

Assignment 7 – Apache Beam

DS 561 – Cloud Computing

Vishwas Bhaktavatsala

U74206902

vishwasb@bu.edu

GitHub Url: <https://github.com/vishwas21/DS561-vishwas-assignments>


Command used to run the Dataflow:

(Can be run by installing all the dependencies in the requirements.txt)

```
○ → DS561-vishwas-assignments git:(master) x python3 -m assignment7
```

Outputs:

Outgoing File:



ffcc410138174527bccdc7330d568568c5b6b3d3dda3990a120475e-apidata.googleusercontent.com/download/storage/v1/b/dataflow-apache-ds56...
Personal Boston University Taxes Games Income Sources
[('4257.html', 249), ('8540.html', 249), ('9800.html', 249), ('5580.html', 249), ('4975.html', 249)]

Incoming File Count:



ff4cd413aaf3c82b3add26ba1eb7c61e3b6174bce28b1760fd82ff4-apidata.googleusercontent.com/download/storage/v1/b/dataflow-apache-ds561-vi...
Personal Boston University Taxes Games Income Sources
[('5984.html', 188), ('1912.html', 166), ('5789.html', 163), ('7885.html', 162), ('7231.html', 162)]