

# Data Migration Project

1. Create a Virtual Machine in AzureCloud, 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.

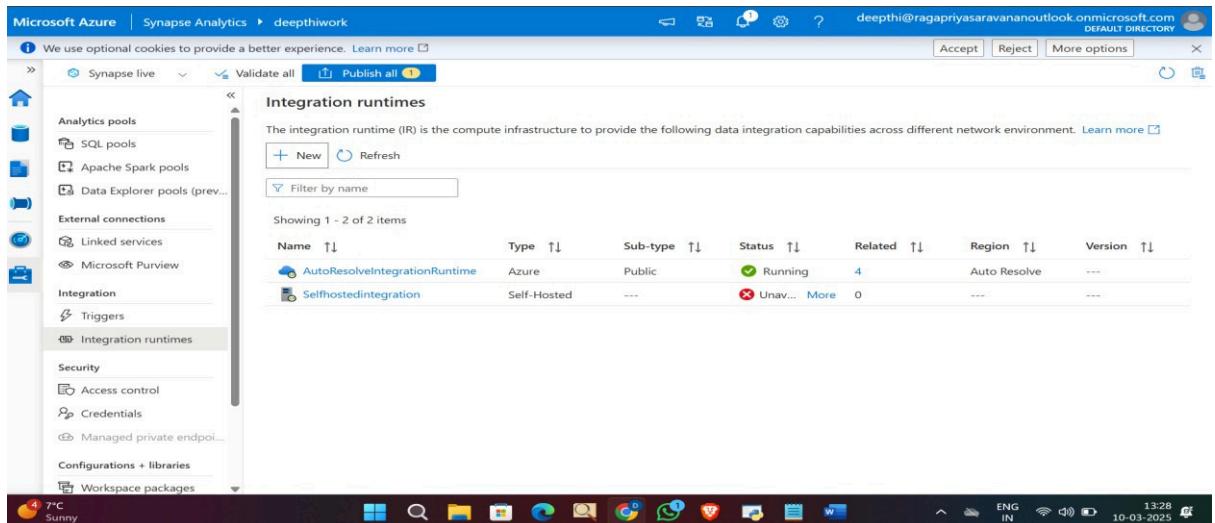
Then go to SQL Server Settings table and enable SQL authentication and review and create.

The screenshot shows the Microsoft Azure portal's deployment overview page. The deployment name is 'CreateVm-microsoftsqlserver.sql2022-ws2022-sqldev-20250310131226'. The status message says 'Your deployment is complete'. Deployment details table:

Resource	Type	Status	Operation details
deepthivm	Microsoft.SqlVirtualM...	OK	<a href="#">Operation details</a>
deepthivm	Microsoft.Compute/vir...	OK	<a href="#">Operation details</a>
deepthivm725	Microsoft.Network/net...	OK	<a href="#">Operation details</a>
deepthivm-nsg	Microsoft.Network/net...	OK	<a href="#">Operation details</a>
deepthivm-vnet	Microsoft.Network/virt...	OK	<a href="#">Operation details</a>
deepthivm-ip	Microsoft.Network/pu...	OK	<a href="#">Operation details</a>
deepthivm_DataDisk_0	Microsoft.Compute/di...	OK	<a href="#">Operation details</a>
deepthivm_DataDisk_1	Microsoft.Compute/di...	OK	<a href="#">Operation details</a>

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

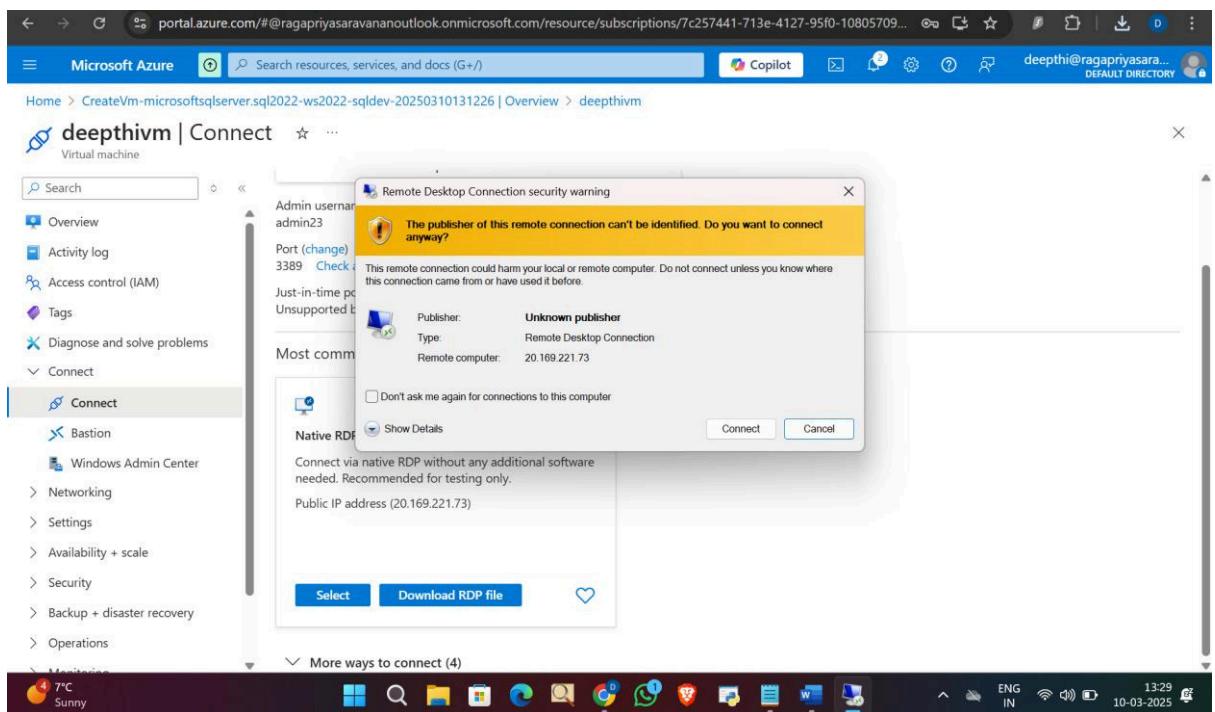
Synapse->Manage->Integration runtime->Create-> Selfhosted-> Give details-> Create->Copy the key



The screenshot shows the Microsoft Azure Synapse Analytics interface. The left sidebar has a tree view with 'Integration runtimes' selected. The main area is titled 'Integration runtimes' and displays a table of two items:

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 to an unknown publisher. The dialog includes fields for Publisher (Unknown publisher), Type (Remote Desktop Connection), and Remote computer (20.169.221.73). There are 'Select' and 'Download RDP file' buttons at the bottom.



Windows Security

X

## Enter your credentials

These credentials will be used to connect to 20.169.221.73.

admin23

Password



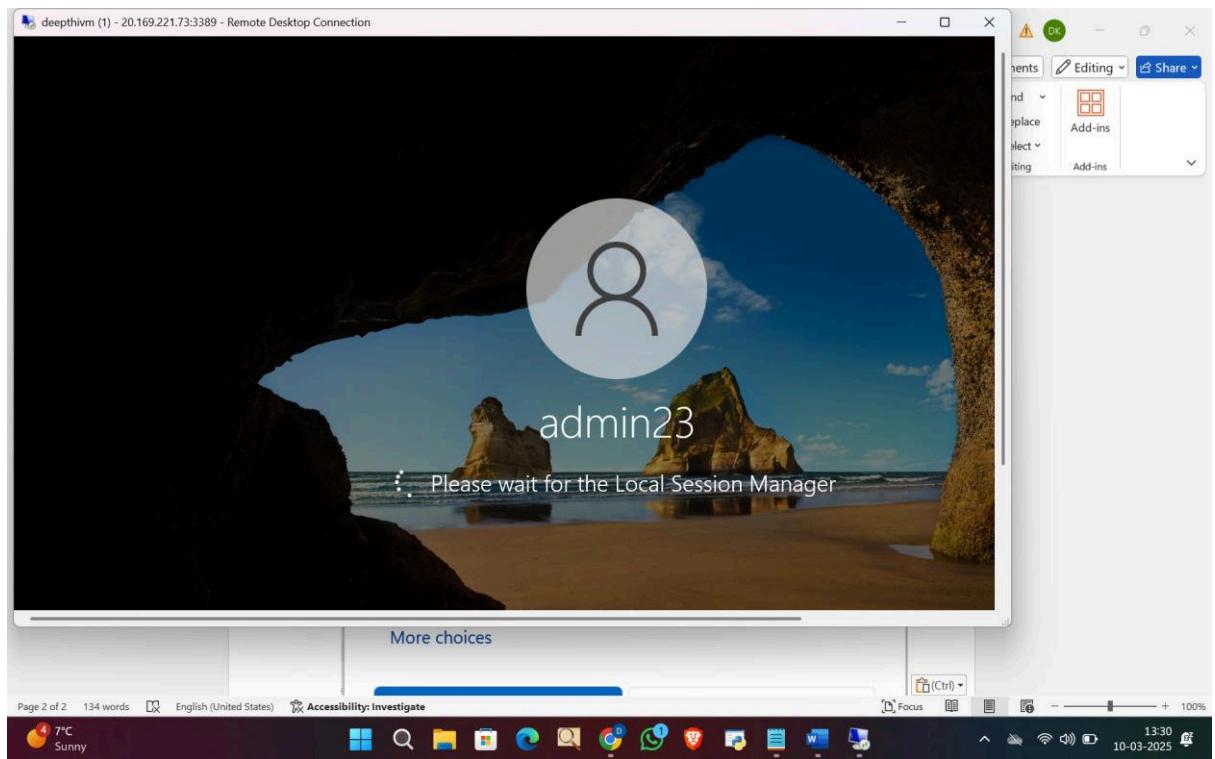
Remember me

[More choices](#)

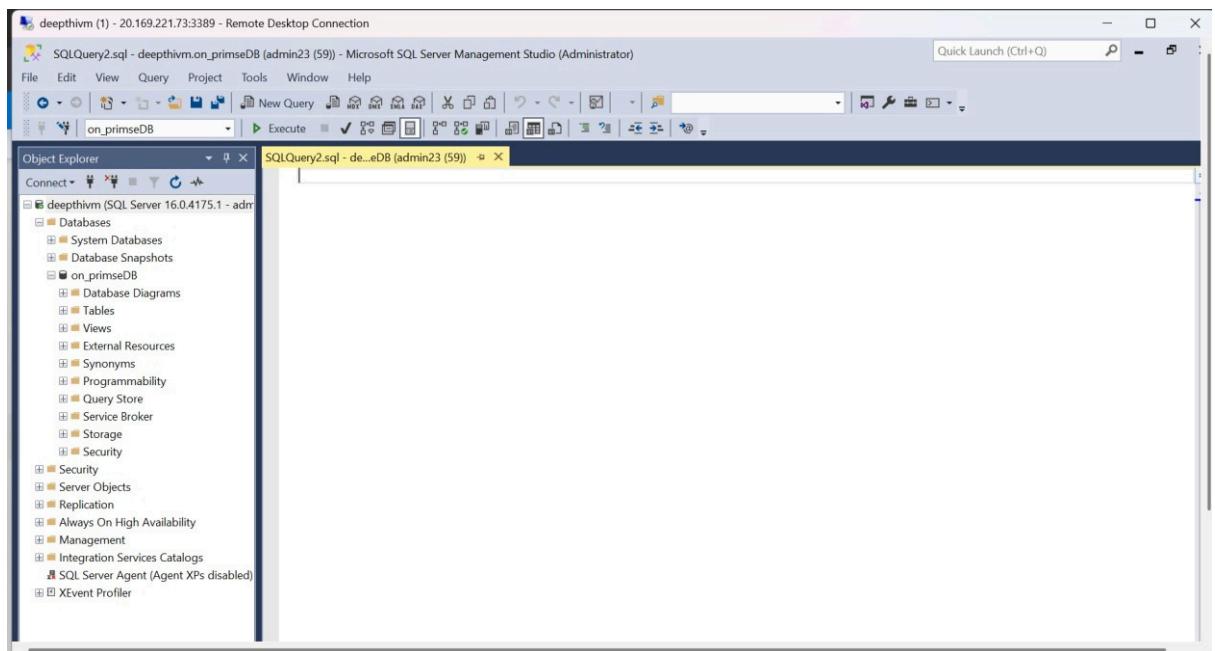
OK

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 results pane shows the message "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 a SQL query window with the same set of insert statements. The results pane displays five rows affected messages for each insert statement, followed by 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

```

The screenshot shows the SQL Query Editor window with the following code:

```

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

```

The code is highlighted with syntax coloring. Below the editor, the status bar shows "100 %". In the bottom left corner, there is a "Messages" pane with the message "Commands completed successfully." and the completion time "Completion time: 2025-03-11T15:53:09.7864831+00:00".

The screenshot shows the SQL Query Editor window with the following code:

```

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 )

The screenshot shows the SQL Query Editor window with the following code:

```

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

```

The code is highlighted with syntax coloring. Below the editor, the status bar shows "0 %". In the bottom left corner, there is a "Messages" pane with the message "Commands completed successfully." and the completion time "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 a SQL query window with three insert 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, the message pane shows three rows affected:

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

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_Phonenum bigint )
```

The screenshot shows a SQL query window with the create table statement for the Customer table:

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

Below the code, the message pane shows the command completed successfully:

Commands completed successfully.

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 a SQL query window with three insert statements into the Customer table:

```
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, the message pane shows three rows affected:

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

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: 2025-03-11T16:18:25.2446182+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

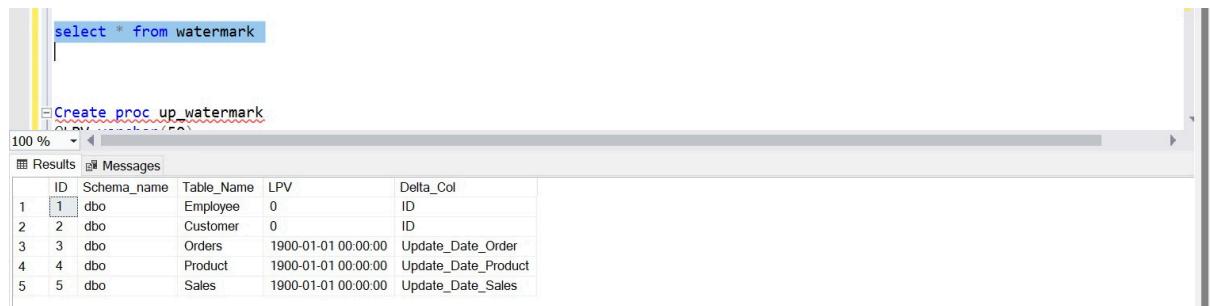
(1 row affected)

(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 window titled "Create\_proc\_up\_watermark" with the message "CREATE PROCEDURE [dbo].[up\_watermark]".

The bottom pane displays the results of the query in a table format:

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

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

SQL Server Integration Services

deepthi@ragapriyasaravananolook.onmicrosoft.com  
DEFAULT DIRECTORY

## 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

Parameters

+ New | Delete

Name	Type	Default value
Schem_name	String	Value
Table_name	String	Value

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.

**Pipeline expression builder**

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

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

**Activity outputs**

- 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 expression builder. A 'Lookup' activity is selected, and its 'Settings' tab is open. In the 'General' section, 'Schema\_name' and 'Table\_name' are specified. Under 'First row only', 'Use query' is selected, and a '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 interface.

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 expression builder. An 'If' condition activity is selected, and its 'Activities (0)' tab is open. The 'Expression' field contains the following expression:

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

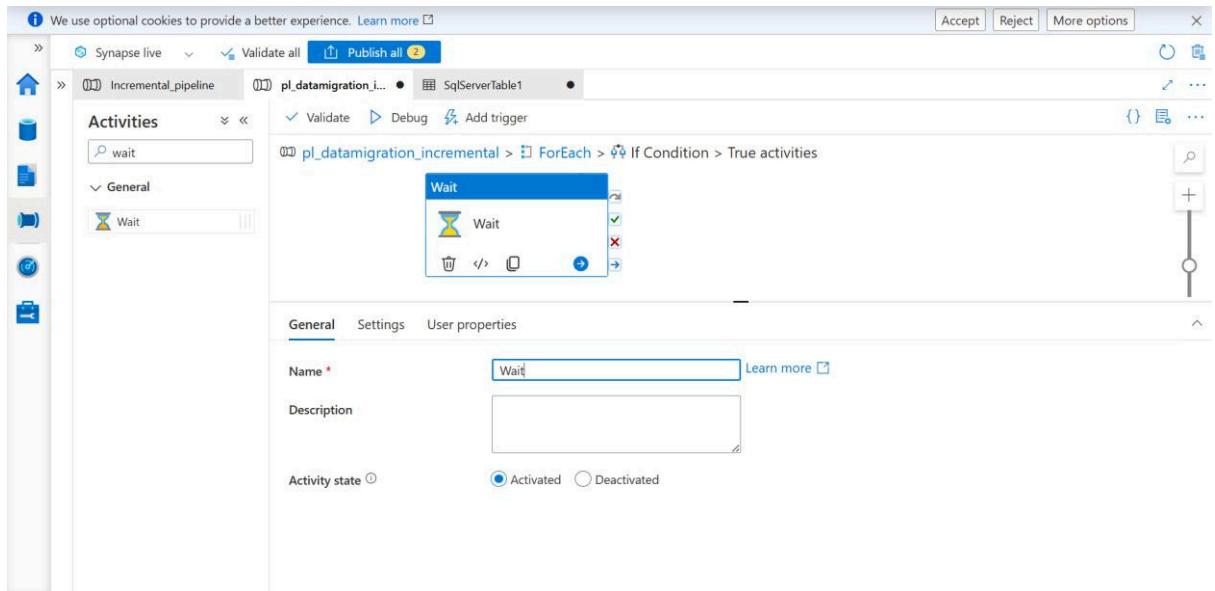
The 'Activity outputs' tab is selected in the pipeline expression builder interface.

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

Drag 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	
File_name	String	

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:

Microsoft Azure | Synapse Analytics > deepthiwork | Search

Pipeline expression builder

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

```
@item().Table_Name
```

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

For File name:

Microsoft Azure | Synapse Analytics > deepthiwork | Search

Pipeline expression builder

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

```
@concat(item().Table_Name,'_',utcNow(),'_.csv')
```

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

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\_j... SqlServerTable1

Activities Stored

Synapse SQL pool stored proc...

General Stored procedure

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

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

Pipeline expression builder

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

```
@item().Table_Name
```

Clear contents

Activity outputs Parameters System variables Functions Variables

Copy data activity output

Lookup

Lookup activity output

Lookup count

Count of the rows

Lookup value array

Array of row data

LookupMax

LookupMax activity output

OK Cancel

## Publish and run the pipeline

We use optional cookies to provide a better experience. [Learn more](#)

Accept Reject More options

Validate all Publish all

Activities Stored

Synapse SQL pool stored proc...

General Stored procedure

Parameters Variables Settings Output

Pipeline run ID: 3e1b7f7a-a817-4cbf-b2f0-7d092172d4a2

Pipeline status In progress

Monitor in Azure Metrics Export to CSV

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

We use optional cookies to provide a better experience. [Learn more](#)

Accept Reject More options

Expand Pipeline live Validate all Publish all

Activities Search activities

Synapse Move and transform Azure Data Explorer Azure Function Batch Service Databricks Data Lake Analytics General HDInsight Iteration & conditionals Machine Learning

SqlServerTable1 Incremental\_pipeline

Parameters Variables Settings Output

Activity	Status	Sub-Activity	Start Time	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),  
@TNAME nvarchar(100)

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 SQL script New data flow New integration dataset Upload Download + New folder Select all Copy link Rename More

ewd

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	
Employees	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

New SQL script New notebook, ...

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')

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 status page. The pipeline ID is ae554a47-2cee-4987-afca-9a611cc53a25. The status is Succeeded. The table below lists the activities and their 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')


```

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

Content Type: 128 B  
Size: 128 B  
65 B

OK

## 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

Overview

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

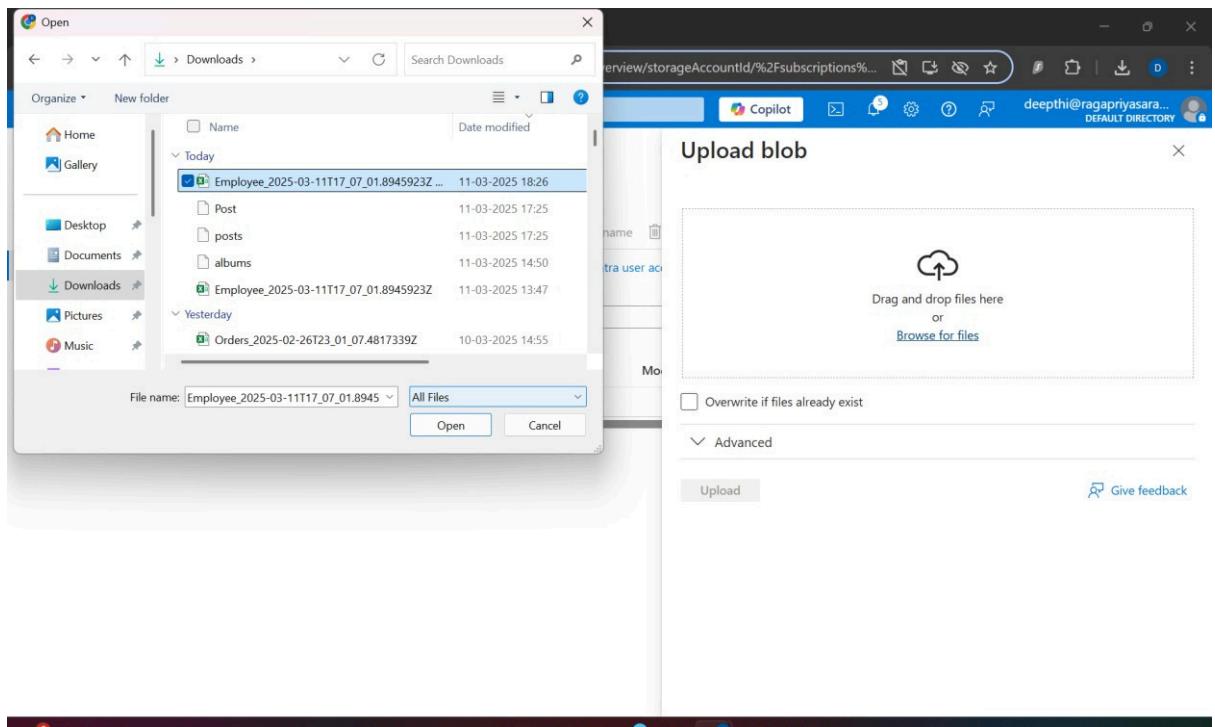
Location: ewd

Search blobs by prefix (case-sensitive)

Show deleted objects

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

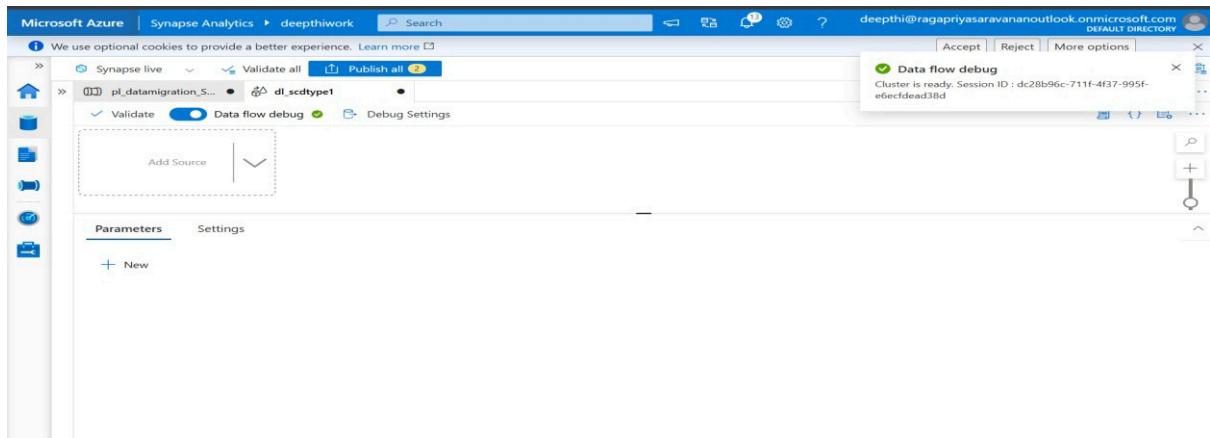


Uploaded the file

A screenshot of the Microsoft Azure Storage account overview page. The top navigation bar includes 'Microsoft Azure', a search bar, and user information. The main navigation shows 'Home &gt; Storage accounts &gt; adlsdeepthi | Containers &gt; ewd'. The 'Overview' tab is selected, displaying details like 'Authentication method: Access key (Switch to Microsoft Entra user account)' and 'Location: ewd / Data\_Migration\_SCD'. A search bar at the top allows searching by blob prefix. Below, a table lists blobs: one entry for '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).

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 )
```

100 %

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
<b>File settings</b> File mode <input checked="" type="radio"/> File <input type="radio"/> Wildcard File path * <input type="text" value="ewd"/> / <input type="text" value="Data_Migration_SCD"/> / <input type="text" value="Employee_2025-03..."/> <input type="button" value="Browse"/> Allow no files found <input type="checkbox"/> Change data capture <input type="checkbox"/> Compression type <input type="text" value="No compression"/> Encoding <input type="text" value="Default(UTF-8)"/> Column delimiter <input type="text" value="Comma (,)"/>					

Got to projection to import the Schema

The screenshot shows the Microsoft Azure Synapse Analytics Data Flow interface. A success message 'Successfully imported' is displayed at the top right. The main area shows a 'source' component with four columns: '\_col0\_', '\_col1\_', '\_col2\_', and '\_col3\_'. The 'Projection' tab is selected, and the 'Schema options' section shows each column as type 'string' with a dropdown menu for 'Format'.

## Data Preview

The 'Data preview' tab is active, showing a preview of three rows from the source. The columns are labeled ID, E\_Name, E\_City, and E\_PhoneNumber. The data is as follows:

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 'Dataflow expression builder' is open, showing a 'Rename' operation. The 'Matching condition' field contains the expression `1==1`. The 'Output column name expression' field contains the expression `concat('Src_', $$)`. The 'Expression elements' panel on the left lists various functions and input schema elements. At the bottom, there are 'Save and finish', 'Cancel', and 'Clear contents' buttons.

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

Microsoft Azure | Synapse Analytics > deepthiwork | Search | ? | deepthi@ragapriyasaravanonoutlook.onmicrosoft.com | DEFAULT DIRECTORY

### 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**: 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 ↴	Change the type of a column	lifted	Statistics	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 \***: AzureSQLDatabase1 [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 'Add lookup' configuration interface. At the top, there are tabs: 'Lookup settings' (selected), 'Optimize', 'Inspect', and 'Data preview'. Below the tabs, the 'Lookup stream' is set to 'Target'. Under 'Match multiple rows', there is a checkbox and a dropdown set to 'Any row'. In the 'Lookup conditions' section, the left column is 'Left: Hashcolumn's column' and the right column is 'Right: Target's column'. A condition is defined as '12s Src\_ID == 123 ID'. There are buttons for '+' and '-' to add or remove conditions.

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

## Insert expression or condition

The screenshot shows the 'Dataflow expression builder' for a 'split' condition. The main area displays the expression 'isNull(ID)'. The 'Expression elements' sidebar lists various options like All, Functions, Input schema, Parameters, Cached lookup, and Data flow library functions. The 'Expression values' sidebar lists variables such as 'Src\_Hashkey', 'ID', and 'E\_Hashkey'. At the bottom, there are buttons for 'Save and finish', 'Cancel', and 'Clear contents'.

## Updating condition

The screenshot shows the 'Dataflow expression builder' for an update condition. The main area displays the expression 'Src\_ID==ID && Src\_Hashkey != E\_Hashkey'. The 'Expression elements' sidebar lists various options like All, Functions, Input schema, Parameters, Cached lookup, and Data flow library functions. The 'Expression values' sidebar lists variables such as 'Src\_Hashkey', 'ID', 'E\_Hashkey', and 'abs(123 numeric\_value)'. At the bottom, there are buttons for 'Save and finish', 'Cancel', and 'Clear contents'.

Conditional split settings

Output stream name \*  Learn more [\[?\]](#)

Description  [Reset](#)

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 &amp;&amp; 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'"/> + <a href="#">Delete</a>
<input type="checkbox"/> Src_CreatedDate	<input type="text" value="currentTimestamp()"/> + <a href="#">Delete</a>
<input type="checkbox"/> Src_Updatedby	<input type="text" value="Dataflow'"/> + <a href="#">Delete</a>
<input type="checkbox"/> Src_Updateddate	<input type="text" value="currentTimestamp()"/> + <a href="#">Delete</a>

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"/>	123 ID
<input type="checkbox"/>	abc Src_E_Name	<input type="checkbox"/>	abc E_Name
<input type="checkbox"/>	abc Src_E_City	<input type="checkbox"/>	abc E_City
<input type="checkbox"/>	12l Src_E_PhoneNumber	<input type="checkbox"/>	12l E_PhoneNumber
<input type="checkbox"/>	12l Src_Hashkey	<input type="checkbox"/>	12l E_Hashkey
<input type="checkbox"/>	abc Src_CreatedBy	<input type="checkbox"/>	abc createdby
<input type="checkbox"/>	⌚ Src_CreatedDate	<input type="checkbox"/>	⌚ created_date
<input type="checkbox"/>	abc Src_UpdatedBy	<input type="checkbox"/>	abc updatedby
<input type="checkbox"/>	⌚ Src_UpdatedDate	<input type="checkbox"/>	⌚ 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 \*

Description

Incoming stream \*

Columns \*

<input type="checkbox"/>	Updatedby	'Dataflow'
<input type="checkbox"/>	Updateddate	currentTimestamp()

## Add Alter row to modify the record.

Alter row settings Optimize Inspect Data preview ●

Output stream name \*

Description

Incoming stream \*

Alter row conditions \*

## 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 ●

Options 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_updateddate

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	<span style="color: green;">Succeeded</span>	Data flow	3/11/2025, 7:27:58 PM	6m 59s	AutoResolveIntegrationRuntime (Canada Centr

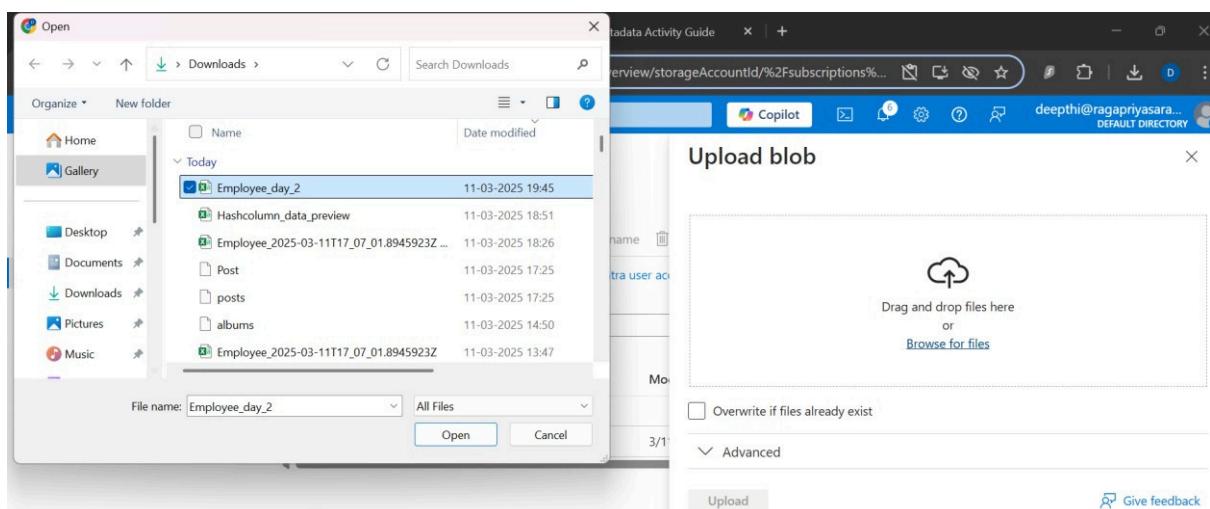
Data added to table

```
select * from DM_Employee
```

Results

ID	E_Name	E_City	E_Phonenumer	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="Delete"/> <input type="button" value="Run"/> <input type="button" value="Succeeded"/>	Succeeded	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 %

Results Messages

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

Source type: Integration dataset

Inline dataset type: DelimitedText

Linked service: AzureDataLakeStorage1

Skip line count: 0

Sampling: Disable

Source options: File mode: File, File path: ewd / Data\_Migration\_SCD / Employee\_2025-03..., Compression type: No compression

Go to projection and import schema

Column name	Type	Format
ID	short	Specify format
E_Name	string	Specify format
E_City	string	Specify format
E_PhoneNumber	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 text '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 'Name as' with the expression '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. A derived column named 'Src\_Hash' is being created. 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 builder interface includes a toolbar with operators like +, -, \*, /, ||, &&, !, ^, ==, ===, <=, !=, >, <, >=, <=, and [ ]. Below the expression is a sidebar for 'Expression elements' with sections for All, Functions, Input schema, Parameters, and Cached lookup. To the right is a 'Expression values' pane showing items like 'Src\_E\_City', 'Src\_E\_PhoneNumber', 'abs()', and 'acos()'. At the bottom are buttons for 'Data preview', '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 SSMS query editor with a single query in the 'SQLQuery1.sql' window:

```
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)
```

In the status bar at the bottom, it says 'Commands completed successfully.' and 'Completion time: 2025-03-11T20:17:56.9553663-04:00'.

## Add target which is Azure SQL database

The screenshot shows the 'Source options' tab selected. Under 'Input', the 'Query' radio button is selected, and the query text is set to 'select E\_ID,E\_Hashkey from DM\_Employee2 where isActive=1'. Other tabs like 'Table' and 'Stored procedure' are also visible.

## Go to projection and import schema

The screenshot shows the 'Projection' tab selected. Under 'Import schema', there are buttons for 'Import schema', 'Clear schema', 'Schema options', and 'Overwrite schema'. Below this, a table lists columns: 'E\_ID' with type 'integer' and 'E\_Hashkey' with type 'long'.

## Add lookup to do left join on source and target

The screenshot shows the 'Lookup settings' tab selected. It includes fields for 'Primary stream' (set to 'Hashkey'), 'Lookup stream' (set to 'Target'), and 'Match multiple rows' (unchecked). Under 'Match on', 'Any row' is selected. In the 'Lookup conditions' section, 'Left: Hashkey's column' is 'Src\_ID' and 'Right: Target's column' is 'E\_ID'. The preview pane at the bottom shows 3 rows inserted, 0 updated, 0 deleted, 0 upserted, 0 looked up, and 0 errors.

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

The screenshot shows the 'Conditional split settings' tab selected. It includes a 'Description' field with the note 'Conditionally distributing the data in E\_ID, Src\_ID, E\_ID, Src\_Hash, E\_Hashkey groups, based on columns '(1)'. The 'Incoming stream' is set to 'lookup'. Under 'Split on', 'First matching condition' is selected. The 'Split condition' table shows two rows: 'Insert' with condition 'isNull(E.ID)' and 'Update' with condition 'Src.ID==E.ID && Src.Hash!=E.Hashkey'.

## Add derived column activity to add columns to update

Derived column's settings

Output stream name \* derivedColumn1

Description Creating/updating the columns 'Src\_ID, Src\_E\_Name, Src\_E\_City, Src\_E\_PhoneNumber, Src\_Hash, E\_ID,'

Incoming stream \* split@Update

**Columns**

Column	Expression
Src_Updatedby	'Dataflow_updated'
Src_Updated_Date	currentTimestamp()
Src_isActive	0

### Add alter row to modify the records

Alter row settings

Output stream name \* alterRow

Description Add expressions to alter rows

Incoming stream \* derivedColumn1

Alter row conditions \* **Update if** 1=1

### Add Update Sink

Sink Settings Errors Mapping Optimize Inspect Data preview

Schema name \* dbo

Table name \* DM\_Employee2

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

123 E\_ID

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

Auto mapping  Add mapping  Delete  Reset Import schema View schema 5 mappings: 5 column(s) from the output schema left unmapped

Input columns	Output columns
123 E_ID	123 E_ID
123 Src_isActive	123 isActive
abc Src_Updatedby	abc updatedby
Src_Updated_Date	updated_date
121 E_Hashkey	121 E_Hashkey

## Add Union activity to union split insert and Split update

**Union settings**

Output stream name \*  Learn more

Description

Incoming stream \*

Union by \*  Name  Position

Union with \*

## Data preview

**Data preview**

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

↑↓	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 \*

**Columns \***  Open expression builder

Column	Expression
Src_createdby	'Dataflow'
Src_createddate	currentTimestamp()
Src_Updatedby	'Dataflow'
Src_Updateddate	currentTimestamp()
Src_IsActive	1

## Data Preview

**Data Preview**

↑↓	Src_ID	↑↓	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 \***

Inline dataset type \*

Linked service \*

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 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 sourceso that it will get file name from GetMeta Data.

Publish and run pipeline

Pipeline ran successfully

[Validate](#)   [Debug](#)   [Add trigger](#)    Data flow debug [✓](#)

[Parameters](#)   [Variables](#)   [Settings](#)   [Output](#)

**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	<a href="#">✓</a> Succeeded	Data flow	3/12/2025, 10:11:23 AM	9m 42s	AutoResolveIntegrationRuntime (Canada Centr)
Get Metadata1	<a href="#">✓</a> 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	1	Robert	Toronto	2499791376	1	2025-03-12 14:14:49.650	Dataflow	2025-03-12 14:14:49.650	1331461762	Dataflow
2	3	John	Montreal	2499793456	1	2025-03-12 14:14:49.650	Dataflow	2025-03-12 14:14:49.650	21035172	Dataflow
3	2	Ann	Brampton	2499799087	1	2025-03-12 14:14:49.650	Dataflow	2025-03-12 14:14:49.650	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 settings <sup>1</sup> User properties

**Dataset \***: Binary2

**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**:

Create a parameter and connect it

Connection Parameters

**Linked service \***: AzureDataLakeStorage1

**Integration runtime \***: AutoResolveIntegrationRuntime

**File path**: ewd / Data\_Migration\_SCD\_2 / @dataset().filename

**Compression type**: No compression

Giving path to delete the file

Microsoft Azure | Synapse Analytics > deepthiwork | Search

Pipeline expression builder

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

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

**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 Run History page. At the top, there are tabs for Validate, Debug, Add trigger, and Data flow debug. Below the tabs is a pipeline diagram with three activities: Get Metadata, Data flow, and Delete. Each activity has a green checkmark indicating success. The pipeline status is listed as Succeeded. Below the pipeline details, a table lists the activities and their run history:

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 / Data\_Migration\_SCD\_2'. A specific file, 'Employee\_day\_2.csv', is shown with a yellow warning icon, indicating it has been deleted. The interface includes standard navigation and management tools like Upload, Refresh, and Delete.

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. The dialog title is 'Upload blob' and it shows '1 file(s) selected: Employee\_day\_2.csv'. The file is being uploaded to the 'ewd / Data\_Migration\_SCD\_2' container. The upload progress bar is visible at the bottom of the dialog.

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	<span style="color: green;">✓ Succeeded</span>	Delete	3/12/2025, 11:02:00 AM	10s	AutoResolveIntegrationRuntime (Canada Ce)
Data flow	<span style="color: green;">✓ Succeeded</span>	Data flow	3/12/2025, 10:52:26 AM	9m 33s	AutoResolveIntegrationRuntime (Canada Ce)
Get Metadata1	<span style="color: green;">✓ Succeeded</span>	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	2499799087	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
...					