

I. 1. Define the *non-serial schedule*.

2p

2. Give a complete description of the *External Merge Sort* algorithm (including the cost computation) using an example.

II. Solve the following problems:

4p

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

(1p)

T4	T5	T6
R(A)		
	W(F)	
R(G)		
		W(B)
	R(A)	
		W(A)
	W(A)	
	R(G)	
W(N)		
		R(M)
		R(B)
		R(G)

time ↓

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. Consider the query:

(1p)

SELECT *

FROM A, B, C, D

WHERE T1 AND T2 AND T3

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 $RF(T)$.

$|A| = 200$, $|B| = 300$, $|C| = 100$, $|D| = 50$, $RF(T1) = 1/3$, $RF(T2) = 1/2$, $RF(T3) = 1/10$

What's the estimated value for the cardinality of the query's result set (in a typical relational query optimizer)? Provide only the final result in the .pdf file (don't include the computation).

3. Let R1 and R2 be 2 relations. R1 has 2.000 records, with 40 records per page. R2 has 100.000 records, with 200 records per page. (2p)

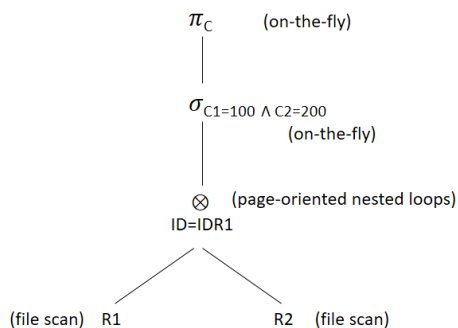
a. Consider the query below :

SELECT R2.C

FROM R1, R2

WHERE R1.ID=R2.IDR1 AND R1.C1=100 AND R2.C2=200

a1. Evaluate the cost of the plan below. R1 is the outer relation.



a2. R1 and R2 are replicated at Cluj-Napoca and Braşov. Describe one factor that must be taken into account when choosing the execution site for the query.

b. 200 buffer pages are available. Compute the cost of $R1 \bowtie_{R1.ID=R2.IDR1} R2$ using *sort-merge join*. R1 and R2 are not sorted beforehand. R1 is the outer relation. Use *external merge sort* to sort R1 and R2. Assume each partition is scanned once during the merging phase of *sort-merge join*.

For a1 and b, provide only the final results 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. A precedence graph for a schedule S contains:
 - a. A node for every aborted transaction in S.
 - b. A node for every committed transaction in S.
 - c. An arc from T1 to T2 if an action in T1 precedes and conflicts with one of the actions in T2.
 - d. An arc from T1 to T2 if an action in T2 precedes and conflicts with one of the actions in T1.
 - e. None of the above answers is correct.

2. Existing deadlocks can be detected using:
 - a. ARIES.
 - b. The Wait-Die policy.
 - c. The waits-for graph.
 - d. The Wound-Wait policy.
 - e. None of the above answers is correct.

3. 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. $\lceil \log_F Q \rceil + 1$ I/Os
 - b. $\lceil \log_Q F \rceil + 1$ I/Os
 - c. $2 * F * \left(\left\lceil \log_{Q-1} \left\lceil \frac{F}{Q} \right\rceil \right\rceil + 1 \right)$ I/Os
 - d. $2 * Q * \left(\left\lceil \log_{F-1} \left\lceil \frac{Q}{F} \right\rceil \right\rceil + 1 \right)$ I/Os
 - e. None of the above answers is correct.

4. The system catalogs in a DBMS maintain the following data about indexes and relations:
 - a. relation cardinality and size
 - b. index cardinality and size
 - c. index height
 - d. index range
 - e. None of the above answers is correct.

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

7. I is an index with search key $\langle K1, K2, K3 \rangle$. If I is a:
 - a. hash index, I matches condition $K1=2 \text{ AND } K2=3 \text{ AND } K3=1$.
 - b. hash index, I matches condition $K1=2 \text{ AND } K2=3 \text{ AND } K3>1$.
 - c. B+ tree index, I matches condition $K2 > 10$.
 - d. B+ tree index, I matches condition $K2 < 10$.
 - e. None of the above answers is correct.

8. A query block:
 - a. Has exactly one FROM clause.
 - b. Has at most one WHERE clause.
 - c. Has at most one GROUP BY clause.
 - d. Has at most one HAVING clause.
 - e. None of the above answers is correct.

9. Let T be a *linear* tree. Then for each join node N in T:
 - a. All children of N must be base relations.
 - b. The right child of N must be a base relation.
 - c. The left child of N must be a base relation.
 - d. The third child of N must be a base relation.
 - e. None of the above answers is correct.

10. Encode the data *non scholae, sed vitae discimus* using the secret encryption key *mariecurie* 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)