THE UNIVERSITY OF DODOMA

COLLEGE OF INFORMATICS AND VIRTUAL EDUCATION

SCHOOL OF INFORMATICS

DEPARTMENT OF COMPUTER SCIENCE

CS300 DISTRIBUTED DATABASE SYSTEMS

UNIVERSITY EXAMINATION

FIRST SEMESTER

2011/2012 ACADEMIC YEAR

Duration: 3 Hours **Time: 08:00-11:00 HRS**

Date: 6th February, 2012

Instructions:

i. This Question paper consists of 7 Questions organized in TWO Sections.
ii. Answer ALL Questions from Section A and THREE Questions from section B.
iii. Questions 6 and 7 are COMPULSORY

SECTION A (24 Marks)

Question One

a) Complete the bellow paragraph with appropriate answers.

The development of DBMSs' helped to fully achieve (i)..... The most important objective of DB technology is (ii)..... and not (iii)..... Distributed database systems mainly aim at achieving (iv).....and (v)..... A DDBMS supports the formation (creation) and (vi) ....., where data are stored at different sites connected through a network. (3 Marks).

b) Carefully study figure 1.1 bellow and answer the questions that follow;

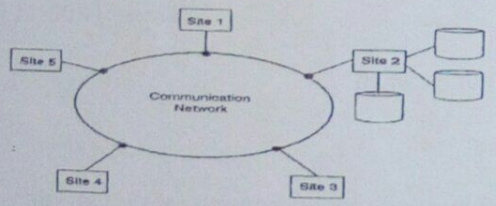

Figure 1.1

i) What type of database is depicted in figure 1.1 above? (0.5 Marks).

ii) Clearly state the reason(s) for your answer in (i) above. (1 Mark).

c) In developing a distributed database, one of the first questions to answer is where to grant system access. According to "*Security Implications of the Choice of Distributed Database Management System Model: Relational vs. Object-Oriented*", Security Problems Unique to Distributed Database Management Systems can be described under two main strategies. Briefly, describe the two strategies. (2 Marks).

d) Which of the two strategies in 1.c) above is easier to handle than the other and why?. (1.5 Marks).

Question Two

a) What are the four (4) main promises of DDBSs'? (2 Marks).

b) Mention any six (6) forms of transparency as used in DDBMSs'. (3 Marks).

c) Briefly, describe any three (3) complicating factors in DDBSs'. (3 Marks).

Question Three

a) Clearly, describe at least three (3) implicit assumptions considered in DDBSs'. (3 Marks).

b) Match the items in List A with corresponding items in List B. (5 Marks).

No.	List A	Lette	List B
i.	Any relation R with R1,R2, . . . ,Rn fragments such that each data item in R can also be found in some Ri.	a)	Query transformation.
ii.	A process that transforms a high-level query into an equivalent and more efficient lower-level query.	b)	Query execution engine.
iii.	Programs storing data in regular files.	c)	Concurrency
iv.	Heterogeneous DDBMSs'	d)	Data processing design
v.	Takes an (optimal) evaluation plan, executes that plan, and returns the answers to the query.	e)	huge overhead, error prone and high running costs.
vi.	Approaches to data processing in DDBSs'.	f)	Distribution design.
vii.	Frequency of queries, site, where query is run, selectivity of the queries, are important information in fragmentation	g)	Decomposition reconstruction.

	decision.		
viii.	The crucial and difficult part of the overall query processing.	h)	Extreme overhead and error prone.
ix.	Separation of the higher-level semantics of the system from the lower-level implementation issues.	i)	Different Sites implement their own databases before integration is considered.
x.	Central part of the design in DDBMSs'.	j)	Qualitative information.
		k)	Decomposition completeness.
		l)	Query execution.
		n)	Database systems and computer networks.
		o)	Query evaluation engine.
		p)	Different Sites implement their own databases after integration has been considered.
		q)	Query processing.
		r)	Transparency
		s)	Data processing systems and computer networks.
		t)	Quantitative information.
		u)	Query optimization.

Note: Provide your answers in a tabular form as shown in an example below;

i.	ii.	iii.	iv.	v.	vi.	vii.	viii.	ix.	x.
					s				

SECTION B (36 Marks)

Question Four

With the help of well drawn and labeled diagrams, briefly describe;

i) The ANSI/SPARC Architecture of DBMSs'. (4 Marks).

ii) The client-server architecture for DDBMSs'. (4 Marks).

iii) Peer-to-Peer architecture for DDBMSs'. (3.5 Marks).

iv) Which one of the above architectures do you think is the base of DDBMSs'? (0.5 Mark).

Question Five

a) What do you consider to be the most important process in "query processing"? (1.5 Marks).

b) Query processing consists of four (4) main steps. In the right order, mention them. (2 Marks).

c) Briefly, describe each of the steps you have mentioned in (b) above. (4 Marks).

d) Consider the database schema below;

EMP(ENO, ENAME, TITLE)

PROJ(PNO, PNAME, BUDGET)

ASG(ENO, PNO, RESP, DUR)

Given the following query (Q1): "Find the names of employees who have been working on project P3 for 8 or 11 months".

i) Provide an SQL statement for the query Q1 above. (1 Mark).

ii) Provide the query qualification in conjunctive normal form. (1 Mark).

iii) Provide the query qualification in disjunctive normal form. (1 Mark).

d) Differentiate between "Query/Connection graph" and "Join graph" as used in the analysis phase of query decomposition. (1.5 Marks).

Question Six

a) Query processing is described as a three (3) steps process that

transforms a **high-level query** into an equivalent and more efficient **lower-level query**. Clearly, describe those three (3) steps. (4.5 Marks).

b) What are the three (3) main factors which make query processing much more difficult in distributed environment when compared to centralized environment? (3 Marks).

c) Consider query (Q2) below being generated from the database schema provided in question 5 above;

```
SELECT ENAME,RESP
FROM EMP, ASG, PROJ
WHERE EMP.ENO = ASG.ENO
AND PNAME = "CAD/CAM" AND DUR>=23 AND
TITLE = "Analyst"
```

i) Sketch the query graph for Q2. (1.5 Marks).

ii) Is Query Q2 semantically correct? (0.5 Marks).

iii) State the reason(s) for your answer in 7 (c) ii. (1 Mark).

iv) What are the three (3) possible solutions to be taken under circumstances in 7 (c) ii.? (1.5 Marks).

Question Seven

a) In your opinion, what do you consider to be a reliable DDBMS? (1 Mark).

b) Recovery algorithms have two main parts. Briefly, describe those two parts. (2 Marks).

c) What are the three (3) most problematic issues in distributed transactions? (1.5 Marks).

d) Briefly and clearly, describe the protocols associated with each of the three (3) issues you have provided in 7 (c) above. (4.5 Marks).

e) Clearly, study the two figures below;

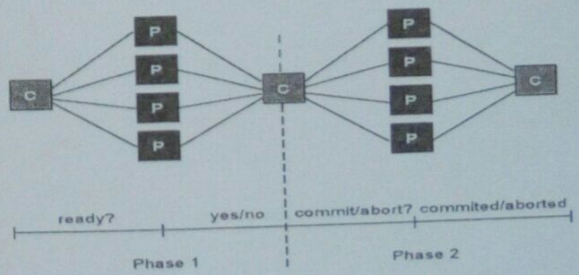

Figure 1.2

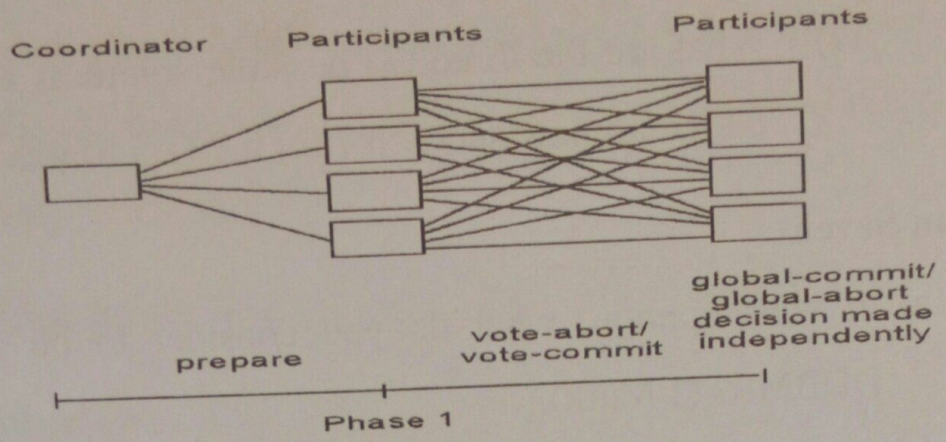

Figure 1.3

i) Which protocol does figure 1.2 above depict? (1 Mark).

ii) Which protocol does figure 1.3 above depict? (1 Mark).

iii) Briefly, differentiate between the two protocols you have mentioned in 8(e) i. & ii. (1 Mark).

THE UNIVERSITY OF DODOMA

COLLEGE OF INFORMATICS AND VIRTUAL EDUCATION

SCHOOL OF INFORMATICS

DEPARTMENT OF COMPUTER SCIENCE

CS300 DISTRIBUTED DATABASE SYSTEMS TEST

SEMESTER I

2012/2013 ACADEMIC YEAR

Time: 11:40-12:30 HRS

Duration: 50 Minutes

Date: 15th January, 2013

Instructions:

i. This Test Consists of 4 Questions.
ii. Answer ALL Questions.
iii. Provide your answers clearly to avoid some inconvenience during marking.
iv. Make sure you follow the instructions for each question.
v. Write your name, registration number and degree programme on each of your answer sheets.

Question One

a) "Query processing in multi-database systems is more complex than in distributed database systems". With at least three (3) reasons, justify the validity of that statement. **(1.5 Marks)**.

b) Differentiate between "mediator" and "wrapper" sites as used in multi-database systems. **(1.0 Mark)**.

Question Two

a) What do you consider to be the fundamental design problem of distributed database systems? **(1.0 Mark)**.

b) Describe the "Correctness Rules" which are normally applied to ensure that the database does not undergo semantic change during fragmentation. **(3.0 Marks)**.

c) Consider the below fragments;

FG

STP	PPR	PPZ	PPD	PPS	PPY
T/UDOM/2009/1986	IDD	MWAIPOSA	1982	M	TANZANIAN
T/UDOM/2010/01919	MICHAEL	DOWSON	1984	M	BRITISH
T/UDOM/20011/0122	SAMI	KHEDIRA	1990	M	GERMAN

FX

ATT1	ATTA
B1245	IPAD
B2225	WINDOWS
B3345	GOOGLECROME
B6677	LCD_TV
B8899	LAPTOP

FT

ATT1	BTX
B1245	LONDON
B2225	TEXAS
B3345	NEW YORK
B6677	BEIJING
B8899	TEXAS

FR

ATT1	BTR	BTY
B1245	APPLE INC	2000
B2225	MICROSOFT	1982
B3345	GOOGLE	2008
B6677	SAMSUNG	2011
B8899	MICROSOFT	2012

FQ

STP	PPR	PPZ	PPD	PPS	PPY
T/UDOM/20010/1114	JAMES	BALOTELLI	1992	M	ITALIAN
T/UDOM/2009/1990	IDRISSA	KAIJAGE	1987	M	TANZANIAN
T/UDOM/2010/1991	WANJIKU	KIPROCHI	1988	F	KENYAN
T/UDOM/2009/1992	VIBROS	JOHN	1985	F	UGANDAN

i. From the fragments above, how many relations do you think can be formulated from those fragments. (0.5 Marks).

ii. Formulate the corresponding relation (s). (1.0 Mark).

iii. Suppose you're required to design a distributed database having only three sites (site1, site2 & site3). By considering only vertical fragments, describe briefly how the fragments will be distributed at the three sites to provide:

A. Easy query processing and very high reliability. (1.5 Marks).

B. Easy concurrency control and low reliability. (1.5 Marks).

Question Three

You are provided with three relations;

Student(Reg#, FName, Surname, Gender, Nationality)

Stud_Course(Reg#, Course_ID, Course_Status, Semester)

Course(Course_ID, Course_Name, Course_Description)

Consider the query below;

"Find the surnames of students enrolled for any course in the ongoing 2012/2013 semester 1". (Note: For on going courses, the course status must be "On Going" and students undertaking courses in 2012/2013 semester I are those enrolled at UDOM from 2009/2010 to 2012/2013).

i) Provide a high level query from the above query. (0.5 Marks).

ii) Using low level queries, provide two possible transformations of the query in Question Three (i) above. (1.0 Mark).

iii) With reason (s), state which of the two transformations you've provided is better than the other. (0.5 Marks).

iv) Think of the following assumptions about data fragmentation;

Data is (horizontally) fragmented:

Site1: Stud_Course1= Reg#≤"T/UDOM/2008"(Stud_Course)

Site2: Stud_Course2= Reg#>"T/UDOM/2008"(Stud_Course)

Site3: Student1= Reg#≤"T/UDOM/2008"(Student)

Site4: Student2= Reg#>"T/UDOM/2008"(Student) and Site5: Result

Relations Student and Stud_Course are fragmented in the same way. (Note: all Students in Stud_Course relation with Course_Status="On Going" have enrolled for at least one course in the current 2012/2013 semester I).

By considering the transformation you have stated in Question Three (iii) above, provide two equivalent distributed execution strategies you would consider in processing the query provided in Question Three (2.5 Marks).

Question Four

a) Provide four (4) main characteristics of the distributed system R* Algorithm. (2.0 Marks).

b) Clearly, describe the five (5) steps of how the "Hill-Climbing query optimization" algorithm works. (2.5 Marks).

c) Draw a well labelled diagram which describes important "Query Optimization Processes". (2.0 Marks).

d) Briefly, describe five (5) additional issues brought by distributed query optimization. (2.5 Marks).

THE UNIVERSITY OF DODOMA

COLLEGE OF INFORMATICS AND VIRTUAL EDUCATION

SCHOOL OF INFORMATICS

UNDERGRADUATE UNIVERSITY EXAMINATION

FIRST SEMESTER 2013/2014

CS 300: DISTRIBUTED DATABASE SYSTEMS

Date: 5th February, 2014.

Time Allocated: 3 Hours

Instructions:

1. *This Exam Consists 7 Questions Organized in Two (2) Sections.*
2. *Answer **ALL** Questions in **Section A** and **TWO** in **Section B**.*
3. *Question **Five (5)** is **Compulsory**.*
4. ***ALL** University of Dodoma Examination **Regulations** Apply.*

SECTION A [40 Marks]

Question One

a) "Query processing in multi-database systems is more complex than in distributed database systems (DDBS)". With at four (4) reasons, justify the validity of that statement. **[6.0 Marks]**

b) Differentiate between "mediator" and "wrapper" sites as used in multi-database systems. **[3.0 Marks]**

Question Two

a) What do you consider to be the "Fundamental design problem of DDBS" and "Reliable DDBMS"? **[4.0 Marks]**

b) With aid of diagrams, differentiate between "Linear Join Tree" and "Bushy Join Tree" as used in query optimization. **[6.0 Marks]**

Question Three

As far as distributed query optimization is concerned, clearly, describe two (2) strategies used to scan the search space. **[6.0 Marks]**

Question Four

a) What is the main benefit of using semi-joins in distributed database systems? **[2.0 Marks]**

b) Consider query Q1 that has to be generated from UDOM-CIVE Database Schema provided on page 3:

Q1: Find the Surnames of students undertaking "Development Studies" course.

i. Provide the join graph for Q1 such that the relations/fragments will be located as follows; Student at Site1, Stud_Course at Site2 and Course at Site3. **[3.0 Marks]**

ii. As far as query optimization and join ordering in fragment queries are concerned, provide at least five (5) different ways by which Q1 can be evaluated. **[5.0 Marks]**

iii. With reference to Q1, provide five (5) important information to be considered in selecting an optimal evaluation plan. **[5.0 Marks]**

SECTION B [60 Marks]

Question Five

You are provided with UDOM-CIVE Database Schema that contains three relations;

Student (Reg#, FName, Surname, Gender, Nationality)
Stud_Course (Reg#, Course_ID, Course_Status, Semester)
Course (Course_ID, Course_Name, Course_Description)

Consider the query below;

a) Find the surnames of students enrolled for any course for the ongoing 2012/2013 semester I? (**Note**: For on going courses, the course status must be "On Going" and students undertaking courses in 2012/2013 semester I are those enrolled at UDOM from 2009/2010 to 2012/2013).

i) Provide a high level query from the above query. **[1.5 Marks]**

ii) Using low level queries; provide two possible transformations of the query in **Question Five (i)** above. **[3.0 Marks]**

iii) With reason (s), state which of the two transformations you've provided is better than the other. **[1.0 Mark]**

Consider the following assumptions about the data fragmentation

Data is (horizontally) fragmented:

Site1:Stud_Course1=Reg#≤"T/UDOM/2008"(Stud_Course)
Site2:Stud_Course2=Reg#>"T/UDOM/2008"(Stud_Course)
Site3: Student1= Reg#≤"T/UDOM/2008"(Student)
Site4: Student2= Reg#>"T/UDOM/2008"(Student)
Site5: Result

Relations Student and Stud_Course are fragmented in the same way. (Note: all Students in Stud_Course relation with

Course_Status="On Going" have enrolled for at least one course in the current 2012/2013 semester I).

b) By considering the transformation you have stated in **Question Five (iii)** above, provide two (2) equivalent distributed execution strategies you would consider in processing the query provided in **Question Five.** **[8.0 Marks]**

c) Calculate the cost of the two strategies stated in **Question Five a)** above under the following assumptions: **[12.0 Marks]**

- Tuples are uniformly distributed to the fragments;
- 35 tuples satisfy *Course_Status="On Going"*
- Size(Student) = 550
- Size(Stud_Course) = 2500
- Tuple access cost = 2 units; tuple transfer cost = 25 units
- Student and Stud_Course have a local index on Reg# and *Course_Status*

d) What are the three (3) main factors which make query processing much more difficult in distributed environment when compared to centralized environment? **[4.5 Marks]**

Question Six

a) Briefly, describe your understanding on "Query Optimization". **[3.0 Marks]**

b) Provide four (4) main characteristics of the distributed system R* query optimization Algorithm. **[12.0 Marks]**

c) Clearly, describe the five (5) steps of how the "Hill-Climbing query optimization" algorithm works. **[15.0 Marks]**

Question Seven

With the help of well drawn and labeled diagrams, describe;

a) The ANSI/SPARC Architecture of DDBMSs'. **[10.5 Marks]**

b) The client-server architecture for DDBMSs'. **[7.5 Marks]**

c) Peer-to-Peer architecture for DDBMSs'. **[10.5 Marks]**

d) Which one of the above architectures do you think is the base of DDBMSs'? **[1.5 Marks]**

THE UNIVERSITY OF DODOMA

COLLEGE OF INFORMATICS AND VIRTUAL EDUCATION

SCHOOL OF INFORMATICS

UNDERGRADUATE UNIVERSITY EXAMINATIONS

FIRST SEMESTER 2014/2015

CS 310: DISTRIBUTED DATABASE SYSTEMS

Date: 13th February, 2015

Time Allocated: 3 Hours

Instructions:

1. *This question paper consist 7 Questions organized in two* ***Sections***.
2. *Answer ALL Questions in* ***Section A*** *and* ***TWO*** *in* ***Section B***.
3. *Question* ***Seven (7)*** *is* ***Compulsory***.
4. *ALL University of Dodoma Examination Regulations Apply.*

SECTION A (40 Marks)

Question One

Describe the "Correctness Rules" applied to ensure that the database does not undergo semantic change during fragmentation. **[6 Marks]**

Question Two

As a key component of a cost-effective distributed database system, Sybase replication server can be used to addresses three (3) critical issues.

a) Briefly, describe them. **[3 Marks]**

b) What are the three (3) key areas within which high availability should be employed to address data availability? **[3 Marks]**

Question Three

a) What do you consider to be the "Fundamental design problem of DDBS" and "Reliable DDBMS"? **[4 Marks]**

b) Clearly, provide five (5) distributed database systems organizations. **[5 Marks]**

Question Four

a) You have been asked by TBL to set up a replication server for its three new distributed branches in Tanzania with Dar es Salaam being the Headquarters (i.e. order entry site: site 1), Tanga (deals with sales: site 2), Mbeya (deals with financial management: site 3) and Arusha (deals with manufacturing: site 4) being the three new branches. With the help of well labeled diagram; briefly, describe how you could set up a replication server environment by providing a **Decision Support Replicates** configuration. **[5 Marks]**

b) Briefly, describe three (3) types of sites which can be involved in the design of a distributed database system that uses a Replication Server. **[6 Marks]**

Question Five

a) Query Optimization is an important process for both centralized and distributed systems. However, there are some additional issues brought by distributed query optimization. Briefly, provide at least four (4) of those issues. **[4 Marks]**

b) Describe two (2) main strategies used to scan the search space **[4 Marks]**

SECTION B (60 Marks)

Question Six

You are provided with UDOM-CIVE Database Schema that contains three relations;

Student (Reg#, FName, Surname, Gender, Nationality)
Stud_Course (Reg#, Course_ID, Course_Status, Semester)
Course (Course_ID, Course_Name, Course_Description)

Consider the query below;

a) Find the surnames of students enrolled for any course for the ongoing 2012/2013 semester I? (**Note**: For on going courses, the course status must be "On Going" and students undertaking courses in 2012/2013 semester I are those enrolled at UDOM from 2009/2010 to 2012/2013).

i) Provide a high level query for the above query. **[1.5 Marks]**

ii) Using low level queries; provide two possible transformations of the query in (i) a) above. **[3 Marks]**

iii) With reason (s), state which of the two transformations you've provided is better than the other. **[1 Mark]**

Consider the following assumptions about the data fragmentation Data is (horizontally) fragmented:

Site1: Stud_Course1= Reg#≤"T/UDOM/2008" (Stud_Course)
Site2: Stud_Course2= Reg#>"T/UDOM/2008" (Stud_Course)
Site3: Student1= Reg#≤"T/UDOM/2008" (Student)
Site4: Student2= Reg#>"T/UDOM/2008" (Student)
Site5: Result

Relations Student and Stud_Course are fragmented in the same way. (Note: all Students in Stud_Course relation with Course_Status="On Going" have enrolled for at least one course in the current 2012/2013 semester I).

b) By considering the transformation you have stated in **Question Six (iii)** above, provide two (2) equivalent distributed execution strategies you would consider in processing the query provided above. **[8 Marks]**

c) Calculate the cost of the two strategies stated in **Question Six** a) above under the following assumptions: **[10 Marks]**

- Tuples are uniformly distributed to the fragments;
- 35 tuples satisfy *Course_Status="On Going"*

- Size(Student) = 550, Size(Stud_Course) = 2500
- Tuple access cost = 2 units; tuple transfer cost = 25 units
- Student and Stud_Course have a local index on Reg# and *Course_Status*

d) With the aid of a well drawn and labeled diagram illustrating the components of the Sybase Replication Server product and their functions, provide four (4) operations areas of replication system. **[6.5 Marks]**

Question Seven

a) With the help of a well drawn and labeled diagram for Distributed DBMS architecture depicting local transaction steps in a distributed environment; provide step-by-step the five (5) main local transaction steps. **[12.5 Marks]**

b) With the help of a well drawn and labeled diagram for Distributed DBMS architecture depicting Global transaction steps in a distributed environment; provide step-by-step the eight (8) main Global transaction steps. **[17.5 Marks]**

Question Eight

a) By providing at least four (4) significant trade-offs in each; differentiate between synchronous and asynchronous distributed databases. **[12 Marks]**

b) With the aid of well labeled diagrams; provide at least four (4) requirements for each of the environments below:

i. Homogeneous, Non-Autonomous Distributed Database Environment. **[9 Marks]**

ii. Heterogeneous Distributed Database Environment. **[9 Marks]**