

# How to load csv files from Azure Data Lake into Azure Synapse

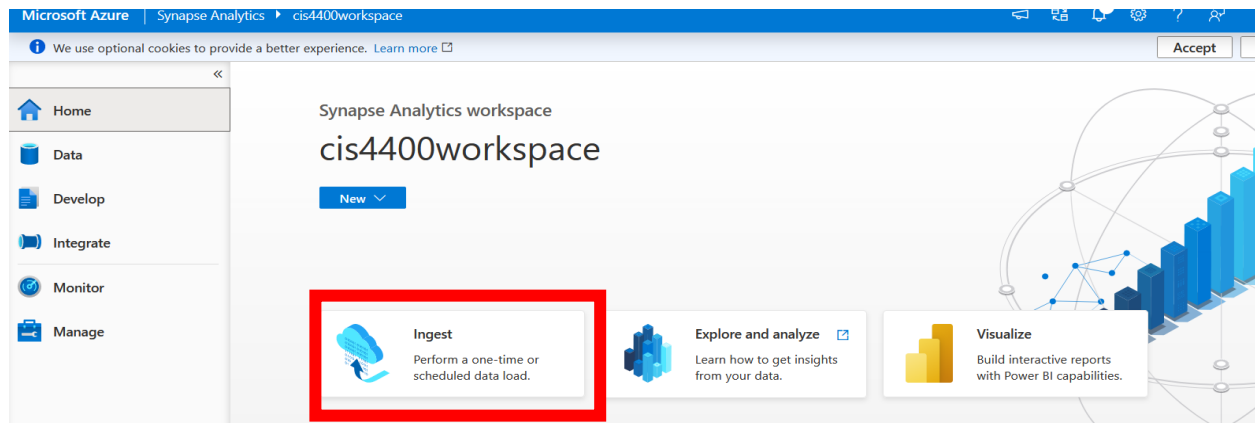
## Pre-work:

1. Prerequisites: Azure Synapse Analytics workspace and new Dedicated SQL pool
2. Ensure that the empty SQL tables are already created inside of the database before reading the csv files into them. This will make everything so much easier.

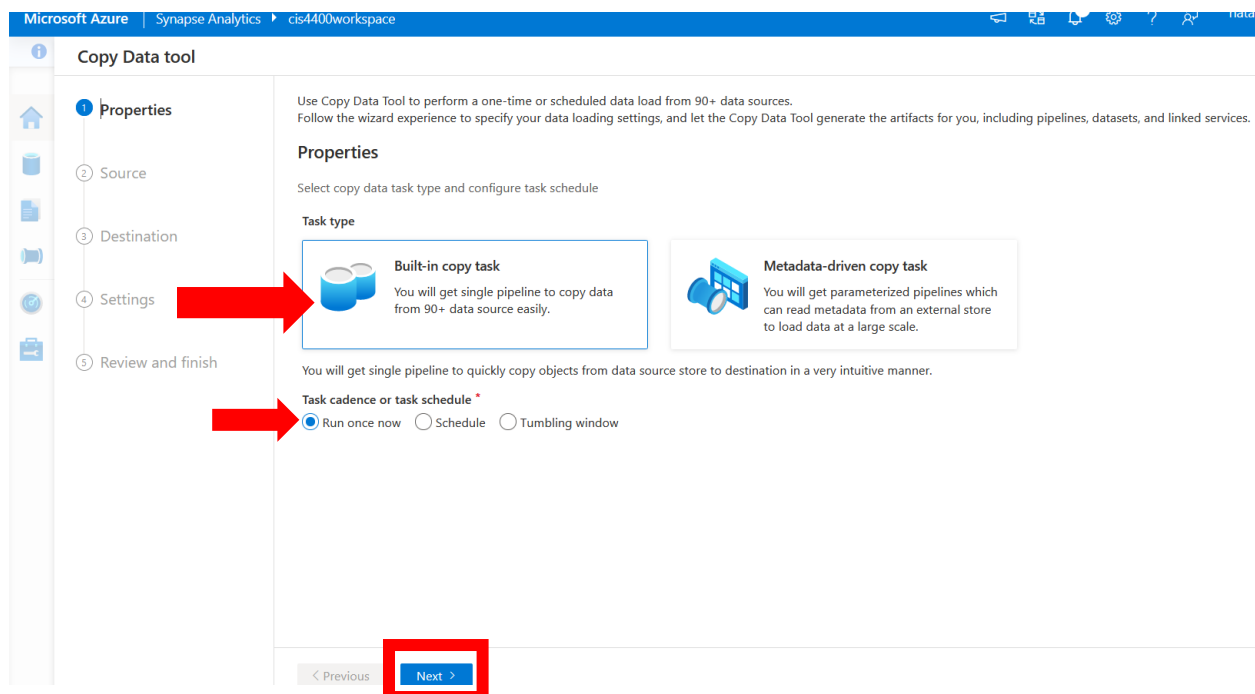
## Process:

Credits: <https://www.youtube.com/watch?v=JnFrY3LzF90> <- great tutorial if you prefer video format

When you launch Synapse Studio, navigate to the “Home” tab, and select “Ingest”.



Select “Built-in copy task”, and “Run once now”. You will repeat the process for each file. Click “Next”.



Click on “ + New Connection”, select the Azure Data Lake icon, from there select your subscription and the name of your Data Lake and “Create” the connection to the Data Lake When your connection is working, select the csv file by clicking on “Browse”. Next, uncheck “Recursively” and hit “Next”.

Microsoft Azure | Synapse Analytics | cis4400workspace

### Copy Data tool

**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** All

**Connection \*** AzureDataLakeStorage1 [Edit](#) [+ New connection](#)

**Integration runtime \*** AutoResolveIntegrationRuntime [Edit](#)

**File or folder**  
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.

datalakefinancialratios/FinancialDataFiles/clean\_dates.csv [Browse](#)

**Options**

☐ Binary copy ⓘ

☐ Recursively ⓘ

☐ Enable partitions discovery ⓘ

**Max concurrent connections ⓘ**

Filter by last modified

**Start time (UTC)** **End time (UTC)**

[< Previous](#) [Next >](#)

On the next page, keep the format settings below for csv files, ensure that “First row as header” is checked. Click on “Preview data”.

Microsoft Azure | Synapse Analytics | cis4400workspace

### Copy Data tool

**File format settings**

**File format** DelimitedText [Detect text format](#) [Preview data](#)

**Column delimiter** Comma (,) [Edit](#)

**Row delimiter** Default (\r\n, or \n) [Edit](#)

☒ First row as header ⓘ

**Advanced**

**Compression type** Select...

**Additional columns ⓘ**

[+ New](#)

If you do not see your data in the new window, go “Source Data Store” page and actually browse and select the file. If you are satisfied with the preview, hit “Next”.

Microsoft Azure | Synapse Analytics | cis4400workspace

**Preview data**

Linked service: AzureDataLakeStorage1

Object: datalakefinancialratios/FinancialDataFiles/clean\_dates.csv

**Preview** Schema

DateKey	DateISO	Year	QuarterNumber	QuarterName	MonthNumber	MonthName	DayNumber	DayN
20230101	2023-01-01	2023	1	1	1	01	1	01
20230102	2023-01-02	2023	1	1	1	01	2	02
20230103	2023-01-03	2023	1	1	1	01	3	03
20230104	2023-01-04	2023	1	1	1	01	4	04
20230105	2023-01-05	2023	1	1	1	01	5	05
20230106	2023-01-06	2023	1	1	1	01	6	06
20230107	2023-01-07	2023	1	1	1	01	7	07
20230108	2023-01-08	2023	1	1	1	01	8	08
20230109	2023-01-09	2023	1	1	1	01	9	09

< Previous **Next >**

In the next “Destination Data Store” page, click on “+ New Connection”, select the “Azure Synapse Analytics workspace” icon, from there select your subscription, the server name, database name (SQL pool name), and authenticate using your Azure Synapse workspace username and password that you created (if you don’t remember, in a new tab open the Synapse workspace and select Overview). Next, click “Create” the connection”.

When prompted to create the Target, click on “Use existing table”, and in the drop-down menu select your target database table that you want to populate. Hit “Next”.

Microsoft Azure | Synapse Analytics | cis4400workspace

### Copy Data tool

**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.

Destination type: All

Connection: AzureSynapseAnalytics1 New connection

Integration runtime: AutoResolveIntegrationRuntime

Source: clean\_dates → Target: dbo.date\_dim Auto-create a destination table with the source schema

☐ Skip column mapping for all tables

< Previous Next >

Then, you will see the mapping between the columns in the csv file and the SQL table columns, verify that the column names match, and click “Next” for now, we will come back to this page.

Microsoft Azure | Synapse Analytics | cis4400workspace

### Copy Data tool

**Column mapping**

Choose how source and destination columns are mapped

Table mappings (1)

- Source: Azure Data Lake Storage Gen2 file
- Target: dbo.date\_dim

Column mappings

Type conversion settings

+ New mapping Clear Reset Delete

Source	Type	Destination	Type
DateKey	abc String	DateKey	123 int
DateISO	abc String	DateISO	date
Year	abc String	Year	123 int
QuarterNumber	abc String	QuarterNumber	123 int
QuarterName	abc String	QuarterName	abc varchar
MonthNumber	abc String	MonthNumber	123 int
MonthName	abc String	MonthName	abc varchar

Azure Synapse Analytics sink properties

Pre-copy script

< Previous Next > Cancel

In the Settings, give the task a meaningful name. Check the box next to “Enable Staging”. Select your Data Lake as Staging account linked service and test the connection. Scroll down and for “Copy method” select “PolyBase”. Then hit “Previous”, to go back to the column mapping.

Microsoft Azure | Synapse Analytics | cis4400workspace

### Copy Data tool

1 Properties  
2 Source  
3 Destination  
4 Settings  
5 Review and finish

#### Settings

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

Task name \*

Task description

Data consistency verification ☐

Fault tolerance

Enable logging ☐

Enable staging ☒

Staging settings

Staging account linked service \*  [Test connection](#) [Edit](#) [New](#)  
Connection successful

Integration runtime \*  [Edit](#)

Storage Path  [Browse](#)

Enable Compression ☐

Advanced

[< Previous](#) [Next >](#)

Microsoft Azure | Synapse Analytics | cis4400workspace

### Copy Data tool

1 Properties  
2 Source  
3 Destination  
4 Settings  
5 Review and finish

#### Settings

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

Enable logging ☐

Enable staging ☒

Staging settings

Staging account linked service \*  [Test connection](#) [Edit](#) [New](#)  
Connection successful

Integration runtime \*  [Edit](#)

Storage Path  [Browse](#)

Enable Compression ☐

Advanced

Copy method ☐ Copy command ☒ PolyBase ☐ Bulk insert ☐ Upsert

Allow PolyBase ☒

Reject type

Reject value

[< Previous](#) [Next >](#)

There, uncheck the “Type conversion” option. Now, you can modify the column datatypes of your csv. Taking the date tables as an example, all columns start off as type String. Let’s change the Quarter, Month and DateNumber columns to Int32, and the DateISO column to DateTime format. Click “Next”.

Microsoft Azure | Synapse Analytics | cis4400workspace | natalia.kurbakova21@login.cuny.edu

**Copy Data tool**

Properties  
Source  
Destination  
Dataset  
Configuration  
Settings  
Review and finish

**Column mapping**  
Choose how source and destination columns are mapped

Table mappings (1)  
Type conversion ☒ **Next**

Source: Azure Data Lake Storage Gen2 file  
Target: dbo.date\_dim

Column mappings

Source	Type	Destination	Type
DateKey	abc String	DateKey (int)	int
DateISO	abc String	DateISO (date)	date
Year	abc String	Year (int)	int
QuarterNumber	abc String	QuarterNumber (int)	int
QuarterName	abc String	QuarterName (varchar)	varchar

Azure Synapse Analytics sink properties

Pre-copy script

Advanced

Previous Next Cancel

Microsoft Azure | Synapse Analytics | cis4400workspace | natalia.kurbakova21@login.cuny.edu

**Copy Data tool**

Properties  
Source  
Destination  
Dataset  
Configuration  
Settings  
Review and finish

**Column mapping**  
Choose how source and destination columns are mapped

Table mappings (1)  
Type conversion ☐

Source: Azure Data Lake Storage Gen2 file  
Target: dbo.date\_dim

Column mappings

Source	Type	Destination	Type
DateKey	123 Int32	DateKey (int)	int
DateISO	DateTime	DateISO (date)	date
Year	123 Int32	Year (int)	int
QuarterNumber	123 Int32	QuarterNumber (int)	int
QuarterName	abc String	QuarterName (varchar)	varchar
MonthNumber	123 Int32	MonthNumber (int)	int
MonthName	abc String	MonthName (varchar)	varchar
DayNumber	123 Int32	DayNumber (int)	int
DayName	abc String	DayName (varchar)	varchar

Azure Synapse Analytics sink properties

Previous **Next** Cancel

Now, you are in the “Settings” again – nothing to change here anymore, click on “Next”.

Microsoft Azure | Synapse Analytics | cis4400workspace

### Copy Data tool

- Properties
- Source
- Destination
- Settings**
- Review and finish

#### Settings

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

Task name \*

Task description

Data consistency verification ☐

Fault tolerance

Enable logging ☐

Enable staging ☒

Staging settings

Staging account linked service \*  [Test connection](#) [Edit](#) [+](#) [New](#)

Integration runtime \*  [Edit](#)

Storage Path  [Browse](#)

Enable Compression ☐

Advanced

[< Previous](#) [Next >](#)

In the Summary page, review the data pipeline and hit “Next”.

Microsoft Azure | Synapse Analytics | cis4400workspace

### Copy Data tool

- Properties
- Source
- Destination
- Settings
- Review and finish**
- Review
- Deployment

#### Summary

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

**Your data pipeline:**

```

graph LR
    A[Azure Data Lake Storage Gen2] --> B[Azure Data Lake Storage Gen2]
    B --> C[Azure Synapse Analytics]
  
```

Properties

Task name

Task description

Source

Connection name

Dataset name

Column delimiter

Escape character

Quote char

First row as header ☒

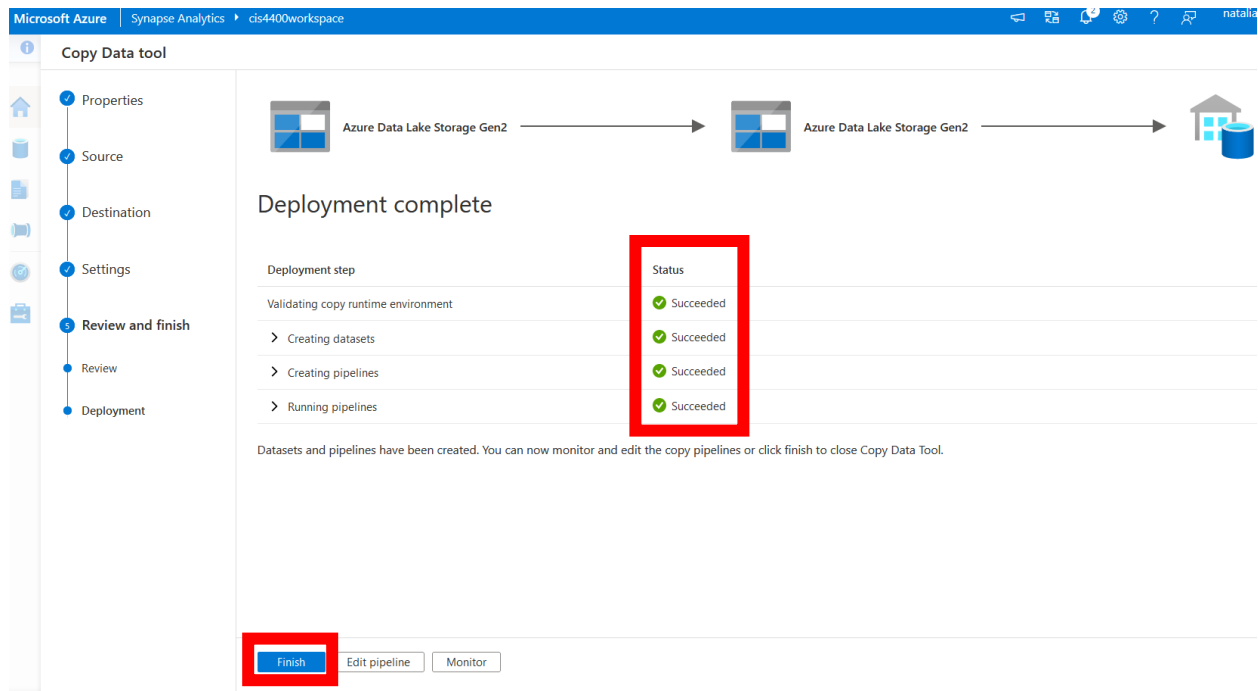
File name

Folder path

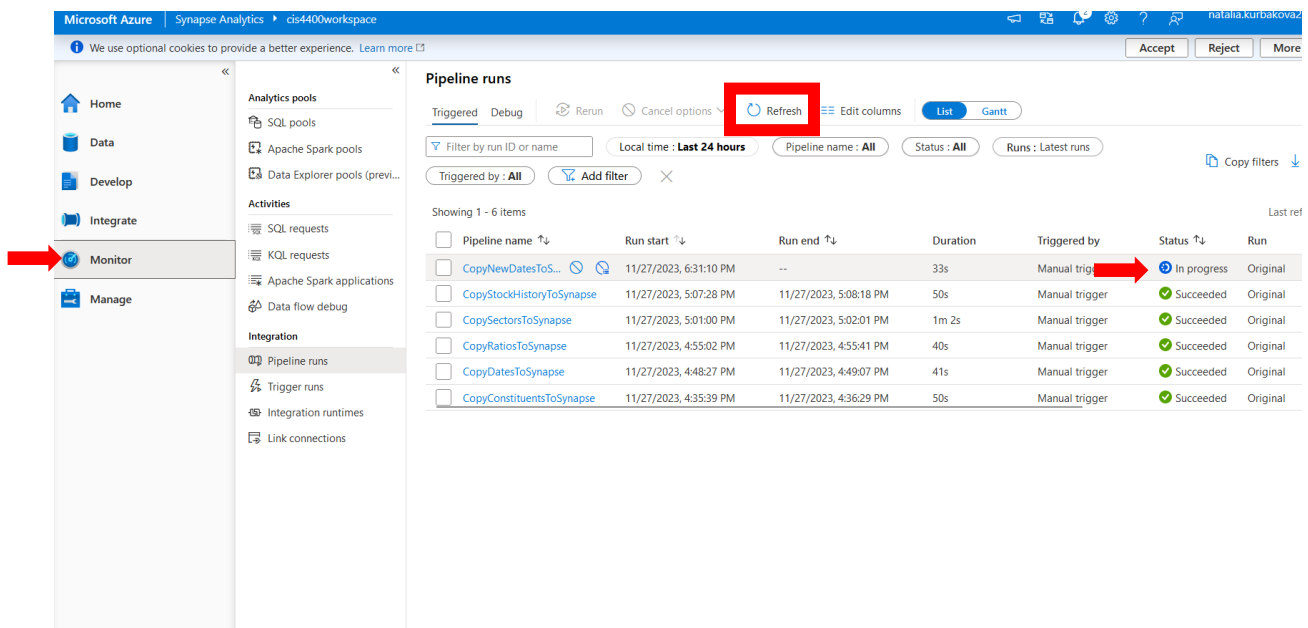
Destination

[< Previous](#) [Next >](#) [Cancel](#)

Once each Development step has a green check mark and a status of succeeded, click “Finish”.



Go to the “Monitor” tab and wait no more than 1-2 minutes for the new pipeline to complete running. Afterwards, click “Refresh”.



Once you see the green checkmark next to the data pipeline you just deployed, proceed to the next step by switching to the “Data” tab.



Microsoft Azure | Synapse Analytics | cis4400workspace

We use optional cookies to provide a better experience. [Learn more](#)

Accept Reject

Home Data Develop Integrate Monitor Manage

Analytics pools

- SQL pools
- Apache Spark pools
- Data Explorer pools (previ...

Activities

- SQL requests
- KQL requests
- Apache Spark applications
- Data flow debug

Integration

- Pipeline runs
- Trigger runs
- Integration runtimes
- Link connections

Pipeline runs

Triggered Debug Rerun Cancel options Refresh Edit columns List Gantt

Filter by run ID or name Local time: Last 24 hours Pipeline name: All Status: All Runs: Latest runs

Triggered by: All Add filter

Showing 1 - 6 items

Pipeline name	Run start	Run end	Duration	Triggered by	Status
CopyNewDatesToSynapse	11/27/2023, 6:31:10 PM	11/27/2023, 6:32:05 PM	56s	Manual trigger	Succeeded
CopyStockHistoryToSynapse	11/27/2023, 5:07:28 PM	11/27/2023, 5:08:18 PM	50s	Manual trigger	Succeeded
CopySectorsToSynapse	11/27/2023, 5:01:00 PM	11/27/2023, 5:02:01 PM	1m 2s	Manual trigger	Succeeded
CopyRatiosToSynapse	11/27/2023, 4:55:02 PM	11/27/2023, 4:55:41 PM	40s	Manual trigger	Succeeded
CopyDatesToSynapse	11/27/2023, 4:48:27 PM	11/27/2023, 4:49:07 PM	41s	Manual trigger	Succeeded
CopyConstituentsToSynapse	11/27/2023, 4:35:39 PM	11/27/2023, 4:36:29 PM	50s	Manual trigger	Succeeded

As the final step, go to the Data tab. Right click the database and create a new empty SQL script (for example, `SELECT * FROM your_database.schema.date_dimension_table;`). Run the script to verify that your previously empty SQL table has been populated with data from the csv file.

Microsoft Azure | Synapse Analytics | cis4400workspace

We use optional cookies to provide a better experience. [Learn more](#)

Accept Reject More options

Home Data Develop Integrate Monitor Manage

Synapse live Validate all Publish all

Workspace Linked

Filter resources by name

SQL database

- financial\_dw (SQL)
  - Tables
    - dbo.company\_dim
    - dbo.date\_dim
    - dbo.fact\_financial\_ratios
    - dbo.fact\_stock\_performance
    - dbo.industry\_dim
  - External tables
  - External resources
  - Views
  - Programmability
  - Schemas
  - Security

SQL script 1

Run Undo Publish Query plan Connect to financial\_dw Use database financial\_dw

`SELECT * FROM financial_dw.dbo.date_dim;`

Results Messages

View Table Chart Export results

DateKey	DateISO	Year	QuarterNumber	QuarterName	MonthNumber	MonthName
20230101	2023-01-01T00:00:00.0000000	2023	1	1	1	01
20230102	2023-01-02T00:00:00.0000000	2023	1	1	1	01
20230103	2023-01-03T00:00:00.0000000	2023	1	1	1	01
20230104	2023-01-04T00:00:00.0000000	2023	1	1	1	01
20230105	2023-01-05T00:00:00.0000000	2023	1	1	1	01
20230106	2023-01-06T00:00:00.0000000	2023	1	1	1	01
20230107	2023-01-07T00:00:00.0000000	2023	1	1	1	01
20230108	2023-01-08T00:00:00.0000000	2023	1	1	1	01

00:00:01 Query executed successfully.

Properties

General Related (0)

Name \* SQL script 1

Description

Type .sql script

Size 83 bytes

Results settings per query

First 5000 rows (default)

All rows

Congrats on loading the SQL tables in Azure Synapse! Repeat the process for all the tables you have.

Now you can serve the data from Azure Synapse Analytics to visualize insights!