**Scenario:** **Moving data from AWS S3 bucket to azure SQL (if file size > 100kb and rows in file > 999)**

**Source**: Amazon S3 Bucket

**Destination**: Azure SQL

**Platform** **Used**: Azure Data Factory

1. First, we need to create an S3 bucket in AWS console
2. We need to add a container
3. Add files in to S3 container
4. Create a user and user group
5. Under user we need to create access key and access secret (it can be used to connect s3 bucket)
6. Give get access, list object access of s3 bucket to created user
7. Create linked service to connect s3 bucket in azure data factory
8. Create another linked service to azure SQL
9. Create data set relevant to S3 and azure SQL
10. Create pipeline with set of activities and apply logic regarding to size > 100kb and rows in a file > 999
11. End point to azure SQL
12. Check in Azure SQL for final output

**AWS🡪**

**Creating amazon S3 Bucket**

* **Sign In to AWS:** 
  + Go to the AWS website and sign in with your credentials.
* **Navigate to S3:**
  + **F**rom the AWS Management Console, use the search bar at the top to type “S3” and select it from the results

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

* Bucket will be displayed like this

A screenshot of a computer

AI-generated content may be incorrect.

* Now creating folder inside s3 bucket and adding files .

A screenshot of a computer

AI-generated content may be incorrect.

* Click on create folder

A screenshot of a computer

AI-generated content may be incorrect.

* Add files inside the folder

A screenshot of a computer

AI-generated content may be incorrect.

* Create a user and user group under IAM.
* Click on IAM

A screenshot of a computer

AI-generated content may be incorrect.

* Click on users
* Create user🡪 s3access

A screenshot of a computer

AI-generated content may be incorrect.

* Create user group or select if already created

A screenshot of a computer

AI-generated content may be incorrect.

* Selecting created user group

A screenshot of a computer

AI-generated content may be incorrect.

* Click on review and create

A screenshot of a computer

AI-generated content may be incorrect.

* User will be displayed like this

A screenshot of a computer

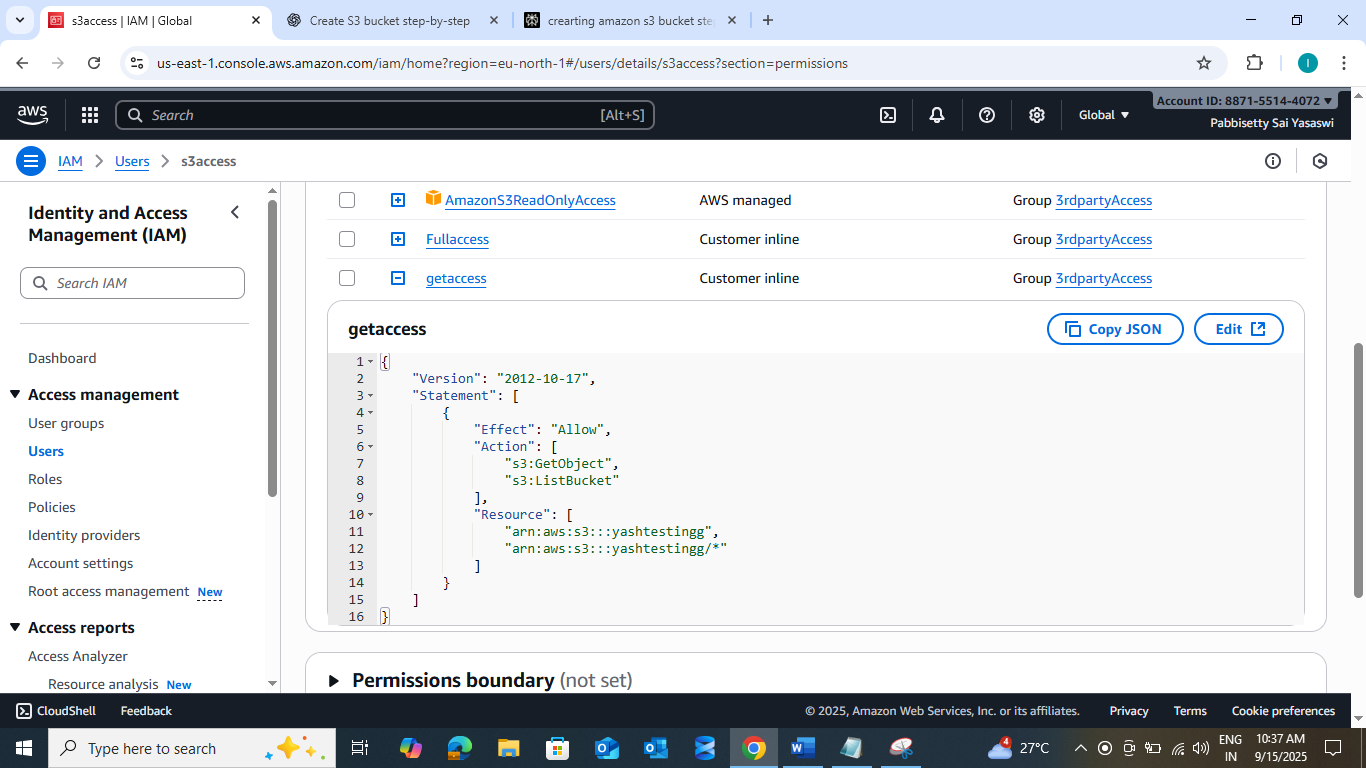
AI-generated content may be incorrect.

* Under the permissions 🡪 click on add permissions 🡪 add inline policy 🡪 select JSON

A screenshot of a computer

AI-generated content may be incorrect.

* Copy and paste the code (get object and listbucket used to select and access files in s3 bucket from azure)



{

"Version": "2012-10-17",

"Statement": [

{

"Effect": "Allow",

"Action": [

"s3:GetObject",

"s3:ListBucket"

],

"Resource": [

"arn:aws:s3:::yashtestingg",

"arn:aws:s3:::yashtestingg/\*"

]

}

]

}

* Create access key under IAM 🡪 select S3 access (created user) 🡪 generate and note user name and password

A screenshot of a computer

AI-generated content may be incorrect.

* **Create linked service in azure to connect amazon s3:**
  + Select s3

A screenshot of a computer

AI-generated content may be incorrect.

* + Give access key and access secret test connection

A screenshot of a computer

AI-generated content may be incorrect.

* + Click on test connection

A screenshot of a computer

AI-generated content may be incorrect.

* **Creating linked service to connect AZURE SQL**
  + Click on manage 🡪 select linked services 🡪 create 🡪 azure sql database🡪
  + Test connection

A screenshot of a computer

AI-generated content may be incorrect.

* **Creating data sets for both amazon s3 and azure sql:**
  + **Ds\_getfileSize\_individual:**

A screenshot of a computer

AI-generated content may be incorrect.

* **Ds\_Azuresql:**

A screenshot of a computer

AI-generated content may be incorrect.

**Pipeline overview:**

A screenshot of a computer

AI-generated content may be incorrect.

Inside Foreach

A screenshot of a computer

AI-generated content may be incorrect.

* **Creating pipeline:**
  + **Getmeta data:**

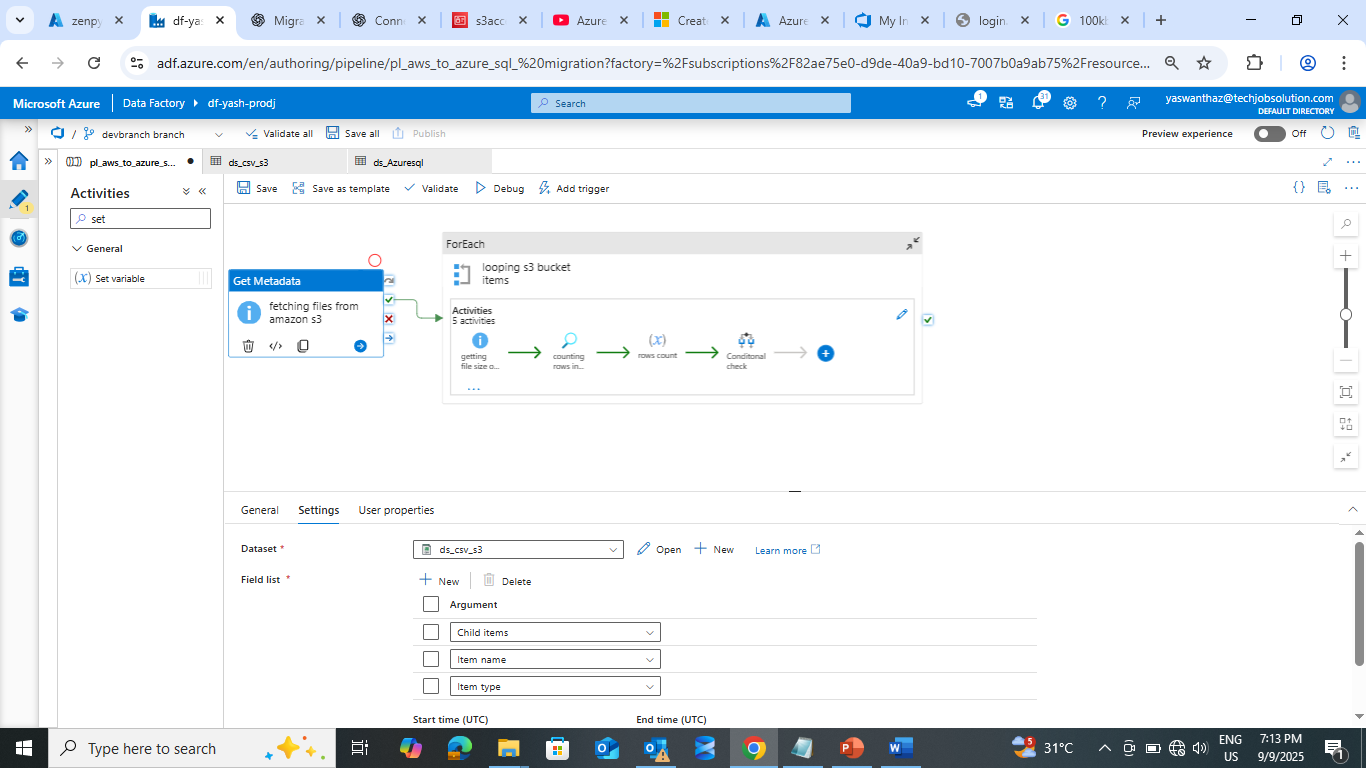
A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

* **For each:**



A screenshot of a computer

AI-generated content may be incorrect.

For each syntax:

@activity('fetching files from amazon s3').output.childItems

* Inside for each getting individual file size

A screenshot of a computer

AI-generated content may be incorrect.

* Counting number of rows in each file for that am using lookup and set variable

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

* Using set variable to count number of rows in each file

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

Code :

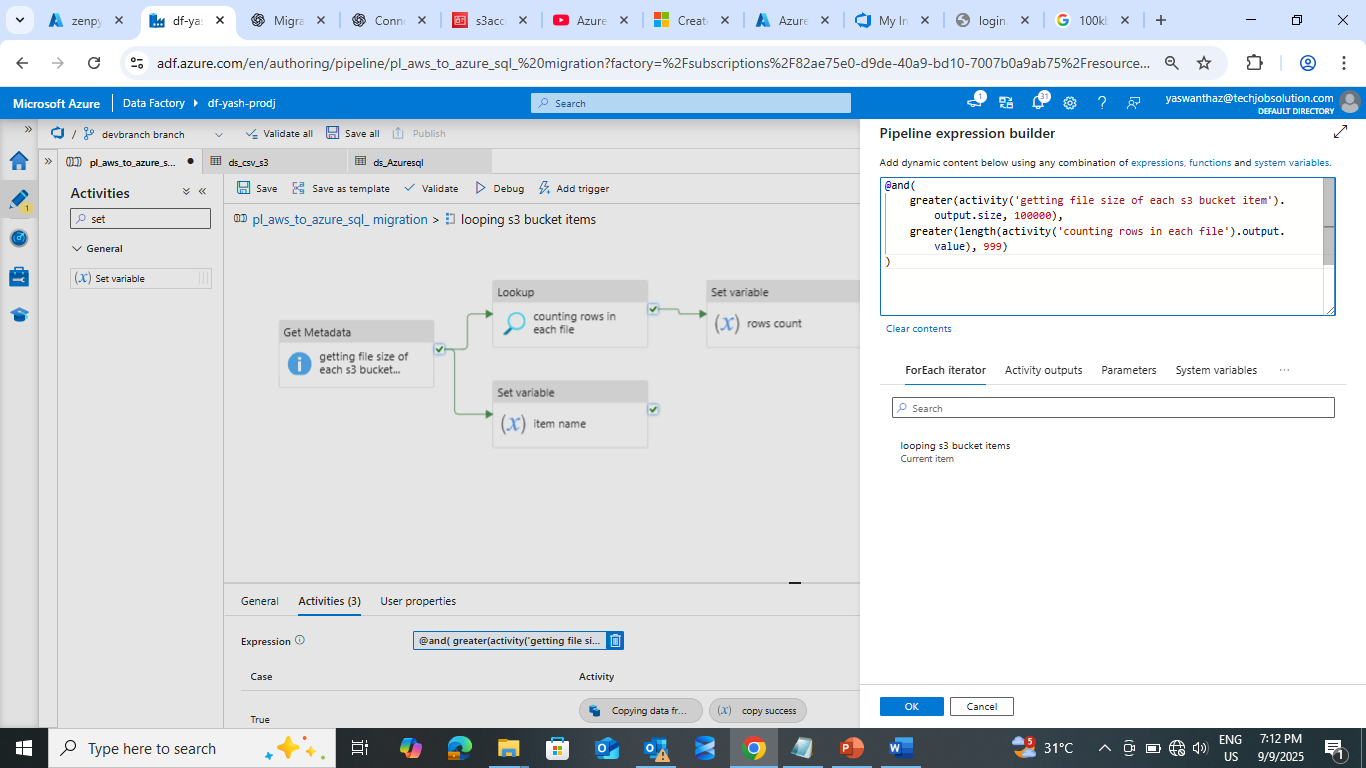
@length(activity('counting rows in each file').output.value)

* Conditional check to check file size > 100KB and rows count > 999 then only it goes to true case other wise goes to false and stores in log table

A computer screen shot of a computer screen

AI-generated content may be incorrect.

* If condition



Logic:

@and(

    greater(activity('getting file size of each s3 bucket item').output.size, 100000),

    greater(length(activity('counting rows in each file').output.value), 999)

)

**Pipeline Success**:

* In execution it will check and filtered the files based on file size > 100kb and rows > 999 then it will store in to AzureSql as a table with same file name

Ex: source file name: bikesales.csv

Destination Azure Sql table name: dbo.bikesales

* True condition need to store data in azure sql🡪 created copy activity

A screenshot of a computer

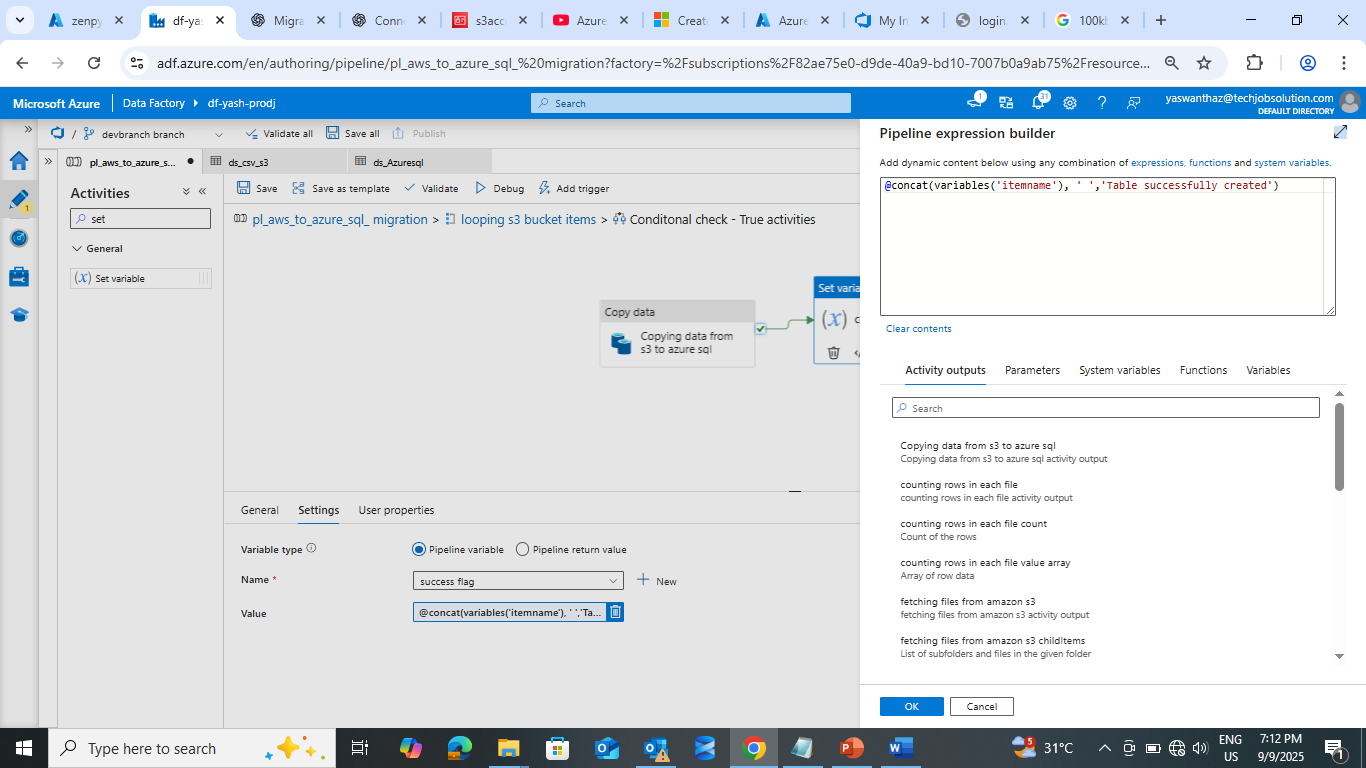
AI-generated content may be incorrect.

* In sink I have mentioned some logic to remove filename extensions of csv files

Example: bikesales.csv (source)🡪 destination bikesales(sql table)

Logic: @split(variables('itemname'), '.')[0]

* Status I am showing with set variable



@concat(variables('itemname'), ' ','Table successfully created')

* Azure sql checking

A computer screen shot of a computer screen

AI-generated content may be incorrect.

* **Pipeline Failure:** In failure scenario it will store the filename and issue of failure in dbo.fileloadlog

A screenshot of a computer

AI-generated content may be incorrect.

* **Creating stored procedure to store the logs**
  + Creating a table in azure

**Table syntax:**

**CREATE TABLE dbo.FileLoadLog**

**(**

**LogID INT NOT NULL PRIMARY KEY,**

**TableName NVARCHAR(200) NULL,**

**NoOfRecords INT NULL,**

**FileSizeKB DECIMAL(18,2) NULL,**

**FailureReason NVARCHAR(500) NULL,**

**LogDate DATE NULL**

**);**

* + Creating stored procedure in azure sql

**Procedure syntax**:

CREATE   PROCEDURE dbo.LogFileFailure

    @TableName NVARCHAR(200),

    @NoOfRecords INT,

    @FileSizeKB DECIMAL(18,2),

    @FailureReason NVARCHAR(500),

    @LogDate DATE

AS

BEGIN

    SET NOCOUNT ON;

    INSERT INTO dbo.FileLoadLog (TableName, NoOfRecords, FileSizeKB, FailureReason, LogDate)

    VALUES (@TableName, @NoOfRecords, @FileSizeKB, @FailureReason, @LogDate);

END;

* Create stored procedure activity in adf and select created adf and mention the parameters values

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

* Parameters syntax:
  + FailureReason:

@concat(

  variables('itemname'),

  ' ',

  string(activity('getting file size of each s3 bucket item').output.size),

  ' bytes (less than 100000), ',

  string(length(activity('counting rows in each file').output.value)),

  ' rows (less than 999)'

)

* + FileSizeKB:

@activity('getting file size of each s3 bucket item').output.size

* + LogDate:

@formatDateTime(utcNow(), 'yyyy-MM-dd')

* + NoOfRecords:

@variables('rowscount')

* + TableName:

@variables('itemname')

* **Error logs inside log table:**

A screenshot of a computer

AI-generated content may be incorrect.