Read Different File Formats from ADLS Gen2 Using SQL Syntax

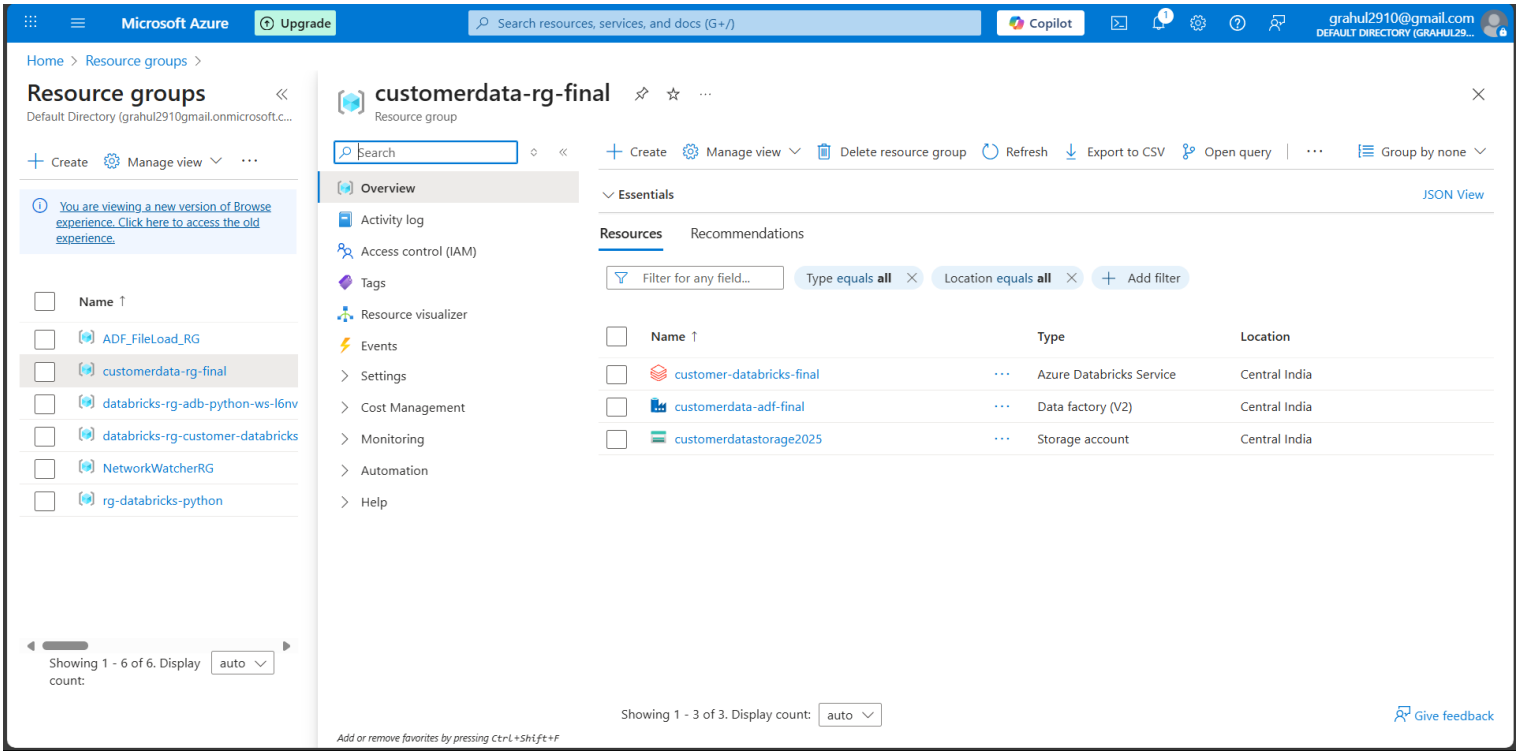
1. Created Sample Data

- Prepared a CSV file customers.csv with 500 rows of sample customer records.
- Included fields: CustomerID, FirstName, LastName, Email, Phone, City, Country.

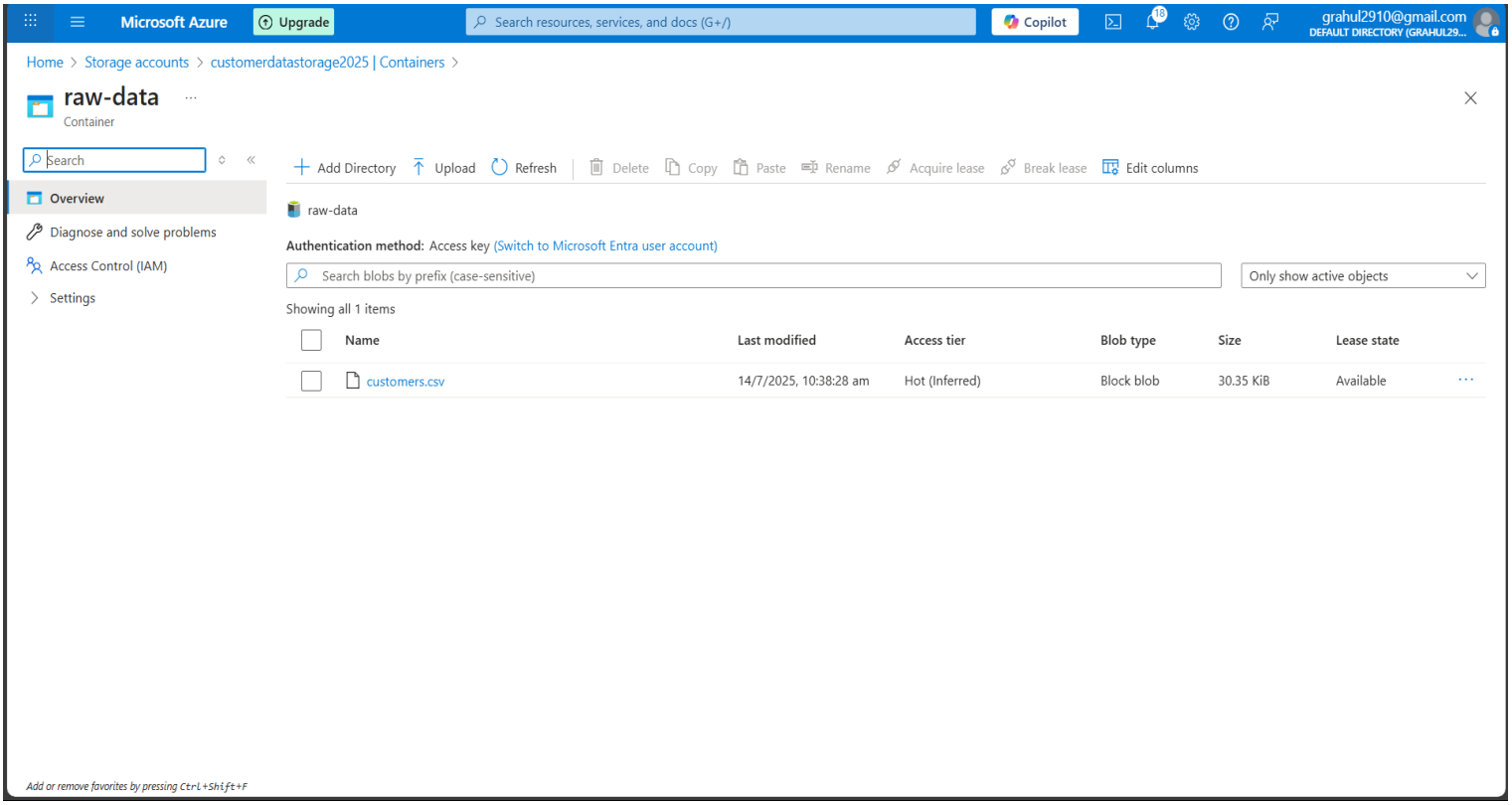
id	first_name	last_name	email	gender	ip_address
1	Deidre	McLeoid	dmcleoid0@github.io	Female	241.79.115.71
2	Jacqui	Gaines	jgaines1@free.fr	Female	195.87.58.156
3	Casandra	Biggam	cbiggam2@thetimes.co.uk	Female	252.241.154.87
4	Morie	Sartain	msartain3@aboutads.info	Male	170.82.69.153
5	Tommie	Longworthy	tlongworthy4@ehow.com	Male	16.12.144.36
6	Oliviero	Parkhouse	oparkhouse5@yolasite.com	Bigender	65.3.164.12
7	Cornie	Sellar	csellar6@sun.com	Male	30.104.150.51
8	Demetris	Prudence	dprudence7@taobao.com	Female	20.53.25.133
9	Francesca	Hamerton	fhamerton8@epa.gov	Female	2.200.9.122
10	Dave	Richardt	drichardt9@biblegateway.com	Male	108.91.69.218
11	Gayleen	Testin	gtestina@newyorker.com	Female	79.77.236.112
12	Tabbi	Alberti	talbertib@google.com.hk	Female	255.161.132.116
13	Becka	Bambrick	bbambrickc@mysql.com	Female	72.242.69.255
14	Ebeneser	Purcell	epurcelld@adobe.com	Male	43.91.219.143
15	Nollie	Tandy	ntandye@clickbank.net	Female	138.160.54.194
16	Clemente	Overall	coverallf@vkontakte.ru	Male	29.9.161.129
17	Vlad	Holby	vholbyg@uol.com.br	Male	109.116.9.221
18	Audrey	Pales	apalesh@shutterfly.com	Female	162.18.171.235
19	Konstance	Longhi	klonghii@friendfeed.com	Female	251.78.234.118
20	Leyla	Curry	lcurryj@dropbox.com	Female	207.77.180.107
21	Vincenty	Coppeard	vcoppeardk@prnewswire.com	Male	127.144.182.48
22	Graehme	Twyning	gtwyningl@adobe.com	Male	250.123.15.166
23	Mathias	Tondeur	mtondeurm@cnbc.com	Male	30.194.99.49
24	Amelie	Antoons	aantoonsn@elpais.com	Female	57.188.242.197
25	Hirsch	Tyers	htyerso@bloomberg.com	Male	202.135.214.221

2. Created Azure Resources

- Resource Group: customerdata-rg-final



- Storage Account: Enabled ADLS Gen2 by checking Hierarchical Namespace
- Containers Created:
    - raw-data: for original customers.csv
    - converted: for output in multiple formats

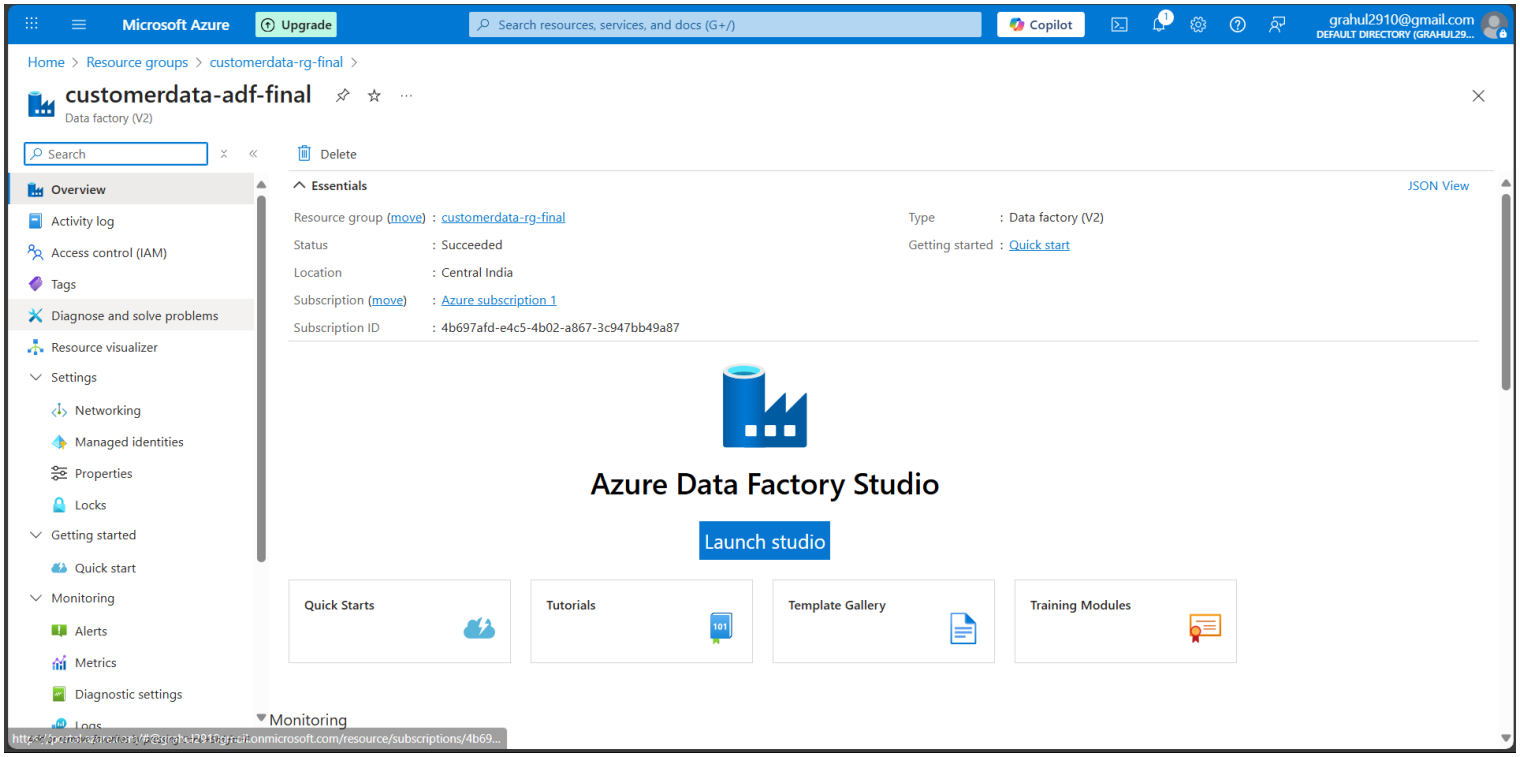


3. Uploaded CSV File

- Uploaded customers.csv to the raw-data container using the Azure portal.

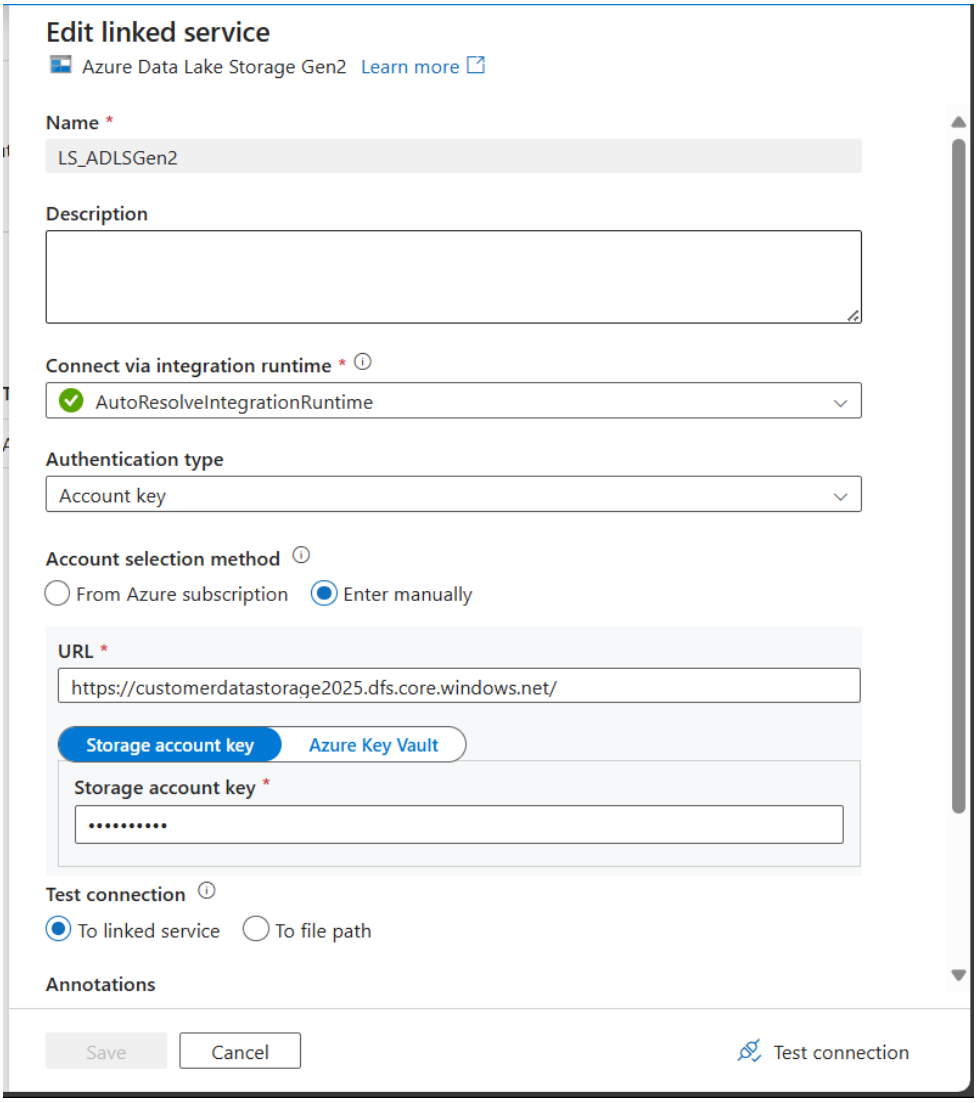
4. Created Azure Data Factory

- Created a data factory named customerdata-adf-final.
- Launched ADF Studio for pipeline development.



5. Created Linked Services

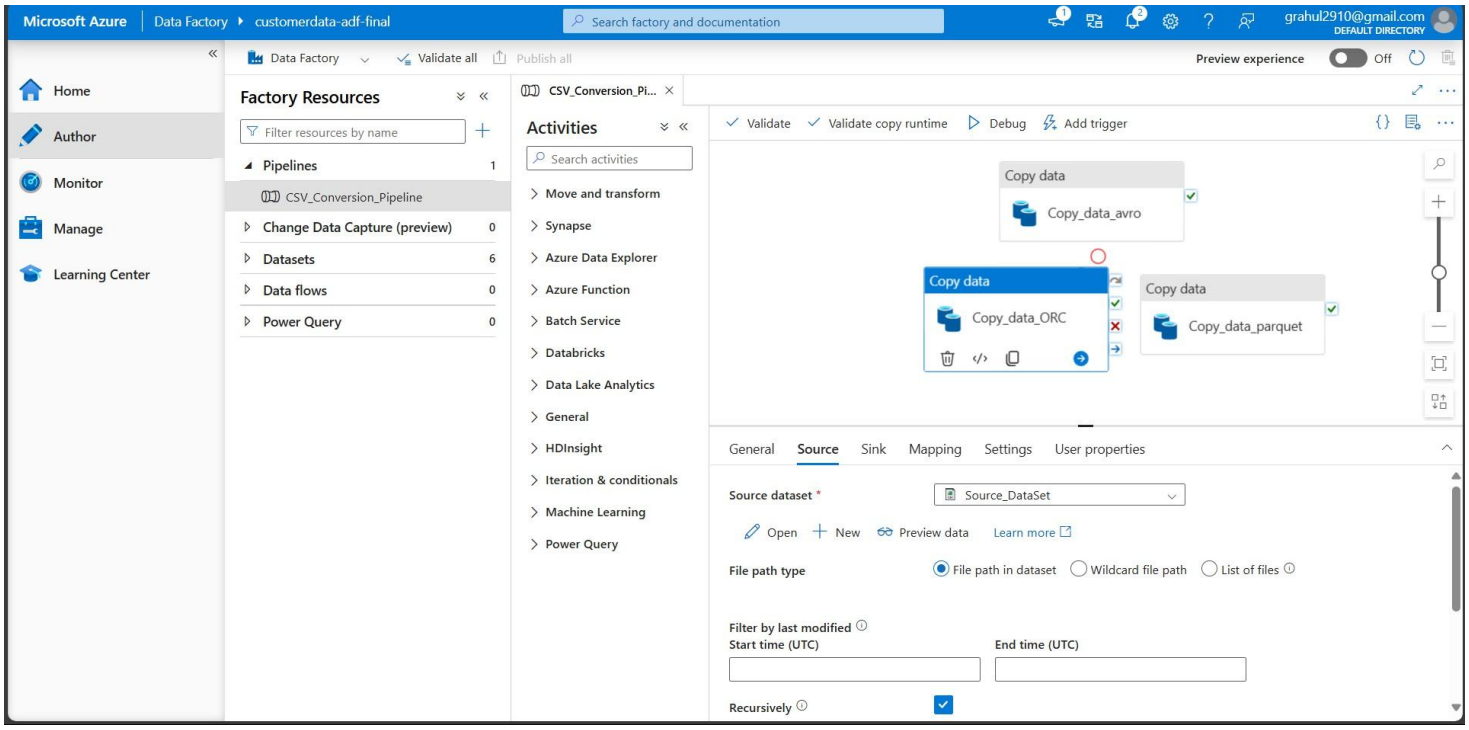
- Created a linked service (LS_ADLSGen2) to connect ADF to the ADLS Gen2 storage.



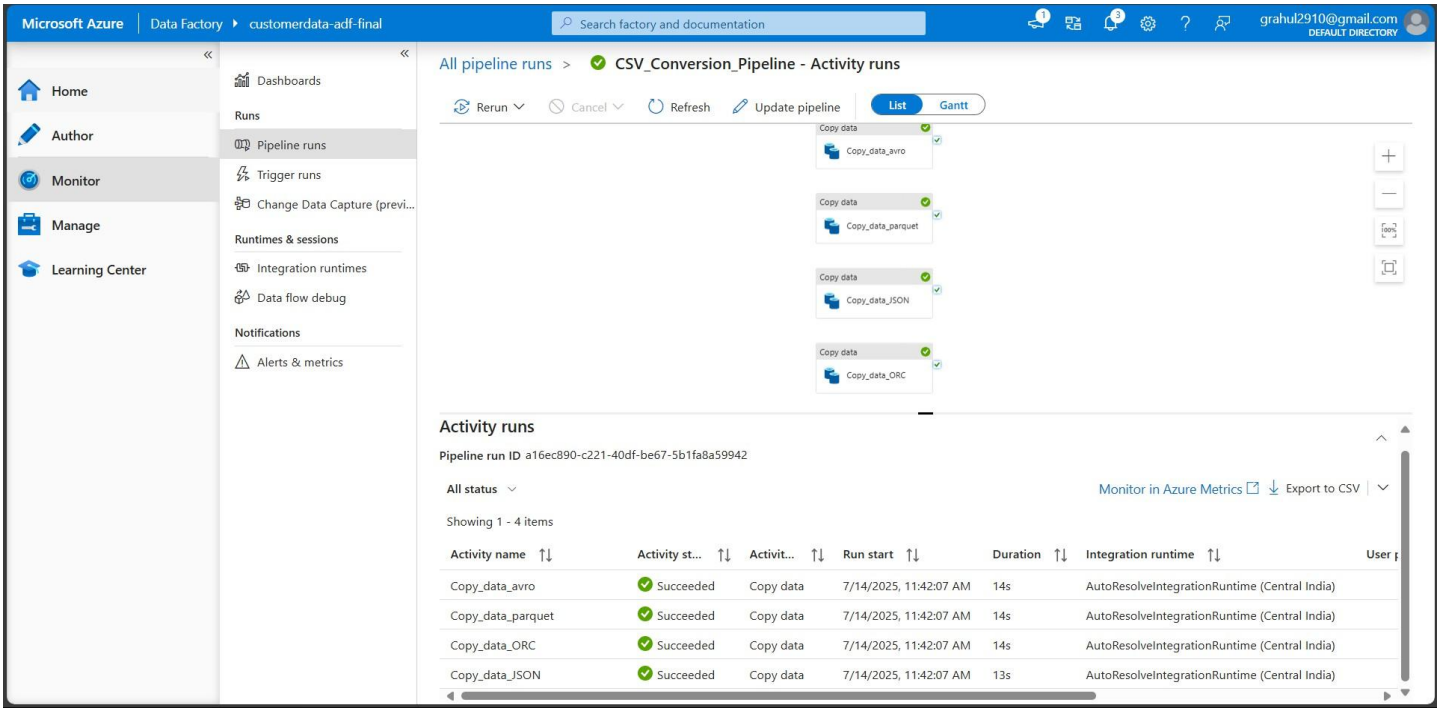
6. Built ADF Pipeline for File Format Conversion

- Pipeline Name: CSV_Conversion_Pipeline
- Created one Copy Data Activity each to convert customers.csv into:
  - Parquet → stored in converted/parquet/
  - Avro → stored in converted/avro/
  - ORC → stored in converted/orc/

Delta was not supported directly in ADF, so it was generated later via Databricks.

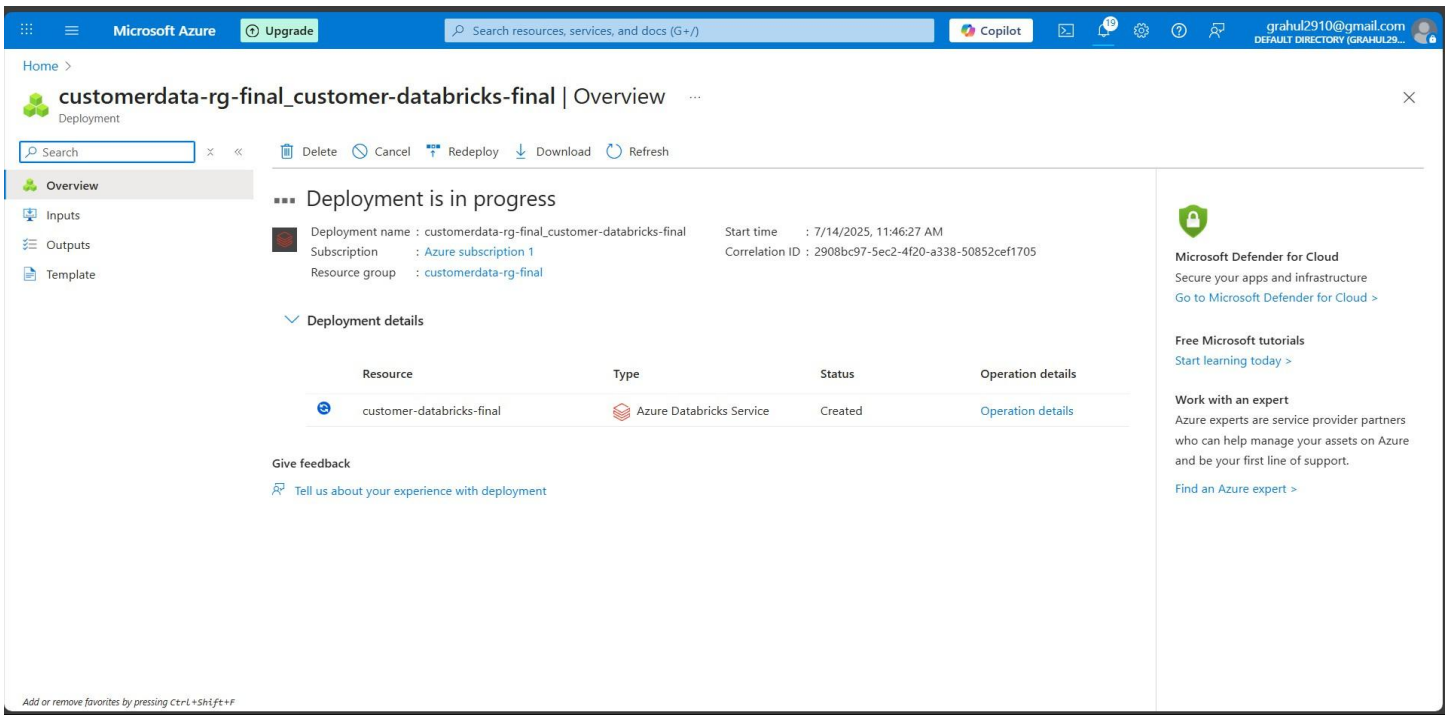


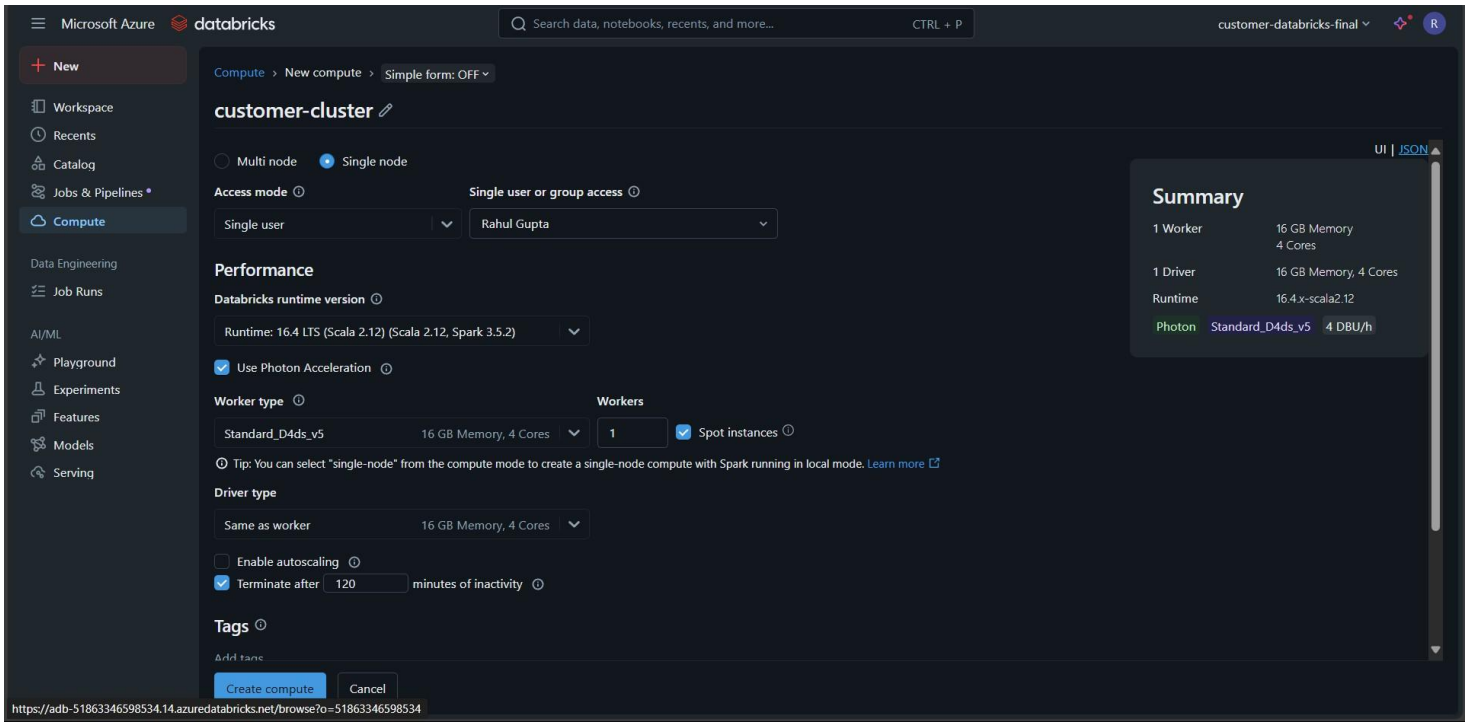
Run the pipeline to convert CSV to various file formats.



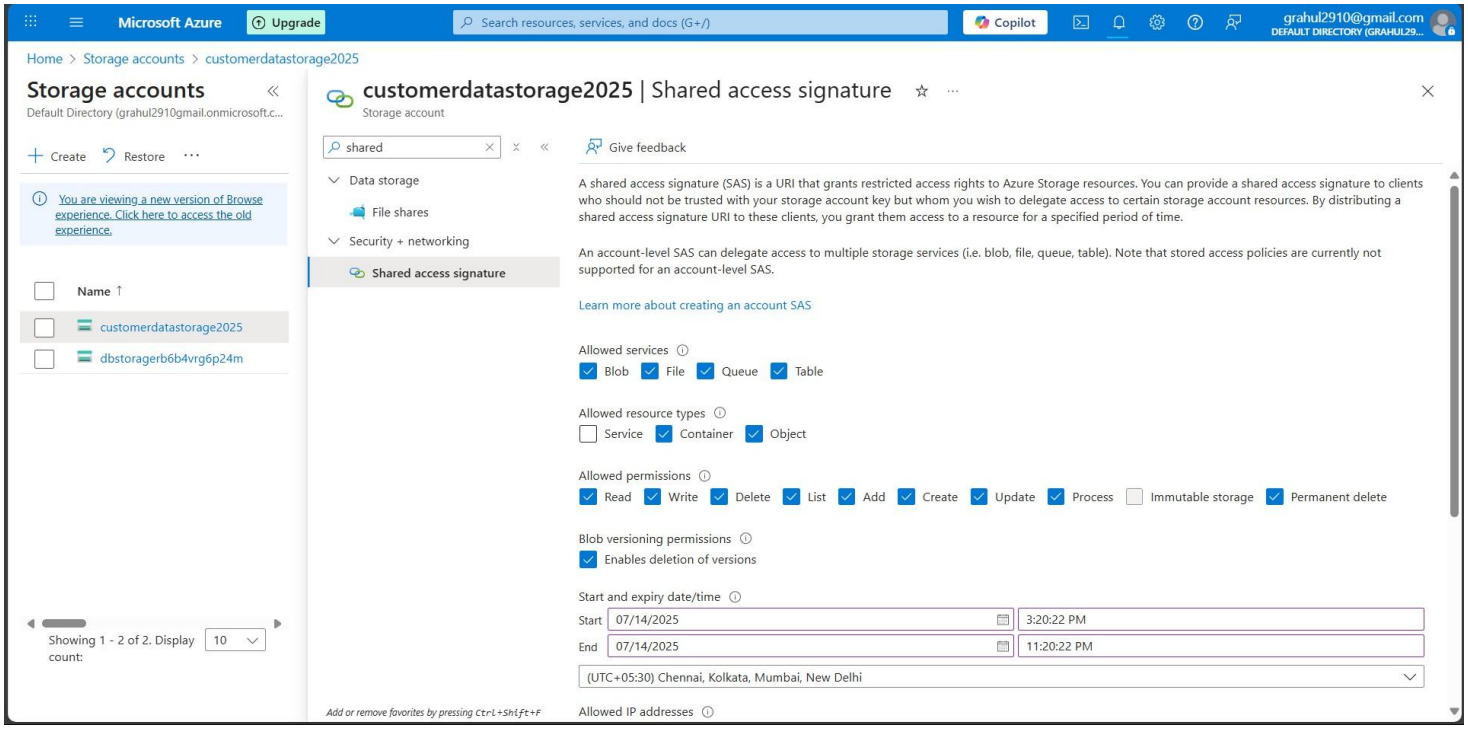
7. Created Databricks Workspace & Cluster

- Workspace: customer-databricks



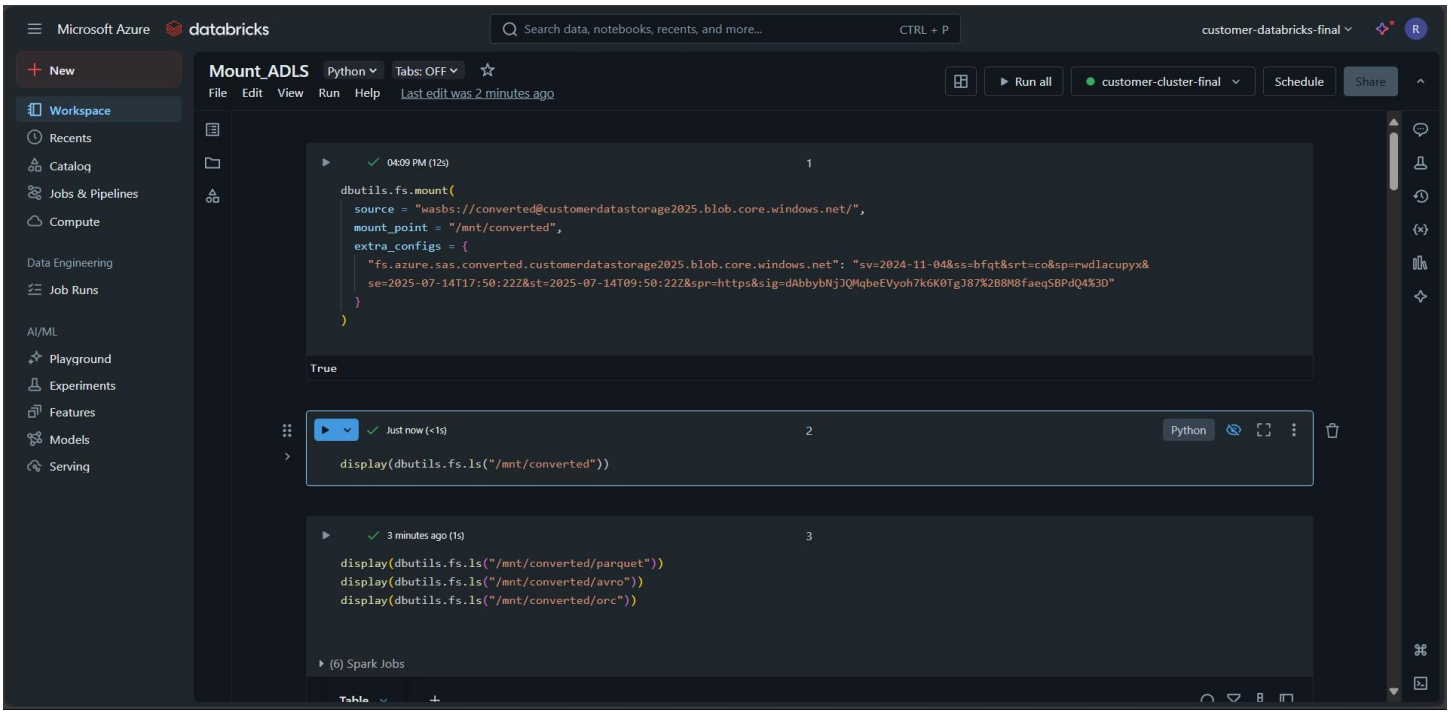


Configure Shared Access Signature, put the generated key in databricks.

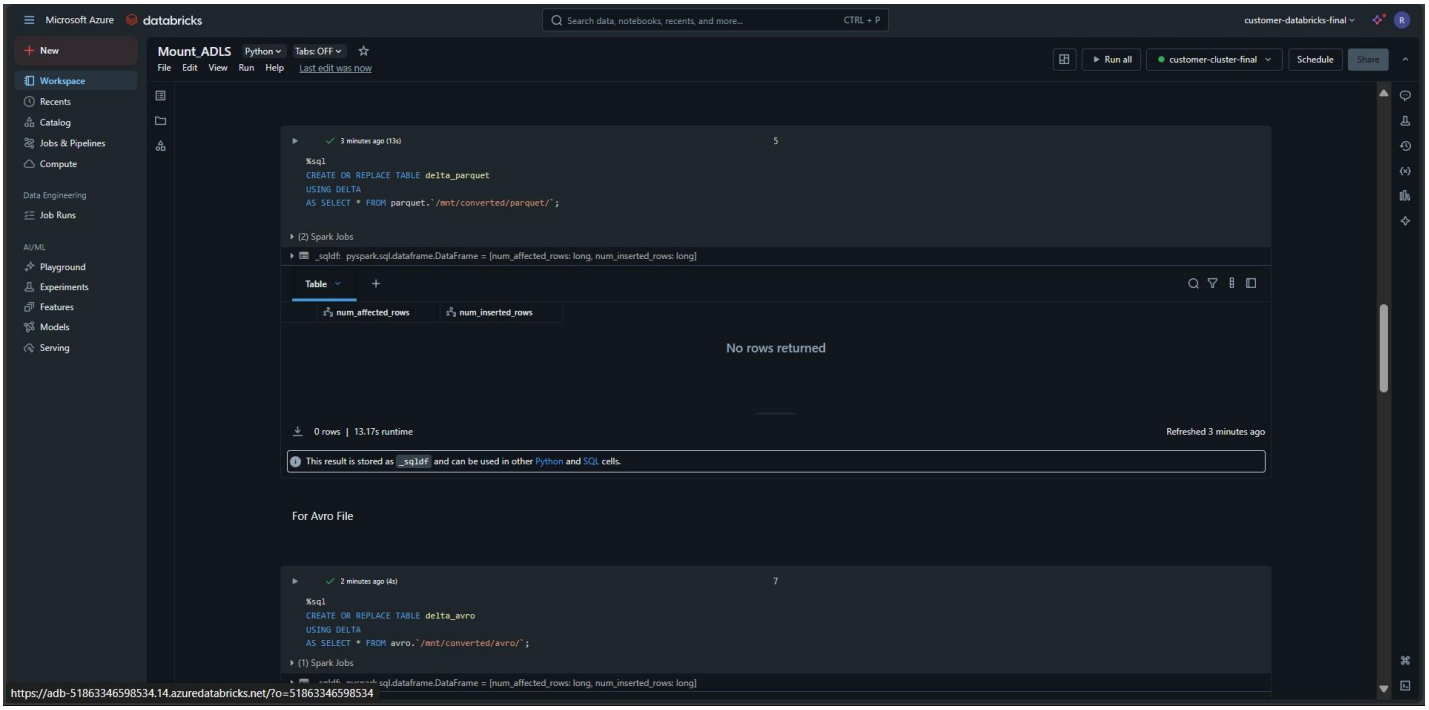


8. Mounted ADLS Containers in Databricks

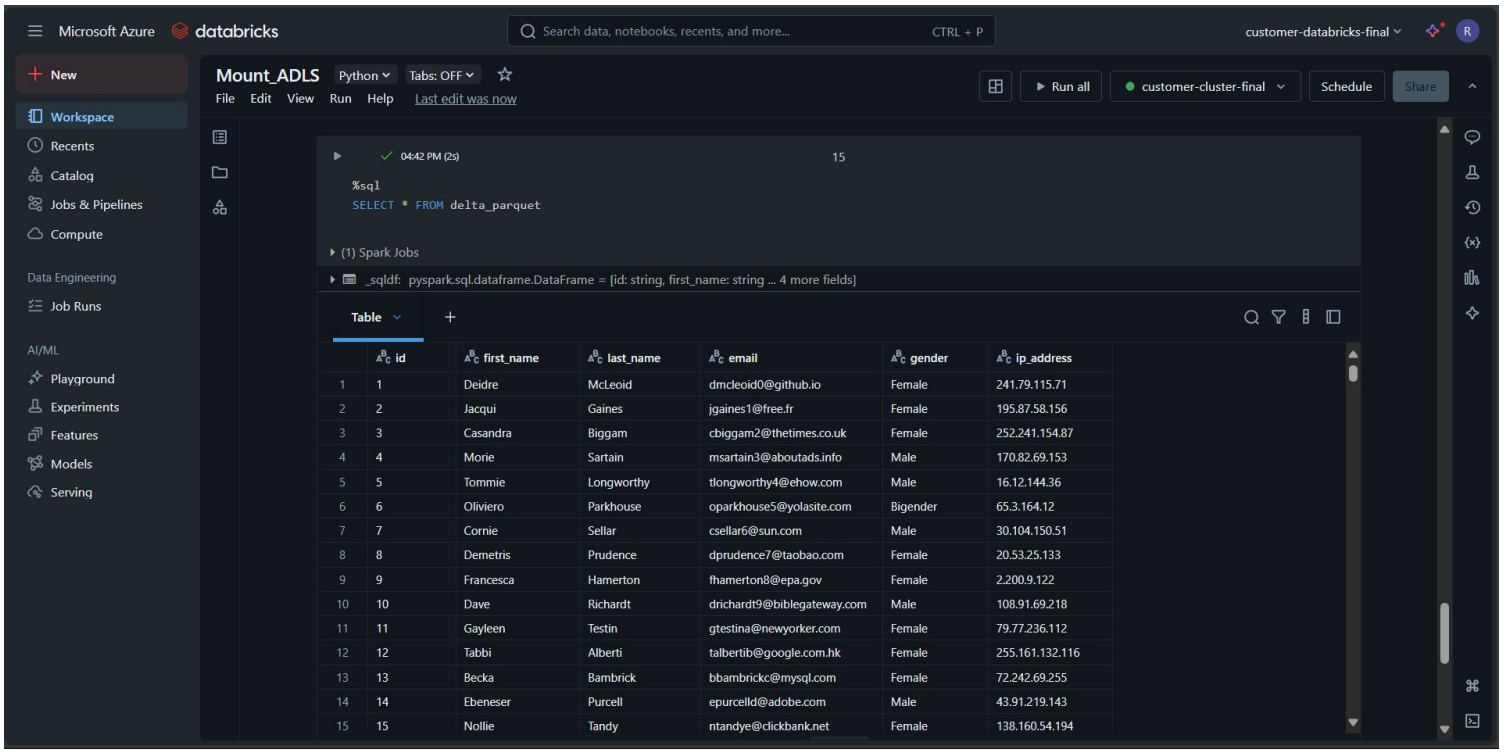
- Mounted both raw-data and converted containers using dbutils.fs.mount()



- Verified mount with dbutils.fs.ls("/mnt/raw-data") and dbutils.fs.ls("/mnt/converted")

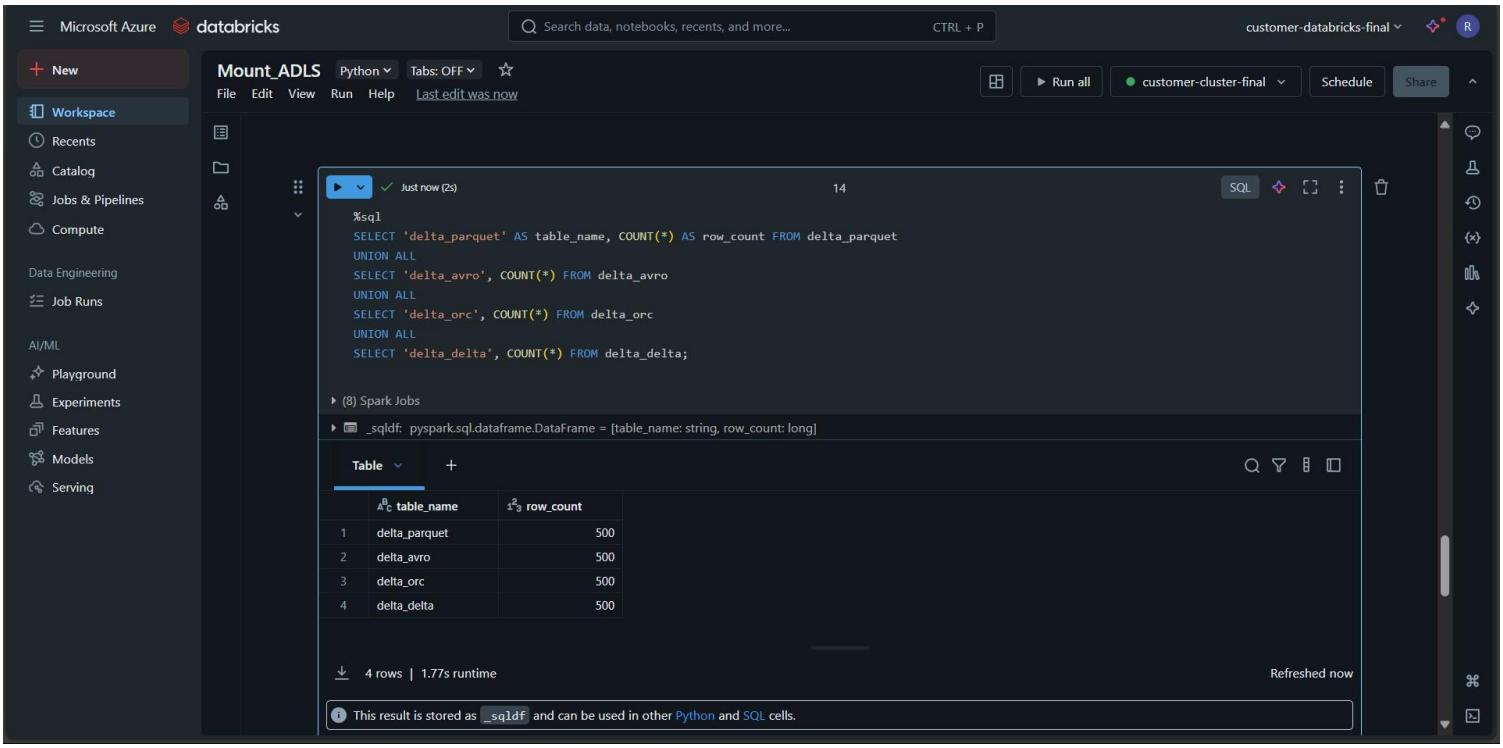


9. Created Delta File Using Databricks (for Delta format)



	id	first_name	last_name	email	gender	ip_address
1	1	Deidre	McLeoid	dmcleoid0@github.io	Female	241.79.115.71
2	2	Jacqui	Gaines	jgaines1@free.fr	Female	195.87.58.156
3	3	Casandra	Biggam	cbiggam2@thetimes.co.uk	Female	252.241.154.87
4	4	Morie	Sartain	msartain3@aboutads.info	Male	170.82.69.153
5	5	Tommie	Longworthy	tlongworthy4@ehow.com	Male	16.12.144.36
6	6	Oliviero	Parkhouse	oparkhouse5@yolasite.com	Bigender	65.3.164.12
7	7	Cornie	Sellar	csellar6@sun.com	Male	30.104.150.51
8	8	Demetris	Prudence	dprudence7@taobao.com	Female	20.53.25.133
9	9	Francesca	Hamerton	fhamerton8@epa.gov	Female	2.200.9.122
10	10	Dave	Richardt	drichardt9@biblegateway.com	Male	108.91.69.218
11	11	Gayleen	Testin	gtestina@newyorker.com	Female	79.77.236.112
12	12	Tabbi	Alberti	talbertib@google.com.hk	Female	255.161.132.116
13	13	Becka	Bambrick	bbambrickc@mysql.com	Female	72.242.69.255
14	14	Ebeneser	Purcell	epurcelld@adobe.com	Male	43.91.219.143
15	15	Nollie	Tandy	ntandye@clickbank.net	Female	138.160.54.194

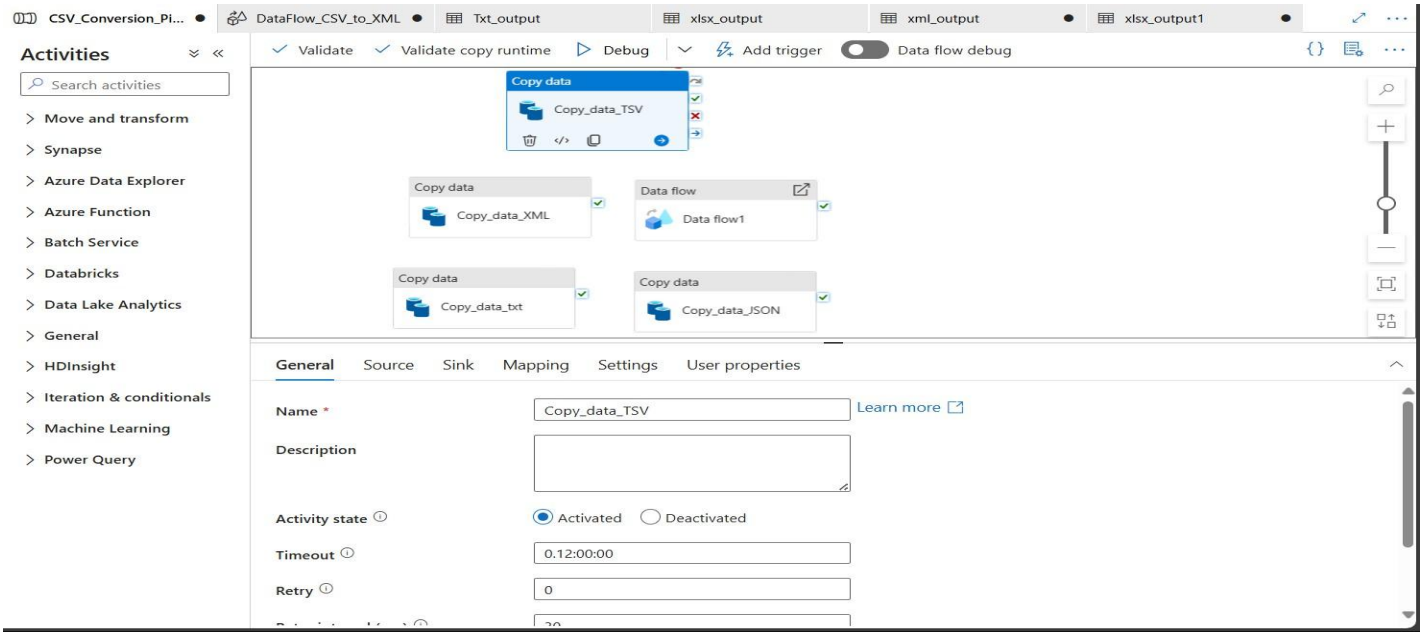
10. Verify and Validate that 500 rows are created



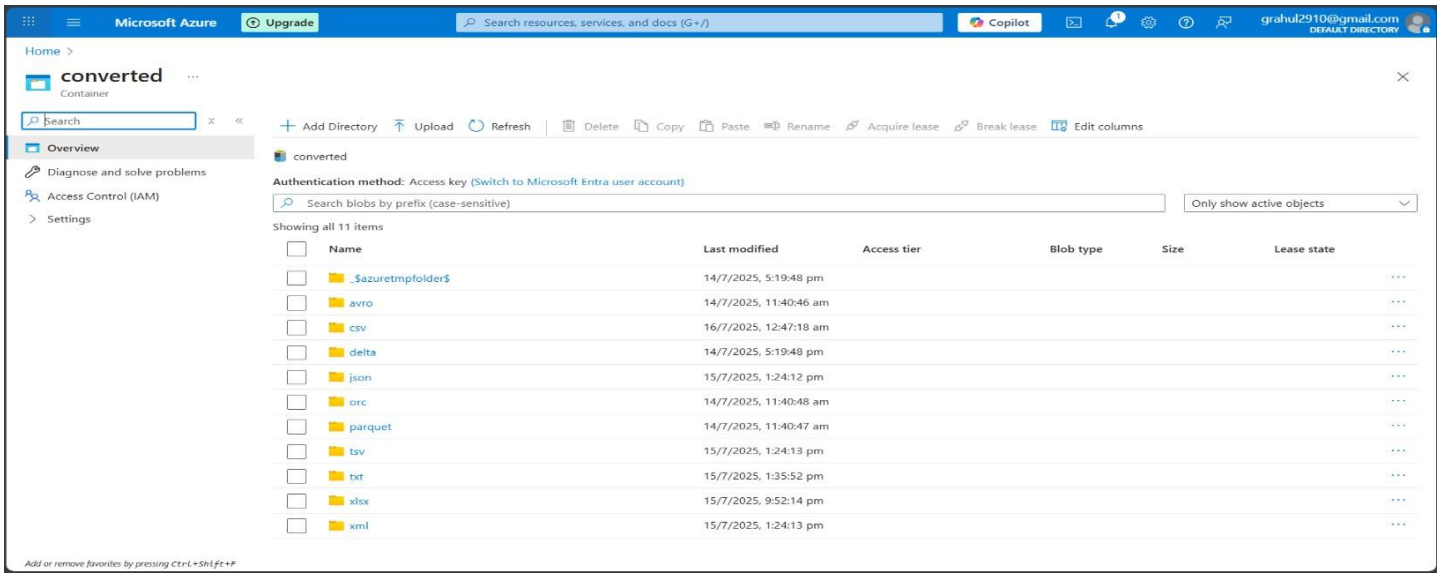
For CSV, TSV, JSON, XML,XLSX, TXT(using temporary view then converting it into Delta table):

1.) Changes to pipeline

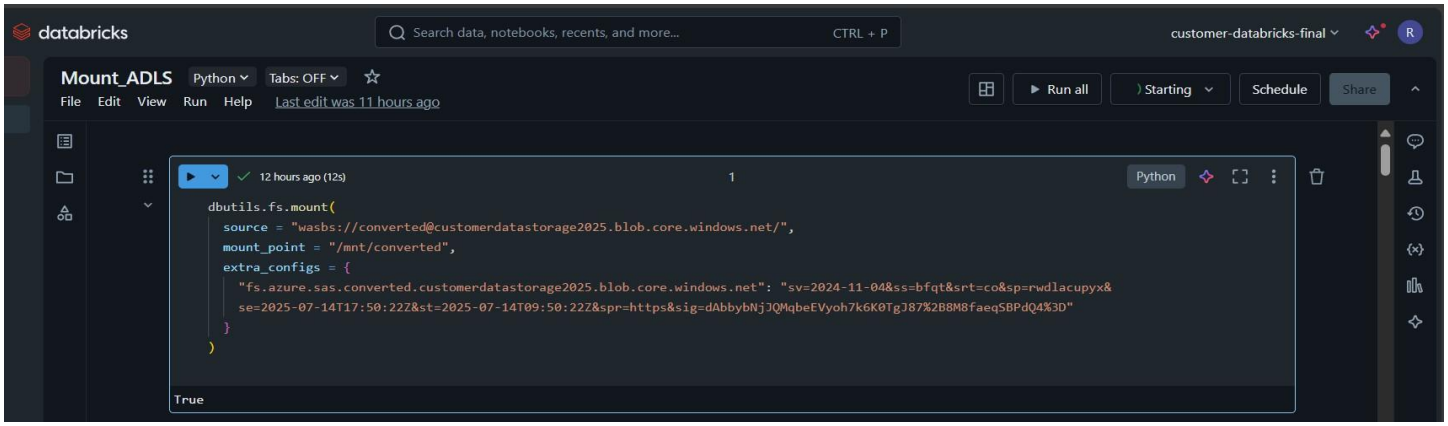
- Make changes to the pipeline so can it can convert sample data to various file formats



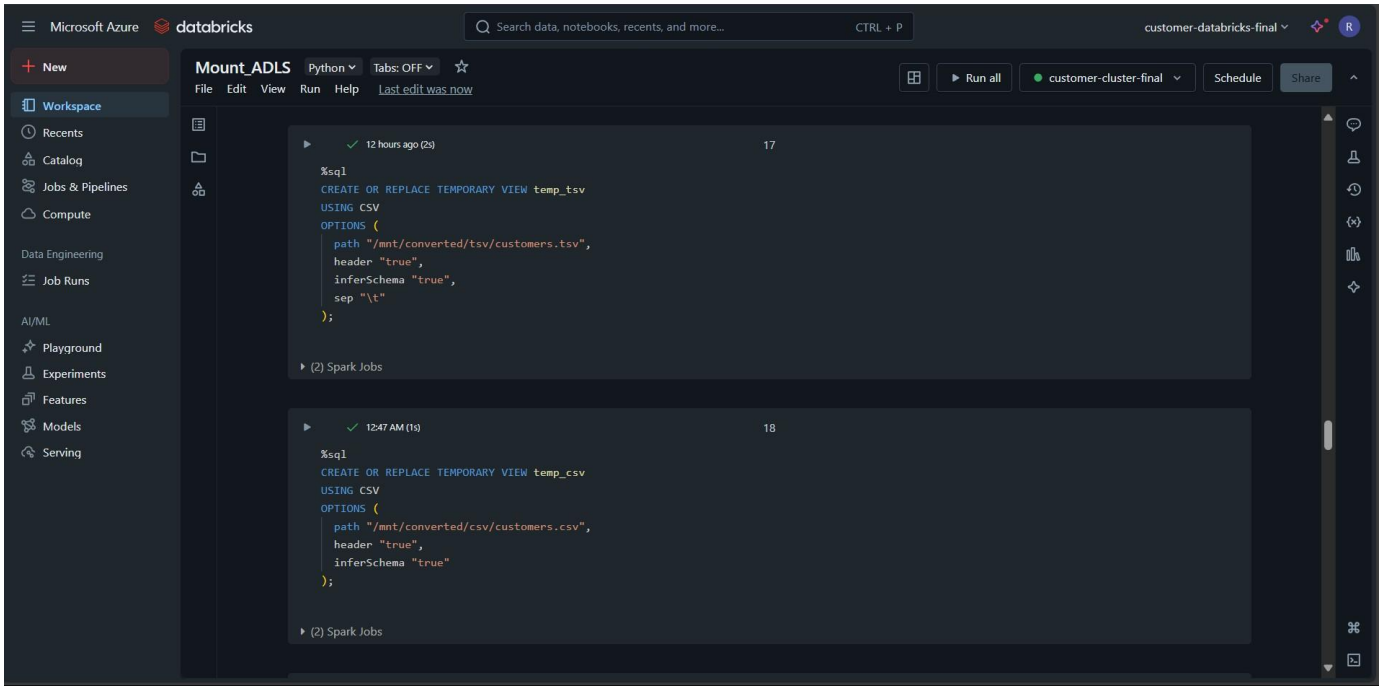
2.) Check the updated output



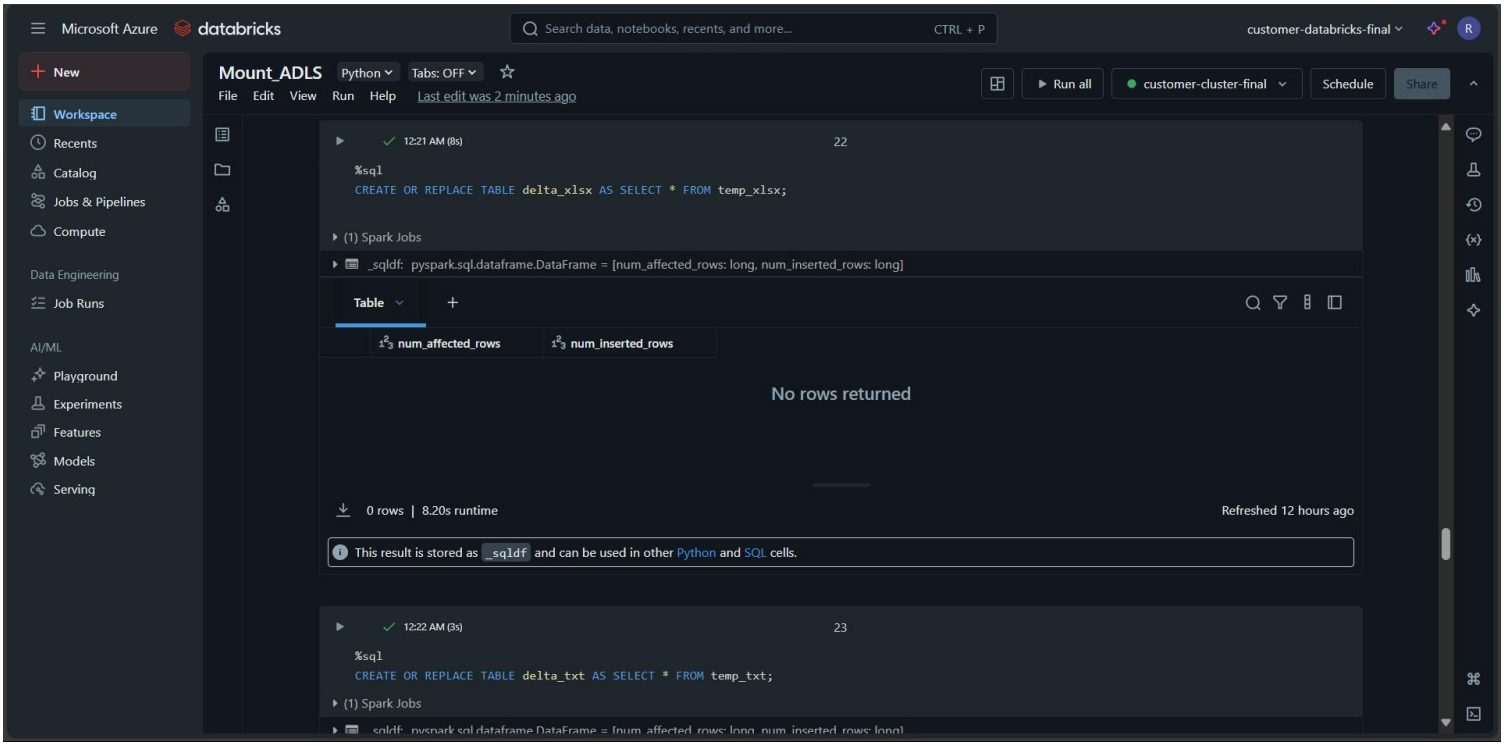
3.) Mount Data on Databricks



4.) Create Temporary View for all the file formats



5.) Create Table As Select using the temporary View



6.) Validate 500 rows