Azure Databricks Coding Challenge

Name : Mitushi Vishwakarma

Question 3 :Execute & explain, Azure datafactory and its copy activity.

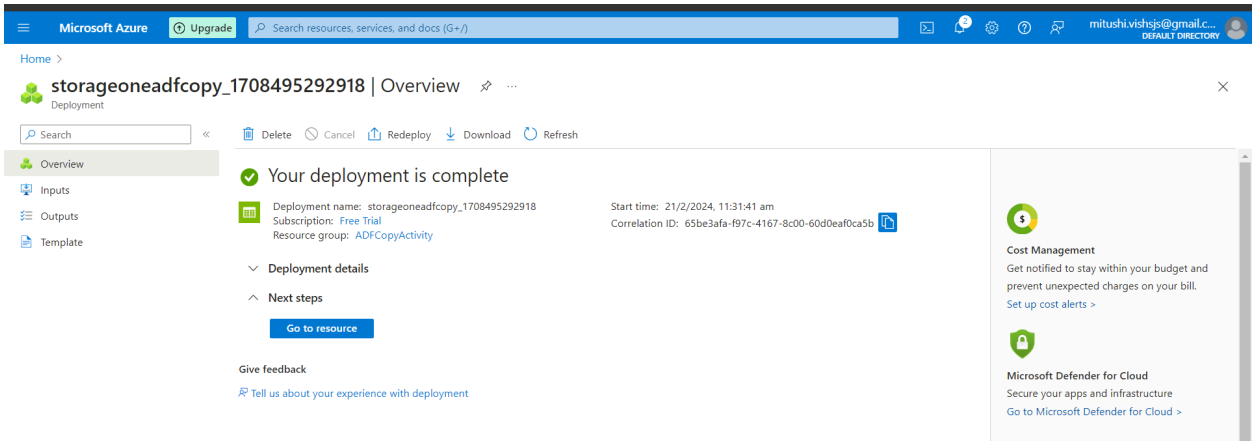
Azure Data Factory is a cloud-based data integration service that allows you to create data-driven workflows in the cloud for orchestrating and automating data movement and data transformation.

- ADF does not store any data itself.
- It allows you to create data-driven workflows to orchestrate the movement of data between supported data stores and then process the data using compute services in other regions or in an on-premise environment.
- It also allows you to monitor and manage workflows using both programmatic and UI mechanisms.

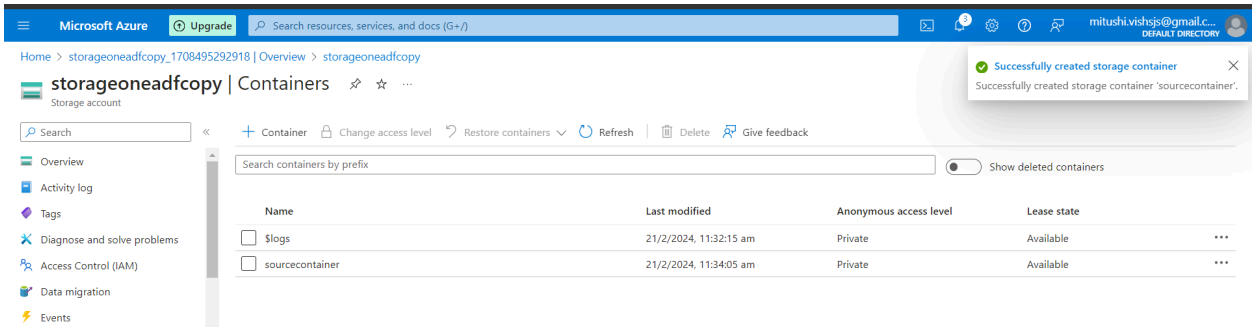
Copy Activity in Azure Data Factory copies data from a source data store to a sink data store.
Azure supports various data stores such as source or sink data stores like
Azure Blob storage,
Azure Cosmos DB (DocumentDB API),
Azure Data Lake Store,
Oracle,
Cassandra, etc.

Copy Activity : Creating two storage accounts “storageoneadfcopy” and “storagesecondadfcopy” and performing copy activity in data factory.

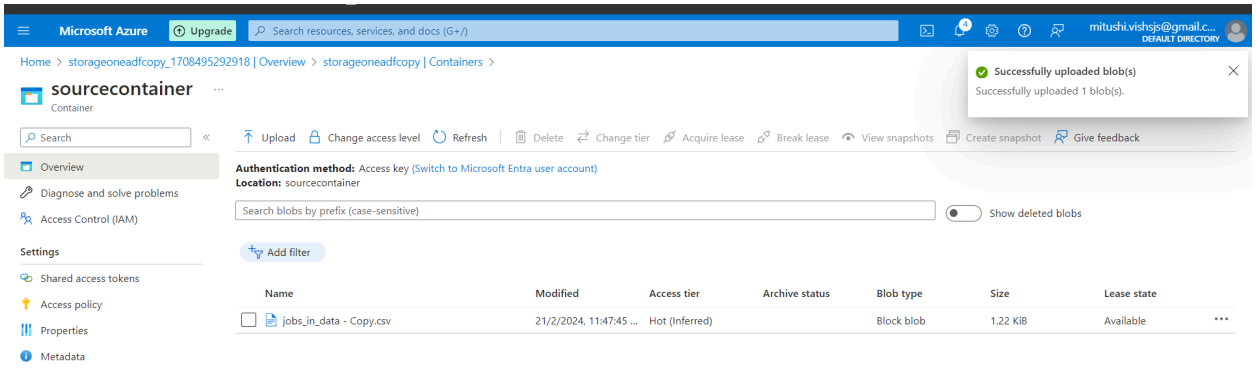
Created first blob storage account and container that will be the source store :



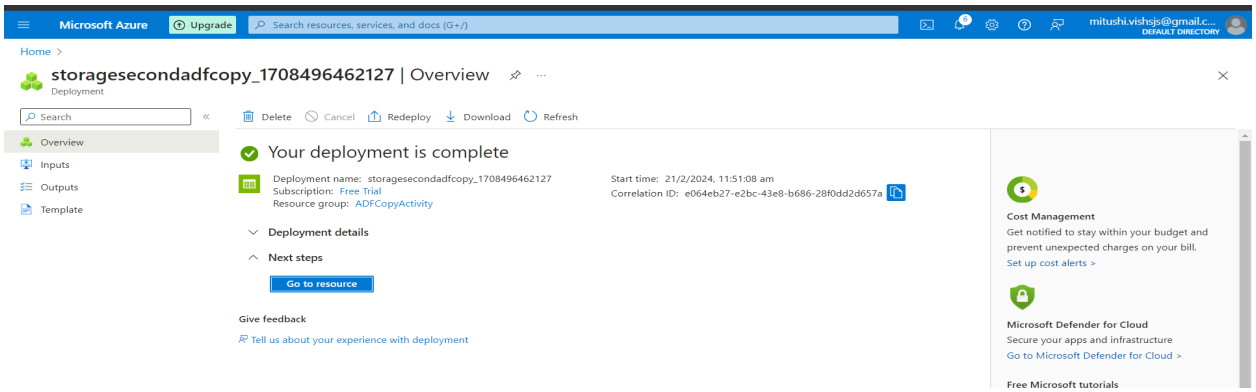
Created container in storage account :



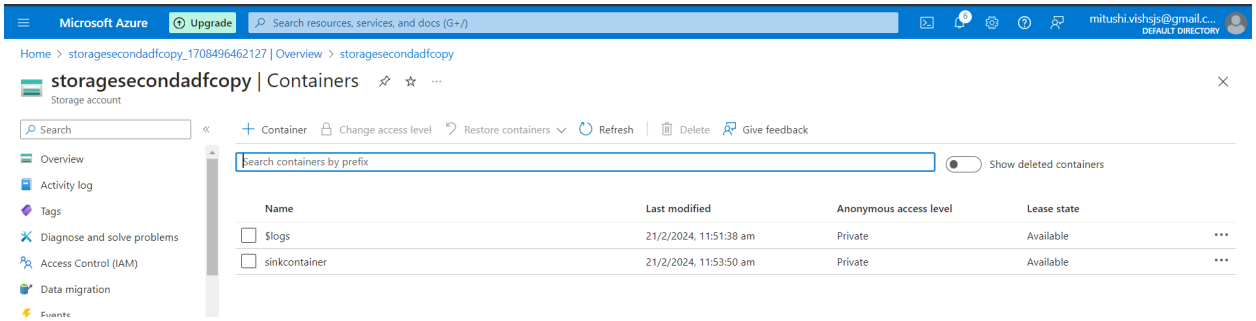
Uploaded file in container :



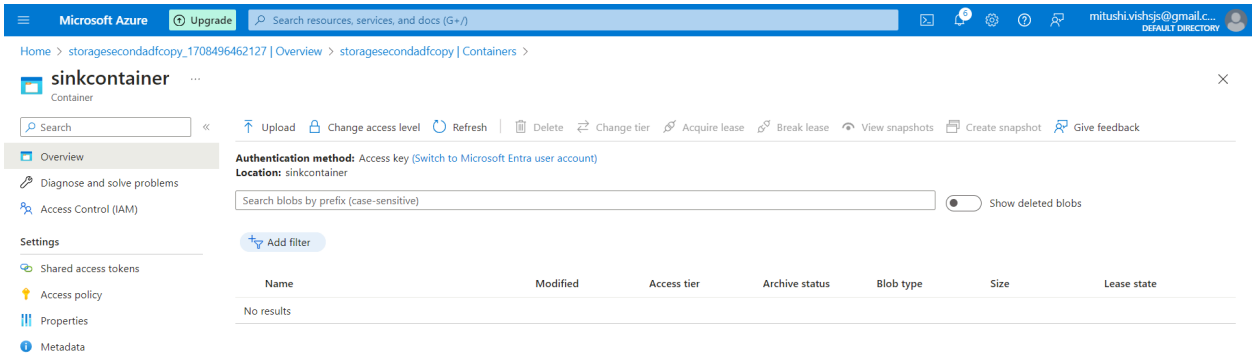
Created second blob storage account and container that will be the sink store :



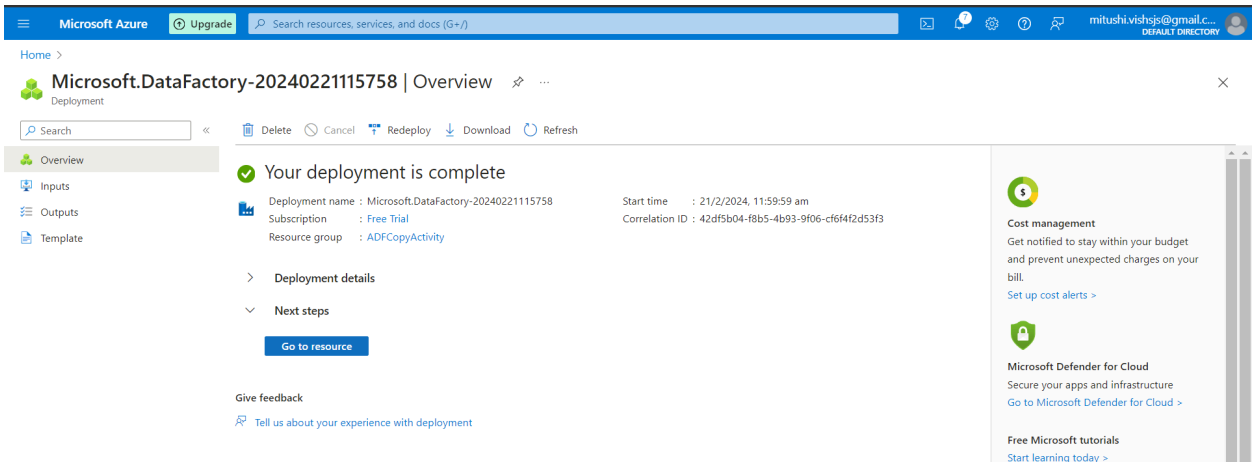
Created container in storage account :



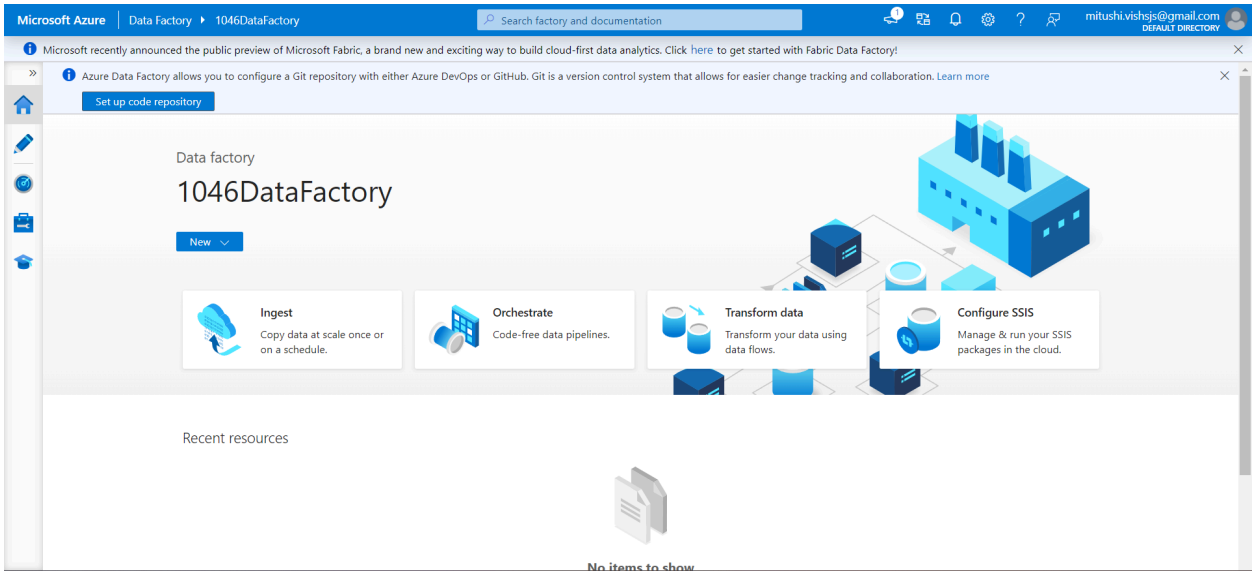
Empty Container :

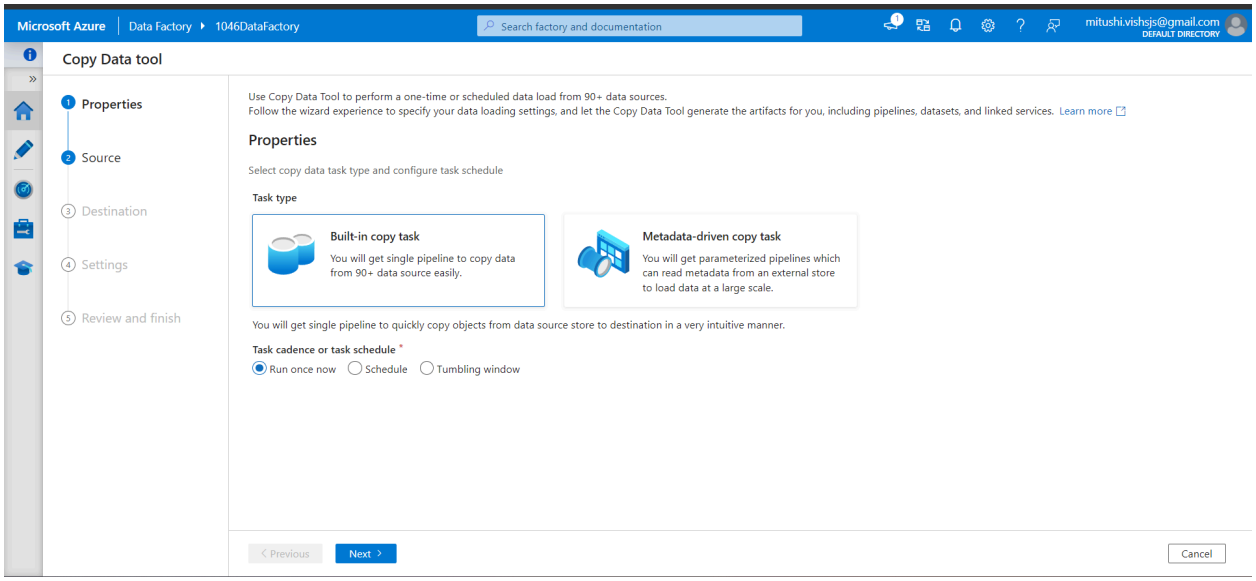


Created Data Factory named “1046DataFactory “ :

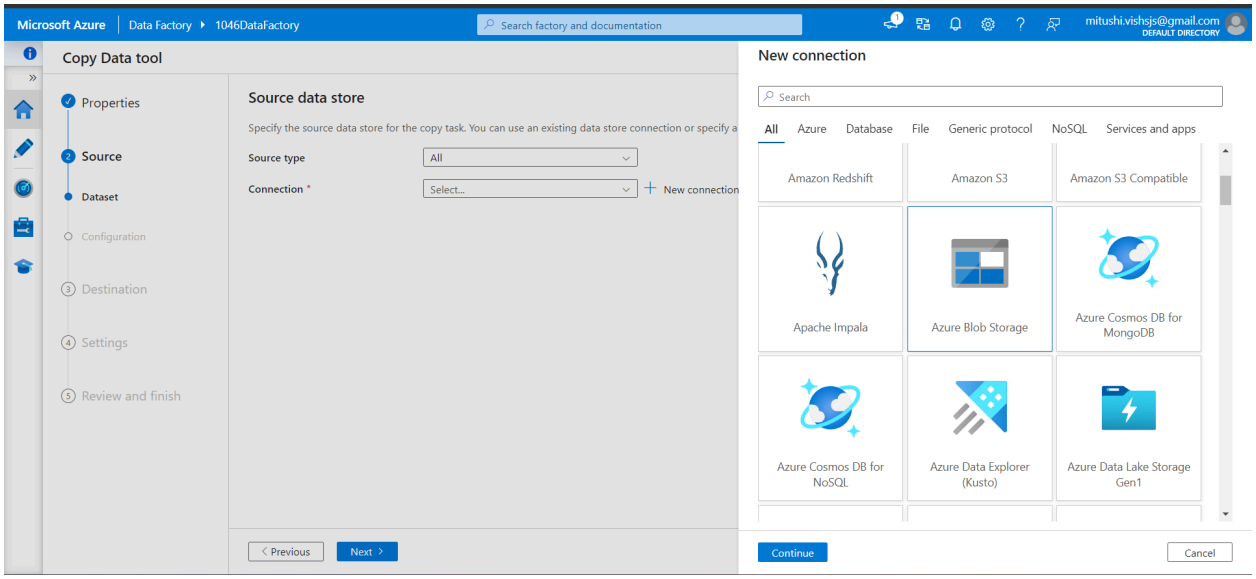


Selecting Ingest to copy data :

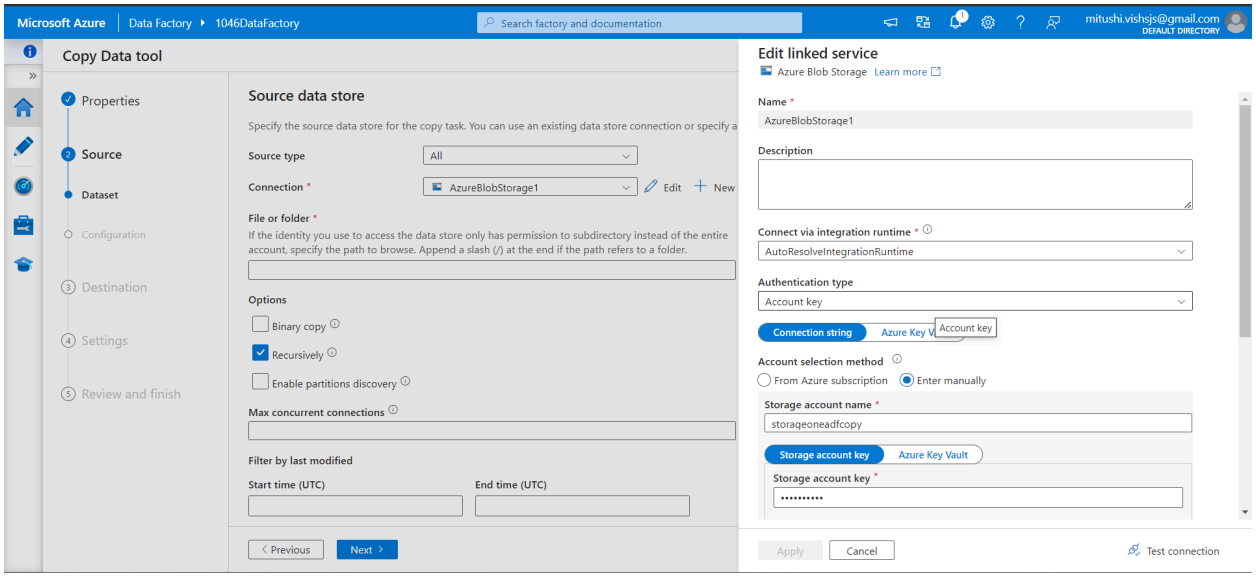




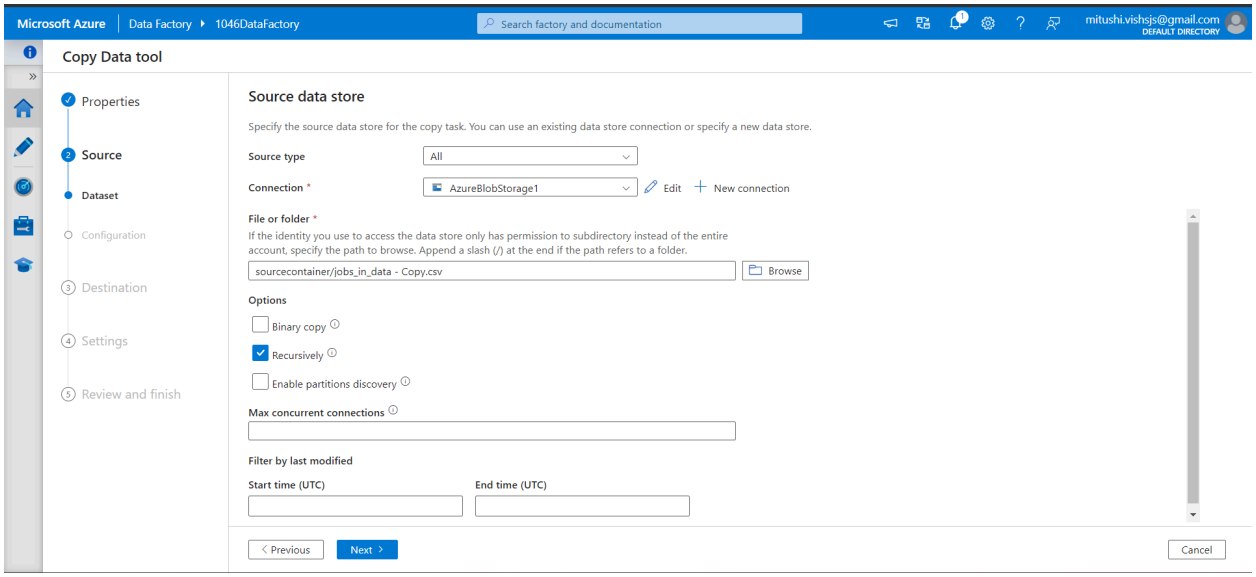
Selecting Source data store :



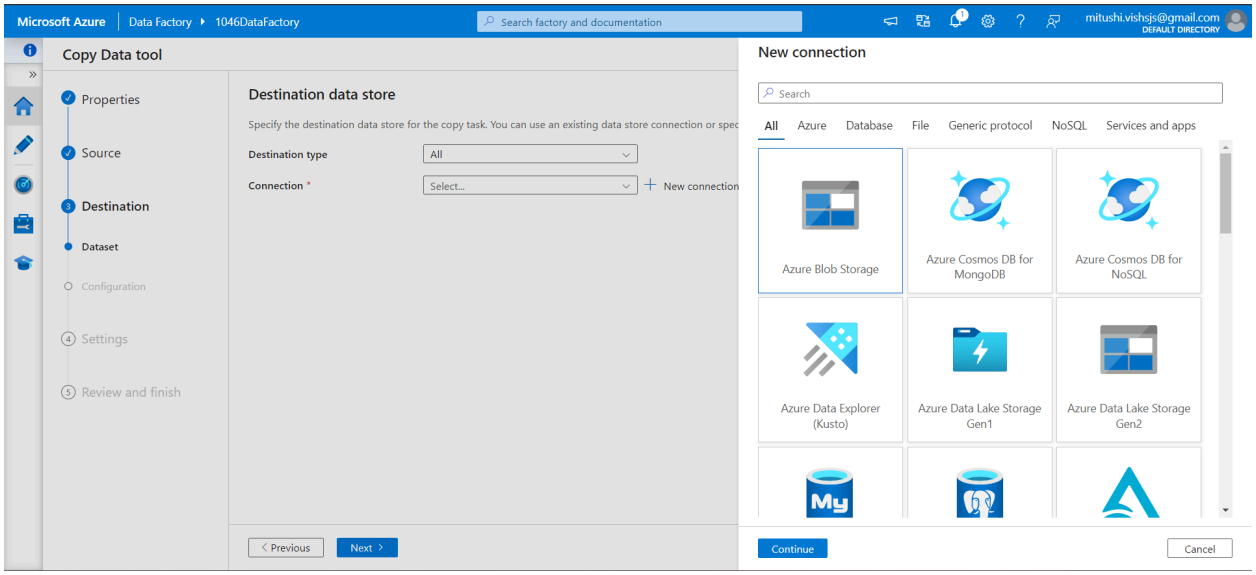
Linked first storage account :



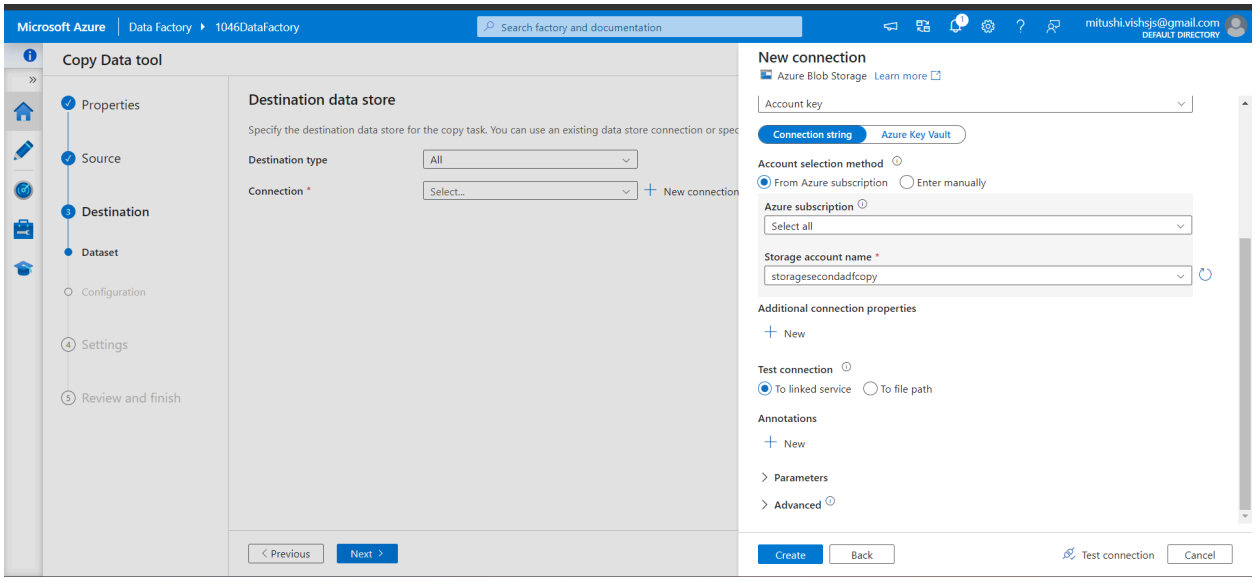
Provided the path for source data store files :



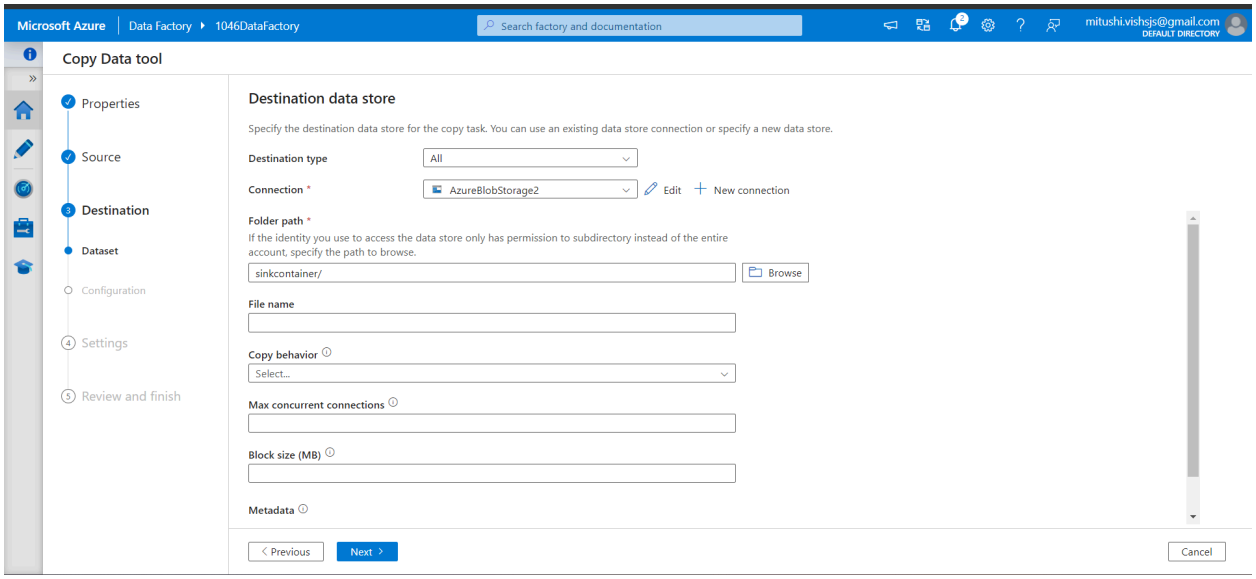
Creating Destination data store connection :



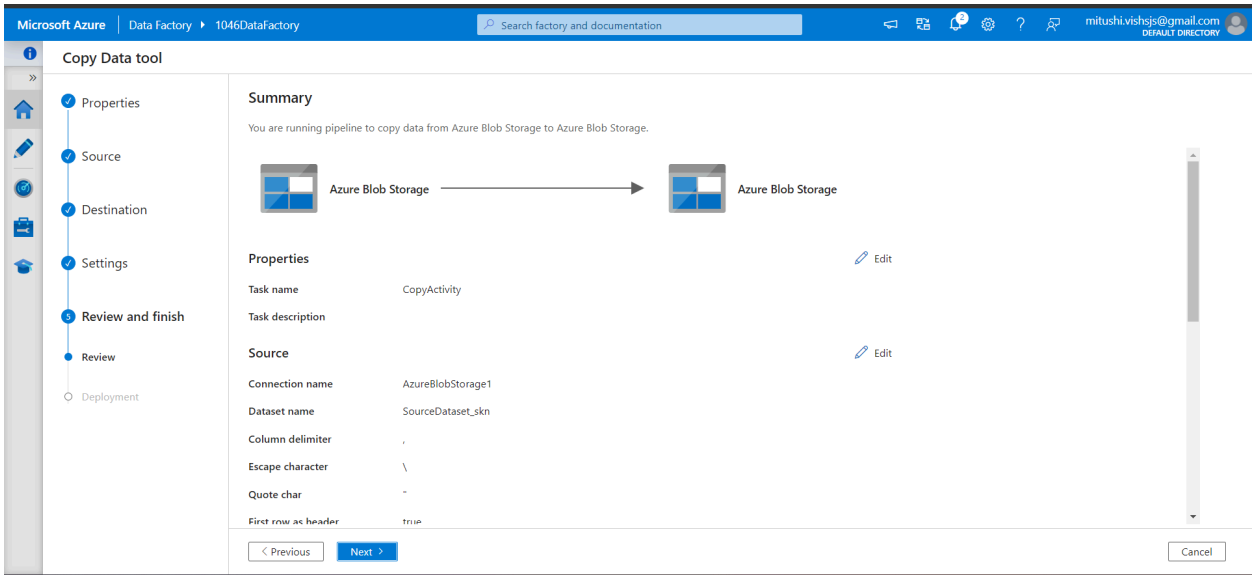
Linked second storage account in the destination connection :



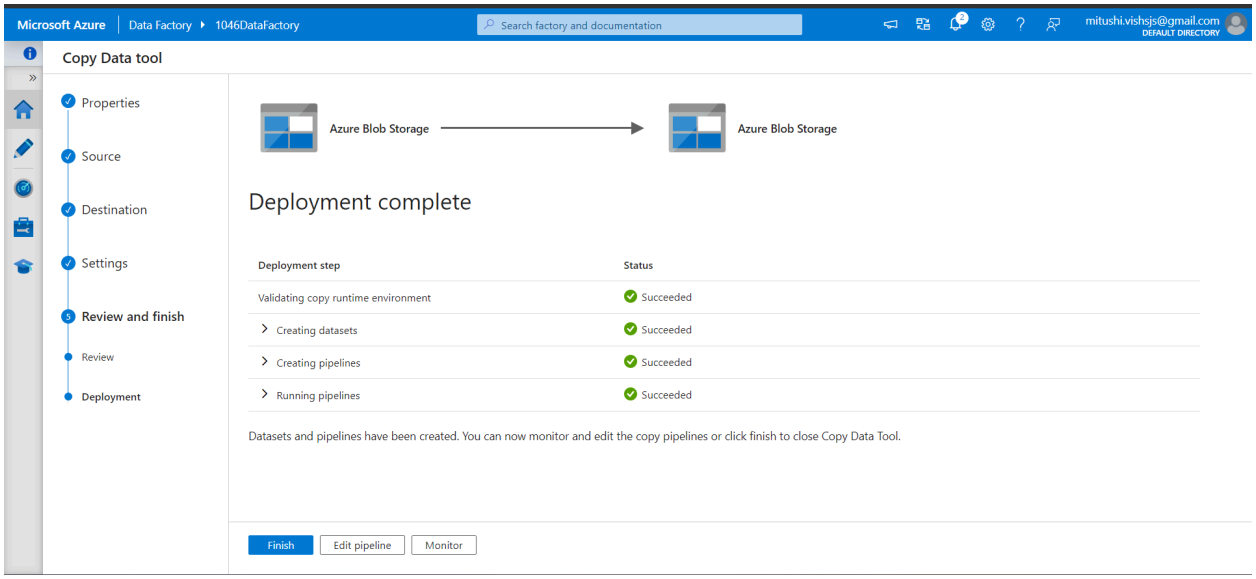
Provided the path where the files will be copied :



Review and submit :



Data Copied successfully :



Checking Copied data file :