**PROJECT-1**

**Overview**

This project aims to use a URL that points to a zip file containing multiple JSON files. These JSON files contain various data structures with multiple documents. The goal is to download the zip file from the URL, extract the data from the JSON files, store it in Amazon S3, and load it into Amazon DocumentDB.

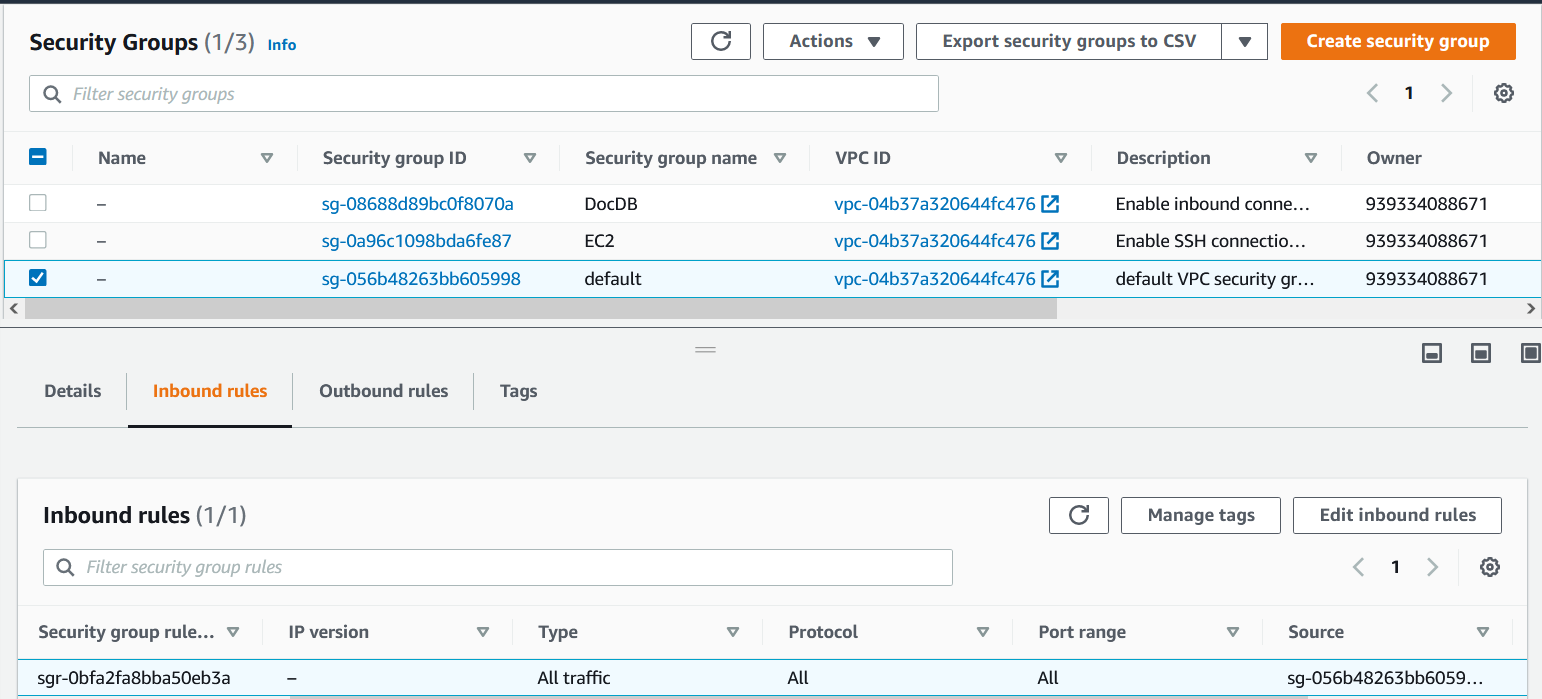
**Project Goals**

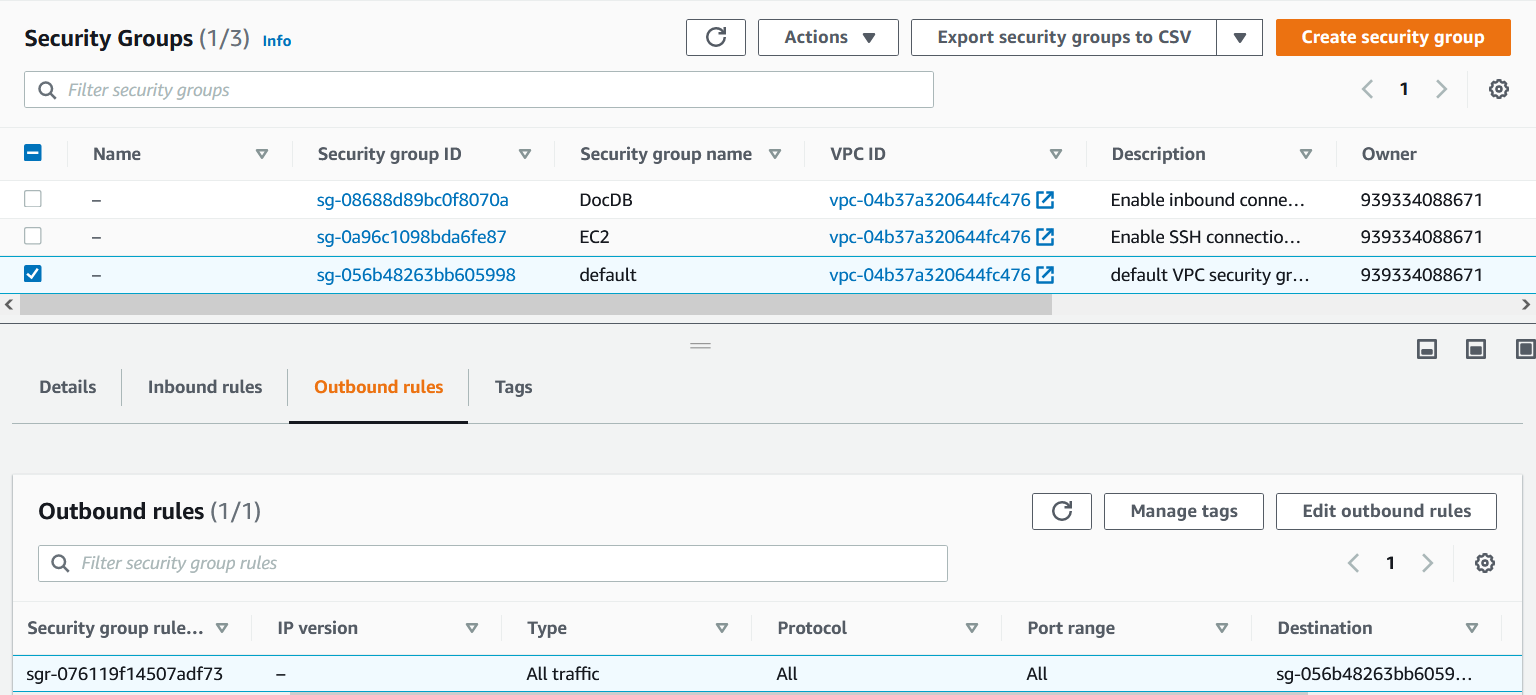
* Use the requests library to download the zip file from the URL.
* Use the zipfile module to extract the data from the zip file.
* Use the boto3 library to store the data in Amazon S3.
* Create a DocumentDB instance and connect it with an EC2 instance in the same VPC. Also, create a SageMaker instance to connect with DocumentDB.
* Load the data from S3 into Amazon DocumentDB (MongoDB).

**Services used**

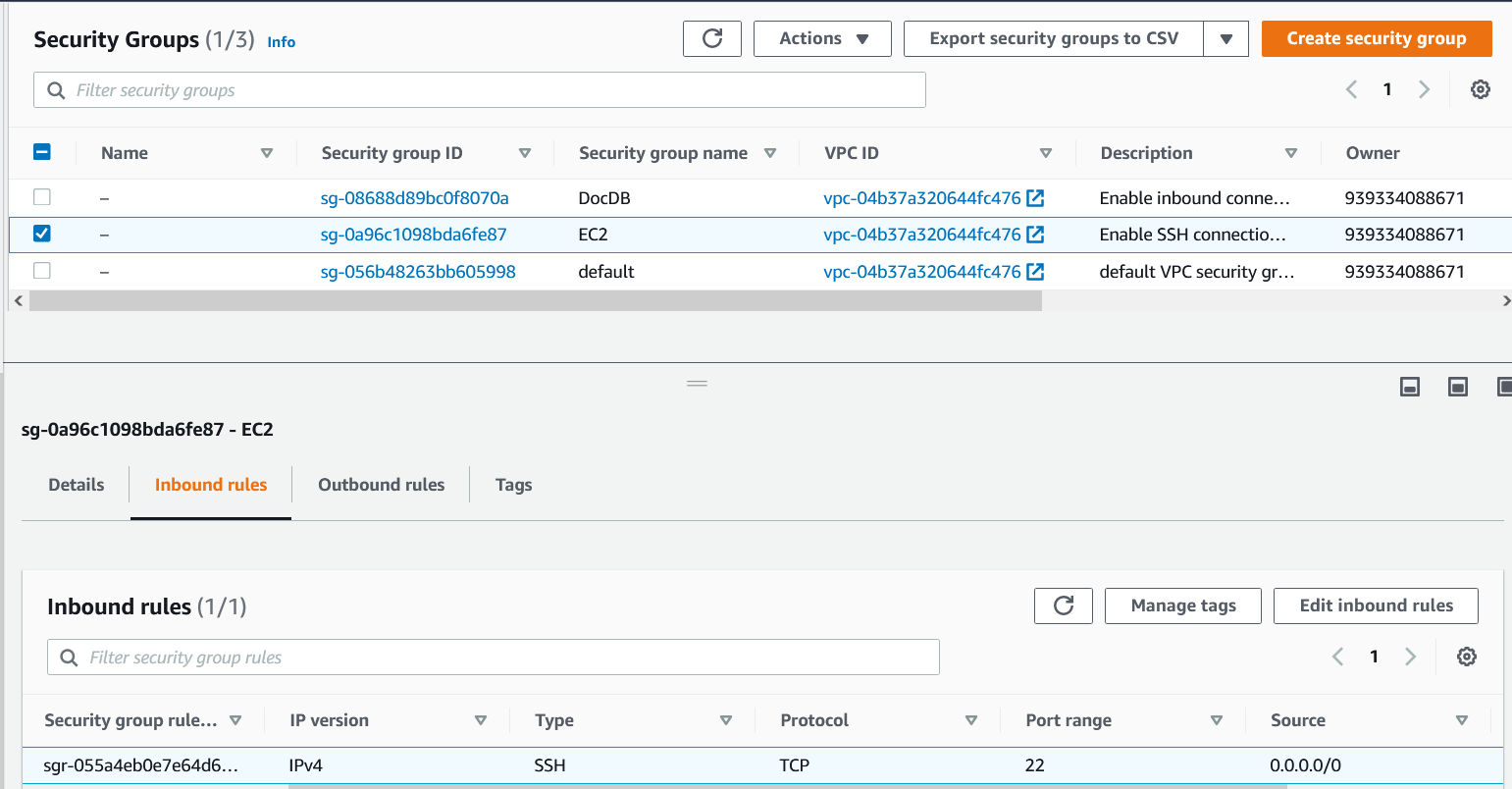
* Amazon S3: Amazon S3 is an object storage service that provides manufacturing scalability, data availability, security, and performance.
* AWS IAM: This is nothing but identity and access management which enables us to manage access to AWS services and resources securely.
* AWS DocumentDB: AWS DocumentDB is a fully managed, non-relational database service that is compatible with MongoDB workloads.
* AWS EC2: Enables you to launch and manage virtual machines, known as instances.
* Amazon SageMaker: An AWS SageMaker Notebook is essentially a Jupyter Notebook in which you can run Python code.

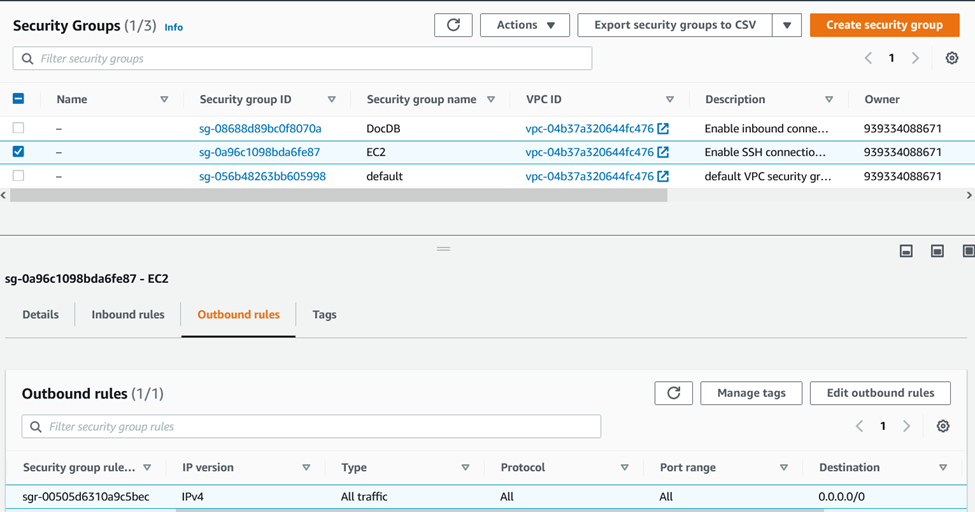
1.Default security groups of inbound and outbound rules in Ohio(us-east-2) region.



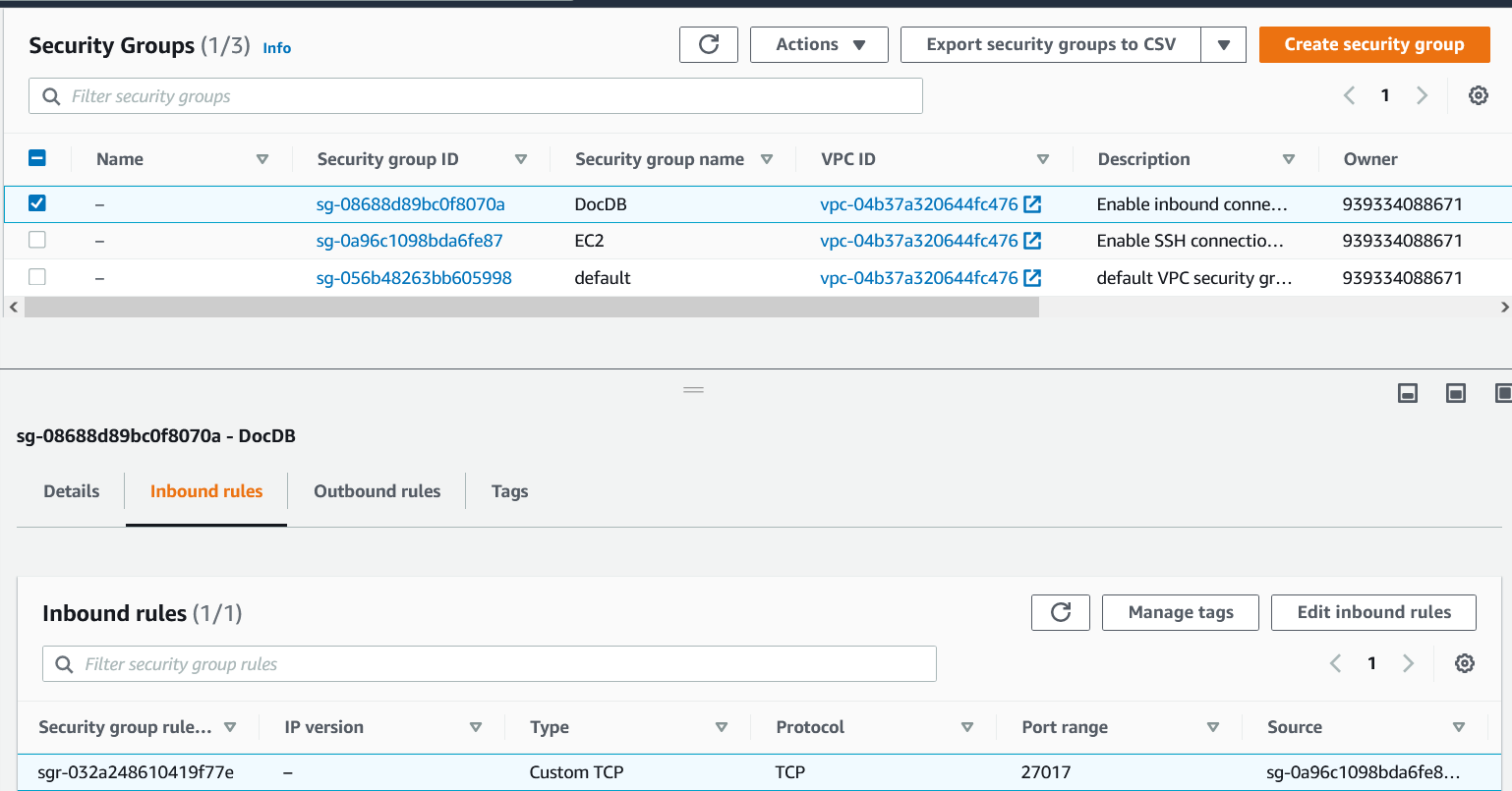


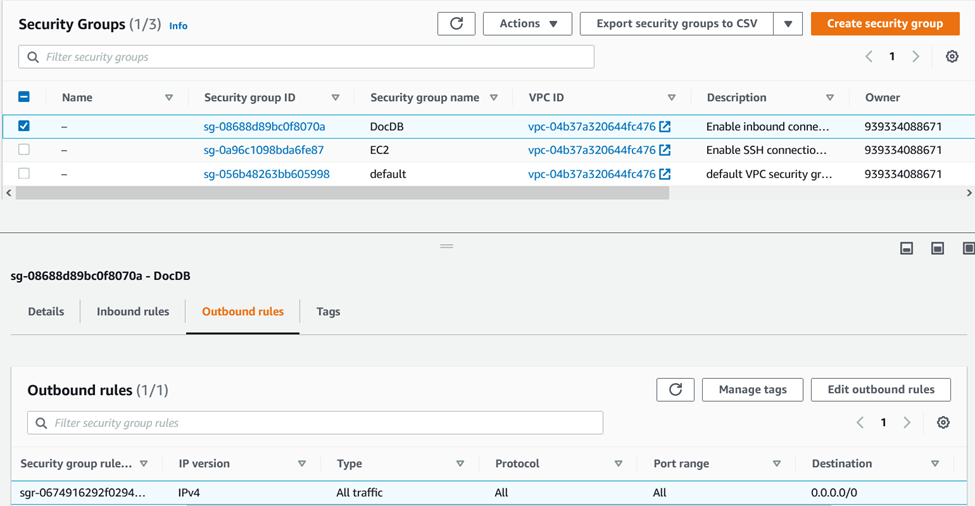
2.The first security group (EC2) allows you to SSH into your EC2 instance from your local machine (client).



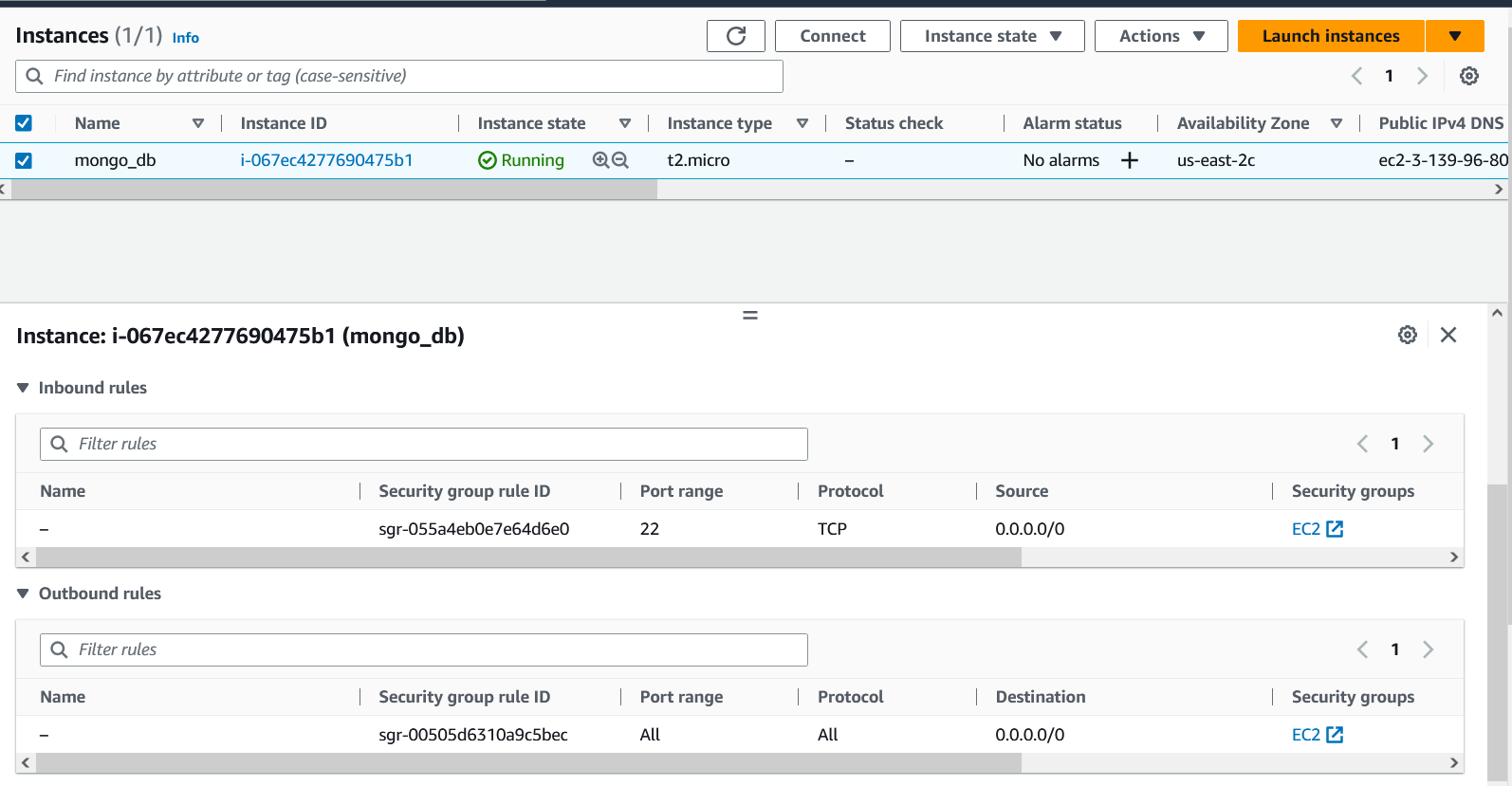


3. The second security group (DocDB) enables you to connect to your Amazon DocumentDB cluster on port 27017 (the default port for Amazon DocumentDB) from your EC2 instance.

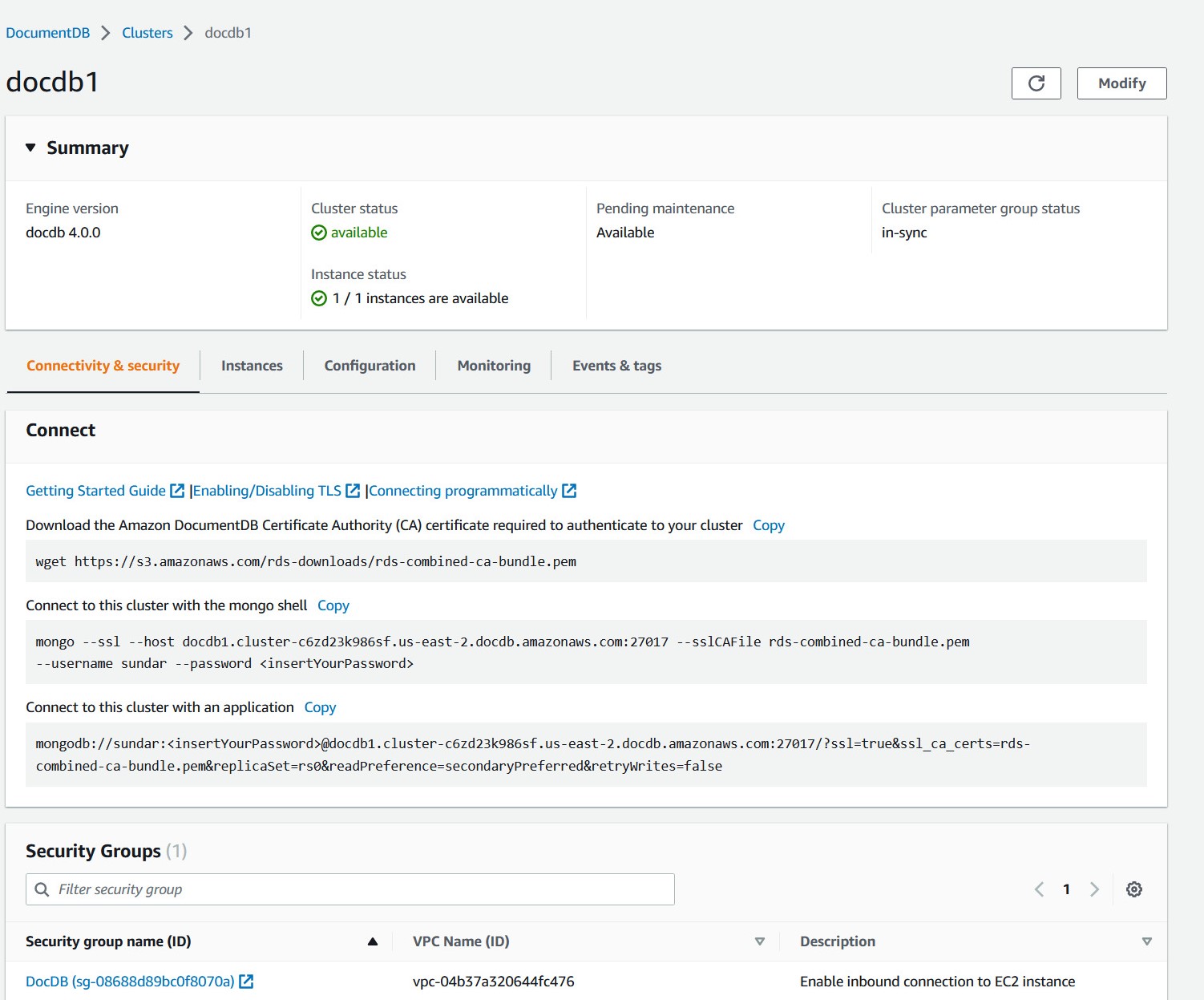




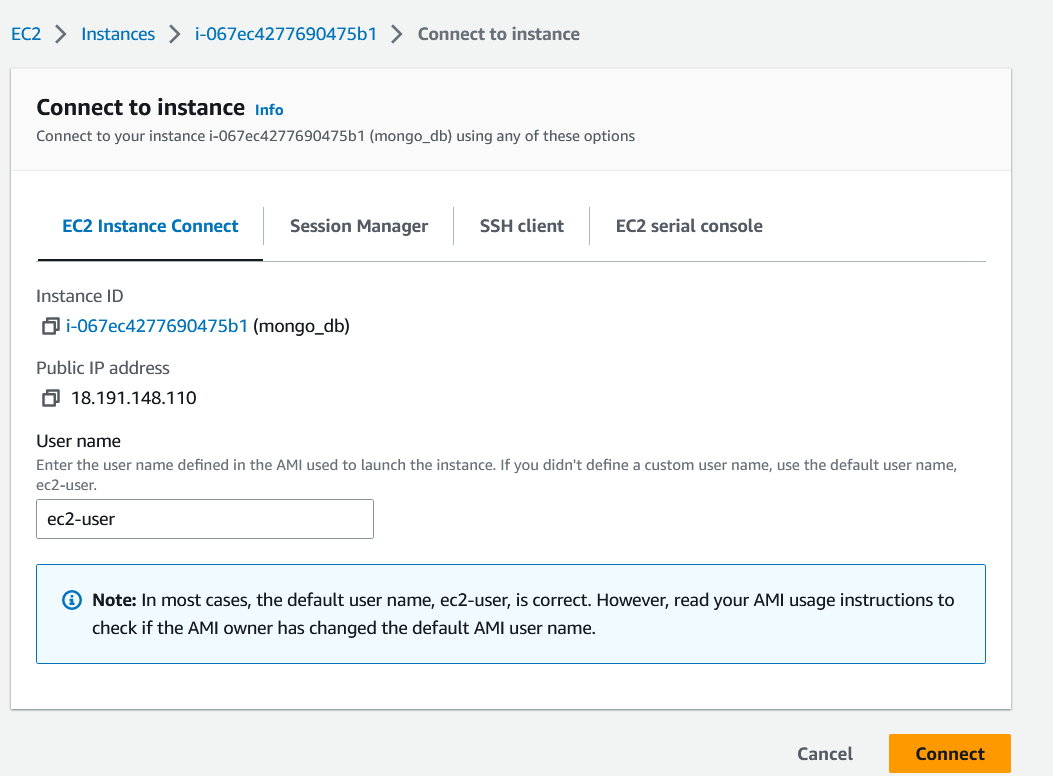
4.Create an EC2 instance in the same Region and VPC that you use to provision your Amazon DocumentDB cluster.

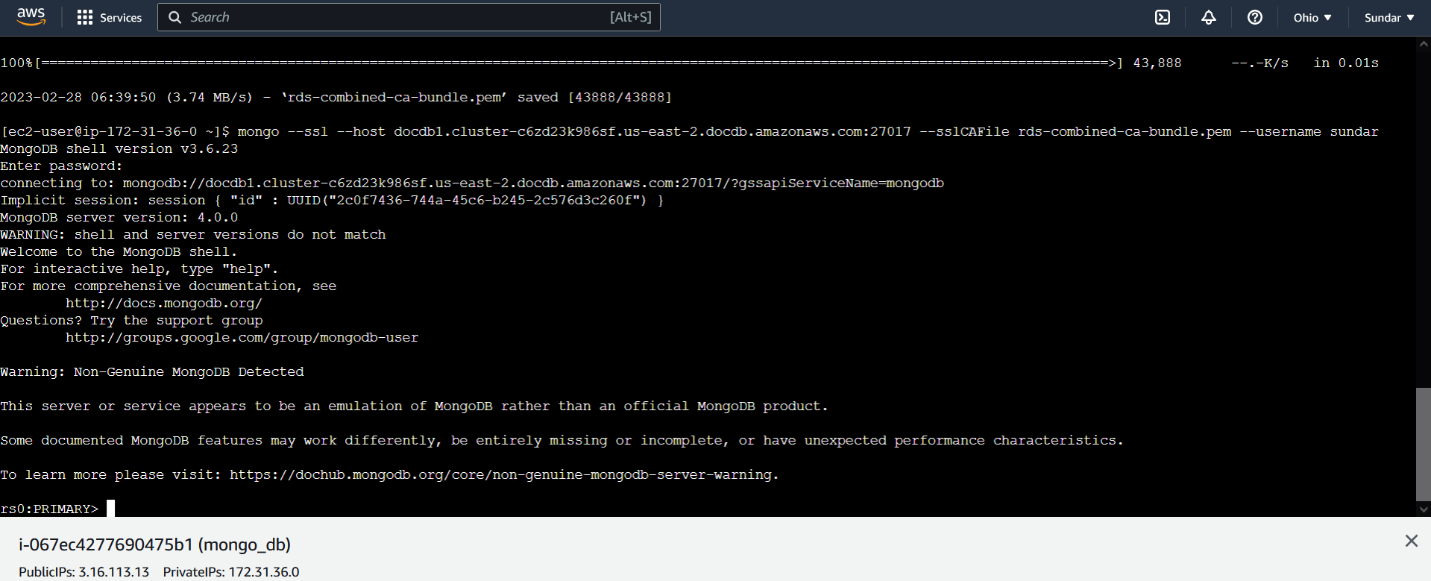


5. Create your Amazon DocumentDB cluster in your default VCP. Because Amazon DocumentDB is a VPC-only service.

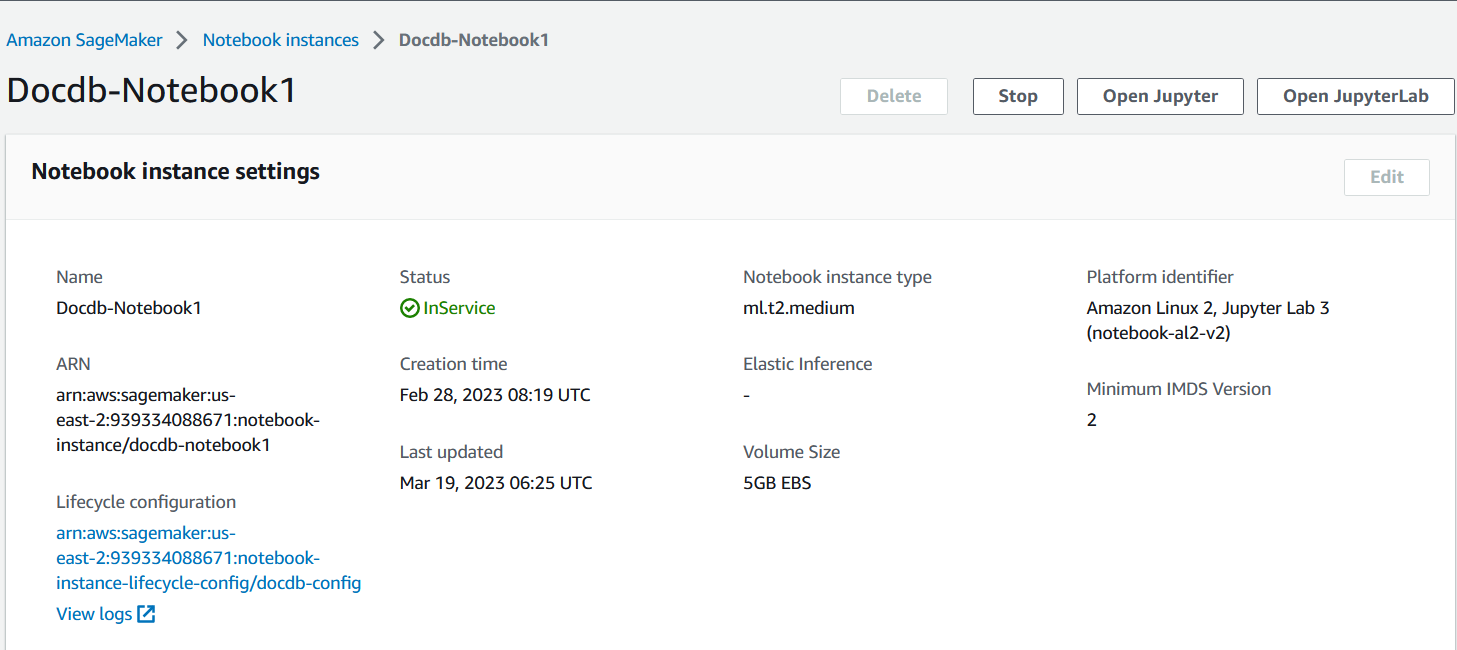


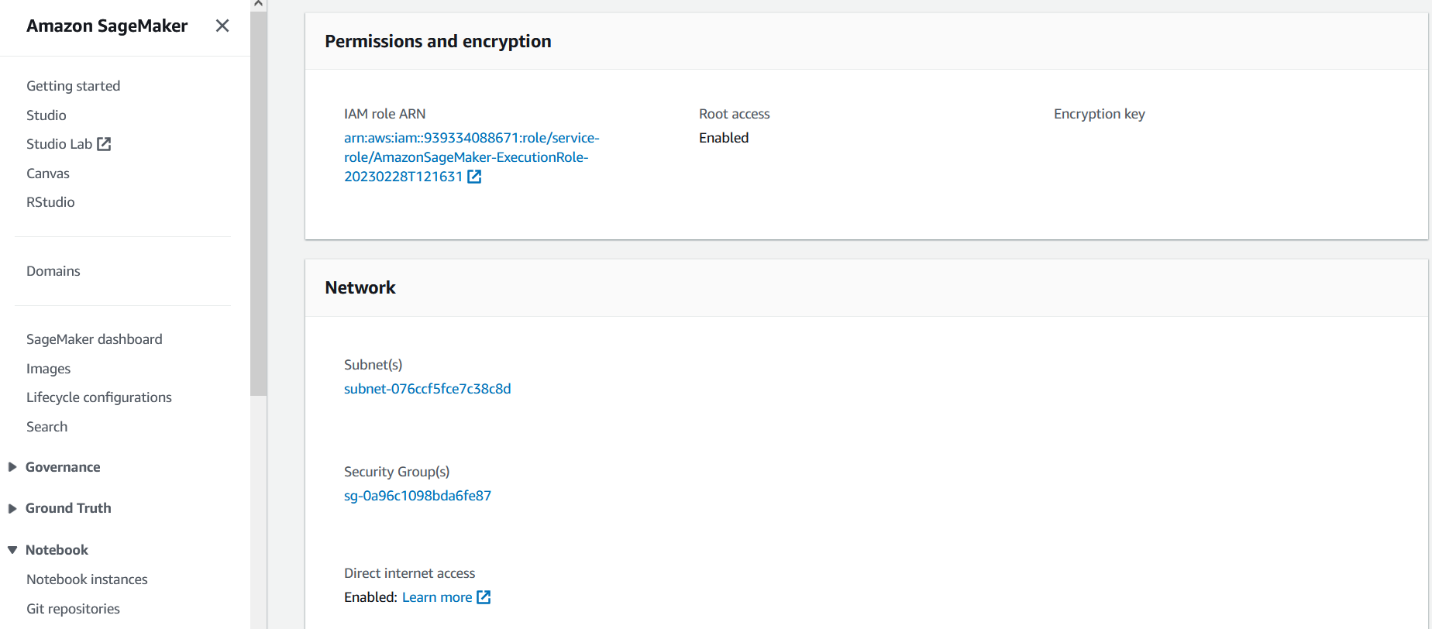
6. To connect to DocumentDB from an EC2 instance, I used the EC2 instance connect feature. Then, I installed the MongoDB shell on the EC2 instance to interact with the DocumentDB cluster.



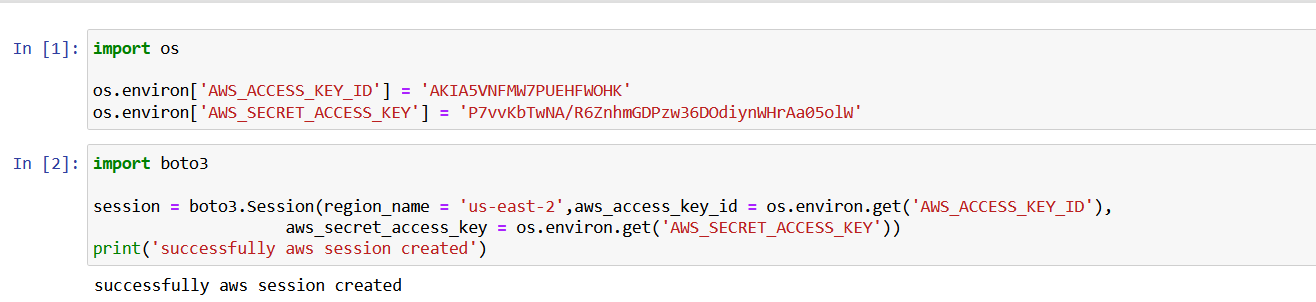


7. Creating an Amazon SageMaker instance within the same VPC as the DocumentDB cluster, using the same security groups as the EC2 instance.





8.Connecting to an AWS S3 session using boto3 library and creating a bucket using the python code.





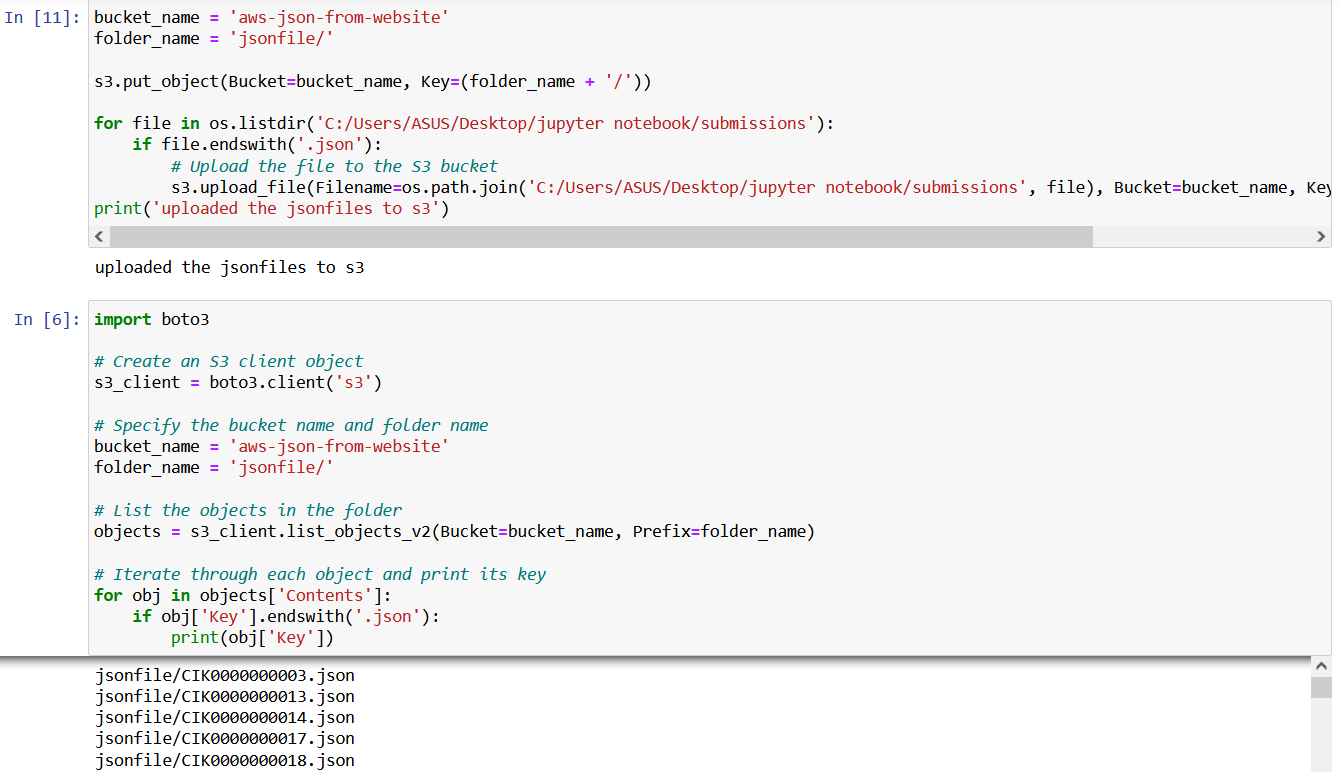
9.Created a python code to download a zipfile from URL.

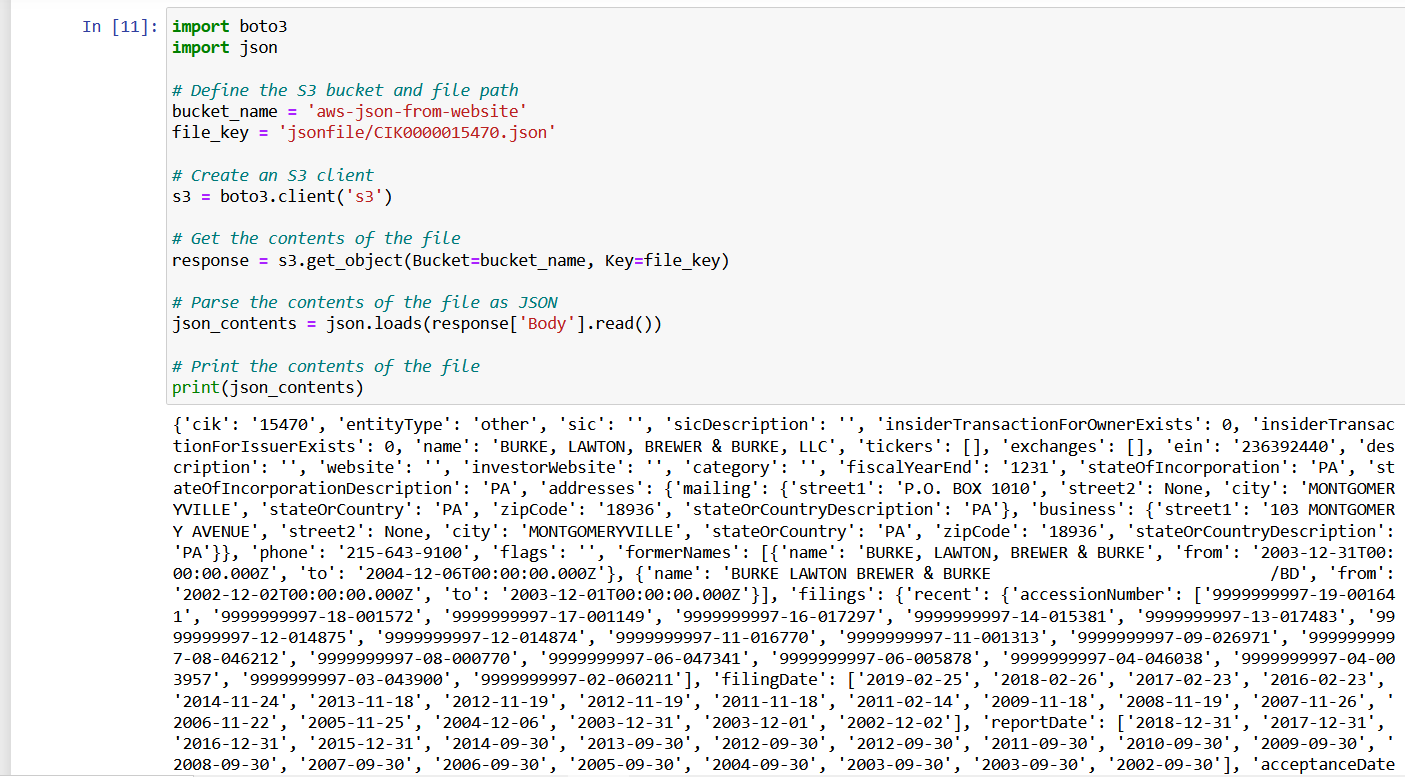


10. Unzipping the file locally and creating a folder in an S3 bucket.



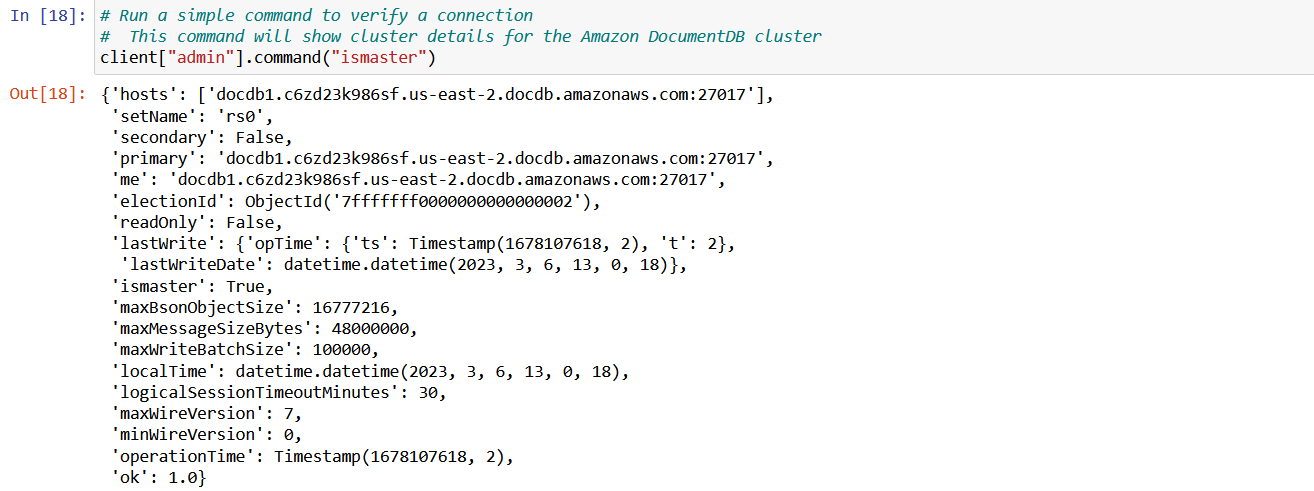
11.Uploading the json files to S3 bucket inside the folder.





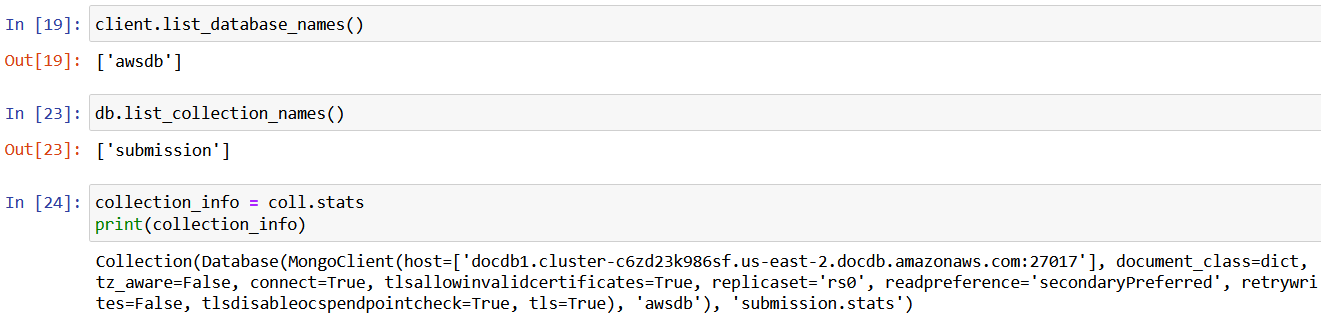
12. Creating a Python code for connection details and credentials for the Amazon DocumentDB to verify the successful connection.





13. Creating a database and collections for Amazon DocumentDB and inserting the JSON files from S3 bucket into the Amazon DocumentDB.





13.Querying the data which is present in the database.

