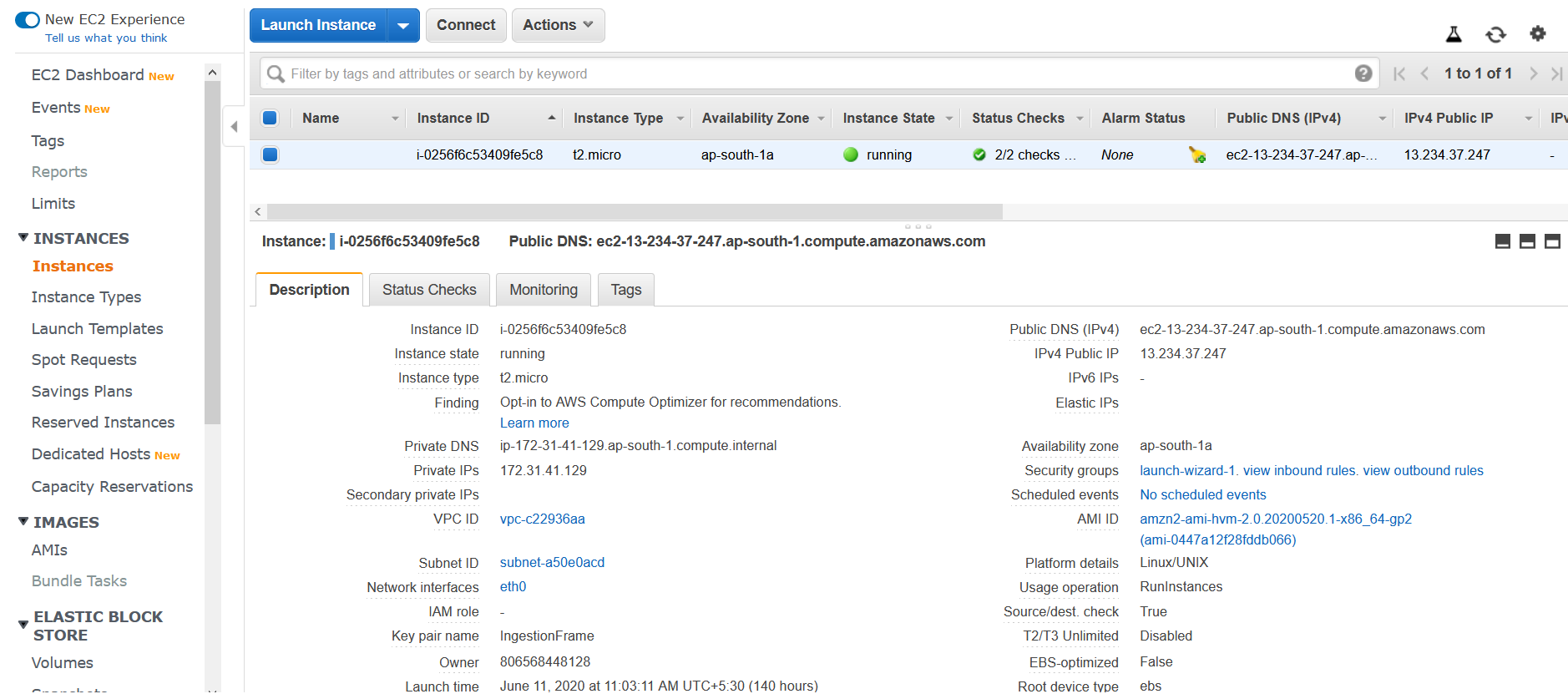
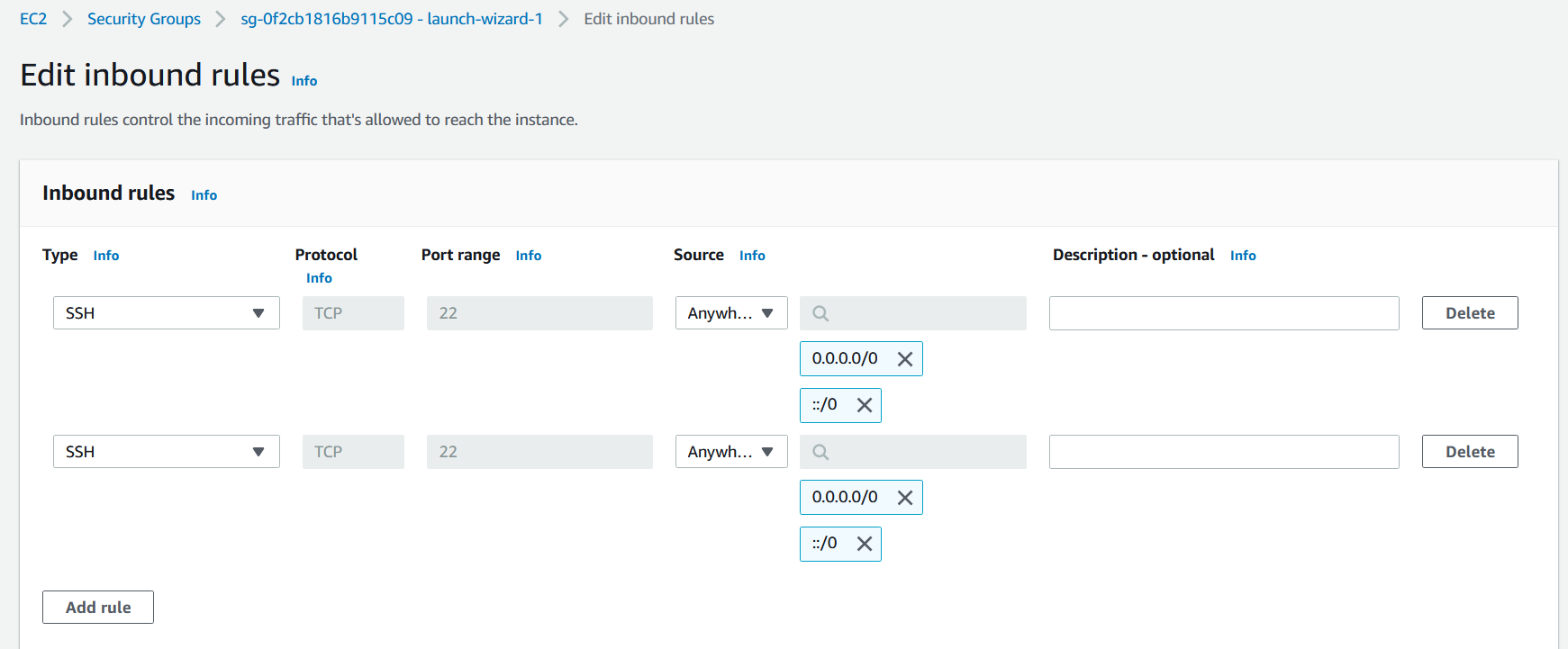
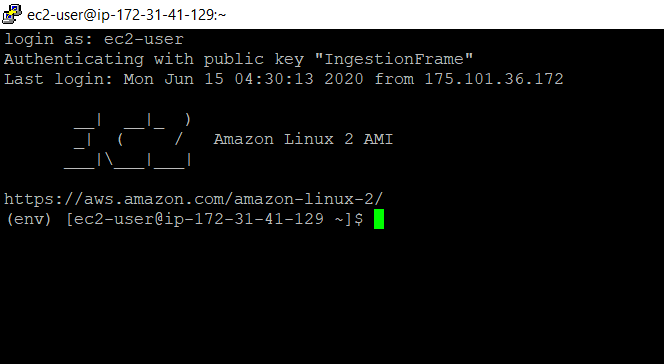
DI Framework on AWS Ec2 compute node execution steps:

Create a compute node Ec2 with key pairs and SSH access points.



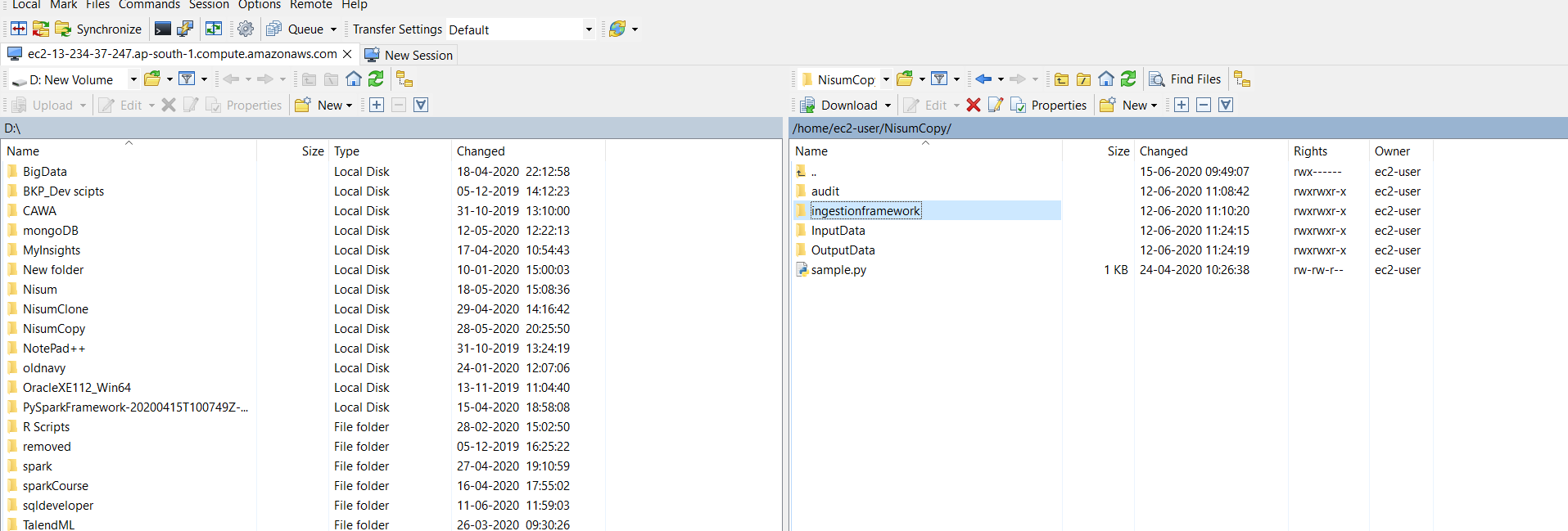
Enable the SSH inbound rules for ec2 node.



Now use putty to connect with EC2 node using public DNS name and SSH authentication.

Using WINSCP you can upload your framework tar file into EC2 node.

Note: - You can use the same authentication(as same as putty) method for WINSCP also.



Now you can install the python and spark into ec2 cluster.

Note: - Ec2 machine have python 2.7 by default. Need to upgrade the python to latest version.

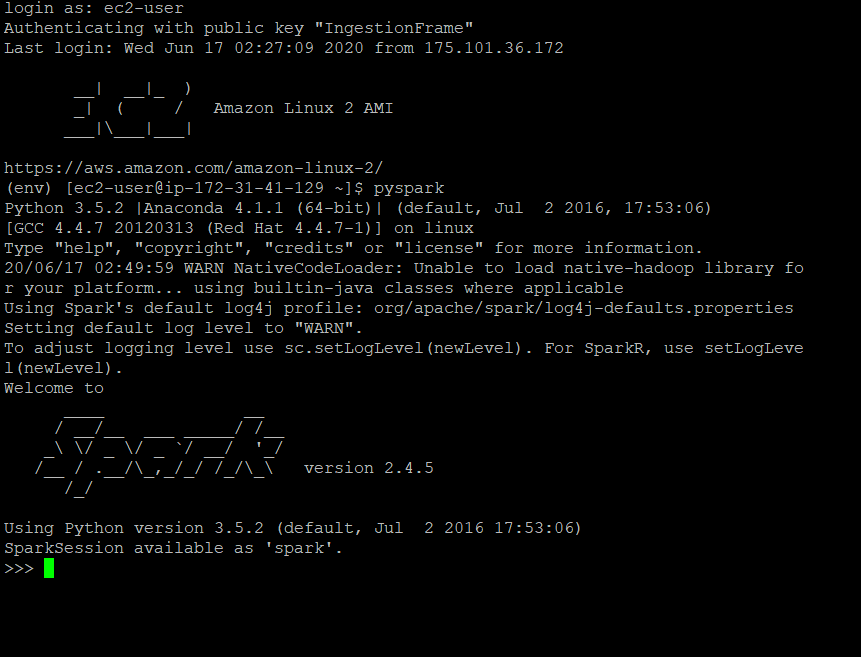
Please command history which used to upgrade the python and to install spark.



Reference websites: -

<https://medium.com/@josemarcialportilla/getting-spark-python-and-jupyter-notebook-running-on-amazon-ec2-dec599e1c297>

<https://aws.amazon.com/premiumsupport/knowledge-center/ec2-linux-python3-boto3/>



Executing Ingestion framework: -

go to config file to change the input file location and output file location.

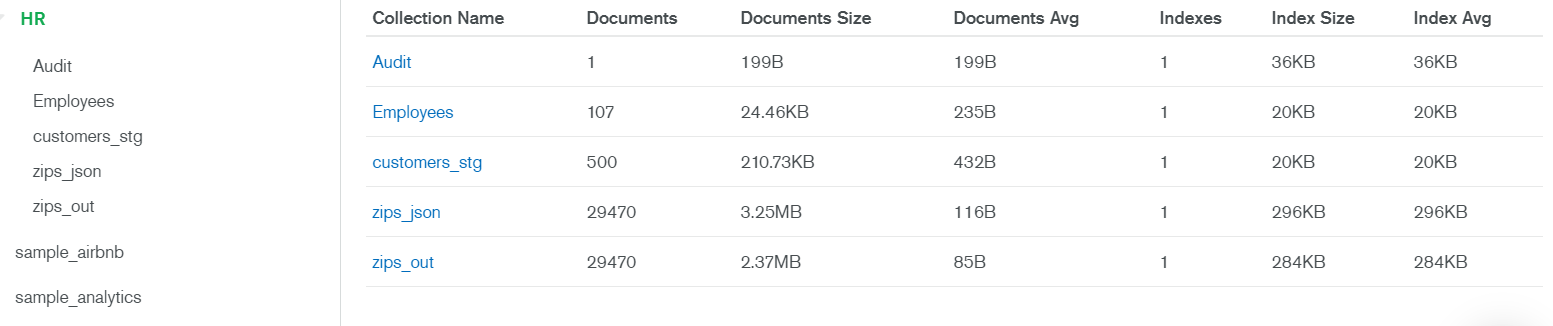


Now execute the spark submit command.

In this execution that I am loading the data from csv to MongoDB.

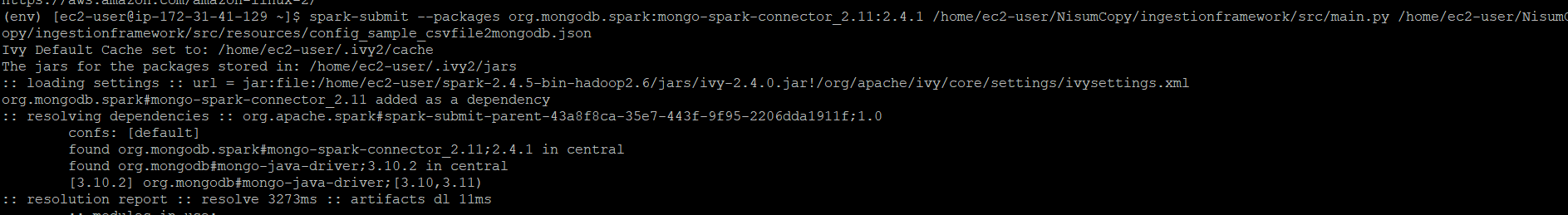
**MongoDb screen shot before executing the job:**

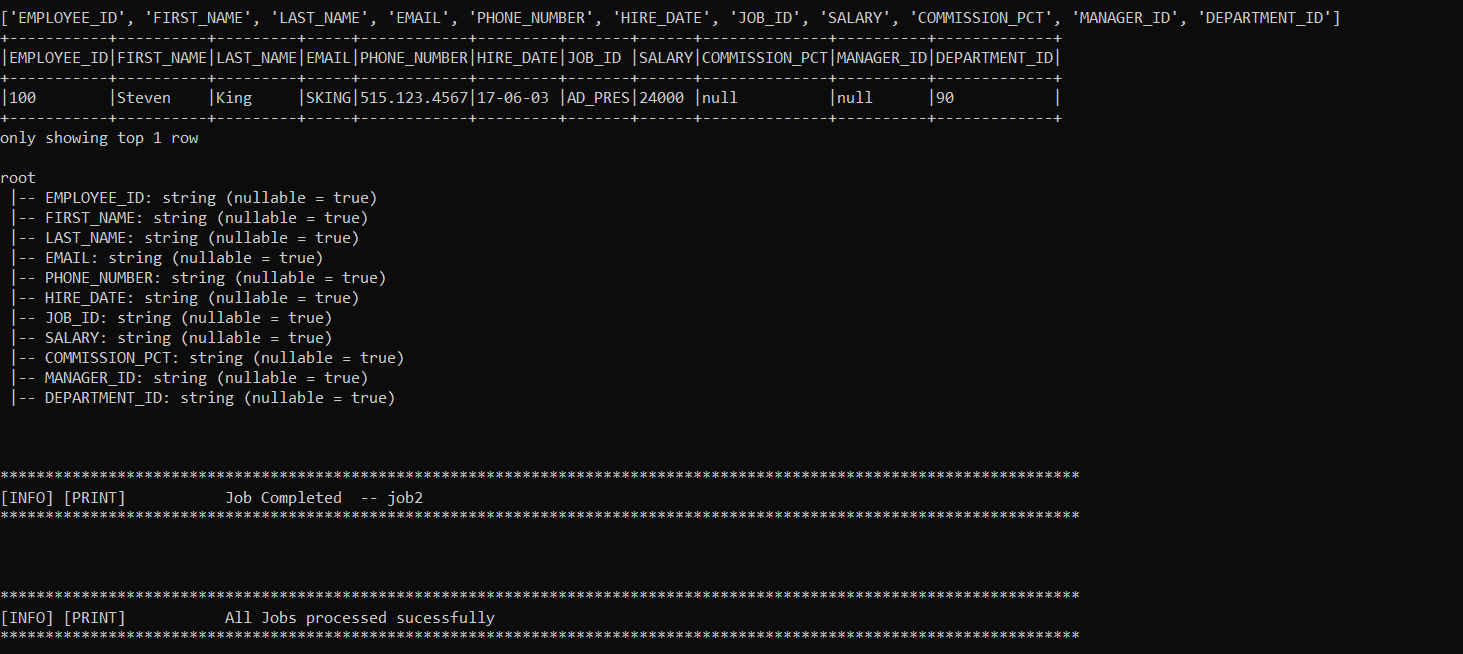
**Employee\_stg** collection is not there in HR database.



**Executing command:**

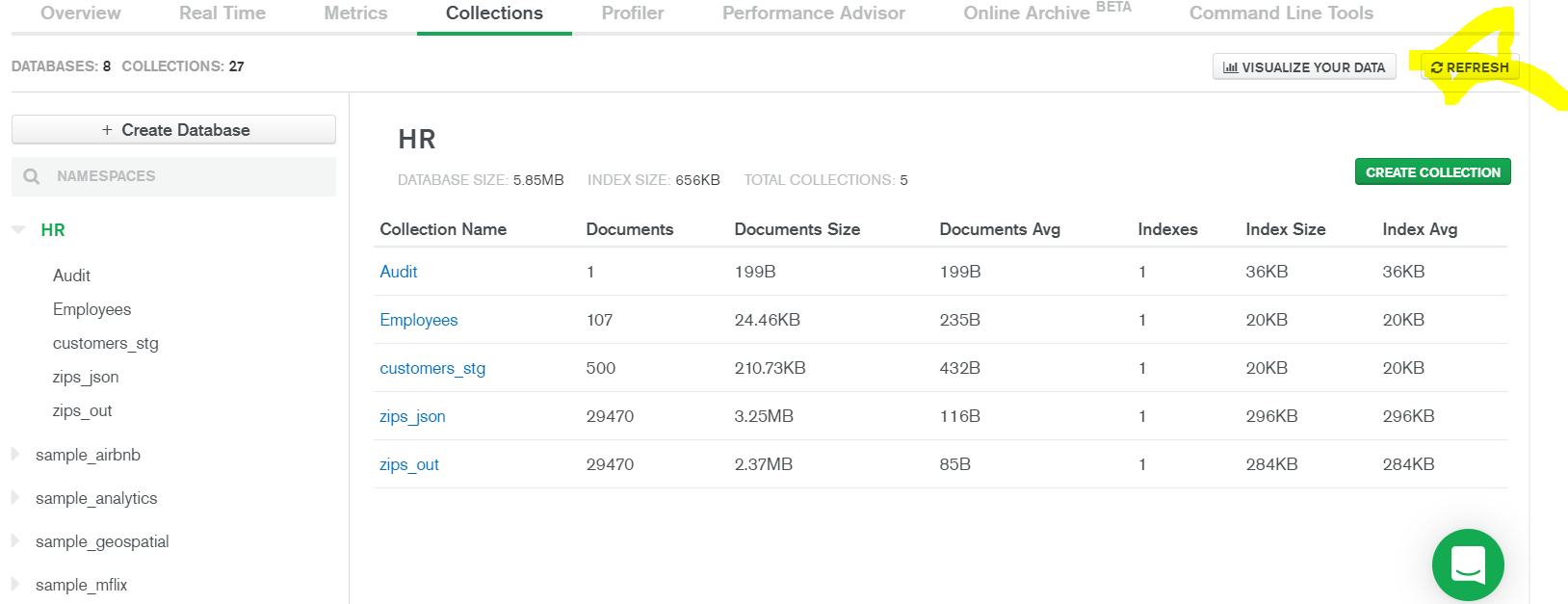
spark-submit --packages org.mongodb.spark:mongo-spark-connector\_2.11:2.4.1 /home/ec2-user/NisumCopy/ingestionframework/src/main.py /home/ec2-user/NisumCopy/ingestionframework/src/resources/config\_sample\_csvfile2mongodb.json

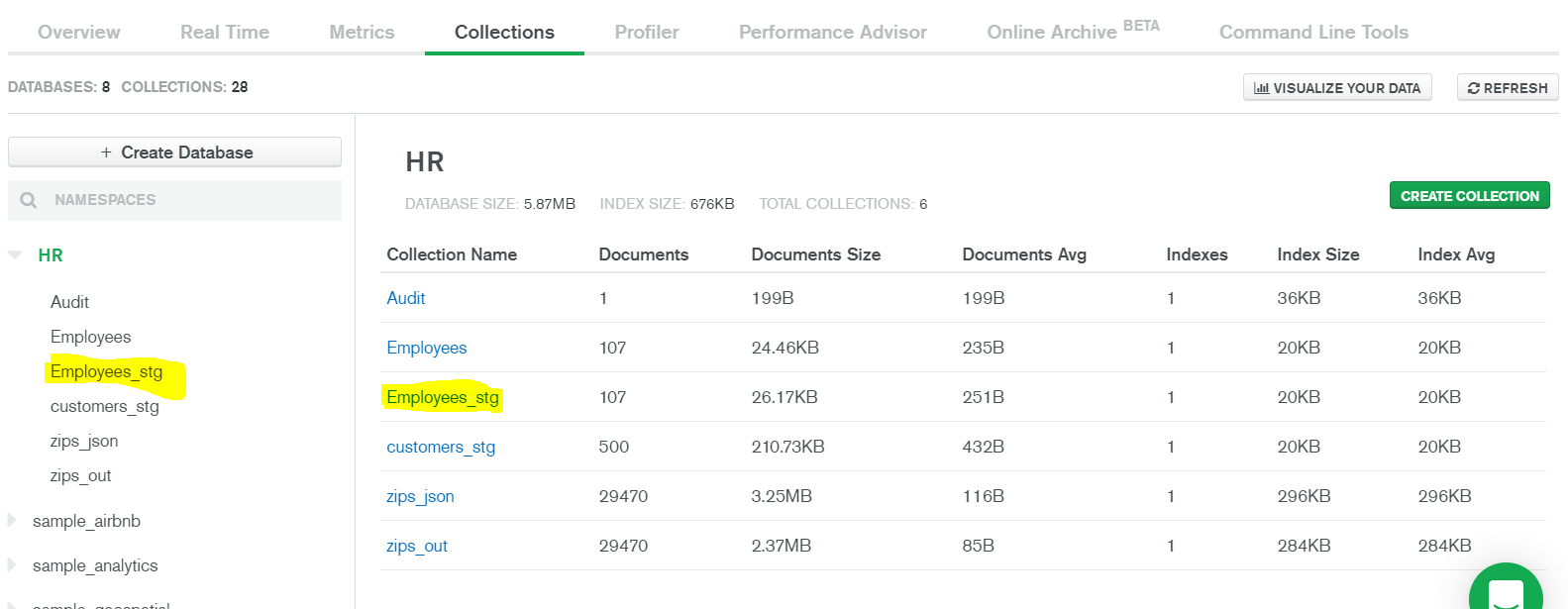


Job Status: -

**After Executing Spark MongoDB screenshot: -**

Refresh collection on MongoDB.





Click on collection to check the documents.

