**Assignment 8: ADF Data Flow 3 Question (Tough)**

**Advanced Questions :**

1. Create a pipeline to copy the customer data from csv file to SQL table which has only 3 columns. While copying the data, ensure that copy all the rows where custid is odd number. Also check if cust id exist in table then update it.

File attached with question.

Table: Create dummy table with column (CID, Name, mailid)

Insert dummy row with id 1,3,5,7,9,8,10

2. Create a pipeline to implement the SCD Type 3 example using the Dataflow.

3. Create a pipeline to remove the duplicate record available in the CSV file and save it as JSON file.a

CSV file attached: **SalesLT.Customer.xRyuaEu2p7**

**Solution:**

**Question1:**

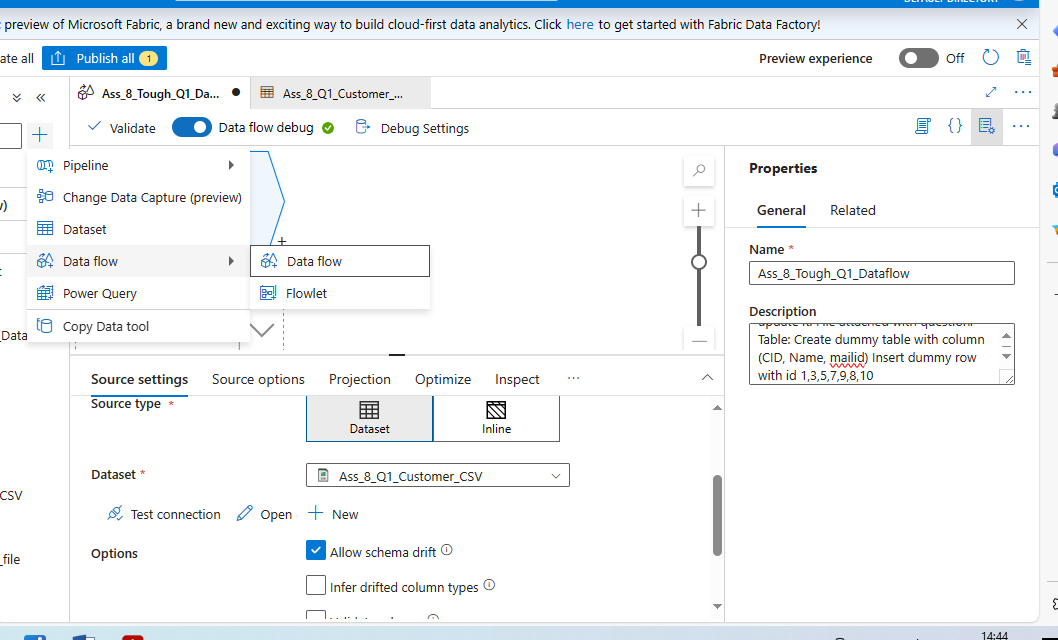
1. Create a pipeline to copy the customer data from csv file to SQL table which has only 3 columns. While copying the data, ensure that copy all the rows where custid is odd number. Also check if cust id exist in table then update it.

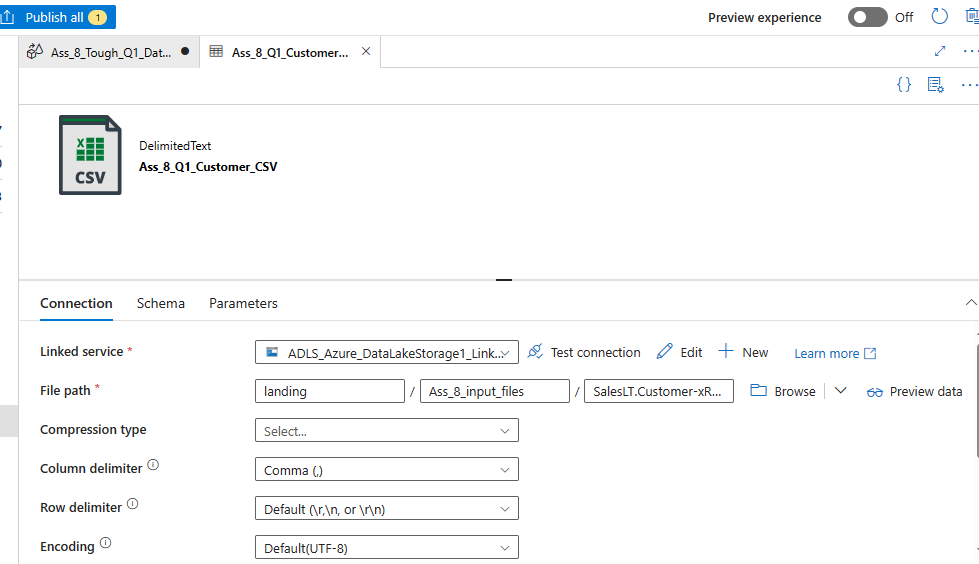
File attached with question.

Table: Create dummy table with column (CID, Name, mailid)

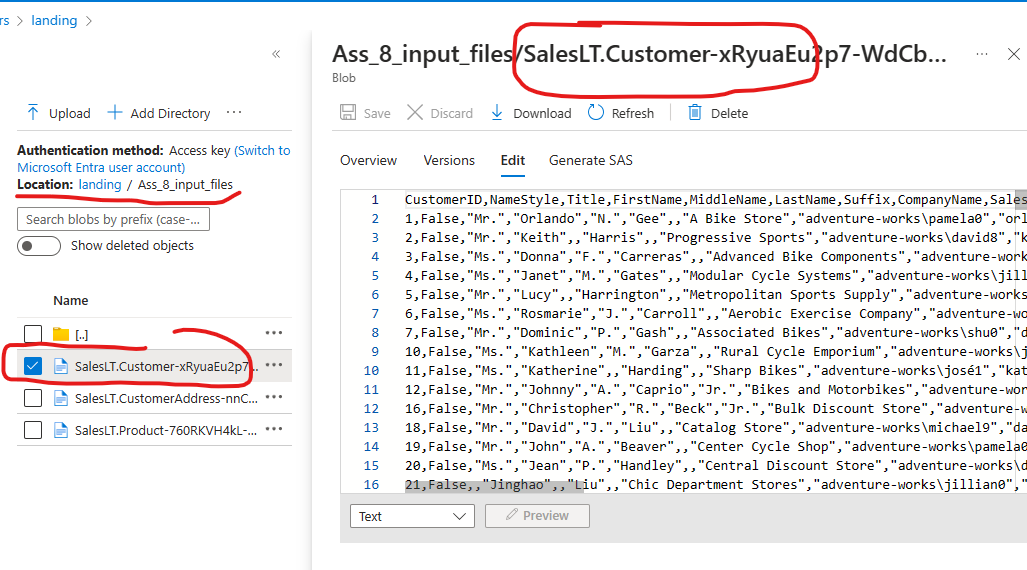
Insert dummy row with id 1,3,5,7,9,8,10

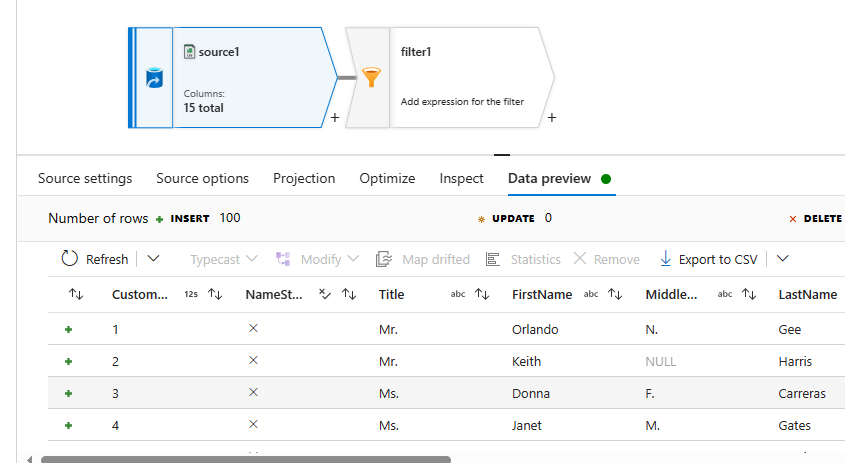
**Solution:**

****

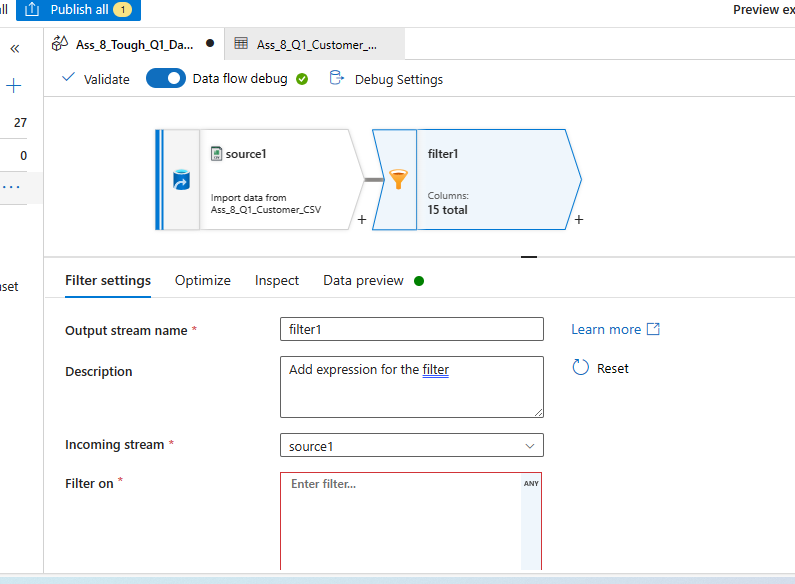
****

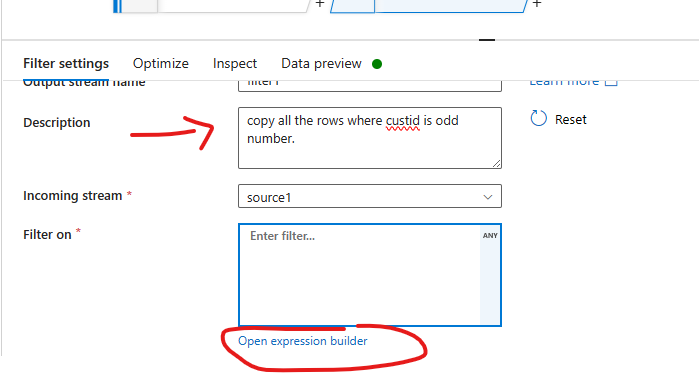
**Customer CSV file in location: (omma separated)**

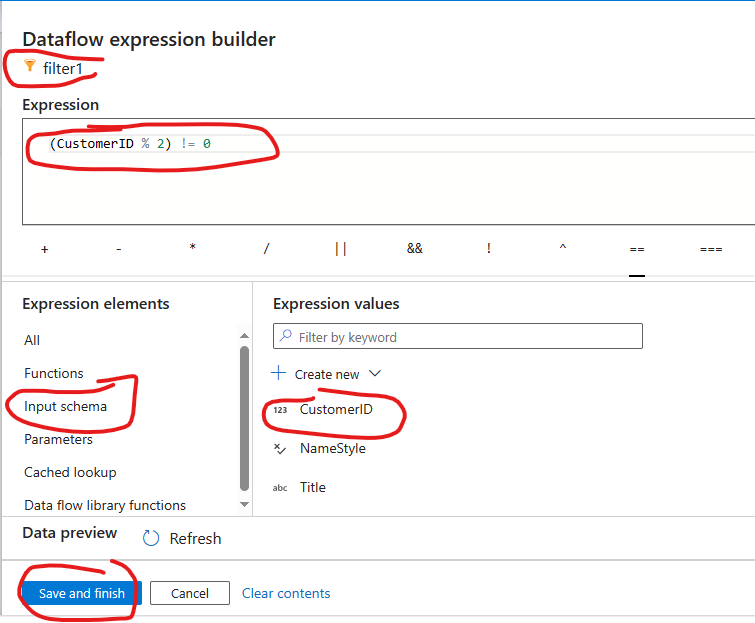
****

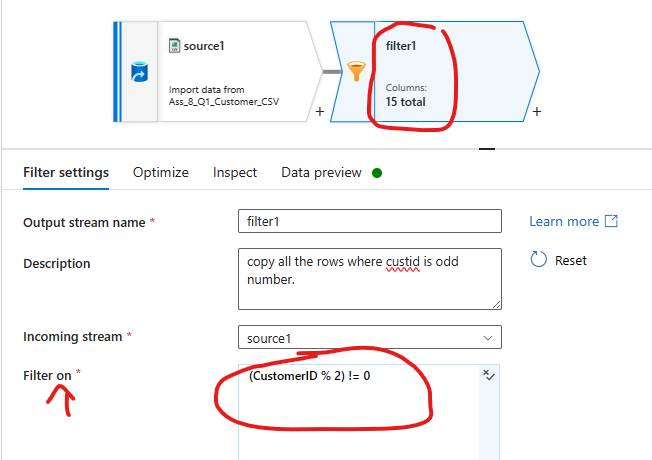
****

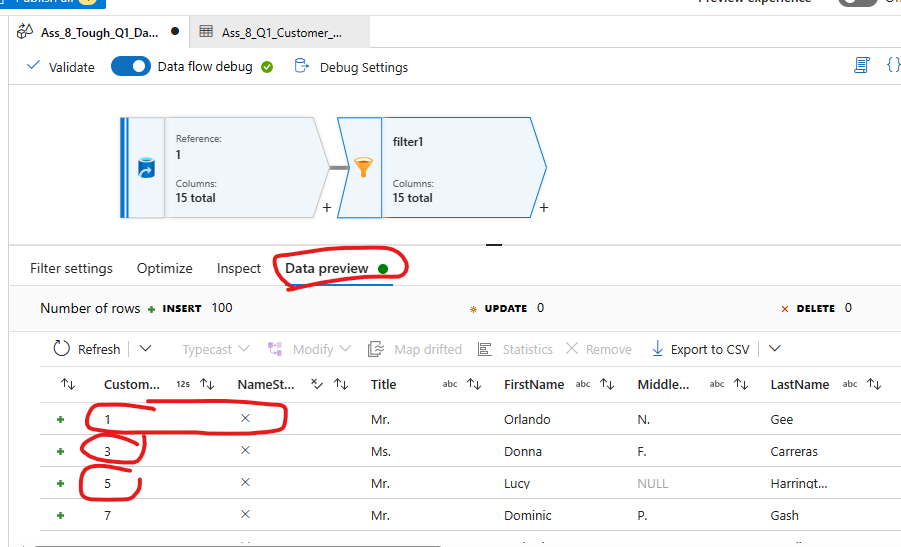
ensure that copy all the rows where custid is odd number.



****

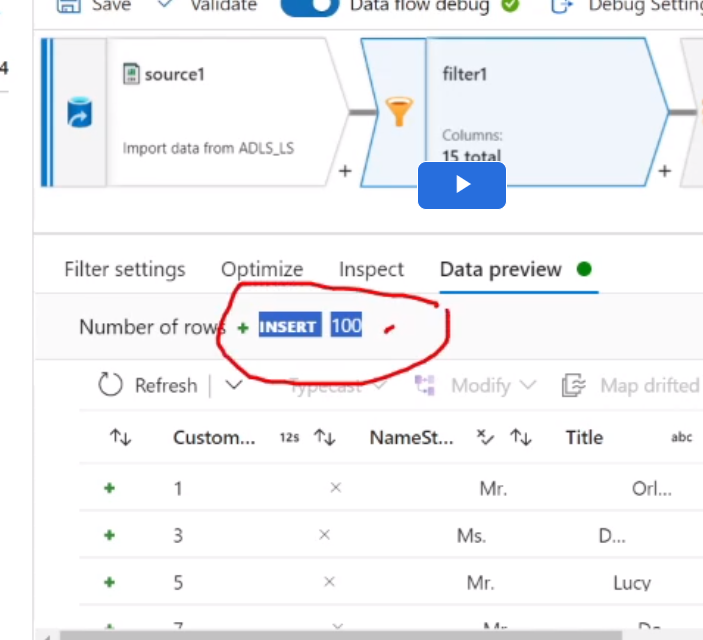
****

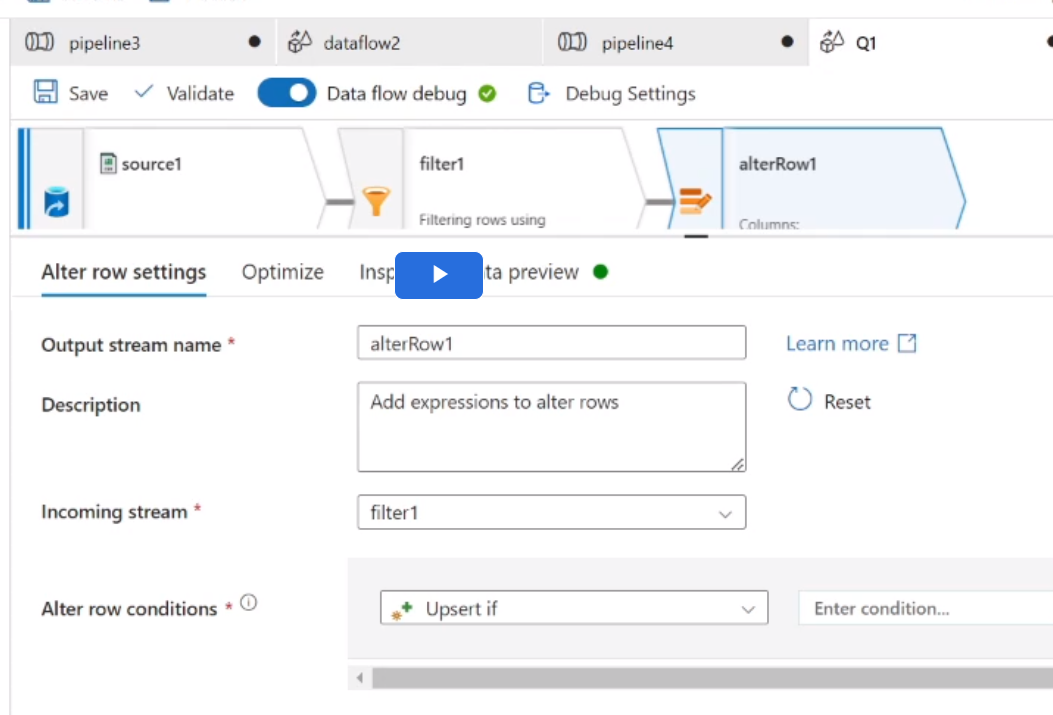
****

****

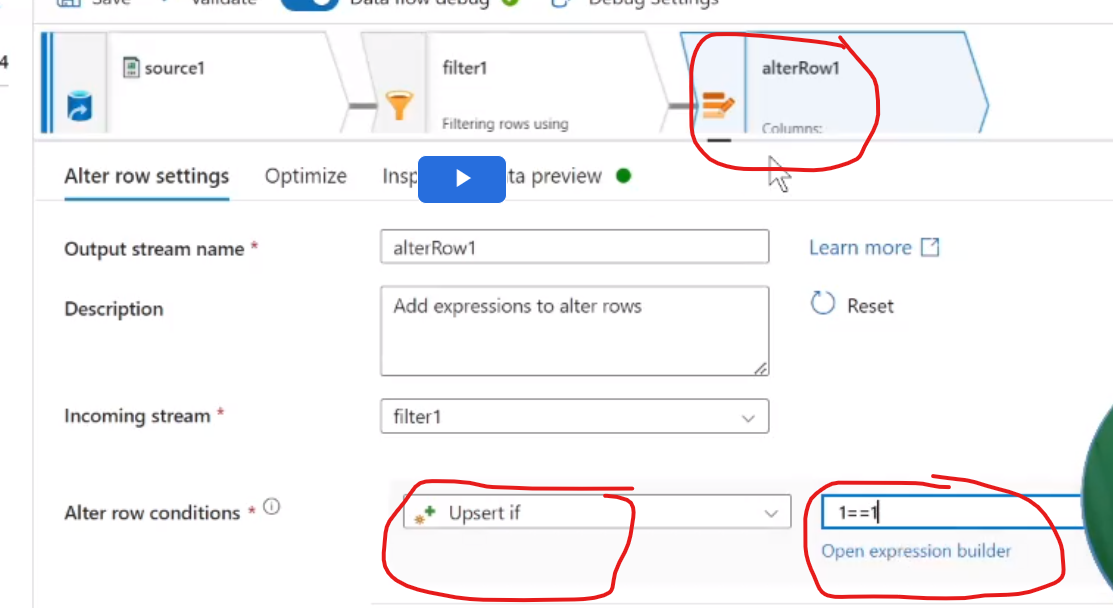
Also check if cust id exist in table then update it.:

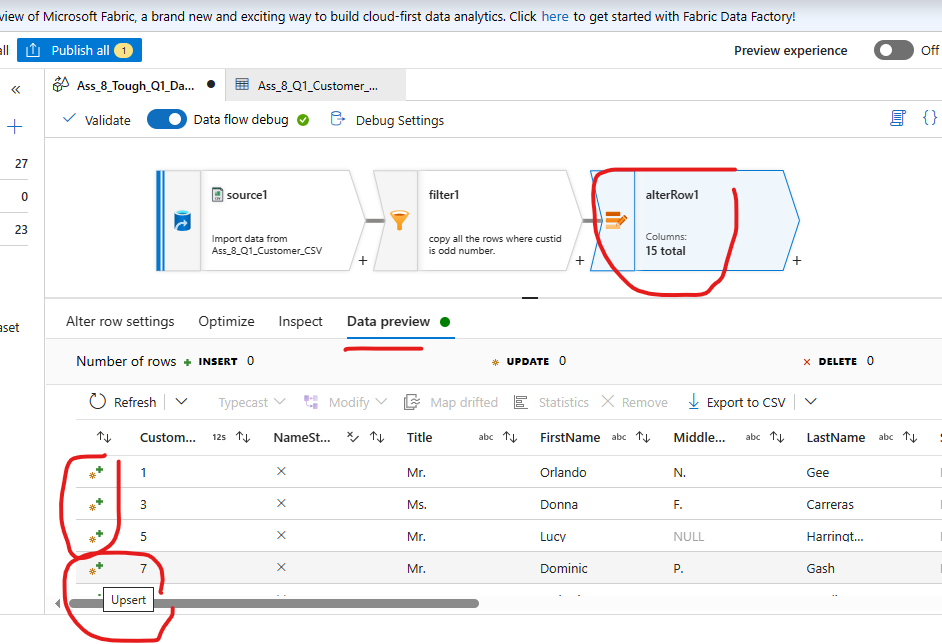
Note all the elements in screenshot has + icon which means all are eligible only for **+INSERT, But not for “UPSERT” – SO I USE “Alter Row” condition and choose property as “UPSERT”**

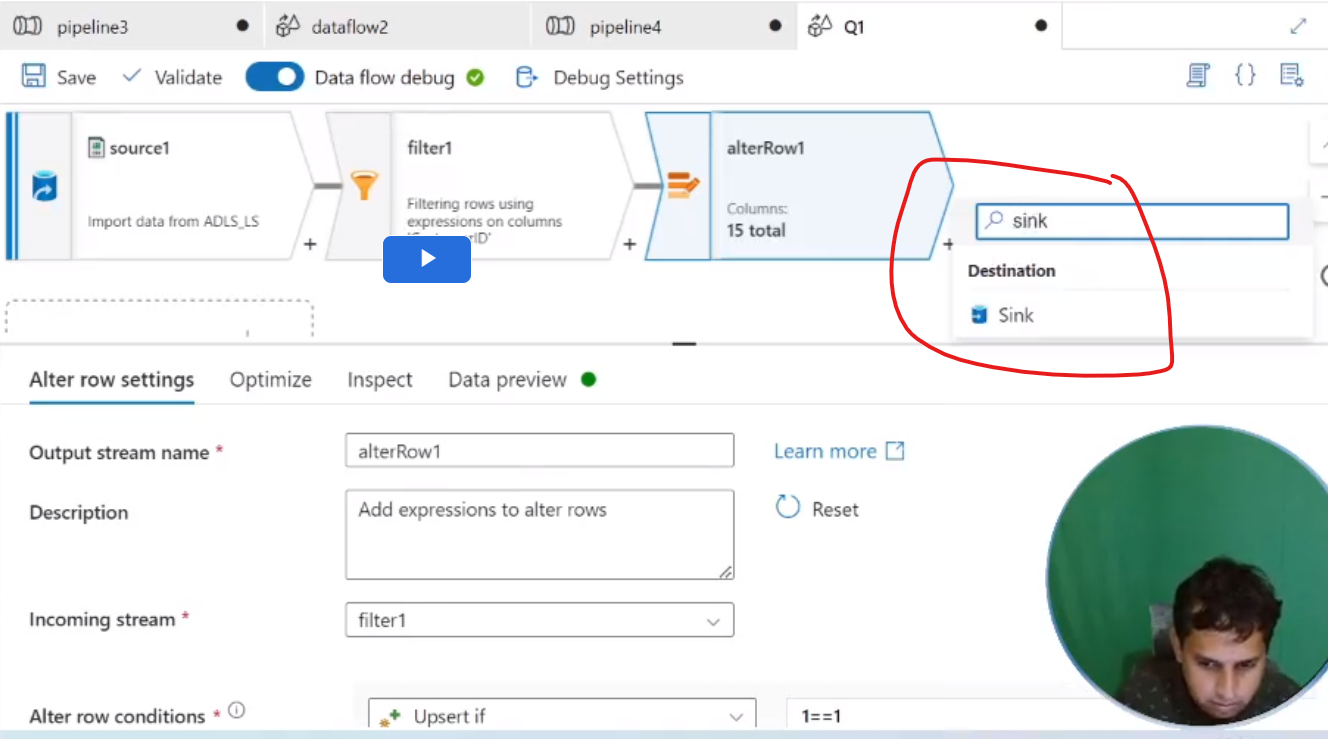
****

****

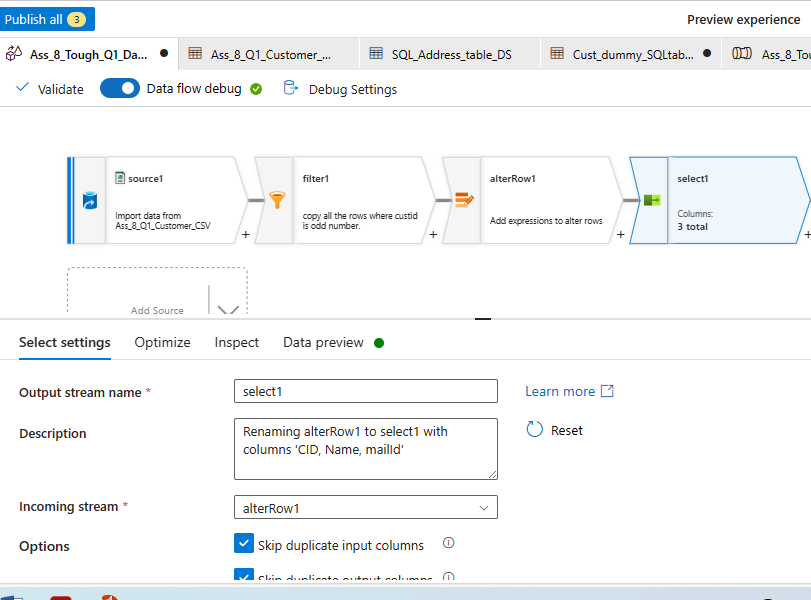
**Give Upsert IFF “1==1”, so all row eligible for upsert**

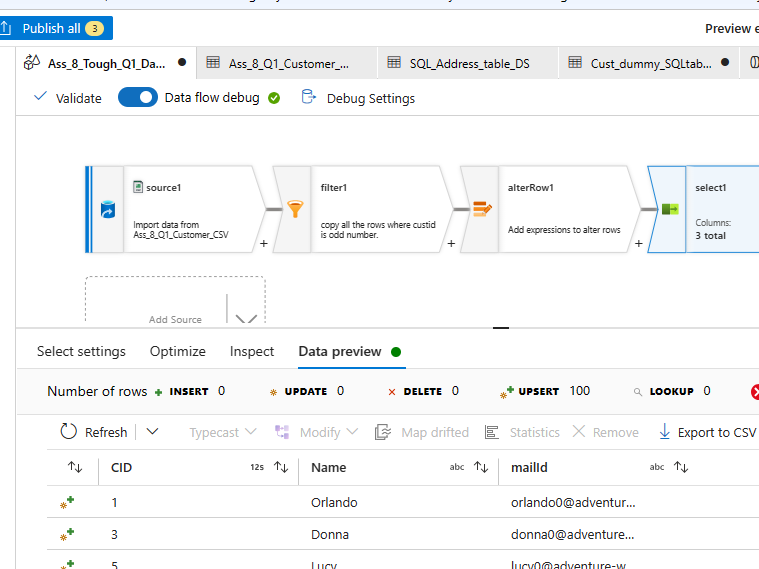
****

****

****

csv file to SQL table which has only 3 columns.



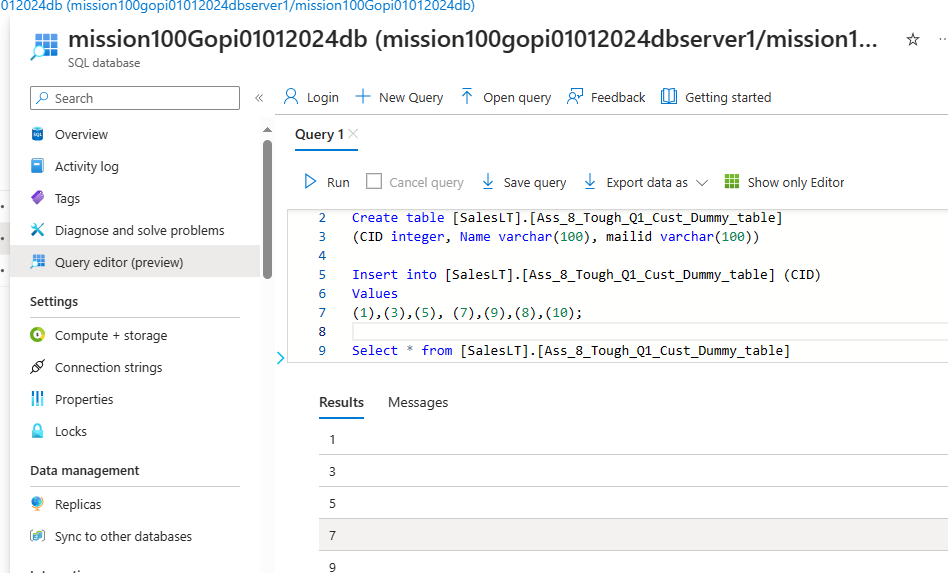


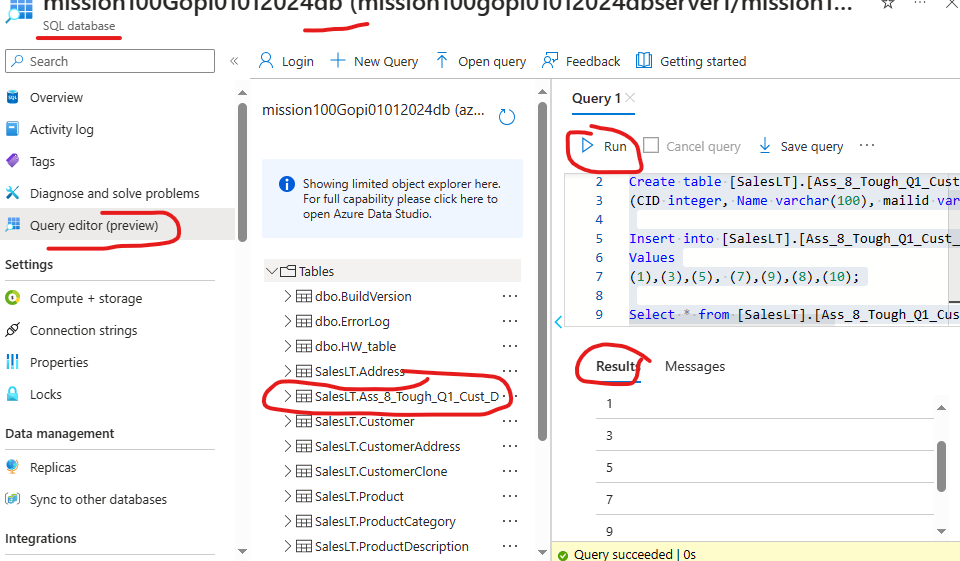
File attached with question.

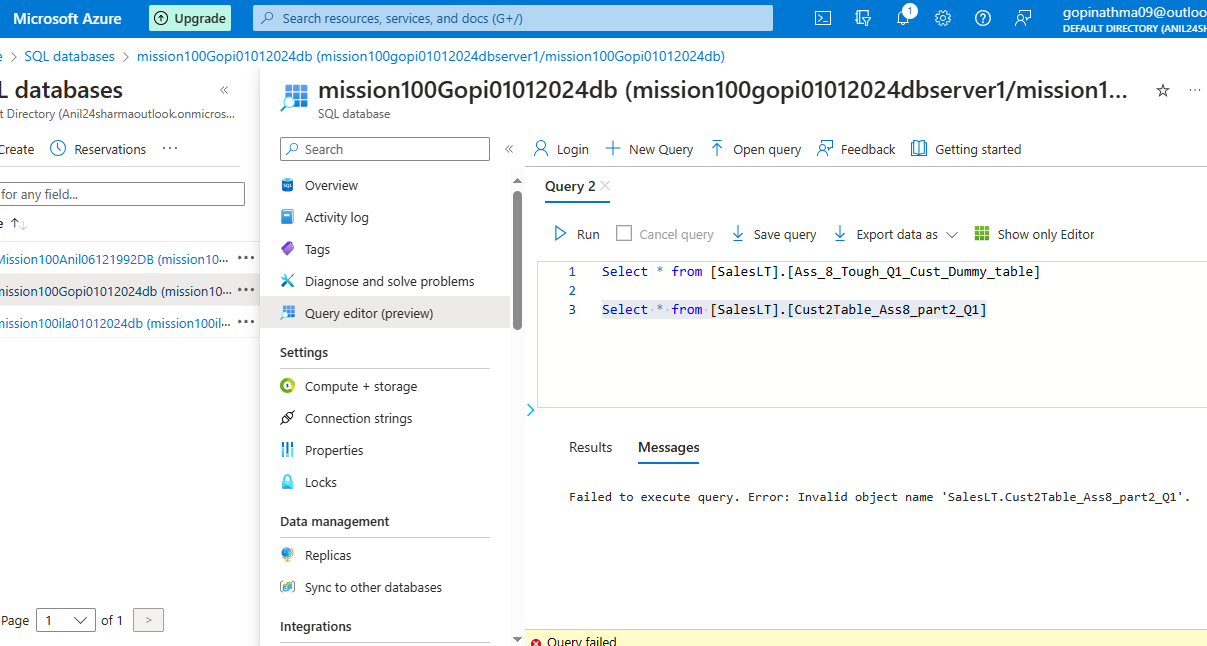
Table: Create dummy table with column (CID, Name, mailid)

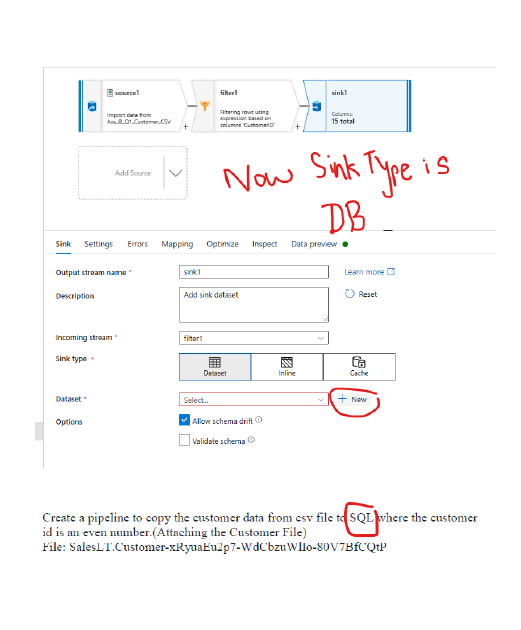
Insert dummy row with id 1,3,5,7,9,8,10

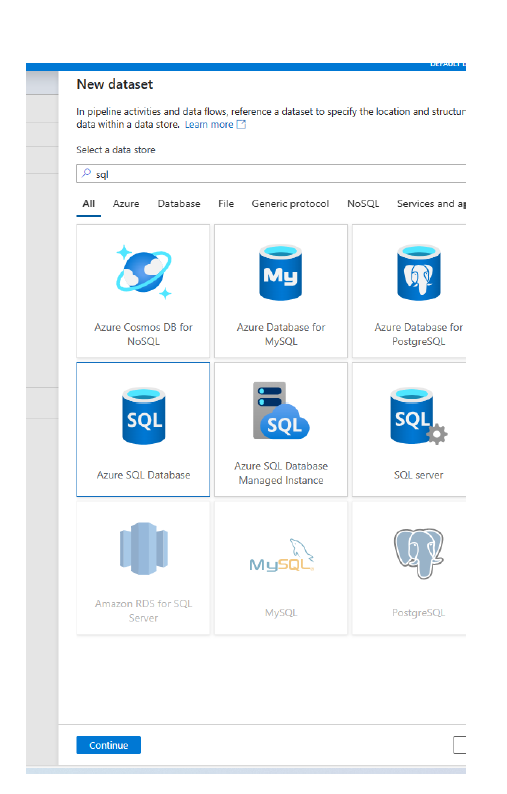
|  |
| --- |
| **Query to create and insert Dummy Table:**  ~~Create table [SalesLT].[Ass\_8\_Tough\_Q1\_Cust\_Dummy\_table]~~  ~~(CID integer, Name varchar(100), mailid varchar(100))~~  ~~Insert into [SalesLT].[Ass\_8\_Tough\_Q1\_Cust\_Dummy\_table] (CID)~~  ~~Values~~  ~~(1),(3),(5), (7),(9),(8),(10);~~  ~~Select \* from [SalesLT].[Ass\_8\_Tough\_Q1\_Cust\_Dummy\_table]~~ |
| Select \* from [SalesLT].[Cust2Table\_Ass8\_part2\_Q1] |
|  |
|  |

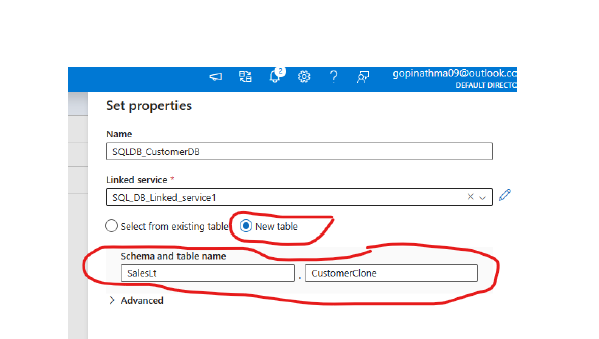
****

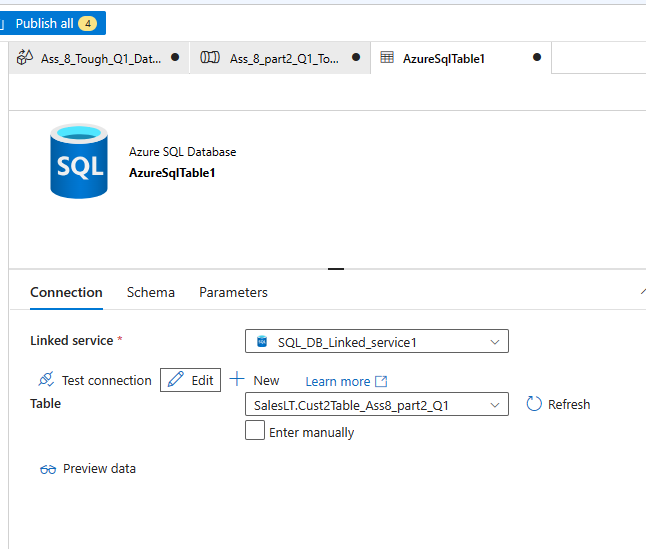
****

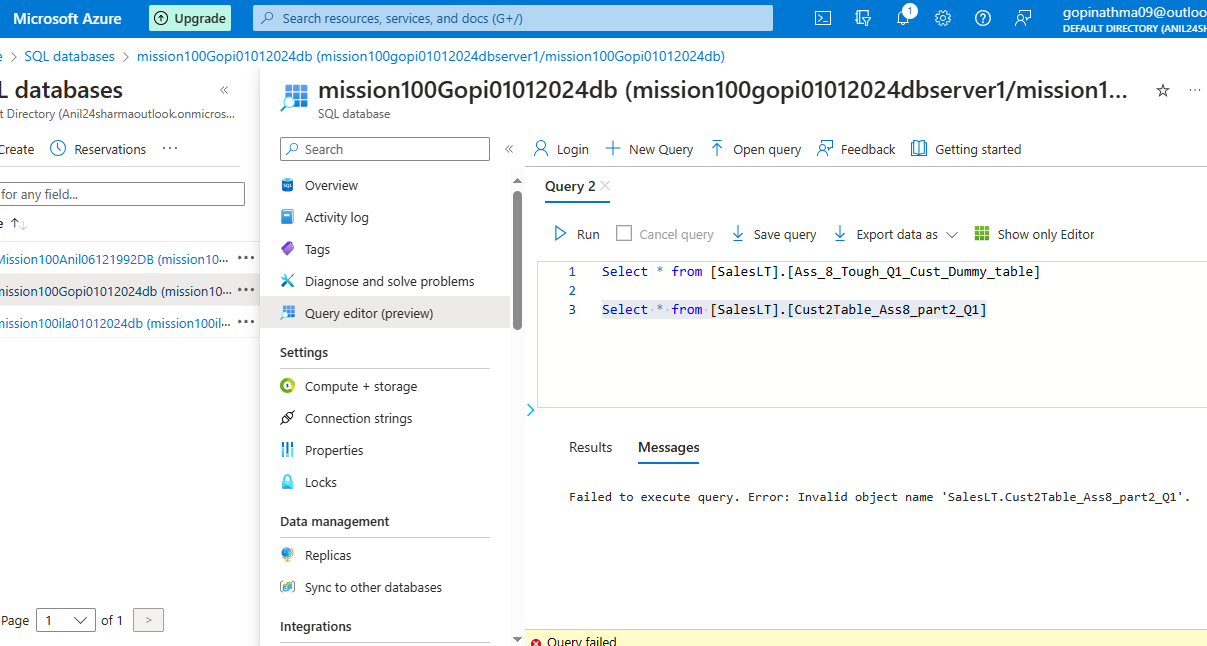
****

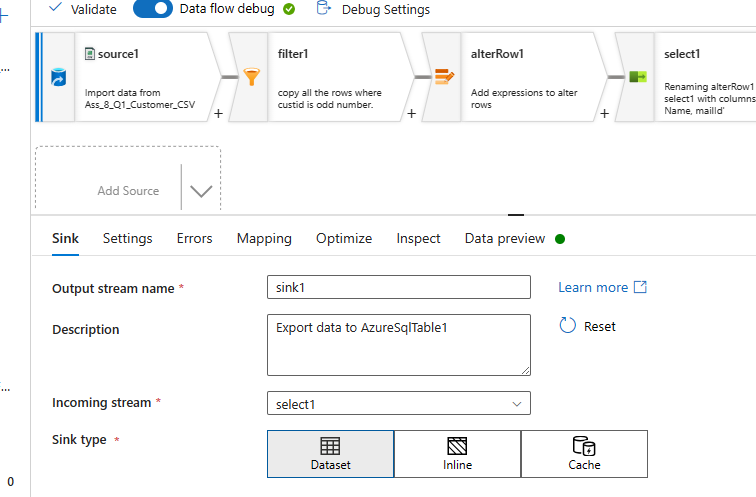
****

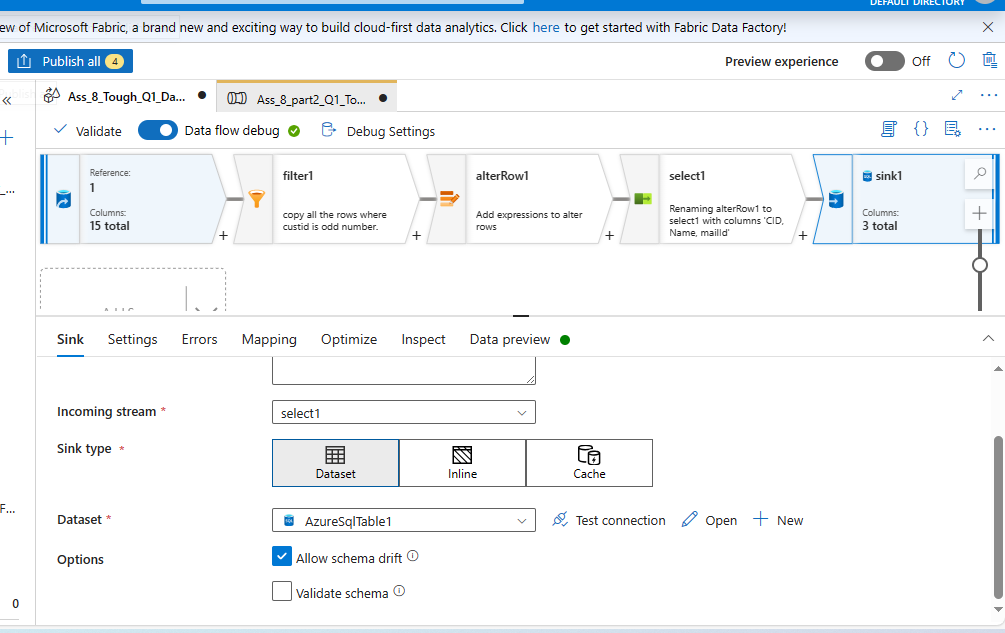
****

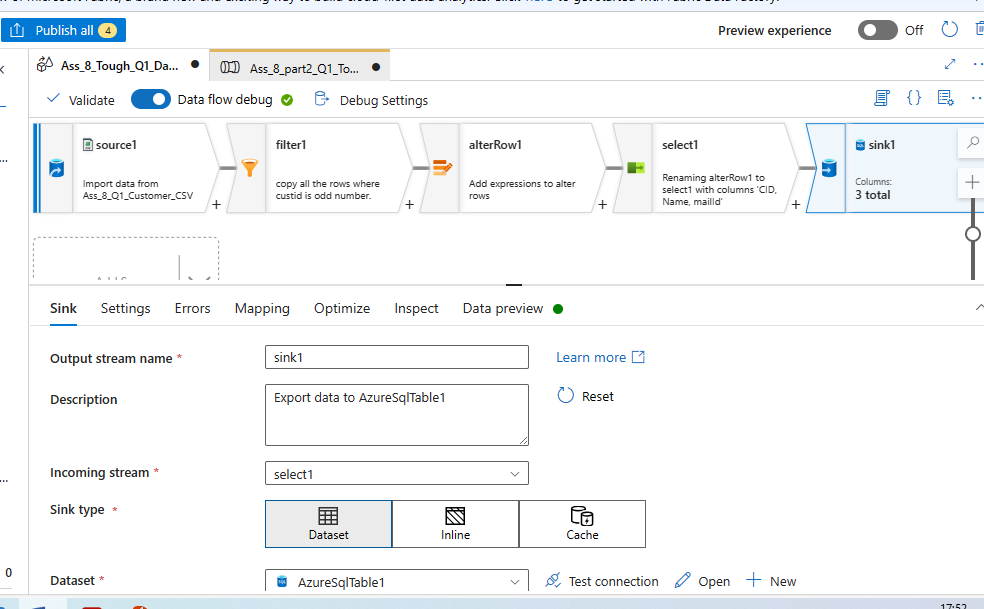
****

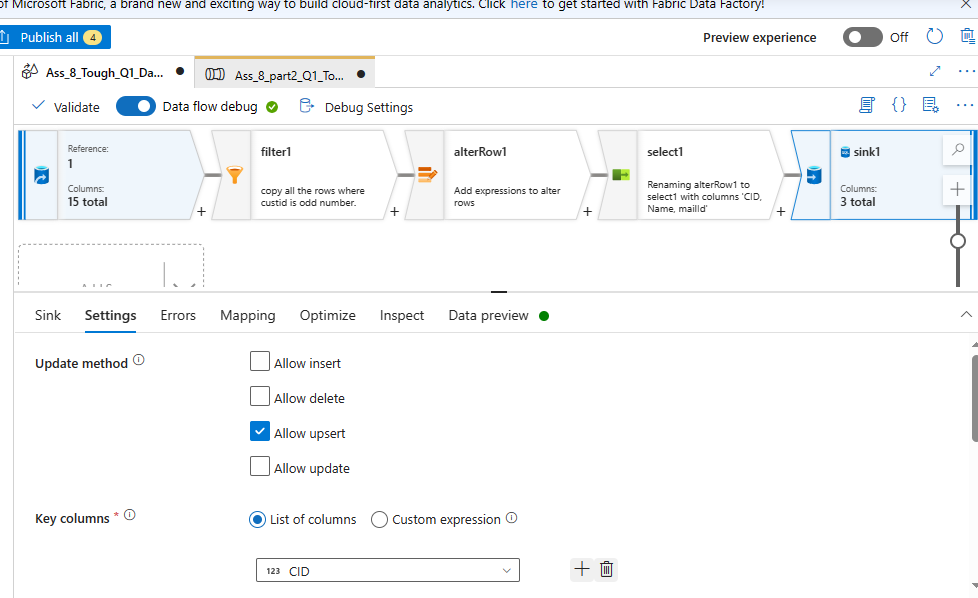
****

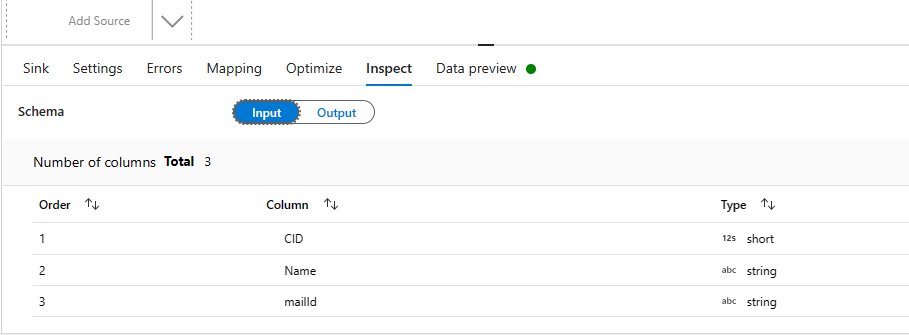
****

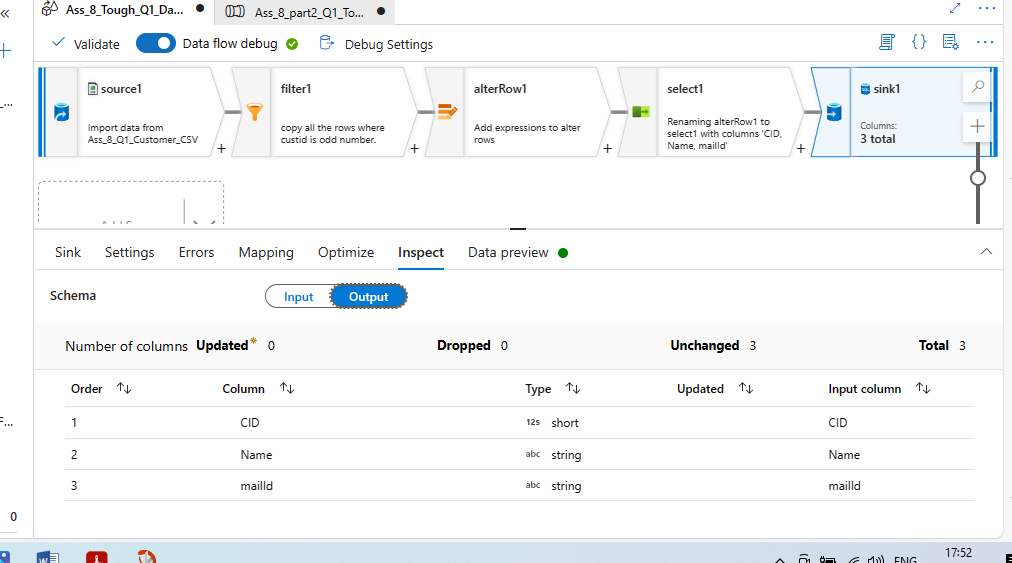
****

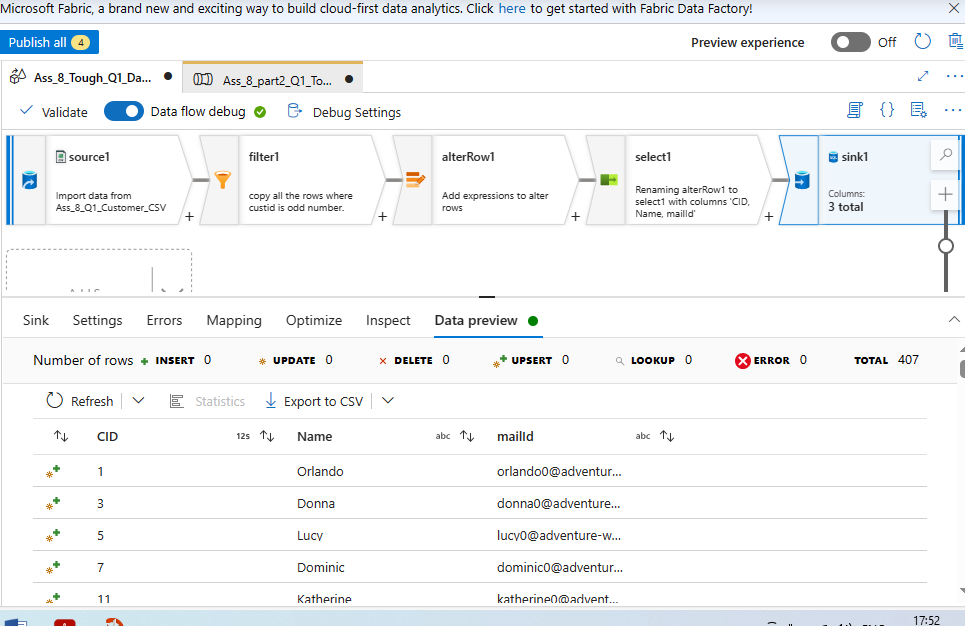
****

****

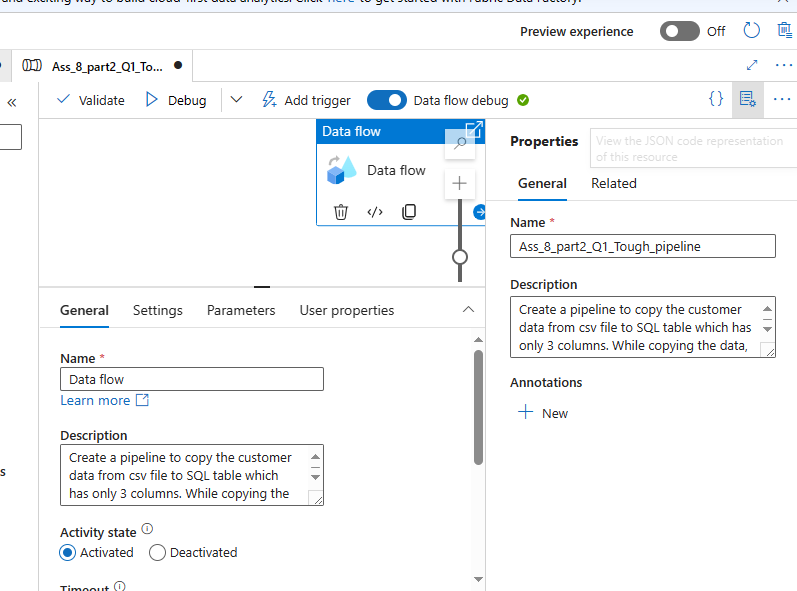
****

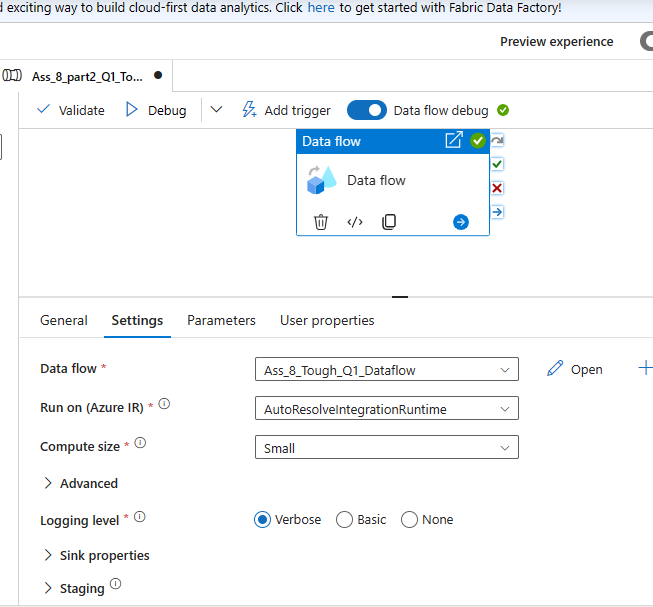
****

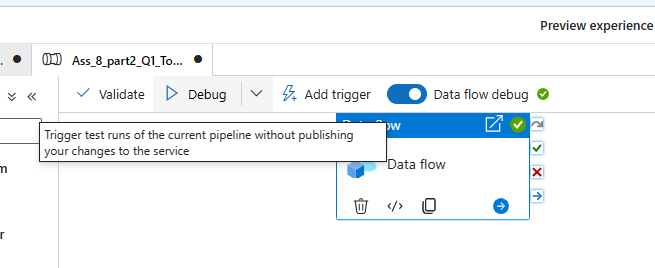
****

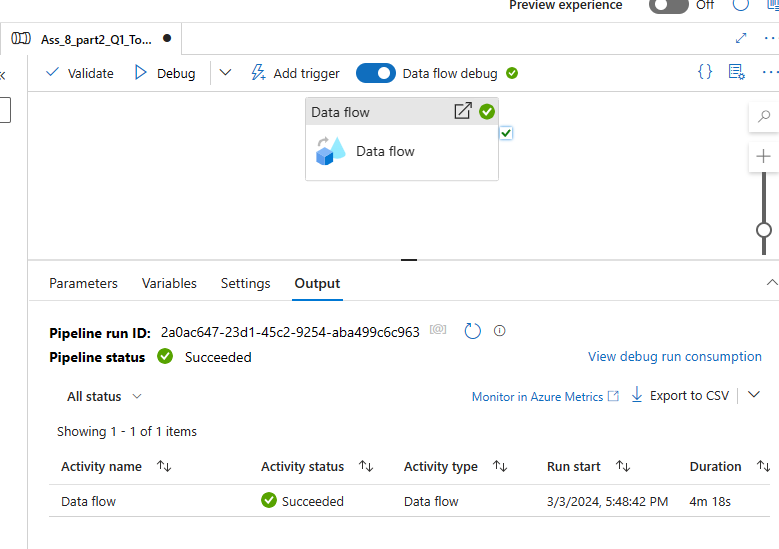
****

**Finally Run pipeline and publish all**

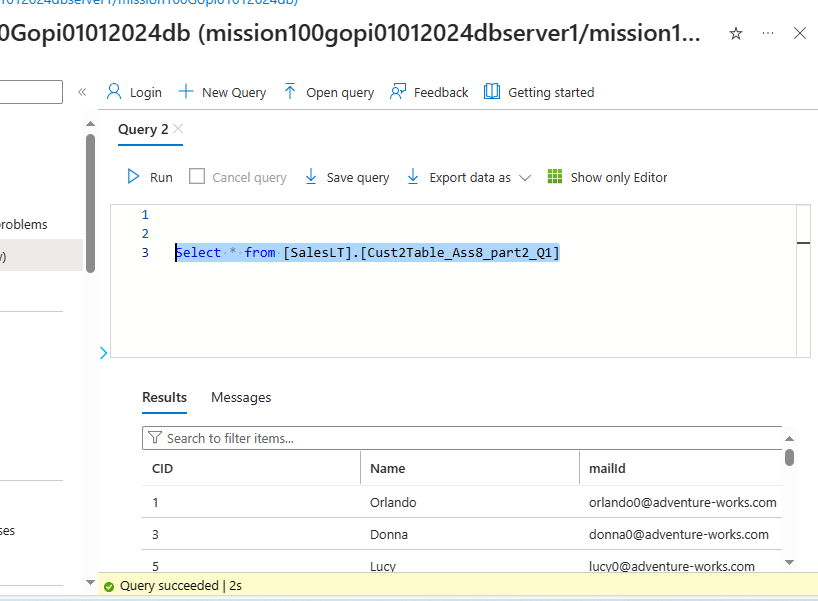
****

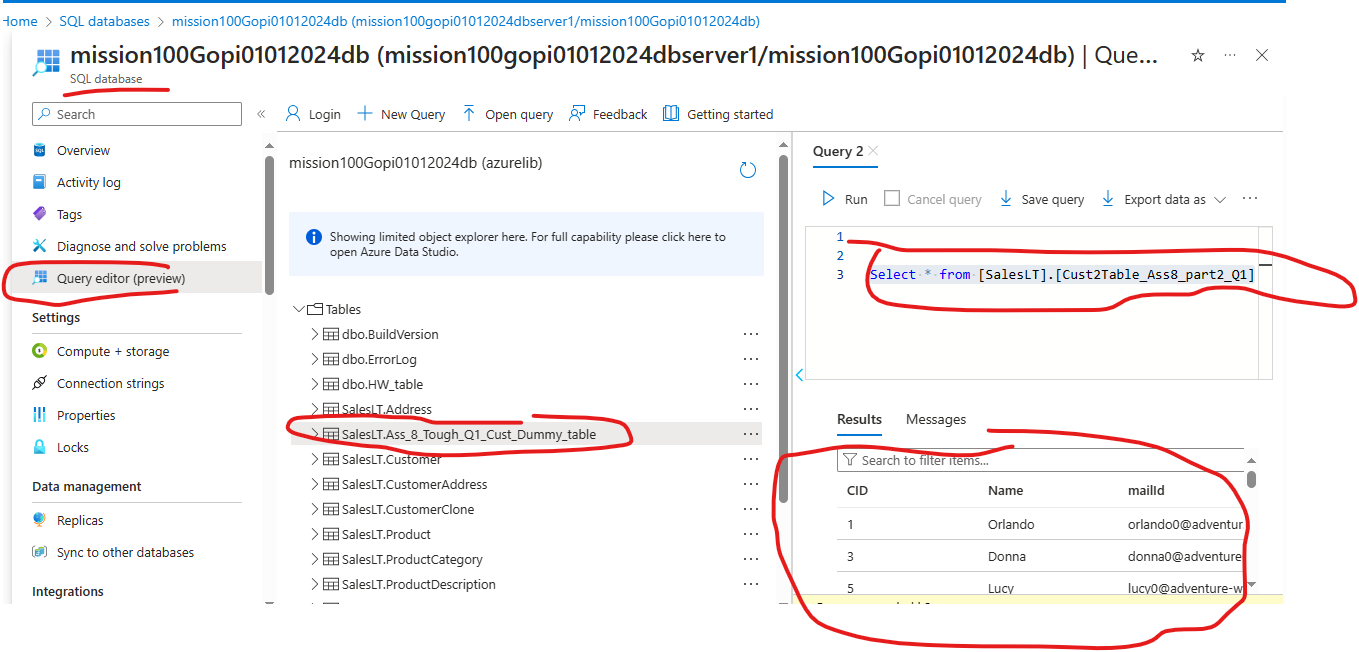
****

****

****

**Once pipeline success , checkif Table got created and rows got inserted**

****

****

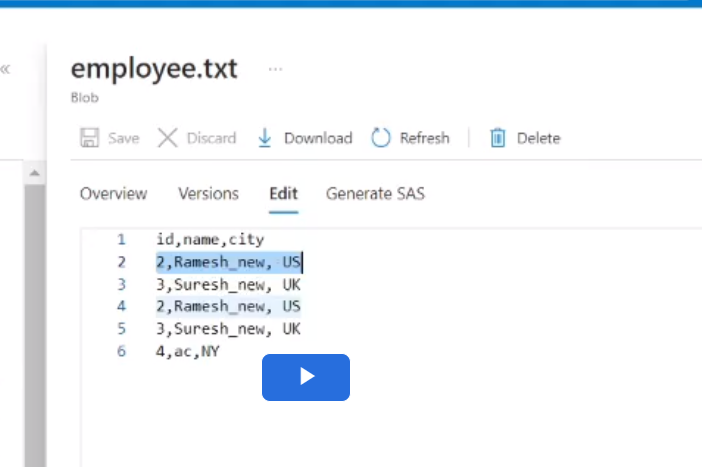
**Finally Click Publish all in Pipeline**

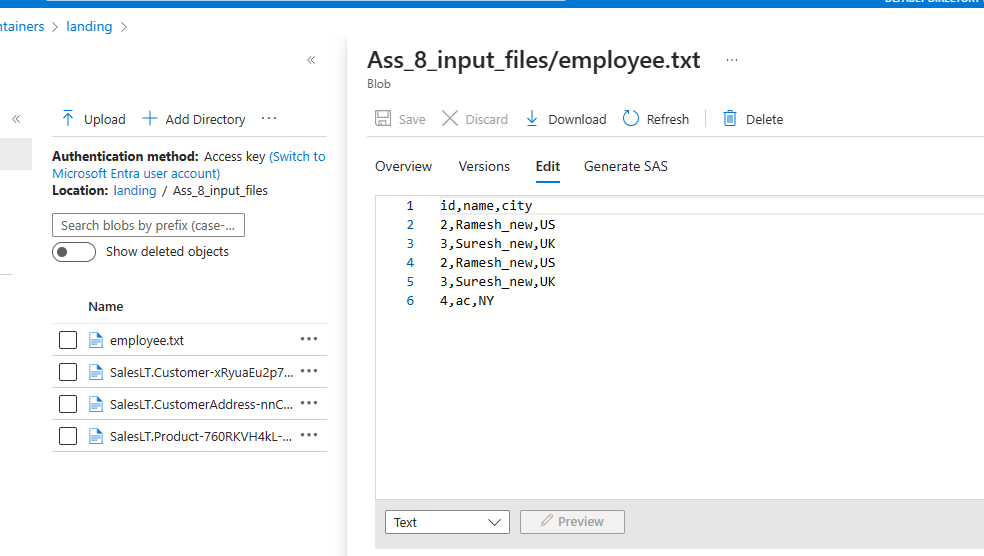
**Question 3:**

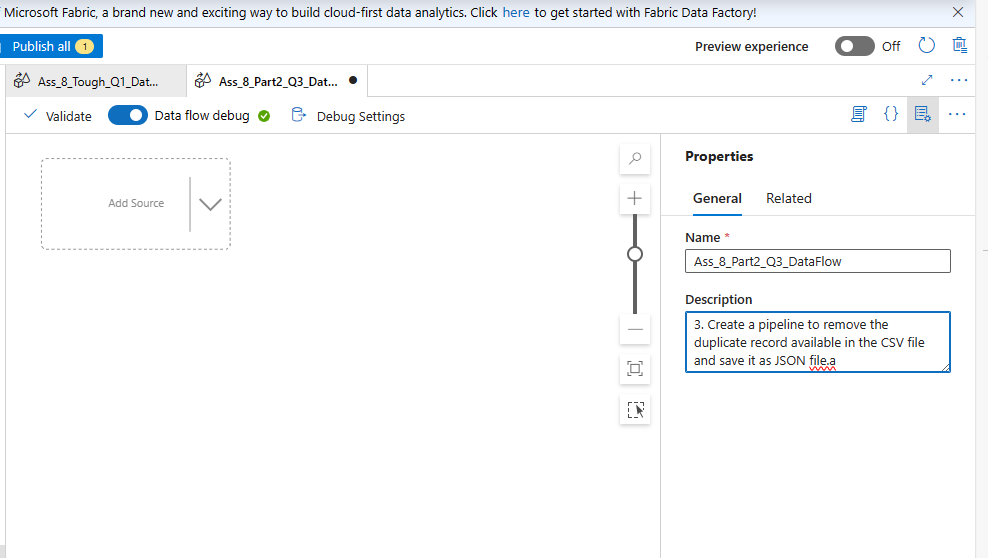
3. Create a pipeline to remove the duplicate record available in the CSV file and save it as JSON file.a

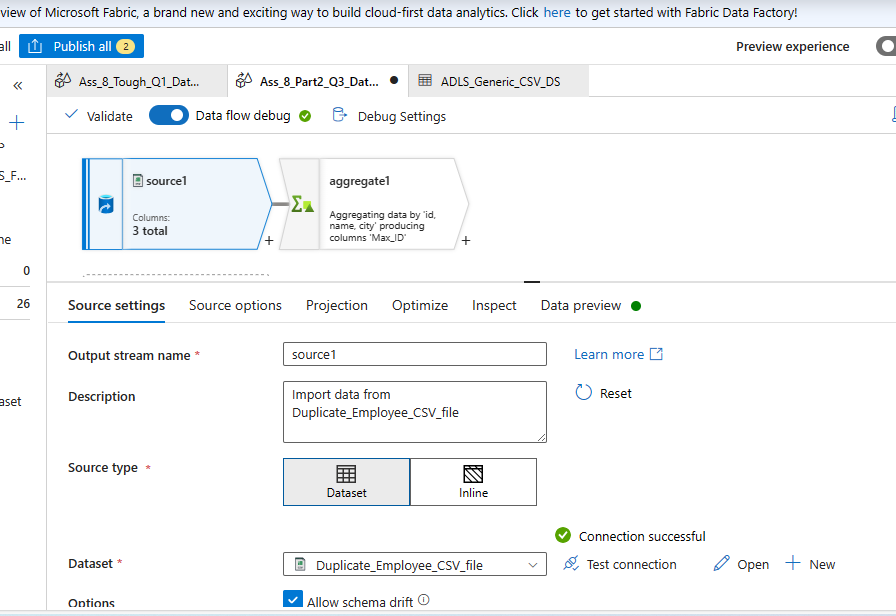
**Solution 3:**

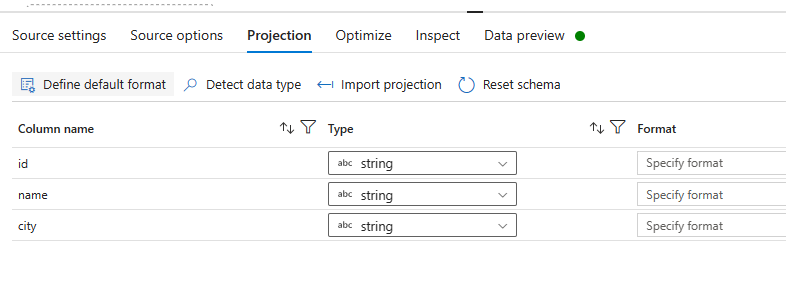
**First load file into Dataflow and then remove duplicate**

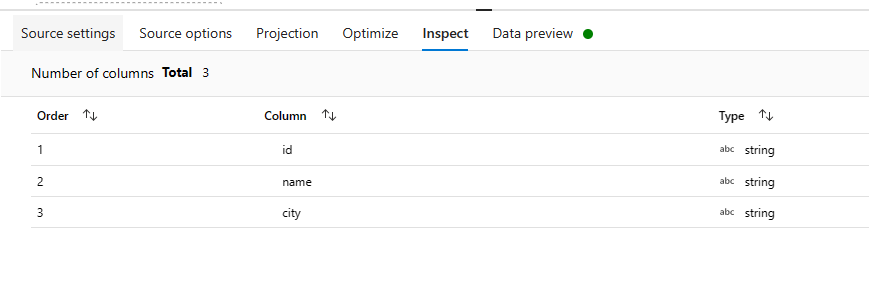
****

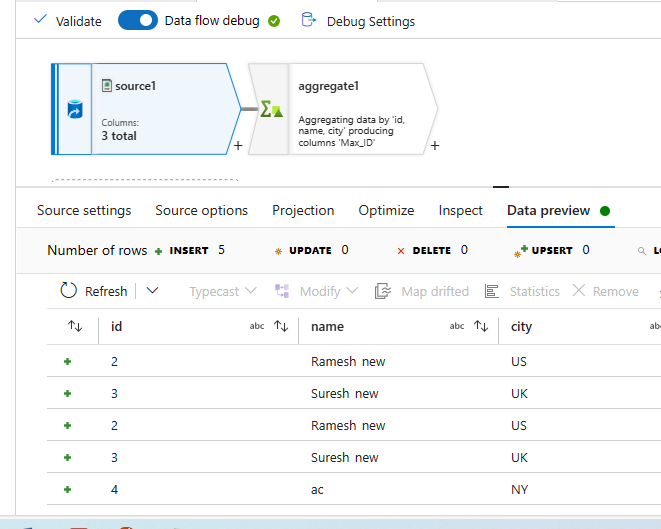
****

****

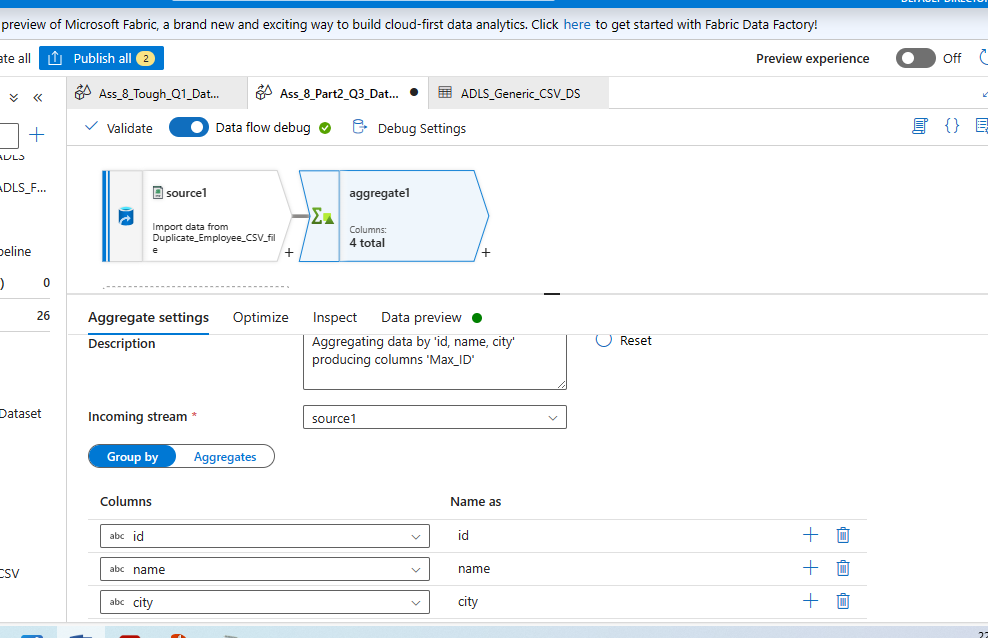
****

****

****

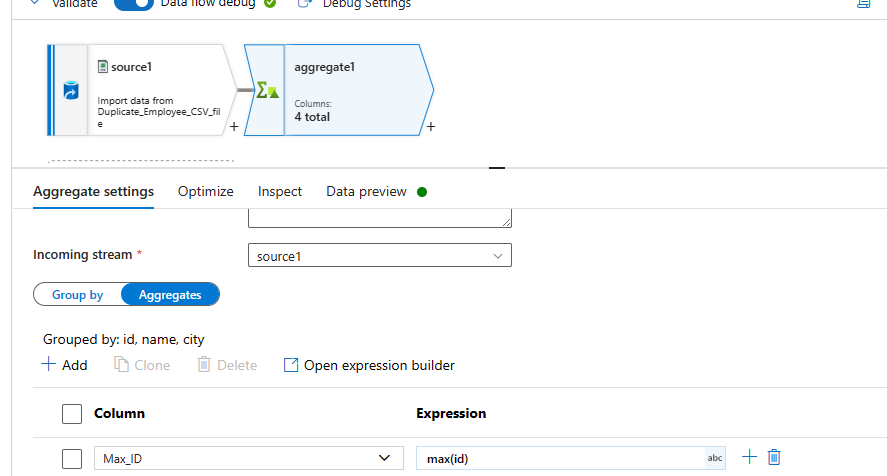
****

**Aggregate on each column will bring duplicate to single unique value**

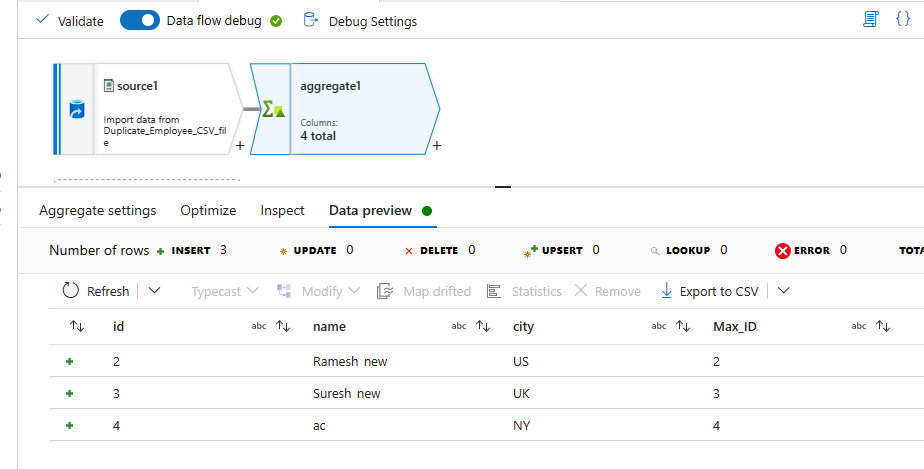
****

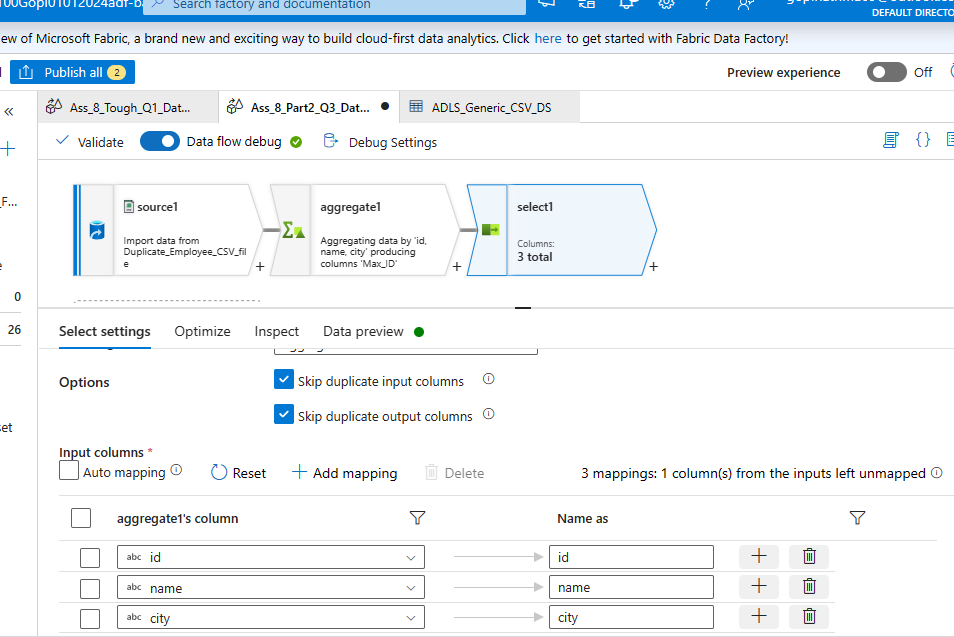
**Fpor namesake, it mis must to use Aggregate function (because we used Group by , so Aggregate is must as per syntax)**

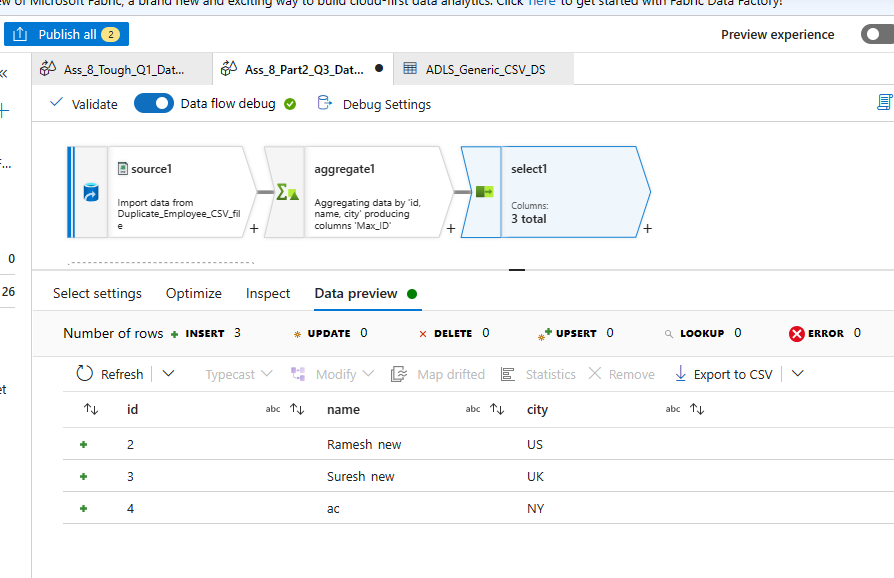
**So we apply some MaxId or so**

****

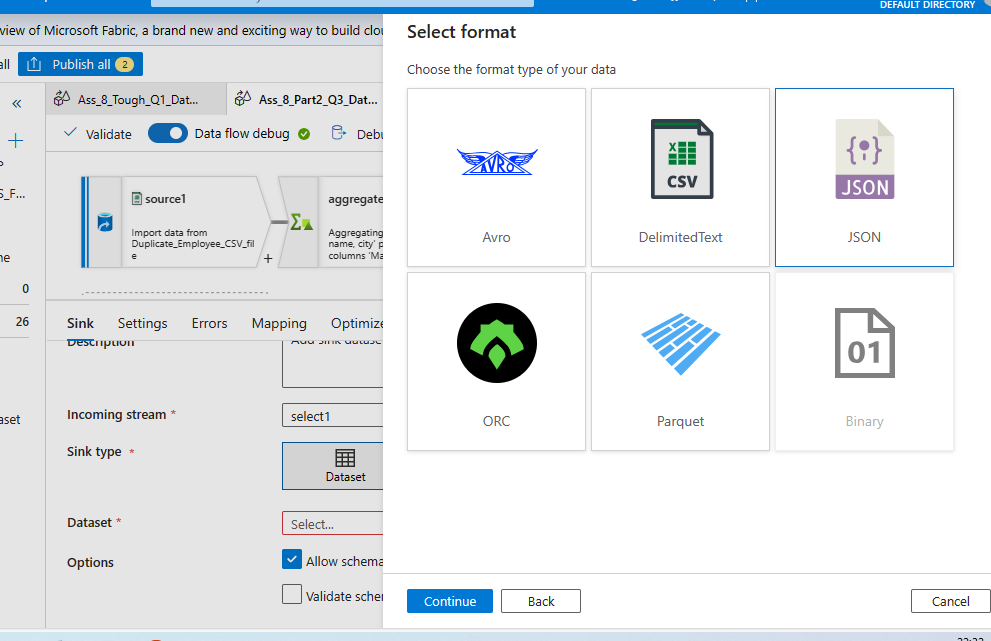
**If you see “Data Prview” tab, u get all unique output along with extra column Max ID, so I again use Select function and just use 3 columns as output**

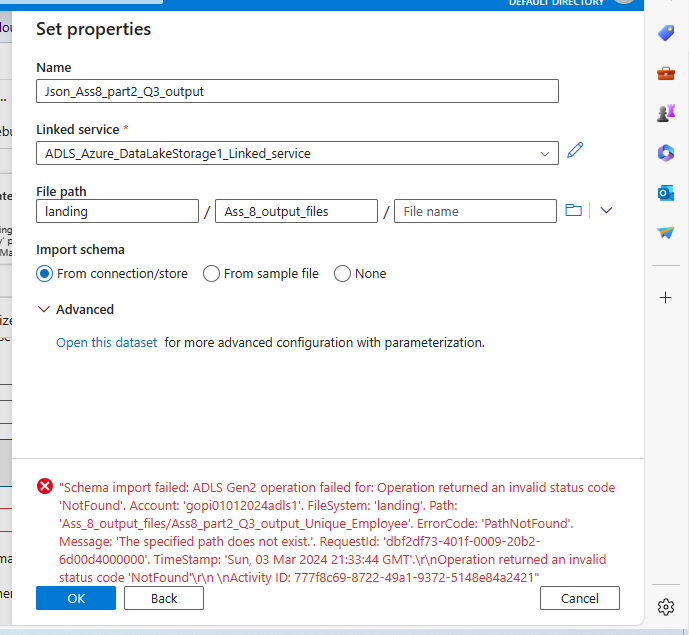
****

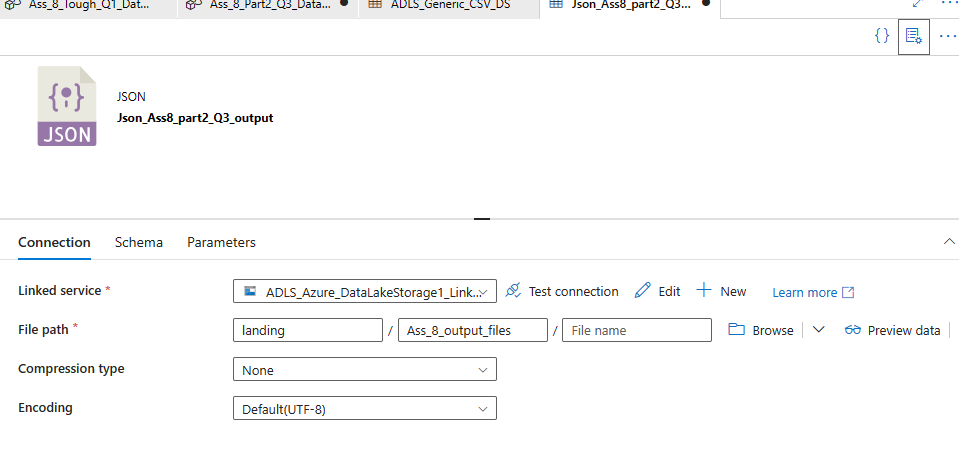
****

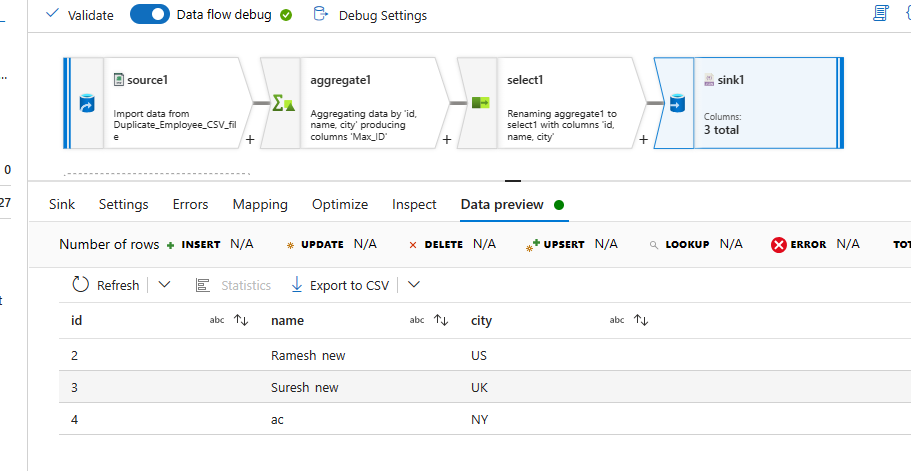
****

and save it as JSON file.a 🡺 We need Destination as **Sink**

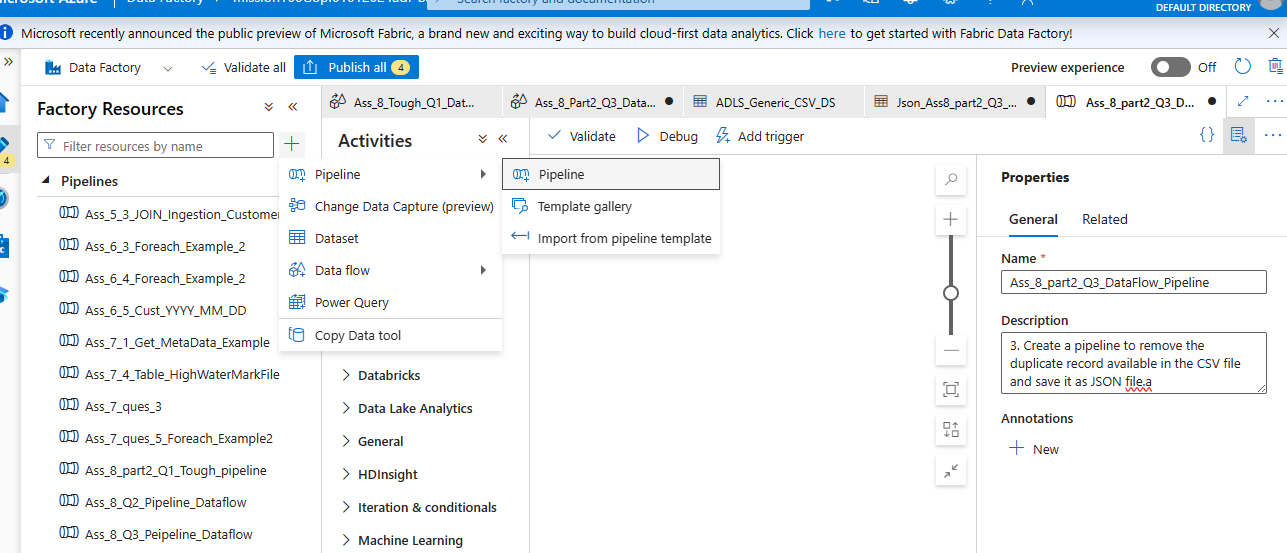


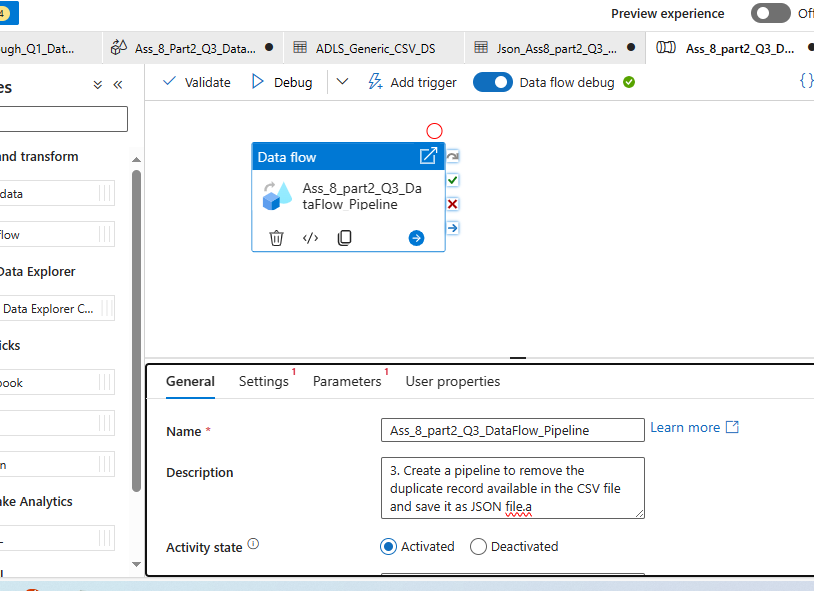
****

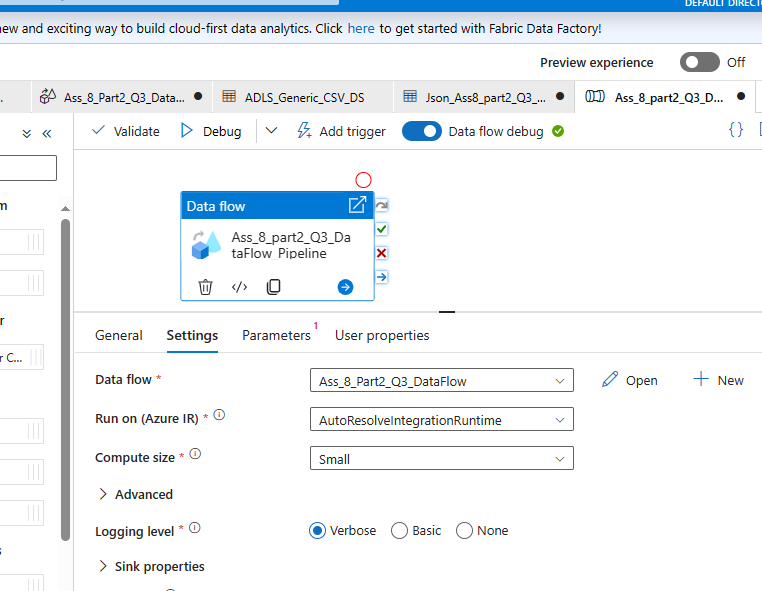
****

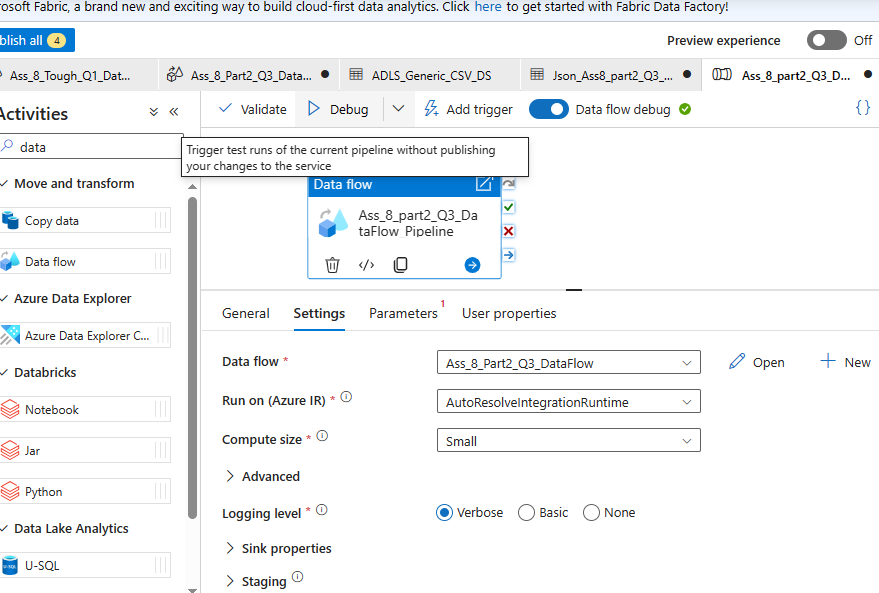
****

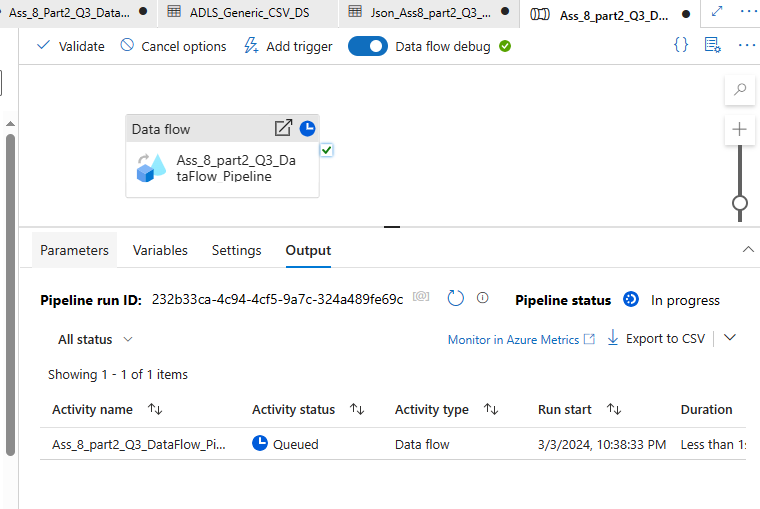
**Finally create a Pipleine to Execute a Daaflow:**

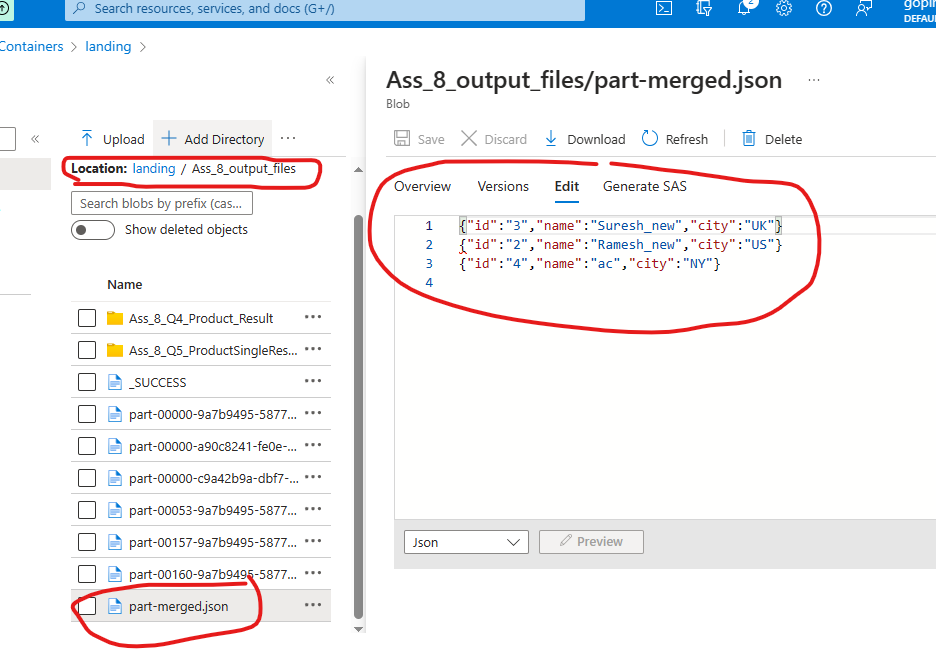
****

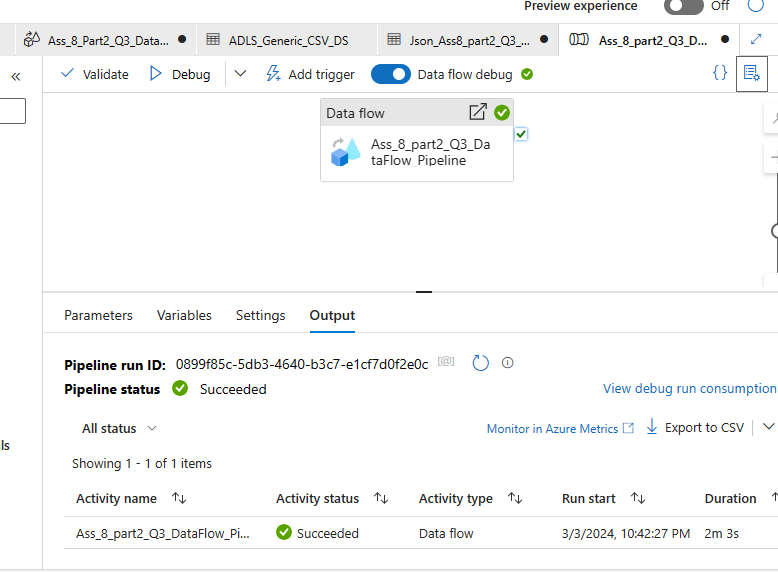
****

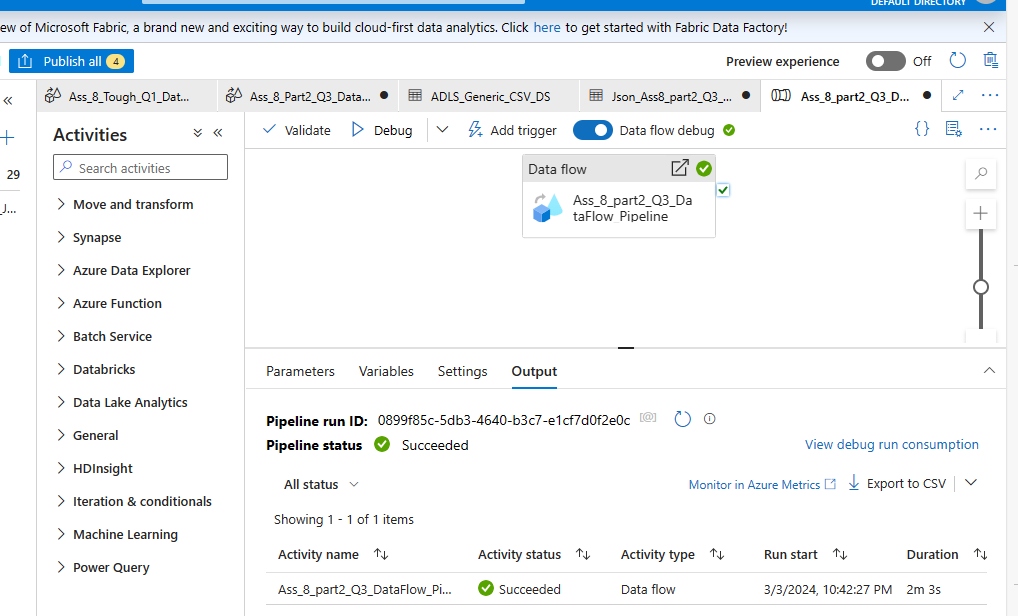
****

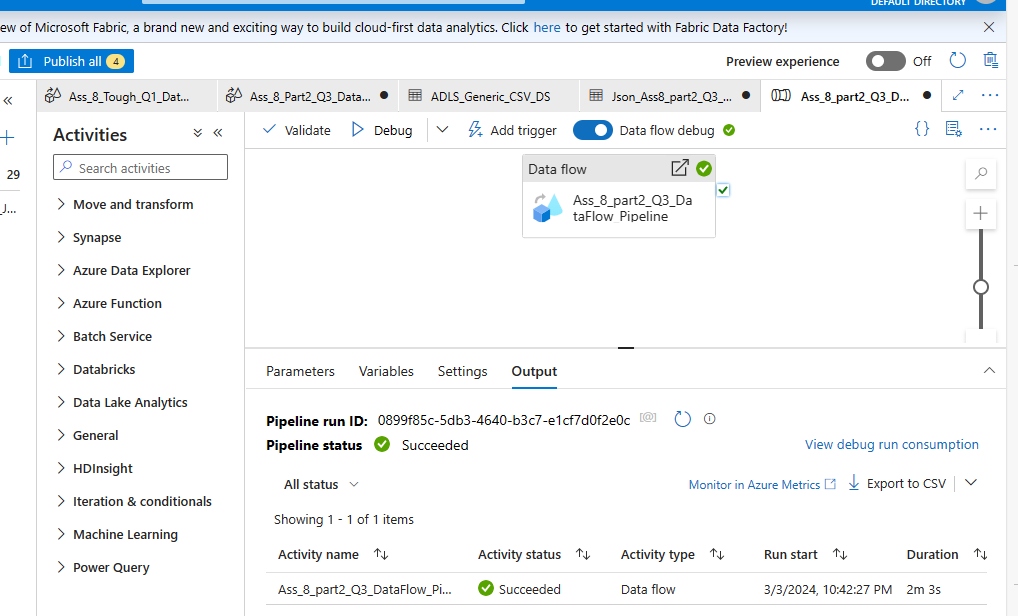
****

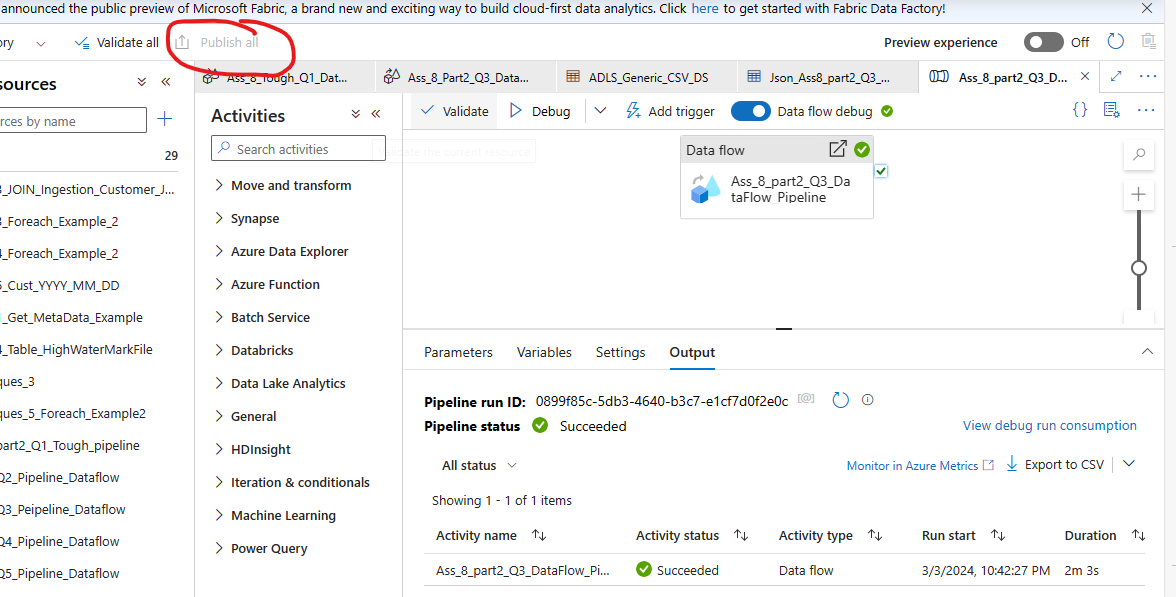
****

****

****

****

****

****