

Corso di Basi di dati
Prova scritta parziale — 31 marzo 2001
Compito A — Soluzioni (cenni)

Nota: le soluzioni per il compito B sono molto simili

Domanda 1

- 1(a)** un guasto del coordinatore (TM) può avere conseguenze molto pesanti anche sulle prestazioni dei partecipanti (RM) perché questi ultimi potrebbero avere transazioni in stato di “ready” (cioè nella “finestra di incertezza”), con i lock acquisiti sulle risorse coinvolte, in attesa di risposta da parte del TM;
- 1(b)** un guasto di un partecipante non ha conseguenze particolari sulle prestazioni degli altri partecipanti (a parte l’eventuale abort di transazioni), perché:
- se il guasto avviene prima del ready, allora il partecipante stesso non risponde al prepare e quindi, scattato il time-out della prima fase, il coordinatore decide per l’abort;
 - se il guasto avviene dopo il ready, l’unica complicazione è la necessità di ripetere la seconda fase
- 2(a)** se riceve almeno una risposta che contenga una decisione, può basarsi su di essa: se riceve almeno una risposta “commit,” può assumere che la transazione sia andata a buon fine, se riceve almeno una risposta “abort,” può assumere che essa sia fallita;
- 2(b)** non è possibile che riceva “abort” da un partecipante e “commit” da un altro; né che riceva “prima-del-ready” da uno e “commit” da un altro.

Domanda 2 Notazione:

- B dimensione del blocco
- l, l_p, l_s lunghezza dei record del file, dell’indice primario e di quello secondario
- N numero di record del file; $N_{p,1}, N_{p,2}, \dots$ numero di record dell’indice primario ai vari livelli; $N_{s,1}, N_{s,2}, \dots$ numero di record dell’indice secondario ai vari livelli;
- L numero di blocchi del file; $L_{p,1}, L_{p,2}, \dots$ numero di blocchi dell’indice primario ai vari livelli; $L_{s,1}, L_{s,2}, \dots$ numero di blocchi dell’indice secondario ai vari livelli;
- f, f_p, f_s fattore di blocco del file, dell’indice primario e di quello secondario

Procedendo sempre in maniera approssimata:

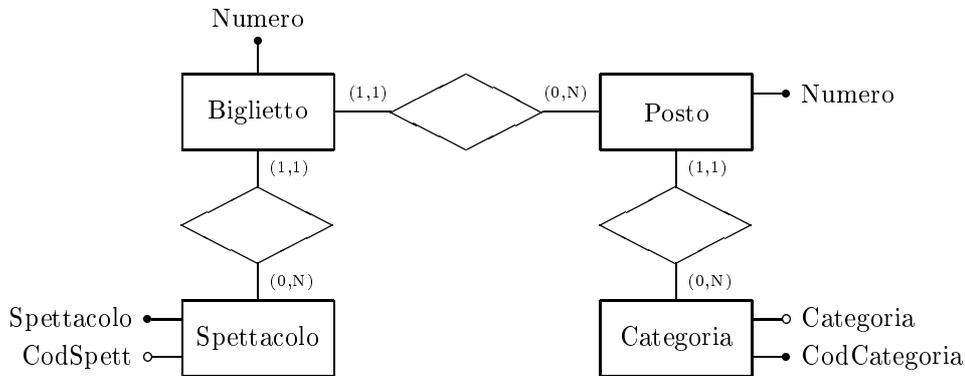
1. il fattore di blocco del file è $f = B/l = 1000/110 = 9$; quindi il file occupa circa $L = N/f = 200.000/9 = 22.000$ blocchi
2. i record dell’indice primario occupano $l_p = 20 + 5 = 25$ byte; quindi il fattore di blocco dell’indice primario è $f_p = B/l_p = 1000/25 = 40$; l’indice può essere sparso, quindi con $N_{p,1} = L = 22.000$ record e quindi il livello più basso occupa circa $L_{p,1} = L/f_p = 22.000/40 = 550$ blocchi; per i livelli successivi si ha $N_{p,i+1} = L_{p,i}$ e $L_{p,i+1} = N_{p,i+1}/f_p = L_{p,i}/f_p$; quindi $L_{p,2} = L_{p,1}/f_p = 550/40 = 14$, $L_{p,3} = L_{p,2}/f_p = 14/40 = 1$. In totale: $L_p = L_{p,1} + L_{p,2} + L_{p,3} = 550 + 14 + 1 = 565$
3. i record dell’indice secondario (riferendoci per semplicità ad un B+-tree) occupano $l_s = 6 + 5 = 11$ byte; quindi il fattore di blocco massimo dell’indice secondario è $f_s = B/l_s = 1000/11 = 90$; però, trattandosi di un B-tree, dobbiamo assumere un riempimento parziale, circa il 70% e quindi possiamo ipotizzare $f_s = 0,7 \times B/l_s = 0,7 \times 1000/11 = 60$; l’indice deve essere denso, quindi con $N_{s,1} = N = 200.000$ record e quindi il livello più basso occupa circa $L_{s,1} = N/f_s = 200.000/60 = 3.300$ blocchi; per i livelli successivi si ha $N_{s,i+1} = L_{s,i}$ e $L_{s,i+1} = N_{s,i+1}/f_s = L_{s,i}/f_s$; quindi $L_{s,2} = L_{s,1}/f_s = 3.300/60 = 50$, $L_{s,3} = L_{s,2}/f_s = 50/60 = 1$. In totale: $L_s = L_{s,1} + L_{s,2} + L_{s,3} = 3.300 + 50 + 1 = 3.350$

Domanda 3

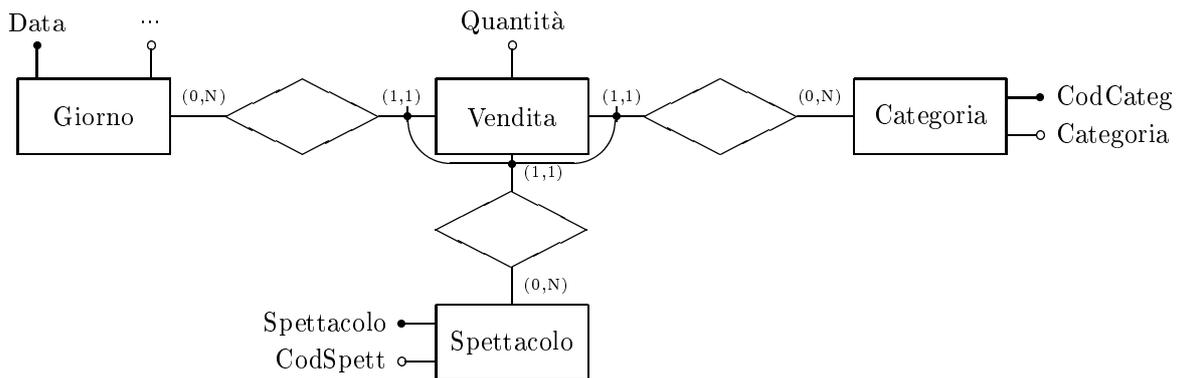
1. la ricerca utilizza l'indice secondario che ha tre livelli e quindi si può assumere che il costo di un'esecuzione dell'operazione sia $c_1 = 4$ (tre blocchi per l'indice più uno per il record vero e proprio); quindi nell'unità di tempo avremo un numero di accessi pari a $c_1 \times f_1 = 4 \times 1000 = 4000$
2. la ricerca utilizza l'indice primario, che ha anch'esso tre livelli; ciascuna ricerca richiede mediamente 25 record e quindi l'accesso a 3 blocchi del file; quindi $c_2 = 6$ e nell'unità di tempo $c_2 \times f_2 = 6 \times 200 = 1200$
3. non cambia niente sulla prima operazione (supponendo che l'indice secondario non sia alterato) ma porta ad un appesantimento la seconda, che risulta realizzabile solo con una scansione sequenziale (una struttura hash non permette, se non in casi molto particolari, ricerche di sottostringhe); c_2 diventerebbe in questo caso pari a 22.000
4. la seconda operazione non subisce modifiche, mentre per la prima si può pensare ad un costo c_2 di poco superiore a 2 (mediamente), in quanto sulla struttura secondaria è mediamente necessario poco più di un accesso cui deve essere aggiunto l'accesso al blocco effettivo del file.

Domanda 4

1 Schema concettuale della base di dati



2(a) Schema concettuale del data warehouse



2(b) Fatto: la vendita di biglietti

Misura: la quantità di biglietti venduti

Dimensioni: il tempo; il tipo di spettacolo; la categoria di posto

2(c) $Vendita(\underline{Data}, \underline{CodSpett}, \underline{CodCateg}, Quantità)$

$Tempo(\underline{Data}, \dots)$

$Spettacolo(\underline{CodSpett}, Spettacolo)$

$Tempo(\underline{CodCateg}, Categoria)$