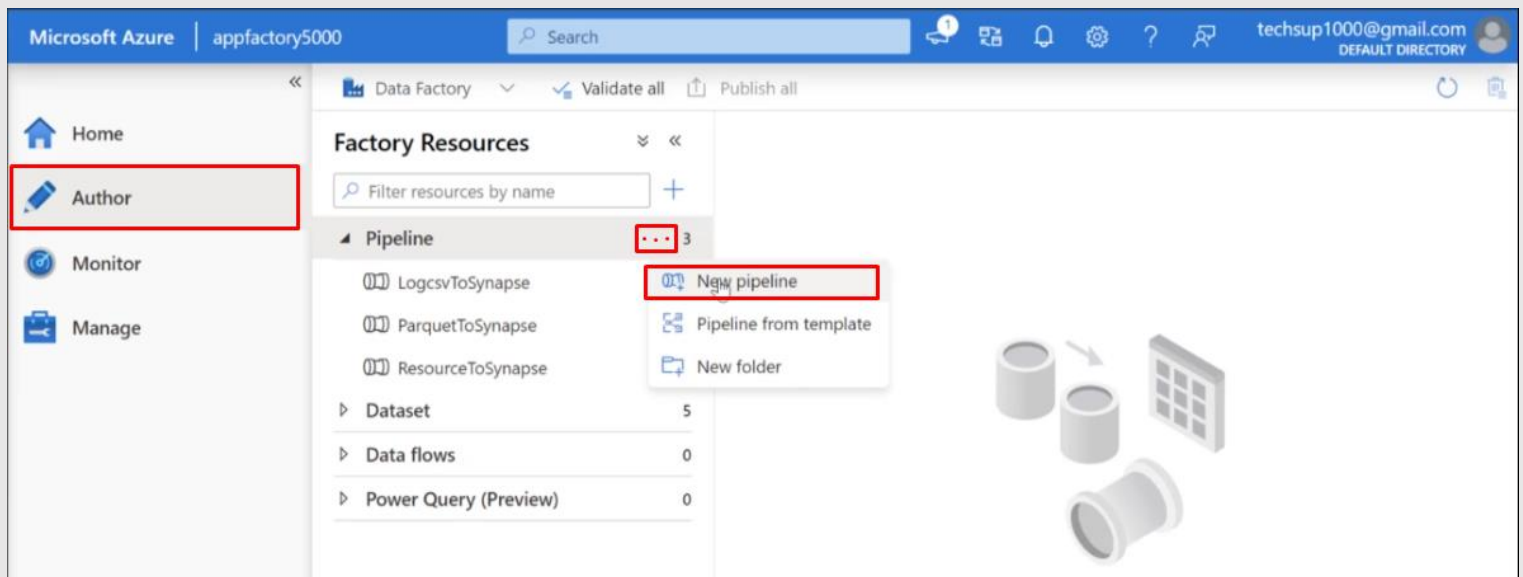
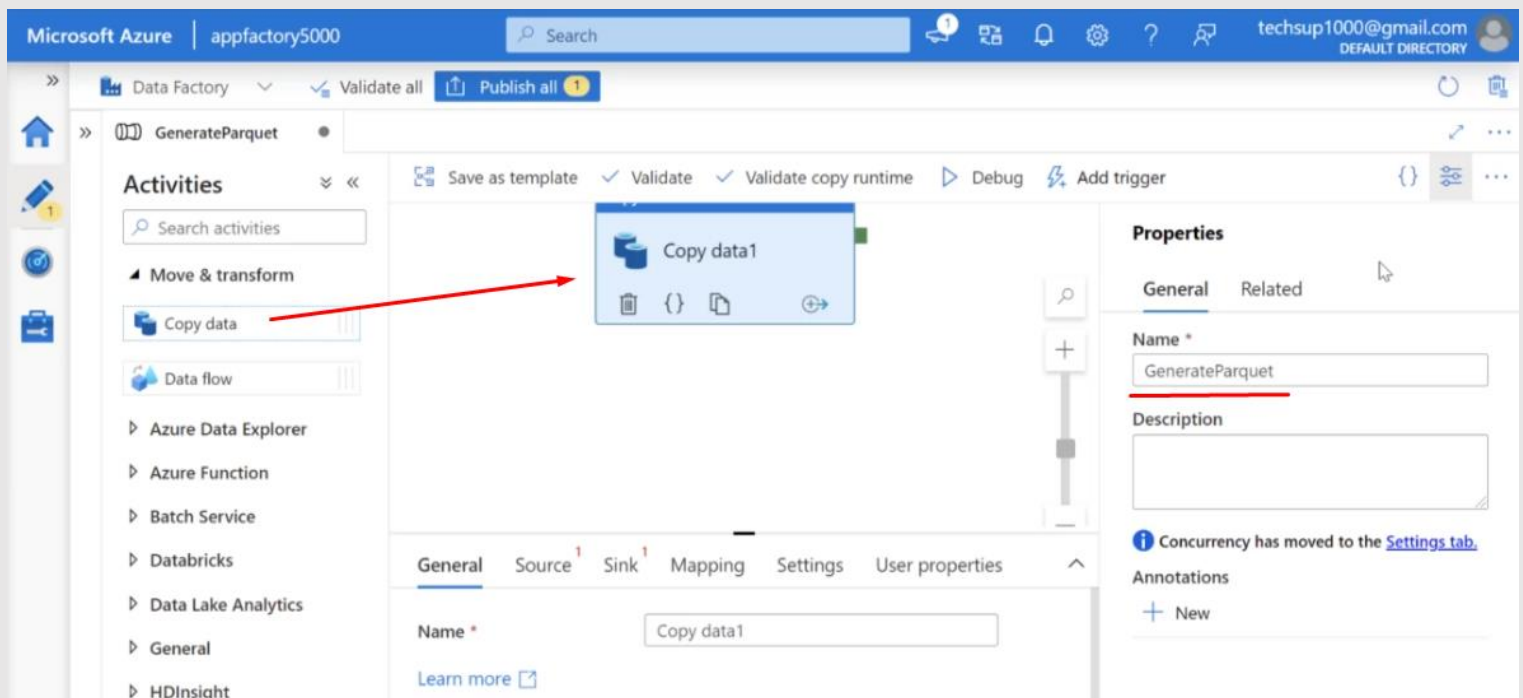
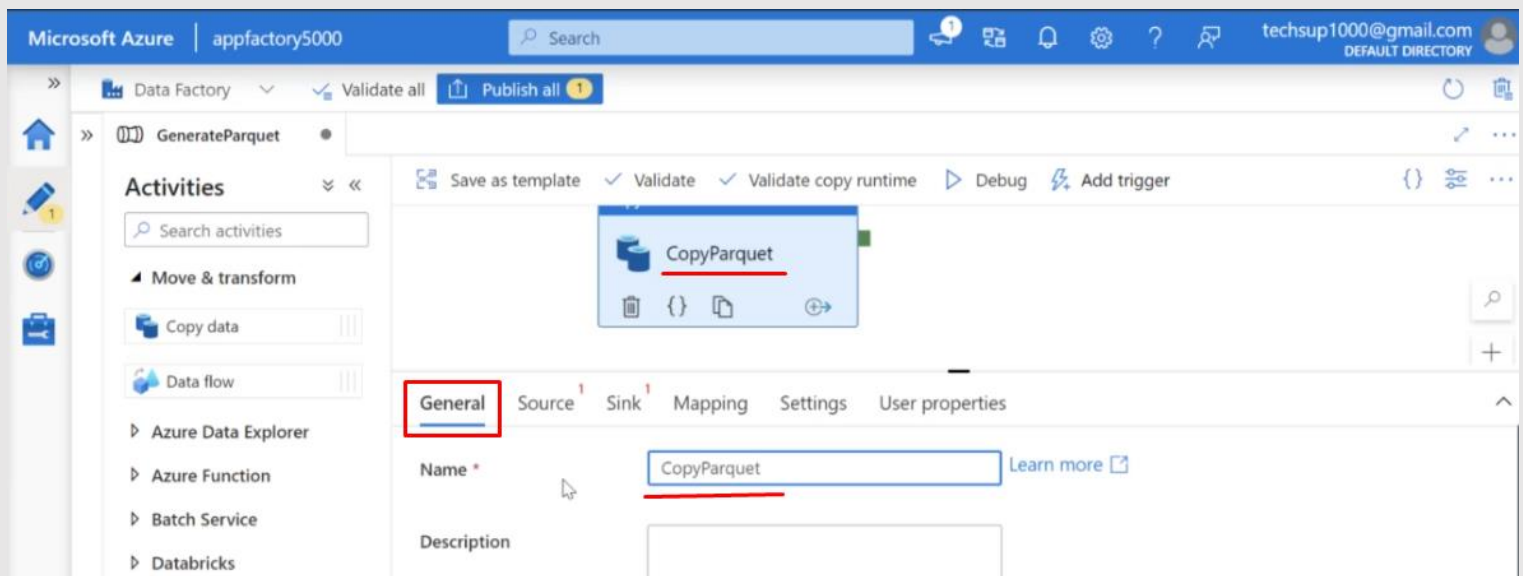


Azure Data Factory Convertir CSV a Parquet

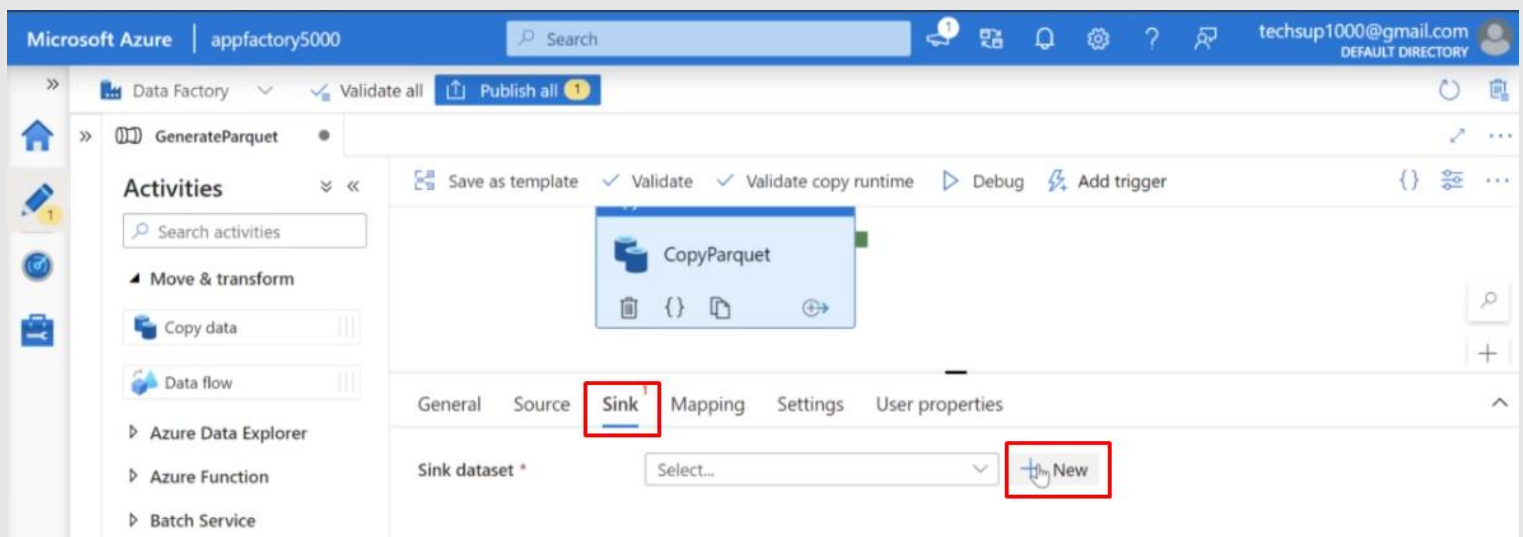
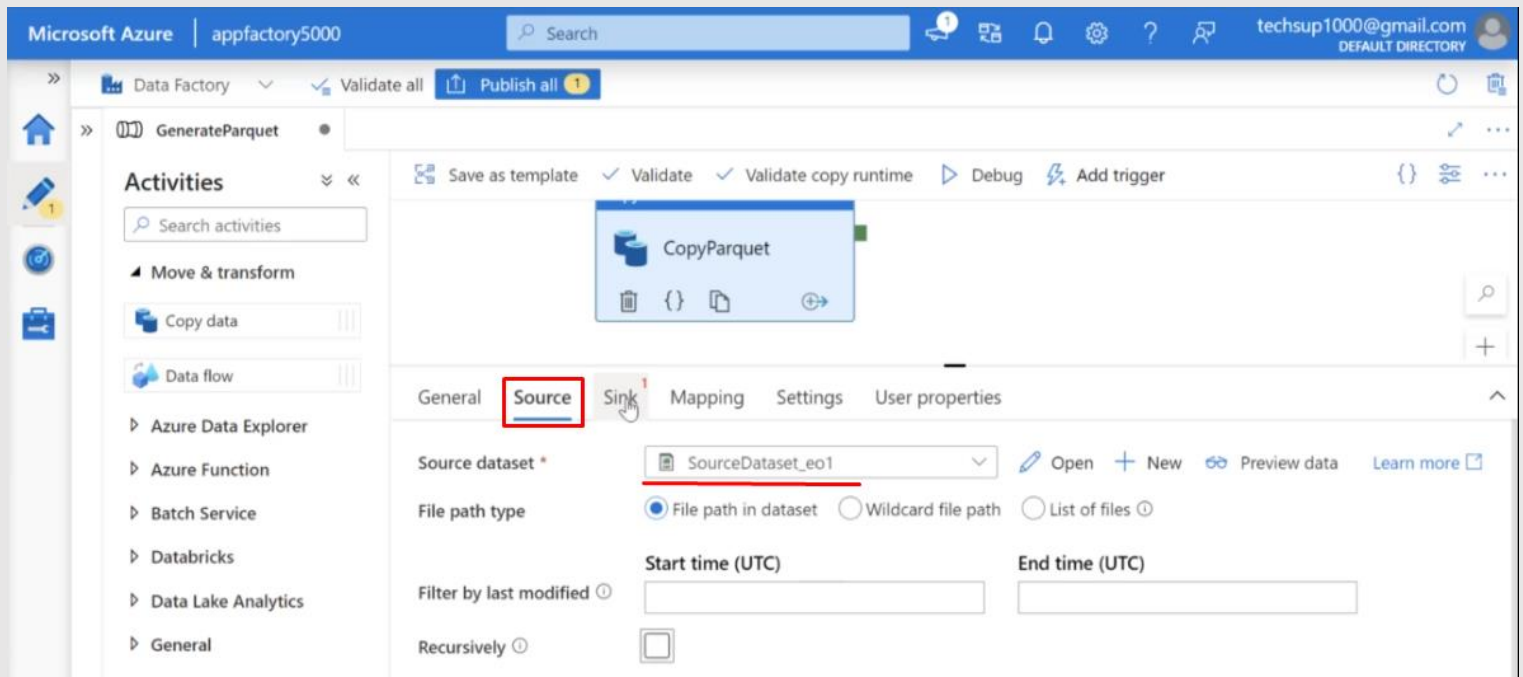


Damos nombre a nuestro Pipeline y movemos la actividad **Copy data** al lienzo





Utilizamos el dataset creado en el punto 1 que apunta al archivo “Log.csv” que se encuentra en el Azure Data Lake Storage /data/raw (contenedor/directorio)



Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | GenerateParquet | Validate all | Publish all

Activities

- Move & transform
- Copy data
- Data flow
- Azure Data Explorer
- Azure Function
- Batch Service
- Databricks
- Data Lake Analytics
- General
- HDInsight
- Iteration & conditionals
- Machine Learning
- Power Query

General Source Sink

Sink dataset *

New dataset

In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. [Learn more](#)

Select a data store

Search

All Azure Database File Generic protocol NoSQL Services and apps

Azure Blob Storage	Azure Cosmos DB (MongoDB API)	Azure Cosmos DB (SQL API)
Azure Data Explorer (Kusto)	Azure Data Lake Storage Gen1	Azure Data Lake Storage Gen2

Continue Cancel

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | GenerateParquet | Validate all | Publish all

Activities

- Move & transform
- Copy data
- Data flow
- Azure Data Explorer
- Azure Function
- Batch Service
- Databricks
- Data Lake Analytics
- General
- HDInsight
- Iteration & conditionals
- Machine Learning
- Power Query

General Source Sink

Sink dataset *

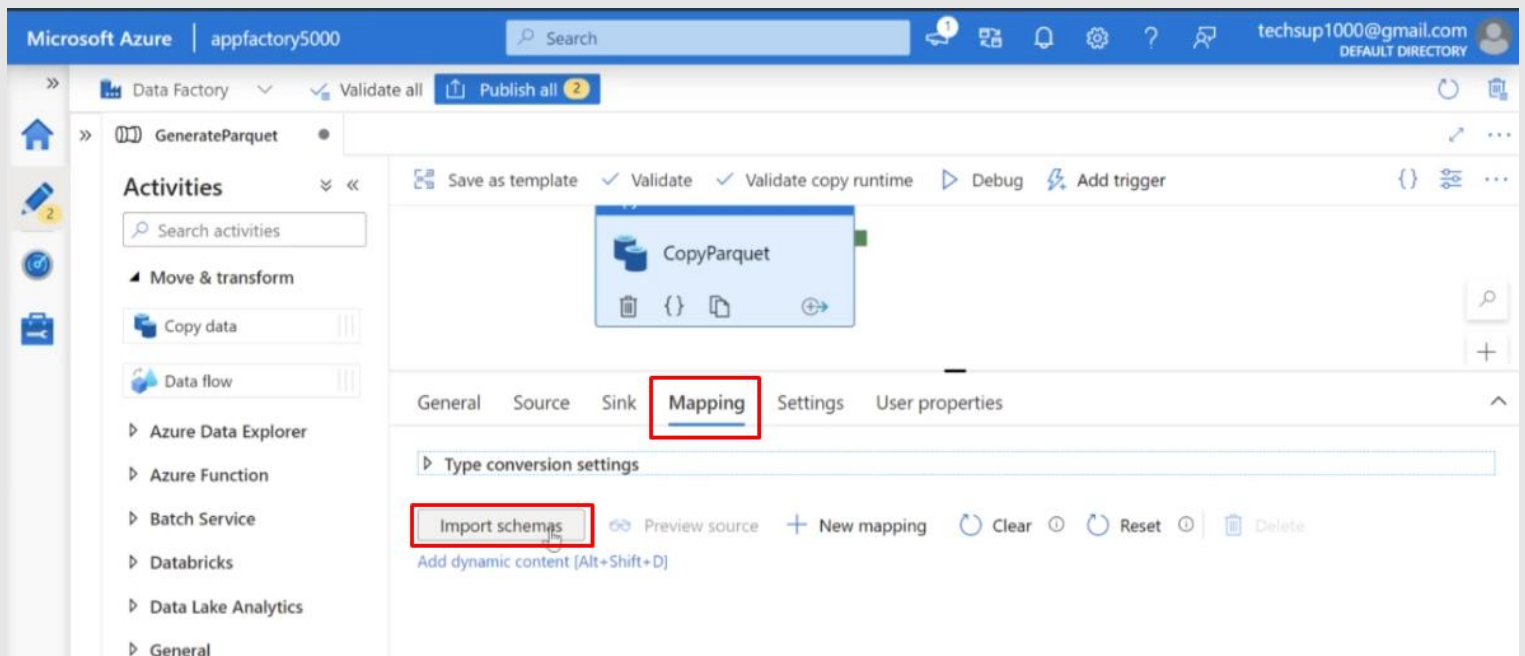
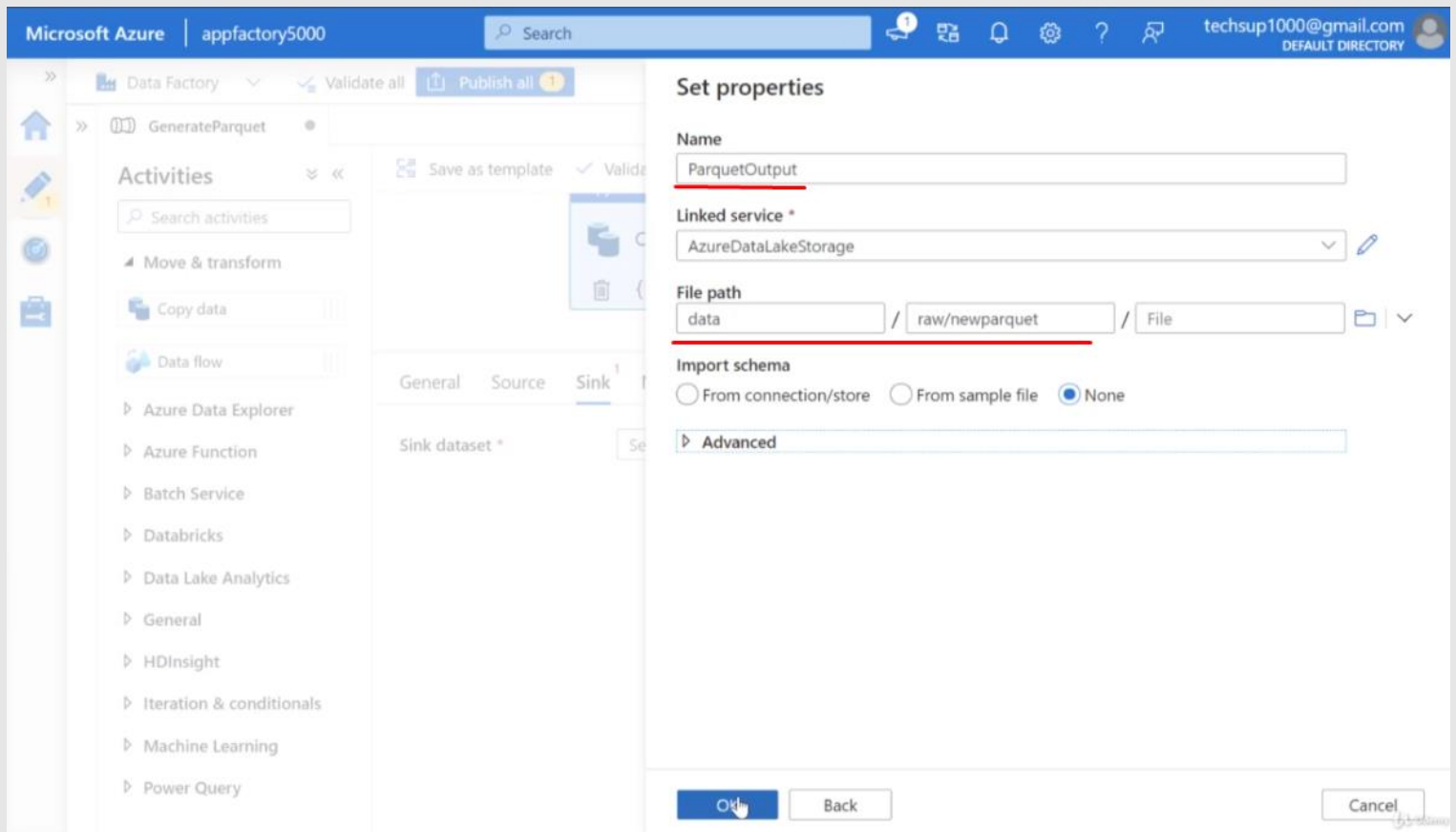
Select format

Choose the format type of your data

Avro	Binary	DelimitedText
JSON	ORC	Parquet

Continue Back Cancel

El destino apuntará al directorio /data/raw/newparquet (contenedor/directorio/subdirectorio)



Debemos indicar manualmente el tipo de datos de la fuente

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all

GenerateParquet

Activities

Search activities

Move & transform

Copy data

Data flow

Azure Data Explorer

Azure Function

Batch Service

Databricks

Data Lake Analytics

General

HDInsight

Iteration & conditionals

Machine Learning

Power Query

CopyParquet

General Source Sink Mapping Settings User properties

Source	Destination	Type
abc String	Id	UTF8
abc String	Correlationid	UTF8
abc String	Operationname	UTF8
abc String	Status	UTF8
abc String	Eventcategory	UTF8
abc String	Level	UTF8
abc String	Time	UTF8

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all

GenerateParquet

Activities

Search activities

Move & transform

Copy data

Data flow

Azure Data Explorer

Azure Function

Batch Service

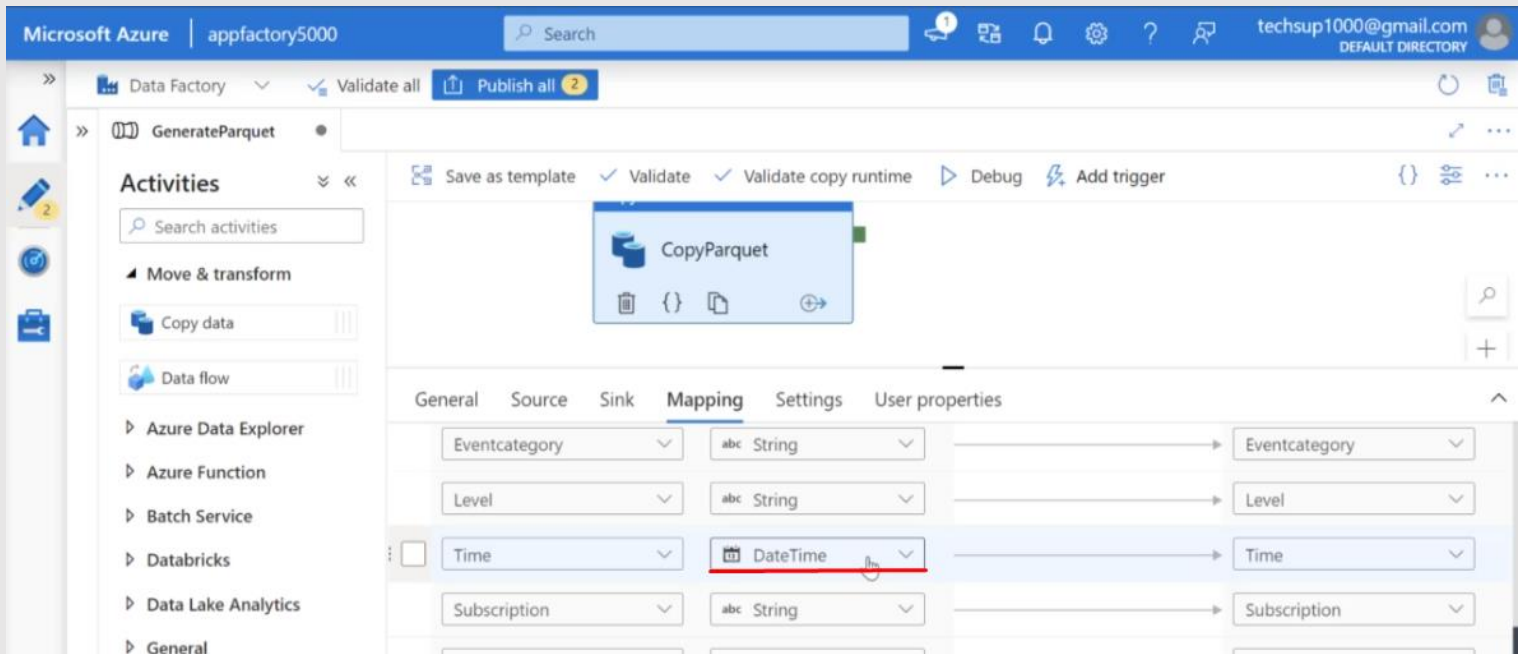
Databricks

Data Lake Analytics

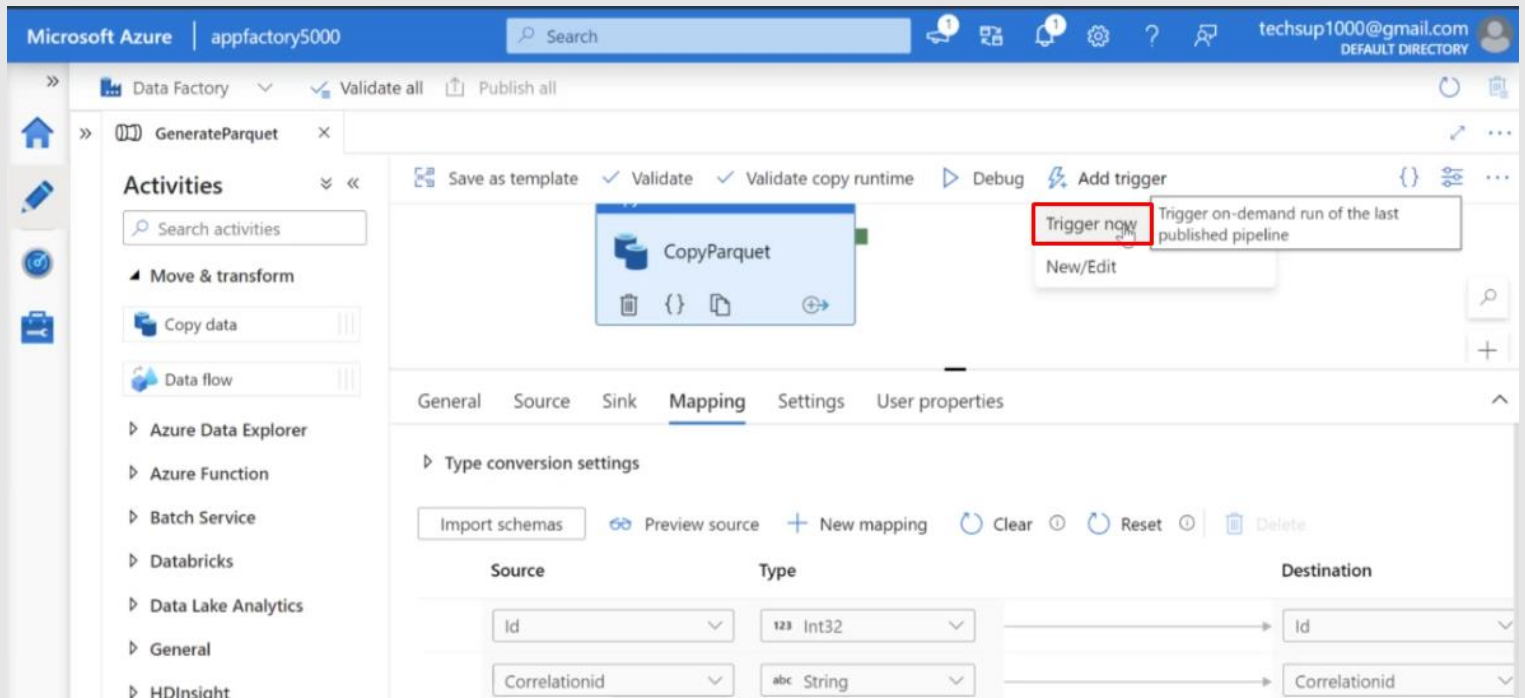
CopyParquet

General Source Sink Mapping Settings User properties

Source	Type	Destination
<input type="checkbox"/> Id	123 Int32	Id
Correlationid	abc String	Correlationid
Operationname	abc String	Operationname



La pestaña **Settings** la dejamos tal cual estaba y luego **Validamos** y **Publicamos**. Por último, ejecutamos.



Microsoft Azure | appfactory5000

techsup1000@gmail.com
DEFAULT DIRECTORY

Pipeline runs

Triggered Debug Rerun Cancel Refresh

Search by run ID or name Abu Dhabi, Muscat (... : Last 24 hours Pipeline name : All

Status : All Runs : Latest runs Add filter Show all

Showing 1 - 4 items

	Duration	Triggered by	Status	Error	Run	Parameters
:54 AM	00:00:09	Manual trigger	✓ Succeeded		Original	
19 AM	00:00:25	Manual trigger	✓ Succeeded		Original	
16 AM	00:00:28	Manual trigger	✓ Succeeded		Original	
:47 PM	00:00:23	Manual trigger	✓ Succeeded		Original	

GenerateParquet

Microsoft Azure Search resources, services, and docs (G+/)

techsup1000@gmail.com
DEFAULT DIRECTORY (TECHSUP1...

Dashboard > Storage accounts > datalake2000 >

data Container

Search (Ctrl+ /)

Upload Add Directory Refresh Rename Delete Change tier

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Shared access signature

Manage ACL

Access policy

Properties

Metadata

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

Location: data / raw / newparquet

Search blobs by prefix (case-sensitive)

Name	Modified	Access tier	Blob type
<input type="checkbox"/> [.]			
<input type="checkbox"/> Log.parquet	7/7/2021, 11:57:52 AM	Hot (Inferred)	Block blob