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 (all transactions commit):

T1	T2	T3
R(A)		
	W(F)	
R(D)		
		W(B)
	R(A)	
	W(A)	
W(E)		
	R(D)	
		W(A)
		R(C)
		R(B)
		R(D)

a. S is conflict serializable.
b. S is not conflict serializable.
c. (R(T1, A), R(T2, A)) belongs to the conflict relation of S.
d. (R(T1, A), W(T2, A)) belongs to the conflict relation of S.
e. None of the above answers is correct.

2. T1 and T2 are 2 concurrent transactions (there are no other concurrent transactions). The final result of their execution must be identical to the result obtained when executing:
a. either transaction T1 or transaction T2, but not both
b. T1 followed by T2 or T2 followed by T1
c. only transaction T1
d. only transaction T2
e. None of the above answers is correct.

3. In vertical fragmentation:
a. fragmentation is performed with the projection operator.
b. fragmentation is performed with the set-difference operator.
c. the reconstruction operator is *set-difference*.
d. the reconstruction operator is *the natural join*.
e. None of the above answers is correct.

4. I is an index with search key <a, b>.
a. If I is a hash index, I matches condition *a=8 AND b=7*.
b. If I is a hash index, I matches condition *b = 5 AND a = 3*.
c. If I is a hash index, I matches condition a=9 AND b<10 AND c<2 AND d<7.
d. If I is a B+ tree index, I matches condition c=9 AND b>10.
e. None of the above answers is correct.

5. Under the READ COMMITTED isolation level:
a. *phantom reads* can't occur
b. *nonrepeatable reads* can't occur
c. *dirty reads* can occur
d. *dirty reads* can't occur
e. None of the above answers is correct.

6. Which of the algorithms below is using the indexing technique:
a. page-oriented nested loops join
b. hash join
c. sort-merge join
d. index nested loops join
e. None of the above answers is correct.

7. In the context of transaction processing, the acronym ACID stands for:
a. atomicity, constituency, idealism, derivability
b. atomicity, consistency, integrity, determinacy
c. atomicity, constituency, isolation, durability
d. atomicity, consistency, indeterminacy, derivability
e. None of the above answers is correct.

8. The reduction factor for condition *Salary > 3000*, assuming data is uniformly distributed and there is an index *I* on *Salary*, can be estimated by:
a. (IHigh(I) – ILow(I)) / 10
b. 10 / (IHigh(I) – ILow(I))
c. (IHigh(I) – 3000) / (IHigh(I) – ILow(I))
d. (IHigh(I) – ILow(I)) / (IHigh(I) – 3000)
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't occur under SERIALIZABLE.
d. *Unrepeatable reads* can occur under SERIALIZABLE.
e. None of the above answers is correct.

10. Let T be a relation with Q pages. The cost of sorting T using *external merge sort* with F pages in the buffer pool is:
a. $2*F*\left(\left\lceil \log_{Q-1}\left\lceil \frac{F}{Q}\right\rceil\right\rceil + 1\right)$ I/Os
b. $2*F*\left(\left\lceil \log_{F-1}\left\lceil \frac{Q}{F}\right\rceil\right\rceil + 1\right)$ I/Os
c. $\lceil \log_F Q \rceil$ +1 I/Os
d. $\lceil \log_Q F \rceil$ +1 I/Os
e. None of the above answers is correct.

11. Consider schedule S below:

T1	T2	T3
read(A)		
A = A – 100		
write(A)		
	read(B)	
	B = B - 100	
	write(B)	
		read(A)
		A = A + 10
		write(A)
read(B)		
B = B + 200		
write(B)		
	read(A)	
	A = A + 50	
	write(A)	
		read(B)
		B = B – 10
		write(B)

time

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.

Before the execution above, A = 300 and B = 400.

12. Encode the data *the day you almost caught jack sparrow* using the secret encryption key *carlsagan* and the table of codes below. Let *M* be the obtained string. Write the substring of *M* that consists of characters on positions <12, 13, 14>. The first character in *M* is on position 1.

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	-
00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27

13. Let T1 and T2 be 2 relations. T1 has 700.000 records; a page can hold 200 T1 records. T2 has 200.000 records; a page can hold 160 T2 records.
a. 202 buffer pages are available. Compute the cost of $T2 \otimes_{T2.ID=T1.IDT2} T1$ using *block nested loops join*. T2 is the outer relation.
b. 202 buffer pages are available. Compute the cost of $T2 \otimes_{T2.ID=T1.IDT2} T1$ using *sort-merge join*. T1 and T2 are not sorted beforehand. T2 is the outer relation. Use *external merge sort* to sort T1 and T2. Assume each partition is scanned once during the merging phase of *sort-merge join.*
c. T1 is stored at Cluj-Napoca, T2 is stored at Bucharest. Compute the cost of $T2 \otimes_{T2.ID=T1.IDT2} T1$ using *page-oriented nested loops join* in Bucharest, without caching. T2 is the outer relation, the query site is Timișoara and the result of $T2 \otimes_{T2.ID=T1.IDT2} T1$ has 3000 pages. Use t_d to denote the time to read / write a page from / to disk; use t_s to denote the time to ship a page from one site to another.

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 that reproduces the deadlock phenomenon.

c. Write a query whose evaluation plan contains a Sort-Merge Join. Draw the evaluation plan and describe the query's evaluation (with concrete algorithms and costs).