CMI/Feb-Mar-2021/End-Semester Exam March 2021

Chennai Mathematical Institute

RDBMS, SQL and Visualization 150 Minutes. 55 Marks.

Instructions

- First section has negative marks.
- Submit your answer sheet as a single pdf file on moodle. For any reason if you cannot upload to moodle, email it to vvtesh.cmi@gmail.com.
- If you make multiple submissions, the last submission will be considered for grading.
- Late submissions (from 12:15 PM onwards) will attract penalty of 0.5 marks per minute.

Section 1: Questions 1 - 10 carry 2 marks each. There is a negative mark of -1 for every wrong answer. If there are multiple correct answers, you need to choose all the correct answers and only the correct answers to avoid getting negative marks.

Question 1. We decided to split the relation R(A,B,C,D,E) into R1(A,D,E) and R2(A,B,C,D). Following functional dependencies exist: $\{AD \to E, C \to BD, E \to AB, AB \to C\}$. Is the decomposition lossless?

Choose the best answer:

(1) **The decomposition is lossless.** ✓ (AD is a key. Exists in both R1 and R2.)
(2) The decomposition is lossy.

Question 2. Find the highest normal form of a relation R(A,B,C,D,E) with FD set as $\{BC \to A, DC \to BE, B \to E\}$.

Choose the best answer:

(1) **2NF** ✓
(2) 3NF
(3) BCNF
(4) 4NF

Question 3. Let R be the set of all binary relations on the set {1,2,3}. Suppose a relation is chosen from R at random. The probability that the chosen relation is asymmetric is _____ ?

(1) $2^6/2^9$
(2) $3^3/2^9$ ✓
(3) $3^6/2^9$
(4) $2^3/2^9$

Question 4. Let A1, A2 and A3 be sets of attributes on relation r such that $A1 \subset A2 \subset A3$. F1 and F2 are boolean expressions over the attributes of r. Which of the following will produce the same output as $\pi_{A1}(\pi_{A2}(\pi_{A3}(\sigma_{F1}(\sigma_{F2}(r)))))$?

Choose all the correct answers:

(1) $\pi_{A1}(\sigma_{F1 \wedge F2}(r))$ ✓
(2) $\pi_{A1}(\sigma_{F1 \vee F2}(r))$
(3) $\pi_{A1}(\pi_{A3}(\sigma_{F1 \wedge F2}(r)))$ ✓
(4) $\pi_{A1}(\pi_{A3}(\sigma_{F1 \vee F2}(r)))$

Question 5. Consider a file of 32768 records. Each record is 32 bytes long and its key field is of size 6 bytes. The file is ordered on a non-key field, and the file organization is unspanned. The file is stored in a file system with block size 1024 bytes, and the size of a block pointer is 10 bytes. If the secondary index is built on the key field of the file, and a multi-level index scheme is used to store the secondary index, the number of first-level and second-level blocks in the multi-level index are respectively _____.

(1) **512, 8** ✓
(2) 256, 4
(3) 512, 4
(4) 256, 8

Question 6. Assume that search key occupies 6 bytes and a block pointer occupies 4 bytes. If the block size is 1024 bytes, what is the maximum order of the B^+ tree?

(1) 51
(2) 52
(3) 102
(4) **103** ✓

Question 7. Which of the following statements are incorrect?

(1) A relation in first normal form cannot have multi-valued attributes.
(2) A relation in fourth normal form cannot have multi-valued attributes.
(3) **A relation in first normal form does not allow multi-determines dependencies.** ✓
(4) A relation in fourth normal form does not allow multi-determines dependencies.

Question 8. What is the canonical cover of the set of dependencies $F = \{A \rightarrow BC, B \rightarrow CE, A \rightarrow E, AC \rightarrow H, D \rightarrow B\}$

Choose all the answers that apply:

(1) $A \rightarrow BH, B \rightarrow CE, D \rightarrow B$ ✓
(2) $A \rightarrow CE, B \rightarrow H, B \rightarrow D$
(3) $A \rightarrow H, B \rightarrow CE, D \rightarrow B$
(4) $A \rightarrow BH, B \rightarrow E, B \rightarrow D$

Question 9. Which of the following is true with respect to the select (σ) and project (π) operations of relational algebra?

(1) σ is commutative, π is commutative.
(2) **σ is commutative, π is not commutative.** ✓
(3) σ is not commutative, π is commutative.
(4) σ is not commutative, π is not commutative.

Question 10. Consider the relation R(A,B,C,D,E). The set of functional dependencies are $F = \{A \to BC, CD \to E, B \to D, E \to A\}$. How many candidate keys exist for R?

(1) 3
(2) **4** ✓ The candidate keys are A, E, CD and BC
(3) 5
(4) 6

Section 2: Questions 11 - 15 carry 5 marks each.

Question 11. We would like to capture daily COVID transmission data in a table. You may make reasonable assumptions on the data items (i.e., columns of the table). Describe a relational schema which is in 3NF but not in 4NF for this purpose. Explain the reasons why you believe the schema you describe is in 3NF but not in 4NF.

A variant may be related to a set of affected countries, and both of them may be related to spreading-speed. For instance see Table 1.

Variant	Country	Speed
SAV	South Africa	Fast
CAV	China	Medium

TABLE 1. Covid Variants

Here, $Variant \twoheadrightarrow Country$ and $Variant \twoheadrightarrow Speed$.

Question 12. Assume that an instance of relation R(A,B,C,D) is as given below. The table is stored using sequential file organization sorted on the attribute D. For a fan-out factor of 3, draw the B^+ tree index over attribute A. Note that the attribute D = 9 occurs twice in R.

A	B	C	D
3	11	d	4
2	11	b	6
1	11	c	6
4	11	a	7
5	11	e	8
6	11	f	9
7	11	g	9
8	11	g	10
9	11	g	11
10	11	g	12

Construct a B+Tree as usual for 3,2,1... 10. Point 9 to the first entry of 9. Since it is sequential file organization, the pointers from one row to other will help us navigate further. A rough example is in Figure 1.

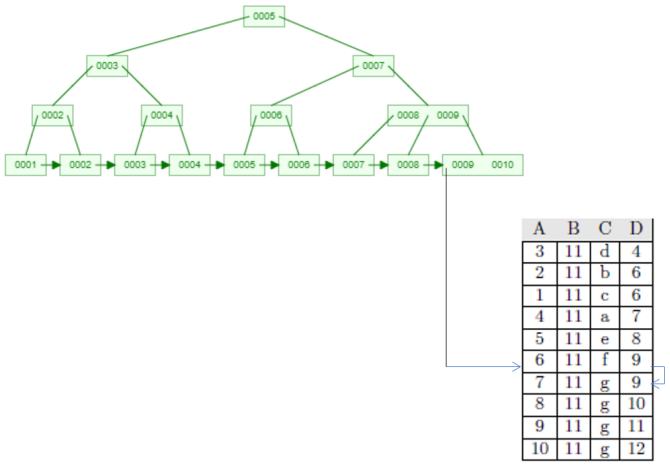

FIGURE 1. B+ Tree

Question 13. Consider the relational schema:

```
student(rollno:int, name:varchar(10), courseCode:char(3),
grade:char(1)) course(courseCode:char(3), courseName:varchar(45))
```

Assume that fail grade is 'F' while answering the following questions.

(a) Write a TRC query to find students who failed in any course. The query should output the student name and the course name. $\{t \mid \exists s \in student, \exists c \in course\,(t[name] = s[name] \wedge t[courseName] = c[courseName] \wedge s[courseCode] = c[courseCode] \wedge s[grade] = \text{'}F\text{'}\}$

(b) Write a DRC query to find all student names who failed in course with code 'SQL'. $< n > \mid < r, n, \text{'}SQL\text{'}, \text{'}F\text{'} > \in student\}$

Question 14. Provide the relational schema for the ER Model shown in Figure 2. Also provide two sample tuples for each relation that you listed in the schema.

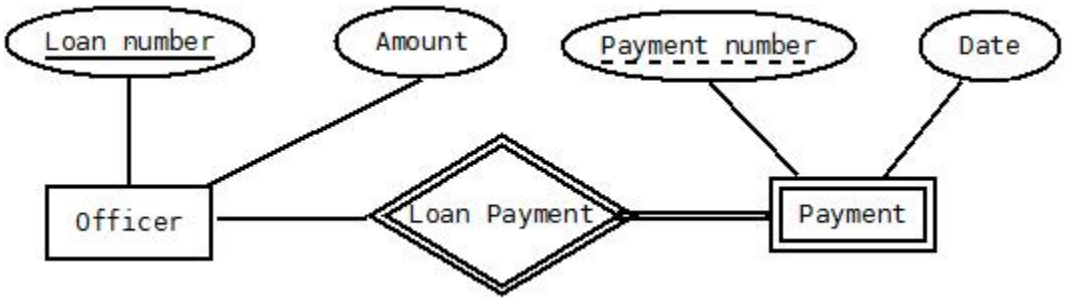

FIGURE 2. ER Model

Schema:

Officer(Loan number, Amount)
Payment (Loan number, Payment number, Date).

Assumes 1:* relationship.

Question 15. Consider a railway reservation system that maintains data in an RDBMS. It holds data such as train schedule, routes, passenger reservations and cancellations. For such a system,

(a) identify two non-trivial functional dependencies.
(b) identify a foreign key constraint.

You may make reasonable assumptions on the relational schema design. Clearly specify the relations that are relevant to the identified functional dependencies and foreign key constraint.

PNR determines seatno, trainno determines haltingstation. A schema could be:

reservation(PNR, seatno, trainno, name, age, gender)
train(trainno, haltingstation)

The trainno in reservation cannot exist without being inserted into train table. Hence it is a FK.

Section 3: Questions 16 - 20 carry 2 marks each.

The database administrator at CMI designs a database which contains the following tables:

```
student(rollno, sname)
course(cid, cname)
registration(rid, cid, year, semester, rollno)
grade(rollno, rid, lettergrade)
```

As an example, consider the following instance of the above mentioned schema:

rollno	sname
MDS201905	Aashish Ranjan

cid	cname
RSV	RDBMS, SQL and Visualization
INFR	Information Retrieval

rid	cid	year	semester	rollno
1	RSV	2019	Oct-Nov	MDS201905
2	INFR	2020	Aug-Sep	MDS201905

rollno	rid	lettergrade
MDS201905	2	A

Using the above schema description, answer all of the following questions.
Please find the SQL script at http://vvtesh.github.io/teaching/dbms2021/dbms2021final.txt

Question 16. Write the SQL query to find the lettergrade that has the most number for students for the RSV course.

Question 17. Write the SQL query to find all the student names who failed in the Oct-Nov semester. The lettergrade 'F' indicates failure.

Question 18. Write the SQL query to find the course id with maximum enrollments across all years.

Question 19. Write the SQL query to update the course id of RSV to SQL.

Question 20. Ashish loved the RSV course. Your instructor decided to offer an RSV2 course. So, he wishes to register for the 2021 Feb-Mar semester offering of RSV2 course. You need to make the necessary inserts to all appropriate tables. Provide the relevant SQL queries.

Section 4: Question 21 is a bonus question for 1 mark. Bonus marks do not apply if you score 55/55 from the rest of the paper.

Question 21. Describe any one idea/concept that you enjoyed the most in this course. Explain why that idea/concept caught your attention.