

Politecnico di Torino

Database Management Systems

Homework 1

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

```
HOTEL (HID, Hotel_Name, City, Nation, Category, HCID)
HOTEL_CHAIN(HCID, Name, Year_of_foundation)
RESERVATION(RID, StartDate, EndDate, Rate, #People, HID)
COMMENT(CID, Description, Vote, Website, HID)
```

Assume the following cardinalities:

- $\text{card}(\text{HOTEL}) = 10^6$ tuples,
distinct values of Nation $\simeq 10$,
 $\text{MIN}(\text{Category}) = 0$, $\text{MAX}(\text{Category}) = 5$,
- $\text{card}(\text{HOTEL_CHAIN}) = 10^4$ tuples,
unique constraint on Name,
 $\text{MIN}(\text{Year_of_foundation}) = 1/1/1970$, $\text{MAX}(\text{Year_of_foundation}) = 31/12/1999$,
- $\text{card}(\text{RESERVATION}) = 10^{10}$ tuples,
 $\text{MIN}(\text{StartDate}) = 1/1/2010$, $\text{MAX}(\text{StartDate}) = 31/12/2012$,
 $\text{MIN}(\text{EndDate}) = 1/1/2010$, $\text{MAX}(\text{EndDate}) = 31/12/2012$,
 $\text{MIN}(\#\text{People}) = 1$, $\text{MAX}(\#\text{People}) = 3$,
- $\text{card}(\text{COMMENT}) = 10^7$ tuples,
 $\text{MIN}(\text{Vote}) = 1$, $\text{MAX}(\text{Vote}) = 10$,
distinct values of Website $\simeq 10^2$,

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

- having $\text{SUM}(\#\text{People}) < 100 \simeq \frac{9}{10}$,
- having $\text{count}(\ast) > 1000 \simeq \frac{1}{10}$,

Consider the following SQL query:

```
select Website, count(*)
from COMMENT
where Vote > 9
      and HID NOT IN (select HID
                      from HOTEL H, HOTEL_CHAIN HC, RESERVATION R
                      where H.HCID=HC.HCID and R.HID=H.HID
                         and StartDate ≥ 1/6/2012 and StartDate ≤ 31/8/2012
                         and #People = 1 and HC.Name <> 'Ibis'
                      group by HID
                      having SUM(#People)<100)
group by Website
having count(*) > 1000;
```

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.).