# Data Engineering Project

- End-to-end automated data Pipeline
- Input data is stored in AWS S3 bucket.
- Data may contain invalid rows.
- Need to separate valid & invalid rows.
- Auto validation of data as soon as data is arrived in S3 bucket.
- Auto scanning of data to create tables.
- Process/analyze data with AWS Athena.

S3 (input data)   Lambda Function   IAM Role

S3 (output data)   Glue Crawler   Athena
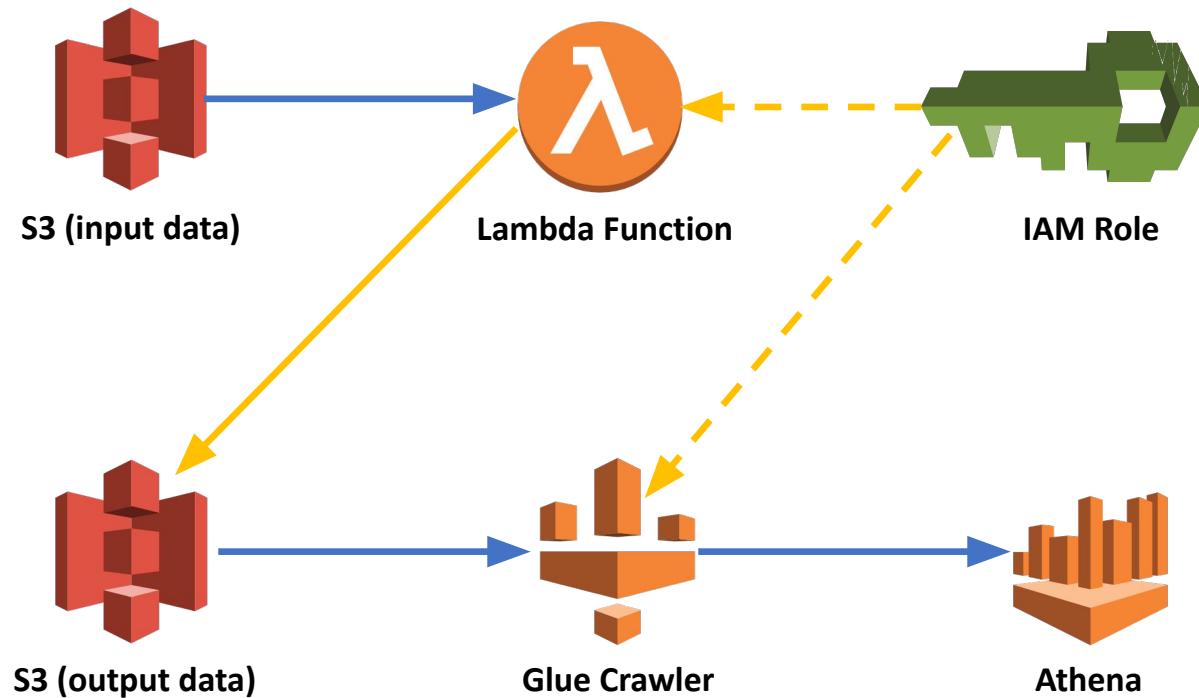
AWS Services used in Project

S3
- Stores input data
- Stores output data

IAM
- IAM permission for Lambda Function
- IAM permission Glue for crawler

Lambda Function
- Reads input data from S3
- Checks for invalid data in the input data
- Writes only valid data into S3.

Glue crawler
- Scans the input data & creates table

Athena
- Analyze data from S3 via Glue data catalog.

IAM Roles/Permissions

IAM Role for Lambda Function:

- Basic Lambda execution permission
- Read/Write permission to S3

IAM Role for Glue Crawler:

- Basic Crawler permission
- Read permission to S3