Politecnico di Torino
Database Management Systems

A.A. 2016-2017
Homework #1

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

$$\text{RETAILER(\underline{RId}, Name, Category, Website)}$$
$$\text{DEAL(\underline{DId}, Title, Description, Price, Region, RId)}$$
$$\text{USER(\underline{UId}, Username, City, Region, Birth\_Date)}$$
$$\text{PURCHASE(\underline{DId}, \underline{UId}, \underline{Date}, Date\_Expire, Date\_Used)}$$

Assume the following cardinalities:

- card(RETAILER)= 10^2 tuples,
  number of Categories $\simeq 10$,

- card(DEAL)= 10^6 tuples,
  number of Regions $\simeq 20$,
  MIN(Price) = 1 Euro, MAX(Price) = 500 Euro,

- card(USER)= 10^7 tuples,
  number of Cities $\simeq 10^5$,
  number of Regions $\simeq 20$,
  MIN(Birth_Date) = 01-01-1950, MAX(Birth_Date) = 31-12-1990,

- card(PURCHASE)= 10^{10} tuples,
  MIN(Date) = 01-01-2015, MAX(Date) = 31-12-2015,
  MIN(Date_Expire) = 01-03-2015, MAX(Date_Expire) = 31-06-2016,
  MIN(Date_Used) = 01-01-2015, MAX(Date_Expire) = 31-06-2016,

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

- having count(distinct UId) $\geq 100 \simeq \frac{1}{100}$.

Consider the following SQL query:

```
select Region, SUM(Price)
from DEAL D, RETAILER R
where D.RId=R.RId and Category <> 'Electronics'
      and DId IN (select DId
                    from PURCHASE P, USER U
                    where P.UId=U.UId
                    and P.Date ≥ 01-12-2015 and P.Date ≤ 31/12/2015
                    and U.Region = 'Piedmont'
                    and U.Birth_Date ≥ 01-01-1970
                    group by DId
                    having count(distinct UId) ≥ 100)
group by Region;
```

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. Execute the following queries using Oracle SQL Developer on the database provided for the laboratory Practice #1 (EMP, DEPT, SALGRADE tables). See instructions of Practice #1 on the course website for connecting to the database and using Oracle SQL Developer in the Lab.

- Query #1
```
select *
from EMP E, DEPT D
where E.deptno=D.deptno AND E.sal NOT IN (
                      select hisal from salgrade S
                      where hisal > 500 and hisal < 1900);
```

- Query #2
```
select *
from EMP E, DEPT D
where E.deptno=D.deptno AND E.sal NOT IN (
                      select hisal from salgrade S
                      where hisal > 500 and hisal < 600);
```

(a) Report the execution plan provided by Oracle (screenshots of Oracle SQL Developer results) without any additional indexes beyond the system indexes that Oracle creates on primary keys.

(b) Create one or more secondary physical structures to improve the query execution plans. Test and report the new execution plans (screenshots of Oracle SQL Developer results), providing justifications of your choices, and describing the changes in the execution plans with respect to the previous point.