

Data Migration Project



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

Data Migration in Azure

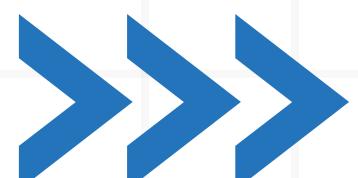
The process of moving data from one storage system to another within the Azure cloud environment, optimizing for accessibility, scalability, and performance.

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.

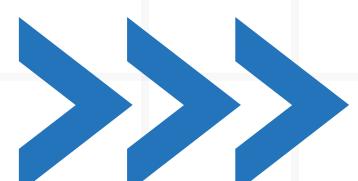


Best practices to follow for this data migration project

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:

```

7  -----
8  -----SCD Type 1-----
9  CREATE TABLE Suppliers (
10    supplier_id INT,
11    name VARCHAR(255),
12    contact_name VARCHAR(255),
13    phone VARCHAR(50),
14    address Varchar(100),
15    supplierUpdatedDate datetime -----Delta Column
16  );
17
18 CREATE TABLE Products (
19    product_id INT,
20    name VARCHAR(255),
21    category VARCHAR(100),
22    price DECIMAL(10, 2),
23    stock INT,
24    supplier_id INT,
25    productUpdatedDate datetime -----Delta Column
26  );
27
28 -----SCD Type 2-----
29 CREATE TABLE Employee (
30    employee_id INT,
31    first_name VARCHAR(255),
32    last_name VARCHAR(255),
33    hire_date DATETIME,
34    last_review_date DATETIME,   -----Delta Column
35    role VARCHAR(100)
36  );
37
38 CREATE TABLE Sales (
39    sale_id INT,
40    sale_date DATETIME,
41    product_id INT,
42    quantity INT,
43    total_amount DECIMAL(10, 2),
44    cashier_id INT,
45    salesUpdatedDate datetime -----Delta Column
46  );
47
48 CREATE TABLE Inventory_Logs (
49    log_id INT,
50    product_id INT,
51    log_updatedDate DATETIME,   -----Delta Column
52    change_quantity INT,
53    remaining_stock INT
54  );

```

-----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

```
56 ----Insert data into Supplier Table
57 INSERT INTO Suppliers (supplier_id, name, contact_name, phone, address, supplierUpdatedDate) VALUES
58 (1, 'Fresh Farms', 'John Doe', '555-3489', '123 Farm Lane', '2023-11-11 00:00:00'),
59 (2, 'Healthy Beverages Co.', 'Emily Stone', '555-7623', '47 Beverage Blvd', '2023-11-11 00:00:00'),
60 (3, 'Premium Meats', 'Alan Smith', '555-9876', '233 Meat St', '2023-11-11 00:00:00');
61
62 ----Insert data into Products Table
63 INSERT INTO Products (product_id, name, category, price, stock, supplier_id, productUpdatedDate) VALUES
64 (1, 'Organic Apples', 'Fruits', 2.99, 150, 1, '2023-09-25 00:00:00'),
65 (2, 'Almond Milk', 'Beverages', 3.49, 85, 2, '2023-09-24 00:00:00'),
66 (3, 'Chicken Breast', 'Meat', 7.99, 60, 3, '2023-09-23 00:00:00');
67
68 ----Insert data into Employee Table
69 INSERT INTO Employee (employee_id, first_name, last_name, hire_date, last_review_date, role) VALUES
70 (1, 'Raj', 'Sharma', '2022-01-05 09:00:00', '2023-09-10 00:00:00', 'Cashier'),
71 (2, 'Harpal', 'Vaghela', '2022-05-15 09:00:00', '2023-09-20 00:00:00', 'Cashier'),
72 (3, 'Amit', 'Singh', '2023-03-23 09:00:00', '2023-09-30 00:00:00', 'Stock Manager');
73
74 ----Insert data into Sales Table
75 INSERT INTO Sales (sale_id, sale_date, product_id, quantity, total_amount, cashier_id, salesUpdatedDate) VALUES
76 (1, '2023-10-01 14:00:00', 1, 10, 29.90, 1, '2023-10-01 14:02:00'),
77 (2, '2023-10-01 14:15:00', 2, 5, 17.45, 2, '2023-10-01 14:17:00'),
78 (3, '2023-10-01 15:00:00', 3, 4, 31.96, 1, '2023-10-01 15:02:00');
79
80 ----Insert data into Inventory_Logs Table
81 INSERT INTO Inventory_Logs (log_id, product_id, log_updatedDate, change_quantity, remaining_stock) VALUES
82 (1, 1, '2023-10-01 08:00:00', 20, 170),
83 (2, 2, '2023-10-01 09:00:00', -10, 75),
84 (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

```
93  CREATE TABLE dbo.WATERMARK(
94    ID INT IDENTITY(1,1),
95    TABLE_NAME VARCHAR(100),
96    SCHEMA_NAME VARCHAR(100),
97    FOLDER_NAME VARCHAR(100),
98    LPV VARCHAR(100),
99    DELTA_COLUMN VARCHAR(100),
100   TABLE_TYPE VARCHAR(100)
101  )
102
103
104
105  Select * From WATERMARK
106
107  -----TableName-----SchemaName-----FolderName-----LPV-----DeltaColumn-----Table Type-----
108  INSERT INTO dbo.WATERMARK VALUES ('Products', 'dbo', 'RetailDB/Products', '1900-01-01 00:00:00', 'productUpdatedDate', 'INCREMENTAL');
109  INSERT INTO dbo.WATERMARK VALUES ('Sales', 'dbo', 'RetailDB/Sales', '1900-01-01 00:00:00', 'salesUpdatedDate', 'INCREMENTAL');
110  INSERT INTO dbo.WATERMARK VALUES ('Employee', 'dbo', 'RetailDB/Employee', '1900-01-01 00:00:00', 'last_review_date', 'SCDTYPE2');
111  INSERT INTO dbo.WATERMARK VALUES ('Suppliers', 'dbo', 'RetailDB/Suppliers', '1900-01-01 00:00:00', 'supplierUpdatedDate', 'SCDTYPE1');
112  INSERT INTO dbo.WATERMARK VALUES ('Inventory_Logs', 'dbo', 'RetailDB/InventoryLogs', '1900-01-01 00:00:00', 'log_updatedDate', 'INCREMENTAL');
113
114
```

```
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
@TableName VARCHAR(100),
@LPV_Value VARCHAR(50)
AS
UPDATE WATERMARK
SET LPV = @LPV_Value
WHERE TABLE_NAME = @TableName
```

Check the data in all tables

RetailDB Schema Overview						
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
Detailed Data for Products Table						
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	
Detailed Data for Suppliers Table						
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
Detailed Data for Employee Table						
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	
Detailed Data for Sales Table						
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
Detailed Data for Inventory_Logs Table						
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

Name	Type	Sub-type	Status	Related	Region	Version
AutoResolveIntegrationRuntime	Azure	Public	Running	8	Auto Resolve	---
SelfhostedIR		Self-Hosted	Running	4	---	5.45.8999.1

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>

Self-hosted node is connected to the cloud service

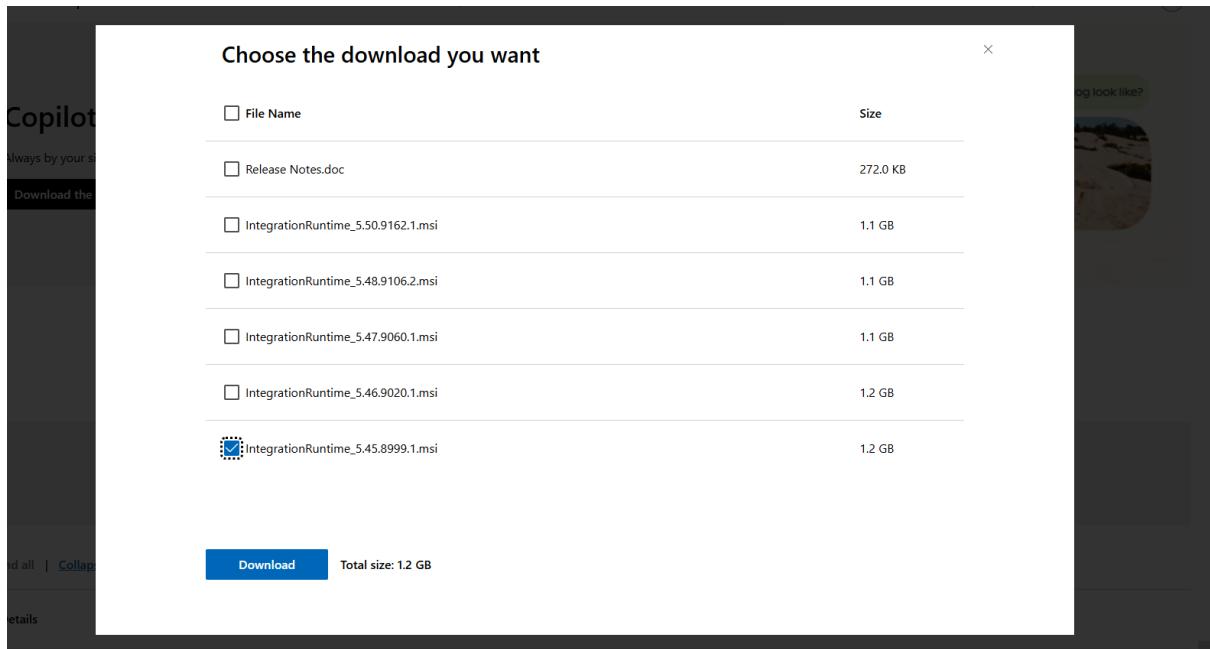
Data Factory: wsp-synapse-harpal
Integration Runtime: SelfhostedIR
Node: HARPAL

Data Source Credential

Credential store: On-premises
Credential status: In sync
Last backup time: N/A

Generate Backup Import Backup

Connected to the cloud service (Data Factory V2)



Let's create the pipeline

Azure Home -> Synapse Workspace -> New Pipeline

Drag and drop Lookup activity, rename as “Watermark”

General Settings User properties

Name * Watermark Learn more ↗

Description

Activity state ⓘ Activated Deactivated

Timeout ⓘ 0.12:00:00

Retry ⓘ 0

Retry interval (sec) ⓘ 30

Secure output ⓘ

Secure input ⓘ

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.

Edit linked service

SQL server [Learn more](#)

Name *
ls_sql_server_sh

Description

Connect via integration runtime * ⓘ
 SelfhostedIR

Version
 Recommended Legacy

Connection string [Azure Key Vault](#)

Server name *
HARPAL

Database name *
NCPL

Authentication type
SQL authentication

User name *
Harpal11

Password [Azure Key Vault](#)
Password *
.....

Always encrypted ⓘ

Parameters

Connection successful ⓘ [Test connection](#)

Apply **Cancel**

Take two parameters:

Parameters				
New	Delete	Name	Type	Default value
<input type="checkbox"/>	<input type="checkbox"/>	SchemaName	String	Value
<input type="checkbox"/>	<input type="checkbox"/>	TableName	String	Value

Go to the Connection tab and assign those parameters

Connection Schema Parameters

Linked service * ls_sql_server_sh Test connection Edit + New Learn more

Integration runtime * SelfhostedIR Edit

Table @dataset().SchemaName . @dataset().TableName Preview data

Enter manually

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

General Settings User properties

Source dataset * ds_sql_server_sh Open + New Preview data Learn more

Dataset properties

Name	Value
SchemaName	dbo
TableName	WATERMARK

First row only

Use query Table Query Stored procedure

Query timeout (minutes) 120

Isolation level Select...

Partition option None Physical partitions of table Dynamic range

Note: Please preview data to validate the partition settings.

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

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



General Settings Activities (4) User properties

Sequential

Batch count

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

Click on the Edit icon on ForEach Activity

Take Lookup activity and rename “GetSource.MaxValue”

General Settings User properties

Name * [Learn more](#)

Description

Activity state Activated Deactivated

Timeout

Retry

Retry interval (sec)

Secure output

Secure input

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

General **Settings** User properties

Source dataset * [Open](#) [New](#) [Preview data](#) [Learn more](#)

Dataset properties

Name	Value
SchemaName	<input type="text"/>
TableName	<input type="text"/>

First row only

Use query Table Query Stored procedure

Query [Delete](#)

Query timeout (minutes)

Isolation level

Partition option None Physical partitions of table Dynamic range

[Please preview data to validate the partition settings.](#)

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

Take If Condition and write this expression

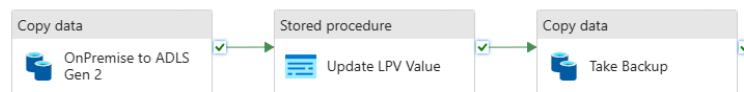


```
@less(string(item().LPV),  
string(activity('GetSourceMaxValue').output.firstRow.MaxValue))
```

Click the edit icon on the True Conditions part

Take these activities

pl_project_Incremental_scd_types_1_2 > ForEach1 > Check for New Records > True 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}'
```

Harpalsinh Vaghela

General Source Sink Mapping Settings User properties

Source dataset * **ds_sql_server_sh** Open New Preview data Learn more

Dataset properties

Name	Value
SchemaName	..
TableName	..

Use query Query Table Stored procedure

Query: `SELECT * FROM @item().SCHEMA_N...`

Query timeout (minutes)

Isolation level

Partition option None Physical partitions of table Dynamic range

Tip: Please preview data to validate the partition settings.

Additional columns

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

Take two parameters

Connection Schema Parameters

+ New Delete

Name	Type	Default value
FolderName	String	Value
FileName	String	Value

Connection tab

Connection Schema Parameters

Linked service * **ls_adls_gen2_sh** Test connection Edit New Learn more

Integration runtime * **SelfhostedIR** Edit

File path `csvfiles / @dataset().FolderName / @dataset().FileName` Browse Preview data Detect format

Compression type **No compression**

Column delimiter **Comma (,)**

Row delimiter **Default (\r,\n, or \r\n)**

Encoding **Default(UTF-8)**

Quote character **Double quote ("")**

Escape character **Backslash (\)**

First row as header

Null value

Write this expression for FolderName

@item().FOLDER_NAME

And this below expression for FileName

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

General Source **Sink** Mapping Settings User properties

Sink dataset *

Dataset properties

Name	Value
FolderName	@item().FOLDER_NAME
FileName	@concat(item().TABLE_NAME,'_',item...

Copy behavior

Max concurrent connections

Block size (MB)

Metadata

Quote all text

File extension

Max rows per file

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,

General **Settings** User properties

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

Linked service *

Integration runtime *

Stored procedure name *

Stored procedure parameters

<input type="checkbox"/>	Name	Type	Value
<input type="checkbox"/>	LPV_Value	String	@activity('GetSource.MaxValue').output...
<input type="checkbox"/>	Table_Name	String	@item().TABLE_NAME

Add dynamic content [Alt+Shift+D]

For LPV Value,

Write this expression

```
@activity('GetSource.MaxValue').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

Parameters tab

Name	Type	Default value
FolderName	String	Value
FileName	String	Value

Connection tab

Connection tab settings:

- Linked service: AzureDataLakeStorage1
- Integration runtime: AutoResolveIntegrationRuntime
- File path: csvfiles / @dataset().FolderName / @dataset().FileName
- Compression type: No compression
- Column delimiter: Comma (,)
- Row delimiter: Default (\r\n, or \n)
- Encoding: Default(UTF-8)
- Quote character: Double quote ("")
- Escape character: Backslash (\)
- First row as header: checked
- Null value:

Source

Source tab settings:

- General tab selected.
- Source dataset: ds_csvfiles
- Dataset properties:
 - FolderName: @item().FOLDER_NAME
 - FileName: @concat(item().TABLE_NAME,'_',item...)
- File path type: File path in dataset (radio button selected).
- Filter by last modified: checked
- Recursively: checked
- Enable partitions discovery: unchecked
- Max concurrent connections:
- Skip line count:

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

The screenshot shows the 'Connection' tab of a sink configuration in Azure Data Factory. The 'Linked service' dropdown is set to 'AzureDataLakeStorage1'. The 'Integration runtime' dropdown is set to 'AutoResolveIntegrationRuntime'. The 'File path' field contains 'csvfiles' followed by two expression placeholders: '@dataset().FolderName' and '@dataset().FileName'. Other settings include 'No compression' for compression type, 'Comma (,) for column delimiter, 'Default (\r\n or \n\r)' for row delimiter, 'Default(UTF-8)' for encoding, 'Double quote (")' for quote character, 'Backslash (\)' for escape character, and 'First row as header' checked.

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

```
12
13 -----SCD Type 1 Table -----
14
15 CREATE TABLE tbl_Suppliers (
16     supplier_id INT,
17     name VARCHAR(255),
18     contact_name VARCHAR(255),
19     phone VARCHAR(50),
20     address Varchar(100),
21     createdBy varchar(100),
22     createdDate datetime,
23     updatedBy varchar(100),
24     updatedDate datetime,
25     hashKey Bigint
26 );
27
28 --TRUNCATE TABLE tbl_Suppliers
29 Select * from tbl_Suppliers
30
31 --Target Source Query in Dataflow SCD Type 1
32 Select supplier_id, hashKey from tbl_Suppliers;
33
34
35 CREATE TABLE tbl_Employee (
36     employee_id INT,
37     first_name VARCHAR(255),
38     last_name VARCHAR(255),
39     hire_date DATETIME,
40     last_review_date DATETIME,
41     role VARCHAR(100),
42     CREATEDBY VARCHAR(100),
43     CREATEDDATE DATETIME,
44     UPDATEDBY VARCHAR(100),
45     UPDATEDDATE DATETIME,
46     HASHKEY BIGINT,
47     ISACTIVE INT
48 );
49
50
51 --TRUNCATE TABLE tbl_Employee
52 Select * from tbl_Employee
53
54
55 --Target Source Query in Dataflow SCD Type 2
56 Select employee_id, HASHKEY from tbl_Employee where ISACTIVE = 1
57
58
59
60
```

```
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 -----

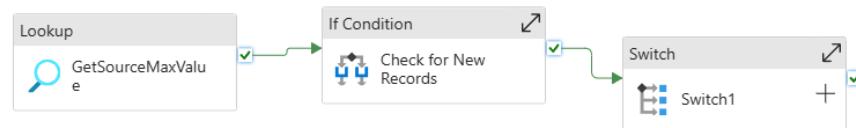
CREATE TABLE tbl_Suppliers (
    supplier_id INT,
    name VARCHAR(255),
    contact_name VARCHAR(255),
    phone VARCHAR(50),
    address Varchar(100),
    createdBy varchar(100),
    createdDate datetime,
    updatedBy varchar(100),
    updatedDate datetime,
    hashKey Bigint
);

--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**

pl_project_Incremental_scd_types_1_2 > ForEach1



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

The screenshot shows the Azure Data Flow pipeline editor. On the left, the 'Activities' sidebar is open, showing various categories like Synapse, Move and transform, Azure Data Explorer, etc. In the center, a 'ForEach1' activity is selected, which contains a 'Switch' activity. The 'Activities' tab in the 'ForEach1' activity details shows four cases: 'Default', 'scdtype2', 'incremental', and 'NoAction'. To the right, the 'Pipeline expression builder' window is open, displaying the following 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'
        )
    )
)

```

```

@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'
        )
    )
)

```

```
    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

General Activities (4) User properties

Expression ⓘ @if(and(eq(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')))))

+ Add case

Case ⓘ

Default

scdtype2

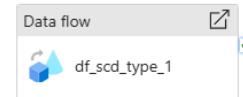
incremental

NoAction

Click on the Default case edit icon

Drag and drop Dataflow Activity

pl_project_Incremental_scd_types_1_2 > ForEach1 > Switch1 - Default



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

General **Settings** Parameters User properties

Data flow * Open New

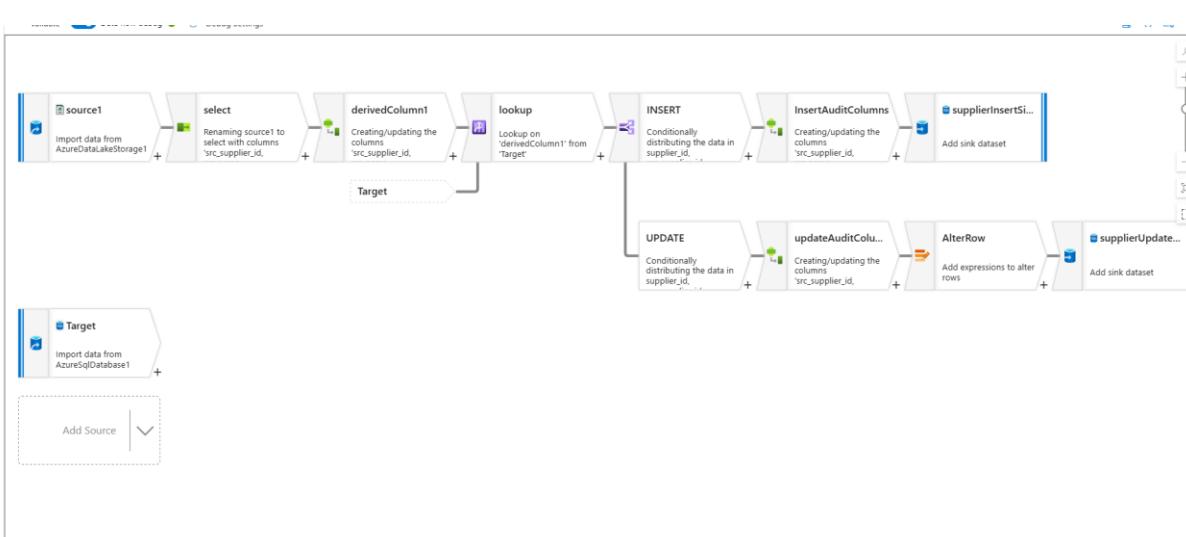
Run on (Azure IR) * Advanced

Compute size * Advanced

Logging level * Verbose Basic None

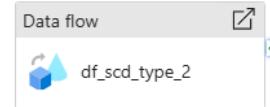
> Sink properties

> Staging



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

pl_project_Incremental_scd_types_1_2 > ForEach1 > Switch1 - scdtype2



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

General **Settings** Parameters User properties

Data flow * Open + New

Run on (Azure IR) * Advanced

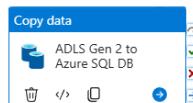
Compute size * Sink properties Staging

Logging level * Verbose Basic None

The screenshot shows the "Settings" tab of the Data Flow configuration. It lists the selected data flow as "df_scd_type_2_design", which runs on "AutoResolveIntegrationRuntime" with a "Small" compute size. The "Logging level" is set to "Verbose". Below this, the main data flow logic is displayed as a complex sequence of transformations: source (AzureDataLakeStorage1) -> select -> derivedColumn (renaming columns) -> lookup (lookup on derivedColumn1 from Target) -> INSERT (conditionally distributing data) -> Union (combining rows from transformation) -> InsertAuditColumns (creating/updating audit columns) -> insertEmployeeSink (add sink dataset). There are also UPDATE and alterFlow steps in the middle of the flow. A target section at the bottom shows the target (AzureSqlDatabase1) and an "Add Source" button.

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

pl_project_Incremental_scd_types_1_2 > ForEach1 > Switch1 - incremental



General Source Sink Mapping Settings User properties

Name * [Learn more](#)

Description

Activity state Activated Deactivated

Timeout

Retry

Retry interval (sec)

Secure output

Secure input

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

Connection Schema Parameters

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

Integration runtime * AutoResolveIntegrationRuntime [Edit](#)

File path / / [Browse](#) [Preview data](#)

Compression type

Column delimiter

Row delimiter

Encoding

Quote character

Escape character

First row as header

Null value

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

The screenshot shows the 'Source' tab selected in the top navigation bar. Under 'Source dataset *', a dropdown menu is set to 'ds_csvfiles'. To the right of the dropdown are buttons for 'Open', 'New', 'Preview data', and 'Learn more'. Below this, under 'Dataset properties', there are two fields: 'FolderName' and 'FileName', both currently showing the value '...'. Under 'File path type', the radio button for 'Wildcard file path' is selected, with the path 'csvfiles / @item().FOLDER_NAME / *_INCREMENTAL.csv' entered. Other settings include 'Start time (UTC)' and 'End time (UTC)' fields, a checked 'Recursively' checkbox, an unchecked 'Enable partitions discovery' checkbox, and a 'Max concurrent connections' field.

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

The screenshot shows the 'Connection' tab selected in the top navigation bar. Under 'Linked service *', the dropdown is set to 'AzureSqlDatabase1'. To the right are buttons for 'Test connection', 'Edit', 'New', and 'Learn more'. Under 'Integration runtime *', the dropdown is set to 'AutoResolveIntegrationRuntime'. Under 'Table', the schema and table names are defined as '@dataset().SchemaName' and '@dataset().TableName' respectively. A checked 'Enter manually' checkbox is also present. To the right of the table fields are buttons for 'Preview data'.

General Source **Sink** Mapping Settings User properties

Sink dataset * Open New Learn more

Dataset properties

Name	Value
SchemaName	@item().SCHEMA_NAME
TableName	@concat('tbl','_',item().TABLE_NAME)

Write behavior Insert Upsert Stored procedure

Bulk insert table lock Yes No

Table option Use existing Auto create table

Pre-copy script

Write batch timeout e.g. 00:30:00

Write batch size

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

pl_project_Incremental_scd_types_1_2 > ForEach1 > Switch1 - NoAction

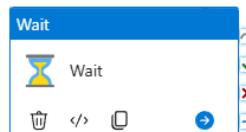
Wait

Wait

General

Case *

pl_project_Incremental_scd_types_1_2 > ForEach1 > Switch1 - NoAction



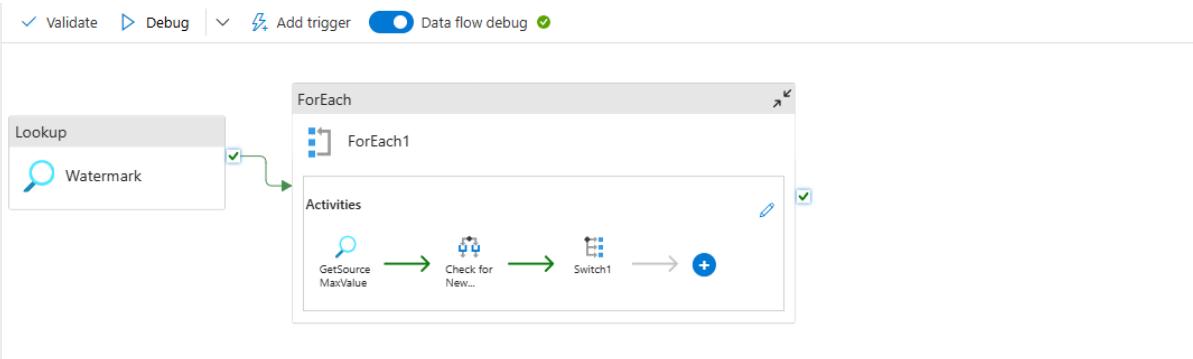
General **Settings** User properties

Wait time in seconds *

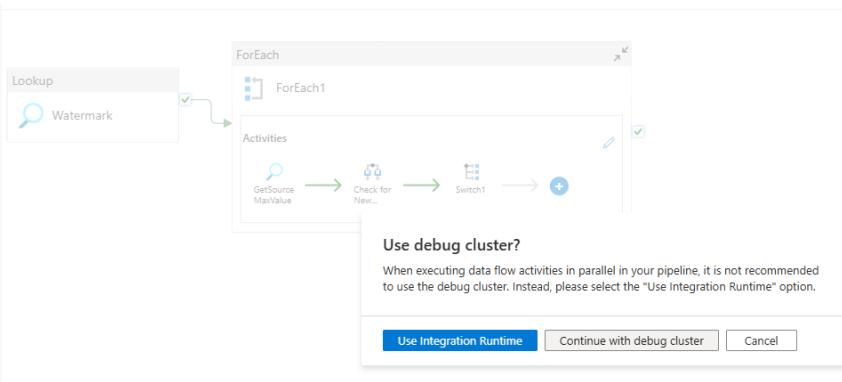
1

Let's run the pipeline

Enable **Data flow Debug** option



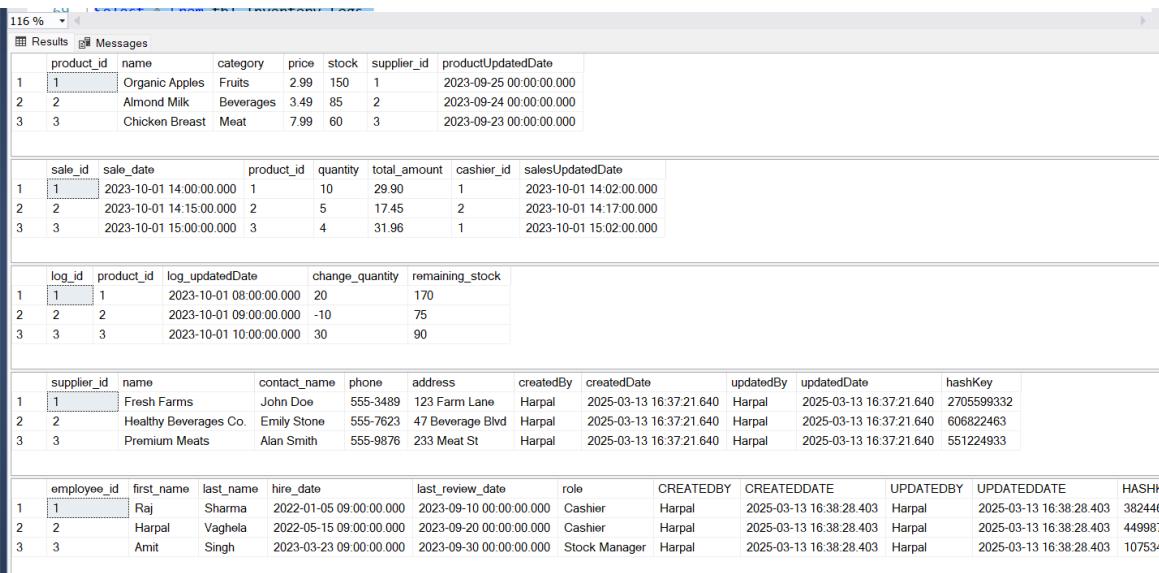
Click on **Continue with debug cluster**



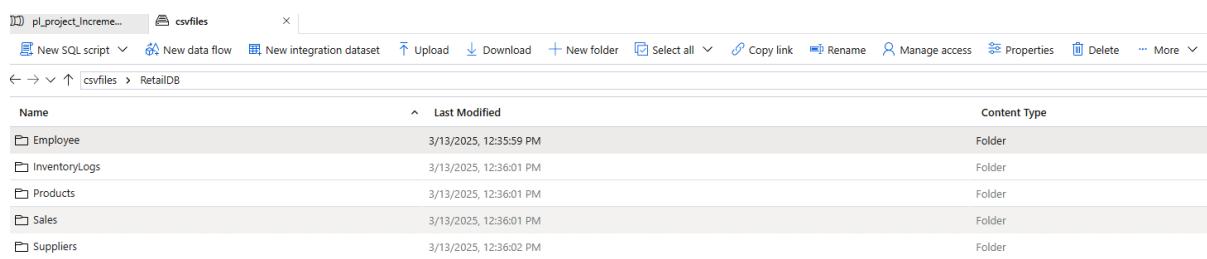
Harpalsinh Vaghela

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-b070-46c4-8673-8dbdefe08e45
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-4cfb-9724-0a30525b478d
OnPremise to ADLS Gen 2			Succeeded	3/13/2025, 12:35:45 PM	19s	9ee25d8b-3df1-4821-93af-6871252caaa3
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-2084-41e7-9a66-6cb2a4f57eb
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-4a24-8375-14922bf8f5ed
Check for New Records			Succeeded	3/13/2025, 12:35:45 PM	1m 2s	71189fd5-d600-4498-8664-0742899968
OnPremise to ADLS Gen 2			Succeeded	3/13/2025, 12:35:46 PM	19s	a25a9024-8b67-41a1-b481-f7ba6ca875b
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-4fcf-92e7-dea52970b255
Check for New Records			Succeeded	3/13/2025, 12:35:45 PM	1m 2s	a7f22c5a-bcef-4c59-9371-953f9e1867fe
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

Results																																																		
																																																		
<table border="1"> <thead> <tr> <th>product_id</th><th>name</th><th>category</th><th>price</th><th>stock</th><th>supplier_id</th><th>productUpdatedDate</th></tr> </thead> <tbody> <tr> <td>1</td><td>Organic Apples</td><td>Fruits</td><td>2.99</td><td>150</td><td>1</td><td>2023-09-25 00:00:00.000</td></tr> <tr> <td>2</td><td>Almond Milk</td><td>Beverages</td><td>3.49</td><td>85</td><td>2</td><td>2023-09-24 00:00:00.000</td></tr> <tr> <td>3</td><td>Chicken Breast</td><td>Meat</td><td>7.99</td><td>60</td><td>3</td><td>2023-09-23 00:00:00.000</td></tr> </tbody> </table>							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																
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																																												
<table border="1"> <thead> <tr> <th>sale_id</th><th>sale_date</th><th>product_id</th><th>quantity</th><th>total_amount</th><th>cashier_id</th><th>salesUpdatedDate</th></tr> </thead> <tbody> <tr> <td>1</td><td>2023-10-01 14:00:00.000</td><td>1</td><td>10</td><td>29.90</td><td>1</td><td>2023-10-01 14:02:00.000</td></tr> <tr> <td>2</td><td>2023-10-01 14:15:00.000</td><td>2</td><td>5</td><td>17.45</td><td>2</td><td>2023-10-01 14:17:00.000</td></tr> <tr> <td>3</td><td>2023-10-01 15:00:00.000</td><td>3</td><td>4</td><td>31.96</td><td>1</td><td>2023-10-01 15:02:00.000</td></tr> </tbody> </table>							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																
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																																												
<table border="1"> <thead> <tr> <th>log_id</th><th>product_id</th><th>log_updatedDate</th><th>change_quantity</th><th>remaining_stock</th></tr> </thead> <tbody> <tr> <td>1</td><td>1</td><td>2023-10-01 08:00:00.000</td><td>20</td><td>170</td></tr> <tr> <td>2</td><td>2</td><td>2023-10-01 09:00:00.000</td><td>-10</td><td>75</td></tr> <tr> <td>3</td><td>3</td><td>2023-10-01 10:00:00.000</td><td>30</td><td>90</td></tr> </tbody> </table>							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																								
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																																														
<table border="1"> <thead> <tr> <th>supplier_id</th><th>name</th><th>contact_name</th><th>phone</th><th>address</th><th>createdBy</th><th>createdAt</th><th>updatedBy</th><th>updatedAt</th><th>hashKey</th></tr> </thead> <tbody> <tr> <td>1</td><td>Fresh Farms</td><td>John Doe</td><td>555-3489</td><td>123 Farm Lane</td><td>Harpal</td><td>2025-03-13 16:37:21.640</td><td>Harpal</td><td>2025-03-13 16:37:21.640</td><td>270599332</td></tr> <tr> <td>2</td><td>Healthy Beverages Co.</td><td>Emily Stone</td><td>555-7623</td><td>47 Beverage Blvd</td><td>Harpal</td><td>2025-03-13 16:37:21.640</td><td>Harpal</td><td>2025-03-13 16:37:21.640</td><td>606822463</td></tr> <tr> <td>3</td><td>Premium Meats</td><td>Alan Smith</td><td>555-9876</td><td>233 Meat St</td><td>Harpal</td><td>2025-03-13 16:37:21.640</td><td>Harpal</td><td>2025-03-13 16:37:21.640</td><td>551224933</td></tr> </tbody> </table>							supplier_id	name	contact_name	phone	address	createdBy	createdAt	updatedBy	updatedAt	hashKey	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	270599332	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	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				
supplier_id	name	contact_name	phone	address	createdBy	createdAt	updatedBy	updatedAt	hashKey																																									
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	270599332																																									
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	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																																									
<table border="1"> <thead> <tr> <th>employee_id</th><th>first_name</th><th>last_name</th><th>hire_date</th><th>last_review_date</th><th>role</th><th>CREATEDBY</th><th>CREATEDDATE</th><th>UPDATEDBY</th><th>UPDATEDDATE</th><th>HASHKEY</th></tr> </thead> <tbody> <tr> <td>1</td><td>Raj</td><td>Sharma</td><td>2022-01-05 09:00:00.000</td><td>2023-09-10 00:00:00.000</td><td>Cashier</td><td>Harpal</td><td>2025-03-13 16:38:28.403</td><td>Harpal</td><td>2025-03-13 16:38:28.403</td><td>382446;</td></tr> <tr> <td>2</td><td>Harpal</td><td>Vaghela</td><td>2022-05-15 09:00:00.000</td><td>2023-09-20 00:00:00.000</td><td>Cashier</td><td>Harpal</td><td>2025-03-13 16:38:28.403</td><td>Harpal</td><td>2025-03-13 16:38:28.403</td><td>449987c</td></tr> <tr> <td>3</td><td>Amit</td><td>Singh</td><td>2023-03-23 09:00:00.000</td><td>2023-09-30 00:00:00.000</td><td>Stock Manager</td><td>Harpal</td><td>2025-03-13 16:38:28.403</td><td>Harpal</td><td>2025-03-13 16:38:28.403</td><td>107534c</td></tr> </tbody> </table>							employee_id	first_name	last_name	hire_date	last_review_date	role	CREATEDBY	CREATEDDATE	UPDATEDBY	UPDATEDDATE	HASHKEY	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	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	449987c	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	107534c
employee_id	first_name	last_name	hire_date	last_review_date	role	CREATEDBY	CREATEDDATE	UPDATEDBY	UPDATEDDATE	HASHKEY																																								
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	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	449987c																																								
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	107534c																																								

Let's check the data in the ADLS Gen 2 storage account

pl_project_Increme...																												
csvfiles																												
																												
<table border="1"> <thead> <tr> <th>Name</th><th>Last Modified</th><th>Content Type</th></tr> </thead> <tbody> <tr> <td>Employee</td><td>3/13/2025, 12:35:59 PM</td><td>Folder</td></tr> <tr> <td>InventoryLogs</td><td>3/13/2025, 12:36:01 PM</td><td>Folder</td></tr> <tr> <td>Products</td><td>3/13/2025, 12:36:01 PM</td><td>Folder</td></tr> <tr> <td>Sales</td><td>3/13/2025, 12:36:01 PM</td><td>Folder</td></tr> <tr> <td>Suppliers</td><td>3/13/2025, 12:36:02 PM</td><td>Folder</td></tr> </tbody> </table>											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
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

Harpalsinh Vaghela

The screenshot shows the Azure Data Explorer interface with the 'Employee' table selected. The table has columns: EMPLOYEE_ID, FIRST_NAME, LAST_NAME, and HIRE_DATE. The data is as follows:

EMPLOYEE_ID	FIRST_NAME	LAST_NAME	HIRE_DATE
1	Raj	Sharma	2022-01-05 09:00:00
2	Harpal	Vaghela	2022-05-15 09:00:00
3	Amit	Singh	2023-03-23 09:00:00
NULL	NULL	NULL	NULL

Inventory Log Data

The screenshot shows the Azure Data Explorer interface with the 'Inventory_Logs_INCREMENTAL' table selected. The table has columns: LOG_ID, PRODUCT_ID, LOG_UPDATE_DATE, and CHANGE_QUANTITY. The data is as follows:

LOG_ID	PRODUCT_ID	LOG_UPDATE_DATE	CHANGE_QUANTITY
1	1	2023-10-01 08:00:00	20
2	2	2023-10-01 09:00:00	-10
3	3	2023-10-01 10:00:00	30
NULL	NULL	NULL	NULL

Products Data

The screenshot shows the Azure Data Explorer interface with the 'Products_INCREMENTAL' table selected. The table has columns: PRODUCT_ID, NAME, CATEGORY, and PRICE. The data is as follows:

PRODUCT_ID	NAME	CATEGORY	PRICE
1	Organic Apples	Fruits	2.99
2	Almond Milk	Beverages	3.49
3	Chicken Breast	Meat	7.99
NULL	NULL	NULL	NULL

Sales Data

The screenshot shows the Azure Data Explorer interface with the 'Sales_INCREMENTAL' table selected. The table has columns: SALE_ID, SALE_DATE, PRODUCT_ID, and QUANTITY. The data is as follows:

SALE_ID	SALE_DATE	PRODUCT_ID	QUANTITY
1	2023-10-01 14:00:00	1	10
2	2023-10-01 14:15:00	2	5
3	2023-10-01 15:00:00	3	4
NULL	NULL	NULL	NULL

Supplier Data

The screenshot shows the Azure Data Explorer interface with the 'Suppliers_SCDTYPE1' table selected. The table has columns: SUPPLIER_ID, NAME, CONTACT_NAME, and PHONE. The data is as follows:

SUPPLIER_ID	NAME	CONTACT_NAME	PHONE
1	Fresh Farms	John Doe	555-3489
2	Healthy Beverage	Emily Stone	555-7623
3	Premium Meats	Alan Smith	555-9876
NULL	NULL	NULL	NULL

Lets check the backup container in storage account

Harpalsinh Vaghela

A screenshot of the Azure Data Explorer interface. The top navigation bar includes 'New SQL script', 'New data flow', 'New integration dataset', 'Upload', 'Download', 'New folder', 'Select all', 'Copy link', 'Rename', and 'Manage access'. The current path is 'csvfiles > backup'. A table below shows a single item: 'Name' (RetailDB) and 'Last Modified' (3/13/2025, 9:19:32 AM).

A screenshot of the Azure Data Explorer interface. The current path is 'csvfiles > backup > RetailDB'. A table shows the contents of the 'RetailDB' folder: 'Employee' (3/13/2025, 9:19:42 AM), 'InventoryLogs' (3/13/2025, 9:19:35 AM), 'Products' (3/13/2025, 9:19:32 AM), 'Sales' (3/13/2025, 9:19:35 AM), and 'Suppliers' (3/13/2025, 9:20:06 AM). The 'Content Type' column for all items shows 'Folder'.

Employee Data Backup

A screenshot of the Azure Data Explorer interface. The current path is 'csvfiles > backup > RetailDB > Employee'. A table shows a single item: 'Name' (Employee_2025-03-13T16:36:18.7263169Z.csv) and 'Last Modified' (3/13/2025, 12:36:34 PM).

A screenshot of the Azure Data Explorer interface. The current path is 'csvfiles > backup > RetailDB > Employee'. A table shows the content of the 'Employee_2025-03-13T16:36:18.7263169Z.csv' file. The file path is 'https://adlsgen2stgharopal.dfs.core.windows.net/csvfiles/backup/RetailDB/03-13T16:36:18.7263169Z.csv'. It was modified on 3/13/2025, 12:36:34 PM. The 'With column header' option is turned on. The data table has columns: EMPLOYEE_ID, FIRST_NAME, LAST_NAME, and HIRE_DATE. The data rows are:

EMPLOYEE_ID	FIRST_NAME	LAST_NAME	HIRE_DATE
1	Raj	Sharma	2022-01-05 09:00:00
2	Harpal	Vaghela	2022-05-15 09:00:00
3	Amit	Singh	2023-03-23 09:00:00
NULL	NULL	NULL	NULL

Inventory Data Backup

A screenshot of the Azure Data Explorer interface. The current path is 'csvfiles > backup > RetailDB > InventoryLogs'. A table shows a single item: 'Name' (Inventory_Logs_2025-03-13T16:36:22.9983517Z.csv) and 'Last Modified' (3/13/2025, 12:36:44 PM). The file path is 'https://adlsgen2stgharopal.dfs.core.windows.net/csvfiles/backup/RetailDB/03-13T16:36:22.9983517Z.csv'. It was modified on 3/13/2025, 12:36:44 PM. The 'With column header' option is turned on. The data table has columns: LOG_ID, PRODUCT_ID, LOG_UPDATE..., and CHANGE_QU. The data rows are:

LOG_ID	PRODUCT_ID	LOG_UPDATE...	CHANGE_QU
1	1	2023-10-01 08:00:00	20
2	2	2023-10-01 09:00:00	-10
3	3	2023-10-01 10:00:00	30
NULL	NULL	NULL	NULL

Product Data Backup

PRODUCT_ID	NAME	CATEGORY	PRICE
1	Organic Apples	Fruits	2.99
2	Almond Milk	Beverages	3.49
3	Chicken Breast	Meat	7.99
NULL	NULL	NULL	NULL

Sales Data Backup

SALE_ID	SALE_DATE	PRODUCT_ID	QUANTITY
1	2023-10-01 14:00:00	1	10
2	2023-10-01 14:10:00	2	5
3	2023-10-01 15:00:00	3	4
NULL	NULL	NULL	NULL

Supplier Data Backup

SUPPLIER_ID	NAME	CONTACT_NAME	PHONE
1	Fresh Farms	John Doe	555-3489
2	Healthy Beverage	Emily Stone	555-7623
3	Premium Meats	Alan Smith	555-9876
NULL	NULL	NULL	NULL

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

```

109 -----Insert new data into Products table
110 INSERT INTO Products (product_id, name, category, price, stock, supplier_id, productUpdatedDate) VALUES
111 (4, 'Vegetable Oil', 'Groceries', 4.50, 100, 2, '2023-12-01 00:00:00');
112
113 -----Insert new data into Sales table
114 INSERT INTO Sales (sale_id, sale_date, product_id, quantity, total_amount, cashier_id, salesUpdatedDate) VALUES
115 (4, '2025-02-25 10:30:00', 5, 20, 90.00, 1, '2025-02-25 10:35:00');
116
117
118 -----Insert new data into Suppliers table --- SCD Type 1
119 INSERT INTO Suppliers (supplier_id, name, contact_name, phone, address, supplierUpdatedDate) VALUES
120 (4, 'Bakers Delight', 'Nora Special Bakes', '111-1111', '88 Baker Rd', '2025-02-22 00:00:00');
121
122
123 -----Insert new data into Employees table ---- SCD Type 2
124 INSERT INTO Employee (employee_id, first_name, last_name, hire_date, last_review_date, role) VALUES
125 (4, 'Nayan', 'Vaghela', '2025-01-01 09:00:00', '2025-02-20 00:00:00', 'Inventory Specialist');
126
127
128
129

```

----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');
```

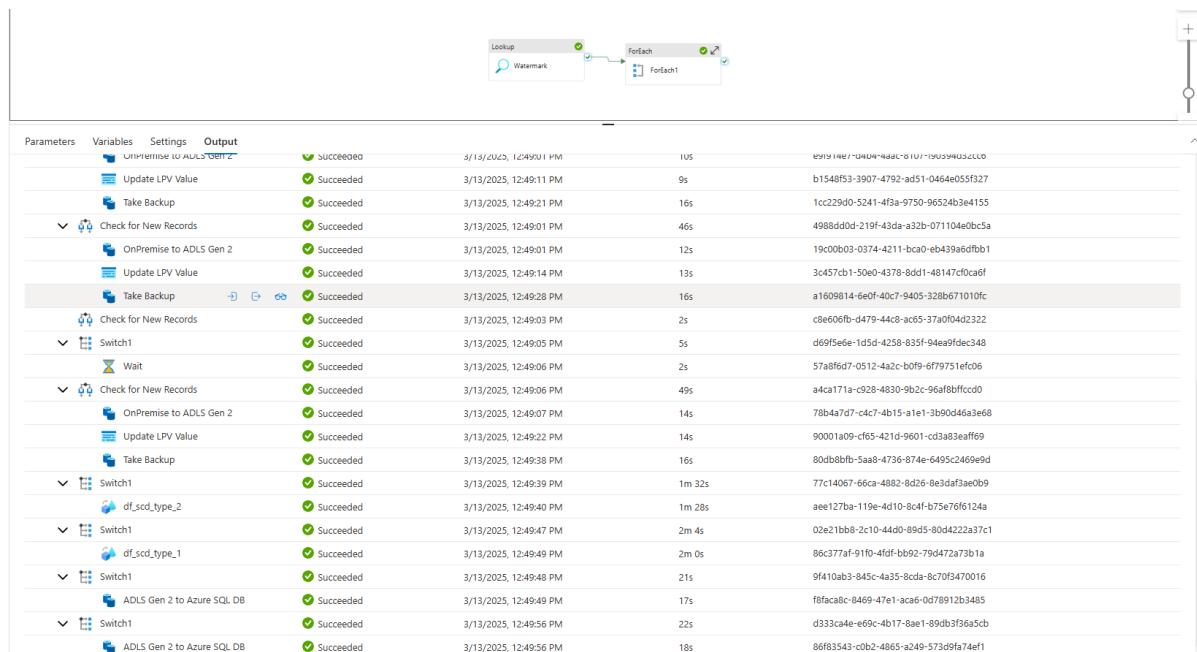
----Insert new data into Suppliers table --- SCD Type 1

```
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

The screenshot shows the Azure Data Explorer interface with the path `csvfiles > RetailDB > Employee`. A file named `Employee_SCDTYPE2.csv` is selected, last modified on 3/13/2025 at 12:49:08 PM. The table has four columns: `EMPLOYEE_ID`, `FIRST_NAME`, `LAST_NAME`, and `HIRE_DATE`. The data shows one row with values: 4, Nayan, Vaghela, and 2025-01-01 09:00:00.

Product Data

The screenshot shows the Azure Data Explorer interface with the path `csvfiles > RetailDB > Products`. A file named `Products_INCREMENTAL.csv` is selected, last modified on 3/13/2025 at 12:49:11 PM. The table has four columns: `PRODUCT_ID`, `NAME`, `CATEGORY`, and `PRICE`. The data shows one row with values: 4, Vegetable Oil, Groceries, and 4.50.

Sales Data

The screenshot shows the Azure Data Explorer interface with the path `csvfiles > RetailDB > Sales`. A file named `Sales_INCREMENTAL.csv` is selected, last modified on 3/13/2025 at 12:49:18 PM. The table has four columns: `SALE_ID`, `SALE_DATE`, `PRODUCT_ID`, and `QUANTITY`. The data shows one row with values: 4, 2025-02-25 10:3..., 5, and 20.

Supplier Data

The screenshot shows the Azure Data Explorer interface with the path `csvfiles > RetailDB > Suppliers`. A file named `Suppliers_SCDTYPE1.csv` is selected, last modified on 3/13/2025 at 12:49:11 PM. The table has four columns: `SUPPLIER_ID`, `NAME`, `CONTACT_NA...`, and `PHONE`. The data shows one row with values: 4, Bakers Delight, Nora Special Bates, and 111-1111.

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

The screenshot shows a database interface with four tables:

- Products:** Contains rows for Organic Apples, Almond Milk, Chicken Breast, and Vegetable Oil.
- Sales:** Contains rows for sales on 2023-10-01 at 14:00, 15:00, and 15:00, and a sale on 2025-02-25 at 10:30.
- Stock Log:** Contains rows for log entries on 2023-10-01 at 09:00, 10:00, and 10:00.
- Suppliers:** Contains rows for Fresh Farms, Healthy Beverages Co., Premium Meats, and Bakers Delight.

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

Harpalsinh Vaghela

	Resource Name	Status	Start Time	End Time	Duration	Log ID
✓	GetResourcePropertyValue	Succeeded	5/15/2020, 1:20:40 PM		8s	4e2e2e2e-e2e2-4e04-9494-747974e2e2e2
✓	Check for New Records	Succeeded	3/13/2025, 12:56:48 PM		39s	03d9b6a1-f311-4aff-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	3c65f159-cd5b-4079-b0de-f5496495e163
↳	Take Backup	Succeeded	3/13/2025, 12:57:10 PM		16s	05e146c2-b108-479c-95e3-7e18aa0ec9d2
↳	Check for New Records	Succeeded	3/13/2025, 12:56:49 PM		2s	9ae80cfd-d99f-481d-99ed-55240da8a1
↳	Check for New Records	Succeeded	3/13/2025, 12:56:49 PM		3s	5294b323-4fe1-4c31-a5f5-202c2039a732
✓	Check for New Records	Succeeded	3/13/2025, 12:56:49 PM		43s	60b0ef0-c83b-4e01-8f06-55428b74fe7b
↳	OnPremise to ADLS Gen 2	Succeeded	3/13/2025, 12:56:50 PM		11s	d17da2ad-ff9f-41cb-9ffb-7cbb88bc7e9
↳	Update LPV Value	Succeeded	3/13/2025, 12:57:02 PM		6s	f8952a15-a158-4d5e-a882-21283241d206
↳	Take Backup	Succeeded	3/13/2025, 12:57:10 PM		20s	a56d72ec-9088-4eb5-b722-526726bf2bd
✓	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-bc51-3d9eaeb7a77
↳	Check for New Records	Succeeded	3/13/2025, 12:56:51 PM		3s	0e3a07d-j725-4f9a-8fb8-5758b74fc2a5
✓	Switch1	Succeeded	3/13/2025, 12:56:52 PM		5s	151b9817-2570-47cf-bfbd-cb307af2a1d
↳	Wait	Succeeded	3/13/2025, 12:56:53 PM		2s	490591a6-c8d0-496d-827e-d63f02d69e5
✓	Switch1	Succeeded	3/13/2025, 12:56:54 PM		5s	f820355-d741-45eb-afe1-a6da245bcdcf
↳	Wait	Succeeded	3/13/2025, 12:56:55 PM		2s	c3d005f-fd78-449f-be74-219d1f389e06
✓	Switch1	Succeeded	3/13/2025, 12:57:28 PM		1m 28s	070217ed-a384-4980-845b-d06ca47f96f
↳	df_scdf_type_2	Succeeded	3/13/2025, 12:57:29 PM		1m 24s	3c29673d-cefc-4af8-8779-794ea4d4012f
✓	Switch1	Succeeded	3/13/2025, 12:57:32 PM		2m 9s	00e2763b-4c03-4d8f-b633-96e9ef0842aa
↳	df_scdf_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

Name	Content Type
Employee_SCDCTYPE2.csv	

Supplier Data

Name
Suppliers_SCDTYPE1.csv
Suppliers_SCDTYPE1.csv
Suppliers_SCDTYPE1.csv
Suppliers_SCDTYPE1.csv

Check the data in Azure SQL Tables

Results		Messages									
supplier_id	name	contact_name	phone	address	createdBy	createdDate	updatedBy	updatedDate	hashKey		
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	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	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	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	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	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	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	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	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.

Harpalsinh Vaghela

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/>