Date: 28 / 09 / 2017

Ahsanullah University of Science and Technology

Department of Computer Science and Engineering

4th Year, 1st Semester, Semester Final Examination (Spring-2017)

Course No: CSE 4125 **Course Title:** Distributed Database Systems

Time: 3 Hours **Full Marks:** 70

- **PART – A contains 5 (Five) questions (no. 1 to 5). Answer any 3 (Three) of them.**
- **PART – B contains 2 (Two) questions (no. 6 and 7). Answer ALL of them.**
- **Marks allotted are indicated in the right margin within '[ ]'.**

PART - A

1. (a) Compare the features of distributed and centralized database. [6]

   (b) Provide descriptions with appropriate diagrams of – [3 x 2 = 6]

   i. A distributed database on geographically dispersed locations.

   ii. A distributed database on a local network.

   iii. A database with multiprocessor system.

   (c) Provide a diagram showing all the components of a distributed database management system. [2]

2. (a) Describe the ISO/ OSI reference architecture. [3]

   (b) Define *relational schema*, *grade* and *tuple* with proper examples. [3]

   (c) Consider the following relations.

ACCOUNT

ACC_ID	*ACC_NAME*	*BRANCH_ID*
111	*Red*	001
112	*Green*	002
113	*Blue*	001
114	*Yellow*	001
115	*White*	003

BRANCH

BRANCH_ID	*LOCATION*
001	*dhk*
002	*dhk*
003	*ctg*
004	*ctg*

Horizontal fragmentation is defined by the following fragmentation schema.

$$ACCOUNT_1 = SL_{BRANCH\_ID="001"}\ ACCOUNT$$

$$ACCOUNT_2 = SL_{BRANCH\_ID="002"}\ ACCOUNT$$

i. Verify the *completeness, reconstruction* and *disjointness* conditions for the above fragmentation. [5]

ii. Write a fragmentation schema to derive the *horizontal fragmentation* of $ACCOUNT$ from $BRANCH$ (based on $LOCATION$). [3]

3. (a) Consider the global relational schema $R\ (ID,\ NAME,\ GENDER,\ AGE)$, given the following fragmentation schema:

$$R_1^1 = PJ_{ID,\ NAME}\ R$$
$$R_2^2 = PJ_{ID,\ GENDER,\ AGE}\ R$$
$$R_3^1 = SL_{GENDER=M}(PJ_{ID,\ GENDER,\ AGE}\ R)$$
$$R_4^2 = SL_{GENDER=F}(PJ_{ID,\ GENDER,\ AGE}\ R)$$
$$R_5^3 = PJ_{ID,\ AGE}(SL_{GENDER=M}(PJ_{ID,\ GENDER,\ AGE}\ R))$$
$$R_6^3 = SL_{AGE>20}(PJ_{ID,\ AGE}(SL_{GENDER=M}(PJ_{ID,\ GENDER,\ AGE}\ R)))$$
$$R_7^4 = SL_{AGE\le 20}(PJ_{ID,\ AGE}(SL_{GENDER=F}(PJ_{ID,\ GENDER,\ AGE}\ R)))$$
$$R_8^3 = SL_{AGE>20}(PJ_{ID,\ AGE}(SL_{GENDER=F}(PJ_{ID,\ GENDER,\ AGE}\ R)))$$
$$R_9^4 = SL_{AGE\le 20}(PJ_{ID,\ AGE}(SL_{GENDER=M}(PJ_{ID,\ GENDER,\ AGE}\ R)))$$

i. Draw the fragmentation tree. [5]

ii. Show the allocation of the fragments to the physical images at different sites. How many of them are *non-redundant* allocation? [5]

(b) Explain two different ways to build *mixed fragmentation* with appropriate diagrams. [4]

4. (a) Consider the following relation: [4]

$Product$	$Warranty\ (months)$	$Price\ (thousands)$
$Laptop$	24	70
$Phone$	12	20
$Watch$	6	10
$Tablet$	12	30

The following min-term predicates for the above relation are given:

$$m_1 : Product = Tablet \ \wedge\ Warranty < 12 \ \wedge\ Price \ge 20$$
$$m_2 : Product = Laptop \ \wedge\ Warranty \ge 12 \ \wedge\ Price \ge 20$$
$$m_3 : Product = Tablet \ \wedge\ Warranty < 12 \ \wedge\ Price < 20$$

$$m_4 : Product \neq Laptop \;\wedge\; Warranty \geq 6 \;\wedge\; Price \geq 32$$

$$m_5 : Product \neq Tablet \;\wedge\; Warranty < 6 \;\wedge\; Price < 20$$

Generate the set P of all possible simple predicates for m_1 to m_5.

(b) Evaluate the following expression using the rules of qualified relation. Show the steps and indicate the rules applied. [5]

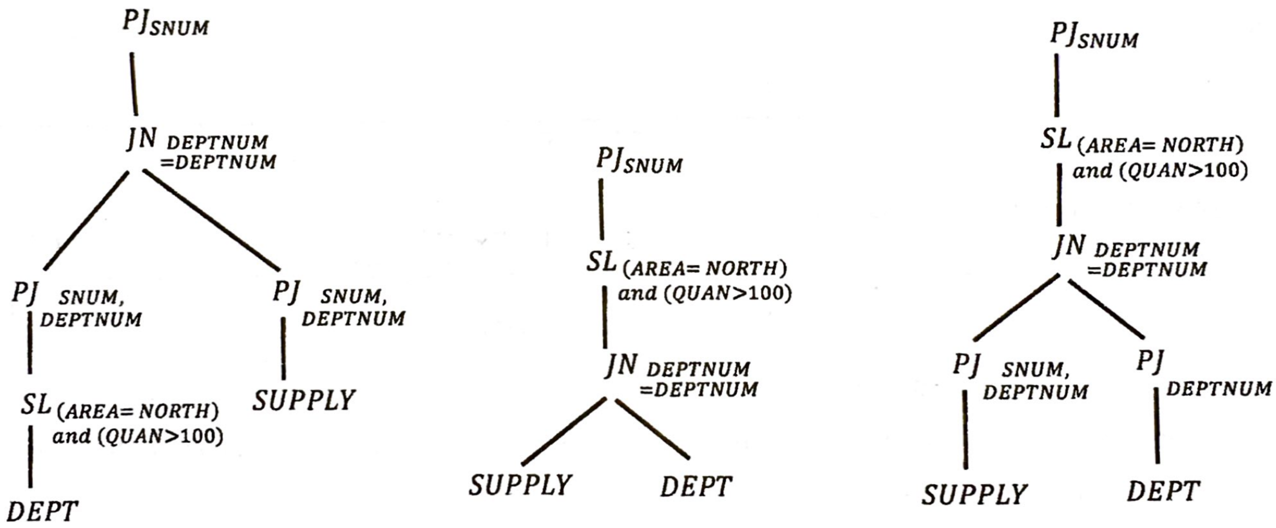
$$[\, M : qm \,] \;\; DF \;\; (([\, R : qr \,] \; UN \; [\, S : qs \,]) \, JN_F ([\, M : qm \,] \; CP \; [\, S : qs \,]))$$

(c) Consider the following relational schema: [5]

$$DEPT\ (DEPTNUM, NAME, AREA, MGRNUM)$$

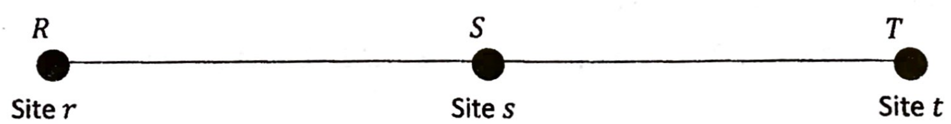
$$SUPPLY\ (SNUM, PNUM, DEPTNUM, QUAN)$$

Find the invalid operator trees from the following trees (Q_1, Q_2 and Q_3). State the reason behind your answer.



(Q_1) (Q_2) (Q_3)

5. (a) Consider the following distributed database with the relations R, S and T over a network.



Let the system dependent constants be $C_0 = 0$ and $C_1 = 1$, and suppose, the following things are known.

$$size\ (R) = 50\ ; \quad card\ (R) = 100\ ; \quad val\ (a[R]) = 50\ ; \quad size\ (a) = 2$$

$$size\ (S) = 20\ ; \quad card\ (S) = 50\ ; \quad val\ (b[S]) = 50\ ; \quad size\ (b) = 2$$

$$size\ (T) = 10\ ; \quad card\ (T) = 50\ ; \quad val\ (c\ [T]) = 50\ ; \quad size\ (c) = 2$$

$R\ SJ_{a=b}\ S$ has selectivity $\rho = 0.1$

$S\ SJ_{b=a}\ R$ has selectivity $\rho = 0.9$

$T\ SJ_{c=b}\ S$ has selectivity $\rho = 0.5$

$S\ SJ_{b=c}\ T$ has selectivity $\rho = 0.5$

The query we want to perform, Q: $(\ \boldsymbol{R}\ JN_{a=b}\boldsymbol{S}\ )\ UN\ (\ \boldsymbol{T}\ JN_{c=b}\boldsymbol{S}\ )$.

Give the total transmission cost of performing Q at *site s* using the semi-join program and without the semi-join program. Which one is the best solution? [5+5+1 =11]

(b) Define *transaction recovery, crash recovery* and *commitment*. [3]

PART - B

---

6. (a) Consider the following global query:

$$(\ (\ (SL_{F1}\ \boldsymbol{R}\ JN_{A=B}\ \ \boldsymbol{S}\ )\ DF\ (SL_{F2}\ \boldsymbol{R}\ JN_{A=B}\ \ \boldsymbol{S}\ )\ )\ NJN\ (\ (\boldsymbol{R}\ JN_{A=B}\ \ \boldsymbol{S}\ )\ UN\ (SL_{F3}\ \boldsymbol{R}\ JN_{A=B}\ \ \boldsymbol{S}\ )\ )\ )$$
$$UN\ (\ SL_{F1\ AND\ NOT\ F2}(\boldsymbol{R}\ JN_{A=B}\ \ \boldsymbol{S}\ )\ )$$

i. Draw the operator tree. [2]

ii. Perform step-by-step transformations to simplify the operator tree, indicating which rule and criterion is applied at each step. [5]

iii. Write the query from the obtained simplified tree. [2]

iv. Transform the simplified query into fragment query by applying canonical expression. Say, R has three fragments, R_1, R_2, R_3, and S has two fragments S_1, S_2. [2]

(b) Prove the following rule of qualified relation: [3]

$$[\ A : a\ ]\ SJ_P\ [\ B : b\ ] \Rightarrow [\ A\ SJ_P\ B\ :\ a\ \ AND\ \ b\ \ AND\ \ P\ ]$$

7. (a) Four fragments R, S, M and T of a relation are given. We want to perform the following query.

$$Q: \left(PJ_a(R\ \ UN\ \ S)\right) JN_{a=a}\left(SL_{m=value}(M\ \ UN\ \ T)\right)$$

Database profiles are provided below.

$card\ (R) = 300$
$site(R) = 1$

	a	b	c	d
$size$	6	7	2	10
val	300	1000	30	50

card (S) = 100
$site(S) = 4$

	a	b	c	d
$size$	6	7	2	10
val	100	10	20	15

$card\ (T) = 2000$
$site(T) = 3$

	a	m	n
$size$	6	5	4
val	2000	5	5

$card\ (M) = 2000$
$site(M) = 2$

	a	m	n
$size$	6	5	4
val	2000	5	5

Assume that, the result of ($R\ \ UN\ \ S$) has no duplicate values for the attribute a, and the same property stands for ($M\ \ UN\ \ T$).

Now answer the following questions.

i. If attribute a is the primary key of S, then $card\ (S) = ?$ 100 [1]

ii. $size\ (R\ SJ_{a=a}\ M) = ?$ [2]

iii. For the simple selection $SL_{m\ =\ value}$ ($M\ \ UN\ \ T$), estimate the selectivity ρ. [2]

iv. Estimate the cardinality of the result of Q. Indicate the formulas applied. [3]

v. Estimate the total size of data in the result of Q. Indicate the formulas applied. [2]

(b) Write a reducer program for the query Q mentioned in 7(a) to optimize the corresponding operator tree. Draw the obtained optimization graph. [4]

- END -

card (M UN T)

= card (M) + card (T)

= 4000

= $\rho \times$ ~~card~~ 4000

$\frac{1}{val[M(m)]}$ $\frac{1}{5}$ =