Project part 1   
Anand Kumar Ratanlal Ahirwar (2331600)

University Canada West

BUSI 653: Cloud Computing Technologies (Sec 01)

Professor- Mahmood Mortazavi Dehkordi

March 4 , 2025

DAP Design and Implementation and Descriptive Analysis

This analysis aims to track changes in number of building permits issued. This analysis helps understand how many permits are issued and what is the issue date, created date , applicant name , years of issue , address of the building and etc.

The data analytic platform is designed based on four key components:

1. Data Collection & Ingestion of Building permit dataset and uploading into S3 bucket.
2. Data Profiling - Using AWS Glue DataBrew to from structured data.
3. Data Cleaning - Using AWS Glue DataBrew to clean the data.
4. Data Cataloging - Organizing the cleaned data into a well-defined schema using AWS Glue and storing it in AWS S3.( catalog file )
5. Data Summarization - Running extraction, time stamp, filter, transform and loading into User and System format, using AWS Glue.

Data analytics platform (DAP) design – This is the road map for the whole AWS process which I have followed to land o the result i.e. Summarization technique

A screenshot of a computer

AI-generated content may be incorrect.

Figure no 1

Data analysis diagram

Vancouver city portal- Its shows value of the building permit issued between to year 2017 to 2015 and other Y- axis line shows permit elapsed days.   
  
A screenshot of a computer

AI-generated content may be incorrect.

Figure No 2

Data Collection & Ingestion -   
  
downloaded the data.csv (Permits\_dataset.csv) from city of Vancouver portal. It is the big data all the permits issued by the city along with issue date, name of the applicant, type of property, value of the property, created date of the permit and other 10-15 columns related to building permit issue.

A screenshot of a computer

AI-generated content may be incorrect.

*Figure no.3*

**Data Profiling - Using AWS Glue DataBrew to structure the data.**

In this step, I analyse the quality of the data by checking for errors, missing values, and patterns. This helps me understand the structure and completeness of the data before using it for analysis. I created the new bucket in S3 by the name of van-city-trf-anand. After adding profiling job the file saved to van-city-trf-anand.

A screenshot of a computer

AI-generated content may be incorrect.

*Figure no.3*

A screenshot of a computer

AI-generated content may be incorrect.

*Figure no 4*

**Data Cleaning - Using AWS Glue DataBrew**data cleaning like rename the column heading for easy understanding.   
A screenshot of a computer

AI-generated content may be incorrect.

*Figure no 5*

A screenshot of a computer

AI-generated content may be incorrect.

*Figure no 6*

**Data Cataloging - Organizing the cleaned data into a well-defined schema**

I organize the data by creating a tabular format data source. This makes it easier to locate and use relevant data when it’s in need .

A screenshot of a computer

AI-generated content may be incorrect.

*Figure no 7*A screenshot of a computer

AI-generated content may be incorrect.*Figure no 8*

**Data Summarization**

I summaries the data so that it is useful to make decision. a field and columns which are not required I removed them to make a concise report in the format of (.csv) for the user and (.parquet) file for the system

A screenshot of a computer

AI-generated content may be incorrect. *Figure no 9*

A screenshot of a computer

AI-generated content may be incorrect. *figure no 10*

*Figure no 11*

A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.*Figure no 12*

Final output of both the report for the user and system

A screenshot of a computer

AI-generated content may be incorrect.*Figure no 13*A screenshot of a computer

AI-generated content may be incorrect.*Figure no 14*

|  |  |  |
| --- | --- | --- |
| **DAP Estimated Cost** | **Monthly cost(USD)** | **Total 12 months cost(USD)** |
| **Team member 2** | | |
| Data Cleaning & Data Summarization | 3.10\*12  (month\*12) | 37.20USD |
|  |  |

*A screenshot of a computer

AI-generated content may be incorrect.  
Figure* no 15

**References**

Dataset – permit issued by city –

<https://opendata.vancouver.ca/explore/dataset/issued-building-permits/export/?dataChart=eyJxdWVyaWVzIjpbeyJjaGFydHMiOlt7InR5cGUiOiJsaW5lIiwiZnVuYyI6IkFWRyIsInlBeGlzIjoicGVybWl0ZWxhcHNlZGRheXMiLCJzY2llbnRpZmljRGlzcGxheSI6dHJ1ZSwiY29sb3IiOiJyYW5nZS1jdXN0b20ifV0sInhBeGlzIjoiaXNzdWVkYXRlIiwibWF4cG9pbnRzIjpudWxsLCJ0aW1lc2NhbGUiOiJ5ZWFyIiwic29ydCI6IiIsInNlcmllc0JyZWFrZG93biI6InBlcm1pdGNhdGVnb3J5IiwiY29uZmlnIjp7ImRhdGFzZXQiOiJpc3N1ZWQtYnVpbGRpbmctcGVybWl0cyIsIm9wdGlvbnMiOnt9fX1dLCJkaXNwbGF5TGVnZW5kIjp0cnVlLCJhbGlnbk1vbnRoIjp0cnVlfQ%3D%3D>

*Understanding and managing Amazon S3 storage classes*.

<https://docs.aws.amazon.com/AmazonS3/latest/userguide/storage-class-intro.html> *creating the job* - <https://docs.aws.amazon.com/databrew/latest/dg/getting-started.01.html>

*cleaning , and preparation job*.

<https://docs.aws.amazon.com/prescriptive-guidance/latest/modern-data-centric-use-cases/data-preparation-cleaning.html>

*Data Catalog- aws library*.

<https://aws.amazon.com/solutions/analytics/data-catalog/>

AWS Glue DataBrew.- *Summarization*

<https://docs.aws.amazon.com/databrew/latest/dg/getting-started.02.html>

*aws piecing calculator -* [*https://calculator.aws/#/*](https://calculator.aws/#/)