CS 340 Programming Assignment 1 - SQLPLUS

Due: 11:59 pm, October 12, 2017

A Simple Oracle Database Design

You are going to use Oracle to design a simple database. You will create tables, views, and implement some queries.

Connecting to Oracle Database

- Open the **.bashrc** file in any text editor (e.g., Emacs).
- Write the following statements at the end of the **.bashrc** file and close it after saving.

```
export PATH=/usr/lib/oracle/11.2/client64/bin:$PATH
export LD_LIBRARY_PATH=/usr/lib/oracle/11.2/client64/lib:$LD_LIBRARY_PATH
```

- Type **source .bashrc** on the Shell prompt to update the PATH and LD_LIBRARY_PATH variables.
- Type echo $PATH on the Shell prompt to print the PATH variable. It should include the path /usr/lib/oracle/11.2/client64/bin
- Type echo $LD_LIBRARY_PATH on the Shell prompt to print the LD_LIBRARY_PATH variable. It should include the path /usr/lib/oracle/11.2/client64/lib
- To connect to the database using SQLPLUS client, use the following statement on the Shell.

sqlplus <db_username/password>@localhost:1521/cs340f17.lums.edu.pk

Your db_user name is the same as your login to the server (203.135.63.67). Your Oracle DB password is initially set to abcd1234. A student with Registration number 15100281 will use the following statement to connect to the database

sqlplus s15100281/abcd1234@localhost:1521/ cs340f17.lums.edu.pk

Once, you are connected to the database you can create tables, insert record in the tables and query the database.

Currently your password to the cs340.lums.edu.pk database is set to abcd1234. To change the password, use the following command on SQL prompt:

ALTER USER db_username IDENTIFIED BY new_password;

Please make sure that you remember your password.

To disconnect from the database, type quit at the SQL prompt.

- Use a text editor you are familiar with to create ".sql" files that contain the necessary SQL commands for creating tables, inserting records or querying the database.
- To execute an sql command in the <filename>.sql file, type @<filename> on the SQL prompt. For example, if your .sql file is saved as queries.sql, then type "@" followed by queries without the .sql extension. It looks like:
  SQL> @queries

Schema for Programming Assignment 1

For this assignment, we will use the modified schema of the U.S. Patent database managed by the National Bureau of Economic Research (http://www.nber.org/patents/). The Patent Database schema used for this assignment is shown in Figure 1.

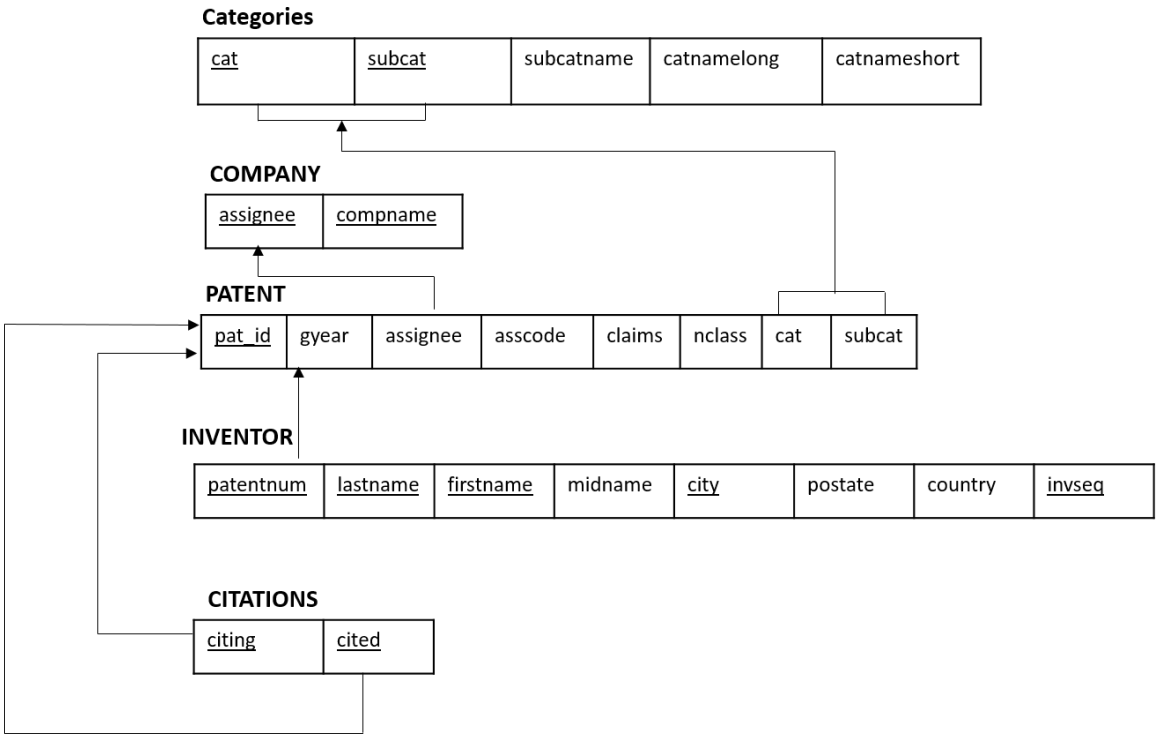

Figure 1. Patent Database schema

Attribute Names

- The categories Table list all the categories and sub-categories (catnamelong – category Name Long; catnameshort – category Name Short, subcatname – sub-category Name)
- The Company table list the identifier (assignee) of the company that has been assigned the patent and the Company Full Name (compname)
- The Patent table includes the following attributes
  - pat_id – patent identifier

- gyear – year in which the patent was granted
- assignee –identifier of the company that was assigned the patent
- asscode – Assignee Type (e.g, Government, private organization, etc.)
- claims – number of claims made in the patent
- nclass – Main Patent Class (3 digit)

- The Inventor table includes the following attributes
  - patnum – patent identifier
  - lastname – last name of the inventor
  - firstname – first name of the inventor
  - midname – Middle name of the inventor, if any
  - city – city of the inventor
  - postate – US State of the inventor, if residing in US
  - country – country of the inventor
  - invseq – inventor sequence number (1^{st} inventor, 2^{nd} inventor, 3^{rd} inventor, etc.)

- The Citations table includes the following attributes
  - citing – identifier of the patent that cites other patents in its list of references
  - cited – identifier of a patent that is included in the reference list of other patents

Tables

Create the following tables described below. Name these tables CATEGORIES, COMPANY, PATENT, INVENTOR, CITATIONS.

- CATEGORIES (**cat** number, **subcat** number, subcatname varchar2(50), catnameshort varchar2(20), catnamelong varchar2(100))

- COMPANY (**assignee** number, compname varchar2(50))

- PATENT (**pat_id** number, gyear number, assignee number, asscode number, claims number, nclass number, cat number, subcat number

- INVENTOR (**patentnum** number, **lastname** varchar2(20), **firstname** varchar2(20), midname varchar2(10), city varchar2(25), postate varchar2(10), country varchar2(20), **invseq** number)

- CITATIONS (**citing** number, **cited** number)

The underlined fields are the primary keys of their respective tables

Queries

Implement the following queries:

1. Find Patent Ids of all patents listed in the 'Chemical' category or 'Drugs AND Medical' category.
2. Find the last name, first name, country, and state of inventors who have published patents either in the 'Chemical' category or 'Drugs AND Medical' category.
3. Find the last name, first name, country, and state of inventors who have published patents in the 'Chemical' category only (i.e., these inventors have not published patents in any other category).
4. Find Patent Ids of all patents that have any of their inventors residing in the US state of California ('CA') or New Jersey ('NJ').
5. Find Patent Ids of all patents that have both their first AND second inventors residing in the US state of California ('CA') or New Jersey ('NJ').
6. Find the name of the company that has been assigned the highest number of patents.
7. Find the name of the company that has been assigned the highest number of patents in the Chemical Category.
8. List the name of companies that have been assigned at least 3 patents in the Chemical Category.
9. For each Subcategory in the Chemical Category, find the name of the company that has been assigned the highest number of patents in that Subcategory.
10. Find the first name and last name of the inventor who has published the highest number of patents in the 'Electrical AND Electronic' category. Also, print the number of patents published by that inventor.
11. For each category, find the first name and last name of the inventor who has published the highest number of patents in that category. Also, print the number of patents published by that inventor.
12. Find the names of companies who have been assigned a patent in every sub category of 'Electrical AND Electronic'.
13. Find the names of all those inventors who have published patents in at least two different sub-categories of 'Chemical' Category.
14. Find the 'Patent Id' of the patent that has the highest number of citations (the citations of a given patent is the number of times the given patent has been cited by other patents).
15. For each category, find the 'Patent Id' of the patent that has the highest number of citations in that category (the citations of a given patent is the number of times the given patent has been cited by other patents).
16. Find the 'Patent Id' of the patent that has cited the highest number of other patents.
17. Find the first name, last name, inventor sequence (invseq), city, and state of the inventor who have received the highest number of citations (the citations of an inventor is the sum of the citations of all patents published by the given inventor).
18. Find the first name, last name, city, and state of the inventor who has published the highest number of patents as First Inventor.
19. For each category, find the first name, last name, city, and state of the inventor who has published the highest number of patents in that category as First Inventor.

20. List the Patent identifiers of all the patents in the 'Chemical' Category that have not been cited by any other patent (i.e., these patents have ZERO citations).
21. For each (category, sub category), find the total number of patents published by inventors residing in the US state of 'California' (i.e., postate of any inventor = 'CA').
22. Find the average number of patents assigned to companies with all of their inventors residing in the New Jersey State (i.e., postate of first inventor = 'NJ').
23. Print the name of the company (companies) whose number of patents is greater than the average number of patents of companies with all of their inventors residing in the New York State (i.e., postate = 'NY').
24. Compute the average number of inventors of patents in the 'Electrical AND Electronic' category.
25. List the first name and last name of all the inventors who have published patents without self-citations (inventor of a patent did not cite his/her prior patent).

Create the following 2 views. Please name them VIEWA and VIEWB.

1. A view that shows the Patent ID, First Name, Last Name, Grant Year (gyear), Company Name, Category Name (long), Sub-category Name (long) of the patents of First Inventor. The inventor sequence (invseq) of the First Inventor is equal to '1' in the INVENTOR Table.

2. A view that shows the Assignee ID, Company Name, Category Name (long), Sub-category Name (long), Number of Patents assigned to the company.

A useful strategy
Here is a useful approach for doing the Programming Assignment.

1. Connect to the Oracle server with your assigned Oracle account.
2. Try a few simple SQL commands until you are comfortable interacting with sqlplus.
3. Work out the SQL commands you need to solve the PATENT database problems.
4. Use a text editor you are familiar with to create ".sql" files that contain the necessary SQL commands for Programming Assignment 1.
5. Test your .sql files. For example, if your .sql file is saved as *queries*.sql, then type "@" followed by *queries* without the .sql extension. It looks like:
   SQL> @queries
6. Please add the following lines to the beginning of your .sql files so that grades can be assigned:
   rem CS 340 Programming Assignment 1
   rem your_first_name Your_last_name
   rem Your login
7. Remember to divide and conquer. Test your .sql file continuously as you add new SQL commands. You can use the Oracle command **spool** to direct the output to a file so that you can check it. However, please **DO NOT** submit the output file.

What to submit

You are going to submit **FOUR** SQL files: ***create.sql, populate.sql, queries.sql***, and ***dropall.sql***.

- create.sql should contain SQL commands to create the tables and defining integrity constraints.
- populate.sql should contain SQL commands (insert into . . .) to populate the tables with your own data.
- queries.sql should contain the SQL commands for the above queries and the SQL commands for displaying the views. (Please use the command DBMS_OUTPUT.PUT_LINE command to display query numbers. E.g., `DBMS_OUTPUT.PUT_LINE('Query #1').` This will help the grader in evaluating the output of your queries.
- dropall.sql should contain the drop table and drop view commands to drop all the tables and views.

How to submit your .sql files

When you are ready to submit your Programming Assignment, put all the files to be submitted in a folder/directory "PA1_[your login]." For example, if your login is "s1510028" put the file in the folder/directory "PA1_s1510028". Compress the folder into a zip file, and submit it on LMS.

Sample data

Sample data is provided on LMS to show the format of data for the corresponding tables. Your queries may not necessarily be evaluated against the sample data.