

Azure Data Factory Cargar un archivo JSON a un Dedicated SQL Pool

Vamos a reutilizar el Pipeline creado en el punto 11 y agregaremos una nueva actividad Copy data para mover los datos desde el archivo JSON que recién creamos a una tabla de un Dedicated SQL Pool.

The screenshot displays the Microsoft Azure Data Factory portal for a workspace named 'appfactory5000'. The top navigation bar includes a search bar, notification icons, and the user profile 'techsup1000@gmail.com'. The main interface shows a pipeline named 'SalesFactPipeline' with a 'Copy data' activity selected. A red box highlights the 'Copy data' activity, and a red arrow points to the 'Copy data' button in the 'Activities' pane on the left. The 'Copy data' activity is configured with 'ParquetJSON' as the source and 'ToSynapse' as the sink. The 'General' tab is active, showing the following configuration:

Property	Value
Name *	ToSynapse
Description	
Timeout	7.00:00:00
Retry	0
Retry interval	30
Secure output	<input type="checkbox"/>

El dataset “LogJson” fue creado en el punto 11 y apunta al Azure Data Lake Storage, a la ruta:
data/raw/log.json

data: contenedor ; **raw:** directorio

This screenshot shows the Microsoft Azure Data Factory interface for a pipeline named 'ParquetToJSON'. The 'Sink' tab is selected, displaying the configuration for the destination dataset. The 'Source dataset' is set to 'LogJson'. The 'File path type' is 'File path in dataset'. The 'Start time (UTC)' and 'End time (UTC)' fields are empty. The 'Filter by last modified' checkbox is checked. The 'Recursively' checkbox is checked. The 'Enable partition discovery' checkbox is unchecked. The 'Max concurrent connections' field is empty. The 'Sink' tab is highlighted with a red circle and a red arrow pointing to it.

This screenshot shows the Microsoft Azure Data Factory interface for the same pipeline. The 'Sink' tab is selected, and the 'Sink dataset' field is highlighted with a red box. The 'Sink dataset' field is currently set to 'Select...'. A red box highlights the '+ New' button next to the 'Sink dataset' field.

El dataset de destino apuntara a Azure Synapse Analytics.

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

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

Copy data

ParquetJSON

General Source Sink

Sink dataset *

Set properties

Name: logdatadt

Linked service *: AzureSynapseAnalytics

Table name: dbo.logdata

Import schema: ☒ From connection/store ☐ None

Advanced

OK Back Cancel

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

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

Copy data

ParquetJSON

Copy data

ToSynapse

General Source Sink Mapping Settings User properties

Import schemas

Collection reference

Map complex values to string

Advanced editor

Name	Type	Collection reference	Column name
------	------	----------------------	-------------

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all 2

SalesFactDataFlow | SalesFactPipeline | Dimensionflow | Dimensionpipeline | ParquetToJSON

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

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

Import schemas | + New mapping | Clear | Delete | Advanced editor

Collection reference

Map complex values to string

Name	Type	Collection reference	Column name	Include
Id	any	any	Id	<input checked="" type="checkbox"/>
Correlationid	any	any	Correlationid	<input checked="" type="checkbox"/>
Operationname	any	any	Operationname	<input checked="" type="checkbox"/>
Status	any	any	Status	<input checked="" type="checkbox"/>
Eventcategory	any	any	Eventcategory	<input checked="" type="checkbox"/>
Level	any	any	Level	<input checked="" type="checkbox"/>

Storage path: en el contenedor **synapse**. Esta opción de “enable staging” no es necesaria, podría en algún escenario mejorar tiempos. Esta casilla hace que genere una copia de los datos en el Data lake antes de subir a la base de datos. Se nota su utilidad en copias dónde el origen está en otro servidor. Pero leyendo archivos del propio data lake puede que no o que sea más lento.

Microsoft Azure | appfactory5000

Search

techsup1000@gmail.com
DEFAULT DIRECTORY

Data Factory | Validate all | Publish all 2

SalesFactDataFlow | SalesFactPipeline | Dimensionflow | Dimensionpipeline | ParquetToJSON

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

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

Enable staging ☒

Staging settings

Staging account linked service | AzureDataLakeStorage | Test connection | Edit | New

Storage Path | synapse | Browse

Enable Compression

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

File Edit View Query Project Tools Window Help

newpool

Execute

SQLQuery2.sql - ap...sqladminuser (119))* SQLQuery1.sql - ap...qladminuser (2408))* Object Explorer

-- 38057
DELETE FROM logdata

SELECT * FROM logdata

100 %

Results Messages

	Id	CorrelationId	OperationName	Status	EventCategory	Level	Time	Subscription	EventInitiatedBy	Resource
1	18995	b121f846-596a-49c3-afe0-235c91db93ea	Delete insights component	Succeeded	Administrative	Informational	2021-04-28 12:06:46.157	20c6eec9-2d80-4700-b0f6-4fde579a8783	techsup1000@gmail.com	Microsoft
2	19137	b6ab09c3-c95d-4407-9591-44c0d83b470f	List Storage Account Keys	Succeeded	Administrative	Informational	2021-04-28 06:03:00.210	20c6eec9-2d80-4700-b0f6-4fde579a8783	techsup1000@gmail.com	Microsoft
3	19279	ec1b4257-2625-421f-a2af-4906a198e0ff	Delete Network Security Group	Failed	Administrative	Error	2021-04-27 19:11:36.307	20c6eec9-2d80-4700-b0f6-4fde579a8783	techsup1000@gmail.com	Microsoft
4	19424	bd2e3c0b-7792-4f92-8603-4210484e54c0	'auditIfNotExists' Policy action.	Succeeded	Policy	Informational	2021-04-27 19:13:50.593	20c6eec9-2d80-4700-b0f6-4fde579a8783	techsup1000@gmail.com	Microsoft
5	2658	c13bf78-a101-48a3-b108-5574d0fa616	Delete Public Ip Address	Failed	Administrative	Error	2021-06-05 17:52:13.167	20c6eec9-2d80-4700-b0f6-4fde579a8783	techsup1000@gmail.com	Microsoft
6	20220	b66d3b74-5b6a-476d-a8a3-7a7c1314b224	Delete Storage Account	Succeeded	Administrative	Informational	2021-04-26 19:35:07.153	20c6eec9-2d80-4700-b0f6-4fde579a8783	techsup1000@gmail.com	Microsoft
7	20354	173a4b8b-0a51-4503-a597-45ff5bb3006	'auditIfNotExists' Policy action.	Started	Policy	Informational	2021-04-26 15:55:46.480	20c6eec9-2d80-4700-b0f6-4fde579a8783	techsup1000@gmail.com	Microsoft
8	5484	ba18d1a2-67e6-41b4-94a2-a396c9a2f09a	Create or Update Public Ip Address	Started	Administrative	Informational	2021-06-04 08:14:04.823	20c6eec9-2d80-4700-b0f6-4fde579a8783	techsup1000@gmail.com	Microsoft
9	5625	efb45c0d-e8f6-4943-9b3c-7cf1aa0f0069	Delete SQL database	Succeeded	Administrative	Informational	2021-06-03 18:48:37.140	20c6eec9-2d80-4700-b0f6-4fde579a8783	techsup1000@gmail.com	Microsoft
10	21528	6f47df7f-0127-4945-aeb4-9c665453bf4d	Write EventSubscription	Accepted	Administrative	Informational	2021-04-24 06:34:31.420	20c6eec9-2d80-4700-b0f6-4fde579a8783	techsup1000@gmail.com	Microsoft
11	18529	68cc49eb-6066-4a1d-834d-a5b6ffee1b0e	write	Succeeded	Administrative	Informational	2021-05-01 06:37:42.920	20c6eec9-2d80-4700-b0f6-4fde579a8783	techsup1000@gmail.com	Microsoft
12	18674	e432bd85-ac6a-434f-bbbf-e40357d3a793	'auditIfNotExists' Policy action.	Succeeded	Policy	Informational	2021-05-01 06:24:03.887	20c6eec9-2d80-4700-b0f6-4fde579a8783	techsup1000@gmail.com	Microsoft
13	18819	a321a3b8-000b-4ad0-b952-c7b5658d424c	Create Deployment	Succeeded	Administrative	Informational	2021-04-30 15:06:38.683	20c6eec9-2d80-4700-b0f6-4fde579a8783	techsup1000@gmail.com	Microsoft
14	18960	7dd574b3-e213-40f8-8e62-e7195f15ead6	Create or Update Network Watcher	Succeeded	Administrative	Informational	2021-04-30 14:51:14.327	20c6eec9-2d80-4700-b0f6-4fde579a8783	Azure Traffic Manager and DNS	Microsoft

Query executed successfully. appworkspace9000.sql.azuresynapse.net sqladminuser (2408) newpool 00:00:04 38,057 rows