

## **AWS NLP Services (Natural Language Processing) for beginners workshop**

Amazon Web Services Japan

Sr. Evangelist, Harunobu Kameda (hkameda@amazon.co.jp)

About this workshop;

You can experience NLP (Natural Language Processing) related services provided by AWS via this workshop. For APAC wide hosted even, this scenario is done by English, but you can customize to your native language if you want later.

- Amazon Translate : translation service by machine learning
- Amazon Polly: Text-to-Speech service
- Amazon Transcribe : Speech-to-Text service
- Amazon Comprehend : obtaining insights from given text.

**Oregon** region is used for this scenario.

### 1. Setting up IAM Role

IAM roles is used to grant access to those NLP services by Lambda function.

#### 1-1. Access top page of IAM, and click [Roles]

## Dashboard

### ▼ Access management

Groups

Users

Roles

Policies

Identity providers

Account settings

1-2. Click [Create role]



1-3. Select [Lambda], and click[Next: Permissions]

### Create role

1234

Select type of trusted entity

**AWS service**  
EC2, Lambda and others

**Another AWS account**  
Belonging to you or 3rd party

**Web identity**  
Cognito or any OpenID provider

**SAML 2.0 federation**  
Your corporate directory

Allows AWS services to perform actions on your behalf. [Learn more](#)

Choose a use case

**Common use cases**

**EC2**  
Allows EC2 instances to call AWS services on your behalf.

**Lambda**  
Allows Lambda functions to call AWS services on your behalf.

**Or select a service to view its use cases**

|                             |                                        |                                           |                                 |                           |
|-----------------------------|----------------------------------------|-------------------------------------------|---------------------------------|---------------------------|
| <a href="#">API Gateway</a> | <a href="#">CodeDeploy</a>             | <a href="#">EMR</a>                       | <a href="#">KMS</a>             | <a href="#">RoboMaker</a> |
| <a href="#">AWS Backup</a>  | <a href="#">CodeGuru</a>               | <a href="#">ElastiCache</a>               | <a href="#">Kinesis</a>         | <a href="#">S3</a>        |
| <a href="#">AWS Chatbot</a> | <a href="#">CodeStar Notifications</a> | <a href="#">Elastic Beanstalk</a>         | <a href="#">Lambda</a>          | <a href="#">SMS</a>       |
| <a href="#">AWS Support</a> | <a href="#">Comprehend</a>             | <a href="#">Elastic Container Service</a> | <a href="#">Lex</a>             | <a href="#">SNS</a>       |
| <a href="#">Amplify</a>     | <a href="#">Config</a>                 | <a href="#">Elastic Transcoder</a>        | <a href="#">License Manager</a> | <a href="#">SWF</a>       |

\* Required

Cancel **Next: Permissions**

1-4. Search [translate], and check [TranslateFullAccess] **Do not click button at**

lower right corner, yet.

Create role

1234

▼ Attach permissions policies

Choose one or more policies to attach to your new role.



Create policy

Refresh

Filter policies ▼

Q translate

Showing 4 results

|                                     | Policy name ▼                                                                                           | Used as                |
|-------------------------------------|---------------------------------------------------------------------------------------------------------|------------------------|
| <input type="checkbox"/>            | ▶ PollyAndTranslateForWordPressPolicy-0308                                                              | None                   |
| <input type="checkbox"/>            | ▶ PollyAndTranslateForWordPressPolicy-0819                                                              | Permissions policy (1) |
| <input checked="" type="checkbox"/> | ▶  TranslateFullAccess | Permissions policy (3) |
| <input type="checkbox"/>            | ▶  TranslateReadOnly   | None                   |

1-4. As like, translate, check [polly], [transcribe], and [comprehend] to setting up full access. After that, click [Next: Tags]






Create policy

Refresh

Filter policies ▼

Q comprehend

Showing 5 results

|                                     | Policy name ▼                                                                                                                           | Used as                |
|-------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|------------------------|
| <input type="checkbox"/>            | ▶  AWSApplicationAutoscalingComprehendEndpointPolicy | None                   |
| <input type="checkbox"/>            | ▶  ComprehendDataAccessRolePolicy                    | None                   |
| <input checked="" type="checkbox"/> | ▶  ComprehendFullAccess                              | Permissions policy (3) |
| <input type="checkbox"/>            | ▶  ComprehendMedicalFullAccess                       | None                   |
| <input type="checkbox"/>            | ▶  ComprehendReadOnly                                | None                   |

▶ Set permissions boundary

\* Required

Cancel

Previous

Next: Tags

1-5. Without input anything, just click [Next: Review]

### Add tags (optional)

IAM tags are key-value pairs you can add to your role. Tags can include user information, such as an email address, or can be descriptive, such as a job title. You can use the tags to organize, track, or control access for this role. [Learn more](#)

| Key                                      | Value (optional)     | Remove |
|------------------------------------------|----------------------|--------|
| <input type="text" value="Add new key"/> | <input type="text"/> |        |

You can add 50 more tags.

1-6. Set up Role name as [YYYYMMDDnlpandson] (YYYYMMDD is date of today.

Ex. 20200412)

### Review

Provide the required information below and review this role before you create it.

**Role name\***









Use alphanumeric and '+=, @, \_' characters. Maximum 64 characters.

**Role description**

Maximum 1000 characters. Use alphanumeric and '+=, @, \_' characters.

**Trusted entities** AWS service: lambda.amazonaws.com

**Policies**

-  [TranslateFullAccess](#) 
-  [AmazonPollyFullAccess](#) 
-  [AmazonTranscribeFullAccess](#) 
-  [ComprehendFullAccess](#) 

**Permissions boundary** Permissions boundary is not set

*At a glance*

1-7. Double check, the role has full access to 4 services, and click [Create role].

If you see the below dialog, role is successfully created.

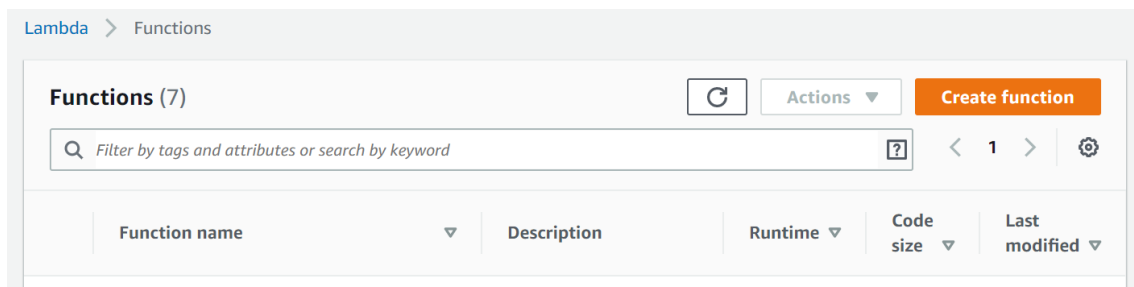


The role **20200412** has been created.

## 2. Amazon Translate

2-1. Let's find an appropriate English News from web. Copy and paste it to your text editor.

2-2. Access AWS Lambda Top page, and click [Create function]



2-3. Input [YYYYMMDDtranslate](YYYYMMDD is date of today) to Function name, and select [Python3.8] as your runtime.

The screenshot shows the 'Basic information' section of the AWS Lambda 'Create function' wizard. It has a title 'Basic information' in bold. Below it, the 'Function name' field is labeled 'Function name' with a description 'Enter a name that describes the purpose of your function.' The input field contains 'YYYYMMDDtranslate'. A note below the field states 'Use only letters, numbers, hyphens, or underscores with no spaces.' The 'Runtime' section is labeled 'Runtime' with an 'Info' link and a description 'Choose the language to use to write your function.' The dropdown menu for runtime is open, showing 'Python 3.8' as the selected option.


2-4. Click [Choose or create an execution role] and check [Use an existing role]. then, you need to select IAM role which you have created above. After you selected role, click [Create function]

▼ **Choose or create an execution role**

**Execution role**  
Choose a role that defines the permissions of your function. To create a custom role, go to the [IAM console](#).

☐ Create a new role with basic Lambda permissions  
☒ Use an existing role  
☐ Create a new role from AWS policy templates

**Existing role**  
Choose an existing role that you've created to be used with this Lambda function. The role must have permission to upload logs to Amazon CloudWatch Logs.

20200412 ▼ 

[View the 20200412 role](#) on the IAM console.

Cancel **Create function**

2-5. Delete sample code at first, and copy and paste [translatelambda.txt] into the editor. Then click [Save] button.

Throttle Qualifiers ▼ Actions ▼ Select a test event ▼ Test Save

Runtime Python 3.8 Handler Info lambda\_function.lambda\_handler

Go Tools Window Save Test ▼

```

1 import boto3
2 import os
3
4 def lambda_handler(event, context):
5
6     SRC_LANG = 'en'
7     TRG_LANG = 'ja'
8
9     translate = boto3.client('translate')
10
11     response = translate.translate_text(Text=event["text"], SourceLanguageCode=SRC_LANG, TargetLanguageCode=TRG_LANG)
12
13     return (response['TranslatedText'])
14
15
  
```

2-6. Click [Test] button.

Throttle Qualifiers ▼ Actions ▼ Select a test event ▼ Test Save

2-7. Type [test] or appropriate name to Event name, and copy and paste with [translatelambdatest.txt].

Configure test event

A function can have up to 10 test events. The events are persisted so you can switch to another computer or web browser and test your function with the same events.

Create new test event

Edit saved test events

Event template

Hello World

Event name

test

1 {

2 "text": "hello"

3 }

2-8. Replace [hello] to your previously copied English news. At the time, you have to wave every line feed code. In other word, the line should be one line.

Then, click [Create]

イベント名

testtranslate

```
1 {  
2   "text": "A mystery pneumonia outbreak in mainland China has sparked fears across eastern Asia"  
3 }
```

キャンセル

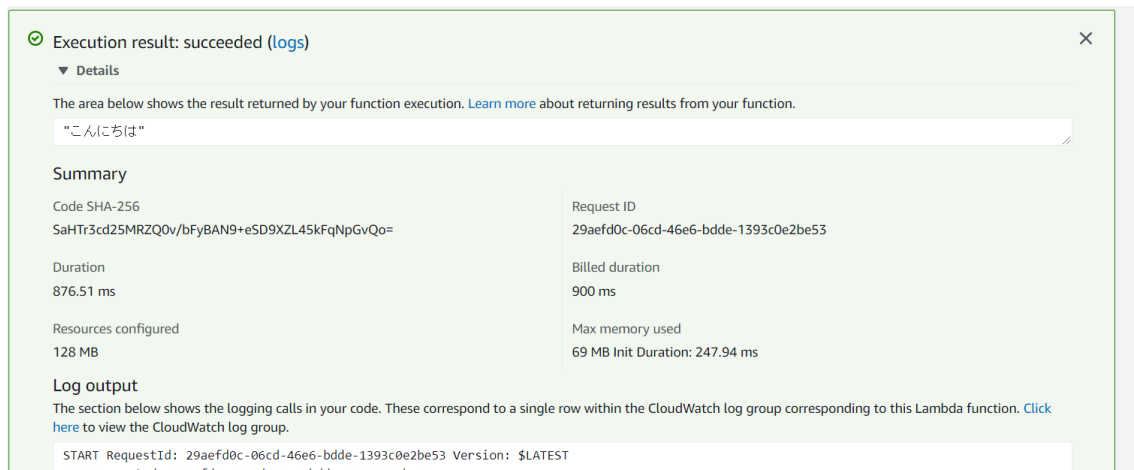
作成

2-9. Click [Test] again, and if you see something green message like below,  
translated job is successful.

✔ Execution result: succeeded ([logs](#))  
▶ Details

2-10. Click [Details], and you can see translated words.





This sample code is translate English to Japanese. You need to change

[TRG\_LANG = 'ja'] to your desired language. You can check if your native

language is supported by translate.

<https://docs.aws.amazon.com/translate/latest/dg/what-is.html#what-is-languages>

### 3. Amazon Polly

You used AWS Lambda for Amazon Translate test. Management console is used

for testing Amazon Polly because this process is done for later purpose recursively.

#### 3-1. Access top page of Amazon Polly

## Text-to-Speech

Listen, customize, and download speech. Integrate when you're ready.

Type or paste your text in the window, choose your language and region, choose a voice, choose Listen to speech, and then integrate it into your applications and services.

With up to 3000 characters you can listen, download, or save immediately. For up to 100,000 characters, your task must be saved to an S3 bucket.

The screenshot shows the Amazon Polly Text-to-Speech interface. At the top, there are tabs for 'Plain text' (selected) and 'SSML'. Below the tabs is a large text input area containing the text 'Hi! My name is Joanna. I will read any text you type here.' Below the input area, it says '58 characters used'. To the right of the input area are links for 'Show default text' and 'Clear text'. Below the input area, there are three main sections: 'Engine', 'Voice', and 'Actions'. The 'Engine' section has two radio buttons: 'Standard' (selected) and 'Neural'. The 'Voice' section has a list of voices: 'Salli, Female', 'Joanna, Female' (selected), 'Ivy, Female', 'Kendra, Female', 'Kimberly, Female', 'Matthew, Male', 'Justin, Male', and 'Joey, Male'. The 'Actions' section has three buttons: 'Listen to speech' (blue), 'Download MP3' (grey), and 'Synthesize to S3' (grey). Below the 'Download MP3' button, it says 'Sample rate: 22050Hz'. Below the 'Synthesize to S3' button, there are links for 'Change file format' and 'Change S3 task settings'.

Plain text SSML ?

Hi! My name is Joanna. I will read any text you type here.

58 characters used Show default text Clear text

Engine ?

☒ Standard  
☐ Neural

Language and Region

English, US

Voice

☐ Salli, Female  
☒ Joanna, Female  
☐ Ivy, Female  
☐ Kendra, Female  
☐ Kimberly, Female  
☐ Matthew, Male  
☐ Justin, Male  
☐ Joey, Male

▶ Listen to speech

Download MP3

Sample rate: 22050Hz

Change file format

Synthesize to S3

Change S3 task settings

3-2. By default, English is selected. Paste any appropriate English words. For later

purpose, please limit length of the word should be less than 150 long. (Yes,

Technically it works, but if you use long voice by Polly, Transcribe job gets longer

to finish)

3-3. Select any voice and click [Download MP3]. You can get mp3 file after a

while.

3-4. Play downloaded mp3 files, and check if it is created correctly. Save the file

for next step use.

If you click [Language and Region], you can check if your native language is

supported or not. You can also obtain full list from below URL. **CAUTION: Region**

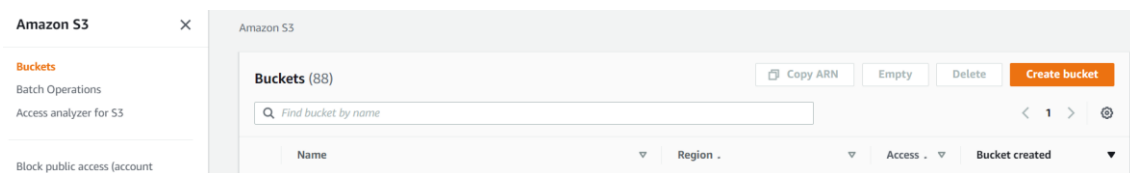
does not mean AWS Region. It is dialects in such country..

<https://docs.aws.amazon.com/polly/latest/dg/SupportedLanguage.html>

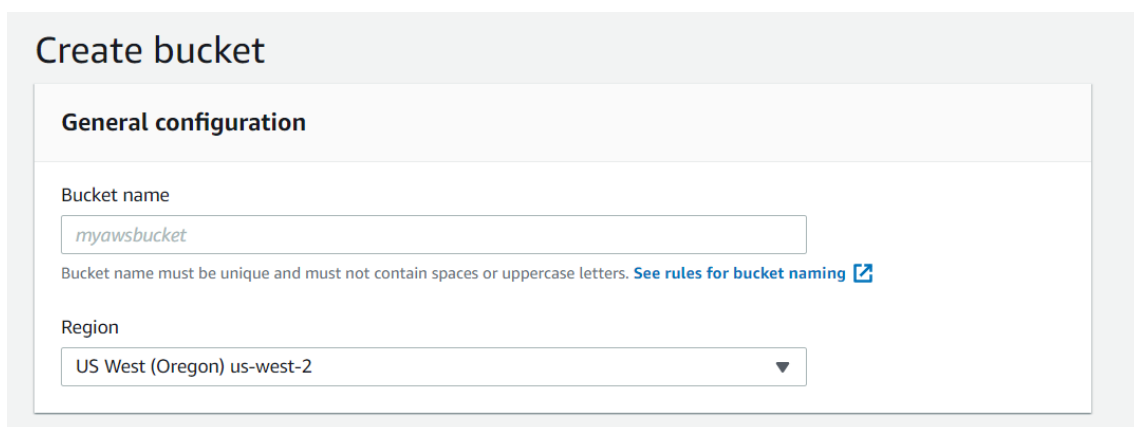
#### 4. Transcribe

Create Speech-to-Text job based on previously create mp3 by Polly.

4-1. Access Amazon S3 management console. Be aware that Region should be Oregon.



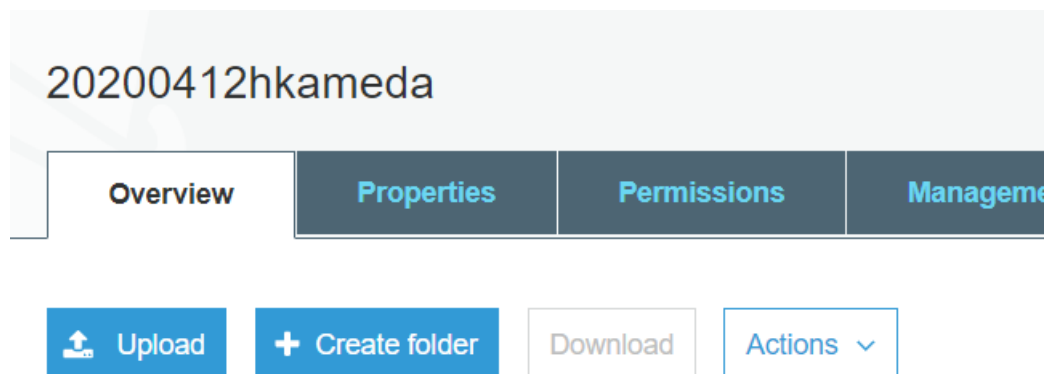
4-2. Click [Create bucket], and input YYYYMMDDName (YYYYMMDD is date of today) to [Bucket name].

A screenshot of the 'Create bucket' form in the Amazon S3 console. The form is titled 'Create bucket' and has a section for 'General configuration'. Inside this section, there is a 'Bucket name' field with the text 'myawsbucket' entered. Below the field is a note: 'Bucket name must be unique and must not contain spaces or uppercase letters. See rules for bucket naming'. There is also a 'Region' dropdown menu with 'US West (Oregon) us-west-2' selected.

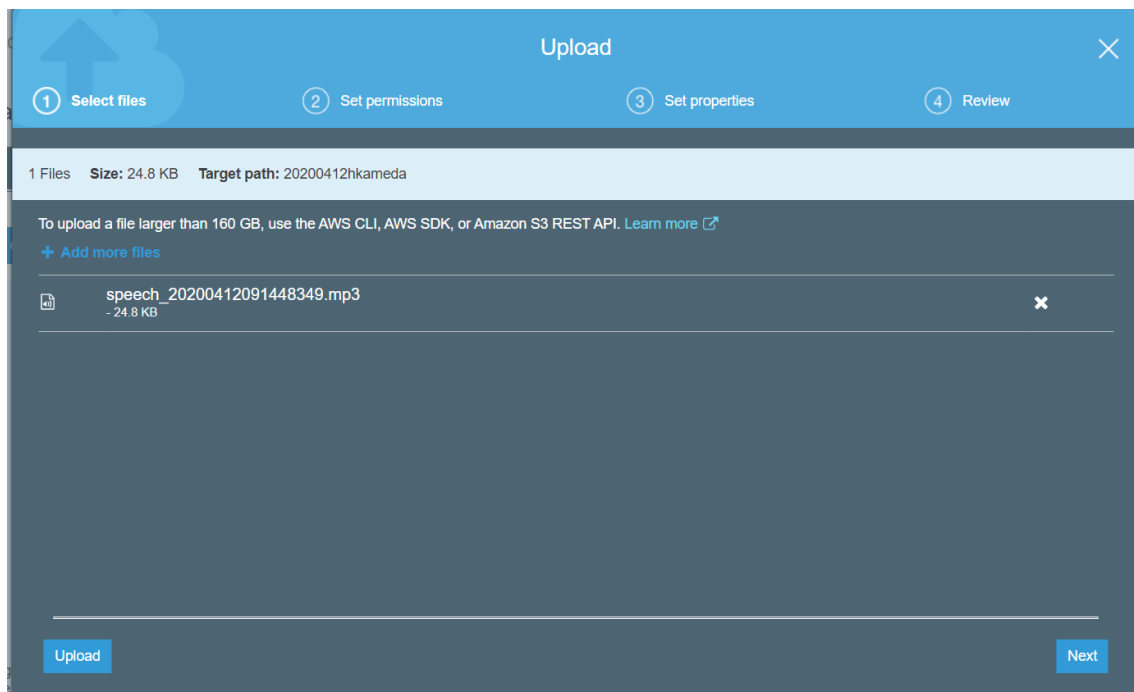
4-3. Click [Create bucket]

✔ **Successfully created bucket 20200412hkameda**  
To upload files and folders, or to configure additional bucket settings, click **Details**.

4-4. Click the bucket, and Click also [Upload]



4-5. Upload mp3 file created by Polly.



4-6. Click [Next] 3 times, and click [Upload] for final step.

Upload


+ Create folder

Download

Actions

US West (Oregon)

Viewing 1 to 1

| <input type="checkbox"/> | Name                                                                                                           | Last modified                    | Size    | Storage class |
|--------------------------|----------------------------------------------------------------------------------------------------------------|----------------------------------|---------|---------------|
| <input type="checkbox"/> |  speech_20200412091448349.mp3 | Apr 12, 2020 6:29:48 PM GMT+0900 | 24.8 KB | Standard      |

Viewing 1 to 1

4-6-1. The IAM role you have created at previous steps does not have access S3. You need to grant access to S3 onto the IAM role. Going back to IAM role management console, identify IAM role.

|                                     |          |                             |
|-------------------------------------|----------|-----------------------------|
| <a href="#">Create role</a>         |          | <a href="#">Delete role</a> |
| <input type="text" value="Q 0412"/> |          |                             |
| Role name ▾                         |          | Trusted entities            |
| <input type="checkbox"/>            | 20200412 | AWS service: lambda         |

4-6-2. Click the role, and click [Attach policies]

[Roles](#) > [20200412](#)
[Delete role](#)

### Summary

|                                  |                                                                                     |
|----------------------------------|-------------------------------------------------------------------------------------|
| Role ARN                         | arn:aws:iam::294963776963:role/20200412                                             |
| Role description                 | Allows Lambda functions to call AWS services on your behalf.   <a href="#">Edit</a> |
| Instance Profile ARNs            |                                                                                     |
| Path                             | /                                                                                   |
| Creation time                    | 2020-04-12 17:51 UTC+0900                                                           |
| Last activity                    | Not accessed in the tracking period                                                 |
| Maximum CLI/API session duration | 1 hour <a href="#">Edit</a>                                                         |

[Permissions](#)
[Trust relationships](#)
[Tags](#)
[Access Advisor](#)
[Revoke sessions](#)

▼ Permissions policies (4 policies applied)
 [Attach policies](#)
[Add inline policy](#)

| Policy name ▾              | Policy type ▾      |   |
|----------------------------|--------------------|---|
| TranslateFullAccess        | AWS managed policy | ✕ |
| AmazonTranscribeFullAccess | AWS managed policy | ✕ |

[Show 2 more](#)

4-6-3. Find and check [AmazonS3FullAccess], and click [Attach policy]

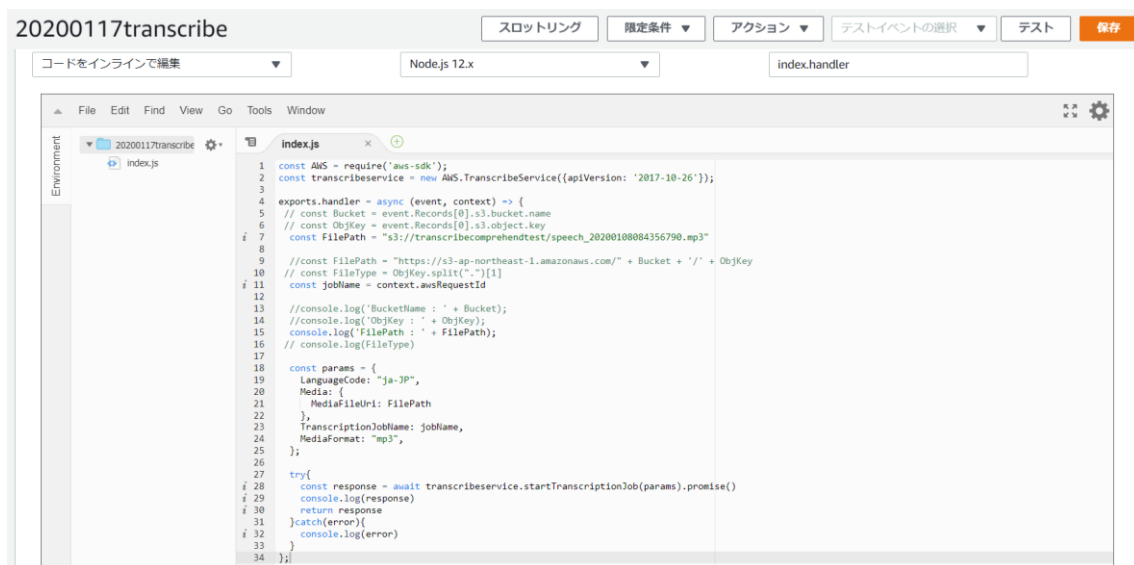
Policy AmazonS3FullAccess has been attached for the 20200412.

4-7. Going back to Lambda management console, and create

[YYYYMMDDtranscribe] functions as you did at Transcribe step. This time,

Runtime should be [Node12.x], not Python. Assign IAM role, which you are working on, to the Lambda function.

4-8. Copy and paste [transcribelambda.txt]



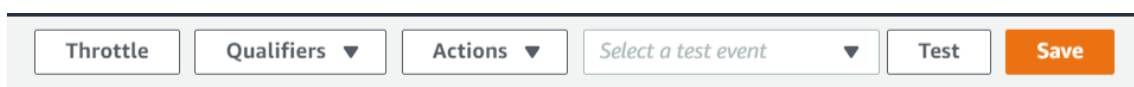
```
1 const AWS = require('aws-sdk');
2 const transcribeservice = new AWS.TranscribeService({apiVersion: '2017-10-26'});
3
4 exports.handler = async (event, context) => {
5   // const Bucket = event.Records[0].s3.bucket.name
6   // const ObjKey = event.Records[0].s3.object.key
7   const FilePath = "s3://transcribecomprehenstest/speech_20200108084356790.mp3"
8
9   //const FilePath = "https://s3-ap-northeast-1.amazonaws.com/" + Bucket + "/" + ObjKey
10  // const FileType = ObjKey.split("-")[1]
11  const jobName = context.awsRequestId
12
13  //console.log('BucketName : ' + Bucket);
14  //console.log('ObjKey : ' + ObjKey);
15  console.log('FilePath : ' + FilePath);
16  // console.log(FileType)
17
18  const params = {
19    LanguageCode: "ja-JP",
20    Media: {
21      MediaFileUri: FilePath
22    },
23    TranscriptionJobName: jobName,
24    MediaFormat: "mp3",
25  };
26
27  try{
28    const response = await transcribeservice.startTranscriptionJob(params).promise()
29    console.log(response)
30    return response
31  }catch(error){
32    console.log(error)
33  }
34  };
```

4-9. Replace [const FilePath] to your S3 bucket name and file name of mp3 file.

The pass should be like below;

s3://20200117kameda/speech\_20200117061325085.mp3

4-10. Click [Save]



4-11. Click [Test]. Input appropriate name to [Event name], and click [Create].

#This function does not require any test input from outside of function, but you need to do this anyway.

**Configure test event** [X]

A function can have up to 10 test events. The events are persisted so you can switch to another computer or web browser and test your function with the same events.

☒ Create new test event  
☐ Edit saved test events

Event template  
Hello World [v]

Event name  
testtranscribe

```
1 {  
2   "key1": "value1",  
3   "key2": "value2",  
4   "key3": "value3"  
5 }
```

4-12. Click [Test] again, and you will see [IN\_PROGRESS] in the case of successfully Transcribe job has been is started.

20200412transcribe [Throttle] [Qualifiers v] [Actions v] testtranscribe [v] [Test] [Save]

✓ Execution result: succeeded (logs) [X]

▼ Details

The area below shows the result returned by your function execution. [Learn more](#) about returning results from your function.

```
{  
  "TranscriptionJob": {  
    "TranscriptionJobName": "24c1829f-68e1-431e-92e0-0b4d357cd7f0",  
    "TranscriptionJobStatus": "IN_PROGRESS",  
    "LanguageCode": "en-US",  
    "MediaFormat": "mp3",  
    "Media": {  
      "MediaFileUri": "s3://20200412hkameda/speech_20200412091448349.mp3"  
    }  
  },  
  "StartTime": "2020-04-12T09:44:19.269Z",  
}
```

4-13. Going back to management console of Transcribe, The project is in progress, and you need to wait for a while until it turns to complete.

Amazon Transcribe > Transcription jobs

### Transcription jobs

Status: All

1

Download

Delete

Copy

Create job

|                       | Name                                 | Language         | Created            | Expires in | Status      |
|-----------------------|--------------------------------------|------------------|--------------------|------------|-------------|
| <input type="radio"/> | 33d6ce0d-a034-4508-a3c3-9ea97846efb1 | Japanese (Japan) | 2020/1/17 15:35:04 | -          | In progress |

[Amazon Transcribe](#) > [Transcription jobs](#) > 33d6ce0d-a034-4508-a3c3-9ea97846efb1




33d6ce0d-a034-4508-a3c3-9ea97846efb1

Download full transcript

Delete

Copy

Job details

|                                                                                            |                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|--------------------------------------------------------------------------------------------|----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Name                                                                                       | Ended                | Vocabulary filter                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| 33d6ce0d-a034-4508-a3c3-9ea97846efb1                                                       | 2020/1/17 15:36:20   | -                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| Status                                                                                     | Input file format    | Vocabulary filtering method                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|  Complete | mp3                  | -                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| Language                                                                                   | Audio sampling rate  | Input data location                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| Japanese (Japan)                                                                           | 22050 Hz             | s3://20200117kameda/speech_20200117061325085.mp3                                                                                                                                                                                                                                                                                                                                 |
| Expiration <a href="#">Info</a>                                                            | Audio identification | Output data location                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| The transcription is available for 89 more days.                                           | None                 | <a href="https://s3.ap-northeast-1.amazonaws.com/aws-transcribe-ap-northeast-1-prod/294963776963/33d6ce0d-a034-4508-a3c3-9ea97846efb1/acf001d9-8a73-47bc-8a4a-2bcf8ddcf912/asrOutput.json">https://s3.ap-northeast-1.amazonaws.com/aws-transcribe-ap-northeast-1-prod/294963776963/33d6ce0d-a034-4508-a3c3-9ea97846efb1/acf001d9-8a73-47bc-8a4a-2bcf8ddcf912/asrOutput.json</a>  |
| Created                                                                                    | Alternative results  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| 2020/1/17 15:35:04                                                                         | Disable              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| Started                                                                                    | Custom vocabulary    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| 2020/1/17 15:35:04                                                                         | None                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |

```

[{"id": "編集(書式) 表示(ヘルプ)", "jobName": "33d0e0ad-a034-4508-ac3c-9ea97846efb1", "accountId": "294963776963", "results": {"transcripts": [{"transcript": "谷村 使用 すれば 実質 と いう 様な タイプ の プリテンション や パーセント にも 管理 を 必要 と せず 実行 でき ます コード の スクリプト すれば 効果 要請 を 実現 し ながら コード を 実行 する ために 必要 と する こと が 減り ます 実行 する こと が 減る こと が 目的 として 自動化 により 力 を 加え ます よう 設定 する こと が 必要 だ と思います", "confidence": 0.4108, "content": "谷村", "type": "pronunciation", "start_time": 0.5, "end_time": 0.59, "alternatives": [{"confidence": 1.0, "content": "谷村", "type": "pronunciation"}], "start_time": 0.59, "end_time": 0.93, "alternatives": [{"confidence": 1.0, "content": "使用", "type": "pronunciation"}], "start_time": 0.93, "end_time": 1.24, "alternatives": [{"confidence": 1.0, "content": "すれば", "type": "pronunciation"}], "start_time": 1.24, "end_time": 1.4, "alternatives": [{"confidence": 1.0, "content": "谷村", "type": "pronunciation"}], "start_time": 1.4, "end_time": 1.6, "alternatives": [{"confidence": 1.0, "content": "使用", "type": "pronunciation"}], "start_time": 1.6, "end_time": 2.42, "alternatives": [{"confidence": 0.9987, "content": "実質", "type": "pronunciation"}], "start_time": 2.42, "end_time": 2.61, "alternatives": [{"confidence": 0.9987, "content": "谷村", "type": "pronunciation"}], "start_time": 2.61, "end_time": 2.73, "alternatives": [{"confidence": 1.0, "content": "谷村", "type": "pronunciation"}], "start_time": 2.73, "end_time": 3.13, "alternatives": [{"confidence": 1.0, "content": "谷村", "type": "pronunciation"}], "start_time": 3.13, "end_time": 3.23, "alternatives": [{"confidence": 1.0, "content": "谷村", "type": "pronunciation"}], "start_time": 3.23, "end_time": 4.12, "alternatives": [{"confidence": 1.0, "content": "アプリケーション", "type": "pronunciation"}], "start_time": 4.12, "end_time": 4.32, "alternatives": [{"confidence": 1.0, "content": "谷村", "type": "pronunciation"}], "start_time": 4.32, "end_time": 5.03, "alternatives": [{"confidence": 0.9538, "content": "谷村", "type": "pronunciation"}], "start_time": 5.03, "end_time": 5.59, "alternatives": [{"confidence": 0.9539, "content": "谷村", "type": "pronunciation"}], "start_time": 5.59, "end_time": 5.84, "alternatives": [{"confidence": 1.0, "content": "谷村", "type": "pronunciation"}], "start_time": 5.84, "end_time": 5.94, "alternatives": [{"confidence": 0.8933, "content": "谷村", "type": "pronunciation"}], "start_time": 5.94, "end_time": 6.1, "alternatives": [{"confidence": 0.8933, "content": "谷村", "type": "pronunciation"}], "start_time": 6.1, "end_time": 6.47, "alternatives": [{"confidence": 0.8993, "content": "管理", "type": "pronunciation"}], "start_time": 6.47, "end_time": 6.57, "alternatives": [{"confidence": 1.0, "content": "谷村", "type": "pronunciation"}], "start_time": 6.57, "end_time": 7.02, "alternatives": [{"confidence": 1.0, "content": "谷村", "type": "pronunciation"}]}]}

```

<https://docs.aws.amazon.com/transcribe/latest/dg/how-it-works.html>



your desired language.

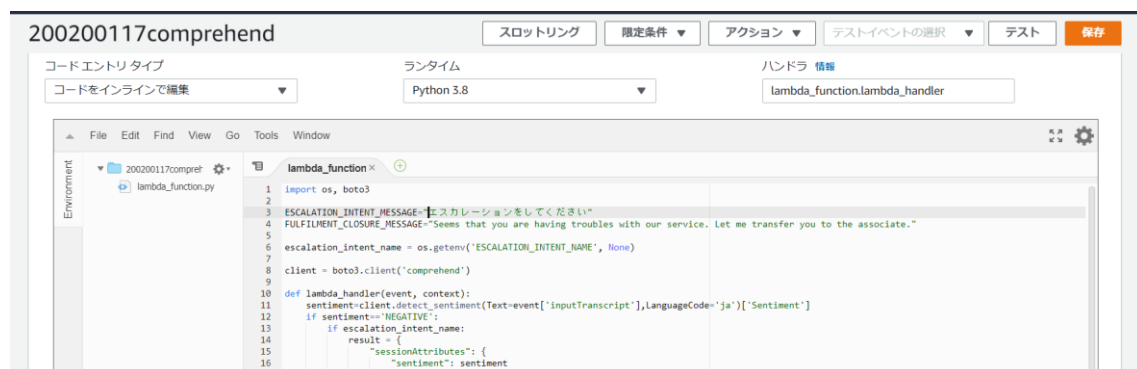
## 5. Comprehend

Copy and paste appropriate English sentence for test purpose.

5-1. Access AWS Lambda management console again, and create function with name of [YYYYMMDDComprehend]. [Runtime] should be Python3.8 this time.

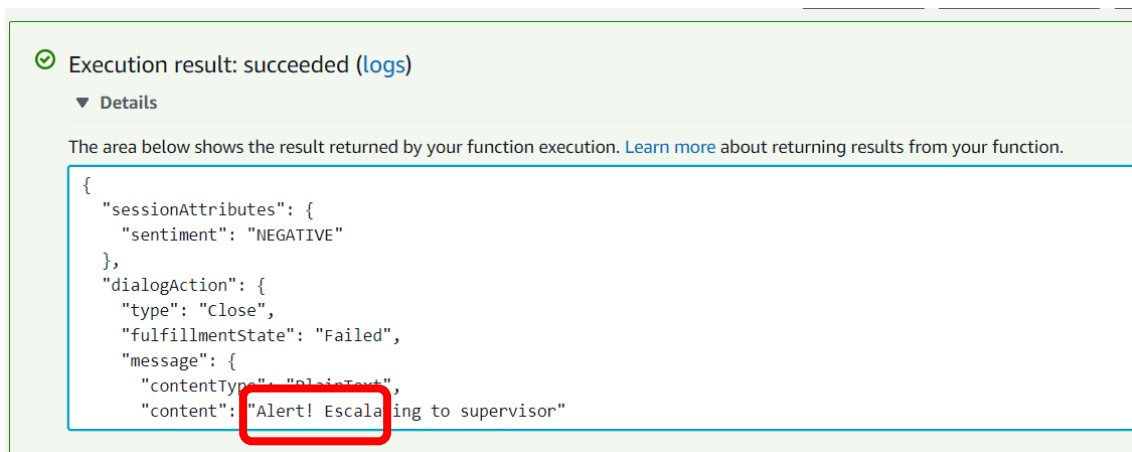
Please be sure to assign IAM role to your function.

5-2. Copy and paste [comprehendlambda.txt]. If input message is negative (like “I hate you”), the script is going to output message of “Alert!”

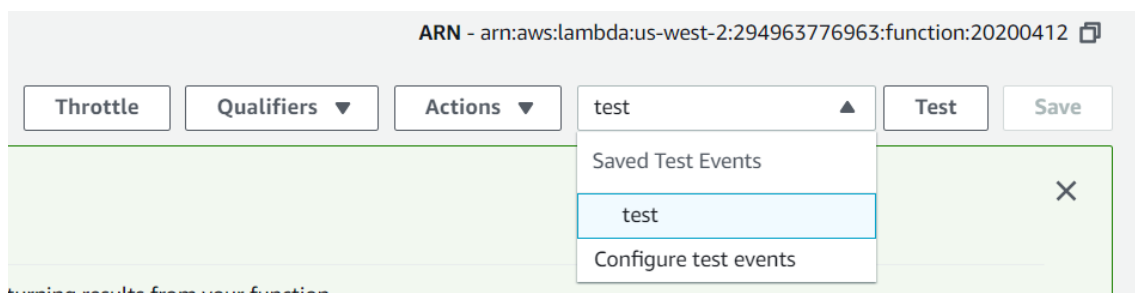


5-3. Click [Save], and Click again [Test]. Input appropriate name to [Event name] and copy and paste [comprehendlambdatest.txt]. Click [Create].

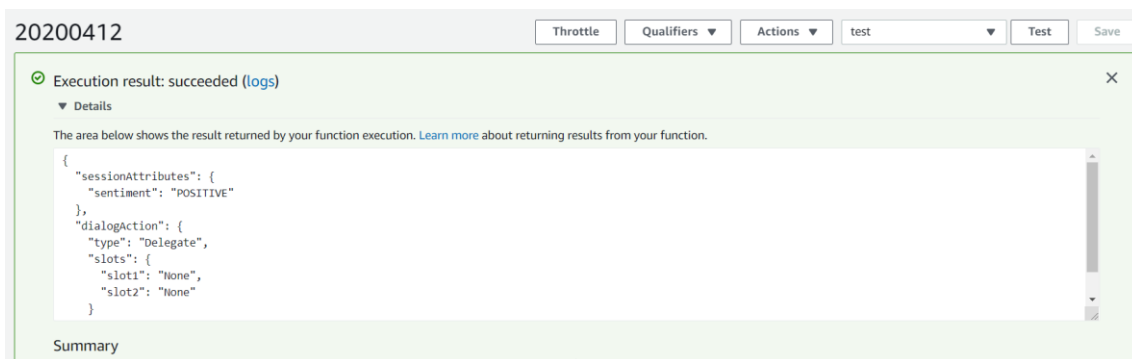
5-4. Test input contains sentence of “I hate you”. If you run the test, you are going to see “Alert”



5-5. Click [Test] again, and replace “I hate you” to “I love you” by selecting [Configure test events].



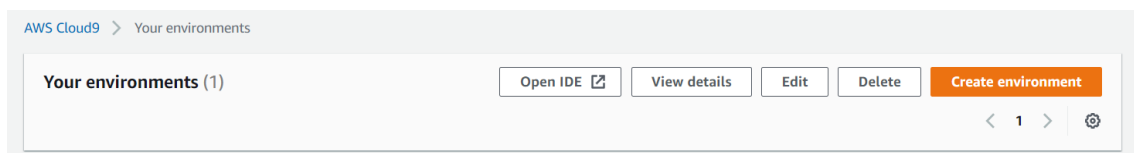
5-6. You may see “POSITIVE” this time.



6. This is final, but very long step here. We Combine Transcribe and Comprehend via Python script, and create dashboard and visualize result of analysis on Amazon

Elasticsearch service. Cloud9 is used for this scenario. Cloud9 takes over IAM rights to access any AWS services from your IAM user you are logging in AWS account by default. Please be sure that your IAM user has access to all Comprehend, Transcribe, Elasticsearch Service, S3, Cloud9.

6-1. Access AWS Cloud9 management console.



6-2. Click [Create Environment]. Input YYYYMMDDhandson (YYYYMMDD is date of today) to the Name, and Click [Next Step]

## Name environment

### Environment name and description

**Name**  
The name needs to be unique per user. You can update it at any time in your environment settings.

Limit: 60 characters

**Description - Optional**  
This will appear on your environment's card in your dashboard. You can update it at any time in your environment settings.

Write a short description for your environment

Limit: 200 characters

CancelNext step

6-3. Select [Network Settings], and Choose VPC which has Public subnet. (in the most case, default VPC works great)

The screenshot shows the 'Network settings (advanced)' section of the AWS Cloud9 console. It contains two main configuration areas: 'Network (VPC)' and 'Subnet'. The 'Network (VPC)' section has a dropdown menu showing 'vpc-0f86d0fce5ac8aca0 (default)' and a 'Create new VPC' button. The 'Subnet' section has a dropdown menu showing 'No preference (default subnet in any Availability Zone)' and a 'Create new subnet' button. At the bottom of the panel are three buttons: 'Cancel', 'Previous step', and 'Next step'.

▼ Network settings (advanced)

**Network (VPC)**  
Launch your EC2 instance into an existing Amazon Private Cloud (VPC) or create a new one.

vpc-0f86d0fce5ac8aca0 (default) [Refresh] [Create new VPC]

**Subnet**  
Select a range of IP addresses in your VPC to isolate EC2 resources from each other.

No preference (default subnet in any Availability Zone) [Refresh] [Create new subnet]

Cancel Previous step Next step

6-4. Click [Next Step]

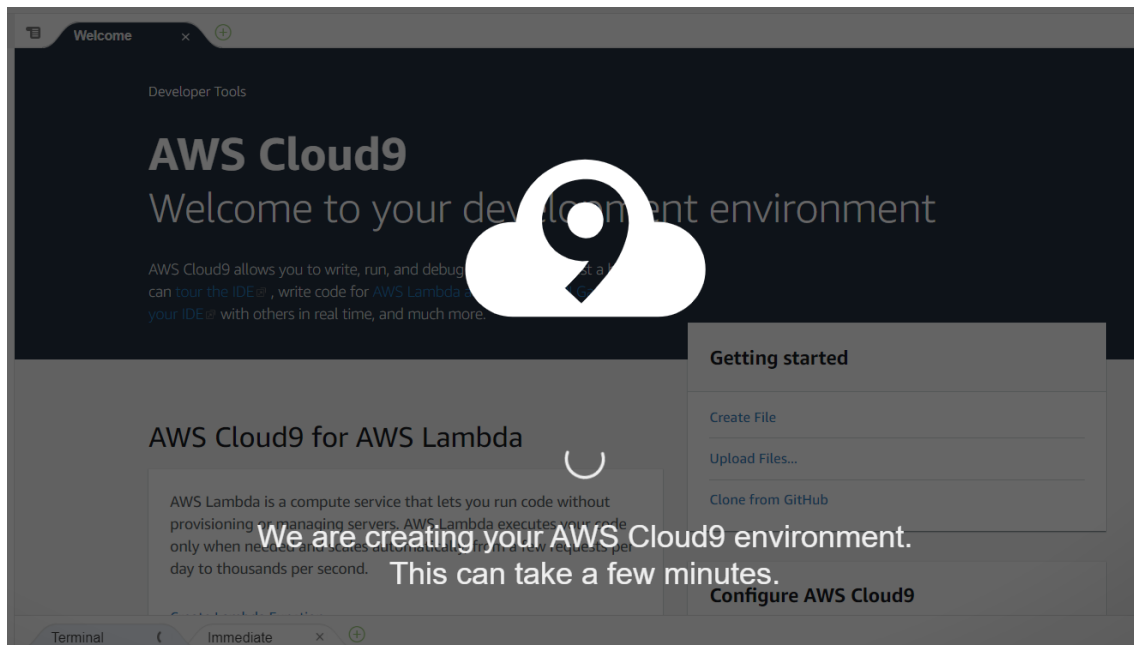
The screenshot shows the 'Create environment' screen in the AWS Cloud9 console. A large light blue box contains a warning icon and the text: 'We recommend the following best practices for using your AWS Cloud9 environment'. Below this are four bullet points with links for more information. At the bottom of the screen are three buttons: 'Cancel', 'Previous step', and 'Create environment'.

**We recommend the following best practices for using your AWS Cloud9 environment**

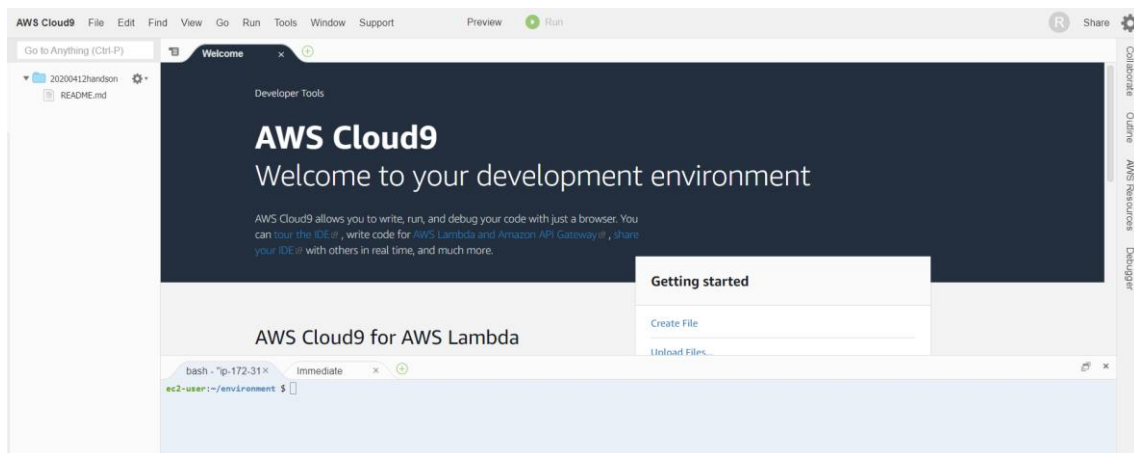
- Use **source control and backup** your environment frequently. AWS Cloud9 does not perform automatic backups.
- Perform regular **updates of software** on your environment. AWS Cloud9 does not perform automatic updates on your behalf.
- **Turn on AWS CloudTrail in your AWS account** to track activity in your environment. [Learn more](#)
- Only share your environment with **trusted users**. Sharing your environment may put your AWS access credentials at risk. [Learn more](#)

Cancel Previous step Create environment

6-5. Click [Create environment], and you will see wait screen



6-6. Launching Cloud9 takes several minutes. After launched, you can access to console. If it takes more than 15mins, You might failed to assign correct VPC with public subnet. Try again with choosing another VPC/Subnet



6-7. Type below command to create S3 bucket at Oregon Region

```
aws s3 mb s3://<yyyymmdd>transcribe<name> --region us-west-
```

(YYYYMMDD is date of today, name is your nickname)

```
ec2-user:~/environment $ aws s3 mb s3://20200117transcribekameda --region us-west-2
make_bucket: 20200117transcribekameda
```


6-8. Check S3 management console if S3 bucket is surely created.

6-9. Access to Elasticsearch service management console

6-10. Click [Create a new domain]

Choose deployment type ?

Deployment types specify common settings for your use case. After creating the domain, you can change these settings at any time.

 UltraWarm <sup>PREVIEW</sup> is available in us-west-2, us-east-1, and us-east-2. Change the region selected to us-west-2, us-east-1, or us-east-2 to create a preview cluster. [Learn more](#)

**Deployment type**

- ☒ **Production**  
Multiple Availability Zones and dedicated master nodes for higher availability.
- ☐ **Development and testing**  
One Availability Zone for when you just need an Elasticsearch endpoint.
- ☐ **Custom**  
Choose settings from all available options.
- ☐ **UltraWarm Preview**  
Choose to preview UltraWarm.

6-11. Select [Development and testing]

Version

Select the version of Elasticsearch for your domain.

Elasticsearch version

[Cancel](#) [Next](#)

6-12. Select 6.2, and click [Next]

6-13. Input [transcribetestYYYYMMDD](YYYYMMDD is date of today) into

Elasticsearchdomain name

## Configure domain



A domain is the collection of resources needed to run Elasticsearch. The domain name will be part of your domain endpoint.

Elasticsearch domain name

transcrietest20200412

The name must start with a lowercase letter and must be between 3 and 28 characters. Valid characters are a-z (lowercase only), 0-9, and - (hyphen).

### 6-14. Click [Next]

### 6-15. Select [Public access] under Network configuration.

#### Network configuration

Choose Internet or VPC access. To enable VPC access, we use private IP addresses from your VPC, which provides an inherent layer of security. You control network access within your VPC using security groups. Optionally, you can add an additional layer of security by applying a restrictive access policy. Internet endpoints are publicly accessible. If you select public access, you should secure your domain with an access policy that only allows specific users or IP addresses to access the domain.

- ☒ VPC access (Recommended)  
☐ Public access

VPC

Select VPC



# You can create domain inside of VPC instead of Public for security. But workshop protocol gets long. Thus, we skip it this time.

### 6-16. Set up Access policy as below; (parameter is [\*])

#### Access policy

Access policies control whether a request is accepted or rejected when it reaches the Amazon Elasticsearch Service domain. If you specify an account, user, or role in this policy, you must sign your requests. [Learn more](#)

Custom policy builder allows at most 10 elements. Use a JSON-defined access policy to define a policy with more than 10 elements.

Domain access policy

Custom access policy

Allow or deny access by AWS account ID, account ARN, IAM user ARN, IAM role ARN, IPv4 address, or CIDR block.

IPv4 address

\*

Allow

Remove element

IAM ARN

\*

Allow

Remove element

[Add element](#)

### 6-17. Stay Encryption setting as default to require HTTPS access to the domain.

## Encryption


These features help protect your data. After creating the domain, you can't change most encryption settings.

- Encryption** ☒ Require HTTPS for all traffic to the domain ⓘ
- ☐ Node-to-node encryption ⓘ
- ☐ Enable encryption of data at rest ⓘ

6-18. Click [Next], and Click [Confirm] at next page.

6-19. The domain is now under construction. When domain is created, you will see

Endpoint and Kibana URL. Copy those for next step.

 You have successfully created an Elasticsearch domain.  
You can update your domain configuration or access policy now. The domain endpoint will be available once the domain has reached an Active status.

transcrietest20200412

[Edit domain](#) [Actions](#)

Your domain is being initialized, which takes about 10 minutes. You cannot load data or run queries against your domain until the initialization is complete. The domain status will change to Active as soon as your domain is ready to use.

Overview

Cluster health

Instance health

Indices

Logs

Upgrade history

Domain status

Loading

Elasticsearch version

6.2

Endpoint

Domain ARN

arn:aws:es:us-west-2:294963776963:domain/transcrietest20200412

Kibana

Availability zones

1

Instance type (data)

r5.large.elasticsearch

Number of nodes

1

6-20. Type below 2 commands at Cloud9 console. Please type directly without copy.

(If you copy and paste at windows environment, [--] changes to [-])

```
python --version
```

```
pip --version
```



If pip version points to version 2.x, type below command.

```
sudo update-alternatives --config python
```

```
Selection      Command
-----
*+ 1           /usr/bin/python2.7
   2           /usr/bin/python3.6

Enter to keep the current selection[+], or type selection number: █
```

Select 2 and press Enter to change pip version from 2.x to 3.x.

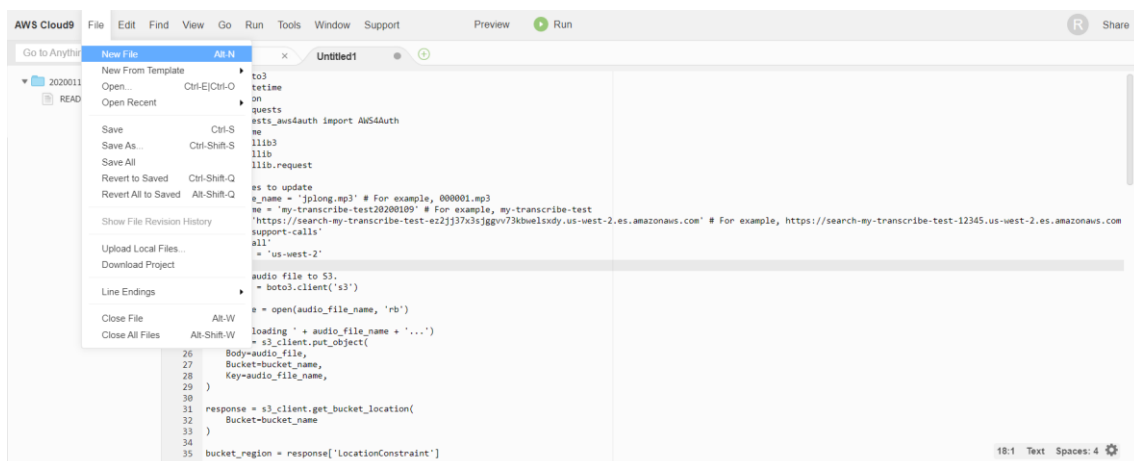
6-21. Type below command to install required libraries.

```
pip install boto3
```

```
pip install requests
```

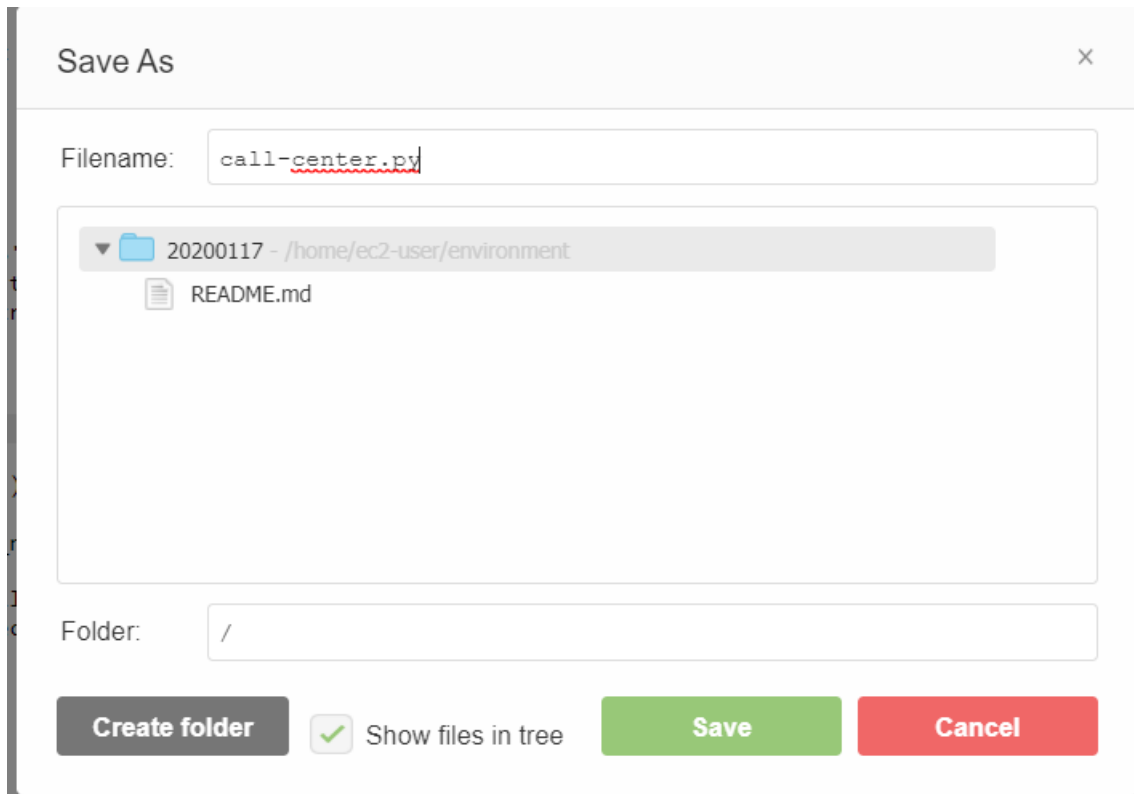
```
sudo pip install requests_aws4auth
```

6-22. Select [File] → [New File], and copy and paste [call-center.py] into the new tab.

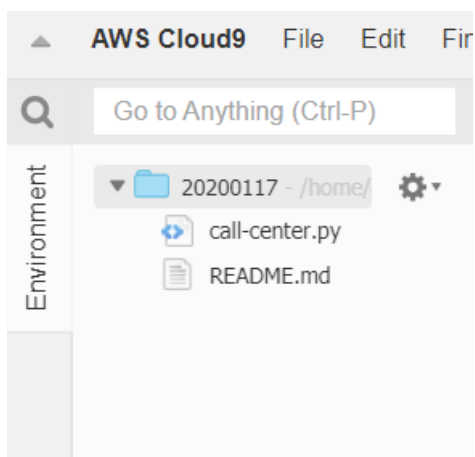


6-23. Click [File] → [Save], and type [call-center.py] to the filename. Then, click

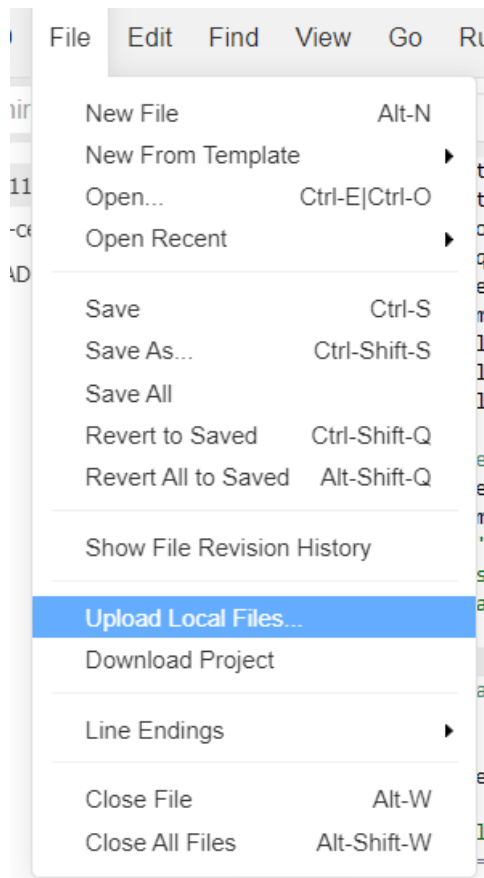
[Save]



6-24. The folder looks like below

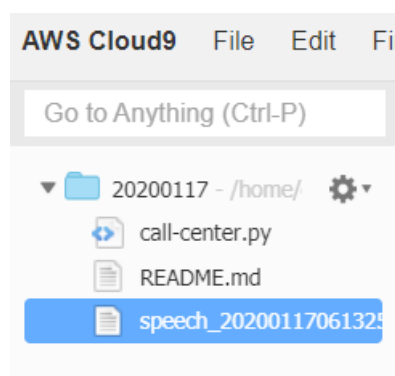


6-25. Select [File] → [Upload Local Files], and upload mp3 file which you created last time. (you can create new mp3 file by Polly if you want)



6-26. Upload dialog does not disappear automatically even though upload is success.

Thus, click [x] to close the dialog. The folder looks like below;



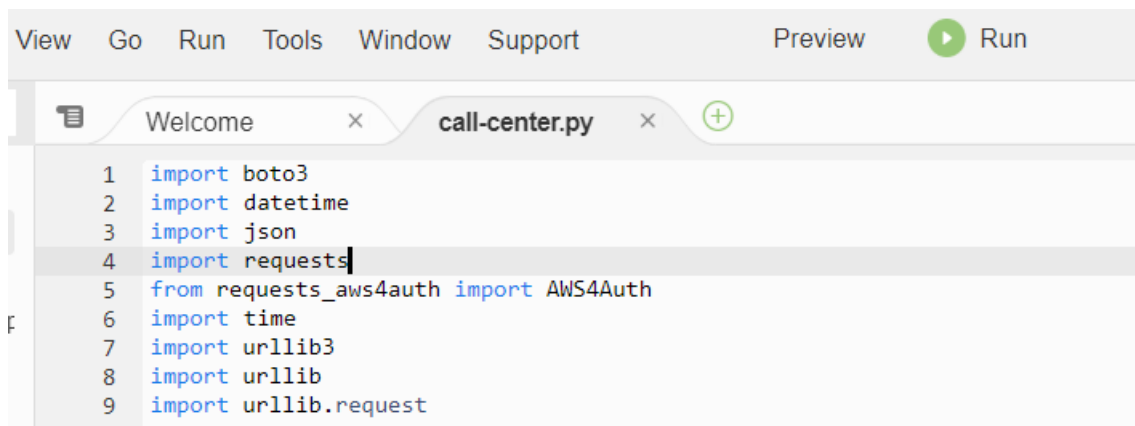
6-27. modify Python script of 12<sup>th</sup>, 13<sup>th</sup>, and 14<sup>th</sup> line.

12<sup>th</sup> line : uploaded mp3 file name

13<sup>th</sup> : S3 bucket name created from Cloud9

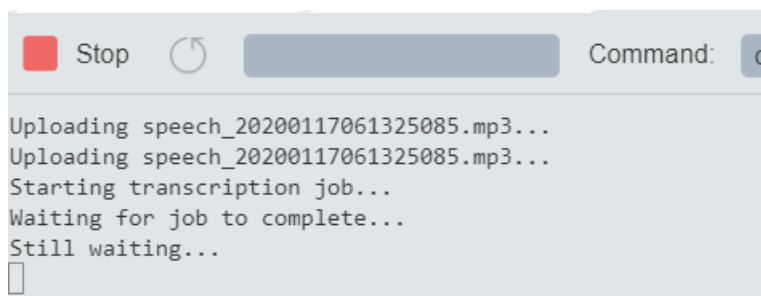
14<sup>th</sup> : Elasticsearch service Endpoint (not Kibana URL)

6-28. Save the file, and Let's Run the script. You can click Run button.



```
View Go Run Tools Window Support Preview Run
Welcome x call-center.py x
1 import boto3
2 import datetime
3 import json
4 import requests
5 from requests_aws4auth import AWS4Auth
6 import time
7 import urllib3
8 import urllib
9 import urllib.request
```

6-29. If you see below, it is working correctly.



```
Stop [refresh] Command: ca
Uploading speech_20200117061325085.mp3...
Uploading speech_20200117061325085.mp3...
Starting transcription job...
Waiting for job to complete...
Still waiting...
[ ]
```

6-30. If you access to Transcribe management console, Job is in progress

Amazon Transcribe > Transcription jobs

Transcription jobs

Download

Delete

Copy

Create job

Q Search job names

Status: All

< 1 > ⚙

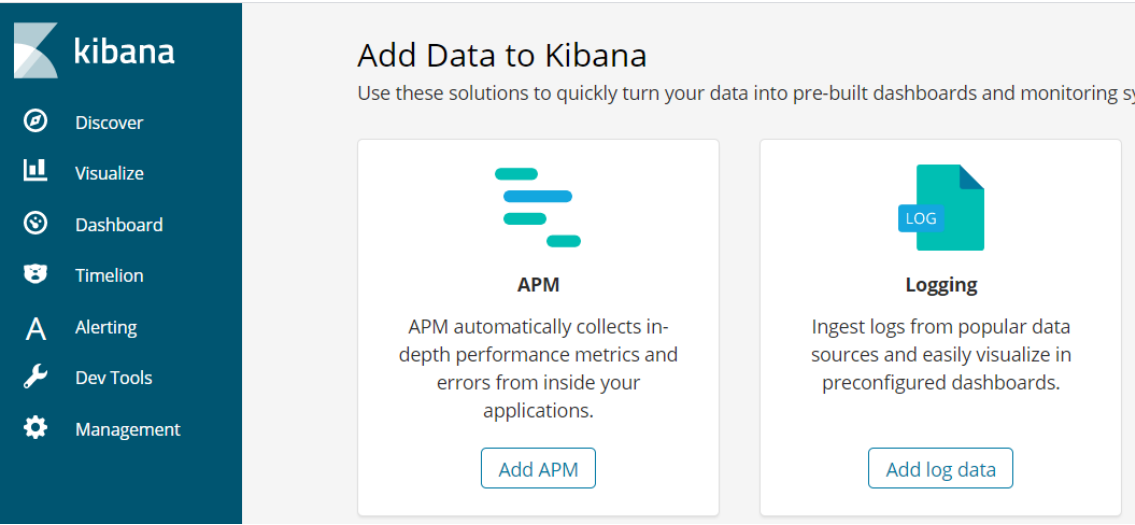
|                       | Name                                         | Language         | Created            | Expires in | Status      |  |
|-----------------------|----------------------------------------------|------------------|--------------------|------------|-------------|--|
| <input type="radio"/> | <a href="#">speech_20200117061325085.mp3</a> | Japanese (Japan) | 2020/1/17 16:54:37 | -          | In progress |  |

z6-31. If you see [Process exited with code: 0], The job is done, and transcribe text is passed to comprehend, and its analysis result was imported to Elasticsearch.

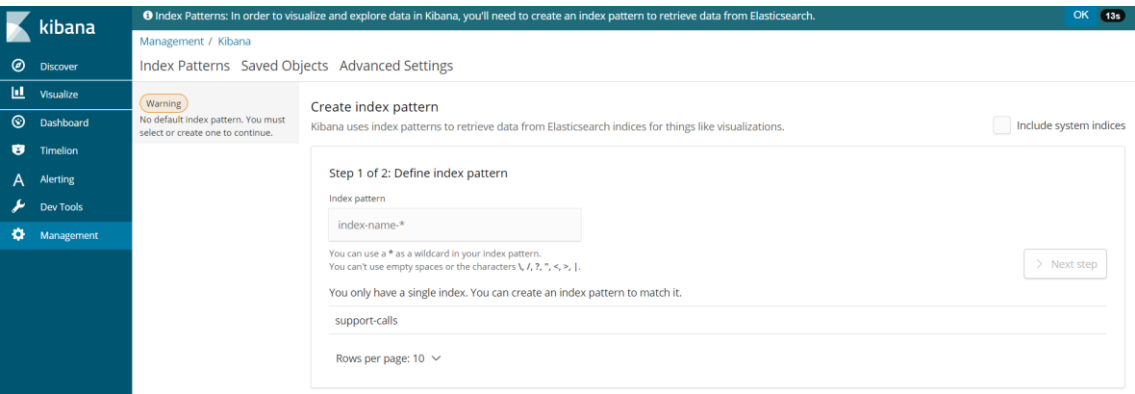
```
Still waiting...
Still waiting...
Still waiting...
Still waiting...
Still waiting...
Still waiting...
Still waiting...
Still waiting...
Detecting key phrases...
Detecting sentiment...
Indexing document...
<Response [201]>
{
  "_id": "speech-20200117061325085",
  "_type": "call",
  "_id": "speech_20200117061325085",
  "_version": 1,
  "result": {
    "created": true,
    "shards": {
      "total": 2,
      "successful": 1,
      "failed": 0
    },
    "_seq_no": 0,
    "_primary_term": 1
  }
}

Process exited with code: 0
```

6-31. Access to Kibana URL via your browser.



6-32. Click [Visualize]



6-32. Type [support\*] to [Index Pattern], and click [Next Step]

### Create index pattern

Kibana uses index patterns to retrieve data from Elasticsearch indices for things like visualizations.

☐ Include system indices

**Step 1 of 2: Define index pattern**

Index pattern

support\*

You can use a \* as a wildcard in your index pattern.  
You can't use empty spaces or the characters \, /, ?, ", <, >, |.

> Next step

✓ **Success!** Your index pattern matches **1 index**.

support-calls

Rows per page: 10 ▾

6-33. Select timestamp under [Time Filter field name], and click [Create Index

Pattern]

### Create index pattern

Kibana uses index patterns to retrieve data from Elasticsearch indices for things like visualizations.

**Step 2 of 2: Configure settings**

You've defined **support\*** as your index pattern. Now you can specify some settings before we create it.

Time Filter field name

Refresh

timestamp ▾

The Time Filter will use this field to filter your data by time.

You can choose not to have a time field, but you will not be able to narrow down your data by a time range.

> Show advanced options

< Back

Create index pattern

6-34. If you see below screen, you correctly set index.

## ★ support\*



Time Filter field name: timestamp

This page lists every field in the **support\*** index and the field's associated core type as recorded by Elasticsearch. While this list allows you to view the core type of each field, changing field types must be done using Elasticsearch's [Mapping API](#).

fields (12)

scripted fields (0)

source filters (0)

Q Filter

All field types ▾

| name ⇅            | type ⇅  | format ⇅ | searchable ⓘ ⇅ | aggregatable ⓘ ⇅ | excluded ⓘ ⇅ | controls |
|-------------------|---------|----------|----------------|------------------|--------------|----------|
| _id               | string  |          | ✓              | ✓                |              |          |
| _index            | string  |          | ✓              | ✓                |              |          |
| _score            | number  |          |                |                  |              |          |
| _source           | _source |          |                |                  |              |          |
| _type             | string  |          | ✓              | ✓                |              |          |
| keywords          | string  |          | ✓              |                  |              |          |
| keywords.keyword  | string  |          | ✓              | ✓                |              |          |
| sentiment         | string  |          | ✓              |                  |              |          |
| sentiment.keyword | string  |          | ✓              | ✓                |              |          |
| timestamp ⓘ       | date    |          | ✓              | ✓                |              |          |

6-35. Click [Visualize] again.

+
0-0 of 0 < >

Looks like you don't have any visualizations. Let's create some!

+ Create a visualization

0 items selected
0-0 of 0 < >

6-36. Click [Create a visualization]

### Select visualization type

Basic Charts

Area

Heat Map

Horizontal Bar

Line

Pie

Vertical Bar

6-37. Select Pie chart

From a New Search, Select Index

Filter... 1 of 1

Name ▲

support\*

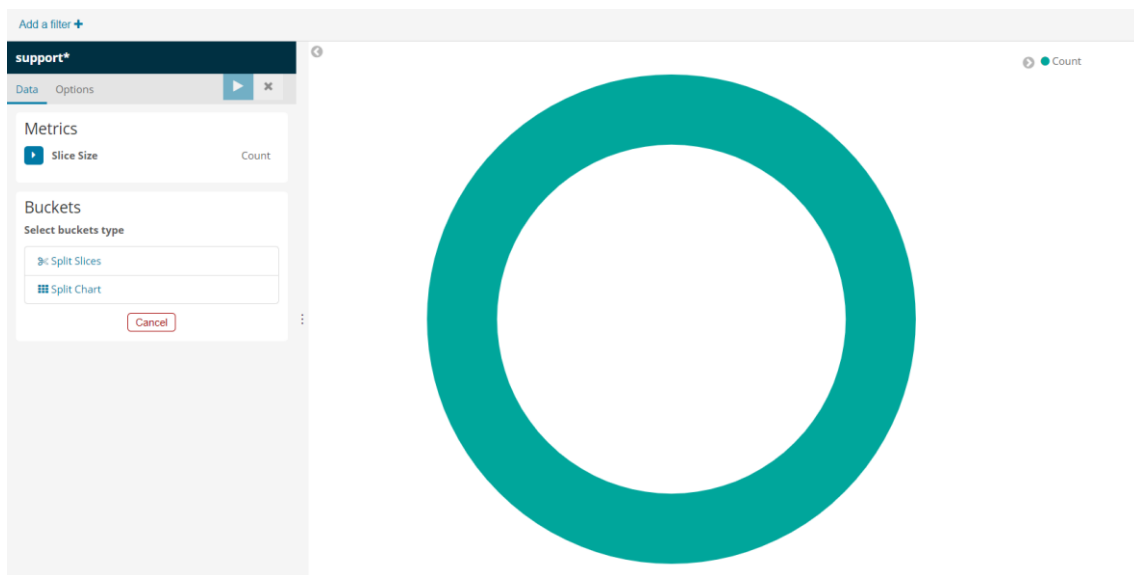
Or, From a Saved Search

Saved Searches Filter... 0-0 of 0 [Manage saved searches](#)

Name ▲

No matching saved searches found.

6-38. Select [support\*\*]. (you will see 2 [\*\*].)



6-39. Select [Split Slices] under the Buckets

**Metrics**

☒ Slice Size Count

**Buckets**

☒ Split Slices ☐ ☒

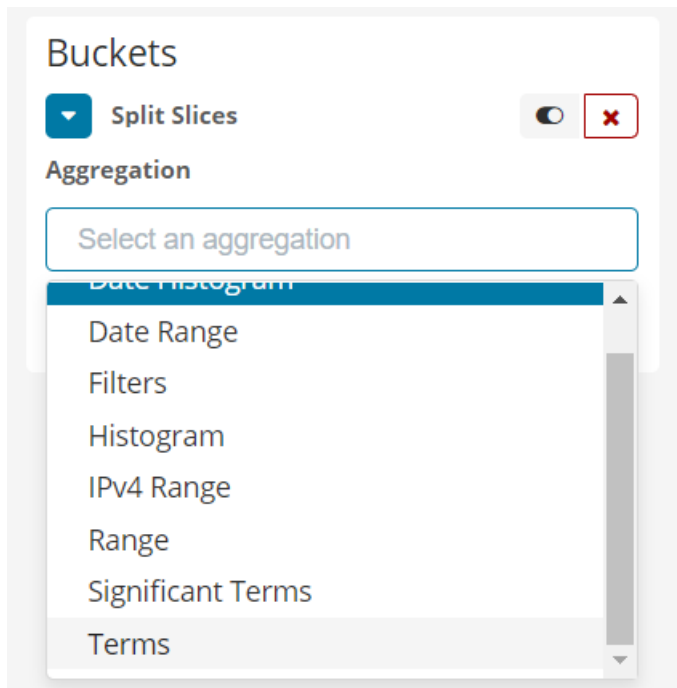
**Aggregation**

Select an aggregation ▼

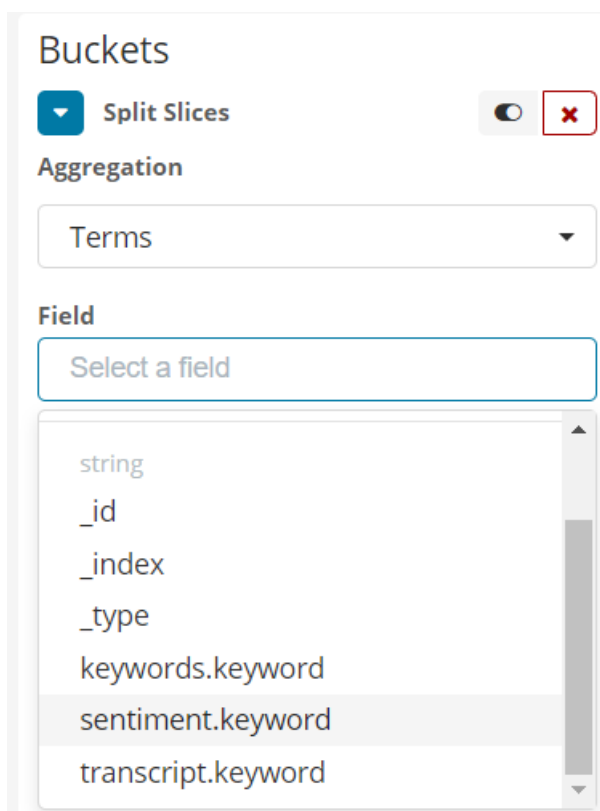
[Add sub-buckets](#)

6-40. Select [Terms] under [Aggregation] field

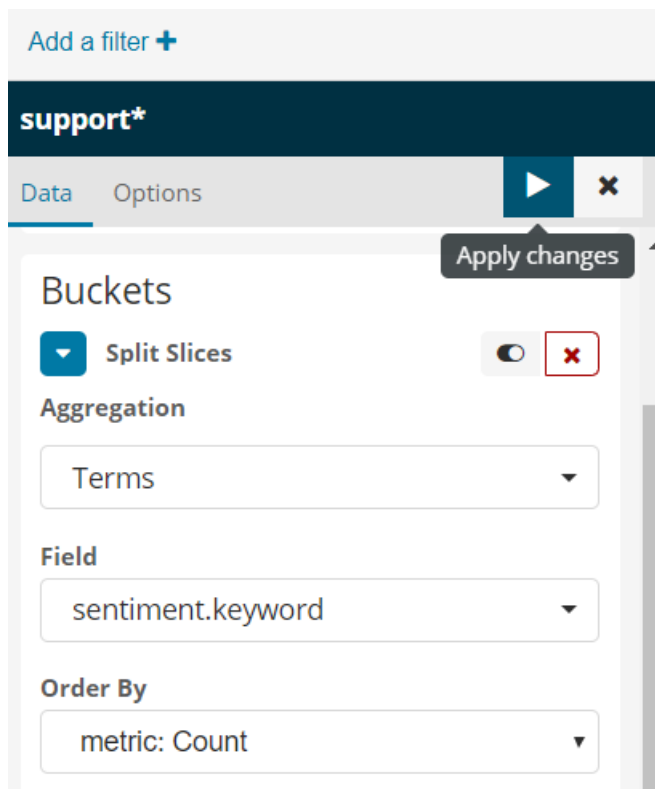




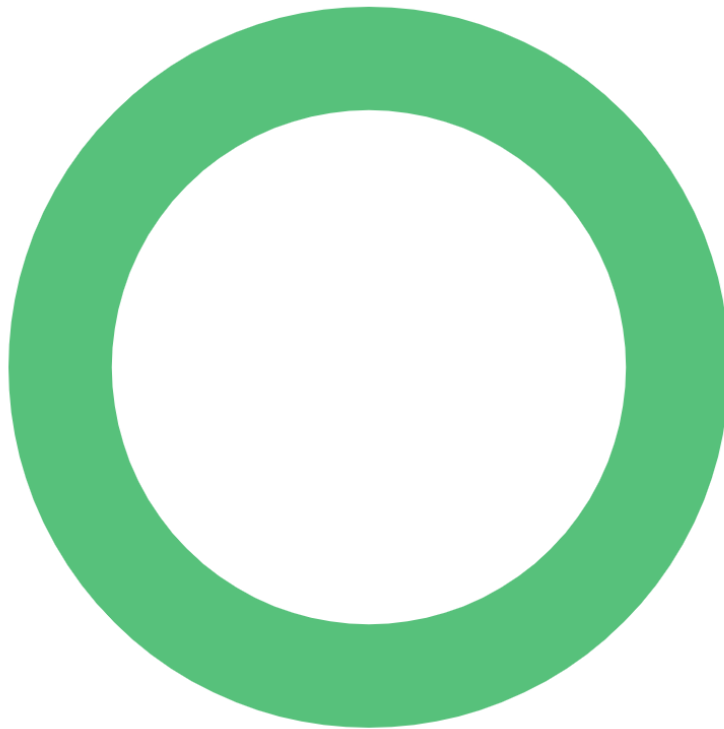
6-41. Select [sentiment.keyword] under [Filed] field.



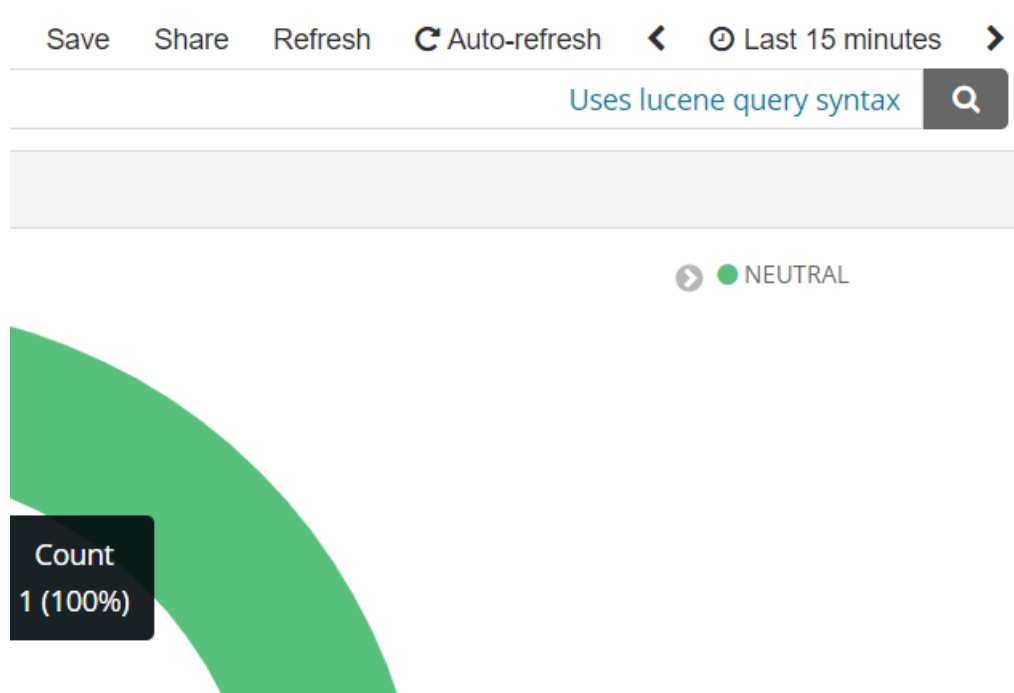
6-42. Click [Apply changes]



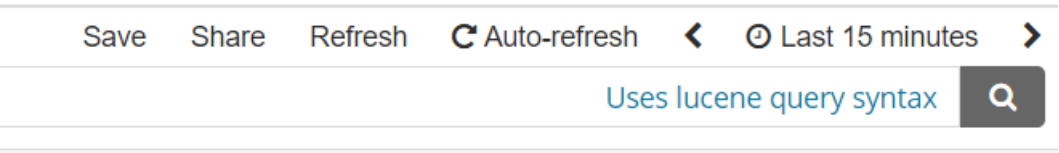
6-43. You will see result of sentiment analysis (Positive,Neutral,Negative) If you run python script with multiple Polly mp3 files, you will see variety of result.



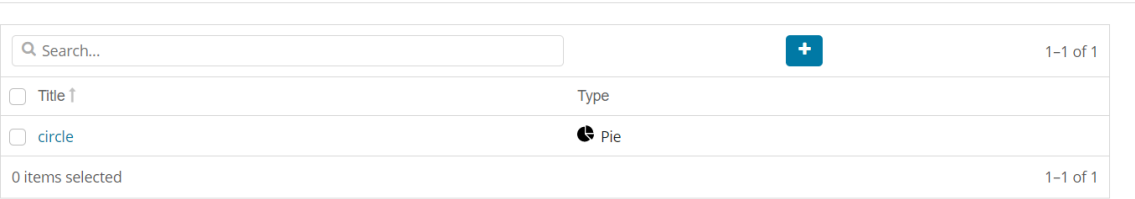
By default, Graph shows only latest 15mins. You can change the range to click [Last 15 minutes].



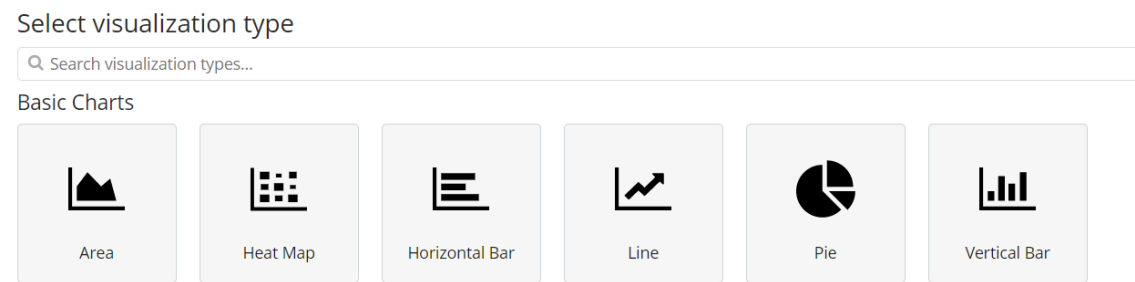
6-44. Click [Save], and type appropriate name.



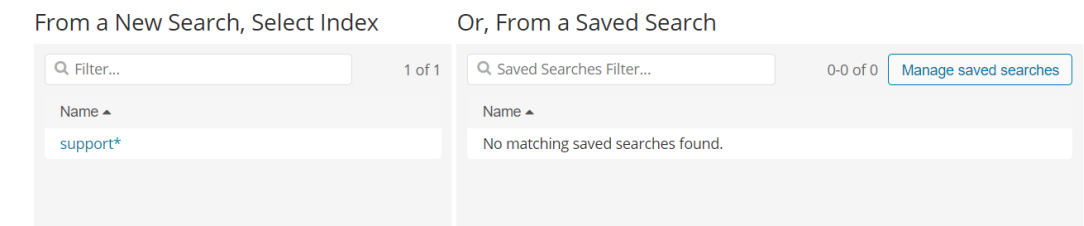
6-45. Click [Visualize] again



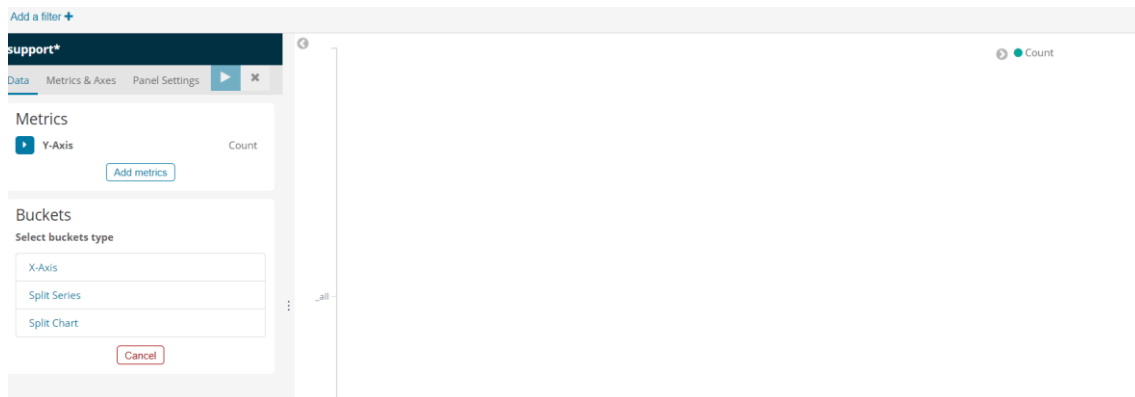
6-46. Press [+] button to create another graph.



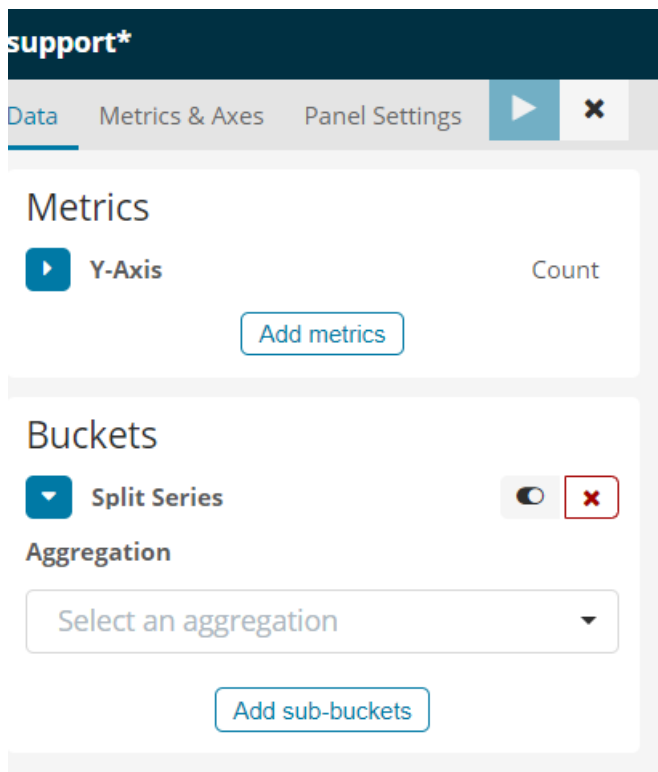
6-47. Select Horizontal Bar



6-48. Select [support\*\*]



6-49. Select [Split Series]



6-50. Set up as below screen, and click [Apply changes]

### Buckets

▼ Split Series

☾ ✕

**Aggregation**

Terms ▼

**Field**

keywords.keyword ▼

**Order By**

metric: Count ▼

Order

Size

Descend ▼

20

6-51. You will see list of keywords analyzed by Comprehend. As like Pie chart, you can

try multiple execution, then you will see more complicated graph.

