

Create an S3 bucket:

Service: Amazon Simple Storage Service (Amazon S3)

AWS provides a scalable object storage solution called Amazon S3. You can use it to store and retrieve any volume of data from any location on the internet.

An object can be stored in a bucket. Every object in S3 is kept in a bucket and is recognized by a special key that is only found in the bucket.

Create a folder in S3 bucket:

Service: Amazon S3

The standard file system's folder structure cannot be replicated in S3, but object keys can be used to do so.

In S3, generating an object with a key that ends in a slash ("/") is equivalent to creating a "folder". When utilizing S3 programmatically or through the S3 interface, this generates a visual representation of a folder.

Upload CSV files into the folder:

Service: Amazon S3

After the folder is made, files can be uploaded into it. These files can be any kind; in this example, CSV files are the kind.

When you upload files, you store items in the designated S3 bucket folder. Every file is uniquely recognized by a key.

Go to IAM (Identity and Access Management):

Service: AWS Identity and Access Management (IAM)

You can securely control who has access to AWS resources and services with IAM.

To manage who is authenticated (signed in) and permitted (has permissions) to use resources, IAM users, groups, roles, and policies are utilized.

Create an IAM role:

Service: IAM

An entity that specifies a set of permissions for submitting requests for AWS services is known as an IAM role.

Roles are like users, except they are intended to be assumed by authorized entities, such as AWS services or users from other AWS accounts.

Create an IAM policy:

Service: IAM

A document that specifies authorization for resources, activities, and conditions is known as an IAM policy.

Policies are JSON documents that define which resources can be accessed by whom and under what circumstances.

Attach the IAM policy to the IAM role:

Service: IAM

When an IAM policy is attached to a role, the role is given the permissions specified in that policy.

By limiting the activities, the role can do on certain resources, this sets the role's permissions boundary.

Navigate to AWS Glue:

Service: AWS Glue

The fully managed extract, transform, and load (ETL) service AWS Glue simplifies the process of getting data ready for analysis and loading it into the system.

For data preparation operations including crawling, converting, and putting data into data lakes and warehouses, Glue offers a console interface and APIs.

Create a database in AWS Glue:

Service: AWS Glue

Tables in Glue have a logical home: a database. It aids in the management and organization of your data's metadata.

To make data organization and querying easier, Glue uses databases to classify and group relevant tables.

Create a table within the database:

Service: AWS Glue

Structured data that can be accessed through SQL or other query languages is represented by a table in Glue.

The structure and schema of the data kept in underlying data sources like S3 are specified via tables in Glue. They provide metadata and column definitions.

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

Service: AWS Glue

A Glue crawler is an automated process that builds metadata tables in the Glue Data Catalog, infers schema, and analyzes several data sources.

To make data easier to retrieve and analyze, crawlers assist in automating the process of finding and classifying data kept in a variety of forms and places.

END