

Create an S3 bucket:

Service: Amazon Simple Storage Service (Amazon S3)

Description: Amazon S3 is a scalable object storage service offered by AWS. It allows you to store and retrieve any amount of data from anywhere on the web.

Terminology: A bucket is a container for storing objects. Each object in S3 is stored in a bucket and is identified by a unique key within the bucket.

Create a folder in S3 bucket:

Service: Amazon S3

Description: In S3, there are no folders like in a traditional file system, but you can emulate a folder structure using object keys.

Terminology: Creating a "folder" in S3 is essentially creating an object with a key ending with a slash ("/"). This creates a visual representation of a folder within the S3 console or when accessing S3 programmatically.

Upload CSV files into the folder:

Service: Amazon S3

Description: Once the folder is created, you can upload files into it. These files can be of any type, in this case, CSV files.

Terminology: Uploading files involves storing objects in the specified folder within the S3 bucket. Each file is identified by a unique key.

Go to IAM (Identity and Access Management):

Service: AWS Identity and Access Management (IAM)

Description: IAM enables you to manage access to AWS services and resources securely.

Terminology: IAM users, groups, roles, and policies are used to control who is authenticated (signed in) and authorized (has permissions) to use resources.

Create an IAM role:

Service: IAM

Description: An IAM role is an entity that defines a set of permissions for making AWS service requests.

Terminology: Roles are similar to users but are meant to be assumed by authorized entities, such as AWS services or users from another AWS account.

Create an IAM policy:

Service: IAM

Description: An IAM policy is a document that defines permissions for actions, resources, and conditions.

Terminology: Policies are JSON documents that specify who has access to which resources and under what conditions.

Attach the IAM policy to the IAM role:

Service: IAM

Description: Attaching an IAM policy to a role grants the permissions defined in that policy to the role.

Terminology: This establishes the permissions boundary for the role, dictating what actions it can perform on which resources.

Navigate to AWS Glue:

Service: AWS Glue

Description: AWS Glue is a fully managed extract, transform, and load (ETL) service that makes it easy to prepare and load data for analysis.

Terminology: Glue provides a console interface and APIs for data preparation tasks such as crawling, transforming, and loading data into data lakes and data warehouses.

Create a database in AWS Glue:

Service: AWS Glue

Description: A database in Glue is a logical container for tables. It helps organize and manage metadata about your data.

Terminology: Databases in Glue are used to categorize and group related tables, making it easier to organize and query data.

Create a table within the database:

Service: AWS Glue

Description: A table in Glue represents structured data that can be queried using SQL or other query languages.

Terminology: Tables in Glue define the schema and structure of the data stored in underlying data sources like S3. They contain column definitions and metadata.

Create a crawler in AWS Glue to discover and catalog the data:

Service: AWS Glue

Description: A crawler in Glue is an automated process that scans various data sources, infers schema, and creates metadata tables in the Glue Data Catalog.

Terminology: Crawlers help automate the process of discovering and cataloging data stored in various formats and locations, making it easier to query and analyse.

End of the process:

Description: This marks the completion of the defined steps for setting up the data ingestion and processing pipeline.

Terminology: Once the setup is complete, data can be accessed, transformed, and analysed according to the requirements of the application or use case.