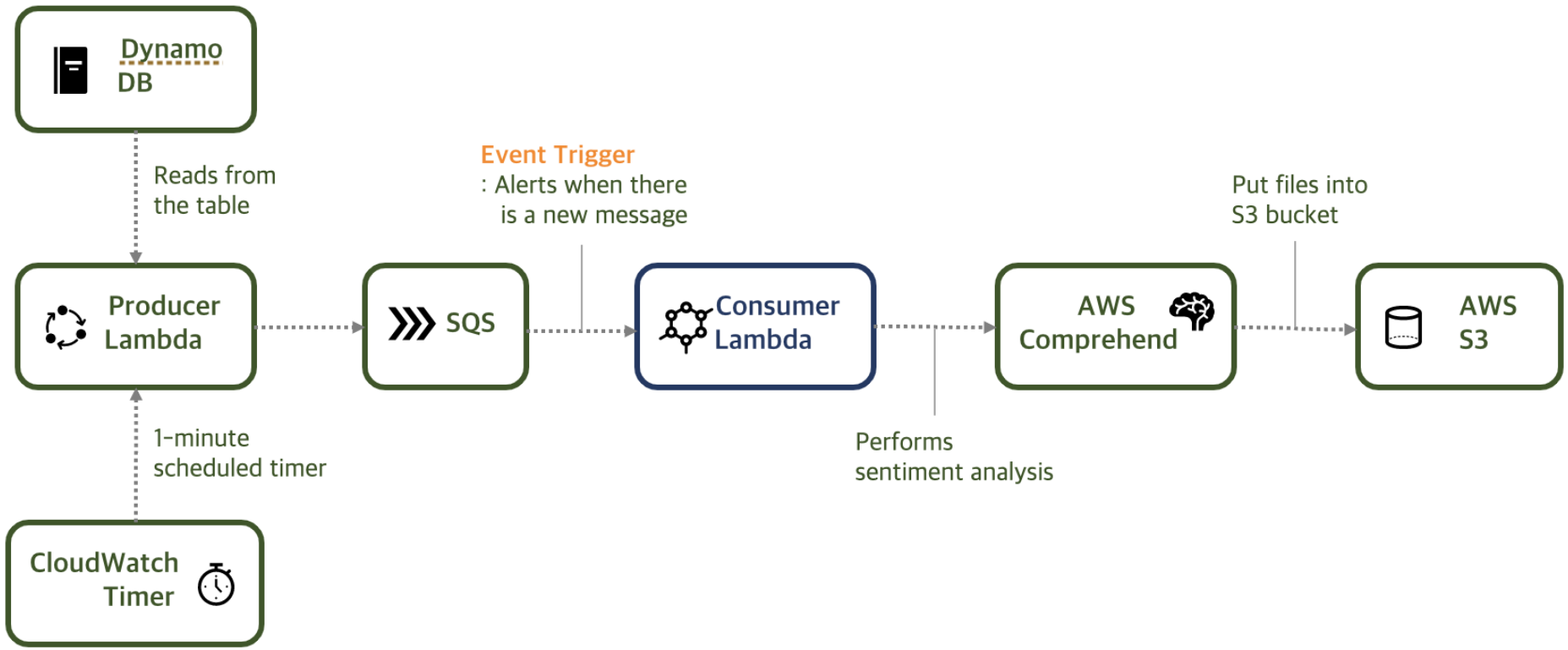
Serverless Data Engineering Pipeline with AWS

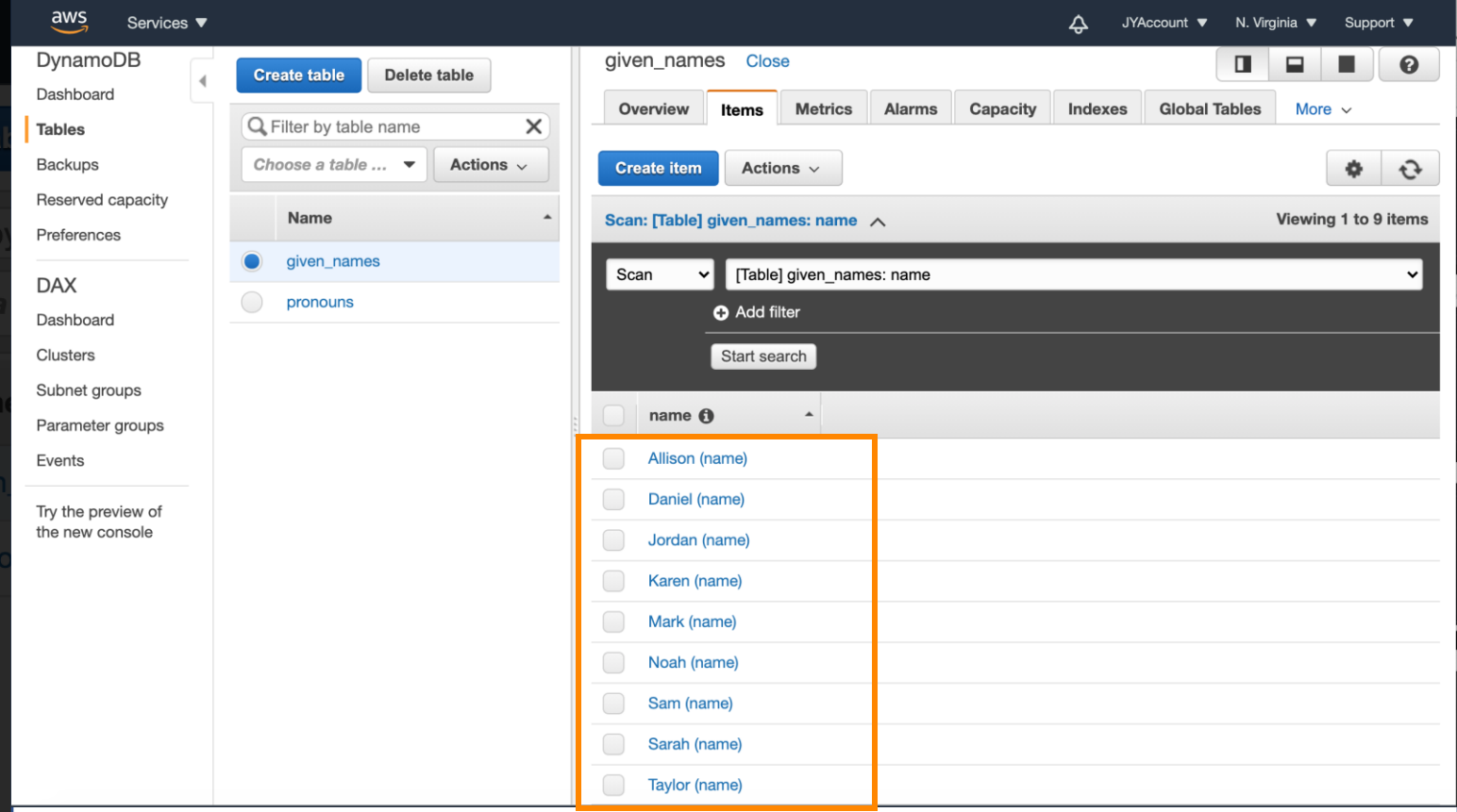
IDS 706. Final Project

Jasmine Young

Rhayoung Park

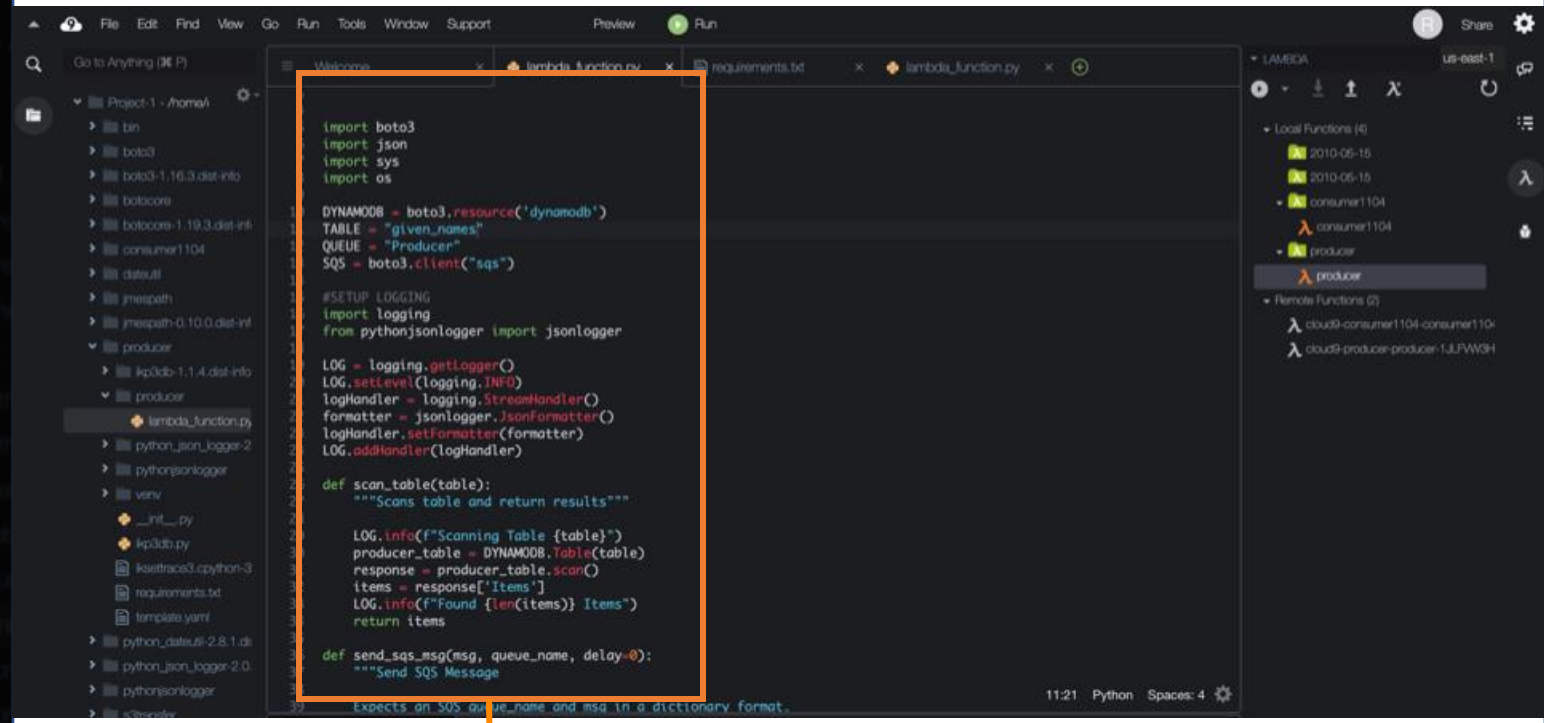

Dynamo DB
Reads from the table
Producer Lambda
1-minute scheduled timer
CloudWatch Timer
SQS
Event Trigger
: Alerts when there is a new message
Consumer Lambda
Performs sentiment analysis
AWS Comprehend
Put files into S3 bucket
AWS S3

DynamoDB Table



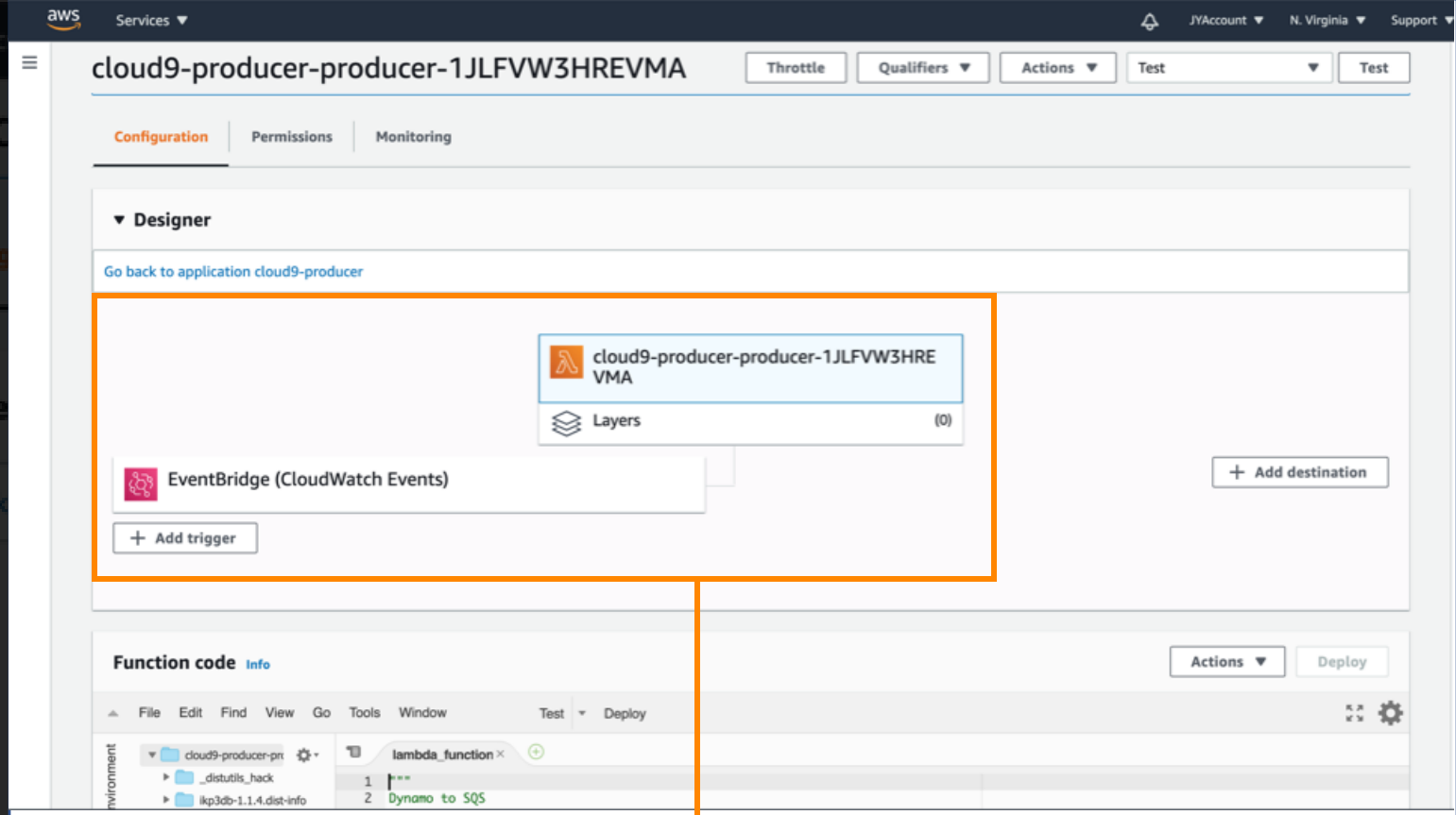
Producer Lambda scans the items in Dynamo DB. The sentiment of these names will be analyzed with AWS Comprehend.

Cloud9 Producer Snapshot



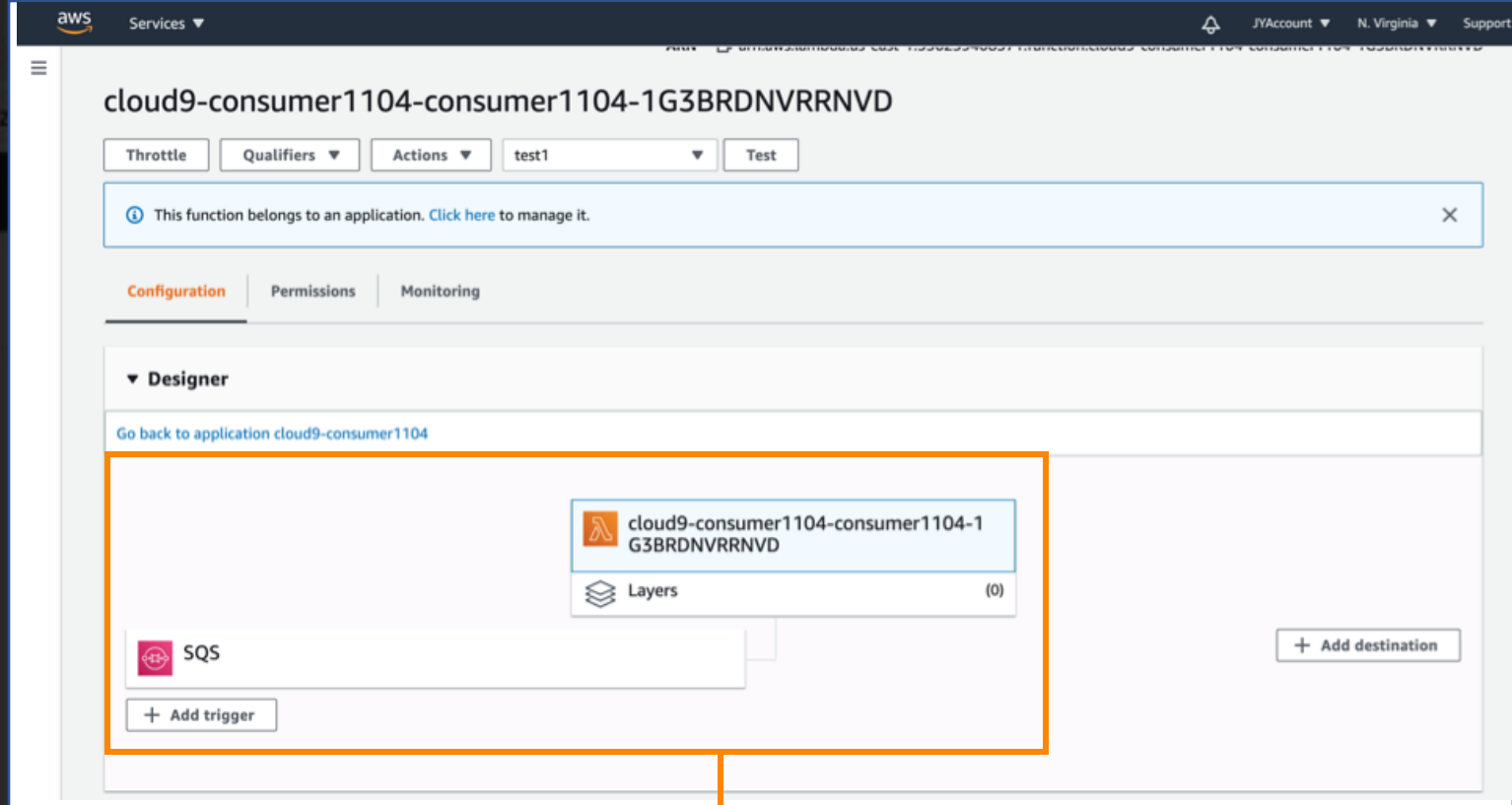
Our lambda functions were written in Cloud9 before deployment.

Producer Lambda Function



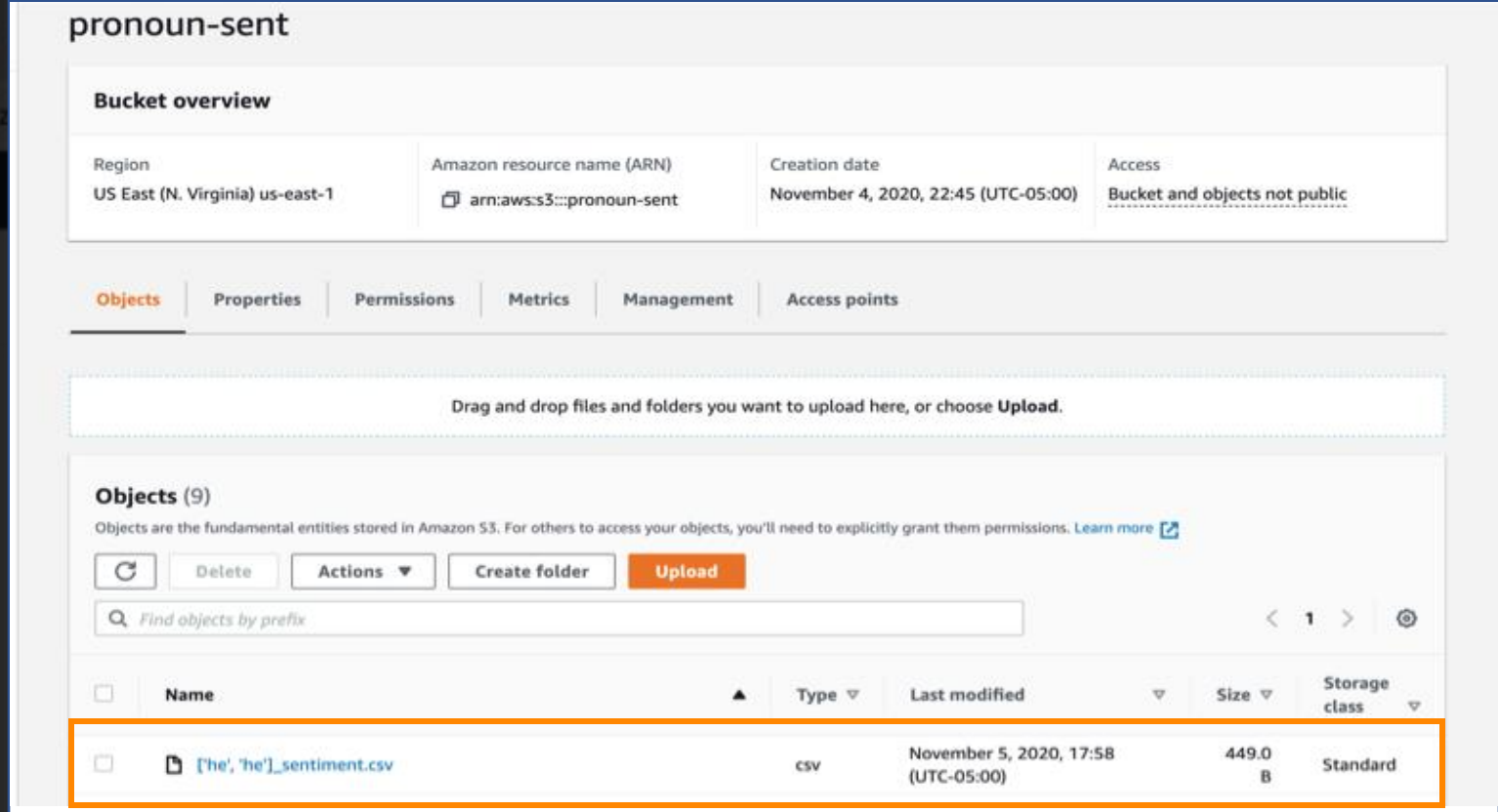
Producer Lambda is triggered by CloudWatch every 1 minute.

Consumer Lambda Function



Consumer Lambda is triggered by queuing system (SQS)

Amazon S3 Bucket



CSV files containing sentiment analysis results are deposited in S3