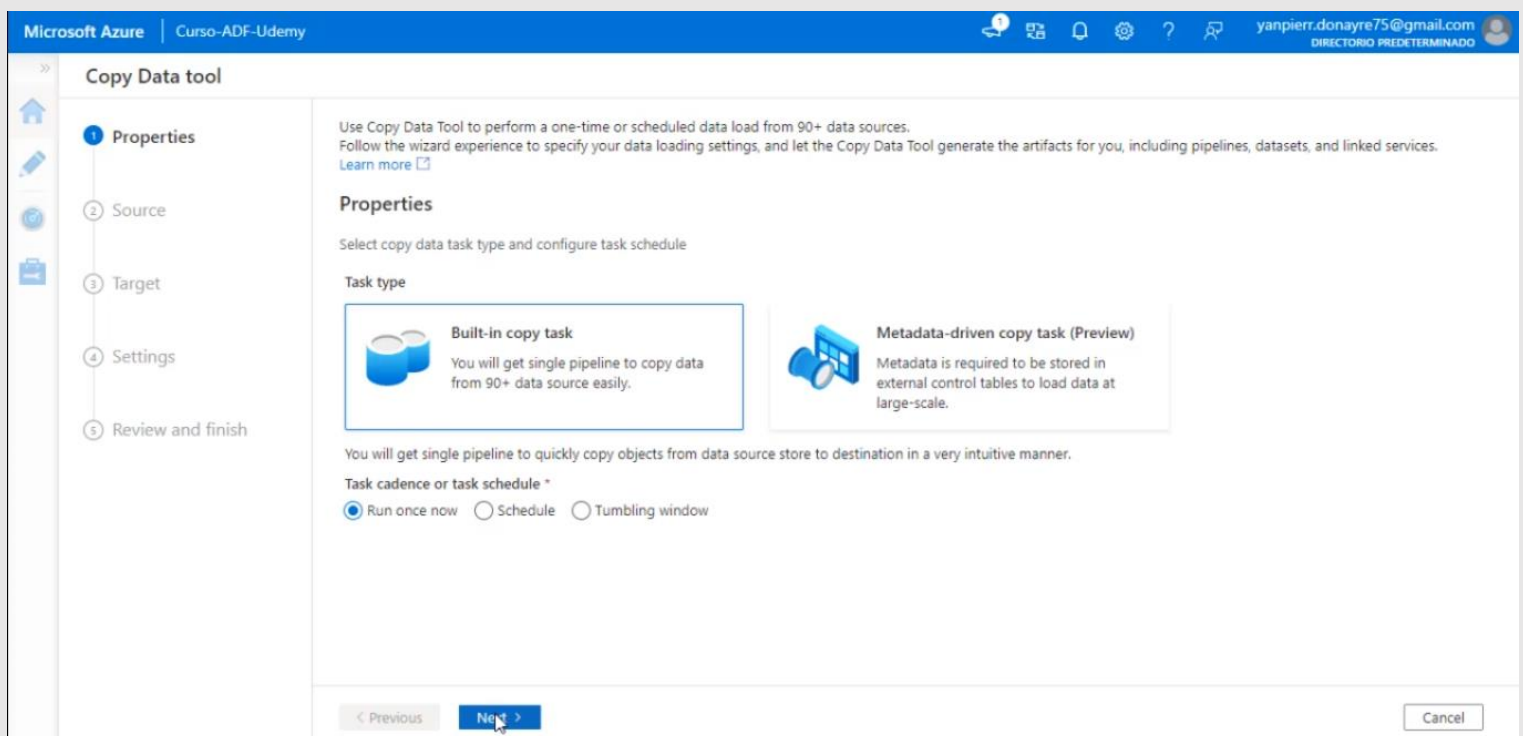
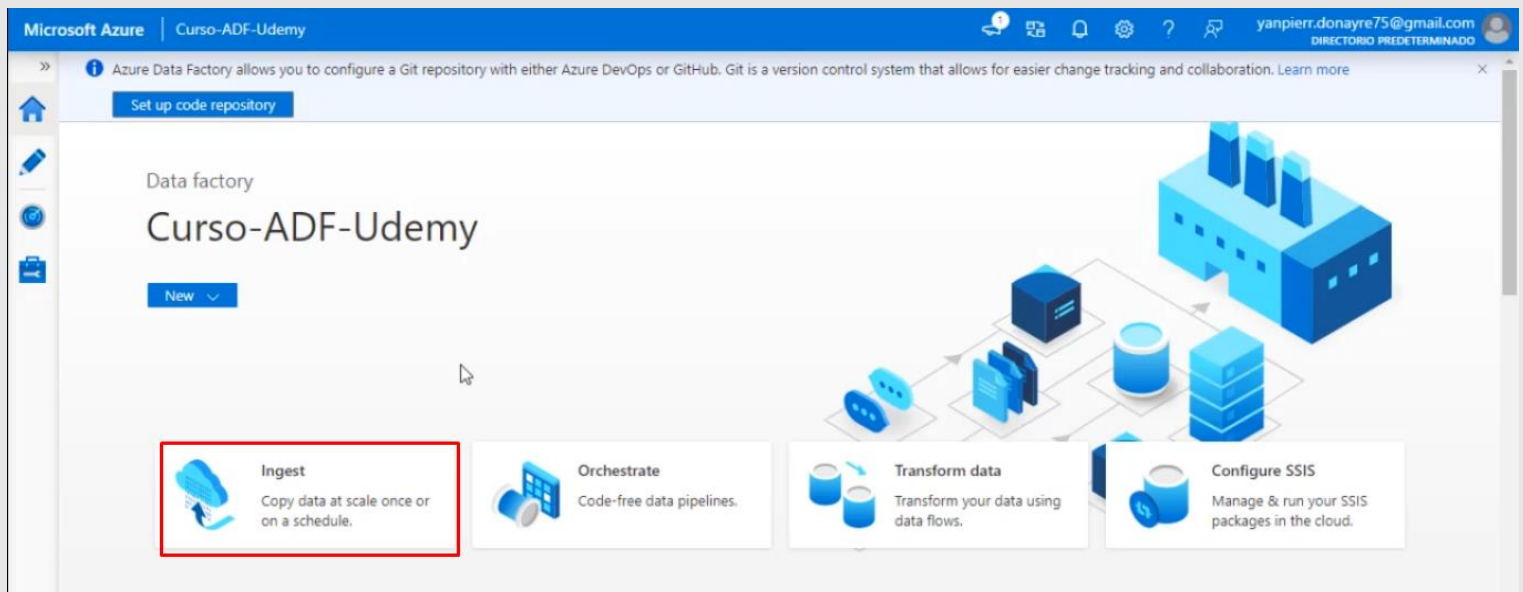


Ingesta de datos en Azure Data Factory desde un origen HTTP a Data Lake



Es probable que crear el Linked Service por este medio nos genere problemas más adelante. Sin embargo, podríamos crearlo antes de generar el proceso de copia de datos, desde la pestaña MANAGE

MN

Mariano

hace 10 meses

1

No, yo lo que hice fue lo siguiente:

1. Antes de realizar el proceso del video, haces click en Manage desde Azure Data Factory (el simbolo del portafolios con la herramienta).
2. Haces click en '+New' para agregar un Linked Service.
3. Agregas un HTTP con el link del video y listo. Luego vas a elegir ese Linked Service sin necesidad de hacerlo.

Microsoft Azure | Curso-ADF-Udemy

yanpierr.donayre75@gmail.com
DIRECTORIO PREDETERMINADO

Copy Data tool

Properties

Source

Dataset

Configuration

Source data store

Specify the source data store for the copy task. You can use an existing data store connection or specify a new data store.

Source type

HTTP

Connection *

Select...

New connection

URL Descarga: [https://cdn.rebrickable.com/media/downloads/](https://cdn.rebrickable.com/media/downloads/themes.csv.gz)
Archivo a descargar: themes.csv.gz

Microsoft Azure | Curso-ADF-Udemy

yanpierr.donayre75@gmail.com
DIRECTORIO PREDETERMINADO

Copy Data tool

Properties

Source

Dataset

Configuration

Target

Settings

Settings

Review and finish

Source data store

Specify the source data store for the copy task. You can use an existing data store connection or specify a new data store.

Source type

HTTP

Connection *

Select...

New connection (HTTP)

Name *

HTTP_Rebrickable

Description

Conexion a Lego Rebrickable

Connect via integration runtime *

AutoResolveIntegrationRuntime

Base URL *

https://cdn.rebrickable.com/media/downloads/

Add dynamic content [Alt+Shift+D]

Server Certificate Validation

Enable

Disable

Authentication type *

Anonymous

Auth headers

New

Annotations

New

Parameters

Create

Cancel

Test connection

Copy Data tool

- 2 Source

Source data store

Specify the source data store for the copy task. You can use an existing data store connection or specify a new data store.

Source type HTTP

Connection * HTTP_Rebrickable Edit New connection

Base URL

<https://cdn.rebrickable.com/media/downloads/>

Relative URL ⓘ

themes.csv.gz

Request method

GET

Additional headers ⓘ

Binary copy

Request timeout ⓘ

Max concurrent connections ⓘ

[← Previous](#)

Next

Cancel

Copy Data tool

- 7 Properties
- 2 Source
- Dataset
- Configuration
- 3 Target
- 4 Settings
- 5 Review and finish

File format settings

File format Text format

Column delimiter

Comma (,)

Row delimiter

Line feed (\n)

☒ First row as header ⓘ

> Advanced

Compression type

gzip (.gz)

Compression level

Optimal

Additional columns ⓘ
+ New

[← Previous](#)

[Next >](#)

Cancel

Microsoft Azure | Curso-ADF-Udemy

yanpierr.donayre75@gmail.com
DIRECTORIO PREDETERMINADO

Copy Data tool

- Properties
- Source
- Dataset
- Configuration
- Target
- Settings
- Review and finish

Preview data

Linked service: HTTP_Rebrickable

Object:

Preview	Schema	
18	Star Wars	1
1	19	Supplemental
1	20	Throwbot Slizer
1	21	Universal Building Set
	22	Creator
22	23	Basic Model
23	34	Building
23	35	Cargo

Cancel

Microsoft Azure | Curso-ADF-Udemy

yanpierr.donayre75@gmail.com
DIRECTORIO PREDETERMINADO

Copy Data tool

- Properties
- Source
- Dataset
- Configuration
- Target
- Settings
- Review and finish

File format settings

File format [ⓘ]
Text format

Column delimiter
Comma (,)

Row delimiter
Line feed (\n)

☒ First row as header [ⓘ]

> Advanced

Compression type
gzip (.gz)

Compression level
Optimal

Additional columns [ⓘ]

Microsoft Azure | Curso-ADF-Udemy

yanpierr.donayre75@gmail.com
DIRECTORIO PREDETERMINADO

Copy Data tool

- Properties
- Source
- Target
- Dataset
- Configuration

Destination data store

Specify the destination data store for the copy task. You can use an existing data store connection or specify a new data store.

Target type
Azure Data Lake Storage Gen2

Connection *
Select...

Microsoft Azure | Curso-ADF-Udemy

yanpierr.donayre75@gmail.com
DIRECTORIO PREDETERMINADO

Copy Data tool

Properties

Source

Target

Dataset

Configuration

Settings

Review and finish

Destination data store

Specify the destination data store for the copy task. You can use an existing

Target type

Azure Data Lake Storage Gen2

Connection *

Select...

New connection (Azure Data Lake Storage Gen2)

Name *

ADLS_DataLakeUdemyLaboratorio

Description

Conexión hacia ADLS

Connect via integration runtime *

AutoResolveIntegrationRuntime

Authentication method

Account key

Account selection method

From Azure subscription

Enter manually

Azure subscription

Azure subscription 1 (fe44dd55-1f86-41f6-8af9-98a293859ba8)

Storage account name *

datalakeudemylaboratio

Test connection

To linked service

To file path

Create

Cancel

Test connection

Microsoft Azure | Curso-ADF-Udemy

yanpierr.donayre75@gmail.com
DIRECTORIO PREDETERMINADO

Copy Data tool

Properties

Source

Target

Dataset

Configuration

Settings

Review and finish

Destination data store

Specify the destination data store for the copy task. You can use an existing data store connection or specify a new data store.

Target type

Azure Data Lake Storage Gen2

Connection *

ADLS_DataLakeUdemyLaboratorio

Folder path

If the identity you use to access the data store only has permission to subdirectory instead of the entire account, specify the path to browse.

lego

File name

themes.txt

Copy behavior

None

Max concurrent connections

Block size (MB)

Next

Cancel

Notar que colocó themes.txt, agregó el tipo de dato con el cual se quiere almacenar el archivo en el Data Lake.


Microsoft Azure | Curso-ADF-Udemy

yanpierr.donayre75@gmail.com
DIRECTORIO PREDETERMINADO

Copy Data tool

- Properties
- Source
- Target
- Dataset
- Configuration
- Settings
- Review and finish

File format settings


File format  Text format

Column delimiter Comma (,)

☐ Edit

Row delimiter Line feed (\n)

☐ Edit

☐ Add header to file 

> Advanced

Compression type None

Max rows per file

File name prefix

< Previous **Next >** Cancel

Al ocasionarse un error o un warning, estos se almacenarán en formato de log en la ruta indicada

Microsoft Azure | Curso-ADF-Udemy

yanpierr.donayre75@gmail.com
DIRECTORIO PREDETERMINADO

Copy Data tool


- Properties
- Source
- Target
- Settings
- Review and finish


Settings

Enter name and description for the copy data task, more options for data movement





Task name * Lego_Copy_Themes_HTTP_ADLS


Task description


Fault tolerance  Skip incompatible rows



Enable logging  ☒


Logging settings

Storage connection name *  ADLS_DataLakeUdemyLaboratorio   

Logging level  Warning

Logging mode  ☐ Reliable ☒ Best effort

Folder path  lego/errors/themes 

Enable staging  ☐

> Advanced

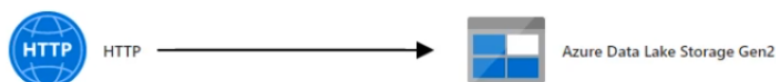
< Previous **Next >** Cancel

Copy Data tool

- ✓ Properties
- ✓ Source
- ✓ Target
- ✓ Settings
- 5 Review and finish
- Review
- Deployment

Summary

You are running pipeline to copy data from HTTP to Azure Data Lake Storage Gen2.



Properties

Task name Lego_Copy_Themes_HTTP_ADLS

Task description

Source

Connection name	HTTP_Rebrickable
-----------------	------------------

Dataset name	SourceDataset_arg
--------------	-------------------

Compression codec

Compression level Optimal

Column delimiter

 Edit Edit

[← Previous](#)

[Next >](#)

Cancel

Copy Data tool

- ✓ Properties
- ✓ Source
- ✓ Target
- ✓ Settings
- 5 Review and finish
- Review
- Deployment

Summary

You are running pipeline to copy data from HTTP to Azure Data Lake Storage Gen2.

Task description

Source

Connection name	HTTP_Rebrickable
-----------------	------------------

Dataset name HTTP_Lego_Themes

Compression codec	gzip
-------------------	------

Compression level Optimal

Column delimiter ,

Row delimiter

Escape character	\
------------------	---

Quote char " "

```
First row as header      true
```

Relative URL	themes.csv.gz
--------------	---------------

 Edit

[← Previous](#)

[Next >](#)

Cancel

Microsoft Azure | Curso-ADF-Udemy

yanpierr.donayre75@gmail.com
DIRECTORIO PREDETERMINADO

Copy Data tool

Properties
Source
Target
Settings
5 Review and finish
Review
Deployment

HTTP → Azure Data Lake Storage Gen2

Deployment complete

Deployment step	Status
Validating copy runtime environment	✓ Succeeded
✓ Creating datasets	✓ Succeeded
✓ Creating pipelines	✓ Succeeded
> Running pipelines	✓ Succeeded

Datasets and pipelines have been created. You can now monitor and edit the copy pipelines or click finish to close Copy Data Tool.

Finish Edit pipeline **Monitor**

Microsoft Azure | Curso-ADU-Udemy

yanpierr.donayre75@gmail.com
DIRECTORIO PREDETERMINADO

Pipeline runs

Triggered Debug Rerun Cancel Refresh Edit columns List Gantt

Filter by run ID or name Bogota, Lima, Quito (... Last 24 hours Pipeline name: Lego_Copy_Themes_HTTP_ADLS Status: All Copy filters

Runs: Latest runs Triggered by: All Add filter

Showing 1 - 1 items

Pipeline name	Run start	Run end	Duration	Triggered by	Status	Error
Lego_Copy_Themes_HTTP_AD...	11/30/21, 4:09:37 PM	11/30/21, 4:09:48 PM	00:00:11	Manual trigger	✓ Succeeded	

Microsoft Azure Storage Explorer

Archivo Editar Ver Ayuda

EXPLORADOR

Buscar recursos Actualizar todo

Contraer todo

- Acceso rápido
- Local y asociado
- Azure subscription 1 (yanpierr.donayre75@gmail.com)
- Cuentas de almacenamiento
 - datalakeudemylaboratorio (ADLS Gen2)
 - Blob Containers
 - dataset
 - lego**
 - publico
 - File Shares
 - Queues
 - Tables
 - storageudemylaboratorio
 - Disks
 - rg-UdemyLab
 - Cuentas de Cosmos DB (en desuso)
 - Data Lake Storage Gen1 (Preview)

lego

Cargar Descargar Abrir Nueva carpeta Seleccionar todo Copiar Pegar Cambiar nombre Mover Administrar ACL Propiedades Eliminar Más

Nombre	Nivel de acceso	Última modificación del nivel de acceso	Última modificación	Tipo de blob	Tipo de contenido	Tamaño	Estado	Id.
errors			30/11/2021 16:09		Carpeta		Activo	
themes.txt	Frecuente (inferido)		30/11/2021 16:09	Blob en bloques	application/octet-stream	10.4 KB	Activo	

Se muestran del 1 al 2 de 2 elementos en caché

Acciones Propiedades Actividades

URL ht
Dominio personalizado Bl
Tipo Bl

Borrado completado Borrado correcto

Microsoft Azure | Buscar recursos, servicios y documentos (G+)

Todos los servicios > Cuentas de almacenamiento > datalakeudemylaboratio

datalakeudemylaboratio | Contenedores

Cuenta de almacenamiento

Buscar (Ctrl+/) << + Contenedor Cambiar nivel de acceso Restaurar contenedores Actualizar Eliminar

Buscar contenedores por prefijo

Mostrar contenedores eliminados

Nombre	Última modificación	Nivel de acceso público	Estado de concesión
<input type="checkbox"/> dataset	26/11/2021 16:15:46	Privada	Available ***
<input type="checkbox"/> lego	30/11/2021 16:09:44	Privada	Available ***
<input type="checkbox"/> publico	26/11/2021 16:26:43	Contenedor	Available ***

Almacenamiento de datos

- Contenedores
- Recursos compartidos de archivos
- Colas
- Tablas

Seguridad y redes

- Redes
- Claves de acceso
- Firma de acceso compartido
- Cifrado
- Seguridad

Administración de datos

Microsoft Azure | Buscar recursos, servicios y documentos (G+)

Todos los servicios > Cuentas de almacenamiento > datalakeudemylaboratio >

lego

Contenedor

Buscar (Ctrl+/) << Cargar + Agregar directorio Actualizar Cambiar nombre Eliminar Cambiar nivel ...

Método de autenticación: Clave de acceso (Cambiar a la cuenta de usuario de Azure AD)

Ubicación: lego

Buscar blobs por prefijo (distingue mayúsculas de minúsculas)

Nombre	Modificado	Nivel de acceso	Estado del archivo	Tipo de blob
<input type="checkbox"/> errors				
<input type="checkbox"/> themes.txt	30/11/2021 16:09:46	Frecuente (inferido)		Block blob

Overview

- Diagnosticar y solucionar problemas
- Control de acceso (IAM)

Configuración

- Tokens de acceso compartido
- Administrar dirección ACL
- Directiva de acceso
- Properties
- Metadatos