Workflow from Airflow to AWS

A screenshot of a computer

Description automatically generated

After running the DAG: reddit extraction is done and uploaded to S3 bucket([dataeng-reddit](https://us-east-1.console.aws.amazon.com/s3/buckets/dataeng-reddit?region=us-east-1&bucketType=general))

##AWS(S3, glue, Athena or Redshift)

S3: [dataeng-reddit](https://us-east-1.console.aws.amazon.com/s3/buckets/dataeng-reddit?region=us-east-1&bucketType=general): first step is to create s3 bucket (before DAG)

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generatedcsv is generated from extraction and uploaded in s3 bucket inside raw from

Then create two other empty folders for next steps

One for transformed data from AWS glue

One for Athena storage

To get transformed data we use AWS glue

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

First we do visual, then set job details and write transformation script according to business logic, then save and run.

After run,  
  
A screenshot of a computer

Description automatically generated

Then you will transformed data in the s3 folder you created  
  
A screenshot of a computer

Description automatically generated

Then you need to create a database to store this transformed data

A screenshot of a computer

Description automatically generated

Open database and add table path

A screenshot of a computer

Description automatically generated

Then we create a crawler

A screenshot of a computer

Description automatically generated

In created crawler, we give IAM policies, and assign created database and run the crawler

A screenshot of a computer

Description automatically generated

Then we go to Athena for query

A screenshot of a computer

Description automatically generated

After this you can also check and see in redshift(visualize data in powerbi, tableu)