**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

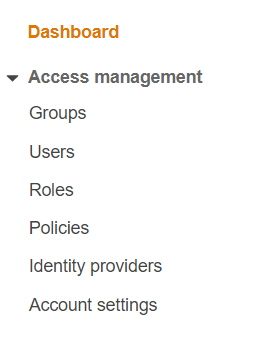
・Amazon Comprehend：obtaining insights from given text.

**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.

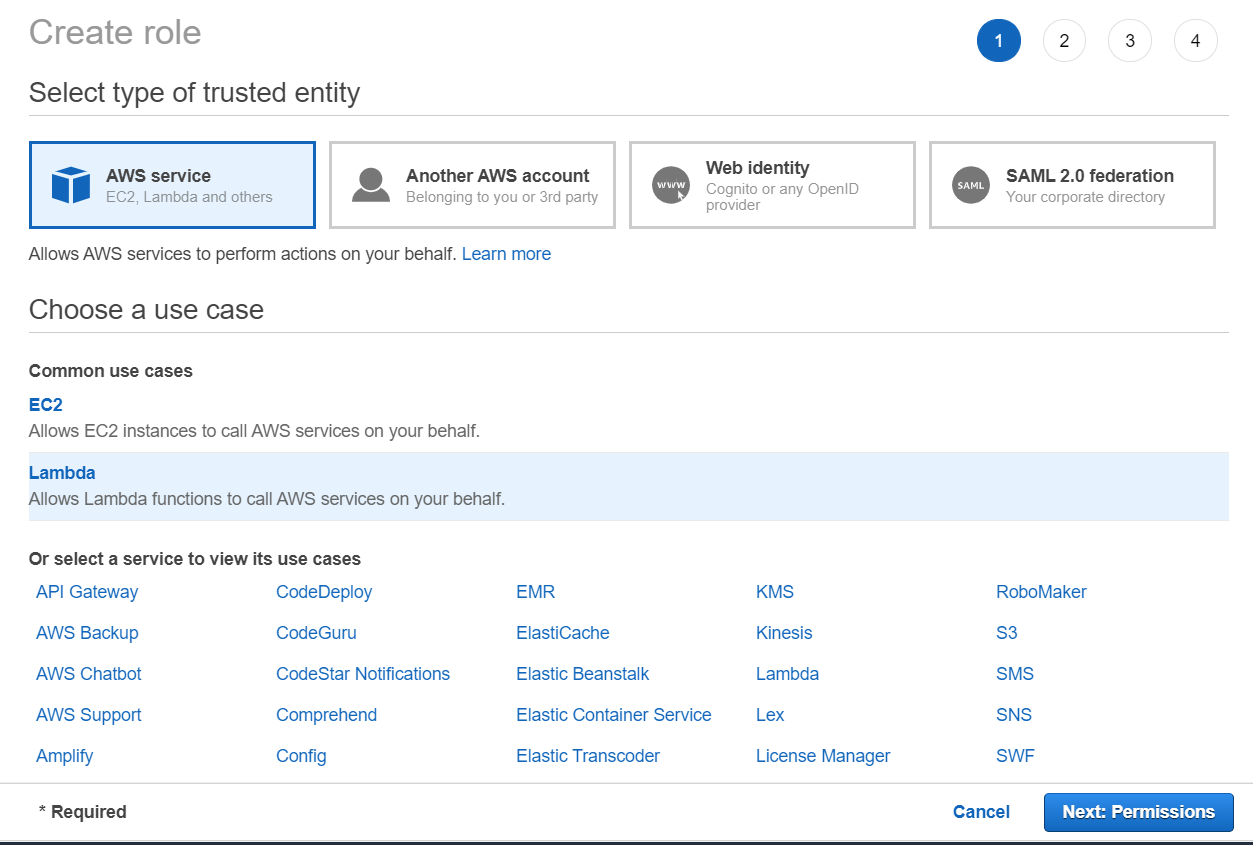
* 1. Access top page of IAM, and click [Roles]



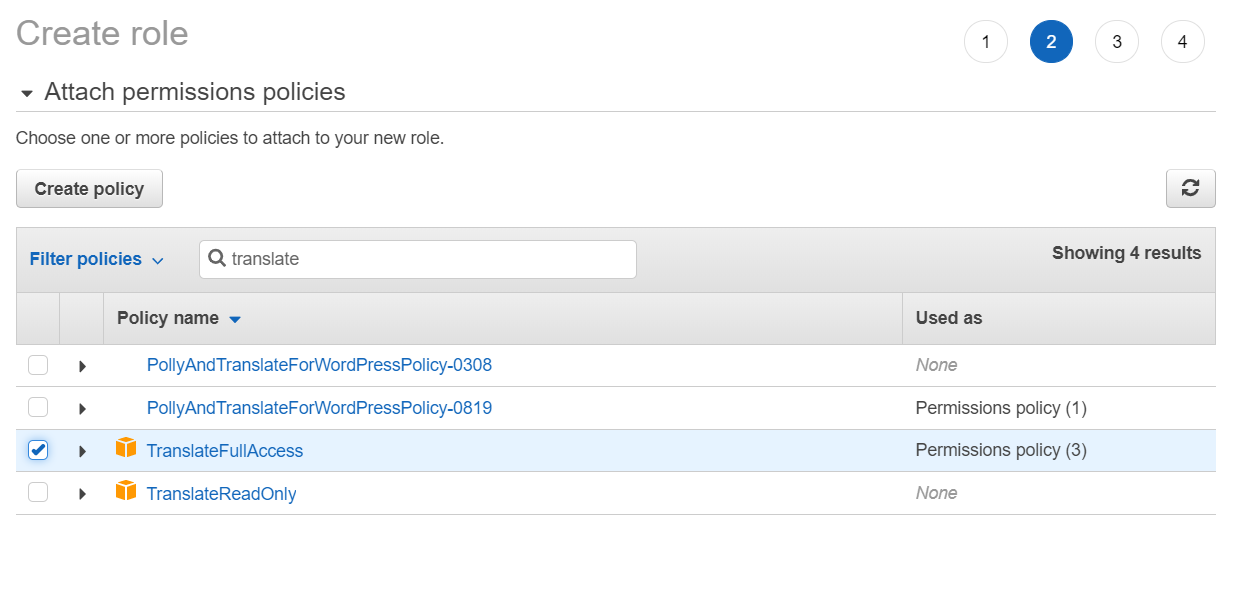
* 1. Click [Create role]



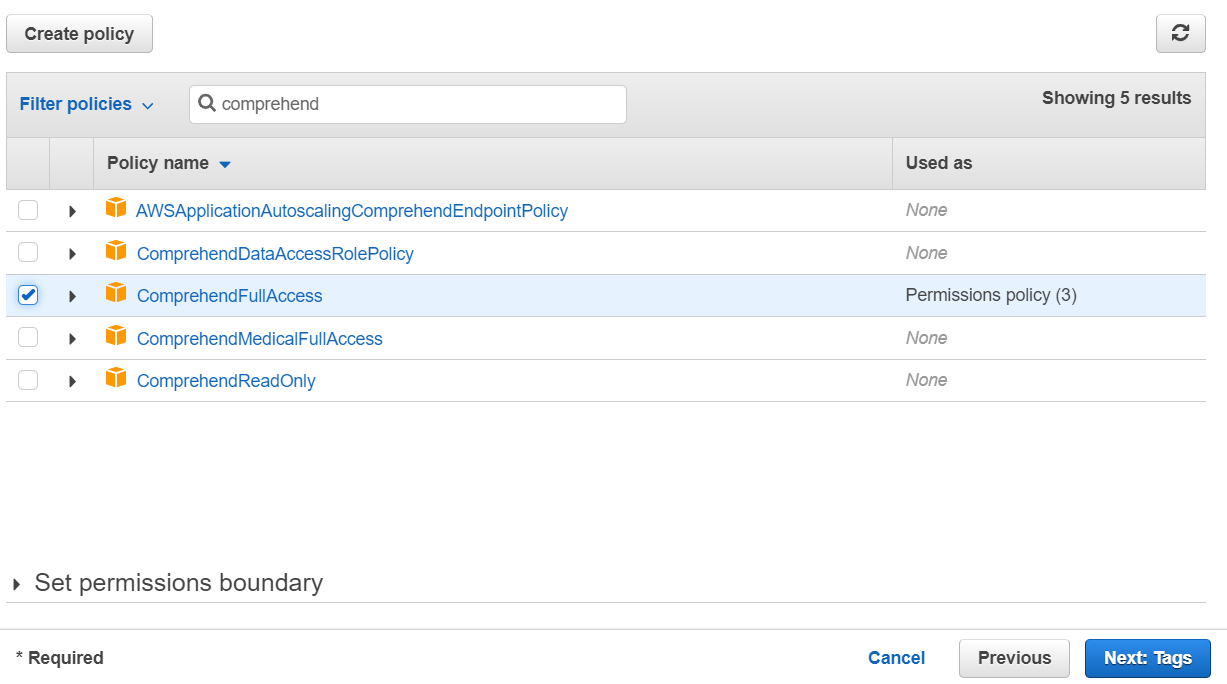
* 1. Select [Lambda], and click[Next: Permissions]



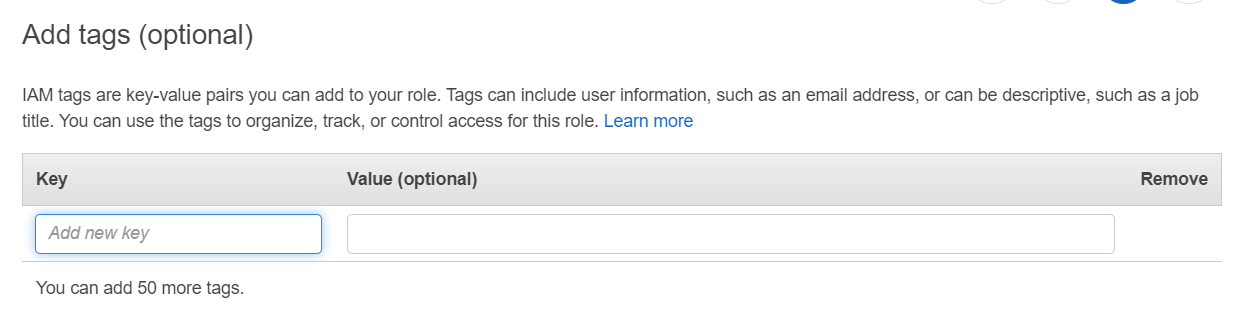
1-4.Search [translate], and check [TranslateFullAccess] **Do not click button at lower right corner, yet.**



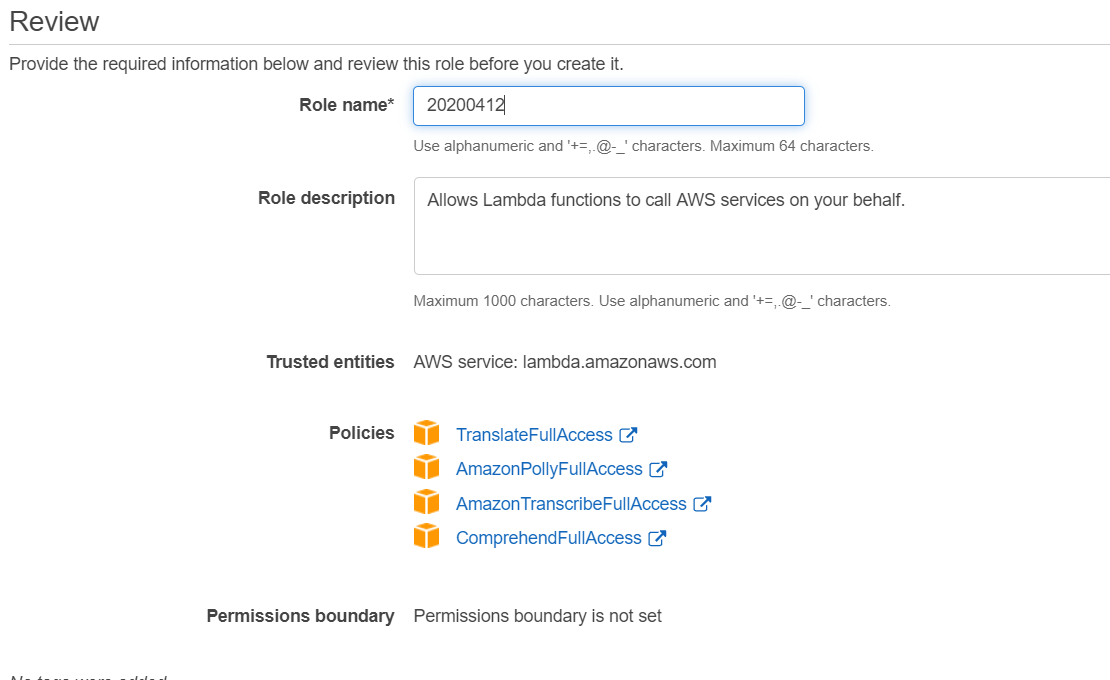
* 1. As like, translate, check [polly], [transcribe], and [comprehend] to setting up full access. After that, click [Next: Tags]



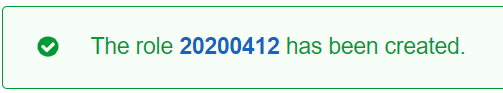
* 1. Without input anything, just click [Next: Review]



1-6.　Set up Role name as [YYYYMMDDnlphandson]（YYYYMMDD is date of today. Ex. 20200412）



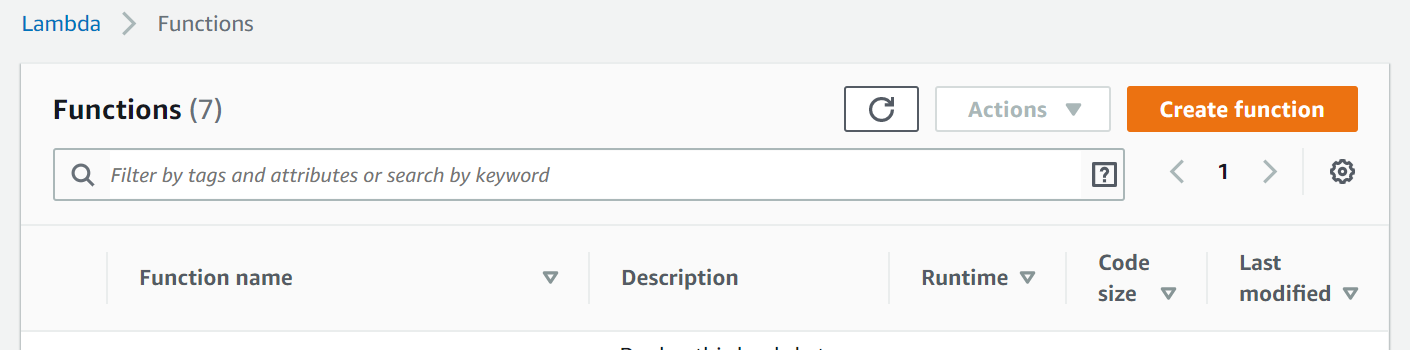
1-7.　Double check, the role has full access to 4 services, and click [Create role]. If you see the below dialog, role is successfully created.



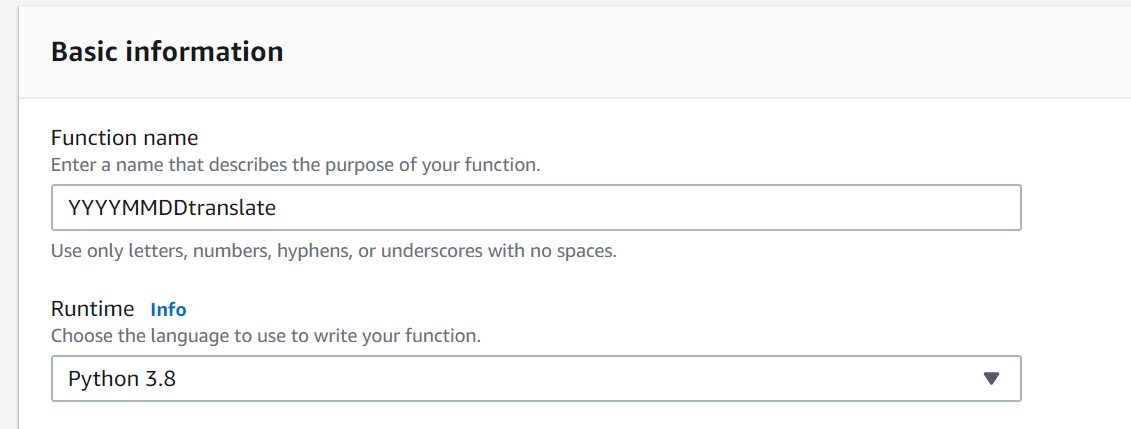
1. Amazon Translate

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

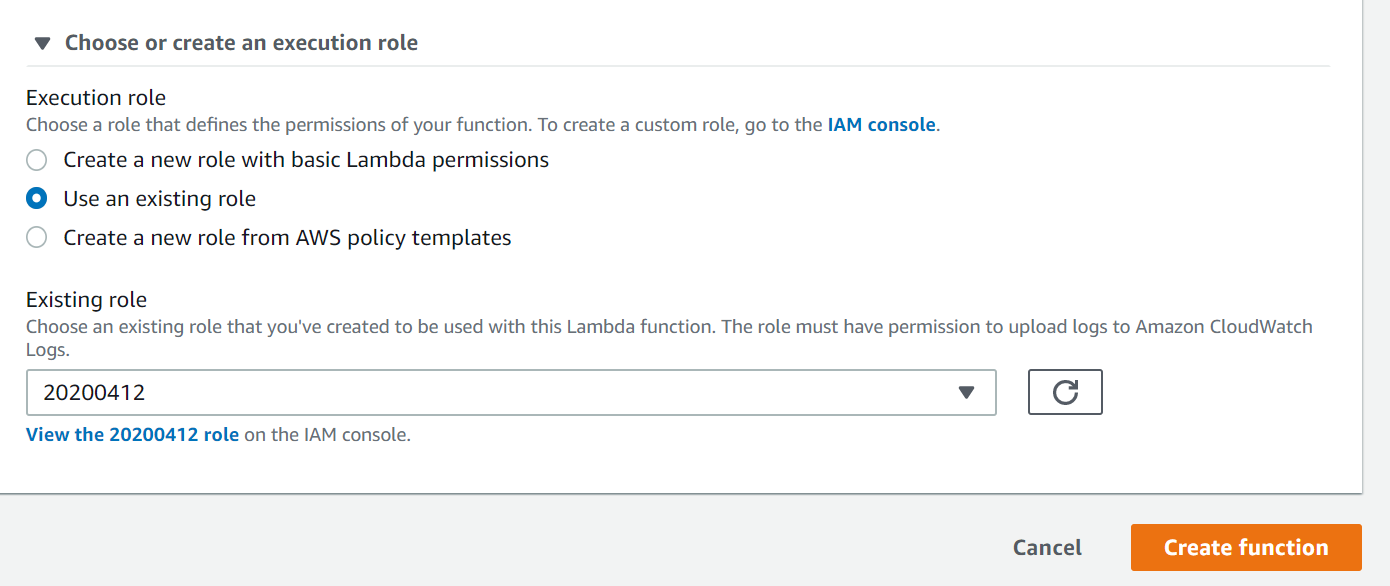
2-2. Access AWS Lambda Top page, and click [Create function]



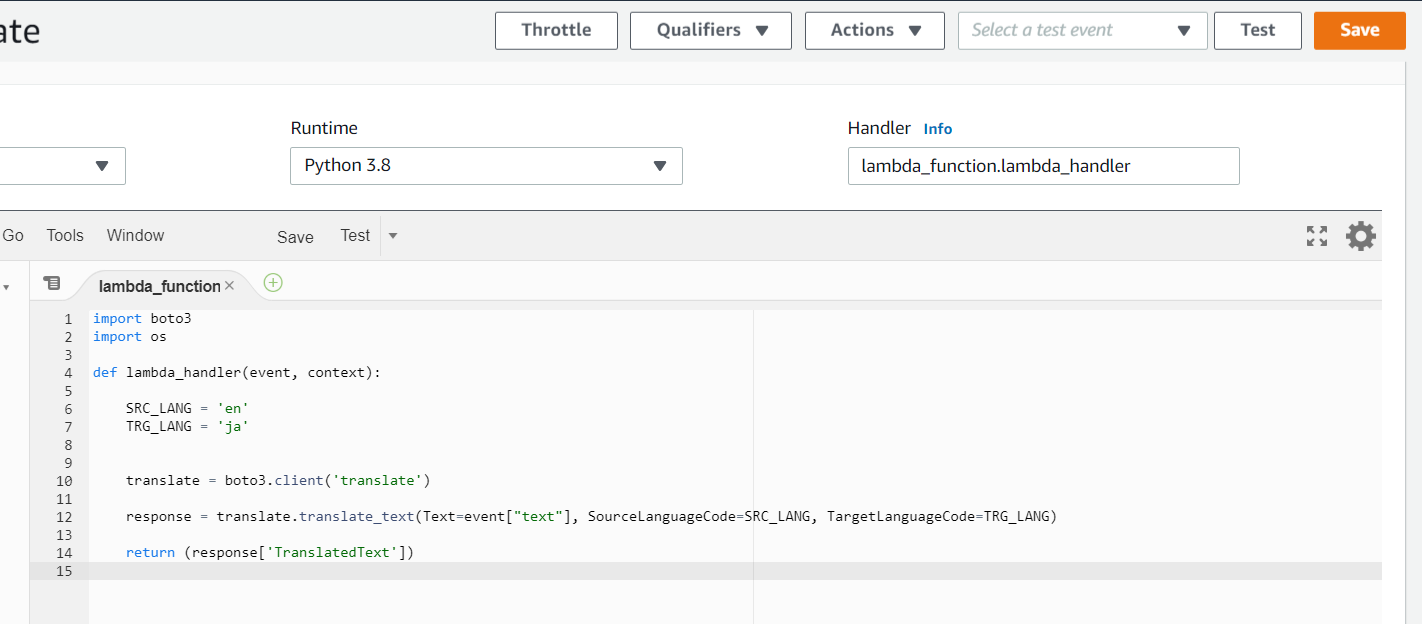
2-3.Input [YYYYMMDDtranslate](YYYYMMDD is date of today) to Function name, and select [Python3.8] as your runtime.



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]



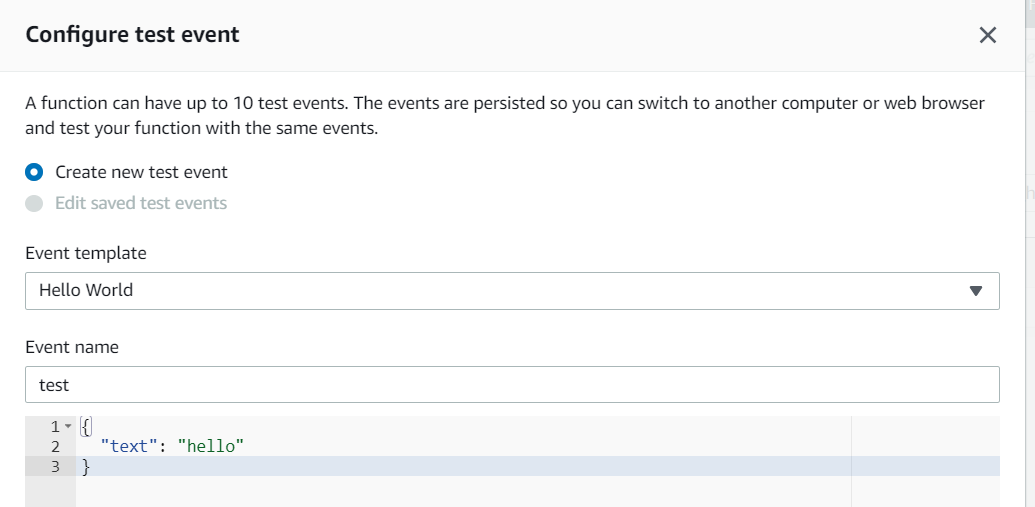
2-5. Delete sample code at first, and copy and paste [translatelambda.txt] into the editor. Then click [Save] button.



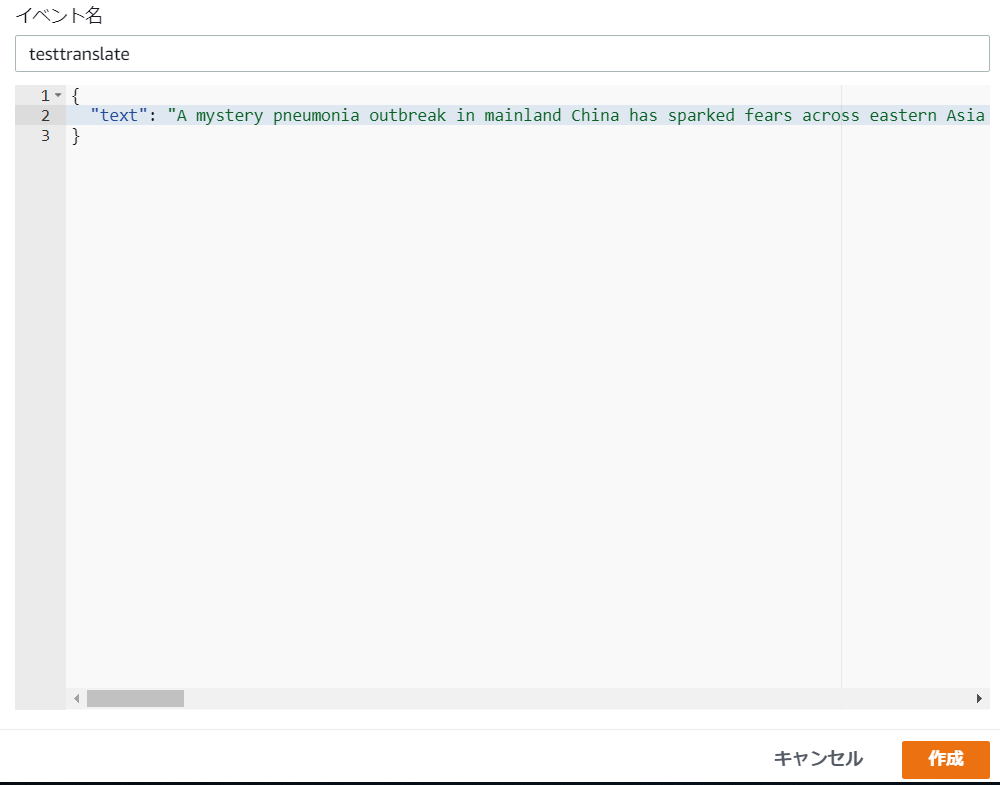
2-6.　Click [Test] button.



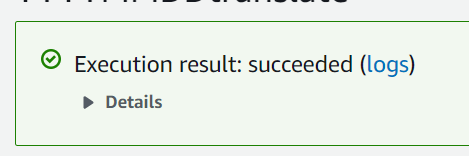
2-7．Type [test] or appropriate name to Event name, and copy and paste with [translatelambdatest.txt].



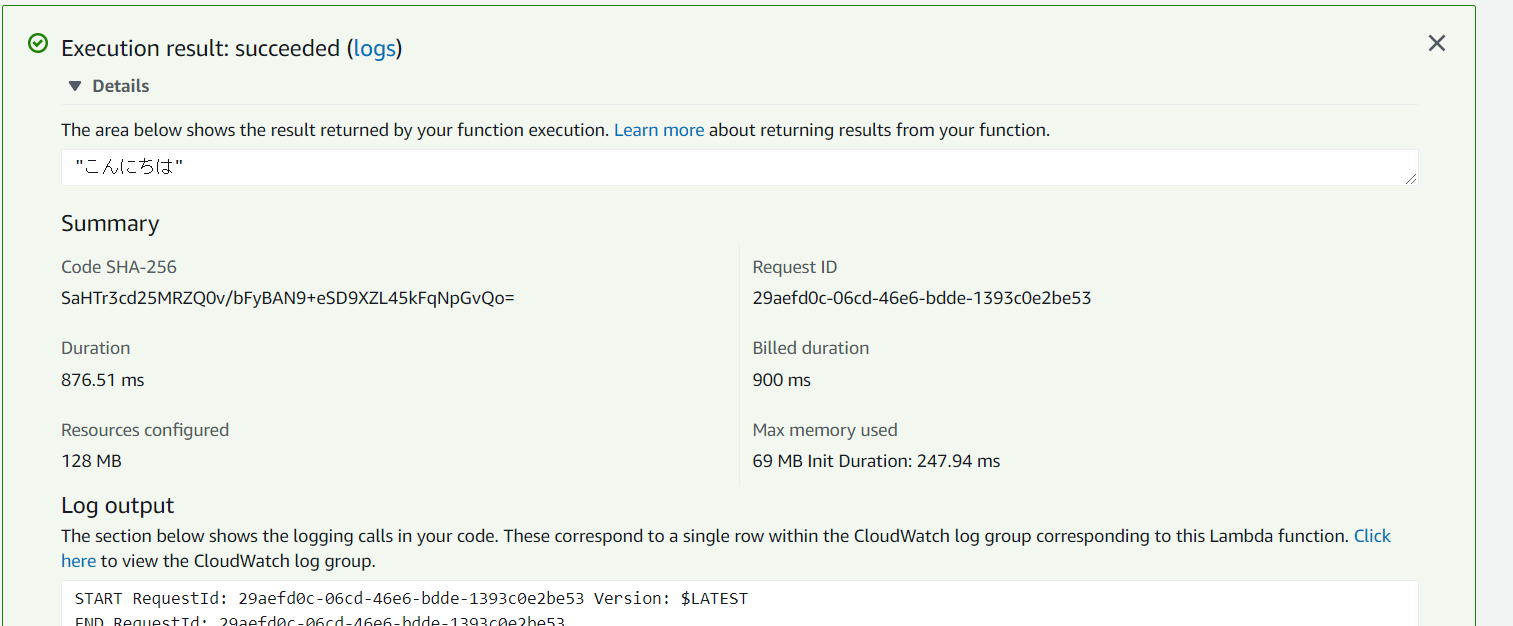
2-8.　Replace [hello] to your previously copied English news. At the time, you have to wave every line feed code. In other word, the line should be one line. Then, click [Create]



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



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

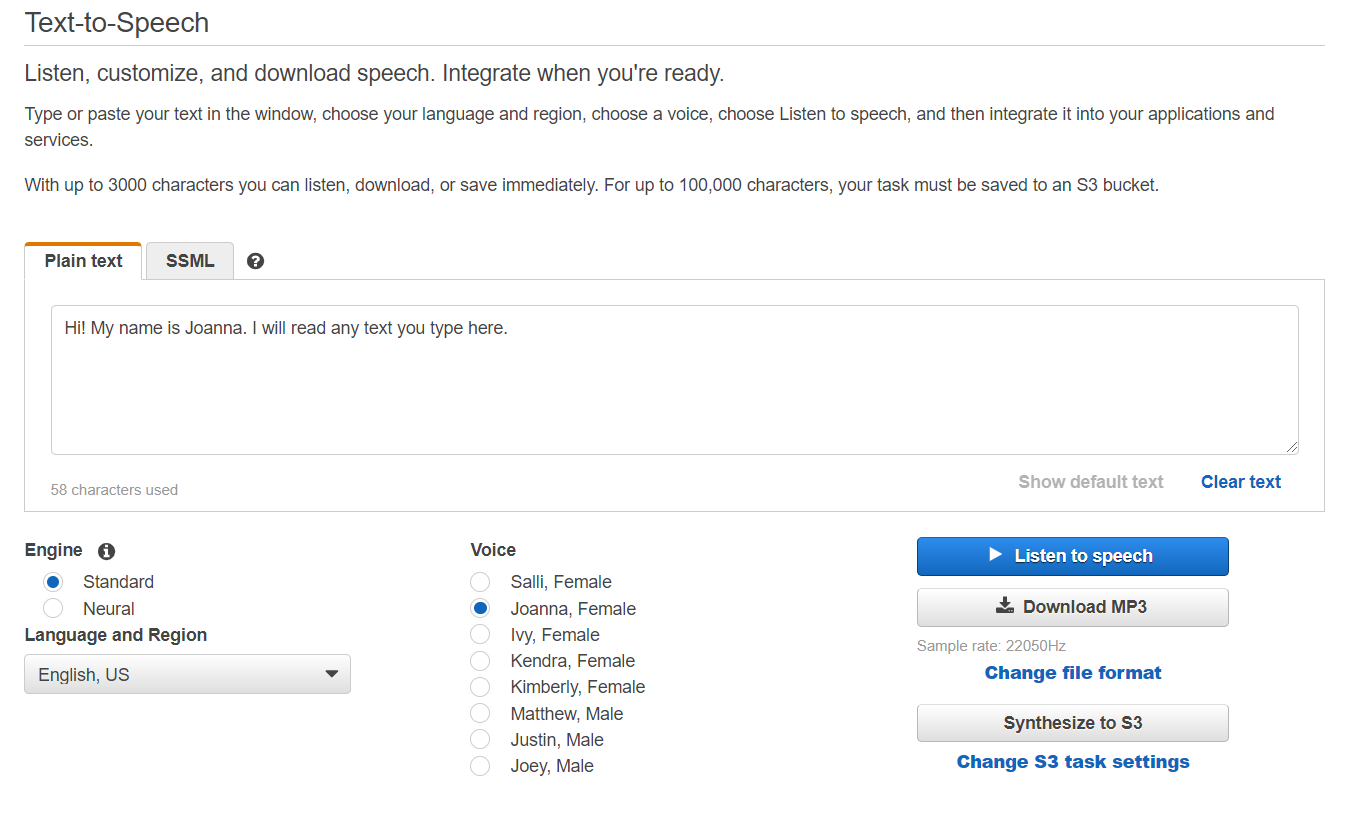


This sample code is translate English to Japanese. You need to change [TRG\_LANG = 'ja'] to your desired language. You can check if your native language is supported by translate. <https://docs.aws.amazon.com/translate/latest/dg/what-is.html#what-is-languages>

1. 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



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.

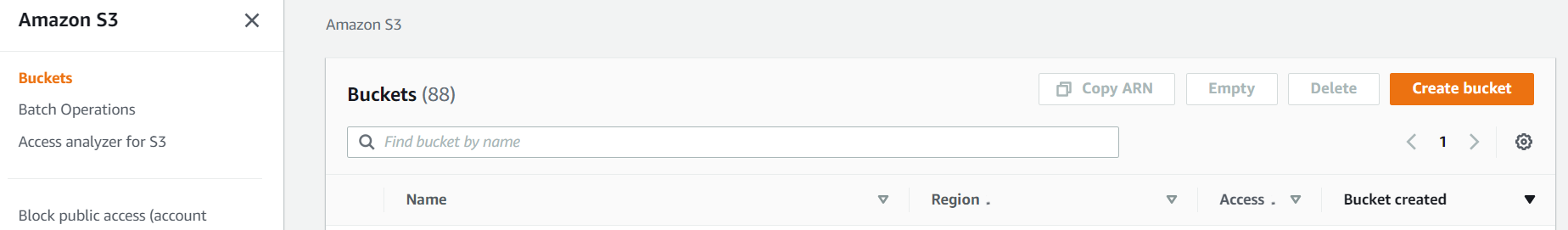
If you click [Language and Region], you can check if your native language is supported or not. You can also obtain full list from below URL. **CAUTION: Region does not mean AWS Region. It is dialects in such country..**

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

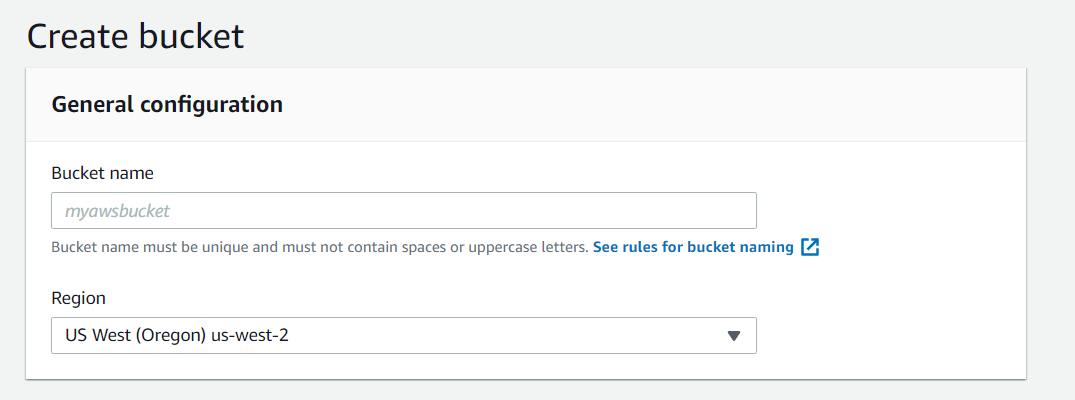
1. Transcribe

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

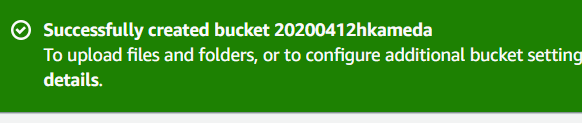
4-1.　Access Amazon S3 management console. Be aware that Region should be Oregon.



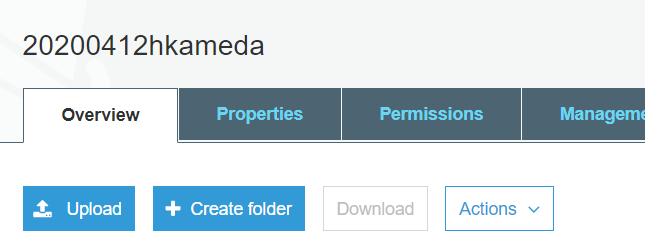
4-2.　Click [Create bucket], and input YYYYMMDDName（YYYYMMDD is date of today） to [Bucket name].



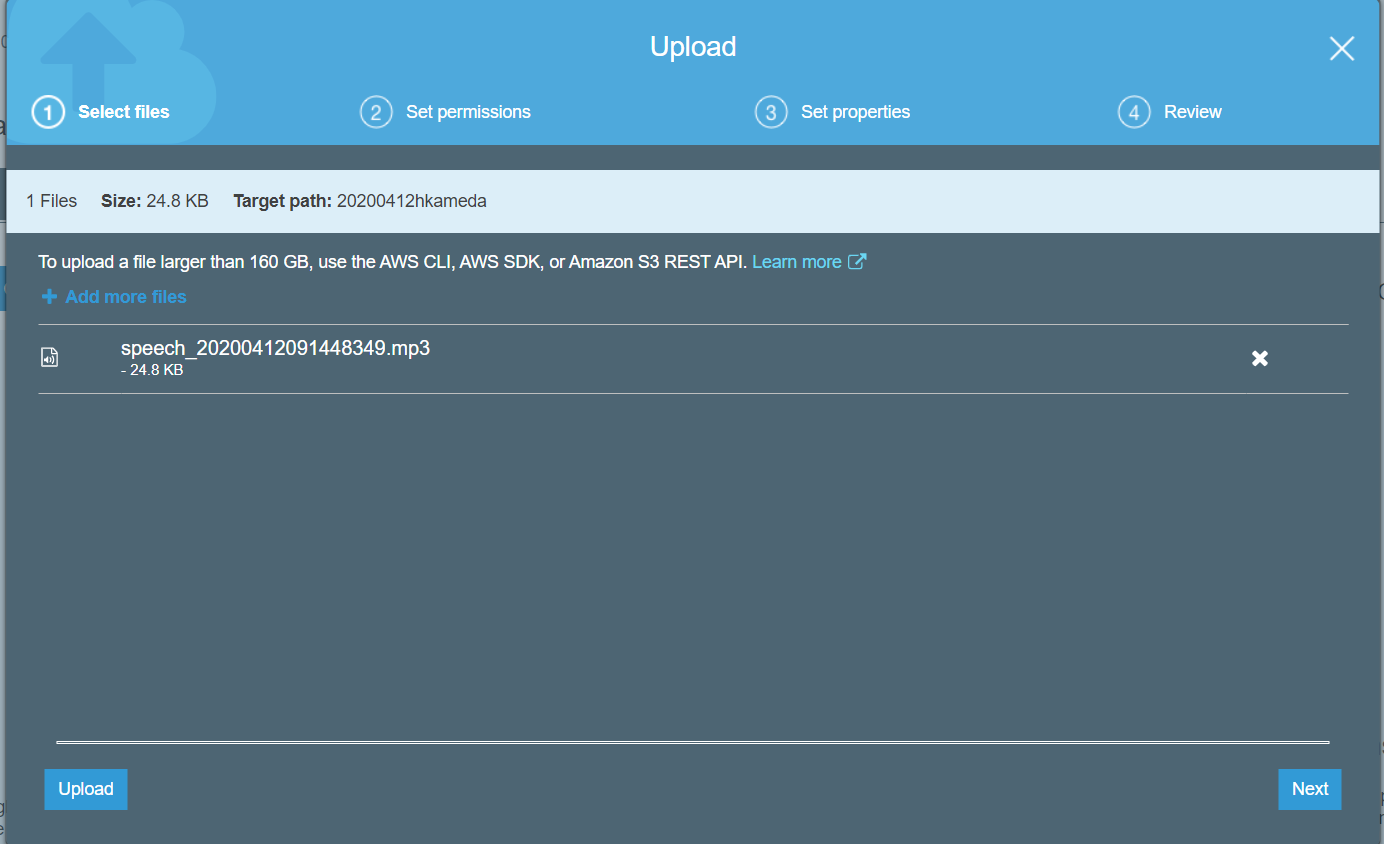
4-3. Click [Create bucket]



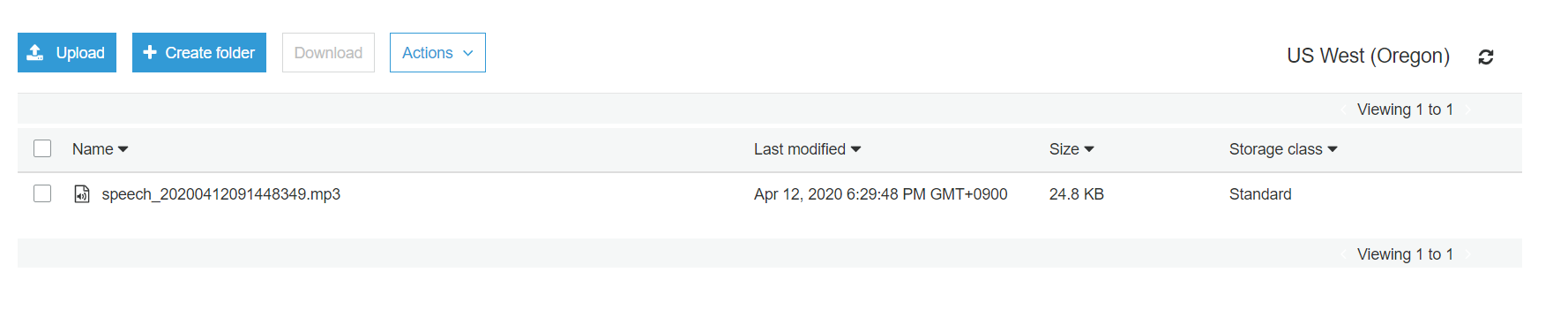
4-4.　Click the bucket, and Click also [Upload]



4-5.　Upload mp3 file created by Polly.



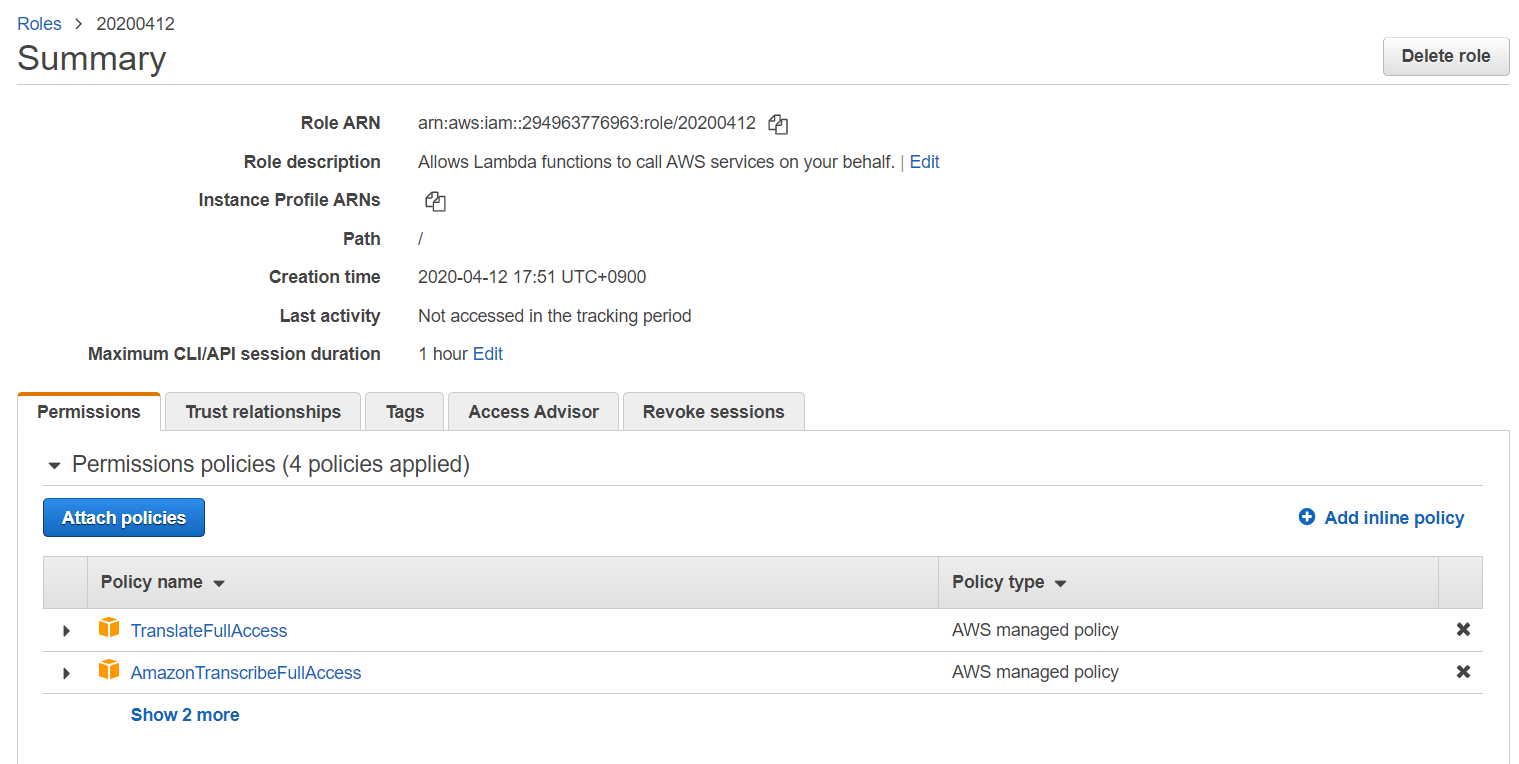
4-6.　Click [Next] 3 times, and click [Upload] for final step.



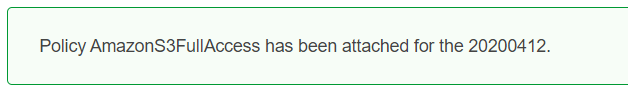
4-6-1．The IAM role you have created at previous steps does not have access S3.You need to grant access to S3 onto the IAM role. Going back to IAM role management console, identify IAM role.



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

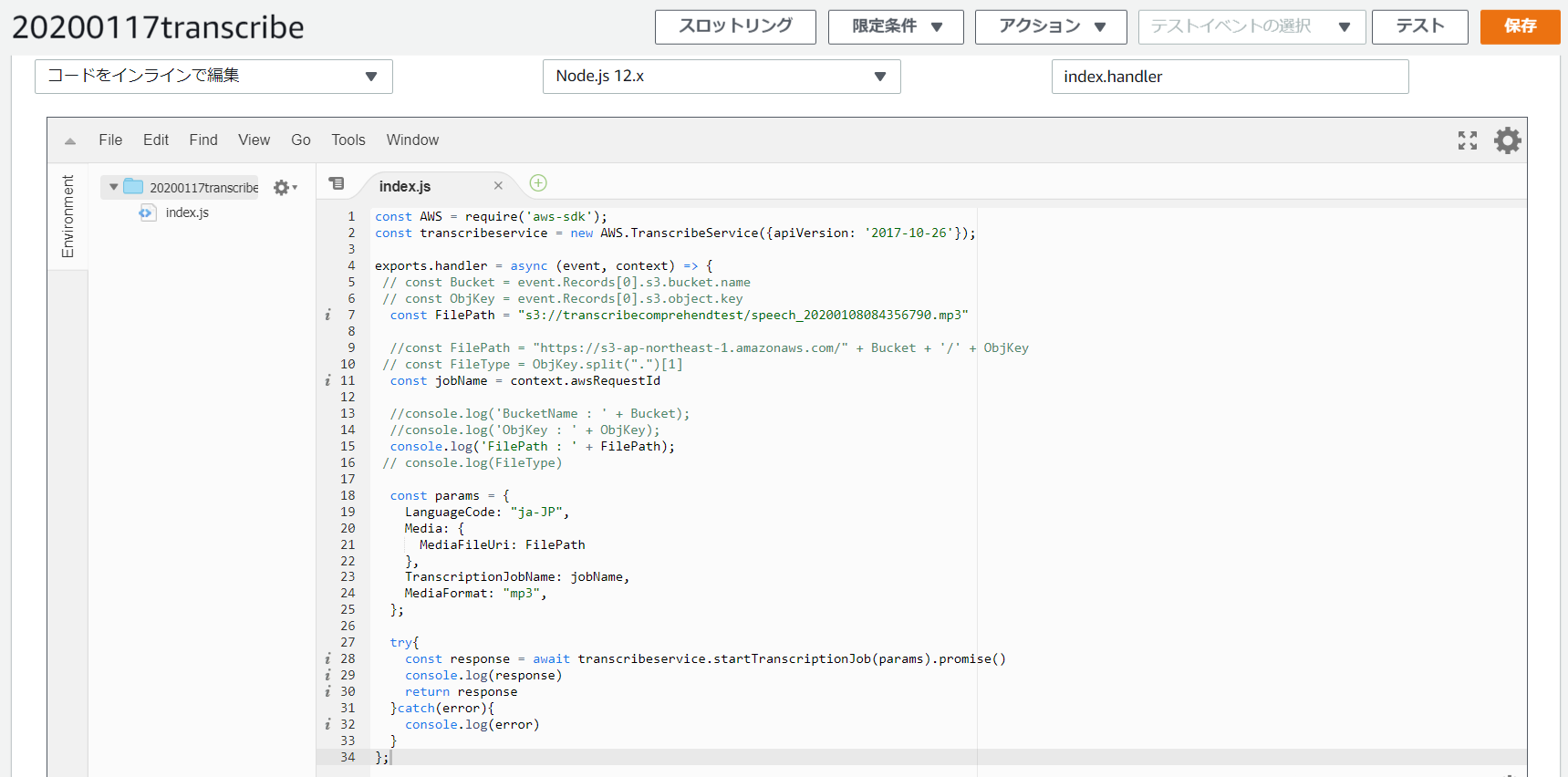


4-6-3.　Find and check [AmazonS3FullAccess], and click [Attach policy]



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.

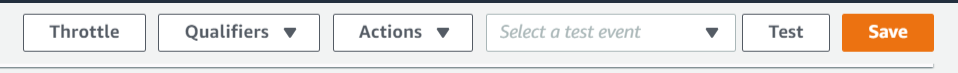
4-8.　Copy and paste [transcribelambda.txt]



4-9.　 Replace [const FilePath] to your S3 bucked name and file name of mp3 file. The pass should be like below;

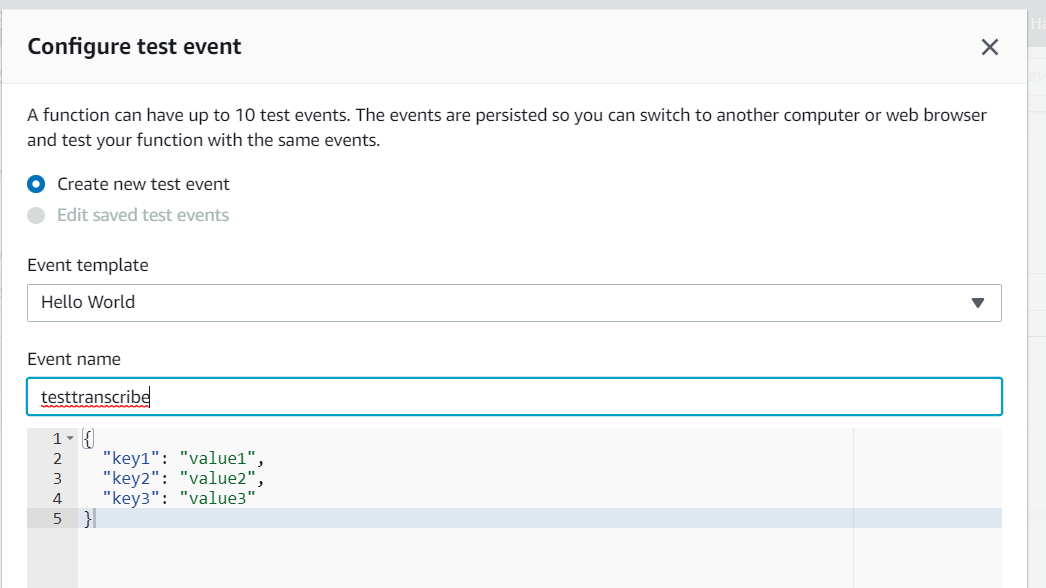
s3://20200117kameda/speech\_20200117061325085.mp3

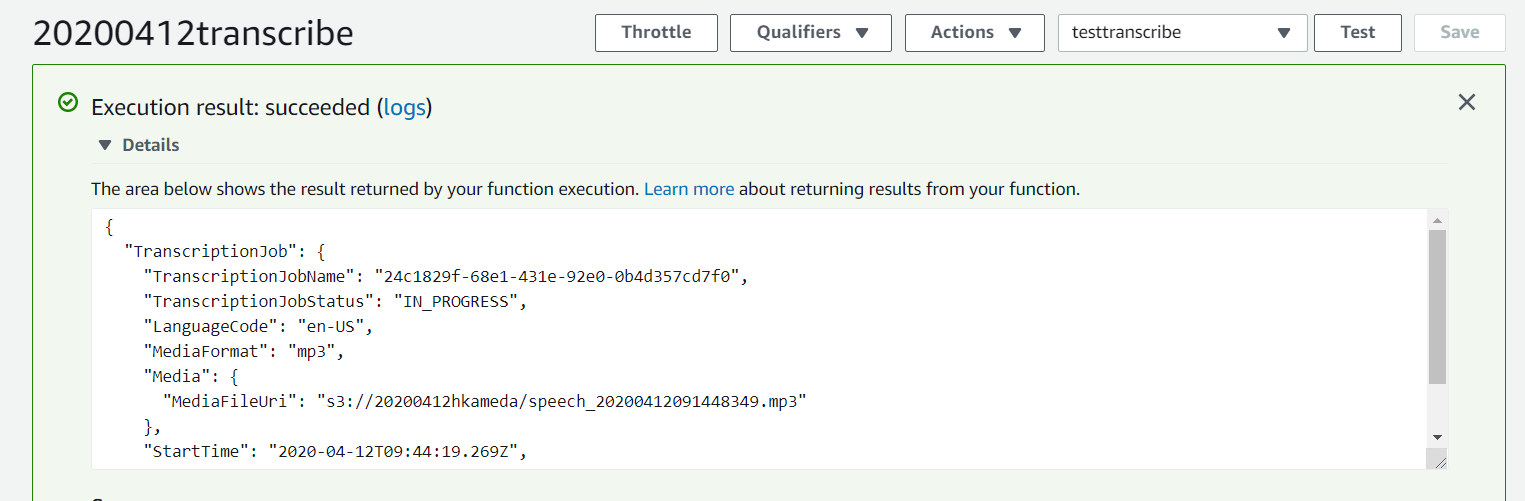
4-10. Click [Save]



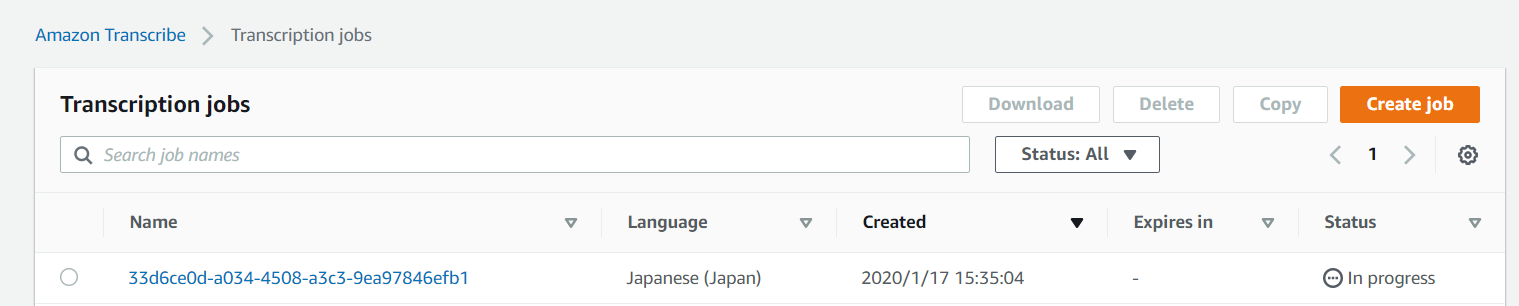
4-11.　Click [Test]. Input appropriate name to [Event name], and click [Create].

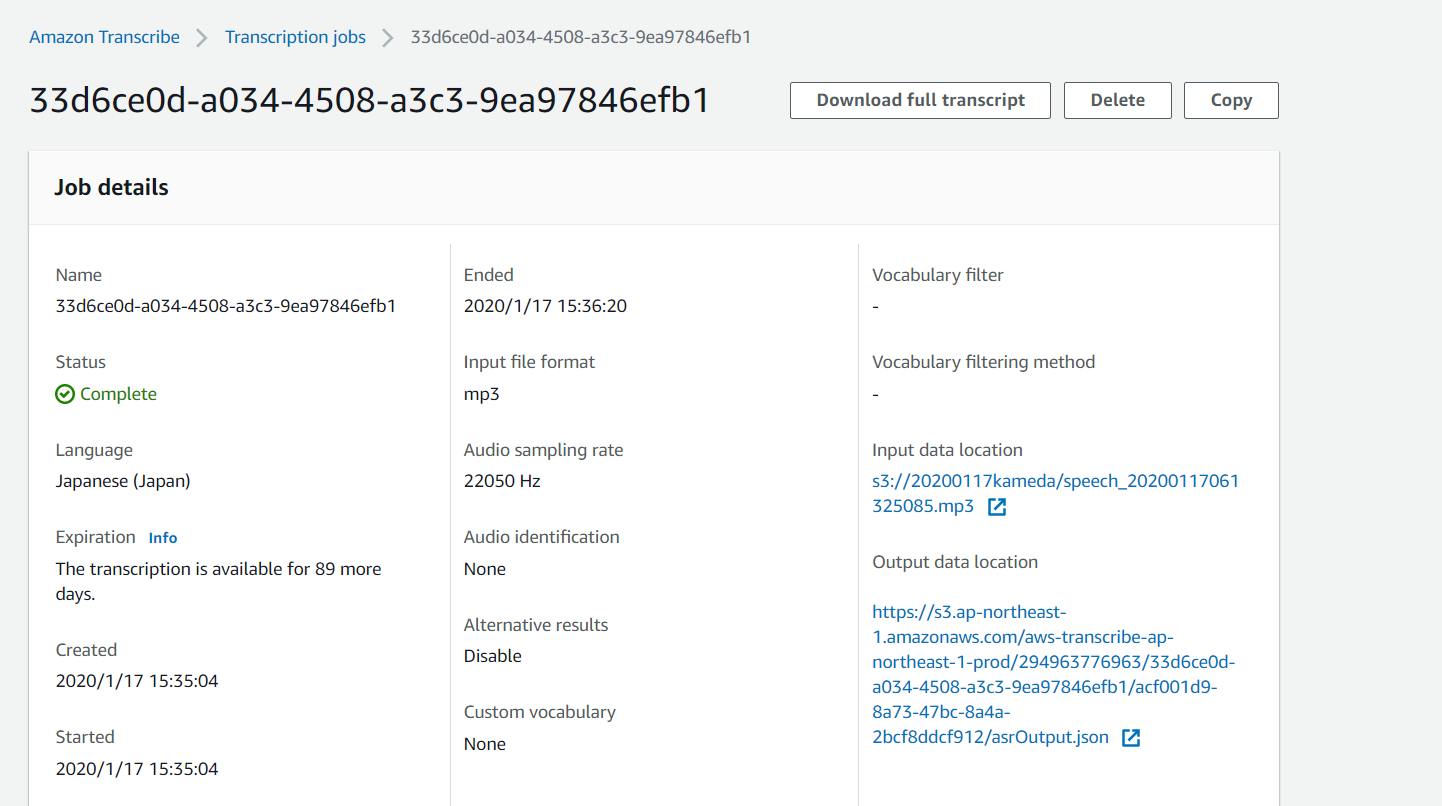
#This function does not require any test input from outside of function, but you need to do this anyway.



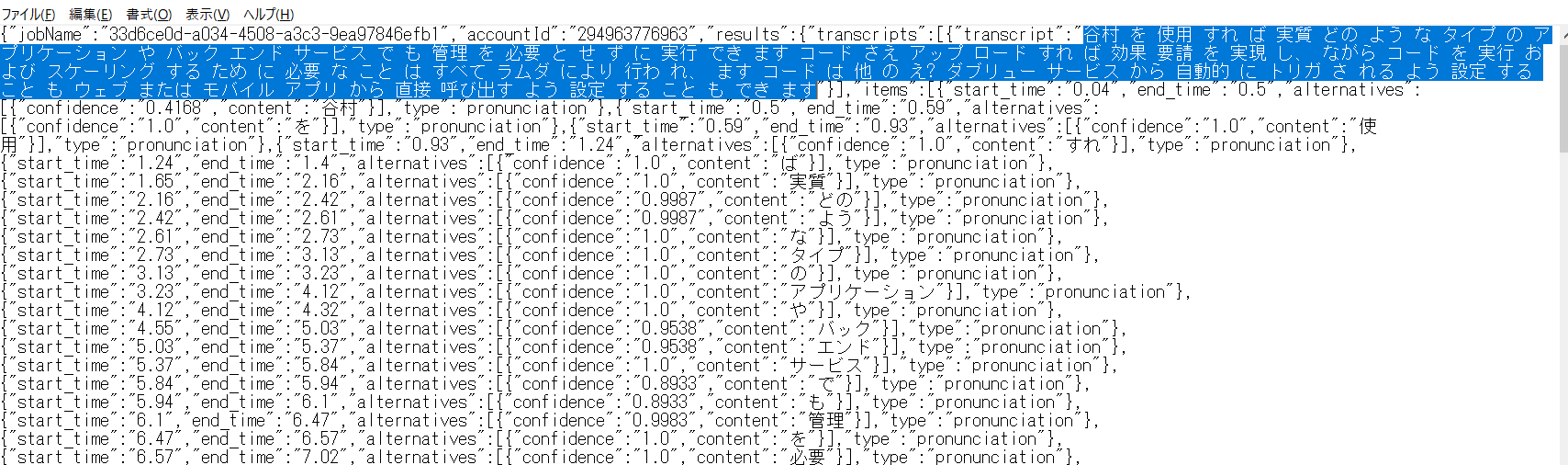
4-12. Click [Test] again, and you will see [IN\_PROGRESS] in the case of successfully Transcribe job has been is started. 

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.



4-14.　Click Name of the job after completed.

4-15.　Clock [Download full transcript]. Then, JSON file is downloaded. Open it with your text editor, and you will see transcribed sentence.



You can obtain full list of supported language from below URL;

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

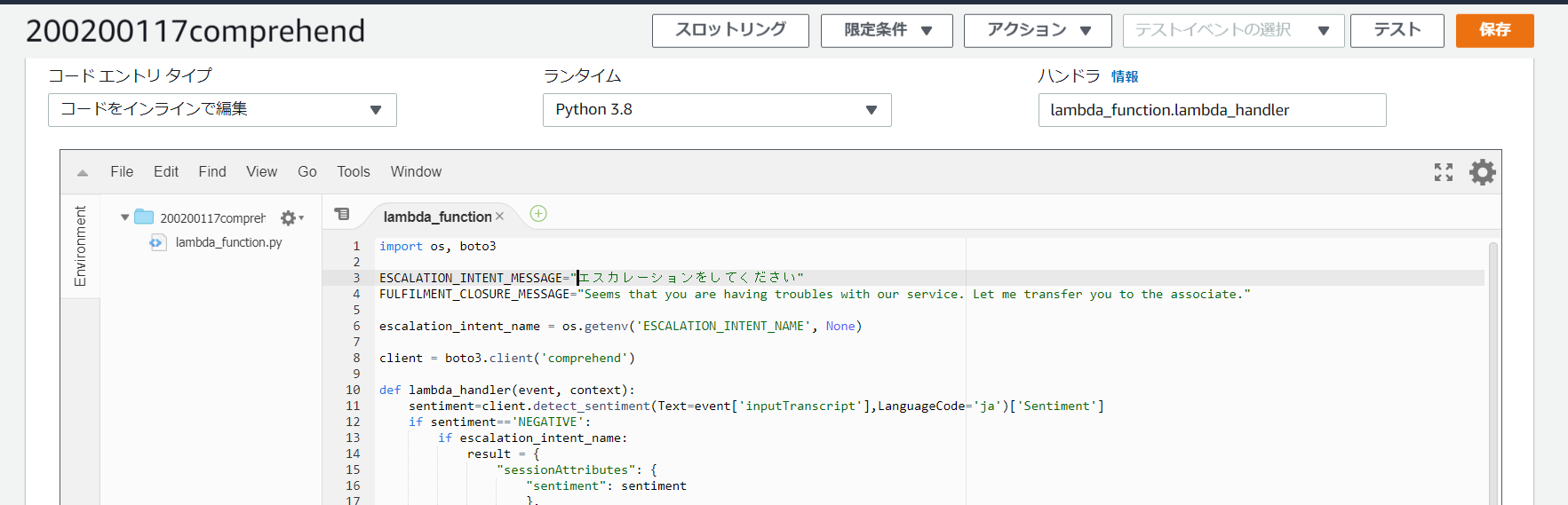
If you native language is supported, you can replace script [LanguageCode: "en-US", ] to your desired language.

1. Comprehend

Copy and paste appropriate English sentence for test purpose.

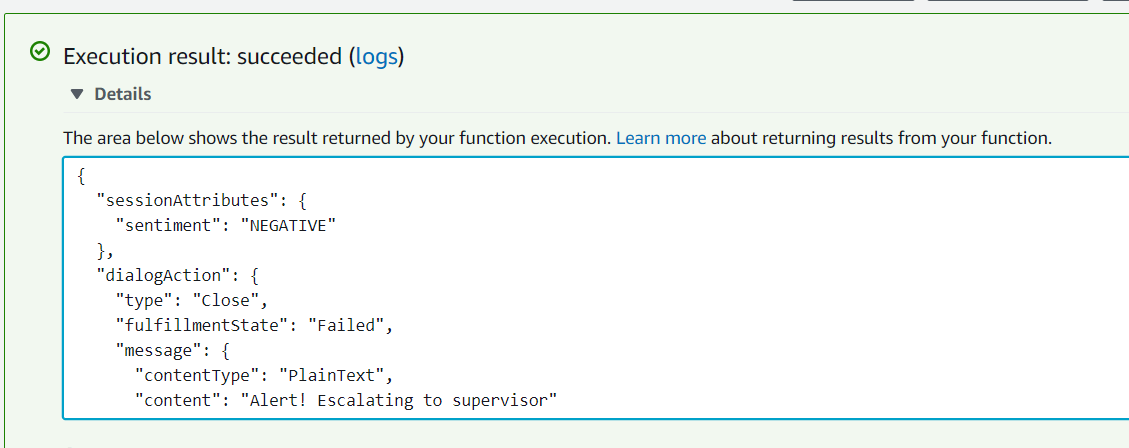
5-1.　Access AWS Lambda management console again, and create function with name of [YYYYMMDDComprehend]. [Runtime] should by Python3.8 this time. Please be sure to assign IAM role to your function.

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!”

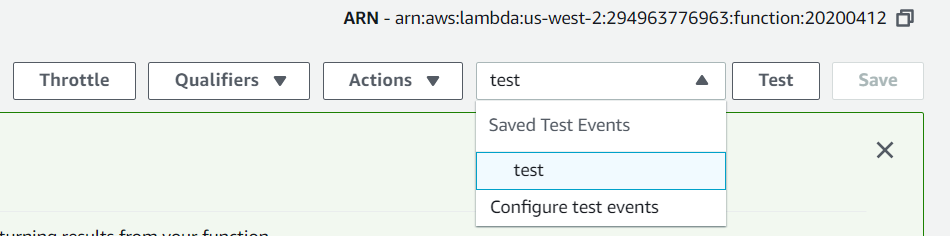


5-3.　Click [Save], and Click again [Test]. Input appropriate name to [Event name] and copy and paste [comprehendlambdatest.txt]. Click [Create].

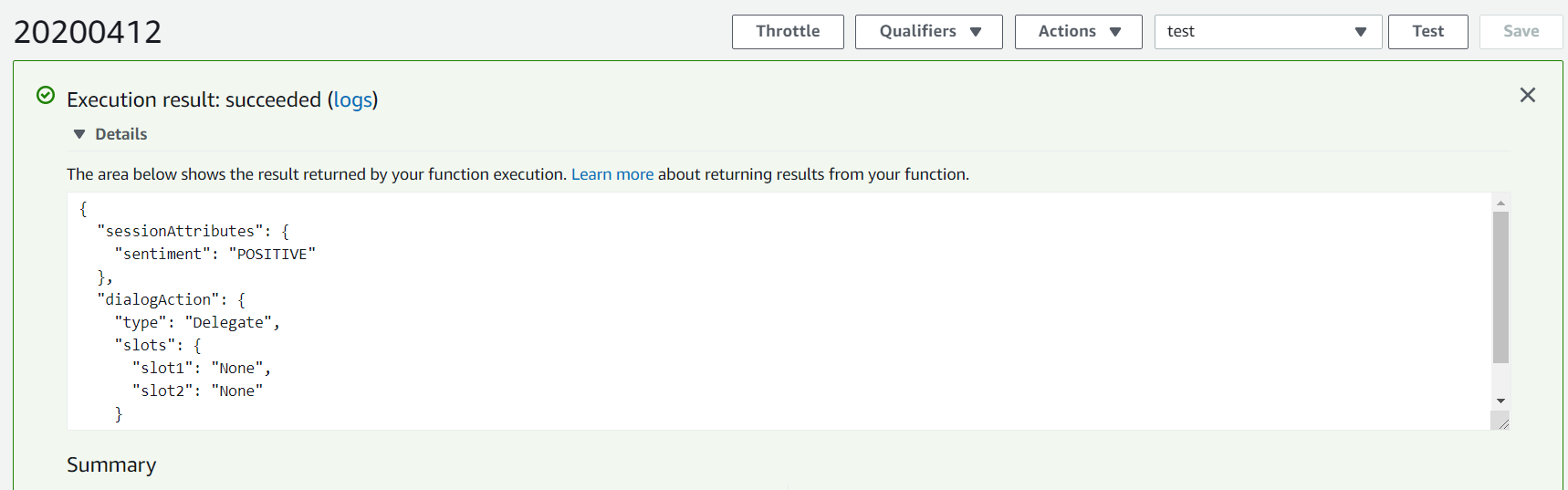
5-4．Test input contains sentence of “I hate you”. If you run the test, you are going to see “Alert”



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

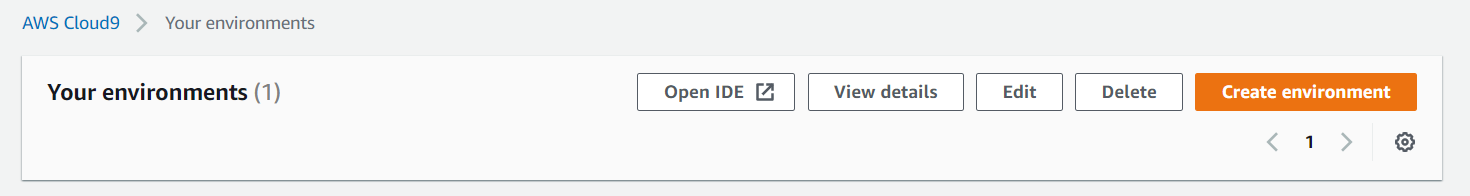


5-6.　You may see “POSITIVE” this time.

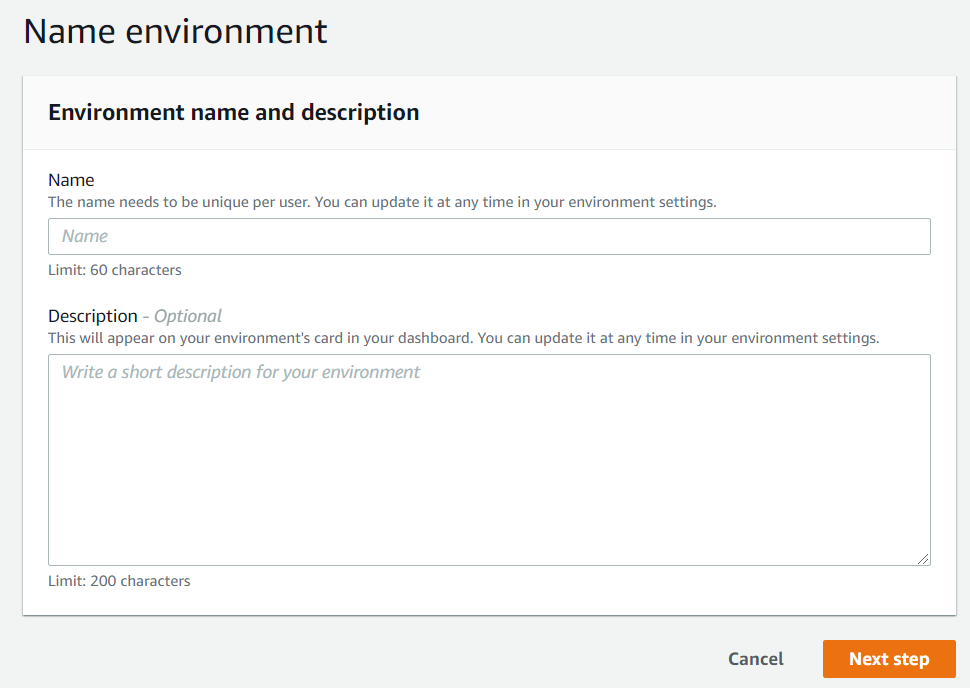


6.　This is final, but very long step here. We Combine Transcribe and Comprehend via Python script, and create dashboard and visualize result of analysis on Amazon Elasticsearch service. **Cloud9 is used for this scenario. Cloud9 takes over IAM righs to access any AWS services from your IAM user you are logging in AWS account by default. Please be sure that your IAM user has access to all Comprehend、Transcribe、Elasticsearch Service、S3、Cloud9.**

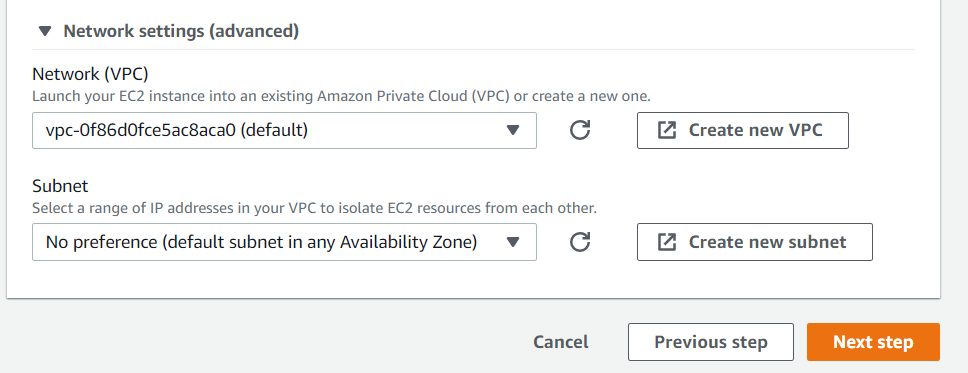
6-1.　Access AWS Cloud9 management console.



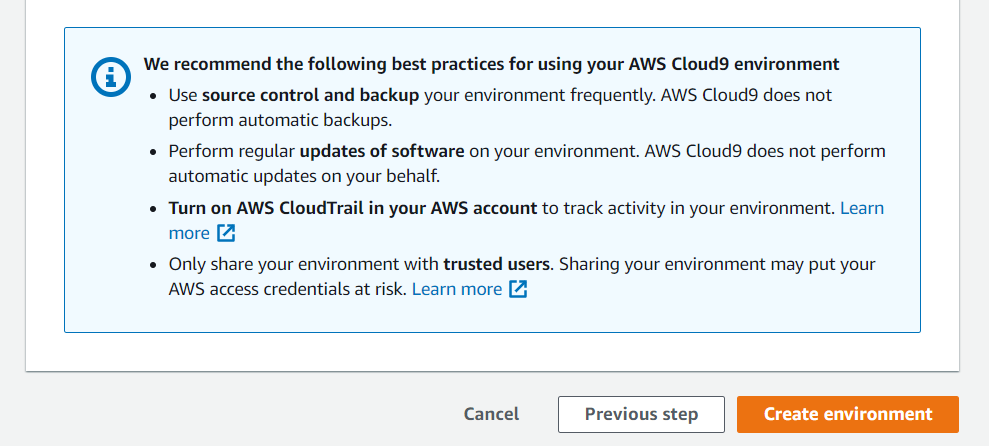
6-2.　Click [Create Environment]. Input YYYYMMDDhandson（YYYYMMDD is dateof today）to the Name, and Click [Next Step]



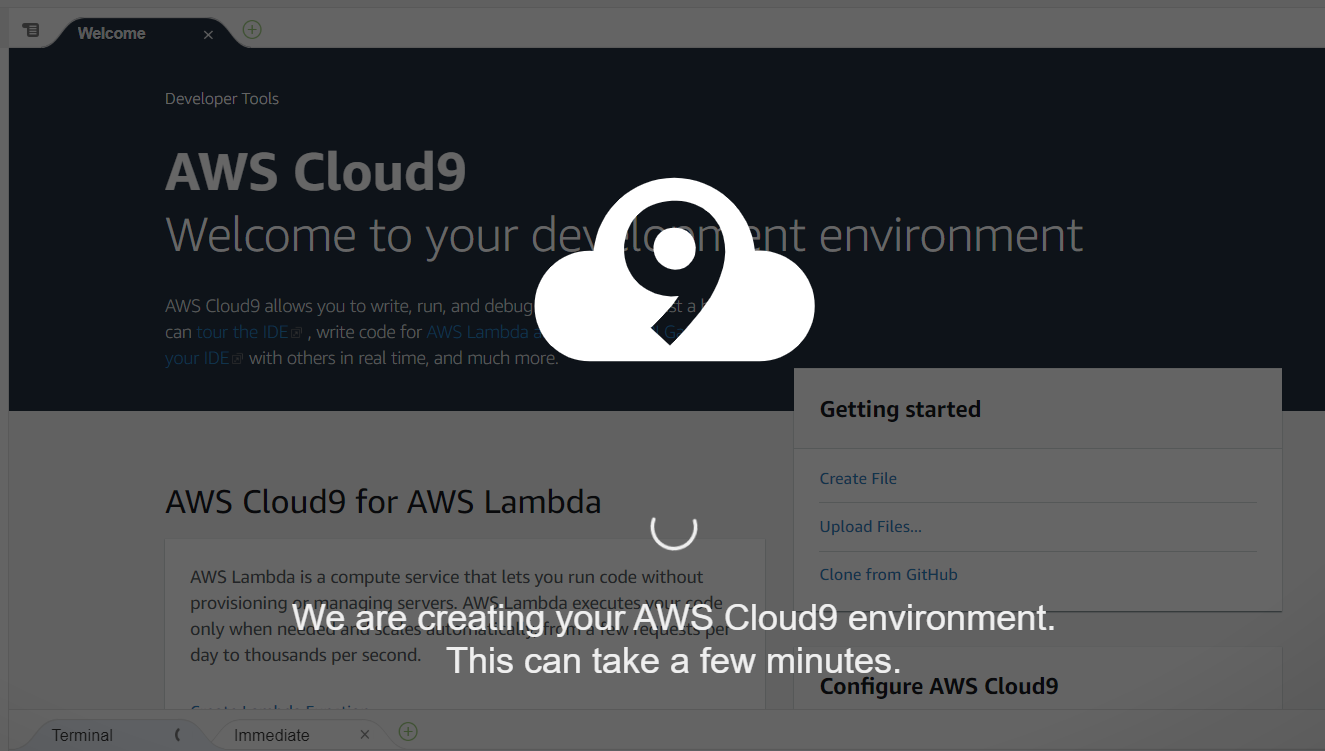
6-3.　Select [Network Settings], and Choose VPC which has Public subnet. (in the most case, default VPC works great)



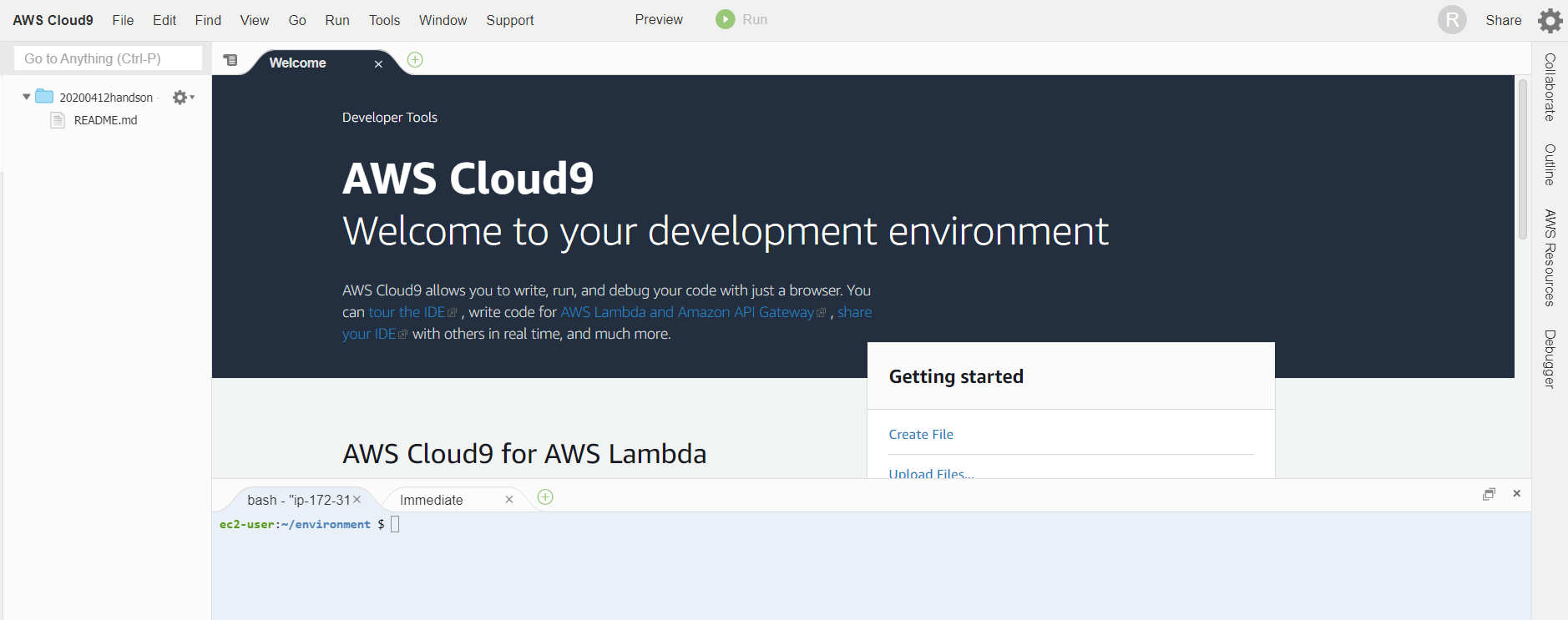
6-4.　Click [Next Step]



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-2*

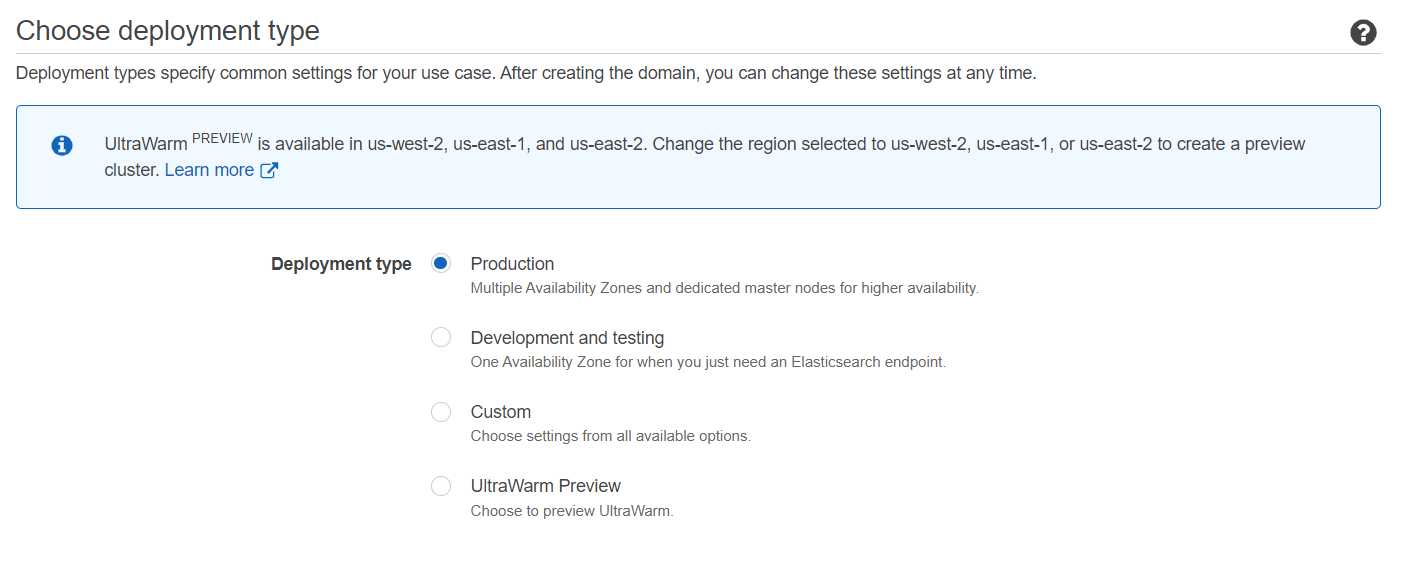
(YYYYMMDD is date of today, name is your nickname)



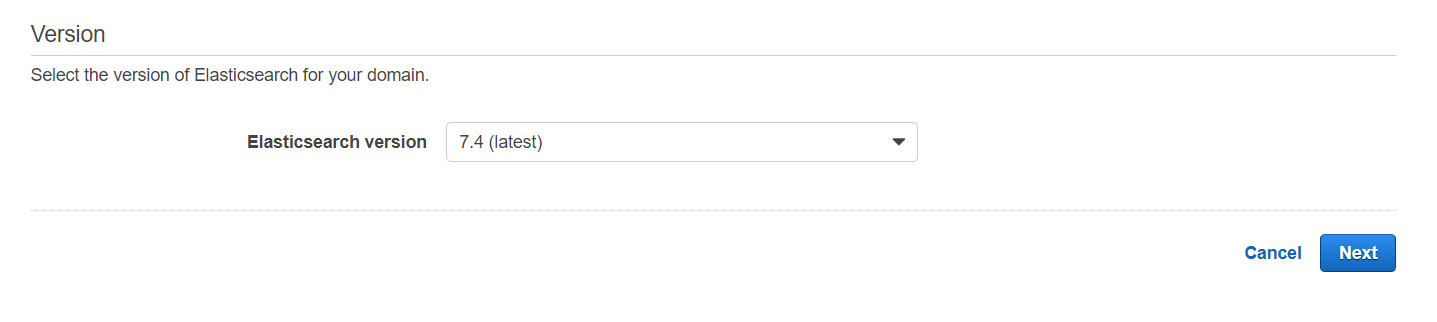
6-8.　Check S3 management console if S3 bucket is surely created.

6-9．　Access to Elasticsearch service management console

6-10.　Click [Create a new domain]

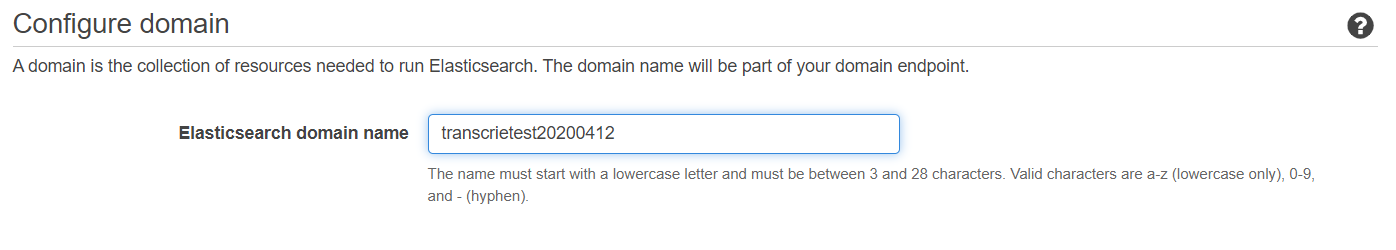


6-11. Select [Development and testing]



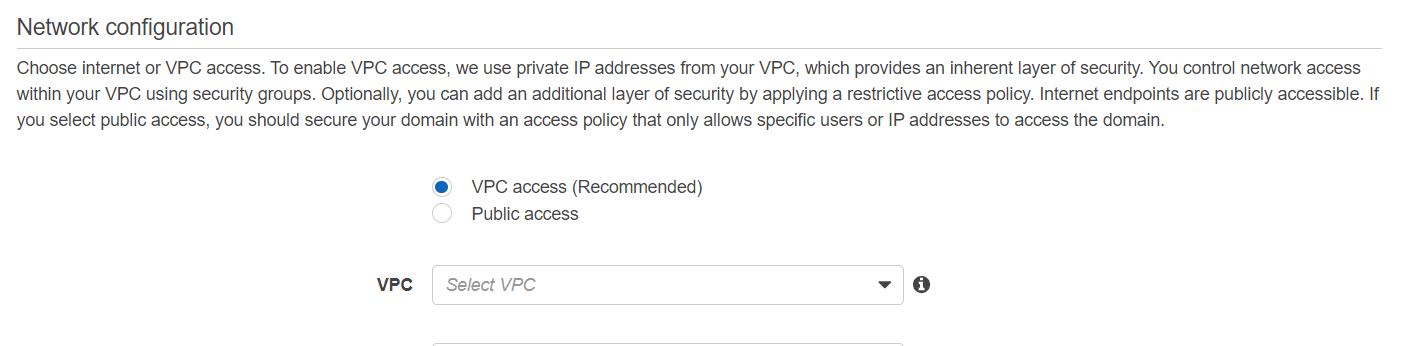
6-12. Select 6.2, and click [Next]

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



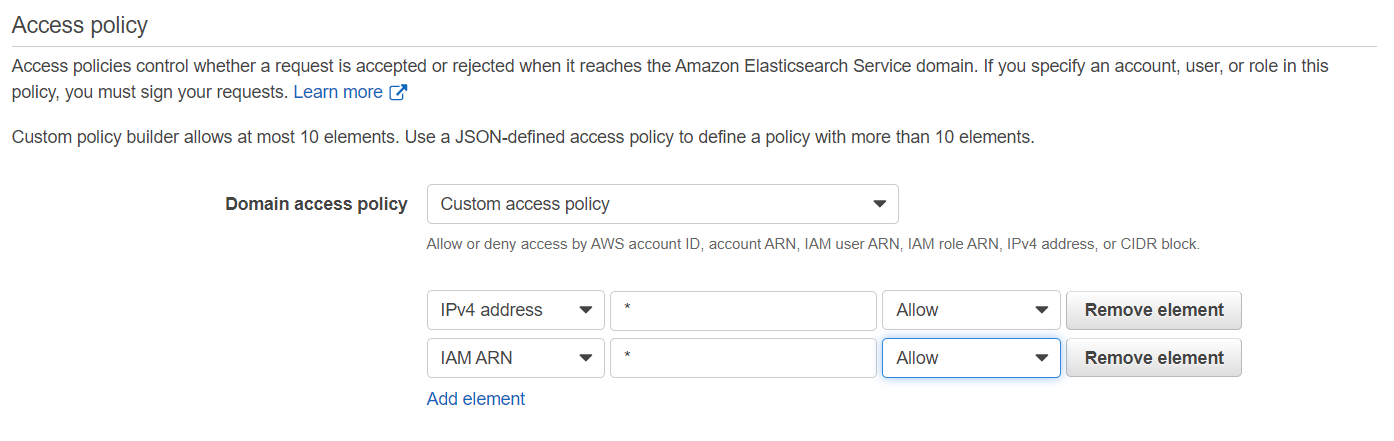
6-14．Click [Next]

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

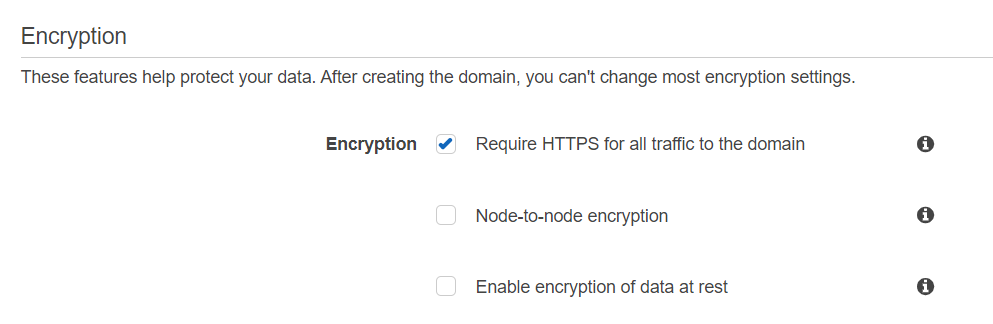


# You can create domain inside of VPC instead of Public for security. But workshop protocol gets long. Thus, we skip it this time.

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

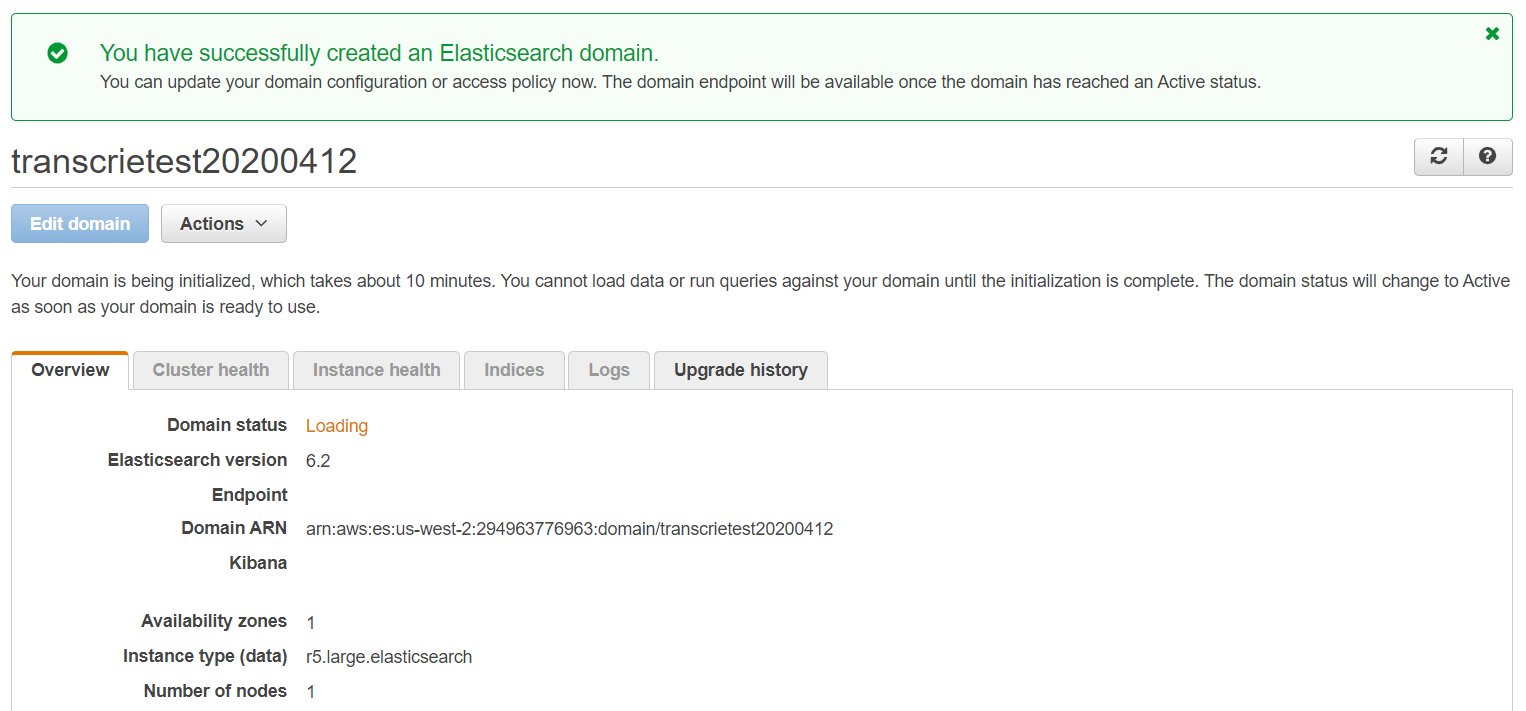


6-17.　Stay Encryption setting as default to require HTTPS access to the domain.



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.



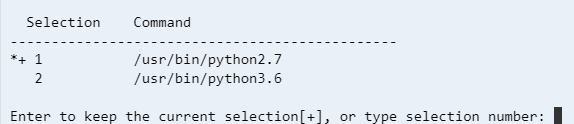
6-20．Type below 2 commands at Cloud9 console. Please type directly without copy. (If you copy and paste at windows environment, [--] changes to [-])

python –-version

pip –-version

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

sudo update-alternatives --config python



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

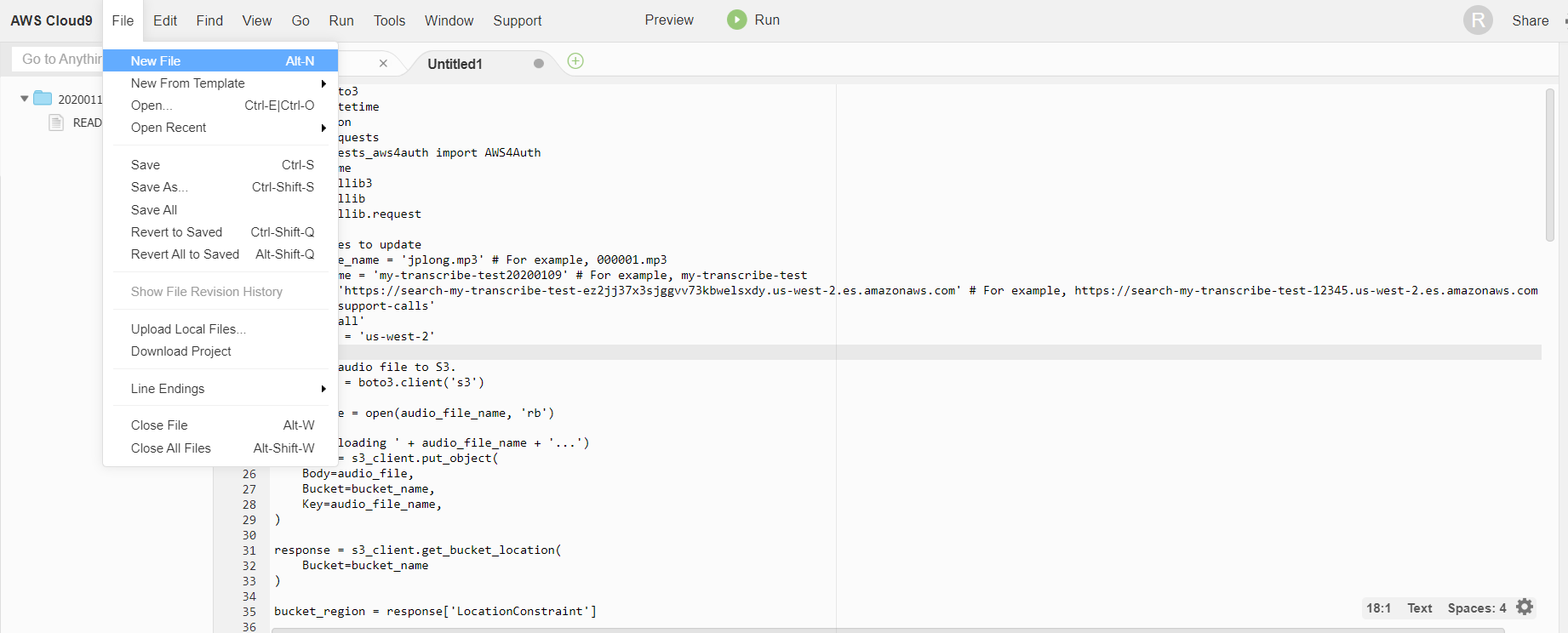
6-21．　Type below command to install required libraries.

pip install boto3

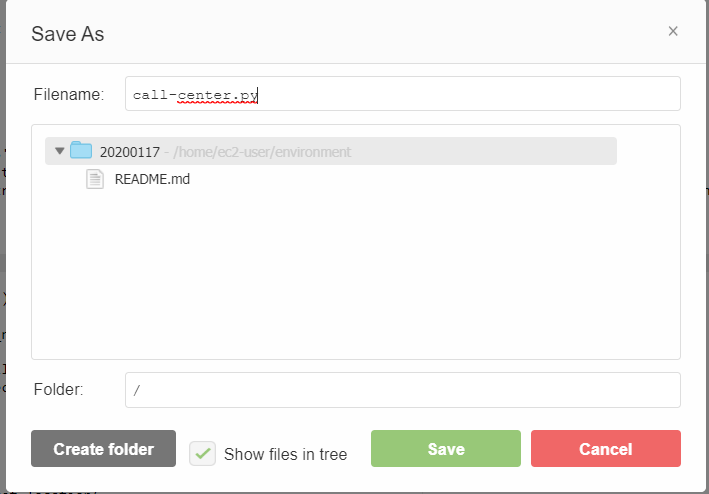
pip install requests

sudo pip install requests\_aws4auth

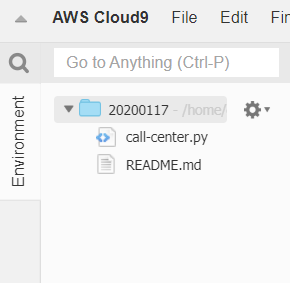
6-22.　Select [File] → [New File], and copy and paste [call-center.py] into the new tab.



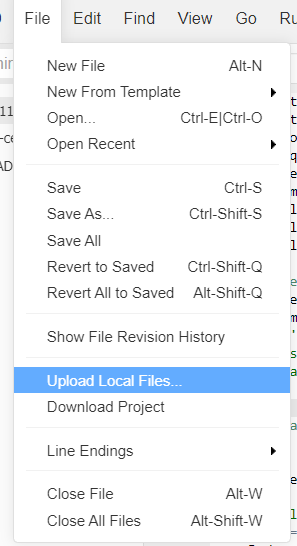
6-23.　Click [File] → [Save], and type [call-center.py] to the filename. Then, click [Save]



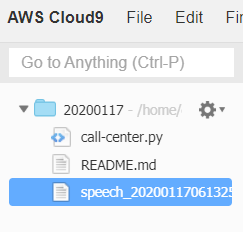
6-24. The folder looks like below



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;



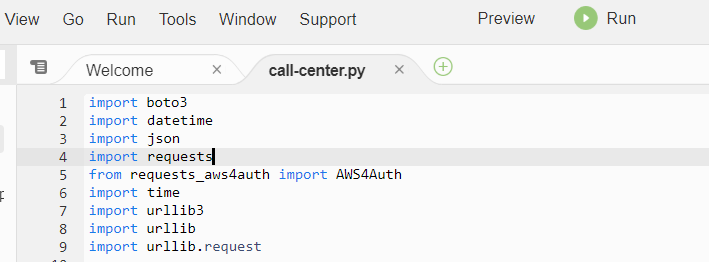
6-27.　modify Python script of 12th,13th, and 14th line.

12th line ：uploaded mp3 file name

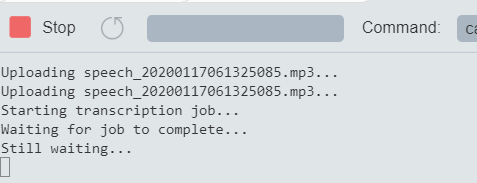
13th ：S3 bucket name created from Cloud9

14th ：Elasticsearch service Endpoint (not Kibana URL)

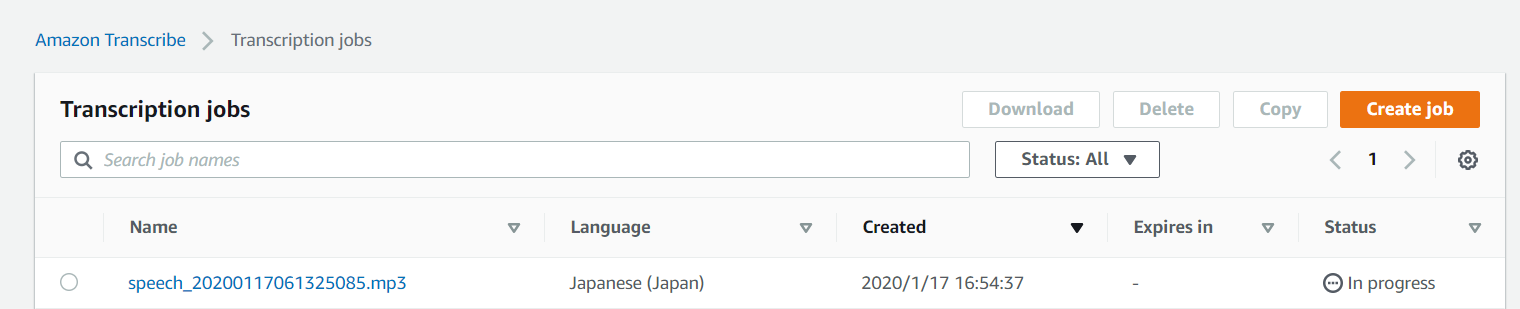
6-28.　Save the file, and Let’s Run the script. You can click Run button.



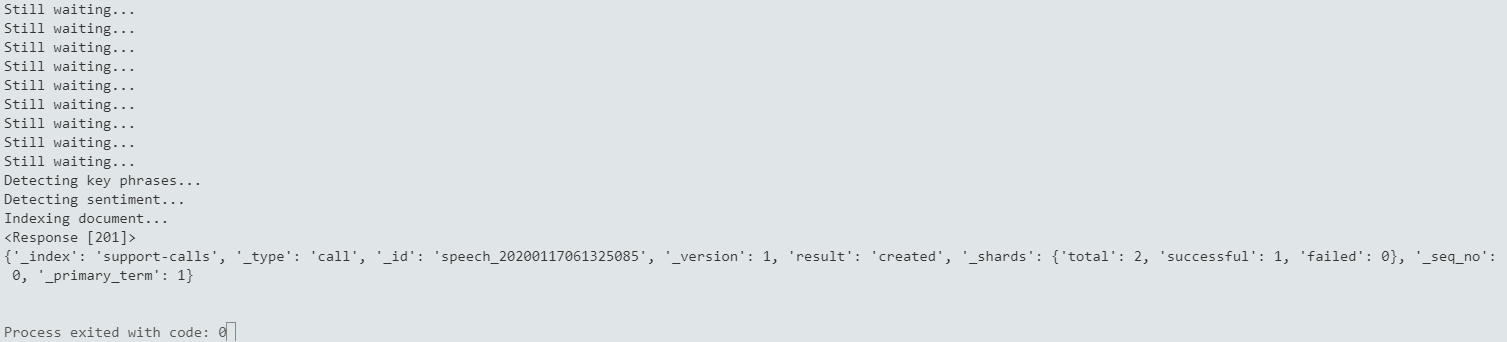
6-29.　If you see below, it is working correctly.

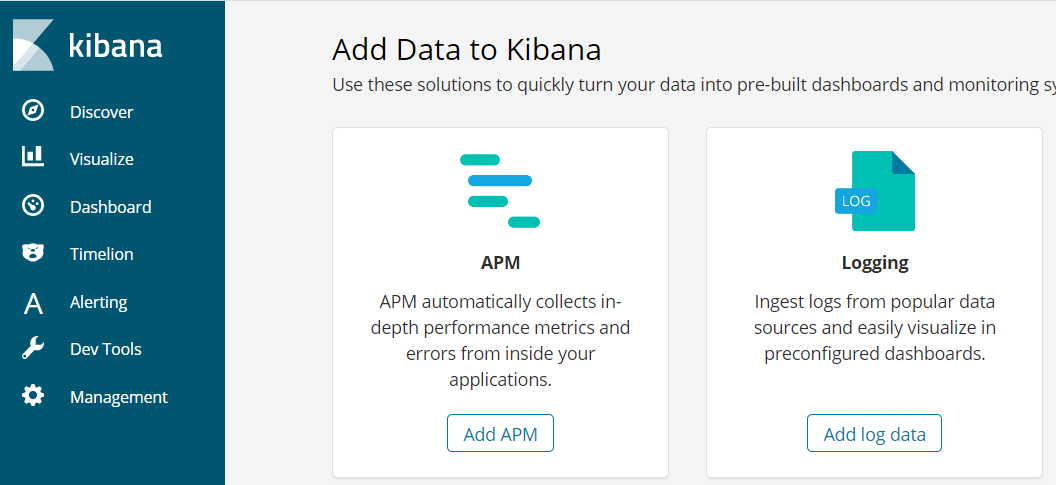


6-30.　If you access to Transcribe management console, Job is in progress

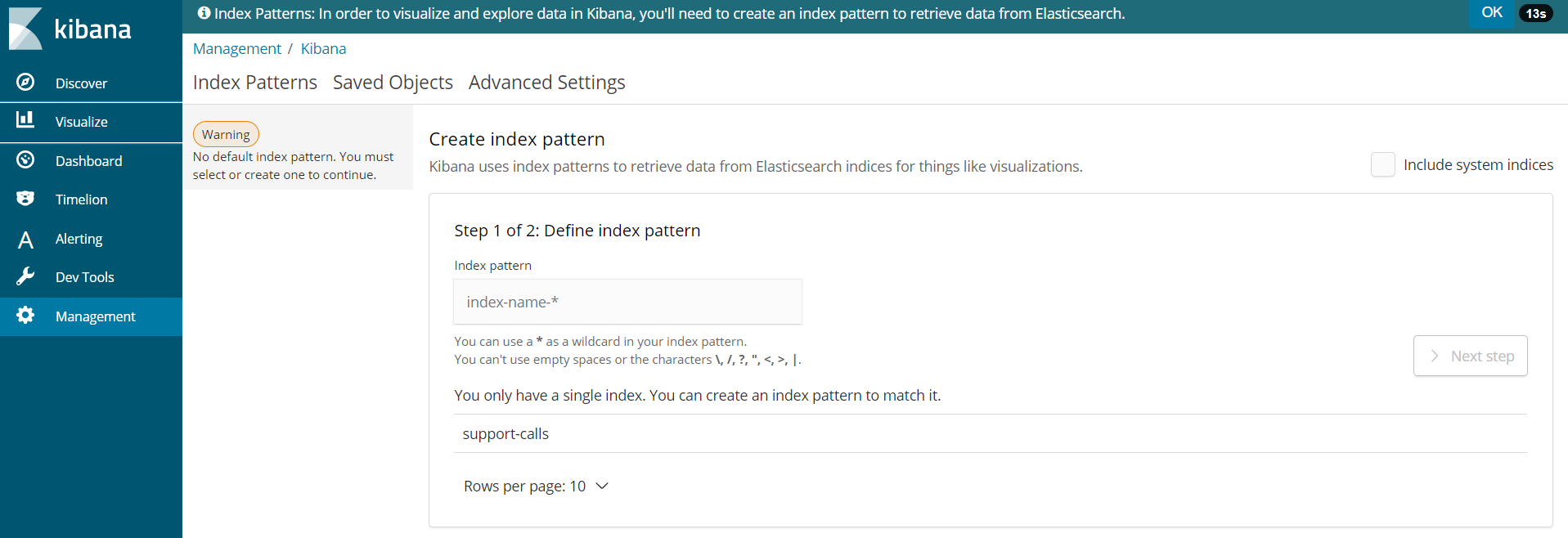


z6-31.　If you see [Process exited with code: 0], The job is done, and transcribe text is passed to comprehend, and its analysis result was imported to Elasticsearch.

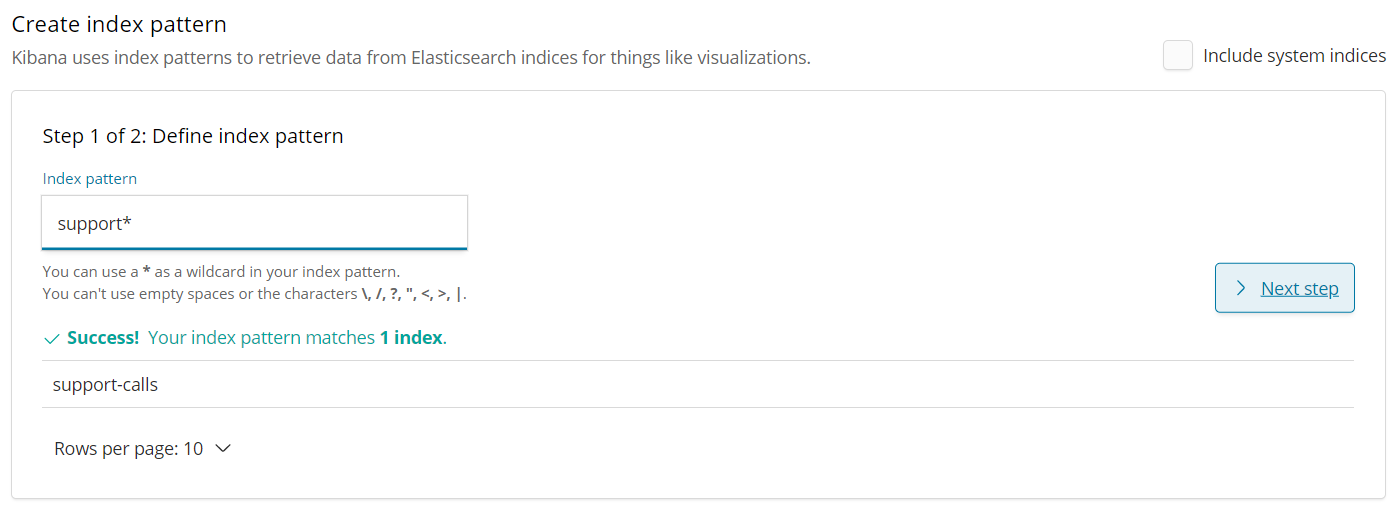


6-31.　Access to Kibana URL via your browser.

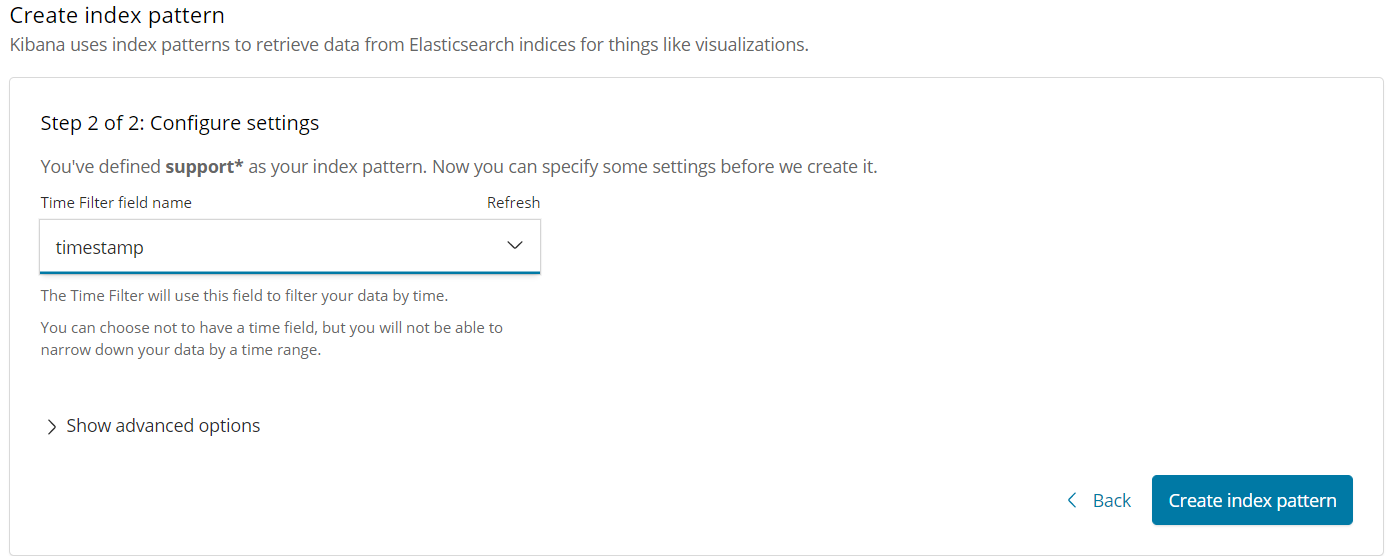
6-32.　Click [Visualize]



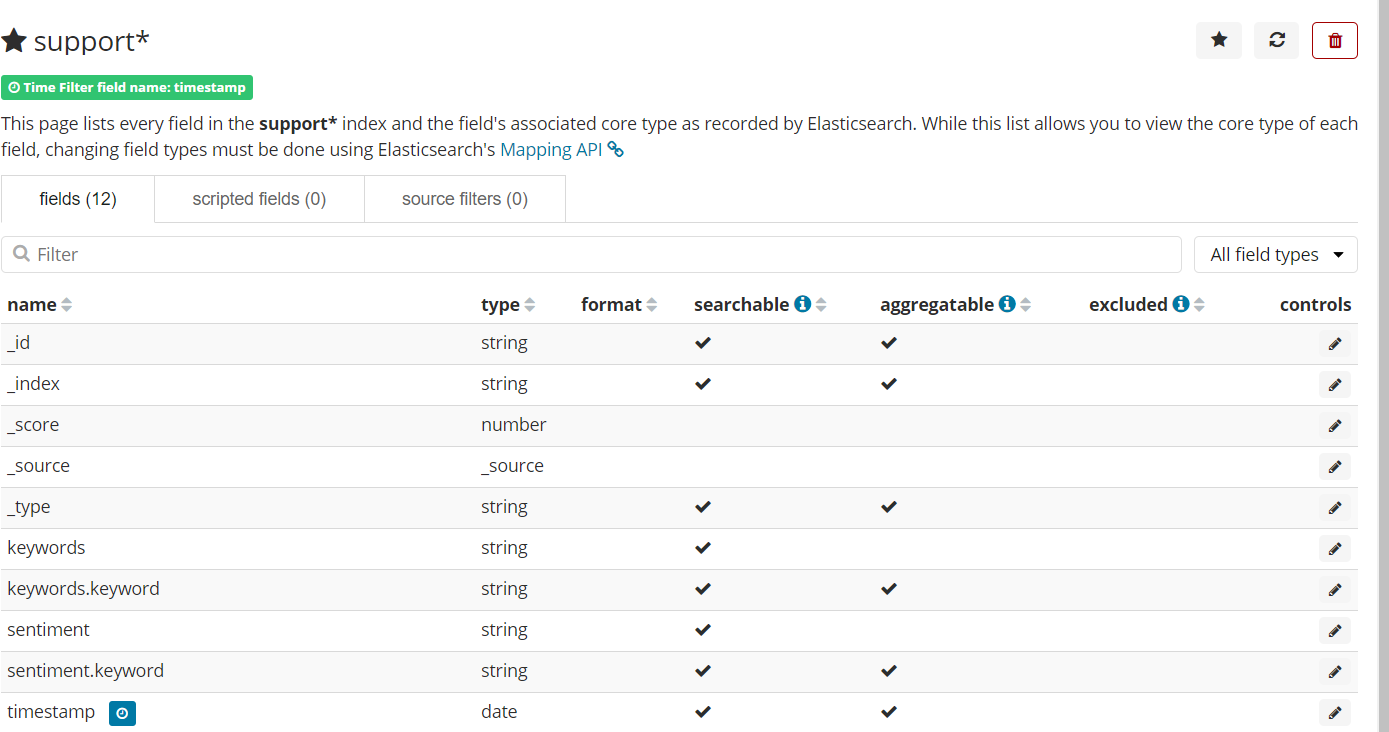
6-32.　Type [support\*] to [Index Pattern], and click [Next Step]



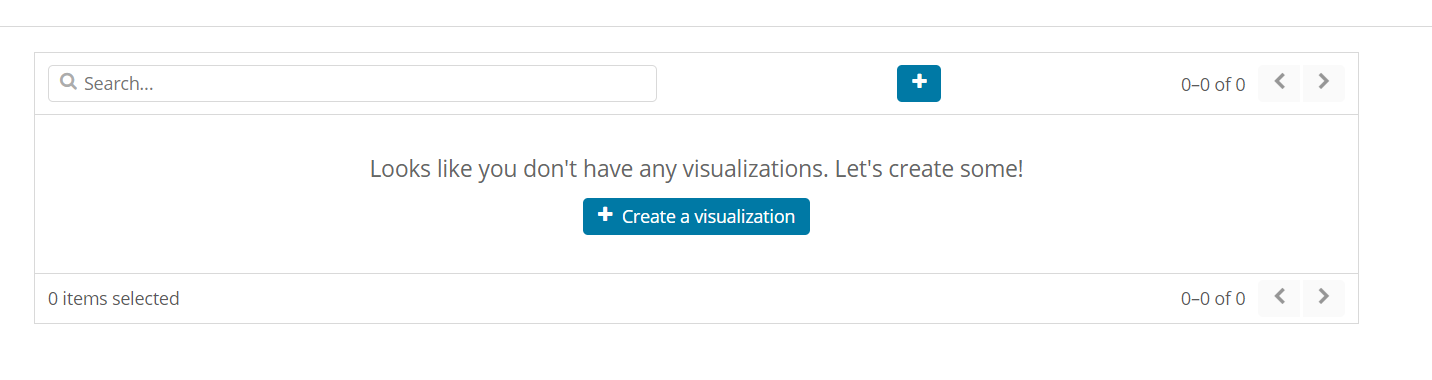
6-33. Select timestamp under [Time Filter field name], and click [Create Index Pattern]



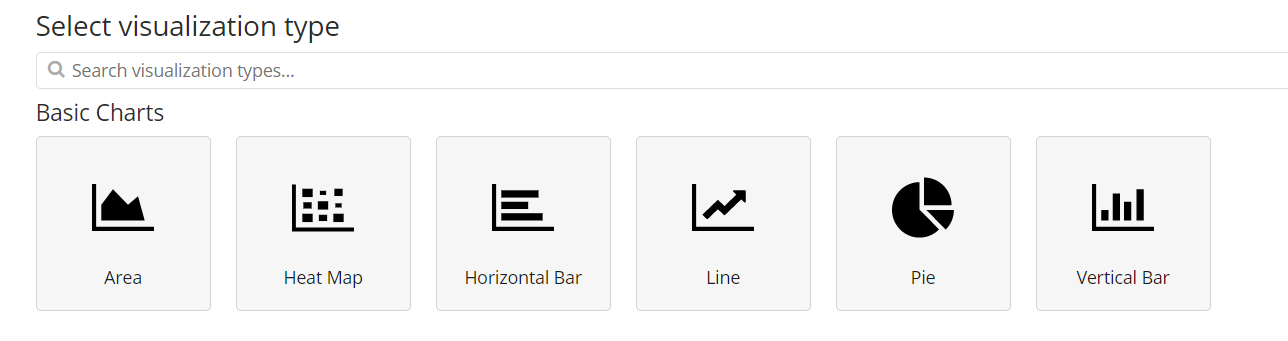
6-34.　If you see below screen, you correctly set index.



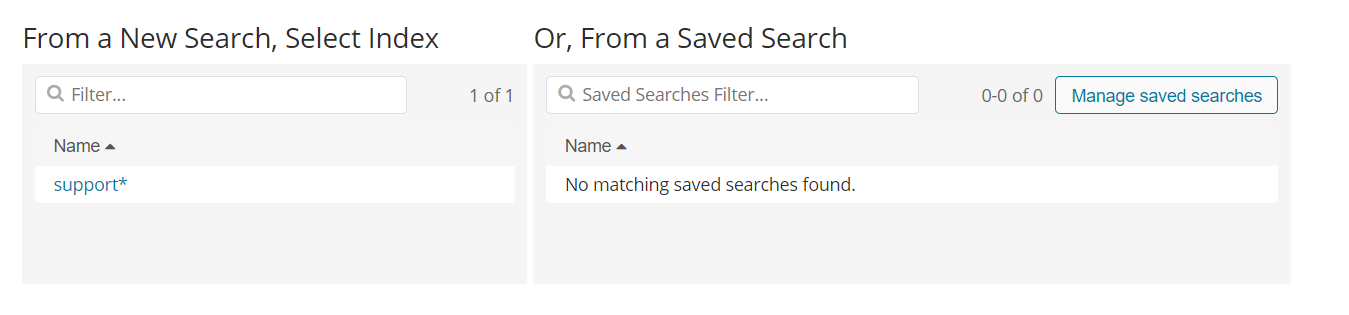
6-35.　Click [Visualize] again.



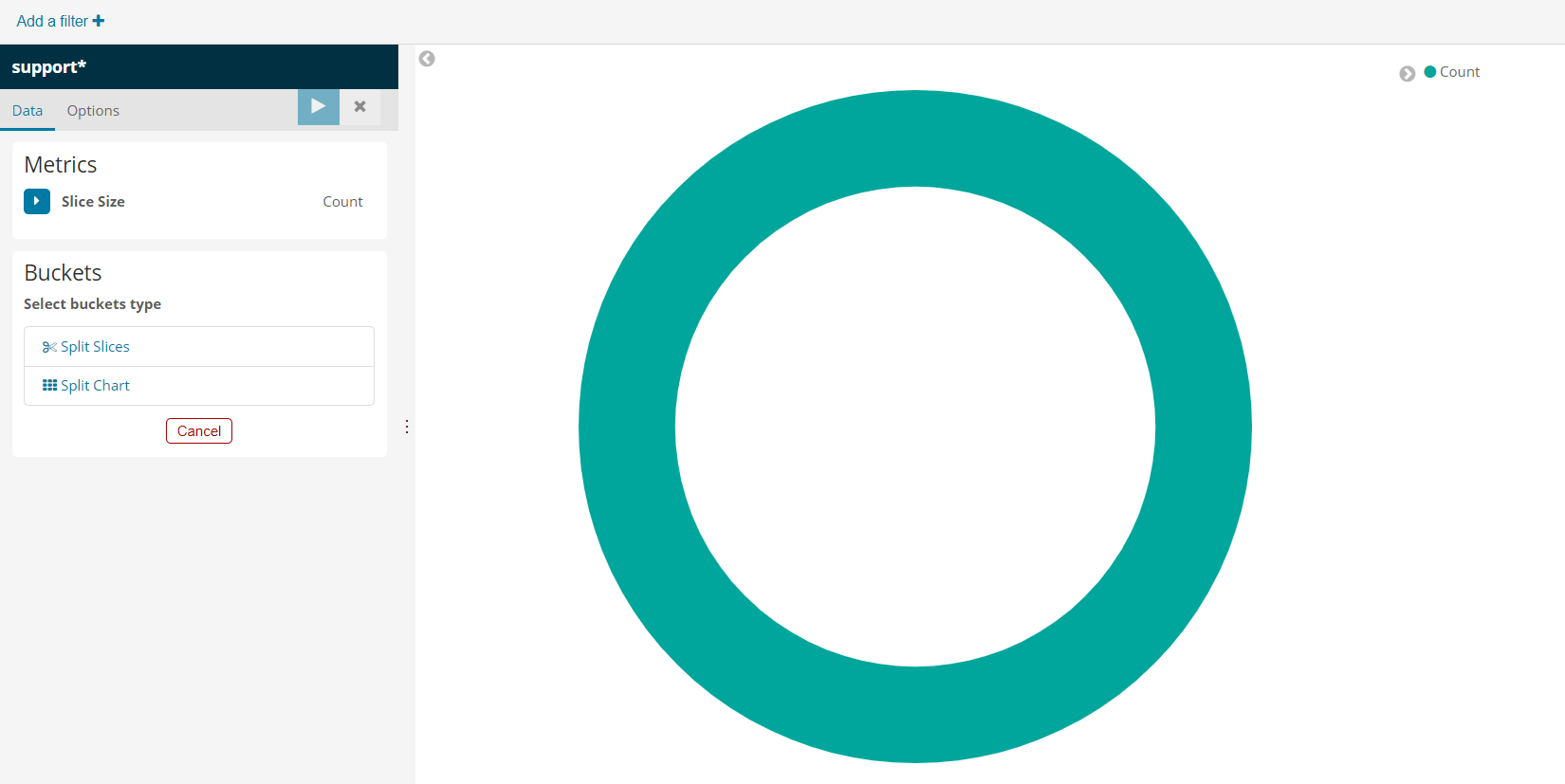
6-36.　Click [Create a visualization]



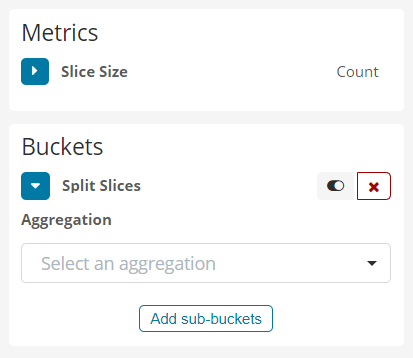
6-37.　Select Pie chart



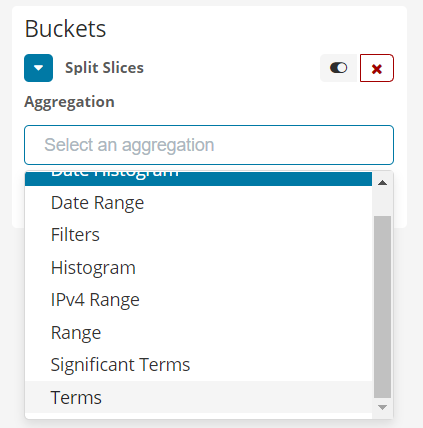
6-38.　Select [support\*\*]. (you will see 2 [\*\*].)



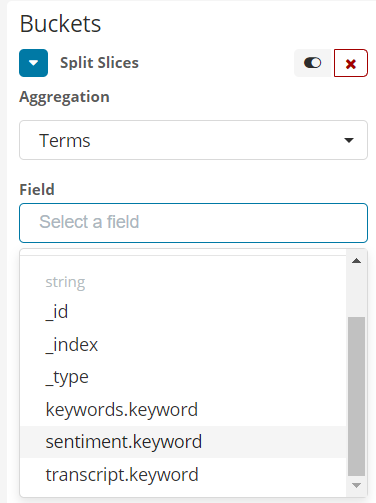
6-39.　Select [Split Slices] under the Buckets



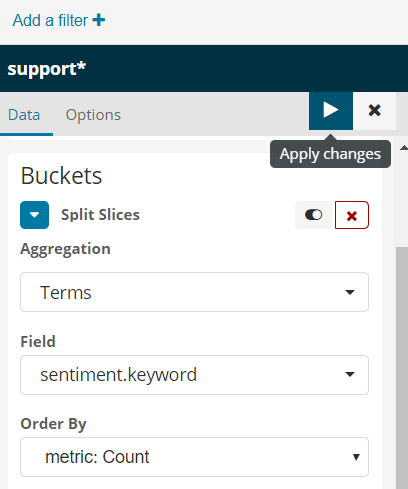
6-40.　Select [Terms] under [Aggregation] field



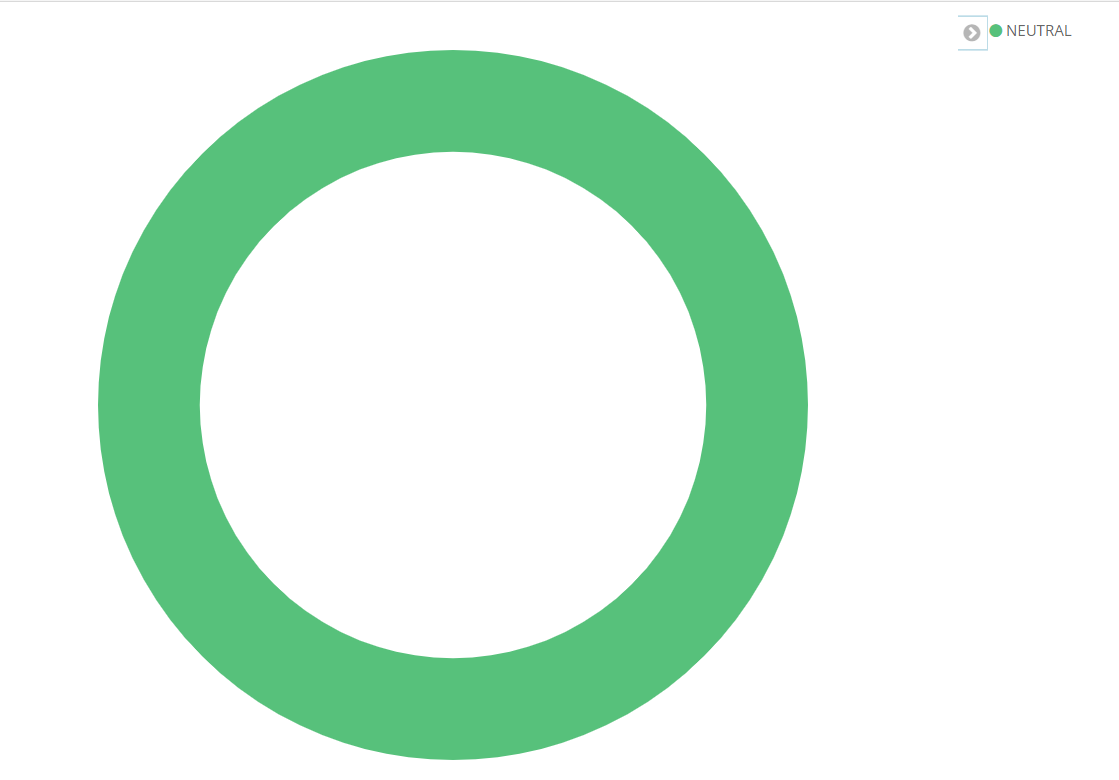
6-41. Select [sentiment.keyword] under [Filed] field.



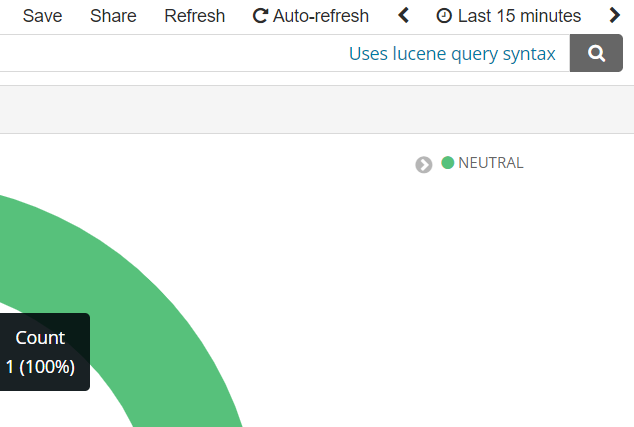
6-42.　Click [Apply changes]



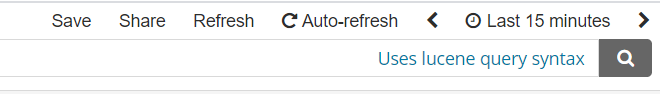
6-43.　You will see result of sentiment analysis (Positive,Neutral,Negative）If you run python script with multiple Polly mp3 files, you will see variety of result.



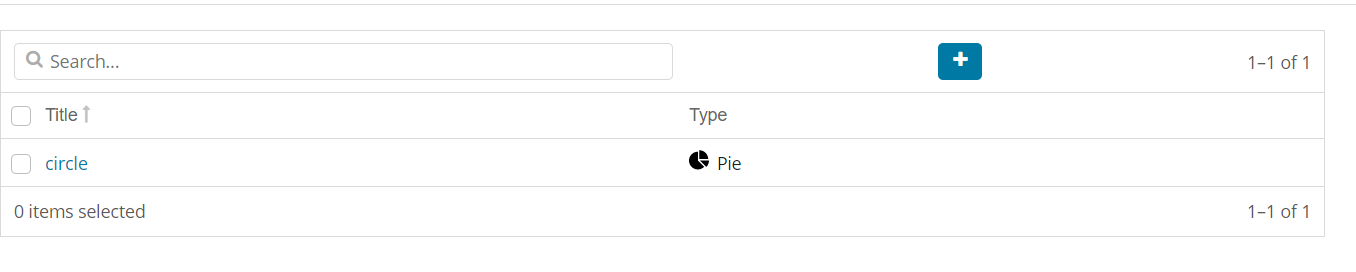
By default, Graph shows only latest 15mins. You can change the range to click [Last 15 minutes].



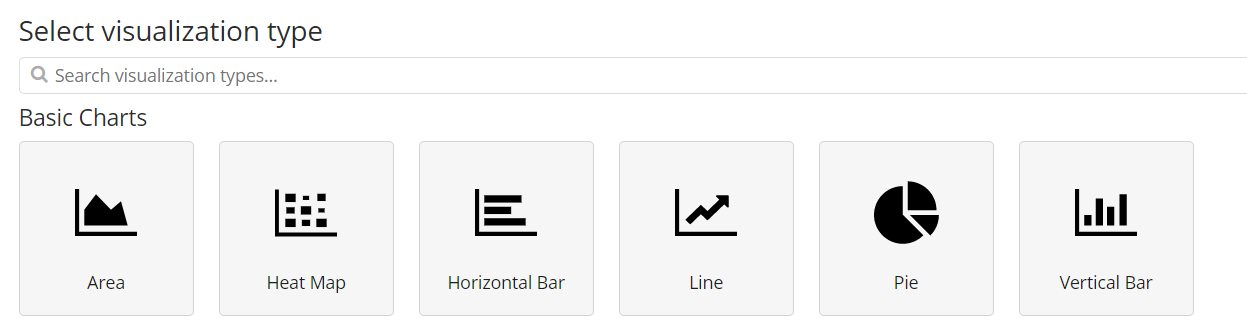
6-44. Click [Save], and type appropriate name.

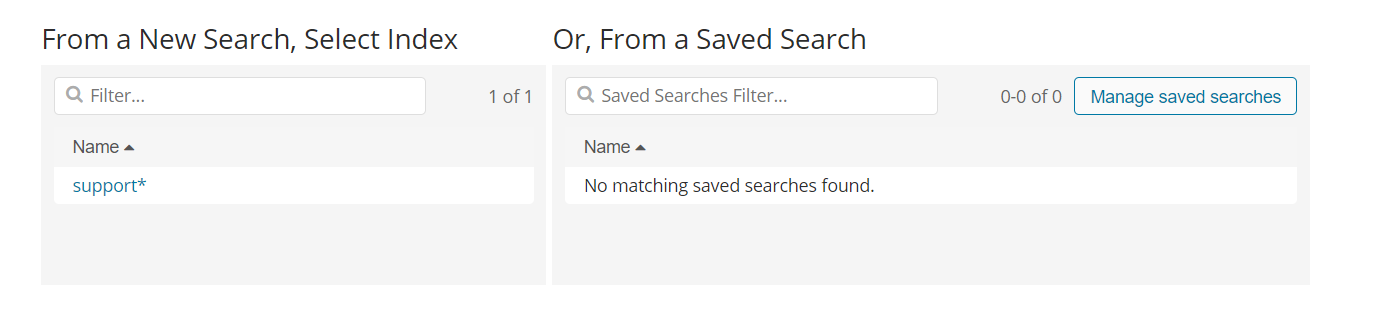


6-45.　Click [Visualize] again

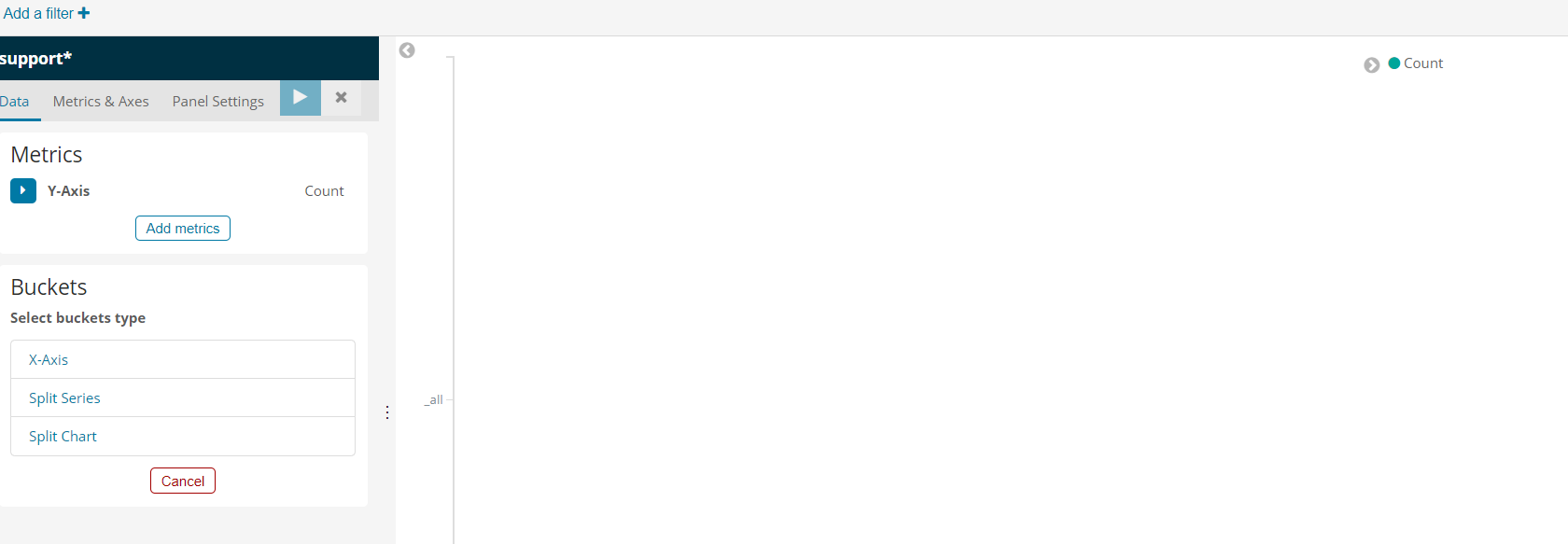


6-46.　Press [+] button to create another graph.

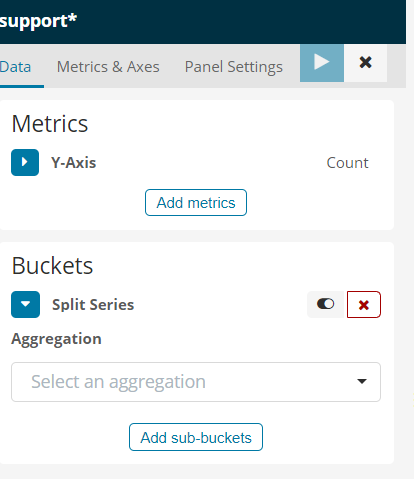


6-47.　Select Horizontal Bar

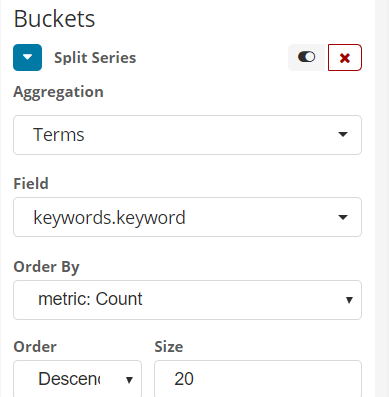
6-48.　Select [support\*\*]



6-49.　Select [Split Series]



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



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.

