

Setting Up Unstructured Data in Data Cloud

Author: [Alex Conway](#), Technical Architect UKI.

Last Update:Lundi 10 Juin

This document contains instructions for setting up a UDLO & UDMO in Data Cloud, and the file notification pipeline to enable ingestion of unstructured data using AWS.

This assumes you have a Data Cloud org with the Unstructured Data Pilot enabled. You can ask PM to enable it using the following sign-up form: <https://salesforce.quip.com/d5x4AtxfoMev>

Set up the AWS CLI

You will need to use the AWS CLI when following these instructions. This assumes you have a user with the appropriate permissions and have obtained an access key & secret access key.

Install the AWS CLI following the instructions here.

<https://docs.aws.amazon.com/cli/latest/userguide/getting-started-install.html#getting-started-install-instructions>

I used the (not recommended) approach of using IAM and long-term credentials to authorise the CLI.

<https://docs.aws.amazon.com/cli/latest/userguide/cli-authentication-user.html>

Alternatively, [see these instructions](#) from [Troy Sellers](#) on how to set this up with SSO & Embark.

Then configure the AWS CLI using the configure command `aws configure` and you will be prompted for the following details.

Your access key

`AWS Access Key ID [None] : AKIAIOSFODNN7EXAMPLE`

Your secret

`AWS Secret Access Key [None] : wJalrXUtnFEMI/K7MDENG/bPxRfiCYEXAMPLEKEY`

Your region (examples: eu-central-1, us-west-1, etc.)

`Default region name [None] : us-west-1`

Output format.

Default output format [None]: json

If you have already configured your AWS CLI and you want to change the default region, use the following command:

```
aws configure set default.region eu-central-1
```

This changes it to eu-central-1. Replace that value as appropriate.

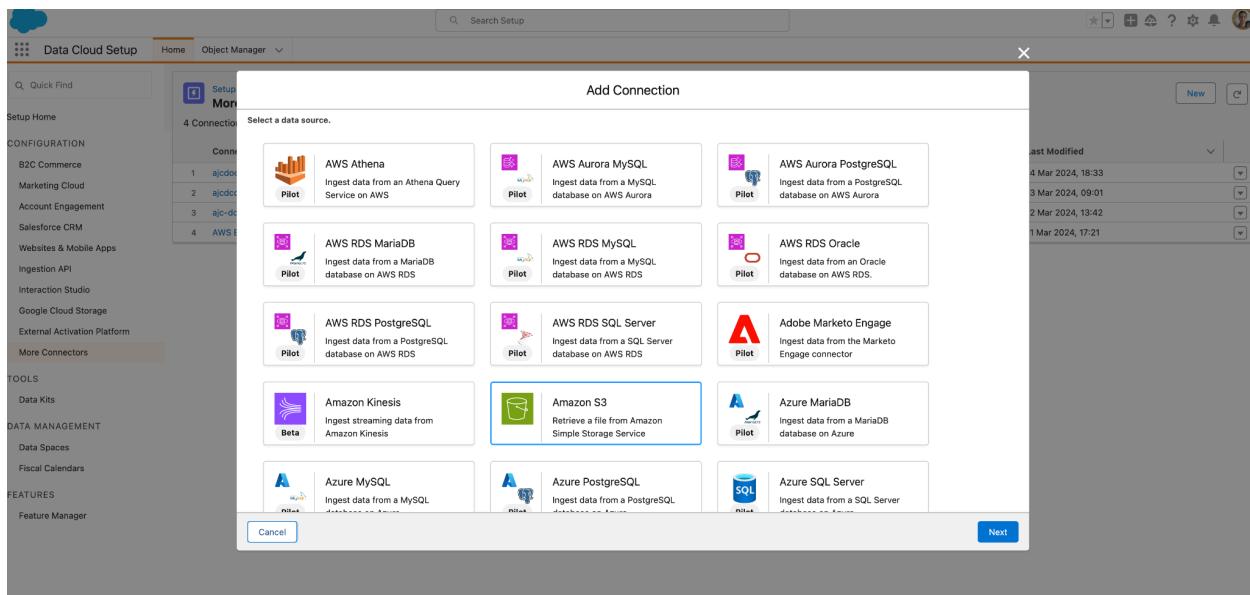
If your AWS CLI default region and the region where you wish to deploy your lambda/secrets/s3 do not match, you can add the --region flag to the aws commands to specify the region, e.g.

```
--region us-west-1
```

Create a Connection to AWS

Go to Data Cloud Setup.

Click on More Connectors and choose Amazon S3.



Put in the details for your bucket. (This is where you will upload your documents.)

Object Manager

New Amazon S3 Source

Connection Name: My AWS Bucket

Connection API Name: My_AWS_Bucket

Authentication Details

AWS access key:
AWS secret access key:

Connection Details

Bucket Name: ajcdcdocs

Parent Directory: /

Test Connection

Back **Save**

In this example, the S3 source is the root directory in the bucket “ajcdcdocs”.

Create a UDLO

Navigate to the Data Cloud app, and go to the Data Lake Objects tab.

Click on New, and then select “From External Files”

(Note: this is different to the normal process where you would create a Data Stream first. For Unstructured Data, you do not create a Data Stream)

Data Cloud Home Data Streams Data Spaces Data Lake Objects Data Model Data Explorer Identity Resolutions Profile Explorer Calculated Insights Data Action Targets Data Actions Segments More

Data Lake Objects Recently Viewed + 1

Search this list... New Update Status

Name	Category	Storage	Data Lake Object Status	Last Updated On	Total Records
AccountWithUltimateParentMatch	Profile	Local	Active	14/03/2024, 18:16	68
Help Policy	Directory Table	Local	Active		0

New Data Lake Object

Choose a method of creating your data lake object.

=
From External Files
 Create an unstructured data lake object from a source of unstructured data.

=
From Existing
 Create a data lake object from an existing object

🔧
New
 Create a data lake object from scratch

⬇️
Choose From Data Kits
 Create a data lake object from installed data kits

Cancel **Next**

ABC CDR-call	Engagement	Local	Active	20/02/2024, 16:07	1,000
Help Documents vector	Vector_EMBEDDING	Local	Active		0
Help Documents chunk	Content	Local	Active		0
Financial Reports vector	Vector_EMBEDDING	Local	Active		0
Financial Reports chunk	Content	Local	Active		0
Financial Reports	Directory Table	Local	Active		0
GenAIGatewayRequestTag	Engagement	Local	Active	15/03/2024, 10:58	175
GenAIContentCategory	Engagement	Local	Active	15/03/2024, 10:58	184

Select S3

Screenshot of the Data Cloud interface showing the "Data Lake Objects" section. A modal window titled "New Data Lake Object" is open, asking "Where is your unstructured data stored?". It lists three options: "Amazon S3" (selected), "Google Cloud Storage", and "Microsoft Azure Storage". Below the modal is a table of existing data lake objects.

Name	Category	Storage	Data Lake Object Status	Last Updated On	Total Records
AccountWithUltimateParentMatch	Profile	Local	Active	14/03/2024, 18:16	68
Help Policy	Directory Table	Local	Active		0

Choose the AWS connection you just created

Screenshot of the Data Cloud interface showing the "New Data Lake Object" configuration screen. The "Connection Details" section shows a dropdown menu with "ajodcdocs" selected. Below it, a "File Name Pattern" field contains ".txt". To the right, a sidebar titled "Data Lake Object Setup Guide" provides information about selecting unstructured data and extracting metadata.

Name	Category	Storage	Data Lake Object Status	Last Updated On	Total Records
ABC CDR-call	Engagement	Local	Active	20/02/2024, 16:07	1,000
Help Documents vector	Vector_EMBEDDING	Local	Active		0
Help Documents chunk	Content	Local	Active		0
Financial Reports vector	Vector_EMBEDDING	Local	Active		0
Financial Reports chunk	Content	Local	Active		0
Financial Reports	Directory Table	Local	Active		0
GenAIGatewayRequestTag	Engagement	Local	Active	15/03/2024, 10:58	175
GenAIContentCategory	Engagement	Local	Active	15/03/2024, 10:58	184

In this example, my files will be stored in the "documents" folder in my S3 bucket. The expected type of file is any document with TXT extension as denoted by using the pattern ***.txt**. (You can use one of either PDF, HTML or TXT).

New Data Lake Object

Set up target objects. Optionally enable semantic search on those objects.

Unstructured Data Lake Object Details
All matching files from all configured folders will be created into a single Unstructured Data Lake Object.

* Object Label: Help Documents
* Object API Name: Help_Documents

Unstructured Data Model Object Mapping
Select a Data Space and choose to map your unstructured data lake object to a new or an existing unstructured data model object. Both objects must be referenced in the same Data Space. You can optionally create a Semantic Search Index Configuration for the unstructured data model object. If you choose not to create one now, you can do so later.

New Existing
* Data Space: default
* Object Name: Help Documents
* Object API Name: Help_Documents

Search Indexation
 Enable semantic search with system defaults

How is data mapping configured?
Selecting Unstructured Data
When an unstructured data connector is created, monitor for changes from the moment it is created and sync them automatically.
Semantic Search Index Configuration
When you create Semantic Search Index Configuration for unstructured data model objects, you break the referenced data into manageable, semantically meaningful chunks that can be embedded as searchable vectors—numerical representations of unstructured data—in your LLMs and prompts.

Back Next

Give your UDLO a label and API name.

Provide a label and API name for your UDMO.

Leave Search Indexation enabled and the search index will be created without needing to do it as a separate step.

Click through and you'll end up with 3 new UDLOs.

Recently Viewed

Name	Category	Storage	Data Lake Object Status	Last Updated On	Total Records
Help Documents	Directory Table	Local	Active	0	0
Help Documents vector	Vector_Embdding	Local	Active	0	0
Help Documents chunk	Content	Local	Active	0	0

You should also have some new UDMOs.

Data Model Objects Mapped

Object Label	Object API Name	Data Streams	Data Lake Objects	Data Space	Type	Status
Help Documents	Help_Documents_dlm		Help Documents	default	Custom	Ready
Help Documents chunk	Help_Documents_chunk_dlm		Help Documents chunk	default	Semantic	Ready
Help Documents vector	Help_Documents_vector_dlm		Help Documents vector	default	Semantic	Ready

A Search Index should also have been created (unless you unticked the Search Indexation box in the previous step).

Search Index
Help_Documents

Data Space default	API Name Help_Documents	Source Data Model Object Help_Documents_dlm	Vector Data Model Object Help_Documents_vector_dlm	Chunk Data Model Object Help_Documents_chunk_dlm	Last Modified Date 14/03/2024, 11:57
---------------------------------------	--	--	---	---	---

Details Configuration

Information

Configuration Name Help_Documents	API Name Help_Documents
Source Data Model Object Help_Documents_dlm	Created Date 14/03/2024, 11:57
Search Index Refreshed On	Search Index Last Run Status In Progress

Post Poll Question

Share an update... Share

Search this feed...

Note: Files will not be ingested automatically. You need to create a notification pipeline so that when you add documents to your S3 bucket, the Data Cloud API is called for each one. The instructions to set up the pipeline for AWS follow below.

Creating the File Notification Pipeline in AWS

Create a Private/Public Key Pair and Certificate

First, create a folder to store these files.

Create private/public key pair

```
openssl genrsa -out keypair.pem 2048
```

Create a digital certificate from the key pair

```
openssl req -new -x509 -nodes -sha256 -days 365 -key keypair.pem -out certificate.crt
```

(You will need to answer a series of prompts)

Create a private key

```
openssl pkcs8 -topk8 -nocrypt -in keypair.pem -out private.key
```

Create a Connected App

Go to App Manager and create a new Connected App.

Enable OAuth settings for the app.

Enable Use Digital Signatures and upload the .crt file you just created.

Add the following scopes:

api / refresh_token,offline_access / cdp_ ingest_api

Put <http://localhost.com> as the callback URL

The screenshot shows the 'API (Enable OAuth Settings)' configuration screen. Key fields include:

- API Name: AWS
- Contact Email: alex.conway@salesforce.com
- Contact Phone: (empty)
- Logo Image URL: (empty)
- Icon URL: (empty)
- Info URL: (empty)
- Description: (empty)

In the 'API (Enable OAuth Settings)' section, the 'Enable OAuth Settings' checkbox is checked. Other settings include:

- Enable for Device Flow: unchecked
- Callback URL: http://localhost.com
- Use digital signatures: checked
- EMAILADDRESS=alex.conway@salesforce.com, CN=http://salesforce.com, OU=TA, O=Salesforce, L=London, ST=London, C=GB 13 Mar 2025 09:44:58 GMT

The 'Selected OAuth Scopes' section contains:

- Access Analytics REST API Charts Geodata resources (clair_api)
- Access Analytics REST API resources (wave_api)
- Access Connect REST API resources (chatter_api)
- Access Einstein GPT services (einstein_gpt_api)
- Access Headless Forgot Password API (forgot_password)
- Access Headless Passwordless Login API (pwdless_login_api)
- Access Headless Registration API (user_registration_api)
- Access Interaction API resources (interaction_api)
- Access Lightning applications (lightning)
- Access Visualforce applications (visualforce)

The 'Available OAuth Scopes' section lists:

- Manage Data Cloud Ingestion API data (cdp_ ingest_api)
- Manage user data via APIs (api)

The 'Selected OAuth Scopes' section shows:

- Manage Data Cloud Ingestion API data (cdp_ ingest_api)
- Manage user data via APIs (api)

Other configuration options include:

- Require Proof Key for Code Exchange (PKCE): unchecked
- Extension for Supported Authorization Flows: (empty)
- Require Secret for Web Server Flow: checked
- Require Secret for Refresh Token Flow: checked

Go to Setup and in OAuth and OpenID Connect Settings, make sure Allow OAuth Username-Password Flows is enabled.

Wait a while (approx. ~10 minutes, sometimes longer) then paste the following into your browser
(*Line-break due to formatting.*)

```
YOUR_ORG_URL/services/oauth2/authorize?response_type=code&client_id=Y0
UR_CONSUMER_KEY&scope=api refresh_token
cdp_ ingest_api&redirect_uri=http://localhost.com&code_challenge=SHA256
```

(Watch out for additional characters when you paste this. There should be a space between api and refresh_token, and a space between refresh_token and cdp_ ingest_api. No other spaces or newlines.)

YOUR_ORG_URL is your org's domain name. Find it by going to Setup → My Domain and copying the value for "Current My Domain URL".

Example of what you should see in your browser:

https://ajc-cdo2-demo.my.salesforce.com/services/oauth2/authorize?response_type=code&client_id=3MVG9IXUyidRC0I1ZZZZZZZiqX2aH.I.TYs2sEFya_ycdQ1NaCxjXnX9KKnR1WSKqL.X0yVMNbFsfq4lhsw&scope=api%20refresh_token%20cdp_ ingest_api&redirect_uri=http://localhost.com&code_challenge=SHA256

Allow when prompted.

Download the Lambda function

The link in the Pilot doc is broken. You can download it from here:

https://drive.google.com/file/d/1m7BQMmSNHn_0zCsfkpxyhcjE7lFlhcQZ/view?usp=sharing

Create some environment variables to reference later.

Open a terminal window to execute the following commands.

Replace alex@ajc-cdo2.demo with your own username

```
export SF_USERNAME=alex@ajc-cdo2.demo
```

This assumes a production org.

```
export SF_LOGIN_URL=https://login.salesforce.com/
```

Replace the account Id with your AWS account Id.

```
export AWS_ACCOUNT_ID=6xxxxxxxxxx1
```

This can be any string. It's just a name.

```
export LAMBDA_ROLE="AJC_Einstein_LX"
```

This can be any string. It's just a name.

```
export LAMBDA_FUNCTION_NAME="AJC_FileNotifier"
```

This will be used to name your secret in AWS Secrets Manager, so call it whatever you like.

```
export RSA_PRIVATE_KEY=einstein1
```

This will be used to name your secret in AWS Secrets Manager, so call it whatever you like.

```
export CONSUMER_KEY=cdo2consumerkey
```

The absolute path is wherever you have stored the PEM file you created earlier. On Mac, you can copy the file in Finder and then paste it into terminal and this will provide the absolute path.

```
export  
LOCAL_ABSOLUTE_PATH_TO_KEYPAIR_PEM_FILE="/Users/alex.conway/Code/awskey/keypair.pem"
```

You will need to upload the Lambda Function to an S3 bucket. Name the bucket whatever you like (though it must be globally unique).

```
export LAMBDA_FUNCTION_S3_BUCKET=ajcdcdc1
```

Save the absolute path for the ZIP file you downloaded.

The absolute path is wherever you have stored the ZIP file you created earlier on your Mac. You can copy the file in Finder and then paste it into terminal and this will provide the absolute path text.

```
export  
PATH_TO_ZIP_FILE="/Users/alex.conway/Downloads/unstructured_package.zip"
```

Create the S3 bucket and upload the Lambda function

Note: Auto formatting converted any double hyphens into a long dash. I've manually corrected them, but might have missed one. For all the AWS CLI commands, the parameters should be prefixed with a double hyphen, not a single dash not a single hyphen. See screenshot for how it should look.

```
aws lambda invoke --region us-west-1 --function-name $LAMBDA_FUNCTION_NAME response.json
```

Create a bucket to store the lambda function.

```
aws s3 mb s3://$LAMBDA_FUNCTION_S3_BUCKET
```

Copy the zip file from local storage to the bucket you just created. Use the absolute path. On Mac, you can copy the file in Finder and then paste it into terminal and this will provide the absolute path.

(This is a single command, line-break due to formatting.)

```
aws s3 cp /Users/alex.conway/Downloads/unstructured_package.zip  
s3://$LAMBDA_FUNCTION_S3_BUCKET/$PATH_TO_ZIPFILE
```

You will need the “key” for the zip file you have uploaded. You can find this in the properties for the file in S3 (see screenshot below for example).

```
export  
LAMBDA_FUNCTION_S3_KEY="/Users/alex.conway/Downloads/unstructured_package.zip"
```

[Amazon S3](#) > [Buckets](#) > [ajcdcdoc1](#) > / > [Users/](#) > [alex.conway/](#) > [Downloads/](#) > [unstructured_package.zip](#)

unstructured_package.zip [Info](#)

[Properties](#)

[Permissions](#)

[Versions](#)

Object overview

Owner

alex.conway

AWS Region

US West (N. California) us-west-1

Last modified

March 14, 2024, 09:50:34 (UTC+00:00)

Size

22.7 MB

Type

zip

Key

 [/Users/alex.conway/Downloads/unstructured_package.zip](#)

Create an execution role for the Lambda function

(This is a single command, line-break due to formatting.)

Unset

```
aws iam create-role --role-name $LAMBDA_ROLE
--assume-role-policy-document '{"Version": "2012-10-17", "Statement": [ { "Effect": "Allow",
```

Save the role ARN to a variable

(This is a single command, line-break due to formatting.)

```
export LAMBDA_ROLE_ARN=$(aws iam get-role --role-name $LAMBDA_ROLE
--query 'Role.Arn' --output text)
```

You can test if the value has been stored correctly by using the echo command.

```
echo $LAMBDA_ROLE_ARN
```

Create secrets in AWS Secrets Manager

Create 1 secret to store the private key and 1 secret to store the consumer key from your connected app.

Private Key

(This is a single command, line-break due to formatting.)

```
aws secretsmanager create-secret --name $RSA_PRIVATE_KEY  
--secret-string file:///$LOCAL_ABSOLUTE_PATH_TO_KEYPAIR_PEM_FILE
```

Consumer Key

(This is a single command, line-break due to formatting.)

```
aws secretsmanager create-secret --name $CONSUMER_KEY --secret-string  
REPLACE_WITH_YOUR_CONSUMER_KEY (where REPLACE_WITH_YOUR_CONSUMER_KEY  
is the actual value for your Connected App consumer key.)
```

You will need to create a policy to attach to the secrets. Create an empty file in the current folder.

```
touch policy.json
```

Open the file in a plain text editor and paste the following policy and save.

```
{"Version": "2012-10-17", "Statement": [ {"Effect": "Allow",  
"Principal": {"AWS":  
"arn:aws:iam::$AWS_ACCOUNT_ID:role/$LAMBDA_ROLE" }, "Action":  
"secretsmanager:GetSecretValue", "Resource": "*" } ]}
```

Replace the variables \$AWS_ACCOUNT_ID and \$LAMBDA_ROLE with the actual values when saving the file. Tip: You've already saved these in the command line console so you can retrieve the values using the following commands and copying the responses:

```
echo $AWS_ACCOUNT_ID  
echo $LAMBDA_ROLE
```

You should end up with something that looks like this:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "AWS": "arn:aws:iam::12345678:role/lambdaRole"
      },
      "Action": "secretsmanager:GetSecretValue",
      "Resource": "*"
    }
  ]
}
```

Attach the policy to each secret (this assumes the policy.json file is in the current working directory).

(This is a single command, line-break due to formatting.)

```
aws secretsmanager put-resource-policy --secret-id $RSA_PRIVATE_KEY
--resource-policy file://policy.json --block-public-policy
```

(This is a single command, line-break due to formatting.)

```
aws secretsmanager put-resource-policy --secret-id $CONSUMER_KEY
--resource-policy file://policy.json --block-public-policy
```

An example of what it should look like is shown below:

The screenshot shows the AWS Secrets Manager 'Overview' page for a secret named 'cd02consumerkey'. In the 'Resource permissions' section, the JSON policy is displayed:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "AWS": "arn:aws:iam::651377595214:role/AIC_Einstein_LX"
      },
      "Action": "secretsmanager:GetSecretValue",
      "Resource": "*"
    }
  ]
}
```

Note: If you accidentally create the secrets in the wrong region, you can migrate them in the console. Click into the secret, then click on the Replicate secret to other Regions button. Make sure to activate after deploying to the new region.

Deploy the Lambda Function

Deploy the Lambda function.

(This is a single command, line-break due to formatting.)

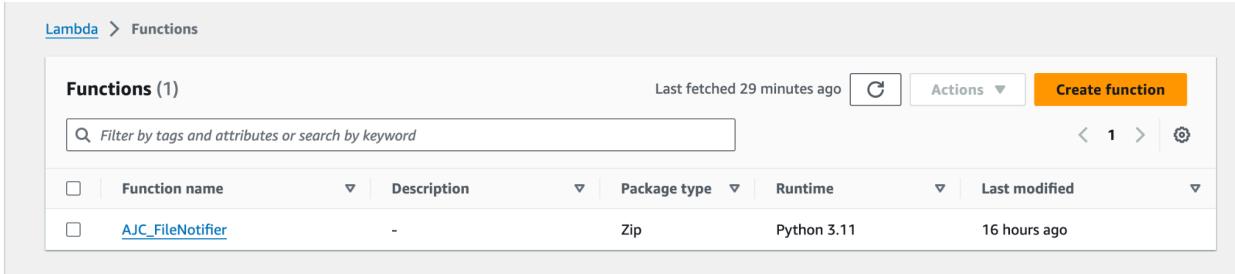
```
aws lambda create-function --function-name $LAMBDA_FUNCTION_NAME  
--runtime python3.11 --handler unstructured_data.s3_events_handler  
--role $LAMBDA_ROLE_ARN --code  
S3Bucket=$LAMBDA_FUNCTION_S3_BUCKET,S3Key=$LAMBDA_FUNCTION_S3_KEY  
--environment  
"Variables={SF_LOGIN_URL=${SF_LOGIN_URL},SF_USERNAME=${SF_USERNAME},RSA_PRIVATE_KEY=${RSA_PRIVATE_KEY},CONSUMER_KEY=${CONSUMER_KEY}}"  
--timeout 60
```

(If this command doesn't seem to be finish and return to the prompt after a while, you should be able to press the 'q' key to terminate in the console).

Create a trigger on the S3 bucket

Go to the AWS Management Console in your web browser.

Navigate to Lambda



The screenshot shows the AWS Lambda Functions list page. At the top, there's a breadcrumb navigation: Lambda > Functions. Below that is a search bar with placeholder text "Filter by tags and attributes or search by keyword". To the right of the search bar are buttons for "Actions" and "Create function". A table below lists the function "AJC_FileNotifier". The table has columns: "Function name", "Description", "Package type", "Runtime", and "Last modified". The "AJC_FileNotifier" row shows: Function name (AJC_FileNotifier), Description (-), Package type (Zip), Runtime (Python 3.11), and Last modified (16 hours ago). The table also includes a header row with sorting icons for each column.

	Function name	Description	Package type	Runtime	Last modified
<input type="checkbox"/>	AJC_FileNotifier	-	Zip	Python 3.11	16 hours ago

You should see your Lambda function listed. (If not, make sure your region is set to the one where your lambda was deployed).

S | Services | Search [Option+S] | N. California | AccountUser

Lambda > Functions > AJC_FileNotifier

AJC_FileNotifier

Function overview Info

Diagram **Template**

```

graph TD
    FN[AJC_FileNotifier]
    S3[S3]
    FN --> S3

```

Description: -

Last modified: 16 hours ago

Function ARN: arn:aws:lambda:us-west-1:651377595214:function:AJC_FileNotifier

Function URL: Info -

Code Test Monitor Configuration Aliases Versions

Code source Info Upload from ▾

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy

Click on Add Trigger

Add trigger

Trigger configuration Info



Bucket

Choose or enter the ARN of an S3 bucket that serves as the event source. The bucket must be in the same region as the function.

X
G

Bucket region: us-west-1

Event types

Select the events that you want to have trigger the Lambda function. You can optionally set up a prefix or suffix for an event. However, for each bucket, individual events cannot have multiple configurations with overlapping prefixes or suffixes that could match the same object key.

All object create events X

Prefix - optional

Enter a single optional prefix to limit the notifications to objects with keys that start with matching characters.

Suffix - optional

Enter a single optional suffix to limit the notifications to objects with keys that end with matching characters.

Select S3 as the service to trigger on.

Select your bucket. *Note, this is the bucket where you will store your documents, not the bucket you uploaded your function to. The bucket must be in the same region as your lambda function.* For event types, All object create events should be selected by default. Leave as is. Leave the prefix and suffix blank.

Enable write permissions to CloudWatch (so you can check what is being executed).
(This is a single command, line-break due to formatting.)

```
aws iam attach-role-policy --role-name $LAMBDA_ROLE --policy-arn arn:aws:iam::aws:policy/service-role/AWSLambdaBasicExecutionRole
```

Test the Lambda function execution

```
aws lambda invoke --function-name $LAMBDA_FUNCTION_NAME response.json
```

If the function is working properly, you should get the response

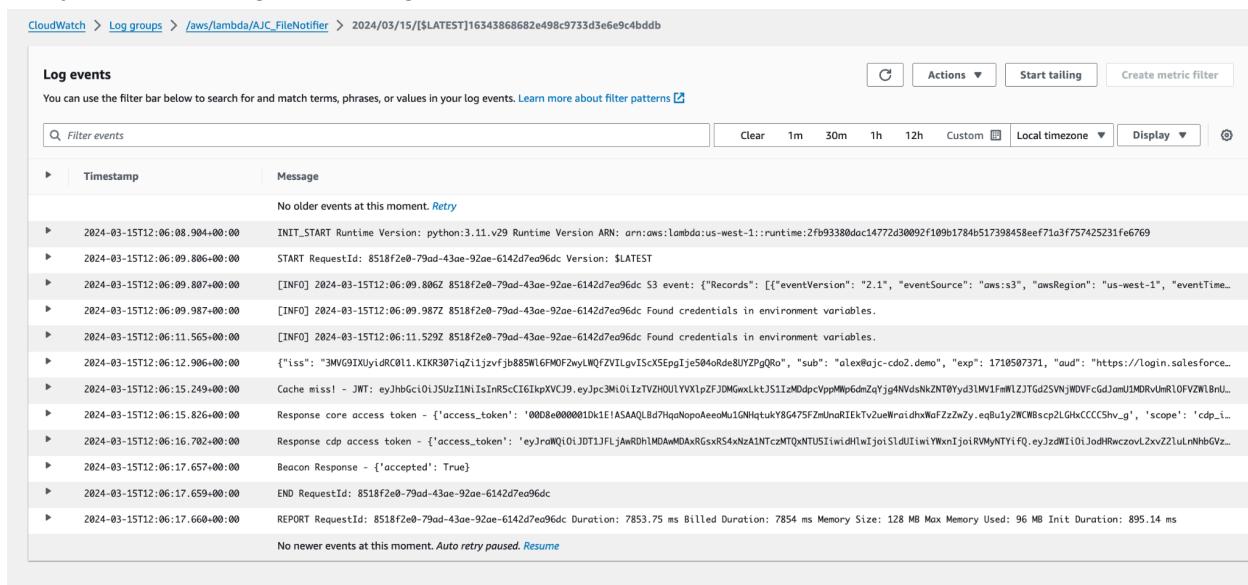
```
{  
  "StatusCode": 200,  
  "ExecutedVersion": "$LATEST"  
}
```

Upload Your Documents to S3

Add documents to your S3 bucket / folder.

Note: You can only upload one type of document to your folder, and it must match the file name pattern you configured when setting up the UDLO.

You can check if the Lambda is running by going to CloudWatch in the console (making sure that you have the right AWS region selected).



The screenshot shows the AWS CloudWatch Log Events interface. The top navigation bar includes 'CloudWatch' > 'Log groups' > '/aws/lambda/AJC_FileNotifier' > '2024/03/15/[\$LATEST]16343868682e498c9733d3e69c4bddd'. Below the navigation is a search bar labeled 'Filter events' and a toolbar with 'Actions', 'Start tailing', 'Create metric filter', and time range buttons for 'Clear', '1m', '30m', '1h', '12h', 'Custom', 'Local timezone', 'Display', and a refresh icon.

The main area is titled 'Log events' and contains a table with columns 'Timestamp' and 'Message'. The table shows the following log entries:

Timestamp	Message
2024-03-15T12:06:08.904+00:00	INIT_START Runtime Version: python:3.11.v29 Runtime Version ARN: arn:aws:lambda:us-west-1:runtime:2fb93380doc14772d30092f109b1784b517398458eef71a3f75742523fe6769
2024-03-15T12:06:09.806+00:00	START RequestId: 8518f2e0-79ad-43ae-92ae-6142d7ea96dc Version: \$LATEST
2024-03-15T12:06:09.807+00:00	[INFO] 2024-03-15T12:06:09.806Z 8518f2e0-79ad-43ae-92ae-6142d7ea96dc S3 event: {"Records": [{"eventVersion": "2.1", "eventSource": "aws:s3", "awsRegion": "us-west-1", "eventTime..."}]}
2024-03-15T12:06:09.987+00:00	[INFO] 2024-03-15T12:06:09.987Z 8518f2e0-79ad-43ae-92ae-6142d7ea96dc Found credentials in environment variables.
2024-03-15T12:06:11.565+00:00	[INFO] 2024-03-15T12:06:11.529Z 8518f2e0-79ad-43ae-92ae-6142d7ea96dc Found credentials in environment variables.
2024-03-15T12:06:11.906+00:00	{"iss": "3MVC9IXUyidRC011.KIR307i1Ij1zv1jb859WlFGFM0Fawy1WQfZV1lgv15c5Epgle5040Rde8UYZPgQRo", "sub": "alex@ajc-cdo2.demo", "exp": 1710507371, "aud": "https://login.salesforce.com", "iat": 1710507371, "nbf": 1710507371, "scope": "openid profile email"}
2024-03-15T12:06:15.249+00:00	Cache miss! - JWT: eyJhbGciOiJSUzIiNiIsInR5cCI6IkpXVCJ9.eyJpc3Mi0IzTVZHOUlYX1pZFJDGMgxLktJS1IzMDpcVppMWpGdmZaYjg4NVNdsNkZNT0Yyd3lMV1fmlZJTGd2SVNjWDVFcGdJAmUjMDRvUmRlOFVZW1BnU...
2024-03-15T12:06:15.826+00:00	Response core access token - {'access_token': '0008e0000010k1E1ASAAQLBd7hqNopoAeoMu1GNHqtkY8G475FZmIhnRIEk1Tv2ueWraidxwFZzZwZy.eqBu1y2WCWBscp2lGHxCCCCShv_g', 'scope': 'cdp_i...'}
2024-03-15T12:06:16.702+00:00	Response cdp access token - {'access_token': 'eyJrdWQl0iJDT1JFLjAnRDh1MDAwMDAxRGsxaRS4xNzA1NTczMTQxNTU5IiwidhIwTjoislduIiwiYXnTjoirVMvNTYifQ.eyJzdWIiOiJodHRwczoL2xVZ2luNhbGVz...'}
2024-03-15T12:06:17.657+00:00	Beacon Response - {'accepted': True}
2024-03-15T12:06:17.659+00:00	END RequestId: 8518f2e0-79ad-43ae-92ae-6142d7ea96dc
2024-03-15T12:06:17.660+00:00	REPORT RequestId: 8518f2e0-79ad-43ae-92ae-6142d7ea96dc Duration: 7853.75 ms Billed Duration: 7854 ms Memory Size: 128 MB Max Memory Used: 96 MB Init Duration: 895.14 ms

At the bottom, a note states 'No newer events at this moment. Auto retry paused. Resume'.

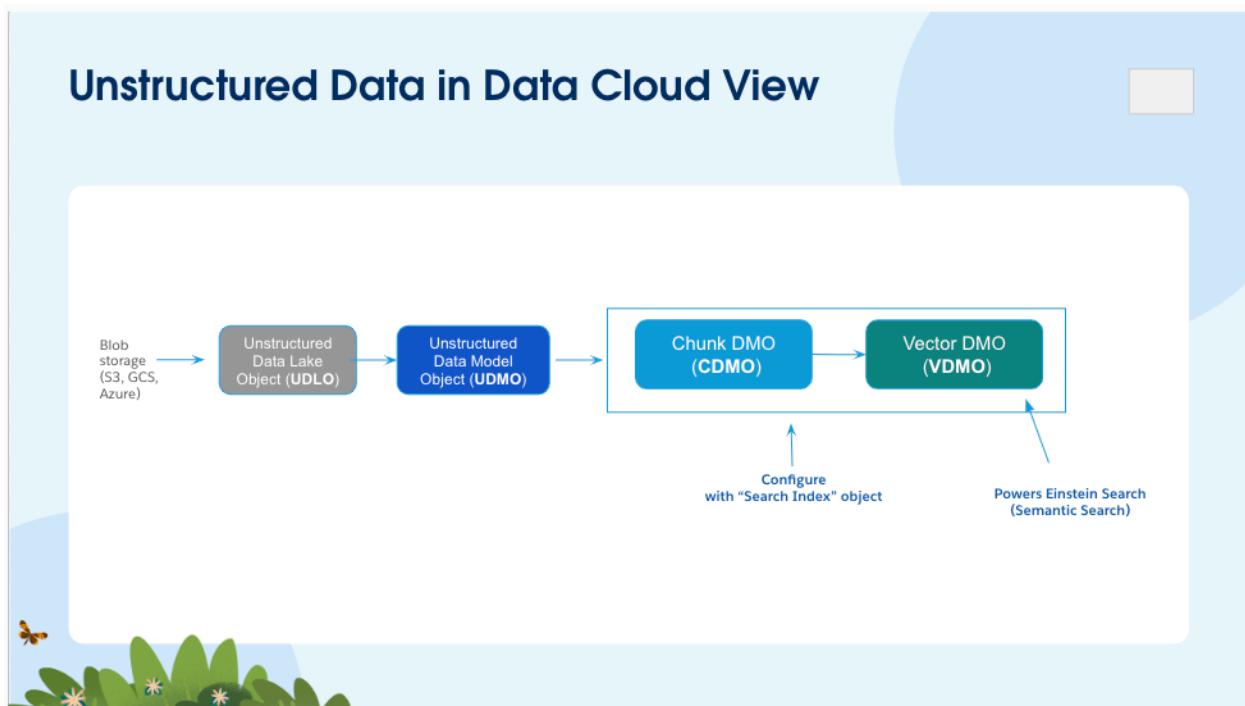
Implementing Vector Search for Unstructured PDFs through S3 & Prompt Builder

Sumit Paliwal
7 Mars 2024

Please note - I will be posting more usecases @ **#sumits-ai-data-explorations** channel.

Objective:

Currently in the pilot stage, the Data Cloud Vector Search's purpose, as outlined in this document, is to illustrate the procedure for uploading unstructured PDF documents into Data Cloud, followed by their vectorization using the Search Index. Once vectorized, these can be accessed through Apex IA & Prompt Builder. This enables Technical Architects to effectively employ this asset in demonstrations to clients.



Steps to Get This to Work

- Make sure unstructured data & search index features are enabled in your org.
- Make sure to setup your AWS S3 or GCP account. Follow this [link](#) to enable to enable your account.
- Create a new or leverage existing S3/GCP bucket.
- Setup a Connected App and OAuth Settings for Data Cloud. Follow this [link](#).
- Create a new UDLO from Unstructured files - **PDF files can not be ingested through Data Streams.**
- Day1 (incremental load) - Now you have two options to ingest unstructured files (pdf) in Data Cloud.
 - Option 1 :- Leverage **Unstructured** event API to notify DC about new files in S3/GCS/Azure Bucket (I followed this approach because I love API's YAY!!). Follow this if you just want to do a quick ingestion.
 - S3 Event API Details [s3ValidEvent.json](#)
 - Azure Event API Details [azureValidEvent.json](#)
 - GCS Event API Details [gcsValidEvent.json](#)
 - Option 2 :- Create a file notification pipeline for AWS S3/GCS/Azure. This is more preferred option as this will automate & leverage lambda function to invoke ingestion on DC as soon incremental file is loaded. Follow this [link](#).
- Once the files ingested in UDLO then map it in custom UDMO.(This can be done with auto mapping)
- Create the search index via easy or advance set-up.
- Run your vector search query and validate the chunking and score on a search "String".
- Create Apex IA and invoke SQL on vector search on a given request.
- Define Prompt template and call this Apex IA.

Detailed Steps :

Make sure unstructured data & search index features are enabled in your org.

Search Index
Recently Viewed

Configuration Name	Data ...	API Name	Source Data Model Obj...	Vector Data Model Object	Chunk Data Model Object	Last Modified D...	Search Index Re...	Search I...
1 MoodyMSADMO	default	MoodyMSADMO	MoodyMSADMO__dim	MoodyMSADMO_vector__dim	MoodyMSADMO_chunk__dim	3/8/2024, 2:26 PM	3/8/2024, 2:55 PM	Success
2 Moodycontractdmo	default	Moodycontractdmo	Moodycontractdmo__dim	Moodycontractdmo_vector__dim	Moodycontractdmo_chunk__dim	3/8/2024, 2:52 PM	In Progress	
3 MoodyMSADMO	default	MoodyMSADMO	MoodyMSADMO__dim	MoodyMSADMO_vector__dim	MoodyMSADMO_chunk__dim	3/5/2024, 4:44 PM	3/7/2024, 2:24 PM	Success

Create a new or leverage existing S3/GCP bucket

Amazon S3

Amazon S3

Account snapshot

Total storage	Object count	Average object size	Metrics
321.8 KB	2	160.9 KB	You can enable advanced metrics in the "default-account-dashboard" configuration.

General purpose buckets (1) [Info](#)

Name	AWS Region	Access	Creation date
sumitdcbucket	US East (N. Virginia) us-east-1	Bucket and objects not public	March 5, 2024, 02:55:31 (UTC+05:30)

Create a new UDLQ from Unstructured files - PDF files can not be ingested through Data Streams

sumitpaliw-230131-702-demo.lightning.force.com/lightning/o/DataLakeObjectInstance/list?filterName=Recent

Salesforce Data Cloud Home Data Streams Data Lake Objects Data Model Data Explorer Identity Resolutions Profile Explorer Calculated Insights Data Action Targets Search Index More

Data Cloud Recently Viewed

13 items - Updated 5 minutes ago

New Data Lake Object

Choose a method of creating your data lake object.

From External Files: Create an unstructured data lake object from a source of unstructured data.

From Existing: Create a data lake object from an existing object.

New: Create a data lake object from scratch.

Cancel Next

Name	Category	Storage	Data Lake Object Status	Last Updated On	Total Records
Contact_00DDn00000CWZgn	Profile	Local	Active	3/2/2024, 12:01 AM	114
Lead_00DDn00000CWZgn	Profile	Local	Active	3/2/2024, 12:05 AM	294
Account_00DDn00000CWZgn	Profile	Local	Active	3/9/2024, 12:31 PM	69

sumitpaliw-230131-702-demo.lightning.force.com/lightning/o/DataLakeObjectInstance/list?filterName=Recent

Salesforce Data Cloud Home Data Streams Data Lake Objects Data Model Data Explorer Identity Resolutions Profile Explorer Calculated Insights Data Action Targets Search Index More

Data Cloud Recently Viewed

13 items - Updated 6 minutes ago

New Data Lake Object

Where is your unstructured data stored

Amazon S3: Retrieve a file from Amazon Simple Storage Service

Google Cloud Storage: Retrieve a file from Google Cloud Storage

Microsoft Azure Storage: Import objects from Microsoft Azure Storage

Back Next

Name	Category	Storage	Data Lake Object Status	Last Updated On	Total Records
Contact_00DDn00000CWZgn	Profile	Local	Active	3/2/2024, 12:01 AM	114
Lead_00DDn00000CWZgn	Profile	Local	Active	3/2/2024, 12:05 AM	294
Account_00DDn00000CWZgn	Profile	Local	Active	3/9/2024, 12:31 PM	69

New Data Lake Object

Reference files from folder in your store

Connection Details

* Select Connection
AWS S3 Connection Vector

Connection Details

You can point to folders and subfolders or a whole directory. All folders under the directory will be included

* Directory /

* File Name Pattern *.pdf

Data Lake Object Setup Guide

Selecting Unstructured Data

When you add files from your external store, Data Cloud extracts three pieces of metadata:

- Content Type
- Total Records
- Last Updated on

Data Cloud stores this metadata in the unstructured data lake objects. Each row within the objects correspond to a single file on the external blob store.



New Data Lake Object

All matching files from all configured folders will be created into a single Unstructured Data Lake Object.

* Object Label YourObjectName_DLO

* Object API Name YourObjectName_DLO

Unstructured Data Model Object Mapping

Select a Data Space and choose to map your unstructured data lake object to a new or an existing unstructured data model object. Both objects must be referenced in the same Data Space. You can optionally create a Semantic Search Index Configuration for the unstructured data model object. If you choose not to create one now, you can do so later.

New Existing

* Data Space default

* Object Name YourObjectName_DMO

* Object API Name YourObjectName_DMO

Search Indexation

Enable semantic search with system defaults

Selecting Unstructured Data

When an unstructured data connector is created, monitor for changes from the moment it is created and sync them automatically.

Semantic Search Index Configurations

When you create Semantic Search Index Configuration for unstructured data model objects, you break the referenced data into manageable, semantically meaningful chunks that can be embedded as searchable vectors—numerical representations of unstructured data—in your LLMs and prompts.



Day1 (incremental load) - Option 1 :- Leverage Unstructured REST API to notify DC about new files in S3/GCP Bucket

The screenshot shows the Postman application interface. On the left, the sidebar lists environments, collections, and histories. The main workspace shows a POST request to <https://mfsdqnlmrdrd0nztgrnd8zddg1.c360a.salesforce.com/api/v1/unstructuredIngest?sourceType=aws>. The request body is a JSON object with various fields like eventVersion, eventSource, awsRegion, etc. The response status is 202 Accepted, time 199 ms, size 723 B. The response body shows a single key "accepted" with the value true.

AWS S3 API Contract Details -

Http Method - POST

EndPoint - [https:// Data Cloud Org>/api/v1/unstructuredIngest?sourceType=aws](https://Data Cloud Org>/api/v1/unstructuredIngest?sourceType=aws)

Body -

```
{
  "Records": [
    {
      "eventVersion": "2.0",
      "eventSource": "aws:s3",
      "awsRegion": "us-west-1",
      "eventTime": "1970-01-01T00:00:00.000Z",
      "eventName": "ObjectCreated:Put",
      "userIdentity": {
        "principalId": "EXAMPLE"
      },
      "requestParameters": {
        "sourceIPAddress": "127.0.0.1"
      },
      "responseElements": {
        "x-amz-request-id": "EXAMPLE123456789",
        "x-amz-id-2": "EXAMPLE123/5678abcdeghijklmnopqrstuvwxyzABCDEFH"
      }
    }
  ]
}
```

```

    "x-amz-id-2": "EXAMPLE123/5678abcdefghijklmbdaisawesome/mnopqrstuvwxyzABCDEFGH"
},
"s3": {
  "s3SchemaVersion": "1.0",
  "configurationId": "testConfigRule",
  "bucket": {
    "name": "sumitdcbucket",
    "ownerIdentity": {
      "principalId": "EXAMPLE"
    },
    "arn": "arn:aws:s3:::sumitdcbucket"
  },
  "object": {
    "key": "sumitdcbucket/Master_Subscription_Agreement1.pdf",
    "size": 3670016,
    "eTag": "976d2e4df7793d72b7de5efeb7604979",
    "sequencer": "1dIDn00000007qLIAQ"
  }
}
}
]
}
}

```

Make sure to replace ““Arn“,“Key”,“eTag” with your S3 bucket file values -

The screenshot shows the AWS S3 console interface. On the left, there's a sidebar with various AWS services like Buckets, Access Grants, and Storage Lens. The main area shows the details of a file named 'Master_Subscription_Agreement1.pdf' located in the 'sumitdcbucket' bucket. The 'Properties' tab is active. Key details shown include:

- Owner:** it-infra-aws-governance+NTY3NTkwMDkyMjlmOTRjMW
- AWS Region:** US East (N. Virginia) us-east-1
- Last modified:** March 9, 2024, 04:19:16 (UTC+05:30)
- Size:** 321.7 KB
- Type:** pdf
- Key:** sumitdcbucket/Master_Subscription_Agreement1.pdf
- Entity tag (Etag):** 976d2e4df7793d72b7de5efeb7604979

Red circles highlight the 'Key' field and the 'Entity tag (Etag)' field.

GCS API Event Details :

```
[  
{  
  "source":  
  "/subscriptions/{subscription-id}/resourceGroups/Storage/providers/Microsoft.Storage/storageAc  
counts/my-storage-account",  
  "subject": "mybucket/folder/HappyFace1.png",  
  "type": "google.cloud.storage.object.v1.finalized",  
  "time": "2020-04-23T07:38:57.230Z",  
  "id": "831e1650-001e-001b-66ab-eeb76e069631",  
  "data": {  
    "bucket": "mybucket",  
    "contentType": "image/png",  
    "crc32c": "rTVTeQ==",  
    "etag": "CNHZkbuF/ugCEAE=",  
    "generation": "1587627537231057",  
    "id": "mybucket/folder/HappyFace1.png/1587627537231057",  
    "kind": "storage#object",  
    "md5Hash": "kF8MuJ5+CTJxvyhHS1xzRg==",  
    "mediaLink":  
    "https://www.googleapis.com/download/storage/v1/b/mybucket/o/folder%2FHappyFace1.png?g  
eneration=1587627537231057&alt=media",  
    "metageneration": "1",  
    "name": "folder/HappyFace1.png",  
    "selfLink": "https://www.googleapis.com/storage/v1/b/mybucket/o/folder/HappyFace1.png",  
    "size": "352",  
    "storageClass": "MULTI_REGIONAL",  
    "timeCreated": "2020-04-23T07:38:57.230Z",  
    "timeStorageClassUpdated": "2020-04-23T07:38:57.230Z",  
    "updated": "2020-04-23T07:38:57.230Z"  
  },  
  "specversion": "1.0"  
},  
{  
  "source":  
  "/subscriptions/{subscription-id}/resourceGroups/Storage/providers/Microsoft.Storage/storageAc  
counts/my-storage-account",  
  "subject": "mybucket/folder/HappyFace2.pdf",  
  "type": "google.cloud.storage.object.v1.deleted",  
  "time": "2020-04-23T07:38:57.230Z",  
  "id": "831e1650-001e-001b-66ab-eeb76e069631",  
  "data": {  
    "bucket": "mybucket",  
  }]
```

```
"contentType": "application/pdf",
"crc32c": "rTVTeQ==",
"etag": "CNHZkbuF/ugCEAE=",
"generation": "1587627537231057",
"id": "mybucket/folder/HappyFace2.pdf/1587627537231057",
"kind": "storage#object",
"md5Hash": "kF8MuJ5+CTJxvyhHS1xzRg==",
"mediaLink":
"https://www.googleapis.com/download/storage/v1/b/mybucket/o/folder%2FHappyFace2.pdf?generation=1587627537231057&alt=media",
"metageneration": "1",
"name": "folder/HappyFace2.pdf",
"selfLink": "https://www.googleapis.com/storage/v1/b/mybucket/o/folder/HappyFace2.pdf",
"size": "423",
"storageClass": "MULTI_REGIONAL",
"timeCreated": "2020-04-23T07:38:57.230Z",
"timeStorageClassUpdated": "2020-04-23T07:38:57.230Z",
"updated": "2020-04-23T07:38:57.230Z"
},
"specversion": "1.0"
},
{
"source":
"/subscriptions/{subscription-id}/resourceGroups/Storage/providers/Microsoft.Storage/storageAccounts/my-storage-account",
"subject": "mybucket/folder/HappyFace3.txt",
"type": "google.cloud.storage.object.v1.deleted",
"time": "2020-04-23T07:38:57.230Z",
"id": "831e1650-001e-001b-66ab-eeb76e069631",
"data": {
"bucket": "mybucket",
"contentType": "text/plain",
"crc32c": "rTVTeQ==",
"etag": "CNHZkbuF/ugCEAE=",
"generation": "1587627537231057",
"id": "mybucket/folder/HappyFace3.txt/1587627537231057",
"kind": "storage#object",
"md5Hash": "kF8MuJ5+CTJxvyhHS1xzRg==",
"mediaLink":
"https://www.googleapis.com/download/storage/v1/b/mybucket/o/folder%2FHappyFace3.txt?generation=1587627537231057&alt=media",
"metageneration": "1",
"name": "folder/HappyFace3.png",
"selfLink": "https://www.googleapis.com/storage/v1/b/mybucket/o/folder/HappyFace3.txt",
```

```
"size": "423",
"storageClass": "MULTI_REGIONAL",
"timeCreated": "2020-04-23T07:38:57.230Z",
"timeStorageClassUpdated": "2020-04-23T07:38:57.230Z",
"updated": "2020-04-23T07:38:57.230Z"
},
"specversion": "1.0"
},
{
"source":
"/subscriptions/{subscription-id}/resourceGroups/Storage/providers/Microsoft.Storage/storageAc
counts/my-storage-account",
"subject": "mybucket/folder/HappyFace4.png",
"type": "google.cloud.storage.object.v1.metadataUpdated",
"time": "2020-04-23T07:38:57.230Z",
"id": "831e1650-001e-001b-66ab-eeb76e069631",
"data": {
"bucket": "mybucket",
"contentType": "image/png",
"crc32c": "rTVTeQ==",
"etag": "CNHZkbuF/ugCEAE=",
"generation": "1587627537231057",
"id": "mybucket/folder/HappyFace4.png/1587627537231057",
"kind": "storage#object",
"md5Hash": "kF8MuJ5+CTJxvyhHS1xzRg==",
"mediaLink":
"https://www.googleapis.com/download/storage/v1/b/mybucket/o/folder%2FHappyFace4.png?g
eneration=1587627537231057\u0026alt=media",
"metageneration": "1",
"name": "folder/HappyFace4.png",
"selfLink": "https://www.googleapis.com/storage/v1/b/mybucket/o/folder/HappyFace4.png",
"size": "423",
"storageClass": "MULTI_REGIONAL",
"timeCreated": "2020-04-23T07:38:57.230Z",
"timeStorageClassUpdated": "2020-04-23T07:38:57.230Z",
"updated": "2020-04-23T07:38:57.230Z"
},
"specversion": "1.0"
},
{
"source":
"/subscriptions/{subscription-id}/resourceGroups/Storage/providers/Microsoft.Storage/storageAc
counts/my-storage-account",
"subject": "mybucket/folder/HappyFace5.png",
```

```

    "type": "google.cloud.storage.object.v1.finalized",
    "time": "2020-04-23T07:38:57.230Z",
    "id": "831e1650-001e-001b-66ab-eeb76e069631",
    "specversion": "1.0"
}
]

```

Azure API Event Details :

```

[
{
  "source":
    "/subscriptions/{subscription-id}/resourceGroups/Storage/providers/Microsoft.Storage/storageAc
counts/my-storage-account",
    "subject": "/blobServices/default/containers/test-container/blobs/HappyFace1.png",
    "type": "Microsoft.Storage.BlobCreated",
    "time": "2017-06-26T18:41:00.9584103Z",
    "id": "831e1650-001e-001b-66ab-eeb76e069631",
    "data": {
      "api": "PutBlockList",
      "clientRequestId": "6d79dbfb-0e37-4fc4-981f-442c9ca65760",
      "requestId": "831e1650-001e-001b-66ab-eeb76e000000",
      "eTag": "0x8D4BCC2E4835CD0",
      "contentType": "image/png",
      "contentLength": 524288,
      "blobType": "BlockBlob",
      "url": "https://my-storage-account.blob.core.windows.net/testcontainer/HappyFace1.png",
      "sequencer": "00000000000044200000000000028963",
      "storageDiagnostics": {
        "batchId": "b68529f3-68cd-4744-baa4-3c0498ec19f0"
      }
    },
    "specversion": "1.0"
},
{
  "source":
    "/subscriptions/{subscription-id}/resourceGroups/Storage/providers/Microsoft.Storage/storageAc
counts/my-storage-account",
    "subject": "/blobServices/default/containers/testcontainer/blobs/HappyFace2.pdf",
    "type": "Microsoft.Storage.BlobDeleted",
    "time": "2017-11-07T20:09:22.5674003Z",
    "id": "4c2359fe-001e-00ba-0e04-58586806d298",
    "data": {
      "api": "DeleteBlob",
    }
}
]
```

```
"requestId": "4c2359fe-001e-00ba-0e04-585868000000",
"contentType": "application/pdf",
"blobType": "BlockBlob",
"url": "https://my-storage-account.blob.core.windows.net/testcontainer/HappyFace2.pdf",
"sequencer": "00000000000002810000000000002F5CA",
"storageDiagnostics": {
    "batchId": "b68529f3-68cd-4744-baa4-3c0498ec19f0"
},
},
{
    "specversion": "1.0"
},
{
    "source": "/subscriptions/{subscription-id}/resourceGroups/Storage/providers/Microsoft.Storage/storageAc
counts/my-storage-account",
    "subject": "/blobServices/default/containers/testcontainer/blobs/HappyFace3.txt",
    "type": "Microsoft.Storage.BlobDeleted",
    "time": "2017-11-07T20:09:22.5674003Z",
    "id": "4c2359fe-001e-00ba-0e04-58586806d298",
    "data": {
        "api": "DeleteBlob",
        "requestId": "4c2359fe-001e-00ba-0e04-585868000000",
        "contentType": "text/plain",
        "blobType": "BlockBlob",
        "url": "https://my-storage-account.blob.core.windows.net/testcontainer/HappyFace3.txt",
        "sequencer": "00000000000002810000000000002F5CA",
        "storageDiagnostics": {
            "batchId": "b68529f3-68cd-4744-baa4-3c0498ec19f0"
        }
    },
    "specversion": "1.0"
},
{
    "source": "/subscriptions/{subscription-id}/resourceGroups/Storage/providers/Microsoft.Storage/storageAc
counts/my-storage-account",
    "subject": "/blobServices/default/containers/testcontainer/blobs/HappyFace4.jpg",
    "type": "Microsoft.Storage.BlobTierChanged",
    "time": "2021-05-04T15:00:00.8350154Z",
    "id": "0fdefc06-b01e-0034-39f6-4016610696f6",
    "data": {
        "api": "SetBlobTier",
        "clientRequestId": "68be434c-1a0d-432f-9cd7-1db90bff83d7",
        "requestId": "0fdefc06-b01e-0034-39f6-401661000000",
        "blobType": "BlockBlob"
    }
}
```

Create the Search Index via easy or advance set-up.

sumitpalw-230131-702-demo.lightning.force.com/lightning/r/DataSemanticSearch/f18Dn000000000GIAQ/view

Salesforce Sumit-Personal Equity Member Ho... GPT

Data Cloud Home Data Streams Data Lake Objects Data Model Data Explorer Identity Resolutions Profile Explorer Calculated Insights Data Action Targets Search Index More

Search Index Moodycontractdmo Delete Refresh Now

Data Space API Name Source Data Model Object Vector Data Model Object Chunk Data Model Object Last Modified Date

default Moodycontractdmo Moodycontractdmo__dim Moodycontractdmo__vector__dim Moodycontractdmo__chunk__dim 3/8/2024, 2:52 PM

Details Configuration

Chunking

File Extension	Chunking Strategy	Chunking Config
html	Passage Extraction	Strip HTML: true, Max Token: 512
txt	Passage Extraction	Strip HTML: true, Max Token: 512
aspx	Passage Extraction	Strip HTML: true, Max Token: 512
pdf	Passage Extraction	Strip HTML: true, Max Token: 512
log	Passage Extraction	Strip HTML: true, Max Token: 512

Vectorization

Embedding model	Dimension
E5 Large V2 Embedding Model	1024
Max Token Limit	Index
512	HNSW
hnswEfConstruction	M
8	4
Similarity Metric	

Post Question Poll

Share an update... Share

Q. Search this feed...

Collaborate here!

Here's where you start talking with your colleagues about this record.

Run your vector search query and validate the chunking and score on a search “String”:

The screenshot shows the Salesforce Query Workspace Vector Query interface. On the left, there's a sidebar with icons for Data Model Objects, Data Lake Objects, Calculated Insights, and Chatter. The main area has tabs for Profile, Engagement, and Related. A search bar at the top says "Search...". Below it, there are buttons for New Query 1 through New Query 6. The New Query 1 tab is active, displaying the following SQL query:

```
1 select v.score__c Score,c.Chunk__c Info FROM vector_search(table(Moodycontractdmo_vector__dlm),'Service Cloud Unlimited Edition', **  
2000) v JOIN Moodycontractdmo_chunk__dlm c ON v.RecordId__c = c.RecordId__c order by Score desc LIMIT 5 |
```

Below the query, a table shows the results:

Score	Info
1 0.8315994739532471	No modification, amendment, or waiver of any provision of this Agreement shall be effective unless in writing and ...
2 0.8084864020347595	"User Guide" means the online user guide for the Services, accessible via login at http://www.salesforce.com , as...
3 0.7963852882385254	Services may be subject to other limitations, such as, for example, limits on disk storage space, on the number of...
4 0.796227359298706	Limitation of Liability. NEITHER PARTY'S LIABILITY WITH RESPECT TO ANY SINGLE INCIDENT ARISING OUT OF ...
5 0.7955498099327087	Unless otherwise specified in the applicable Order Form, (i) Services are purchased as User subscriptions and m...

Create Apex IA and invoke SQL on vector search on a given request

```
public class GetMyAnswerSP {  
  
    @InvocableMethod(CapabilityType = 'PromptTemplateType://einstein_gpt__fieldCompletion'  
category = 'Prompt')  
    public static List<Response> searchforContext(List<Request> requests) {  
        List<Response> responses = new List<Response>();  
        Response response = new Response();  
  
        String topic = requests[0].RelatedEntity.Question__c;  
  
        ConnectApi.CdpQueryInput input = new ConnectApi.CdpQueryInput();  
        input.sql = 'select v.score__c Score, c.Chunk__c Info FROM  
vector_search(table(Moodycontractdmo_vector__dlm), \'' + String.escapeSingleQuotes(topic) +  
\', \'\', 2000) v JOIN Moodycontractdmo_chunk__dlm c ON v.RecordId__c = c.RecordId__c  
LIMIT 15';  
  
        ConnectApi.CdpQueryOutput output = ConnectApi.CdpQuery.queryANSISql(input);  
        List<Object> data = output.data;
```

```

String scs = "";
for (Object searchRecord : data) {
    Map<String, Object> myMap = (Map<String, Object>)
        JSON.deserializeUntyped(JSON.serialize(searchRecord));
    // check for access of case record for the current user
    if (GetMyAnswerSP.getUserRecordAccess((String) myMap.get('Id__c'))) {
        Map<String, String> sc = new Map<String, String>();
        //sc.put('Link', 'https://rag-tdx-brk-demo.lightning.force.com/lightning/r/Article__c/' +
        String.valueOf(myMap.get('Link')) + '/view' );
        sc.put('Info', String.valueOf(myMap.get('Info')));
        scs = scs + JSON.serialize(sc);
    }
}
response.Prompt = scs;
responses.add(response);
return responses;
}

public static boolean getUserRecordAccess(String recordId) {
    return true;
}

// All inputs will be passed into the IA
public class Request {
    @InvocableVariable
    public Question__c RelatedEntity;
}

public class Response {
    @InvocableVariable
    public String Prompt;
}

}

```

Define Prompt template and call this Apex IA

sumitpaliw-230131-702-demo.lightning.force.com/lightning/setup/EinsteinPromptStudio/0hfDn000000GmaYAS/edit?c__versionId=3vNDn00000GmasMAC

Salesforce Salesforce Sumit-Personal Equity Member Ho... GPT

Cloud Search Setup

Setup Home Object Manager

Prompt Builder Find Contract Details Version 2

Help Settings

Activate Save As Save

Prompt Template Workspace

You will be an expert that answers a general knowledge question.

Here is the question I want you to answer and find the discount percentage number:

Input:Question_c.Question_c

Here is the information you can use to answer the question and generate the response in below format:

Discount Percentage: Apex: GetMyAnswerSP

Configuration

Template Properties

Model Type Standard

Model Standard OpenAI GPT 3.5 Turbo ...

View this model

Preview

Resolution

upon the end of the then-current Order Term the prices set forth in this Pricing Addendum.Table 1: Discounted Products Product Monthly Fee per User in USD% Discount from ListPriceSales Cloud - Unlimited Edition \$82.00 67%Sales Cloud - Unlimited EditionService Cloud Unlimited Edition +Knowledge Pack \$111.58 64%Unlimited Edition - Knowledge Only \$17.80 64%Unlimited Edition - EmployeeCommunity Only\$7.83 61%Customer Community Plus 1,000 logins/month\$1400 40%Customer Community Plus 2,000 logins/month\$1600 46%Service Cloud Unlimited Edition \$92.84% {"Info": "Each party hereby waives any right to jury trial in connection with any Screenshot in any way arising out of or related to this Agreement.Back

Related Record

Resolution Enabled First Question

Preview

Response

The discount percentage for the Service Cloud Unlimited Edition product is 64%.