The Data ingestions process simply involves 3 services from AWS like AWS S3, AWS IAM, AWS Glue

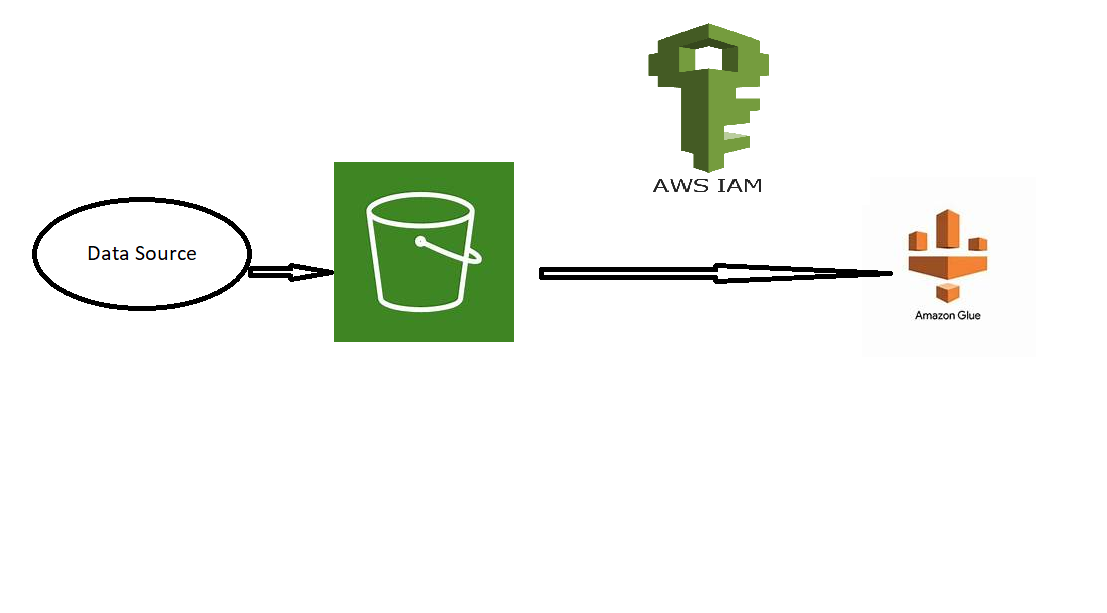
AWS S3 acts as the data source which stores csv file for the ingestion process

AWS IAM acts as security between 2 services by creating roles and policies by adding arn of S3 service in role & policies

AWS glue acts as data ingestion engine to ingest files from S3 to database as reference tables through running glue crawler

Reference Table which are accessed through other AWS services like AWS athena for querying according the business requirements

Design flow



Lets see the detailed process from creation, upload & other steps involved

1. Creating bucket in S3 :

Description: Amazon S3 provides scalable object storage, allowing storage and retrieval of any amount of data from

Anywhere on the web

Terminology : An S3 bucket is the fundamental container within AWS S3 that holds your data. Its similar to folder in file

System but it's a top-level container. Buckets can be thought of as a root folder that can store objects, which could be

anything from images, videos, documents, or even raw data files.

1. In the S3 bucket, create a folder:   
   Provider: Amazon S3  
   Despite the lack of conventional folders in S3, object keys may be used to mimic a folder structure.   
     
   Terminology: In order to create a "folder" in S3, one must first create an object with a key that ends in a slash ("/"). This object may be created programmatically or visually within the S3 interface.
2. Put CSV files into the designated folder:   
   Provider: Amazon S3   
   Description: Files can be uploaded into the folder when it is created. Any kind of file, including CSV files, may be used for these.   
   Terminology: When files are uploaded, they are stored in the designated folder of the S3 bucket, with a unique key assigned to each file.
3. Open the Identity and Access Management (IAM) page.  
   AWS Identity and Access Management (IAM) is the service.  
   The safe administration of access to AWS resources and services is made possible via IAM.  
   Terminology: Users, groups, roles, and policies are examples of IAM entities. These entities manage permission and authentication for AWS resources.
4. Establish an IAM role.  
   Provided: I am  
   IAM roles delineate a range of authorizations for submitting requests to AWS services, which are intended to be assumed by authorized entities such as AWS services or users originating from another account.  
   Words used: Permissions to act on resources in AWS environments are granted with the assistance of roles.
5. Draft an IAM Policy:  
   Provided: IAM  
   IAM policies specify which resources, activities, and circumstances are subject to which permissions.  
   Terminology: Policies are JSON documents that outline the resources that can be accessed by whom and under what circumstances.
6. Attach IAM policy in the IAM role:  
   Provided: I am  
   When a policy is attached to a role, the role is given the permissions specified in the policy.  
   Terminology: By defining the role's permission boundaries and the actions it may take on which resources, this action sets the role's permissions.
7. Open AWS Glue: Service. The purpose of AWS Glue is to prepare and load data for analysis. It is a fully managed ETL (extract, transform, load) service.  
   Terminology: To load, transform, and crawl data into data lakes and warehouses, utilize Glue's capabilities.
8. Use AWS Glue to create a database:  
   AWS Glue is the service.  
   Glue databases serve as logical containers for tables, facilitating the management and organization of data information.  
   Terminology: Related tables are grouped and classified by databases.
9. In the database, make a table:  
   Provider: AWS Glue Description: SQL or other query languages may be used to query structured data stored in Glue tables.  
   Words used: Tables, which include column definitions and metadata, establish the schema and organization of data stored in underlying data sources such as S3.
10. Use AWS Glue to create a crawler that will find and catalog the data:  
    AWS Glue is the service.  
    Glue crawlers are automated procedures that generate metadata tables in the Glue Data Catalog, infer schemas, and scan data sources.  
    Terminology: Crawlers make it easier to retrieve and analyze data by automatically finding and classifying material that is stored in a variety of forms and places.