

## Homework-4: NoSQL on the Cloud

Deadline: October 12<sup>th</sup>, 11:59PM ET/8:59pm PT.

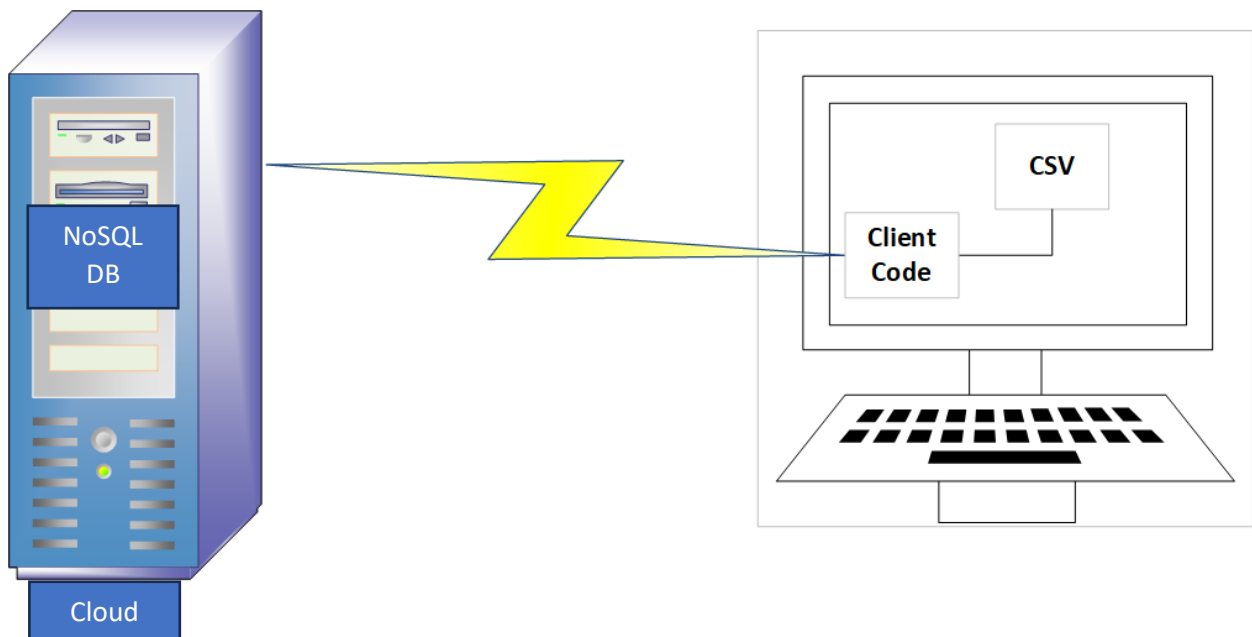
In this homework, you are expected to practice NoSQL implementation on the Cloud. It's expected you will apply the practices and understandings provided in the lecture materials.

Use the following GitHub classroom to access the assignment and create your assignment repository: [https://classroom.github.com/a/D\\_g83xyY](https://classroom.github.com/a/D_g83xyY)

**You should submit the URL for your GitHub repository on Canvas. Grading penalty will be applied if otherwise. Students are responsible for making sure they have uploaded all materials to GitHub on-time. For 3 days, a standard fixed 20 points penalty is applied for delays beyond the Due date.**

### General Description:

In this homework, you will create a NoSQL database table and populate it with data via a script that is run on your local machine.



## Problem Description

You have deployed three models of sensors to capture different weather related information.

- Your oldest model (oldnist-sensor) captures the humidity percentage in the air.
- Your mid-age model (mednist-sensor) captures humidity and wind information.
- Your newer model (modnist-sensor) captures humidity, wind and temperature.

You would like to store their readings in a proper database and you have identified that the best solution for this problem is to use a **columnar store**.

To save your data, you must optimize your data model to ensure efficient operation.

In this assignment, you are tasked to develop an application to upload your data to one NoSQL database in a single application run. Creating the table and populating the data in the table must occur via code.

To do so, your code is expected to conduct the following activities:

- Create a NoSQL table and all metadata required for it.
- Read your sensor data
- Upload the data to your NoSQL DB and Fill your NoSQL DB with experiment data
- Query the data at the end to ensure they were saved successfully

Notes:

- You can choose any programming language to complete this assignment. If you choose Python, Pandas library may help you with reading data from the CSVs.
- You may hardcode 1) column headers and 2) model names in your code if this will make the code easier. It is not expected that you should hardcode any other pieces of the data file in your code.

- The data file (sensor.csv) is uploaded with this homework. This file contains sample data but your solution should be scalable to handle as many records as possible as long as they belong to one of the sensor models identified earlier.
- Sensor\_ID is a unique column in the dataset.
- Hint: an efficient way to optimize your data storage starts with grouping and dividing your dataset into a set of columns.
- Hint: don't store any N/A values in the database table.

### Submission Guidelines:

- For this homework, submit the following:
  1. Screenshot of your NoSQL database instance on the Cloud. You may create your instance via GUI. (20%)
  2. Your Code file(s) (or Jupyter Notebook) (50%)
  3. Query all the data records in your database table and show the output of the query in your Jupyter Notebook or as an uploaded screenshot (30%).
    - Your query screenshot may look like this screenshot:

```
Scanning for all sensors:
Reading data for AAAAAA:
  Temperature_in_Fahrenheit: 45 F @2023-10-03 21:29:48.750000+00:00
  Humidity_Percentage: 50% @2023-10-03 21:29:48.750000+00:00
  Model_Used: modnist-sensor @2023-10-03 21:29:48.749000+00:00
  Notes: newer type of sensor @2023-10-03 21:29:48.750000+00:00
  Wind_Direction: NW @2023-10-03 21:29:48.750000+00:00
  Wind_Speed_in_mph: 4 mph @2023-10-03 21:29:48.750000+00:00

Reading data for AAAAAB:
  Humidity_Percentage: 40% @2023-10-03 21:29:48.750000+00:00
  Model_Used: mednist-sensor @2023-10-03 21:29:48.750000+00:00
  Notes: our mid-age quality sensor @2023-10-03 21:29:48.750000+00:00
  Wind_Direction: SW @2023-10-03 21:29:48.750000+00:00
  Wind_Speed_in_mph: 5 mph @2023-10-03 21:29:48.750000+00:00

Reading data for AAAAAC:
  Humidity_Percentage: 80% @2023-10-03 21:29:48.751000+00:00
  Model_Used: oldnist-sensor @2023-10-03 21:29:48.750000+00:00
  Notes: our oldest sensor @2023-10-03 21:29:48.751000+00:00
```

**Optional Useful Readings:**

- <https://www.geeksforgeeks.org/different-ways-to-iterate-over-rows-in-pandas-dataframe/>
- <https://cloud.google.com/python/docs/reference/bigtable/latest/column-family>
- <https://cloud.google.com/bigtable/docs/reading-data>