

## Basi di dati II, primo modulo

### Esercizi di autovalutazione — 8 aprile 2011

**Domanda 1** Il *semijoin* è un'operazione simile al join, in cui del secondo operando interessano solo gli attributi di join. In concreto, se il join su  $A_1 = A_2$  di  $r_1(X_1)$  e  $r_2(X_2)$  (con  $A_1 \in X_1$ ,  $A_2 \in X_2$  e  $X_1 \cap X_2 = \emptyset$ ) è definito come

$$r_1 \text{ JOIN}_{A_1=A_2} r_2 = \{ t \text{ su } X_1 X_2 \mid \text{esistono } t_1 \in r_1 \text{ e } t_2 \in r_2 \text{ con } t[X_1] = t_1, t[X_2] = t_2 \text{ e } t_1[A_1] = t_2[A_2] \}$$

il semijoin corrispondente è definito come

$$r_1 \text{ SEMIJOIN}_{A_1=A_2} r_2 = \{ t_1 \mid t_1 \in r_1 \text{ ed esiste } t_2 \in r_2 \text{ con } t_1[A_1] = t_2[A_2] \}$$

Ad esempio:

$R_1$		$R_2$		$R_1 \text{ SEMIJOIN}_{D=G} R_2$																																		
<table style="width: 100%; border-collapse: collapse;"> <tr><th>A</th><th>B</th><th>C</th><th>D</th></tr> <tr><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>2</td><td>2</td><td>2</td><td>1</td></tr> <tr><td>3</td><td>2</td><td>2</td><td>2</td></tr> </table>	A	B	C	D	1	1	1	1	2	2	2	1	3	2	2	2		<table style="width: 100%; border-collapse: collapse;"> <tr><th>G</th><th>E</th></tr> <tr><td>1</td><td>1</td></tr> <tr><td>3</td><td>2</td></tr> </table>	G	E	1	1	3	2		<table style="width: 100%; border-collapse: collapse;"> <tr><th>A</th><th>B</th><th>C</th><th>D</th></tr> <tr><td>1</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>2</td><td>2</td><td>2</td><td>1</td></tr> </table>	A	B	C	D	1	1	1	1	2	2	2	1
A	B	C	D																																			
1	1	1	1																																			
2	2	2	1																																			
3	2	2	2																																			
G	E																																					
1	1																																					
3	2																																					
A	B	C	D																																			
1	1	1	1																																			
2	2	2	1																																			

Mostrare algoritmi per l'esecuzione del semijoin, come modifica degli algoritmi noti per il join, indicandone il costo (e sottolineando i casi in cui tale costo è diverso da quello del join).

Inoltre, specificare quali possono essere i vantaggi dell'uso del semijoin in contesti distribuiti (con le due relazioni memorizzate in nodi diversi), anche come passo preliminare per l'esecuzione di join.

**Domanda 2** Spiegare perché, scrivendo un programma che accede a due basi di dati diverse utilizzando JDBC, non è possibile garantire il commit a due fasi. Indicare quali funzionalità aggiuntive sono necessarie per i server locali che vogliono realizzare tale servizio e come si può procedere con un sistema che non disponga di tali funzionalità.

**Domanda 3** Si consideri il protocollo di commit a a due fasi (2PC).

1. Spiegare perché
  - (a) un guasto del coordinatore (TM) può avere conseguenze molto pesanti anche sulle prestazioni dei partecipanti (RM);
  - (b) un guasto di un partecipante non ha conseguenze particolari sulle prestazioni degli altri partecipanti (a parte l'eventuale abort di transazioni).

Per ovviare alle conseguenze negative di un guasto del coordinatore, alcune implementazioni del 2PC prevedono la possibilità di comunicazione fra i partecipanti (mentre la versione base prevede solo comunicazione fra il coordinatore e ciascuno dei partecipanti). In particolare, un partecipante che abbia una transazione in stato di "ready" può chiedere agli altri partecipanti informazioni sullo stato di tale transazione (che può essere "prima-del-ready", "ready", "commit" o "abort") presso di loro.

2. In tale contesto indicare
  - (a) quali insiemi di risposte il partecipante può ricevere e quali invece no;
  - (b) come (e in quali casi) il partecipante può trarre profitto dalle risposte ottenute.

**Domanda 4** Si supponga che Napoleone abbia fatto utilizzare il commit a due fasi per organizzare le attività in battaglia. Si consideri il seguente scenario. Esistono:

- (a) generali che possono coordinare azioni
- (b) reparti di riserva disponibili per azioni
- (c) messaggeri utilizzati dai generali e dai reparti di riserva per comunicare gli uni con gli altri

In particolare, il 2PC viene utilizzato dai generali per organizzare azioni che coinvolgono (*contemporaneamente, ad un certo orario*) due o più reparti di riserva (ma ciascun reparto potrebbe ricevere richieste da più generali). Ad esempio, il generale Murat, alle ore 10, decide di voler organizzare un attacco alle ore 12 con il coinvolgimento del quarto squadrone di cavalleria e della prima batteria di artiglieria pesante.

Descrivere brevemente il protocollo in questo contesto, sottolineando le criticità, dovute al fatto che si tratta di una battaglia e quindi tutti i soggetti coinvolti (generali, comandanti dei reparti di riserva e messaggeri) possono essere colpiti. In particolare, individuare quali ipotesi sarebbero necessarie (anche se non sempre soddisfatte in una battaglia) per permettere l'utilizzo del protocollo.