

Seconda prova di verifica del 17/12/2007

1. Si consideri lo schema relazionale $R(B, C, D, F, G, H)$ con le seguenti DF, già in forma canonica:

$BG \rightarrow C, G \rightarrow F, C \rightarrow H, C \rightarrow G, F \rightarrow D$

- (a) La dipendenza $BC \rightarrow DH$ è derivabile dalle DF?
- (b) La decomposizione $BCDGH$ e $DFGH$ preserva i dati?
- (c) Si trovino le chiavi di R .
- (d) Dire se lo schema è in 3FN o in FNBC.
- (e) Si applichi allo schema l'algoritmo di sintesi per ottenere una decomposizione in 3FN che preserva i dati.
- (f) Se nella decomposizione del passo precedente esistono schemi non in FNBC, decomporli in FNBC e si dica se la nuova decomposizione di R preserva le dipendenze.
- (g) (Opzionale) Si porti il seguente schema relazionale in forma canonica: $R\langle(A, B, C, D, E, F, G), \{ABE \rightarrow CD, AE \rightarrow F, F \rightarrow B, A \rightarrow DG, AD \rightarrow G\}\rangle$.

2. Si consideri lo schema relazionale:

$R(\underline{Rpk}: \text{integer}, RA: \text{integer}, RB: \text{integer})$
 $S(\underline{Spk}: \text{integer}, SA: \text{string}, SB: \text{integer}, Rfk*: \text{integer})$
 $U(\underline{Sfk*}: \text{integer}, \underline{UA}: \text{string})$

Si supponga che la relazione S sia *ordinata* sulla chiave primaria Spk . Si consideri l'interrogazione

```
SELECT    DISTINCT Spk, COUNT(*) AS c
FROM      R, S, U
WHERE     R.Rpk = S.Rfk AND S.Spik = U.Sfk
          AND RA = 10 AND RB = 100 AND SB = 20
GROUP BY  Spk
HAVING    COUNT(*) > 5
ORDER BY  c ;
```

- (a) Disegnare l'albero di sintassi astratta di un'espressione algebrica (albero logico) per l'interrogazione.
- (b) Si dica se il seguente piano d'accesso è corretto e produce il risultato cercato. Se non va bene si dica perché.

- (c) Se non va bene, lo si modifichi in tre modi: (a) rendendolo corretto facendo solo le modifiche necessarie, (b) riscrivendo poi il piano corretto anticipando restrizioni e proiezioni ed eliminando operatori inutili, e (c) riscrivendo il piano usando solo due indici a vostra scelta.

Sort
({c})

Distinct

Filter
(RA = 10 AND RB = 100 AND SB = 20)

GroupBy
({Spk},{COUNT(*)>5})

NestedLoop
(R.Rpk = S.Rfk)

TableScan
(U)

TableScan **TableScan**
(R) (S)

3. Considerate lo schema:

StudentiEsami(Matricola, Nome, Cognome, Voto, Materia, Data)

- (a) Spiegate, con una o al più due frasi, perché l'uso di questo schema presenta dei problemi, con un linguaggio comprensibile al committente che deve usare la tabella (ovvero, senza parlare di dipendenze funzionali).
- (b) Con lo stesso linguaggio non tecnico, e altrettanto brevemente, spiegate perché la decomposizione indicata sotto non va bene:

Studenti(Matricola, Nome, Cognome)
Esami(Cognome, Voto, Materia, Data)

- (c) Specificate due dipendenze funzionali che immaginate valgano nello schema considerato, prima in italiano, e poi con la notazione $A \rightarrow B$.

4. Enunciate, con una o al più due frasi, che cosa è un indice.

Seconda prova di verifica del 17/12/2007 - Soluzioni

1. Si consideri lo schema relazionale $R(B, C, D, F, G, H)$ con le seguenti DF, già in forma canonica:

$BG \rightarrow C, G \rightarrow F, C \rightarrow H, C \rightarrow G, F \rightarrow D$

- (a) La dipendenza $BC \rightarrow DH$ è derivabile dalle DF?

Sì: $BC_F^+ = BCFGDH \supseteq DH$

- (b) La decomposizione $BCDGH$ e $DFGH$ preserva i dati?

Sì: $BCDGH \cap DFGH = DGH$, e $DGH_F^+ = DGHF$

- (c) Si trovino le chiavi di R .

Sono BC e BG .

- (d) Dire se lo schema è in 3FN o in FNBC.

No: la dipendenza $G \rightarrow F$ viola la 3FN, e quindi anche la FNBC.

- (e) Si applichi allo schema l'algoritmo di sintesi per ottenere una decomposizione in 3FN che preserva i dati.

$R1 \langle (B, C, G), \{BG \rightarrow C\} \rangle$,

$R2 \langle (F, G), \{G \rightarrow F\} \rangle$,

$R3 \langle (C, G, H), \{C \rightarrow G, C \rightarrow H\} \rangle$,

$R4 \langle (F, D), \{F \rightarrow D\} \rangle$

- (f) Se nella decomposizione del passo precedente esistono schemi non in FNBC, decomporli in FNBC e si dica se la nuova decomposizione di R preserva le dipendenze.

$R1 \langle (B, C, G), \{BG \rightarrow C, C \rightarrow G\} \rangle$ non è in FNBC, va decomposta in $R5 \langle (C, G), \{C \rightarrow G\} \rangle$ ed $R6 \langle (C, B), \{ \} \rangle$, e si perde la dipendenza $BG \rightarrow C$.

- (g) (Opzionale) Si porti il seguente schema relazionale in forma canonica: $R \langle (A, B, C, D, E, F, G), \{ABE \rightarrow CD, AE \rightarrow F, F \rightarrow B, A \rightarrow DG, AD \rightarrow G\} \rangle$.

Si ottiene: $R \langle (A, B, C, D, E, F, G), \{AE \rightarrow C, AE \rightarrow F, F \rightarrow B, A \rightarrow D, A \rightarrow G\} \rangle$

2. Si consideri lo schema relazionale:

$R(\underline{Rpk}: \text{integer}, RA: \text{integer}, RB: \text{integer})$

$S(\underline{Spk}: \text{integer}, SA: \text{string}, SB: \text{integer}, Rfk*: \text{integer})$

$U(\underline{Sfk*}: \text{integer}, \underline{UA}: \text{string})$

Si supponga che la relazione S sia ordinata sulla chiave primaria Spk . Si consideri l'interrogazione

```

SELECT    DISTINCT Spk, COUNT(*) AS c
FROM      R, S, U
WHERE     R.Rpk = S.Rfk AND S.SpK = U.Sfk
          AND RA = 10 AND RB = 100 AND SB = 20
GROUP BY  Spk
HAVING    COUNT(*) > 5
ORDER BY  c ;

```

- (a) *Disegnare l'albero di sintassi astratta di un'espressione algebrica (albero logico) per l'interrogazione. (Figura 1)*

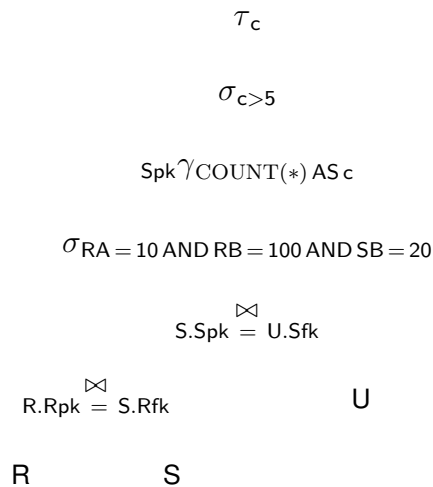


Figura 1: Albero logico

- (b) *Si dica se il seguente piano d'accesso è corretto e produce il risultato cercato. Se non va bene si dica perché.*

Il piano di accesso è scorretto perché (1) la GroupBy ammette solo un operando, mentre qui ne ha due (NestedLoop e TableScan), e calcola funzioni di aggregazione senza fare restrizioni (COUNT(*) > 5), (2) l'operatore Filter è applicato dopo che la GroupBy ha già eliminato gli attributi RA, RB, SB, (3) la GroupBy è applicata a dati non ordinati su Spk, (4) manca la condizione di HAVING e (5) il Sort ordina su c che va prodotto con una Project. (Figura 2)

- (c) *Se non va bene, lo si modifichi in tre modi: (a) rendendolo corretto facendo solo le modifiche necessarie, (b) riscrivendo poi il piano corretto anticipando restrizioni e proiezioni ed eliminando operatori inutili, e (c) riscrivendo il piano usando solo due indici a vostra scelta.*

(a) Piano di Figura 3. (b) Piano di Figura 4. L'operatore Distinct è inutile e usando come relazione esterna S diventa inutile anche il Sort prima di GroupBy. (c) Molte soluzioni sono possibili (Figura 5).

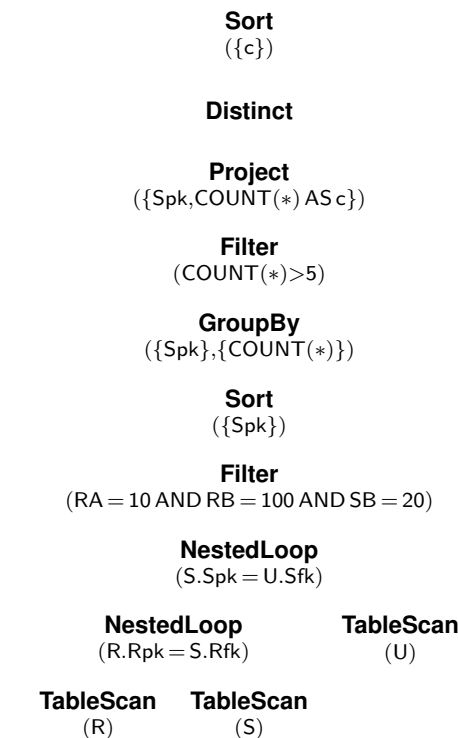


Figura 2: Piano d'accesso corretto

3. Considerate lo schema:

StudentiEsami(Matricola, Nome, Cognome, Voto, Materia, Data)

- (a) *Spiegate, con una o al più due frasi, perché l'uso di questo schema presenta dei problemi, con un linguaggio comprensibile al committente che deve usare la tabella (ovvero, senza parlare di dipendenze funzionali).*

Se uno studente supera più esami, il suo Nome e Cognome sono ripetuti in più righe, con la possibilità che si presentino valori diversi per errori di immissione o di aggiornamento. Inoltre, con questo schema, uno studente senza esami deve essere rappresentato in modo incoerente rispetto agli studenti con esami.

- (b) *Con lo stesso linguaggio non tecnico, e altrettanto brevemente, spiegate perché la decomposizione indicata sotto non va bene:*

Studenti(Matricola, Nome, Cognome)

Esami(Cognome, Voto, Materia, Data)

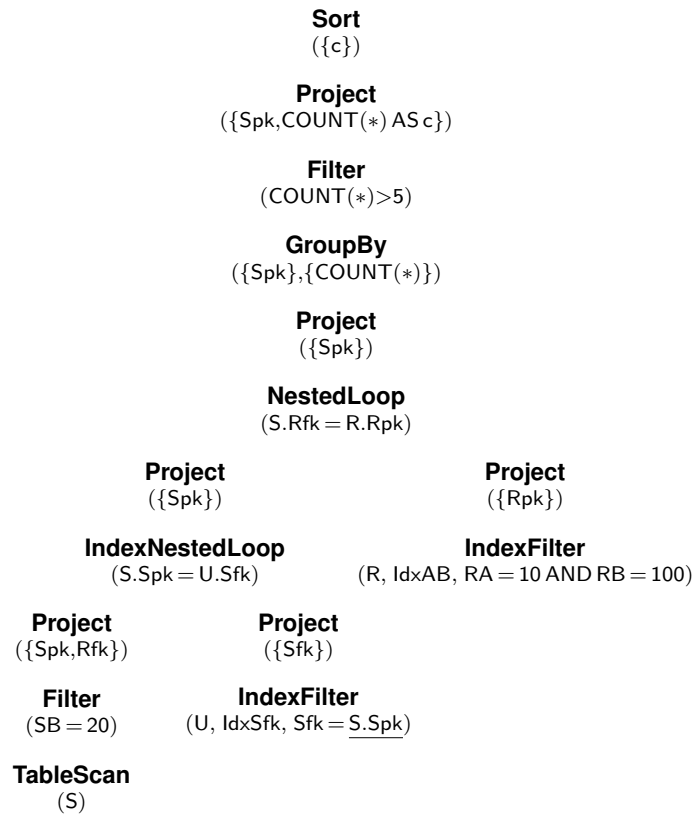


Figura 4: Piano d'accesso con un indice multiattributo su RA e RB ed un indice sulla chiave esterna di U

utilizzato per accedere rapidamente alle righe di R che soddisfano ad una condizione $A = k$ oppure $k_1 \leq A \leq k_2$.