

Extracting and loading data from S3 to sql using ADF

- Go to IAM access and Now go to Users and select Create User

User name	Path	Group	Last activity	MFA	Password age	Console last sign-in	Access key ID	Active key age
admin	/	0	18 days ago	-	-	-	Active - AKIA4MTWNR...	19 days

- Now enter the details

User details

User name
s3-bucket-user

The user name can have up to 64 characters. Valid characters: A-Z, a-z, 0-9, and + = , . @ _ - (hyphen)

Provide user access to the AWS Management Console - optional
If you're providing console access to a person, it's a best practice [to manage their access in IAM Identity Center](#).

If you are creating programmatic access through access keys or service-specific credentials for AWS CodeCommit or Amazon Keyspaces, you can generate them after you create this IAM user. [Learn more](#)

Cancel Next

- Select next

Set permissions

Add user to an existing group or create a new one. Using groups is a best-practice way to manage user's permissions by job functions. [Learn more](#)

Permissions options

Add user to group
Add user to an existing group, or create a new group. We recommend using groups to manage user permissions by job function.

Copy permissions
Copy all group memberships, attached managed policies, and inline policies from an existing user.

Attach policies directly
Attach a managed policy directly to a user. As a best practice, we recommend attaching policies to a group instead. Then, add the user to the appropriate group.

Cancel Next

- Now select the s3 read only access

Filter by Type		
<input type="text" value="s3"/>	<input type="button" value="X"/>	All types
	Policy name	Type
<input type="checkbox"/>	<input checked="" type="checkbox"/> AmazonDMSRedshiftS3Role	AWS managed
<input type="checkbox"/>	<input checked="" type="checkbox"/> AmazonS3fullAccess	AWS managed
<input type="checkbox"/>	<input checked="" type="checkbox"/> AmazonS3ObjectLambdaExecutionRolePolicy	AWS managed
<input type="checkbox"/>	<input checked="" type="checkbox"/> AmazonS3OutpostsFullAccess	AWS managed
<input type="checkbox"/>	<input checked="" type="checkbox"/> AmazonS3OutpostsReadOnlyAccess	AWS managed
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> AmazonS3ReadOnlyAccess	AWS managed
<input type="checkbox"/>	<input checked="" type="checkbox"/> AWSBackupServiceRolePolicyForS3Backup	AWS managed
<input type="checkbox"/>	<input checked="" type="checkbox"/> AWSBackupServiceRolePolicyForS3Restore	AWS managed
<input type="checkbox"/>	<input checked="" type="checkbox"/> AWSS3OnOutpostsServiceRolePolicy	AWS managed
<input type="checkbox"/>	<input checked="" type="checkbox"/> IVSRecordToS3	AWS managed
<input type="checkbox"/>	<input checked="" type="checkbox"/> QuickSightAccessForS3StorageManagementAnalyticsRe...	AWS managed
<input type="checkbox"/>	<input checked="" type="checkbox"/> S3StorageLensServiceRolePolicy	AWS managed

➤ Select next and select create user

➤ Now select on the user created

IAM > Users > s3-bucket-user

s3-bucket-user [Info](#) [Delete](#)

Summary		
ARN arn:aws:iam::851725617514:user/s3-bucket-user	Console access Disabled	Access key 1 Create access key
Created August 11, 2024, 15:25 (UTC-05:00)	Last console sign-in -	
Permissions Groups Tags Security credentials Access Advisor		

➤ Now select the create access key

➤ Select this option

Application running outside AWS
You plan to use this access key to authenticate workloads running in your data center or other infrastructure outside of AWS that needs to access your AWS resources.

Other
Your use case is not listed here.

➤ Enter the description

Set description tag - *optional* [Info](#)

The description for this access key will be attached to this user as a tag and shown alongside the access key.

Description tag value
Describe the purpose of this access key and where it will be used. A good description will help you rotate this access key confidently later.

Maximum 256 characters. Allowed characters are letters, numbers, spaces representable in UTF-8, and: _ . : / = + ~ @

[Cancel](#) [Previous](#) [Create access key](#)

➤ You will get like this

Retrieve access keys Info

Access key

If you lose or forget your secret access key, you cannot retrieve it. Instead, create a new access key and make the old key inactive.

Access key

Secret access key

AKIA4MTWNRFVFC7RCFIY

***** [Show](#)

Access key best practices

- Never store your access key in plain text, in a code repository, or in code.
- Disable or delete access key when no longer needed.
- Enable least-privilege permissions.
- Rotate access keys regularly.

For more details about managing access keys, see the [best practices for managing AWS access keys](#).

[Download .csv file](#)

[Done](#)

- Now make a connection with ads with this details
➤ Select s3 for the linked service and enter the details from aws

New linked service

 Amazon S3 [Learn more](#) 

Name *

AmazonS31

Description

Connect via integration runtime * 

AutoResolveIntegrationRuntime

Authentication type

Access key

[Access key ID](#)

[Azure Key Vault](#)

Access key ID *

AKIA4MTWNRFVFC7RCFIY

[Secret access key](#)

[Azure Key Vault](#)

Secret access key *

- After entering please check the test connection for testing, make sure the connection works

> Parameters

✓ Connection successful

Create

Back

🔗 Test connection

Cancel

➤ Now create data set for s3

🔍 s3

< All Azure Database File Generic protocol >



Amazon S3



Amazon S3 Compatible



Oracle Cloud Storage (S3 API)

➤ Select csv and continue

Select format

Choose the format type of your data



Avro



Binary



DelimitedText



➤ Please make sure you give the path of s3

Set properties

Name

DelimitedText1

Linked service *

AmazonS31



File path

customer991

/ customer

/ File name



First row as header



Import schema

From connection/store From sample file None

➤ Source is created

Connection Schema Parameters

Linked service *	AmazonS31	Test connection	Edit	New	Learn more
File path *	customer991	/ customer	/ File name	Browse	Preview data
Compression type	Select...	Detect format			
Column delimiter	Comma (,)				

➤ Now creating the target

We need to store it in sql

- Please make sure sql server is created before
- Now we need to create the table according to data in s3 table
- Please create table according to table

Query 1 ×

Run Cancel query Save query Export data as Show

```
1 CREATE TABLE Customers (
2     CustomerID INT,
3     FirstName VARCHAR(50),
4     LastName VARCHAR(50),
5     Email VARCHAR(100),
6     PhoneNumber VARCHAR(20),
7     Address VARCHAR(100),
8     City VARCHAR(50),
9     State VARCHAR(50),
10    ZipCode VARCHAR(20),
11    Country VARCHAR(50)
12 );
```

Results Messages

Query succeeded: Affected rows: 0

- Table is created according to csv data from s3
- Now creating connection for destination, select new link and select azure sql database

New linked service

 Azure SQL Database [Learn more](#) 

Database name *

Authentication type *

User name *

Password Azure Key Vault

Password *

Always encrypted 

Encrypt 

Trust server certificate 

Host name in certificate

- Please test connection

 Connection successful

 Test connection

- Now create dataset for target

New dataset

In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. [Learn more](#)

Select a data store

sql da

All Azure Database File Generic protocol >

Azure SQL Database

Azure SQL Database Managed Instance

- Select continue
Set properties

Name

AzureSqlTable1

Linked service *

AzureSqlDatabase1



Table name

dbo.Customers



Enter manually

Import schema

From connection/store None

Connection Schema Parameters

Linked service *

AzureSqlDatabase1

Test connection

Edit

New

[Learn more](#)

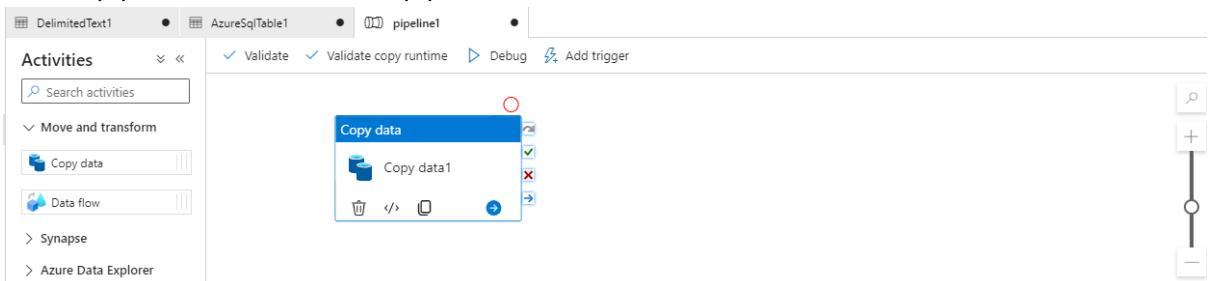
Table

dbo.Customers

Refresh Preview data

Enter manually

- Connections and the datasets are created now we need to create pipeline for flow
-select pipeline and create new pipeline



- Now select source

General Source Sink¹ Mapping Settings User properties

Source dataset *

File path type File path in dataset Prefix Wildcard file path List of files

Wildcard paths customer991 / /

Start time (UTC)
End time (UTC)

Filter by last modified
Recursively

- Please select preview data for review data, now select sink

Link dataset *

Write behavior Insert Upsert Stored procedure

bulk insert table lock Yes No

Table option Use existing Auto create table

re-copy script

Write batch timeout

Write batch size

- Now select the validate



Your pipeline has been validated.

No errors were found.

➤ Select debug

Showing 1 - 1 of 1 items

Activity name ↑↓	Activity status ↑↓	Activity type ↑↓	Run start ↑
Copy data1	✓ Succeeded	Copy data	8/11/2024, 3:

➤ Now go to sql database and select the table

The screenshot shows the Microsoft SQL Server Management Studio (SSMS) interface. On the left, the Object Explorer displays a tree view of the database structure, including tables like 'dbo.Customers' which contains columns such as CustomerID, FirstName, LastName, Email, PhoneNumber, Address, City, State, ZipCode, and Country. In the center, a query editor window is open with the following SQL command:

```
1  SELECT TOP (1000) * FROM [dbo].[Customers]
```

Below the query editor, the results pane shows a grid of data from the 'Customers' table. The data includes 7 rows of customer information:

CustomerID	FirstName	LastName	Email	PhoneNumber	Address	City
1	John	Doe	john.doe@example.com	555-1234	123 Elm St	Springfield
2	Jane	Smith	jane.smith@example.com	555-5678	456 Oak St	Rivertown
3	Emily	Jones	emily.jones@example.com	555-8765	789 Pine St	Laketown
4	Michael	Brown	michael.brown@example.com	555-4321	101 Maple St	Hilltop
5	Alice	Johnson	alice.johnson@example.com	555-6789	202 Cedar St	Brookfield
6	Robert	Davis	robert.davis@example.com	555-3456	303 Birch St	Creekside
7	Linda	Miller	linda.miller@example.com	555-7890	404 Walnut St	Meadowville

➤ We can see that we successfully loaded data into table

- Now creating the lambda function to trigger the ADF to take data when ever file insert into s3
- Select lambda in search and select the create function

Create function Info

Choose one of the following options to create your function.

Author from scratch
Start with a simple Hello World example.

Use a blueprint
Build a Lambda application from sample code and configuration presets for common use cases.

Container image
Select a container image to deploy for your function.

Basic information

Function name
Enter a name that describes the purpose of your function.

triggerforADF

Use only letters, numbers, hyphens, or underscores with no spaces.

Runtime Info
Choose the language to use to write your function. Note that the console code editor supports only Node.js, Python, and Ruby.

Python 3.12

▼

Architecture Info
Choose the instruction set architecture you want for your function code.

x86_64
 arm64

▼ Change default execution role

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

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

ⓘ Role creation might take a few minutes. Please do not delete the role or edit the trust or permissions policies in this role.

Role name
Enter a name for your new role.

s3access

Use only letters, numbers, hyphens, or underscores with no spaces.

Policy templates - *optional* Info
Choose one or more policy templates.

Amazon S3 object read-only permissions

X

- Now create function

triggerforADF

Throttle Copy ARN Actions ▾

Function overview Info

Diagram Template

triggerforADF

Layers (0)

+ Add trigger + Add destination

Description

Last modified now

Function ARN arn:aws:lambda:us-east-1:851725617514:function:triggerforADF

Function URL Info

Code Test Monitor Configuration Aliases Versions

- Now select the add trigger

Add trigger

Trigger configuration Info

Select a source

s3

Batch/bulk data processing



S3

aws asynchronous storage

Cancel

Add

Add trigger

Trigger configuration Info



S3

aws asynchronous storage

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.

s3/customer991

X



Bucket region: us-east-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.

customer/

Suffix - optional

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

.csv

Recursive invocation

If your function writes objects to an S3 bucket, ensure that you are using different S3 buckets for input and output. Writing to the same bucket increases the risk of creating a recursive invocation, which can result in increased Lambda usage and increased costs. [Learn more](#)

- I acknowledge that using the same S3 bucket for both input and output is not recommended and that this configuration can cause recursive invocations, increased Lambda usage, and increased costs.

Lambda will add the necessary permissions for AWS S3 to invoke your Lambda function from this trigger. [Learn more](#) about the Lambda permissions model.

Cancel

Add

➤ Please select add

▼ Function overview [Info](#) [Export to A](#)

Diagram [Template](#)

 triggerforADF

 S3 [+ Add destination](#)

[+ Add trigger](#)

Description

Last modified 17 minutes ago

Function ARN [arn:aws:lambda:](#)

Function URL [Info](#)

➤ Now upload the some file with .csv and now watch in monitor and select view cloudwatch

Test **Monitor** **Configuration** **Aliases**

Info

Log streams (1)

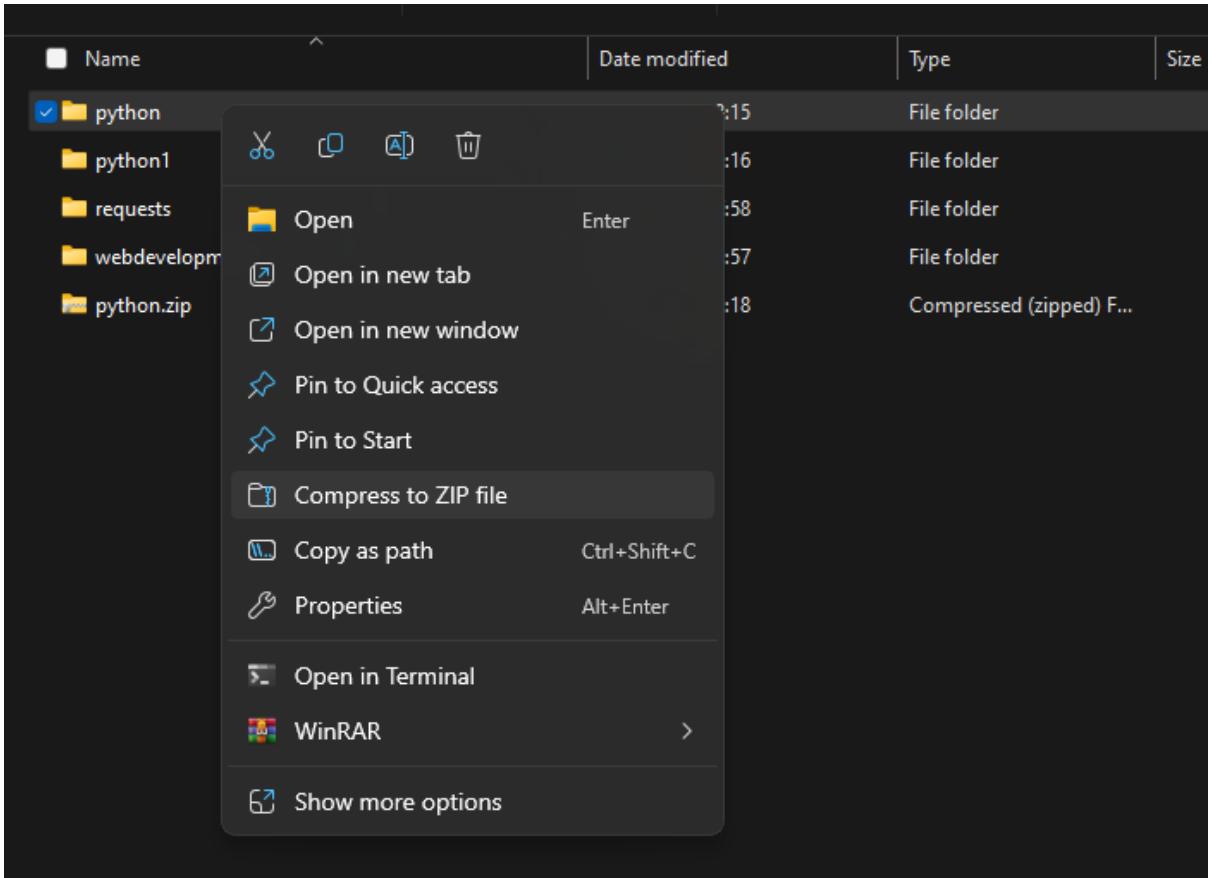
Filter log streams or try prefix search Exact match Show expired [Info](#)

Log stream Last event time 2024-08-11 21:21:18 (UTC)

[Create log stream](#) [Search all log streams](#)

➤ I got a log when file is uploaded, please check the logs for any errors

- Now are writing code to call the data factory using api call
- For this we need to make sure requests is library in “python” we need to add It to lambda function using layers concept . I recommend to name the dir as python only
- If windows open power shell and please create dir name python, please make sure you know the location. Now go to that dir and enter this cmd -→ “pip install requests -t .“
- Now go to dir and check for the files
- Next step is zip the complete folder, right click on the created dir and do this



- Make sure zip file is created
- Now go to aws lambda function page and select lambda on left

▼ Additional resources

[Code signing configurations](#)

[Event source mappings](#)

[Layers](#)

[Replicas](#)

- Now select create layer
- Now upload the zip file created in local machine

Layer configuration

Name
requests

Description - optional
Description

Upload a .zip file
 Upload a file from Amazon S3

 Upload

For files larger than 10 MB, consider uploading using Amazon S3.

➤ Select the next fields make sure you select the python

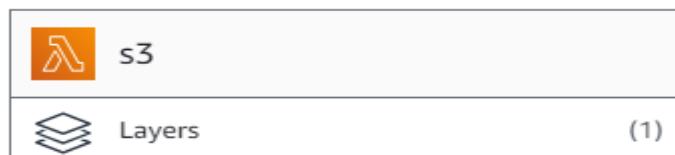
Compatible architectures - optional [Info](#)
Choose the compatible instruction set architectures for your layer.
 x86_64
 arm64

Compatible runtimes - optional [Info](#)
Choose up to 15 runtimes.

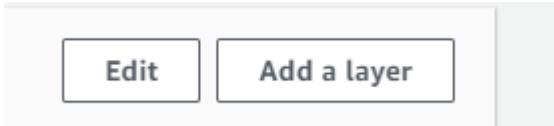
Runtimes ▾ C
Python 3.12 X

License - optional [Info](#)

- And select create
- Now go to functions created before and select the layer option



- Select the add a layer



Function runtime settings

Runtime Python 3.12	Architecture x86_64
------------------------	------------------------

Choose a layer

Layer source [Info](#)
Choose from layers with a compatible runtime and instruction set architecture or specify the Amazon Resource Name (ARN) of a layer version. You can also [create a new layer](#).

AWS layers
Choose a layer from a list of layers provided by AWS.

Custom layers
Choose a layer from a list of layers created by your AWS account or organization.

Specify an ARN
Specify a layer by providing the ARN.

Custom layers
Layers created by your AWS account or organization that are compatible with your function's runtime.
▼

Version
▼

[Cancel](#) [Add](#)

- Select the custom layer and select the layer you created before and select the version also and select the add option
- Here is the code to send the latest file uploaded into s3 and trigger the data factory with the file name to do ETL with this file

```
import json
import urllib3
import os
import requests

def lambda_handler(event, context):
    object_key = event['Records'][0]['s3']['object']['key']
    tenant_id = "808cc83e-a546-47e7-a03f-73alebba24f3"
    client_id = "5bf380e3-f965-4dfb-b344-be467ca93de9"
    client_secret = "1oJ8Q~QNSKc-BEkmh6FCLPUp_idkRriUXhwIKcNR"
    resource_group = "dataengineering"
    adf_factory_name = "data-from-sql-adls"
    adf_pipeline_name = "ProcessFilePipeline"
    adf_url = (
        f"https://management.azure.com/subscriptions/cfb46a3a-9f6d-4b13-"
        f"9c3f-2dab3ca2016d/"
        f"resourceGroups/{resource_group} /"
        f"providers/Microsoft.DataFactory/factories/{adf_factory_name} /"
        f"pipelines/{adf_pipeline_name}/createRun?api-version=2018-06-01"
    )

    data = {"sourceFileName": object_key}

    headers = {
        'Content-Type': 'application/json',
        'Authorization': f'Bearer {os.getenv("ADF_ACCESS_TOKEN")}'
    }

    try:
        # Enable SSL verification (default behavior)
        response = requests.post(adf_url, json=data, headers=headers)

        if response.status_code == 200:
            print(f"Successfully triggered ADF pipeline: {response.text}")
        else:
            print(f"Failed to trigger ADF pipeline: {response.status_code}")
    except Exception as e:
        print(f"Error triggering ADF pipeline: {str(e)}")
```

We can see in this link we gave directly the subscription id

<https://management.azure.com/subscriptions/cfb46a3a-9f6d-4b13-9c3f-2dab3ca2016d/>"

this we can get from → go to home and select subscription and there we can see the column with subscription ID

- Now please copy and paste here, make sure you copy the subscription you working with

➤ Here is the code we need get the details of this to insert into this code

```
tenant_id =  
client_id =  
client_secret =  
resource_group =  
adf_factory_name =  
  
adf_pipeline_name =
```

➤ We need this details to run this code , we need to get all this details from azure portal let start with basic details like pipeline name and factory name

```
resource_group = // the resource group name your data factory is present  
adf_factory_name = // you can find it in data factory page
```

```
adf_pipeline_name = // this name you will find it after creating pipeline
```

adf_factory_name :

The screenshot shows the Azure Data Factory blade. At the top, there's a blue header bar. Below it, the URL is shown as "Home > Data factories". The main title is "Data factories" with a gear icon and three dots. Below the title, it says "K L University (kluniversity.in)". There are navigation buttons for "Create", "Manage view", and "Refresh". A search bar at the top right contains the placeholder "Filter for any field..." and a "Subscription equ" dropdown. Below the search bar, it says "Showing 1 to 1 of 1 records." There is one record listed: "data-from-sql-adls". Each record has a checkbox and a "Name" column with an up-down arrow icon.

Name
<input type="checkbox"/> data-from-sql-adls

Adf_pipeline_name :

The screenshot shows the 'Factory Resources' interface. At the top, there is a search bar labeled 'Filter resources by name' and a '+' button. Below the search bar, there is a section titled 'Pipelines' with two items listed: 'pipeline1' and 'ProcessFilePipeline'. Each pipeline item has a small icon and a '...' button to its right.

- Now we want
tenant_id =
client_id =
client_secret =
- For this search for this

The screenshot shows a search results page with the query 'aad' in the search bar. The results are categorized into 'All', 'Services (5)', and 'Marketplace (6)'. Under the 'Services' category, there is a result for 'Microsoft Entra ID' with the keyword 'AAD'. There is also a 'See more' link.

- Select entra id
- Under manage select app registration

Administrative units

- Delegated admin partners
- Enterprise applications
- Devices
- App registrations

➤ Select new registration

+ New registration

Register an application

* Name
The user-facing display name for this application (this can be changed later).

MyLambdaFunction ✓

Supported account types
Who can use this application or access this API?

- Accounts in this organizational directory only (K L University only - Single tenant)
- Accounts in any organizational directory (Any Microsoft Entra ID tenant - Multitenant)
- Accounts in any organizational directory (Any Microsoft Entra ID tenant - Multitenant) and personal Microsoft accounts (e.g. Skype, Xbox)
- Personal Microsoft accounts only

➤ SELECT this 2 and continue register

➤ After creation you will see the overview page or select the app you created

➤ Now under mage on left side select certificates and secrets

▽ Manage

- Branding & properties
- Authentication
- Certificates & secrets
- Token configuration

➤ Select create new client secret

Add a client secret

Description	lambda-secret
Expires	Recommended: 180 days (6 months)

Home > KL University | App registrations > MyLambdaFunction

MyLambdaFunction | Certificates & secrets

Search Got feedback?

Overview Quickstart Integration assistant Diagnose and solve problems

Credentials enable confidential applications to identify themselves to the authentication service when receiving token scheme). For a higher level of assurance, we recommend using a certificate (instead of a client secret) as a credential.

- You will get like this please make sure you note it down some where it will not show again if you leave that page

+ New client secret

Description	Expires	Value ⓘ	Secret ID
lambda-secret	09/02/2025	YS78Q~jhZrXGgxYP.VFpYiOJ4ZlqKeLECN...	8d942797-d6ea-4a11-9e0f-3727d58540d9

We need the value here copy and paste some where

the above value in the pic is client_secret

now we need :

tenant_id =

client_id =

- Go to overview same page left top

Application (client) ID : c623958e-7d64-412d-bb6e-903dde651d79

Object ID : 957948c5-3611-4d2e-a378-cc894e60c802

Directory (tenant) ID : 808cc83e-a546-47e7-a03f-73a1ebba24f3

- Here are the both client_id and tenant_id

-make sure you note all this in notepad or text for easy access to this

- Role is created now we need to attach to data factory and actions

- For this now go to data factory page in azure portal, not data factory workspace
this is home page of now select IAM access

The screenshot shows the Azure Data Factory V2 Overview page for a resource named "data-from-sql-adls".

Essentials:

- Resource group ([move](#)) [dataengineering](#)
- Status: Succeeded
- Location: Central US
- Subscription ([move](#)) [Azure for Students](#)
- Subscription ID: cfb46a3a-9f6d-4b13-9c3f-2dab.

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Settings

- Networking
- Managed identities

Properties

Now select role access

Add role assignment

Add co-administrator

My access

Select this and continue

Grant access to Azure resources based on job function, such as the ability to create virtual machines.

factory factor 🏭

data facto X Type : All Category :

Name ↑↓	Description ↑↓
Data Factory Contributor	Create and manage data factories, as well as datasets and pipelines.

Showing 1 - 1 of 1 results.

[Review + assign](#) [Previous](#) [Next](#)

- Select members
 - User, group, or service principal
 - Managed identity
- Please enter the name we created at app registration

Select members X

Select ⓘ my lambda

mylambda

 MyLambdaFunction

- Select and continue
- Select review and assign

- Now we need to do one more step in this process
 - Go to home and search for subscription

sub

All Services (15) Marketplace (31)

Services

Subscriptions

- Select the sub where you are working
- Select IAM in this

Azure for Students Subscription

Search Upgrad

To check

Overview Activity log Access control (IAM) Tags

Subscription | cfb46a3a-9f6

Add role assignment

Name	Description
Data Factory Contributor	Create and manage data factories, as well as child resou

- Select new role

Add role assignment

Type : All Category : All

Name ↑↓	Description ↑↓
Data Factory Contributor	Create and manage data factories, as well as child resou

Showing 1 - 1 of 1 results.

- Same as before select --> select members

Data Factory Contributor

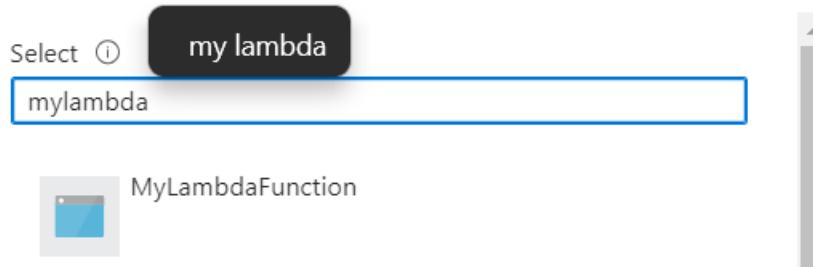
- User, group, or service principal
 Managed identity

[+ Select members](#)

- Now enter the app name (if you don't find the app name then please refer to app management in AAD)

Select members

X



- Select this and continue review and assign
➤ Now we gave access to app to access the subscription and the data factory

- We need to change minor for in the pipeline we are taking file name as dynamic from s3
We need to make sure it changes every time for this we are going to give dynamic value

S3 ----→ lambda ----- sends latest file name ----→ data factory collects and use it

We need to change how this file name is stored in data factory and use it in source file name

- Now go data factory workspace and go to pipeline
- Select pipeline which is created and select white space(beside copy data) then you will get this options

The screenshot shows the Azure Data Factory Pipeline Editor. On the left, the 'Activities' pane is open, listing various activity types like Move and transform, Synapse, Azure Data Explorer, etc. In the main workspace, a 'Copy data' activity named 'Copy data1' is selected. Below the workspace, the 'Parameters' tab is active in the 'Variables' section. It shows a single parameter named 'sourceFileName' of type String with a default value field.

Name	Type	Default value
sourceFileName	String	

- We are send the parameter with name sourceFileName , leave default name
- Make sure same thing in variables section

The screenshot shows the 'Variables' tab in the 'Variables' section. It displays the same 'sourceFileName' variable as the previous screen, with its properties: Name, Type (String), and Default value.

Name	Type	Default value
sourceFileName	String	

- We need to change in data sets also for dynamic data file name
 - Go to pipeline and select source

General **Source** Sink Mapping Settings User properties

File path type File path in dataset Prefix Wildcard file path List of files ⓘ

Wildcard paths customer991 / Wildcard folder path / Wildcard file name

Start time (UTC) End time (UTC)

Filter by last modified ⓘ

Recurisvely

- Now select wildcard file name
 - Now we can see dynamic , select that
- Enter this @pipeline().parameters.sourceFileName

Pipeline expression builder

Add dynamic content below using any combination of [expressions](#), [functions](#) and [system variables](#).

```
@pipeline().parameters.sourceFileName
```

[Clear contents](#)

Parameters System variables Functions Variables

Search

+

- Select continue

Now when ever lambda function calls ADF with file name this will collect and add it dynamically

- Now there is one more step we need to do for get one value in this code

```

> import json
> import urllib3
> import os
> import requests

def lambda_handler(event, context):
    object_key = event['Records'][0]['s3']['object']['key']
    tenant_id = "808cc83e-a546-47e7-a03f-73a1ebba24f3"
    client_id = "5bf380e3-f965-4dfb-b344-be467ca93de9"
    client_secret = "1oJ8Q~QNSKc-BEkmh6FCLPUp_idkRriUXhwIKeNR"
    resource_group = "dataengineering"
    adf_factory_name = "data-from-sql-adls"
    adf_pipeline_name = "ProcessFilePipeline"
    adf_url = (
        f"https://management.azure.com/subscriptions/cfb46a3a-9f6d-
4b13-9c3f-2dab3ca2016d/"
        f"resourceGroups/{resource_group}/"
        f"providers/Microsoft.DataFactory/factories/{adf_factory_name}/"
        f"pipelines/{adf_pipeline_name}/createRun?api-version=2018-
06-01"
    )

    data = {"sourceFileName": object_key}

    headers = {
        'Content-Type': 'application/json',
        'Authorization': f'Bearer {os.getenv("ADF_ACCESS_TOKEN")}'
    }

    try:
        # Enable SSL verification (default behavior)
        response = requests.post(adf_url, json=data, headers=headers)

        if response.status_code == 200:
            print(f"Successfully triggered ADF pipeline: "
{response.text}")
        else:
            print(f"Failed to trigger ADF pipeline: "
{response.status_code}")
    except Exception as e:
        print(f"Error triggering ADF pipeline: {str(e)}")

```

➤ 'Authorization': f'Bearer {os.getenv("ADF_ACCESS_TOKEN")}'

Here we are taking this value from env

First we will see how to generate this value

- To get this value we need some details

```

➤ import requests

# Replace these variables with your actual values
tenant_id = "808cc83e-a546-47e7-a03f-73a1ebba24f3"
client_id = "5bf380e3-f965-4dfb-b344-be467ca93de9"
client_secret = "1oJ8Q~QNSKc-BEkmh6FCLPUp_idkRriUXhw1KcNR"
scope = "https://management.azure.com/.default" # This is the scope
for Azure Management API

# Azure AD Token endpoint for OAuth2
token_url =
f"https://login.microsoftonline.com/{tenant_id}/oauth2/v2.0/token"

# Prepare the data for the token request
data = {
    'grant_type': 'client_credentials',
    'client_id': client_id,
    'client_secret': client_secret,
    'scope': scope
}

# Make the request to get the token
response = requests.post(token_url, data=data)

# Parse the JSON response to extract the access token
token = response.json().get('access_token')

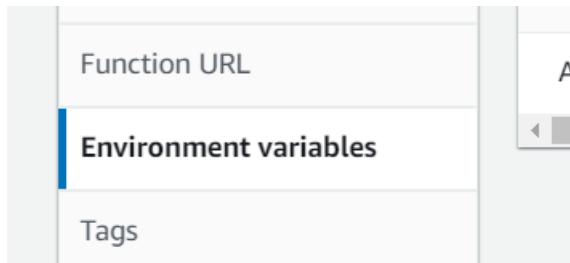
if token:
    print(f"ADF_ACCESS_TOKEN: {token}")
else:
    print("Failed to obtain access token.")
    print(response.json())

```

- this is the another code to generate the ADF_ACCESS_TOKEN
- please give the value we got before
- please run in localmachine python, Jupyter Notebook recommended
- now we need to insert into lambda functions for this go to lambda function

-go to configuration in function

- now select environment variables at left side



- select edit or new

please give according the name given in lambda function code

You can define environment variables as key-value pairs that are accessible from your function code. These are useful to store configuration settings without the need to change function code. [Learn more](#)

Key

ADF_ACCESS_TOKEN

Value

eyJ0eXAiOiJKV1QiLCJhbGciOiJSUzI1NiI...

Remove

Add environment variable

▶ Encryption configuration

- don't add '' for key or value

now start running 😊

Checking the pipeline

Go to data base and clear the data in the table using truncate

Run Cancel query Save query Export data as

```
1 select * from [dbo].[Customers]
2 truncate table [dbo].[Customers]
```

- now add csv file with suitable schema into bucket connected to ADF
- just add file_name.csv into s3 location
- go to function and select monitor and select cloudwatch
- there we can see the log for lambda function status

	Timestamp	Message
No older events at this moment. Retry		
▶	2024-08-13T23:18:50.977Z	INIT_START Runtime Version: python:3.12.v29 Runtime Version ARN: arn:aws:lambda:us-east-1::runtime:667e80dac4962c4a1f927
▶	2024-08-13T23:18:51.395Z	START RequestId: 62085b19-1bbc-466e-9521-965d115b31cf Version: \$LATEST
▶	2024-08-13T23:18:52.106Z	Successfully triggered ADF pipeline: {
▶	2024-08-13T23:18:52.106Z	"runId": "39666614-8651-4758-9ffa-aaab010ff495"
▶	2024-08-13T23:18:52.106Z	}
▶	2024-08-13T23:18:52.121Z	END RequestId: 62085b19-1bbc-466e-9521-965d115b31cf
▶	2024-08-13T23:18:52.121Z	REPORT RequestId: 62085b19-1bbc-466e-9521-965d115b31cf Duration: 725.08 ms Billed Duration: 726 ms Memory Size: 128 MB
No newer events at this moment. Auto retry paused. Resume		

➤ this is the log that we can monitor the logs

➤ now we will check in the database

Query 1

 Run Cancel query  Save query  Export data as  Show only Editor 

```
1 select * from [dbo].[Customers]
2
```

Results

 Search to filter items...

CustomerID	FirstName	LastName
1	John	Doe
2	Jane	Smith

➤ we can see data stored successfully

Mostly the code will work here are few things we need to keep in mind

- make sure the s3 location is correct to trigger
- lambda functions are correct and please check layer for request lib and env variables
- schema of the table and data in the csv should match
- follow triggers for any errors
- if function is giving errors, then try this take the complete code and try it on jupyter note book. Comment the finding the latest object and give file name already presented in the s3 bucket in data code

:)