

Qualche esercizio

Sia dato un file F con 16000 record di 56 byte ciascuno. Sia inoltre dato un disco con settori di 256 byte, un fattore di bloccaggio pari a 4, 24 settori per traccia, n tracce per cilindro, dove n è esattamente il numero minimo di tracce necessarie per memorizzare tutto F . In termini di record (di F), è più capiente il suddetto disco o un nastro magnetico lungo 600 pollici, con inter-record gap di 0.25 pollici, stesso fattore di bloccaggio del disco e densità di 800 bit per pollice?

Sia dato un indice unclustered organizzato come un B^+ -tree che usa un posting file separato per la memorizzazione delle liste di TID. La dimensione dei nodi è 1KB (1 nodo = 1 blocco), la lunghezza del puntatore al posting file è 4 byte, la lunghezza della chiave è 20 byte, ci sono un milione di record memorizzati in 40000 pagine, il numero di valori distinti di chiave è 10000, l'utilizzazione pari a $\ln 2$, l'ordine $g = 21$.

Data una query che richiede di reperire i record relativi a 10 valori di chiave distinti, stimare il numero di accessi necessari utilizzando l'indice suddetto, assumendo che, in media, ogni posting file occupi 1 KB.

Siano dati:

- un file F organizzato ad accesso sequenziale che, oltre alle classiche operazioni logiche, permette anche la `update` per la modifica del record corrente; il file, memorizzato su disco in modo contiguo, contiene 16000 record allocati in blocchi la cui capacità è di 32 record;
- una cache di 32 pagine (1 pagina = 1 blocco);
- un controllore che opera prefetching su tutta la cache e che adotta la politica read-after-write.

Si assuma che l'ultima richiesta di lettura sia stata per il record 12801 e che F risulti ancora aperto. Determinare il costo per accedere e modificare il record 14529.

Un disco è caratterizzato da:

- dimensione di un blocco: 4 KB
- tempo di seek: 12.6 ms
- tempo di rotazione: 13.9 ms
- tempo di trasferimento: 1.6 MB/s

Determinare il tempo medio di lettura di un blocco nel caso in cui la velocità di rotazione del disco sia raddoppiata.

Definire uno schema E/R che descriva i dati relativi all'archivio di un amministratore di condomini, secondo le seguenti specifiche:

- ogni condominio ha un indirizzo e comprende un insieme di appartamenti raggruppati in scale;
- all'interno di un condominio le scale sono numerate;
- in ogni condominio, gli appartamenti sono identificati dalla scala di appartenenza e dall'interno;
- agli appartamenti sono associati i millesimi, il proprietario e l'inquilino (che possono essere distinti), ciascuno con codice fiscale, cognome e nome.

Nello schema vanno definite cardinalità e identificatori.

Realizzare la semplificazione e la traduzione in uno schema relazionale.

Definire uno schema E/R che descriva i dati relativi al listino prezzi di un salone automobilistico, secondo le seguenti specifiche:

- le auto appartengono a case produttrici con nome (es. Saab, Opel), indirizzo di riferimento, nazionalità;
- per ogni casa sono disponibili in salone diversi modelli (es. Corsa, Calibra) con nome, anno di lancio, tipologia di vettura (es. berlina, utilitaria), numero di versioni;
- per ogni versione, si conoscono prezzo, combustibile, cilindrata, numero di porte, velocità massima e la eventuale presenza di due optional: climatizzatore e ABS; le versioni si identificano attraverso nome, modello e casa di produzione (es. Opel Corsa Swing);
- ogni auto è identificata dal numero di telaio e si distingue in auto nuova, usata o a chilometri zero, nel qual caso interessa sapere quanti chilometri ha.

Discutere le scelte progettuali sapendo che ai clienti interessano in primo luogo le auto climatizzate.

Sia dato il seguente schema relazionale:

IMPIEGATI(CF, *Cognome*, *Nome*, *DataNascita*, *Reparto*, *Stipendio*)
REPARTI(Codice, *Nome*, *Indirizzo*)
PROGETTI(Sigla, *Titolo*, *Valore*)
PARTECIPAZIONE(Impiegato, Progetto, *DataInizio*, *DataFine*).

Esistono vincoli di integrità referenziale tra *Reparto* e la relazione REPARTI, tra *Progetto* e la relazione PROGETTI, tra *Impiegato* e la relazione IMPIEGATI. Risolvere i seguenti problemi:

1. Definire in SQL la vista V come join delle relazioni IMPIEGATI, PARTECIPAZIONE e PROGETTI relative ad impiegati che sono ancora attivi in qualche progetto.
2. Definire una query SQL che restituisce *CF*, *Titolo*, e *DataInizio* relativamente a impiegati che hanno iniziato a lavorare a progetti nel 2002 e che vi lavorano ancora.
3. Definire in algebra relazionale la formula che restituisce *Nome* e *Cognome* degli impiegati che partecipano (o hanno partecipato) ad almeno due progetti.
4. Definire in calcolo relazionale l'espressione che restituisce *Nome* e *Indirizzo* dei reparti dove lavorano almeno tre persone coinvolte ancora attivamente nello stesso progetto.

Siano date le seguenti relazioni:

$R(\underline{A}, B, C)$
 $S(\underline{D}, E, F)$

di cardinalità n e m tuple, rispettivamente. Inoltre, esiste un vincolo di integrità referenziale tra C e la chiave primaria di S .

1. Quante tuple contiene la relazione che esprime il θ -join tra le due relazioni, dove il predicato di join è $C = D$?
2. Quante invece se il predicato di join è $B = E$?
3. Esprimere in algebra relazionale, calcolo relazionale, ed SQL, l'espressione che calcola il θ -join tra le due relazioni, con predicato di join $B < F$ e tale per cui sono escluse le tuple di R il cui valore di C è unico (ovvero presente una sola volta nella relazione).

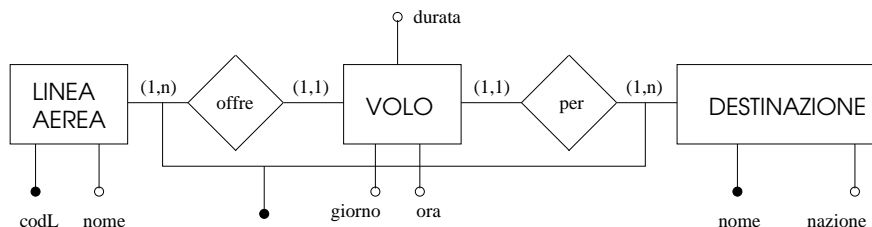
Discutere le seguenti transazioni concorrenti.

T1	Tempo	T2
read[X]	1	
write[X]	2	
	3	read[Y]
	4	read[X]
	5	write[Y]
commit	6	
	7	commit

T3	Tempo	T4
write[X]	1	
	2	read[X]
	3	write[Y]
read[Z]	4	
	5	commit
rollback	6	

T5	Tempo	T6
write[X]	1	
	2	write[X]
	3	write[Y]
read[Z]	4	
	5	commit
rollback	6	

Sia dato il seguente schema E/R.



Si disegni lo schema di navigazione delle seguenti query:

- Data una destinazione X ed un giorno Y, si visualizzino i nomi delle linee aeree che effettuano voli per X nel giorno Y.
- Data una destinazione X, si calcoli la durata media di un volo per X.

Si effettui il progetto logico relazionale di tale schema E/R, evidenziando i passi di semplificazione prima della fase di traduzione.

Si formuli in algebra relazionale, calcolo relazionale e SQL la query a) e in SQL la query b).

Sia dato il seguente schema relazionale:

$R1(A, B, C)$

$R2(A, D, F)$

$R3(D, E, F)$

Formulare le query SQL relative alle seguenti espressioni di algebra relazionale:

- $\pi_{AF}(R1 \bowtie R2 \bowtie R3)$
- $(\pi_A(R1) \cup \pi_A(R2)) - \pi_A(R2 \bowtie R3)$
- $(\pi_A(R1) \cap \pi_A(R2)) \bowtie_{A>D} R3$
- $R2 \div \pi_{DF}(R3)$

Sia dato il seguente schema di relazione:

$ORDINI(NumeroOrdine, CodiceCliente, IndirizzoCliente, Data, CodiceProdotto, DescrizioneProdotto, PrezzoUnitario, Quantit\grave{a}, Totale, TotaleOrdine)$.

Si suppone quanto segue:

- ogni ordine \u00e8 relativo ad un cliente, ma ad uno o pi\u00f9 prodotti, ciascuno dei quali compare una sola volta nell'ordine;
- ogni cliente ha un indirizzo;

- ogni prodotto, identificato dal codice, ha una sola descrizione ed un unico prezzo unitario;
- *Totale* indica la spesa totale per un prodotto in un ordine ($= \text{PrezzoUnitario} \times \text{Quantità}$), mentre *TotaleOrdine* indica la spesa totale per un ordine.

Trasformare lo schema ORDINI in uno schema normalizzato.

Sia dato il seguente schema relazionale:

CITTÀ(nome, regione, abitanti)
 FIUME(nome, lunghezza, affluente_di)
 SCORRE(città, fiume)

Esistono vincoli di integrità referenziale tra *città* e la relazione CITTÀ, tra *fiume* e la relazione FIUME e tra *affluente_di* e la relazione FIUME.

- Se le relazioni CITTÀ e FIUME hanno cardinalità N_1 e N_2 , rispettivamente, stabilire la cardinalità massima teorica della relazione SCORRE.
- Formulare le seguenti query in algebra relazionale e calcolo relazionale:
 - a) Visualizzare nome e regione delle città con più di 50000 abitanti e attraversate dal Po o da uno dei suoi affluenti.
 - b) Trovare le città che sono attraversate da almeno due fiumi più lunghi di 100 km.
- Dichiarare in MYSQL le tabelle CITTÀ, FIUME, SCORRE e risolvere le query a) e b) più le seguenti due:
 - c) Trovare il fiume più lungo che attraversa città piemontesi.
 - d) Elencare gli affluenti del Po che non passano per città lombarde.

Sia dato il seguente schema relazionale:

STUDENTI(Matricola, Cognome, Nome)
 ESAMI(Studente, Materia, Voto, Data).

Esiste un vincolo di integrità referenziale tra *Studente* e la relazione STUDENTI. Risolvere i seguenti problemi:

1. Qual è la cardinalità di $\pi_{\text{Matricola, Cognome}}(\text{STUDENTI})$ se la cardinalità di STUDENTI è n_1 ?
2. Qual è la cardinalità di $\text{STUDENTI} \bowtie_{\text{Matricola=Studente}} \text{ESAMI}$ se la cardinalità di ESAMI è n_2 ?

3. Trovare matricola, nome e cognome degli studenti che hanno preso almeno un 30 (esprimere in algebra relazionale, calcolo relazionale e SQL).
4. Trovare le matricole che hanno superato almeno due esami dopo il 20/04/2004 (esprimere in algebra relazionale, calcolo relazionale e SQL).
5. Trovare nome e cognome degli studenti che hanno superato almeno un esame dopo il 20/04/2004 (esprimere in algebra relazionale, calcolo relazionale e SQL).
6. Trovare le matricole che hanno preso almeno due 30 (esprimere in algebra relazionale, calcolo relazionale e SQL).
7. Trovare matricola, cognome e numero di esami superati per ogni studente (esprimere in MySQL).

Sia data la seguente query:

```
SELECT * FROM Veicoli, Modelli
WHERE Veicoli.Cod_Modello = Modelli.Cod_Modello
AND (Veicoli.Cilindrata > 1500 OR Veicoli.Velocità > 160)
AND Veicoli.Cod_Categoria = '01'
AND Modelli.Cod_Fabbrica = '001'
```

e i seguenti dati:

- un unclustered B⁺-tree su *Veicoli.Cod_Categoria*, con NL = 30 e $h = 3$
- NT_{Veicoli} = 5000, NP_{Veicoli} = 5000,
NK_{Cod_Categoria} = 20,
NK_{Cilindrata} = 20, con valori compresi tra 900 e 2800,
NK_{Velocità} = 10, con valori compresi tra 110 e 200
- NT_{Modelli} = 100, NP_{Modelli} = 4,
NK_{Cod_Fabbrica} = 10

Discutere la scelta tra le possibili modalità per risolvere la query in modo ottimale (metodo di join: nested-block join).