[Descriptive Analysis 2](#_Toc201726286)

[Project Description 2](#_Toc201726287)

[Project Title 2](#_Toc201726288)

[Objective 2](#_Toc201726289)

[Dataset 3](#_Toc201726290)

[Methodology 4](#_Toc201726291)

[Tools and Technologies 6](#_Toc201726292)

[Deliverables 7](#_Toc201726293)

[AWS Deployment and Service Models 9](#_Toc201726294)

[AWS Cost Analysis 11](#_Toc201726295)

[AWS Global infrastructure 12](#_Toc201726296)

[AWS IAM 13](#_Toc201726297)

[AWS VPC 14](#_Toc201726298)

[AWS Lambda 15](#_Toc201726299)

[AWS EBS 15](#_Toc201726300)

# Descriptive Analysis

The duality of the dataset used in this project is that is has the voting history of the council members. This project aims to conduct a descriptive analysis based on various AWS services in order to learn more about the occurrence of the voting trends, the people who tend to vote similarly, and whether there exist or not any patterns in votes. On the basis of this information, as well, we will know which issues are more consensual and which ones are more contested. The information contains votes cast by various members in different times and different matters. Such kind of analysis can be applicable in informing the council decisions to be more open and clear to the citizens. To explain what the data is saying we will apply simple statistics and visualization tools. This is to say, we shall get the number of voters and the voting they cast as to the question yes or no and the number of times a particular issue appeared.

Project Description

This project concerns the perception of the council voting records with cloud-based supplies and services. Our data will be stored with AWS services, such as S3, consisting of cleaning and transformation steps, where we will employ AWS Glue and AWS Glue DataBrew and query and visualize the data with the help of Athena and QuickSight, respectively. Secure access to data will be assisted by IAM and Key Management Service. CloudWatch will be used as an aid to logging and utilization. In case more work is required, EC2 can also be applied. With these tools, the entire project will exist on the cloud and it will facilitate uploading, analyzing, and processing large data volumes. Our techniques will not be very complicated but rather would aim at cleaning data, organize it well, and making it easy to interpret. This will help us provide users or policy makers with an improved image of how the council votes on various issues in the long run. The last objective is to create a basic but useful report and dashboard on displaying voting patterns.

Project Title

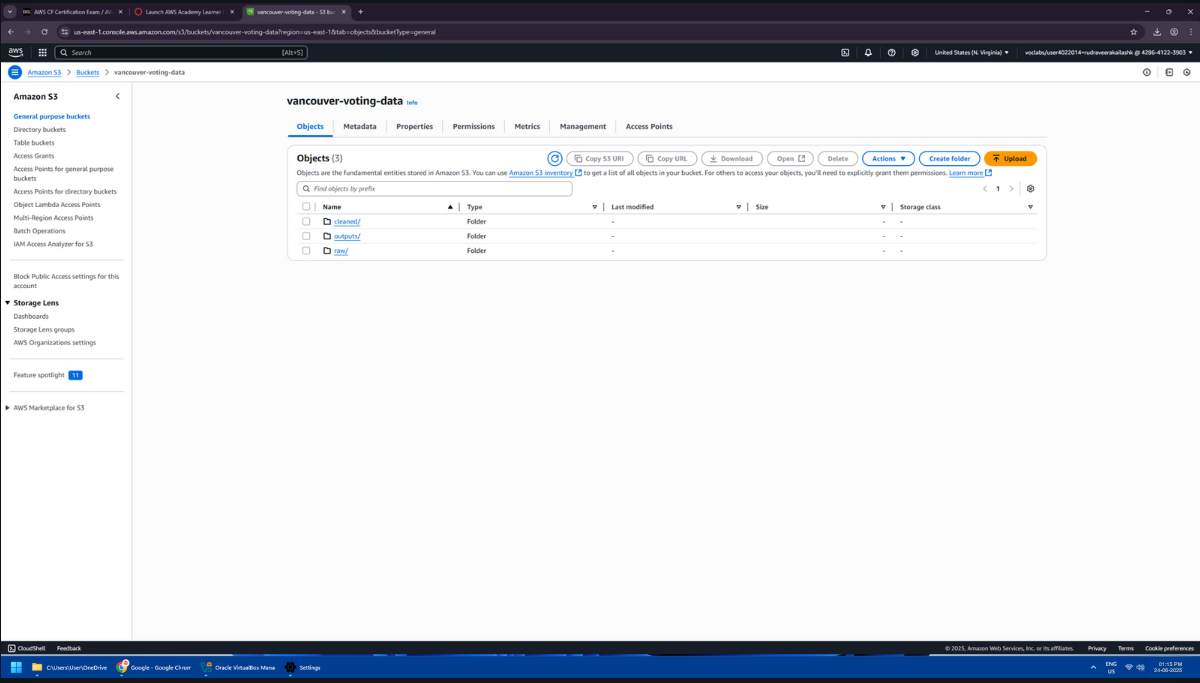
Cloud-Based Analysis of Council Voting Patterns Using AWS Services.

Objective

The overall objective of the project is to research and learn the voting pattern of council members through the cloud. The project is interested to demonstrate the frequency of the people voting similarly and the sort of topic an individual views more or less consensus. It is also concerned with the voting patterns during transition. It could make people understand how the council makes decisions and on which side people stand. The idea is to have the data ready to read and meaningful to the people and the workforce at government. The project also demonstrates how it is possible to work with real data in a modern safe way by means of cloud services.

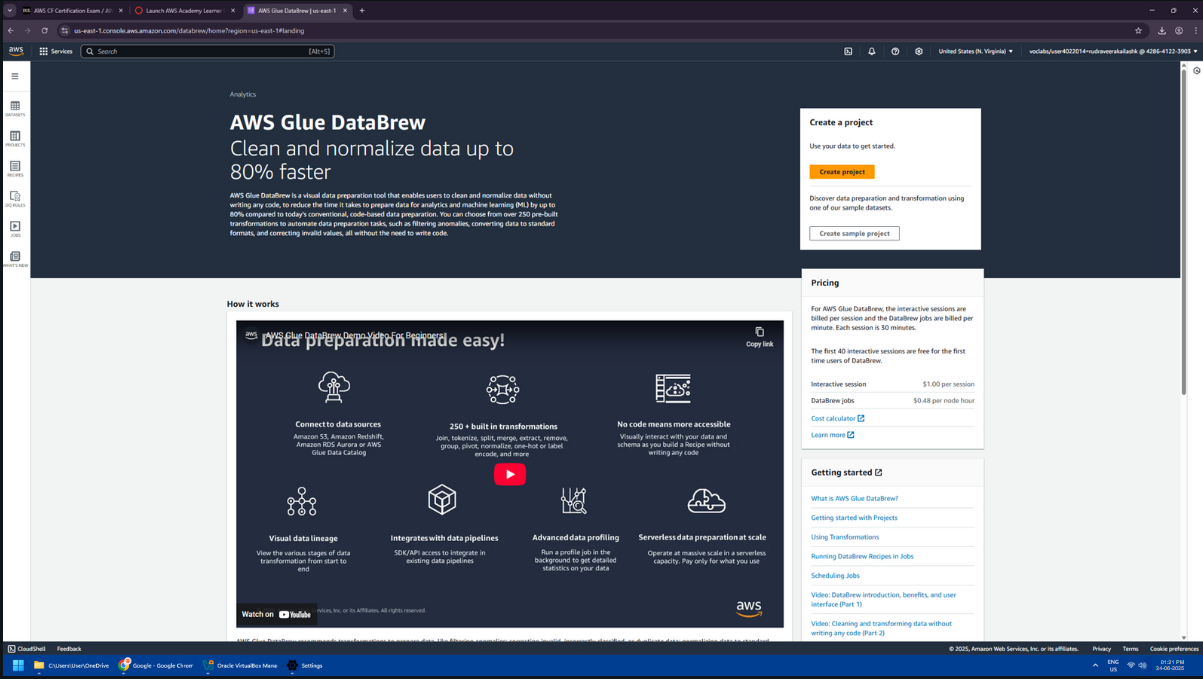
Dataset

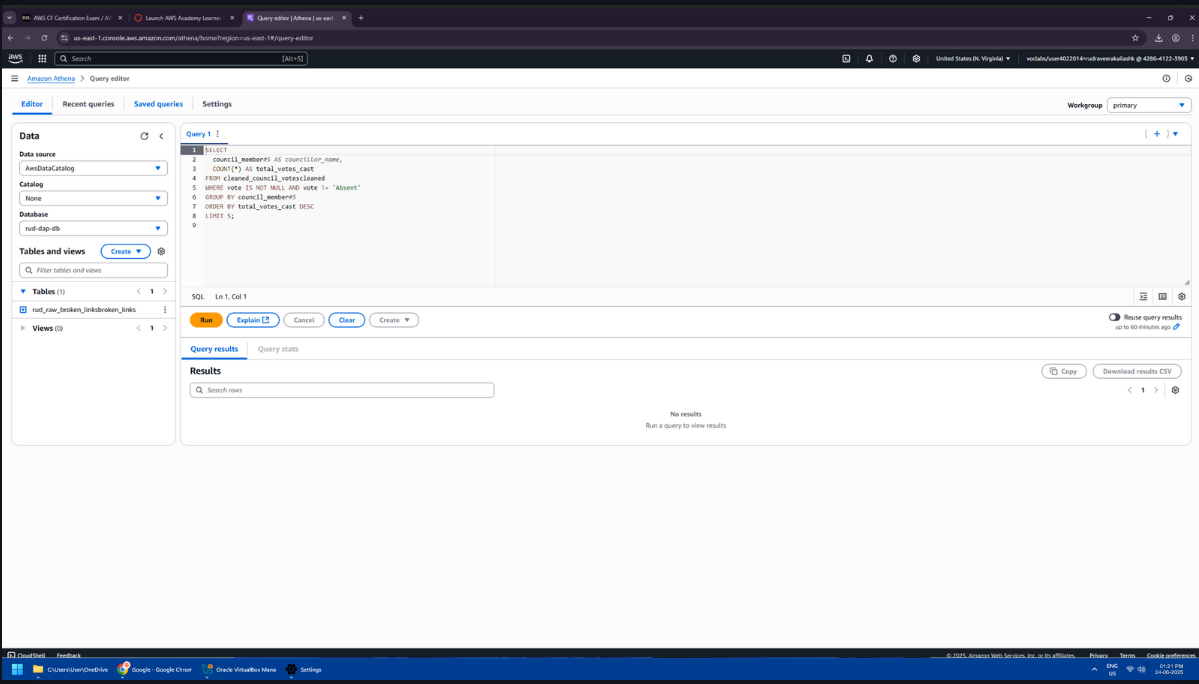
The file used in this project contains the information about voting records of a council. Every record contains the information such as who voted, how they voted and on what. This information is saved in a CSV form and uploaded on Amazon S3 to process. The data is both clean and it contains sufficient data to comprehend the voting behavior. It addresses a lot of various votes and issues. This enables the project to examine the frequency of agreement among the members, the frequency at which they vote as well as the nature of decisions made frequently. Such data will suffice in both the preliminary analysis and detailed analysis when necessary.



## Methodology

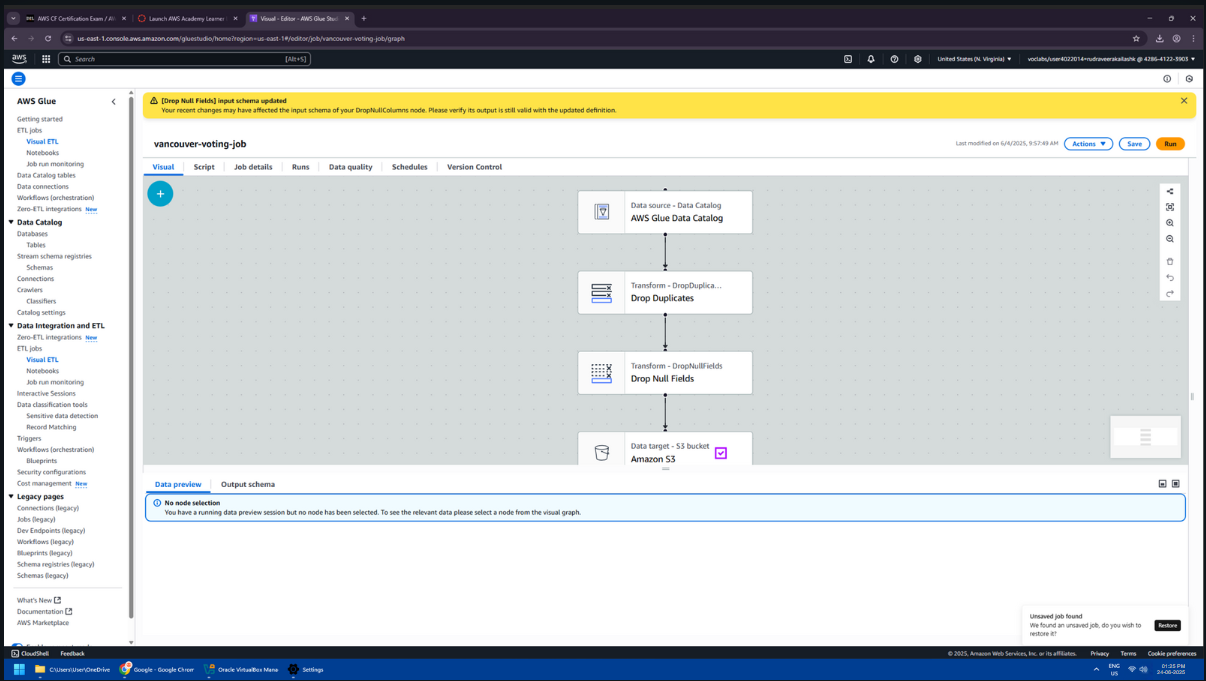
The project will begin by uploading the dataset on Amazon S3 to store safely. Afterward, data preparation takes place by utilizing AWS Glue. Glue assists in the cleaning and transformation of the data into a superior format. Next, a further data cleaning is performed with the AWS Glue DataBrew that uses an easy to use user interface. Next, SQL queries are run by means of Amazon Athena to obtain practical information in the data. Such queries assist in answering questions such as the number of times a person votes or proposes in agreement with other people. IAM and KMS are employed to ensure that only the appropriately disposed can access the data or view the information. Lastly, the outcomes are presented in a report and dashboard so that people can read and comprehend.

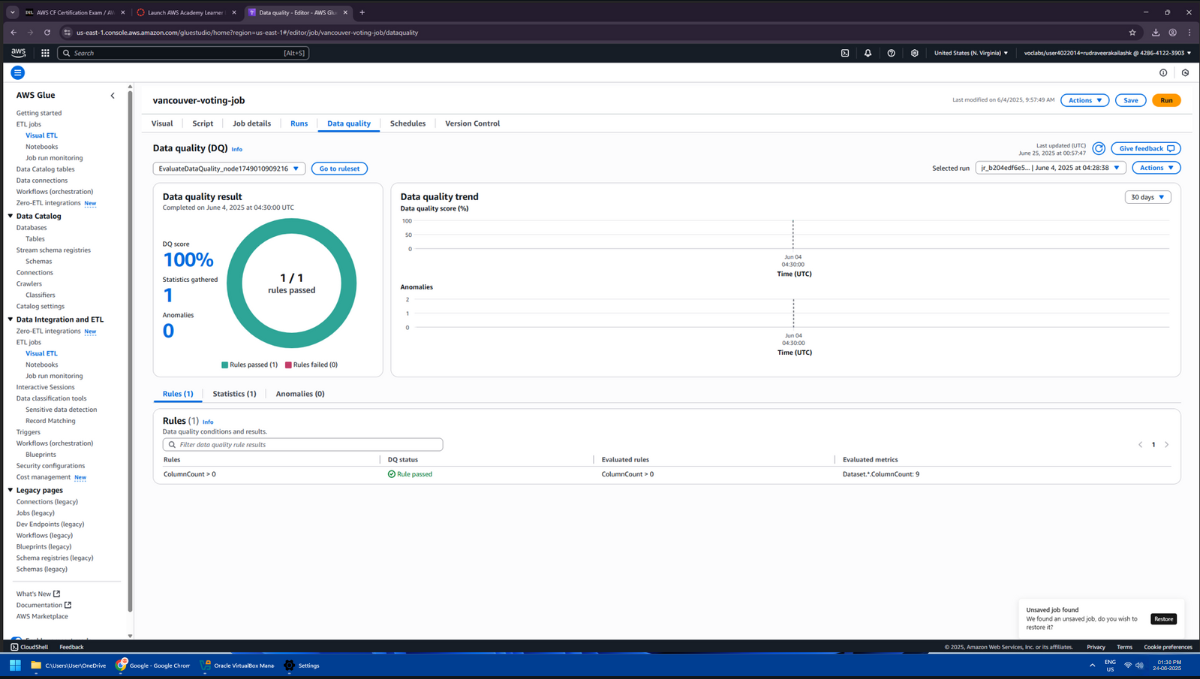




## Tools and Technologies

In this project, numerous tools and services provided by AWS are utilized in order to perform every stage of the work. The data is put in an S3 bucket that is secure. AWS Glue is the platform with the help of which data is prepared and cleaned to utilize it in the analysis. AWS Glue DataBrew assists in the data cleansing process in simple steps and provides more visibility to the dataset. To execute SQL commands directly against the data in the S3 store, you have Amazon Athena. This is useful in extraction of useful results without the movement of data. It also checks the logs and the services performance using the CloudWatch feature which aids in the troubleshooting process within a short time. The use of AWS Identity and Access Management (IAM) and AWS Key Management Service (KMS) have been used to ensure the data is secure and only the authorized personnel will be in a position to access the data. Other jobs may also be supported with Amazon EC2 including the possible running of custom applications. All these tools enable one to perform all the tasks in the cloud end to end. They are also used to make the project simple to operate, and safe to operate upon and can be operated again in future.



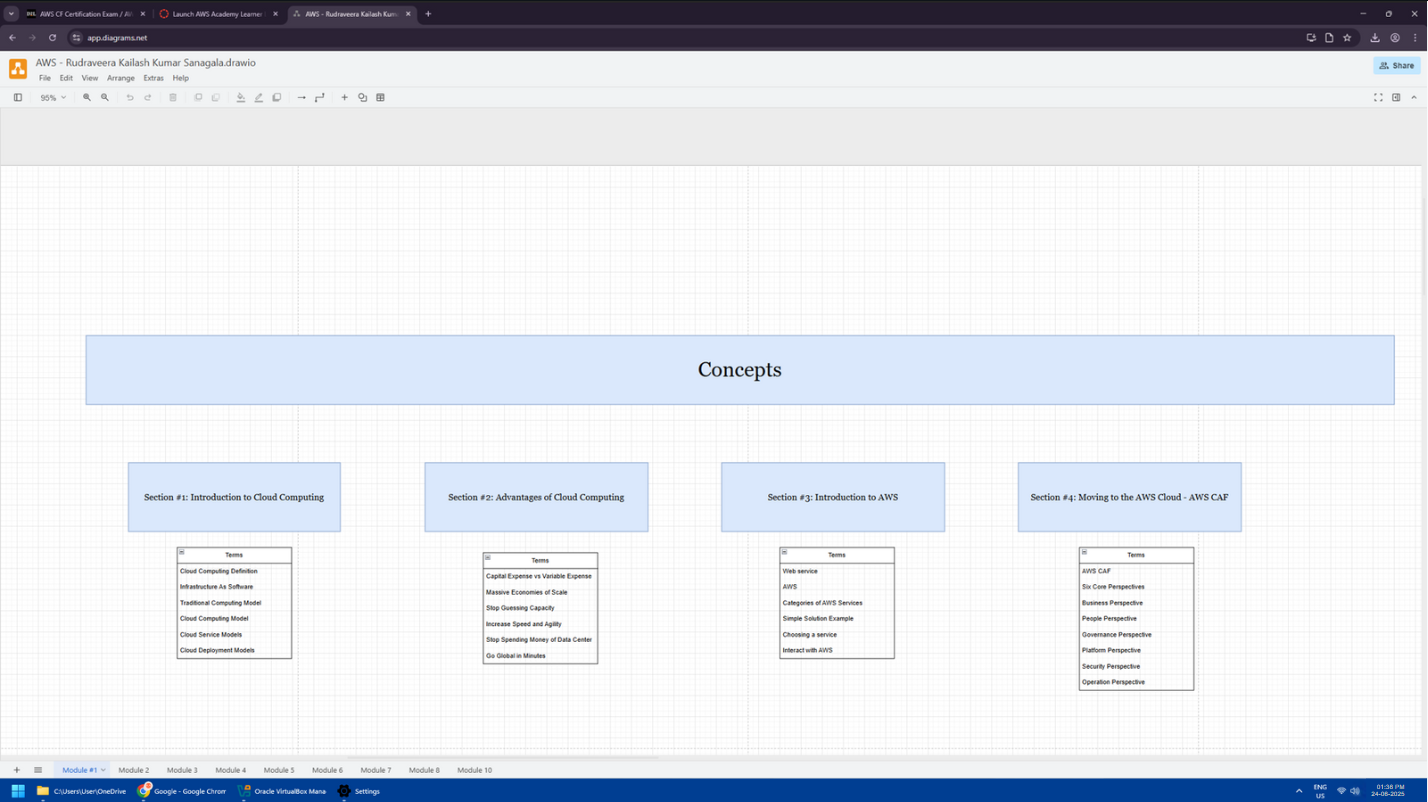


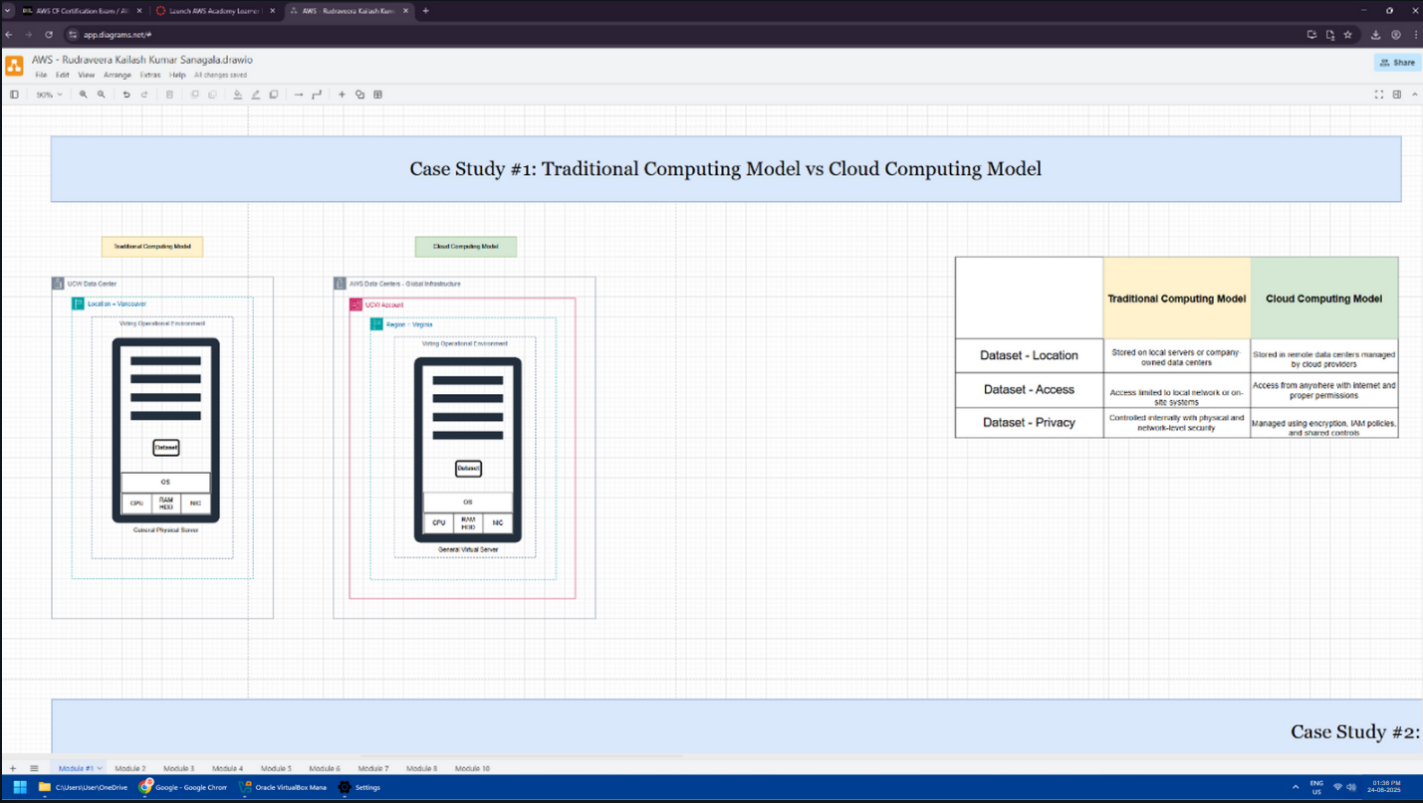
## Deliverables

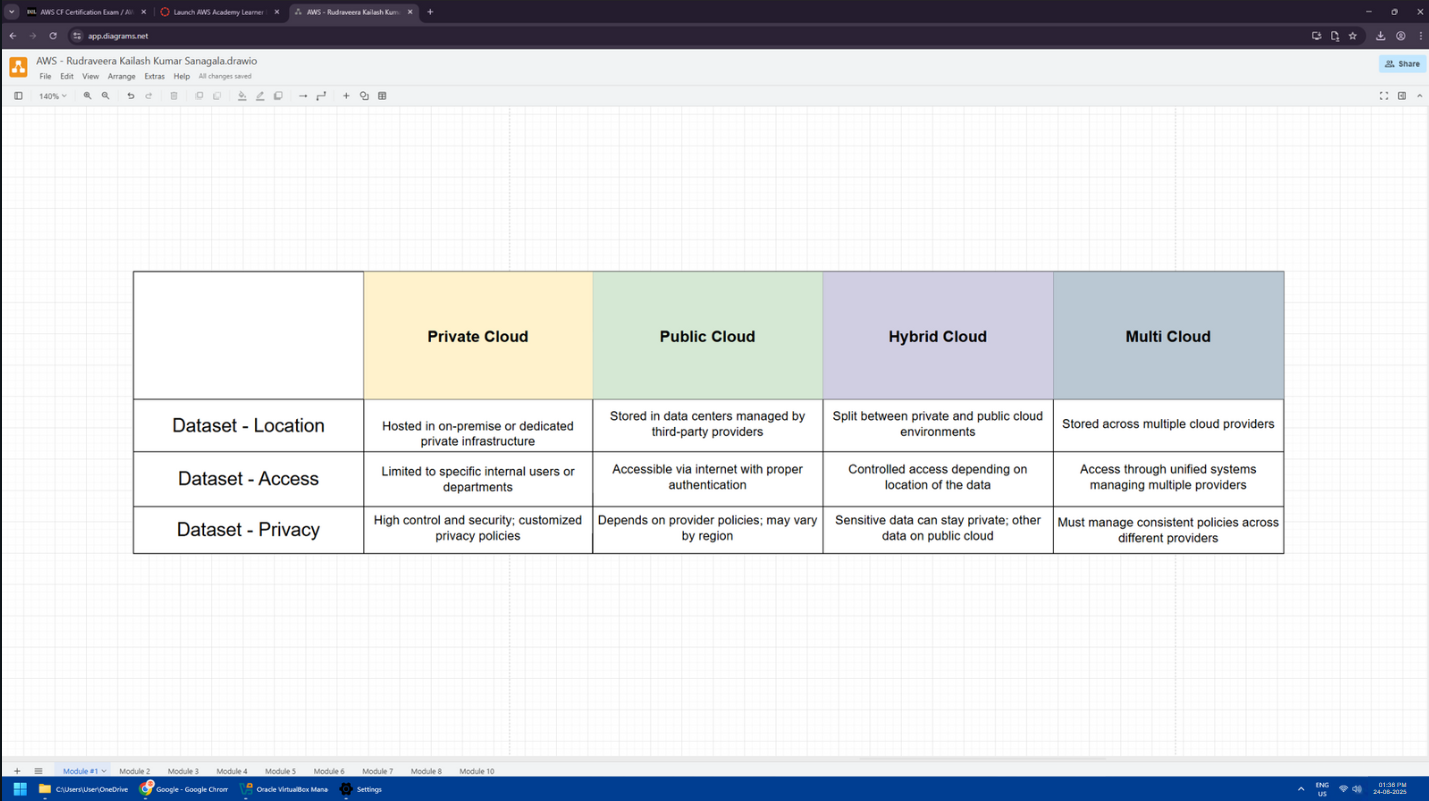
Some of the things that will come out of the project will be useful. First, there will be clean and well-structured data which can be analyzed. This may be applied in other studies. Then there will be summary reports which will display what has been discovered in the data; that is, how many people vote in the same way or what are the topics which are the most supported. Such reports will be readable and straightforward. Finally, SQL queries will also be found in Athena to give different answers to the data. Such searches will assist others to carry out searches in the future. The project will also produce easy visual dashboards in case they are required to make people visualize the trends on voting. All this will make the voting information more productive and comprehensible by any learner who wants to know more.

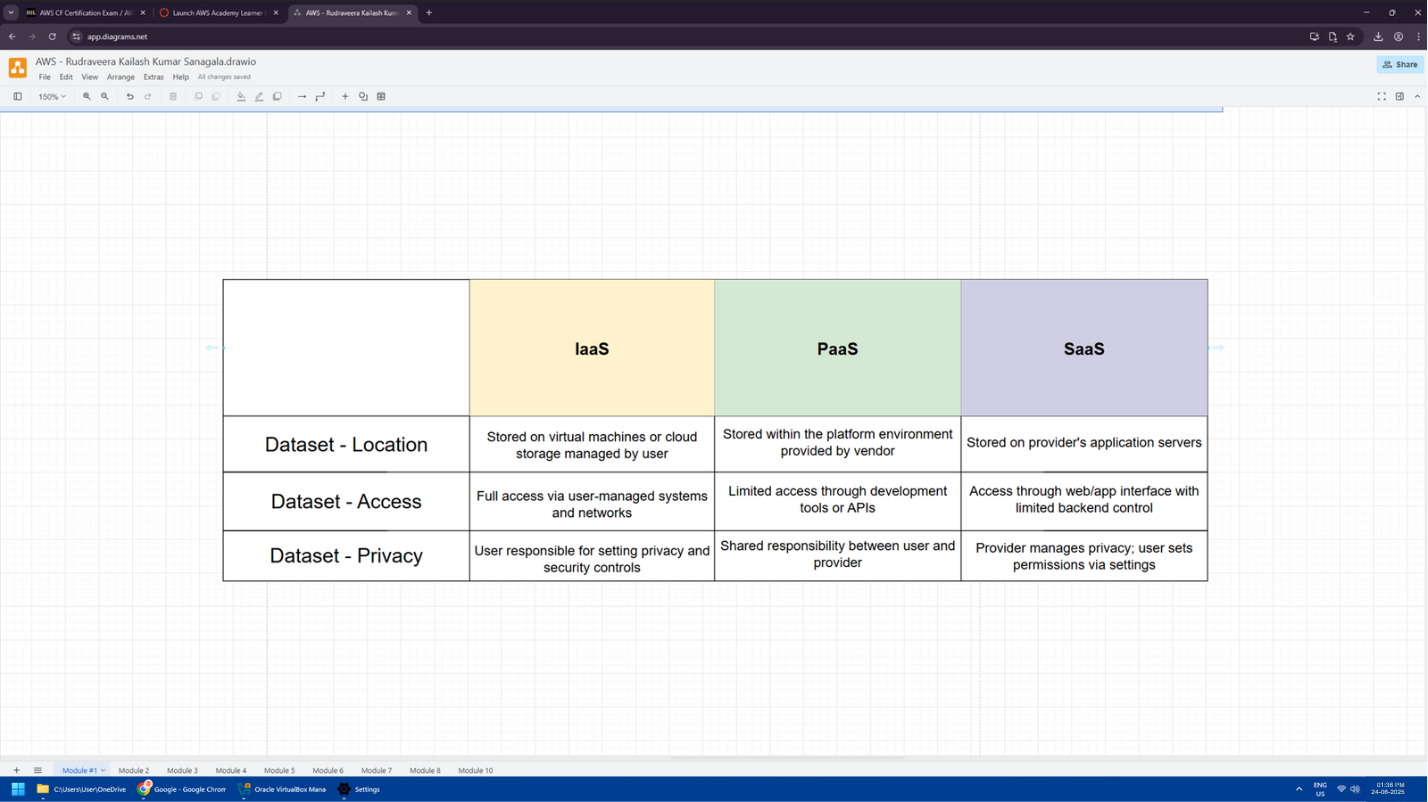
# AWS Deployment and Service Models

In this project the AWS deployment model that will be adopted is the public cloud. The AWS provides all the services such as Amazon S3, AWS Glue, Athena and many other services through the internet. This implies that we do not have to purchase our own servers and install them. We directly make use of cloud services available at AWS. The service model that is employed here is primarily Platform as a Service (PaaS), and Software as a Service (SaaS). Data services such as AWS Glue and Athena enable us to run data jobs and SQL with no need to deal with the hardware or software provisions. This enables the user to deal with the data and not on how to operate servers. These models make us complete the project within a short period and with less difficulties during setup. It is also through the cloud that we get to scale the project in case the data grows larger in the future.



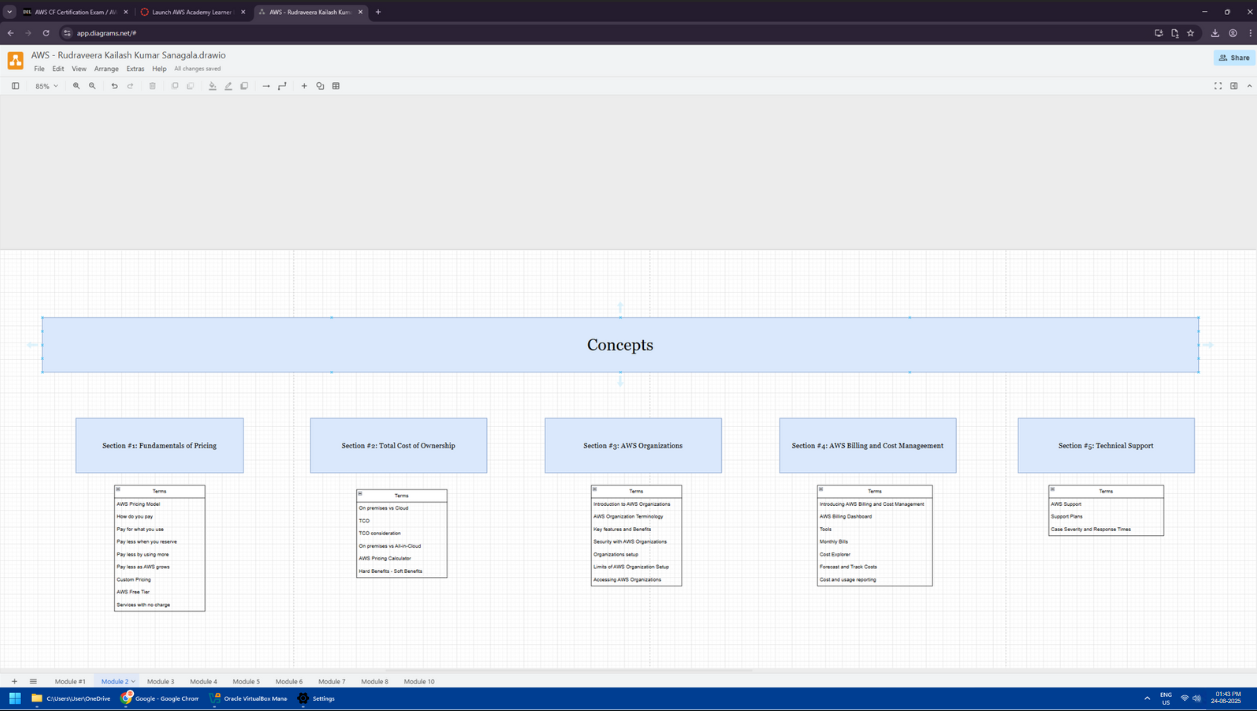






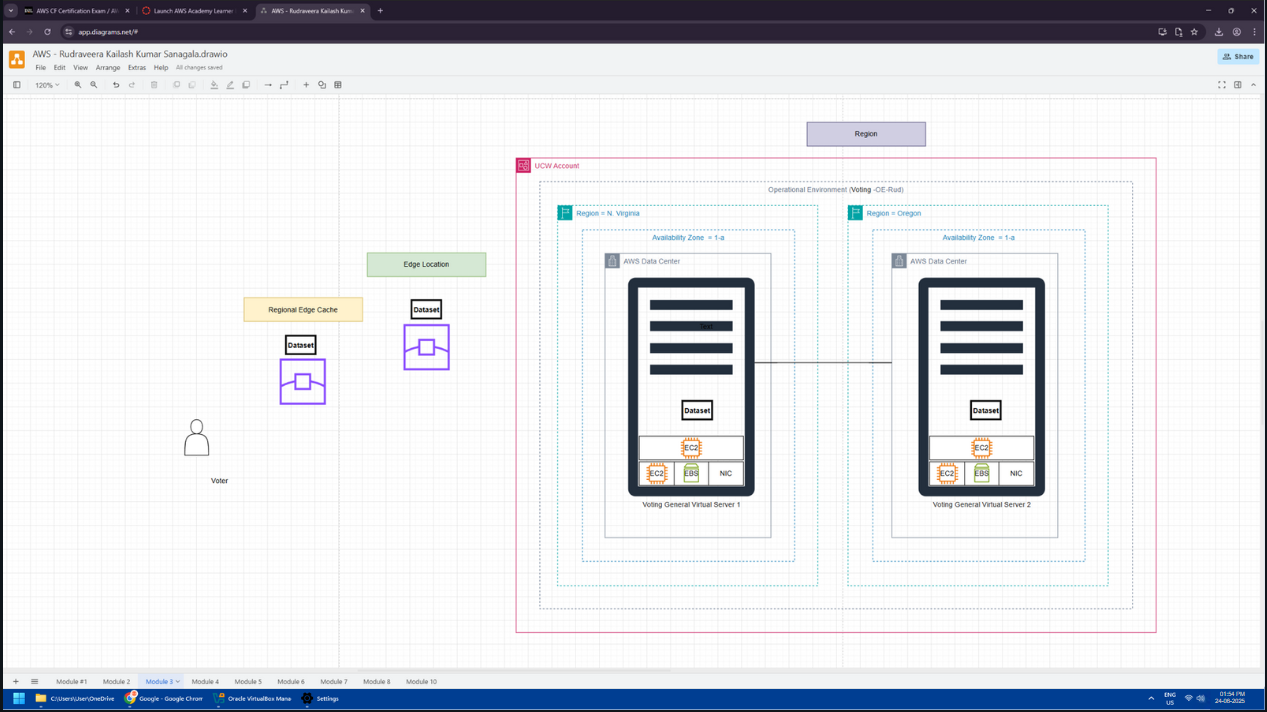
# AWS Cost Analysis

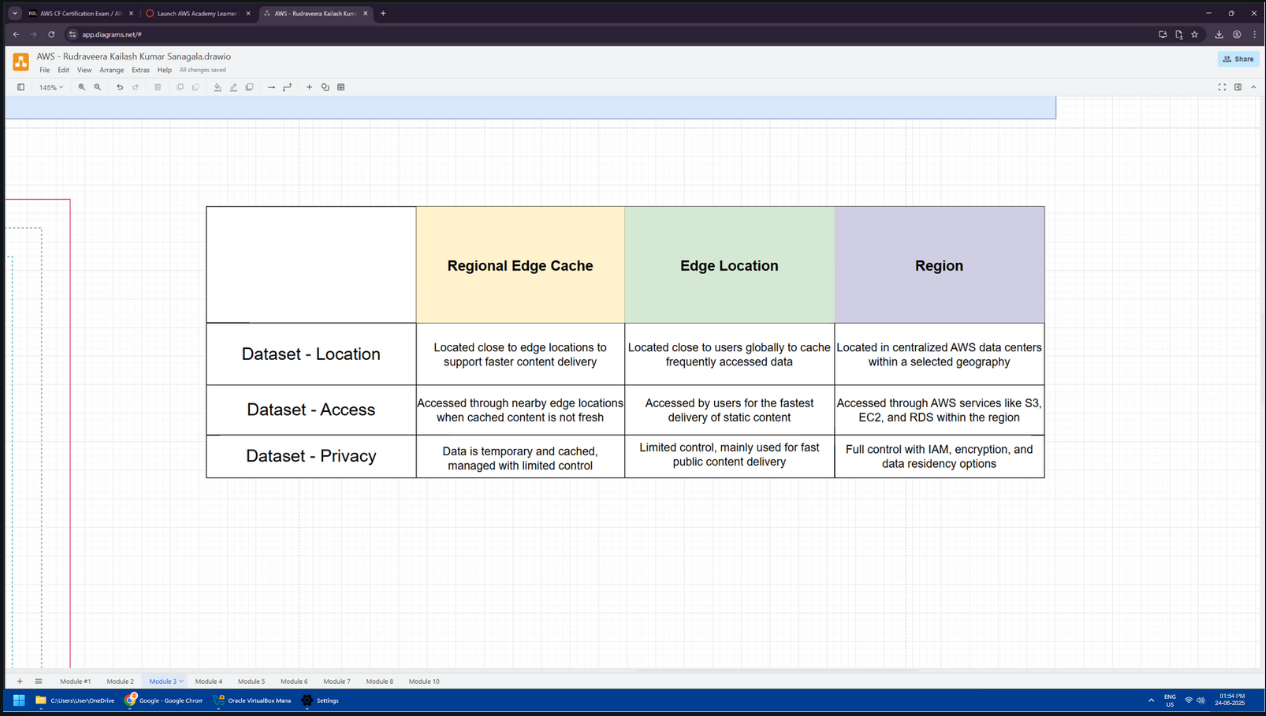
This project cost is calculated depending on the amount consumed of each AWS service. As an example, the non-use of the dataset in the S3 will imply minimal costs since the file is not large after all. The cost of AWS Glue depends on the duration it is running, and the number of jobs to be utilized. Another charge that AWS Glue DataBrew uses is the number of data processing sessions. The Athena charges money when performing a query and this is determined by the amount of data being read by it. The EC2 is only expensive together with other hours we operate it. The basic level of IAM and KMS is free, but it is at a cost when the advanced functionality is implemented. There are free capacities on CloudWatch and additional logs might be more expensive. The cost of this project in general is low-medium since the data will be small as well as the services used in a simple manner. We may also make use of AWS cost control tools with the help of which the cost may be decreased.



# AWS Global infrastructure

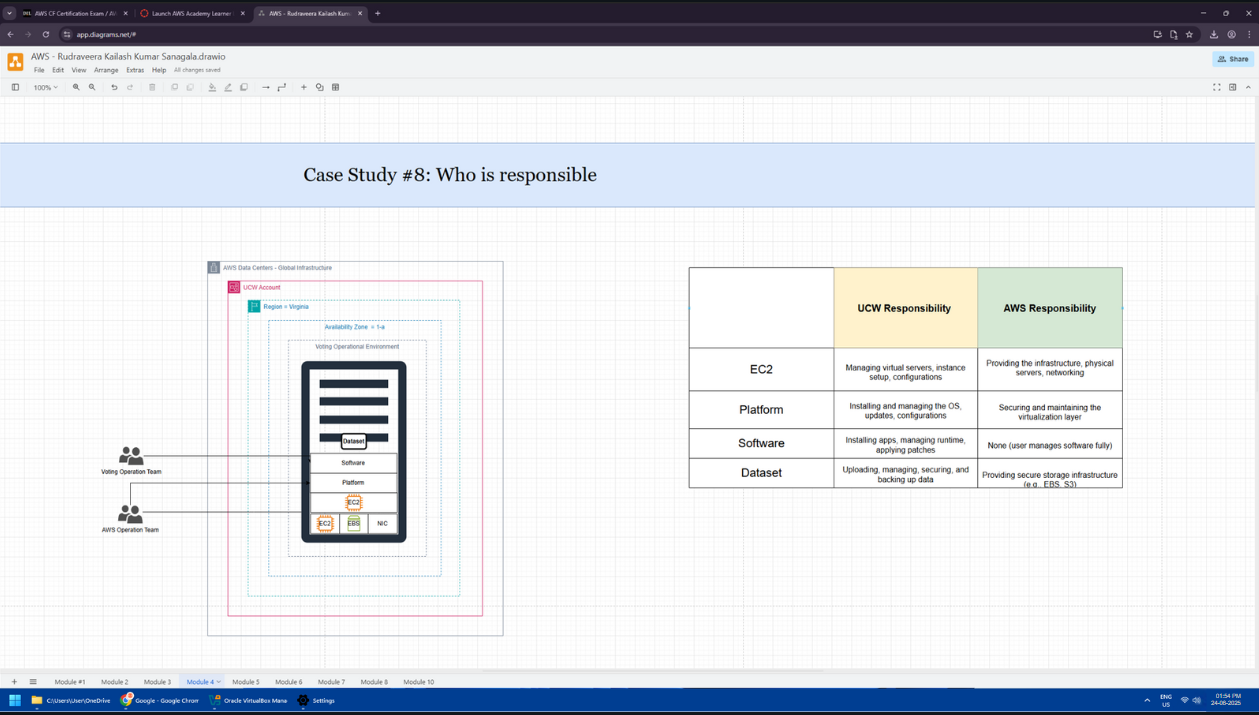
The strength of AWS global infrastructure is high and assists this project to run better. Data centers of AWS are present in numerous segments of the globe. These are aligned in the regions and availability zones. Every region possesses multiple zones and in the case that a specific zone has an issue the remaining zone is still operational. This provides upper availability and proper backup. We have the option of selecting the AWS area that is near the users in this project. This makes the uploading, processing as well as downloading of data quick. It also assists in the adherence of regulations regarding the tenure of data. Utilization of global infrastructure also spells that we have improved security, speed and low delay. It will be useful later when the project grows in size or when additional individuals will use it. Such a robust international network can serve the cloud services, such as S3, Glue, and Athena, very well. At every center, AWS takes care of everything, including power, cooling, and internet connection; hence we do not need to worry. The worldwide arrangement is also beneficial when we are required to expand the project or transfer data to other regions of the planet. It provides a solid foundation to the development of secure and rapid cloud such as this one.





# AWS IAM

In this project, the AWS Identity and Access Management also known as IAM is deployed to manage access by users. It assists in the security of the data and tools. IAM allows creating users and groups and assigning permission to them. Indicatively, we may give just one individual to read the data but not to make any changes whereas another man is capable of running jobs or queries. This assists in handling the project in a better manner and prevents misuse of the data. IAM also has role-based support, so we can provide a temporary access when required such as a job on EC2 or Glue. Multi-factor login can be employed as well in order to make it safer. IAM also logs all activities being carried out by every person, and in the event there is an issue, it is possible to ascertain who did what. In this project, IAM ensures that only the correct users are able to access the S3, Glue, Athena, and other available services. It collaborates with KMS to safeguard information using encryption. The use of IAM is an important component of the project since it ensures the development of trust and safety in case of real work with cloud tools.



# AWS VPC

VPC or AWS Virtual Private Cloud is a subsystem of the AWS network that allows us to manage the area in which our services are being executed. It also aids us to establish a personal network within the cloud on our project. VPC would be relevant in this project in case we are utilising the services such as EC2 or we desire additional flexibility in data relating to flow between the tools. Using VPC, we are able to configure matters such as IP addresses, subnets as well as gateways. This makes maintenance of the project more secure and organized. Engaging security groups and access rules we can permit only some traffic out of or into the VPC. In this manner, Project is secured against the external attacks. VPN can also be used to connect to on-premise networks through VPC in case one is required to do so. Although most of these AWS services are relatively high level and do not require detailed control of the network, VPC can be still helpful in the future additional changes or larger projects, or even running secure EC2 jobs. It provides an excellent basis to handle cloud resources in a secure and exclusive fashion.

# AWS Lambda

AWS Lambda is a service allowing you to run code without scaling, provisioning and administering servers. AWS Lambda, in this project, will suffice to perform small tasks, such as checking the data, converting format, or triggering an action upon adding a new file to S3. As a use case, we may be interested in triggering a Glue job or send a message when a new dataset is uploaded. It just charges when operating, thus we do not settle money when not in operation. This makes Lambda a versatile as well as economical utility. It has numerous language support such as Python, Node.js and Java. Integration with other services on AWS can be enabled simply with the help of Lambda. In our project he might not be in front processing but Lambda is quite helpful in smaller automated task which enables the entire thing to flow smoothly. It does not require that much work but makes the system more efficient.

# AWS EBS

EBS or Amazon Elastic Block Store is applied in storing data that executes with EC2. It becomes such a hard drive in the cloud. EBS is not the copy of the primary storage in this project since we make use of S3 as the primary storage. However, in case we intend to perform special tasks with the help of EC2 or deploy software requiring local storage, we will require EBS. It may contain files, logs or even intermediate outcomes of data analysis. EBS provides rapid and readily accessible storage and we also have the option of adjusting according to the level of large or rapid we require. It also stores data in collections including backups and snapshots. In case EC2 is halted then the EBS data can be retained and we can reuse it. EBS is useful in the present project to sponsor local ones, or operations that are not efficiently operated on S3. It allows increased control and aids when additional storage facilities are required. Therefore, despite the lack of its use in the beginning, EBS is an effective tool to expand the project in the future.