

Basi di Dati – Modulo 1

24 Ottobre 2022

1) Dato il seguente schema di una **base di dati contenente dati relativi a magazzini**

PRODOTTI(Codice,Descrizione,Prezzo_unit)

MAGAZZINI(Codice,Indirizzo, Città)

SCORTE(Codice-P,Codice-M,N-pezzi)

NOTE:

SCORTE contiene i dati relativi alla presenza (**N-pezzi** ≥ 1) dei prodotti nei vari magazzini

esprimere in **algebra relazionale** le seguenti interrogazioni:

1a) Per ogni prodotto che costa più di 100 euro e del quale sono presenti meno di 50 pezzi in almeno un magazzino, si desidera conoscere: i dati del prodotto e l'indirizzo dei magazzini in cui sono presenti meno di 50 pezzi del prodotto.

1b) Per ogni prodotto trovare il magazzino di Roma in cui abbiamo le scorte maggiori (restituire Codice prodotto, Descrizione, Codice magazzino, Indirizzo, N-pezzi)

2) Dato il seguente schema $R = ABCDEH$ sul quale è definito il seguente insieme di dipendenze funzionali:

$F = \{A \rightarrow C, AB \rightarrow D, ABH \rightarrow CE, BH \rightarrow AD, D \rightarrow C, DE \rightarrow H, E \rightarrow B\}$

2a) trovare le chiavi dello schema motivando la risposta*

2b) dire se lo schema è in 3NF motivando la risposta*

2c) trovare una decomposizione dello schema*, in modo tale che ogni sottoschema sia in 3NF, e che la decomposizione preservi F e abbia un join senza perdita.

(* scrivere le definizioni formali degli elementi teorici a cui si fa riferimento e lo pseudo codice degli algoritmi utilizzati)

3) E' dato un file di 168.600 record. Ogni record occupa 426 byte, di cui 32 per la chiave. Un blocco contiene 4096 byte. Un puntatore a blocco occupa 5 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) Dato il seguente schema di una base di dati contenente dati relativi a magazzini

PRODOTTI(Codice, Descrizione, Prezzo_unit)

MAGAZZINI(Codice, Indirizzo, Città)

SCORTE(Codice-P, Codice-M, N-pezzi)

NOTE:

SCORTE contiene i dati relativi alla presenza (N-pezzi >= 1) dei prodotti nei vari magazzini

esprimere in algebra relazionale le seguenti interrogazioni:

1a) Per ogni prodotto che costa più di 100 euro e del quale sono presenti meno di 50 pezzi in almeno un magazzino, si desidera conoscere: i dati del prodotto e l'indirizzo dei magazzini in cui sono presenti meno di 50 pezzi del prodotto.

1b) Per ogni prodotto trovare il magazzino di Roma in cui abbiamo le scorte maggiori (restituire Codice prodotto, Descrizione, Codice magazzino, Indirizzo, N-pezzi)

$$1a) Costori = \left(\sigma_{\substack{\text{Prezzo-unit} \\ > 100}} (Prodotti) \right)$$

$$OUT = \pi_{\substack{\text{Codice P,} \\ \text{Descrizione,} \\ \text{Magazzino,} \\ \text{Indirizzo,} \\ \text{Città}}} \left(\sigma_{\substack{\text{N-Pezzi} \\ < 50}} \left(Costori \bowtie_{\substack{\text{Codice P} = \text{Codice P} \\ \text{Codice M} = \text{Codice M}}} Scorte \bowtie_{\text{Codice M}} Magazzini \right) \right)$$

$$1b) MAGAZZINI R = \sigma_{\substack{\text{Città} \\ = \text{ROMA}}} (MAGAZZINI)$$

$$SC = Scorte \rho_{\substack{\text{Codice P} = CP, \text{ Codice M} = CM, \text{ N-Pezzi} = CN}}$$

$$MAG SC = MAGAZZINI R \bowtie_{\text{Codice M} = CM} SC$$

$$MAG S = MAGAZZINI R \bowtie_{\text{Codice M} = \text{Codice M}} Scorte$$

$$MAG SC \bowtie_{\substack{\text{N-Pezzi} \\ < 50}} P = \left(\sigma_{\substack{\text{N-Pezzi} \\ < 50}} (MAG S \bowtie_{\text{Codice P} = CP} P) \right)$$

$$Scorte T = (MAG S - MAG SC \bowtie_{\text{Codice P} = CP} P)$$

$$OUT = \pi_{\substack{\text{Codice P,} \\ \text{Descrizione,} \\ \text{Città}}} \left(Prodotti \bowtie_{\text{Codice P} = \text{Codice P}} Scorte T \right)$$

2) Dato il seguente schema $R = ABCDEH$ sul quale è definito il seguente insieme di dipendenze funzionali:

$F = \{A \rightarrow C, AB \rightarrow D, ABH \rightarrow CE, BH \rightarrow AD, D \rightarrow C, DE \rightarrow H, E \rightarrow B\}$

2a) trovare le chiavi dello schema motivando la risposta*

2b) dire se lo schema è in 3NF motivando la risposta*

2c) trovare una decomposizione dello schema*, in modo tale che ogni sottoschema sia in 3NF, e che la decomposizione preservi F e abbia un join senza perdita.

(* scrivere le definizioni formali degli elementi teorici a cui si fa riferimento e lo pseudo codice degli algoritmi utilizzati)

21)

$$(BH)_F^+ = BH ADC E \quad \checkmark$$

$$(DE)_F^+ = DEH BAC \quad \checkmark$$

$$(EH)_F^+ = EBH ADC \quad \checkmark$$

$$(EA)_F^+ = EBACDH \quad \checkmark$$

22) Now in 3NF per $A \rightarrow C$

$$2c) I^a \quad F = \{A \rightarrow C, AB \rightarrow D, ABH \rightarrow C, ABH \rightarrow E, BH \rightarrow A, BH \rightarrow D, D \rightarrow C, DE \rightarrow H, E \rightarrow B\}$$

II

- $AB \rightarrow D$

$$(A)_F^+ = AC \quad \checkmark \quad D$$

$$(B)_F^+ = B \quad \checkmark \quad D$$

- $ABH \rightarrow C$

$$(A)_F^+ = AC \supset C \Rightarrow \quad \cancel{ABH \rightarrow C} \quad \text{not redundant} \Rightarrow \text{ok}$$

- $ABH \rightarrow E$

$$(A)_F^+ = AC \quad \checkmark \quad E$$

$$(B)_F^+ = B \quad \checkmark \quad E$$

$$(H)_F^+ = H \quad \checkmark \quad E$$

- $BH \rightarrow A$

$$(B)_F^+ = B \quad \checkmark \quad A$$

$$(H)_F^+ = H \quad \checkmark \quad A$$

- $BH \rightarrow D$

$$(B)_F^+ = B \quad \checkmark \quad D$$

$$(H)_F^+ = H \quad \checkmark \quad D$$

- $DE \rightarrow H$

$$(D)_F^+ = D \quad \checkmark \quad H$$

$$(E)_F^+ = EB \quad \checkmark \quad H$$

$$F = \{A \rightarrow C, AB \rightarrow D, \cancel{ABH \rightarrow E}, BH \rightarrow A, \cancel{BA \rightarrow D}, D \rightarrow X, DE \rightarrow H, E \rightarrow B\}$$

III^a

$$- A \rightarrow C$$

$$(A)^+_{F-A \rightarrow C} = A \cup C$$

$$- AB \rightarrow D$$

$$(AB)^+_{F-AB \rightarrow D} = ABC \cup D$$

$$- ABH \rightarrow E$$

$$(ABH)^+_{F-ABH \rightarrow E} = ABH \cup AD \cup C \cup E$$

$$- BA \rightarrow A$$

$$(BA)^+_{F-BA \rightarrow A} = BA \cup X \cup E \cup A$$

$$- BH \rightarrow D$$

$$(BH)^+_{F-BH \rightarrow D} = BH \cup A \cup C \cup D \cup X \cup E$$

$$F = \{A \rightarrow C, AB \rightarrow D, \cancel{ABH \rightarrow E}, BH \rightarrow A, \cancel{BA \rightarrow D}, D \rightarrow X, DE \rightarrow H, E \rightarrow B\}$$

gl: alt: nur \neg & \cup

$$F = \{A \rightarrow C, AB \rightarrow D, BH \rightarrow E, BH \rightarrow A, D \rightarrow X, DE \rightarrow H, E \rightarrow B\}$$

$$\mathcal{S} = \{A, AB, BH, BH, D, D, DE, E, B\}$$

3) E' dato un file di 168.600 record. Ogni record occupa 426 byte, di cui 32 per la chiave. Un blocco contiene 4096 byte. Un puntatore a blocco occupa 5 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?

$$(3a) \text{ Bucket} \times \text{Bc dir} = \frac{\sqrt{N^{\circ} \text{ Bucket}}}{\text{Pointer} \text{ Block}} = \frac{\sqrt{250}}{819} = 2$$

$$\text{Pointer} \times \text{Block} = \frac{\text{Blocksize}}{\sqrt{Psize}} = 819$$

$$\text{Bucket} \times \text{Block} = \frac{\sqrt{\text{Record} \times \text{Block}}}{\text{Record} \times \text{Block}} = \frac{\sqrt{675}}{9} = 25$$

$$\text{Record} \times \text{Block} = \frac{\sqrt{N^{\circ} \text{ Record}}}{N^{\circ} \text{ Bucket}} = 675$$

$$\text{Record} \times \text{Block} = \left\lfloor \frac{\text{Blocksize} - \text{Psize}}{Rsize} \right\rfloor = 9$$

$$(3b) \text{ Avg Access} = \frac{\sqrt{25}}{2} = 25$$

$$(3c) \text{ Avg Access} \leq 20$$

$$\frac{\text{Blocksize}}{2} \leq 20$$

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

$$\text{Record} \times \text{Block} \leq 360$$

$$\frac{N^{\circ} \text{ Record}}{N^{\circ} \text{ Bucket}} \leq 360$$

$$N^{\circ} \text{ Bucket} \geq \frac{N^{\circ} \text{ Record}}{360} \geq 469$$