Project 4:   
Working with NoSQL Systems, in particular, MongoDB

DS503/CS585  
Big Data Management  
Team 6

Jiaxing (Jack) Zhang & Dennis Silva, Jr.

# Overview

This project is meant to deepen our understanding of NoSQL databases, in particular, MongoDB. As we are writing several, and at times, complex queries, we wanted to use a programming language instead of the MongoDB shell. This way, we can display our flow and though process for each question. We chose to use Python to write our MongoDB queries in (Prof. Elke said using Python was acceptable). We use the official Python Mongodb distribution PyMongo which is the recommended way to work with MongoDB from Python. One can see the PyMongo Documentation [here](https://api.mongodb.com/python/current/). The syntax is very similar to the MongoDB shell or Java implementation, except some of the ‘short cuts’ from the shell are not available.

# Relative Tasks Accomplished

Our groups pair (Dennis and Jack) first selected Python as our programming language, and then both completed the entirety of the project with a Jupyter notebook. Thus to be specific, both Dennis and Jack independently completed problems 1, 2, and 3.

# Team Methodology

When the project was first given our, Jack and Dennis exchanged contact information via email and WeChat. During the first week, we talked, in person before or after class, through the general methodology for solving some of the problems before actually coding. After the first week, Dennis went on a business trip for the entirety of the next (last) week. Dennis and Jack communicated via WeChat when issues came up with our solutions about every other day, and exchanged code snippets via email and/or our Google Drive we set up. A few days before the project was due, Jack sent Dennis his code and Dennis merged the two together after checking if the two produced the same output. Code solutions to be sent in for grading was selected by both code elegance and efficiency. For example, some code snippets by one individual was optimized by the other individual and vice versa. Dennis then sent Jack the finalized version of the code, who then approved it for final submission. We then drafted this report via Google Docs together.

# Methods for Running our Project Solutions

We chose to use Jupyter Notebooks to write our code when connecting/querying to MongoDB. Jupyter notebooks are useful for testing purposes, but also to view and organize the output for each individual query. This way, the grader can view input and output very easily to make sure our code is doing what it is supposed to do, even without the option of running it for themselves. We also include comments in the notebook for each query. Follow one of the viewing methods below depending on if you want to run the code or simply view it.

If one chooses to actually run our code, make sure to first start up MongoDB. If one does not have MongoDB installed, download from the [MongoDB website](https://www.mongodb.com/download-center#community). We recommend using the current stable release (3.4.10), which is the version we used for the project. Once downloaded, go to wherever you chose to download MongoDB (default is in C:\Program Files\MongoDB in Windows) and run mongo.exe and mongod.exe to spin up a local server. Our code will connect to the local instance of MongoDB and create our database and collections.

## Run Jupyter Notebook Locally

We wrote our code in Python, particularly using an Ipython (Jupyter) notebookIf one does not have Python and or Jupyter installed, we recommend downloading the Anaconda Distrubtion (with Python version 3.6). This distribution includes Python 3, Jupyter notebook, and several other packages used in Data Science practices. Download and follow the in-depth tutorial [here](https://www.anaconda.com/download/).

## Run Jupyter Notebook Online

If one does not have Python installed, one can view the notebook online in a semi-interactive method. Visit <https://try.jupyter.org/>, the official online testing space for Jupyter notebooks. Click ‘upload’ on the right hand side, then our project ‘proj4-dennis-silva.ipynb’. Then click on the newly uploaded notebook and view the notebook interactively.

## View Notebook via HTML in Browser

If one does not want to interact with our notebook, but solely check on the code descriptions, the code inputs, and the NoSQL document outputs, one can view our notebook in HTML in your browser (we recommend Chrome). We included in our zip a file called ‘project4-dennis-jack.html’. Open this file in a browser, and an html non-interactive version of our code can be seen. This version looks exactly like it would in a Jupyter notebook environment, except it is static.