Name _____   RUID _____

CS 198:336 Midterm Exam 1 - Spring 2019
Section 4, 5, 6 (Circle one)

Please silence your cell phone and put all phones away (either put them in the bag, or give them to the proctors). Proctors will not collect your test if found that your cell phone(s) is/are with you either in your pocket, or on the table.

- Please put your initials on all pages

- All relational table columns or ER attributes are non-null unless explicitly specified in a question part.

- Two queries are equivalent if they return the same answer (including duplicates) over all possible database states

Part 1: U.S Athletes Database [Total: 20 points]

We want to design a database for storing the information of U.S. athletes and their participation in Olympic games.

- An athlete has his/her SSN, the name, and the birth date.

- Each Olympiad has the year, the country, and the city. For example, the 2008 Olympic game was held in Beijing, China.

- An athlete can participate in several sports in an Olympiad. For example, Michael Armstrong participated in the swimming, diving, and water polo events.

- Every athlete in our database participated in at least one Olympiad. Every Olympiad has at least one athlete (and, actually, several thousand).

Question 1.1 Draw an ER diagram for this database. Make sure to indicate primary keys, cardinality constraints, weak entities (if any), and participation constraints. You don't really need a separate sport entity. Instead, a simple attribute on the participate relation would suffice. Same assumption applies to Question 1.2 **[10 points]**

Question 1.2 Turn the ER diagram into tables: give SQL statements to create the tables. Make sure to indicate primary keys, foreign keys, if any. Specify the appropriate integrity constraints to enforce the cardinality and participation constraints. **[10 points]**

Part 2: Basic SQL [Total 25 points]
For questions below, consider a table `states`, with three columns `name`, `income`, and `population`, representing attributes of U.S. states.
Question 2.1 Write a SQL query that returns the population-weighted average income over all states. **[5 points]**

Question 2.2 Write a SQL statement that adjusts the income of each state: if it is positive, then it should be increased by 3%, otherwise left unchanged. **[5 points]**

Question 2.3 Execute a query that displays all employees who earn more than 40000 and are either in department 3 or are male. Display all the attributes. **[5 points]**
Schema: `Employee(emp_first_name, emp_last_name, emp_gender, emp_dpt_number, emp_salary)`

Question 2.4 Execute a query that diplays all employees whose last name contains the letter 'o'. List each employee's first and last name. **[5 points]**
Schema: `Employee(emp_first_name, emp_last_name, emp_gender, emp_dpt_number)`

Question 2.5 The company's Executive Director for project manager needs to know the number of projects each department controls based on the information stored in the project table. **[5 points]**

- You need to refer to the `pro_dept_number` column that stores the department number and the `pro_number` column that stores the project number information.

- The result table should show the department number and the number of projects that each department works on.

Part 3: Relational Algebra [Total 10 points]
A network intrusion refers to an activity where an attacker sends malicious packets to a victim to do malicious things (e.g. crash the victim machine, or gain root access of the victim machine). A network security company monitors the network and records the network intrusion events. For the purpose, it has the following two tables in its database. The table `MACHINE` contains the attacker or victim information and it has the following three fields:

- `mname`: the machine name of the attacker or the victim; e.g. `andromeda.rutgers.edu`.

- `ipaddr`: the IP address of the attacker or the victim; e.g. `9.9.9.9`.

- `os`: the operating system of the attacker or the victim; e.g. Linux, Windows.

The primary key of the MACHINE table is ipaddr.
The table EVENT contains the attack information and it has the following three fields:

- a-ipaddr: the IP address of the attacker; e.g. 9.9.9.9.

- v-ipaddr: the IP address of the victim; e.g. 10.10.10.2.

- date: the date the attack was performed; e.g. 2/2/2013.

The primary key of the EVENT table includes all three fields.

EVENT		
a-ipaddr	v-ipaddr	date
9.9.9.1	10.10.10.2	2/1/2013
9.9.9.2	10.10.10.1	2/2/2013
9.9.9.2	10.10.10.3	2/2/2013
9.9.9.9	10.10.10.4	2/3/2013
9.9.9.9	10.10.10.3	2/4/2013
9.9.9.10	10.10.10.3	2/5/2013
9.9.9.10	10.10.10.4	2/6/2013

Consider the following expression:

$$\pi_{\text{a-ipaddr, v-ipaddr}}(\text{EVENT}) \div \pi_{\text{v-ipaddr}}[\sigma_{\text{v-ipaddr}=\text{"9.9.9.9"}}(\text{EVENT})]$$

Question 3.1 Describe in English what the query tries to do. A (possibly incorrect) answer is "It gives all the attackers on 2/1/2013." **[2 points]**

Question 3.2 How many columns are returned in the output? **[2 points]**

Question 3.3 What are the column headers of the output? **[2 points]**

Question 3.4 How many rows are returned in the output? **[2 points]**

Question 3.5 Give the contents of the returned rows. **[2 points]**

PART 4: NATURAL AND OUTERJOIN [5 points each; total 25 points]

| TABLE T1 | | | TABLE T2 | | |

P	Q	R
10	a	5
15	b	8
25	a	6

A	B	C
10	b	6
15	c	3
25	b	5

Consider the two tables T1 and T2 shown in figure above. Show the results of the following operations:

a. T1 **NATURAL JOIN** T1.P = T2.A T2

P	Q	R	A	B	C

b. T1 **NATURAL JOIN** T1.Q = T2.B T2

P	Q	R	A	B	C

c. T1 **LEFT JOIN** T2.P = T2.A T2

P	Q	R	A	B	C

d. T1 **RIGHT JOIN** T1.Q = T2.B T2

P	Q	R	A	B	C

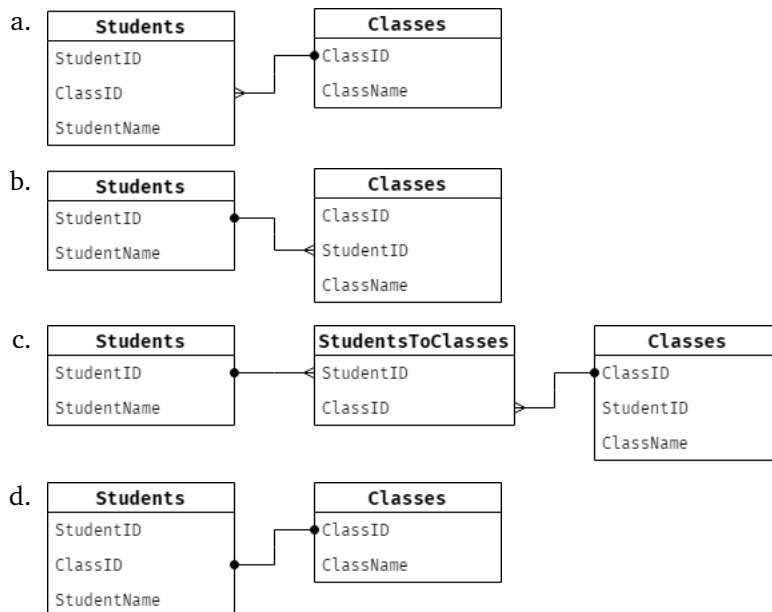e. Tl **NATURAL JOIN** (T1.P = T2.A AND T1.R = T2.C) T2

P	Q	R	A	B	C

Part 5: Multiple Choice [2 points each, total 10 points]

Question 5.1 Which of the following diagrams best represents the relationship between a table of customers and a table of orders placed by customers?

a.

Customers		Orders
CustomerID		OrderID
OrderID		OrderDate
CustomerName		

b.

Customers		Orders
CustomerID		OrderID
CustomerName		CustomerID
		OrderDate

c.

Customers	OrdersToCustomers	Orders
CustomerID	CustomerID	OrderID
CustomerName	OrderID	OrderDate

d.

| Customers		Orders
CustomerID		OrderID
OrderID		OrderDate
CustomerName		
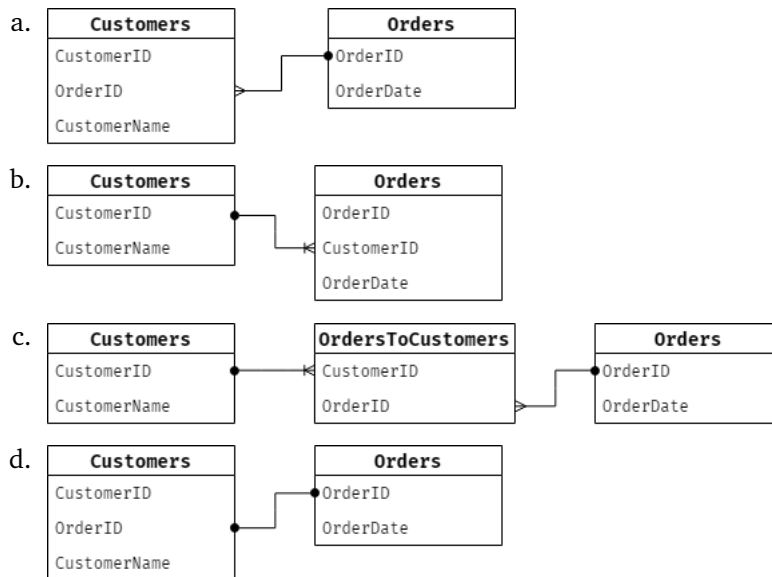
Question 5.2 Which of the following diagrams best represents the relationship between a table of students and a table of classes for which the student is registered?

a.

Students		Classes
StudentID		ClassID
ClassID		ClassName
StudentName		

b.

Students		Classes
StudentID		ClassID
StudentName		StudentID
		ClassName

c.

Students	StudentsToClasses	Classes
StudentID	StudentID	ClassID
StudentName	ClassID	StudentID
		ClassName

d.

Students		Classes
StudentID		ClassID
ClassID		ClassName
StudentName		

Question 5.3 Based on the figure, which column or columns in each table should be defined as the primary key?

a. Orders: OrderID
   OrderLineItems: OrderID
   Products: ProductID

b. Orders: OrderID and OrderDate
   OrderLinelItems: OrderID and OrderSequence
   Products: ProductID

c. Orders: OrderID
   OrderLineItems: OrderID and OrderSequence
   Products: ProductID

d. Orders: OrderID
   OrderLineItems: OrderID, OrderSequence, and ProductID
   Products: ProductID and ProductName

Orders	OrderLineItems	Products
OrderID	OrderID	ProductID
OrderDate	OrderSequence	ProductName
	ProductID	
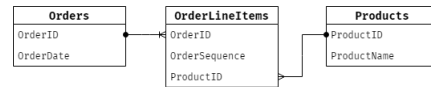
Question 5.4 A join that returns records from related tables only if their related fields match is called

a. an outer join

b. an inner join

c. a cross join

d. a data join

Question 5.5 To add a column to a table after a database has been created, you can use the

a. ALTER TABLE statement

b. ALTER COLUMN statement

c. ALTER DATABASE statement

d. ADD COLUMN statement

PART 6: REIFICATION [Total 10 points]

1. Take a simple relationship S between entities A and B, where B's participation is total, and S has attribute p. Give its diagram as a reified relationship. Also state one reason why reified relationships are useful. You can make an ER diagram to show the concept of the S as a relationship attribute. Now S is a new table but how to get the key for the new table S? (Be brief) **(5 points)**

2. How would you solve the Ternary relationship (Developer, Customer, and Project) (Be brief) **(5 points)**