition, discussion, or film projection, and is characterized by an implementation cost (attribute `EventCost`). Each event can be repeated more times in different dates. The `CALENDAR_OF_EVENTS` table reports the scheduling of events in various dates and locations in Torino. Write the triggers managing the following activities.

- (1) *Update the CATEGORY_SUMMARY table.* The `CATEGORY_SUMMARY` table reports, for each event category and for each date, the total number of scheduled events and the total cost to implement them. When a new event is scheduled (a new record is inserted in `CALENDAR_OF_EVENTS`), the modifications on the `CALENDAR_OF_EVENTS` table should be propagated to the `CATEGORY_SUMMARY` table. Write the trigger managing this activity.
- (2) *Integrity constraint on the maximum cost of an event.* The cost of an event belonging to the film projection category (attribute `EventCategory`) cannot exceed 1500 euro. If a cost higher than 1500 is entered in the `EVENT` table, the attribute `EventCost` should be set to 1500. Write the trigger enforcing this integrity constraint.

3. Data Warehouse design

The Piedmont Region wants to analyze the admissions and the usage of the hospitals to identify the structures requiring financial investments or cost cuts. To perform these analyses, a suitable data warehouse must be designed.

Each hospital is divided into several departments (e.g., department of internal medicine, emergency department, cardiology department, etc.). For each department, the number of available beds for the admissions is known for each room. Each bed can be equipped with different features (e.g., oxygen, automatic reclining, orthopedic mattress, etc.).

The Piedmont Region is interested in analyzing the percentage of free beds according to:

- the date, the month, the 2-month period, the trimester, the 6-month period and the year;
- the day of the year (from 1 to 366), the day of the week (monday-sunday), the day of the month (1-31);
- the department, the hospital, the city district, the city, and the province;
- the bed equipment (oxygen, automatic reclining, orthopedic mattress, etc.).

The following are some of the frequent analysis the Region is interested in:

- (a) Considering only year 2010, for each hospital and each month, select the percentage of free beds. Associate a rank to the results (rank 1 is the highest percentage).
- (b) For each bed equipment, select the total number of free beds during the different days of the week and the percentage of free beds in each day of the week respect to all the days of the week. Associate a rank to total number of free beds in descending order.
- (c) Considering only the beds with oxygen and orthopedic mattress, select the total number of free beds for each hospital and each month.

The data warehouse will store information about the last five full years. The following cardinalities are known (suppose data is uniformly distributed):

- Departments for each hospital: ~ 20
 - Hospitals: ~ 100
 - Cities: ~ 30
 - number of different bed equipments: 10
- (a) (7 Points) Design the data warehouse to address the described issues. In particular, the designed data warehouse must allow efficient execution of all the queries described in the specifications.
 - (b) (4 Points) Write the frequent query (a) of the “problem specifications” using the extended SQL language.
 - (c) (*Optional*: 5 Points) Write the frequent query (b) of the “problem specifications” using the extended SQL language.