

Basi di Dati – Modulo 1

23 gennaio 2023 - Traccia A

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(CF, 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=0000)
- 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 che hanno avuto l'ultima revisione meno di 5 anni dopo l'anno di costruzione e con cui sono stati effettuati voli diretti a Berlino nel 2022

1b) Trovare nome e cognome e Codice Fiscale di membri del personale che non hanno mai volato con partenze diverse 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) Dire se la decomposizione $\rho=\{ACD, BCE, AEF\}$ preserva le dipendenze e descrivere il procedimento utilizzato giustificando i passaggi

3) E' dato un file di 134.700 record. Ogni record occupa 357 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 in blocchi 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 15 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=0000)
- 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 che hanno avuto l'ultima revisione meno di 5 anni dopo l'anno di costruzione e con cui sono stati effettuati voli diretti a Berlino nel 2022

1b) Trovare nome e cognome e Codice Fiscale di membri del personale che non hanno mai volato con partenze diverse da Roma.

1A)

$$\text{Anno Rev min 5} = \sigma_{\text{Aereo}} \left(\text{AnnoR} \leq \text{AnnoC} + 5 \right)$$

$$\text{Voli Berlino 2022} = \pi_{\text{AEREO}} \left(\sigma_{\text{data} = 31/12/2022} \left(\left(\text{VIAGGIO} \bowtie \text{Volo} \right) \bowtie \text{Aeroporto} \right) \right)$$

$\text{SIGLA volo} = \text{SIGLA}$ $\text{ARRIVO} = \text{SIGLA}$
 $\text{CITTÀ} = \text{"Berlino"}$

$$\text{OUT} = \pi_{\text{Modello}} \left(\text{Voli Berlino 2022} \bowtie \text{Anno Rev min 5} \right)$$

$\text{AEREO ID} =$

1B)

$$\text{PARTENZA NcRoma} = \pi_{\text{VIAGGIO ID}} \left(\sigma_{\text{CITTÀ} \neq \text{ROMA}} \left(\left(\text{VIAGGIO} \bowtie \text{Volo} \right) \bowtie \text{Aeroporto} \right) \right)$$

$\text{SIGLA volo} = \text{SIGLA}$ $\text{ARRIVO} = \text{SIGLA}$

$$\text{Equip NcRoma} = \pi_{\text{Nome, Cognome, CF}} \left(\text{EQUIPAGGIO} \bowtie \text{PARTENZA NcRoma} \right) \bowtie \text{Personale}$$

$\text{VIAGGIO ID} =$ $\text{PERSON ID} =$

$$\text{Tot Equip} = \pi_{\text{Nome, Cognome, CF}} \left(\text{EQUIPAGGIO} \bowtie \text{Personale} \right)$$

$\text{PERSON ID} =$

$$\text{OUT} = \text{Tot Equip} - \text{Equip NcRoma}$$

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) Dire se la decomposizione $\rho=\{ACD, BCE, AEF\}$ preserva le dipendenze e descrivere il procedimento utilizzato giustificando i passaggi

2a)

$$(AB)^+_F = ABCDEF \Rightarrow \text{CHIAVE}$$

$$(BC)^+_F = BC AEDF \Rightarrow \text{CHIAVE}$$

$$(BD)^+_F = BD AC EF \Rightarrow \text{CHIAVE}$$

2b) Non in 3NF per $A \rightarrow EF$

A non superchiave EF non primo

2c)

$$\rho = (ACD, BCE, AEF)$$

$AB \rightarrow CD$ problematica

$BC \rightarrow EA$ //

Parto da $AB \rightarrow CD$ usando l'algoritmo

$$Z_0 = AB$$

$$S_0 = (AB \cap ACD)^+_F \cap ACD \cup (AB \cap BCE)^+_F \cap BCE \cup (AB \cap AEF)^+_F \cap AEF =$$

$$= (A)^+_F \cap ACD \cup (B)^+_F \cap BCE \cup (A)^+_F \cap AEF =$$

$$= AEF \cap ACD \cup B \cap BCE \cup AEF \cap AEF =$$

$$A \cup B \cup AEF = AB \cup EF$$

$$Z_1 = AB \cup EF$$

$$\begin{aligned}
S_1 &= (ABEF \cap ACB)^+_F \cap ACB \cup (ABEF \cap BCE)^+_F \cap BCE \cup (ABEF \cap AEF)^+_F \cap AEF \\
&= (A)^+_F \cap ACB \cup (BE)^+_F \cap BCE \cup (AEF)^+_F \cap AEF \\
&= AEF \cap ACB \cup BE \cap BCE \cup AEF \cap AEF \\
&= A \cup BE \cup AEF = ABEF
\end{aligned}$$

$S_1 \subset \mathcal{L}_1 \Rightarrow M_1$ FERMO, $ABEF$ NON CONTIENE $CD \Rightarrow \rho$ NON PRESERVA

3) E' dato un file di 134.700 record. Ogni record occupa 357 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 in blocchi 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 15 accessi?

2a)

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

$$\text{Blocchi} \times \text{Bucket} = \frac{\lceil \text{Record} \times \text{Bucket} \rceil}{\text{Record} \times \text{Block dir}} = \frac{539}{5} = 108$$

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

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

$$\text{Record} \times \text{Bucket} = \frac{\lceil N^{\circ} \text{Record} \rceil}{N^{\circ} \text{Bucket}} = \frac{\lceil \frac{134700}{250} \rceil}{1} = 539$$

2b)

$$\text{Avg Time Search} = \frac{\lceil \text{Blocchi} \times \text{Bucket} \rceil}{2} = \frac{108}{2} = 54$$

2c)

$$\frac{\text{Blocchi} \times \text{Bucket}}{2} \leq 15$$

$$\text{Blocchi} \times \text{Bucket} \leq 30$$

$$\frac{\text{Record} \times \text{Bucket}}{\text{Record} \times \text{Block dir}} \leq 30$$

$$\text{Record} \times \text{Bucket} \leq 150$$

$$\frac{134700}{N^{\text{Bucket}}} \leq 150$$

$$134700 \leq 150 \cdot N^{\text{Bucket}}$$

$$N^{\text{Bucket}} \geq \frac{134700}{150}$$

$$N^{\text{Bucket}} \geq 898$$