**8-1 Assignment: Data Aggregation Pipeline**

Christine R. Emerson

Department of Computer Science, Southern New Hampshire University

CS-340: Client/Server Development

Professor. Vanderburgh

August 8, 2023

1. Using the mongoimport tool, **create the database** “companies” by loading the documents found in the “companies.json” file into the “research” collection. This file is located in the “/usr/local/datasets/” directory in Apporto.

A screen shot of a computer

Description automatically generated

**Verify your load by issuing the following queries:**

* 1. db.research.find({"name" : "AdventNet"})

A screenshot of a computer program

Description automatically generated

* 1. db.research.find({"founded\_year" : 1996},{"name" : 1}).limit(10)

A screenshot of a computer program

Description automatically generated

1. Perform the following tasks **using MongoDB queries**:
   1. List only the first 20 names of companies founded after the year 2010, ordered alphabetically.

A computer screen with a black background

Description automatically generated

* 1. List only the first 20 names of companies with offices in either California or Texas, ordered by the number of employees and sorted largest to smallest.

**A screenshot of a computer

Description automatically generated**

1. **Design and implement a MongoDB aggregation pipeline** to show the total number of offices by state for all companies that have offices in the United States. Be sure that you account for the fact that some companies have offices in several states. Explain your aggregation pipeline.

A screenshot of a computer

Description automatically generated

A black rectangular object with white text

Description automatically generated

A screenshot of a computer

Description automatically generated

**Explanation:**

The aggregation pipeline starts by matching documents with offices in the United States. Then, the offices array is unwound to create separate documents for each office. Next, the documents are grouped by state\_code, and the total number of offices in each state is calculated. Finally, the results are sorted by state code. The output of the aggregation pipeline provides a list of states with the corresponding total number of offices for companies that have offices in the United States.