

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

```

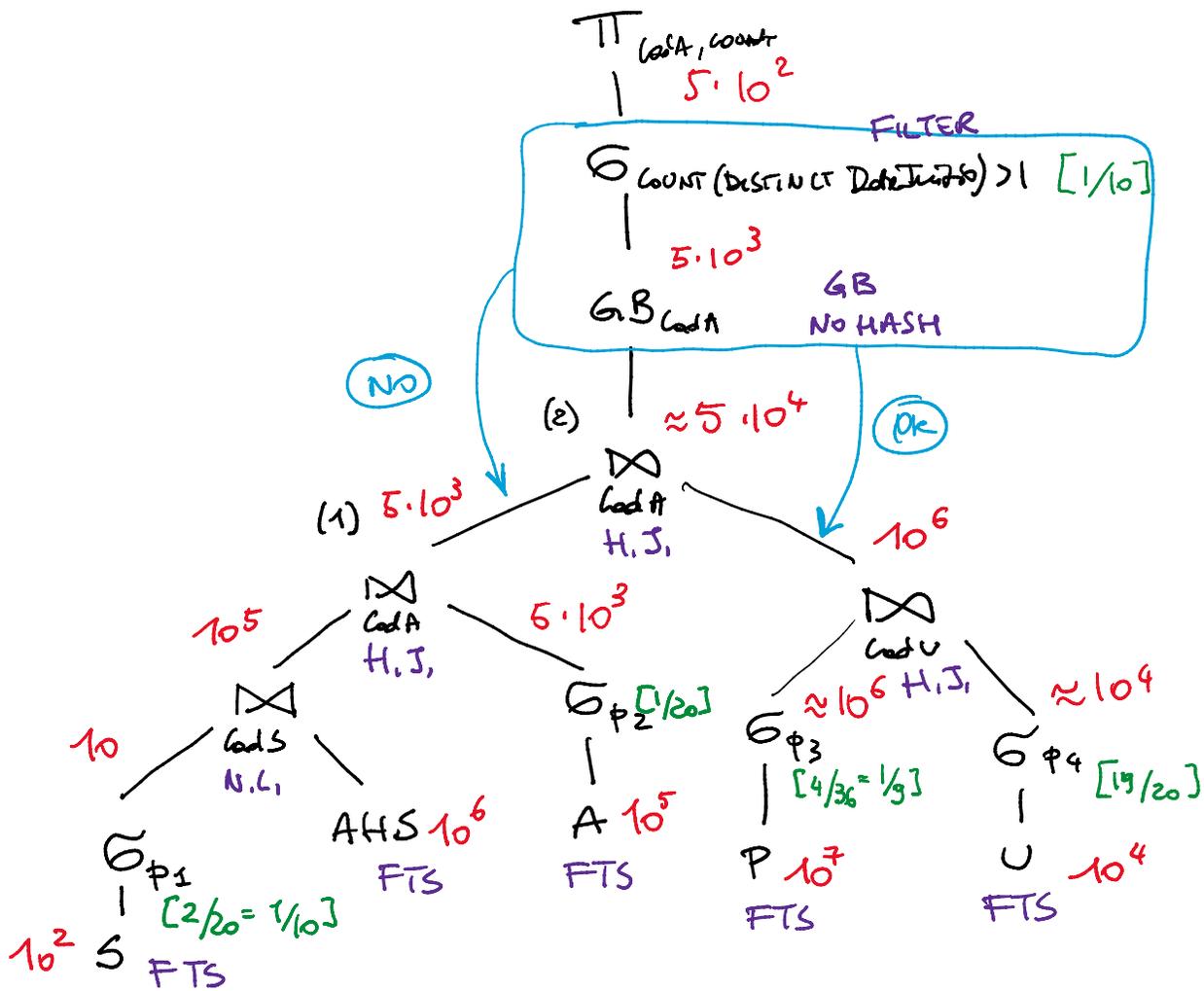
Homework tasks

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.).

Soluzione 1

$\Pi_{\text{CodA, count}}$



P_1 : TipoServizio = 'Per cheggio' OR
 TipoServizio = 'Elettrodomestici'

P_2 : Regione = 'Liguria'

P_3 : DataInizio \geq 1/5/2020 AND
 DataInizio \leq 31/8/2020

P_4 : Regione \neq 'Piemonte'

NO cond: $1/10 \cdot 10^6 = 10^5$
 CodS

(2) 10^5 Alloggi
 10^6 Servizi x Alloggi } \Rightarrow ogni alloggio ha
 ≈ 10 servizi

Riduzione 1/10 sui servizi

\hookrightarrow Rimane ≈ 1 servizio per alloggio

... .. 0 - 1 servizio per alloggio

riduzione $\times 10$...

↳ Rimane ≈ 1 servizio per alloggio

AHS dopo il join ha 1 servizio per alloggio

Per calcolare \bowtie si può ipotizzare lo stesso fattore di riduzione che si è applicato ad Alloggi.

⇒ Riduzione su AHS di $1/20$

(2) Nell'ipotesi di 1 servizio per alloggio si può applicare per il join \bowtie lo stesso fattore di riduzione che opera su alloggi ($1/20$).

Indici

Indice su ALLOGGIO (Regione)

Hash

⇒ indice coprente perché Alloggio non viene letto se riesco a utilizzare i rowid per avere i CodA

Soluzione 2

