

Politecnico di Torino

Database Management Systems

July 8th 2011

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

```
USER(UId, Name, Surname, City, State, BirthDate)
PHOTO(PId, UId, Resolution, KBs)
UPLOAD(PId, Date, Time, Description)
TAG(PId, Tag)
```

Assume the following cardinalities:

- $\text{card}(\text{USER}) = 10^6$ tuples,
MIN(BirthDate) = 1-1-1941, MAX(BirthDate) = 31-12-1990,
number of City $\simeq 10^2$,
number of State $\simeq 10$,
- $\text{card}(\text{PHOTO}) = 10^8$ tuples,
number of Resolution $\simeq 10$,
MIN(KBs) = 10^2 , MAX(KBs) = $2 \cdot 10^3$,
- $\text{card}(\text{UPLOAD}) = 5 \cdot 10^8$ tuples,
MIN(Date) = 01-01-2010, MAX(Date) = 31-12-2010,
- $\text{card}(\text{TAG}) = 10^9$ tuples

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

- $\text{having count}(\text{distinct tag}) \leq 10 \simeq \frac{99}{100}$.

Consider the following SQL query:

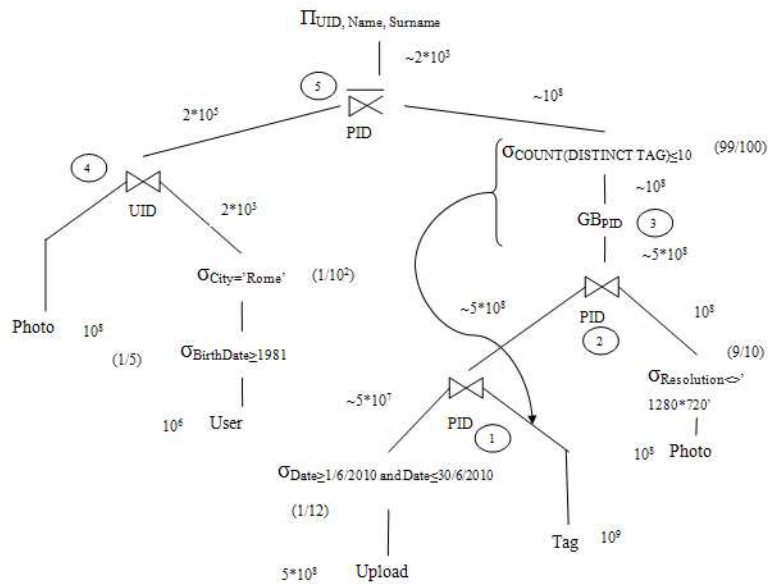
```
select UId, Name, Surname
from USER U, PHOTO P
where U.UId=P.PId and BirthDate > 1-1-1981
      and City = 'Rome'
      and PId NOT IN (select UP.PId
                      from UPLOAD UP, TAG T, PHOTO P1
                      where UP.PId=T.PId and UP.PId=P1.PId
                      and Date  $\geq$  01-06-2010 and Date  $\leq$  30/06/2010
                      and Resolution <> '1280x720'
                      group by UP.PId
                      having count(distinct Tag)  $\leq$ 10);
```

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.).

Join and group by discussion:

- Hash Join



- (2) Hash Join
- (3) GB No Hash
- (4) Hash Join or Nested loop
- (5) Hash Join

Indexes:

- Table Upload: B⁺-Tree on Date
- Table User: Hash on City