Advanced Database Systems (CMP 401)
Final Exam – 4th Year - Spring 2009
(2 Hours) - Total Marks: 70

Question 1 (15)	Question 2 (10)	Question 3 (15)	Question 4 (10)	Question 5 (10)	Question 6 (10)	Total (70)

Question 1: [15 points]

1.a) Suppose that a disk unit has a block size B=2400 bytes. An EMPLOYEE file has the following fields:

SSN, 9 bytes; NAME, 30 bytes; BIRTHDATE, 10 bytes; ADDRESS, 45 bytes; PHONE, 12 bytes; Gender 1 byte; SUPERVISORSSN, 9 bytes; DEPARTMENT, 4 bytes; deletion marker, 1 byte. The EMPLOYEE file has r=30000 STUDENT records, fixed-length format, and unspanned blocking. Write down appropriate formulas and calculate the following values for the above EMPLOYEE file:

1) [2 pt] Calculate the wasted space in each disk block because of the unspanned organization.

2) [2 pt] Suppose we use a spanned record organization, where each block has a 5-byte pointer to the next block (this space is not used for record storage). Calculate the number of disk blocks.

3) [2 pt] Given that index record size = 10 bytes, Calculate the average number of block accesses to run the query COUNT(SSN) knowing that SSN is the primary index in this table.

4) [2 pt] Given that index record size = 10 bytes, calculate the number of levels in case of a multi-level primary index.

1.b) [4 pt] Given the following relations for the entities Professor and Course:

Professor (PID, Name, DeptID)
Course(Code, CName, DeptID, PID,Semester)

Given the following SQL query Q, draw the initial query tree and then show how the query
tree is optimized using heuristics.

SELECT C.CName, C.Semester
FROM Professor P, Course C
WHERE P.PID = C.PID AND P.Name = "Sayed"

1.c) [3 pt] Consider the B+-tree shown below with order p=4 and p leaf = 4, show how the
tree will expand and what the final tree looks like when inserting letter N.
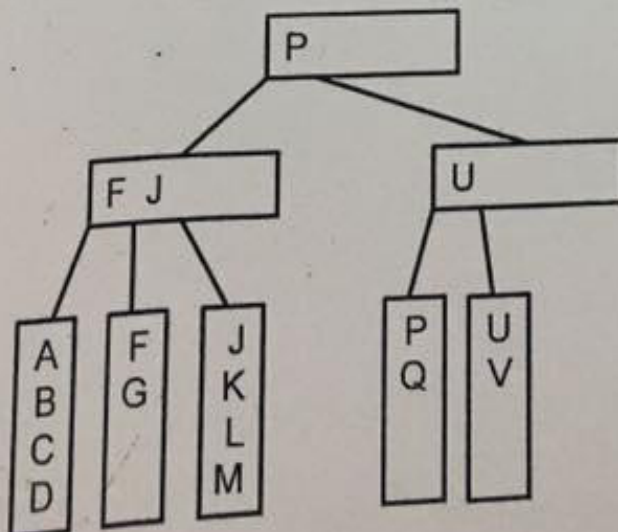
Question 2: [10 points]

2.a) [6 pt] Differentiate between rollback and historical database by giving an example (table) for each. Give two examples of the data that can be retrieved only if you have bi-temporal database?

2.b) [4 pt] Assume that the number of desired clusters is 2, show how the k-Means will cluster the following 2D data. (Hint: RID is not considered).

RID	Age	Yeas of Experience
1	30	5
2	50	25
3	50	15
4	25	5

Question 3: [15 points]

3.a) [3 pt] What is the meaning of Transaction? Give an example for a transaction.

3.b) [4 pt] Give an example for two of the three problems that may occur when concurrent execution is uncontrolled.

3.c) [2 pt] For the following serial schedule, can you find a serializable conflict equivalent schedule.

r1(x) w1(x) r1(y) w1(y) r2(x) w2(x) r2(y) w2(y)

3.d) [6 pt] Consider the three transactions T1, T2, and T3, and the schedules S1 and S2 given below. Draw the serializibility (precedence) graphs for S1 and S2 and state whether each schedule is serializable or not. If a schedule is serializable, write down the equivalent serial schedule(s).

T1: r1(x); r1(z); w1(x)
T2: r2(z); r2(y); w2(z); w2(y)
T3: r3(x); r3(y); w3(y)

S1: r1(x); r2(z); r1(x); r3(x); r3(y); w1(x); w3(y); r2(y); w2(z); w2(y)
S2: r1(x); r2(z); r3(x); r1(z); r2(y); r3(y); w1(x); w2(z); w3(y); w2(y)

Question 4: [10 points]

4.a) [4 pt] Why in the Multi-version Two-Phase Locking Algorithm, the abort of transaction T occurs if a version with W_TS ≤ TS(T) has a R_TS > TS(T)? Support your answer with an example of a schedule.

4. b) [2 pt] Explain how certify locks improves concurrency.

4.c) [2 pt] What are the common things between Wound-Wait and Wait-Die scheme?

4.d) [2 pt] How Wound-Wait and Wait-Die algorithms avoid starvation of transactions.

Question 5: [10 points]

5.a) [4 pt] List and explain briefly two of the three main principles of the ARIES recovery algorithm. What is the reasoning behind each principle?

5.b) [6 pt] For recovery in ARIES, study the shown log at time of crash, show the transaction table and the dirty page table at time of check point and after the analysis phase of the recovery process. Explain what will happen in the other two phases of the recovery process.

LSN	Last LSN	Trans Id	Type	Page Id
1	0	T1	Update	A
2	0	T2	Update	B
3	2	T2	Commit	
4	Begin checkpoint			
5	End checkpoint			
6	1	T1	Commit	
7	0	T3	Update	C
8	7	T3	Update	A

Question 6: [10 points]

6.a) [4 pt] What are the two restrictions enforced on data access based on the subject/object classifications? What is the philosophy behind each of them?

6.b) [2 pt]   Consider the relation shown below. How would it appear to a user with classification U? Suppose a classification U user tries to update the salary of "Ahmed" to 50,000; what would be the result of this action?

National ID	Name	Salary	Job Performance	Tuple Classification
27001012300054 C	Ahmed Hamdy U	40,000 C	Fair S	C
27211232100043 C	Yousef Abdo C	60,000 S	Good S	S

6.c) [4 pt] What types of queries are allowed in Statistical database security and what is not allowed?