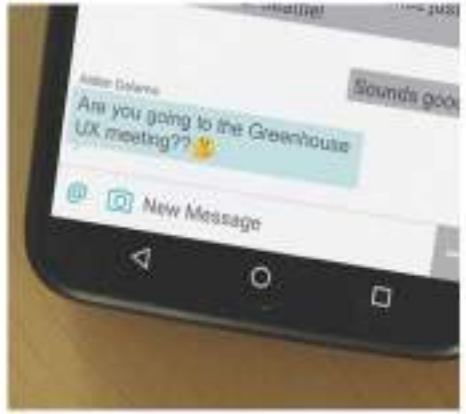


Cloud Foundations

GURURAJ, JASON, LUIS

aws

Presentation Link

https://broadcast.amazon.com/videos/1163159

Scenario

As you know, Washington is looking for solutions to maintain the road infrastructure as more people switch to electric vehicles. The gas tax, which traditionally funded road maintenance, is declining. The proposed road charge system aims to address this by charging drivers per mile driven.

To implement the road charge system effectively, we need a clear picture of the electric vehicle landscape in Washington. We need to know where EVs are concentrated, how fast they're growing in different regions, and how this might change in the future. That's exactly where this project comes in. We've created visualizations that show the current distribution of electric vehicles across Washington, and if this idea wants to be expanded, it also includes California, Virginia, Maryland, and Texas.

Services

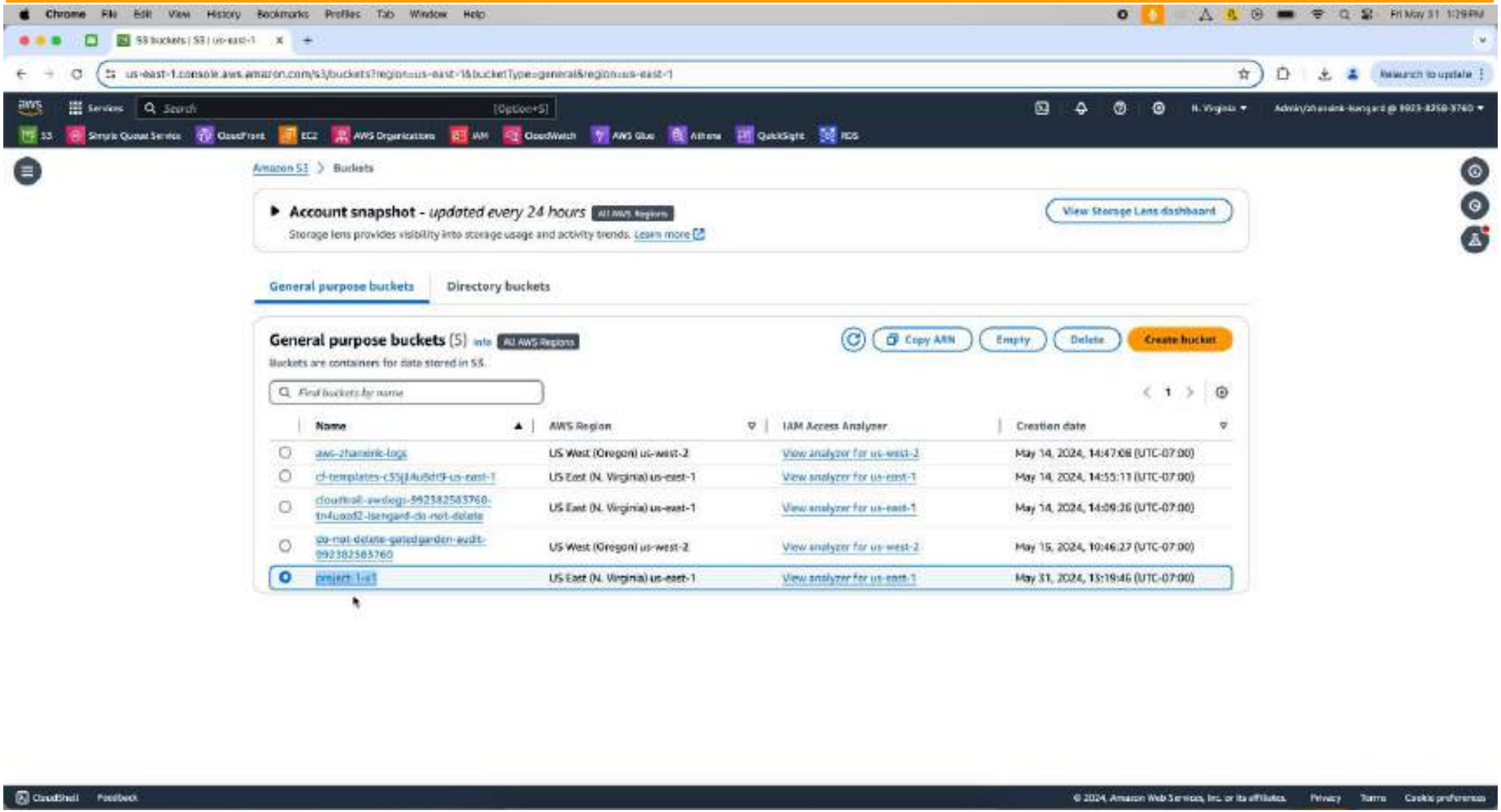
S3

Glue

Athena

Quicksight

Demo Video



https://broadcast.amazon.com/videos/1163141

S3

- **Object storage service**
  - Metadata
  - Key is composed of prefix + object name
    - S3://my-bucket/my_folder/my_file.txt
- **Region level**

S3

- **Durability and availability**
  - 99.99% availability
- **Scalability**
  - Auto growing and shrinking as adding and removing data
- **Security and data protection**
  - User-Based
  - Resource-based
- **Lowest price and highest performance**
  - Multiple storage classes

99.999999999%

Quicksight

- **Serverless Architecture**
- **Integrated with**
  - RDS, Aurora, Redshift, **S3, Athena**
- **Pay-per-Session pricing**

- **Use case**
  - Business analytics
  - Building visualizations
  - Get business insights using data

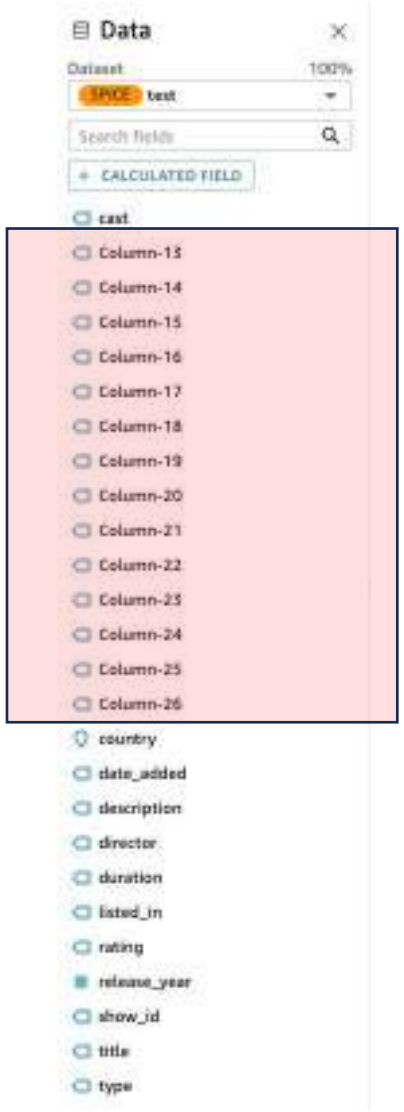
74%

Cons

S3

+

Quicksight



Glue

Fully Managed ETL Service

- No servers to provision or manage
- Automatic provisioning and scaling of resources

Data Catalog

- Persistent metadata store
- Stores table definitions, schemas, and partition information Integrates with Athena, Redshift Spectrum, and EMR

Data Integration

- Supports various data sources (S3, RDS, DynamoDB, etc.)
- Built-in data transformations (filter, join, dedup, etc.)

Serverless and Pay-per-use

- No upfront costs or long-term commitments
- Pay only for the resources used during job runs

Integration with Other AWS Services

- Works seamlessly with S3, Athena, Redshift, EMR, and more
- Enables building end-to-end data pipelines

Athena

Scalability:
•Athena automatically adjusts resources based on your query needs, so you don't have to worry about capacity planning.

Flexibility:
•Athena can adapt to changing data formats and structures without needing to define schemas upfront.
Durability and Availability:
•Your data stored in S3 is highly durable and available, thanks to Amazon's robust infrastructure.

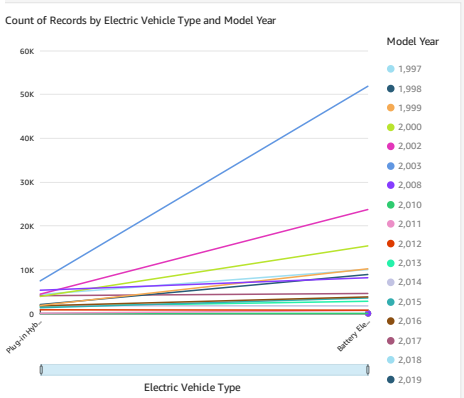
Security:
•Athena allows you to control access to your data through user-based and resource-based permissions.

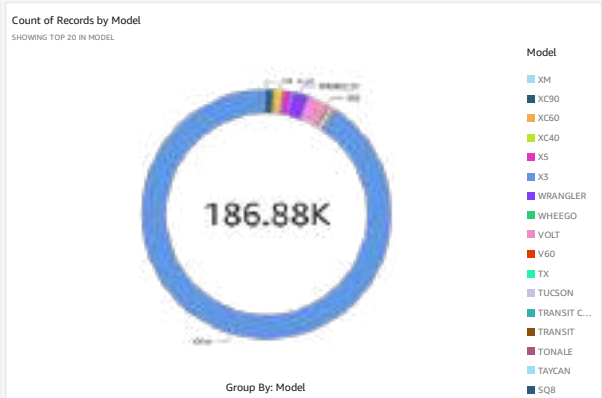
Cost-Effective:
•You only pay for the queries you run, eliminating the need to maintain expensive infrastructure.

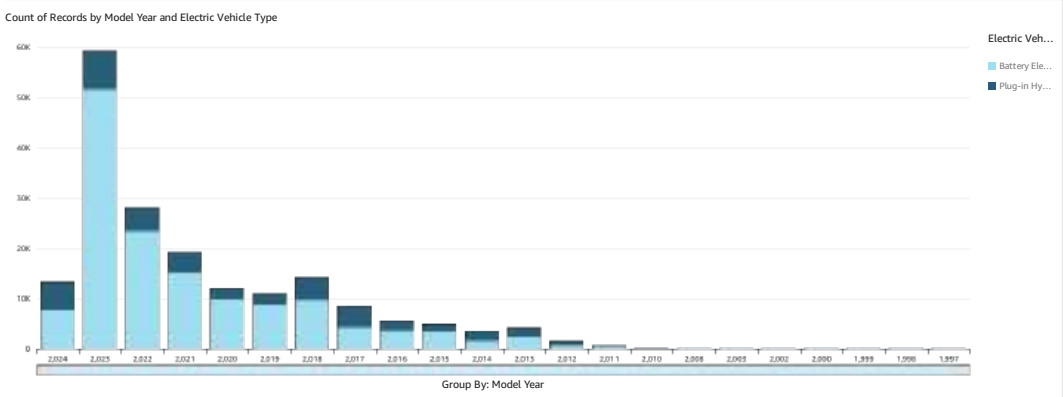
High Performance:
•Athena efficiently executes queries in parallel, providing fast results even on large datasets.

Storage Options:
•Athena can query data stored in different S3 storage classes, giving you cost-effective options for various data types.

Quicksight Dashboard







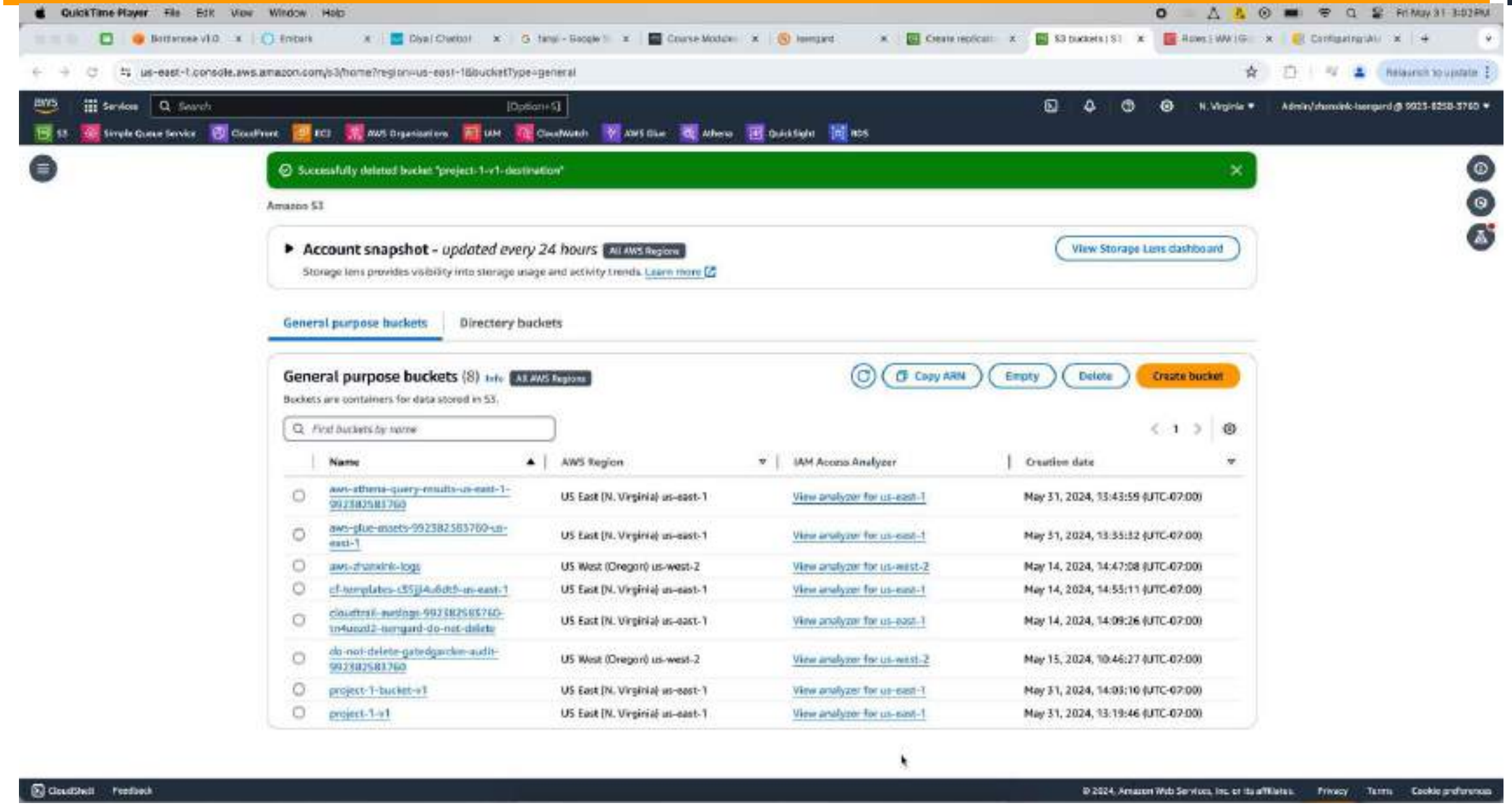
Summary

Service	 Amazon S3	 AWS Glue	 Amazon Athena	 Amazon QuickSight
Description	Highly scalable Durable object storage service	Fully managed extract Transform and load (ETL) service	Interactive query service to analyze data in S3 using SQL	Business intelligence and data visualization service
Key Features	Scalable data storage Durable data storage Object-based storage Accessible from anywhere	Automated data discovery Data transformation capabilities Serverless ETL processing Integration with other AWS services	Serverless SQL querying Leverages Glue Data Catalog No infrastructure management	Fully managed BI service, Interactive dashboards and visualizations Connects to various data sources Supports collaboration and sharing
Use Cases	Data lake storage Backup and archiving Static website hosting Content delivery	Data preparation for analytics ETL pipelines for data warehouses Automated schema creation Data cataloging and governance	Data analysis Querying large datasets in S3 Generating reports and insights Integrating with BI tools	Self-service business intelligence Interactive dashboards and visualizations Collaborative data exploration Embedding analytics in applications
Integrations	Integration with Glue and Athena Used as a data source for other AWS services Supports various data formats and protocols	Integrates with S3 for data sources and targets Generates data catalog used by Athena Can load data into other AWS services	Integrates with S3 for data sources and targets, Generates data catalog used by Athena, Can load data into other AWS services Leverages Glue Data Catalog for data discovery Can directly query data in S3, Integrates with various data visualization tools	Connects to data sources like S3 Athena, and other AWS services Provides interactive visualizations and dashboards Integrates with other BI and visualization tools

Disaster

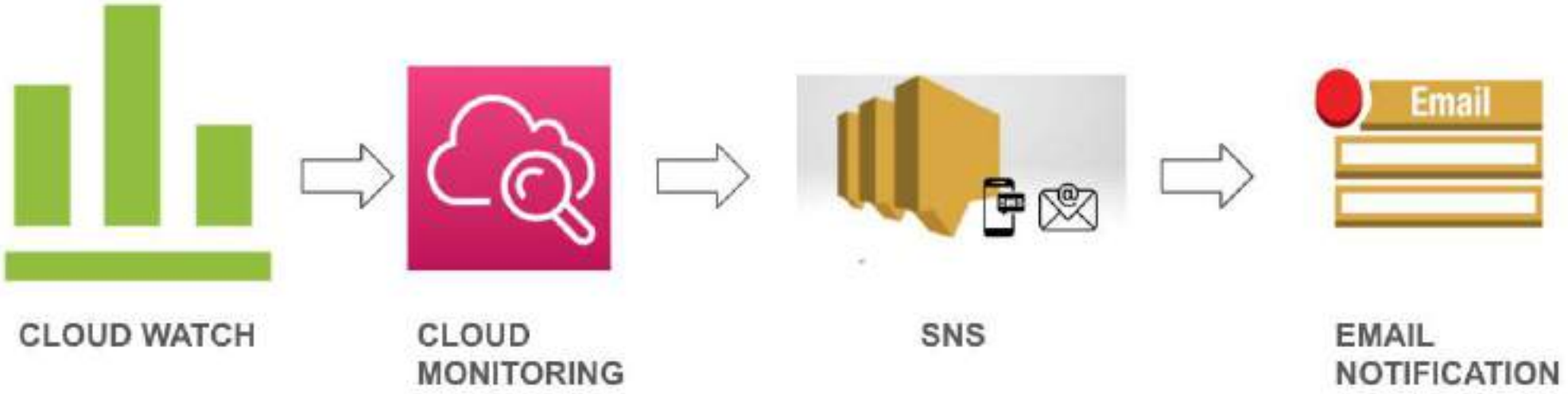
https://distributedalgorithm.wordpress.com/2016/02/25/disaster-recovery-for-solr-a-story-from-the-field/

Demo Video - Replication Rule



https://broadcast.amazon.com/videos/1163156

Notification

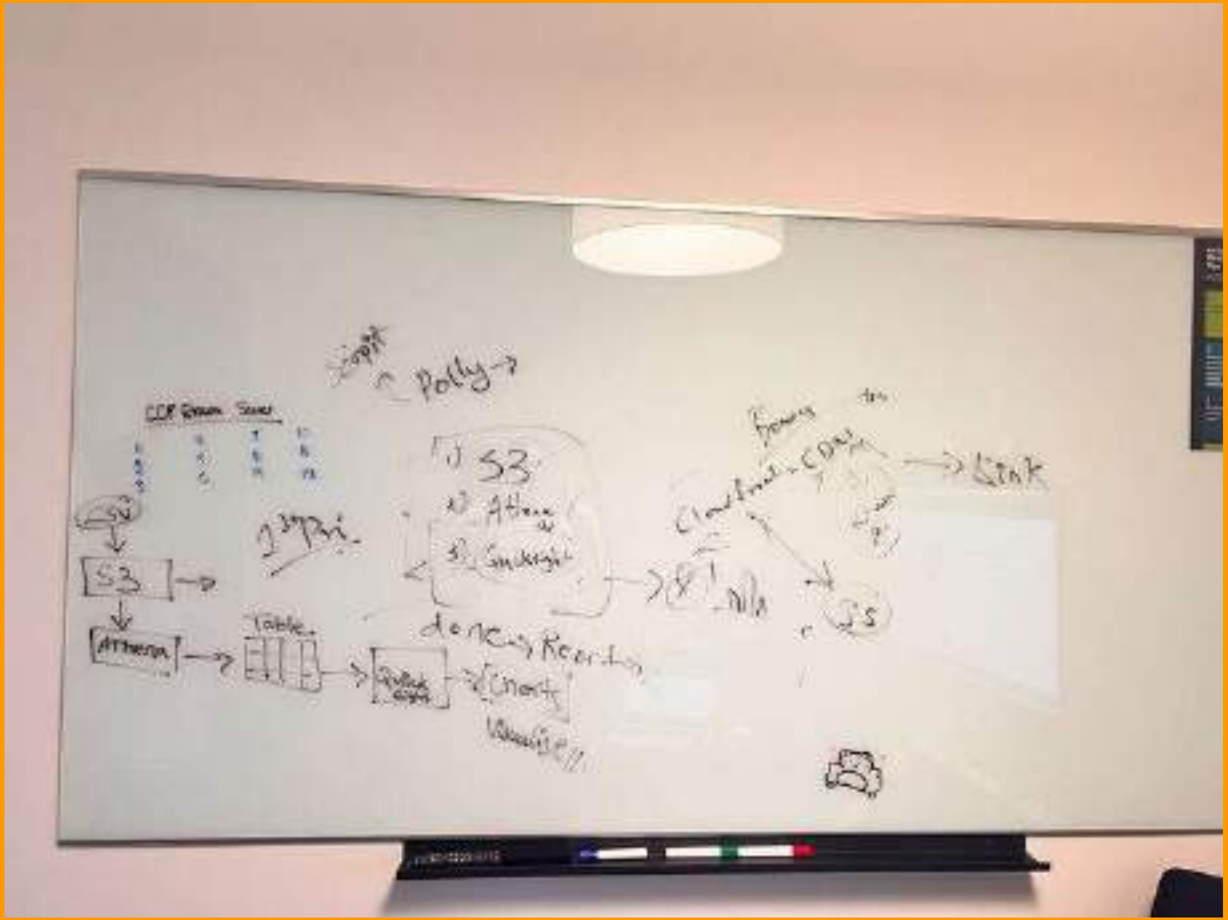


Problems we got

- **Problem 1:** Encounters issues importing data from Athena to Quicksight; tables do not appear in Quicksight
  - **Solution 1:** To ensure Athena and Quicksight are in the same region.

- **Problem 2:** Unable to connect Quicksight to S3 bucket
  - **Solution 2:** Need to select the specific s3 bucket to share permission. (*Manage Quicksight -> Security & permissions -> Manage(Access granted to services) -> select the specific s3 bucket)*

- **Problem 3***:* Glue crawler fails to discover data in S3 bucket.
  - Solution 1: Double-check the IAM role assigned to the Glue crawler. The role should have the necessary permissions to access the S3 bucket and read the data files. Additionally, ensure that the data files are stored in a supported format (e.g., CSV, JSON, Parquet) and that the crawler is configured to include the correct file extensions or prefixes.



Q&A