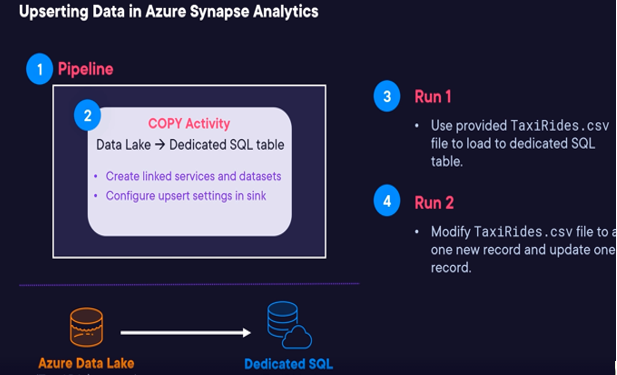
**Upserting Data in Azure Synapse Analytics**

**Introduction**

Upsert allows you to insert and update data in the target table as a single transaction. In this guide, we will demonstrate how to perform upsert operations using a Synapse pipeline in a dedicated SQL pool table. During the first run, we will load all the records into the table, and in the second run, we will perform insert and update operations as a single transaction.



**Solution**

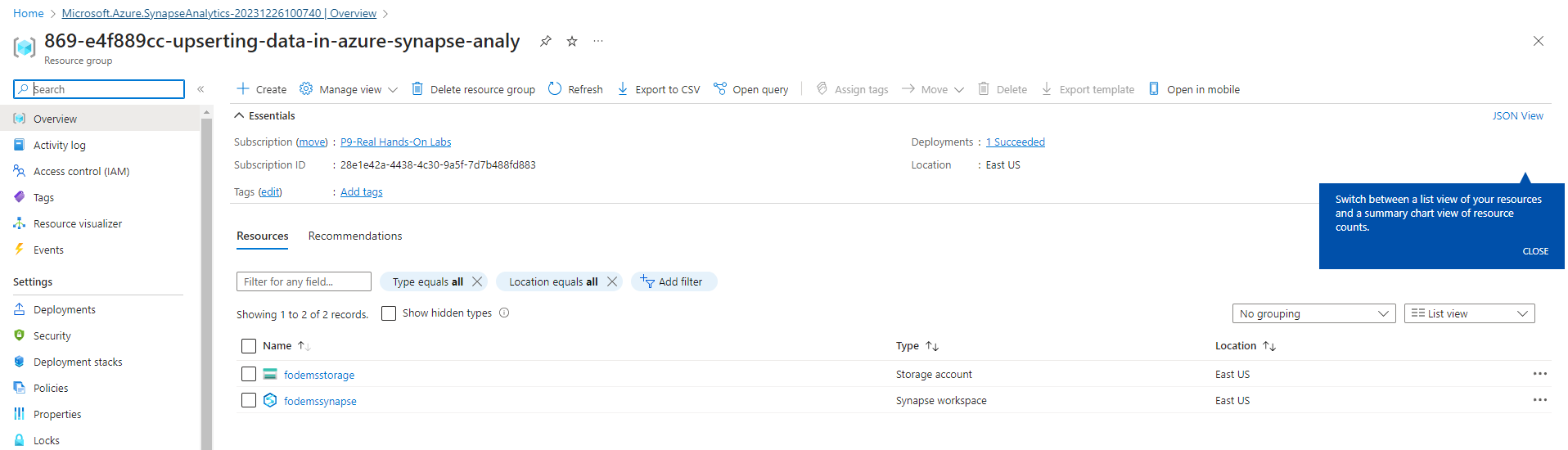
**Log in to the Azure Portal**

To get started, log in to the Azure portal using the credentials provided on the lab page. Ensure that you use an incognito or private browser window to guarantee you're using the lab account and not your personal one.

**Set Up the Environment**

**Note:** For this objective, ensure that you have downloaded the TaxiRides.csv file located in this GitHub repository.

1. Navigate to Azure Synapse Analytics.
2. In the upper left corner, click on +Create.
3. For the Resource group, select the existing resource group.
4. In the Workspace name field, enter acgmssynapseXXXX, replacing the Xs to make it unique.
5. Choose the Region as East US 2.
6. For the Select Data Lake Storage Gen2 > Account name, click Create new and enter the following:
   * Name: Enter acgmsstorage.
   * Click OK.
7. For File system name, click Create new and enter the following:
   * Name: Enter synapsecontainer.
   * Click OK.
8. Click Next: Security.
9. For Authentication method, select Use both local and Azure Active Directory (Azure AD) authentication.
10. Set the following:
    * SQL Server admin login: Enter acgadmin.
    * SQL Password: Create a password.
    * Confirm password: Confirm the password.
11. Click Review + create.
12. Click Create.
    * Note: This may take a few minutes to deploy.



**Create a Container**

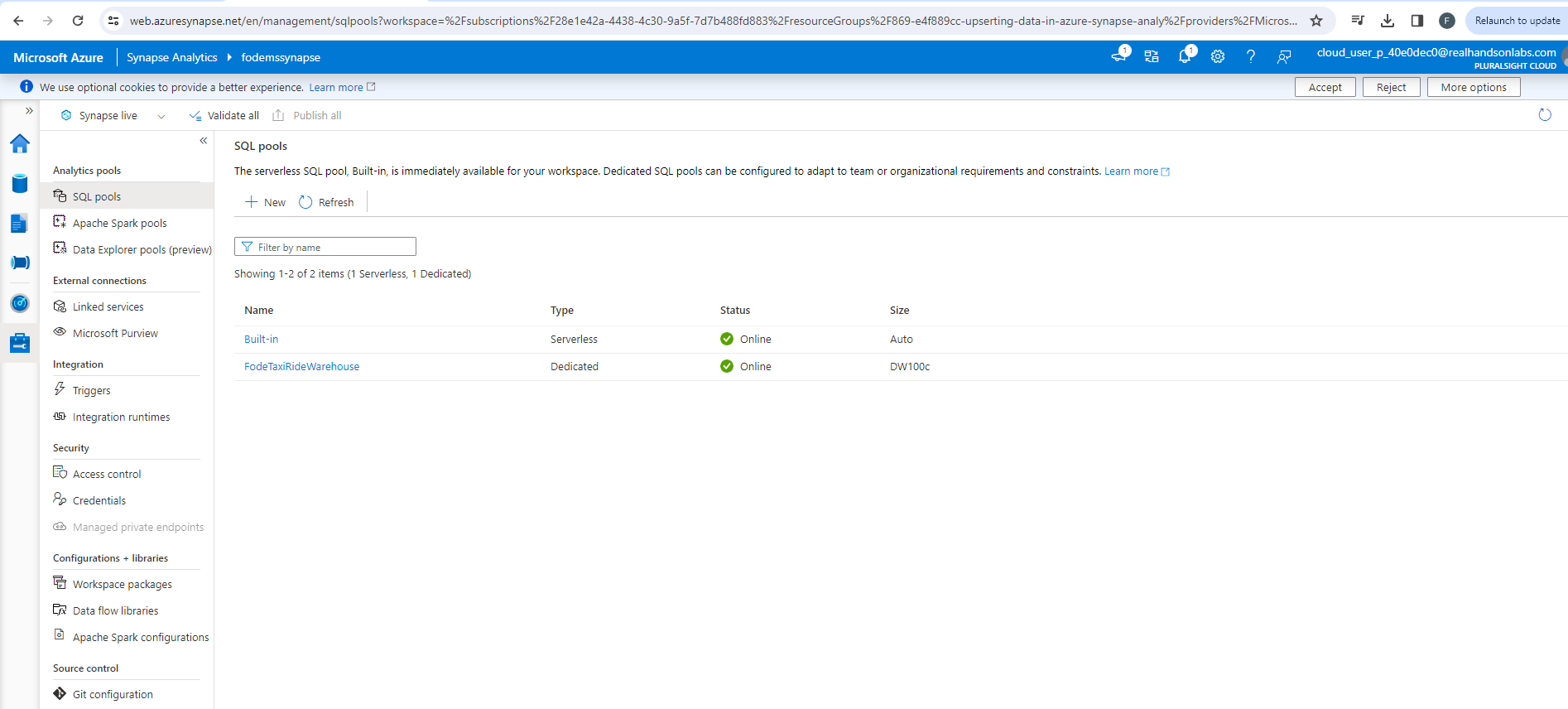
1. Once the instance is deployed, click on Go to resource group.
2. Click on the acgmsstorage storage account.
3. In the left navigation menu, under Data storage, click Containers.
4. Click + Container.
5. In the New container menu that appears on the right, enter taxidata for Name.
6. Click Create.
7. Click on the taxidata container.
8. In the toolbar, click + Add Directory.
9. In the Add Directory menu that appears on the right, enter Raw for Name.
10. Click Save.
11. Repeat steps 8-10, but this time, name the folder "Output."

**Upload a File**

1. Click on the Raw folder.
2. Click Upload.
3. Click Browse for files, and select the TaxiRides.csv file you downloaded earlier.
4. Click Upload.

**Set Up Dedicated SQL Pool Instance**

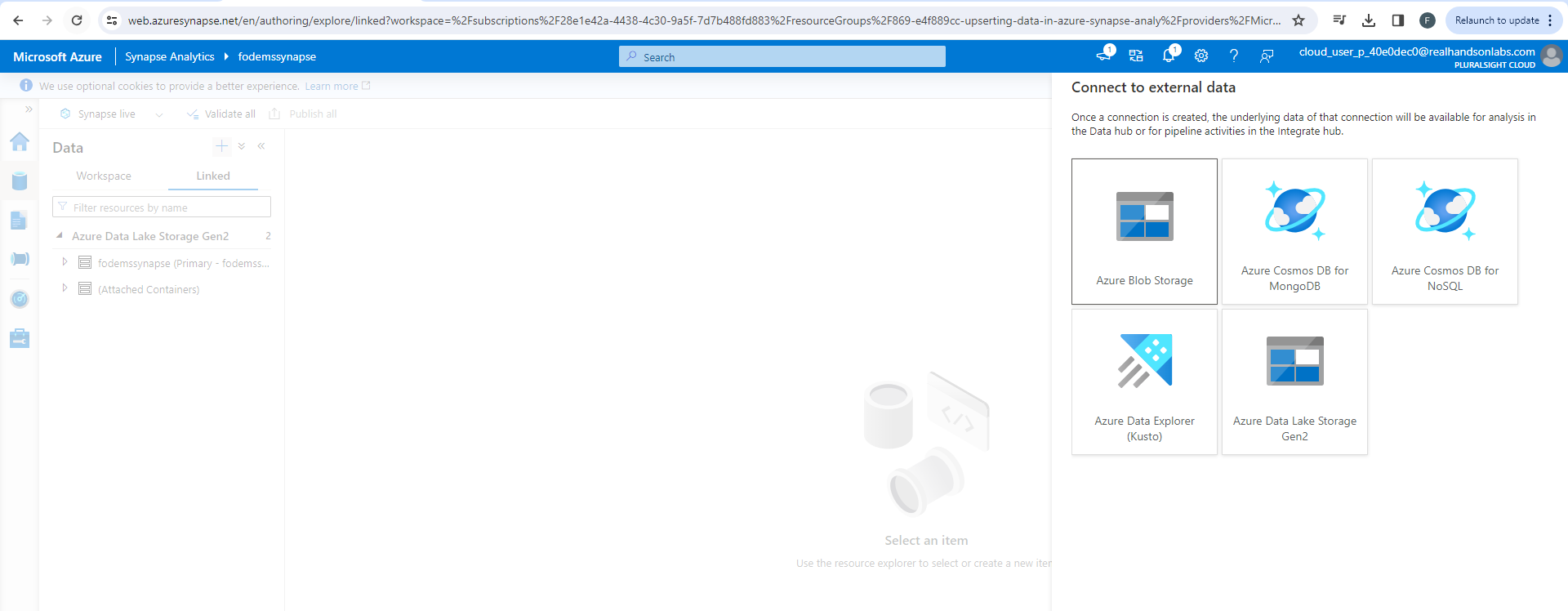
1. In the upper left breadcrumb trail, click on the playground-sandbox link to navigate back to the Resource group page.
2. Click on acgmssynapse.
3. Scroll down, and under Open Synapse Studio, click Open.
4. In the Synapse workspace, click on the Manage icon in the left navigation menu.
5. In the expanded Manage menu, select SQL pools.
6. Click +New to create a new SQL pool.
7. Under Dedicated SQL pool details, set the following:
   * Dedicated SQL pool name: Enter TaxiRidesWarehouse.
   * Performance level: Adjust the setting to DW100c.
8. Click Review + create.
9. Click Create.
   * Note: This may take a few minutes to deploy.



**Create a Pipeline to Copy Data from Data Lake File to Dedicated SQL Table**

Create a Linked Service

1. In the left Analytics pools navigation menu, under External connections, click Linked services. (You should see a default linked service for Azure Synapse, but due to lab environment restrictions, you need to create a new one.)
2. In the left navigation menu, click on the Data icon.
3. In the Data sub-navigation menu, click on the Linked tab.
4. Next to Data, click on the + icon, and select Connect to external data.



1. Select Azure Data Lake Storage Gen2.
2. Click Continue.
3. For New linked service, enter the following:
   * Name: Enter AzureDataLakeLinkedService.
   * Storage account name: Select acgmsstorage.
4. In the bottom right corner, click Test connection, you should see the Connection successful notification.
5. Click Create.

Create a Pipeline

1. In the left navigation menu, click on the Develop icon.
2. Next to Develop, click on the + icon, and select SQL script.
3. Expand the SQL script 1 window.
4. In the toolbar, click on the Connect to dropdown menu and select TaxiRidesWarehouse.
5. Ensure Use database is set to TaxiRidesWarehouse.
6. Create a TaxiRides table in the dedicated SQL pool:

CREATE TABLE TaxiRides

(

RideId INT,

VendorId INT,

PickupTime VARCHAR(100),

DropTime VARCHAR(100),

PickupLocationId INT,

DropLocationId INT,

CabNumber VARCHAR(100),

DriverLicenseNumber VARCHAR(100),

PassengerCount INT,

TripDistance FLOAT,

RateCodeId INT,

PaymentType INT,

TotalAmount FLOAT,

FareAmount FLOAT,

Extra FLOAT,

MtaTax FLOAT,

TipAmount FLOAT,

TollsAmount FLOAT,

ImprovementSurcharge FLOAT

)

WITH

(

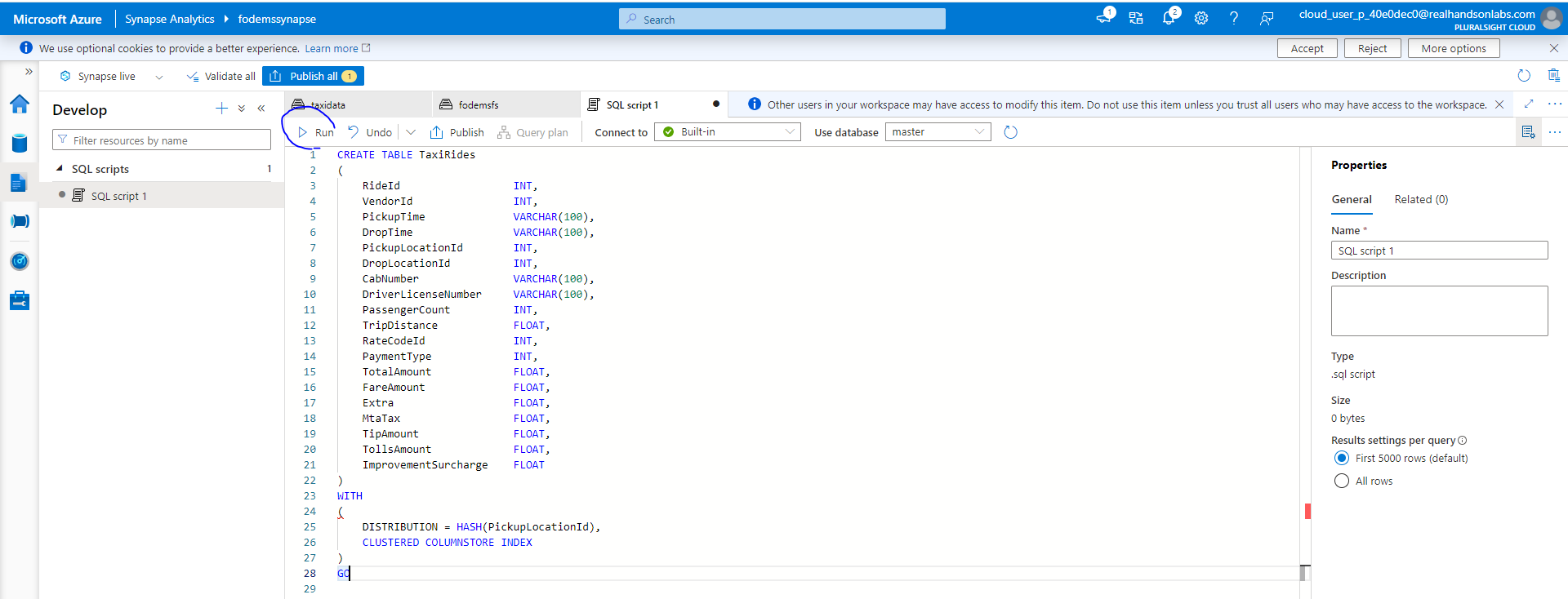
DISTRIBUTION = HASH(PickupLocationId),

CLUSTERED COLUMNSTORE INDEX

)

GO

1- In the upper left corner, click Run

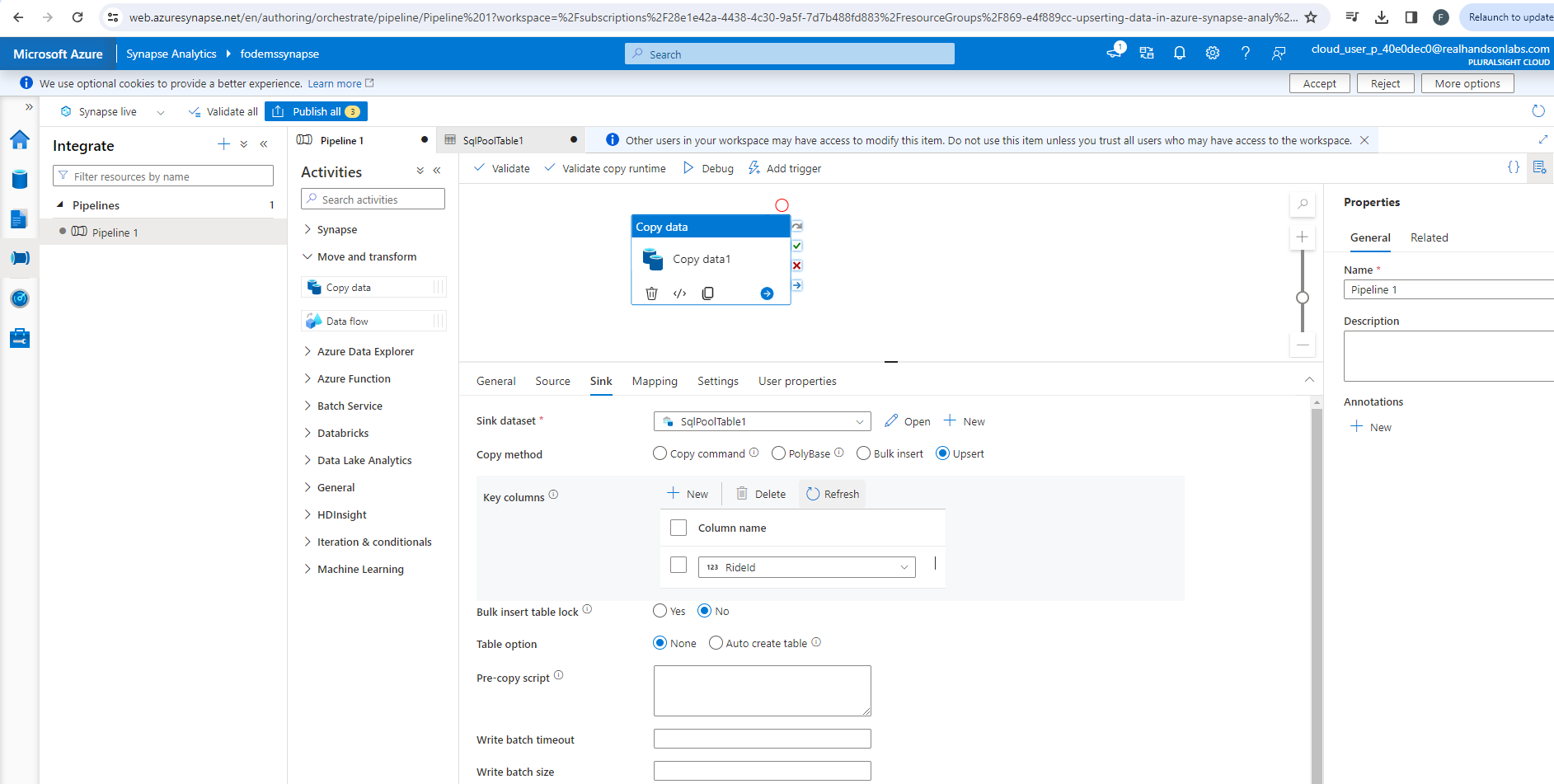


Update the Pipeline

1. Minimize the SQL script 1 window, and in the left navigation menu, click on the Integrate icon.
2. Next to Integrate, click on the + icon, and select Pipeline.
3. Expand the Pipeline 1 window.
4. In the left navigation menu, under Move and transform, select Copy data, and drag and drop it onto the canvas.
5. Click on the Source tab.
6. For Source dataset, click + New.
7. Under New integration set, search for and select Azure Data Lake Storage Gen2.
8. Click Continue.
9. Select Delimited Text.
10. Click Continue.
11. For Set properties > Linked service, select AzureDataLakeLinkedService.
12. Next to File path, click on the folder icon to browse, and select taxidata > Raw > TaxiRides.csv.
13. Click OK.
14. Leave the settings at default, and click OK.

Enter Sink Properties

1. Click on the Sink tab.
2. For Sink dataset, click + New.
3. Under New integration set, search for and select Azure Synapse dedicated SQL pool.
4. Click Continue.
5. For Set properties, enter the following:
   * SQL pool: Select TaxiRidesWarehouse.
   * Table name: Select dbo.TaxiRides.



1. Leave the other settings at default, and click OK.
2. In the Sink tab, for Copy method, select Upsert.
3. For Key columns, select + New.

