HW2: SQL

TOTAL POINTS: 1+1+1+1+2=6 [AND 1+1 BONUS POINTS!]

IN THIS ASSIGNMENT, YOU WILL CODE SOLUTIONS TO THE **FIVE** SQL PROBLEMS GIVEN BELOW.

ALL THE SQL KNOWLEDGE/COMMANDS YOU NEED TO ANSWER THE QUESTIONS HAVE BEEN COVERED IN CLASS! YOU **DO NOT** NEED TO LEARN MORE COMMANDS OR TECHNIQUES (EG. USE OF 'TRIGGERS') ETC. ON YOUR OWN IN ORDER TO DO THIS HW SET.

TO RUN SQL CODE, YOU CAN USE ONE OF THE THREE WAYS MENTIONED IN THE LECTURE NOTES [A LOCALLY INSTALLED DB, OR A REMOTE SERVER-BASED DB VIA AN ONLINE SHELL RUNNING IN A BROWSER PAGE, OR A CLOUD DB] TO DO THE PROBLEMS.

One more cloud-y way, not mentioned in class, to do your homework (and continue learning and practicing SQL) is to use https://livesql.oracle.com/ (https://livesql.oracle.com/); after you sign up for a free account, you can create tables, insert rows and do queries - you can save all your work in separate sessions, reload [any/all of] them after logging in again in order to recreate your tables+data and rerun queries, and also add more tables/data/code to the mix - it's very convenient and cool, try it!

What you need to submit are text files with the SQL commands that you come up with, one file for each question (Q1.sql, Q2.sql.. Q5.sql). PLEASE MENTION AT THE TOP OF EACH FILE, which database (eg. Oracle, SQLite..) you used for that question! Your grader(s) will execute the SQL commands from the text files you submit, using the same software you used, to see if they produce the expected results.

YOU CAN TALK TO YOUR FRIENDS/CLASSMATES TO INFORMALLY DISCUSS APPROACHES TO THE PROBLEMS, BUT DO NOT {'SHARE'/LOOK UP/ASK OTHERS FOR} ACTUAL CODE - THAT WOULD BE PLAGIARISM, AND WILL GET YOU AN 'F' IN THE COURSE.

Please see your TAs/graders if you need one-on-one help (or see me). You can also post (and reply!) in the Piazza 'hw2' forum. Don't wait, start early. Good luck, have fun!

Q1 (1 POINT). Here is a table for recording guests' stays at a hotel (arrDate denotes arrival date, depDate is departure date):

```
CREATE TABLE HotelStays
(roomNum INTEGER NOT NULL,
arrDate DATE NOT NULL,
depDate DATE NOT NULL,
guestName CHAR(30) NOT NULL,
PRIMARY KEY (roomNum, arrDate));
```

THERE ARE TWO PROBLEMS (ISSUES) WITH THE ABOVE. FIRST, THE ARRIVAL DATE COULD BE INCORRECTLY ENTERED TO BE LATER THAN THE DEPARTURE DATE.

SECOND, A NEW ENTRY (FOR A NEW GUEST) COULD BE ACCIDENTALLY PUT IN FOR A ROOM NUMBER, EVEN BEFORE THE EXISTING GUEST IN THAT ROOM HAS CHECKED OUT:

```
INSERT INTO HotelStays(roomNum, arrDate, depDate, guestName)
VALUES
(123, to_date('20160202', 'YYYYYMMDD'), to_date('20160206','YYYYYMMDD'),
(123, to_date('20160204', 'YYYYYMMDD'), to_date('20160208','YYYYYMMDD'),
(201, to_date('20160210', 'YYYYYMMDD'), to_date('20160206','YYYYYMMDD'),
;
```

NOTE THAT THE TO_DATE() FUNCTION IN THE ABOVE INSERT STATEMENT IS SPECIFIC TO POSTGRESQL (I USED V.9.3). SO IF YOU WANT TO USE THE ABOVE LINE AS-IS, YOU NEED TO RUN IT UNDER POSTGRES; ALTERNATELY, YOU CAN USE ANY OTHER RDB (ORACLE, MYSQL..) - JUST BE SURE TO USE THE APPROPRIATE DATE REPRESENTATION. IN OTHER WORDS, THE ABOVE STATEMENT IS THERE JUST AS AN EXAMPLE, YOU CAN MODIFY IT HOWEVER YOU WANT.

How would you **redesign the table** to fix both these issues? For your answer, you can either provide a textual explanation, and/or provide SQL statements. Hint - "do not be concerned with efficiency" - ANY working solution is acceptable:)

Q2 (1 POINT). GIVEN THE FOLLOWING ENROLLMENT TABLE, **WRITE A QUERY** TO CREATE A LISTING THAT INCLUDES COURSE NAME AND THE NUMBER OF STUDENTS ENROLLED IN THE COURSE, SORTED IN REVERSE ORDER OF ENROLLMENT.

SID	ClassName	e Grade
123	ARCH201	А
123	QUAN432	В
662	AMER352	В
662	ECON966	Α
662	QUAN432	В
345	QUAN432	Α
345	ECON966	D

GIVEN SOMETHING LIKE THE ABOVE, THE OUTPUT COULD BE:

ClassName	Total
QUAN432	3
ECON966	2
AMER352	1
ARCH201	1

Q3 (1 POINT). BELOW IS A SMALL TABLE THAT TRACKS WORK ORDERS FOR PROJECTS. WE HAVE A PROJECT ID COLUMN ON THE LEFT, A 'STEP' COLUMN IN THE MIDDLE (0,1,2.. DENOTE STEPS OF THE PROJECT), AND A STATUS COLUMN ON THE RIGHT (WHERE 'W' DENOTES 'WAITING', 'C' DENOTES 'COMPLETED'). SUCH A TABLE LETS PROJECT MANAGERS GET A QUICK STATUS ON THE VARIOUS ASPECTS (STEPS) OF THEIR PROJECTS ("WHERE THEY'RE AT", IN COLLOQUIAL, UNGRAMMATICAL LANGUAGE).

ProjectID	Step	Status
P100	0	С
P100	1	W
P100	2	W
P201	0	С
P201	1	С
P333	0	W
P333	1	W
P333	2	W
P333	3	W

WRITE A QUERY TO OUTPUT THE PROJECT(S) WHERE ONLY STEP O HAS BEEN COMPLETED, IE. THE PROJECT GOTTEN STARTED BUT THE REST OF THE STEPS ARE IN WAITING MODE. IN THE ABOVE TABLE, SUCH A QUERY WOULD OUTPUT JUST 'P100'. YOU CAN ASSUME THAT STEPS GET COMPLETED IN ORDER, IE. P333 WILL NEVER HAVE C,W,C,W FOR EXAMPLE [ALL CS WILL OCCUR BEFORE ALL WS].

Q4 (1 POINT). BELOW IS A SMALL SAMPLE OF A JUNKMAIL DATABASE WE OWN, IE. PEOPLE WE WANT TO 'SPAM' VIA POSTAL MAIL [LOL].

Name	Address	ID	SameFam
'Alice'	'A'	10	NULL
'Bob'	'B'	15	NULL
'Carmen'	'C'	22	NULL
'Diego'	'A'	9	10
'Ella'	'B'	3	15
'Farkhad'	'D'	11	NULL

EACH ENTRY CONSISTS OF A NAME, ADDRESS, ID, AND WHETHER THERE IS A PRIOR FAMILY MEMBER ALREADY IN THE DB; IN THE LAST COLUMN, NULL MEANS THAT ENTRY IS THE FIRST FAMILY MEMBER IN OUR DB, AND A NON-NULL VALUE IS THE ID OF THE FIRST FAMILY MEMBER (EG. DIEGO POINTS TO ALICE, AND ELLA POINTS TO BOB).

WRITE A QUERY TO DELETE FROM THE TABLE, NAMES THAT HAVE 'NULL' FOR SAMEFAM AND ANOTHER FAMILY MEMBER IN THE DB. SO IN OUR EXAMPLE ABOVE, THE QUERY WOULD DELETE ALICE AND BOB, BUT NOT CARMEN AND FARKHAD. IN OTHER WORDS, WE WANT TO SAVE POSTAGE BY ONLY SENDING ONE MAIL PER HOUSEHOLD IF WE HAVE TWO PEOPLE FROM A HOUSE IN OUR DB (WE ASSUME THAT OUR DB CONTAINS EITHER ONE PERSON PER HOUSEHOLD, OR TWO).

Q5 (2 POINTS).WE ARE A DESSERT CATERING COMPANY. BELOW IS A TABLE OF CHEFS WHOSE SERVICES WE ACCESS ON AN AS-NEEDED 'GIG' BASIS, AND THE YUMMY DESSERTS THEY KNOW TO MAKE [FOR FUN, GOOGLE (IMAGE SEARCH) ONES YOU AREN'T FAMILIAR WITH!]:

Chef	Dish
'A'	'Apple pie'
'A'	'Beignet'
'A'	'Andes Chocolate Cake'
'B'	'Tiramisu'
'B'	'Creme brulee'
'B'	'Beignet'
'C'	'Tiramisu'
'C'	'Creme brulee'
'D'	'Apple pie'
'D'	'Tiramisu'
'D'	'Creme brulee'
'E'	'Apple pie'
'E'	'Bananas Foster'
'E'	'Creme brulee'
'E'	'Tiramisu'

WRITE A QUERY THAT WILL PICK OUT CHEFS WHO CAN MAKE EVERY ITEM IN THE TABLE BELOW (AN ORDER FROM SOMEONE WHO WANTS US TO CATER TO A WEEKEND PARTY; WE'D LIKE TO WORK WITH JUST THOSE CHEFS WHO CAN EACH MAKE AND DELIVER ALL ITEMS):

```
'Creme brulee'
'Apple pie'
'Tiramisu'
```

WITH THE ABOVE DATA, THE QUERY WOULD OUTPUT

Chef			
'D' 'E'			
'E'			

You can hardcode the dishes just for submission purposes, but your query should work for ANY such table! Using comments in the code, **YOU NEED TO EXPLAIN IN YOUR OWN WORDS, WHAT THE QUERY DOES** (How it works); don't just say things like "Now I'm using a WHERE condition", instead explain what it's for (why you're using it).

Bonus(1(+1) Point(s)). You will get 1 extra point if you can reformulate the query in a very different way! If you do this, submit a separate text file with the code, eg. Q5_v2.sql. If you come up with YET ANOTHER very different way, you can get 1 more bonus point (submit yet another file, eg. Q5_v3.sql). 'Very different' means just that - the approaches do have to be totally distinct, eg. you can't use NOT to invert an existing solution, or use IN() instead of OR, etc. Sooo... is this actually possible [to do it in two, or three, different ways]? Yes!

TO REITERATE, EVERYTHING YOU NEED IS IN THE SLIDES (THE MATERIAL WE WENT THROUGH IN CLASS) - YOU DO NOT NEED TO READ AHEAD OR LOOK UP MORE COMMANDS ONLINE! LOOK AT EACH RELATIONAL OPERATOR, EACH COMMAND, EACH FUNCTION, EACH KEYWORD THAT WE COVERED, AND ASK YOURSELF HOW IT COULD BE OF USE IN CONSTRUCTING YOUR QUERY.

YOU'LL **LOSE POINTS** IF YOU:

- · DON'T NAME YOUR FILES PROPERLY
- · SUBMIT A .ZIP (YOU NEED TO SUBMIT INDIVIDUAL FILES INSTEAD)
- · NEGLECT TO MENTION WHAT DB SOFTWARE YOU USED
- · DON'T EXPLAIN PROPERLY, IN Q5, WHAT THE QUERY DOES
- · SUBMIT/RESUBMIT AFTER THE DEADLINE (BE SURE TO DO IT AHEAD OF TIME, AND MAKE SURE YOU DID)

GOOD LUCK, **HAVE FUN** FIGURING IT ALL OUT!!