Dominic Drury

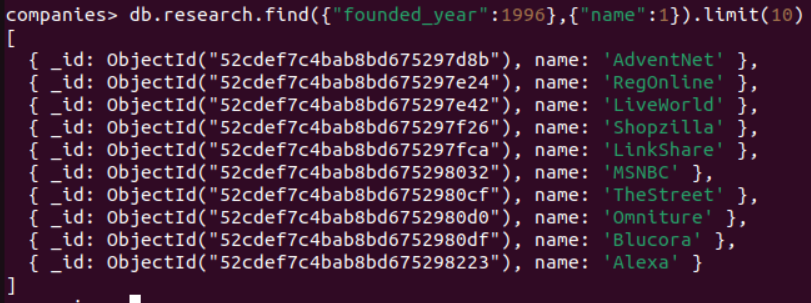
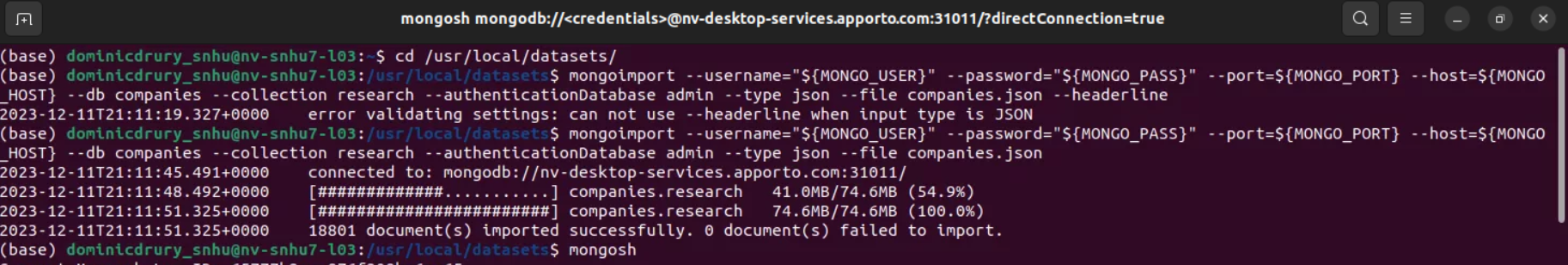
12/11/2023

CS-340

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. Verify your load by issuing the following queries:

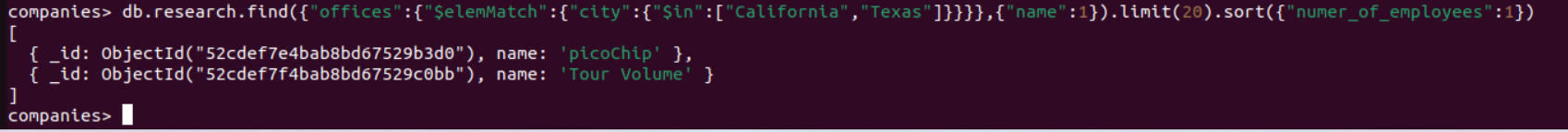
* db.research.find({"name" : "AdventNet"})
* db.research.find({"founded\_year" : 1996},{"name" : 1}).limit(10)



Perform the following tasks using MongoDB queries:

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





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.



For my pipeline I chose to use unwind to create documents for each office a company has, I used match to filter out all companies without an office inside of the USA, I used group to organize the documents by the state code and used sum to total the number of offices in each code, finally in order to sort the results from most offices to least offices I used sort with a –1 so the results were in descending order (largest number to smallest number).