

Homework 3

Data Science e Tecnologie per le basi di dati

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

ALLOGGIO(CodA, NumeroDiOspiti, Indirizzo, Citta',
Regione)
SERVIZIO(CodS, NomeServizio, TipoServizio)
ALLOGGIO-HA-SERVIZIO(CodA, CodS)
UTENTE(CodU, Nome, Cognome, BusinessAccount, DataNascita,
Indirizzo, Citta', Regione)
PRENOTAZIONE(CodA, DataInizio, CodU, DataFine)

Assumere le seguenti cardinalità:

- $\text{card}(\text{ALLOGGIO}) = 10^5$ tuple,
Valori distinti di Regione = 20
- $\text{card}(\text{SERVIZIO}) = 10^2$ tuple,
Valori distinti di TipoServizio = 20
- $\text{card}(\text{ALLOGGIO-HA-SERVIZIO}) = 10^6$ tuple,
- $\text{card}(\text{UTENTE}) = 10^4$ tuples,
 $\text{MIN}(\text{DATE}(\text{DataNascita})) = 1/1/1930$,
 $\text{MAX}(\text{DATE}(\text{DataNascita})) = 31/12/2009$,
Valori distinti di Regione = 20,
Valori distinti di of BusinessAccount = 2 ("True",
"False")
- $\text{card}(\text{PRENOTAZIONE}) = 10^7$ tuple,
 $\text{MIN}(\text{DataInizio}) = 1/9/2017$, $\text{MAX}(\text{DataInizio}) =$
31/08/2020

Inoltre, assumere il seguente fattore di riduzione per la condizione di group by:

- Having COUNT(Distinct DataInizio) > 1 $\approx 1/10$

Si consideri la seguente query SQL:

```
select A.CodA, count(Distinct DataInizio)
from SERVIZIO S, ALLOGGIO-HA-SERVIZIO AHS, ALLOGGIO A,
PRENOTAZIONE P, UTENTE U
where S.CodS=AHS.CodS and A.CodA=AHS.CodA and
U.CodU=P.CodU and P.CodA=A.CodA
and (S.TipoServizio="Parcheggio" or
S.TipoServizio="Elettrodomestici")
and A.Regione="Liguria" and P.DataInizio>=1/5/20 and
P.DataInizio<=31/8/20
and U.Regione<>"Piemonte"
group by CodA
Having COUNT(Distinct DataInizio)>1
```

Homework tasks

Per l'interrogazione SQL

1. 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.
2. Si scelgano le strutture fisiche accessorie per migliorare le prestazioni dell'interrogazione. Si motivi la scelta e si definisca il piano di esecuzione (ordine e tipo dei join, accesso alle tabelle e/o indici, etc.).