CS585 Midterm

Spring term, 2/22/19
Duration: 1 hour

Instructions/notes

- the exam is closed books/notes/devices/neighbors, and open mind :)
- there are 8 questions, a 'non-data-related' bonus, total = 35+1 points
- there are no 'trick' questions, or ones with long calculations or formulae
- **please do NOT cheat;** you get a 0 if you are found to have cheated
- **when time is up, stop your work**; you get a 0 if you continue

Q	Your score	Max possible score
1		4
2		5
3		4
4		3
5		4
6		5
7		5
8		5
Bonus		1
Total		**36**

Q1 (4 points).
Suppose an online vendor maintains its customer list like so:

firstName	lastName	address	city	state	ZIP	phoneNumber	SkypeID	emailAddress
A	B	123 Main St	Los Angeles	CA	90089	213-543-6543		AB@mail.com
Fam	Act	222 Burton Way	Beverly Hills	CA	90210		RichNFamous	RNF@imdb.com
MoreFam	Act	108 Roxbury St	Beverly Hills	CA	90210	323-654-1002		TheBest@BevHills.us
Grad	Student	154 Adams St	Los Angeles	CA	90089			DontBugMe@usc.edu

What two problems do you see with the above scheme, and how would you fix them?
Your answer can be in the form of E-R (using any notation), or in table format (like above) or even SQL. And, feel free to create any new attributes that might be necessary.

Q2 (5 points). Parents in a wealthy family want to create a DB of all their assets. For each asset, they would like to name benefactors - some or all of their five children who would get the asset. Each asset has a financial value associated with it, and a maturity date (when the kid(s) can cash in). They'd like to track the following diverse set of assets they own: bank accounts, real estate, stocks, jewelry, life insurance. **What would be a good design (using an ER diagram) for this?** You can make any assumptions you want about the assets, create whatever descriptors (columns) you need, etc.

Q3 (4 points). A reality company keeps track of its home sales like so:

Seller	Buyer	LendingBank
S1	B1	BofA
S2	B1	Chase
S1	B2	Chase

Things seem fine (redundancy and all), until they hire you to 'clean up' their table. After analysis, you come up with these three separate tables [all linked properly with FK/PK], which makes for good design:
Table 'SellerBuyer', with rows such as (S1,B1).
Table 'BuyerBank', with (B2,Chase) as a sample row.
Table 'SellerBank', eg. with (S2,Chase) as a row.

You write the following three-way 'join' query just for fun, to see if you can recreate the original triplets (eg. S1,B1,BofA):

```
SELECT SB.Buyer, SN.Seller, BN.LendingBank
FROM SellerBuyer as SB, SellerBank as SN, BuyerBank as BN
WHERE BN.Buyer=SB.Buyer
AND BN.LendingBank=SN.LendingBank
AND SN.Seller=SB.Seller
```

Question: what, if any, is the problem with the above query?

Q4 (2+1=3 points). You pull out your smartphone, log on to your banking app, and proceed to transfer $7200 (to pay for a 4-unit 'SC course!) from your savings account into your checking account. Prior to the transfer, you had $20,000 in savings and $800 in checking. While you are in the middle of doing this, due to poor DB design, a report generator (that would produce a monthly statement to email you) starts to run on the bank's server. **What could go wrong, and what is such a scenario called?**

Q5 (2+2=4 points). How would you optimize (by rewriting) the following two queries?

a. `SELECT * FROM TBL WHERE substr(STATE,1,1)='C'`
[we want to select all rows containing US states CA, CO, or CT; substr(<string>,1,1) returns just the first character of a string]

b. `SELECT * FROM TBL WHERE AGE>21`
[the AGE column stores ages as 0..99 integers; assume it has been indexed]

Q6 (3+2 = 5 points). For a while now, NASA has been conceptualizing a network called the Interplanetary Internet, which could come in handy 'someday' when we colonize Mars [when pigs fly out of our butts :)]. If that were to come to fruition, Eric Brewer's 'CAP theorem' would be highly relevant and applicable to such a distributed system of nodes. As per the CAP theorem, 'you can't always get what you want' (at least not C,A,P all at once, all equally guaranteed).

In an Interplanetary Internet, how would you rank C,A,P in terms of concerns? In other words, which would we worry about most, and relatively which, the least? You need to state why (justify your ordering).

Where might nodes be located, for an Interplanetary Internet? And, what disaster scenarios can you envision (that affect the network)?

Q7 (5 points). What operation does the following SQL query implement?

```
SELECT DISTINCT c
FROM A as tA
WHERE EXISTS (SELECT *
              FROM B as tB
              WHERE tA.c = tB.c);
```

Q8 (5 points). Here are a pair of tables – a PRODUCTS table that lists products a company sells, and SALES, which records sales of the products (each unit of a product that is sold, gets a separate row in SALES):

```
PRODUCTS(PRODUCT_ID, PRODUCT_NAME);
SALES(SALE_ID, YEAR, PRODUCT_ID, PRICE);
```

Consider the following three queries, we're calling them Q1, Q2, Q3. In Q2, fyi, 'SELECT 1' returns a 1, which we can ignore (it is not essential to our query).

```
SELECT S.PRODUCT_ID,SUM(PRICE)
FROM SALES S
 JOIN
 PRODUCTS P
 ON (S.PRODUCT_ID = P.PRODUCT_ID)
GROUP BY S.PRODUCT_ID;
```

```
SELECT S.PRODUCT_ID,SUM(PRICE)
FROM SALES S
WHERE EXISTS
(
 SELECT  1
 FROM PRODUCTS P
 WHERE P.PRODUCT_ID = S.PRODUCT_ID
)
GROUP BY S.PRODUCT_ID;
```

```
SELECT S.PRODUCT_ID,SUM(PRICE)
FROM SALES S
WHERE S.PRODUCT_ID IN
(
 SELECT  PRODUCT_ID
 FROM PRODUCTS P
)
GROUP BY S.PRODUCT_ID;
```

Circle the correct choice below:

a. Q1, Q2, Q3 are all different (they produce different results)
b. Q1, Q2, Q3 are all identical
c. Q1 and Q2 are identical
d. Q1 and Q3 are identical
e. Q2 and Q3 are identical

Bonus (1 point). Look at the flattened cube below on the left, and four cubes on the right - which of the four would produce the flattening?