UCLA
Computer Science Department
Winter 2021

Instructor: J. Cho

CS143 Midterm: 1 Hour 50 minutes

Student Name: ____

Student ID: ____

(IMPORTANT PLEASE READ **):**

- The exam is *closed book* and *closed notes*. You may use *one double-sided cheat-sheets*. You can use a calculator.
- *Simplicity and clarity of your solutions will count.* You may get as few as 0 point for a problem if your solution is far more complicated than necessary, or if we cannot understand your solution.
- If you need to make any assumption to solve a question, *please write down your assumptions*. To get partial credits, you may want to write down how you arrived at your answer step by step.
- If a question asks for a numeric answer, you don't have to calculate. You may just write down a numeric expression.
- Please, write your answers neatly.

Problem	**Score**	
1	25	
2	30	
3	15	
4	15	
5	15	
Total	100	

Problem 1: Relational Algebra (25 points)

Consider a relation $R(A, B)$ that contains r tuples and a relation $S(B, C)$ that contains s tuples; assume $r \geq s > 0$. Make no assumptions about keys. For each of the following relational algebra expressions, write down the minimum and maximum number of tuples that could be in the result of the expression using r, s, and/or numbers.

Expression	minimum #tuples	maximum #tuples
$R \cup \rho_{S(A,B)}(S)$		
$\pi_{A,C}(R \bowtie S)$		
$\pi_B(R) - (\pi_B(R) - \pi_B(S))$		
$(R \bowtie R) \bowtie R$		
$\sigma_{A>B}(R) \cup \sigma_{A<B}(R)$		

Problem 2: Query Equivalence (30 points)

Two queries are considered equivalent if they return exactly the same results for all database instances. For each of the pair of queries listed below, write "YES" if the two queries are equivalent and "NO" otherwise.

Assume the following for the two relations R and S referenced in the queries.

- Both R and S have two columns, $R(A, B)$ and $S(A, B)$.
- The attribute A is a key for each relation.
- The attribute B is not a key in either relation.
- NULL values are not allowed in any attribute.

Do not make any other assumptions about the data.

Again, remember that A is a key, B is not, and NULL values are not allowed. These assumptions are crucial to derive the correct answers.

Queries	Equivalent?
(a) $\pi_A(R - S)$ (b) $\pi_A(R) - \pi_A(S)$	
(a) $\pi_{R.A}(\sigma_{R.A=S.A}(R \times S))$ (b) `SELECT R.A FROM R,S WHERE R.A=S.A`	
(a) $\pi_{R1.B}(\sigma_{R1.B=R2.B \wedge R1.A \neq R2.A}(\rho_{R1}(R) \times \rho_{R2}(R)))$ (b) `SELECT B FROM R GROUP BY B HAVING COUNT(*) > 1`	
(a) `SELECT SUM(B)/COUNT(*) FROM R` (b) `SELECT AVG(B) FROM R`	
(a) `SELECT B FROM R` `WHERE NOT EXISTS(SELECT * FROM S WHERE R.B = S.B)` (b) `(SELECT B FROM R) EXCEPT (SELECT B FROM S)`	
(a) `SELECT B FROM R WHERE A > 2 OR B < 2` (b) `(SELECT B FROM R WHERE A > 2) UNION` `(SELECT B FROM R WHERE B < 2)`	

Problem 3: Expressive Power of Relational Algebra (15 points)

Consider two relations $R(A, B)$ and $S(A, B)$. You would like to compute their intersection $R \cap S$, but unfortunately you only have three relational algebra operators at your disposal: σ, π, and $\times$. Is it possible to compute $R \cap S$ using just these three operators? If so, show the simplest equivalent expression you can come up with. If not, briefly explain why not.

Problem 4: Handling NULL (15 points)

Given the following instances of SQL tables $R(A, B)$ and $S(B, C)$,

R.A	*R.B*
2	NULL
1	1
1	2

S.B	*S.C*
1	3
1	NULL
NULL	2
2	NULL

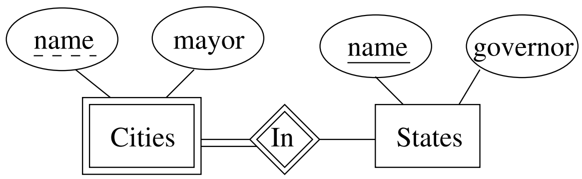
write down the result of the following SQL SELECT statement.

```
SELECT S.B, AVG(C)
FROM R, S
WHERE R.B = S.B
GROUP BY A, S.B
```

In order to get full credit, explain, step by step, how you arrived at your answer.

Problem 5: ER Model (15 points)

Which of the following is necessarily true about the City and State entity sets and their relationship In?



Mark the statements with TRUE or FALSE. Your answer should be purely based on what is documented in the ER model, nothing else.

1. Each City can be In at most one State. TRUE/FALSE
2. Each City has at most one mayor. TRUE/FALSE
3. No two Cities In the same State can have the same name. TRUE/FALSE
4. No two States can have the same name. TRUE/FALSE
5. Two Cities with the same name cannot be In two different States. TRUE/FALSE

1.
   1. Min: r (when $S \subseteq R$), Max: $r + s$ (when $R \cap S =$)
   2. Min: 0 (when all R.B values are different from S.B values), Max: $r \times s$ (when all R.B = S.B = b)
   3. Min: 0, Max: s. This expression is equivalent to $\pi_B(R) \cap \pi_B(S)$
   4. Min: r, Max: r. R ⋈ R is always R
   5. Min: 0 (when A=B for every tuple in R), Max: r (when A≠B for every tuple in R)

2.
   1. No. Consider $R = \{(1, 2)\}, S = \{(1, 3)\}$
   2. Yes
   3. Yes
   4. Yes
   5. No. Consider $R = \{(1, 0), (2, 0)\}$.
   6. No. Consider $R = \{(2, 1), (3, 1)\}$.

3. $\pi_{R.A,R.B}(\sigma_{R.A=S.A \wedge R.B=S.B}(R \times S))$

4. {(1,3), (2, NULL)}. 'NULL = NULL' is Unknown, so when the WHERE clause condition is applied, the R tuple (1,1) will join with S tuples {(1, 3), (1, NULL)} and the R tuple (1,2) will join with S tuple (2, NULL). Then after GROUP BY A, S.B, we have two groups ((1,1), {(1,3), (1,NULL)}) and ((1,2), {(2, NULL)}), and the result of AVG(C) will be 3 and NULL for each group, respectively.

5.
   1. False (double lines is at least, not at most).
   2. True (Each state has one mayor)
   3. True (state.name, city.name) is the key of a City, so city.name must be unique within a state
   4. True (state name is the key of State entity set)
   5. False (No such constraint is implied by the given ER model)