

Basi di dati II — Prova parziale — 29 maggio 2014 — Compito A

Tempo a disposizione: un'ora e trenta minuti.

Cognome _____ Nome _____ Matricola _____

Domanda 1 (20%) Considerare un sistema distribuito su cui viene eseguita una transazione che coinvolge tre nodi, un coordinatore C e due partecipanti P1 e P2. Dopo la richiesta di **prepare** da parte del coordinatore C, i due partecipanti ricevono e rispondono correttamente, e uno dei due, P1, va in *crash* subito dopo aver risposto. Il coordinatore riceve le risposte, prende la decisione, invia i relativi messaggi e subito dopo va anch'esso in *crash* (quindi senza fare in tempo a ricevere le conferme). Indicare, nello schema sottostante, una possibile sequenza di scritture sui log e invio di messaggi, supponendo che entrambi i nodi siano ripristinati abbastanza presto (ma che il coordinatore perda i messaggi di risposta inviati ad esso a seguito del commit). Per semplicità, si fa riferimento ad una sola transazione e quindi non c'è bisogno di indicarla. Per i messaggi si usi la notazione *tipo*→*destinatari* (come nell'esempio: **prepare**→P1,P2). Supporre che nel log del coordinatore si scrivano solo i record di **prepare**, **commit** e **complete**, mentre i messaggi sono gestiti in memoria.

Nodo C		Nodo P1		Nodo P2	
Log	Messaggi	Log	Messaggi	Log	Messaggi
prepare(P1,P2)	prepare→P1,P2		<i>crash</i>		
<i>crash</i>					
<i>restart</i>			<i>restart</i>		

Eventuali commenti:

Basi di dati II — 29 maggio 2014 — Compito A

Domanda 2 (40%) Considerare uno schema dimensionale relativo a presenze e reti di giocatori di calcio, che utilizzi, come tabella dei fatti e come una delle dimensioni, relazioni come quelle qui schematizzate:

<u>KGiocatore</u>	<u>KSquadra</u>	<u>KTorneo</u>	<u>KStagione</u>	Presenze	Reti
301	201	401	21
301	202	402	28
302	201	401	30
302	203	403	22
...

<u>KTorneo</u>	Nome	...
401	Serie A	...
402	Coppa dei Campioni	...
403	Coppa Italia	...
...

Supporre che si presentino le seguenti esigenze di modifica:

- I tornei possono cambiare nome nel tempo: per esempio, il torneo nella prima enupla potrebbe ad un certo punto cambiare nome da “Coppa dei Campioni” in “Champions League”; interessano selezioni e aggregazioni relative alle presenze e alle reti tanto con riferimento al nome del torneo (distinguendo quindi la “Coppa dei Campioni” dalla “Champions League”) quanto alla sua identità (“Coppa dei Campioni” e “Champions League” in questo senso avrebbero la stessa identità; allo scopo si usa un codice, ma gli analisti talvolta preferiscono usare il nome attuale del torneo, quindi nell’esempio “Champions League” anche per le vecchie edizioni). Le modifiche sono rare, ma è possibile che ci siano tornei con vari cambiamenti di nome. Mostrare modifiche alle relazioni (una o entrambe) che permettano di soddisfare questa esigenza (mostrare i dati, con riferimento a quelli presenti negli esempi sopra, aggiungendo nuovi dati ragionevoli, che permettano di comprendere le modifiche).

Basi di dati II — 29 maggio 2014 — Compito A

- Per ogni giocatore, interessa rappresentare anche l'allenatore sotto la guida del quale ha giocato (per sommare ad esempio presenze e reti segnate con un certo allenatore); supporre (anche se l'ipotesi non è realistica) che ogni squadra abbia, in ciascuna stagione, un solo allenatore e che siano disponibili, nella staging area, tutti i dati relativi agli allenatori nelle varie stagioni. Anche in questo caso, mostrare le modifiche da apportare a una o a entrambe le relazioni.

- Con riferimento al punto precedente, considerare invece il caso, più realistico, in cui gli allenatori possono cambiare nel corso della stagione. Indicare quali modifiche sarebbero necessarie allo schema dimensionale e, soprattutto, quali dati sarebbero necessari nella staging area per poter riorganizzare il tutto.

Basi di dati II — 29 maggio 2014 — Compito A

Domanda 3 (20%) Supporre di voler organizzare una riunione in una certa data, seguendo un protocollo simile al commit a due fasi, procedendo attraverso chiamate telefoniche ai partecipanti, nel modo seguente:

- il coordinatore telefona a ciascuno dei partecipanti, chiedendo la disponibilità; se tutti dichiarano di essere disponibili, decide di svolgere la riunione, altrimenti no
- dopo la decisione, il coordinatore telefona di nuovo a tutti coloro che hanno dato disponibilità, per informarli sulla decisione.

Rispondere alle domande negli spazi sottostanti.

Quale potrebbe essere un punto debole della procedura (pensare al caso in cui per partecipare alla riunione i partecipanti debbano organizzare un viaggio)? Quale condizione sarebbe necessaria per garantire l'efficacia?

È sufficiente ritelefonare solo a coloro che hanno dato disponibilità? Perché?

Domanda 4 (20%) Per ciascuno degli schedule sotto riportati, indicare, scrivendo **sì** o **no** in ciascuna casella, a quali classi appartiene: S (seriale, rispetto a letture e scritture, ignorare commit e abort), CSR (conflict-serializzabile), S2PL (cioè generabile da uno scheduler basato su 2PL stretto), MV (cioè generabile da uno scheduler multiversion con controllo di serializzabilità: “a serializable transaction cannot modify or lock rows changed by other transactions after the serializable transaction began”). Negli schedule, s_i indica l'inizio della transazione i e c_i il suo commit.

	S	CSR	S2PL	MV
$s_1, r_1(x), w_1(x), s_2, r_2(x), w_2(x), c_2, s_3, r_3(y), w_3(y), c_3, r_1(y), w_1(y), c_1$				
$s_2, r_2(x), w_2(x), c_2, s_1, r_1(x), w_1(x), c_1$				
$s_1, r_1(x), s_2, r_2(x), w_1(x), c_1, w_2(x), c_2$				
$s_1, r_1(x), s_2, r_2(x), w_2(x), r_2(y), w_2(y), c_2, r_1(y), c_1$				

Basi di dati II — Prova parziale — 29 maggio 2014 — Compito A

Cenni sulle soluzioni (solo Compito A, le varianti del testo sono in rosso)

Tempo a disposizione: un'ora e trenta minuti.

Cognome _____ Nome _____ Matricola _____

Domanda 1 (20%) Considerare un sistema distribuito su cui viene eseguita una transazione che coinvolge tre nodi, un coordinatore **C** e due partecipanti **P1** e **P2**. Dopo la richiesta di **prepare** da parte del coordinatore **C**, i due partecipanti ricevono e rispondono correttamente, e uno dei due, **P1**, va in crash subito dopo aver risposto. Il coordinatore riceve le risposte, prende la decisione, invia i relativi messaggi e subito dopo va anch'esso in crash (quindi senza fare in tempo a ricevere le conferme). Indicare, nello schema sottostante, una possibile sequenza di scritture sui log e invio di messaggi, supponendo che entrambi i nodi siano ripristinati abbastanza presto (ma che il coordinatore perda i messaggi di risposta inviati ad esso a seguito del commit). Per semplicità, si fa riferimento ad una sola transazione e quindi non c'è bisogno di indicarla. Per i messaggi si usi la notazione *tipo*→*destinatari* (come nell'esempio: **prepare**→**P1,P2**). Supporre che nel log del coordinatore si scrivano solo i record di **prepare**, **commit** e **complete**, mentre i messaggi sono gestiti in memoria.

Nodo C		Nodo P1		Nodo P2	
Log	Messaggi	Log	Messaggi	Log	Messaggi
prepare(P1,P2)	prepare→ P1,P2	ready	ready→ C	ready	ready→ C
		<i>crash</i>			
commit	commit→ P1,P2			commit	ack→ C
	<i>crash</i>				
	<i>restart</i>				ack→ C
	commit→ P1,P2		<i>restart</i>		
	commit→ P1	commit			
complete			ack→ C		

Eventuali commenti:

Basi di dati II — 29 maggio 2014 — Compito A

Domanda 2 (40%) Considerare uno schema dimensionale relativo a presenze e reti di giocatori di calcio, che utilizzi, come tabella dei fatti e come una delle dimensioni, relazioni come quelle qui schematizzate:

<u>KGiocatore</u>	<u>KSquadra</u>	<u>KTorneo</u>	<u>KStagione</u>	Presenze	Reti
301	201	401	21
301	202	402	28
302	201	401	30
302	203	403	22
...

<u>KTorneo</u>	Nome	...
401	Serie A	...
402	Coppa dei Campioni	...
403	Coppa Italia	...
...

Supporre che si presentino le seguenti esigenze di modifica:

- I tornei possono cambiare nome nel tempo: per esempio, il torneo nella prima ennupla potrebbe ad un certo punto cambiare nome da “Coppa dei Campioni” in “Champions League”; interessano selezioni e aggregazioni relative alle presenze e alle reti tanto con riferimento al nome del torneo (distinguendo quindi la “Coppa dei Campioni” dalla “Champions League”) quanto alla sua identità (“Coppa dei Campioni” e “Champions League” in questo senso avrebbero la stessa identità; allo scopo si usa un codice, ma gli analisti talvolta preferiscono usare il nome attuale del torneo, quindi nell’esempio “Champions League” anche per le vecchie edizioni). Le modifiche sono rare, ma è possibile che ci siano tornei con vari cambiamenti di nome. Mostrare modifiche alle relazioni (una o entrambe) che permettano di soddisfare questa esigenza (mostrare i dati, con riferimento a quelli presenti negli esempi sopra, aggiungendo nuovi dati ragionevoli, che permettano di comprendere le modifiche).

La dimensione Torneo va gestita come una “slowly changing dimension”, con un doppio accorgimento:

- va aggiunta una ennupla per ogni nuova versione di torneo
- vanno aggiunti un attributo con il nome attuale anche per la versione precedente e uno con un codice

<u>KTorneo</u>	Codice	NomeAlMomento	NomeAttuale	...
401	T1	Serie A	Seria A	...
402	T2	Coppa dei Campioni	Champions League	...
403	T3	Coppa Italia	Coppa Italia	...
...
409	T2	Champions League	Champions League	...

Basi di dati II — 29 maggio 2014 — Compito A

- Per ogni giocatore, interessa rappresentare anche l'allenatore sotto la guida del quale ha giocato (per sommare ad esempio presenze e reti segnate con un certo allenatore); supporre (anche se l'ipotesi non è realistica) che ogni squadra abbia, in ciascuna stagione, un solo allenatore e che siano disponibili, nella staging area, tutti i dati relativi agli allenatori nelle varie stagioni. Anche in questo caso, mostrare le modifiche da apportare a una o a entrambe le relazioni.

Va aggiunta, per l'allenatore, una dimensione secondaria (o supplementare), cioè una dimensione la cui chiave, nella tabella dei fatti, dipende funzionalmente da altre. Il valore da inserire per ogni ennupla viene univocamente determinato dai valori delle chiavi di squadra e stagione.

<u>KGiocatore</u>	<u>KSquadra</u>	<u>KTorneo</u>	<u>KStagione</u>	<u>KAllenatore</u>	Presenze	Reti
301	201	401	21
301	202	402	28
302	201	401	30
302	203	403	22
...

(Seguendo la consuetudine, KAllenatore è sottolineato, anche se in effetti non sarebbe corretto).

Va anche aggiunta una tabella per la dimensione allenatore.

<u>KAllenatore</u>	Cognome
...

- Con riferimento al punto precedente, considerare invece il caso, più realistico, in cui gli allenatori possono cambiare nel corso della stagione. Indicare quali modifiche sarebbero necessarie allo schema dimensionale e, soprattutto, quali dati sarebbero necessari nella staging area per poter riorganizzare il tutto.

In questo caso l'allenatore non è univocamente individuato dalle altre dimensioni e quindi la relativa dimensione non è secondaria. Lo schema rimane apparentemente lo stesso, con l'unica differenza che KAllenatore diventa parte effettiva della chiave della tabella dei fatti.

La vera differenza è nel contenuto: la grana è diversa, e ci possono essere più ennuple per un giocatore e una stagione. La tabella dei fatti va quindi ricalcolata e l'unica possibilità ragionevole per farlo è la disponibilità, nella staging area, dei dati di tutte le partite (qualunque altra ipotesi, ad esempio sui periodi è poco realistica).

Basi di dati II — 29 maggio 2014 — Compito A

Domanda 3 (20%) Supporre di voler organizzare una riunione in una certa data, seguendo un protocollo simile al commit a due fasi, procedendo attraverso chiamate telefoniche ai partecipanti, nel modo seguente:

- il coordinatore telefona a ciascuno dei partecipanti, chiedendo la disponibilità; se tutti dichiarano di essere disponibili, decide di svolgere la riunione, altrimenti no
- dopo la decisione, il coordinatore telefona di nuovo a tutti coloro che hanno dato disponibilità, per informarli sulla decisione.

Rispondere alle domande negli spazi sottostanti.

Quale potrebbe essere un punto debole della procedura (pensare al caso in cui per partecipare alla riunione i partecipanti debbano organizzare un viaggio)? Quale condizione sarebbe necessaria per garantire l'efficacia?

Possibile risposta

Il protocollo è asincrono. La prima fase può essere gestita con timeout ma la seconda no. Una garanzia per l'efficacia potrebbe essere la certezza che tutti rispondano al telefono quando li cerchiamo.

È sufficiente ritelefonare solo a coloro che hanno dato disponibilità? Perché?

Possibile risposta

Sì, perché sono i soli ad avere "impegnato" il proprio tempo, gli altri, avendo detto di no, sanno già che la riunione non ci sarà.

Domanda 4 (20%) Per ciascuno degli schedule sotto riportati, indicare, scrivendo **sì** o **no** in ciascuna casella, a quali classi appartiene: S (seriale, rispetto a letture e scritture, ignorare commit e abort), CSR (conflict-serializzabile), S2PL (cioè generabile da uno scheduler basato su 2PL stretto), MV (cioè generabile da uno scheduler multiversion con controllo di serializzabilità: "a serializable transaction cannot modify or lock rows changed by other transactions after the serializable transaction began"). Negli schedule, s_i indica l'inizio della transazione i e c_i il suo commit.

Soluzioni per il compito A

	S	CSR	S2PL	MV
$s_1, r_1(x), w_1(x), s_2, r_2(x), w_2(x), c_2, s_3, r_3(y), w_3(y), c_3, r_1(y), w_1(y), c_1$	no	sì	no	no
$s_2, r_2(x), w_2(x), c_2, s_1, r_1(x), w_1(x), c_1$	sì	sì	sì	sì
$s_1, r_1(x), s_2, r_2(x), w_1(x), c_1, w_2(x), c_2$	no	no	no	no
$s_1, r_1(x), s_2, r_2(x), w_2(x), r_2(y), w_2(y), c_2, r_1(y), c_1$	no	no	no	sì *

(*) in questo caso $r_1(y)$ legge il valore di y all'inizio della transazione 1, cioè prima della scrittura $w_2(y)$