

Git location:

<https://github.com/singhsa3/azuredf/tree/dev>

The learner will be able to create a Linked Service to configure a connection to Azure Data Lake Gen2 containing employee and payroll data

A Linked Service object is present in the data pipeline repository of type "AzureBlobFS" that configures a connection to Azure Data Lake Gen2 containing payroll data.

The learner will be able to create a Linked Service in Azure Data Factory from Azure SQL Database to configure a connection to Azure SQL Database containing employee and payroll data

A Linked Service object is present in the data pipeline repository of type "AzureSQLDatabase" that configures a connection to Azure SQL Database containing employee and payroll data.

The learner will be able to create a Linked Service to configure a connection to Azure Synapse Analytics	A Linked Service object is present in the data pipeline repository of type "AzureSqlDW" that configures a connection to Azure Synapse Analytics to land the data for the NYC Synapse Data Warehouse analytics platform
--	--

**Microsoft Azure | Data Factory | sanjdf**

Search

dev branch | Validate all | Save all | Publish

**Linked services**

Linked service defines the connection information to a data store or compute. [Learn more](#)

+ New

Filter by name Annotations: Any

Showing 1 - 3 of 3 items

Name	Type	Related
AzureDataLakeStorage1	Azure Data Lake Storage Gen2	10
AzureSqlDatabas...	Azure SQL Database	1
AzureSynapseAnalytics1	Azure Synapse Analytics	5

**Related**

AzureSynapseAnalytics1

This linked service is being used by the following:

**Datasets**

- NYC\_Payroll\_AGENCY\_MD
- NYC\_Payroll\_Data
- NYC\_Payroll\_EMP\_MD
- NYC\_Payroll\_Summary
- NYC\_Payroll\_TITLE\_MD

The learner will be able to create datasets to provide views of employee and payroll data in Azure Data Lake Gen2.	Multiple dataset objects are present in the data pipeline repository of type "AzureBlobFSLocation" with schemas from "AgencyMaster.csv", "TitleMaster.csv", "nycpayroll_2021.csv", and "EmpMaster.csv" to provide datasets for data views from Azure Data Lake Gen2.
--	--

**Microsoft Azure | Synapse Analytics | project74**

Search

Synapse live | Validate all | Publish all

**Data**

Workspace Linked

Filter resources by name

- Azure Data Lake Storage Gen2
  - project74 (Primary - sanjacc1)
    - adlsnycpayrollsk (Primary)
      - sanjefs1
        - (Attached Containers)

**adlsnycpayrollsk**

New SQL script | New notebook | New data flow | More

adlsnycpayrollsk > dirpayrollfiles

Name	Last Modified	Content Type	Size
AgencyMaster.csv	12/16/2022, 10:02:55 PM		4.9 KB
EmpMaster.csv	12/16/2022, 10:01:16 PM		23.0 KB
nycpayroll_2021.csv	12/16/2022, 10:02:55 PM		16.4 KB
TitleMaster.csv	12/16/2022, 10:02:55 PM		49.0 KB

Microsoft Azure | Data Factory | sanjdf

Search

sknwar3@gatech.edu  
GEORGIA INSTITUTE OF TECHNOLOGY

dev branch | Validate all | Save all | Publish

Preview experience | Off

Factory Resources

Filter resources by name

Pipelines7

pipeline\_Agency\_master\_to\_synapse

pipeline\_aggregate

pipeline\_employee\_master\_to\_synapse

pipeline\_Load\_2021\_payroll\_into\_SQLDB

pipeline\_Load\_All\_Data\_Into\_Synapse

pipeline\_Load\_current\_year\_from\_SQLDB\_to\_synapse

pipeline\_Title\_master\_to\_synapse

Datasets11

AgencyMaster...

EmpMaster

NYC\_Payroll\_AGENCY\_MD

NYC\_Payroll\_Data

NYC\_Payroll\_EMP\_MD

NYC\_Payroll\_Summary

NYC\_Payroll\_TITLE\_MD

nycpayroll\_2021

payroll\_history

TitleMaste

transact\_current

Data flows6

Power Query0

pipeline\_Load\_2021... | AgencyMaster

Saved

DelimitedText  
AgencyMaster

CSV

Connection | Schema | Parameters

Linked service \*  
AzureDataLakeStorage1

Test connection | Edit | + New | Learn more

File path \*  
adlsnycpayrollsk / dirpayrollfiles / AgencyMaster

Compression type  
None

Column delimiter  
Comma (,) | Edit

Row delimiter  
Default (\r,\n, or \r\n) | Edit

Encoding  
Default(UTF-8)

Escape character  
Backslash (\) | Edit

Quote character  
Double quote (")

Microsoft Azure | Data Factory | sanjdf

Search

sknwar3@gatech.edu  
GEORGIA INSTITUTE OF TECHNOLOGY

dev branch | Validate all | Save all | Publish

Preview experience | Off

Factory Resources

Filter resources by name

Pipelines7

pipeline\_Agency\_master\_to\_synapse

pipeline\_aggregate

pipeline\_employee\_master\_to\_synapse

pipeline\_Load\_2021\_payroll\_into\_SQLDB

pipeline\_Load\_All\_Data\_Into\_Synapse

pipeline\_Load\_current\_year\_from\_SQLDB\_to\_synapse

pipeline\_Title\_master\_to\_synapse

Datasets11

AgencyMaster

EmpMaster

NYC\_Payroll\_AGENCY\_MD

NYC\_Payroll\_Data

NYC\_Payroll\_EMP\_MD

NYC\_Payroll\_Summary

NYC\_Payroll\_TITLE\_MD

nycpayroll\_2021

payroll\_history

TitleMaste...

transact\_current

Data flows6

line\_Load\_2021...

AgencyMaster

EmpMaster

TitleMaste

Saved

DelimitedText

TitleMaste

Connection | Schema | Parameters

Linked service \*

AzureDataLakeStorage1

Test connection | Edit | New | Learn more

File path \*

adlsnycpayrollsk / dirpayrollfiles / TitleMaster.csv

Compression type

None

Column delimiter

Comma (,)

Edit

Row delimiter

Default (\r,\n, or \n)

Edit

Encoding

Default(UTF-8)

Escape character

Backslash (\)

Edit

Microsoft Azure

Data Factory | sanjdf

Search

skewar3@gatech.edu  
GEORGIA INSTITUTE OF TECHNOLOGY

dev branch

Validate all

Save all

Publish

Preview experience

Off

Factory Resources

Filter resources by name

Pipelines7

pipeline\_Agency\_master\_to\_synapse

pipeline\_aggregate

pipeline\_employee\_master\_to\_synapse

pipeline\_Load\_2021\_payroll\_into\_SQLDB

pipeline\_Load\_All\_Data\_Into\_Synapse

pipeline\_Load\_current\_year\_from\_SQLDB\_to\_synapse

pipeline\_Title\_master\_to\_synapse

Datasets11

AgencyMaster

EmpMaster

NYC\_Payroll\_AGENCY\_MD

NYC\_Payroll\_Data

NYC\_Payroll\_EMP\_MD

NYC\_Payroll\_Summary

NYC\_Payroll\_TITLE\_MD

nycpayroll\_2021

payroll\_history

TitleMaste

transact\_current

Data flows6

cyMaster

EmpMaster

TitleMaste

nycpayroll\_2021

Saved

DelimitedText

nycpayroll\_2021

Connection

Schema

Parameters

Linked service \*

AzureDataLakeStorage1

Test connection Edit + New Learn more

File path \*

adlsnycpayrollsk

 / 

dirpayrollfiles

 / 

nycpayroll\_2021.csv

Compression type

None

Column delimiter 

Comma (,)

Edit

Row delimiter 

Default (\r,\n, or \r\n)

Edit

Encoding 

Default(UTF-8)

Escape character 

Backslash (\)

Edit

Microsoft Azure

Data Factory > sanjdf

Search

skewar3@gatech.edu

GEORGIA INSTITUTE OF TECHNOLOGY

dev branch

Validate all

Save all

Publish

Preview experience

Off

Factory Resources

Filter resources by name

Pipelines

7

pipeline\_Agency\_master\_to\_synapse

pipeline\_aggregate

pipeline\_employee\_master\_to\_synapse

pipeline\_Load\_2021\_payroll\_into\_SQLDB

pipeline\_Load\_All\_Data\_Into\_Synapse

pipeline\_Load\_current\_year\_from\_SQLDB\_to\_synapse

pipeline\_Title\_master\_to\_synapse

Datasets

11

AgencyMaster

EmpMaster

NYC\_Payroll\_AGENCY\_MD

NYC\_Payroll\_Data

NYC\_Payroll\_EMP\_MD

NYC\_Payroll\_Summary

NYC\_Payroll\_TITLE\_MD

nycpayroll\_2021

payroll\_history

TitleMaste

transact\_current

Data flows

6

Power Query

0

Templates

0

pipeline\_Load\_2021...

AgencyMaster

EmpMaster

Saved

DelimitedText

EmpMaster

Connection

Schema

Parameters

Linked service \*

AzureDataLakeStorage1

Test connection

Edit

New

Learn more

File path \*

adlsnycpayrollsk

dirpayrollfiles

EmpMaster.csv

Compression type

None

Column delimiter

Comma (,)

Edit

Row delimiter

Default (\r\n, or \n)

Edit

Encoding

Default(UTF-8)

Escape character

Backslash (\)

Edit

Quote character

Double quote (")

The learner will be able to create a dataset to provide a view of employee and payroll data in the Azure SQL DB table.

A dataset object is present in the data pipeline repository of type "AzureSqlTable" with schemas from the NYC Payroll Data SQL DB table to provide a dataset for a data view.

Microsoft Azure | Data Factory | sanjdf

Search

dev branch | Validate all | Save all | Publish

Preview experience: Off

### Factory Resources

Filter resources by name

- Pipelines** (7)
  - pipeline\_Agency\_master\_to\_synapse
  - pipeline\_aggregate
  - pipeline\_employee\_master\_to\_synapse
  - pipeline\_Load\_2021\_payroll\_into\_SQLDB
  - pipeline\_Load\_All\_Data\_Into\_Synapse
  - pipeline\_Load\_current\_year\_from\_SQLDB\_to\_synapse
  - pipeline\_Title\_master\_to\_synapse
- Datasets** (11)
  - AgencyMaster
  - EmpMaster
  - NYC\_Payroll\_AGENCY\_MD
  - NYC\_Payroll\_Data
  - NYC\_Payroll\_EMP\_MD
  - NYC\_Payroll\_Summary
  - NYC\_Payroll\_TITL\_MD
  - nycpayroll\_2021
  - payroll\_history
  - TitleMaste
  - transact\_current**
- Data flows** (6)

transact\_current

Azure SQL Database

transact\_current

Connection | Schema | Parameters

Linked service \* | AzureSqlDatabase1

Test connection | Edit | New | Learn more

Table | dbo.NYC\_Payroll\_Data | Refresh

Edit

Preview data

The learner will be able to create datasets to provide views of data to the Synapse Data Warehouse platform

Multiple datasets are present in the project repository dataset directory of type "AzureSqlDWTable" to land the data from Azure Data Lake Gen2 into the Synapse Data Warehouse platform.

Microsoft Azure | Data Factory | sanjdf

Search

dev branch | Validate all | Save all | Publish

Preview experience: Off

### Factory Resources

Filter resources by name

- Pipelines** (7)
  - pipeline\_Agency\_master\_to\_synapse
  - pipeline\_aggregate
  - pipeline\_employee\_master\_to\_synapse
  - pipeline\_Load\_2021\_payroll\_into\_SQLDB
  - pipeline\_Load\_All\_Data\_Into\_Synapse
  - pipeline\_Load\_current\_year\_from\_SQLDB\_to\_synapse
  - pipeline\_Title\_master\_to\_synapse
- Datasets** (11)
  - AgencyMaster
  - EmpMaster
  - NYC\_Payroll\_AGENCY\_MD**
  - NYC\_Payroll\_Data
  - NYC\_Payroll\_EMP\_MD

NYC\_Payroll\_AGENCY\_MD

Azure Synapse Analytics

NYC\_Payroll\_AGENCY\_MD

Connection | Schema | Parameters

Linked service \* | AzureSynapseAnalytics1

Test connection | Edit | New | Learn more

Table | dbo.NYC\_Payroll\_AGENCY\_MD | Refresh

Edit

Preview data



Microsoft Azure | Data Factory | sanjdf

dev branch | Validate all | Save all | Publish | Preview experience: Off

### Factory Resources

Filter resources by name

- Pipelines** (7)
  - pipeline\_Agency\_master\_to\_synapse
  - pipeline\_aggregate
  - pipeline\_employee\_master\_to\_synapse
  - pipeline\_Load\_2021\_payroll\_into\_SQLDB
  - pipeline\_Load\_All\_Data\_Into\_Synapse
  - pipeline\_Load\_current\_year\_from\_SQLDB\_to\_synapse
  - pipeline\_Title\_master\_to\_synapse
- Datasets** (11)
  - AgencyMaster
  - EmpMaster
  - NYC\_Payroll\_AGENCY\_MD
  - NYC\_Payroll\_Data
  - NYC\_Payroll\_EMP\_MD**
  - NYC\_Payroll\_Summary

Connection | Schema | Parameters

Linked service \*: AzureSynapseAnalytics1

Test connection | Edit | + New | Learn more

Table: Loading... | Refresh

☐ Edit

Preview data

Microsoft Azure | Data Factory | sanjdf

dev branch | Validate all | Save all | Publish | Preview experience: Off

### Factory Resources

Filter resources by name

- Pipelines** (7)
  - pipeline\_Agency\_master\_to\_synapse
  - pipeline\_aggregate
  - pipeline\_employee\_master\_to\_synapse
  - pipeline\_Load\_2021\_payroll\_into\_SQLDB
  - pipeline\_Load\_All\_Data\_Into\_Synapse
  - pipeline\_Load\_current\_year\_from\_SQLDB\_to\_synapse
  - pipeline\_Title\_master\_to\_synapse
- Datasets** (11)
  - AgencyMaster
  - EmpMaster
  - NYC\_Payroll\_AGENCY\_MD
  - NYC\_Payroll\_Data
  - NYC\_Payroll\_EMP\_MD
  - NYC\_Payroll\_Summary
  - NYC\_Payroll\_TITLE\_MD**

Connection | Schema | Parameters

Linked service \*: AzureSynapseAnalytics1

Test connection | Edit | + New | Learn more

Table: dbo.NYC\_Payroll\_TITLE\_MD | Refresh

☐ Edit

Preview data



Microsoft Azure | Data Factory | sanjdf

dev branch | Validate all | Save all | Publish

Preview experience | Off

### Factory Resources

Filter resources by name

- Pipelines** (7)
  - pipeline\_Agency\_master\_to\_synapse
  - pipeline\_aggregate
  - pipeline\_employee\_master\_to\_synapse
  - pipeline\_Load\_2021\_payroll\_into\_SQLDB
  - pipeline\_Load\_All\_Data\_Into\_Synapse
  - pipeline\_Load\_current\_year\_from\_SQLDB\_to\_synapse
  - pipeline\_Title\_master\_to\_synapse
- Datasets** (11)
  - AgencyMaster
  - EmpMaster
  - NYC\_Payroll\_AGENCY\_MD
  - NYC\_Payroll\_Data**
  - NYC\_Payroll\_EMP\_MD
  - NYC\_Payroll\_Summary
  - NYC\_Payroll\_TITLE\_MD
  - nycpayroll\_2021

**NYC\_Payroll\_Data**

Connection | Schema | Parameters

Linked service \* | AzureSynapseAnalytics1

Test connection | Edit | New | Learn more

Table | dbo.NYC\_Payroll\_Data | Refresh

Preview data

Microsoft Azure | Data Factory | sanjdf

dev branch | Validate all | Save all | Publish

Preview experience | Off

### Factory Resources

Filter resources by name

- Pipelines** (7)
  - pipeline\_Agency\_master\_to\_synapse
  - pipeline\_aggregate
  - pipeline\_employee\_master\_to\_synapse
  - pipeline\_Load\_2021\_payroll\_into\_SQLDB
  - pipeline\_Load\_All\_Data\_Into\_Synapse
  - pipeline\_Load\_current\_year\_from\_SQLDB\_to\_synapse
  - pipeline\_Title\_master\_to\_synapse
- Datasets** (11)
  - AgencyMaster
  - EmpMaster
  - NYC\_Payroll\_AGENCY\_MD
  - NYC\_Payroll\_Data
  - NYC\_Payroll\_EMP\_MD
  - NYC\_Payroll\_Summary**
  - NYC\_Payroll\_TITLE\_MD

**NYC\_Payroll\_Summary**

Connection | Schema | Parameters

Linked service \* | AzureSynapseAnalytics1

Test connection | Edit | New | Learn more

Table | dbo.NYC\_Payroll\_Summary | Refresh

Preview data

The learner will be able to create data flows to aggregate payroll data from Azure SQL DB and NYC Payroll history files to the Synapse Data Warehouse.

A Dataflow object is present in the data pipeline repository of type "MappingDataFlow" with a union to create a derived aggregated column with the total amount paid to an employee ( $\text{TotalPaid} = \text{RegularGrossPaid} + \text{TotalOTPaid} + \text{TotalOtherPay}$ ). The data sources for this aggregate column should be the data from Azure SQL DB table and the historical NYC payroll data.

Microsoft Azure | Data Factory | sanjdf

dev branch | Validate all | Save all | Publish

Factory Resources

- Pipelines (7)
  - pipeline\_Agency\_master\_to\_synapse
  - pipeline\_aggregate
  - pipeline\_employee\_master\_to\_synap...
  - pipeline\_Load\_2021\_payroll\_into\_SQ...
  - pipeline\_Load\_All\_Data\_into\_Synapse
  - pipeline\_Load\_current\_year\_from\_SQ...
  - pipeline\_title\_master\_to\_synapse
- Datasets (11)
- Data flows (6)
  - dataflow\_agency\_master\_to\_synapse
  - dataflow\_aggregate
  - dataflow\_employee\_master\_to\_syna...
  - dataflow\_load\_curent\_year\_data\_fro...
  - dataflow\_payroll\_Load\_2021\_payroll...
  - dataflow\_title\_master\_to\_synapse
- Power Query (0)
- Templates (0)

NYC\_Payroll\_Summary | dataflow\_aggregate

current | history | union1 | filter1 | TotalPaid | aggregate1 | sink1

Derived column's settings

Output stream name: TotalPaid

Description: Creating/updating the columns 'FiscalYear, PayrollNumber, AgencyID, AgencyName, EmployeeID, LastName,

Incoming stream: filter1

Columns:

Column	Expression
TotalPaid	RegularGrossPaid + TotalOTPaid + TotalOtherPay

CRITERIA	MEETS SPECIFICATIONS
The student will be able to create a data pipeline containing Dataflow activities.	Multiple pipeline objects are present in the data pipeline repository with activities of type "ExecuteDataFlow" in the pipeline directory which contain Dataflow objects.
The learner should be able to trigger a pipeline and execute the Dataflows in it.	A screenshot is present showing a successful pipeline execution in Azure Data Factory

Microsoft Azure | Data Factory | sanjdf

>> / dev branch Validate all Save all

### Factory Resources

Filter resources by name

- Pipelines** 7
  - pipeline\_Agency\_master\_to\_synapse
  - pipeline\_aggregate
  - pipeline\_employee\_master\_to\_synapse
  - pipeline\_Load\_2021\_payroll\_into\_SQLDB
  - pipeline\_Load\_All\_Data\_Into\_Synapse
  - pipeline\_Load\_current\_year\_from\_SQLDB\_to\_synapse
  - pipeline\_Title\_master\_to\_synapse
- Datasets** 11
- Data flows** 6
  - dataflow\_agency\_master\_to\_synapse
  - dataflow\_aggregate
  - dataflow\_employee\_master\_to\_synapse
  - dataflow\_load\_curent\_year\_data\_from\_sqldb\_to\_synapse
  - dataflow\_payroll\_load\_2021\_payroll\_into\_SQLDB
  - dataflow\_title\_master\_to\_synapse
- Power Query** 0
- Templates** 0

Microsoft Azure | Data Factory | sanjdf

Pipeline runs

Triggered Debug Return Cancel options Refresh Edit columns Last Queue

Filter by run ID or name Local time: Last 24 hours Pipeline name: All Status: All Runs: Latest runs Triggered by: All Add filter Copy filters Export to CSV

Showing 1 - 15 items

Pipeline name	Run start	Run end	Duration	Triggered by	Status	Error	Run	Parameters	Annotations	Run ID
pipeline_title_master_to_synapse	12/17/2022, 3:31:04 PM	12/17/2022, 3:37:03 PM	00:05:59	Manual trigger	Succeeded		Original			5e887d21-24eb-47c0-1
pipeline_Load_current_year_from_SQLDB_to_synapse	12/17/2022, 3:30:57 PM	12/17/2022, 3:36:42 PM	00:05:44	Manual trigger	Succeeded		Original			a6700e3b-c509-44e4-t
pipeline_Load_All_Data_Into_Synapse	12/17/2022, 3:30:53 PM	12/17/2022, 3:41:20 PM	00:10:26	Manual trigger	Succeeded		Original			80f87846-9197-4c8a-9
pipeline_Load_2021_payroll_into_SQLDB	12/17/2022, 3:30:45 PM	12/17/2022, 3:35:28 PM	00:04:43	Manual trigger	Succeeded		Original			38d6a83c-c151-4dd3-t
pipeline_employee_master_to_synapse	12/17/2022, 3:30:39 PM	12/17/2022, 3:35:32 PM	00:04:52	Manual trigger	Succeeded		Original			71a3f582-3f6b-4dd5-9
pipeline_aggregate	12/17/2022, 3:30:33 PM	12/17/2022, 3:35:49 PM	00:05:15	Manual trigger	Succeeded		Original			d4d424c3-2c92-4587-5
pipeline_Agency_master_to_synapse	12/17/2022, 3:30:27 PM	12/17/2022, 3:35:17 PM	00:04:50	Manual trigger	Succeeded		Original			d4e102d6-c06a-467b-5

Last refreshed 0 minutes ago

Microsoft Azure | Data Factory | sanjdf

Search

skewar3@gatech.edu  
GEORGIA INSTITUTE OF TECHNOLOGY

» <<

dashboards

Runs

Pipeline runs

Trigger runs

Runtimes & sessions

Integration runtimes

Data flow debug

Notifications

Alerts & metrics

All pipeline runs > pipeline\_Load\_All\_Data\_Into\_Synapse - Activity runs

pipeline\_Load\_All\_Data\_Into\_Synapse

List Gantt

Rerun Rerun from activity Rerun from failed activity Refresh Update pipeline

dataflow\_employee\_master\_to\_synapse

dataflow\_title\_master\_to\_synapse

dataflow\_agency\_master\_to\_synapse

dataflow\_payroll\_load\_2021\_payro...

dataflow\_load\_current\_year\_data...

Activity runs

Pipeline run ID fc0856b5-6842-439c-bcc6-e9a2d9ce7071

All status

Export to CSV

Showing 1 - 5 items

Activity name	Activity type	Run start	Duration	Status	Error
dataflow_load_current_year_data_from_sqldb_to_syn...	Data flow	12/17/2022, 2:03:22 PM	00:04:27	Succeeded	
dataflow_title_master_to_synapse	Data flow	12/17/2022, 1:58:29 PM	00:04:47	Succeeded	
dataflow_employee_master_to_synapse	Data flow	12/17/2022, 1:58:28 PM	00:04:53	Succeeded	
dataflow_agency_master_to_synapse	Data flow	12/17/2022, 1:58:28 PM	00:04:49	Succeeded	
dataflow_payroll_load_2021_payro...	Data flow	12/17/2022, 1:58:28 PM	00:04:49	Succeeded	

Microsoft Azure | Data Factory | sanjdf

Search

skewar3@gatech.edu  
GEORGIA INSTITUTE OF TECHNOLOGY

» <<

dashboards

Runs

Pipeline runs

Trigger runs

Runtimes & sessions

Integration runtimes

Data flow debug

Notifications

Alerts & metrics

All pipeline runs > pipeline\_aggregate - Activity runs

pipeline\_aggregate

List Gantt

Rerun Rerun from activity dataflow\_aggregate Rerun from failed activity Refresh Update pipeline

dataflow\_aggregate

Activity runs

Pipeline run ID 4030d04e-072c-4147-ba3c-fa72c1cfd2e1

All status

Export to CSV

Showing 1 - 1 items

Activity name	Activity type	Run start	Duration	Status	Error	Log
dataflow_aggregate	Data flow	12/17/2022, 2:49:20 PM	00:05:13	Succeeded		