😊 Filtering Rows

1. In this lab we are going to filter some of our rows from the Parquet table.
2. As if you will look at the data that we have in Parquet file we can see that in the Resource group column some of the data is Null.
3. So, we used the below command and we get 191 rows that are null

SELECT * FROM [logdata_parquet]
WHERE Resourcegroup IS NULL

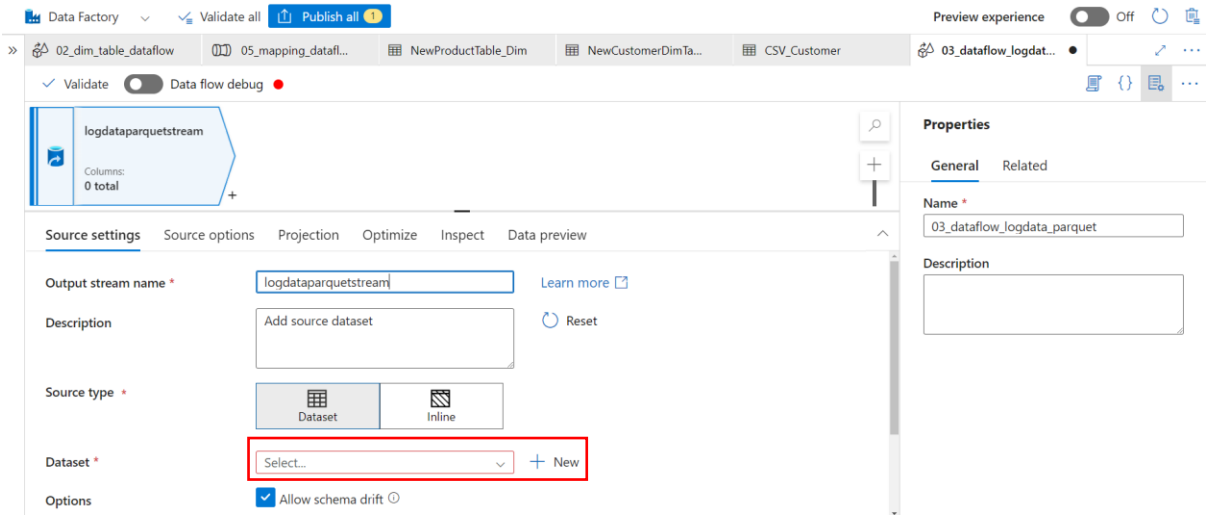
Resourcetype	Resourcegroup	Resource
NULL	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07
NULL	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
NULL	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
Microsoft.RecoveryServices/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
Microsoft.RecoveryServices/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
Microsoft.RecoveryServices/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
Microsoft.RecoveryServices/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
Microsoft.RecoveryServices/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
Microsoft.RecoveryServices/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
Microsoft.RecoveryServices/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
NULL	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
NULL	NULL	/SUBSCRIPTIONS/6912D7A0-BC28-459A-9407-33BBBA641...
Microsoft.RecoveryServices/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
Microsoft.RecoveryServices/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
Microsoft.RecoveryServices/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
Microsoft.RecoveryServices/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
NULL	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
NULL	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
Microsoft.RecoveryServices/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
Microsoft.RecoveryServices/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
Microsoft.RecoveryServices/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
Microsoft.RecoveryServices/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
Microsoft.RecoveryServices/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
Microsoft.Network/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
Microsoft.RecoveryServices/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
Microsoft.RecoveryServices/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
Microsoft.RecoveryServices/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
Microsoft.RecoveryServices/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
NULL	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
Microsoft.RecoveryServices/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
Microsoft.RecoveryServices/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
Microsoft.RecoveryServices/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
Microsoft.RecoveryServices/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro...
Microsoft.RecoveryServices/locations	NULL	/subscriptions/6912d7a0-bc28-459a-9407-33bbba641c07/pro

datasynapse1234.sql.azuresy... | sqladminuser (0) | demodb | 00:00:01 | 191 rows

4. As you can see below this is the data which we'll get when we remove the Null values.

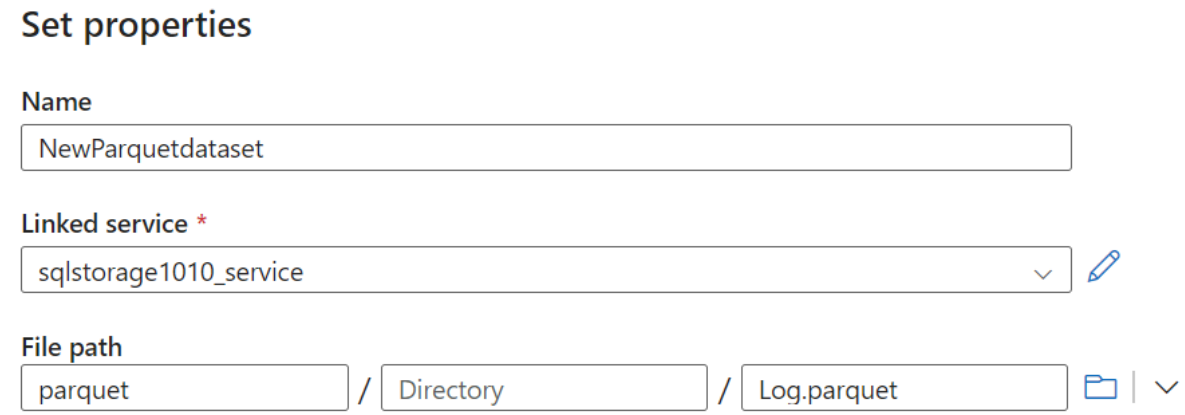
```
-- Original Table - 19,229 rows
-- With NULL values - 191
-- Rows in the end - 19,038 rows
```

5. Now move to Data Factory wizard and create a new data flow.
6. Here first we will give name to our data flow then we choose to create a new resource then name it.
7. After that you need to choose the dataset.

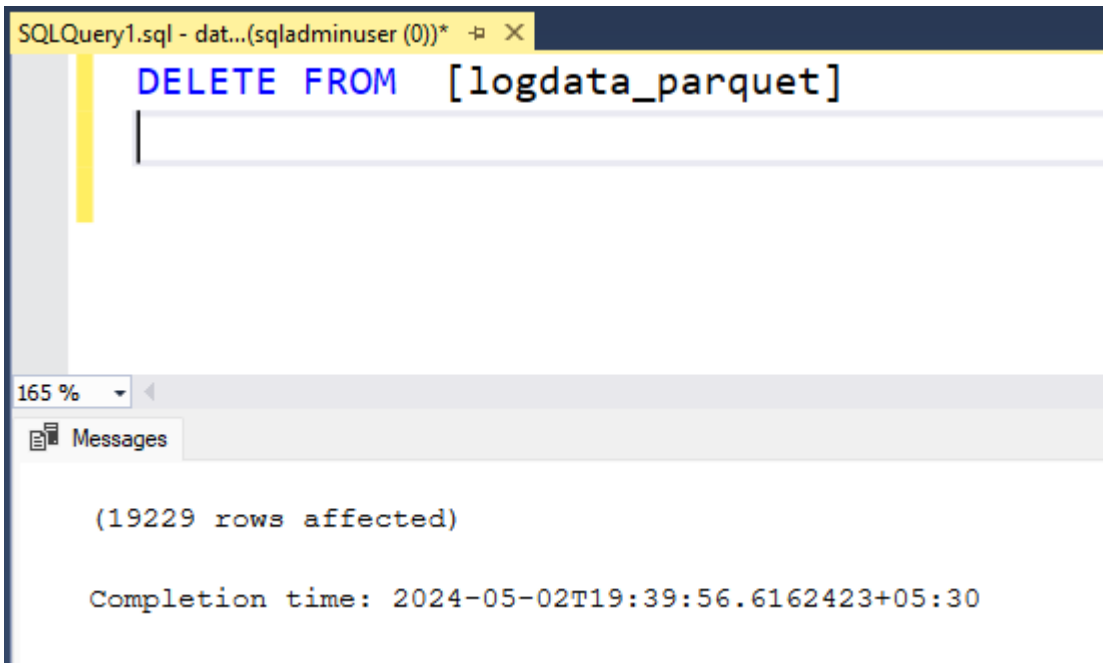


8. Now we need to create the dataset for that click on new.

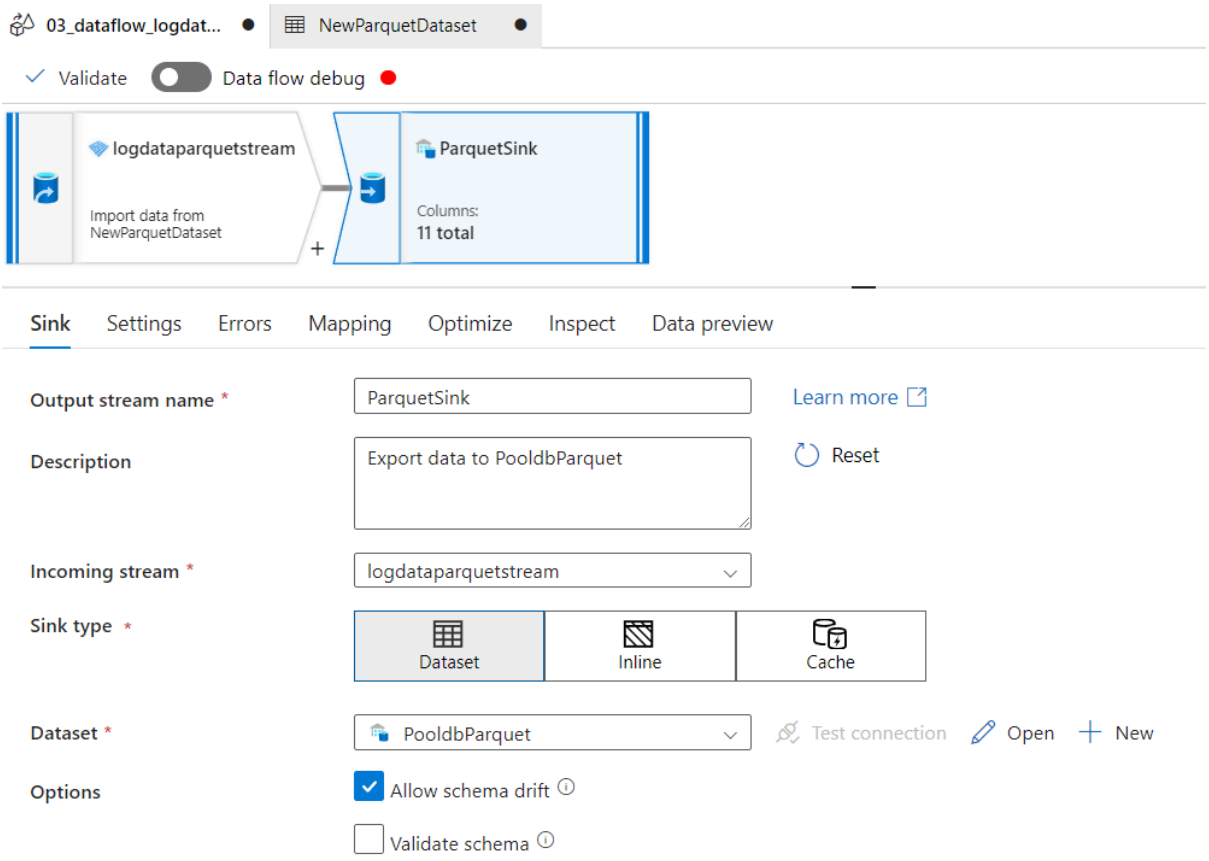
9. Then you have to choose your Azure data lake gen2 storage account, then you need to choose Parquet as your file format and after that in the set properties option you need to give it a name choose your linked service and give the file path for your parquet container and the file stored in it.



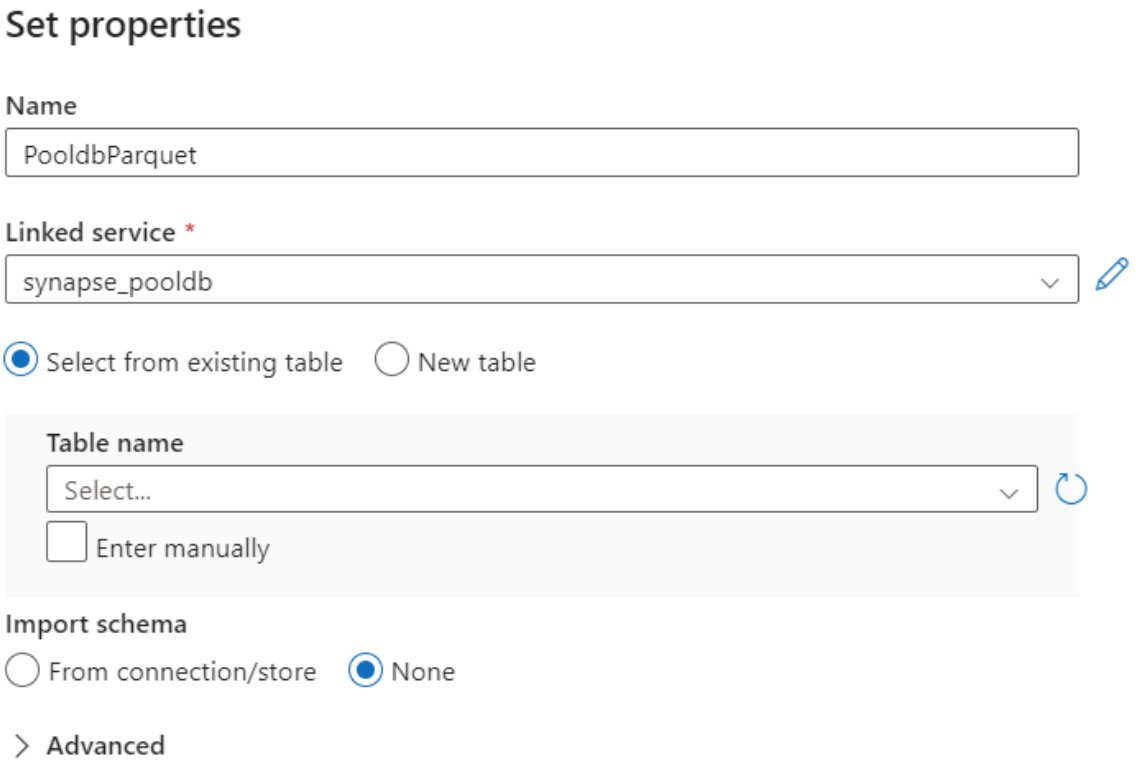
10. After that go back to SSMS and delete the data in your Parquet file.



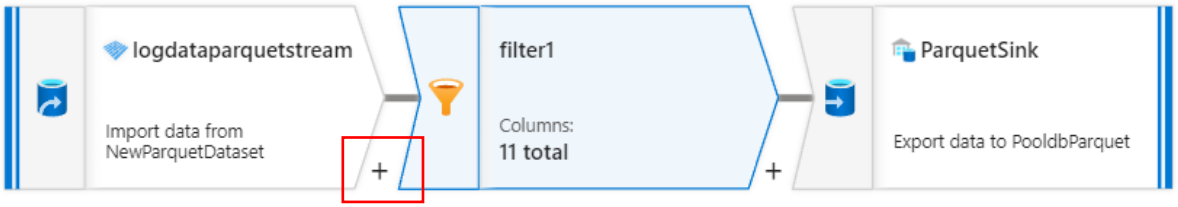
11. Then you need to create a sink in the data flow.
12. Now in your sink give it a name then for the dataset click on new and then choose Synapse Analytics.



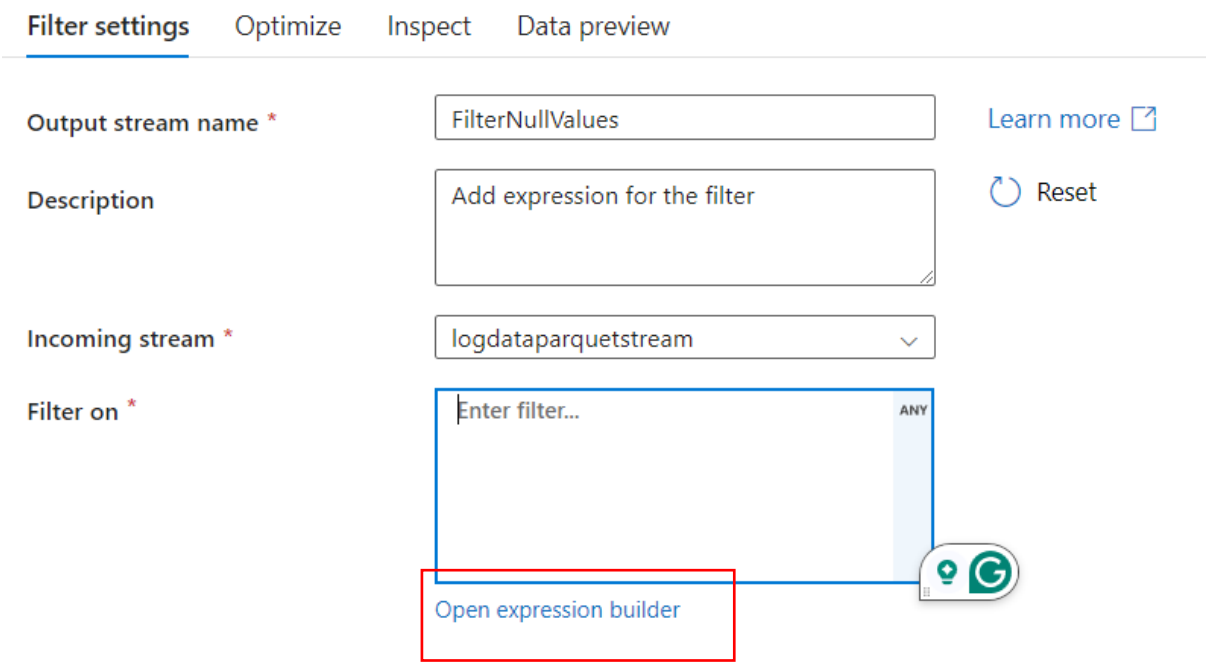
13. After that in the set properties you need to give it a name then in the linked service choose your pool db. Then you have to choose your parquet table.



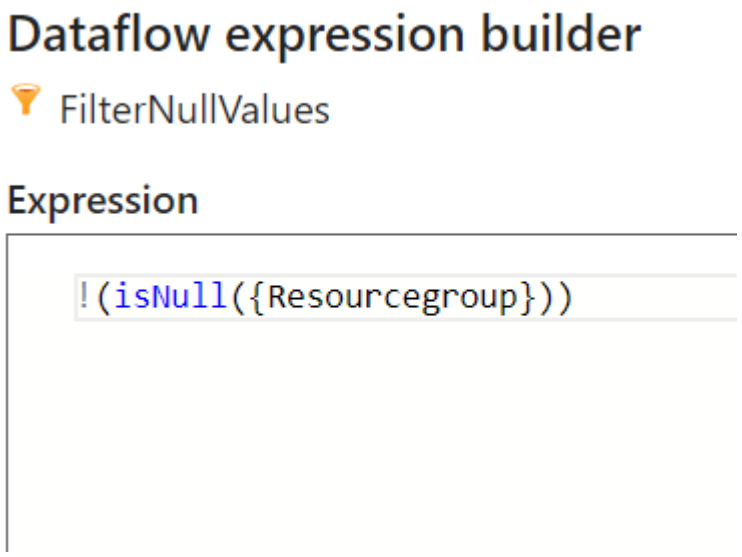
14. Then click on the highlighted plus icon in your dataflow and choose filter option from the drop down menu.



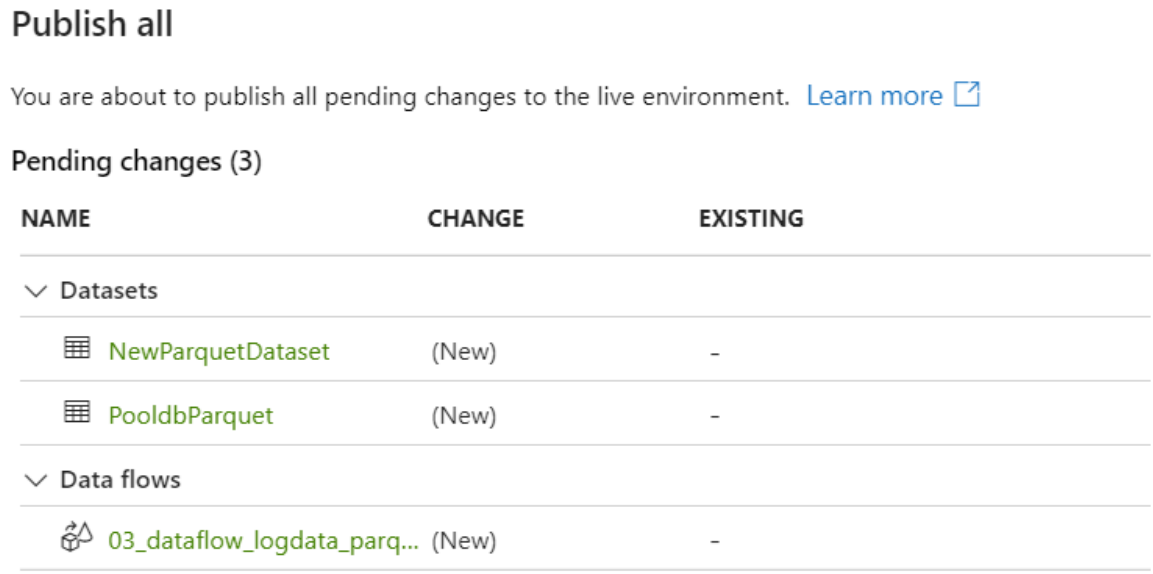
15. Then give your filter a name and in the filter on option click on open expression builder.



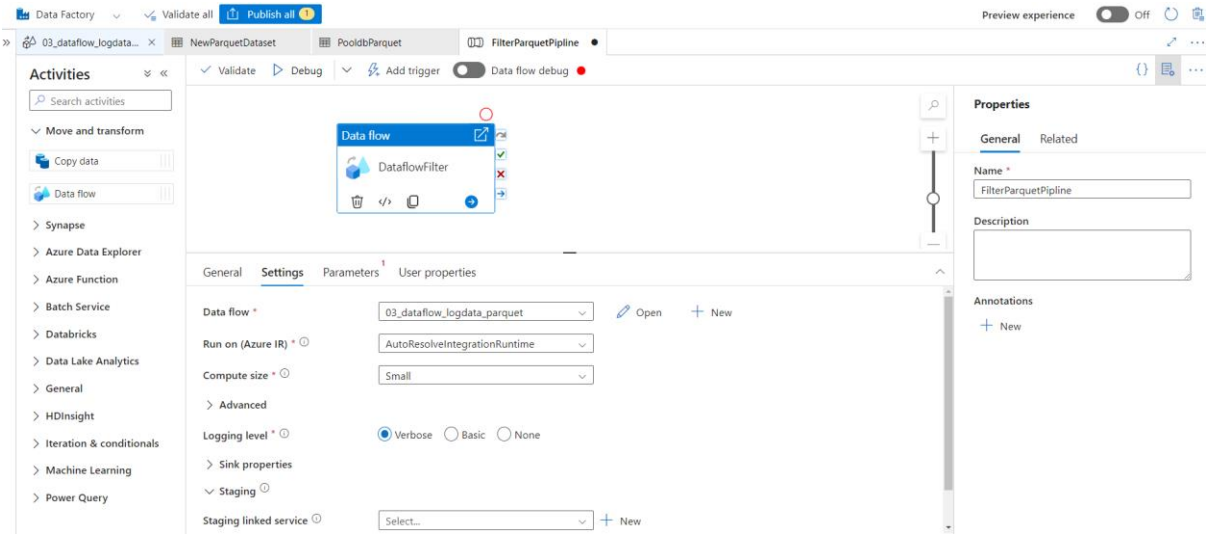
16. In the expression builder you have to write the same expression as shown below then click on save and finish.



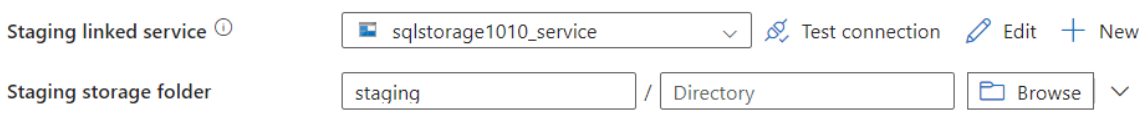
17. Then just click on publish all and then go to the pipelines and create a new pipeline.



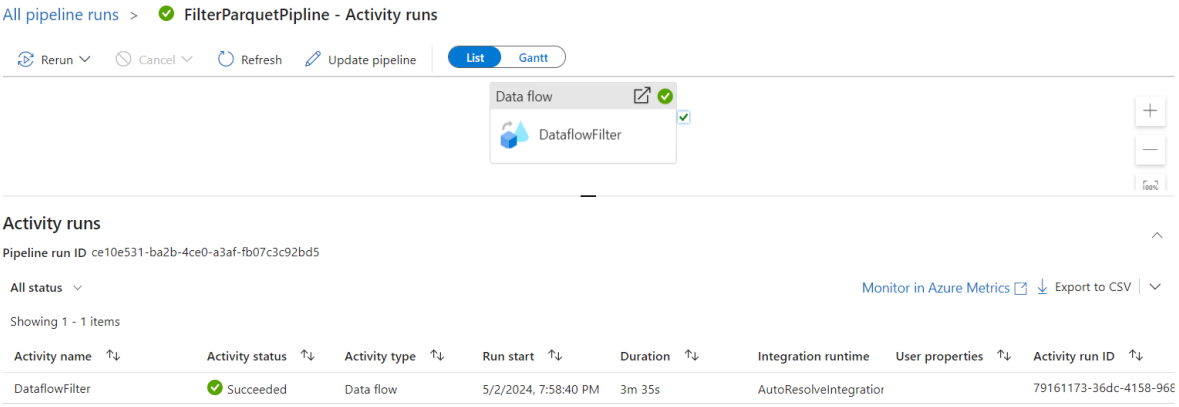
18. Create a pipeline for dataflow and then give it a name then choose your dataflow.



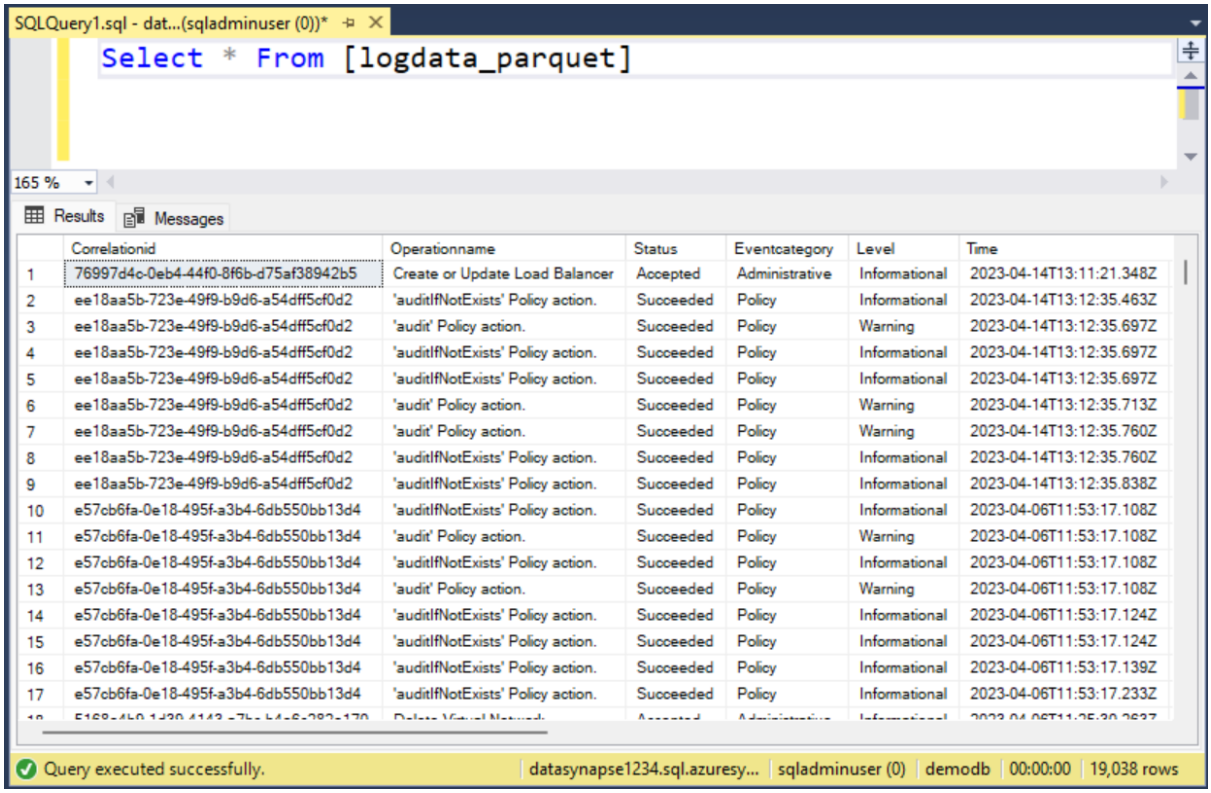
19. In staging choose your linked service and the container.



20. After that just publish your pipeline and click on trigger now.
21. Below you can see that our pipeline run was successful.



22. Now if we go to our SSMS and run the select statement for the parquet file then we'll get the filtered data.
23. As expected, we get the data and the number of rows are 19038



	Correlationid	Operationname	Status	Eventcategory	Level	Time
1	76997d4c-0eb4-44f0-8f6b-d75af38942b5	Create or Update Load Balancer	Accepted	Administrative	Informational	2023-04-14T13:11:21.348Z
2	ee18aa5b-723e-49f9-b9d6-a54dff5cf0d2	'auditIfNotExists' Policy action.	Succeeded	Policy	Informational	2023-04-14T13:12:35.463Z
3	ee18aa5b-723e-49f9-b9d6-a54dff5cf0d2	'audit' Policy action.	Succeeded	Policy	Warning	2023-04-14T13:12:35.697Z
4	ee18aa5b-723e-49f9-b9d6-a54dff5cf0d2	'auditIfNotExists' Policy action.	Succeeded	Policy	Informational	2023-04-14T13:12:35.697Z
5	ee18aa5b-723e-49f9-b9d6-a54dff5cf0d2	'auditIfNotExists' Policy action.	Succeeded	Policy	Informational	2023-04-14T13:12:35.697Z
6	ee18aa5b-723e-49f9-b9d6-a54dff5cf0d2	'audit' Policy action.	Succeeded	Policy	Warning	2023-04-14T13:12:35.713Z
7	ee18aa5b-723e-49f9-b9d6-a54dff5cf0d2	'audit' Policy action.	Succeeded	Policy	Warning	2023-04-14T13:12:35.760Z
8	ee18aa5b-723e-49f9-b9d6-a54dff5cf0d2	'auditIfNotExists' Policy action.	Succeeded	Policy	Informational	2023-04-14T13:12:35.760Z
9	ee18aa5b-723e-49f9-b9d6-a54dff5cf0d2	'auditIfNotExists' Policy action.	Succeeded	Policy	Informational	2023-04-14T13:12:35.838Z
10	e57cb6fa-0e18-495f-a3b4-6db550bb13d4	'auditIfNotExists' Policy action.	Succeeded	Policy	Informational	2023-04-06T11:53:17.108Z
11	e57cb6fa-0e18-495f-a3b4-6db550bb13d4	'audit' Policy action.	Succeeded	Policy	Warning	2023-04-06T11:53:17.108Z
12	e57cb6fa-0e18-495f-a3b4-6db550bb13d4	'auditIfNotExists' Policy action.	Succeeded	Policy	Informational	2023-04-06T11:53:17.108Z
13	e57cb6fa-0e18-495f-a3b4-6db550bb13d4	'audit' Policy action.	Succeeded	Policy	Warning	2023-04-06T11:53:17.108Z
14	e57cb6fa-0e18-495f-a3b4-6db550bb13d4	'auditIfNotExists' Policy action.	Succeeded	Policy	Informational	2023-04-06T11:53:17.124Z
15	e57cb6fa-0e18-495f-a3b4-6db550bb13d4	'auditIfNotExists' Policy action.	Succeeded	Policy	Informational	2023-04-06T11:53:17.124Z
16	e57cb6fa-0e18-495f-a3b4-6db550bb13d4	'auditIfNotExists' Policy action.	Succeeded	Policy	Informational	2023-04-06T11:53:17.139Z
17	e57cb6fa-0e18-495f-a3b4-6db550bb13d4	'auditIfNotExists' Policy action.	Succeeded	Policy	Informational	2023-04-06T11:53:17.233Z