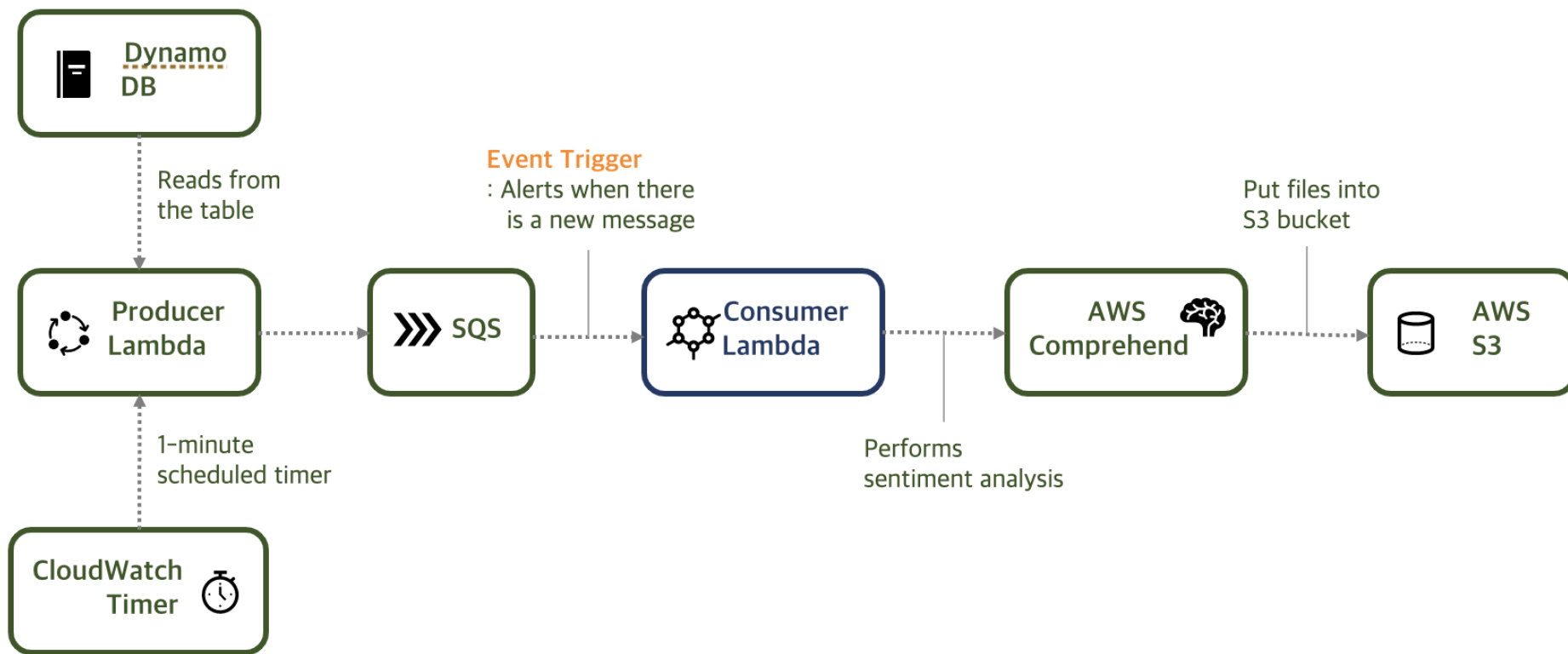




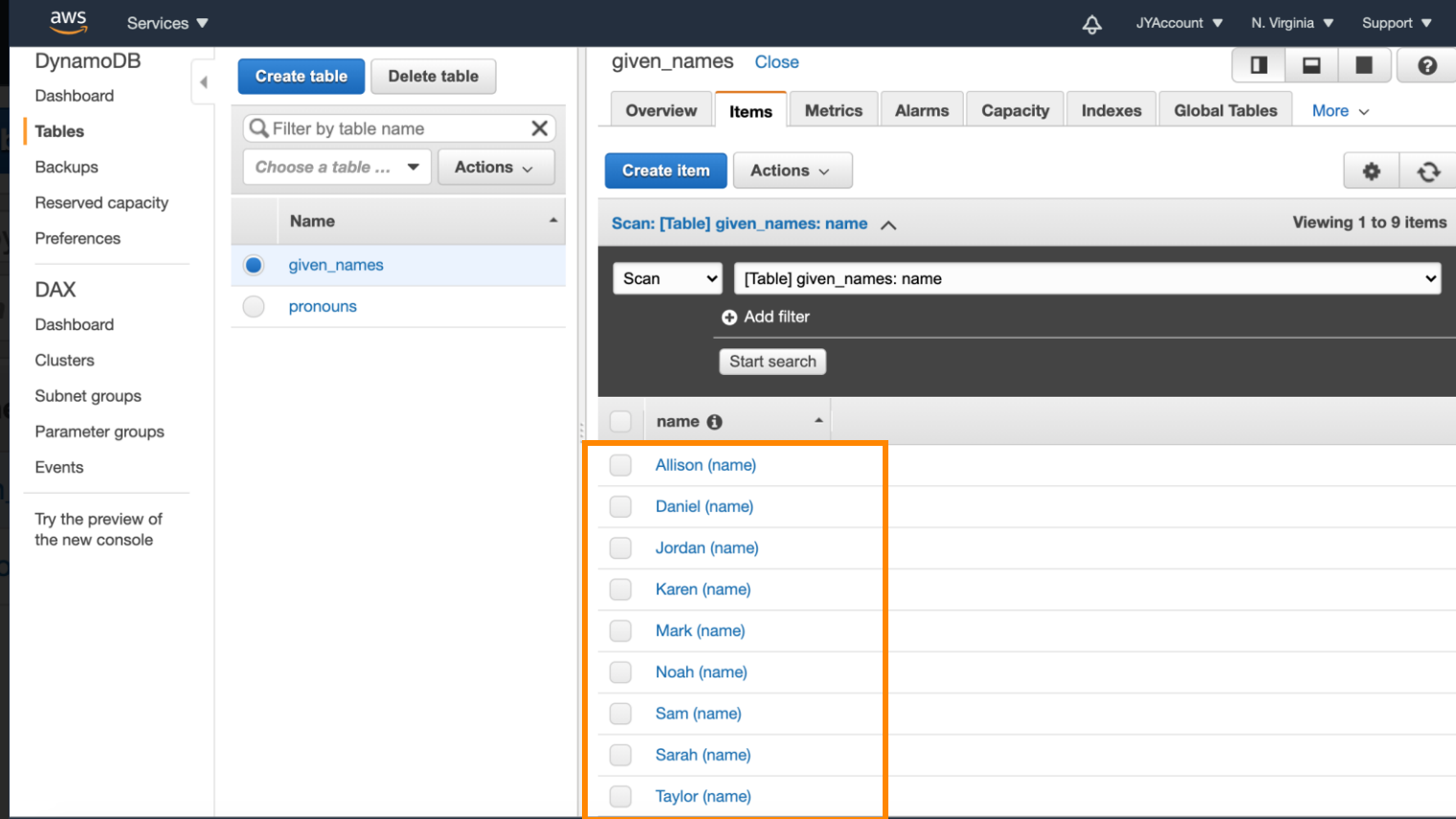
Serverless Data Engineering Pipeline with AWS

IDS 706. Final Project

Jasmine Young
Rhayoung Park



DynamoDB Table



The screenshot displays the AWS Management Console interface for a DynamoDB table named 'given_names'. The left sidebar shows the navigation menu with 'DynamoDB' selected. The main content area is divided into two panes. The left pane shows the 'Tables' section with a filter by table name and a list of tables: 'given_names' (selected) and 'pronouns'. The right pane shows the 'Items' tab for the 'given_names' table. It includes a 'Scan' button and a search bar. Below the search bar, a list of items is displayed, each with a checkbox and a name: Allison (name), Daniel (name), Jordan (name), Karen (name), Mark (name), Noah (name), Sam (name), Sarah (name), and Taylor (name). The list is highlighted with an orange border.

given_names Close

Overview Items Metrics Alarms Capacity Indexes Global Tables More

Create Item Actions

Scan: [Table] given_names: name Viewing 1 to 9 items

Scan [Table] given_names: name

+ Add filter

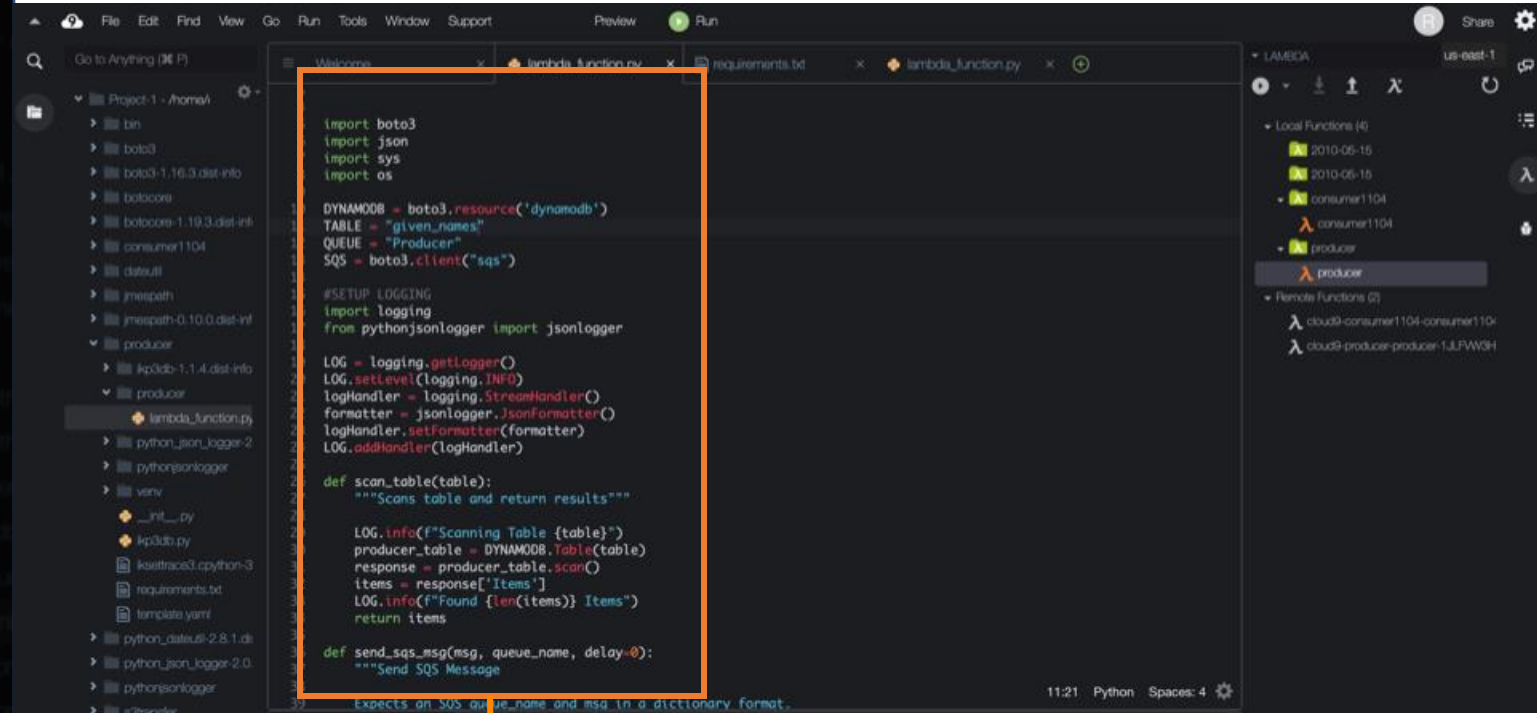
Start search

name

- ☐ Allison (name)
- ☐ Daniel (name)
- ☐ Jordan (name)
- ☐ Karen (name)
- ☐ Mark (name)
- ☐ Noah (name)
- ☐ Sam (name)
- ☐ Sarah (name)
- ☐ Taylor (name)

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

Cloud9 Producer Snapshot



```
import boto3
import json
import sys
import os

DYNAMODB = boto3.resource('dynamodb')
TABLE = "given_names"
QUEUE = "Producer"
SQS = boto3.client("sqs")

#SETUP LOGGING
import logging
from pythonjsonlogger import jsonlogger

LOG = logging.getLogger()
LOG.setLevel(logging.INFO)
logHandler = logging.StreamHandler()
formatter = jsonlogger.JsonFormatter()
logHandler.setFormatter(formatter)
LOG.addHandler(logHandler)

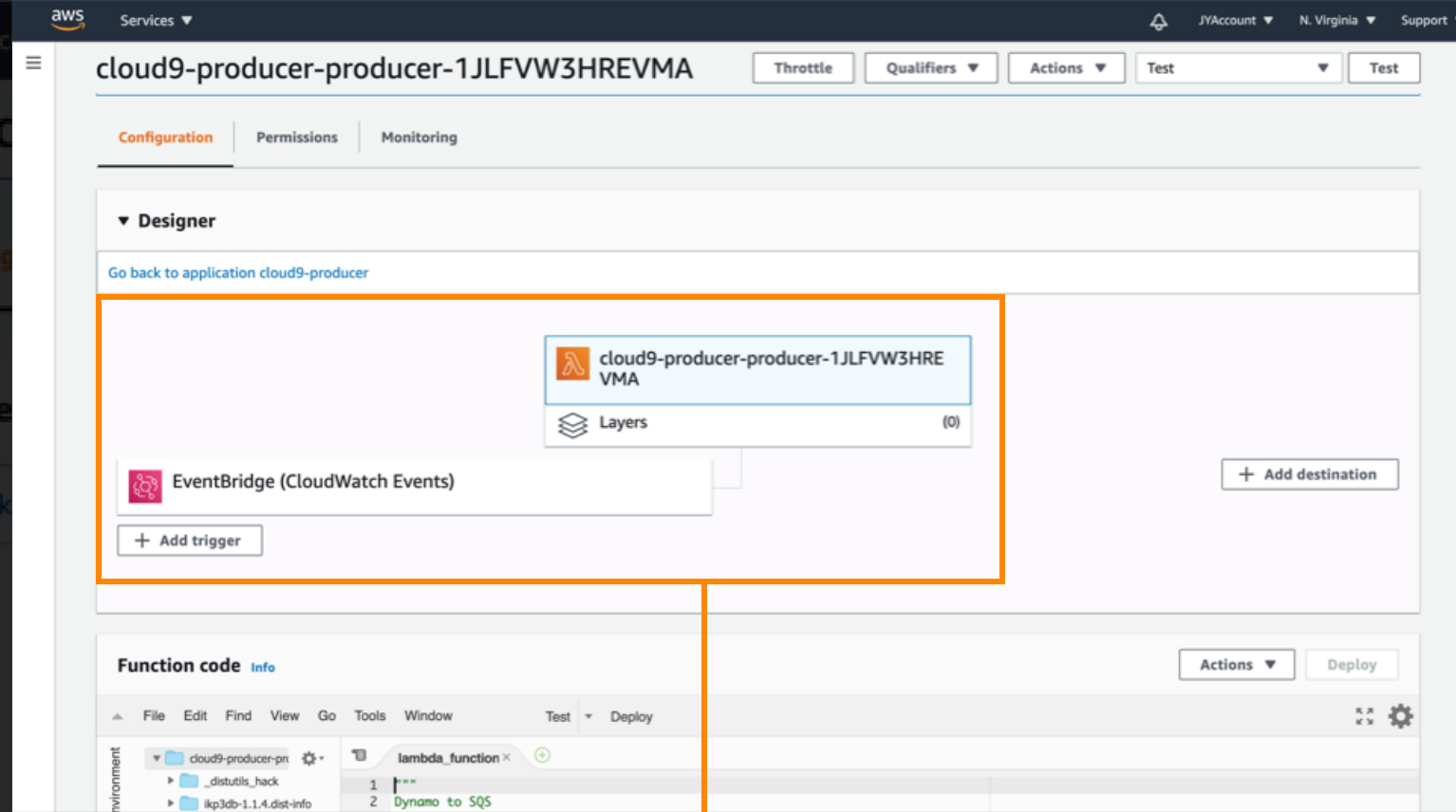
def scan_table(table):
    """Scans table and return results"""

    LOG.info(f"Scanning Table {table}")
    producer_table = DYNAMODB.Table(table)
    response = producer_table.scan()
    items = response['Items']
    LOG.info(f"Found {len(items)} Items")
    return items

def send_sqs_msg(msg, queue_name, delay=0):
    """Send SQS Message
    Expects an SQS msg, queue_name and msg in a dictionary format.
```

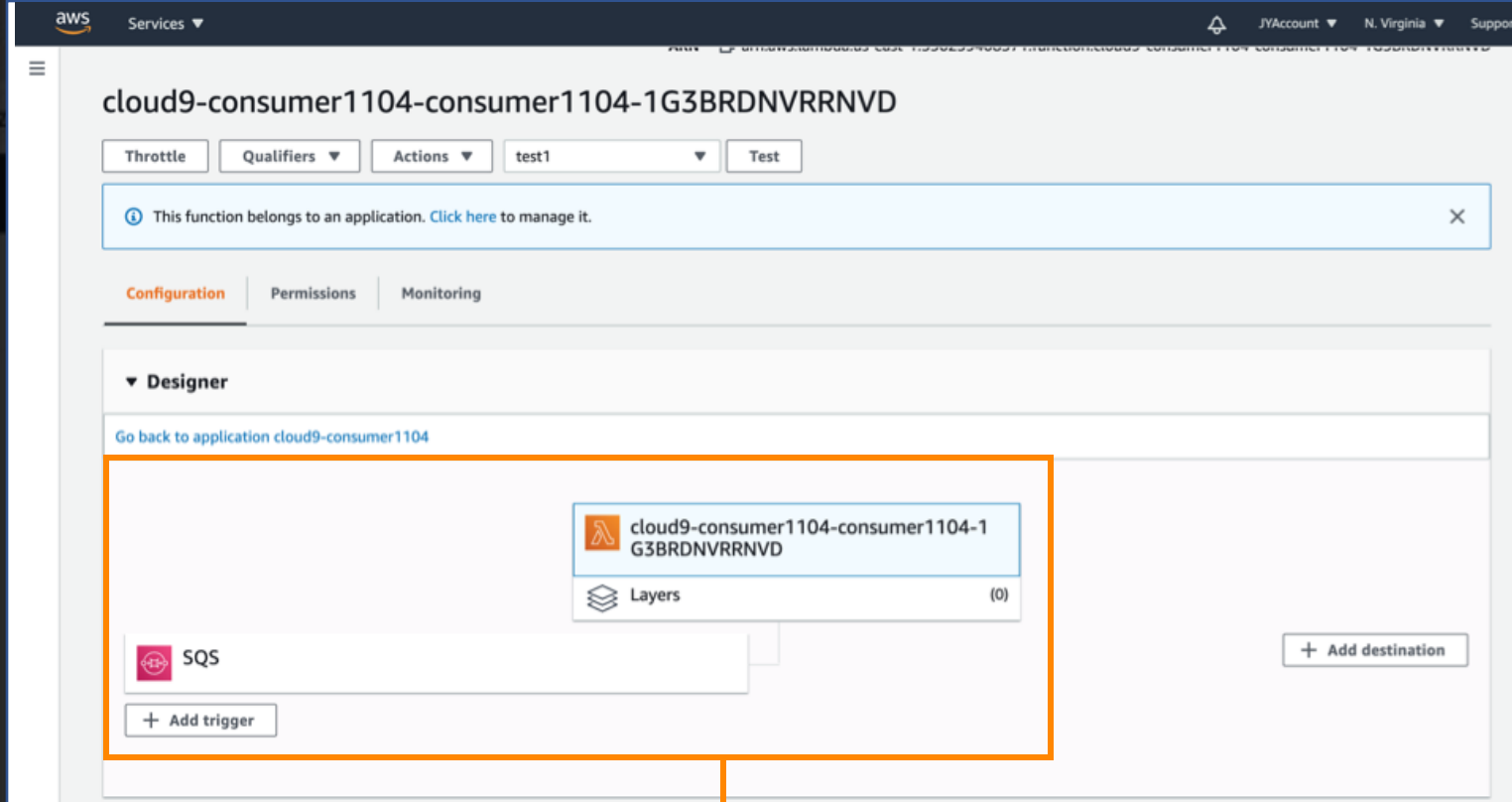
Our lambda functions were written in Cloud9 before deployment.

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

pronoun-sent




Bucket overview


Region US East (N. Virginia) us-east-1	Amazon resource name (ARN)  arn:aws:s3:::pronoun-sent	Creation date November 4, 2020, 22:45 (UTC-05:00)	Access Bucket and objects not public
---	---	--	---


Objects | Properties | Permissions | Metrics | Management | Access points

Drag and drop files and folders you want to upload here, or choose **Upload**.

Objects (9)
Objects are the fundamental entities stored in Amazon S3. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

  **Actions** ▾  **Upload**

 Find objects by prefix

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	 ['he', 'he']_sentiment.csv	csv	November 5, 2020, 17:58 (UTC-05:00)	449.0 B	Standard

CSV files containing sentiment analysis results are deposited in S3