<div align="center">

CptS -451 Introduction to Database Systems
Spring 2021

Project Milestone-1

Due Date: Wednesday February 17th, 11:59pm

</div>

Summary:

In this milestone you will parse the Yelp JSON data and develop a simple database application. The goal of this exercise is to get you started in database programming early on. In Milestone3 you will develop a larger application with all required features.

Milestone Description:

Task 1:

Download the Yelp dataset from https://eecs.wsu.edu/~arslanay/CptS451/project/yelp_dataset/yelp_CptS451_2020.zip. Look at each JSON file and understand what information the JSON objects provide. Pay attention to the data items in JSON objects that you will need for your application.

Download the sample JSON Parser program (Python) from Canvas (*TermProject page*). This programs provides example code for:

- o reading JSON objects form a file and extracting certain key and value pairs from JSON objects,
- o writing extracted data into a text file.

Please note that the sample code includes examples of extracting simple key values only. In a JSON object the key value can be an array or another JSON object (for example: hours,attributes), therefore you need to recursively parse those nested objects until you extract all data stored in JSON objects. You will write the code for parsing business, tips, user, and checkin JSON objects.

- ▪ In `yelp_business.json` : Parse all keys. Flatten the nested attrite and hours values.
- ▪ In `yelp_user.json` : Parse all keys.
- ▪ In `yelp_tip.json`: Parse all keys.
- ▪ In `yelp_checkin.json` : Parse all keys.

About Check-in Data: Each check-in object has a "date" key whose value includes the timestamps of the check-ins to the corresponding business. All check-in timestamps are included in a long string, separated by commas. You need the split this string and extract the check-in timestamps for the business.

Task 2:

i. Design a database schema that models the database for the described application scenario in Appendix-A and provide the ER diagram for your database design. Your schema should be precise but complete. It should be designed in such a way that all queries/data retrievals on/from the database run efficiently and effectively. In Milestone2 you will revise your ER model.

ii. Translate your ER model into relations and produce DDL SQL statements for creating the corresponding tables in a relational DBMS. Note the constraints, including key constraints, referential integrity constraints, not NULL constraints, etc. needed for the relational schema to capture and enforce the semantics of your ER design.

Task 3:

i. Download and unzip the "milestone1db.zip" file from Canvas (on the "Term Project" page).

ii. Create a database on PostgreSQL with name "`milestone1db`" and create a table named "`business`". The schema of the *business* table should comply with the columns of the CSV file, i.e., there should be an attribute for each column of the CSV file. Please define the type and domain of each attribute based on the possible values that appear in the corresponding column.

The "`milestone1db.csv`" file includes 4 columns: *business_id* (unique id of the business), *name* (name of the business), *state*, and *city*.

iii. Import the CSV file into this table by executing the following statement in `pgadmin`. Please replace <path> with the directory where CSV file is located.

```
copy business (business_id,name,state,city) FROM
'<path>/milestone1db.csv' DELIMITER ',' CSV
```

Note-1: *In the above command, replace* `<path>` *with the full path of the directory where* `milestone1db.csv` *file is located. On Windows machines, if the file is located under a directory created by Windows OS (such as "Downloads", "Documents", "One Drive", etc.) you may get "permission denied" errors. A possible solution is to create a directory under* `C:\` *drive and copy the* `milestone1db.csv` *file to that directory. For example, assume the file is under* `'C:\temp'`*, then you should run the command:*
```
copy business (business_id,name,state,city) FROM 'C:\temp\milestone1db.csv'
DELIMITER ',' CSV
```

Note-2: *If you run the above command on the PostgreSQL command-line client (psql), include a backslash (i.e., '\') in the beginning, i.e.,*
```
\copy business (business_id,name,state,city) FROM '<path>/ milestone1db.csv'
DELIMITER ',' CSV
```

iv. Build a simple application (either web or standalone) which connects to the `milestone1db` database and runs simple queries on the `business` table. A sample screenshot for your milestone1 application is shown below. The application will:
  - o list the states that appear in business table and allow user to select a state;
  - o when a state is selected, the cities in that state will be listed;
  - o when a city is selected, the list of the businesses in that state and city will be listed;
  - o when a business is selected, the business details are displayed on a new window.
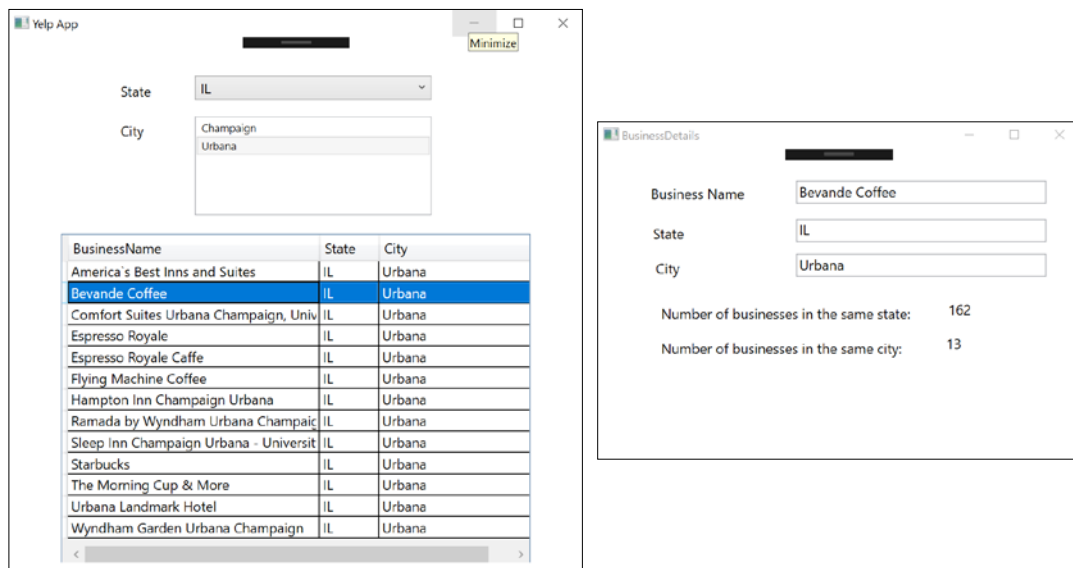
A video tutorial on how to establish connectivity with the PostgreSQL in C# using Npgsql is available on Canvas.

You need to run the following queries on the *business* table:
```
SELECT DISTINCT state
FROM business
ORDER BY state

SELECT DISTINCT city
FROM business
WHERE state= selected_state
ORDER BY city

SELECT name
FROM business
WHERE city= selected_city AND state= selected_state
ORDER BY name

SELECT name, state, city
FROM business
WHERE business_id = id_of_the_selected_business

SELECT COUNT(*)
FROM business
WHERE state = (SELECT state FROM business
               WHERE business_id = id_of_the_selected_business>)

SELECT COUNT(*)
FROM business
WHERE city = (SELECT city FROM business
              WHERE business_id = id_of_the_selected_business >)
```

Milestone-1 Deliverables:

1. (25%) Source code for parsing all JSON data. Only submit your source code, **not the data files**.
2. (40%) The E-R diagram and relations (CREATE TABLE statements) for your database design. To create your ER diagram, I suggest you to use draw.io tool (https://www.draw.io/). You may also use your favorite drawing tool (e.g., Visio, Word, PowerPoint). Should be submitted in .pdf format. Name the diagram "*<your-team-name>_ER_v1.pdf*" and the SQL statements "*<your-team-name>_schema.sql*".
3. (35%) Source code for your application. Only submit your source code, **not the data files**.

Create a zip archive *"<your-team-name>_milestone1.zip"* that includes your source code for JSON parsing and your sample application. Upload your milestone-1 submission on Canvas until the deadline.

References:
1. Yelp Dataset Challenge, http://www.yelp.com/dataset_challenge/
2. Samples for users of the Yelp Academic Database, https://github.com/Yelp/dataset-examples
3. Yelp Challenge, University of Washington Student Paper 1
   http://courses.cs.washington.edu/courses/cse544/13sp/final-projects/p08-fants.pdf
4. Yelp Challenge, University of Washington Student Paper 2,
   http://courses.cs.washington.edu/courses/cse544/13sp/final-projects/p10-michelmj.pdf