



Project 1: Using Amazon Redshift and Simple Storage Service

BY: BERNARD SUMTER AND PRIYA JOSHI

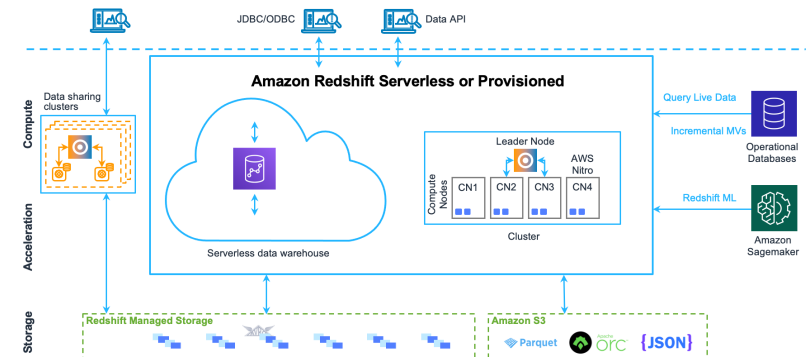
What is a Bucket and how will we be using it?

- ▶ A virtual container within Amazon's simple storage service
- ▶ **Centralized Data repository:** S3 is a scalable storage solution for large amounts of raw, unstructured, and structure data.
- ▶ Data we will be uploading: A CSV file

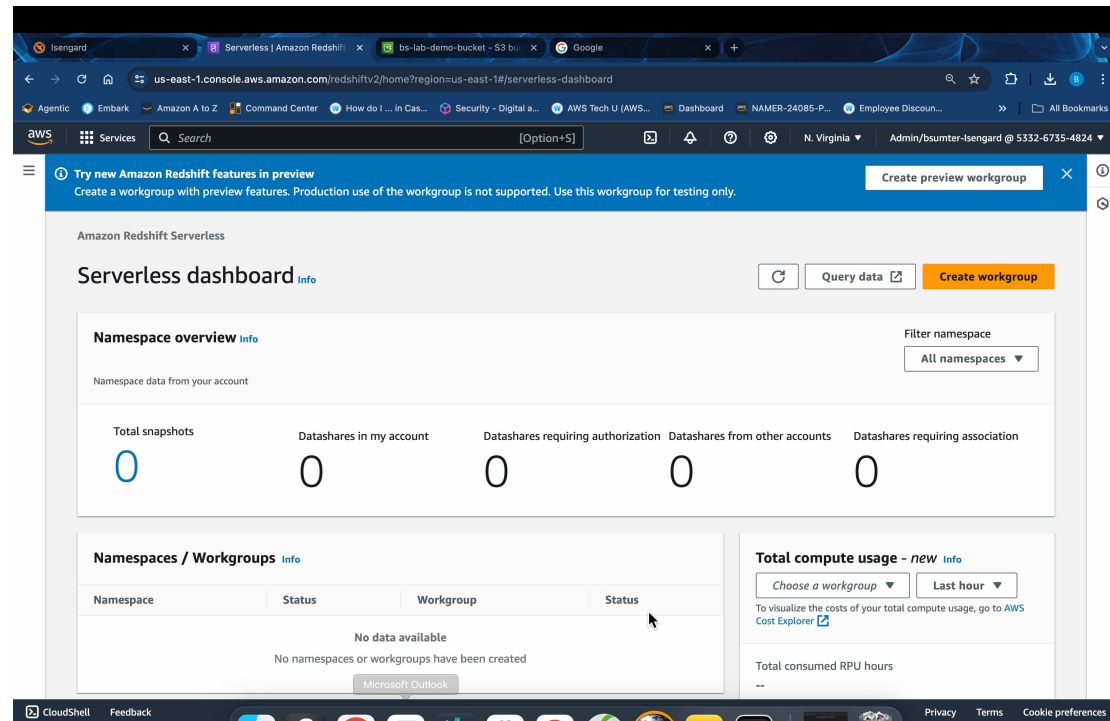


What is AWS Redshift and how will we be using it?

- ▶ Amazon Redshift is a cloud-based database storage that can be used for business that need to store and analyze large datasets.
- ▶ We will be using
 - AWS Redshift Sever-Less Cluster: automated cluster that scale up and scale down based workload of data
- ▶ Using Redshift sever-less cluster will allow us to load and analyze our large dataset that we stored in our S3 bucket.



Demo Video



How did we use these services together?

- ▶ Import a large CSV file into our redshift data warehouse to effectively manage and analyze the data from our CSV file
- ▶ Use cases:
 - Data analysis: loading archived data from S3 into Redshift when needed for historical analysis, reporting, ECT.
 - Data backup/ Disaster recovery: Using Snapshots stored in S3 to restore redshift cluster to a specific point in time in case of data loss.
 - Orchestrating ETL workflows by using services such as Glue with redshift for data integration and warehousing.
 - Redshift can be used to create predictive models