

Basi di Dati – Canali unificati

04 settembre 2023

Dato il seguente schema di una **base di dati contenente dati dei progetti di un'azienda**

DIPENDENTE(Id, SettId, Nome, Cognome, DataN)

PROGETTO(Id, SettId, Titolo, DataI)

TEAM(ProgId, DipenId, Ruolo)

SETTORE(Id, CapoId, Nome)

NOTE:

- Gli attributi o insiemi di attributi sottolineati costituiscono le chiavi delle relazioni
- DataN e DataI in DIPENDENTE e PROGETTO sono, rispettivamente, la data di nascita di un dipendente e la data di inizio di un progetto
- In DIPENDENTE, PROGETTO e SETTORE l'attributo Id contiene il codice identificativo di, rispettivamente, un dipendente, un progetto e un settore
- In TEAM, che contiene i dati dei team che stanno lavorando sui progetti dell'azienda, il dipendente il cui codice identificativo è DipenId lavora sul progetto il cui identificativo è ProgId
- In TEAM, il dominio dell'attributo Ruolo è {"ProjectLeader", "Designer", "Programmer"}
- In SETTORE, l'attributo CapoId contiene l'Id del dipendente che è a capo di un settore e Nome è il nome del settore
- In PROGETTO, l'attributo SettId contiene il codice identificativo del settore al quale appartiene il progetto (ogni progetto appartiene ad un solo settore ma un settore può contenere più progetti)

- 1) Elenco (nome, cognome e data di nascita) dei dipendenti che hanno lavorato su progetti più vecchi di 3 anni (iniziati prima dell'1/9/2020) nel il ruolo di "Desogner" oppure di "Programmer"
- 2) Titolo del progetto o dei progetti più vecchi (meno recenti) di cui ogni dipendente è stato Project Leader (nome e cognome del Project Leader e titolo del progetto).

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

$F = \{BCD \rightarrow AH, BD \rightarrow CH, A \rightarrow D, C \rightarrow E, CD \rightarrow H, H \rightarrow E, AH \rightarrow B\}$

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 145.950 record. Ogni record occupa 255 byte. Un blocco contiene 2048 byte. Un puntatore a blocco occupa 5 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 10 accessi?

DIPENDENTE(Id, SettlId, Nome, Cognome, DataN)
 PROGETTO(Id, SettlId, Titolo, DataI)
 TEAM(Progid, DipenId, Ruolo)
 SETTORE(Id, Capold, Nome)

NOTE:

- Gli attributi o insiemi di attributi sottolineati costituiscono le chiavi delle relazioni
- DataN e DataI in DIPENDENTE e PROGETTO sono, rispettivamente, la data di nascita di un dipendente e la data di inizio di un progetto
- In DIPENDENTE, PROGETTO e SETTORE l'attributo Id contiene il codice identificativo di, rispettivamente, un dipendente, un progetto e un settore
- In TEAM, che contiene i dati dei team che stanno lavorando sui progetti dell'azienda, il dipendente il cui codice identificativo è DipenId lavora sul progetto il cui identificativo è Progid
- In TEAM, il dominio dell'attributo Ruolo è {"ProjectLeader", "Designer", "Programmer"}
- In SETTORE, l'attributo Capold contiene l'Id del dipendente che è a capo di un settore e Nome è il nome del settore
- In PROGETTO, l'attributo SettlId contiene il codice identificativo del settore al quale appartiene il progetto (ogni progetto appartiene ad un solo settore ma un settore può contenere più progetti)

- 1) Elenco (nome, cognome e data di nascita) dei dipendenti che hanno lavorato su progetti più vecchi di 3 anni (iniziati prima dell'1/9/2020) nel il ruolo di "Desogner" oppure di "Programmer"
- 2) Titolo del progetto o dei progetti più vecchi (meno recenti) di cui ogni dipendente è stato Project Leader (nome e cognome del Project Leader e titolo del progetto).

$$\begin{aligned} \uparrow 1A) \\ \text{Proy Vecchi} &= \sigma_{\substack{\text{DataI} < 2020 \\ \text{Ruolo} \in \{\text{Designer}, \text{Programmer}\}}}(\text{Progetto}) \\ \text{TPV} &= \pi_{\text{DipenId}} \left(\sigma_{\substack{\text{Ruolo} \in \{\text{Designer}, \text{Programmer}\} \\ \text{Progid} = \text{Id}}}(\text{Team K Proy Vecchi}) \right) \end{aligned}$$

$$\text{OUT} = \pi_{\substack{\text{Nome,} \\ \text{Cognome,} \\ \text{DataN}}}(\text{DIPENDENTE K TPV})$$

$$\uparrow 1B) \\ \text{TEAM DIP} = \sigma_{\substack{\text{Ruolo} = \text{"Project Leader"} \\ \text{Id} = \text{DipenId}}}(\text{DIPENDENTE K TEAM})$$

$$\text{CP} = \text{PROGETTO} \quad \rho_{\text{Id} \rightarrow \text{CID}, \text{SettlId} \rightarrow \text{CSET}, \text{Titolo} \rightarrow \text{CTitolo}, \text{DataI} \rightarrow \text{CDataI}}$$

$$Q = \pi_{\text{Id}} \left(\sigma_{\substack{\text{DataI} < 2020 \\ \text{Id} = \text{CID}}}(\text{P K CP}) \right)$$

$$\text{OUT} = \pi_{\substack{\text{Nome,} \\ \text{Cognome,} \\ \text{Titolo}}}(\text{TEAM DIP K Q})$$

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

$F = \{BCD \rightarrow AH, BD \rightarrow CH, A \rightarrow D, C \rightarrow E, CD \rightarrow H, H \rightarrow E, AH \rightarrow B\}$

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)

$$2a) \quad (BCD)_F^+ = BCDCHAE \Rightarrow \text{CHIAVE} \quad (AB)_F^+ = R$$

$$(AH)_F^+ = AHBDCE \Rightarrow \text{CHIAVE} \quad (AC)_F^+ = R$$

2b) Non è in 3NF per $C \rightarrow E$ poiché C non è superchiave ed E non è primo

$$2c) \quad F = \{ \cancel{BCD \rightarrow A}, \cancel{BCD \rightarrow H}, BD \rightarrow C, BD \rightarrow H, A \rightarrow D, C \rightarrow E, CD \rightarrow H, H \rightarrow E, AH \rightarrow B \}$$

$$I \quad BCD \rightarrow A$$

$$(B)_F^+ = B \nrightarrow A$$

$$(C)_F^+ = CE \nrightarrow A$$

$$(D)_F^+ = D \nrightarrow A$$

$$- BD \rightarrow C$$

$$(B)_F^+ = B \nrightarrow C$$

$$(D)_F^+ = D \nrightarrow C$$

$$BCD \rightarrow H$$

$$(B)_F^+ = B \nrightarrow H$$

$$(C)_F^+ = CE \nrightarrow H$$

$$(D)_F^+ = D \nrightarrow H$$

$$- BD \rightarrow H$$

$$(B)_F^+ = B \nrightarrow H$$

$$(D)_F^+ = D \nrightarrow H$$

$$CD \rightarrow H$$

$$(C)_F^+ = C$$

$$(D)_F^+ =$$

$$II \quad BCD \rightarrow A$$

$$(BCD)_F^+ = BCDH \nrightarrow A \Rightarrow \text{non va bene}$$

$$B \rightarrow D \rightarrow H$$

$$(BCD)^+_F \vdash B \rightarrow D \rightarrow H = B \rightarrow D \rightarrow H \Rightarrow \neg \text{LC}$$

$$F = \{ B \rightarrow D \rightarrow A, \cancel{B \rightarrow D \rightarrow H}, B \rightarrow D \rightarrow C, B \rightarrow D \rightarrow H, A \rightarrow D, C \rightarrow E, CD \rightarrow H, H \rightarrow E, AH \rightarrow B \}$$

$$B \rightarrow D \rightarrow C$$

$$(BCD)^+_F \vdash B \rightarrow D \rightarrow C = B \rightarrow D \rightarrow C \Rightarrow \text{now } \neg \text{LC}$$

$$B \rightarrow D \rightarrow H$$

$$(BCD)^+_G \vdash B \rightarrow D \rightarrow H = B \rightarrow D \rightarrow H \Rightarrow \neg \text{LC}$$

$$F = \{ B \rightarrow D \rightarrow A, \cancel{B \rightarrow D \rightarrow H}, B \rightarrow D \rightarrow C, \cancel{B \rightarrow D \rightarrow H}, A \rightarrow D, C \rightarrow E, CD \rightarrow H, H \rightarrow E, AH \rightarrow B \}$$

$$\neg A \rightarrow D$$

$$(A)^+_G \vdash \neg A \rightarrow D = A \rightarrow D \Rightarrow \text{now } \neg \text{LC}$$

$$\neg C \rightarrow E$$

$$(C)^+_G \vdash \neg C \rightarrow E = C \rightarrow E \Rightarrow \text{now } \neg \text{LC}$$

$$\neg CD \rightarrow H$$

$$(CD)^+_G \vdash \neg CD \rightarrow H = CD \rightarrow H \Rightarrow \text{now } \neg \text{LC}$$

$$\neg H \rightarrow E$$

$$(H)^+_F \vdash \neg H \rightarrow E = H \rightarrow E \Rightarrow \text{now } \neg \text{LC}$$

$$\neg AH \rightarrow B$$

$$(AH)^+_F \vdash \neg AH \rightarrow B = AH \rightarrow B \Rightarrow \text{now } \neg \text{LC}$$

$$F = \{ BD \rightarrow C, A \rightarrow D, C \rightarrow E, CD \rightarrow H, H \rightarrow G, AH \rightarrow B \}$$

$$\rho = \{ BDA, BDC, AD, CE, CDH, HE, AHB \}$$

POE SNA E' IN 3NF E HA SCN SENZA PERDITA

E' dato un file di 145.950 record. Ogni record occupa 255 byte. Un blocco contiene 2048 byte. Un puntatore a blocco occupa 5 byte. Si utilizza una organizzazione hash con record distribuiti uniformemente tra 250 bucket.

- Calcolare l'occupazione in blocchi della bucket directory e dei bucket
- Calcolare il costo medio di una ricerca considerando chiavi di hash univoche
- Quanti bucket occorrerebbero per avere un costo medio di ricerca minore o uguale a 10 accessi?

$$a) \text{ Blochi} \times \text{Bucket dir} = \frac{\lceil \text{N}^\circ \text{Record} \rceil}{\text{Punter} \times \text{Blocco}} = \frac{\lceil \frac{250}{409} \rceil}{1} = 1$$

$$\text{Punter} \times \text{Blocco} = \frac{\text{Bucket size}}{\lceil \frac{\text{Punter size}}{5} \rceil} = \frac{2048}{5} = 409$$

$$\text{Blochi} \times \text{Bucket} = \frac{\lceil \text{Record} \times \text{Bucket} \rceil}{\text{Record} \times \text{Blocco}} = \frac{\lceil \frac{589}{5} \rceil}{1} = 23$$

$$\text{Record} \times \text{Bucket} = \frac{\lceil \text{N}^\circ \text{Record} \rceil}{\text{N}^\circ \text{Bucket}} = 584$$

$$\text{Record} \times \text{Blocco} = \frac{\lceil \frac{\text{Block size} - \text{Punter}}{\text{Record size}} \rceil}{1} = \frac{\lceil \frac{2048 - 5}{255} \rceil}{1} = 8$$

$$b) \text{ Avg Time} = \frac{\lceil \frac{23}{2} \rceil}{2} = 39$$

$$c) \text{ Avg time} \leq 10 \quad \frac{\text{Record} \times \text{Bucket}}{\text{Record} \times \text{Blocco}} \leq 20$$

$$\text{Noord} \times \text{Zuid} \leq 160$$

$$\frac{\text{Noord}}{\text{Noord} + \text{Zuid}} \leq 160$$

$$\text{Noord} \geq \frac{\text{Noord}}{160} = 913$$