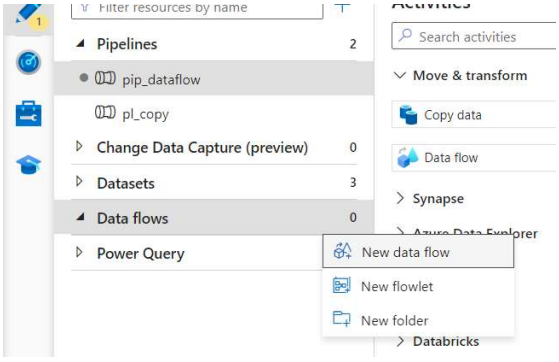
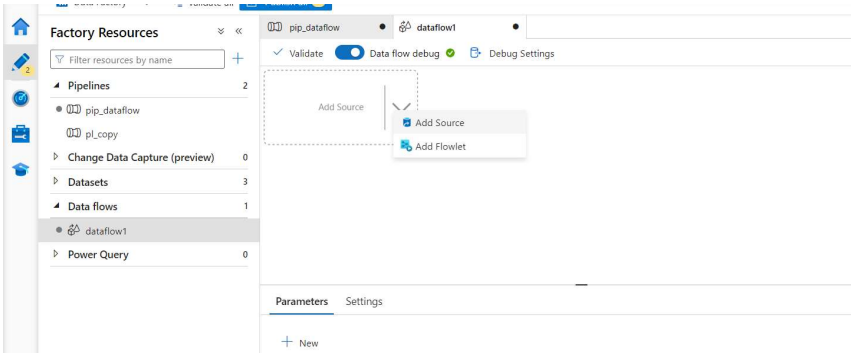
Context

Mapping Data Flow

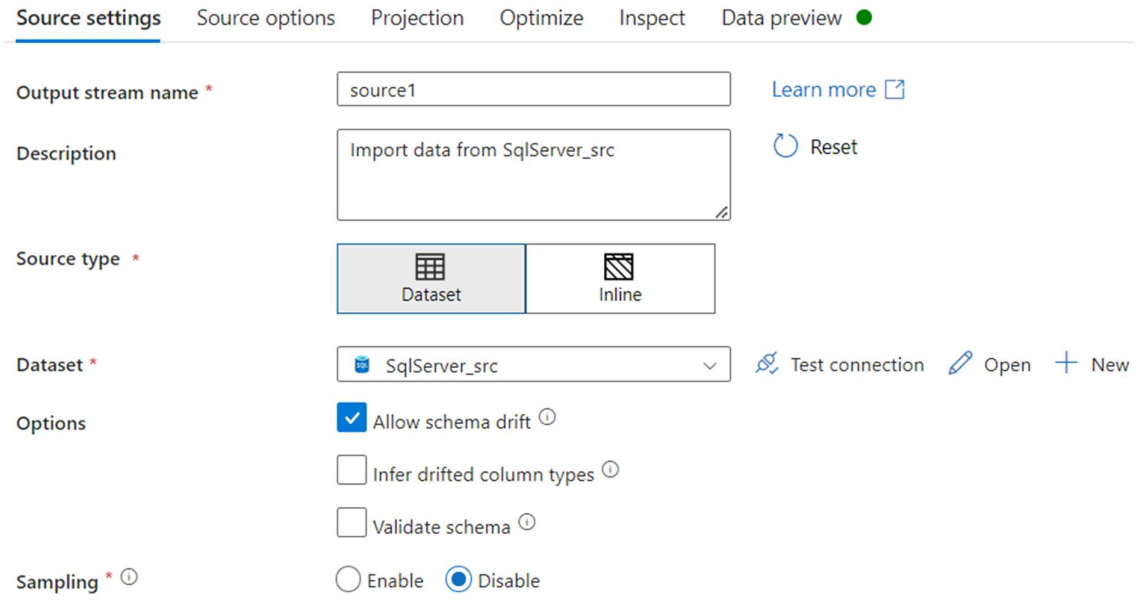
1. In this example we are going to create a Data Flow. To load the SQL Server table into Cloud Storage.
2. Go to Data Factory, go to Author, right-click on Data Flow, and click on New data flow.



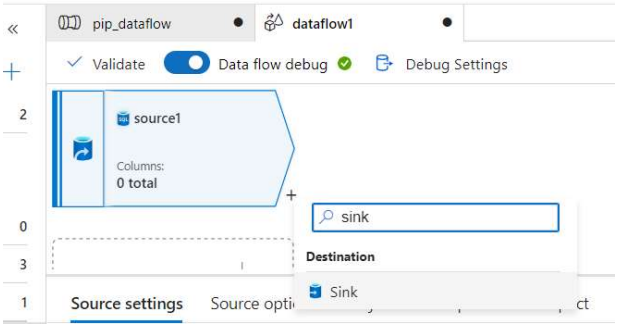
3. Switch on the data flow debug. Click on the drop-down and click on Add Source.



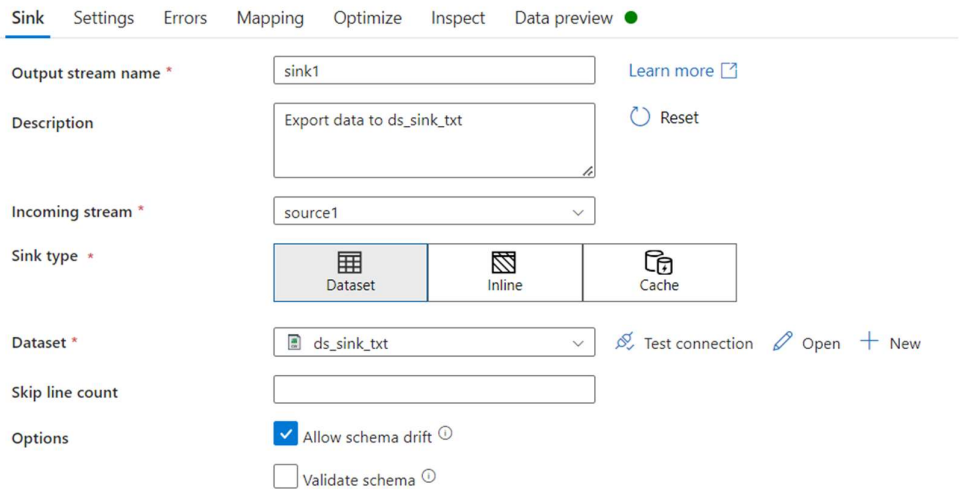
4. In the Source Setting give the Name and select the SqlServer dataset that we created in the previous session and check the below properties as shown below.



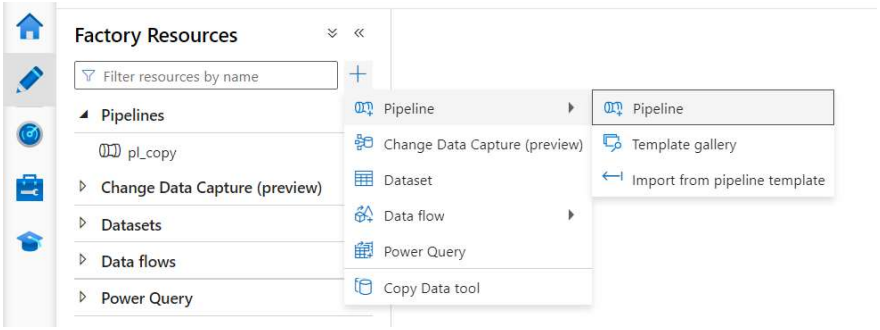
5. Now click on the Plus symbol and click on Sink as shown below.



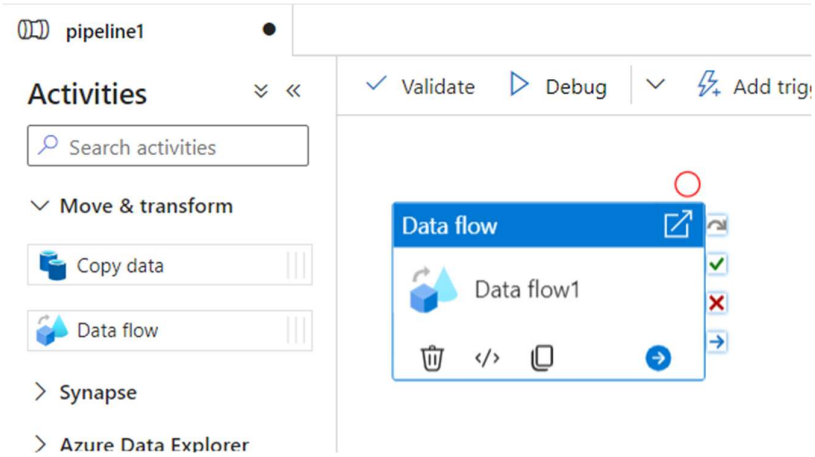
6. Now in the sink, give the name and select the sink dataset that we created in the previous session and check the below properties as shown below.



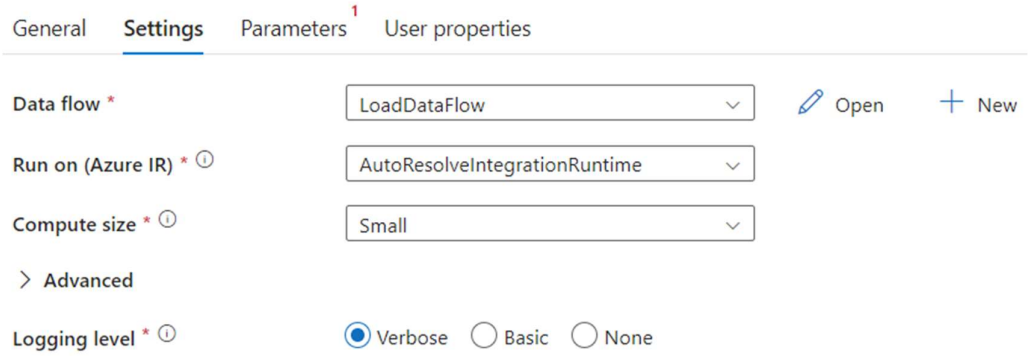
7. Create a Pipeline



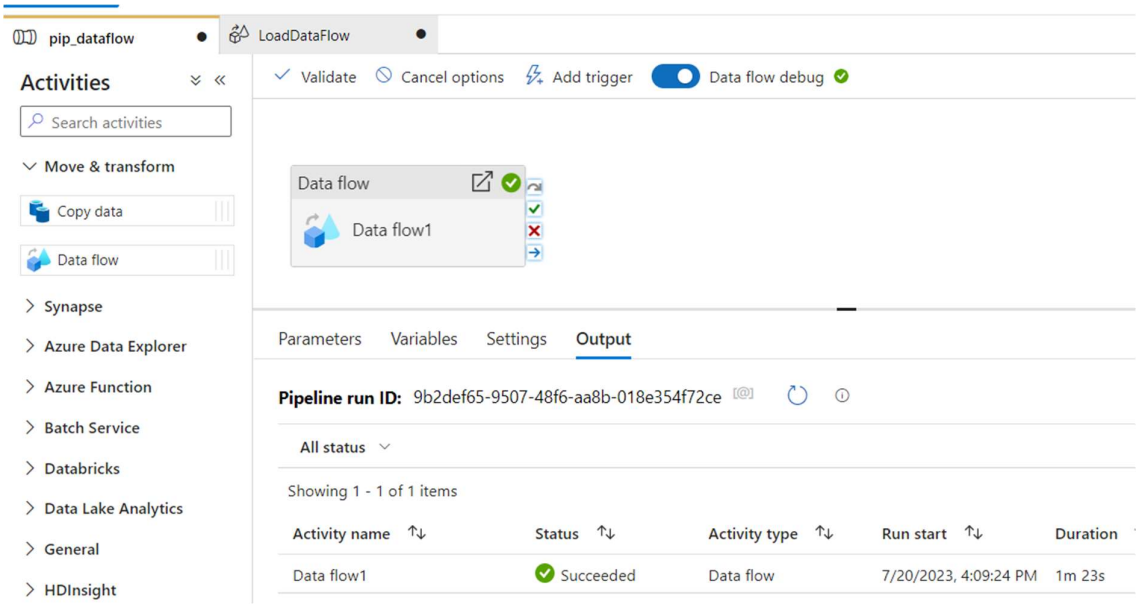
8. Next Under Activity Drag and drop the Data Flow as shown below.



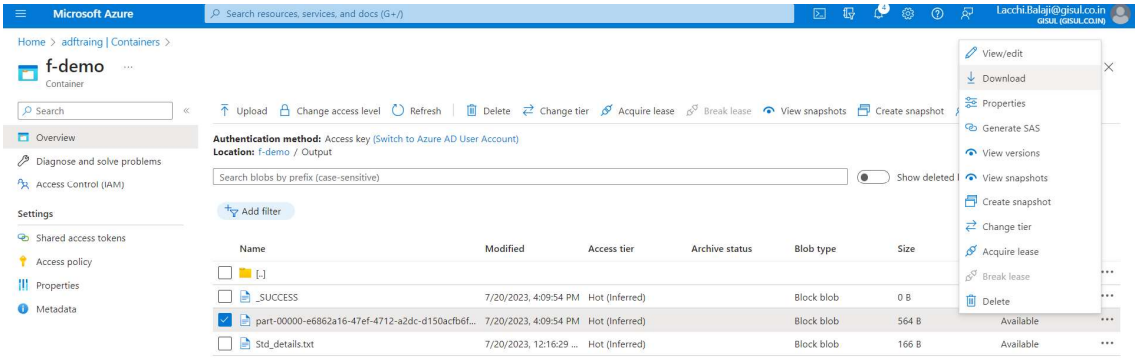
9. Under settings select the Data flow that we created.



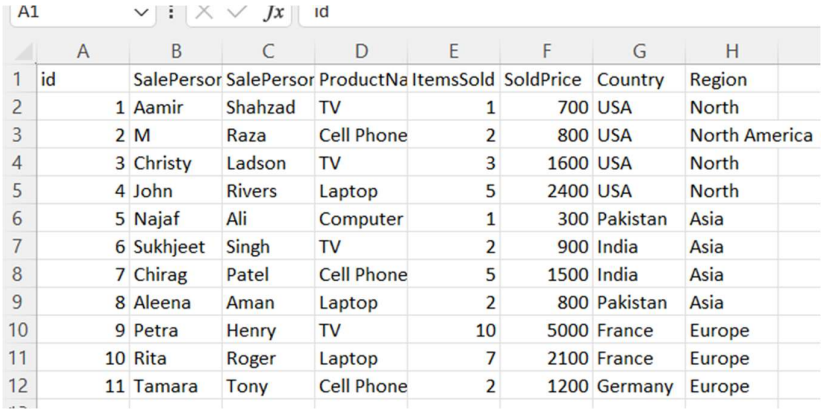
10. Next validate the Pipeline and click on Debug.
11. Our pipeline is Successfully executed.



12. Now go to Storage account and the container folder. We have one output folder and under that, we have a new file download.



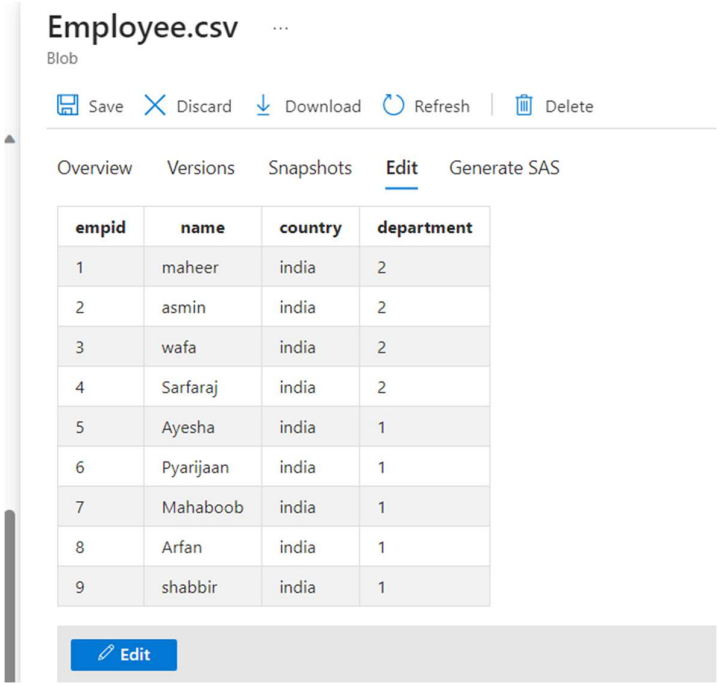
13. Open it you will see our TotalSale table.



	A	B	C	D	E	F	G	H
1	id	SalePersor	SalePersor	ProductNa	ItemsSold	SoldPrice	Country	Region
2	1	Aamir	Shahzad	TV	1	700	USA	North
3	2	M	Raza	Cell Phone	2	800	USA	North America
4	3	Christy	Ladson	TV	3	1600	USA	North
5	4	John	Rivers	Laptop	5	2400	USA	North
6	5	Najaf	Ali	Computer	1	300	Pakistan	Asia
7	6	Sukhjeet	Singh	TV	2	900	India	Asia
8	7	Chirag	Patel	Cell Phone	5	1500	India	Asia
9	8	Aleena	Aman	Laptop	2	800	Pakistan	Asia
10	9	Petra	Henry	TV	10	5000	France	Europe
11	10	Rita	Roger	Laptop	7	2100	France	Europe
12	11	Tamara	Tony	Cell Phone	2	1200	Germany	Europe

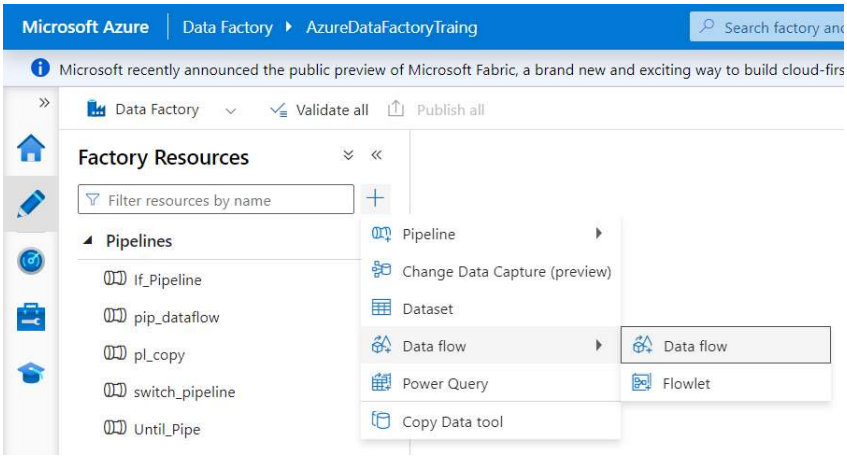
Select Transformation

1. In this example we are going to use the below file. I have uploaded the file for the local system to my source connection.

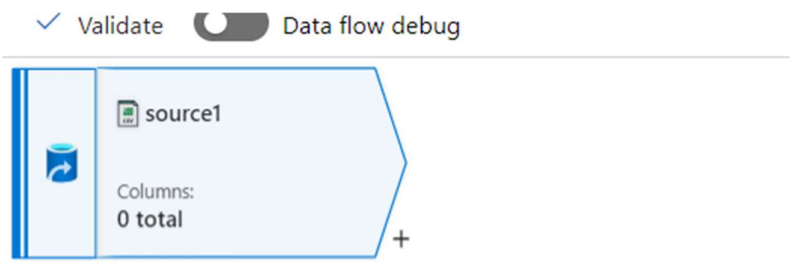


empid	name	country	department
1	maheer	india	2
2	asmin	india	2
3	wafa	india	2
4	Sarfaraj	india	2
5	Ayesha	india	1
6	Pyarijaan	india	1
7	Mahaboob	india	1
8	Arfan	india	1
9	shabbir	india	1

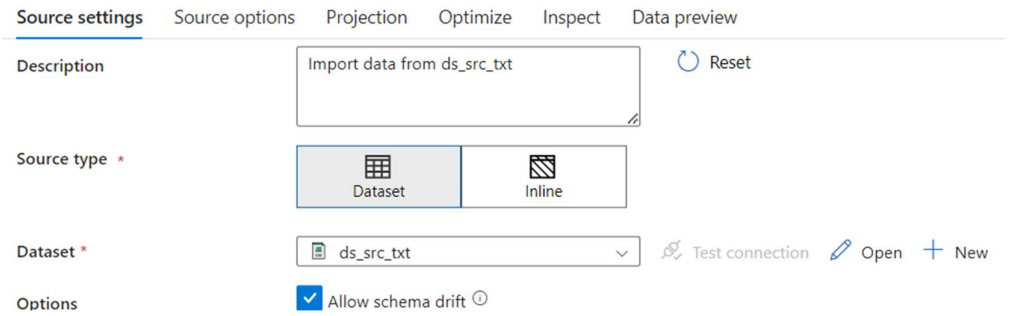
2. Create a Data flow.



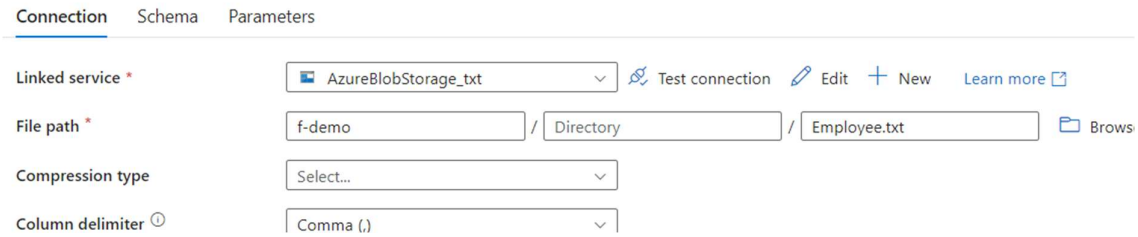
3. Add Data source.



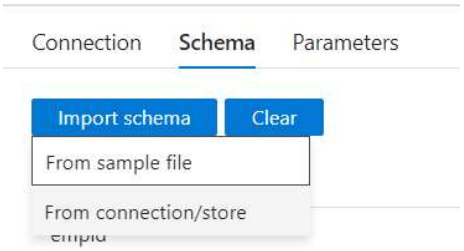
4. Under source select the source connection and click on open.



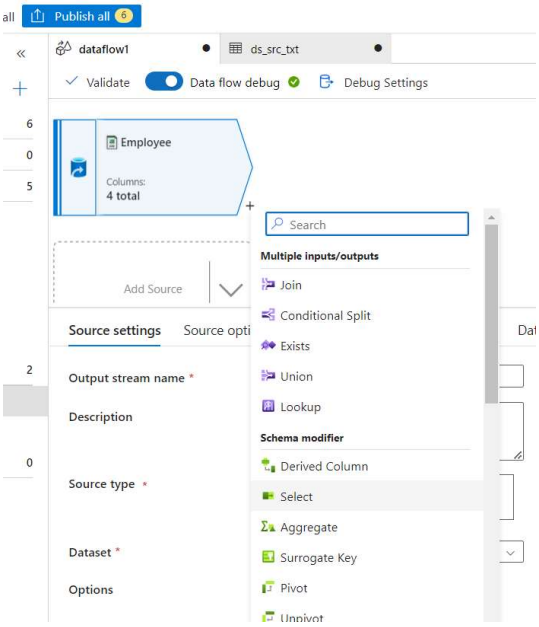
5. Select the Employee file as shown below.



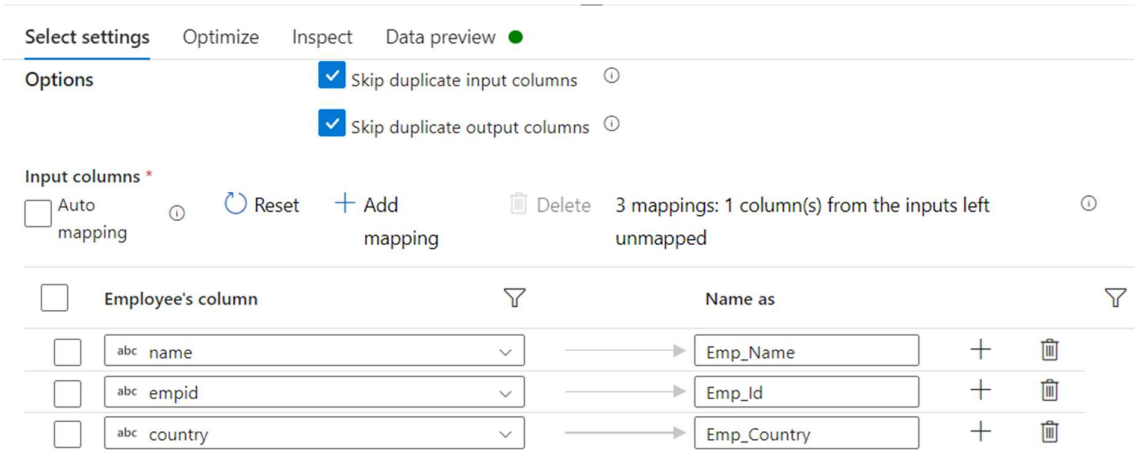
6. Now under schema click on the Import schema and click on From connection/store.



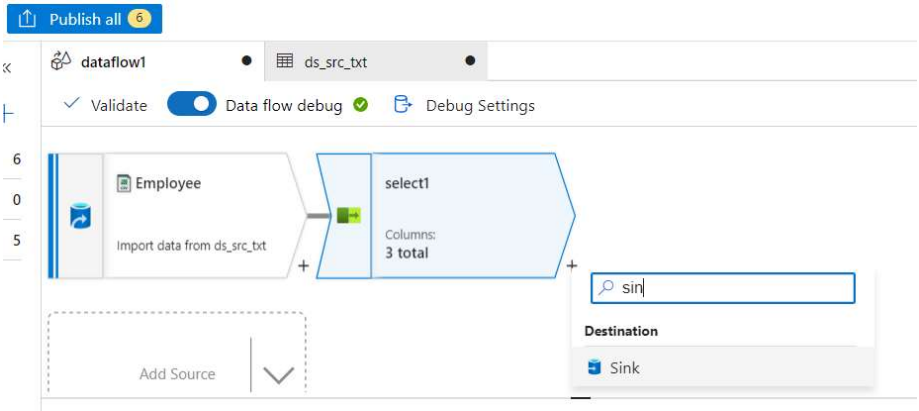
7. Now go to Data flow click on the plus symbol and click on the select option as shown below.



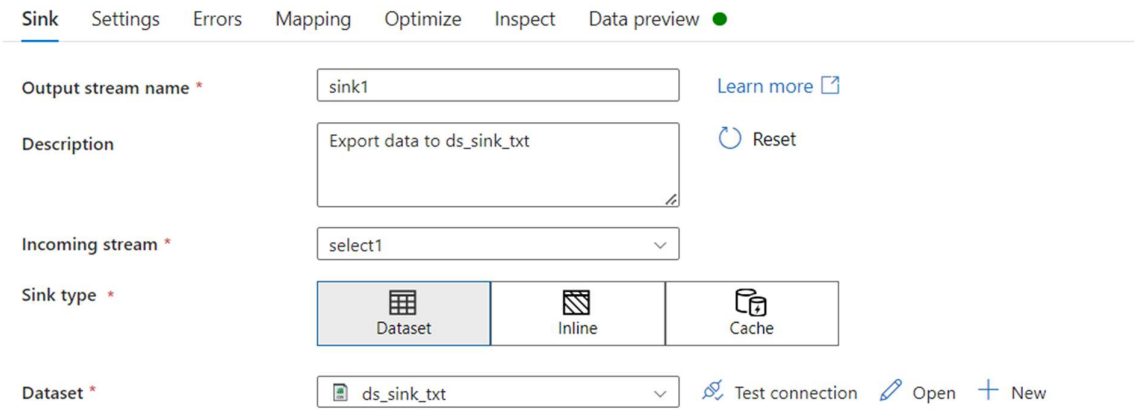
8. Under the setting we have 4 columns but I deleted the department column by clicking on the delete icon button.
9. And I rearrange the order but click on the drop-down from the Employee's column and change the column names also as shown below.



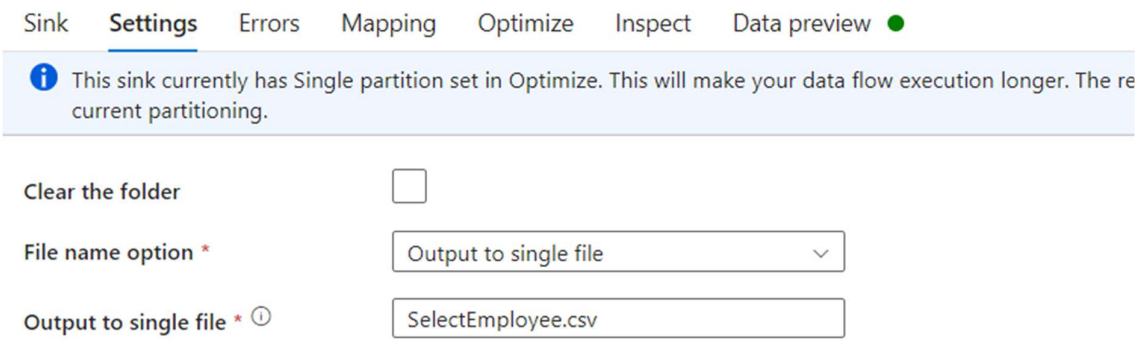
10. Now click on the select plus symbol and click on Sink as shown below.



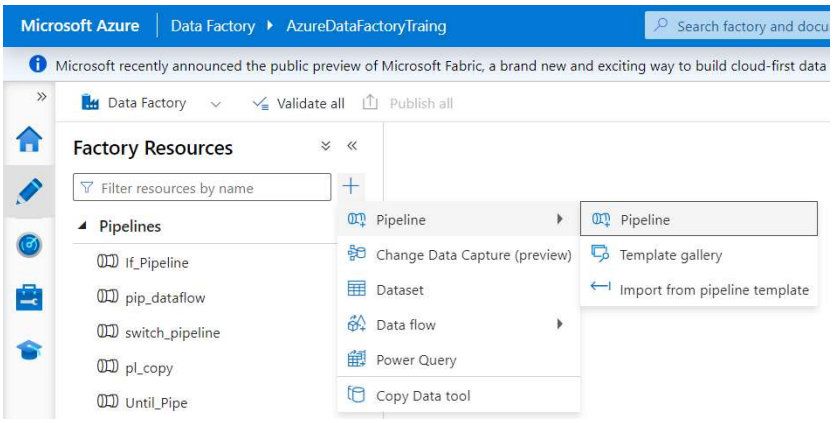
11. Under the sink select the destination dataset.



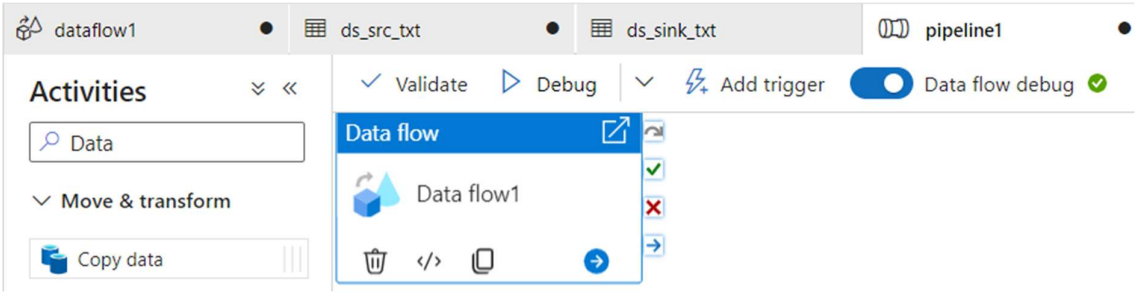
12. Under settings select the File name option and give the file name as shown below.



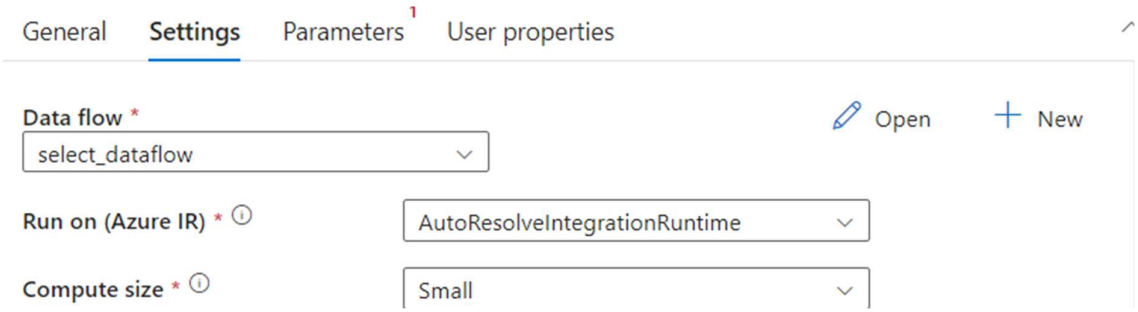
13. Create a pipeline.



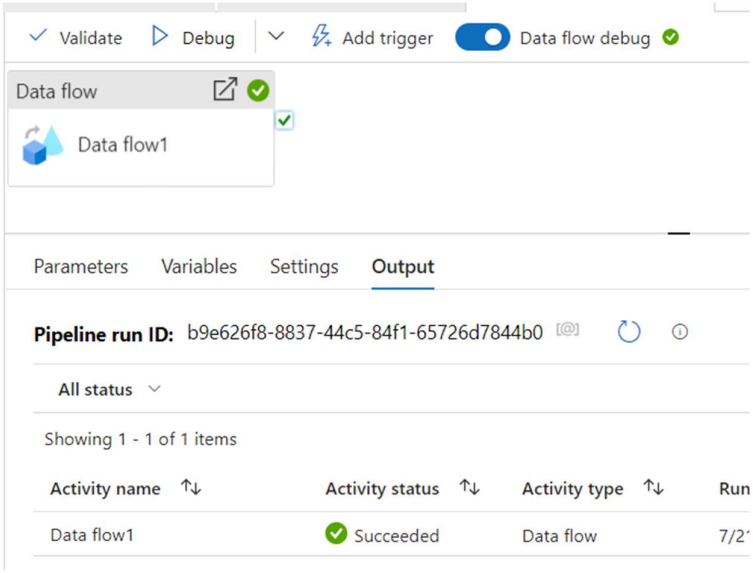
14. Drag and drop the Data flow activity.



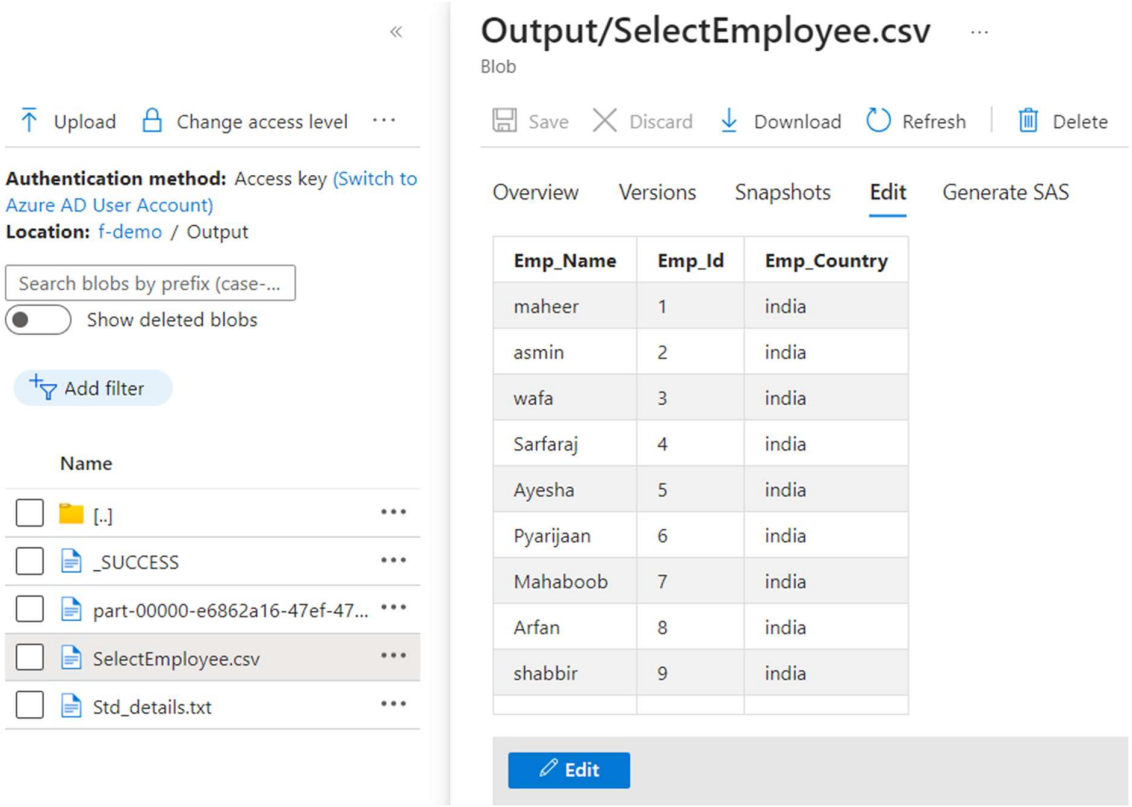
15. Under the settings select the dataflow.



16. Next validate the pipeline and click on Debug.
17. Here our pipeline is executed successfully.

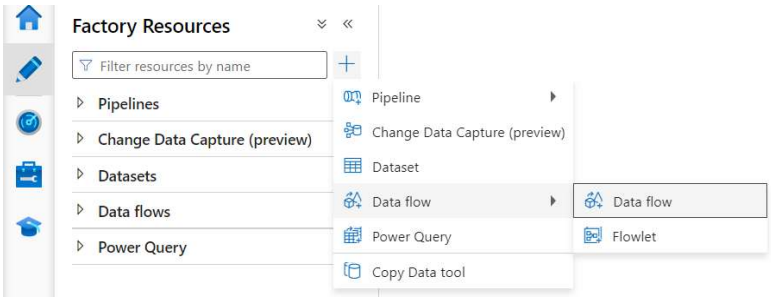


18. Now go to a destination and check the file and the data.

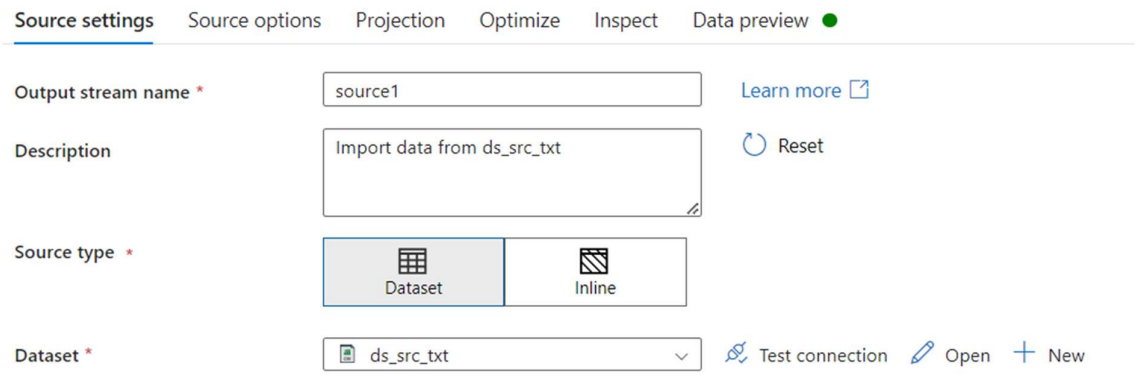


Conditional Split Transformation

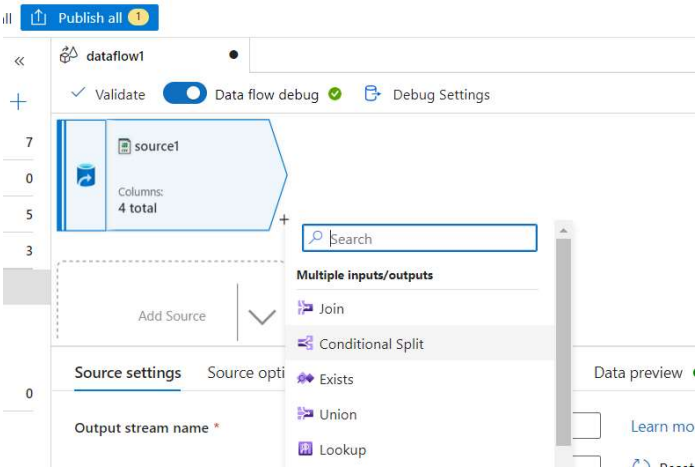
1. For this example, also we are using the same employee table.
2. Go to Azure Data Factory, and create a data flow.



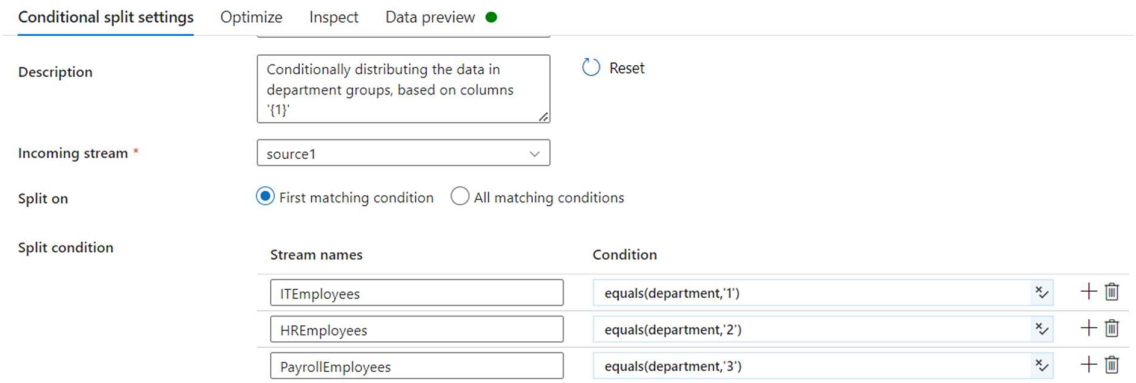
3. Select the source and employee table as shown below.



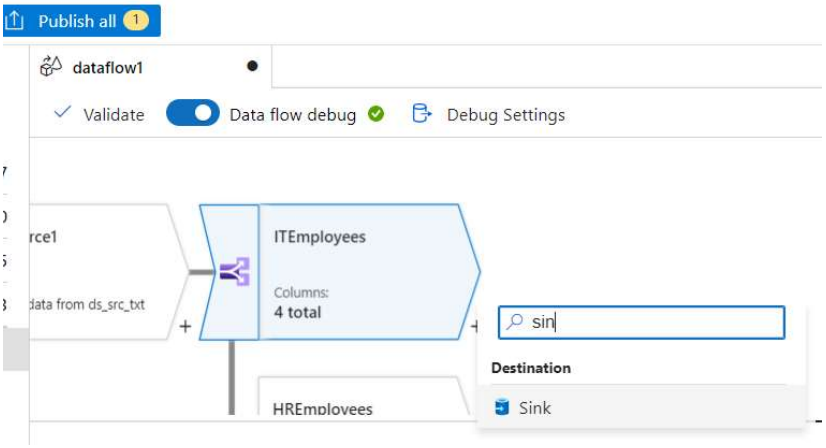
4. Click on the source plus symbol and click on a conditional split as shown below.



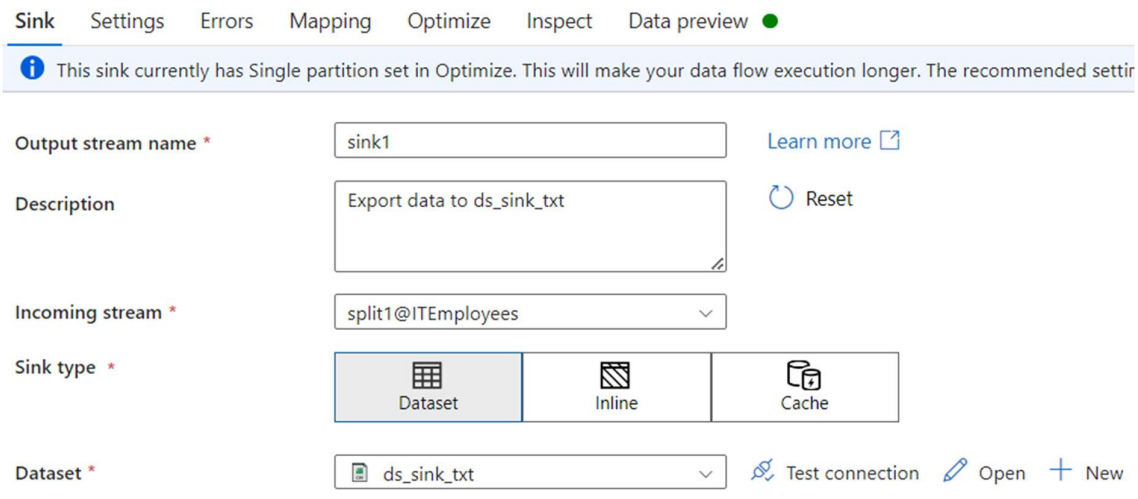
5. Under the split condition add the three conditions as shown below.
6. Delete the default condition.



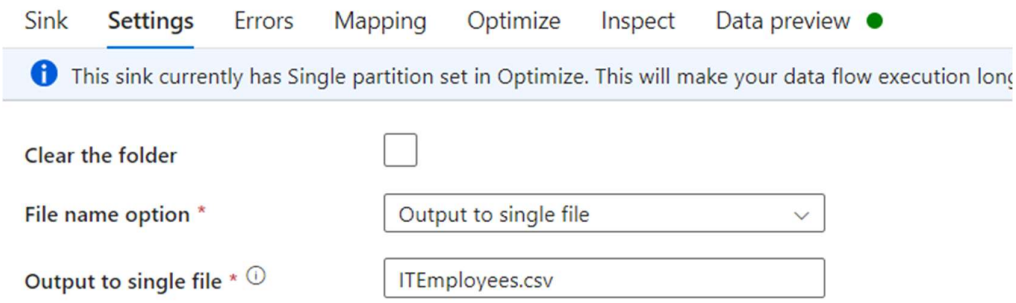
7. Click on the IT Employee plus symbol and click on the sink.



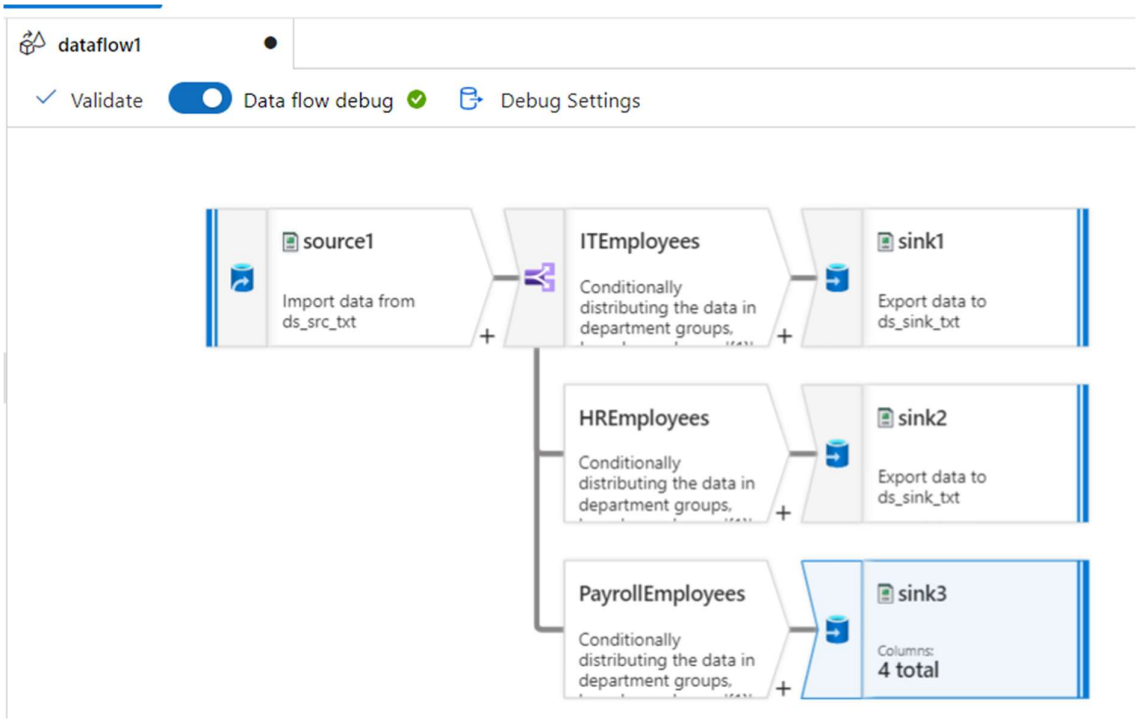
8. Under the sink select the destination dataset.



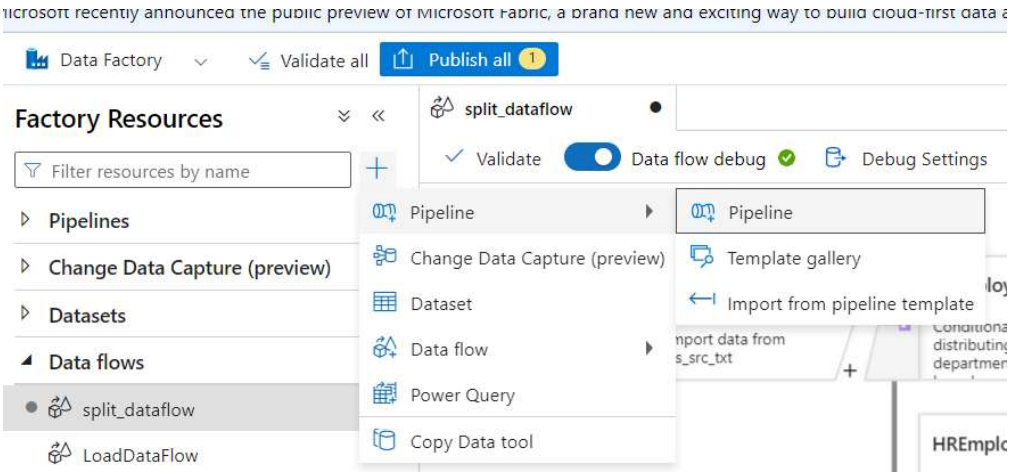
9. Under setting select the File option and file name as shown below.



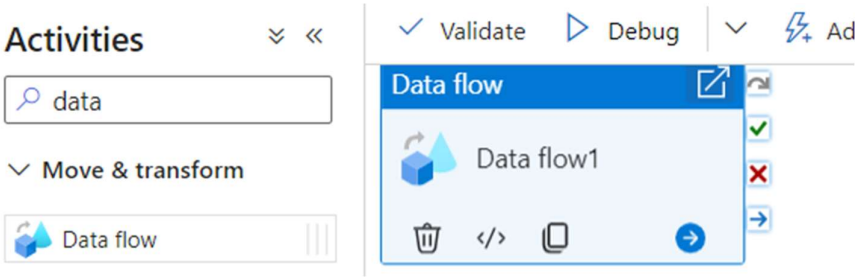
10. Similarly add the sink for other conditions also as shown below.



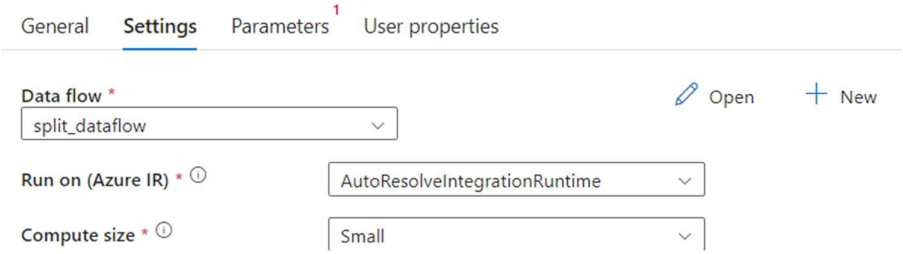
11. Create a pipeline.



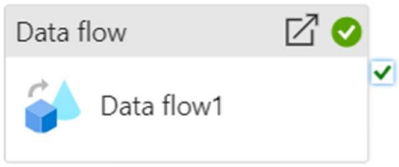
12. Drag and drop the Data Flow activity.

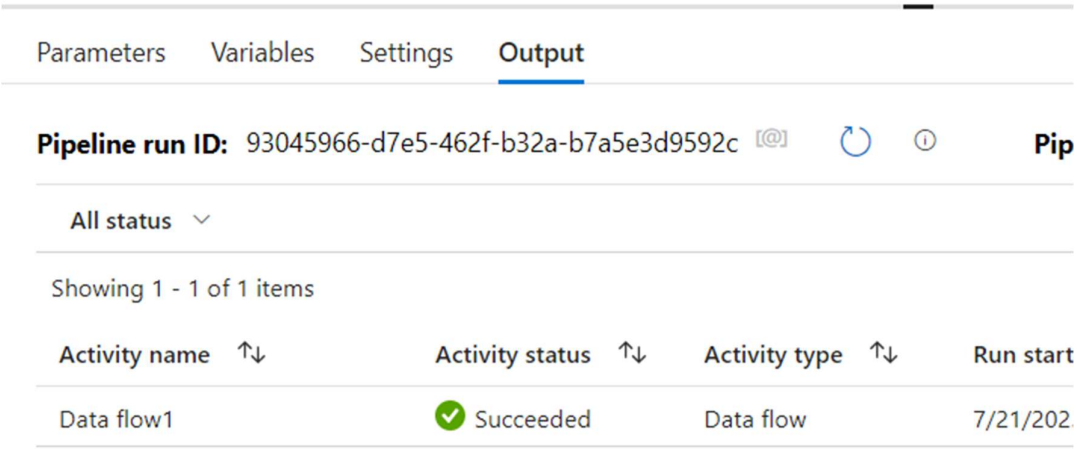


13. Under the setting select the data flow.

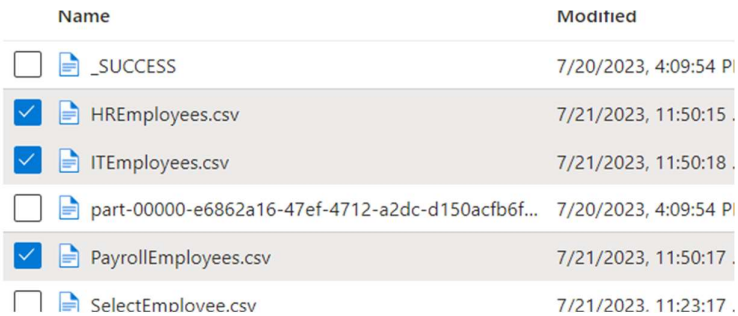


14. Now validate the pipeline and click on debug.
15. Our pipeline was executed successfully.



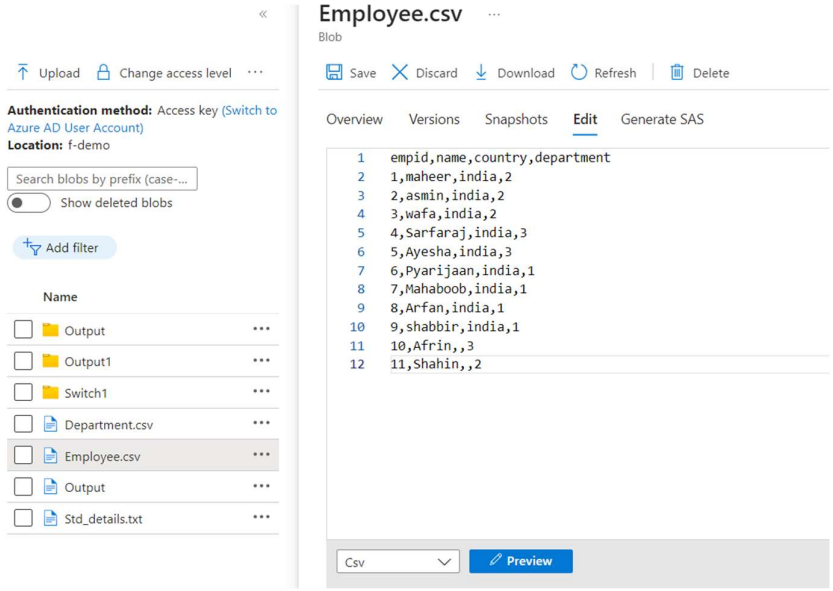


16. Now go to a destination and check whether the files are created or not.

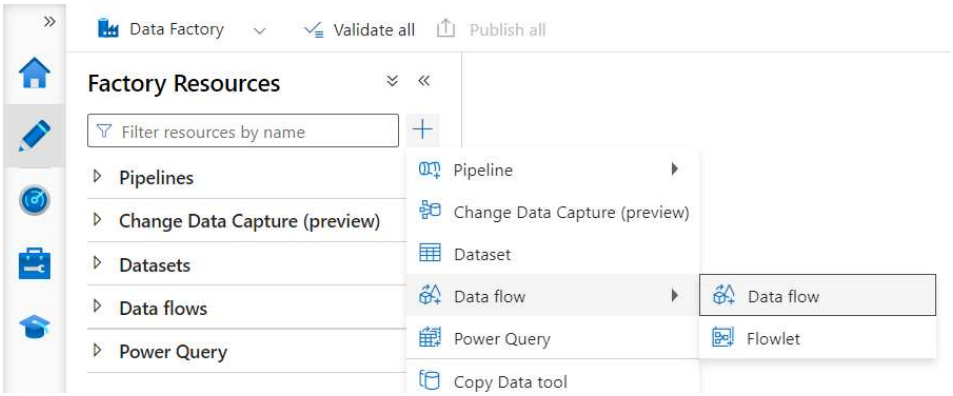


Derived Column Transformation

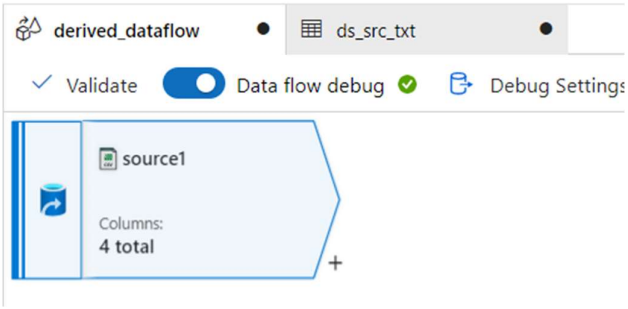
1. In this example also we are using the same Employee file but here we are editing this file.
2. Here I have added two more columns and I didn't give the country name for the new column.
3. Change on your side also and click on save.



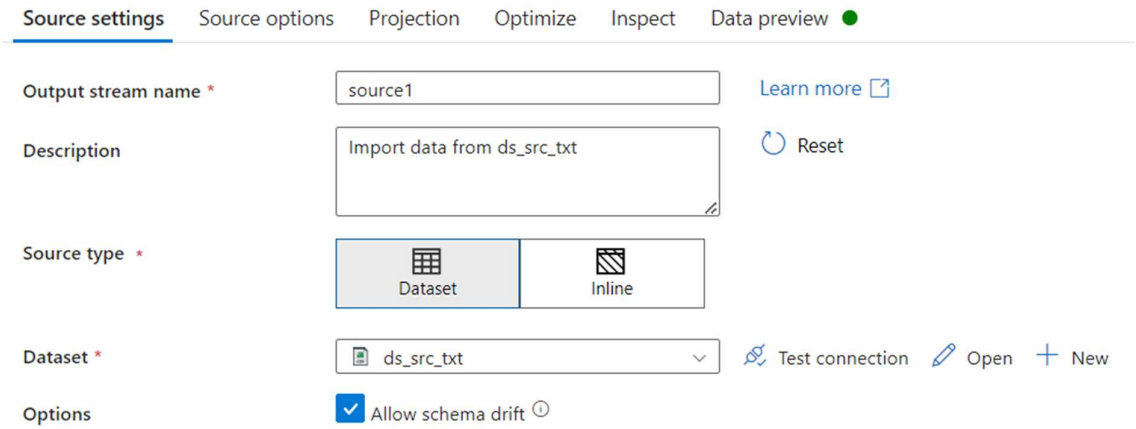
4. Create a Data flow.



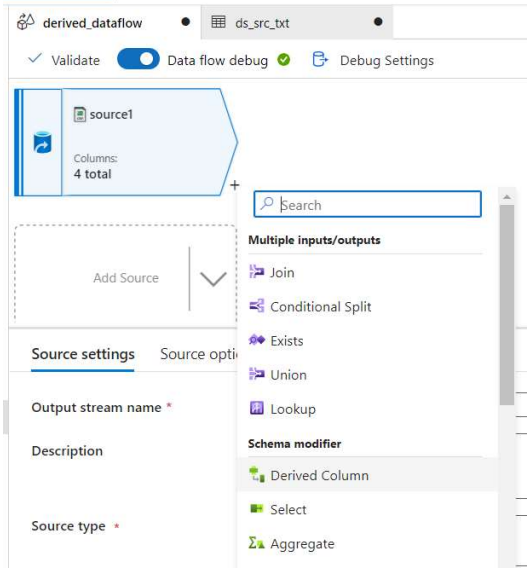
5. Get the Source.



6. Select the dataset and employee file as shown below.



7. Click on the Source plus symbol and click on the Derived column as shown below.

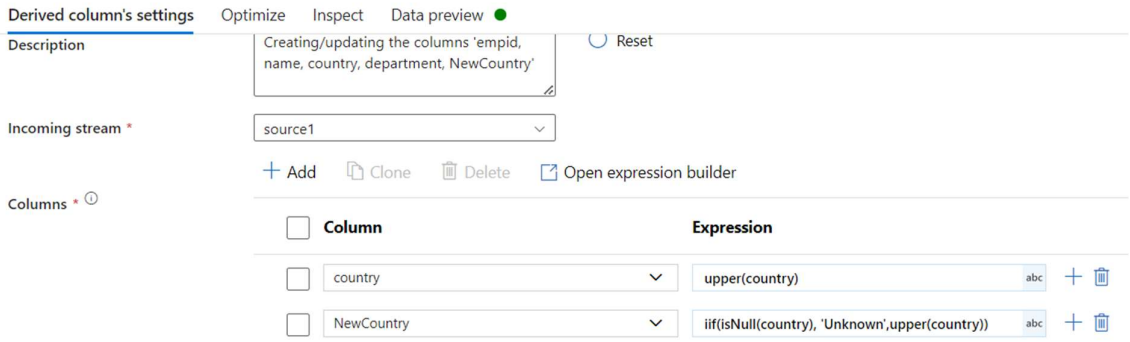


8. Under Columns add two columns and give the below expressions.
9. For the first column just click on the drop-down and select the country that gives the below expression

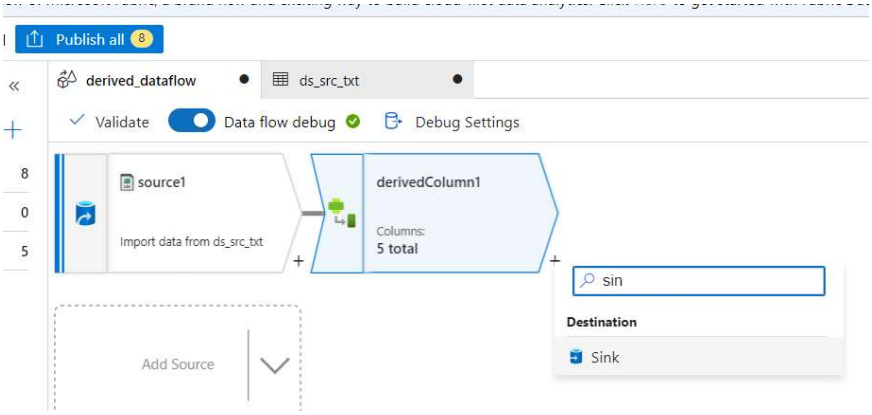
Exp: upper(country)

10. For the second column type a NewCountry and give the below expression.

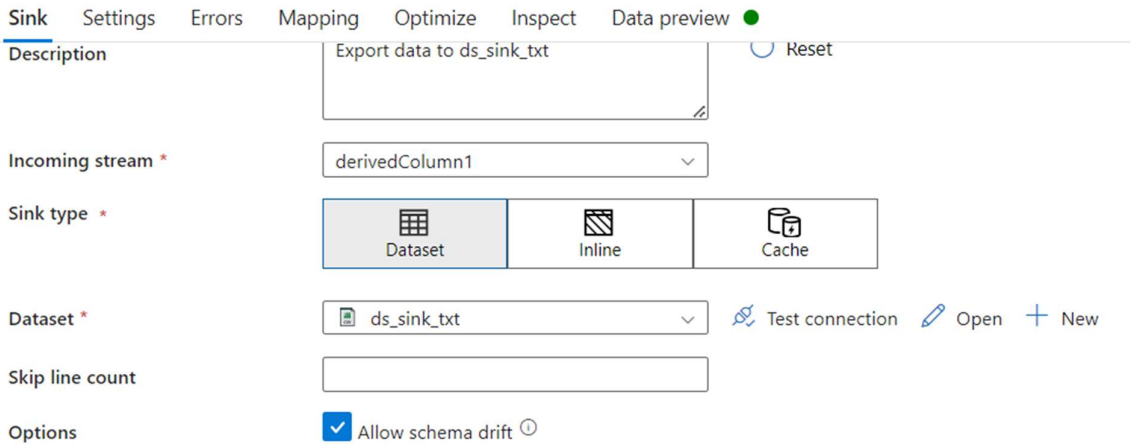
Exp: iif(isNull(country), 'Unknown',upper(country))



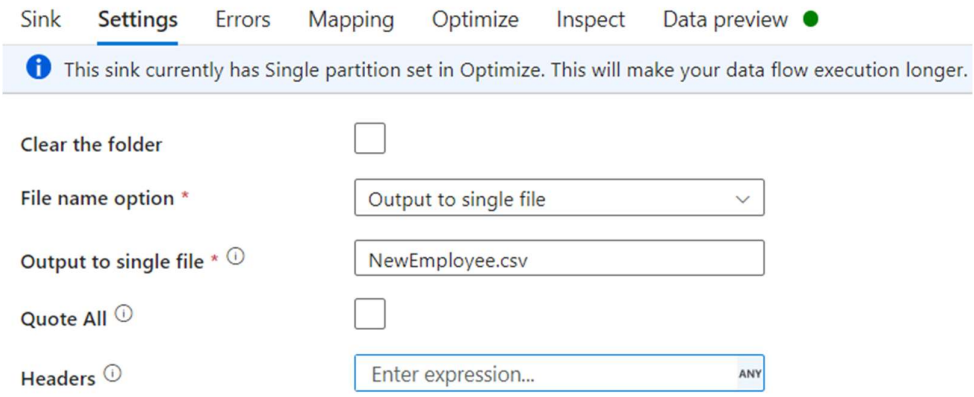
11. Now click on the plus symbol and add a sink as shown below.



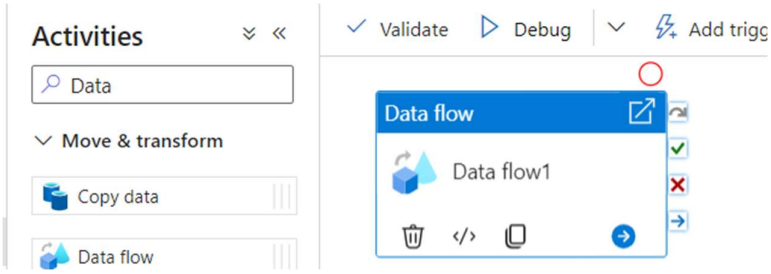
12. Under the sink, select the destination dataset.



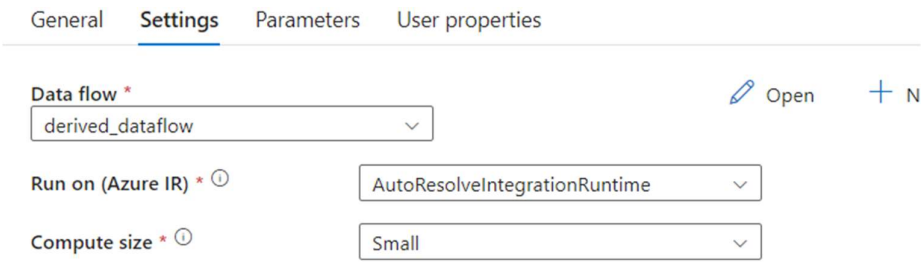
13. Under setting select the file option and file name as shown below.



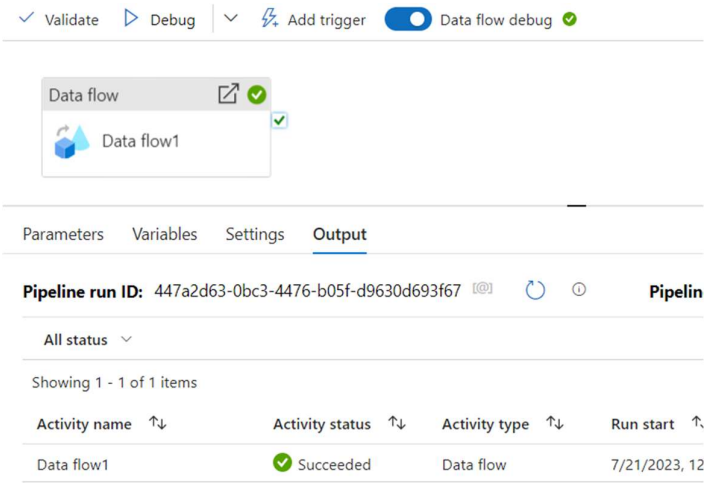
14. Create a Pipeline and drag and drop the data flow activity.



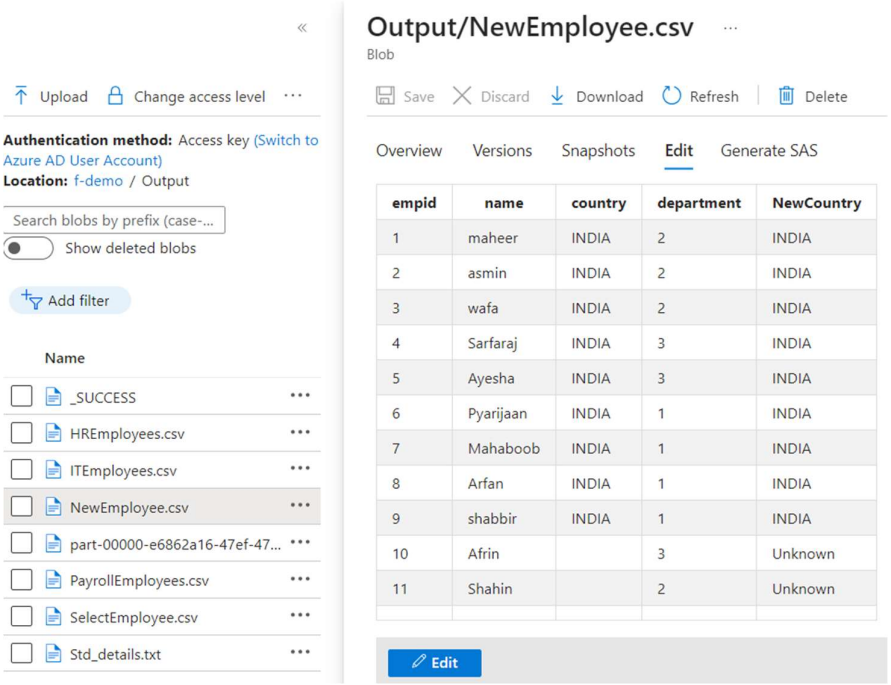
15. Select the Data flow.



16. Now validate the data flow and click on debug.
17. Here our pipeline was executed successfully.



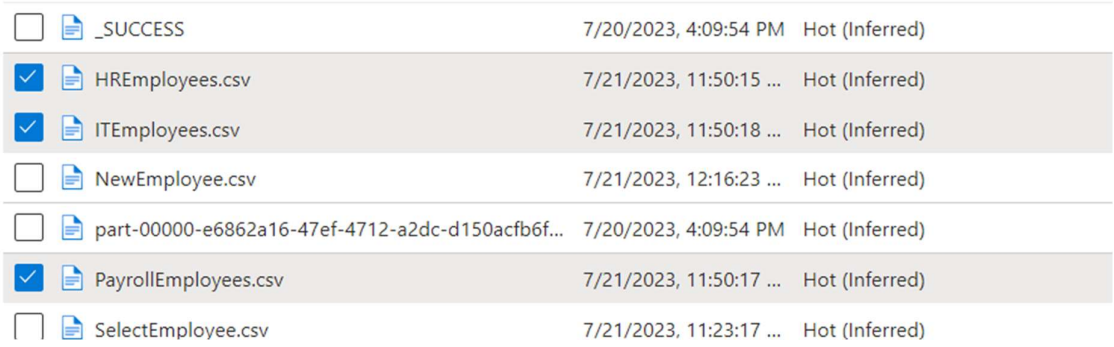
18. Now go to a destination and check the data.



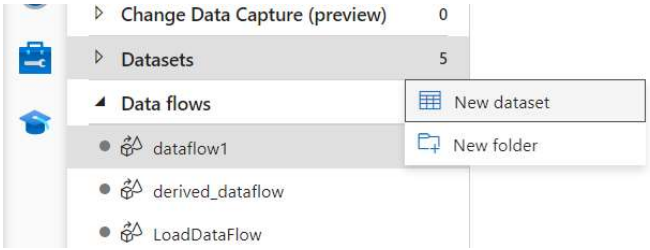
empid	name	country	department	NewCountry
1	maheer	INDIA	2	INDIA
2	asmin	INDIA	2	INDIA
3	wafa	INDIA	2	INDIA
4	Sarfaraj	INDIA	3	INDIA
5	Ayesha	INDIA	3	INDIA
6	Pyarijaan	INDIA	1	INDIA
7	Mahaboob	INDIA	1	INDIA
8	Arfan	INDIA	1	INDIA
9	shabbir	INDIA	1	INDIA
10	Afrin		3	Unknown
11	Shahin		2	Unknown

Union Transformation

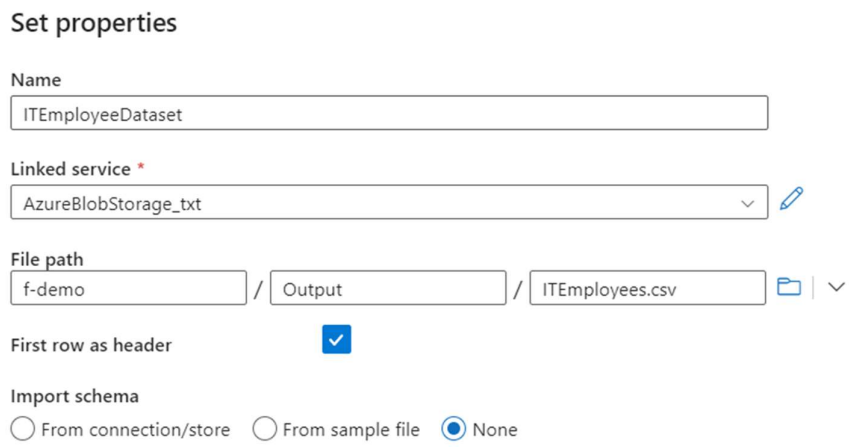
1. In this example we are using the Output of conditional split transformation.
2. These outputs are from Conditional split transformation.
3. Now In our example there are the source data we are going to use a union to combine and stored in one file.



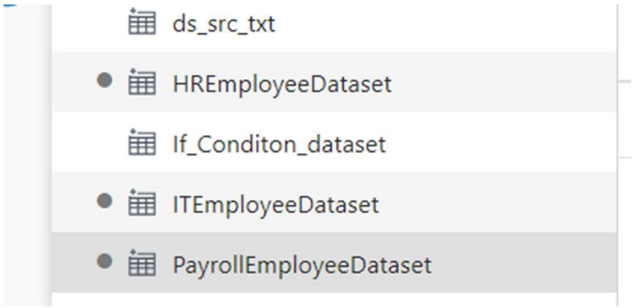
4. We need to create three datasets for these three files.



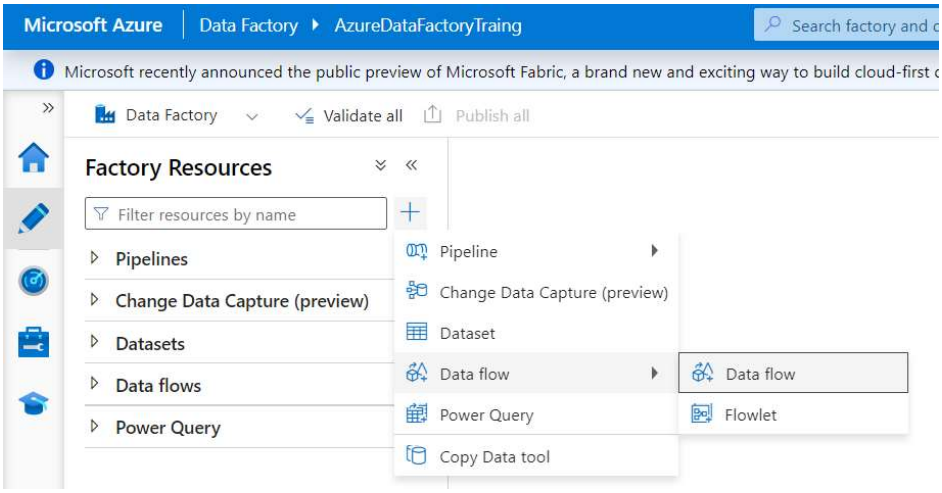
5. Select the Azure Blob Storage and click on continue then select the delimited text and click on continue.
6. Then give the name, select the linked service, and select the file as shown below.



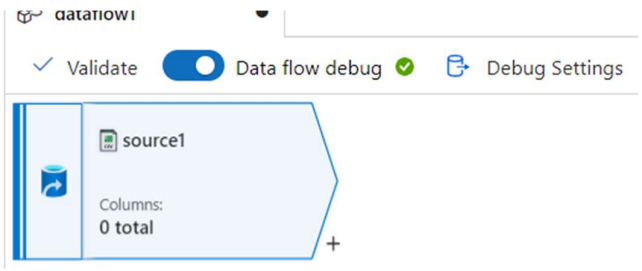
7. Similarly create datasets for the other two files also.



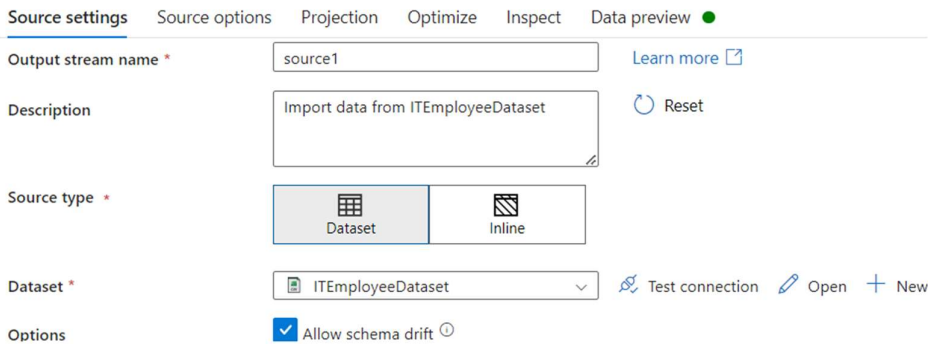
8. Create a Data flow.



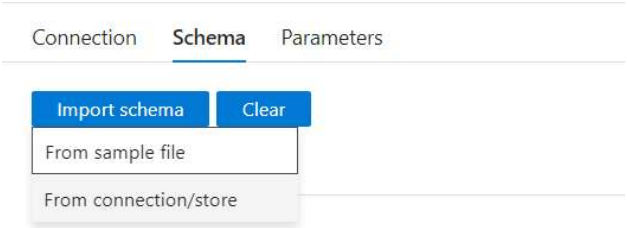
9. Add a Source.



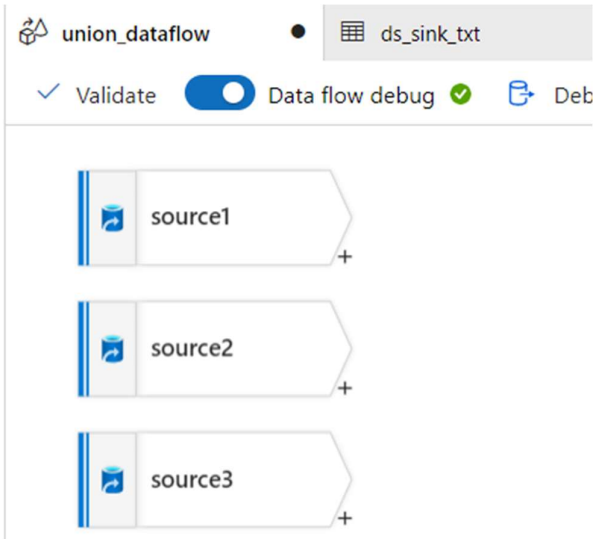
10. Under source select the destination dataset because our tables are present in destination and click on open.



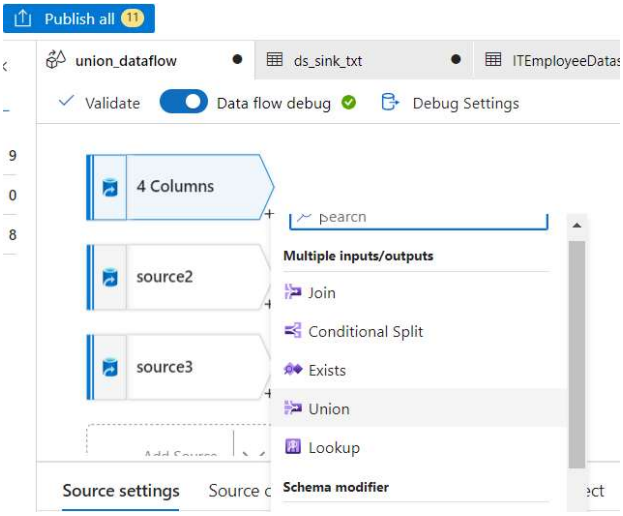
11. Under schema click on Import schema and click on From connection/store.



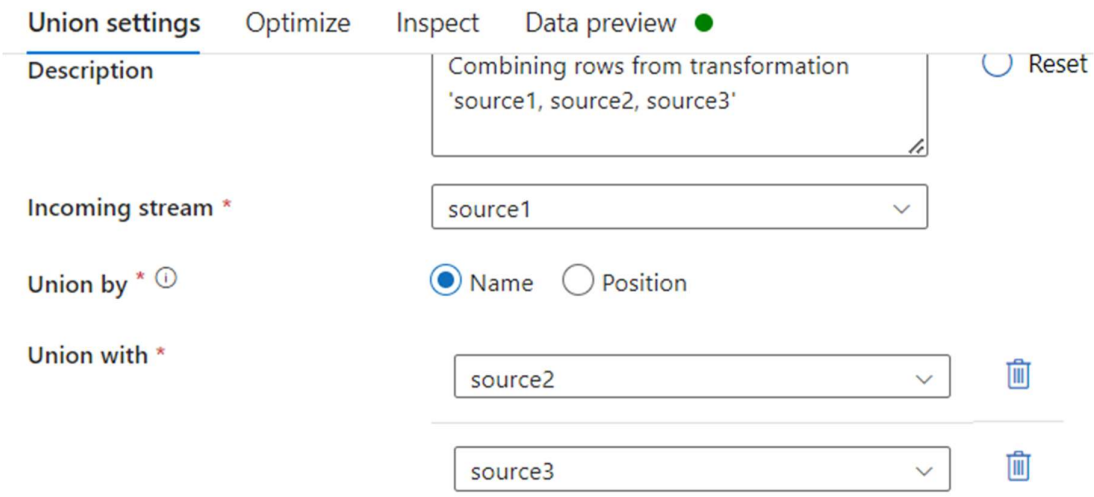
12. Go to Data Flow and add two more data sources. Similarly configure the source file with the datasets we created for each department like the above process.



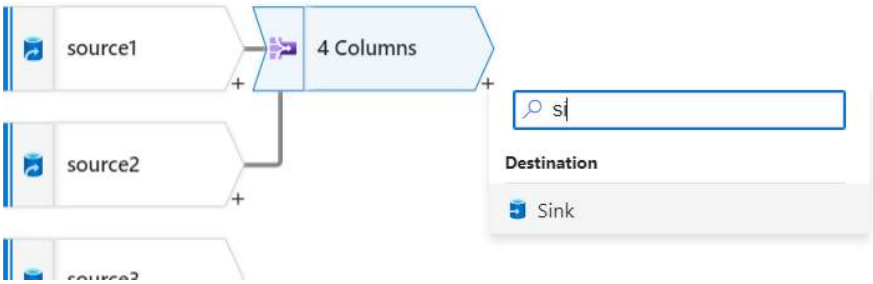
13. Now click on the plus symbol and click on the Union as shown below.



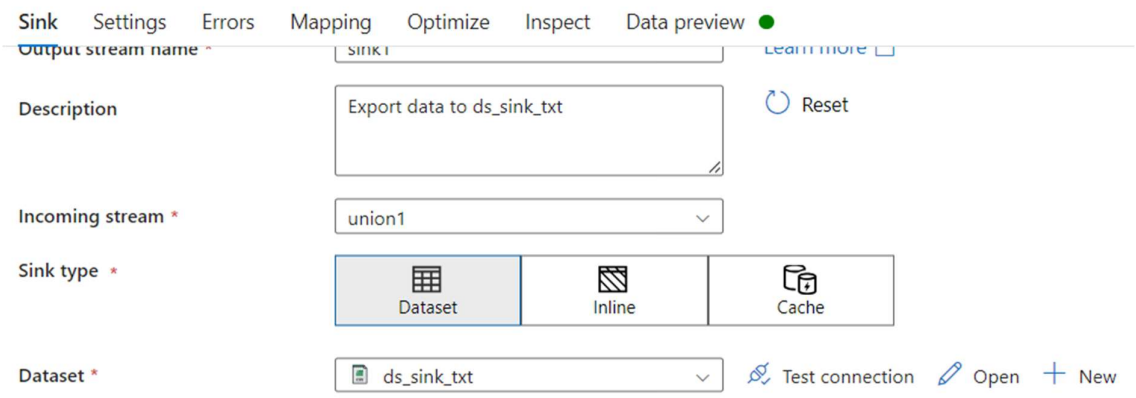
14. Under the union setting add select source 2 and source 3 as shown below.



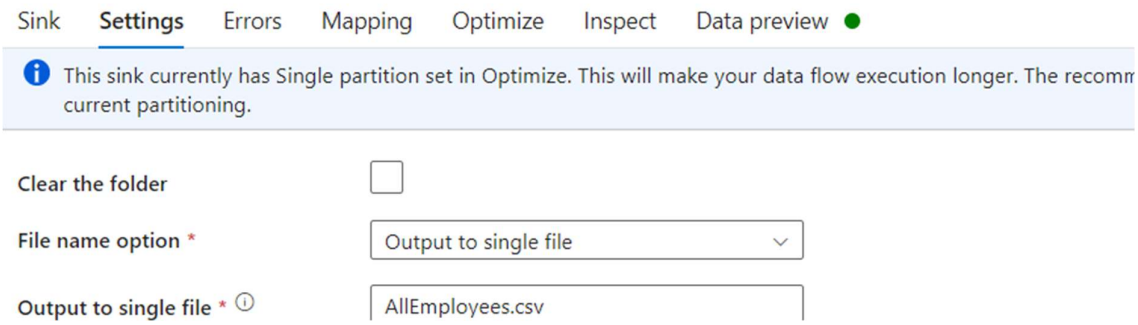
15. Now click on the plus symbol and click on the sink.



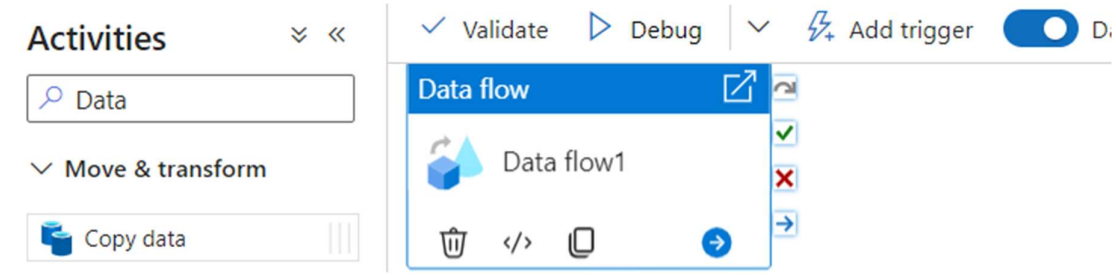
16. Under the sink select the destination dataset.



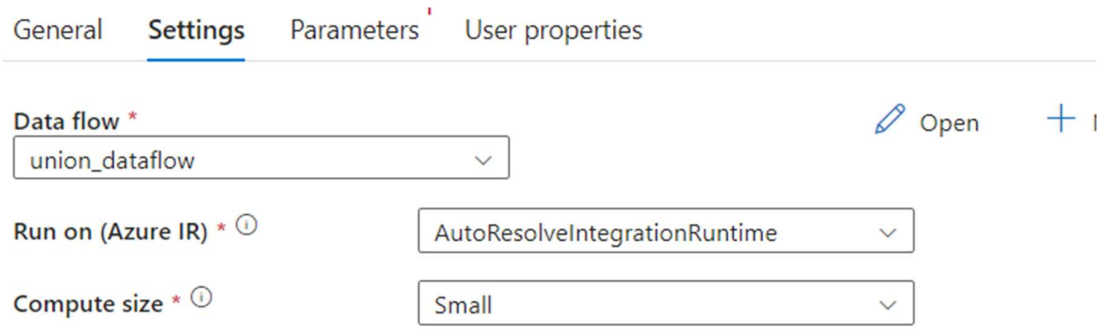
17. Under setting select the file option and file name as shown below.



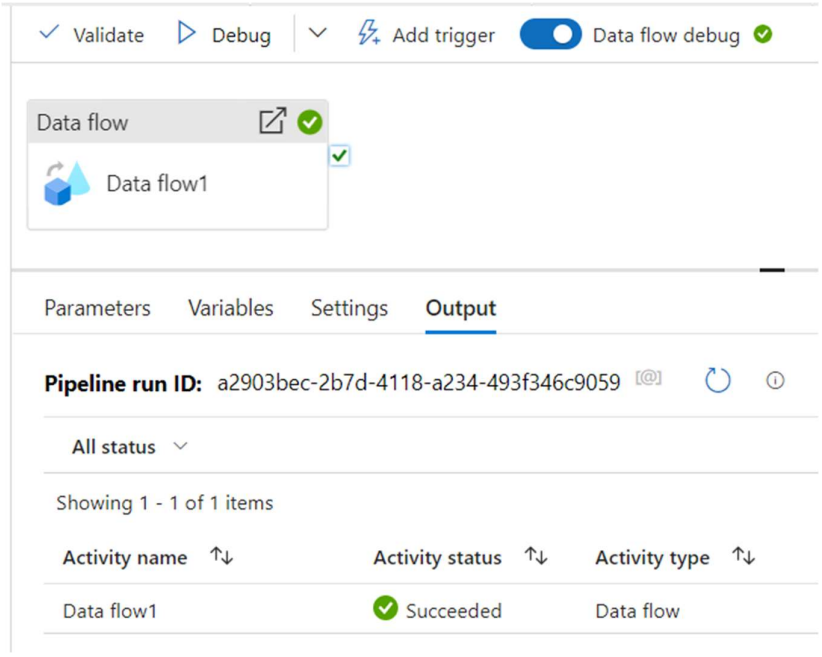
18. Create a pipeline then drag and drop the data flow activity.



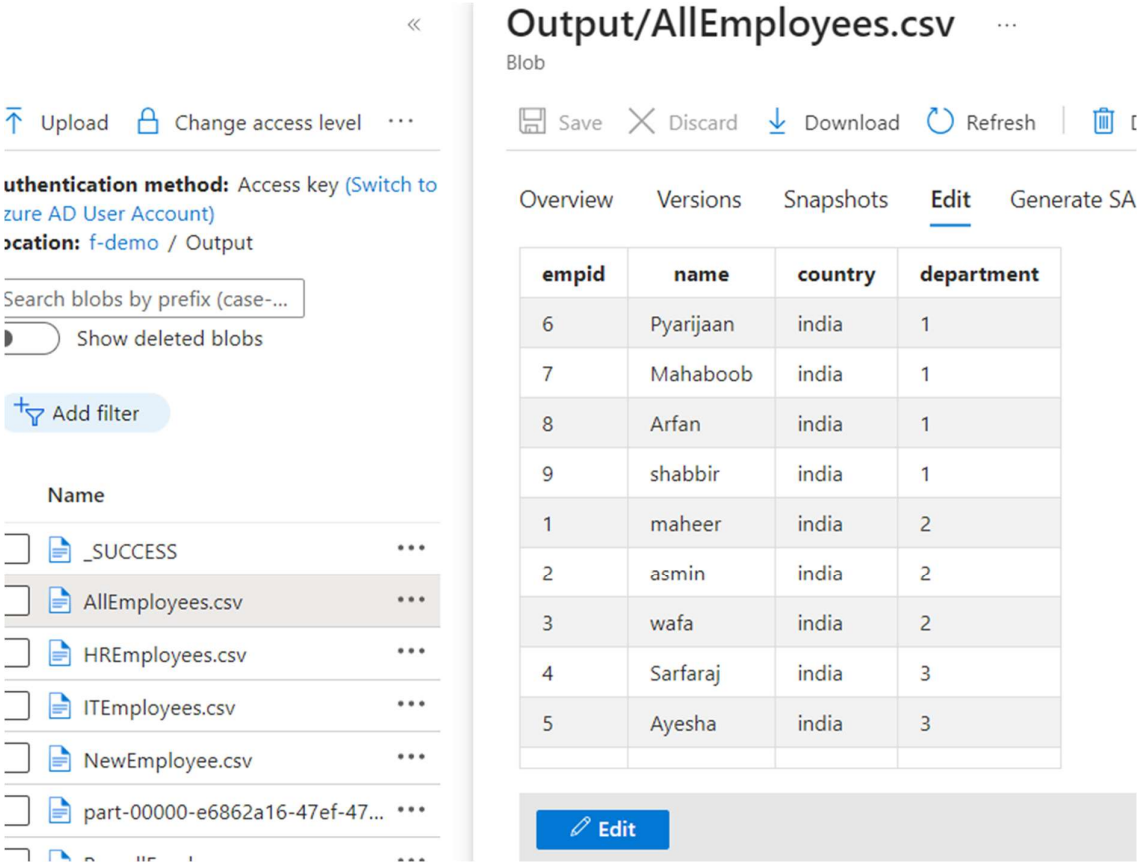
19. Next select the dataflow.



20. Validate the pipeline and click on debug.
21. Here our pipeline was executed successfully.

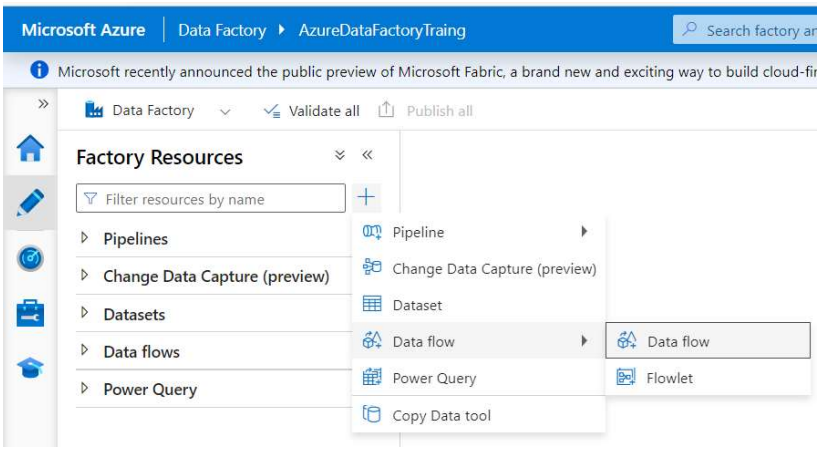


22. Next check the data in the destination.

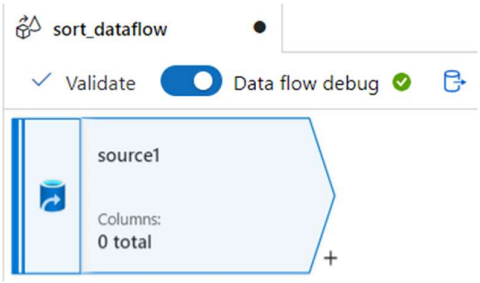


Sort Transformation

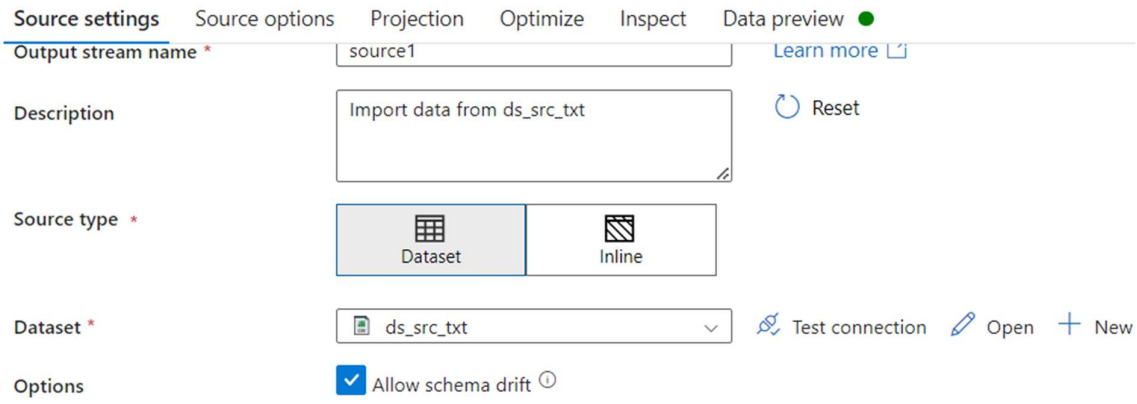
1. In this example we are using the Employee file. So here we are going to sort the emp name column.
2. Create a Data flow.



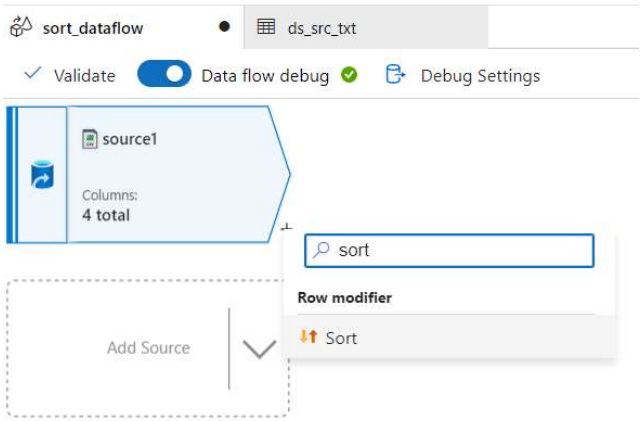
3. Add data source.



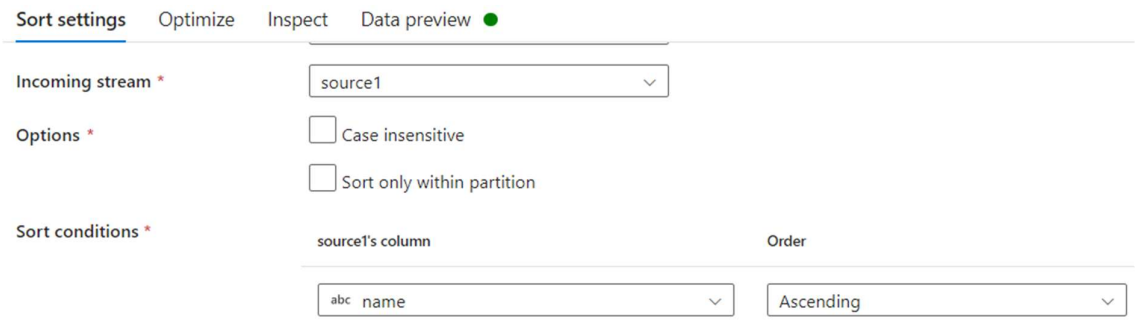
4. Under the source select the source dataset and file as shown below.



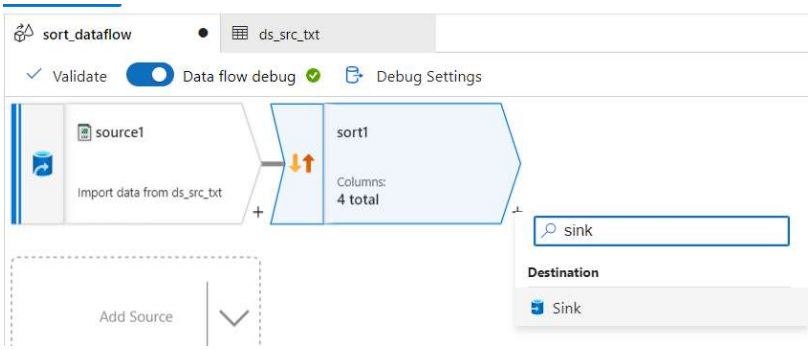
5. Click on the plus symbol and click on sort as shown below.



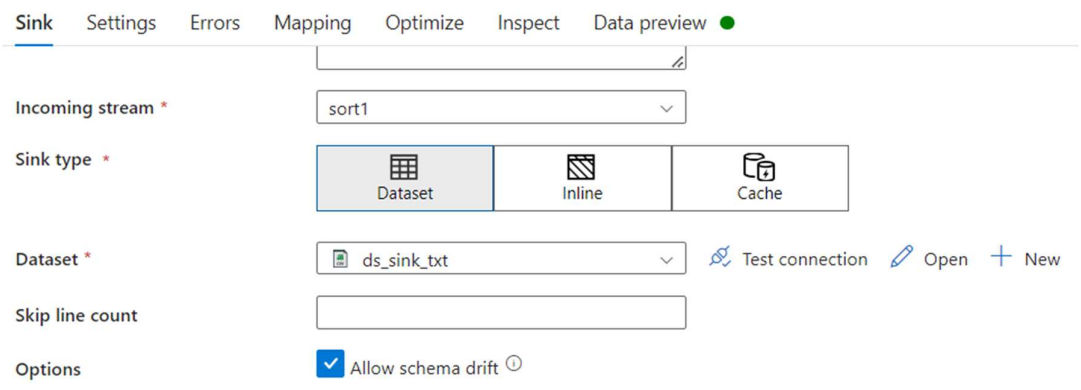
6. Under the sort setting select the name and give an Ascending order as shown below.



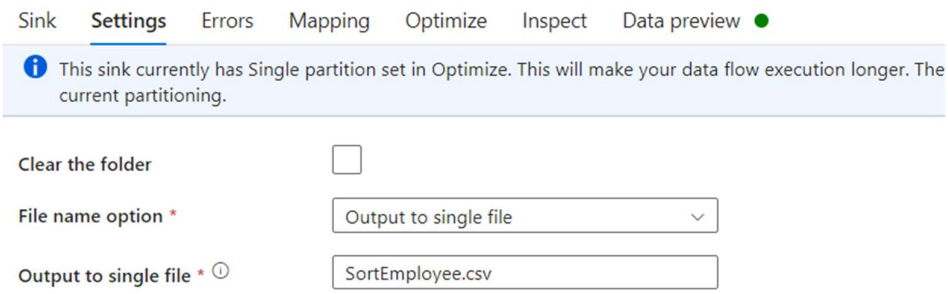
7. Now click on the plus symbol and click on the sink.



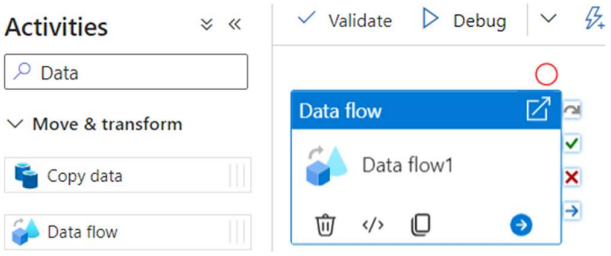
8. Under sink select the destination dataset as shown below.



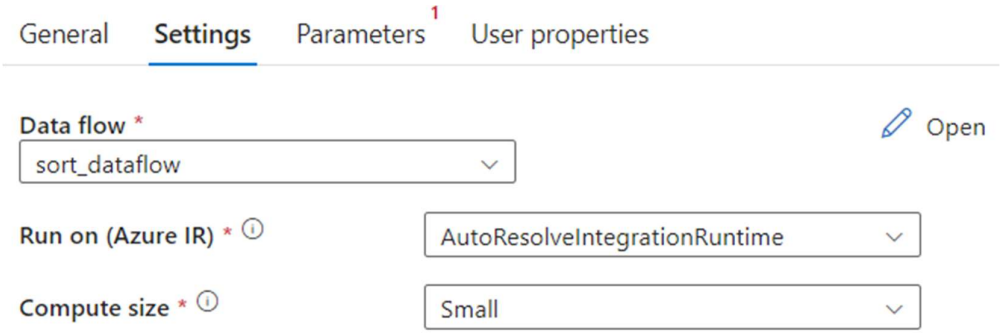
9. Under setting select the select the file option and file name as shown below.



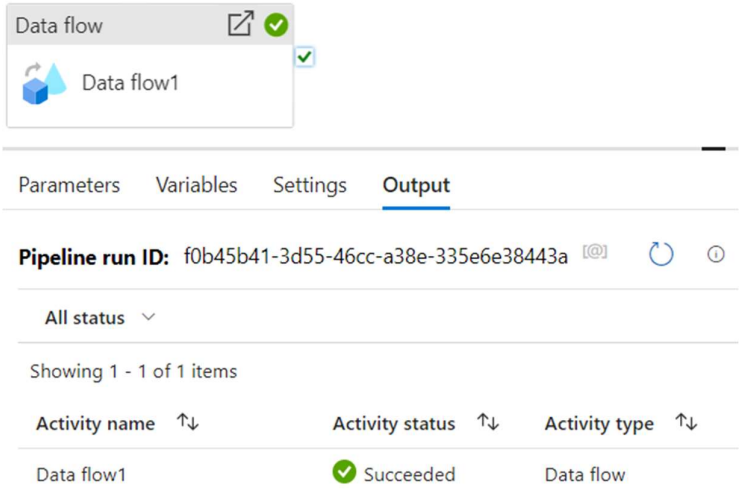
10. Create a pipeline then drag and drop the data flow activity.



11. Select the Data flow.

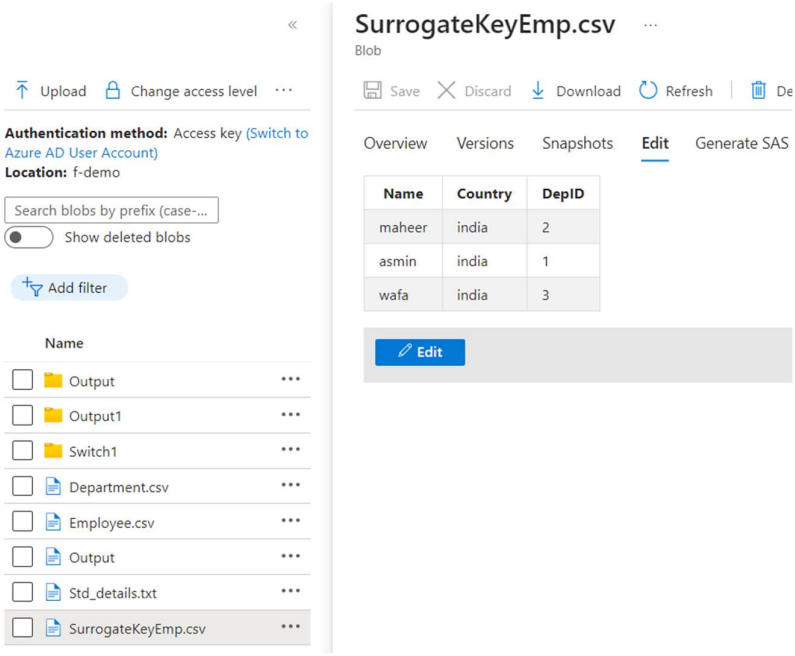


12. Now validate the pipeline and click on debug.
13. Here our pipeline was executed successfully.

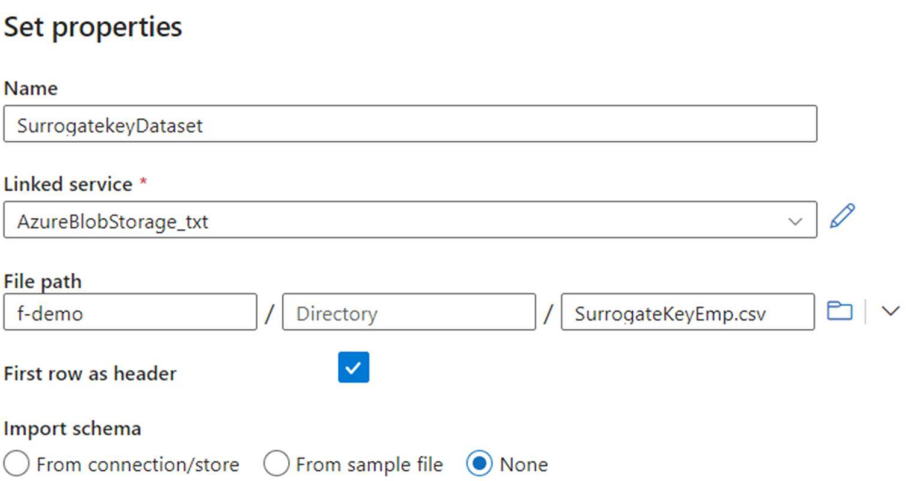


Surrogate Key Transformation

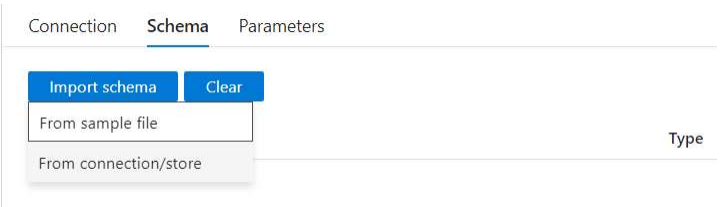
1. In this example we are using different data files as shown below.



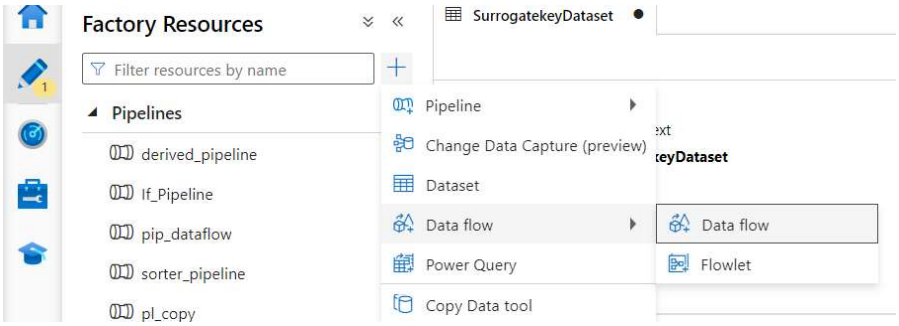
2. In the Azure Data Factory create a Data set for this file.



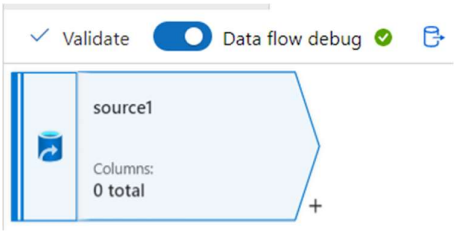
3. Next under schema. Click on Import schema and click on From connection/store.



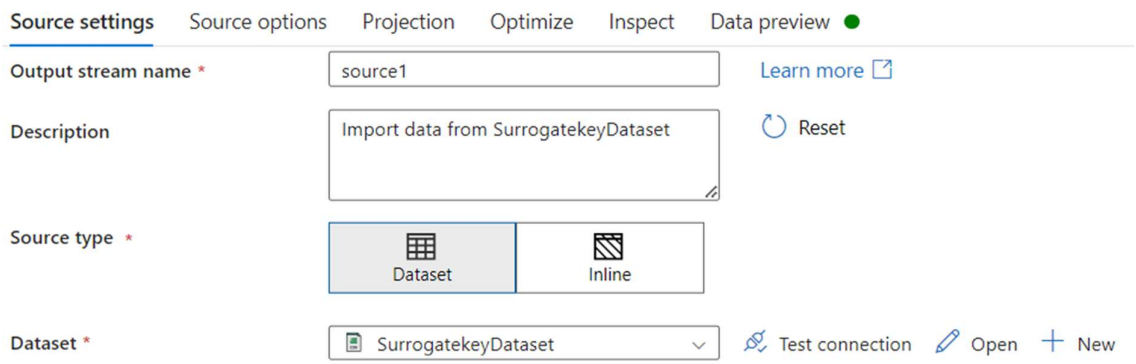
4. Create a data flow.



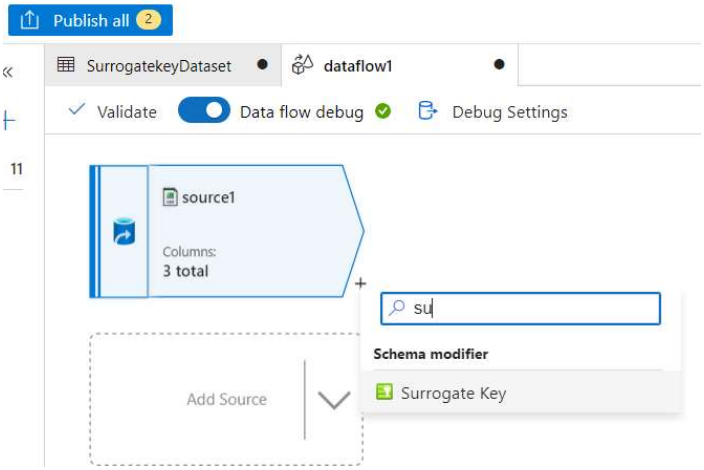
5. Add a data source.



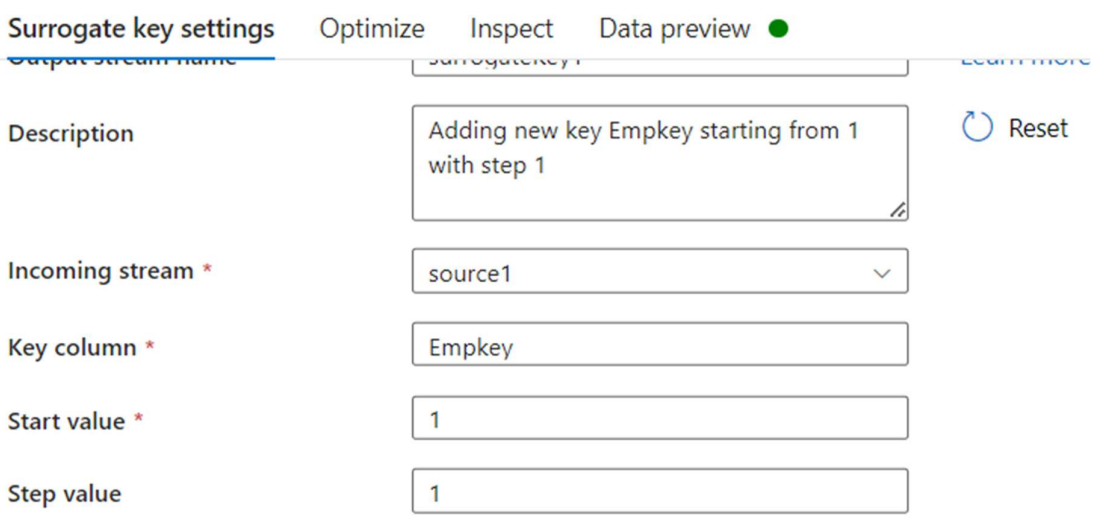
6. Select the surrogate data set that we created in the source.



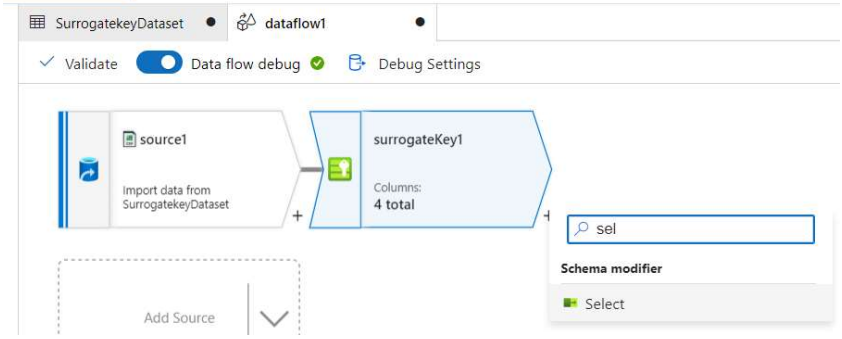
7. Click on the plus symbol and click on the surrogate key.



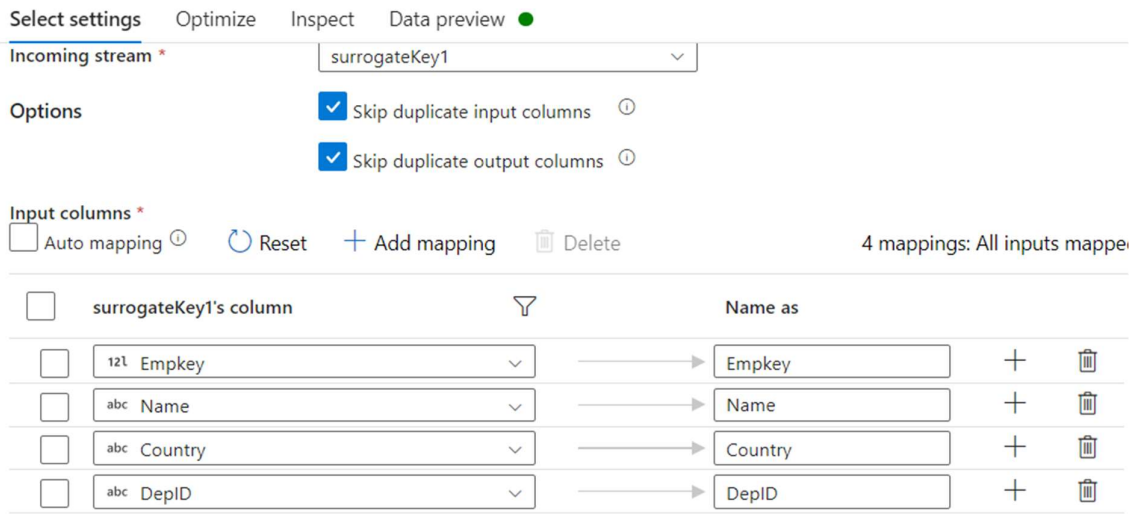
8. Set the key column as Empkey.



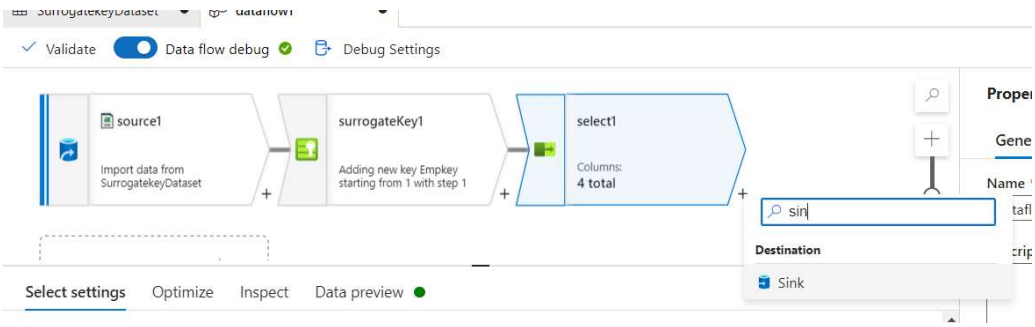
9. Click on the plus symbol and click on Select.



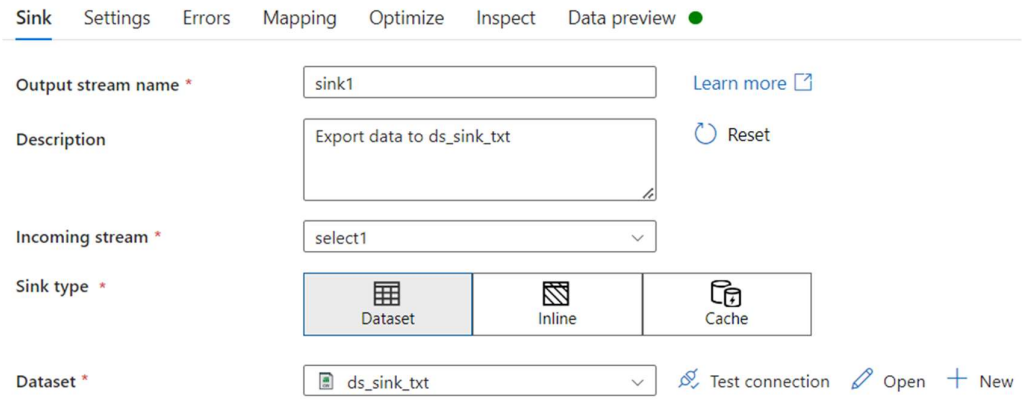
10. Under Select set the properties as shown below.



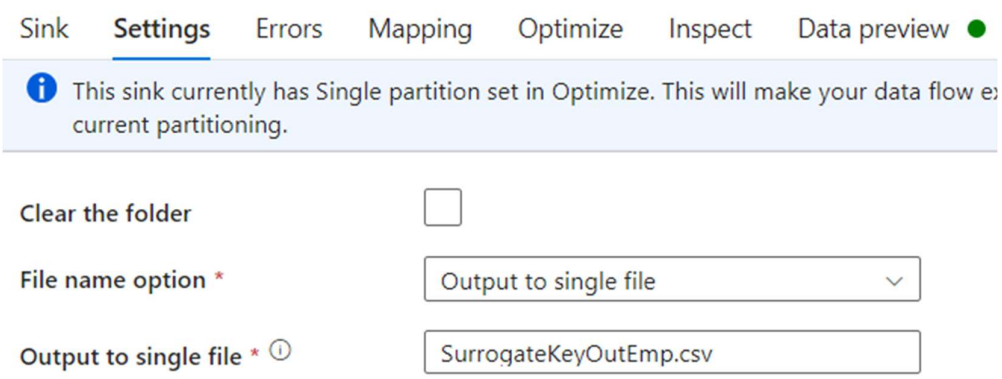
11. Click on the plus symbol and select the sink.



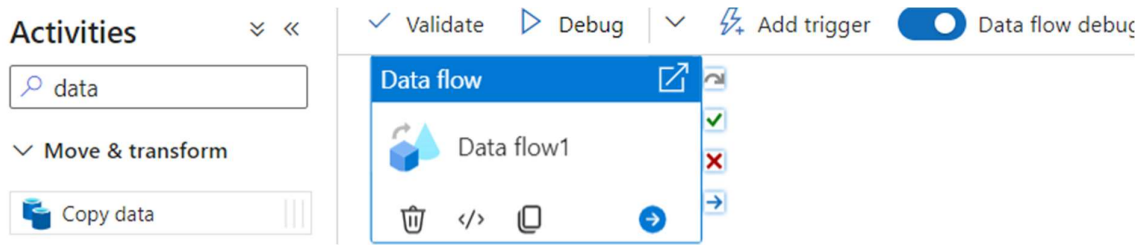
12. Select the destination dataset.



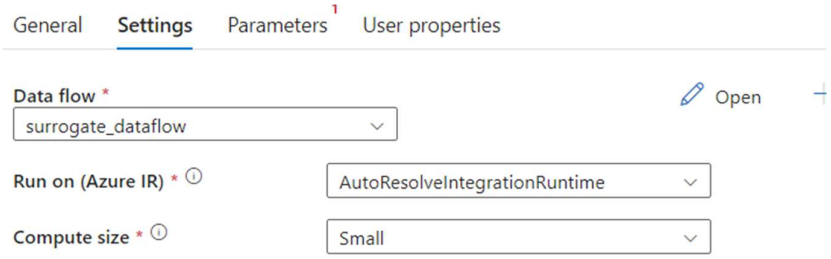
13. In the setting select the file option and file name as shown below.



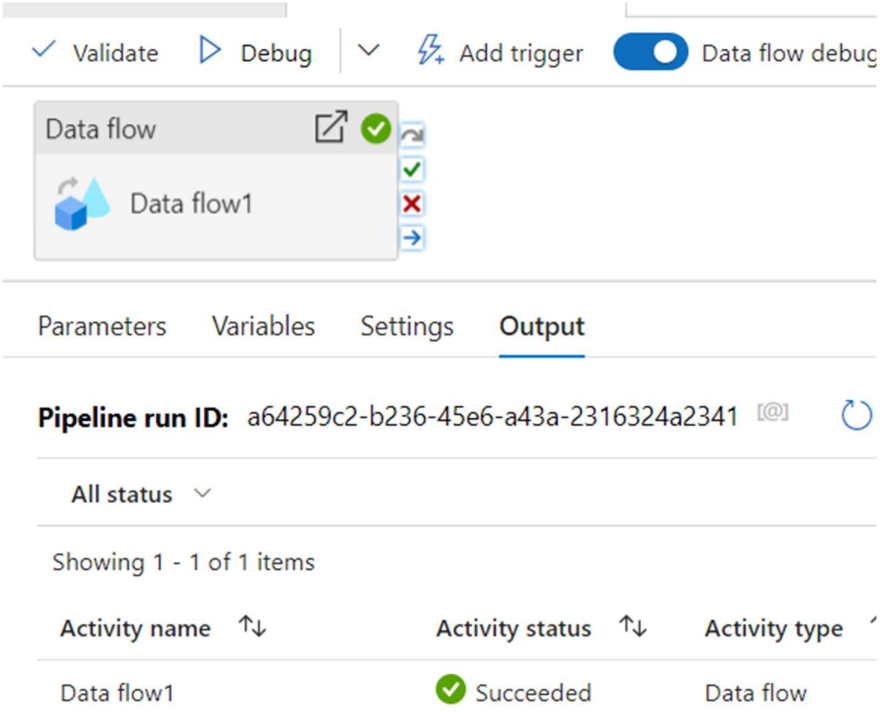
14. Create a pipeline then drag and drop the data flow activity.



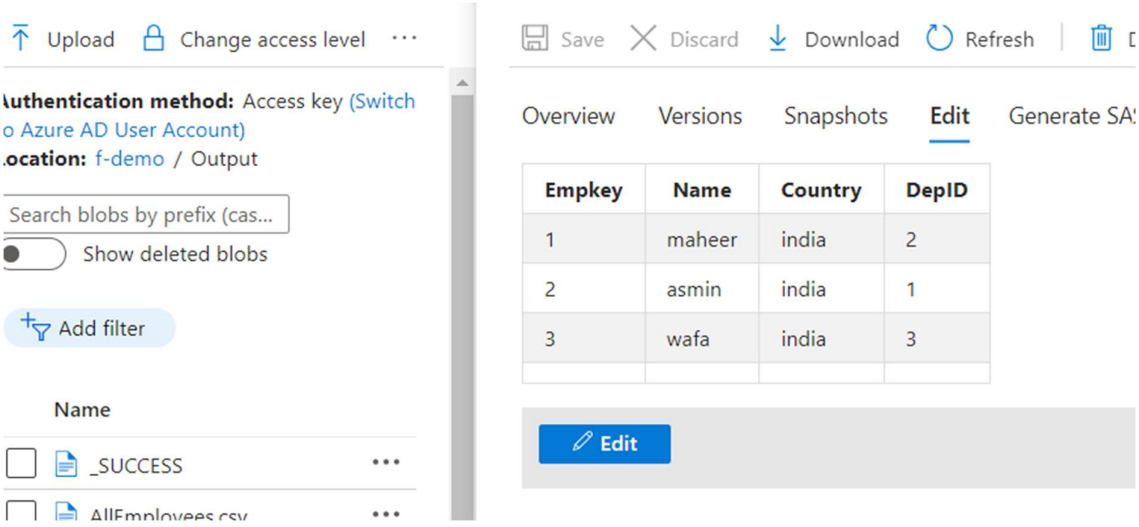
15. Select the data flow.



16. Validate the pipeline and click on debug.
17. Here our pipeline was executed successfully.

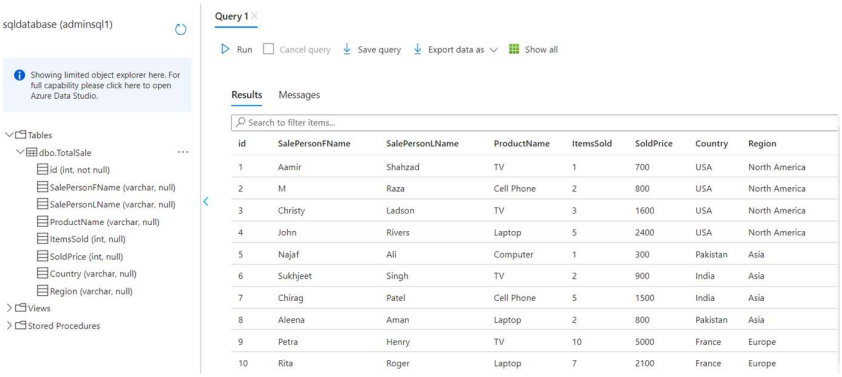


18. Now go to a destination and check the data.



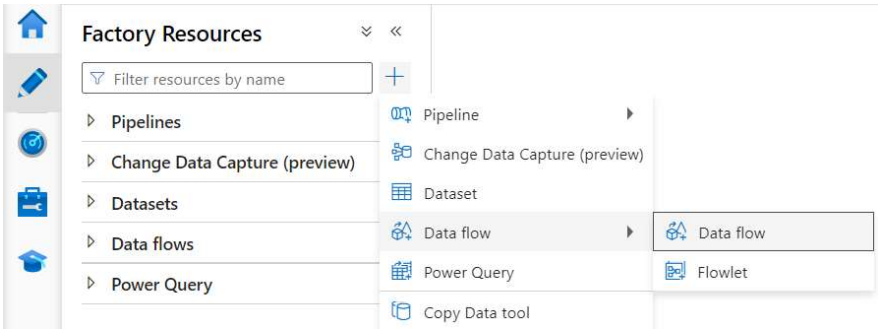
Filter Transformation

1. In this example we are going to use Azure SQL Database TotalSale table that we created in Day 1 session.

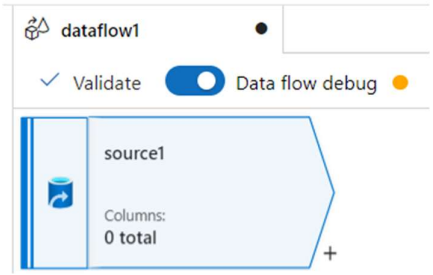


id	SalePersonFName	SalePersonLName	ProductName	ItemsSold	SoldPrice	Country	Region
1	Aamir	Shahzad	TV	1	700	USA	North America
2	M	Raza	Cell Phone	2	800	USA	North America
3	Christy	Ladson	TV	3	1600	USA	North America
4	John	Rivers	Laptop	5	2400	USA	North America
5	Najaf	Ali	Computer	1	300	Pakistan	Asia
6	Sukhjeet	Singh	TV	2	900	India	Asia
7	Chirag	Patel	Cell Phone	5	1500	India	Asia
8	Aleena	Aman	Laptop	2	800	Pakistan	Asia
9	Petra	Henry	TV	10	5000	France	Europe
10	Rita	Roger	Laptop	7	2100	France	Europe

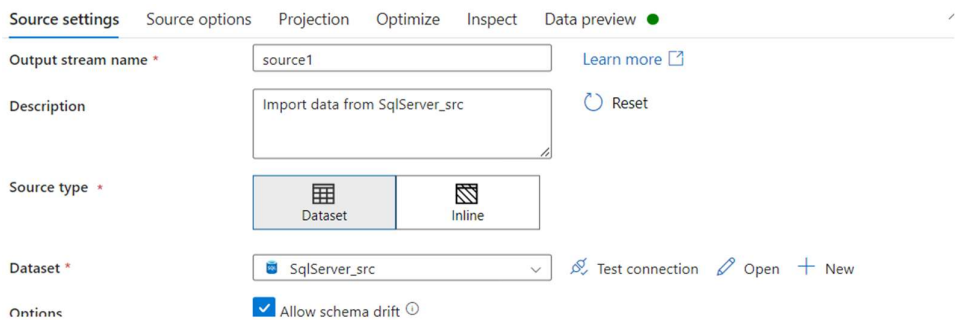
2. Create a Data flow.



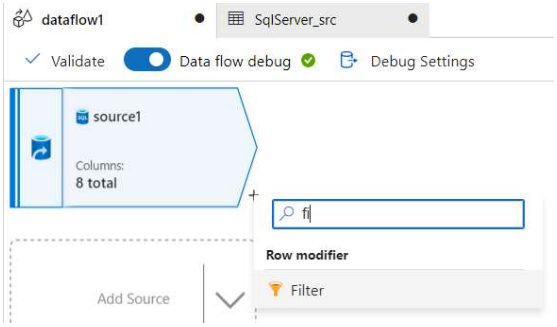
3. Add data source.



4. Under the source select the sql connection and table. Click on Open and go to Schema then click on Import Schema.

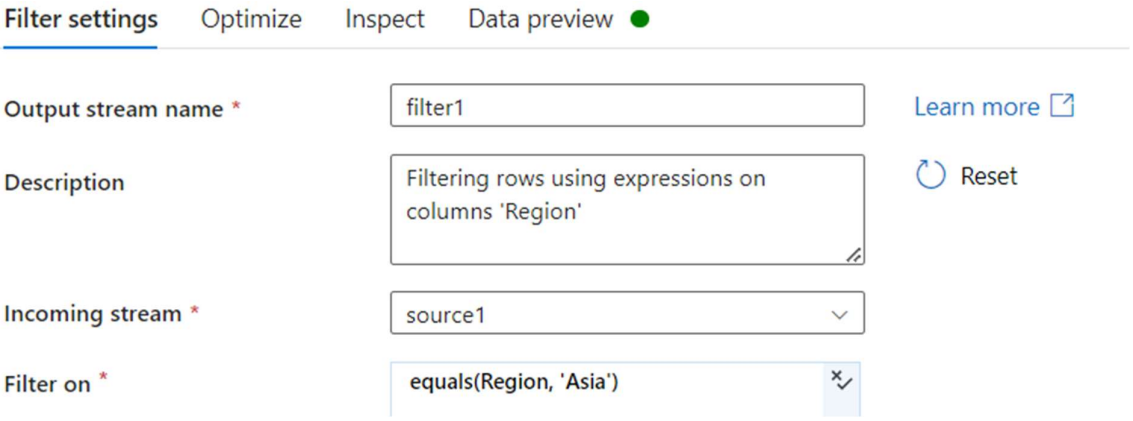


5. Click on the Plus symbol and click on Filter.

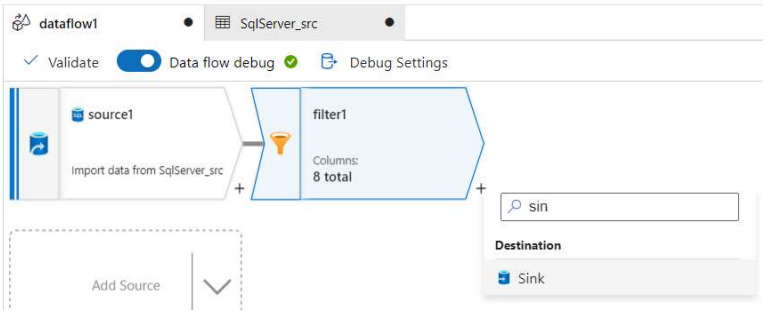


6. Under filter give the below expression.

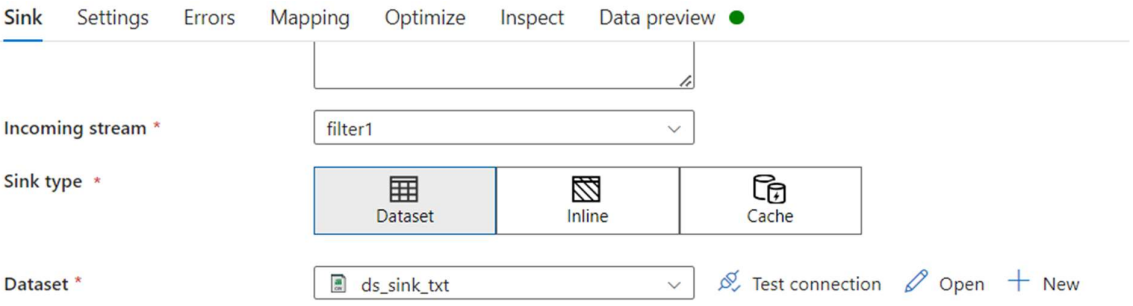
Exp: equals(Region, 'Asia')



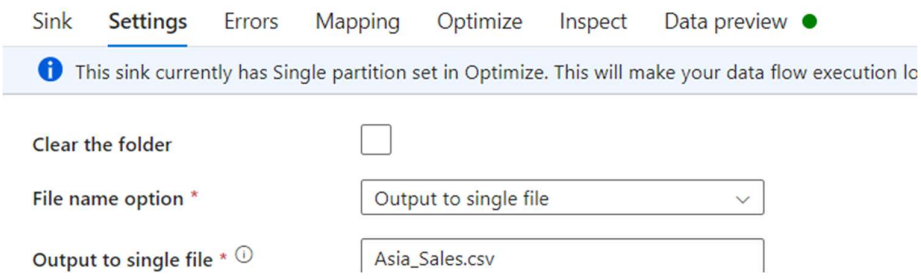
7. Click on the plus symbol and click on the sink.



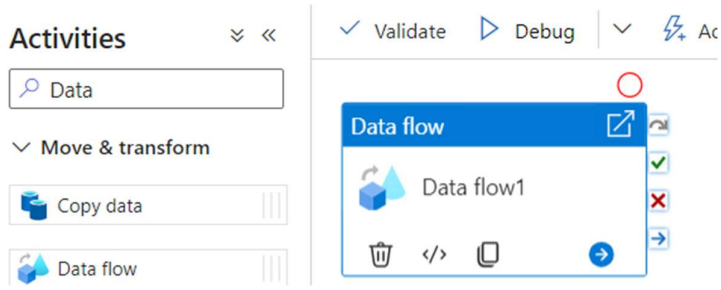
8. Under sink select the destination dataset.



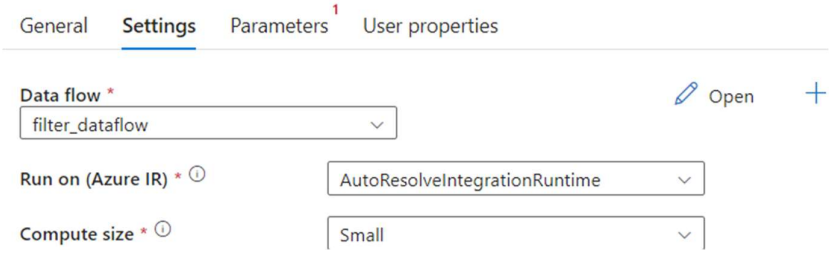
9. Under settings give the file option and file name as shown below.



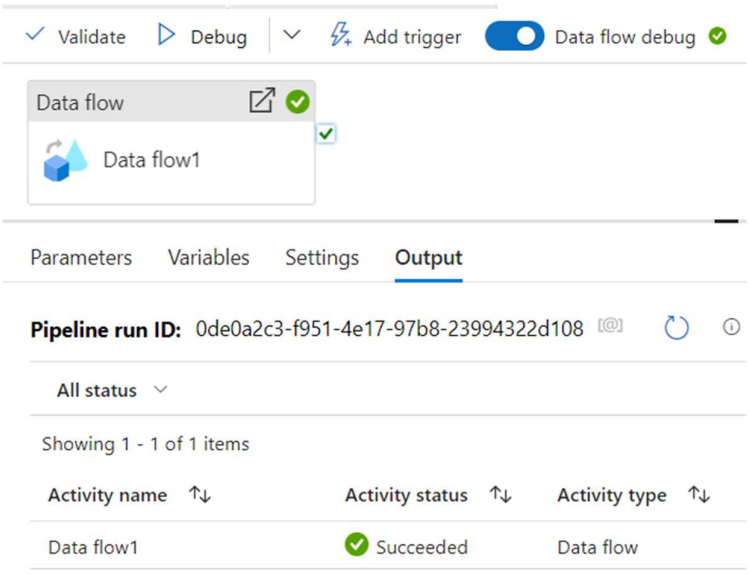
10. Create a Pipeline.
11. Drag and drop the Data flow activity.



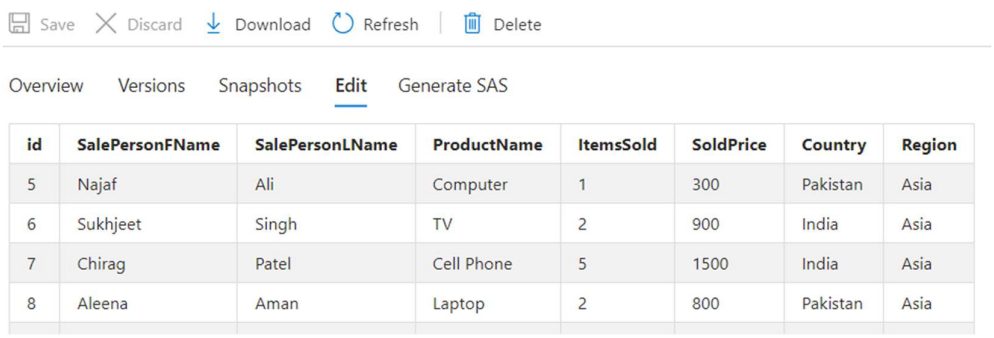
12. Select the Data flow as shown below.



13. Now validate the pipeline and click on debug.
14. Here our pipeline was executed successfully.



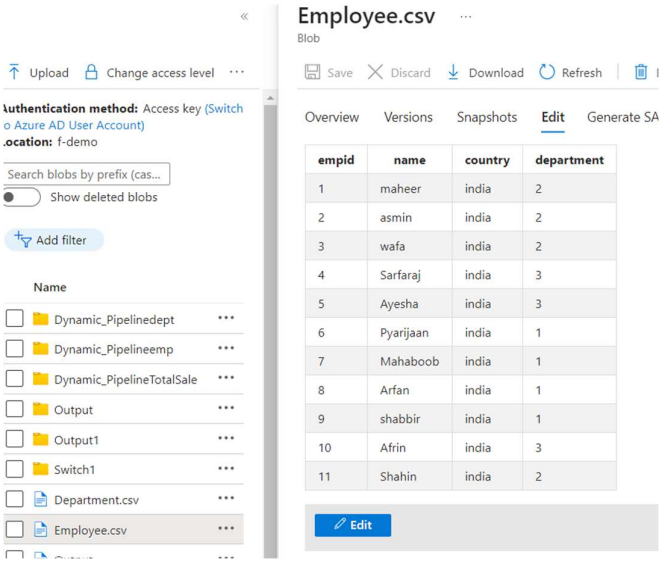
15. Now go to a destination and check the data.



id	SalePersonFName	SalePersonLName	ProductName	ItemsSold	SoldPrice	Country	Region
5	Najaf	Ali	Computer	1	300	Pakistan	Asia
6	Sukhjeet	Singh	TV	2	900	India	Asia
7	Chirag	Patel	Cell Phone	5	1500	India	Asia
8	Aleena	Aman	Laptop	2	800	Pakistan	Asia

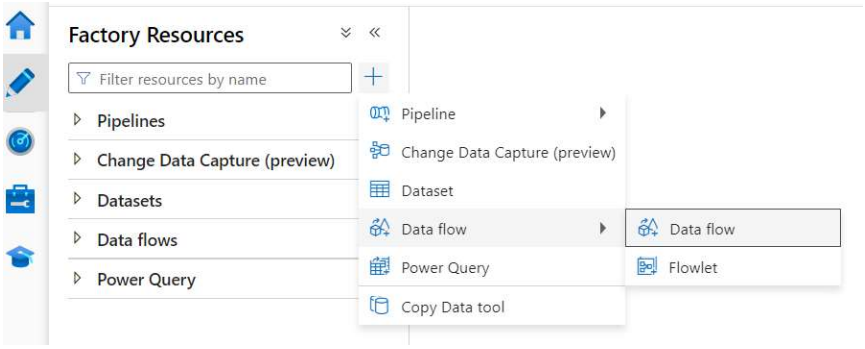
Aggregator Transformation

1. In this example we are going to use the Employee file.

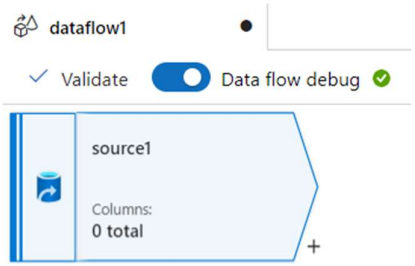


empid	name	country	department
1	maheer	india	2
2	asmin	india	2
3	wafa	india	2
4	Sarfaraj	india	3
5	Ayesha	india	3
6	Pyarijaan	india	1
7	Mahaboob	india	1
8	Arfan	india	1
9	shabbir	india	1
10	Afrin	india	3
11	Shahin	india	2

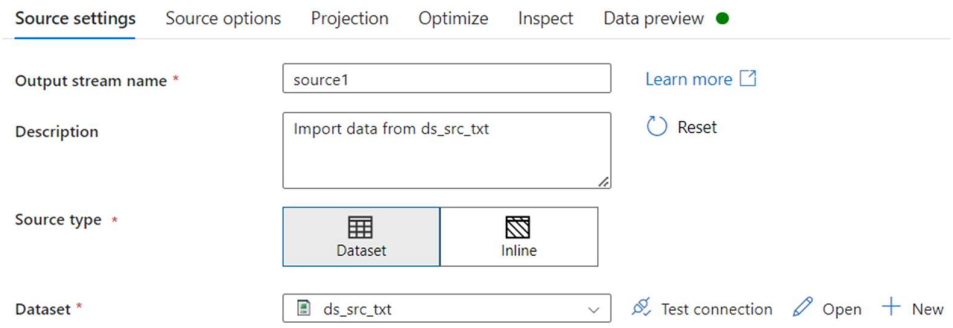
2. Create a Dataflow.



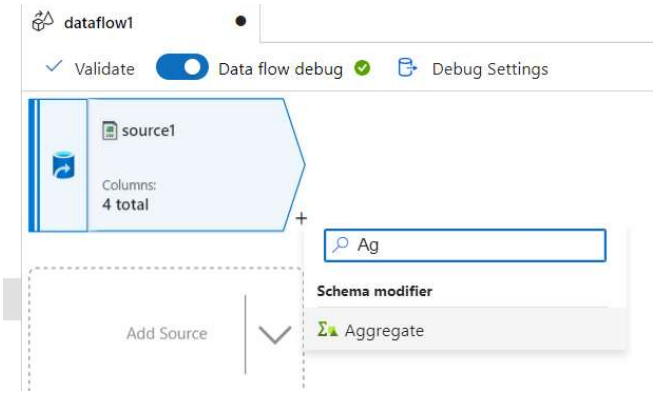
3. Add Data source.



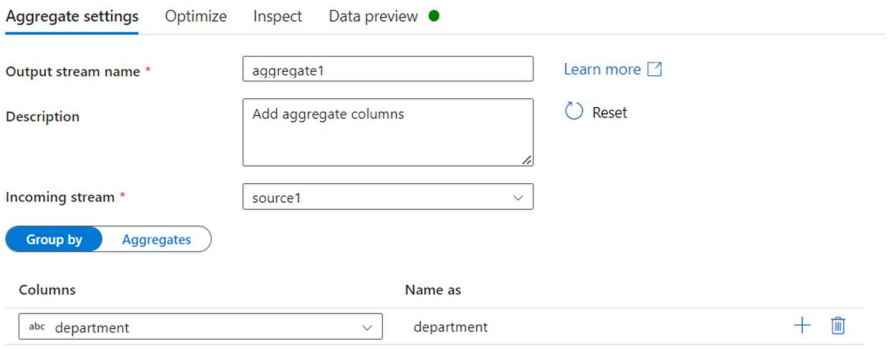
4. Under the source select the dataset and file.



5. Click on the plus symbol and click on the Aggregate.

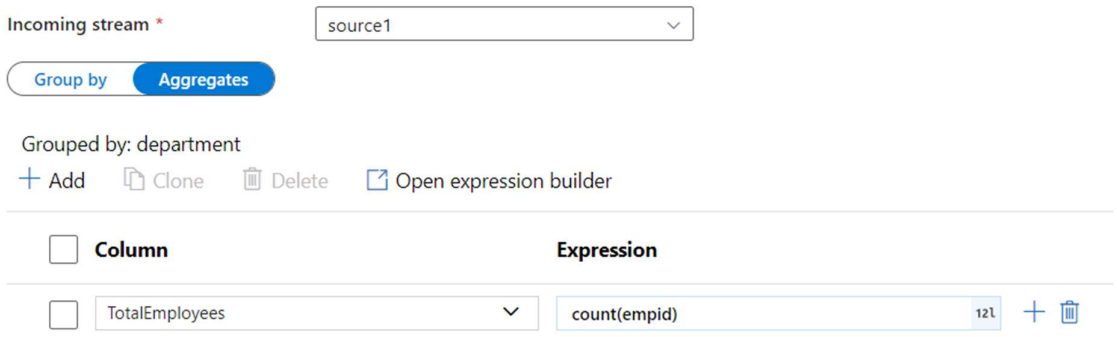


6. Under the Group select the Department column as shown below.

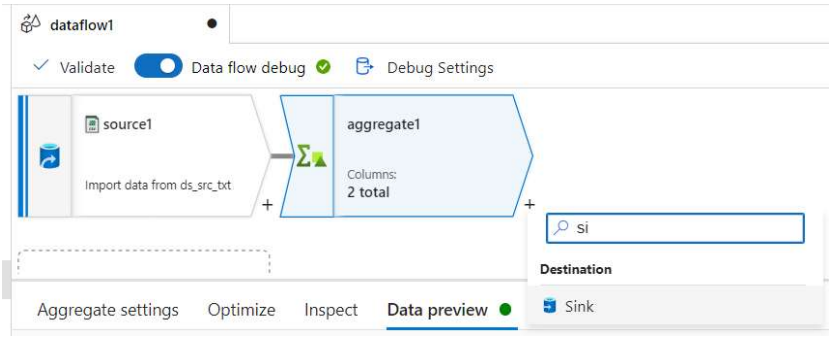


7. Under Aggregates type TotalEmployees and give the below expression.

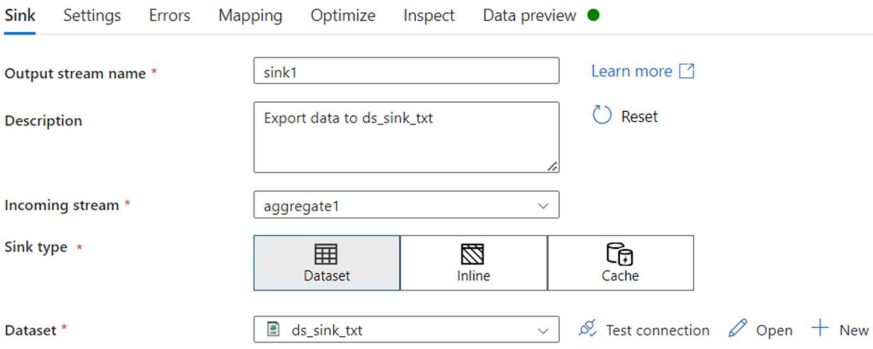
Exp: count(empid)



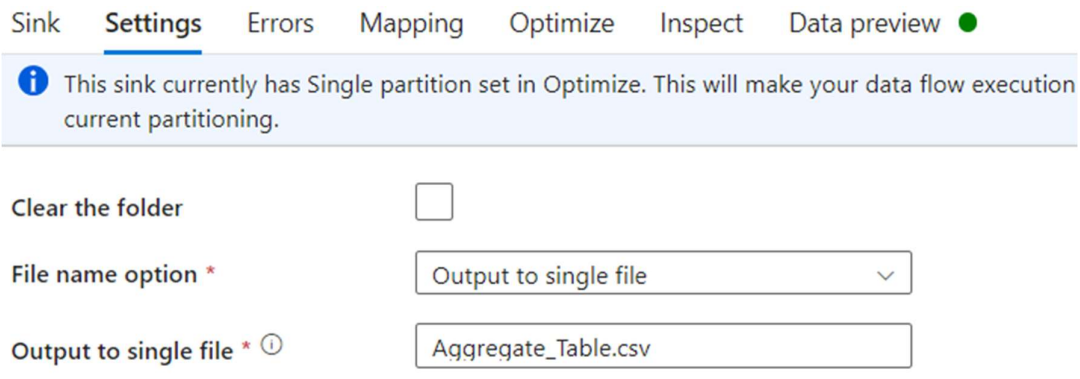
8. Click on the Plus symbol and click on Sink.



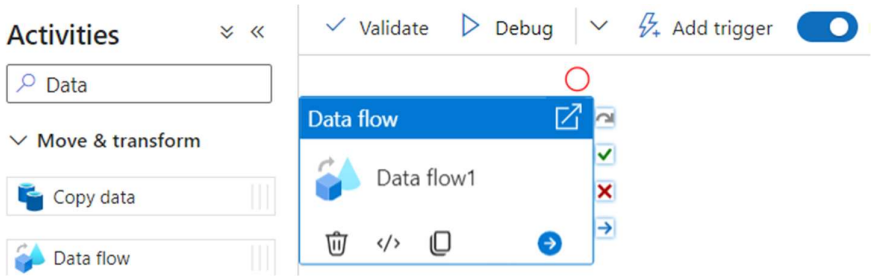
9. Under sink select the destination dataset.



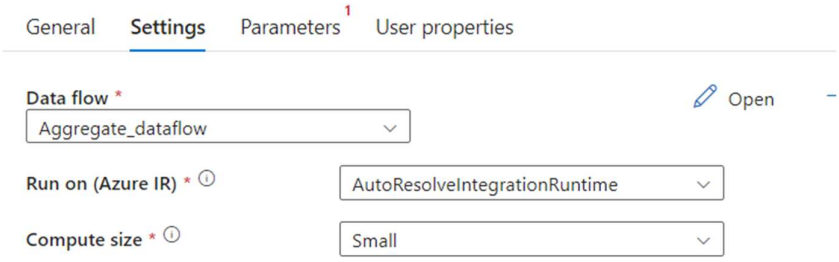
10. Under setting select the file option and name as shown below.



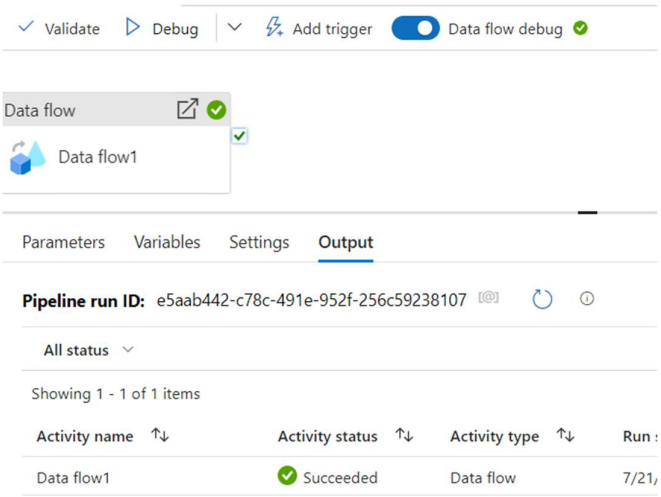
11. Create a pipeline.
12. Drag and drop the Data flow activity.



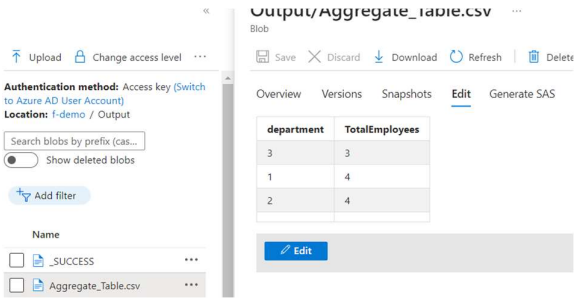
13. Select the data flow.



14. Validate the pipeline and click on debug.
15. Here our pipeline was executed successfully.

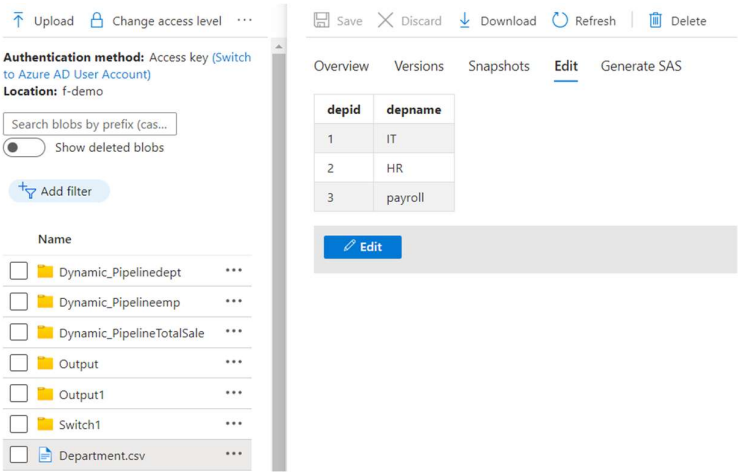


16. Now go to a destination and check the data.

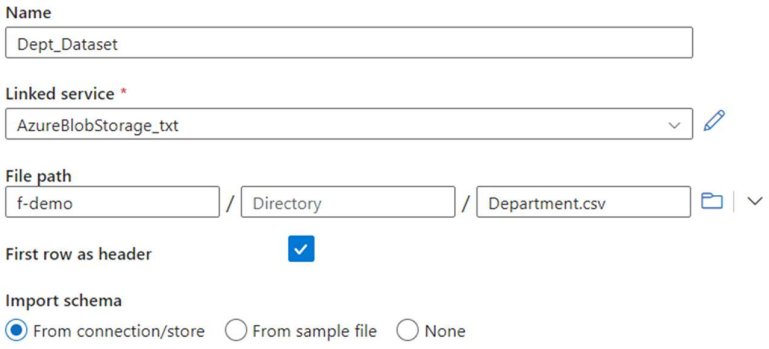


Join Transformation

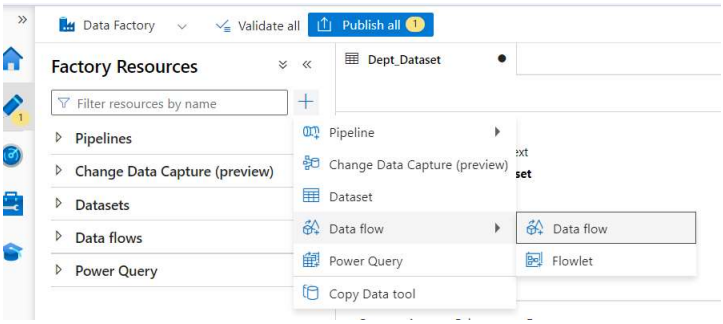
1. In this example, we are going to use two sources one is the Employee file, and the other one is the department file.



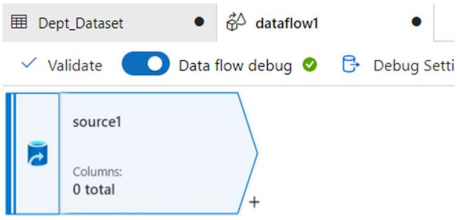
2. Create a dataset for this department file.



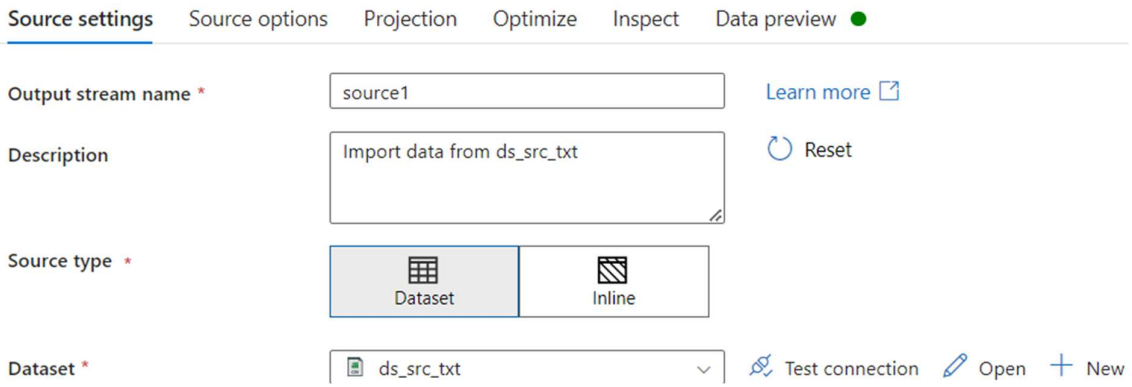
3. Create a Data flow.



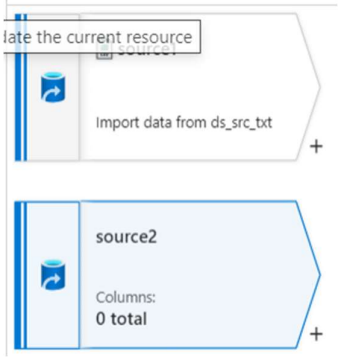
4. Add a data source.



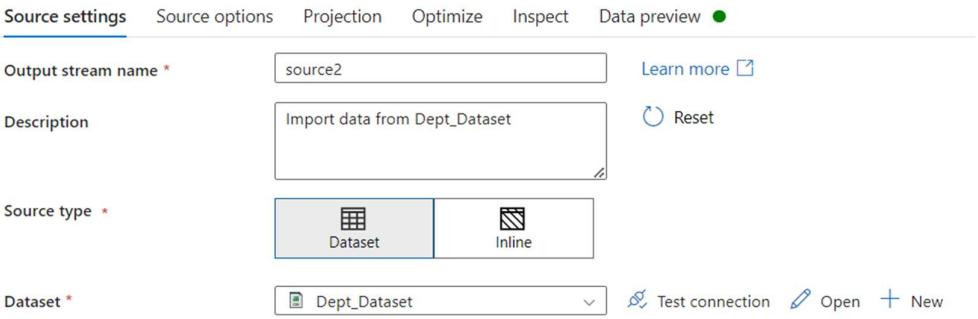
5. Under the source select the Employee file dataset.



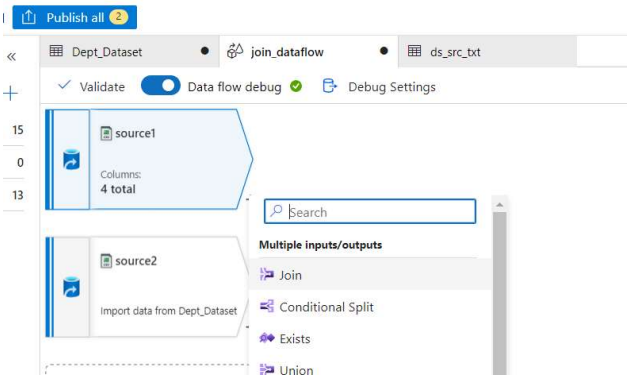
6. Add another data source.



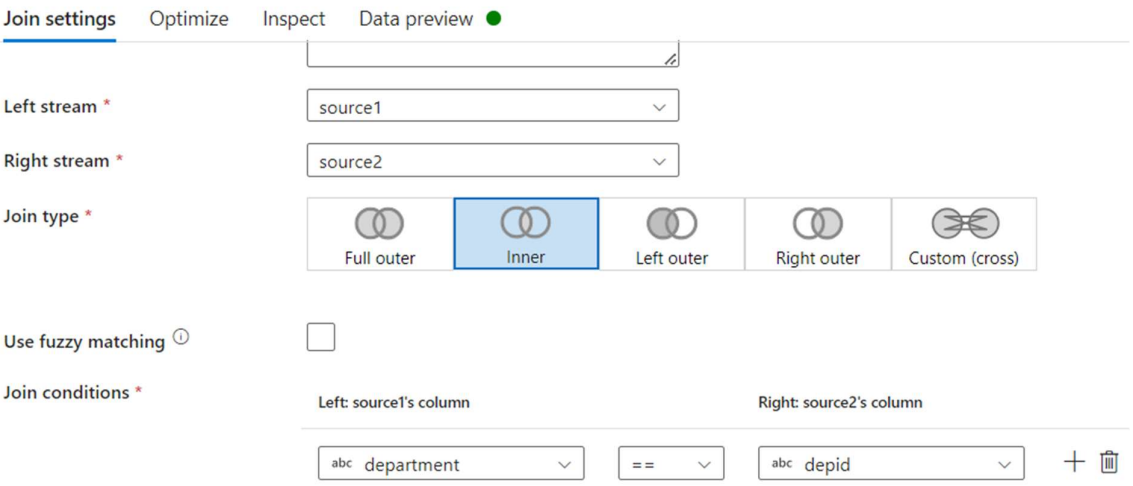
7. Select the department dataset that we created before.



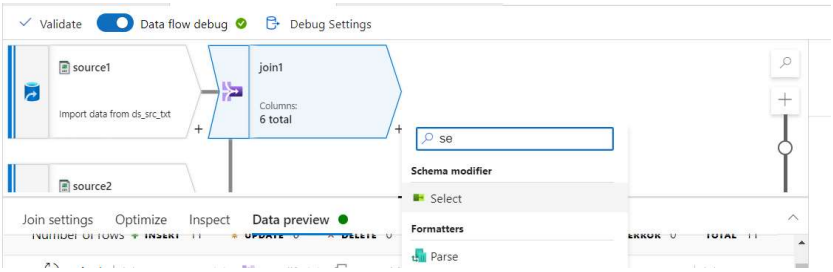
8. Click on the plus symbol and click on join.



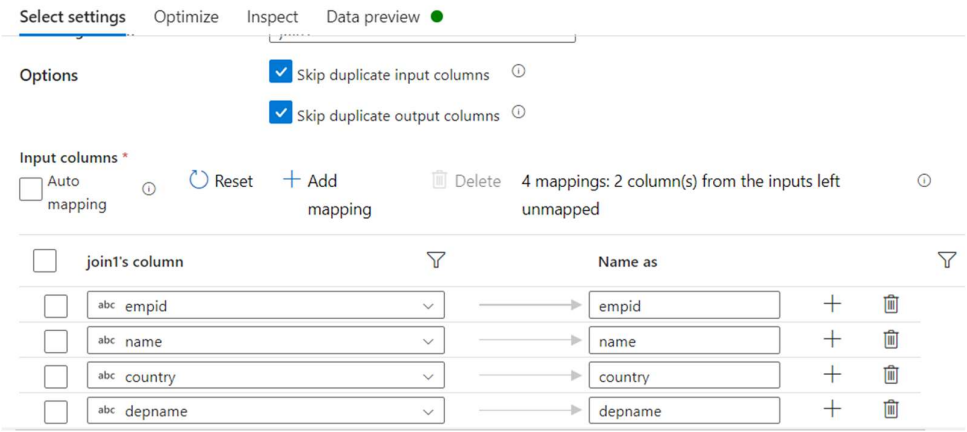
9. Under the join setting select the Right stream and give the below condition.



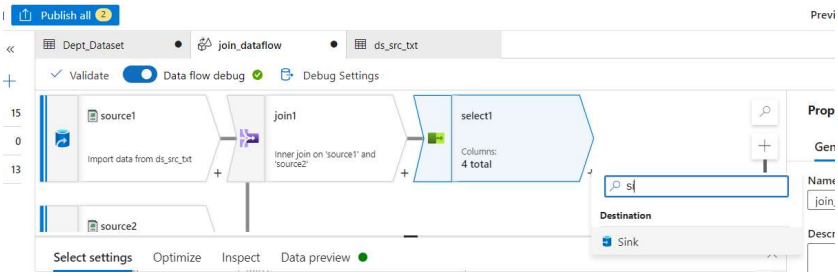
10. Click on the Plus symbol and click on Select.



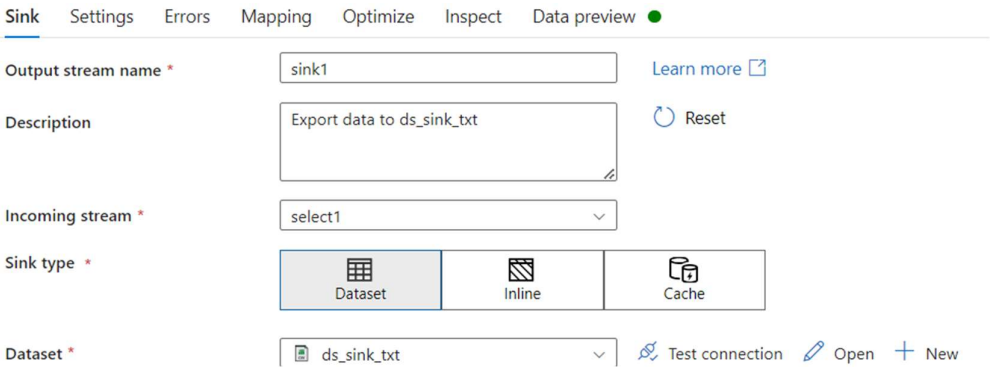
11. Under settings delete the department and depid column as shown below.



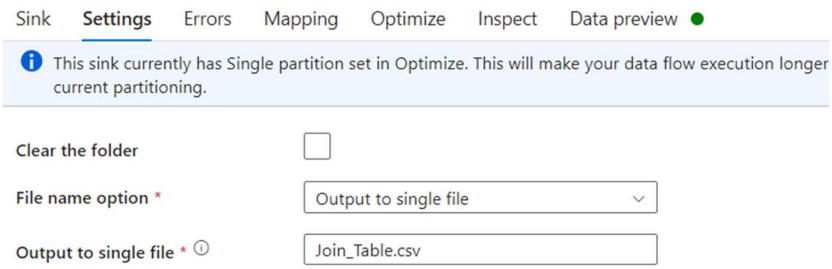
12. Click on the plus symbol and click on the sink.



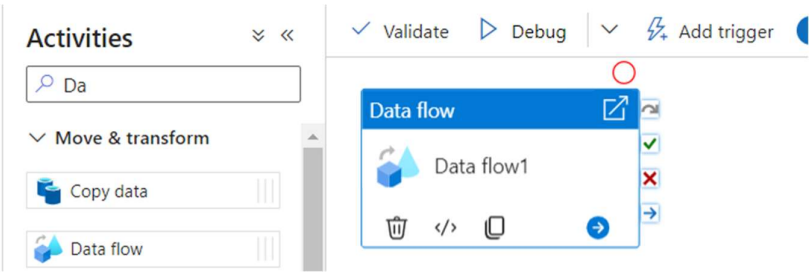
13. Under sink select the destination dataset.



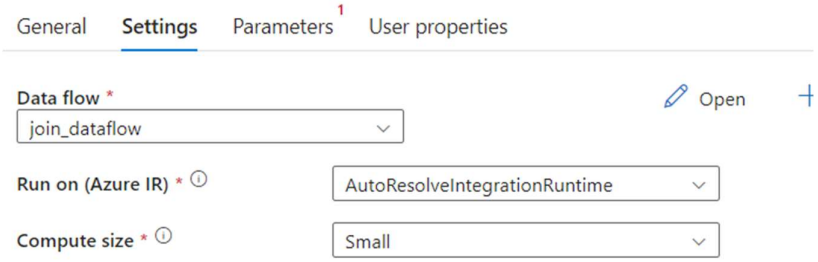
14. Under setting select the file option and file name as shown below.



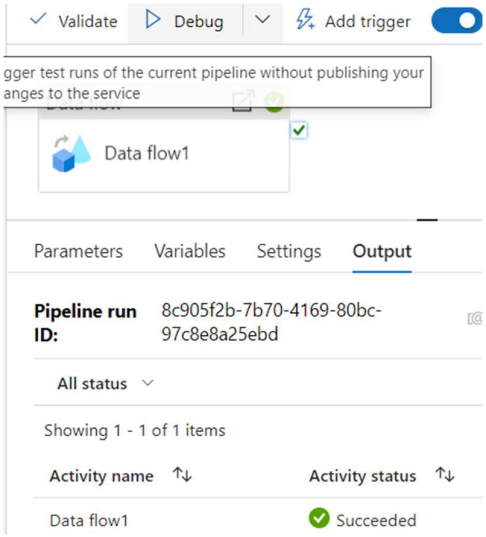
15. Create a pipeline and drag and drop the data flow activity.



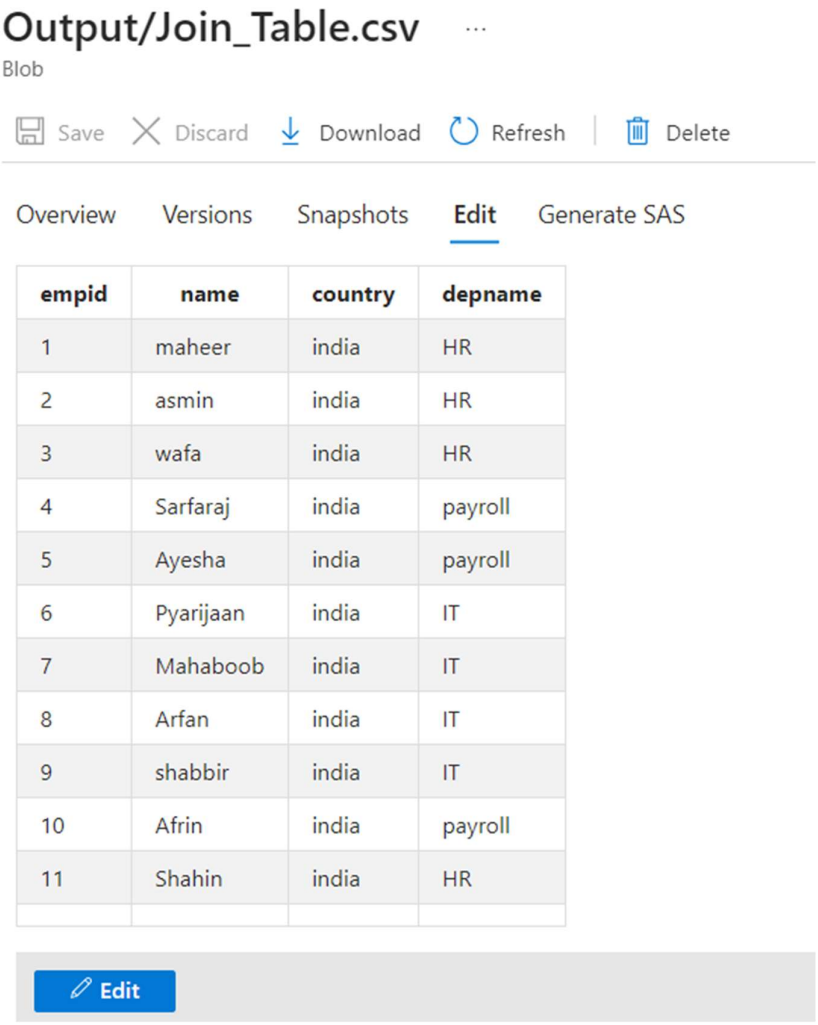
16. Select the data flow.



17. Validate it and click on debug.



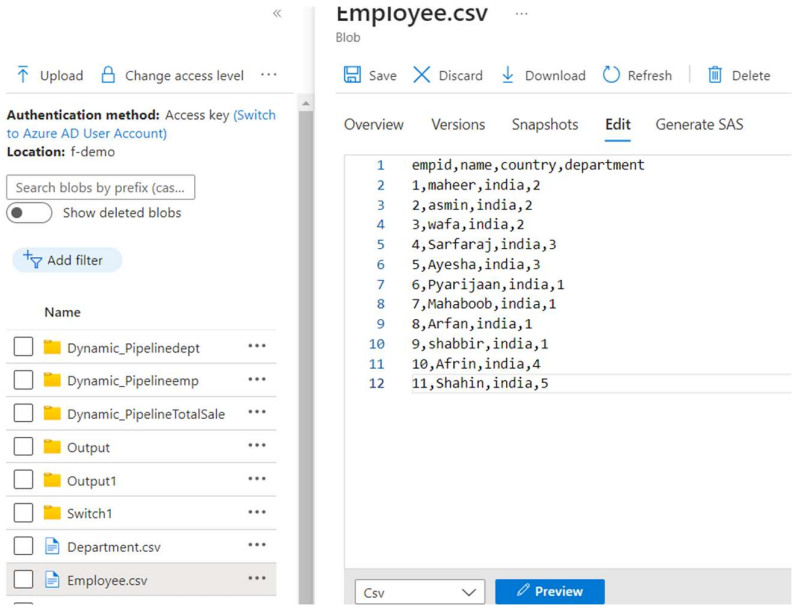
18. Go to the destination and check the data.



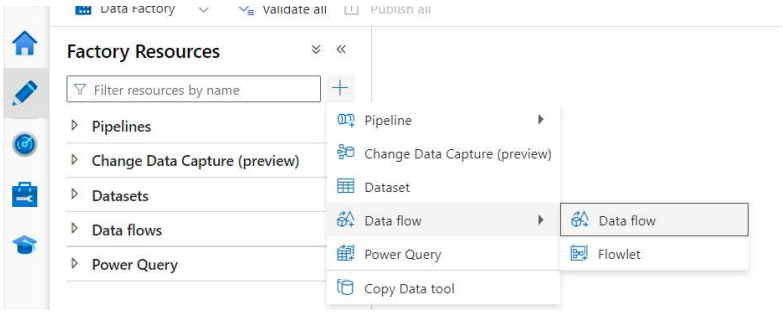
empid	name	country	depname
1	maheer	india	HR
2	asmin	india	HR
3	wafa	india	HR
4	Sarfaraj	india	payroll
5	Ayesha	india	payroll
6	Pyarijaan	india	IT
7	Mahaboob	india	IT
8	Arfan	india	IT
9	shabbir	india	IT
10	Afrin	india	payroll
11	Shahin	india	HR

Lookup Transformation

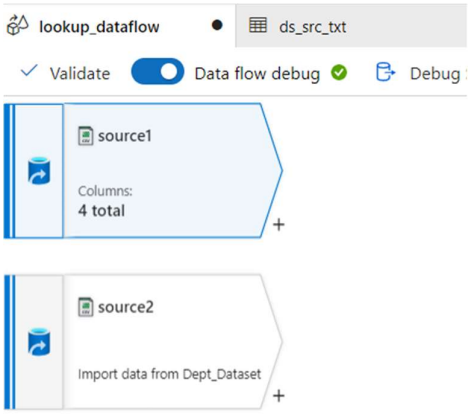
1. In this example also we are using department and employee files. But in the employee file, we are going to edit a few values.
2. Here I edited the department value for the last two rows.



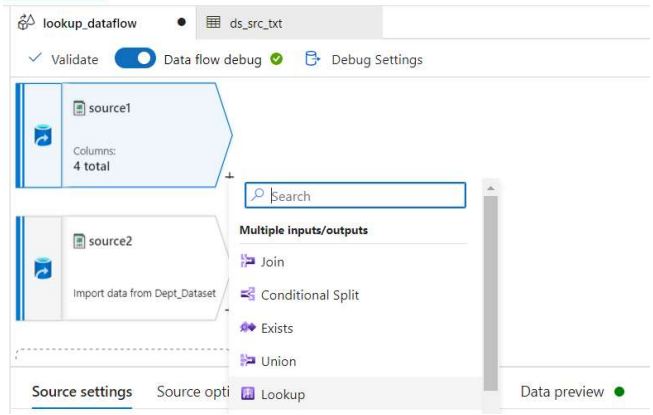
3. Create a data flow.



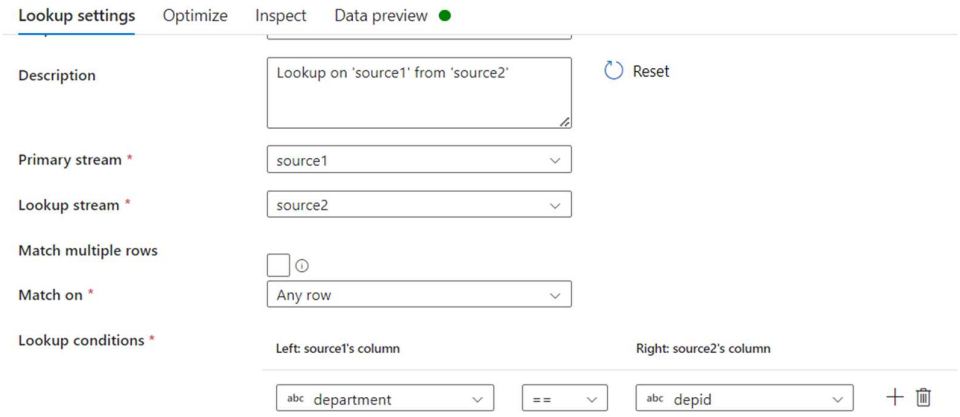
4. Add a source two sources and load the source file for the Employee and department.



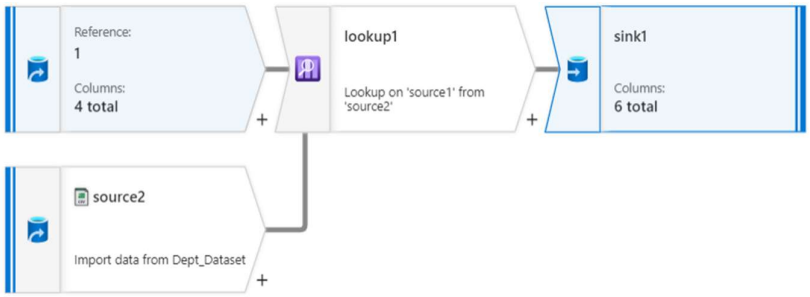
5. Click on the plus symbol and click on lookup as shown below.



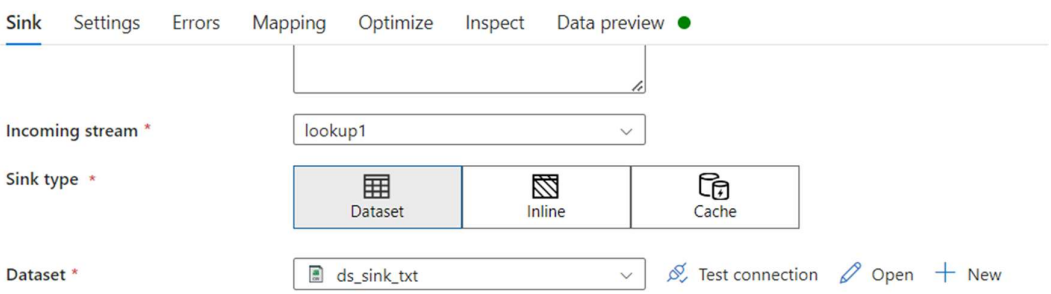
6. Under lookup select the lookup stream and give the below condition.



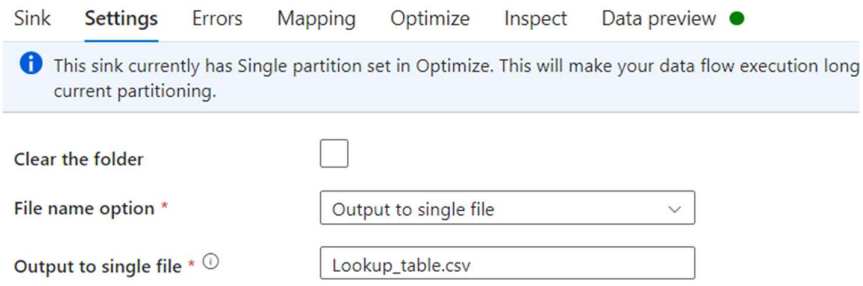
7. Add sink.



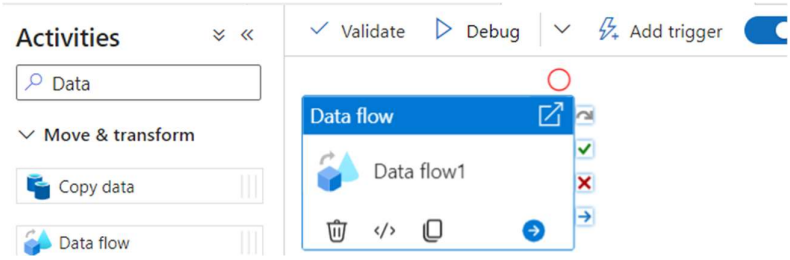
8. Under sink select the destination dataset.



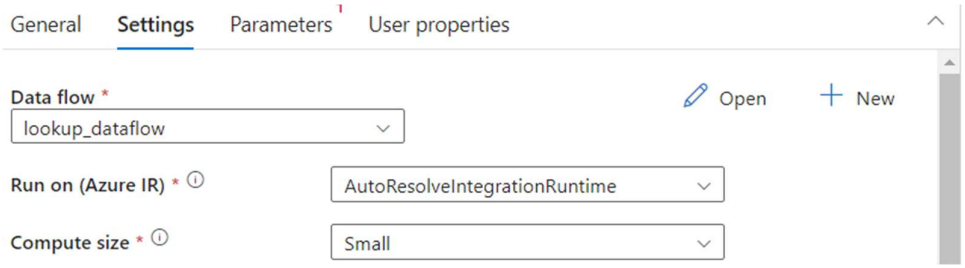
9. Under setting select the file option and give the file name.



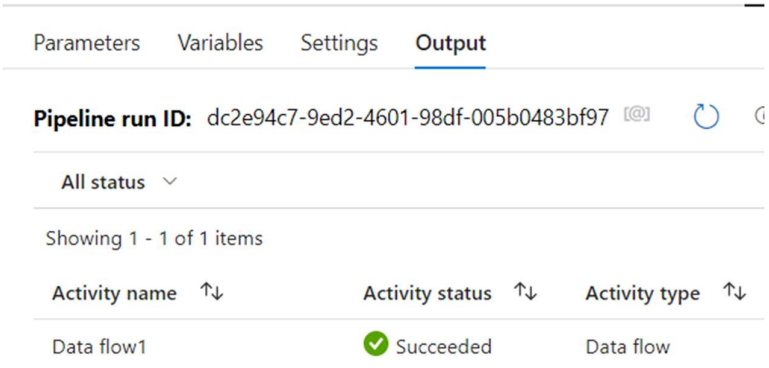
10. Create a lookup and drag and drop the data flow.



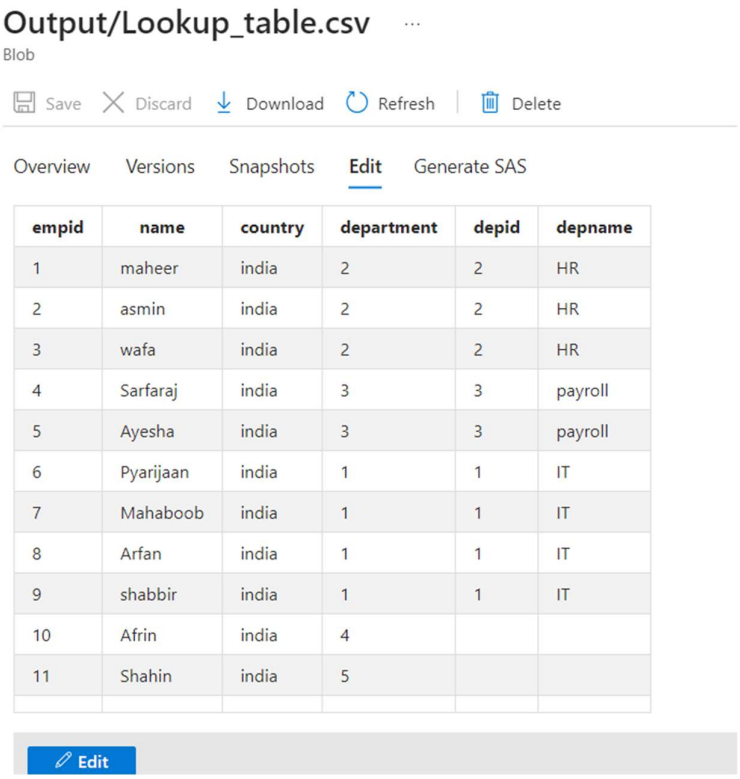
11. Select the data flow.



12. Validate the pipeline and click on debug.
13. Here our Pipeline was executed successfully.



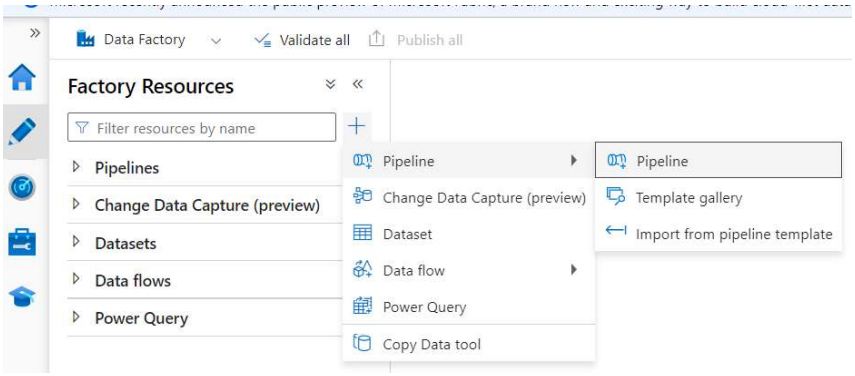
14. Go to the destination and check the data.



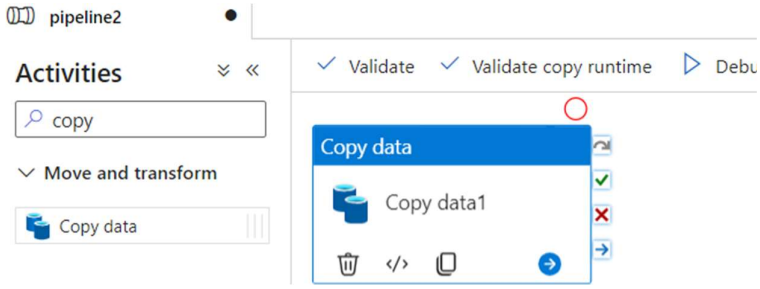
empid	name	country	department	depid	depname
1	maheer	india	2	2	HR
2	asmin	india	2	2	HR
3	wafa	india	2	2	HR
4	Sarfaraj	india	3	3	payroll
5	Ayesha	india	3	3	payroll
6	Pyarijaan	india	1	1	IT
7	Mahaboob	india	1	1	IT
8	Arfan	india	1	1	IT
9	shabbir	india	1	1	IT
10	Afrin	india	4		
11	Shahin	india	5		

Pipeline Using REST

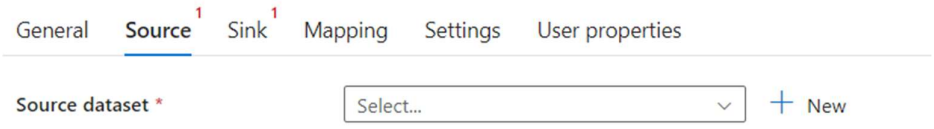
1. Open Azure Data Factory.
2. Create a Pipeline.



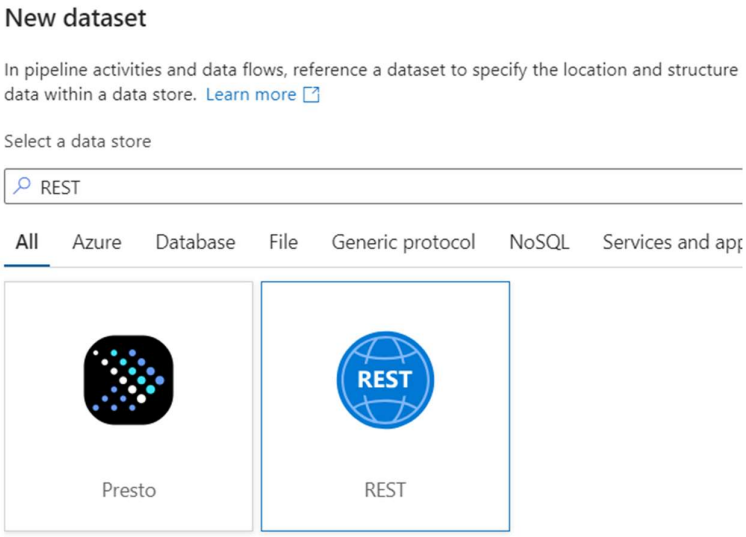
3. Drag and drop the Copy Data Activity as shown below.



4. Under source click on New.



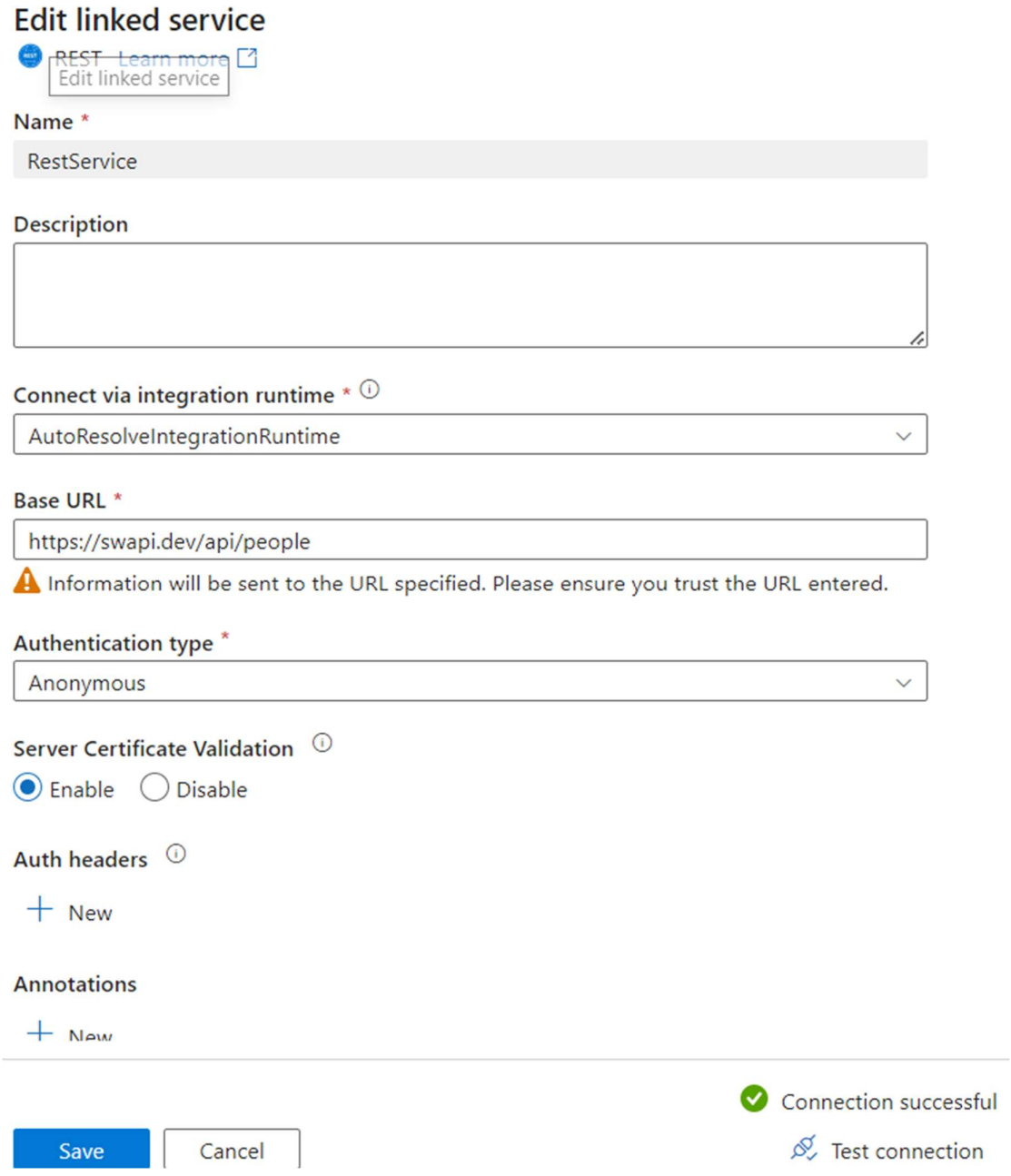
5. Select REST and click on continue.



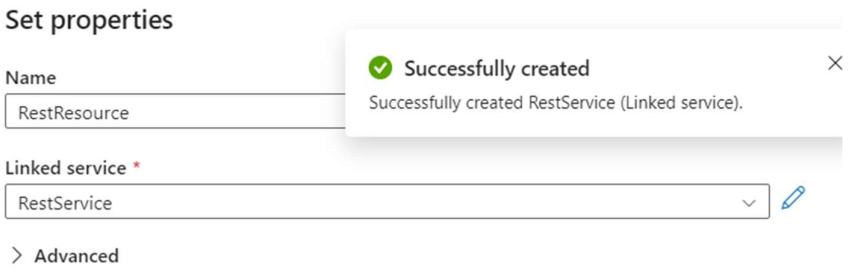
6. Give the Name and click on New.

7. In the Linked Service under basic URL give the below URL.

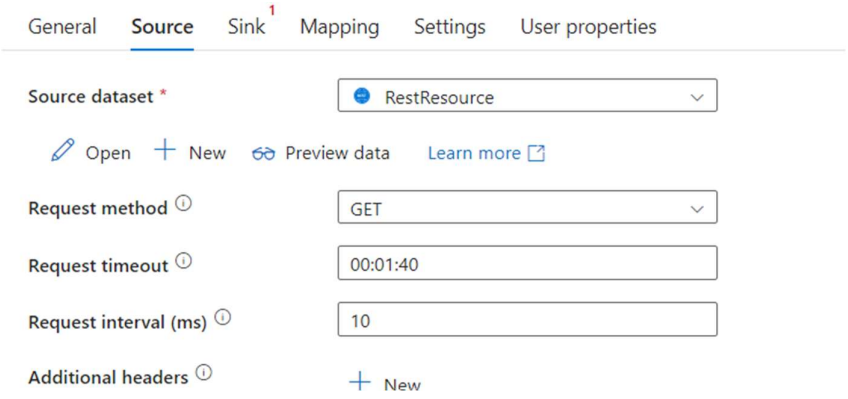
URL: https://swapi.dev/api/people



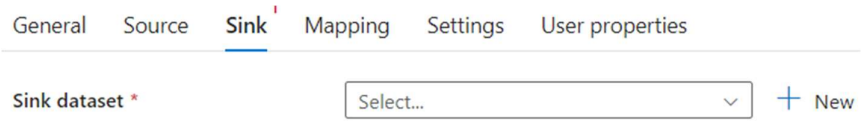
8. After creating a linked service select it and click on Ok.



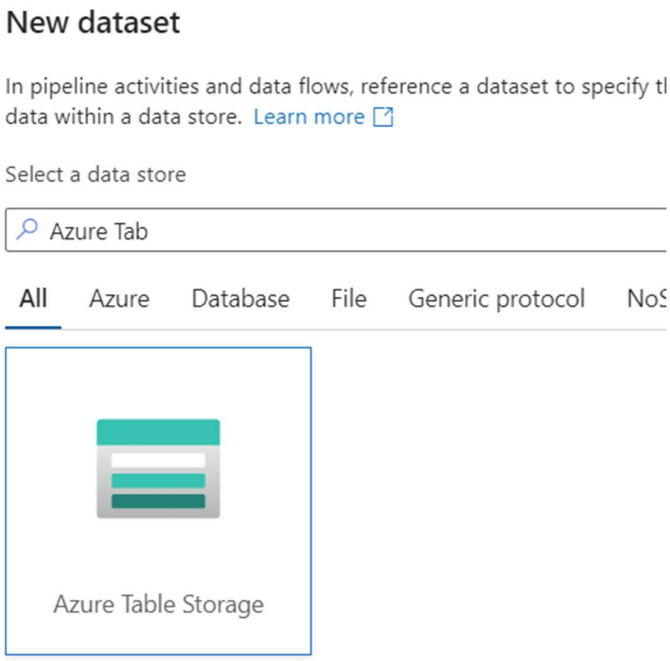
9. We have configured our source part.



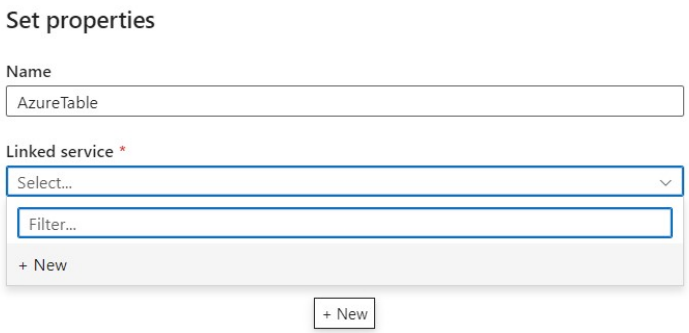
10. Now under Sink click on New.



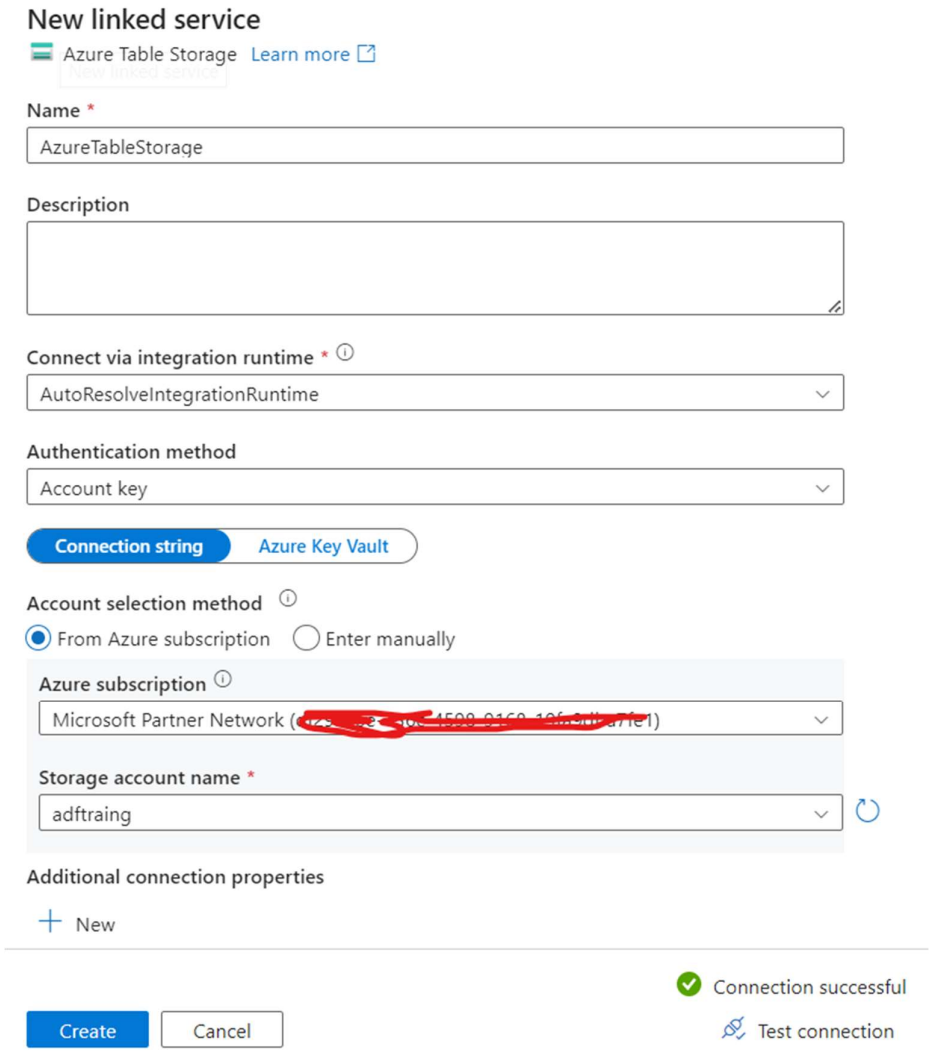
11. Select Azure table Storage and click on continue.



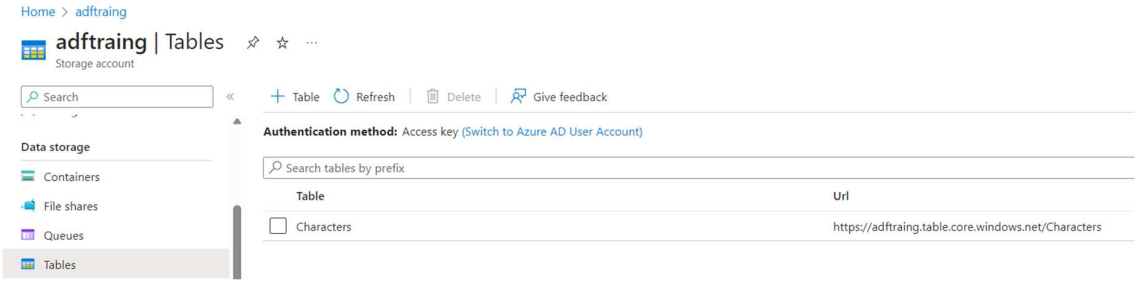
12. Give the Name and click on New.



13. Create a Linked Service with the below properties.

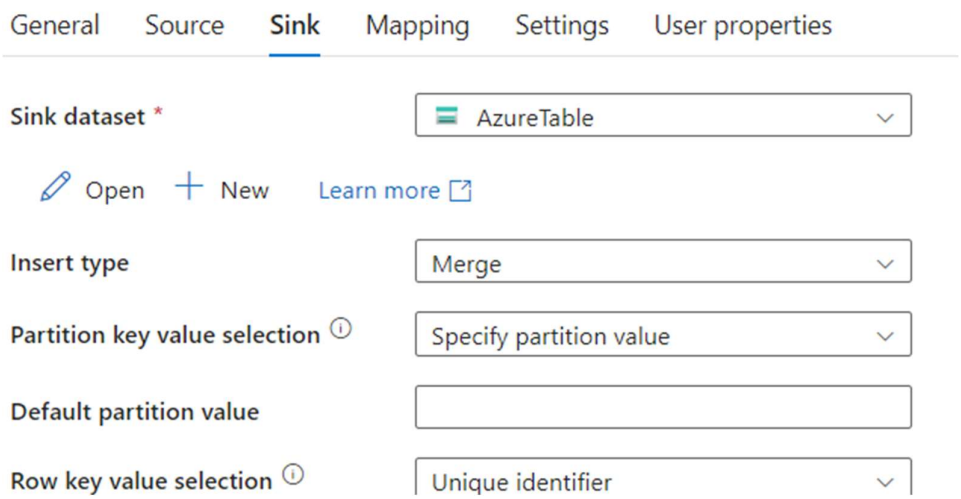


14. Under Storage Account, go to Tables and click on plus symbol then create a Table.



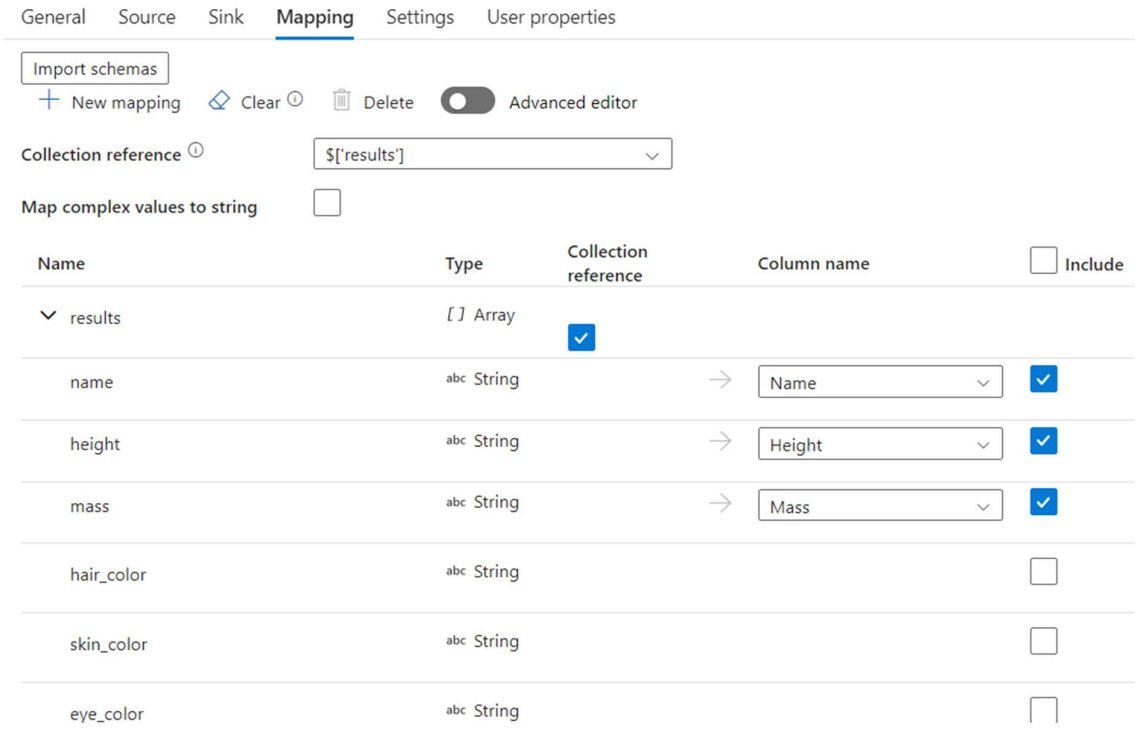
15. Select the Linked Service and select the table and click on Ok.

16. We have configured our Sink part.

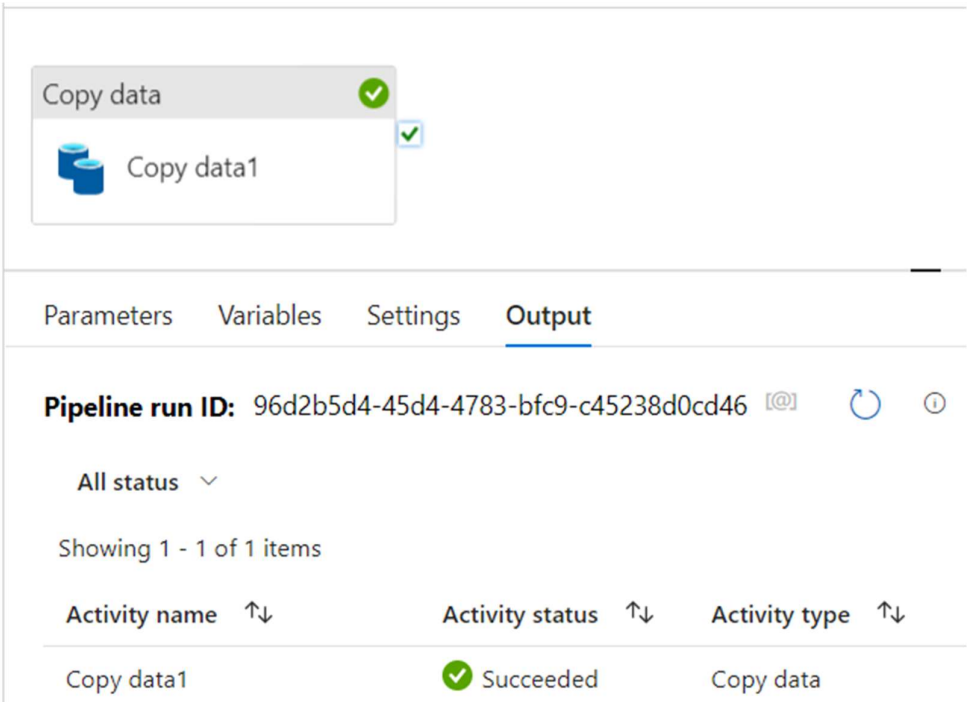


17. Under Mapping click Import schemas.
18. After a few seconds we should see the format of the API response appear.
19. Notice the response layout includes properties such as count, next, previous, and results. ADF intuitively knows that the results property is an array.
20. Next, we'll set our Collection reference to the array named results. To do this, either select it from the drop-down list or check the box under the collection reference column.
21. Then we set up the Column names for the items we're mapping for import.
22. We'll map the fields: name, height, mass, and homeworld.
23. Uncheck the Include column for everything else.

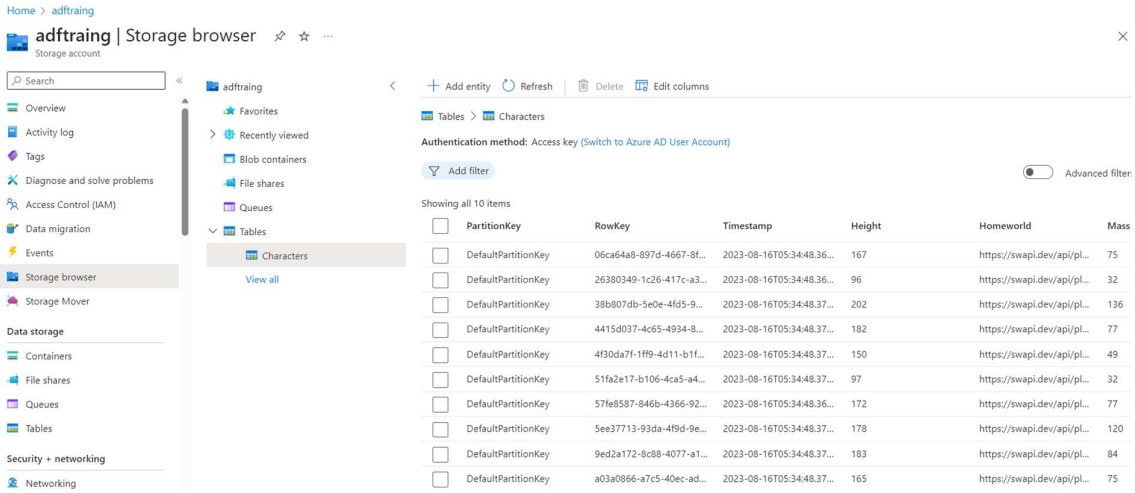
24. Finally, type in the name we want our columns to be called (name, height, mass, homeworld.)



25. Next validate the Pipeline and Click on Debug.
26. Here our pipeline was executed successfully.

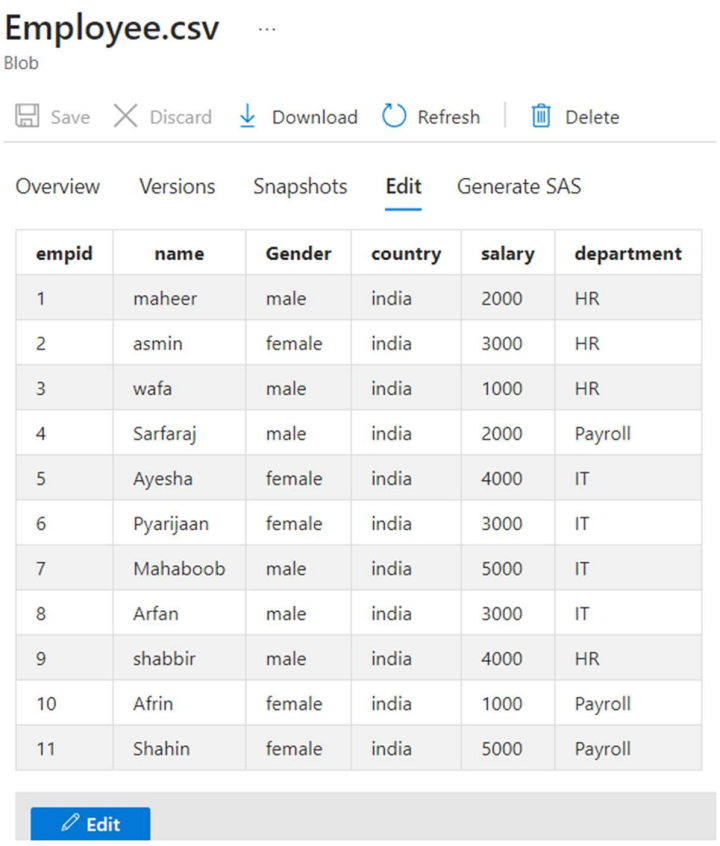


27. We built our ADF pipeline with a task to copy data from an external API and save the results into Azure Table Storage.

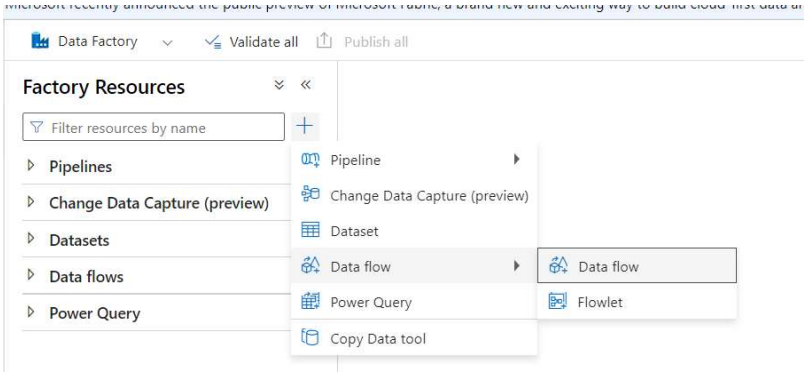


Parameterize

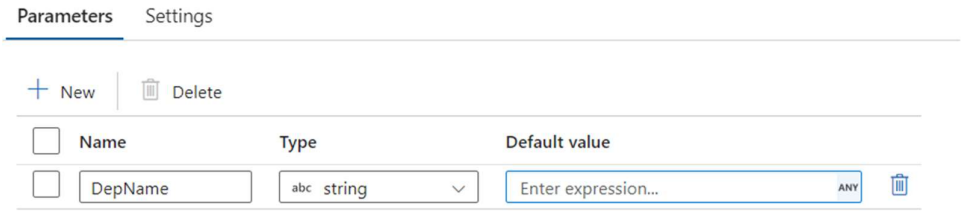
1. In this Example we are using Employee data.



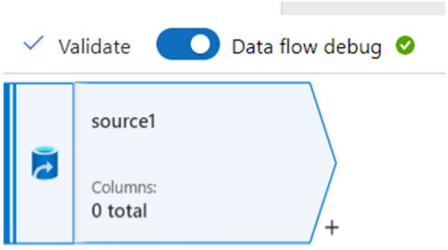
2. Create a Data flow.



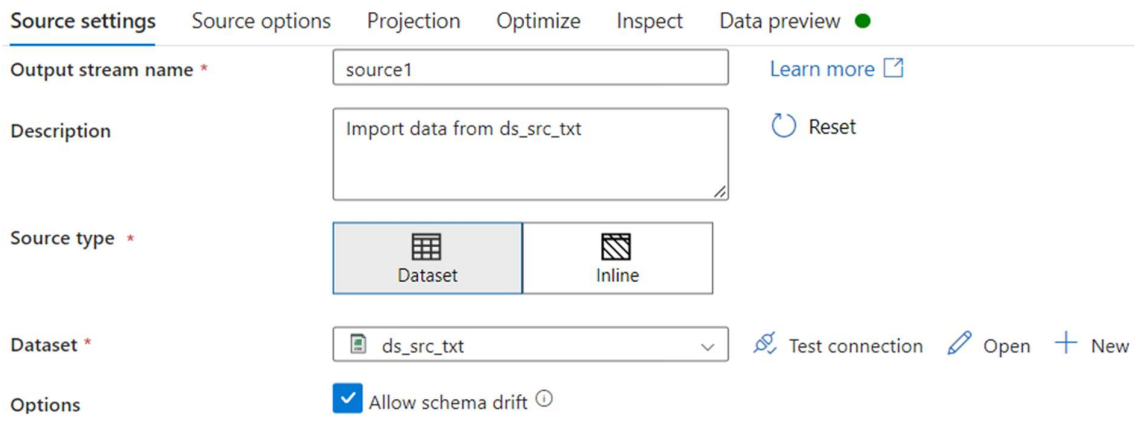
3. Create the below parameter.



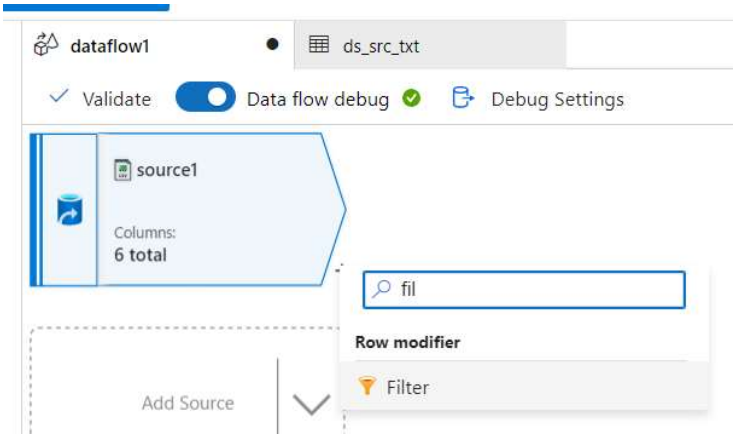
4. Add a Data Source.



5. Under source select the source dataset.

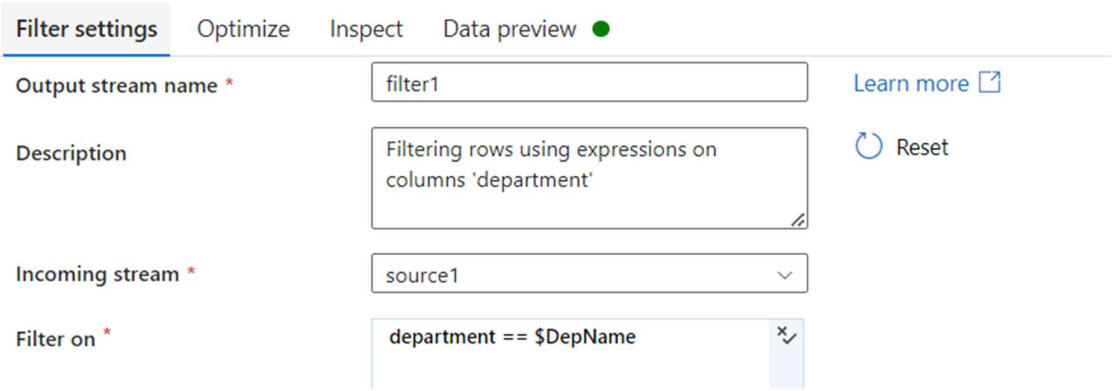


6. Click on the plus symbol and click on the filter.

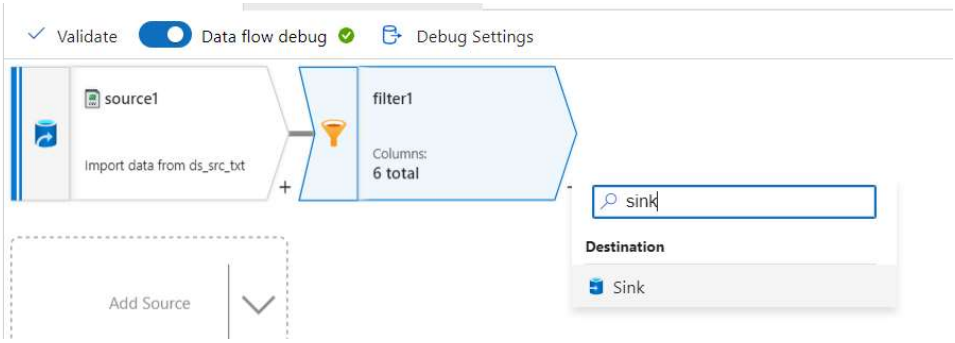


7. Under filter give the below condition.

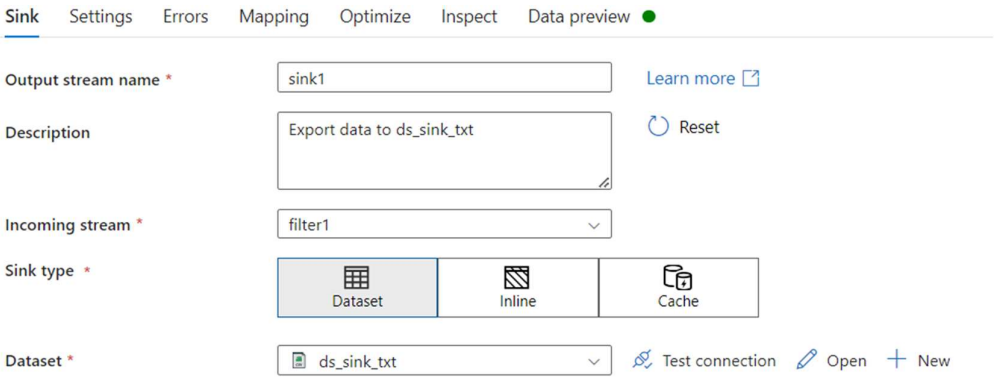
Con: department == $DepName



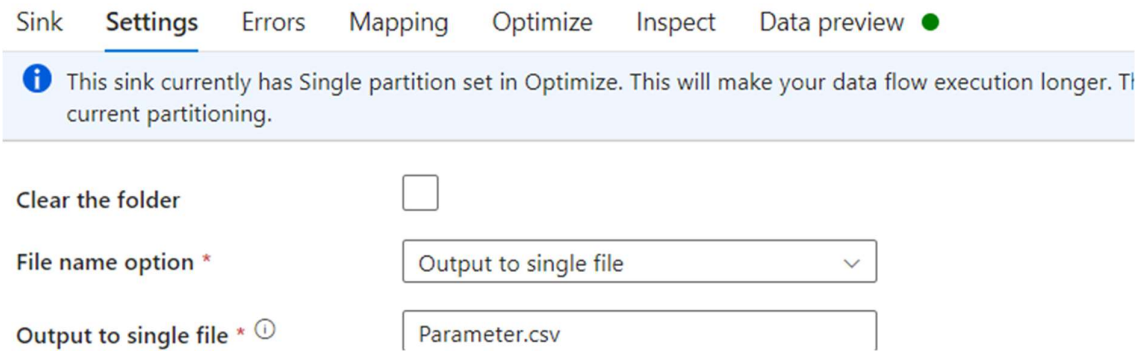
8. Click on the Plus symbol and click on Sink.



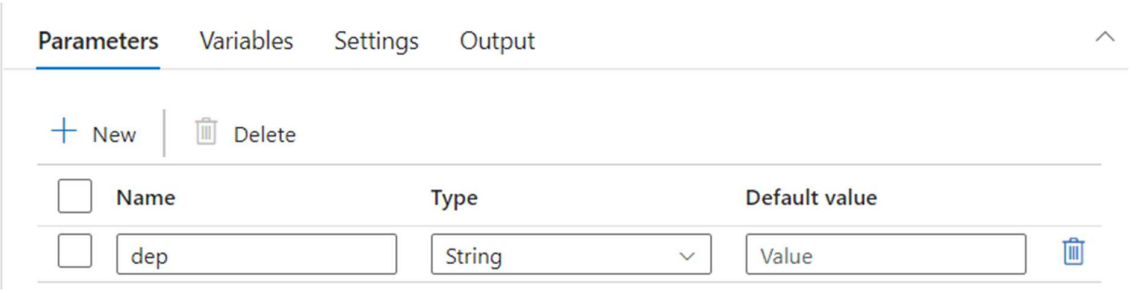
9. Under select the destination dataset.



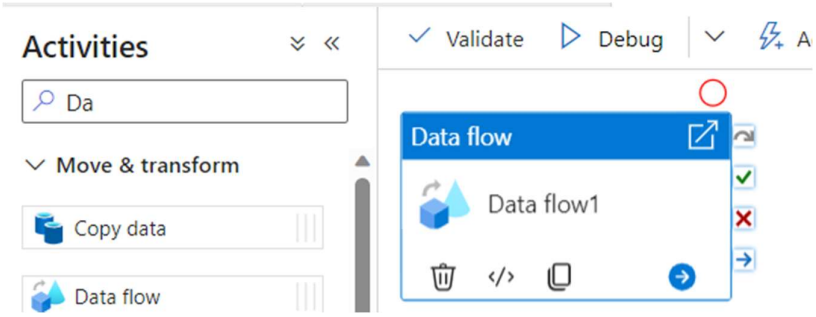
10. Under setting select the file option and give the filename.



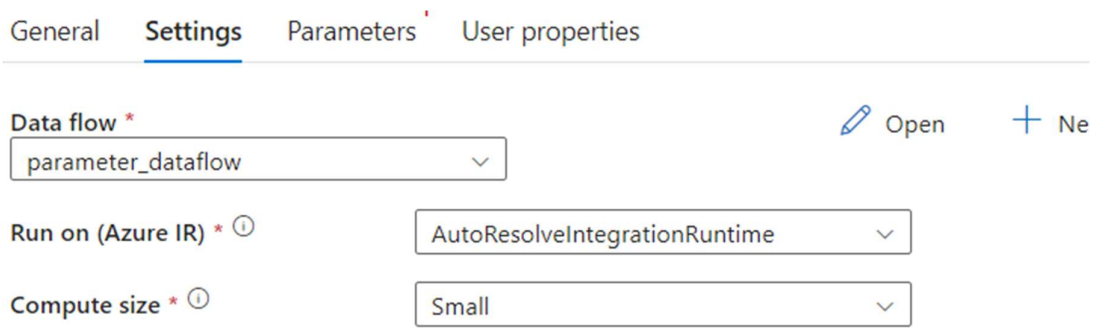
11. Create a Pipeline and create a parameter as shown below.



12. Drag and drop the Data flow.

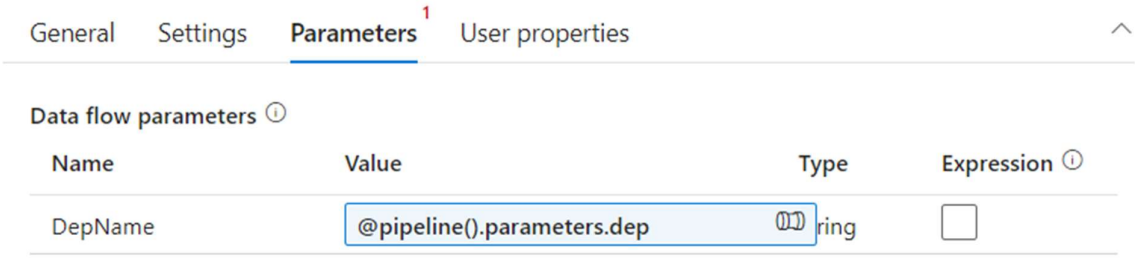


13. Under setting select the dataflow.

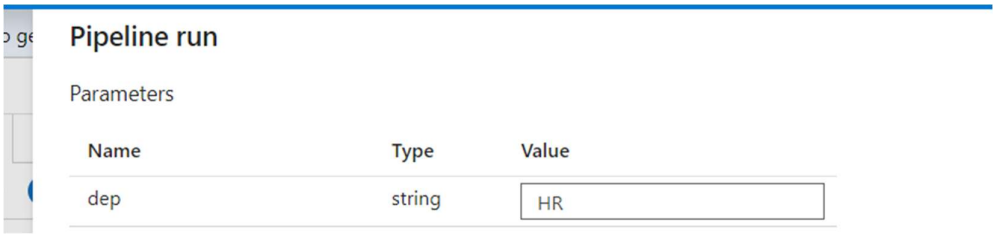


14. Under parameter click on the pipeline expression and give the below expression.

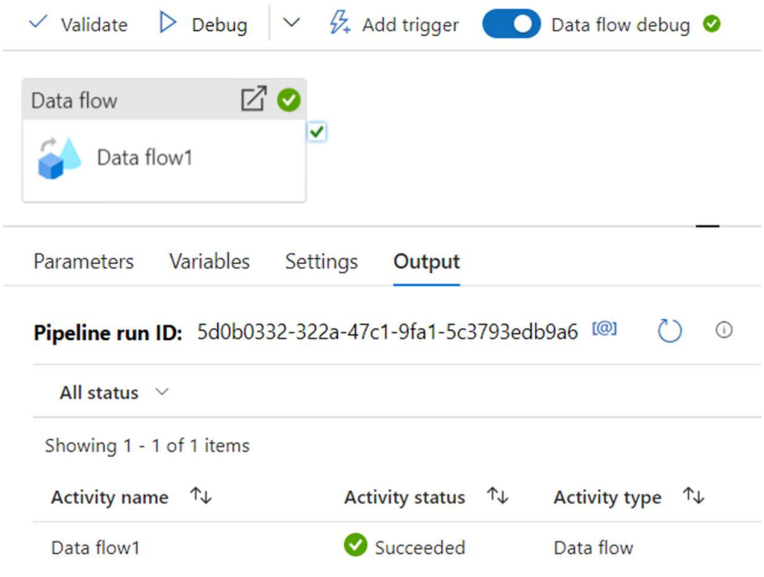
Exp: @pipeline().parameters.dep



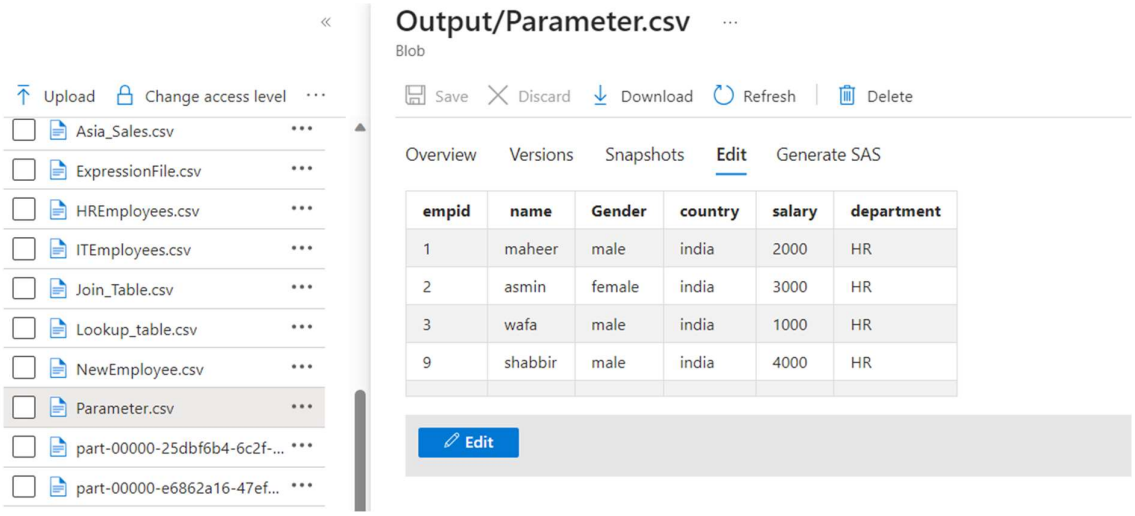
15. Now validate the pipeline and click on debug. Then it will ask for the parameter value and give it as HR.



16. Here our pipeline was executed successfully.

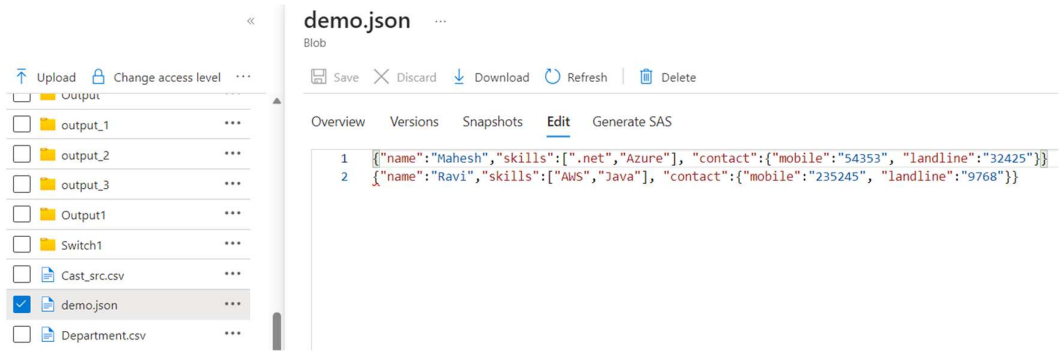


17. Go to Destination and check the data.

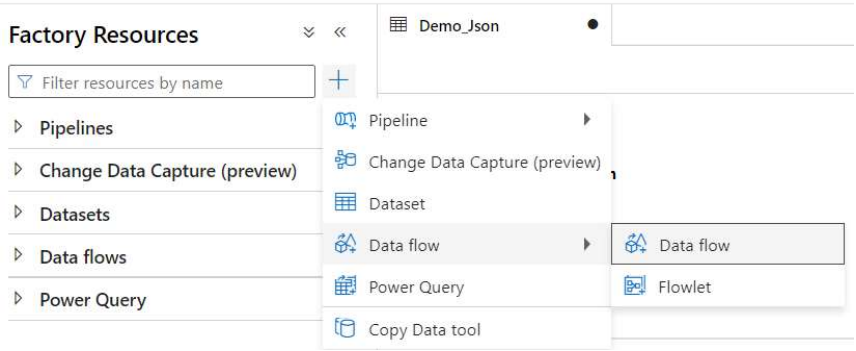


Stringify Transformation

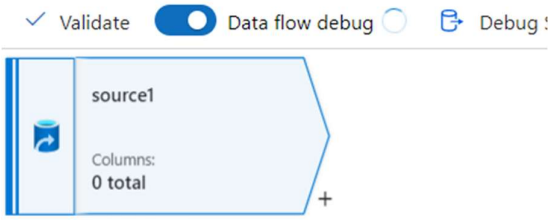
1. In this example we are going to use below json data.



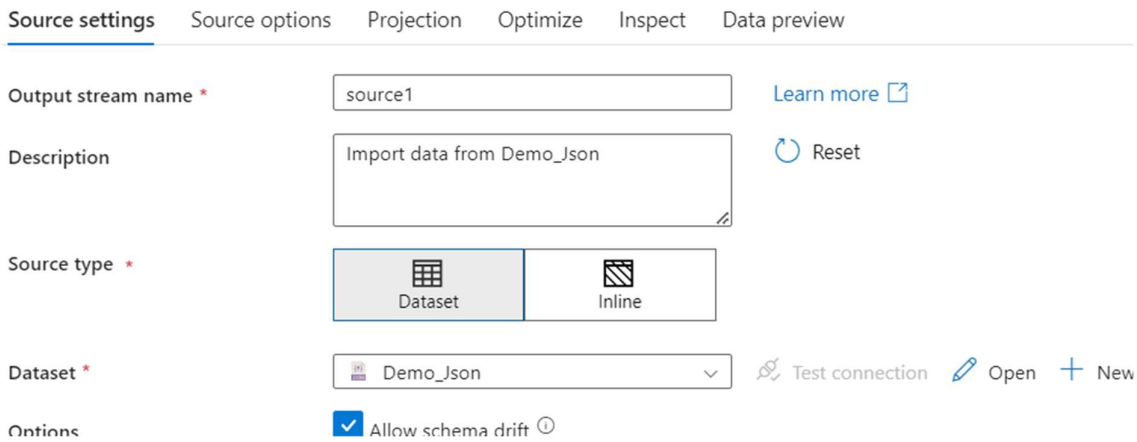
2. Create a Dataset for the above json file.
3. Create a Data flow.



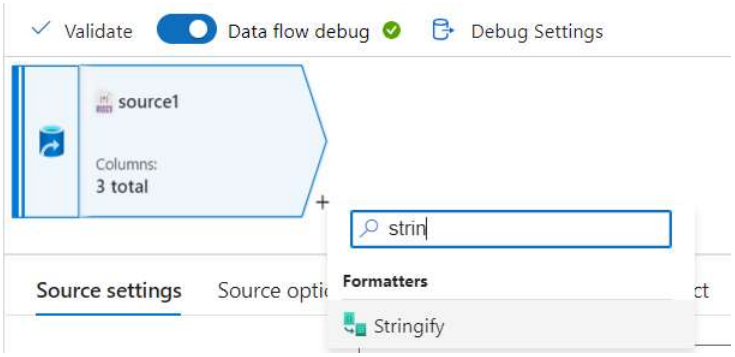
4. Add a Data source.



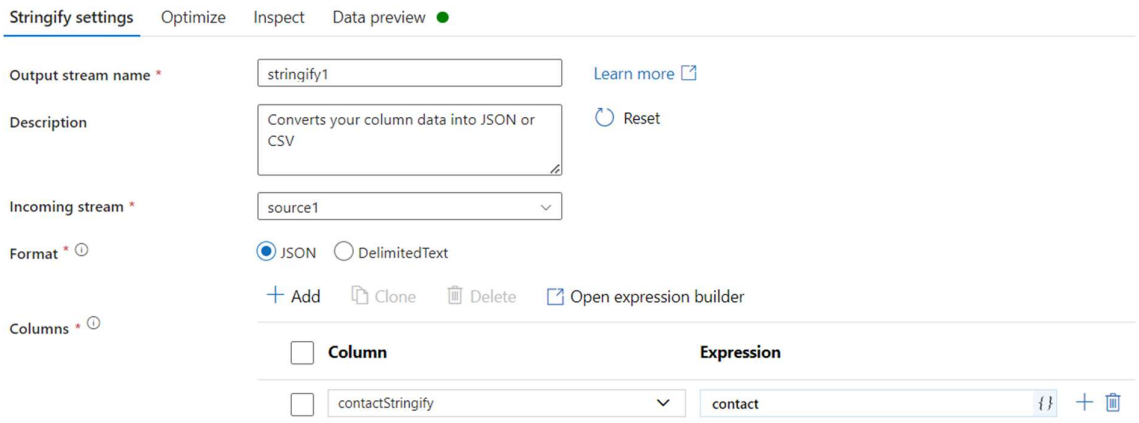
5. Select the Source dataset.



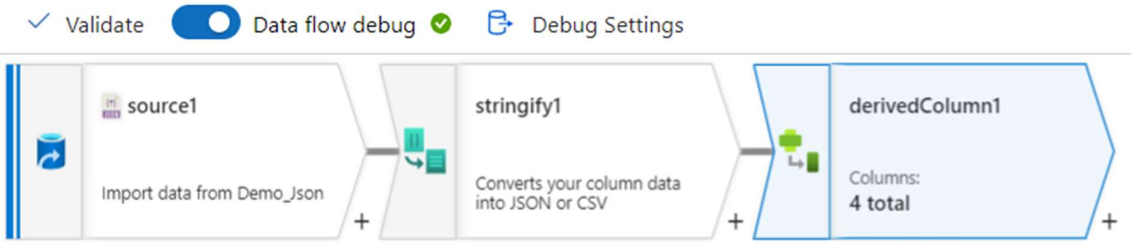
6. Click on the plus symbol and click on Stringify.



7. Under settings give the new column name and give the below expression as contact.

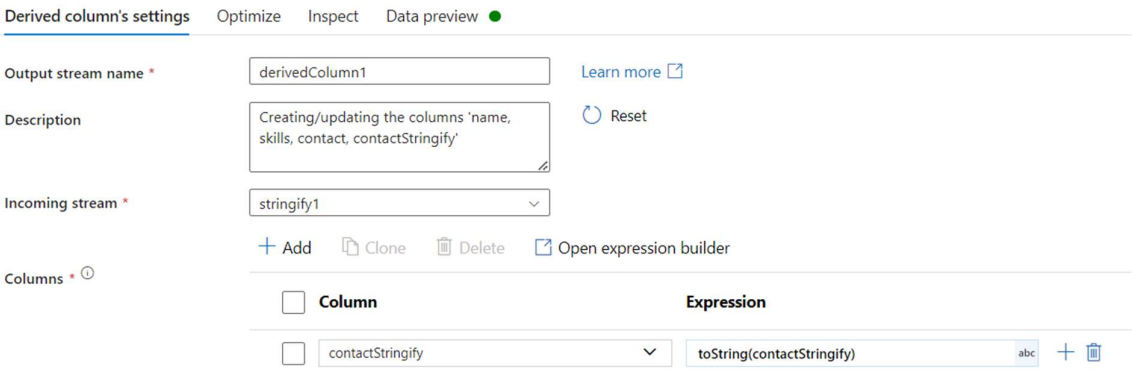


8. Add derived column.

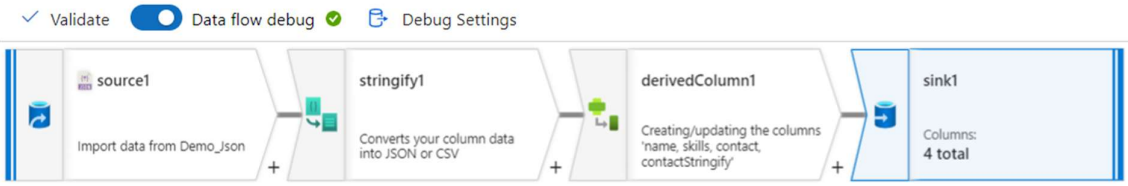


9. Under settings select the column and give the below expression.

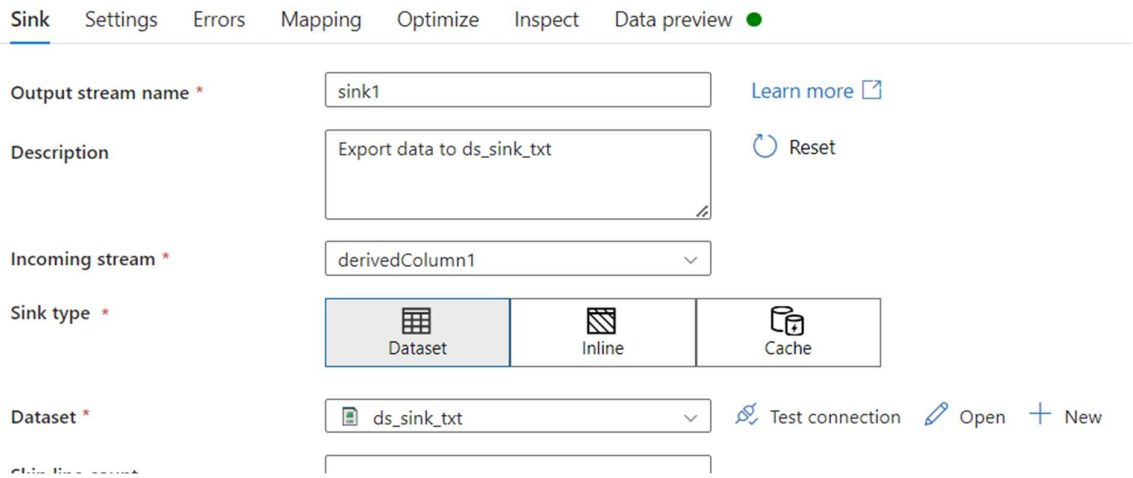
Exp: toString(contactStringify)



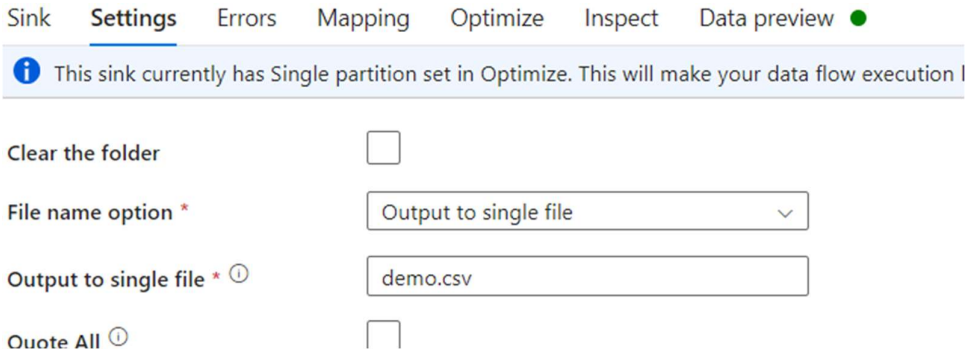
10. Add Sink.



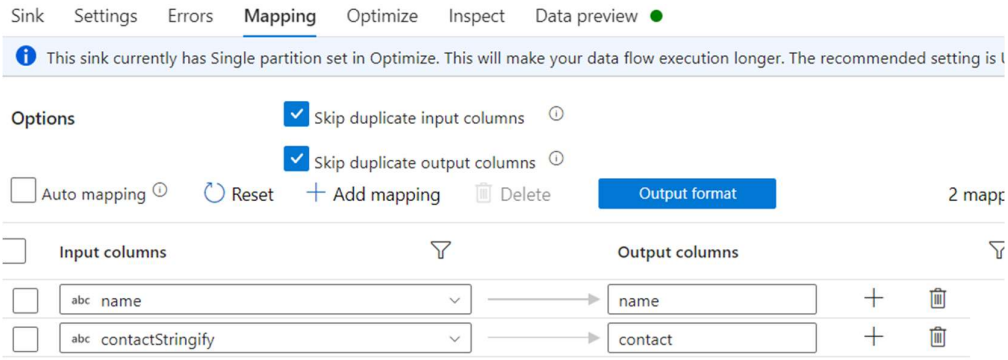
11. Add the target dataset.



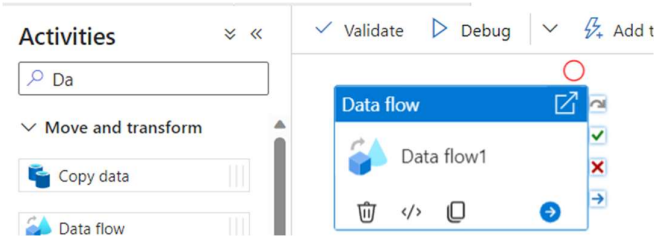
12. Under settings select the file option and give the file name as shown below.



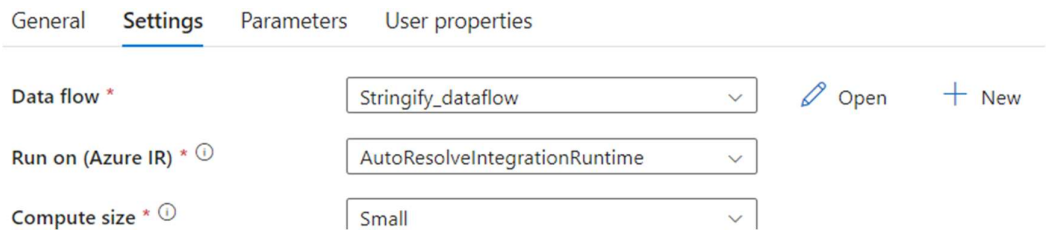
13. Under mapping delete the fields except below two.



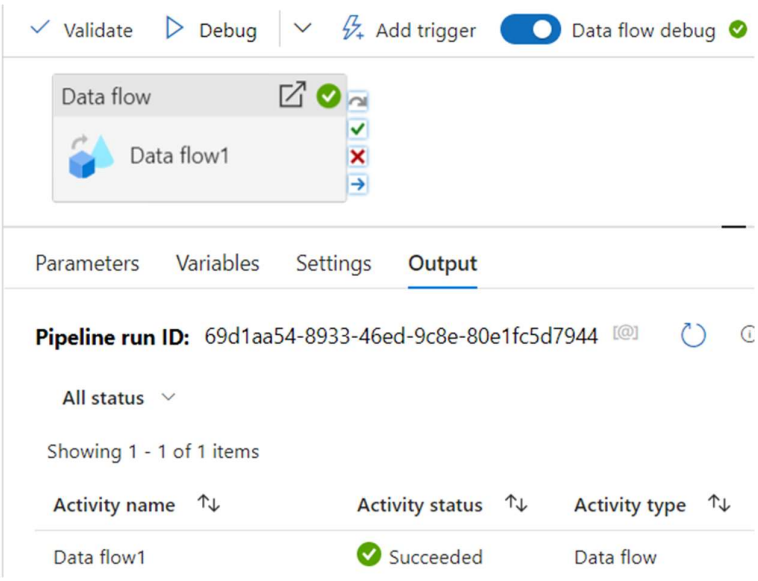
14. Create a Pipeline and Drag and drop the Data flow.



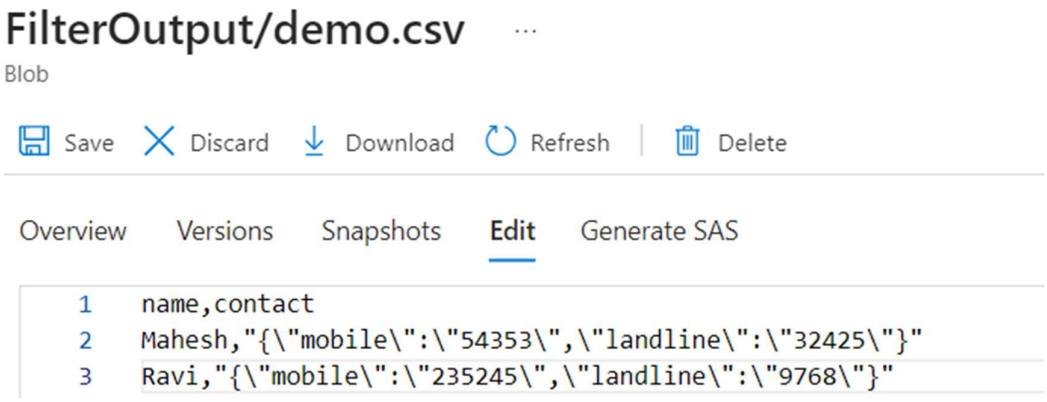
15. Under settings select the Data Flow.



16. Validate the Pipeline and click on Debug.
17. Here our pipeline executed successfully.

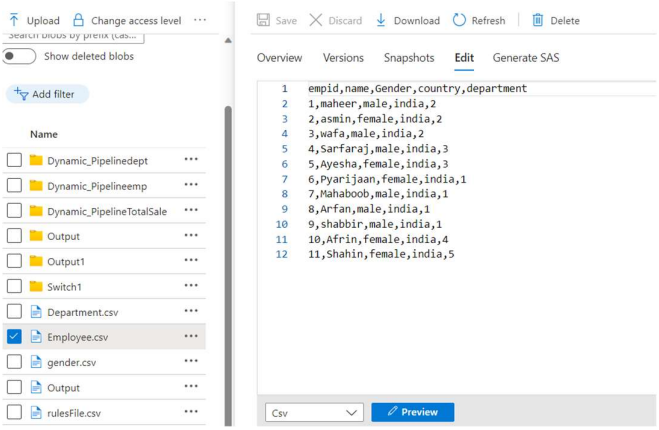


18. Go to Destination and check the Data.

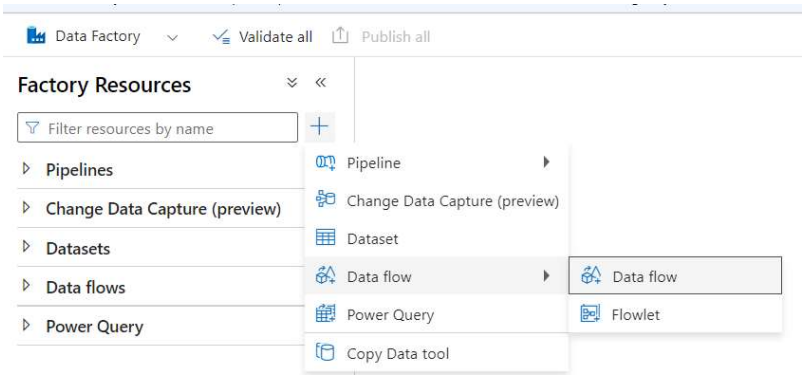


Pivot Transformation

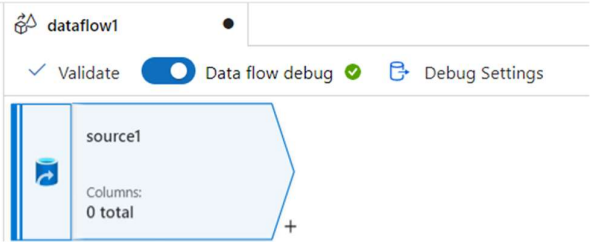
1. In this example we are going to use below employee table as shown below.



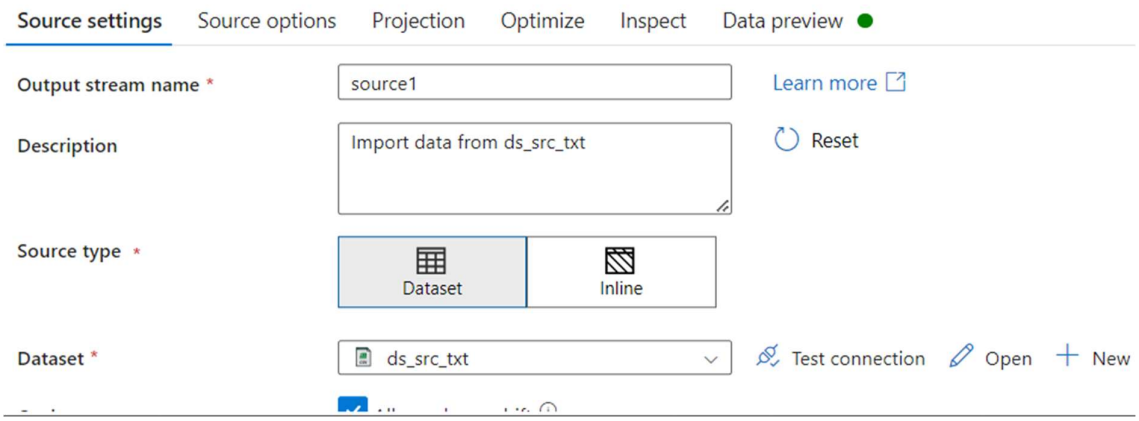
2. Go to Azure Data Factory and create a Data flow.



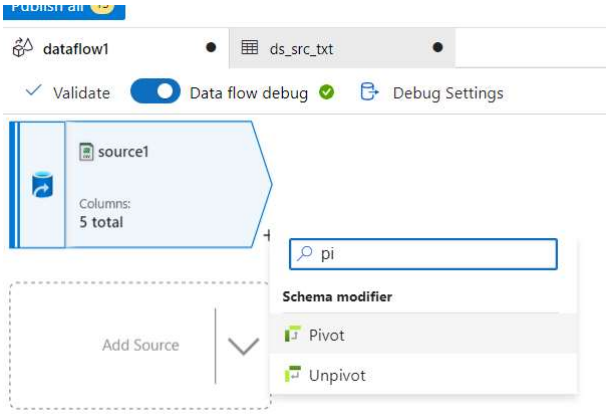
3. Add a Data Source.



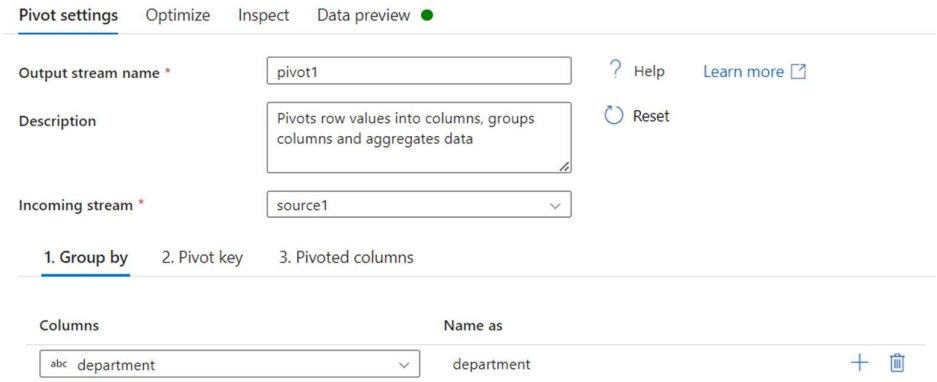
4. Select the Employee data set as shown below.



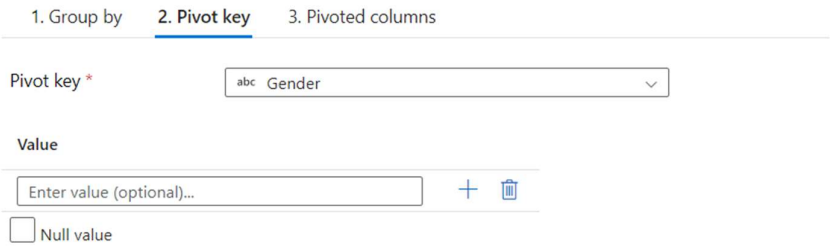
5. Now click on the plus symbol and click on Pivot as shown below.



6. Under Pivot Setting for the Group by select Department.

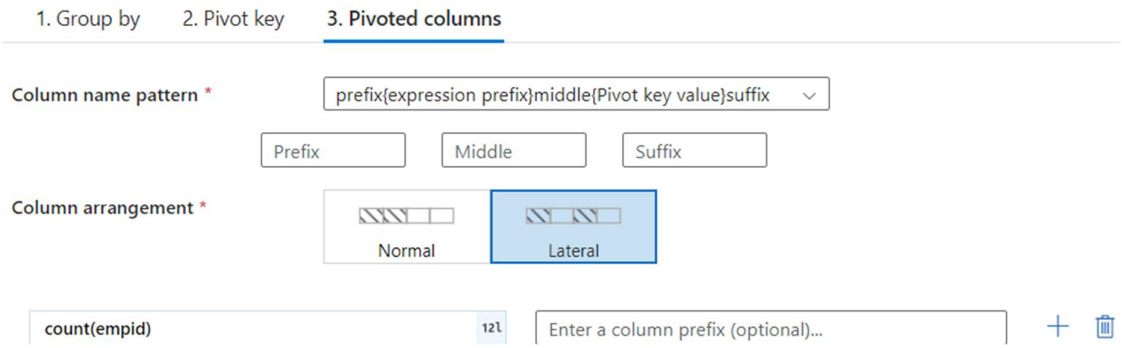


7. For the Pivot key select Gender as shown below.

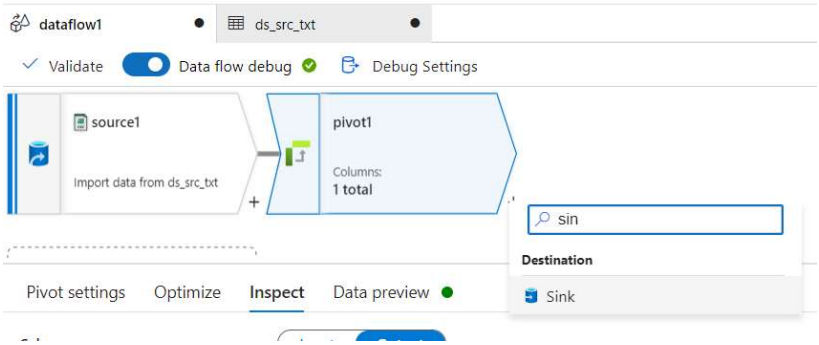


8. Under Pivoted columns give the below expression.

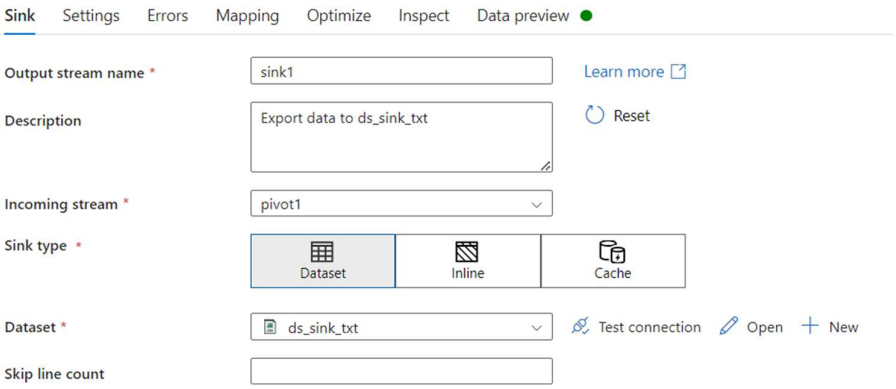
Exp: count(empid)



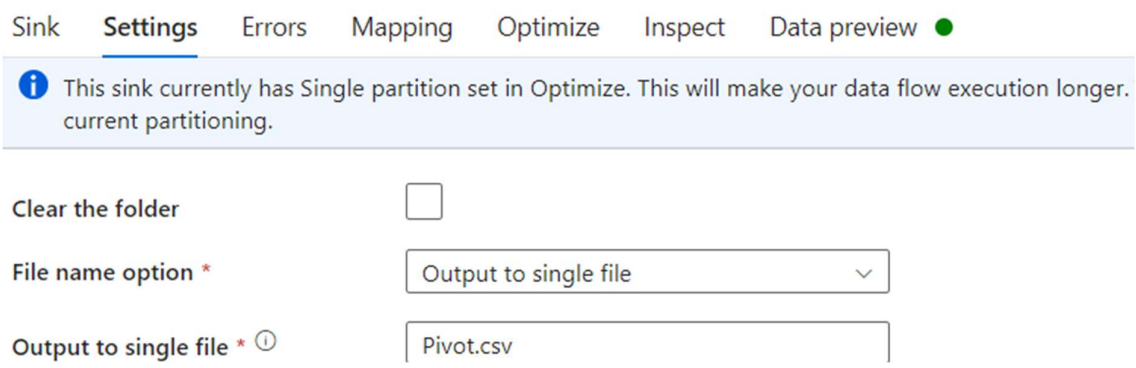
9. Now click on the Plus symbol and click on Sink.



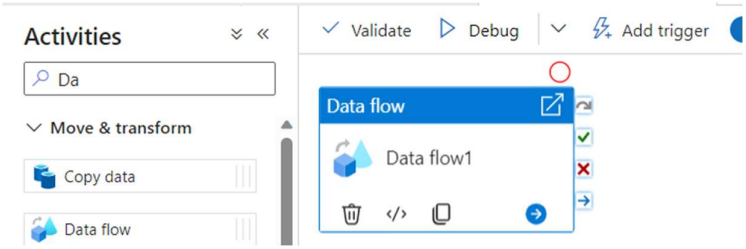
10. Under sink select the destination dataset.



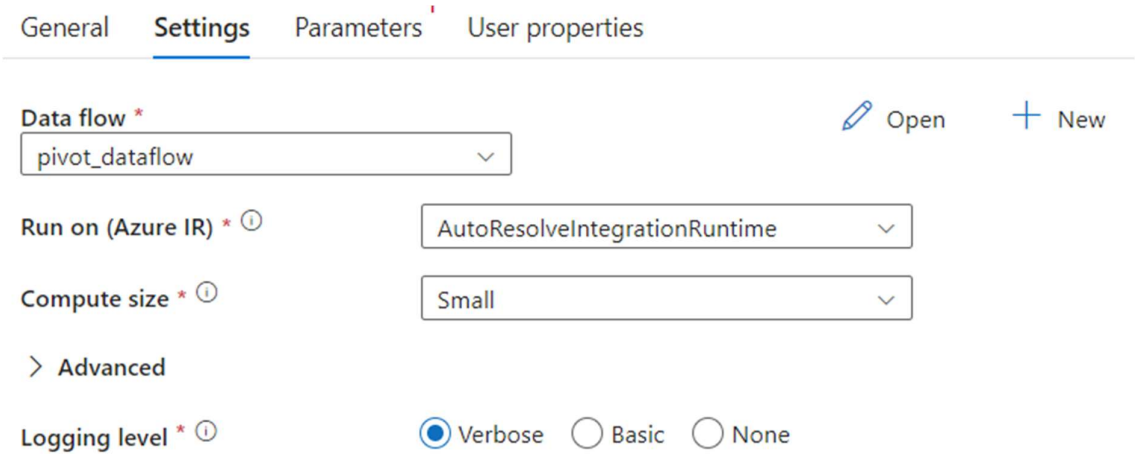
11. Under the Setting set the file option and file name.



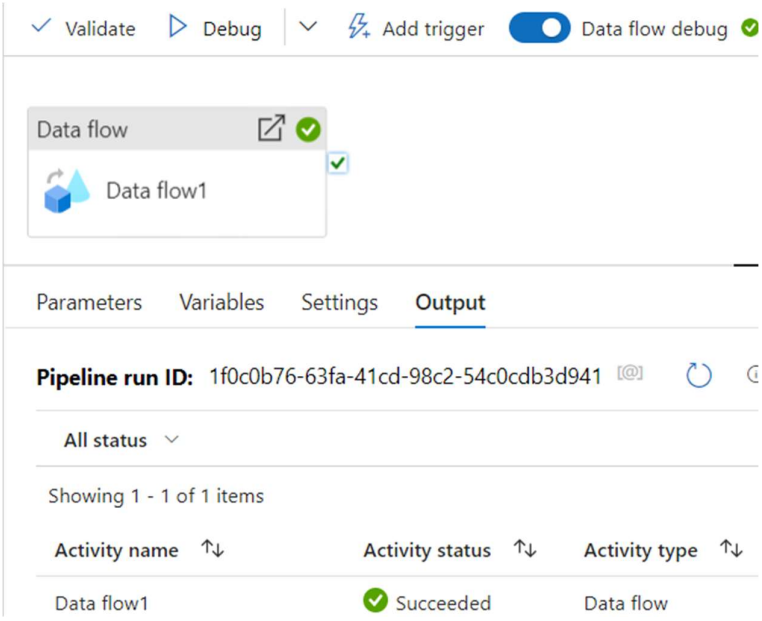
12. Create a Pipeline and drag and drop the Data flow as shown below.



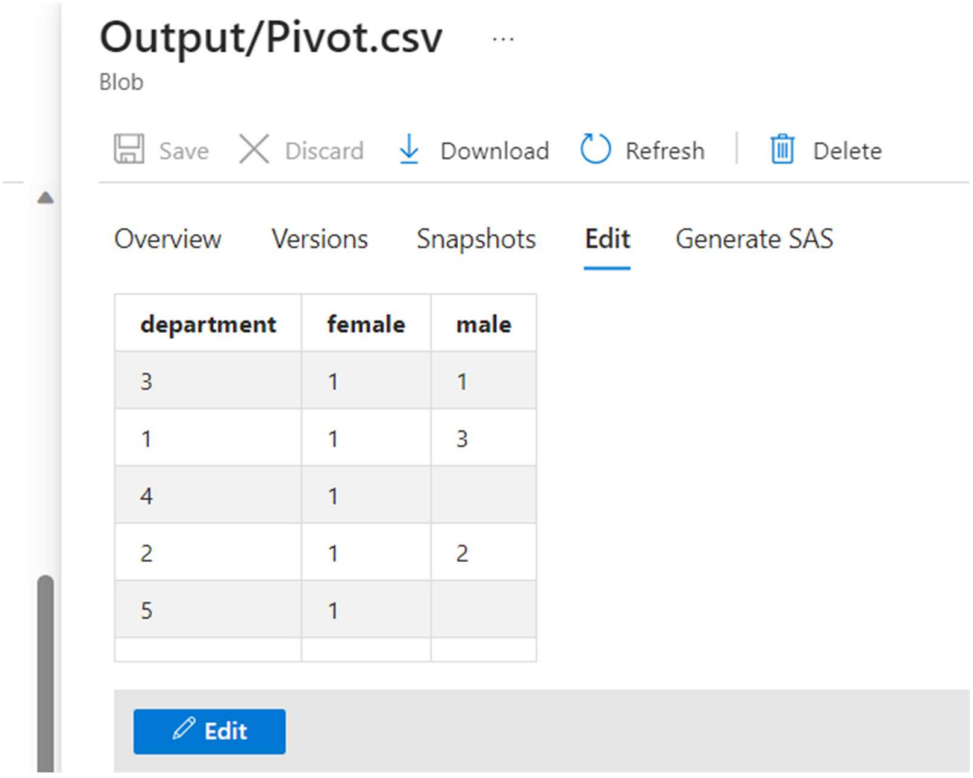
13. Select the Data flow.



14. Validate and click on debug.
15. Here our pipeline was executed successfully.

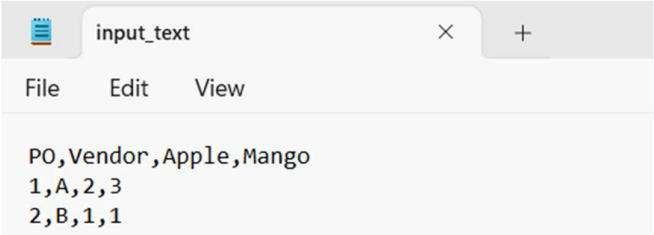


16. Now go to Destination and check the data.

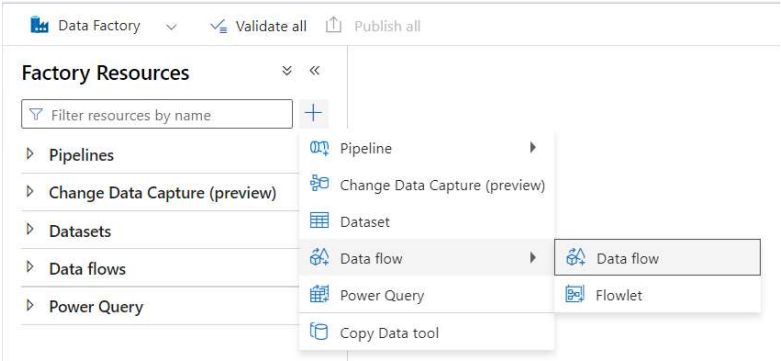


Unpivot Transformation

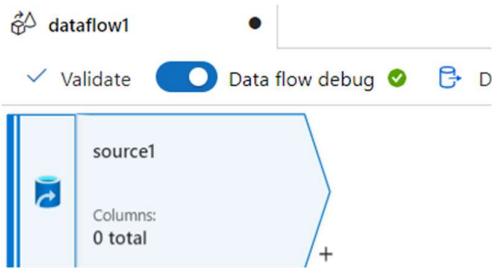
1. In this example we are going to use the below data.



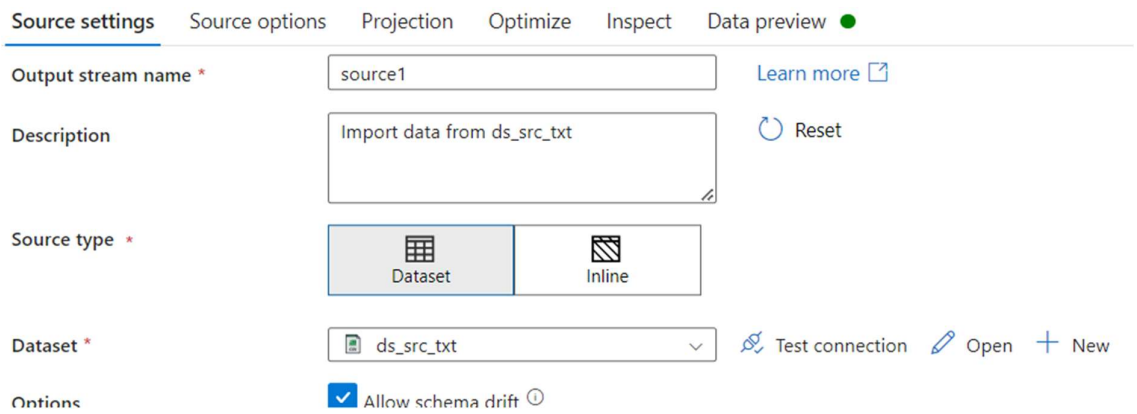
2. Upload the file into the container and create a Dataset for this data.
3. Next create a Data flow.



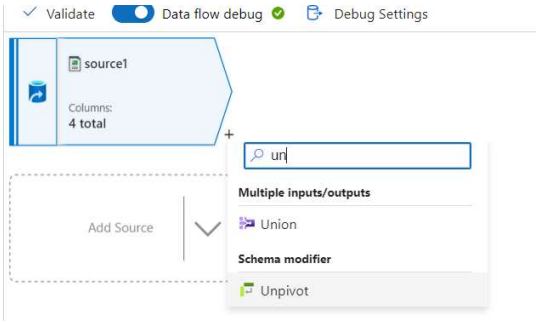
4. Add a Data Source.



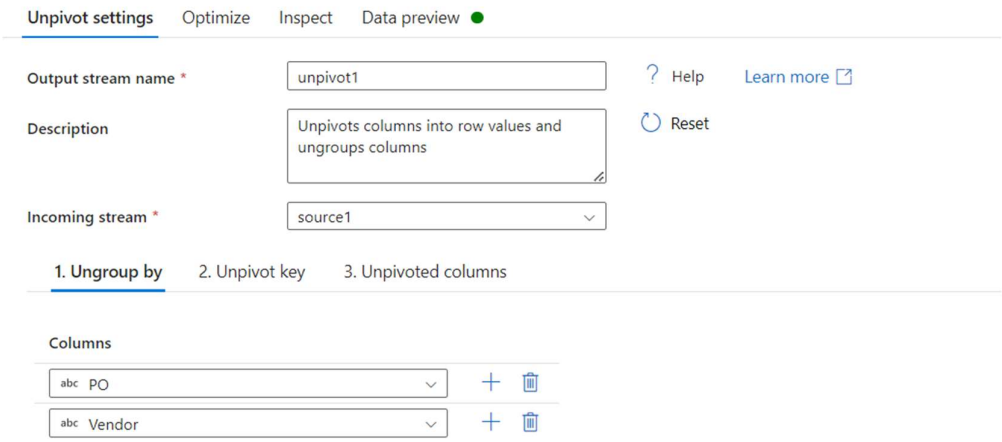
5. Select the source Dataset.



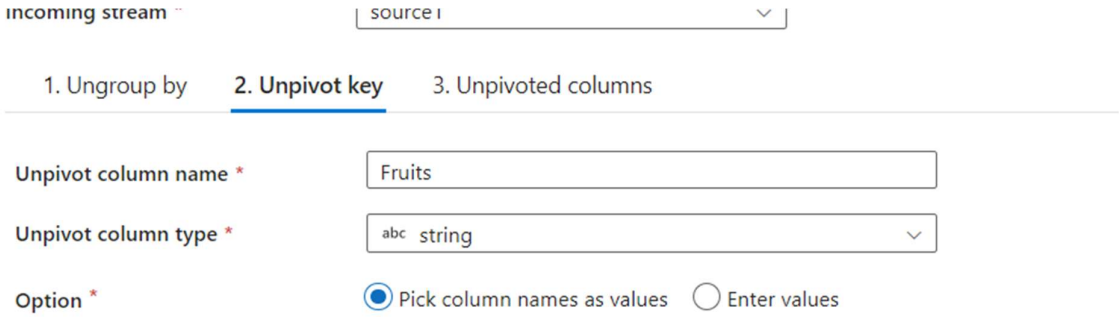
6. Click on the plus symbol and click on Unpivot.



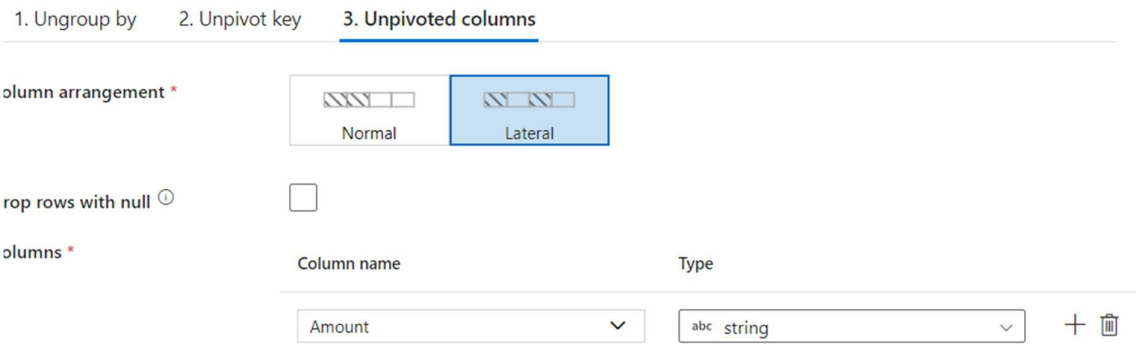
7. Under Unpivot settings, for the Ungroup add the below columns.



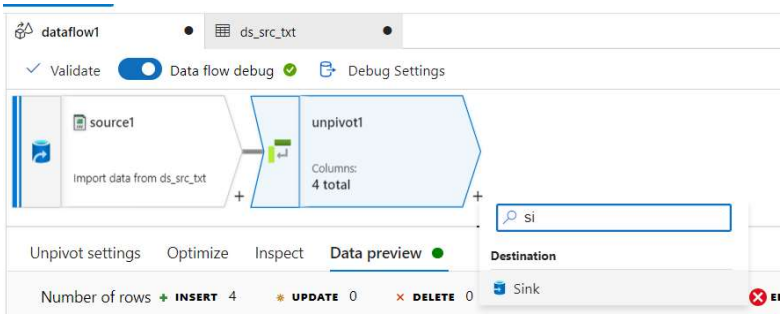
8. For the Unpivot key give the column name and column type as shown below.



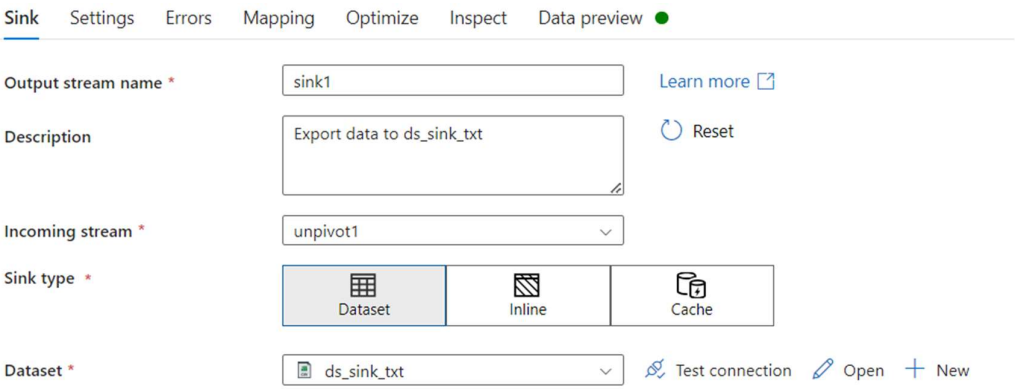
9. For the Unpivoted columns give the column name and column type.



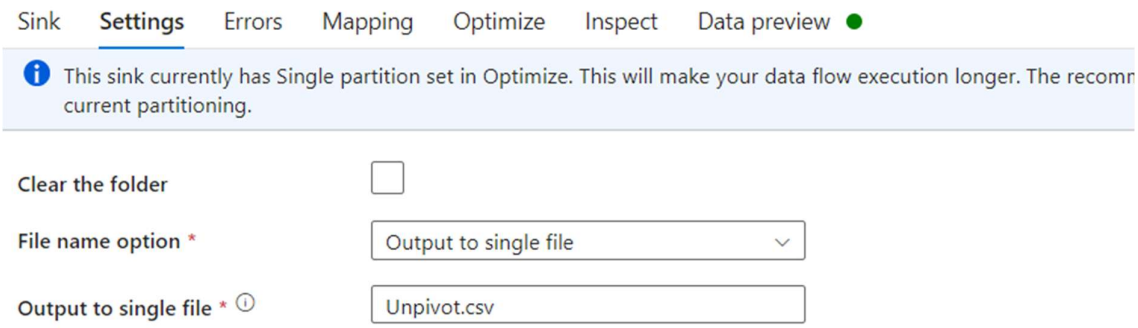
10. Click on the plus symbol and click on Sink.



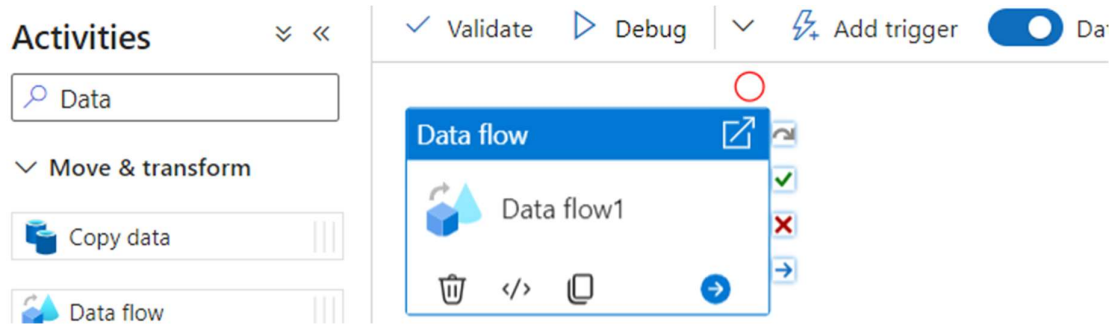
11. Select the Destination dataset.



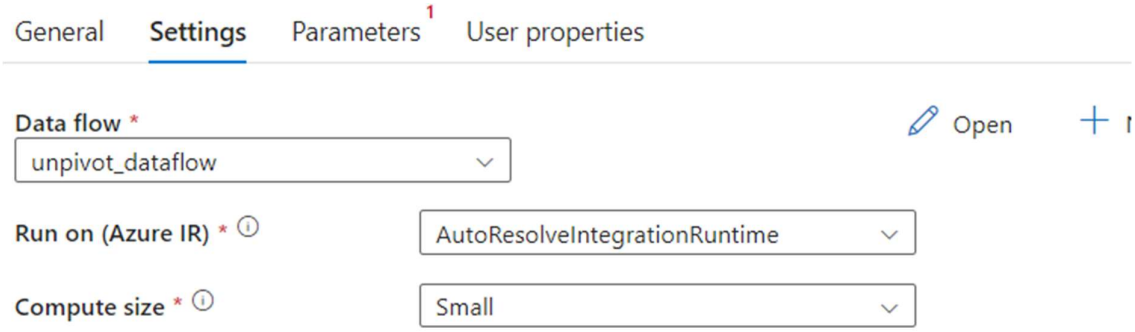
12. Under setting set the file option and file name.



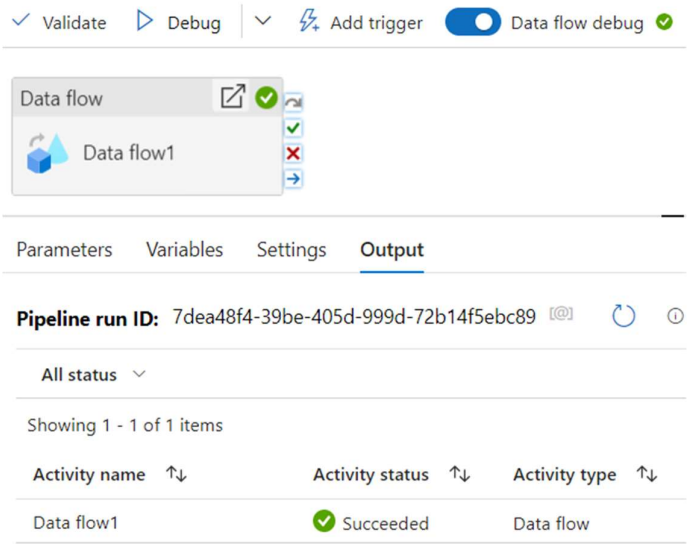
13. Create a Pipeline and drag and drop the data flow.



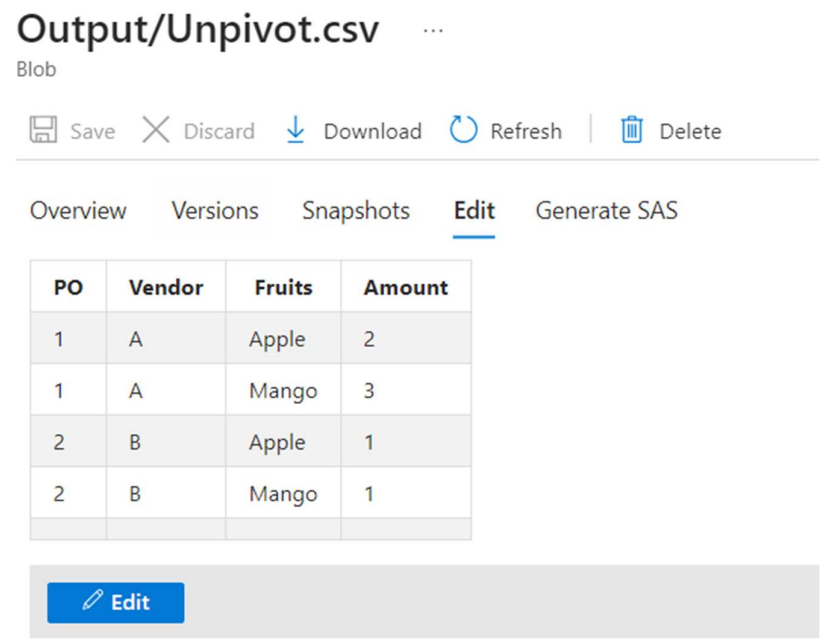
14. Under setting select the dataflow.



15. Now validate it and click on debug.
16. Here our pipeline was executed successfully.

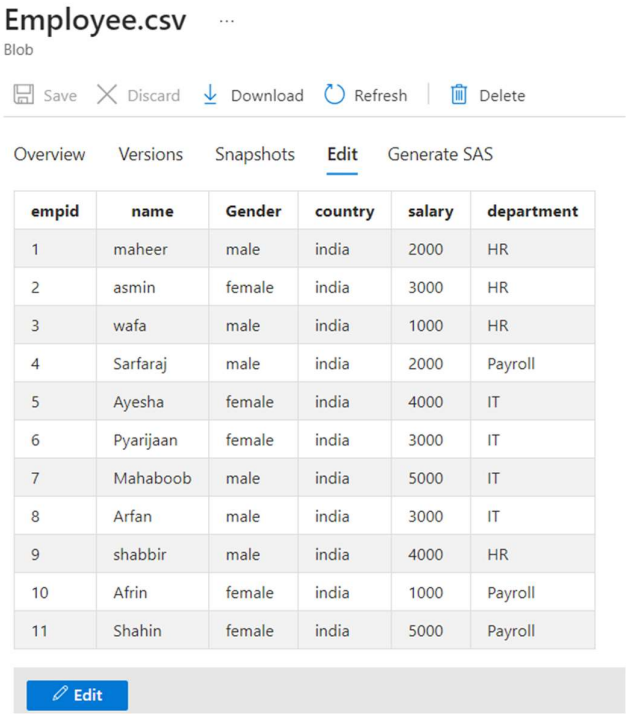


17. Go to Destination and check the data.



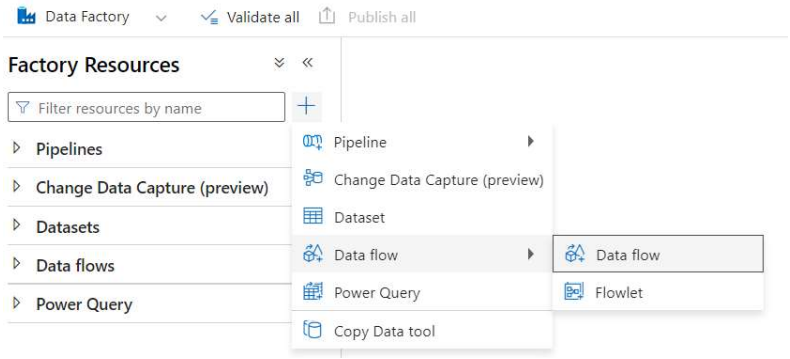
Window Transformation

1. In this example we are going to use the below data.

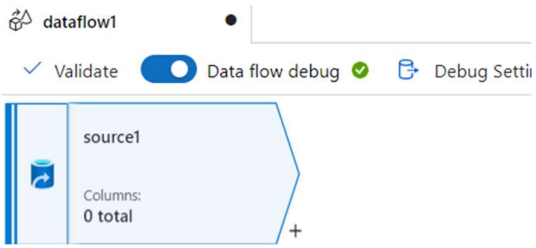


empid	name	Gender	country	salary	department
1	maheer	male	india	2000	HR
2	asmin	female	india	3000	HR
3	wafa	male	india	1000	HR
4	Sarfaraj	male	india	2000	Payroll
5	Ayesha	female	india	4000	IT
6	Pyarijaan	female	india	3000	IT
7	Mahaboob	male	india	5000	IT
8	Arfan	male	india	3000	IT
9	shabbir	male	india	4000	HR
10	Afrin	female	india	1000	Payroll
11	Shahin	female	india	5000	Payroll

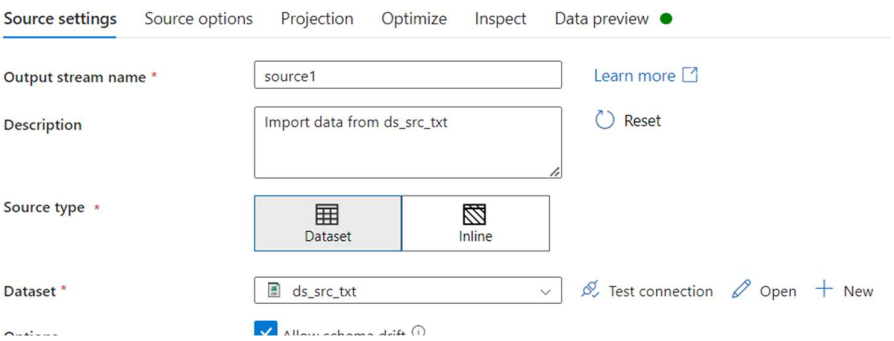
2. Create a Data flow.



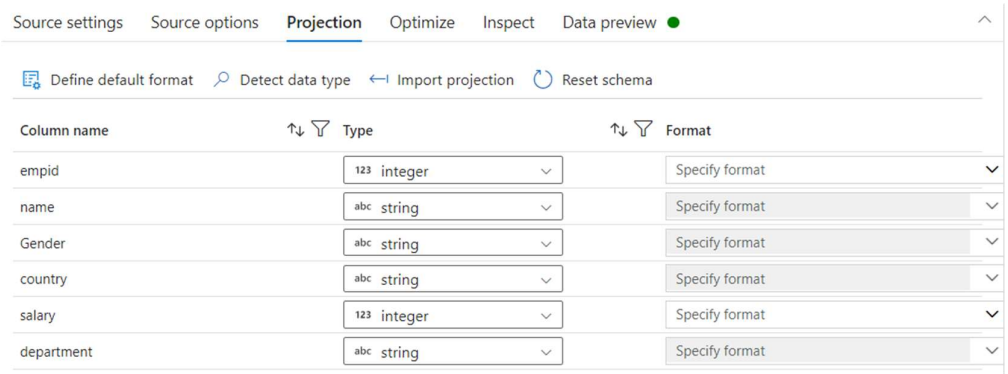
3. Add data source.



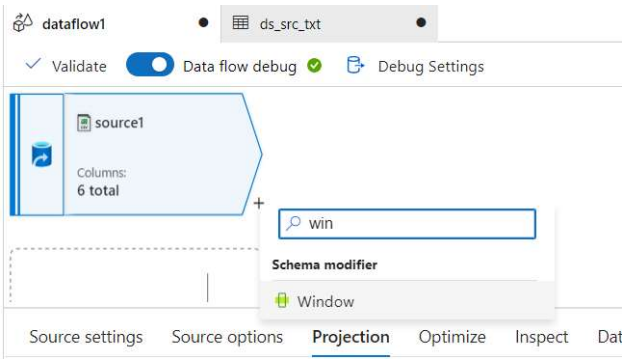
4. Select the source file dataset.



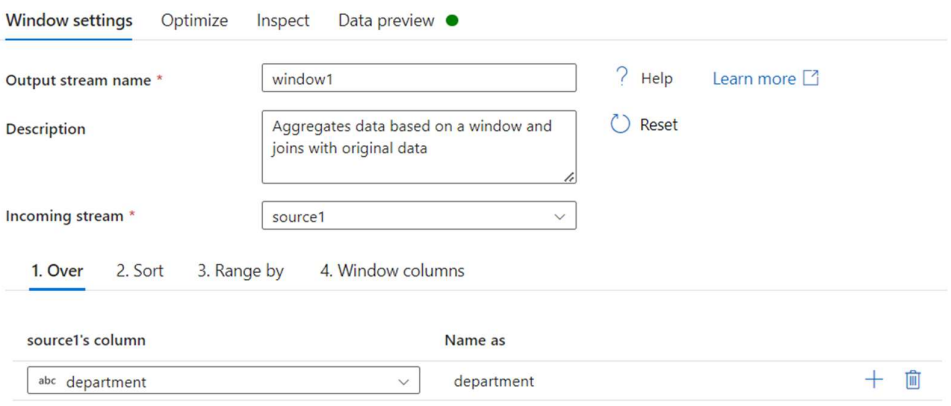
5. Under the Projection set the data types.



6. Click on the plus symbol and click on Window.



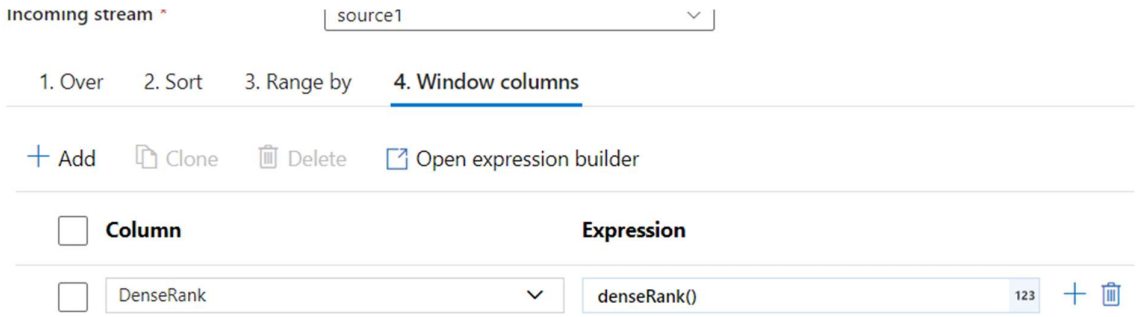
7. Under Setting for the Over set the department column.



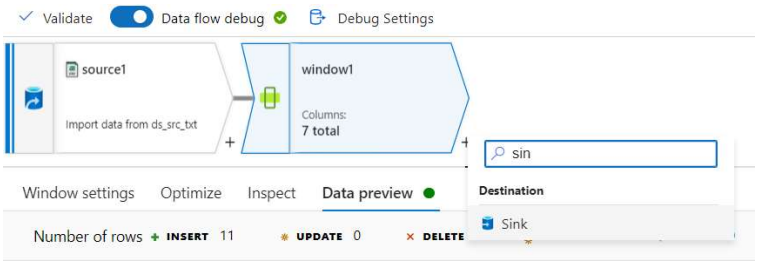
8. For the Sort set the salary column and order as shown below.

9. Under Window columns give the column name and expression.

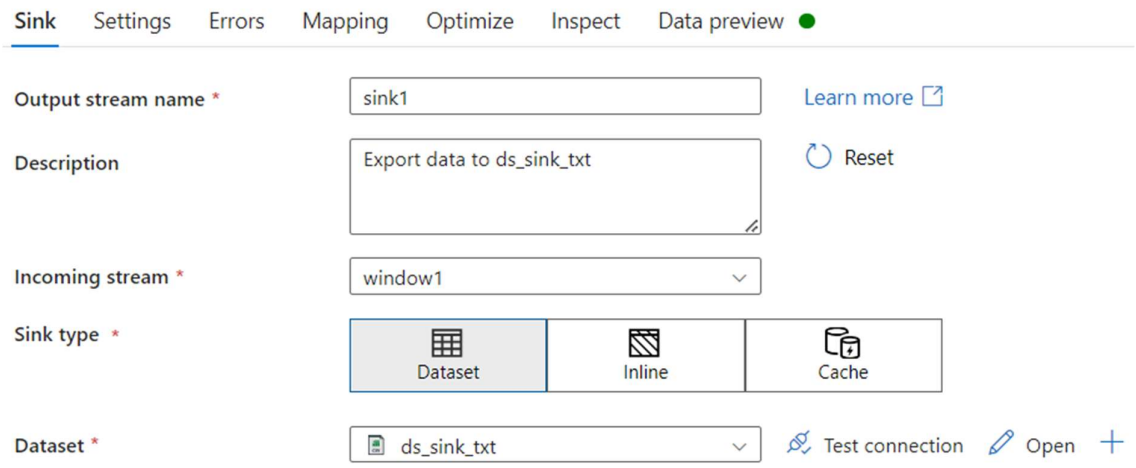
Exp: denseRank()



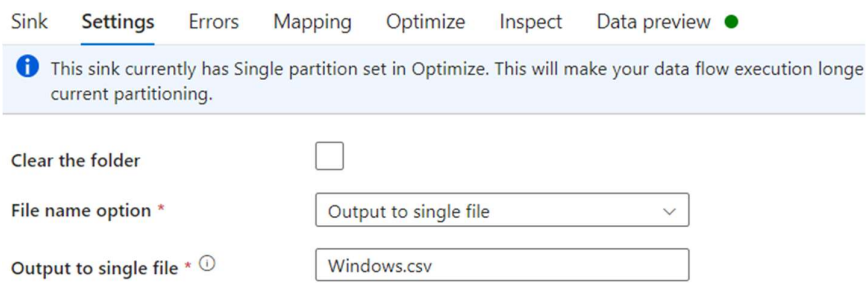
10. Now click on the plus symbol and click on the sink.



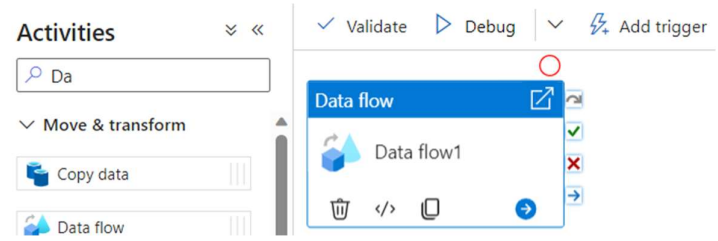
11. Under sink select the destination dataset.



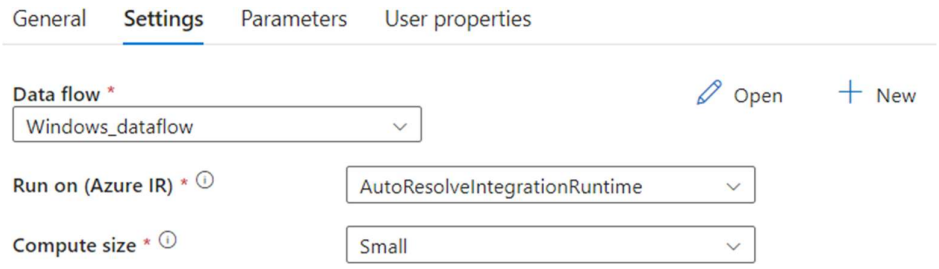
12. Under Settings set the file type and file name as shown below.



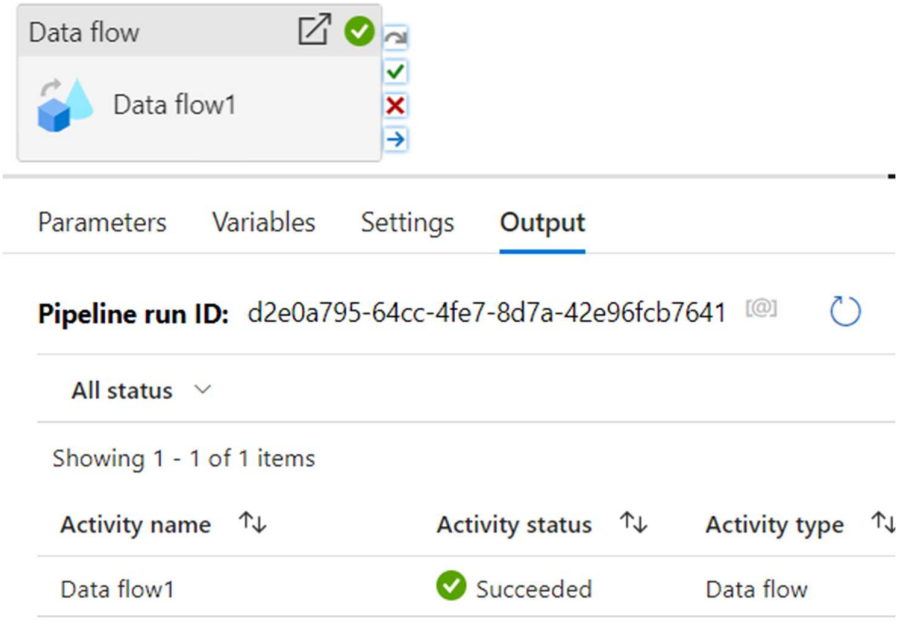
13. Create a Pipeline and drag and drop the Data flow as shown below.



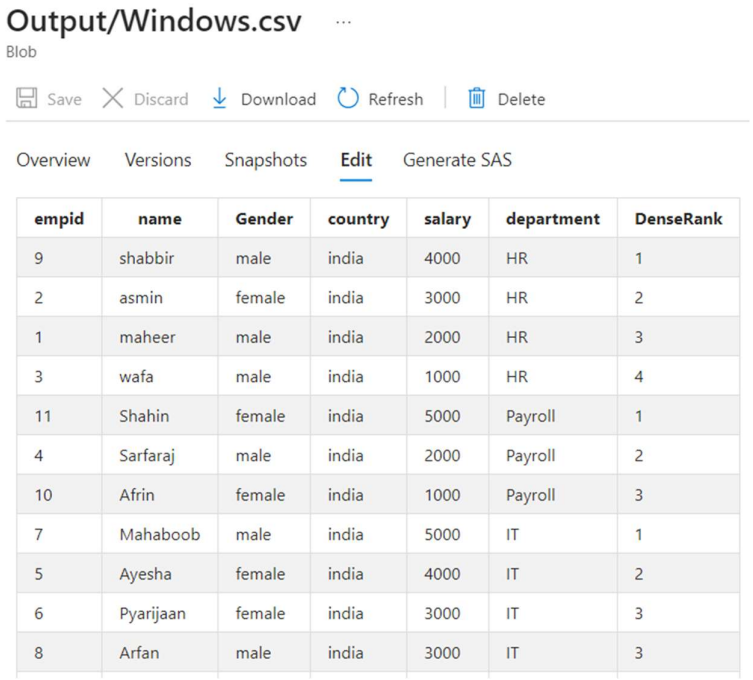
14. Under setting select the Data flow.



15. Validate the pipeline and click on debug.
16. Here our pipeline was executed successfully.



17. Go to Destination and check the data.



Output/Windows.csv

Blob

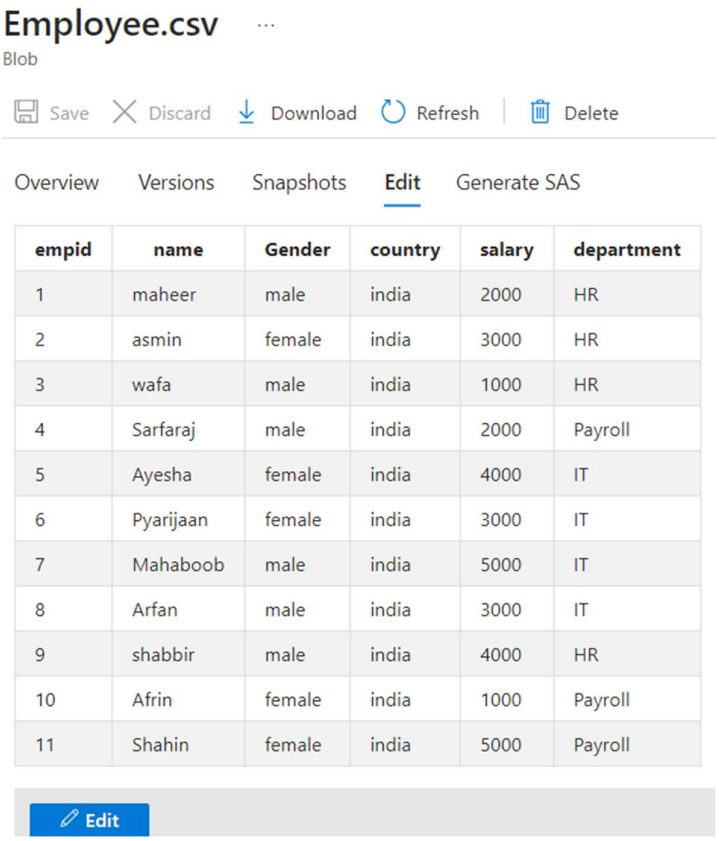
Save Discard Download Refresh Delete

Overview Versions Snapshots Edit Generate SAS

empid	name	Gender	country	salary	department	DenseRank
9	shabbir	male	india	4000	HR	1
2	asmin	female	india	3000	HR	2
1	maheer	male	india	2000	HR	3
3	wafa	male	india	1000	HR	4
11	Shahin	female	india	5000	Payroll	1
4	Sarfaraj	male	india	2000	Payroll	2
10	Afrin	female	india	1000	Payroll	3
7	Mahaboob	male	india	5000	IT	1
5	Ayesha	female	india	4000	IT	2
6	Pyarijaan	female	india	3000	IT	3
8	Arfan	male	india	3000	IT	3

Alter Row Transformation

1. In this example our is below employee details.



Employee.csv

Blob

Save Discard Download Refresh Delete

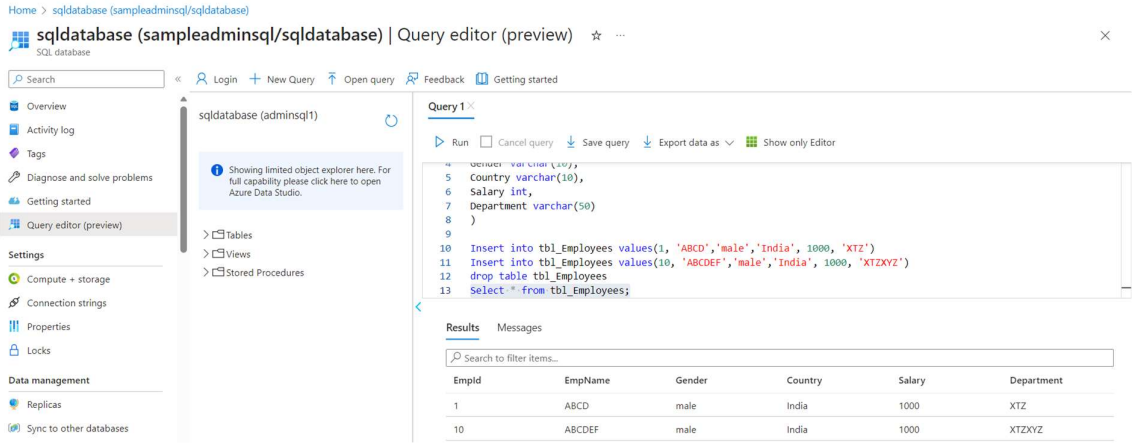
Overview Versions Snapshots Edit Generate SAS

empid	name	Gender	country	salary	department
1	maheer	male	india	2000	HR
2	asmin	female	india	3000	HR
3	wafa	male	india	1000	HR
4	Sarfaraj	male	india	2000	Payroll
5	Ayesha	female	india	4000	IT
6	Pyarijaan	female	india	3000	IT
7	Mahaboob	male	india	5000	IT
8	Arfan	male	india	3000	IT
9	shabbir	male	india	4000	HR
10	Afrin	female	india	1000	Payroll
11	Shahin	female	india	5000	Payroll

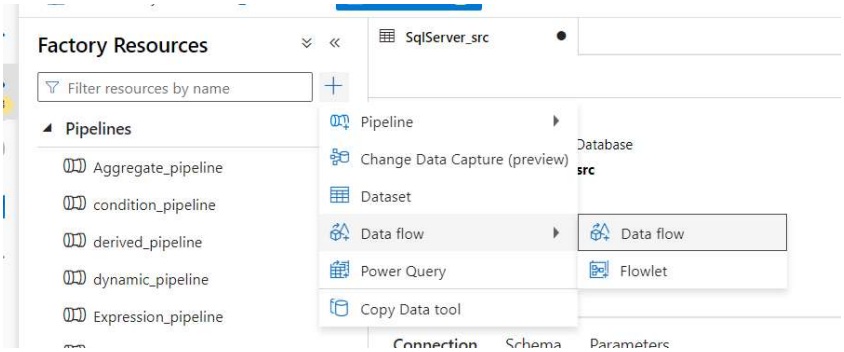
Edit

2. In the Sql Database we have created a sample table with the sample data.
3. This is our destination table.

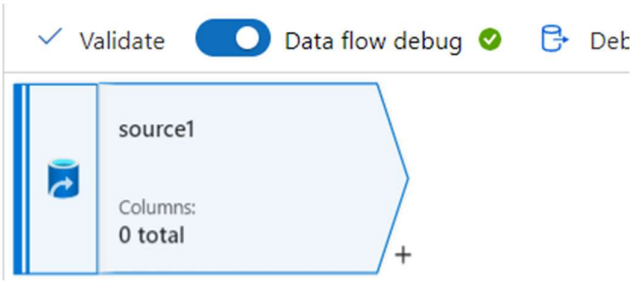
```
Code: Create table tbl_Employees(
EmpId int,
EmpName varchar(50),
Gender varchar(10),
Country varchar(10),
Salary int,
Department varchar(50)
)

Insert into tbl_Employees values(1, 'ABCD','male','India', 1000, 'XTZ')
Insert into tbl_Employees values(10, 'ABCDEF','male','India', 1000, 'XTZXYZ')
```



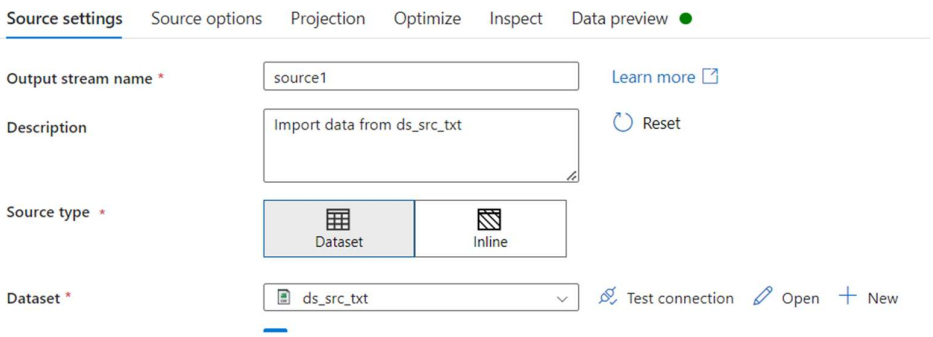
4. Next create a dataset for this table.
5. Create a Data flow.



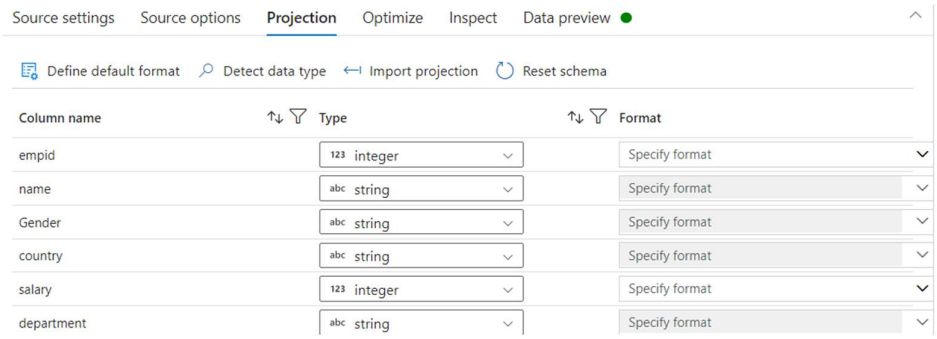
6. Add Data Source.



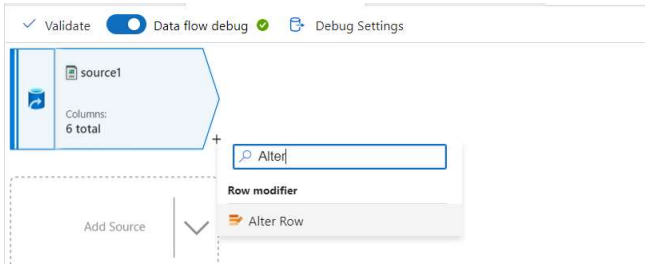
7. Under Source select the source dataset.



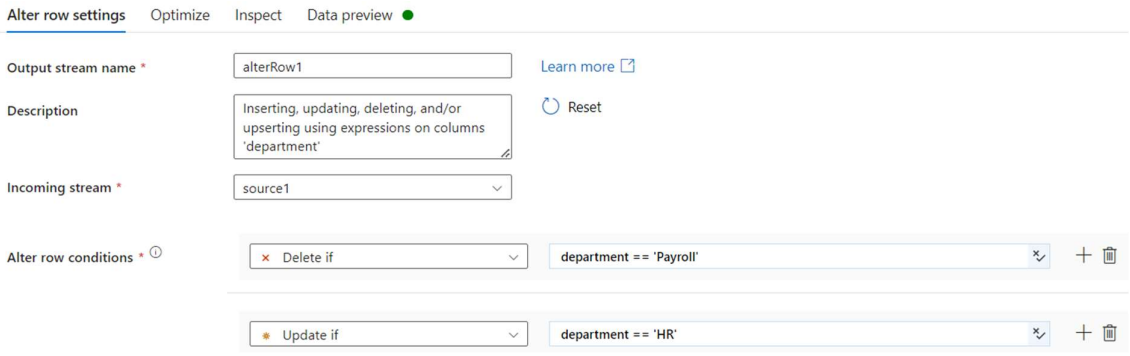
8. Under projection set the datatypes as shown below.



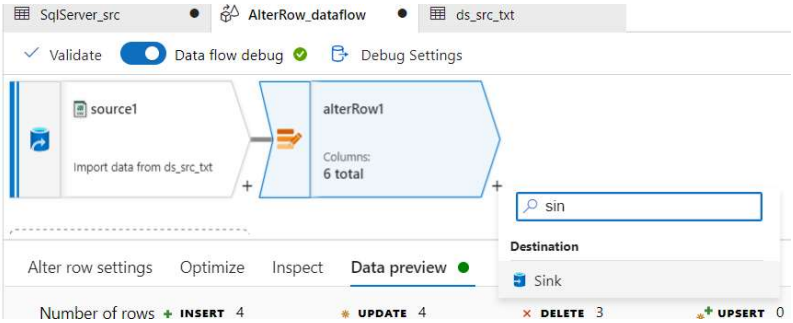
9. Click on the plus and click on Alter Row.



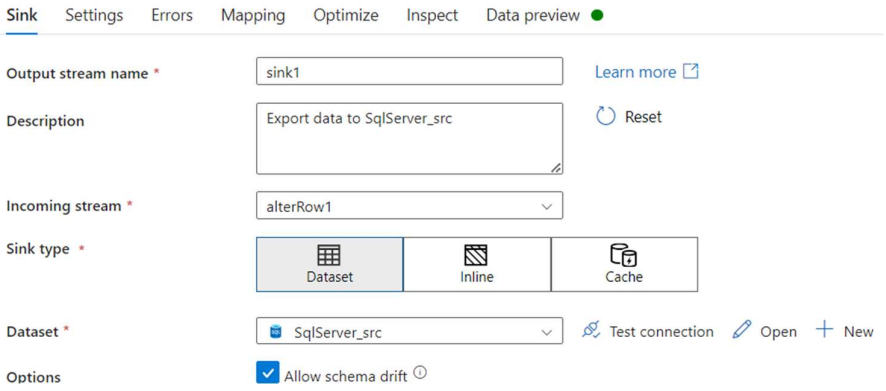
10. Under alter row condition give the below conditions.



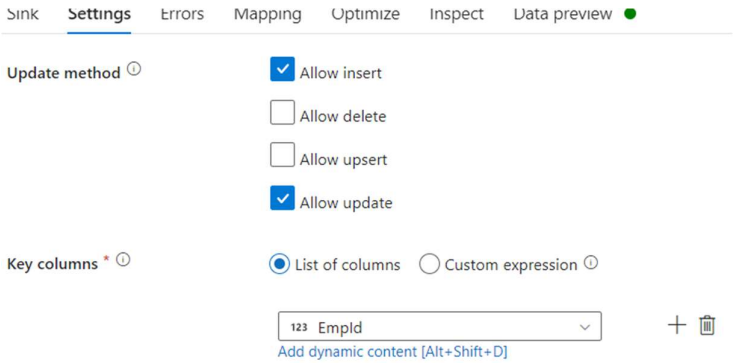
11. Click on the Plus symbol and click on Sink.



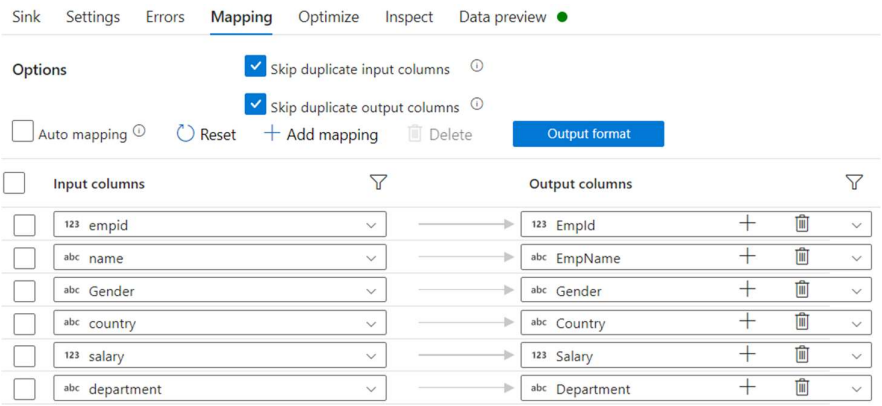
12. Under sink select the destination dataset.



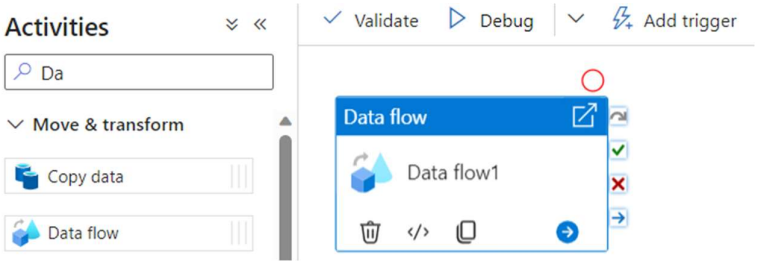
13. Under settings check the below properties.



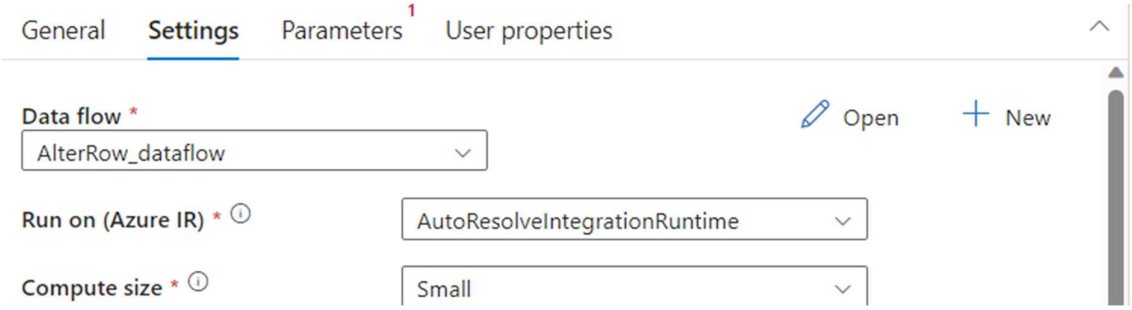
14. Under Mapping, Map the fields as shown below.



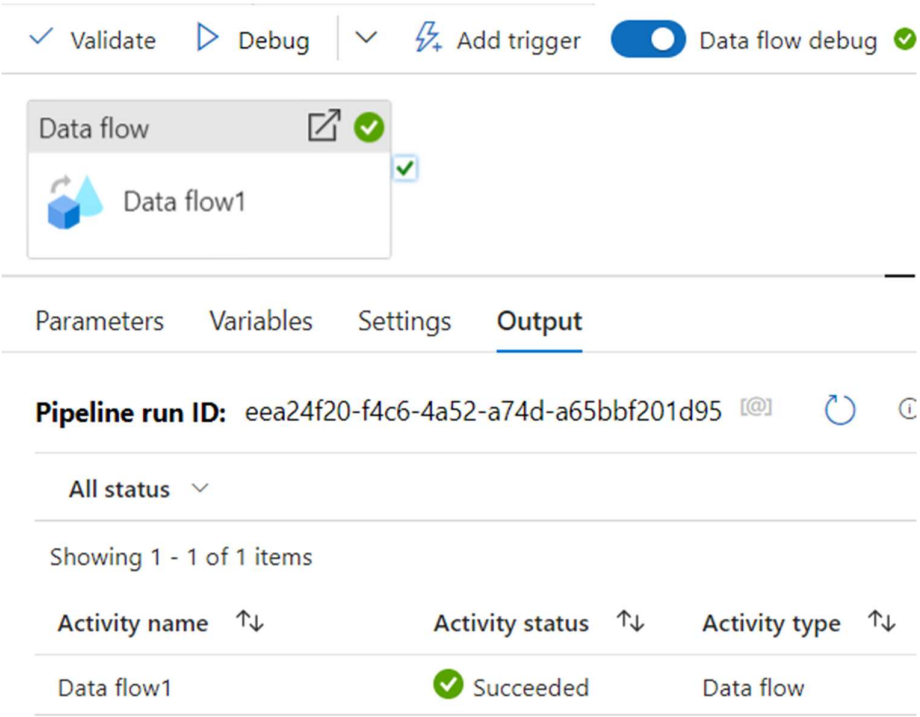
15. Create a Pipeline and Drag and Drop the Dataflow as shown below.



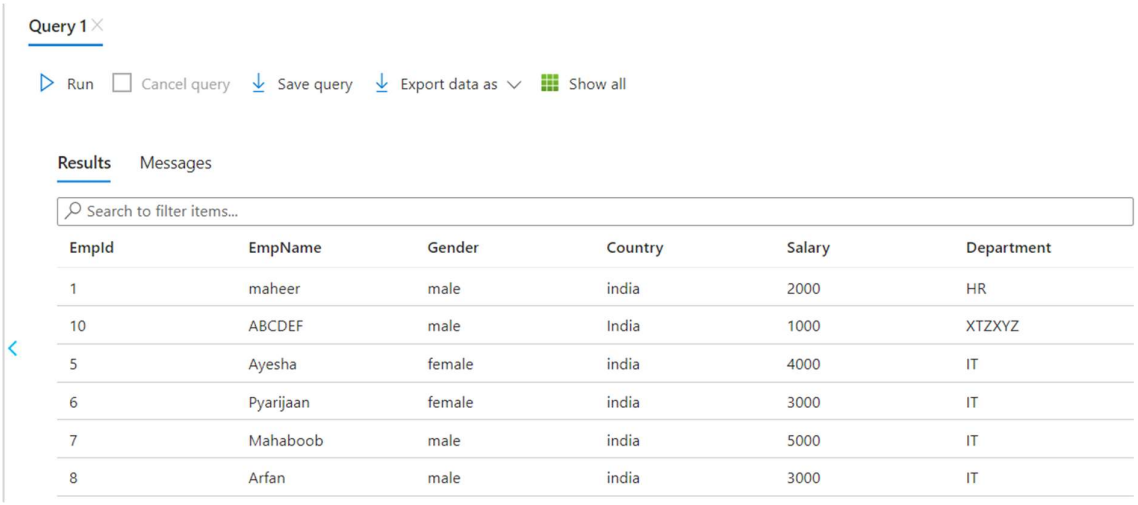
16. Select Dataflow.



17. Validate the pipeline and click on debug.
18. Here our pipeline was executed successfully.



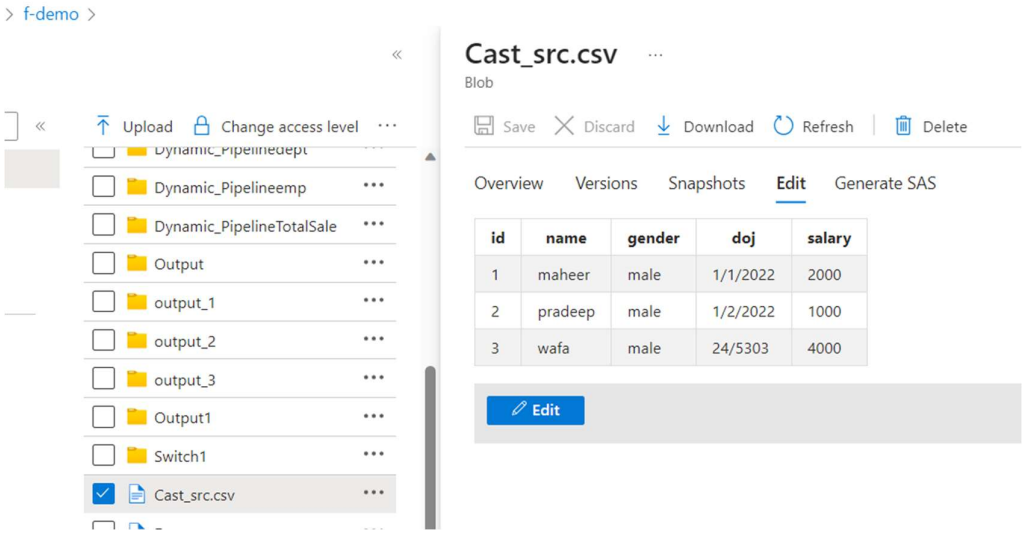
19. Now go to a destination and check the data.



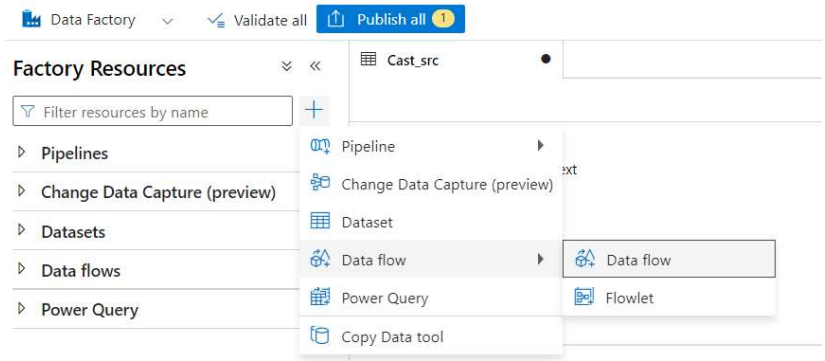
EmpId	EmpName	Gender	Country	Salary	Department
1	maheer	male	india	2000	HR
10	ABCDEF	male	India	1000	XTZXYZ
5	Ayesha	female	india	4000	IT
6	Pyarijaan	female	india	3000	IT
7	Mahaboob	male	india	5000	IT
8	Arfan	male	india	3000	IT

Cast Transformation

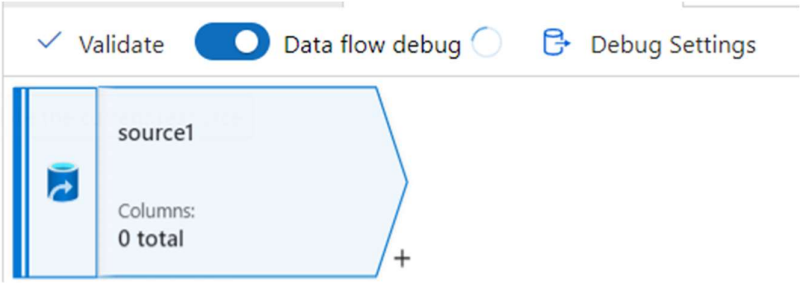
1. In this example we are using the below data.



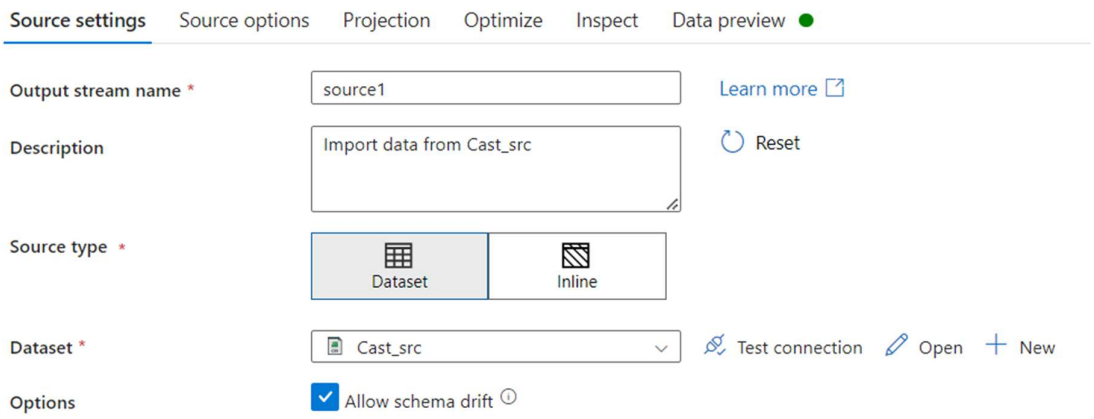
2. Create a Dataset for this data.
3. Create a Dataflow.



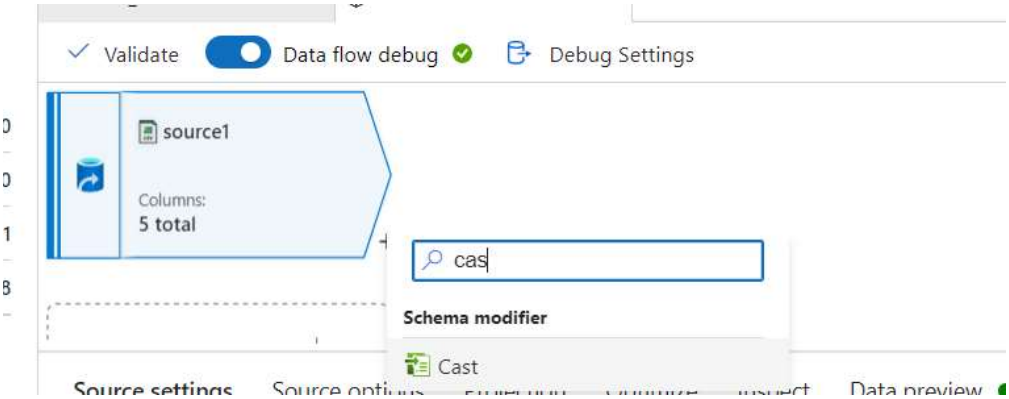
4. Add a Source.



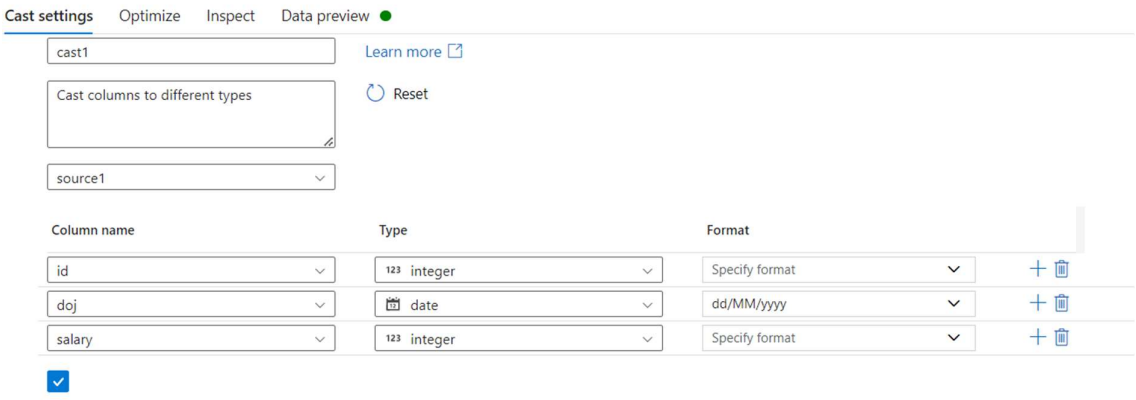
5. Under source setting select the Source dataset.



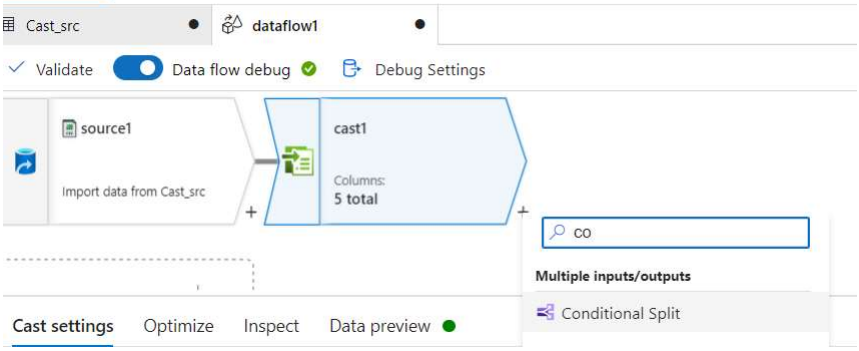
6. Click on the plus symbol and click on the cast.



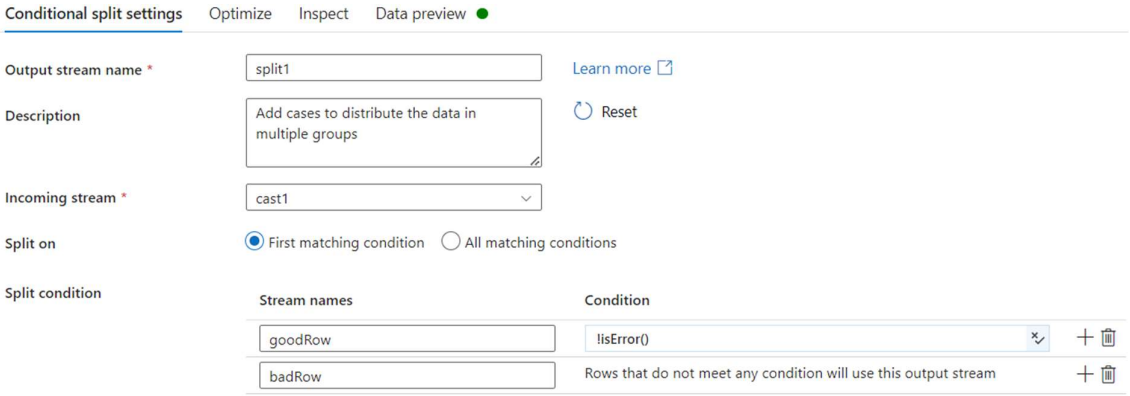
7. Under Cast Settings add the below column names and specify the Type and Format as shown below.



8. Click on the plus symbol and click on the Conditional Split option as shown below.

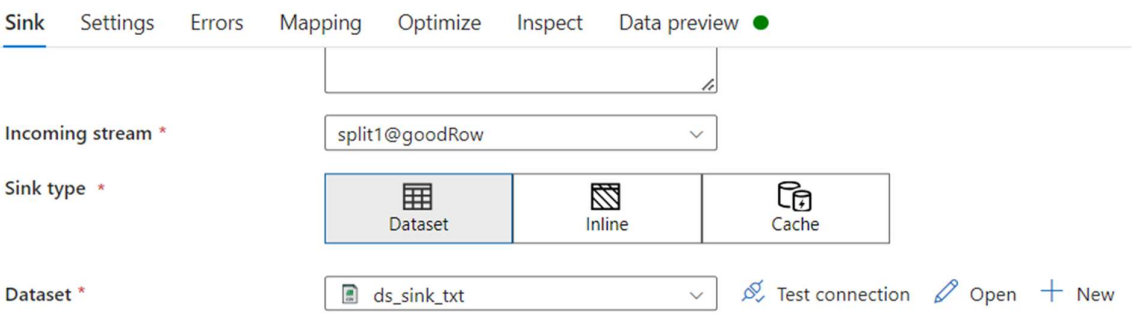


9. Under the Conditional Split give the below conditions.
10. For the goodRow give this expression: !isError()

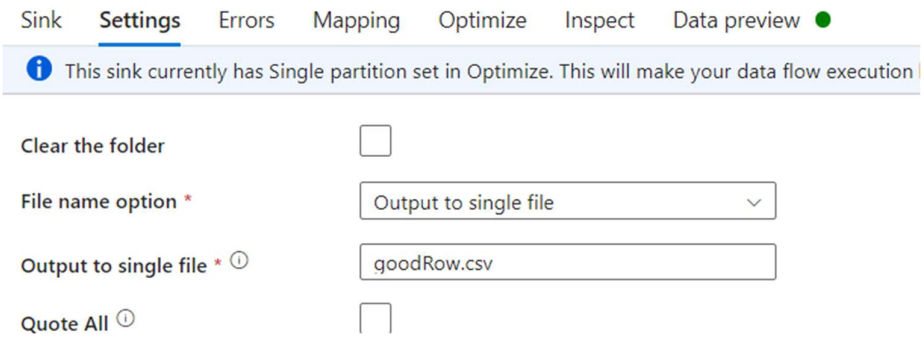


11. Now add Two sinks as shown below.

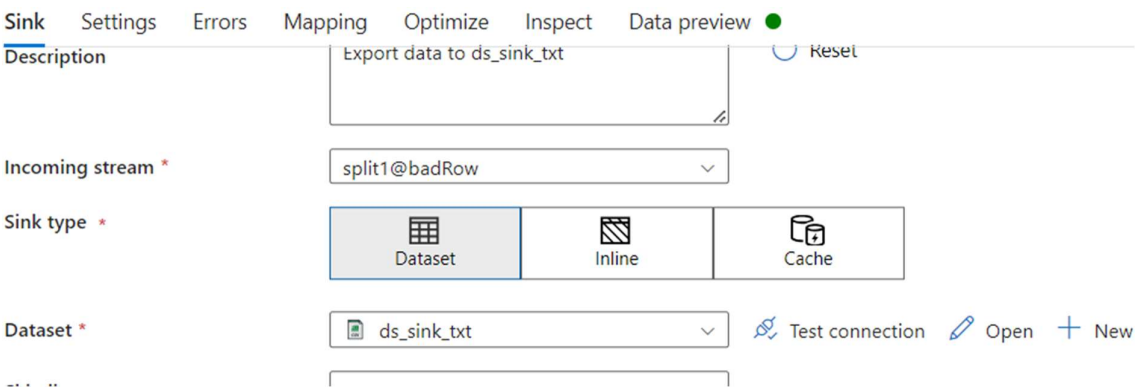
12. For sink1 select the target dataset.



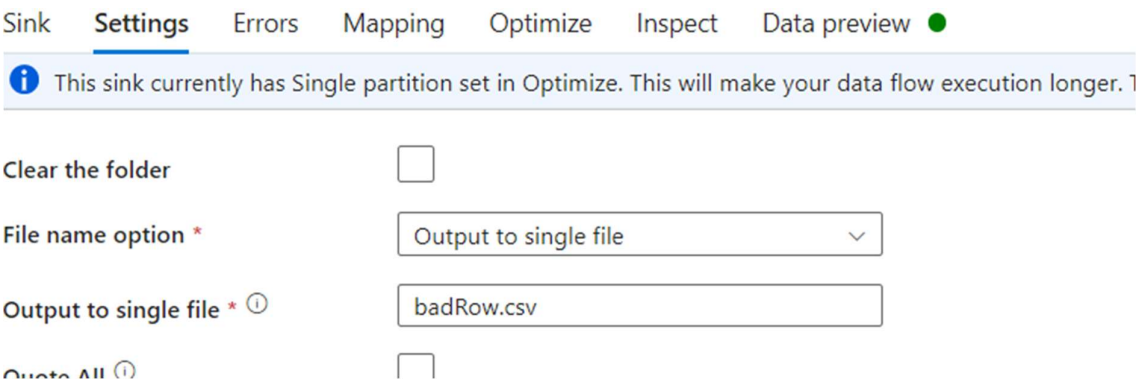
13. And give file type and name as shown below.



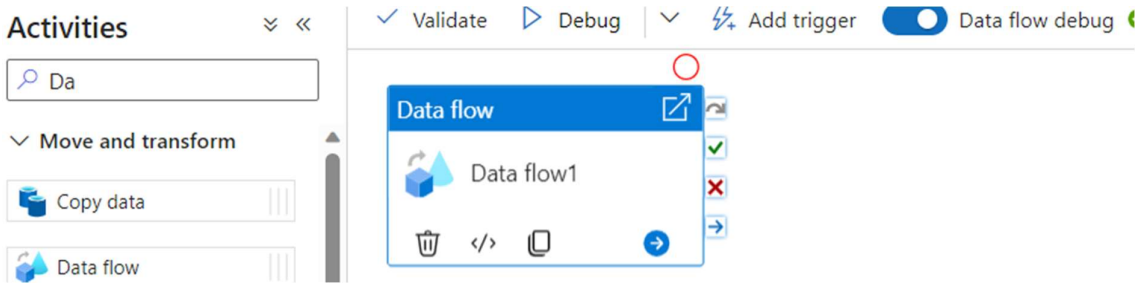
14. For sink2 select the target dataset.



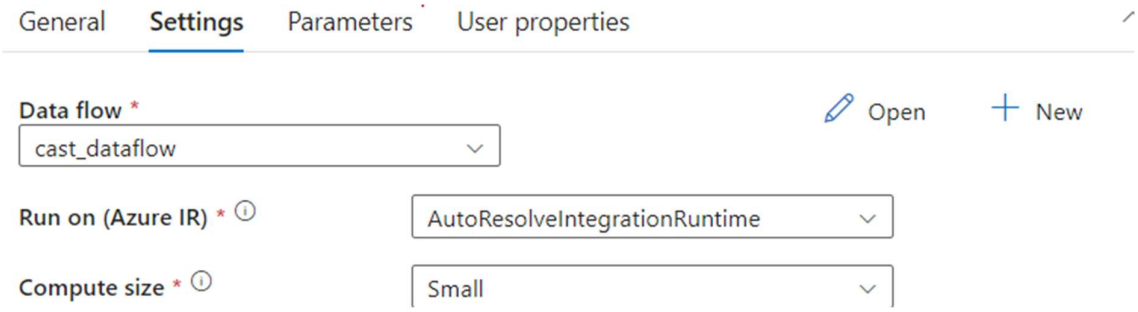
15. And give file type and name as shown below.



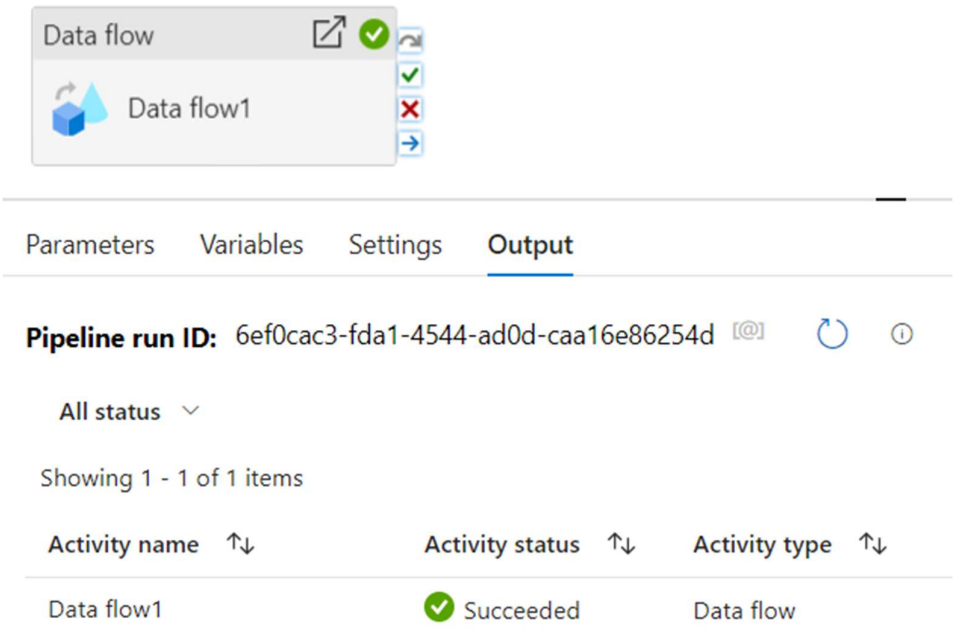
16. Create a Pipeline, Drag and drop the Dataflow activity.



17. And select the Dataflow.



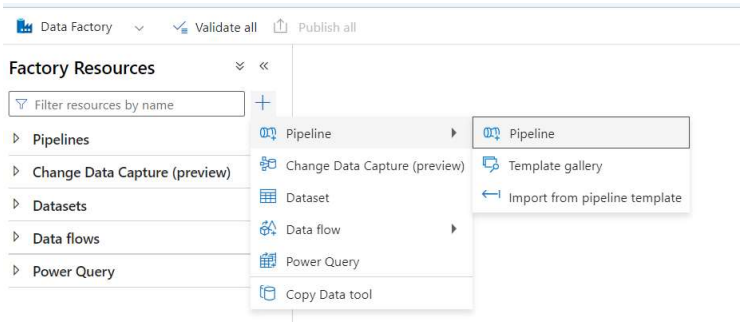
18. Now validate the pipeline and click on debug.
19. Here our pipeline was executed successfully.



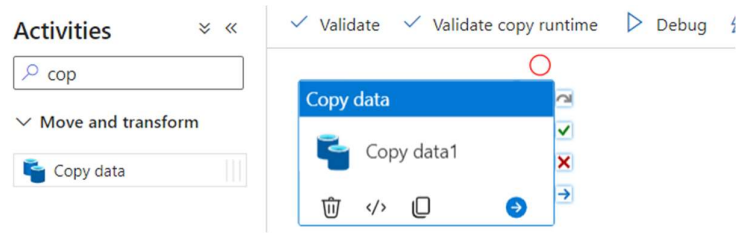
20. Now go to the destination and check the files and data.

Web Page Usage

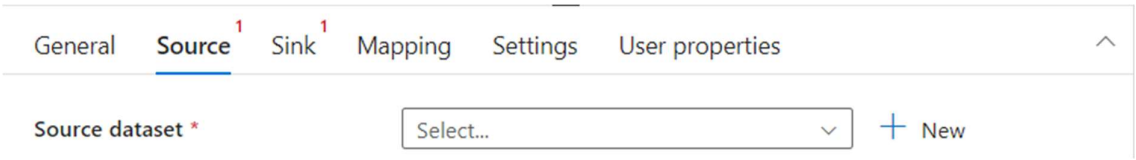
1. Create a Pipeline.



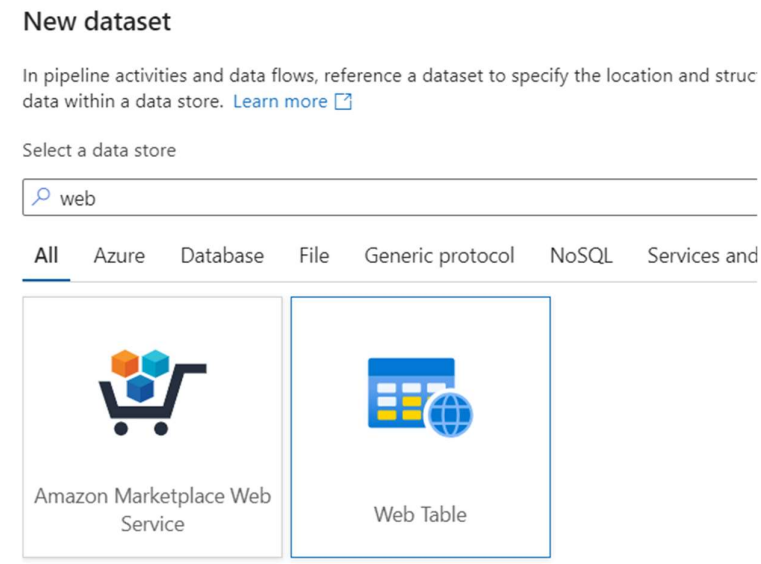
2. Drag and drop the copy data activity.



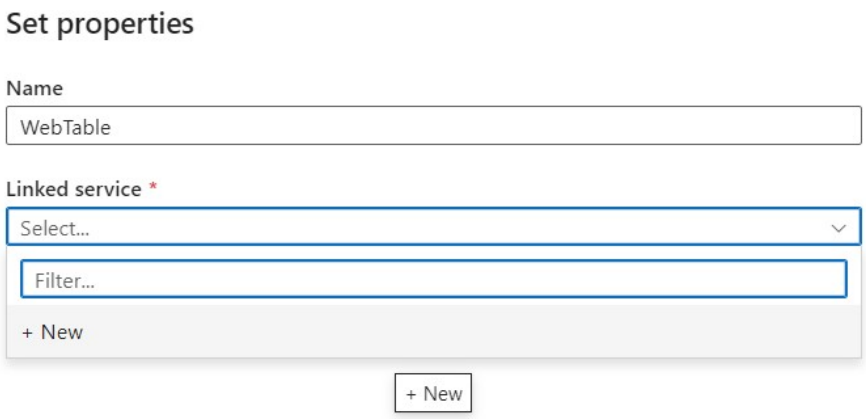
3. Under source click on the plus symbol.



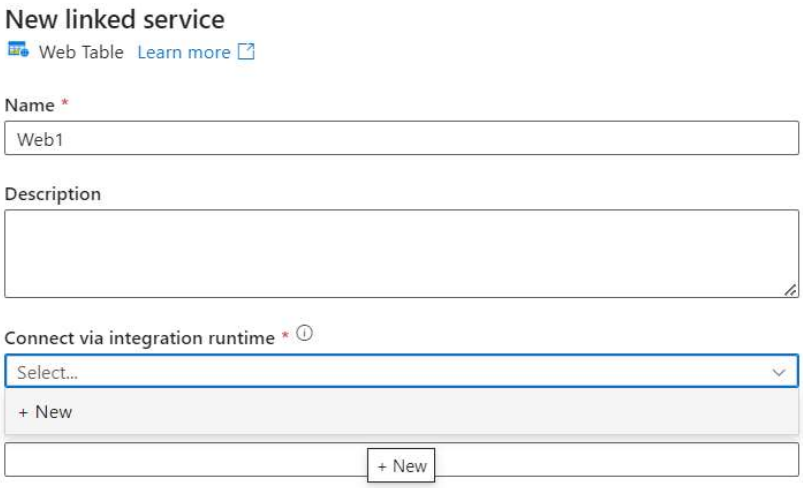
4. Select the web table and click on continue.



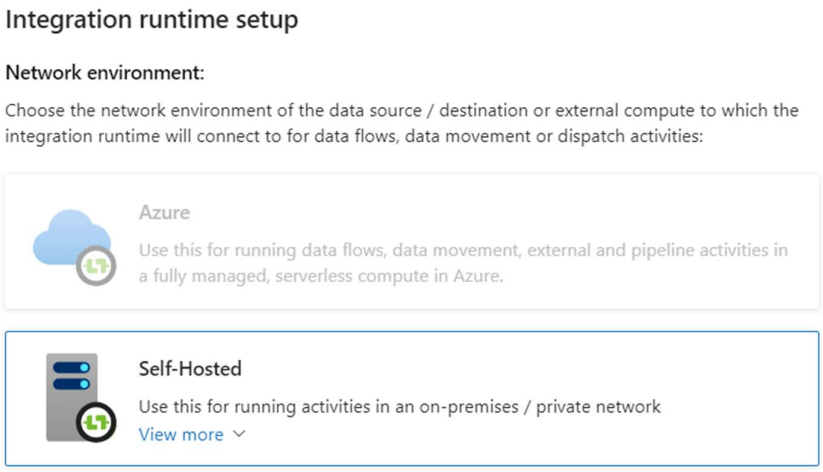
5. Give the name and click on New.



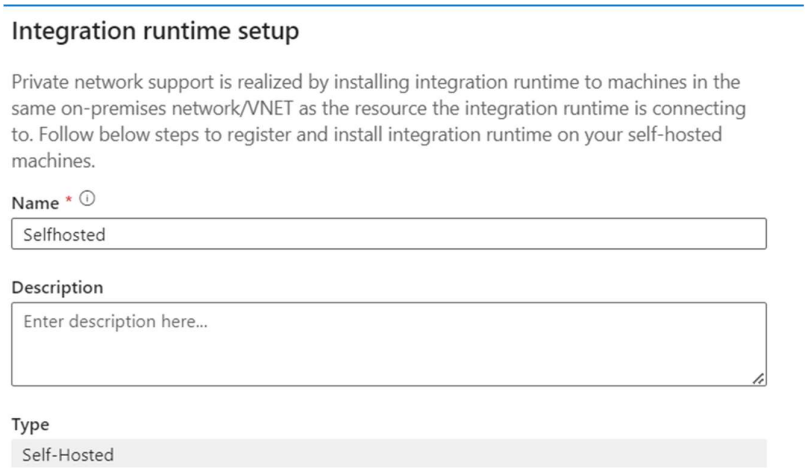
6. Under Integration Runtime click on New.



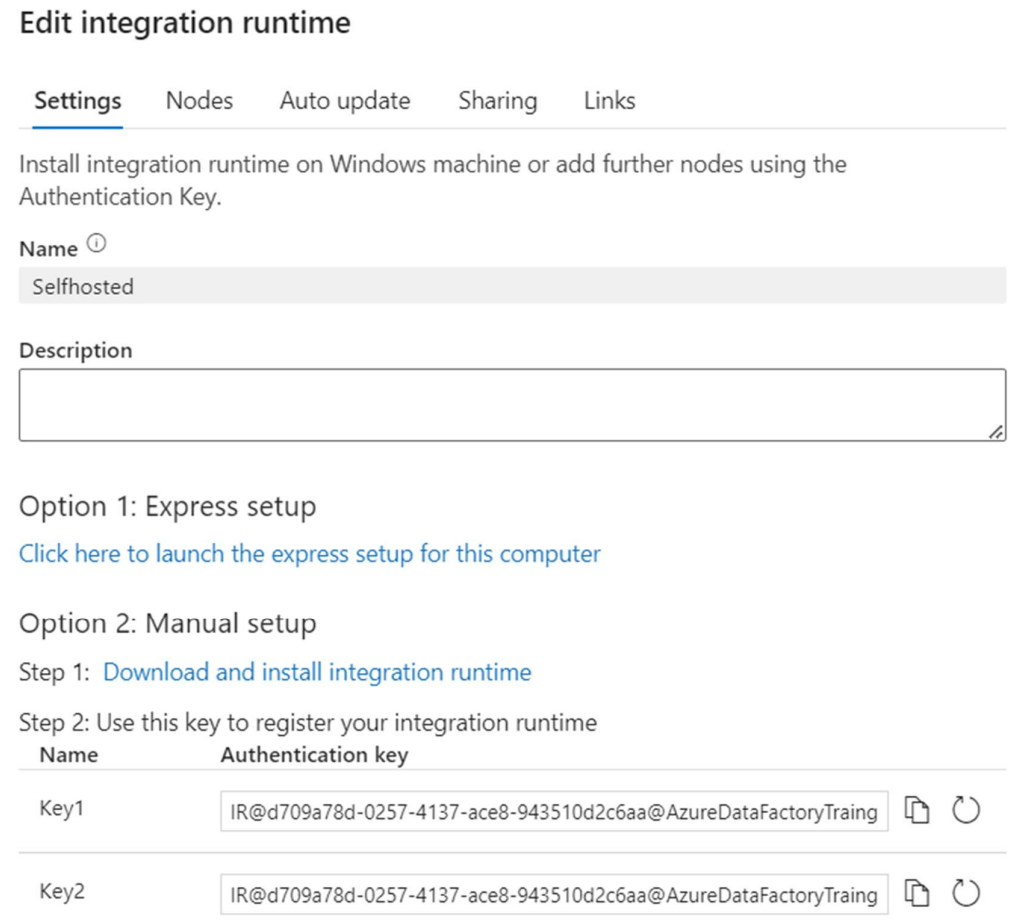
7. Select Selfhosted and click on continue.



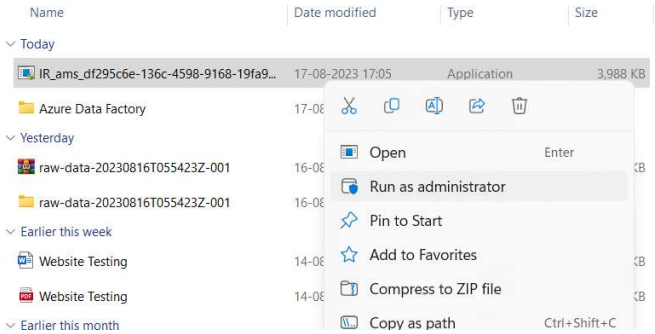
8. Give the Name and click on create.



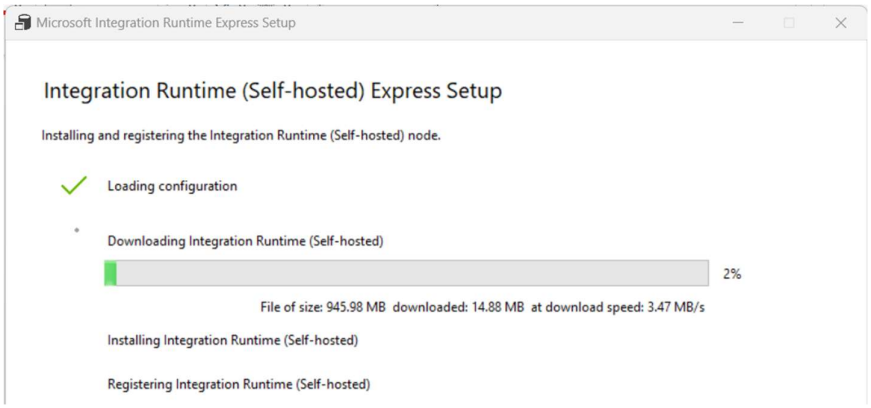
9. Next, we need to Install Integration Runtime. In our case we are using Option 1 so click on the Option 1 link.
10. A file will be downloaded.



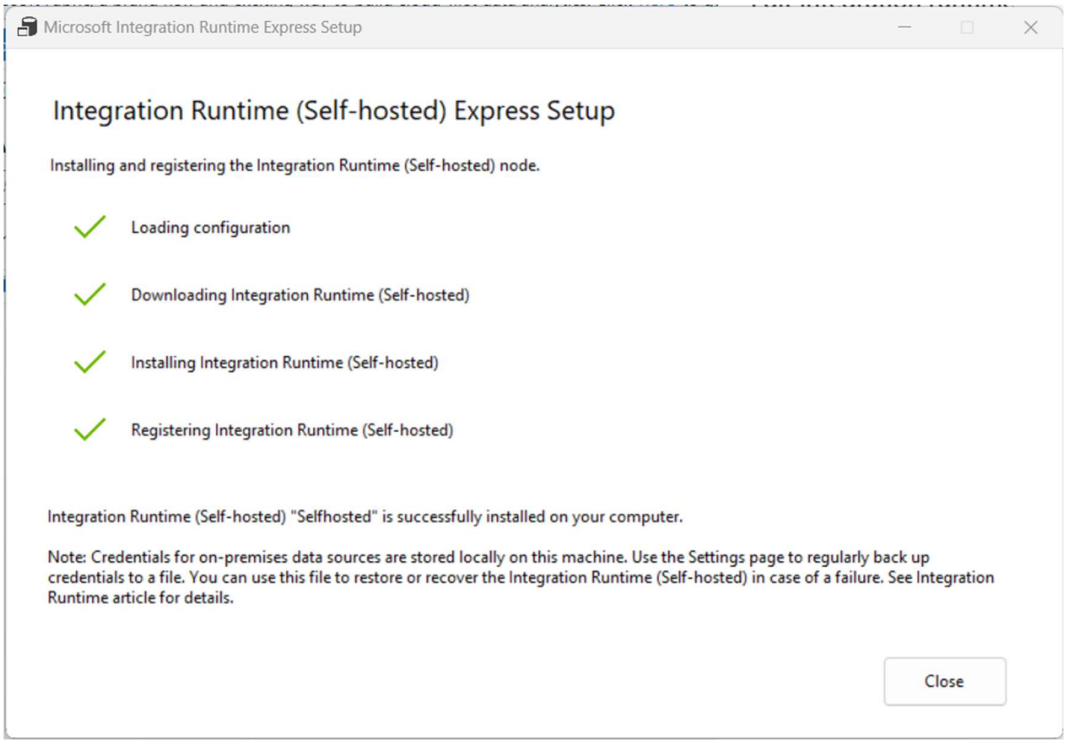
11. Right-click on the file and click on Run as administration as shown below.



12. It will start downloading and installing. It will take some time.

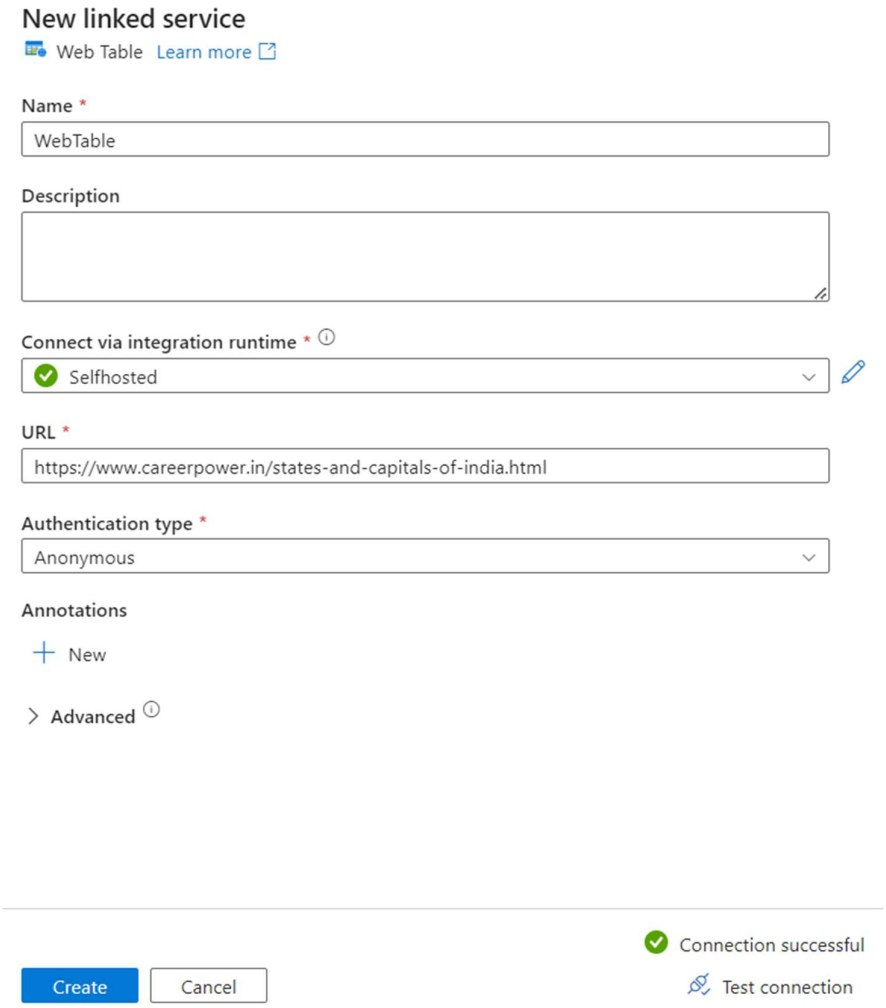


13. After Installing click on the close option.

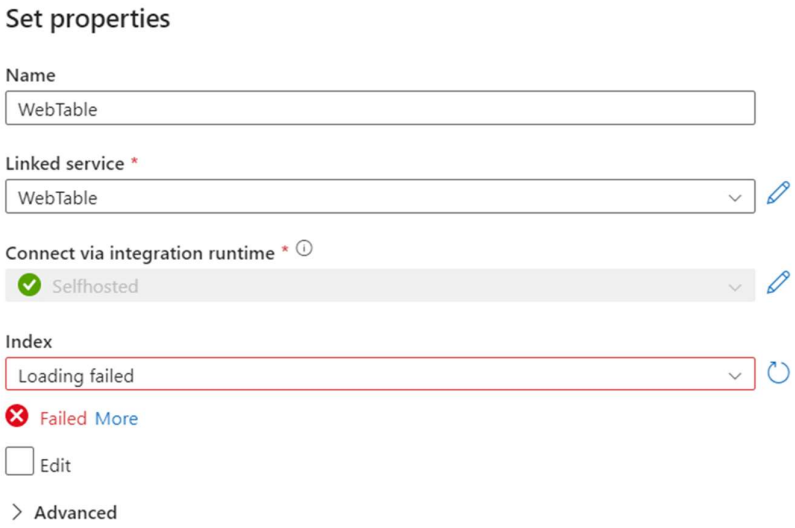


14. Next in the Linked service select the Integration Runtime and give the below URL.

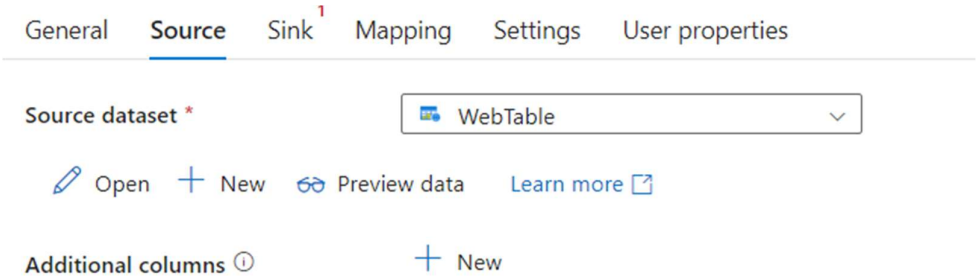
URL: https://www.careerpower.in/states-and-capitals-of-india.html



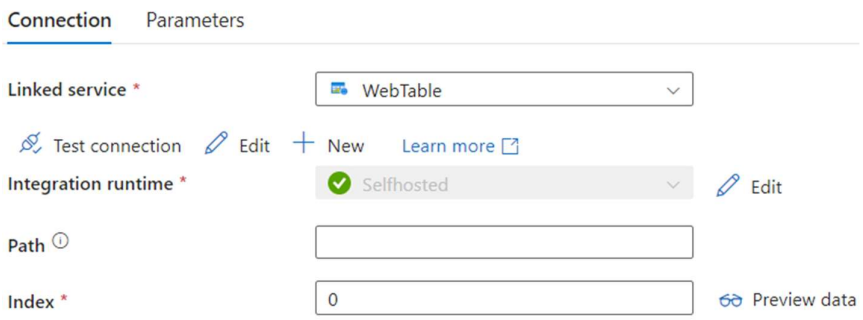
15. You will see it like this but click on Ok.



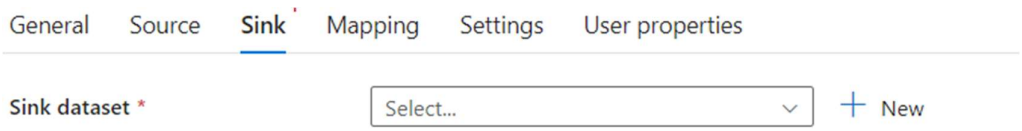
16. Click on Open.



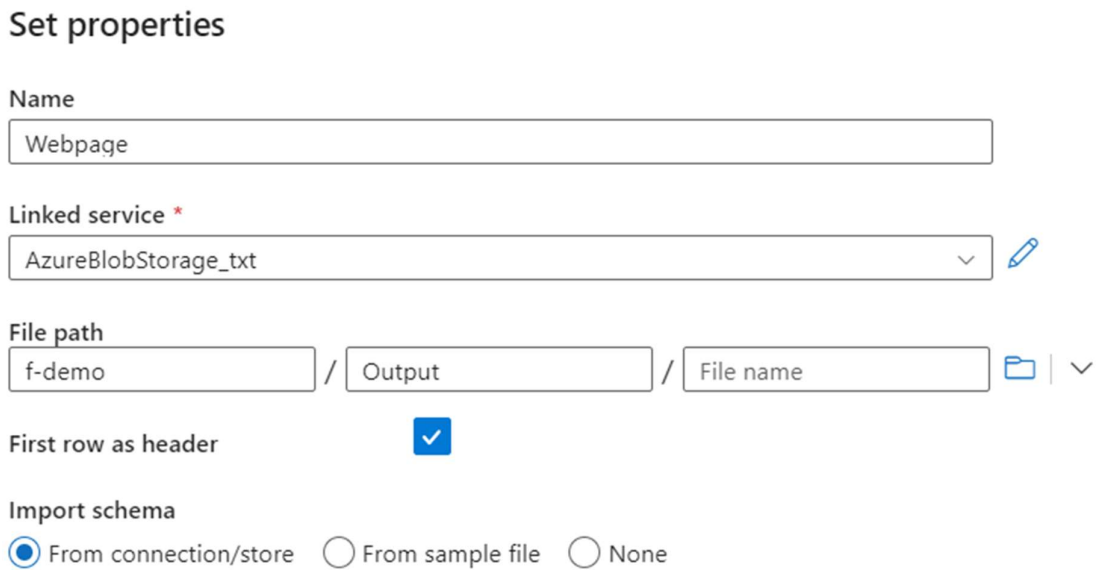
17. Under Index give the value as 0.



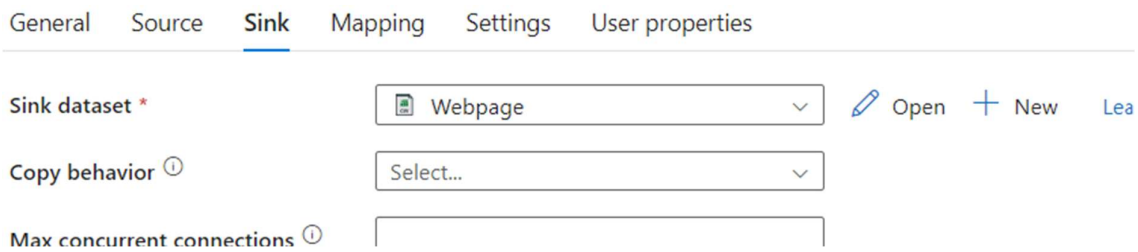
18. Now go back to the pipeline.
19. Under sink click on new.



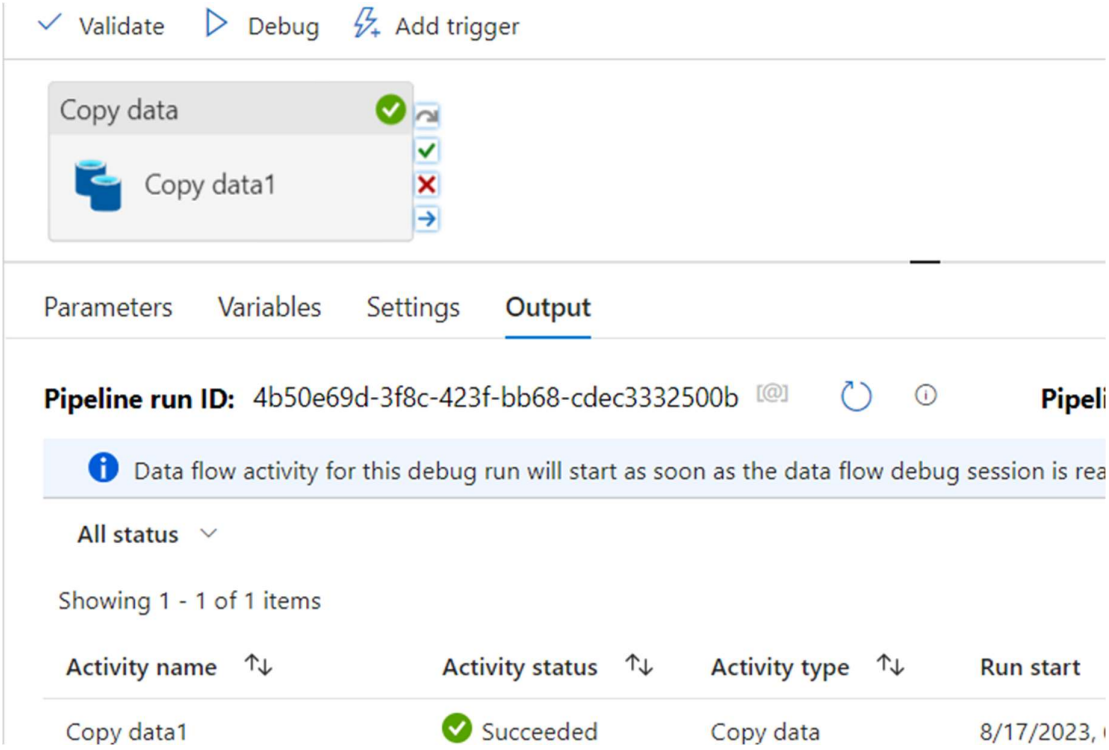
20. Select the Azure Blob Storage and click on continue then select Delimited Text and click on continue.
21. Next give the Name, select the Linked service, and give the File path, Name.



22. We have configured our target part.

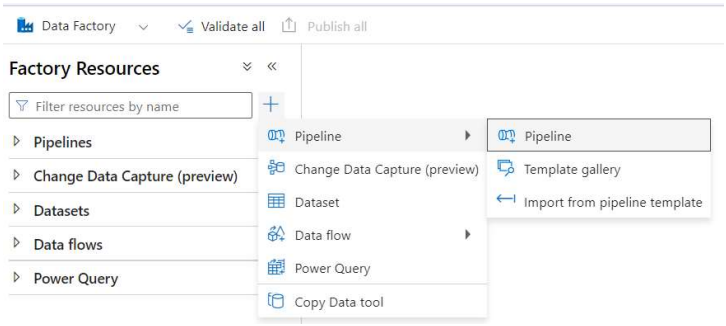


23. Validate the pipeline and click on debug.
24. Our pipeline was executed successfully.

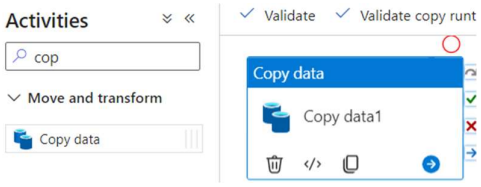


Rest Pagination

1. Create a Pipeline.

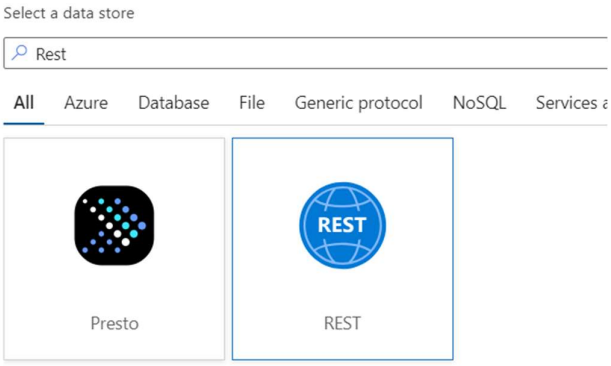


2. Drag and drop the copy data activity as shown below.

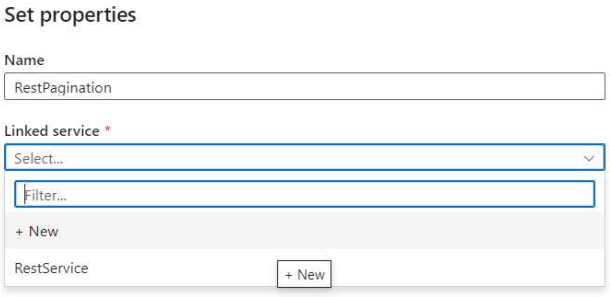


3. Under the source click on the plus symbol.

4. Select the Rest and click on Continue as shown below.

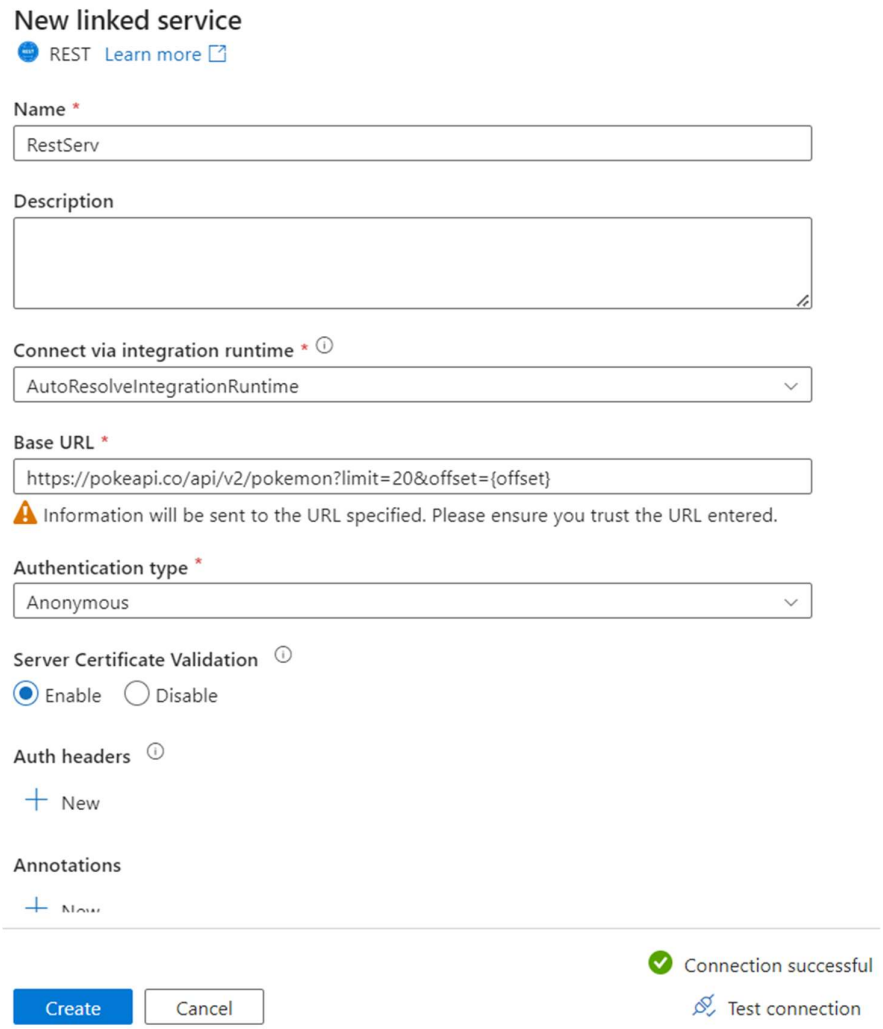


5. Give the Name and click on New.

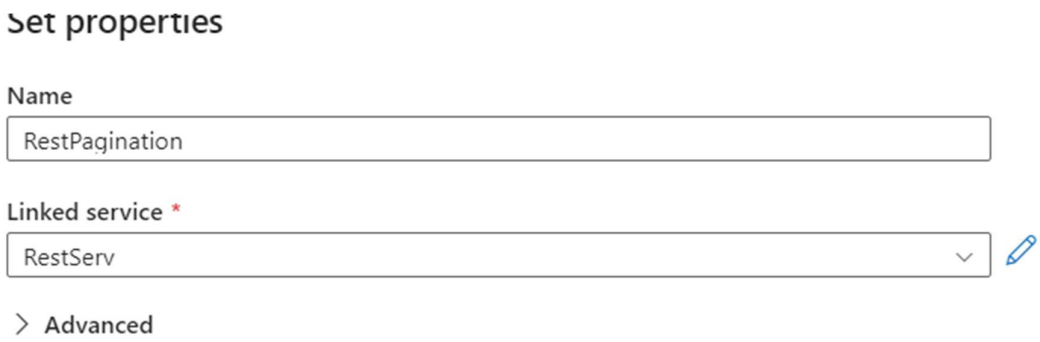


6. Give the Below URL and click on Create.

URL: https://pokeapi.co/api/v2/pokemon?limit=20&offset={offset}

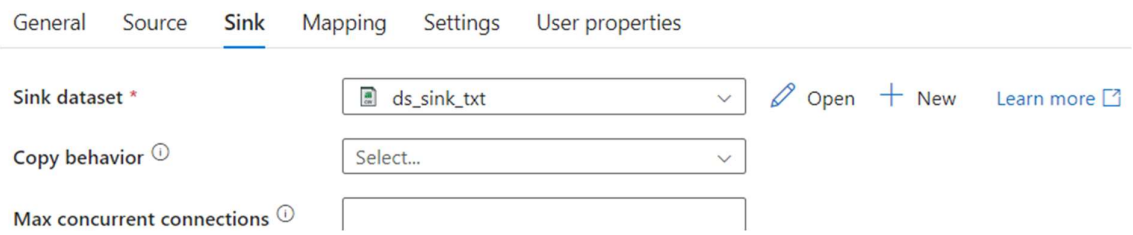


7. Next click on Ok.

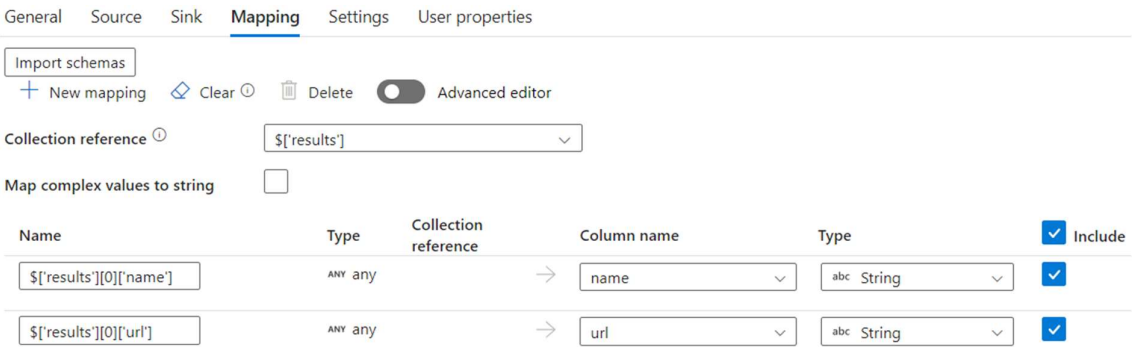


8. Under Source, in the Pagination Rule set up the below Rule with the below Range.

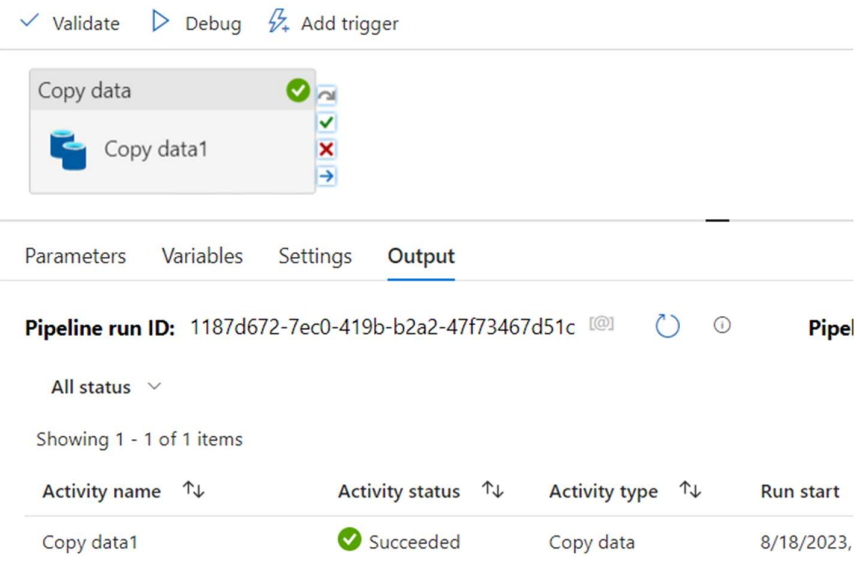
9. Under sink select the Destination.



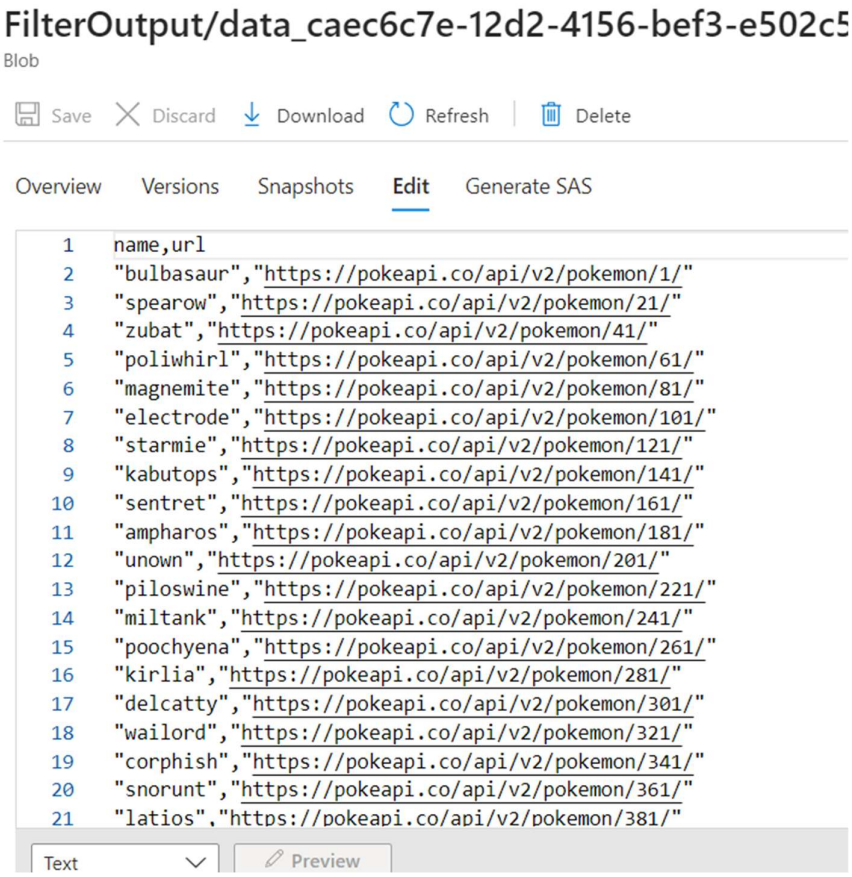
10. Under mapping click on Import Schema.
11. Then select the reference.
12. Under Name delete every column except below two.



13. Now validate the pipeline and click on Debug.
14. Here our pipeline was executed successfully.

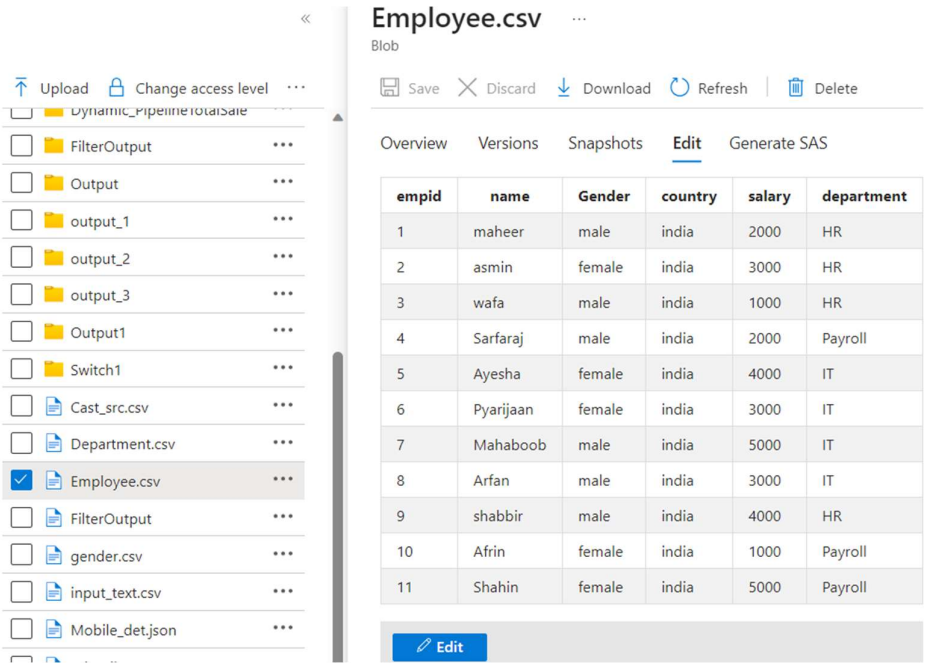


15. Go to Destination and check the Data.

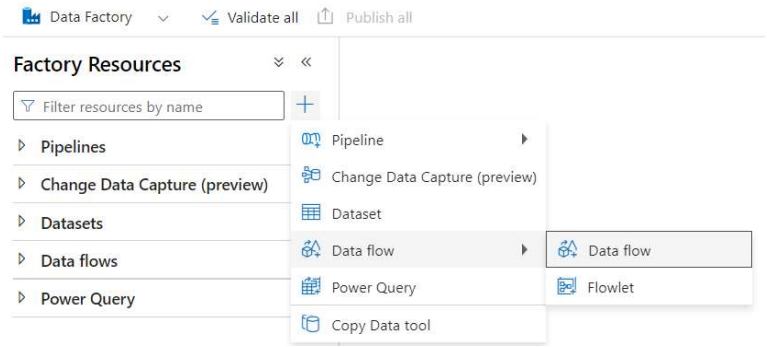


Rank Transformation

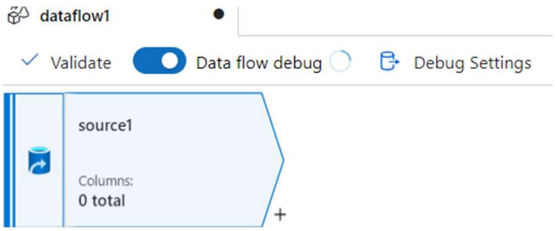
1. In this example we are using the below data.



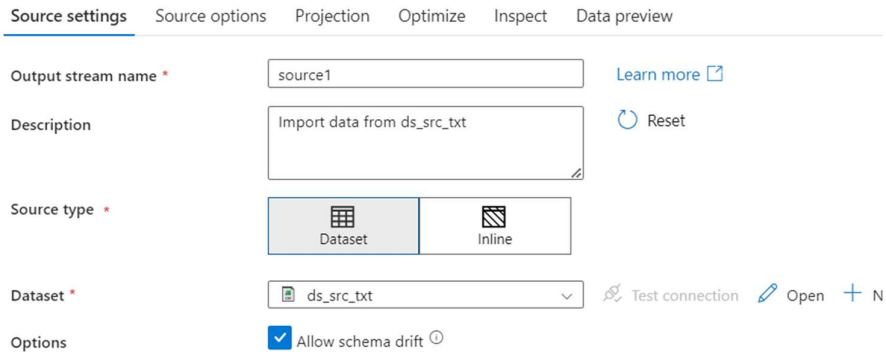
2. Create a Dataflow.



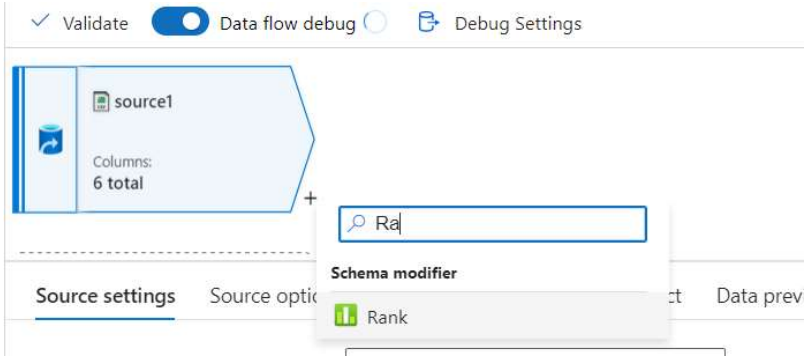
3. Add a Source.



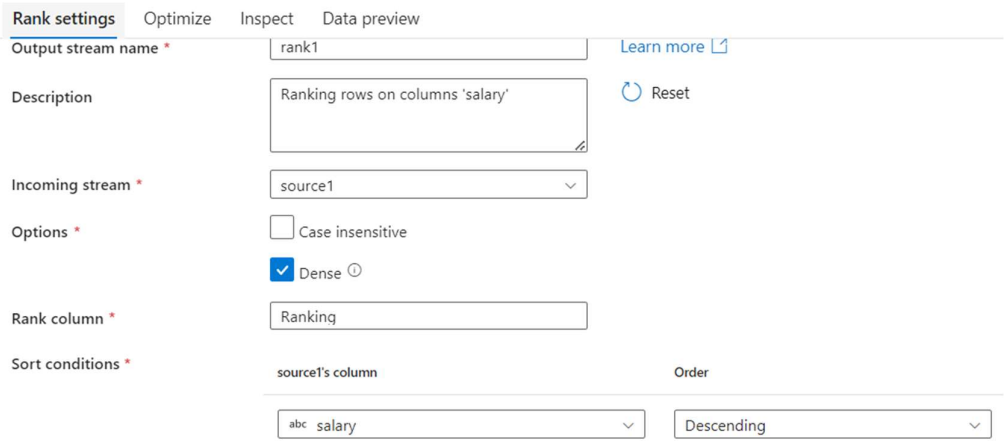
4. Under the Source setting select the Source dataset.



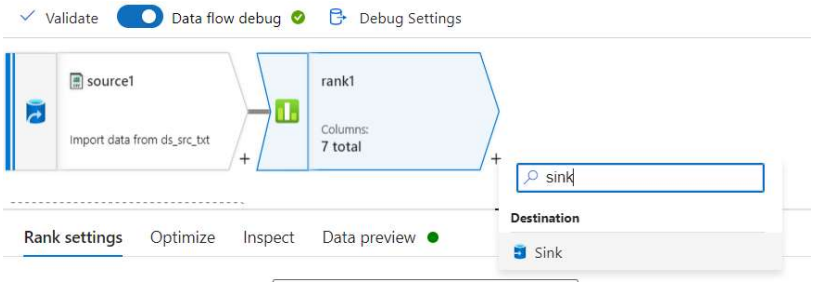
5. Click on the Plus symbol and click on Rank as shown below.



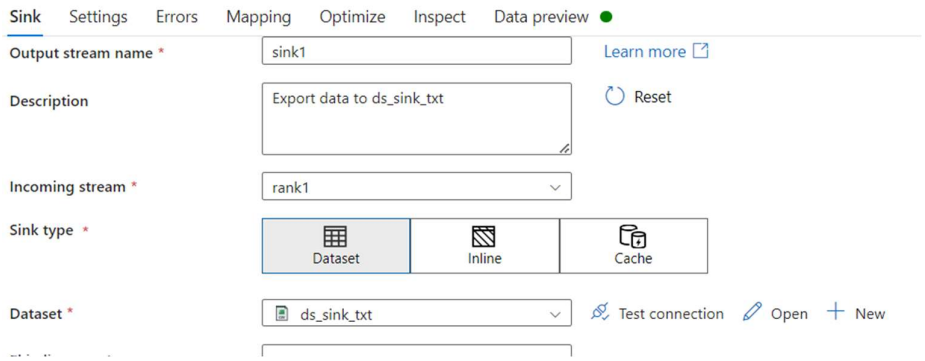
6. Under Rank setting, select the Options, give the Rank column name, and set the below condition.



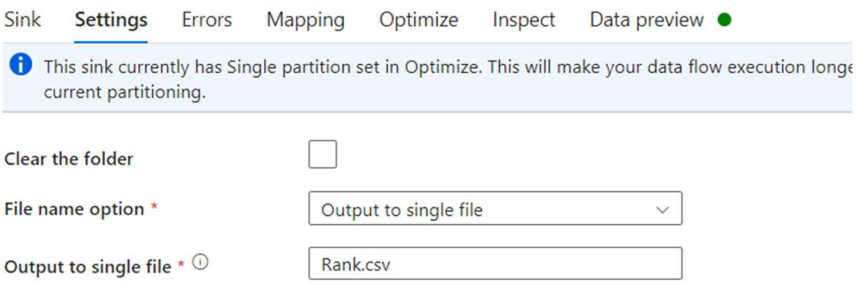
7. Click on the plus symbol and click on Sink.



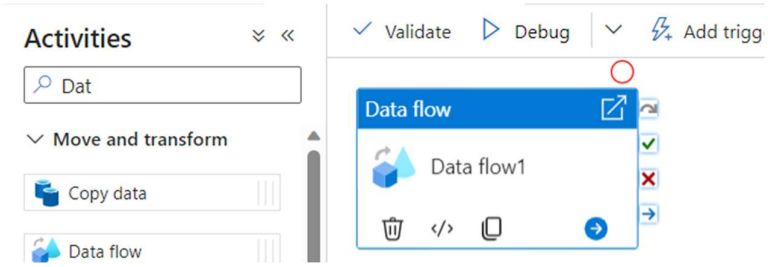
8. Select the Destination dataset.



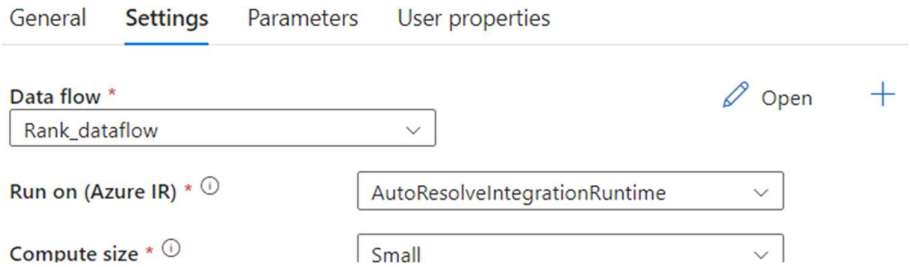
9. Under settings select the file name option and give the file name as shown below.



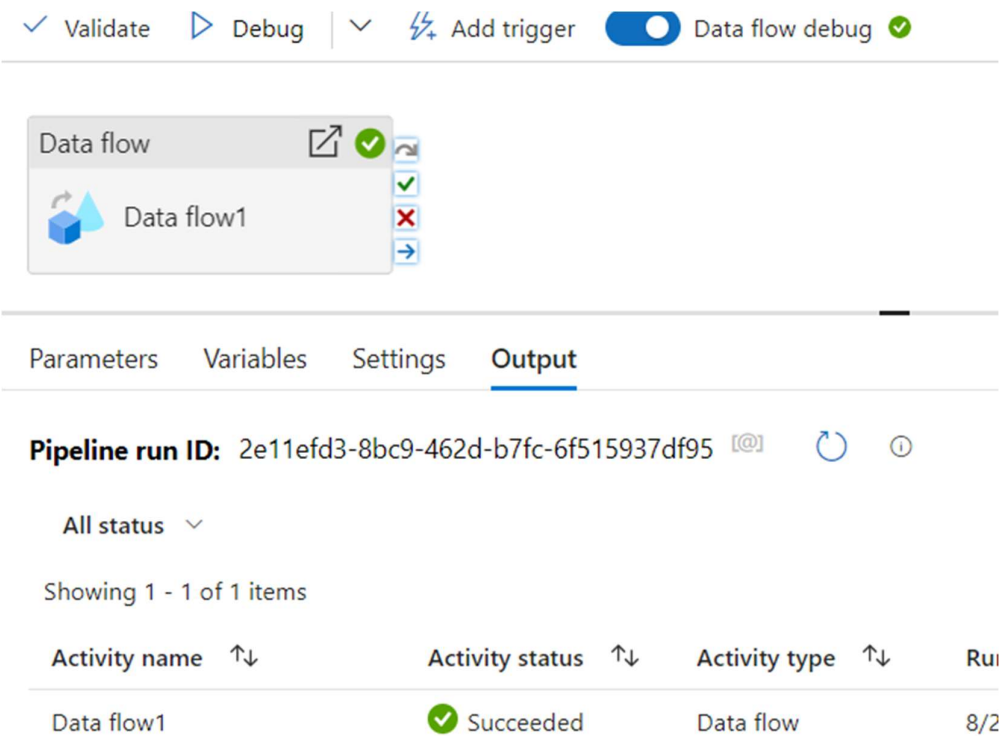
10. Create a Pipeline and drag and drop the Data Flow Activity as shown below.



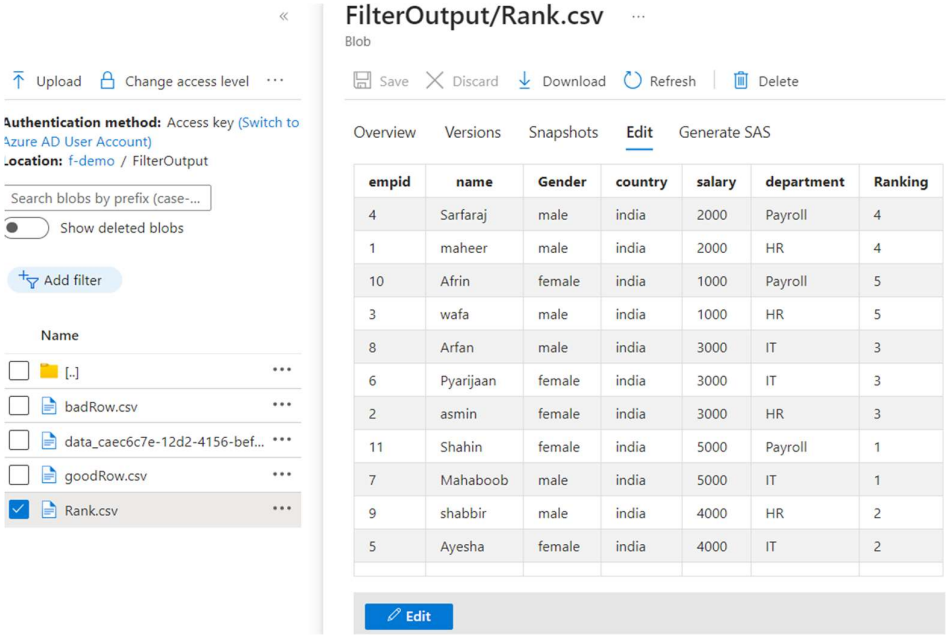
11. Under settings select Dataflow.



12. Validate the pipeline and click on Debug.
13. Here our pipeline was executed successfully.



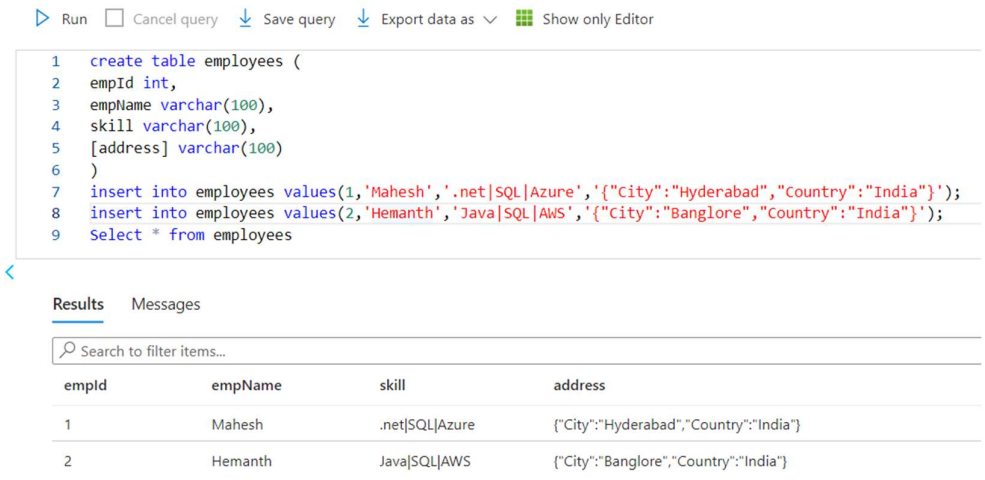
14. Now go to Destination and check the Data.



Parse Transformation

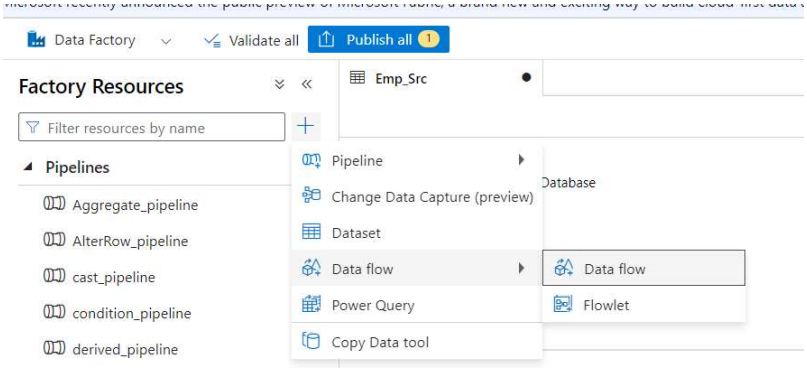
1. In this we are using the below table from the SQL database.

Code: Create table employees (
empId int,
empName varchar(100),
skill varchar(100),
[address] varchar(100)
)
insert into employees
values(1,'Mahesh','.net|SQL|Azure','{"City":"Hyderabad","Country":"India"}');
insert into employees
values(2,'Hemanth','Java|SQL|AWS','{"City":"Banglore","Country":"India"}');
Select * from employees

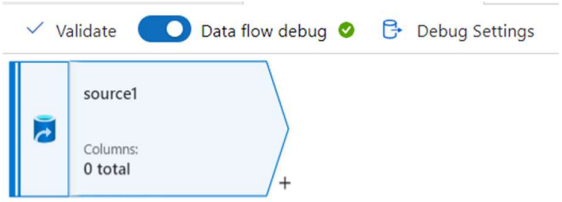


2. Create a Dataset for the Above table.

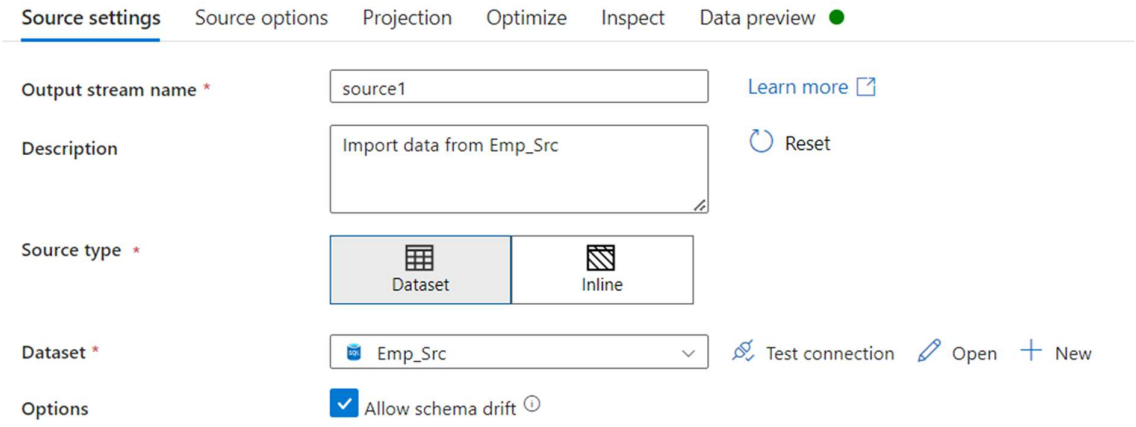
3. Create a Dataflow.



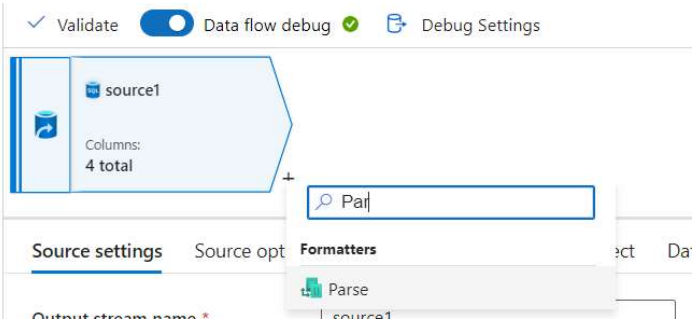
4. Add a Source.



5. Select the Source Dataset.

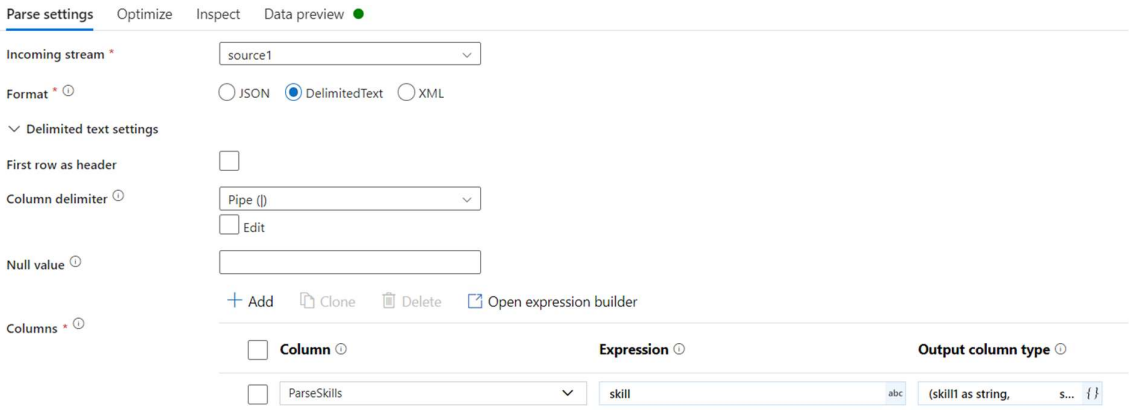


6. Click on the plus symbol and click on Parse.

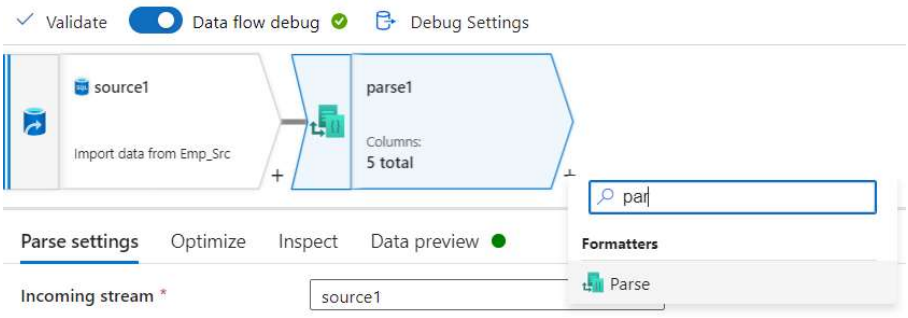


7. Under the Parse setting, set the below properties and for the Output, column type gives the below expression.

Exp: (skill1 as string, skill2 as string, skill3 as string)

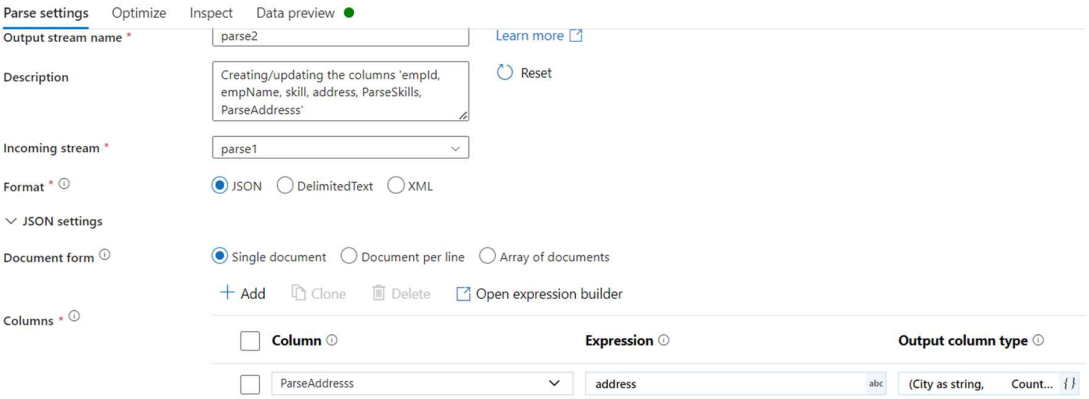


8. Click on the Plus symbol and click on Parse again.

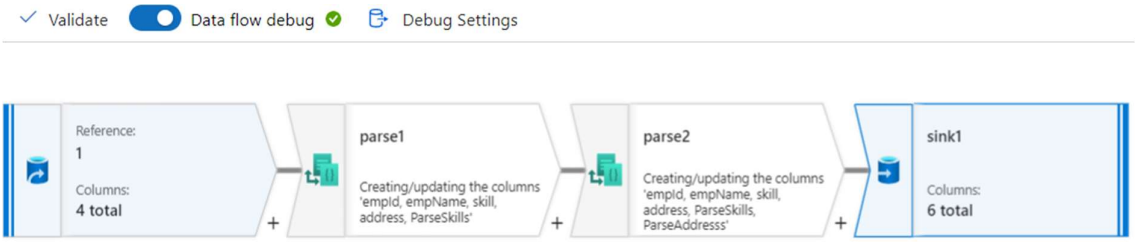


9. Under the Parse setting, set the below properties and for the Output, column type gives the below expression.

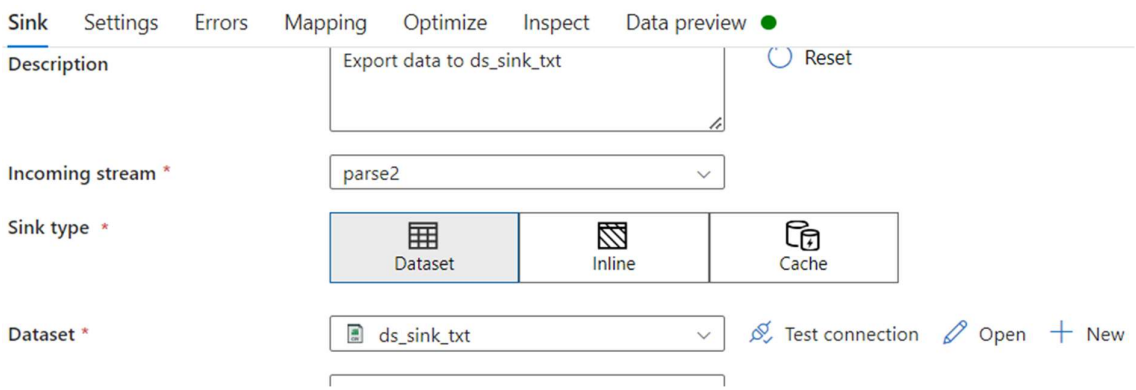
Exp: (City as string, Country as string)



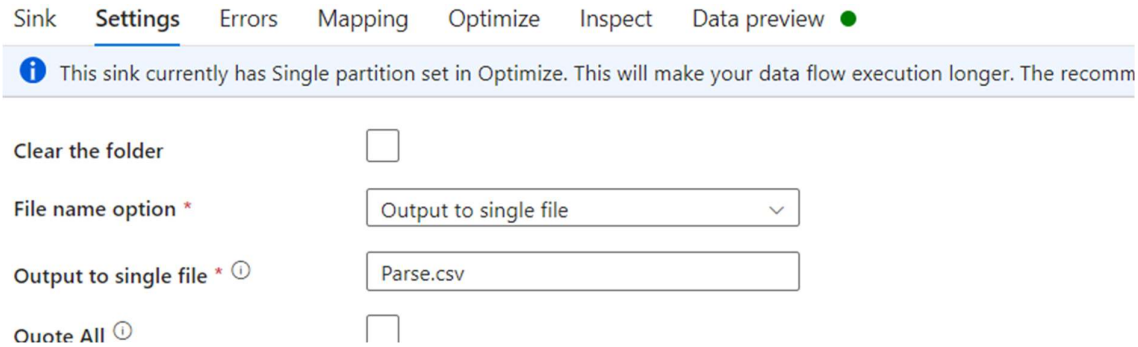
10. Next Add a sink.



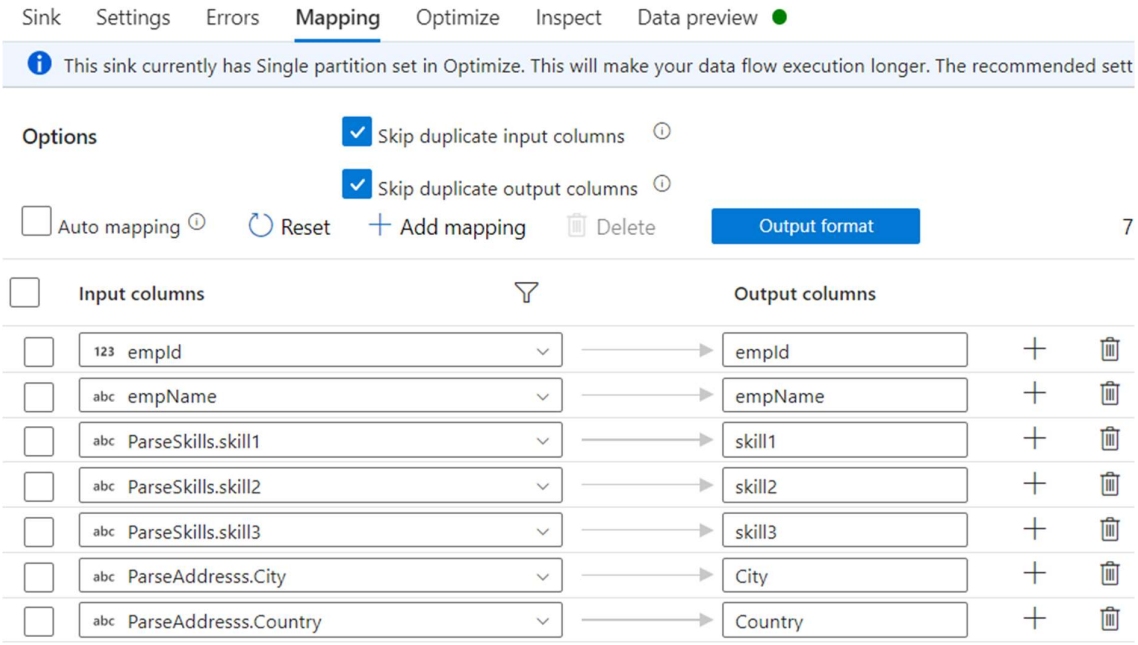
11. Select the target dataset.



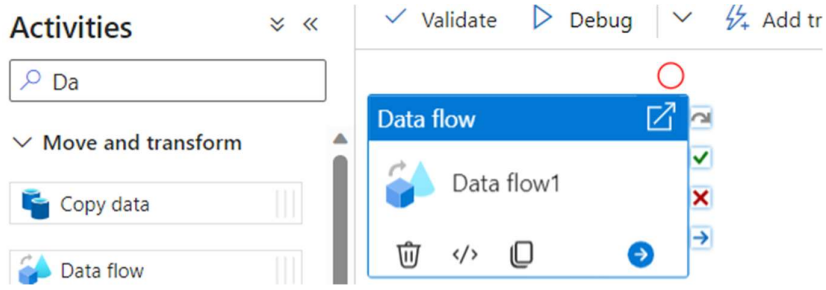
12. Under settings select the file option and file name as shown below.



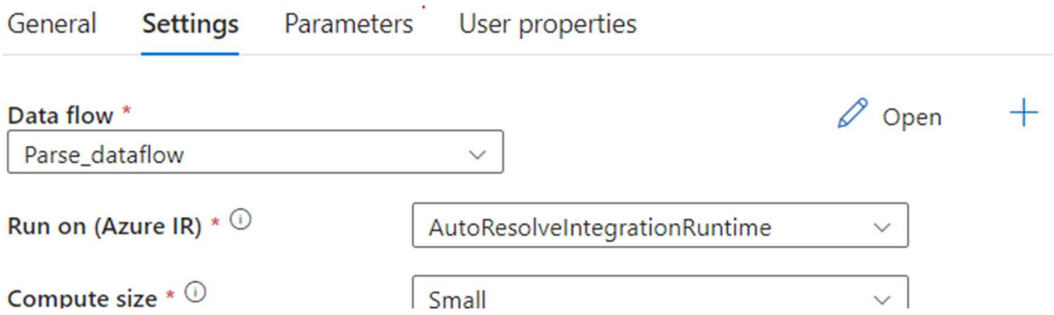
13. Under mapping delete the unnecessary fields and add the below fields as shown below.



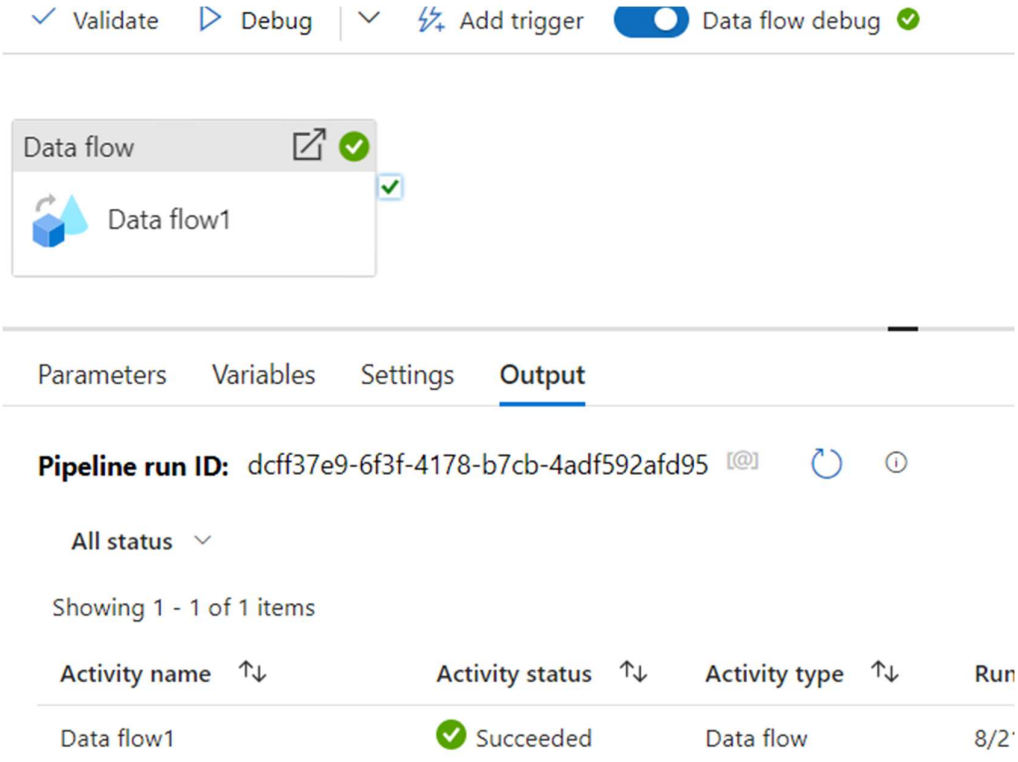
14. Create a Pipeline and drag and drop the Data flow.



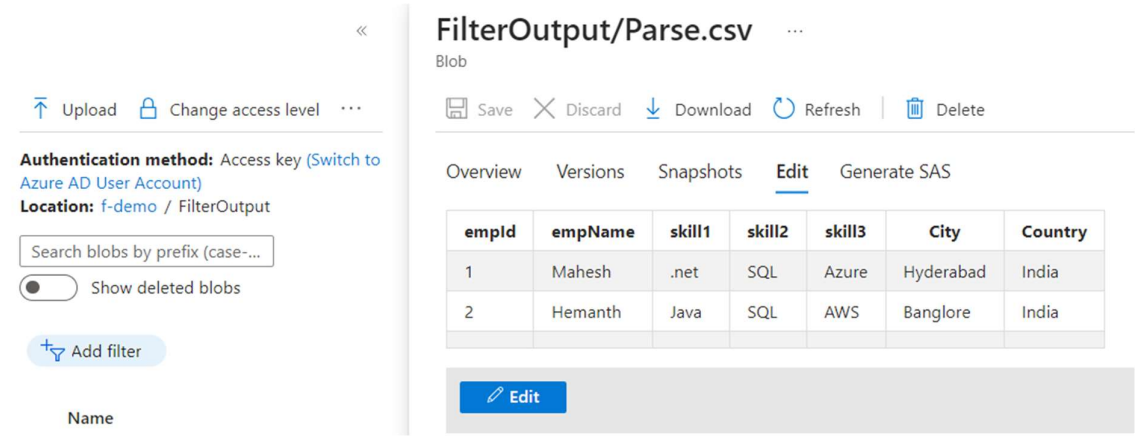
15. Select the Dataflow.



16. Validate the pipeline and click on Debug.
17. Here our pipeline was executed successfully.

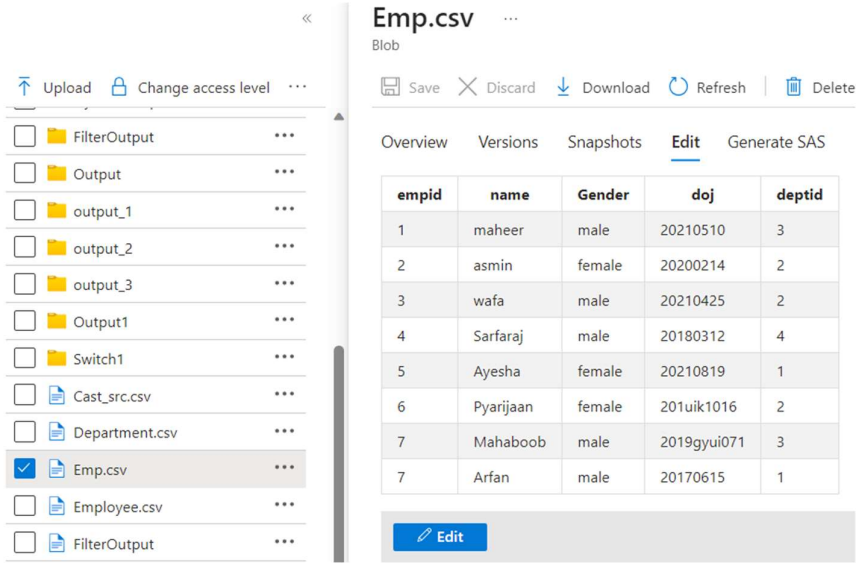


18. Go to Destination and check the target Data.

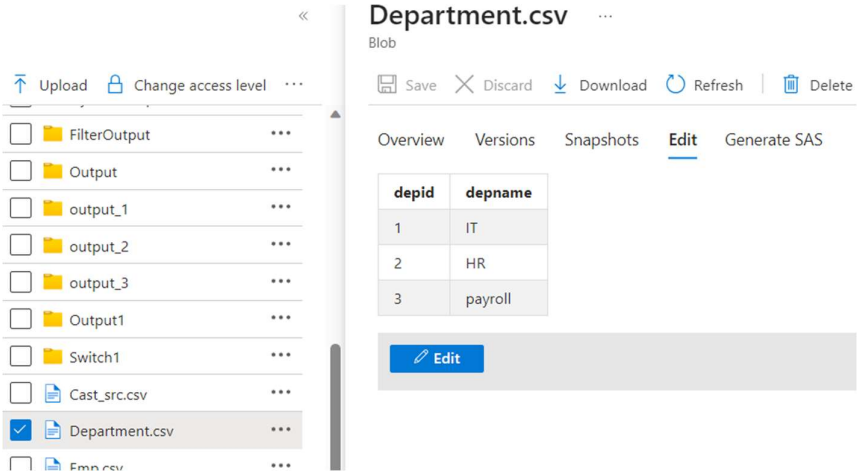


Assert Transformation

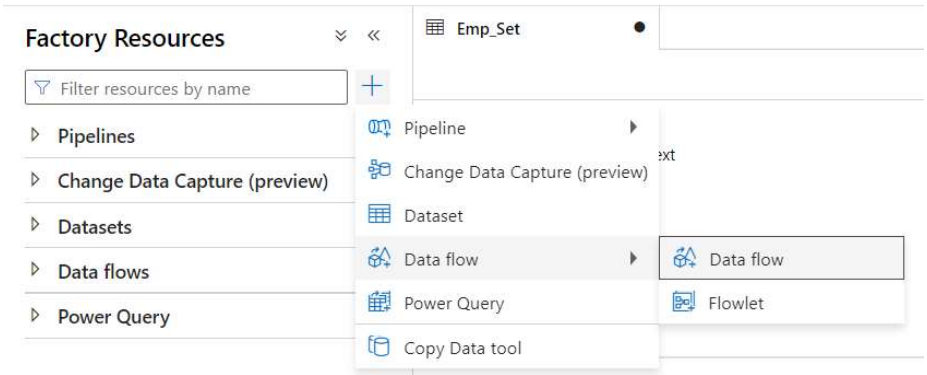
1. For this we are using the below Emp data.



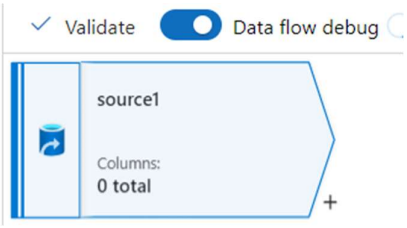
2. And also, below Department data.



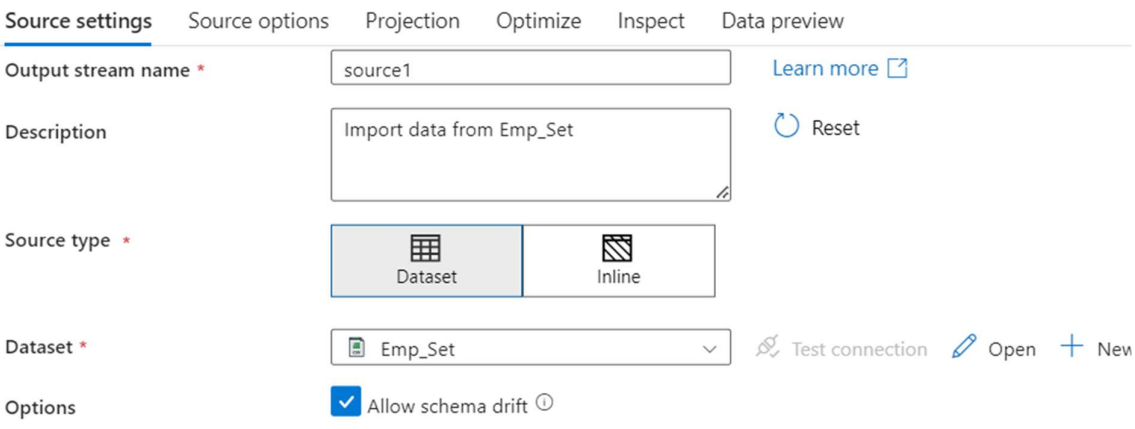
3. Create a Dataset for both files.
4. Create a Dataflow.



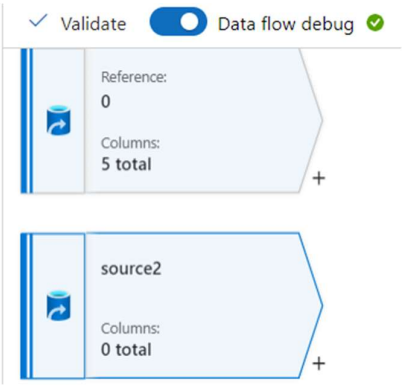
5. Add a Source.



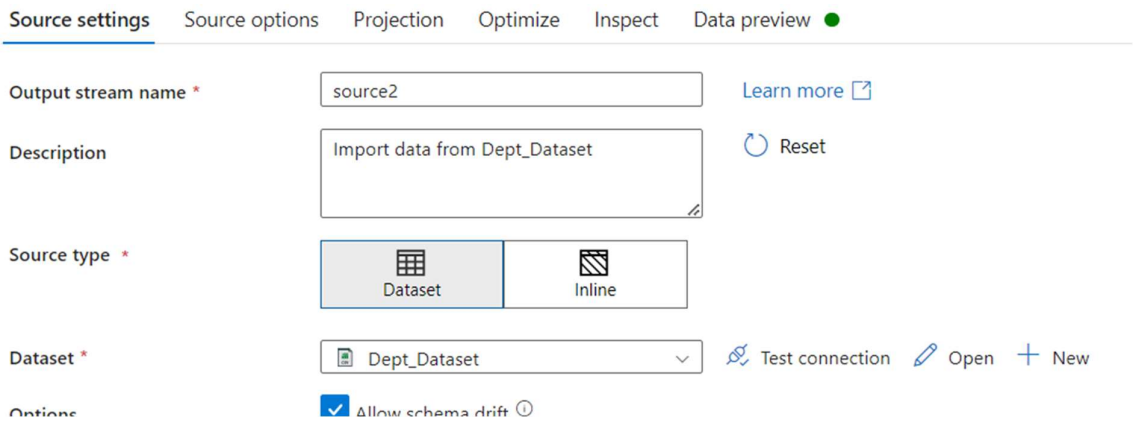
6. Select the Emp source Dataset.



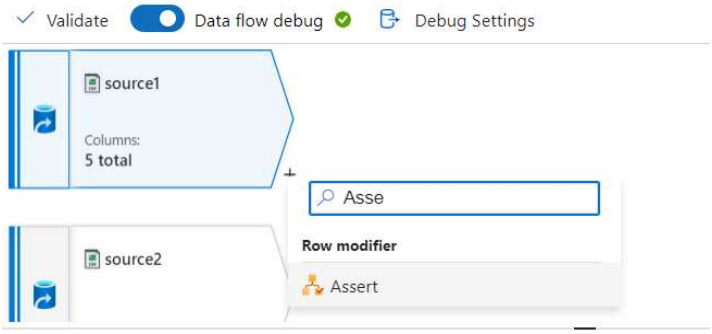
7. Add another Source.



8. Select the Department source dataset.



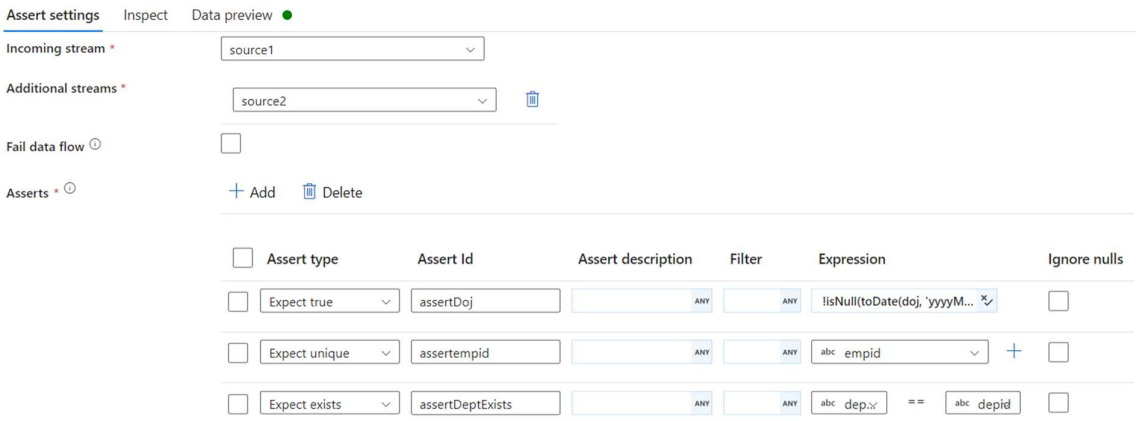
9. Click on the plus symbol and click on Assert as shown below.



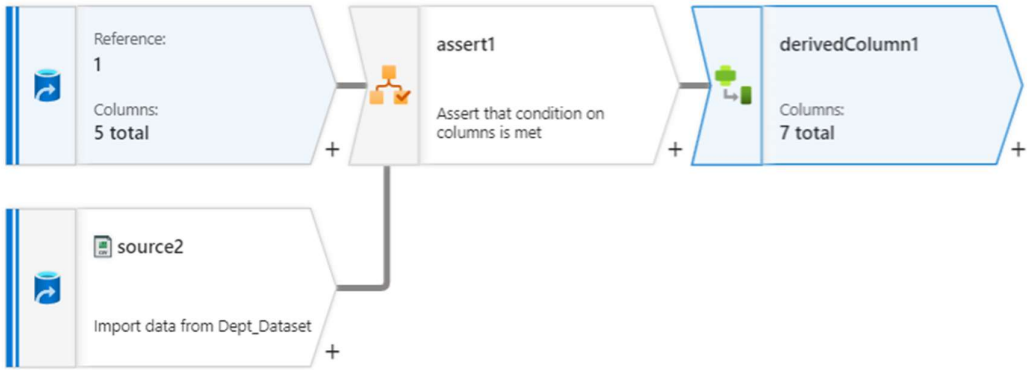
10. Under Assert settings, set the below properties then add below Asserts.
11. For the assertDoj give the below expression.

Exp: !isNull(toDate(doj, 'yyyyMMdd'))

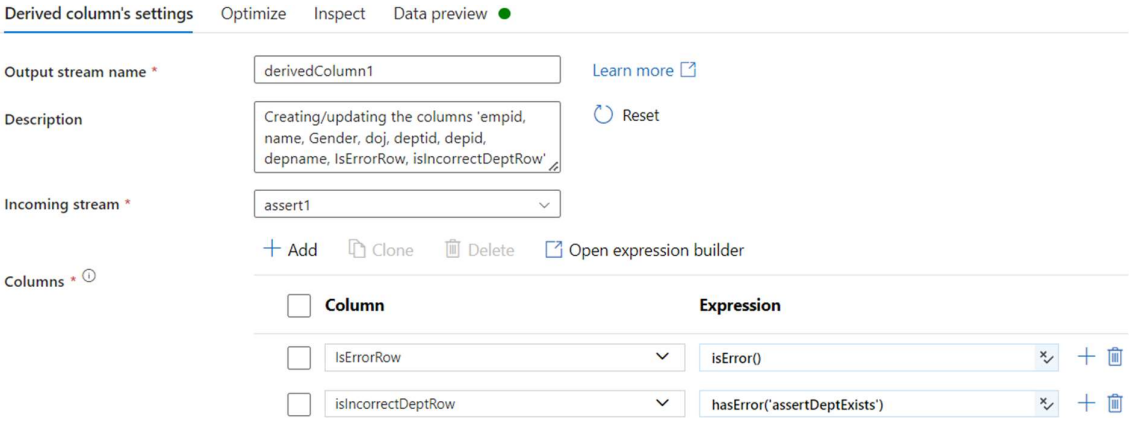
12. For the assertDeptExists select the fields are Deptid and Depid



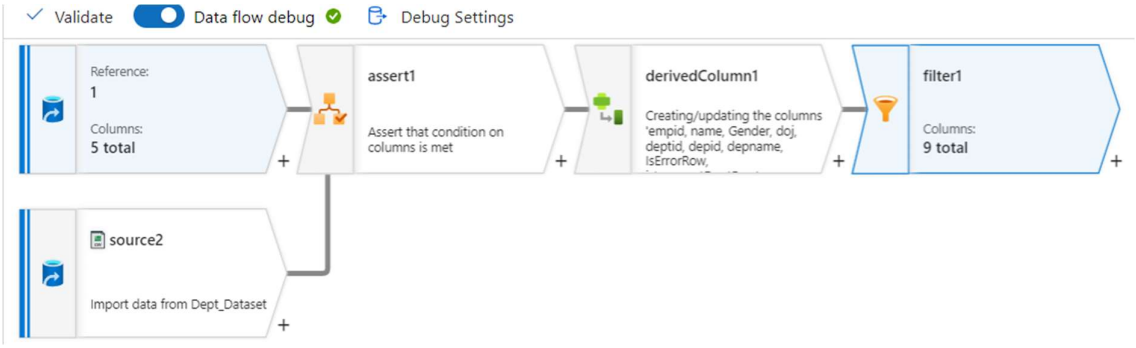
13. Add the Derived column.



14. Under the Derived column's settings add the below columns and conditions.

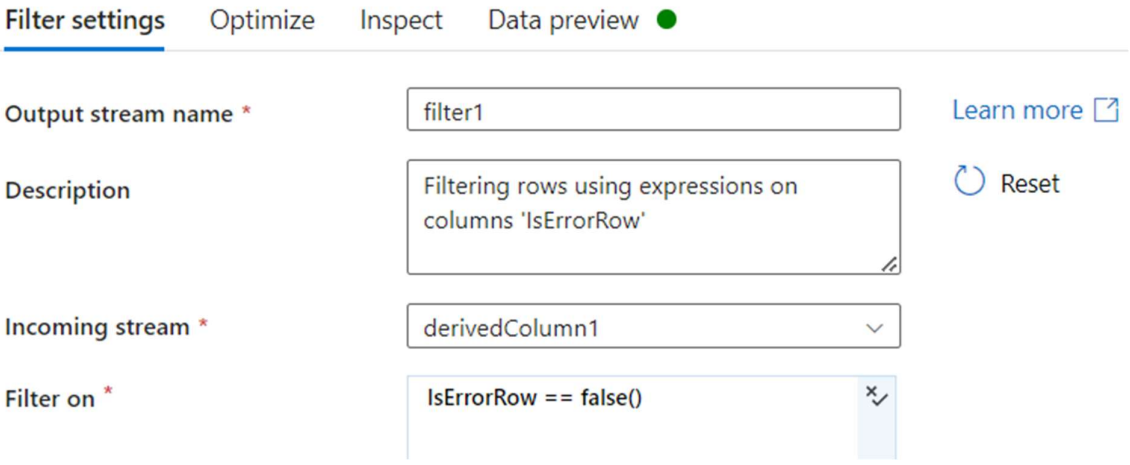


15. Add a Filter.

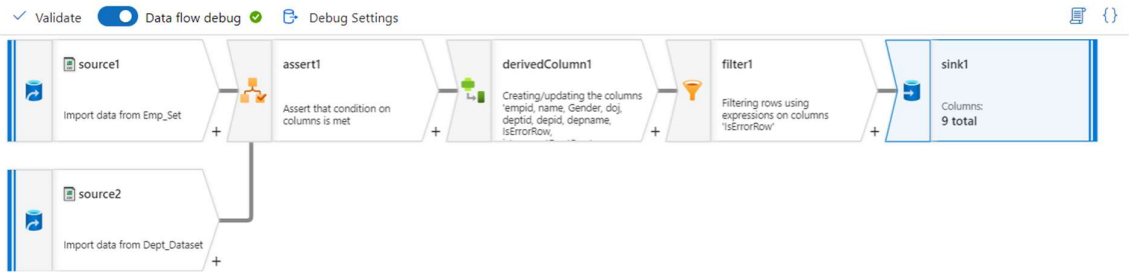


16. Under filter settings give the below Filter on Expression.

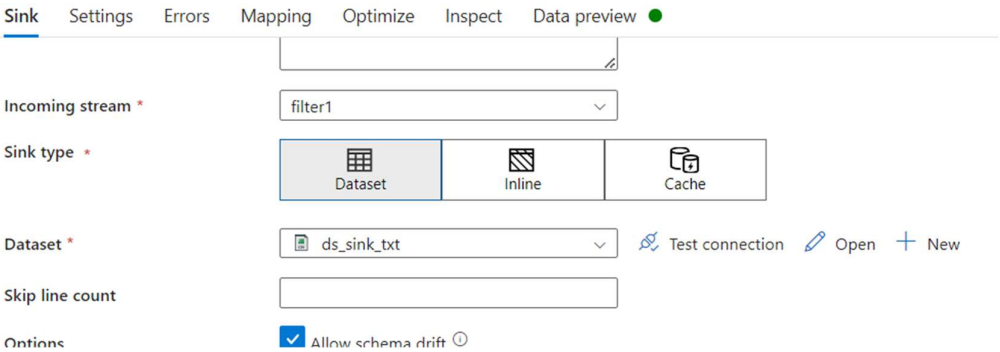
Exp: IsErrorRow == false()



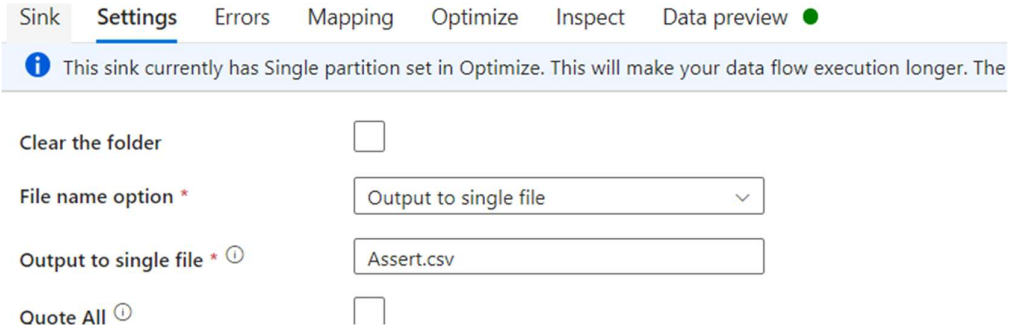
17. Add Sink.



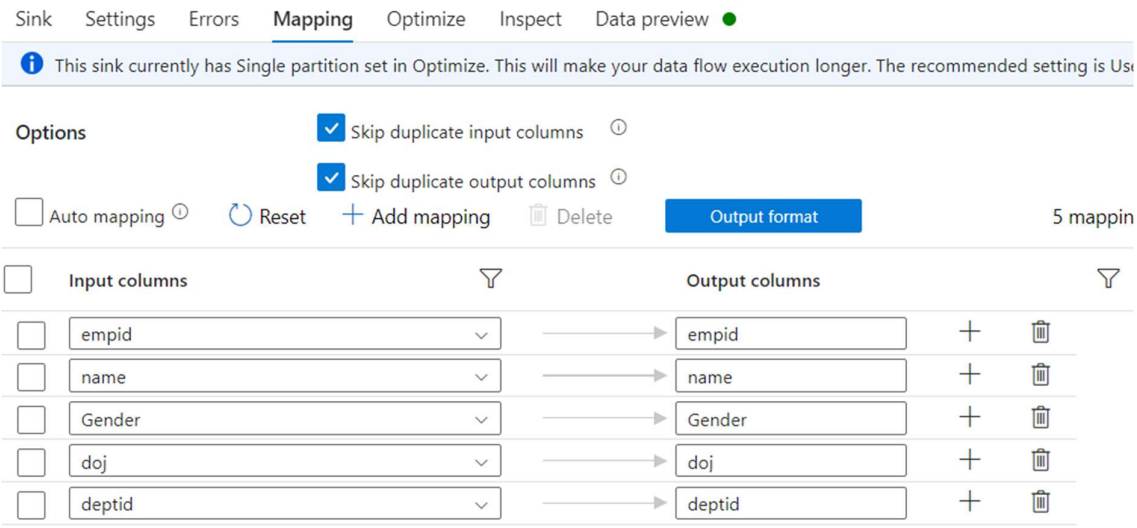
18. Select the Destination Dataset.



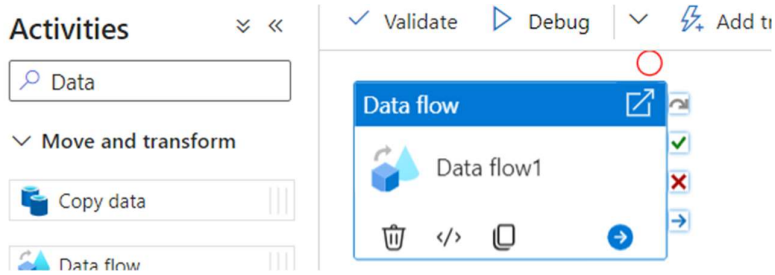
19. Under settings select the File Option and File Name.



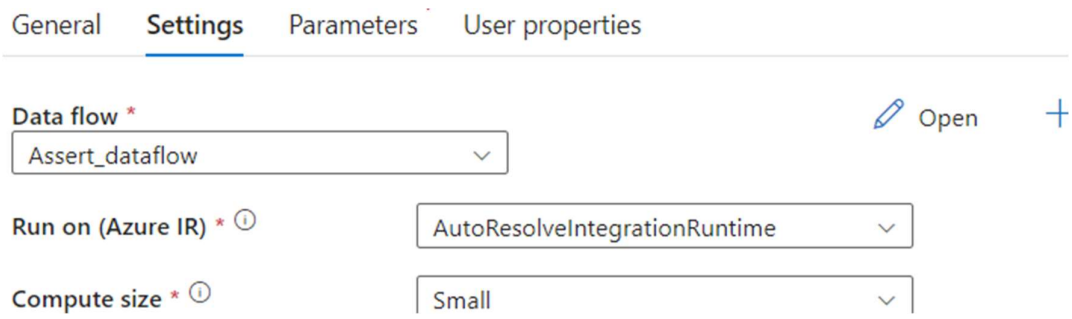
20. Under Mapping keep the below columns and delete others.



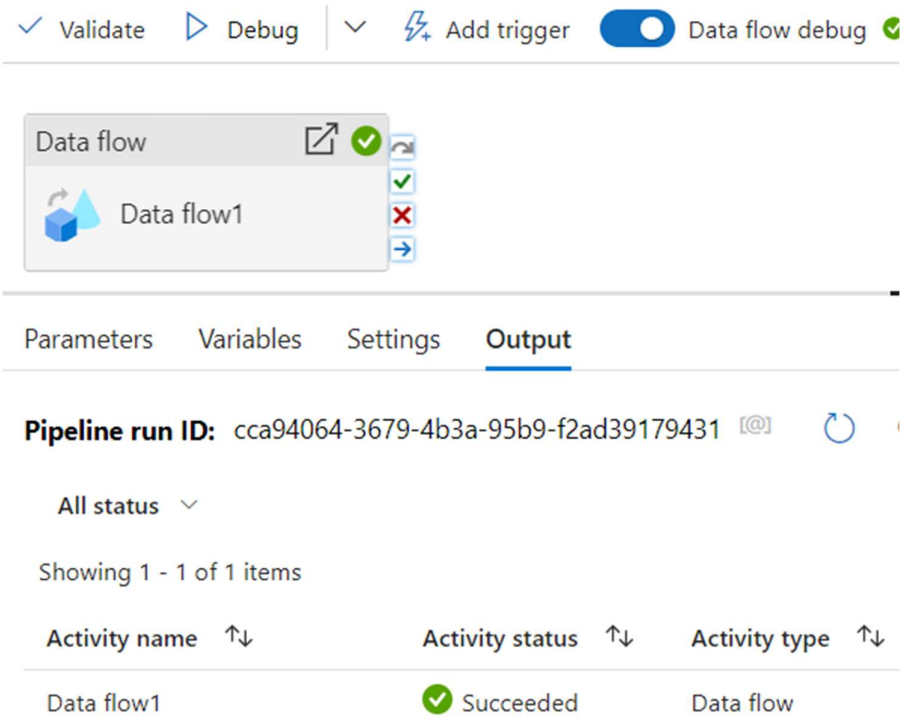
21. Create a Pipeline
22. Drag and drop the Data flow.



23. Under setting select the Dataflow.



24. Validate the pipeline and click on Debug.
25. Here our pipeline is Executed successfully.



26. Go to Destination and check the Data.