CMPS 180, Final Exam, Fall 2016, Shel Finkelstein

Student Name: _____

Student ID: _____

UCSC Email: _____

Final Points

Part	Max Points	Points
I	18	
II	24	
III	18	
IV	20	
V	21	
Total	101	

The exam is double-sided, so you have extra space to write answers. If you use that extra space, please be sure to write the number of the problem that you're solving next to your answer.

At the end of the exam, please show your UCSC id when you hand in your exam booklet.

Part I: (18 Points, 3 each)

For questions in Part I, use relations with the following schemas. Underlined attributes are keys.

Sailors(<u>sid</u>, sname, rating, age)
        // For each sid, gives the sailor's name, rating and age

Boats(<u>bid</u>, bname, color)
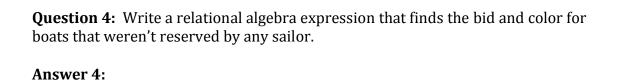   // For each bid, give the boat's name and color

Reservations(<u>sid, bid</u>, day)
        // For each reservation of a boat by a sailor, gives day sailor reserved boat

Some questions ask you to write SQL queries. Others ask you to write relational algebra expressions.

Question 1: Write a SQL statement that finds the sid and age for sailors who reserved at least one boat.

Answer 1:

Question 2: Write a relational algebra expression that finds the sid and age for sailors who reserved at least one boat.

Answer 2:

Question 3: Write a SQL statement that finds the bid and color for boats that weren't reserved by any sailor.

Answer 3:

Question 4: Write a relational algebra expression that finds the bid and color for boats that weren't reserved by any sailor.

Answer 4:

Question 5: Explain what the following relational algebra expression computes, where ⋈ is Natural Join.

      (π_{rating} (($\sigma_{color='red'}$ Boats) ⋈ ($\sigma_{day='11/07/2016'}$ Reservations) ⋈ Sailors))

  ∩ (π_{rating} (($\sigma_{color='green'}$ Boats) ⋈ ($\sigma_{day='11/07/2016'}$ Reservations) ⋈ Sailors))

Answer 5:

Question 6: For each equality, answer YES if it is <u>always</u> true, and answer NO if it isn't always true.

6a): $\sigma_{C1} (\sigma_{C2} (R)) = \sigma_{C2} (\sigma_{C1} (R))$

where C1 and C2 are conditions on relation R's attributes.

Answer 6a): _____

6b): $\pi_{A1} (\pi_{A2} (R)) = \pi_{A2} (\pi_{A1} (R))$

where A1 and A2 are sets of attributes of relation R.

Answer 6b): _____

6c): $(R - S) - T = R - (S - T)$

where R, S and T are union-compatible relations.

Answer 6c): _____

Part II: (24 Points, 4 each)

Question 7: A database has the relation Employees, with primary key emp_id, and other attributes giving name, department and salary of an employee.

Employees(emp_id, ename, edept, esalary)

A programmer writes a SQL query intending to find the name, salary and department for employees who make the most money in their departments.

```
SELECT e1.ename, e1.salary, e1.edept
FROM Employees e1
WHERE e1.salary > ALL
        ( SELECT e2.salary
          FROM Employees e2
          WHERE e1.edept = e2.edept );
```

7a): What error has the programmer made in writing this query? Fix the query so that it does what the programmer intended, making as few changes as possible.

Answer 7a):

7b): Suppose that there is only one department ('Marketing') in Employees, and the employees in that department are 'Smith', 'Jones' and 'Brown'. The salary of 'Smith' is 50000 and the salaries of 'Jones' and 'Brown" are NULL. What will be the output of the corrected 7a) query that the programmer intended to write?

Answer 7b):

Question 8a): If S(A,B,C) is a relation where A's domain is (a1, a2, a3, a4) but A can't be NULL, and B's domain is (b1, b2, b3, b4) but B <u>can</u> be NULL, and C domain is (c1, c2) but c <u>can</u> be NULL, what is the maximum number of <u>different</u> tuples that can be in relation S?

Answer 8a: _____

8b): Write a SQL statement that updates tuples of S in which B has the value NULL but C does not, setting the B value for those tuples to be b2.

Answer 8b:

Question 9: Define the Transaction Isolation levels Read Committed, and Repeatable Read, explaining the difference between them.

Answer 9:

Question 10: Briefly explain two different advantages of using Stored Procedures versus other Application Programming approaches.

Answer 10:

Question 11: For the following table.

```
CREATE TABLE Beers (
       name   CHAR(20) PRIMARY KEY,
       manf   CHAR(20) );

CREATE TABLE Sells (
       bar     CHAR(20),
       beer    CHAR(20),
       price   REAL );
```

Rewrite the CREATE statement for Sells (don't do an ALTER) so that the beer attribute of Sells is a Foreign Key that correspond to name, the primary key of BEERS. Do the rewrite in 2 different ways for parts a) and b):

11a): First, rewrite the CREATE so that a Beers row can't be deleted if there is a bar that sells the beer.
11b) Second, rewrite the CREATE so that if a Beers row is deleted, all the Sells rows for bars that sell that beer also are deleted.

(Don't be concerned about handling of Updates of the name attribute of Beers.)

Answer 11a):

Answer 11b):

Question 12:

12a): Explain in words exactly what the regular expression (in bold) for person in the following address book DTD requires.

```
<!DOCTYPE addressbook [
  <!ELEMENT addressbook (person*)>
  <!ELEMENT person
     (name, address?, ( homephone | ( workphone, mobile )* ), email+)>
  <!ELEMENT name          (#PCDATA)>
  <!ELEMENT address       (#PCDATA)>
  <!ELEMENT homephone   (#PCDATA)>
  <!ELEMENT workphone    (#PCDATA)>
  <!ELEMENT mobile        (#PCDATA)>
  <!ELEMENT email          (#PCDATA)>
]>
```

Answer 12a):

12b): Does the following data conform to that DTD?   (YES or NO)

```
<person>
      <name> MacNiel, John </name>
      <addr> Rome, OH 98765 </addr>
       <homephone> 718 555 3089 </homephone>
      <workphone> (903) 555 9876 </workphone>
      <mobile> (321) 555 2543 </mobile>
      <email> macniel@ucsc.edu </email>
</person>
```

Answer 12b): _____

Part III: (18 points, 3 each):

Answer the following questions with **YES/TRUE** or **NO/FALSE**.

Question 13: For a database with the following relations, with primary keys underlined:

Employees(<u>emp_id</u>, ename, edept, salary)
Departments(<u>dept_id</u>, dmanager, daddress)

Are the following two queries always equivalent?

SELECT ename, dmanager, SUM(salary),
FROM Employees, Departments
WHERE edept = dept_id
  AND dmanager<> 'Turing'
GROUP BY ename, dmanager;

SELECT ename, dmanager, SUM(salary),
FROM Employees, Departments
WHERE edept = dept_id
GROUP BY ename, dmanager
HAVING dmanager <> 'Turing';

Answer 13: _____

Question 14: When optimizing a query tree for execution, it is a good idea to push selection predicates and projections as far down in the tree as possible, to avoid extraneous rows and attributes. But you can't rid of an attribute if it will be needed later to evaluate a predicate, or as part of a query answer.

Answer 14: _____

Question 15: The 5 relational algebra operators product (x), projection (π), selection (σ), union ($\cup$) and difference (–) are independent of each other, meaning that none of those can be expressed used the other. However, it's possible to express join, natural join, intersection and division using the 5 relational algebra operators.

Answer 15: _____

Question 16: Is every binary relation R(A,B) in Boyce Codd Normal Form, no matter what functional dependencies R has?

Answer 16: _____

Question 17: A set F of Functional Dependencies (FDs) implies an FD F (written $F \vDash$ F) if every instance r that satisfies FDs F must also satisfy the FD F.

Now suppose that we find an instance r that satisfies F and that also satisfies F. YES or NO, does that prove that $F \vDash$ F?

Answer 17: _____

Question 18: For OLAP, with a star schema, there is a foreign key constraint between the Fact table and each of the Dimension tables. However, there can be combinations of Dimension table values for which there is no row in the Fact table.

Answer 18: _____

Part IV: (20 points, 5 each):

The questions in Part III are about the following tables, which should be familiar:

Stores (<u>store_id</u>, name, address, manager)
Customers (<u>customer_id</u>, name, address, email)
Products (<u>product_id</u>, name, category, manufacturer)
Sales (<u>store_id, customer_id, product_id, purchase_date</u>, quantity, unit_price, shipped)

Question 19: Write a SQL statement that creates the Products table with product_id as a primary key. Ensure that there can't be two products that have both the same manufacturer and name. The default manufacturer should be 'Apple'. Category can be NULL, but the other attributes can't be NULL. product_id is an integer, and name category and manufacturer are character strings of length 20.

Answer 19:

Question 20:

Here is a Create Statement for the Sales Table.

CREATE TABLE Sales
(store_id INT,
  customer_id INT,
  product_id INT,
  purchase_date DATE,
  quantity INT,
  unit_price DECIMAL(6,2),
  shipped BOOLEAN,
  PRIMARY KEY(store_id, customer_id, product_id, purchase_date));

Rewrite this CREATE statement (don't do an ALTER) so that store_id, customer_id and product_id are Foreign Keys that correspond to the keys of Stores, Customers and Products, respectively. The policies for referential integrity should be:

- Deleting a customer also deletes all Sales involving that customer.
- Products can't be deleted if there are Sales of that product.
- Deleting a store deletes all Sales that were made in that store.

Also, the total price for any sale (which is quantity multiplied by unit_price) must be less than 2000.

Answer 20:

Question 21:

21a): Write a view BacklogProds that finds the product_id and product name for all products whose manufacturer is 'Kelloggs' where that product has at least one sale in Sales that has not been shipped. Each product should be listed only once, even if there are multiple unshipped sales of that product.

Answer 21a):

21b): Write a query that does the following: For each product that is in the BacklogProds view, give the product_id, product name, and the total quantity that **has** been shipped for that product.

Answer 21b):

Question 22:

22a): Write a statement that creates an index on the attributes purchase_date and unit_price of the Sales table.

Answer 22a):

22b): For which of these queries would the index you created in part a) work best?

Query 1:

  SELECT *
  FROM Sales
  WHERE purchase_date = '11/29/2016'
    AND unit_price < 20;

Query 2:

  SELECT *
  FROM Sales
  WHERE unit_price = 20
    AND purchase_date < '11/29/2016';

Answer 22b): _____

22c: Indexes help process queries faster, but they also have significant disadvantages. Provide two different disadvantages associated with having indexes.

Answer 22c):

Part V: (21 points, 7 each):

Question 23: Suppose that you have a relation R(A,B,C), and that an instance of R has rows (1,2,3), (4,2,3) and (5,3,3).

23a) Put an X next to a listed functional dependency if you can determine that it <u>does not hold</u> for R? Otherwise, leave the line blank.

_____   A → B

_____   BC → A

_____   B → C

23b): Are there any functional dependencies that you can determine <u>do hold</u> for R? If so, give an example. If not, explain why not.

Answer 23b):

Question 24:

24a): As part of design theory, we discussed anomalies, which database designers often try to avoid. Using an example, explain the deletion anomaly problem.

Answer 24a):

24b): If you have a relation R and a set of functional dependencies F, what is a lossless-join decomposition for R with respect to F?

Answer 24b):

24c): Give an example of a relation R with no functional dependencies and a decomposition of R that is <u>not</u> a lossless-join decomposition.

Answer 24c):

Question 25: Suppose that you have a relation R(A,B,C,D,E), with the following functional dependencies: A → B, BC → E, and ED → A.

25a): Is R in BCNF? Explain your answer.

Answer 25a):

25b): Is R in 3NF?   Explain your answer.

Answer 25b):

25c): For each of the following, indicate whether it's a candidate key for R by writing YES or NO in the blank.

Answer 25c):

_____   ABC

_____   ACDE

_____   BCD