**Data Pipeline Assignment**

**1. Python Extraction Code**

The following script was deployed to AWS Glue to extract data from a public API and store it in an S3 bucket.  
  
**import requests**

**import boto3**

**import json**

**import io**

**API\_URL = "https://jsonplaceholder.typicode.com/users"**

**BUCKET\_NAME = "vishal-data-pipeline-assignment"**

**FILE\_KEY = "raw\_data/users.json"**

**print("Starting data extraction...")**

**try:**

**response = requests.get(API\_URL)**

**response.raise\_for\_status()**

**data = response.json()**

**print(f"Successfully extracted {len(data)} records from API.")**

**json\_data = json.dumps(data, indent=4)**

**json\_buffer = io.BytesIO(json\_data.encode('utf-8'))**

**s3\_client = boto3.client('s3')**

**s3\_client.upload\_fileobj(**

**json\_buffer,**

**BUCKET\_NAME,**

**FILE\_KEY**

**)**

**print(f"Successfully stored data in S3 at s3://{BUCKET\_NAME}/{FILE\_KEY}")**

**except requests.exceptions.RequestException as e:**

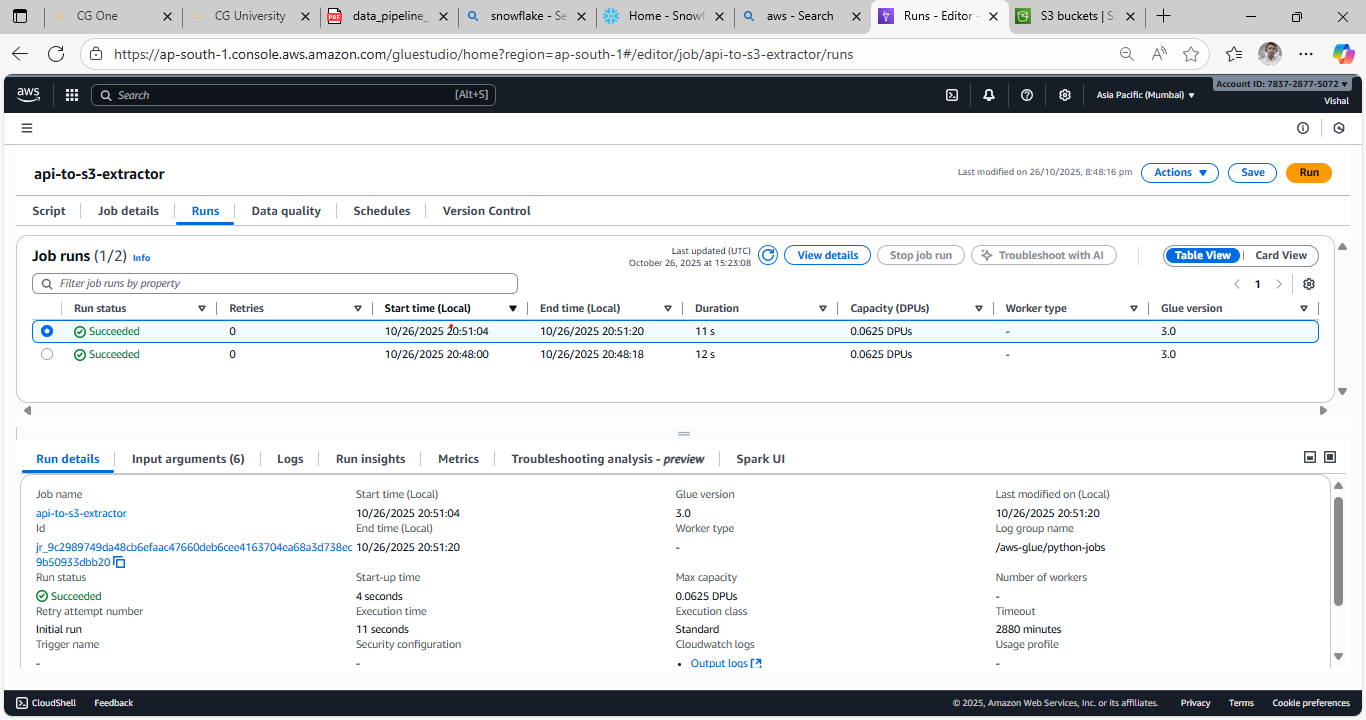
**print(f"Error during API request: {e}")**

**except Exception as e:**

**print(f"An error occurred: {e}")**

**2. Screenshot: Python Code Running on the Cloud**

This image shows the AWS Glue job api-to-s3-extractor with a "Succeeded" status.



**3. Screenshot: File Stored in Cloud Storage**

This image shows the users.json file stored in the raw\_data folder of the S3 bucket.

A screenshot of a computer

AI-generated content may be incorrect.

**4. Screenshot: Snowflake Raw Data Table**

This image shows the result of querying the raw data table (RAW\_USERS), with each row containing the JSON for a single user.

Query: SELECT \* FROM RAW\_USERS LIMIT 10;

A screenshot of a computer

AI-generated content may be incorrect.

**5. Screenshot: Snowflake Transformed Data View**

This image shows the result of querying the final view (VW\_USERS), which parses the JSON into clean, separate columns.

Query: SELECT \* FROM VW\_USERS LIMIT 10;

A screenshot of a computer

AI-generated content may be incorrect.