# CptS -451 Introduction to Database Systems
# Everett
# Spring 2020

# Project Milestone-1

This project is for the CptS451 students majoring in Data Analytics

Due Date: Thursday February 13th, 11:59pm

## 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:

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 Blackboard (*Project\Sample JSON Parsing Code*). This programs provides example code for:

- reading JSON objects form a file and extracting certain key and value pairs from JSON objects,
- 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 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 **except open/close times**.
- In `yelp_user.json` : Parse all keys.
- In `yelp_tip.json`: Parse all keys.
- In `yelp_checkin.json` : Parse all keys.

*Parsing 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. You need the split this string and extract the check-in timestamps for the business.

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 yet 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.

3) (i) Download the “milestone1DB.csv” file from the link
[http://www.eecs.wsu.edu/~arslanay/CptS451/project/yelp_dataset/milestone1DB.csv](http://www.eecs.wsu.edu/~arslanay/CptS451/project/yelp_dataset/milestone1DB.csv)

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.

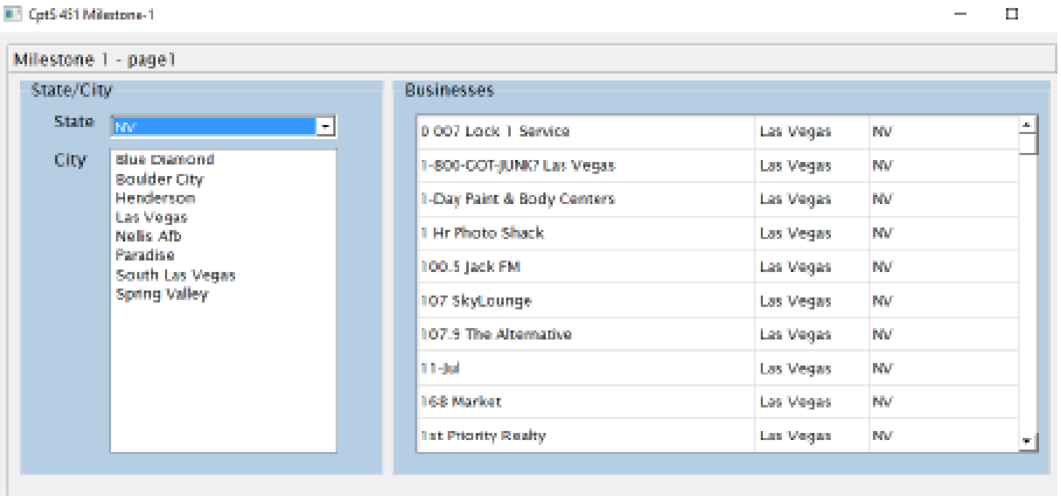
The “`milestone1DB.csv`” file includes 3 columns: *name* (name of the business), *state*, and *city*.

Import the CSV file into this table by executing the following statement in the PostgreSQL command line. Please replace <path> with the directory path for the `milestone1DB.csv` file.

```
\copy business (name,state,city) FROM '<path>/ milestone1DB.csv'
DELIMITER ',' CSV
```

(ii) Write 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:

- list the states that appear in business table and allow user to select a state;
- when a state is selected, the zipcodes in that state will be listed;
- when a zipcode is selected the list of the businesses will be listed.



State NV

City
Blue Diamond
Boulder City
Henderson
Las Vegas
Nellis Afb
Paradise
South Las Vegas
Spring Valley

| | | |
|---|---|---|
| 0 007 Lock 1 Service | Las Vegas | NV |
| 1-800-GOT-JUNK? Las Vegas | Las Vegas | NV |
| 1-Day Paint & Body Centers | Las Vegas | NV |
| 1 Hr Photo Shack | Las Vegas | NV |
| 100.5 Jack FM | Las Vegas | NV |
| 107 SkyLounge | Las Vegas | NV |
| 107.9 The Alternative | Las Vegas | NV |
| 11-Jul | Las Vegas | NV |
| 168 Market | Las Vegas | NV |
| 1st Priority Realty | Las Vegas | NV |

A video tutorial on how to establish connectivity with the PostgreSQL in Python using `psycopg2` will be available on Blackboard.

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;
```

*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 for your database design. To create your ER diagram, I suggest you to use Edraw Max (https://www.edrawsoft.com/download-edrawmax.php) . You may also use your favorite drawing tool (e.g., Visio, Word, PowerPoint). Should be submitted in .pdf format. Name this file "*<your-team-name>_ER_v1.pdf*"
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 Blackboard until the deadline.

**You will demonstrate your Milestone1 to the instructor and the TA.**

**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