Fall 2000 COMP334 Midterm Solutions

Question 1: (20%)

a) What are the similarities and differences between parallel and distributed database management systems?

Answer:
Similarities:

- Both make access to DDBS transparent to users
- Both support inter-query and intra-query parallelism, reducing the query response time

Differences:

- PDBMS is homogeneous, while DDBMS may be homogeneous or heterogeneous
- PDBMS is usually under one operating system; DDBMS is not.

b) Why is giving database objects unique names more complicated in a distributed DBMS?

Answer:

- Each assignment of name would require to check all DBMS for the existence of that name.
- Synchronization problems: each site lacks the instantaneous information of other sites.

c) Briefly discuss the ajor differences between the classical distributed database systems and the multi-database systems.

Answer:
From two aspects,

- Architectural models of their management systems
    i. Autonomy: DDBS is tightly integrated; MDBS is total isolation.
    ii. Heterogeneity: Usually DDBS is homogeneous, but MDBS is heterogeneous.
- Design
    i. Strategy: DDBS is top-down; MDBS is bottom-up.
    ii. Structure: DDBS must have a GCS; MDBS may not have it.

d) There are two possible architectures for multi-database systems: with or without the global schema. List the major differences between these two alternatives.

Answer:

- With GCS, global external schemas are defined from GCS or the global database structure; Without GCS, global external schemas are defined from LCSs of related sites.

- With GCS, MDBMS explicitly handles the global part of databases; without GCS, MDBMS does not need to do so.

Question 2: (25%)

A supply-part database consists of fur relations as shown below:

SUPPLIER		
SNO	SNAME	COUNTRY
S1	SN1	USA
S2	SN2	INDIA
S3	SN3	CANADA
S4	SN4	CHINA
S5	SN5	INDIA
S6	SN6	USA
S7	SN7	CANADA
S8	SN8	INDIA

SUPPLY		
SNO	PNO	QTY
S1	P1	60
S2	P1	70
S2	P2	60
S3	P3	55
S3	P4	96
S4	P2	65
S5	P2	70
S6	P4	96
S7	P3	72
S8	P3	80

PARTS		
PNO	PNAME	COLOR
P1	PC	30
P2	CAMERA	27
P3	VIDEO	50
P4	HI-FI	62

COUNTRY	
COUNTRY	REGION
CHINA	ASIA
INDIA	ASIA
USA	AMERICA
CANADA	AMERICA

Let p1: REGION = "ASIA", p2:REGION = "AMERICA" be two simple predicates. Let us further assume that there are only two regions.

a) **Perform PHF of COUNTRY with respect to these predicates. Show the detailed steps and the result fragments.**

Answer:

Set of simple predicates = {p1, p2}
Set of minterm predicates = {$p1 \wedge p2$, $\neg p1 \wedge p2$, $p1 \wedge \neg p2$, $\neg p1 \wedge \neg p2$}
Set of implications = {$p1 \Leftrightarrow \neg p2$, $p2 \Leftrightarrow \neg p1$}

After elimination and simplfication,
Set of minterm predicates = {p1, p2}

Therefore, there are two result gragments:

COUNTRY1 = $\sigma_{\text{REGION="ASIA"}}$(COUNTRY)
COUNTRY2 = $\sigma_{\text{REGION="AMERCIA"}}$(COUNTRY)

b) **Perform further DHF for relations in the database based on the fragmentation obtained in step(a).**

Answer:

Step 1: Since SUPPLIER.COUNTRY is a foreign key attribute referencing COUNTRY.COUNTRY, SUPPLIER should be perform DHF:

SUPPLIER1 = SUPPLIER $\ltimes$ COUNTRY1
SUPPLIER2 = SUPPLIER $\ltimes$ COUNTRY2

Step 2: Since SUPPLY.SNO is a foreign key attribute referencing SUPPLIER.SNO, SUPPLY should be perform DHF:

SUPPLY1 = SUPPLY $\ltimes$ SUPPLIER1
SUPPLY2 = SUPPLY $\ltimes$ SUPPLIER2

Question 3: (15%)

The following three application access the database in Question 2.

Q1: Print SNO of supplier with PNAME of the parts supplied by the supplier.
Q2: For each supplier, print SNAME and total quantity of parts that he supplies.
Q3: For each part, print PNAME and the total quantity supplied.

a) **Construct the U(Q, A) matrix for the attributes for all relations accessed by the applications.**

Answer:

```
Q1: select SNO, PNAME
      from SUPPLY, PARTs
      where SUPPLY.PNO = PARTS.PNO

Q2: select SNAME, sum(QTY)
      from SUPPLIER, SUPPLY
      where SUPPLIER.SNO = SUPPLY.SNO
      group by SUPPLIER.SNO

Q3: select PNAME, sym(QTY)
      from PARTS, SUPPLY
      where PARTs.PNO = SUPPLY.PNO
      group buy PARTS.PNO
```

	SUPPLIER			COUNTRY		PARTS			SUPPLY		
	SNO	SNAME	COUNTRY	COUNTRY	REGION	PNO	PNAME	COLOR	SNO	PNO	QTY
Q1	0	0	0	0	0	1	1	0	1	1	0
Q2	1	1	0	0	0	0	0	0	1	0	1
Q3	0	1	0	0	0	1	1	0	0	1	1

b) **Let us further assume that Q1 is executed at sites 1 and 2, with frequencies 10 and 20, respectively; Q2 is executed at sites 2 and 3 with frequencies 20 and 10, respectively, and Q3 is executed at sites 1 and 3**

with frequencies 10 and 5, respectively. Construct the affinity matrix for all attributes accessed (Assume that the number of accesses for each execution at all site is 1).

Answer:
Actually, we do not need to construct the affinity values for comparing any two attributes of different relations. Instead, the affinity matrix for each relation's attributes is given:

AA(SUPPLIER):

	SNO	SNAME	COUNTRY
SNO	30	30	0
SNAME	30	30	0
COUNTRY	0	0	0

AA(COUNTRY):

	COUNTRY	REGION
COUNTRY	0	0
REGION	0	0

AA(PARTS)

	PNO	PAME	COLOR
PNO	45	45	0
PNAME	45	45	0
COLOR	0	0	0

AA(SUPPLY)

	SNO	PNO	QTY
SNO	60	30	30
PNO	30	45	15
QTY	30	15	45

Question 4: (40%)

Given the following use-matrix U and the access frequency matrix ACC, find a vertical fragmentation using the BEA and Partitioning algorithm (Assume the number of access for each execution at all sites is 1)

U =

	A1	A2	A3	A4
q1	0	1	1	1
q2	1	1	1	1
q3	1	0	0	1
q4	0	0	1	0
q5	1	1	1	0

ACC =

	S1	S2	S3
q1	10	20	0
q2	5	0	10
q3	0	35	5
q4	0	10	0
q5	0	15	0

Answer:

Step 1: Finding the AA matrix

	A1	A2	A3	A4
A1	70	30	30	55
A2	30	60	60	45
A3	30	60	70	45
A4	55	45	45	85

Step 2: Find the CA matrix (by using the Bond Energy Algorithm)

Iteration 1:Placing A3 into CA.

Cont(0, A3, A1) = 16950
Cont(A1, A3, A2) = 22050
Cont(A2, A3, 0) = 24150

Therefore, A3 should be placed between A1 and A2.

Iteration 2: Placing A4 into CA.

Cont(0, A4, A1) = 22450
Cont(A1, A4, A3) = 28150
Cont(A3, A4, A2) = 22950
Cont(A2, A4, 0) = 21750

Therefore, A4 should be placed between A1 and A3.

Iteration 3: Row Reorganization

The final CA is:

	A1	A4	A3	A2
A1	70	55	30	30
A4	55	85	45	45
A3	30	45	70	60
A2	30	45	60	60

Step 3: Performing a binary partition (by using the Partitioning Algorithm)

Shift 0: ordering is A1 A4 A3 A2
Iteration 1: TA = {A1}, BA = {A4, A3, A2}
TQ = {}, BQ = {q1, q4}, OQ = {q2, q3, q5}
CTQ = 0, CBQ = 40, COQ = 70, z = -4900.
Iteration 2: TA = {A1, A4}, BA = {A3, A2}
TQ = {q3}, BQ = {q4}, OQ = {q1, q2, q5}
CTQ = 40, CBQ = 10, COQ = 60, z = -3200
Iteration 3: TA = {A1, A4, A3}, BA = {A2}
TQ = {q3, q4}, BQ = {}, OQ = {q1, q2, q5}
CTQ = 50, CBQ = 0, COQ = 60, z = -3600

Shift 1: ordering is A4 A3 A2 A1
Iteration 1: TA = {A4}, BA = {A3, A2, A1}
TQ = {}, BQ = {q4, q5}, OQ = {q1, q2, q3}
CTQ = 0, CBQ = 25, COQ = 85, z = -7225
Iteration 2: TA = {A4, A3}, BA = {A2, A1}
TQ = {q4}, BQ = {}, OQ = {q1, q2, q3, q5}
CTQ = 10, CBQ = 0, COQ = 100, z = -10000
Iteration 3: repeated

Shift 2: ordering is A3 A2 A1 A4
Iteration 1: TA = {A3}, BA = {A2, A1, A4}
TQ = {q4}, BQ = {q3}, OQ = {q1, q2, q5}
CTQ = 10, CBQ = 40, COQ = 60, z = -3200
Iteration 2: repeated
Iteration 3: repeated
Shift 3: ordering is A2 A1 A4 A3 -> no need since all are repeated.

Therefore, Shift 0 Iteration 2 is chosen: TA = {A1, A4}, BA = {A3, A2}
Let A0 be the primary key attribute of relation R = (A0, A1, A2, A3, A4). A vertical fragmentation of R is

$R1 = \Pi_{A0, A1, A4}(R)$
$R2 = \Pi_{A0, A2, A3}(R)$