| NGUYỄN HUY BÁCH  **IMPORT DATA FROM NDJSON FILES INTO TABLE**  USER GUIDE |  |
| --- | --- |

horizontal line

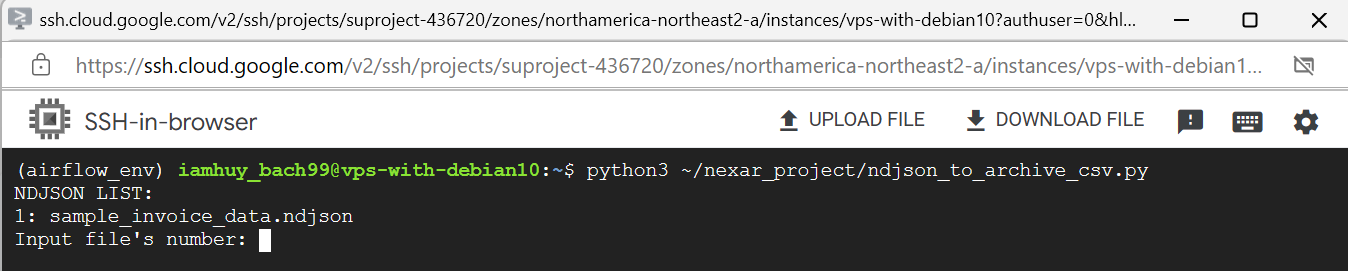
# Program 1: Convert data from NDJSON files to CSV files and compress

## Manual run:

**Run manually, User can select the file to be converted**

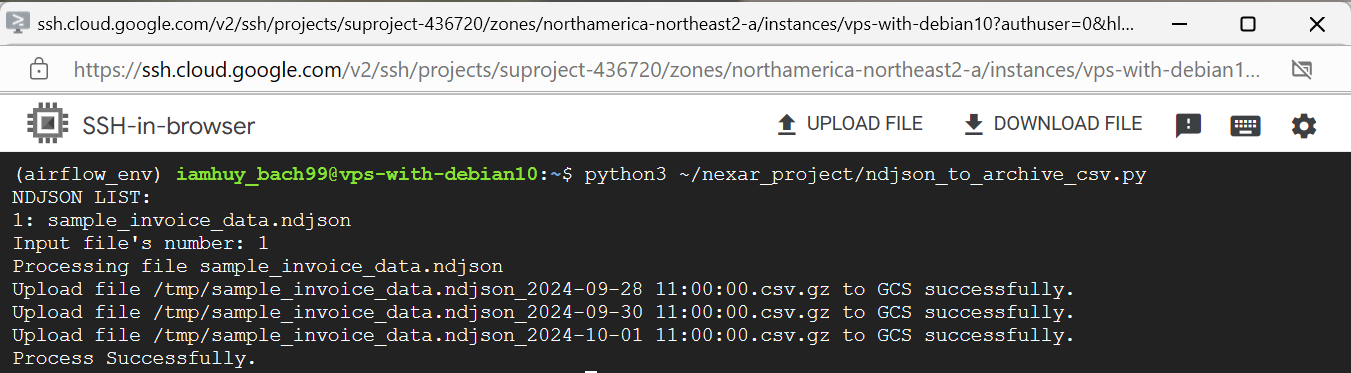
**Step 1:** Run by command “python3 ~/nexar\_project/ndjson\_to\_archive\_csv.py” or “python3 ~/nexar\_project/ndjson\_to\_archive\_csv.py manual” .

Screen will show list file of NDJSON in bucket and user can choose to be converted.

****

**Step 2:** Choose file will be converted.

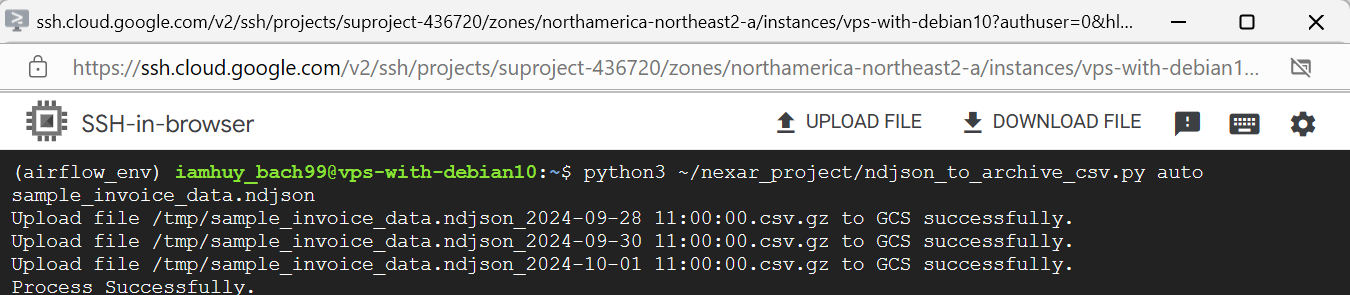
After selecting the file to convert, ndjson file will be converted to csv and split based on the timestamp (at the time of data addition), and the csv files will be compressed.



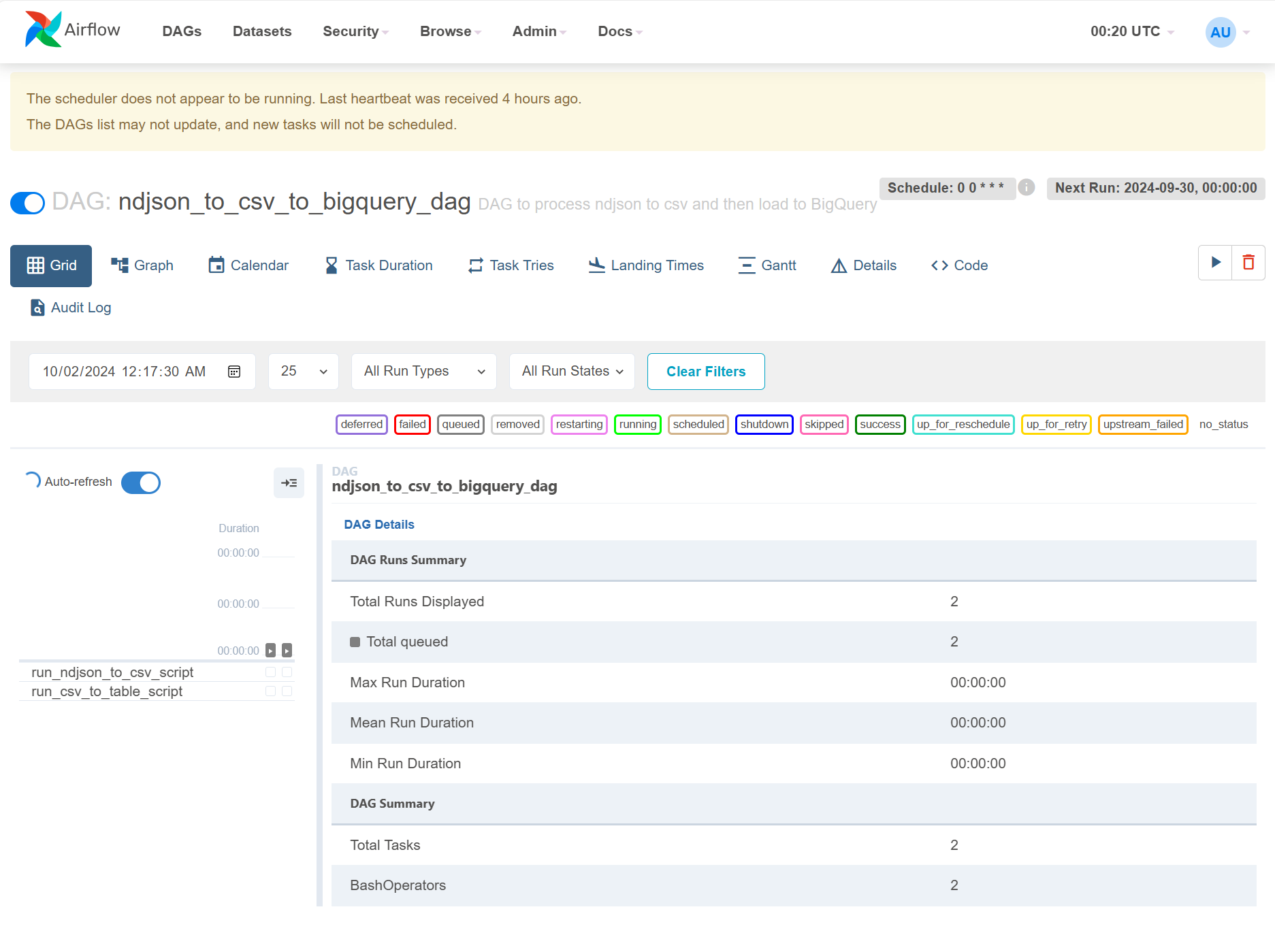
## Automatic run:

**The process runs automatically to serve the creation of jobs on Airflow, or users can use it to run all NDJSON files available in the bucket**

Run by command “python3 ~/nexar\_project/ndjson\_to\_archive\_csv.py auto” => File will be converted automatically.



Run by airflow:



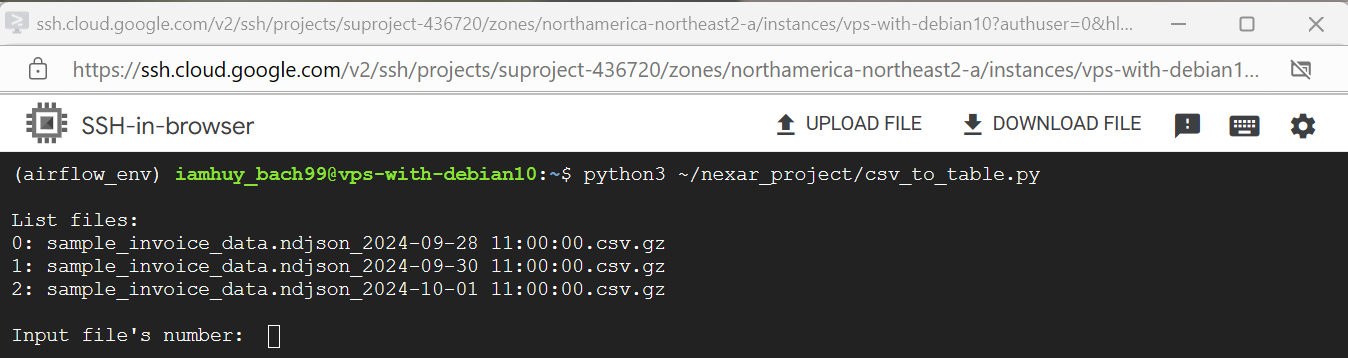
# Program 2: Import data from CSV files into Table in Bigquery

## Manual run:

**Run manually, User can select the file to be converted**

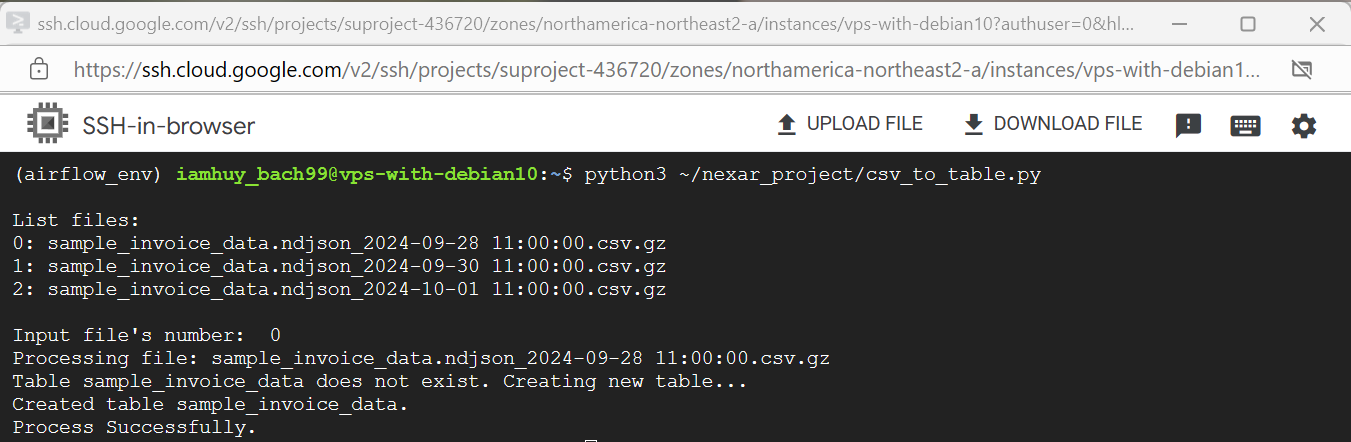
**Step 1:** Run by command “python3 ~/nexar\_project/csv\_to\_table.py” or “python3 ~/nexar\_project/csv\_to\_table.py manual” .

Screen will show list file of CSV was compressed in bucket and user can choose to be imported into table of Bigquery.

****

**Step 2:** Choose file will be imported.

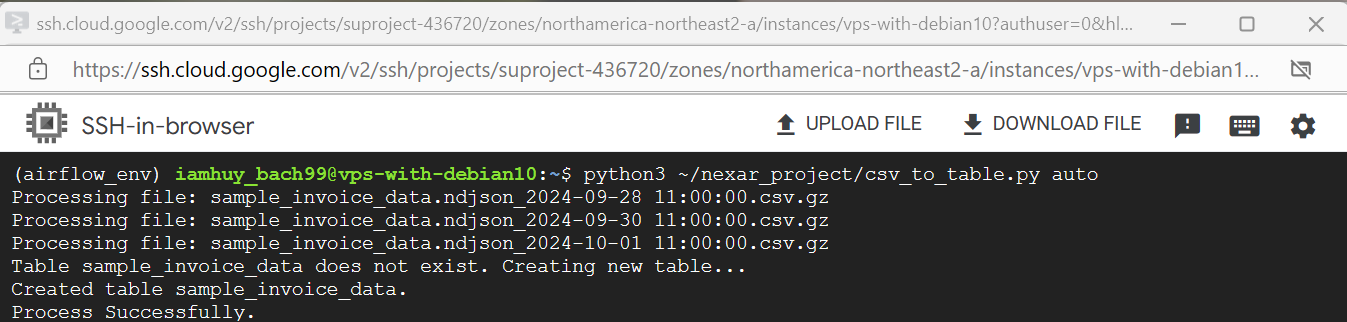
After selecting the file to import, CSV file will be imported to table of Bigquery. If the table does not exist, a new table will be created.



## Automatic run:

**The process runs automatically to serve the creation of jobs on Airflow, or users can use it to run all CSV files was compressed available in the bucket**

Run by command “python3 ~/nexar\_project/csv\_to\_table.py auto” => File will be imported automatically.



Run by airflow:

