

Amber Foster

April 19, 2022

8-1 Assignment

## DATA AGGREGATION PIPELINE

### 1. CREATING THE DATABASE:

```
(base) amberfoster3_snhu@msnv-snhu3-1002:~$ cd /usr/local/datasets
(base) amberfoster3_snhu@msnv-snhu3-1002:~/usr/local/datasets$ /usr/local/bin/mongodb_ctl start-noauth
+++ Starting MongoDB (NOAUTH): Port=54917 Unix Socket=/tmp/mongodb-54917.sock Dir=/home/amberfoster3_snhu/mongodb
(base) amberfoster3_snhu@msnv-snhu3-1002:~/usr/local/datasets$ mongoimport --port 54917 --db companies --collection research ./companies.json
2022-04-19T21:13:02.343+0000 connected to: mongodb://localhost:54917/
2022-04-19T21:13:05.343+0000 [#####] companies.research 68.8MB/74.6MB (92.1%)
2022-04-19T21:13:05.665+0000 [#####] companies.research 74.6MB/74.6MB (100.0%)
2022-04-19T21:13:05.665+0000 18801 document(s) imported successfully. 0 document(s) failed to import.
(base) amberfoster3_snhu@msnv-snhu3-1002:~/usr/local/datasets$ mongo
MongoDB shell version v4.2.6
connecting to: mongodb://127.0.0.1:54917/?compressors=disabled&gssapiServiceName=mongodb
Implicit session: session { "id" : "UUID("462b4e5b-8519-48b0-90f5-134661c33aa")" }
MongoDB server version: 4.2.6
Server has startup warnings:
2022-04-19T21:08:21.421+0000 I CONTROL [initandlisten]
2022-04-19T21:08:21.421+0000 I CONTROL [initandlisten] ** WARNING: Access control is not enabled for the database.
2022-04-19T21:08:21.421+0000 I CONTROL [initandlisten] ** Read and write access to data and configuration is unrestricted.
2022-04-19T21:08:21.421+0000 I CONTROL [initandlisten]
2022-04-19T21:08:21.421+0000 I CONTROL [initandlisten] ** WARNING: This server is bound to localhost.
2022-04-19T21:08:21.421+0000 I CONTROL [initandlisten] ** Remote systems will be unable to connect to this server.
2022-04-19T21:08:21.421+0000 I CONTROL [initandlisten] ** Start the server with --bind ip <address> to specify which IP
2022-04-19T21:08:21.421+0000 I CONTROL [initandlisten] ** addresses it should serve responses from, or with --bind ip all to
2022-04-19T21:08:21.421+0000 I CONTROL [initandlisten] ** bind to all interfaces. If this behavior is desired, start the
2022-04-19T21:08:21.421+0000 I CONTROL [initandlisten] ** server with --bind ip 127.0.0.1 to disable this warning.
2022-04-19T21:08:21.421+0000 I CONTROL [initandlisten]
...
Enable MongoDB's free cloud-based monitoring service, which will then receive and display
metrics about your deployment (disk utilization, CPU, operation statistics, etc).

The monitoring data will be available on a MongoDB website with a unique URL accessible to you
and anyone you share the URL with. MongoDB may use this information to make product
improvements and to suggest MongoDB products and deployment options to you.

To enable free monitoring, run the following command: db.enableFreeMonitoring()
To permanently disable this reminder, run the following command: db.disableFreeMonitoring()
...
> use companies
switched to db companies
```

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

```
> db.research.find({"name":"AdventNet"})
{ "_id" : ObjectId("52cdef7c4bab8bd675297d8b"), "name" : "AdventNet", "permalink" : "abc3", "crunchbase url" : "http://www.crunchbase.com/company/adventnet", "homepage url" : "http://adventnet.com", "blog url" :
"", "blog feed url" : "", "twitter username" : "manageengine", "category code" : "enterprise", "number of employees" : 600, "founded year" : 1996, "deadpooled year" : 2, "tag list" : "", "alias list" : "Zoho Ma
nagEngine", "email address" : "pr@adventnet.com", "phone number" : "925-924-9500", "description" : "Server Management Software", "created at" : ISODate("2007-05-25T19:24:22Z"), "updated at" : "Wed Oct 31 18:26
:09 UTC 2012", "overview" : "<p>AdventNet is now <a href='\"/company/zoho-manageengine/\"'>Zoho ManageEngine</a> rel='\"nofollow\">Zoho ManageEngine</p><p>AdventNet has served a diver
se range of enterprise IT, networking and telecom customers.</p><p>AdventNet supplies server and network management software.</p>", "image" : { "available sizes" : [ [ 150, 55 ], "assets/images/resized/000
1/9732/19732v1-max-150x150.png" ], [ 150, 55 ], "assets/images/resized/0001/9732/19732v1-max-250x250.png" ], [ 150, 55 ], "assets/images/resized/0001/9732/19732v1-max-450x450.png" ] ], "products" : [ ], "r
elationships" : [ { "is past" : true, "title" : "CEO and Co-Founder", "person" : { "first name" : "Sridhar", "last name" : "Venbu", "permalink" : "sridhar-venbu" } }, { "is past" : true, "title" : "VP of Busines
s Dev", "person" : { "first name" : "Neil", "last name" : "Butani", "permalink" : "neil-butani" }, { "is past" : true, "title" : "Usability Engineer", "person" : { "first name" : "Bharath", "last name" : "Balas
ubramanian", "permalink" : "bharath-balasubramanian" }, { "is past" : true, "title" : "Director of Engineering", "person" : { "first name" : "Rajendran", "last name" : "Dandapani", "permalink" : "rajendran-dan
dapani" }, { "is past" : true, "title" : "Market Analyst", "person" : { "first name" : "Aravind", "last name" : "Natarajan", "permalink" : "aravind-natarajan" }, { "is past" : true, "title" : "Director of Pr
oduct Management", "person" : { "first name" : "Myther", "last name" : "Nizam", "permalink" : "myther-nizam" }, { "is past" : true, "title" : "Western Regional QEM Sales Manager", "person" : { "first name" : "I
an", "last name" : "Wenig", "permalink" : "ian-wenig" } }, "competitions" : [ ], "providerships" : [ { "title" : "DHFH", "is past" : true, "provider" : { "name" : "A Small Orange", "permalink" : "a-small-oran
ge" } }, "total money raised" : [ ], "funding rounds" : [ ], "investments" : [ ], "acquisitions" : null, "acquisitions" : [ ], "offices" : [ { "description" : "Headquarters", "address1" : "4900 Hopyard Rd.", "
address2" : "Suite 310", "zip code" : "94588", "city" : "Pleasanton", "state code" : "CA", "country code" : "USA", "latitude" : 37.692934, "longitude" : -121.984945 }, { "milestones" : [ ], "video embeds" : [ ],
"screenshots" : [ { "available sizes" : [ [ 150, 94 ], "assets/images/resized/0004/3400/43400v1-max-150x150.png" ], [ 150, 94 ], "assets/images/resized/0004/3400/43400v1-max-250x250.png" ], [ 150, 282 ],
"assets/images/resized/0004/3400/43400v1-max-450x450.png" ] }, "attribution" : null }, "external links" : [ ], "partners" : [ ] }
>
```

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

```
> db.research.find({"founded_year":1996},{ "name":1}).limit(10)
{ "_id" : ObjectId("52cdef7c4bab8bd675297d8b"), "name" : "AdventNet" }
{ "_id" : ObjectId("52cdef7c4bab8bd675297e24"), "name" : "RegOnline" }
{ "_id" : ObjectId("52cdef7c4bab8bd675297e42"), "name" : "LiveWorld" }
{ "_id" : ObjectId("52cdef7c4bab8bd675297f26"), "name" : "Shoptilla" }
{ "_id" : ObjectId("52cdef7c4bab8bd675297fca"), "name" : "LinkShare" }
{ "_id" : ObjectId("52cdef7c4bab8bd675298032"), "name" : "MSNBC" }
{ "_id" : ObjectId("52cdef7c4bab8bd6752980cf"), "name" : "TheStreet" }
{ "_id" : ObjectId("52cdef7c4bab8bd6752980d0"), "name" : "Omniure" }
{ "_id" : ObjectId("52cdef7c4bab8bd6752980ef"), "name" : "Blucora" }
{ "_id" : ObjectId("52cdef7c4bab8bd675298223"), "name" : "Alexa" }
```

## 2. USING MongoDB QUERIES:

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

```
> db.research.find({"founded_year":{"$gt":2010}},{"name":1}).sort({"name":1}).limit(20)
{ "_id" : ObjectId("52cdef7e4ab8bd67529b2f3"), "name" : "dshared" }
{ "_id" : ObjectId("52cdef7c4ab8bd67529b2d4"), "name" : "Advaliant" }
{ "_id" : ObjectId("52cdef7d4ab8bd67529b3b8"), "name" : "Advisor" }
{ "_id" : ObjectId("52cdef7d4ab8bd67529a048"), "name" : "Baveq" }
{ "_id" : ObjectId("52cdef7d4ab8bd6752992f7"), "name" : "Bling Easy" }
{ "_id" : ObjectId("52cdef7d4ab8bd675298da5"), "name" : "Carfeine" }
{ "_id" : ObjectId("52cdef7c4ab8bd675297f78"), "name" : "CirctUp" }
{ "_id" : ObjectId("52cdef7e4ab8bd67529b3bf"), "name" : "Clowdy" }
{ "_id" : ObjectId("52cdef7e4ab8bd67529b3bf"), "name" : "CompareChecker" }
{ "_id" : ObjectId("52cdef7e4ab8bd67529a38a"), "name" : "Cyphercor" }
{ "_id" : ObjectId("52cdef7f4ab8bd67529c6eb"), "name" : "DocASAP" }
{ "_id" : ObjectId("52cdef7e4ab8bd67529bb4d"), "name" : "Easel" }
{ "_id" : ObjectId("52cdef7d4ab8bd67529b31ff"), "name" : "EasyBib" }
{ "_id" : ObjectId("52cdef7d4ab8bd67529a0b4"), "name" : "Emotive Communications" }
{ "_id" : ObjectId("52cdef7e4ab8bd67529ae2a"), "name" : "FAT Media" }
{ "_id" : ObjectId("52cdef7d4ab8bd675299f92"), "name" : "FamilyDen" }
{ "_id" : ObjectId("52cdef7d4ab8bd67529b9f3"), "name" : "FirstString" }
{ "_id" : ObjectId("52cdef7c4ab8bd675297fec"), "name" : "Fixya" }
{ "_id" : ObjectId("52cdef7e4ab8bd67529b1ef"), "name" : "Fliggo" }
{ "_id" : ObjectId("52cdef7c4ab8bd67529830a"), "name" : "Fluc" }
```

b. List only the first 20 names of companies with offices in either California or Texas, ordered by the number of employees & sorted large to small.

```
> db.research.find({$or:[{"offices.state_code":"CA"}, {"offices.state_code":"TX"}]}, {"name":1}).sort({"number_of_employees":-1}).limit(20)
{ "_id" : ObjectId("52cdef7c4ab8bd675297e89"), "name" : "PayPal" }
{ "_id" : ObjectId("52cdef7d4ab8bd675298aa4"), "name" : "Samsung Electronics" }
{ "_id" : ObjectId("52cdef7d4ab8bd67529b0b9"), "name" : "Accenture" }
{ "_id" : ObjectId("52cdef7e4ab8bd67529aa51"), "name" : "Flextronics International" }
{ "_id" : ObjectId("52cdef7d4ab8bd675299156"), "name" : "Safeway" }
{ "_id" : ObjectId("52cdef7c4ab8bd675297e6ff"), "name" : "Sony" }
{ "_id" : ObjectId("52cdef7c4ab8bd675297dc4"), "name" : "Intel" }
{ "_id" : ObjectId("52cdef7c4ab8bd67529b1b9"), "name" : "Apple" }
{ "_id" : ObjectId("52cdef7c4ab8bd6752983b9"), "name" : "Dell" }
{ "_id" : ObjectId("52cdef7e4ab8bd67529b998"), "name" : "ExxonMobil" }
{ "_id" : ObjectId("52cdef7f4ab8bd67529c29e"), "name" : "Affiliated Computer Services" }
{ "_id" : ObjectId("52cdef7c4ab8bd675297da2"), "name" : "Cisco" }
{ "_id" : ObjectId("52cdef7c4ab8bd675297ee9"), "name" : "Sun Microsystems" }
{ "_id" : ObjectId("52cdef7c4ab8bd6752987ff"), "name" : "Texas Instruments" }
{ "_id" : ObjectId("52cdef7c4ab8bd675297dba"), "name" : "Google" }
{ "_id" : ObjectId("52cdef7c4ab8bd675297e8e"), "name" : "The Walt Disney Company" }
{ "_id" : ObjectId("52cdef7d4ab8bd675298cb5"), "name" : "Avaya" }
{ "_id" : ObjectId("52cdef7d4ab8bd67529921e"), "name" : "AMD" }
{ "_id" : ObjectId("52cdef7c4ab8bd6752982a8"), "name" : "Experian" }
{ "_id" : ObjectId("52cdef7c4ab8bd675297d9b"), "name" : "eBay" }
> db.research.find({"offices":{"$elemMatch":{"city":{"$in":["California","Texas"]}}}}, {"name":1}).sort({"number_of_employees":-1}).limit(20).pretty()
{ "_id" : ObjectId("52cdef7e4ab8bd67529b3b0"), "name" : "picoChip" }
{ "_id" : ObjectId("52cdef7f4ab8bd67529c0bb"), "name" : "Tour Volume" }
```

## DESIGN & IMPLEMENT A MongoDB AGGREGATION PIPELINE:

```
> db.research.aggregate([{$match:{'offices.country_code':"USA"}},{$group:{'_id':'offices.state_code',sumQuantity:{$sum:'$number_of_employees'}}}]
{ "_id" : "NC", "sumQuantity" : 85 }
{ "_id" : null, "sumQuantity" : 98 }
{ "_id" : "VA", "sumQuantity" : 0 }
{ "_id" : "OR", "CA", "sumQuantity" : 165 }
{ "_id" : "IN", "IN", "CA", "TX", "FL", "GA", "SC", "GA", "TN", "NY", "CA", "NV", "sumQuantity" : 0 }
{ "_id" : "CA", "sumQuantity" : 797 }
{ "_id" : null, null, null, null, null, null, null, null, null, null, null, null, null, null, null, null, "CA", "sumQuantity" : 800 }
{ "_id" : null, "NY", "VA", "sumQuantity" : 200 }
{ "_id" : null, null, "sumQuantity" : 1245 }
{ "_id" : "CA", "ID", "sumQuantity" : 0 }
{ "_id" : "MD", "PA", "sumQuantity" : 6 }
{ "_id" : "IL", "sumQuantity" : 5 }
{ "_id" : null, "WA", "NV", "sumQuantity" : 50 }
{ "_id" : "CA", "NV", "AZ", "CA", "sumQuantity" : 148 }
{ "_id" : null, null, "WA", "sumQuantity" : 4 }
{ "_id" : "NV", "sumQuantity" : 206 }
{ "_id" : "CA", "TX", "sumQuantity" : 1624 }
{ "_id" : null, "CA", "NV", "sumQuantity" : 150 }
{ "_id" : "CA", "NV", "IL", "sumQuantity" : 0 }
{ "_id" : "CA", "AL", "sumQuantity" : 58 }
Type "it" for more
{ "_id" : "IL", "NV", "sumQuantity" : 500 }
{ "_id" : "MD", "CA", "NV", "sumQuantity" : 0 }
{ "_id" : "IL", "NV", "CA", "AZ", "sumQuantity" : 90 }
{ "_id" : "CA", "WA", "CO", "sumQuantity" : 0 }
{ "_id" : "NC", "sumQuantity" : 137 }
{ "_id" : "MD", "NV", "WA", "sumQuantity" : 0 }
{ "_id" : "CA", "NV", "CO", "sumQuantity" : 225 }
{ "_id" : "CA", "PA", "sumQuantity" : 0 }
{ "_id" : null, null, "NV", "sumQuantity" : 75 }
{ "_id" : "FL", "WA", "sumQuantity" : 0 }
{ "_id" : "CA", "NV", "sumQuantity" : 55 }
{ "_id" : null, "TN", "sumQuantity" : 0 }
{ "_id" : "NC", "NC", "sumQuantity" : 70 }
{ "_id" : "CA", "WA", "sumQuantity" : 9 }
{ "_id" : "NV", "TX", "sumQuantity" : 52 }
{ "_id" : "CA", "CA", "sumQuantity" : 700 }
{ "_id" : "GA", "sumQuantity" : 90 }
{ "_id" : "CO", "IL", "sumQuantity" : 0 }
{ "_id" : "IA", "IL", "sumQuantity" : 0 }
{ "_id" : "OR", "CA", "sumQuantity" : 250 }
Type "it" for more
> db.research.aggregate([{$match:{'offices':{'$elemMatch':{'country_code':"USA'}}}},{$group:{'_id':null,total_number_of_employees:{$sum:1}}}]
{ "_id" : null, "total_number_of_employees" : 9266 }
```

## EXIT & STOP MongoDB:

```
> exit
bye
(base) amberfoster3@snhu3-lb02: /usr/local/dataset4 $ ./usr/local/bin/mongodb_c11 stop
*** Stopping MongoDB for amberfoster3@snhu
(base) amberfoster3@snhu3-lb02: /usr/local/dataset4 $
```