JARAMOGI OGINGA ODINGA UNIVERSITY OF SCIENCE AND TECHNOLOGY

SCHOOL OF INFORMATICS AND INNOVATIVE SYSTEMS

UNIVERSITY EXAMINATION FOR THE DEGREE OF MASTER OF SCIENCE

MANAGEMENT INFORMATION SYSTEM

1ST YEAR 2ND SEMESTER 2013/2014 ACADEMIC YEAR

CENTRE: KISUMU

COURSE CODE: IIS 5121

COURSE TITLE: DISTRIBUTED DATABASE SYSTEM

EXAM VENUE: **STREAM: BSc. Library Information Science**

DATE: 6/12/2013 **EXAM SESSION: 9.00 – 12 NOON**

TIME: 2 HOURS

Instructions:

1. **Answer question 1(Compulsory) and ANY other 2 questions**
2. **Candidates are advised not to write on the question paper.**
3. **Candidates must hand in their answer booklets to the invigilator while in the examination room.**

QUESTION 1:(20 MARKS)

a) Explain how transparent management of distributed and replicated data is
. achieved through the different transparencies involved. (10 marks)

b) Explain the correctness rules of fragmentation. (10 marks)

QUESTION 2(20 MARKS)

1. a.) what are the categories of cloud services?Explain each of them. What benefits do the user stand to gain if they apply all these categories of cloud services? (18 marks)

b) What is server virtualization? (2 marks)

QUESTION 3(20 MARKS)

2. a) Consider a database schema with the following two relations:
S(eid, dob, project, hours)

R(eid, name, sex)

where eid is the employee id, dob is the date-of-birth, project is the project that the employee participates in, hours is the number of hours the employee spent on the project, name is the

employee name, and sex is the sex of the employee. The key attributes are underlined (i.e.,

(eid,dob) is a composite key of S, and eid is the key of R). Moreover, sex can only be M(male) or F (female). The following predicates appear in the majority of queries:

hours $\leq$ 5, hours $\leq$ 15, sex = M, sex = F

We would like to fragment R and S using the different techniques for fragmentation. Choose

which one you think will work best for you. Say if these statements are true of false.
. (4 marks)

i.) S1 = σ hours $\leq$ 15 S is one fragment for S.

ii.) There are in total 2 fragments of S and 2 fragments of R.

iii.) Relation R can be further fragmented using derived horizontal fragmentation based on the fragmentations of S.

iv.) The vertical fragmentation of S into S1(eid, project) and S2(eid, dob, hours) is desirable if queries frequently access (eid, dob, hours).

b) Consider an Engineering database that maintains three tables:

EMP (ENO, NAME, TITLE, SAL)

PROJ (PNO, PNAME, BUDGET, LOC)

ASG (PNO, ENO, RESP, DUR)

EMP stores employee information, including the name, title and salary of the employee.

PROJ stores project information, including the name, budget and location of the project.

ASG keeps information about the assignment between projects and employees. RESP is the responsibility of the employee, while DUR is the duration that the employee works on the project. ENO is the employee id. PNO is the project id. The keys are underlined.

The following predicates appear in the majority of queries.

LOC = “New York”, LOC = “Montreal”, LOC = “Paris”

SAL > 5000, SAL ≤ 5000

i.) Write down the fragments if we perform primary fragmentation on EMP and PROJ,given the predicates above. (6 marks)

ii.) Write down the fragments if we perform derived horizontal fragmentation on ASG based on the fragments of PROJ in a). (4 marks)

iii.) Write down the fragments, if we further fragment EMP based on fragmented ASG. Is this kind of fragmentation (on the three tables) a good one? Why? (Hint: Analyze the desired properties of fragmentation.) (6 marks)

QUESTION 4 (20 MARKS)

3. In a shared nothing database system with 4 homogeneous nodes, relation R(X,Y) is partitioned on the attribute X across node 1 and node 2, while relation S(Y,Z) is partitioned across node 3 and node 4 on attribute Y. We want to perform equal-join between R and S on these 4 nodes.
The query optimizer has the following statistical information:

Node	table	NO OF ROWS	Min Y value	Max Y value
1	R	16K	1	4000
2	R	24K	4001	16000
3	S	24K	1	12000
4	S	16K	12001	16000

Assume that data in each bucket is uniformly distributed.

a) If communication cost is the only concern, suggest a scheme to reallocate R and S. Calculate the total communication cost. Communication cost is measured in the number of tuples transmitted. (4 marks)

b) Suppose load-balance is the major concern. Suggest a scheme to reallocate R and S that minimize the communication cost while achieving the best balance between nodes. The workload is defined as the number of join results estimated. (10 marks)

c) After redistribution, partitions of R and S are hash-joined locally in each node. Each node has a hash table with 1000 entries. If the workload is defined as the number of comparisons during the hash-join, calculate the workload (total number of comparisons) in the above two schemes. Suppose the hash functions used are completely random. (6 marks)

QUESTION 5(20 MARKS)

5. Assume that a) the cost of storing fragments at sites is negligible, and b) the cost of accessing a local fragment is 0 unit and c) the cost of READing a remote fragment is 1 unit and the cost of WRITEing a remote fragment is 2 units.

The access profiles of six applications are as follows:

Application	Fragment											
	1		2		3		4		5		6	
	R	W	R	W	R	W	R	W	R	W	R	W
1	-	30	-	10	10	-	30	-	-	-	50	10
2	20	-	40	-	-	30	-	50	10	40	-	-
3	10	20	-	50	-	-	20	-	30	-	30	-
4	20	10	20	-	-	-	50	20	50	10	-	30
5	-	-	10	10	100	-	-	20	50	80	10	40
6	10	60	10	-	-	-	20	40	-	-	-	-

The frequency of applications on sites is as follows:

Application	sites			
	A	B	C	D
1	20	-	-	30
2	-	10	10	-
3	10	30	-	20
4	-	-	50	20
5	20	10	-	-
6	-	10	20	-

Allocate the six fragments to the four sites using the “best-fit” method.