

STEP 1: Setup Data and Resources in Azure

Microsoft Azure | Search resources, services, and docs (G+)

Home > ODL-DataEng-240197 | Resource group

Search

Essentials

Resources Recommendations

Filter for any field... Type equals all Location equals all Add filter

Showing 1 to 5 of 5 records. Show hidden types No grouping List view

Name	Type	Location
db_nycpayroll (project4sqladb/db_nycpayroll)	SQL database	East US
pipelineproject4	Synapse workspace	East US
project4sqladb	SQL server	East US
udacitydatafactory	Data factory (V2)	East US
udacityproject4	Storage account	East US

Activate Windows
Go to Settings to activate Windows.
Give feedback

Microsoft Azure | Search resources, services, and docs (G+)

Home > udacityproject4_1691923630362 | Overview > udacityproject4 | Containers

adlsnycpayroll-oreva-e | Container

Search

Authentication method: Access key (Switch to Azure AD User Account)
Location: adlsnycpayroll-oreva-e

Search blobs by prefix (case-sensitive) Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
dirhistoryfiles						
dirpayrollfiles						
dirstaging						

Microsoft Azure Search resources, services, and docs (G+/)

Home > udacityproject4 | Containers >

adlsnycpayroll-oreva-e

Container

Search

Upload Add Directory Refresh Rename Delete Change tier Acquire lease Break lease Give feedback

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Shared access tokens

Manage ACL

Access policy

Properties

Metadata

Authentication method: Access key (Switch to Azure AD User Account)

Location: adlsnycpayroll-oreva-e / dirhistoryfiles

Search blobs by prefix (case-sensitive) Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[-]						...
nycpayroll_2020.csv	8/13/2023, 11:56:44 ...	Hot (inferred)		Block blob	17.07 KiB	Available

Microsoft Azure Synapse Analytics sapipelineproject

Synapse live Validate all Publish all

adlsnycpayroll-oreva-e SQL script 1

Run Undo Publish Query plan Connect to pipelineproject Use database pipelineproject

```
1 CREATE TABLE [dbo].[NYC_Payroll_EMP_MD](
2     [EmployeeID] [varchar](10) NULL,
3     [LastName] [varchar](20) NULL,
4     [FirstName] [varchar](20) NULL
5 )
6 GO
7
8 CREATE TABLE [dbo].[NYC_Payroll_TITLE_MD](
9     [TitleCode] [varchar](10) NULL,
10    [TitleDescription] [varchar](100) NULL
11 )
12 GO
13
14 CREATE TABLE [dbo].[NYC_Payroll_AGENCY_MD](
15     [AgencyID] [varchar](10) NULL,
16     [AgencyName] [varchar](50) NULL
17 )
18 GO
19
20
```

Results Messages

00:00:05 Query executed successfully.

Properties

General Related (0)

Name * SQL script 1

Description

Type .sql script

Size 0 bytes

Results settings per query

First 5000 rows (default)

All rows

Activate Windows Go to Settings to activate Windows.

Microsoft Azure | Search resources, services, and docs (G+/)

Home > project4sqladb > db_nycpayroll (project4sqladb/db_nycpayroll)

db_nycpayroll (project4sqladb/db_nycpayroll) | Query editor (preview) ☆

SQL database

Search

Login + New Query ↑ Open query Feedback Getting started

Overview
Activity log
Tags
Diagnose and solve problems
Query editor (preview)

Settings

Compute + storage
Connection strings
Properties
Locks
Data management
Replicas
Sync to other databases
Integrations

db_nycpayroll (oreva)

Showing limited object explorer here. For full capability please click here to open Azure Data Studio.

Tables

dbo.NYC_Payroll_Data

FiscalYear (int, null)
PayrollNumber (int, null)
AgencyID (varchar, null)
AgencyName (varchar, null)
EmployeeID (varchar, null)
LastName (varchar, null)
FirstName (varchar, null)
AgencyStartDate (date, null)
WorkLocationBorough (varchar, null)

Query 1

Run Cancel query Save query Export data as Show only Editor

```
1 CREATE TABLE [dbo].[NYC_Payroll_Data](
2     [FiscalYear] [int] NULL,
3     [PayrollNumber] [int] NULL,
4     [AgencyID] [varchar](10) NULL,
5     [AgencyName] [varchar](50) NULL,
6     [EmployeeID] [varchar](10) NULL,
7     [LastName] [varchar](20) NULL,
8     [FirstName] [varchar](20) NULL
```

Results Messages

Query succeeded: Affected rows: 0

Query succeeded | 0s

Activate Windows
Go to Settings to activate Windows.

Step 2: Create Linked Services

Microsoft Azure | Data Factory | udacitydatafactory | Search factory and documentation

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Preview experience Off

General
Factory settings
Connections
Linked services
Integration runtimes
Microsoft Purview
Source control
Git configuration
ARM template
Author
Triggers
Global parameters
Data flow libraries
Security
Credentials

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	Annotations
AzureDataLakeStorage1	Azure Data Lake Storage Gen2	0	
AzureSqlDatabase1	Azure SQL Database	0	
AzureSynapseAnalytics1	Azure Synapse Analytics	0	

Activate Windows
Go to Settings to activate Windows.

Step 3: Create Datasets in Azure Data Factory

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' pane is expanded to 'Datasets', showing a list of datasets including 'AgencyMaster_dls', 'EmpMaster_dls', 'NYCpayroll2021_dls', 'TitleMaster_dls', 'Azure sqldb', 'NYC_Payroll_data_sqldb', 'Azure synapse analytics', 'NYC_Payroll_EmpMD_syn', 'NYC_Payroll_TitleMD_syn', 'NYC_PayrollAgency_syn' (selected), and 'NYC_PayrollData_syn'. The main workspace displays the 'NYC_PayrollAgency_syn' dataset configuration. The 'Connection' tab is active, showing 'Linked service' as 'AzureSynapseAnalytics1' and 'Table' as 'dbo.NYC_Payroll_AGENCY_MD'. The 'Properties' pane on the right shows the 'Name' as 'NYC_PayrollAgency_syn' and a 'Description' field. The 'Annotations' section is also visible.

Step 4: Create Data Flows

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' pane is expanded to 'Data flows', showing a list of data flows including 'dataflow2021SQLdb to synapse' (selected), 'dataflow payroll aggregate data', 'dataflowAgencyMaster to synapse', 'dataflowEmpMaster to synapse', 'dataflowPayroll2021tosqldb', and 'dataflowTitleMD to synapse'. The main workspace displays the 'dataflow2021SQLdb to synapse' data flow configuration. The 'Parameters' tab is active, showing a 'load2021SQLdbtosyn...' activity and a 'sink1' activity. The 'Add Source' button is visible. The 'Preview experience' toggle is set to 'Off'.

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

Pipeline runs

Triggered Debug Rerun Cancel options Refresh Edit columns List Gantt

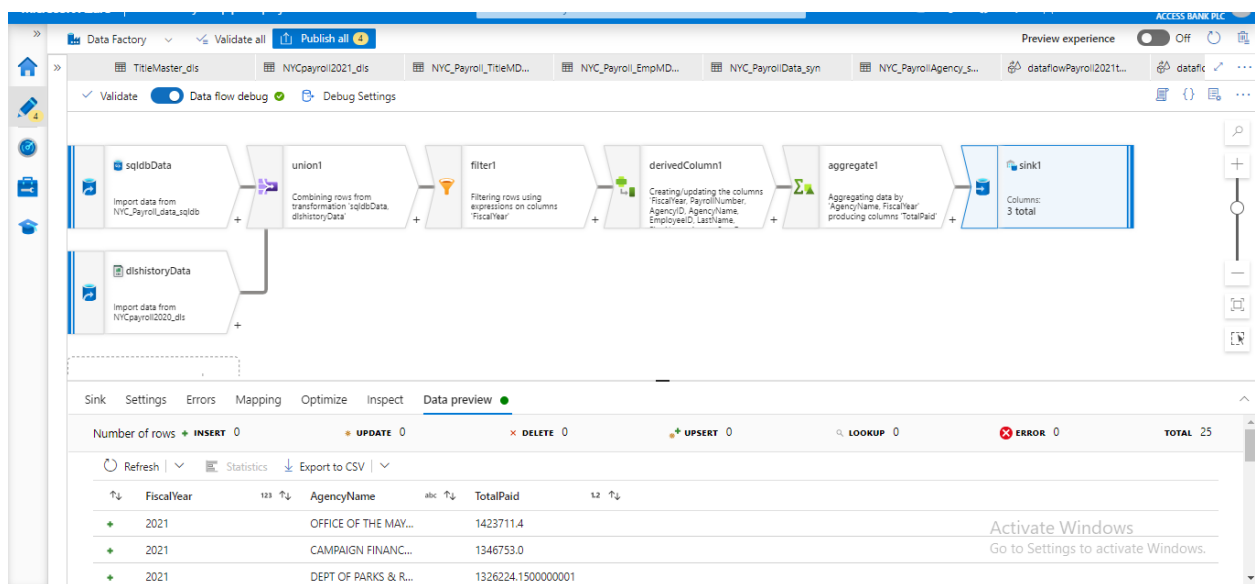
Filter by run ID or name West Central Africa (... : Last 24 hours Pipeline name : All Status : All Runs : Latest runs Copy filters Export to CSV

Triggered by : All Add filter

Showing 1 - 6 items Last refreshed 0 minutes ago

Pipeline name	Run start	Run end	Duration	Triggered by	Status	Run	Parameters
pipeline1	8/15/2023, 11:45:43 PM	8/15/2023, 11:49:44 PM	4m 1s	Manual trigger	Succeeded	Original	
pipeline load 2021data into S...	8/15/2023, 8:59:47 PM	8/15/2023, 9:05:50 PM	6m 4s	Manual trigger	Succeeded	Original	
pipeline load 2021data into S...	8/15/2023, 8:59:34 PM	8/15/2023, 9:05:43 PM	6m 9s	Manual trigger	Succeeded	Original	
pipeline load 2021data into S...	8/15/2023, 8:59:19 PM	8/15/2023, 9:05:49 PM	6m 31s	Manual trigger	Succeeded	Original	
pipeline load 2021data into S...	8/15/2023, 8:59:00 PM	8/15/2023, 9:03:06 PM	4m 6s	Manual trigger	Succeeded	Original	
pipeline load 2021data into S...	8/15/2023, 8:48:11 PM	8/15/2023, 8:52:09 PM	3m 58s	Manual trigger	Succeeded	Original	

Step 5: Data Aggregation and Parameterization



Synapse live Validate all Publish all

Develop

Filter resources by name

- SQL scripts 1
 - SQL script 1

SQL script 1

```
1 CREATE TABLE [dbo].[NYC_Payroll_Summary]
2 (
3   [FiscalYear] [int] NULL,
4   [AgencyName] [varchar](50) NULL,
5   [TotalPaid] [float] NULL )
```

Run Undo Publish Query plan Connect to pipelineproject Use database pipelineproject

Properties

General Related (0)

Name * SQL script 1

Description

Type .sql script

Size 0 bytes

Results settings per query

☒ First 5000 rows (default)

☐ All rows

Activate Windows
Go to Settings to activate Windows.

Results Messages

No results to show
Your query yielded no displayable results

00:00:02 Query executed successfully.

Synapse live Validate all Publish all

Data

Workspace Linked

Filter resources by name

- SQL database 2
 - pipelineproject (SQL)
 - Tables
 - dbo.NYC_Payroll_AGENCY_...
 - dbo.NYC_Payroll_Data
 - dbo.NYC_Payroll_EMP_MD
 - dbo.NYC_Payroll_Summary
 - Columns
 - dbo.NYC_Payroll_TITLE_MD
 - External tables
 - External resources
 - Views
 - Programmability
 - Schemas

SQL script 1

```
1 SELECT TOP (100) [FiscalYear]
2 , [AgencyName]
3 , [TotalPaid]
4 FROM [dbo].[NYC_Payroll_Summary]
```

Run Undo Publish Query plan Connect to pipelineproject Use database pipelineproject

Results Messages

View Table Chart Export results

Search

FiscalYear	AgencyName	TotalPaid
2021	OFFICE OF THE COMPTROLLER	4138318.7
2021	COMMUNITY COLLEGE (LAGUAR...	1401301
2021	BOARD OF ELECTION	1361195.45
2021	POLICE DEPARTMENT	3889557.15
2021	DEPT OF ENVIRONMENT PROTE...	4294012
2021	DEPT OF SOCIAL SERVICES	2222222.22

00:00:01 Query executed successfully.

Activate Windows
Go to Settings to activate Windows.