😊 Copy data using the Copy Command

In these labs, the primary objective is to transfer data from various sources into a dedicated SQL pool within Azure Synapse Analytics. Two different methods are explored: the Copy Command and Polybase.

First, using the Copy Command, data is copied into the SQL pool. Challenges arise due to the presence of an additional column, requiring adjustments to the pipeline and SQL table schema. Once resolved, the data is successfully transferred.

Next, Polybase is employed for data transfer, offering parallel processing capabilities for enhanced efficiency. Staging configurations are enabled, necessitating the creation of a storage container for temporary data storage. After configuring the pipeline, data is transferred using Polybase, enabling faster parallel data movement.

The end goal of these labs is to provide practical experience in data transfer operations within Azure Synapse Analytics, showcasing different techniques and their respective advantages. By completing these exercises, users gain insights into data transfer processes and acquire skills in configuring pipelines for efficient data movement into Azure Synapse Analytics.

😄 To begin with the Lab:

1. In this lab we will use the copy command to copy our data into our dedicated SQL Pool. So, again we will make use of the same pipeline.
2. First, we will issue the delete command to delete all the data from our rows.

```
DELETE FROM [logdata_parquet]
```

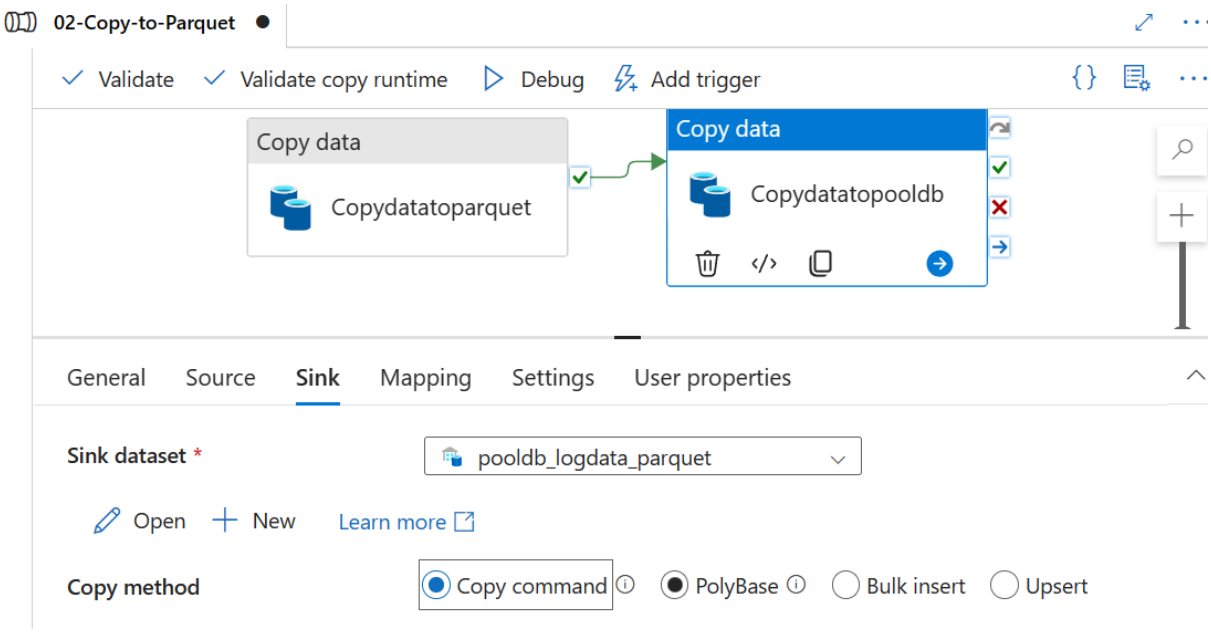
150 %

Messages

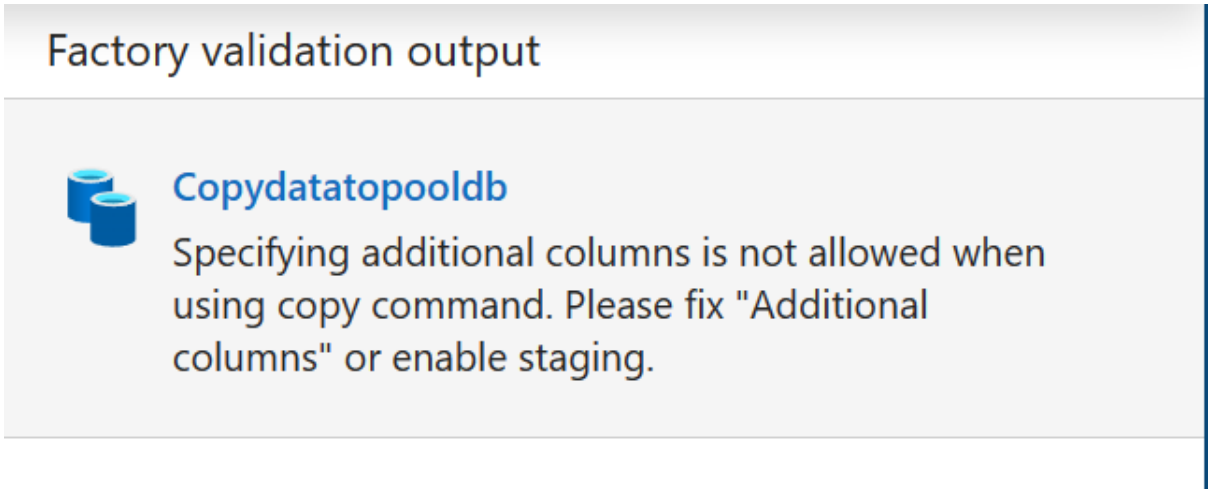
```
(19229 rows affected)

Completion time: 2024-04-24T17:38:59.4185057+05:30
```

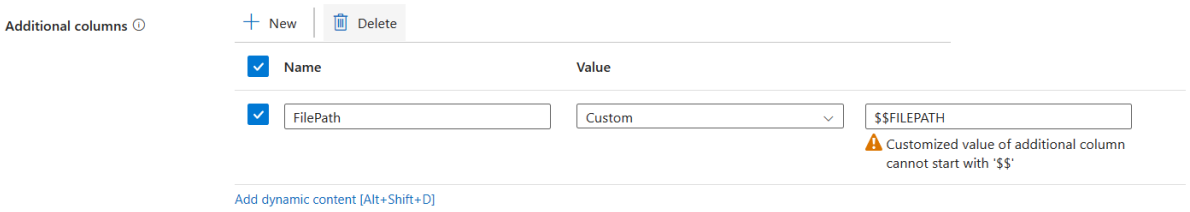
3. Then go back to Data Factory and select your pipeline then go to sink and in copy method choose copy command.



4. After that you have to click on validate all. Below you can see that our validation has failed. Because it says that specifying additional columns is not allowed when using the copy command.



5. So, first we are going to delete our additional column. For that go to source and scroll down to bottom then select your column and hit delete.



6. From our pipeline we have deleted that additional column, but it is still present on our table so, we have to drop our table and then recreate it.
7. Below you can see that we have recreated our table without a file path column.

```
DROP TABLE [logdata_parquet]

CREATE TABLE [logdata_parquet]
(
    [Correlationid] [varchar](200) NULL,
    [Operationname] [varchar](200) NULL,
    [Status] [varchar](100) NULL,
    [Eventcategory] [varchar](100) NULL,
    [Level] [varchar](100) NULL,
    [Time] [datetime] NULL,
    [Subscription] [varchar](200) NULL,
    [Eventinitiatedby] [varchar](1000) NULL,
    [Resourcetype] [varchar](1000) NULL,
    [Resourcegroup] [varchar](1000) NULL,
    [Resource] [varchar](2000) NULL
)
```

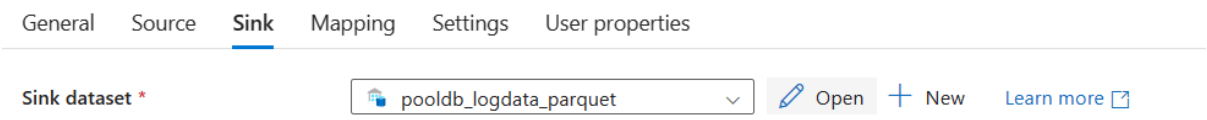
150 %

Messages

```
Commands completed successfully.

Completion time: 2024-04-24T17:46:08.3778689+05:30
```

8. Then again, the steps are similar to before, go to sink and click on open.



9. Then go onto the schema and click on import schema. There you will see the updated schema without that additional column.

Connection **Schema** Parameters

Import schema Clear

Column name	Type
Correlationid	varchar
Operationname	varchar
Status	varchar
Eventcategory	varchar
Level	varchar
Time	datetime
Subscription	varchar
Eventinitiatedby	varchar
Resourcetype	varchar
Resourcegroup	varchar
Resource	varchar

10. After that go to mapping, first click on clear and clear out your mappings then click on import schemas and there you will see your updated schemas.

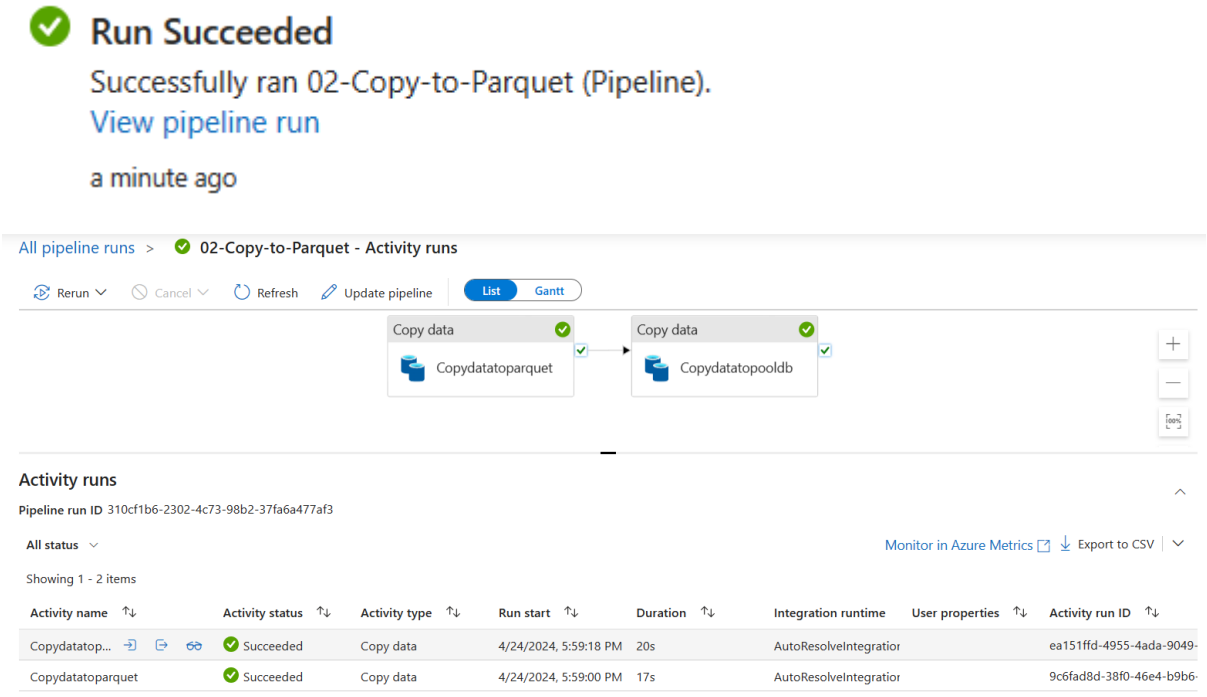
General Source Sink **Mapping** Settings User properties

Import schemas Preview source New mapping Clear Reset Delete

Source	Type	Destination	Type
Correlationid	abc UTF8	Correlationid	abc varchar
Operationname	abc UTF8	Operationname	abc varchar
Status	abc UTF8	Status	abc varchar
Eventcategory	abc UTF8	Eventcategory	abc varchar
Level	abc UTF8	Level	abc varchar
Time	abc UTF8	Time	datetime
Subscription	abc UTF8	Subscription	abc varchar
Eventinitiatedby	abc UTF8	Eventinitiatedby	abc varchar
Resourcetype	abc UTF8	Resourcetype	abc varchar
Resourcegroup	abc UTF8	Resourcegroup	abc varchar
Resource	abc UTF8	Resource	abc varchar

Add dynamic content [Alt+Shift+D]

11. Once it is done then click on validate all, after that click on publish all.
12. After publishing you have to trigger your pipeline.
13. Then wait for some time. After that, you see your pipeline too in the monitor section.



14. Then run the Select statement to view your data.

```
select * from [logdata_parquet]
```

150 %

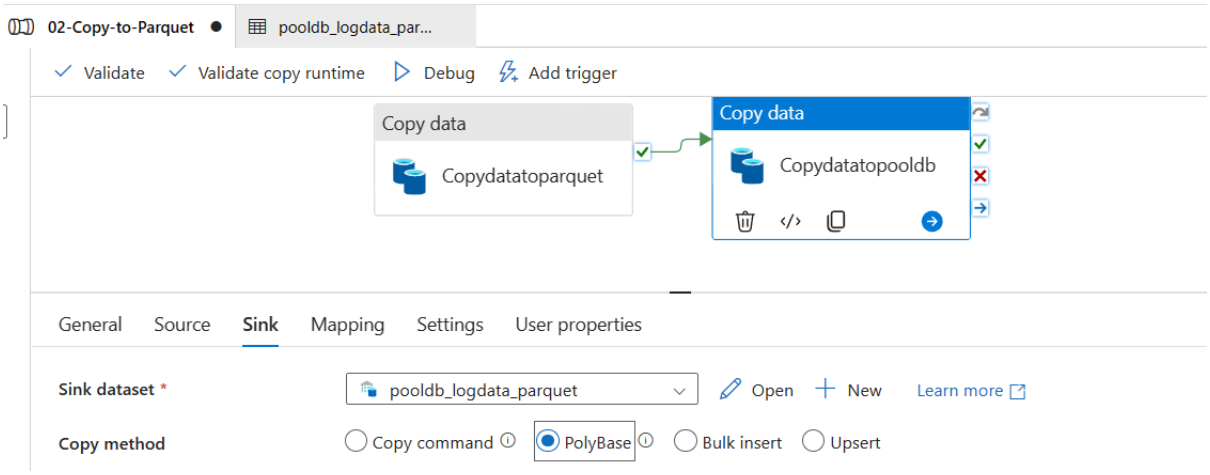
Results | Messages

	Correlationid	Operationname	Status	Eventcategory	Level	Time	Subscription
1	a4927ba1-fe0e-460b-8c0c-8d78718ee847	'audit' Policy action.	Succeeded	Policy	Warning	2023-04-17T12:59:00.470Z	6912d7a0-bc28-459a-9407-33bbba641c07
2	99fe9c3a-e36e-44e0-acd4-58272ab10c7e	Update SQL database	Succeeded	Administrative	Informational	2023-04-25T03:36:59.503Z	6912d7a0-bc28-459a-9407-33bbba641c07
3	b5247d6a-f509-41a1-9ae7-05a8fc566e42	Validate Deployment	Started	Administrative	Informational	2023-04-25T03:23:48.335Z	6912d7a0-bc28-459a-9407-33bbba641c07
4	3cbc1247-a562-4336-885c-e5be7045a10c	Delete Virtual Network	Failed	Administrative	Error	2023-04-17T15:49:49.125Z	6912d7a0-bc28-459a-9407-33bbba641c07
5	87b423cf-9aa3-44b7-8649-05ee4da94c8d	Create Deployment	Succeeded	Administrative	Informational	2023-04-17T15:43:51.854Z	6912d7a0-bc28-459a-9407-33bbba641c07
6	84fb2d66-c063-4573-b973-e3709d781462	'auditIfNotExists' Policy action.	Succeeded	Policy	Informational	2023-04-17T15:32:39.557Z	6912d7a0-bc28-459a-9407-33bbba641c07
7	304f186a-198f-4c52-8a1c-1baa98oec9cd	Create Deployment	Succeeded	Administrative	Informational	2023-04-17T05:32:38.445Z	6912d7a0-bc28-459a-9407-33bbba641c07
8	f4c54d30-0b68-4c60-94fc-69349e63b788	Create or Update Disk	Accepted	Administrative	Informational	2023-04-17T05:34:01.966Z	6912d7a0-bc28-459a-9407-33bbba641c07
9	70d4f886-7d32-42d8-a191-2183f22aecdc	Delete Virtual Network	Accepted	Administrative	Informational	2023-04-16T14:14:04.290Z	6912d7a0-bc28-459a-9407-33bbba641c07
10	ef83d4df-1952-4490-bce2-7d0bd65b9b69	Delete Disk	Failed	Administrative	Error	2023-04-16T14:12:54.409Z	6912d7a0-bc28-459a-9407-33bbba641c07
11	420e3dc7-c154-49b1-9e49-f2f44de74cd6	Create or Update Disk	Succeeded	Administrative	Informational	2023-04-16T06:14:39.816Z	6912d7a0-bc28-459a-9407-33bbba641c07
12	75887772-299d-432e-9827-86b2260c351b	'audit' Policy action.	Succeeded	Policy	Warning	2023-04-14T05:39:22.099Z	6912d7a0-bc28-459a-9407-33bbba641c07

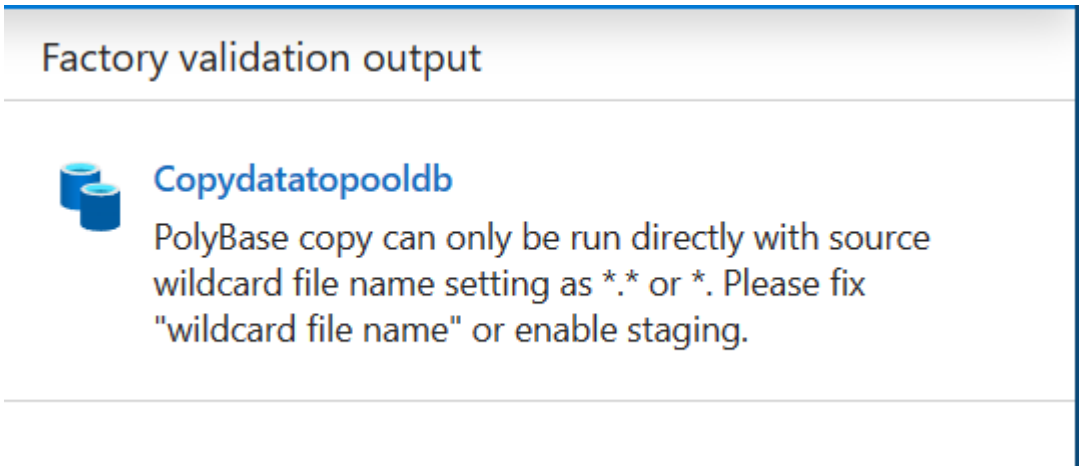
Query executed successfully.

😄 Copy data using Polybase

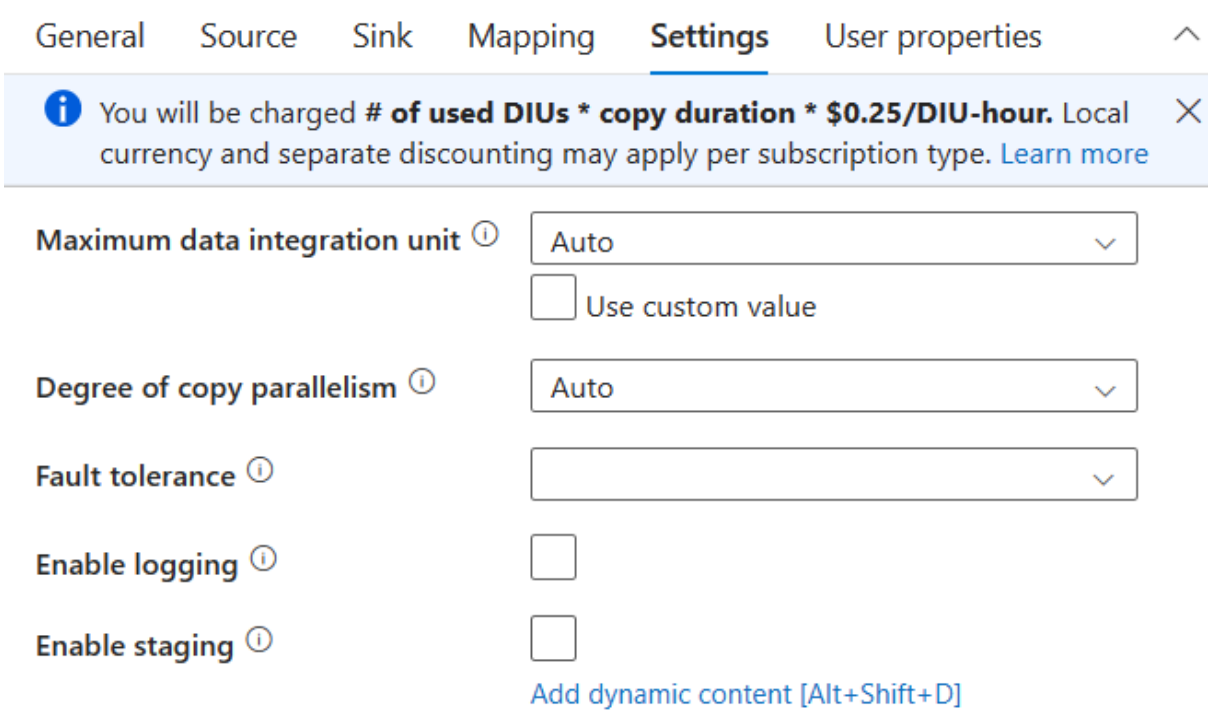
1. This time we will be using Polybase to modify our pipeline.
2. Now come back to your pipeline. In terms of sink we will choose Polybase.



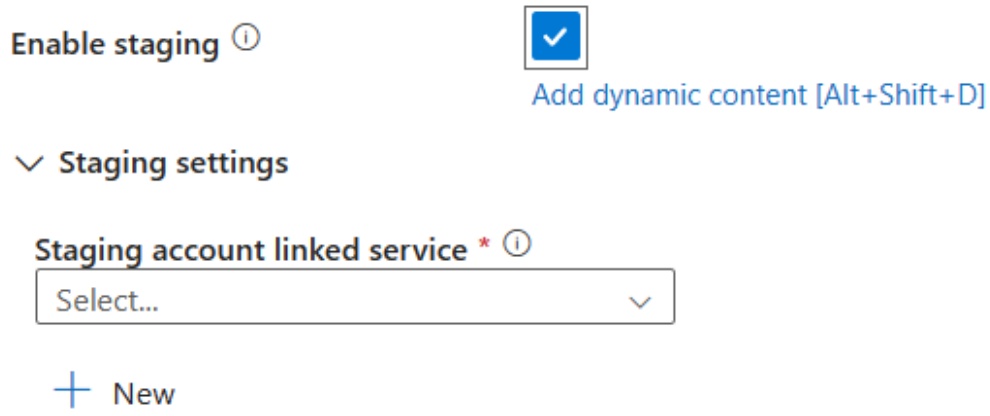
3. Then click on validate. You will get an error.



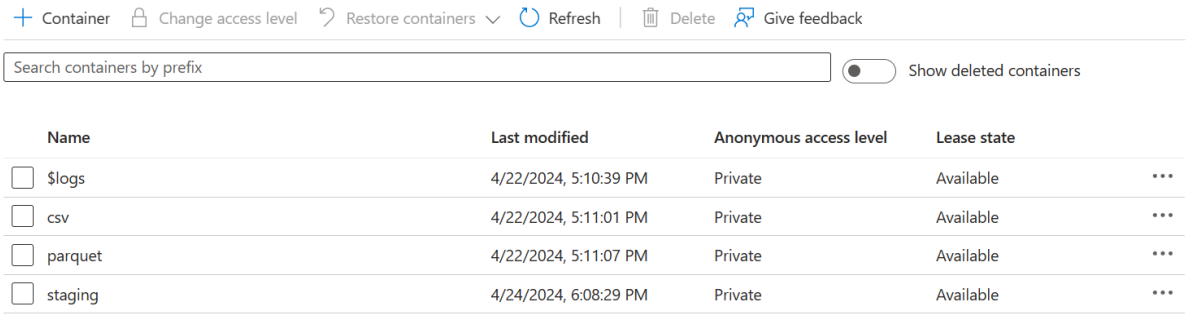
4. To get rid of this error we will use staging. So go to settings and enable staging.



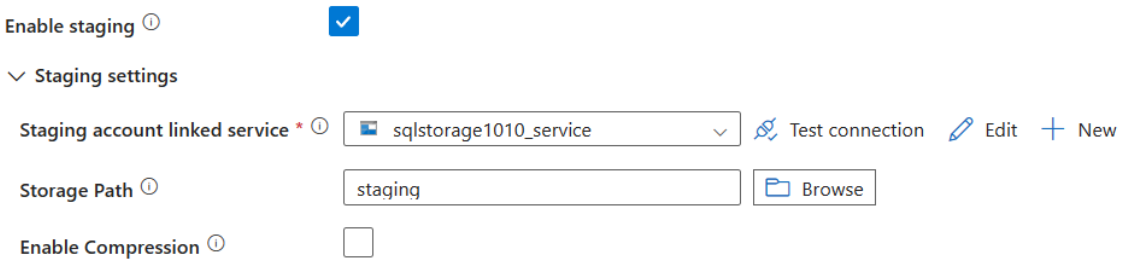
5. For staging, it needs to have a storage account in place. So, this is going to be a temporary place for the entire transfer activity.



6. Now go back to your storage account and navigate to containers. There you have to create a new container for staging.



7. In terms of linked service we will choose our existing service and in the path we will choose our newly created container.



8. So, we'll always use this container when it comes to the staging area. Let Azure Data Factory put whatever data it wants in this container when it comes onto the staging area.
9. Now go back to SSMS and delete the data from your table.

```
delete from [logdata_parquet]
```

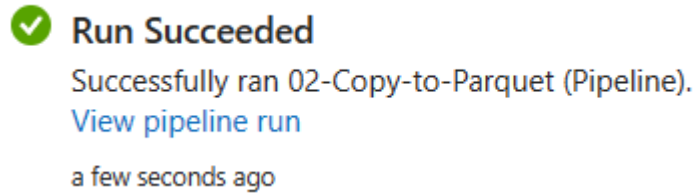
150 %

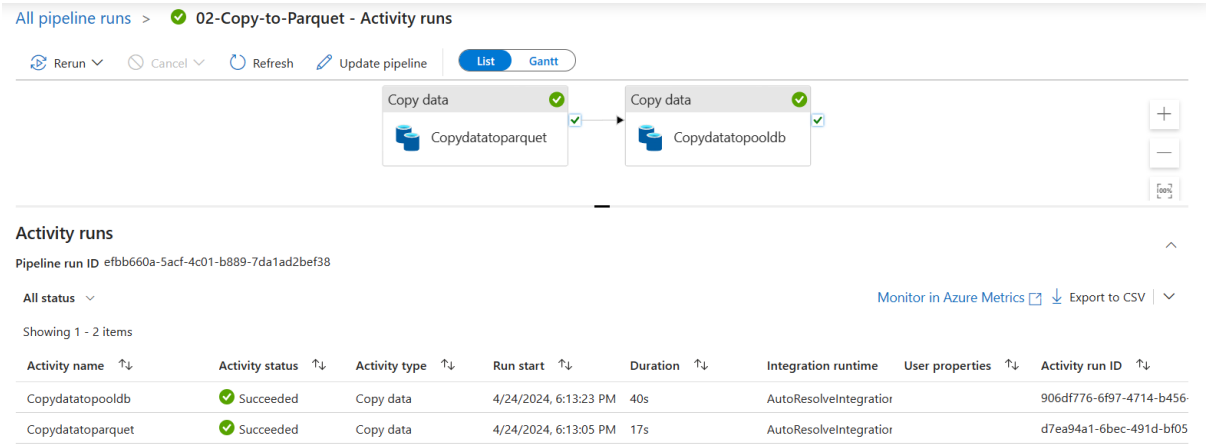
Messages

```
(19229 rows affected)

Completion time: 2024-04-24T18:11:16.8902817+05:30
```

10. After that click on validate all, then click on publish all.
11. Once it is published then you have to trigger the pipeline and wait for some time.
12. When your run is successful you can view your pipeline.





13. Now go back to SSMS and run the Select statement. And you can see the data accordingly.

```
select * from [logdata_parquet]
```

150 %

Results | Messages

	Correlationid	Operationname	Status	Eventcategory	Level	Time	Subscription
1	3cbc1247-a562-4336-885c-e5be7045a10c	Delete Virtual Network	Failed	Administrative	Error	2023-04-17T15:49:49.125Z	6912d7a0-bc28-459a-9407-33bbba641c07
2	87b423cf-9aa3-44b7-8649-05ee4da94c8d	Create Deployment	Succeeded	Administrative	Informational	2023-04-17T15:43:51.854Z	6912d7a0-bc28-459a-9407-33bbba641c07
3	84fb2d66-c063-4573-b973-e3709d781462	'auditIfNotExists' Policy action.	Succeeded	Policy	Informational	2023-04-17T15:32:39.557Z	6912d7a0-bc28-459a-9407-33bbba641c07
4	2127d1c8-ba1f-47df-804a-d57ed00c8c70	'auditIfNotExists' Policy action.	Succeeded	Policy	Informational	2023-04-17T15:28:25.174Z	6912d7a0-bc28-459a-9407-33bbba641c07
5	9054a1fe-efa2-436a-bcce-2b9fca0e6e71	Validate Deployment	Started	Administrative	Informational	2023-04-16T11:37:10.039Z	6912d7a0-bc28-459a-9407-33bbba641c07
6	a265ddd3-d275-46ec-9bc1-f2633cf37b90	'audit' Policy action.	Succeeded	Policy	Warning	2023-04-16T06:18:39.005Z	6912d7a0-bc28-459a-9407-33bbba641c07
7	9803cd2e-60a0-4aa6-a22f-951f69daf260	Create or Update Disk	Succeeded	Administrative	Informational	2023-04-17T11:28:26.364Z	6912d7a0-bc28-459a-9407-33bbba641c07
8	4edf0175-66f4-42d5-8494-4fcba324c2bf	'auditIfNotExists' Policy action.	Succeeded	Policy	Informational	2023-04-17T12:55:48.601Z	6912d7a0-bc28-459a-9407-33bbba641c07
9	891eacee-4531-463c-8fc3-9f41969cad6a	Delete Network Security Group	Started	Administrative	Informational	2023-04-17T11:54:54.846Z	6912d7a0-bc28-459a-9407-33bbba641c07
10	4c79dfd7-8b07-433d-b059-cc0e8e817f3d	'audit' Policy action.	Succeeded	Policy	Warning	2023-04-17T11:51:24.286Z	6912d7a0-bc28-459a-9407-33bbba641c07
11	85a0e0aa-bd69-4d07-8cfe-ede9e0779f81	Create or Update Virtual Machine	Succeeded	Administrative	Informational	2023-04-17T11:30:48.486Z	6912d7a0-bc28-459a-9407-33bbba641c07
12	9803cd2e-60a0-4aa6-a22f-951f69daf260	'auditIfNotExists' Policy action.	Succeeded	Policy	Informational	2023-04-17T11:28:28.277Z	6912d7a0-bc28-459a-9407-33bbba641c07
13	9803cd2e-60a0-4aa6-a22f-951f69daf260	Create or Update Virtual Machine Extension	Succeeded	Administrative	Informational	2023-04-17T11:21:32.642Z	6912d7a0-bc28-459a-9407-33bbba641c07
14	e0197b66-3b79-4eb0-b27e-8f65e97de59f	'audit' Policy action.	Succeeded	Policy	Warning	2023-04-14T10:57:22.582Z	6912d7a0-bc28-459a-9407-33bbba641c07
15	52083e58-f82e-42e4-aa1f-210e8da6b8f7	Create or Update Disk	Accepted	Administrative	Informational	2023-04-14T10:37:46.561Z	6912d7a0-bc28-459a-9407-33bbba641c07
16	52083e58-f82e-42e4-aa1f-210e8da6b8f7	'auditIfNotExists' Policy action.	Succeeded	Policy	Informational	2023-04-14T10:47:50.411Z	6912d7a0-bc28-459a-9407-33bbba641c07
17	52083e58-f82e-42e4-aa1f-210e8da6b8f7	'auditIfNotExists' Policy action.	Started	Policy	Informational	2023-04-14T10:37:30.923Z	6912d7a0-bc28-459a-9407-33bbba641c07
18	8b0be2de-b751-4e3a-a5df-c1d2c0073679	Delete Network Interface	Succeeded	Administrative	Informational	2023-04-17T06:07:20.226Z	6912d7a0-bc28-459a-9407-33bbba641c07
19	83a28383-0fe4-4967-ae48-4ac5e4aa7aa4	Delete Network Interface	Failed	Administrative	Error	2023-04-17T06:04:41.423Z	6912d7a0-bc28-459a-9407-33bbba641c07

Query executed successfully. datasynaps

So, a very important note. When you are looking for an effective way of copying data into Azure Synapse, always consider using PolyBase. So, PolyBase can make use of a parallel mechanism, wherein data can be shifted in parallel onto your, or the use of multiple compute nodes if you do have them configured for your dedicated SQL pool. So, it makes it much more efficient when it comes to the data transfer.