Philip Trinh

SNHU/ CS-340

18 Oct 2023

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

1. A screen shot of a computer

   Description automatically generatedUsing 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. Verify your load by issuing the following queries:
   1. A screenshot of a computer program

      Description automatically generateddb.research.find({"name" : "AdventNet"})

A screenshot of a computer screen

Description automatically generatedA computer screen shot of a program

Description automatically generatedA screen shot of a computer program

Description automatically generated

* 1. A screenshot of a computer program

     Description automatically generateddb.research.find({"founded\_year" : 1996},{"name" : 1}).limit(10)

1. Perform the following tasks **using MongoDB queries**:
   1. A screenshot of a computer screen

      Description automatically generatedList only the first 20 names of companies founded after the year 2010, ordered alphabetically.
   2. A screenshot of a computer program

      Description automatically generatedList 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 program

Description automatically generatedA screenshot of a computer program

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.

Explanation of each stage in the aggregation:

Unwind Stage: Companies with offices in multiple states will have an array of office locations. To work with these individual locations, I used the $unwind stage to deconstruct the offices array.

Match Stage: Filter companies that have offices in the United States. I used the $match stage to do this by checking if the offices.country is "United States."

Group Stage: Group the data by the offices.state\_code field. I used the $group stage to group by offices.state\_code and calculate the count of offices in each state.

Sort Stage: To display the results in a meaningful order, I used the $sort stage to sort the states by the count of offices in descending order.

The result of this aggregation will be a list of states (identified by their state codes) with the total number of employees for companies that have offices in each state. It provides a breakdown of employee counts by state for all U.S.-based offices.

A screenshot of a computer program

Description automatically generated