Database Management Systems - Written Exam
Computer Science in English, 2nd year

I. Choose the correct answer(s) for multiple choice questions 1 to 11. Each question has at least one correct answer. For questions 12 and 13, write only the final result(s). **Enter all answers in file Answers.docx. Only the answers in this file will be considered for the grade.**

1. Consider schedule S below:

T1	**T2**	**T3**
	read(B)	
	B = B - 200	
	write(B)	
		read(A)
		A = A + 20
		write(A)
read(A)		
A = A – 50		
write(A)		
		read(B)
		B = B – 10
		write(B)
	read(A)	
	A = A + 50	
	write(A)	
read(B)		
B = B + 200		
write(B)		

time

Before the execution above, A = 600 and B = 700.

a. S is not serializable.
b. S is serializable. A serial schedule that's equivalent to S (in its effect on the database) is: T1 followed by T2 followed by T3.
c. S is serializable. A serial schedule that's equivalent to S (in its effect on the database) is: T2 followed by T1 followed by T3.
d. S is serializable. A serial schedule that's equivalent to S (in its effect on the database) is: T3 followed by T1 followed by T2.
e. None of the above answers is correct.

2. The Index Nested Loops Join algorithm is an instance of the:
a. iteration technique
b. partitioning technique
c. indexing technique
d. supervision technique
e. None of the above answers is correct.

3. Let R be a relation with 2000 pages. There are 50 pages in the buffer pool. R is sorted with *external merge sort*. Then:
a. Pass 0 produces 2 runs.
b. Each run produced in pass 0 has 2000 pages.
c. Pass 0 produces 40 runs.
d. Each run produced in pass 0 has 50 pages.
e. None of the above answers is correct.

4. The *atomicity* property of a transaction ensures that:
a. The transaction never acquires X locks.
b. The transaction never acquires S locks.
c. The transaction preserves the consistency of the database after execution.
d. The transaction is protected from the effects of concurrently scheduling other transactions.
e. None of the above answers is correct.

5. In the Two-Phase Locking protocol:
a. A transaction can write object O without acquiring an X lock on O.
b. A transaction can release locks before it completes execution.
c. A transaction can read object O without acquiring an S lock on O.
d. Once a transaction releases a lock, it cannot request other locks.
e. None of the above answers is correct.

6. If a relation R is fragmented into 3 vertical fragments R1, R2, R3:
a. None of the 3 fragments can be replicated.
b. Only R1 and R2 can be replicated.
c. Only R1 and R3 can be replicated.
d. Only R2 and R3 can be replicated.
e. None of the above answers is correct.

7. Consider the following statement:

```
SELECT *
FROM table
WHERE column = v
```

Which of the following values for *v* indicates / indicate a SQL injection attack:
a. 5; INSERT INTO users VALUES('x', 'y')
b. 5; UPDATE users SET pswd = 'JHGrmO65v' WHERE user LIKE '%admin%'
c. 100
d. 3; DROP TABLE users
e. None of the above answers is correct.

8. I is an index with search key <A, B, C>. If I is a:
a. B+ tree index, I matches condition A=3.
b. hash index, I matches condition A=1 AND B=2 AND D=3.
c. hash index, I matches condition A=1 AND B=2 AND C=3.
d. B+ tree index, I matches condition B > 10.
e. None of the above answers is correct.

9. In SQL Server:
a. Under REPEATABLE READ, a transaction must acquire an exclusive lock to write an object.
b. Under REPEATABLE READ, a transaction doesn't need to acquire an exclusive lock to write an object.
c. *Dirty reads* can occur under SERIALIZABLE.
d. *Unrepeatable reads* can't occur under SERIALIZABLE.
e. None of the above answers is correct.

10. The reduction factor for condition *Age = 20*, assuming data is uniformly distributed and there is an index *I* on *Age*, can be estimated by:
a. 1/NKeys(I)
b. NKeys(I)
c. INPages(I)
d. 1/INPages(I)
e. None of the above answers is correct.

11. Consider schedule S below (all transactions commit):

T4	T5	T6
read(A)		
	read(B)	
read(C)		
	write(A)	
write(E)		
		read(C)
		write(D)
write(F)		
		read(A)

time

a. S is conflict serializable.
b. (read(T4, A), write(T5, A)) belongs to the conflict relation of S.
c. (write(T5, A), read(T6, A)) belongs to the conflict relation of S.
d. the following serial schedule is conflict equivalent with S: T4 followed by T5 followed by T6.
e. None of the above answers is correct.

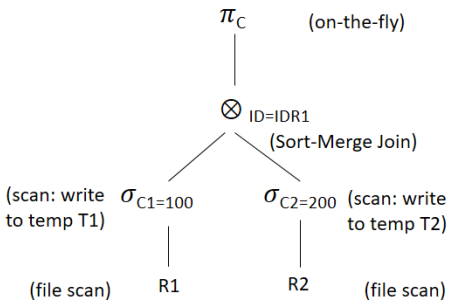
12. Consider the query:
SELECT *
FROM P, Q, S
WHERE T7 AND T5 AND T9 AND T10
The conditions tested by the terms in the WHERE clause are statistically independent.
The cardinality of relation R is denoted by |R|. The reduction factor associated with term T is denoted by RF(T).
|P| = 400, |Q| = 500, |S| = 6000, RF(T7) = 1/15, RF(T5) = 1/20, RF(T9) = 1/10, RF(T10) = 1/5

What's the estimated value for the cardinality of the query's result set (in a typical relational query optimizer)?

13. Let R1 and R2 be 2 relations. R1 has 3.000 records, with 75 records per page. R2 has 50.000 records, with 50 records per page.
a. Consider the following query:

```
SELECT R2.C
FROM R1, R2
WHERE R1.ID=R2.IDR1 AND R1.C1=100 AND R2.C2=200
```

Evaluate the cost of the plan below. The size of T1 is estimated at 30 pages, the size of T2 - at 90 pages. T1 and T2 are not sorted beforehand. There are 60 pages available in the buffer pool. For *sort-merge join*: T1 is the outer relation; use *external merge sort* to sort T1 and T2; assume each partition is scanned once during the merging phase.



b. Compute the cost of $R1 \otimes_{R1.ID=R2.IDR1} R2$ using *hash join*. Assume each partition fits into memory during probing.
c. 400 buffer pages are available. R2 is not sorted beforehand. Compute the cost of $\pi_{ID,C,D}(R2)$ using *projection based on sorting* (basic version, without improvement). Use *external merge sort.* The size of a tuple in the result of $\pi_{ID,C,D}(R2)$ is 1/4 times the size of a tuple in R2.

II. Think of an application that's powered by a relational database. In this context:

a. Draw the database diagram (at least 4 interrelated tables, with primary keys and foreign keys).

b. Describe one real-world scenario in which different users are trying to access and manipulate the data concurrently, along with the problems that can occur.

c. Write a query with one SELECT clause, one FROM clause, one WHERE clause and one JOIN clause. The SELECT clause must contain two columns. The WHERE clause must have two terms of the form *AttributeName Operator Value*, where *Operator* $\in \{<, \leq, =, >, \geq, <>\}$; the terms must be connected by AND. Draw an evaluation plan for the query and describe the query's evaluation (with concrete algorithms and costs).