

Ottimizzatore

1. Sono date le relazioni seguenti (le chiavi primarie sono sottolineate):

```
ATTORE(AId, Nome, Cognome, Cittadinanza, DataNascita)
FILM(FId, Titolo, Regista, Tipo, Anno)
RECITAZIONE(FId, AId, Ruolo)
CINEMA(CId, Sala, #Posti, TipoSchermo)
PROGRAMMAZIONE-FILM-CINEMA(CId, Sala, FId, Data, OraInizio, OraFine)
```

Si ipotizzino le seguenti cardinalità:

- $\text{card}(\text{ATTORE}) = 10^4$ tuple,
numero di Cittadinanze $\simeq 100$,
- $\text{card}(\text{FILM}) = 10^6$ tuple,
numero di Tipi $\simeq 10$,
numero di Anni $\simeq 10$,
numero di Registri $\simeq 100$,
- $\text{card}(\text{RECITAZIONE}) = 10^8$ tuple,
- $\text{card}(\text{CINEMA}) = 10^3$ tuple,
numero Di TipoSchermo $\simeq 10$,
- $\text{card}(\text{PROGRAMMAZIONE-FILM-CINEMA}) = 10^8$ tuple,
 $\text{MIN}(\text{Data}) = 1-1-2009$, $\text{MAX}(\text{Data}) = 31-12-2009$,

Inoltre si ipotizzino i seguenti fattori di riduzione per le condizioni di group by:

- per la query (a), $\text{having count}(\ast) \geq 10 \simeq \frac{1}{10}$.
- per la query (b), $\text{having count}(\ast) \geq 100 \simeq \frac{1}{10}$.

Si considerino le seguenti query SQL:

- (a)

```
select distinct Nome, Cognome
from ATTORE A, RECITAZIONE R, FILM F
where A.AId=R.AId and F.FId=R.FId
and A.Cittadinanza='Italiana' and F.Regista= 'Woody Allen'
and exists (select * from PROGRAMMAZIONE-FILM-CINEMA PFC
            where PFC.Data ≥ '01/08/2009' and PFC.Data ≤ '31/08/2009' and F.FId=PFC.FId)
group by AId, Nome, Cognome
having count(*) ≥ 10
```
- (b)

```
select distinct PFC.FId, PFC.CId
from PROGRAMMAZIONE-FILM-CINEMA PFC, CINEMA C
where PFC.CId=C.CId and PFC.Sala=C.Sala
and PFC.Data='01/07/2009' and C.TipoSchermo='LED'
and PFC.FId not in (select F.FId
                    from FILM F, RECITAZIONE R
                    where F.FId=R.FId and F.Tipo='Commedia'
                    group by F.FId
                    having count(*) ≥ 100)
```

Separatamente per ciascuna delle due interrogazioni SQL

- (a) Si scriva l'espressione algebrica corrispondente, indicando le operazioni svolte, la cardinalità e la selettività di ogni operazione. Dove necessario, si ipotizzi la distribuzione dei dati. Discutere la possibilità di anticipare l'operatore GROUP BY.

- (b) Si scriva il piano di esecuzione che potrebbe essere scelto dall'ottimizzatore in assenza di strutture fisiche accessorie (ordine e tipo dei join, accesso alle tabelle).
- (c) Si scelgano le strutture fisiche accessorie per migliorare le prestazioni dell'interrogazione. Si motivi la scelta e si definisca il nuovo piano di esecuzione ipotizzato.