Homework 2 – SQL

Due date: June 11, 2018 at 11:59PM.

Total Point: 8

Note: The deadline is firm. DO NOT submit the homework at the last minute because you may fail. Some students experienced this for homework 1. Please submit a version of your homework at least 1 hour before the deadline to make sure that you have at least one version to be graded. You can re-submit an infinite number of versions before the deadline. We will not allow late submissions this time!

In this assignment, we will use Google Cloud SQL to work with SQL queries. This will help us learn how to use cloud services as well as run code on SQL. The document is divided into several parts. Parts 1, 2 and 3 help you set up the platform. You have to write SQL queries for questions at Part 4. Good luck!

Hint: For complex queries, you should start with simple query and increase its complexity incrementally.

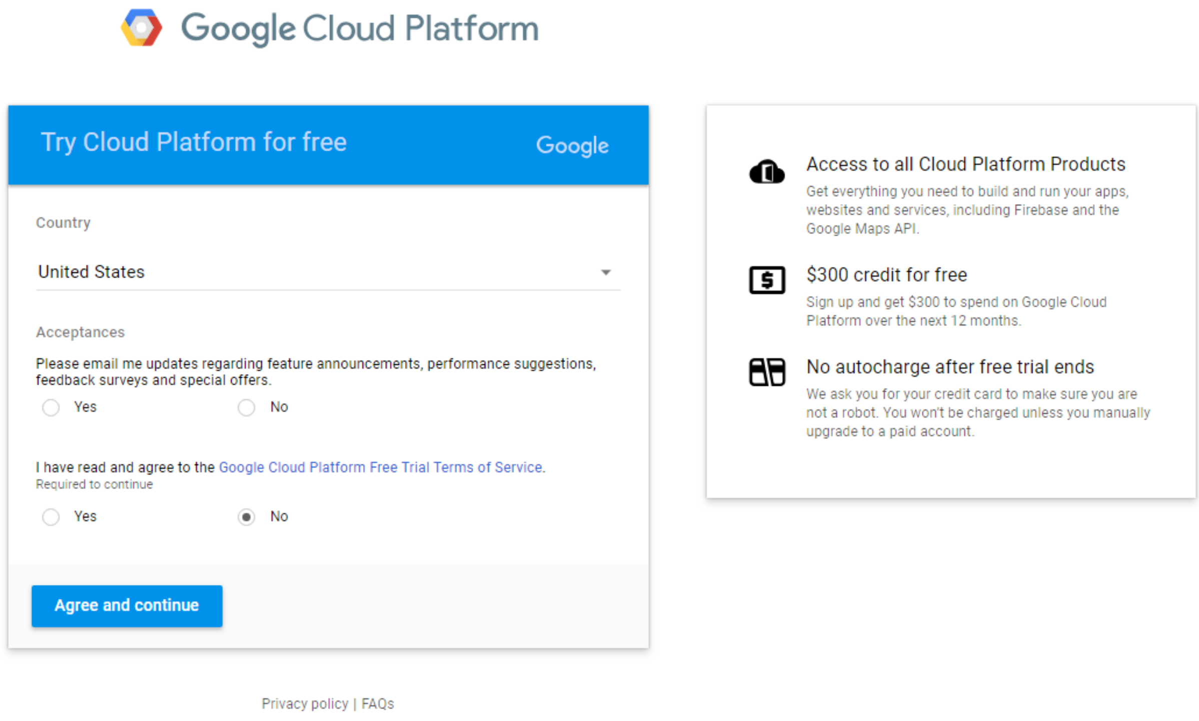
Part 1: Setting up Google Cloud Platform

Google Cloud Platform helps you to run your work on Google Compute Engine (GCE) and to use its core infrastructure, data analytics and machine learning.

To set up GCP, follow the steps below.

1. Go to https://console.cloud.google.com/freetrial using your personal Google account. (Do not use the USC account.) This gives you a $300 credit to spend on the GCP for the next 12 months. Agree to the acceptances and click on "Agree and Continue", as in Figure 1.1.
2. In the Customer info screen, select Account Type: Individual and fill in all the details. (You will be required to enter credit/debit card details, but you will not be charged as long as you don't exceed the $300, which is sufficient for all the database assignments.)
3. Click on "Start my Free Trial" and wait for Google Cloud Platform to set up.

Congratulations! You just finished the first part of the assignment.



Google Cloud Platform

Try Cloud Platform for free — Google

Country

United States

Acceptances

Please email me updates regarding feature announcements, performance suggestions, feedback surveys and special offers.

○ Yes ○ No

I have read and agree to the Google Cloud Platform Free Trial Terms of Service.
Required to continue

○ Yes ● No

Agree and continue

Access to all Cloud Platform Products
Get everything you need to build and run your apps, websites and services, including Firebase and the Google Maps API.

$300 credit for free
Sign up and get $300 to spend on Google Cloud Platform over the next 12 months.

No autocharge after free trial ends
We ask you for your credit card to make sure you are not a robot. You won't be charged unless you manually upgrade to a paid account.

Privacy policy | FAQs

Figure 1Google Cloud Platform Sign-up Page.

Notes on Google Cloud Pricing:

If you go to https://cloud.google.com/free/docs/always-free-usage-limits you will see that there are certain usage items that are always free. For example, under the Google App Engine 28 instance hours/day and 5GB Cloud Storage are just two of several items. The Google Cloud Datastore offers 1GB storage and 50,000 reads, 20,000 writes and 20,000 deletes for free. There are many other aspects of the Google Cloud that include free elements. Unfortunately, the MySQL and PostgreSQL are not one of them. As a result, the $300 free credit will kick in immediately.

Part 2: Setting up Cloud SQL

Cloud SQL is a part of the GCE to run PostgreSQL and MySQL scripts.

Go to https://cloud.google.com/sql/.

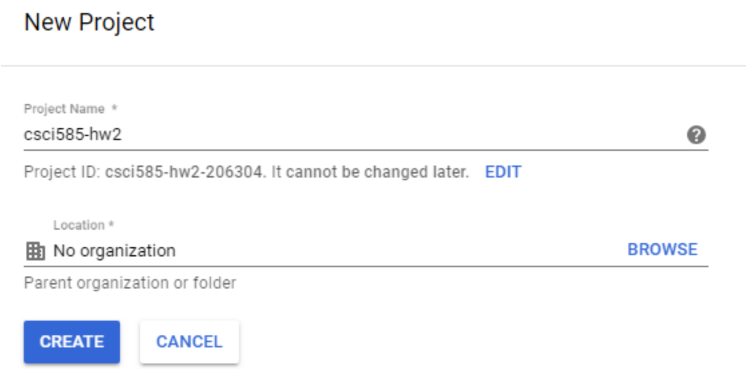
If you prefer to use MySQL for this assignment, you can find the Quick Start guide at: https://cloud.google.com/sql/docs/mysql/quickstart.

If you prefer to use PostgreSQL instead, visit https://cloud.google.com/sql/docs/postgres/quickstart.

The pages are self-explanatory, and in case you do not face any problem setting it up, feel free to skip the rest of Part 2. Below are detailed steps from the same page.

Before you begin:

1. Select or create a Cloud platform project:
   Go to: https://console.cloud.google.com/start. At the top, click on 'Select a project', and click on "New Project".
2. In the next screen, (as in Figure 2), enter a project name.
3. Enable billing for your project. In the navigation drawer on the left, click on Billing and add a billing account. Follow the steps and make a billing profile. The details would already be pre-filled as you had entered card details before. Click on 'Submit and enable billing'.



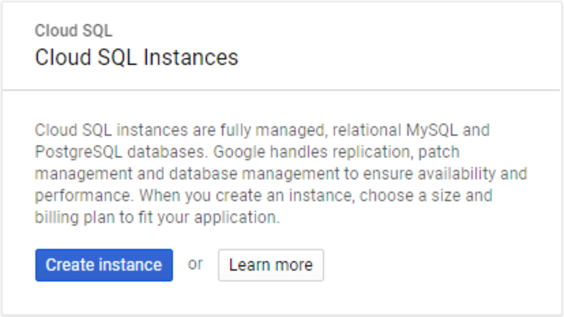
New Project

Project Name *
csci585-hw2

Project ID: csci585-hw2-206304. It cannot be changed later. EDIT

Location *
No organization
Parent organization or folder

Figure 2 New Project Screen



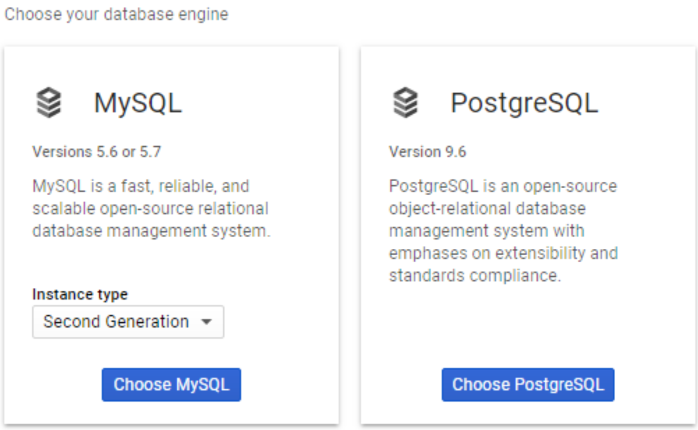
Cloud SQL
Cloud SQL Instances

Cloud SQL instances are fully managed, relational MySQL and PostgreSQL databases. Google handles replication, patch management and database management to ensure availability and performance. When you create an instance, choose a size and billing plan to fit your application.

Figure 3 Create Instance

Create a Cloud SQL Instance

1. Click on “Create Instance” button (Figure 3).
2. Select either MySQL or PostgreSQL (Figure 4). For example, I select MySQL. On the next screen, I select MySQL Development (Figure 5).



Choose your database engine

MySQL

Versions 5.6 or 5.7

MySQL is a fast, reliable, and scalable open-source relational database management system.

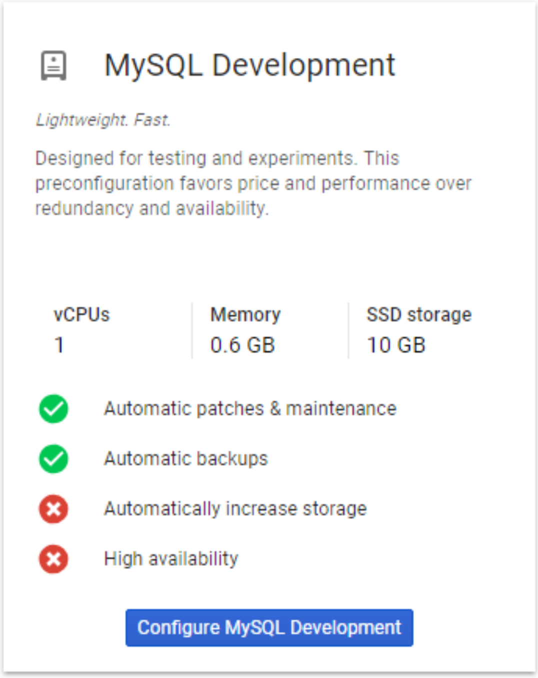
Instance type: Second Generation

PostgreSQL

Version 9.6

PostgreSQL is an open-source object-relational database management system with emphases on extensibility and standards compliance.

Figure 4 MySQL and PostgreSQL.



MySQL Development

Lightweight. Fast.

Designed for testing and experiments. This preconfiguration favors price and performance over redundancy and availability.

vCPUs	Memory	SSD storage
1	0.6 GB	10 GB

✓ Automatic patches & maintenance

✓ Automatic backups

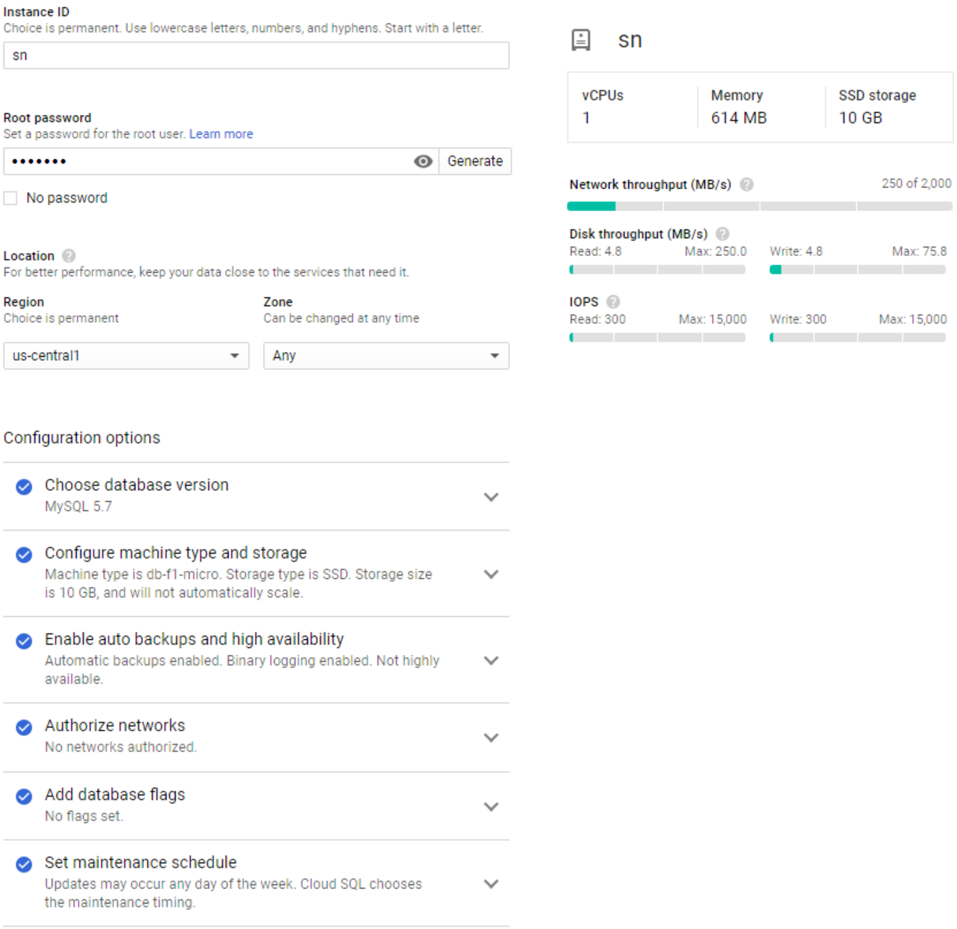
✗ Automatically increase storage

✗ High availability

Figure 5MySQL Development

The next steps are explained with MySQL.

3. Provide an Instance ID and a root password. Leave the rest as they are:



Instance ID
Choice is permanent. Use lowercase letters, numbers, and hyphens. Start with a letter.
sn

Root password
Set a password for the root user. Learn more
••••••• Generate
□ No password

Location
For better performance, keep your data close to the services that need it.

Region
Choice is permanent
us-central1

Zone
Can be changed at any time
Any

Configuration options

- Choose database version
  MySQL 5.7
- Configure machine type and storage
  Machine type is db-f1-micro. Storage type is SSD. Storage size is 10 GB, and will not automatically scale.
- Enable auto backups and high availability
  Automatic backups enabled. Binary logging enabled. Not highly available.
- Authorize networks
  No networks authorized.
- Add database flags
  No flags set.
- Set maintenance schedule
  Updates may occur any day of the week. Cloud SQL chooses the maintenance timing.

sn

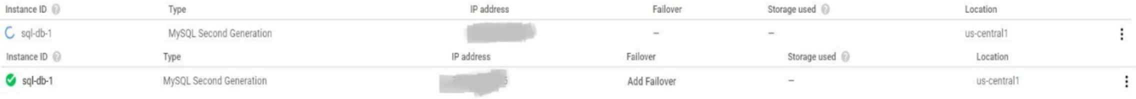
vCPUs	Memory	SSD storage
1	614 MB	10 GB

Network throughput (MB/s) 250 of 2,000

Disk throughput (MB/s)
Read: 4.8 Max: 250.0 Write: 4.8 Max: 75.8

IOPS
Read: 300 Max: 15,000 Write: 300 Max: 15,000

Click on “Create”. You will see “Instance is being created”. Wait until the left most wheel turns into a green tick. (It may take up to 15 minutes or more. Be patient!)

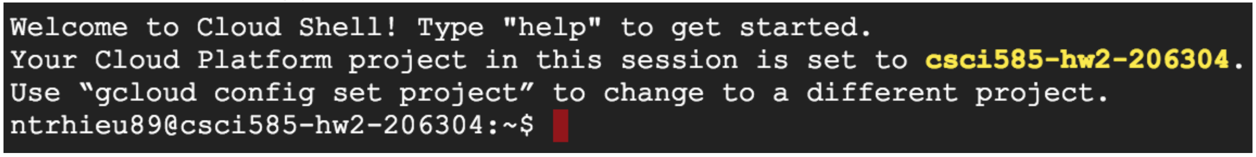


Instance ID	Type	IP address	Failover	Storage used	Location	
sql-db-1	MySQL Second Generation	[redacted]	–	–	us-central1	⋮

Instance ID	Type	IP address	Failover	Storage used	Location	
✓ sql-db-1	MySQL Second Generation	[redacted]	Add Failover	–	us-central1	⋮

Note: On the right-hand side, the three-dot menu has a “Delete” option. Be sure to delete this instance once you are done with the homework to avoid extra charges on the instance.

4. Click on the instance ID name to open the “Instance details” page, and then click on “Connect using Cloud Shell.” At the Cloud Shell prompt, connect to your Cloud SQL instance. When the Cloud shell finishes initializing you should see:



```
Welcome to Cloud Shell! Type "help" to get started.
Your Cloud Platform project in this session is set to csci585-hw2-206304.
Use "gcloud config set project" to change to a different project.
ntrhieu89@csci585-hw2-206304:~$
```

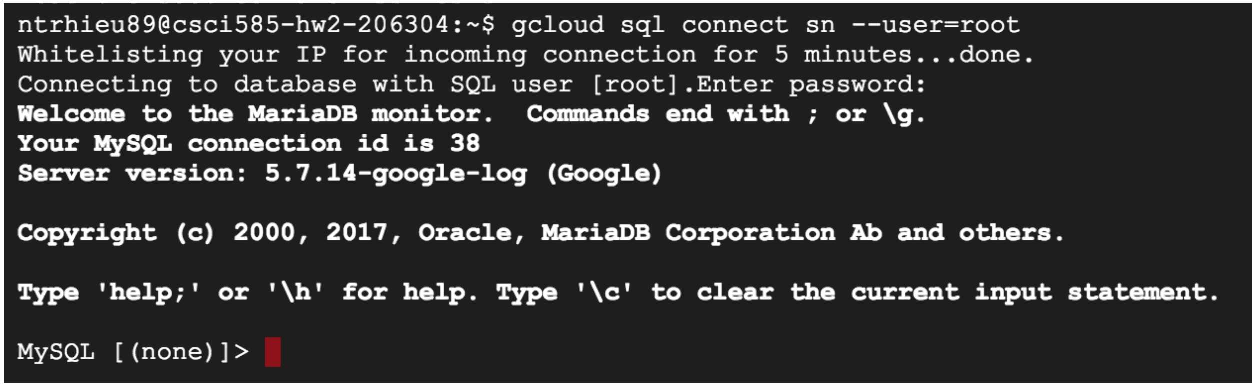
db-hw-sql would be replaced with the name of your project.

5. At the Cloud Shell prompt, connect to your cloud SQL instance.

```
gcloud sql connect myinstance --user=root
```

Replace myinstance with the name of your instance, (in this example, sn.)

Enter your password (it is a linux terminal so you won’t see it being typed). You should now be able to see the mysql prompt.



```
ntrhieu89@csci585-hw2-206304:~$ gcloud sql connect sn --user=root
Whitelisting your IP for incoming connection for 5 minutes...done.
Connecting to database with SQL user [root].Enter password:
Welcome to the MariaDB monitor.  Commands end with ; or \g.
Your MySQL connection id is 38
Server version: 5.7.14-google-log (Google)

Copyright (c) 2000, 2017, Oracle, MariaDB Corporation Ab and others.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

MySQL [(none)]> 
```

Congratulations! You just finished the second part of the assignment.

Part 3: Working with SQL (Optional)

In this part of the assignment, we will build a database with one table and run queries to see if MySQL works as expected.

1. Create a SQL database on your Cloud SQL instance:

```
CREATE DATABASE test;
```

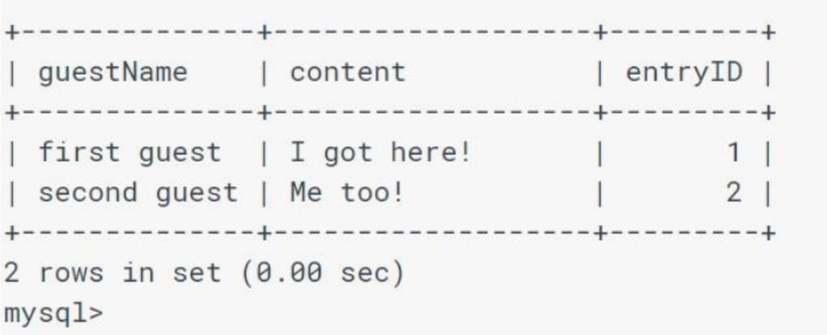
2. Insert sample data into the guestbook database:

```
USE test;
CREATE TABLE entries (guestName VARCHAR(255), content
VARCHAR(255), entryID int not null AUTO_INCREMENT, PRIMARY
KEY(entryID));
INSERT intO entries (guestName, content) values ("first guest",
"I got here!");
INSERT intO entries (guestName, content) values ("second guest",
"Me too!");
```

3. Retrieve the data.

```
SELECT * FROM entries;
```

You should see:



```
+--------------+-------------------+---------+
| guestName    | content           | entryID |
+--------------+-------------------+---------+
| first guest  | I got here!       |       1 |
| second guest | Me too!           |       2 |
+--------------+-------------------+---------+
2 rows in set (0.00 sec)
mysql>
```

Congratulations! You are now ready to solve the assignment.

Part 4: Programming Assignment

A database for a social networking application consists of the following tables:

USERS (**USER_ID**, FNAME, LNAME, GENDER, DATE_OF_BIRTH)

FRIENDSHIPS (**INVITER_ID, INVITEE_ID**, STATUS)

POSTS (**POST_ID**, USER_ID, TEXT)

COMMENTS (**COMMENT_ID**, POST_ID, USER_ID, TEXT)

The primary key for each table is bolded.

INVITER_ID and INVITEE_ID attributes of the FRIENDSHIPS table are foreign keys referencing the USERS table. USER_ID attribute of the POSTS table is the foreign key referencing the USER_ID attribute of the USERS table. USER_ID attribute of the COMMENTS table is the foreign key referencing the USER_ID attribute of the USERS table. POST_ID attribute of the COMMENTS table is the foreign key referencing the POST_ID of the POSTS table.

Notes:

- All attributes of tables are NOT NULL.
- With FRIENDSHIPS table, INVITER_ID is the USER_ID of the user who invites the user with INVITEE_ID as friend.
- The STATUS attribute of the FRIENDSHIPS table has value 0 if inviter invited invitee to be friend and has not been accepted. Once accepted, the STATUS has value 1.
- FNAME and LNAME are text attributes, each with a maximum of 50 characters.
- The GENDER column of the USERS is represented as ‘F’ for female users and ‘M’ for male users.
- The format for the DATE_OF_BIRTH column is ‘YYYY-MM-DD’.
- We haven't provided any tables of data. You are responsible to make your own with the schema given above and do the query tests on them. We will have our own tables to test your queries.
- Assume the corresponding data for every query exists and that it must return some records.
- Don't use views. They are **NOT** allowed in this assignment.
- **Temporary** tables, **dummy** tables... are **not** allowed. All the questions should be answered in one query (with as many subqueries as you need) for that question.
- **For loops** are **not** allowed either.

Instructions:

- For each query provide the SQL query and an explanation of why the query works the way it does. Make any assumptions that are not conflicting. Only use the mentioned attributes (we will make queries only with the exact given names of tables and attributes. Also they should be all CAPITAL letters).

Please provide the SQL queries for the questions below (Each question from 1-4 is worth 1.5 point. Questions 5 and 6 each is worth 1 point):

1. List the USER_ID, FNAME, LNAME of friends of the user whose id is 5.
2. List the USER_ID, FNAME, LNAME, GENDER, DAY_OF_BIRTH of all pending friends (users that invited this user as friend but have not yet been accepted) of the user whose id is 1.
3. List the USER_ID, FNAME, LNAME of female mutual friends between users 1 and 2.
4. List the USER_ID of female users who were born after '1990-12-20' and commented on posts of USER_ID=10. Show their friends count in a separate column.
5. List the user ids of up to 10 pairs of users where their distances are exactly 2 (i.e., they do not have direct friendship and share at least one common friend).
6. List the user ids of up to 10 pairs of users where one is male and the other is female, and each comments on the other's posts at least 5 times.

Submission Guidelines:

- The submission MUST be a pdf file named [Student First Name] _[Student Last Name]_HW2.pdf. Different file format will not be graded.
- If you have any general questions about the homework, please post your questions on HW2 discussion on USC DEN course forum only.
- Please don't email the TAs directly with questions. Let other students in the class also benefit from answers and questions.
- Before asking, please check the forum to see whether similar questions were asked and answered.
- Note that your programs will be automatically checked against other codes. This will be in the amount of percentage similarities. Any percentage can risk your code credibility. So don't show/share your code with anyone.
- Write your codes in any script (MySQL, PostgreSQL) you feel more comfortable with. However, MySQL is preferred.