PHASE 4 Optimization, Scalability and Deployment (10%)
AWS Cost Analysis and Optimization:
o Analyze the cost of running your AWS resources.
o Suggest cost optimization strategies (e.g., using AWS Cost
Explorer and AWS Trusted Advisor).
Scaling the Pipeline (optional):
Investigate options for automating and scaling the pipeline, such as AWS
Lambda for triggering ETL jobs.
Deployment
Phase 4 Deliverable:
Project Repository on GitHub (Updated Table of Contents)
Deliverable 4 Document accessible in Github (include screenshots)

PROJECTED COST ESTIMATE FOR AWS SERVICES USED:

❖ Amazon S3 buckets:

S3 Standard storage

| 10 | | GB per month ▼ |

How will data be moved into S3 Standard?
Automatically calculates PUT, COPY, POST costs for moving data into S3 Standard initially. To compare the cost of current storage in S3 Standard to lifecycling this data to another storage class, you can specify that your storage is already stored in S3 Standard while selecting Lifecycle under the new storage class to capture the upfront cost of moving your data.

The specified amount of data is already stored in S3 Standard ▼

PUT, COPY, POST, LIST requests to S3 Standard
Ongoing monthly number of PUT, COPY, POST or LIST requests

500

GET, SELECT, and all other requests from S3 Standard
Ongoing monthly number of GET, SELECT and all other requests

500

Data returned by S3 Select
Ongoing monthly volume of data returned by S3 Select requests
Value

| 1 | | GB per month ▼ |

Data scanned by S3 Select
Ongoing monthly volume of data scanned by S3 Select requests
Value

| 1 | | GB per month ▼ |

▼ Show calculations

Tiered price for: 10 GB
10 GB x 0.023 USD = 0.23 USD
Total tier cost = 0.23 USD (S3 Standard storage cost)
500 PUT requests for S3 Standard Storage x 0.000005 USD per request = 0.0025 USD (S3 Standard PUT requests cost)
500 GET requests in a month x 0.0000004 USD per request = 0.0002 USD (S3 Standard GET requests cost)
1 GB x 0.0007 USD = 0.0007 USD (S3 select returned cost)
1 GB x 0.002 USD = 0.002 USD (S3 select scanned cost)
0.23 USD + 0.0002 USD + 0.0025 USD + 0.0007 USD + 0.002 USD = 0.24 USD (Total S3 Standard Storage, data requests, S3 select cost)
S3 Standard cost (monthly): 0.24 USD

S3 Standard cost (upfront): 0.00 USD

❖ Amazon Athena:

Configure Amazon Athena Info   ✕

With per query pricing, you pay only for the queries you run. To estimate your monthly cost, enter the total number of queries you expect to run in a day, week, or month and the amount of data scanned per query.

Total number of queries

| 100 |

Unit

| per day ▼ |

Amount of data scanned per query

| 50 |

Unit

| MB ▼ |

SQL queries with capacity based pricing

With capacity based pricing, you can pay an hourly price for compute resources to run your queries. You pay only for the capacity you need, measured in Data Processing Units (DPUs), and the time capacity is active in your account. To estimate your monthly cost, enter the number of DPUs and time needed.

Number of DPUs
A single Data Processing Unit (DPU) provides 4 vCPU and 16 GB of memory. Minimum of 24 DPUs per Capacity Reservation and minimum increment of 4 DPUs.

| *Enter amount* |

Length of time (hours) capacity is active
Minimum billed duration per Capacity Reservation is 1 hour.

Value

| 7 |

Unit

| hours per month ▼ |

Configure Amazon Athena Info   ✕

▼ Show calculations

Unit conversions

Total number of queries: 100 per day * (730 hours in a month / 24 hours in a day) = 3041.67 queries per month

Amount of data scanned per query: 50 MB x 9.5367432e-7 TB in a MB = 0.000047683716 TB

Pricing calculations

Rounding (3041.67) = 3042 Rounded total number of queries
3,042 queries per month x 0.000047683716 TB x 5.00 USD = 0.73 USD
SQL queries with per query cost (monthly): 0.73 USD
Max (7 hours per month, 1) = 7.00 hour(s) per month
Total SQL queries cost (monthly): 0.73 USD

❖ Amazon Quicksight:

Configure Amazon QuickSight Info ✕

Reader Pricing Info

Number of readers

```
10
```

Number of working days per month

```
7
```

Usage breakdown

Percent of active readers (twice a day)

```
5
```
%

Percent of frequent readers (twice a week)

```
50
```
%

Percent of occasional readers (twice a month)

```
25
```
%

Percent of inactive readers (less than monthly)

```
20
```
%

Configure Amazon QuickSight Info ✕

Unit conversions

Percent of active readers (twice a day): 5 / 100 = 0.05

Percent of frequent readers (twice a week): 50 / 100 = 0.5

Percent of occasional readers (twice a month): 25 / 100 = 0.25

Percent of inactive readers (less than monthly): 20 / 100 = 0.2

Pricing calculations

10 Reader(s) x 0.05 = 0.50 Active Reader(s)

RoundUp (0.50) = 1 billable active readers

2 active reader sessions per day x 7 work days per month x 0.30 USD = 4.20 USD per active reader

Min (4.20 USD, 5 USD maximum amount billable per reader) = 4.20 USD billable per reader

1 Reader(s) x 4.20 USD = 4.20 USD for active readers

Active reader cost (monthly): 4.20 USD

10 Reader(s) x 0.50 = 5.00 Frequent reader(s)

RoundUp (5.00) = 5 billable frequent readers

5 reader(s) x 2 reader sessions per week x 4 weeks per month x 0.30 USD = 12.00 USD for frequent readers

Frequent reader cost (monthly): 12.00 USD

10 Reader(s) x 0.25 = 2.50 Occasional reader(s)

RoundUp (2.50) = 3 billable occasional readers

3 Reader(s) x 2 reader sessions per month x 0.30 USD = 1.80 USD for occasional readers

Occasional reader cost (monthly): 1.80 USD

10 Reader(s) x 0.20 = 2.00 Inactive Reader(s)

RoundUp (2.00) = 2 billable inactive readers

2 Reader(s) x 0 reader sessions per month = 0.00 USD for inactive readers

Inactive reader cost (monthly): 0.00 USD

4.20 USD (active reader price estimate) + 12.00 USD (frequent reader price estimate) + 1.80 USD (occasional reader price estimate) = 18.00 USD per month

Cost of readers (monthly): 18.00 USD

Configure Amazon QuickSight Info ✕

Unit conversions

Percent of active readers (twice a day): 5 / 100 = 0.05

Percent of frequent readers (twice a week): 50 / 100 = 0.5

Percent of occasional readers (twice a month): 25 / 100 = 0.25

Percent of inactive readers (less than monthly): 20 / 100 = 0.2

Pricing calculations

10 Reader(s) x 0.05 = 0.50 Active Reader(s)

RoundUp (0.50) = 1 billable active readers

2 active reader sessions per day x 7 work days per month x 0.30 USD = 4.20 USD per active reader

Min (4.20 USD, 5 USD maximum amount billable per reader) = 4.20 USD billable per reader

1 Reader(s) x 4.20 USD = 4.20 USD for active readers

Active reader cost (monthly): 4.20 USD

10 Reader(s) x 0.50 = 5.00 Frequent reader(s)

RoundUp (5.00) = 5 billable frequent readers

5 reader(s) x 2 reader sessions per week x 4 weeks per month x 0.30 USD = 12.00 USD for frequent readers

Frequent reader cost (monthly): 12.00 USD

10 Reader(s) x 0.25 = 2.50 Occasional reader(s)

RoundUp (2.50) = 3 billable occasional readers

3 Reader(s) x 2 reader sessions per month x 0.30 USD = 1.80 USD for occasional readers

Occasional reader cost (monthly): 1.80 USD

10 Reader(s) x 0.20 = 2.00 Inactive Reader(s)

RoundUp (2.00) = 2 billable inactive readers

2 Reader(s) x 0 reader sessions per month = 0.00 USD for inactive readers

Inactive reader cost (monthly): 0.00 USD

4.20 USD (active reader price estimate) + 12.00 USD (frequent reader price estimate) + 1.80 USD (occasional reader price estimate) = 18.00 USD per month

Cost of readers (monthly): 18.00 USD

❖ AWS Glue:

❖ Amazon Sagemaker:

Configure Amazon QuickSight Info   ✕

Unit conversions

Percent of active readers (twice a day): 5 / 100 = 0.05
Percent of frequent readers (twice a week): 50 / 100 = 0.5
Percent of occasional readers (twice a month): 25 / 100 = 0.25
Percent of inactive readers (less than monthly): 20 / 100 = 0.2

Pricing calculations
10 Reader(s) x 0.05 = 0.50 Active Reader(s)
RoundUp (0.50) = 1 billable active readers
2 active reader sessions per day x 7 work days per month x 0.30 USD = 4.20 USD per active reader
Min (4.20 USD, 5 USD maximum amount billable per reader) = 4.20 USD billable per reader
1 Reader(s) x 4.20 USD = 4.20 USD for active readers
Active reader cost (monthly): 4.20 USD
10 Reader(s) x 0.50 = 5.00 Frequent reader(s)
RoundUp (5.00) = 5 billable frequent readers
5 reader(s) x 2 reader sessions per week x 4 weeks per month x 0.30 USD = 12.00 USD for frequent readers
Frequent reader cost (monthly): 12.00 USD
10 Reader(s) x 0.25 = 2.50 Occasional reader(s)
RoundUp (2.50) = 3 billable occasional readers
3 Reader(s) x 2 reader sessions per month x 0.30 USD = 1.80 USD for occasional readers
Occasional reader cost (monthly): 1.80 USD
10 Reader(s) x 0.20 = 2.00 Inactive Reader(s)
RoundUp (2.00) = 2 billable inactive readers
2 Reader(s) x 0 reader sessions per month = 0.00 USD for inactive readers
Inactive reader cost (monthly): 0.00 USD
4.20 USD (active reader price estimate) + 12.00 USD (frequent reader price estimate) + 1.80 USD (occasional reader price estimate) = 18.00 USD per month
Cost of readers (monthly): 18.00 USD

Configure Amazon SageMaker Info   ✕

Selected Instance:
ml.c5.12xlarge

Compute Type: Compute Optimized Instances

V CPU: 48

Memory: 96 GiB

Clock Speed: 3 GHz

GPU: 0

Network Performance: N/A

Storage: EBS only

GPU Memory: 0

▼ Show calculations

5 data scientist(s) x 1 Studio Notebook instance(s) = 5.00 Studio Notebook instance(s)
5.00 Studio Notebook instance(s) x 7 hours per day x 1 days per month = 35.00 SageMaker Studio Notebook hours per month
35.00 hours per month x 2.448 USD per hour instance cost = 85.68 USD(monthly On-Demand cost)
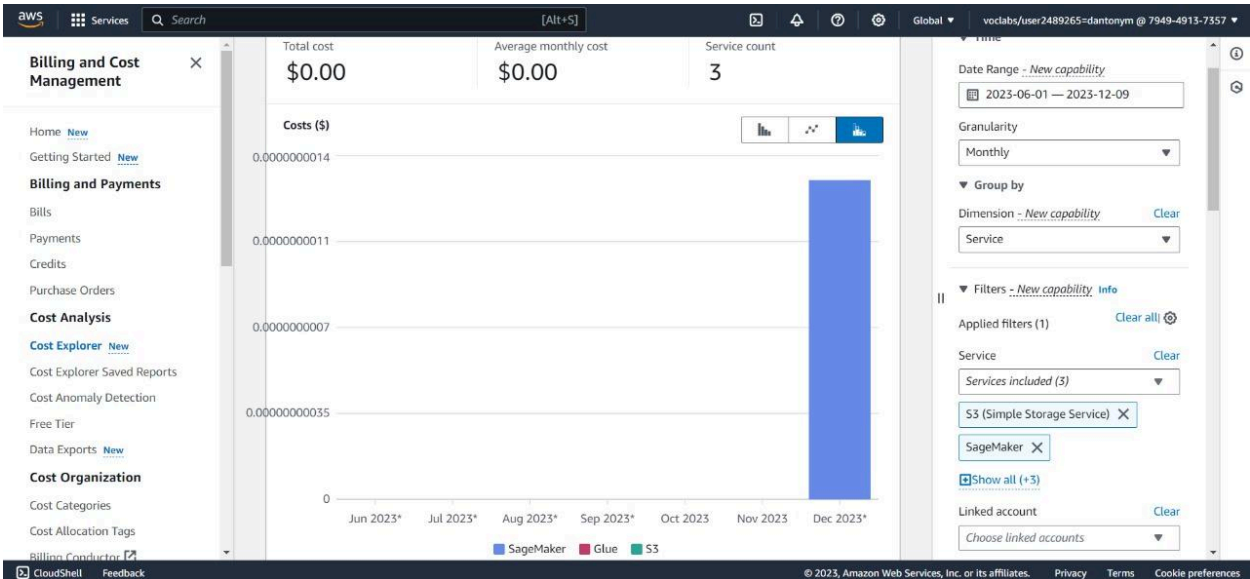**Total cost for Studio Notebooks (monthly): 85.68 USD**

TOTAL ESTIMATE : $124.85 per month

Monthly cost	Total 12 months cost	Currency						
124.85000000000000	1498.20	USD						
	* Includes upfront cost							
Region	Description	Service	Upfront	Monthly	First 12 months total	Currency	Status	Configuration summary
US East (N. Virginia)	Predict students' dropout and academic success Dataset	S3 Standard	0	0.24	2.88	USD		S3 Standard storage (10 GB per month), PUT, COPY, POST, LIST requests to S3 Standard (500), GET, SELECT, and all other requests from S3 Standard (500), Data returned by S3 Select (1 GB per month), Data scanned by S3 Select (1 GB per month)
US East (N. Virginia)	Predict students' dropout and academic success Dataset	Data Transfer	0	0	0.00	USD		
US East (N. Virginia)	Predict students' dropout and academic success Dataset	Amazon Athena	0	0.73	8.76	USD		Total number of queries (100 per day), Amount of data scanned per query (50 MB), Length of time (hours) capacity is active (7 hours per month)
US East (N. Virginia)	Predict students' dropout and academic success Visualization	Amazon QuickSight Readers, Authors, SPICE	0	36	432.00	USD		Number of working days per month (7), SPICE capacity in gigabytes (GB) (10), Number of authors (1), Number of readers (10)
US East (N. Virginia)	Predict students' dropout and academic success Glue job	AWS Glue ETL jobs and interactive sessions	0	2.2	26.40	USD		Number of DPUs for Apache Spark job (10), Number of DPUs for Python Shell job (0.0625)

US East (N. Virginia)	Predict students' dropout and academic success Notebook	SageMaker Studio Notebooks	0	85.68	1028.16	USD		Instance name (ml.c5.12xlarge), Number of data scientist(s) (5), Number of Studio Notebook instances per data scientist (1), Studio Notebook hour(s) per day (7), Studio Notebook day(s) per month (1)

OPTIMIZATION STRATEGIES:

- **Amazon S3:** Includes costs for standard storage (10 GB per month), various requests, and data handling via S3 Select.
- **Amazon Athena:** Costs are based on the number of queries (100 per day) and data scanned per query.
- **Amazon Quicksight:** The cost reflects the usage for readers, authors, and SPICE capacity.
- **AWS Glue:** Involves costs for ETL jobs and interactive sessions, calculated based on the number of Data Processing Units (DPUs).
- **Amazon SageMaker:** The largest cost component, associated with SageMaker Studio Notebooks for data scientists.

For optimization:

- We can review the instance sizes and scale patterns in SageMaker and other services that could yield significant savings.
- We can further investigate data transfer and storage patterns in S3 and Athena, which can lead to opportunities for cost reduction.
- The usage pattern in Quicksight is over provisioned and could be optimized.

Submitted by Group14:

Ramit Aditya

Anirudh Cheruvu

Mounisha Bolla

Amirthavarshini Dhanavel

Delphine Antony Muthu