1. **Data Migration and Transformation Tool for Amazon RDS Data Warehouses**

**Overview**

Goal:

Extraction of the data from a zip file that is available at a URL and uploading it into Amazon S3 and Amazon RDS.

Technologies used:

Python, Requests library, Zipfile library, boto3 library, psycopg2 library pandas, Amazon S3, Amazon RDS

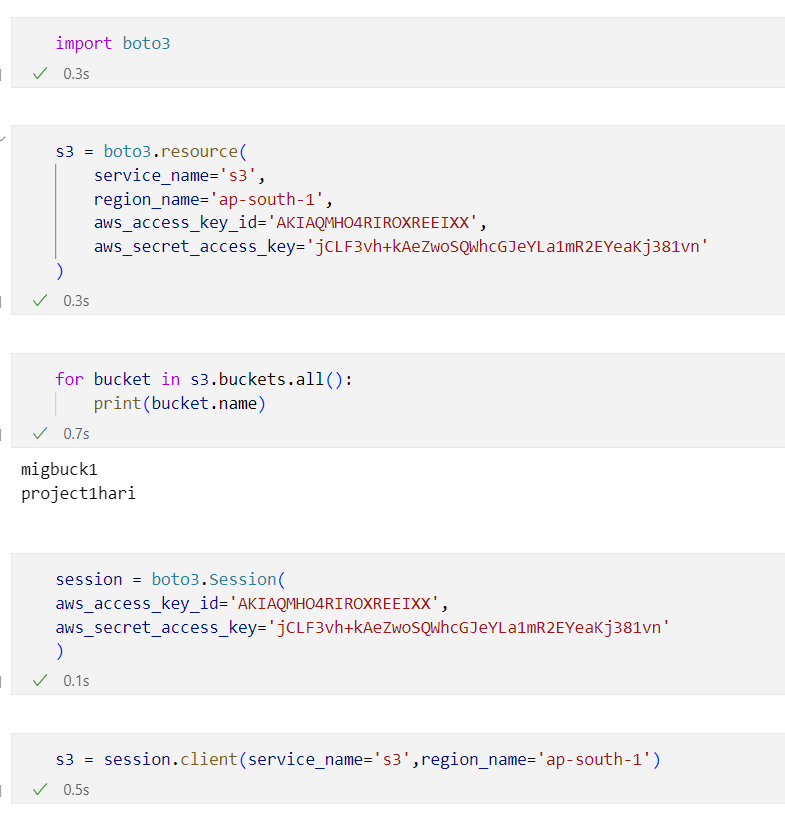
Steps:

1. Using Python libraries like [requests, urlib, wget, curl] to download the zip file from the URL.
2. Zipfilelibrary in Python to extract the data from the zip file. This will create a file or directory containing the extracted data.
3. AWS SDK for Python (Boto3) to create an S3 bucket.
4. Using put\_object function of Boto3 to upload the extracted data to the S3 bucket.
5. Boto3 is used to create an RDS instance.
6. Using psycopg2 library in Python to connect to the RDS instance.
7. Using pandas libraryin Python to read the extracted data into a DataFrame and this data is pushed to RDS instance.

The codes used and screenshots of the task done is given below.

**Codes used:**

Using boto3 a connection was established with the local system and amazon s3 services. After that a session and client were created in order to push the object from URL to s3 bucket

****

Using the requests library, the zip file was downloaded from the URL.

Using put\_object() the zipfile was uploaded to s3 and then using the zipfile module the zipfile was extracted into another folder inside s3.



**Issues faced while doing the project:**

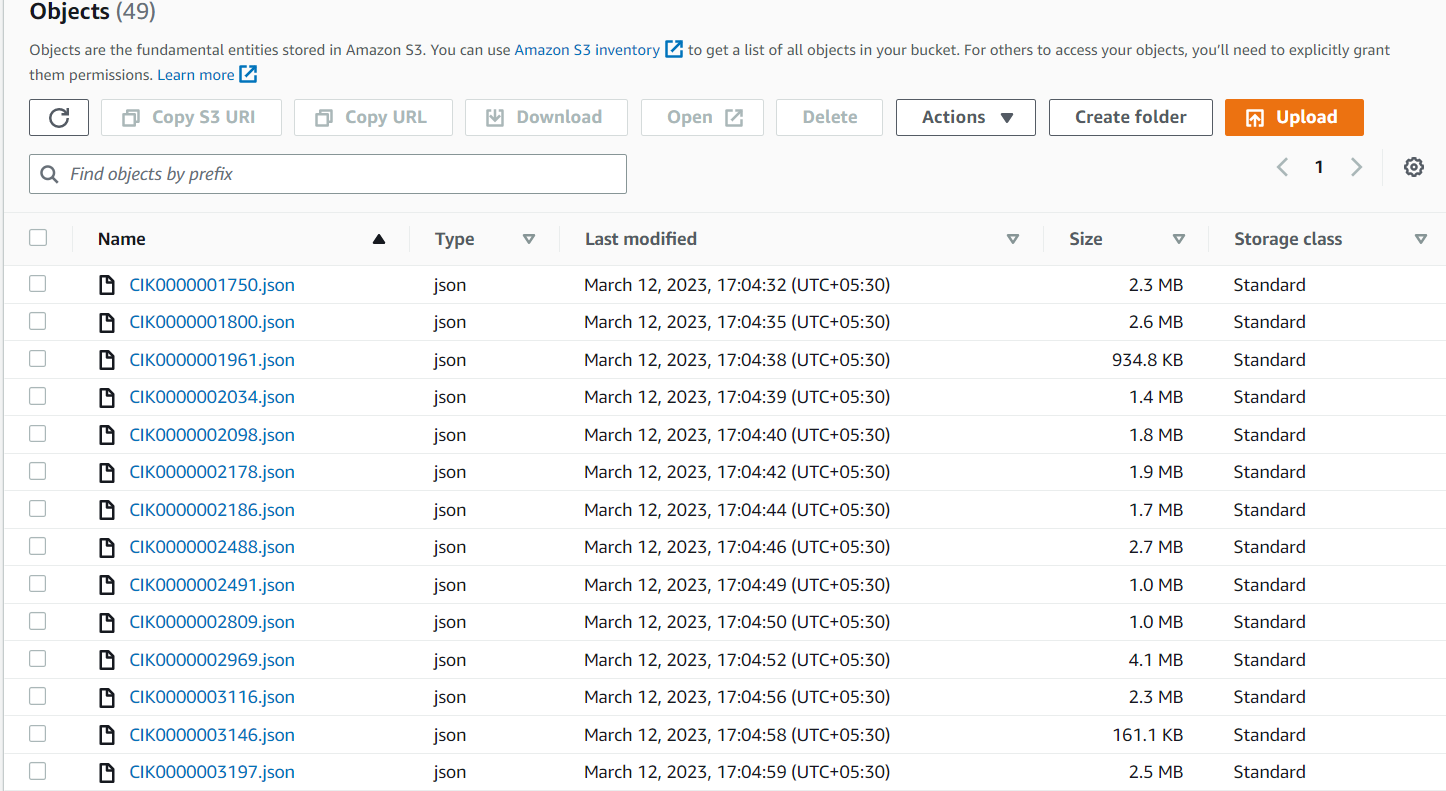
The zip file downloaded from the given URL was not accessible and it had badzip file error. More over the URL was giving a 403 forbidden error. Hence the zip file was downloaded manually to the local system and only 50 files were uploaded into the s3 from the local system to check the code and complete the project.

****

From the local system using Boto3, fifty files were uploaded to S3 using the code below

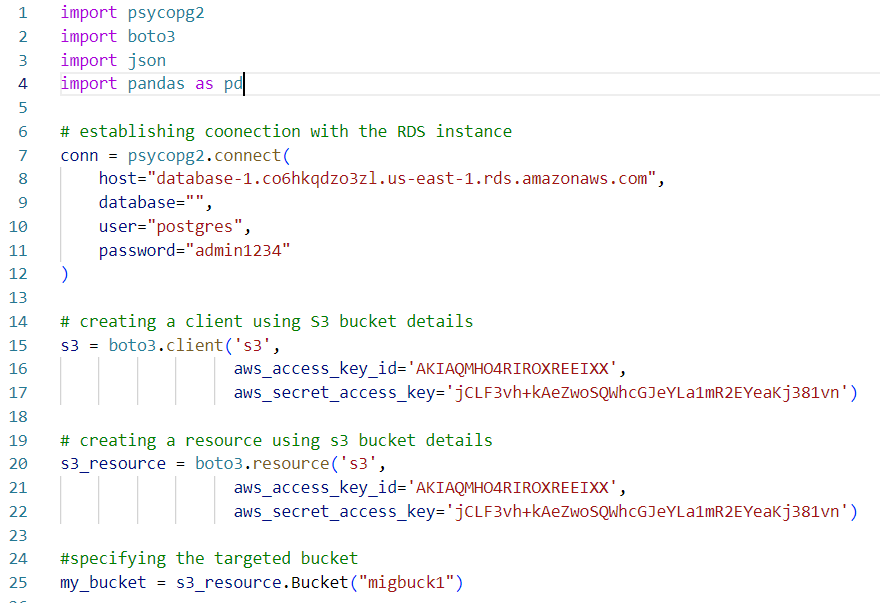


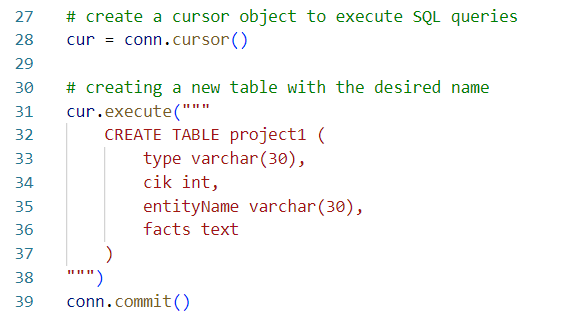
Result for the above code

****

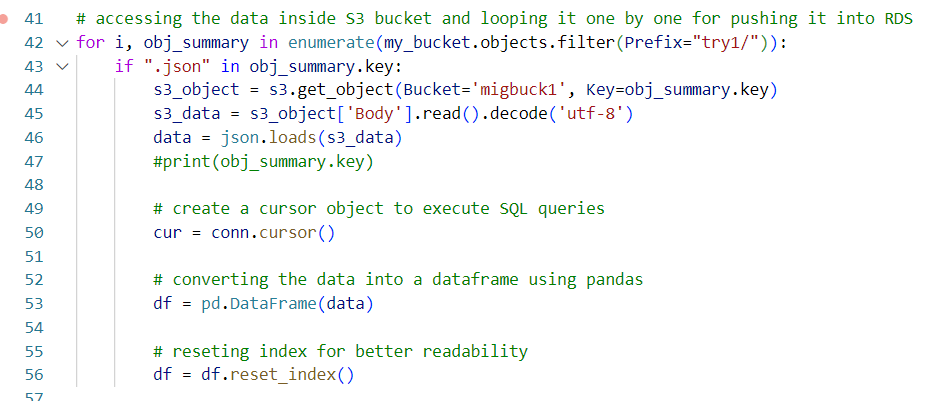
Pushing the data from S3 to RDS instance using the following code

Here using the below code, we are establishing a connection between RDS and S3 using psycopg2 and boto3 libraries.

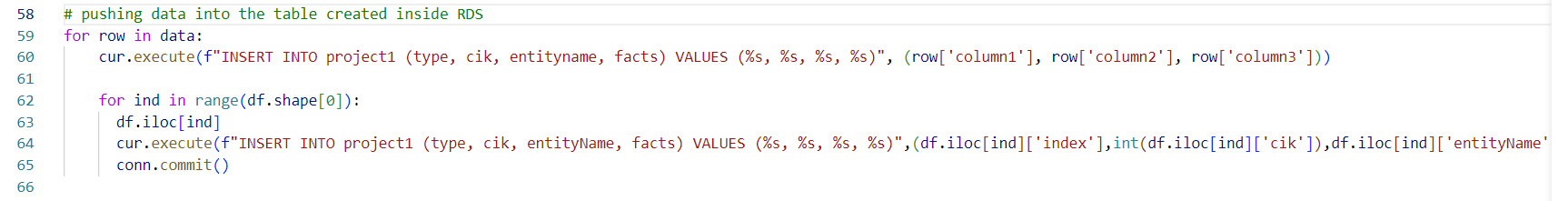


After that we are creating a table of our desire inside RDS for accommodating the data.

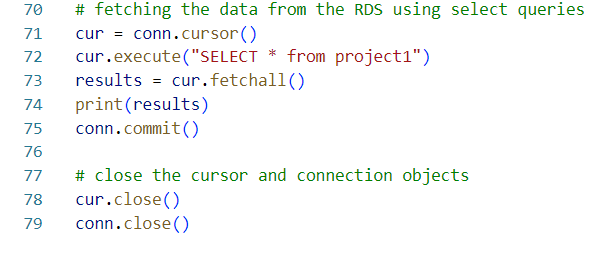
From S3 we are accessing the data in order to push it inside the RDS



Here we are pushing data inside the table created in the RDS using for loop

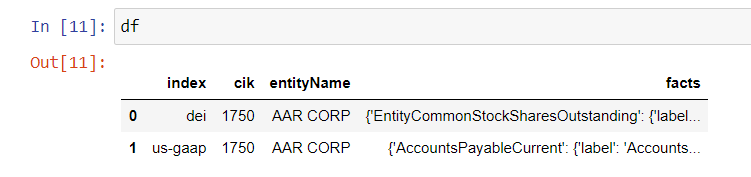


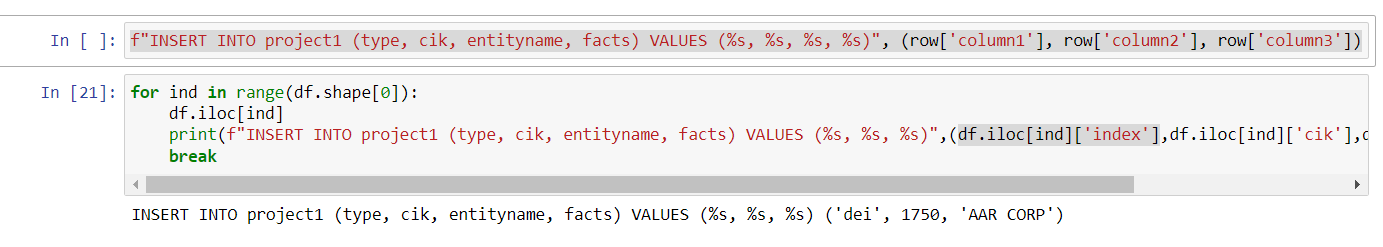
Using select query, we are able to access the data stored inside the RDS

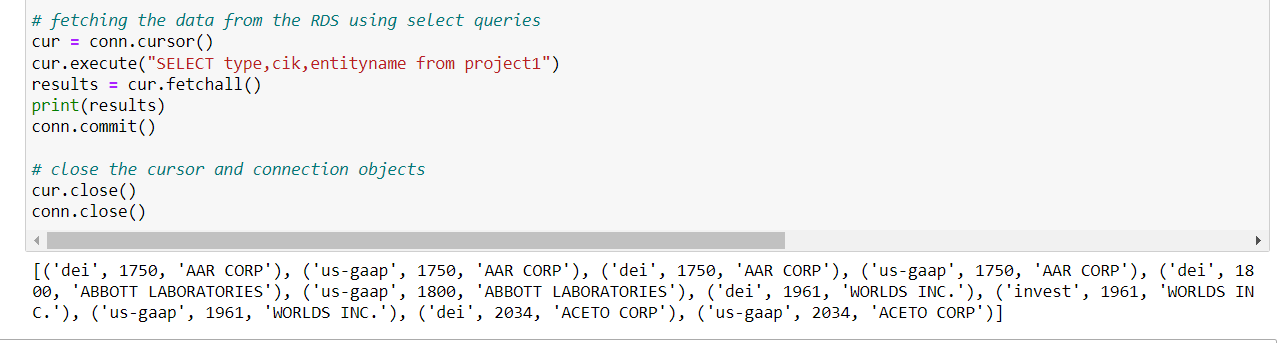


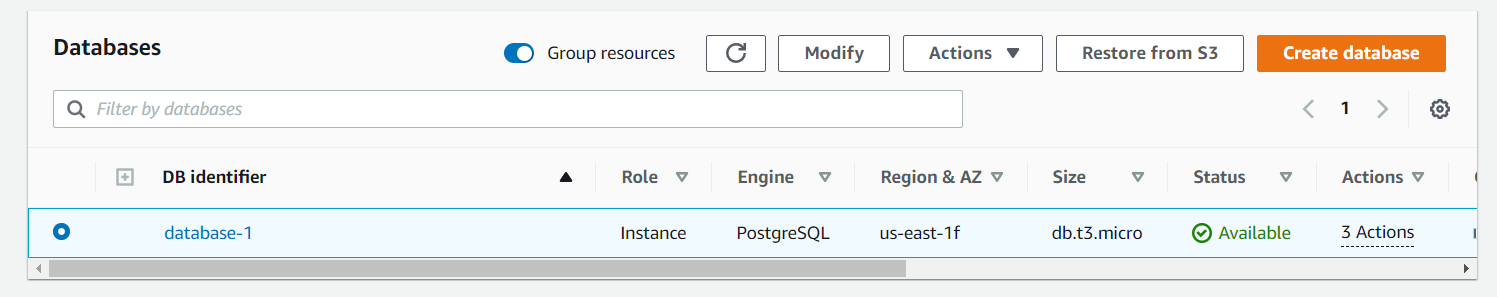
**Results:**

Converting the data into DataFrames

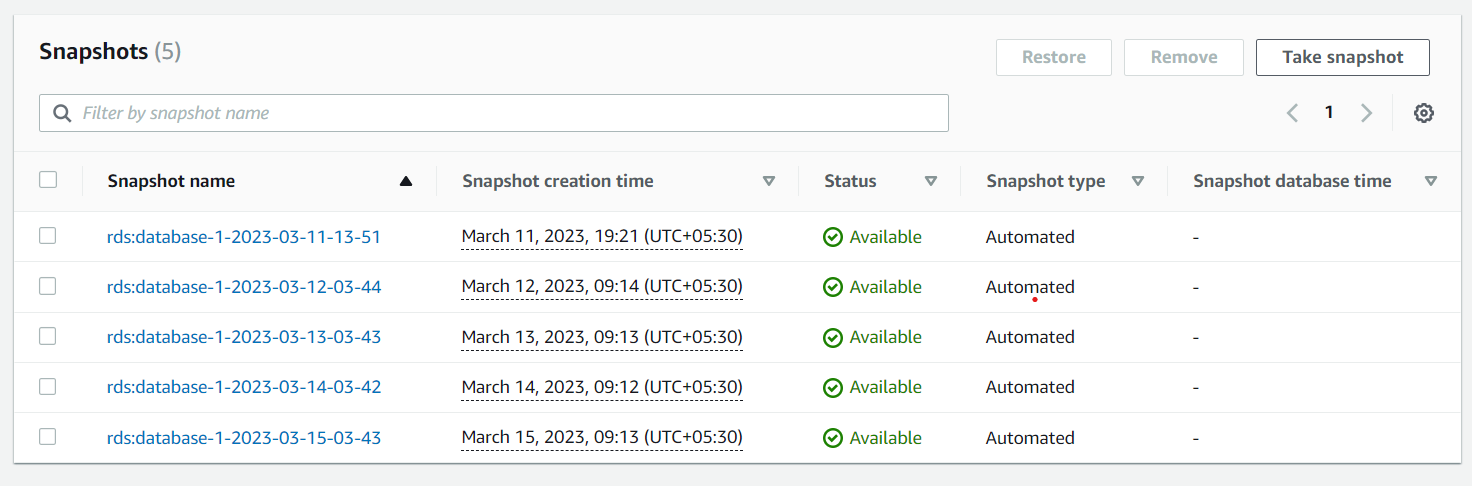
****

Looping the data into the table created inside RDS

Fetching the data from the RDS using select query

****RDS database that was created

Snapshots of the RDS database created

****