

Data Science e Tecnologie per le Basi di Dati

Homework 4

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

```
TURISTA(CodFiscT, Nome, DataNascita, CartaCredito, Nazione)
VILLAGGIO_TURISTICO(CodVT, NomeSocietà, Nome, Città, #Stelle)
SERVIZI_DISPONIBILI(CodVT, NomeServizio)
PRENOTAZIONE_SOGGIORNO(CodFiscT, DataInizioSoggiorno, CodVT,
DataFineSoggiorno, Importo, #Adulti, #Bambini, Acconto)
```

Si ipotizzino le seguenti cardinalità:

- $\text{card}(\text{TURISTA}) = 10^5$ tuple,
 $\text{MIN}(\text{DataNascita}) = 1-1-1940$, $\text{MAX}(\text{DataNascita}) = 31-12-1999$,
valori distinti di Nazionalità $\simeq 100$,
- $\text{card}(\text{VILLAGGIO_TURISTICO}) = 10^4$ tuple,
valori distinti di Città $\simeq 100$,
- $\text{card}(\text{SERVIZI_DISPONIBILI}) = 10^4$ tuple,
valori distinti di NomeServizio $\simeq 10$,
- $\text{card}(\text{PRENOTAZIONE_SOGGIORNO}) = 10^9$ tuple,
 $\text{MIN}(\text{DataInizioSoggiorno}) = 01-01-2017$, $\text{MAX}(\text{DataInizioSoggiorno}) = 31-12-2017$,
 $\text{MIN}(\text{Importo}) = 1$, $\text{MAX}(\text{Importo}) = 10,000$

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

- $\text{having COUNT}(\ast) \geq 2 \simeq \frac{1}{10}$.

Si consideri la seguente query SQL:

```
select VT.CodVT, VT.Nome, SUM(Importo)
from VILLAGGIO_TURISTICO VT, PRENOTAZIONE_SOGGIORNO PS1
where VT.CodVT=PS1.CodVT and VT.Città<>'Santo Domingo'
      and PS1.CodFiscT IN (select PS.CodFiscT
                          from PRENOTAZIONE_SOGGIORNO PS, SERVIZI_DISPONIBILI SD
                          where PS.CodVT=SD.CodVT and PS.DataInizioSoggiorno ≥ 01/07/2017 and
                          (SD.NomeServizio='Wi-Fi' or SD.NomeServizio='Campo da tennis')
                          group by PS.CodFiscT
                          having COUNT(*) ≥ 2)
group by VT.CodVT, VT.Nome
```

Per l'interrogazione SQL:

- 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.
- 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.).