ASSIGNMENT 4 - 601.315/415/615 - Databases

Due Date: No later than Sunday, December 13, 7:30PM
(no extensions are allowed because solution sets will be distributed on the 13th in preparation for the final exam on December 14th from 6-9PM).

This assignment is intended primarily for review and exam preparation. It is composed of questions given on previous final exams and will also serve as an effective sample final. To incentivize students to do these practice/review exercises for exam pre paration, it will constitute a relatively small 3% of your grade, worth doing overall, but with mini mal impact on your course grade.

Questions

‘

(1) Consider the following relational schema and SQL query (next page).

EMPLOYEE	SSN	Name	DNO	Salary	Hobby	Sex
	339-18-4886	Milton Kent	520	65000	Golf	M

DEPARTMENT	DNO	DName	Budget	Location	MGRSSN
	520	Accounting	10000000	NEB	339-18-4886

WORKS_ON	ESSN	PNUM
	339-18-4886	109

PROJECT	PNUM	PName	Budget	Location	Goal
	520	F16Engine	30000000	NEB	Profit

Assume that there are 50,000 employees, 1,000 projects, 30 departments 50% of employees are male, 0.01% of employees have a yodeling hobby, the Relational Mechanics department only has 6 members and the median project budget is $1,000,000. 5 people work on the F16Engine project.

(a) Convert the following SQL query into the relational algebra and draw a query tree for it for the most efficient execution ordering you can think of.

```
SELECT E.LName
FROM   Employee E, Department D, WorksON W, Project P
WHERE  E.DNO=D.DNO and E.SSN=W.ESSN and P.PNUM=W.PNUM
       and P.Budget > $50 and E.Sex='M' and E.hobby='Yodeling'
       and D.Dname='Rational Mechanics';
```

(b) Convert the following SQL query into the relational algebra and draw a query tree for it for the most efficient execution ordering you can think of.

```
SELECT E.LName, D.DNO
FROM   Employee E, Department D, WorksON W, Project P
WHERE  E.DNO=D.DNO and E.SSN=W.ESSN and P.PNUM=W.PNUM
       and P.Fname='F16Engine' and E.Sex='M'
```

(2) Consider the following two transactions:

T_0	T_1
read_item(A)	read_item(B)
read_item(B)	read_item(A)
if A=0 then B :=B+1	if B=0 then A := A+1
write_item(B)	write_item(A)

(a) Show a concurrent execution of T_0 and T_1 which produces a serializable schedule (Represent concurrent executions as shown in the table in problem 7).

(b) Show a concurrent execution of T_0 and T_1 which produces a non-serializable schedule

(3) Consider the following 3 transactions, consisting of a sequence of read and write operations, with (omitted) intermediate code.

T_0	T_1	T_2
read_item(X)	read_item(Z)	read_item(Y)
write_item(X)	read_item(Y)	read_item(Z)
read_item(Y)	write_item(Y)	write_item(Y)
write_item(Y)	read_item(X)	write_item(Z)
	write_item(X)	

Now consider the following concurrent schedule of these 3 transactions. Is this schedule conflict serializable? Why or why not (create a labeled precedence graph to help you argue). If it is conflict serializable, give an equivalent serial schedule.

(See table on the following page)

T_0	T_1	T_2
	read_item(Z)	
	read_item(Y)	
	write_item(Y)	
		read_item(Y)
		read_item(Z)
read_item(X)		
write_item(X)		
		write_item(Y)
		write_item(Z)
	read_item(X)	
read_item(Y)		
write_item(Y)		
	write_item(X)	

(4) Consider the 3 transactions given in problem 3, along with the following concurrent schedule. Is this schedule conflict serializable? Why or why not (create a labeled precedence graph to help you argue). If it is conflict serializable, give a equivalent serial schedule.

T_0	T_1	T_2
		read_item(Y)
		read_item(Z)
read_item(X)		
write_item(X)		
		write_item(Y)
		write_item(Z)
	read_item(Z)	
read_item(Y)		
write_item(Y)		
	read_item(Y)	
	write_item(Y)	
	read_item(X)	
	write_item(X)	

(5) Consider the precedence graph below. Is the corresponding transaction history conflict serializable? If so, give an equivalent valid serial schedule for the transactions.

PLEASE SEE ATTACHMENT ON CLASS WEBSITE FOR DIAGRAM.

(6) Consider the labeled precedence graph below. Is the corresponding transaction history conflict serializable? If so, give an equivalent valid serial schedule for the transactions.

PLEASE SEE ATTACHMENT ON CLASS WEBSITE FOR DIAGRAM.

(7) Consider a relation that is fragmented horizontally by *PlantLocation*:

EMPLOYEE	Name	Title	Salary	PlantLocation
	Joe Steel	Foreman	65000	Edmonton

Assume each fragment has two replicas: one stored at the New York office and one stored locally at each plant location. Describe a good processing strategy for the following queries entered at the San Jose plant location.

(a) Find all employees at the *'Boca'* plant.

(b) Find the average salary of all employees in the company.

(c) Find the highest-paid employee at each of the following locations Toronto, Edmonton, Vancouver, Montreal.

(c) Find the lowest paid employee in the entire company.

(8) Consider the relations:

EMPLOYEE	Name	Title	Salary	PlantLocation
	Joe Steel	Foreman	65000	Edmonton

MACHINE	MachineNumber	Type	PlantLocation
	117	Infusinator	Edmonton

Assume the EMPLOYEE relation is fragmented horizontally by PlantLocation and each fragment is store locally at its corresponding plant site. Assume the MACHINE relation is stored in its entirty at the Armonk site. Describe a good strategy for processing each of the following queries.

(a) Find all employees at the plant containing machine number 1130.

(b) Find all employees at plants containing machines whose type is "Milling Machine"

(c) Find all machines at the Almaden plant.

(c) Compute EMPLOYEE $\bowtie$ MACHINE.

(9) Compute $r \bowtie s$ and s **semijoin** r for the following relations:

r:

A	B	C
9	1	8
2	3	5
6	4	3
5	6	7
2	3	4

s:

C	D	E
4	5	6
3	4	3
1	3	4
1	5	2
4	7	9

(10) If a hash structure is used on a search key for which range queries are likely, what property should the hash function have?

(11) Recall that the hash join algorithm (for the natural *inner* join $A \bowtie B$) basically adds the join attribute of B to the hash table, then hashes on A into the same table. If a match is found, then the join on that attribute succeeds. If no match is found, the join on that attribute does not succeed, but the attribute of A is *not* added to the hash table.

Briefly list the changes necessary to this algorithm for to handle a full *outer* join.

(12) Suppose that the bank used in the textbook's running example maintains a statistical database containing the average balances of all customers. The scheme for this relation is *(customer-name, customer-city, avg-balance)*. Assume that for security reasons, the following restrictions are imposed on queries against this data:

- Every query must involve at least 10 customers
- The intersection of any pair of queries must be at most 5.

Construct a series of queries to find the average balance of a customer. (Hint: this can be done in fewer than 7 queries.)

(13) Suppose that we decompose the schema r(A,B,C,D,E) into r_1(A,B,C) and r_2(A,D,E). Show that this decomposition is a lossless decomposition if the following set F of functional dependencies holds:

$A \rightarrow BC$

$CD \rightarrow E$

$B \rightarrow D$

$E \rightarrow A$

(14) Compute the closure of the set of functional dependencies from problem 13.

(15) Compute the canonical cover F_c of the set of functional dependencies from problem 13.

(16) Compute B^+ given the set of functional dependencies from problem 13.

(17) Use Armstrongs's axioms to prove the soundness of the pseudotransitivity rule.

(18) Given the databse schema $R(a, b, c)$, and a relation r on the schema R, write a SQL query to test whether the funciontal dpendency $b \rightarrow c$ holds on relation r. There is no need to deploy/test this SQL code.

(19) Write a SQL assertion that enforces the functional dependency from Question 17; assume that no null values are present. There is no need to deploy/test this SQL code.

(20) Show that the decomposition in problem 13 is not a dependency preserving decomposition.

(21) Give an example of a relational schema R' and the set F' of functional dependencies such that there are at least three distinct lossless decompositions of R' into BCNF.