

# Azure Data Factory Convertir Parquet a JSON

Comenzamos creando un nuevo Pipeline

The screenshot displays the Microsoft Azure Data Factory web interface. The top navigation bar shows the user is logged in as 'techsup1000@gmail.com' and is in the 'appfactory5000' workspace. The main area is divided into three sections: 'Activities', 'Parameters', and 'Properties'.

- Activities:** A list of activities is shown on the left, including 'Move & transform', 'Azure Data Explorer', 'Azure Function', 'Batch Service', 'Databricks', 'Data Lake Analytics', 'General', 'HDInsight', 'Iteration & conditionals', 'Machine Learning', and 'Power Query'. The 'General' category is selected.
- Parameters:** A tab labeled 'Parameters' is active, showing a '+ New' button to create a new parameter.
- Properties:** The 'Properties' tab is active, showing the 'Name' field set to 'ParquetToJSON' and the 'Description' field empty. A note indicates that 'Concurrency has moved to the Settings tab.'

The interface also includes a search bar at the top and a sidebar with navigation icons for home, data factory, and other resources.

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com  
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all 1

SalesFactDataFlow | SalesFactPipeline | Dimensionflow | Dimensionpipeline | ParquetToJSON

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

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

ParquetJSON

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

Name \* ParquetJSON [Learn more](#)

Description

Timeout 7.00:00:00

Retry 0

Retry interval 30

Secure output ☐

Secure input ☐

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com  
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all 1

SalesFactDataFlow | SalesFactPipeline | Dimensionflow | Dimensionpipeline | ParquetToJSON

Activities

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

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

ParquetJSON

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 1

SalesFactDataFlow

SalesFactPipeline

DimensionalFactFlow

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

General

Source 1

Sink 1

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 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

Azure Databricks

Azure HDInsight

Azure IoT Hub

Azure Key Vault

Azure Machine Learning

Azure Monitor

Azure Notification Hubs

Azure OpenAI

Azure Redis

Azure Search

Azure Service Bus

Azure Storage

Azure Stream Analytics

Azure Synapse Analytics

Azure Traffic Manager

Azure Virtual Machine

Azure Virtual Network

Azure Web App

Azure Webhook

Continue

Cancel

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com  
DEFAULT DIRECTORY

» Data Factory

Validate all

Publish all 1

SalesFactDataFlow

SalesFactPipeline

DimensionalFactFlow

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

General

Source 1

Sink 1

Source dataset \*

Select

Select format

Avro

Binary

DelimitedText

Excel

JSON

ORC

Parquet

XML

Continue

Back

Cancel

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com  
DEFAULT DIRECTORY

Data Factory > Validate all > Publish all 1

SalesFactDataFlow > SalesFactPipeline > Dimen

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

General Source Sink

Source dataset \*

### Set properties

Name: ParquetLog

Linked service \*: AzureDataLakeStorage

File path: data / raw/parquet / File

Import schema: ☒ From connection/store ☐ From sample file ☐ None

Advanced

OK Back Cancel

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com  
DEFAULT DIRECTORY

Data Factory > Validate all > Publish all 2

SalesFactDataFlow > SalesFactPipeline > Dimensionflow > Dimensionpipeline > ParquetToJSON

Activities

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

Save as template Validate Validate copy runtime Debug Add trigger

ParquetJSON

General Source Sink Mapping Settings User properties

Source dataset \*: ParquetLog

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

Wildcard paths: data / Wildcard folder path / \*.parquet

Start time (UTC) End time (UTC)

Filter by last modified

Recursively ☒

Enable partition discovery ☐

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com  
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all 2

SalesFactDataFlow | SalesFactPipeline | Dimensionflow | Dimensionpipeline | ParquetToJSON

Activities

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

ParquetJSON

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

Sink dataset \* Select... **+ New**

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com  
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all 2

SalesFactDataFlow | SalesFactPipeline | Dimensionflow | Dimensionpipeline | ParquetToJSON

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

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

Sink 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 Blob Storage	Azure Cosmos DB (MongoDB API)	Azure Cosmos DB (SQL API)
Azure Data Explorer (Kusto)	Azure Data Lake Storage Gen1	<b>Azure Data Lake Storage Gen2</b>

Continue Cancel

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com  
DEFAULT DIRECTORY

» Data Factory » Validate all Publish all 2

SalesFactDataFlow SalesFactPipeline Dimen

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

Save as template Validate

General Source Sink

Sink dataset \*

Select format

Choose the format type of your data

Avro Binary DelimitedText

JSON ORC Parquet

Continue Back Cancel

El nombre del archivo en la ruta lo escribimos nosotros.

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com  
DEFAULT DIRECTORY

» Data Factory » Validate all Publish all 2

SalesFactDataFlow SalesFactPipeline Dimen

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

Save as template Validate

General Source Sink

Sink dataset \*

Set properties

Name

LogJson

Linked service \*

AzureDataLakeStorage

File path

data / raw / log.json

Import schema

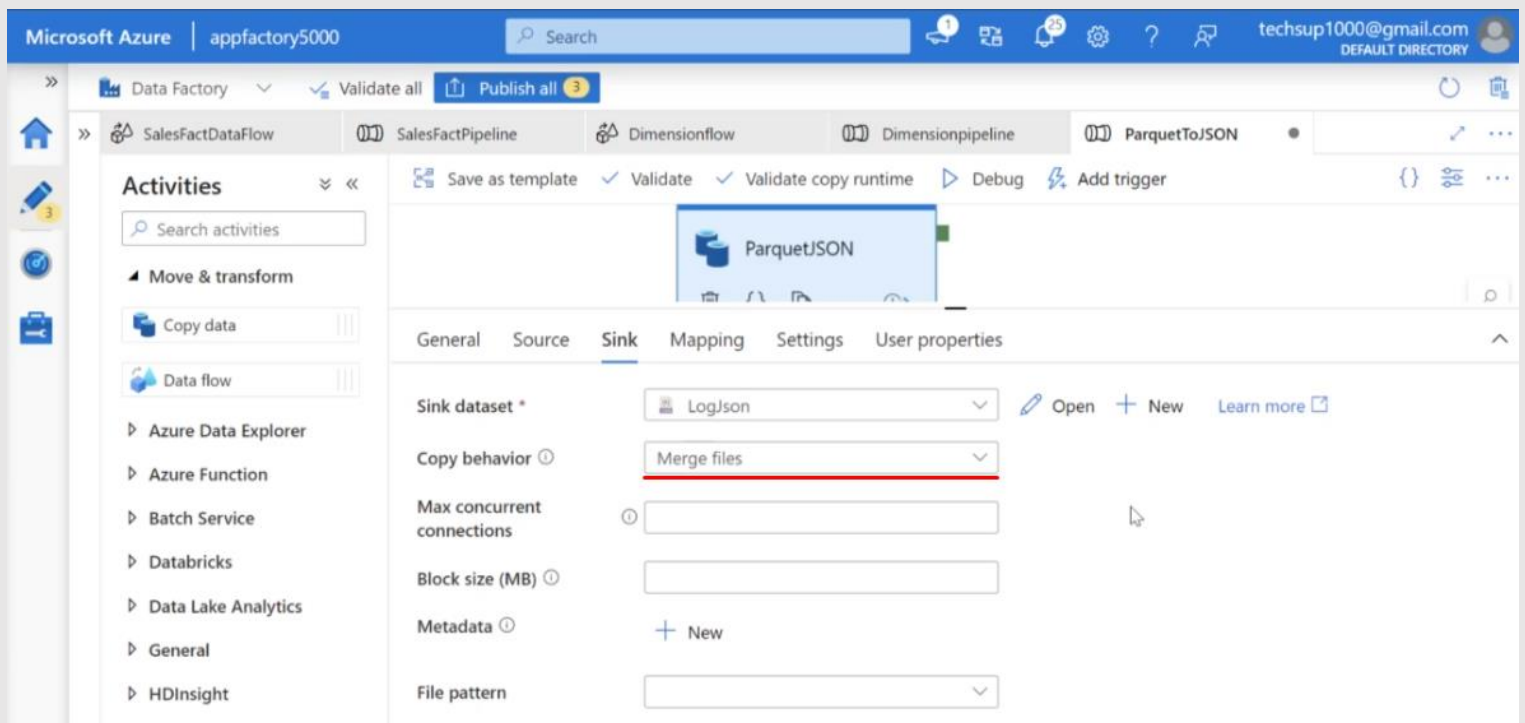
☐ From connection/store ☐ From sample file ☒ None

Advanced

OK Back Cancel

✖ "Schema import failed: ADLS Gen2 operation failed for: Operation returned an invalid status code 'NotFound'. Account: 'datalake2000'. FileSystem: 'data'. Path: 'raw/log.json'. ErrorCode: 'PathNotFound'. Message: 'The specified path does not exist.'. RequestId: '92ee49fe-b01f-0058-0530-73d929000000'. TimeStamp: 'Wed, 07 Jul 2021 13:00:47 GMT'. Operation returned an invalid status code 'NotFound'. Activity ID: 62afdd4a-3e93-4abf-908b-4651f70346bd"





Luego, **Validamos, Publicamos y Ejecutamos**. Ahora podemos verificar si el archivo Json se encuentra en el storage.

