Big Data Integration and Storage– Lab 3

Hardik Rathod

8797380

PROG8450

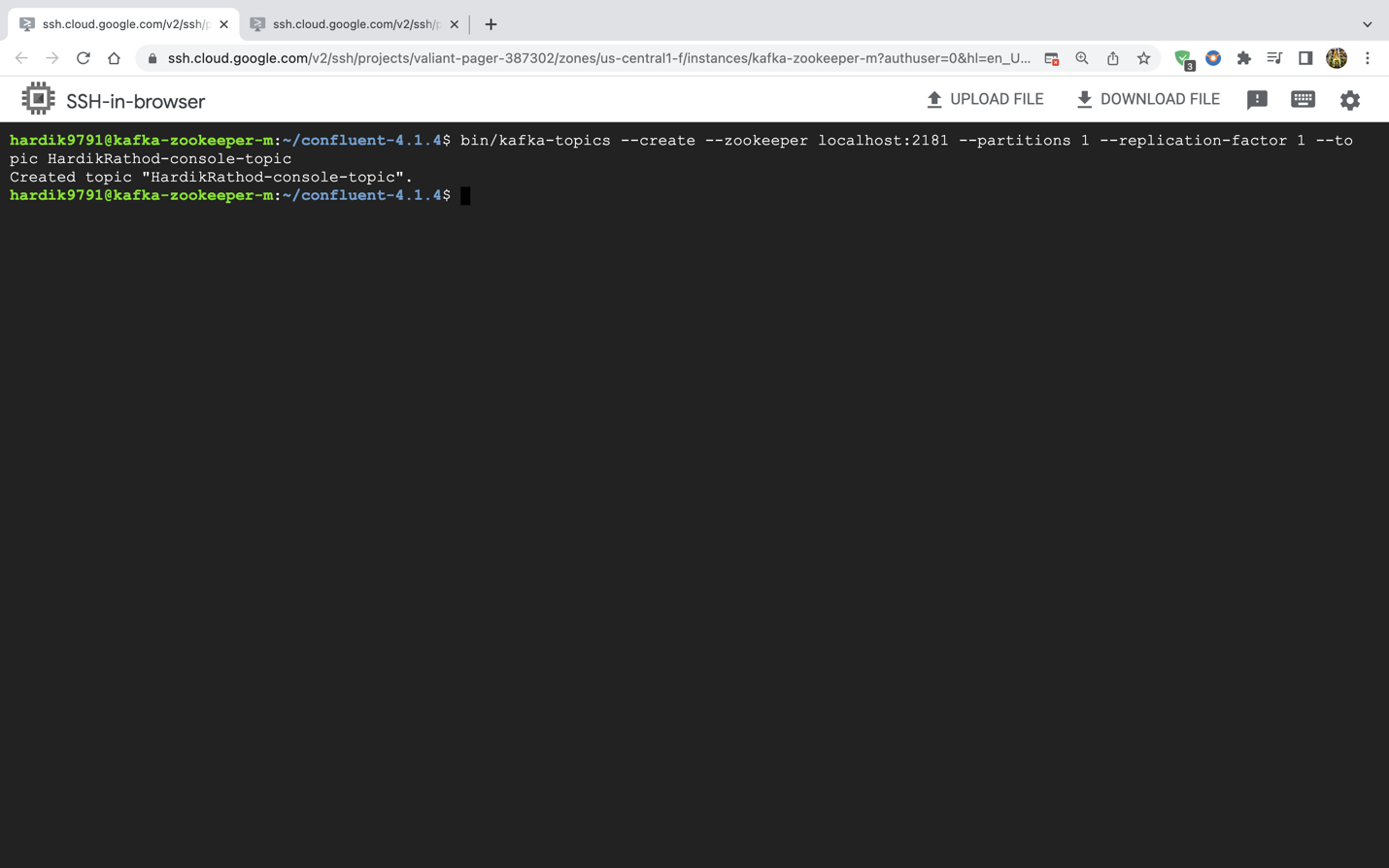
Prof. Saber Amini

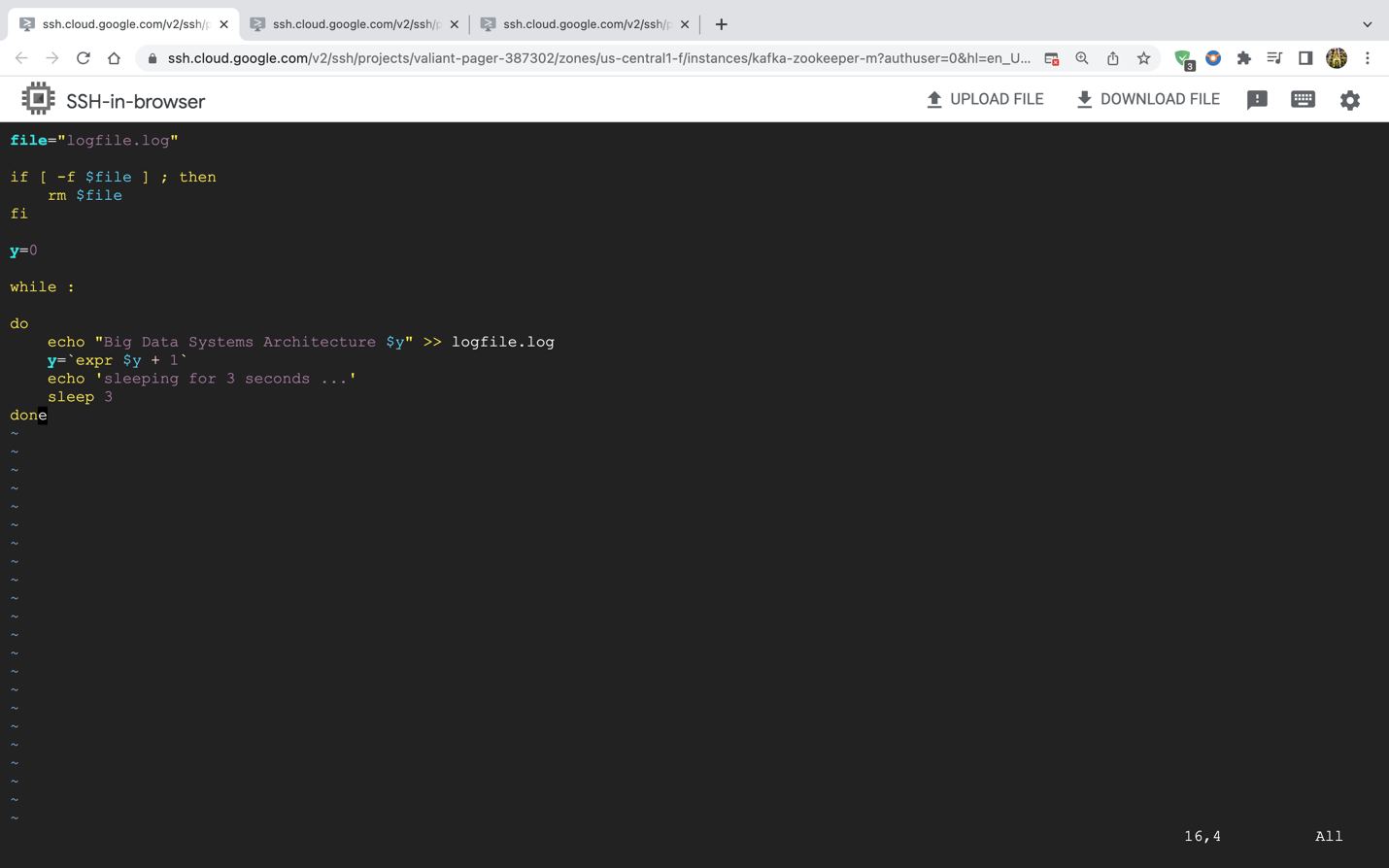
July 4th, 2023

We must first install and launch Zookeeper and Kafka in an SSH terminal.

The second step is to create a topic, and in another SSH terminal, my topic is called “HardikRathod-console-topic”.

Since the partition in this case is set to 1, the same information will be displayed in both places. If we set the partition to 2, some data will be allocated to Consumer 1 and some to Consumer 2.



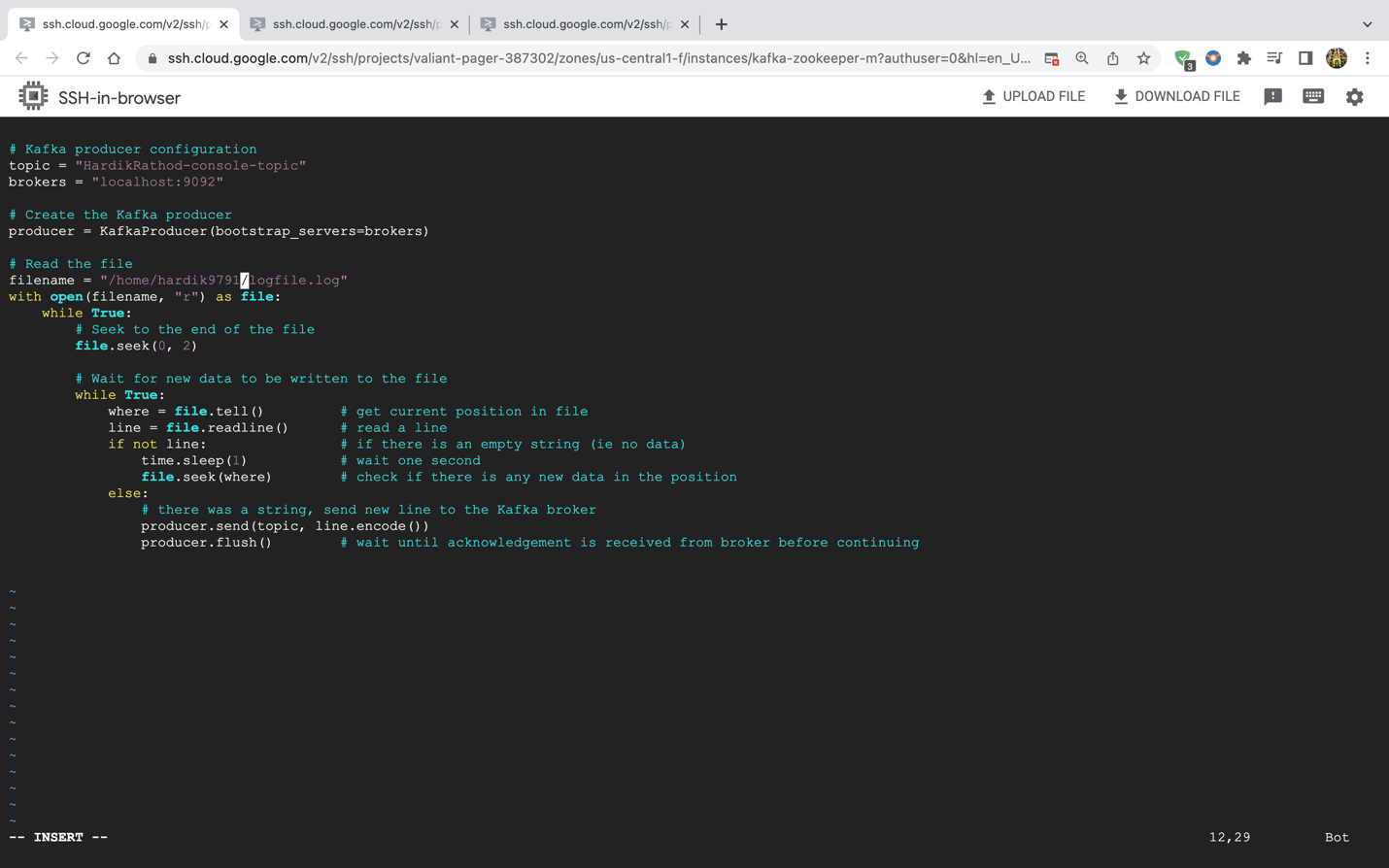
Afterward, I created write-to-file.sh

I then created the “consumer-one.py” file.



Then created the “consumer-two.py” file.

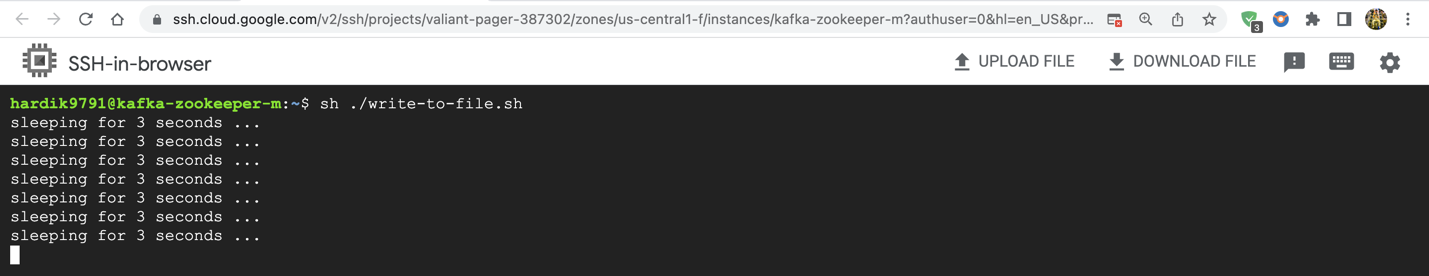
Then, I created “producer-file.py” file.



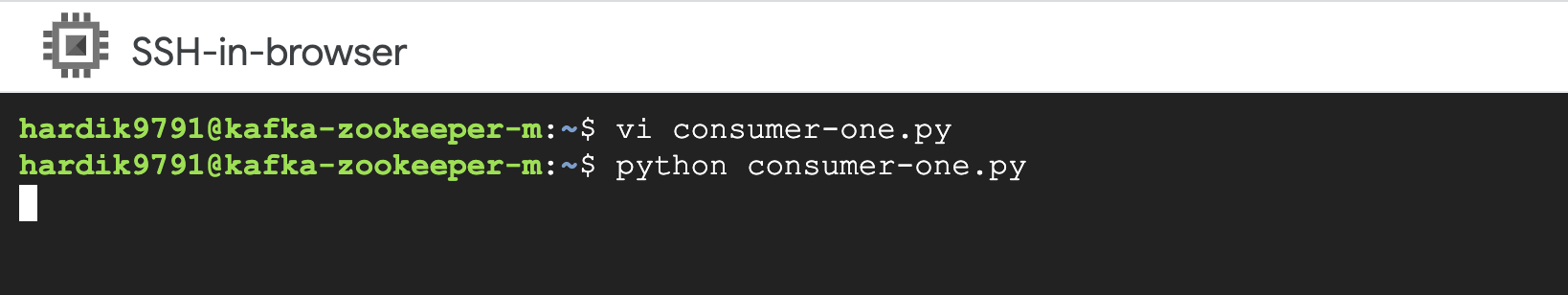
Both consumer files and producer file are shown here.

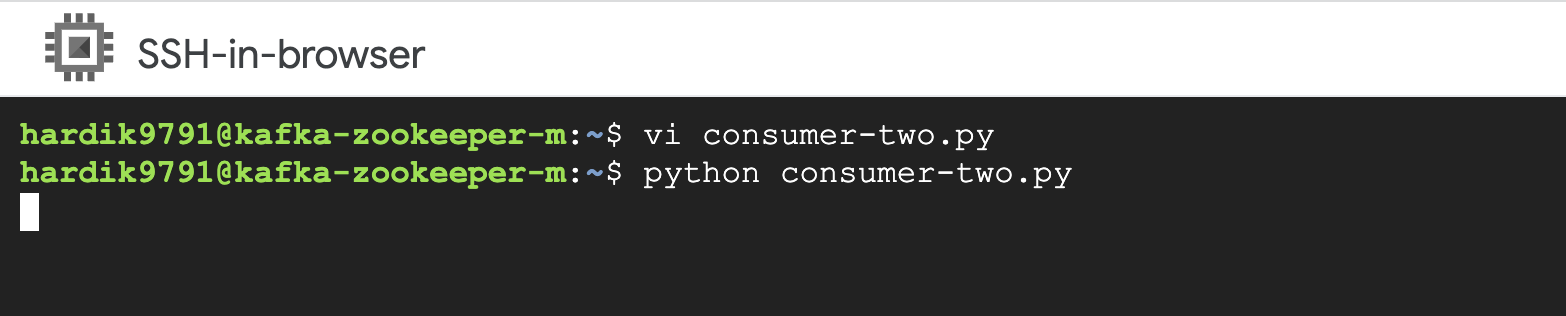
The following steps will demonstrate how I started by first running the write-to-file.sh file, then the consumers and then the producer python files.

write-to-file.sh



First consumer python file



Second consumer python file

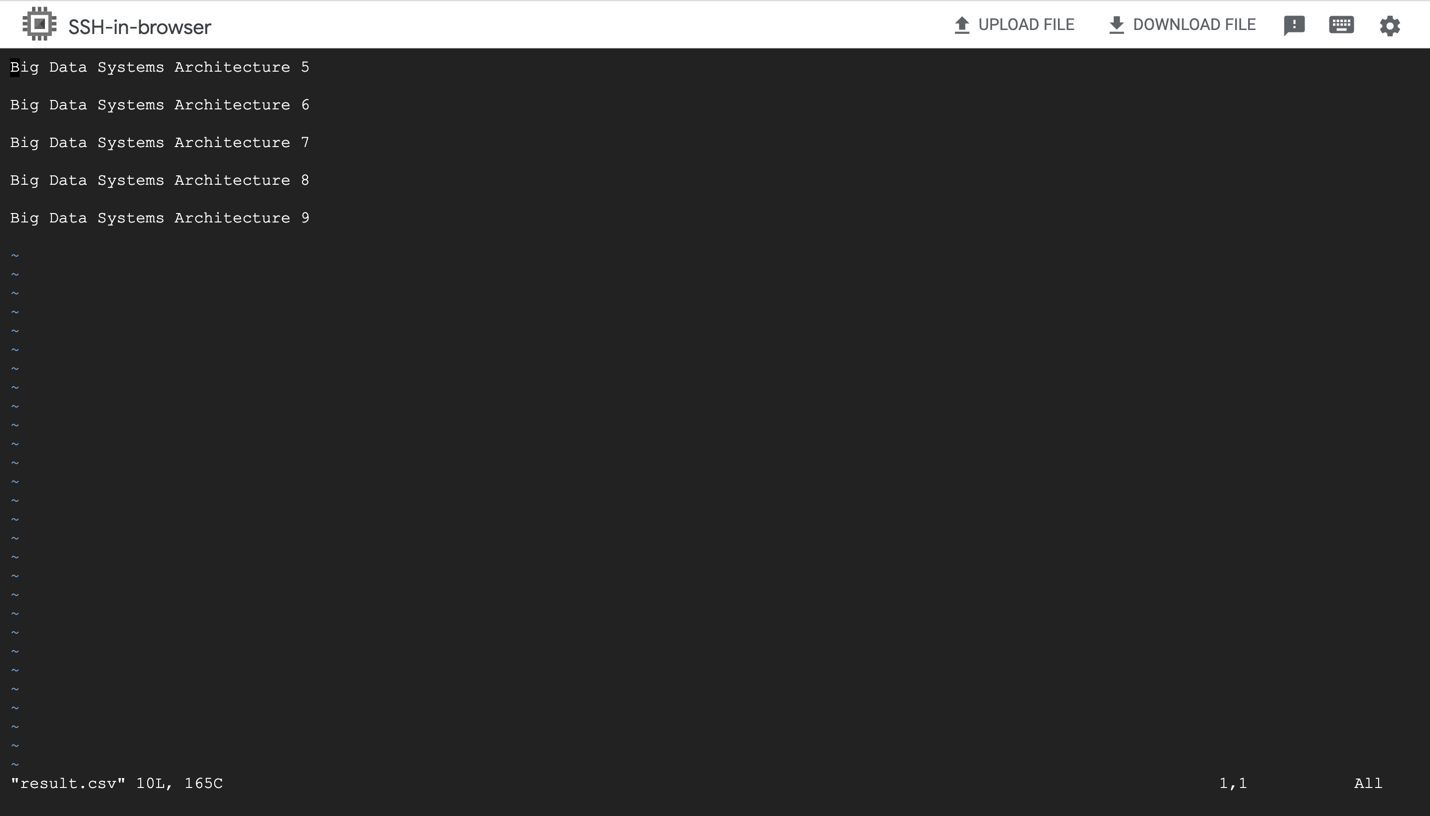
Producer python file

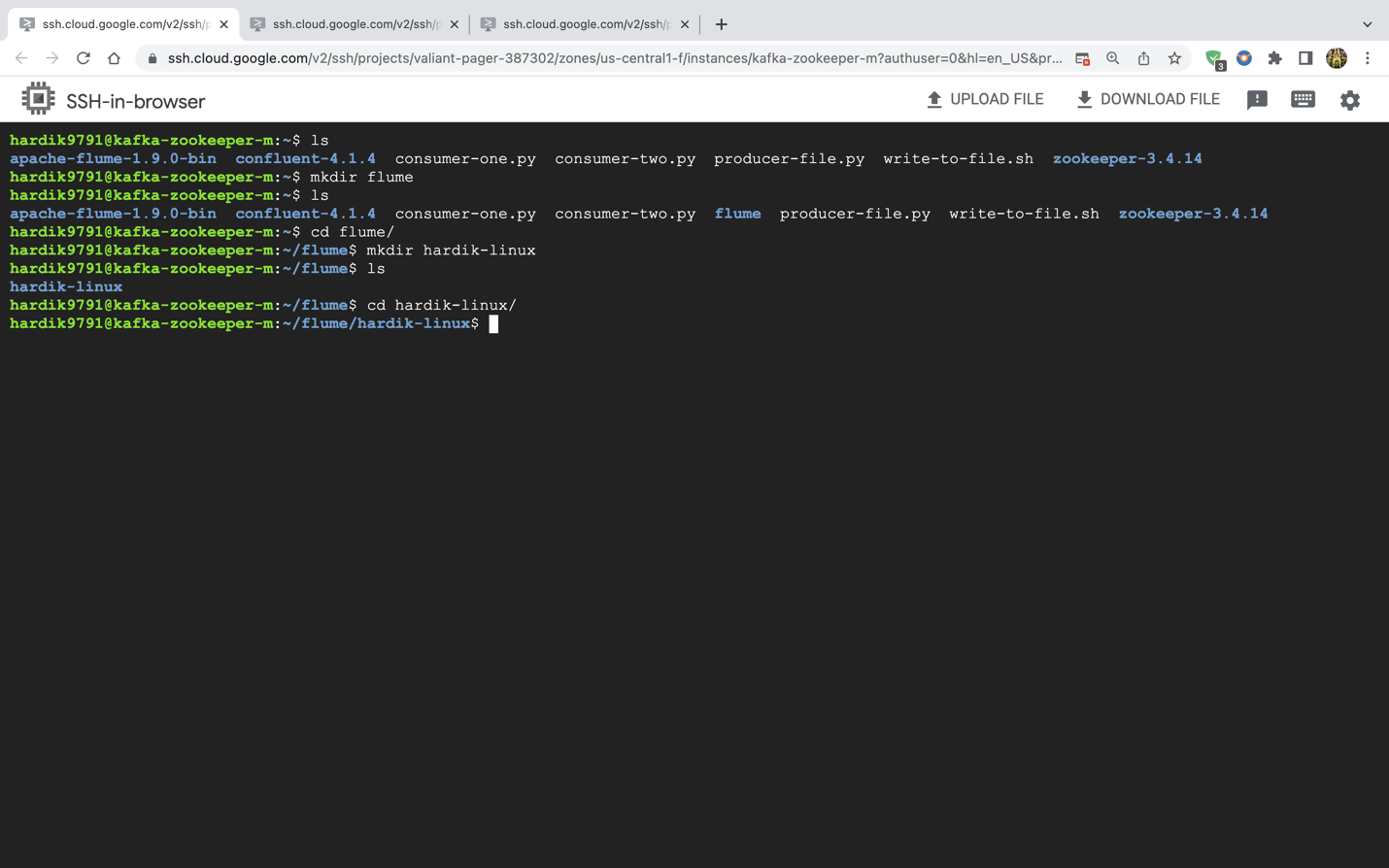


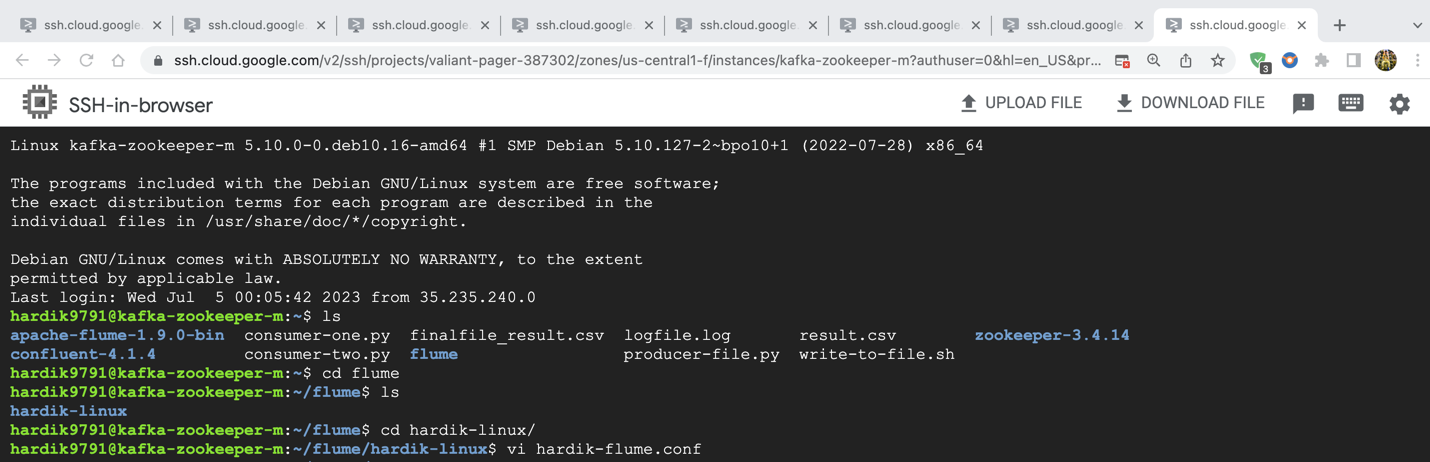
The producer file has been executed, and the results are starting to show in both consumers files.

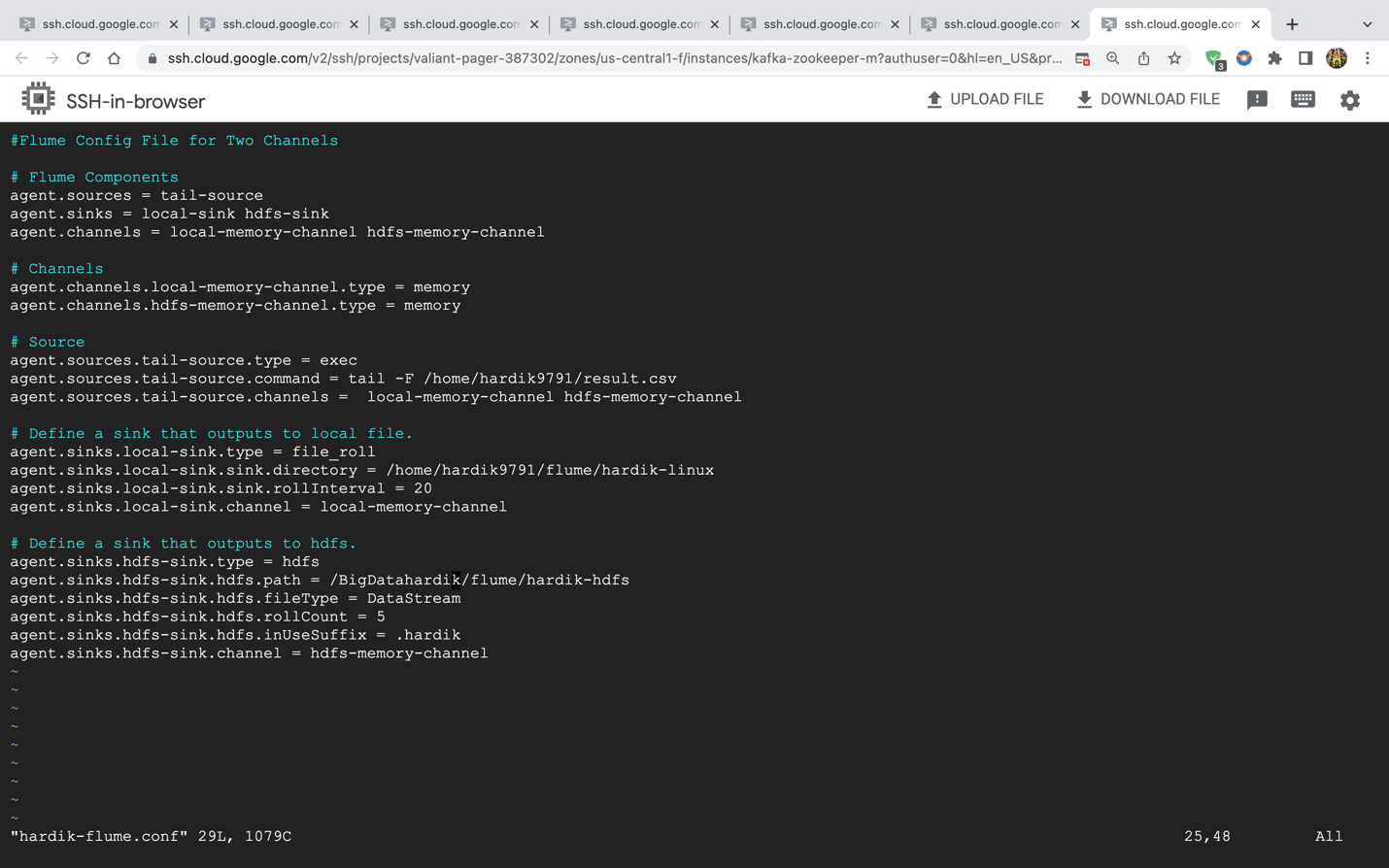
First consumer output: 

There will be output produced every three seconds. The second consumer uses the same process, but the outcome is saved in the “result.csv” file.

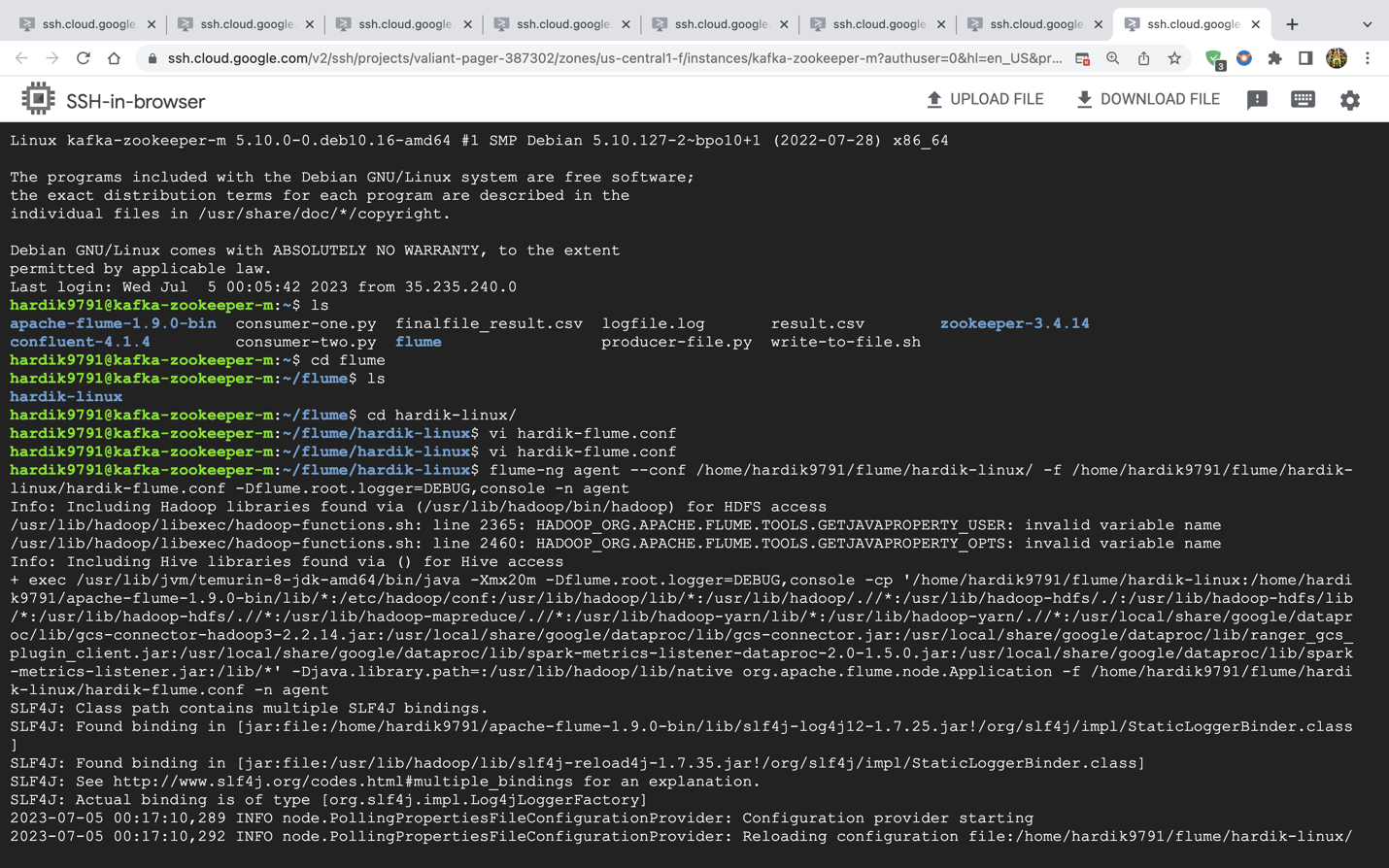
Second consumer output: this output is stored in a csv file

Next, we will install and launch flume then, create a directory called “flume” and in that directory we are going to create another directory called “Hardik-linux” and in that directory we are going to create a config file called “Hardik-flume.conf”.

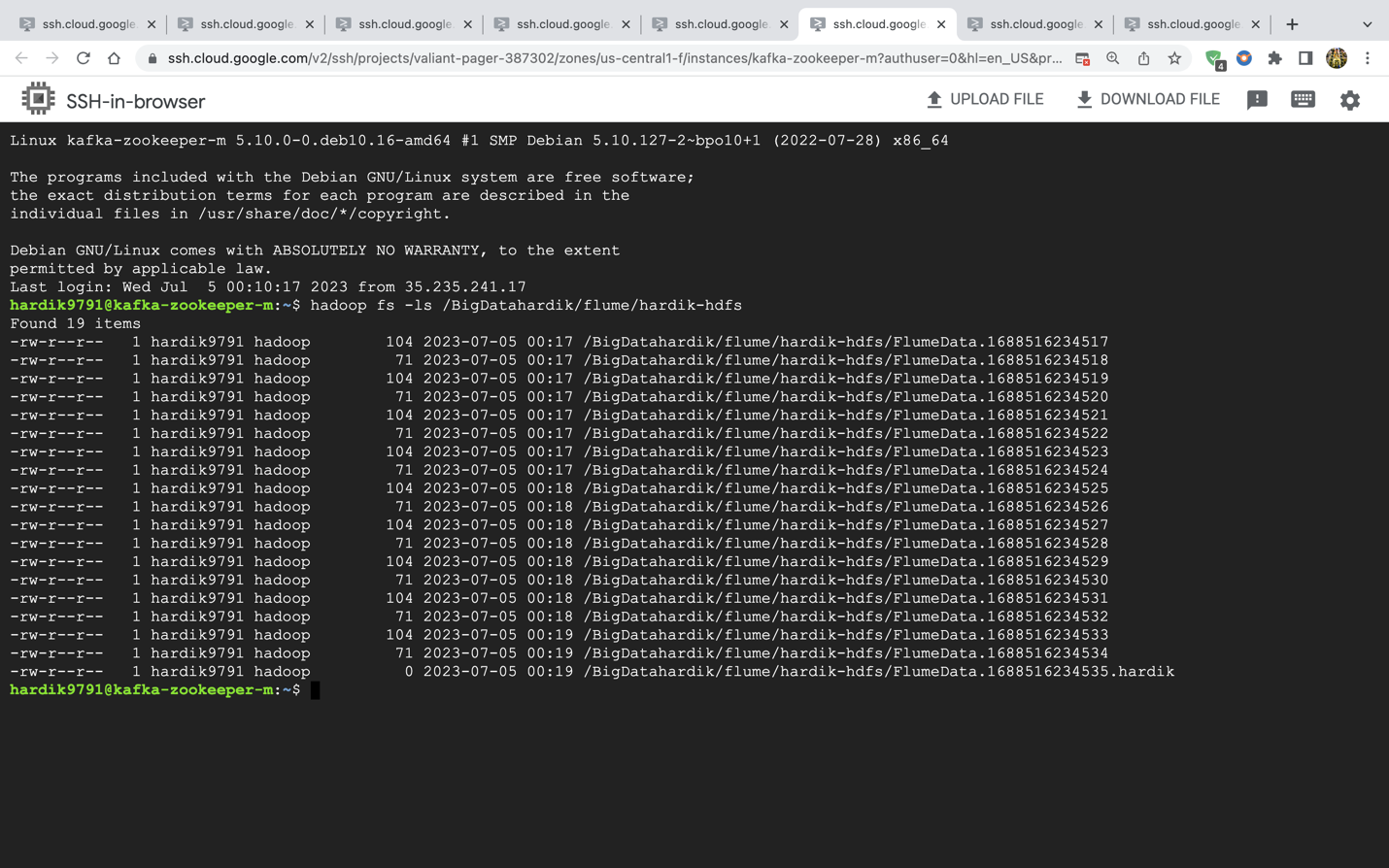


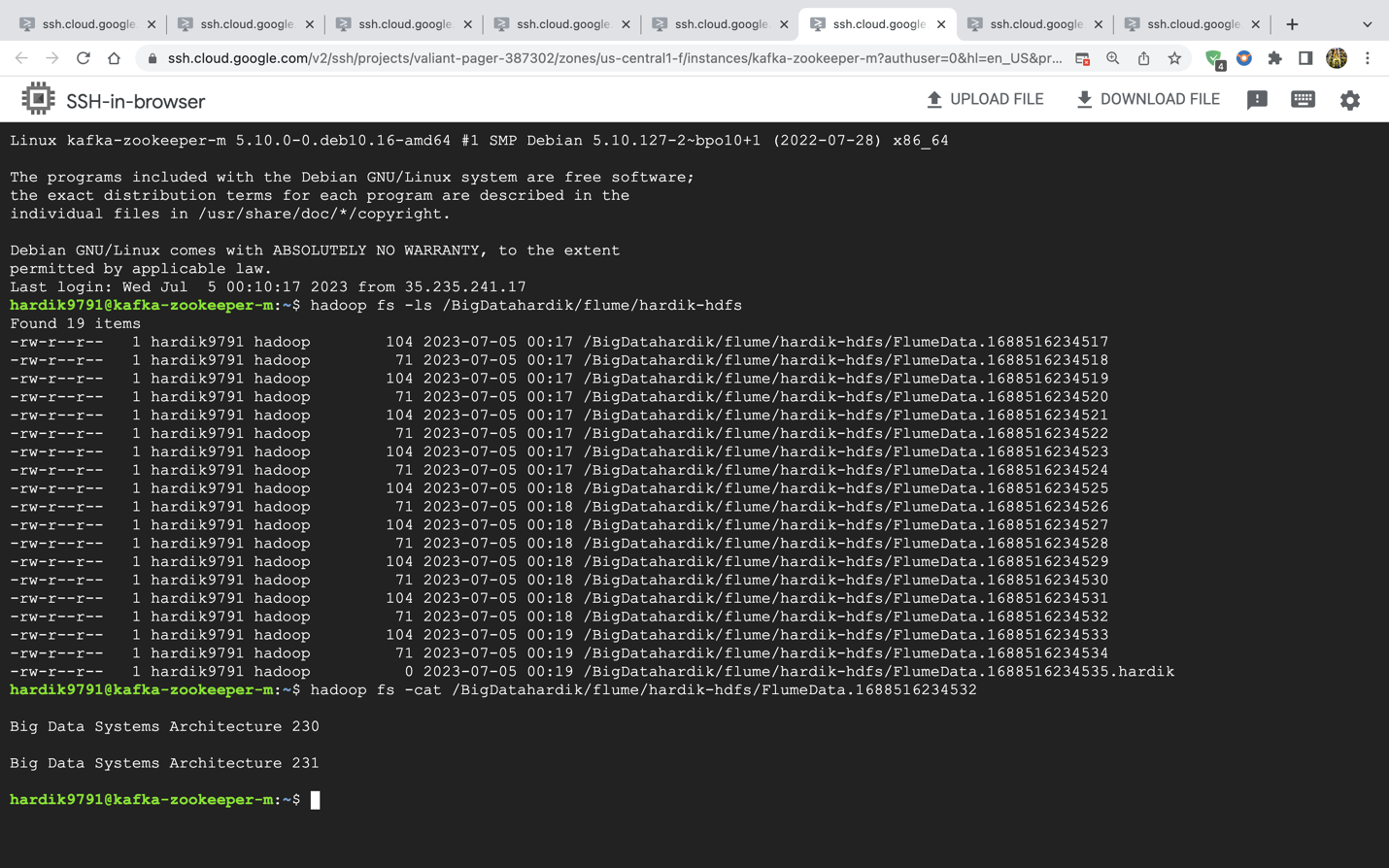
Below is the code for “hardik-flume.conf” file 

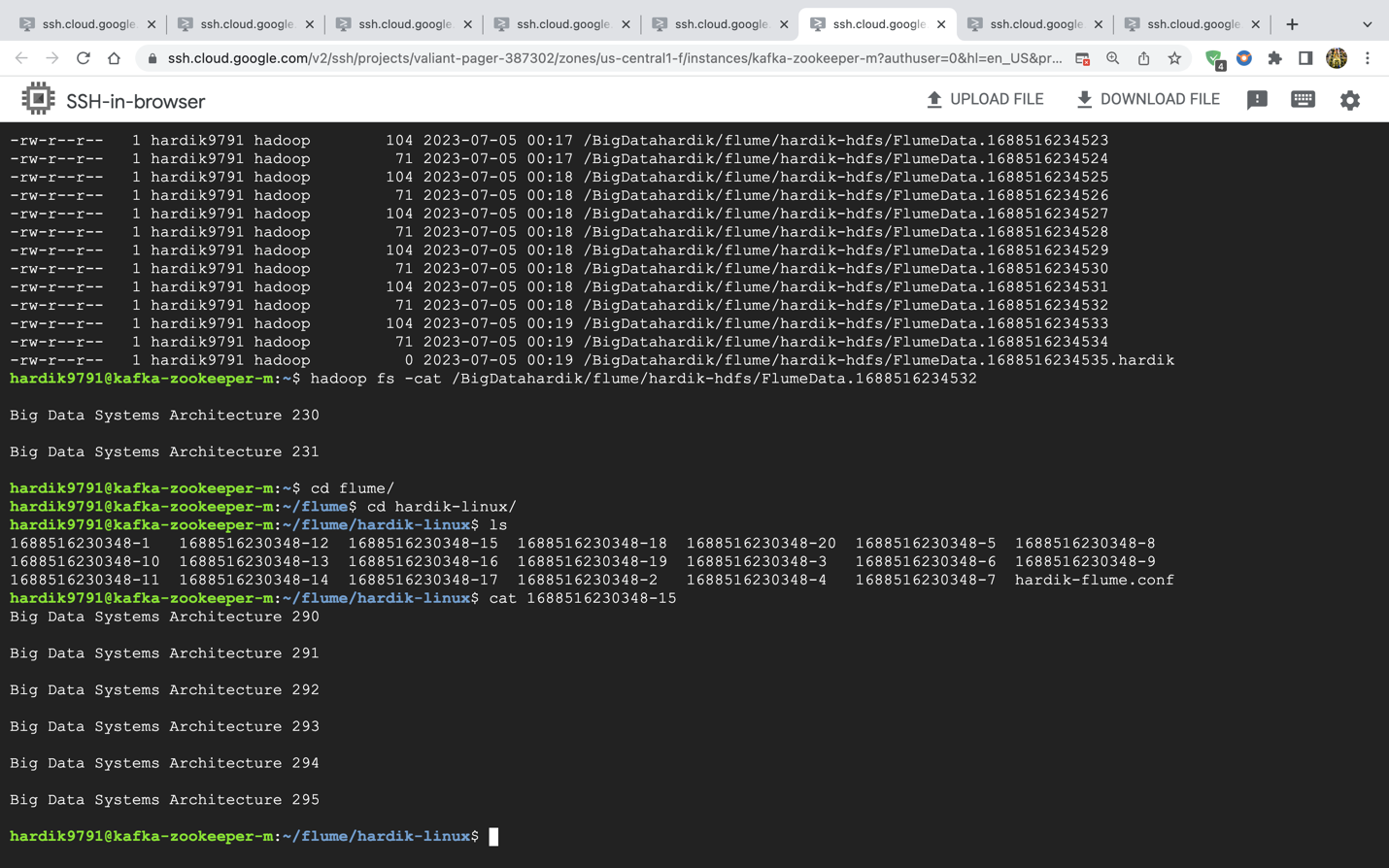
I will now launch the Flume Agent, and the command that was used is shown in the screenshot below:



I will demonstrate several hdfs and linux file outputs in this step:



Screenshot of a few HDFS-stored files 

A couple of the linux files outputs are listed below

Video Recording Link: <https://conestogac.zoom.us/rec/share/7y7_jDhsuJAw6Ej2U4pK6tCq7yj1LQ9hD9OmQojU8vPhiyksGhhbN-1qpOSXaOxi.LUsbRjKRwn1iwJEu?startTime=1688526747000>