Cloud Computing and Big Data - Spring 2020

Homework Assignment 3

Assignment:

Implement a photo album web application, that can be searched using natural language through both text and voice. You will learn how to use Lex, ElasticSearch, and Rekognition to create an intelligent search layer to query your photos for people, objects, actions, landmarks and more.

Outline:

This assignment has five components:

1. **Launch an ElasticSearch instance**[1]:
   a. Using AWS ElasticSearch service[2], create a new domain called **"photos"**.
   b. Make note of the Security Group (**SG1**) you attach to the domain.
   c. Deploy the service inside a VPC[3].
      i. This prevents unauthorized internet access to your service.
2. **Upload & index photos**
   a. Create a S3 bucket (**B2**) to store the photos.
   b. Create a Lambda function (**LF1**) called **"index-photos"**.
      i. Launch the Lambda function inside the same VPC as ElasticSearch. This ensures that the function can reach the ElasticSearch instance.
      ii. Make sure the Lambda has the same Security Group (**SG1**) as ElasticSearch.
   c. Set up a PUT event trigger[4] on the photos S3 bucket (**B2**), such that whenever a photo gets uploaded to the bucket, it triggers the Lambda function (**LF1**) to index it.
      i. To test this functionality, upload a file to the photos S3 bucket (**B2**) and check the logs of the indexing Lambda function (**LF1**) to see if it got invoked. If it did, your setup is complete.

[1] https://www.elastic.co/webinars/getting-started-elasticsearch?elektra=home&storm=sub1
[2] https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/es-createupdatedomains.html
[3] https://docs.aws.amazon.com/elasticsearch-service/latest/developerguide/es-vpc.html
[4] https://docs.aws.amazon.com/AmazonS3/latest/dev/NotificationHowTo.html

- If the Lambda (**LF1**) did not get invoked, check to see if you set up the correct permissions[5] for S3 to invoke your Lambda function.

d. Implement the indexing Lambda function (**LF1**):
   i. Given a S3 PUT event (**E1**) detect labels in the image, using Rekognition[6] ("detectLabels" method).
   ii. Store a JSON object in an ElasticSearch index ("photos") that references the S3 object from the PUT event (**E1**) and an array of string labels, one for each label detected by Rekognition.

Use the following schema for the JSON object:

```
{
      "objectKey": "my-photo.jpg",
      "bucket": "my-photo-bucket",
      "createdTimestamp": "2018-11-05T12:40:02",
      "labels": [
            "person",
            "dog",
            "ball",
            "park"
      ]
}
```


3. Search

a. Create a Lambda function (**LF2**) called **"search-photos"**.
   i. Launch the Lambda function inside the same VPC as ElasticSearch. This ensures that the function can reach the ElasticSearch instance.
   ii. Make sure the Lambda has the same Security Group (**SG1**) as ElasticSearch.

b. Create an Amazon Lex bot to handle search queries.
   i. Create one intent named "SearchIntent".
   ii. Add training utterances to the intent, such that the bot can pick up both keyword searches ("trees", "birds"), as well as sentence searches ("show me trees", "show me photos with trees and birds in them").

---

[5] https://docs.aws.amazon.com/lambda/latest/dg/with-s3-example.html (see Configure Amazon S3 to Publish Events)
[6] https://aws.amazon.com/rekognition/

- You should be able to handle at least one or two keywords per query.

c. Implement the Search Lambda function (**LF2**):
   i. Given a search query “q”, disambiguate the query using the Amazon Lex bot.
   ii. If the Lex disambiguation request yields any keywords (**K_1, …, K_n**), search the “photos” ElasticSearch index for results, and return them accordingly (as per the API spec).
      - You should look for ElasticSearch SDK libraries to perform the search.
   iii. Otherwise, return an empty array of results (as per the API spec).

4. **Build the API layer**
   a. Build an API using API Gateway.
      i. The Swagger API documentation for the API can be found here: https://github.com/001000001/ai-photo-search-columbia-f2018/blob/master/swagger.yaml
   b. The API should have two methods:
      i. PUT /photos

         Set up the method as an Amazon S3 Proxy.[7] This will allow API Gateway to forward your PUT request directly to S3.

      ii. GET /search?q={query text}

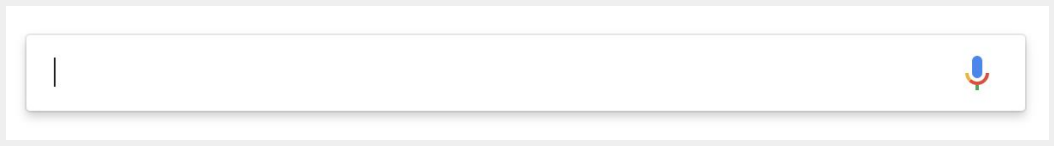
         Connect this method to the search Lambda function (**LF2**).
   c. Setup an API key for your two API methods.
   d. Deploy the API.
   e. Generate a SDK for the API (**SDK1**).
5. **Frontend**
   a. Build a simple frontend application that allows users to:
      i. Make search requests to the GET /search endpoint
      ii. Display the results (photos) resulting from the query
      iii. Upload new photos using the PUT /photos
   b. Create a S3 bucket for your frontend (**B1**).
   c. Set up the bucket for static website hosting (same as HW1).
   d. Upload the frontend files to the bucket (**B2**).

---

[7] https://docs.aws.amazon.com/apigateway/latest/developerguide/integrating-api-with-aws-services-s3.html

e. Integrate the API Gateway-generated SDK (**SDK1**) into the frontend, to connect your API.

6. **Implement Voice accessibility in the frontend**
   a. Give the frontend user the choice to use voice rather than text to perform the search.
   b. Use Amazon Transcribe[8] on the frontend to transcribe speech to text (STT) in real time[9], then use the transcribed text to perform the search, using the same API like in the previous steps.
   c. Note: You can use a Google-like UI (see below) for implementing the search: 1. input field for text searches and 2. microphone icon for voice interactions.



7. **Deploy your code using AWS CodePipeline[10]**
   a. Define a pipeline (P1) in AWS CodePipeline that builds and deploys the code for/to all your Lambda functions
   b. Define a pipeline (P2) in AWS CodePipeline that builds and deploys your frontend code to its corresponding S3 bucket
8. **Create a AWS CloudFormation[11] template for the stack**
   a. Create a CloudFormation template (T1) to represent all the infrastructure resources (ex. Lambdas, ElasticSearch, API Gateway, CodePipeline, etc.) and permissions (IAM policies, roles, etc.).

At this point you should be able to:

1. Visit your photo album application using the S3 hosted URL.
2. Search photos using natural language via voice and text.
3. See relevant results (ex. If you searched for a cat, you should be able to see photos with cats in them) based on what you searched.
4. Upload new photos and see them appear in the search results.

[8] https://aws.amazon.com/transcribe/
[9] https://docs.aws.amazon.com/transcribe/latest/dg/streaming.html
[10] https://aws.amazon.com/codepipeline/
[11] https://aws.amazon.com/cloudformation/

Acceptance criteria:

1. Using the CloudFormation template (T1) you should be able to stand up the entire functional stack for this assignment.
2. Once a new commit is pushed to GitHub (both for frontend and backend repos), CodePipeline should build and deploy your code to the corresponding AWS infrastructure.
3. For a given photo and a given search query, a correct search (as defined in the assignment) should be able to return every photo that matches the query. Specifically, if Rekognition returns 12 labels for a given photo, your search should return the photo for any one of those 12 labels, if searched independently ("show me **dogs**") or in groups ("show me **cats** and **dogs**").
4. All other functionality should be working as described above.

ANNEX

Architecture Diagram