Project-3

Problem Statement: You have a URL that points to a zip file. The zip file contains multiple JSON files. The JSON files contain multiple documents with various data structures. Your 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 RDS (NoSQL). You want to use Python or PySpark to perform these tasks. You may use any libraries or tools that are necessary to complete the task. Approach: To extract the data from a zip file that is available at a URL and load it into Amazon S3 and Amazon RDS (NoSQL), you can follow these steps: 1. Use the requests library to download the zip file from the URL. 2. Use the zipfile module to extract the data from the zip file. 3. Use the boto3 library or PySpark to store the data in Amazon S3. 4. Use the pandas library and sqlalchemy or PySpark to load the data from S3 into Amazon RDS (NoSQL).

SOLUTION:

1.Downloadin the zip from url to s3





2.unipping the zip file from s3 bucket



3.Read the each json files from the unzipped files and store it in Mongodb/documentdb

