

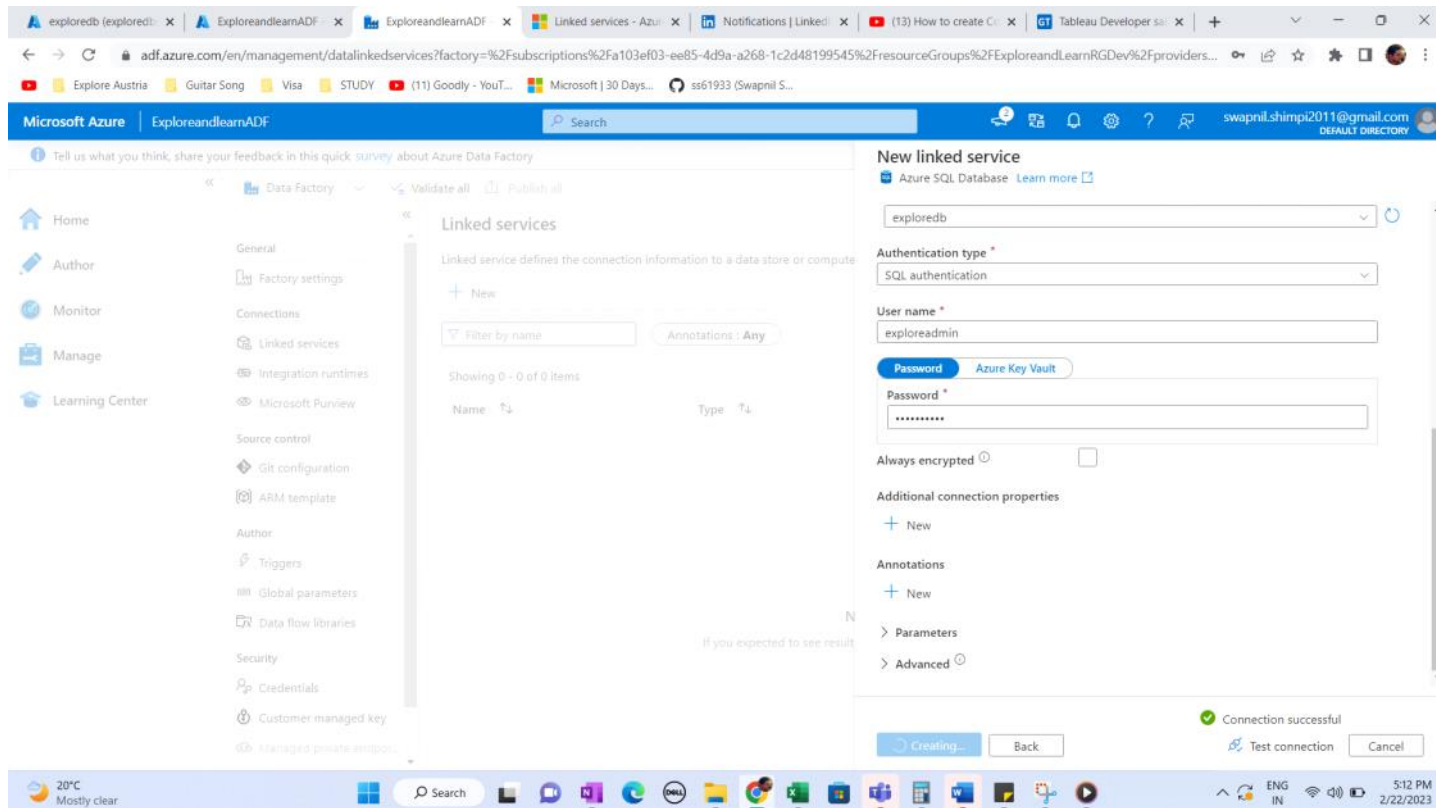
Step1 Creating Linked Services

Wednesday, February 22, 2023 4:41 PM

Linked service defines the connection information to a data store or compute.

Creating Linked service to connect to Azure Sql database

1. Click on Mange -> Linked services -> New -> Azure sql Database
2. Provide Azure subscription -> Server name -> DB name -> Credentials
3. Keep default Integration run time environment value which is **AutoResolveIntegrationRuntime**
4. Click on test connection
5. Name the link service as **link_to_exploredb**



Creating Linked service to connect to ADLS

1. Click on New-> ADLS Gen2->and follow same steps
2. Name the service as **link_to_ADLS**

22

2

swapnil.shimpi2011@gmail.com
DEFAULT DIRECTORY

re

Edit linked service

Azure Data Lake Storage Gen2 [Learn more](#)

Name *
link_to_ADLS

Description

Connect via integration runtime * ¹
AutoResolveIntegrationRuntime

Authentication type
Account key

Account selection method ¹
☐ From Azure subscription ☒ Enter manually

URL *
https://exploreandlearnsgdev.dfs.core.windows.net/

Storage account key

Azure Key Vault

Storage account key *

Test connection ¹
☒ To linked service ☐ To file path

Annotations

+ New

Save

Cancel

Test connection

Step2 Creating dataset

Wednesday, February 22, 2023 5:21 PM

Create a dataset structure:

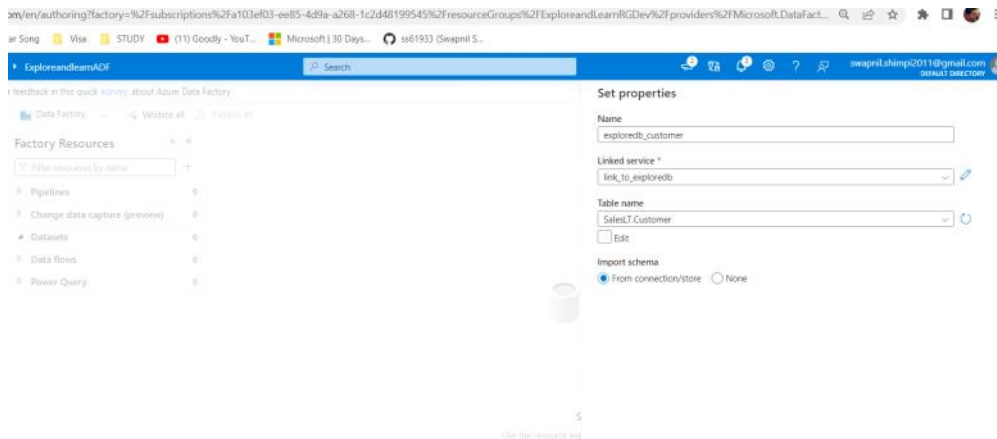
InputDatasets:

1. To read from customer tables from sql database
2. To read from Employees file from ADLS

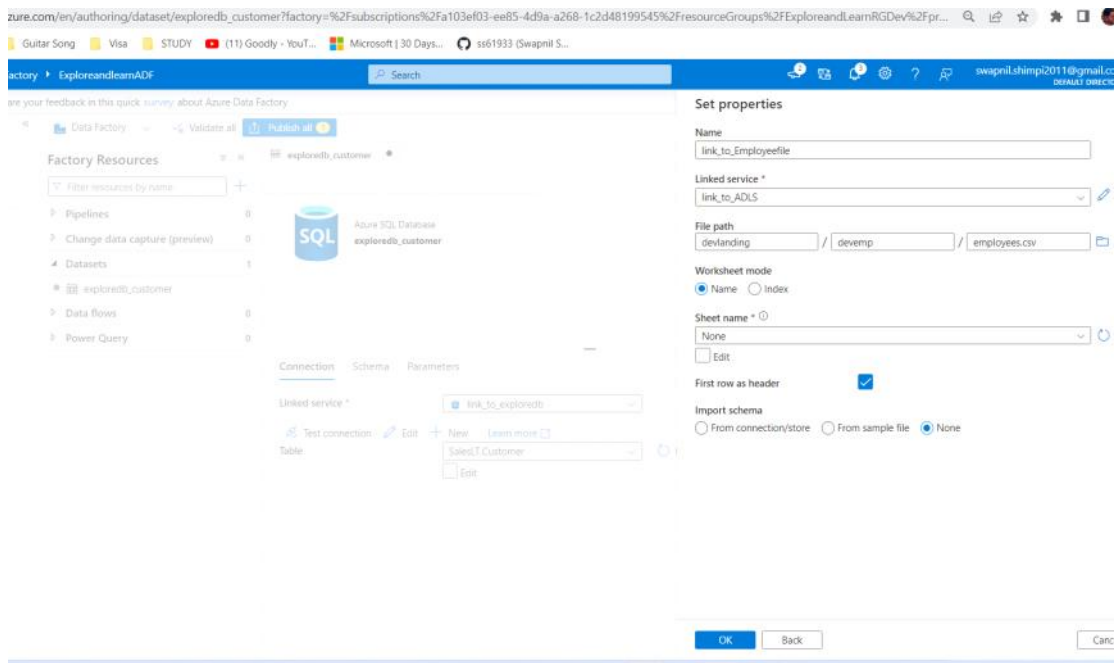
OutputDatasets:

1. To write a data into CustomerData file

1. Create dataset for connecting to customer table in sql db
2. Click on new dataset and provide below details . Name the dataset as **ip_exploredb_customer**
3. We will consume the link_to_exploredb linked service



1. Creating a dataset to connect to file system
2. Click on new data set ->ADLS gen2->Excel->provide below details. Name dataset as **ip_ADLS_Employeefile**
3. We will consume link_to_adls linked service
4. Select the underlying container and file
5. Set First Row as a Header property



Similarly we can create a dataset for storing output as a csv file with name **op_CustomerFile**

Step3 Create a copy pipeline

Wednesday, February 22, 2023 5:49 PM

USE CASE :

Read a data from Customer table and load it to CustomerData file

1. Create a pipeline with name **LoadCustomerPipeline**
2. Add a Copy activity
3. Source - Use **ip_exploredb_customer** dataset
4. Sink - Use **op_CustomerFile** dataset

The screenshot displays the Azure Data Factory (ADF) portal interface for configuring a pipeline. The pipeline is named 'LoadCustomerPipeline' and is currently in the 'Source' tab. The 'Source dataset' is set to 'ip_exploredb_customer'. The 'Use query' option is set to 'Table'. The 'Query timeout (minutes)' is set to 120. The 'Isolation level' is set to 'None'. The 'Partition option' is set to 'None'. The 'Sink dataset' is set to 'op_CustomerFile'. The 'Copy behavior' is set to 'None'. The 'Max concurrent connections' and 'Block size (MB)' are both set to their default values. The 'Sink' tab is selected, showing the 'Sink dataset' and 'Copy behavior' settings. The 'Properties' panel on the right shows the 'General' tab with the 'Name' set to 'LoadCustomerPipeline'.

General Source Sink Mapping Settings User properties

Source dataset * [Open](#) [New](#) [Preview data](#) [Learn more](#)

Use query ☒ Table ☐ Query ☐ Stored procedure

Query timeout (minutes)

Isolation level

Partition option ☒ None ☐ Physical partitions of table ☐ Dynamic range

Please preview data to validate the partition settings are correct before you trigger a run or publish the pipeline.

Validate all Publish all

Preview exp

General Source Sink Mapping Settings User properties

Sink dataset * [Open](#) [New](#) [Learn more](#)

Copy behavior

Max concurrent connections

Block size (MB)

Properties

General

Name *

Description

Annotations [+ New](#)

Once setup is completed click on debug. On success view the input , output and logs to debug

Details Refresh

Activity run id: d07cc04a-fe34-41f5-964c-fb71f3c689b0

Copy duration → **Azure SQL Database** → **Azure Data Lake Storage Gen2**

Source: Azure SQL Database
 Region: North Europe
 Data read: 278.379 KB
 Rows read: 847
 Peak connections: 1

Destination: Azure Data Lake Storage Gen2
 Region: Switzerland North
 Data written: 217.829 KB
 Files written: 1
 Rows written: 847
 Peak connections: 1

Copy duration: 00:00:07
 Throughput: 139.189 KB/s

Start time: 2/22/2023, 6:09:24 PM
Used DUs: 4
Used parallel copies: 1

Duration: 00:00:07

Details	Working duration	Total duration
Queue	00:00:00	00:00:04
Transfer	00:00:00	00:00:02

Data consistency verification: Not verified

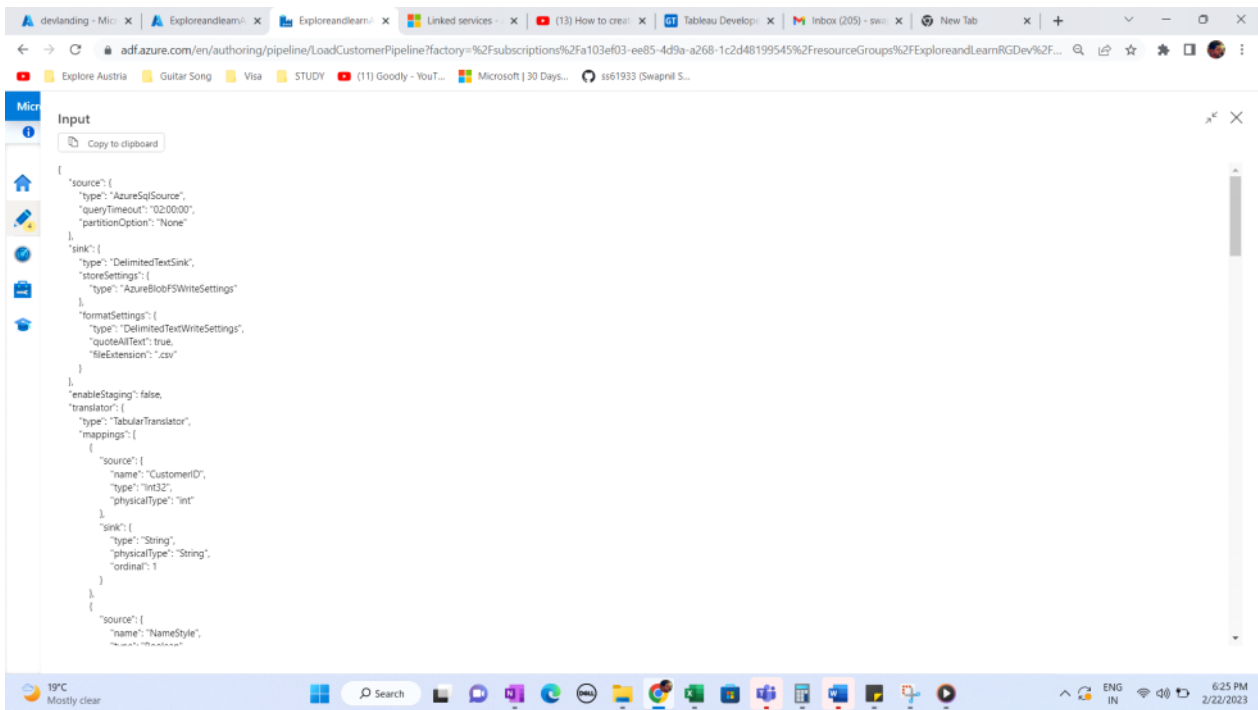
How satisfied or dissatisfied are you with the performance of this copy activity?
 ★ ★ ★ ★ ★

Check the output

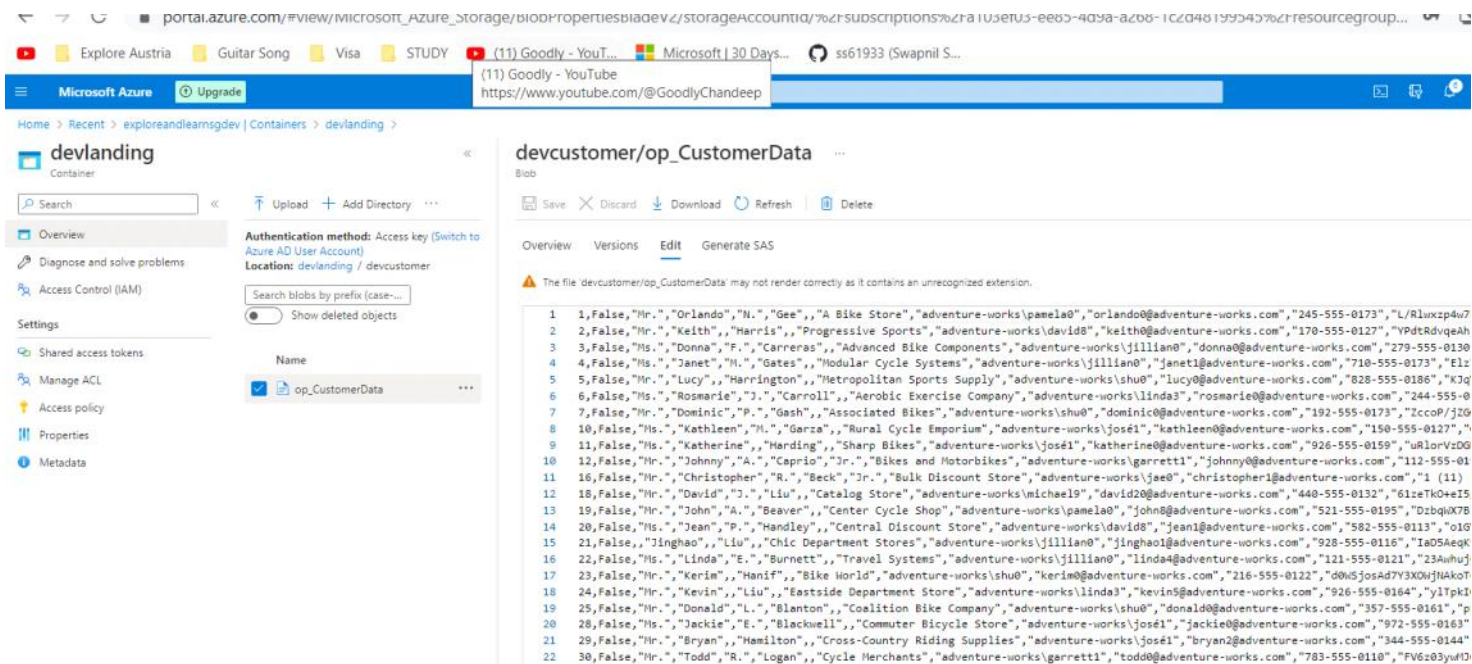
Output Copy to clipboard Learn more on output details

```
{
  "dataRead": 278379,
  "dataWritten": 217829,
  "filesWritten": 1,
  "sourcePeakConnections": 1,
  "sinkPeakConnections": 1,
  "rowsRead": 847,
  "rowsCopied": 847,
  "copyDuration": 7,
  "throughput": 139.189,
  "errors": [],
  "effectiveIntegrationRuntime": "AutoResolveIntegrationRuntime (North Europe)",
  "usedDataIntegrationUnits": 4,
  "billingReference": {
    "activityType": "DataMovement",
    "billableDuration": {
      "meterType": "AzureIR",
      "duration": 0.06666666666666667,
      "unit": "DPUHours"
    }
  },
  "usedParallelCopies": 1,
  "executionDetails": {
    "source": {
      "type": "AzureSqlDatabase",
      "region": "North Europe"
    },
    "sink": {
      "type": "AzureBlobFS",
      "region": "Switzerland North"
    },
    "status": "Succeeded",
    "start": "2/22/2023, 6:09:24 PM",
    "finish": "2/22/2023, 6:09:31 PM"
  }
}
```

Check the input



Navigate to the folder to view the file



Step4 Publish all the changes

Wednesday, February 22, 2023 5:59 PM

our feedback in this quick [survey](#) about Azure Data Factory

Data Factory

Validate all

Publishing

ip_exploredb_custom...

LoadCustomerPipel...

op_CustomerFile

Factory Resources

Filter resources by name

Pipelines1

LoadCustomerPipeline

Change data capture (preview)0

Datasets3

InputDatasets2

ip_ADLS_Employeefile

ip_exploredb_customer

OutputDatasets1

op_CustomerFile

Data flows0

Power Query0

DelimitedText

ip_ADLS_Employeefile

CSV

ConnectionSchemaParameters

Linked service *link_to_ADLS

Test connection

Edit

New

Learn more

Connection successful

File path *

Automation

Parameters

Publish all

You are about to publish all pending changes to the live environment. [Learn more](#)

Pending changes (4)

NAME	CHANGE	EXISTING
Pipelines		
LoadCustomerPipeline	(New)	-
Datasets		
ip_exploredb_customer	(New)	-
op_CustomerFile	(New)	-
ip_ADLS_Employeefile	(New)	-

Step5 Setting up GitHub

Wednesday, February 22, 2023 8:31 PM

Go to git configuration and link it to either Azure devops or link it to your GitHub .
Here I linked it to personal GitHub <https://github.com/ss61933/Azure-Data-Factory>

The screenshot shows the Azure Data Factory portal interface. The top navigation bar includes the Microsoft Azure logo, 'Data Factory', and the specific instance name 'ExploreandlearnADF'. A search bar and user profile 'ss61933 (Swapnil S...)' are also visible. Below the navigation bar, a message prompts the user to provide feedback. The main content area is titled 'Configure a repository' and includes instructions to connect the workspace to a Git repository. A table lists the configured repository details:

Repository type	GitHub
GitHub account	ss61933
Repository name	Azure-Data-Factory
Collaboration branch	main
Publish branch	adf_publish
Root folder	/Code
Last published commit	c033704e68c64e0db5a61a3bc2fdcd02d8605e13
Publish (from ADF Studio)	Enabled

The left-hand sidebar contains a navigation menu with categories like General, Connections, Source control, and Security. The 'Git configuration' option under 'Source control' is currently selected.