| **Project Title** | **Data Migration and Transformation Tool for Amazon RDS Data Warehouses** |
| --- | --- |
| **Technologies** | **Python/PySpark,Requests,Zipfile,boto3,pandas,sqlalchemy, Amazon S3,Amazon RDS** |

**Problem Statement:**

You have a [URL](https://www.sec.gov/edgar/sec-api-documentation) 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. 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).

**Results:**

The result of following these steps should be that the data from the zip file is extracted and stored in a list of dictionaries (if you are using Python) or a DataFrame (if you are using PySpark). Each dictionary or DataFrame row will represent a document from one of the JSON files in the zip file.

The data in the list or DataFrame will then be stored in Amazon S3 as JSON files. You will be able to access these JSON files using the boto3 library or the Amazon S3 web interface.

The data from the JSON files will also be loaded into Amazon RDS (NoSQL). You will be able to access the data in RDS using SQL queries. The data will be stored in a table in RDS, and the schema of the table will be determined by the structure of the JSON documents.

I hope this helps. Let me know if you have any further questions or need more assistance.