Task 2 – Cost Estimation

Objective

The goal of this task is to estimate the monthly cost of deploying the data pipeline on Microsoft Azure based on the following plan:
• The DuckDB data warehouse is updated once per day.
• The Streamlit dashboard remains always available to users.
This helps determine whether continuing to deploy the data pipeline on Azure is cost-effective.

Verified Monthly Cost (from Azure Cost Management)

• Total Monthly Cost: ≈ 73.34 SEK (≈ $6.50 USD)
• Average Daily Cost: 1.56 SEK/day
• Subscription: Azure for Students (covered by free credits)

Cost Breakdown by Service

Service	Monthly Cost (SEK)	Purpose
Azure Container Instance (Dagster)	≈ 61.36 kr	Runs daily data pipeline and updates DuckDB
Azure Container Registry	≈ 6.45 kr	Stores Docker images used in deployment
Azure Storage	≈ 2.85 kr	Stores data and artifacts
Recovery Services Vault	≈ 2.77 kr	Backup and monitoring
Azure App Service (Streamlit)	≈ 0.18 kr	Dashboard (Free Tier, always online)
Total Verified Cost	≈ 73.34 kr	≈ $6.50 USD per month

Analysis

• Dagster container is the main cost driver (≈80% of total).
• Streamlit dashboard runs on Free Tier with negligible cost.
• Other services like registry, storage, and backup add minimal overhead.
• Total cost remains within free student credits – no real payment required.

Conclusion

The verified monthly cost of ≈73 SEK demonstrates that the Azure deployment is cost-efficient, sustainable, and suitable for continuous daily data updates and dashboard availability. The company can confidently continue using Azure for this project.

Task 4 – Cost Comparison to Cloud Data Warehouse (Snowflake)

Objective

This task estimates the additional cost that would occur if DuckDB were replaced by the cloud data warehouse Snowflake. Although Snowflake was not deployed, the estimation is based on typical usage for this project's pipeline workload.

Current Setup (Azure + DuckDB)

Service	Monthly Cost (SEK)	Purpose
Azure Container Instance (Dagster)	≈ 61.36 kr	Runs daily data pipeline and updates DuckDB
Azure Container Registry	≈ 6.45 kr	Stores Docker images used in deployment
Azure Storage	≈ 2.85 kr	Stores data and artifacts
Recovery Services Vault	≈ 2.77 kr	Backup and monitoring
Azure App Service (Streamlit)	≈ 0.18 kr	Dashboard (Free Tier, always online)
Total Verified Cost	≈ 73.34 kr	≈ $6.50 USD per month

Alternative Setup (Azure + Snowflake)

Component	Estimated Monthly Cost (SEK)	Description
Snowflake Compute	≈ 170 kr	1 X-Small virtual warehouse (~$15 USD) running 1 hr/d
Snowflake Storage	≈ 30 kr	10 GB managed storage (~$3 USD)
Existing Azure Resources	≈ 73 kr	Containers, registry, storage, dashboard
Total (Approx.)	≈ 200–220 SEK / month	≈ $18–20 USD/month

Cost Comparison

Setup	Monthly Cost (SEK)	Increase	Ratio
Azure + DuckDB	≈ 73 kr	–	1×
Azure + Snowflake	≈ 200 kr	+127 kr	≈ 3–5× higher

Pros and Cons of Using Snowflake

Advantages
• Scalable and elastic – handles large datasets and many users.
• Fully managed service – no maintenance required.
• High performance for analytics and BI.
• Supports data sharing and concurrency.

Disadvantages
• 3–5× higher cost for small workloads.
• Idle compute costs can still occur.
• More complex setup and integration.
• Overkill for student or lightweight pipelines.

Conclusion

Switching from DuckDB to Snowflake would increase monthly costs from about 73 SEK to around 200 SEK. While Snowflake offers scalability and performance, it is not cost-effective for this small-scale project.

Recommendation: Continue using DuckDB within Azure. It provides an optimal balance of performance, simplicity, and cost-efficiency for this deployment.