Important !!

This assignment will be evaluated by an **AUTOMATED** system. It is important that you follow the guidelines described in this section in preparing the answers. As a result, you either get complete points for a particular question or zero ! There will not be any partial points. Additionally if the guidelines are not followed and TAs have to manually edit your submissions to make it work with the evaluation tool, points will be deducted very generously !!

All assignment questions will have to work either on **DB2 or on POSTGRES** database systems. Do not submit both versions. Only **ONE** should be submitted.

Your queries should output **ONLY** those columns that is specified in the question in the requested format. Do not **RENAME** columns unless explicitly specified. If an **ORDERING** of rows is requested in the question, make sure that it is performed. All ordering is implied to be **ASCENDING** unless explicitly stated otherwise.

Make sure your queries are always terminated with a **SEMILCOLON** (**;**) . In order for semicolons to be used in DB2, you will have to start the client by typing.

```
$ db2 -t
```

You may write comments in the SQL file, as long as it is standard SQL comments (i.e. starts with --).

Example 1

Q You need to find all the employees who live in Laval and their department. The output rows of the query will consists first the column NAME from the employee table followed by ADDR_CITY from the department table. Order them based on ADDR_CITY and then on NAME.

Correct SQL 1

```
SELECT NAME, ADDR_CITY
....
ORDER BY 2, 1
;
```

Correct SQL 2

```
SELECT NAME, ADDR_CITY
....
```

```
ORDER BY ADDR_CITY, NAME
;
```

Correct SQL 3

```
SELECT E.NAME, E.ADDR_CITY
....
ORDER BY ADDR_CITY, NAME
;
```

In-Correct SQL 1

```
SELECT E.NAME EMPLOYEE_NAME, E.ADDR_CITY
....
ORDER BY ADDR_CITY, EMPLOYEE_NAME
;
```

Incorrect - because the employee name column is renamed.

In-Correct SQL 2

```
SELECT E.NAME, E.ADDR_CITY
....
ORDER BY NAME, ADDR_CITY
;
```

Incorrect - because the ordering should be first by ADDR_CITY and then NAME and not the other way around.

Example 2

Q You need to find the city in which the most number of employees live and the number of employees that live in that city. Your output should rows should contain the column ADDR_CITY followed by the number of employees who live in that city, name this second column EMPLOYEE_COUNT . To handle the situation where you have more than one city with most number of employees in it, make sure the output is ordered by ADDR_CITY.

Correct SQL 1

```
SELECT ADDR_CITY, COUNT(*) EMPLOYEE_COUNT
....
ORDER BY 1
;
```

In-Correct SQL 1

```
SELECT ADDR_CITY, COUNT(*)
....
ORDER BY 1
;
```

Incorrect : COUNT(*) should have been renamed EMPLOYEE_COUNT

Setup Instructions

If you are using **DB2** download the assignment package , download the package DB2-XXXXXXXXX.tar.gz

If you are using **POSTGRES** download the assignment package , download the package PG-XXXXXXXXX.tar.gz

These scripts are meant to be used in the comp421.cs.mcgill.ca . If you are using your own (say) installation of database or connecting from elsewhere to the database server, you will have to edit the connection settings accordingly.

untar the contents and **RENAME** the folder to replace XXXXXXXXX with your student id number. Say you student id is 342567987 and you are using postgres, this can be accomplished by executing the command

```
$ mv  XXXXXXXXX  PG-342567987
```

change directory

```
$ cd PG-342567987/data
```

and execute the script to create tables (alternatively you can copy the create table statements into your database GUI of choice and execute them)

```
$ ./setup.sh
```
Enter your socs unix password if requested.

There is a script `delNdrop`.sh that is supplied in case you want to drop the tables that was created as part of the assignment.

You can execute it by running the command.

```
$ ./delNdrop.sh
```

For each question, once you have verified your query is working, save the query into the corresponding sql file in the folder. For example, if you are using postgresql, and want to save the query for Q1, edit the file pgsql/q01.sql and put the contents inside of it. Make sure to end your query with a semicolon. Similarly, if you are using db2, save the query into the file db2/q01.sql , and end it with a semicolon. **DO NOT** include the “db2 connect to cs421” command in the query file. We will be already connected before executing the queries, so this should not be included.

You can test your query output again by executing from your shell prompt

in Postgres

```
$ psql cs421 <  q01.sql
```

or in DB2

```
$ db2 connect to cs421
$ db2 -t <  q01.sql
```

Q1

……...

What to Submit !

Please compress your entire folder and submit it in my courses. DO NOT submit the outputs . We don’t need the data folder, but it is ok, if you cannot remove it. Following are some examples of how you can compress what we need. Make sure the parent directory structure with your student id in it is preserved. So if you chose postgres, following the above examples,

you will submit a compressed version of `PG-342567987` or if you choose db2, you will submit a compressed version of `DB2-342567987` .

Correct

```
$ tar -zcvf PG-342567987.tar.gz PG-342567987/pgsql/
```

Submit the file `PG-342567987.tar.gz to mycourses.`

Correct (db2 example)

```
$ tar -zcvf PG-342567987.tar.gz DB2-342567987/db2/
```

Submit the file `DB2-342567987.tar.gz to mycourses.`

Correct

```
$ tar -zcvf PG-342567987.tar.gz PG-342567987/
```

This includes the data folder and other things that we do not need, but is still ok as it preserves the parent folder as well as includes the contents of the sql folder which contains your answer queries.

Incorrect

```
$ cd PG-342567987
$ tar -zcvf PG-342567987.tar.gz  pgsql/
```

incorrect because now you lost the parent directory information which is `PG-342567987` that is needed for us to know your student id.