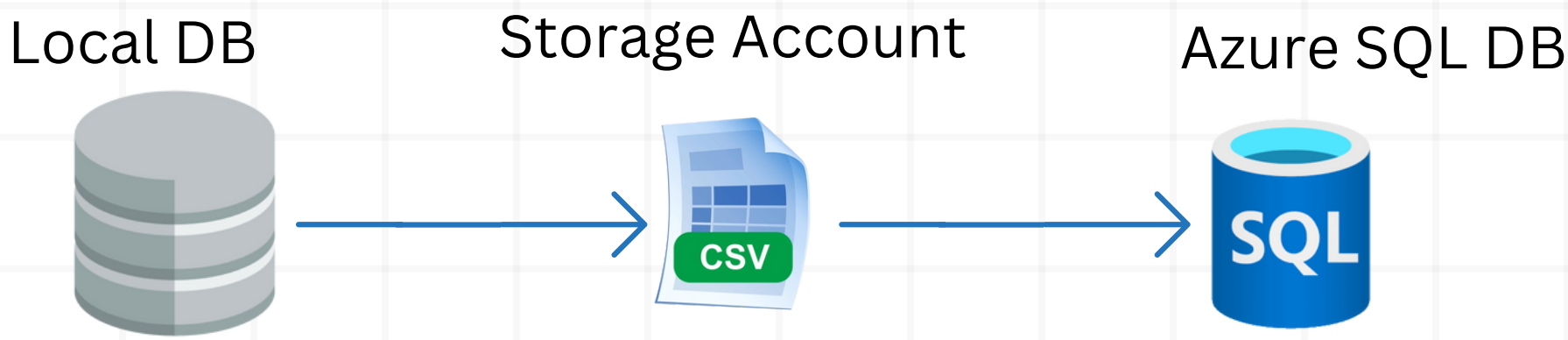



# Data Migration Project

Incremental Data, File Backup, SCD Type 1 and Type 2 Implementation in Azure Synapse Workspace

## Data Migration in Azure

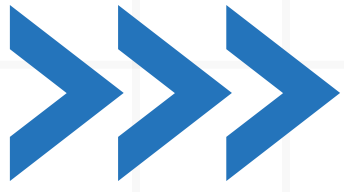
The process of moving data from one storage system to another within the Azure cloud environment, optimizing for accessibility, scalability, and performance.

## Why do we need this concept?

To efficiently manage and scale resources, ensuring data is accessible and optimized for cloud-based applications.

## What are SCD and why are they important in data warehousing?

SCDs are methods used in data warehousing to manage and track changes to record data over time, enabling historical data analysis and accurate reporting.



## Best practices to follow for this data migration project

Key practices included thorough initial planning, continuous monitoring, using staging environments for tests, and incremental loading to minimize system impact.

## Want to see the step by step guide?

Let’s go!

# Data Migration and SCD Implementation in Azure SQL Database

Create 5 source tables and one Watermark table in the **on-premises** database as shown below:

```
----------SCD Type 1--------------
CREATE TABLE Suppliers (
    supplier_id INT,
    name VARCHAR(255),
    contact_name VARCHAR(255),
    phone VARCHAR(50),
    address Varchar(100),
    supplierUpdatedDate datetime -------------Delta Column
);

CREATE TABLE Products (
    product_id INT,
    name VARCHAR(255),
    category VARCHAR(100),
    price DECIMAL(10, 2),
    stock INT,
    supplier_id INT,
    productUpdatedDate datetime ------------Delta Column
);

----------SCD Type 2--------------
CREATE TABLE Employee (
    employee_id INT,
    first_name VARCHAR(255),
    last_name VARCHAR(255),
    hire_date DATETIME,
    last_review_date DATETIME,  ----------------Delta Column
    role VARCHAR(100)
);

CREATE TABLE Sales (
    sale_id INT,
    sale_date DATETIME,
    product_id INT,
    quantity INT,
    total_amount DECIMAL(10, 2),
    cashier_id INT,
    salesUpdatedDate datetime  -------------------Delta Column
);

CREATE TABLE Inventory_Logs (
    log_id INT,
    product_id INT,
    log_updatedDate DATETIME,    ---------------------Delta Column
    change_quantity INT,
    remaining_stock INT
);
```

----------SCD Type 1--------------
CREATE TABLE Suppliers (
  supplier_id INT,

```
  name VARCHAR(255),
  contact_name VARCHAR(255),
  phone VARCHAR(50),
  address Varchar(100),
        supplierUpdatedDate datetime ----------Delta Column
);

CREATE TABLE Products (
  product_id INT,
  name VARCHAR(255),
  category VARCHAR(100),
  price DECIMAL(10, 2),
  stock INT,
  supplier_id INT,
        productUpdatedDate datetime ----------Delta Column
);

----------SCD Type 2----------
CREATE TABLE Employee (
  employee_id INT,
  first_name VARCHAR(255),
  last_name VARCHAR(255),
  hire_date DATETIME,
  last_review_date DATETIME,  ----------Delta Column
  role VARCHAR(100)
);

CREATE TABLE Sales (
  sale_id INT,
  sale_date DATETIME,
  product_id INT,
  quantity INT,
  total_amount DECIMAL(10, 2),
  cashier_id INT,
        salesUpdatedDate datetime  ----------Delta Column
);

CREATE TABLE Inventory_Logs (
  log_id INT,
  product_id INT,
  log_updatedDate DATETIME,   ----------Delta Column
  change_quantity INT,
  remaining_stock INT
);
```

**Insert records into these source tables**

```
----Insert data into Supplier Table
INSERT INTO Suppliers (supplier_id, name, contact_name, phone, address, supplierUpdatedDate) VALUES
(1, 'Fresh Farms', 'John Doe', '555-3489', '123 Farm Lane', '2023-11-11 00:00:00'),
(2, 'Healthy Beverages Co.', 'Emily Stone', '555-7623', '47 Beverage Blvd','2023-11-11 00:00:00'),
(3, 'Premium Meats', 'Alan Smith', '555-9876', '233 Meat St', '2023-11-11 00:00:00');

----Insert data into Products Table
INSERT INTO Products (product_id, name, category, price, stock, supplier_id, productUpdatedDate) VALUES
(1, 'Organic Apples', 'Fruits', 2.99, 150, 1, '2023-09-25 00:00:00'),
(2, 'Almond Milk', 'Beverages', 3.49, 85, 2, '2023-09-24 00:00:00'),
(3, 'Chicken Breast', 'Meat', 7.99, 60, 3, '2023-09-23 00:00:00');

----Insert data into Employee Table
INSERT INTO Employee (employee_id, first_name, last_name, hire_date, last_review_date, role) VALUES
(1, 'Raj', 'Sharma', '2022-01-05 09:00:00', '2023-09-10 00:00:00', 'Cashier'),
(2, 'Harpal', 'Vaghela', '2022-05-15 09:00:00', '2023-09-20 00:00:00', 'Cashier'),
(3, 'Amit', 'Singh', '2023-03-23 09:00:00', '2023-09-30 00:00:00', 'Stock Manager');

----Insert data into Sales Table
INSERT INTO Sales (sale_id, sale_date, product_id, quantity, total_amount, cashier_id, salesUpdatedDate) VALUES
(1, '2023-10-01 14:00:00', 1, 10, 29.90, 1,'2023-10-01 14:02:00'),
(2, '2023-10-01 14:15:00', 2, 5, 17.45, 2, '2023-10-01 14:17:00'),
(3, '2023-10-01 15:00:00', 3, 4, 31.96, 1, '2023-10-01 15:02:00');

----Insert data into Inventory_Logs Table
INSERT INTO Inventory_Logs (log_id, product_id, log_updatedDate, change_quantity, remaining_stock) VALUES
(1, 1, '2023-10-01 08:00:00', 20, 170),
(2, 2, '2023-10-01 09:00:00', -10, 75),
(3, 3, '2023-10-01 10:00:00', 30, 90);
```

----Insert data into Supplier Table
INSERT INTO Suppliers (supplier_id, name, contact_name, phone, address, supplierUpdatedDate) VALUES
(1, 'Fresh Farms', 'John Doe', '555-3489', '123 Farm Lane', '2023-11-11 00:00:00'),
(2, 'Healthy Beverages Co.', 'Emily Stone', '555-7623', '47 Beverage Blvd','2023-11-11 00:00:00'),
(3, 'Premium Meats', 'Alan Smith', '555-9876', '233 Meat St', '2023-11-11 00:00:00');

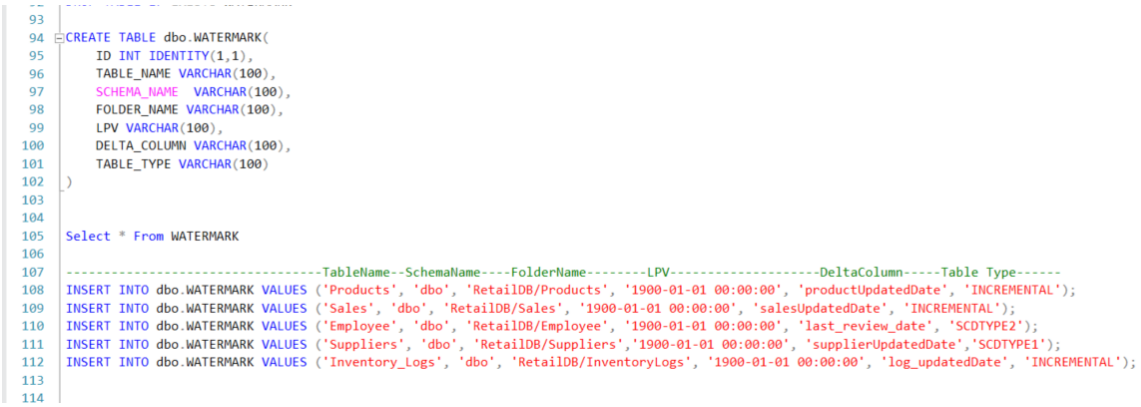
----Insert data into Products Table
INSERT INTO Products (product_id, name, category, price, stock, supplier_id, productUpdatedDate) VALUES
(1, 'Organic Apples', 'Fruits', 2.99, 150, 1, '2023-09-25 00:00:00'),
(2, 'Almond Milk', 'Beverages', 3.49, 85, 2, '2023-09-24 00:00:00'),
(3, 'Chicken Breast', 'Meat', 7.99, 60, 3, '2023-09-23 00:00:00');

----Insert data into Employee Table
INSERT INTO Employee (employee_id, first_name, last_name, hire_date, last_review_date, role) VALUES
(1, 'Raj', 'Sharma', '2022-01-05 09:00:00', '2023-09-10 00:00:00', 'Cashier'),
(2, 'Harpal', 'Vaghela', '2022-05-15 09:00:00', '2023-09-20 00:00:00', 'Cashier'),
(3, 'Amit', 'Singh', '2023-03-23 09:00:00', '2023-09-30 00:00:00', 'Stock Manager');

----Insert data into Sales Table
INSERT INTO Sales (sale_id, sale_date, product_id, quantity, total_amount, cashier_id, salesUpdatedDate) VALUES
(1, '2023-10-01 14:00:00', 1, 10, 29.90, 1,'2023-10-01 14:02:00'),
(2, '2023-10-01 14:15:00', 2, 5, 17.45, 2, '2023-10-01 14:17:00'),
(3, '2023-10-01 15:00:00', 3, 4, 31.96, 1, '2023-10-01 15:02:00');

----Insert data into Inventory_Logs Table
INSERT INTO Inventory_Logs (log_id, product_id, log_updatedDate, change_quantity, remaining_stock) VALUES
(1, 1, '2023-10-01 08:00:00', 20, 170),
(2, 2, '2023-10-01 09:00:00', -10, 75),
(3, 3, '2023-10-01 10:00:00', 30, 90);

**Create Watermark table and insert data**



```
CREATE TABLE dbo.WATERMARK(
    ID INT IDENTITY(1,1),
    TABLE_NAME VARCHAR(100),
    SCHEMA_NAME  VARCHAR(100),
    FOLDER_NAME VARCHAR(100),
    LPV VARCHAR(100),
    DELTA_COLUMN VARCHAR(100),
    TABLE_TYPE VARCHAR(100)
)


Select * From WATERMARK

---------------------------------TableName--SchemaName----FolderName--------LPV--------------------DeltaColumn-----Table Type------
INSERT INTO dbo.WATERMARK VALUES ('Products', 'dbo', 'RetailDB/Products', '1900-01-01 00:00:00', 'productUpdatedDate', 'INCREMENTAL');
INSERT INTO dbo.WATERMARK VALUES ('Sales', 'dbo', 'RetailDB/Sales', '1900-01-01 00:00:00', 'salesUpdatedDate', 'INCREMENTAL');
INSERT INTO dbo.WATERMARK VALUES ('Employee', 'dbo', 'RetailDB/Employee', '1900-01-01 00:00:00', 'last_review_date', 'SCDTYPE2');
INSERT INTO dbo.WATERMARK VALUES ('Suppliers', 'dbo', 'RetailDB/Suppliers','1900-01-01 00:00:00', 'supplierUpdatedDate','SCDTYPE1');
INSERT INTO dbo.WATERMARK VALUES ('Inventory_Logs', 'dbo', 'RetailDB/InventoryLogs', '1900-01-01 00:00:00', 'log_updatedDate', 'INCREMENTAL');
```

CREATE TABLE dbo.WATERMARK(
    ID INT IDENTITY(1,1),
    TABLE_NAME VARCHAR(100),
    SCHEMA_NAME  VARCHAR(100),
    FOLDER_NAME VARCHAR(100),
    LPV VARCHAR(100),
    DELTA_COLUMN VARCHAR(100),
    TABLE_TYPE VARCHAR(100)
)

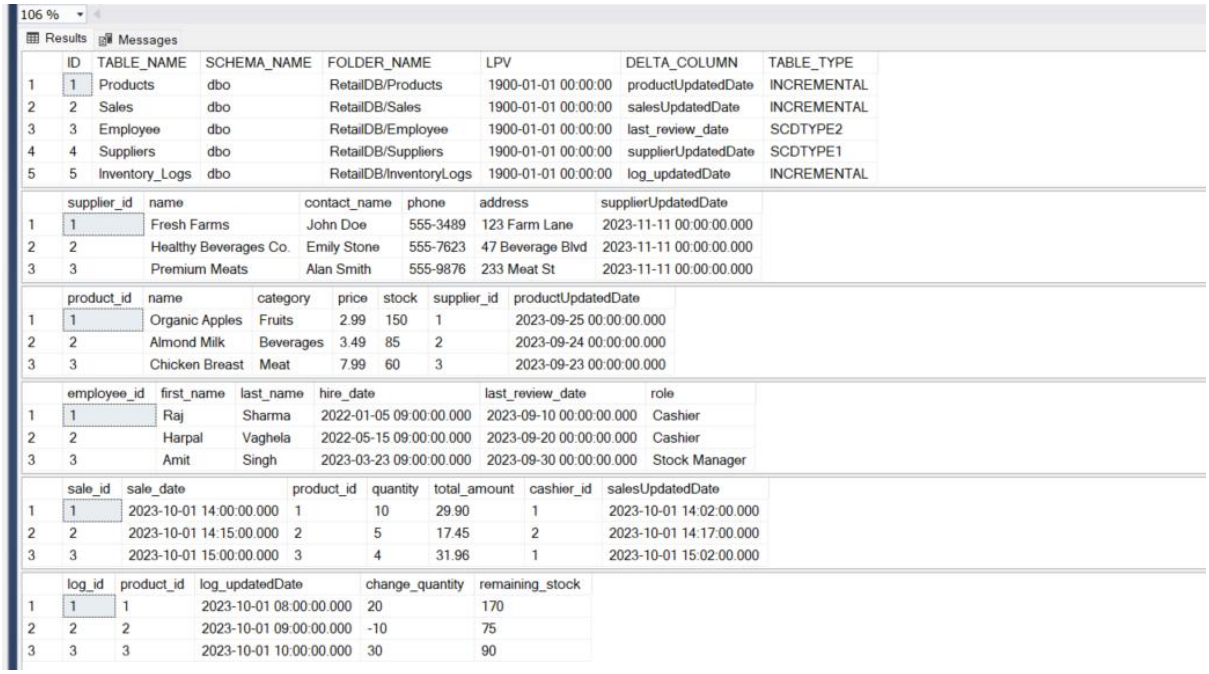
INSERT INTO dbo.WATERMARK VALUES ('Products', 'dbo', 'RetailDB/Products', '1900-01-01 00:00:00', 'productUpdatedDate', 'INCREMENTAL');
INSERT INTO dbo.WATERMARK VALUES ('Sales', 'dbo', 'RetailDB/Sales', '1900-01-01 00:00:00', 'salesUpdatedDate', 'INCREMENTAL');
INSERT INTO dbo.WATERMARK VALUES ('Employee', 'dbo', 'RetailDB/Employee', '1900-01-01 00:00:00', 'last_review_date', 'SCDTYPE2');
INSERT INTO dbo.WATERMARK VALUES ('Suppliers', 'dbo', 'RetailDB/Suppliers','1900-01-01 00:00:00', 'supplierUpdatedDate','SCDTYPE1');
INSERT INTO dbo.WATERMARK VALUES ('Inventory_Logs', 'dbo', 'RetailDB/InventoryLogs', '1900-01-01 00:00:00', 'log_updatedDate', 'INCREMENTAL');

### Create a Stored Procedure to Update LPV Value

```
114
115  ----------Stored Procedure----------
116
117  CREATE PROC USP_WATERMARK_UPDATE
118  @Table_Name VARCHAR(100),
119  @LPV_Value VARCHAR(50)
120  AS
121  UPDATE WATERMARK
122  SET LPV = @LPV_Value
123  WHERE TABLE_NAME = @Table_Name
124
125
```

```
CREATE PROC USP_WATERMARK_UPDATE
@Table_Name VARCHAR(100),
@LPV_Value VARCHAR(50)
AS
UPDATE WATERMARK
SET LPV = @LPV_Value
WHERE TABLE_NAME = @Table_Name
```

### Check the data in all tables



Results | Messages

| ID | TABLE_NAME | SCHEMA_NAME | FOLDER_NAME | LPV | DELTA_COLUMN | TABLE_TYPE |
|---|---|---|---|---|---|---|
| 1 | Products | dbo | RetailDB/Products | 1900-01-01 00:00:00 | productUpdatedDate | INCREMENTAL |
| 2 | Sales | dbo | RetailDB/Sales | 1900-01-01 00:00:00 | salesUpdatedDate | INCREMENTAL |
| 3 | Employee | dbo | RetailDB/Employee | 1900-01-01 00:00:00 | last_review_date | SCDTYPE2 |
| 4 | Suppliers | dbo | RetailDB/Suppliers | 1900-01-01 00:00:00 | supplierUpdatedDate | SCDTYPE1 |
| 5 | Inventory_Logs | dbo | RetailDB/InventoryLogs | 1900-01-01 00:00:00 | log_updatedDate | INCREMENTAL |

| supplier_id | name | contact_name | phone | address | supplierUpdatedDate |
|---|---|---|---|---|---|
| 1 | Fresh Farms | John Doe | 555-3489 | 123 Farm Lane | 2023-11-11 00:00:00.000 |
| 2 | Healthy Beverages Co. | Emily Stone | 555-7623 | 47 Beverage Blvd | 2023-11-11 00:00:00.000 |
| 3 | Premium Meats | Alan Smith | 555-9876 | 233 Meat St | 2023-11-11 00:00:00.000 |

| product_id | name | category | price | stock | supplier_id | productUpdatedDate |
|---|---|---|---|---|---|---|
| 1 | Organic Apples | Fruits | 2.99 | 150 | 1 | 2023-09-25 00:00:00.000 |
| 2 | Almond Milk | Beverages | 3.49 | 85 | 2 | 2023-09-24 00:00:00.000 |
| 3 | Chicken Breast | Meat | 7.99 | 60 | 3 | 2023-09-23 00:00:00.000 |

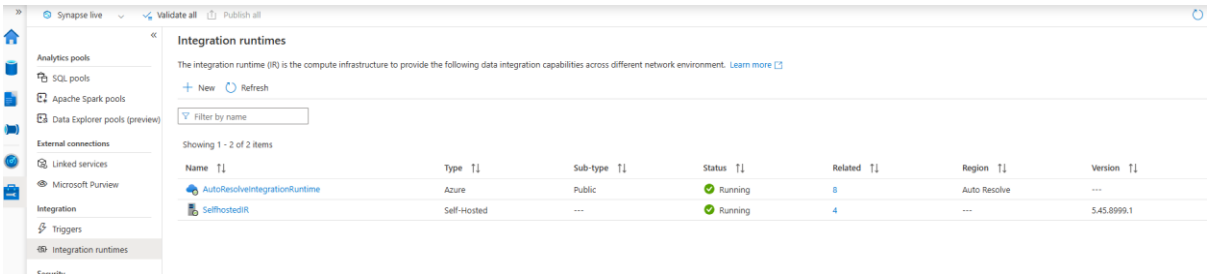
| employee_id | first_name | last_name | hire_date | last_review_date | role |
|---|---|---|---|---|---|
| 1 | Raj | Sharma | 2022-01-05 09:00:00.000 | 2023-09-10 00:00:00.000 | Cashier |
| 2 | Harpal | Vaghela | 2022-05-15 09:00:00.000 | 2023-09-20 00:00:00.000 | Cashier |
| 3 | Amit | Singh | 2023-03-23 09:00:00.000 | 2023-09-30 00:00:00.000 | Stock Manager |

| sale_id | sale_date | product_id | quantity | total_amount | cashier_id | salesUpdatedDate |
|---|---|---|---|---|---|---|
| 1 | 2023-10-01 14:00:00.000 | 1 | 10 | 29.90 | 1 | 2023-10-01 14:02:00.000 |
| 2 | 2023-10-01 14:15:00.000 | 2 | 5 | 17.45 | 2 | 2023-10-01 14:17:00.000 |
| 3 | 2023-10-01 15:00:00.000 | 3 | 4 | 31.96 | 1 | 2023-10-01 15:02:00.000 |

| log_id | product_id | log_updatedDate | change_quantity | remaining_stock |
|---|---|---|---|---|
| 1 | 1 | 2023-10-01 08:00:00.000 | 20 | 170 |
| 2 | 2 | 2023-10-01 09:00:00.000 | -10 | 75 |
| 3 | 3 | 2023-10-01 10:00:00.000 | 30 | 90 |

```
Select * From WATERMARK
Select * From Suppliers
Select * From Products
Select * From Employee
Select * From Sales
Select * From Inventory_Logs
```

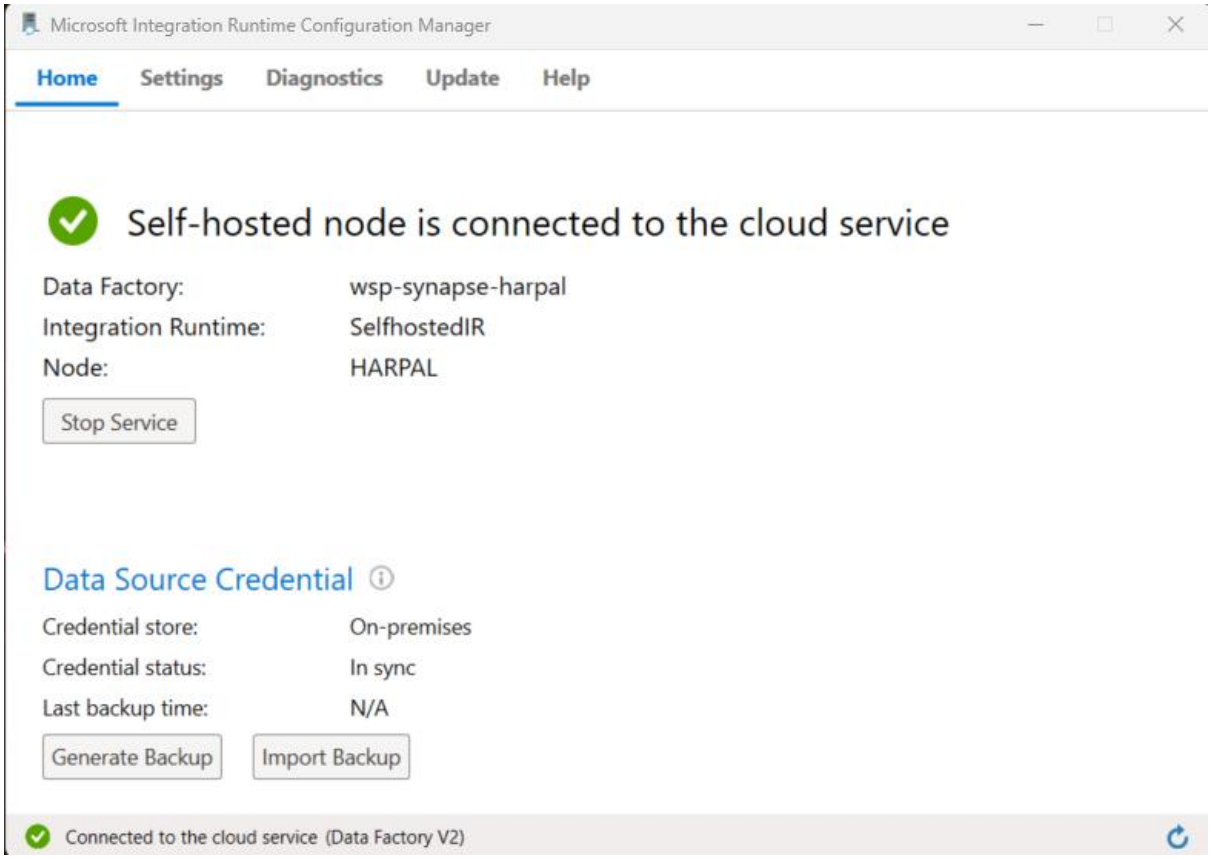
## Create Self-Hosted IR in the Manage tab in Synapse

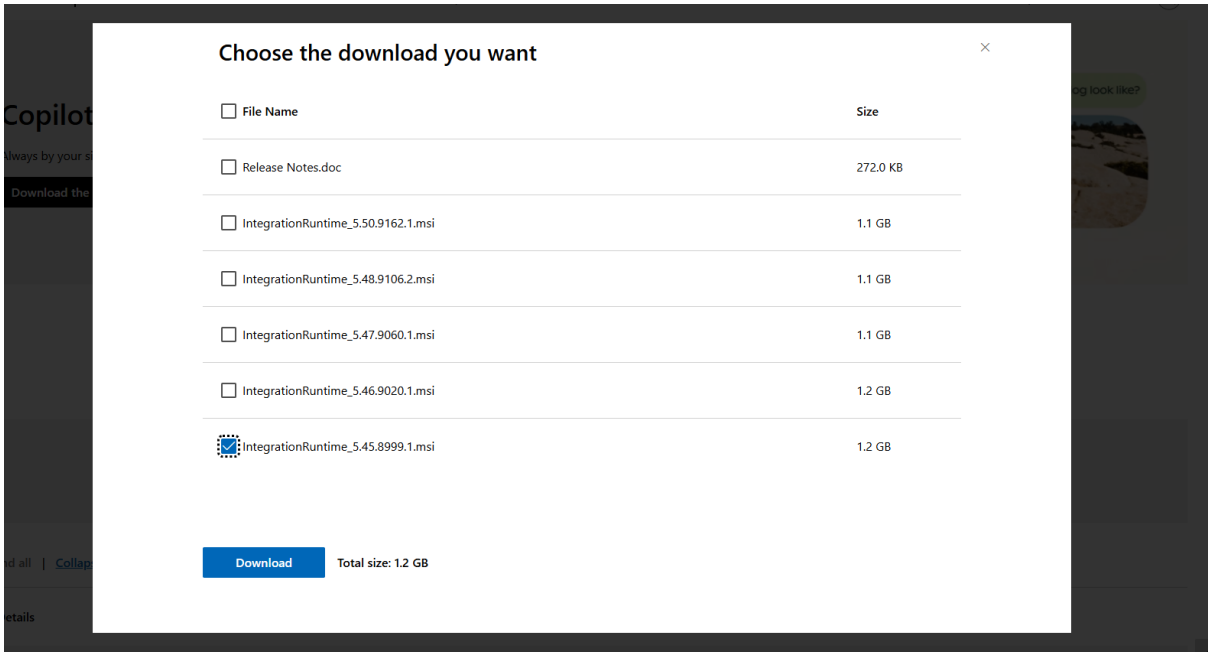



Install Integration Runtime software from the Microsoft Website and configure it to support Self-Hosted IR, copy and paste Key from synapse Self-Hosted IR in Microsoft Integration Runtime Configuration Manager

Website to download Integration Runtime:

https://www.microsoft.com/en-us/download/details.aspx?id=39717

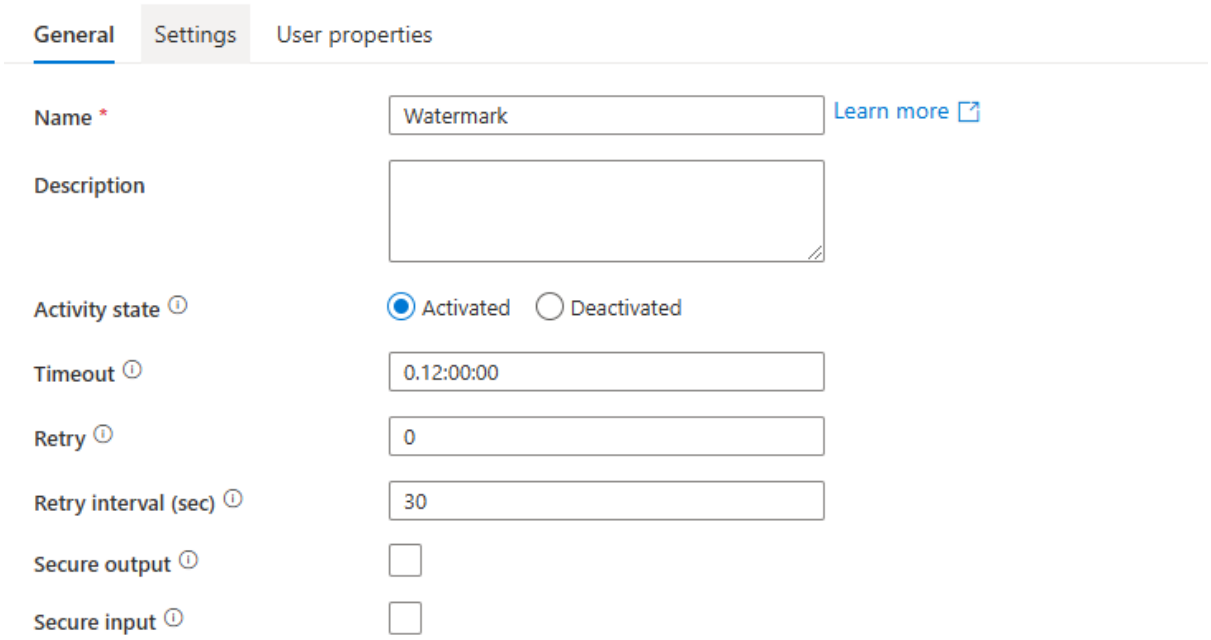





**Let's create the pipeline**

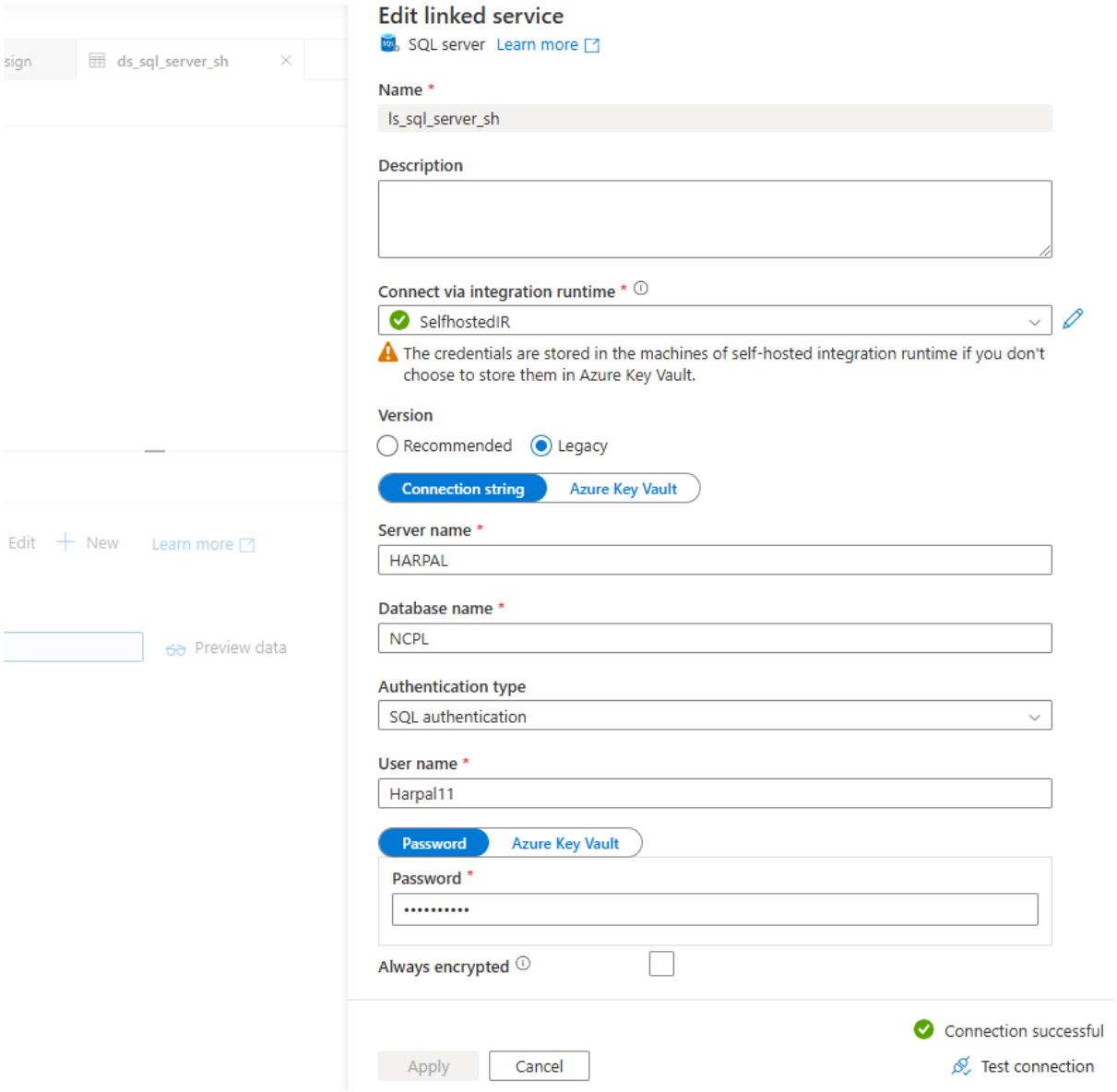
Azure Home -> Synapse Workspace -> New Pipeline

Drag and drop Lookup activity, rename as "Watermark"

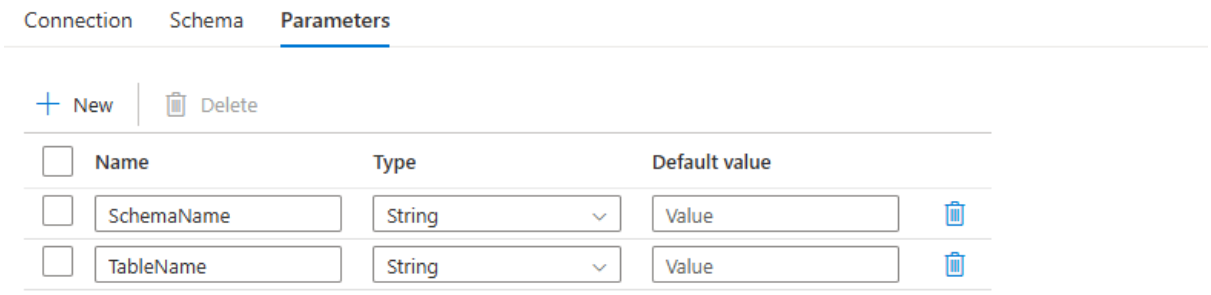



Create new Source Dataset as SQL Server, as we are connecting with on-premises database.

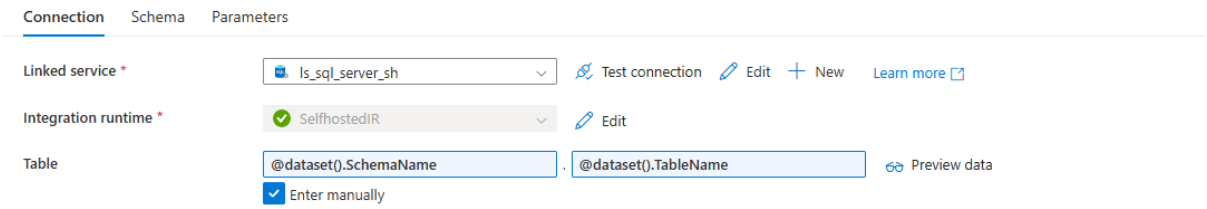
Select the linked service, make sure to select Self Hosted IR.



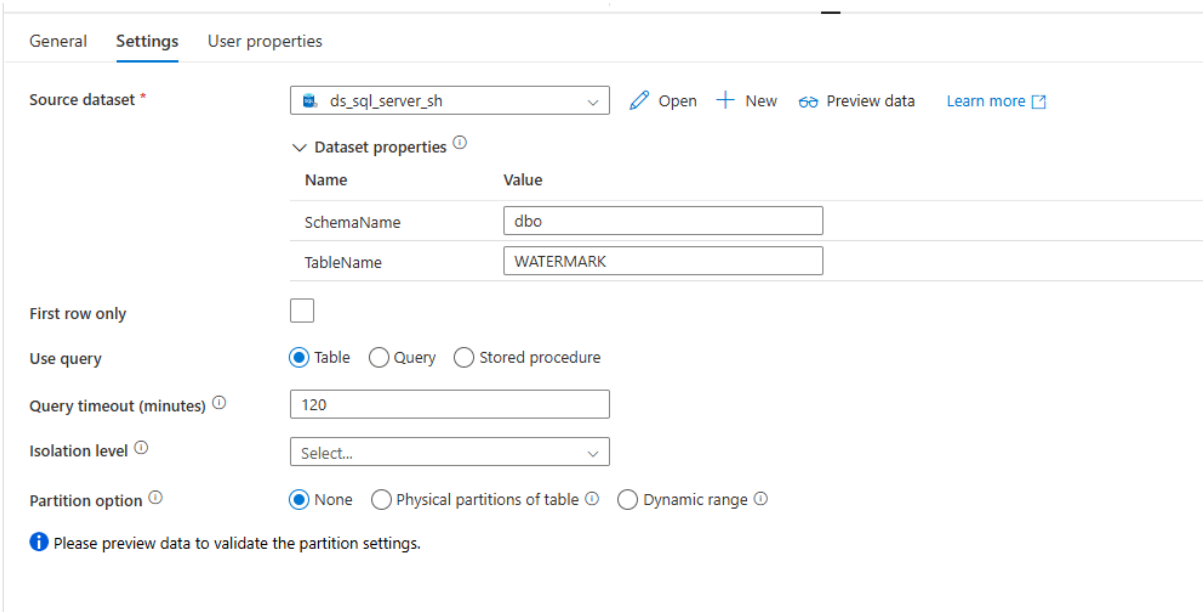

Take two parameters:



Go to the Connection tab and assign those parameters

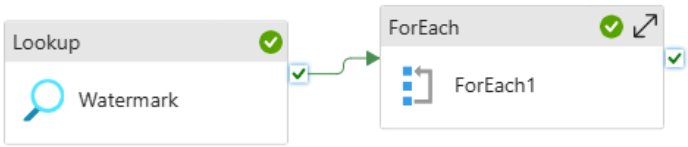


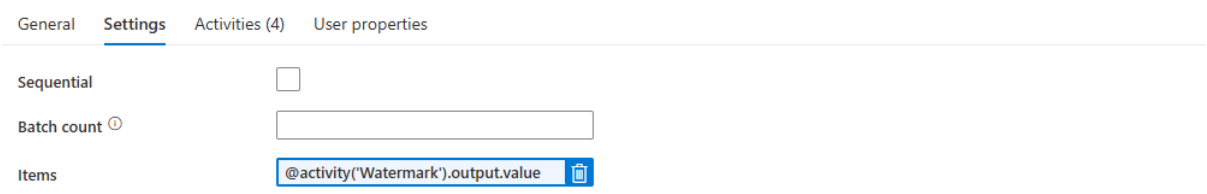

In Settings tab, mention the schema name and table name as shown below




Take ForEach Activity and write this expression in settings tab -> Item

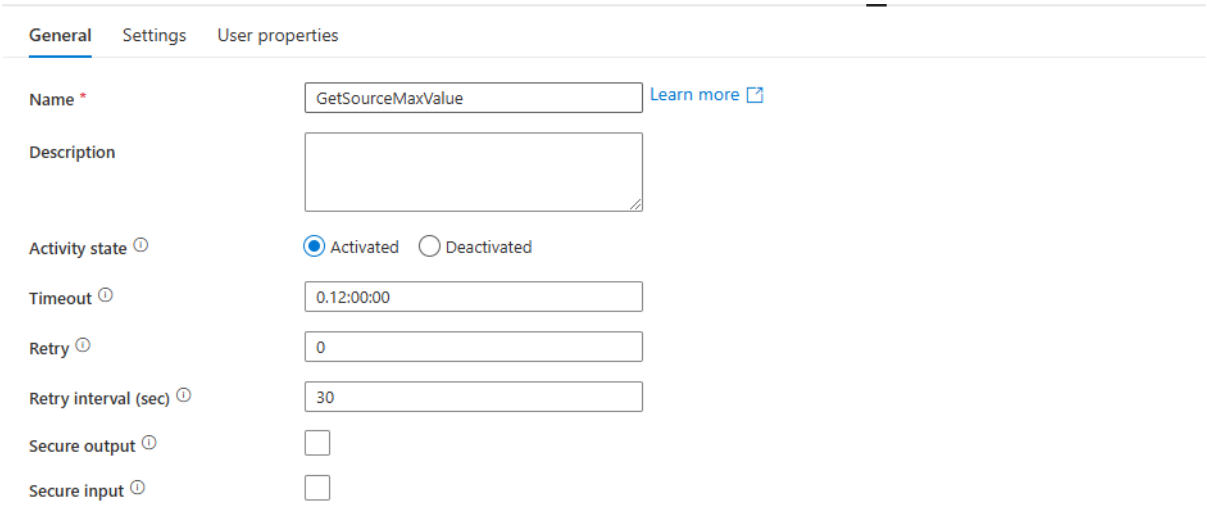
**@activity('Watermark').output.value**



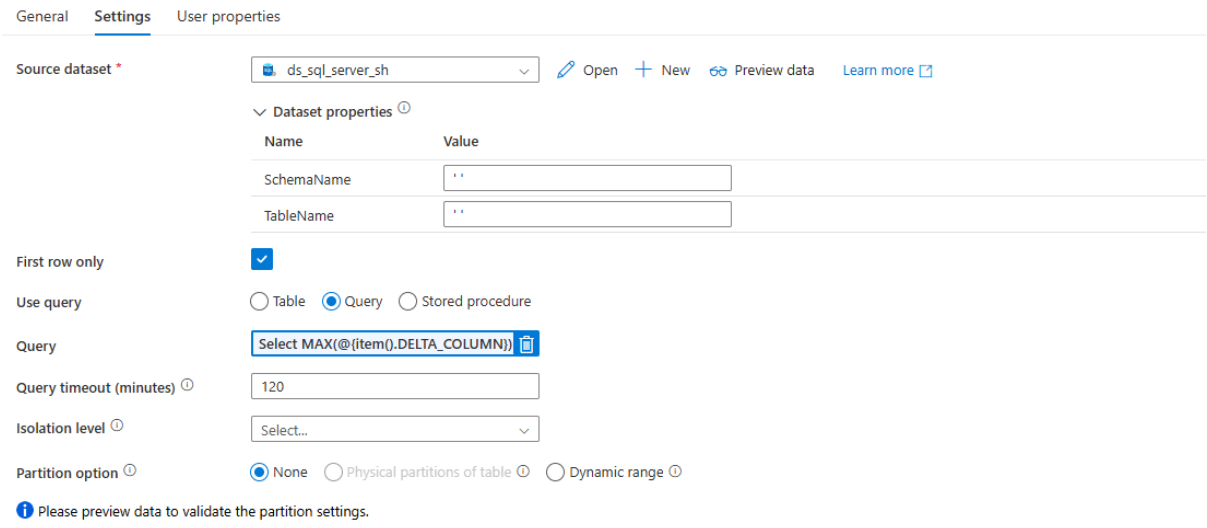



Click on the Edit icon on ForEach Activity

Take Lookup activity and rename “GetSourceMaxValue”

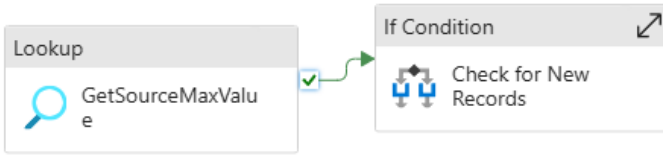



Select the Source Dataset as an on-premises database and linked service (Self-hosted IR) And Write this expression in the Query option




**Select MAX(@{item().DELTA_COLUMN}) as MaxValue from @{item().SCHEMA_NAME}.@{item().TABLE_NAME}**

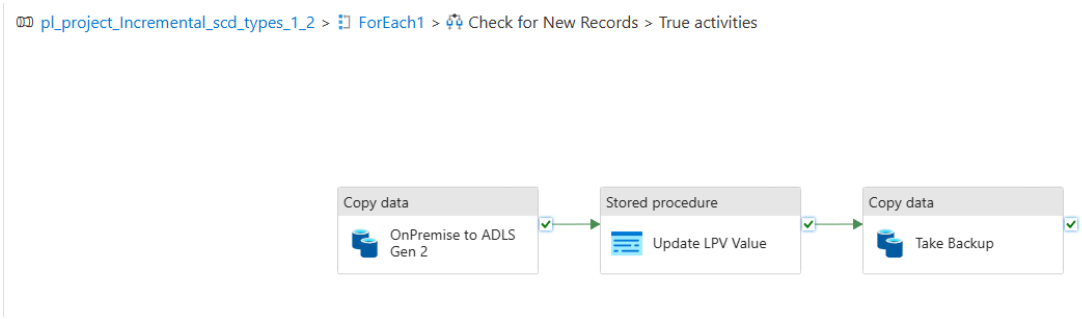
Take If Condition and write this expression




**@less(string(item().LPV), string(activity('GetSourceMaxValue').output.firstRow.MaxValue))**

Click the edit icon on the True Conditions part

Take these activities

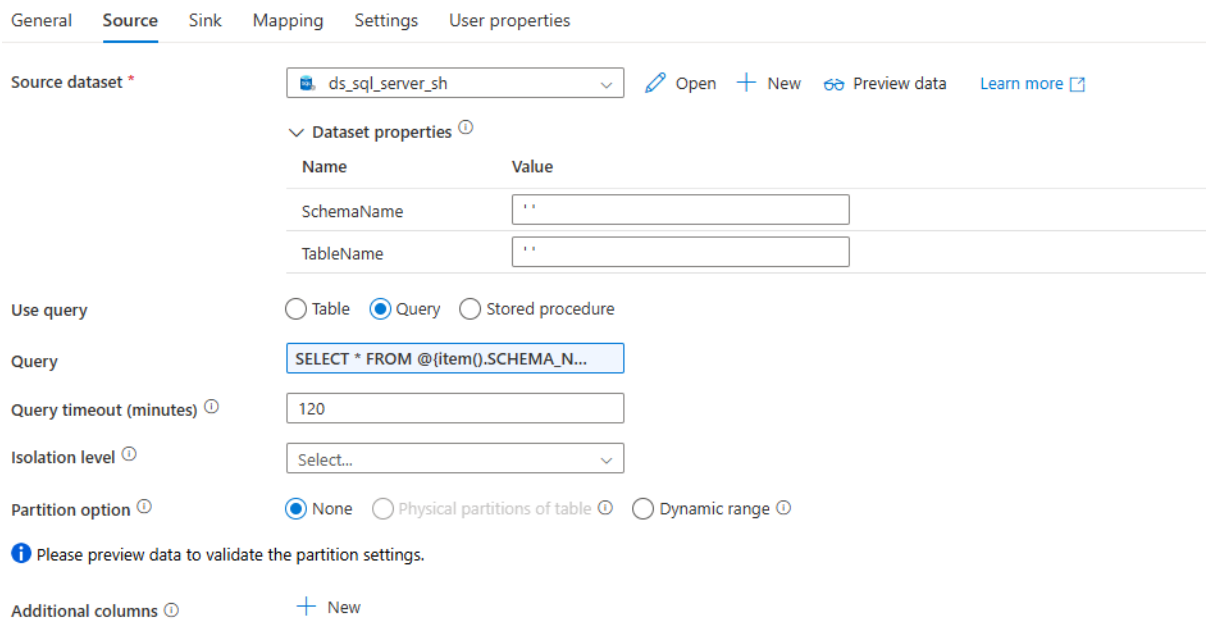



For Copy Data Activity:

Take Source Dataset as an on-premises database and Self Hosted Linked Service

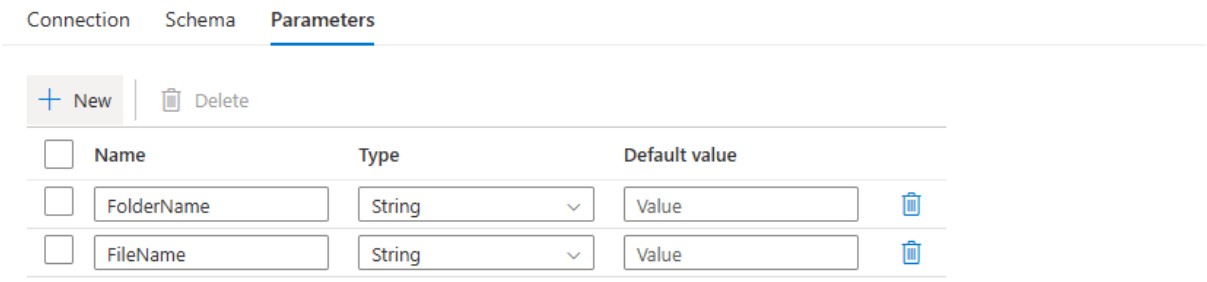
Write this query in the source section in the query option

**SELECT * FROM @{item().SCHEMA_NAME}.@{item().TABLE_NAME} WHERE @{item().DELTA_COLUMN}>'@{item().LPV}'**

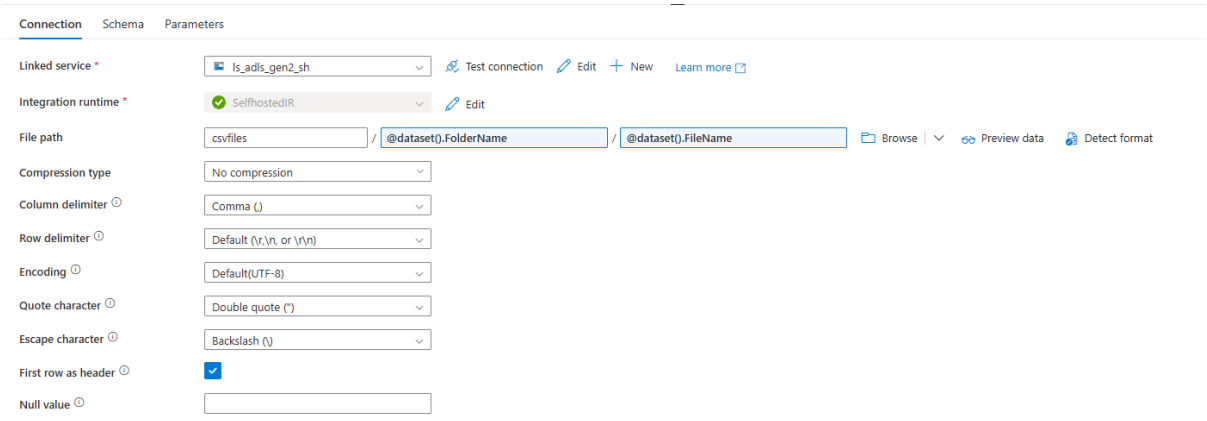



On the Sink side, select delimited Text source dataset and Auto Resolve Integration Runtime service.

Take two parameters




Connection tab

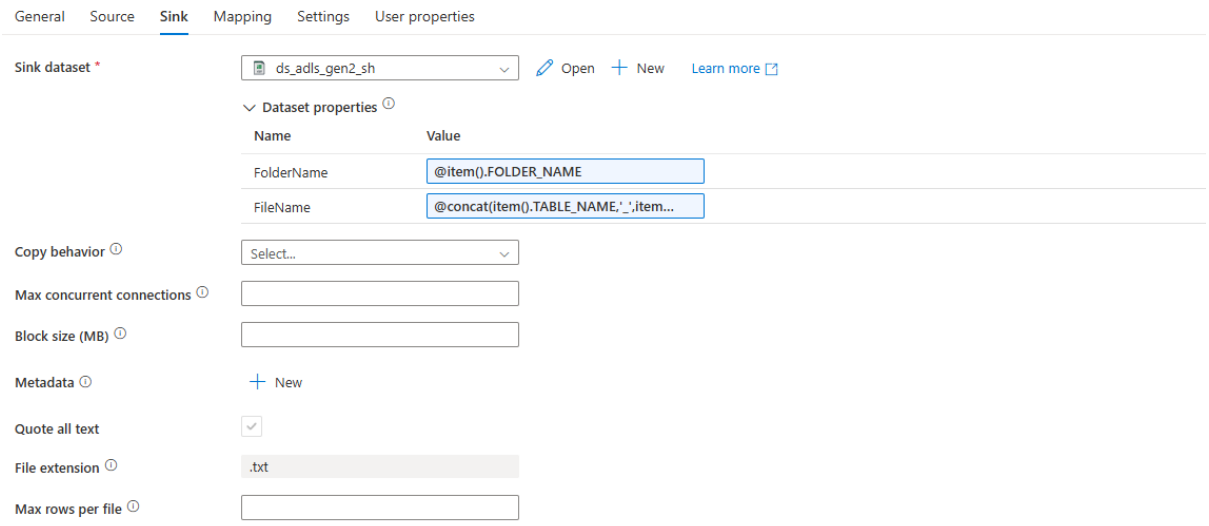



Write this expression for FolderName

**@item().FOLDER_NAME**

And this below expression for FileName

**@concat(item().TABLE_NAME,'_',item().TABLE_TYPE,'.csv')**

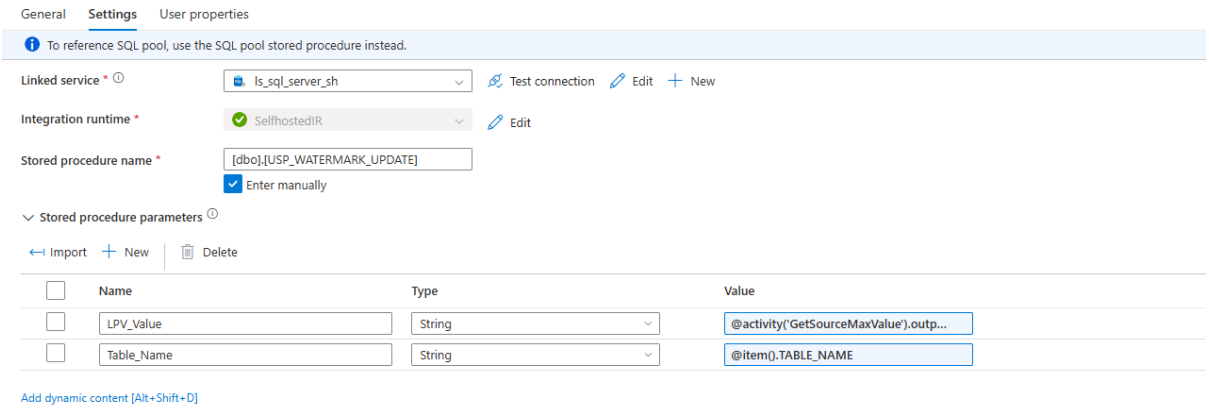



**For Stored Procedure Activity:**

Rename it as “Update LPV Value”

Select the linked service, select the stored procedure and click on import parameters

Two parameters will be imported from our stored procedure,




For LPV Value,

Write this expression

**@activity('GetSourceMaxValue').output.firstRow.MaxValue**

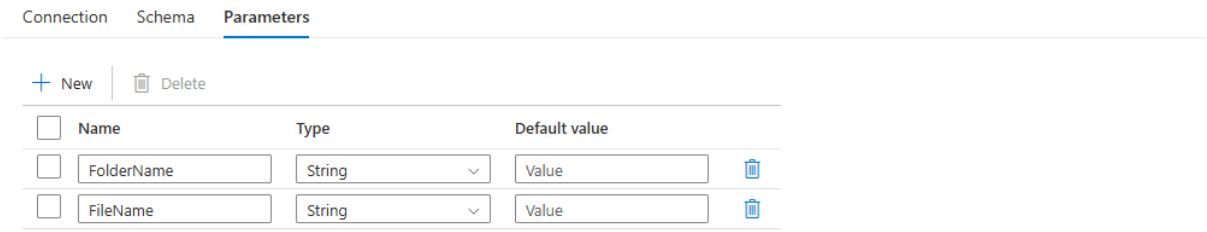
For the Table Name parameter, write this below expression,

**@item().TABLE_NAME**

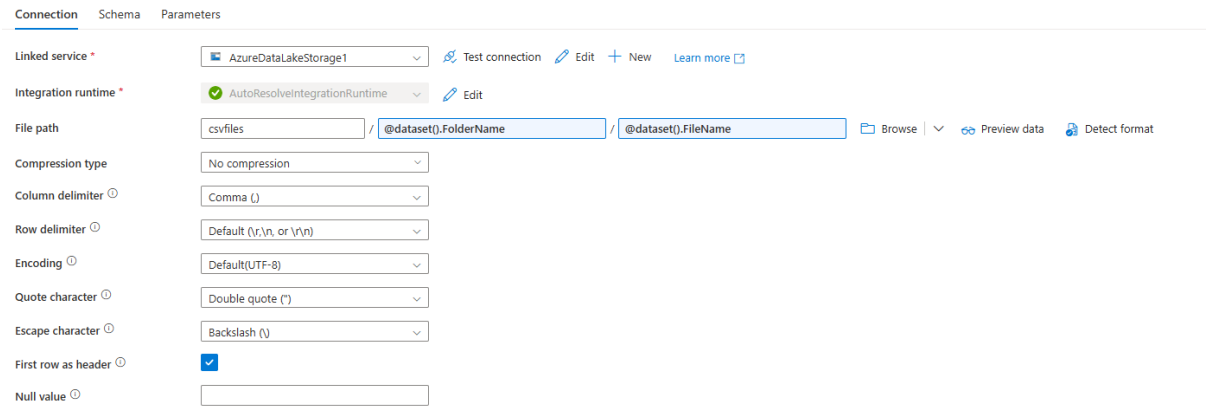
For CopyData Activity (Take Backup):

Source:

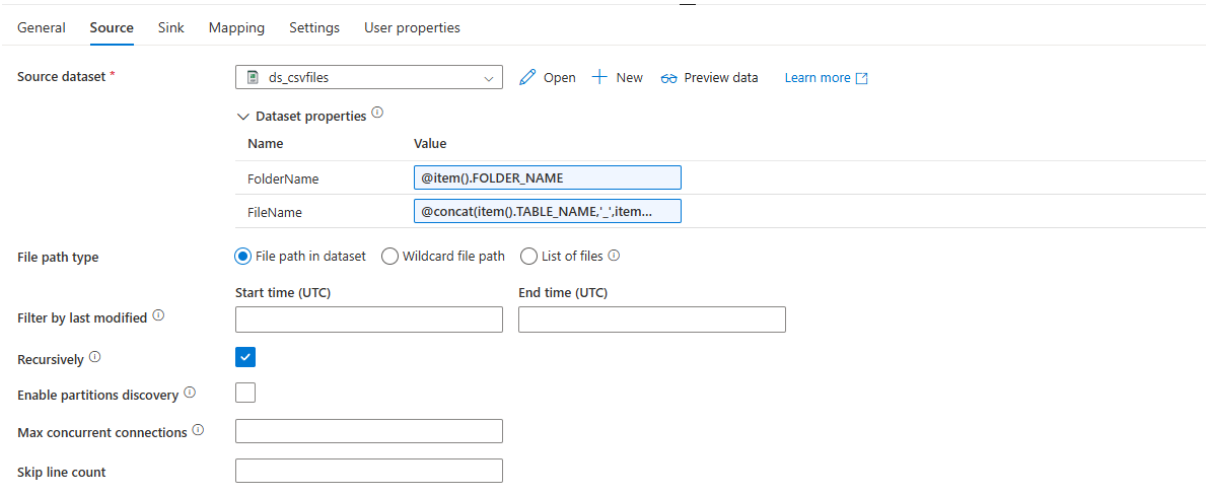
Select the source dataset as delimited text and write two parameters




Connection tab




Source



Provide the expression for Folder Name and File Name parameters as follows:

Folder Name:

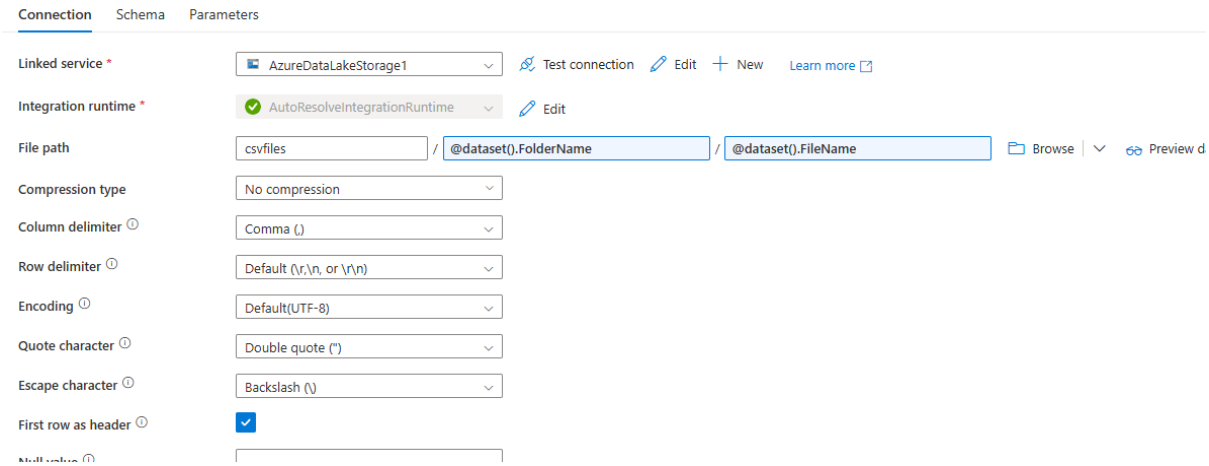
**@item().FOLDER_NAME**

File Name:

**@concat(item().TABLE_NAME,'_',item().TABLE_TYPE,'.csv')**

Go to the Sink,

Again take the source dataset as delimited text as shown below, take two parameters and give the connection in the connection tab




Provide the expression for Folder Name and File Name parameters as follows:

Folder Name:

**@concat('backup/',item().FOLDER_NAME)**

File Name:

**@concat(item().TABLE_NAME,'_',utcNow(),'.csv')**

**Now, Create 5 tables in Azure SQL Database as destination**

```
-------SCD Type 1 Table ----------------

CREATE TABLE tbl_Suppliers (
    supplier_id INT,
    name VARCHAR(255),
    contact_name VARCHAR(255),
    phone VARCHAR(50),
    address Varchar(100),
    createdBy varchar(100),
    createdDate datetime,
    updatedBy varchar(100),
    updatedDate datetime,
    hashKey Bigint
);

--TRUNCATE TABLE tbl_Suppliers
Select * from tbl_Suppliers

--Target Source Query in Dataflow SCD Type 1
Select supplier_id, hashKey from tbl_Suppliers;
```

```
CREATE TABLE tbl_Employee (
employee_id INT,
first_name VARCHAR(255),
last_name VARCHAR(255),
hire_date DATETIME,
last_review_date DATETIME,
role VARCHAR(100),
CREATEDBY VARCHAR(100),
CREATEDDATE DATETIME,
UPDATEDBY VARCHAR(100),
UPDATEDDATE DATETIME,
HASHKEY BIGINT,
ISACTIVE INT
);

--TRUNCATE TABLE tbl_Employee
Select * from tbl_Employee

--Target Source Query in Dataflow SCD Type 2
Select employee_id, HASHKEY from tbl_Employee where ISACTIVE = 1
```

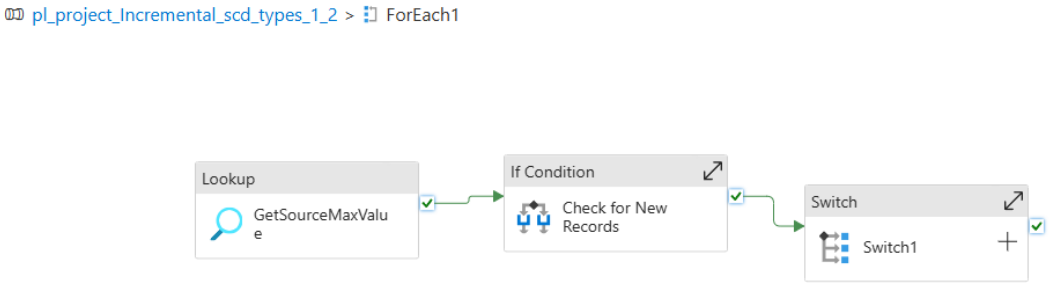
```
74  CREATE TABLE tbl_Products (
75      product_id INT,
76      name VARCHAR(255),
77      category VARCHAR(100),
78      price DECIMAL(10, 2),
79      stock INT,
80      supplier_id INT,
81      productUpdatedDate datetime
82  );
83
84  CREATE TABLE tbl_Sales (
85      sale_id INT,
86      sale_date DATETIME,
87      product_id INT,
88      quantity INT,
89      total_amount DECIMAL(10, 2),
90      cashier_id INT,
91      salesUpdatedDate datetime
92  );
93
94  CREATE TABLE tbl_Inventory_Logs (
95      log_id INT,
96      product_id INT,
97      log_updatedDate DATETIME,
98      change_quantity INT,
99      remaining_stock INT
100 );
```

```
CREATE TABLE tbl_Products (
  product_id INT,
  name VARCHAR(255),
  category VARCHAR(100),
  price DECIMAL(10, 2),
  stock INT,
  supplier_id INT,
        productUpdatedDate datetime
);

CREATE TABLE tbl_Sales (
  sale_id INT,
  sale_date DATETIME,
  product_id INT,
  quantity INT,
  total_amount DECIMAL(10, 2),
  cashier_id INT,
        salesUpdatedDate datetime
);
```

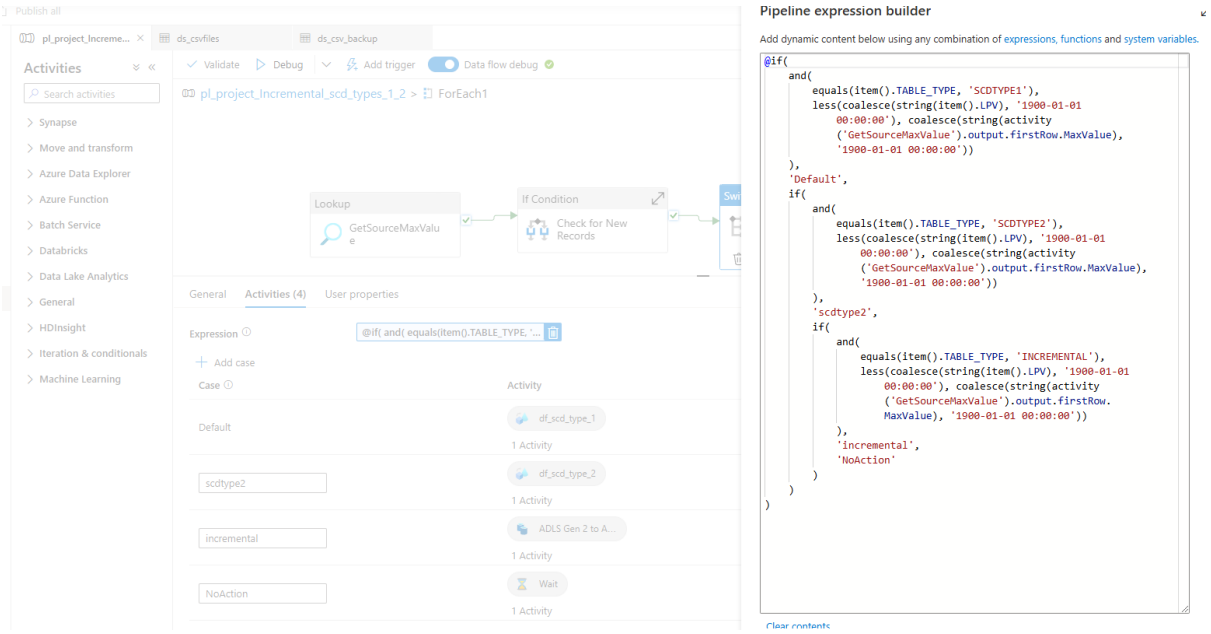
```
CREATE TABLE tbl_Inventory_Logs (
   log_id INT,
   product_id INT,
   log_updatedDate DATETIME,
   change_quantity INT,
   remaining_stock INT
);

CREATE TABLE tbl_Employee (
employee_id INT,
first_name VARCHAR(255),
last_name VARCHAR(255),
hire_date DATETIME,
last_review_date DATETIME,
role VARCHAR(100),
CREATEDBY VARCHAR(100),
CREATEDDATE DATETIME,
UPDATEDBY VARCHAR(100),
UPDATEDDATE DATETIME,
HASHKEY BIGINT,
ISACTIVE INT
);

--TRUNCATE TABLE tbl_Employee
Select * from tbl_Employee

--Target Source Query in Dataflow SCD Type 2
Select employee_id, HASHKEY from tbl_Employee where ISACTIVE = 1
-------SCD Type 1 Table ----------------

CREATE TABLE tbl_Suppliers (
   supplier_id INT,
   name VARCHAR(255),
   contact_name VARCHAR(255),
   phone VARCHAR(50),
   address Varchar(100),
         createdBy varchar(100),
         createdDate datetime,
         updatedBy varchar(100),
         updatedDate datetime,
         hashKey Bigint
);

--TRUNCATE TABLE tbl_Suppliers
Select * from tbl_Suppliers
```

--Target Source Query in Dataflow SCD Type 1
Select supplier_id, hashKey from tbl_Suppliers;

Now go outside of For Each activity, drag and drop **Switch Activity**



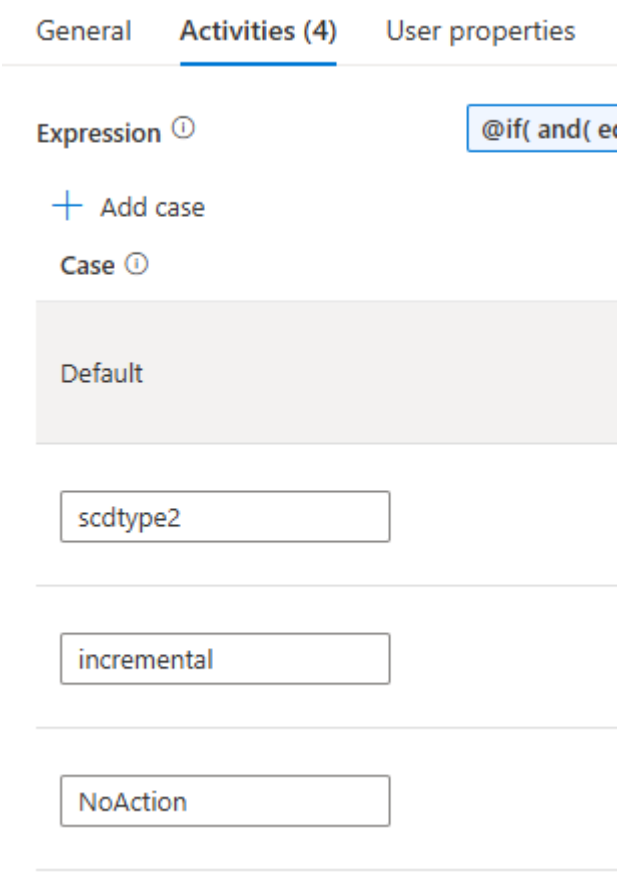

Click on Switch Activity, and go to Activities -> Expression and write this below expression




```
@if(
  and(
    equals(item().TABLE_TYPE, 'SCDTYPE1'),
    less(coalesce(string(item().LPV), '1900-01-01 00:00:00'),
coalesce(string(activity('GetSourceMaxValue').output.firstRow.MaxValue), '1900-
01-01 00:00:00'))
  ),
  'Default',
  if(
    and(
      equals(item().TABLE_TYPE, 'SCDTYPE2'),
```

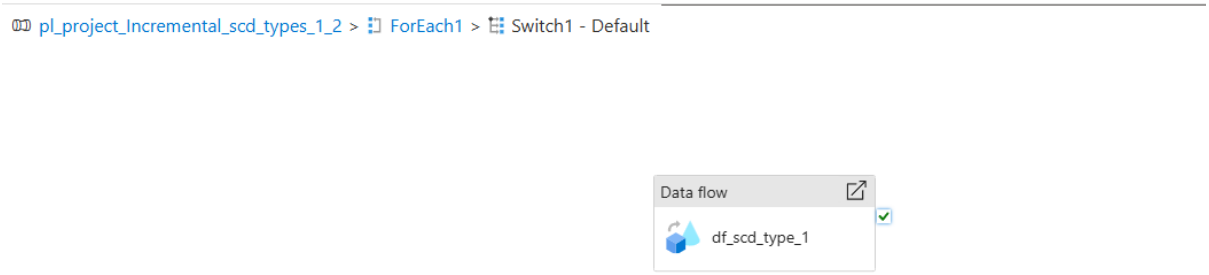
```
        less(coalesce(string(item().LPV), '1900-01-01 00:00:00'),
coalesce(string(activity('GetSourceMaxValue').output.firstRow.MaxValue), '1900-
01-01 00:00:00'))
      ),
      'scdtype2',
      if(
        and(
          equals(item().TABLE_TYPE, 'INCREMENTAL'),
          less(coalesce(string(item().LPV), '1900-01-01 00:00:00'),
coalesce(string(activity('GetSourceMaxValue').output.firstRow.MaxValue), '1900-
01-01 00:00:00'))
        ),
        'incremental',
        'NoAction'
      )
    )
)
```

Click on **Add case** and create 3 new cases as below

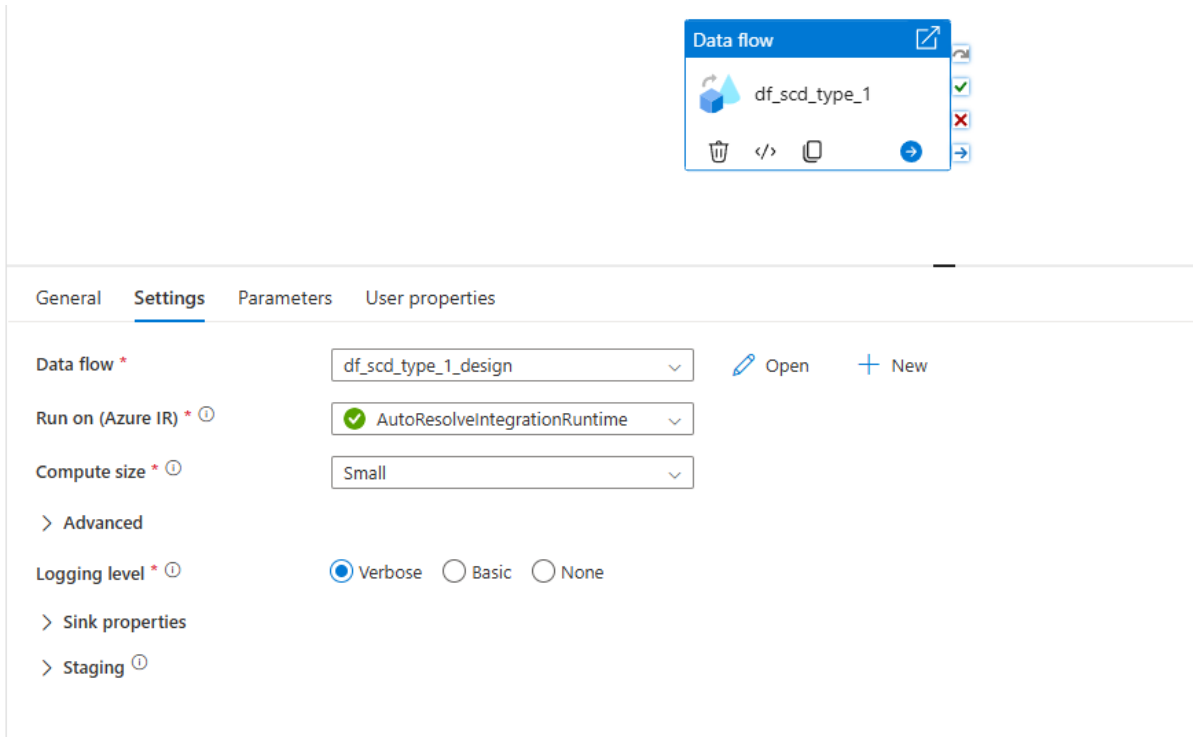


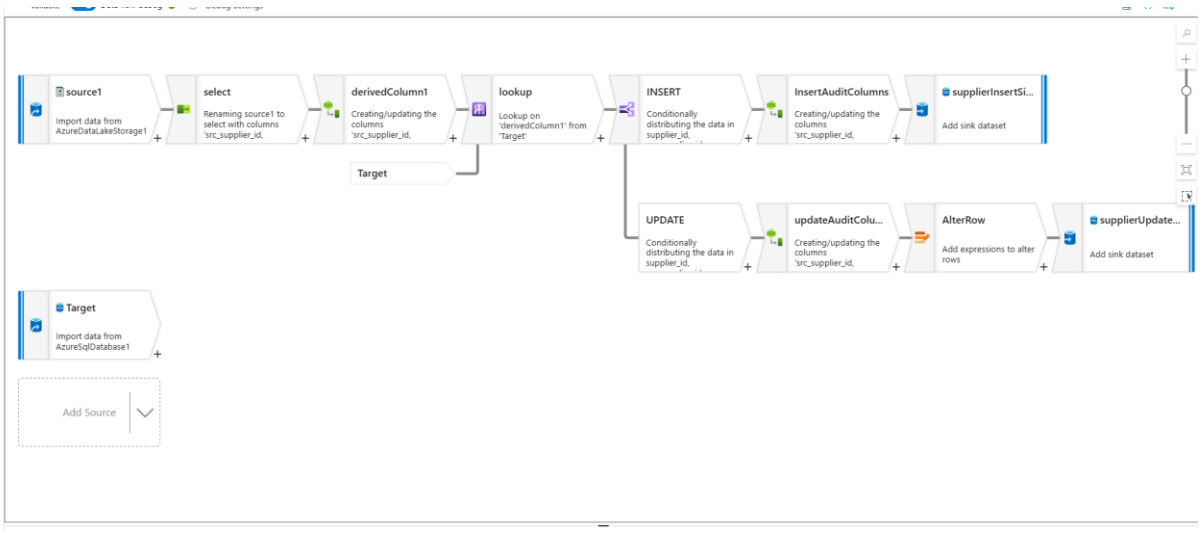

Click on the Default case edit icon

## Drag and drop Dataflow Activity

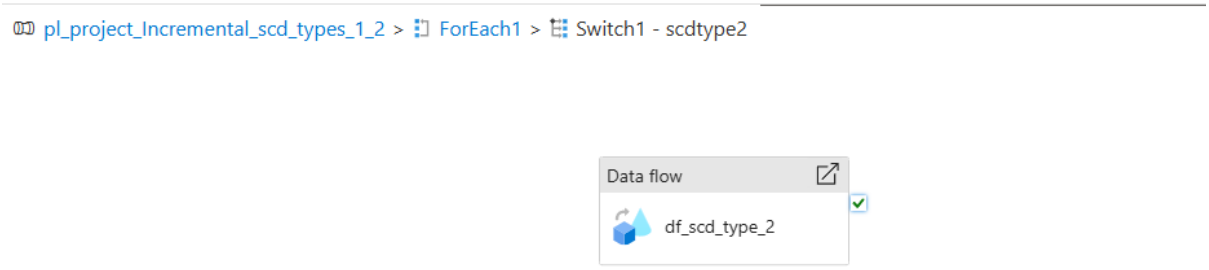



## Go to settings and select the dataflow we have already designed for SCD type 1 logic

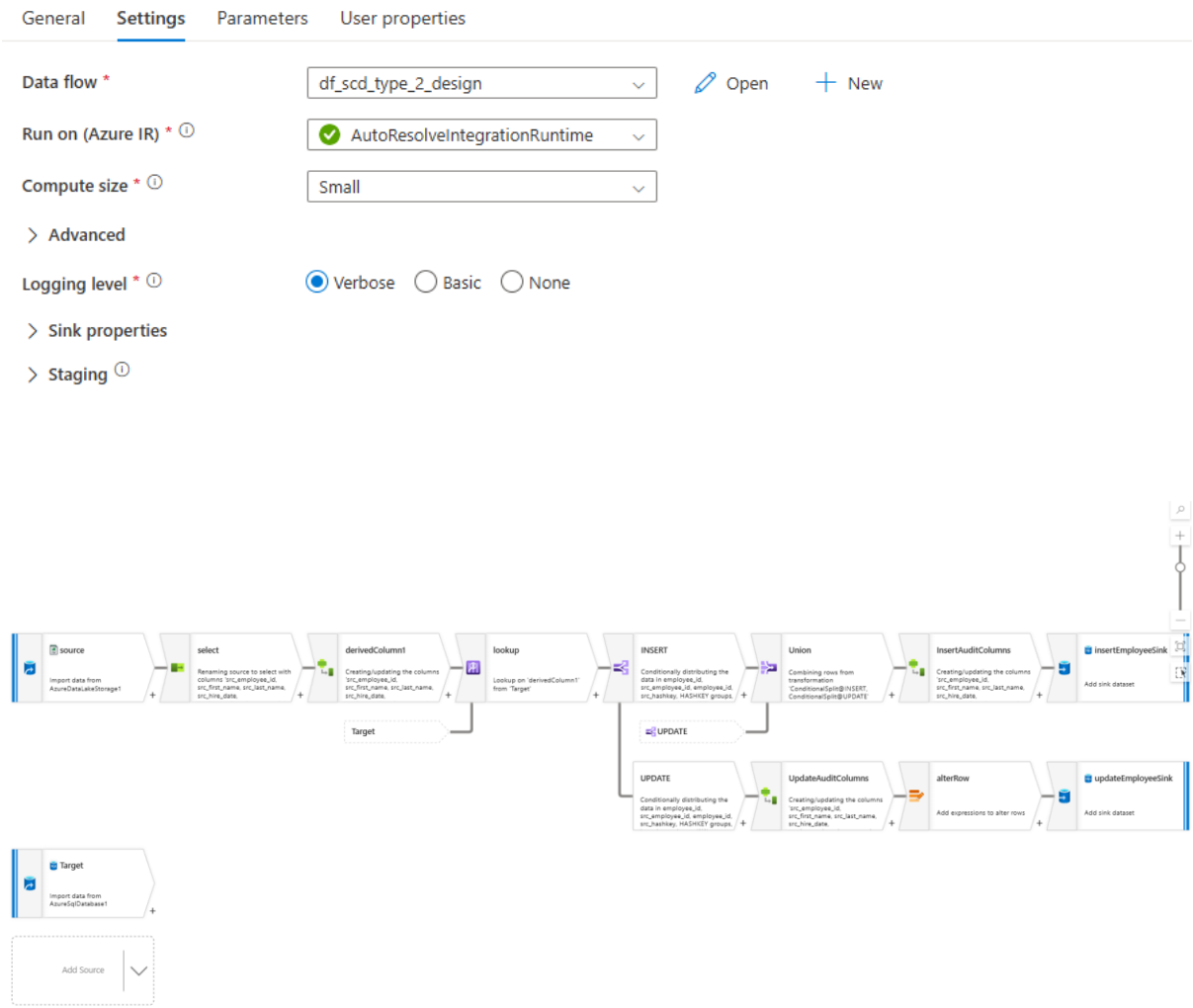


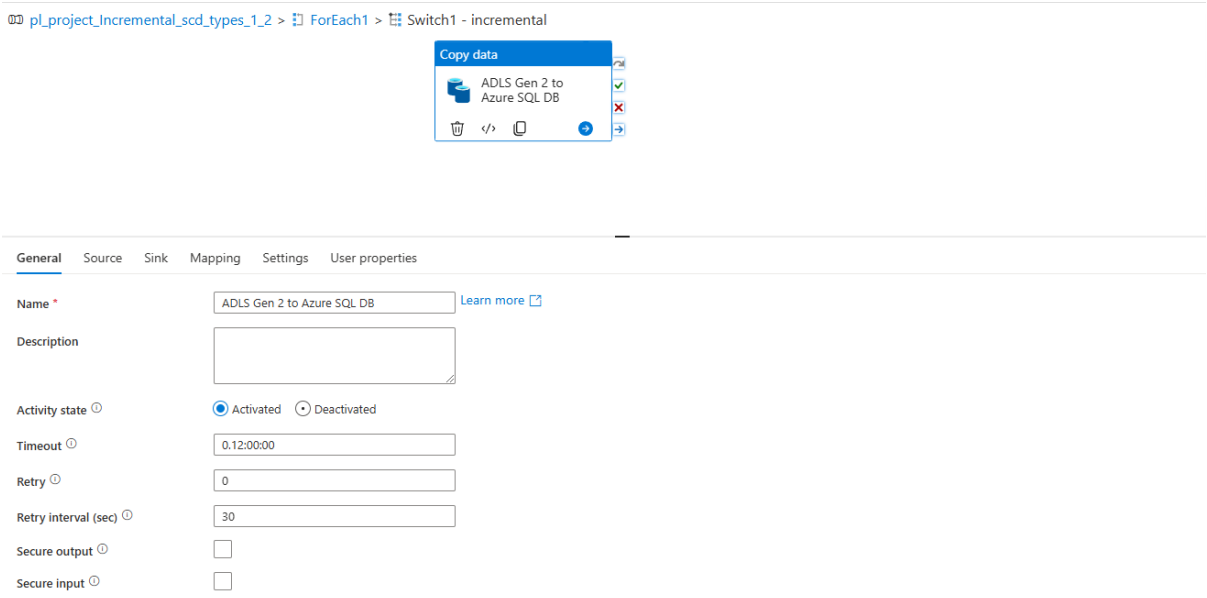
Go back to switch case -> **scdtype2** edit icon and drag and drop dataflow activity



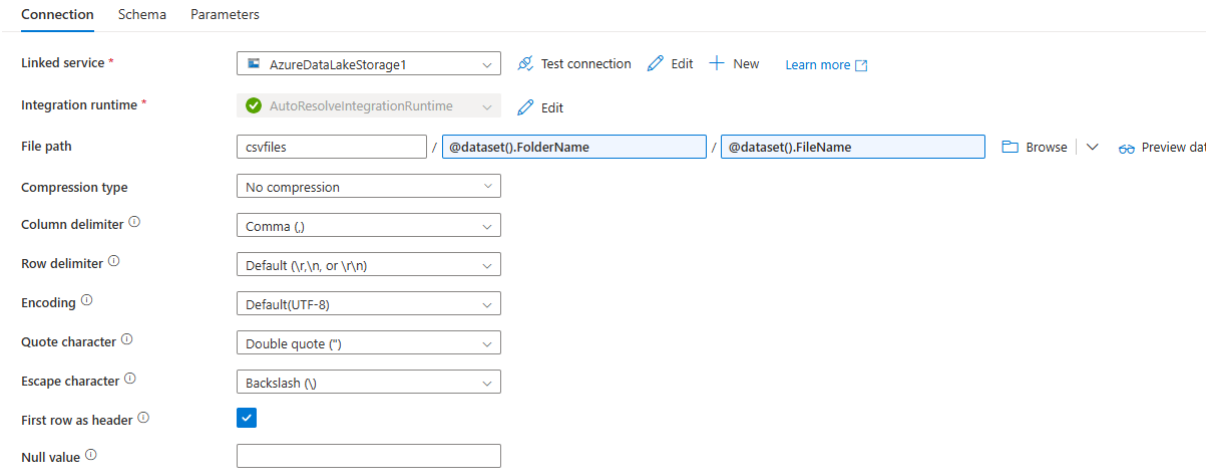

Go to settings and select the dataflow we have already designed for SCD type 2 logic



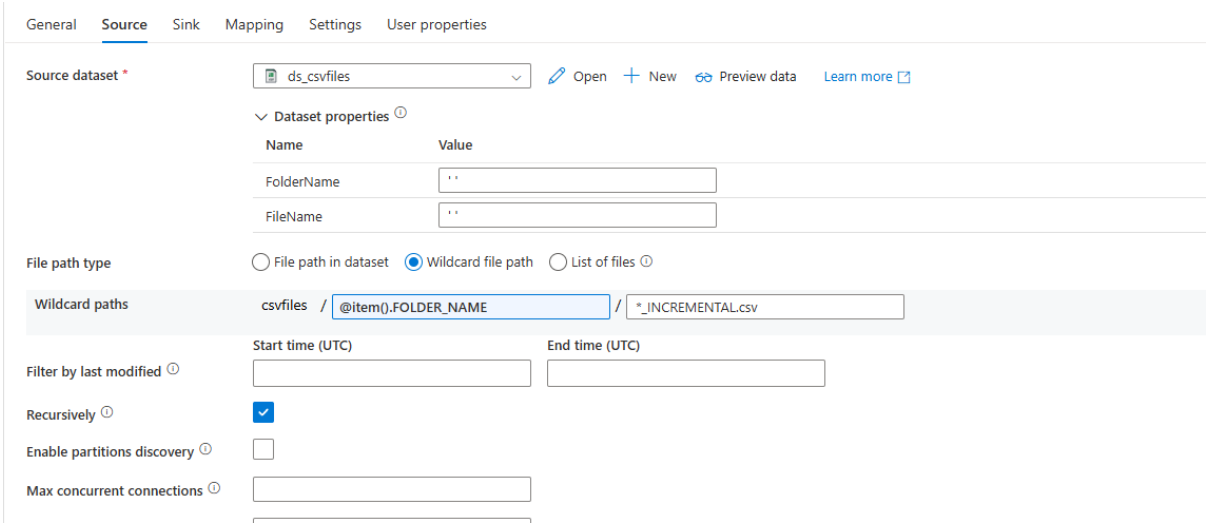

Go back to switch case -> **incremental** edit icon and drag and drop copy data activity




Select the source dataset as delimited text file and take two parameters and build the connection




Select the Wildcard paths option and write these two expression and value

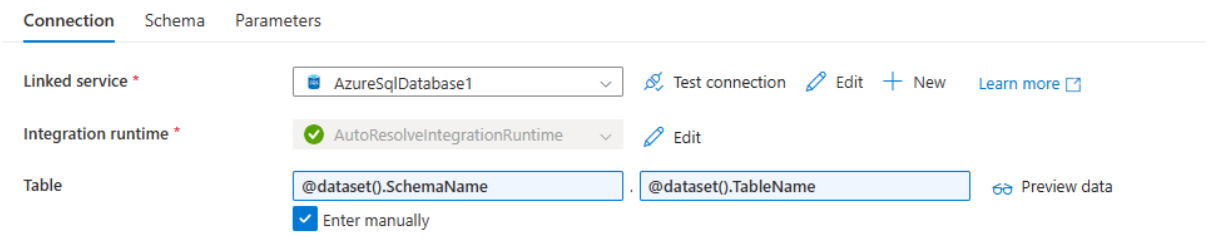


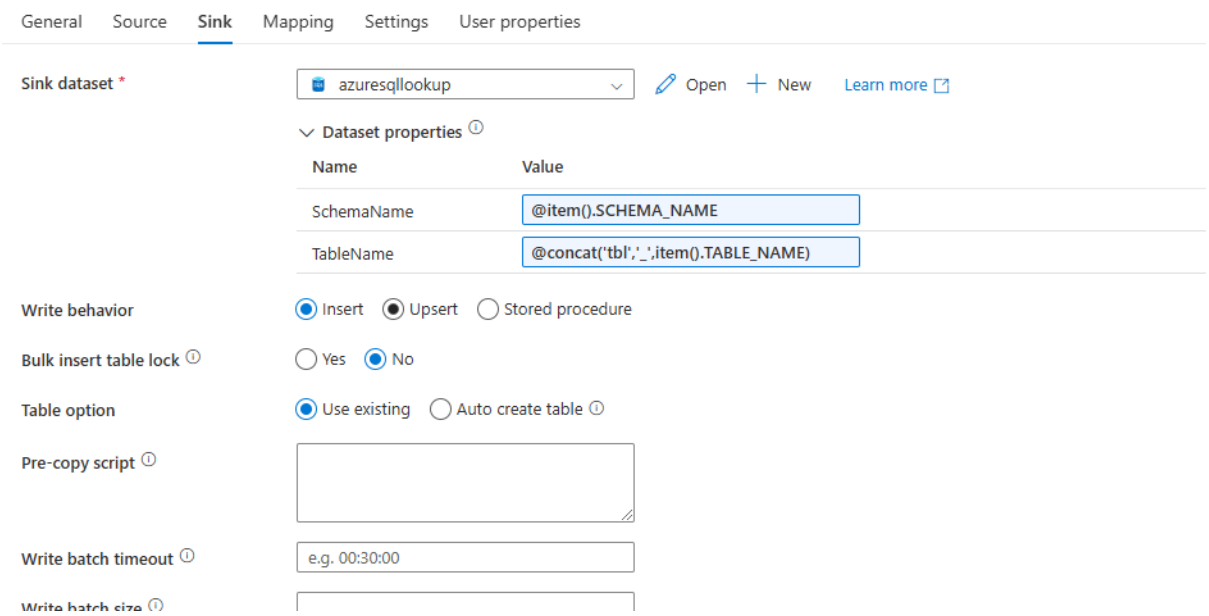

Folder name: **@item().FOLDER_NAME**

Filename: ***_INCREMENTAL.csv**

Go to the Sink

Select the source dataset as Azure SQL Database, take two parameters and build the connection






Write this two expressions in Schema Name and Table Name

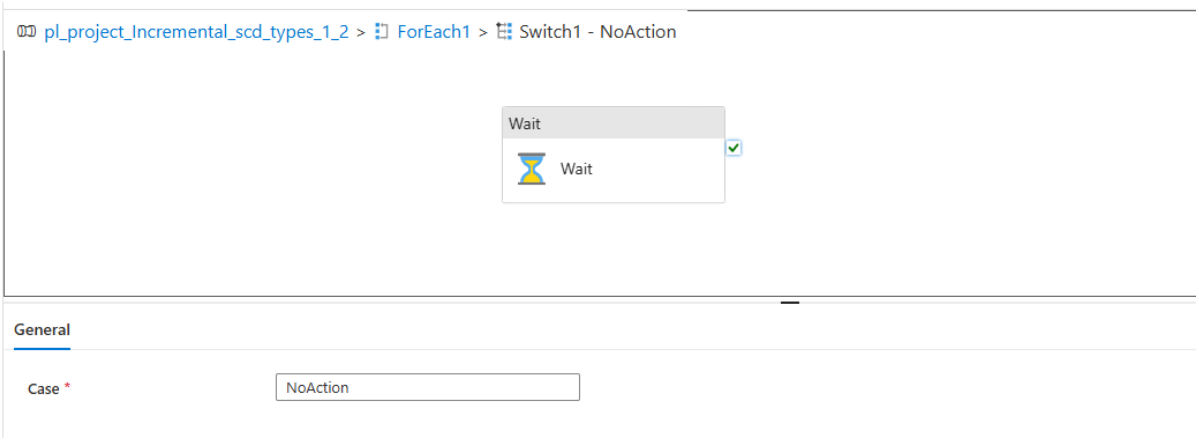
Schema Name:

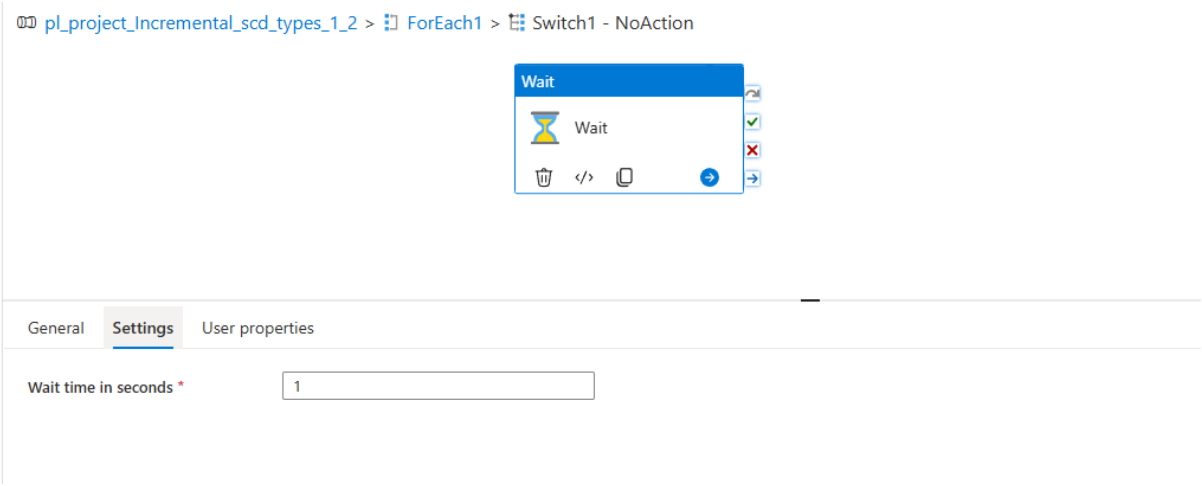
**@item().SCHEMA_NAME**

Table Name:

**@concat('tbl','_',item().TABLE_NAME)**

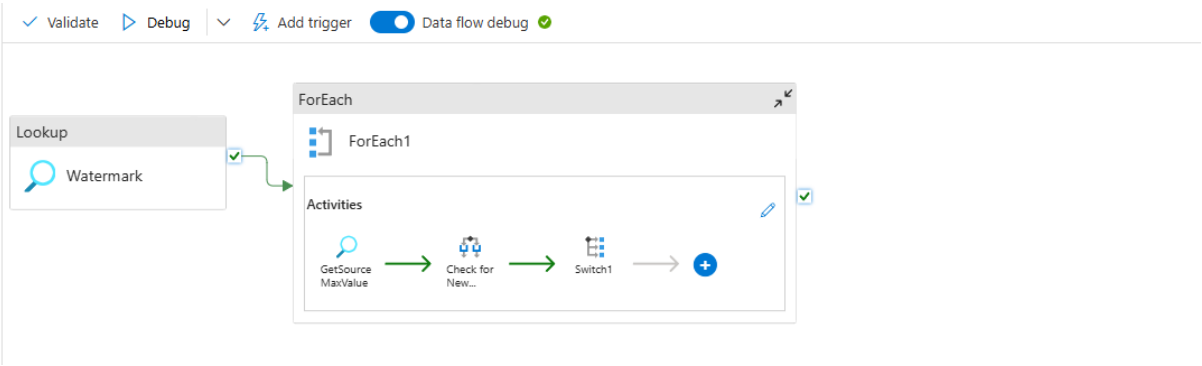
Go back to switch case -> **No Action** edit icon and drag and drop Wait Activity



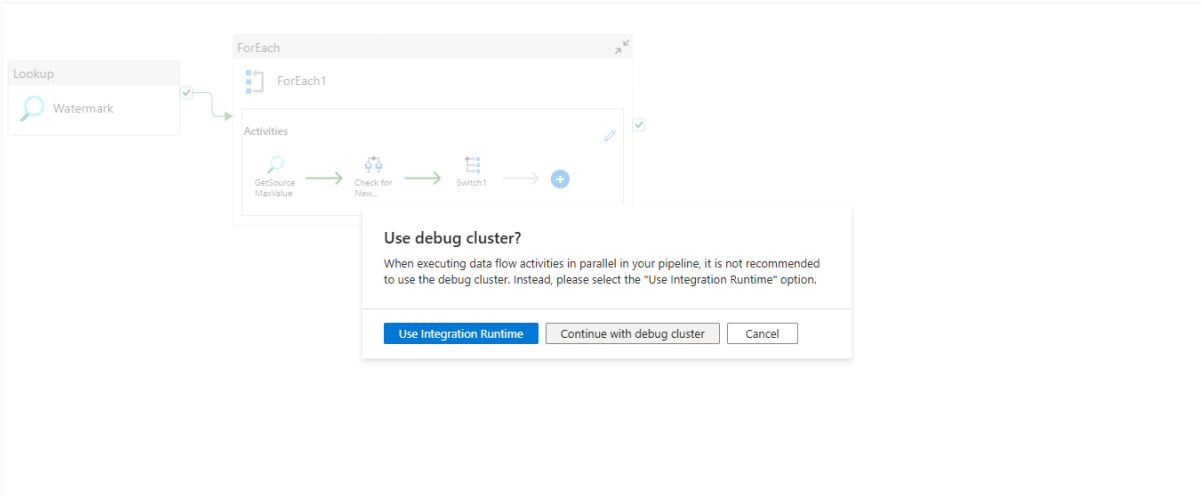


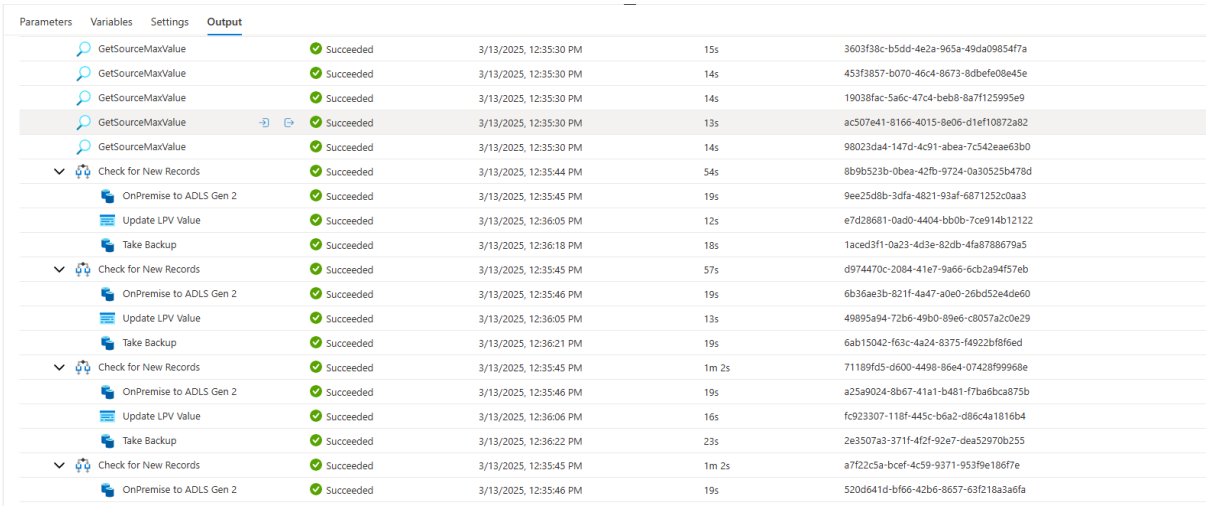

Let's run the pipeline

Enable **Data flow Debug** option




Click on **Continue with debug cluster**

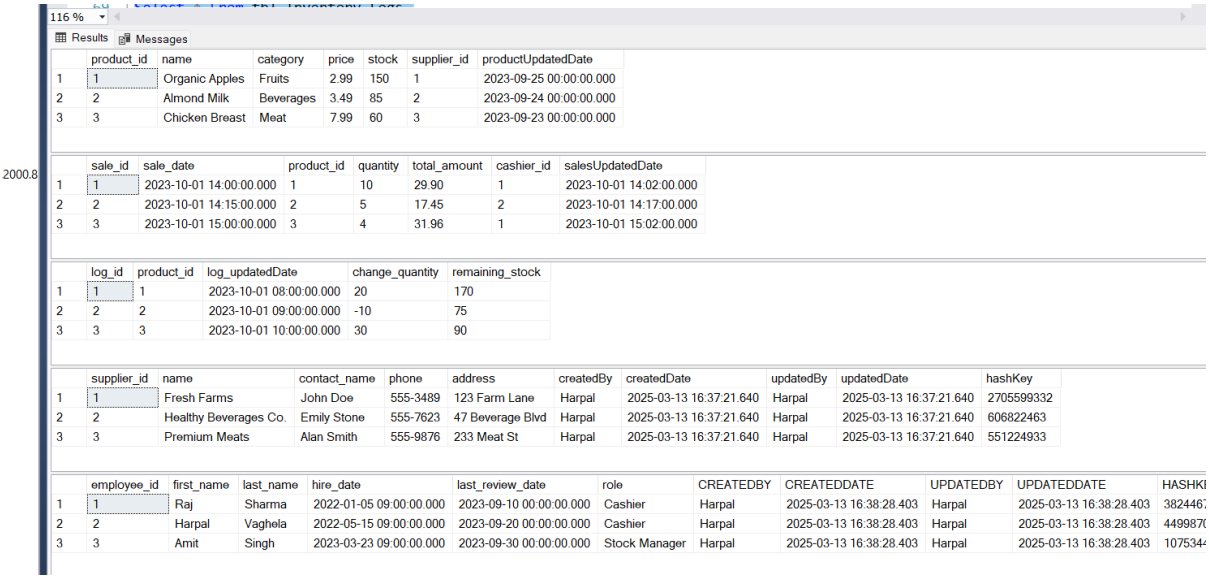




Parameters | Variables | Settings | Output

| | | | | |
|---|---|---|---|---|
| GetSourceMaxValue | Succeeded | 3/13/2025, 12:35:30 PM | 15s | 3603f38c-b5dd-4e2a-965a-49da09854f7a |
| GetSourceMaxValue | Succeeded | 3/13/2025, 12:35:30 PM | 14s | 453f3857-b070-46c4-8673-8dbefe08e45e |
| GetSourceMaxValue | Succeeded | 3/13/2025, 12:35:30 PM | 14s | 19038fac-5a6c-47c4-beb8-8a7f125995e9 |
| GetSourceMaxValue | Succeeded | 3/13/2025, 12:35:30 PM | 13s | ac507e41-8166-4015-8e06-d1ef10872a82 |
| GetSourceMaxValue | Succeeded | 3/13/2025, 12:35:30 PM | 14s | 98023da4-147d-4c91-abea-7c542eae63b0 |
| Check for New Records | Succeeded | 3/13/2025, 12:35:44 PM | 54s | 8b9b523b-0bea-42fb-9724-0a30525b478d |
| OnPremise to ADLS Gen 2 | Succeeded | 3/13/2025, 12:35:45 PM | 19s | 9ee25d8b-3dfa-4821-93af-6871252c0aa3 |
| Update LPV Value | Succeeded | 3/13/2025, 12:36:05 PM | 12s | e7d28681-0ad0-4404-bb0b-7ce914b12122 |
| Take Backup | Succeeded | 3/13/2025, 12:36:18 PM | 18s | 1aced3f1-0a23-4d3e-82db-4fa8788679a5 |
| Check for New Records | Succeeded | 3/13/2025, 12:35:45 PM | 57s | d974470c-2084-41e7-9a66-6cb2a94f57eb |
| OnPremise to ADLS Gen 2 | Succeeded | 3/13/2025, 12:35:46 PM | 19s | 6b36ae3b-821f-4a47-a0e0-26bd52e4de60 |
| Update LPV Value | Succeeded | 3/13/2025, 12:36:05 PM | 13s | 49895a94-72b6-49b0-89e6-c8057a2c0e29 |
| Take Backup | Succeeded | 3/13/2025, 12:36:21 PM | 19s | 6ab15042-f63c-4a24-8375-f4922bf8f6ed |
| Check for New Records | Succeeded | 3/13/2025, 12:35:45 PM | 1m 2s | 71189fd5-d600-4498-86e4-07428f99968e |
| OnPremise to ADLS Gen 2 | Succeeded | 3/13/2025, 12:35:46 PM | 19s | a25a9024-8b67-41a1-b481-f7ba6bca875b |
| Update LPV Value | Succeeded | 3/13/2025, 12:36:06 PM | 16s | fc923307-118f-445c-b6a2-d86c4a1816b4 |
| Take Backup | Succeeded | 3/13/2025, 12:36:22 PM | 23s | 2e3507a3-371f-4f2f-92e7-dea52970b255 |
| Check for New Records | Succeeded | 3/13/2025, 12:35:45 PM | 1m 2s | a7f22c5a-bcef-4c59-9371-953f9e186f7e |
| OnPremise to ADLS Gen 2 | Succeeded | 3/13/2025, 12:35:46 PM | 19s | 520d641d-bf66-42b6-8657-63f218a3a6fa |

## Let’s check the data in Azure SQL Database tables



116 % | Results | Messages

| | product_id | name | category | price | stock | supplier_id | productUpdatedDate |
|---|---|---|---|---|---|---|---|
| 1 | 1 | Organic Apples | Fruits | 2.99 | 150 | 1 | 2023-09-25 00:00:00.000 |
| 2 | 2 | Almond Milk | Beverages | 3.49 | 85 | 2 | 2023-09-24 00:00:00.000 |
| 3 | 3 | Chicken Breast | Meat | 7.99 | 60 | 3 | 2023-09-23 00:00:00.000 |

2000.8

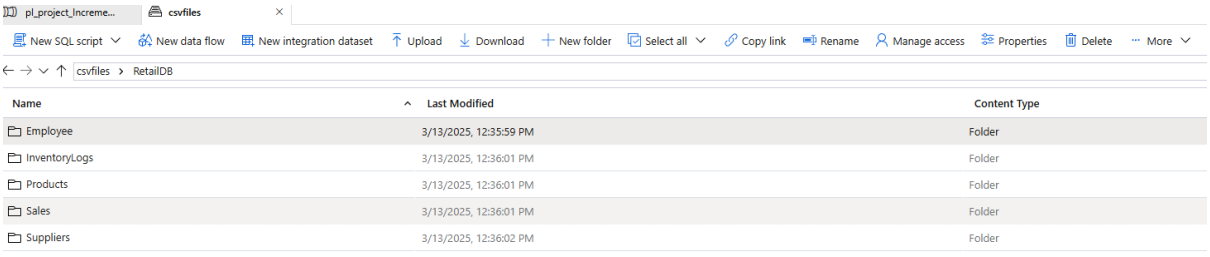
| | sale_id | sale_date | product_id | quantity | total_amount | cashier_id | salesUpdatedDate |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 2023-10-01 14:00:00.000 | 1 | 10 | 29.90 | 1 | 2023-10-01 14:02:00.000 |
| 2 | 2 | 2023-10-01 14:15:00.000 | 2 | 5 | 17.45 | 2 | 2023-10-01 14:17:00.000 |
| 3 | 3 | 2023-10-01 15:00:00.000 | 3 | 4 | 31.96 | 1 | 2023-10-01 15:02:00.000 |

| | log_id | product_id | log_updatedDate | change_quantity | remaining_stock |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 2023-10-01 08:00:00.000 | 20 | 170 |
| 2 | 2 | 2 | 2023-10-01 09:00:00.000 | -10 | 75 |
| 3 | 3 | 3 | 2023-10-01 10:00:00.000 | 30 | 90 |

| | supplier_id | name | contact_name | phone | address | createdBy | createdDate | updatedBy | updatedDate | hashKey |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | Fresh Farms | John Doe | 555-3489 | 123 Farm Lane | Harpal | 2025-03-13 16:37:21.640 | Harpal | 2025-03-13 16:37:21.640 | 2705599332 |
| 2 | 2 | Healthy Beverages Co. | Emily Stone | 555-7623 | 47 Beverage Blvd | Harpal | 2025-03-13 16:37:21.640 | Harpal | 2025-03-13 16:37:21.640 | 606822463 |
| 3 | 3 | Premium Meats | Alan Smith | 555-9876 | 233 Meat St | Harpal | 2025-03-13 16:37:21.640 | Harpal | 2025-03-13 16:37:21.640 | 551224933 |

| | employee_id | first_name | last_name | hire_date | last_review_date | role | CREATEDBY | CREATEDDATE | UPDATEDBY | UPDATEDDATE | HASHKI |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | Raj | Sharma | 2022-01-05 09:00:00.000 | 2023-09-10 00:00:00.000 | Cashier | Harpal | 2025-03-13 16:38:28.403 | Harpal | 2025-03-13 16:38:28.403 | 382446 |
| 2 | 2 | Harpal | Vaghela | 2022-05-15 09:00:00.000 | 2023-09-20 00:00:00.000 | Cashier | Harpal | 2025-03-13 16:38:28.403 | Harpal | 2025-03-13 16:38:28.403 | 449987 |
| 3 | 3 | Amit | Singh | 2023-03-23 09:00:00.000 | 2023-09-30 00:00:00.000 | Stock Manager | Harpal | 2025-03-13 16:38:28.403 | Harpal | 2025-03-13 16:38:28.403 | 107534 |

## Let’s check the data in the ADLS Gen 2 storage account



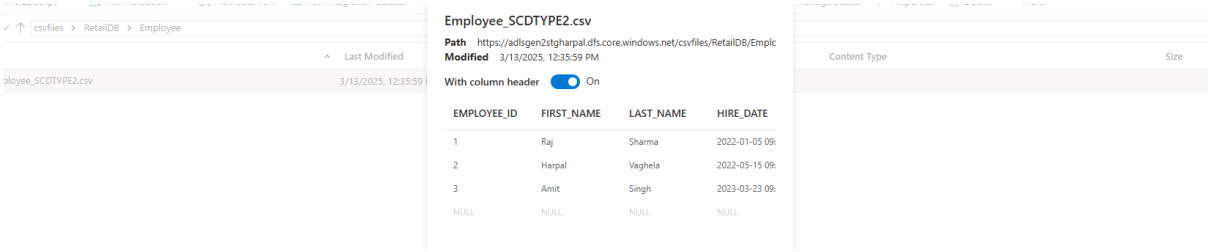
pl_project_Increme... | csvfiles

New SQL script | New data flow | New integration dataset | Upload | Download | New folder | Select all | Copy link | Rename | Manage access | Properties | Delete | More

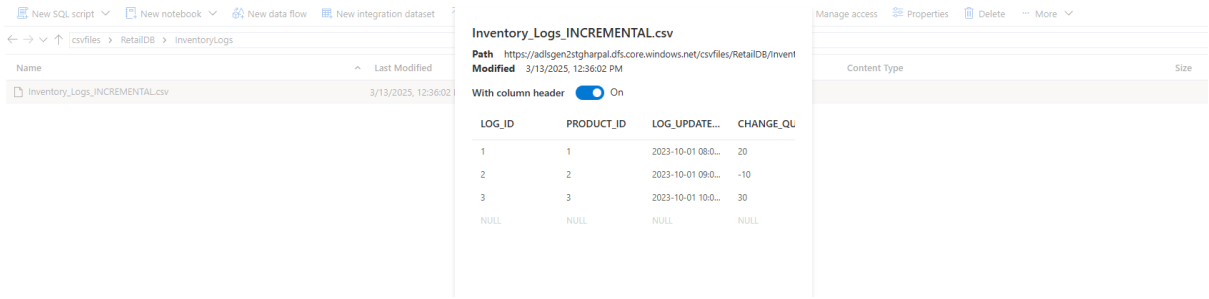
csvfiles › RetailDB

| Name | Last Modified | Content Type |
|---|---|---|
| Employee | 3/13/2025, 12:35:59 PM | Folder |
| InventoryLogs | 3/13/2025, 12:36:01 PM | Folder |
| Products | 3/13/2025, 12:36:01 PM | Folder |
| Sales | 3/13/2025, 12:36:01 PM | Folder |
| Suppliers | 3/13/2025, 12:36:02 PM | Folder |

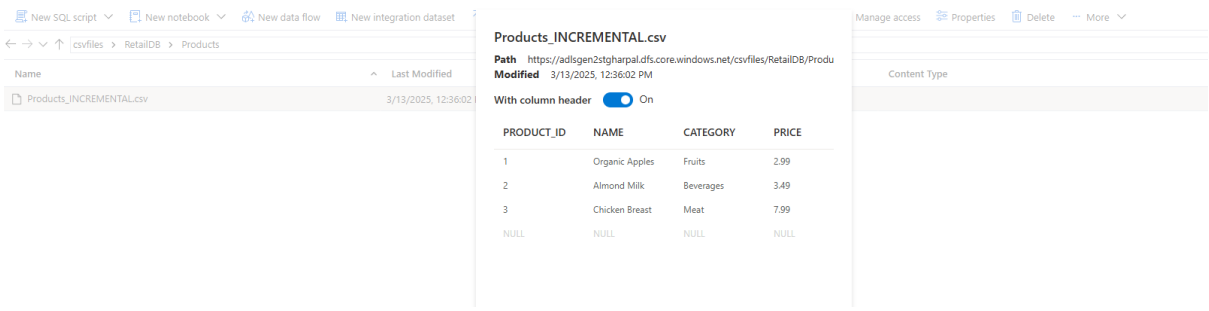
## Employee Data



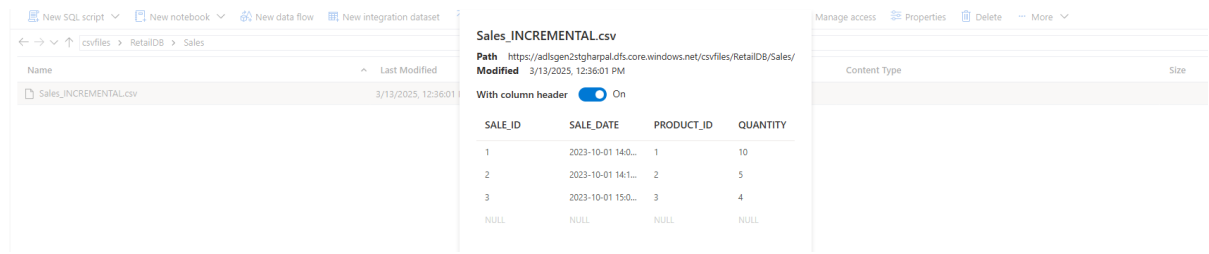

## Inventory Log Data



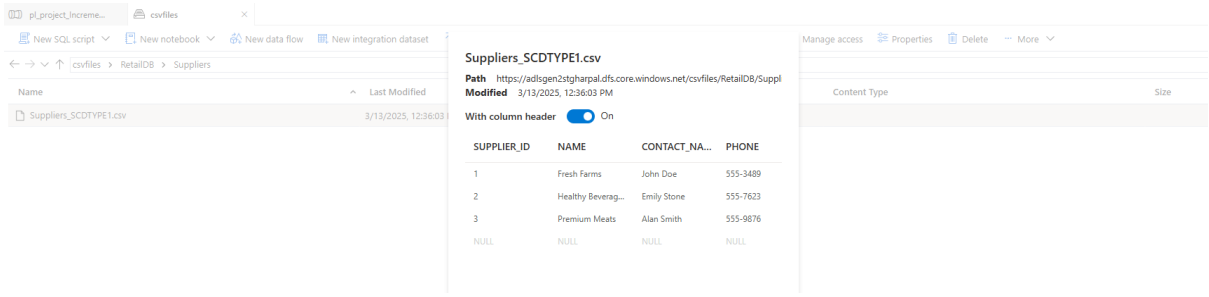

## Products Data



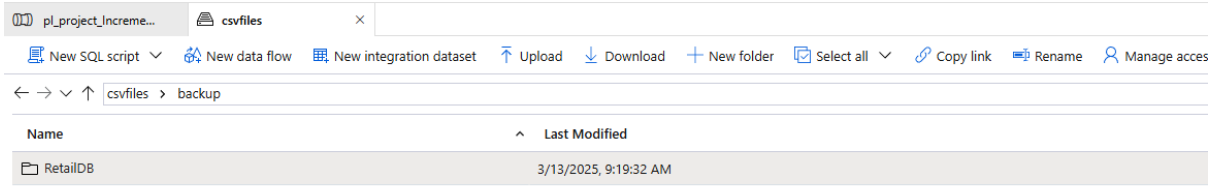

## Sales Data

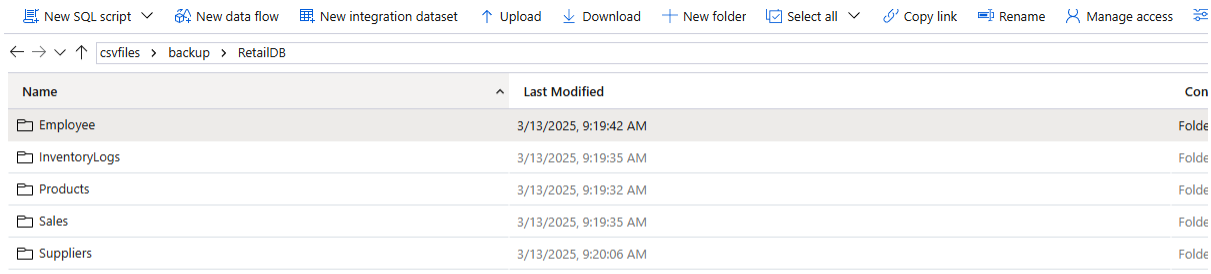



## Supplier Data

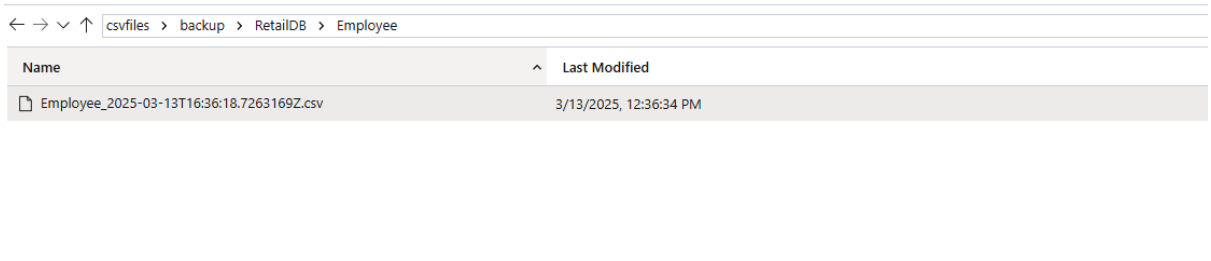


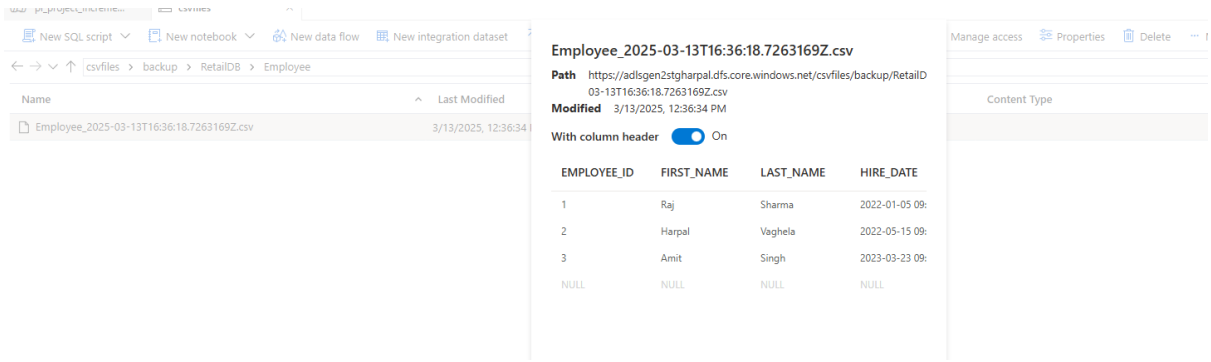

Lets check the backup container in storage account



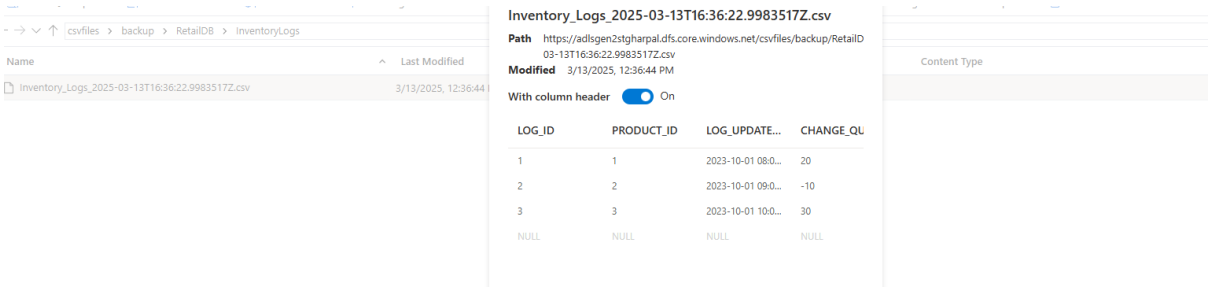




## Employee Data Backup

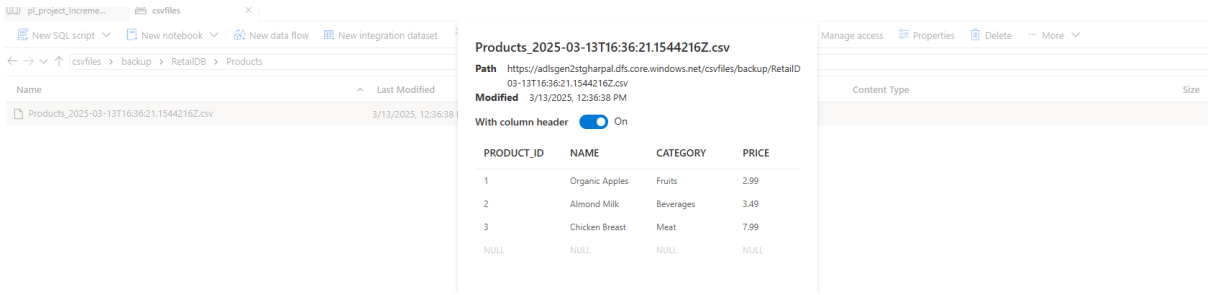


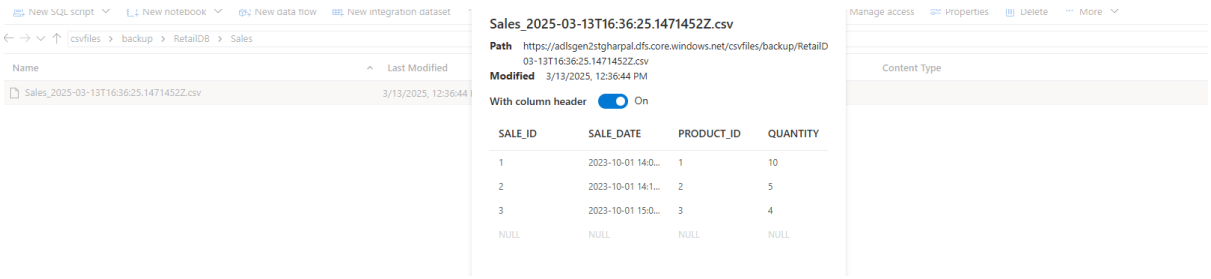

## Inventory Data Backup



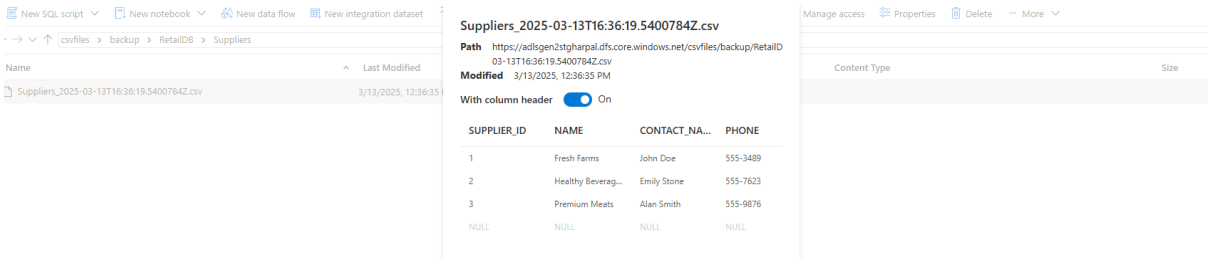
## Product Data Backup




## Sales Data Backup




## Supplier Data Backup




**Let's insert new records in on-premise database tables**

```
----------------------------------------------------------------------------------------------
----Insert new data into Products table
INSERT INTO Products (product_id, name, category, price, stock, supplier_id, productUpdatedDate) VALUES
(4, 'Vegetable Oil', 'Groceries', 4.50, 100, 2, '2023-12-01 00:00:00');

----Insert new data into Sales table
INSERT INTO Sales (sale_id, sale_date, product_id, quantity, total_amount, cashier_id, salesUpdatedDate) VALUES
(4, '2025-02-25 10:30:00', 5, 20, 90.00, 1, '2025-02-25 10:35:00');


----Insert new data into Suppliers table --- SCD Type 1
INSERT INTO Suppliers (supplier_id, name, contact_name, phone, address, supplierUpdatedDate) VALUES
(4, 'Bakers Delight', 'Nora Special Bates', '111-1111', '88 Baker Rd', '2025-02-22 00:00:00');


----Insert new data into Employees table ---- SCD Type 2
INSERT INTO Employee (employee_id, first_name, last_name, hire_date, last_review_date, role) VALUES
(4, 'Nayan', 'Vaghela', '2025-01-01 09:00:00', '2025-02-20 00:00:00', 'Inventory Specialist');
```

----Insert new data into Products table
INSERT INTO Products (product_id, name, category, price, stock, supplier_id, productUpdatedDate) VALUES
(4, 'Vegetable Oil', 'Groceries', 4.50, 100, 2, '2023-12-01 00:00:00');

----Insert new data into Sales table

```
INSERT INTO Sales (sale_id, sale_date, product_id, quantity, total_amount, cashier_id, salesUpdatedDate) VALUES
(4, '2025-02-25 10:30:00', 5, 20, 90.00, 1, '2025-02-25 10:35:00');
```

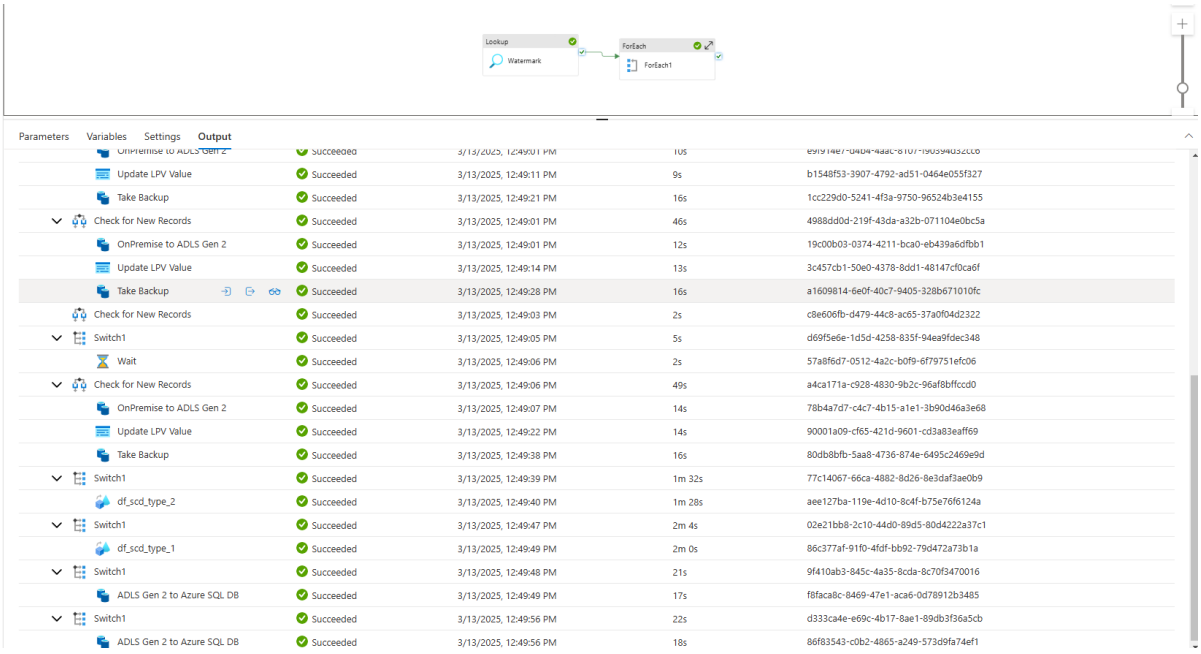
----Insert new data into Suppliers table --- SCD Type 1

```
INSERT INTO Suppliers (supplier_id, name, contact_name, phone, address, supplierUpdatedDate) VALUES
(4, 'Bakers Delight', 'Nora Special Bates', '111-1111', '88 Baker Rd', '2025-02-22 00:00:00');
```

----Insert new data into Employees table ---- SCD Type 2

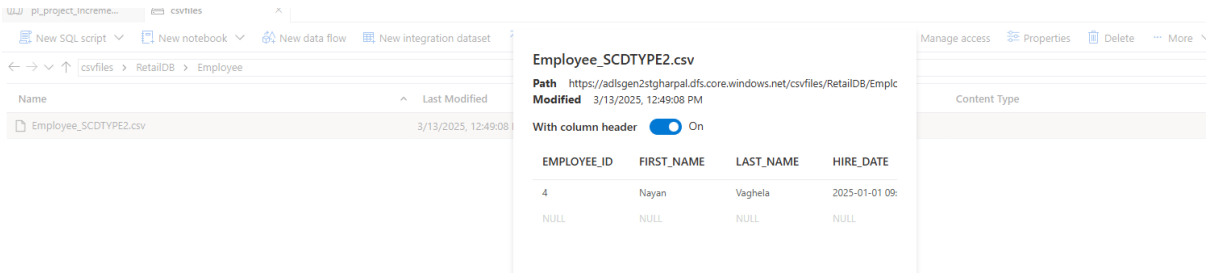
```
INSERT INTO Employee (employee_id, first_name, last_name, hire_date, last_review_date, role) VALUES
(4, 'Nayan', 'Vaghela', '2025-01-01 09:00:00', '2025-02-20 00:00:00', 'Inventory Specialist');
```

**Debug the pipeline**

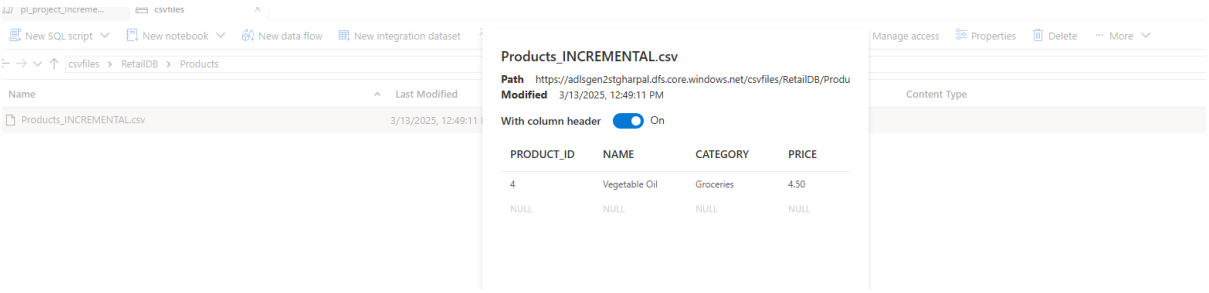



Now, let's check the data in the ADLS Gen 2 storage account

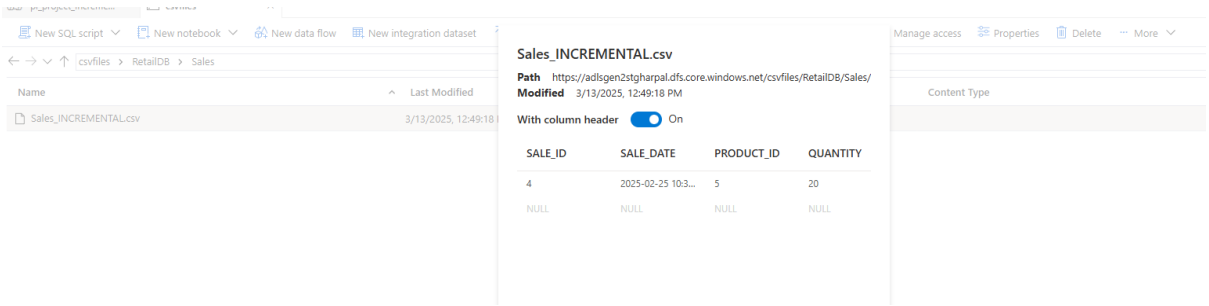
Employee Data



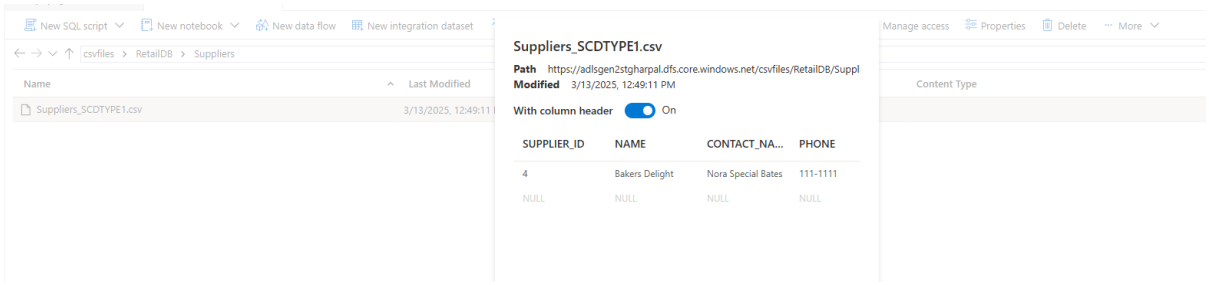

## Product Data




## Sales Data



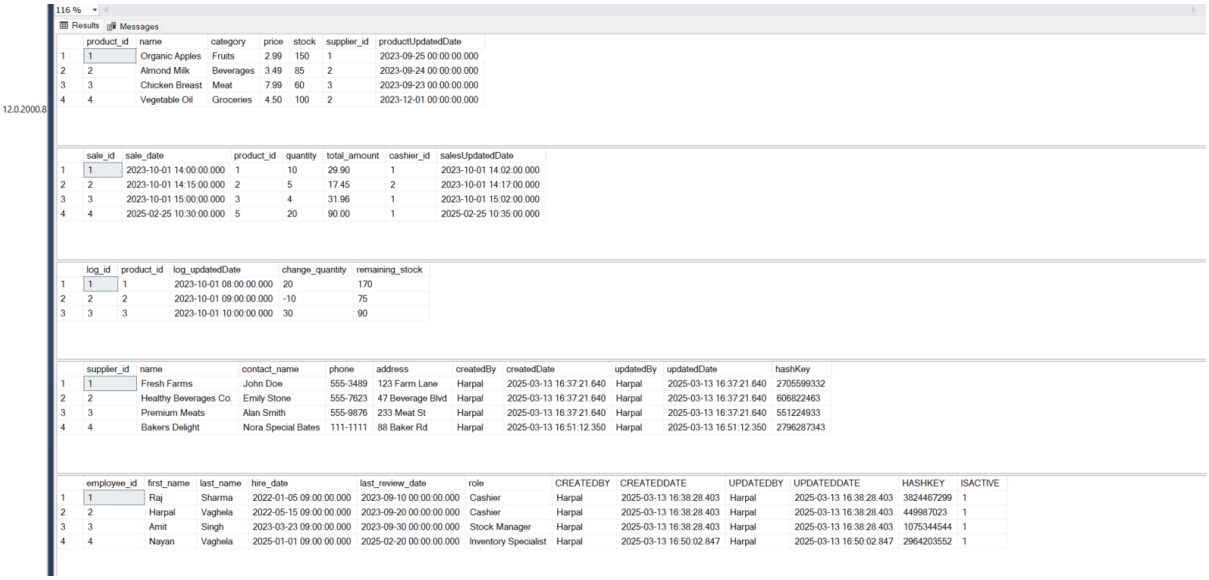

## Supplier Data




Let's check the output of Azure SQL tables

**Select * From tbl_Products**
**Select * From tbl_Sales**
**Select * From tbl_Inventory_Logs**

**Select * from tbl_Suppliers order by 1**
**Select * from tbl_Employee order by 1**



116 %

Results Messages

| | product_id | name | category | price | stock | supplier_id | productUpdatedDate |
|---|---|---|---|---|---|---|---|
| 1 | 1 | Organic Apples | Fruits | 2.99 | 150 | 1 | 2023-09-25 00:00:00.000 |
| 2 | 2 | Almond Milk | Beverages | 3.49 | 85 | 2 | 2023-09-24 00:00:00.000 |
| 3 | 3 | Chicken Breast | Meat | 7.99 | 60 | 3 | 2023-09-23 00:00:00.000 |
| 4 | 4 | Vegetable Oil | Groceries | 4.50 | 100 | 2 | 2023-12-01 00:00:00.000 |

12.0.2000.8

| | sale_id | sale_date | product_id | quantity | total_amount | cashier_id | salesUpdatedDate |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 2023-10-01 14:00:00.000 | 1 | 10 | 29.90 | 1 | 2023-10-01 14:02:00.000 |
| 2 | 2 | 2023-10-01 14:15:00.000 | 2 | 5 | 17.45 | 2 | 2023-10-01 14:17:00.000 |
| 3 | 3 | 2023-10-01 15:00:00.000 | 3 | 4 | 31.96 | 1 | 2023-10-01 15:02:00.000 |
| 4 | 4 | 2025-02-25 10:30:00.000 | 5 | 20 | 90.00 | 1 | 2025-02-25 10:35:00.000 |

| | log_id | product_id | log_updatedDate | change_quantity | remaining_stock |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 2023-10-01 08:00:00.000 | 20 | 170 |
| 2 | 2 | 2 | 2023-10-01 09:00:00.000 | -10 | 75 |
| 3 | 3 | 3 | 2023-10-01 10:00:00.000 | 30 | 90 |

| | supplier_id | name | contact_name | phone | address | createdBy | createdDate | updatedBy | updatedDate | hashKey |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | Fresh Farms | John Doe | 555-3489 | 123 Farm Lane | Harpal | 2025-03-13 16:37:21.640 | Harpal | 2025-03-13 16:37:21.640 | 2705599332 |
| 2 | 2 | Healthy Beverages Co. | Emily Stone | 555-7623 | 47 Beverage Blvd | Harpal | 2025-03-13 16:37:21.640 | Harpal | 2025-03-13 16:37:21.640 | 606822463 |
| 3 | 3 | Premium Meats | Alan Smith | 555-9876 | 233 Meat St | Harpal | 2025-03-13 16:37:21.640 | Harpal | 2025-03-13 16:37:21.640 | 551224933 |
| 4 | 4 | Bakers Delight | Nora Special Bates | 111-1111 | 88 Baker Rd | Harpal | 2025-03-13 16:51:12.350 | Harpal | 2025-03-13 16:51:12.350 | 2796287343 |

| | employee_id | first_name | last_name | hire_date | last_review_date | role | CREATEDBY | CREATEDDATE | UPDATEDBY | UPDATEDDATE | HASHKEY | ISACTIVE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | Raj | Sharma | 2022-01-05 09:00:00.000 | 2023-09-10 00:00:00.000 | Cashier | Harpal | 2025-03-13 16:38:28.403 | Harpal | 2025-03-13 16:38:28.403 | 3824467299 | 1 |
| 2 | 2 | Harpal | Vaghela | 2022-05-15 09:00:00.000 | 2023-09-20 00:00:00.000 | Cashier | Harpal | 2025-03-13 16:38:28.403 | Harpal | 2025-03-13 16:38:28.403 | 449987023 | 1 |
| 3 | 3 | Amit | Singh | 2023-03-23 09:00:00.000 | 2023-09-30 00:00:00.000 | Stock Manager | Harpal | 2025-03-13 16:38:28.403 | Harpal | 2025-03-13 16:38:28.403 | 1075344544 | 1 |
| 4 | 4 | Nayan | Vaghela | 2025-01-01 09:00:00.000 | 2025-02-20 00:00:00.000 | Inventory Specialist | Harpal | 2025-03-13 16:50:02.847 | Harpal | 2025-03-13 16:50:02.847 | 2964203552 | 1 |

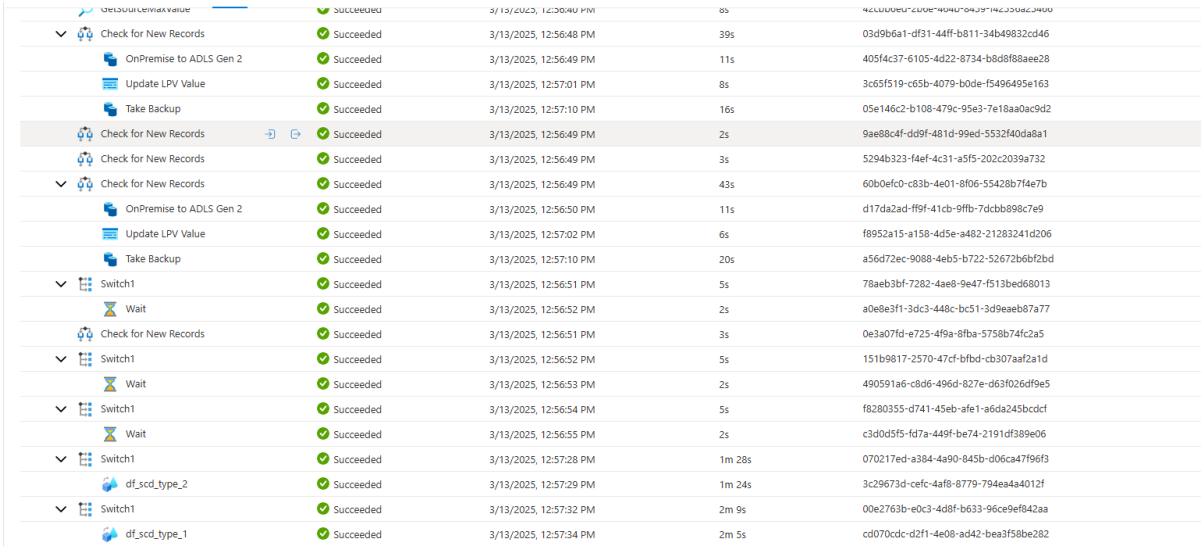
Let’s test the SCD Type 1 and 2 logic by modifying the data

```
133
134  ----Insert new data into Suppliers table --- SCD Type 1 -- 2nd Round
135  INSERT INTO Suppliers (supplier_id, name, contact_name, phone, address, supplierUpdatedDate) VALUES
136  (4, 'Bakers Magic Delight', 'Nora Magic Bates', '555-2222', '88 Baker Rd', '2025-02-23 00:00:00');
137
138  ----Update Employee Table
139  Update Employee
140  Set first_name = 'Nayansinh', last_name = 'Patel', role = 'Inventory Manager', last_review_date = '2025-02-27 00:00:00'
141  Where employee_id = 4;
142
143
144  -------------------------------------
```

----Insert new data into Suppliers table --- SCD Type 1 -- 2nd Round
INSERT INTO Suppliers (supplier_id, name, contact_name, phone, address, supplierUpdatedDate) VALUES
(4, 'Bakers Magic Delight', 'Nora Magic Bates', '555-2222', '88 Baker Rd', '2025-02-23 00:00:00');

----Update Employee Table
Update Employee
Set first_name = 'Nayansinh', last_name = 'Patel', role = 'Inventory Manager', last_review_date = '2025-02-27 00:00:00'
Where employee_id = 4;

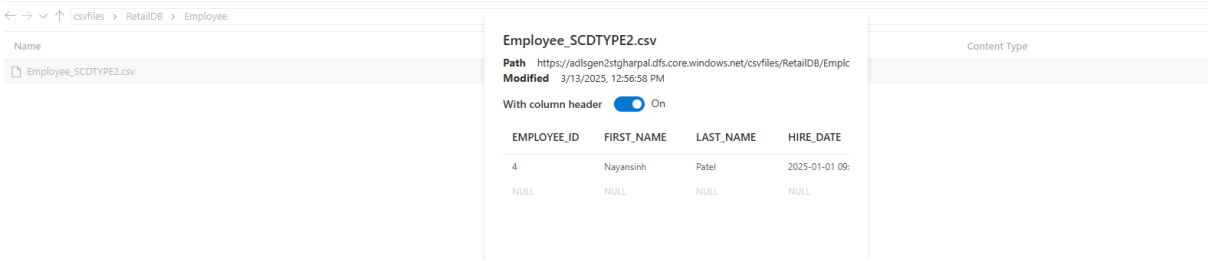
**Debug the Pipeline**



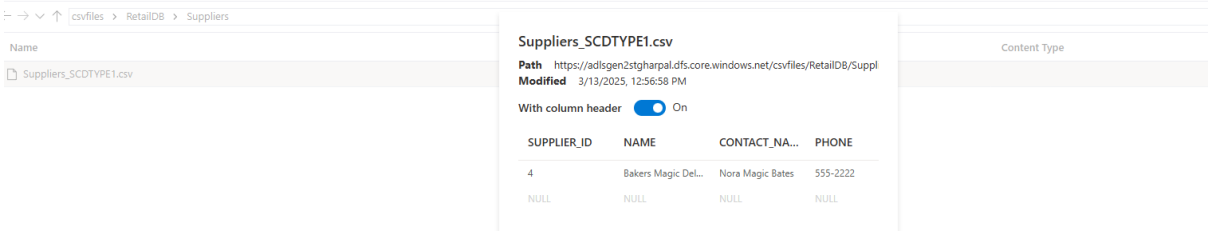
| | | | | |
|---|---|---|---|---|
| Check for New Records | Succeeded | 3/13/2025, 12:56:48 PM | 39s | 03d9b6a1-df31-44ff-b811-34b49832cd46 |
| OnPremise to ADLS Gen 2 | Succeeded | 3/13/2025, 12:56:49 PM | 11s | 405f4c37-6105-4d22-8734-b8d8f88aee28 |
| Update LPV Value | Succeeded | 3/13/2025, 12:57:01 PM | 8s | 3c65f519-c65b-4079-b0de-f5496495e163 |
| Take Backup | Succeeded | 3/13/2025, 12:57:10 PM | 16s | 05e146c2-b108-479c-95e3-7e18aa0ac9d2 |
| Check for New Records | Succeeded | 3/13/2025, 12:56:49 PM | 2s | 9ae88c4f-dd9f-481d-99ed-5532f40da8a1 |
| Check for New Records | Succeeded | 3/13/2025, 12:56:49 PM | 3s | 5294b323-f4ef-4c31-a5f5-202c2039a732 |
| Check for New Records | Succeeded | 3/13/2025, 12:56:49 PM | 43s | 60b0efc0-c83b-4e01-8f06-55428b7f4e7b |
| OnPremise to ADLS Gen 2 | Succeeded | 3/13/2025, 12:56:50 PM | 11s | d17da2ad-ff9f-41cb-9ffb-7dcbb898c7e9 |
| Update LPV Value | Succeeded | 3/13/2025, 12:57:02 PM | 6s | f8952a15-a158-4d5e-a482-21283241d206 |
| Take Backup | Succeeded | 3/13/2025, 12:57:10 PM | 20s | a56d72ec-9088-4eb5-b722-52672b6bf2bd |
| Switch1 | Succeeded | 3/13/2025, 12:56:51 PM | 5s | 78aeb3bf-7282-4ae8-9e47-f513bed68013 |
| Wait | Succeeded | 3/13/2025, 12:56:52 PM | 2s | a0e8e3f1-3dc3-448c-bc51-3d9eaeb87a77 |
| Check for New Records | Succeeded | 3/13/2025, 12:56:51 PM | 3s | 0e3a07fd-e725-4f9a-8fba-5758b74fc2a5 |
| Switch1 | Succeeded | 3/13/2025, 12:56:52 PM | 5s | 151b9817-2570-47cf-bfbd-cb307aaf2a1d |
| Wait | Succeeded | 3/13/2025, 12:56:53 PM | 2s | 490591a6-c8d6-496d-827e-d63f026df9e5 |
| Switch1 | Succeeded | 3/13/2025, 12:56:54 PM | 5s | f8280355-d741-45eb-afe1-a6da245bcdcf |
| Wait | Succeeded | 3/13/2025, 12:56:55 PM | 2s | c3d0d5f5-fd7a-449f-be74-2191df389e06 |
| Switch1 | Succeeded | 3/13/2025, 12:57:28 PM | 1m 28s | 070217ed-a384-4a90-845b-d06ca47f96f3 |
| df_scd_type_2 | Succeeded | 3/13/2025, 12:57:29 PM | 1m 24s | 3c29673d-cefc-4af8-8779-794ea4a4012f |
| Switch1 | Succeeded | 3/13/2025, 12:57:32 PM | 2m 9s | 00e2763b-e0c3-4d8f-b633-96ce9ef842aa |
| df_scd_type_1 | Succeeded | 3/13/2025, 12:57:34 PM | 2m 5s | cd070cdc-d2f1-4e08-ad42-bea3f58be282 |

## Check the data in the ADLS Gen 2 storage account

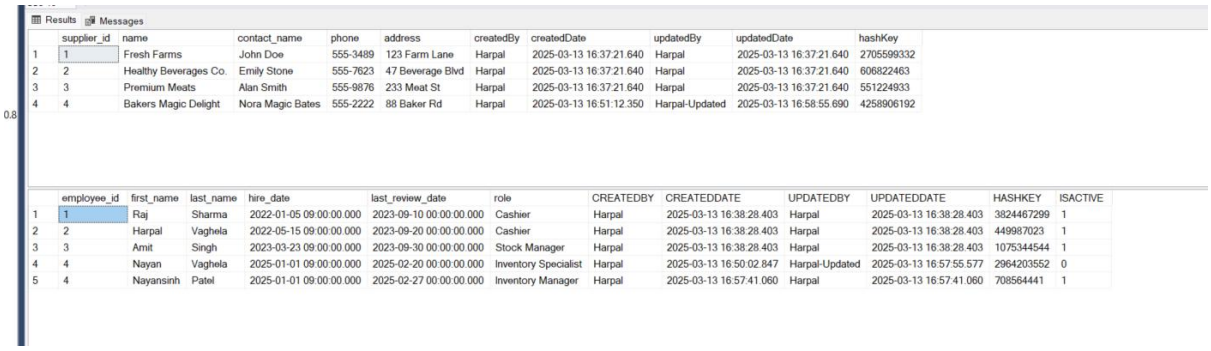
Employee Data




Supplier Data




Check the data in Azure SQL Tables



Results Messages

| | supplier_id | name | contact_name | phone | address | createdBy | createdDate | updatedBy | updatedDate | hashKey |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | Fresh Farms | John Doe | 555-3489 | 123 Farm Lane | Harpal | 2025-03-13 16:37:21.640 | Harpal | 2025-03-13 16:37:21.640 | 2705599332 |
| 2 | 2 | Healthy Beverages Co. | Emily Stone | 555-7623 | 47 Beverage Blvd | Harpal | 2025-03-13 16:37:21.640 | Harpal | 2025-03-13 16:37:21.640 | 606822463 |
| 3 | 3 | Premium Meats | Alan Smith | 555-9876 | 233 Meat St | Harpal | 2025-03-13 16:37:21.640 | Harpal | 2025-03-13 16:37:21.640 | 551224933 |
| 4 | 4 | Bakers Magic Delight | Nora Magic Bates | 555-2222 | 88 Baker Rd | Harpal | 2025-03-13 16:51:12.350 | Harpal-Updated | 2025-03-13 16:58:55.690 | 4258906192 |

| | employee_id | first_name | last_name | hire_date | last_review_date | role | CREATEDBY | CREATEDDATE | UPDATEDBY | UPDATEDDATE | HASHKEY | ISACTIVE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | Raj | Sharma | 2022-01-05 09:00:00.000 | 2023-09-10 00:00:00.000 | Cashier | Harpal | 2025-03-13 16:38:28.403 | Harpal | 2025-03-13 16:38:28.403 | 3824467299 | 1 |
| 2 | 2 | Harpal | Vaghela | 2022-05-15 09:00:00.000 | 2023-09-20 00:00:00.000 | Cashier | Harpal | 2025-03-13 16:38:28.403 | Harpal | 2025-03-13 16:38:28.403 | 449987023 | 1 |
| 3 | 3 | Amit | Singh | 2023-03-23 09:00:00.000 | 2023-09-30 00:00:00.000 | Stock Manager | Harpal | 2025-03-13 16:38:28.403 | Harpal | 2025-03-13 16:38:28.403 | 1075344544 | 1 |
| 4 | 4 | Nayan | Vaghela | 2025-01-01 09:00:00.000 | 2025-02-20 00:00:00.000 | Inventory Specialist | Harpal | 2025-03-13 16:50:02.847 | Harpal-Updated | 2025-03-13 16:57:55.577 | 2964203552 | 0 |
| 5 | 4 | Nayansinh | Patel | 2025-01-01 09:00:00.000 | 2025-02-27 00:00:00.000 | Inventory Manager | Harpal | 2025-03-13 16:57:41.060 | Harpal | 2025-03-13 16:57:41.060 | 708564441 | 1 |

We got the expected output in these tables.

Thank you

You can connect with me on these profiles:

**LinkedIn**: https://www.linkedin.com/in/harpalvaghela/

**Medium Blog**: https://medium.com/@harpalvaghela

**Website**: https://www.harpalvaghela.com/