AWS Lambda Role: `de-on-youtube-raw-us-east-1-s3-lambda-role`

Role Overview

The `de-on-youtube-raw-us-east-1-s3-lambda-role` is an AWS Identity and Access Management (IAM) role designed to allow Lambda functions to interact with Amazon S3 and AWS Glue. This role is essential for Lambda-based serverless workflows, enabling these functions to read from, write to, and process data stored in S3, as well as interact with the Glue Data Catalog.

Policy Attachments

1. AmazonS3FullAccess

- **Type**: AWS Managed Policy
- **Purpose**:
  - Provides full access to Amazon S3, enabling Lambda functions to:
    - Read data files from S3 buckets.
    - Write processed data back to S3.
    - List objects in S3 buckets for identifying input and output paths.
- **Key Permissions**:
  - `s3:GetObject`
  - `s3:PutObject`
  - `s3:ListBucket`
  - `s3:DeleteObject`

---

2. AWSGlueServiceRole

- **Type**: AWS Managed Policy
- **Purpose**:
  - Grants Lambda functions the ability to interact with the AWS Glue service, such as:
    - Querying the Glue Data Catalog for table definitions.
    - Updating metadata in the Glue Data Catalog.
- **Key Permissions**:
  - Access to Glue Data Catalog resources:
    - `glue:GetTable`
    - `glue:GetTableVersion`
    - `glue:BatchGetPartition`
  - Job and crawler operations (if required):
    - `glue:StartCrawler`
    - `glue:UpdateCrawler`
    - `glue:StartJobRun`

Role Usage

1. **Lambda Functions**:
   - Lambda functions executing serverless data pipelines use this role to:
     - Fetch raw data from S3 for processing.
     - Write transformed or processed data back to S3.
     - Interact with the Glue Data Catalog for metadata or schema validation.
2. **Data Processing Workflows**:
   - The role supports Lambda functions in workflows that require:
     - Validating data against Glue schema definitions.
     - Transforming raw data into a cleaned format.
     - Triggering downstream Glue operations (e.g., starting crawlers or ETL jobs).

Benefits of This Role

1. **Comprehensive S3 Access**:
   - Ensures Lambda can seamlessly read and write to any required S3 bucket.
2. **Glue Integration**:
   - Enables metadata-driven workflows by allowing Lambda to query and utilize the Glue Data Catalog.
3. **Serverless Workflow Support**:
   - Facilitates serverless data processing pipelines by combining S3 data access and Glue catalog interactions in one role.

Summary

The `de-on-youtube-raw-us-east-1-s3-lambda-role` is a vital IAM role for enabling serverless data workflows. It provides:

- Full S3 access for reading and writing data.
- Integration with the Glue Data Catalog for metadata management. This combination ensures efficient data processing and metadata-driven workflows in AWS Lambda.