## Install MongoDB on your EC2 instance

Follow the instruction on

<https://docs.aws.amazon.com/zh_cn/dms/latest/sbs/CHAP_MongoDB2DocumentDB.02.html>

<https://docs.mongodb.com/manual/tutorial/install-mongodb-on-amazon/>

**Verify Linux Distribution:**



**Configure the package management system (yum).:**

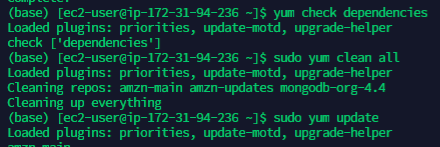


Input:

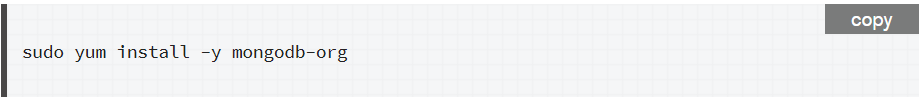


**Install the MongoDB packages.:**

Update **glibc**: if you glibc is old version <= 2.17







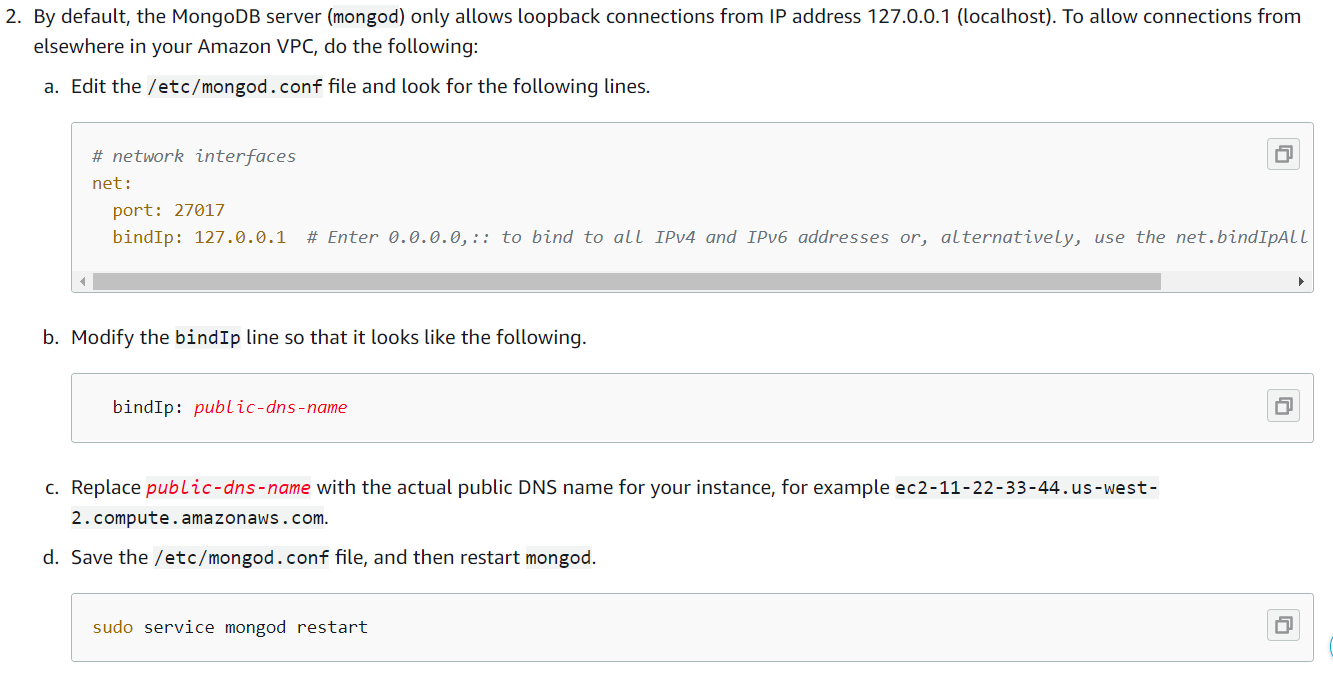
Follow the instructions in the remaining part: **Run MongoDB Community Edition**

And remember to **ulimit** the **Resource Utilization**



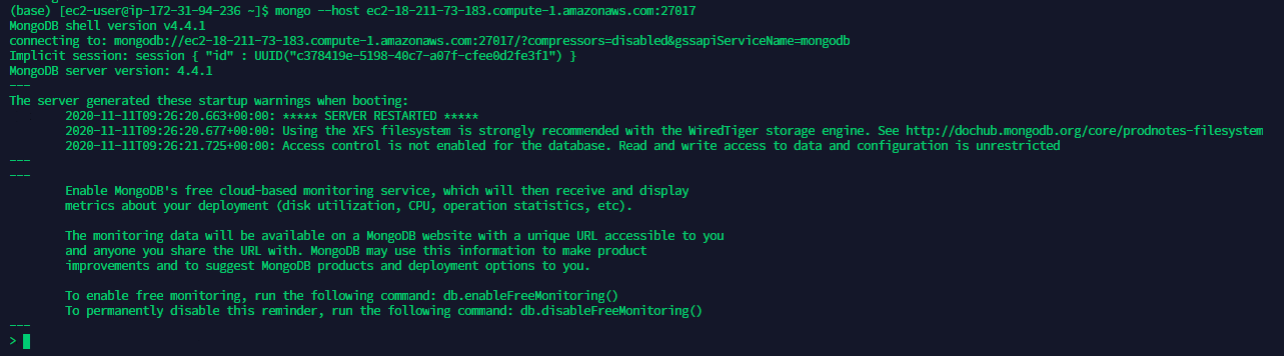
Configure host name to public DNS:

sudo vi /etc/mongod.conf



**Run MongoDB:**

mongo --host ec2-18-211-73-183.compute-1.amazonaws.com:27017



**Next step please see the next page**

## Download GSE13355.zip (attached)

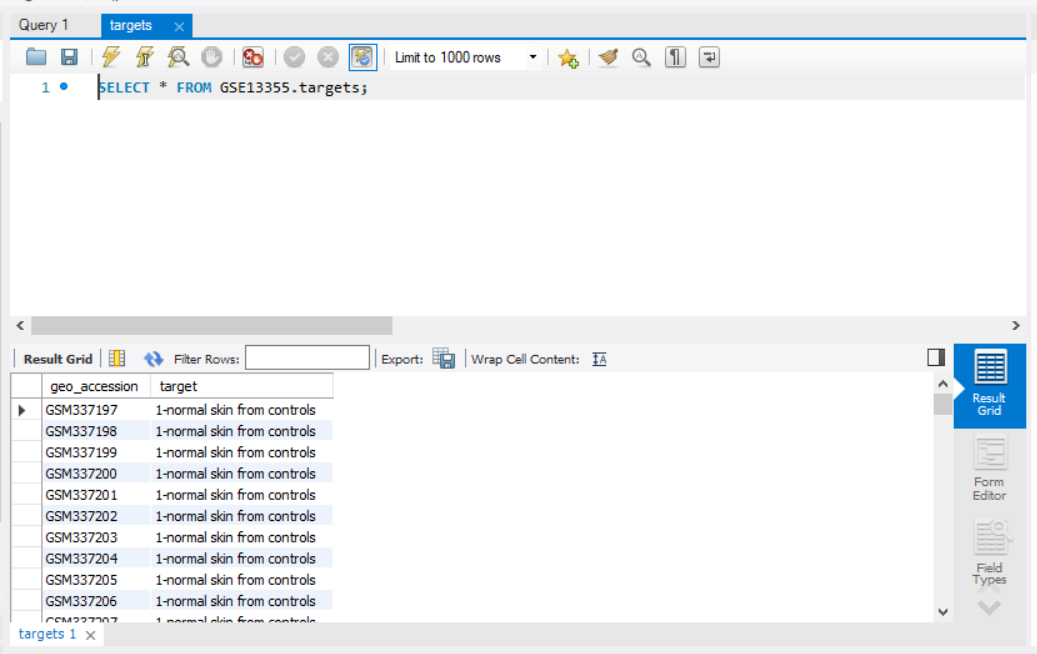
## Create GSE13355 database on MySQL and MongoDB

## Upload GSE13355\_targets.csv to MySQL and MongoDB

**MySQL:**

CREATE SCHEMA `GSE13355` ;

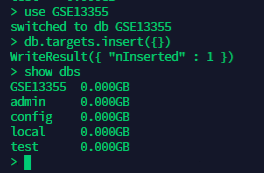
Use MySQL Workbench to create table and import data. Result:



**MongoDB:**

use GSE13355

db.targets.insert({})



Import data by shell:

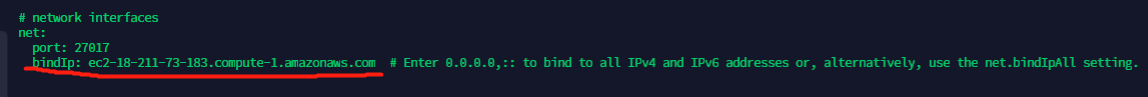
mongoimport --db GSE13355 --collection targets --type csv --headerline --file /home/ec2-user/GSE13355\_targets.csv --host ec2-18-211-73-183.compute-1.amazonaws.com:27017

Reference: <https://blog.csdn.net/qq_32447321/article/details/79223332>

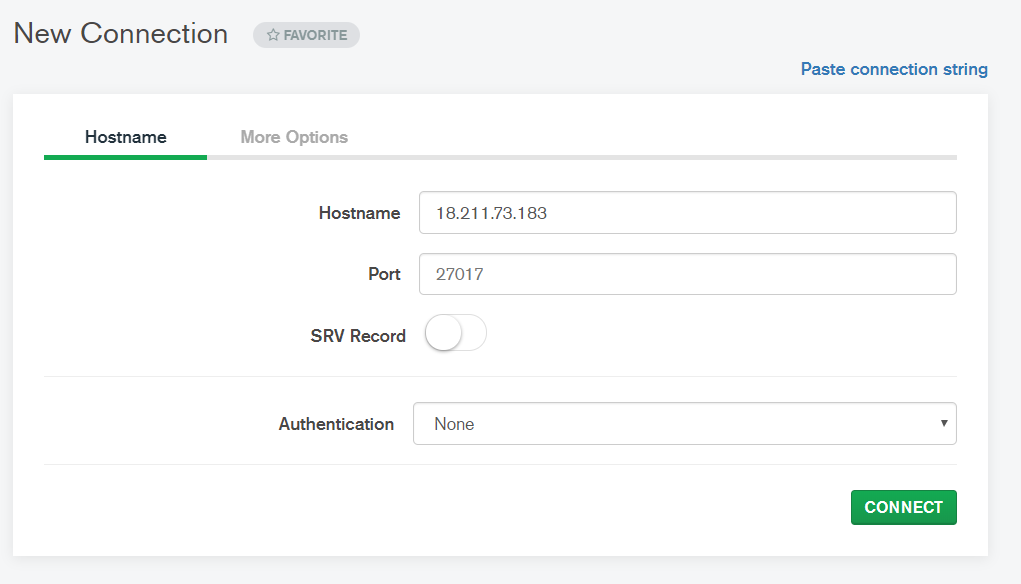
Valid import data using UI interface: **MongoDB Compass**

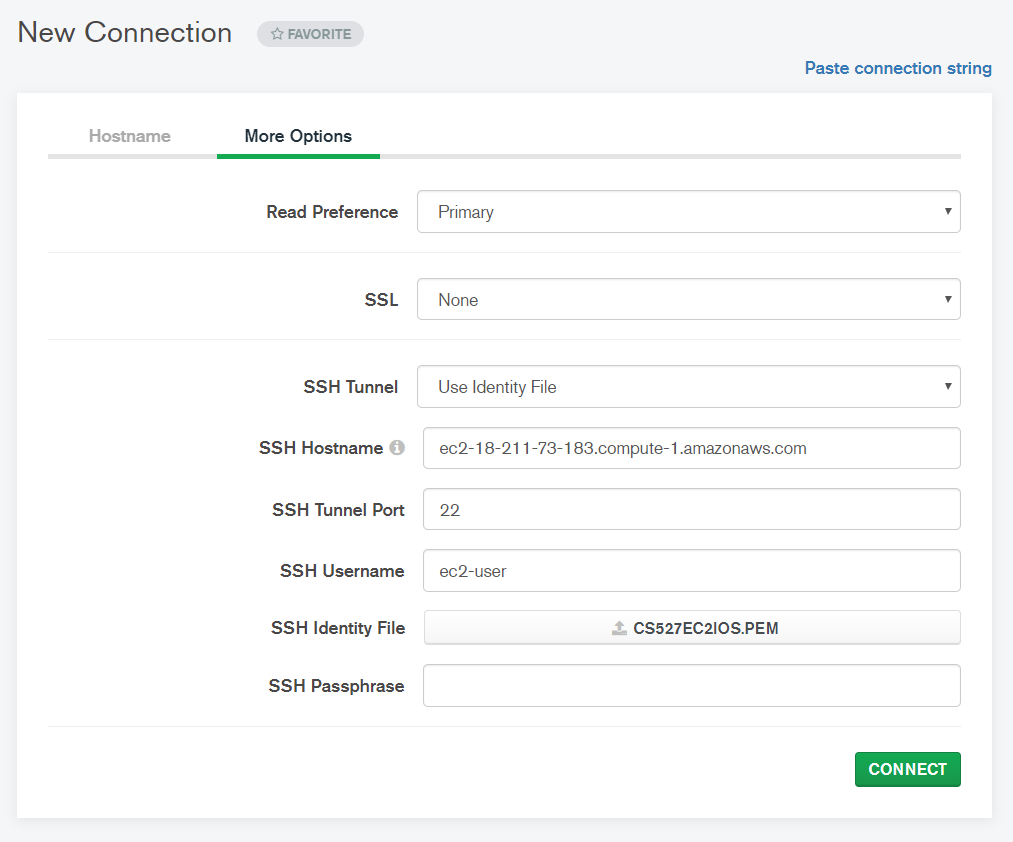
Firstly config the MongoDB server:

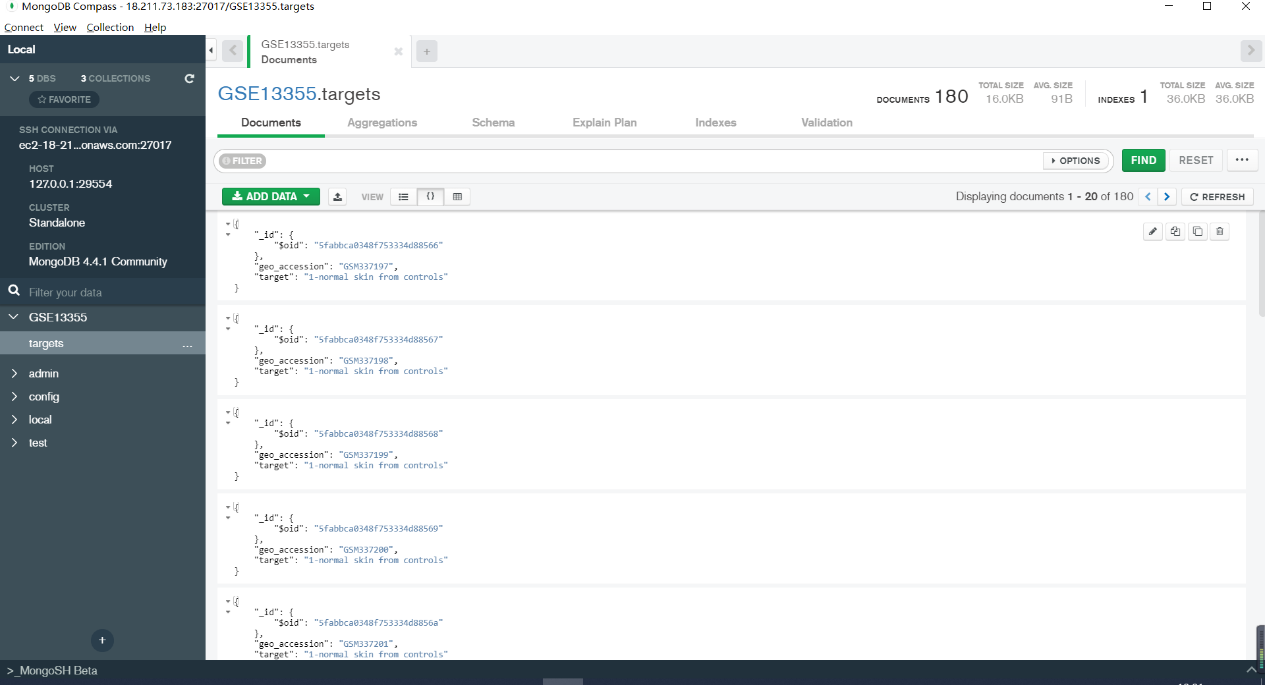
sudo vi /etc/mongod.conf



Then create connection in MongoDB Compass:



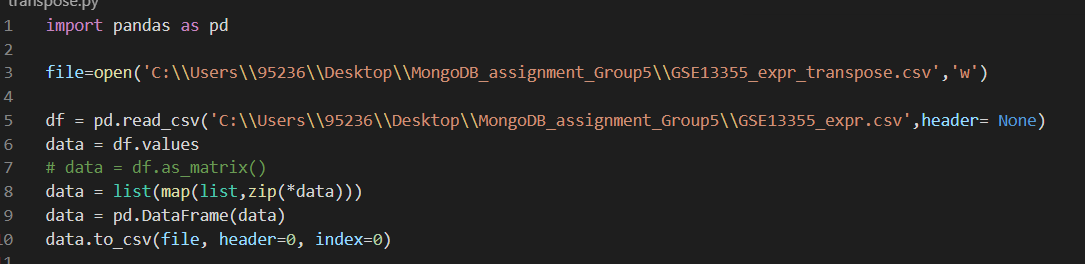




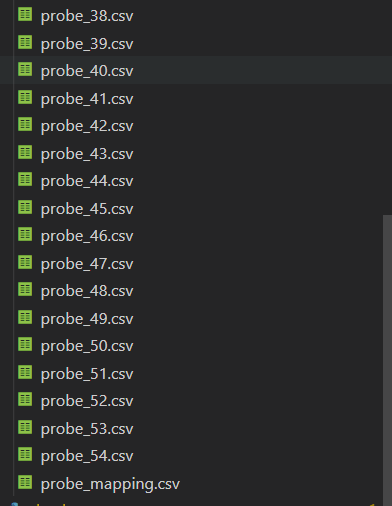
## Transpose GSE13355\_expr.csv and upload the transposed file to MySQL and MongoDB

**MySQL:**

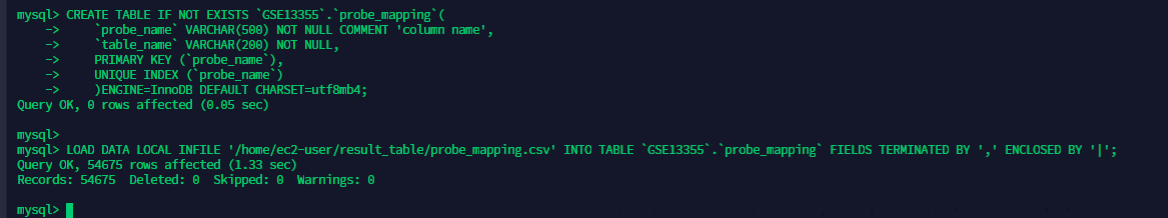
Use **transpose.py** to transpose the csv file and get raw data.



Split the csv into multiple tables by **split\_table.py** each has 1001 columns, since the max number of columns in MySQL is 4096, that one is the id related to the **geo\_info** like GSM337197 and others are **probe\_name** and the mapping between the **probe\_name** and **table\_name** is stored in table **probe\_mapping**. And write the **geo\_info** into the table **expr** with **geo\_id** and **geo\_name**

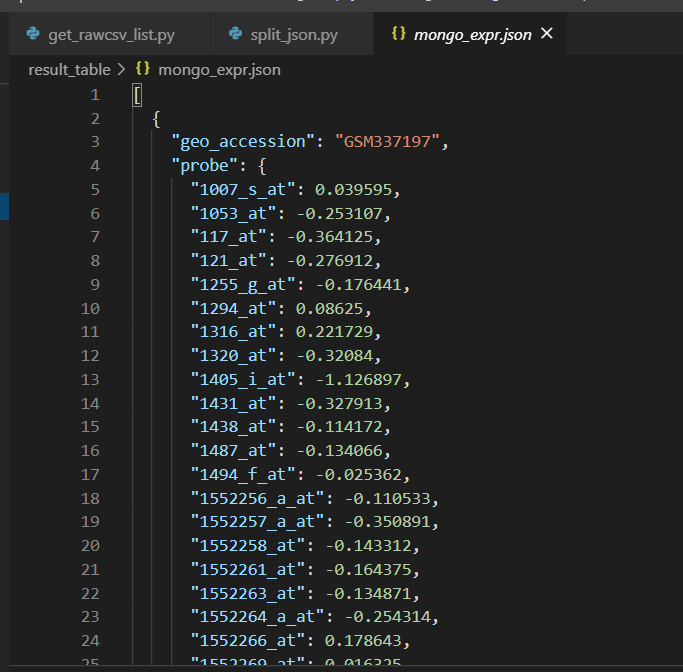
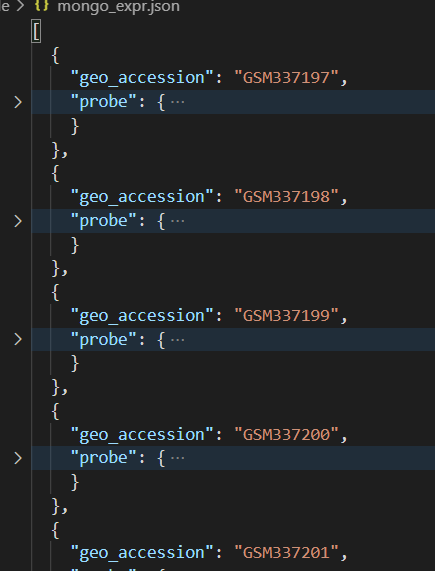


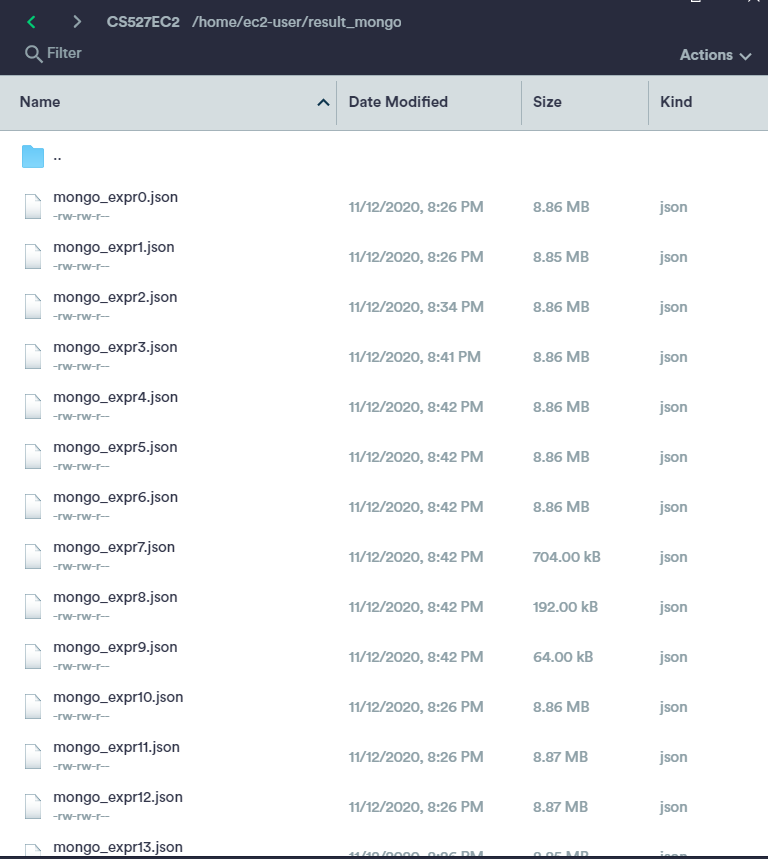
then use the **load** method to import the data from ec2 to rds:



**MongoDB:**

We use **split\_json.py** to split the data to the json file as the following format, because the json result is too large and in order to import more convenient, we split the json file to 36 json files and each of it is less than 10 MB so that we can use **mongoimport –jsonArray** to import the data:

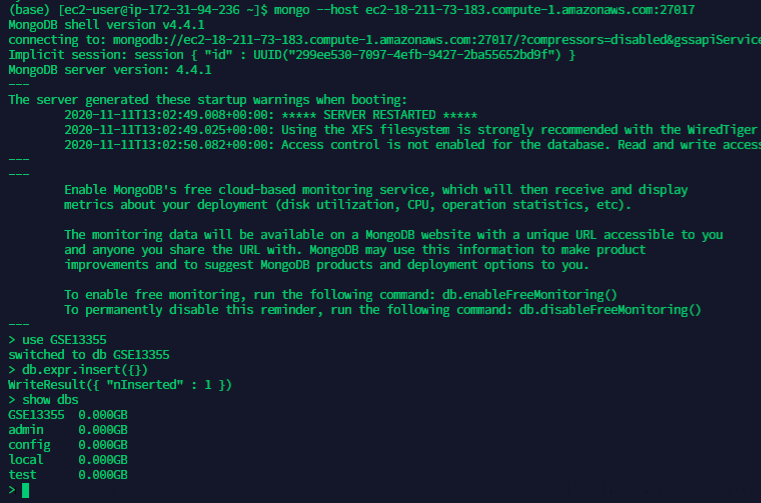


And then as the above method, use the **mongoimport** to import the json data into MongoDB:

mongo --host ec2-18-211-73-183.compute-1.amazonaws.com:27017

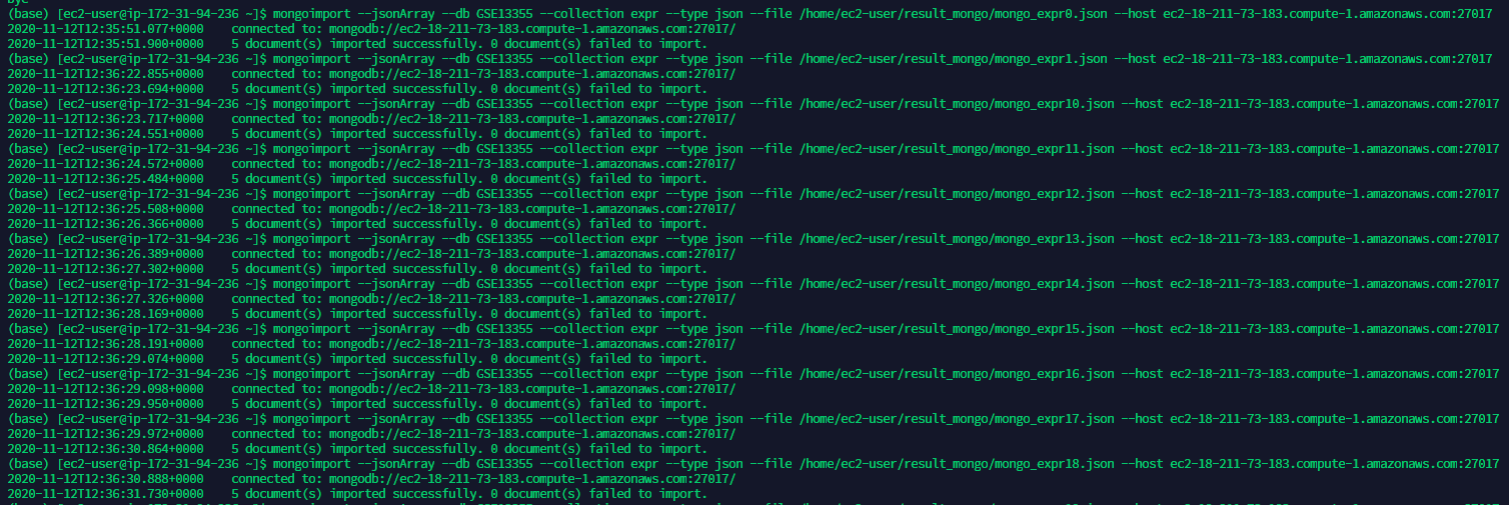
use GSE13355

db.targets.insert({})



Example:

mongoimport --jsonArray --db GSE13355 --collection expr --type json --file /home/ec2-user/result\_mongo/mongo\_expr1.json --host ec2-18-211-73-183.compute-1.amazonaws.com:27017



## Up to 5 teams will be selected to present their assignment