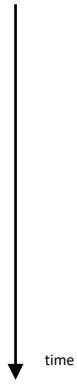
Database Management Systems - Written Exam - I
Computer Science in English, 2nd year

I. 1. Define the notion of *schedule*. **2p**
2. Give a complete description of the *Index Nested Loops Join* algorithm (including the cost computation) using an example.

II. Solve the following problems: **4p**
1. Consider schedule S below (all transactions commit): (1p)

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)



Compute the conflict relation of S.
Is S conflict serializable? Justify your answer.
If S is conflict serializable, find a serial schedule S_{ser} such that $S \equiv_c S_{ser}$.

2. Express the SQL query below in the extended relational algebra, using a σ π $\times$ expression. Include only the final expression in the .pdf file. (1p)

```
SELECT A.IDA, MIN(C.Value2)
FROM A, B, C
WHERE A.IDA = B.IDA AND B.IDC = C.IDC AND A.Name = 'Ionescu' AND C.Value1 = 100
GROUP BY A.IDA
HAVING COUNT(*) <= 100
```

3. Let T1 and T2 be 2 relations. T1 has 500.000 records; a page can hold 100 T1 records. T2 has 100.000 records; a page can hold 100 T2 records. (2p)
a. 102 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. 102 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 10000 pages.

Provide only the final results for a, b and c in the .pdf file (don't include the computation).

III. Choose the correct answer(s) for the following 9 multiple choice questions. Each question has at least one correct answer. Fill in the encoded data for question 10. **3p**

1. In the context of transaction processing, the acronym ACID stands for:
a. atomicity, consistency, idealism, durability
b. atomicity, consistency, idiosyncrasy, durability
c. atomicity, cardinality, isolation, durability
d. acidity, consistency, isolation, durability
e. None of the above answers is correct.

2. Choose the correct answer(s):
a. *Dirty reads* can occur under READ UNCOMMITTED.
b. *Dirty reads* can occur under REPEATABLE READ.
c. *Unrepeatable reads* can occur under SERIALIZABLE.
d. *Unrepeatable reads* can occur under REPEATABLE READ.
e. None of the above answers is correct.

3. The reduction factor for condition *Age > 30*, assuming data is uniformly distributed and there is an index *I* on *Age*, can be estimated by:
a. INPages(I)
b. 1/INPages(I)
c. 1/NKeys(I)
d. NKeys(I)
e. None of the above answers is correct.

4. To prevent an SQL injection attack:
a. Data validation is performed using regular expressions.

b. Statements are parameterized.
c. String separators are preceded with "^&^".
d. Users are asked nicely not to commit an attack.
e. None of the above answers is correct.

5. The *Block Nested Loops Join* algorithm is an instance of the:
a. iteration technique
b. partitioning technique
c. indexing technique
d. None of the above answers is correct.

6. Choose the correct answer(s):
a. *Primary site replication* is an asynchronous replication technique.
b. *Primary site replication* is a synchronous replication technique.
c. *Read-any write-all* is a synchronous replication technique.
d. *Read-any write-all* is an asynchronous replication technique.
e. None of the above answers is correct.

7. Phantom deadlocks:
a. Are detected using the Wound-Wait policy.
b. Can lead to unnecessary aborts.
c. Are detected using the Wait-Die policy.
d. Can occur when there are delays in propagating local information.
e. None of the above answers is correct.

8. Consider a query with one relation in the FROM clause. Which of the following access paths can be used for the query?
a. single-index access path
b. multiple-index access path
c. sorted index access path
d. index-only access path
e. None of the above answers is correct.

9. In vertical fragmentation:
a. A fragment is a subset of columns.
b. A fragment is a subset of rows.
c. Two different fragments can be stored at different sites.
d. Fragments can't be replicated.
e. None of the above answers is correct.

10. Encode the data *the day you almost caught jack sparrow* using the secret encryption key *anaaslan* and the table of codes:

	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

(0.3p / question)

1p of