DI- FCT NOVA **June 5th, 2023**

Database Systems

Test-2 2022/23 – Version A

Duration: 2 hours (limited open book)

Group 1 (each question is worth 2,5 values out of 20)

Consider part of a database for a bus company for a European network covering 32 countries (where the attributes constituting the primary key are underlined):

Locations({<u>locCode</u>,name,country,population,...}) Routes({<u>numberR</u>,locOrigin,locDestination})

Trips({<u>numberR,date</u>,driverCode,licensePlate}) Drivers(<u>driverCode</u>,name,sex,birthdate,...)

For each of these tables there is a B+ tree non-clustering index on the primary key attribute(s), created with the column order indicated in the tables. If primary key attributes of a table T occur in a table O, then they are a foreign key in O referencing T.

The adopted DBMS uses blocks of 8KiB (8192 bytes). The records of all tables have a variable size, and, on average, a record of the locations table occupies 0.5KiB, a driver occupies 1KiB, a record of Routes and Trips occupy 128 bytes. At any given time, the locations table has 3,200 tuples, the routes table 12,000 tuples, the trips table 3,000,000 tuples and the drivers table 1,000 tuples. A B+ tree node can contain about 100 search keys, and it is known that a seek time is 10ms and the transfer time of a block t_T is 1ms, while the memory only holds 100 blocks.

Note: In this group, whenever examples are requested, these must be exclusively about this database, Additionally, all the answers must contain a brief justification.

1 a) Estimate the number of tuples returned by the following query, assuming that the route numbers are sequential and consecutive, starting at 1.

```
SELECT * FROM  Trips NATURAL INNER JOIN Routes
        WHERE numberR <= 10 AND date = '2023-05-31'
```

1 b) The transactions in the schedule below may be executed in READ UNCOMMITTED, READ COMMITED or SNAPSHOT ISOLATION modes. Assuming that the database tables are empty at the beginning, indicate which values are written by the DBMS_OUTPUT instructions knowing that n1 and n2 are numeric variables. Justify whether one of the previous modes corresponds to a serializable execution of the transactions.

Step	Transaction T_1:	Transaction T_2:

1		SELECT COUNT(*) INTO n2
		FROM Routes;

2		DBMS_OUTPUT.PUT_LINE(n2);
3		INSERT INTO Routes VALUES(12001,...);
4	INSERT INTO Routes(12002,...);	

5		SELECT COUNT(*) INTO n2
		FROM Routes;

6		DBMS_OUTPUT.PUT_LINE(n2);
7		COMMIT;

8	SELECT COUNT(*) INTO n1
	FROM Routes;	

9	DBMS_OUTPUT.PUT_LINE(n1);	
10	COMMIT;	

1 c) Using the example database present an unrecoverable schedule containing two transactions, as well as the necessary changes to make it recoverable and cascadeless.

1 d) Explicate the outcome of the timestamp ordering protocol applied to the following three transactions, where each transaction index is the timestamp of the start of the transaction. Assume that all read and write timestamps are initialized with the value 1.

	T_{107}	**T_{205}**	**T_{220}**

1	UPDATE Routes
	SET locOrigin = 2
	WHERE numberR= 2;		

2			SELECT locOrigin
			FROM Routes
			WHERE numberR = 3

3		SELECT locOrigin
		FROM Routes
		WHERE numberR = 2	

4	UPDATE Routes
	SET locOrigin = 7
	WHERE numberR= 3;		

1 e) Consider the following list of events involving three transactions (the only ones running on the system). Present the log record according to the studied algorithm knowing that at the beginning the locOrigin attribute is 0 for every Route tuple.

	T_1	**T_2**	**T_3**	**system**

1	begin transaction
	UPDATE Routes
	SET locOrigin = 1
	WHERE numberR = 1;			

2		begin transaction
		UPDATE Routes
		SET locOrigin = 2
		WHERE numberR= 2;		

			begin transaction
			UPDATE Routes
			SET locOrigin = 2
			WHERE numberR = 3;	

3	COMMIT;			
4				**CHECKPOINT**
5			ROLLBACK**;**	
6				**CRASH!**
7				**RECOVER**

Group 2 (each question is worth 2,5 values out of 20)

Note: the response to each of the items in this group cannot exceed one page.

2 a) Explain why some database management systems optimizers only consider left-deep join trees instead of arbitrary trees.

2 b) Describe what it is the write-ahead logging rule and in which conditions must be used.

2 c) Indicate the purpose of each of the phases of the two-phase commit algorithm (2 Phase Commit – 2PC), explaining the major drawback of this algorithm.