**AWS Portfolio Project**

Mandeep Kaur (2236017)

University Canada West

BUSI 653: Cloud Computing Technologies (HBD-SPRING25-03)

Mahmood Mortazavi Dehkordi

24th June 2025

[Project Description: 3](#_Toc1430229208)

[Project Title: 3](#_Toc1649424798)

[Objective: 3](#_Toc1985794350)

[Dataset: 3](#_Toc253398588)

[Methodology: 3](#_Toc1805645163)

[Tools and Technologies: 4](#_Toc242875872)

[Deliverables: 4](#_Toc1995117632)

[AWS Deployment and Service Models 4](#_Toc914194637)

[AWS Cost Analysis 6](#_Toc2032893621)

[AWS Global infrastructure 9](#_Toc1070462277)

[AWS IAM 10](#_Toc1138330697)

[AWS VPC 12](#_Toc158185041)

[AWS Lambda 13](#_Toc164704258)

[AWS EBS 14](#_Toc1021751728)

**Introduction**

Cloud platforms today have emerged as the critical elements of growth in scalable, secure, and affordable data operations in the modern data-driven environment. Usage of services like Amazon Web Services (AWS) enables organizations to manage the complexity of data that is increasing in the quest for achieving operational excellence. During Weeks 2-4, I had a guided learning experience that connected academic theory to practice and led me to learn the best practices of cloud-based data analysis work rather thoroughly. The two steps of the root cause identification and the end-to-end ETL execution process carried out in this project have made me capable of providing analytical knowledge based on the strong cloud infrastructure. This reflection describes my development in terms of business analysis, infrastructure, data preparation, pipeline monitoring, and governance.

# Descriptive Analysis

**Week 2 - Cloud Infrastructure Planning and Business Analysis**

During Week 2, I gained a comprehensive knowledge of how to approach cloud adoption by critiquing practical business issues applying scholarly instruments. I made a Fishbone diagram to decompose and pinpoint the core of inefficiencies in operations. This activity allowed me to criticize the place of digital change required. It is on this basis that I came up with a Data Lake architecture that could fit both structured and unstructured data, but also is essential for cloud-friendly analytics at scale.

I had practical experience with AWS S3, structuring logical folders and using ethical principles of tagging, which exposed me to the system of cloud structure management and access control system practices. I also got acquainted with AWS service structures and related pricing models, and that helped me to understand how to design cost-efficient and secure cloud infrastructure.

The weekly discussions in the forums enabled me to input professional tips, especially on issues related to cloud migration and failure to embrace change in organizations, as I further test my capacity to apply theoretical knowledge in practical situations.

**Week 3: Dataset Ingestion and Cleaning**

In Week 3, I was concerned with raw datasets ingestion on a cloud platform and cleaning. The cost of dataset ingestion was estimated by using the AWS Pricing Calculator, which helped me know how to plan the usage of the cloud resources, considering the budget accordingly. I subsequently analyzed the dataset on missing values, incorrect data format, and duplicates, and created a data cleaning pipeline in AWS Glue and Python.

This practical allowed me to develop a practical data preparation pipeline, which is cloud-native and provides clean, reliable, and analysis-ready data. In the process, the significance of automating the cleaning activities to eliminate human error, save time of processing, and facilitate large-scale analytics was highlighted. This week helped my knowledge base with regards to the direct business intelligence effects of data quality.

**Week 4: Why Cloud-Based ETL Pipeline and Monitoring?**

During Week 4, I constructed a professional-grade ETL (Extract, Transform, Load) pipeline on AWS Glue with the detection of a schema through the Glue Crawler and metadata recording in the Glue Data Catalog. I used Amazon Athena to query the curated data, and this step was the central point of my data transformation and analysis.

I reused the same instance of the AWS Pricing Calculator to guess the cost of different ETL jobs, and this served the objective of cost-optimized infrastructure. I also adjusted AWS Cloud Watch to monitor the main metrics, such as the size of buckets and the results of the ETL jobs and configured real-time notifications to establish the visibility of operations and availability of the system.

This week, I enhanced my skills to create end-to-end cloud data pipelines that are efficient and scalable, as well as capable of good governance and monitoring. These are the skills that can be widely used in real-life cloud data engineering work and have brought me closer to the business strategy in my technical work process.

Project Description: Descriptive Analysis of Customer Purchase Patterns

Project Title: Understanding Call Center Patterns – City of Vancouver

Objective:

The current project conducts a descriptive statistical study of call center work in the City of Vancouver between 2020 and 2025. Using cloud tools and AWS services, it is aimed:

* Analysis of call volumes every month.
* Determine waves of seasonality and workload.
* In power workforce scheduling and support responsiveness in operations with data-based insights.

Dataset: The dataset presents monthly records on call-centers' performance throughout the City of Vancouver, and the primary fields are the following:

**Name Description**

Date (Month, Year) Time record of each month

Calls Offered Total inbound calls (unanswered 304 + answered calls 38)

Call Volume Count Summation of the number of calls provided

Annual division Aggregate data between 2020-2025

This architecture is like the one I applied to the analysis of the workloads in the public service with AWS Glue, S3, and DataBrew.

## Methodology:

**Step 1: Data Ingestion and Preparation:**

* Raw CSV data is used by Amazon S3.
* Cleaned data using AWS Glue and DataBrew (e.g., removing duplicates, standardising data formats).
* Set lifecycle policies up to optimize my storage through the AWS pricing calculator and approximately estimate my monthly costs of ETL.

**Step 2- Descriptive Statistics:**

Number of Total Call Offerings within 20 months: 3,391

Mean: The average calls per month are

=20 Months/3391 Calls​=169.55 Calls/Month

**Step 3 Data Visualization**:

I have developed:

* Time-series graphs on the current trend in calls by month.
* Bar graphs of comparison of volumes of calls between years.
* Heatmaps (optional future improvement) indicating weekdays/hour needles (weekday/hour spikes).

**4. ETL and Cataloging**

To get the data ready quickly to be analyzed, I employed AWS Glue Crawler to generate metadata tables of the data framework and automatically read through the dataset to find the schema. The following queries were made by querying the data through Amazon Athena using such tables.

Also, I created a whole pipeline of ETL based on AWS Glue that could process the whole pipeline of having data and treating it, turning it into the form needed by the analysis, and storing the results in a structured form.

**5. MONITORING, GOVERNANCE, and data security**

To achieve high-level data security, I enabled AWS KMS (Key Management Service) with customer-managed keys to encrypt the contents of the S3 bucket. I have also enabled bucket versioning and set replication across buckets to help create proper backups and avoid any loss of my data.

Data governance includes applying quality rules with the help of AWS Glue, which divides records into passed and failed ones and saves them in separate folders. This helped in providing more control and review of data.

I also configured CloudTrail to log API activity to be used as a part of the audit and implemented CloudWatch dashboards to be able to monitor important metrics, which were bucket size and Glue job performance. Operational issues were monitored, and alerts were set to respond to them on a real-time basis.

**6. Analysis and Cost Control**

I have assessed my cloud infrastructure design based on the AWS Well-Architected Framework regarding the chief pillars of the framework, which include Operational Excellence, Security, Reliability, Performance Efficiency, and Cost Optimization.

I calculated the annual price of operating the solution with the help of the AWS Pricing Calculator. The mentioned results proved that the pipeline was not simply effective and scalable but also cost-efficient to be used on a long-term basis.

## Tools and Technologies:

* Amazon S3 - secure storage of data
* Glue DataBrew- data profiling and data cleaning
* Glue Crawler is used to detect and catalog the schema
* Glue ETL Jobs- to construct the data pipeline
* Amazon Athena - to make SQL queries over catalogued data
* CloudTrail - AWS API calls auditing
* CloudWatch- to observe and notify
* Athena SQL- Analysis and Reporting
* AWS Pricing Calculator- cost estimation

## Deliverables:

* A trusted, cloud ETL pipeline with managed AWS services
* Data governance model: A classification of the records into the pass/fail groups according to the quality checks.
* Interactive dashboards to include Athena and/or QuickSight to make informative visualizations
* A cost breakdown report proving the low cost of operations of the pipeline

# AWS Deployment and Service Models

**A screenshot of a computer

AI-generated content may be incorrect.**

**A screenshot of a computer

AI-generated content may be incorrect.**

A diagram of a cloud

AI-generated content may be incorrect.

A screen shot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

# AWS Cost Analysis

A screenshot of a document

AI-generated content may be incorrect.

**Case Study #5: AWS Pricing Calculator**

**Data Storage**

**A screenshot of a document

AI-generated content may be incorrect.**

**Data Profiling and Cleaning**

**A screenshot of a document

AI-generated content may be incorrect.**

**Data Analysis**

**A screenshot of a document

AI-generated content may be incorrect.**

**A screenshot of a computer

AI-generated content may be incorrect.**

**A screenshot of a computer

AI-generated content may be incorrect.**

# AWS Global infrastructure

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

# AWS IAM

**Case Study #8: Who is responsible**

**A screenshot of a computer

AI-generated content may be incorrect.**

**A white box with text on it

AI-generated content may be incorrect.**

**Case Study #9: IAM Practice: Lab 1**

**A screenshot of a computer

AI-generated content may be incorrect.**

**Module 4 Knowledge Check**

**A screenshot of a computer

AI-generated content may be incorrect.**

# AWS VPC

**Case Study #10: Build your VPC: Lab 2**

**A screenshot of a computer

AI-generated content may be incorrect.**

**Module 5 Knowledge Check**

**A screenshot of a computer

AI-generated content may be incorrect.**

# AWS Lambda

**Case Study #11: Create an AWS Lambda function**

**A screenshot of a computer

AI-generated content may be incorrect.**

**Module 6 Knowledge Check**

**A screenshot of a computer

AI-generated content may be incorrect.**

# AWS EBS

**Case Study #12: Working with Amazon EBS: Lab 4**

**A screen shot of a graph

AI-generated content may be incorrect.**

**Module 7 Knowledge Check**

**A screenshot of a computer

AI-generated content may be incorrect.**

**Conclusion**

This development of business problem analysis to technical implementation within these three weeks gave a whole picture of cloud-based data management. I gained infrastructure planning and pipeline automation, quality assurance, and cost-optimisation skills, which are considered the foundational blocks of contemporary data engineering. Above all, this experience made me realize how to integrate technical design into the overall business strategy and how to achieve procedural effectiveness and value additions. With these skills, I am in a better position to implement data initiatives in the real world with precision, vision, and confidence.

**References**

https://awsacademy.instructure.com/courses/118864/modules/items/11235287