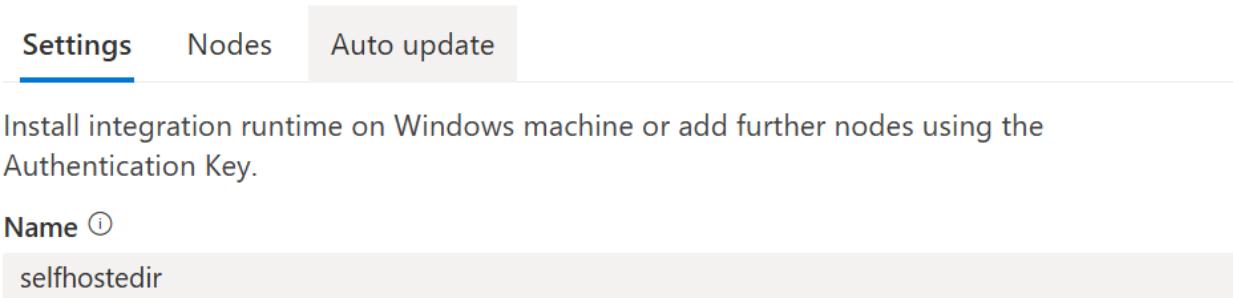


END TO END PROJECT

TO CREATE SELF HOSTED

Step 1 : login to synapse

Step 2 : manage → click IR → click on new → click self-hosted → create
Integration runtime setup



The screenshot shows a user interface for managing integration runtimes. At the top, there are three tabs: 'Settings' (underlined in blue), 'Nodes', and 'Auto update'. Below the tabs, a message reads: 'Install integration runtime on Windows machine or add further nodes using the Authentication Key.' Under the 'Name' label, the value 'selfhostedir' is displayed in a text input field.

Option 1: Express setup

[Click here to launch the express setup for this computer](#)

Option 2: Manual setup

Step 1: [Download and install integration runtime](#)

Step 2: Use this key to register your integration runtime

Name	Authentication key	Actions
Key1	IR@ef645529-26aa-444e-91a2-2317db0221b4@wspabi@aue@9l+j7YIC	 
Key2	IR@ef645529-26aa-444e-91a2-2317db0221b4@wspabi@aue@z9cU2X	 

Step 3 : click for the option 2 and download and install the IR



Register Integration Runtime (Self-hosted)

Welcome to Microsoft Integration Runtime Configuration Manager. Before you start, register your Integration Runtime (Self-hosted) node using a valid Authentication Key.



Show Authentication Key

[Learn how to find the Authentication Key](#)

HTTP Proxy

Current Proxy: No proxy

[Change](#)

Diagnostic Tool

[Troubleshoot problems \(preview\)](#)

[Register](#)

[Cancel](#)

Step 4 : register and install successfully



Self-hosted node is connected to the cloud service

Data Factory: wspabi

Integration Runtime: selfhostedir

Node: DESKTOP-3STF8QU

[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)



Step 5 : in linked services → in azsqlserver → click selfhostedir

Edit linked service

 SQL server [Learn more](#) 

Name *

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

[Connection string](#)

[Azure Key Vault](#)

Server name *

Database name *

Authentication type

User name *

[Password](#)

[Azure Key Vault](#)

Password *

[Apply](#)

[Cancel](#)

 [Test connection](#)

INCREMENTAL LOAD

TABLE 1

```

CREATE TABLE Company1 (
    CompanyID INT, --deltacolumn
    Name NVARCHAR(255),
    Address NVARCHAR(255),
    PhoneNumber NVARCHAR(50),
    Email NVARCHAR(255),
)

```

```

insert into Company1 values
(1,'palainteractive','toronto',1234567890,'pala@gmail.com')

insert into Company1 values (2,'apple','toronto',8967564534,'apple@gmail.com')

insert into Company1 values (3,'google','toronto',2345432890,'google@gmail.com')

insert into Company1 values (4,'ebay','toronto',0987567890,'ebay@gmail.com')

insert into Company1 values (5,'nike','toronto',9988567890,'nike@gmail.com')

```

	CompanyID	Name	Address	PhoneNumber	Email
1	1	palainteractive	toronto	1234567890	pala@gmail.com
2	2	apple	toronto	8967564534	apple@gmail.com
3	3	google	toronto	2345432890	google@gmail.com
4	4	ebay	toronto	987567890	ebay@gmail.com
5	5	nike	toronto	9988567890	nike@gmail.com

TABLE 2

```

CREATE TABLE Department1 (
    DepartmentID INT, --deltacolumn
    Name NVARCHAR(255) NOT NULL,
    ManagerID INT NULL,
    joindatedep datetime
)

```

```
insert into Department1 values (1,'it',100,'2020-01-01 00:00:00')
```

```

insert into Department1 values (2,'finance',101,'2020-01-01 00:00:00')
insert into Department1 values (3,'customer',102,'2020-01-01 00:00:00')
insert into Department1 values (4,'devops',103,'2020-01-01 00:00:00')
insert into Department1 values (5,'sap',104,'2020-01-01 00:00:00')

```

	DepartmentID	Name	ManagerID	joindatedep
1	1	it	100	2020-01-01 00:00:00.000
2	2	finance	101	2020-01-01 00:00:00.000
3	3	customer	102	2020-01-01 00:00:00.000
4	4	devops	103	2020-01-01 00:00:00.000
5	5	sap	104	2020-01-01 00:00:00.000

TABLE 3

```

create TABLE Employee1 (
    EmployeeID INT,
    Name NVARCHAR(255),
    Email NVARCHAR(255),
    PhoneNumber NVARCHAR(50),
    joindate datetime
)

```

```

insert into Employee1 values (1,'abhi','abhi@gamil.com',2343454567,'2020-01-01
00:00:00')
insert into Employee1 values (2,'john','john@gamil.com',9988454567,'2020-01-01
00:00:00')
insert into Employee1 values (3,'smith','smith@gamil.com',9943454567,'2020-01-01
00:00:00')
insert into Employee1 values (4,'kanguva','kanguva@gamil.com',8843454567,'2020-01-
01 00:00:00')
insert into Employee1 values (5,'amaran','amaran@gamil.com',2343454567,'2020-01-
01 00:00:00')

```

	EmployeeID	Name	Email	PhoneNumber	joindate
1	1	abhi	abhi@gamil.com	2343454567	2020-01-01 00:00:00.000
2	2	john	john@gamil.com	9988454567	2020-01-01 00:00:00.000
3	3	smith	smith@gamil.com	9943454567	2020-01-01 00:00:00.000
4	4	kanguva	kanguva@gamil.com	8843454567	2020-01-01 00:00:00.000
5	5	amaran	amaran@gamil.com	2343454567	2020-01-01 00:00:00.000

TABLE 4

```
create TABLE product1 (
    productid int,
    Name NVARCHAR(255),
    joindate datetime
)
```

```
insert into product1 values (1,'tv','2020-01-01 00:00:00')
insert into product1 values (2,'pears','2020-01-01 00:00:00')
insert into product1 values (3,'radio','2020-01-01 00:00:00')
insert into product1 values (4,'yogurt','2020-01-01 00:00:00')
insert into product1 values (5,'banana','2020-01-01 00:00:00')
```

	productid	Name	joindate
1	1	tv	2020-01-01 00:00:00.000
2	2	pears	2020-01-01 00:00:00.000
3	3	radio	2020-01-01 00:00:00.000
4	4	yogurt	2020-01-01 00:00:00.000
5	5	banana	2020-01-01 00:00:00.000

TABLE 5

```
create TABLE product2 (
    productid int,
    Name NVARCHAR(255),
```

```
    joindate datetime  
)  
;
```

```
insert into product2 values (1,'remote','2020-01-01 00:00:00')  
insert into product2 values (2,'charger','2020-01-01 00:00:00')  
insert into product2 values (3,'candle','2020-01-01 00:00:00')  
insert into product2 values (4,'laptop','2020-01-01 00:00:00')  
insert into product2 values (5,'desktop','2020-01-01 00:00:00')
```

	productid	Name	joindate
1	1	remote	2020-01-01 00:00:00.000
2	2	charger	2020-01-01 00:00:00.000
3	3	candle	2020-01-01 00:00:00.000
4	4	laptop	2020-01-01 00:00:00.000
5	5	desktop	2020-01-01 00:00:00.000

CREATE WATER MARK TABLE

```
create table watermark2  
(  
id int,  
tablename varchar(100),  
schemaname varchar(100),  
foldername varchar(100),  
lpv varchar(100),  
deltacolumn varchar(100)  
)  
insert into watermark2 values(1,'Company1','dbo','silver/Company1',0,'CompanyID')  
insert into watermark2  
values(2,'Department1','dbo','silver/Department1',0,'DepartmentID')
```

```
insert into watermark2 values(3,'Employee1','dbo','silver/Employee1','1900-01-01  
00:00:00','joindate')
```

```
insert into watermark2 values(4,'product1','dbo','silver/product1','1900-01-01  
00:00:00','joindate')
```

```
insert into watermark2 values(5,'product2','dbo','silver/product2','1900-01-01  
00:00:00','joindate')
```

	id	tablename	schemaname	filename	lpv	deltacolumn
1	1	Company1	dbo	silver/Company1	5	CompanyID
2	2	Department1	dbo	silver/Department1	5	DepartmentID
3	3	Employee1	dbo	silver/Employee1	2020-01-01T00:00:00	joindate
4	4	product1	dbo	silver/product1	2020-01-01T00:00:00	joindate
5	5	product2	dbo	silver/product2	2020-01-01T00:00:00	joindate

CREATE STORED PROCEDURE

```
create procedure watermark2_update  
  
@tablename varchar(100),  
  
@lpv varchar(100)  
  
as  
  
update watermark2  
  
set lpv=@lpv  
  
where tablename=@tablename
```

TO CREATE INCREMENTAL PIPELINE

TO CREATE INCREMENTAL PIPELINE IN THE SYNAPSE

[Create lookup activity](#), from source get data from sqldatabase (ssms)

Step 1 : Getdata from sources (sqlserver)

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.

I didn't select specific table because will use parameter to fetch the table dynamically

To do dynamically follow this steps

Step 2: click on the open from sources

Source dataset *

SQL ServerTable1

Step 3: add parameters

Connection Schema Parameters

+ New | Delete

Name	Type	Default value
tablename	String	Value
schemaname	String	Value

Step 4: to add schemaname and tablename dynamically

Click on add dynamic content and fetch the values

Linked service * Is_SQLServer_sh Test connection Edit + New Learn more

Integration runtime * selfhostedir Edit

Table @dataset().schemaname . @dataset().tablename Preview data

Enter manually

Step 5 : in my getwatermark sources the tablename and schemaname will dynamically fetch

Step 6 : By clicking preview data, can see this table list

	id	tablename	schemaname	foldername	lpv	deltacolumn
1	1	Company1	dbo	silver/Company1	0	CompanyID
2	2	Department1	dbo	silver/Department1	0	DepartmentID
3	3	Employee1	dbo	silver/Employee1	1900-01-01 00:00:00	joindate
4	4	product1	dbo	silver/product1	1900-01-01 00:00:00	joindate
5	5	product2	dbo	silver/product2	1900-01-01 00:00:00	joindate

Create foreach activity

Step 7 : link the lookup and foreach activity

Step 8 : in setting, items click on add dynamic content and select the getwatermark data value array

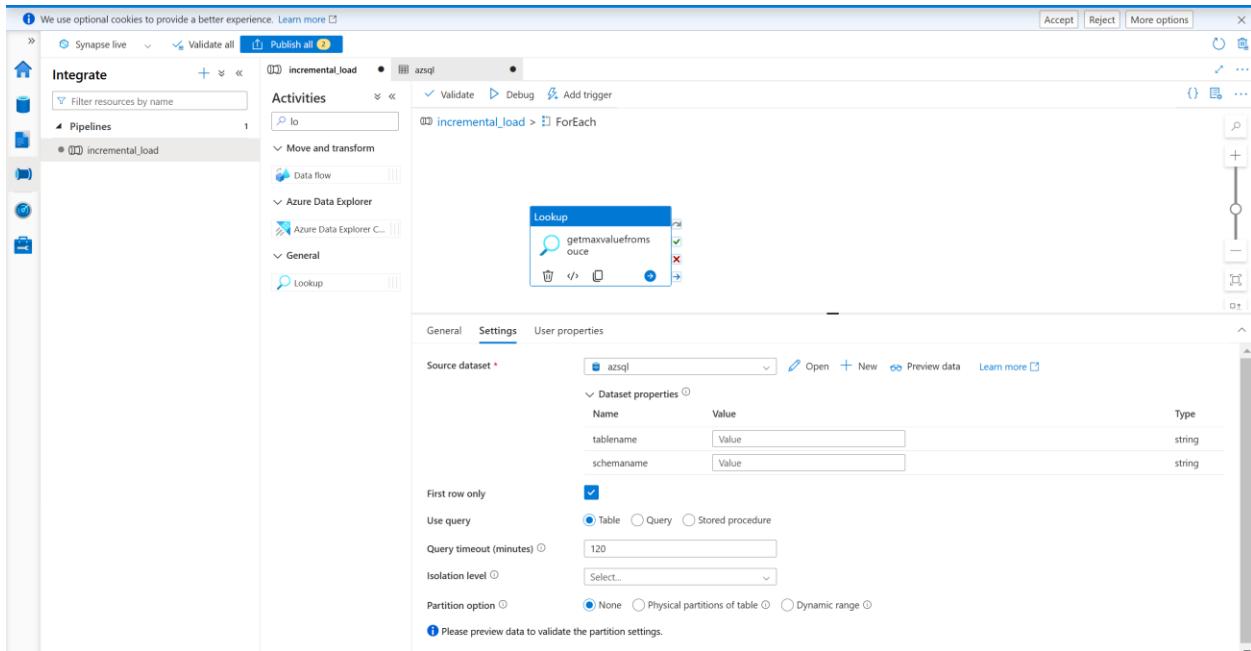
Items

@activity('Lookup1').output.value

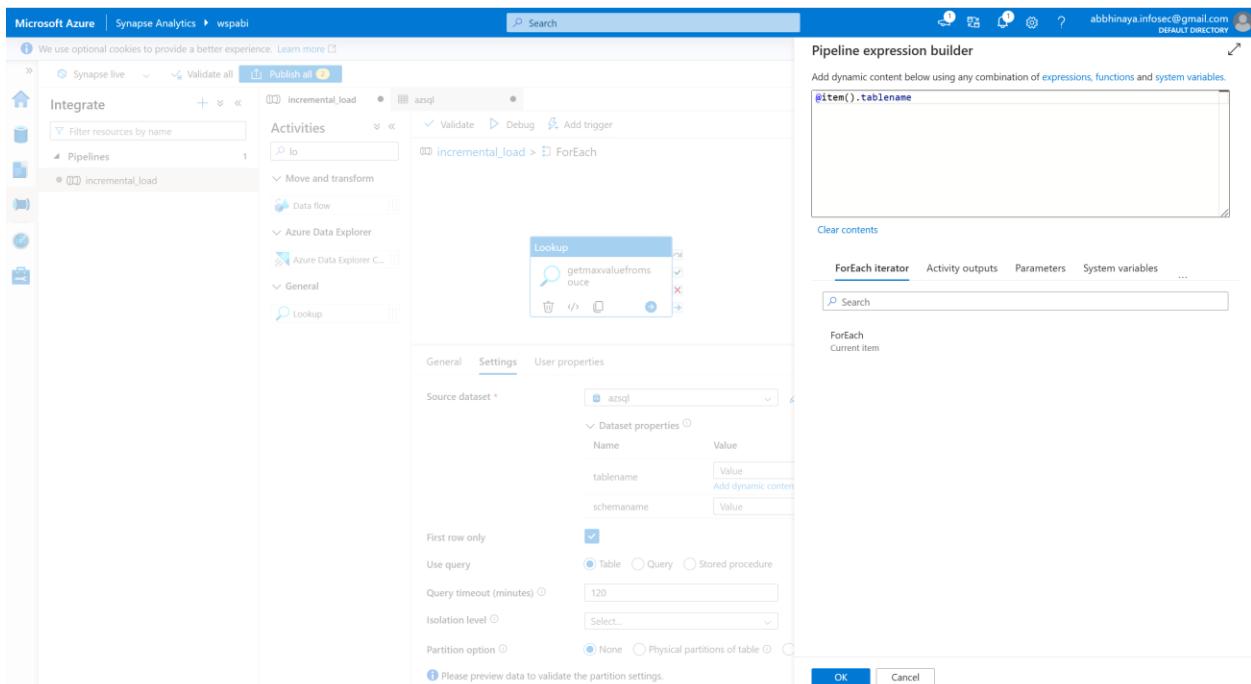
Step 9 : Inside the foreach activity, add one lookup activity

The lookup activity will give max value from source (my sources is the 5 table)

Click setting and add dynamic content in tablename and schemaname



Add tablename



Add schemaname

The screenshot shows the Microsoft Azure Synapse Analytics pipeline editor. On the left, the 'Integrate' sidebar is open, showing 'Pipelines' selected. In the main workspace, there is a pipeline named 'incremental_load'. A 'Lookup' activity is selected, which is part of a 'ForEach' loop. The 'Pipeline expression builder' pane on the right shows the expression '@item().schemaname' entered into the 'Add dynamic content below using any combination of expressions, functions and system variables' field.

Step 10 : Value is passed inside the foreach (lookup activity) and By clicking first row only

The screenshot shows the 'Dataset properties' dialog for 'SqlServerTable1'. The 'Source dataset' dropdown is set to 'SqlServerTable1'. Under 'Dataset properties', two entries are listed: 'tablename' with value '@item().tablename' and 'schemaname' with value '@item().schemaname'. Both values are highlighted with a blue selection bar. The 'First row only' checkbox is checked.

Name	Type
tablename	string
schemaname	string

Step 11 : get the max value in each table (click query and add dynamic content)

Query *

Edit

Add dynamic content [Alt+Shift+D]

Query to get max value in each table

```
select max(@{item().deltacolumn}) as maxvalue from  
@{item().schemaname}.@{item().tablename}
```

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

```
select max(@{item().deltacolumn}) as maxvalue from @{item().  
schemaname}.@{item().tablename}
```

[Clear contents](#)

ForEach iterator

Activity outputs

Parameters

System variables

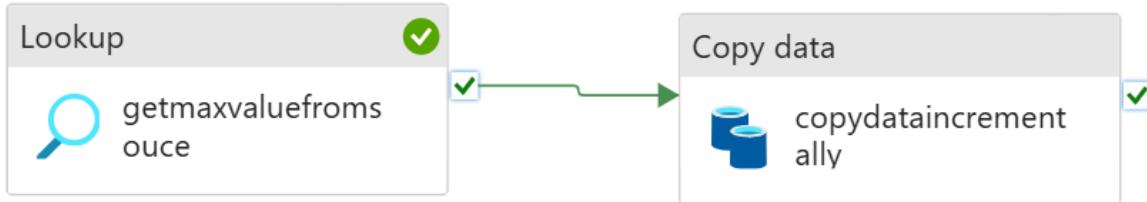
...

 Search

ForEach

Current item

Inside the foreach already we have created lookup and now we are [creating copydata activity](#)



Step 12 : by giving " it will get only incremental rows

tablename: ''

schemaname: ''

Add dynamic content [Alt+Shift+D]

Step 13 : pass the query

select * from @{{item().tablename}} where @{{item().deltacolumn}}>'@{{item().Ipv}}'

Pipeline expression builder

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

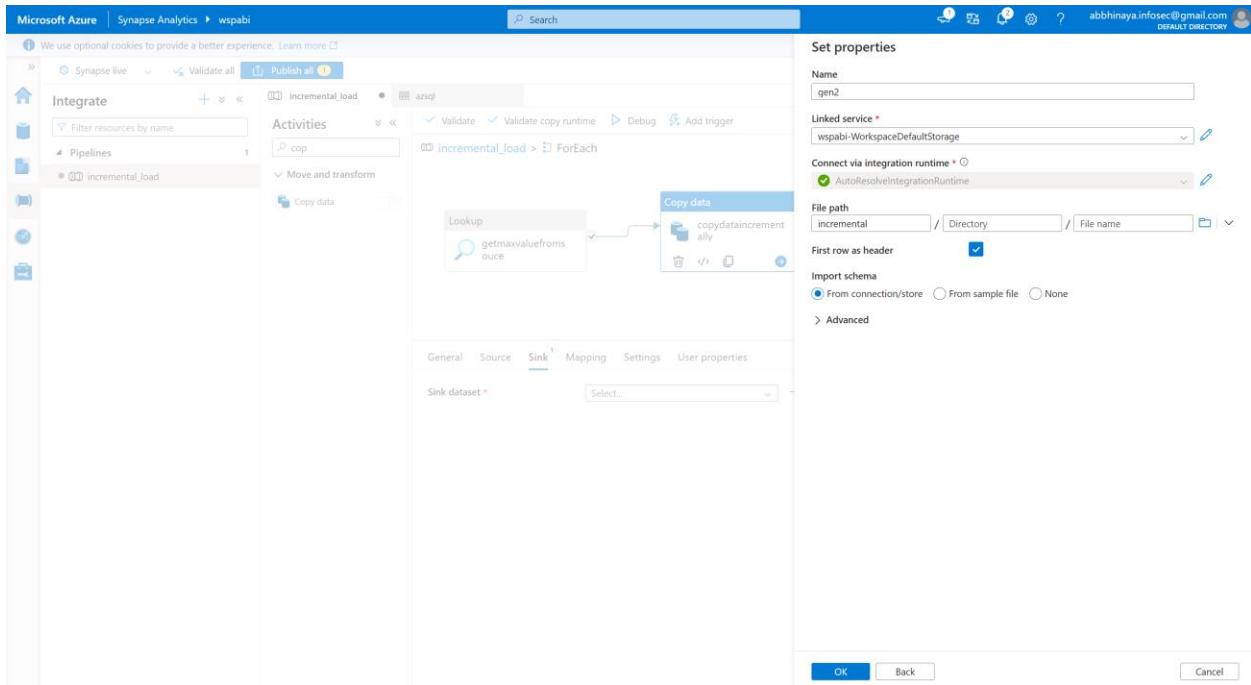
select * from @{{item().tablename}} where @{{item().deltacolumn}}>'@{{item().Ipv}}'

ForEach iterator

Source dataset: azsql

Query: select * from @{{item().tablename}} wh...

Step 14 : in sink I wanna load the data in ADLS gen 2



Step 15 : in sink create the parameter

Parameters		
<input type="checkbox"/>	Name	Type
<input type="checkbox"/>	filename	String
<input type="checkbox"/>	filename	String

Step 16 : click on connection and in directory click add dynamic content and give foldername and do as for filename

Linked service *	<input type="button" value="wspabi-WorkspaceDefaultStorage"/> Test connection <input type="button" value="Edit"/> <input type="button" value="New"/> Learn more
Integration runtime *	<input type="button" value="AutoResolveIntegrationRuntime"/> <input type="button" value="Edit"/>
File path	incremental / <input type="text" value="directory"/> / <input type="text" value="File name"/> <input type="button" value="Browse"/> Add dynamic content [Alt+Shift+D]

```
@dataset().filename
```

[Clear contents](#)

Parameters

Functions

 Search

filename

filename

Step 17 : both done

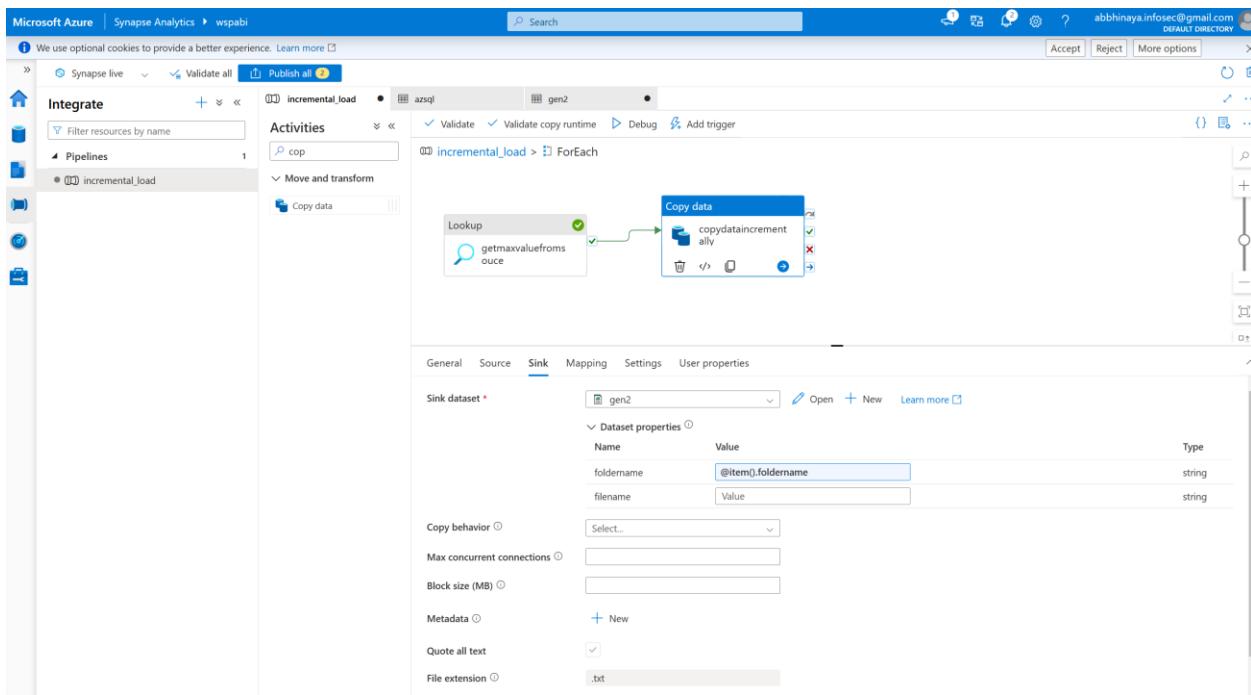
File path

incremental

/ @dataset().filename

/ @dataset().filename

Step 18 : give the filename in sink

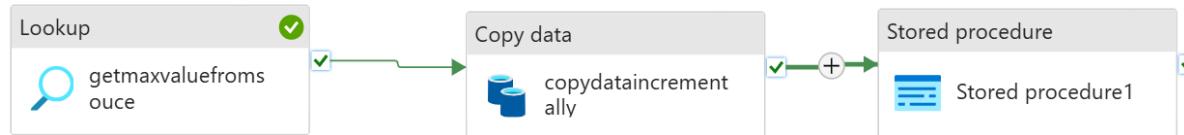


Step 19 : for file name use concat function

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

filename	<code>@item().filename</code>
filename	<code>@concat(item().tablename,'_',utcNow...)</code>

Inside the foreach already we have created lookup and now we are [creating Stored procedure activity](#)



Step 20 : by clicking on linked service, click the existing db which you have created in ssms and in stored procedure give the name which you have created.

Step 21 : By clicking import will get this

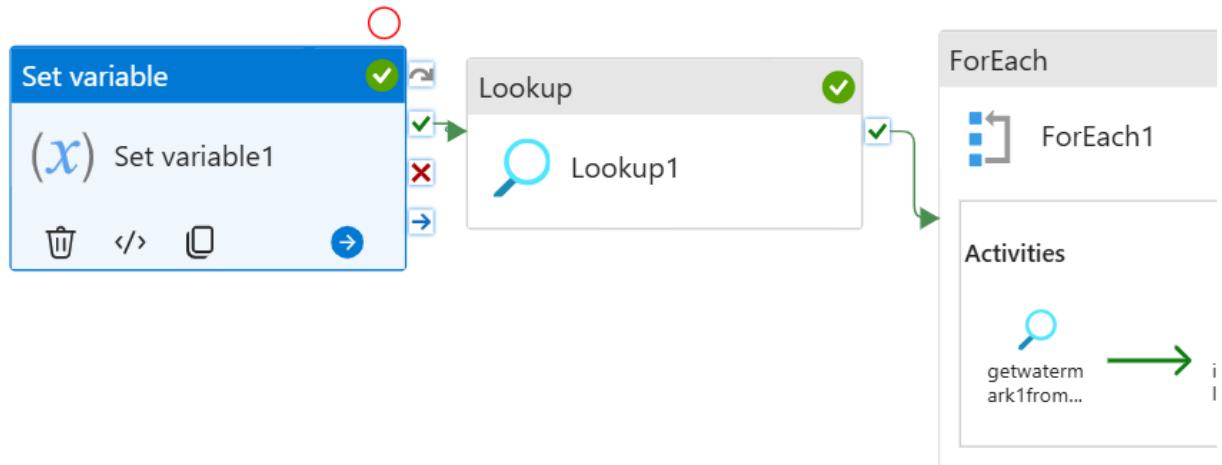
<input type="checkbox"/>	Name	Type	Value	<input type="checkbox"/> Treat as null
<input type="checkbox"/>	Ipv	String	<input type="text"/>	<input type="checkbox"/>
<input type="checkbox"/>	tablename	String	<input type="text"/>	<input type="checkbox"/> Treat as null

Step 22 : Click on add dynamic content give the query

@activity('getmaxvaluefromsource').output.firstRow.maxvalue

<input type="checkbox"/>	Name	Type	Value
<input type="checkbox"/>	Ipv	String	<input type="text"/> @activity('getmaxvaluefromsource').o...
<input type="checkbox"/>	tablename	String	<input type="text"/> @item().tablename

Step 23 : create setvariable 1 and in value put utc()



General **Settings** User properties

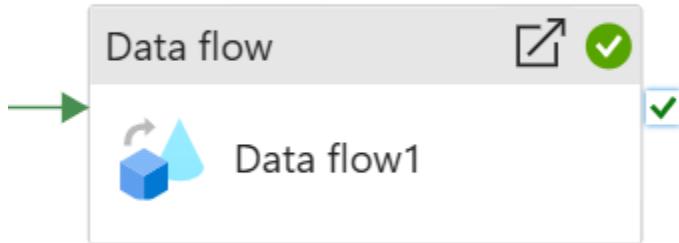
Variable type (i)

Pipeline variable Pipeline return value

Name *

Value

Step 24 : Create the dataflow scdtype1



Step 25 : SCDTYPE1

```
CREATE TABLE Company_scdtype1 (
    CompanyID int,
    Name VARCHAR(50),
    Address VARCHAR(255),
    PhoneNumber bigint,
    Email NVARCHAR(255),
    createdby varchar(100), createddate datetime, updateby varchar(100),
    updatedate datetime, hashkey bigint
)
```

This is the file which I have already created for my incremental load after debug the file got generated from my db to gen2

Name	Last Modified	Content Type	Size
Company_2025-03-09T17:38:13.4154517Z.csv	3/9/2025, 1:38:27 PM		358 B
Company_2025-03-09T17:51:45.9705466Z.csv	3/9/2025, 1:52:00 PM		102 B

STEPS TO CREATE DATA FLOW IN SYNAPSE

STEP 1 : To create dataflow before starting the process click on data flow debug

The screenshot shows a configuration panel with two main sections. The first section, "Compute size", has a dropdown menu set to "Small". The second section, "Debug time to live", also has a dropdown menu set to "1 hour". Both sections include a small downward arrow icon indicating they can be expanded.

STEP 2 : drag and drop the data flow and in setting →dataflow →click for new →give name for the data flow

STEP 3 : My source is gen2 storage account which I have upload the 3 csv files

In sources setting →source type click for inline →inline data type change into delimited text → and in linked services give synapse storage account where you upload 3 csv files.

The screenshot shows the "Source settings" tab selected in the top navigation bar. The page contains several configuration fields:

- Output stream name ***: A text input field containing "source".
- Description**: A text input field containing "Import data from wspabi-WorkspaceDefaultStorage".
- Source type ***: A selection button group with three options: "Integration dataset" (disabled), "Inline" (selected), and "Workspace DB".
- Inline dataset type ***: A dropdown menu showing "DelimitedText".
- Linked service ***: A dropdown menu showing "wspabi-WorkspaceDefaultStorage". To its right are "Test connection", "Edit", and "New" buttons.
- Skip line count**: An empty text input field.
- Sampling ***: A radio button group with "Disable" selected (indicated by a blue outline).

STEP 4 : in source options →I gave my file path where my files are located

File path * / /

STEP 5: In source options check first row as header

First row as header



STEP 6 : In projection → click on import the schema

STEP 7 : Click on data preview (I have imported day1 csv file)

Source settings		Source options		Projection		Optimize		Inspect		Data preview			
												TOTAL 6	
Number of rows	+ INSERT 6	*	UPDATE 0	×	DELETE 0	*	+ UPSERT 0	🔍	LOOKUP 0	✖	ERROR 0		
		⟳ Refresh ↻	Typecast ↻	⤿ Modify ↻	⤿ Map drifted ↻	⤿ Statistics ↻	X Remove	⤿ Export to CSV ↻					
↑↓	DepartmentID	12s ↑↓	Name	abc ↑↓	ManagerID	12s ↑↓	joindatedep	abc ↑↓					
+	1		it		100		2020-01-01 00:00:00.0000000						
+	2		finance		101		2020-01-01 00:00:00.0000000						
+	3		customer		102		2020-01-01 00:00:00.0000000						
+	4		devops		103		2020-01-01 00:00:00.0000000						
+	5		sap		104		2020-01-01 00:00:00.0000000						
+	6		cyber		105		2024-08-01 00:00:00.0000000						

STEP 8 : By clicking + in the sources I'm gonna start new transformation called **select**

In select transformation I wanna rename the column name by doing this step

Step1 : delete all the column in here

<input type="checkbox"/> sourcegen's column	⤿	Name as	⤿
<input type="checkbox"/> 12s id	⤿	→ id	+
<input type="checkbox"/> abc name	⤿	→ name	+
<input type="checkbox"/> abc city	⤿	→ city	+
<input type="checkbox"/> 12l phononenumber	⤿	→ phononenumber	+

Step 2 : to do this delete all the column and click on + symbol click on rule-bases mapping.

Concat('src_',\$\$)

By doing this it will change all the column name into
update_id, update_name, update_city, update_phonenumber

<input type="checkbox"/>	1==1	⤿	⤿	⤿	concat('src_',\$\$)	abc	+	⠀	⠀
--------------------------	------	---	---	---	---------------------	-----	---	---	---

STEP 9 : click on preview data and refresh (column name changed)

Select settings Optimize Inspect Data preview ●

Number of rows + INSERT 6 UPDATE 0 DELETE 0 UPSERT 0 LOOKUP 0 ERROR 0 TOTAL 6

Refresh | Typecast | Modify | Map drifted | Statistics | Remove | Export to CSV |

	src_DepartmentID	src_Name	src_ManagerID	src_joindatedep
+	1	it	100	2020-01-01 00:00:00.0000000
+	2	finance	101	2020-01-01 00:00:00.0000000
+	3	customer	102	2020-01-01 00:00:00.0000000
+	4	devops	103	2020-01-01 00:00:00.0000000
+	5	sap	104	2020-01-01 00:00:00.0000000
+	6	cyber	105	2024-08-01 00:00:00.0000000

STEP 10 : By clicking + in the **select(rename)** Im gonna start new transformation called **drivedcolumn**

STEP 11 : in derived column , have to create a new column (hashkey) and use concat expressions

```
crc32(concat(toString(src_DepartmentID),src_Name,toString(src_ManagerID),toString(src_joindatedep)))
```



STEP 12 : I wanna add my target (sink)

In source setting → source type → inline → inline dataset type → sql db
(mylocalserver) → linked services → servername (ssms)

Source settings Source options Projection Optimize Inspect Data preview ●

Output stream name * target Learn more ↗

Description Import data from AzureSqlDatabase1 Reset

Source type * Integration dataset Inline Workspace DB

Inline dataset type * Azure SQL Database

Linked service * AzureSqlDatabase1 Test connection Edit New Connection successful

Sampling * ⓘ Enable Disable

STEP 13 : I wanna compare this 2 column whether this is existing or not

`select empid, hashkey from emp_scdtype1`

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

Input Table Query Stored procedure

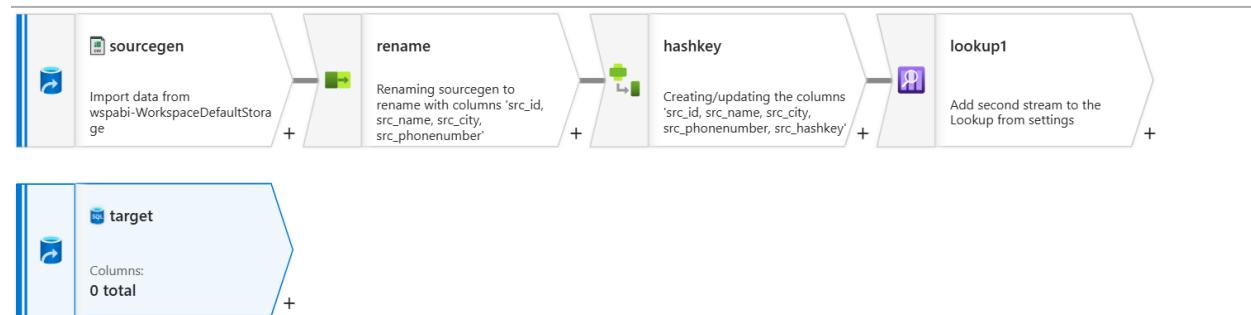
Query * ⓘ `select DepartmentID, hashkey from Department1_scdtype1` 

Incremental column ⓘ

Isolation level ⓘ Read uncommitted

STEP 14 : click on projection → click on import schema

STEP 15: By clicking + in the **derived(hashkey)** → start new transformation called **lookup**



STEP 16 : in lookup setting → lookup stream (select target (sql db) and in lookup conditions → select update_id for left and empid for right

Lookup settings Optimize Inspect Data preview ●

Output stream name: lookup · [Learn more](#)

Description: Lookup on 'hashkey' from 'target'

Primary stream *: hashkey

Lookup stream *: target

Match multiple rows: ⓘ

Match on *: Any row

Lookup conditions *: Left: hashkey's column Right: target's column

12s src_DepartmentID == 123 DepartmentID

+

STEP 17 : in data preview→refresh

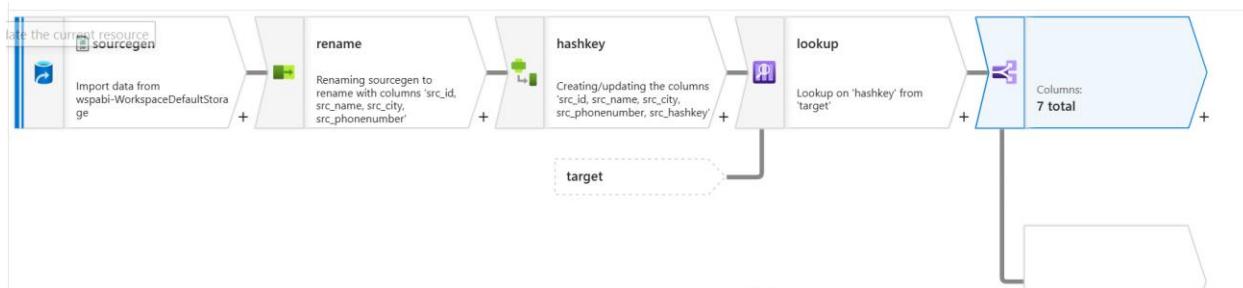
Lookup settings Optimize Inspect **Data preview** ●

Number of rows: + **INSERT** 6 * **UPDATE** 0 × **DELETE** 0 *+ **UPSERT** 0 ⚡ **LOOKUP** 0 ✖ **ERROR** 0

Refresh | Typecast | Modify | Map drifted | Statistics | Remove | Export to CSV |

	src_DepartmentID	src_Name	src_ManagerID	src_jointeddep	src_hashkey	DepartmentID
+	1	it	100	2020-01-01 00:00:00.000000	2301808087	NULL
+	2	finance	101	2020-01-01 00:00:00.000000	411533361	NULL
+	3	customer	102	2020-01-01 00:00:00.000000	1867630511	NULL
+	4	devops	103	2020-01-01 00:00:00.000000	3852105114	NULL
+	5	sap	104	2020-01-01 00:00:00.000000	135914675	NULL
+	6	cyber	105	2024-08-01 00:00:00.000000	1212600369	NULL

STEP 18 : By clicking + in the **lookup**→ start new transformation called **conditional split**



STEP 19 : delete the default stream and add the new (add conditional stream)

STEP 20 : if the custid is null consider this as a new record

`isNull(DepartmentId)`

If the custid is not null consider this as a existing record and update this

Hashkey is already generated and it will generate another hashkey use the below code
code

`src_DepartmentId==DepartmentId && src_hashkey!=hashkey`

Conditional split settings Optimize Inspect Data preview ●

Output stream name * Learn more ↗

Description ⌂ Reset

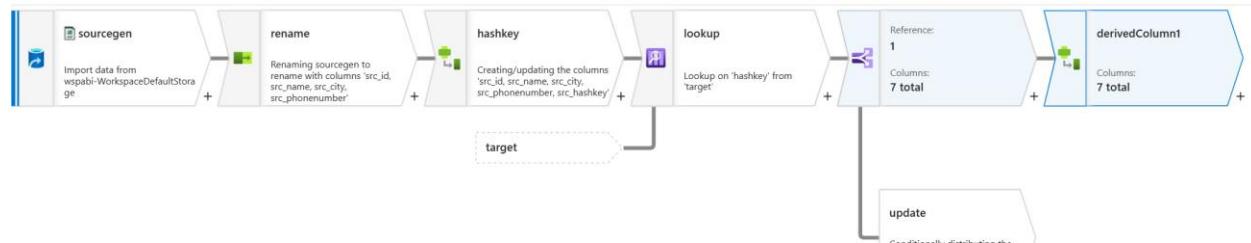
Incoming stream *

Split on First matching condition All matching conditions

Split condition

Stream names	Condition
<input type="text" value="insert"/>	<code>isNull(src_DepartmentID)</code> <input type="button" value="x"/> <input type="button" value="+"/>
<input type="text" value="update"/>	<code>src_DepartmentID==DepartmentID && src_hashkey!=hashkey</code> <input type="button" value="x"/> <input type="button" value="+"/>

STEP 21 : By clicking + in the insert Im gonna start new transformation **called derived column** (to insert audit column)



STEP 22 : I wanna insert audit column (createdby, createddate, updateby, updatedate)

Add →new columns

Derived column's settings Optimize Inspect Data preview ●

Description: Creating/updating the columns 'src_DepartmentID, src_Name, src_ManagerID, src_JointedDep'

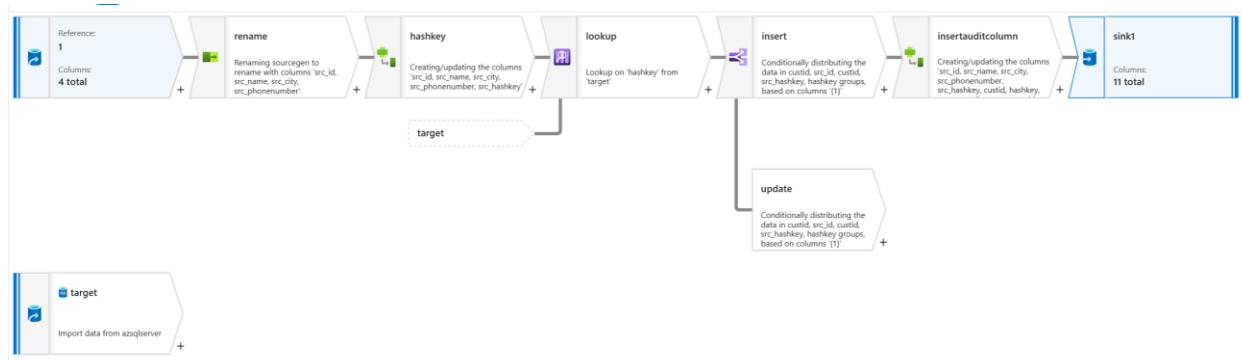
Incoming stream *: split1@insert

Columns *

Column	Expression
src_createdby	'dataflow'
src_createddate	currentTimestamp()
src_updatedby	'dataflow'
src_updatedate	currentTimestamp()

Add → expression as well

STEP 23 : By clicking + in the **derived** (audit) add new transformation called **sink**



STEP 23 : In sink → sink type click inline → inline dataset type click sqldb → in linked services click your ssms(server)

Sink Settings Errors Mapping Optimize Inspect Data preview

Output stream name * Learn more

Description Reset

Incoming stream *

Sink type * Integration dataset Workspace DB Cache

Inline dataset type * Azure SQL Database

Linked service * AzureSQLDatabase1 Test connection Edit New Connection successful

Options Allow schema drift
 Validate schema

STEP 24 : In setting if you unable to see the scheman refresh

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

Schema name * Refresh Success

STEP 25 : update the schema and table name and uncheck the tempdb

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

Schema name *

Table name *

Table action None Recreate table Truncate table

Update method Allow insert
 Allow delete
 Allow upsert
 Allow update

Use tempdb

STEP 26 : in mapping → uncheck the auto mapping and click → importschema

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

Skip duplicate output columns

Auto mapping Add mapping Delete Reset Import schema View schema 9 mappings: All outputs mapped

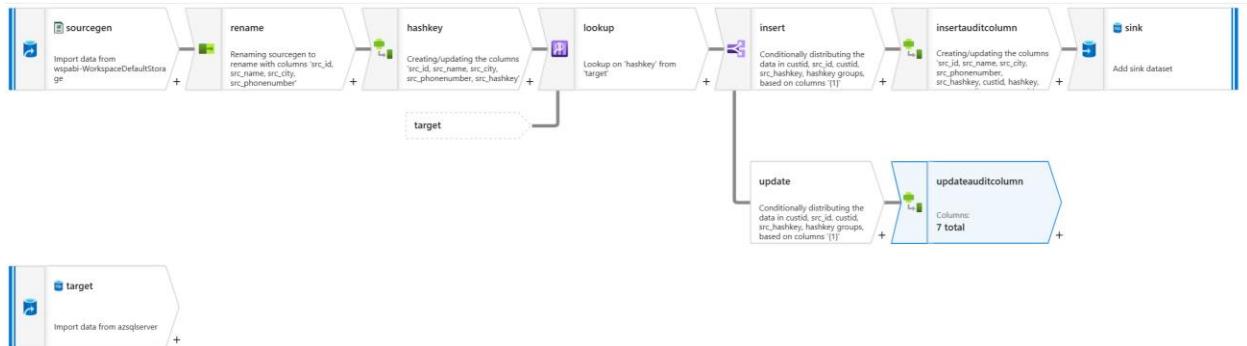
Input columns	Output columns
src_DepartmentID	DepartmentID
src_Name	Name
src_ManagerID	ManagerID
src_joindatedep	joindatedep
src_createdby	createdby
src_createddate	createddate
src_updatedby	updateby
src_updatedate	updatedate

Once the importschema inserted click on reset

STEP 27 : give the values

STEP 28 : click on data preview

STEP 30 : By clicking + in the **update** I'm gonna start new transformation called **derived column** (to update audit column)



STEP 31 : in add column just give the updateby and update date

Conditional split settings

Output stream name * Learn more [🔗](#)

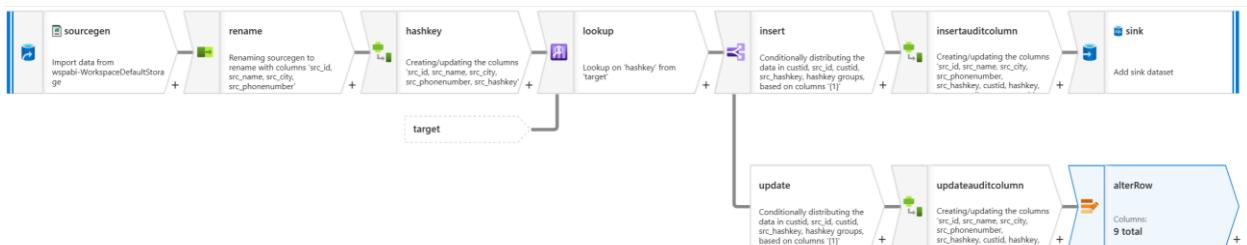
Description [Reset](#)

Incoming stream *

Split on First matching condition All matching conditions

Split condition	Stream names	Condition
insert	<input type="text" value="insert"/>	<input type="text" value="isNull(src_DepartmentID)"/> x [edit]
update	<input type="text" value="update"/>	<input type="text" value="src_DepartmentID == DepartmentID && src_hashkey != hashkey"/> x [edit]

STEP 32 : By clicking + in the **derivedcolumn** I'm gonna start new transformation called **alter row**



STEP 33 : Permission to modify the data

Alter row settings Optimize Inspect Data preview

Output stream name * alterRow [Learn more](#)

Description Add expressions to alter rows [Reset](#)

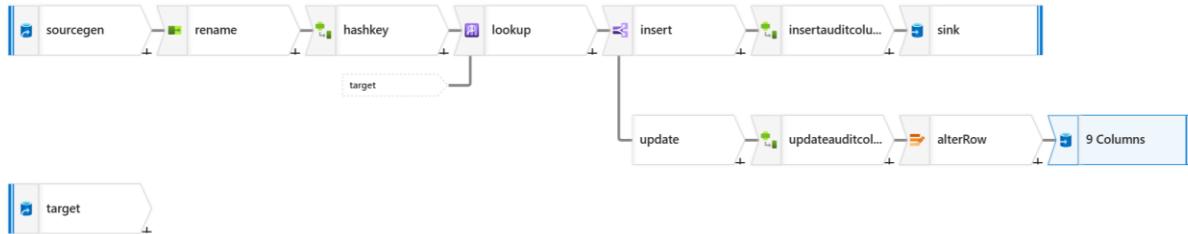
Incoming stream * updateauditcolumn

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

* Update if **1==1**

Updateif 1==1

STEP 34 : By clicking + in the **alterrow** Im gonna start new transformation called **sink**



STEP 35 : give sql db and server name (ssms)

Sink Settings Errors Mapping Optimize Inspect Data preview

Output stream name * sink2 [Learn more](#)

Description Add sink dataset [Reset](#)

Incoming stream * alterRow1

Sink type *

Integration dataset	Inline	Workspace DB	Cache
---------------------	--------	--------------	-------

Inline dataset type * Azure SQL Database

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

Connection successful

STEP 36 : in setting → give schema and table name

Schema name *

dbo

Table name *

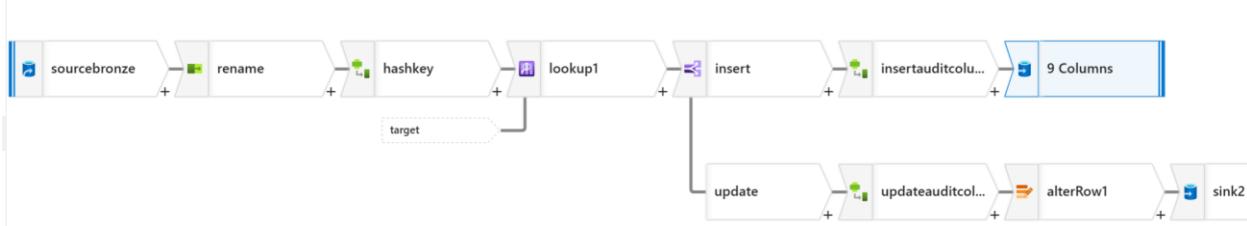
Department1_scdtype1

STEP 37 : in mapping → uncheck auto mapping → click on importschema and click reset

STEP 38 : delete the createby and createdate I only need updateby and updatedate

STEP 39 : in setting give key column as empid (is my keycolumn based on empid I will update data)

STEP 40 : published successfully (THIS IS THE FLOW)



STEP 41 : pipeline debug

Parameters Variables Settings Output

Pipeline run ID: 63da5865-099e-42c1-9277-506fb0c10274 [@](#) [⟳](#) [ⓘ](#) Pipeline status ✓ Succeeded [View debug run consumption](#)

All status [▼](#) [Monitor in Azure Metrics](#) [Export to CSV](#) [▼](#)

Showing 1 - 1 of 1 items

Activity name	Activity st...	Activit...	Run start	Duration	Integration runtime
Data flow	✓ Succeeded	Data flow	3/5/2025, 3:38:17 PM	1m 22s	AutoResolveIntegrationRuntime (Australia)

STEP 44 : output in ssms

Results Messages

	DepartmentID	Name	ManagerID	joindatedep	createdby	createddate	updateby	updatedate	hashkey
1	1	it	100	2020-01-01 00:00:00.000	dataflow	2025-03-13 14:27:56.237	dataflow	2025-03-13 14:27:56.237	2301808087
2	6	cyber	105	2024-08-01 00:00:00.000	dataflow	2025-03-13 14:27:56.237	dataflow	2025-03-13 14:27:56.237	1212600369
3	6	devops	105	2024-08-01 00:00:00.000	dataflow	2025-03-13 14:27:56.237	dataflow	2025-03-13 14:27:56.237	3919136540
4	6	devops	1066	2024-08-01 00:00:00.000	dataflow	2025-03-13 14:27:56.237	dataflow	2025-03-13 14:27:56.237	3456138492
5	3	customer	102	2020-01-01 00:00:00.000	dataflow	2025-03-13 14:27:56.237	dataflow	2025-03-13 14:27:56.237	1867630511
6	4	devops	103	2020-01-01 00:00:00.000	dataflow	2025-03-13 14:27:56.237	dataflow	2025-03-13 14:27:56.237	3852105114
7	5	sap	104	2020-01-01 00:00:00.000	dataflow	2025-03-13 14:27:56.237	dataflow	2025-03-13 14:27:56.237	135914675
8	2	finance	101	2020-01-01 00:00:00.000	dataflow	2025-03-13 14:27:56.237	dataflow	2025-03-13 14:27:56.237	411533361

SCDTYPE 2

```
create TABLE product2_scdtype2 (
    productid int,
    Name NVARCHAR(255),
    joindate datetime, createdby varchar(100), createddate datetime, updateby
    varchar(100), updatedate datetime, hashkey bigint, isactive int
)
```

This the file which I have already created for my incremental load after debug the file got generated from my db to gen2

dbo > Department

Name	Last Modified	Content Type	Size
Department_2025-03-09T17:38:11.8997446Z.csv	3/9/2025, 1:38:28 PM		134 B
Department_2025-03-09T17:51:43.8732056Z.csv	3/9/2025, 1:51:59 PM		46 B

STEPS TO CREATE DATA FLOW IN SYNAPSE

STEP 1 : To create dataflow before starting the process click on data flow debug

The screenshot shows two configuration sections. The first section, "Compute size", has a dropdown menu set to "Small". The second section, "Debug time to live", has a dropdown menu set to "1 hour".

STEP 2 : drag and drop the data flow and in setting →dataflow →click for new →give name for the data flow

The screenshot shows the "Settings" tab selected in a data flow configuration. It includes tabs for General, Settings (selected), Parameters, and User properties. A "Data flow *" dropdown is set to "Select...", and a "New" button is available.

STEP 3 : My source is gen2 storage account which I have upload the 3 csv files

In sources setting →source type click for inline →inline data type change into delimited text → and in linked services give synapse storage account where you upload 3 csv files.

The screenshot shows the "Source settings" tab in a data flow configuration. It includes tabs for Source settings (selected), Source options, Projection, Optimize, Inspect, and Data preview. Under "Source settings", there are fields for "Output stream name" (set to "source"), "Description" (set to "Add source dataset"), and "Source type" (set to "Inline"). Other options include "Integration dataset", "Workspace DB", and "DelimitedText" for "Inline dataset type". A "Linked service" dropdown is set to "wspabi-WorkspaceDefaultStorage", with "Test connection" and "Edit" buttons, and a "New" button.

STEP 4 : in source options → I gave my file path where my files are located

File Wildcard

silver / product2 / product2_2025-03-12T15:34... [Browse](#)

STEP 5: In source options check first row as header

First row as header



STEP 6 : In projection → click on import the schema

Import schema [Clear schema](#) [Schema options](#)

Column name	Type	Format
productid	12s short	Specify format
Name	abc string	Specify format
joindate	abc string	Specify format

STEP 7 : Click on data preview (I have imported day1 csv file)

Number of rows		INSERT 5	UPDATE 0	DELETE 0	UPSERT 0	LOOKUP 0	ERROR 0	TOTAL N/A
↻ Refresh	Typecast	Modify	Map drifted	Statistics	X Remove	Export to CSV	①	
↑↓	productid	12s ↑↓	Name	abc ↑↓	joindate			abc ↑↓
+	1		remote		2020-01-01 00:00:00.0000000			
+	2		charger		2020-01-01 00:00:00.0000000			
+	3		candle		2020-01-01 00:00:00.0000000			
+	4		laptop		2020-01-01 00:00:00.0000000			
+	5		desktop		2020-01-01 00:00:00.0000000			

STEP 8 : By clicking + in the sourcesgen Im gonna start new transformation called **select**

In select transformation I wanna rename the column name by doing this step

Step1 : delete all the column in here

sourcegen's column	Name as	
12s id	id	+
abc name	name	+
abc city	city	+
12l phononenumber	phononenumber	+

Step 2 : to do this delete all the column and click on + symbol click on rule-bases mapping.

`Concat('src_',$$)`

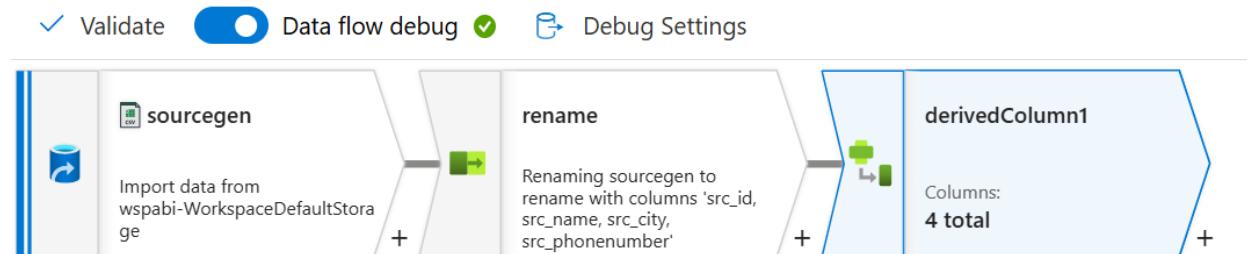
By doing this it will change all the column name into
`src_id,src_name,src_city,src_phonenumber`



STEP 9 : click on preview data and refresh (column name changed)

Select settings	Optimize	Inspect	Data preview	▲
Number of rows	+ INSERT 5	* UPDATE 0	✗ DELETE 0	*+ UPSERT 0
				✗ LOOKUP 0
				✗ ERROR 0
				TOTAL 5
⟳ Refresh ↴	Typecast ↴	Modify ↴	Map drifted ↴	Statistics ↴
↑↓ src_productid	12s ↑↓	src_Name	abc ↑↓	src_joindate
+	1	remote		2020-01-01 00:00:00.0000000
+	2	charger		2020-01-01 00:00:00.0000000
+	3	candle		2020-01-01 00:00:00.0000000
+	4	laptop		2020-01-01 00:00:00.0000000
+	5	desktop		2020-01-01 00:00:00.0000000

STEP 10 : By clicking + in the **select(rename)** Im gonna start new transformation called **derivedcolumn**



STEP 11 : in derived column , have to create a new column (hashkey) and use concat expressions

`crc32(concat(toString(src_productid),src_Name,toString(src_joindate)))`



STEP 12 : I wanna add my target (sink)

In source setting → source type → inline → inline dataset type → sql db → linked services → servername (ssms)

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

- Output stream name ***: target
- Description**: Add source dataset
- Source type ***: Inline (selected)
- Inline dataset type ***: Azure SQL Database
- Linked service ***: azsqlserver
- Sampling ***: Disable (selected)

STEP 13 : I wanna compare this 2 column whether this is existing or not

select custid,hashkey from cust_scptype2 where isactive=1

The configuration for the query input is as follows:

- Input**: Table (radio button unselected), Query (radio button selected), Stored procedure (radio button unselected).
- Query ***:
select productid, hashkey from product2_scptype2 where isactive=1
- Incremental column**: (empty field)
- Isolation level**: Read uncommitted

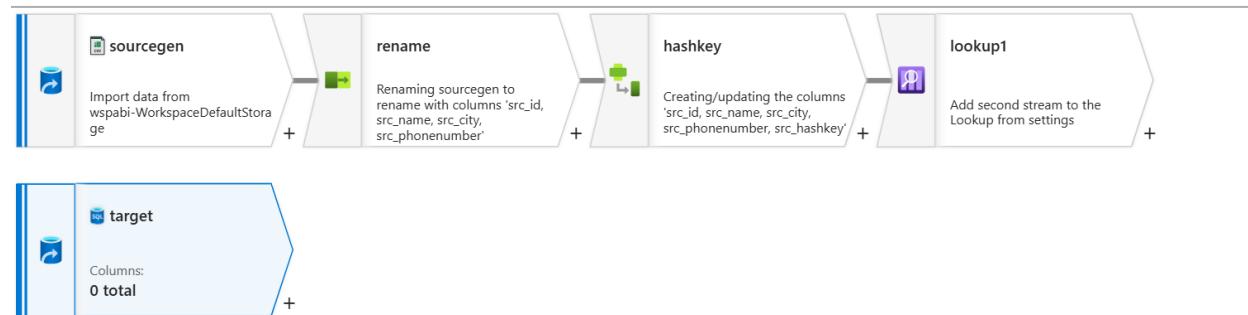
STEP 14 : click on projection → click on import schema

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

Import schema **Clear schema** **Schema options** **Overwrite schema**

Column name	Type
productid	integer
hashkey	long

STEP 15: By clicking + in the **derived(hashkey)** → start new transformation called **lookup**



STEP 16 : in lookup setting → lookup stream (select target (sql db) and in lookup conditions → select src_productid for left and productid for right

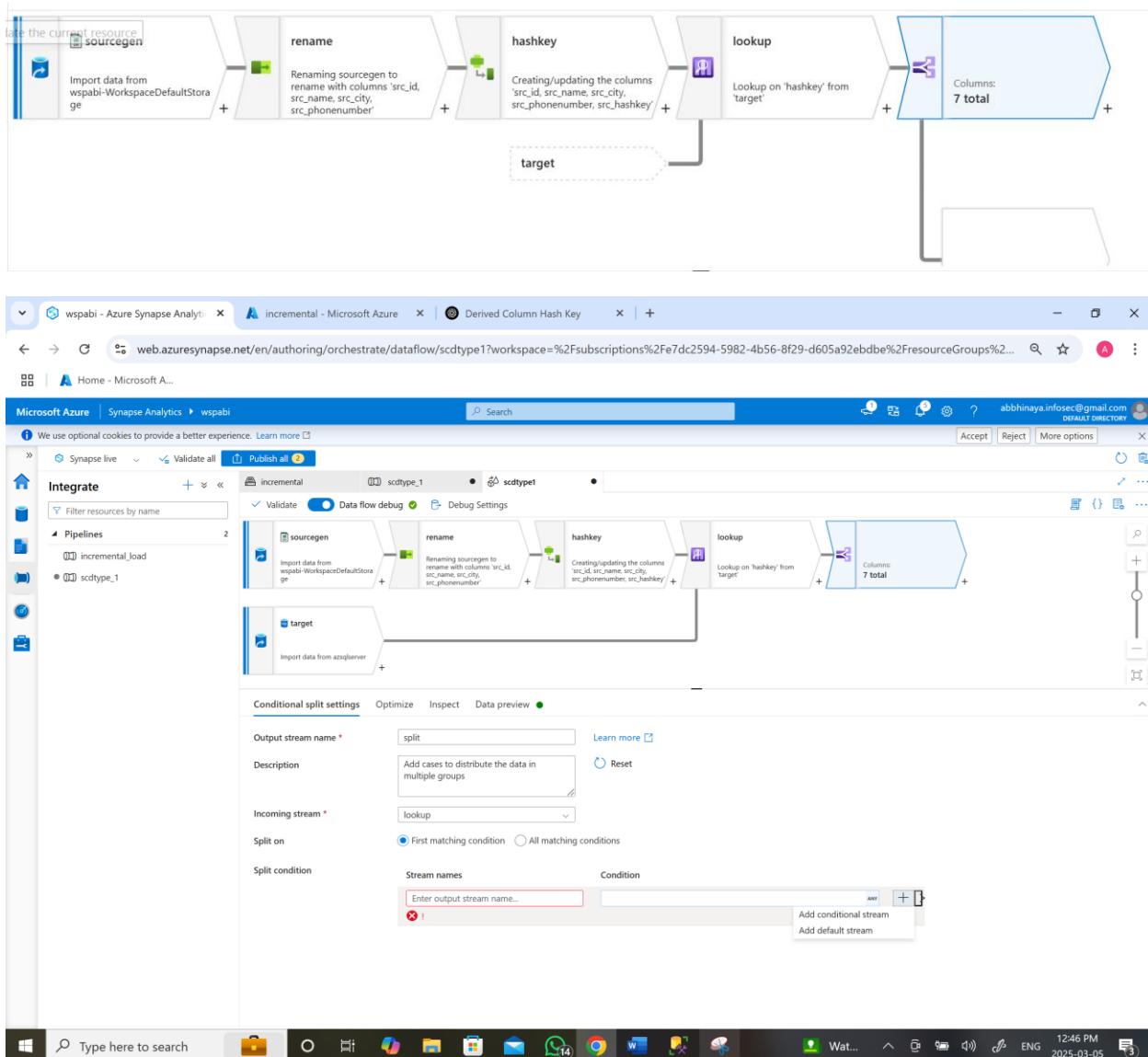


STEP 17 : in data preview→refresh

Data Preview									
		Typecast		Modify		Map drifted		Statistics	
↑↓	src_productid 12s ↑↓	src_Name abc ↑↓	src_joindate abc ↑↓	src_hashkey 12l ↑↓	productid 123 ↑↓	hashkey 12l ↑↓	isactive 123		
🔍 +	1	remote	2020-01-01 00:00:00.000000	4261544782	1	4261544782	1		
🔍 +	2	charger	2020-01-01 00:00:00.000000	49971383	2	49971383	1		
🔍 +	3	candle	2020-01-01 00:00:00.000000	1851679364	3	1851679364	1		
🔍 +	4	laptop	2020-01-01 00:00:00.000000	186749103	4	186749103	1		
🔍 +	5	desktop	2020-01-01 00:00:00.000000	2452544535	5	2452544535	1		

In the above picture I can only see the src_id,src_name,src_city,src_phonenumber this is my csv file and in the custid,hashkey is null(I don't inserted data in the sqldb), to bring data use conditional split

STEP 18 : By clicking + in the **lookup** → start new transformation called **conditional split**



STEP 19 : if the custid is null consider this as a new record

`isNull(productid)`

If the custid is not null consider this as a existing record and update this

Hashkey is already generated and it will generate another hashkey use the below code
code

`src_productid==productid && src_hashkey!=hashkey`

Conditional split settings

Optimize Inspect Data preview

Output stream name * [Learn more](#)

Description

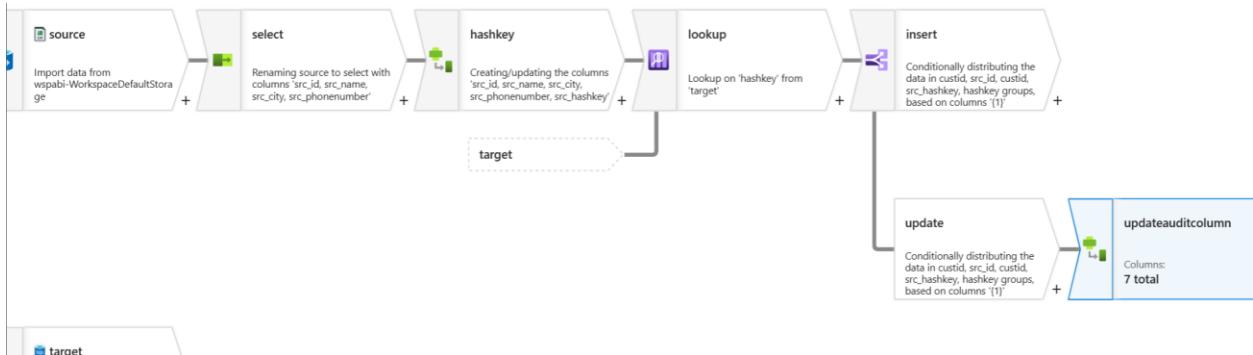
Incoming stream *

Split on First matching condition All matching conditions

Split condition

Stream names	Condition
insert	isNull(productid)
update	src_productid==productid && src_hashkey!=hashkey

STEP 20 : By clicking + in the **update** I'm gonna start new transformation called **derived column** (to update audit column)



STEP 21 : I wanna update audit column (src_updateby, src_updatedate)

Add → new columns

Derived column's settings

Optimize Inspect Data preview

Output stream name * [Learn more](#)

Description

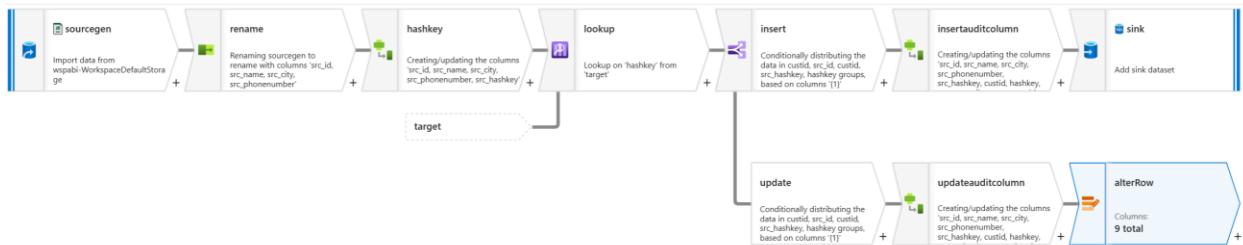
Incoming stream *

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

Columns * [①](#)

Column	Expression
<input type="checkbox"/> src_updateby	'dataflow-update'
<input type="checkbox"/> src_updatedate	currentTimestamp()
<input type="checkbox"/> src_isactive	0

STEP 22 : By clicking + in the **derivedcolumn** Im gonna start new transformation called **alter row**



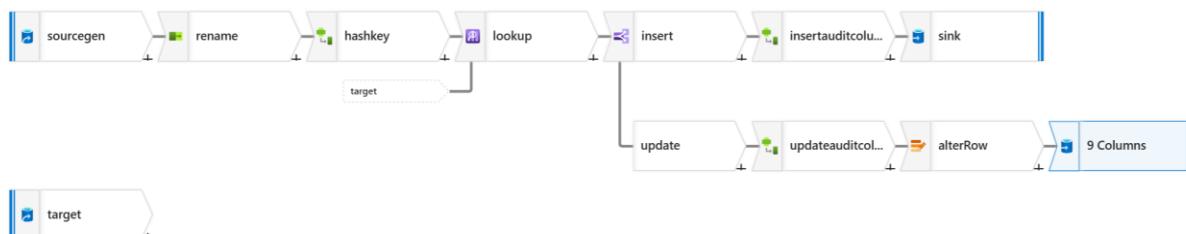
STEP 23 : Permission to modify the data

Alter row settings Optimize Inspect Data preview

Output stream name *	<input type="text" value="alterRow"/>	Learn more
Description	<input type="text" value="Add expressions to alter rows"/> 	
Incoming stream *	<input type="text" value="updateauditcolumn"/>	
Alter row conditions *	<input type="text" value="* Update if"/>	<input type="text" value="1==1"/>
Open expression builder		

Updateif 1==1

STEP 24 : By clicking + in the **alterrow** Im gonna start new transformation called **sink**



STEP 25 : give sql db and server name (ssms)

Sink Settings Errors Mapping Optimize Inspect Data preview

Output stream name * sinkupdate [Learn more](#)

Description Add sink dataset [Reset](#)

Incoming stream * alterRow1

Sink type *

Integration dataset	Inline	Workspace DB	Cache
---------------------	--------	--------------	-------

Inline dataset type * Azure SQL Database

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

Options

Allow schema drift [①](#)

Validate schema [①](#)

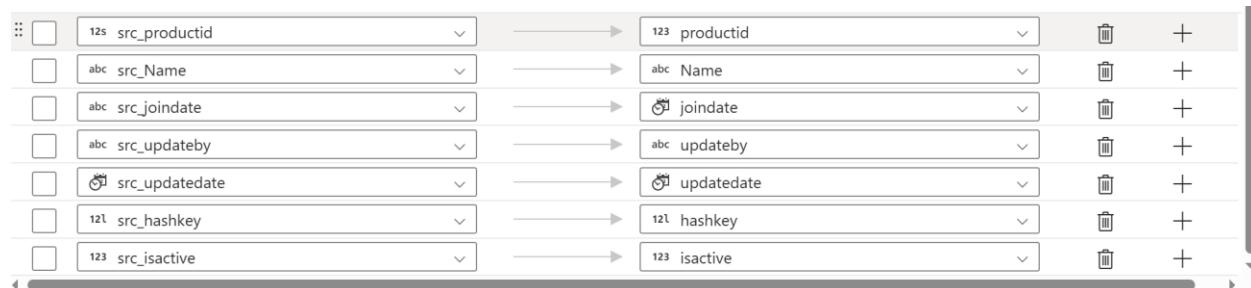
STEP 26 : in setting → give schema and table name

Schema name * dbo

Table name * product2_scdtype2

STEP 27 : in mapping → uncheck auto mapping → click on importschema and click reset

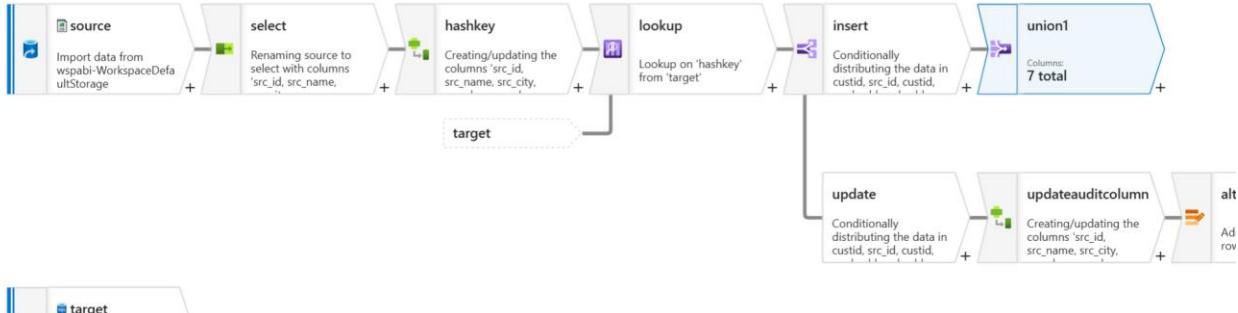
STEP 28 : reset



STEP 29 : in setting give key column as custid (is my keycolumn based on productid I will update data)

123 productid	<input type="button" value="▼"/>	<input type="button" value="+"/>	<input type="button" value="Delete"/>
123 hashkey	<input type="button" value="▼"/>	<input type="button" value="+"/>	<input type="button" value="Delete"/>

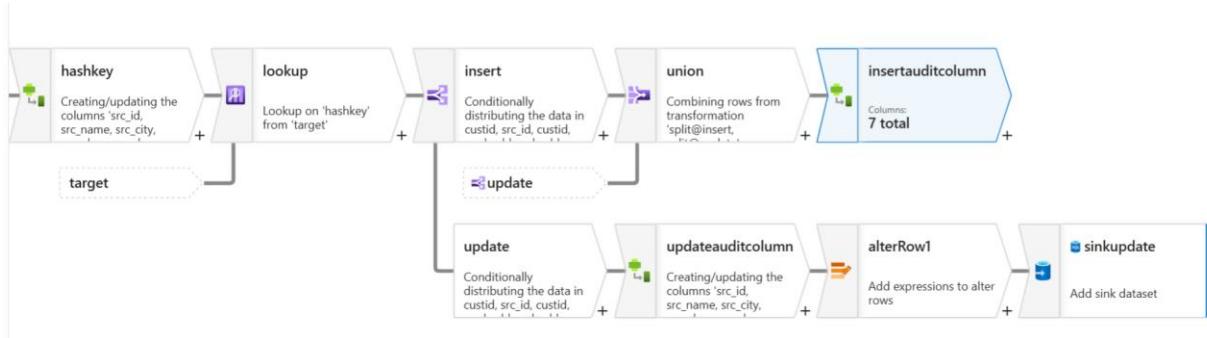
STEP 30 : By clicking + in the **insert** I'm gonna start new transformation **called union**



STEP 31 : In the union settings → union with → give split@update

Union settings	Optimize	Inspect	Data preview
Output stream name *	union	Learn more	
Description	Combining rows from transformation 'split@insert, split@update'	Reset	
Incoming stream *	split@insert		
Union by * <small>(i)</small>	<input checked="" type="radio"/> Name <input type="radio"/> Position		
Union with *	split@update	<input type="button" value="+"/> <input type="button" value="Delete"/>	

STEP 32 : By clicking + in the **union** I'm gonna start new transformation called **derivedcolumn (insert audit column)**



STEP 33 : I wanna insert audit column (createdby, createddate, updateby, updatedate)

Add → new columns

Derived column's settings Optimize Inspect Data preview

Output stream name * [Learn more](#)

Description

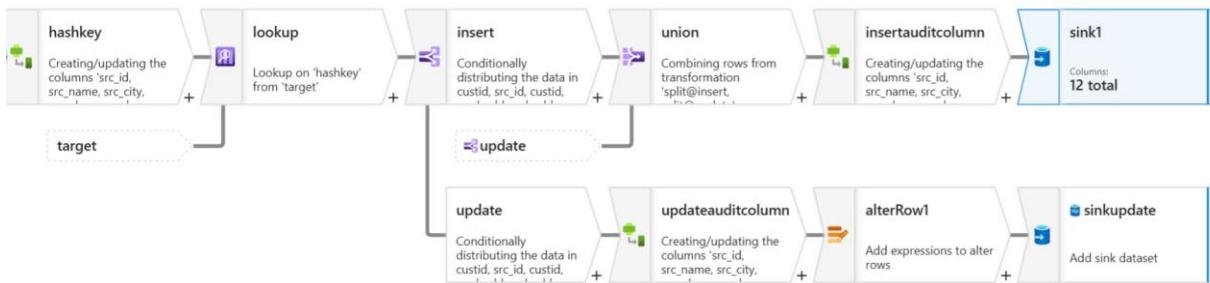
Incoming stream * [Add](#) [Clone](#) [Delete](#) [Open expression builder](#)

Columns * ⓘ

Column	Expression
<input type="checkbox"/> src_createdby	'dataflow'
<input type="checkbox"/> src_createddate	currentTimestamp()
<input type="checkbox"/> src_updateby	'dataflow'
<input type="checkbox"/> src_updatedate	currentTimestamp()
<input type="checkbox"/> src_isactive	1

[Open expression builder](#)

STEP 34 : By clicking + in the **derived column** (insert audit column) add new transformation called **sink**



STEP 35 : In sink →sink type click inline → inline dataset type click sqldb →in linked services click your ssms(server)

Sink Settings Errors Mapping Optimize Inspect Data preview ●

Output stream name * [Learn more](#)

Description [Reset](#)

Incoming stream *

Sink type *
 Integration dataset Inline Workspace DB Cache

Inline dataset type *
 [Azure SQL Database](#)

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

Options
 Allow schema drift [①](#)
 Validate schema [①](#)

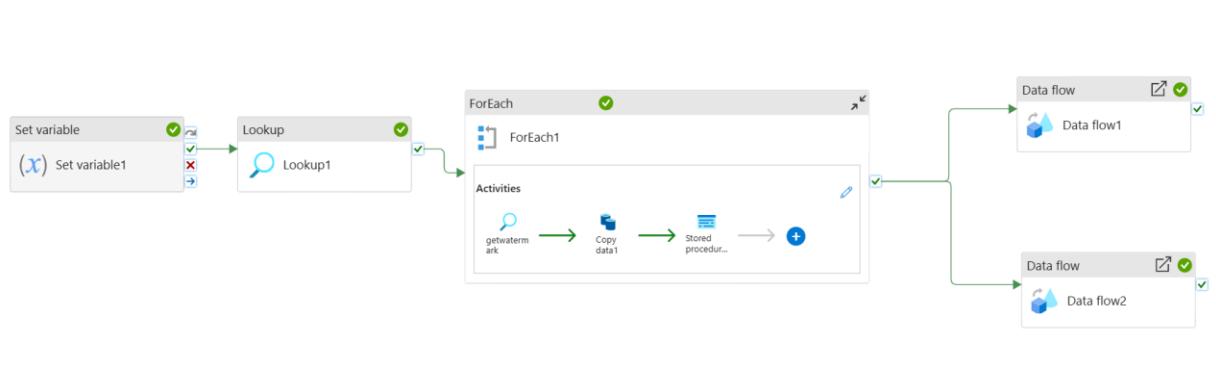
Sink Settings Errors **Mapping** Optimize Inspect Data preview ●

Skip duplicate output columns ⚙

Auto mapping ⚙ 9 mappings: All outputs mapped

Input columns	Output columns
12s src_productid	123 productid
abc src_Name	abc Name
abc src_joindate	⌚ joindate
abc src_createdby	abc createdby
⌚ src_createddate	⌚ createddate
abc src_updateby	abc updateby
⌚ src_updatedate	⌚ updatedate
12l src_hashkey	12l hashkey
123 src_isactive	123 isactive

STEP 36 : publish



STEP 37 : output of ssms

Results Messages

	productid	Name	joindate	createdby	createddate	updateby	updatedate	hashkey	isactive
1	1	remote	2020-01-01 00:00:00.000	dataflow	2025-03-12 17:51:20.377	dataflow	2025-03-12 17:51:20.377	4261544782	1
2	3	candle	2020-01-01 00:00:00.000	dataflow	2025-03-12 17:51:20.377	dataflow	2025-03-12 17:51:20.377	1851679364	1
3	4	laptop	2020-01-01 00:00:00.000	dataflow	2025-03-12 17:51:20.377	dataflow	2025-03-12 17:51:20.377	186749103	1
4	5	desktop	2020-01-01 00:00:00.000	dataflow	2025-03-12 17:51:20.377	dataflow	2025-03-12 17:51:20.377	2452544535	1
5	2	charger	2020-01-01 00:00:00.000	dataflow	2025-03-12 17:51:20.377	dataflow	2025-03-12 17:51:20.377	49971383	1
6	1	tv	2020-01-01 00:00:00.000	dataflow	2025-03-12 21:26:04.967	dataflow-update	2025-03-12 21:33:35.183	3185841148	0
7	3	radio	2020-01-01 00:00:00.000	dataflow	2025-03-12 21:26:04.967	dataflow-update	2025-03-12 21:33:35.183	4016701354	0
8	4	yogurt	2020-01-01 00:00:00.000	dataflow	2025-03-12 21:26:04.967	dataflow-update	2025-03-12 21:33:35.183	1972015168	0
9	5	banana	2020-01-01 00:00:00.000	dataflow	2025-03-12 21:26:04.967	dataflow-update	2025-03-12 21:33:35.183	2392473047	0
10	2	pears	2020-01-01 00:00:00.000	dataflow	2025-03-12 21:26:04.967	dataflow-update	2025-03-12 21:33:35.183	2316910759	0
11	1	tv	2020-01-01 00:00:00.000	dataflow	2025-03-12 21:33:52.970	dataflow	2025-03-12 21:33:52.970	3185841148	1
12	3	radio	2020-01-01 00:00:00.000	dataflow	2025-03-12 21:33:52.970	dataflow	2025-03-12 21:33:52.970	4016701354	1
13	4	yogurt	2020-01-01 00:00:00.000	dataflow	2025-03-12 21:33:52.970	dataflow	2025-03-12 21:33:52.970	1972015168	1
14	5	banana	2020-01-01 00:00:00.000	dataflow	2025-03-12 21:33:52.970	dataflow	2025-03-12 21:33:52.970	2392473047	1
15	2	pears	2020-01-01 00:00:00.000	dataflow	2025-03-12 21:33:52.970	dataflow	2025-03-12 21:33:52.970	2316910759	1

Microsoft Azure | Synapse Analytics > wspabi

Synapse live Validate all Publish all

Integrate + <

Pipelines finalyendtoend 11

Activities

Search activities

finalyendtoend Dataflow10 Dataflow9

Validate Debug Add trigger Data flow debug

finalyendtoend Dataflow10 Dataflow9

Set variable (X) Set variable1 Lookup1 ForEach1 ForEach1 Activities Data flow Data Row1

Parameters Variables Settings Output

Pipeline run ID: fa58f5e1-3f99-43fe-9153-6662c3f92c7d

Pipeline status Succeeded View debug run consumption

All status List Monitor in Azure Metrics Export to CSV

Showing 1 - 20 of 20 items

Activity name	Activity st...	Activit...	Run start	Duration	Integration
Data flow2	Succeeded	Data flow	3/12/2025, 5:44:21 PM	4m 45s	AutoResolv
Data flow1	Succeeded	Data flow	3/12/2025, 5:44:21 PM	4m 58s	AutoResolv
Stored procedure1	Succeeded	Stored procedu	3/12/2025, 5:44:09 PM	5s	selfhostedII
Stored procedure1	Succeeded	Stored procedu	3/12/2025, 5:44:01 PM	16s	selfhostedII
Stored procedure1	Succeeded	Stored procedu	3/12/2025, 5:44:00 PM	16s	selfhostedII
Stored procedure1	Succeeded	Stored procedu	3/12/2025, 5:44:00 PM	14s	selfhostedII
Stored procedure1	Succeeded	Stored procedu	3/12/2025, 5:43:59 PM	6s	selfhostedII

Type here to search

Windows Taskbar: Search, File Explorer, Mail, Edge, Google Chrome, File History, Task View, Weather (-1°C), System, Date (5:51 PM, 2025-03-12)