Mitchell Fontaine

CS-340

12/17/2024

8-1 Assignment: Data Aggregation Pipeline

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 computer screen with white text

Description automatically generated

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

A screenshot of a computer

Description automatically generated

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

A screenshot of a computer screen

Description automatically generated

Perform the following tasks **using MongoDB queries**:

1. List only the first 20 names of companies founded after the year 2010, ordered alphabetically.

A screenshot of a computer screen

Description automatically generated

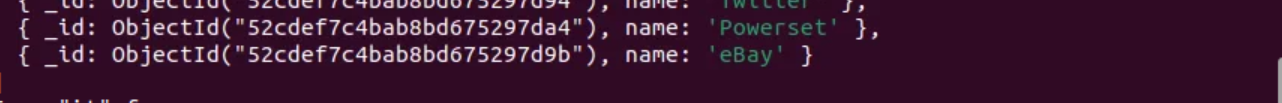
A computer screen with white text

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 program

Description automatically generated



**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.

$unwind: "$offices" Deconstruct the offices array.

$match: { "offices.country\_code": "USA" } Filters for US offices.

$match: { "offices.state\_code": { $exists: true, $ne: null } } Filters out missing or null state codes. I initially didn’t have this and had a null state with 252 offices.

$group: { \_id: "$offices.state\_code", totalOffices: { $sum: 1 } } Group by state and count the number of offices.

$sort: { totalOffices: -1 } Sorts by office count.

A screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated