Ira A. Fulton Schools of Engineering
Arizona State University

CSE 511: Data Processing at Scale

Project 1: NoSQL

Purpose

You must carry out transformations and actions on a No-SQL database using python files for this project. Here, data is not saved in relations like it is in standard relational databases.

Objectives

Learners will be able to:

- Recognize fundamental concepts and features of NoSQL databases.
- Demonstrate retrieving data from NoSQL databases.
- Explain how to convert mathematical concepts into code.

Technology Requirements

- Python 3.5.x version
- cython
- unqlite

Project Description

Through this project, you will gain experience working with NoSQL databases and understand how to use Python to conduct several fundamental operations on them. Unlike conventional relational databases, NoSQL allows data to be stored more freely without the use of rigid schemas. Additionally, they are capable of handling massive amounts of data quickly.

Directions

Implement the functions provided in the Jupyter Notebook to perform the operations as listed below. You may use the included sample.db to test your functions with the Python code provided or you may use the Jupyter Notebook environment, which has the sample database embedded within.

A. **FindBusinessBasedOnCity(cityToSearch, saveLocation1, collection)** – This function searches the 'collection' given to find all the business present in the city provided in 'cityToSearch' and save it to 'saveLocation1'. For each business you found, you should store the name, full address, city, and state of the business in the following format. Each line of the saved file will contain: Name$FullAddress$City$State. ($ is the separator and must be present.)

B. **FindBusinessBasedOnLocation(categoriesToSearch, myLocation, maxDistance, saveLocation2, collection)** – This function searches the 'collection' given to find the name of all the businesses present in the 'maxDistance' from the given 'myLocation' (please use the distance algorithm attached in the Coursera project description titled "...sample") and save them to 'saveLocation2'. Each line of the output file will contain the name of the business only.

Distance Algorithm:

A distance algorithm will need to be used. Given two pairs of latitude and longitude as [lat2, lon2] and [lat1, lon1], you can calculate the distance between them using the formula given below:

DistanceFunction(lat2, lon2, lat1, lon1):

- var R = 3959; // miles
- var φ1 = lat1.toRadians();
- var φ2 = lat2.toRadians();
- var Δφ = (lat2-lat1).toRadians();
- var Δλ = (lon2-lon1).toRadians();
- var a = Math.sin(Δφ/2) * Math.sin(Δφ/2) + Math.cos(φ1) * Math.cos(φ2) * Math.sin(Δλ/2) * Math.sin(Δλ/2);
- var c = 2 * Math.atan2(Math.sqrt(a), Math.sqrt(1-a));
- var d = R * c;
- d is the distance between the given pair of latitude and longitude. The distance is in **miles.**

Reference: http://www.movable-type.co.uk/scripts/latlong.html

Test Cases:

Test cases are provided in the file attached in the Coursera project description titled "...Additional Test Cases".

Submission Directions for Project Deliverables

Programming Assignment:

Submit a **.ipynb file**.

1. Go to "**Programming Assignment: Project 1: NoSQL**".
2. Click on the "**My submissions**" tab (located at the top of the page under the deadline date).
3. Click on "**Create submission**."
4. Upload one file for the assignment and click "**Submit**."
    a. If there is an error in your assignment, you may make corrections and resubmit.

Report Submission:

For this project deliverable, you must write a **2-3 page report detailing your work on the project**. Your report should include the following elements:

1. **Reflection**: How did you approach the project? What did you specifically do? (Write two function codes along with distance function and explain about them)
2. **Lessons Learned**: What did you learn by doing this project?
3. **Output**: Screenshot shows parts of your TXT output files.
4. **Result**: Result of your code by passing the test cases.

This is a manually graded task and you will get credit for the report if you covered **all** the mentioned parts.

Your report should be in **doc, docx, or pdf**. Please title your report file with **"Last Name_First Name_Fall B 2022_Project 1 NoSQL Report**."

When you are ready to submit the report:

Double check that the content of your report is complete, your file is in **doc, docx, or pdf** and titled with **"Last Name_First Name_Fall B 2022_Project 1 NoSQL Report**." Then submit your report:

1. Go to "**Graded Assignment: Project 1: NoSQL Report Submission**".
2. Click "**Start Submission**".
3. Click "**Upload a File**."
4. Locate and select your report file.
5. Run your file through **Plagiarism Detection**.
6. Click "**Submit**."

Evaluation

There is one test for a total of one point. If some part of your data is incorrect, you will get partial scores of 0.25, 0.50, or 0.75. **If the submission fails, you will see the corresponding error logs that indicate *where* the error occurred.**

Common Errors:

1. Error: Submission was not a well-formed Jupyter Notebook file.
2. Save the file as IPYNB directly from Jupyter Notebook rather than saving it in another format and then replacing their properties as it makes the file unstable.
3. Runtime errors happen in your submission invalid syntax (<string>, line 22)
4. Error: Distance function, FindBusinessBasedOnCity, FindBusinessBasedOnLocation
5. Your submission did not define the FindBusinessBasedOnLocation function, but passed all tests for FindBusinessBasedOnCity.
6. Runtime errors happen in your submission name 'data' is not defined.
7. The original function needs to remain untouched and students need to write code inside the function definition block.

8. When you modify your code in the function and run it, make sure to delete "output_city.txt" and "output_loc.txt" if generated from the file directory of the notebook to receive correct output for the testcases.

Additional Common Errors:

A. Malformed feedback error

Solutions:

- Grader won't accept new functions to be defined, so the code should be within the already mentioned functions. use normal libraries in 3.5.
- You need to import any library in the cell with function code.
- Delete all the self-defined 'print' functions.
- Also please make sure that all your code is in the predefined cell.

B. The following error message: *Your submission did not define the FindBusinessBasedOnLocation function, but passed all tests for FindBusinessBasedOnCity.*

Solution:

- Think like an auto-grader. Only the graded cell is examined and the other cells are ignored.

Grading Rubric for Report

Rubrics communicate specific criteria for evaluation. Prior to starting any graded coursework, learners are expected to read through the rubric, so they know how they will be assessed. You are encouraged to self-assess your responses and make informed revisions before submitting your final report. Engaging in this learning practice will support you in developing your best work. Points may be deducted at the discretion of the faculty for disorganized submissions that convolute the grading process.

Component	0	1
Reflection	There is nothing about reflection	Completely talks about the 3 functions

Lessons Learned	There is nothing about Lesson Learned	Completely talked about what they learned from this project
Output	There is nothing about output	Shows part of two outputs
Result	There is nothing about the result	Show/talk about the output

Learner Checklist

Prior to submitting, read through the Learner Checklist to ensure you are ready to submit your best work.

- ☐ Did you title your file correctly and convert it into a single **.ipynb file**?
- ☐ Did you include your legal first and last name in the designated area on the **report**?
- ☐ Did you title your **report document** correctly and convert it into a single **doc, docx, or pdf**?
    - **Last Name_First Name_Semester Year_Name of Project Name**
- ☐ Did you answer all of the questions to the best of your ability?
- ☐ Did you make sure your answers directly address the prompt(s) in an organized manner that is easy to follow?
- ☐ Did you self-assess your open-ended responses using the rubric and make any necessary revisions?
- ☐ Did you proofread your work?