

## Task 1: Export MySQL Data to CSV, Parquet, and Avro

- 1.) Set up Microsoft Azure
- 2.) Create an azure data factory and created linked services using ngrok (mysql\_local\_ngrok) to create a tunnel using ngrok which enables us to send the data from our local MYSQL Workbench to Microsoft Aure
- 3.) Created an azure storage account, used it to make another linked service azure\_blob\_storage. Used a contained called output-data for storing the files.

### All resources

Default Directory

+ Create Manage view Refresh Export to CSV Open query Assign tags Delete Group by none

You are viewing a new version of Browse experience. Some features may be missing. Click here to access the old experience.

Filter for any field... Subscription equals all Resource Group equals all Type equals all Location equals all Add filter

<input type="checkbox"/>	Name ↑	Type	Resource Group	Location	Subscription
<input type="checkbox"/>	celebalinternstorage	Storage account	internship-rg	Central India	Azure subscription 1
<input type="checkbox"/>	internship-adf-celebal	Data factory (V2)	internship-rg	Central India	Azure subscription 1

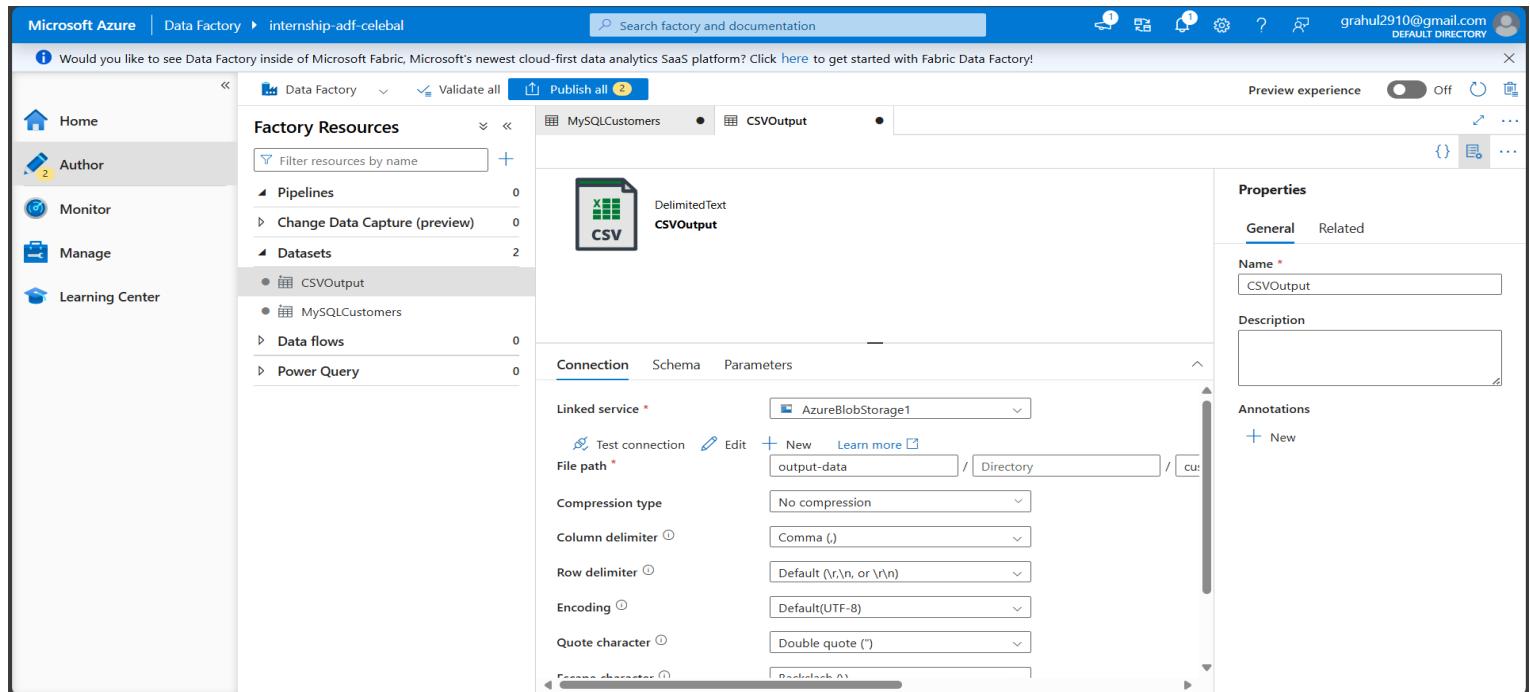
- 4.) Created a source dataset called MYSQL\_Source and used the customers table in MYSQL Workbench to feed data into it using ngrok
- 5.) Then created a pipeline named CopyToAllFormats that did the conversion of MYSQL data to CSV, Parquet and Avro respectively when the trigger command was sent.

The screenshot displays the Azure Data Factory (ADF) console interface. On the left, the 'Activities' pane shows a search bar and a list of activity categories including 'Move and transform', 'Synapse', 'Azure Data Explorer', 'Azure Function', 'Batch Service', 'Databricks', 'Data Lake Analytics', 'General', 'HDInsight', 'Iteration & conditionals', 'Machine Learning', and 'Power Query'. The main workspace shows a pipeline named 'pipeline1' with three activities: 'Copy data' (CopyToCSV), 'Copy data' (CopyToParquet), and 'Copy data' (CopyToAvro). The pipeline status is 'Succeeded'. The 'Output' tab is selected, showing a table of activity results.

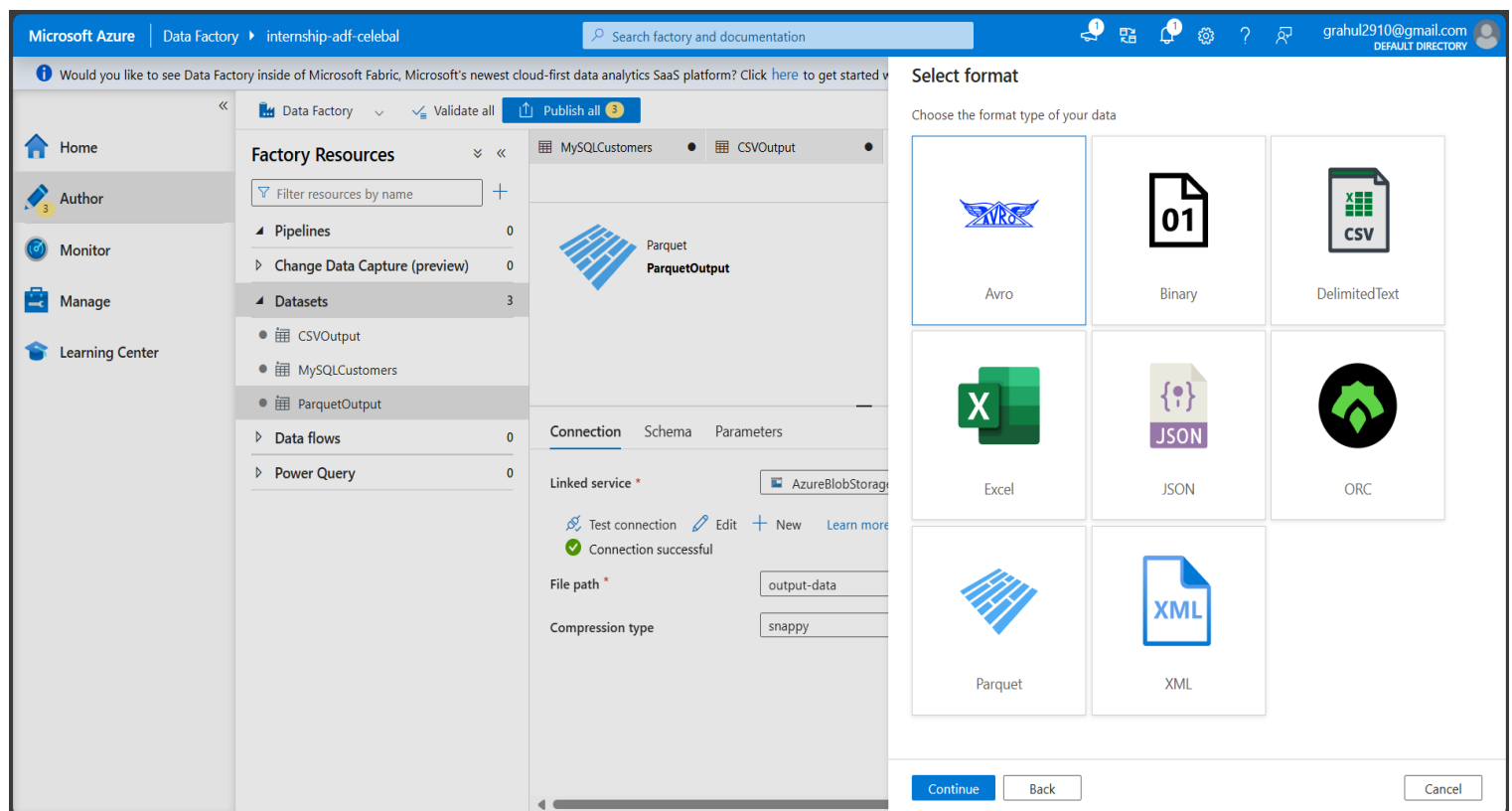
Activity name	Activity st...	Activit...
CopyToCSV	Succeeded	Copy data
CopyToAvro	Succeeded	Copy data
CopyToParquet	Succeeded	Copy data

## 6.) Created Sink Datasets

- CSV Output
  - Created a dataset: AzureBlob\_CSV
  - Format: DelimitedText



- Parquet Output
  - Dataset: AzureBlob\_Parquet
  - Format: Parquet



- Avro Output
  - Dataset: AzureBlob\_Avro
  - Format: Avro

This screenshot shows the Microsoft Azure Data Factory Author interface. The left sidebar contains navigation options: Home, Author (selected), Monitor, Manage, and Learning Center. The main workspace is divided into three panes. The 'Factory Resources' pane on the left shows a tree view with 'Pipelines' (1 item: pipeline1) and 'Datasets' (4 items: AvroOutput, CSVOutput, MySQLCustomers, ParquetOutput). The 'Activities' pane in the center shows a list of activities under 'Move and transform', including 'Copy data'. The 'Copy data' activity is selected, and its configuration pane on the right is visible. The 'General' tab of the 'Copy data' activity shows the 'Name' as 'CopyToAvro' and the 'Description' as an empty text box. The 'Activity state' is set to 'Activated'. The 'Timeout' is set to '0'. The 'Properties' pane on the far right shows the 'General' tab for the selected pipeline, with the 'Name' as 'pipeline1' and the 'Description' as an empty text box. The 'Annotations' section is also empty.

This screenshot shows the Microsoft Azure Data Factory Author interface, specifically the configuration for the 'AvroOutput' dataset. The left sidebar is the same as the previous screenshot. The 'Factory Resources' pane shows the 'AvroOutput' dataset selected. The main workspace displays the 'AvroOutput' dataset configuration. The 'Connection' tab is active, showing the 'Linked service' as 'AzureBlobStorage1'. The 'File path' is set to 'output-data / Directory / custo'. The 'Compression type' is set to 'No compression'. The 'Properties' pane on the right shows the 'General' tab for the dataset, with the 'Name' as 'AvroOutput' and the 'Description' as an empty text box. The 'Annotations' section is also empty.

**Output-** The output-data file is the blob file that stores all the data that is converted from MYSQL to CSV, Avro and parguet.

Microsoft Azure

Search resources, services, and docs (G+)

Copilot

grahul2910@gmail.com  
DEFAULT DIRECTORY

Home > All resources > celebalinternstorage | Containers >

output-data  
Container

Search

+ Add Directory ↑ Upload 🔒 Change access level ↻ Refresh 🗑️ Delete 📄 Copy 📄 Paste 🔄 Rename ⚙️ Acquire lease ⚙️ Break lease 🛠️ Edit columns

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

output-data

Authentication method: Access key (Switch to Microsoft Entra user account)

Add filter

Search blobs by prefix (case-sensitive)

Only show active blobs

Showing all 4 items

<input type="checkbox"/>	Name	Last modified	Access tier	Blob type	Size	Lease state
<input type="checkbox"/>	customers.avro	4/7/2025, 4:25:24 pm	Hot (Inferred)	Block blob	507 B	Available
<input type="checkbox"/>	customers.csv	4/7/2025, 2:30:23 pm	Hot (Inferred)	Block blob	208 B	Available
<input type="checkbox"/>	customers.parquet	4/7/2025, 4:25:26 pm	Hot (Inferred)	Block blob	1017 B	Available
<input type="checkbox"/>	customers_output.csv	4/7/2025, 4:25:22 pm	Hot (Inferred)	Block blob	251 B	Available

Add or remove favorites by pressing Ctr+L+Shift+F

## Notes

- MySQL connection was done via Ngrok
- Output files are stored in Azure Blob Storage container 'output-data'