

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 di aerei mai revisionati con cui sono stati effettuati voli diretti a New York nel 2021

1b) Trovare nome e cognome di membri del personale che non hanno mai volato verso Milano partendo da Roma.

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

$F=\{AB \rightarrow CD, BC \rightarrow AE, D \rightarrow AC, A \rightarrow EF\}$

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 134.700 record. Ogni record occupa 357 byte, di cui 57 per la chiave. Un blocco contiene 2048 byte. Un puntatore a blocco occupa 4 byte. Si utilizza una organizzazione hash con record distribuiti uniformemente tra 200 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 di aerei mai revisionati con cui sono stati effettuati voli diretti a New York nel 2021

1b) Trovare nome e cognome di membri del personale che non hanno mai volato verso Milano partendo da Roma

1A)

$$V_{di \text{ New York } 2021} = \left(\begin{array}{c} \sigma \\ \text{Data} \\ \text{Arrivo} \\ \text{"New York"} \end{array} \left(\begin{array}{c} V_{do} \bowtie V_{viaggio} \\ \text{Volo. Sigla} \\ \text{Via Geo. Sigla Vola} \end{array} \right) \right)$$

$$OUT = \pi_{Modello} \left(\begin{array}{c} \sigma_{AnnoR} \\ \text{"00/00/00"} \end{array} \left(\begin{array}{c} Aereo \bowtie V_{di \text{ New York } 2021} \\ \text{Aereo. id} \\ \text{V}_{di \text{ New York } 2021}. Aereo \end{array} \right) \right)$$

1B)

$$V_{\text{viaggio da Roma}} = \left(\begin{array}{c} \sigma_{Aeroporto} \\ \text{Città} \\ \text{"Roma"} \end{array} \left(\begin{array}{c} V_{do} \bowtie V_{viaggio} \\ \text{Volo. Sigla} \\ \text{Via Geo. Sigla Vola} \end{array} \right) \bowtie \begin{array}{c} Aereo \text{ parte} \\ \text{VIAGGIO. PARTENZA} \\ \text{AEROPORTO. Sigla} \end{array} \right)$$

$$V_{\text{viaggio per Milano}} = \left(\begin{array}{c} \sigma_{Aeroporto} \\ \text{Città} \\ \text{"Milano"} \end{array} \left(\begin{array}{c} V_{do} \bowtie V_{viaggio} \\ \text{Volo. Sigla} \\ \text{Via Geo. Sigla Vola} \end{array} \right) \bowtie \begin{array}{c} Aereo \text{ parte} \\ \text{VIAGGIO. ARRIVO} \\ \text{AEROPORTO. Sigla} \end{array} \right)$$

$$V_{\text{viaggio RM}} = V_{\text{viaggio da Roma}} \bowtie V_{\text{viaggio per Milano}}$$

$\text{Viaggio da Roma. ID}$
 $\text{Viaggio per Milano. ID}$

Passare da $V_{\text{viaggio RM}}$

$$V_{\text{viaggio RM}} \bowtie \text{Equipaggio} \bowtie \text{Personale}$$

$\text{ID} = \text{viaggio}$
 $\text{Pers} = \text{ID}$

$$\omega T = \pi_{\substack{\text{non} \\ \text{cage}}} \quad (\text{Peso} - \text{Peso} \text{ del Uroglio DM})$$

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

$F=\{AB \rightarrow CD, BC \rightarrow AE, D \rightarrow AC, A \rightarrow EF\}$

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)

$$(AB)_F^+ = AB \ CD \ EF \ \checkmark$$

$$(BC)_F^+ = BC \ AE \ FD \ \checkmark$$

$$(BD)_F^+ = BD \ AC \ EF \ \checkmark$$

2b)

No per $A \rightarrow EF$ A non super, EF non primo

2c)

$$F = \{ AB \rightarrow C, AB \rightarrow D, BC \rightarrow A, BC \rightarrow E, D \rightarrow A, D \rightarrow C, A \rightarrow E, A \rightarrow F \}$$

\mathbb{I}^0

• $AB \rightarrow C$

$$(A)_F^+ = AEF \ \not\subset C \Rightarrow \text{rimuove}$$

$$(B)_F^+ = B \ \not\subset C \Rightarrow \text{rimuove}$$

• $AB \rightarrow D$

$$(A)_F^+ = AEF \ \not\subset D$$

$$(B)_F^+ = B \ \not\subset D$$

• $BC \rightarrow A$

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

$$(C)_F^+ = C \ \not\subset A$$

• $BC \rightarrow E$

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

$$(C)_F^+ = C \ \not\subset E$$

Fine 2° Passo $F_2 = \{ AB \rightarrow C, AB \rightarrow D, BC \rightarrow A, BC \rightarrow E, D \rightarrow A, D \rightarrow C, A \rightarrow E, A \rightarrow F \}$

III°

- $AB \rightarrow C$

$$(AB)^+_F = ABDC \quad \Rightarrow \text{Tot Go}$$

$$F_2 = \{ AB \rightarrow D, BC \rightarrow A, BC \rightarrow E, D \rightarrow A, D \rightarrow C, A \rightarrow E, A \rightarrow F \}$$

- $AB \rightarrow D$

$$(AB)^+_F = ABED \quad \text{Now Tot Go}$$

- $BC \rightarrow A$

$$(BC)^+_F = BCE \quad \text{Now Tot Go}$$

- $D \rightarrow A$

$$(D)^+_F = DC \quad \text{Now Tot Go}$$

- $D \rightarrow C$

$$(D)^+_F = DAEF \quad \text{Now Tot Go}$$

- $A \rightarrow E$

$$(A)^+_F = AF \quad \text{Now Tot Go}$$

- $A \rightarrow F$

$$(A)^+_F = AE \quad \text{Now Tot Go}$$

Alla fine del 3° passo

$$F_2 = \{ AB \rightarrow D, BC \rightarrow A, BC \rightarrow E, D \rightarrow A, D \rightarrow C, A \rightarrow E, A \rightarrow F \}$$

↓

$$\rho = \{ ABCD, BCDA, BCED, DA, DC, AE, AF \} \quad \text{E' IN UN POCKET}$$

HA TUTTI I SEGMENTI DI ρ E LA SEGA SENZA ADESSA POCA E' HA CHIAVE = AB IN ABD

3) E' dato un file di 134.700 record. Ogni record occupa 357 byte, di cui 57 per la chiave. Un blocco contiene 2048 byte. Un puntatore a blocco occupa 4 byte. Si utilizza una organizzazione hash con record distribuiti uniformemente tra 200 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?

$$3a) \text{Block} + \text{Bucket Dir} = \frac{\lceil \frac{\text{N}^\circ \text{Bucket} \rceil}{\text{Record} + \text{Block dir}}} = \frac{\lceil \frac{200}{512} \rceil}{1} = 1$$

$$\text{Record} + \text{Block dir} = \left\lfloor \frac{\text{Block Size}}{\text{P size}} \right\rfloor = \frac{2048}{4} = 512$$

$$\text{Block} + \text{Bucket} = \frac{\lceil \text{Record} + \text{Bucket} \rceil}{\text{Record} + \text{Block}} = \frac{\lceil 679 \rceil}{5} = 135$$

$$\text{Record} + \text{Bucket} = \frac{\lceil \frac{\text{N}^\circ \text{Record}}{\text{N}^\circ \text{Bucket}} \rceil}{1} = \frac{\lceil \frac{134700}{200} \rceil}{1} = 679$$

$$\text{Record} + \text{Block} = \left\lfloor \frac{\text{Block size} - \text{P size}}{\text{R size}} \right\rfloor = \frac{2048 - 4}{357} = 5$$

$$3b) \text{Avg Time} = \frac{\lceil \text{Block} + \text{Bucket} \rceil}{2} = \frac{\lceil 135 \rceil}{2} = 68 \text{ Accessi}$$

$$3c) \frac{\text{Block} + \text{Bucket}}{2} < 20 \quad \text{Block} + \text{Bucket} < 40$$

$$\frac{\text{Record} + \text{Bucket}}{\text{Record} + \text{Block}} < 40 \quad \text{Record} + \text{Bucket} < 200$$

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

$$\text{N}^\circ \text{Record} < 200 \cdot \text{N}^\circ \text{Bucket}$$

$$134\,700 < 200 \cdot \text{N}^\circ \text{Bucket}$$

$$\text{N}^\circ \text{Bucket} > \frac{134\,700}{200}$$

$$\text{N}^\circ \text{Bucket} > 674$$