

## Creating an ADLS GEN 2 Account and the CSV files are in the folders

The screenshot shows the Microsoft Azure Storage Container list for the account 'adlsnycpayrollsashyab'. The left sidebar includes links for Overview, Activity log, Tags, Diagnose and solve problems, Access Control (IAM), Data migration, Events, Storage browser, and Storage Mover. Under Data storage, 'Containers' is selected. The main area displays a table of containers:

Name	Last modified	Anonymous access level	Lease state
\$logs	9/10/2023, 10:06:03 AM	Private	Available
dirhistoryfiles	9/10/2023, 10:08:13 AM	Private	Available
dirpayrollfiles	9/10/2023, 10:08:05 AM	Private	Available
dirstaging	9/10/2023, 10:08:21 AM	Private	Available

## History container

The screenshot shows the Microsoft Azure Storage Container details for the 'dirhistoryfiles' container. The left sidebar includes links for Overview, Diagnose and solve problems, Access Control (IAM), Shared access tokens, Access policy, Properties, and Metadata. The main area displays a table of blobs:

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
nycpayroll_2020.csv	9/10/2023, 10:11:09 ...	Hot (Inferred)		Block blob	17.07 KiB	Available

## Payroll Container

The screenshot shows the Azure Storage Container blade for the 'dirpayrollfiles' container. The container has four blob items:

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
AgencyMaster.csv	9/10/2023, 10:09:41 AM	Hot (Inferred)		Block blob	4.9 KiB	Available
EmpMaster.csv	9/10/2023, 10:09:41 AM	Hot (Inferred)		Block blob	23.04 KiB	Available
nycpayroll_2021.csv	9/10/2023, 10:09:41 AM	Hot (Inferred)		Block blob	16.45 KiB	Available
TitleMaster.csv	9/10/2023, 10:09:41 AM	Hot (Inferred)		Block blob	49.04 KiB	Available

## CREATING AZURE DATA FACTORY RESOURCE

The screenshot shows the Microsoft.DataFactory-20230910101932 | Overview page. Deployment details are as follows:

Deployment name	Start time
Microsoft.DataFactory-20230910101932	9/10/2023, 10:20:04 AM

Next steps include:

- Deployment details
- Next steps

Callouts on the right side provide links to Cost management, Microsoft Defender for Cloud, Free Microsoft tutorials, and Work with an expert.

## CREATING SYNAPSE AND SQL POOLS AND TABLES

The screenshot shows the Microsoft Azure Synapse Analytics deployment overview page. The deployment name is "Microsoft.Azure.SynapseAnalytics-20230910102418". The status is "Deployment succeeded". Deployment details include a start time of 9/10/2023, 10:26:19 AM, and a correlation ID of ce042c27-419c-41ce-96c3-7a8c691b4ebd. The resource group is "udacity\_final\_project\_resource\_group". There are sections for "Deployment details" and "Next steps". A "Go to resource group" button is present. On the right, there are promotional cards for "Cost management", "Microsoft Defender for Cloud", "Free Microsoft tutorials", and "Work with an expert". The browser toolbar at the top shows multiple tabs and the URL https://portal.azure.com/#view/HubsExtension/DeploymentDetailsBlade~/overview/id%2Fsu... The status bar at the bottom indicates Sun Sep 10 10:30 AM.

The screenshot shows the Azure Synapse Analytics workspace under the "udacity-project-synapse" workspace. The left sidebar navigation includes "Synapse live", "Validate all", and "Publish all". The main area is titled "SQL pools" and contains the following text: "The serverless SQL pool, Built-in, is immediately available for your workspace. Dedicated SQL pools can be configured to adapt to team or organizational requirements and constraints. Learn more". Below this is a table showing two items:

Name	Type	Status	Size
Built-in	Serverless	Online	Auto
mysqlpool	Dedicated	Deploying	DW100c

The left sidebar also lists other workspace components: Analytics pools, SQL pools (selected), Apache Spark pools, Data Explorer pools (preview), External connections, Linked services, Microsoft Purview, Integration, Triggers, Integration runtimes, Security, Access control, Credentials, Managed private endpoints, Configurations + libraries, Workspace packages, Data flow libraries, Apache Spark configurations, and Source control.

**(ALL MY TABLES ARE ON THE LEFT SIDE)**

## SQL DB AND THE TABLE

## CREATE LINKED SERVICES

The screenshot shows the Microsoft Azure Data Factory interface within a Firefox browser window. The URL is <https://adf.azure.com/en/management/datalinkedservices?factory=%2Fsubscriptions%2F...>. The browser status bar indicates it's Sunday, September 10, at 11:26 AM, with a battery level of 52%.

The left sidebar menu includes options like General, Factory settings, Connections, Linked services (which is selected), Integration runtimes, Microsoft Purview, Source control, Git configuration, ARM template, Author, Triggers, Global parameters, Data flow libraries, Security, Credentials, Customer managed key, Outbound rules, Managed private endpoints, Workflow orchestration manager, and Apache Airflow.

The main content area is titled "Linked services" and contains a table listing three existing linked services:

Name	Type	Related	Annotations
adls_ls	Azure Data Lake Storage Gen2	0	
azure_synapse_linked_service	Azure Synapse Analytics	0	
sql_db_linked_service	Azure SQL Database	0	

## CREATE DATA SETS

---

### ◀ Datasets

11

-  adls\_ds\_2021
  -  adls\_ds\_agency\_master
  -  adls\_ds\_emp\_master
  -  adls\_ds\_title\_master
  -  adls\_historical\_2020\_s
  -  sql\_db\_ds\_payroll ...
  -  synaps sql\_db\_ds\_payroll
  -  synapse\_ds\_payroll\_agency
  -  synapse\_ds\_payroll\_data
  -  synapse\_ds\_summary
  -  synapse\_ds\_title\_md
- 

## CREATING DATA FLOWS AND PIPELINES

## ◀ Data flows

6

↗️ dataflowagencytosynapse

...

↗️ dataflowaggregation

↗️ dataflowemptosynapse

↗️ dataflowpayrolltosqlldb

...

↗️ dataflowsqldbtosynapse

↗️ dataflowtitletosynapse

## Pipelines

## ◀ Pipelines

...

⌚ agency\_to\_synapse\_pipeline

⌚ all\_data\_to\_synapse\_pipeline

⌚ current\_year\_from\_sql\_db\_to\_synaps...

⌚ employee\_to\_synapse\_pipeline

⌚ load2021tosqlldb

⌚ title\_to\_synapse\_pipeline

Pipeline name ↑↓	Run start ↑↓	Duration	Triggered by	Status ↑↓	Parameters	Run ID
title_to_synapse_pipeline	9/10/2023, 6:06:55 PM	3m 40s	Manual trigger	✓ Succeeded	e48f9b5d-f4d0-42b5-8afb-581bc810cb3f	
current_year_from_sql_db_to_...	9/10/2023, 6:06:50 PM	3m 54s	Manual trigger	✓ Succeeded	c7a8af9a-4dd4-4d0e-80d3-4faea6041a3b	
load2021tosqldb	9/10/2023, 6:06:45 PM	3m 24s	Manual trigger	✓ Succeeded	e6ead9f0-b362-4117-be22-e54218e6ac96	
employee_to_synapse_pipeline	9/10/2023, 6:06:36 PM	3m 57s	Manual trigger	✓ Succeeded	01609ba9-3d24-49b3-9edc-43b53286cbdb	
all_data_to_synapse_pipeline	9/10/2023, 6:06:30 PM	3m 56s	Manual trigger	✓ Succeeded	1c6248bd-ad72-4277-a432-3993b5e448a2	
agency_to_synapse_pipeline	9/10/2023, 6:06:22 PM	3m 17s	Manual trigger	✓ Succeeded	c0e9577b-3b0e-4834-8505-0fc698ef9f85	
current_year_from_sq..._to_...	9/10/2023, 5:59:19 PM	3m 25s	Manual trigger	✓ Succeeded	062cb8f3-d52c-4534-99a5-48268c57d1f8	
all_data_to_synapse_pipeline	9/10/2023, 5:57:49 PM	3m 38s	Manual trigger	✓ Succeeded	a52554af-68e2-4210-8720-1cda9c34fab6	
load2021tosqldb	9/10/2023, 5:47:50 PM	3m 30s	Manual trigger	✓ Succeeded	415c9ad7-0b0d-49af-9c2b-8a684f0b43ab	

## CREATE NYC SUMMARY TABLE IN SYNPASE

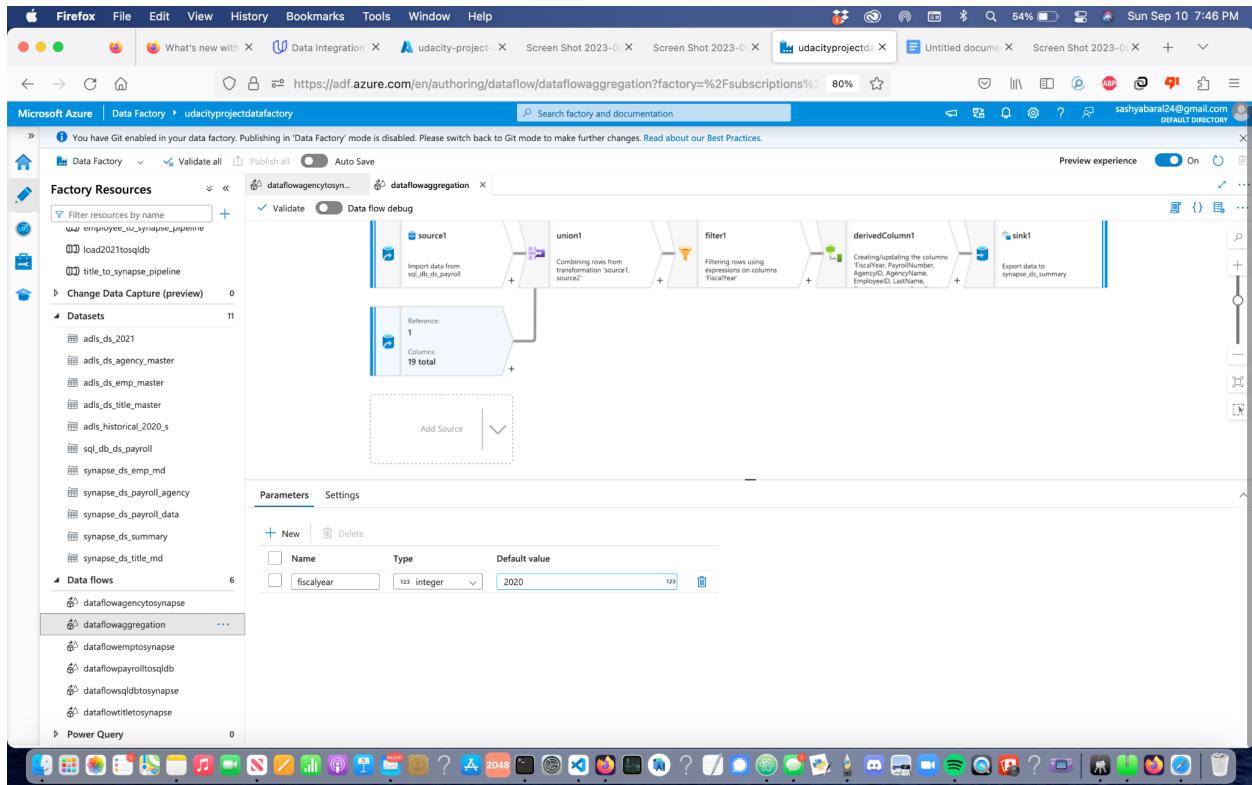
## CREATE dataset for the historical files

### ◀ Datasets

11

- grid adls\_ds\_2021
- grid adls\_ds\_agency\_master
- grid adls\_ds\_emp\_master
- grid adls\_ds\_title\_master
- grid adls\_historical\_2020\_s
- grid sql\_db\_ds\_payroll
- grid synapse\_ds\_emp\_md
- grid synapse\_ds\_payroll\_agency
- grid synapse\_ds\_payroll\_data
- grid synapse\_ds\_summary
- grid synasynapse\_ds\_summary

## CREATE AGGREGATE DATA FLOW



## PIPELINE

### ► Pipelines 6

agency\_to\_synapse\_pipeline

all\_data\_to\_synapse\_pipeline

current\_year\_from\_sql\_db\_to\_synaps...

employee\_to\_synapse\_pipeline

load2021tosqldb

title\_to\_synapse\_pipeline

## PIPELINE RUNS

<input type="checkbox"/>	Pipeline name ↑↓	Run start ↑↓	Duration	Triggered by	Status ↑↓	Parameters	Run ID
<input type="checkbox"/>	<a href="#">title_to_synapse_pipeline</a>	9/10/2023, 6:06:55 PM	3m 40s	Manual trigger	<span>✓ Succeeded</span>		e48f9b5d-f4d0-42b5-8aft-581bc810cb3f
<input type="checkbox"/>	<a href="#">current_year_from_sql_db_to_...</a>	9/10/2023, 6:06:50 PM	3m 54s	Manual trigger	<span>✓ Succeeded</span>		c7a8af9a-4dd4-4d0e-80d3-4faea6041a3b
<input type="checkbox"/>	<a href="#">load2021tosqldb</a>	9/10/2023, 6:06:45 PM	3m 24s	Manual trigger	<span>✓ Succeeded</span>		e6ead9f0-b362-4117-be22-e54218e6ac96
<input type="checkbox"/>	<a href="#">employee_to_synapse_pipeline</a>	9/10/2023, 6:06:36 PM	3m 57s	Manual trigger	<span>✓ Succeeded</span>		01609ba9-3d24-49b3-9edc-43b53286cbde
<input type="checkbox"/>	<a href="#">all_data_to_synapse_pipeline</a>	9/10/2023, 6:06:30 PM	3m 56s	Manual trigger	<span>✓ Succeeded</span>		1c6248bd-ad72-4277-a432-3993b5e448a2
<input type="checkbox"/>	<a href="#">agency_to_synapse_pipeline</a>	9/10/2023, 6:06:22 PM	3m 17s	Manual trigger	<span>✓ Succeeded</span>		c0e9577b-3b0e-4834-8505-0fc698ef9f85
<input type="checkbox"/>	<a href="#">current_year_from_sql_db_to_...</a>	9/10/2023, 5:59:19 PM	3m 25s	Manual trigger	<span>✓ Succeeded</span>		062cb8f3-d52c-4534-99a5-48268c57d1f8
<input type="checkbox"/>	<a href="#">all_data_to_synapse_pipeline</a>	9/10/2023, 5:57:49 PM	3m 38s	Manual trigger	<span>✓ Succeeded</span>		a52554af-68e2-4210-8720-1cda9c34fab6
<input type="checkbox"/>	<a href="#">load2021tosqldb</a>	9/10/2023, 5:47:50 PM	3m 30s	Manual trigger	<span>✓ Succeeded</span>		415c9ad7-0b0d-49af-9c2b-8a684f0b43ab