

Basi di dati II, primo modulo Prova parziale — 22 aprile 2010 — Compito A

Rispondere su questo fascicolo. Consegnare anche i fogli usati come brutta copia, che però non verranno corretti.
Tempo a disposizione: un'ora e trenta minuti.

Cognome _____ Nome _____ Matricola _____ Ordin. _____

Domanda 1 Considerare un disco con una velocità di rotazione di $v = 10.000$ giri al minuto e un tempo medio di posizionamento della testina (tempo di seek) $t_S = 5$ msec. Ogni traccia contiene $N = 240$ blocchi e ogni blocco contiene $B = 4$ KB. Rispondere alle seguenti domande mostrando formula e valore numerico numero (N.B. non servono calcolatrici, i risultati sono semplici, approssimare $1 \text{ MB} = 1000 \text{ KB}$).

1. Qual è tempo medio di latenza (attesa dovuta alla rotazione) t_L ?

2. Qual è la massima velocità di trasferimento (in MB al secondo)?

3. Qual è il tempo minimo t_B necessario per leggere un blocco?

4. Qual è il tempo medio necessario per leggere un blocco?

5. Qual è il tempo medio necessario per leggere $n = 40$ blocchi contigui?

Domanda 2 Una catena di supermercati ha una base di dati dei propri clienti che dispongono di una “tessera fedeltà,” con varie informazioni su ciascun cliente, fra cui (a) il totale dei punti acquisiti attraverso l’uso della tessera e (b) il negozio della catena cui fa riferimento (ad esempio, quello presso cui ha inizialmente richiesto la tessera). Si vuole eseguire su di essa l’interrogazione che calcola, per ciascun negozio, il numero dei clienti, la somma dei punti fedeltà dei clienti e la relativa media per cliente. Indicare quale livello di isolamento (READ UNCOMMITTED, READ COMMITTED, REPEATABLE READ o SERIALIZABLE) si potrebbe scegliere in ciascuno dei seguenti casi (si supponga che, in generale, sia stato rilevato che, nel corso degli inserimenti e delle modifiche, vengono inseriti valori sbagliati anche di vari ordini di grandezza, che sono poi corretti prima del commit):

1. L’operazione è eseguita mentre vengono inseriti alcuni nuovi clienti (per ciascun negozio pochi rispetto a quelli già presenti), con la finalità di acquisire informazioni approssimate ma ragionevolmente indicative sugli andamenti complessivi.
2. L’operazione è eseguita mentre vengono inseriti molti nuovi clienti, con la finalità di acquisire informazioni approssimate ma ragionevolmente indicative sugli andamenti complessivi.
3. L’operazione è eseguita mentre vengono inseriti alcuni nuovi clienti (per ciascun negozio pochi rispetto a quelli già presenti), con la finalità di individuare i primi tre negozi da premiare in una campagna promozionale sulla base dei punti acquisiti dai rispettivi clienti.
4. L’operazione è eseguita mentre vengono modificati i valori dei punti fedeltà di tutti i clienti (a seguito di una ridefinizione dei criteri di assegnazione dei punti stessi), con la finalità di individuare i primi tre negozi da premiare in una campagna promozionale sulla base dei punti acquisiti dai rispettivi clienti.
5. L’operazione è eseguita in un momento in cui non ci sono aggiornamenti di alcun genere, con la finalità di acquisire informazioni approssimate ma ragionevolmente indicative sugli andamenti complessivi.

| Risposte | | | | |
|----------|----|----|----|----|
| 1. | 2. | 3. | 4. | 5. |
| | | | | |

Domanda 3 Molti DBMS utilizzano, oltre a quelli basati sui B⁺-tree, anche indici basati su una tecnica detta "bitmap." Un indice bitmap, su un attributo *A* con *a* valori diversi, per una relazione con N_R ennuple, è costituito da *a* vettori (uno per ciascun valore di *A*) di N_R bit ciascuno (un bit per ogni ennupla della relazione): l'*i*-esimo bit del vettore associato al valore a_j è 1 se il valore della *i*-esima ennupla sull'attributo *A* è a_j e 0 se il valore è diverso. L'accesso ai vettori è organizzato per mezzo di un albero. In sostanza, rispetto ad un B⁺-tree, abbiamo una struttura di accesso simile, ma con foglie diverse: nel B⁺-tree, per ogni valore dell'attributo abbiamo una lista di indirizzi, mentre nell'indice bitmap abbiamo un vettore di bit. Esiste poi una struttura (tabella, indice o altro) che associa gli indirizzi delle ennuple ai numeri da 1 a N_R . Sulla base degli elementi forniti, con riferimento ad un sistema con blocchi di dimensione $B = 4$ KB e indirizzi dei record di lunghezza $b = 8$ B e ad una relazione R con $N_R = 1.600.000$ ennuple e (fra gli altri) quattro attributi, A, C, E, F ciascuno con $a = 20$ valori diversi e una lunghezza $c = 6$ byte

1. valutare lo spazio SBM (in blocchi) necessario per le foglie di un indice bitmap su A; valutare anche lo spazio S_I necessario per una tabella (una sorta di array in memoria secondaria, ad esempio) che permetta di associare un indirizzo a ciascun valore compreso fra 1 e N_R ; per confronto, calcolare lo spazio ST necessario per le foglie di un B⁺-tree (con riferimenti ai record) per un indice su A

Poi, valutare (sia nel caso di indici bitmap, sia nel caso di B⁺-tree) il costo (in termini di numero di accessi a memoria secondaria, supponendo che tutti i nodi intermedi di tutti gli indici siano nei buffer; basta quindi contare il numero di foglie da visitare e l'accesso alle ennuple, quando necessario) per le seguenti interrogazioni

2. `SELECT COUNT(*) FROM R WHERE A = 5`
3. `SELECT * FROM R WHERE A = 5`
4. `SELECT * FROM R WHERE A = 15 AND C = 14 AND E = 13 AND F = 9` (considerare i valori indipendenti l'uno dall'altro; quindi assumere che il risultato contenga circa N_R/a^4 ennuple)

Riportare le risposte nella tabella sottostante, nelle righe da 1 a 4, indicando formula e valore numerico.

| | | Risposte | |
|---|--|---------------|-----------------------------|
| | | Indici bitmap | Indici B ⁺ -tree |
| 1 | | | |
| 2 | | | |
| 3 | | | |
| 4 | | | |

Domanda 4 Specificare per ciascuno dei seguenti schedule, la sua appartenenza alle varie classi (scrivere “sì” o “no” nelle caselle della tabella; assumere che gli identificatori delle transazioni corrispondano ai timestamp).

| | Seriale | VSR | CSR | TS | 2PL |
|--|---------|-----|-----|----|-----|
| $r_1(y)r_2(z)r_2(y)w_2(y)w_2(z)r_1(z)$ | | | | | |
| $r_1(x)w_1(x)r_2(x)w_2(x)r_0(y)w_1(y)$ | | | | | |
| $r_1(x)r_2(z)w_2(z)w_1(x)r_2(x)w_2(x)$ | | | | | |
| $r_2(x)w_2(x)r_1(x)w_1(x)$ | | | | | |

Domanda 5 Considerare uno schema dimensionale relativo agli esami, che utilizzi, come tabella dei fatti e come una delle dimensioni, le relazioni come quelle qui schematizzate:

| <u>K</u> Studente | <u>K</u> Corso | <u>K</u> Data | Voto | ... |
|-------------------|----------------|---------------|------|-----|
| 301 | 201 | 405 | 25 | ... |
| 301 | 202 | 406 | 28 | ... |
| 302 | 201 | 407 | 30 | ... |
| 302 | 203 | 407 | 22 | ... |
| ... | ... | ... | ... | ... |

| <u>K</u> Corso | Titolo | Crediti | ... |
|----------------|-----------|---------|-----|
| 201 | Fisica I | 6 | ... |
| 202 | Chimica | 9 | ... |
| 203 | Geometria | 6 | ... |
| ... | ... | ... | ... |

Supporre che si presentino le seguenti esigenze di modifica:

- i corsi cambiano nome nel tempo: per esempio, il corso nella prima ennupla potrebbe ad un certo punto cambiare nome da “Fisica I” in “Meccanica”; interessano selezioni e aggregazioni relative agli esami tanto con riferimento al nome del corso (al momento dell’esame) quanto alla sua identità (un codice che viene introdotto allo scopo, ma non sempre viene utilizzato, perché alcuni analisti preferiscono fare riferimento al nome corrente del corso); le modifiche sono rare, ma è possibile che ci siano corsi con vari cambiamenti di nome;
- per ogni corso, interessa rappresentare anche il docente, per supportare analisi sugli esami svolti da ciascun docente; i docenti cambiano nel tempo e passano da un corso all’altro (e possono anche tenere più corsi nello stesso momento, ma ogni corso ha, in un certo giorno, un solo docente); è disponibile l’informazione relativa ai docenti dei corsi nel tempo (per tutto il periodo, anche passato, di interesse).

Mostrare nuove versioni delle due tabelle che permettano di soddisfare le esigenze sopra citate (mostrare anche i dati, con riferimento a quelli presenti negli esempi sopra, aggiungendo nuovi dati ragionevoli, che permettano di comprendere le modifiche).

Basi di dati II, primo modulo Prova parziale — 22 aprile 2010 — Compito B

Rispondere su questo fascicolo. Consegnare anche i fogli usati come brutta copia, che però non verranno corretti.
Tempo a disposizione: un'ora e trenta minuti.

Cognome _____ Nome _____ Matricola _____ Ordin. _____

Domanda 1 Considerare un disco con una velocità di rotazione di $r = 10.000$ giri al minuto e un tempo medio di posizionamento della testina (tempo di seek) $t_S = 5$ msec. Ogni traccia contiene $P = 300$ blocchi e ogni blocco contiene $B = 4$ KB. Rispondere alle seguenti domande mostrando formula e valore numerico numero (N.B. non servono calcolatrici, i risultati sono semplici, approssimare $1 \text{ MB} = 1000 \text{ KB}$).

1. Qual è tempo medio di latenza (attesa dovuta alla rotazione) t_L ?

2. Qual è la massima velocità di trasferimento (in MB al secondo)?

3. Qual è il tempo minimo t_B necessario per leggere un blocco?

4. Qual è il tempo medio necessario per leggere un blocco?

5. Qual è il tempo medio necessario per leggere $c = 50$ blocchi contigui?

Domanda 2 Una catena di supermercati ha una base di dati dei propri clienti che dispongono di una “tessera fedeltà,” con varie informazioni su ciascun cliente, fra cui (a) il totale dei punti acquisiti attraverso l’uso della tessera e (b) il negozio della catena cui fa riferimento (ad esempio, quello presso cui ha inizialmente richiesto la tessera). Si vuole eseguire su di essa l’interrogazione che calcola, per ciascun negozio, il numero dei clienti, la somma dei punti fedeltà dei clienti e la relativa media per cliente. Indicare quale livello di isolamento (READ UNCOMMITTED, READ COMMITTED, REPEATABLE READ o SERIALIZABLE) si potrebbe scegliere in ciascuno dei seguenti casi (si supponga che, in generale, sia stato rilevato che, nel corso degli inserimenti e delle modifiche, vengono inseriti valori sbagliati anche di vari ordini di grandezza, che sono poi corretti prima del commit):

1. L’operazione è eseguita mentre vengono inseriti molti nuovi clienti, con la finalità di acquisire informazioni approssimate ma ragionevolmente indicative sugli andamenti complessivi.
2. L’operazione è eseguita mentre vengono modificati i valori dei punti fedeltà di alcuni clienti (in ciascun negozio una piccola percentuale), con la finalità di acquisire informazioni approssimate ma ragionevolmente indicative sugli andamenti complessivi.
3. L’operazione è eseguita in un momento in cui non ci sono aggiornamenti di alcun genere, con la finalità di individuare i primi tre negozi da premiare in una campagna promozionale sulla base dei punti acquisiti dai rispettivi clienti.
4. L’operazione è eseguita mentre vengono inseriti molti nuovi clienti, con la finalità di individuare i primi tre negozi da premiare in una campagna promozionale sulla base dei punti acquisiti dai rispettivi clienti.
5. L’operazione è eseguita mentre vengono modificati i valori dei punti fedeltà di alcuni clienti (in ciascun negozio una piccola percentuale), con la finalità di individuare i primi tre negozi da premiare in una campagna promozionale sulla base dei punti acquisiti dai rispettivi clienti.

| Risposte | | | | |
|----------|----|----|----|----|
| 1. | 2. | 3. | 4. | 5. |
| | | | | |

Domanda 3 Molti DBMS utilizzano, oltre a quelli basati sui B⁺-tree, anche indici basati su una tecnica detta "bitmap." Un indice bitmap, su un attributo *A* con *v* valori diversi, per una relazione con N_R ennuple, è costituito da *v* vettori (uno per ciascun valore di *A*) di N_R bit ciascuno (un bit per ogni ennupla della relazione): l'*i*-esimo bit del vettore associato al valore a_j è 1 se il valore della *i*-esima ennupla sull'attributo *A* è a_j e 0 se il valore è diverso. L'accesso ai vettori è organizzato per mezzo di un albero. In sostanza, rispetto ad un B⁺-tree, abbiamo una struttura di accesso simile, ma con foglie diverse: nel B⁺-tree, per ogni valore dell'attributo abbiamo una lista di indirizzi, mentre nell'indice bitmap abbiamo un vettore di bit. Esiste poi una struttura (tabella, indice o altro) che associa gli indirizzi delle ennuple ai numeri da 1 a N_R . Sulla base degli elementi forniti, con riferimento ad un sistema con blocchi di dimensione $D = 8$ KB e indirizzi dei record di lunghezza $p = 8$ B e ad una relazione R con $N_R = 3.200.000$ ennuple e (fra gli altri) quattro attributi, A, C, E, F ciascuno con $v = 20$ valori diversi e una lunghezza $a = 6$ byte

1. valutare lo spazio SBM (in blocchi) necessario per le foglie di un indice bitmap su A; valutare anche lo spazio S_I necessario per una tabella (una sorta di array in memoria secondaria, ad esempio) che permetta di associare un indirizzo a ciascun valore compreso fra 1 e N_R ; per confronto, calcolare lo spazio ST necessario per le foglie di un B⁺-tree (con riferimenti ai record) per un indice su A

Poi, valutare (sia nel caso di indici bitmap, sia nel caso di B⁺-tree) il costo (in termini di numero di accessi a memoria secondaria, supponendo che tutti i nodi intermedi di tutti gli indici siano nei buffer; basta quindi contare il numero di foglie da visitare e l'accesso alle ennuple, quando necessario) per le seguenti interrogazioni

2. `SELECT COUNT(*) FROM R WHERE A = 5`
3. `SELECT * FROM R WHERE A = 5`
4. `SELECT * FROM R WHERE A = 15 AND C = 14 AND E = 13 AND F = 9` (considerare i valori indipendenti l'uno dall'altro; quindi assumere che il risultato contenga circa N_R/v^4 ennuple)

Riportare le risposte nella tabella sottostante, nelle righe da 1 a 4, indicando formula e valore numerico.

| | | Risposte | |
|---|--|---------------|-----------------------------|
| | | Indici bitmap | Indici B ⁺ -tree |
| 1 | | | |
| 2 | | | |
| 3 | | | |
| 4 | | | |

Domanda 4 Specificare per ciascuno dei seguenti schedule, la sua appartenenza alle varie classi (scrivere “sì” o “no” nelle caselle della tabella; assumere che gli identificatori delle transazioni corrispondano ai timestamp).

| | Seriale | VSR | CSR | TS | 2PL |
|--|---------|-----|-----|----|-----|
| $r_2(x)w_2(x)r_1(x)w_1(x)$ | | | | | |
| $r_1(x)w_1(x)r_2(x)w_2(x)r_0(y)w_1(y)$ | | | | | |
| $r_1(x)r_2(z)w_2(z)w_1(x)r_2(x)w_2(x)$ | | | | | |
| $r_1(y)r_2(z)r_2(y)w_2(y)w_2(z)r_1(z)$ | | | | | |

Domanda 5 Considerare uno schema dimensionale relativo agli esami, che utilizzi, come tabella dei fatti e come una delle dimensioni, le relazioni come quelle qui schematizzate:

| <u>K</u> Studente | <u>K</u> Corso | <u>K</u> Data | Voto | ... |
|-------------------|----------------|---------------|------|-----|
| 201 | 301 | 405 | 26 | ... |
| 201 | 302 | 406 | 28 | ... |
| 202 | 301 | 407 | 30 | ... |
| 202 | 303 | 407 | 22 | ... |
| ... | ... | ... | ... | ... |

| <u>K</u> Corso | Titolo | Crediti | ... |
|----------------|-----------|---------|-----|
| 301 | Fisica I | 6 | ... |
| 302 | Chimica | 9 | ... |
| 303 | Geometria | 6 | ... |
| ... | ... | ... | ... |

Supporre che si presentino le seguenti esigenze di modifica:

- i corsi cambiano nome nel tempo: per esempio, il corso nella prima ennupla potrebbe ad un certo punto cambiare nome da “Fisica I” in “Meccanica”; interessano selezioni e aggregazioni relative agli esami tanto con riferimento al nome del corso (al momento dell’esame) quanto alla sua identità (un codice che viene introdotto allo scopo, ma non sempre viene utilizzato, perché alcuni analisti preferiscono fare riferimento al nome corrente del corso); le modifiche sono rare, ma è possibile che ci siano corsi con vari cambiamenti di nome;
- per ogni corso, interessa rappresentare anche il docente, per supportare analisi sugli esami svolti da ciascun docente; i docenti cambiano nel tempo e passano da un corso all’altro (e possono anche tenere più corsi nello stesso momento, ma ogni corso ha, in un certo giorno, un solo docente); è disponibile l’informazione relativa ai docenti dei corsi nel tempo (per tutto il periodo, anche passato, di interesse).

Mostrare nuove versioni delle due tabelle che permettano di soddisfare le esigenze sopra citate (mostrare anche i dati, con riferimento a quelli presenti negli esempi sopra, aggiungendo nuovi dati ragionevoli, che permettano di comprendere le modifiche).

Basi di dati II, primo modulo Prova parziale — 22 aprile 2010 — Compito C

Rispondere su questo fascicolo. Consegnare anche i fogli usati come brutta copia, che però non verranno corretti.
Tempo a disposizione: un'ora e trenta minuti.

Cognome _____ Nome _____ Matricola _____ Ordin. _____

Domanda 1 Considerare un disco con una velocità di rotazione di $v = 10.000$ giri al minuto e un tempo medio di posizionamento della testina (tempo di seek) $t_S = 5$ msec. Ogni traccia contiene $P = 240$ blocchi e ogni blocco contiene $B = 4$ KB. Rispondere alle seguenti domande mostrando formula e valore numerico numero (N.B. non servono calcolatrici, i risultati sono semplici, approssimare $1 \text{ MB} = 1000 \text{ KB}$).

1. Qual è tempo medio di latenza (attesa dovuta alla rotazione) t_L ?

2. Qual è la massima velocità di trasferimento (in MB al secondo)?

3. Qual è il tempo minimo t_B necessario per leggere un blocco?

4. Qual è il tempo medio necessario per leggere un blocco?

5. Qual è il tempo medio necessario per leggere $c = 40$ blocchi contigui?

Domanda 2 Una catena di supermercati ha una base di dati dei propri clienti che dispongono di una “tessera fedeltà,” con varie informazioni su ciascun cliente, fra cui (a) il totale dei punti acquisiti attraverso l’uso della tessera e (b) il negozio della catena cui fa riferimento (ad esempio, quello presso cui ha inizialmente richiesto la tessera). Si vuole eseguire su di essa l’interrogazione che calcola, per ciascun negozio, il numero dei clienti, la somma dei punti fedeltà dei clienti e la relativa media per cliente. Indicare quale livello di isolamento (READ UNCOMMITTED, READ COMMITTED, REPEATABLE READ o SERIALIZABLE) si potrebbe scegliere in ciascuno dei seguenti casi (si supponga che, in generale, sia stato rilevato che, nel corso degli inserimenti e delle modifiche, vengono inseriti valori sbagliati anche di vari ordini di grandezza, che sono poi corretti prima del commit):

1. L’operazione è eseguita mentre vengono inseriti alcuni nuovi clienti (per ciascun negozio pochi rispetto a quelli già presenti), con la finalità di acquisire informazioni approssimate ma ragionevolmente indicative sugli andamenti complessivi.
2. L’operazione è eseguita mentre vengono modificati i valori dei punti fedeltà di tutti i clienti (a seguito di una ridefinizione dei criteri di assegnazione dei punti stessi), con la finalità di acquisire informazioni approssimate ma ragionevolmente indicative sugli andamenti complessivi.
3. L’operazione è eseguita in un momento in cui non ci sono aggiornamenti di alcun genere, con la finalità di individuare i primi tre negozi da premiare in una campagna promozionale sulla base dei punti acquisiti dai rispettivi clienti.
4. L’operazione è eseguita mentre vengono inseriti alcuni nuovi clienti (per ciascun negozio pochi rispetto a quelli già presenti), con la finalità di individuare i primi tre negozi da premiare in una campagna promozionale sulla base dei punti acquisiti dai rispettivi clienti.
5. L’operazione è eseguita mentre vengono modificati i valori dei punti fedeltà di tutti i clienti (a seguito di una ridefinizione dei criteri di assegnazione dei punti stessi), con la finalità di individuare i primi tre negozi da premiare in una campagna promozionale sulla base dei punti acquisiti dai rispettivi clienti.

| Risposte | | | | |
|----------|----|----|----|----|
| 1. | 2. | 3. | 4. | 5. |
| | | | | |

Domanda 3 Molti DBMS utilizzano, oltre a quelli basati sui B⁺-tree, anche indici basati su una tecnica detta "bitmap." Un indice bitmap, su un attributo *A* con *v* valori diversi, per una relazione con N_R ennuple, è costituito da *v* vettori (uno per ciascun valore di *A*) di N_R bit ciascuno (un bit per ogni ennupla della relazione): l'*i*-esimo bit del vettore associato al valore a_j è 1 se il valore della *i*-esima ennupla sull'attributo *A* è a_j e 0 se il valore è diverso. L'accesso ai vettori è organizzato per mezzo di un albero. In sostanza, rispetto ad un B⁺-tree, abbiamo una struttura di accesso simile, ma con foglie diverse: nel B⁺-tree, per ogni valore dell'attributo abbiamo una lista di indirizzi, mentre nell'indice bitmap abbiamo un vettore di bit. Esiste poi una struttura (tabella, indice o altro) che associa gli indirizzi delle ennuple ai numeri da 1 a N_R . Sulla base degli elementi forniti, con riferimento ad un sistema con blocchi di dimensione $B = 8$ KB e indirizzi dei record di lunghezza $p = 8$ B e ad una relazione R con $N_R = 3.200.000$ ennuple e (fra gli altri) quattro attributi, A, C, E, F ciascuno con $v = 20$ valori diversi e una lunghezza $a = 6$ byte

1. valutare lo spazio SBM (in blocchi) necessario per le foglie di un indice bitmap su A; valutare anche lo spazio S_I necessario per una tabella (una sorta di array in memoria secondaria, ad esempio) che permetta di associare un indirizzo a ciascun valore compreso fra 1 e N_R ; per confronto, calcolare lo spazio ST necessario per le foglie di un B⁺-tree (con riferimenti ai record) per un indice su A

Poi, valutare (sia nel caso di indici bitmap, sia nel caso di B⁺-tree) il costo (in termini di numero di accessi a memoria secondaria, supponendo che tutti i nodi intermedi di tutti gli indici siano nei buffer; basta quindi contare il numero di foglie da visitare e l'accesso alle ennuple, quando necessario) per le seguenti interrogazioni

2. `SELECT COUNT(*) FROM R WHERE A = 5`
3. `SELECT * FROM R WHERE A = 5`
4. `SELECT * FROM R WHERE A = 15 AND C = 14 AND E = 13 AND F = 9` (considerare i valori indipendenti l'uno dall'altro; quindi assumere che il risultato contenga circa N_R/v^4 ennuple)

Riportare le risposte nella tabella sottostante, nelle righe da 1 a 4, indicando formula e valore numerico.

| | | Risposte | |
|---|--|---------------|-----------------------------|
| | | Indici bitmap | Indici B ⁺ -tree |
| 1 | | | |
| 2 | | | |
| 3 | | | |
| 4 | | | |

Domanda 4 Specificare per ciascuno dei seguenti schedule, la sua appartenenza alle varie classi (scrivere “sì” o “no” nelle caselle della tabella; assumere che gli identificatori delle transazioni corrispondano ai timestamp).

| | Seriale | VSR | CSR | TS | 2PL |
|--|---------|-----|-----|----|-----|
| $r_1(x)r_2(z)w_2(z)w_1(x)r_2(x)w_2(x)$ | | | | | |
| $r_1(y)r_2(z)r_2(y)w_2(y)w_2(z)r_1(z)$ | | | | | |
| $r_2(x)w_2(x)r_1(x)w_1(x)$ | | | | | |
| $r_1(x)w_1(x)r_2(x)w_2(x)r_0(y)w_1(y)$ | | | | | |

Domanda 5 Considerare uno schema dimensionale relativo agli esami, che utilizzi, come tabella dei fatti e come una delle dimensioni, le relazioni come quelle qui schematizzate:

| KStudente | KCorso | KData | Voto | ... |
|-----------|--------|-------|------|-----|
| 401 | 201 | 305 | 27 | ... |
| 401 | 202 | 306 | 28 | ... |
| 402 | 201 | 307 | 30 | ... |
| 402 | 203 | 307 | 22 | ... |
| ... | ... | ... | ... | ... |

| KCorso | Titolo | Crediti | ... |
|--------|-----------|---------|-----|
| 201 | Fisica I | 6 | ... |
| 202 | Chimica | 9 | ... |
| 203 | Geometria | 6 | ... |
| ... | ... | ... | ... |

Supporre che si presentino le seguenti esigenze di modifica:

- i corsi cambiano nome nel tempo: per esempio, il corso nella prima ennupla potrebbe ad un certo punto cambiare nome da “Fisica I” in “Meccanica”; interessano selezioni e aggregazioni relative agli esami tanto con riferimento al nome del corso (al momento dell’esame) quanto alla sua identità (un codice che viene introdotto allo scopo, ma non sempre viene utilizzato, perché alcuni analisti preferiscono fare riferimento al nome corrente del corso); le modifiche sono rare, ma è possibile che ci siano corsi con vari cambiamenti di nome;
- per ogni corso, interessa rappresentare anche il docente, per supportare analisi sugli esami svolti da ciascun docente; i docenti cambiano nel tempo e passano da un corso all’altro (e possono anche tenere più corsi nello stesso momento, ma ogni corso ha, in un certo giorno, un solo docente); è disponibile l’informazione relativa ai docenti dei corsi nel tempo (per tutto il periodo, anche passato, di interesse).

Mostrare nuove versioni delle due tabelle che permettano di soddisfare le esigenze sopra citate (mostrare anche i dati, con riferimento a quelli presenti negli esempi sopra, aggiungendo nuovi dati ragionevoli, che permettano di comprendere le modifiche).

Basi di dati II, primo modulo Prova parziale — 22 aprile 2010 — Compito D

Rispondere su questo fascicolo. Consegnare anche i fogli usati come brutta copia, che però non verranno corretti.
Tempo a disposizione: un'ora e trenta minuti.

Cognome _____ Nome _____ Matricola _____ Ordin. _____

Domanda 1 Considerare un disco con una velocità di rotazione di $r = 10.000$ giri al minuto e un tempo medio di posizionamento della testina (tempo di seek) $t_S = 5$ msec. Ogni traccia contiene $N = 300$ blocchi e ogni blocco contiene $B = 4$ KB. Rispondere alle seguenti domande mostrando formula e valore numerico numero (N.B. non servono calcolatrici, i risultati sono semplici, approssimare $1 \text{ MB} = 1000 \text{ KB}$).

1. Qual è tempo medio di latenza (attesa dovuta alla rotazione) t_L ?

2. Qual è la massima velocità di trasferimento (in MB al secondo)?

3. Qual è il tempo minimo t_B necessario per leggere un blocco?

4. Qual è il tempo medio necessario per leggere un blocco?

5. Qual è il tempo medio necessario per leggere $n = 50$ blocchi contigui?

Domanda 2 Una catena di supermercati ha una base di dati dei propri clienti che dispongono di una “tessera fedeltà,” con varie informazioni su ciascun cliente, fra cui (a) il totale dei punti acquisiti attraverso l’uso della tessera e (b) il negozio della catena cui fa riferimento (ad esempio, quello presso cui ha inizialmente richiesto la tessera). Si vuole eseguire su di essa l’interrogazione che calcola, per ciascun negozio, il numero dei clienti, la somma dei punti fedeltà dei clienti e la relativa media per cliente. Indicare quale livello di isolamento (READ UNCOMMITTED, READ COMMITTED, REPEATABLE READ o SERIALIZABLE) si potrebbe scegliere in ciascuno dei seguenti casi (si supponga che, in generale, sia stato rilevato che, nel corso degli inserimenti e delle modifiche, vengono inseriti valori sbagliati anche di vari ordini di grandezza, che sono poi corretti prima del commit):

1. L’operazione è eseguita mentre vengono modificati i valori dei punti fedeltà di alcuni clienti (in ciascun negozio una piccola percentuale), con la finalità di acquisire informazioni approssimate ma ragionevolmente indicative sugli andamenti complessivi.
2. L’operazione è eseguita mentre vengono modificati i valori dei punti fedeltà di tutti i clienti (a seguito di una ridefinizione dei criteri di assegnazione dei punti stessi), con la finalità di acquisire informazioni approssimate ma ragionevolmente indicative sugli andamenti complessivi.
3. L’operazione è eseguita mentre vengono inseriti molti nuovi clienti, con la finalità di individuare i primi tre negozi da premiare in una campagna promozionale sulla base dei punti acquisiti dai rispettivi clienti.
4. L’operazione è eseguita mentre vengono modificati i valori dei punti fedeltà di alcuni clienti (in ciascun negozio una piccola percentuale), con la finalità di individuare i primi tre negozi da premiare in una campagna promozionale sulla base dei punti acquisiti dai rispettivi clienti.
5. L’operazione è eseguita in un momento in cui non ci sono aggiornamenti di alcun genere, con la finalità di acquisire informazioni approssimate ma ragionevolmente indicative sugli andamenti complessivi.

| Risposte | | | | |
|----------|----|----|----|----|
| 1. | 2. | 3. | 4. | 5. |
| | | | | |

Domanda 3 Molti DBMS utilizzano, oltre a quelli basati sui B⁺-tree, anche indici basati su una tecnica detta “bitmap.” Un indice bitmap, su un attributo *A* con *v* valori diversi, per una relazione con N_R ennuple, è costituito da *v* vettori (uno per ciascun valore di *A*) di N_R bit ciascuno (un bit per ogni ennupla della relazione): l’*i*-esimo bit del vettore associato al valore a_j è 1 se il valore della *i*-esima ennupla sull’attributo *A* è a_j e 0 se il valore è diverso. L’accesso ai vettori è organizzato per mezzo di un albero. In sostanza, rispetto ad un B⁺-tree, abbiamo una struttura di accesso simile, ma con foglie diverse: nel B⁺-tree, per ogni valore dell’attributo abbiamo una lista di indirizzi, mentre nell’indice bitmap abbiamo un vettore di bit. Esiste poi una struttura (tabella, indice o altro) che associa gli indirizzi delle ennuple ai numeri da 1 a N_R . Sulla base degli elementi forniti, con riferimento ad un sistema con blocchi di dimensione $D = 4$ KB e indirizzi dei record di lunghezza $p = 8$ B e ad una relazione R con $N_R = 1.600.000$ ennuple e (fra gli altri) quattro attributi, A, C, E, F ciascuno con $v = 20$ valori diversi e una lunghezza $a = 6$ byte

1. valutare lo spazio SBM (in blocchi) necessario per le foglie di un indice bitmap su A; valutare anche lo spazio S_I necessario per una tabella (una sorta di array in memoria secondaria, ad esempio) che permetta di associare un indirizzo a ciascun valore compreso fra 1 e N_R ; per confronto, calcolare lo spazio ST necessario per le foglie di un B⁺-tree (con riferimenti ai record) per un indice su A

Poi, valutare (sia nel caso di indici bitmap, sia nel caso di B⁺-tree) il costo (in termini di numero di accessi a memoria secondaria, supponendo che tutti i nodi intermedi di tutti gli indici siano nei buffer; basta quindi contare il numero di foglie da visitare e l’accesso alle ennuple, quando necessario) per le seguenti interrogazioni

2. `SELECT COUNT(*) FROM R WHERE A = 5`
3. `SELECT * FROM R WHERE A = 5`
4. `SELECT * FROM R WHERE A = 15 AND C = 14 AND E = 13 AND F = 9` (considerare i valori indipendenti l’uno dall’altro; quindi assumere che il risultato contenga circa N_R/v^4 ennuple)

Riportare le risposte nella tabella sottostante, nelle righe da 1 a 4, indicando formula e valore numerico.

| | | Risposte | |
|---|--|---------------|-----------------------------|
| | | Indici bitmap | Indici B ⁺ -tree |
| 1 | | | |
| 2 | | | |
| 3 | | | |
| 4 | | | |

Domanda 4 Specificare per ciascuno dei seguenti schedule, la sua appartenenza alle varie classi (scrivere “sì” o “no” nelle caselle della tabella; assumere che gli identificatori delle transazioni corrispondano ai timestamp).

| | Seriale | VSR | CSR | TS | 2PL |
|--|---------|-----|-----|----|-----|
| $r_2(x)w_2(x)r_1(x)w_1(x)$ | | | | | |
| $r_1(y)r_2(z)r_2(y)w_2(y)w_2(z)r_1(z)$ | | | | | |
| $r_1(x)w_1(x)r_2(x)w_2(x)r_0(y)w_1(y)$ | | | | | |
| $r_1(x)r_2(z)w_2(z)w_1(x)r_2(x)w_2(x)$ | | | | | |

Domanda 5 Considerare uno schema dimensionale relativo agli esami, che utilizzi, come tabella dei fatti e come una delle dimensioni, le relazioni come quelle qui schematizzate:

| KStudente | KCorso | KData | Voto | ... |
|-----------|--------|-------|------|-----|
| 401 | 301 | 105 | 28 | ... |
| 401 | 302 | 106 | 28 | ... |
| 402 | 301 | 107 | 30 | ... |
| 402 | 303 | 107 | 22 | ... |
| ... | ... | ... | ... | ... |

| KCorso | Titolo | Crediti | ... |
|--------|-----------|---------|-----|
| 301 | Fisica I | 6 | ... |
| 302 | Chimica | 9 | ... |
| 303 | Geometria | 6 | ... |
| ... | ... | ... | ... |

Supporre che si presentino le seguenti esigenze di modifica:

- i corsi cambiano nome nel tempo: per esempio, il corso nella prima ennupla potrebbe ad un certo punto cambiare nome da “Fisica I” in “Meccanica”; interessano selezioni e aggregazioni relative agli esami tanto con riferimento al nome del corso (al momento dell’esame) quanto alla sua identità (un codice che viene introdotto allo scopo, ma non sempre viene utilizzato, perché alcuni analisti preferiscono fare riferimento al nome corrente del corso); le modifiche sono rare, ma è possibile che ci siano corsi con vari cambiamenti di nome;
- per ogni corso, interessa rappresentare anche il docente, per supportare analisi sugli esami svolti da ciascun docente; i docenti cambiano nel tempo e passano da un corso all’altro (e possono anche tenere più corsi nello stesso momento, ma ogni corso ha, in un certo giorno, un solo docente); è disponibile l’informazione relativa ai docenti dei corsi nel tempo (per tutto il periodo, anche passato, di interesse).

Mostrare nuove versioni delle due tabelle che permettano di soddisfare le esigenze sopra citate (mostrare anche i dati, con riferimento a quelli presenti negli esempi sopra, aggiungendo nuovi dati ragionevoli, che permettano di comprendere le modifiche).