Tenant

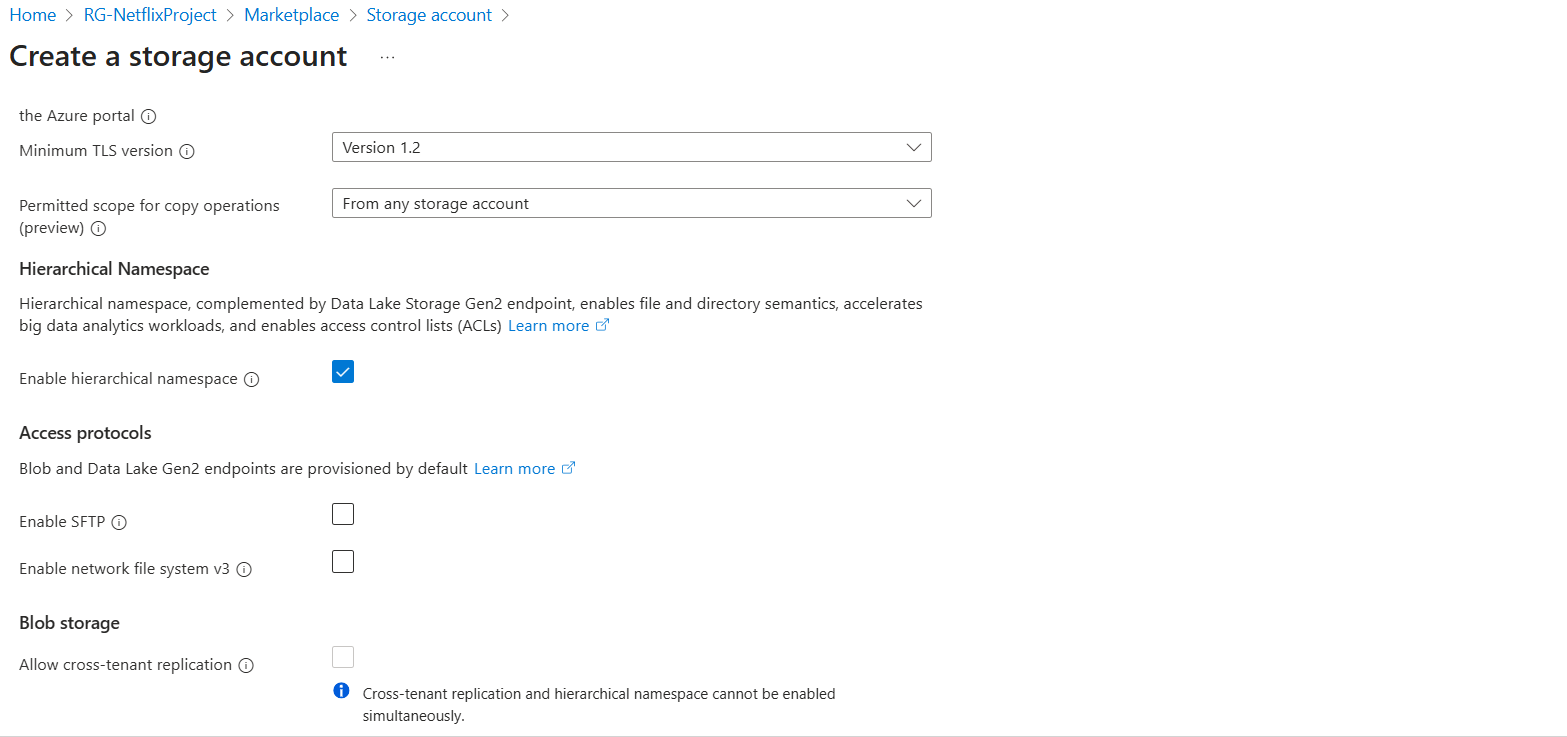
Subscriptions

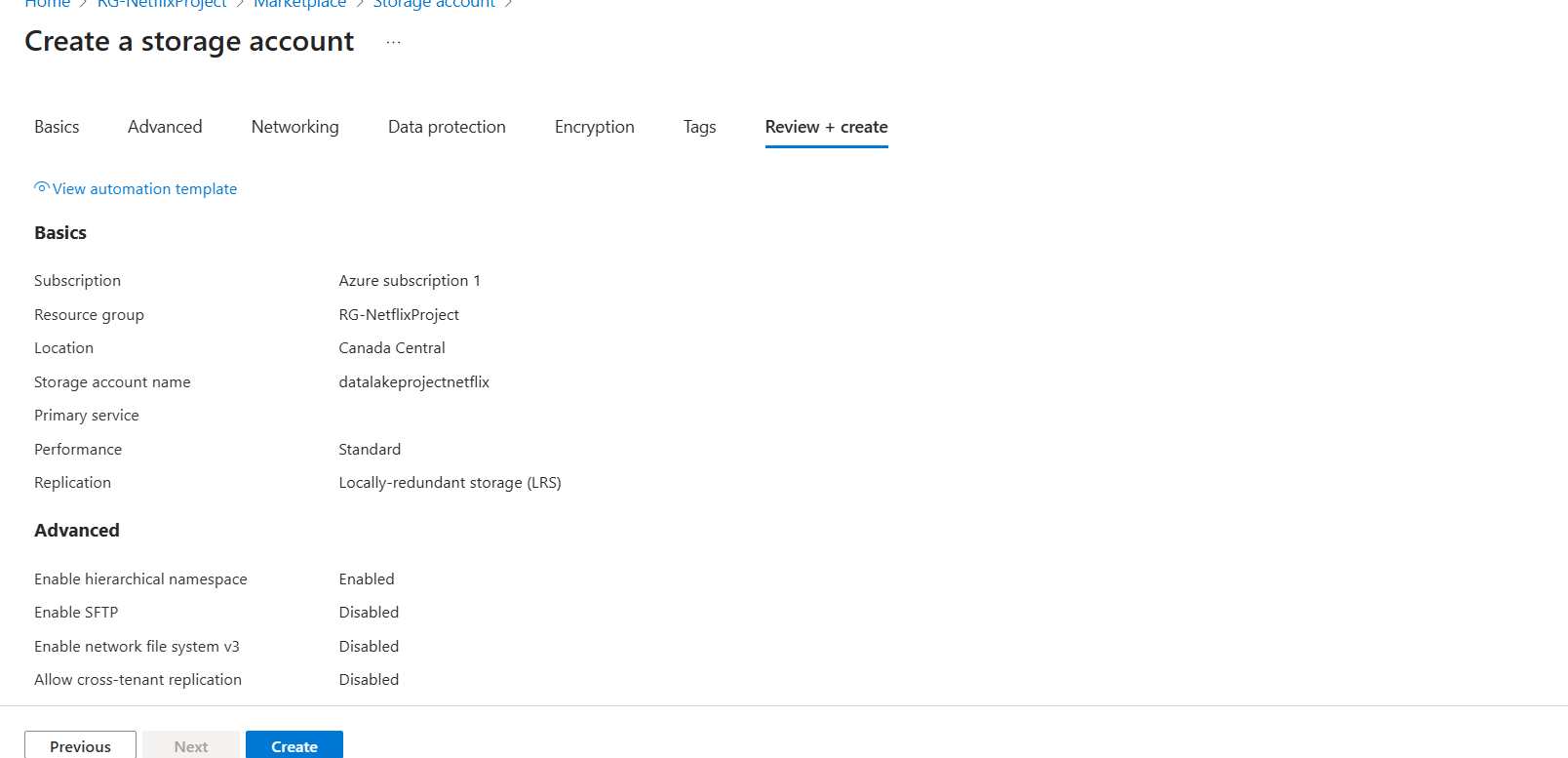
Resource Group

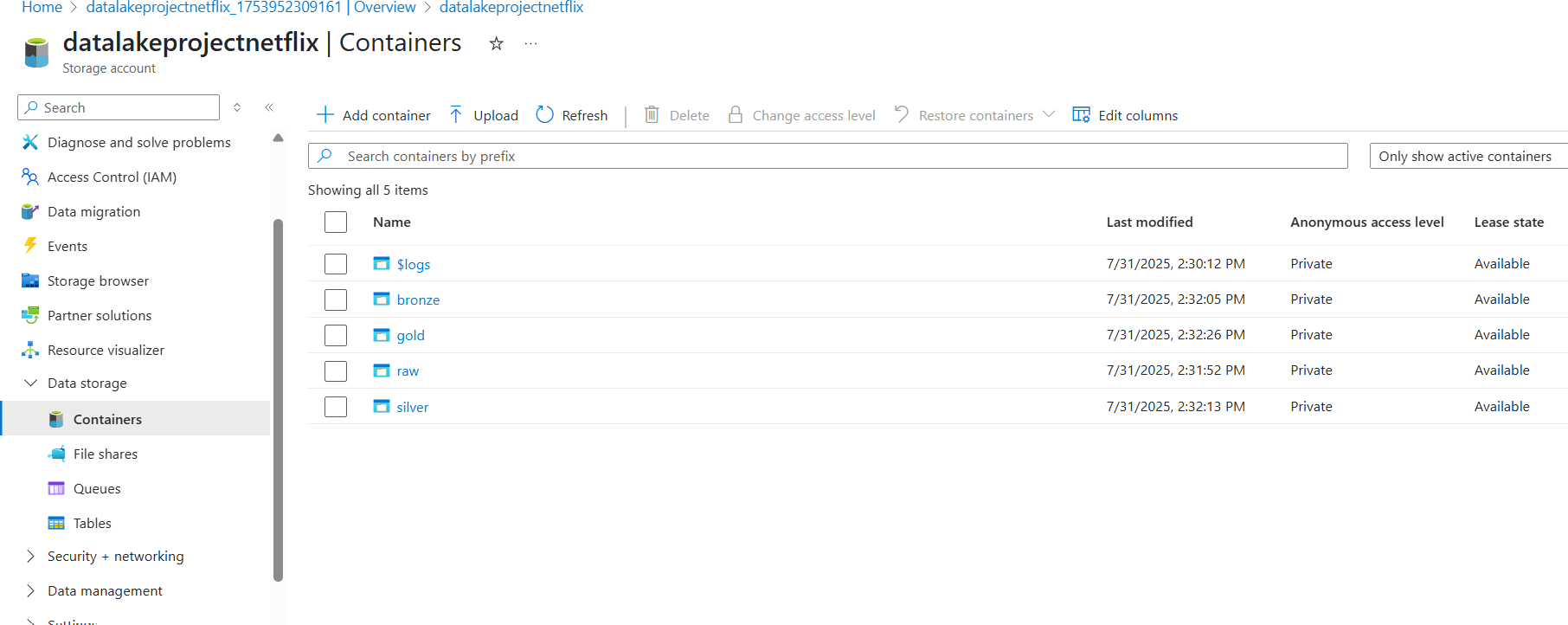
Resource groups

Resources

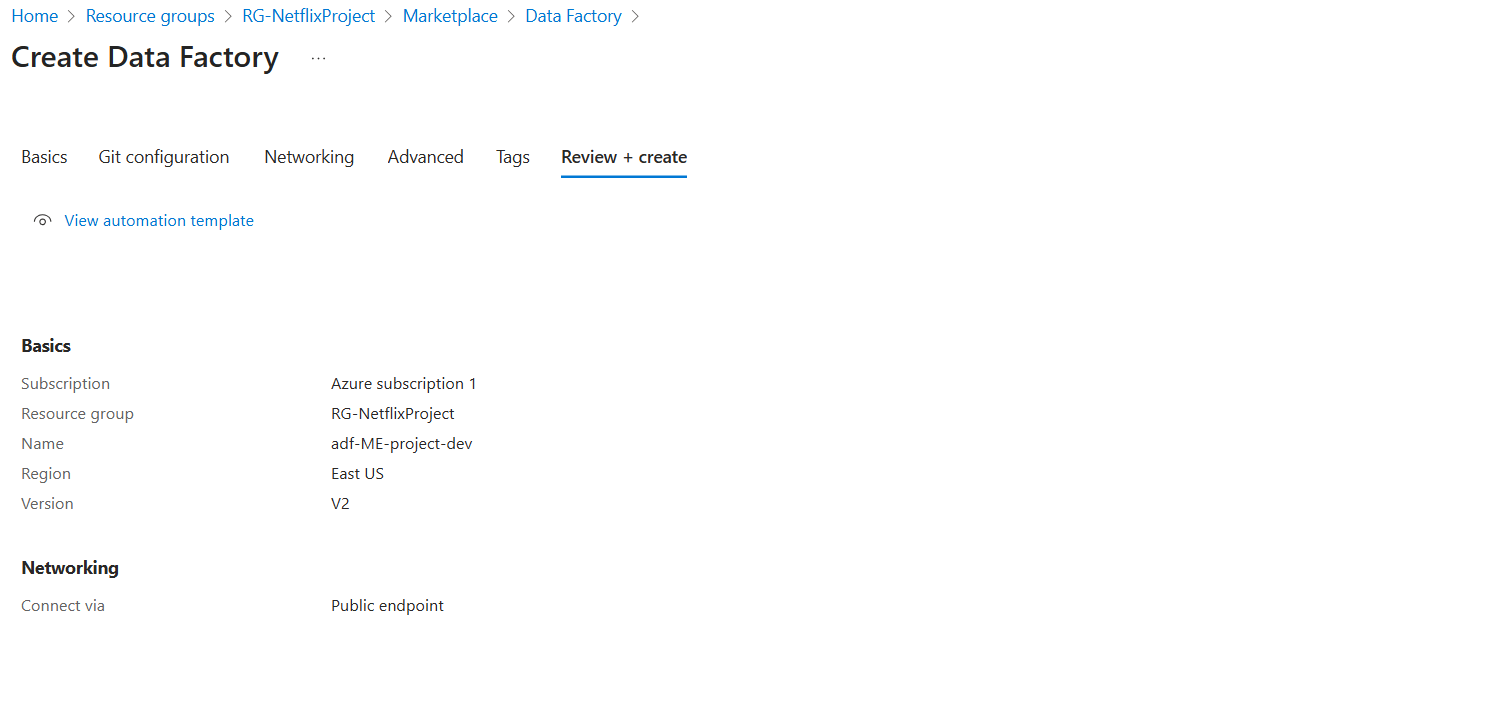
Storage account – datalakeprojectnetflix

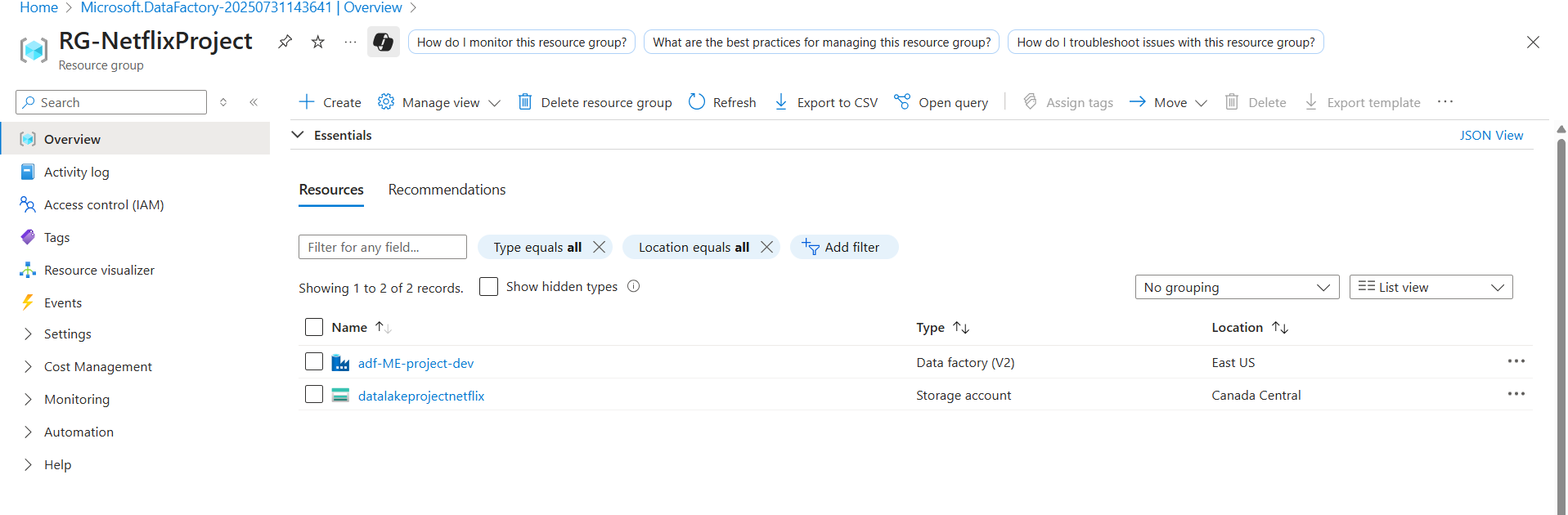


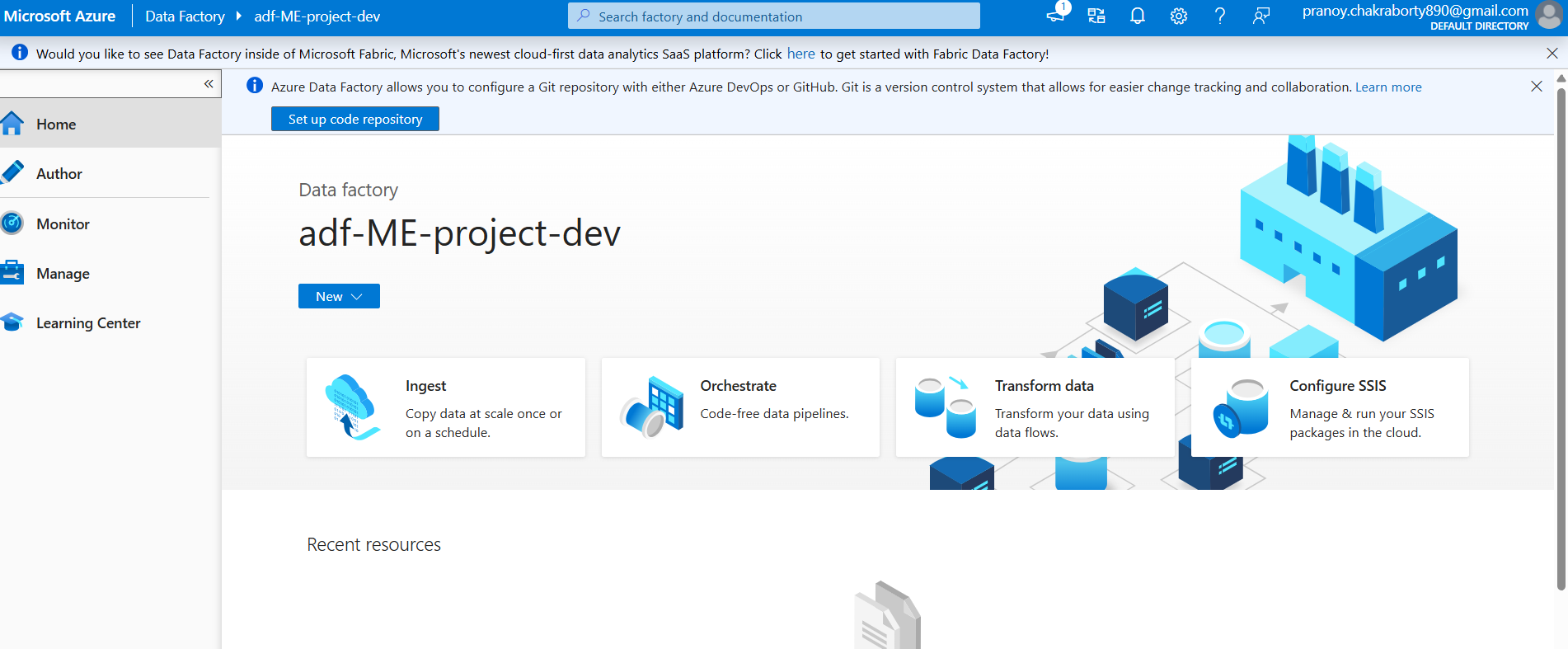




Azure Data Factory



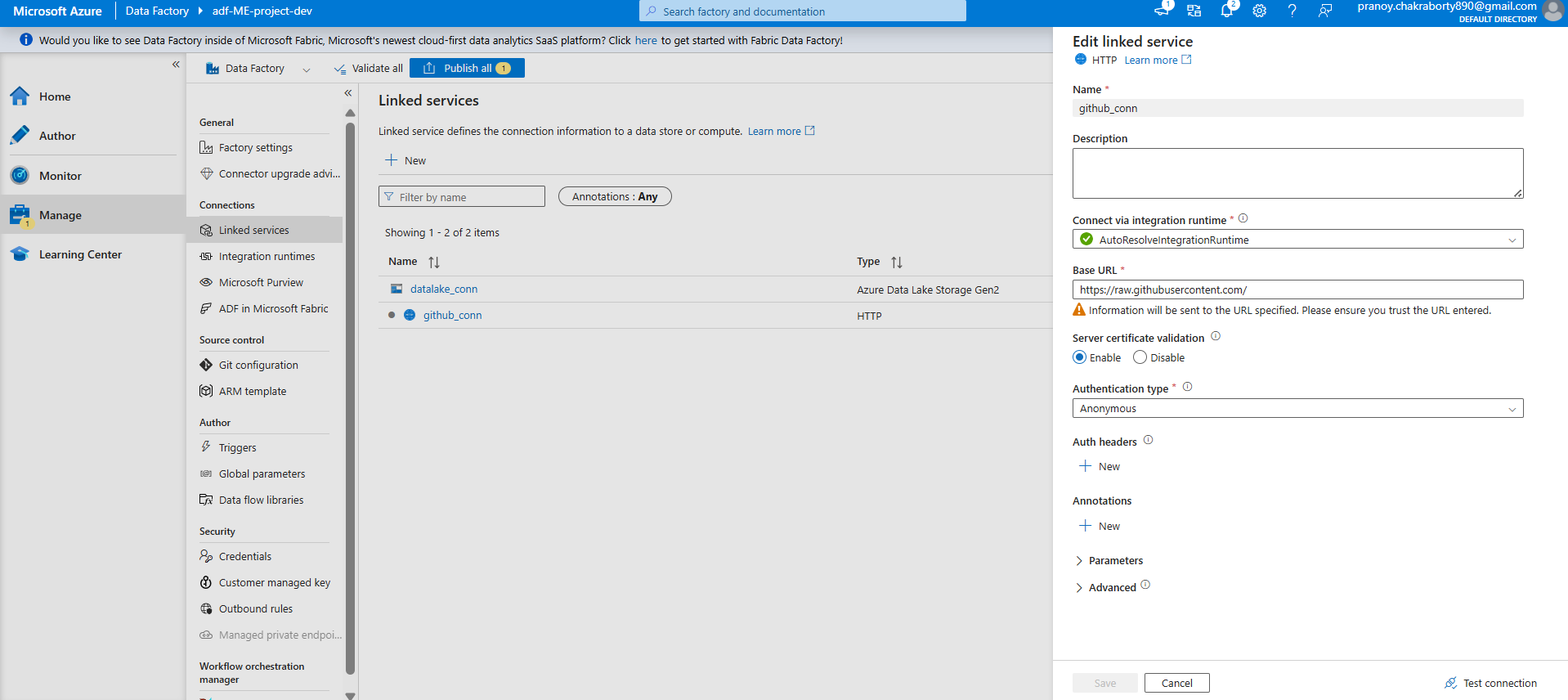




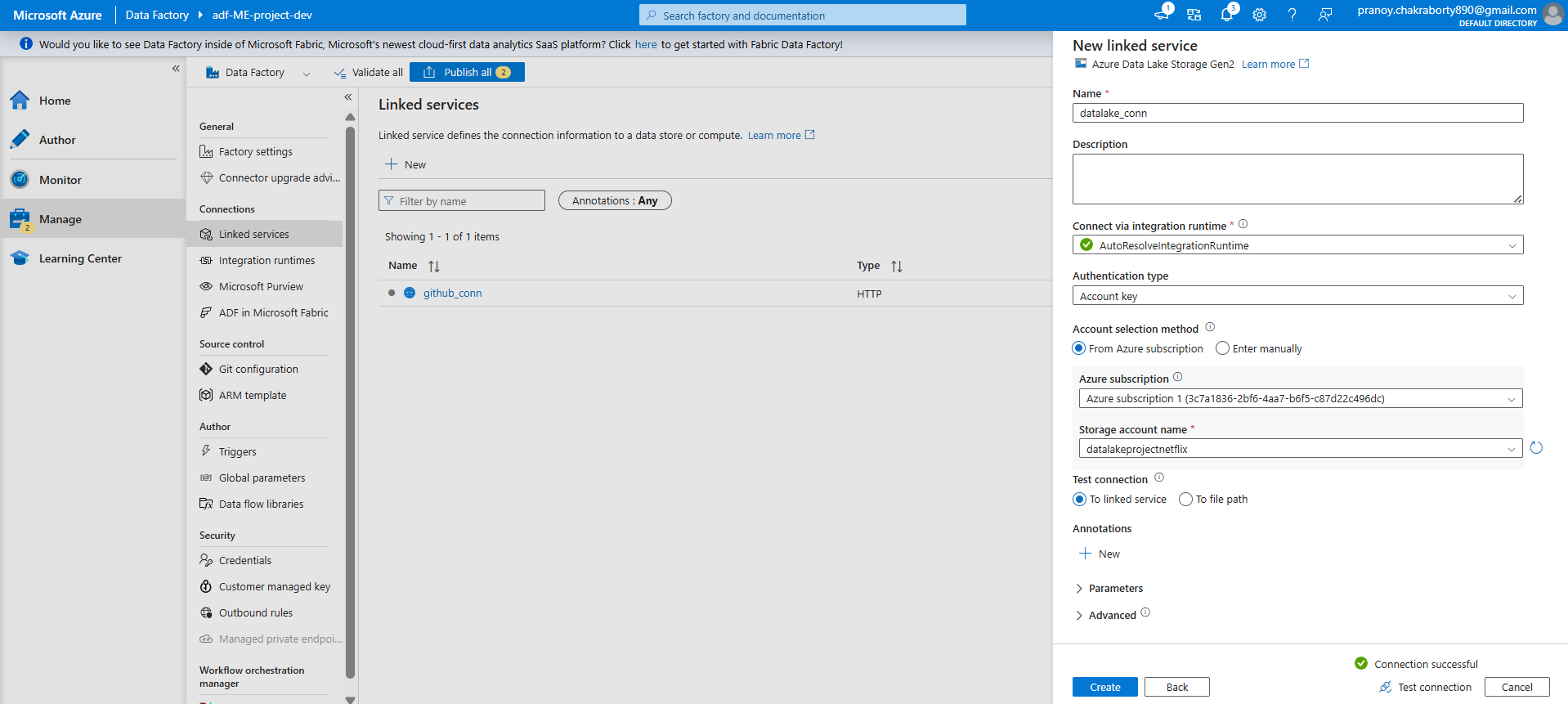
GitHub 🡪 ADF 🡪 Data Lake (storage account)

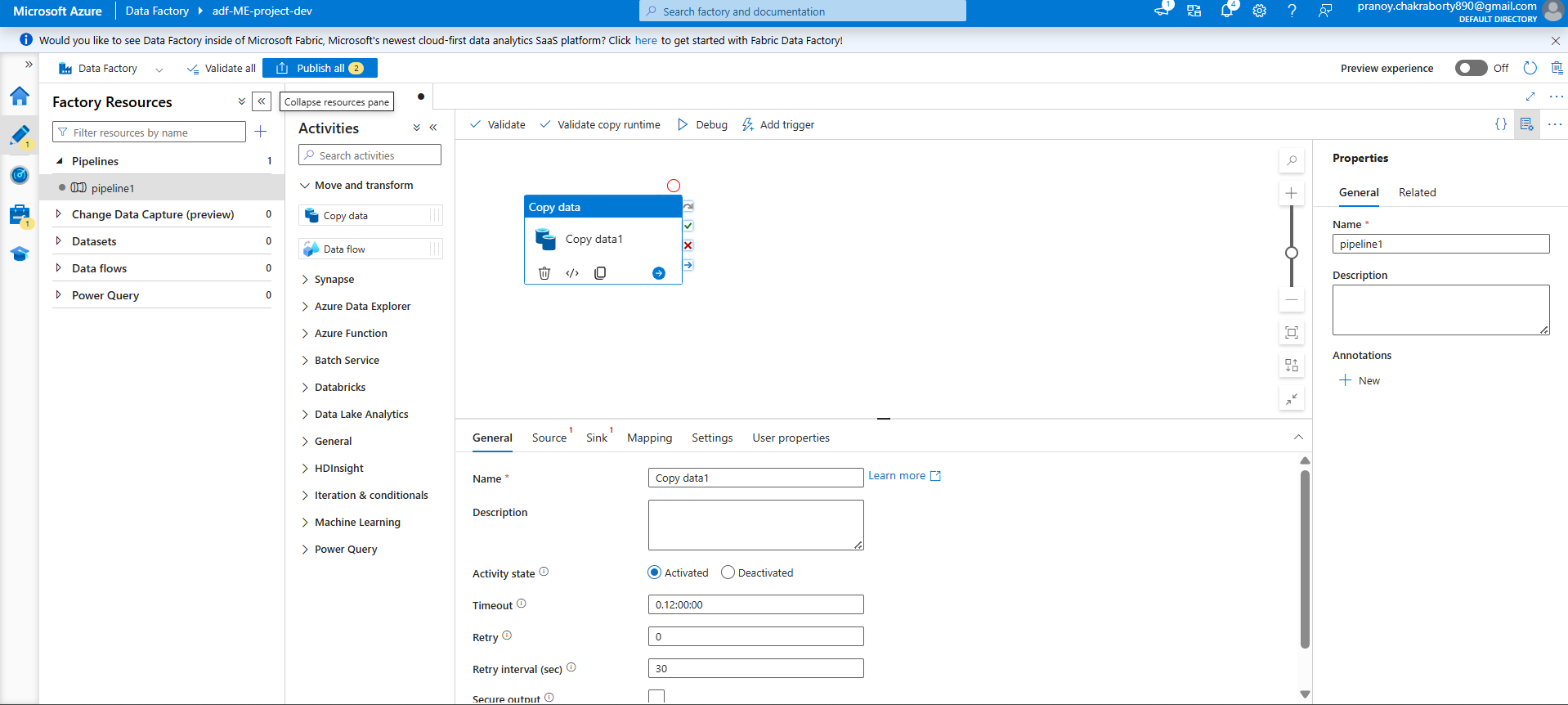
Create connection

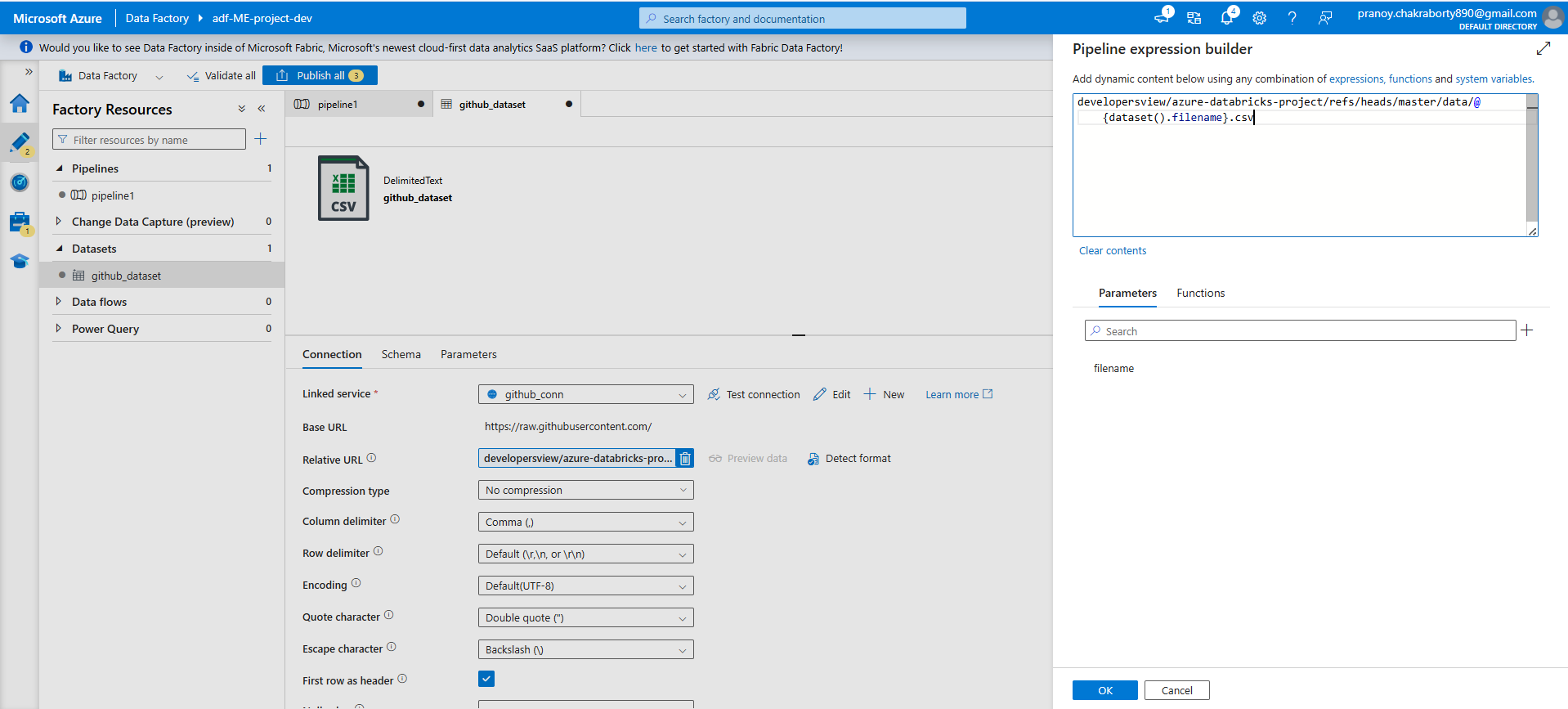
For github use http

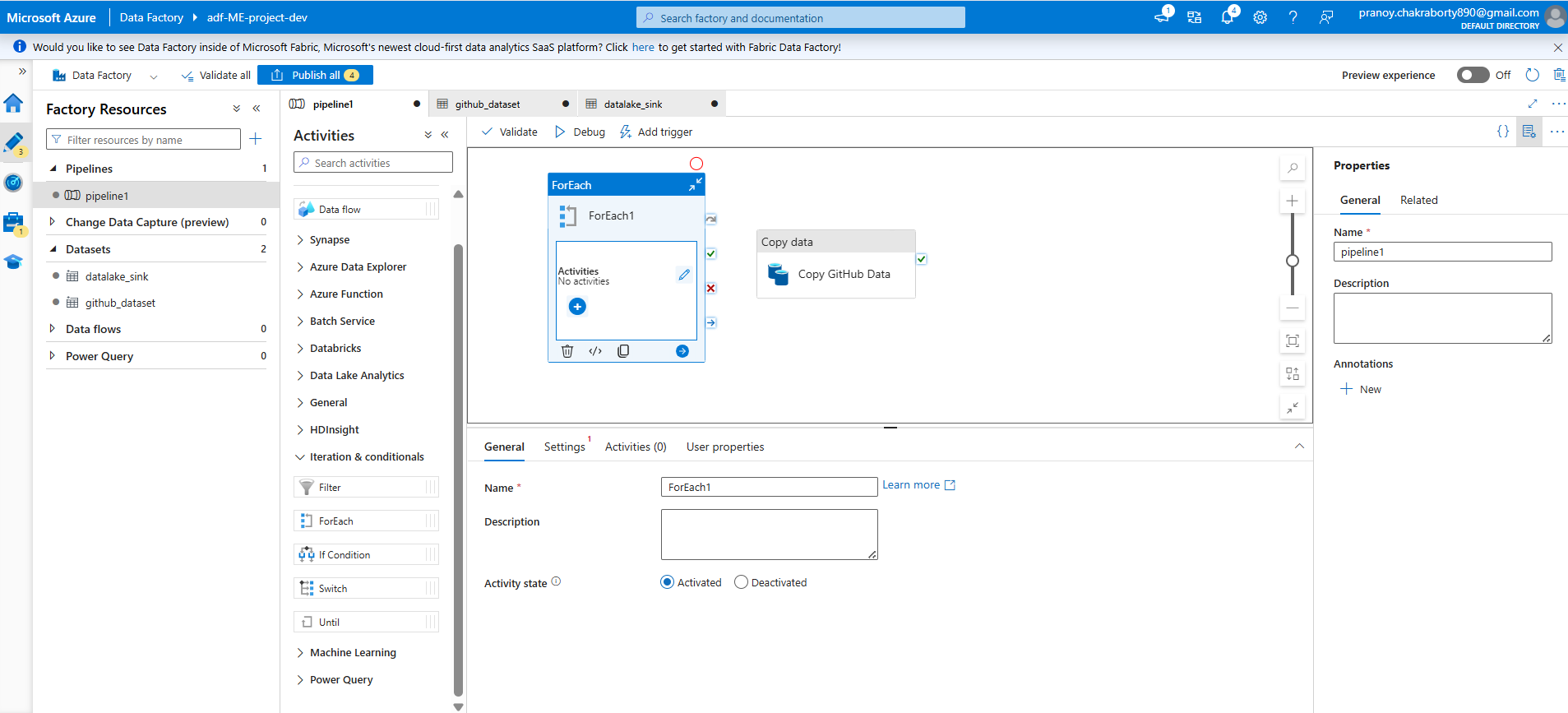


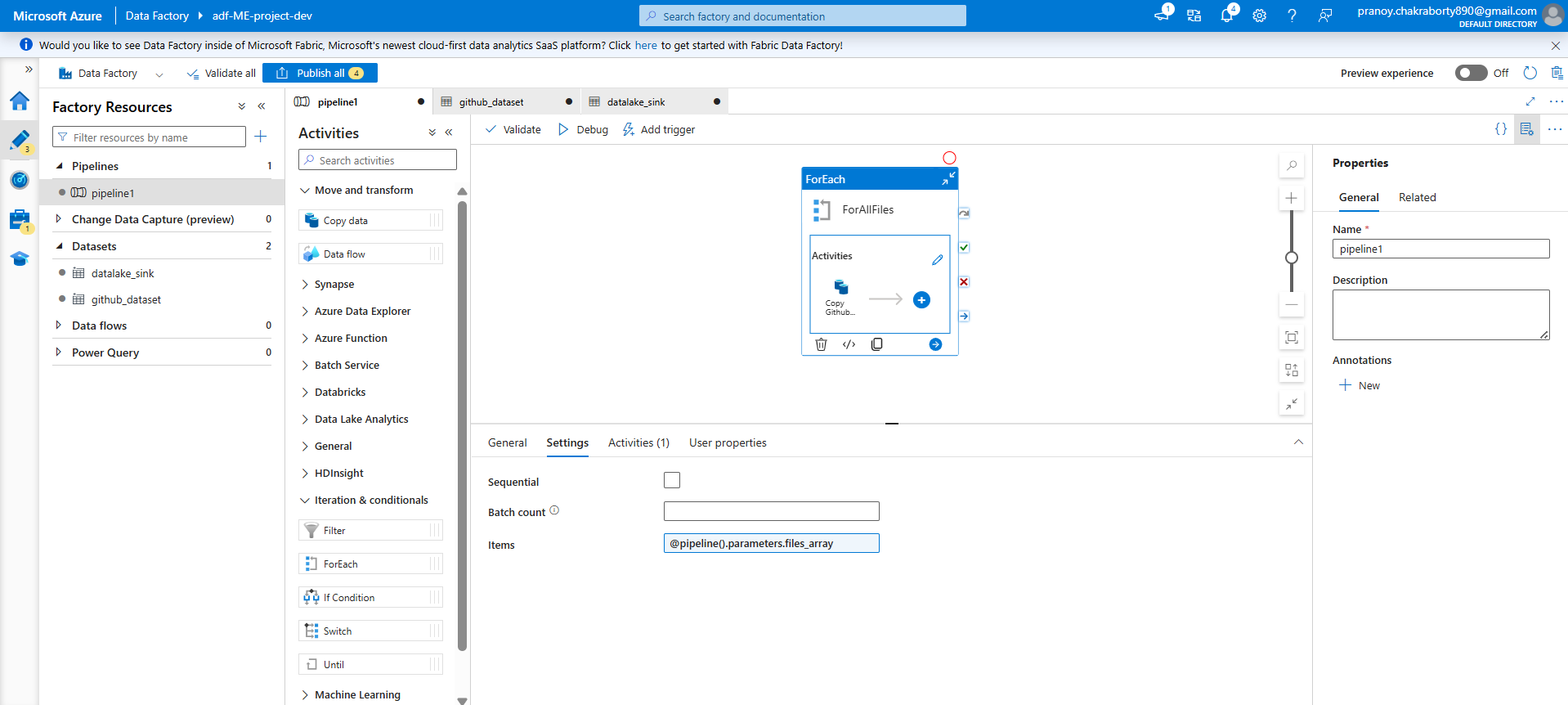
For data lake, use Azure Data Lake Storage Gen 2

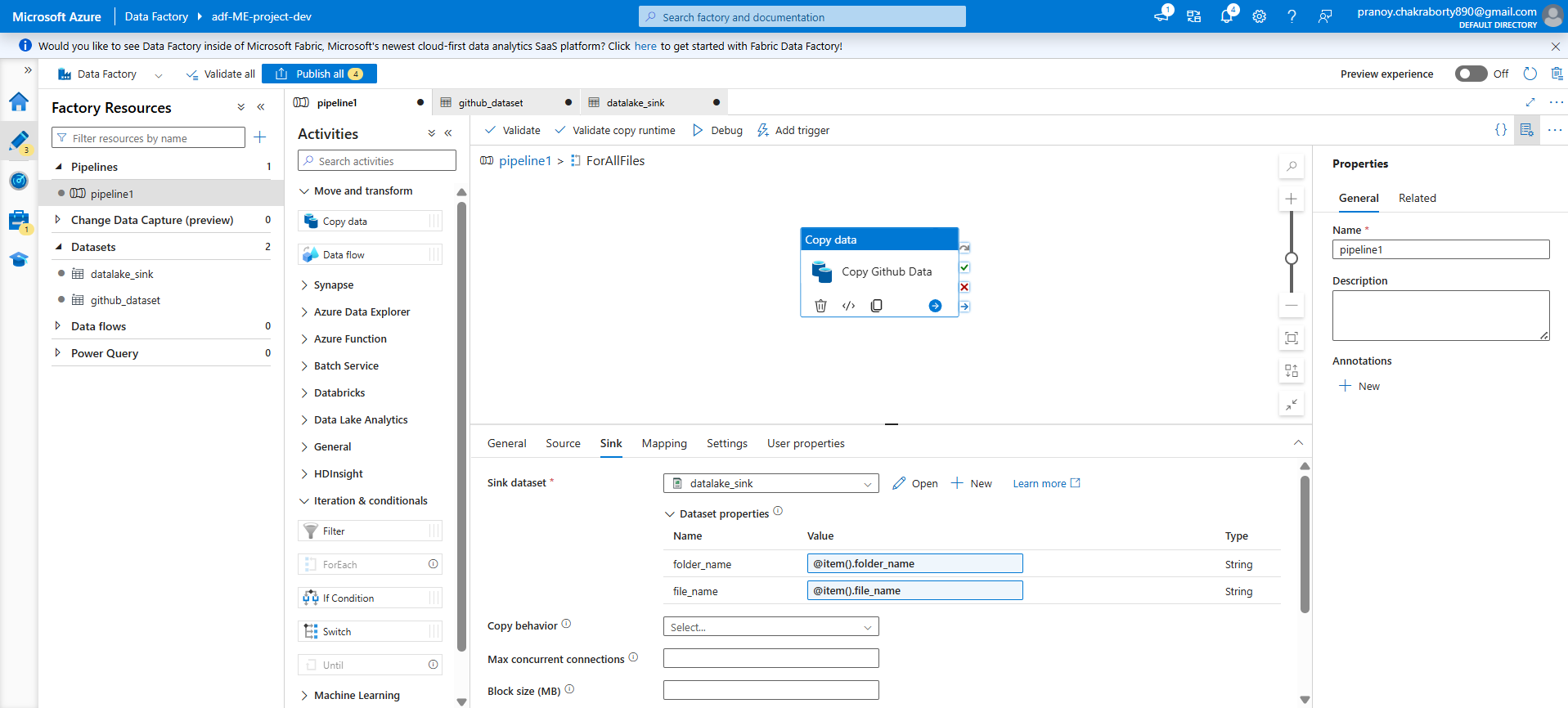


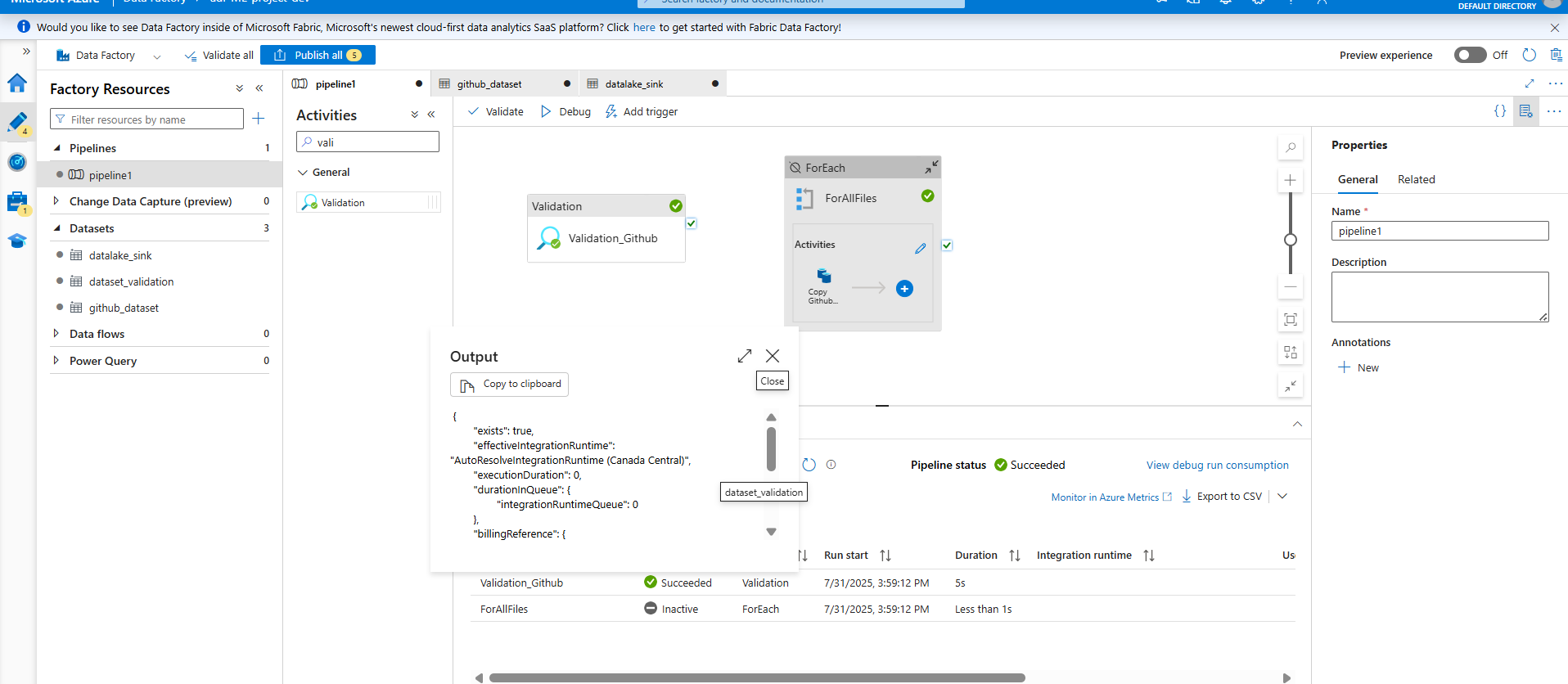


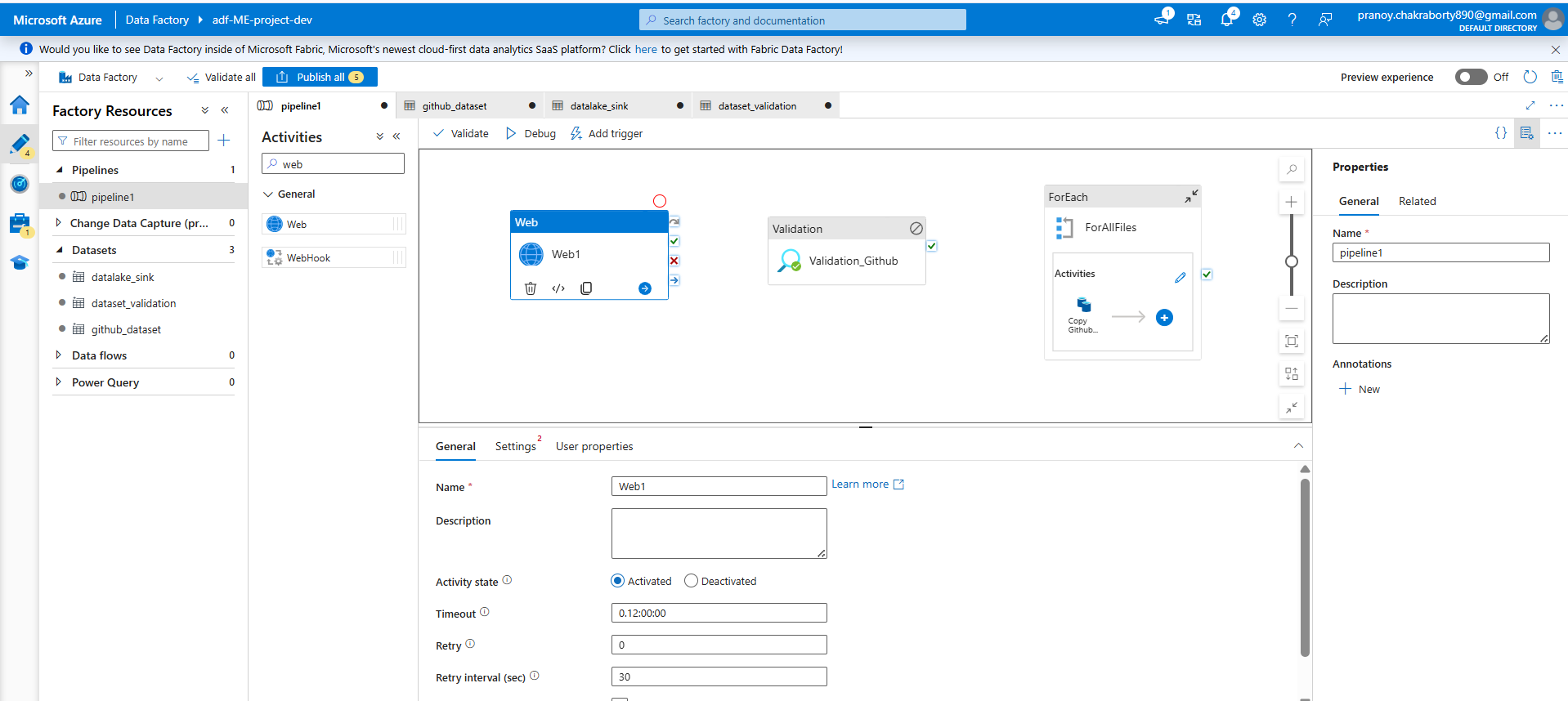


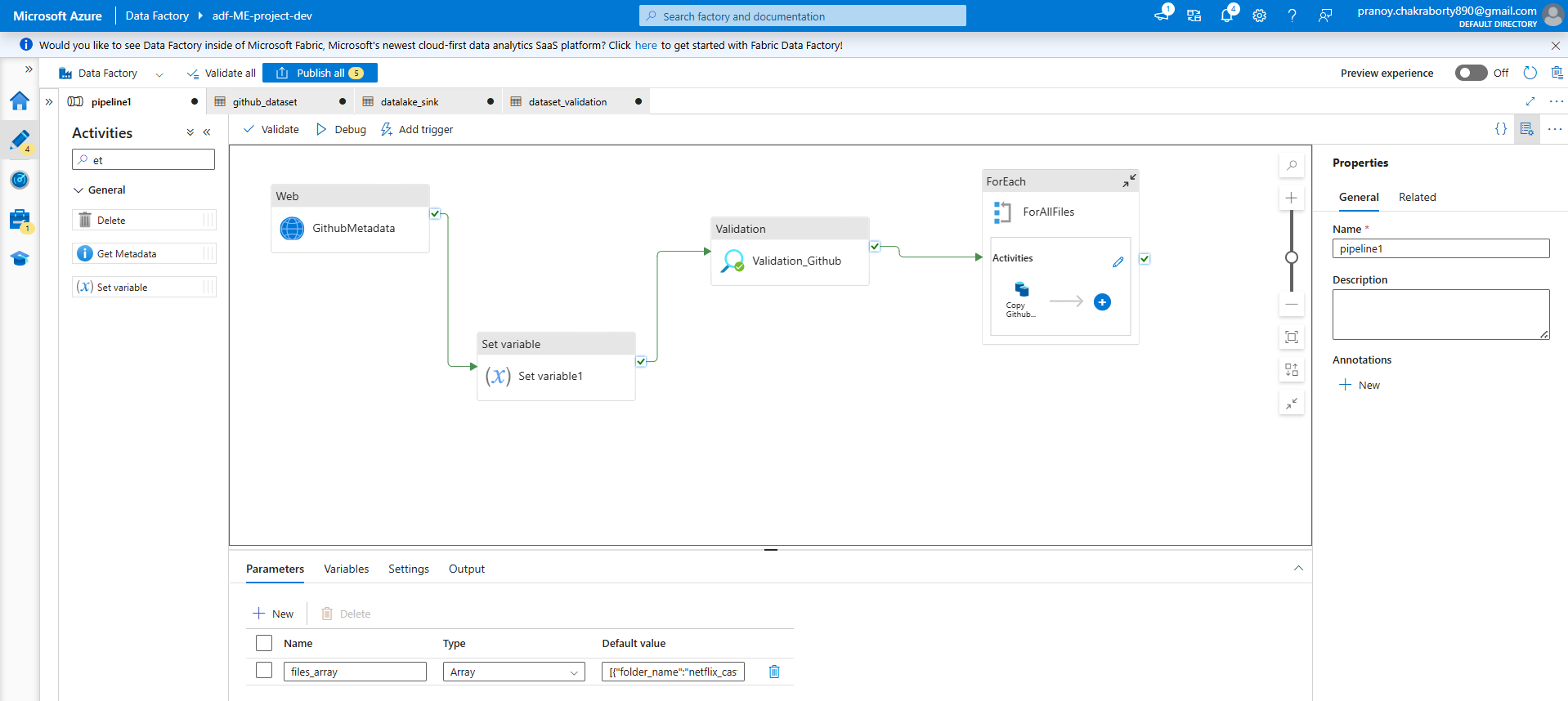








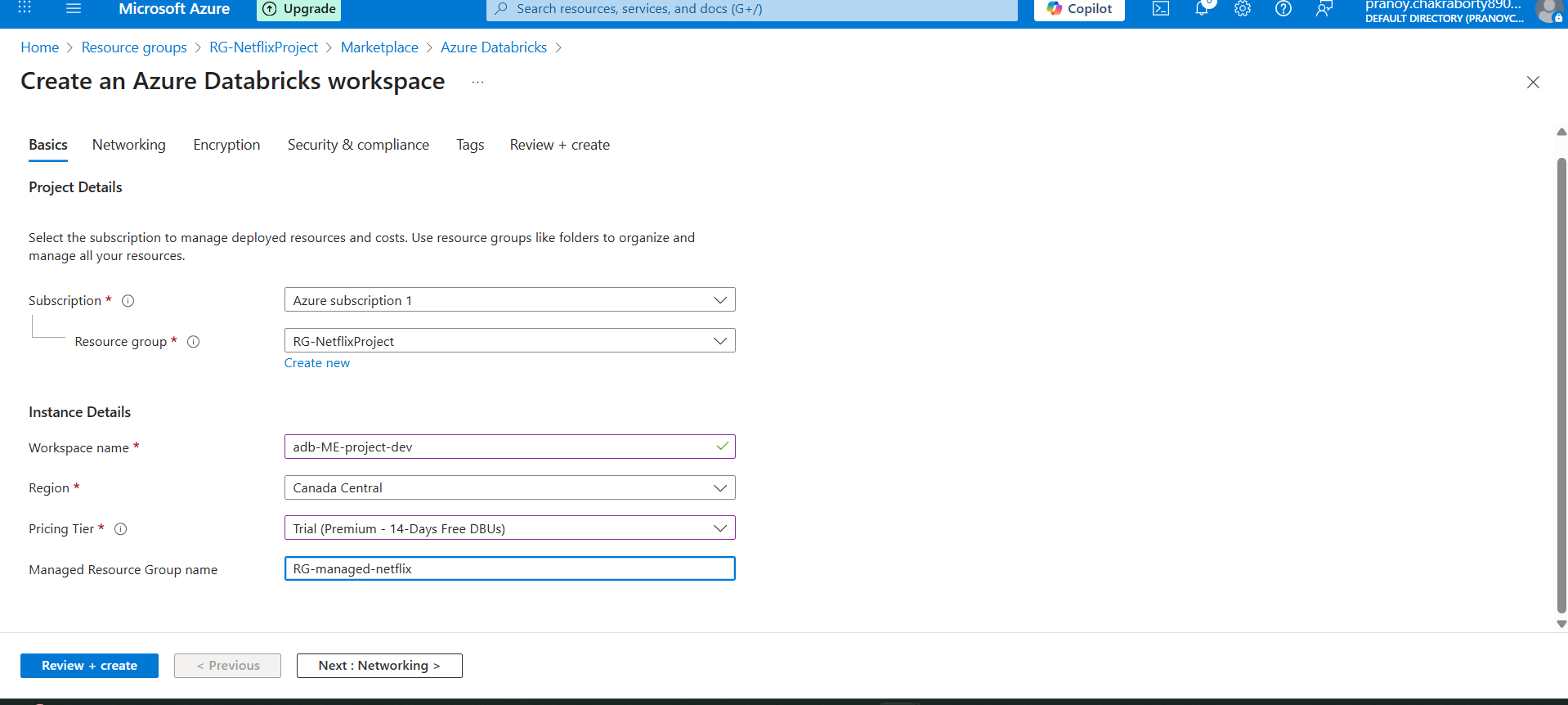
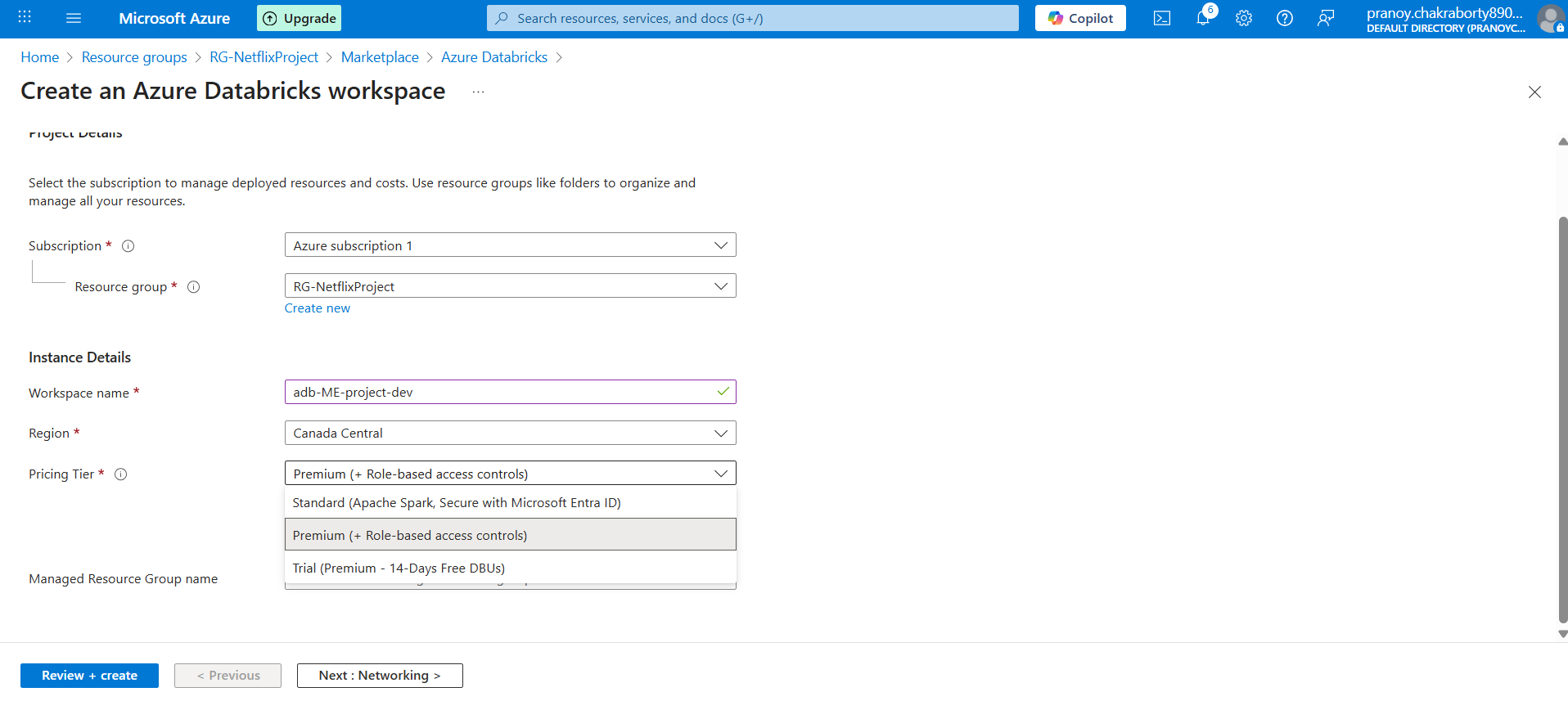


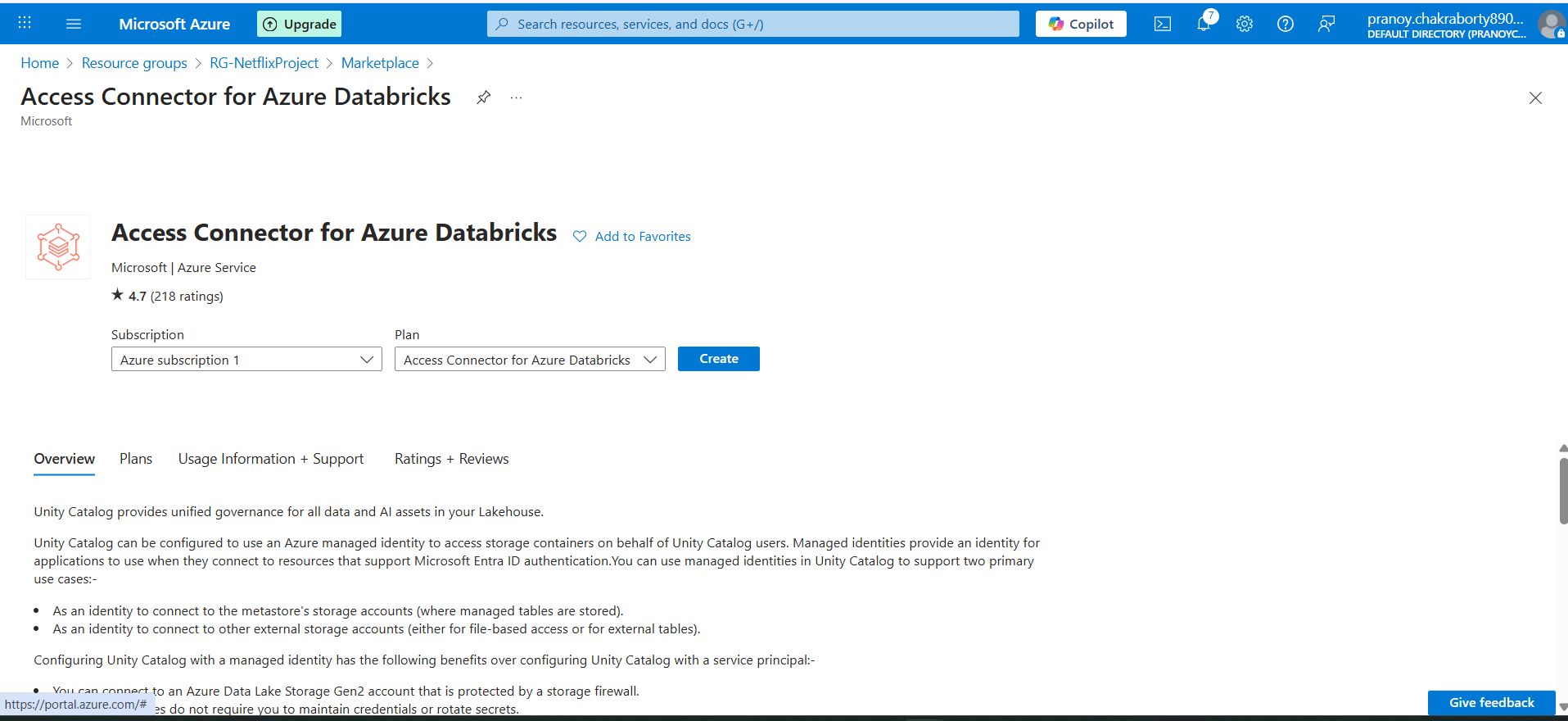


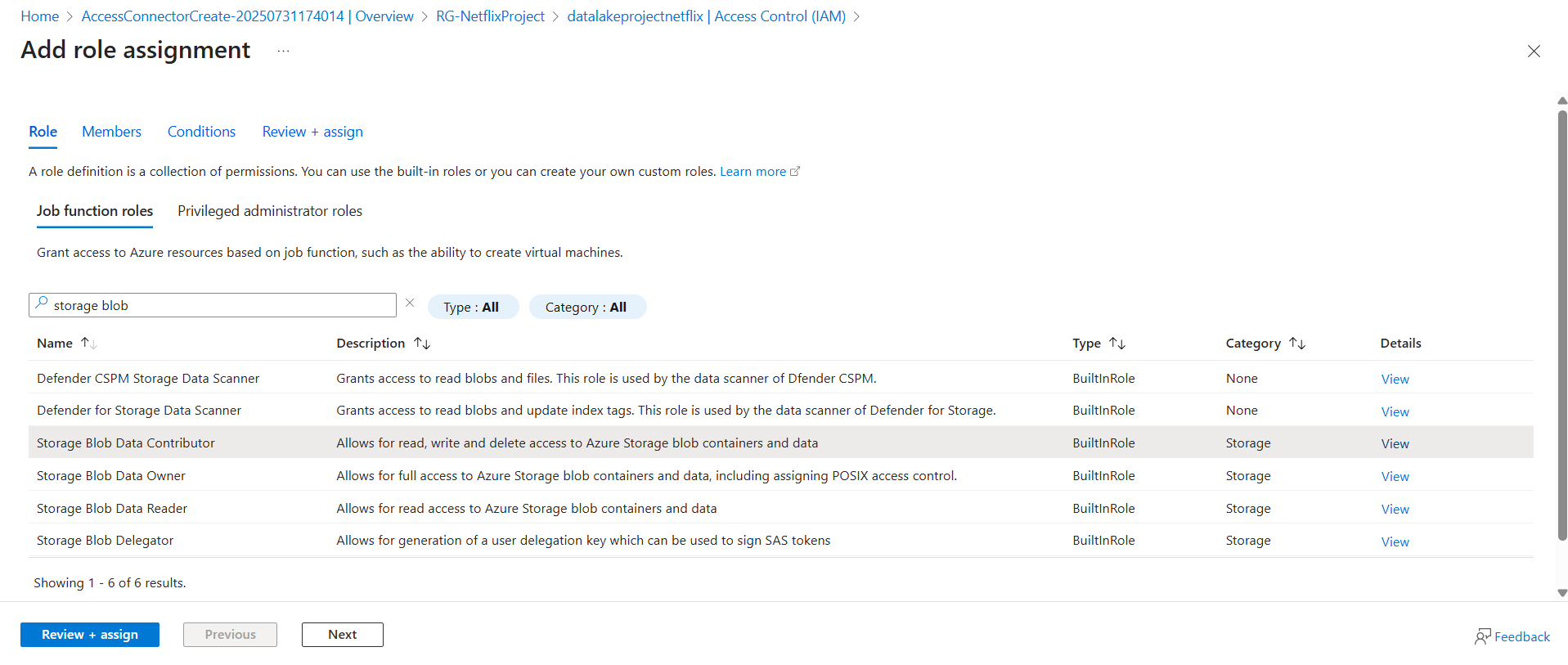
Files Array:

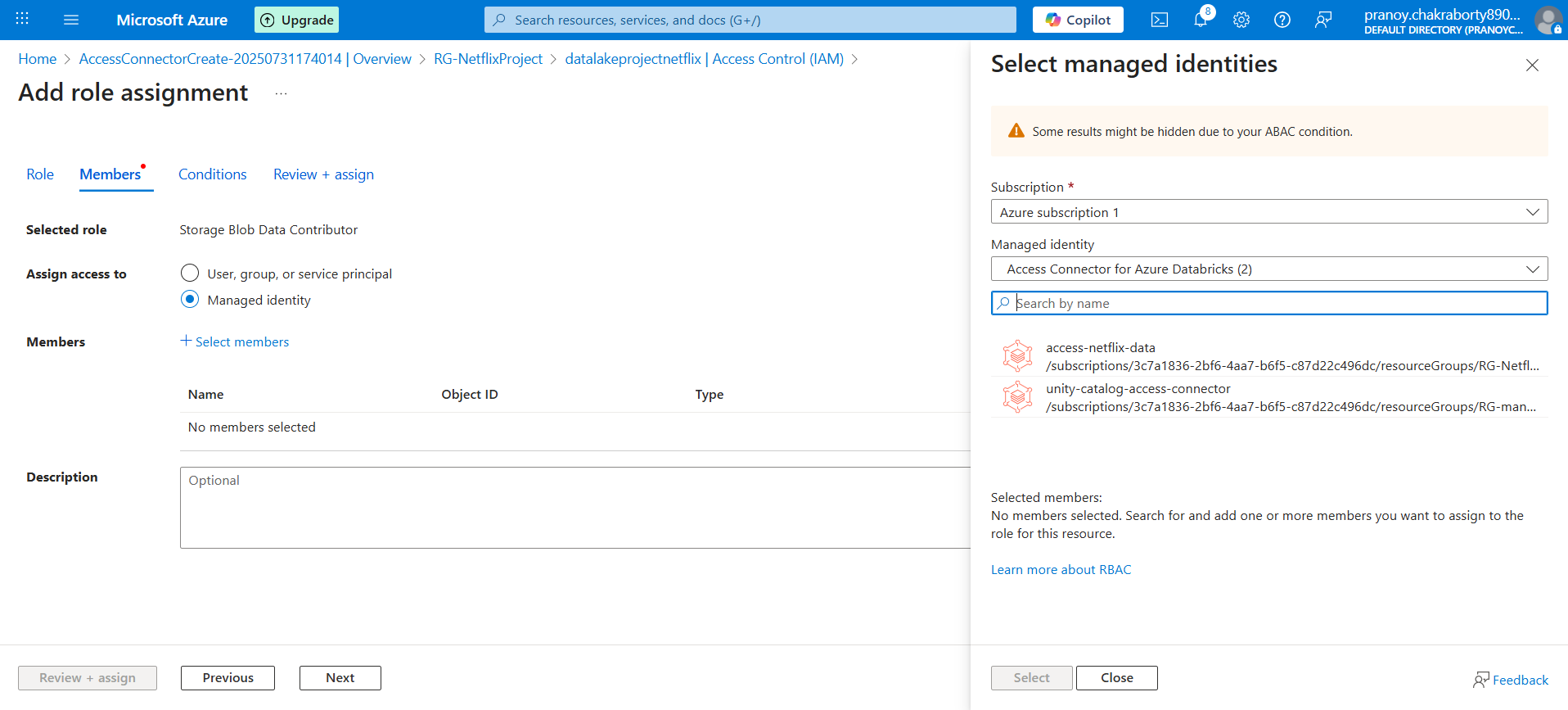
|  |
| --- |
| [  {  "folder\_name":"netflix\_cast",  "file\_name":"netflix\_cast.csv"  },  {  "folder\_name":"netflix\_category",  "file\_name":"netflix\_category.csv"  },  {  "folder\_name":"netflix\_countries",  "file\_name":"netflix\_countries.csv"  },  {  "folder\_name":"netflix\_directors",  "file\_name":"netflix\_directors.csv"  }  ] |

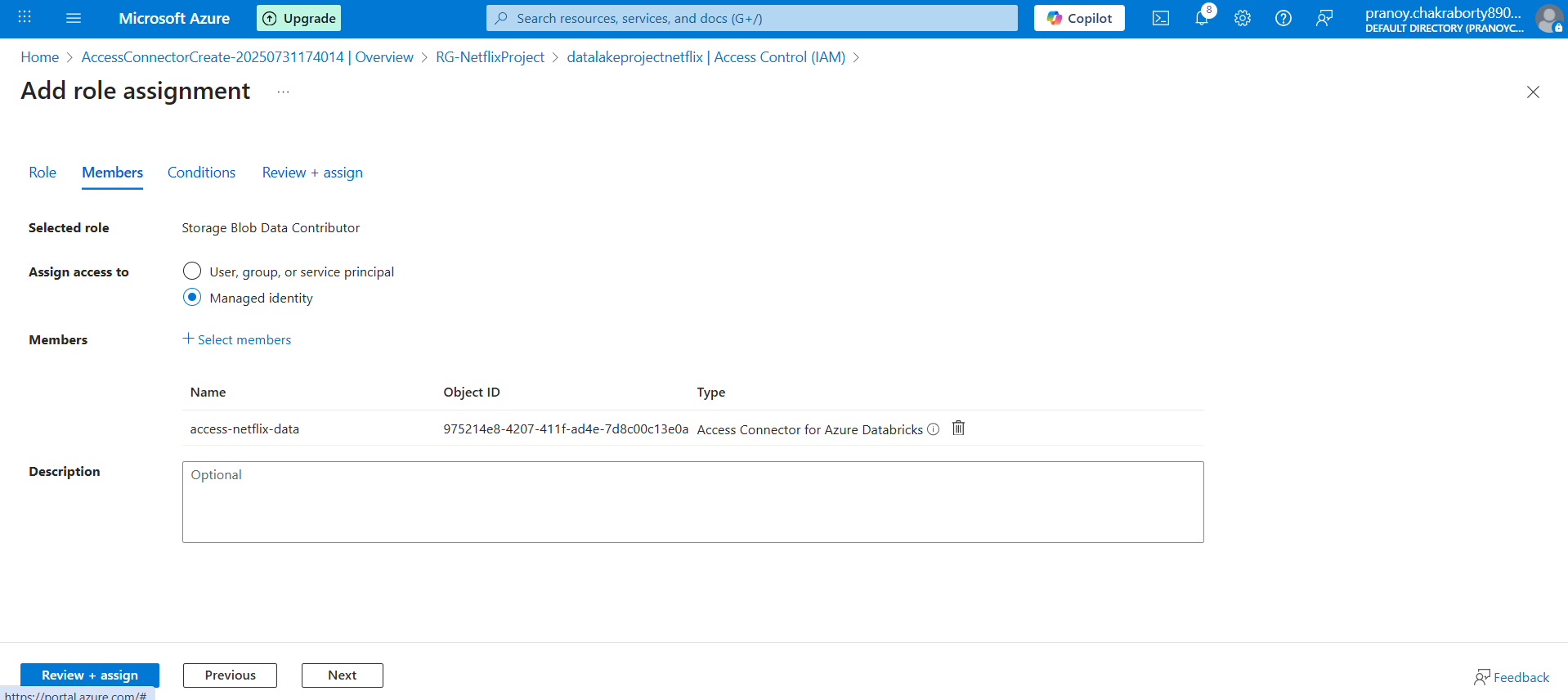
Azure Databricks

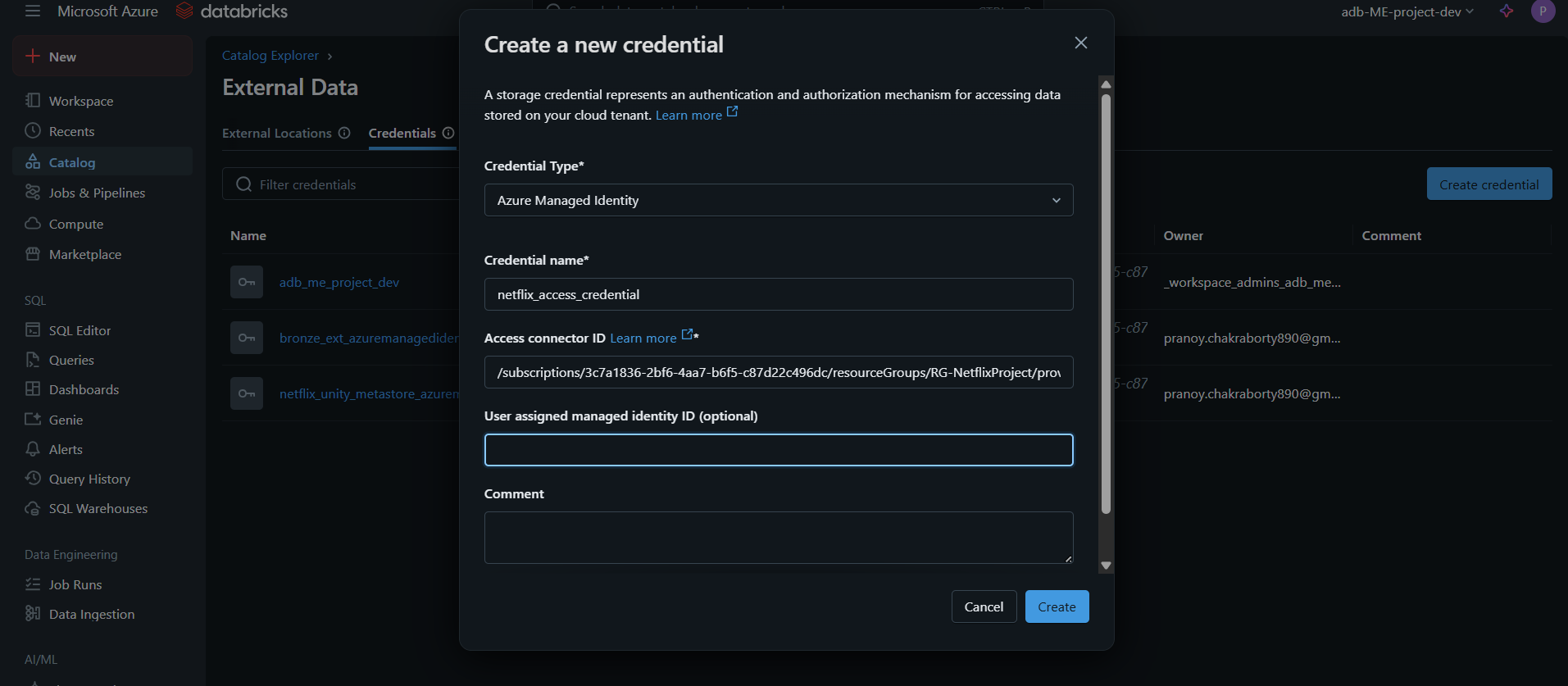


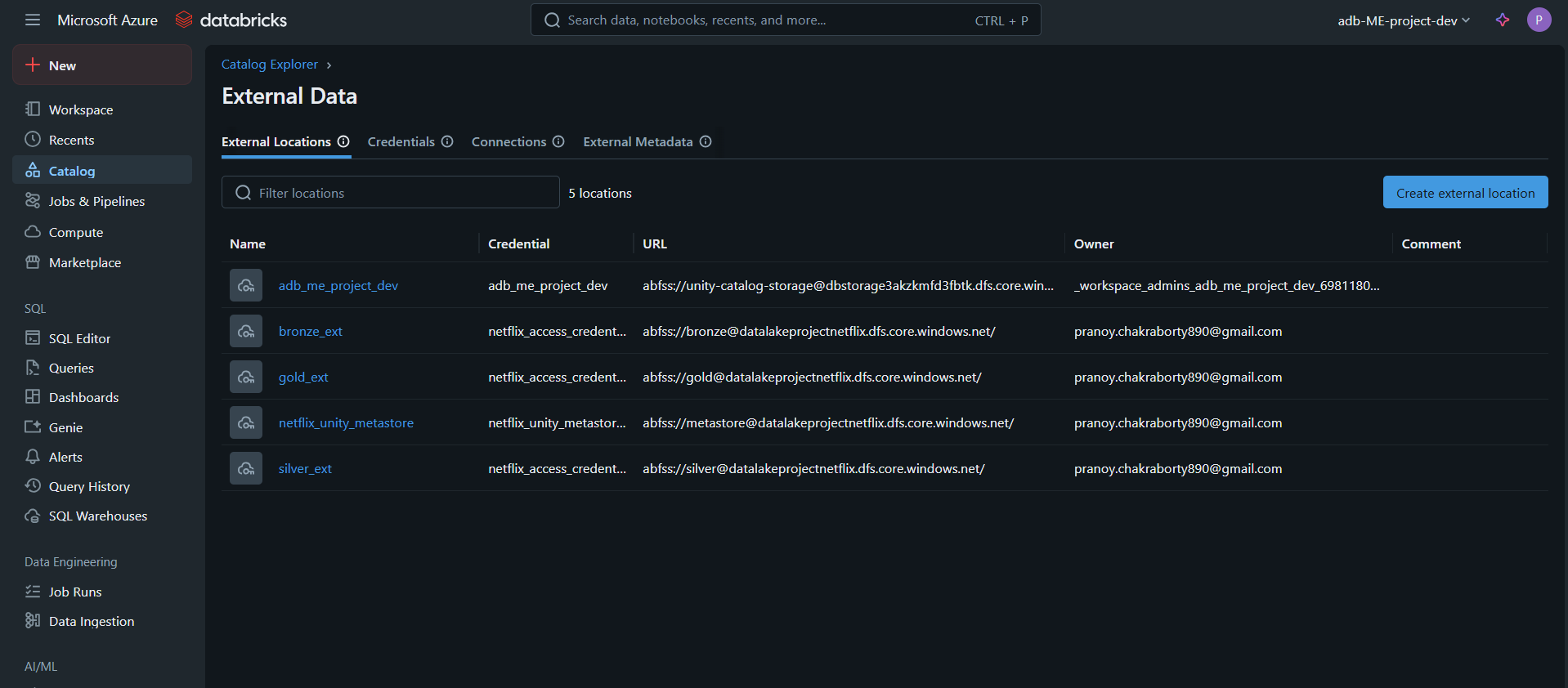




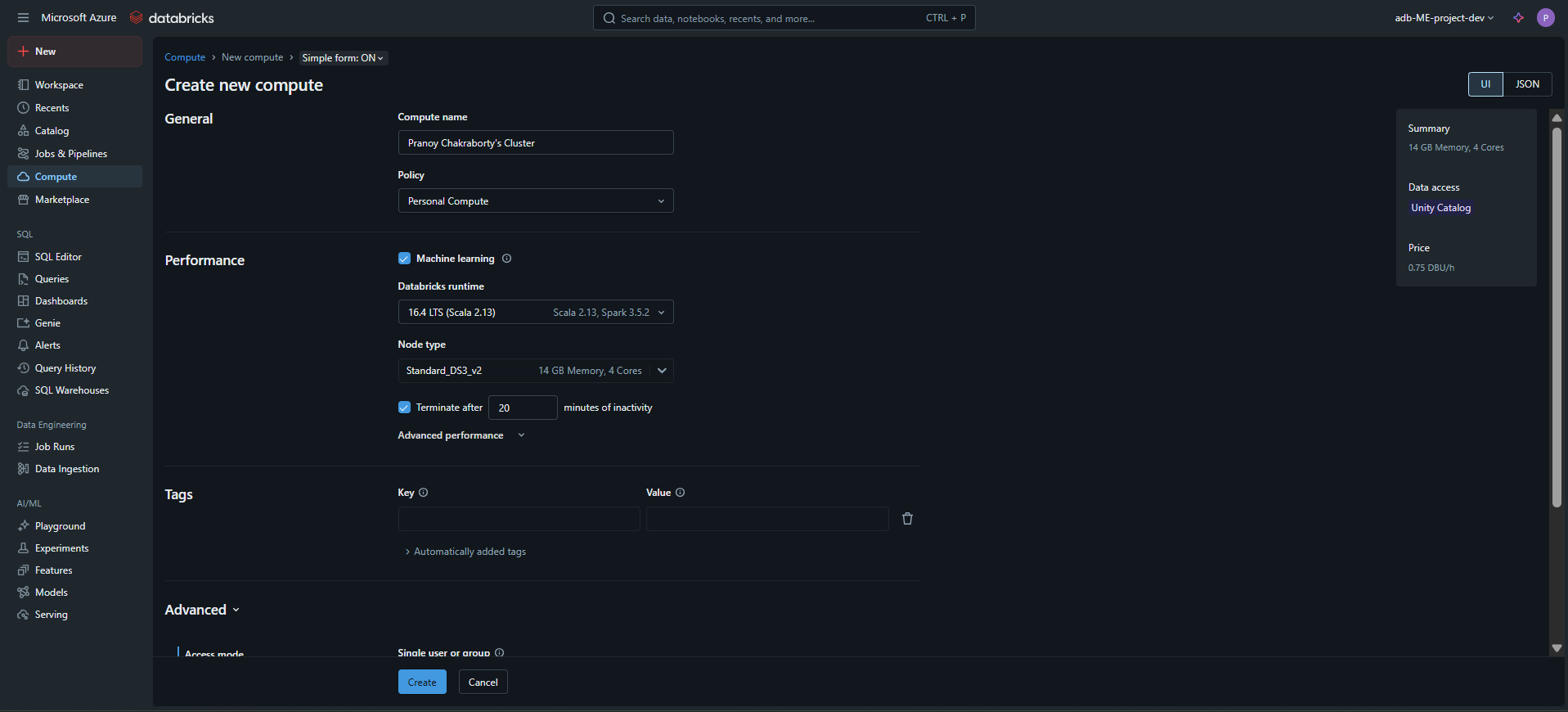








Create Compute:



Create Workspace, folder and notebook

Autoloader

Autoloader – Directory Listing, File Notifications

ADLS ----InferSchema (schema location) --- DB ------- ADLS (checkpoint location)

In file