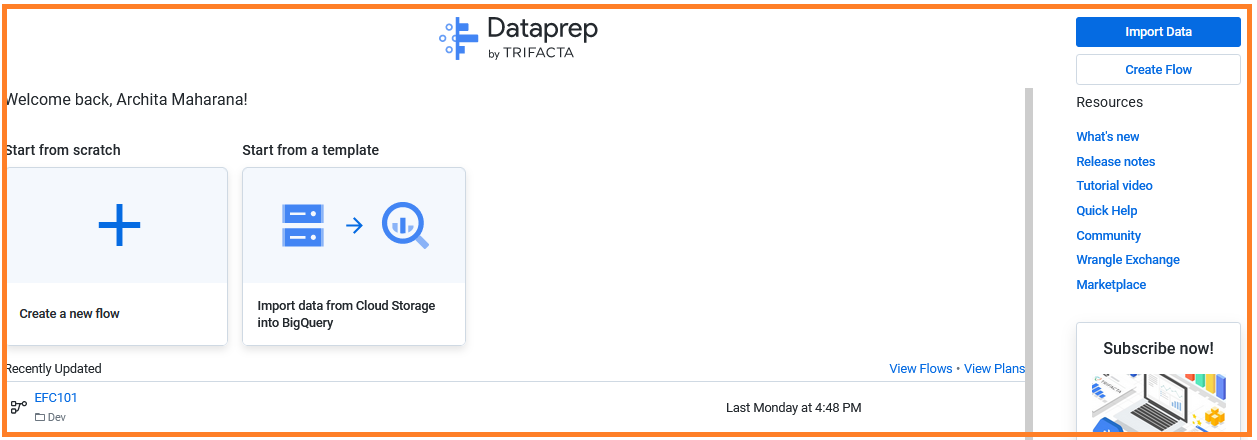
How to create flows in Dataprep

1. Extract data
2. Transform data
3. Load data

First go to BigQuery and select the project name.

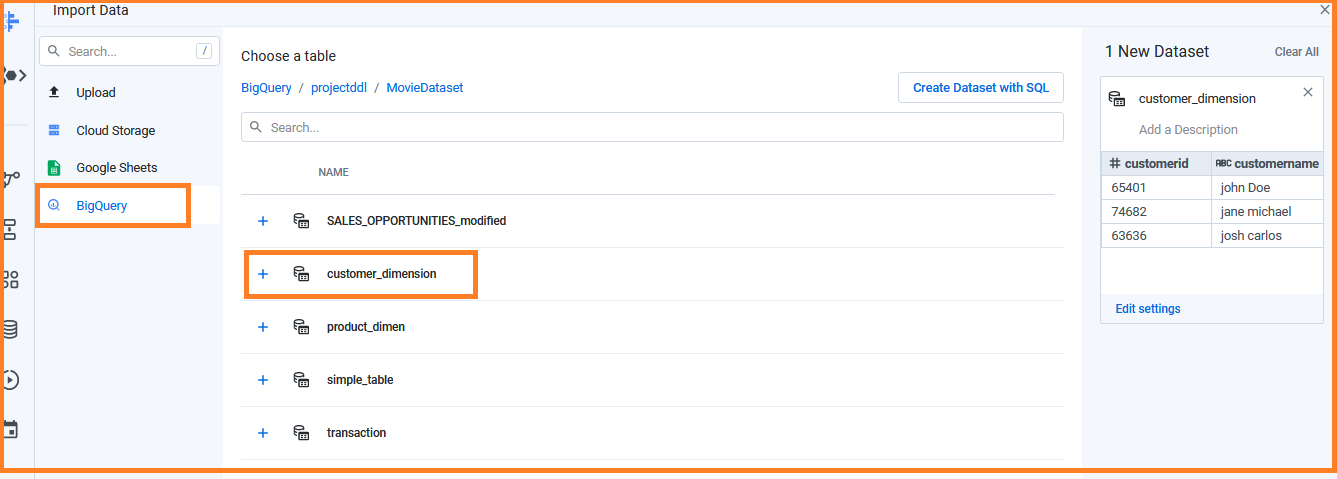
Goto Dataprep and connect.



Click on Import data

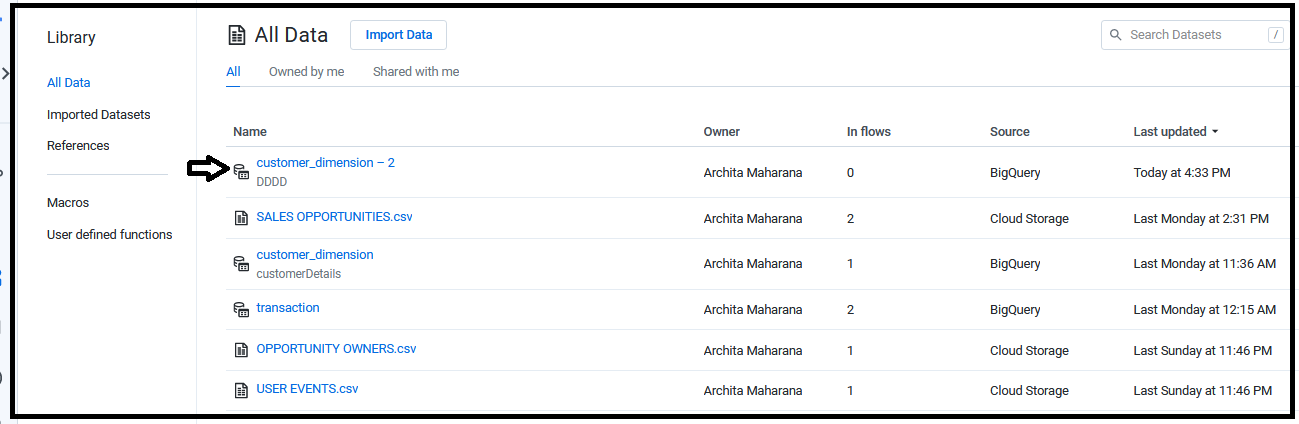
Select the option BigQuery

Select the dataset rom database

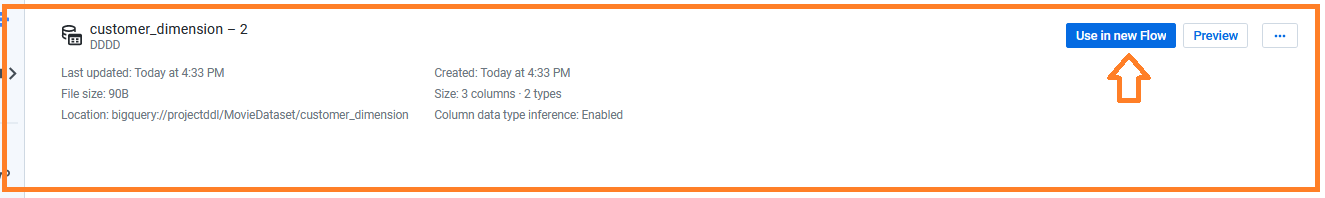


Click on Continue.

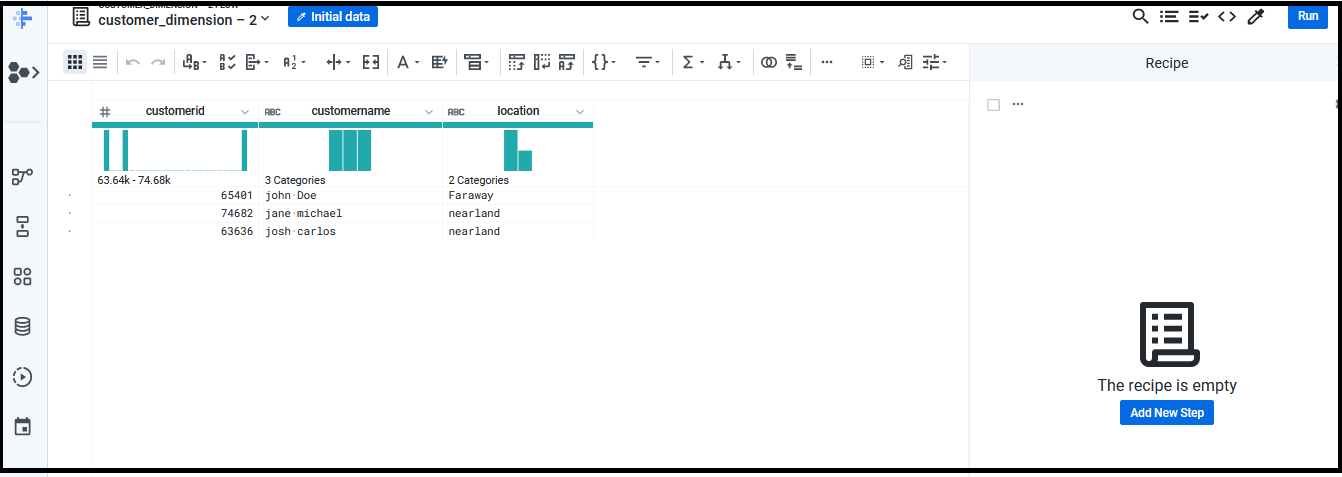
Click on the dataset(customer\_dimension)



The page will redirect to…



Click on Use in Flow.



Click on Add New Steps.

=============================================

Case-2

**Import data from cloud storage and add flow**

**Select the option flow**

Go to create

Create having 3 options as

* Blank flow
* Move data from cloud storage to BigQuery
* New folder

Select the option Move data from cloud storage to BigQuery.

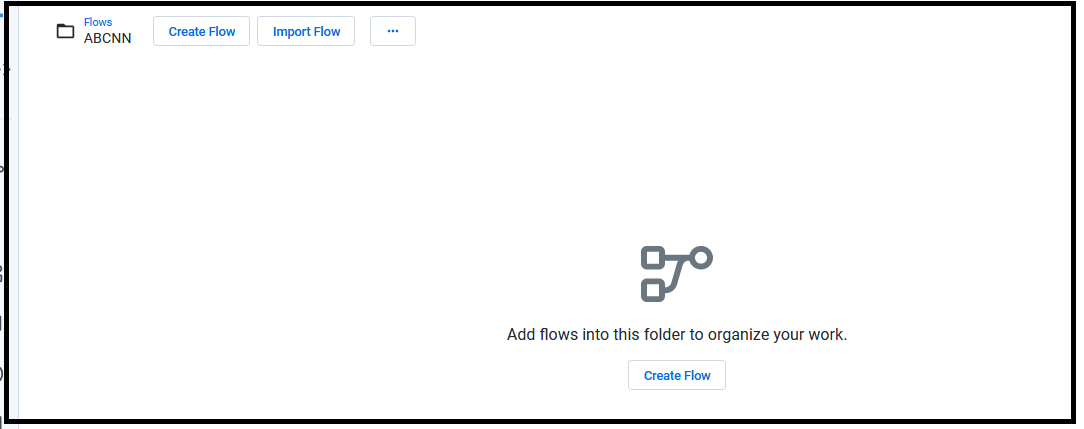
* Go to cloud storage bucket
* Select the job run option
* Select the dataset.
* Click Import and Add flow

Case 3

**New Folder**

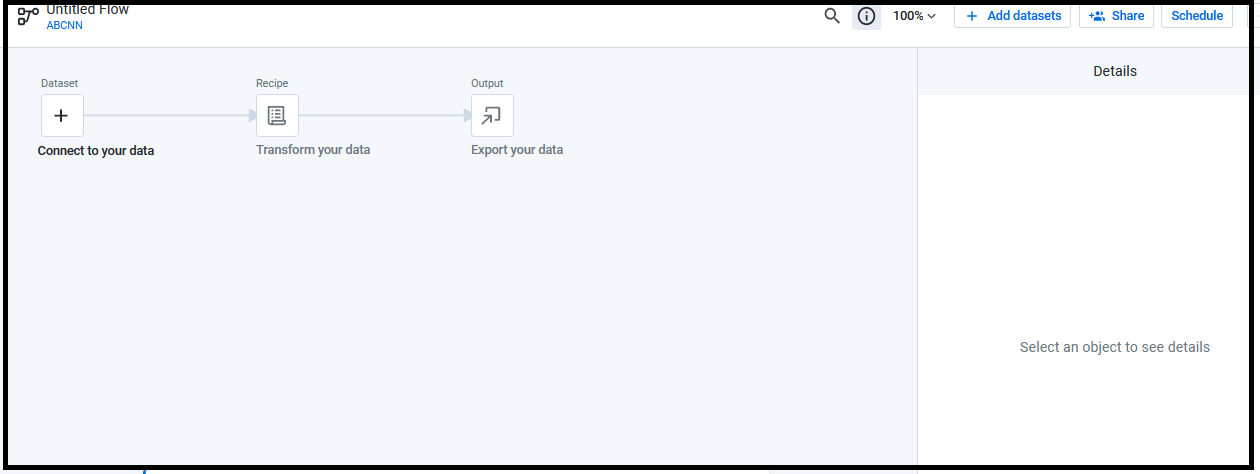
Create a new folder ABCNN

Click on the new folder(ABCNN)

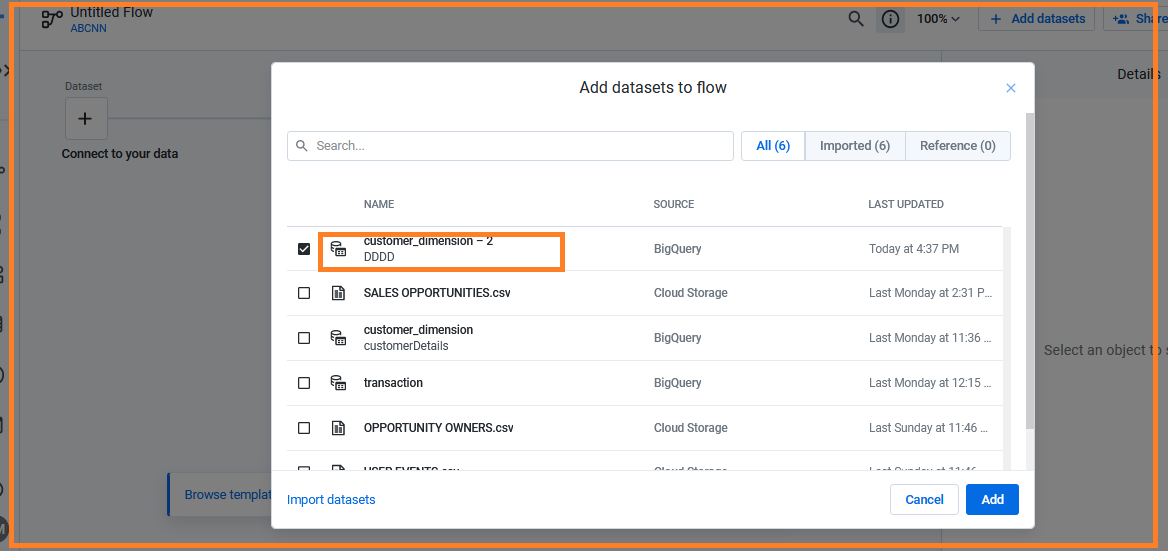


Click on create flow

The page redirect to …

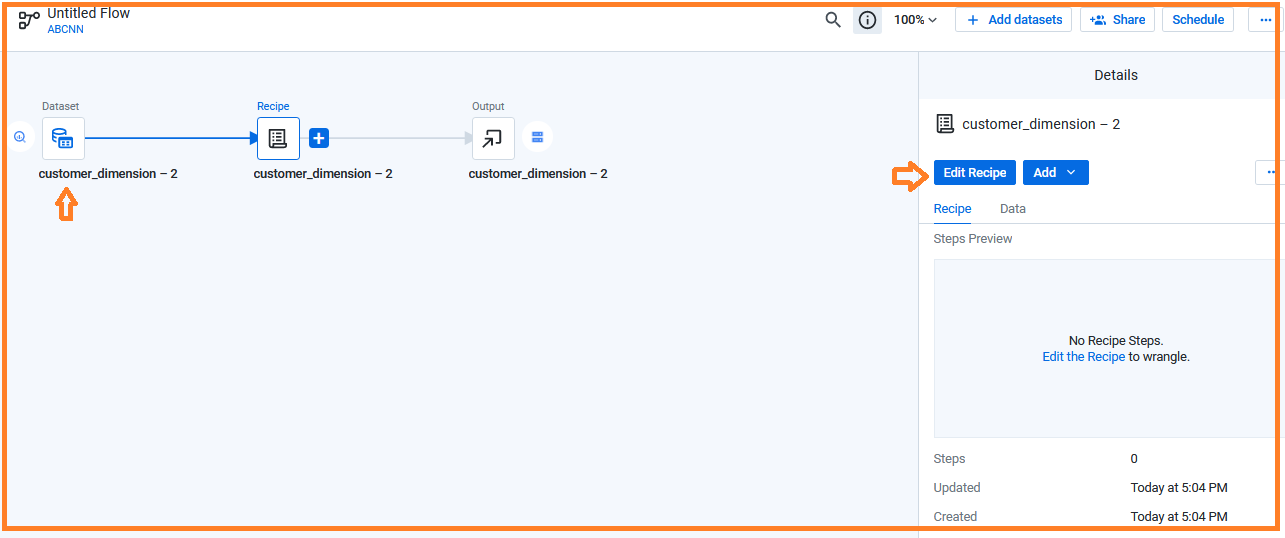


Click on connect to your dataset.



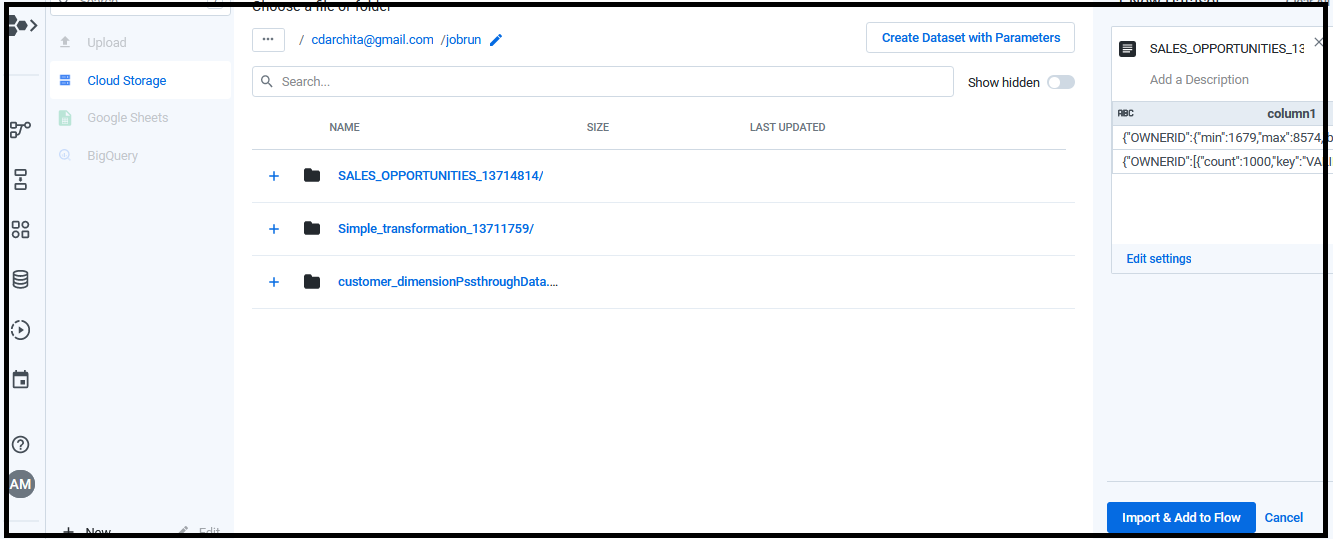
Click on Add.

Now we can add recipe to the dataset and to some transformations.



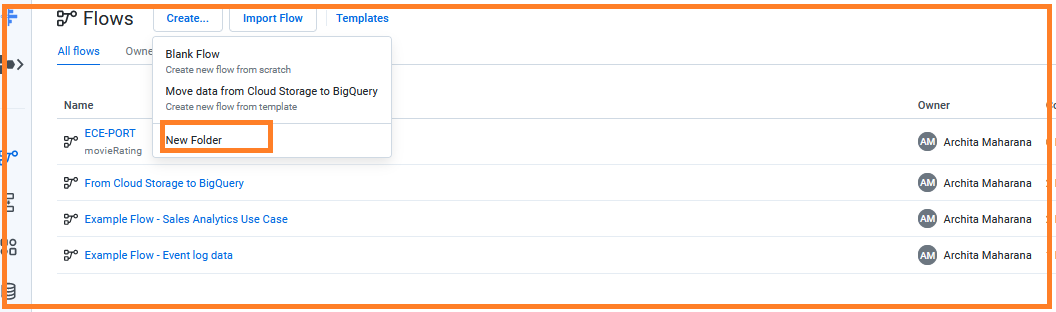
=====================================

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

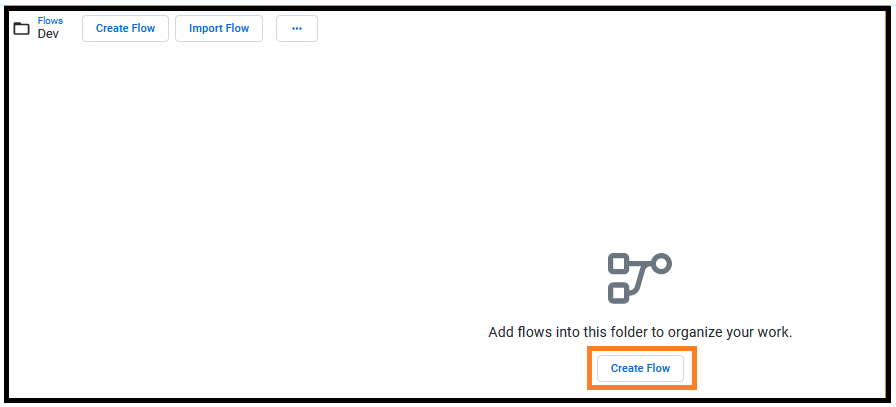


Here we will extract data from BigQuery and load data into Google cloud storage.

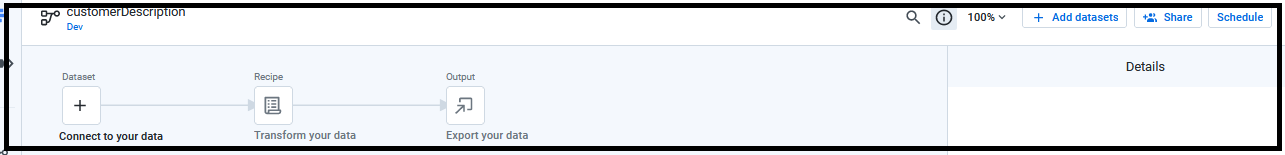
* Go to dataprep and create the job.
* First create a folder



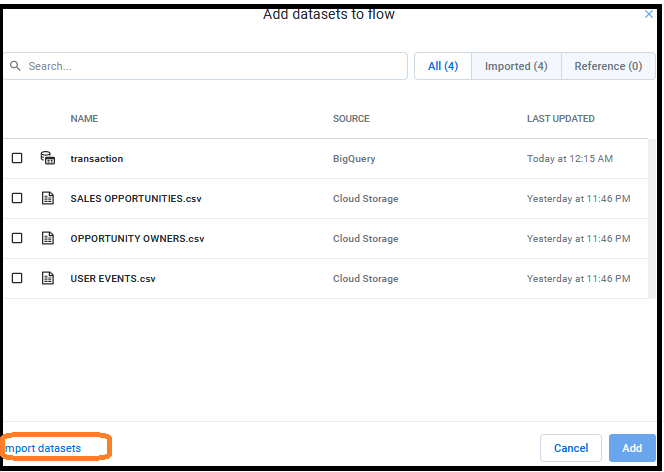
* Go to Dev folder and click on create flow.



* Click on create flow and assign a name to it.
* **Note**: before u select a dataset from BigQuery for flow/transformation, first select a valid table from project/dataset in BigQuery.
* Then go to dataprep flow and do the above steps.
  + Click on create flow.



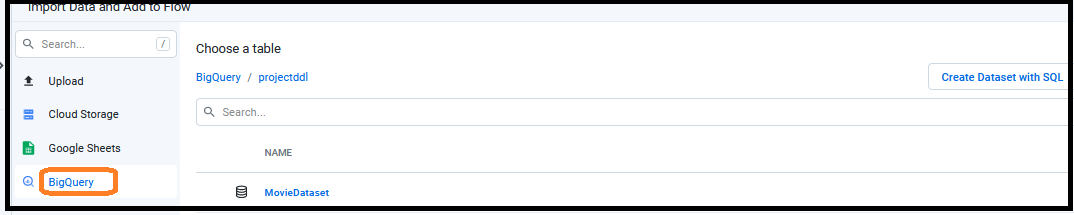
* Click on Add dataset



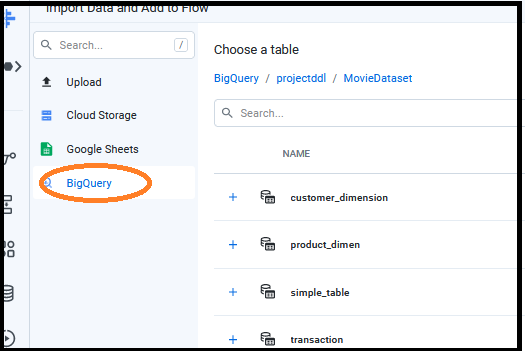
* Click on Import datasets

**Note Add** option adds the dataset to flow…recipe format.

* The page view comes as below and then select BigQuery option from left column bar.



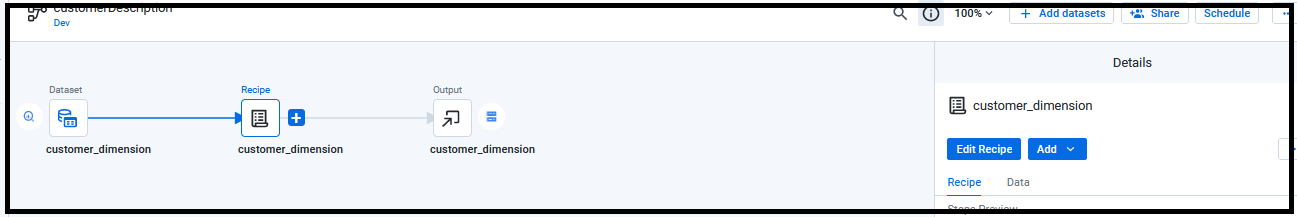
* Click on database “MovieDataset”
* From here you can select your respective dataset or a table.

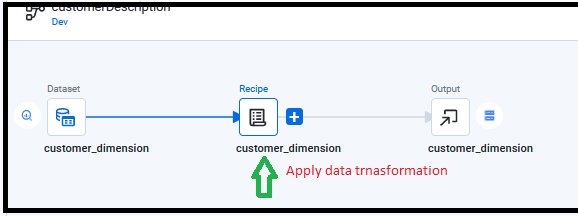


* Click on the dataset(which I have chosen as “customer\_dimension”
* U can do some changes on right side of your selected table

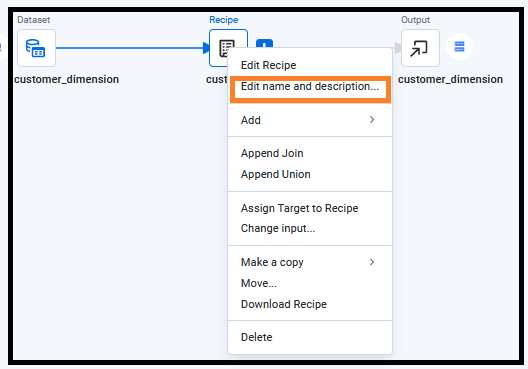


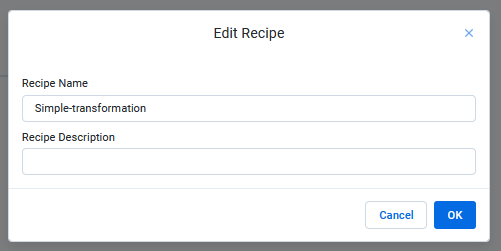
* Click Import and Add to Flow.
* The view comes as





* Right click on Recipe(customer\_dimension) and select the option Edit name and description.



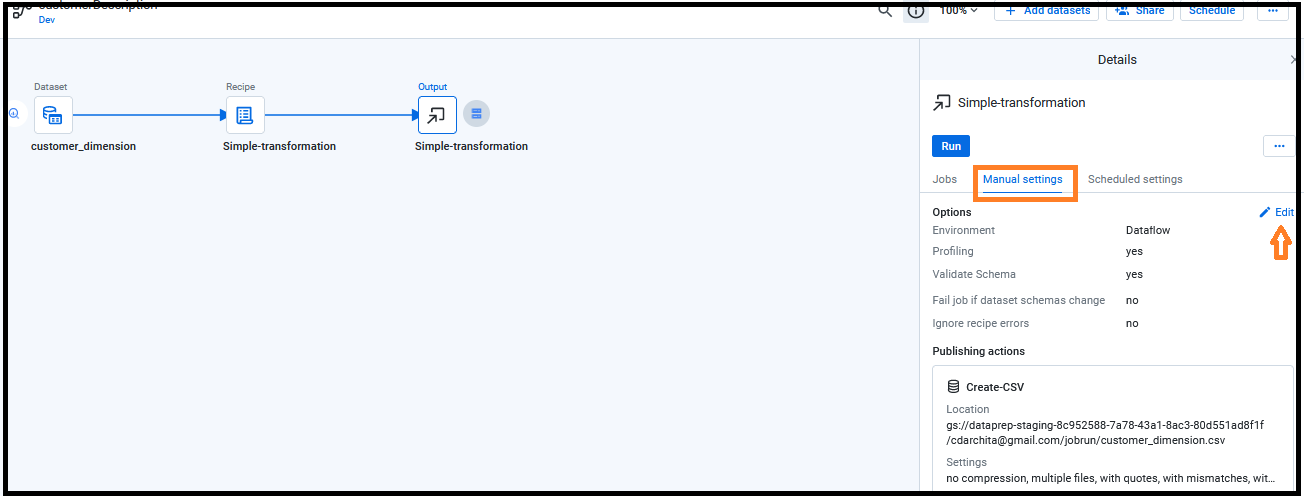


* Click OK

Publishing action

Here we can change the file format.

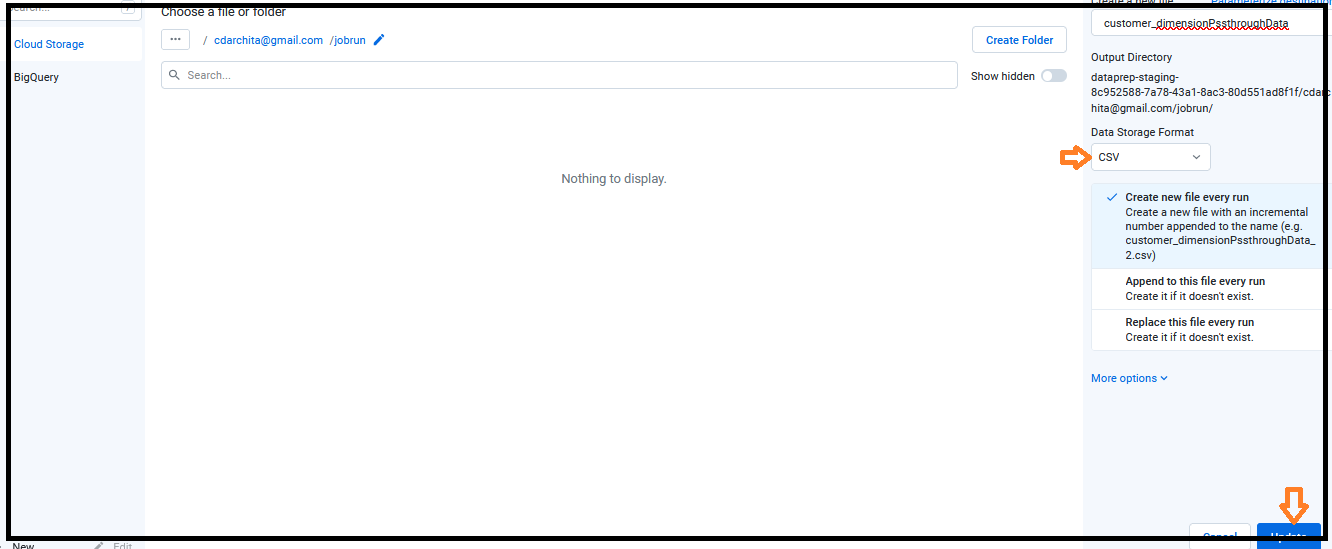
* Click on output channel
* From right bar option we can select manual operation….it has the options as
  + In which google storage location you want to save your job
  + If you want to change your file format.



The view page shows as



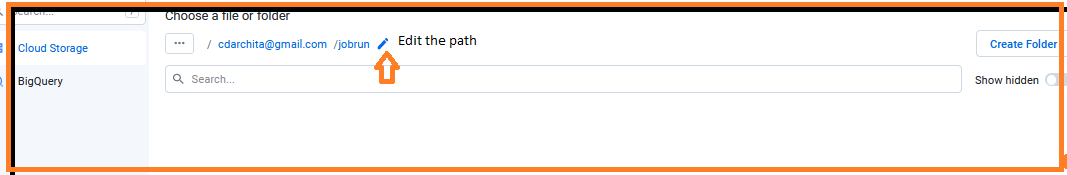
* Click on Edit option.



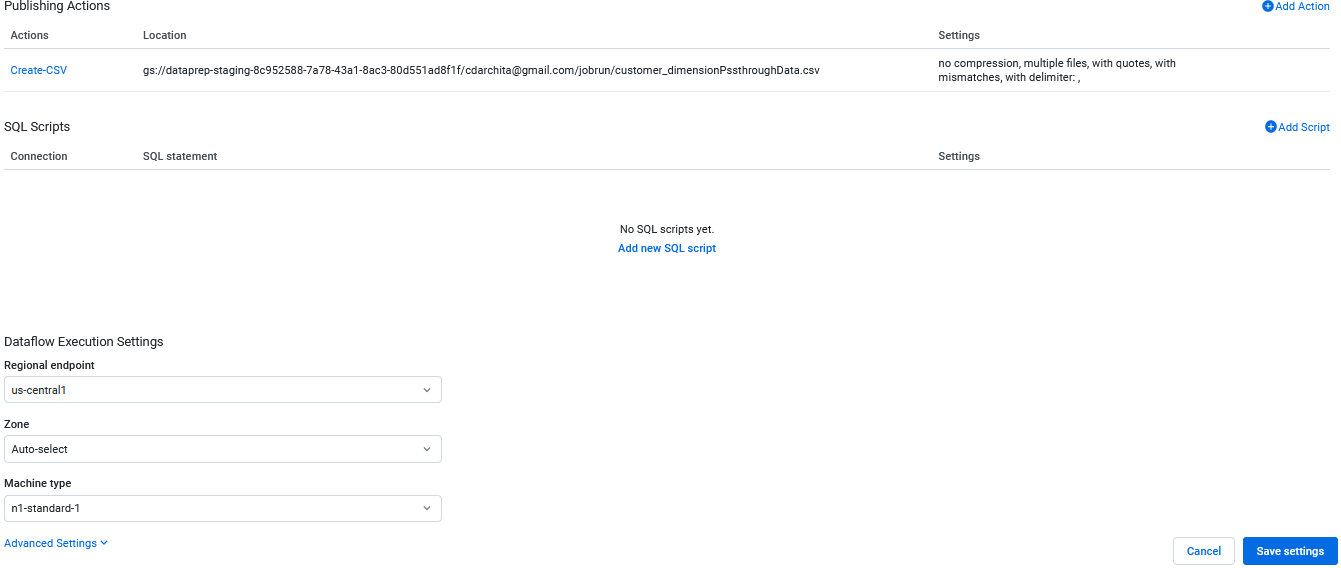
* Click on More options
  + Delimeter( give ,)
  + Compression
    - Choose the option according your need

You can check from above left corner that the file stores in cloud storage under “jobrun”.

If u want to edit then click on Edit or if you want to create a folder then do changes.



* Click on update.
* The flow shows as

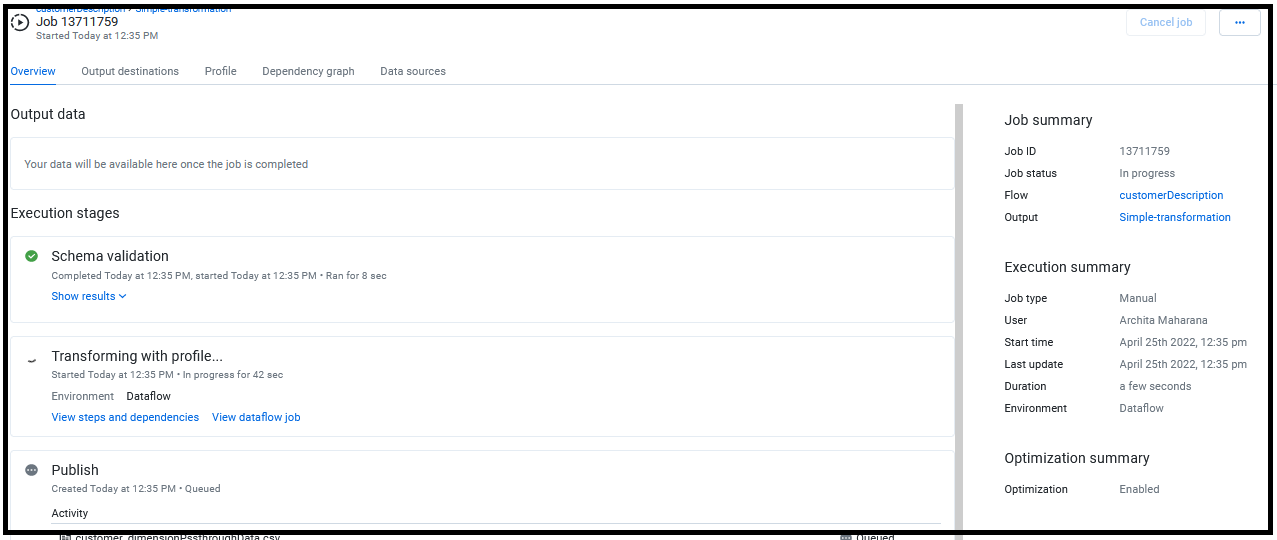


We have the options as

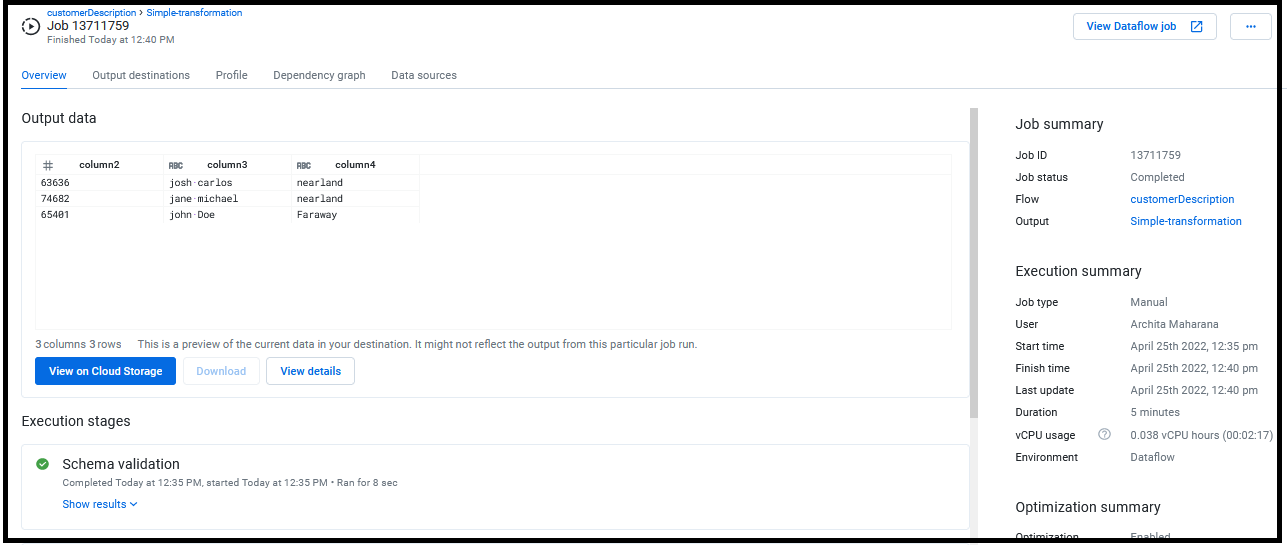
* Regional explore
* Zone
* Machine type

Once you have done everything click on save settings.

* Click on Run jobs.

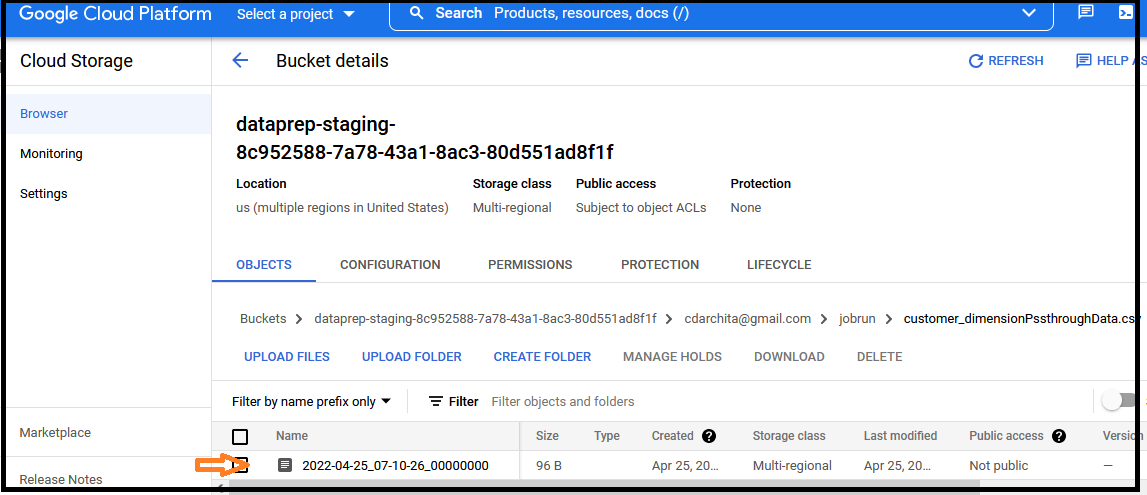


It takes some time to complete all these steps. Once the job is completed the output shows as

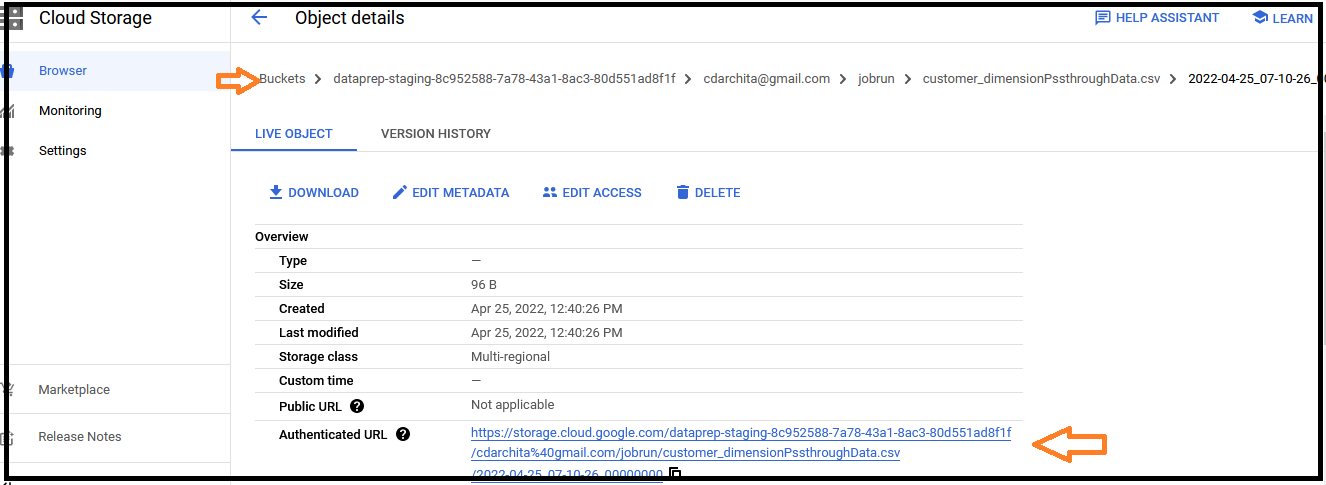


Now go to your google cloud storage and check your file

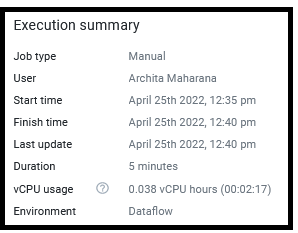
Click on view on cloud storage



Click on the file, we will get details about the file created in CLOUD SORAGE.



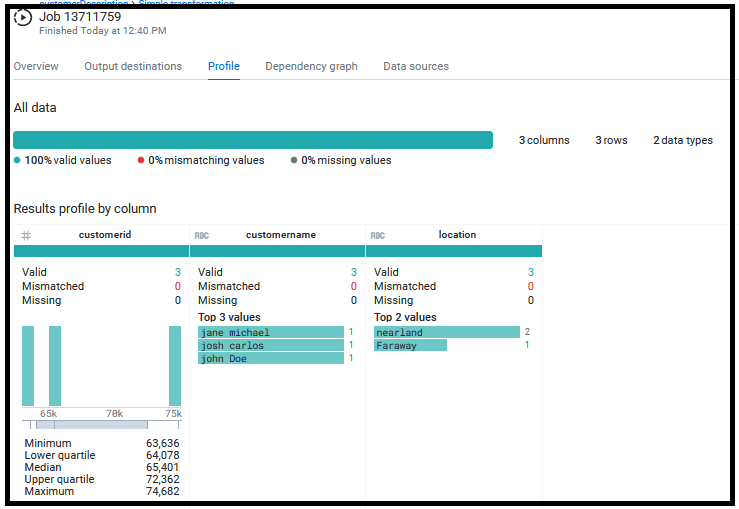
You can click on the link and download the file.



Go to dataprep dashboard and check the options

* Overview
* Profile
* Data sources
* Dependency graph

Under profile option we get the details about dataset…as



End