Certainly! Let's summarize the process step by step, highlighting the services involved and their respective terminologies:

1. Create an S3 bucket:

Service: Amazon Simple Storage Service (Amazon S3)

Description: Amazon S3 provides scalable object storage, allowing storage and retrieval of any amount of data from anywhere on the web.

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

2. Create a folder in the S3 bucket:

Service: Amazon S3

Description: Although S3 doesn't have traditional folders, you can simulate a folder structure using object keys.

Terminology: Creating a "folder" in S3 involves creating an object with a key ending with a slash ("/"), visually representing a folder within the S3 console or programmatically.

3. Upload CSV files into the folder:

Service: Amazon S3

Description: Once the folder is created, files can be uploaded into it. These files can be of any type, such as CSV files.

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

4. Go to IAM (Identity and Access Management):

Service: AWS Identity and Access Management (IAM)

Description: IAM enables secure management of access to AWS services and resources.

Terminology: IAM entities include users, groups, roles, and policies, which control authentication and authorization to AWS resources.

5. Create an IAM role:

Service: IAM

Description: IAM roles define a set of permissions for making AWS service requests, meant to be assumed by authorized entities like AWS services or users from another account.

Terminology: Roles help grant permissions for actions on resources within AWS environments.

6. Create an IAM policy:

Service: IAM

Description: IAM policies define permissions for actions, resources, and conditions.

Terminology: Policies are JSON documents specifying who has access to which resources under defined conditions.

7. Attach the IAM policy to the IAM role:

Service: IAM

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

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

8. Navigate to AWS Glue:

Service AWS Glue

Description AWS Glue is a fully managed ETL (extract, transform, load) service for preparing and loading data for analysis.

Terminology: Glue provides tools for crawling, transforming, and loading data into data lakes and data warehouses.

9. Create a database in AWS Glue:

Service: AWS Glue

Description: A Glue database is a logical container for tables, helping organize and manage metadata about the data.

Terminology: Databases categorize and group related tables, facilitating organization and querying of data.

10. Create a table within the database:

Service: AWS Glue

Description: Glue tables represent structured data that can be queried using SQL or other query languages.

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

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

Service: AWS Glue

Description: Glue crawlers are automated processes that scan data sources, infer schemas, and create metadata tables in the Glue Data Catalog.

Terminology: Crawlers automate the discovery and cataloging of data stored in various formats and locations, simplifying querying and analysis.

12. End of the process:

Description: Marks the completion of setting up the data ingestion and processing pipeline, enabling data access, transformation, and analysis as required by the application or use case.