CSE 444 Midterm Exam

November 13, 2009

Name ________Sample Solution__________

Question 1	/ 24
Question 2	/ 22
Question 3	/ 22
Question 4	/ 12
Question 5	/ 20
Total	/ 100

Question 1. SQL (24 points, 8 each part) With the holiday gift-giving season upon us, we'd like to set up a database to keep track of different products and stores that stock them. Our database has the following tables:

PRODUCT(pid, pname, manufacturer, color)
STORE(sid, sname, address, city, phone)
INVENTORY(pid, sid, price, quantity)

Table PRODUCT has a unique product id number, pid, for each product, and gives the name, manufacturer, and color of that product. Table STORE has a unique id number for each store, sid, and gives the store name, address, city, and phone number. Note that there may be more than one store with the same name in the same city – for example, there may be several 'Costco' stores in 'Seattle' – but the stores will have different addresses and store id numbers. The INVENTORY table gives a list of which stores carry which products, how many of each product are currently in stock at each store, and the product price at each different store. The pid and sid entries in this table are foreign keys referring to the PRODUCT and STORE tables respectively. The quantity in an INVENTORY table entry might be 0 if a particular store carries a product but currently has none in stock. Also, the same product may have different prices at different stores.

(a) Write a SQL query that gives a list of store names, addresses, and cities of all stores that sell iPods (pname is 'iPod') that have the color 'red'. The list should include all such stores, even if they currently have none in stock. The list should be sorted by city; beyond that, the other information about stores in a given city may appear in any order.

```
SELECT s.sname, s.address, s.city
FROM PRODUCT p, STORE s, INVENTORY i
WHERE p.pid = i.pid AND s.sid = i.sid AND p.pname = 'iPod' AND p.color = 'red'
ORDER BY s.city
```

(continued next page)

Question 1. (cont) Schemas repeated for convenience:

PRODUCT(pid, pname, manufacturer, color)
STORE(sid, sname, address, city, phone)
INVENTORY(pid, sid, price, quantity)

(b) Write a SQL query that gives the minimum available price of the product named 'Droid' from any store that has one or more 'Droid's in stock – don't consider any store that doesn't have at least one 'Droid' in its current inventory. You only need to produce the minimum price; you don't need to identify the store.

SELECT min(i.price)
FROM INVENTORY i, PRODUCT p
WHERE p.pname = 'Droid' AND i.pid = p.pid AND i.quantity > 0

(c) Write a SQL query that produces a table showing the availability by city of products with the name 'Xbox'. The results should have one entry for each city where an 'Xbox' is available. The entry for each city should show the city name, the total number of 'Xbox's currently available in all stores in that city and the average selling price of a 'Xbox' in that city (just compute the average of the individual store prices – don't consider the number of units in stock at each store). The results should be sorted by city.

SELECT s.city, SUM(s.quantity), AVG(i.price)
FROM PRODUCT p, STORE s, INVENTORY i
WHERE p.pid = i.pid AND i.sid = s.sid AND p.pname = 'Xbox'
GROUP BY s.city
ORDER BY s.city

Question 2. Database design (22 points) Continuing with our holiday theme, Santa has been having trouble keeping track of all of his workshops and elves. He has decided that he needs a database to organize all this information and has asked you to design it.

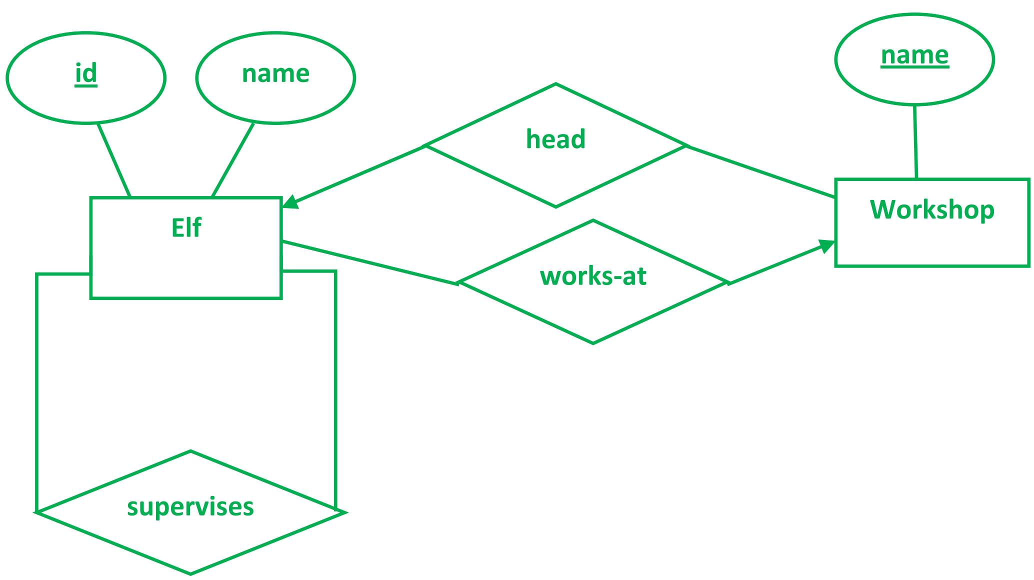
(a) Give an E/R diagram that captures the following information.

- Santa has several workshops. Each workshop has a unique name (North Pole, Northwest Branch, Toy Factory, Hoboken …), and a head elf (the elf in charge of the workshop). A single elf may be the head of more than one workshop. Although all workshops normally have a head elf, these positions are vacant from time to time.

- Santa has many elves working for him. An elf only works at one particular workshop, but some elves are between jobs and are not assigned to any particular workshop at the moment. Each elf has a name and a unique employee number.

- Some elves supervise other elves, and those elves might themselves have a supervisor who is a different elf. Most elves are just workers who don't supervise anyone. An individual elf may have more than one supervisor if that is appropriate for the particular elf's job. There are a few elves that don't have a supervisor. An elf who is the head of a workshop (see the first paragraph above) does not necessarily supervise other elves, but he/she might do so.

Draw your E/R diagram below:



(continued next page)

Question 2. (cont.) (b) Give a relational schema that captures your E/R diagram from part (a). You should give the table and attribute names and clearly indicate which attributes are keys and foreign keys in the various tables. Your tables should be in BCNF (i.e., contain no bad anomalies or functional dependencies), but you do not need to demonstrate this or prove that your design is in BCNF. You do **not** need to give SQL CREATE TABLE statements for your tables.

Notation: tbl.key means a foreign key in the given table.

Elf(<u>id</u>, name, Workshop.name)

Workshop(<u>name</u>, Elf.id)

Supervises(<u>Elf.id</u>, <u>Elf.id</u>)

Question 3. BCNF (22 points) Consider the relation R(A, B, C, D, E) that has the following functional dependencies:

$AB \rightarrow C$
$CD \rightarrow E$
$DE \rightarrow B$

Decompose this relation, if necessary, into a collection of relations that are in BCNF. For full credit, show your work, the intermediate decomposition steps, and show which dependency violations you are correcting at each step.

There are three different ways to get a BCNF decomposition, depending on which dependency is used first.

I. Use AB+ = {A, B, C} to decompose R into R1(A, B, C) and R2(A, B, D, E). Table R2 is not in BCNF because DE+ = {B, D, E} $\neq$ {A, B, D, E}. Use this to decompose R2 into R21(B, D, E) and R22(A, D, E)

II. Use CD+ = {B, C, D, E} to decompose R into R1(B, C, D, E) and R2(A, C, D). R1 is not in BCNF because DE+ = {B, D, E} $\neq$ {B, C, D, E}, so decompose it to get R11(B, D, E) and R12(C, D, E).

III. Use DE+ = {B, D, E} to decompose R into R1(B, D, E) and R2(A, C, D, E). R2 is not in BCNF beause of CD+ = {C, D, E}, so decompose it to get R21(C, D, E) and R22(A, C, D).

Question 4. Concurrency Control / Serialization (12 points) We have looked at several ways a database scheduler can manage transactions to either ensure that their operations are consistent with some conflict-serializable schedule, or detect operations that would violate serialization and either delay or roll back one or more transactions to prevent the problem.

Each of the different concurrency control schemes determines the effective serial order of transactions based on some operation(s) issued by the transactions. For example, one possibility is to ensure that the effective serial order is the same order in which the transactions issue their first read operation. Or the serial order might reflect the order in which the transactions commit or abort.

The significant events that occur during a transaction include some or all of the following:

(a) Start transaction
(b) First read operation issued by the transaction
(c) Last read operation issued by the transaction
(d) First write operation issued by the transaction
(e) Last write operation issued by the transaction
(f) All write operations have completed (i.e., are written to stable storage)
(g) First lock acquired by transaction
(h) Last lock acquired by transaction
(i) First lock released by transaction
(j) Last lock released by transaction
(k) Beginning of validation phase
(l) End of validation phase
(m) Transaction commit or abort (rollback) issued

Write the correct letter in front of each concurrency control scheme below to indicate the event that determines the effective serial order of transactions using that scheme. If a scheme doesn't absolutely guarantee to produce a serializable schedule, use the answer that best determines the effective order of the transactions.

_____**i**_____ Two Phase Locking (2PL)

_____**a**_____ Timestamps

_____**l**_____ Validation

_____**a**_____ Snapshot Isolation

Question 5. ARIES (20 points) After a crash, an ARIES log contains the following entries. There are no checkpoints in the log, so this is all of the information that is available.

LSN	prevLSN	Transaction ID	Type	pageID	Data
1	0	T1	U	P2	Y
2	0	T2	U	P2	Y
3	2	T2	U	P1	X
4	3	T2	C	--	--
5	0	T3	U	P2	Y

(Note: the log entries in this question go from top to bottom, unlike the example in class where they ran from left to right.)

(a) Using the information in this log, show the results of the analysis pass that is the first part of the ARIES crash recovery. Fill in the transaction table and dirty page table below with the information discovered during the analysis pass. If an entry in a table is created and then later changed or deleted during the analysis pass, show the entry and draw a line through all or part of it as appropriate to indicate the changes or deletions.

Transaction table

Transaction	lastLSN
T1	**1**
~~**T2**~~	~~**2, 3, 4**~~
T3	**5**

Dirty page table

Page	RecoveryLSN
P2	**1**
P1	**3**

(b) What is the FirstLSN, i.e., the LSN of the first log entry that needs to be redone during the redo phase of the crash recovery that follows the analysis?

1

(c) Which transaction or transactions are "loser" transactions that need to be undone during the undo phase of the crash recovery that follows the redo phase?

T1, T3