# Database Management Systems

## Homework 1

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

> PLAY-ACTOR(<u>AId</u>, Name, Nationality, BirthDate)
> COMEDY(<u>ComId</u>, Title, Director, SceneNumber, Year)
> PLAY-ACTOR-IN-COMEDY(<u>ComId</u>, <u>AId</u>, Role)
> PLANNING(<u>ComId</u>, <u>Theater</u>, <u>Date</u>, StartTime, LengthOfTime)

Assume the following cardinalities:

- card(PLAY-ACTOR)= $10^4$ tuples,
  MIN(BirthDate) = 1-1-1960, MAX(BirthDate) = 31-12-1999,

- card(COMEDY)= $10^3$ tuples,
  distinct values of SceneNumber $\simeq 15$,

- card(PLAY-ACTOR-IN-COMEDY)= $10^6$ tuples,
  distinct values of Role $\simeq 30$,

- card(PLANNING)= $10^8$ tuples,
  MIN(Date) = 1-1-2010, MAX(Date) = 31-12-2010,
  MIN(LengthOfTime) = 81, MAX(LengthOfTime) = 180,

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

- having count(DISTINCT Theater)$\geq$50 $\simeq \frac{1}{10}$.

Consider the following SQL query:
```
select Title, Director
from COMEDY C, PLANNING P, PLAY-ACTOR-IN-COMEDY AC
where P.ComId=C.ComId and AC.ComId=C.ComId
      and LengthOfTime=180 and SceneNumber>12
      and AC.Aid in (select Aid from PLAY-ACTOR
                       where BirthDate ≥ 1996)
group by ComId, Title, Director
having count(DISTINCT Theater) ≥ 50
```
For the SQL query:

1. 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. Analyze also the group by anticipation.

2. 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.).