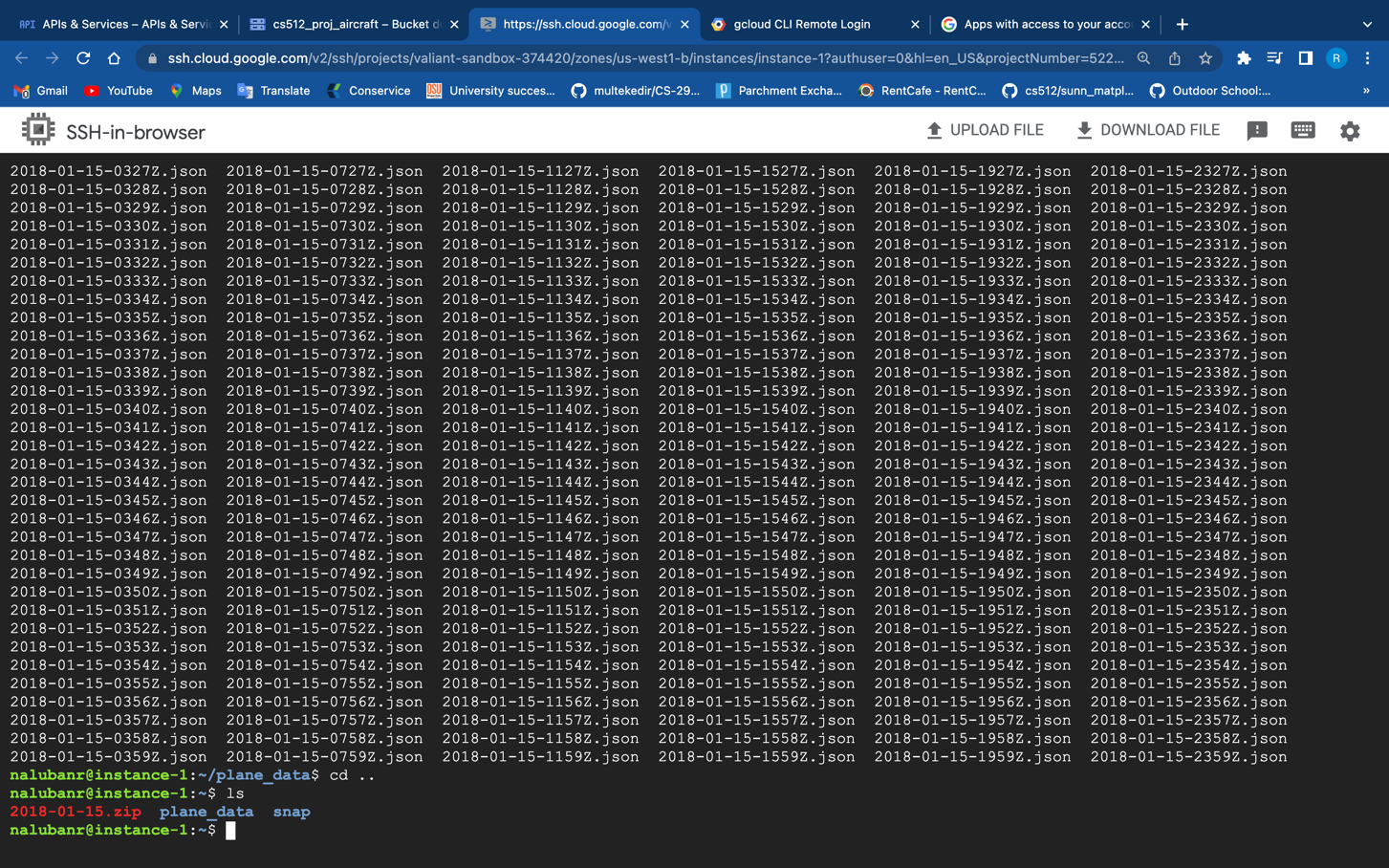
Big Query Assignment

Step 1. Get Zip file onto Compute Engine Instance.



Step 2: Load JSON files into Google Cloud Storage

Graphical user interface, text, application

Description automatically generated

Step 3: Load JSON files as a data set into Google Cloud Data prep.

Graphical user interface, application, Word

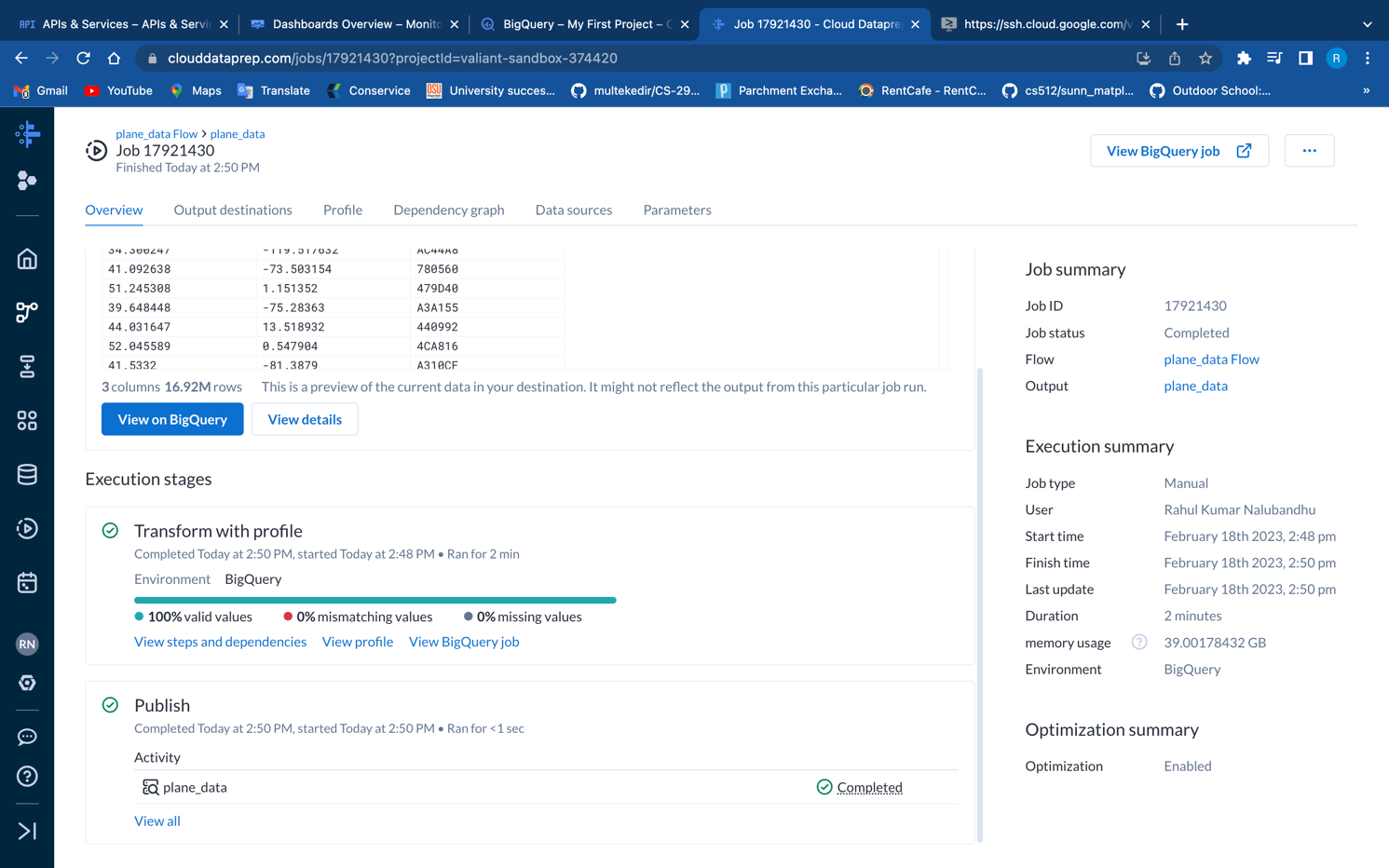
Description automatically generated

Step 4: Parse JSON into appropriate columns in Dataprep

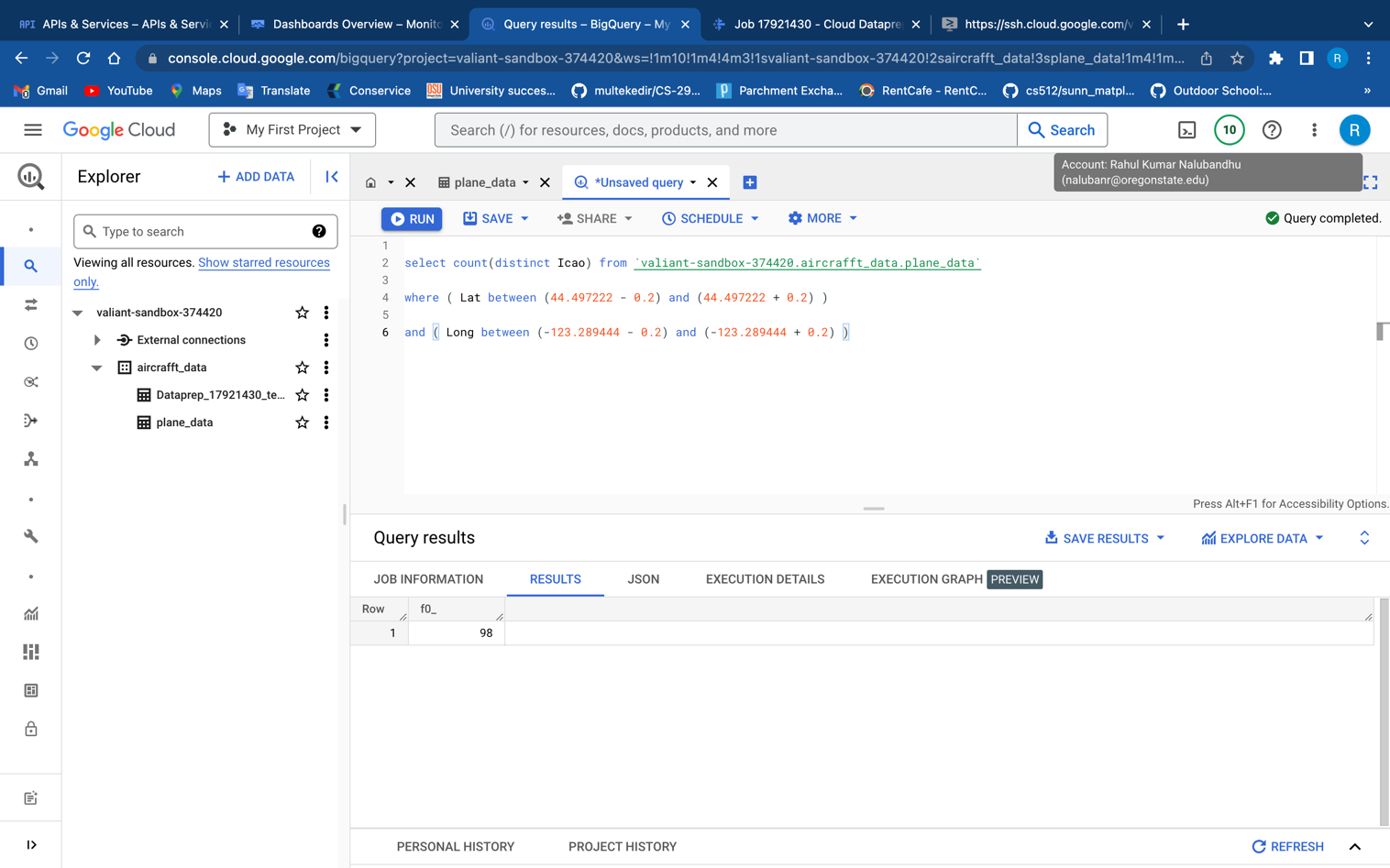
Graphical user interface, application, table, Excel

Description automatically generated

Step 5: Export DataPrep job into BigQuery



Step 6: Write Big Query SQL to compute the answer.



Detailed Description of the process :

First, I accessed the Google Cloud Console and created an instance in the Compute Engine. Initially, I followed the steps in the video, but the default machine size was running too slowly. So, I stopped the instance and updated it to 2vcpu and 8gb, which resulted in better performance.

Once the instance was created, I logged in via SSH and installed wget and unzip. I created a directory called "plane\_data" and used wget to download the aircraft\_data file, which took around 12-15 minutes. Once downloaded, I unzipped the file in the same directory and moved the downloaded .zip file to the root directory. After obtaining all the .json files, I copied them to a cloud storage bucket that I created in the Google Cloud Console, called "cs512\_proj\_aircraft."

Once the copying was complete, I paused the instance and loaded the JSON files as a dataset into Google Cloud Dataprep. However, I encountered an error in the browser in JSON format that said, "Error: socket hang up." After researching and discussing with my partner, we discovered that it was a server issue, and we tried again later, successfully loading the dataset into Google Cloud Dataprep.

After loading the dataset, I encountered difficulty previewing the data and sought help in the ED discussion forum. With the help of Joshua Magana's advice, I resolved the issue and was able to view the data. Next, I created a data flow to filter all the unnecessary columns and rows. I followed the video and successfully created all the recipes but had difficulty in creating the first recipe because I couldn't find the transformation "delete row." After selecting the first column to delete, I found the "delete row" recommendation under the recipe, added it to the step, and updated it according to the video. From there, I added all the necessary steps and filtered the data, and I run it to Export DataPrep job into BigQuery.