



## Adding an Extra Column

In this lab, we're modifying an existing pipeline in Azure Data Factory to include an additional column with a static value during data transfer. The end goal is to enrich the transferred data by appending a new column, facilitating better analysis and tracking of the data's origin.

We start by navigating to the pipeline configuration, where we add the extra column with a predefined value, such as the file path. Next, we update the database schema in SQL Server Management Studio to accommodate the new column. Back in Azure Data Factory, we adjust the sink settings to reflect the schema changes and update the mapping accordingly.

After validating and publishing the pipeline changes, we execute the pipeline to transfer the data with the added column. Finally, we verify the success of the pipeline run and check the transferred data in SQL Server Management Studio to ensure the new column's presence and accuracy.

Overall, the aim is to enhance data processing and analysis capabilities by incorporating additional metadata into the transferred data.



### To begin with the Lab:

1. Now in this lab, we will go through how you can kind of add a static value for an additional column when it comes onto your pipeline.
2. Now we will move to our existing pipeline in which we were copying and converting data to parquet and then transferring it into our dedicated SQL Pool.
3. Here in your existing pipeline if you select your pipeline where we were copying the data onto our dedicated pool. In the source if you scroll down to the bottom. Then here you can see an option to add additional columns.

The screenshot shows the Azure Data Factory interface for a pipeline named "02-Copy-to-Parquet". The "Source" tab is active. The source dataset is "Copydatatoparquet" and the sink dataset is "Copydatatopooldb". The "Additional columns" section is highlighted with a red border, containing a "New" button.

- Now if you click on new then you will see that in the value you have the file path. Surely, we can add different values but for this lab, we will be specifying only the file path. Then in the same write the file path as written below.

Additional columns		
	<a href="#">New</a>	
<input type="checkbox"/>	Name	
<input type="checkbox"/>	FilePath	Value
<a href="#">Add dynamic content [Alt+Shift+D]</a>		

- Now in SSMS we will drop our parquet table and recreate it with a new additional column that is the file path. Just copy the commands and paste them, then just execute them.

```

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,
    [FilePath] [varchar](400) NULL
)

```

150 % ▶

Messages

Commands completed successfully.

Completion time: 2024-04-24T17:20:58.0944731+05:30

6. As we are making a change now come back to the Data Factory wizard.
7. Go to sink and click on open.

General   Source   **Sink**   Mapping   Settings   User properties

---

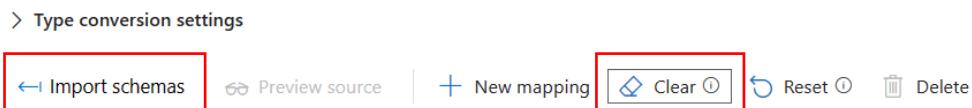
**Sink dataset \***

**Copy method**  Copy command  PolyBase  Bulk insert  Upsert

8. Here you have to go to the schema and click on import schema. Once you have done it then you will be able to see the new column for the file path.

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

9. Now go back onto your pipeline and open mapping. First you have to click on clear then click on import schemas.



Add dynamic content [Alt+Shift+D]

10. And there you will also see your new column.

General	Source	Sink	Mapping	Settings	User properties
<input type="checkbox"/> Source	Type		Destination	Type	
<input type="checkbox"/> Correlationid	abc UTF8		→ Correlationid	abc varchar	<a href="#">+</a> <a href="#">Delete</a>
<input type="checkbox"/> Operationname	abc UTF8		→ Operationname	abc varchar	<a href="#">+</a> <a href="#">Delete</a>
<input type="checkbox"/> Status	abc UTF8		→ Status	abc varchar	<a href="#">+</a> <a href="#">Delete</a>
<input type="checkbox"/> Eventcategory	abc UTF8		→ Eventcategory	abc varchar	<a href="#">+</a> <a href="#">Delete</a>
<input type="checkbox"/> Level	abc UTF8		→ Level	abc varchar	<a href="#">+</a> <a href="#">Delete</a>
<input type="checkbox"/> Time	abc UTF8		→ Time	abc datetime	<a href="#">+</a> <a href="#">Delete</a>
⚠️ Copying from column Time to column Time may have data truncation.					
<input type="checkbox"/> Subscription	abc UTF8		→ Subscription	abc varchar	<a href="#">+</a> <a href="#">Delete</a>
<input type="checkbox"/> Eventinitiatedby	abc UTF8		→ Eventinitiatedby	abc varchar	<a href="#">+</a> <a href="#">Delete</a>
<input type="checkbox"/> Resourctype	abc UTF8		→ Resourctype	abc varchar	<a href="#">+</a> <a href="#">Delete</a>
<input type="checkbox"/> Resourcegroup	abc UTF8		→ Resourcegroup	abc varchar	<a href="#">+</a> <a href="#">Delete</a>
<input type="checkbox"/> Resource	abc UTF8		→ Resource	abc varchar	<a href="#">+</a> <a href="#">Delete</a>
<input type="checkbox"/> FilePath	abc String	Additional	→ FilePath	abc varchar	<a href="#">+</a> <a href="#">Delete</a>

Add dynamic content [Alt+Shift+D]

11. Then the steps are the same as before.
12. First click on Validate all, then click on publish all. After that you need to trigger the pipeline.
13. Once your pipeline run is successful. Then click on view pipeline.

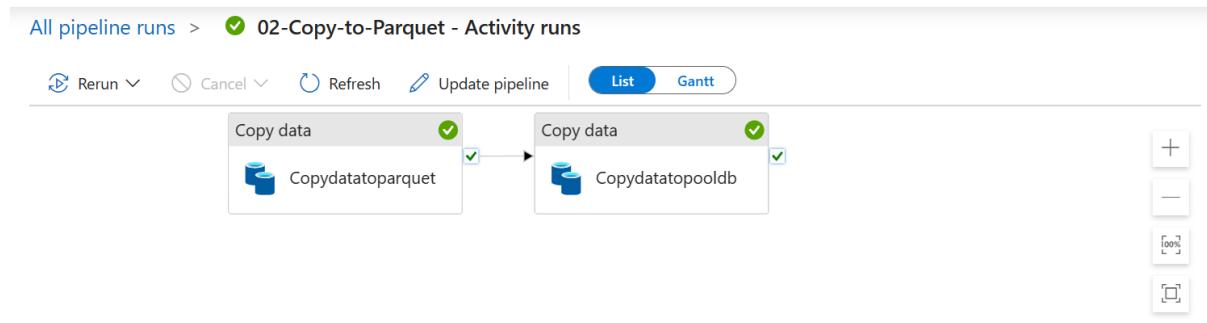
## Run Succeeded

Successfully ran 02-Copy-to-Parquet (Pipeline).

[View pipeline run](#)

a few seconds ago

14. Below you can see your pipeline.



All pipeline runs >  02-Copy-to-Parquet - Activity runs

Rerun Cancel Refresh Update pipeline List Gantt

Activity runs

Pipeline run ID 4521bd1a-119d-4806-806a-29f678f40a8c

All status ▾ Monitor in Azure Metrics Export to CSV ▾

Showing 1 - 2 items

Activity name	Activity status	Activity type	Run start	Duration	Integration runtime	User
Copydatatopooldb	 Succeeded	Copy data	4/24/2024, 5:28:22 PM	21s	AutoResolveIntegration	
Copydatatoparquet	 Succeeded	Copy data	4/24/2024, 5:28:03 PM	18s	AutoResolveIntegration	

15. Now go back to SSMS and run the Select statement to view your data and below you can see your additional column in place.

3