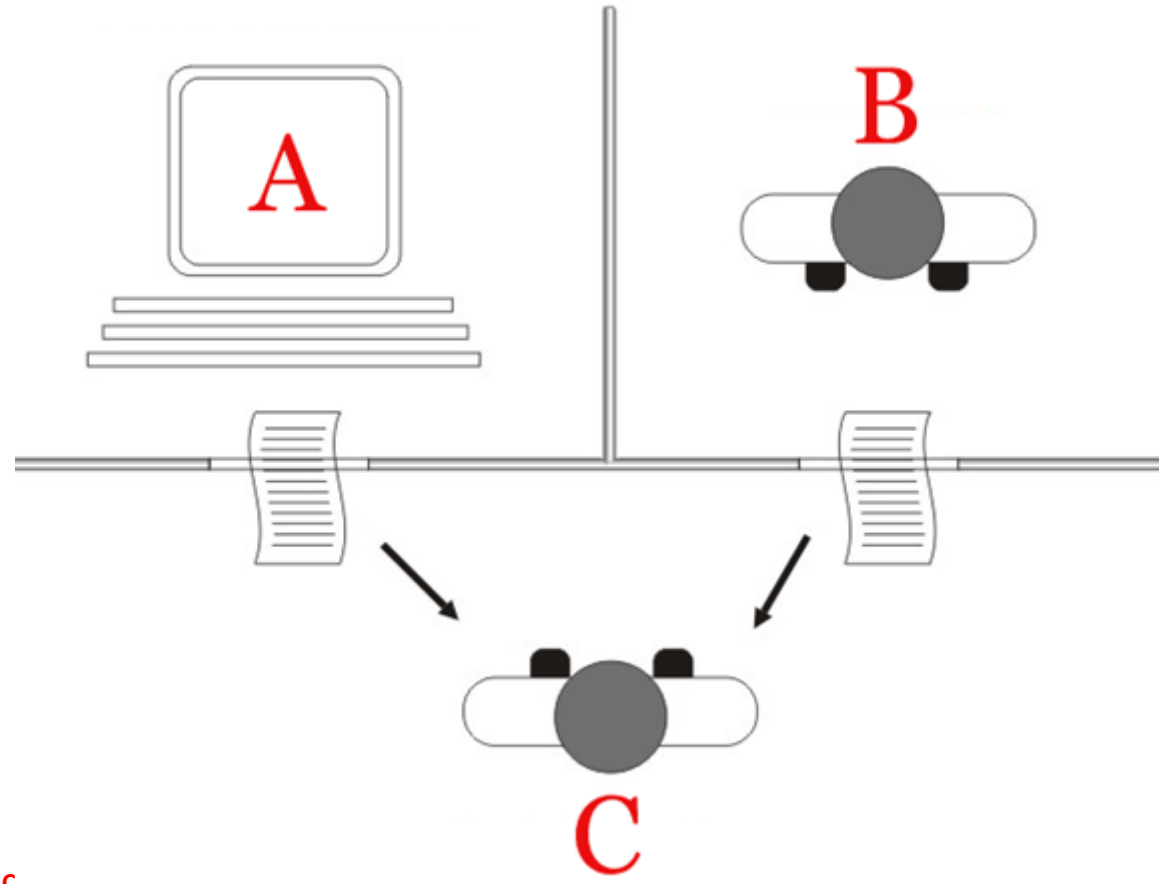
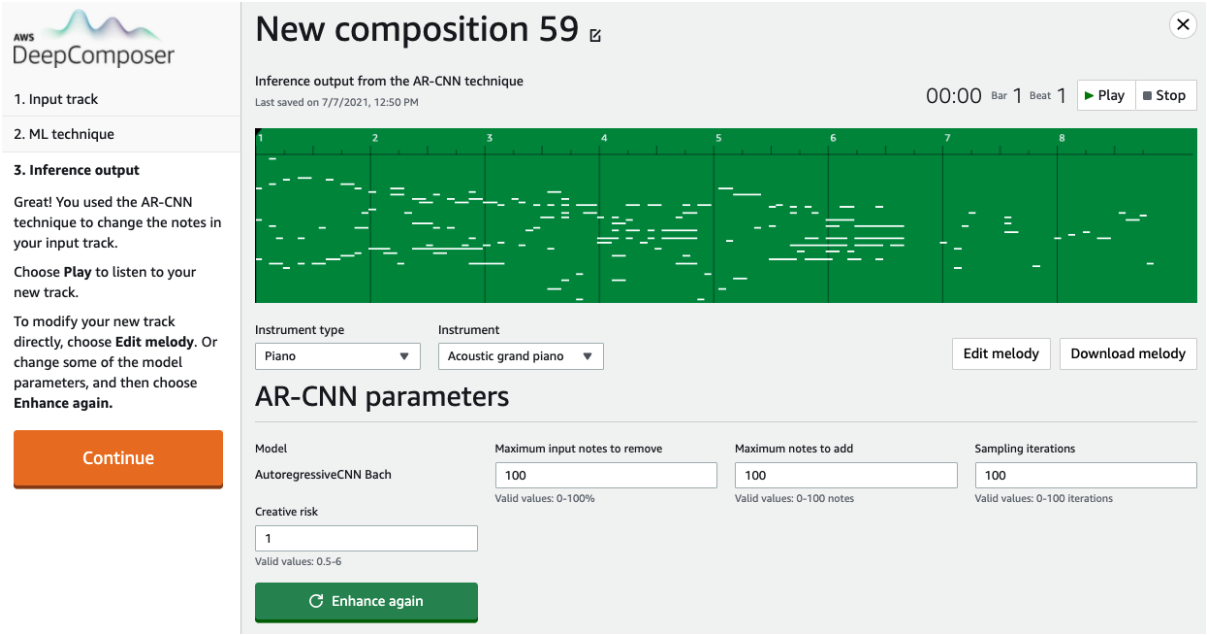
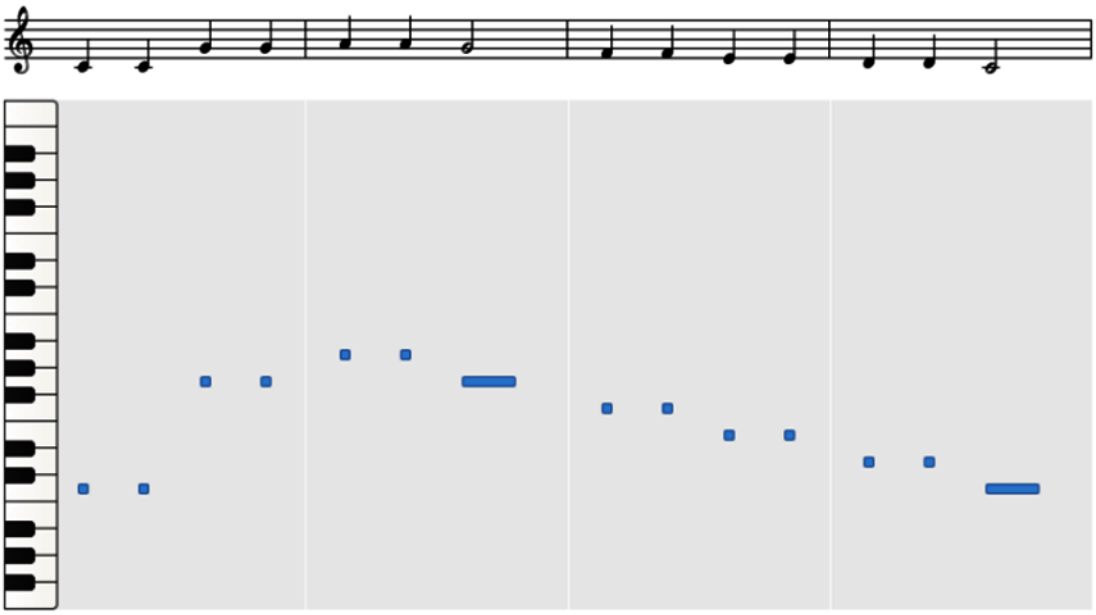
Chapter 1: NLP in the Business Context and Introduction to AWS AI Services

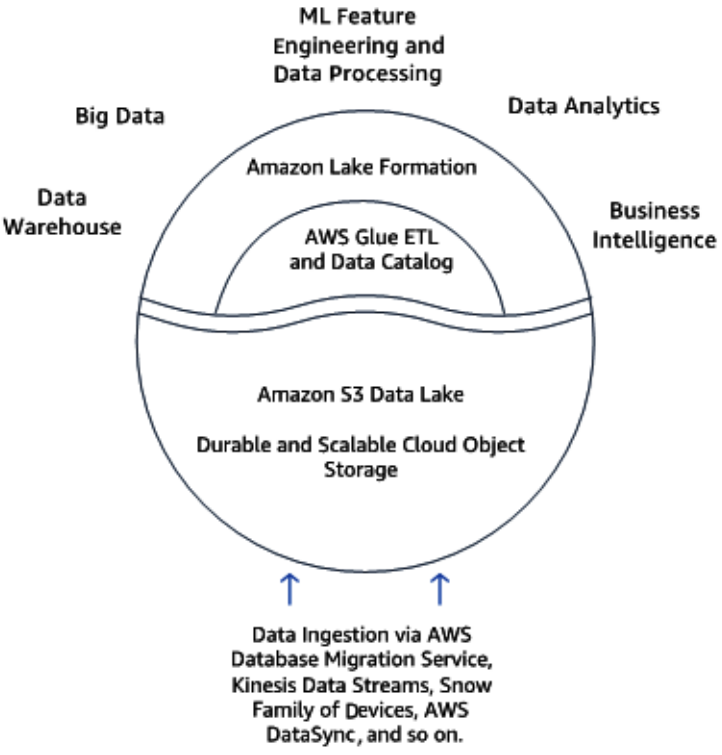


C

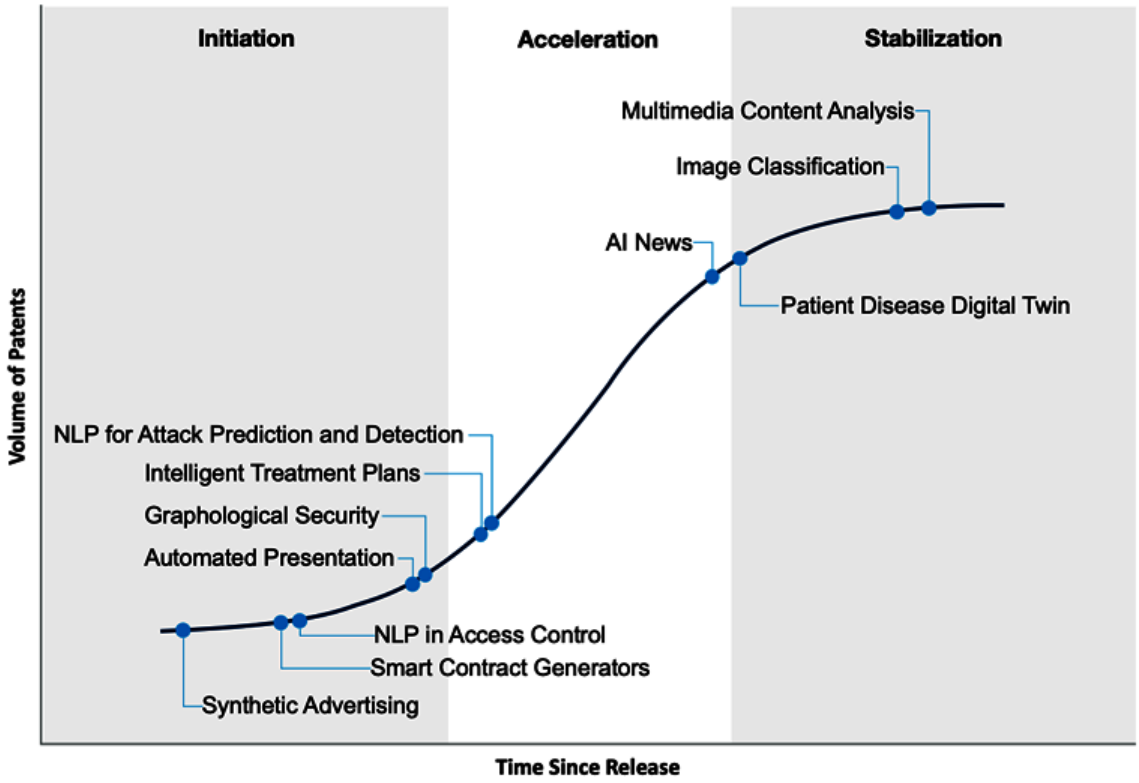




Starting Word	CATS
Step 1	MATS
Step 2	MATE
Step 3	MACE
Ending Word	MICE

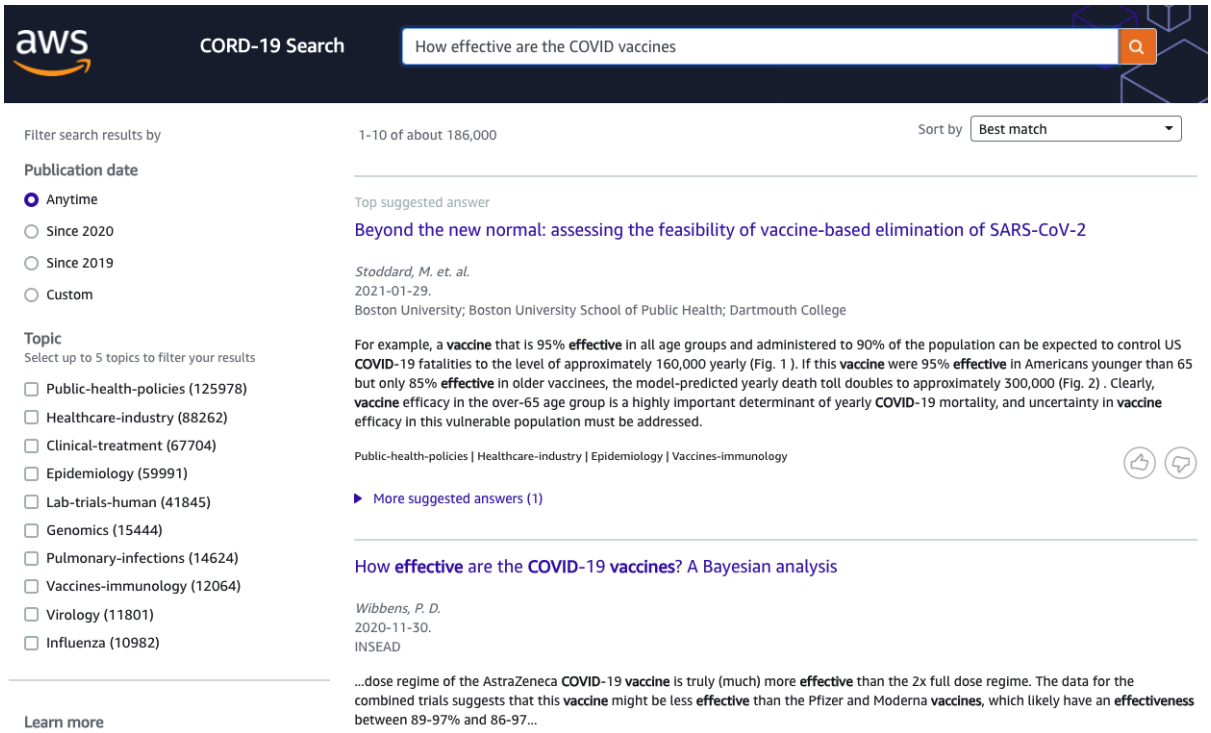


NLP Emergence Cycle: Intelligent Applications & Augmented/Text Analytics



Source: Gartner, August 2020

729322



Prepare	Build	Train & Tune	Deploy & Manage
SageMaker Ground Truth	**SageMaker Studio**	**One-Click Training**	**One-Click Deployment**
SageMaker Data Wrangler	**Built-In and Bring-Your-Own-Algorithms**	**SageMaker Experiments**	**Kubernetes and Kubeflow Integration**
SageMaker Processing	**Local Mode**	**Automatic Model Tuning**	**Multi-Model Endpoints**
SageMaker Feature Store	**SageMaker Autopilot**	**Distributed Training Libraries**	**SageMaker Model Monitor**
SageMaker Clarify	**SageMaker JumpStart**	**SageMaker Debugger**	**SageMaker Edge Manager**
		Managed Spot Training	**SageMaker Pipelines**

Prepare	Build	Train & Tune	Deploy & Manage
SageMaker Ground Truth	**SageMaker Studio**	**One-Click Training**	**One-Click Deployment**
SageMaker Data Wrangler	**Built-In and Bring-Your-Own-Algorithms**	**SageMaker Experiments**	**Kubernetes and Kubeflow Integration**
SageMaker Processing	**Local Mode**	**Automatic Model Tuning**	**Multi-Model Endpoints**
SageMaker Feature Store	**SageMaker Autopilot**	**Distributed Training Libraries**	**SageMaker Model Monitor**
SageMaker Clarify	**SageMaker JumpStart**	**SageMaker Debugger**	**SageMaker Edge Manager**
		Managed Spot Training	**SageMaker Pipelines**

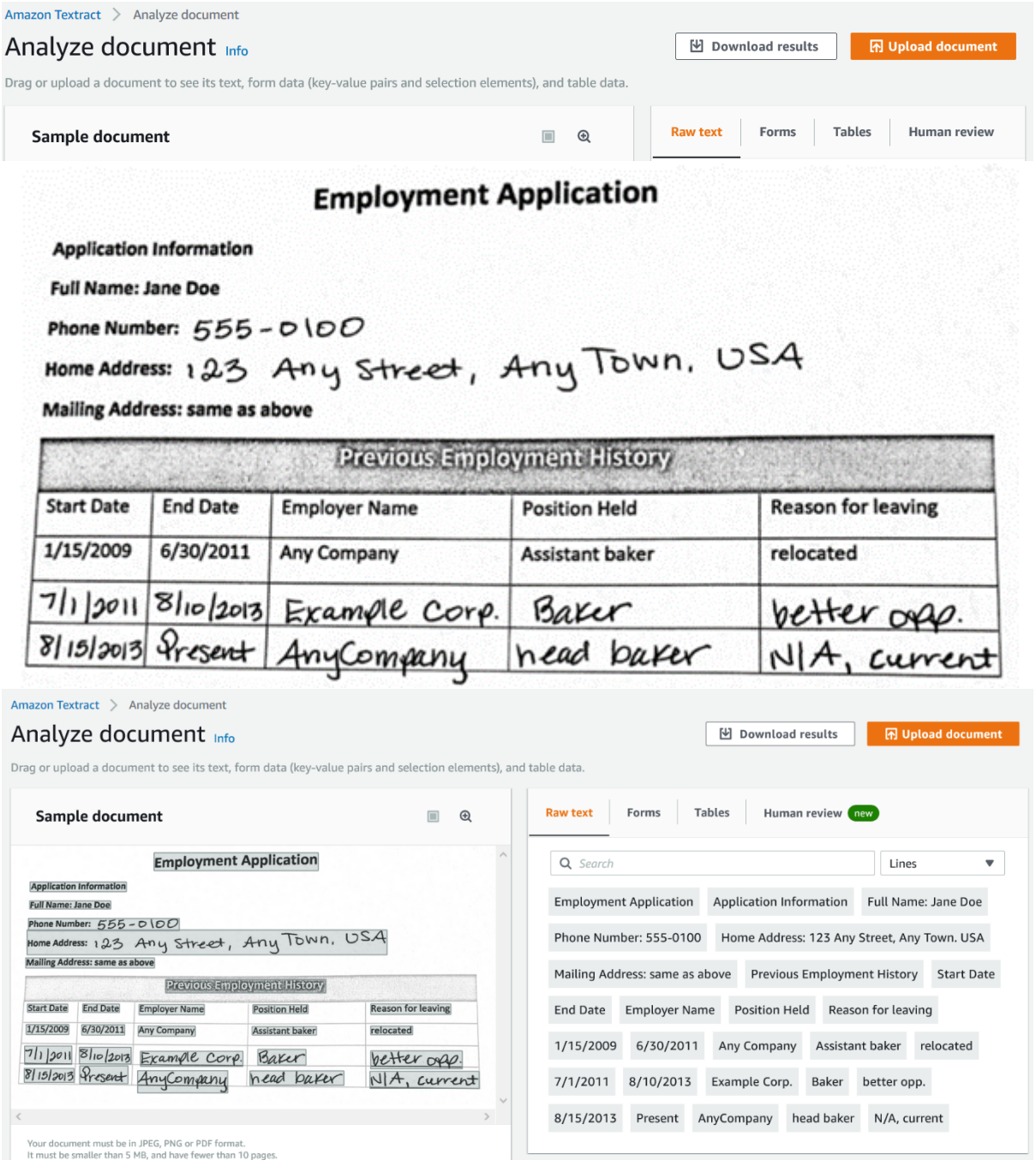
Health AI	Industrial AI	Fraud	Anomaly Detection	Contact Centers	Code and DevOps
• Amazon HealthLake • Amazon Transcribe Medical • Amazon Comprehend Medical	• AWS Panorama • Amazon Monitron • Amazon Lookout for Equipment • Amazon Lookout for Vision	• Amazon Fraud Detector	• Amazon Lookout for Metrics	• Contact Lens for Amazon Connect	• DevOps Guru • Code Guru

Vision	Speech	Text	Search	Chatbots	Personalization	Forecasting
• Amazon Rekognition	• Amazon Polly • Amazon Transcribe	• Amazon Textract • Amazon Translate • Amazon Comprehend	• Amazon Kendra	• Amazon Lex	• Amazon Personalize	• Amazon Forecast

Health AI	Industrial AI	Fraud	Anomaly Detection	Contact Centers	Code and DevOps
• Amazon HealthLake • Amazon Transcribe Medical • Amazon Comprehend Medical	• AWS Panorama • Amazon Monitron • Amazon Lookout for Equipment • Amazon Lookout for Vision	• Amazon Fraud Detector	• Amazon Lookout for Metrics	• Contact Lens for Amazon Connect	• DevOps Guru • Code Guru

Vision	Speech	Text	Search	Chatbots	Personalization	Forecasting
• Amazon Rekognition	• Amazon Polly • Amazon Transcribe	• Amazon Textract • Amazon Translate • Amazon Comprehend	• Amazon Kendra	• Amazon Lex	• Amazon Personalize	• Amazon Forecast

Chapter 2: Introducing Amazon Textract



Raw text | Forms | Tables | Human review new

Search | Lines

Employment Application | Application Information | Full Name: Jane Doe

Phone Number: 555-0100 | Home Address: 123 Any Street, Any Town. USA

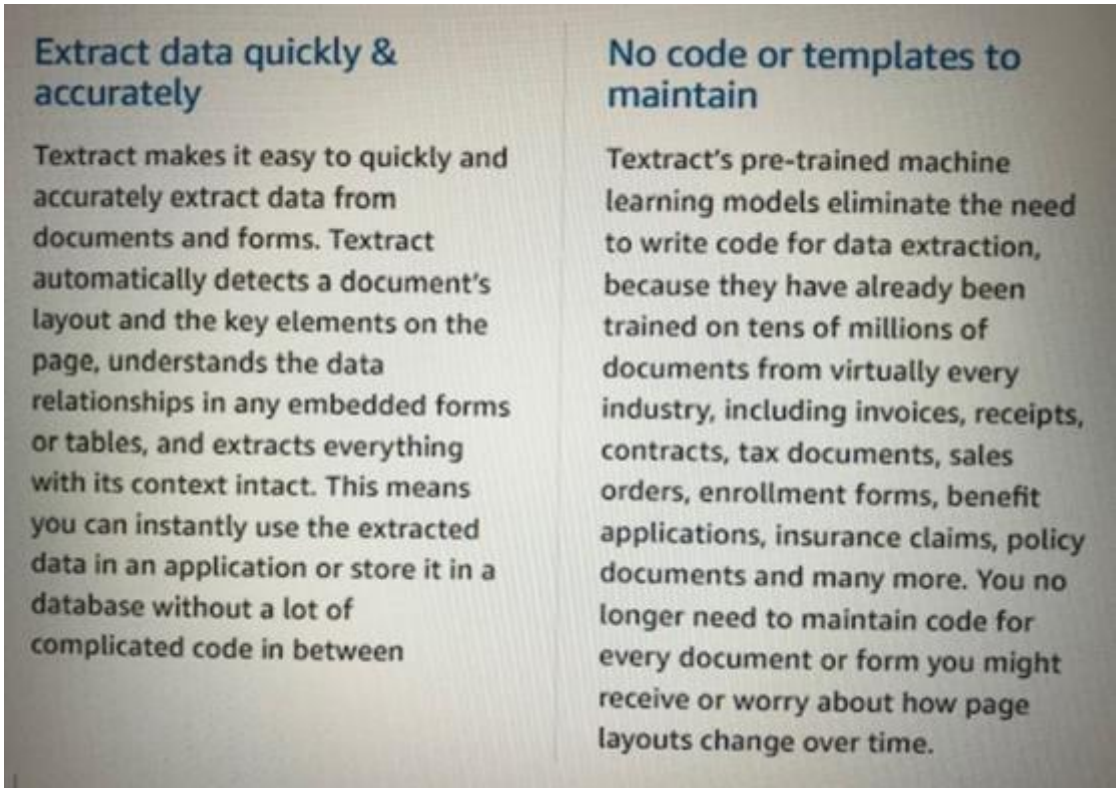
Mailing Address: same as above | Previous Employment History | Start Date

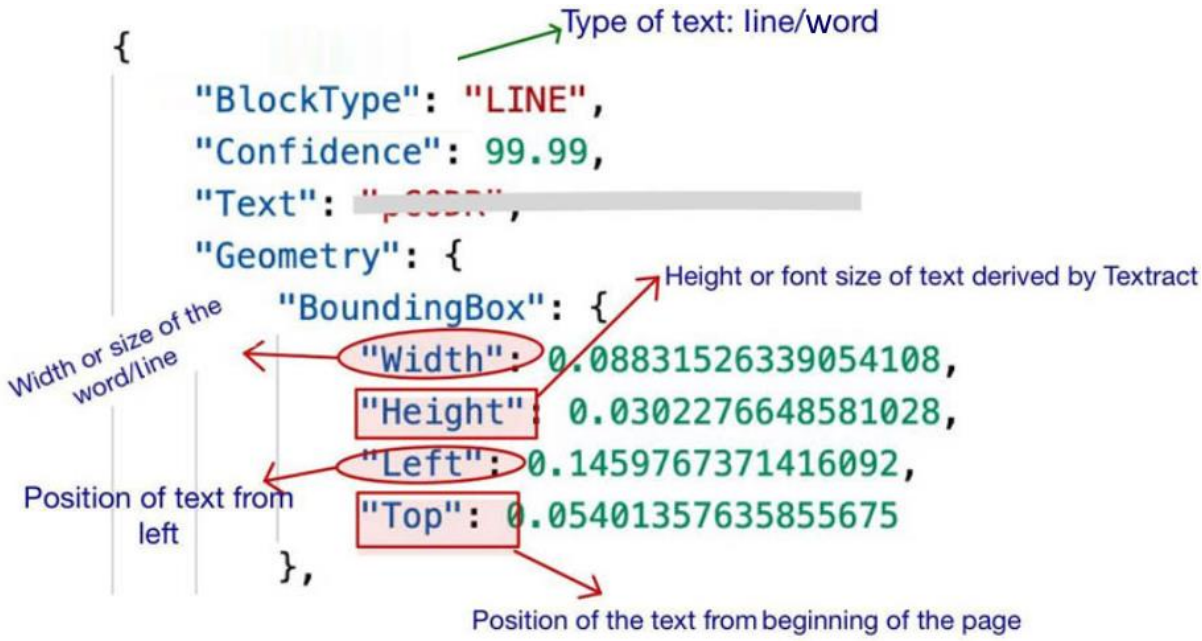
End Date | Employer Name | Position Held | Reason for leaving

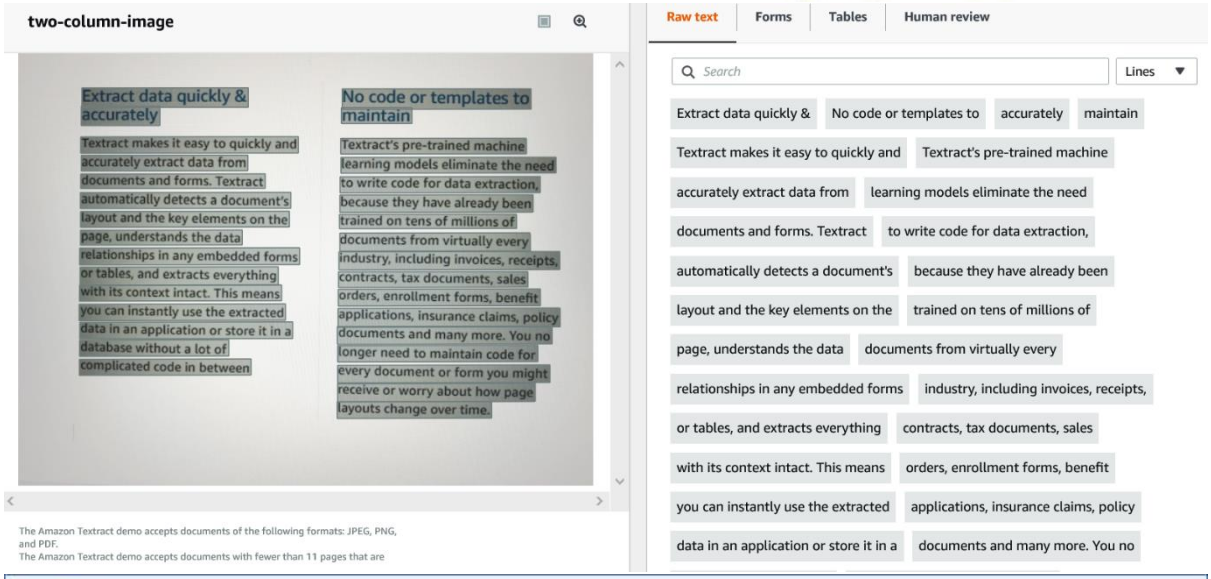
1/15/2009 | 6/30/2011 | Any Company | Assistant baker | relocated

7/1/2011 | 8/10/2013 | Example Corp. | Baker | better opp.

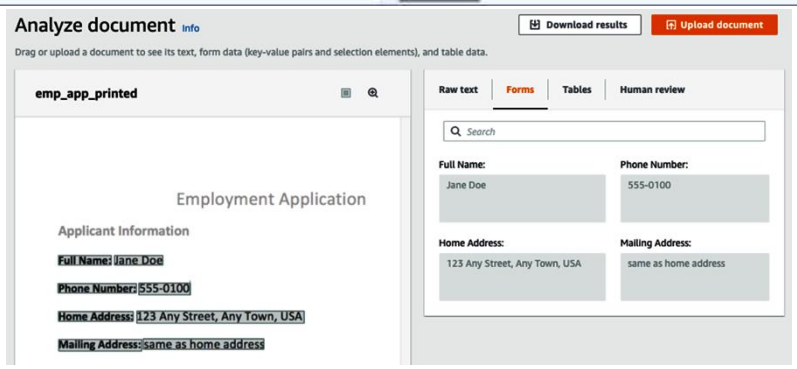
8/15/2013 | Present | AnyCompany | head baker | N/A, current

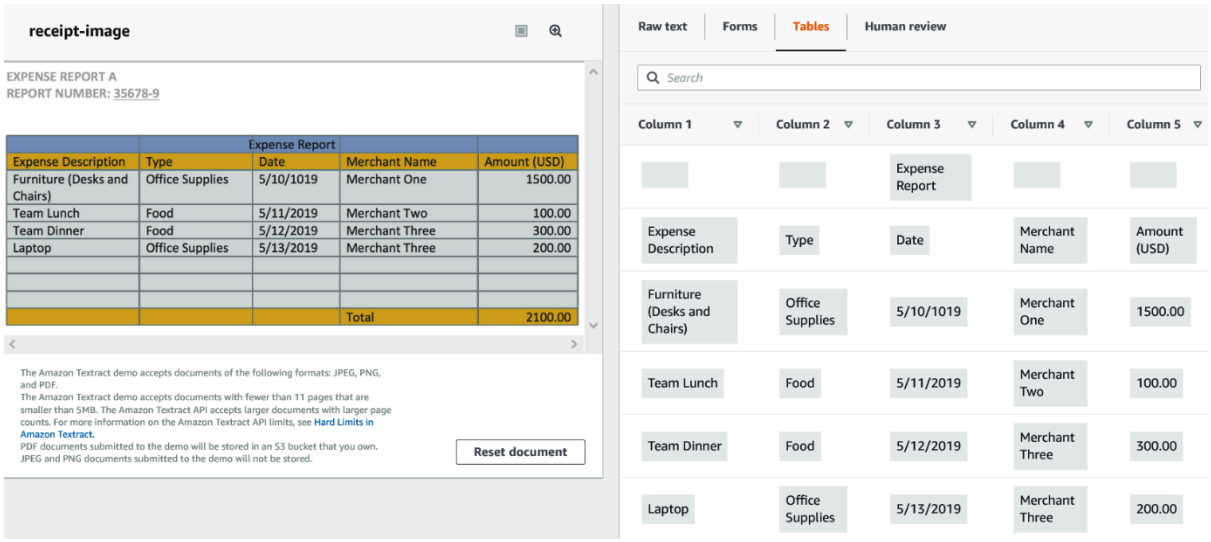



Type of text: line/word
{
"BlockType": "LINE",
"Confidence": 99.99,
"Text":
"Geometry": {
"BoundingBox": {
"Width": 0.08831526339054108,
"Height": 0.0302276648581028,
"Left": 0.1459767371416092,
"Top": 0.05401357635855675
},
Width or size of the word/line
Height or font size of text derived by Textract
Position of text from left
Position of the text from beginning of the page


two-column-image
Extract data quickly & accurately
Textract makes it easy to quickly and accurately extract data from documents and forms. Textract automatically detects a document's layout and the key elements on the page, understands the data relationships in any embedded forms or tables, and extracts everything with its context intact. This means you can instantly use the extracted data in an application or store it in a database without a lot of complicated code in between
No code or templates to maintain
Textract's pre-trained machine learning models eliminate the need to write code for data extraction, because they have already been trained on tens of millions of documents from virtually every industry, including invoices, receipts, contracts, tax documents, sales orders, enrollment forms, benefit applications, insurance claims, policy documents and many more. You no longer need to maintain code for every document or form you might receive or worry about how page layouts change over time.
The Amazon Textract demo accepts documents of the following formats: JPEG, PNG, and PDF.
The Amazon Textract demo accepts documents with fewer than 11 pages that are
Raw text
Forms
Tables
Human review
Search
Lines
Extract data quickly &
No code or templates to
accurately
maintain
Textract makes it easy to quickly and
Textract's pre-trained machine
accurately extract data from
learning models eliminate the need
documents and forms. Textract
to write code for data extraction,
automatically detects a document's
because they have already been
layout and the key elements on the
trained on tens of millions of
page, understands the data
documents from virtually every
relationships in any embedded forms
industry, including invoices, receipts,
or tables, and extracts everything
contracts, tax documents, sales
with its context intact. This means
orders, enrollment forms, benefit
you can instantly use the extracted
applications, insurance claims, policy
data in an application or store it in a
documents and many more. You no

Amazon Textract > Analyze document
Analyze document Info
Download results
Upload document
Drag or upload a document to see its text, form data (key-value pairs and selection elements), and table data.
Sample document
Raw text
Forms
Tables
Human review

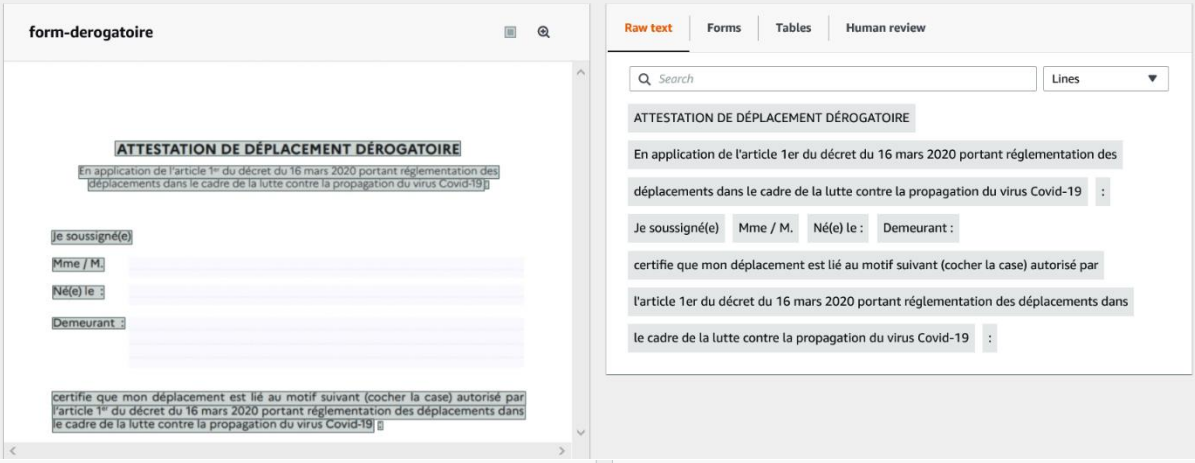

Analyze document Info
Download results
Upload document
Drag or upload a document to see its text, form data (key-value pairs and selection elements), and table data.
emp_app_printed
Employment Application
Applicant Information
Full Name: Jane Doe
Phone Number: 555-0100
Home Address: 123 Any Street, Any Town, USA
Mailing Address: same as home address
Raw text
Forms
Tables
Human review
Search
Full Name:
Jane Doe
Phone Number:
555-0100
Home Address:
123 Any Street, Any Town, USA
Mailing Address:
same as home address

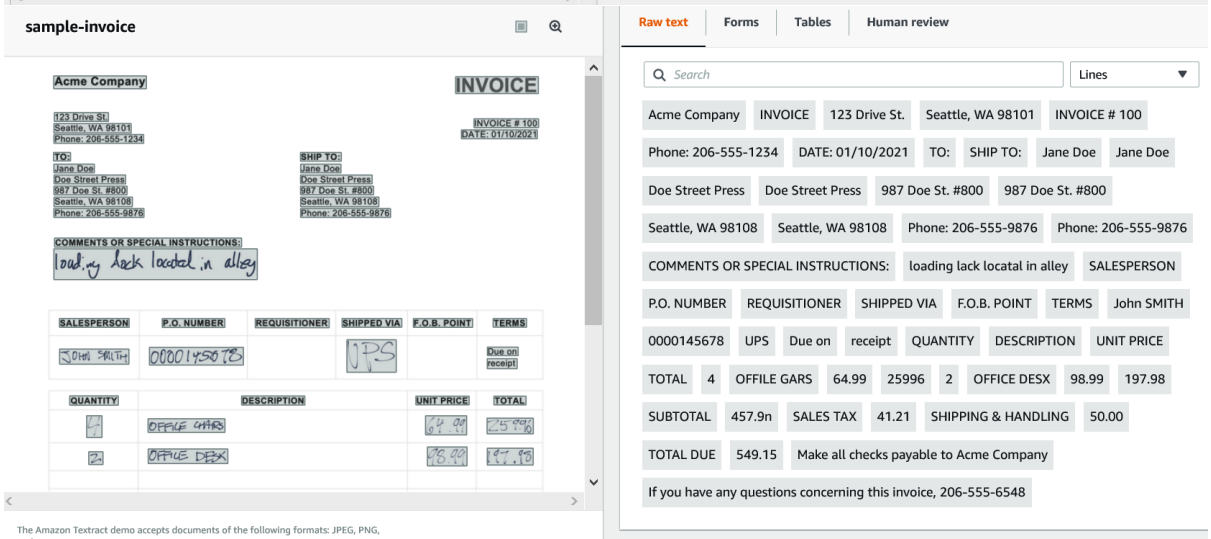


Analyze document Info

Download results | Upload document

Drag or upload a document to see its text, form data (key-value pairs and selection elements), and table data.





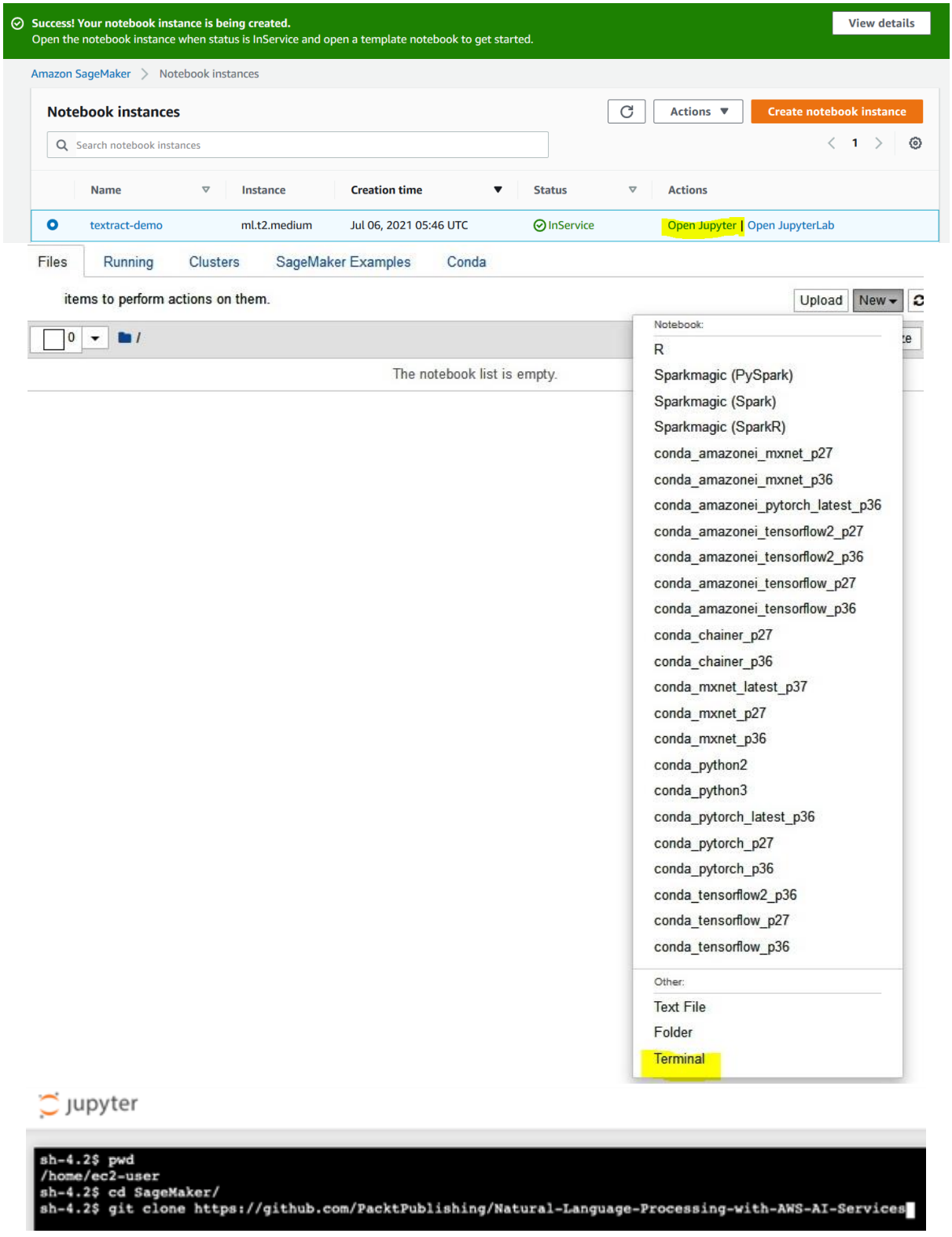
Analyze document Info

Download results | Upload document

Drag or upload a document to see its text, form data (key-value pairs and selection elements), and table data.

Sample document

Raw text | Forms | Tables | Human review


Success! Your notebook instance is being created.
Open the notebook instance when status is InService and open a template notebook to get started.
View details
Amazon SageMaker > Notebook instances
Notebook instances
Actions
Create notebook instance
Search notebook instances
Name
Instance
Creation time
Status
Actions
textract-demo
ml.t2.medium
Jul 06, 2021 05:46 UTC
InService
Open Jupyter
Open JupyterLab
Files
Running
Clusters
SageMaker Examples
Conda
items to perform actions on them.
Upload
New
The notebook list is empty.
Notebook:
R
Sparkmagic (PySpark)
Sparkmagic (Spark)
Sparkmagic (SparkR)
conda_amazonei_mxnet_p27
conda_amazonei_mxnet_p36
conda_amazonei_pytorch_latest_p36
conda_amazonei_tensorflow2_p27
conda_amazonei_tensorflow2_p36
conda_amazonei_tensorflow_p27
conda_amazonei_tensorflow_p36
conda_chainer_p27
conda_chainer_p36
conda_mxnet_latest_p37
conda_mxnet_p27
conda_mxnet_p36
conda_python2
conda_python3
conda_pytorch_latest_p36
conda_pytorch_p27
conda_pytorch_p36
conda_tensorflow2_p36
conda_tensorflow_p27
conda_tensorflow_p36
Other:
Text File
Folder
Terminal
jupyter
sh-4.2$ pwd
/home/ec2-user
sh-4.2$ cd SageMaker/
sh-4.2$ git clone https://github.com/PacktPublishing/Natural-Language-Processing-with-AWS-AI-Services

Acme Company

INVOICE

123 Drive St.
Seattle, WA 98101
Phone: 206-555-1234

INVOICE # 100
DATE: 01/10/2021

TO:
Jane Doe
Doe Street Press
987 Doe St. #800
Seattle, WA 98108
Phone: 206-555-9876

SHIP TO:
Jane Doe
Doe Street Press
987 Doe St. #800
Seattle, WA 98108
Phone: 206-555-9876

COMMENTS OR SPECIAL INSTRUCTIONS:
load. my dock located in alley

SALESPERSON	P.O. NUMBER	REQUISITIONER	SHIPPED VIA	F.O.B. POINT	TERMS
JOHN SMITH	0000145078		UPS		Due on receipt

QUANTITY	DESCRIPTION	UNIT PRICE	TOTAL
4	OFFICE CHAIRS	64.99	259.96
2	OFFICE DESK	98.99	197.98

SUBTOTAL	457.94
SALES TAX	41.21
SHIPPING & HANDLING	50.00
TOTAL DUE	549.15

Make all checks payable to Acme Company
If you have any questions concerning this invoice, 206-555-6548

Acme Company
123 Drive St.
Seattle, WA 98101
Phone: 206-555-1234
TO:
Jane Doe
Doe Street Press
987 Doe St. #800
Seattle, WA 98108
Phone: 206-555-9876
COMMENTS OR SPECIAL INSTRUCTIONS:
loading lack locatal in alley
SALESPERSON
John SMITH
QUANTITY
4
2
Make all checks payable to Acme Company
If you have any questions concerning this invoice, 206-555-6548
INVOICE
INVOICE # 100
DATE: 01/10/2021
TERMS
Due on
receipt

Company Name / Logo

Job Application

Street XYZ
City, St Zip ABC 0000

Phone 000-000-0000
Fax 000-000-0000

Personal Information

Last	First	MI	SSN#	Email
Jane	**Doe**		xxxxxxx	fake_email@mail.com

Street Address	City	ST	Zip	Home Phone	Mobile
1234 Main Street	ABC	XYZ	000000	555-88	

Are you entitled to work in the United States?	Are you 18 or older?	If yes, Date of Birth
Yes	yes	

Have you been convicted of a felony or been incarcerated in connection with a felony in the past seven years?	If yes, please explain:
no	no

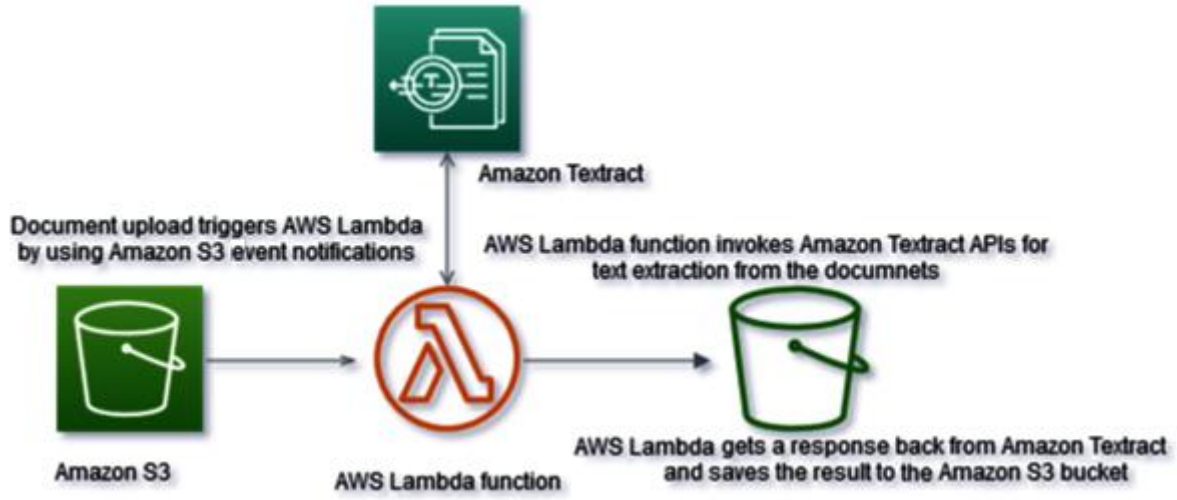
Military Service?	Branch	Are you a veteran?	War

What position are you applying for?	How did you hear about this position?

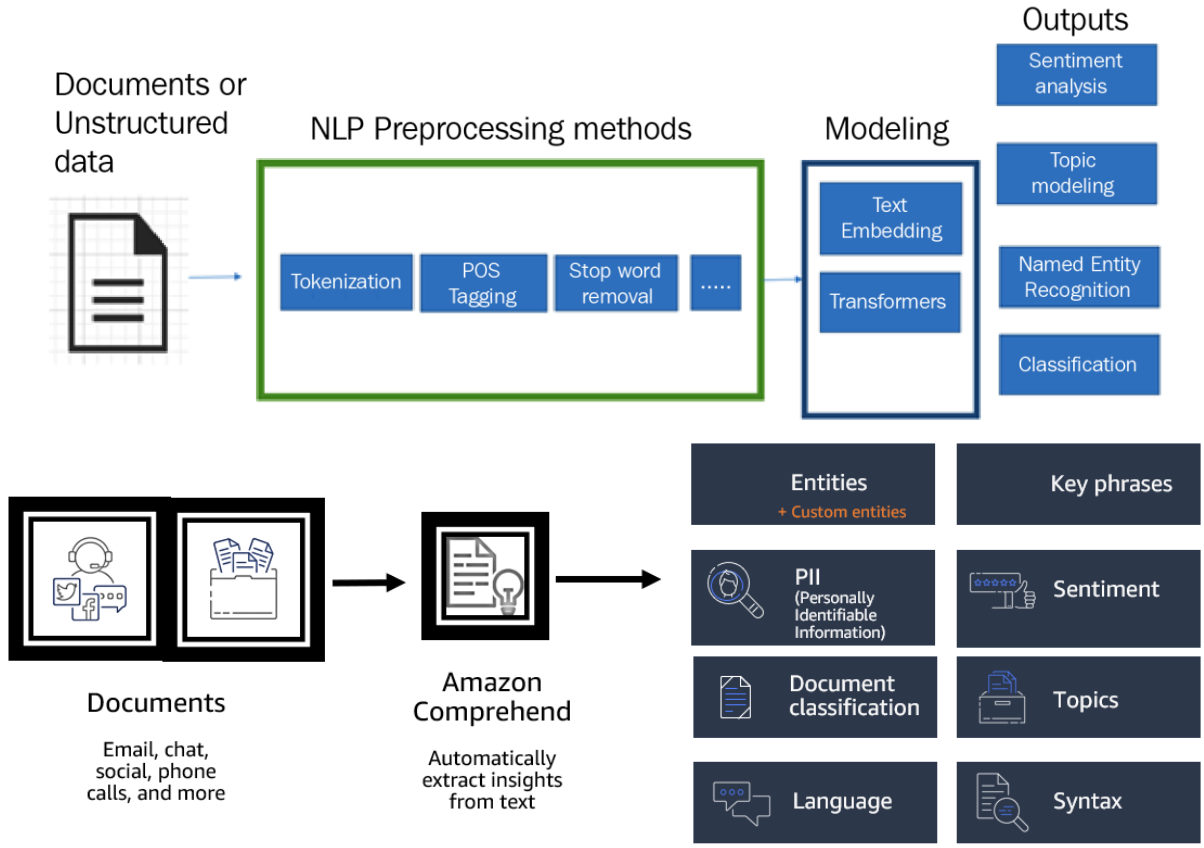
Expected Hourly Rate	Expected Weekly Earnings	Date Available

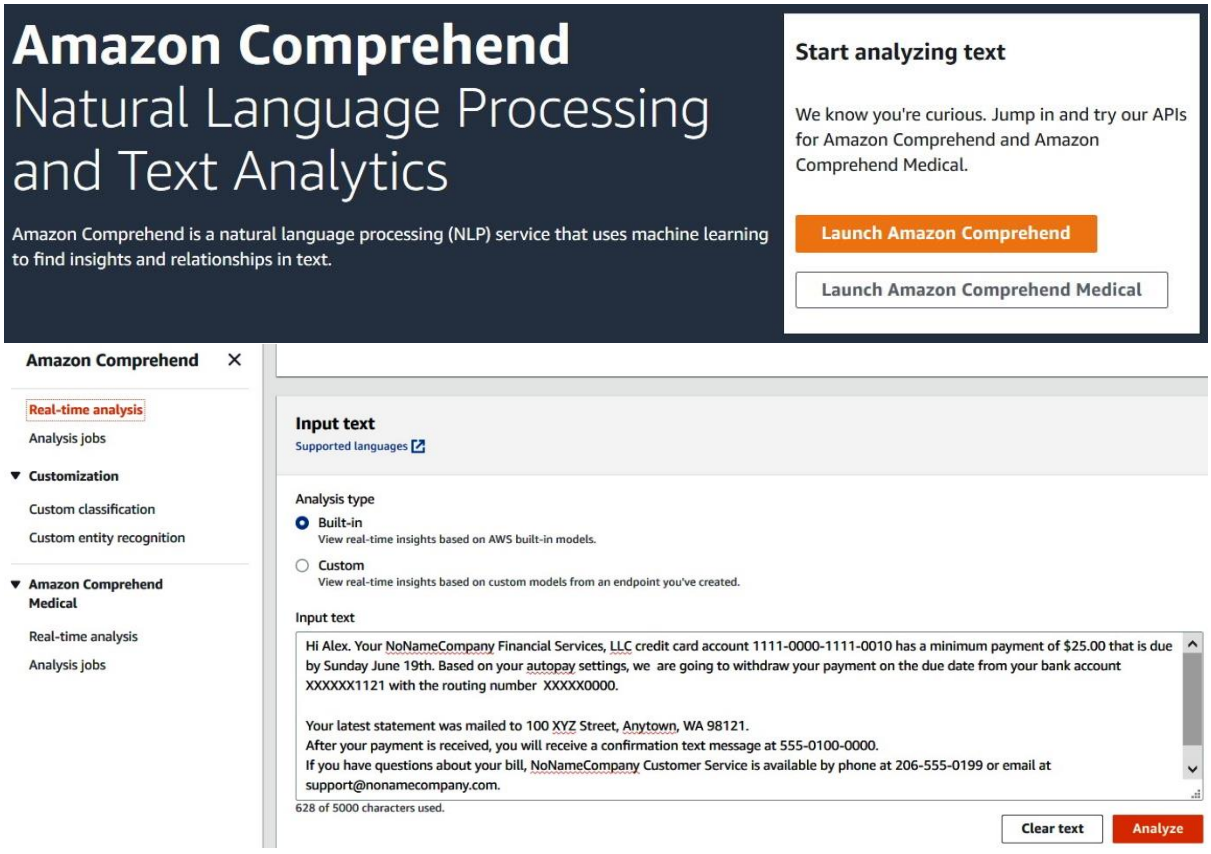
Prior Work Experience

	Current or Most Recent	Prior	Prior
Employer	ABC Corp		


Amazon Textract
Document upload triggers AWS Lambda by using Amazon S3 event notifications
AWS Lambda function invokes Amazon Textract APIs for text extraction from the documnets
Amazon S3
AWS Lambda function
AWS Lambda gets a response back from Amazon Textract and saves the result to the Amazon S3 bucket

Chapter 3: Introducing Amazon Comprehend





Insights Info

Entities | Key phrases | Language | PII | Sentiment | Syntax

Analyzed text

Hi Alex. Your NoNameCompany Financial Services, LLC credit card account 1111-0000-1111-0010 has a minimum payment of $25.00 that is due by Sunday June 19th. Based on your autopay settings, we are going to withdraw your payment on the due date from your bank account XXXXXX1121 with the routing number XXXXX0000.

Your latest statement was mailed to 100 XYZ Street, Anytown, WA 98121.
After your payment is received, you will receive a confirmation text message at 555-0100 (tel:2065550100)-0000.
If you have questions about your bill, NoNameCompany Customer Service is available by phone at 206-555-0199 or email at support@nonamecompany.com.

▼ **Results**

Search 1 2

Entity	Type	Confidence
Alex	Person	0.99+
NoNameCompany Financial Services, LLC	Organization	0.99+
1111-0000-1111-0010	Other	0.97
$25.00	Quantity	0.99+
Sunday June 19th	Date	0.99+
XXXXXX1121	Other	0.96
XXXXX0000	Other	0.89
100 XYZ Street, Anytown, WA 98121	Location	0.99+
555-0100-0000	Other	0.99+
NoNameCompany	Organization	0.93

▼ **Application integration**

API call and API response of DetectEntities API. Info

API call

```
{
    "Text": "Hi Alex. Your NoNameCompany Financial
        Services, LLC credit card account 1111-0000-1111
        -0010 has a minimum payment of $25.00 that is due
        by Sunday June 19th. Based on your autopay
        settings, we  are going to withdraw your payment
        on the due date from your bank account XXXXXX1121
        with the routing number  XXXXX0000.\n\nYour latest
        statement was mailed to 100 XYZ Street, Anytown,
        WA 98121.\nAfter your payment is received, you
        will receive a confirmation text message at 555
        -0100-0000.\nIf you have questions about your bill
        , NoNameCompany Customer Service is available by
        phone at 206-555-0199 or email at
        support@nonamecompany.com.\n",
    "LanguageCode": "en"
}
```

Copy

API response

```
{
    "Entities": [
        {
            "Score": 0.9991238713264465,
            "Type": "PERSON",
            "Text": "Alex",
            "BeginOffset": 3,
            "EndOffset": 7
        },
        {
            "Score": 0.9981185793876648,
            "Type": "ORGANIZATION",
            "Text": "NoNameCompany Financial Services,
                LLC",
            "BeginOffset": 14,
            "EndOffset": 51
        },
        {
            "Score": 0.9798830151557922,
            "Type": "OTHER",
            "Text": "1111-0000-1111-0010",
            "BeginOffset": 72,
            "EndOffset": 91
        },
        {
            "Score": 0.9965561032295227,
```

Copy

Insights Info

Entities | Key phrases | Language | PII | Sentiment | Syntax

Analyzed text

Hi Alex. Your NoNameCompany Financial Services, LLC credit card account 1111-0000-1111-0010 has a minimum payment of $25.00 that is due by Sunday June 19th. Based on your autopay settings, we are going to withdraw your payment on the due date from your bank account XXXXXX1121 with the routing number XXXXX0000.

Your latest statement was mailed to 100 XYZ Street, Anytown, WA 98121.
After your payment is received, you will receive a confirmation text message at 555-0100-0000.
If you have questions about your bill, NoNameCompany Customer Service is available by phone at 206-555-0199 or email at support@nonamecompany.com.

▼ **Results**

Search 1 2 3

Key phrases	Confidence
Hi Alex	0.99+
Your NoNameCompany Financial Services	0.99+
LLC credit card account 1111-0000-1111-0010	0.90

Insights Info

Entities | Key phrases | Language | PII | Sentiment | Syntax

Analyzed text

Hi Alex. Your NoNameCompany Financial Services, LLC credit card account 1111-0000-1111-0010 has a minimum payment of $25.00 that is due by Sunday June 19th. Based on your autopay settings, we are going to withdraw your payment on the due date from your bank account XXXXXX1121 with the routing number XXXXX0000.

Your latest statement was mailed to 100 XYZ Street, Anytown, WA 98121.
After your payment is received, you will receive a confirmation text message at 555-0100-0000.
If you have questions about your bill, NoNameCompany Customer Service is available by phone at 206-555-0199 or email at support@nonamecompany.com.

▼ **Results**

Language

English, en
0.99 confidence

Insights Info

Entities | Key phrases | Language | PII | Sentiment | Syntax

Personally identifiable information (PII) analysis mode

- (●) Offsets
  Identify the location of PII in your text documents.
- ( ) Labels
  Label text documents with PII.

Analyzed text

Hi Alex. Your NoNameCompany Financial Services, LLC credit card account 1111-0000-1111-0010 has a minimum payment of $25.00 that is due by Sunday June 19th. Based on your autopay settings, we are going to withdraw your payment on the due date from your bank account XXXXXX1121 with the routing number XXXXX0000.

Your latest statement was mailed to 100 XYZ Street, Anytown, WA 98121.
After your payment is received, you will receive a confirmation text message at 555-0100-0000.
If you have questions about your bill, NoNameCompany Customer Service is available by phone at 206-555-0199 or email at support@nonamecompany.com.

▼ **Results**

Search 1

Entity	Type	Confidence
Alex	Name	0.99+
1111-0000-1111-0010	Credit debit number	0.99+
Sunday June 19th	Date time	0.99+
XXXXXX1121	Bank account number	0.99+
XXXXX0000	Bank routing	0.99+
100 XYZ Street, Anytown, WA 98121	Address	0.99+
555-0100-0000	Phone	0.99+
206-555-0199	Phone	0.99+
support@nonamecompany.com	Email	0.99+

Insights Info

Entities | Key phrases | Language | PII | Sentiment | Syntax

Personally identifiable information (PII) analysis mode

- ○ Offsets
  Identify the location of PII in your text documents.
- ◉ Labels
  Label text documents with PII.

▼ Results

Search 1

Type	Confidence
Date time	1.00
Email	0.99+
Address	0.91
Bank routing	0.64
Phone	1.00

Insights Info

Entities | Key phrases | Language | PII | Sentiment | Syntax

Analyzed text

Hi Alex. Your NoNameCompany Financial Services, LLC credit card account 1111-0000-1111-0010 has a minimum payment of $25.00 that is due by Sunday June 19th. Based on your autopay settings, we are going to withdraw your payment on the due date from your bank account XXXXXX1121 with the routing number XXXXX0000.

Your latest statement was mailed to 100 XYZ Street, Anytown, WA 98121.
After your payment is received, you will receive a confirmation text message at 555-0100-0000.
If you have questions about your bill, NoNameCompany Customer Service is available by phone at 206-555-0199 or email at support@nonamecompany.com.

▼ Results

Sentiment

Neutral	Positive	Negative	Mixed
0.97 confidence	0.00 confidence	0.02 confidence	0.00 confidence

Insights Info

Entities | Key phrases | Language | PII | Sentiment | Syntax

Analyzed text

Hi Alex. Your NoNameCompany Financial Services, LLC credit card account 1111-0000-1111-0010 has a minimum payment of $25.00 that is due by Sunday June 19th. Based on your autopay settings, we are going to withdraw your payment on the due date from your bank account XXXXXX1121 with the routing number XXXXX0000.

Your latest statement was mailed to 100 XYZ Street, Anytown, WA 98121.
After your payment is received, you will receive a confirmation text message at 555-0100-0000.
If you have questions about your bill, NoNameCompany Customer Service is available by phone at 206-555-0199 or email at support@nonamecompany.com.

▼ **Results**

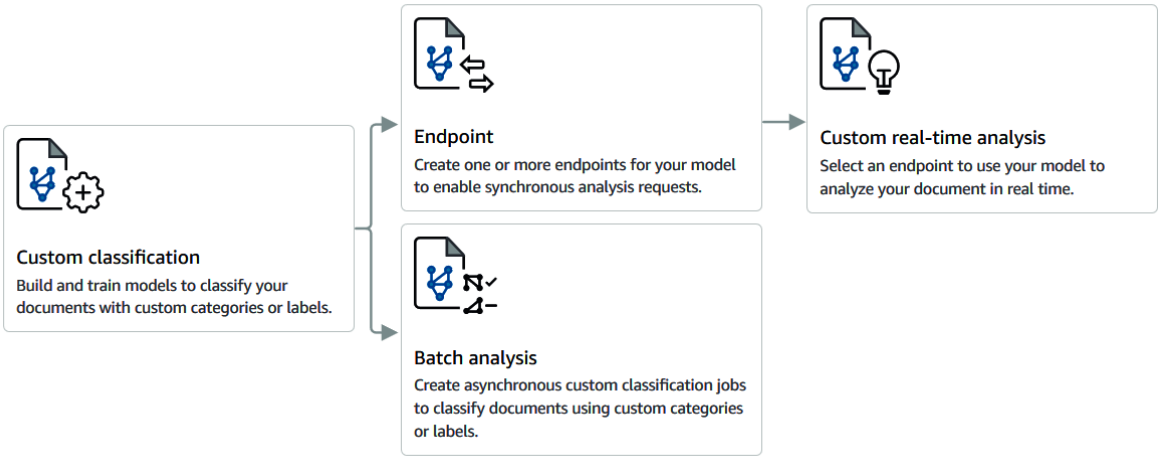
Search 1 2 3 4 5 6 7 ... 11

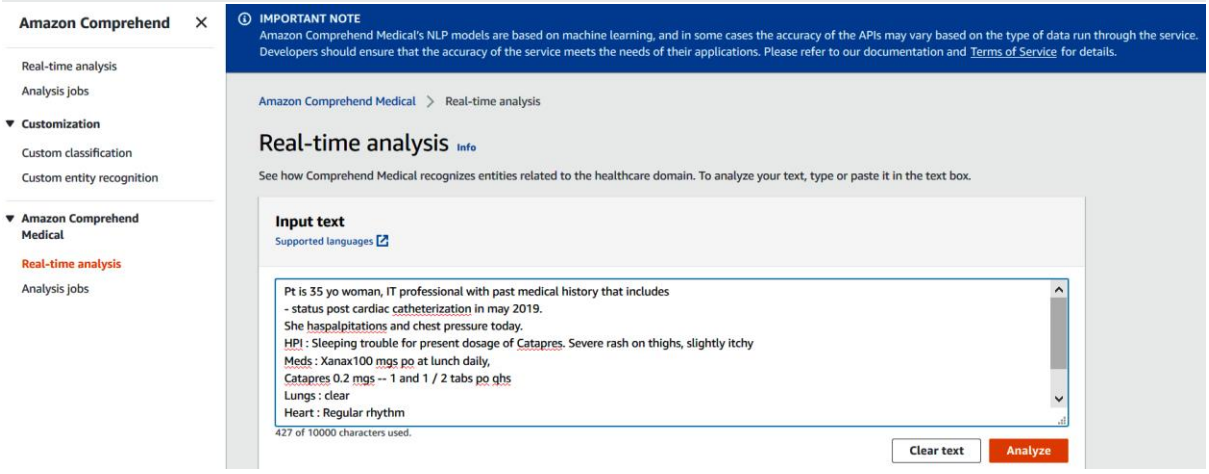
Word	Part of speech	Confidence
Hi	Interjection	0.98
Alex	Proper noun	0.98
.	Punctuation	0.99+

Custom classification Info

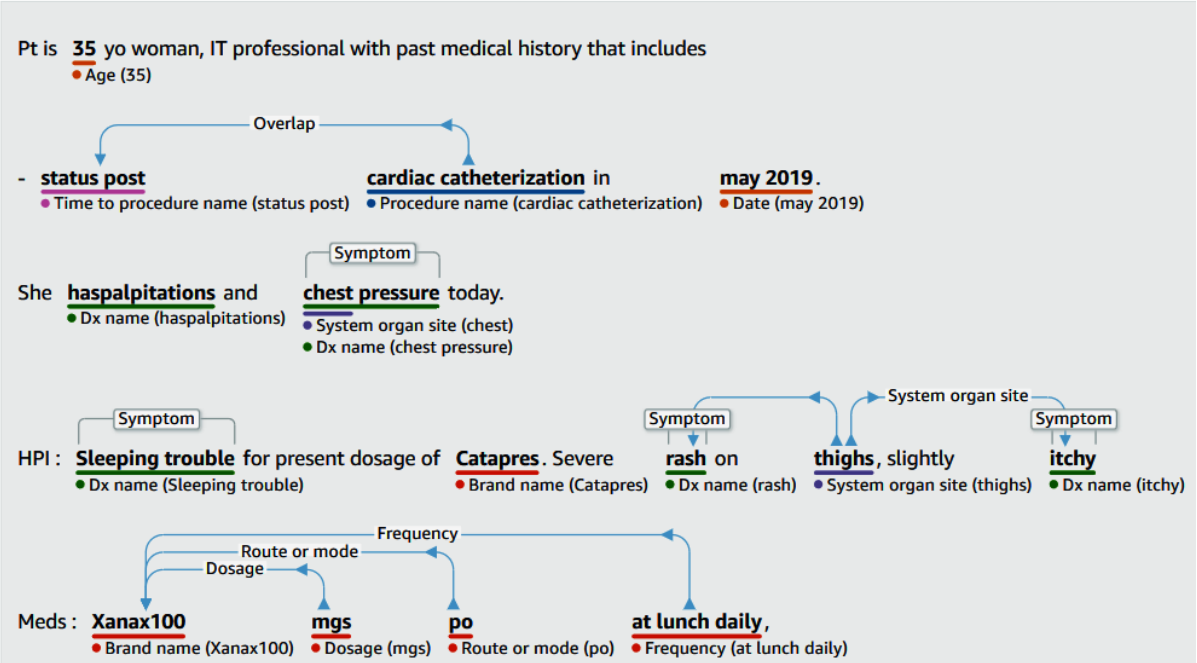
Classify documents using custom created categories.

▼ **Overview**





Analyzed text



Entity	Type	Category	Traits
35 0.9969 score	Age	Protected health information	-
status post 0.6577 score	Time to procedure name	Time expression	-
cardiac catheterization 0.9182 score	Procedure name	Test treatment procedure	-
may 2019 0.9997 score	Date	Protected health information	-
haspalpitations 0.4688 score	Dx name	Medical condition	-
chest 0.9425 score	System organ site	Anatomy	-
chest pressure 0.5286 score	Dx name	Medical condition	Symptom 0.5457 score
Sleeping trouble 0.8365 score	Dx name	Medical condition	Symptom 0.6399 score
Catapres 0.7436 score	Brand name	Medication	-
rash 0.9967 score	Dx name	Medical condition	Symptom 0.7731 score
thighs 0.9688 score	System organ site	Anatomy	-

Entities | RxNorm concepts | ICD-10-CM concepts

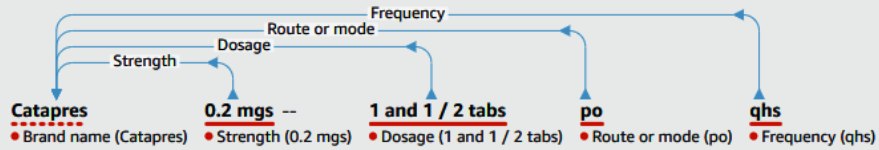
RxNorm source is the UMLS Metathesaurus 2017AB provided by NLM.

Analyzed text

Pt is 35 yo woman, IT professional with past medical history that includes

- status post cardiac catheterization in may 2019. She haspalpitations and chest pressure today. HPI : Sleeping trouble for present dosage of

Meds : Xanax100 mgs po at lunch daily,



Lungs : clear

Heart : Regular rhythm

Next follow up as scheduled on 06/19/2021

Entities | RxNorm concepts | ICD-10-CM concepts

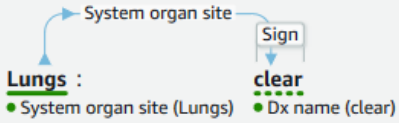
10-CM source is the 2019 Release of ICD-10 Clinical Modification provided by CDC.

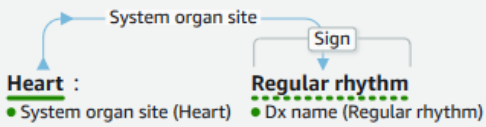
alyzed text

- status post cardiac catheterization in may 2019. She haspalpitations and chest pressure today.
Symptom: chest pressure — Dx name (chest pressure)

HPI : Sleeping trouble for present dosage of Catapres. Severe rash on thighs, slightly itchy
Symptom: Sleeping trouble — Dx name (Sleeping trouble)
Symptom: rash — Dx name (rash)
System organ site: thighs — System organ site (thighs)
Symptom: itchy — Dx name (itchy)

Meds : Xanax100 mgs po at lunch daily, Catapres 0.2 mgs -- 1 and 1 / 2 tabs po qhs





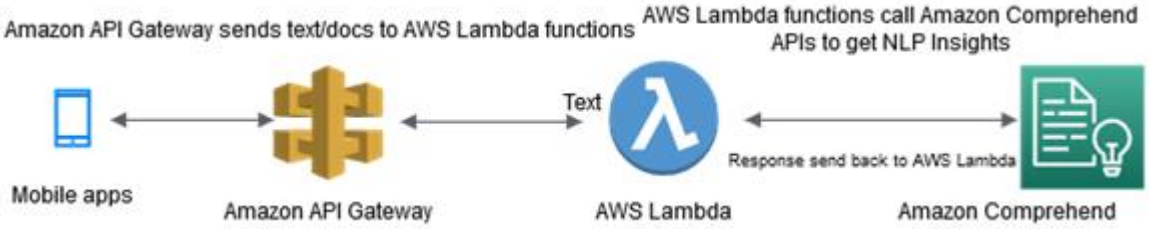
Next follow up as scheduled on 06/19/2021

```
{
    "Entities": [
        {
            "BeginOffset": 0,
            "EndOffset": 5,
            "Score": 0.995110809803009,
            "Text": "Packt",
            "Type": "ORGANIZATION"
        },
        {
            "BeginOffset": 41,
            "EndOffset": 45,
            "Score": 0.9951934218406677,
            "Text": "2003",
            "Type": "DATE"
        },
        {
            "BeginOffset": 63,
            "EndOffset": 77,
            "Score": 0.9813098311424255,
            "Text": "Birmingham, UK",
            "Type": "LOCATION"
        },
        {
            "BeginOffset": 95,
            "EndOffset": 108,
            "Score": 0.9864111542701721,
            "Text": "Mumbai, India",
            "Type": "LOCATION"
        },
        {
            "BeginOffset": 110,
            "EndOffset": 115,
            "Score": 0.9877771139144897,
            "Text": "Packt",
            "Type": "ORGANIZATION"
        }
```

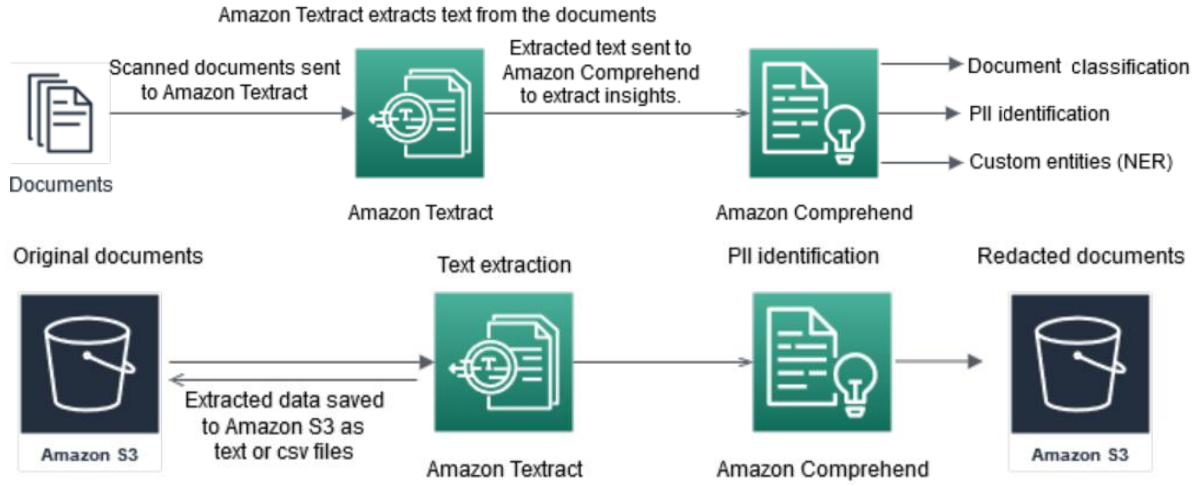
```
{
    "KeyPhrases": [
        {
            "BeginOffset": 0,
            "EndOffset": 5,
            "Score": 1.0,
            "Text": "Packt"
        },
        {
            "BeginOffset": 10,
            "EndOffset": 31,
            "Score": 0.9999998211860657,
            "Text": "une soci\u00e9t\u00e9 d'\u00e9dition"
        },
        {
            "BeginOffset": 42,
            "EndOffset": 46,
            "Score": 0.9999995231628418,
            "Text": "2003"
        },
        {
            "BeginOffset": 47,
            "EndOffset": 60,
            "Score": 0.9999982714653015,
            "Text": "dont le si\u00e8ge"
        },
        {
            "BeginOffset": 67,
            "EndOffset": 93,
            "Score": 0.9880826473236084,
            "Text": "Birmingham, au Royaume-Uni"
        },
        {
            "BeginOffset": 100,
            "EndOffset": 121,
            "Score": 0.8736910223960876,
            "Text": "des bureaux \u00e0 Mumbai,"
        },
        {
            "BeginOffset": 125,
            "EndOffset": 129,
            "Score": 0.9220616221427917,
            "Text": "Inde"
        },
```

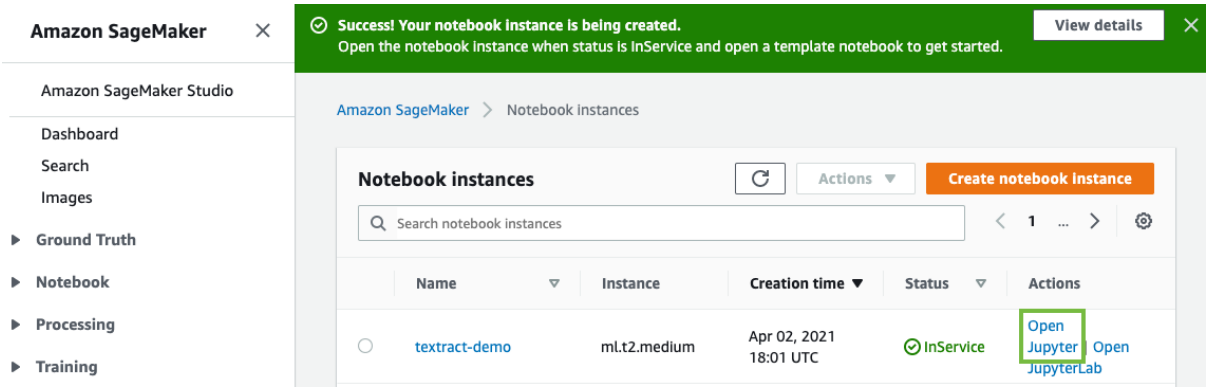
```
    },
    "ResultList": [
        {
            "Index": 0,
            "Sentiment": "POSITIVE",
            "SentimentScore": {
                "Mixed": 0.004781718365848064,
                "Negative": 0.00489041255787015,
                "Neutral": 0.028603093698620796,
                "Positive": 0.9617248177528381
            }
        },
        {
            "Index": 1,
            "Sentiment": "POSITIVE",
            "SentimentScore": {
                "Mixed": 0.043538015335798264,
                "Negative": 0.0020733894780278206,
                "Neutral": 0.005934751592576504,
                "Positive": 0.9484539031982422
            }
        },
        {
            "Index": 2,
            "Sentiment": "POSITIVE",
            "SentimentScore": {
                "Mixed": 0.006463758181780577,
                "Negative": 0.001142512890510261,
                "Neutral": 0.00479789637029171,
                "Positive": 0.9875959157943726
            }
        },
        {
            "Index": 3,
            "Sentiment": "NEUTRAL",
            "SentimentScore": {
                "Mixed": 0.00010260937415296212,
                "Negative": 0.004924369975924492,
                "Neutral": 0.6275114417076111,
                "Positive": 0.36746159195899963
            }
        }
    ]
}
```

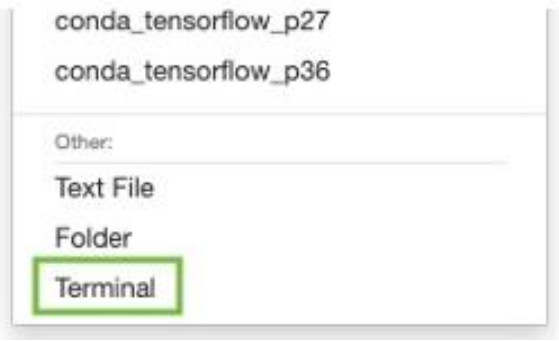
```
{
    "ErrorList": [],
    "ResponseMetadata": {
        "HTTPHeaders": {
            "content-length": "179",
            "content-type": "application/x-amz-json-1.1",
            "date": "Tue, 13 Apr 2021 16:03:59 GMT",
            "x-amzn-requestid": "88d1ffc8-13c3-4cde-9c2b-ba47aba8d360"
        },
        "HTTPStatusCode": 200,
        "RequestId": "88d1ffc8-13c3-4cde-9c2b-ba47aba8d360",
        "RetryAttempts": 0
    },
    "ResultList": [
        {
            "Index": 0,
            "Languages": [
                {
                    "LanguageCode": "en",
                    "Score": 0.9827539324760437
                }
            ]
        },
        {
            "Index": 1,
            "Languages": [
                {
                    "LanguageCode": "fr",
                    "Score": 0.99605792760849
                }
            ]
        }
    ]
}
```

Chapter 4: Automating Document Processing Workflows







jupyter

```
sh-4.2$ pwd
/home/ec2-user
sh-4.2$ cd SageMaker/
sh-4.2$ git clone https://github.com/PacktPublishing/Natural-Language-Processing-with-AWS-AI-Services
```

TemplateLAB

Bank of America

P.O. Box 15284
Wilmington, DE 19850

Customer service information

Customer service 1.800.432.1000
TDD/TTY users only: 1.800.288.4408
En Español: 1.800.688.6086
bankofamerica.com
Bank of America, N.A.
P.O. Box 25118
Tampa, FL 33622-5118

WILLIAM SMITH
3314 E KELTON LN UNIT 105
PHOENIX, AZ 85032-2762

Your BofA Core Checking

for December 20, 2016 to January 19, 2017

Account Number: 4570 3064 5567

WILLIAM SMITH

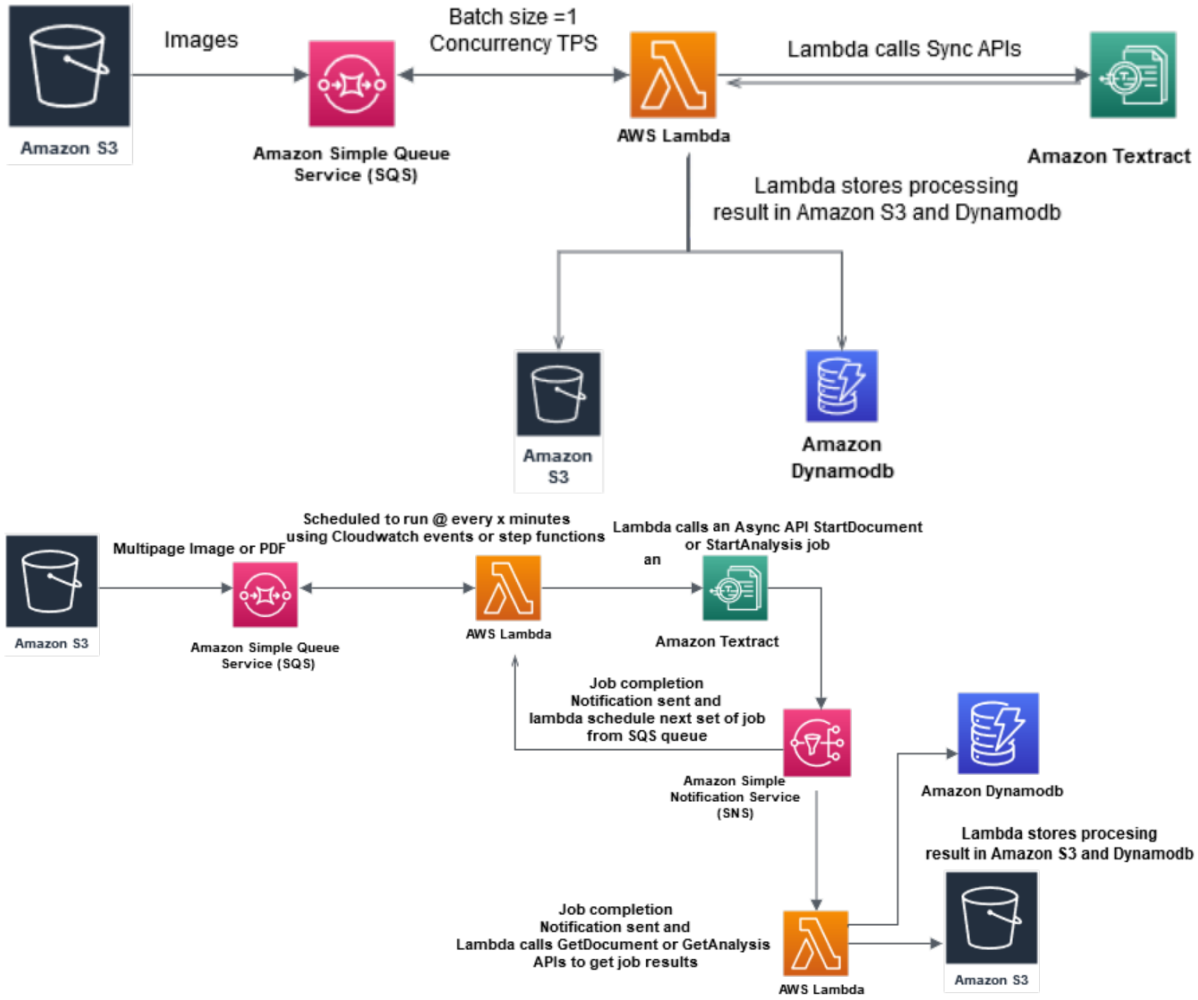
Account Summary

Beginning balance on December 20, 2016	$356.41
Deposits and other additions	3,921.69
ATM and debit card subtractions	-3,318.28
Other subtractions	-284.36
Checks	-85.00
Service fees	-12.00
Ending balance on January 19, 2017	**$578.46**

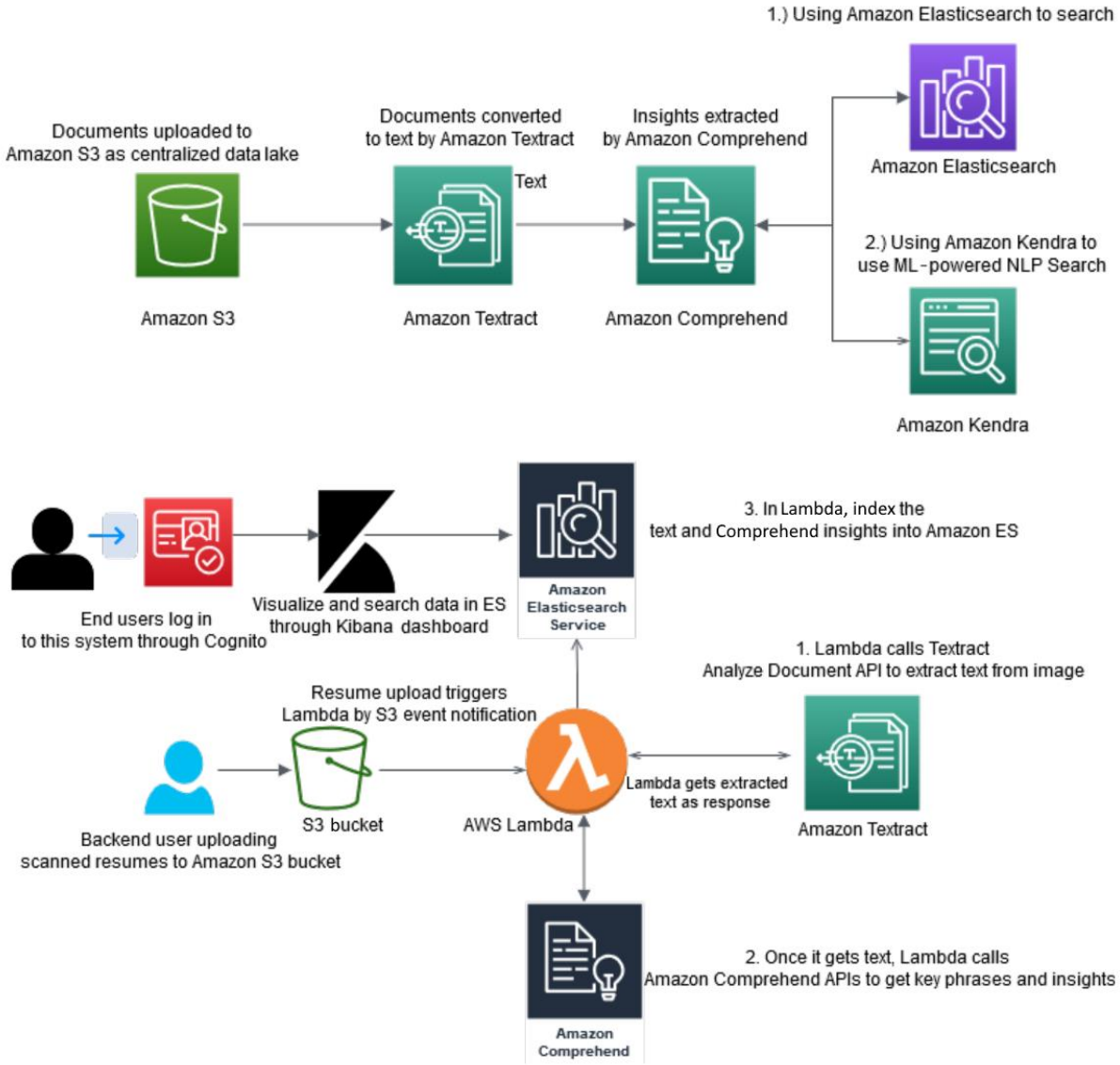
```
print(page_string)
```

```
CREATED S'TemplateLABBank of AmericaP.O. Box 15284Wilmington, DE 19850Customer service information1)Customer service
1.800.432.1000TDD/TTY users only: 1.800.288.4488WILLIAM SMITHEn Español: 1.800.688.60863314 E KELTON LN UNIT 105banko
famerica.comPHOENIX, AZ 85032-2762Bank of America, N.A.P.O. Box 25118Tampa, FL 33622-5118Your BofA Core Checkingfor D
ecember 20, 2016 to January 19, 2017Account Number: 457030645567WILLIAM SMITHAccount SummaryBeginning balance on Dece
mber 20, 2016$356.41Deposits and other additions3,921.69ATM and debit card subtractions-3,318.28Other subtractions-28
4.36Checks-85.00Service fees--12.00Ending balance on January 19, 2017$578.46Here's a tipDon't miss important account
notifications - keep your contact information updated.It's quick and easy to keep your phone number, email and mailin
g address up to date. Go to Profile &Settings and review your information. You'll help make sure you receive all of y
our notices and helpstay on top of your account.Is your contact info up to date? Check now in Online Banking at banko
famerica.com.
```

```
Finding PII text...
ADDRESS
detected as type 'ADDRESS' and will be redacted.
PHONE
detected as type 'PHONE' and will be redacted.
PHONE
detected as type 'PHONE' and will be redacted.
NAME
detected as type 'NAME' and will be redacted.
PHONE
detected as type 'PHONE' and will be redacted.
ADDRESS
detected as type 'ADDRESS' and will be redacted.
ADDRESS
detected as type 'ADDRESS' and will be redacted.
ADDRESS
detected as type 'ADDRESS' and will be redacted.
DATE_TIME
detected as type 'DATE_TIME' and will be redacted.
DATE_TIME
detected as type 'DATE_TIME' and will be redacted.
BANK_ACCOUNT_NUMBER
detected as type 'BANK_ACCOUNT_NUMBER' and will be redacted.
NAME
detected as type 'NAME' and will be redacted.
DATE_TIME
detected as type 'DATE_TIME' and will be redacted.
DATE_TIME
detected as type 'DATE_TIME' and will be redacted.
URL
detected as type 'URL' and will be redacted.
Found 15 text boxes to redact.
0 additional text boxes were detected, but did not meet the confidence score threshold.
s3://sagemaker-u[redacted]pii-detection-redaction/186389221476-PII-f9b911d0720328937333b5652876072d/output/pii_data.txt.out
pii-detection-redaction/[redacted]-PII-f9b911d0720328937333b5652876072d/output/pii_data.txt.out
b"CREATED S'TemplateLABBank of ********** 19850Customer service information1)Customer service **********TDD/TTY users only: **********WILLIAM ******* Espa\xc3\xb1ol: ********** ********** **********, ** 85032-2762Bank of America, ********** BofA Core Checkingfor ********** to ********** Number: ********** ********** SummaryBeginning balance on **********$356.41Deposits and other additions3,921.69ATM and debit card subtractions-3,318.28Other subtractions-284.36Checks-85.00Service fees--12.00Ending balance on **********$578.46Here's a tipDon't miss important account notifications - keep your contact information updated.It's quick and easy to keep your phone number, email and mailing address up to date. Go to Profile &Settings and review your information. You'll help make sure you receive all of your notices and helpstay on top of your account.Is your contact info up to date? Check now in Online Banking at **********.\n\n"
```


Amazon S3
Images
Amazon Simple Queue Service (SQS)
Batch size =1
Concurrency TPS
AWS Lambda
Lambda calls Sync APIs
Amazon Textract
Lambda stores processing result in Amazon S3 and Dynamodb
Amazon S3
Amazon Dynamodb
Amazon S3
Multipage Image or PDF
Amazon Simple Queue Service (SQS)
Scheduled to run @ every x minutes using Cloudwatch events or step functions
AWS Lambda
Lambda calls an Async API StartDocument or StartAnalysis job
an
Amazon Textract
Job completion Notification sent and lambda schedule next set of job from SQS queue
Amazon Simple Notification Service (SNS)
Amazon Dynamodb
Lambda stores procesing result in Amazon S3 and Dynamodb
Job completion Notification sent and Lambda calls GetDocument or GetAnalysis APIs to get job results
AWS Lambda
Amazon S3

Chapter 5: Creating NLP Search



CloudFormation > Stacks > Create stack

Quick create stack

Template

Template URL

https://[redacted]emplate-export-textract.yml

Stack description

-

Stack name

Stack name

document-search

Stack name can include letters (A-Z and a-z), numbers (0-9), and dashes (-).

Parameters

Parameters are defined in your template and allow you to input custom values when you create or update a stack.

CognitoAdminEmail

E-mail address of the Cognito admin name

[redacted]

DOMAINNAME

Name for the Amazon ES domain that this template will create. Domain names must start with a lowercase letter and must be between 3 and 28 characters. Valid characters are a-z (lowercase only), 0-9.

documentsearchapp

Capabilities and transforms

Transforms might require access capabilities

A transform might add Identity and Access Management (IAM) resources that could provide entities access to make changes to your AWS account. If a transform adds IAM resources, you must acknowledge their capabilities to create or update them. Ensure that you want to create or update the IAM resources, and that they have the minimum required permissions. In addition, if they have custom names, check that the names are unique within your AWS account. Learn more

- [x] **I acknowledge that AWS CloudFormation might create IAM resources.**
- [x] **I acknowledge that AWS CloudFormation might create IAM resources with custom names.**
- [x] **I acknowledge that AWS CloudFormation might require the following capability: CAPABILITY_AUTO_EXPAND**

Cancel | Create change set | Create stack

document-search

Delete | Update | Stack actions | Create stack

Stack info | Events | Resources | Outputs | Parameters | Template | Change sets

Overview

Stack ID
arn:aws:cloudformation:us-east-1:791176201651:stack/document-search/de724e60-b1a7-11eb-a123-0a6046b66d41

Description
-

Status
CREATE_COMPLETE

Status reason
-

document-search

Delete | Update | Stack actions | Create stack

Stack info | Events | Resources | Outputs | Parameters | Template | Change sets

Outputs (2)

Search outputs

Key	Value	Description	Export name
KibanaLoginURL	https://search-documentsearchapp-is2d57tgxywbn5w47viw6vly6e.us-east-1.es.amazonaws.com/_plugin/kibana/	Kibana login URL	-
S3KeyPhraseBucket	https://console.aws.amazon.com/s3/buckets/document-search-s3-mizd51is2yuc/?region=us-east-1	-	-

Amazon S3 > document-search-s3-mizd51is2yuc

document-search-s3-mizd51is2yuc

Objects | Properties | Permissions | Metrics | Management | Access Points

Objects (1)

Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more

Copy URL | Open | Download | Delete | Actions | Create folder | Upload

Find objects by prefix

1

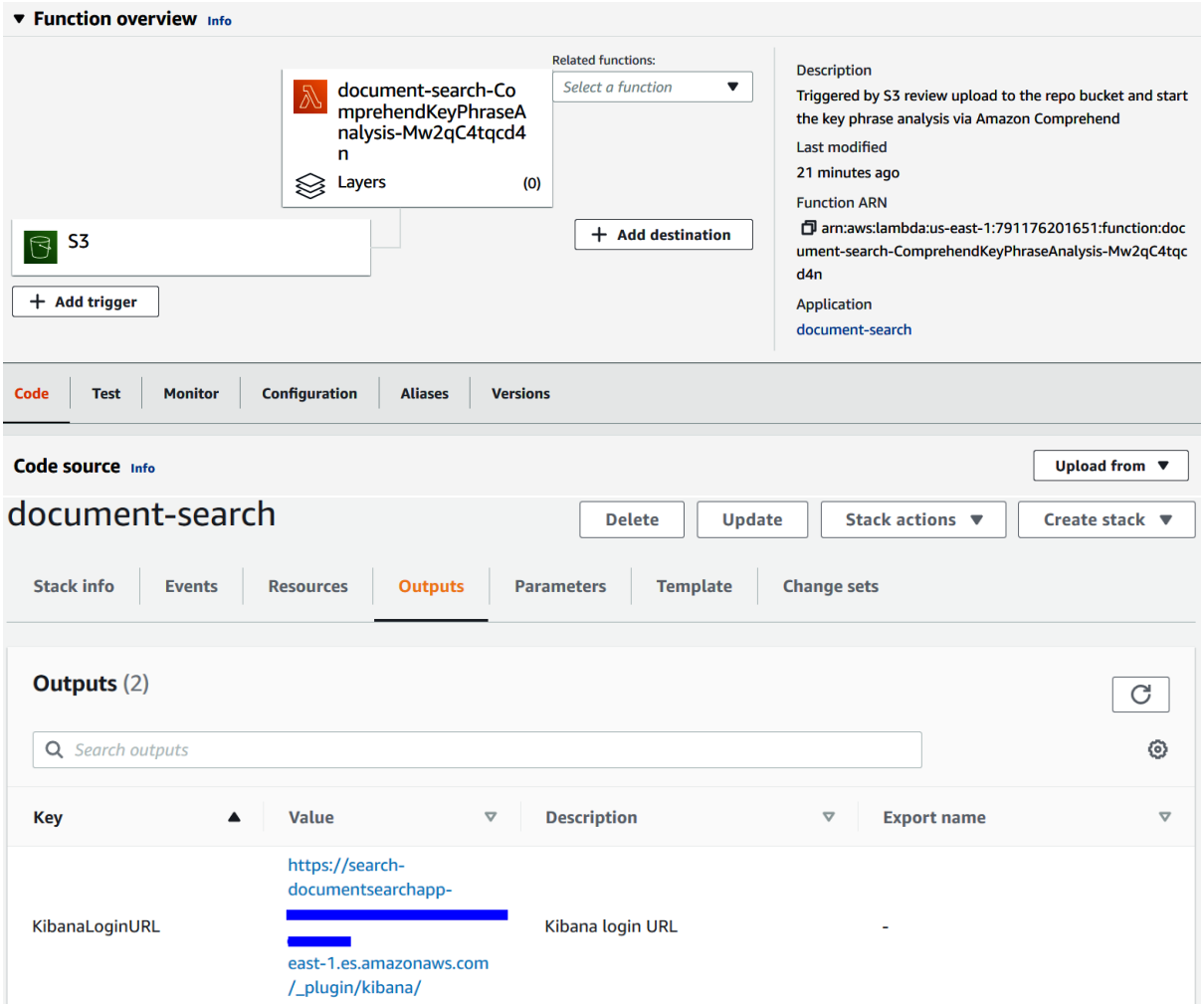
Name	Type	Last modified	Size	Storage class
resume_sample.PNG	PNG	May 10, 2021, 12:28:22 (UTC-04:00)	59.7 KB	Standard

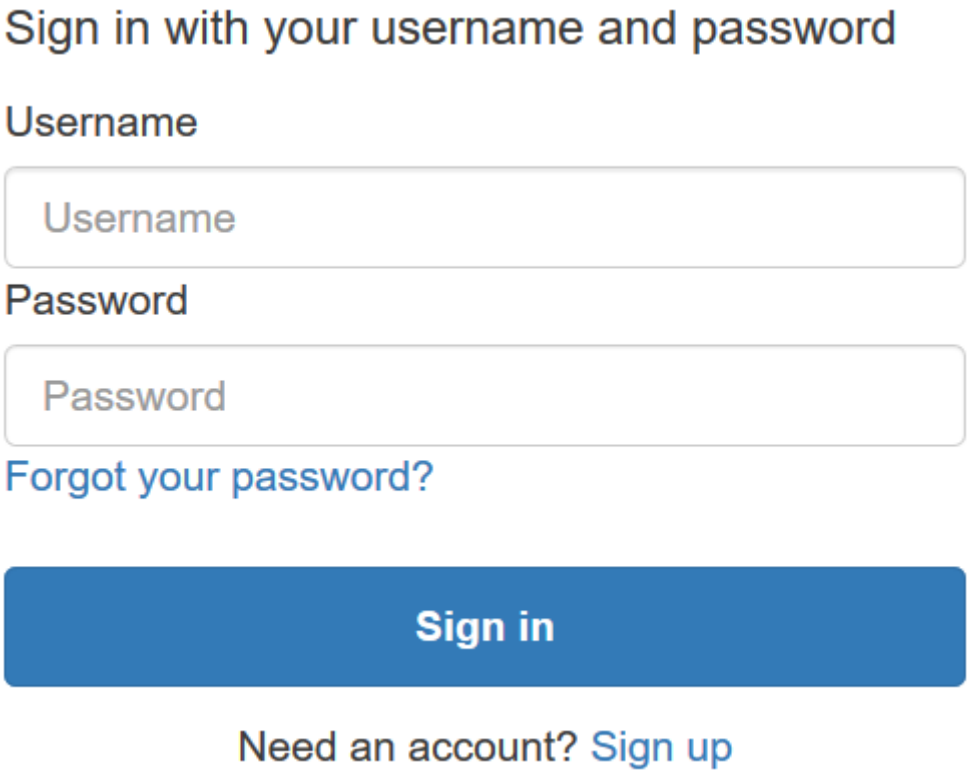
Event notifications (1)

Send a notification when specific events occur in your bucket. Learn more

Edit | Delete | Create event notification

Name	Event types	Filters	Destination type	Destination
S3ObjectCreatedEvent	All object create events	-	Lambda function	document-search-ComprehendKeyPhraseAnalysis-Mw2qC4tqcd4n


Function overview Info
Related functions:
Select a function
document-search-ComprehendKeyPhraseAnalysis-Mw2qC4tqcd4n
Layers (0)
S3
Add trigger
Add destination
Description
Triggered by S3 review upload to the repo bucket and start the key phrase analysis via Amazon Comprehend
Last modified
21 minutes ago
Function ARN
arn:aws:lambda:us-east-1:791176201651:function:document-search-ComprehendKeyPhraseAnalysis-Mw2qC4tqcd4n
Application
document-search
Code
Test
Monitor
Configuration
Aliases
Versions
Code source Info
Upload from
document-search
Delete
Update
Stack actions
Create stack
Stack info
Events
Resources
Outputs
Parameters
Template
Change sets
Outputs (2)
Search outputs
Key
Value
Description
Export name
KibanaLoginURL
https://search-documentsearchapp-
east-1.es.amazonaws.com/_plugin/kibana/
Kibana login URL
-


Sign in with your username and password
Username
Username
Password
Password
Forgot your password?
Sign in
Need an account? Sign up

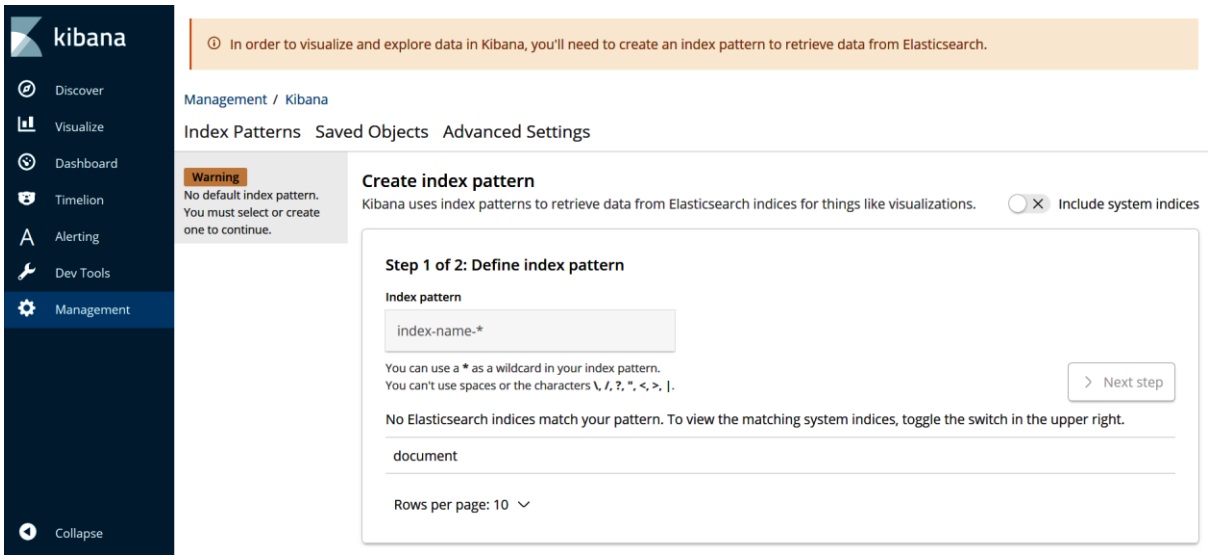
Reply Reply All Forward

no-reply@verificationemail.com [redacted] 11:54 AM

[EXTERNAL] Your temporary password

CAUTION: This email originated from outside of the organization. Do not click links or open attachments unless you can confirm the sender and know the content is safe.

Your username is [redacted] and temporary password is [redacted]



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

document*

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

Next step

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

document

Rows per page: 10

Create index pattern

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

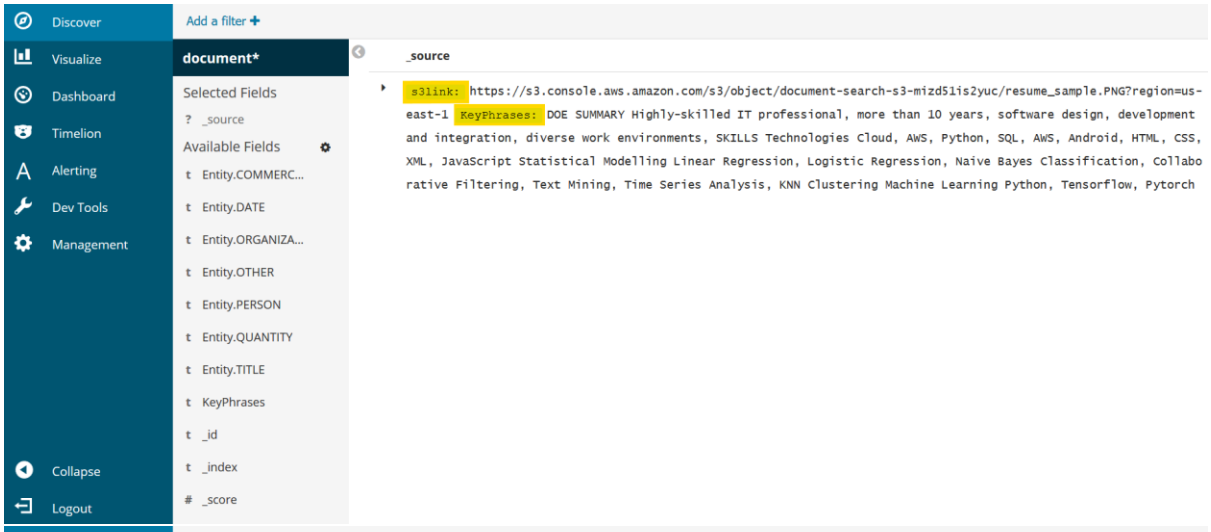
Step 2 of 2: Configure settings

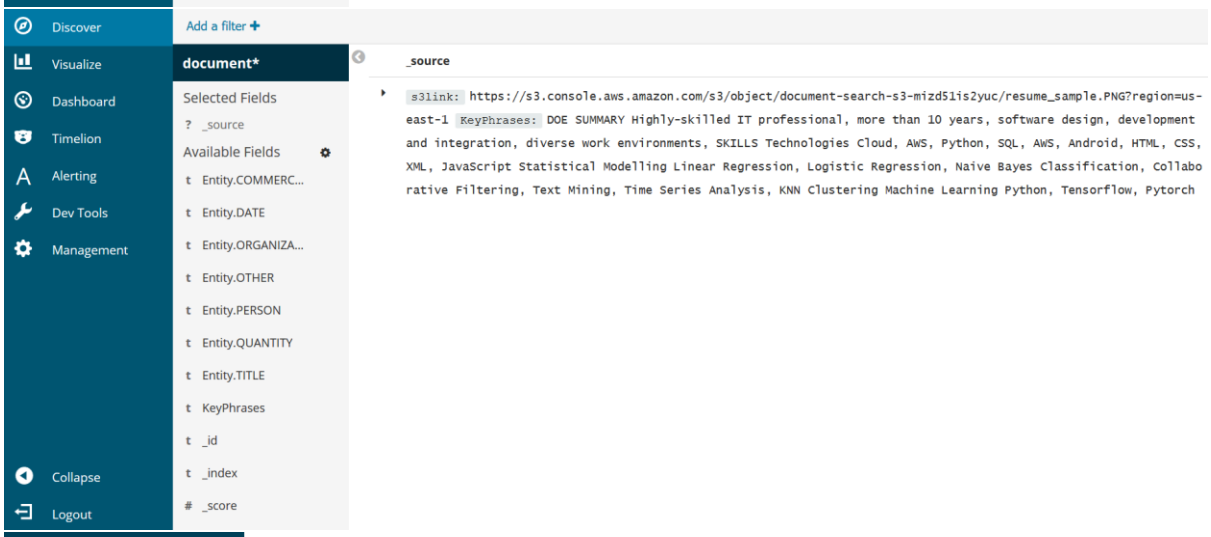
You've defined **document*** as your index pattern. Now you can specify some settings before we create it.

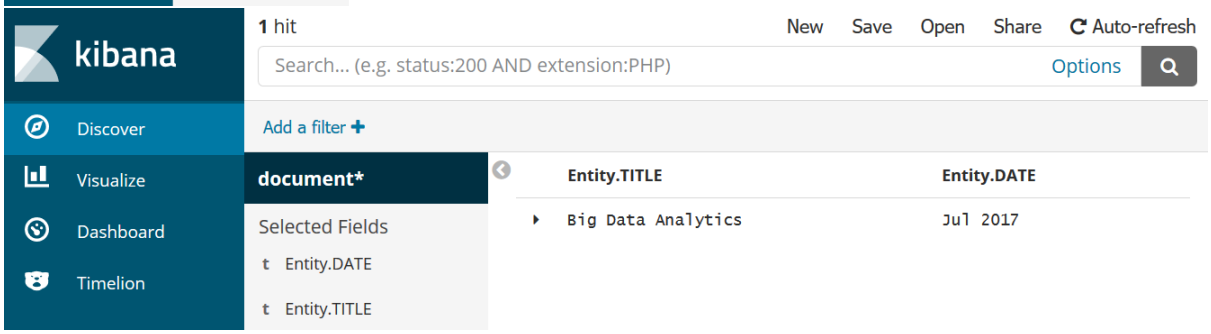
The indices which match this index pattern don't contain any time fields.

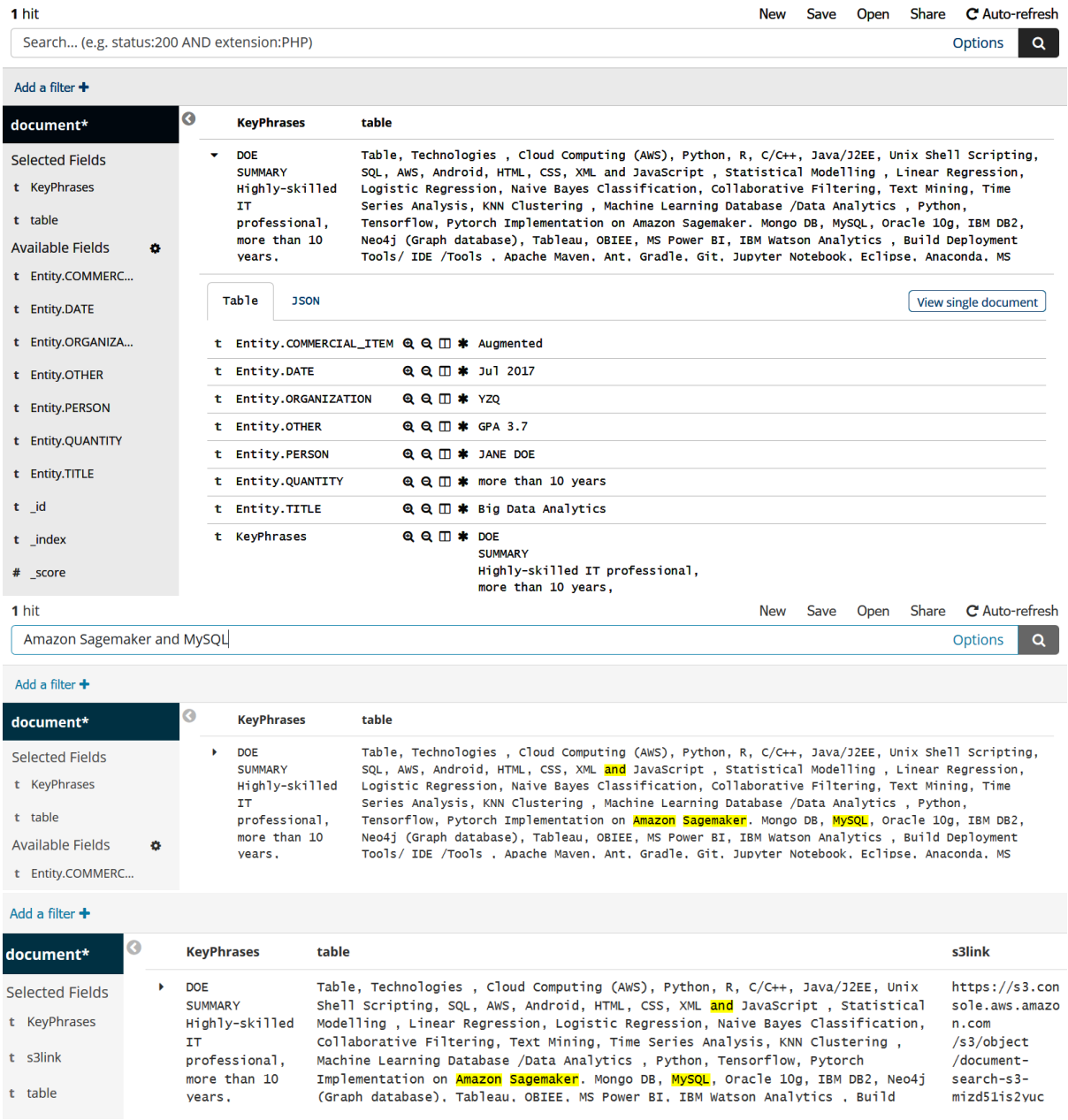
Show advanced options

Back Create index pattern


Discover
Visualize
Dashboard
Timelion
Alerting
Dev Tools
Management
Collapse
Logout
Add a filter +
document*
Selected Fields
? _source
Available Fields
t Entity.COMMERC...
t Entity.DATE
t Entity.ORGANIZA...
t Entity.OTHER
t Entity.PERSON
t Entity.QUANTITY
t Entity.TITLE
t KeyPhrases
t _id
t _index
_score
_source
s3link: https://s3.console.aws.amazon.com/s3/object/document-search-s3-mizd51is2yuc/resume_sample.PNG?region=us-east-1 KeyPhrases: DOE SUMMARY Highly-skilled IT professional, more than 10 years, software design, development and integration, diverse work environments, SKILLS Technologies Cloud, AWS, Python, SQL, AWS, Android, HTML, CSS, XML, JavaScript Statistical Modelling Linear Regression, Logistic Regression, Naive Bayes Classification, Collaborative Filtering, Text Mining, Time Series Analysis, KNN Clustering Machine Learning Python, Tensorflow, Pytorch


Discover
Visualize
Dashboard
Timelion
Alerting
Dev Tools
Management
Collapse
Logout
Add a filter +
document*
Selected Fields
? _source
Available Fields
t Entity.COMMERC...
t Entity.DATE
t Entity.ORGANIZA...
t Entity.OTHER
t Entity.PERSON
t Entity.QUANTITY
t Entity.TITLE
t KeyPhrases
t _id
t _index
_score
_source
s3link: https://s3.console.aws.amazon.com/s3/object/document-search-s3-mizd51is2yuc/resume_sample.PNG?region=us-east-1 KeyPhrases: DOE SUMMARY Highly-skilled IT professional, more than 10 years, software design, development and integration, diverse work environments, SKILLS Technologies Cloud, AWS, Python, SQL, AWS, Android, HTML, CSS, XML, JavaScript Statistical Modelling Linear Regression, Logistic Regression, Naive Bayes Classification, Collaborative Filtering, Text Mining, Time Series Analysis, KNN Clustering Machine Learning Python, Tensorflow, Pytorch


kibana
1 hit
New
Save
Open
Share
Auto-refresh
Search... (e.g. status:200 AND extension:PHP)
Options
Discover
Visualize
Dashboard
Timelion
Add a filter +
document*
Selected Fields
t Entity.DATE
t Entity.TITLE
Entity.TITLE
Entity.DATE
Big Data Analytics
Jul 2017


1 hit
New Save Open Share Auto-refresh
Search... (e.g. status:200 AND extension:PHP)
Options
Add a filter +
document*
Selected Fields
t KeyPhrases
t table
Available Fields
t Entity.COMMERC...
t Entity.DATE
t Entity.ORGANIZA...
t Entity.OTHER
t Entity.PERSON
t Entity.QUANTITY
t Entity.TITLE
t _id
t _index
_score
KeyPhrases
table
DOE SUMMARY Highly-skilled IT professional, more than 10 years.
Table, Technologies , Cloud Computing (AWS), Python, R, C/C++, Java/J2EE, Unix Shell Scripting, SQL, AWS, Android, HTML, CSS, XML and JavaScript , Statistical Modelling , Linear Regression, Logistic Regression, Naive Bayes Classification, Collaborative Filtering, Text Mining, Time Series Analysis, KNN Clustering , Machine Learning Database /Data Analytics , Python, Tensorflow, Pytorch Implementation on Amazon Sagemaker. Mongo DB, MySQL, Oracle 10g, IBM DB2, Neo4j (Graph database), Tableau, OBIEE, MS Power BI, IBM Watson Analytics , Build Deployment Tools/ IDE /Tools , Apache Maven, Ant, Gradle, Git, Jupyter Notebook, Eclipse, Anaconda, MS
Table
JSON
View single document
t Entity.COMMERCIAL_ITEM Augmented
t Entity.DATE Jul 2017
t Entity.ORGANIZATION YZQ
t Entity.OTHER GPA 3.7
t Entity.PERSON JANE DOE
t Entity.QUANTITY more than 10 years
t Entity.TITLE Big Data Analytics
t KeyPhrases DOE SUMMARY Highly-skilled IT professional, more than 10 years,
1 hit
New Save Open Share Auto-refresh
Amazon Sagemaker and MySQL
Options
Add a filter +
document*
Selected Fields
t KeyPhrases
t table
Available Fields
t Entity.COMMERC...
KeyPhrases
table
DOE SUMMARY Highly-skilled IT professional, more than 10 years.
Table, Technologies , Cloud Computing (AWS), Python, R, C/C++, Java/J2EE, Unix Shell Scripting, SQL, AWS, Android, HTML, CSS, XML and JavaScript , Statistical Modelling , Linear Regression, Logistic Regression, Naive Bayes Classification, Collaborative Filtering, Text Mining, Time Series Analysis, KNN Clustering , Machine Learning Database /Data Analytics , Python, Tensorflow, Pytorch Implementation on Amazon Sagemaker. Mongo DB, MySQL, Oracle 10g, IBM DB2, Neo4j (Graph database), Tableau, OBIEE, MS Power BI, IBM Watson Analytics , Build Deployment Tools/ IDE /Tools , Apache Maven, Ant, Gradle, Git, Jupyter Notebook, Eclipse, Anaconda, MS
Add a filter +
document*
Selected Fields
t KeyPhrases
t s3link
t table
KeyPhrases
table
s3link
DOE SUMMARY Highly-skilled IT professional, more than 10 years.
Table, Technologies , Cloud Computing (AWS), Python, R, C/C++, Java/J2EE, Unix Shell Scripting, SQL, AWS, Android, HTML, CSS, XML and JavaScript , Statistical Modelling , Linear Regression, Logistic Regression, Naive Bayes Classification, Collaborative Filtering, Text Mining, Time Series Analysis, KNN Clustering , Machine Learning Database /Data Analytics , Python, Tensorflow, Pytorch Implementation on Amazon Sagemaker. Mongo DB, MySQL, Oracle 10g, IBM DB2, Neo4j (Graph database), Tableau, OBIEE, MS Power BI, IBM Watson Analytics , Build
https://s3.console.aws.amazon.com/s3/object/document-search-s3-mizd51is2yuc

Amazon S3 > Create bucket

Create bucket

Buckets are containers for data stored in S3. Learn more

General configuration

Bucket name

myawsbucket

Bucket name must be unique and must not contain spaces or uppercase letters. **See rules for bucket naming**

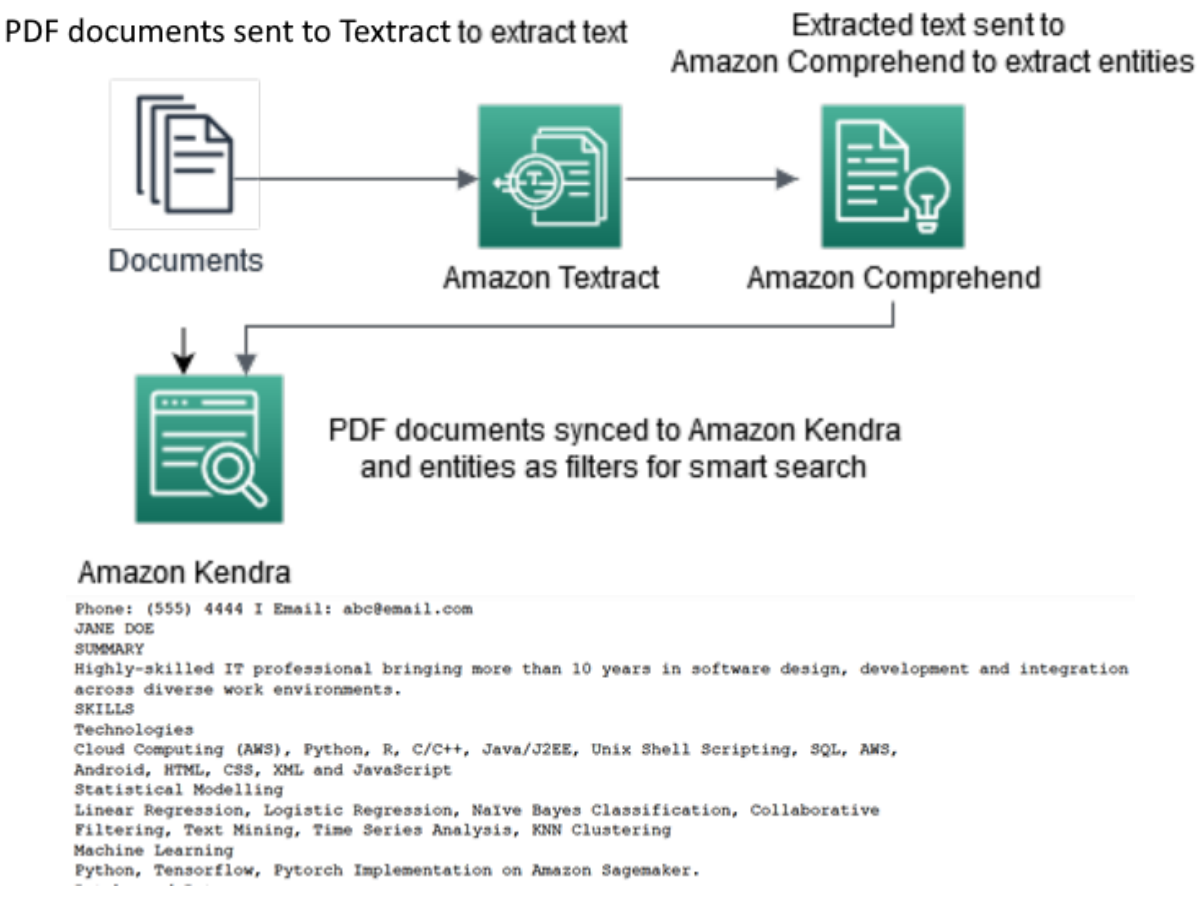
AWS Region

US East (N. Virginia) us-east-1

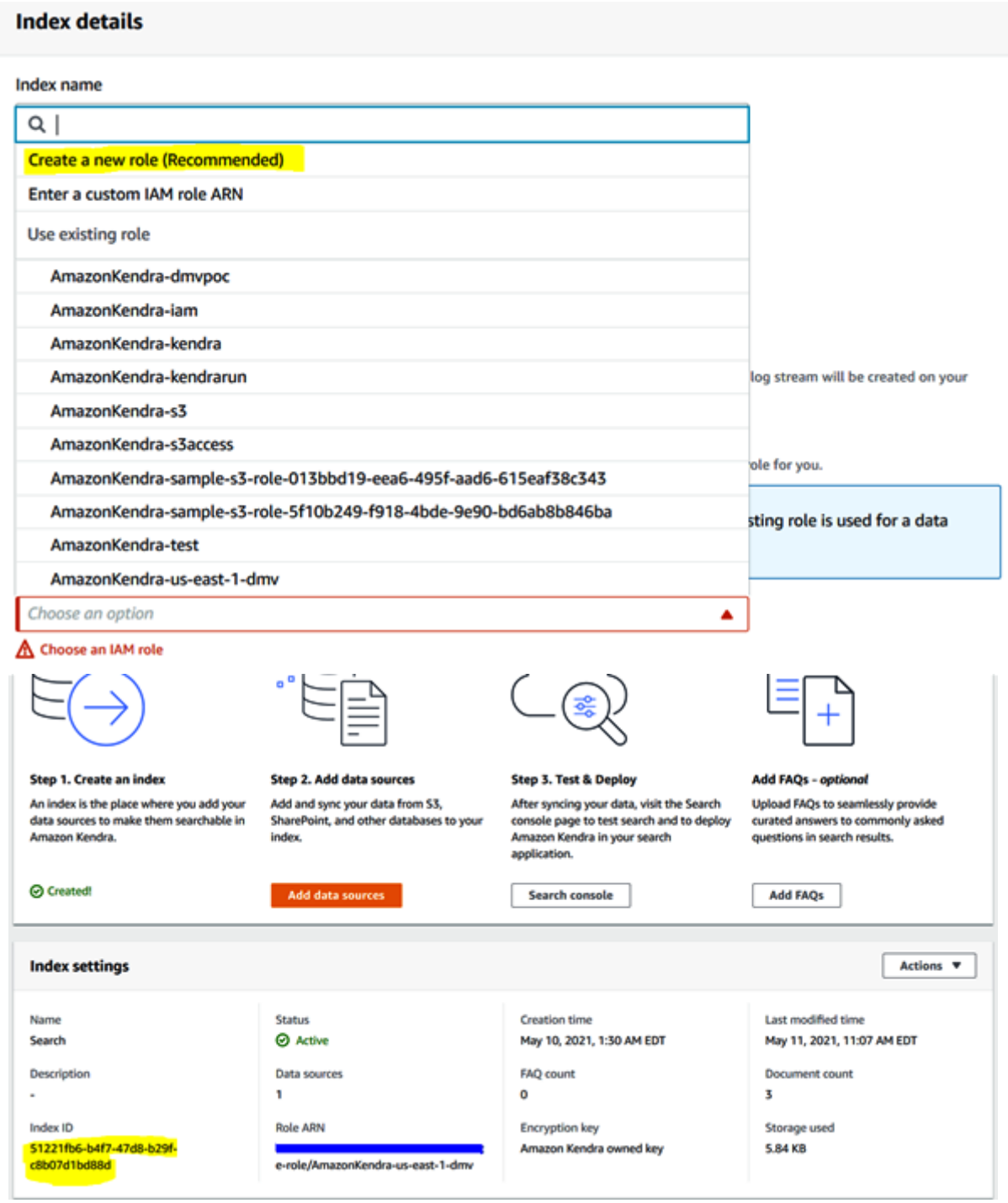
Copy settings from existing bucket - *optional*

Only the bucket settings in the following configuration are copied.

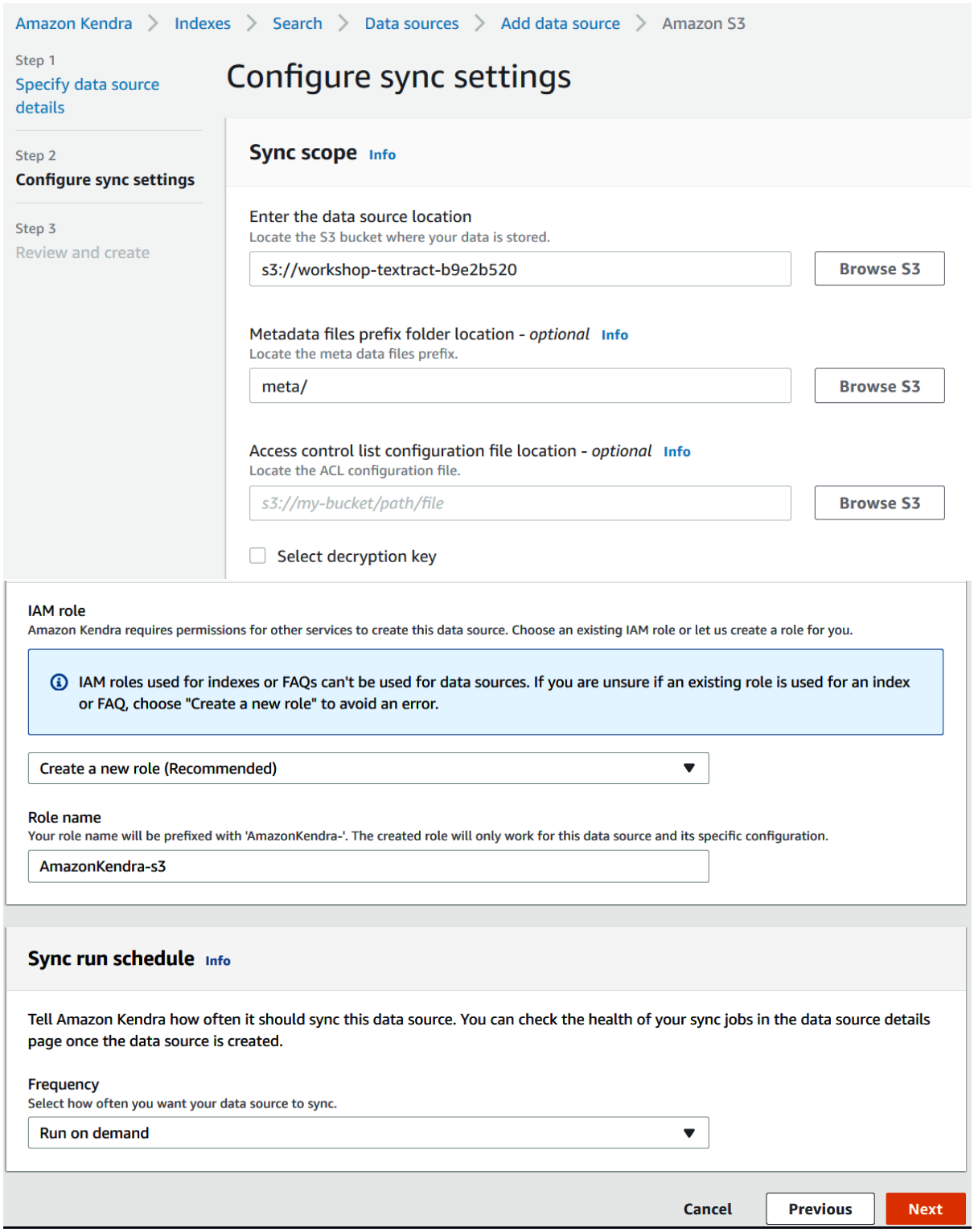
Choose bucket

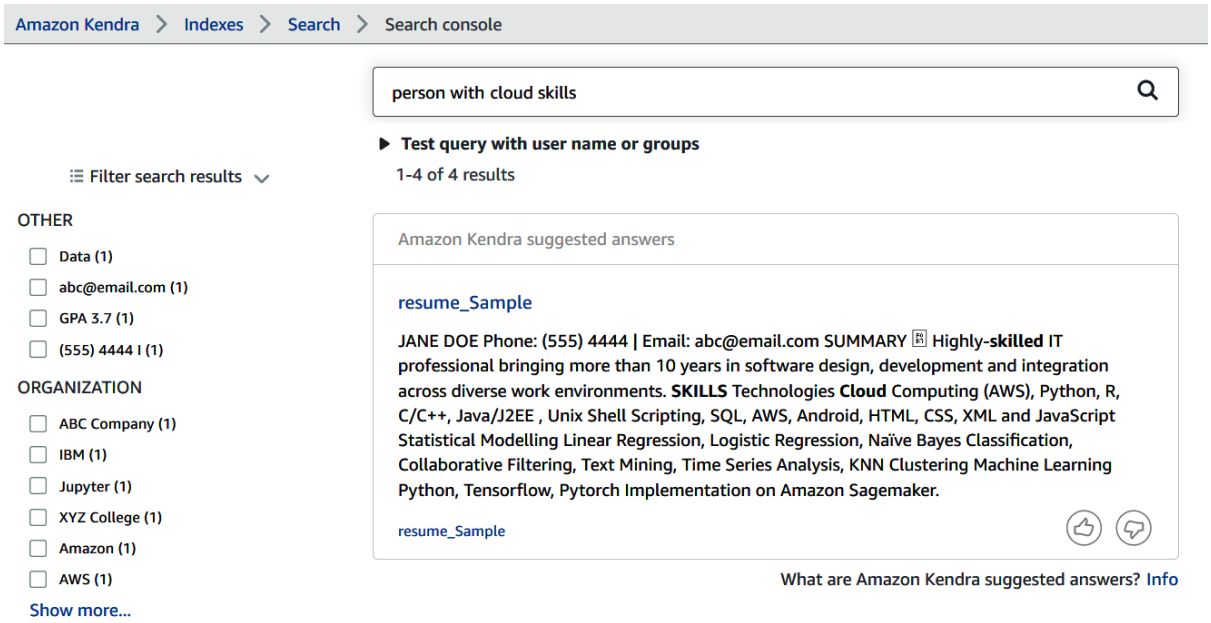


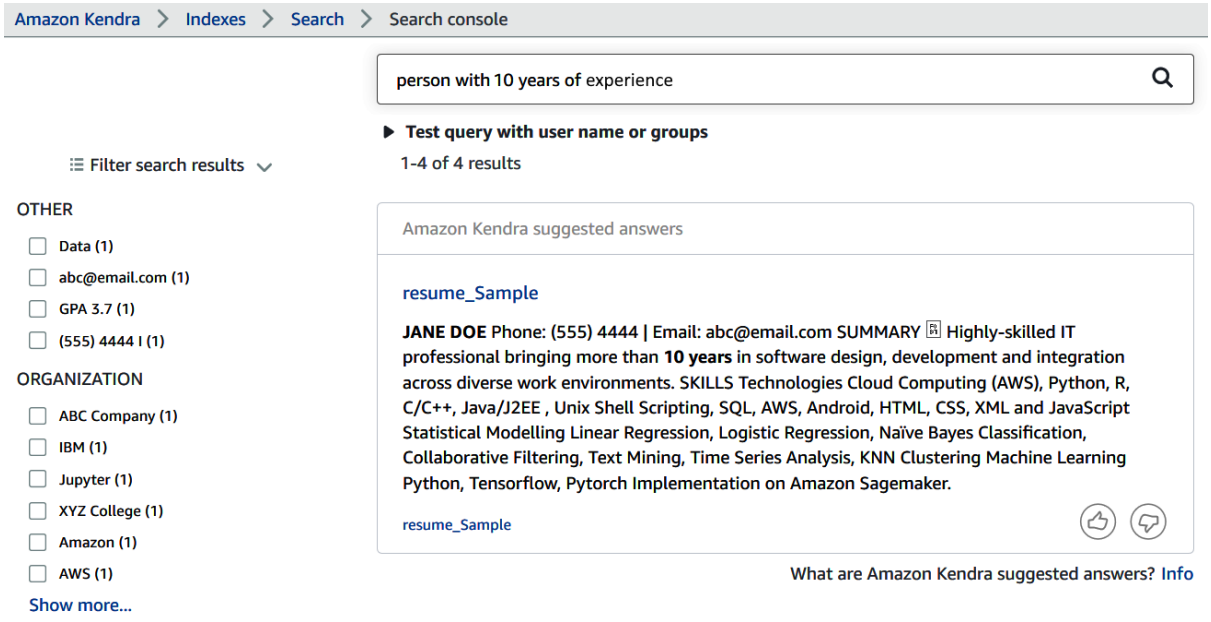

```
Phone: (555) 4444 I Email: abc@email.com
JANE DOE
SUMMARY
Highly-skilled IT professional bringing more than 10 years in software design, development and integration
across diverse work environments.
SKILLS
Technologies
Cloud Computing (AWS), Python, R, C/C++, Java/J2EE, Unix Shell Scripting, SQL, AWS,
Android, HTML, CSS, XML and JavaScript
Statistical Modelling
Linear Regression, Logistic Regression, Naïve Bayes Classification, Collaborative
Filtering, Text Mining, Time Series Analysis, KNN Clustering
Machine Learning
Python, Tensorflow, Pytorch Implementation on Amazon Sagemaker.
```



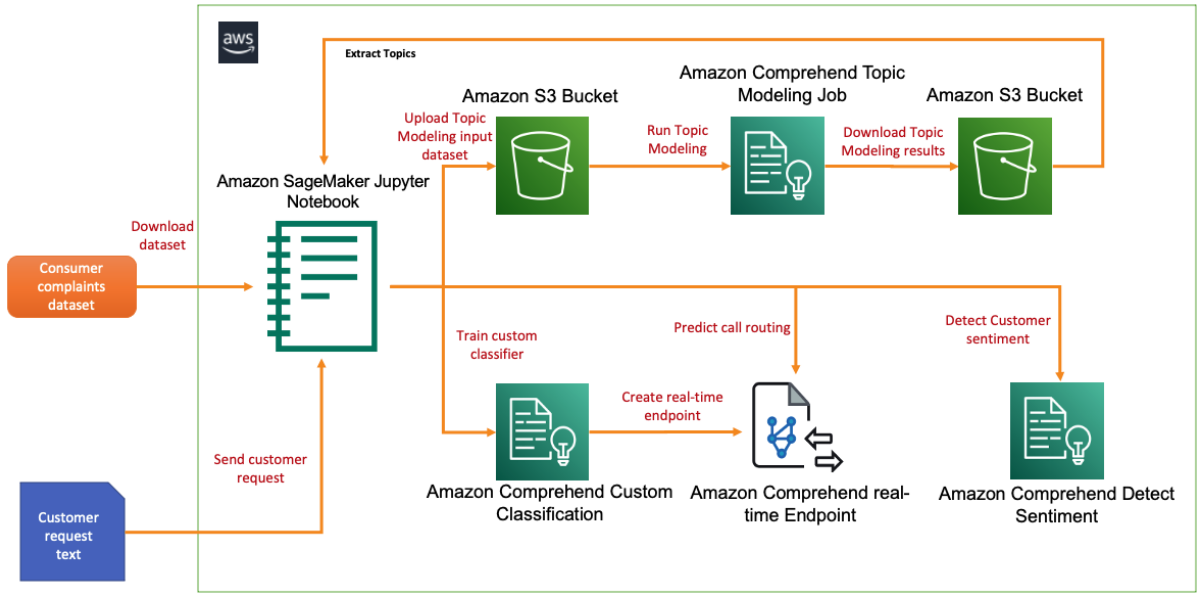
{'ORGANIZATION': {'Amazon NLP Al Services', 'IBM', 'Amazon Athena', 'YZQ Organization', 'Amazon', 'ABC Company', 'Jupyter', 'Amazon Quicksight', 'XYZ College', 'AWS', 'Coursera'}, 'PERSON': {'JANE DOE'}, 'DATE': {'Aug 2016', 'August 2017', 'May 2011', 'Jul 2017'}, 'COMMERCIAL_ITEM': {'Notebook', 'Augmented', 'Code Star'}, 'OTHER': {'GPA 3.7', 'abc@email.com', 'Data', '(555) 4444 I'}, 'TITLE': {'Glue', 'Lakeformation', 'Mongo DB', 'Outlook', 'R', 'IDE_', 'Java/J2EE', 'Database', 'Lambda', 'PowerPoint', 'Tools', 'Oracle 10g', 'Neo4j', 'OBIEE', 'Watson Analytics', 'Naïve', 'AI/ML Stack', 'Ant', 'KNN Clustering', 'HTML', 'Git', 'MS Power BI', 'Python', 'Anaconda', 'Word', 'Analytics', 'Big Data Analytics', 'Tensorflow', 'MySQL', 'Excel', 'JIRA', 'Android', 'CSS', 'Coursera', 'XML', 'JavaScript', 'Unix Shell Scripting', 'C/C++', 'Pytorch', 'Access', 'Big Data', 'Sagemaker', 'Apache Maven', 'Tableau', 'DB2', 'Eclipse', 'SQL', 'Gradle'}, 'QUANTITY': {'more than 10 years'}}

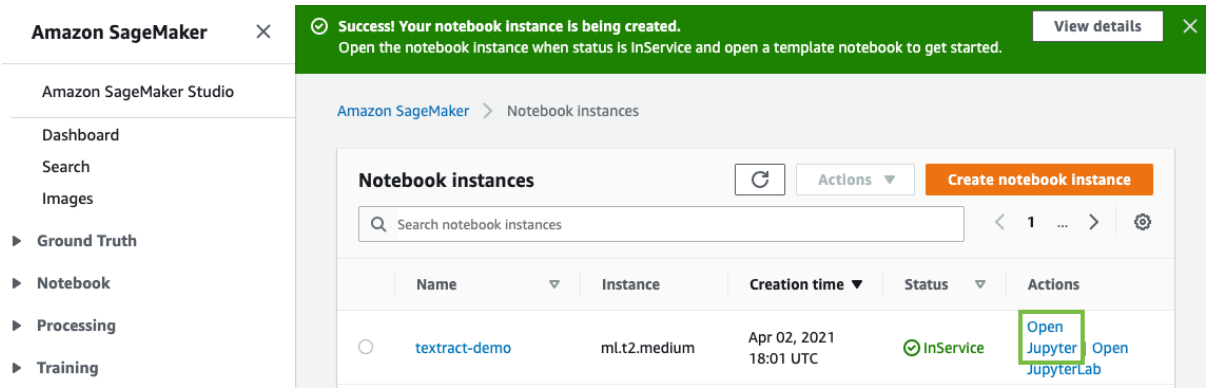

Amazon Kendra > Indexes > Search > Data sources > Add data source > Amazon S3
Step 1
Specify data source details
Step 2
Configure sync settings
Step 3
Review and create
Configure sync settings
Sync scope Info
Enter the data source location
Locate the S3 bucket where your data is stored.
s3://workshop-textract-b9e2b520
Browse S3
Metadata files prefix folder location - optional Info
Locate the meta data files prefix.
meta/
Browse S3
Access control list configuration file location - optional Info
Locate the ACL configuration file.
s3://my-bucket/path/file
Browse S3
Select decryption key
IAM role
Amazon Kendra requires permissions for other services to create this data source. Choose an existing IAM role or let us create a role for you.
IAM roles used for indexes or FAQs can't be used for data sources. If you are unsure if an existing role is used for an index or FAQ, choose "Create a new role" to avoid an error.
Create a new role (Recommended)
Role name
Your role name will be prefixed with 'AmazonKendra-'. The created role will only work for this data source and its specific configuration.
AmazonKendra-s3
Sync run schedule Info
Tell Amazon Kendra how often it should sync this data source. You can check the health of your sync jobs in the data source details page once the data source is created.
Frequency
Select how often you want your data source to sync.
Run on demand
Cancel
Previous
Next

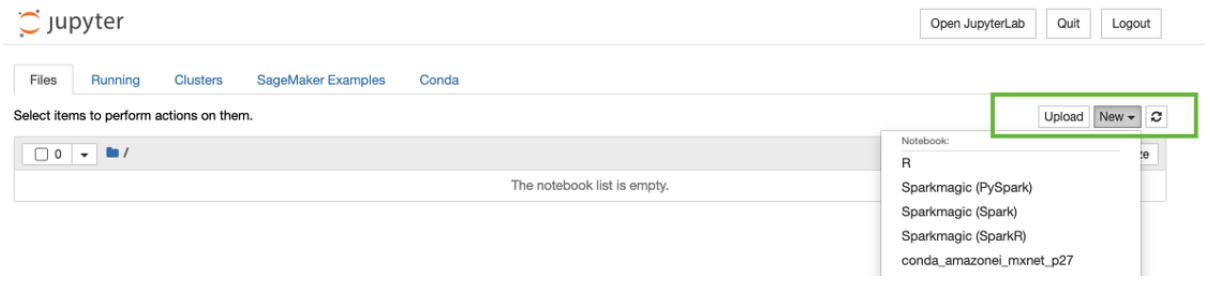

Amazon Kendra > Indexes > Search > Search console
person with cloud skills
Test query with user name or groups
1-4 of 4 results
Filter search results
OTHER
Data (1)
abc@email.com (1)
GPA 3.7 (1)
(555) 4444 I (1)
ORGANIZATION
ABC Company (1)
IBM (1)
Jupyter (1)
XYZ College (1)
Amazon (1)
AWS (1)
Show more...
Amazon Kendra suggested answers
resume_Sample
JANE DOE Phone: (555) 4444 | Email: abc@email.com SUMMARY Highly-skilled IT professional bringing more than 10 years in software design, development and integration across diverse work environments. SKILLS Technologies Cloud Computing (AWS), Python, R, C/C++, Java/J2EE , Unix Shell Scripting, SQL, AWS, Android, HTML, CSS, XML and JavaScript Statistical Modelling Linear Regression, Logistic Regression, Naïve Bayes Classification, Collaborative Filtering, Text Mining, Time Series Analysis, KNN Clustering Machine Learning Python, Tensorflow, Pytorch Implementation on Amazon Sagemaker.
resume_Sample
What are Amazon Kendra suggested answers? Info

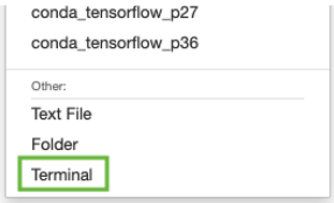

Amazon Kendra > Indexes > Search > Search console
person with 10 years of experience
Test query with user name or groups
1-4 of 4 results
Filter search results
OTHER
Data (1)
abc@email.com (1)
GPA 3.7 (1)
(555) 4444 I (1)
ORGANIZATION
ABC Company (1)
IBM (1)
Jupyter (1)
XYZ College (1)
Amazon (1)
AWS (1)
Show more...
Amazon Kendra suggested answers
resume_Sample
JANE DOE Phone: (555) 4444 | Email: abc@email.com SUMMARY Highly-skilled IT professional bringing more than 10 years in software design, development and integration across diverse work environments. SKILLS Technologies Cloud Computing (AWS), Python, R, C/C++, Java/J2EE , Unix Shell Scripting, SQL, AWS, Android, HTML, CSS, XML and JavaScript Statistical Modelling Linear Regression, Logistic Regression, Naïve Bayes Classification, Collaborative Filtering, Text Mining, Time Series Analysis, KNN Clustering Machine Learning Python, Tensorflow, Pytorch Implementation on Amazon Sagemaker.
resume_Sample
What are Amazon Kendra suggested answers? Info

Chapter 6: Using NLP to Improve Customer Service Efficiency



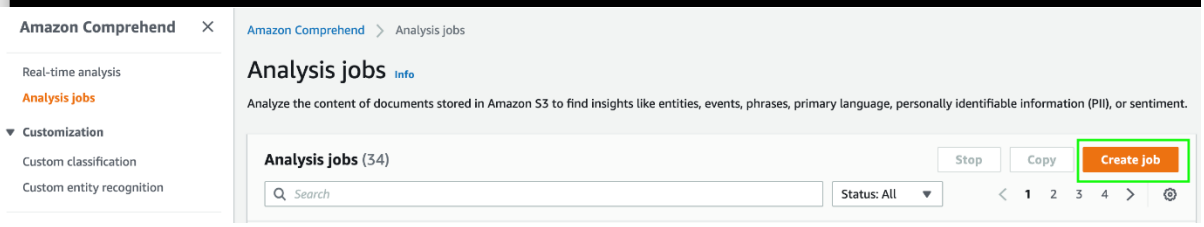






jupyter

```
sh-4.2$ pwd
/home/ec2-user
sh-4.2$ cd SageMaker/
sh-4.2$ git clone https://github.com/PacktPublishing/Natural-Language-Processing-with-AWS-AI-Services
```



Job settings

Name

chapter6-topic-modeling-1

The name can have up to 256 characters. Valid characters: A-Z, a-z, 0-9, space, and % + . / : = @ _ - (hyphen)

Analysis type Info

Topic modeling

Job encryption Info

Input data Info

Data source

- My documents
  We recommend providing at least 1,000 documents containing at least 100 words each.
- Example documents
  Example documents are available only in English

S3 location

Paste the URL of an input data file in S3, or select a bucket or folder location in S3.

s3://nlp-book-samples/chapter6/topic_modeling/input/topic_input.csv

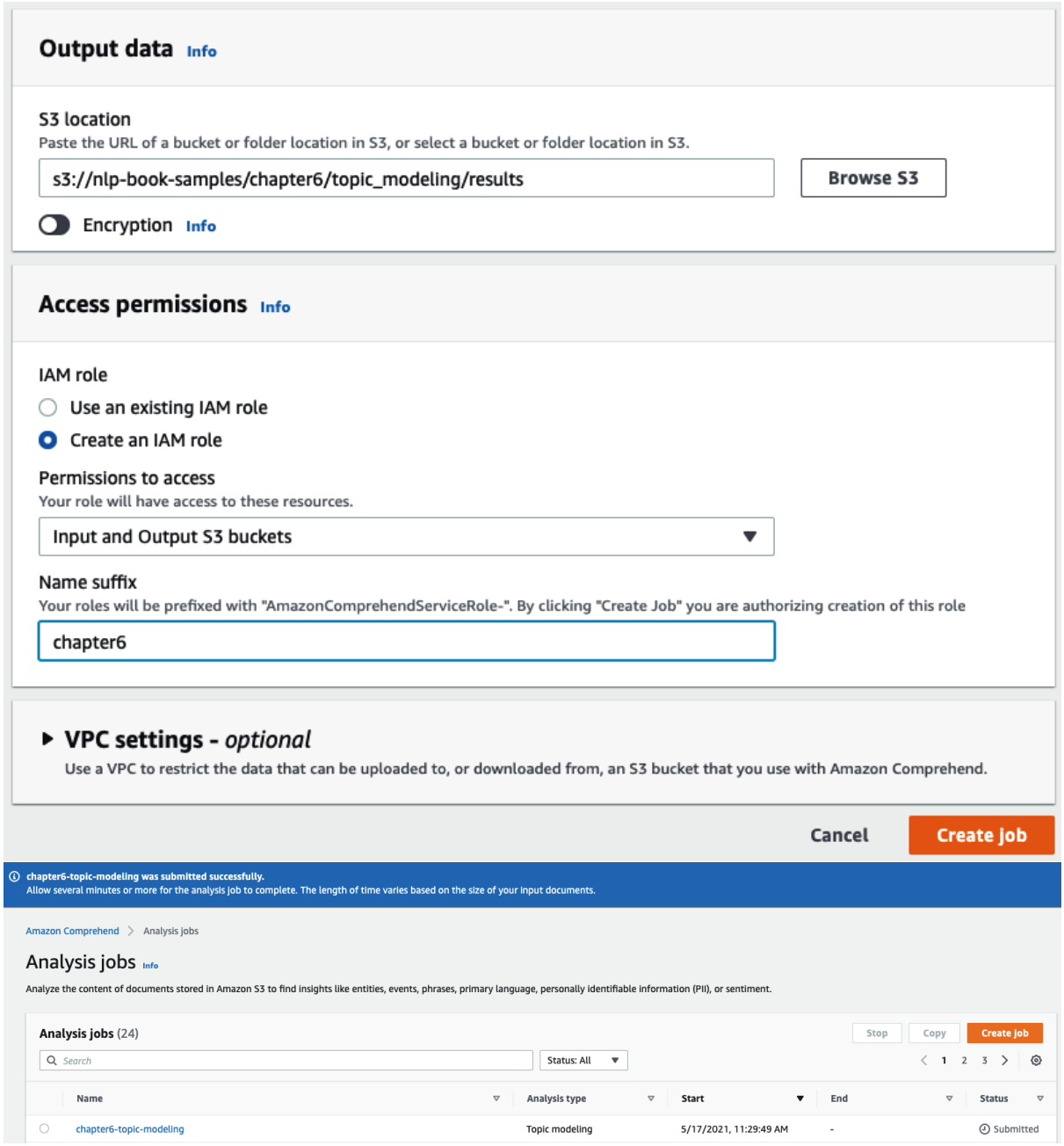
Browse S3

Input format - *optional* Info

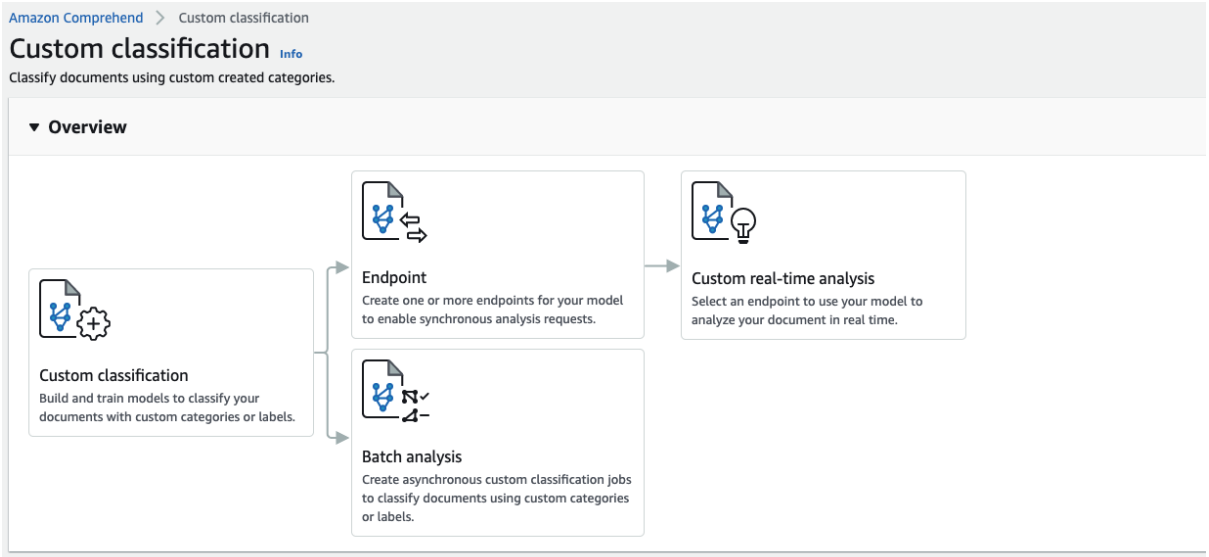
Choose input format

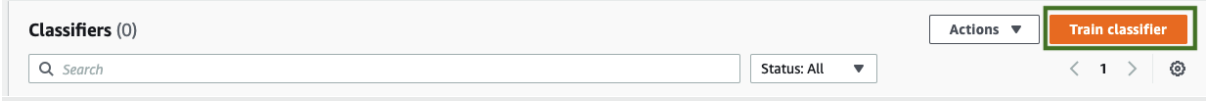
Number of topics Info

8


Output data Info
S3 location
Paste the URL of a bucket or folder location in S3, or select a bucket or folder location in S3.
s3://nlp-book-samples/chapter6/topic_modeling/results
Browse S3
Encryption Info
Access permissions Info
IAM role
Use an existing IAM role
Create an IAM role
Permissions to access
Your role will have access to these resources.
Input and Output S3 buckets
Name suffix
Your roles will be prefixed with "AmazonComprehendServiceRole-". By clicking "Create Job" you are authorizing creation of this role
chapter6
VPC settings - optional
Use a VPC to restrict the data that can be uploaded to, or downloaded from, an S3 bucket that you use with Amazon Comprehend.
Cancel
Create job

chapter6-topic-modeling was submitted successfully.
Allow several minutes or more for the analysis job to complete. The length of time varies based on the size of your input documents.
Amazon Comprehend > Analysis jobs
Analysis jobs Info
Analyze the content of documents stored in Amazon S3 to find insights like entities, events, phrases, primary language, personally identifiable information (PII), or sentiment.
Analysis jobs (24)
Stop
Copy
Create job
Search
Status: All
1 2 3
Name
Analysis type
Start
End
Status
chapter6-topic-modeling
Topic modeling
5/17/2021, 11:29:49 AM
-
Submitted





Train classifier Info

Classifier settings

Name

chapter6-cust-class-1

The name can have up to 63 characters, and it must be unique. Valid characters: A-Z, a-z, 0-9, and - (hyphen)

Language

English

Classifier encryption Info

Training data Info

Classifier mode

Using Multi-class mode
The training data file must have one class and one document on each line. It must have at least 10 documents for each class.

Example

COMEDY	document text 1
COMEDY	document text 2
DRAMA	document text 3

Using Multi-label mode
The training data file must have one or more classes and one document on each line. It must have at least 10 documents for each class.

Example

COMEDY	document text 1
DRAMA	document text 2
COMEDY | DRAMA	document text 3

Training data format

To train your custom model, you must provide training data. This data must be formatted as either a CSV file or as one or more augmented manifest files.

CSV file Info
A two-column CSV file that contains classes in one column and training data in the other. The required format depends on the classifier mode.

s3://nlp-book-samples/chapter6/custom_classification/train/train.csv

Browse S3

Output data - *optional* Info

S3 location

Paste the URL of a bucket or folder location in S3, or select a bucket or folder location in S3.

s3://MyBucketName

Browse S3

Encryption Info

Access permissions Info

IAM role

- (•) Use an existing IAM role
- ( ) Create an IAM role

Role name

A role that grants access to the S3 input and output locations.

AmazonComprehendServiceRole-chapter6

▸ VPC settings - *optional*

Use a VPC to restrict the data that can be uploaded to, or downloaded from, an S3 bucket that you use with Amazon Comprehend.

▸ Tags - *optional* Info

A tag is a label that you can add to a resource as metadata to help you organize, search, or filter your data. Each tag consists of a key and an optional value.

Cancel **Train classifier**

chapter6-cust-class-1 Info

Stop Copy

Classifier details

Name chapter6-cust-class-1	ARN arn:aws:comprehend:us-east-1[redacted]document-classifier/chapter6-cust-class-1	Number of labels -	Input data location s3://nlp-book-samples/chapter6/custom_classification/train/train.csv
Status Training	Training started 5/18/2021, 6:31:36 PM	Number of trained documents -	Language English
Mode Multi-class	Training ended -	Number of test documents -	Classifier encryption -

Classifier performance Info

Classifier training is in progress. Classifier performance will become available once training is complete.

chapter6-cust-class-1 Info

Copy | Delete | Create job

Classifier details

Name chapter6-cust-class-1	ARN arn:aws:comprehend:us-east-1:[redacted]:document-classifier/chapter6-cust-class-1	Number of labels 6	Input data location s3://nlp-book-samples/chapter6/custom_classification/train/train.csv
Status Trained	Training started 5/18/2021, 6:31:36 PM	Number of trained documents 4387	Language English
Mode Multi-class	Training ended 5/18/2021, 7:42:07 PM	Number of test documents 487	Classifier encryption -

Classifier performance Info

Accuracy 0.9055	Precision 0.8745	Recall 0.9261	F1 score 0.89
Hamming loss 0.0945	Micro precision 0.9055	Micro recall 0.9055	Micro F1 score 0.9055

Endpoints (0)

Use endpoints to gain real-time insights.

Copy | Delete | Edit | Use in real-time analysis | Create endpoint

Search | Status: All | 1

Name	Creation time	Inference units	Status

Empty resources

No resource to display

Create endpoint

Tags (0)

Manage tags

Key	Value

Empty tags

No tags associated with the resource

Add tag

Application integration

Use Amazon Comprehend APIs to manage your models. Learn more

Create endpoint Info

Endpoint settings

Endpoint name

chapter5-cust-class-endpoint

The name can have up to 40 characters, and it must be unique. Valid characters: A-Z, a-z, 0-9, and - (hyphen)

Inference units Pricing information

You may incur additional charges depending on duration of use.

1

The number of inference units to provision the endpoint. Each unit represents a throughput of 100 characters per second.

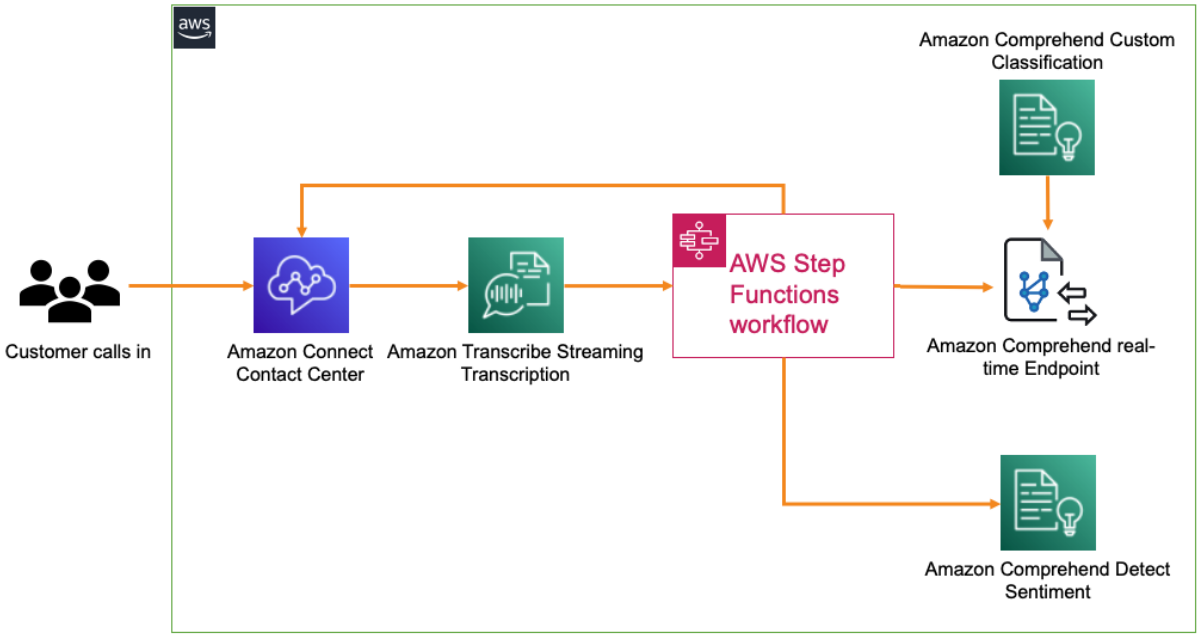
▶ **Tags - *optional*** Info

A tag is a label that you can add to a resource as metadata to help you organize, search, or filter your data. Each tag consists of a key and an optional value.

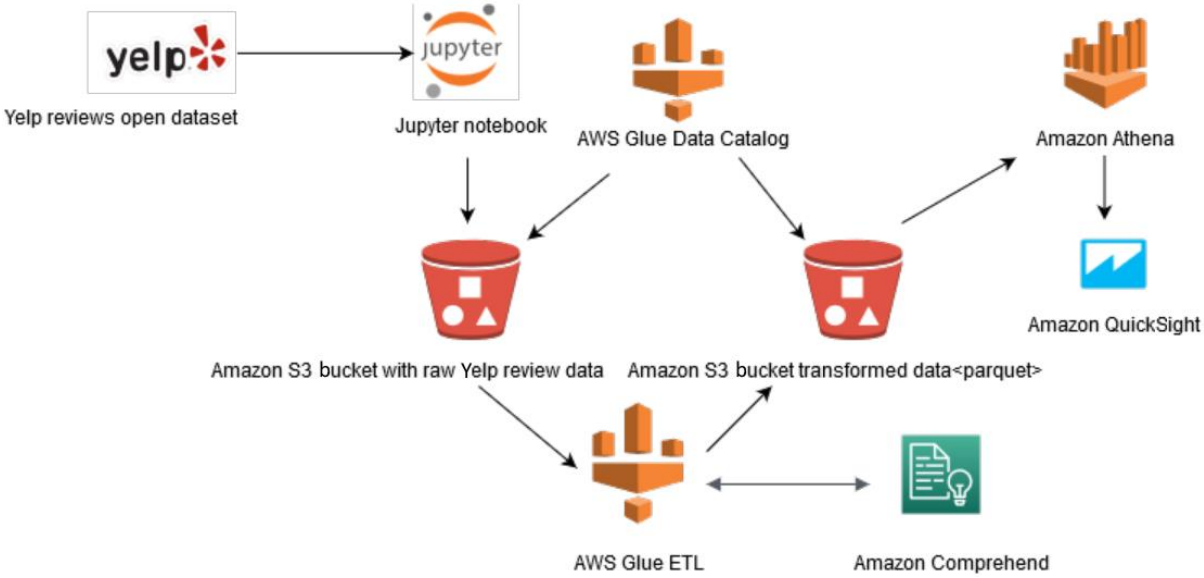
Cancel | Create endpoint

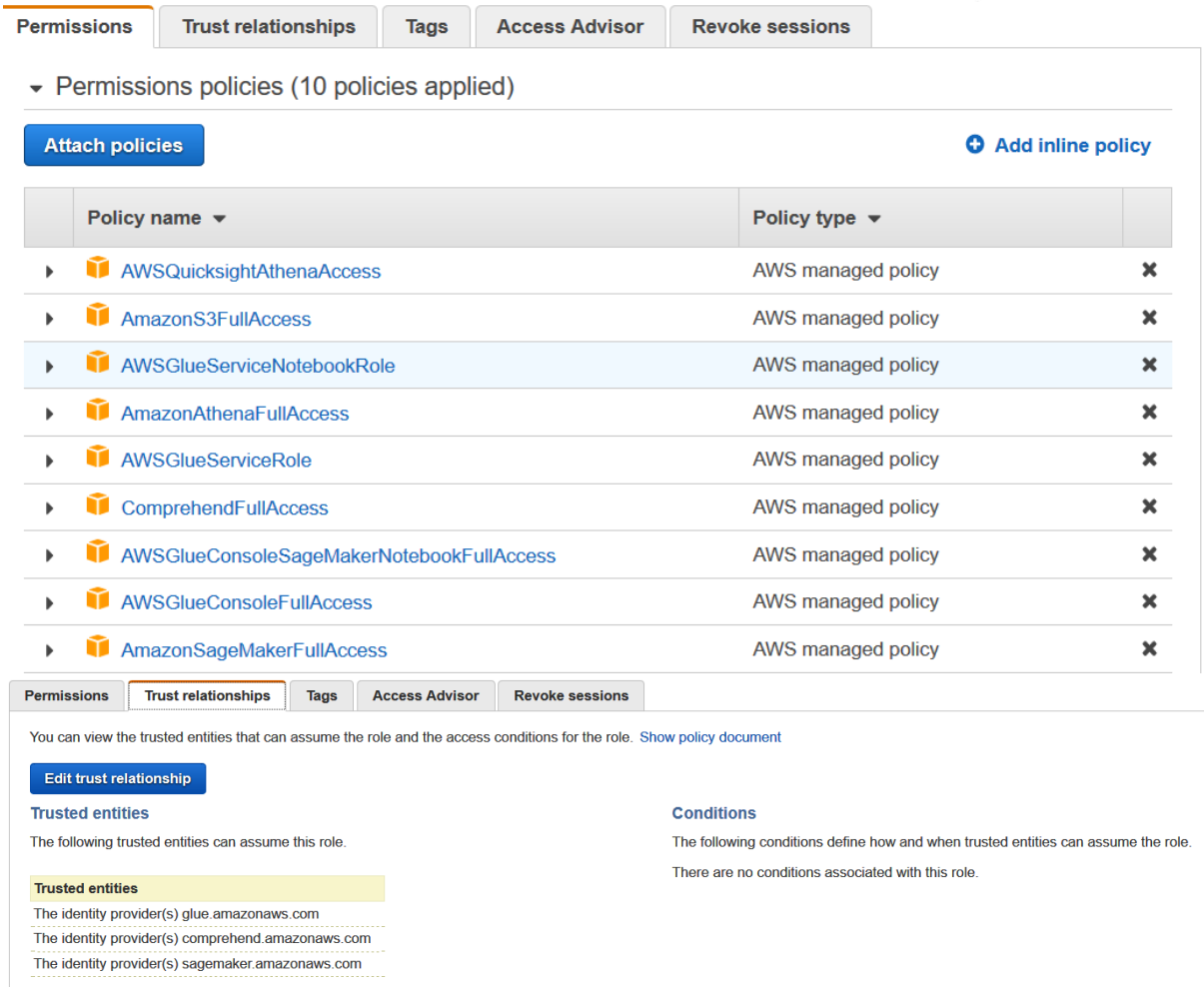
Endpoint details

Name	ARN	Creation time
chapter5-cust-class-endpoint	arn:aws:comprehend:us-east-[redacted]:document-classifier-endpoint/chapter5-cust-class-endpoint	5/18/2021, 8:26:32 PM
Status	**Inference units** Pricing information	**Last modified time**
Ready	1	5/18/2021, 8:26:32 PM



Chapter 7: Understanding the Voice of Your Customer Analytics





	0	1
0	5	dr. goldberg offers everything i look for in a general practitioner. he's nice and easy to talk to without being patronizing; he's always on time in seeing his patients; he's affiliated with a top-notch hospital (nyu) which my parents have explained to me is very important in case something happens and you need surgery; and you can get referrals to see specialists without having to see him first. really, what more do you need? i'm sitting here trying to think of any complaints i have about him, but i'm really drawing a blank.
1	2	Unfortunately, the frustration of being Dr. Goldberg's patient is a repeat of the experience I've had with so many other doctors in NYC -- good doctor, terrible staff. It seems that his staff simply never answers the phone. It usually takes 2 hours of repeated calling to get an answer. Who has time for that or wants to deal with it? I have run into this problem with many other doctors and I just don't get it. You have office workers, you have patients with medical needs, why isn't anyone answering the phone? It's incomprehensible and not work the aggravation. It's with regret that I feel that I have to give Dr. Goldberg 2 stars.
2	4	Been going to Dr. Goldberg for over 10 years. I think I was one of his 1st patients when he started at MHMG. He's been great over the years and is really all about the big picture. It is because of him, not my now former gyn Dr. Markoff, that I found out I have fibroids. He explores all options with you and is very patient and understanding. He doesn't judge and asks all the right questions. Very thorough and wants to be kept in the loop on every aspect of your medical health and your life.
3	4	Got a letter in the mail last week that said Dr. Goldberg is moving to Arizona to take a new position there in June. He will be missed very much. \n\nI think finding a new doctor in NYC that you actually like might almost be as awful as trying to find a date!
4	1	I don't know what Dr. Goldberg was like before moving to Arizona, but let me tell you, STAY AWAY from this doctor and this office. I was going to Dr. Johnson before he left and Goldberg took over when Johnson left. He is not a caring doctor. He is only interested in the co-pay and having you come in for medication refills every month. He will not give refills and could less about patients's financial situations. Trying to get your 90 days mail away pharmacy prescriptions through this guy is a joke. And to make matters even worse, his office staff is incompetent. 90% of the time when you call the office, they'll put you through to a voice mail, that NO ONE ever answers or returns your call. Both my adult children and husband have decided to leave this practice after experiencing such frustration. The entire office has an attitude like they are doing you a favor. Give me a break! Stay away from this doc and the practice. You deserve better and they will not be there when you really need them. I have never felt compelled to write a bad review about anyone until I met this pathetic excuse for a doctor who is all about the money.

	rating	review
0	4	I love that NY NY's food court isn't set up li...
1	3	I had the brisket & pastrami. most ny type del...
2	4	best meat to bread ratio i've experienced. ...
3	5	OMG! Greenberg's serves up some of the most d...
4	3	The System: nn1 Star: Almost impossible to get...

Crawlers > YelpCuratedCrawler

Run crawler | Edit

Name	YelpCuratedCrawler
Description	Crawler for the Parquet Yelp Reviews with Sentiment
Create a single schema for each S3 path	false
Security configuration	
Tags	-
State	Ready
Schedule	
Last updated	Tue May 25 11:02:26 GMT-400 2021
Date created	Tue May 25 11:02:26 GMT-400 2021
Database	yelp
Table prefix	reviews_
Service role	ComprehendEventsBlog-SageMakerRole-1PKY0AT53UQJ4
Selected classifiers	
Data store	S3
Include path	s3://datalake-e37e8e2a-ffde-4a1f-b7c3-959a76f0128f/yelp/parquet/
Connection	
Exclude patterns	

Out[34]:

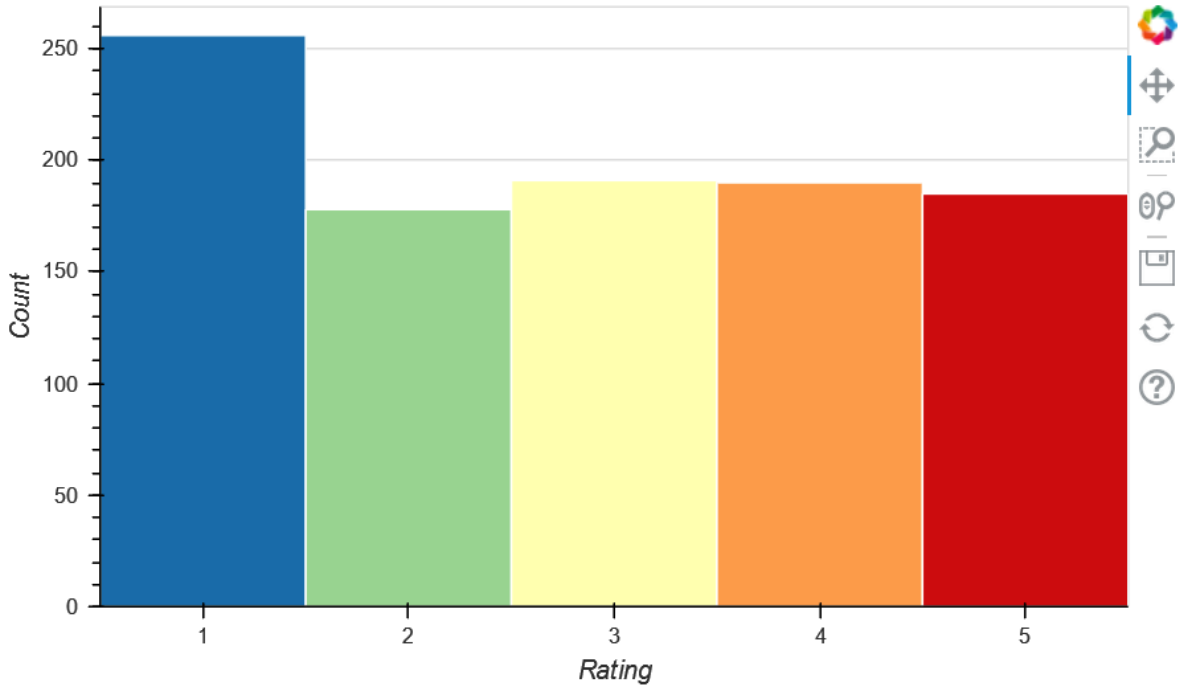
	rating	review	sentiment
0	2	Meet up with some friends Sat night to have so...	POSITIVE
1	5	these guys are great! not only did the tech se...	POSITIVE
2	5	I've been going to this place for 29 years. Th...	POSITIVE
3	4	Super duper friendly staff!!! The people playi...	POSITIVE
4	5	5 stars, this place is as good as it gets! I d...	POSITIVE
5	4	Very nice place to eat! Friendly device, great...	POSITIVE
6	4	I was pleasantly surprised by the amazing food...	POSITIVE
7	3	Timing is what it's all about people!! \nThe ...	POSITIVE
8	4	The food is good. I took it to go. When I got...	POSITIVE
9	4	Have the chicken avocado burrito. It was great.	POSITIVE

Out[35]:

	rating				review			
	count	unique	top	freq	count	unique	top	freq
sentiment								
MIXED	220	5	3	81	220	220	A neighbor recommended Franks all about time a...	1
NEGATIVE	333	5	1	211	333	333	Back in the day before there were $14 gourmet ...	1
NEUTRAL	12	3	1	6	12	12	I would try to help a real non-profit that is ...	1
POSITIVE	435	5	5	176	435	435	This place is a little fat Mexican girl's drea...	1

Out[37]:

	review				sentiment			
	count	unique	top	freq	count	unique	top	freq
rating								
1	256	256	I called Central Towing last night because I h...	1	256	4	NEGATIVE	211
2	178	178	10pm on a Friday night. Wanted a hassle free ...	1	178	3	NEGATIVE	97
3	191	191	Alright store and what you would normally expe...	1	191	4	POSITIVE	87
4	190	190	Food was delicious, has a store too.	1	190	3	POSITIVE	146
5	185	185	I stopped at this waffle house it was really g...	1	185	4	POSITIVE	176



QuickSight access to AWS services

QuickSight can connect to the selected AWS products & services below for all users & groups:

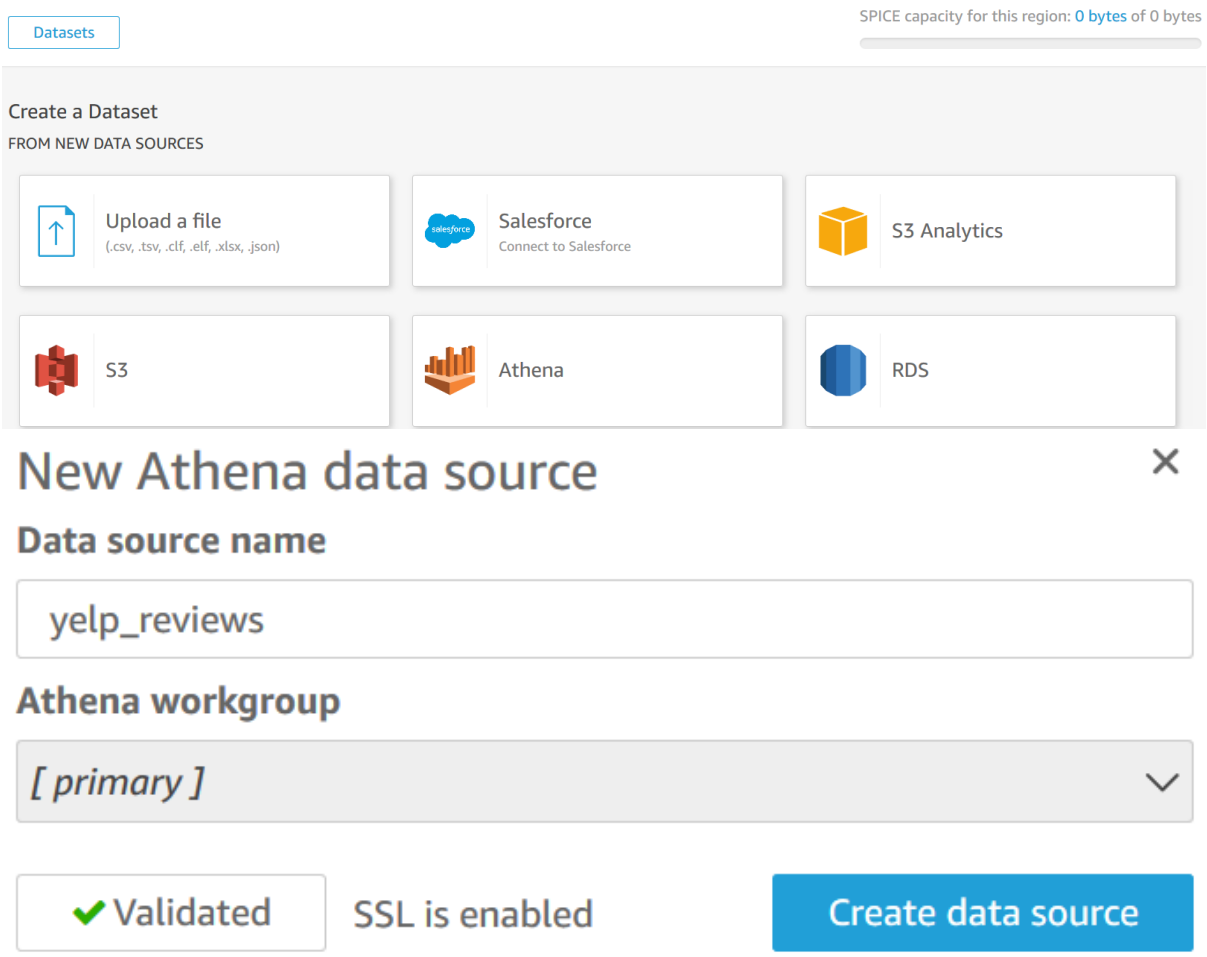
Amazon Redshift
Enables QuickSight to auto-discover clusters ☑

Amazon RDS
Enables QuickSight to auto-discover instances ☑

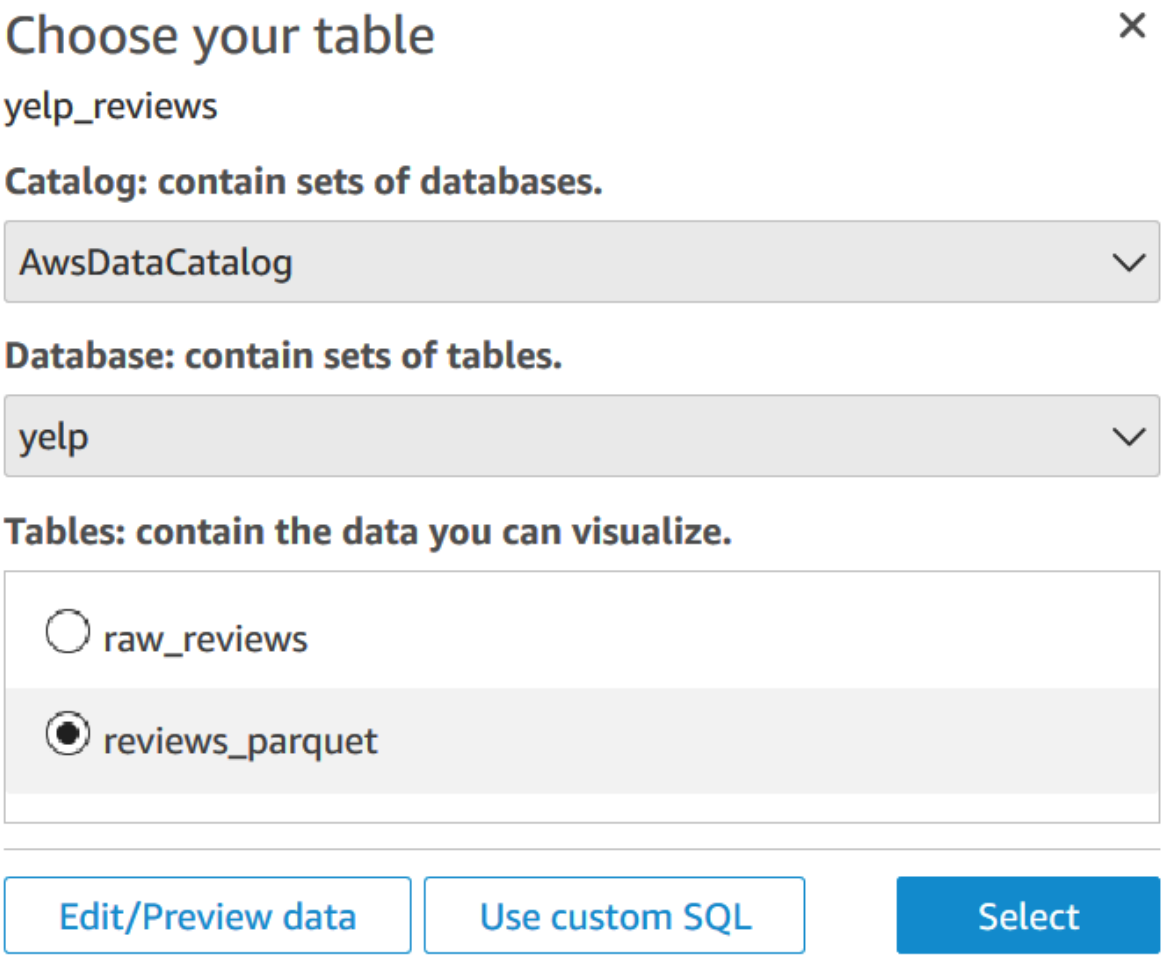
IAM
Enables you to invite IAM users from this AWS Account to access QuickSight ☑

Amazon S3
Enables QuickSight to auto-discover your Amazon S3 buckets Hide ☑

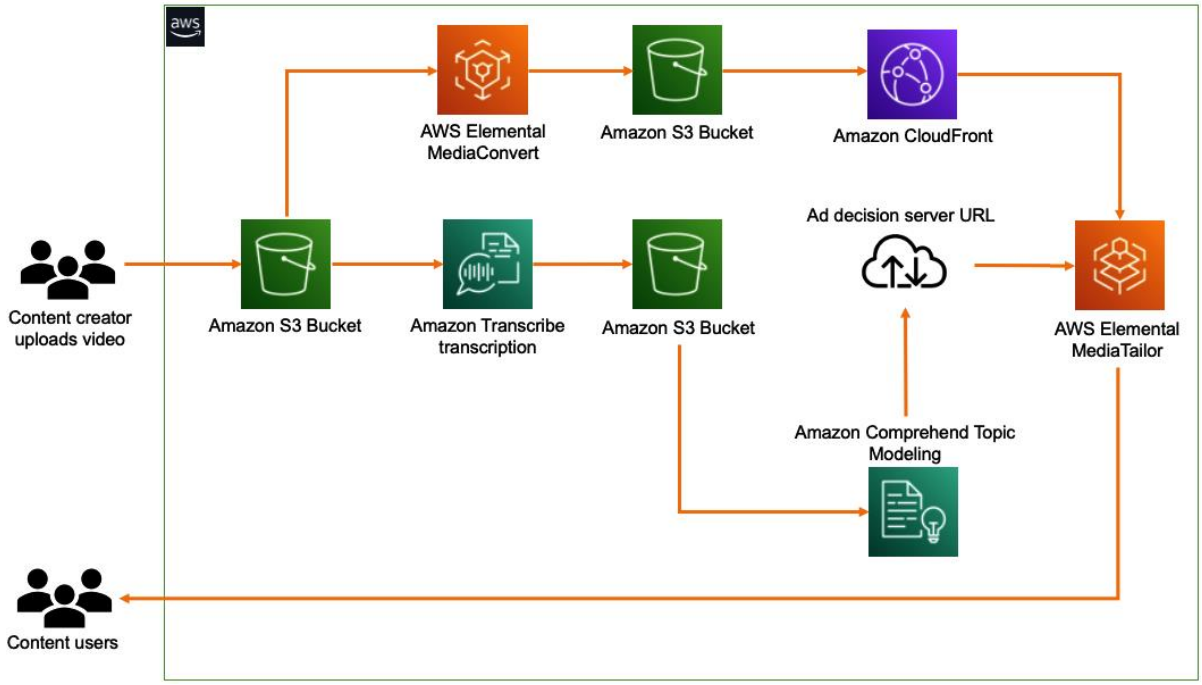
Select S3 buckets — 58 buckets selected

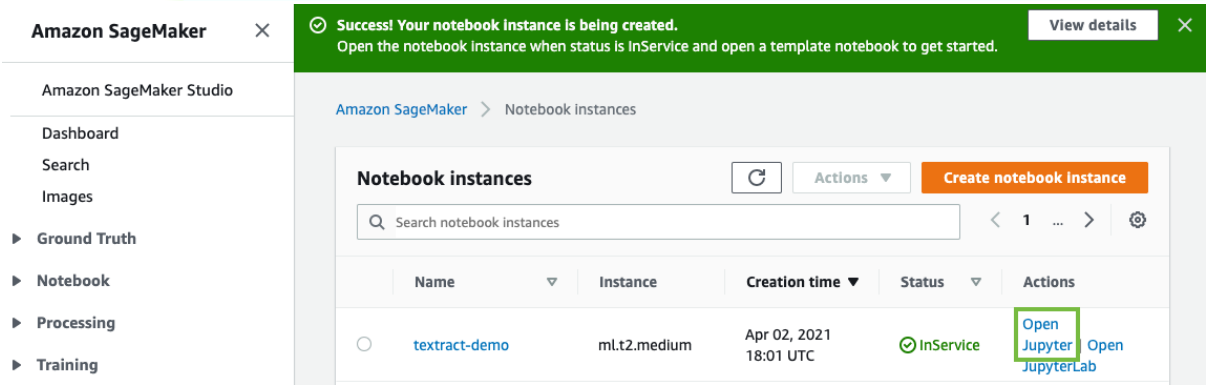

Datasets
SPICE capacity for this region: 0 bytes of 0 bytes
Create a Dataset
FROM NEW DATA SOURCES
Upload a file
(.csv, .tsv, .clf, .elf, .xlsx, .json)
Salesforce
Connect to Salesforce
S3 Analytics
S3
Athena
RDS

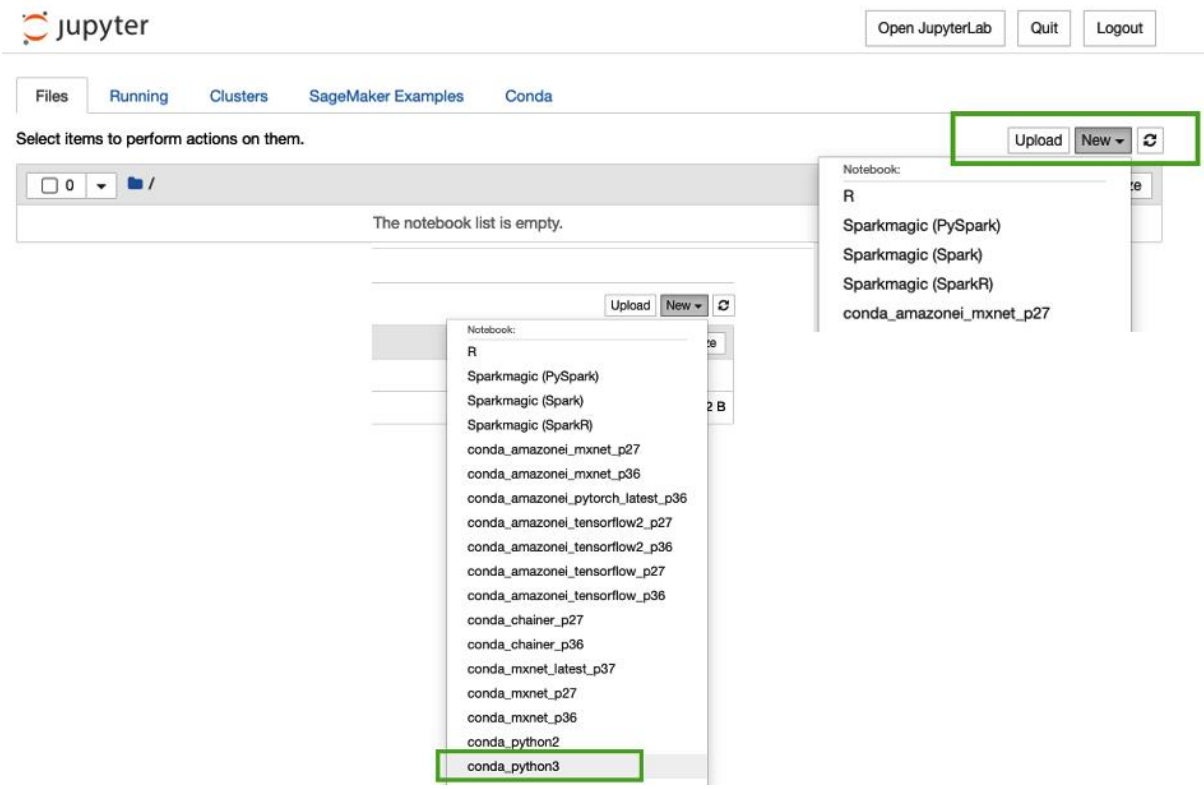
New Athena data source
Data source name
yelp_reviews
Athena workgroup
[primary]
Validated
SSL is enabled
Create data source

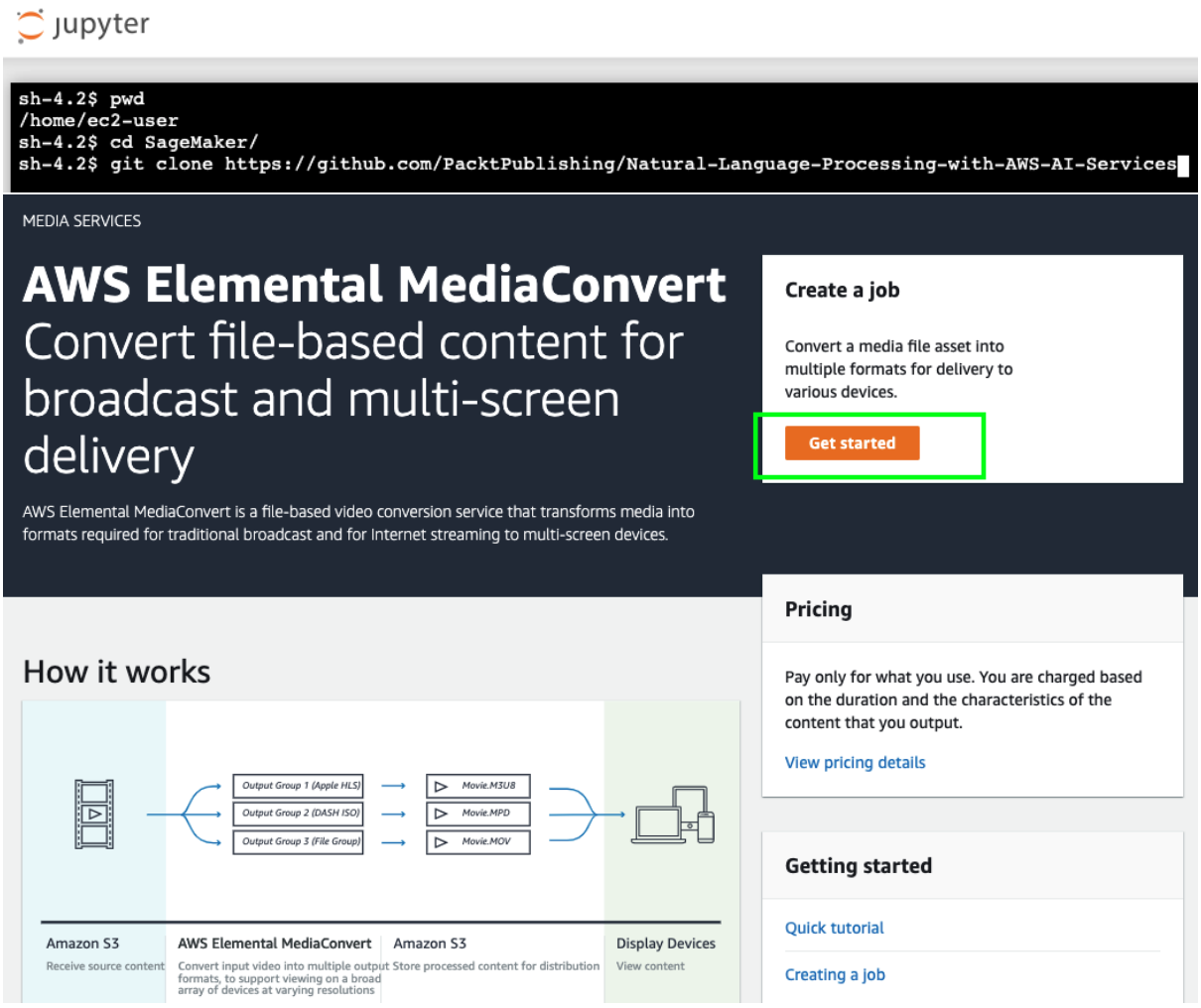

Choose your table
yelp_reviews
Catalog: contain sets of databases.
AwsDataCatalog
Database: contain sets of tables.
yelp
Tables: contain the data you can visualize.
raw_reviews
reviews_parquet
Edit/Preview data
Use custom SQL
Select

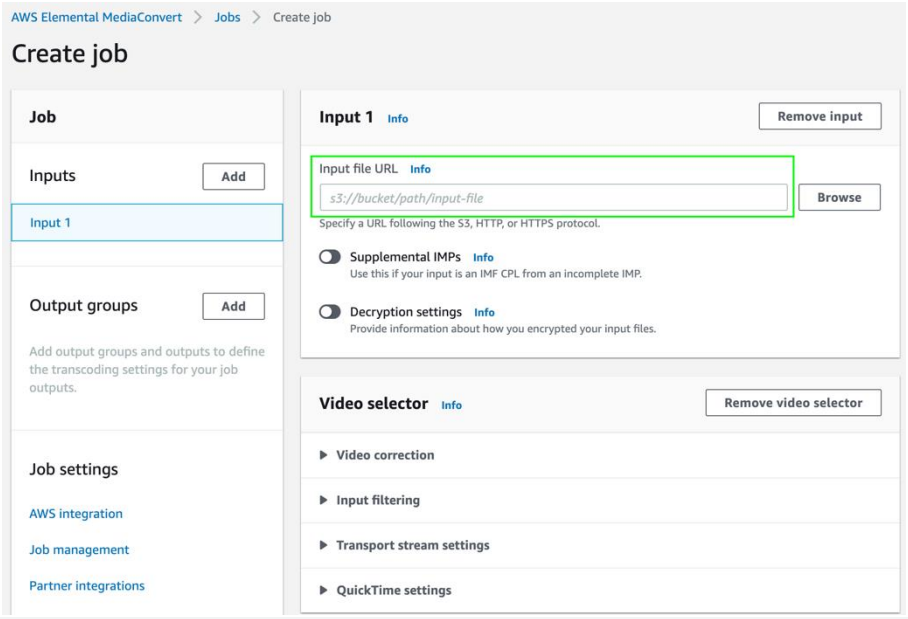
Chapter 8: Leveraging NLP to Monetize Your Media Content

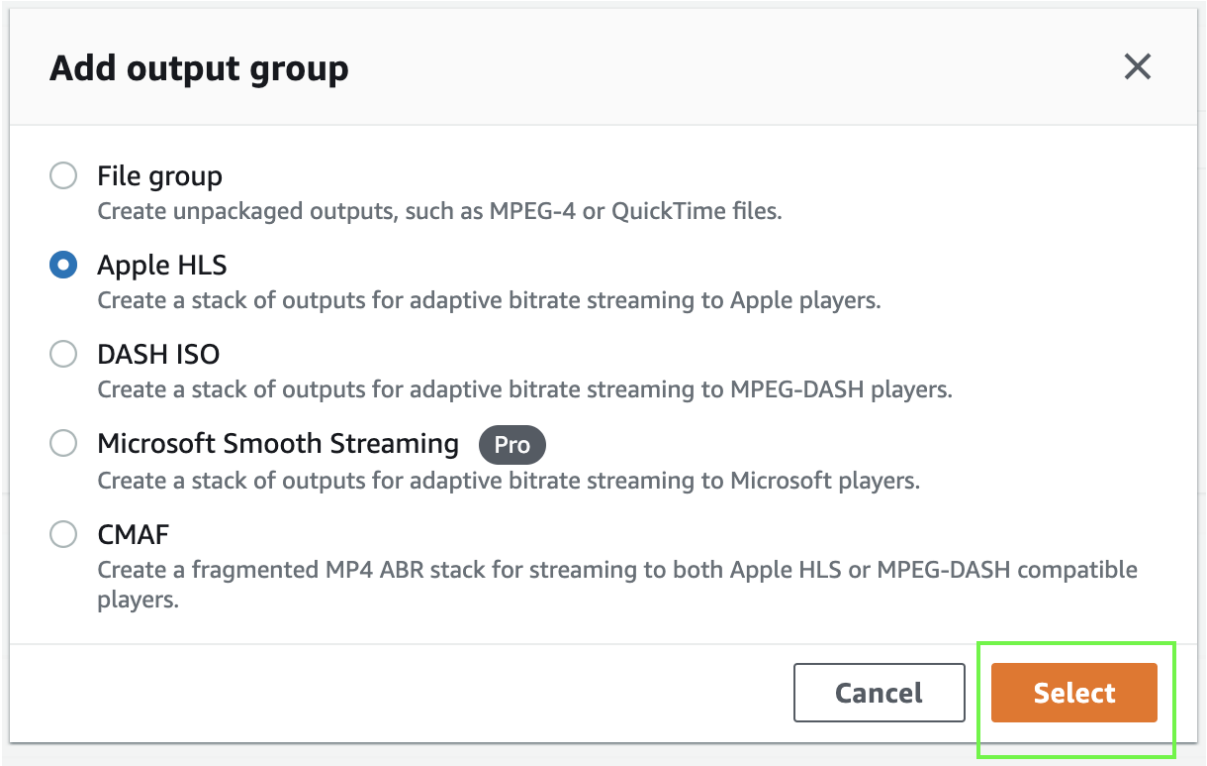


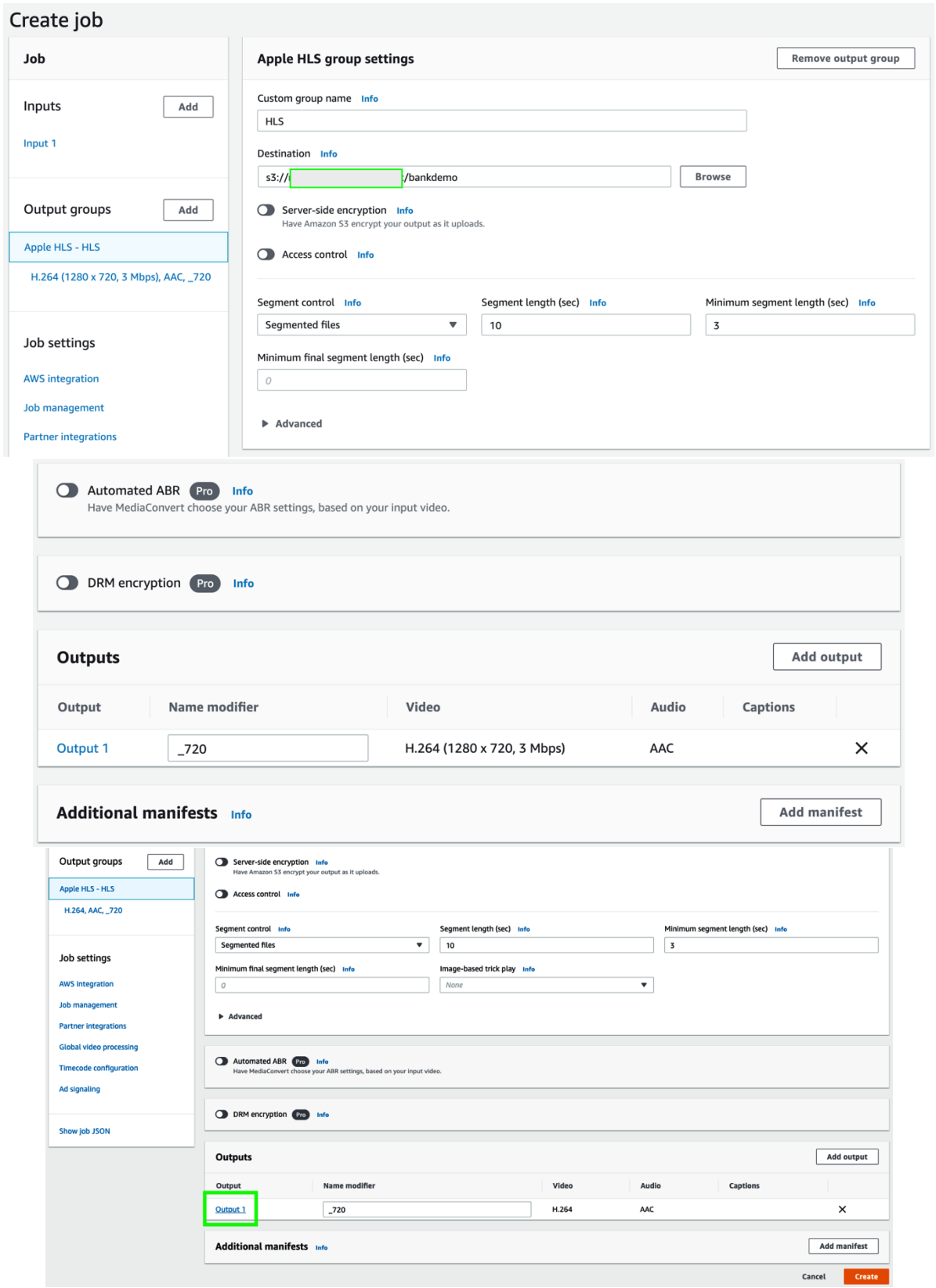


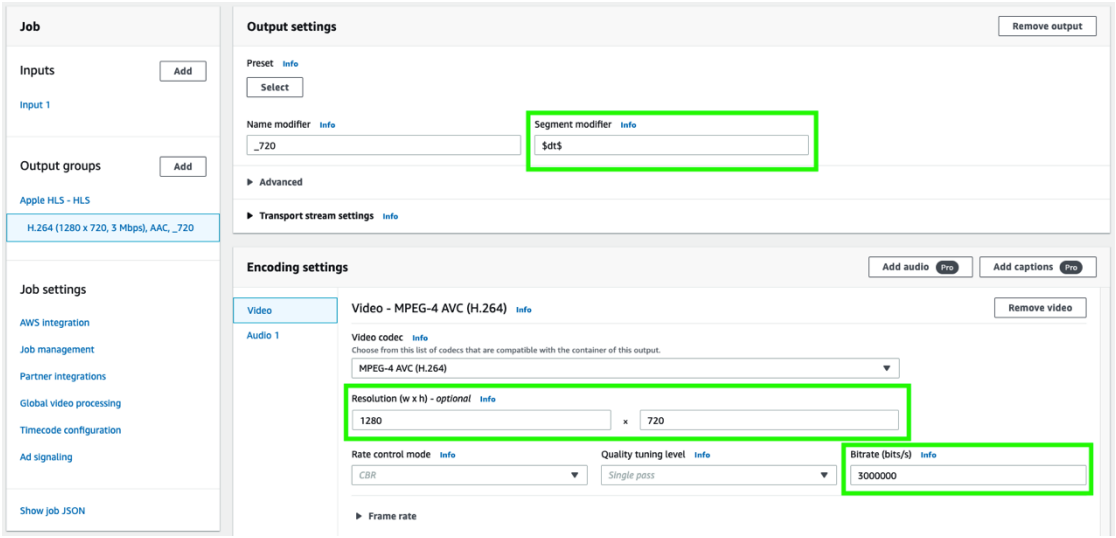

jupyter
Open JupyterLab
Quit
Logout
Files
Running
Clusters
SageMaker Examples
Conda
Select items to perform actions on them.
Upload
New
0
/
The notebook list is empty.
Notebook:
R
Sparkmagic (PySpark)
Sparkmagic (Spark)
Sparkmagic (SparkR)
conda_amazonei_mxnet_p27
conda_amazonei_mxnet_p36
conda_amazonei_pytorch_latest_p36
conda_amazonei_tensorflow2_p27
conda_amazonei_tensorflow2_p36
conda_amazonei_tensorflow_p27
conda_amazonei_tensorflow_p36
conda_chainer_p27
conda_chainer_p36
conda_mxnet_latest_p37
conda_mxnet_p27
conda_mxnet_p36
conda_python2
conda_python3

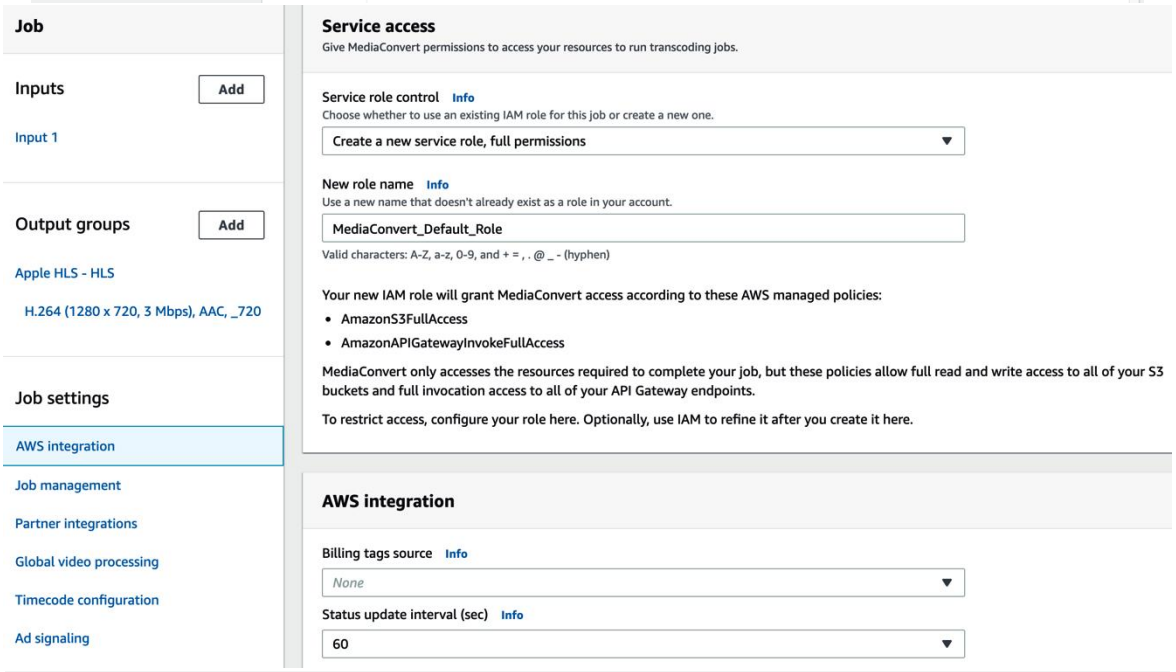

jupyter
sh-4.2$ pwd
/home/ec2-user
sh-4.2$ cd SageMaker/
sh-4.2$ git clone https://github.com/PacktPublishing/Natural-Language-Processing-with-AWS-AI-Services
MEDIA SERVICES
AWS Elemental MediaConvert
Convert file-based content for broadcast and multi-screen delivery
AWS Elemental MediaConvert is a file-based video conversion service that transforms media into formats required for traditional broadcast and for Internet streaming to multi-screen devices.
Create a job
Convert a media file asset into multiple formats for delivery to various devices.
Get started
Pricing
Pay only for what you use. You are charged based on the duration and the characteristics of the content that you output.
View pricing details
How it works
Output Group 1 (Apple HLS)
Output Group 2 (DASH ISO)
Output Group 3 (File Group)
Movie.M3U8
Movie.MPD
Movie.MOV
Amazon S3
Receive source content
AWS Elemental MediaConvert
Convert input video into multiple output formats, to support viewing on a broad array of devices at varying resolutions
Amazon S3
Store processed content for distribution
Display Devices
View content
Getting started
Quick tutorial
Creating a job

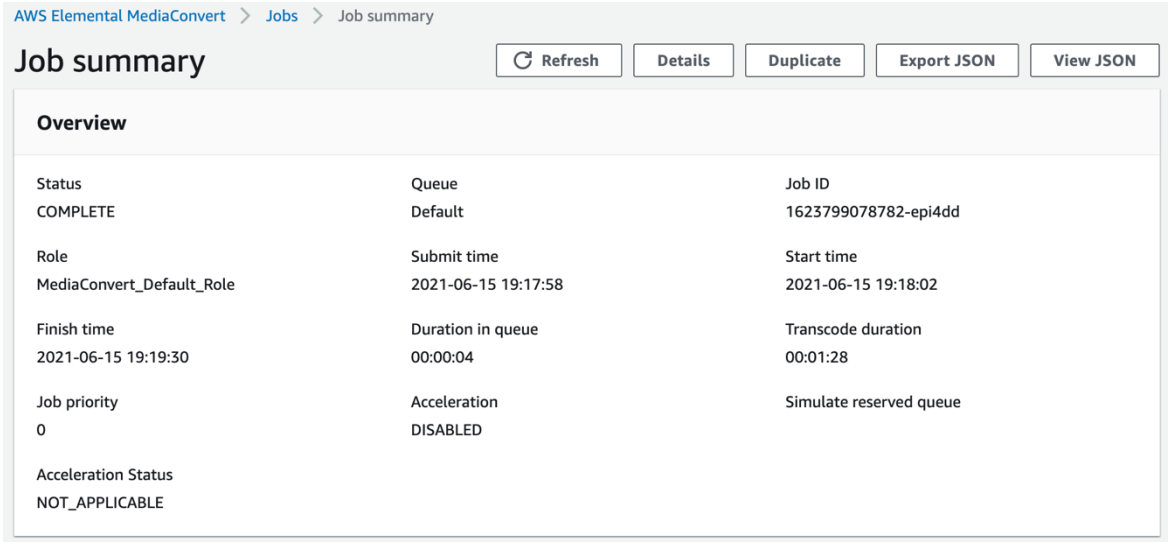

AWS Elemental MediaConvert > Jobs > Create job
Create job
Job
Inputs
Add
Input 1
Output groups
Add
Add output groups and outputs to define the transcoding settings for your job outputs.
Job settings
AWS integration
Job management
Partner integrations
Input 1 Info
Remove input
Input file URL Info
s3://bucket/path/input-file
Browse
Specify a URL following the S3, HTTP, or HTTPS protocol.
Supplemental IMPs Info
Use this if your input is an IMF CPL from an incomplete IMP.
Decryption settings Info
Provide information about how you encrypted your input files.
Video selector Info
Remove video selector
Video correction
Input filtering
Transport stream settings
QuickTime settings

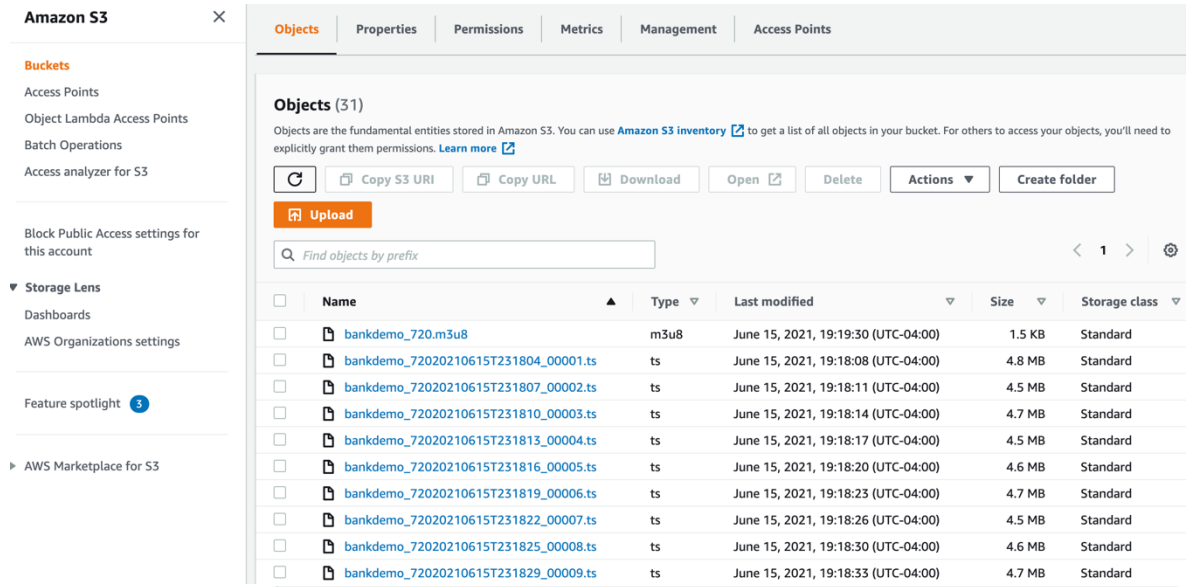

Add output group
File group
Create unpackaged outputs, such as MPEG-4 or QuickTime files.
Apple HLS
Create a stack of outputs for adaptive bitrate streaming to Apple players.
DASH ISO
Create a stack of outputs for adaptive bitrate streaming to MPEG-DASH players.
Microsoft Smooth Streaming
Pro
Create a stack of outputs for adaptive bitrate streaming to Microsoft players.
CMAF
Create a fragmented MP4 ABR stack for streaming to both Apple HLS or MPEG-DASH compatible players.
Cancel
Select

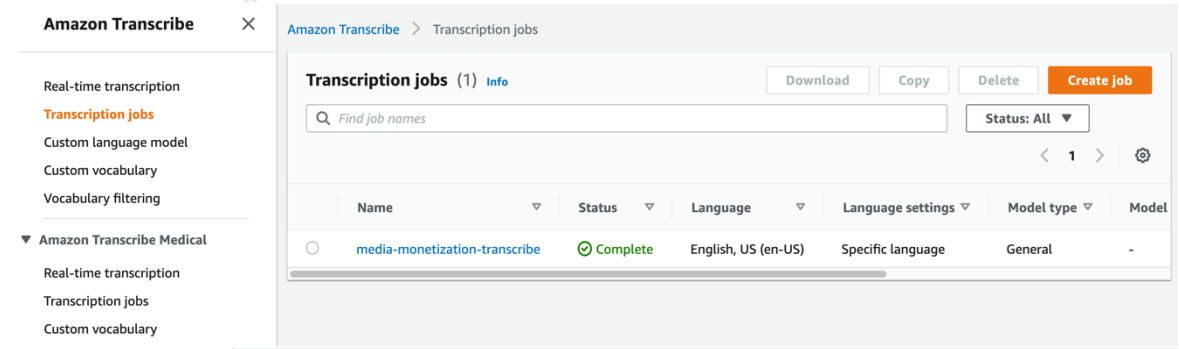

Create job
Job
Inputs
Add
Input 1
Output groups
Add
Apple HLS - HLS
H.264 (1280 x 720, 3 Mbps), AAC, _720
Job settings
AWS integration
Job management
Partner integrations
Apple HLS group settings
Remove output group
Custom group name Info
HLS
Destination Info
s3://
/bankdemo
Browse
Server-side encryption Info
Have Amazon S3 encrypt your output as it uploads.
Access control Info
Segment control Info
Segmented files
Segment length (sec) Info
10
Minimum segment length (sec) Info
3
Minimum final segment length (sec) Info
0
Advanced
Automated ABR Pro Info
Have MediaConvert choose your ABR settings, based on your input video.
DRM encryption Pro Info
Outputs
Add output
Output
Name modifier
Video
Audio
Captions
Output 1
_720
H.264 (1280 x 720, 3 Mbps)
AAC
Additional manifests Info
Add manifest
Output groups
Add
Apple HLS - HLS
H.264, AAC, _720
Job settings
AWS integration
Job management
Partner integrations
Global video processing
Timecode configuration
Ad signaling
Show job JSON
Server-side encryption Info
Have Amazon S3 encrypt your output as it uploads.
Access control Info
Segment control Info
Segmented files
Segment length (sec) Info
10
Minimum segment length (sec) Info
3
Minimum final segment length (sec) Info
0
Image-based trick play Info
None
Advanced
Automated ABR Pro Info
Have MediaConvert choose your ABR settings, based on your input video.
DRM encryption Pro Info
Outputs
Add output
Output
Name modifier
Video
Audio
Captions
Output 1
_720
H.264
AAC
Additional manifests Info
Add manifest
Cancel
Create

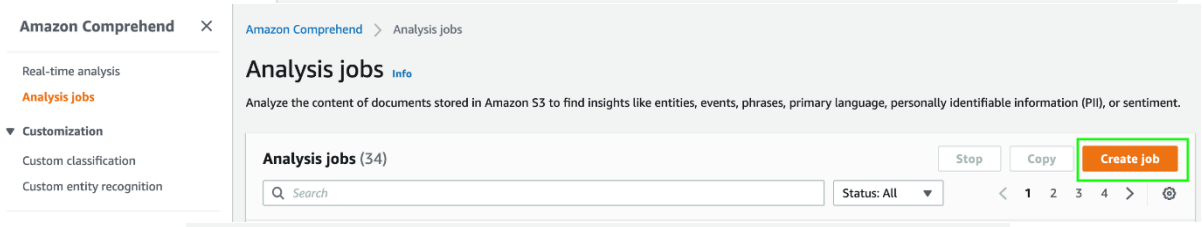

Job
Inputs
Add
Input 1
Output groups
Add
Apple HLS - HLS
H.264 (1280 x 720, 3 Mbps), AAC, _720
Job settings
AWS integration
Job management
Partner integrations
Global video processing
Timecode configuration
Ad signaling
Show job JSON
Output settings
Remove output
Preset Info
Select
Name modifier Info
_720
Segment modifier Info
dt
Advanced
Transport stream settings Info
Encoding settings
Add audio Pro
Add captions Pro
Video
Audio 1
Video - MPEG-4 AVC (H.264) Info
Remove video
Video codec Info
Choose from this list of codecs that are compatible with the container of this output.
MPEG-4 AVC (H.264)
Resolution (w x h) - optional Info
1280
x
720
Rate control mode Info
CBR
Quality tuning level Info
Single pass
Bitrate (bits/s) Info
3000000
Frame rate

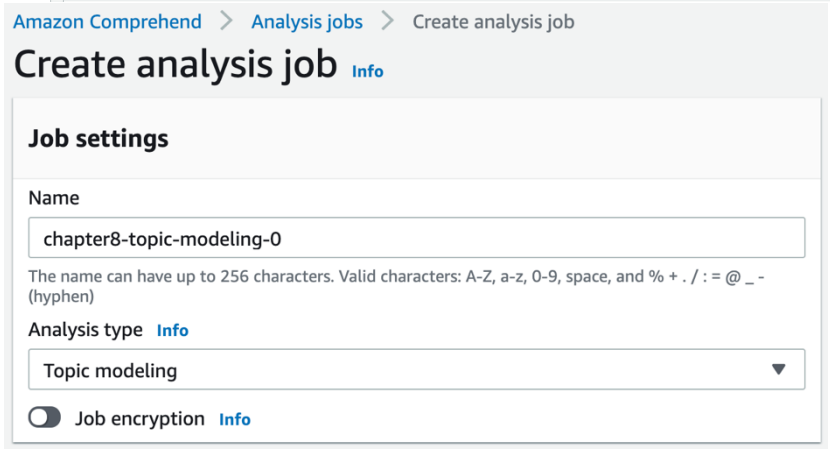

Job
Inputs
Add
Input 1
Output groups
Add
Apple HLS - HLS
H.264 (1280 x 720, 3 Mbps), AAC, _720
Job settings
AWS integration
Job management
Partner integrations
Global video processing
Timecode configuration
Ad signaling
Service access
Give MediaConvert permissions to access your resources to run transcoding jobs.
Service role control Info
Choose whether to use an existing IAM role for this job or create a new one.
Create a new service role, full permissions
New role name Info
Use a new name that doesn't already exist as a role in your account.
MediaConvert_Default_Role
Valid characters: A-Z, a-z, 0-9, and + = , . @ _ - (hyphen)
Your new IAM role will grant MediaConvert access according to these AWS managed policies:
AmazonS3FullAccess
AmazonAPIGatewayInvokeFullAccess
MediaConvert only accesses the resources required to complete your job, but these policies allow full read and write access to all of your S3 buckets and full invocation access to all of your API Gateway endpoints.
To restrict access, configure your role here. Optionally, use IAM to refine it after you create it here.
AWS integration
Billing tags source Info
None
Status update interval (sec) Info
60


AWS Elemental MediaConvert
Jobs
Job summary
Job summary
Refresh
Details
Duplicate
Export JSON
View JSON
Overview
Status
COMPLETE
Queue
Default
Job ID
1623799078782-epi4dd
Role
MediaConvert_Default_Role
Submit time
2021-06-15 19:17:58
Start time
2021-06-15 19:18:02
Finish time
2021-06-15 19:19:30
Duration in queue
00:00:04
Transcode duration
00:01:28
Job priority
0
Acceleration
DISABLED
Simulate reserved queue
Acceleration Status
NOT_APPLICABLE


Amazon S3
Buckets
Access Points
Object Lambda Access Points
Batch Operations
Access analyzer for S3
Block Public Access settings for this account
Storage Lens
Dashboards
AWS Organizations settings
Feature spotlight 3
AWS Marketplace for S3
Objects
Properties
Permissions
Metrics
Management
Access Points
Objects (31)
Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. Learn more
Copy S3 URI
Copy URL
Download
Open
Delete
Actions
Create folder
Upload
Find objects by prefix
Name
Type
Last modified
Size
Storage class
bankdemo_720.m3u8 m3u8 June 15, 2021, 19:19:30 (UTC-04:00) 1.5 KB Standard
bankdemo_720202106151231804_00001.ts ts June 15, 2021, 19:18:08 (UTC-04:00) 4.8 MB Standard
bankdemo_720202106151231807_00002.ts ts June 15, 2021, 19:18:11 (UTC-04:00) 4.5 MB Standard
bankdemo_720202106151231810_00003.ts ts June 15, 2021, 19:18:14 (UTC-04:00) 4.7 MB Standard
bankdemo_720202106151231813_00004.ts ts June 15, 2021, 19:18:17 (UTC-04:00) 4.5 MB Standard
bankdemo_720202106151231816_00005.ts ts June 15, 2021, 19:18:20 (UTC-04:00) 4.6 MB Standard
bankdemo_720202106151231819_00006.ts ts June 15, 2021, 19:18:23 (UTC-04:00) 4.7 MB Standard
bankdemo_720202106151231822_00007.ts ts June 15, 2021, 19:18:26 (UTC-04:00) 4.5 MB Standard
bankdemo_720202106151231825_00008.ts ts June 15, 2021, 19:18:30 (UTC-04:00) 4.6 MB Standard
bankdemo_720202106151231829_00009.ts ts June 15, 2021, 19:18:33 (UTC-04:00) 4.7 MB Standard


Amazon Transcribe
Real-time transcription
Transcription jobs
Custom language model
Custom vocabulary
Vocabulary filtering
Amazon Transcribe Medical
Real-time transcription
Transcription jobs
Custom vocabulary
Amazon Transcribe > Transcription jobs
Transcription jobs (1) Info
Download
Copy
Delete
Create job
Find job names
Status: All
Name
Status
Language
Language settings
Model type
Model
media-monetization-transcribe Complete English, US (en-US) Specific language General -


Amazon Comprehend
Real-time analysis
Analysis jobs
Customization
Custom classification
Custom entity recognition
Amazon Comprehend > Analysis jobs
Analysis jobs Info
Analyze the content of documents stored in Amazon S3 to find insights like entities, events, phrases, primary language, personally identifiable information (PII), or sentiment.
Analysis jobs (34)
Stop
Copy
Create job
Search
Status: All


Amazon Comprehend > Analysis jobs > Create analysis job
Create analysis job Info
Job settings
Name
chapter8-topic-modeling-0
The name can have up to 256 characters. Valid characters: A-Z, a-z, 0-9, space, and % + . / : = @ _ - (hyphen)
Analysis type Info
Topic modeling
Job encryption Info

Input data Info

Data source

- My documents
  We recommend providing at least 1,000 documents containing at least 100 words each.
- Example documents
  Example documents are available only in English

S3 location

Paste the URL of an input data file in S3, or select a bucket or folder location in S3.

s3:// [redacted] /chapter8/topic-modeling/job-input/tm-input.csv Browse S3

Input format - *optional* Info

One document per line

Number of topics Info

2

Maximum of 100

Output data Info

S3 location

Paste the URL of a bucket or folder location in S3, or select a bucket or folder location in S3.

s3://[redacted]/chapter8/topic-modeling Browse S3

Encryption Info

Access permissions Info

IAM role

- Use an existing IAM role
- Create an IAM role

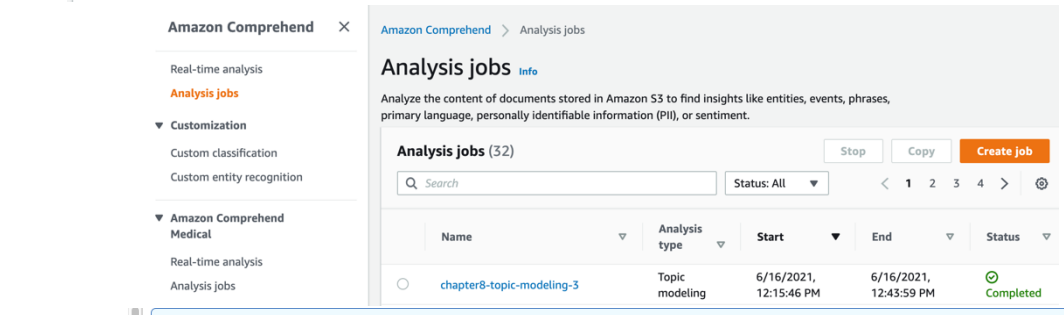
Permissions to access

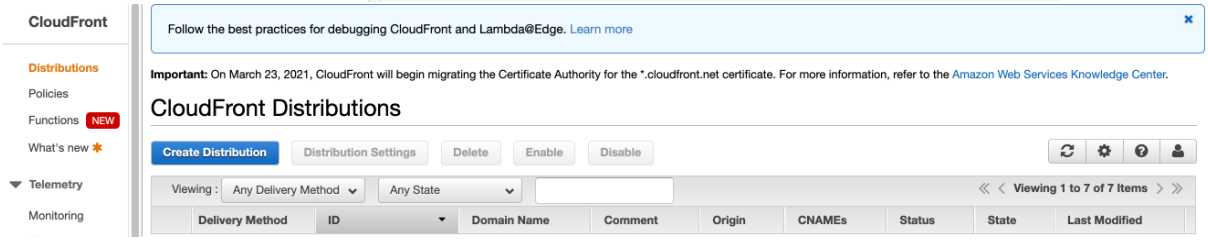
Your role will have access to these resources.

Any S3 bucket

Name suffix

Your roles will be prefixed with "AmazonComprehendServiceRole-". By clicking "Create Job" you are authorizing creation of this role





Create Distribution

Origin Settings

Origin Domain Name [redacted].s3.amazonaws.com

Origin Path

Enable Origin Shield
- ○ Yes
- ◉ No

Origin ID S3-[redacted]

Restrict Bucket Access
- ◉ Yes
- ○ No

Origin Access Identity
- ◉ Create a New Identity
- ○ Use an Existing Identity

Comment access-identity-[redacted].s3.ama

Grant Read Permissions on Bucket
- ◉ Yes, Update Bucket Policy
- ○ No, I Will Update Permissions

Origin Connection Attempts 3

Origin Connection Timeout 10

Origin Custom Headers | **Header Name** | **Value**

Step 1: Select delivery method

Step 2: Create distribution

Default Cache Behavior Settings

Path Pattern Default (*)

Viewer Protocol Policy
- ○ HTTP and HTTPS
- ◉ Redirect HTTP to HTTPS
- ○ HTTPS Only

Allowed HTTP Methods
- ◉ GET, HEAD
- ○ GET, HEAD, OPTIONS
- ○ GET, HEAD, OPTIONS, PUT, POST, PATCH, DELETE

Field-level Encryption Config

Cached HTTP Methods GET, HEAD (Cached by default)

Cache and origin request settings
- ◉ Use a cache policy and origin request policy
- ○ Use legacy cache settings

Cache Policy Managed-Elemental-MediaPackage | Create a new policy

View policy details

Learn More

Origin Request Policy | Create a new policy

View policy details

Learn More

Smooth Streaming
- ○ Yes
- ◉ No

Restrict Viewer Access (Use Signed URLs or Signed Cookies)
- ○ Yes
- ◉ No

Compress Objects Automatically
- ○ Yes
- ◉ No

Distribution Settings

Price Class Use Only U.S., Canada and Europe

Amazon WAF Web ACL None

Alternate Domain Names (CNAMEs)

SSL Certificate

- (•) Default CloudFront Certificate (*.cloudfront.net)

  Choose this option if you want your users to use HTTPS or HTTP to access your content with the CloudFront domain https://d111111abcdef8.cloudfront.net/logo.jpg).
  Important: If you choose this option, CloudFront requires that browsers or devices support TLSv1 or later to access

- ( ) Custom SSL Certificate (example.com):

  Choose this option if you want your users to access your content by using an alternate domain name, such as
  You can use a certificate stored in Amazon Certificate Manager (ACM) in the US East
  (N. Virginia) Region, or you can use a certificate stored in IAM.

  Request or Import a Certificate with ACM

  Learn more about using custom SSL/TLS certificates with CloudFront.
  Learn more about using ACM.

Supported HTTP Versions

- (•) HTTP/2, HTTP/1.1, HTTP/1.0
- ( ) HTTP/1.1, HTTP/1.0

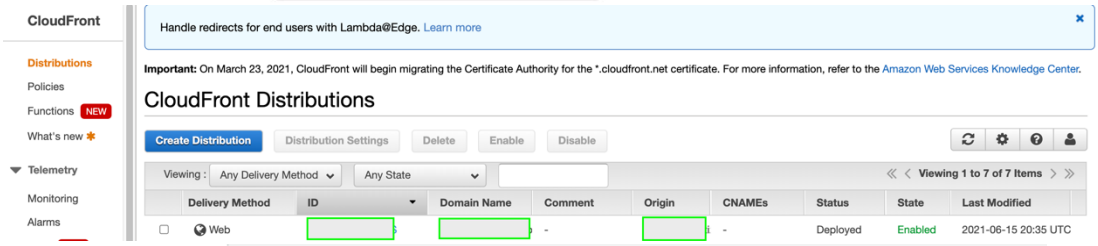
Default Root Object

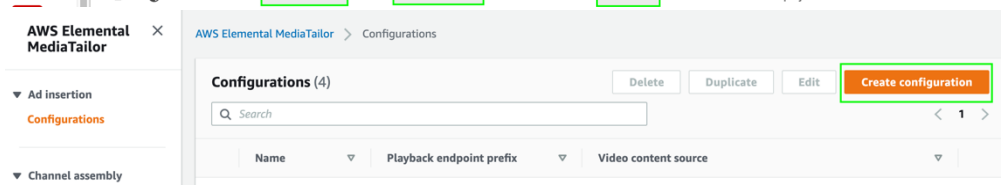
Standard Logging

- ( ) On
- (•) Off

S3 Bucket for Logs

Log Prefix





AWS Elemental MediaTailor > Configurations > Create configuration

Create configuration

Required settings

Every configuration requires a name, a content source, and an ad decision server (ADS) that returns VAST or VMAP responses.

Name

A configuration identifier that you use for API and console interactions.

chapter8-ad-campaign

Supported characters are numbers, letters, underscores (), and dashes (-). Must be unique per AWS account per Region.

Content source Info

The origin server that's providing content to AWS Elemental MediaTailor.

https://[redacted].cloudfront.net

If using HTTPS, the certificate can't be self-signed.

Ad decision server Info

The URL for the ad decision server (ADS).

https://pubads.g.doubleclick.net/gampad/ads?sz=640x480&iu=/124319096/external/ad_rule_samples&ciu_szs=300x250&ad_rule=1&impl=s&gdfp_req=1&env=vp&output=vmap&unviewed_position_start=1&cust_params=deployment%3Ddevsite%26sample_ar%3Dpremidpost&cmsid=176&vid=short_tencue&correlator=[avail.random]

If using HTTPS, the certificate can't be self-signed. Maximum 25,000 characters.

Playback endpoint prefixes

Session initialization prefix [redacted]

HLS playback prefix [redacted]

DASH playback prefix [redacted]

Video settings

Video content source

https://[redacted]

Ad settings

Ad decision server

https://pubads.g.doubleclick.net/gampad/ads?sz=640x480&iu=/124319096/external/ad_rule_samples&ciu_szs=300x250&ad_rule=1&impl=s&gdfp_req=1&env=vp&output=vmap&unviewed_position_start=1&cust_params=deployment%3Ddevsite%26sample_ar%3Dpremidpost&cmsid=496&vid=short_onecue&correlator=[avail.random]

Slate ad

Not defined

Personalization threshold

Not defined

Bumper start URL

Not defined

Live preroll ad decision server

Not defined

Live preroll maximum allowed duration

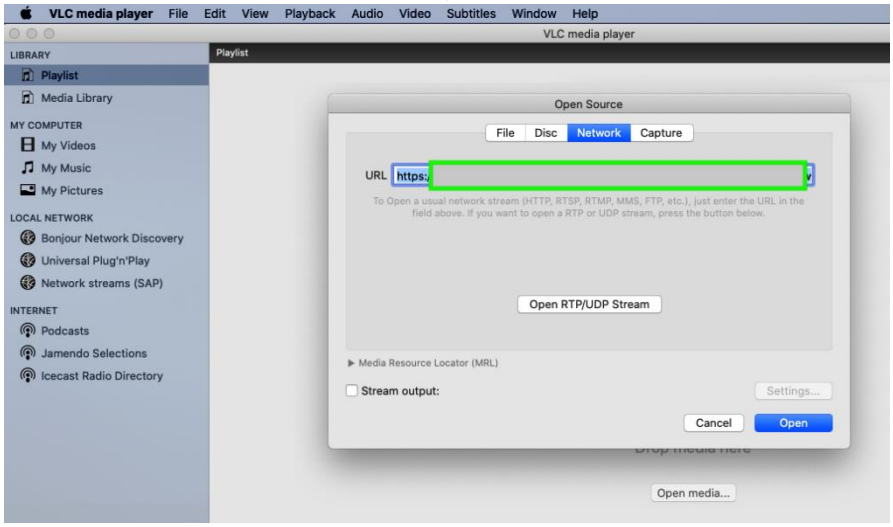
Not defined

Avail suppression mode

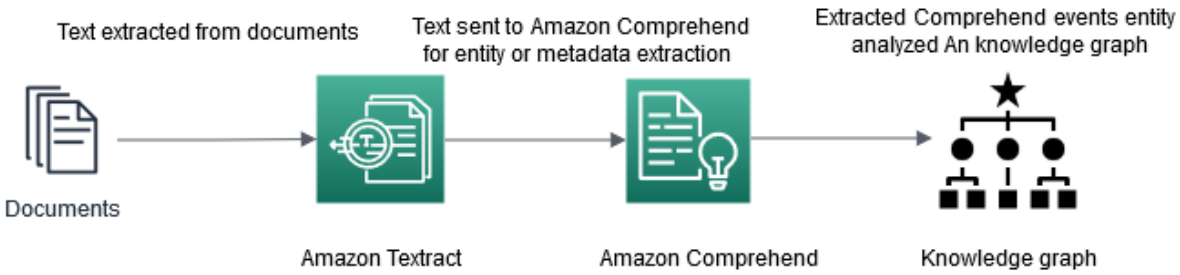
OFF

Avail suppression value

Not defined

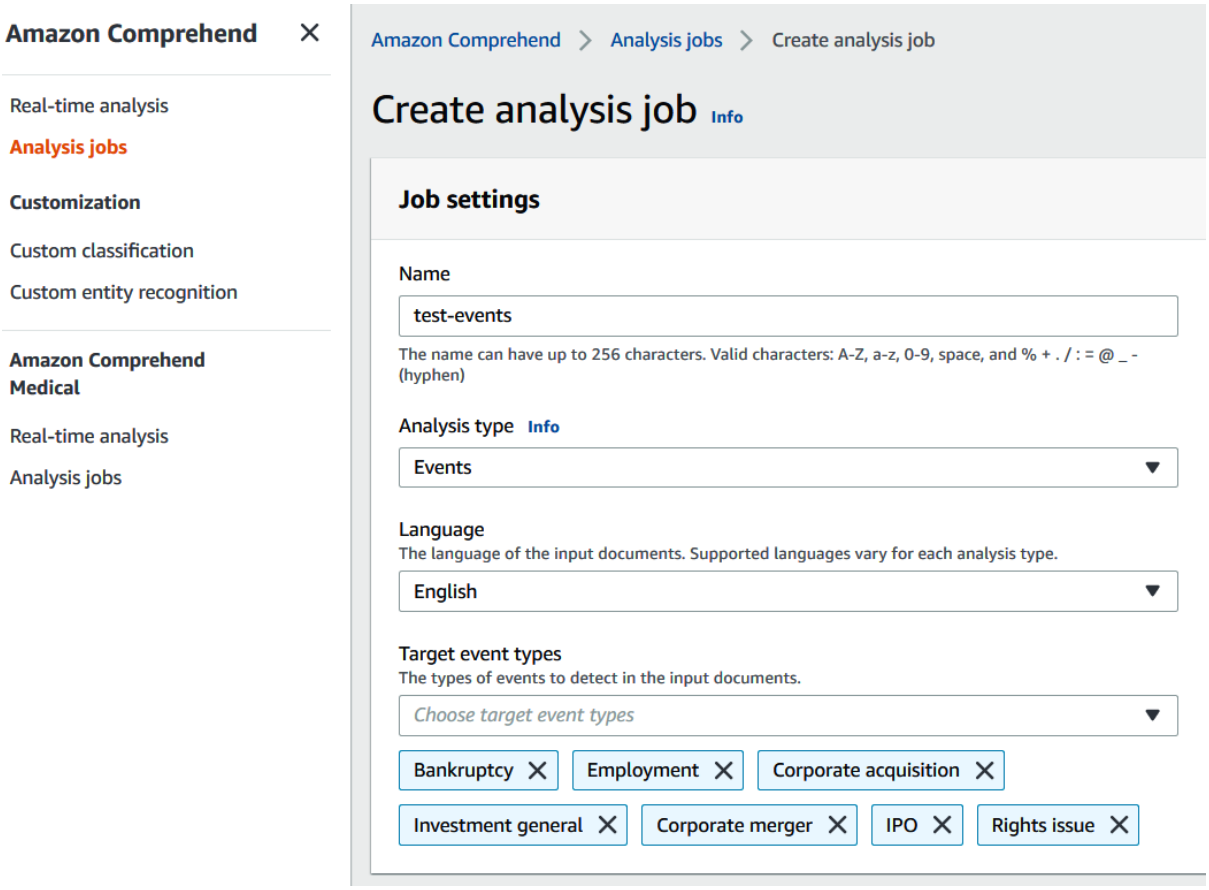


Chapter 9: Extracting Metadata from Financial Documents




```
Started job with id: 5e1770681a137478b04edd3016151865a257a1d8a369f9b35e74157d830ef3e1
Job status: IN_PROGRESS
Job status: IN_PROGRESS
Job status: IN_PROGRESS
```

```
Amazon (NASDAQ:AMZN) and Whole Foods Market, Inc. (NASDAQ:WFM) today announced that they
have entered into a definitive merger agreement under which Amazon will acquire Whole Foods
Market
for $42 per share in an all-cash transaction valued at approximately $13.7 billion, includi
ng Whole Foods
Market's net debt. "Millions of people love Whole Foods Market because they offer the best
natural
and organic foods, and they make it fun to eat healthy," said Jeff Bezos, Amazon founder an
d CEO.
"Whole Foods Market has been satisfying, delighting and nourishing customers for nearly fou
r decades -
they're doing an amazing job and we want that to continue." "This partnership presents an o
pportunity
to maximize value for Whole Foods Market's shareholders, while at the same time extending o
ur
mission and bringing the highest quality, experience, convenience and innovation to our cus
```



Input data Info

Data source

- (•) My documents
- ( ) Example documents
  Example documents are available only in English

S3 location
Paste the URL of an input data file in S3, or select a bucket or folder location in S3.

s3://[redacted]/sample_finance_ Browse S3

Input format - *optional* Info

One document per line

Output data Info

S3 location
Paste the URL of a bucket or folder location in S3, or select a bucket or folder location in S3.

s3://[redacted]/ Browse S3

Encryption Info

Access permissions Info

IAM role

- ( ) Use an existing IAM role
- (•) Create an IAM role

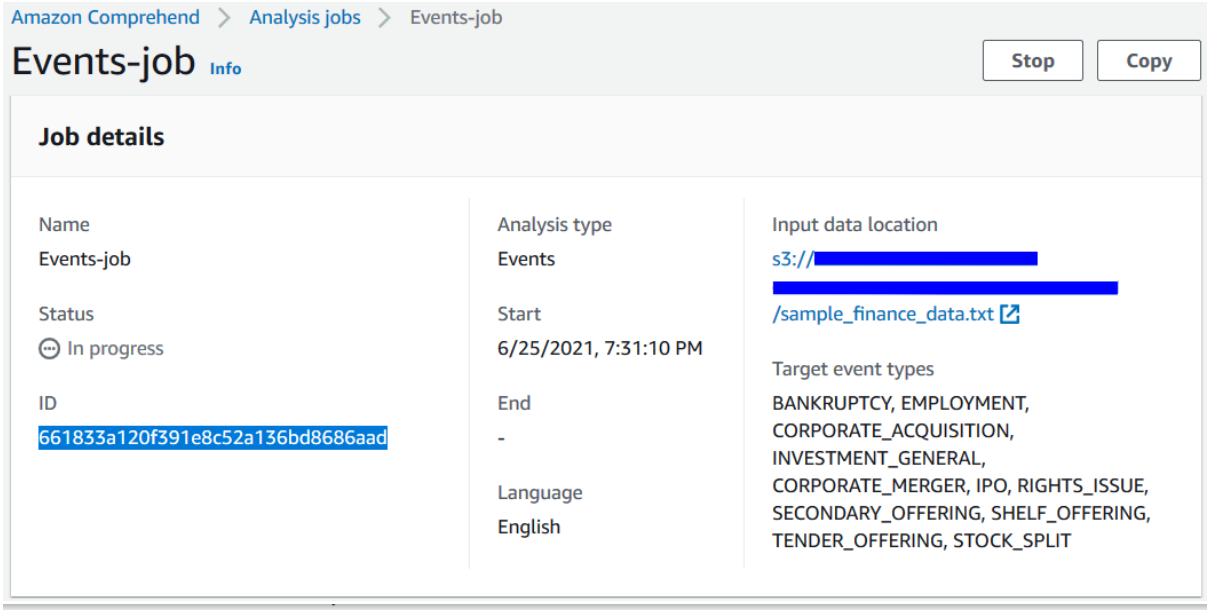
Permissions to access
Your role will have access to these resources.

Input and Output S3 buckets

Name suffix
Your roles will be prefixed with "AmazonComprehendServiceRole-". By clicking "Create Job" you are authorizing creation of this role

events-role

Cancel Create job




```
    'Type': 'EMPLOYMENT'}],
  'Type': 'EMPLOYMENT'},
 {'Arguments': [{'EntityIndex': 2, 'Role': 'INVESTEE', 'Score': 0.999677},
   {'EntityIndex': 11, 'Role': 'DATE', 'Score': 0.977672}],
  'Triggers': [{'BeginOffset': 2795,
    'EndOffset': 2806,
    'GroupScore': 1.0,
    'Score': 0.660339,
    'Text': 'transaction',
    'Type': 'CORPORATE_ACQUISITION'},
   {'BeginOffset': 2960,
    'EndOffset': 2971,
    'GroupScore': 0.999965,
    'Score': 0.678594,
    'Text': 'transaction',
    'Type': 'CORPORATE_ACQUISITION'}],
  'Type': 'CORPORATE_ACQUISITION'}],
'File': 'sample_finance_data.txt',
'Line': 0}
```

Out[34]:

```
[{'BeginOffset': 160,
  'EndOffset': 167,
  'GroupScore': 1.0,
  'Score': 0.999959,
  'Text': 'acquire',
  'Type': 'CORPORATE_ACQUISITION'},
 {'BeginOffset': 210,
  'EndOffset': 213,
  'GroupScore': 0.999974,
  'Score': 0.828479,
  'Text': 'all',
  'Type': 'CORPORATE_ACQUISITION'},
 {'BeginOffset': 219,
  'EndOffset': 230,
  'GroupScore': 0.501301,
  'Score': 0.98162,
  'Text': 'transaction',
  'Type': 'CORPORATE_ACQUISITION'},
```


Out[35]:

```
[{'EntityIndex': 5, 'Role': 'AMOUNT', 'Score': 0.998241},
 {'EntityIndex': 4, 'Role': 'DATE', 'Score': 0.994679},
 {'EntityIndex': 2, 'Role': 'INVESTEE', 'Score': 0.999588},
 {'EntityIndex': 0, 'Role': 'INVESTOR', 'Score': 0.999565},
 {'EntityIndex': 3, 'Role': 'INVESTEE', 'Score': 0.975579},
 {'EntityIndex': 8, 'Role': 'DATE', 'Score': 0.994265}]
```


Out[36]:

```
[{'BeginOffset': 190,
  'EndOffset': 193,
  'GroupScore': 1.0,
  'Score': 0.998241,
  'Text': '$42',
  'Type': 'MONETARY_VALUE'},
 {'BeginOffset': 255,
  'EndOffset': 268,
  'GroupScore': 0.871696,
  'Score': 0.999512,
  'Text': '$13.7 billion',
  'Type': 'MONETARY_VALUE'}]
```

Amazon ORGANIZATION (NASDAQ:AMZN STOCK_CODE) and Whole Foods Market, Inc. ORGANIZATION (NASDAQ:WFM STOCK_CODE) today DATE announced that theyhave entered into a definitive merger CORPORATE_MERGER agreement under which Amazon ORGANIZATION will acquire CORPORATE_ACQUISITION Whole Foods Marketfor ORGANIZATION $42 MONETARY_VALUE per share in an all CORPORATE_ACQUISITION -cash transaction CORPORATE_ACQUISITION valued at approximately $13.7 billion MONETARY_VALUE , including Whole FoodsMarket ORGANIZATION 's net debt. "Millions of people love Whole Foods Market ORGANIZATION because they ORGANIZATION offer the best naturaland organic foods, and they PERSON make it fun to eat healthy," said Jeff Bezos, Amazon ORGANIZATION founder and CEO." Whole Foods Market ORGANIZATION has been satisfying, delighting and nourishing customers PERSON for nearly four decades - they ORGANIZATION 're doing an amazing job and we ORGANIZATION want that to continue." "This partnership CORPORATE_MERGER presents an opportunityto maximize value for Whole Foods Market ORGANIZATION 's shareholders, while at the same time extending ourmission and bringing the highest quality, experience, convenience and innovation to our customers,"said John Mackey PERSON , Whole Foods Market ORGANIZATION co-founder PERSON_TITLE and CEO PERSON_TITLE . Whole Foods Market ORGANIZATION will continue tooperate stores under the Whole Foods Market ORGANIZATION brand and source from trusted vendors and partnersaround the world. John Mackey PERSON will remain EMPLOYMENT as CEO PERSON_TITLE of Whole Foods Market ORGANIZATION and Whole Foods Market'sheadquarters will stay in Austin, Texas. Completion of the transaction CORPORATE_ACQUISITION is subject to approval by

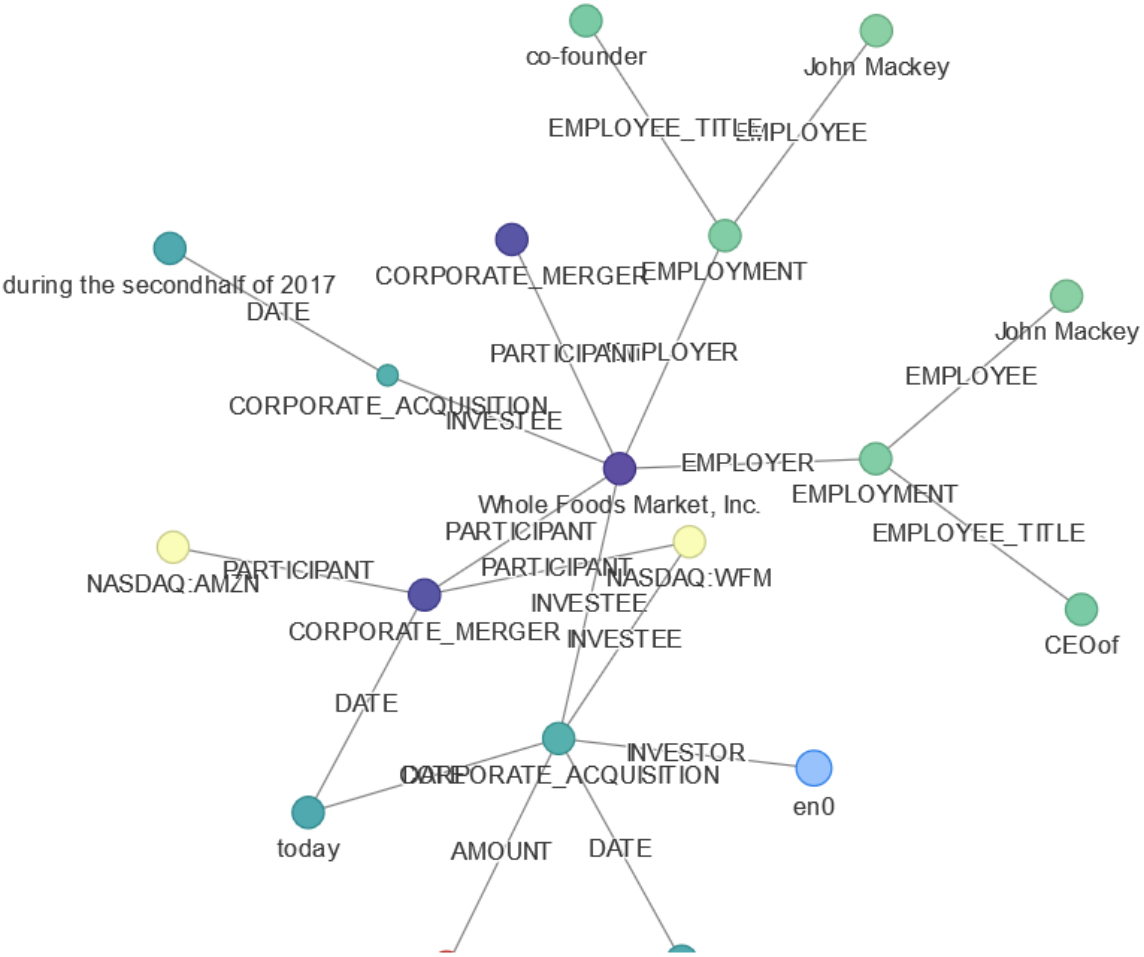
Out[49]:

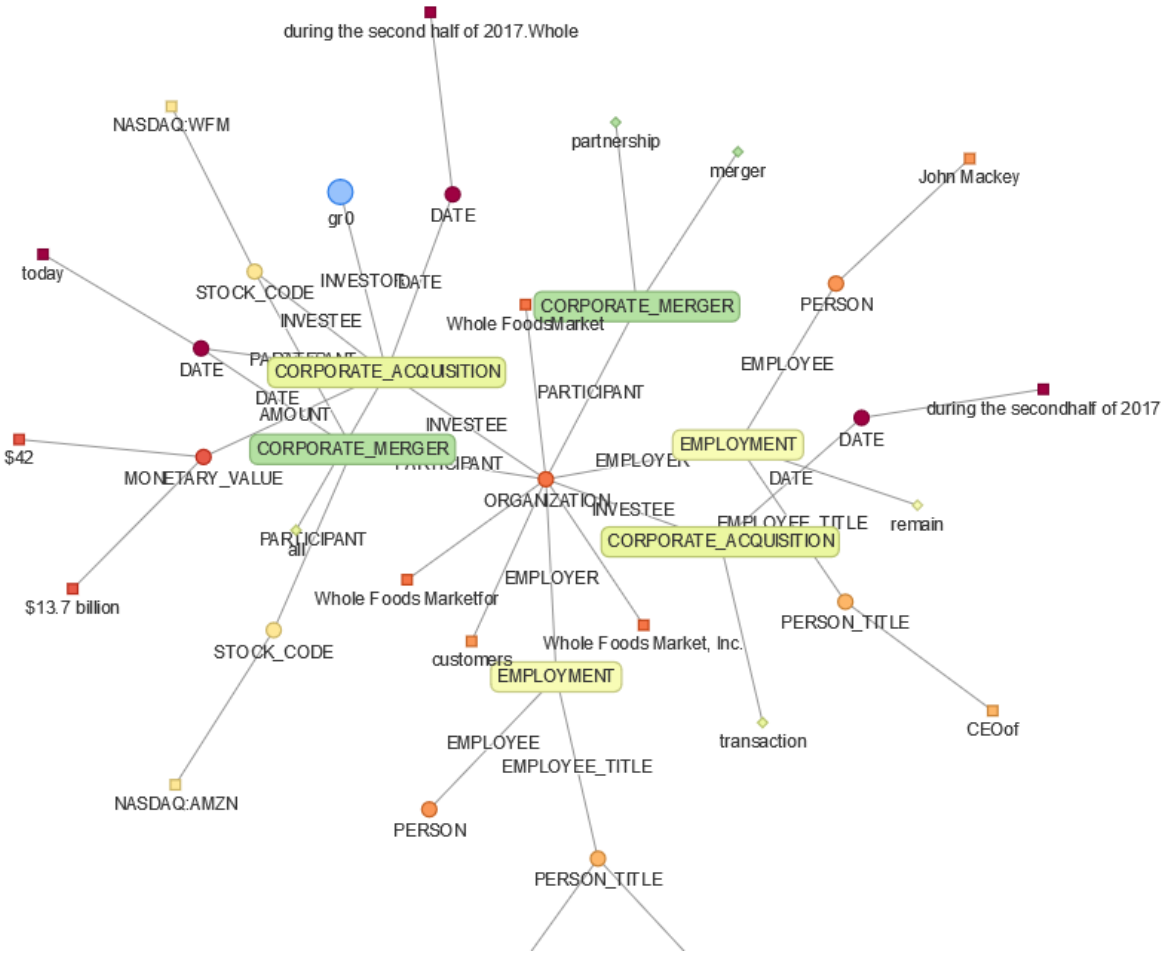
	EventIndex	EntityIndex	Role	ScoreEvent	BeginOffsetEvent	EndOffsetEvent	GroupScoreEvent	TextEvent	TypeEvent	Begin
0	0	4	DATE	0.999686	119	125	1.000000	merger	CORPORATE_MERGER	
1	0	4	DATE	0.999686	119	125	1.000000	merger	CORPORATE_MERGER	
2	0	4	DATE	0.999914	654	665	0.999946	partnership	CORPORATE_MERGER	
3	0	4	DATE	0.999914	654	665	0.999946	partnership	CORPORATE_MERGER	
4	1	4	DATE	0.999959	160	167	1.000000	acquire	CORPORATE_ACQUISITION	
...	...	...	...	...	...	...	...	...	...	
675	4	9	EMPLOYEE	0.999974	2677	2683	1.000000	remain	EMPLOYMENT	
676	4	9	EMPLOYEE	0.999974	2677	2683	1.000000	remain	EMPLOYMENT	
677	4	10	EMPLOYEE_TITLE	0.999974	2677	2683	1.000000	remain	EMPLOYMENT	
678	5	11	DATE	0.660339	2795	2806	1.000000	transaction	CORPORATE_ACQUISITION	
679	5	11	DATE	0.678594	2960	2971	0.999965	transaction	CORPORATE_ACQUISITION	

680 rows × 15 columns

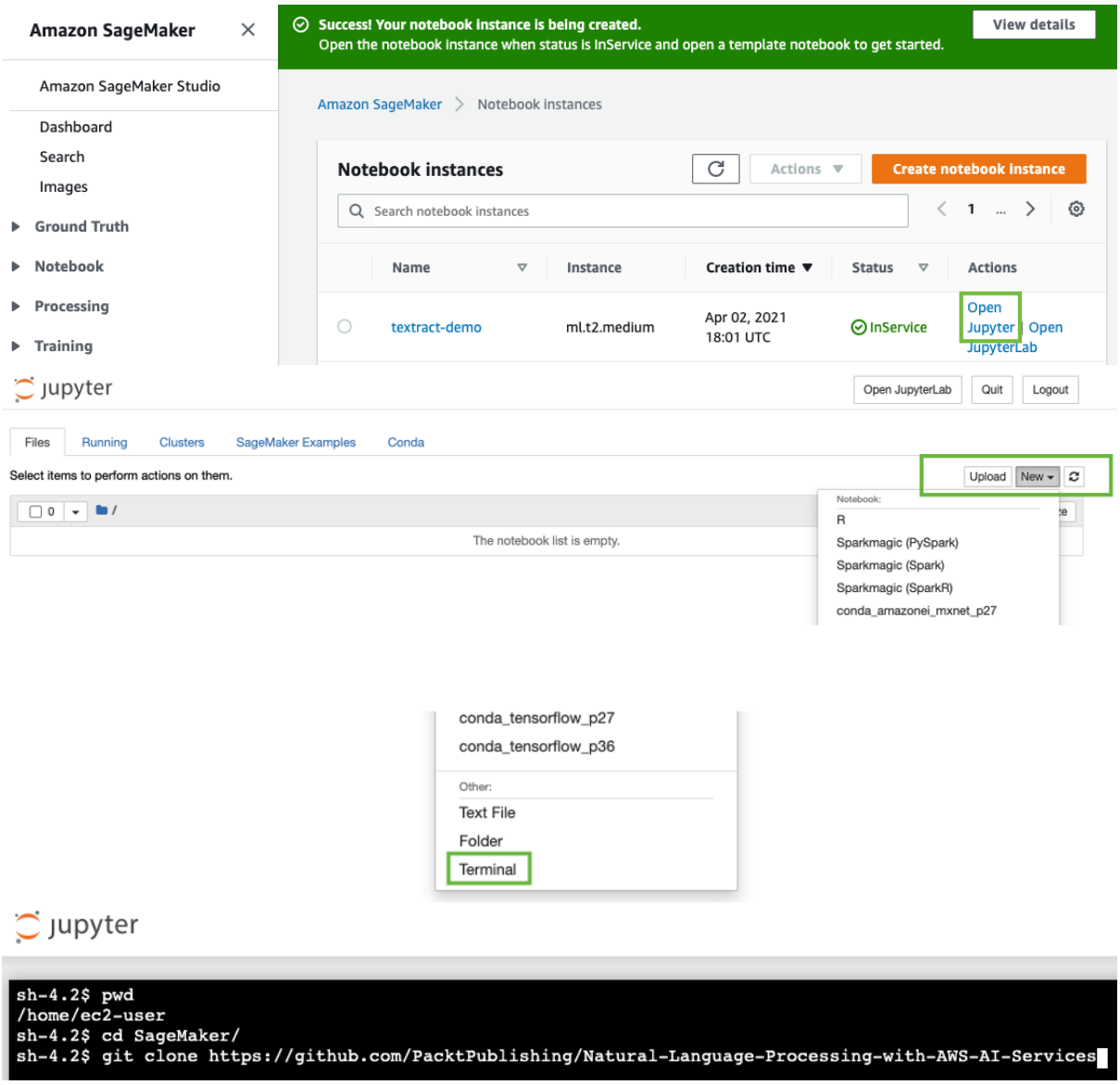
Out[51]:

EventType	Triggers	DATE	PARTICIPANT	INVESTEE	AMOUNT	INVESTOR	EMPLOYER	EMPLOYEE	EMPLOYEE_TITLE
CORPORATE_MERGER	{partnership, merger}	{today}	{NASDAQ:AMZN, they, WholeFoods Market, Foods M...						
CORPORATE_ACQUISITION	{all, transaction, acquire}	{today, during the second half of 2017.Whole}		{they, WholeFoods Market, Foods Market, Inc., ...	{ 13.7billion , 42}	{Amazon, www.amazon.com/about, we, AmazonNews}			
EMPLOYMENT	{remain}						{they, WholeFoods Market, Foods Market, Inc., ...	{John Mackey}	{CEO, co-founder}
CORPORATE_MERGER	{partnership, merger}		{they, WholeFoods Market, Foods Market, Inc., ...						
EMPLOYMENT	{remain}						{they, WholeFoods Market, Foods Market, Inc., ...	{John Mackey}	{CEOof}


co-founder
John Mackey
EMPLOYEE_TITLE
EMPLOYEE
during the secondhalf of 2017
DATE
CORPORATE_MERGER
EMPLOYMENT
PARTICIPANT
EMPLOYER
John Mackey
EMPLOYEE
CORPORATE_ACQUISITION
INVESTEE
EMPLOYER
EMPLOYMENT
Whole Foods Market, Inc.
PARTICIPANT
EMPLOYEE_TITLE
NASDAQ:AMZN
PARTICIPANT
PARTICIPANT
NASDAQ:WFM
INVESTEE
CORPORATE_MERGER
INVESTEE
CEOof
DATE
INVESTOR
CORPORATE_ACQUISITION
en0
today
AMOUNT
DATE


during the second half of 2017.Whole
NASDAQ:WFM
partnership
merger
John Mackey
gr0
DATE
today
STOCK_CODE
INVESTOR
DATE
CORPORATE_MERGER
PERSON
INVESTEE
Whole FoodsMarket
DATE
CORPORATE_ACQUISITION
EMPLOYEE
PARTICIPANT
DATE
during the secondhalf of 2017
AMOUNT
INVESTEE
EMPLOYMENT
DATE
$42
CORPORATE_MERGER
EMPLOYER
MONETARY_VALUE
DATE
ORGANIZATION
INVESTEE
EMPLOYEE_TITLE
remain
PARTICIPANT
all
CORPORATE_ACQUISITION
EMPLOYER
Whole Foods Marketfor
$13.7 billion
PERSON_TITLE
STOCK_CODE
customers
Whole Foods Market, Inc.
EMPLOYMENT
transaction
CEOof
EMPLOYEE
EMPLOYEE_TITLE
NASDAQ:AMZN
PERSON
PERSON_TITLE

Chapter 10: Reducing Localization Costs with Machine Translation



Family Bank Holdings

Date: 06/29/2021

Who we are and what we do

A wholly owned subsidiary of LiveRight, we are the nation's largest bank for SMB owners and cooperative societies, with more than 4500 branches spread across the nation, servicing more than 5 million customers and continuing to grow. We offer a number of lending products to our customers including checking and savings accounts, lending, credit cards, deposits, insurance, IRA and more. Started in 1787 as a family owned business providing low interest loans for farmers struggling with poor harvests, LiveRight helped these farmers design long distance water channels from lakes in neighboring districts to their lands. The initial success helped these farmers invest their wealth in LiveRight and later led to our cooperative range of products that allowed farmers to own a part of LiveRight. In 1850 we moved our HeadQuarters to New York city to help build the economy of our nation by providing low interest lending products to small to medium business owners looking to start or expand their business. From 2 branches then to 4500 branches today, the trust of our customers helped us grow to become the nation's largest SMB bank.

Beteiligungen für Familienbanken

Date: 06/29/2021

Wer wir sind und was wir machen

Als hundertprozentige Tochtergesellschaft von Liveright sind wir die größte Bank des Landes für SMB-Eigentümer und Genossenschaftsgesellschaften mit mehr als 4500 Filialen im ganzen Land, betreuen mehr als 5 Millionen Kunden und wachsen weiter. Wir bieten unseren Kunden eine Reihe von Kreditprodukten an, darunter Scheck- und Sparkonten, Kredite, Kreditkarten, Einlagen, Versicherungen, IRA und mehr. LiveRight wurde 1787 als Familienunternehmen gegründet, das Zinsdarlehen für Landwirte, die mit schlechten Ernten zu kämpfen hatten, und half diesen Landwirten, Fernwasserkanäle von Seen in benachbarten Bezirken zu entwerfen in ihr Land. Der anfängliche Erfolg half diesen Landwirten, ihr Vermögen in Liveright zu investieren, und führte später zu unserer kooperativen Produktpalette, die es den Landwirten ermöglichte, einen Teil von LiverRight zu besitzen. Im Jahr 1850 verlegten wir unseren Hauptsitz nach New York City, um zum Aufbau der Wirtschaft unseres Landes beizutragen, indem wir kleinen bis mittleren Unternehmern, die ihr Geschäft beginnen oder ausbauen möchten, Produkte mit niedrigen Zinsen zur Verfügung stellen. Von 2 Filialen bis hin zu 4500 Filialen heute hat uns das Vertrauen unserer Kunden geholfen, zur größten KMB-Bank des Landes zu werden.

Participaciones bancarias familiares

Date: 06/29/2021

Quiénes somos y qué hacemos

Una filial de propiedad total de Liveright, somos el banco más grande del país para propietarios de PYMES y sociedades cooperativas, con más de 4500 sucursales repartidas por todo el país, prestando servicios a más de 5 millones de clientes y sigue creciendo. Ofrecemos una serie de productos de préstamo a nuestros clientes, incluidas cuentas de cheques y ahorros, préstamos, tarjetas de crédito, depósitos, seguros, IRA y más. Comenzado en 1787 como una empresa familiar que otorgaba préstamos a bajo interés para los agricultores que luchan con cosechas deficientes, Liveright ayudó a estos agricultores a diseñar canales de agua de larga distancia desde lagos de los distritos vecinos. a sus tierras. El éxito inicial ayudó a estos agricultores a invertir su riqueza en Liveright y posteriormente llevó a nuestra gama cooperativa de productos que permitieron a los agricultores poseer una parte de Liveright. En 1850 trasladamos nuestra sede central a la ciudad de Nueva York para ayudar a construir la economía de nuestra nación proporcionando productos de préstamos de bajo interés a propietarios de pequeñas y medianas empresas que desean iniciar o expandir su negocio. Desde 2 sucursales hasta 4500 sucursales en la actualidad, la confianza de nuestros clientes nos ayudó a crecer hasta convertirnos en el banco de pymes más grande del país.

फॅमिली बैंक होल्डिंग्स

Date: 06/29/2021

हम कौन हैं और हम क्या करते हैं

Liveright की एक पूर्ण स्वामित्व वाली सहायक कंपनी, हम एसएमबी मालिकों और सहकारी समितियों के लिए देश का सबसे बड़ा बैंक हैं, देश भर में 4500 से अधिक शाखाएं फैली हुई हैं, 5 मिलियन से अधिक ग्राहकों की सेवा कर रही हैं और बढ़ती जा रही हैं। हम अपने ग्राहकों को चेकिंग और बचत खाते, उधार, क्रेडिट कार्ड, जमा, बीमा, आईआरए और बहुत कुछ सहित कई उधार उत्पाद प्रदान करते हैं। 1787 में शुरू किया गया एक परिवार के स्वामित्व वाले व्यवसाय के रूप में गरीब फसल के साथ संघर्ष करने वाले किसानों के लिए कम ब्याज ऋण प्रदान करते हुए, Liveright ने इन किसानों को पड़ोसी जिलों में झीलों से लंबी दूरी के पानी के चैनल डिजाइन करने में मदद की उनकी भूमि के लिए। प्रारंभिक सफलता ने इन किसानों को Liveright में अपनी संपत्ति का निवेश करने में मदद की और बाद में हमारे सहकारी उत्पादों की श्रृंखला का नेतृत्व किया जिसने किसानों को Liveright का एक हिस्सा रखने की अनुमति दी। 1850 में हमने अपने मुख्यालय को न्यूयॉर्क शहर में स्थानांतरित कर दिया ताकि छोटे से मध्यम व्यापार मालिकों को कम ब्याज उधार देने वाले उत्पाद प्रदान करके हमारे देश की अर्थव्यवस्था का निर्माण करने में मदद मिल सके, जो अपने व्यवसाय को शुरू करने या विस्तारित करने के लिए देख रहे हैं। आज दो शाखाओं से लेकर 4500 शाखाओं तक, हमारे ग्राहकों के विश्वास ने हमें देश का सबसे बड़ा एसएमबी बैंक बनने में मदद की।

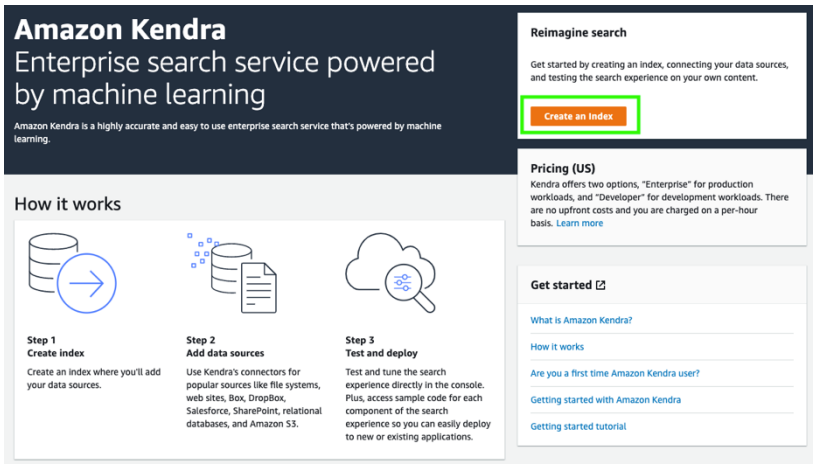
குடும்ப வங்கி ஹோல்டிங்ஸ்

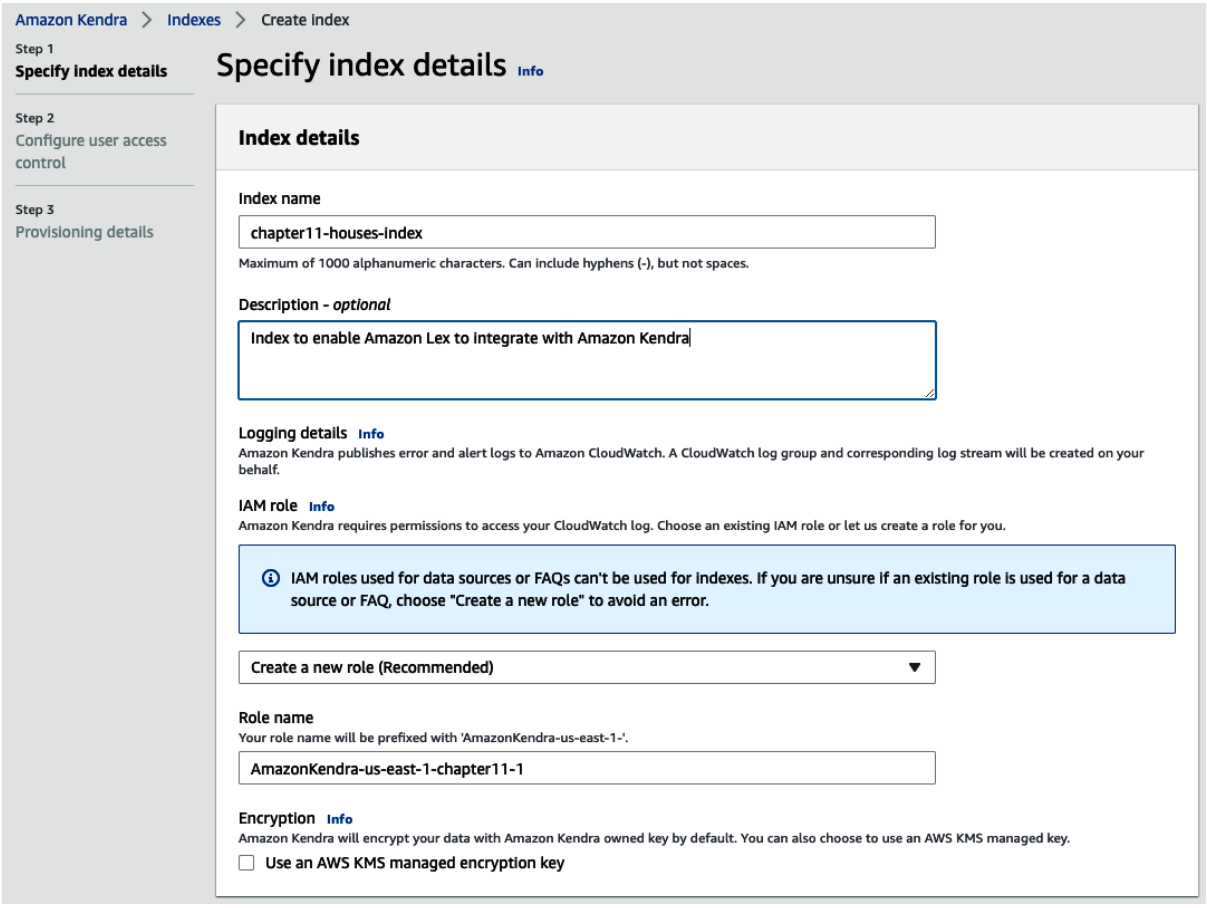
Date: 06/29/2021

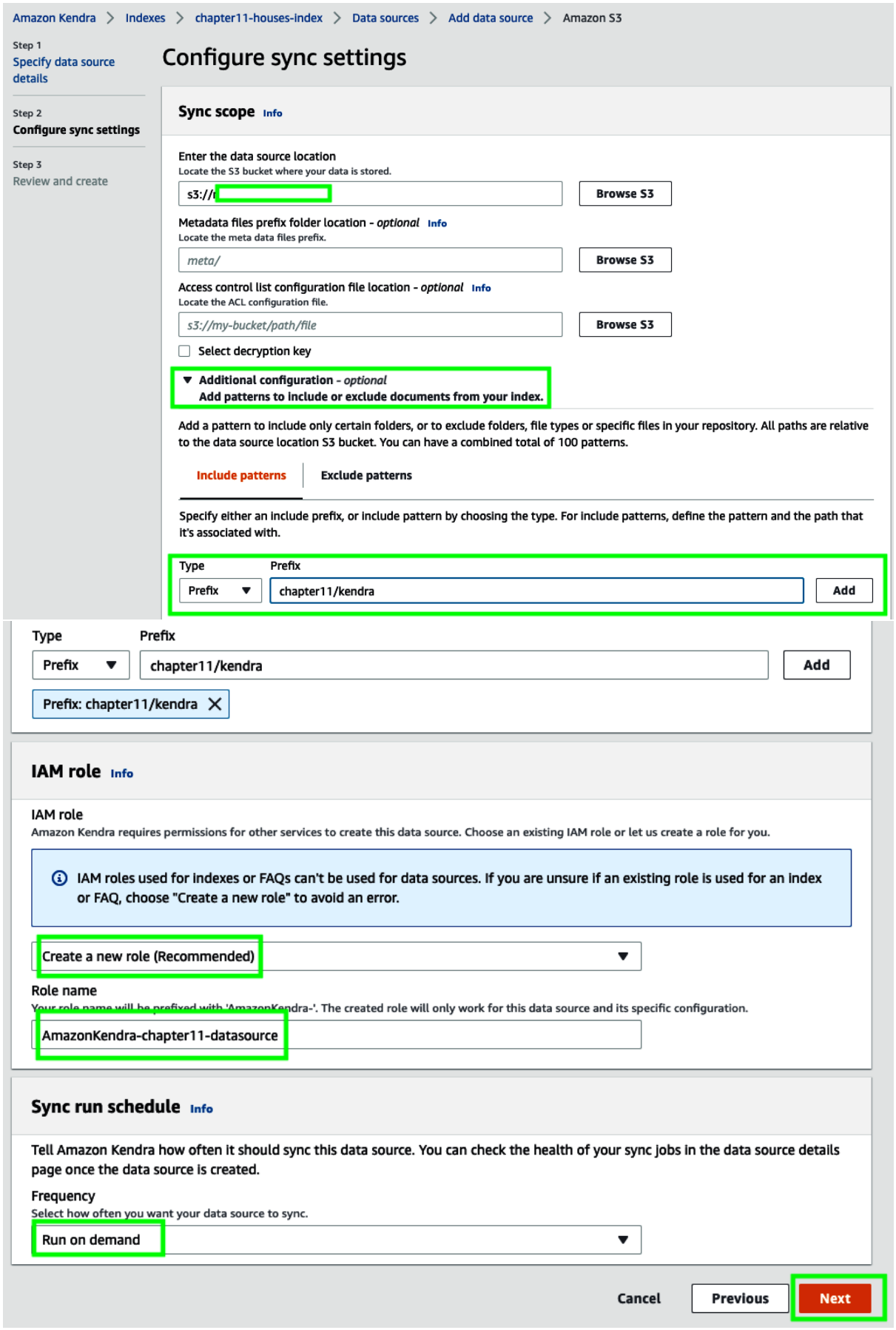
நாம் யார், நாம் என்ன செய்கிறோம்

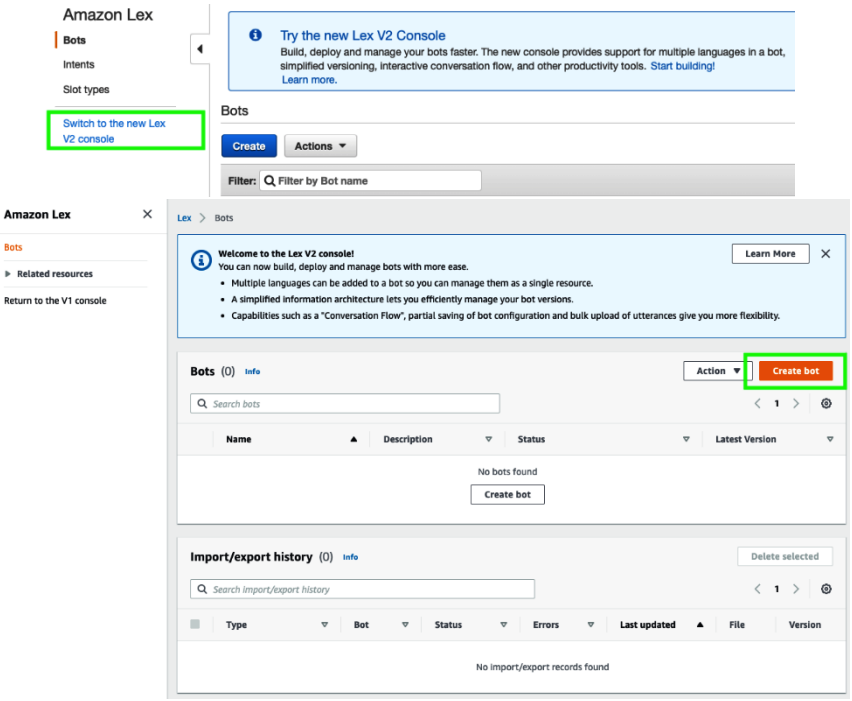
லிவரைலைட் முழுவதுமாக சொந்தமான துணை நிறுவனமாக நாங்கள் SMB உரிமையாளர்கள் மற்றும் கூட்டுறவு சங்கங்களுக்கு நாட்டின் மிகப்பெரிய வங்கியாக இருக்கிறோம், மேலும் 4500 க்கும் மேற்பட்ட கிளைகள் நாடு முழுவதும் பரவியுள்ளன, மேலும் 5 மில்லியன் வாடிக்கையாளர்களுக்கு சேவை செய்து வளர்ந்தன. நாங்கள் சோதனை மற்றும் சேமிப்பு கணக்குகள், கடன், கடன் அட்டைகள், வைப்பு, காப்பீடு, IRA மற்றும் பல உட்பட எங்கள் வாடிக்கையாளர்களுக்கு கடன் வழங்கும் பல தயாரிப்புகளை வழங்குகிறோம். ஏழை அறுவடைகளுடன் போராடும் விவசாயிகளுக்கு குறைந்த வட்டி கடன்களை வழங்கும் குடும்பத்திற்கு சொந்தமான வணிகமாக 1787 ஆம் ஆண்டில் தொடங்கப்பட்டது, லிவரைட் இந்த விவசாயிகள் அண்டை மாவட்டங்களில் ஏரிகளிலிருந்து நீண்ட தூர நீர் சேனல்களை வடிவமைக்க உதவியது தங்கள் நிலங்களுக்கு. ஆரம்ப வெற்றி இந்த விவசாயிகள் லிவரைட்டில் தங்கள் செல்வத்தை முதலீடு செய்ய உதவியது, பின்னர் விவசாயிகள் லிவர்லைட் ஒரு பகுதியை சொந்தமாக அனுமதித்த எங்கள் கூட்டுறவு வரம்பிற்கு வழிவகுத்தது. 1850 ஆம் ஆண்டில் நாங்கள் எங்கள் தலைமையகத்தை நியூயார்க் நகரத்திற்கு நகர்த்தினோம், எங்கள் நாட்டின் பொருளாதாரத்தை உருவாக்க உதவுவதற்காக குறைந்த வட்டி கடன் வழங்கும் பொருட்களை சிறிய முதல் நடுத்தர வணிக உரிமையாளர்களுக்கு தங்கள் வியாபாரத்தை ஆரம்பிக்கவோ அல்லது விரிவுபடுத்தவோ விரும்பினோம். இன்று 2 கிளைகள் முதல் 4500 கிளைகள் வரை, எங்கள் வாடிக்கையாளர்களின் நம்பிக்கை நாட்டின் மிகப்பெரிய SMB வங்கியாக வளர உதவியது.

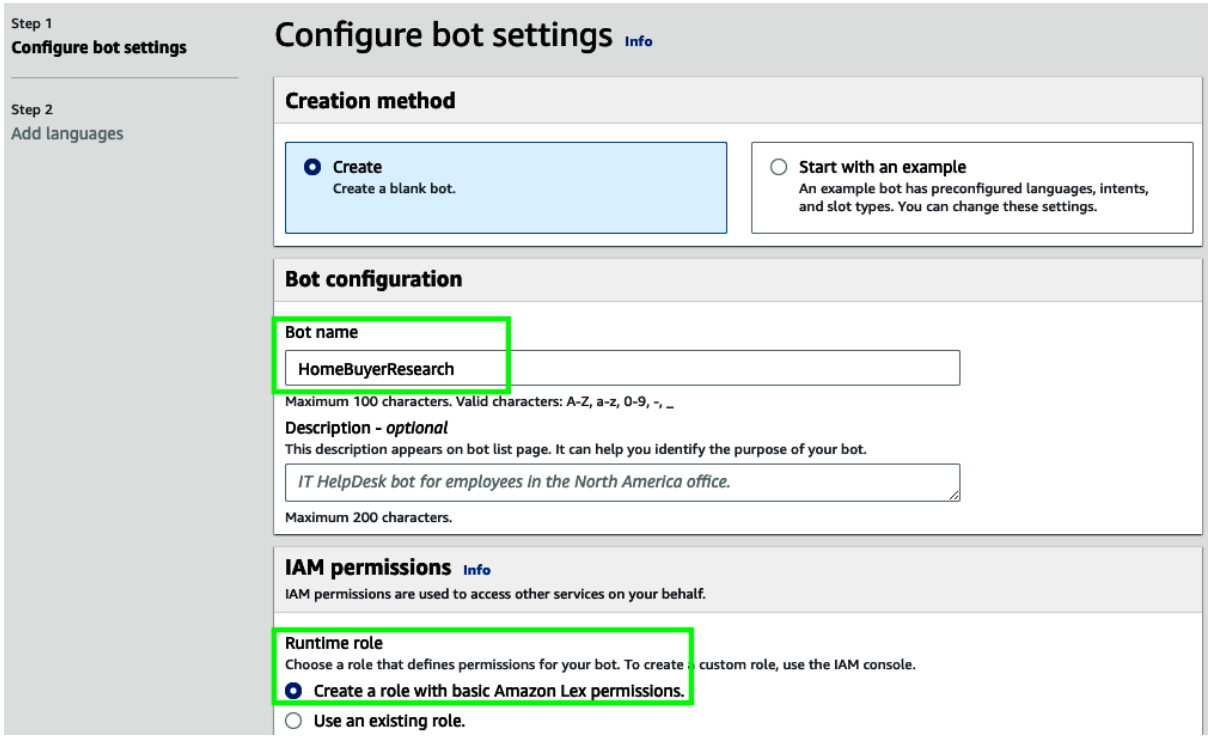
Chapter 11: Using Chatbots for Querying Documents

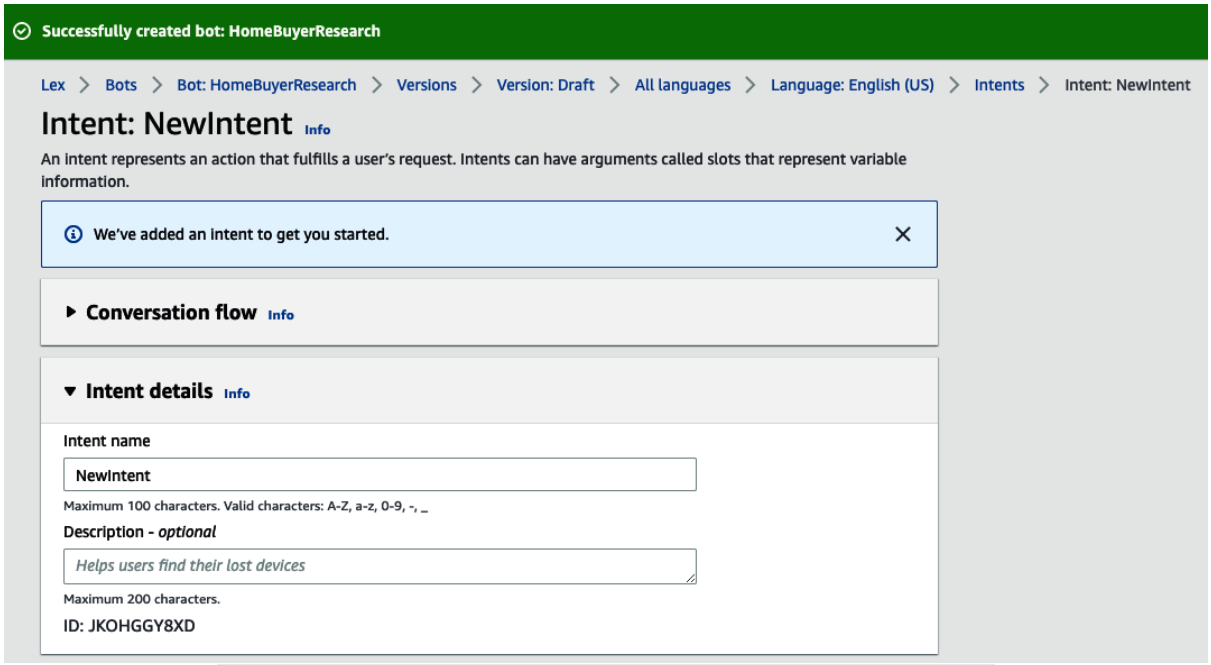


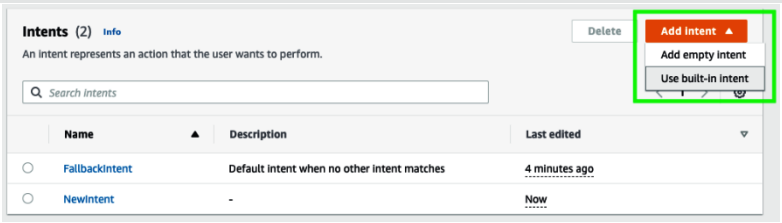


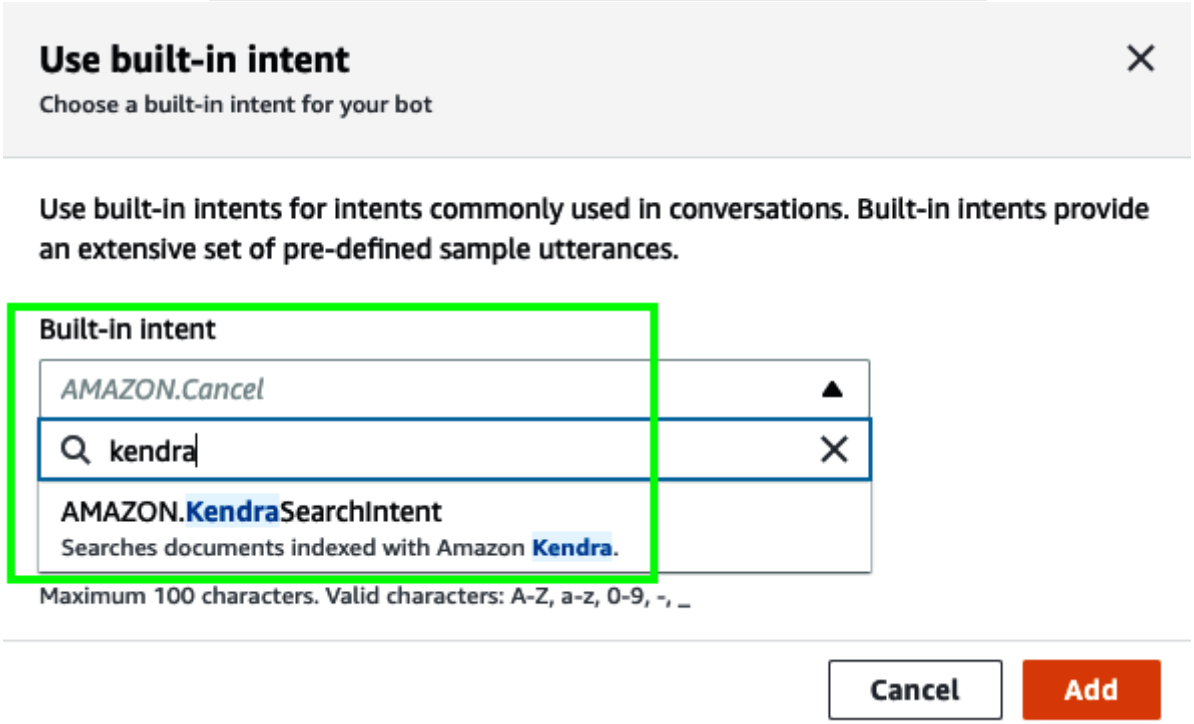

Amazon Kendra > Indexes > chapter11-houses-index > Data sources > Add data source > Amazon S3
Step 1
Specify data source details
Step 2
Configure sync settings
Step 3
Review and create
Configure sync settings
Sync scope Info
Enter the data source location
Locate the S3 bucket where your data is stored.
s3://
Browse S3
Metadata files prefix folder location - optional Info
Locate the meta data files prefix.
meta/
Browse S3
Access control list configuration file location - optional Info
Locate the ACL configuration file.
s3://my-bucket/path/file
Browse S3
Select decryption key
Additional configuration - optional
Add patterns to include or exclude documents from your index.
Add a pattern to include only certain folders, or to exclude folders, file types or specific files in your repository. All paths are relative to the data source location S3 bucket. You can have a combined total of 100 patterns.
Include patterns
Exclude patterns
Specify either an include prefix, or include pattern by choosing the type. For include patterns, define the pattern and the path that it's associated with.
Type
Prefix
Prefix
chapter11/kendra
Add
Type
Prefix
Prefix
chapter11/kendra
Add
Prefix: chapter11/kendra
IAM role Info
IAM role
Amazon Kendra requires permissions for other services to create this data source. Choose an existing IAM role or let us create a role for you.
IAM roles used for indexes or FAQs can't be used for data sources. If you are unsure if an existing role is used for an index or FAQ, choose "Create a new role" to avoid an error.
Create a new role (Recommended)
Role name
Your role name will be prefixed with 'AmazonKendra-'. The created role will only work for this data source and its specific configuration.
AmazonKendra-chapter11-datasource
Sync run schedule Info
Tell Amazon Kendra how often it should sync this data source. You can check the health of your sync jobs in the data source details page once the data source is created.
Frequency
Select how often you want your data source to sync.
Run on demand
Cancel
Previous
Next

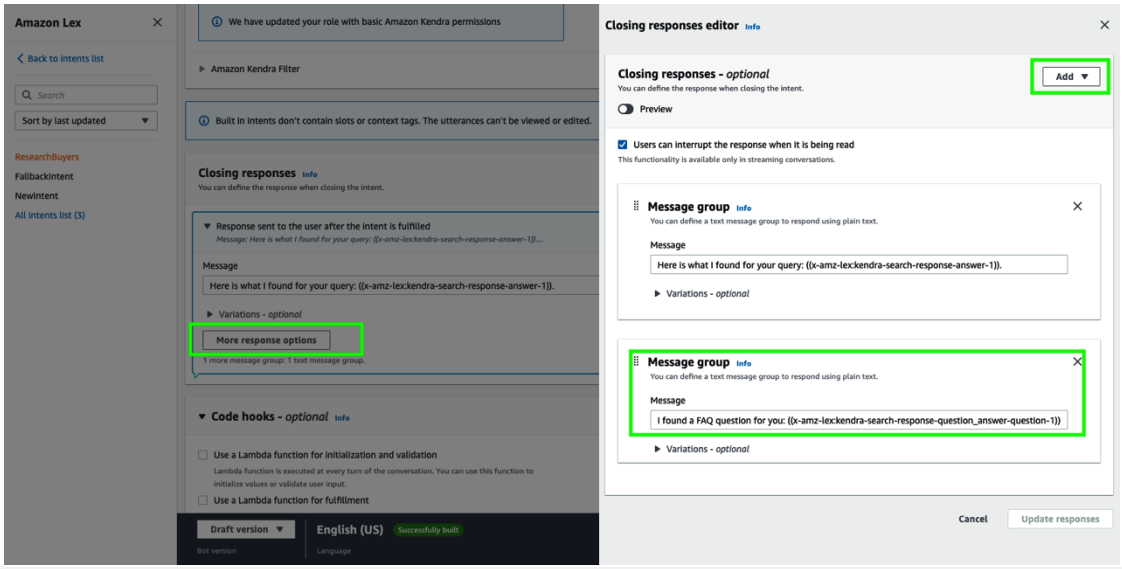

Amazon Lex
Bots
Intents
Slot types
Switch to the new Lex V2 console
Try the new Lex V2 Console
Build, deploy and manage your bots faster. The new console provides support for multiple languages in a bot, simplified versioning, interactive conversation flow, and other productivity tools. Start building!
Learn more.
Bots
Create
Actions
Filter: Filter by Bot name
Amazon Lex
Bots
Related resources
Return to the V1 console
Lex > Bots
Welcome to the Lex V2 console!
You can now build, deploy and manage bots with more ease.
Multiple languages can be added to a bot so you can manage them as a single resource.
A simplified information architecture lets you efficiently manage your bot versions.
Capabilities such as a "Conversation Flow", partial saving of bot configuration and bulk upload of utterances give you more flexibility.
Learn More
Bots (0) Info
Action
Create bot
Search bots
Name
Description
Status
Latest Version
No bots found
Create bot
Import/export history (0) Info
Delete selected
Search import/export history
Type
Bot
Status
Errors
Last updated
File
Version
No import/export records found

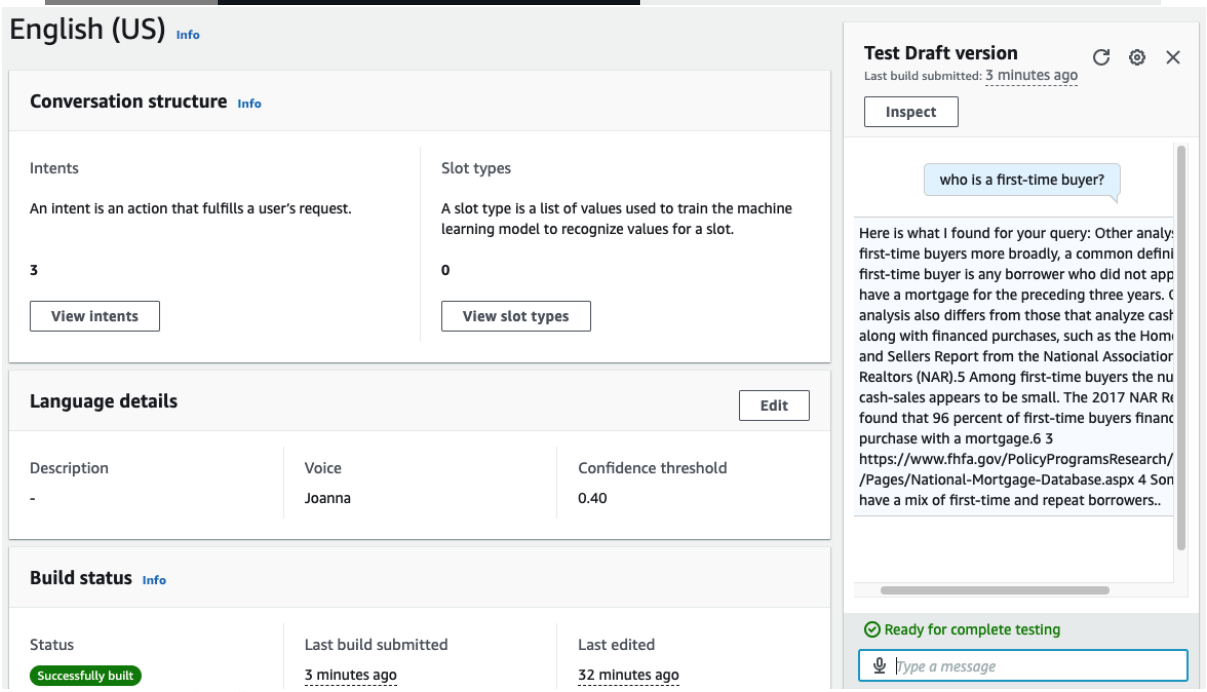

Step 1
Configure bot settings
Step 2
Add languages
Configure bot settings Info
Creation method
Create
Create a blank bot.
Start with an example
An example bot has preconfigured languages, intents, and slot types. You can change these settings.
Bot configuration
Bot name
HomeBuyerResearch
Maximum 100 characters. Valid characters: A-Z, a-z, 0-9, -, _
Description - optional
This description appears on bot list page. It can help you identify the purpose of your bot.
IT HelpDesk bot for employees in the North America office.
Maximum 200 characters.
IAM permissions Info
IAM permissions are used to access other services on your behalf.
Runtime role
Choose a role that defines permissions for your bot. To create a custom role, use the IAM console.
Create a role with basic Amazon Lex permissions.
Use an existing role.


Successfully created bot: HomeBuyerResearch
Lex > Bots > Bot: HomeBuyerResearch > Versions > Version: Draft > All languages > Language: English (US) > Intents > Intent: NewIntent
Intent: NewIntent Info
An intent represents an action that fulfills a user's request. Intents can have arguments called slots that represent variable information.
We've added an intent to get you started.
Conversation flow Info
Intent details Info
Intent name
NewIntent
Maximum 100 characters. Valid characters: A-Z, a-z, 0-9, -, _
Description - optional
Helps users find their lost devices
Maximum 200 characters.
ID: JKOHGGY8XD


Intents (2) Info
An intent represents an action that the user wants to perform.
Delete
Add intent
Add empty intent
Use built-in intent
Search intents
Name
Description
Last edited
FallbackIntent
Default intent when no other intent matches
4 minutes ago
NewIntent
-
Now


Use built-in intent
Choose a built-in intent for your bot
Use built-in intents for intents commonly used in conversations. Built-in intents provide an extensive set of pre-defined sample utterances.
Built-in intent
AMAZON.Cancel
kendra
AMAZON.KendraSearchIntent
Searches documents indexed with Amazon Kendra.
Maximum 100 characters. Valid characters: A-Z, a-z, 0-9, -, _
Cancel
Add





Launch

To launch a stack using the CodeBuild Mode (faster and easier), click this button:

Launch Stack

Click the following button to launch a stack using the Pipeline Mode:

Launch Stack

CloudFormation Resources

The CloudFormation stack can create resources in your AWS account including:

- Amazon Lex bot. You can optionally pass the bot name of an existing one to avoid creating a new one.
- Cognito Identity Pool used to pass temporary AWS credentials to the web app. You can optionally pass the ID of an existing Cognito Identity Pool to avoid creating a new one.
- CodeBuild project to configure and deploy to S3 when using the CodeBuild Deployment Mode. If using the Pipeline Deployment Mode, a CodeBuild project is created to bootstrap a CodeCommit repository whit the application source.
- S3 buckets to host the web application and to store build artifacts.
- Lambda functions used as CloudFormation Custom Resources to facilitate custom provisioning logic
- CloudWatch Logs groups automatically created to log the output of the Lambda functions
- Associated IAM roles for the stack resources

If using the Pipeline Deployment Mode, the stack also creates the following resources:

- CodeCommit repository loaded with the source code in this project. This is only created when using the pipeline deployment mode
- A continuous delivery pipeline using CodePipeline and CodeBuild. The pipeline automatically builds and deploys changes to the app committed to the CodeCommit repo

BootstrapBucket
S3 bucket containing pre-staged nested templates and source artifacts

aws-bigdata-blog

BootstrapPrefix
S3 prefix where the templates and source are stored under

artifacts/aws-lex-web-ui/artifacts

Lex V1 Bot Configuration Parameters

BotName
Name of an existing Lex Bot to be used by the web ui. NOTE: You must also enter your published bot alias in the BotAlias field below.

BotAlias
WARNING: For production deployments, use your bot's published alias here. The $LATEST alias should only be used for manual testing. Amazon Lex limits the number of runtime requests that you can make to the $LATEST version of the bot.

$LATEST

Lex V2 Bot Configuration Parameters

LexV2BotId
Bot ID (not bot name) of an existing Lex V2 Bot to be used by the web ui. NOTE: You must also enter your Bot alias ID in the LexV2BotAliasId field below.

F66TWFSDTY

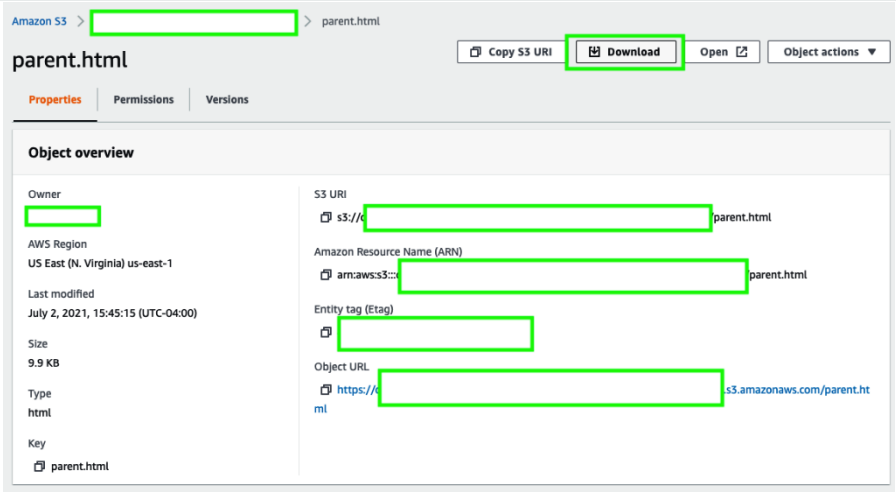
LexV2BotAliasId
Use your Lex V2 bot's alias id (not alias name) here

QJSXLKSAPV

LexV2BotLocaleId
Specify your bot's supported locale ids. By default this list contains only en_US. Other Lex V2 supported values are de_DE, en_AU, en_GB, es_419, es_ES, es_US, fr_CA, fr_FR, it_IT, ja_JP. A comma separated list of values can be supplied with the first value in the list being the default value. The remaining items can be selected in the Lex Web Ui menu. See "https://docs.aws.amazon.com/lexv2/latest/dg/lex2.0.pdf" for details on supported languages and locales.

en_US

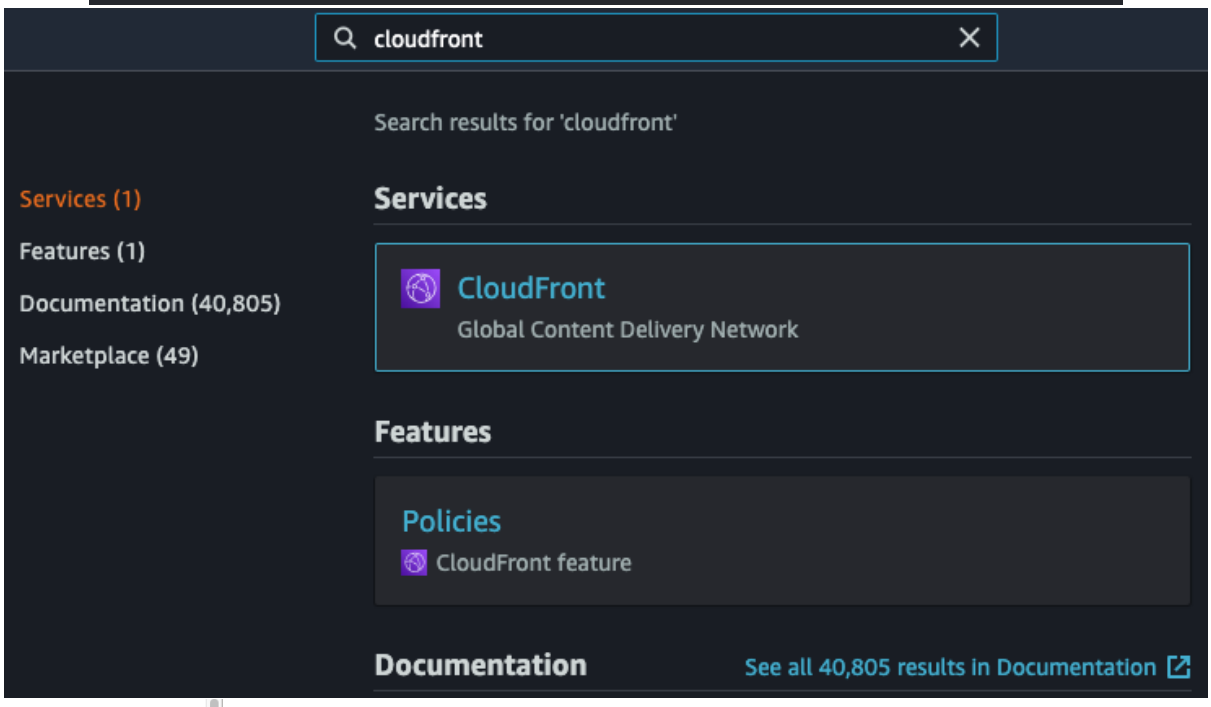
WebAppBucket chapter11-lex-web-ui-

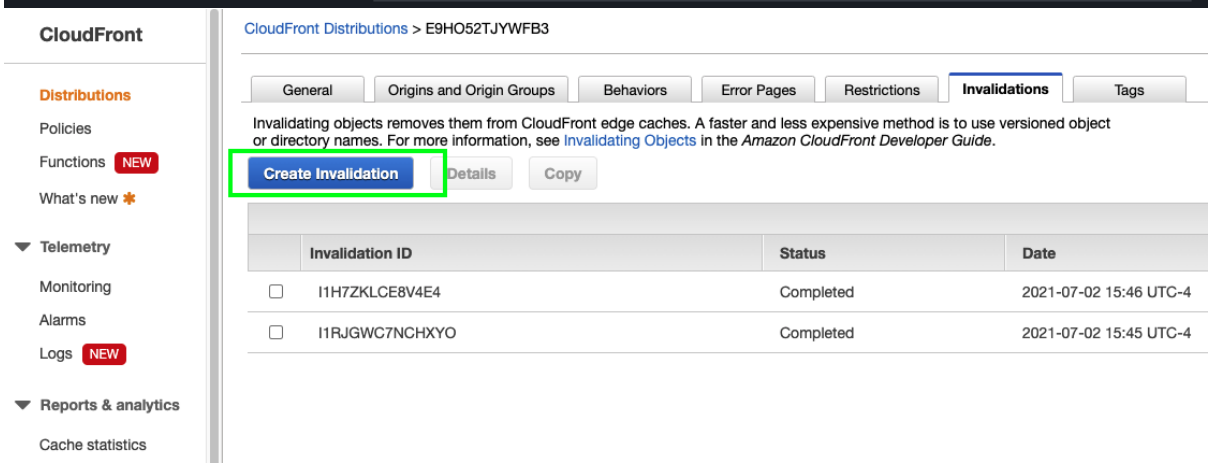


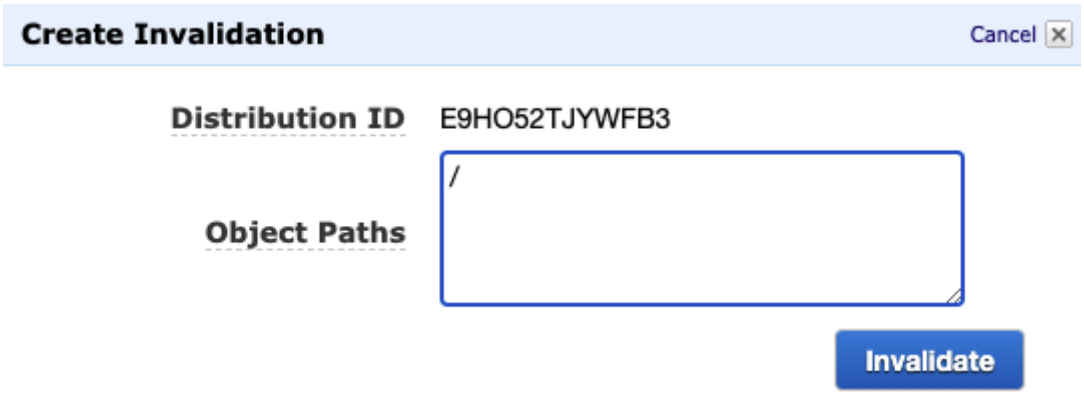

```
<link rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/bootstrap/3.3.7/css/bootstrap.min.css" integrity="s
"panel panel-primary"><div class="panel-heading">Send "what is important for buyers" utterance to iframe</div><d
></div></div><div class="row"><div class="col-md-6"><div class="panel panel-primary"><div class="panel-heading">
 panel-primary"><div class="panel-heading">Lex Session Attributes</div><div class="panel-body"><div><pre id="ses
el-body"><p>This application was created using the <a href="https://github.com/awslabs/aws-lex-web-ui">aws-lex-w
```

```
// Once the chatbot UI is ready, it sends a 'lexWebUiReady' event
$(document).on('lexWebUiReady', function onUpdateLexState(evt) {
  // We are just sending a ping request here as an example
  // This example uses an event instead of calling
  // iframeLoader.api.ping() to show the asynchronous
  // event API alternative
  var event = new CustomEvent(
    'lexWebUiMessage',
    { detail: {message: {event: 'ping'} } }
  );
  document.dispatchEvent(event);
});

// setup Send button handler
$('#send-intent').on('click', function(event) {
  event.preventDefault();
  sendUtterance('what is important for buyers?');
});
```





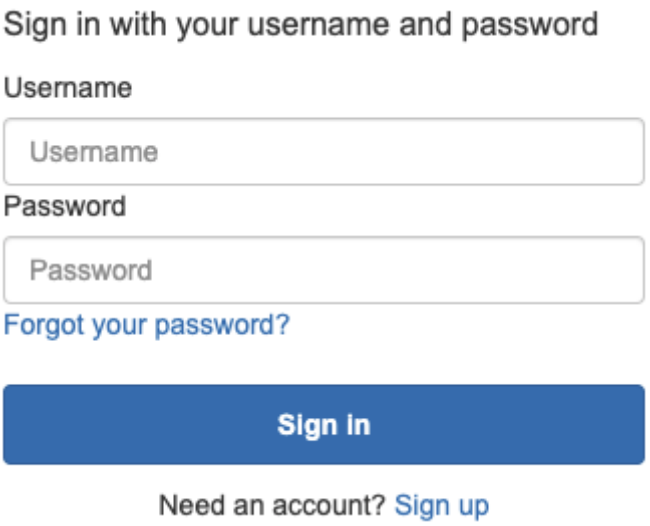


Stack info | Events | Resources | Outputs | Parameters | Template | Change sets

Outputs (10)

Search outputs

Key	Value	Description	Export name
CodeBuildUrl	https://console.aws.amazon.com/codebuild/home?region=us-east-1#/projects/chap11-lex-web-ui/view	Monitor the pipeline URL to see when the application has been fully built and deployed.	-
CognitoIdentityPoolId		Cognito Identity Pool Id	-
CognitoUserPoolClientId		Cognito User Pool Client Id	chap11-lex-web-ui:CognitoUserPoolClientId
CognitoUserPoolPubKey	https://cognito-idp.us-east-1.amazonaws.com/us-east-	Include Cognito User Pool Public Key URL in stack outputs (needed for QnABot token auth)	-
LoaderScriptUrl	https:// /lex-web-ui-loader.min.js	URL of the loader script This script will be available after the pipeline/deployment completes.	-
ParentPageUrl	https:// /parent.html	URL of the iframe based sample web application This page will be available after the pipeline/deployment completes.	-
SnippetUrl	https:// /iframe-snippet.html	URL of a page showing the snippet to load the chatbot UI as an iframe	-
WebAppBucket		S3 bucket hosting lexwebui artifacts	chap11-lex-web-ui:WebAppBucket
WebAppDomainName		DomainName of the web application	chap11-lex-web-ui:WebAppDomainName
WebAppUrl	https:// /index.html	URL of the stand-alone sample web application. This page will be available after the pipeline/deployment completes.	-



We have sent a code by email to r***@a***.com.
Enter it below to confirm your account.

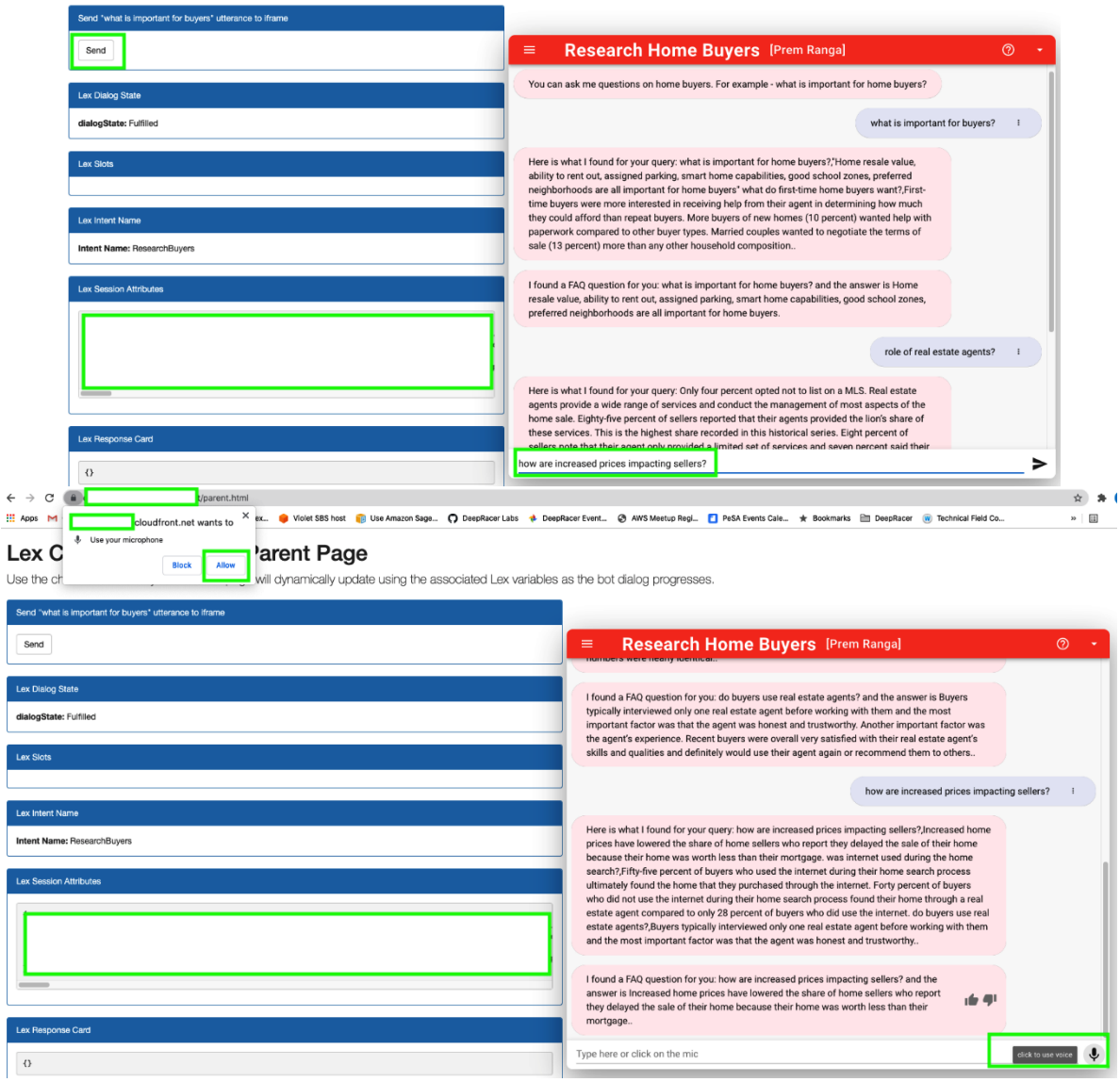
Verification Code

Confirm Account

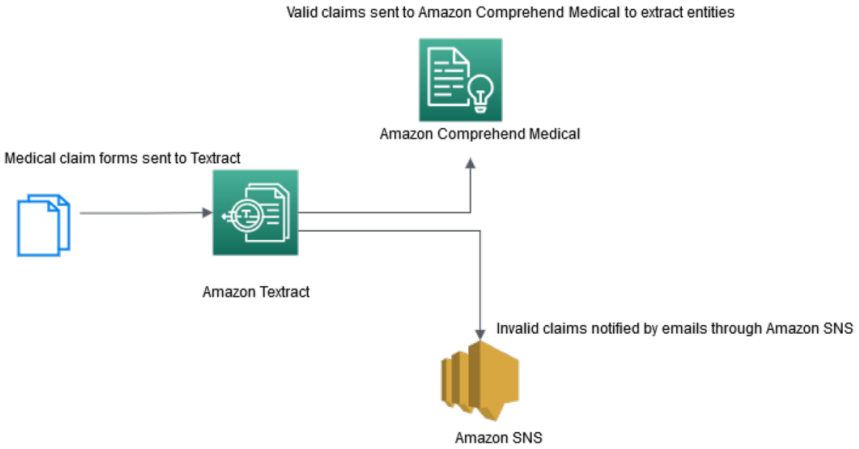
Didn't receive a code? Resend it

Lex Chatbot UI Sample Parent Page

Use the chatbot UI to drive your bot. This page will dynamically update using the associated Lex variables as the bot dialog progresses.



Chapter 12: AI and NLP in Healthcare



MEDICARE ☑ MEDICAID ☐

NAME Test1 Name1	BIRTH DATE 7-20-1975
ADDRESS 1 Test1 St	
CITY Testcity1 / STATE MA	Single ☐ Married ☑ Other ☐
ZIP CODE 01742 / TELEPHONE 712-674-3619	
ID NUMBER a-184054-222	
INSURANCE PLAN NAME myplan1	
PROCEDURE Blood glucose/reagent strips	
EMPLOYER MyEmployer	

READ BACK OF FORM BEFORE COMPLETING & SIGNING THIS FORM.

PATIENT'S OR AUTHORIZED PERSON'S SIGNATURE I authorize the release of any medical or other information necessary to process this claim. I also request payment of government benefits either to myself or to the party who accepts assignment below.

SIGNED Test1 Name1 DATE 07-25-2019

SYMPTOM DATE 06-03-2019	
DIAGONOSIS Diabetic	
RESERVED FOR LOCAL USE	

Image loaded validmedicalform.png

{'DocumentMetadata': {'Pages': 1}, 'Blocks': [{'BlockType': 'PAGE', 'Geometry': {'BoundingBox': {'Width': 1.0, 'Height': 1.0, 'Left': 0.0, 'Top': 0.0}, 'Polygon': [{'X': 1.2692531957743554e-16, 'Y': 0.0}, {'X': 1.0, 'Y': 1.1816080497439928e-16}, {'X': 1.0, 'Y': 1.0}, {'X': 0.0, 'Y': 1.0}]}, 'Id': '4af9d333-866f-4e3f-ab3b-74b4c6924970', 'Relationships': [{'Type': 'CHILD', 'Ids': ['ede5028d-ed05-40e0-94d0-b5515ecda492', '99bad1eb-479b-4e26-840c-d4eeb1a6c2a1', 'e1d79e29-d4b5-491c-ad7d-a1df5bd37c75', '4d1de960-a208-448c-912b-7a2be7cb0b49', '4d9eb38c-e0de-4e41-835f-9fa8ea1dd788', '00a69976-330e-490e-9f8d-b4fe17e12ba3', 'd7195ecc-923b-4ce5-9187-f36902256f37', 'd365cbd5-3c1c-4a3f-85e3-e936b7ca89d1', '71a46b5c-29af-4470-9480-186c5b907b25', '4a5a9876-458f-4d34-8b15-2ddf19ea067c', '65b32a81-4b40-4152-8604-09d58f446932', '8d25cf11-9813-4127-9cc7-7434b138ce7f', 'adc55859-866c-4416-bbbe-e71084f3a045', '121899e0-b058-46f4-94ac-96c19dc162a5', 'b5362726-5bad-440b-afca-9f5ed263e0d4', '18491ae7-b578-4c68-b5b7-a30393d8bd7a', '8cb6d837-d391-4814-8fdc-bd85eea1bbc5', '99ed1313-6a06-44d1-8f7a-9b3c7df02a77', 'df91db51-0410-4a0c-8636-7725ea9b02ea', '307a3750-19e8-474b-a1e0-d167eadd51e3', 'deda6d4d-4f07-4717-99aa-80bd963c241a', 'a7114686-9d77-4b73-9e1c-d1e37d15fd1e', '49742fd4-822a-4bf8-a9df-cf9f4932fe42', '49b7b800-ba01-4cf8-86e4-b0a816395a7f', '2cc368ce-afe9-4705-aa16-105e2980b473', '295591ac-2bb4-42d5-beca-5a060a7ccec2', 'ab39abad-c866-4bc2-b161-1e8aa5e3a117', 'ceb116b4-e8df-4dba-8af5-5ee293279475', '829691d9-0336-41e5-8433-29d6d699a168', 'f1f640c1-70da-46b5-90a2-35d3e53bafe4', '74273f3f-cb47-47a8-8d91-3f5c4afc0586', 'e4c5530d-9e57-4b45-a8a6-48d13f741bde', '5d7d491d-7c47-4d2f-82da-dfdc86fa949c', '2e1caa52-7f91-402a-b709-d52323e44a61', '14c5937f-d62f-461f-9271-afb9ca9917e5', 'f1b2117b-8d4e-4609-9b8c-ed33cfc08980', '287fa619-0a2f-4b11-85e5-6c5844491574', '1337f531-c99e-4a80-ace1-93d1f096530d', '6f989a73-3bc5-4169-b54c-09ae935d618e', '78c02bb6-93a8-42b3-b865-8b1167514993', '63f2a3b9-2694-4e54-946c-afc7a7337039', '74bc3196-59e7-4903-bf43-5dfc745c7a53', '31f6a7
{'Married': 'SELECTED', 'Other': 'NOT_SELECTED', 'MEDICAID': 'NOT_SELECTED', 'Single': 'NOT_SELECTED', 'TELEPHONE': '111-111-1111', 'ZIP CODE': '111111', 'STATE': 'MA', 'INSURANCE PLAN NAME': 'myplan1', 'ID NUMBER': 'a-184054-6661', 'ADDRESS': '1 Test St', 'NAME': 'Failed Test', 'DATE': '08-22-2019', 'SIGNED': 'Failed Test', 'MEDICARE': 'SELECTED', 'DIAGONOSIS': 'Allergy', 'BIRTH DATE': '7-22-1979', 'CITY': 'Testcity', 'SYMPTOMDATE': '06-03-2019', 'PROCEDURE': 'Oxygen concentrator', 'EMPLOYER': 'MyEmployer'}

{'MEDICAID': 'NOT_SELECTED', 'Single': 'NOT_SELECTED', 'TELEPHONE': '712-674-3619', 'Other': 'NOT_SELECTED', 'Married': 'SELECTED', 'ADDRESS': '1 Test1 St', 'ZIP CODE': '01742', 'ID NUMBER': 'a-184054-222', 'STATE': 'MA', 'INSURANCE PLAN NAME': 'myplan1', 'DATE': '07-25-2019', 'MEDICARE': 'SELECTED', 'NAME': 'Test1 Name1', 'CITY': 'Testcity1', 'PROCEDURE': 'Blood glucose/reagent strips', 'SIGNED': 'Test1 Name1', 'BIRTH DATE': '7-20-1975', 'DIAGONOSIS': 'Diabetic', 'SYMPTOM.DATE': '06-03-2019', 'EMPLOYER': 'MyEmployer'}

Ok

a-184054-222

Medical Entities

========

- 111-111-1111
  Type: PHONE_OR_FAX
  Category: PROTECTED_HEALTH_INFORMATION

- 111111
  Type: ID
  Category: PROTECTED_HEALTH_INFORMATION

- MA
  Type: NAME
  Category: PROTECTED_HEALTH_INFORMATION

- myplan1
  Type: NAME
  Category: PROTECTED_HEALTH_INFORMATION

- a-184054-6661
  Type: ID
  Category: PROTECTED_HEALTH_INFORMATION

- 08-22-2019
  Type: DATE
  Category: PROTECTED_HEALTH_INFORMATION

- Allergy
  Type: DX_NAME
  Category: MEDICAL_CONDITION

MEDICARE ☑ MEDICAID ☐

NAME Failed Test	BIRTH DATE 7-22-1979
ADDRESS 1 Test St	
CITY Testcity / STATE MA	Single ☐ Married ☑ Other ☐
ZIP CODE 111111 / TELEPHONE 111-111-1111	
ID NUMBER a-184054-6661	
INSURANCE PLAN NAME myplan1	
PROCEDURE Oxygen concentrator	
EMPLOYER MyEmployer	

READ BACK OF FORM BEFORE COMPLETING & SIGNING THIS FORM.

PATIENT'S OR AUTHORIZED PERSON'S SIGNATURE I authorize the release of any medical or other information necessary to process this claim. I also request payment of government benefits either to myself or to the party who accepts assignment below.

SIGNED Failed Test DATE 08-22-2019

SYMPTOM DATE 06-03-2019	
DIAGONOSIS Allergy	
RESERVED FOR LOCAL USE	

{'Married': 'SELECTED', 'Other': 'NOT_SELECTED', 'MEDICAID': 'NOT_SELECTED', 'Single': 'NOT_SELECTED', 'TELEPHONE': '111-111-1111', 'ZIP CODE': '111111', 'STATE': 'MA', 'INSURANCE PLAN NAME': 'myplan1', 'ID NUMBER': 'a-184054-6661', 'ADDRESS': '1 Test St', 'NAME': 'Failed Test', 'DATE': '08-22-2019', 'SIGNED': 'Failed Test', 'MEDICARE': 'SELECTED', 'DIAGONOSIS': 'Allergy', 'BIRTH DATE': '7-22-1979', 'CITY': 'Testcity', 'SYMPTOMDATE': '06-03-2019', 'PROCEDURE': 'Oxygen concentrator', 'EMPLOYER': 'MyEmployer'}

{'Married': 'SELECTED', 'Other': 'NOT_SELECTED', 'MEDICAID': 'NOT_SELECTED', 'Single': 'NOT_SELECTED', 'TELEPHONE': '111-111-1111', 'ZIP CODE': '111111', 'STATE': 'MA', 'INSURANCE PLAN NAME': 'myplan1', 'ID NUMBER': 'a-184054-6661', 'ADDRESS': '1 Test St', 'NAME': 'Failed Test', 'DATE': '08-22-2019', 'SIGNED': 'Failed Test', 'MEDICARE': 'SELECTED', 'DIAGONOSIS': 'Allergy', 'BIRTH DATE': '7-22-1979', 'CITY': 'Testcity', 'SYMPTOMDATE': '06-03-2019', 'PROCEDURE': 'Oxygen concentrator', 'EMPLOYER': 'MyEmployer'}

Invalid claim Id
a-184054-6661
False

Create topic

Details

Type Info

Topic type cannot be modified after topic is created

- ○ **FIFO (first-in, first-out)**
  - Strictly-preserved message ordering
  - Exactly-once message delivery
  - High throughput, up to 300 publishes/second
  - Subscription protocols: SQS
- ◉ **Standard**
  - Best-effort message ordering
  - At-least once message delivery
  - Highest throughput in publishes/second
  - Subscription protocols: SQS, Lambda, HTTP, SMS, email, mobile application endpoints

Name

invalid-claims-notify

Maximum 256 characters. Can include alphanumeric characters, hyphens (-) and underscores (_).

Display name - *optional*

To use this topic with SMS subscriptions, enter a display name. Only the first 10 characters are displayed in an SMS message. Info

My Topic

Maximum 100 characters, including hyphens (-) and underscores (_).

Amazon SNS > Subscriptions > Create subscription

Create subscription

Details

Topic ARN

arn:aws:sns:us-east-[redacted]

Protocol

The type of endpoint to subscribe

Email

Endpoint

An email address that can receive notifications from Amazon SNS.

test@example.com

After your subscription is created, you must confirm it. Info

Amazon SNS > Topics

Topics (3)

Edit | Delete | Publish message | Create topic

Search

1

Name	Type	ARN
invalid-claims-notify	Standard	arn:aws:sns:us-east-1:[redacted]nvalid-claims-notify


```
{'MessageId': '767633de-0fe1-5614-8c34-6aa0cdcdcbd8', 'ResponseMetadata': {'RequestId': 'ec96366c-d9c1-53f8-9ca7-6f121bed3e34', 'HTTPStatusCode': 200, 'HTTPHeaders': {'x-amzn-requestid': 'ec96366c-d9c1-53f8-9ca7-6f121bed3e34', 'content-type': 'text/xml', 'content-length': '294', 'date': 'Sat, 31 Jul 2021 03:13:47 GMT'}, 'RetryAttempts': 0}}
```

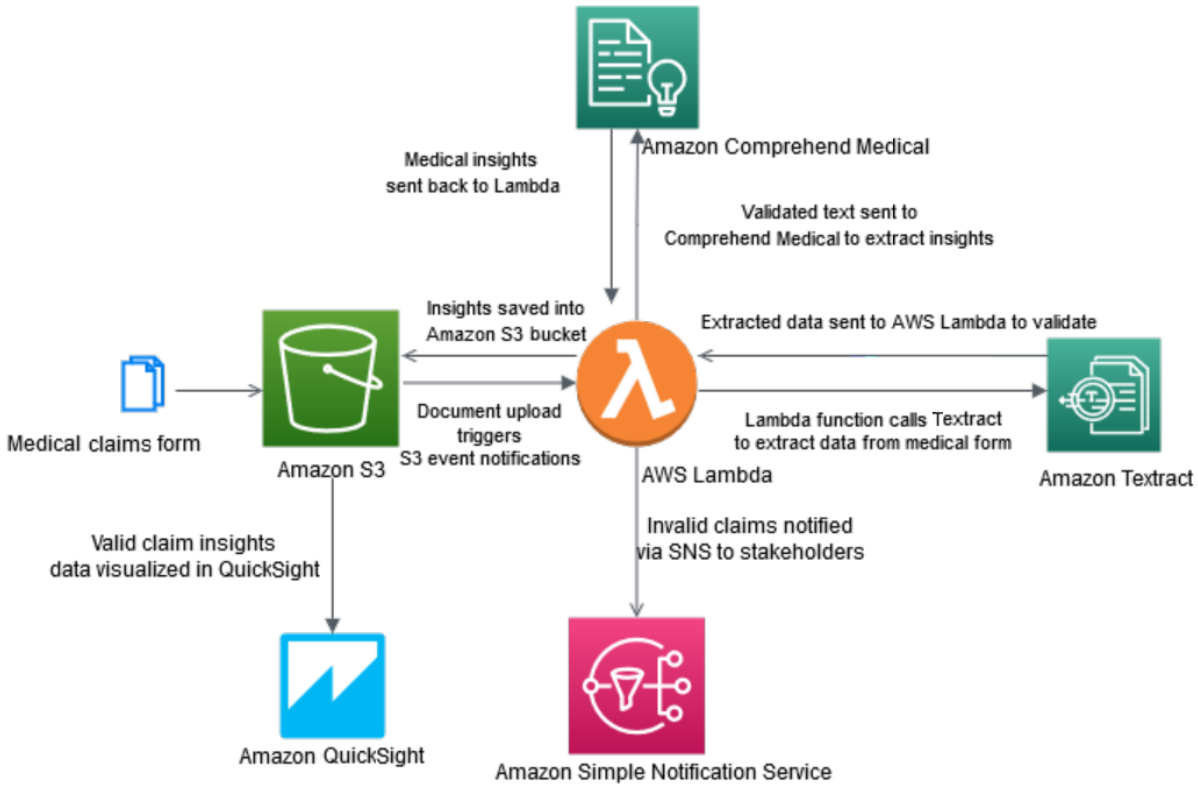
Reply Reply All Forward

AN

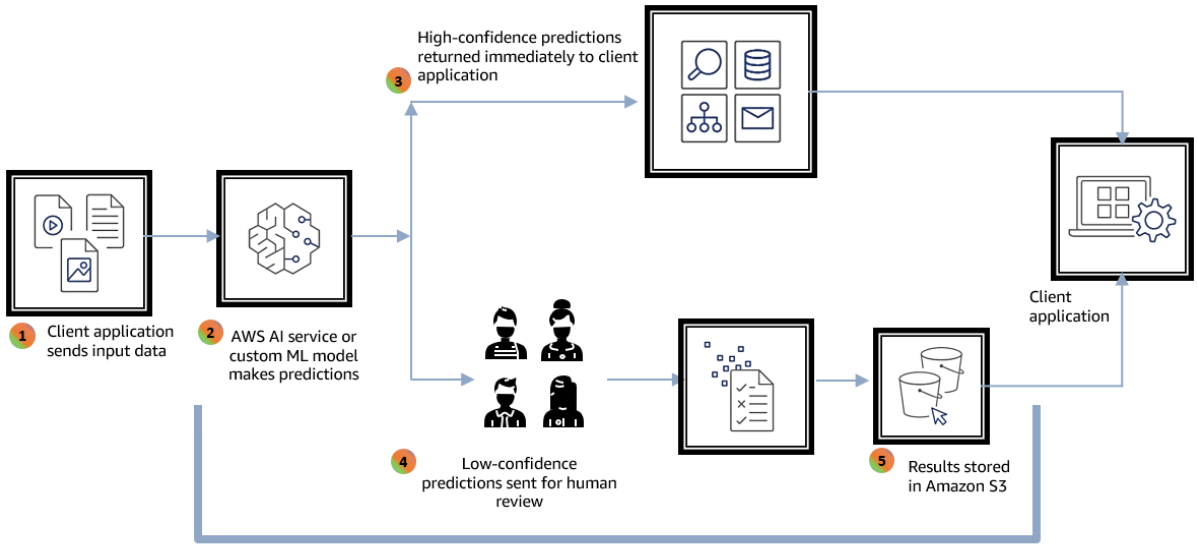
AWS Notifications <no-reply@sns.amazonaws.com>

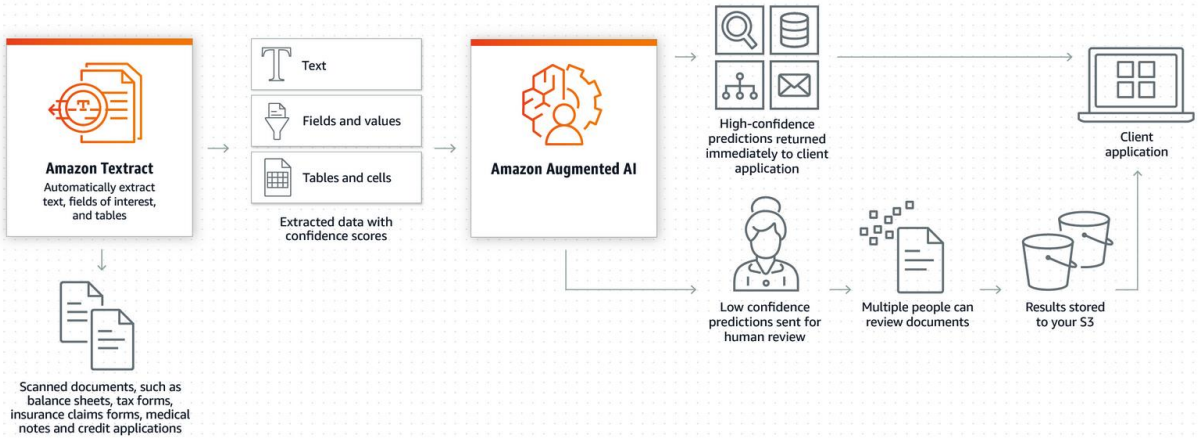
AWS Notification Message

"ADDRESS": "1 Test St",
"NAME": "Failed Test",
"DATE": "08-22-2019",
"SIGNED": "Failed Test",
"MEDICARE": "SELECTED",
"DIAGONOSIS": "Allergy",
"BIRTH DATE": "7-22-1979",
"CITY": "Testcity",
"SYMPTOMDATE": "06-03-2019",
"PROCEDURE": "Oxygen concentrator",
"EMPLOYER": "MyEmployer"
}Reason:Invalid claim Id



Chapter 13: Improving the Accuracy of Document Processing Workflows





Cross-origin resource sharing (CORS)

The CORS configuration, written in JSON, defines a way for client web applications that are loaded in one domain to interact with resources in a different domain. Learn more

Edit

Copy

```
[
  {
    "AllowedHeaders": [],
    "AllowedMethods": [
      "GET"
    ],
    "AllowedOrigins": [
      "*"
    ],
    "ExposeHeaders": []
  }
]
```

Add workers Info

Add workers to your private work team by adding worker email addresses or importing workers from existing Amazon Cognito user groups.

- (•) Invite new workers by email
- ( ) Import workers from existing Amazon Cognito user groups

Email addresses

Preview invitation

We send an invitation with instructions to each of the worker email addresses that you add here.

abc@gmail.com

Use a comma between addresses. You can add up to 50 workers.

Organization name

We use this information to customize the email that we send to the workers.

AWS

Contact email

Workers use this address to report issues related to the task.

Enter the contact email

Workflow settings

Name

textract-check

The name must be lowercase, unique within the Region in your account, and can have up to 63 characters. Valid characters: a-z, 0-9, and - (hyphen)

S3 bucket

Enter the path to the Amazon S3 bucket where you want to store the output of the human review. Open Amazon S3 console

s3://a2i-demos[redacted]

The path must have the following format: s3://bucket name/folder name.

IAM role

This IAM role is used to grant Augmented AI permission to call other services on your behalf. If you want to use this role to start and manage human loops using Augmented AI Runtime, Amazon Rekognition or Amazon Textract API operations, you can attach the AmazonAugmentedAIIntegratedAPIAccess policy to the role in the IAM console Learn more

A2ISageMake[redacted]

IAM role created

The role you just created only allows Augmented AI to call other services on your behalf. If you want to use this role to start and manage human loops using Augmented AI Runtime, Amazon Rekognition or Amazon Textract API operations, you can attach the AmazonAugmentedAIIntegratedAPIAccess policy to the role in the IAM console Learn more

Amazon Textract form extraction - Conditions for invoking human review

When Amazon Textract extracts information from a document, it returns a confidence score. You can use these confidence scores to define business conditions that trigger human review.

Identification confidence
The confidence score for key-value pairs detected within a form.

Qualification confidence
The confidence score for text contained within key and value in a form.

You can define a range for Identification confidence and Qualification confidence thresholds. A human review will be triggered when the confidence score falls within the defined range.

Learn more about using Amazon Augmented AI with Amazon Textract

☑ Trigger a human review for specific form keys based on the form key confidence score or when specific form keys are missing.
The form key and value will be sent for human review.

Key name

$

Trigger human review when this form key is missing,

or when its identification confidence threshold is between 0 and 90

or when its qualification confidence threshold is between 0 and 90

Add key

☐ Trigger human review for all form keys identified by Amazon Textract with confidence scores in a specified range.
The form key and value will be sent for human review.

☐ Randomly send a sample of forms to humans for review.
For each form sent, all key-value pairs identified by Amazon Textract for that form will be sent for human review.

Workers

Worker types

○ **Amazon Mechanical Turk**
An on-demand 24x7 workforce of over 500,000 independent contractors worldwide, powered by Amazon Mechanical Turk.

◉ **Private**
A team of workers that you have sourced yourself, including your own employees or subject matter experts.

○ **Vendor**
A curated list of third party vendors that specialize in providing human review services, available via the AWS market place.

Private teams

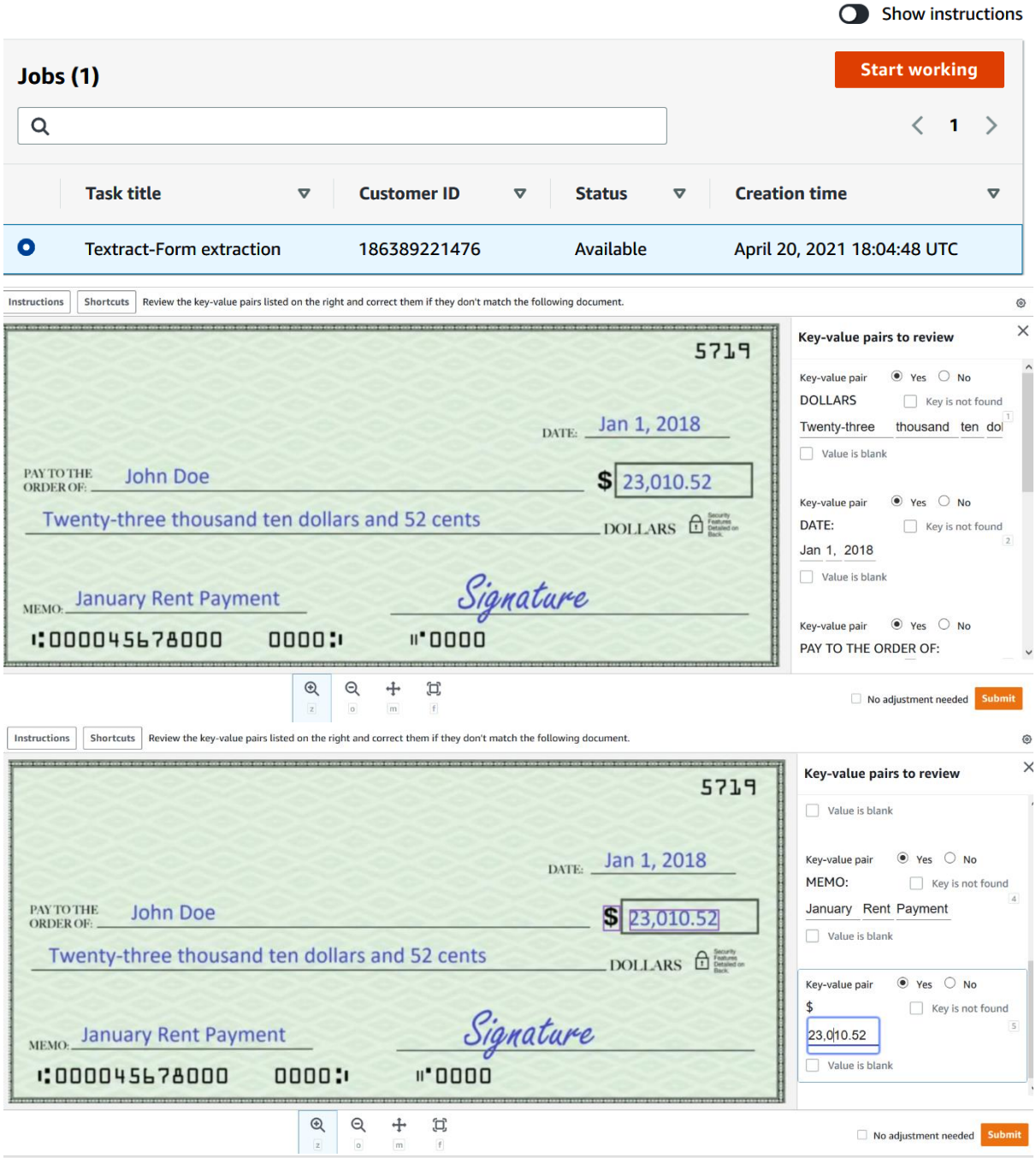
Choose from the teams you created in the private workforce or **create a new team** on the Human review workforces page.

Choose team

⚠ Choose a team from your private Workforces.

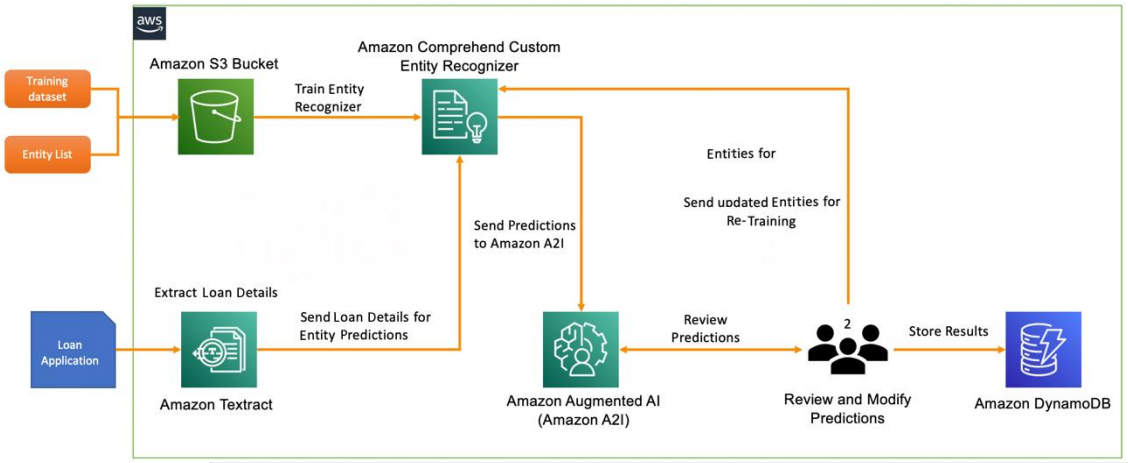
```
print(response)
```

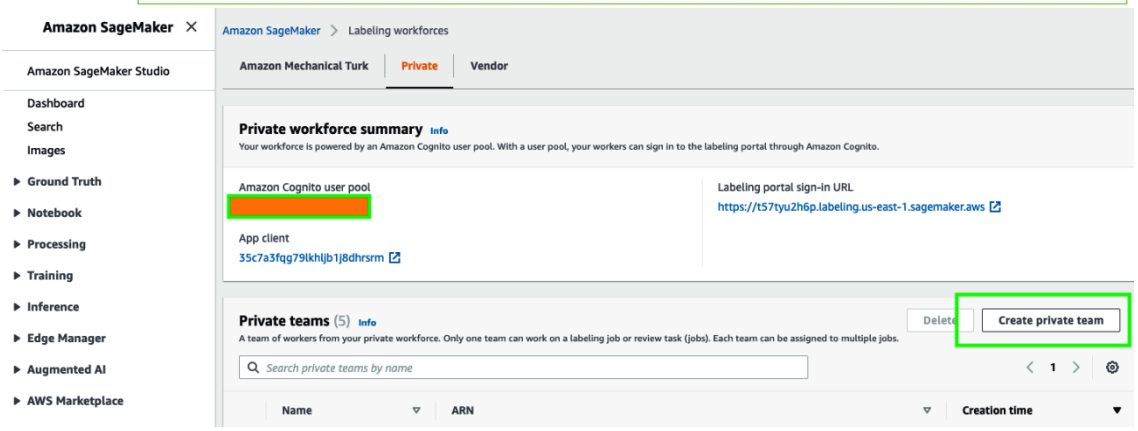

```
tionships': [{'Type': 'VALUE', 'Ids': ['9f84a49c-210e-4371-8832-3959dcf0ef50']}, {'Type': 'CHILD', 'Ids': ['6d365fd
2-e381-48a5-8eeb-60b82b61272c']}], 'EntityTypes': ['KEY']}, {'BlockType': 'KEY_VALUE_SET', 'Confidence': 77.0, 'Geo
metry': {'BoundingBox': {'Width': 0.5739195346832275, 'Height': 0.06345107406377792, 'Left': 0.044734030961990356,
'Top': 0.5416923761367798}, 'Polygon': [{'X': 0.044734030961990356, 'Y': 0.5416923761367798}, {'X': 0.6186535954475
403, 'Y': 0.5416923761367798}, {'X': 0.6186535954475403, 'Y': 0.6051434278488159}, {'X': 0.044734030961990356, 'Y':
0.6051434278488159}]}, 'Id': '9f84a49c-210e-4371-8832-3959dcf0ef50', 'Relationships': [{'Type': 'CHILD', 'Ids': ['8
2b39d45-6779-4099-ac4f-2a6d55e8bc2d', 'cf4510f9-6cfd-413b-8898-ee00ff209f1e', '144192d5-a534-4aad-aa56-ab14391b231d
', 'bf27afd9-7163-4a8e-9505-6a9b70e0c2ec', '378cbde0-47e4-4e2b-9b18-9285ef364c93', '703cee3c-47dd-43cf-b0c2-8d90302
6fca4', '3e7ef823-ecfc-48f6-98fa-a0a338f00837']}], 'EntityTypes': ['VALUE']}], 'HumanLoopActivationOutput': {'Human
LoopArn': 'arn:aws:sagemaker:us-east-1:186389221476:human-loop/textract-10', 'HumanLoopActivationReasons': ['Condit
ionsEvaluation'], 'HumanLoopActivationConditionsEvaluationResults': '{"Conditions":[{"And":[{"ConditionType":"Impor
tantFormKeyConfidenceCheck","ConditionParameters":{"ImportantFormKey":"*","ImportantFormKeyAliases":[],"KeyValueBlo
ckConfidenceLessThan":99.0,"WordBlockConfidenceLessThan":99.0},"EvaluationResult":true},{"ConditionType":"Important
FormKeyConfidenceCheck","ConditionParameters":{"ImportantFormKey":"*","ImportantFormKeyAliases":[],"KeyValueBlockCo
nfidenceGreaterThan":0.0,"WordBlockConfidenceGreaterThan":0.0},"EvaluationResult":true}],"EvaluationResult":tru
e}]}'}, 'AnalyzeDocumentModelVersion': '1.0', 'ResponseMetadata': {'RequestId': '67ae6246-bcb1-4ef2-89d4-2c4cb2890d
88', 'HTTPStatusCode': 200, 'HTTPHeaders': {'x-amzn-requestid': '67ae6246-bcb1-4ef2-89d4-2c4cb2890d88', 'content-ty
pe': 'application/x-amz-json-1.1', 'content-length': '38256', 'date': 'Sun, 18 Jul 2021 18:00:41 GMT'}, 'RetryAttem
pts': 0}}
```

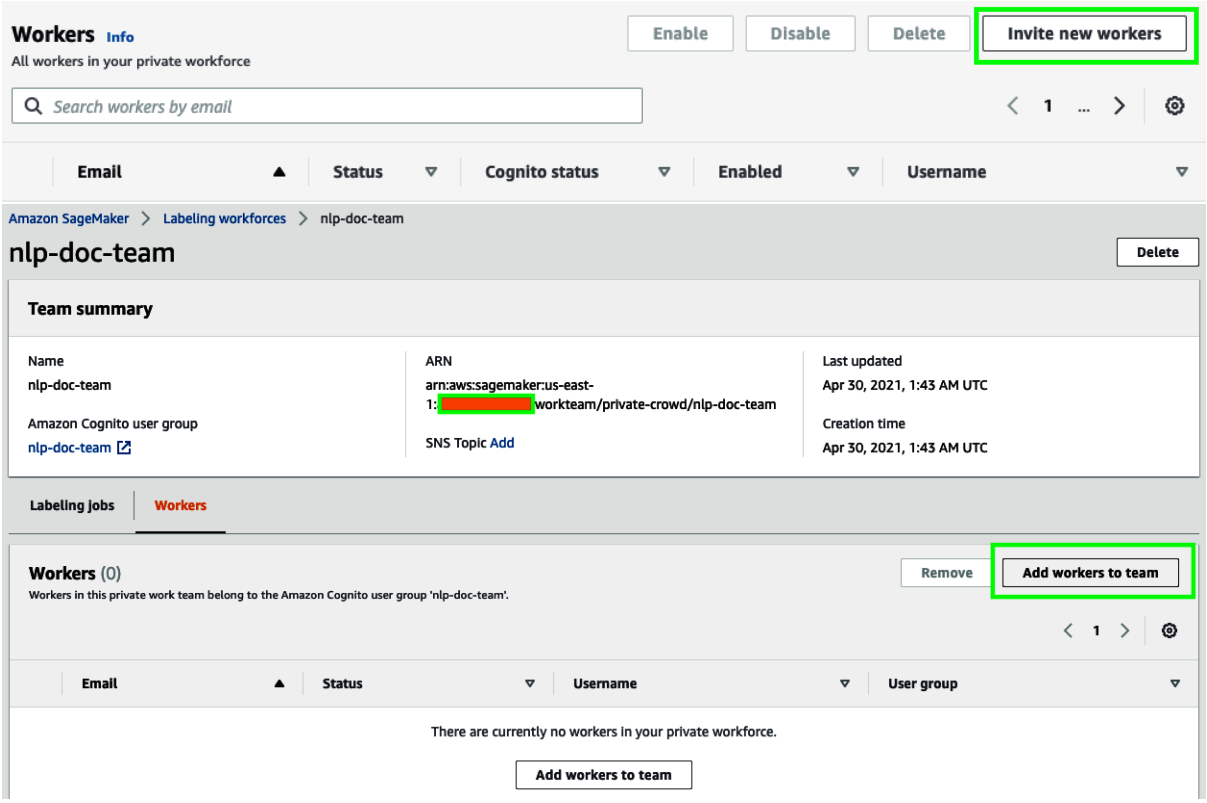


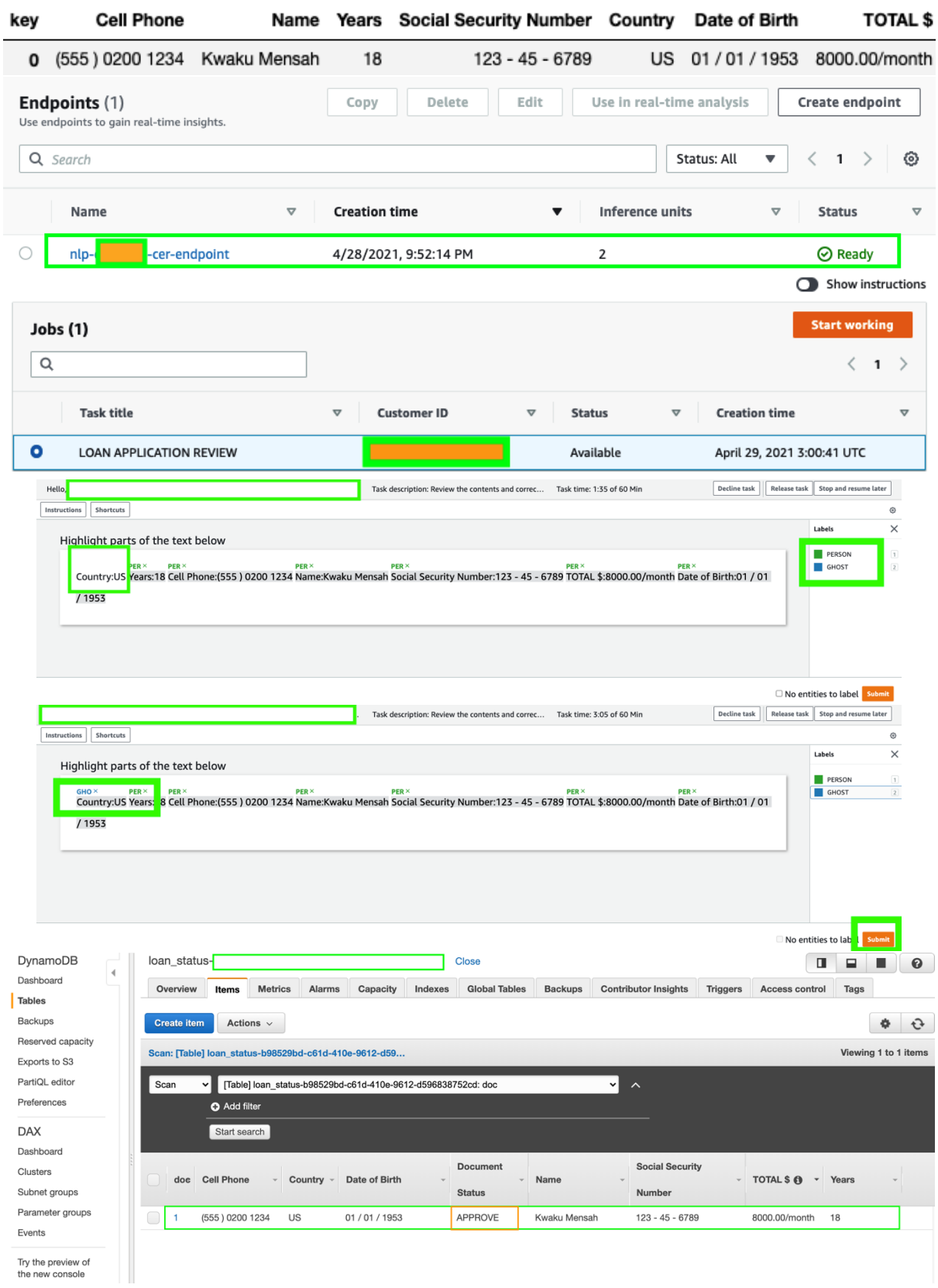

```
textarct-check/2021/07/18/23/29/16/7ecedc0b-d80c-4918-acfc-3b53e3b09326/output.json
{   'awsManagedHumanLoopRequestSource': 'AWS/Textract/AnalyzeDocument/Forms/V1',
    'flowDefinitionArn': 'arn:aws:sagemaker:us-east-1:[redacted]:flow-definition/textarct-check',
    'humanAnswers': [   {   'acceptanceTime': '2021-07-18T23:30:47.742Z',
                            'answerContent': {   'AWS/Textract/AnalyzeDocument/Forms/V1': {   'blocks': [   {   'bl
ockType': 'WORD',
                                                                                                'id
': '242eb680-49a8-4262-bc06-4493934af2e2',
                                                                                                'te
xt': 'THE'},
                                                                                            {   'bl
ockType': 'KEY_VALUE_SET',
                                                                                                'co
nfidence': 77,
                                                                                                'en
tityTypes': [   'KEY'],
                                                                                                'id
': '74bc7b7c-88fe-4430-9997-94018fb6974a',
                                                                                                're
```

Chapter 14: Auditing Named Entity Recognition Workflows

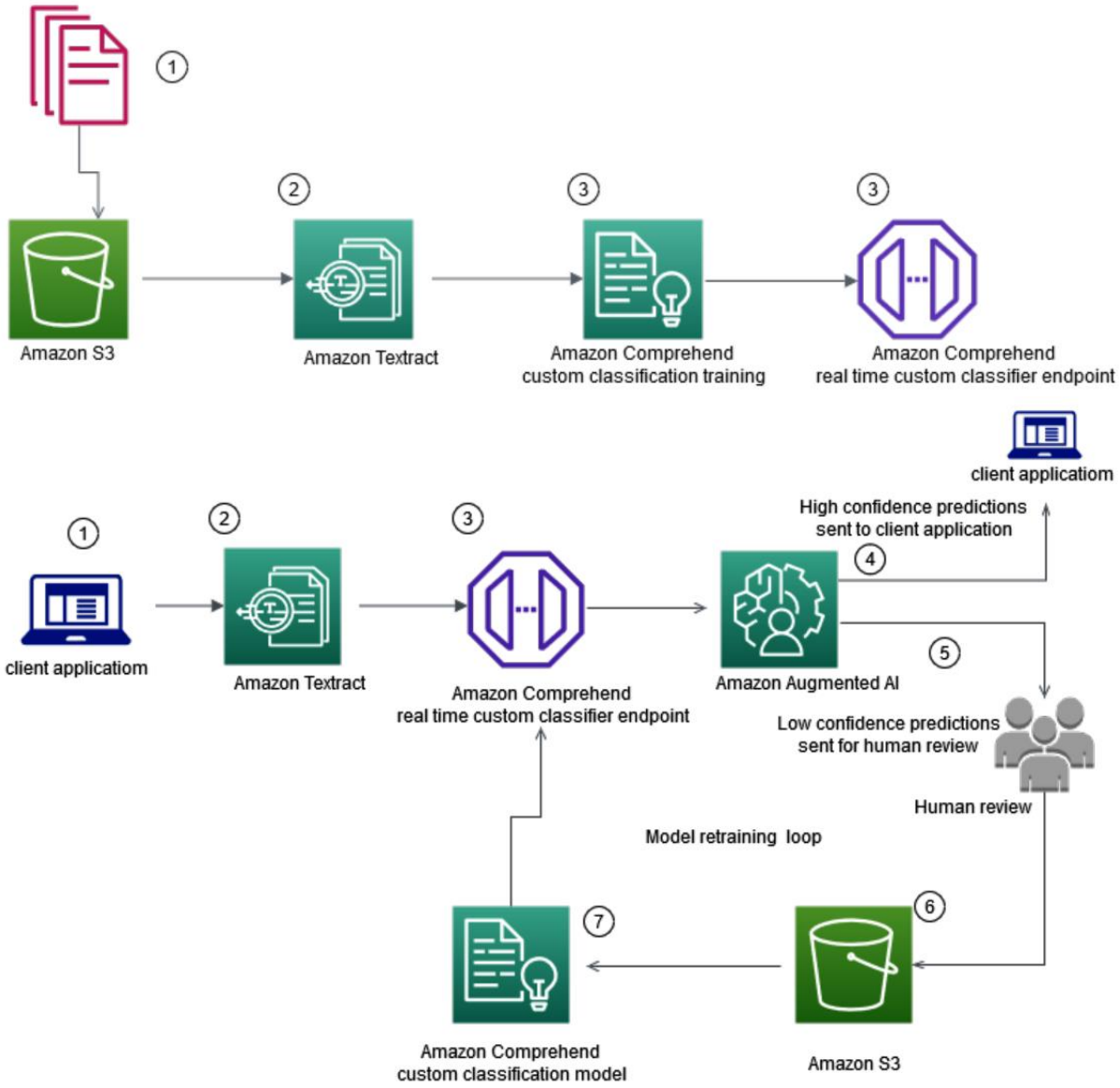







key
Cell Phone
Name
Years
Social Security Number
Country
Date of Birth
TOTAL $
0
(555) 0200 1234
Kwaku Mensah
18
123 - 45 - 6789
US
01 / 01 / 1953
8000.00/month
Endpoints (1)
Use endpoints to gain real-time insights.
Copy
Delete
Edit
Use in real-time analysis
Create endpoint
Search
Status: All
Name
Creation time
Inference units
Status
-cer-endpoint
4/28/2021, 9:52:14 PM
2
Ready
Show instructions
Jobs (1)
Start working
Task title
Customer ID
Status
Creation time
LOAN APPLICATION REVIEW
Available
April 29, 2021 3:00:41 UTC
Task description: Review the contents and correc...
Task time: 1:35 of 60 Min
Decline task
Release task
Stop and resume later
Instructions
Shortcuts
Highlight parts of the text below
Country:US Years:18 Cell Phone:(555) 0200 1234 Name:Kwaku Mensah Social Security Number:123 - 45 - 6789 TOTAL $:8000.00/month Date of Birth:01 / 01 / 1953
Labels
PERSON
GHOST
No entities to label
Submit
Task time: 3:05 of 60 Min
Country:US Years: 8 Cell Phone:(555) 0200 1234 Name:Kwaku Mensah Social Security Number:123 - 45 - 6789 TOTAL $:8000.00/month Date of Birth:01 / 01 / 1953
DynamoDB
Dashboard
Tables
Backups
Reserved capacity
Exports to S3
PartiQL editor
Preferences
DAX
Dashboard
Clusters
Subnet groups
Parameter groups
Events
Try the preview of the new console
loan_status-
Close
Overview
Items
Metrics
Alarms
Capacity
Indexes
Global Tables
Backups
Contributor Insights
Triggers
Access control
Tags
Create item
Actions
Scan: [Table] loan_status-b98529bd-c61d-410e-9612-d59...
Viewing 1 to 1 items
Scan
[Table] loan_status-b98529bd-c61d-410e-9612-d596838752cd: doc
Add filter
Start search
doc
Cell Phone
Country
Date of Birth
Document Status
Name
Social Security Number
TOTAL $
Years
1
(555) 0200 1234
US
01 / 01 / 1953
APPROVE
Kwaku Mensah
123 - 45 - 6789
8000.00/month
18

Chapter 15: Classifying Documents and Setting up Human in the Loop for Active Learning

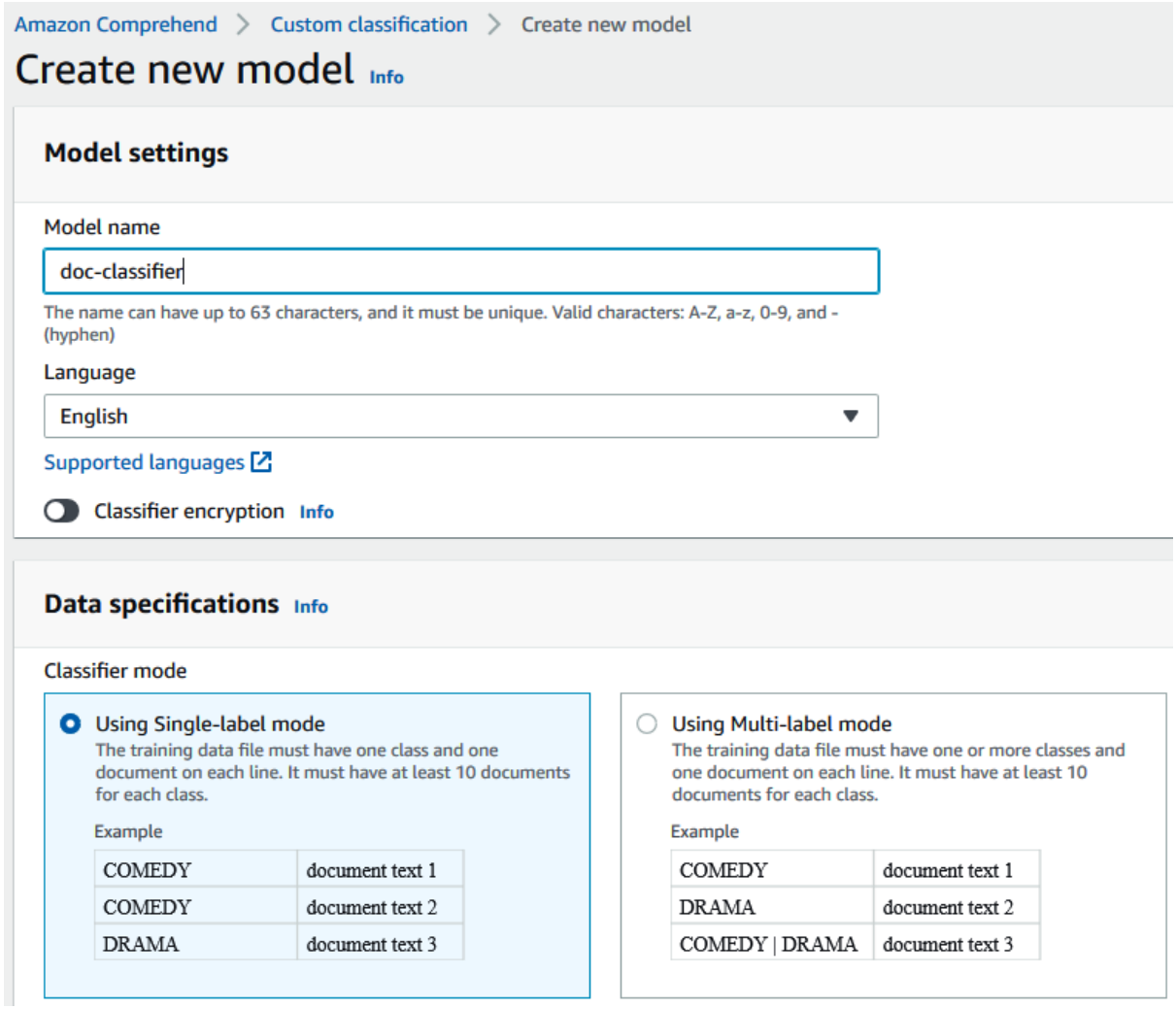


```
/home/ec2-user/SageMaker/SAMPLE8/WORDS/
train/Pay Stubs
/home/ec2-user/SageMaker/SAMPLE8/WORDS/
train/Pay Stubs
/home/ec2-user/SageMaker/SAMPLE8/WORDS/
train/Pay Stubs
/home/ec2-user/SageMaker/SAMPLE8/WORDS/
train/Pay Stubs
/home/ec2-user/SageMaker/SAMPLE8/WORDS/
train/Pay Stubs
/home/ec2-user/SageMaker/SAMPLE8/WORDS/
train/Pay Stubs
/home/ec2-user/SageMaker/SAMPLE8/WORDS/
train/Pay Stubs
/home/ec2-user/SageMaker/SAMPLE8/WORDS/
train/Pay Stubs
/home/ec2-user/SageMaker/SAMPLE8/WORDS/
train/Pay Stubs
/home/ec2-user/SageMaker/SAMPLE8/WORDS/
train/Pay Stubs
--- 74.77762579917908 seconds for extracting ---
```

Out[39]:

	label	document
0	Pay Stubs	CO FILE DEPT. CLOCK VCHR NO. 046 LV3 000342 00...
1	Pay Stubs	Acme LLC Earnings Statement 555 Woolworth Driv...
2	Pay Stubs	Acme LLC Earnings Statement 555 Woolworth Driv...
3	Pay Stubs	Acme LLC Earnings Statement 555 Woolworth Driv...
4	Pay Stubs	Acme LLC Earnings Statement 555 Woolworth Driv...
...	...	...
72	Bank Statements	This is a sample checking account bank stateme...
73	Bank Statements	Bank name COMPANY NAME AND Address 1 Address 2...
74	Bank Statements	CHASE July 1, 2008 through July 31, 2008 Prima...
75	Bank Statements	CHASE July 1, 2008 through July 31, 2008 Prima...
76	Bank Statements	CHASE July 1, 2008 through July 31, 2008 Prima...

77 rows × 2 columns


Amazon Comprehend > Custom classification > Create new model
Create new model Info
Model settings
Model name
doc-classifier
The name can have up to 63 characters, and it must be unique. Valid characters: A-Z, a-z, 0-9, and - (hyphen)
Language
English
Supported languages
Classifier encryption Info
Data specifications Info
Classifier mode
Using Single-label mode
The training data file must have one class and one document on each line. It must have at least 10 documents for each class.
Example
COMEDY document text 1
COMEDY document text 2
DRAMA document text 3
Using Multi-label mode
The training data file must have one or more classes and one document on each line. It must have at least 10 documents for each class.
Example
COMEDY document text 1
DRAMA document text 2
COMEDY | DRAMA document text 3

Training data location on S3

Paste the URL of an input data file in S3, or select a bucket or folder location in S3.

s3://doc-processing-bucket[redacted]comprehend_train_data.csv

Browse S3

Output data - *optional* Info

S3 location

Paste the URL of a bucket or folder location in S3, or select a bucket or folder location in S3.

s3://doc-processing-bucket[redacted]

Browse S3

Encryption Info

Access permissions Info

IAM role

- ○ Use an existing IAM role
- ◉ Create an IAM role

Permissions to access

Your role will have access to these resources.

Input and output (if specified) S3 bucket

Name suffix

Your roles will be prefixed with "AmazonComprehendServiceRole-". By clicking "Train classifier" you are authorizing creation of this role.

classifydoc

Amazon Comprehend > Create endpoint

Create endpoint Info

Endpoint settings

Endpoint name

classify-doc

The name can have up to 40 characters, and it must be unique. Valid characters: A-Z, a-z, 0-9, and - (hyphen)

Custom model

Choose an existing model you've previously trained or go to Custom classification or Custom entity recognition to create a model.

doc-classifier
Custom classifier, 1 endpoint

▶ Tags - *optional* Info

A tag is a label that you can add to a resource as metadata to help you organize, search, or filter your data. Each tag consists of a key and an optional value.

Inference units

IU estimator

Number of inference units (IUs)

A single IU can analyze one hundred characters per second. IU is a measure of the endpoint's throughput. You can adjust the IU of an endpoint anytime.

1 IU ≤ 100 characters/second

Range from 1 to 10

Amazon Comprehend > Custom classification > doc-classifier > classify-doc

classify-doc

Duplicate | Delete | Edit | Use in real-time analysis

Endpoint details

Name classify-doc	Creation time [redacted]2:20:46 PM	Total active time [redacted]	Classifier name doc-classifier
Status Active	Last modified time [redacted]:20:46 PM	Estimated hourly cost [redacted]	ARN arn:aws:comprehend:us-east-[redacted]document-classifier-endpoint/classify-doc
Inference units Pricing information 1 IU			

ABC Company
123 Main Street,
Anytown, NY 10000.
EIN: 51-2322286

Earnings Statement

Pay Period: May 15, 2020 to May 16, 2020
Pay Date: May 16, 2020

SSN: XXX-XXX-1425
Stub No.: 1112

Peter Sims
Emp.ID: 263
102 Main Street,Anytown, NY 10000.

Earnings	Rate	Hours	This Period	YTD
Regular Earnings	$100.00	40.00 hrs	$4,000.00	$20,000.00
Gross Pay			**$4,000.00**	**$20,000.00**

Taxes / Deductions	Type	This Period	YTD
	Federal Withholding	$339.54	$1,697.70
	FICA - Social Security	$248.00	$1,240.00
	FICA - Medicare	$58.00	$290.00
	State Withholding	$264.00	$1,320.00
	Employer Taxes		
	FUTA	$1.00	$11.00

{'Classes': [{'Name': 'Pay Stubs', 'Score': 0.9994999766349792}, {'Name': 'Bank Statements', 'Score': 0.0005000000237487257}], 'ResponseMetadata': {'RequestId': '60d202a0-0fce-4185-a96a-e9bbfc53a81f', 'HTTPStatusCode': 200, 'HTTPHeaders': {'x-amzn-requestid': '60d202a0-0fce-4185-a96a-e9bbfc53a81f', 'content-type': 'application/x-amz-json-1.1', 'content-length': '117', 'date': 'Wed, [redacted]}, 'RetryAttempts': 0}}

No remaining tasks in the current job.
Select another job and choose **Start working** to work on new tasks.

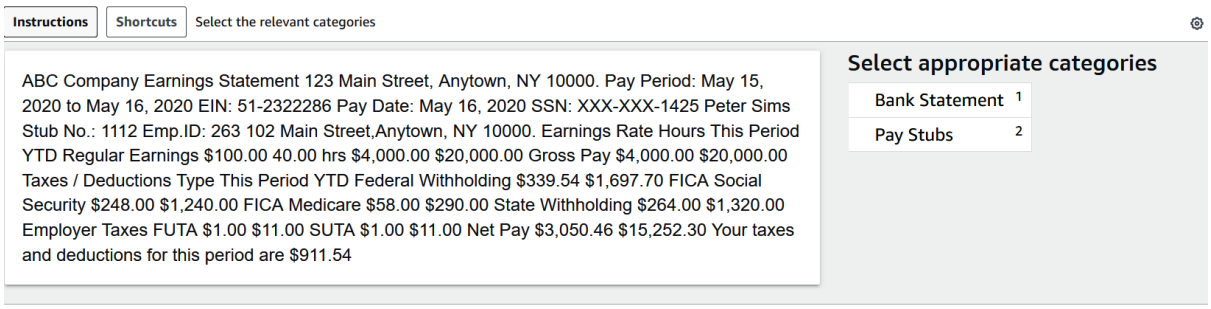
Show instructions

Jobs (1)

Start working

1

Task title	Customer ID	Status	Creation time
Classify the text	[redacted]	Available	[redacted]



Treat the data in this task as confidential.


```
output/custom-classify4d1a13be-0ad3-11ec-a906-06c3864a6c5f-fd-a2i/[redacted]/1ac1e472-572b-4167-ae34-f8d139f
08dc7/output.json
doc-processing-bucket-[redacted]
{   'flowDefinitionArn': 'arn:aws:sagemaker:us-east-1:[redacted]:flow-definition/custom-classify4d1a13be-0ad3-11ec-
a906-06c3864a6c5f-fd-a2i',
    'humanAnswers': [   {   'acceptanceTime': '2021-09-01T03:51:17.363Z',
                            'answerContent': {   'category': {   'labels': [   'Pay '
                                                                               'Stubs']}},
                            'submissionTime': '2021-09-01T03:52:49.959Z',
                            'timeSpentInSeconds': 92.596,
                            'workerId': 'a4d8700afb91b37f',
                            'workerMetadata': {   'identityData': {   'identityProviderType': 'Cognito',
                                                                      'issuer': 'https://cognito-idp.us-east-1.amazon
aws.com/us-east-1_VHJkg7Dra',
                                                                      'sub': 'd0b307cb-3a7d-483f-9933-e6bb2f2a2aec
'}}}}],
    'humanLoopName': '1ac1e472-572b-4167-ae34-f8d139f08dc7',
    'inputContent': {   'taskObject': 'ABC Company\n'
                                      'Earnings Statement\n'
                                      '123 Main Street,\n'
                                      'Anytown, NY 10000.\n'
                                      'Pay Period: May 15, 2020 to May 16, '
                                      '2020\n'
                                      'EIN: 51-2322286\n'
                                      'Pay Date: May 16, 2020\n'
                                      'SSN: XXX-XXX-1425\n'
                                      'Peter Sims\n'
```

Chapter 16: Improving the Accuracy of PDF Batch Processing

	DocName	LineNr	DetectedText	Confidence	CorrectedText	Comments
0	form-s20-LRHL-registration.pdf	9	Holdings	99.5		
1	form-s20-LRHL-registration.pdf	11	123	99.7		
2	form-s20-LRHL-registration.pdf	12	Unicoumstleet Supernova City, NY 99999 and Pho...	91.9		
3	form-s20-LRHL-registration.pdf	13	(Address, including zip code, and telephone nu...	99.8		
4	form-s20-LRHL-registration.pdf	15	@###	96.4		
5	form-s20-LRHL-registration.pdf	16	(Name, address, including zip code, and teleph...	99.5		
6	form-s20-LRHL-registration.pdf	17	Approximate date of commencement of proposed s...	99.9		
7	form-s20-LRHL-registration.pdf	18	October 20th, 2022	99.8		
8	form-s20-LRHL-registration.pdf	19	Calculation of Registration Fee	99.8		
9	form-s20-LRHL-registration.pdf	20	Title: LfRHL	76.1		
10	form-s20-LRHL-registration.pdf	21	Amount: 1000033	72.4		
11	form-s20-LRHL-registration.pdf	22	Fee per unit:	99.6		
12	form-s20-LRHL-registration.pdf	23	Maximum Fee:	99.6		
13	form-s20-LRHL-registration.pdf	24	Registration Fee:	99.7		
14	form-s20-LRHL-registration.pdf	25	$130	99.0		
15	form-s20-SUBS1-registration.pdf	9	LiveRight Holdings Subsidiary. 1	90.1		
16	form-s20-SUBS1-registration.pdf	11	123-Unicorn Street, Supemovar City; NY.99999 a...	68.1		

Hello, [redacted] Customer ID: [redacted] Task description: Review the contents and co... Task time: 22:06 of 59 Min 17 Sec

Instructions

Please review the SEC registration form inputs, and make corrections where appropriate.

Original Registration Form - Page 1

UNITED STATES
SECURITIES AND EXCHANGE COMMISSION
Washington, D.C. 20549

OMB APPROVAL
Not subject to -
P.L. 96-511

FORM S-20

REGISTRATION STATEMENT UNDER THE SECURITIES ACT OF 1933

LiveRight Holdings
(Exact name of registrant as specified in its charter)

123 Unicorn Street, Supernova City, NY 99999 and Phone: 999-123-4567
(Address, including zip code, and telephone number, including area code, of registrant's principal executive offices)

LiveRight Brokerage, 456, Pathless Road, Lightness City, NY 88888 and Phone: 888-456-7890 @###
(Name , address, including zip code, and telephone number, including area code, of agent for service)

Approximate date of commencement of proposed sale to the public October 20th, 2022

Calculation of Registration Fee

Title: LfRHL	Amount: 1000033	Fee per unit: $130	Maximum Fee: $10M	Registration Fee: $10K

Note: Specific details relating to the fee calculation shall be furnished in notes to the table, including references to provisions of Rule 457 (§230.457 of this chapter) relied upon, if the basis of the calculation is not otherwise evident from the information presented in the table.

Please enter your modifications below

Line Nr	Detected Text	Confidence	Change Required	Corrected Text
9	LiveRight Holdregs Subsidiary 2	86.3	○ Correct ◉ Incorrect	LiveRight Holdings Subsidiary 2
11	123 Unicorn Street, Supernova Rity, 99999 and Phone: 999-123-4567	86.9	○ Correct ◉ Incorrect	123 Unicorn Street, Supernova City, 99999 and Phone: 999-123-4567
18	Title: SUB555	77.5	○ Correct ◉ Incorrect	Title: SUBS1
19	Amgyst: 1999000	21.6	○ Correct ◉	Amount: 1999000

	DocName	LineNr	DetectedText	Confidence	CorrectedText	Comments
0	form-s20-LRHL-registration.pdf	9	Holdings	99.5		
1	form-s20-LRHL-registration.pdf	11	123	99.7		
2	form-s20-LRHL-registration.pdf	12	Unicoumstleet Supernova City, NY 99999 and Pho...	91.9	123 Unicorn Street, Supernova City, NY 99999 a...	Street name was missing
3	form-s20-LRHL-registration.pdf	13	(Address, including zip code, and telephone nu...	99.8		
4	form-s20-LRHL-registration.pdf	15	@###	96.4		
5	form-s20-LRHL-registration.pdf	16	(Name, address, including zip code, and teleph...	99.5		
6	form-s20-LRHL-registration.pdf	17	Approximate date of commencement of proposed s...	99.9		
7	form-s20-LRHL-registration.pdf	18	October 20th, 2022	99.8		
8	form-s20-LRHL-registration.pdf	19	Calculation of Registration Fee	99.8		
9	form-s20-LRHL-registration.pdf	20	Title: LfRHL	76.1	Title: LRHL	typo in title
10	form-s20-LRHL-registration.pdf	21	Amount: 1000033	72.4	Amount: 100000	incorrect amount

Items returned (45)

Actions ▼ | Create item

Find items

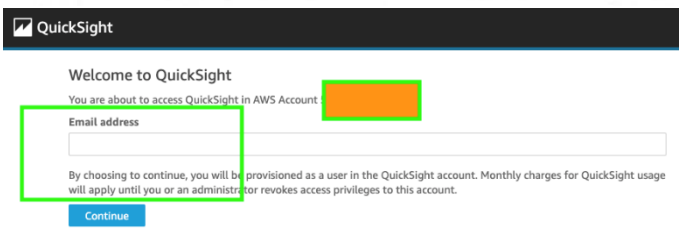
1

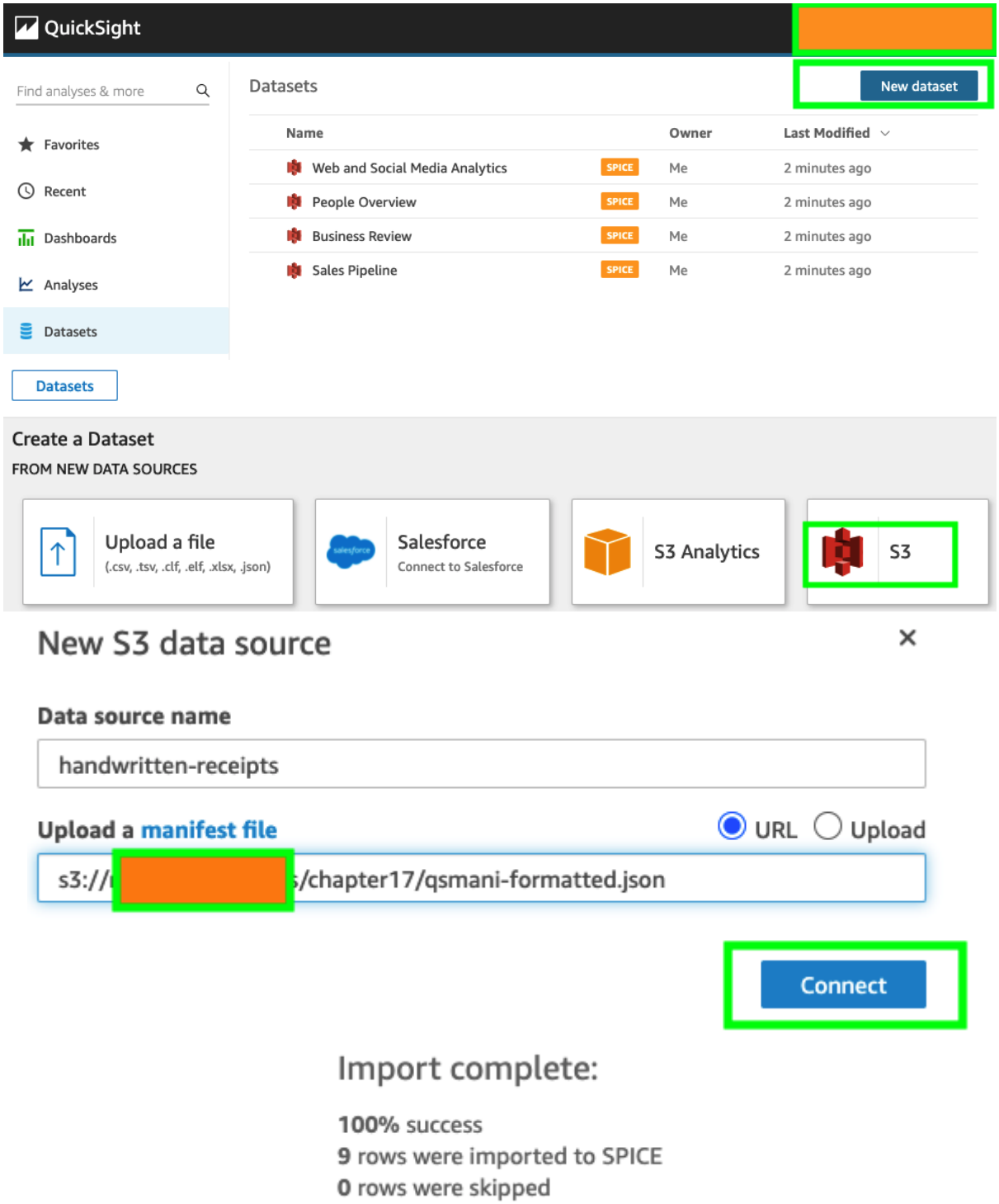
row_nr ▲	change_comments	confidence	corrected_line	detected_line	doc_name	line_nr
0	<empty>	99.5	<empty>	Holdings	form-s20-LRHL-registrat...	9
1	<empty>	99.7	<empty>	123	form-s20-LRHL-registrat...	11
2	Street name was missing	91.9	123 Unicorn Street, Supernova...	Unicoumstleet Supernova City, N...	form-s20-LRHL-registrat...	12
3	<empty>	99.8	<empty>	(Address, including zip code, and ...	form-s20-LRHL-registrat...	13
4	<empty>	96.4	<empty>	@###	form-s20-LRHL-registrat...	15
5	<empty>	99.5	<empty>	(Name, address, including zip cod...	form-s20-LRHL-registrat...	16
6	<empty>	99.9	<empty>	Approximate date of commence...	form-s20-LRHL-registrat...	17
7	<empty>	99.8	<empty>	October 20th, 2022	form-s20-LRHL-registrat...	18
8	<empty>	99.8	<empty>	Calculation of Registration Fee	form-s20-LRHL-registrat...	19
9	typo in title	76.1	Title: LRHL	Title: LfRHL	form-s20-LRHL-registrat...	20
10	incorrect amount	72.4	Amount: 100000	Amount: 1000033	form-s20-LRHL-registrat...	21

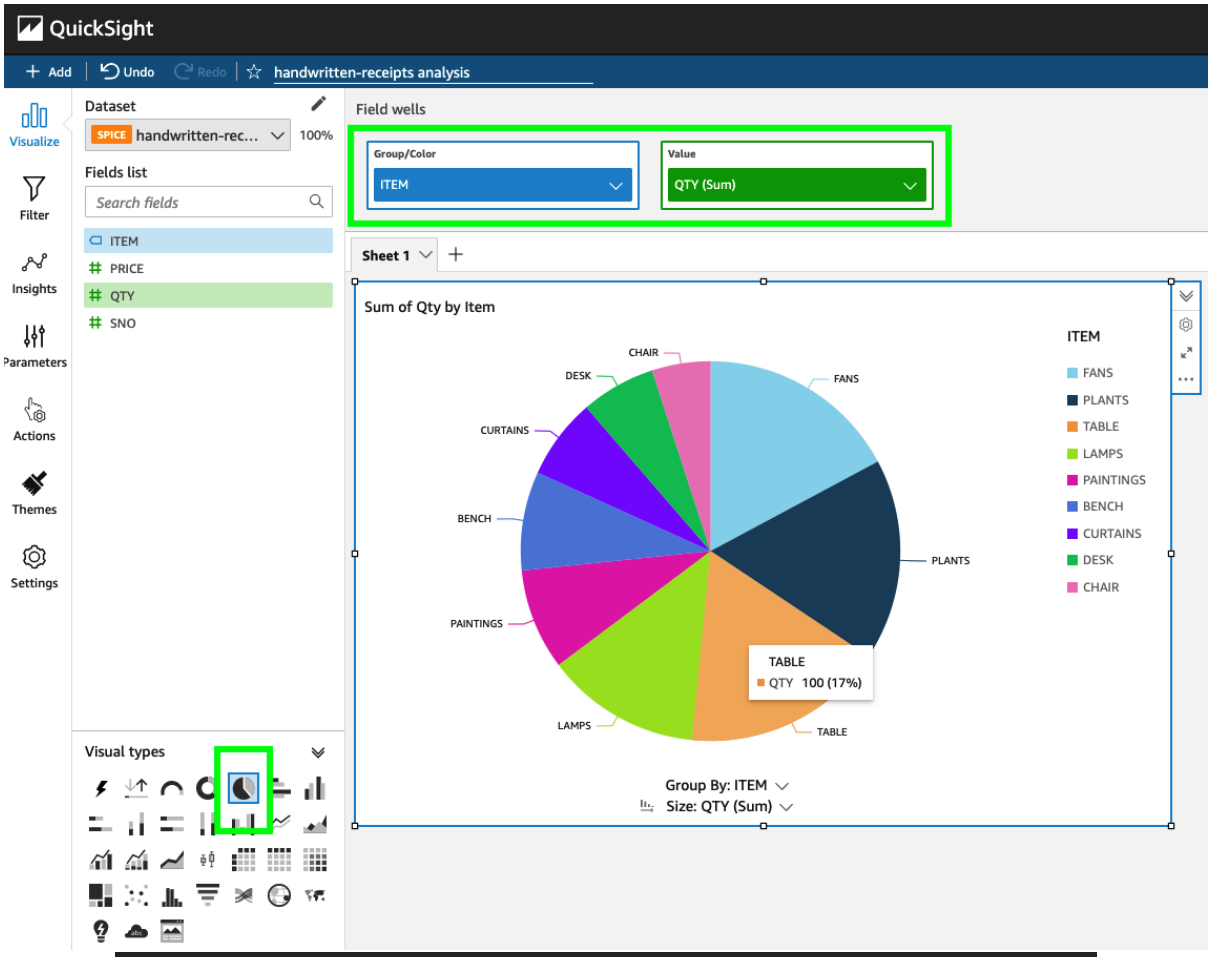
Chapter 17: Visualizing Insights from Handwritten Content

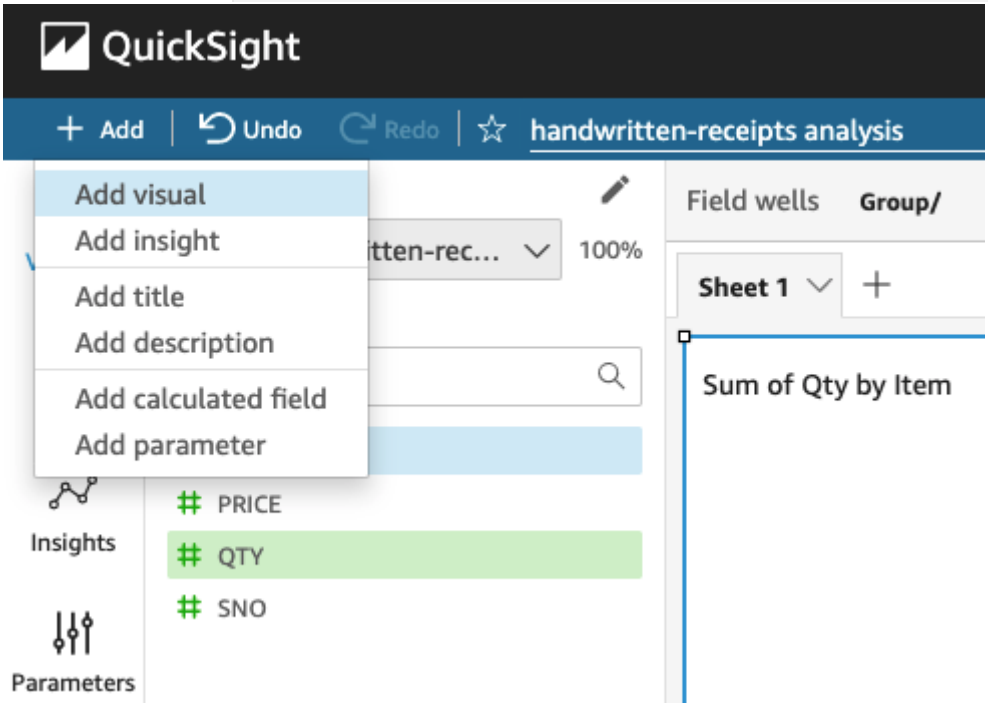
ABC Furnitures

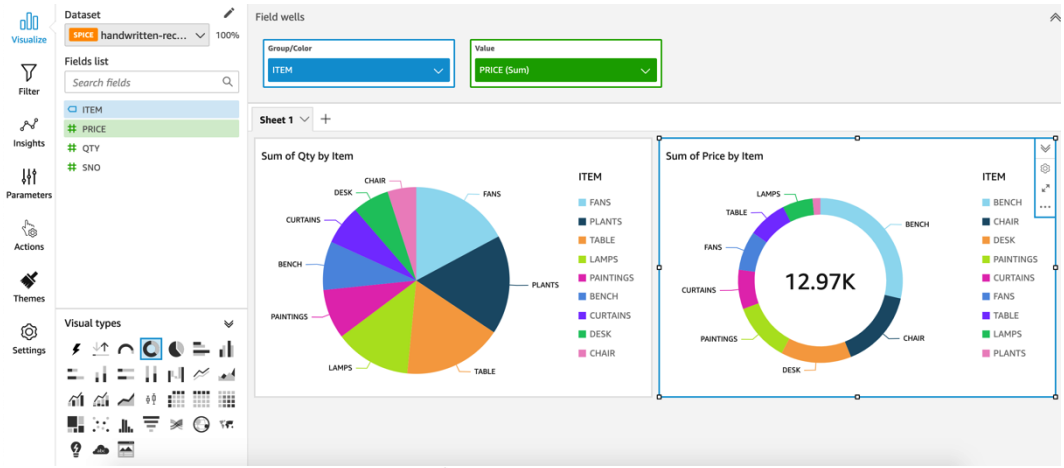
SNO	ITEM	QTY	PRICE
1	TABLE	100	$1000
2	CHAIR	29	$2000
3	DESK	37	$1800
4	BENCH	49	$3700

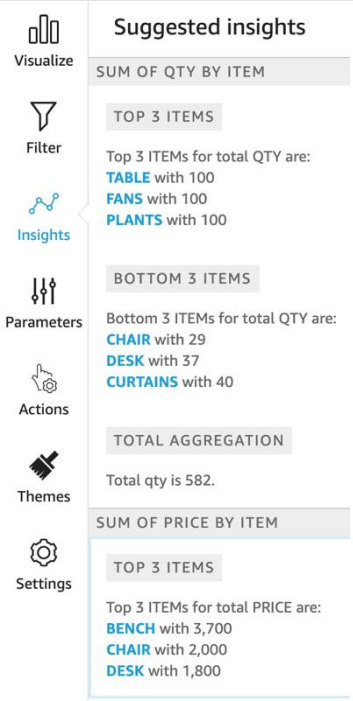



QuickSight
Find analyses & more
Favorites
Recent
Dashboards
Analyses
Datasets
Datasets
Datasets
New dataset
Name
Owner
Last Modified
Web and Social Media Analytics
SPICE
Me
2 minutes ago
People Overview
SPICE
Me
2 minutes ago
Business Review
SPICE
Me
2 minutes ago
Sales Pipeline
SPICE
Me
2 minutes ago
Create a Dataset
FROM NEW DATA SOURCES
Upload a file
(.csv, .tsv, .clf, .elf, .xlsx, .json)
Salesforce
Connect to Salesforce
S3 Analytics
S3
New S3 data source
Data source name
handwritten-receipts
Upload a manifest file
URL
Upload
s3://
/chapter17/qsmani-formatted.json
Connect
Import complete:
100% success
9 rows were imported to SPICE
0 rows were skipped


QuickSight
Add
Undo
Redo
handwritten-receipts analysis
Visualize
Filter
Insights
Parameters
Actions
Themes
Settings
Dataset
SPICE handwritten-rec...
100%
Fields list
Search fields
ITEM
PRICE
QTY
SNO
Field wells
Group/Color
ITEM
Value
QTY (Sum)
Sheet 1
Sum of Qty by Item
CHAIR
DESK
CURTAINS
BENCH
PAINTINGS
LAMPS
FANS
PLANTS
TABLE
TABLE
QTY 100 (17%)
ITEM
FANS
PLANTS
TABLE
LAMPS
PAINTINGS
BENCH
CURTAINS
DESK
CHAIR
Group By: ITEM
Size: QTY (Sum)
Visual types

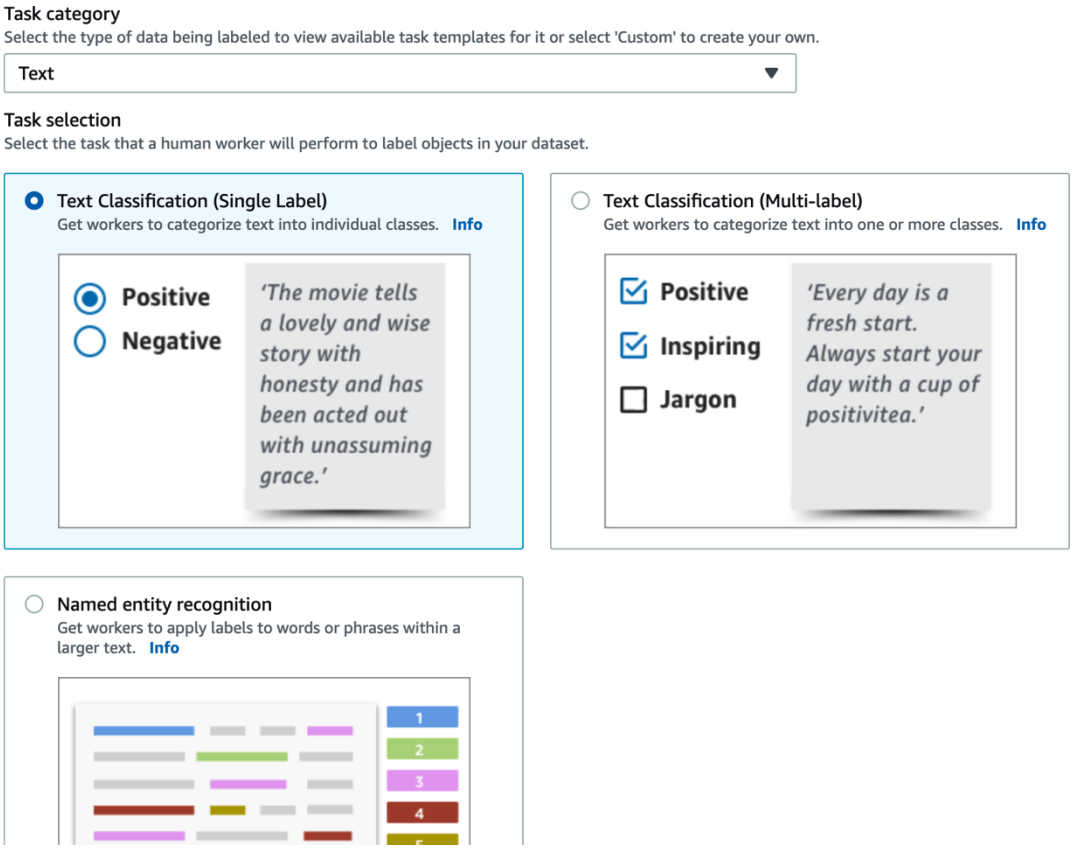

QuickSight
Add
Undo
Redo
handwritten-receipts analysis
Add visual
Add insight
Add title
Add description
Add calculated field
Add parameter
100%
Field wells
Group/
Sheet 1
Sum of Qty by Item
Insights
Parameters
PRICE
QTY
SNO


Visualize
Filter
Insights
Parameters
Actions
Themes
Settings
Dataset
SPICE handwritten-rec...
100%
Fields list
Search fields
ITEM
PRICE
QTY
SNO
Visual types
Field wells
Group/Color
ITEM
Value
PRICE (Sum)
Sheet 1
Sum of Qty by Item
CHAIR
DESK
FANS
CURTAINS
BENCH
PLANTS
PAINTINGS
LAMPS
TABLE
ITEM
FANS
PLANTS
TABLE
LAMPS
PAINTINGS
BENCH
CURTAINS
DESK
CHAIR
Sum of Price by Item
LAMPS
TABLE
BENCH
FANS
12.97K
CURTAINS
PAINTINGS
CHAIR
DESK
ITEM
BENCH
CHAIR
DESK
PAINTINGS
CURTAINS
FANS
TABLE
LAMPS
PLANTS


Visualize
Filter
Insights
Parameters
Actions
Themes
Settings
Suggested insights
SUM OF QTY BY ITEM
TOP 3 ITEMS
Top 3 ITEMs for total QTY are:
TABLE with 100
FANS with 100
PLANTS with 100
BOTTOM 3 ITEMS
Bottom 3 ITEMs for total QTY are:
CHAIR with 29
DESK with 37
CURTAINS with 40
TOTAL AGGREGATION
Total qty is 582.
SUM OF PRICE BY ITEM
TOP 3 ITEMS
Top 3 ITEMs for total PRICE are:
BENCH with 3,700
CHAIR with 2,000
DESK with 1,800

Chapter 18: Building Secure, Reliable, and Efficient NLP Solutions

AWS Well-Architected Framework Pillars					
Stages of NLP development	Operational excellence	Security	Reliability	Performance efficiency	Cost optimization
Document pre-processing	1.1a – Collect data from disparate sources into an S3 Data Lake.	1.2a – Use S3 Object Lambda Access Points for PII. 1.2b – Encrypt data at rest. 1.2c – Enforce least privilege access.	1.3a – Use an S3 Data Lake.	1.4a – Use AWS Glue for pre-processing. 1.4b – Use SageMaker Ground Truth for annotations. 1.4c – Use Amazon Comprehend with PDF, Word directly.	1.5a – Save costs with serverless pre-processing using AWS Glue.
Prediction	2.1a – Use API Gateway for throttling requests to Amazon Comprehend custom real-time endpoints.	2.2a – Encrypt data in transit.	2.3a – Set up Auto Scaling for Amazon Comprehend real-time endpoints. 2.3b – Use Amazon A2I to review predictions.	2.4a – Use .async batch for bulk predictions.	2.5a – Save on training costs with pre-trained Textract and Amazon Comprehend models for common NLP tasks.
Document post-processing	3.1a – Use AWS Glue for transformation.	3.2a – Encrypt data at rest. 3.2b – Enforce least privilege access.	3.3a – Persist results.	3.4a – Use Textract Response Parser Library. 3.4b – Use AWS Glue for serverless transformation.	3.5a – Save costs with AWS Glue for serverless transformation.
Overall NLP solution build	4.1a – Use CloudFormation templates. 4.1b – Use AWS Step Function for orchestration.	4.2a – Use least privilege IAM policies. 4.2b – Obfuscate sensitive data. 4.3b – Protect data at rest and in transit.	4.3a – Automate monitoring of custom training metrics.	4.4a – Use more async APIs to build a loosely coupled architecture.	4.5a – Save costs using managed services when available, such as Amazon Textract and Amazon Comprehend for NLP.



Annotation and data format

Configure how you are providing your data.

Data format

To train your custom model, you must provide training data. This data must be formatted as either a CSV file or as one or more augmented manifest files.

- ○ **CSV file** Info
  The CSV file that contains either the annotations or the entity lists for your training data. The required format depends on the type of CSV file that you provide.
- ◉ **Augmented manifest** Info
  A labeled training dataset that is produced by Amazon SageMaker Ground Truth. You can provide up to 5 augmented manifest files. To create an augmented manifest file, you can create a labeling job in Amazon SageMaker Ground Truth.

Training and test dataset

Training model type

○ Plain text documents	◉ PDF, Word documents
Choose this option if your augmented manifest labeling job was used to label plain text documents.	Choose this option if your augmented manifest labeling job was used to label PDF documents.

Augmented manifest locations and attribute names

Collapse all

You can provide up to 5 augmented manifest files. Each file can reside in a separate location. For each file, you can name up to 5 attributes to use as training data.

▼ **Input location** Remove

SageMaker Ground Truth augmented manifest file S3 location

Paste the URL of an input data file in S3, or select a bucket or folder location in S3.

s3://MyBucketName/output.manifest Browse S3

Amazon Comprehend ×

- Real-time analysis
- Analysis jobs
- ▼ Customization
  - Custom classification
  - Custom entity recognition
  - Endpoints
- ▼ Amazon Comprehend Medical
  - Real-time analysis
  - Analysis jobs

Hello Zhang Wei, I am John. Your AnyCompany Financial Services, LLC credit card account 1111-0000-1111-0008 has a minimum payment of $24.53 that is due by July 31st. Based on your autopay settings, we will withdraw your payment on the due date from your bank account number XXXXXX1111 with the routing number XXXXX0000.

Your latest statement was mailed to 100 Main Street, Any City, WA 98121.
After your payment is received, you will receive a confirmation text message at 206-555-0100.
If you have questions about your bill, AnyCompany Customer Service is available by phone at 206-555-0199 or email at support@anycompany.com.

▼ Results

Search 1 2

Entity	Type	Confidence
Zhang Wei	Name	0.99+
John	Name	0.99+
1111-0000-1111-0008	Credit debit number	0.99+
July 31st	Date time	0.99+
XXXXXX1111	Bank account number	0.99+
XXXXX0000	Bank routing	0.99+

Default encryption

Automatically encrypt new objects stored in this bucket. **Learn more**

Server-side encryption

- ○ Disable
- ◉ Enable

Encryption key type

To upload an object with a customer-provided encryption key (SSE-C), use the AWS CLI, AWS SDK, or Amazon S3 REST API.

- ◉ Amazon S3 key (SSE-S3)
  An encryption key that Amazon S3 creates, manages, and uses for you. Learn more
- ○ AWS Key Management Service key (SSE-KMS)
  An encryption key protected by AWS Key Management Service (AWS KMS). Learn more

Cancel | Save changes

Select custom model type

Choose a custom model type that you would like to assign this endpoint to. After you create an endpoint, you will be able to edit the version it's pointing to but the model type must be the same. To train a custom model, visit **Custom classification** or **Custom entity recognition.**

- ○ Custom classification
- ◉ Custom entity recognition

Select recognizer model

Text-Analysis-Custom-Entity-Recognizer840370 - No Version Name
April 15, 2021, 18:58 (UTC-04), F1 score: 98.99 %, Active endpoints: 0

▶ **Tags - *optional*** Info

A tag is a label that you can add to a resource as metadata to help you organize, search, or filter your data. Each tag consists of a key and an optional value.

Inference units

IU estimator

Number of inference units (IUs)

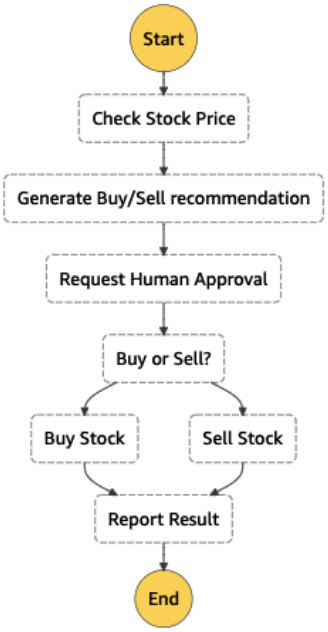
A single IU can analyze one hundred characters per second. IU is a measure of the endpoint's throughput. You can adjust the IU of an endpoint anytime.

1 IU ≤ 100 characters/second

Range from 1 to 10

Definition

Sample project code is read-only, but you can edit it after creation. State machines are defined using Amazon States Language. Learn more

Export ▼ Layout ▼

```
{
    "StartAt": "Check Stock Price",
    "Comment": "An example of itegrating lambda functtions in Step Functions state machine.",
    "States": {
        "Check Stock Price": {
            "Type": "Task",
            "Resource": "<CHECK_STOCK_PRICE_LAMBDA_ARN>",
            "Next": "Generate Buy/Sell recommendation
        },
        "Generate Buy/Sell recommendation": {
            "Type": "Task",
            "Resource": "<GENERATE_BUY_SELL_RECOMMENDATION_LAMBDA_ARN>",
            "ResultPath": "$.recommended_type",
            "Next": "Request Human Approval"
        },
        "Request Human Approval": {
            "Type": "Task",
            "Resource": "arn:<PARTITION>:states:::sqs:sendMessage.waitForTaskToken
            "Parameters": {
                "QueueUrl": "<REQUEST_HUMAN_APPROVAL_SQS_URL>",
```