INFS7907 Advanced Database Systems

Tutorial 2: Distributed Query Processing and Transactions

Semester 1, 2004

Question 1: **[PRIMARY HORIZONTAL FRAGMENTATION]**

Consider S given below. It is the subject of primary horizontal fragmentation.

S

TITLE	**SAL**
Elect. Eng.	40000
Syst. Anal.	34000
Mech. Eng.	27000
Programmer	24000

(a) Suppose that there is only 1 application that accesses S. The application checks the salary information and determines a raise accordingly. Assume that employee records are managed in two places, one handling the records of those with salaries less than or equal to $30,000 and the other handling the records of those who earn more than $30,000. Therefore the query is issued at two sites.

The simple predicates that would be used to partition S are:

P_1: SAL $\leq$ 30000
P_2: SAL > 30000

thus giving the initial set of simple predicates Pr = $\{p_1, p_2\}$.

(a) How do we check if Pr is complete and minimal?
(b) Apply the PHORIZONTAL algorithm.

Question 2: A company has branches in Brisbane, Sydney and Melbourne. Its company database consists of two tables:

```
branch(name, location, manager)

employee(name, office, salary, home_address)
```

where primary keys are underlined, `branch.manager` is a foreign key to `employee.name`, and `employee.office` is a foreign key to `branch.name`. The database is fragmented as the following:

Brisbane site: $\sigma_{location = \text{"Brisbane"}}$ branch, $\Pi_{name, office, home\_address}$ employee

Sydney site: $\sigma_{location = \text{"Sydney"}}$ branch

Melbourne site: $\sigma_{location \neq \text{"Brisbane"} \text{ and } location \neq \text{"Sydney"}}$ branch, $\Pi_{name,\ salary}$ employee

For the following query

select   e.name, salary
from   branch b, employee e
where location = "Brisbane" and e.name = manager

Give operations trees for the above query 1) after query decomposition using global schema, 2) after localization, and 3) after reduction.

Question 3: Let $r(\overline{A}B)$ and $s(\overline{B}CD)$ be two following relations

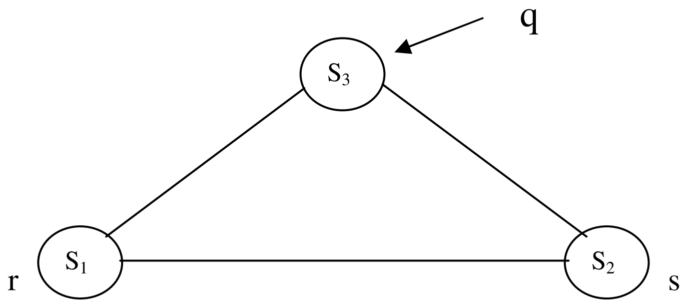
$r(\overline{A}$, B)			$s(\overline{B}$ C D)		
1	4		4	5	0
1	5		4	7	8
2	4		5	0	1
2	6		5	2	1
3	7				

(a) compute $r \ltimes s$
(b) compute $s \ltimes r$
(c) Assume r is at site 1 and s is at site 2, and a query $r \bowtie s$ has been issues at site 2. Give steps for a query processing strategy using semijoin, and check if the semijoin is a beneficial option in this case (ignore local processing cost).

Question 4: Assume that a computing network consists of three sites S_1, S_2 and S_3. There are two relations:

$r(\overline{A}$ B C) located at site S_1

$s(\overline{C}$ D E) located at site S_2



The query q ← r ⋈ s has been issued at site S_3.

For the following population of relations:

r ($\overline{A}$ B C)

A	B	C
2	4	5
1	2	3
4	5	6
7	8	9

s ($\overline{C}$ D E)

C	D	E
5	5	6
6	4	3
8	5	1
1	2	1
7	1	0

Is the semijoin a beneficial option in this distributed system? Ignore the computation costs at each site, or assume that they are homogeneous sites.

Question 5:

Given the following database schema

WARD

<u>NO</u>	WNAME	LOC	BEDS	AIR-COND
ITC1	Intensive Care 1	Level 1	45	Y
ITC2	Intensive Care 2	Level 3	30	Y
GW1	General Ward 1	Ground	60	N
EW1	Emergency Ward 1	Level 2	12	Y
MAT1	Maternity Ward	Ground	58	Y
EW2	Emergency Ward 2	Ground	45	N
GW2	General Ward 2	Level 1	120	N

PATIENT

<u>PAT-NO</u>	PNAME	WARD	AGE	INIT-DIAG
23-678-99	John	ITC1	56	Heart Attack
12-976-00	Harry	GW1	32	Kidney
49-654-99	David	EW1	18	Concussion
76-862-98	Barbara	MAT1	27	Labour
43-853-00	Bruce	EW2	21	Fracture
12-657-99	Kate	GW2	39	Appendicitis
34-098-99	Anthony	GW1	78	Flu

a) Write an SQL query to find the name and initial diagnosis of patients in air-conditioned wards

b) What is the minterm predicate for the primary horizontal fragmentation of the WARD relation for air-conditioned wards

c) How many fragments of WARD are generated from applying the predicates

P1: AIR-COND = 'Y'

P2: AIR-COND = 'N'

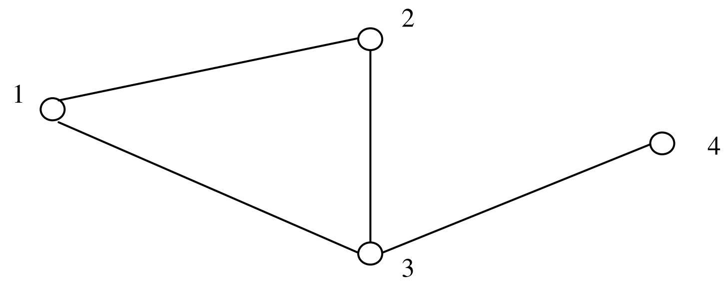
d) What is the relationship between WARD and PATIENT, that is which is owner and which is member

e) What is the join predicate of the relations WARD and PATIENT

f) Find the derived horizontal fragments of the PATIENT relation on the basis of the predicates used in (c)

Question 6:

Let the following be a network of computing nodes, and data allocation.



Node 1 → (r_1, r_3)
Node 2 → (r_2, r_3, r_4)
Node 3 → (r_5, r_4)
Node 4 → (r_1, r_6)

Discuss query execution strategies for the following queries. How will the size of the tables affect the choice of a 'good' strategy?

(a) $q_1 = r_2 >< r_3 >< r_6$ at node 4

(b) $q_2 = \sigma_P(r_4) >< r_5$ at node 1

(c) $q_3 = \sigma_P(r_6) >< \sigma_S(r_1) >< \pi_X(r_4)$ at node 1

Distributed Transactions

Question 7: Insert appropriate *lock* and *unlock* operations in the following transactions, following the read/write locking scheme.

T1:

read-item (X); X:= X +K; write-item (X);	

T2:

read-item (Y); read-item (X); X:= X + Y; write-item (X);	

T3:

read-item (Z); Z:= Z + N; read-item (X); X:= X + Z; read item (Y); Z:= Z + Y; write-item (Z); write-item (X);	

Question 8: Do the following transactions follow the two-phase locking scheme. If not, convert to comply with the scheme.

T1:

read-lock (X); write-lock (Y); read-item (Y); Y:= Y + M; write-item (Y); unlock (Y); read-item (X); write-lock (X); X:=X + N; unlock (X);	

T2:

read-lock (Y); read-item (Y); write-lock (X); X:= X + Y; write-item (X); unlock (X); unlock (Y);	

T3

write-lock (Y); read-item (Y); Y:= Y + M; write-item (Y); read-lock (Y); write-lock (X); X:= X + Y; unlock (Y); write-item (X); unlock (X);	

Question 9: Find a schedule of the following transactions that may cause a deadlock

T1	T2
read-lock (X); read-item (X); write-lock (Y); unlock (X); read-item (Y); Y:= Y + X; write-item (Y); unlock (Y);	read-lock (Y); read-item (Y); write-lock (X); unlock (Y); read-item (X); X:= X + Y; write-item (X); unlock (X);

T1	T2