Deliverable 4

AWS Cost Analysis and Optimization:

Analyzing the cost of running AWS resources involves several factors, including the type and quantity of resources used, data transfer costs, storage costs, and compute costs. Here's a breakdown of potential cost considerations based on the activities required:

- **S3 Bucket**: Cost considerations include storage costs based on the amount of data stored, requests made (PUT, GET, etc.), data transfer costs (if data is transferred out of the AWS region), and any additional features used (like versioning or replication).

- **AWS Athena:** Athena charges are based on the amount of data scanned by queries. The cost varies with the volume of data processed during queries.

- **Data Analysis:** Cost here might involve the use of compute resources if complex analysis is performed using services like AWS Glue or Athena.

- **AWS Glue:** Glue's costs involve the resources used during ETL jobs, including the time taken for processing, as well as any additional features utilized.

- **Data Processing:** This includes costs for processing resources and any additional services used during data filtering, removal of duplicates, and column addition.

- **AWS SageMaker**: SageMaker costs are typically based on the type and quantity of instances used for model training and deployment, data storage, and any additional services or features utilized.

Cost Analysis:

1. Amazon S3:
   Monthly Cost: $1.29 USD
   Usage: Standard storage, requests, S3 Select data

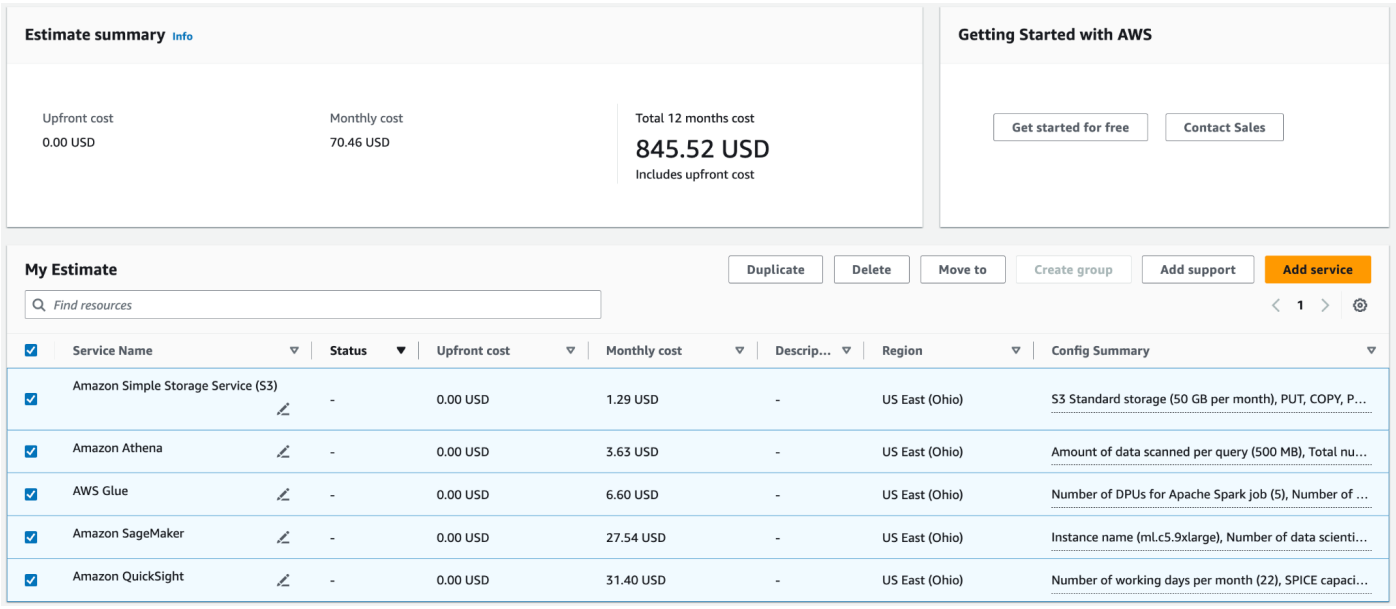
2. Amazon Athena:
   Monthly Cost: $3.63 USD
   Usage: Data scanned per query, total number of queries

3. AWS Glue:
   Monthly Cost: $6.60 USD
   Usage: DPUs for Spark and Python Shell jobs

4. Amazon SageMaker:
   Monthly Cost: $27.54 USD
   Usage: ML instance, Studio Notebook usage

5. Amazon QuickSight:
   Monthly Cost: $31.40 USD
   Usage: SPICE capacity, number of authors/readers

   Monthly cost - 70.46 USD
   Total 12 months cost - 845.52 USD
   Includes upfront cost



Cost Optimization Strategies:

S3 Bucket:

- Lifecycle Policies: Implement policies to transition data to lower-cost storage classes like Glacier or Glacier Deep Archive based on access frequency.
- Enable S3 Intelligent-Tiering: Automatically moves objects between frequent access and infrequent access tiers to optimize costs.

AWS Athena:

- Partitioning and Data Organization: Partition data to reduce the amount scanned during queries, minimizing costs.
- Optimize Queries: Use efficient SQL queries to minimize data scanning, thus reducing costs.

AWS Glue:

- Monitor DPU Usage: Scale resources based on actual usage to avoid overprovisioning.
- Optimize ETL Jobs: Review and optimize ETL processes for efficiency to reduce processing time and costs.

AWS SageMaker:

- Use Spot Instances: Utilize Spot Instances for training jobs to reduce costs significantly.
- Right-size Instances: Choose instance types based on workload requirements to avoid over-provisioning.

Amazon QuickSight:

- Usage Analysis: Monitor user access and usage to adjust user permissions and reduce unnecessary access.
- Optimize Data Queries: Optimize queries to reduce data retrieval and processing costs.

By employing these strategies and continuously monitoring resource usage, we can effectively optimize costs across the AWS services. We can use the tools like AWS Cost Explorer and Trusted Advisor for detailed insights and recommendations tailored to the specific usage patterns.