SECTION: _____ NAME: _____ ROLL NUMBER: _____

ADVANCED DATABASES FINAL EXAM SUBJECTIVE PART

[Spring 2014] [Total Points: 60] [Time: 150 min.]

NOTE: Calculators are permitted. Please write your solutions in the spaces provided on the exam. You may use the blank areas and backs of the exam pages for scratch work. Please do not staple any additional scratch paper. Write your roll no in the upper right corner of every page.

Question 1 **(4points)**

Look at the schedule below. Here two transactions T1 and T2 are running. The statements are executed in the sequence as given below. No concurrency control is used.

T1: START TRANSACTION;
T2: START TRANSACTION;
T1: UPDATE product SET No_Items=3 WHERE Product_Id=1;
T2: SELECT No_Items FROM product WHERE Product_Id=1;
...

a) The given execution of the two transactions leads to a problem. What is this problem called like? (Just write the name of the problem.)

b) Which of the properties of a transaction which are guaranteed by the DBMS is violated? (Write the name of the property and one sentence why it is violated.)

Question 2 **(2+6= 8 points)**

Consider the schedule:

	T1	T2	T3
1	read(A)		
2		read(C)	
3			read(A)
4		write(C)	
5	read(C)		
6			read(B)
7		write(B)	

a) Show, by drawing the precedence graph, that the schedule is serializable.

b) Show that the schedule will be accepted/rejected by the below protocols:

i) The basic two-phase locking protocol (add locks to the transactions)

ii) The timestamp-ordering protocol (you have T1 < T2 < T3)

Question 3 (8 points)

Consider the following two transactions:
T1: read(B); (if B >=100 then B:=B-100); write(B).
T2: read(B); (if B >=100 then B:=B-100); write(B).
The consistency requirement is that (B>=0). Let B = 100 be the initial value.
a) Show that every serial execution involving T1 and T2 preserves consistency of the database.
b) Show a concurrent execution of T1 and T2 that produces a nonserializable schedule.
c) Is the final result of your nonserializable schedule satisfies the consistency requirement? Is it correct? Why or why not?
d) Add locks to the transactions and show how strict 2-phase locking would prevent your nonserializable execution.

Question 4 (4 points)

Assume a relation R (A, B, C) is given; R is stored as an ordered file (un-spanned) on non-key field C and contains 1000000 (1 million) tuples. Attributes A, B and C need 4 byte of storage each, and blocks have a size of 4000 Byte. Each A value occurs at an average 20 times in the database, each B value occurs 100 times in the database, and each C value occurs 100000 times in the database. Assume there is no index structure exists.

Estimate the number of block fetches needed to compute the following queries (where C_a and C_c are integer constants):

a) SELECT B, C FROM R WHERE A = C_a;

b) SELECT B, C FROM R WHERE C = C_c;

Question 5 (2+2= 4 points)

Consider the student table:

RollNo	Name	Address	Gender	Age	Grade
1001	Khadija	Faisal Town	F	16	B
1002	Tahreem	Model Town	F	16	C
1003	Isbah	DHA	F	18	A
1004	Izaan	Model Town	M	18	B
1005	Alia	Faisal Town	F	20	A
1006	Tahreem	DHA	F	17	B
1007	Ismail	Johar Town	M	19	A
1008	Izaan	DHA	M	17	D

List the name of extremely high selectivity column and extremely low selectivity column of the above table. One sentence why you choose the column.

Question 6 (2 points)

Suppose that the most often used query on the Customer database is:
SELECT Product_Name FROM product, buys WHERE product.Product_ID=buys.Product_ID;
On which column(s) would you create an index? Write down the column name(s) and one sentence why you choose the column(s).

Question 7 (4 points)

Assume a relation S (A, B, C) is given; S is stored as an un-ordered file (un-spanned) on key field A and contains 3000 data blocks. Assume there is B^+- tree access structure (index) on A of height x=4 (root, 2 intermediate layer, leaf). Moreover, one node of the B^+-tree is stored in one block on the disk.
Estimate the number of block fetches needed to compute the following queries (where C_a, C_{a2}, C_b and C_c are integer constants):

a) SELECT B, C FROM S WHERE A = C_a;
b) SELECT B, C FROM S WHERE A = C_a AND B = C_b;
c) SELECT B, C FROM S WHERE A = C_a OR A = C_{a2};
d) SELECT B, C FROM S WHERE A > C_a ;

Question 8 (6 points)

Figure below shows the log corresponding to a particular schedule at the point of a system crash for five transactions. Suppose that we use the immediate update (undo/redo) protocol with checkpointing. Describe the recovery process from the system crash. Assume that the initial values of items are A=10, B=20, and C=30. Isolation level of all transactions is READ COMMITTED.

a) Identify which transactions need undo/ redo operation(s)?
b) Specify which operations in the log are redone (in correct order) and which are undone.
c) Write down the values of items A, B, and C after system recovery.

[start_transaction, T1]
[read_item, T1, B, 20]
[start_transaction, T2]
[read_item, T2, A, 10]
[write_item, T2, A, 10, 5]
[read_item, T2, B, 20]
[write_item, T2, B, 20, 12]
[commit, T2]
[start_transaction, T3]
[read_item, T1, A, 5]
[write_item, T1, A, 5, 2]
[read_item, T3, C, 30]
[write_item, T3, C, 30, 25]
[checkpoint]
[commit, T3]
[start_transaction, T4]
[read_item, T4, C, 25]
[write_item, T4, C, 25, 22]
[start_transaction, T5]
[read_item, T5, C, 25]
[write_item, T5, C, 25, 21]
[commit, T5]
System crash

Question 9 **(5points)**

List four methods for implementing join operations and briefly explain one.

Question 10 **(5 points)**

List any four types of parameters that are used in cost functions. Where is this information kept?

Question 11 (10 points)

Draw the initial query tree for the query below, and then using heuristic rule, show query tree that is efficient to execute (i.e. optimized).

SELECT S.name FROM student S, course C, Takes T
WHERE S.id=T.Stu-Id AND C.number=T.course-No AND C.title='ADB' AND T.grade>'C';