# REQUIREMENT FROM (OUR SIMULATED) CUSTOMER

* Launching new data-driven campaign
* Main advertising channel: **YouTube**
* Initial questions to answer:
* "How to categorize videos, based on their comments and statistics"
* "What factors affect how popular a YouTube video will be"

# Goals and Success Criteria

## Data Ingestion

* Ingest data, one-offs and incrementally

## Data Lake

* Design and build a new Data Lake architecture

## AWS Cloud

* AWS as the cloud provider

## ETL Design

* Extract, transform and load data efficiently

## Scalability

* The data architecture should be scaled efficiently

## Reporting

* Build a Business Intelligence tier, incl. Dashboards

Steps:

1. Build IAM user, name “AarshMehtani\_JI”
2. Log in through IAM user.
3. Configure the aws through cmd.

A computer screen with white text

Description automatically generated

1. Get the data, and create s3 bucket, for our landing bucket. Copy the data to s3, using aws cli

**aws s3 cp . s3://de-youtube-raw-ap-south-1/youtube/raw\_statistics\_reference\_data/ --recursive --exclude "\*" --include "\*.json”**

****

1. Now I’m going to build AWS glue catalog.

* As IAM user, we don’t have direct permissions for one service accessing to another service.
* For that we must build a role.
  + Give permission:
    - S3 full version
    - Glue full version
* Use that role while building the crawler.
* Run the crawler.

1. View the table, it will open ATHENA automatically.
2. Set the **Query result** in new bucket

**‘**[**de-youtube-raw-ap-south-1-athena-output**](https://ap-south-1.console.aws.amazon.com/s3/buckets/de-youtube-raw-ap-south-1-athena-output?region=ap-south-1)**’**

1. To solve the error:

**A screenshot of a computer

Description automatically generated**

* Go to the **lambda** service.
* Create new function.
* Build new role

**A screenshot of a computer

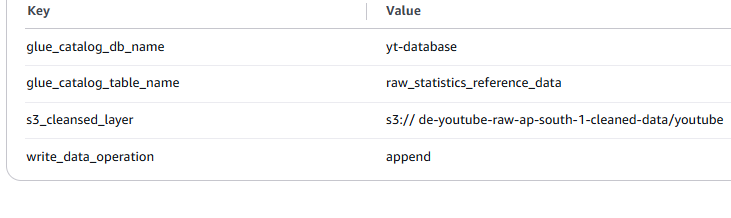
Description automatically generated**

* Give permissions:
  + **S3 full access**

**A screenshot of a computer

Description automatically generated**

1. Set the environment variables.

****

1. Replace the bucket name and url under Test

**A screenshot of a computer code

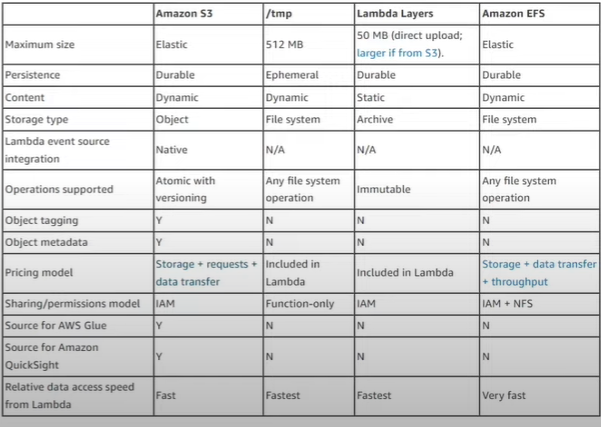
Description automatically generated**

* Run the test
* Showing the error:

**A screenshot of a computer

Description automatically generated**

\* To understand the error, first understand the lambda service.



To solve the error add lambda layer.