



Mapping Data Flow – Dimension Table

In this lab, we embark on a journey to create dimension tables using Mapping Data Flow within Azure Data Factory. Our primary focus is on crafting two dimension tables: the Dimension Customer Table and the Dimension Product Table. Here's a step-by-step breakdown of our endeavor:

😊 Dimension Customer Table:

1. We initiate the process by creating an empty dimension customer table in SQL Server Management Studio (SSMS), laying the groundwork for our data flow.
2. Navigating to the Azure Data Factory wizard, we create a new data flow and name it. Then, we add the customer table as a resource, configuring it as our source dataset from the Azure SQL Database.
3. Next, we create a sink for the customer table, specifying Azure Synapse Analytics as the destination dataset.
4. After turning off auto-mapping, we meticulously map the columns between the source and sink datasets.
5. Upon validation, we publish the data flow, integrating it into our pipeline for execution.

😊 Dimension Product Table:

1. Following a similar process, we create the product table in SSMS, preparing it for inclusion in our data flow.
2. Back in Azure Data Factory, we add the product table as a source dataset, repeating the steps for configuring the dataset properties.
3. Additionally, we add the product model table as another source dataset, establishing a left join between the product and product model tables.
4. Extending our data flow, we introduce the product category table as a further source dataset, joining it with the existing product and product model tables.
5. With meticulous attention to mapping details, we ensure the seamless flow of data between sources and sink.
6. Upon validation and publication, our dimension tables are ready for action within the pipeline.

😊 Pipeline Execution:

1. With the dimension tables integrated into our pipeline, we configure the staging settings and validate the pipeline.
2. Following validation, we publish the pipeline, marking its readiness for execution.
3. Triggering the pipeline, we initiate the data flow process, monitoring its progress through the Azure Data Factory interface.
4. Upon completion, we verify the success of the pipeline run by querying the tables in SSMS, confirming the seamless flow of data from source to destination.

In summary, this guide provides a comprehensive walkthrough of creating dimension tables using Mapping Data Flow in Azure Data Factory, from table creation to pipeline execution. By harnessing the power of Mapping Data Flow, organizations can streamline their data transformation workflows, enabling efficient data integration and analytics operations.

To begin with the Lab:

Now we are creating table on the behalf of below scripts. Below script is for customer table and we are creating table accordingly.

The script below is a SQL query that selects specific columns from the [SalesLT].[Customer] table, aliasing it as ct. It retrieves data related to customers from the specified table.

Here's a breakdown of the query:

- **SELECT ct.[CustomerID], ct.[CompanyName], ct.[SalesPerson]:** This part of the query selects three columns (CustomerID, CompanyName, and SalesPerson) from the ct table (which is an alias for [SalesLT].[Customer]).
 - **FROM [SalesLT].[Customer] as ct:** This part specifies the source table from which the data is being selected. It selects data from the [SalesLT].[Customer] table and aliases it as ct, allowing us to reference the table using the shorter alias ct within the query.
-

```
SELECT ct.[CustomerID],ct.[CompanyName],ct.[SalesPerson]  
FROM [SalesLT].[Customer] as ct
```

Dimension Customer Table

1. In this lab we will create a Dimension table using mapping data flow.
2. Here we will create two tables dimension customer table and dimension product table.
3. For that first we need to create our dimension customer table in SSMS. Below you can see that we have created the table and it is empty.

SQLQuery1.sql - dat... (sqladminuser (0)) * X

```

CREATE TABLE [dbo].[DimCustomer](
    [CustomerID] [int] NOT NULL,
    [CompanyName] [varchar](200) NOT NULL,
    [SalesPerson] [varchar](300) NOT NULL
)
WITH
(
    DISTRIBUTION = REPLICATE
)

SELECT * FROM [dbo].[DimCustomer]

```

150 %

Results Messages

CustomerID	CompanyName	SalesPerson
------------	-------------	-------------

4. Now we are going to navigate to the Data Factory wizard and create a new data flow.
5. Here you need to give it a name then from add resource add your resource for the customer table.

Factory Resources

- Pipelines: 4
 - 01-Copy-data-to-logdata
 - 02-Copy-to-Parquet
 - 03-copy-using-query
 - 04_mappingdataflow_factsalestable
- Change Data Capture (preview): 0
- Datasets: 9
- Data flows: 2
 - 01_dataflow_factsalestable
 - 02_Dataflow_Dimension
- Power Query: 0

Properties

General Related

Name * 02_Dataflow_Dimension

Description

Source settings Source options Projection ...

Output stream name * Customertablestream

Learn more Description Add source dataset

Source type * Dataset Inline

Dataset * Select... + New

Options Allow schema drift Infer drifted column types

6. Then you have to add the Dataset for it which is going to be your Azure SQL database. For that click on new then choose Azure SQL Database. After that in set properties first give it a name then choose your linked service then choose customer table as shown below and click on OK.

Set properties

Name

Linked service *

Table name

  Enter manually

Import schema

 From connection/store None Advanced

- After that you need to create a sink from your customer table and then give your sink a name and then you have to choose the data sink which will be your Azure Synapse Analytics.

02_Dataflow_Dimen... ●

✓ Validate Data flow debug

Import data from Customertablestr... + DimCustomerStream
Columns: 15 total

Sink Settings Errors Mapping Optimize Inspect Data preview

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

Description Add sink dataset [Reset](#)

Incoming stream * Customertablestream

Sink type *  Dataset  Inline  Cache

Dataset * [New](#)

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

- Now in set properties, first give it a name then choose your linked service. After that you have to choose the table name then just click on OK.

Set properties

Name

DimCustomerTable

Linked service *

datasynapse1234_pooldb

X ▾



Select from existing table New table

Table name

dbo.DimCustomer

▼



Enter manually

Import schema

From connection/store None

> Advanced

9. After that go to mappings and turn off auto-mapping because we don't want it to auto map.

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

Options Skip duplicate input columns ⓘ Skip duplicate output columns ⓘ

Auto mapping ⏪ ↻ Reset + Add mapping Delete Output format **3 mappings: All outputs mapped**

Input columns	Output columns
123 CustomerID	123 CustomerID
abc CompanyName	abc CompanyName
abc SalesPerson	abc SalesPerson

10. After that just click on Publish all.

Publish all

You are about to publish all pending changes to the live environment. [Learn more](#)

Pending changes (3)

NAME	CHANGE	EXISTING
▽ Datasets		
Customertable	(New)	-
DimCustomertable	(New)	-
▽ Data flows		
02_Dataflow_Dimension	(New)	-

Dimension Product Table

Here the script use is for customer product table.

The script below is commonly referred to as a "dimension table query" or simply a "dimension query." This type of query is used to retrieve data from multiple related tables in a database, typically in a data warehousing or dimensional modeling context.

In this specific query:

- The SalesLT.Product table represents the main dimension table, containing information about products.
- The SalesLT.ProductModel table is joined to the main dimension table (SalesLT.Product) using a left join, indicating that every record from the main dimension table will be included in the result set, regardless of whether there is a matching record in the joined table.
- Similarly, the SalesLT.ProductCategory table is also joined to the main dimension table (SalesLT.Product) using a left join.
- The SELECT statement specifies the columns to be included in the result set, including columns from each of the joined tables. Aliases are used to provide more descriptive names for certain columns, such as ProductName, ProductModelName, and ProductCategoryName.

```
SELECT prod.[ProductID],prod.[Name] as
ProductName,model.[ProductModelID],model.[Name] as
ProductModelName,category.[ProductcategoryID],category.[Name] AS
ProductCategoryName

FROM [SalesLT].[Product] prod
```

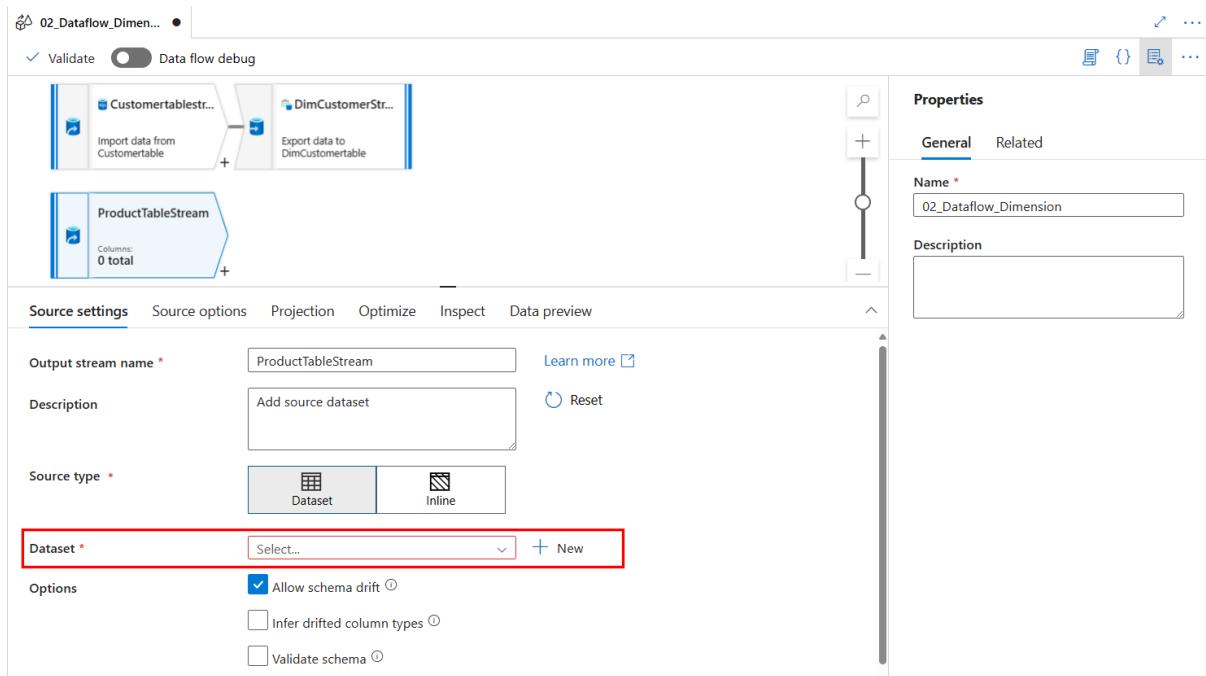
```
LEFT JOIN [SalesLT].[ProductModel] model ON prod.[ProductModelID] =  
model.[ProductModelID]
```

```
LEFT JOIN [SalesLT].[ProductCategory] category ON  
prod.[ProductcategoryID]=category.[ProductcategoryID]
```

1. Now we are going to build another table which is the Product table.
2. First go to SSMS and create your Product table in our dedicated SQL Pool.

```
SQLQuery1.sql - dat...(sqladminuser (0))*  ↗ X  
CREATE TABLE [dbo].[DimProduct](  
    [ProductID] [int] NOT NULL,  
    [ProductModelID] [int] NOT NULL,  
    [ProductcategoryID] [int] NOT NULL,  
    [ProductName] varchar(50) NOT NULL,  
    [ProductModelName] varchar(50) NULL,  
    [ProductCategoryName] varchar(50) NULL  
)  
WITH  
(  
    DISTRIBUTION = REPLICATE  
)  
  
150 %  ◀  
Messages  
Commands completed successfully.  
Completion time: 2024-04-25T14:57:48.4695027+05:30
```

3. After that you have to go to the Data Factory wizard and add your third source in the same data flow. Here you have to give it a name then for the data choose SQL Database and then choose your product table.



4. Now in the set properties first give it a name and then choose your linked service and then the table as shown below. Then click on OK.

Set properties

Name
Product_table

Linked service *
demodb_service

Table name
SalesLT.Product

Enter manually

Import schema
 From connection/store None

> Advanced

5. Then we need to add our fourth source which will be our product model table. Give it a name then for the dataset again add SQL Database as your dataset.

The screenshot shows the 'Source settings' tab selected in the top navigation bar. The main area contains the following fields:

- Output stream name ***: ProductModelStream
- Description**: Add source dataset
- Source type ***: Dataset (selected)
- Dataset ***: Select... (dropdown menu)
- Options**:
 - Allow schema drift ⓘ
 - Infer drifted column types ⓘ
 - Validate schema ⓘ
- Sampling ***:
 - Enable
 - Disable

6. Then again the same steps, this time you have to choose the product model table and then click on OK.

Set properties

Name
ProductModel_table

Linked service *
demodb_service

Table name
SalesLT.ProductModel

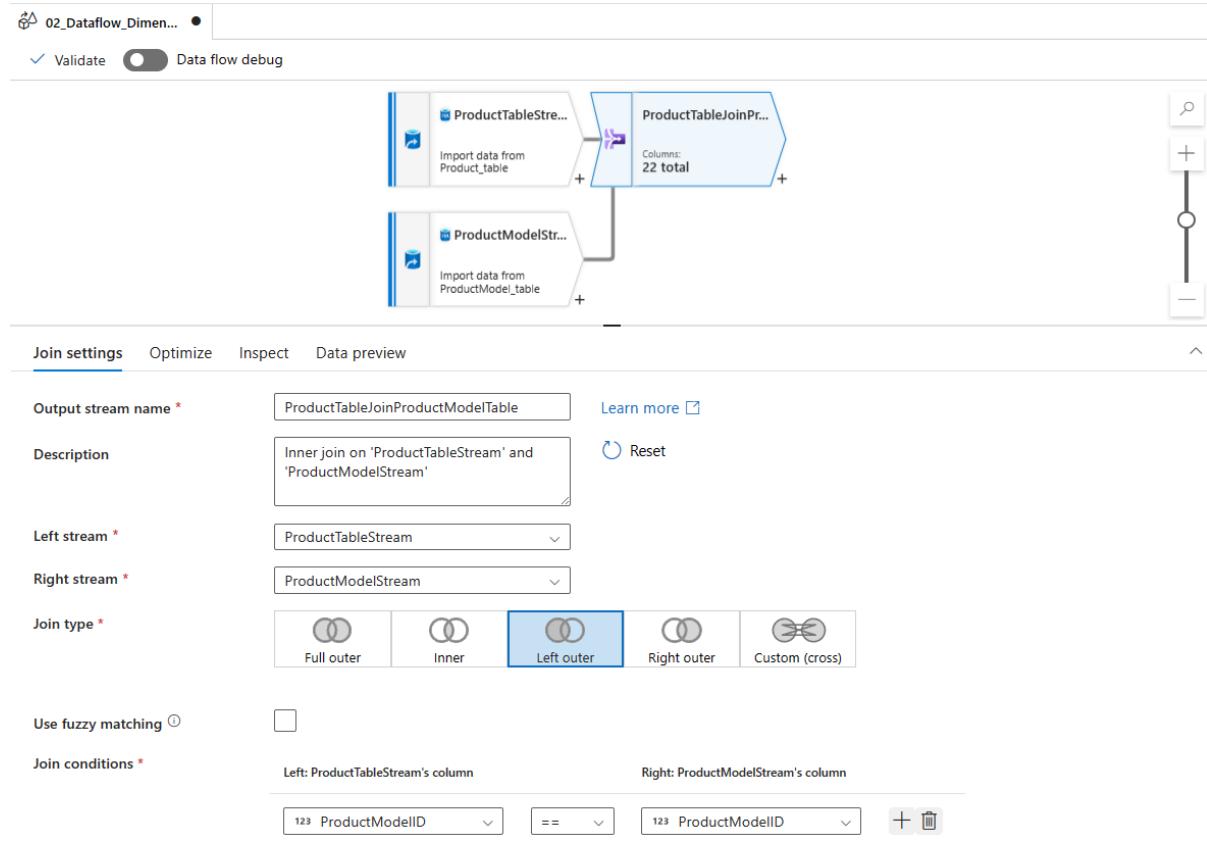
Enter manually

Import schema
 From connection/store None

Advanced

7. Now between these two tables we want to have a join.
8. Here we created a join from the Product table to the product model table. First, we gave it a name then in the left stream we have the product table and in the right

stream, we have the product model table. In the join type, we choose left outer. Then for the join conditions choose Product model ID.



- Now we need to add one more source which will be for our Product category table. Here give it a name first then for the dataset again you have to choose SQL Database.

The screenshot shows the 'Source settings' tab selected in the navigation bar. The main area displays a dataset named 'ProductCategorySt...' with a preview icon showing a single row of data. Below the preview, the 'Output stream name' is set to 'ProductCategoryStream'. The 'Description' field contains the placeholder 'Add source dataset'. Under 'Source type', the 'Dataset' option is selected. In the 'Dataset' dropdown, 'Select...' is chosen, and a '+ New' button is available. Under 'Options', the 'Allow schema drift' checkbox is checked. There are also three unchecked checkboxes for 'Infer drifted column types', 'Validate schema', and 'Sampling'. The sampling section has 'Disable' selected.

10. Then in the set properties, we need to give it a name first, after that we need choose our linked service then the table name as shown below. Then just click on Ok.

Set properties

Name
ProductCategory_table

Linked service *
demodb_service

Table name
SalesLT.ProductCategory

Enter manually

Import schema
 From connection/store None

> Advanced

11. Now we'll need a join and we choose it from our previous join which as product table and product model table.

12. Below you can see that we created a join. First, we gave it a name then in the right stream we chose the Product category table. After that, the join type is left outer. Then in the join conditions we choose Product category ID.

13. After that we are going to add a sink from the join of the product category table.

14. Below you can see that we created the sink. Then we gave it a name after that we need to choose the dataset which will be Azure Synapse Analytics.

15. Here in set properties, give it a name first, then we choose our linked service or say our workspace then we choose the table as shown below. Then just click on Ok.

Set properties

Name
DimProduct_table

Linked service *

datasynapse1234_pooldb

Select from existing table New table

Table name
dbo.DimProduct

Enter manually

Import schema
 From connection/store None

> Advanced

16. Once it is done then you have to go to mappings and disable auto-mapping do the mapping yourself this time.

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

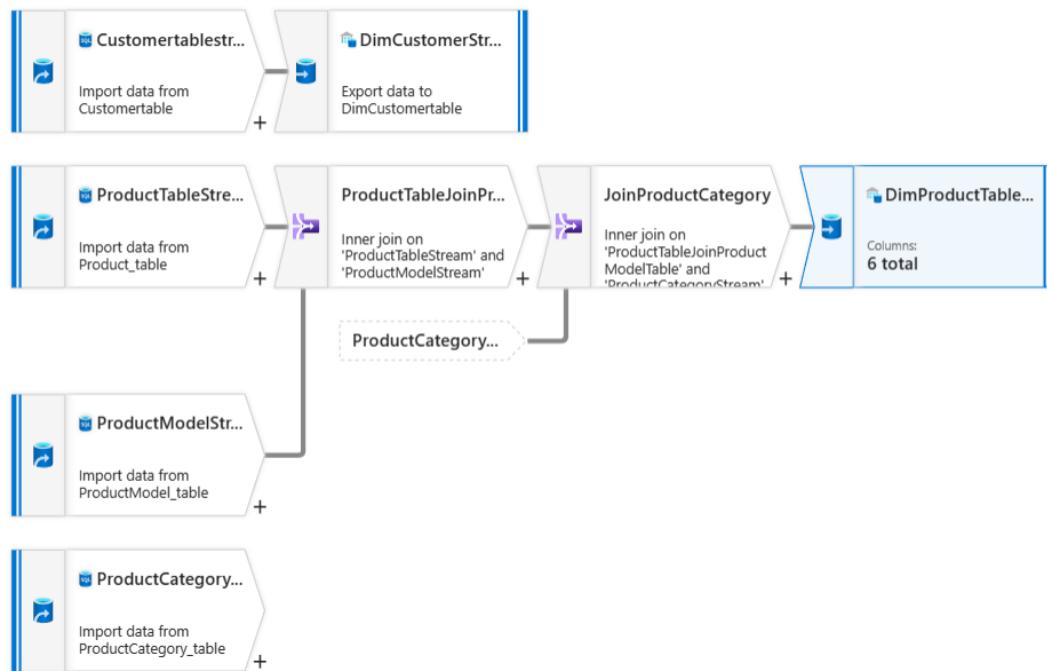
Options Skip duplicate input columns Skip duplicate output columns Auto mapping Reset Output format 6 mappings: 5 column(s) from the output schema left unmapped

Input columns	Output columns
123 ProductID	123 ProductID
	123 ProductModelID
	123 ProductcategoryID
	abc ProductName
	abc ProductModelName
	abc ProductCategoryName

17. Now you have to do the same mappings as shown below.

Input columns			Output columns	
<input type="checkbox"/>	123 ProductID	→	123 ProductID	+ <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
<input type="checkbox"/>	123 ProductTableStream@ProductModelID	→	123 ProductModelID	+ <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
⋮ <input type="checkbox"/>	123 ProductTableStream@ProductCategoryID	→	123 ProductcategoryID	+ <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
<input type="checkbox"/>	abc ProductTableStream@Name	→	abc ProductName	+ <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
<input type="checkbox"/>	abc ProductModelStream@Name	→	abc ProductmodelName	+ <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
<input type="checkbox"/>	abc ProductCategoryStream@Name	→	abc ProductCategoryName	+ <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

18. Below you can see the entire data flow mapping diagram.



19. Now click on validate all then click on publish all. Below you can see that what it is going to publish.

Publish all

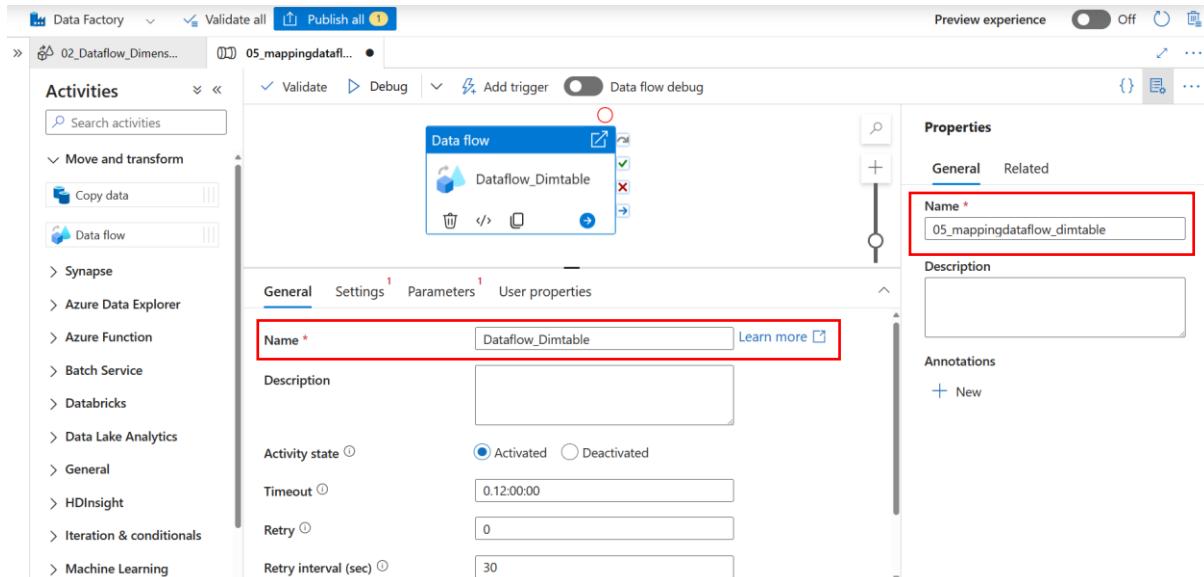
You are about to publish all pending changes to the live environment. [Learn more](#)

Pending changes (5)

NAME	CHANGE	EXISTING
Datasets		
Product_table	(New)	-
ProductModel_table	(New)	-
ProductCategory_table	(New)	-
DimProduct_table	(New)	-
Data flows		
02_Dataflow_Dimension	(Edited)	02_Dataflow_Dimension

20. Once it is published then you have to create a new pipeline and give a name to your pipeline.

21. After creating that you need to expand move and transform then drag the Data flow to the canvas. Then give it a name.



22. After that you need to go to settings and choose your data flow. Then you have to choose your linked service in staging and choose the staging storage container from your service like you did in the previous lab.

General **Settings** Parameters ¹ User properties

Data flow * **02_Dataflow_Dimension**

Run on (Azure IR) * ⓘ AutoResolveIntegrationRuntime

Compute size * ⓘ Small

> Advanced

Logging level * ⓘ Verbose Basic None

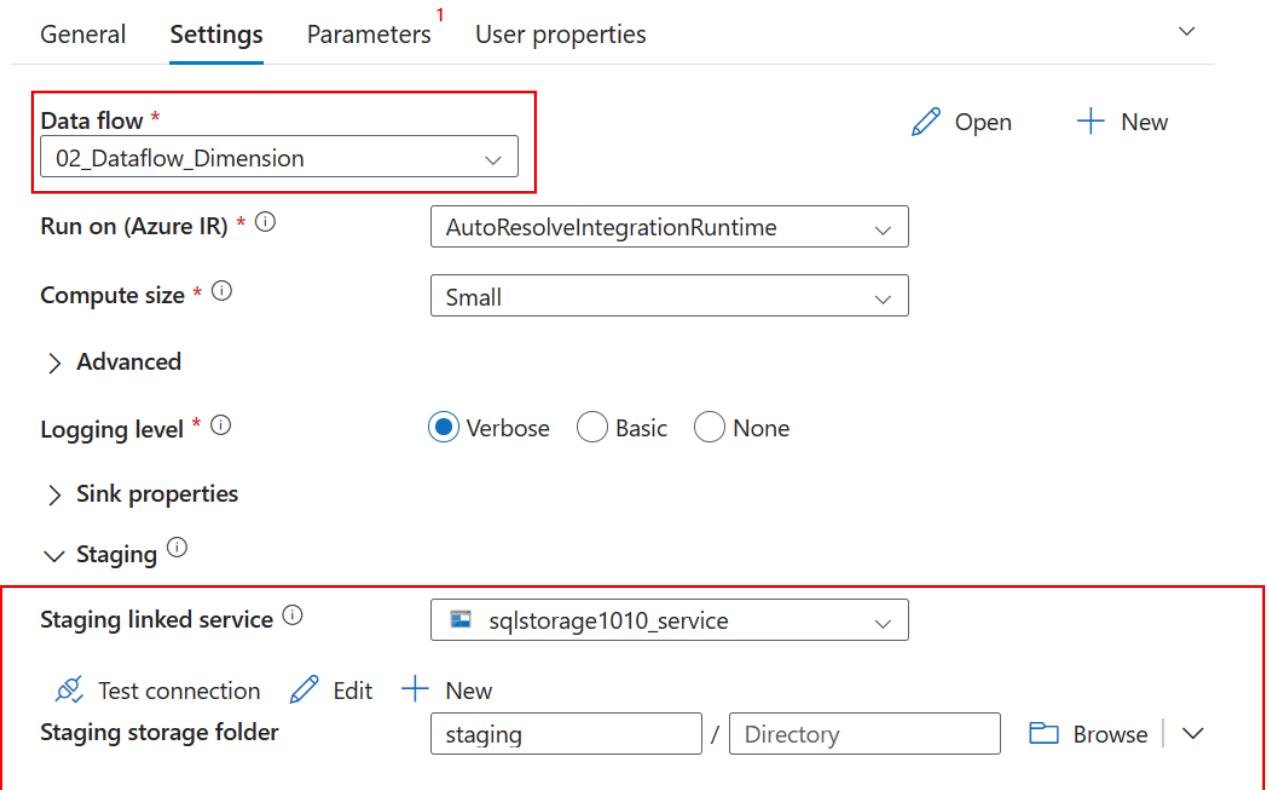
> Sink properties

▽ Staging ⓘ

Staging linked service ⓘ **sqlstorage1010_service**

 Test connection Edit + New

Staging storage folder **staging** / **Directory** Browse | ▽



23. Then you have to click on validate all and then click on publish all.

Publish all

You are about to publish all pending changes to the live environment. [Learn more](#)

Pending changes (1)

NAME	CHANGE	EXISTING
▽ Pipelines		

00 05_mappingdataflow_dimt... (New) -

24. After that you have to click on the trigger and run your pipeline. Below you can see that the pipeline run has been completed.

All pipeline runs > 05_mappingdataflow_dimtable - Activity runs

Rerun Cancel Refresh Update pipeline List Gantt

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User pr
Dataflow_Dimtable	Succeeded	Data flow	4/25/2024, 3:34:20 PM	3m 54s	AutoResolveIntegratio	

25. Now if you go back to SSMS and run the Select statement for both the tables then you get the output accordingly.

```
SELECT * FROM [dbo].[DimProduct]
```

```
SELECT * FROM [dbo].[DimCustomer]
```

150 %

Results Messages

	ProductID	ProductModelID	ProductcategoryID	ProductName	ProductmodelName	ProductCategoryName
1	846	108	37	Taillights - Battery-Powered	Taillight	Lights
2	928	85	41	LL Mountain Tire	LL Mountain Tire	Tires and Tubes
3	855	12	22	Men's Bib-Shorts, S	Men's Bib-Shorts	Bib-Shorts
4	803	105	14	ML Fork	ML Fork	Forks
5	977	31	6	Road-750 Black, 58	Road-750	Road Bikes
6	980	22	5	Mountain-400-W Silver, 38	Mountain-400-W	Mountain Bikes
7	876	118	30	Hitch Rack - 4-Bike	Hitch Rack - 4-Bike	Bike Racks
8	847	109	37	Headlights - Dual-Beam	Headlights - Dual-Beam	Lights
9	934	91	41	Touring Tire	Touring Tire	Tires and Tubes
10	856	12	22	Men's Bib-Shorts, M	Men's Bib-Shorts	Bib-Shorts
11	804	106	14	HL Fork	HL Fork	Forks
12	997	31	6	Road-750 Black, 44	Road-750	Road Bikes
13	981	22	5	Mountain-400-W Silver, 40	Mountain-400-W	Mountain Bikes
14	680	6	18	HL Road Frame - Black, 58	HL Road Frame	Road Frames
15	848	110	37	Headlights - Weatherproof	Headlights - Weatherproof	Lights
16	922	93	41	Road Tire Tube	Road Tire Tube	Tires and Tubes
17	857	12	22	Men's Bib-Shorts, L	Men's Bib-Shorts	Bib-Shorts
18	802	104	14	LL Fork	LL Fork	Forks
19	998	31	6	Road-750 Black, 48	Road-750	Road Bikes
20	982	22	5	Mountain-400-W Silver, 42	Mountain-400-W	Mountain Bikes
21	706	6	18	HL Road Frame - Red, 58	HL Road Frame	Road Frames
22	870	111	32	Water Bottle - 30 oz.	Water Bottle	Bottles and Cages
23	929	86	41	ML Mountain Tire	ML Mountain Tire	Tires and Tubes
24	852	38	28	Women's Tights, S	Women's Tights	Tights

Query executed successfully.

```
SELECT * FROM [dbo].[DimCustomer]
```

150 %

Results Messages

	CustomerID	CompanyName	SalesPerson
1	29822	Family Cycle Store	adventure-works\garrett1
2	29911	Consolidated Messenger	adventure-works\garrett1
3	30004	Some Discount Store	adventure-works\pamela0
4	30105	Convenient Sales and Service	adventure-works\pamela0
5	507	Global Sporting Goods	adventure-works\jillian0
6	608	Golf and Cycle Store	adventure-works\shu0
7	29499	Valley Toy Store	adventure-works\jose1
8	29600	Online Bike Sellers	adventure-works\michael9
9	29705	Rewarding Activities Company	adventure-works\garrett1
10	29797	Metro Manufacturing	adventure-works\jose1
11	1	A Bike Store	adventure-works\pamela0
12	97	Mountain Bike Center	adventure-works\shu0
13	198	Field Trip Inc	adventure-works\michael9
14	293	Fashionable Bikes and Accessories	adventure-works\shu0
15	388	Road Way Sales and Rental	adventure-works\jose1
16	487	Bike Experts	adventure-works\pamela0
17	29824	Superb Sales and Repair	adventure-works\garrett1
18	29913	Field Trip Store	adventure-works\linda3
19	30005	Variety Cycling	adventure-works\jae0
20	30106	Solid Bike Parts	adventure-works\jillian0
21	508	Designer Department Stores	adventure-works\jillian0
22	609	Good Toys	adventure-works\linda3
23	29502	Major Sport Suppliers	adventure-works\david8
24	29601	Custom Accessories Company	adventure-works\garrett1

Query executed successfully.