# Option Analysis

## I/ Option Overview

* **Bronze Layer:** Ingest data through Azure Data Lake Storage Gen2 (ADLS2) REST API.
* **Silver Layer:** Azure Function to transform and store data
* **Gold Layer:** Azure Function to transform and store data
* **Orchestration:** Github Actions
* **Storage:** ADLS2 for all layers

The related Azure products and their functionality for the GenAI solution are highlighted in *Table 1.*

| Categorisation | Function | Type of Solution | Version |
| --- | --- | --- | --- |
| Artificial Intelligence | AI & ML Hub Studio | Azure AI Studio | Baseline |
| Artificial Intelligence | Azure Open AI Services | GPT-4o-mini GD |
| Machine Learning | Azure Machine Learning | E8a v4 |

Table 1. Azure products and their functionality.

## II/ Option Justification

Azure Functions is very cost-effective for small to medium sized workloads with the first 400,000 GB/s of execution and 1,000,000 executions being free. As our initial dataset is small, around 10,000 rows of data which is around 5MB of data for our Bronze layer. Cost will only incur if our data volume or execution amounts increase however, even then for the scale of this project the costs will be negligible.

Azure Functions is very easy to set up and deploy, thus, efficient for small-scale projects like ours. Furthermore, you pay as you go so you pay for the resources as you use them increasing cost efficiency.

As for our storage solution, we will be using ADLS2 for the first MVP as it is based on Azure Blob storage which is very cheap and can store an ‘infinite’ amount of data. So both cost and scalability will not be an issue with this storage solution.

The Azure AI Studio would be the main platform the AI Team will utilise for the project lifecycle as it acts as a hub to connect LLM, ML and the cleaned data. In addition, the cost of Azure AI Studios is free and scalable as additional AI models can be integrated depending on the scenario.

The LLM selected by the AI team is the GPT-4o-mini global deployment, as it enables the same features of the enterprise GPT-4o model at a cheaper rate. If scalability becomes an issue due to increased data processing, the Team can upgrade to a GPT-4o enterprise plan to ensure satisfactory input and output times for each prompt.

The Azure ML E8a V4 is important in the GenAI’s capability to detect fraudulent claims. The ML allows for fraud pattern recognition that would be integrated into the LLM for the claims officer to interact with.

## 

## III/ Option Pricing

Azure Data Lake Storage Gen2:

* combines the power of a Hadoop-compatible file system with integrated hierarchical namespace with the massive scale and economy of Azure Blob Storage to help speed your transition from proof of concept to production
* Costs are billed based on the number of GB of storage needed and these costs are minimal
* Scalable - As we progress through the bronze, silver and gold layers the amount of data also increases. Gen2 allows easy scalability of storage as we can increase for each GB of storage needed
* Storage - First 51,200 GB / month - $0.031 per GB
* Write operations - $0.10942 (every 4 MB, per 10,000)
* Read operations - $0.00874 (every 4 MB, per 10,000)

Azure Functions:

* Consumption Plan: Pay only for the time your functions run. Billing is based on the number of executions, execution time, and memory used.
* The first 400,000 GB/s of execution and 1,000,000 executions are free.
* Each 1,000,000 executions more is AUD$0.31

Application Insights

* Data Ingestion: $5.11/GB of daily logs ingested
* Data Retention: The first 3 months of retention are free
  + $0.23/GB per month for 0.9GB total monthly ingestion

Azure AI Studios

* Free platform, only requires payment for the AI and ML models that are utilised.

Azure Open AI Services

* GPT-4o-mini Global Deployment
  + Input cost: A$0.00023 per 1,000
  + Output cost: A$0.0010 per 1,000

Azure Machine Learning:

* E8a v4
  + A$0.925/hour

## IV/ Pricing summary

In summary, our cost can be categorised into 3 sections:

| **Category** | **Service** | **Price /month** | **Calculation/Explanation** |
| --- | --- | --- | --- |
| Storage | ADLS2 | **$**0.44748 | * Storage - First 51,200 GB / month - $0.031 per GB * Write operations - $0.10942 (every 4 MB, per 10,000) * Read operations - $0.00874 (every 4 MB, per 10,000)   Calculation:   * The total of the operations come to 0.14916 * Then we multiply by 3 for each of our bronze, silver, and gold layers storage * **Total Cost: $0.44748/month** |
| Compute | Azure Functions | $0 | First 400,000 GB/s of execution and 1,000,000 executions are free  We will be processing 10Mb of data, so it is unrealistic to ever use up all of the free-tier capacity. Therefore, Azure Functions will be free. |
| Monitoring | Application Insights | Likely: $0  Unlikely: $5.11 | We intend to turn off all system logging from Application Insights, but only log the custom logs in our code. This means we will turn off all telemetry like requests, dependencies, and exceptions.  Assuming we can even log 30Mb of data daily for 30 days, we will still not even log 1Gb of data per month which will not be charged anything. Realistically, we will log a lot lower than this amount so we are confident that we will be using Application Insights for free.  Another aspect is log retention, since Application Insights provides us with 3 months of free log retention which align with how long this project will run for, we will choose to retain our log for 3 months.  So in the worst case scenario where we log more than 1Gb of data per month, we will be charged $5.11 per month. |
| Orchestrate | Github Actions | $0 | Free because we have a public repository. |
| AI & ML Hub Studio | Azure AI Studio | $0 | Free as it only charges for each model that is used. |
| LLM | Azure Open AI Services (GPT-4o-mini GD) | $0.00123 | The cost includes both input and output cost per 1,000 actions. The Team agrees the number of actions appropriate for the project lifecycle. |
| ML Algorithm | Azure Machine Learning (E8a v4) | $45.25 | The cost factors in 50 hours of usage for E8a v4. The Team agrees on the number of actions appropriate for the project lifecycle. Moreover, the E8a v4 allows for scalability for scenarios where the data size is to increase. |
| Total | **Likely: $46.70 per month**  Unlikely: $52.26 per month - this is when we log more than expected, but since we intend to only log custom logs, it will be very low. | | |