

BASI DI DATI II, 26 gennaio 2021
(Svolto a distanza)

Domanda 1

Si considerino due relazioni $R_1(A, C, D, E)$, $R_2(A, F, G, H)$, in cui gli attributi hanno tutti la stessa dimensione $L = 10$ Byte, molto più piccola della dimensione del blocco pari a $B = 8000$ Byte. Si supponga che

- l'attributo A sia chiave per entrambe le relazioni,
- i valori per A nelle due relazioni siano gli stessi
- le relazioni abbiano entrambe $N = 1.000.000$ ennuple
- ciascuna relazione abbia un indice primario su A (e quindi sia ordinata su A)

e che le operazioni più frequenti su di essa siano le seguenti:

- o1: join delle due relazioni su A, con frequenza $f_1 = 10.000$;
- o2: scansione dell'intera relazione R_1 , con frequenza $f_2 = 100$

Valutare le due seguenti alternative di memorizzazione, calcolando il costo complessivo (riportare la formula che indica il numero di accessi nell'unità di tempo, in base alle variabili sopra citate, e poi il valore numerico):

- A: memorizzazione separata delle due relazioni
- B: memorizzazione delle due relazioni in un cluster plurirelazionale

Rispondere a video, commentando brevemente la soluzione. Spiegare anche, senza entrare nei dettagli, come cambierebbe la situazione se le frequenze fossero diverse, con f_2 molto maggiore di f_1 (anziché molto minore).

Domanda 2

Si consideri la seguente porzione dello schema dell'archivio delle carriere degli studenti di una anagrafe ministeriale:

- Studenti(CodiceFiscale, Cognome, Nome, DataNascita, TipoMaturità)
- Iscrizioni(CodiceFiscale, AnnoAccademico, CorsoDiStudio, AnnoDiCorso, AnnodiImmatricolazione)
- CorsiDiStudio(CodiceCdS, Titolo, Livello, Classe, Università)
- Lauree(CodiceFiscale, CodiceCdS, Data, Voto)

Le chiavi delle relazioni sono le seguenti

- CodiceFiscale per Studenti
- CodiceFiscale, AnnoAccademico per Iscrizioni
- CodiceCdS per CorsiDiStudio
- CodiceFiscale, CodiceCdS per Lauree

Progettare uno schema dimensionale che permetta di rispondere, fra le altre, alle seguenti interrogazioni:

- calcolare il numero di studenti (con la relativa media dei voti e l'età media) che si sono laureati in un certo corso di studio (inteso come corso di studio presso una università) in un certo anno accademico (si supponga che, per la data di laurea, l'unico dettaglio rilevante sia l'anno accademico e che esista un modo univoco per associare un anno accademico ad una data di laurea)
- calcolare il numero di laureati per una classe di corsi studio, distinto per tipo di maturità e per numero di anni impiegati per conseguire il titolo (ad esempio, 2, 3, 4, 5, 6, più di 6)
- calcolare il numero di laureati per una classe di corsi studio, distinto per "età alla laurea" (ad esempio, 21, 22, 23, ...26, più di 26)

Assumere che, per ragioni di privatezza e di compattezza, sia opportuno limitare la cardinalità della tabella dei fatti, a patto di permettere la risposta alle precedenti interrogazioni.

Specificare (discorsivamente ma con precisione) la grana scelta e mostrare lo schema dimensionale (con gli attributi essenziali).

Descrivere anche, sempre discorsivamente, come si può generare il contenuto della tabella dei fatti a partire dalle relazioni sopra mostrate.

Domanda 3 (25%)

Considerare una relazione **conti** (**numero**, **saldo**) e il seguente scenario in cui due client inviano richieste ad un gestore del controllo di concorrenza. Ciascun client può inviare una richiesta solo dopo che è stata eseguita o rifiutata la precedente (se invece una richiesta viene bloccata da un lock, allora il client rimane inattivo fino alla concessione o allo scadere del timeout). Si supponga che, in caso di stallo, abortisca la transazione che ha avanzato la richiesta per prima.

T1	T2
<pre> start transaction isolation level ...; select * from conti where numero = 3; update conti set saldo = 20 where numero = 3; commit; </pre>	<pre> start transaction isolation level ...; select * from conti where numero = 3; update conti set saldo = 10 where numero = 3; commit; </pre>

Considerare uno scheduler con controllo di concorrenza basato su **Multiversioni** (come in Postgres). Per ciascuno dei tre possibili valori per il livello di isolamento (**read committed**, **repeatable read**, **serializable**), indicare: l'esito (**commit** o **abort**) di ciascuna delle due transazioni e l'ordine in cui i due eventi si verificano. Rispondere su tre righe, in ogni riga indicare il livello di isolamento cui ci si riferisce e poi le due azioni, nell'ordine in cui si verificano (ad esempio *livello* **commit(T2) abort(T1)**). Si possono aggiungere eventualmente brevi commenti.

Basi di dati II — 26 gennaio 2021

Domanda 4 (25%)

Considerare un sistema distribuito su cui viene eseguita una transazione T che coinvolge un coordinatore N1 e due partecipanti N2 e N3. Come schematizzato sotto, coordinatore N1 invia il messaggio di **prepare** all'istante t_0 ; i due partecipanti hanno il tempo di rispondere entro l'istante t_1 e poi uno dei due, N2, va in crash. Il coordinatore ha il tempo di prendere la decisione all'istante t_2 e di ricevere una risposta da N3 all'istante t_3 , ma va in crash poco dopo. Supporre che N1 e N2 siano ripristinati abbastanza presto.

	Nodo N1		Nodo N2		Nodo N3	
	Log	Messaggi	Log	Messaggi	Log	Messaggi
t_0	prepare(T,N2,N3)	prepare(T) → N2,N3				
t_1						
t_2				crash		
t_3		crash		restart		
		restart				

Supponendo che il coordinatore non scriva nel log quali messaggi di ack ha ricevuto, indicare quali record scrivono i vari nodi (coordinatore e partecipanti) nei rispettivi log e quali messaggi inviano. Cercare di specificare l'ordine in cui le varie azioni avvengono