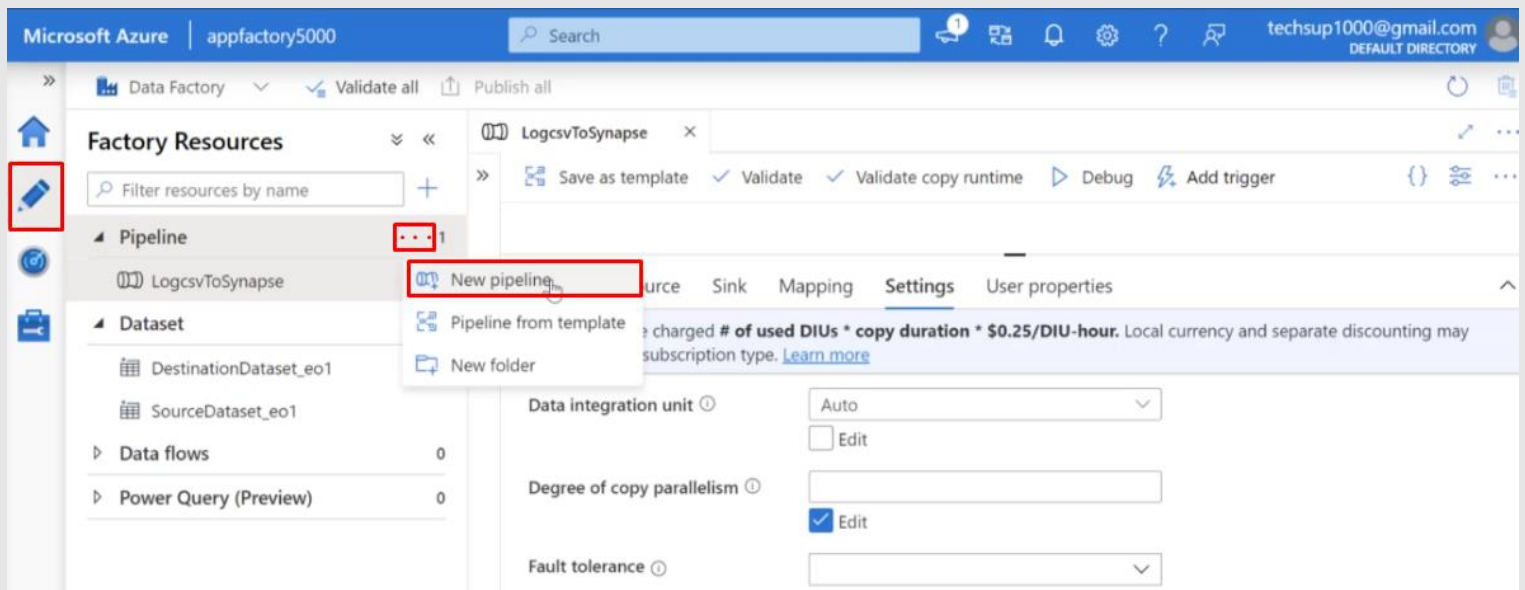
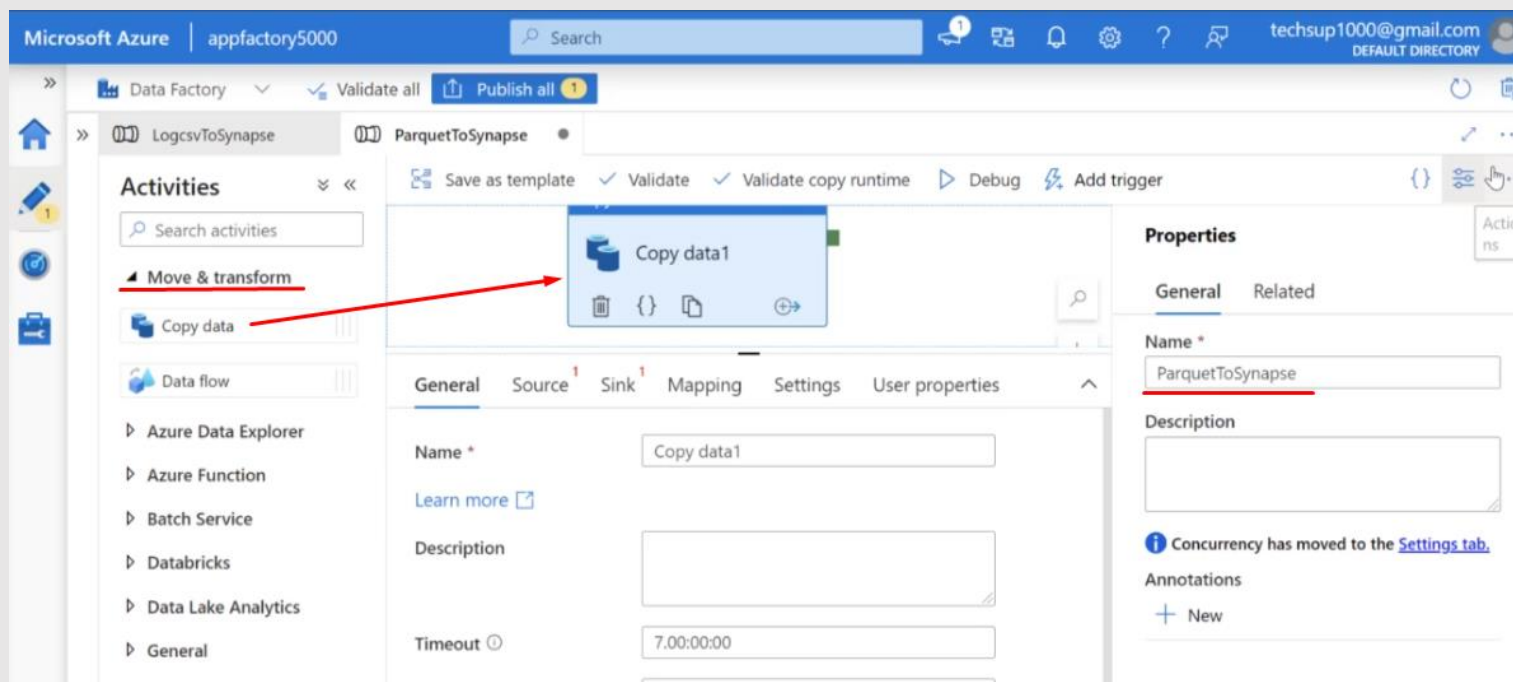


Azure Data Factory Copiar datos desde Azure Data Lake Storage a un Dedicated SQL Pool

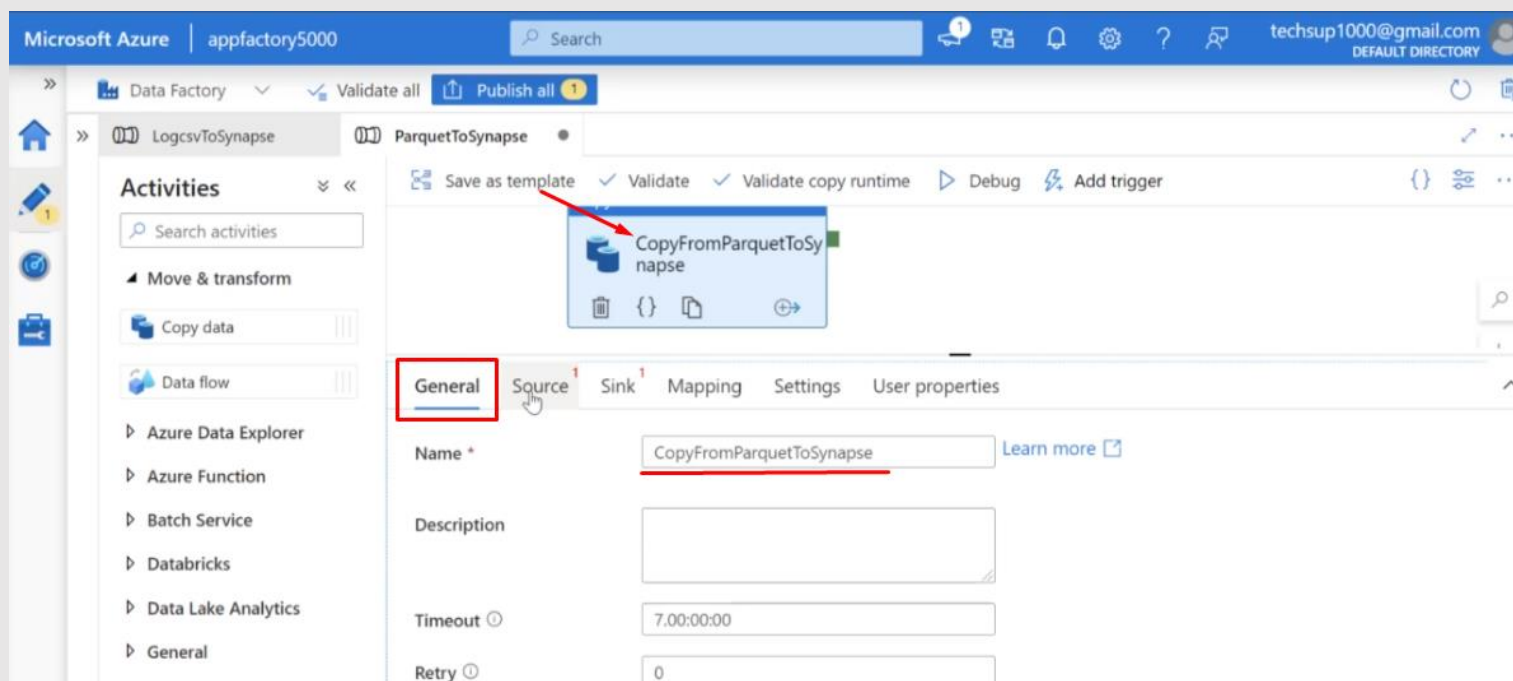
Copiaremos un archivo PARQUET:



Damos un nombre a nuestro Pipeline de copia de datos.
Arrastramos la actividad COPY DATA al lienzo.



Damos un nombre a nuestra actividad



Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory > Validate all > Publish all

LogcsvToSynapse > ParquetToSynapse

Activities

- Move & transform
- Copy data
- Data flow
- Azure Data Explorer
- Azure Function
- Batch Service
- Databricks

CopyFromParquetToSynapse

General Source Sink Mapping Settings User properties

Source dataset * Select... **New**

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory > Validate all > Publish all

LogcsvToSynapse > ParquetToSynapse

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

CopyFromParquetToSynapse

General Source Sink Mapping Settings User properties

Source dataset * Select...







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](#)

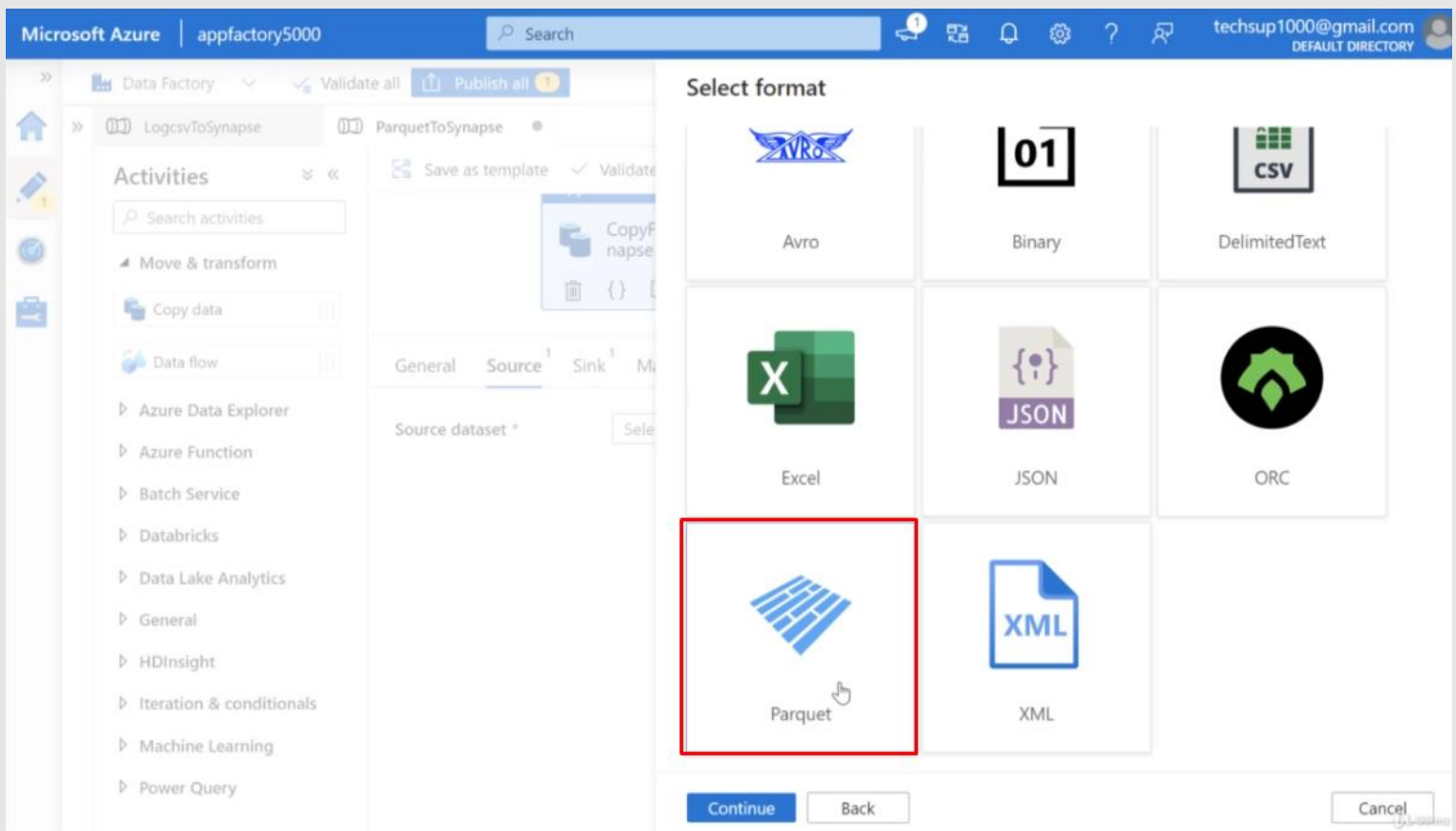
Select a data store

Search

All **Azure** Database File Generic protocol NoSQL Services and apps

| | | |
|--|---|--|
|  Azure Data Explorer (Kusto) |  Azure Data Lake Storage Gen1 |  Azure Data Lake Storage Gen2 |
|  Azure Database for MariaDB |  Azure Database for MySQL |  Azure Database for PostgreSQL |

Continue Cancel

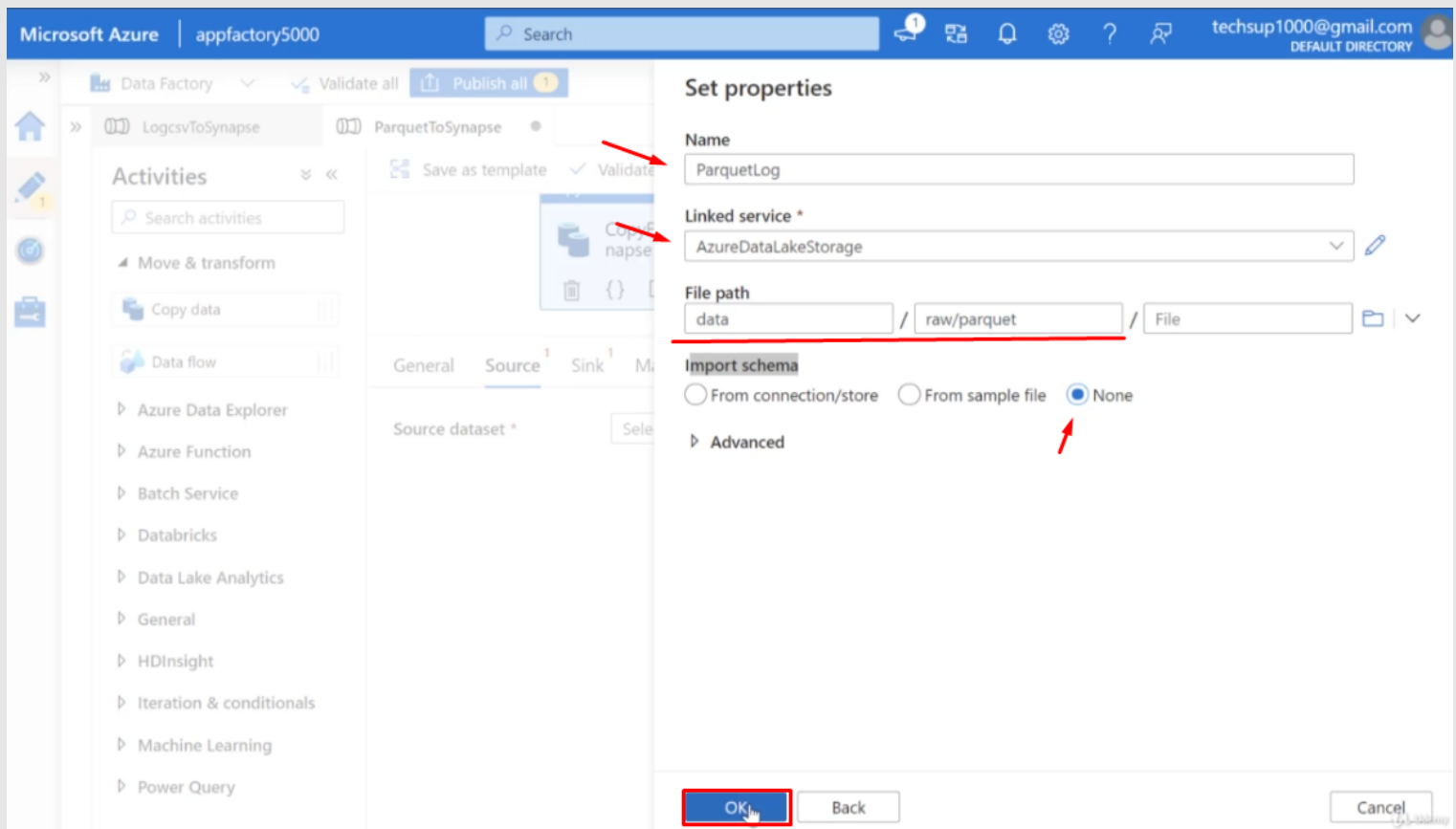


Name: nombre del Source dataset

Linked service: conexión al Azure Data Lake Storage (creado en el punto 1)

Filepath: data --> container ; raw --> directorio ; parquet --> subdirectorio con el archivo parquet

Import schema: None, dado que el archivo parquet trae incorporados el schema y tipos de datos



Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all 2

LogcsvToSynapse | ParquetToSynapse

Activities

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

CopyFromParquetToSynapse

General | **Source** | Sink 1 | Mapping | Settings | User properties

Source dataset * ParquetLog **Open** + New Preview data Learn more

File path type ☐ File path in dataset ☒ Wildcard file path ☐ List of files

Wildcard paths data / Wildcard folder path / *.parquet
Add dynamic content [Alt+Shift+D]

Filter by last modified Start time (UTC) End time (UTC)

Recursively ☒

Enable partition discovery ☐

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all 2

LogcsvToSynapse | ParquetToSynapse | ParquetLog

Parquet
ParquetLog

Connection | **Schema** | Parameters

Import schema Clear

From sample file

From connection/store

Type

Properties

General Related (1)

Name * ParquetLog

Description

Annotations
+ New

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all 2

LogcsvToSynapse | ParquetToSynapse | **ParquetLog**

Parquet
ParquetLog

| Connection | Schema | Parameters |
|------------------|--------|------------|
| Id | | INT32 |
| Correlationid | | UTF8 |
| Operationname | | UTF8 |
| Status | | UTF8 |
| Eventcategory | | UTF8 |
| Level | | UTF8 |
| Time | | INT96 |
| Subscription | | UTF8 |
| Eventinitiatedby | | UTF8 |

Properties

General Related (1)

Name *
ParquetLog

Description

Annotations
+ New

Utilizamos el mismo DATASET DE DESTINO que el punto 1, el cual apunta a la tabla "logdata". Para efectos de este ejercicio, se borraron todos los registros de la tabla.

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all 2

ParquetToSynapse | **ParquetLog**

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

CopyFromParquetToSynapse

General Source **Sink** Mapping Settings User properties

Sink dataset *
DestinationDataset_eo1

Copy method
☒ PolyBase ☐ Copy command ☐ Bulk insert

Allow PolyBase ☒

Reject type
Value

Reject value
0

Use type default ☒

Table option
☒ None ☐ Auto create table

Pre-copy script

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all

ParquetToSynapse | ParquetLog

Activities

- Move & transform
- Copy data
- Data flow
- Azure Data Explorer
- Azure Function
- Batch Service
- Databricks

CopyFromParquetToSynapse

General | Source | Sink | Mapping | Settings | User properties

Import schemas | Preview source | New mapping | Clear | Reset | Delete

Add dynamic content (Alt+Shift+D)

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all

ParquetToSynapse | ParquetLog

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

CopyFromParquetToSynapse

General | Source | Sink | Mapping | Settings | User properties

| Source | Field | Target | Field | Target |
|--------|-------|------------------|-------|----------|
| INT32 | | Id | | int |
| UTF8 | | Correlationid | | varchar |
| UTF8 | | Operationname | | varchar |
| UTF8 | | Status | | varchar |
| UTF8 | | Eventcategory | | varchar |
| UTF8 | | Level | | varchar |
| INT96 | | Time | | datetime |
| UTF8 | | Subscription | | varchar |
| UTF8 | | Eventinitiatedby | | varchar |

Storage path: synapse --> directorio

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all

ParquetToSynapse | ParquetLog

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

CopyFromParquetToSynapse

General Source Sink Mapping **Settings** User properties

Save as template | Validate | Validate copy runtime | Debug | Add trigger

Degree of copy parallelism

Fault tolerance

Enable logging

Enable staging

Staging settings

Staging account linked service: AzureDataLakeStorage

Storage Path: synapse

Enable Compression

Luego tenemos que VALIDAR que todo este encaminado correctamente

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all

ParquetToSynapse | ParquetLog

Activities

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

CopyFromParquetToSynapse

General Source Sink Mapping **Settings** User properties

Save as template | **Validate** | Validate copy runtime | Debug | Add trigger

You will be charged # of used DIUs * copy duration * \$0.25/DIU-hour. Local currency and separate discounting may apply per subscription type. [Learn more](#)

Data integration unit

Degree of copy parallelism

Fault tolerance

Luego lo PUBLICAMOS, que es una suerte de guardar los cambios. Debemos publicar siempre y cuando después de 'validar' no existan errores.

The screenshot shows the Microsoft Azure Data Factory interface for a workspace named 'appfactory5000'. The top navigation bar includes the 'Publish all' button, which is highlighted with a red box. The main workspace area displays a pipeline named 'ParquetToSynapse' with a 'CopyFromParquetToSynapse' activity. The 'Settings' tab is selected, showing a warning about charges based on Data Integration Units (DIUs) and a dropdown menu for the 'Data integration unit' set to 'Auto'.

The screenshot shows the 'Publish all' dialog box in the Microsoft Azure Data Factory interface. The dialog box contains a table of pending changes, with two items listed: 'ParquetToSynapse' (Pipeline) and 'ParquetLog' (Dataset). Both are marked as '(New)' and have no existing versions. The 'Publish' button at the bottom is highlighted with a red box. Red arrows point to the 'CHANGE' column and the 'ParquetLog' dataset entry.

| NAME | CHANGE | EXISTING |
|----------------------------|--------|----------|
| Pipelines | | |
| ParquetToSynapse | (New) | - |
| Datasets | | |
| ParquetLog | (New) | - |

Finalmente

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory > Validate all > Publish all

ParquetToSynapse x ParquetLog

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

CopyFromParquetToSynapse

Save as template

Validate

Validate copy runtime

Debug

Add trigger

Trigger now

Trigger on-demand run of the last published pipeline

New/Edit

General Source Sink Mapping Settings User properties

You will be charged # of used DIUs * copy duration * \$0.25/DIU-hour. Local currency and separate discounting may apply per subscription type. [Learn more](#)

Data integration unit

Auto

Edit

Degree of copy parallelism

Edit

Fault tolerance

Enable logging

Microsoft Azure | appfactory5000

techsup1000@gmail.com
DEFAULT DIRECTORY

Dashboards

Runs

Pipeline runs

Trigger runs

Runtimes & sessions

Integration runtimes

Data flow debug

Notifications

Alerts & metrics

Pipeline runs

Triggered Debug Rerun Cancel Refresh Edit columns List Gantt

Search by run ID or name

Abu Dhabi, Muscat (... : Last 24 hours

Pipeline name : All

Copy filters

Status : All

Runs : Latest runs

Add filter

Showing 1 - 2 items

| Duration | Triggered by | Status | Error | Run | Parameters | Annotations |
|----------|----------------|-----------|-------|----------|------------|-------------|
| 00:00:28 | Manual trigger | Succeeded | | Original | | |
| 00:00:23 | Manual trigger | Succeeded | | Original | | |

Verificamos si los datos fueron cargados correctamente

SQLQuery1.sql - appworkspace9000.sql.azure.synapse.net.newpool (sqladminuser (2408)) - Microsoft SQL Server Management Studio

File Edit View Query Project Tools Window Help

New Query vCREATE VIEW SelectColor AS

newpool Execute

SQLQuery1.sql - ap...qladminuser (2408)) Object Explorer

SELECT * FROM logdata

100 %

Results Messages

| Id | CorrelationId | OperationName | Status | Eventcategory | Level | Time | Subscription | Eventinitiatedby | |
|----|---------------|--------------------------------------|---|---------------|----------------|---------------|-------------------------|--------------------------------------|-------------------------|
| 1 | 11663 | fd3d4e81-4239-443f-ab7a-b71ee3fa26d | Create or Update Public Ip Address | Succeeded | Administrative | Informational | 2021-05-21 12:39:14.507 | 20cbeec9-2d80-4700-b0f6-4fde579a8783 | techsup1000@gmail.com |
| 2 | 11804 | 903ca2f2-ab19-4081-bf12-76f1926325dd | Create or Update Virtual Machine | Succeeded | Administrative | Informational | 2021-05-21 07:48:34.370 | 20cbeec9-2d80-4700-b0f6-4fde579a8783 | techsup1000@gmail.com |
| 3 | 11947 | b7311d0c-bed1-4aed-93b2-e8e60f0c3d86 | 'audit' Policy action. | Succeeded | Policy | Warning | 2021-05-21 07:43:32.363 | 20cbeec9-2d80-4700-b0f6-4fde579a8783 | techsup1000@gmail.com |
| 4 | 12086 | 23531004-2719-4c40-97a5-f25bb1c30162 | Delete Public Ip Address | Succeeded | Administrative | Informational | 2021-05-21 07:35:02.667 | 20cbeec9-2d80-4700-b0f6-4fde579a8783 | techsup1000@gmail.com |
| 5 | 3263 | 33b5fe2-1291-4aef-afd0-52be727bfb59 | Create or Update Virtual Network | Succeeded | Administrative | Informational | 2021-06-05 14:22:53.947 | 20cbeec9-2d80-4700-b0f6-4fde579a8783 | techsup1000@gmail.com |
| 6 | 3317 | c04eodf9-9d7a-42f0-87a0-d6124e3316cd | Get token | Succeeded | Administrative | Informational | 2021-06-05 13:41:56.147 | 20cbeec9-2d80-4700-b0f6-4fde579a8783 | techsup1000@gmail.com |
| 7 | 3456 | b82d44b1-27e0-4727-91ae-c964d214d935 | 'auditIfNotExists' Policy action. | Started | Policy | Informational | 2021-06-05 12:03:31.307 | 20cbeec9-2d80-4700-b0f6-4fde579a8783 | techsup1000@gmail.com |
| 8 | 3600 | 0a661e9e-9a59-430c-b009-8ff1f9fcd66 | List Storage Account Keys | Started | Administrative | Informational | 2021-06-05 11:39:22.647 | 20cbeec9-2d80-4700-b0f6-4fde579a8783 | Microsoft.StorageSync |
| 9 | 14774 | 6c2558cd-fba-4403-bf6c-852ff7b1efe5 | Validate Features | Started | Administrative | Informational | 2021-05-12 15:05:22.597 | 20cbeec9-2d80-4700-b0f6-4fde579a8783 | techsup1000@gmail.com |
| 10 | 14901 | 47a86857-16a6-4845-9739-77acc7ad6698 | List clusterMonitoringUser credential | Started | Administrative | Informational | 2021-05-12 14:44:49.287 | 20cbeec9-2d80-4700-b0f6-4fde579a8783 | techsup1000@gmail.com |
| 11 | 16894 | 0c5c6403-c137-467d-87e4-12d280fd535 | Create or Update Network Security Group | Started | Administrative | Informational | 2021-05-05 18:12:18.930 | 20cbeec9-2d80-4700-b0f6-4fde579a8783 | 8d329043-8cfb-4462-8c20 |
| 12 | 17021 | 87da52b4-9f2d-4ee7-88df-88ba1e52e91c | List clusterAdmin credential | Succeeded | Administrative | Informational | 2021-05-05 17:30:08.603 | 20cbeec9-2d80-4700-b0f6-4fde579a8783 | techsup1000@gmail.com |
| 13 | 17148 | d2bbae44-145a-4152-820b-adaf95e542d7 | Validate Deployment | Succeeded | Administrative | Informational | 2021-05-05 16:28:19.457 | 20cbeec9-2d80-4700-b0f6-4fde579a8783 | AzureContainerService |
| 14 | 24726 | fcd98d67-5dcc-4b56-8d29-56afa2c5080 | Update SQL database | Succeeded | Administrative | Informational | 2021-04-14 14:36:58.490 | 20cbeec9-2d80-4700-b0f6-4fde579a8783 | techsup1000@gmail.com |

Query executed successfully.

appworkspace9000.sql.azure... sqladminuser (2408) newpool 00:00:03 38,057 rows