8-1 Assignment: Data Aggregation Pipeline

Vincent Yanez

8/20/2024

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. Verify your load by issuing the following queries:

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

A screen shot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a computer program

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

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.

A screenshot of a computer program

Description automatically generated

Explain your aggregation pipeline:

Explanation:

* $match: Filters companies with offices in the USA.
* $unwind: Splits office locations into separate entries.
* $group: Counts the offices in each state.
* $sort: Orders the states by the number of offices, from most to least.

This pipeline will provide the total number of offices by state for companies in the United States.