CMI/Oct-Nov-2019/End-Semester Exam 28/11/2019

Chennai Mathematical Institute

RDBMS, SQL AND VISUALIZATION 180 MINUTES. 70 MARKS.

Instructions

- This is a closed book test.
- You are allowed to carry one page (A4 size only) of hand written or printed notes. Write your name and roll number on these notes. Submit the notes along with this booklet.
- Please switch off your mobile phones and any other digital equipment you may have (like laptops, calculators and smart watches).
- No negative marks.
- You are encouraged to make your assumptions if any, explicit.

Section 1: Questions 1 - 5 carry 2 mark each.

Question 1. An instance of a relation R(A,B,C,D) contains four records as given below. What is the result of the query $\Pi_A(\sigma_{(A=B)\vee(C<30)})$ on the given instance?

A	B	C	D
m	n	23	4
m	n	33	5
n	n	23	4
n	m	33	5

Question 2. A relation R($A_1, A_2, ..., A_n$) has n attributes. If "A_1A_2" together form the candidate key, how many superkeys are possible for this relation?

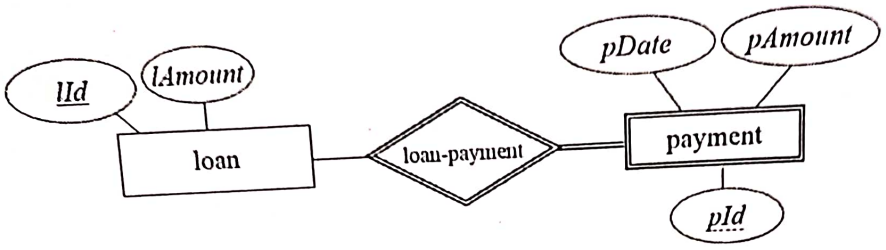
Question 3. Consider the relation scheme R(E, F, G, H, I, J, K, L, M, N) and the set of functional dependencies $\{EF \to G, F \to IJ, EH \to KL, K \to M, L \to N\}$ on R. What is the key for R?

Question 4. Consider a schema R(A, B, C, D, E) and functional dependencies $A \to B$ and $C \to DE$. Then the decomposition of R into R1(A, B) and R2(C, D, E) is lossless. True/False?

Question 5. If we had customer as outer relation and depositor as inner relation in a nested loop join scenario, what would be the worst-case cost? Assume that the customer has 1000 records, and depositor has 100 records. Customer relation is stored in 40 blocks. The relation depositor uses 2 blocks.

Section 2: Questions 6 - 10 carry 4 mark each.

Question 6. Convert the ER Diagram given below to a relational model. Provide the relational schema. List the functional dependencies if you find any. Remember to make your assumptions explicit.



Question 7. Find the highest normal form of a relation R(A,B,C,D,E) with FD set as $\{BC \rightarrow D, AC \rightarrow BE, B \rightarrow E\}$.

Question 8. Consider the following four schedules due to three transactions (indicated by the subscript) using read and write on a data item x, denoted by r(x) and w(x) respectively. Which of these are conflict serializable?

(1) $r_1(X); r_2(X); w_1(X); r_3(X); w_2(X)$
(2) $r_2(X); r_1(X); w_2(X); r_3(X); w_1(X)$
(3) $r_2(X); w_2(X); r_3(X); r_1(X); w_1(X)$
(4) $r_3(X); r_2(X); r_1(X); w_2(X); w_1(X)$

Question 9. Assume that an instance of relation R(A,B,C,D) is as given below. For a fan-out factor of 3, draw the B+ tree index over attribute C.

A	B	C	D
1	11	d	21
2	11	b	21
3	11	c	21
4	11	a	21
5	11	e	21
5	11	f	21

Question 10. Consider the following schema: suppliers(sid, sname, saddress), parts(pid, pname, color), catalog(sid, pid, cost).
The catalog relation gives the parts sold by suppliers and the cost at which those parts are sold.

Find the names of the suppliers who supply red parts. Write the query in tuple (2 marks) and domain relational calculus (2 marks) form.

Section 3: Questions 11 - 14 carry 2 marks each. Questions 15, 16 and 17 carry 4 marks each.

Assume that CMI wants to maintain student information in a database. 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

TABLE 1. student

cid	cname
RSV	RDBMS, SQL and Visualization
INFR	Information Retrieval

TABLE 2. course

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

TABLE 3. registration

rollno	rid	lettergrade
MDS201905	1	B
MDS201905	2	A

TABLE 4. grade

Using the above schema description, answer all of the following questions.

Question 11. Write the SQL query to insert *Tanmey Rawal* and *MDS201938* into the student table.

Question 12. Write the SQL query required to register *Tanmey Rawal* for *2019*, *Oct-Nov* offering of the *RSV* course.

Question 13. We know that *Aashish Ranjan* is a brilliant student. It was a mistake to have his grade as *B*. Write the SQL query to change *Aashish Ranjan*'s grade for the *RSV* (*2019*, *Oct-Nov*) to *A*.

Question 14. Assume we registered several students for several courses. Write the SQL query to count the number of registrations for each course.

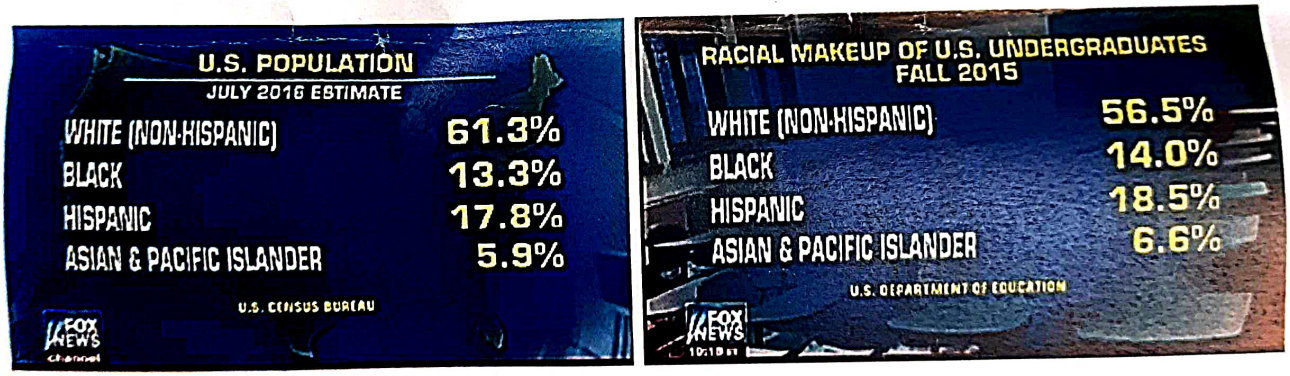
Question 15. Our academic office has an objection to pre-fill grades for a course that is yet to happen. Please write the SQL query(ies) to delete all grades of *INFR* course. Let the course record remain in the registration table.

Question 16. Write the SQL query to find the course that has maximum students with 'A' grade.

Question 17. Write the SQL query to list the *rollno* of all students who have scored *A* grade in all their registered courses.

Section 4: Questions 18 -21 carry 5 marks each.

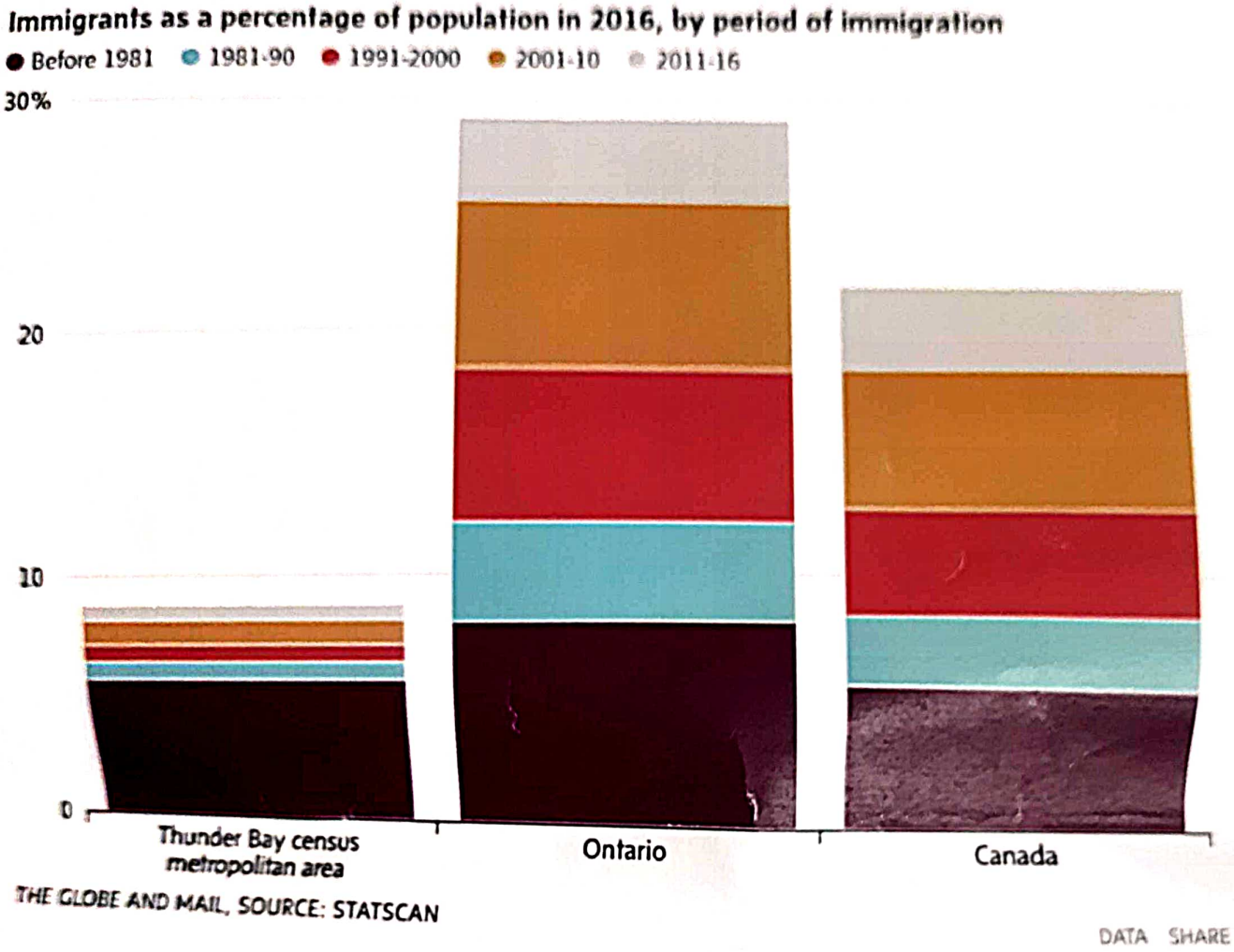
Question 18. On August 2, 2017 during a discussion on affirmative action, the anchor of a show of Fox News in USA, presented two tables, first the table on the left, then the right, stating:



> "OK, I want to put up some numbers here just so people have a little bit of data in front of them to look at the official population estimates [left table]. This is the overall U.S. population. You can see the statistics there and you have the white population at 61.3 percent, and then a breakdown between black, Hispanic, and Asian and Pacific Islanders. Now when you look at those [right table] the racial makeup of U.S. undergraduate students, it's about 5 percent lower for white students and slightly higher for each of the other groups represented there."

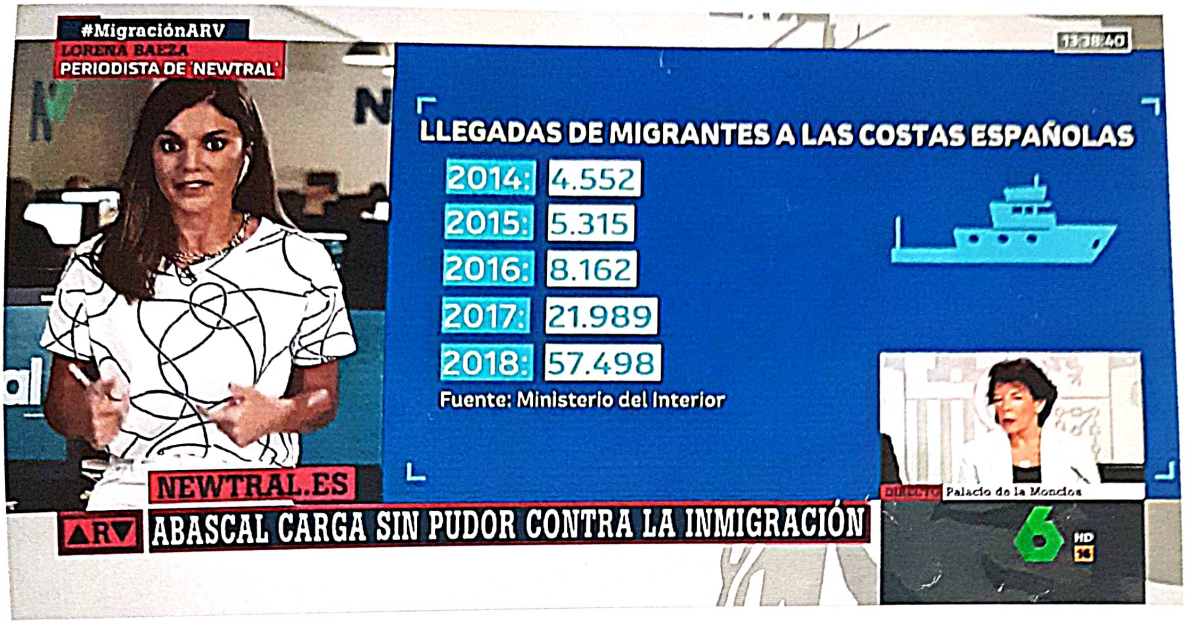
There is a problem in the analysis presented by Fox News anchor. Explain the problem briefly.

Question 19. Following figure presents the immigrants as a percentage in 2016, by period of immigration in two areas of Canada and in overall Canada.



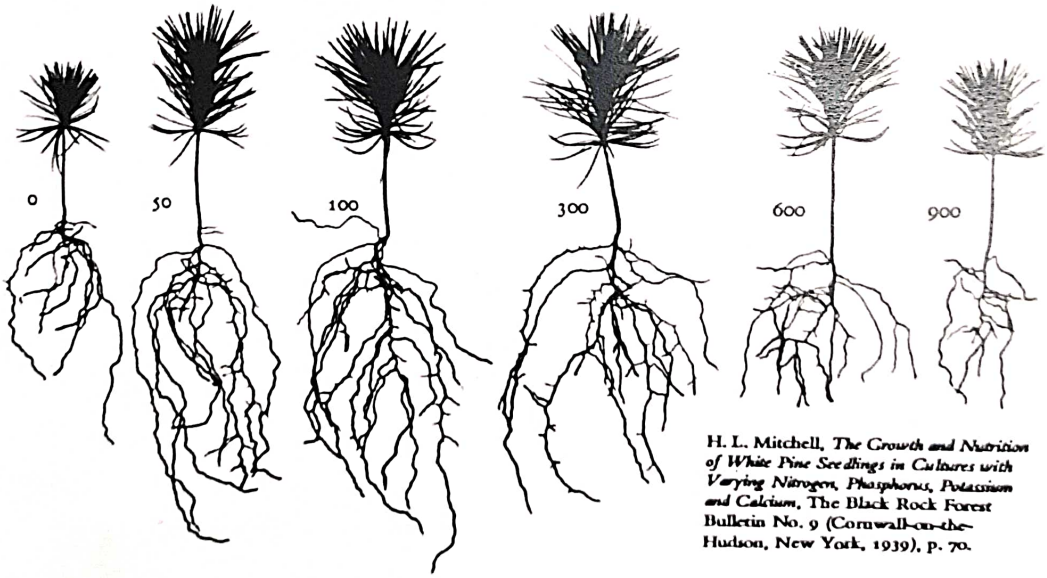
(1) This is an example of bad graphics. Why?
(2) Fix the graphical analysis by presenting it alternately. (Hint: You can make some approximation from the current graph.)

Question 20. Explain the problem in the following graph



Question 21. Write a report in one paragraph based on the following graph.

And similarly, the varying sizes of white pine seedlings after growing for one season in sand containing different amounts of calcium, in parts per million in nutrient-sand cultures:



H. L. Mitchell, *The Growth and Nutrition of White Pine Seedlings in Cultures with Varying Nitrogen, Phosphorus, Potassium and Calcium*, The Black Rock Forest Bulletin No. 9 (Cornwall-on-the-Hudson, New York, 1939), p. 70.