Project Deliverable 2

Teammates:

Nishanth Dangethi

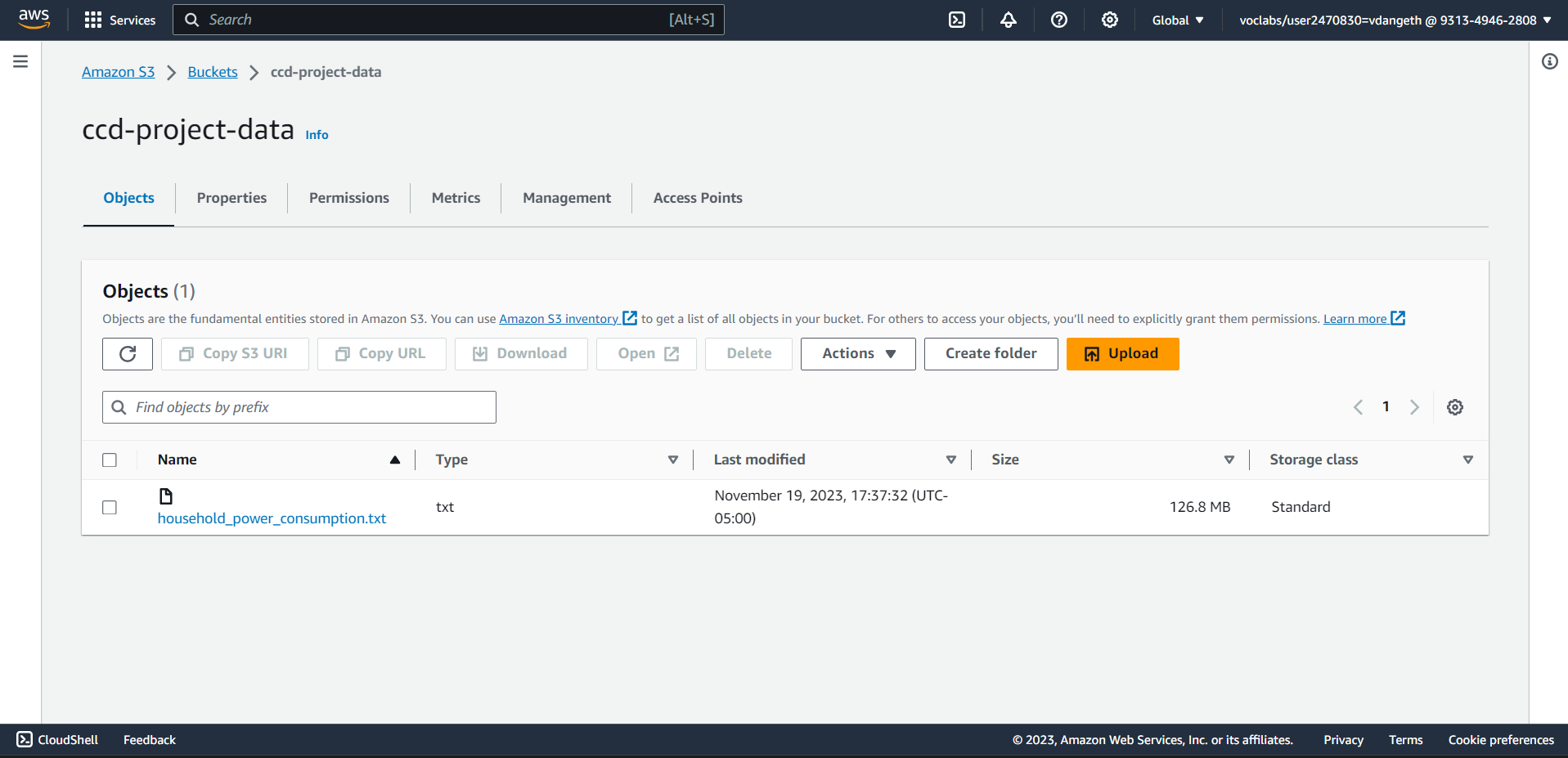
Ujwala Mallela

Sushant Sagar Saka

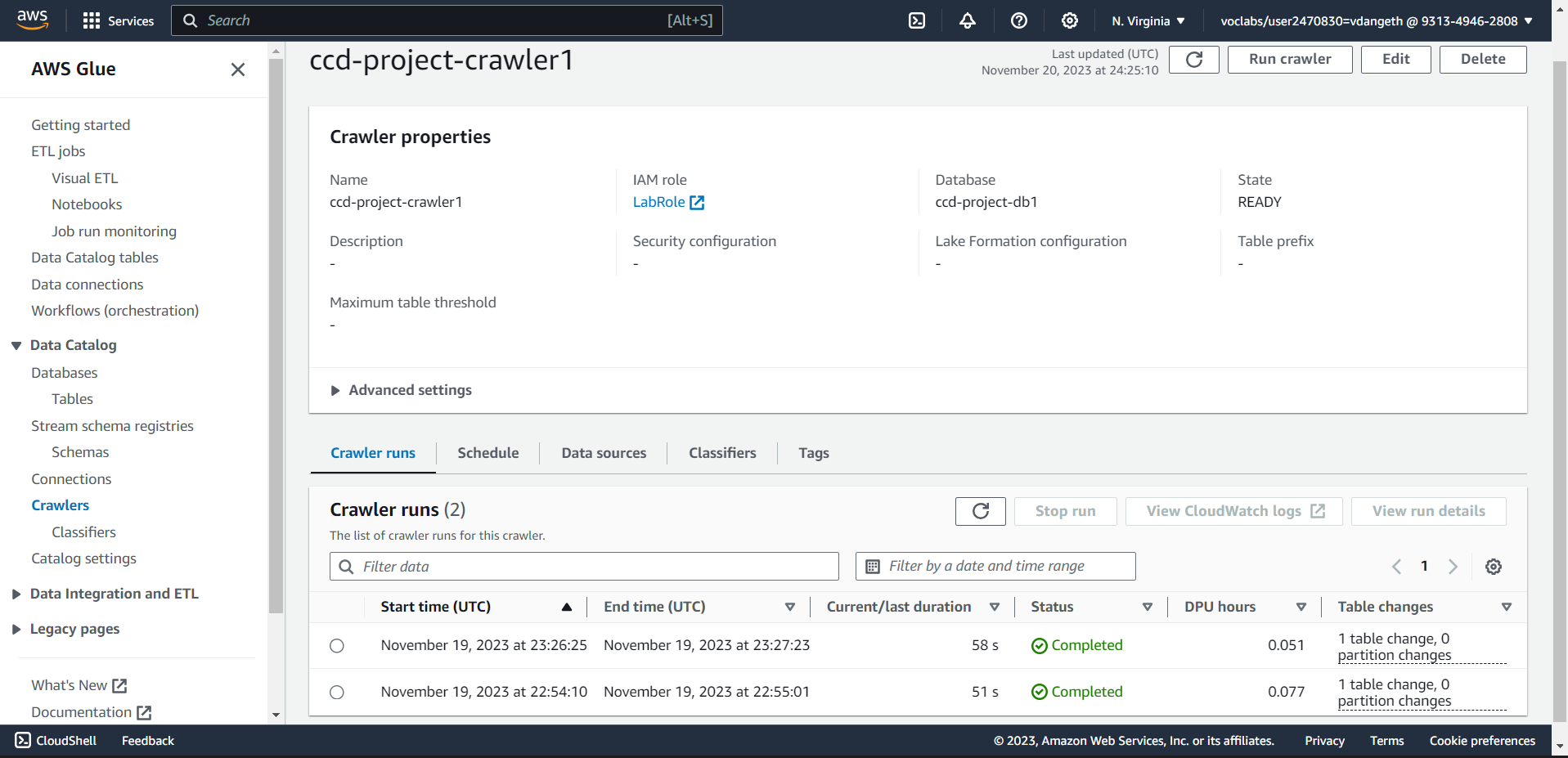
Jaideep Gurrapu

Pavani Venkata Satya Mahathi Lanka

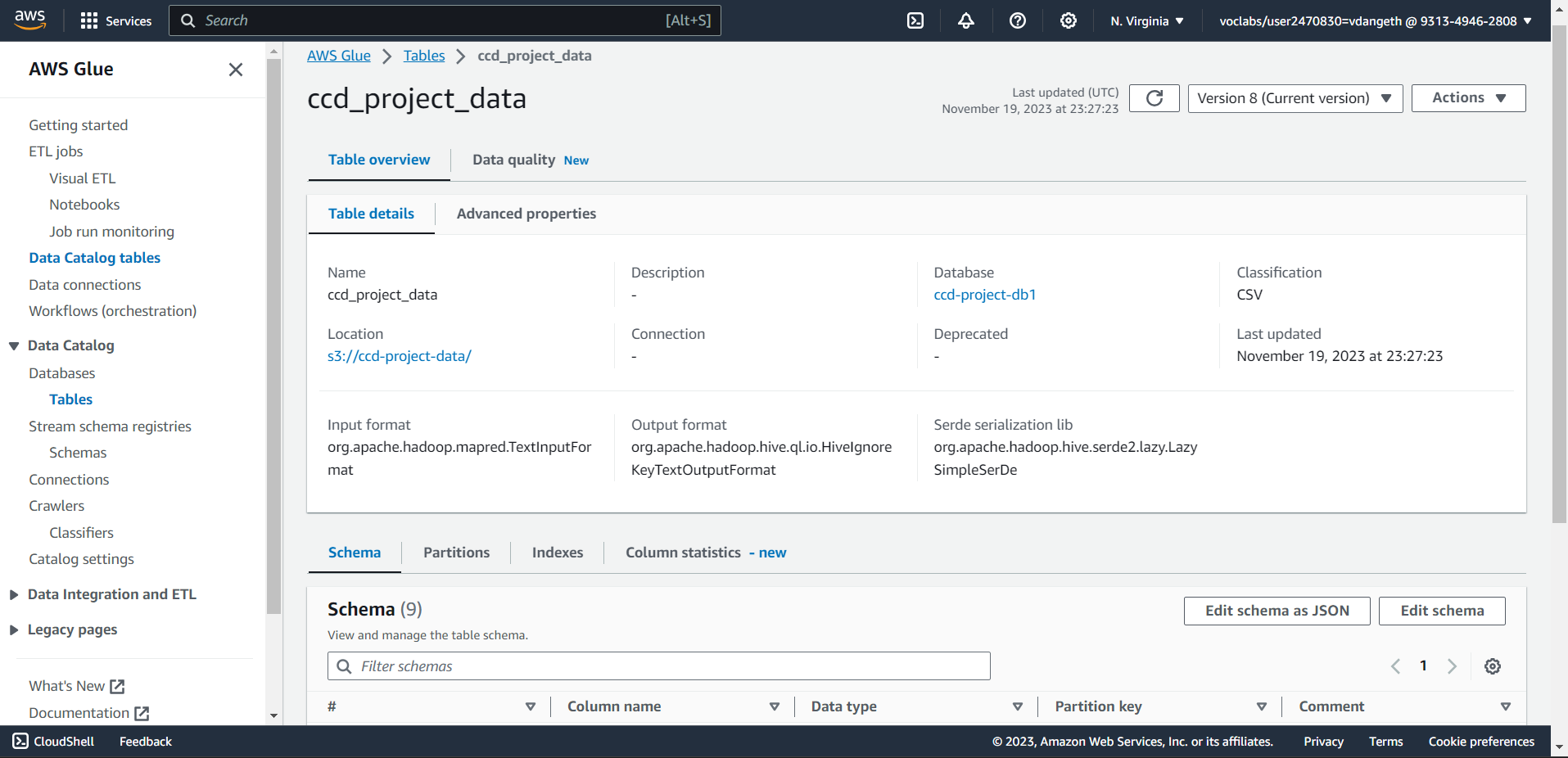
S3 bucket creation



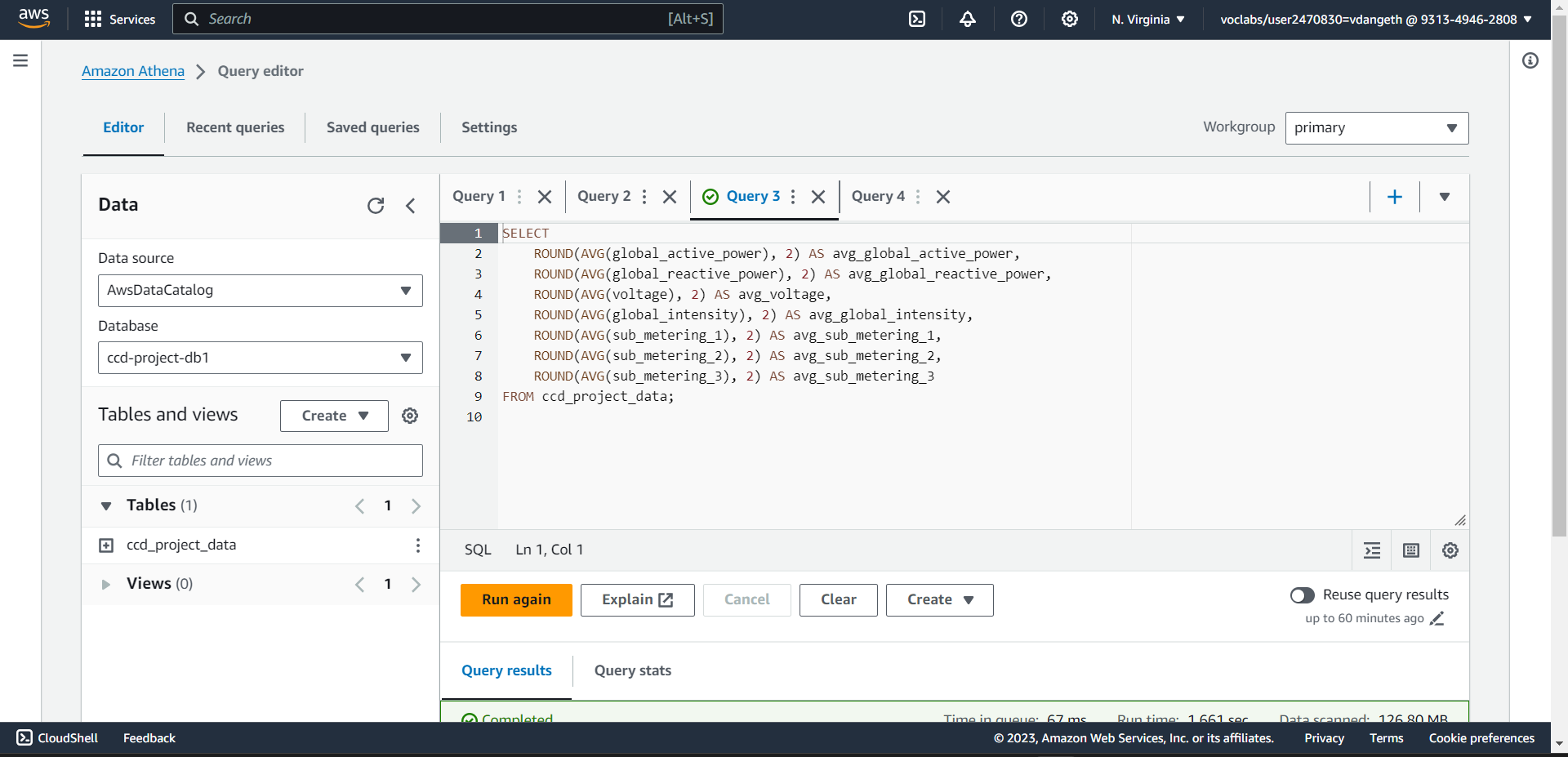
Creation of Crawler using Glue

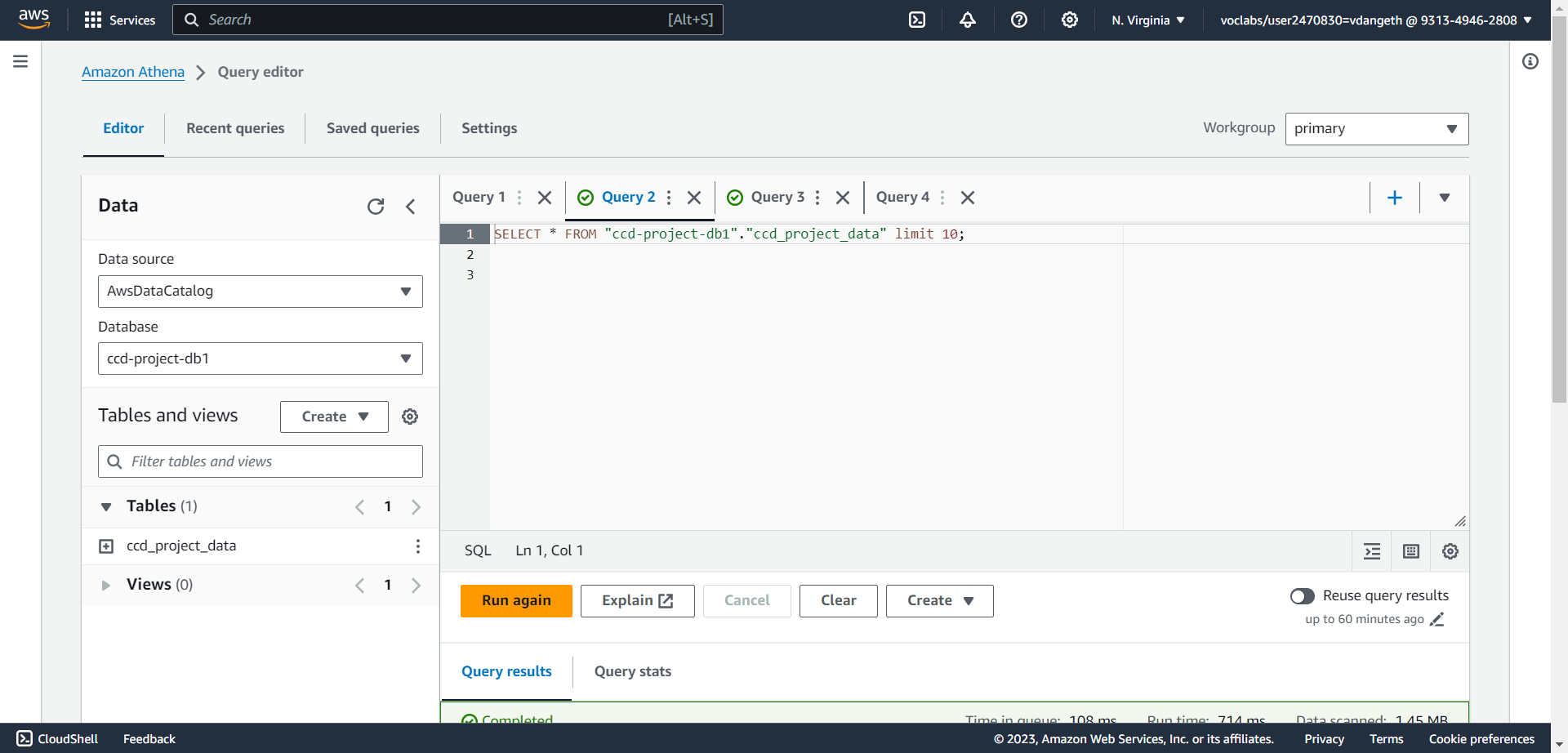


Data Catalog tables in Glue

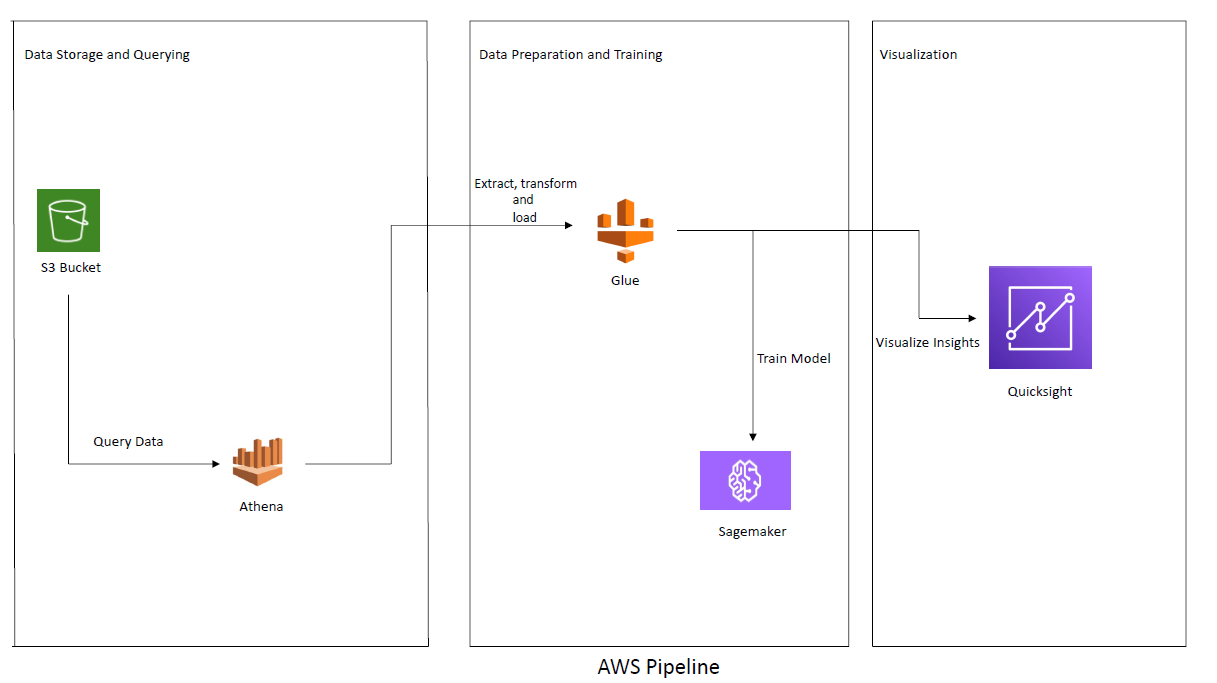


Sample Queries in Athena





Pipeline Diagram/Project Flow Diagram



**AWS Services Utilized till present**

**S3 (Simple Storage Service):**

**Usage:** Utilized as a central storage repository for project datasets, model artifacts, and other resources.

**Implementation:** Uploaded project datasets and related files to S3 buckets, ensuring secure and scalable data storage accessible across various services.

**Athena:**

**Usage:** Facilitates SQL-based querying directly on S3-stored data without the need for complex data warehousing.

**Implementation:** Defined dataset schemas, created corresponding tables in Athena, and executed SQL queries for exploratory analysis and data insights.

**Glue:**

**Usage:** Automates ETL (Extract, Transform, Load) processes to prepare data for machine learning tasks.

**Implementation:** Implemented Glue jobs to extract, transform, and load data from S3 into suitable formats for ML model training and analysis.

These AWS services - S3 for storage, Athena for SQL-based querying, and Glue for ETL—were integrated into the project workflow to ensure efficient data management, exploration, and preparation for machine learning tasks within the AWS ecosystem.