

# ADF Copy Data Activity

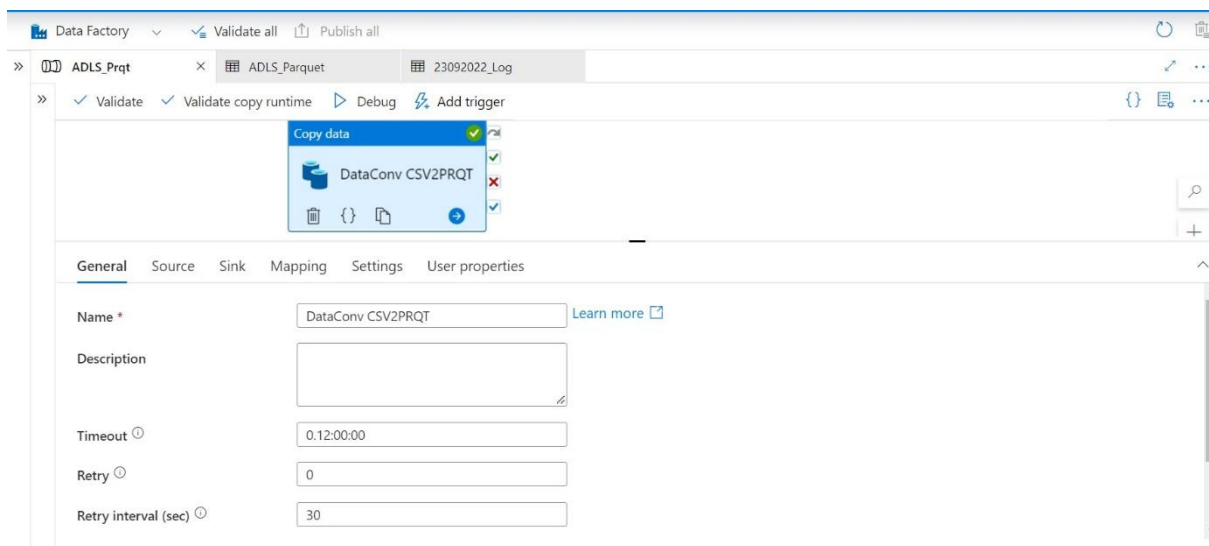
## (On Prem SQL to ADLS .csv to Parquet format)

### Problem Statement:

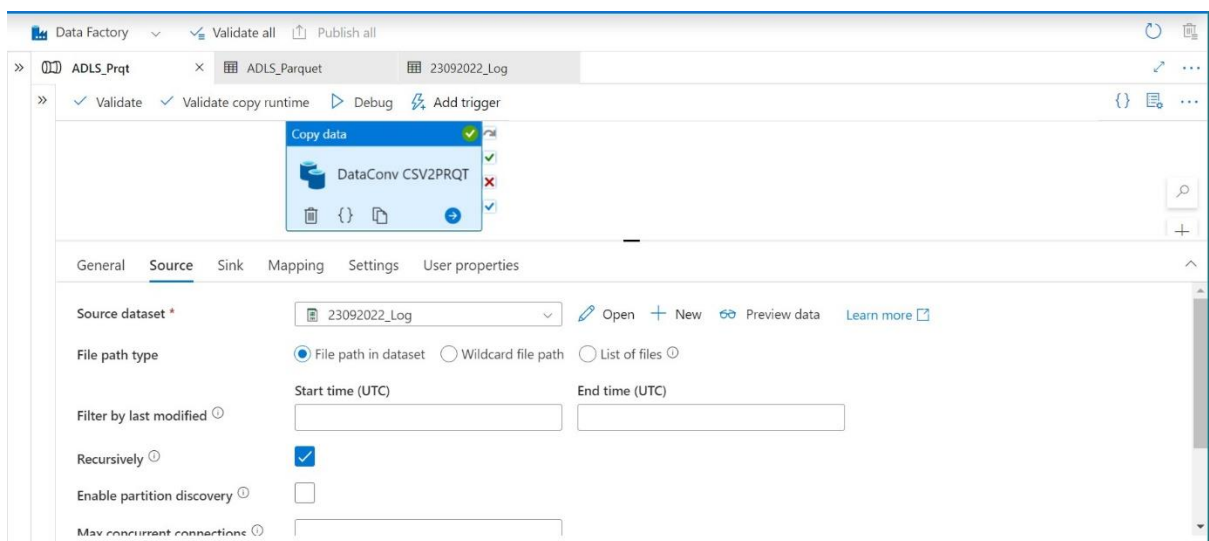
Copy data from MS SQL server data base to Azure Data Lake Storage gen2 and store data in Parquet format at ADLS.

### Solution:

#### ADF Copy Activity Pipeline – General



#### ADF Copy Activity Pipeline – Source



## ADF Copy Activity Pipeline – Preview Source data set



### Preview data

Linked service: ADLS gen2 ADF Scenarios

Object: 15092022\_log\_csv.csv

	EMPLOYEEID	FIRSTNAME	LASTNAME	OPERATTIONNAME	TIMESTAMP
1	1	Rosy	Ava	List Storage Account Keys	26-06-2021 14:24
2	2	Mary	Amelia	Delete SqlPools	25-06-2021 14:22
3	3	Jack	Haland	Pause SQL Analytics pools.	25-06-2021 14:22
4	4	Raees	khan	Pause SQL Analytics pools.	23-06-2021 07:23
5	5	Ruby	Jasline	List Storage Account Keys	20-06-2021 07:23
6	6	Romeo	Lukaku	Pause a Datawarehouse database.	19-06-2021 07:23
7	7	Mathew	Hayden	Create Pipeline Run	17-06-2021 07:23
8	8	Mark	Alexander	Create Pipeline Run	15-06-2021 14:22
9	9	Merlin	Olivia	Create or Update any Pipeline.	13-06-2021 07:23



## ADF Copy Activity Pipeline – Sink

The screenshot shows the 'Sink' tab of an ADF Copy Activity configuration. The 'Sink dataset' is set to 'ADLS\_Parquet'. The 'Copy behavior' is set to 'None'. The 'Max concurrent connections' is set to 1. The 'Block size (MB)' is set to 1. The 'Metadata' is set to '+ New'. The 'Max rows per file' is set to 1. The 'Copy data' button is visible, and the 'DataConv CSV2PRQT' converter is selected. The 'Copy data' button is also visible in the top right corner.

Copy data

DataConv CSV2PRQT

Sink dataset \* ADLS\_Parquet

Copy behavior None

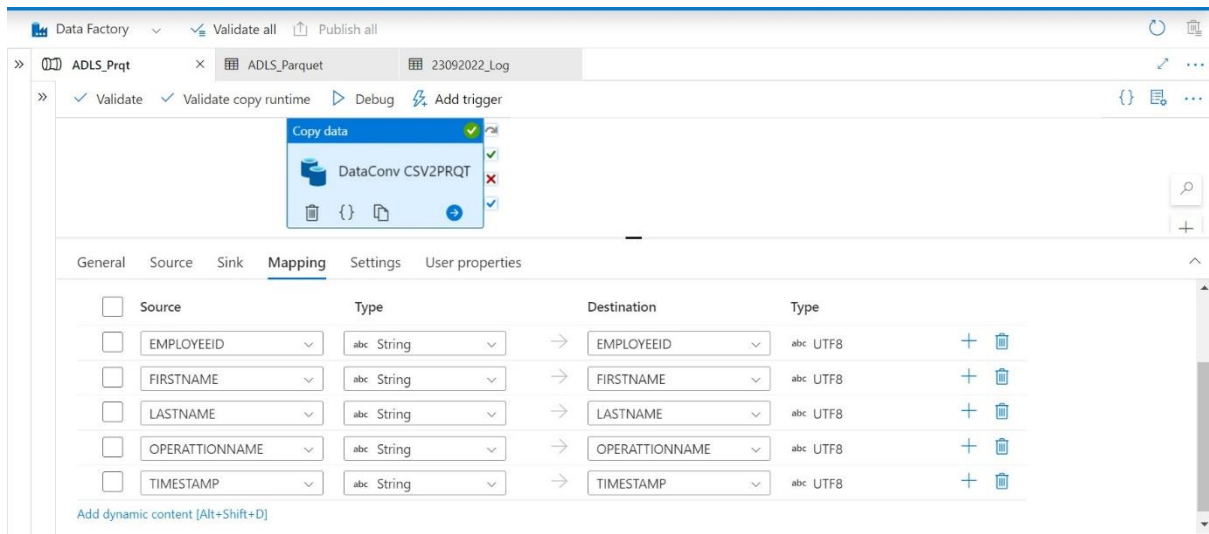
Max concurrent connections 1

Block size (MB) 1

Metadata + New

Max rows per file 1

## ADF Copy Activity Pipeline – Mapping

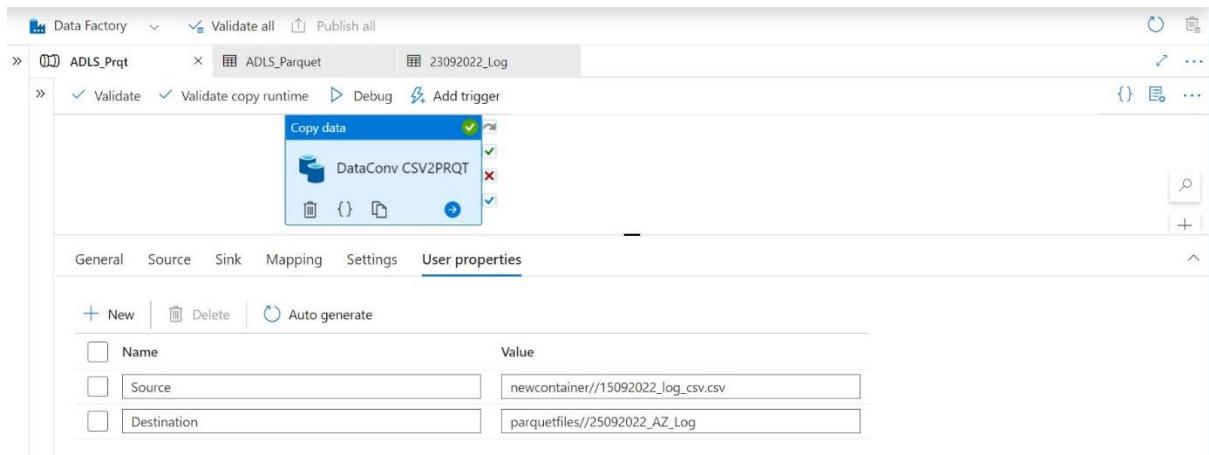


The screenshot shows the 'Mapping' tab of the 'Copy data' activity in the Azure Data Factory portal. The activity is named 'DataConv CSV2PRQT'. The mapping table below shows the source and destination columns and their data types.

Source	Type	Destination	Type
EMPLOYEEID	abc String	EMPLOYEEID	abc UTF8
FIRSTNAME	abc String	FIRSTNAME	abc UTF8
LASTNAME	abc String	LASTNAME	abc UTF8
OPERATIONNAME	abc String	OPERATIONNAME	abc UTF8
TIMESTAMP	abc String	TIMESTAMP	abc UTF8

Buttons for '+', '-', and 'X' are available for each row. A link 'Add dynamic content [Alt+Shift+D]' is at the bottom left.

## ADF Copy Activity Pipeline – User Properties

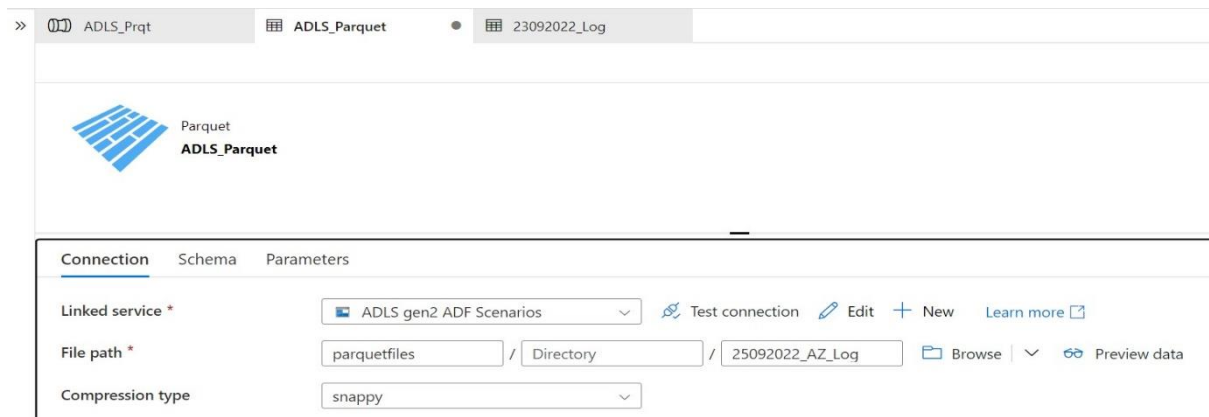


The screenshot shows the 'User properties' tab of the 'Copy data' activity. It displays a table with user-defined properties.

Name	Value
Source	newcontainer//15092022_log_csv.csv
Destination	parquetfiles//25092022_AZ_Log

Buttons for '+ New', 'Delete', and 'Auto generate' are at the top left.

## ADF Copy Activity Pipeline – Sink Data set Properties (used snappy)



The screenshot shows the 'Sink Data set Properties' for the 'ADLS\_Parquet' data set. The 'Connection' tab is selected, showing the following configuration:

- Linked service: ADLS gen2 ADF Scenarios
- File path: parquetfiles / Directory / 25092022\_AZ\_Log
- Compression type: snappy

Buttons for 'Test connection', 'Edit', '+ New', and 'Learn more' are available for the linked service. A 'Browse' button is next to the file path, and a 'Preview data' button is at the bottom right.

Dashboards
Runs
Pipeline runs
Trigger runs
Runtimes & sessions
Integration runtimes
Data flow debug
Notifications
Alerts & metrics

## ADLS\_Prqt

List
Gantt

Rerun
Rerun from activity
Rerun from failed activity
Refresh
Up

+
-
100%

Succeeded
Successfully ran ADLS\_Prqt (Pipeline).  
[View pipeline run](#)

### Activity runs

Pipeline run ID c3388142-088d-45f6-85a0-a1f3181e1999

All status ▾

Showing 1 - 1 items

Activity name ↑↓	Activity type ↑↓	Run start ↑↓	Duration ↑↓	Status ↑↓	Error
DataConv CSV2PRQT	Copy data	Sep 25, 2022, 7:19:48 pm	00:00:10	✓ Succeeded	

Home >

parquetfiles

Container

Search

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Shared access tokens

Access policy

Properties

Metadata

Upload

Change access level

Refresh

Delete

Change tier

Acquire lease

Break lease

View snapshots

Authentication method: Access key (Switch to Azure AD User Account)

Location: parquetfiles

Search blobs by prefix (case-sensitive)

Show deleted blobs

Add filter

Name	Modified	Access tier	Archive status	Blob type
<div>25092022_AZ_Log</div>	9/25/2022, 7:19:56 PM	Hot (Inferred)		Block blob