

Basi di Dati – Modulo 1

Prof. De Marsico

18 gennaio 2022

1) Sia data la seguente base di dati relativa ad una compagnia aerea

AEREO(ID, Costruttore, Modello, NPosti, AnnoC, AnnoR)
VOLO(Sigla, Partenza, Arrivo, Orario)
AEROPORTO(Sigla, Città)
VIAGGIO(ID, SiglaVolo, Aereo, Data)
PERSONALE(CE, ID, Nome, Cognome, DataNascita)
EQUIPAGGIO(Viaggio, Pers)

NOTE

- AEREO.AnnoC e AEREO.AnnoR sono rispettivamente l'anno di costruzione e l'anno di ultima revisione di un aereo (se l'aereo non è stato mai revisionato AnnoR=00/00/00)
- VOLO.Sigla è la sigla di un volo
- VOLO.Partenza e VOLO.Arrivo sono sigle di aeroporti
- Le città più grandi possono avere più aeroporti
- Un volo viene effettuato sempre alla stessa ora in più date
- VIAGGIO.SiglaVolo è la sigla di un volo effettuato in data VIAGGIO.Data
- VIAGGIO.Aereo è l'ID di un aereo
- EQUIPAGGIO.Viaggio è l'ID di un viaggio
- EQUIPAGGIO.Pers è l'ID di un membro del personale
- Un membro del personale può partecipare a più viaggi in date diverse.

1a) Trovare i modelli e numero posti di aerei revisionati prima del 2019 con cui sono stati effettuati voli in partenza da Milano nel 2021

1b) Trovare i modelli di aerei che non hanno mai volato verso New York partendo Napoli nel 2018.

2) Siano dati lo schema $R=ABCDEF$ e l'insieme di dipendenze funzionali

$F=\{AE \rightarrow BD, BE \rightarrow AC, D \rightarrow AC, B \rightarrow CF\}$

2a) Determinare le tre chiavi dello schema

2b) Dire se lo schema è 3NF e giustificare l'affermazione

2c) Trovare una decomposizione dello schema tale che:

- ogni sottoschema è 3NF
 - la decomposizione preserva le dipendenze
 - la decomposizione ha un join senza perdita.
-

3) E' dato un file di 145.500 record. Ogni record occupa 400 byte. Un blocco contiene 2048 byte. Un puntatore a blocco occupa 4 byte. Si utilizza una organizzazione hash con record distribuiti uniformemente tra 250 bucket.,

3a) Calcolare l'occupazione della bucket directory e dei bucket

3b) Calcolare il costo medio di una ricerca considerando chiavi di hash univoche

3c) Quanti bucket occorrerebbero per avere un costo medio di ricerca minore o uguale a 20 accessi?

1) Sia data la seguente base di dati relativa ad una compagnia aerea

AEREO(ID, Costruttore, Modello, NPosti, AnnoC, AnnoR)
 VOLO(Sigla, Partenza, Arrivo, Orario)
 AEROPORTO(Sigla, Città)
 VIAGGIO(ID, SiglaVolo, Aereo, Data)
 PERSONALE(CE, ID, Nome, Cognome, DataNascita)
 EQUIPAGGIO(Viaggio, Pers)

NOTE

- AEREO.AnnoC e AEREO.AnnoR sono rispettivamente l'anno di costruzione e l'anno di ultima revisione di un aereo (se l'aereo non è stato mai revisionato AnnoR=00/00/00)
- VOLO.Sigla è la sigla di un volo
- VOLO.Partenza e VOLO.Arrivo sono sigle di aeroporti
- Le città più grandi possono avere più aeroporti
- Un volo viene effettuato sempre alla stessa ora in più date
- VIAGGIO.SiglaVolo è la sigla di un volo effettuato in data VIAGGIO.Data
- VIAGGIO.Aereo è l'ID di un aereo
- EQUIPAGGIO.Viaggio è l'ID di un viaggio
- EQUIPAGGIO.Pers è l'ID di un membro del personale
- Un membro del personale può partecipare a più viaggi in date diverse.

1a) Trovare i modelli e numero posti di aerei revisionati prima del 2019 con cui sono stati effettuati voli in partenza da Milano nel 2021

1b) Trovare i modelli di aerei che non hanno mai volato verso New York partendo Napoli nel 2018.

1A)

$$A_{mod} = \sigma_{\substack{AnnoC < 2019 \\ AnnoR > 2019}}(A_{aereo}) \quad A_{aeroportoM_1} = \sigma_{\substack{Città = \\ "Milano"}}(A_{aeroporto})$$

$$V_{daM_1} = \pi_{A_{aereo}} \left(\sigma_{\substack{AnnoC < 2019 \\ AnnoR > 2019}} \left(\left(\left(V_{sigla} \bowtie V_{viaggio} \right) \bowtie A_{aeroportoM_1} \right) \right) \right)$$

$$OUT = A_{mod} \bowtie_{ID = A_{aereo}} V_{daM_1}$$

1B) $TotAerei = \left(A_{aereo} \bowtie_{\substack{ID = A_{aereo} \\ SiglaVolo \\ Sigla}} V_{viaggio} \bowtie_{\substack{SiglaVolo \\ Sigla}} V_{da} \right)$

$$V_{daNY} = \left(\sigma_{\substack{AnnoC < 2018 \\ Data < \\ AnnoR > 2018 \\ Città = "Napoli"}} \left(\left(\left(V_{da} \bowtie V_{viaggio} \right) \bowtie A_{aeroporto} \right) \right) \right)$$

$$V_{daNY} = \left(\sigma_{\substack{AnnoC < 2018 \\ Data < \\ AnnoR > 2018 \\ Città = "Napoli"}} \left(\left(\left(V_{da} \bowtie V_{viaggio} \right) \bowtie A_{aeroporto} \right) \right) \right)$$

$$DA_{NA} \wedge NY = V_{daNY} \cap V_{daNY}$$

$$OUT = \pi_{\substack{Modello}} (TotAerei - DA_{NA} \wedge NY)$$

2) Siano dati lo schema $R=ABCDEF$ e l'insieme di dipendenze funzionali

$F=\{AE \rightarrow BD, BE \rightarrow AC, D \rightarrow AC, B \rightarrow CF\}$

2a) Determinare le tre chiavi dello schema

2b) Dire se lo schema è 3NF e giustificare l'affermazione

2c) Trovare una decomposizione dello schema tale che:

- ogni sottoschema è 3NF
- la decomposizione preserva le dipendenze
- la decomposizione ha un join senza perdita.

2a)

Perché "E" mai a dx sana parte delle chiavi

Primo

$$- (AE)^+_F = AEBDCF \quad \checkmark \Rightarrow AE \text{ chiave}$$

$$- (BE)^+_F = BEACDF \quad \checkmark \Rightarrow BE \text{ chiave}$$

$$- (DE)^+_F = DEACBF \quad \checkmark \Rightarrow DE \text{ chiave}$$

2b)

Non in 3NF per $B \rightarrow CF$, B non superchiave e C non primo

2c)

I°

$$F = \{AE \rightarrow B, AE \rightarrow D, BE \rightarrow A, BE \rightarrow C, D \rightarrow A, D \rightarrow C, B \rightarrow C, B \rightarrow F\}$$

II° Capotero Minimo

$$- AE \rightarrow B$$

$$(A)^+_F = A \not\rightarrow B$$

$$(E)^+_F = E \not\rightarrow B$$

$$- AE \rightarrow D \text{ non serve per che già causa le chiusure}$$

$$- BE \rightarrow A$$

$$(B)^+_F = BCF \not\rightarrow A$$

$$(E)^+_F = E \not\rightarrow A$$

$$- BE \rightarrow C$$

$$(B)^+_F = BCF \supset C \Rightarrow BF \rightarrow C \text{ ma già lo ha } \Rightarrow \text{Idgo}$$

Alla fine 2° passo $F = \{AE \rightarrow B, AE \rightarrow D, BE \rightarrow A, D \rightarrow A, D \rightarrow C, B \rightarrow C, B \rightarrow F\}$

III° passo CEDC RIDUZIONE

- $AE \rightarrow B$

$$(AE)^+_{F-AE \rightarrow B} = AEDC \quad \& B \Rightarrow \text{NEW TAGO}$$

- $AE \rightarrow D$

$$(AE)^+_{F-AE \rightarrow D} = AEB CF \quad \& D \Rightarrow \text{NEW TAGO}$$

- $BE \rightarrow A$

$$(BE)^+_{F-BE \rightarrow A} = BECF \quad \& A \Rightarrow \text{NEW TAGO}$$

- $D \rightarrow A$

$$(D)^+_{F-D \rightarrow A} = DC \quad \& A \Rightarrow \text{NEW TAGO}$$

- $D \rightarrow C$

$$(D)^+_{F-D \rightarrow C} = DA \quad \& C \Rightarrow \text{NEW TAGO}$$

- $B \rightarrow C$

$$(B)^+_{F-B \rightarrow C} = BF \quad \& C \Rightarrow \text{NEW TAGO}$$

- $B \rightarrow F$

$$(B)^+_{F-B \rightarrow F} = BC \quad \& F \Rightarrow \text{NEW TAGO}$$

Fino III° passo $F = \{AE \rightarrow B, AE \rightarrow D, BE \rightarrow A, D \rightarrow A, D \rightarrow C, B \rightarrow C, B \rightarrow F\}$

$\rho = \{AEB, AED, BEA, DA, DC, BC, BF\}$ E' in 3NF poiche' contiene

tutti gli elementi di R, HA JON SENZA PERDITA poiche' contiene chiave AE

CONTIENE AE
✓

3) E' dato un file di 145.500 record. Ogni record occupa 400 byte. Un blocco contiene 2048 byte. Un puntatore a blocco occupa 4 byte. Si utilizza una organizzazione hash con record distribuiti uniformemente tra 250 bucket.,

3a) Calcolare l'occupazione della bucket directory e dei bucket

3b) Calcolare il costo medio di una ricerca considerando chiavi di hash univoche

3c) Quanti bucket occorrerebbero per avere un costo medio di ricerca minore o uguale a 20 accessi?

$$2a) \quad N^{\circ} \text{Blocchi} \times \text{Bucket Dir} = \frac{\lceil N^{\circ} \text{Bucket} \rceil}{\text{Record} \times \text{Block Dir}} = \frac{\lceil \frac{250}{512} \rceil}{1} = 1$$

$$\text{Record} \times \text{Block Dir} = \left\lfloor \frac{\text{Block size}}{P_{\text{size}}} \right\rfloor = \frac{2048}{4} = 512$$

$$\text{Blocchi} \times \text{Bucket} = \frac{\lceil \text{Record} \times \text{Bucket} \rceil}{\text{Record} \times \text{Block}} = \frac{\lceil 582 \rceil}{5} = 117$$

$$\text{Record} \times \text{Block} = \left\lfloor \frac{\text{Block size} - P_{\text{size}}}{\text{Record size}} \right\rfloor = \left\lfloor \frac{2048 - 4}{400} \right\rfloor = 5$$

$$\text{Record} \times \text{Bucket} = \frac{\lceil N^{\circ} \text{Record} \rceil}{N^{\circ} \text{Bucket}} = \frac{\lceil \frac{145500}{250} \rceil}{1} = 582$$

$$3b) \quad \text{Avg time} = \frac{\lceil \text{Blocchi} \times \text{Bucket} \rceil}{2} = \frac{\lceil 117 \rceil}{2} = 59 \text{ accessi}$$

$$3c) \quad \frac{\text{Blocchi} \times \text{Bucket}}{2} \leq 20$$

$$\text{Blocchi} \times \text{Bucket} \leq 40$$

$$\frac{\text{Record} \times \text{Bucket}}{\text{Record} \times \text{Block}} \leq 40$$

near per
univ

$$\text{Record} \times \text{Bucket} \leq 40 \cdot 5$$

$$\text{Record} \times \text{Bucket} \leq 200$$

$$\frac{N^{\circ} \text{Record}}{N^{\circ} \text{Bucket}} \leq 200$$

$$N^{\circ} \text{record} \leq 200 \cdot N^{\circ} \text{Bucket}$$

$$N^{\circ} \text{Bucket} \geq \frac{N^{\circ} \text{Record}}{200} \geq 728$$