

Data Migration Project

Name: Deepthi

1. Create a Virtual Machine in Azure Cloud, if you have on premise database can use it directly if not have to create virtual machine and use SQL database from there.

To create virtual machine

Go to Azure portal login and click on create a resource.

Click on Virtual machine

Select Subscription name, resource group and give VM name and select a region

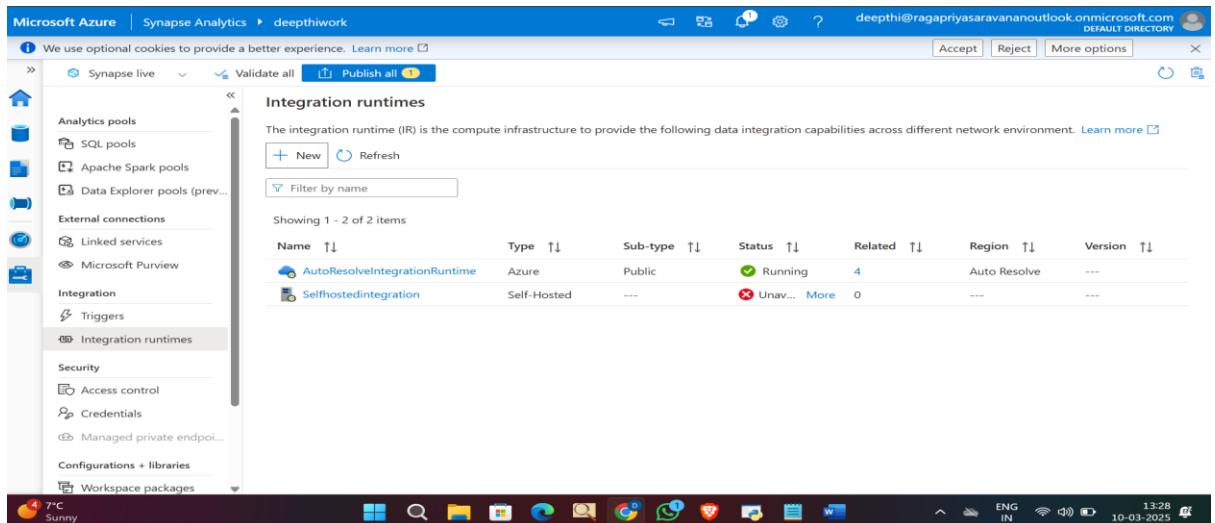
Select SQL server 2022 on Windows Server on 2022 as the machine and for the size select as per our need.

Then go to SQL Server Settings table and enable SQL authentication and review and create.

| Resource | Type | Status | Operation details |
|----------------------|---------------------------|--------|-----------------------------------|
| deepthivm | Microsoft.SqlVirtualM... | OK | Operation details |
| deepthivm | Microsoft.Compute/vir... | OK | Operation details |
| deepthivm725 | Microsoft.Network/net... | OK | Operation details |
| deepthivm-nsg | Microsoft.Network/net... | OK | Operation details |
| deepthivm-vnet | Microsoft.Network/virt... | OK | Operation details |
| deepthivm-ip | Microsoft.Network/pu... | OK | Operation details |
| deepthivm_DataDisk_0 | Microsoft.Compute/di... | OK | Operation details |
| deepthivm_DataDisk_1 | Microsoft.Compute/di... | OK | Operation details |

After this Go to Synapse to create Self hosted integration run time.

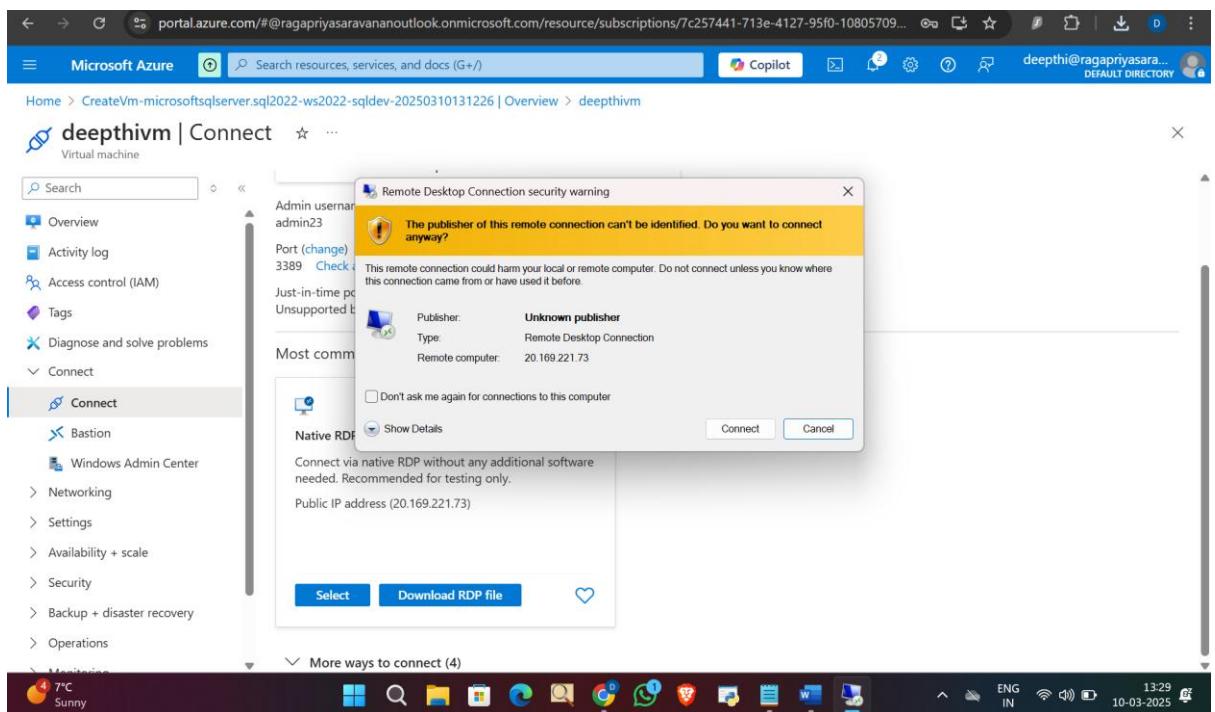
Synapse->Manage->Integration runtime->Create-> Selfhosted-> Give details-> Create->Copy the key



The screenshot shows the Microsoft Azure Synapse Analytics interface. The left sidebar has 'Integration runtimes' selected. The main area displays 'Integration runtimes' with two items listed:

| Name | Type | Sub-type | Status | Related | Region | Version |
|-------------------------------|-------------|----------|---------|---------|--------------|---------|
| AutoResolveIntegrationRuntime | Azure | Public | Running | 4 | Auto Resolve | --- |
| SelfhostedIntegration | Self-Hosted | --- | Unav... | More | 0 | --- |

Next we need to connect to VM, go to create VM -> Connect-> Download RDP



The screenshot shows the Microsoft Azure portal with a virtual machine named 'deepthivm'. The 'Connect' option is selected in the left sidebar. A 'Remote Desktop Connection security warning' dialog box is open, asking if the user wants to connect anyway. The dialog box contains the following information:

The publisher of this remote connection can't be identified. Do you want to connect anyway?

This remote connection could harm your local or remote computer. Do not connect unless you know where this connection came from or have used it before.

Publisher: Unknown publisher
Type: Remote Desktop Connection
Remote computer: 20.169.221.73

Don't ask me again for connections to this computer
 Show Details
Connect Cancel

Below the dialog box, there are 'Select' and 'Download RDP file' buttons.



Windows Security

X

Enter your credentials

These credentials will be used to connect to 20.169.221.73.

admin23

Password

 Password

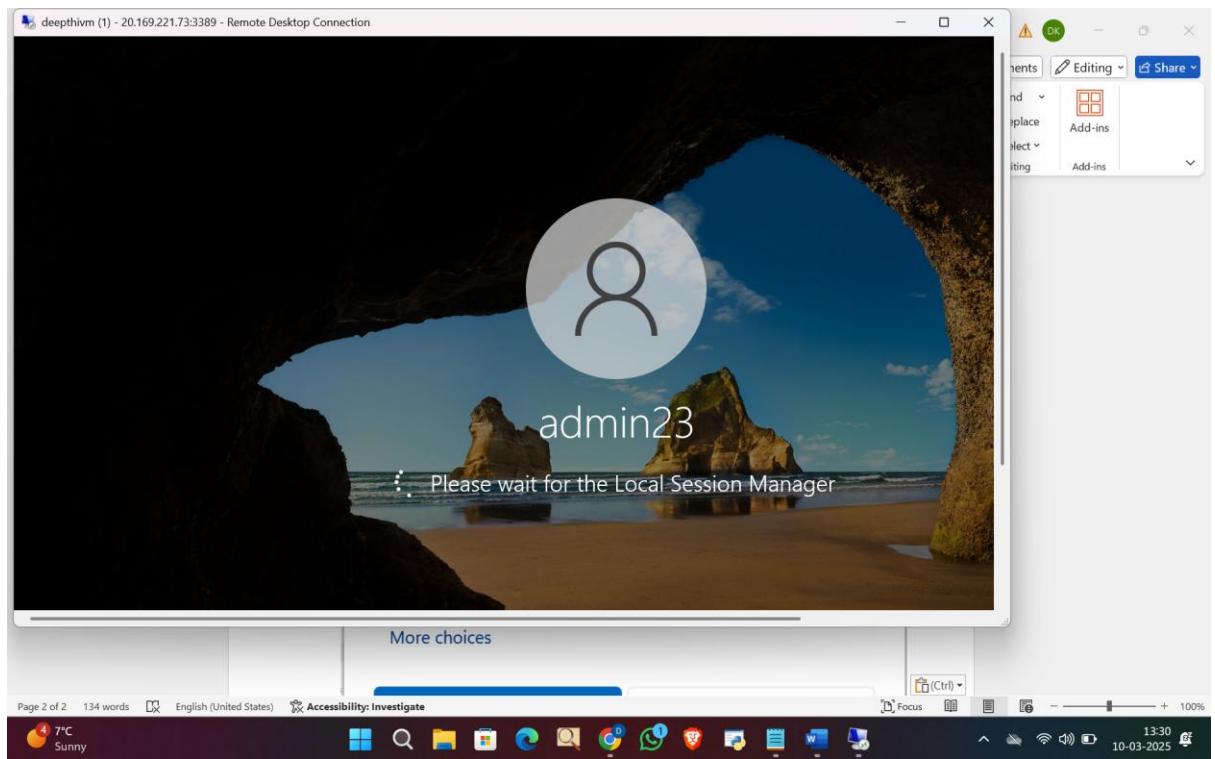
Remember me

[More choices](#)

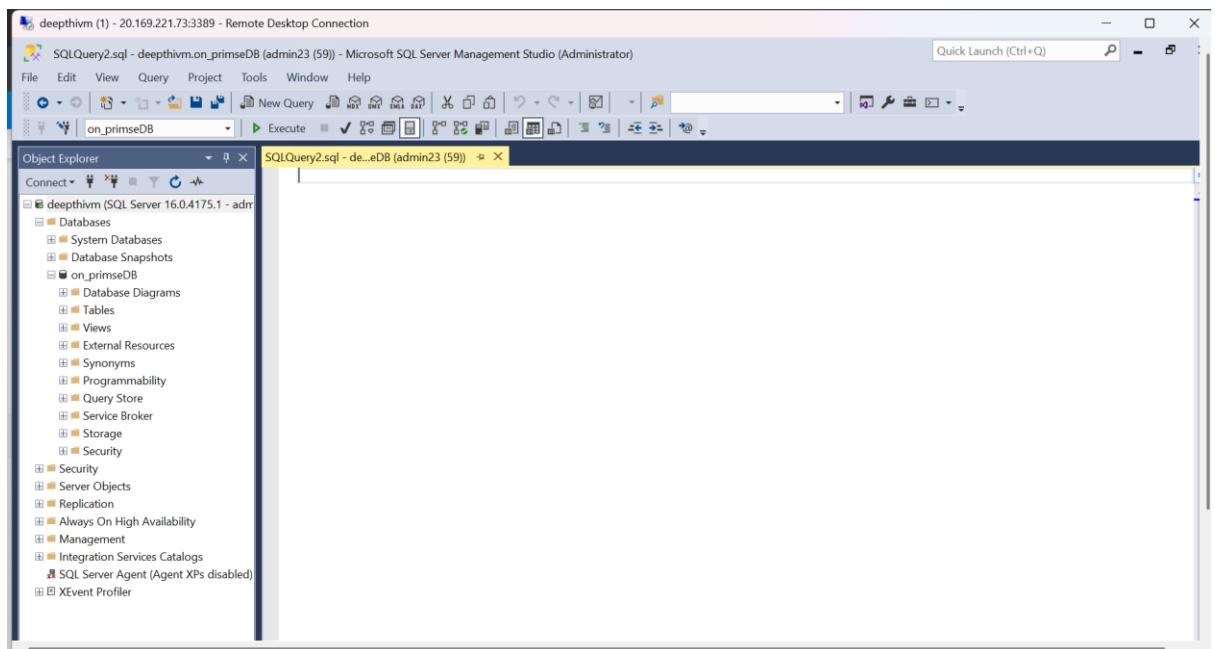
OK

Cancel

VM is loading



Then Download Self hosted runtime in VM and add the key ad register and then go to SSMS and connect it is VM SQL Server and create a DB



Creating a watermark table

```
create table watermark(
    ID int identity(1,1), Schema_name varchar(20), Table_Name varchar(20), LPV varchar(50),
    Delta_Col varchar(20))
```

The screenshot shows a SQL query window titled "SQLQuery1.sql - de...eDB (admin23 (60))". The query is:`create table watermark(
 ID int identity(1,1), Schema_name varchar(20), Table_Name varchar(20), LPV varchar(50),
 Delta_Col varchar(20))`

The status bar at the bottom indicates "Commands completed successfully." and the completion time: 2025-03-11T15:45:31.9799277+00:00.

Inserting values into watermark table

```
insert into watermark values('dbo','Employee',0,'ID')
insert into watermark values('dbo','Customer',0,'ID')
insert into watermark values('dbo','Orders','1900-01-01 00:00:00','Update_Date_Order')
insert into watermark values('dbo','Product','1900-01-01 00:00:00','Update_Date_Product')
insert into watermark values('dbo','Sales','1900-01-01 00:00:00','Update_Date_Sales')
```

The screenshot shows the same SQL query window with the five insert statements highlighted in blue. The status bar at the bottom indicates multiple rows affected and the completion time: 2025-03-11T15:49:12.9745265+00:00.

Creating a Stored procedure

```
Create proc up_watermark
@LPV varchar(50),
@Table_name varchar(50)
as
update watermark set LPV=@LPV where Table_Name = @Table_name
```

```
create proc up_watermark
@LPV varchar(50),
@Table_name varchar(50)
as
update watermark set LPV=@LPV where Table_Name = @Table_name
```

Completion time: 2025-03-11T15:53:09.7864831+00:00

```
SQLQuery1.sql - de...eDB (admin23 (60))*
create table watermark(
ID int identity(1,1), Schema_name varchar(20), Table_Name varchar(20), LPV varchar(50), Delta_Col varchar(20)
)

insert into watermark values('dbo','Employee',0,'ID')
insert into watermark values('dbo','Customer',0,'ID')
insert into watermark values('dbo','Orders','1900-01-01 00:00:00','Update_Date_Order')
insert into watermark values('dbo','Product','1900-01-01 00:00:00','Update_Date_Product')
insert into watermark values('dbo','Sales','1900-01-01 00:00:00','Update_Date_Sales')

Create_proc up_watermark
@LPV varchar(50),
@Table_name varchar(50)
as
update watermark set LPV=@LPV where Table_Name = @Table_name
```

Creating Tables

Table 1: dbo.Employee

```
Create table dbo.Employee (
```

```
ID int, E_Name varchar(20), E_City varchar(20), E_Phonenumbe bigint )
```

```
Create table dbo.Employee (
ID int, E_Name varchar(20), E_City varchar(20), E_Phonenumbe bigint )
```

Completion time: 2025-03-11T16:00:13.3231907+00:00

Inserting values into dbo.Employee table

```
insert into dbo.Employee values (1,'Robert','Toronto','2499791376')
```

```
insert into dbo.Employee values (2,'Ann','Brampton','2499799087')
```

```
insert into dbo.Employee values (3,'John','Montreal','2499793456')
```

The screenshot shows the SQL Server Management Studio interface. In the center pane, there is a code editor window containing three SQL statements:

```
insert into dbo.Employee values (1,'Robert','Toronto','2499791376')
insert into dbo.Employee values (2,'Ann','Brampton','2499799087')
insert into dbo.Employee values (3,'John','Montreal','2499793456')
```

Below the code editor, the message pane displays the results of the execution:

```
(1 row affected)
(1 row affected)
(1 row affected)
```

At the bottom of the interface, the status bar shows the completion time: 2025-03-11T16:02:23.7936746+00:00.

Table 2: dbo.Customer

```
Create table dbo.Customer (
```

```
    ID int, C_Name varchar(20), C_City varchar(20), C_Phonenumbr bigint )
```

The screenshot shows the SQL Server Management Studio interface. In the center pane, there is a code editor window containing the SQL statement to create the Customer table:

```
Create table dbo.Customer (
    ID int, C_Name varchar(20), C_City varchar(20), C_Phonenumbr bigint )
```

Below the code editor, the message pane displays the results of the execution:

```
Commands completed successfully.
```

At the bottom of the interface, the status bar shows the completion time: 2025-03-11T16:03:58.2949116+00:00.

Inserting values into dbo.Customer

```
insert into dbo.Customer values (1,'Deepthi','Sudbury','4169791376')
```

```
insert into dbo.Customer values (2,'Rahul','Brampton','4169799087')
```

```
insert into dbo.Customer values (3,'Priya','Montreal','4169793456')
```

The screenshot shows the SQL Server Management Studio interface. In the center pane, there is a code editor window containing three SQL statements:

```
insert into dbo.Customer values (1,'Deepthi','Sudbury','4169791376')
insert into dbo.Customer values (2,'Rahul','Brampton','4169799087')
insert into dbo.Customer values (3,'Priya','Montreal','4169793456')
```

Below the code editor, the message pane displays the results of the execution:

```
(1 row affected)
(1 row affected)
(1 row affected)
```

At the bottom of the interface, the status bar shows the completion time: 2025-03-11T16:05:12.6550564+00:00.

Table 3: dbo.Orders

```
Create table dbo.Orders (OID int, O_Name varchar(50), O_Type varchar(50),
Update_Date_Order datetime)
```

```
Create table dbo.Orders (
    OID int, O_Name varchar(50), O_Type varchar(50), Update_Date_Order datetime)
```

00 %

Messages

Commands completed successfully.

Completion time: 2025-03-11T16:11:40.2856805+00:00

Inserting values into dbo.Orders

```
insert into dbo.Orders values (1,'Facewash','Cosmetic','2025-01-01 00:00:00')
```

```
insert into dbo.Orders values (2,'Cookies','Food','2025-02-10 00:00:00')
```

```
insert into dbo.Orders values (3,'Pen','Stationary','2025-02-15 00:00:00')
```

```
insert into dbo.Orders values (1,'Facewash','Cosmetic','2025-01-01 00:00:00')
insert into dbo.Orders values (2,'Cookies','Food','2025-02-10 00:00:00')
insert into dbo.Orders values (3,'Pen','Stationary','2025-02-15 00:00:00')
```

00 %

Messages

(1 row affected)

(1 row affected)

(1 row affected)

Completion time: 2025-03-11T16:15:42.3514313+00:00

Table 3: dbo.Product

```
Create table dbo.Product (PID int, P_Name varchar(50), P_Category varchar(50),
Update_Date_Product datetime)
```

```
Create table dbo.Product (
    PID int, P_Name varchar(50), P_Category varchar(50), Update_Date_Product datetime)
```

00 %

Messages

Commands completed successfully.

Completion time: 2025-03-11T16:18:25.2448182+00:00

Inserting values into table

```
insert into dbo.Product values (1,'sneakers','Footwear','2025-01-01 00:00:00')
```

```
insert into dbo.Product values (2,'Hoodie','Clothing','2025-02-10 00:00:00')
```

```
insert into dbo.Product values (3,'Finger rings','Jewellery','2025-02-15 00:00:00')
```

```
insert into dbo.Product values (1,'sneakers','Footwear','2025-01-01 00:00:00')
insert into dbo.Product values (2,'Hoodie','Clothing','2025-02-10 00:00:00')
insert into dbo.Product values (3,'Finger rings','Jewellery','2025-02-15 00:00:00')

10 %
Messages
(1 row affected)
(1 row affected)
(1 row affected)
Completion time: 2025-03-11T16:21:11.3838816+00:00
```

Table 5: dbo.Sales

Create table dbo.Sales (

SID int, Sales_Person_Name varchar(50), S_Item varchar(50), Update_Date_Sales datetime)

```
Create table dbo.Sales (
SID int, Sales_Person_Name varchar(50), S_Item varchar(50), Update_Date_Sales datetime)

0 %
Messages
Commands completed successfully.
Completion time: 2025-03-11T16:22:52.2956461+00:00
```

Inserting into dbo.Sales

insert into dbo.Sales values (1,'Ann','sneakers','2025-02-22 02:00:00')

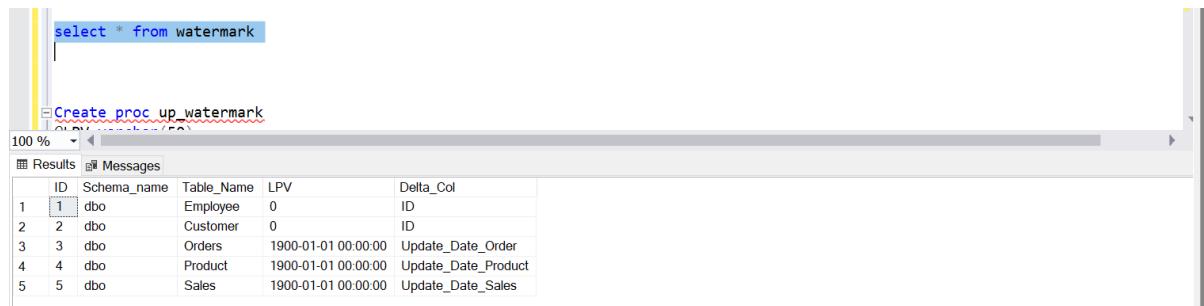
insert into dbo.Sales values (2,'Robert','Hoodie','2024-12-10 10:00:00')

insert into dbo.Sales values (3,'John','Finger rings','2025-03-15 00:00:00')

```
insert into dbo.Sales values (1,'Ann','sneakers','2025-02-22 02:00:00')
insert into dbo.Sales values (2,'Robert','Hoodie','2024-12-10 10:00:00')
insert into dbo.Sales values (3,'John','Finger rings','2025-03-15 00:00:00')

0 %
Messages
(1 row affected)
(1 row affected)
(1 row affected)
Completion time: 2025-03-11T16:26:27.8141683+00:00
```

Values of water mark table



The screenshot shows a SQL Server Management Studio (SSMS) window. In the top-left pane, there is a code editor with the following SQL query:

```
select * from watermark
```

In the top-right pane, there is a message box titled "Create proc up_watermark" with the text "CREATE PROCEDURE up_watermark AS". Below the message box, there is a progress bar labeled "0%".

The bottom half of the window is a results grid titled "Results". It displays the following data:

| ID | Schema_name | Table_Name | LPV | Delta_Col |
|----|-------------|------------|---------------------|---------------------|
| 1 | dbo | Employee | 0 | ID |
| 2 | dbo | Customer | 0 | ID |
| 3 | dbo | Orders | 1900-01-01 00:00:00 | Update_Date_Order |
| 4 | dbo | Product | 1900-01-01 00:00:00 | Update_Date_Product |
| 5 | dbo | Sales | 1900-01-01 00:00:00 | Update_Date_Sales |

Creating a pipeline to migrate data between on_premise SQL DB to Azure ADLS storage account

Go to Synapse click on Integrate and click + to create new pipeline

Drap and drop lookup activity and in setting for Dataset click new and add SQL and create new linked service



New linked service

SQL server [Learn more](#)

Connect via integration runtime * ⓘ

Selfhostedintegration

⚠ The credentials are stored in the machines of self-hosted integration runtime if you don't choose to store them in Azure Key Vault.

Version

Recommended Legacy

[Import from connection string](#)

Server name *

deepthivm

Database name *

on_prismeDB

Authentication type

SQL authentication

User name *

✓ Connection successful

[Create](#)

[Cancel](#)

🔗 [Test connection](#)

After creating the linked service create parameters for Schema and Table name

This screenshot shows the "Parameters" tab of the linked service configuration. It includes a toolbar with "Connection", "Schema", "Parameters", a "New" button, and a "Delete" button. Below is a table with columns: Name, Type, and Default value. Two rows are present: one for "Schema_name" (String type) and another for "Table_name" (String type).

| Name | Type | Default value |
|-------------|--------|---------------|
| Schema_name | String | |
| Table_name | String | |

Then go to connection to connect this parameter

Connection Schema Parameters ^

Linked service * **SqlServer** Test connection Edit + New Learn more ▾

Integration runtime * **Selfhostedintegration** Edit

Table **@dataset().Schema_name** . **@dataset().Table_name**

Enter manually

Preview data

Uncheck First row only

General **Settings** User properties ^

Source dataset * **SqlServerTable1** Open + New Preview data Learn more ▾

Dataset properties ⓘ

| Name | Value | Type |
|------------|-----------|--------|
| Schem_name | dbo | string |
| Table_name | watermark | string |

First row only

Use query Table Query Stored procedure

Query timeout (minutes) ⓘ **120**

Drag on drop ForEach activity so it will loop through lookup, connect lookup to forEach on success.

Microsoft Azure | Synapse Analytics > deepthiwork | Search

We use optional cookies to provide a better experience. Learn more ▾

Synapse live Incremental_pipeline pl.datamigration_i... SqlServerTable1

Activities foreac

Iteration & conditionals ForEach

Lookup

General Settings Activities (0) User properties

Sequential

Batch count ⓘ

Items *

Add dynamic content [A]

Pipeline expression builder

Add dynamic content below using any combination of **expressions, functions and system variables**.

`@activity('Lookup').output.value`

Clear contents

Activity outputs Parameters System variables Functions Variables

Search

Lookup Lookup activity output

Lookup count Count of the rows

Lookup value array Array of row data

OK Cancel

Inside forEach drag and drop another lookup activity which is used to find the max value for the given table, use the below query

```
select max(@{item().Delta_Col}) as maxvalue from
@{item().Schema_Name}.@{item().Table_Name}
```

The screenshot shows the Microsoft Azure Synapse Analytics pipeline editor. A 'Lookup' activity named 'LookupMax' is selected. The 'Settings' tab is open, showing 'Schema_name' and 'Table_name'. Under 'First row only', 'Use query' is selected, and the 'Query' field contains the following SQL:

```
select max(@{item().Delta_Col}) as maxvalue from @{item().Schema_Name}:@{item().Table_Name}
```

The 'ForEach iterator' tab is selected in the pipeline expression builder. The expression is set to check if the maximum value from the previous LookupMax activity equals the current item's LPV value.

Then drag and drop if condition to avoid creating empty files when there is no new record or update in tables

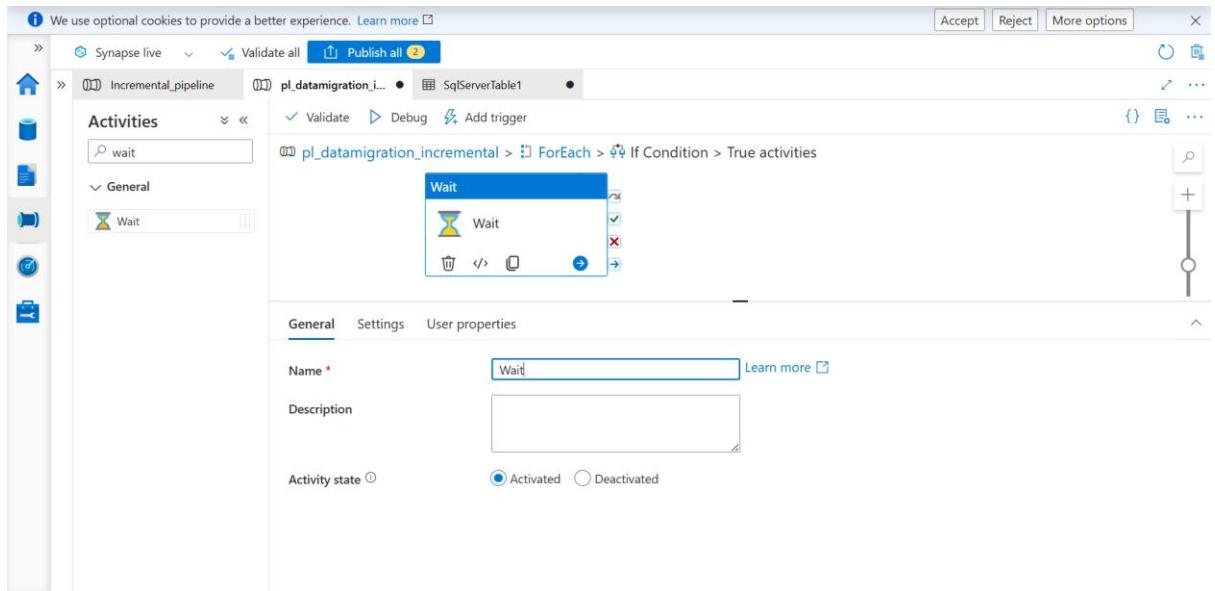
The screenshot shows the Microsoft Azure Synapse Analytics pipeline editor. An 'If' condition activity is selected. The 'Activities (0)' tab is selected in the pipeline expression builder. The expression is set to check if the maximum value from the previous LookupMax activity equals the current item's LPV value.

If max value and LPV value are same it will wait and if it false it will perform copy activity, query used is

`@equals(string(activity('LookupMax').output.firstRow.maxvalue),item().LPV)`

If True wait

Drap and drop wait activity



If false, perform other activities

Drag and drop copy activity first and here source is On-Premise SQL DB and we are bring data using query

```
select * from @{item().Schema_name}.@{item().Table_Name} where
@{item().Delta_Col}>@{item().LPV}
```

For sink our target is ADSL gen2, connect to linked services of ADSL and for folder name and file name we are creating parameters

| Name | Type | Default value |
|-------------|--------|---------------|
| Folder_name | String | Value |
| File_name | String | Value |

Connection Schema Parameters

Linked service * AzureDataLakeStorage1 Test connection Edit + New Learn more

Integration runtime * AutoResolveIntegrationRuntime Edit

File path ewd / @dataset().Folder_name / @dataset().File_name Browse Preview data

Compression type No compression

For folder name:

For File name:

Now drag and drop stored procedure which will be used to update LPV values of water mark table.

pl_datamigration_incremental > ForEach > If Condition - False activities

General Settings User properties

To reference SQL pool, use the SQL pool stored procedure instead.

Linked service * Test connection Edit New

Integration runtime * Edit

Stored procedure name * Refresh Enter manually

Stored procedure parameters

| <input type="checkbox"/> | Name | Type | Value |
|--------------------------|------------|--------|---|
| <input type="checkbox"/> | LPV | String | @activity('LookupMax').output.firstR... |
| <input type="checkbox"/> | Table_name | String | @item().Table_Name |

LPV and Table values we will get by using below expression

Microsoft Azure | Synapse Analytics > deepthiwork | Search

We use optional cookies to provide a better experience. Learn more ↗

Synapse live Validate all Publish all

Incremental_pipeline pl_datamigration_i... SqlServerTable1

Activities Stored

Synapse SQL pool stored proc...

General Stored procedure

pl_datamigration_incremental > ForEach >

General Settings User properties

To reference SQL pool, use the SQL pool stored procedure instead.

Linked service * Test connection Edit New

Integration runtime * Edit

Stored procedure name * Refresh Enter manually

Stored procedure parameters

| <input type="checkbox"/> | Name | Type | Value |
|--------------------------|------------|--------|---|
| <input type="checkbox"/> | LPV | String | @activity('LookupMax').output.firstR... |
| <input type="checkbox"/> | Table_name | String | @item().Table_Name |

Pipeline expression builder

Add dynamic content below using any combination of expressions, functions and system variables.

```
activity('LookupMax').output.firstRow maxValue
```

Clear contents

Activity outputs Parameters System variables Functions Variables

- Lookup activity output
- Lookup count Count of the rows
- Lookup value array Array of row data
- LookupMax LookupMax activity output
- LookupMax first row Data of the first row

OK Cancel

Publish and run the pipeline

| Activity name | Activity st... | Activit... | Run start | Duration | Integration runtime |
|---------------|----------------|------------|-----------------------|----------|-----------------------|
| LookupMax | Queued | Lookup | 3/11/2025, 1:06:34 PM | 2s | |
| LookupMax | Queued | Lookup | 3/11/2025, 1:06:34 PM | 2s | |
| LookupMax | Queued | Lookup | 3/11/2025, 1:06:34 PM | 2s | |
| LookupMax | Queued | Lookup | 3/11/2025, 1:06:34 PM | 2s | |
| ForEach | In progress | ForEach | 3/11/2025, 1:06:33 PM | 3s | |
| Lookup | Succeeded | Lookup | 3/11/2025, 1:04:59 PM | 1m 32s | Selfhostedintegration |

Pipeline ran successfully

| Activity name | Activity st... | Activit... | Run start | Duration | Integration runtime |
|---------------|----------------|--------------|-----------------------|----------|-----------------------|
| Copy data | Succeeded | Copy data | 3/11/2025, 1:38:07 PM | 17s | Selfhostedintegration |
| If Condition | Succeeded | If Condition | 3/11/2025, 1:38:06 PM | 1m 3s | |
| Copy data | Succeeded | Copy data | 3/11/2025, 1:38:06 PM | 24s | Selfhostedintegration |
| If Condition | Succeeded | If Condition | 3/11/2025, 1:38:05 PM | 59s | |
| Copy data | Succeeded | Copy data | 3/11/2025, 1:37:58 PM | 15s | Selfhostedintegration |
| If Condition | Succeeded | If Condition | 3/11/2025, 1:37:57 PM | 41s | |
| LookupMax | Succeeded | Lookup | 3/11/2025, 1:37:39 PM | 23s | Selfhostedintegration |
| LookupMax | Succeeded | Lookup | 3/11/2025, 1:37:39 PM | 35s | Selfhostedintegration |
| LookupMax | Succeeded | Lookup | 3/11/2025, 1:37:39 PM | 18s | Selfhostedintegration |
| LookupMax | Succeeded | Lookup | 3/11/2025, 1:37:39 PM | 34s | Selfhostedintegration |
| LookupMax | Succeeded | Lookup | 3/11/2025, 1:37:39 PM | 27s | Selfhostedintegration |
| ForEach | Succeeded | ForEach | 3/11/2025, 1:37:38 PM | 1m 56s | |
| Lookup | Succeeded | Lookup | 3/11/2025, 1:37:23 PM | 14s | Selfhostedintegration |

Watermark table is updated with LPV values

```

update watermark set LPV='1900-01-01 00:00:00' where ID=5
select * from watermark

Create proc up_watermark
@LPV varchar(50),

```

Results

| ID | Schema_name | Table_Name | LPV | Delta_Col |
|----|-------------|------------|---------------------|---------------------|
| 1 | dbo | Employee | 3 | ID |
| 2 | dbo | Customer | 3 | ID |
| 3 | dbo | Orders | 2025-02-15T00:00:00 | Update_Date_Order |
| 4 | dbo | Product | 2025-02-15T00:00:00 | Update_Date_Product |
| 5 | dbo | Sales | 2025-03-15T00:00:00 | Update_Date_Sales |

New folders were created in ADLS storage account

Microsoft Azure | Synapse Analytics > deepthiwork

New folder

| Name | Last Modified | Content Type | Size |
|------------|-----------------------|--------------|------|
| Customers | 2/26/2025, 4:51:27 PM | Folder | |
| Delta_file | 3/10/2025, 2:43:09 PM | Folder | |
| Empl | 2/26/2025, 4:31:00 PM | Folder | |
| ewd | 3/11/2025, 1:07:28 PM | Folder | |
| new | 2/21/2025, 6:52:39 PM | Folder | |
| NoColoured | 2/24/2025, 9:46:39 PM | Folder | |
| Orders | 2/26/2025, 6:01:31 PM | Folder | |
| Product | 2/26/2025, 9:03:13 PM | Folder | |
| Sales | 2/26/2025, 6:51:04 PM | Folder | |
| SCD_Type | 3/2/2025, 10:21:24 PM | Folder | |
| synapse | 2/21/2025, 6:48:43 PM | Folder | |

Showing 1 to 13 of 13 cached items

Microsoft Azure | Synapse Analytics > deepthiwork

Employee_2025-03-11T17:07:01.8945923Z.csv

Path: https://adlsdeepthi.dfs.core.windows.net/ewd/Employee/Employee_2025-03-11T17:07:01.8945923Z.csv

Modified: 3/11/2025, 1:07:28 PM

With column header: On

| ID | E_NAME | E_CITY | E_PHONE | Content Type | Size |
|------|--------|----------|------------|--------------|-------|
| 1 | Robert | Toronto | 2499791376 | | 128 B |
| 2 | Ann | Brampton | 2499799087 | | 128 B |
| 3 | John | Montreal | 2499793456 | | |
| NULL | NULL | NULL | NULL | | |

OK

Showing 1 to 2 of 2 cached items

Now will be inserting new records to each table

```

insert into dbo.Sales values (4,'Flex','Waterbottle','2025-03-15 10:00:00')
insert into dbo.Product values (4,'earbuds','electronic','2025-03-25 10:00:00')
insert into dbo.Orders values (4,'Stuffed bear','Toys','2025-03-15 10:00:00')
insert into dbo.Employee values(4,'Flex','Vancouver','2499795678')
insert into dbo.Customer values(4,'Raj','Halifax','4169797865')

100 % ▶ Messages

(1 row affected)

Completion time: 2025-03-11T17:46:25.6923541+00:00

```

Running pipeline again as new records were inserted into tables

Pipeline ran successfully

The screenshot shows the Microsoft Azure Synapse Analytics pipeline run status page. The pipeline run ID is ae554a47-2cee-4987-afca-9a611cc53a25. The Pipeline status is Succeeded. The table below lists the activities and their execution details:

| Activity name | Activity st... | Activit... | Run start | Duration | Integration runtime |
|------------------|----------------|----------------|-----------------------|----------|-----------------------|
| Stored procedure | Succeeded | Stored procedu | 3/11/2025, 1:50:56 PM | 2m 9s | Selfhostedintegration |
| Stored procedure | Succeeded | Stored procedu | 3/11/2025, 1:50:48 PM | 42s | Selfhostedintegration |
| Stored procedure | Succeeded | Stored procedu | 3/11/2025, 1:50:47 PM | 23s | Selfhostedintegration |
| Stored procedure | Succeeded | Stored procedu | 3/11/2025, 1:50:44 PM | 2m 7s | Selfhostedintegration |
| Stored procedure | Succeeded | Stored procedu | 3/11/2025, 1:50:37 PM | 25s | Selfhostedintegration |
| Copy data | Succeeded | Copy data | 3/11/2025, 1:50:31 PM | 16s | Selfhostedintegration |
| If Condition | Succeeded | If Condition | 3/11/2025, 1:50:30 PM | 1m 1s | |
| Copy data | Succeeded | Copy data | 3/11/2025, 1:50:27 PM | 28s | Selfhostedintegration |
| If Condition | Succeeded | If Condition | 3/11/2025, 1:50:26 PM | 2m 43s | |

Watermark table is updated with new LPV values

```

select * from watermark

--insert into dbo.Sales values (4,'Flex','Waterbottle','2025-03-15 10:00:00')
--insert into dbo.Product values (4,'earbuds','electronic','2025-03-25 10:00:00')
--insert into dbo.Orders values (4,'Stuffed bear','Toys','2025-03-15 10:00:00')
--insert into dbo.Employee values(4,'Flex','Vancouver','2499795678')
--insert into dbo.Customer values(4,'Raj','Halifax','4169797865')

100 % ▶ Results

Results Messages

ID Schema_name Table_Name LPV Delta_Col
1 1 dbo Employee 4 ID
2 2 dbo Customer 4 ID
3 3 dbo Orders 2025-03-15T10:00:00 Update_Date_Order
4 4 dbo Product 2025-03-25T10:00:00 Update_Date_Product
5 5 dbo Sales 2025-03-15T10:00:00 Update_Date_Sales

```

New value file is also updated in ADLS storage account

Employee_2025-03-11T17:50:11.6190498Z.csv

Path: https://adlsdeepthi.dfs.core.windows.net/ewd/Employee/Employee_2025-03-11T17:50:11.6190498Z.csv

Modified: 3/11/2025, 1:50:27 PM

With column header: On

| ID | E_NAME | E_CITY | E_PHONE |
|------|--------|-----------|------------|
| 4 | Flex | Vancouver | 2499795678 |
| NULL | NULL | NULL | NULL |

OK

Showing 1 to 3 of 3 cached items

Creating SCD Type 1 and 2

SCD Type 1:

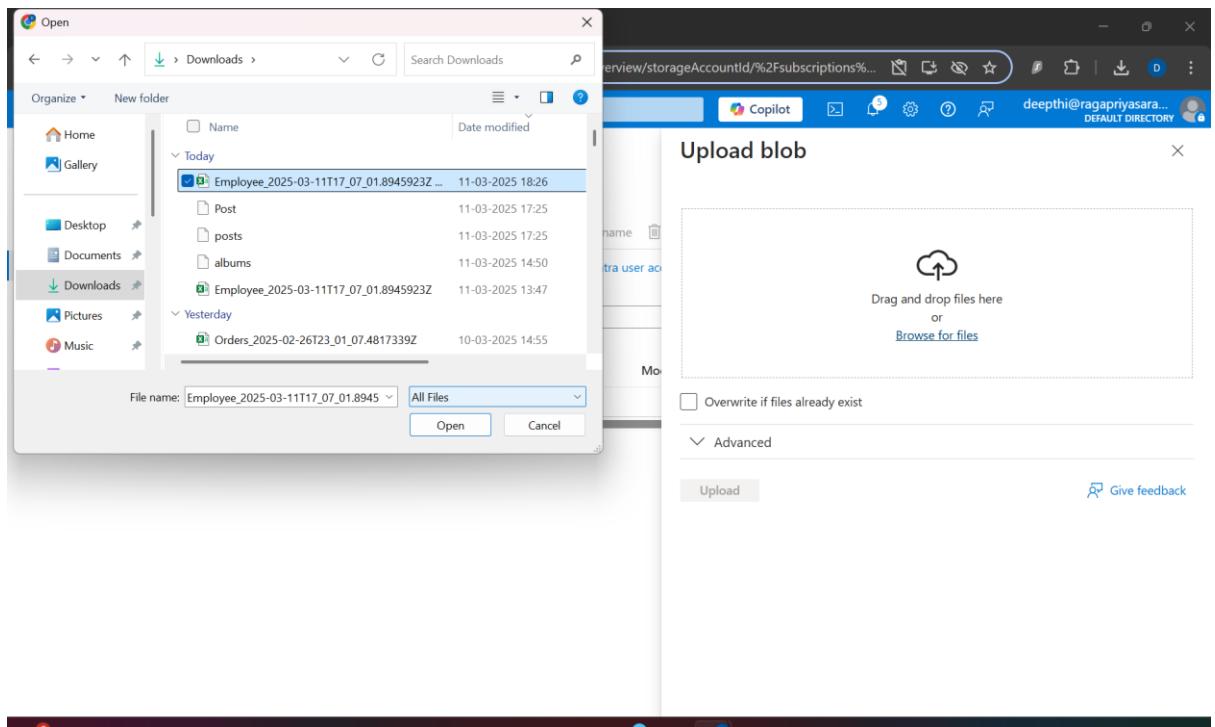
To build a pipeline for SCD type 1 first create a different folder in ADLS and copy a file which we copied using incremental load from on-premise DB.

Created a folder named Data_Migration_SCD and download a file from Employee folder

ewd Container

| Name | Modified | Access tier | Archive status | Blob type | Size |
|--------------------|-----------------------|----------------|----------------|------------|------|
| Customer | 3/11/2025, 1:07:50 PM | | | | |
| Customers | 2/26/2025, 4:51:27 PM | | | | |
| Data_Migration_SCD | 3/11/2025, 6:26:28 PM | | | | |
| Delta_file | 3/10/2025, 2:43:09 PM | | | | |
| Employee | 3/11/2025, 1:07:28 PM | | | | |
| Orders | 2/26/2025, 6:01:31 PM | | | | |
| Product | 2/26/2025, 9:03:13 PM | | | | |
| Rest | 3/11/2025, 2:16:03 PM | | | | |
| Rest_Ap | 3/11/2025, 5:32:09 PM | | | | |
| Sales | 2/26/2025, 6:51:04 PM | | | | |
| post | 3/11/2025, 5:37:54 PM | Hot (Inferred) | | Block blob | 20. |

Uploading downloaded file to created folder

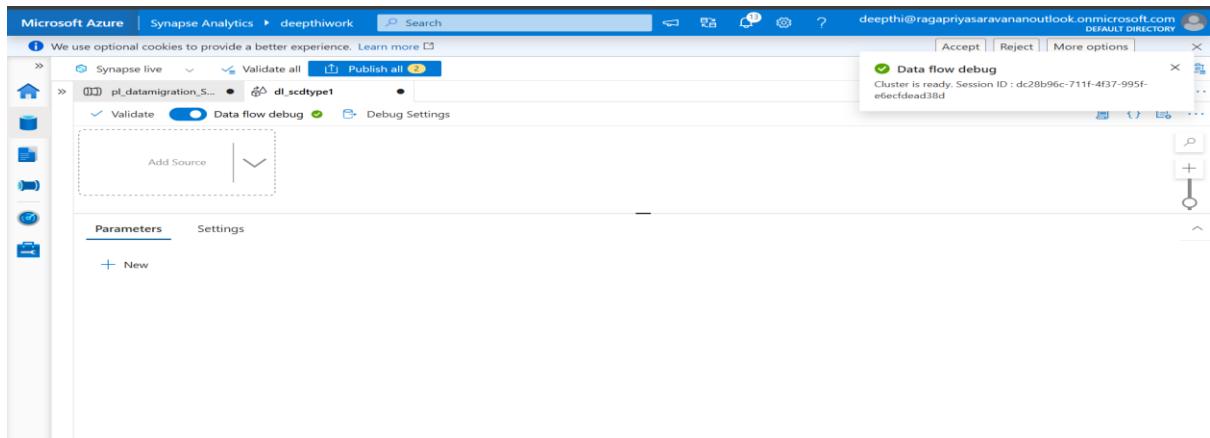


Uploaded the file

The screenshot shows the Microsoft Azure Storage account overview page for the container 'ewd'. The 'Overview' tab is selected. It displays the file 'Employee_2025-03-11T17_07_01.8945923Z (1).csv' with details: Name, Modified (3/11/2025, 6:28:05 PM), Access tier (Hot (Inferred)), Archive status (None), Blob type (Block blob), and Size (128). Other tabs include 'Diagnose and solve problems' and 'Access Control (IAM)'. Top navigation includes 'Search resources, services, and docs (G+)', 'Copilot', and user info 'deepthi@ragapriyasa... DEFAULT DIRECTORY'.

Create pipeline:

Drag and Drop Dataflow and turn on Data flow Debug



Created a table in SSMS

```
create table DM_Employee(
```

```
    ID int, E_Name varchar(20), E_City varchar(20), E_Phonenum bigint, E_Hashkey bigint )
```

```
SQLQuery1.sql - de...iDB (admin23 (80)) * 
create table DM_Employee(
    ID int, E_Name varchar(20), E_City varchar(20), E_Phonenum bigint, E_Hashkey bigint
)
```

Messages

Commands completed successfully.

Completion time: 2025-03-11T18:37:51.4400897-04:00

Go to Pipeline add source

Choose source dataset as Delimited text and linked service as ADLS, then add path

Source settings Source options Projection Optimize Inspect Data preview ●

File settings

File mode File Wildcard

File path * ewd / Data_Migration_SCD / Employee_2025-03-...

Allow no files found

Change data capture

Compression type No compression

Encoding Default(UTF-8)

Column delimiter Comma (,)

Got to projection to import the Schema

The screenshot shows the Microsoft Azure Synapse Analytics Data Flow interface. At the top, there is a message: "We use optional cookies to provide a better experience. Learn more" with options to "Accept", "Reject", or "More options". A success message "Successfully imported" is displayed, stating "Successfully imported the schema for source (Source)". The main area shows a data flow pipeline with three stages: "source", "pl_datamigration_S...", and "df_scctype1". The "source" stage has 4 total columns. Below the pipeline, tabs include "Source settings", "Source options", "Projection", "Optimize", "Inspect", and "Data preview". The "Projection" tab is selected. Under "Projection", there are sections for "Import schema", "Clear schema", and "Schema options". The "Schema options" section lists four columns: "_col0_ (string)", "_col1_ (string)", "_col2_ (string)", and "_col3_ (string)". Each column has a dropdown menu labeled "Specify format".

Data Preview

The screenshot shows the "Data preview" tab. At the top, it displays the number of rows: "Number of rows: 3" with "INSERT 3", "UPDATE 0", "DELETE 0", "UPSERT 0", "LOOKUP 0", "ERROR 0", and "TOTAL 3". Below this, there is a table with four columns: "ID", "E_Name", "E_City", and "E_PhoneNumber". The data rows are:

| ID | E_Name | E_City | E_PhoneNumber |
|----|--------|----------|---------------|
| 1 | Robert | Toronto | 2499791376 |
| 2 | Ann | Brampton | 2499799087 |
| 3 | John | Montreal | 2499793456 |

Add select to rename the column names in source.

The screenshot shows the "Dataflow expression builder" interface. It includes sections for "Matching condition" (containing the expression "1==1") and "Output column name expression" (containing the expression "concat('Src_', \$\$)"). Below these, there is a "Expression elements" table and a "Expression values" table. The "Expression elements" table includes rows for "All", "Functions", "Input schema", "Parameters", "Cached lookup", and "Data flow library functions". The "Expression values" table includes rows for "abc \$\$", "abc \$0", and "123 \$#". At the bottom, there are buttons for "Save and finish", "Cancel", and "Clear contents".

Add Derived Column to add hash key column, to generate hash key we are using CRC32 hash function

Microsoft Azure | Synapse Analytics > deepthiwork

Search

Expression reference documentation

Dataflow expression builder

Hashcolumn

Derived Columns

Create new

ANY Src_Hashkey

Column name *

Src_Hashkey

Expression

`crc32(concat(toString(Src_ID),Src_E_Name,toString(Src_E_PhoneNumber)))`

Save

Expression elements

All

Functions

Input schema

Parameters

Cached lookup

Expression values

Filter by keyword

Create new

abc Src_E_City

123 Src_E_PhoneNumber

123 abs(123 numeric_value)

Data preview Refresh

Save and finish Cancel Clear contents

Data preview, hash column is generated.

| Number of rows | | + INSERT 3 | ● UPDATE 0 | ✗ DELETE 0 | + UPSERT 0 | ○ LOOKUP 0 | ✖ ERROR 0 | TOTAL 3 | | |
|----------------|--------|------------|------------|-----------------------------|------------|------------|-------------------|---------------|-------------|--------|
| ↻ Refresh | ↴ | Typecast | v | Change the type of a column | lifted | Statistics | X Remove | Export to CSV | ↴ | |
| ↑↓ | Src_ID | 12s ↑↓ | Src_E_Name | abc ↑↓ | Src_E_City | abc ↑↓ | Src_E_PhoneNumber | 12l ↑↓ | Src_Hashkey | 12l ↑↓ |
| + | 1 | | Robert | | Toronto | | 2499791376 | | 3942289821 | |
| + | 2 | | Ann | | Brampton | | 2499799087 | | 861896313 | |
| + | 3 | | John | | Montreal | | 2499793456 | | 1029932612 | |

Add Target

Source settings Source options Projection Optimize Inspect Data preview

Output stream name * Target Learn more

Description Add source dataset Reset

Source type * Integration dataset Inline Workspace DB

Inline dataset type * Azure SQL Database

Linked service * AzureSqlDatabase Test connection Edit + New

Sampling * Enable Disable

Source settings Source options Projection Optimize Inspect Data preview

Input Table Query Stored procedure

Query * select ID, E_Hashkey from dbo.DM_Employee

Incremental column

Isolation level Read uncommitted

Add lookup

The screenshot shows the 'Lookup settings' tab selected in the top navigation bar. The configuration includes:

- Lookup stream ***: Target
- Match multiple rows**: ⓘ
- Match on ***: Any row
- Lookup conditions ***:
 - Left: Hashcolumn's column
 - Right: Target's column
 - Condition: `12s Src_ID == 123 ID`

Then add conditional Split to check if data need to be inserted or updated.

Insert expression or condition

The screenshot shows the 'Expression' section of the Dataflow expression builder with the following expression:

```
isNull(ID)
```

The 'Expression elements' sidebar shows:

- All
- Functions
- Input schema
- Parameters
- Cached lookup
- Data flow library functions

The 'Expression values' sidebar shows:

- Filter by keyword
- Create new
- 123 Src_Hashkey
- 123 ID
- 123 E_Hashkey

Updating condition

The screenshot shows the 'Expression' section of the Dataflow expression builder with the following expression:

```
Src_ID==ID && Src_Hashkey!=E_Hashkey
```

The 'Expression elements' sidebar shows:

- All
- Functions
- Input schema
- Parameters
- Cached lookup
- Data flow library functions

The 'Expression values' sidebar shows:

- Filter by keyword
- Create new
- 123 Src_Hashkey
- 123 ID
- 123 E_Hashkey
- 123 abs(123 numeric_value)

Conditional split settings

Output stream name * Learn more

Description

Incoming stream *

Split on First matching condition All matching conditions

Split condition

| Stream names | Condition |
|--------------|---|
| Insert | <code>isNull(ID)</code> |
| Update | <code>Src_ID==ID && Src_Hashkey!=E_Hashkey</code> |

Add derived column to both insert and update to add columns

Added new column in insert condition

Derived column's settings

Optimize Inspect Data preview

Incoming stream *

+ Add Clone Delete Open expression builder

Columns *

| Column | Expression |
|--|---|
| <input type="checkbox"/> Src_Createdby | <input type="text" value="Dataflow'"/> |
| <input type="checkbox"/> Src_CreatedDate | <input type="text" value="currentTimestamp()"/> |
| <input type="checkbox"/> Src_Updatedby | <input type="text" value="Dataflow'"/> |
| <input type="checkbox"/> Src_Updateddate | <input type="text" value="currentTimestamp()"/> |

Insert sink:

Dataset as Azure SQL and give schema name and table name which we created and check Allow insert.

Sink Settings Errors Mapping Optimize Inspect Data preview

Schema name * Refresh Success

Table name *

Table action None Recreate table Truncate table

Update method

Allow insert
 Allow delete
 Allow upsert
 Allow update

Use tempdb

Pre SQL scripts List of scripts Custom expression

Then go to Mapping uncheck Add mapping then import schema and reset and map all the columns

Sink Settings Errors Mapping Optimize Inspect Data preview ●

Auto mapping ⚡ 9 mappings: All outputs mapped

| Input columns | | Output columns | |
|--|--|---|--|
| <input type="checkbox"/> 12s Src_ID | <input type="checkbox"/> abc Src_E_Name | <input type="checkbox"/> 123 ID | <input type="checkbox"/> abc E_Name |
| <input type="checkbox"/> abc Src_E_City | <input type="checkbox"/> 121 Src_E_PhoneNumber | <input type="checkbox"/> abc E_City | <input type="checkbox"/> 121 E_PhoneNumber |
| <input type="checkbox"/> 121 Src_Hashkey | <input type="checkbox"/> abc Src_CreatedBy | <input type="checkbox"/> 121 E_Hashkey | <input type="checkbox"/> abc createdby |
| <input type="checkbox"/> abc Src_CreatedDate | <input type="checkbox"/> abc Src_UpdatedBy | <input type="checkbox"/> 121 created_date | <input type="checkbox"/> abc updatedby |
| <input type="checkbox"/> abc Src_UpdatedDate | <input type="checkbox"/> abc Src_UpdatedDate | <input type="checkbox"/> 121 updated_date | <input type="checkbox"/> abc updated_date |

Data preview

Sink Settings Errors Mapping Optimize Inspect Data preview ●

Number of rows TOT

| ID | E_Name | E_City | E_PhoneNumber | E_Hashkey | createdby | created_date | updatedby | updated_date |
|----|--------|----------|---------------|------------|-----------|-------------------------|-----------|--------------------|
| 1 | Robert | Toronto | 2499791376 | 3942289821 | Dataflow | 2025-03-11 23:20:34.889 | Dataflow | 2025-03-11 23:20:3 |
| 2 | Ann | Brampton | 2499799087 | 861896313 | Dataflow | 2025-03-11 23:20:34.889 | Dataflow | 2025-03-11 23:20:3 |
| 3 | John | Montreal | 2499793456 | 1029932612 | Dataflow | 2025-03-11 23:20:34.889 | Dataflow | 2025-03-11 23:20:3 |

Adding column for update

Derived column's settings Optimize Inspect Data preview ●

Output stream name * Learn more

Description

Incoming stream *

Columns *

| Column | Expression |
|-------------|--------------------|
| Updatedby | 'Dataflow' |
| Updateddate | currentTimestamp() |

Add Alter row to modify the record.

Alter row settings Optimize Inspect Data preview ●

Output stream name * Learn more

Description

Incoming stream *

Alter row conditions *

Update if

Add sink to update

Sink Settings Errors Mapping Optimize Inspect Data preview ●

Schema name * dbo Refresh Success

Table name * DM_Employee

Table action None Recreate table Truncate table

Update method Allow insert Allow delete Allow upsert Allow update

Skip writing key columns

Key columns * List of columns Custom expression

Sink Settings Errors **Mapping** Optimize Inspect Data preview ●

Skip duplicate input columns Skip duplicate output columns

Auto mapping Add mapping Delete Reset Import schema View schema 7 mappings: 2 column(s) from the output schema left unmapped

| Input columns | Output columns |
|------------------|------------------|
| Src_ID | 123 ID |
| Src_E_Name | abc_E_Name |
| Src_E_City | abc_E_City |
| Src_E_Phonenumer | 123_E_Phonenumer |
| Src_Hashkey | 123_E_Hashkey |
| Src_Updatedby | abc_updatedby |
| Src Updateddate | abc_updated date |

There is no data in DM_Employee table

```
select * from DM_Employee
```

.00 %

Results Messages

| ID | E_Name | E_City | E_Phonenumer | E_Hashkey | createdby | created_date | updatedby | updated_date |
|----|--------|--------|--------------|-----------|-----------|--------------|-----------|--------------|
|----|--------|--------|--------------|-----------|-----------|--------------|-----------|--------------|

Publish pipeline and run

Pipeline ran successfully

Parameters Variables Settings **Output**

Pipeline run ID: e5dc55bd-f82d-4b6a-9f79-960383a7a91a Monitor in Azure Metrics Export to CSV

Pipeline status Succeeded

All status ▾

Showing 1 - 1 of 1 items

| Activity name | Activity st... | Activit... | Run start | Duration | Integration runtime |
|---------------|---|------------|-----------------------|----------|---|
| Data flow | <input checked="" type="checkbox"/> Succeeded | Data flow | 3/11/2025, 7:27:58 PM | 6m 59s | AutoResolveIntegrationRuntime (Canada Centr |

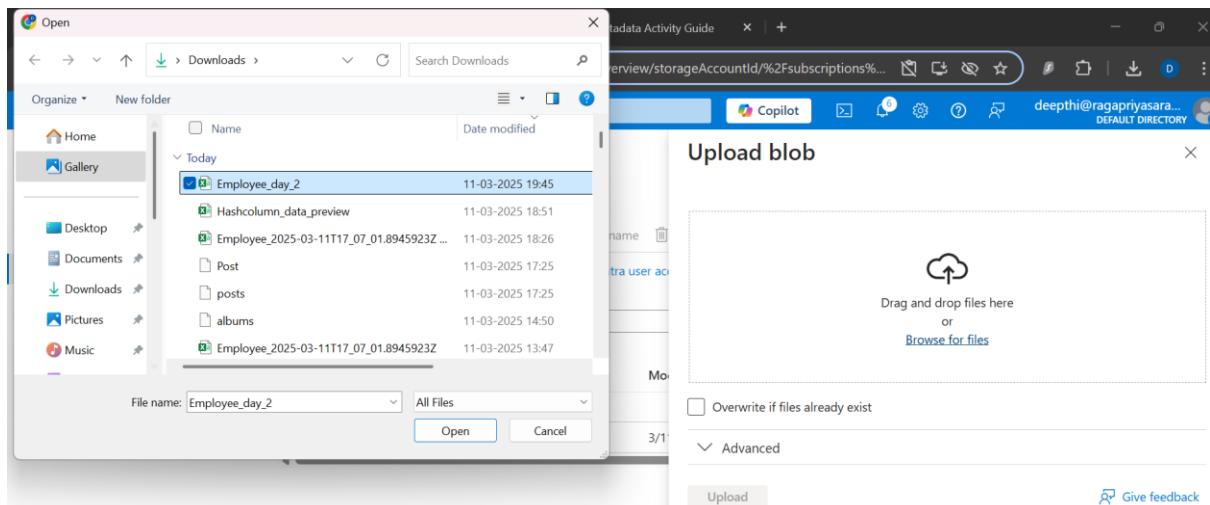
Data added to table

```
select * from DM_Employee
```

Results

| ID | E_Name | E_City | E_Phononenumber | E_Hashkey | createdby | created_date | updatedby | updated_date |
|----|--------|----------|-----------------|------------|-----------|-------------------------|-----------|-------------------------|
| 1 | Robert | Toronto | 2499791376 | 3942289821 | Dataflow | 2025-03-11 23:31:19.767 | Dataflow | 2025-03-11 23:31:19.767 |
| 3 | John | Montreal | 2499793456 | 1029932612 | Dataflow | 2025-03-11 23:31:19.767 | Dataflow | 2025-03-11 23:31:19.767 |
| 2 | Ann | Brampton | 2499799087 | 861896313 | Dataflow | 2025-03-11 23:31:19.767 | Dataflow | 2025-03-11 23:31:19.767 |

Adding second day data to storage container



Path changed in Source

Source settings **Source options** Projection Optimize Inspect Data preview

File settings

File mode File Wildcard

File path * / /

Publish and run pipeline: Pipeline ran successfully

Parameters Variables Settings **Output**

Pipeline run ID: 4e5bf1ff-338f-46e1-a3e1-37e8ff421ebf **Pipeline status:** ✓ Succeeded [View debug run consumption](#)

All status

Showing 1 - 1 of 1 items

| Activity name | Activity st... | Activit... | Run start | Duration | Integration runtime | | |
|---------------|-------------------------------------|------------------------------------|---|-----------|-----------------------|--------|--|
| Data flow | <input type="button" value="Edit"/> | <input type="button" value="Run"/> | <input checked="" type="checkbox"/> Succeeded | Data flow | 3/11/2025, 7:47:15 PM | 6m 50s | AutoResolveIntegrationRuntime (Canada Centr) |

Id 1 data is updated and ID 4 data is inserted which is SCD Type 1

```
select * from DM_Employee
```

100 %

| ID | E_Name | E_City | E_PhoneNumber | E_Hashkey | createdby | created_date | updatedby | updated_date |
|----|---------|----------|---------------|------------|-----------|-------------------------|-----------|-------------------------|
| 1 | Robert | Sudbury | 4169793456 | 3911387829 | Dataflow | 2025-03-11 23:31:19.767 | Dataflow | 2025-03-11 23:50:28.043 |
| 2 | John | Montreal | 2499793456 | 1029932612 | Dataflow | 2025-03-11 23:31:19.767 | Dataflow | 2025-03-11 23:31:19.767 |
| 3 | Ann | Brampton | 2499799087 | 861896313 | Dataflow | 2025-03-11 23:31:19.767 | Dataflow | 2025-03-11 23:31:19.767 |
| 4 | charlie | Montreal | 7896054327 | 2859192875 | Dataflow | 2025-03-11 23:47:36.293 | Dataflow | 2025-03-11 23:47:36.293 |

SCD Type 2

Go to synapse drag and drop Dataflow and then turn on the data flow debug

Then add source

Description: Add source dataset Reset

Source type *: Integration dataset

Inline dataset type *: DelimitedText

Linked service *: AzureDataLakeStorage1 Test connection Edit New

Skip line count:

Sampling *: Enable Disable

Source settings Source options Projection Optimize Inspect Data preview ●

File settings:

- File mode: File Wildcard
- File path: ewd / Data_Migration_SCD / Employee_2025-03...
- Allow no files found:
- Change data capture:
- Compression type: No compression

Go to projection and import schema

| Column name | Type | Format |
|---------------|------------|----------------|
| ID | 12s short | Specify format |
| E_Name | abc string | Specify format |
| E_City | abc string | Specify format |
| E_PhoneNumber | 12l long | Specify format |

Data preview

| Number of rows | | INSERT 3 | UPDATE 0 | DELETE 0 | UPSERT 0 | LOOKUP 0 | ERROR 0 | TOTAL 3 |
|----------------|---------|----------|----------|-------------|------------|----------|---------------|---------|
| ↻ | Refresh | Typecast | Modify | Map drifted | Statistics | Remove | Export to CSV | |
| ↑↓ | ID | 12s ↑↓ | E_Name | abc ↑↓ | E_City | abc ↑↓ | E_PhoneNumber | 12l ↑↓ |
| + | 1 | | Robert | | Toronto | | 2499791376 | |
| + | 2 | | Ann | | Brampton | | 2499799087 | |
| + | 3 | | John | | Montreal | | 2499793456 | |

Add select activity to rename all the columns

The screenshot shows the 'Select settings' blade for a 'RenameColumns' activity. The 'Output stream name' is set to 'RenameColumns'. The 'Description' field contains the note: 'Renaming source to RenameColumns with columns 'Src_ID, Src_E_Name, Src_E_City, Src_E_PhoneNumber''. Under 'Incoming stream', it is set to 'source'. In the 'Options' section, 'Skip duplicate input columns' and 'Skip duplicate output columns' are checked. The 'Input columns' section shows a single mapping: 'source's column' is mapped to 'concat('Src_', \$\$)'. A note at the bottom right says '1 mappings: All inputs mapped'.

Add derived column to add Hash key to source

The screenshot shows the 'Dataflow expression builder' for a derived column named 'Src_Hash'. The 'Column name' is 'Src_Hash' and the 'Expression' is 'crc32(concat(toString(Src_ID), Src_E_Name, Src_E_City, toString(Src_E_PhoneNumber)))'. The 'Expression elements' sidebar lists 'All', 'Functions', 'Input schema', 'Parameters', and 'Cached lookup'. The 'Expression values' sidebar lists 'abc Src_E_City', '123 Src_E_PhoneNumber', '123 abs(123 numeric_value)', and '123 acos(123 numeric_value)'. At the bottom, there are buttons for 'Save and finish', 'Cancel', and 'Clear content'.

Creating a table in SSMS to store SCD type 2 data

```
create table DM_Employee2
```

```
E_ID int, E_name varchar(50), E_City varchar(50), E_PhoneNumber bigint, isActive int, createdby
varchar, created_date datetime,
```

```
updatedby varchar(50), updated_date datetime)
```

The screenshot shows the SQL Server Management Studio (SSMS) query editor with the following code in the query pane:

```
SQLQuery1.sql - de...DB (admin23 (70)) 100 %
create table DM_Employee2(
    E_ID int, E_name varchar(50), E_City varchar(50), E_PhoneNumber bigint, isActive int, createdby varchar, created_date datetime,
    updatedby varchar(50), updated_date datetime)
```

The status bar at the bottom indicates 'Commands completed successfully.' and 'Completion time: 2025-03-11T20:17:56.9553663-04:00'.

Add target which is Azure SQL database

Source settings Source options Projection Optimize Inspect Data preview

Input

Query* Table Query Stored procedure

select E_ID,E_Hashkey from
DM_Employee2 where isActive=1

Add dynamic content [Alt+Shift+D]

Incremental column

Isolation level

Go to projection and import schema

Source settings Source options **Projection** Optimize Inspect Data preview

Import schema Clear schema Schema options Overwrite schema

| Column name | Type |
|-------------|---------|
| E_ID | integer |
| E_Hashkey | long |

Add lookup to do left join on source and target

Lookup settings Optimize Inspect Data preview

Primary stream * Hashkey

Lookup stream * Target

Match multiple rows

Match on * Any row

Lookup conditions * Left: Hashkey's column Right: Target's column

| Src_ID | == | E_ID |
|--------|----|------|
| 123 | = | 123 |

Number of rows **INSERT 3** **UPDATE 0** **DELETE 0** **UPSERT 0** **LOOKUP 0** **ERROR 0** **TOTAL 3**

| Src_ID | Src_E_Name | Src_E_City | Src_E_Phonenumer | Src_Hash | E_ID | E_Hashkey |
|--------|------------|------------|------------------|------------|------|-----------|
| 1 | Robert | Toronto | 2499791376 | 1331461762 | NULL | NULL |
| 2 | Ann | Brampton | 2499799087 | 3761180994 | NULL | NULL |
| 3 | John | Montreal | 2499793456 | 21035172 | NULL | NULL |

Add conditional split for insert and update and give conditions for each

Conditional split settings Optimize Inspect Data preview

Description Conditionally distributing the data in E_ID, Src_ID, E_ID, Src_Hash, E_Hashkey groups, based on columns '{1}'

Incoming stream * lookup

Split on First matching condition All matching conditions

| Split condition | Stream names | Condition |
|-----------------|-------------------------------------|---|
| Insert | isNull(E_ID) | <input type="button" value="X"/> <input type="button" value="+"/> |
| Update | Src_ID==E_ID && Src_Hash!=E_Hashkey | <input type="button" value="X"/> <input type="button" value="+"/> |

Add derived column activity to add columns to update

Derived column's settings Optimize Inspect Data preview

Output stream name * Learn more

Description

Incoming stream *

[+ Add](#) [Clone](#) [Delete](#) [Open expression builder](#)

Columns *

| Column | Expression |
|---|--------------------|
| <input type="checkbox"/> Src_Updatedby | 'Dataflow_updated' |
| <input type="checkbox"/> Src_Updated_Date | currentTimestamp() |
| <input type="checkbox"/> Src_isActive | 0 |

Add alter row to modify the records

Alter row settings Optimize Inspect Data preview

Output stream name * Learn more

Description

Incoming stream *

Alter row conditions * Update if [Open expression builder](#)

Add Update Sink

Sink Settings Errors Mapping Optimize Inspect Data preview

Schema name * Refresh

Table name *

Table action None Recreate table Truncate table

Update method Allow insert
 Allow delete
 Allow upsert
 Allow update

Skip writing key columns

Key columns * List of columns Custom expression [Import schema](#)

[+](#) [Delete](#)

Go to mapping and uncheck add mapping then import schema and then rest and map the required columns

Auto mapping [Import schema](#) [View schema](#) 5 mappings: 5 column(s) from the output schema left unmapped

[+ Add mapping](#) [Delete](#) [Reset](#)

| Input columns | Output columns |
|---|---|
| <input type="checkbox"/> 123 E_ID | <input type="checkbox"/> 123 E_ID |
| <input type="checkbox"/> 123 Src_isActive | <input type="checkbox"/> 123 isActive |
| <input type="checkbox"/> abc Src_Updatedby | <input type="checkbox"/> abc updatedby |
| <input type="checkbox"/> ⏪ Src_Updated_Date | <input type="checkbox"/> ⏪ updated_date |
| <input type="checkbox"/> 121 E_Hashkey | <input type="checkbox"/> 121 E_Hashkey |

Add Union activity to union split insert and Split update

Union settings

Output stream name * [Learn more](#)

Description [Reset](#)

Incoming stream *

Union by * Name Position

Union with * [+](#) [Edit](#)

Data preview

Data preview

| Number of rows | | INSERT 3 | UPDATE 0 | DELETE 0 | UPSERT 0 | LOOKUP 0 | ERROR 0 | TOTAL 3 | | | | | | |
|-------------------------|----------|------------------------|-----------------------------|----------------------------|------------------------|-------------------------------|-------------------|---------|------------|--------|------|--------|-----------|--------|
| Refresh | Typecast | Modify | Map drifted | Statistics | Remove | Export to CSV | | | | | | | | |
| ↑↓ | Src_ID | ↑↓ | Src_E_Name | abc ↑↓ | Src_E_City | abc ↑↓ | Src_E_PhoneNumber | 121 ↑↓ | Src_Hash | 121 ↑↓ | E_ID | 123 ↑↓ | E_Hashkey | 121 ↑↓ |
| + | 1 | | Robert | | Toronto | | 2499791376 | | 1331461762 | | NULL | | NULL | |
| + | 2 | | Ann | | Brampton | | 2499799087 | | 3761180994 | | NULL | | NULL | |
| + | 3 | | John | | Montreal | | 2499793456 | | 21035172 | | NULL | | NULL | |

Add derived column to add columns to insert

Derived column's settings

Incoming stream *

[Add](#) [Clone](#) [Delete](#) [Open expression builder](#)

Columns *

| Column | Expression |
|-----------------|--------------------|
| Src_createdby | 'Dataflow' |
| Src_createddate | currentTimestamp() |
| Src_updatedby | 'Dataflow' |
| Src_updateddate | currentTimestamp() |
| Src_IsActive | 1 |

Data Preview

[Refresh](#) [Typecast](#) [Modify](#) [Map drifted](#) [Statistics](#) [Remove](#) [Export to CSV](#)

| ↑↓ | Src_ID | 12s ↑↓ | Src_E_Name | abc ↑↓ | Src_E_City | abc ↑↓ | Src_E_PhoneNumber | 121 ↑↓ | Src_Hash | 121 ↑↓ | E_ID | 123 ↑↓ | E_Hashkey | 121 ↑↓ | Src_createdby | abc ↑↓ | Src_createdd |
|----|--------|--------|------------|--------|------------|--------|-------------------|--------|------------|--------|------|--------|-----------|--------|---------------|--------|--------------|
| + | 1 | | Robert | | Toronto | | 2499791376 | | 1331461762 | | NULL | | NULL | | Dataflow | | 2025-03-12 (|
| + | 2 | | Ann | | Brampton | | 2499799087 | | 3761180994 | | NULL | | NULL | | Dataflow | | 2025-03-12 (|
| + | 3 | | John | | Montreal | | 2499793456 | | 21035172 | | NULL | | NULL | | Dataflow | | 2025-03-12 (|

Sink for insert

Sink type * Integration dataset Inline Workspace DB Cache

Inline dataset type *

Linked service * [Test connection](#) [Edit](#) [New](#)

Options Allow schema drift [?](#) Validate schema [?](#)

Sink Settings Errors Mapping Optimize Inspect Data preview

Schema name * dbo Refresh Success

Table name * DM_Employee2

Table action None Recreate table Truncate table

Update method

- Allow insert
- Allow delete
- Allow upsert
- Allow update

Mapping

Auto mapping Add mapping Delete Reset Import schema View schema 10 mappings: All outputs mapped

| Input columns | Output columns |
|----------------------|------------------|
| 12s Src_ID | 123 E_ID |
| abc Src_E_Name | abc E_name |
| abc Src_E_City | abc E_City |
| 12l Src_E_Phonenumer | 12l E_Phonenumer |
| 123 Src_isActive | 123 isActive |
| abc Src_createdby | abc createdby |
| Src_createddate | Src_createddate |
| abc Src_Updatedby | abc updatedby |
| Src_Updateddate | Src_updateddate |
| 12l Src_Hash | 12l E_Hashkey |

Data preview

| Number of rows | INSERT 0 | UPDATE 0 | DELETE 0 | UPSERT 0 | LOOKUP 0 | ERROR 0 |
|--|----------|----------|----------|----------|----------|---------|
| Refresh Statistics Export to CSV | | | | | | |
| ↓ E_ID 123 ↓ E_name abc ↑↓ E_City abc ↑↓ E_Phonenumer 12l ↑↓ isActive 123 ↑↓ createdby abc ↑↓ created_date ↗↑↓ updatedby abc ↑↓ updated_date ↗ | | | | | | |
| + 1 Robert Toronto 2499791376 1 Dataflow 2025-03-12 02:28:26.334 Dataflow 2025-03-12 02:28:26 | | | | | | |
| + 2 Ann Brampton 2499799087 1 Dataflow 2025-03-12 02:28:26.334 Dataflow 2025-03-12 02:28:26 | | | | | | |
| + 3 John Montreal 2499793456 1 Dataflow 2025-03-12 02:28:26.334 Dataflow 2025-03-12 02:28:26 | | | | | | |

Add Get data activity to get every data dynamically

Validate Debug Add trigger Data flow debug

Get Metadata Data flow

Get Metadata1

General Settings User properties

Dataset * DelimitedText8 Open New Learn more

Field list * New Delete Argument Child items Item name

Filter by last modified Start time (UTC) End time (UTC)

Skip line count

Give this expression in pipeline

Pipeline expression builder

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

```
@activity('Get Metadata1').output.childItems[0].name
```

[Clear contents](#)

Activity outputs [Parameters](#) [System variables](#) [Functions](#) [Variables](#)

[Search](#)

Get Metadata1
Get Metadata1 activity output

Get Metadata1 childItems
List of subfolders and files in the given folder

Get Metadata1 exists
Whether a file, folder, or table exists

[OK](#) [Cancel](#)

Create a parameter in dataflow and give that parameter as filename in source so that it will get file name from GetMeta Data.

Publish and run pipeline

Pipeline ran successfully

The screenshot shows the Azure Data Factory Pipeline Run history page. At the top, there are navigation links for Validate, Debug, Add trigger, and Data flow debug. The Data flow debug link is selected, indicated by a blue circle with a checkmark. Below the links is a search bar and a plus sign icon for adding new components.

The main area displays a pipeline diagram with two activities: "Get Metadata" and "Data flow". The "Get Metadata" activity has a green checkmark and is connected to the "Data flow" activity, which also has a green checkmark. A green arrow points from the "Get Metadata" activity to the "Data flow" activity.

Below the diagram, there are tabs for Parameters, Variables, Settings, and Output. The Output tab is selected, showing the following details:

- Pipeline run ID:** 76f8975c-9d48-4ed1-99bb-c4fb13f76e26
- Pipeline status:** Succeeded
- View debug run consumption**
- All status:** ▾
- Showing 1 - 2 of 2 items**

| Activity name | Activity st... | Activit... | Run start | Duration | Integration runtime |
|---------------|----------------|--------------|------------------------|----------|--|
| Data flow | Succeeded | Data flow | 3/12/2025, 10:11:23 AM | 9m 42s | AutoResolveIntegrationRuntime (Canada Centr) |
| Get Metadata1 | Succeeded | Get Metadata | 3/12/2025, 10:11:11 AM | 11s | AutoResolveIntegrationRuntime (Canada Centr) |

Table DM_Employee2 is updated with data

```

ALTER TABLE DM_Employee2 DROP COLUMN createdby,
select * from DM_Employee2

```

.00 % ▶

Results Messages

| | E_ID | E_name | E_City | E_PhoneNumber | isActive | created_date | updatedBy | updated_date | E_Hashkey | createdby |
|---|--------|----------|------------|---------------|-------------------------|--------------|-------------------------|--------------|------------|-----------|
| 1 | Robert | Toronto | 2499791376 | 1 | 2025-03-12 14:14:49.650 | Dataflow | 2025-03-12 14:14:49.650 | 1331461762 | 1331461762 | Dataflow |
| 2 | John | Montreal | 2499793456 | 1 | 2025-03-12 14:14:49.650 | Dataflow | 2025-03-12 14:14:49.650 | 21035172 | 21035172 | Dataflow |
| 3 | Ann | Brampton | 2499799087 | 1 | 2025-03-12 14:14:49.650 | Dataflow | 2025-03-12 14:14:49.650 | 3761180994 | 3761180994 | Dataflow |

To get second day data we need to delete first day data for that we need to use Delete activity, drag and drop the delete activity and create a dataset and linked service

General Source Logging settings¹ User properties

Dataset * Binary2 Open + New Learn more

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

Filter by last modified Start time (UTC) End time (UTC)

Recursively

Max concurrent connections 1

Create a parameter and connect it

Connection Parameters

Linked service * AzureDataLakeStorage1 Test connection Edit + New Learn more

Integration runtime * AutoResolveIntegrationRuntime Edit

File path ewd / Data_Migration_SCD_2 / @dataset().filename Browse

Compression type No compression

Giving path to delete the file

Microsoft Azure | Synapse Analytics > deepthiwork | Search

We use optional cookies to provide a better experience. Learn more ▾

Synapse live Validate all Publish all 3

Activities d

Get Metadata

General Source Logging settings User properties

Dataset * Binary2

File path type File path in dataset

Filter by last modified Start time (UTC)

Recursively

Max concurrent connections 1

Pipeline expression builder

Add dynamic content below using any combination of expressions, functions and system variables.

```
@activity('Get Metadata1').output.childItems[0].name
```

Clear contents

Activity outputs Parameters System variables Functions Variables

Search

Data flow Data flow activity output

Get Metadata1 Get Metadata1 activity output

Get Metadata1 childItems List of subfolders and files in the given folder

OK Cancel

Publish and run pipeline, pipeline ran successfully

The screenshot shows the Azure Data Factory Pipeline Editor interface. At the top, there are buttons for Validate, Debug, Add trigger, and Data flow debug. Below the toolbar is a pipeline canvas with three activities: Get Metadata, Data flow, and Delete. The Data flow activity has a green checkmark and a delete icon. A green checkmark is also present next to the Delete activity. Below the canvas, tabs for Parameters, Variables, Settings, and Output are visible, with Output selected. Pipeline run details show a Pipeline run ID of bb72d609-9452-4632-bae4-493b1278f378, a Pipeline status of Succeeded, and a run start time of 3/12/2025, 10:41:13 AM. A table below lists the activities and their details:

| Activity name | Activity st... | Activit... | Run start | Duration | Integration runtime |
|---------------|----------------|--------------|------------------------|----------|---|
| Delete | Succeeded | Delete | 3/12/2025, 10:41:13 AM | 3s | AutoResolveIntegrationRuntime (Canada Ce) |
| Data flow | Succeeded | Data flow | 3/12/2025, 10:31:44 AM | 9m 28s | AutoResolveIntegrationRuntime (Canada Ce) |
| Get Metadata1 | Succeeded | Get Metadata | 3/12/2025, 10:31:38 AM | 5s | AutoResolveIntegrationRuntime (Canada Ce) |

File deleted from storage account

The screenshot shows the Azure Storage Explorer interface. It displays a list of blobs in a container named 'ewd' under the path 'Data_Migration_SCD_2'. A yellow warning icon indicates that a file has been deleted. The file 'Employee_day_1.csv' is shown with a delete icon. The interface includes standard navigation and management tools like Upload, Add Directory, Refresh, Rename, Delete, Change tier, Acquire lease, Break lease, and Give feedback.

Now for data 2 add file to storage account and run pipeline again

The screenshot shows the Azure Storage Explorer interface with a file upload dialog open over the storage account view. The dialog is titled 'Upload blob' and shows a preview area with a single file selected: 'Employee_day_2.csv'. The dialog includes options to Overwrite if files already exist and an Advanced section. The background shows the storage account overview with the 'ewd' container selected.

Uploaded Dat 2 file

| Name | Modified | Access tier | Archive status | Blob type | Size |
|--------------------|-------------------------|----------------|----------------|------------|------|
| ... | | | | | |
| Employee_day_2.csv | 3/12/2025, 10:51:47 ... | Hot (Inferred) | | Block blob | 92 B |

Run pipeline

Pipeline ran successfully

Parameters Variables Settings Output

Pipeline run ID: 18676e2f-99b5-4736-8286-7d145a73b1a9 Pipeline status: Succeeded View debug run consumption

All status Monitor in Azure Metrics Export to CSV

Showing 1 - 3 of 3 items

| Activity name | Activity st... | Activit... | Run start | Duration | Integration runtime |
|---------------|----------------|--------------|------------------------|----------|---|
| Delete | Succeeded | Delete | 3/12/2025, 11:02:00 AM | 10s | AutoResolveIntegrationRuntime (Canada Ce) |
| Data flow | Succeeded | Data flow | 3/12/2025, 10:52:26 AM | 9m 33s | AutoResolveIntegrationRuntime (Canada Ce) |
| Get Metadata1 | Succeeded | Get Metadata | 3/12/2025, 10:52:20 AM | 5s | AutoResolveIntegrationRuntime (Canada Ce) |

New data is inserted and old data is updated with new value and before updated record is still present with isActive value as 0

```
ALTER TABLE DM_Employee2 DROP COLUMN createdby,
select * from DM_Employee2
```

.00 %

Results Messages

| E_ID | E_name | E_City | E_Phonenumer | isActive | created_date | updatedby | updated_date | E_Hashkey | createdby |
|------|---------|----------|--------------|----------|-------------------------|------------------|-------------------------|------------|-----------|
| 1 | Robert | Toronto | 2499791376 | 0 | 2025-03-12 14:14:49.650 | Dataflow_updated | 2025-03-12 14:52:47.040 | 1331461762 | Dataflow |
| 2 | John | Montreal | 2499793456 | 1 | 2025-03-12 14:14:49.650 | Dataflow | 2025-03-12 14:14:49.650 | 21035172 | Dataflow |
| 3 | Ann | Brampton | 24997799087 | 1 | 2025-03-12 14:14:49.650 | Dataflow | 2025-03-12 14:14:49.650 | 3761180994 | Dataflow |
| 4 | charlie | Montreal | 7896054327 | 1 | 2025-03-12 14:55:37.387 | Dataflow | 2025-03-12 14:55:37.387 | 2885929265 | Dataflow |
| 5 | Robert | Sudbury | 4169793456 | 1 | 2025-03-12 14:55:37.387 | Dataflow | 2025-03-12 14:55:37.387 | 1431409828 | Dataflow |

Day 2 file is also deleted from the storage account

Upload Add Directory Refresh Rename Delete Change tier Acquire lease Break lease Give feedback

Authentication method: Access key (Switch to Microsoft Entra user account)

Location: ewd / Data_Migration_SCD_2

Search blobs by prefix (case-sensitive) Show deleted objects

| Name | Modified | Access tier | Archive status | Blob type | Size |
|------|----------|-------------|----------------|-----------|------|
| ... | | | | | |

