Politecnico di Torino
Database Management Systems

September 21^{st} 2011

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

   PLAY-ACTOR(AId, Name, Nationality, BirthDate)
   COMEDY(ComId, Title, Director, SceneNumber, Year)
   PLAY-ACTOR-IN-COMEDY(ComId, AId, Role)
   PLANNING(ComId, Theater, Date, 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:

   (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.).

   Join and group by discussion:

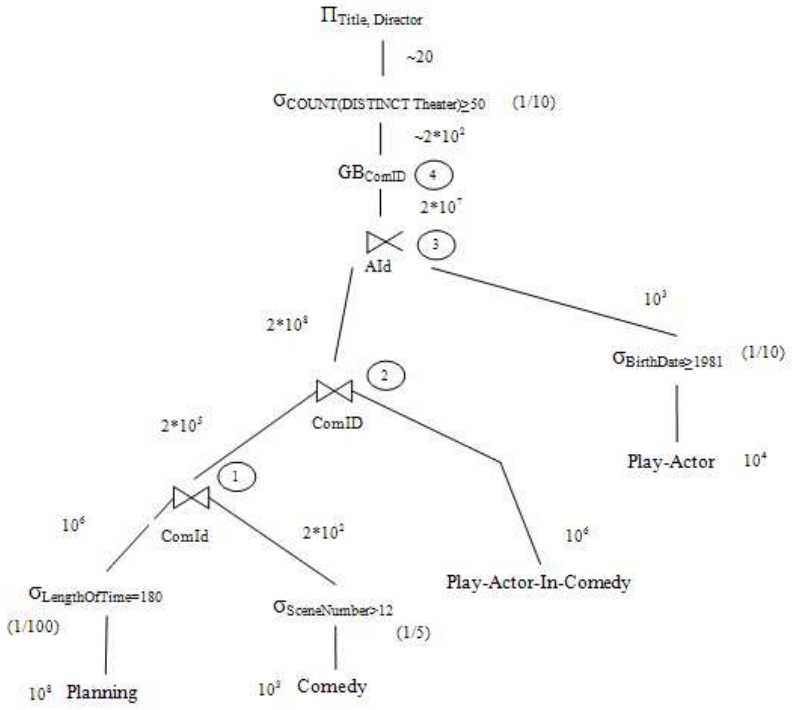
   (1) Nested loop

   (2) Hash Join

   (3) Hash Join or Nested loop

   (4) GB Hash

   Indexes:



- Table PLANNING: B$^+$-Tree on LengthOfTime
- Table PLAY-ACTOR: B$^+$-Tree on BirthDate

Group by anticipation:

Pay attention: The distinct statement should be insert into the select clause since a many-to-many relationship holds among comedy and actor tables.

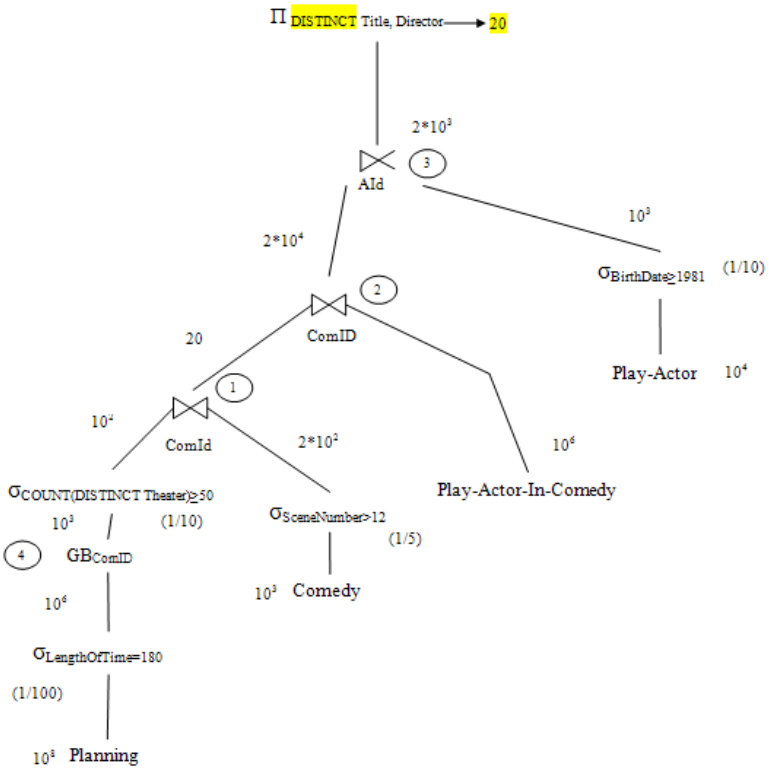
Join and group by discussion:

(1) Nested loop

(2) Nested loop

(3) Hash Join or Nested loop

(4) GB Hash


Π DISTINCT Title, Director → 20
2*10³
3
AId
10³
2*10⁴
σBirthDate≥1981 (1/10)
2
ComID
20
Play-Actor 10⁴
1
10⁴
ComId
2*10²
10⁶
σCOUNT(DISTINCT Theater)≥50
Play-Actor-In-Comedy
10³ (1/10)
σSceneNumber>12
(1/5)
4
GBComID
10³ Comedy
10⁶
σLengthOfTime=180
(1/100)
10⁵ Planning