



Venue \_\_\_\_\_

Seat Number \_\_\_\_\_

Student Number | | | | | | | | |

Family Name \_\_\_\_\_

First Name \_\_\_\_\_

## Semester Two Final Examinations, 2021

*This paper is for St Lucia Campus students.*

Reading Time: 10 minutes

**For Examiner Use Only**

Casio FX82 series or a calculator on the UQ approved list

During reading time (= planning time) - students are encouraged to review and plan responses to the exam questions

This examination paper will be released to the Library

### Materials Permitted In The Exam Venue:

**(No electronic aids are permitted e.g. laptops, phones)**

None

**Materials To Be Supplied To Students:**

None

### Instructions To Students:

**Additional exam materials (e.g. answer booklets, rough paper) will be provided upon request.**

Please answer all questions on the examination paper

Total marks: 100 (to be scaled down to 60)

[illegible]

Total

**Question 1 [10 marks]** Consider the following relations in a *University* database:

Student (SID, SName, Age, Type)

Course (CID, CName)

Enrolment (SID, CID, Semester, Grade)

*SID* is the primary key of the *Student* relation; *CID* is the primary key of the *Course* relation.

**1.1) [2 marks]** Assume that a student can take the same course multiple times in different semesters. What is the most suitable **primary key** for *Enrolment* relation? Justify your answer.

**1.2) [3 marks]** Assume the *Type* field in relation *Student* is defined as CHAR(15), and its value must be either 'Domestic' or 'International'. Write the SQL statement to CREATE DOMAIN StudentType that enforces this constraint.

**1.3) [5 marks]** Assume a student can take at most 4 courses in each semester. Write the SQL statement to define an **assertion** that ensures this constraint.

**Question 2 [4 marks]** Briefly explain the structure of **bitmap index**. What are the advantages and limitations of bitmap indexing?

**Question 3 [11 marks]** Consider the following *Movie* relation that contains  $N = 50,000$  records:

*Movie* (MovieID, Title, DirectorID, Category, Rating)

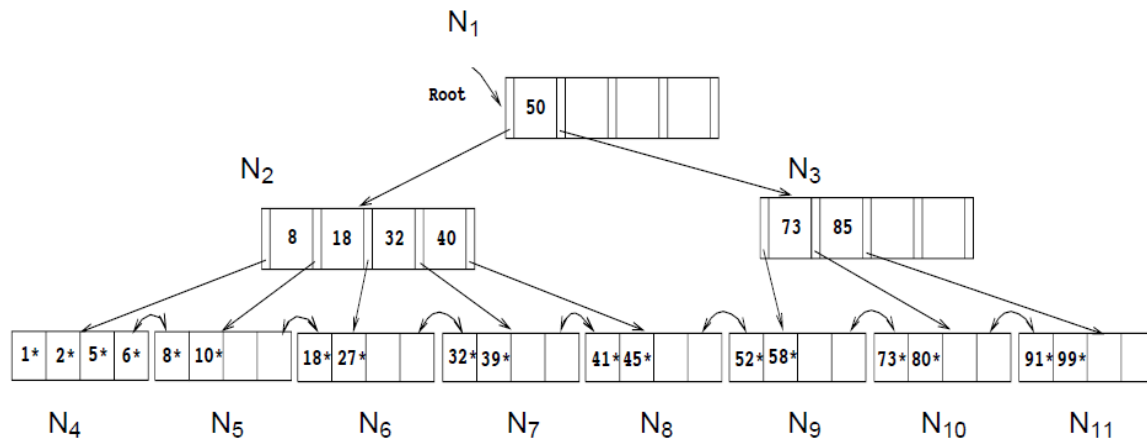
The file is unsorted, and a **B+ tree** index is constructed on the non-key attribute *DirectorID* (integer, 4 bytes). There are 1000 distinct values of *DirectorID*, and the *Movie* records are evenly distributed among these values. Each tree node is 60% full on average, and the B+ tree is stored on a disk with the following configuration:

- Block size (B) = 500 bytes
- Block pointer size (P) = 6 bytes

**3.1) [5 marks]** What is the height of the B+ tree? Show your calculation process and result.

**3.2) [6 marks]** What is the maximum cost (i.e., number of block accesses) of doing an equality search on the non-key field *DirectorID* using the B+ tree? Show your calculation process and result. (**Note:** You should consider the cost of both index access and data access)

**Question 4 [10 marks]** Consider a B+ tree as shown in the figure below, where the tree nodes are labelled as  $N_1, N_2, \dots, N_{11}$ . Assume the following rule applies for redistributing keys after a leaf node split: **Three keys** stay in the old node and the remaining keys move to a new node.



**4.1) [5 marks]** What is the minimum number of tree nodes that must be visited to answer the query: “Get all records with key greater than 15 and less than 75”? List all the visited nodes.

**4.2) [5 marks]** Show the updated B+ tree after inserting an entry with key “7”. For simplicity, you can show only the updated or newly-created tree nodes.

**Question 5 [11 marks]** Consider two relations  $R(A, B)$  and  $S(A, C)$ . Assume that **page-oriented nested-loop join** strategy is used for  $R * S$  (natural join). Let  $R$  be the outer relation and  $S$  be the inner relation.

**5.1) [4 marks]** Given the following tuples of  $R$  and  $S$ , what are the first 4 tuples produced by  $R * S$  using the page-oriented nested-loop join? Assume each block can store only 2  $R$  tuples or  $S$  tuples.

- 1<sup>st</sup> tuple:
- 2<sup>nd</sup> tuple:
- 3<sup>rd</sup> tuple:
- 4<sup>th</sup> tuple:

A	B	A	C
7	x	8	1
2	z	4	3
9	y	2	6
8	y	1	5
3	w	3	7
9	x	2	1
1	w	7	8
3	y	9	2
6	z		

R                      S

**5.2) [5 marks]** Assume that  $R$  and  $S$  are stored on a disk with block size = 1000 bytes. Relation  $R$  contains 200,000 records with each record occupying 50 bytes. Relation  $S$  contains 10,000 records with each record occupying 20 bytes. Estimate the number of block accesses required for  $R * S$  using the page-oriented nested-loop join. Show your calculation process and result.

**5.3) [2 marks]** Can we improve the efficiency of  $R * S$  by changing the join order, i.e.  $S$  as the outer relation and  $R$  as the inner relation? Justify your answer.

**Question 6 [11 marks] Operator evaluation** determines the best algorithm for implementing each relational algebra operator. This is an important step in query optimization. Both rule-based and cost-based strategies can be used for operator evaluation.

**6.1) [3 marks]** Consider the selection operator with a single condition. Explain each of the three main heuristics used for **rule-based** selection operator evaluation.

**6.2) [8 marks]** The **Cost-based** operator evaluation strategy needs to estimate the size of the operator result. Given two relations R1 and R2, where R1 contains N1 tuples and R2 contains N2 tuples ( $N2 > N1 > 0$ ), what is the minimum and maximum possible size (i.e., number of tuples) respectively for the resulting relation produced by each of the following relational algebra operators? Justify your answer (i.e., in what situation would that minimum or maximum size happen).

- **[2 marks]**  $R1 \times R2$  (Cross Product)
  
  
  
  
  
  
  
  
  
  
- **[2 marks]**  $R1 * R2$  (Natural Join)
  
  
  
  
  
  
  
  
  
  
- **[2 marks]**  $\pi_A(R1)$ , assuming A is a field in R1 (Projection)
  
  
  
  
  
  
  
  
  
  
- **[2 marks]**  $\sigma_{A=5}(R1)$ , assuming A is a field in R1 (Selection)

**Question 7 [13 marks]** Consider the following relations:

Sailor (SID, SName, Age, Rating)

Boat (BID, BName, Capacity)

Reserve (SID, BID, Date, Price)

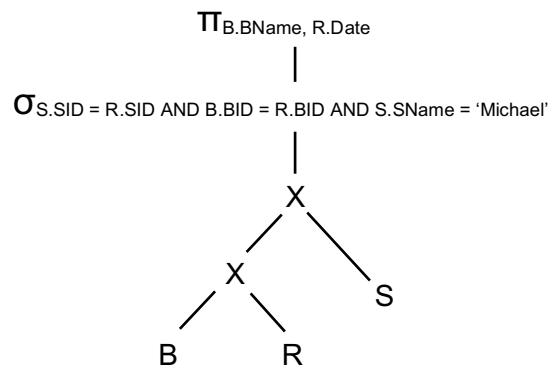
Assume there are 10,000 *Sailor* tuples and 100 *Boat* tuples in the database. *SName* is a unique field in the *Sailor* relation. Given the following SQL query:

```
SELECT B.BName, R.Date
```

```
FROM Sailor S, Boat B, Reserve R
```

```
WHERE S.SID = R.SID AND B.BID = R.BID AND S.SName = 'Michael';
```

The initial **query tree** is illustrated as below:



**7.1) [3 marks]** Based on the above initial query tree, show the equivalent query tree after pushing down selection operators.

**7.2) [2 marks]** Based on the query tree obtained in Question 7.1, show the equivalent query tree after converting cross-products into joins.

**7.3) [2 marks]** Based on the query tree obtained in Question 7.2, show the equivalent query tree after rearranging leaf nodes so as to execute the most restrictive selection operators first.

**7.4) [6 marks]** Based on the query tree obtained in Question 7.3, show the equivalent query tree after pushing down projection operators.



**Question 8 [10 marks]** Lock-based concurrency control is widely used in relational database systems. A lock on data item X indicates that a transaction is performing some operation (read or write) on X.

**8.1) [5 marks]** Consider the following lock table where each entry keeps the information about a locked data item. Construct a **wait-for-graph (WFG)** based on this lock table. Is there any deadlock in the system? Justify your answer.

Data Item	Locks Granted	Queued Requests
A	<T1, read_lock>, <T5, read_lock>	<T4, write_lock>
B	<T2, write_lock>	<T1, read_lock>
C	<T3, write_lock>	<T1, write_lock>, <T2, read_lock>
D	<T4, read_lock>	<T2, write_lock>

**8.2) [2 marks]** **Victim selection** is used to decide which transaction on a WFG should be aborted when a deadlock is detected. List at least two factors that should be considered in victim selection.

**8.3) [3 marks]** Explain the pros and cons of fine-grained locking granularity compared with coarse-grained locking granularity.

**Question 9 [10 marks]** Database recovery is necessary when the system encounters failures, such as transaction failure, system failure, media failure, etc.

**9.1) [4 marks]** Which **ACID** properties of transactions can database recovery guarantee? For each guaranteed property, explain its meaning and specify what recovery operation is used to ensure that property.

**9.2) [2 marks]** Assume that **steal/no-force** buffer management is adopted. What recovery operations are needed for database recovery? Justify your answer.

**9.3) [4 marks]** Assume that the **checkpoint** mechanism is adopted, i.e., the recovery engine periodically adds a checkpoint in the log indicating that all the modified buffers have been flushed to disk at that point. Given the following database log, what recovery operations are needed for T1, T2, T3, and T4, respectively? Justify your answer.

[start_transaction, T1]
[write_item, T1, D, 10, 20]
[commit, T1]
[checkpoint]
[start_transaction, T2]
[write_item, T2, B, 2, 15]
[abort, T2]
[start_transaction, T3]
[write_item, T3, C, 1, 5]
[start_transaction, T4]
[write_item, T4, D, 20, 25]
[commit, T4]
[write_item, T3, A, 10, 5] ← System crash

**Question 10 [10 marks]** Consider two transactions T1, T2 and two data items X, Y.

T1: R(X) ; W(X) ; R(Y) ; W(Y)

T2: R(X) ; W(X)

Given the following schedule of interleaved operations from T1 and T2:

S: R1(X) ; W1(X) ; R2(X) ; R1(Y) ; W2(X) ; C2 ; W1(Y) ; A1

“A1” means transaction T1 aborts, and “C2” means transaction T2 commits.

**10.1) [2 marks]** What type of **anomaly** occurs in this schedule? Explain your answer.

**10.2) [2 marks]** Two-Phase Locking (2PL) is widely used for concurrency control in the relational database systems. Briefly describe the **strict 2PL** protocol.

**10.3) [4 marks]** For each of T1 and T2, insert all the **lock** and **unlock** operations to make the transaction satisfy the strict 2PL protocol.

**10.4) [2 marks]** Show that modifying the schedule according to strict 2PL can prevent the anomaly. Justify your answer.

**END OF EXAMINATION**