

Homework for Distributed Database Systems

(Issue: 2012 / 3 / 14 Due: 2012 / 3 / 28)

1. Fragmentation

A supply-part database consists of three relations as shown below.

SUPPLIER		
SNO	SNAME	COUNTRY
S1	SN1	USA
S2	SN2	INDIA
S3	SN3	CHINA
S4	SN4	CHINA
S5	SN5	INDIA
S6	SN6	USA

SUPPLY		
SNO	PNO	QTY
S1	P1	60
S1	P3	70
S2	P2	60
S3	P3	55
S3	P4	96
S4	P2	65
S6	P2	70
S6	P4	96

PARTS		
PNO	PNAME	PRICE
P1	PC	10000
P2	CAMERA	8000
P3	VIDEO	5000
P4	HI-HI	3000

There are three applications:

Q1: Print SNO of suppliers who supply parts with price less than 6000.

Q2: For each supplier in USA, print SNAME, and PNO of the parts that s/he supplies.

Q3: For each supplier, print SNO, SNAME, the number of parts s/he supplies.

- a) Determine a set of simple predicates that is complete and minimal. Justify your answer.
- b) Derive a horizontal fragmentation step by step, and show the contents of each fragmentation.

2. Allocation

- a) You are given a fragment F , a set of sites S_1, S_2, \dots, S_m , and a set of queries, Q_1, Q_2, \dots, Q_n . Workload of the system is represented by two arrays: $Size[1..n]$, and $Freq[1..n, 1..m]$. $Size[i]$ is the size of data to be transferred for query i when the fragment is not allocated at the site where the query is issued. $Freq[i,j]$ is the frequency of query i issued at site j . Write the procedure (in pseudo-code form) that determines the site where F should be allocated so that the total data transferring cost is minimized.
- b) Follow the above question. If you are also given an array $Update[1..n]$, where $Update[i]$ represents the update cost of query i ($Update[i] = 0$ indicates that query i does not involve any updates). Write the procedure (in pseudo-code form) that determines whether F should be replicated another site.