➢ **Quality movie Data analysis - POC**
   Tech stack - AWS Glue, redshift, s3, eventbridge, SNS, SQL

Project execution steps:

- Daily data would come into S3 bucket.
- Created crawlers to add source and target table data catalog which would scan and update the metadata.
- Created Data quality check for source S3 table in Glue visual ETL.
- Created ETL pipeline with visual ETL . Extract the raw data from S3, perform the data quality checks like IMDB rating > 8.5 , dump the failed records to separate folder in s3. For passed data quality records, load that refined data with specific columns into redshift table which would contain only the quality movies above 8.5.
- Created Rules in eventbridge which would track data quality evaluation results from cloudwatch logs generated from GLUE ETL pipeline and send email notifications by SNS.

Glue job :