

Local DB



Storage Account



Azure SQL DB



Data Migration Project



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

Data Migration in Azure

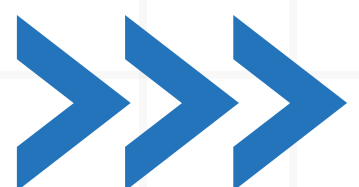
The process of moving data from one storage system to another within the Azure cloud environment, optimizing for accessibility, scalability, and performance.

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.

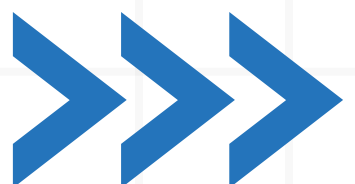


Best practices to follow for this data migration project

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
94 CREATE TABLE dbo.WATERMARK(
95     ID INT IDENTITY(1,1),
96     TABLE_NAME VARCHAR(100),
97     SCHEMA_NAME VARCHAR(100),
98     FOLDER_NAME VARCHAR(100),
99     LPV VARCHAR(100),
100     DELTA_COLUMN VARCHAR(100),
101     TABLE_TYPE VARCHAR(100)
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
@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

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

Name	Type	Sub-type	Status	Related	Region	Version
AutoResolveIntegrationRuntime	Azure	Public	Running	8	Auto Resolve	---
SelfhostedIR	Self-Hosted	---	Running	4	---	5.45.0999.1

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

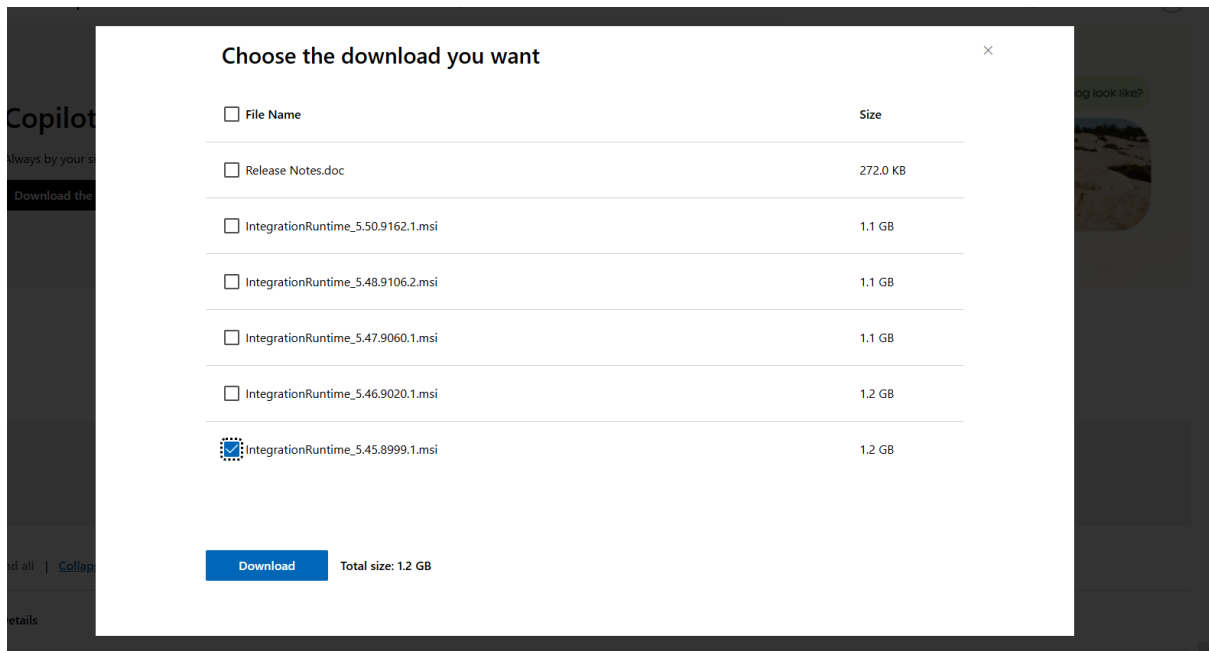
Stop Service

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 * [Learn more](#)

Description

Activity state ☒ Activated ☐ Deactivated

Timeout

Retry

Retry interval (sec)

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.

sign

ds_sql_server_sh

×

Edit

+

New

Learn more

Preview data

Edit linked service

SQL server [Learn more](#)

Name *

ls_sql_server_sh

Description

Connect via integration runtime * ⓘ

SelfhostedIR

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

Version

☐ Recommended
 ☒ Legacy

Server name *

HARPAL

Database name *

NCPL

Authentication type

SQL authentication

User name *

Harpal11

Password *

.....

Always encrypted ⓘ ☐

☒ Connection successful
[Test connection](#)

Take two parameters:

Connection

Schema

Parameters

+

New

Delete

<input type="checkbox"/>	Name	Type	Default value	
<input type="checkbox"/>	SchemaName	String	Value	
<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 ^①

Name	Value
SchemaName	dbo
TableName	WATERMARK

First row only ☐

Use query ☒ Table ☐ Query ☐ Stored procedure

Query timeout (minutes) ^① 120

Isolation level ^① Select...

Partition option ^① ☒ None ☐ Physical partitions of table ^① ☐ Dynamic range ^①

i Please preview data to validate the partition settings.

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

@activity('Watermark').output.value

Lookup Watermark

ForEach ForEach1

General Settings Activities (4) User properties

Sequential ☐

Batch count ^①

Items @activity('Watermark').output.value [🗑️](#)

Click on the Edit icon on ForEach Activity

Take Lookup activity and rename “GetSourceMaxValue”

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" value=""/>
TableName	<input type="text" value=""/>

First row only ☒

Use query ☐ Table ☒ Query ☐ Stored procedure

Query [Copy](#)

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

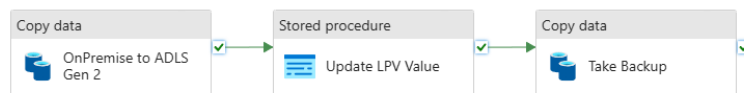


**@less(string(item().LPV),
string(activity('GetSourceMaxValue').output.firstRow.MaxValue))**

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}'**

General **Source** Sink Mapping Settings User properties

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

Dataset properties ^①

Name	Value
SchemaName	''
TableName	''

Use query ☐ Table ☒ Query ☐ Stored procedure

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

Query timeout (minutes) ^① 120

Isolation level ^① Select...

Partition option ^① ☒ None ☐ Physical partitions of table ^① ☐ Dynamic range ^①

i Please preview data to validate the partition settings.

Additional columns ^① [+ New](#)

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

Take two parameters

Connection Schema **Parameters**

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

<input type="checkbox"/>	Name	Type	Default value	
<input type="checkbox"/>	FolderName	String	Value	Delete
<input type="checkbox"/>	FileName	String	Value	Delete

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 ^① Comma (,)

Row delimiter ^① Default (\r\n, or \n)

Encoding ^① Default(UTF-8)

Quote character ^① Double quote (")

Escape character ^① Backslash (\)

First row as header ^① ☒

Null value ^①

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 * [Open](#) [+ New](#) [Learn more](#)

Dataset properties [ⓘ]

Name	Value
FolderName	<input type="text" value="@item().FOLDER_NAME"/>
FileName	<input type="text" value="@concat(item().TABLE_NAME,'_',item()..."/>

Copy behavior [ⓘ]

Max concurrent connections [ⓘ]

Block size (MB) [ⓘ]

Metadata [ⓘ] [+ New](#)

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 * [ⓘ] [Test connection](#) [Edit](#) [+ New](#)

Integration runtime * ☒ SelfhostedIR [Edit](#)

Stored procedure name * ☒ Enter manually

Stored procedure parameters [ⓘ]

[← Import](#) [+ New](#) [Delete](#)

<input type="checkbox"/>	Name	Type	Value
<input type="checkbox"/>	LPV_Value	String	<input type="text" value="@activity('GetSourceMaxValue').outp..."/>
<input type="checkbox"/>	Table_Name	String	<input type="text" value="@item().TABLE_NAME"/>

[Add dynamic content \[Alt+Shift+D\]](#)

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 Schema **Parameters**

+ New | Delete

<input type="checkbox"/>	Name	Type	Default value	
<input type="checkbox"/>	FolderName	String	Value	
<input type="checkbox"/>	FileName	String	Value	

Connection tab

Connection Schema Parameters

Linked service * AzureDataLakeStorage1 Test connection Edit + New Learn more

Integration runtime * AutoResolveIntegrationRuntime 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 \n)

Encoding Default(UTF-8)

Quote character Double quote (")

Escape character Backslash (\)

First row as header ☒

Null value

Source

General **Source** Sink Mapping Settings User properties

Source dataset * ds_csvfiles Open + New Preview data Learn more

Dataset properties

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

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

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

Recursively ☒

Enable partitions discovery ☐

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 the Sink configuration in Azure Data Factory. The 'Linked service' is set to 'AzureDataLakeStorage1'. The 'Integration runtime' is set to 'AutoResolveIntegrationRuntime'. The 'File path' is configured with three segments: 'csvfiles', '@dataset().FolderName', and '@dataset().FileName'. The 'Compression type' is 'No compression', 'Column delimiter' is 'Comma (,)', 'Row delimiter' is 'Default (\r,\n, or \r\n)', 'Encoding' is 'Default(UTF-8)', 'Quote character' is 'Double quote (")', 'Escape character' is 'Backslash (\)', and 'First row as header' is 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
36
37
38
39
40 CREATE TABLE tbl_Employee (
41     employee_id INT,
42     first_name VARCHAR(255),
43     last_name VARCHAR(255),
44     hire_date DATETIME,
45     last_review_date DATETIME,
46     role VARCHAR(100),
47     CREATEDBY VARCHAR(100),
48     CREATEDDATE DATETIME,
49     UPDATEDBY VARCHAR(100),
50     UPDATEDDATE DATETIME,
51     HASHKEY BIGINT,
52     ISACTIVE INT
53 );
54
55 --TRUNCATE TABLE tbl_Employee
56 Select * from tbl_Employee
57
58 --Target Source Query in Dataflow SCD Type 2
59 Select employee_id, HASHKEY from tbl_Employee where ISACTIVE = 1
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 -----
```

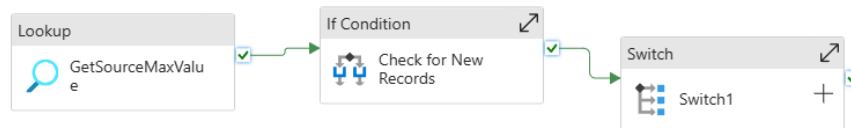
```
CREATE TABLE tbl_Suppliers (  
    supplier_id INT,  
    name VARCHAR(255),  
    contact_name VARCHAR(255),  
    phone VARCHAR(50),  
    address Varchar(100),  
        createdBy varchar(100),  
        createdDate datetime,  
        updatedBy varchar(100),  
        updatedDate datetime,  
        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 'Pipeline expression builder' window with 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'
    )
)
)
)

```

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

General
Activities (4)
User properties

Expression ⓘ
@if(and(eq

+ Add case

Case ⓘ

Default

scdtype2

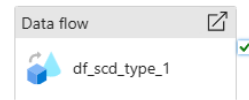
incremental

NoAction

Click on the Default case edit icon

Drag and drop Dataflow Activity

pl_project_Incremental_scd_types_1_2 > ForEach1 > Switch1 - Default

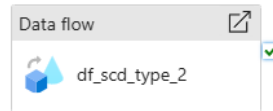


Go to settings and select the dataflow we have already designed for SCD type 1 logic

A screenshot of the Data flow settings and design view. The top part shows the 'Data flow' dropdown menu with 'df_scd_type_1' selected. Below it, the 'Settings' tab is active, showing various configuration options: 'Data flow' is set to 'df_scd_type_1_design', 'Run on (Azure IR)' is set to 'AutoResolveIntegrationRuntime', 'Compute size' is set to 'Small', 'Logging level' is set to 'Verbose', and 'Sink properties' and 'Staging' are expanded. The bottom part shows the data flow design view, which includes a 'source1' activity, a 'select' activity, a 'derivedColumn1' activity, a 'lookup' activity, an 'INSERT' activity, an 'InsertAuditColumns' activity, a 'supplierInsertSi...' activity, a 'Target' activity, an 'UPDATE' activity, an 'updateAuditColu...' activity, an 'AlterRow' activity, and a 'supplierUpdate...' activity. The 'Target' activity is highlighted with a dashed border.

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 * df_scd_type_2_design Open New

Run on (Azure IR) * AutoResolveIntegrationRuntime

Compute size * Small

> Advanced

Logging level * Verbose Basic None

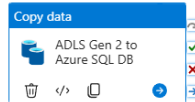
> Sink properties

> Staging 1

The diagram illustrates a dataflow for SCD type 2 logic. It starts with a 'source' activity (Import data from AzureDataLakeStorage1) followed by a 'select' activity (Renaming source to select with columns src_employee_id, src_first_name, src_last_name, src_hire_date). This is followed by a 'derivedColumn1' activity (Creating/updating the columns src_employee_id, src_first_name, src_last_name, src_hire_date). A 'lookup' activity (Lookup on 'derivedColumn1' from 'Target') follows. The main logic is split into two paths: an 'INSERT' path (Conditionally distributing the data in employee_id, src_employee_id, employee_id, src_hire_date, HASHSET groups) and an 'UPDATE' path (Conditionally distributing the data in employee_id, src_employee_id, employee_id, src_hire_date, HASHSET groups). Both paths lead to a 'Union' activity (Combining rows from transformation Condition1@src@INSERT, Condition1@src@UPDATE). This is followed by an 'InsertAuditColumns' activity (Creating/updating the columns src_employee_id, src_first_name, src_last_name, src_hire_date), an 'InsertEmployeeSink' activity (Add sink dataset), and an 'UpdateEmployeeSink' activity (Add sink dataset). The diagram also shows a 'Target' activity (Import data from AzureDataLakeStorage1) 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 * [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

General **Source** Sink Mapping Settings User properties

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

Dataset properties [ⓘ]

Name	Value
FolderName	**
FileName	**

File path type ☐ File path in dataset ☒ Wildcard file path ☐ List of files [ⓘ]

Wildcard paths csvfiles / @item().FOLDER_NAME / *_INCREMENTAL.csv

Filter by last modified [ⓘ] Start time (UTC) End time (UTC)

Recursively ☒ [ⓘ]

Enable partitions discovery ☐ [ⓘ]

Max concurrent connections [ⓘ]

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

Connection **Schema** Parameters

Linked service * AzureSqlDatabase1 [Test connection](#) [Edit](#) [New](#) [Learn more](#)

Integration runtime * AutoResolveIntegrationRuntime [Edit](#)

Table @dataset().SchemaName . @dataset().TableName [Preview data](#)

☒ Enter manually

General Source **Sink** Mapping Settings User properties

Sink dataset * azuresqllookup 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

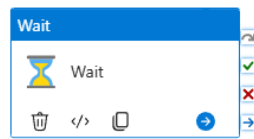
Wait

Wait

General

Case * NoAction

pl_project_Incremental_scd_types_1_2 > ForEach1 > Switch1 - NoAction



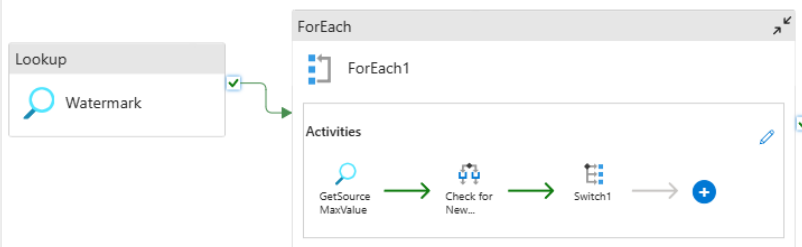
General **Settings** User properties

Wait time in seconds *

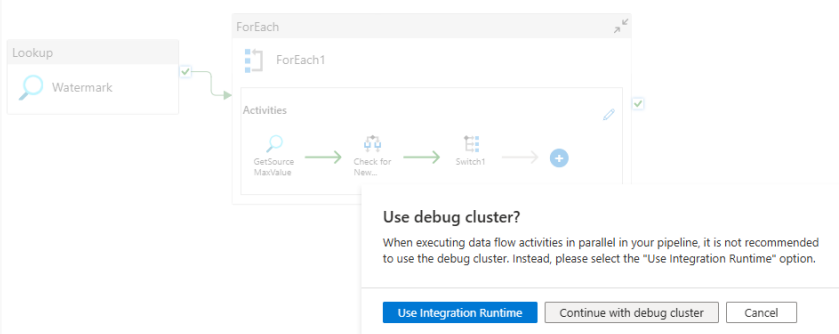
Let's run the pipeline

Enable **Data flow Debug** option

✓ Validate ▶ Debug ▼ ⚡ Add trigger ☒ Data flow debug ✓



Click on **Continue with debug cluster**



Harpalsinh Vaghela

Parameters	Variables	Settings	Output
	GetSourceMaxValue		Succeeded
	GetSourceMaxValue		Succeeded
	GetSourceMaxValue		Succeeded
	GetSourceMaxValue		Succeeded
	Check for New Records		Succeeded
	OnPremise to ADLS Gen 2		Succeeded
	Update LPV Value		Succeeded
	Take Backup		Succeeded
	Check for New Records		Succeeded
	OnPremise to ADLS Gen 2		Succeeded
	Update LPV Value		Succeeded
	Take Backup		Succeeded
	Check for New Records		Succeeded
	OnPremise to ADLS Gen 2		Succeeded
	Update LPV Value		Succeeded
	Take Backup		Succeeded
	Check for New Records		Succeeded
	OnPremise to ADLS Gen 2		Succeeded

Let’s check the data in Azure SQL Database tables

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

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

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

employee_id	first_name	last_name	hire_date	last_review_date	role	CREATEDBY	CREATEDDATE	UPDATEDBY	UPDATEDDATE	HASHKI
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	3824467
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	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

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

Employee_SCDTYPE2.csv

Path https://adfsigen3tgharpal.dfs.core.windows.net/csvfiles/RetailDB/Emplic
Modified 3/13/2025, 12:35:59 PM

With column header ☒ On

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

Content Type	Size
--------------	------

Inventory Log Data

New SQL script

New notebook

New data flow

New integration dataset

←

→

↶

↷

csvfiles

RetailDB

InventoryLogs

Name	Last Modified
Inventory_Logs_INCREMENTAL.csv	3/13/2025, 12:36:02

Inventory_Logs_INCREMENTAL.csv

Path

https://adlsgen2stgghpal.dfs.core.windows.net/csvfiles/RetailDB/invent

Modified

3/13/2025, 12:36:02 PM

With column header

On

LOG_ID	PRODUCT_ID	LOG_UPDATE...	CHANGE_QU
1	1	2023-10-01 08:0...	20
2	2	2023-10-01 09:0...	-10
3	3	2023-10-01 10:0...	30
NULL	NULL	NULL	NULL

Manage access

Properties

Delete

More

Content Type

Size

Products Data

New SQL script

New notebook

New data flow

New integration dataset

←

→

↶

csvfiles

RetailDB

Products

Name

Last Modified

Products_INCREMENTAL.csv

3/13/2025, 12:36:02 PM

Products_INCREMENTAL.csv

Path

https://adisgen2stgharpal.dfs.core.windows.net/csvfiles/RetailDB/Products_INCREMENTAL.csv

Modified

3/13/2025, 12:36:02 PM

With column header

On

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

Manage access

Properties

Delete

More

Content Type

Sales Data

New SQL script

New notebook

New data flow

New integration dataset

←

→

↶

↷

csfiles

RetailDB

Sales

Name

Last Modified

📄

Sales_INCREMENTAL.csv

3/13/2025, 12:36:01 PM

Sales_INCREMENTAL.csv

Path

https://adlsgen2stgharpaldfs.core.windows.net/csfiles/RetailDB/Sales/

Modified

3/13/2025, 12:36:01 PM

With column header

On

SALE_ID	SALE_DATE	PRODUCT_ID	QUANTITY
1	2023-10-01 14:0...	1	10
2	2023-10-01 14:1...	2	5
3	2023-10-01 15:0...	3	4
NULL	NULL	NULL	NULL

Manage access

Properties

Delete

More

Content Type

Size

Supplier Data

The screenshot shows the Databricks file browser interface. On the left, the breadcrumb navigation shows 'csvfiles' > 'RetailDB' > 'Suppliers'. The file 'Suppliers_SCDDTYPE1.csv' is selected. The file details show the path 'https://adlsgen2stgharpal.dfs.core.windows.net/csvfiles/RetailDB/Suppliers_SCDDTYPE1.csv', the modification date '3/13/2025, 12:36:03 PM', and the 'With column header' toggle is set to 'On'. The file content is displayed as a table with the following data:

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

Lets check the backup container in 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 acces

←

→

↑

csvfiles

>

backup

Name	Last Modified
RetailDB	3/13/2025, 9:19:32 AM

New SQL script

New data flow

New integration dataset

Upload

Download

New folder

Select all

Copy link

Rename

Manage access

←

→

↑

csvfiles

>

backup

>

RetailDB

Name	Last Modified	Con
Employee	3/13/2025, 9:19:42 AM	Folde
InventoryLogs	3/13/2025, 9:19:35 AM	Folde
Products	3/13/2025, 9:19:32 AM	Folde
Sales	3/13/2025, 9:19:35 AM	Folde
Suppliers	3/13/2025, 9:20:06 AM	Folde

Employee Data Backup

←

→

↑

csvfiles

>

backup

>

RetailDB

>

Employee

Name	Last Modified
Employee_2025-03-13T16:36:18.7263169Z.csv	3/13/2025, 12:36:34 PM

pl_project_Increme...

csvfiles

New SQL script

New notebook

New data flow

New integration dataset

←

→

↑

csvfiles

>

backup

>

RetailDB

>

Employee

Name	Last Modified
Employee_2025-03-13T16:36:18.7263169Z.csv	3/13/2025, 12:36:34 PM

Employee_2025-03-13T16:36:18.7263169Z.csv

Path

https://adlsgen2stgharpal.dfs.core.windows.net/csvfiles/backup/RetailD03-13T16:36:18.7263169Z.csv

Modified

3/13/2025, 12:36:34 PM

With column header

On

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

Manage access

Properties

Delete

Content Type

Inventory Data Backup

←

→

↑

csvfiles

>

backup

>

RetailDB

>

InventoryLogs

Name	Last Modified
Inventory_Logs_2025-03-13T16:36:22.9983517Z.csv	3/13/2025, 12:36:44 PM

Inventory_Logs_2025-03-13T16:36:22.9983517Z.csv

Path

https://adlsgen2stgharpal.dfs.core.windows.net/csvfiles/backup/RetailD03-13T16:36:22.9983517Z.csv

Modified

3/13/2025, 12:36:44 PM

With column header

On

LOG_ID	PRODUCT_ID	LOG_UPDATE...	CHANGE_QU
1	1	2023-10-01 08:0...	20
2	2	2023-10-01 09:0...	-10
3	3	2023-10-01 10:0...	30
NULL	NULL	NULL	NULL

Content Type

Product Data Backup

Products_2025-03-13T16:36:21.1544216Z.csv

Path: https://adlsgen2stgharpal.dfs.core.windows.net/csvfiles/backup/RetailD
03-13T16:36:21.1544216Z.csv
Modified: 3/13/2025, 12:36:38 PM

With column header: ☒ On

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

Sales_2025-03-13T16:36:25.1471452Z.csv

Path: https://adlsgen2stgharpal.dfs.core.windows.net/csvfiles/backup/RetailD
03-13T16:36:25.1471452Z.csv
Modified: 3/13/2025, 12:36:44 PM

With column header: ☒ On

SALE_ID	SALE_DATE	PRODUCT_ID	QUANTITY
1	2023-10-01 14:0...	1	10
2	2023-10-01 14:1...	2	5
3	2023-10-01 15:0...	3	4
NULL	NULL	NULL	NULL

Supplier Data Backup

Suppliers_2025-03-13T16:36:19.5400784Z.csv

Path: https://adlsgen2stgharpal.dfs.core.windows.net/csvfiles/backup/RetailD
03-13T16:36:19.5400784Z.csv
Modified: 3/13/2025, 12:36:35 PM

With column header: ☒ On

SUPPLIER_ID	NAME	CONTACT_NAME	PHONE
1	Fresh Farms	John Doe	555-3489
2	Healthy Beverag...	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 | -----
110 | ----Insert new data into Products table
111 | INSERT INTO Products (product_id, name, category, price, stock, supplier_id, productUpdatedDate) VALUES
112 | (4, 'Vegetable Oil', 'Groceries', 4.50, 100, 2, '2023-12-01 00:00:00');
113 |
114 | ----Insert new data into Sales table
115 | INSERT INTO Sales (sale_id, sale_date, product_id, quantity, total_amount, cashier_id, salesUpdatedDate) VALUES
116 | (4, '2025-02-25 10:30:00', 5, 20, 90.00, 1, '2025-02-25 10:35:00');
117 |
118 |
119 | ----Insert new data into Suppliers table --- SCD Type 1
120 | INSERT INTO Suppliers (supplier_id, name, contact_name, phone, address, supplierUpdatedDate) VALUES
121 | (4, 'Bakers Delight', 'Nora Special Bates', '111-1111', '88 Baker Rd', '2025-02-22 00:00:00');
122 |
123 |
124 | ----Insert new data into Employees table ---- SCD Type 2
125 | INSERT INTO Employee (employee_id, first_name, last_name, hire_date, last_review_date, role) VALUES
126 | (4, 'Nayan', 'Vaghela', '2025-01-01 09:00:00', '2025-02-20 00:00:00', 'Inventory Specialist');
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');
```

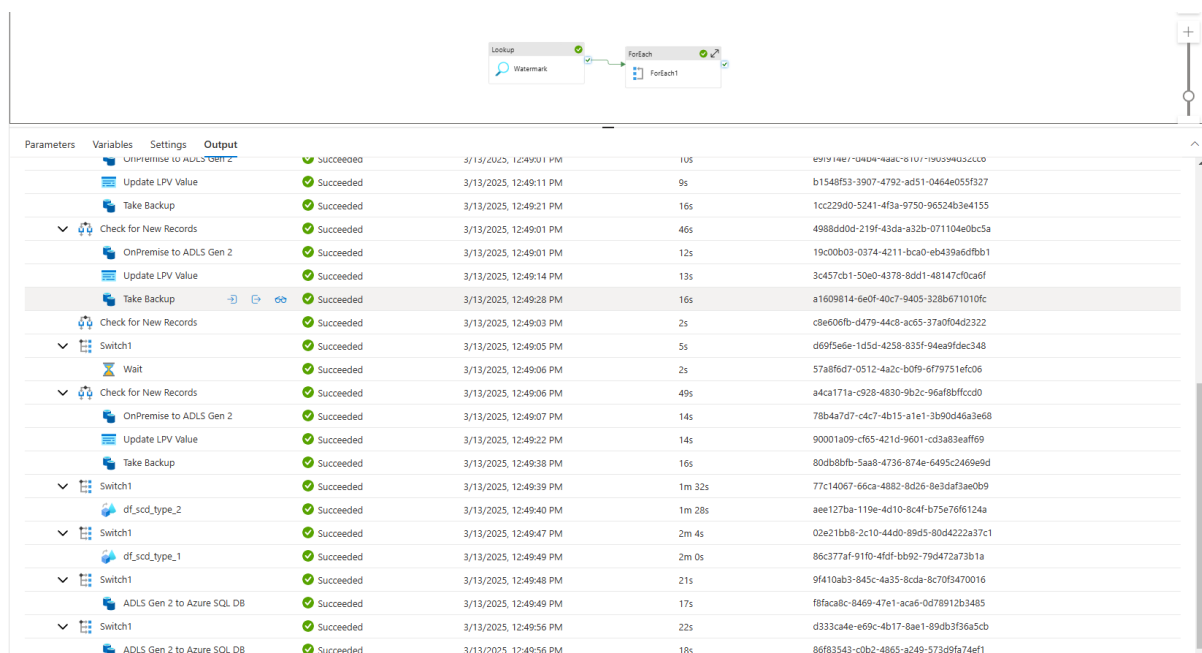
----Insert new data into Suppliers table --- SCD Type 1

```
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



Parameters	Variables	Settings	Output
OnPremise to ADLS Gen 2			✓ Succeeded 3/13/2025, 12:49:01 PM 10s e919146e-7040d4-848c-b101-790394032c00
Update LPV Value			✓ Succeeded 3/13/2025, 12:49:11 PM 9s b1548f53-3907-4792-ad51-0464e055f327
Take Backup			✓ Succeeded 3/13/2025, 12:49:21 PM 16s 1cc229d0-5241-4f3a-9750-96524b3e4155
Check for New Records			✓ Succeeded 3/13/2025, 12:49:01 PM 46s 4988dd0d-219f-43da-a32b-071104e0bc5a
OnPremise to ADLS Gen 2			✓ Succeeded 3/13/2025, 12:49:01 PM 12s 19c00b03-0374-4211-bca0-eb439a6dfb1
Update LPV Value			✓ Succeeded 3/13/2025, 12:49:14 PM 13s 3c457cb1-50e0-4378-8dd1-48147c0ca6f
Take Backup			✓ Succeeded 3/13/2025, 12:49:28 PM 16s a1609814-6e0f-40c7-9405-328b671010fc
Check for New Records			✓ Succeeded 3/13/2025, 12:49:03 PM 2s c8e606fb-d479-44c8-ac05-37a0f04d2322
Switch1			✓ Succeeded 3/13/2025, 12:49:05 PM 5s d69f5e6e-1d5d-4258-835f-94ea9fdec348
Wait			✓ Succeeded 3/13/2025, 12:49:06 PM 2s 57a8fed7-0512-4a2c-b0f9-6f79751efc06
Check for New Records			✓ Succeeded 3/13/2025, 12:49:06 PM 49s a4ca171a-c928-4830-9b2c-96af8bfcccd0
OnPremise to ADLS Gen 2			✓ Succeeded 3/13/2025, 12:49:07 PM 14s 78b4a7d7-c4c7-4b15-a1e1-3b90d46a3e68
Update LPV Value			✓ Succeeded 3/13/2025, 12:49:22 PM 14s 90001a09-cf65-421d-9601-cd3a83eaff69
Take Backup			✓ Succeeded 3/13/2025, 12:49:38 PM 16s 80db8bf0-5aa8-4736-874e-6495c2409e9d
Switch1			✓ Succeeded 3/13/2025, 12:49:39 PM 1m 32s 77c14067-66ca-4882-8d26-8e3daf3ae0b9
df_scd_type_2			✓ Succeeded 3/13/2025, 12:49:40 PM 1m 28s aee127ba-119e-4d10-8c4f-b75e76f6124a
Switch1			✓ Succeeded 3/13/2025, 12:49:47 PM 2m 4s 02e21bb8-2c10-44d0-89d5-80d4222a37c1
df_scd_type_1			✓ Succeeded 3/13/2025, 12:49:49 PM 2m 0s 86c377af-91f0-4fdf-bb92-79d472a73b1a
Switch1			✓ Succeeded 3/13/2025, 12:49:48 PM 21s 9f410ab3-845c-4a35-8cda-8c70f3470016
ADLS Gen 2 to Azure SQL DB			✓ Succeeded 3/13/2025, 12:49:49 PM 17s f8fac8c-8469-47e1-acaf-0d78912b3485
Switch1			✓ Succeeded 3/13/2025, 12:49:56 PM 22s d333ca4e-e69c-4b17-8ae1-89db3f36a5cb
ADLS Gen 2 to Azure SQL DB			✓ Succeeded 3/13/2025, 12:49:56 PM 18s 86f83543-c0b2-4865-a249-573d9fa74ef1

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

Employee Data

Harpalsinh Vaghela

The screenshot shows the Azure Data Explorer interface. On the left, a file explorer pane displays the file 'Employee_SCDTYPE2.csv' under the path 'csvfiles > RetailDB > Employee'. The main pane shows the file's details: Path is 'https://adlsgen2stgharpal.dfs.core.windows.net/csvfiles/RetailDB/Emplc', Modified is '3/13/2025, 12:49:08 PM', and 'With column header' is set to 'On'. A table preview shows columns EMPLOYEE_ID, FIRST_NAME, LAST_NAME, and HIRE_DATE with data for employee 4 (Nayan Vaghela) and NULL values.

EMPLOYEE_ID	FIRST_NAME	LAST_NAME	HIRE_DATE
4	Nayan	Vaghela	2025-01-01 09:
NULL	NULL	NULL	NULL

Product Data

The screenshot shows the Azure Data Explorer interface. On the left, a file explorer pane displays the file 'Products_INCREMENTAL.csv' under the path 'csvfiles > RetailDB > Products'. The main pane shows the file's details: Path is 'https://adlsgen2stgharpal.dfs.core.windows.net/csvfiles/RetailDB/Produ', Modified is '3/13/2025, 12:49:11 PM', and 'With column header' is set to 'On'. A table preview shows columns PRODUCT_ID, NAME, CATEGORY, and PRICE with data for product 4 (Vegetable Oil) and NULL values.

PRODUCT_ID	NAME	CATEGORY	PRICE
4	Vegetable Oil	Groceries	4.50
NULL	NULL	NULL	NULL

Sales Data

The screenshot shows the Azure Data Explorer interface. On the left, a file explorer pane displays the file 'Sales_INCREMENTAL.csv' under the path 'csvfiles > RetailDB > Sales'. The main pane shows the file's details: Path is 'https://adlsgen2stgharpal.dfs.core.windows.net/csvfiles/RetailDB/Sales/', Modified is '3/13/2025, 12:49:18 PM', and 'With column header' is set to 'On'. A table preview shows columns SALE_ID, SALE_DATE, PRODUCT_ID, and QUANTITY with data for sale 4 and NULL values.

SALE_ID	SALE_DATE	PRODUCT_ID	QUANTITY
4	2025-02-25 10:3...	5	20
NULL	NULL	NULL	NULL

Supplier Data

The screenshot shows the Azure Data Explorer interface. On the left, a file explorer pane displays the file 'Suppliers_SCDTYPE1.csv' under the path 'csvfiles > RetailDB > Suppliers'. The main pane shows the file's details: Path is 'https://adlsgen2stgharpal.dfs.core.windows.net/csvfiles/RetailDB/Suppl', Modified is '3/13/2025, 12:49:11 PM', and 'With column header' is set to 'On'. A table preview shows columns SUPPLIER_ID, NAME, CONTACT_NA..., and PHONE with data for supplier 4 (Bakers Delight) and NULL values.

SUPPLIER_ID	NAME	CONTACT_NA...	PHONE
4	Bakers Delight	Nora Special Bates	111-1111
NULL	NULL	NULL	NULL

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

12.0.2000.8

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
4	Vegetable Oil	Groceries	4.50	100	2	2023-12-01 00:00:00.000

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
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	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

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 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	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	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

Harpalsinh Vaghela

	✓	Check for New Records	✓ Succeeded	3/13/2025, 12:30:40 PM	8s	9c3d6a0f7e3c7404749b1274c2326ac29050
▼	🔗	Check for New Records	✓ Succeeded	3/13/2025, 12:56:48 PM	398s	03d9b6a1-df31-44ff-b811-34b49832cd46
	💻	OnPremise to ADLS Gen 2	✓ Succeeded	3/13/2025, 12:56:49 PM	11s	405f4c37-6105-4d22-8734-bbd8f8baee28
	📄	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-5532f40daaba1
	🔗	Check for New Records	✓ Succeeded	3/13/2025, 12:56:49 PM	3s	5294b323-44ef-4c31-a5f5-20c20399a732
▼	🔗	Check for New Records	✓ Succeeded	3/13/2025, 12:56:49 PM	43s	60b0efc0-c83b-4e01-8f06-55428cb7f4e7b
	💻	OnPremise to ADLS Gen 2	✓ Succeeded	3/13/2025, 12:56:50 PM	11s	d17da2ad-f9f1-41cb-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-9088-4ee5-b722-52672bbfb2bd
▼	⚙️	Switch1	✓ Succeeded	3/13/2025, 12:56:51 PM	5s	78aab3bf-7282-4ae8-9e47-f513bed58013
	⌛	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-8ffa-5758b074fc2a5
▼	⚙️	Switch1	✓ Succeeded	3/13/2025, 12:56:52 PM	5s	151b9817-2570-47cf-bfbd-cb307aar2a12
	⌛	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-a6eda245bcdcf
	⌛	Wait	✓ Succeeded	3/13/2025, 12:56:55 PM	2s	c3dd5f5-fd7a-449f-be74-219f4df389e06
▼	⚙️	Switch1	✓ Succeeded	3/13/2025, 12:57:28 PM	1m 28s	070217ed-a384-4a90-845b-d06ca47f96f3
	🔄	fd_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-b633-96cec9ef842aa
	🔄	fd_scd_type_1	✓ Succeeded	3/13/2025, 12:57:34 PM	2m 5s	cd070cdc-d2f1-4e08-ad42-b6ea3f8be282

Check the data in the ADLS Gen 2 storage account

Employee Data

Supplier Data

→ ↶ ↷ ↴

csvfiles > RetailDB > Suppliers

Name

Suppliers_SCDTYPE1.csv

Suppliers_SCDTYPE1.csv

Path

https://adlsgen2stghparp.dfs.core.windows.net/csvfiles/RetailDB/Suppl

Modified

3/13/2025, 12:56:58 PM

With column header

☒ On

SUPPLIER_ID	NAME	CONTACT_NA...	PHONE
4	Bakers Magic DeL...	Nora Magic Bates	555-2222
NULL	NULL	NULL	NULL

Content Type

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		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	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	3024467299	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	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.

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/>