

# Chirag Khandhar

A20438926 | CSP 554 – Big Data Technologies | Fall 2020 | Assignment 10

## Uploading **consume.py** and **log4j.properties**

The image shows two terminal windows side-by-side. The left window, labeled 'EC2 -1', shows a terminal session on a hadoop node (ip-172-31-91-84) where the user lists files in the current directory, showing 'consume.py' and 'log4j.properties'. The right window, labeled 'EC2 -2', shows a terminal session on a Windows machine (MINGW64) where the user runs two scp commands to upload files to an EC2 instance. The first command for 'consume.py' fails with an 'invalid format' error. The second command for 'log4j.properties' succeeds, showing a progress bar at 100% with a transfer rate of 72.9KB/s.

```
hadoop@ip-172-31-91-84:~$ ls
consume.py  log4j.properties
hadoop@ip-172-31-91-84:~$

Chirag@Chirag-Khandhar MINGW64 /f/MS/SEM3/BDT/Git/Keys (master)
$ scp -i F:/MS/SEM3/BDT/Git/Keys/emr-key-pair.pem F:/MS/SEM3/BDT/Git/Assignments/Questions/10/consume.py hadoop@ec2-184-72-200-92.compute-1.amazonaws.com:/home/hadoop
load pubkey "F:/MS/SEM3/BDT/Git/Keys/emr-key-pair.pem": invalid format
consume.py                                100% 690    17.9KB/s   00:00

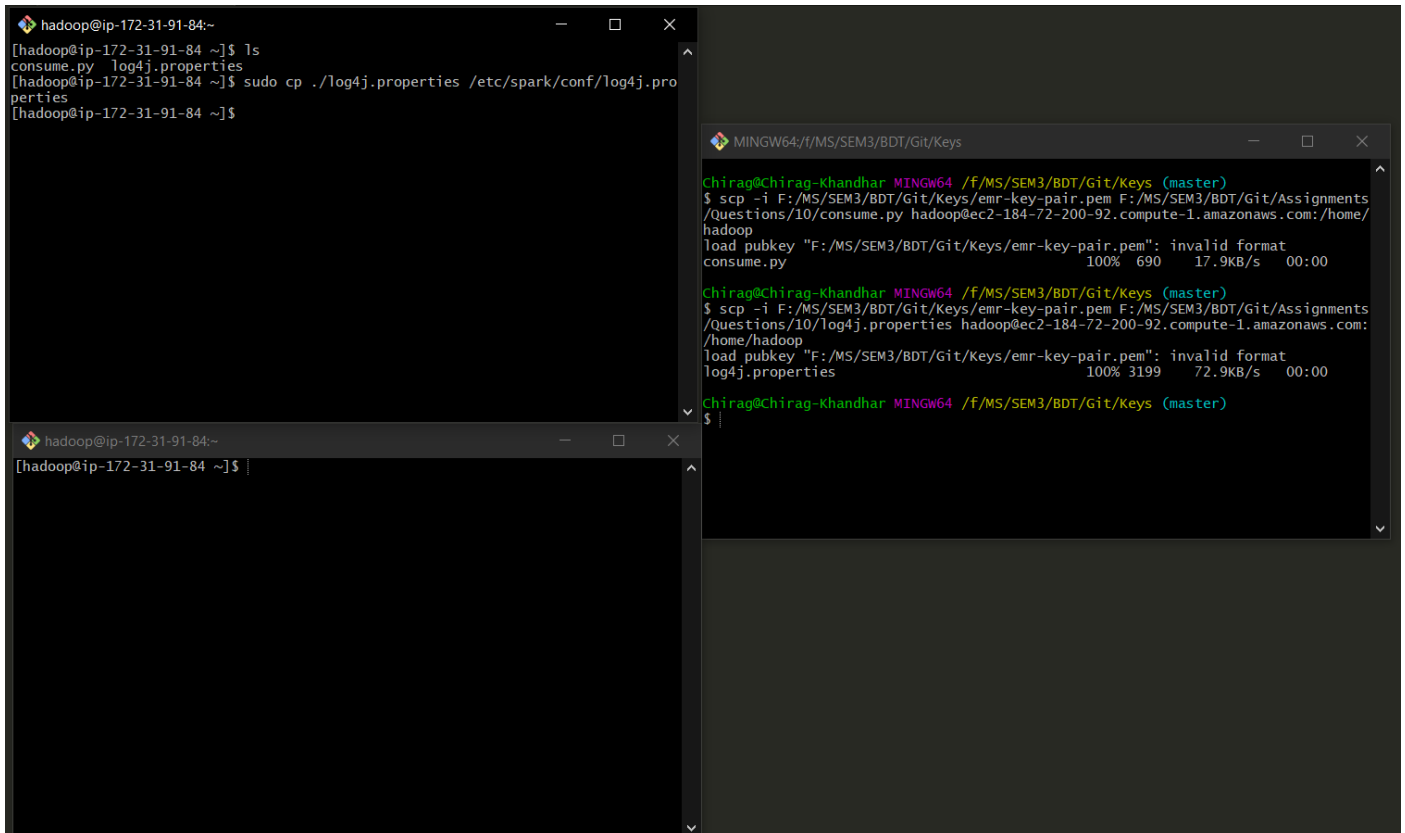
Chirag@Chirag-Khandhar MINGW64 /f/MS/SEM3/BDT/Git/Keys (master)
$ scp -i F:/MS/SEM3/BDT/Git/Keys/emr-key-pair.pem F:/MS/SEM3/BDT/Git/Assignments/Questions/10/log4j.properties hadoop@ec2-184-72-200-92.compute-1.amazonaws.com:/home/hadoop
load pubkey "F:/MS/SEM3/BDT/Git/Keys/emr-key-pair.pem": invalid format
log4j.properties                         100% 3199   72.9KB/s   00:00

Chirag@Chirag-Khandhar MINGW64 /f/MS/SEM3/BDT/Git/Keys (master)
$
```

In the EC2-1 window enter the following:

```
sudo cp ./log4j.properties /etc/spark/conf/log4j.properties
```

This changes the logging properties to turn off “INFO” messages to allow easier viewing of the results of the stream processing job. But it is not something you always want to disable.



```
hadoop@ip-172-31-91-84:~$ ls
consume.py  log4j.properties
hadoop@ip-172-31-91-84:~$ sudo cp ./log4j.properties /etc/spark/conf/log4j.pro
erties
hadoop@ip-172-31-91-84:~$

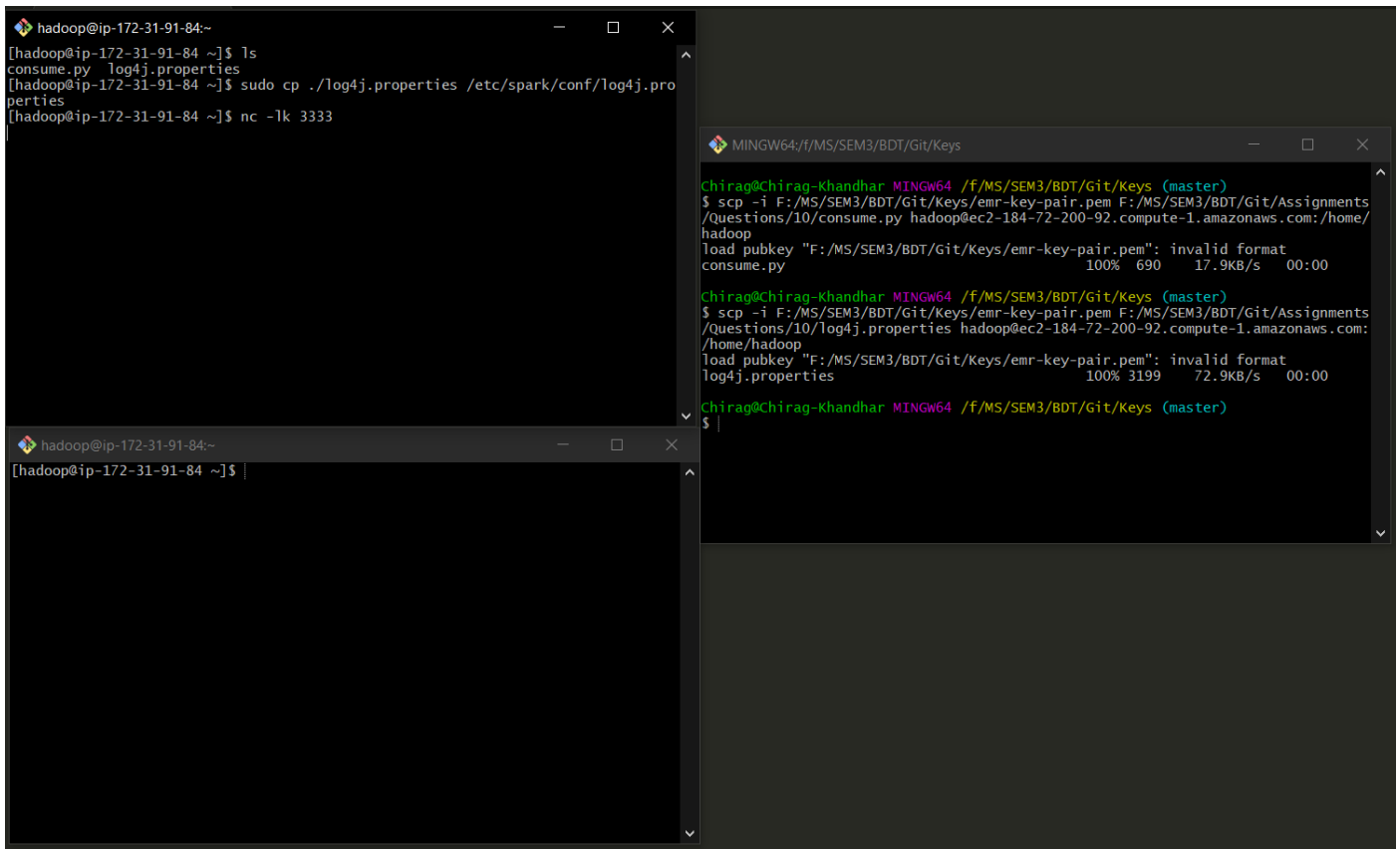
Chirag@Chirag-Khandhar MINGW64 /f/MS/SEM3/BDT/Git/Keys (master)
$ scp -i F:/MS/SEM3/BDT/Git/Keys/emr-key-pair.pem F:/MS/SEM3/BDT/Git/Assignments
/Questions/10/consume.py hadoop@ec2-184-72-200-92.compute-1.amazonaws.com:/home/
hadoop
load pubkey "F:/MS/SEM3/BDT/Git/Keys/emr-key-pair.pem": invalid format
consume.py                                100% 690   17.9KB/s   00:00

Chirag@Chirag-Khandhar MINGW64 /f/MS/SEM3/BDT/Git/Keys (master)
$ scp -i F:/MS/SEM3/BDT/Git/Keys/emr-key-pair.pem F:/MS/SEM3/BDT/Git/Assignments
/Questions/10/log4j.properties hadoop@ec2-184-72-200-92.compute-1.amazonaws.com:
/home/hadoop
load pubkey "F:/MS/SEM3/BDT/Git/Keys/emr-key-pair.pem": invalid format
log4j.properties                         100% 3199   72.9KB/s   00:00

Chirag@Chirag-Khandhar MINGW64 /f/MS/SEM3/BDT/Git/Keys (master)
$
```

In the EC2-1 window enter the following command to open a TCP (socket) connection on port 3333

**nc -lk 3333**



```
hadoop@ip-172-31-91-84:~$ ls
consume.py  log4j.properties
hadoop@ip-172-31-91-84:~$ sudo cp ./log4j.properties /etc/spark/conf/log4j.properties
hadoop@ip-172-31-91-84:~$ nc -lk 3333

MINGW64:~/MS/SEM3/BDT/Git/Keys
Chirag@Chirag-Khandhar MINGW64 /f/MS/SEM3/BDT/Git/Keys (master)
$ scp -i F:/MS/SEM3/BDT/Git/Keys/emr-key-pair.pem F:/MS/SEM3/BDT/Git/Assignments/Questions/10/consume.py hadoop@ec2-184-72-200-92.compute-1.amazonaws.com:/home/hadoop
load pubkey "F:/MS/SEM3/BDT/Git/Keys/emr-key-pair.pem": invalid format
consume.py                                100% 690    17.9KB/s   00:00

Chirag@Chirag-Khandhar MINGW64 /f/MS/SEM3/BDT/Git/Keys (master)
$ scp -i F:/MS/SEM3/BDT/Git/Keys/emr-key-pair.pem F:/MS/SEM3/BDT/Git/Assignments/Questions/10/log4j.properties hadoop@ec2-184-72-200-92.compute-1.amazonaws.com:/home/hadoop
load pubkey "F:/MS/SEM3/BDT/Git/Keys/emr-key-pair.pem": invalid format
log4j.properties                          100% 3199   72.9KB/s   00:00

Chirag@Chirag-Khandhar MINGW64 /f/MS/SEM3/BDT/Git/Keys (master)
$
```

In the EC2-2 window enter the following command:

**spark-submit consume.py**

```
hadoop@ip-172-31-91-84:~$ ls
consume.py  log4j.properties
hadoop@ip-172-31-91-84:~$ sudo cp ./log4j.properties /etc/spark/conf/log4j.pro
erties
hadoop@ip-172-31-91-84:~$ nc -lk 3333

hadoop@ip-172-31-91-84:~$ spark-submit consume.py
20/11/03 22:39:39 WARN StreamingContext: Dynamic Allocation is enabled for this
application. Enabling Dynamic allocation for Spark Streaming applications can ca
use data loss if Write Ahead Log is not enabled for non-replayable sources like
Flume. See the programming guide for details on how to enable the Write Ahead Lo
g.
-----
Time: 2020-11-03 22:39:50
-----
```

```
hadoop@ip-172-31-91-84:~$ ls
consume.py  log4j.properties
hadoop@ip-172-31-91-84:~$ sudo cp ./log4j.properties /etc/spark/conf/log4j.pro
erties
hadoop@ip-172-31-91-84:~$ nc -lk 3333
hi this is chirag and he is good man

hadoop@ip-172-31-91-84:~$
-----
Time: 2020-11-03 22:40:40
-----
Time: 2020-11-03 22:40:50
-----
Time: 2020-11-03 22:41:00
-----
('this', 1)
('is', 2)
('he', 1)
('', 1)
('good', 1)
('hi', 1)
('chirag', 1)
('and', 1)
('man', 1)
```