**ASSIGNMENT – 2**

A screenshot of a computer

Description automatically generated

Architecture Diagram

Data Upload to S3 Buckets: The process begins with uploading files from a local machine into AWS Cloud. These files are stored in S3 buckets. S3 (Simple Storage Service) is a scalable

IAM Role Creation: Before AWS Glue can access the data in the S3 buckets, an IAM (Identity and Access Management) service role needs to be created. This role serves as a way to securely delegate permissions to AWS services. In this case, we create a custom IAM role that includes the AWSGlueServiceRole policy along with a custom policy that grants specific access to the S3 buckets containing the data. This ensures that AWS Glue has the necessary permissions to access and process the data in the S3 buckets.

Database Creation: After setting up the IAM role, a database is created. This database will be used to store the reference table generated by the AWS Glue crawler. In AWS Glue, a database is a logical container for tables.

Crawler Configuration: With the IAM role and database in place, an AWS Glue crawler is configured. The purpose of the crawler is to analyze the data in the S3 buckets and create metadata tables in the AWS Glue Data Catalog. The crawler examines the data schema and generates a reference table based on the structure of the data.

Reference Table Creation: Once the crawler has completed its job, a reference table is created in the database. This table contains metadata information about the data stored in the S3 buckets. This metadata includes details such as column names, data types, and file locations. The reference table serves as a convenient way to query and analyze the data using tools like Amazon Athena or to perform transformations using AWS Glue jobs.

Overall, this system engineering flow diagram outlines the process of ingesting data from local files into S3 buckets, configuring IAM roles to provide access permissions, creating a database to store metadata, using an AWS Glue crawler to analyze the data and generate a reference table, and finally utilizing this reference table for querying and transformation purposes.

JSON Code Explanation:

Version: Specifies the version of the policy language. In this case, it's "2012-10-17", which is the date when the policy language version was released.

Statement: This is an array containing one or more individual policy statements. Each statement represents a set of permissions.

Action: Lists the AWS actions allowed or denied by this statement. Actions are specific tasks that can be performed on AWS resources. In this policy, actions related to S3 operations such as Get, Put, Delete, List, etc., are allowed.

Resource: Specifies the AWS resources to which the actions apply. The "arn" (Amazon Resource Name) is used to uniquely identify AWS resources. In this policy, it allows access to specific S3 buckets (query-result-65d41890, glue-bucket-65d41890) and any objects within those buckets. It also allows access to S3 buckets with names starting with query-result- and glue-bucket-.

Effect: Indicates whether the specified actions are allowed or denied. In this policy, the effect is set to "Allow", meaning the actions listed under each statement are permitted for the specified resources.

Overall, this policy grants permissions to perform various S3 operations on specific buckets and objects, as well as on S3 buckets with specific naming patterns. It ensures that the AWS Glue service, or any entity associated with this IAM policy, can interact with the specified S3 resources according to the defined permissions.