DI- FCT/UNL   **12 de junho de 2025**

Database Systems
2nd test 2024/25 – Version C
Duration: 2 hours (limited consultation)

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

Consider part of a database for a mobile phone operator that provides roaming services

Persons({personID,name,fiscalNR,countryID,...})   Countries({countryID,countryName,code})

Devices({deviceID,number,operatorID,personID})   Operators(operatorID,operatorName,countryID)

Registration(regID,deviceID,operatorID,start,end)   Calls(callID,regCaller,regDest,time,duration,cost)

Persons (personID as primary key) stores user data and includes a foreign key countryID referencing **Countries** (countryID as primary key), which holds country names and dial codes. **Operators** (operatorID as primary key) also reference **Countries** to indicate their home country. **Devices** (deviceID as primary key), and have foreign keys to personID and operatorID toPersons and Operators (respectively), indicating ownership and issuing operator of the device contract with the operator. **Registration** (regID as primary key) logs device roaming sessions using foreign keys deviceID and operatorID, where start and end are timestamps (end might be null if the roaming session is active). **Calls** (callID as primary key) track roaming calls, with foreign keys regCaller and regDest to caller and destination of a call, referencing **Registration**, and include details like time, duration (integer in seconds), and cost. For each of these tables there is a secondary B+ tree (non-clustering) index on the primary key attribute(s), created with the column order indicated in the tables. Additionally, there is a composite secondary index on Calls(regCaller,regDest).

The adopted DBMS uses blocks of 4KiB (4096 bytes). The records of all tables have a variable size, and at any given time, the Persons table has 100,000 tuples (tuple size 108 bytes, total 2640 disk blocks), and Devices 200,000 tuples (62 bytes/3028 blocks), the Countries table 200 tuples (58 bytes/3 disk blocks), the Operators table 800 tuples (58 bytes, total 12 disk blocks), the Registration table has 800,000 tuples (28 bytes/5469 blocks), and the Calls table 10,000,000 tuples (28 bytes per tuple/68,360 disk blocks).

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)   Determine the estimated number of tuples returned for the following query, that the primary key of country is numbered sequentially and starting from 1

SELECT *
FROM (SELECT * FROM Countries WHERE CountryID >= 180)
**   INNER JOIN Persons USING (CountryID)**
**   INNER JOIN Devices USING (PersonID)**

1 b) Consider the following schedule of two SQL transactions. Place the lock-S, lock-X, and unlock instructions in this schedule knowing that the exclusive locks are released immediately after the command requiring them is executed, and shared locks are kept till transaction end. Indicate whether the obtained scheduled obeys to two-phase locking policy, and if it is recoverable.

	Transaction T1:	Transaction T2:
1		UPDATE Countries SET countryName='Portugal' WHERE code = 351
2	UPDATE Countries SET countryName='Portugallo' WHERE code = 351	
3		DELETE Countries WHERE countryName='Portugallo'
4		COMMIT
5	SELECT COUNT(*) INTO n3 FROM Countries	
6	ABORT	

1 c) Consider the following recovery log record of a DBMS with memory buffering of disk blocks in which disk writes are only performed when checkpoints are performed (you can assume that there is no log buffering). The memory is initially empty. Complete the log knowing that the system has recovered from the recorded failure, indicating for the items referred to in the log the status of the disk and memory at each moment since the start of operation, as well as the values of X at the checkpoint and the contents of record 11 resulting from the rollback of transaction T3.

	Log
1	<T1, start>
2	<T1,Registration(123).end,NULL,'2025-06-12 08:30'>
3	<T2,start>
4	<T2,Calls(727).cost,0.0,1.75>
5	<T1,Calls(726).cost,0.0,0.35>
6	<T2,Calls(727).cost,1.75,2.05>
7	<checkpoint X>
8	<T3, start>
9	<T3,Registration(124).end,NULL,'2025-06-12 08:32'>
10	<T1, commit>
11	<T3, ... >
12	<T3, abort>
13	CRASH

1 d) Transactions in the schedule below are executed in SNAPSHOT ISOLATION mode. Assuming the database tables are initially empty, display the contents of table N at the end of the run knowing that it only has one column and no integrity constraints are in effect. Verify that the indicated schedule is serializable.

Passo	Transação T1:	Transação T2:	Transação T3:
1		begin transaction INSERT INTO Countries VALUES(1,'Portugal',351)	
2		INSERT INTO Countries VALUES(2,'Spain,34)	
3	begin transaction INSERT INTO Countries VALUES(3,'France',33)		
4		SELECT COUNT(*) INTO N FROM Countries;	
5			begin transaction INSERT INTO Countries VALUES(4,'Germany',49)
6	COMMIT		
7			COMMIT
		DBMS_OUTPUT.PUT_LINE(N);	
8		SELECT COUNT(*) INTO N FROM Countries;	
		DBMS_OUTPUT.PUT_LINE(N);	
9		COMMIT	

1 e) Consider the following distributed transaction that takes place at three different locations. Indicate the concurrency control log records (in order) of the coordinator and of the transaction managers of the several sites knowing that the transaction is coordinated by site 3, that the transaction coordinator failed after communicating to site 1 the decision to commit (which received it and logged it), and that it is used 2 phase commit for concurrency control. Explain what is the decision of site 2 when it discovers that the coordinator has failed and before it recovers.

#	Site 1	Site 2	Site 3
1			begin transaction
2	UPDATE Countries SET countryName='Portugal' WHERE code = 351		
3			UPDATE Persons SET … WHERE personID = 1234
4		SELECT COUNT(*) FROM Operators	
5			COMMIT

Group2 (each question is worth 2,5 values out of 20)

Note: the answer to each question cannot exceed 1 page

2 a)   The estimated number of tuples of a complex selection $\sigma_{\theta 1 \vee \theta 2 \vee \ldots \vee \theta n}(r)$ is given by the expression below, where s_i is the number of satisfying tuples of condition θ_i and n_r is the total number of tuples of relation r.

$$n_r * \left(1 - (1 - \frac{s_1}{n_r}) * (1 - \frac{s_2}{n_r}) * \ldots * (1 - \frac{s_n}{n_r}) \right)$$

HINT: Use De Morgan Laws

2 b)   Oracle implements SNAPSHOT ISOLATION with "First-Updater Wins" rule. Compare this approach with the "First-Commit Wins" rules, addressing advantages and disadvantages of the two approaches. Provide an example where the behaviour would be different.

2 c)   Explain what is the log force operation and why it is used.

THE END