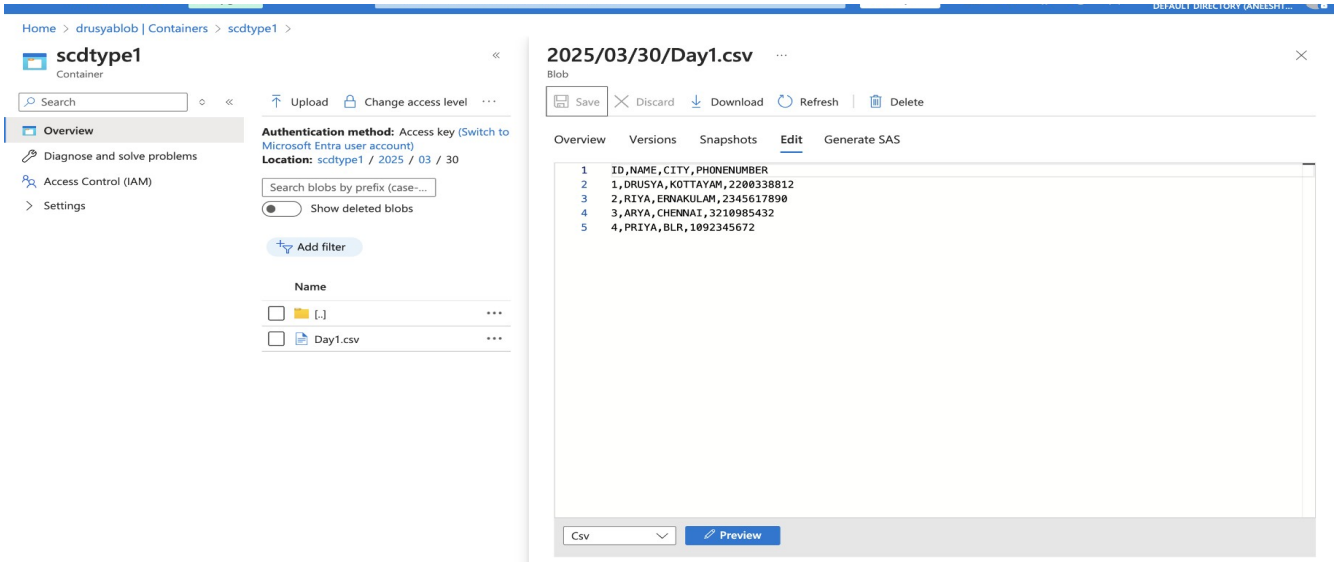
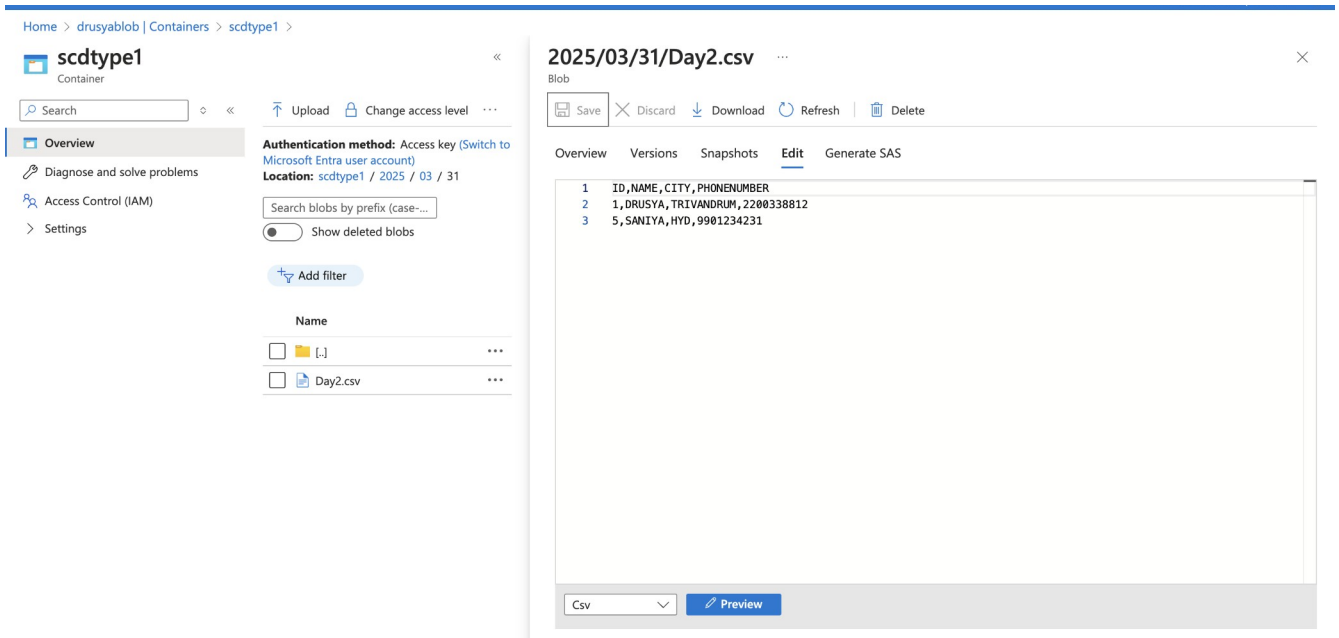
SCD TYPE 1 LOGIC ON DELTA TABLE IN DATABRICKS

Step 1:

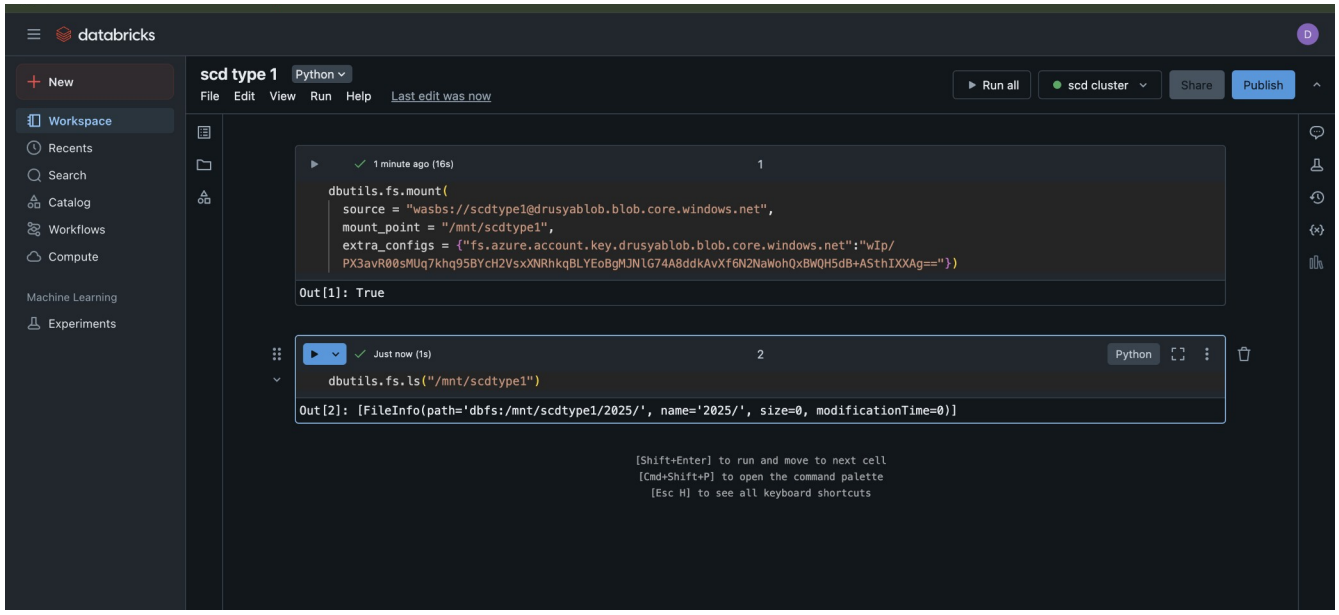
I have uploaded my day 1 and day 2 csv files in my blob storage account.





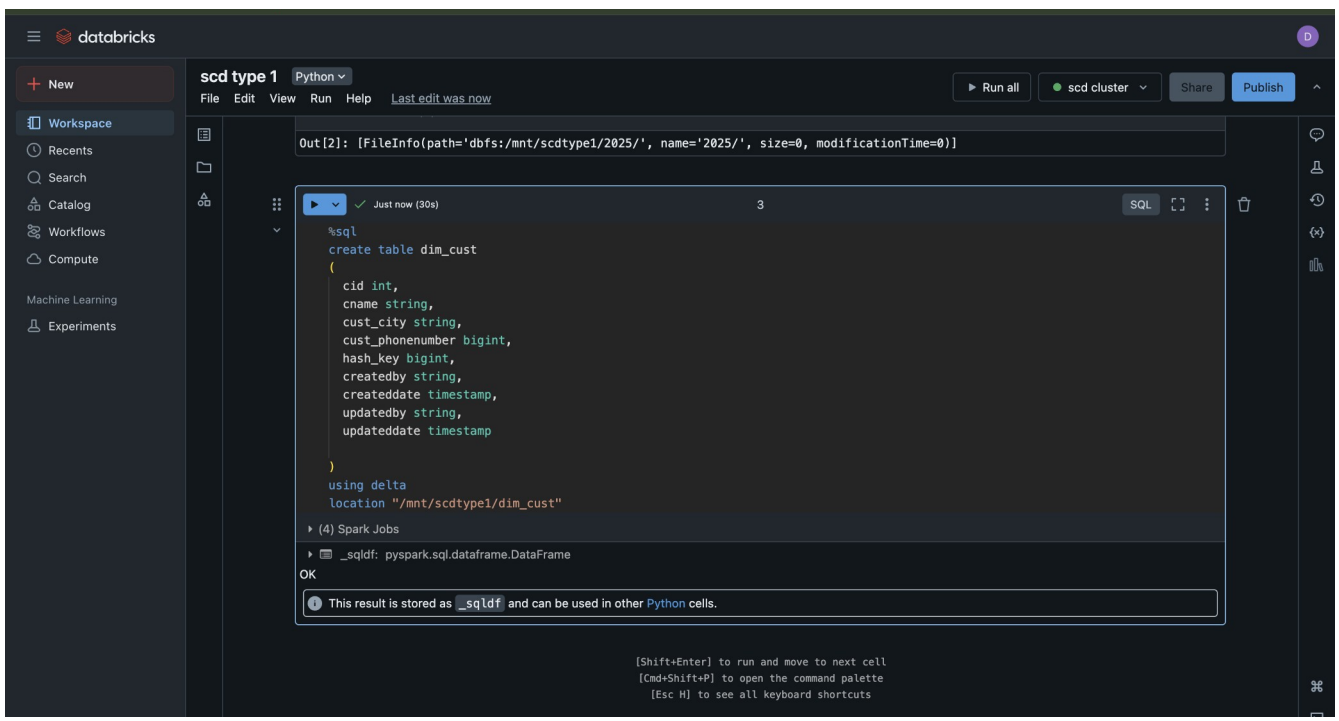
Step 2:

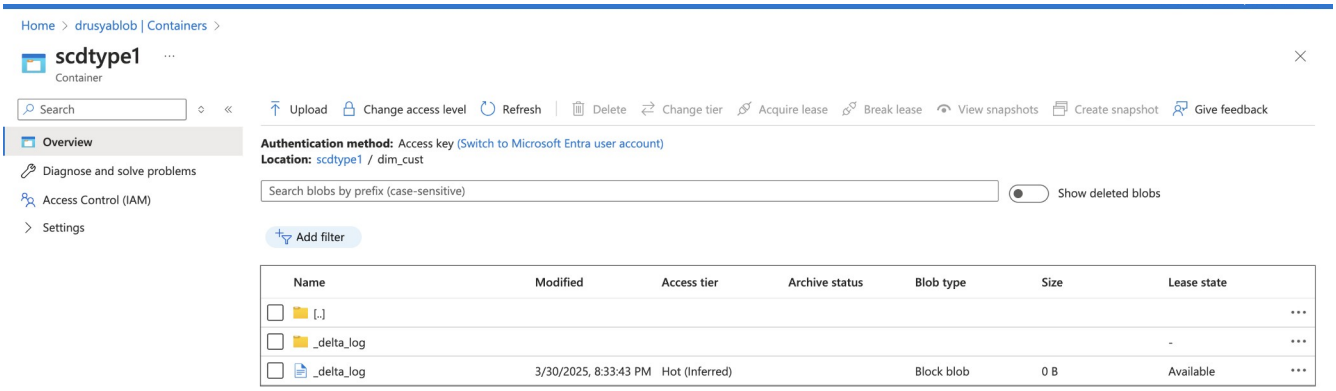
Mounted my storage account using the access key to the databricks community edition my account.



Step 3:

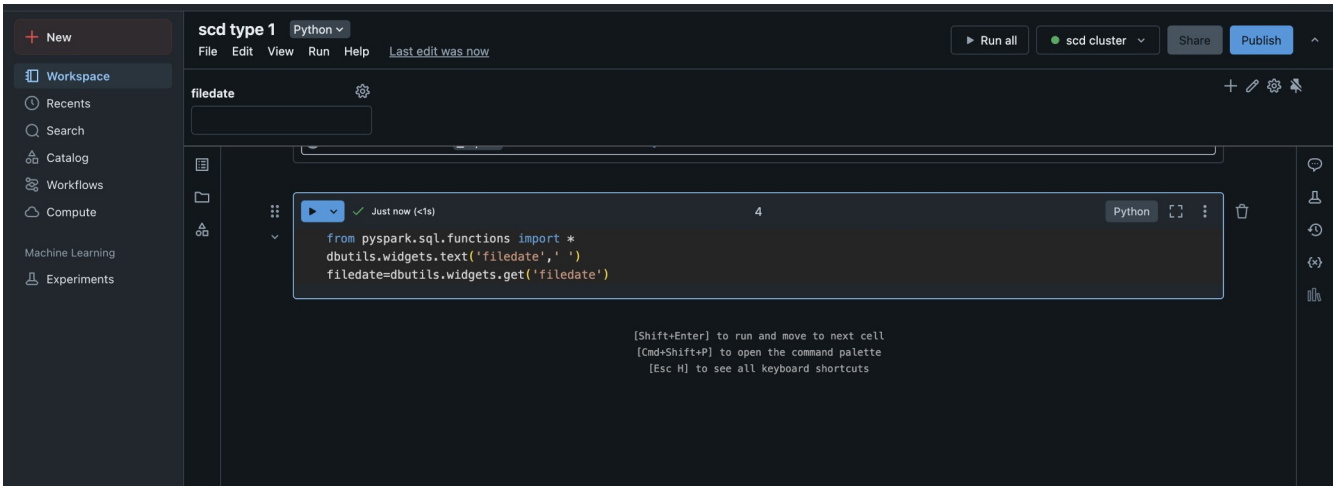
Creating a delta table in my account for storing the data.





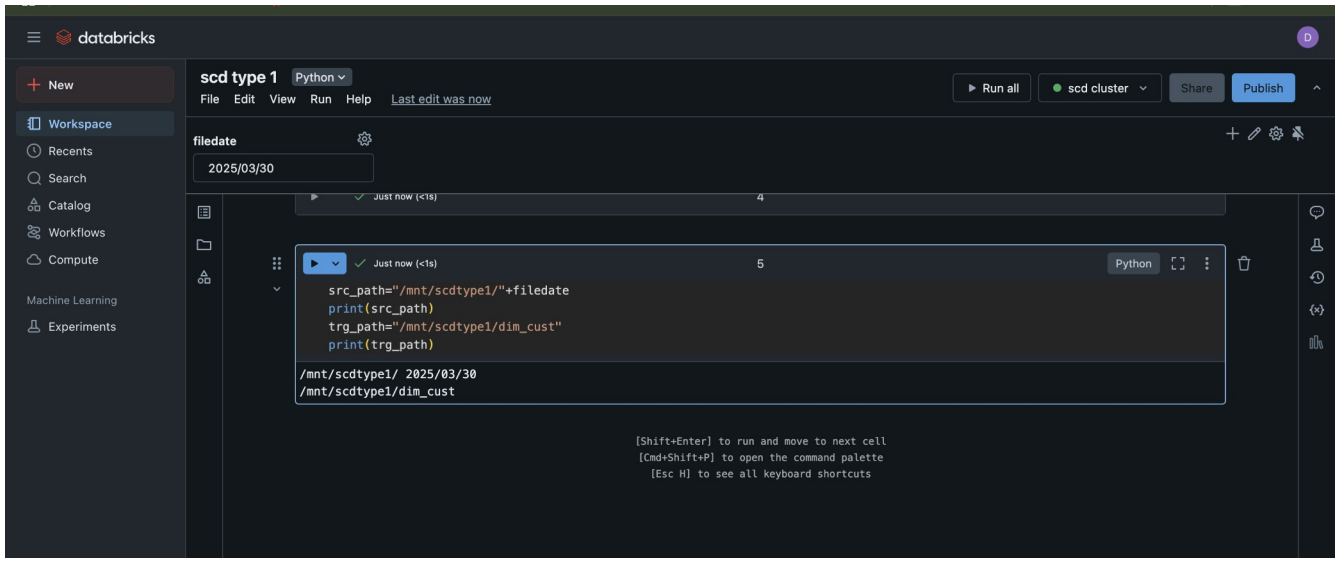
Step 4:

Created a widget as file date.



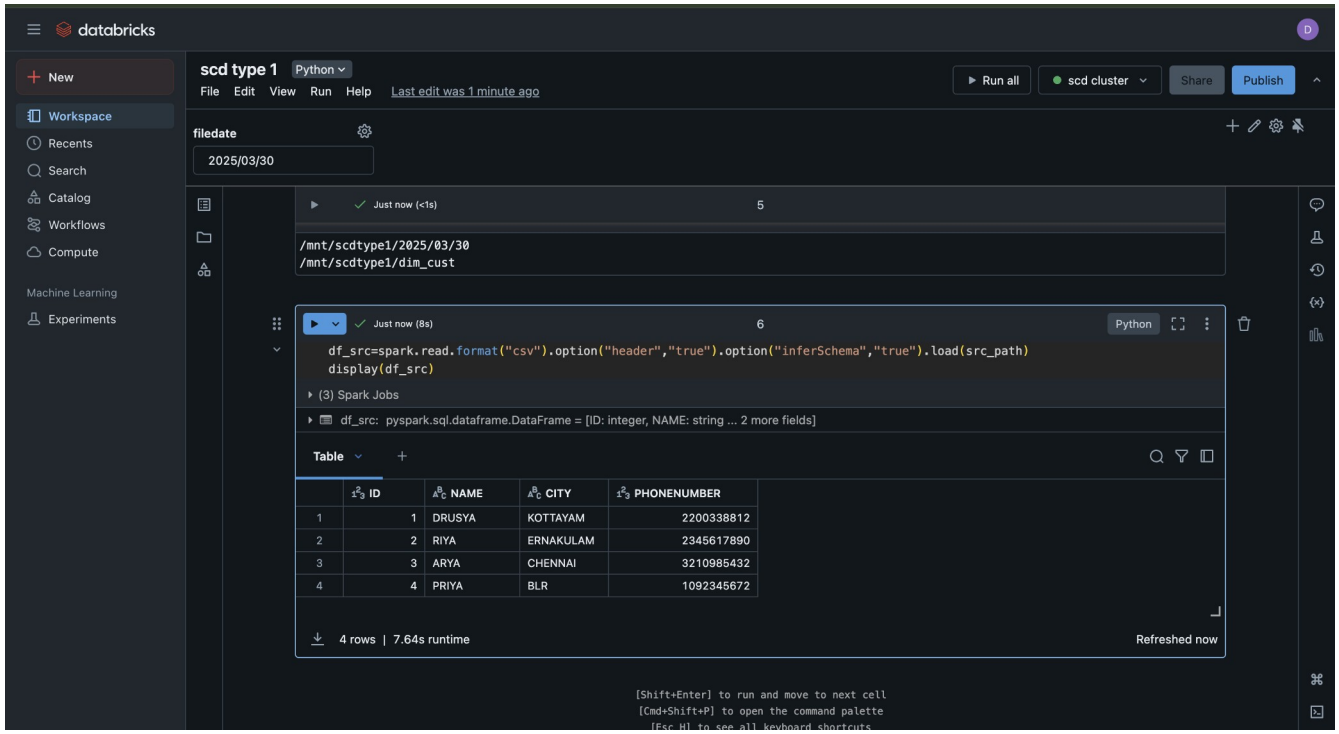
Step 5:

Now I have entered my source path and target path using the file date that I created earlier.



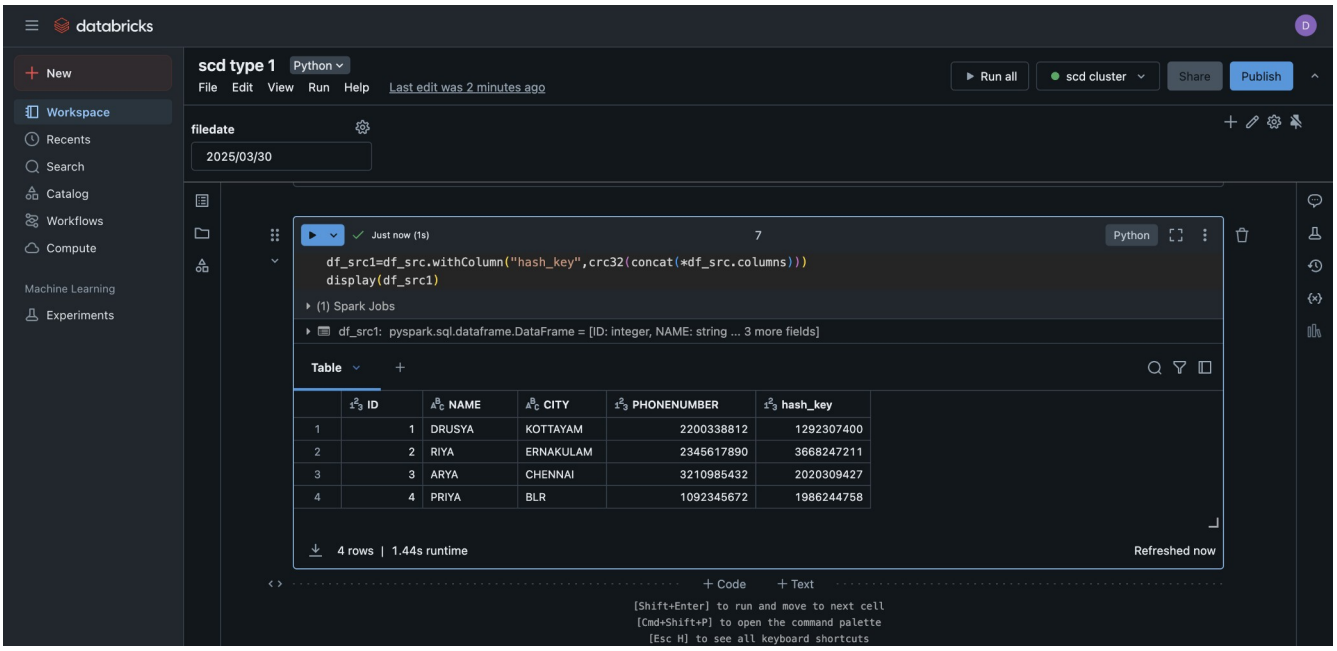
Step 6:

To dispaly my source day 1 file I use the read command.



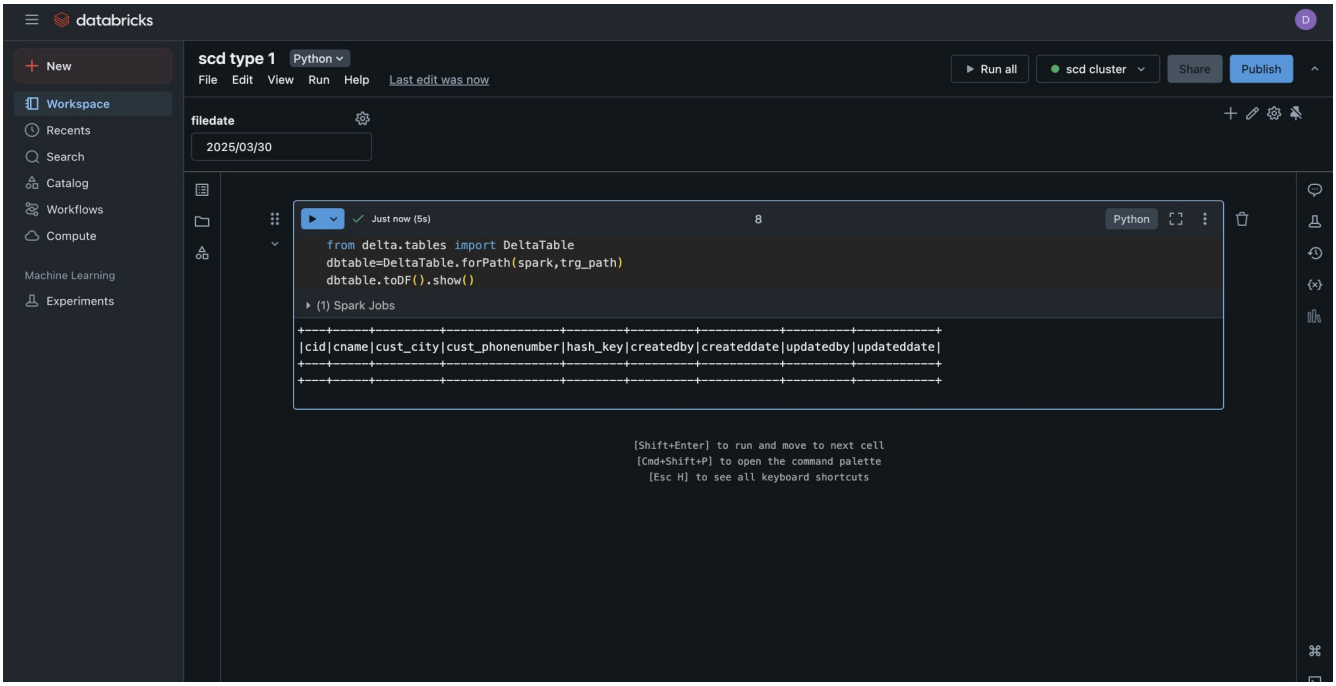
Step 7:

Now I add a new column haskey so that we can compare the data with other data files to check if it is a new one or updated one, with concating all the column values together to integer. For that I am using the crc32 command.



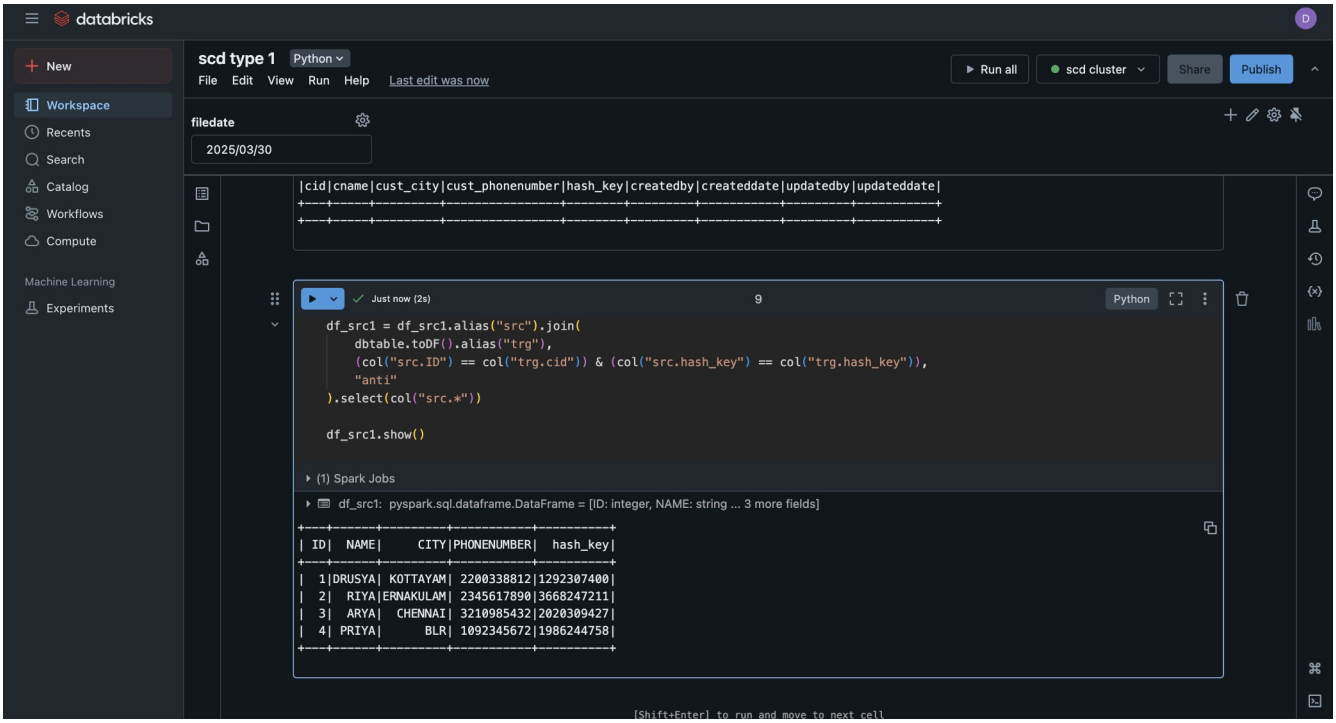
Step 8:

Now we are converting our target delta table into a dataframe and is desplaying it.



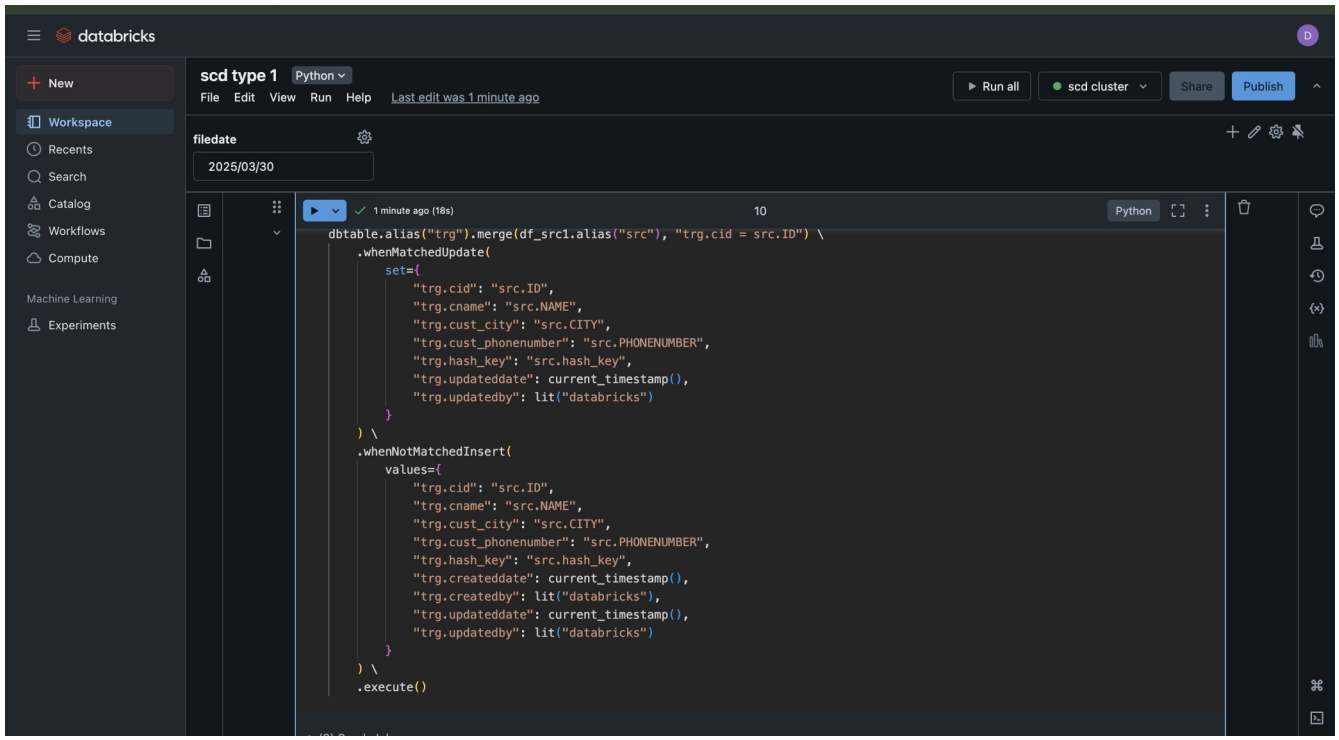
Step 9:

Now I join my source file and target based on the ID so that in the next step I can compare the hashkey and ID to determine it is a insert or update.



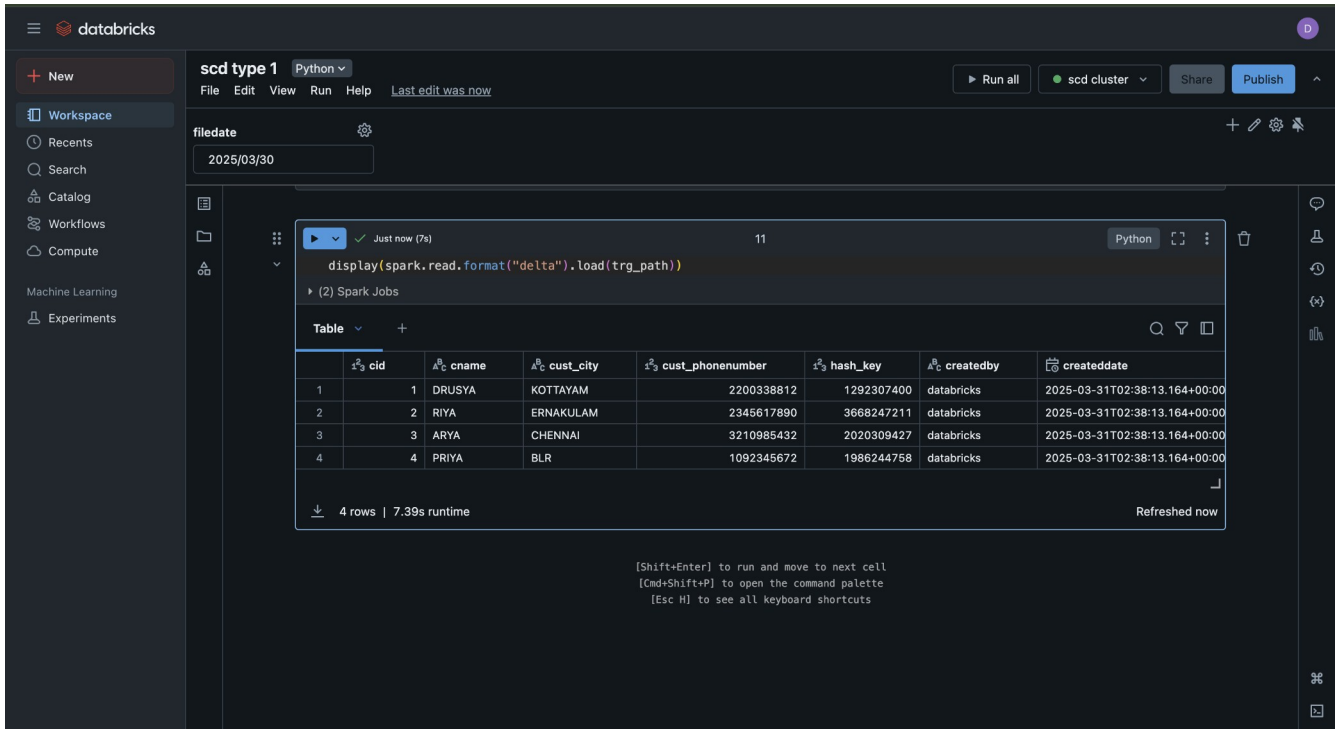
Step 10:

Based on update or insert new records condition merge the files:



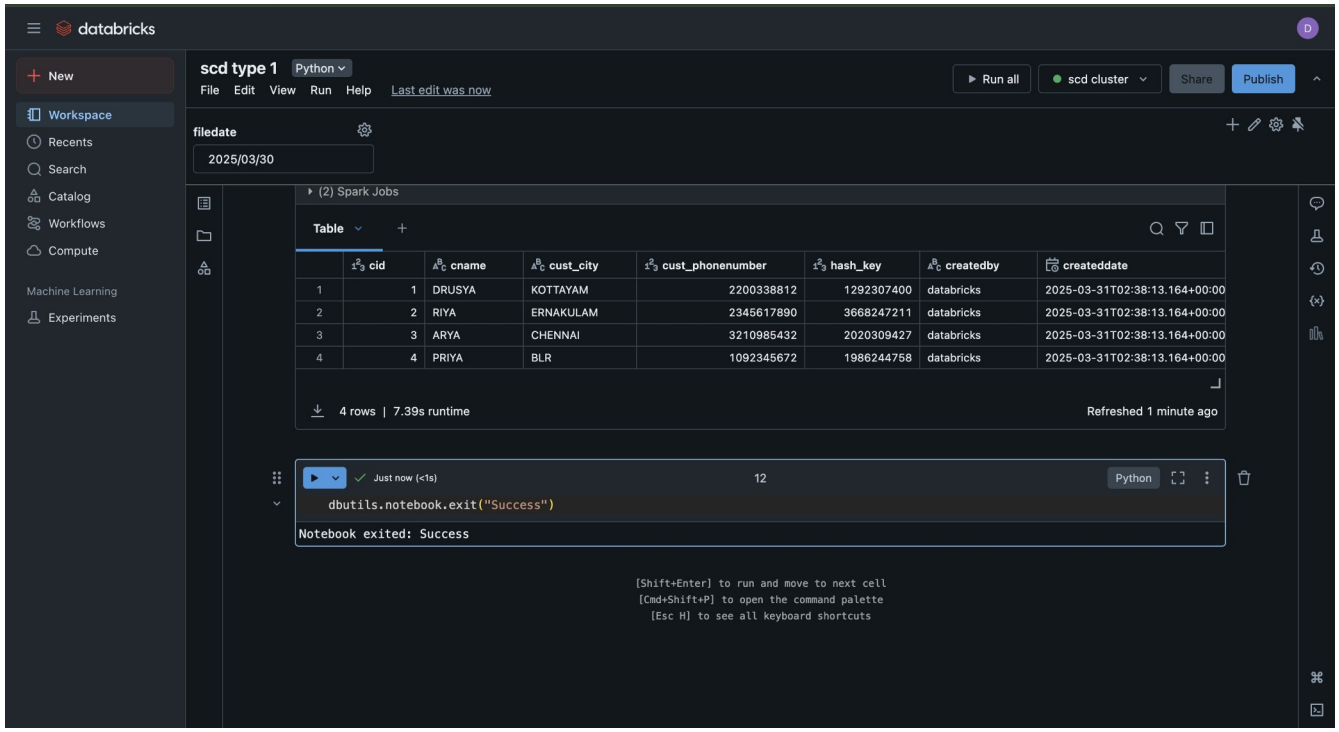
Step 11:

Now reading the target file.



Step 12:

Exit the notebook when success.



Step 13:

Passing day 2 file and check the output.