

Họ và tên: Bùi Thiện Nhân

MSSV: 2274802010592

LAB 08 NHẬP MÔN PHÂN TÍCH DỮ LIỆU LỚN

Kiểm tra file input.txt và đẩy lên HDFS

```
hadoop@hadoop-2274802010592:~/iDragonCloud/71ITDS40303/Thực hành/HadoopWithPython/python/HDFS$ ls
Python-2.7.18.tgz copy_to_local.py delete.py get-pip.py list_directory.py list_file.py mkdir.py text.py
t/doop@hadoop-2274802010592:~/iDragonCloud/71ITDS40303/Thực hành/HadoopWithPython/python/HDFS$ hdfs dfs -put input.txt /user/hadoop/input

hadoop@hadoop-2274802010592:~/iDragonCloud/71ITDS40303/Thực hành/HadoopWithPython/python/HDFS$ ls
Python-2.7.18.tgz copy_to_local.py delete.py get-pip.py input.txt list_directory.py list_file.py mkdir.py text.py
hadoop@hadoop-2274802010592:~/iDragonCloud/71ITDS40303/Thực hành/HadoopWithPython/python/HDFS$
```

Cài snakebite cho python 3

```
hadoop@hadoop-2274802010592:~/iDragonCloud/71ITDS40303/Thực hành/HadoopWithPython/python/HDFS$ pip install snakebite-py3
Defaulting to user installation because normal site-packages is not writeable
Collecting snakebite-py3
  Downloading snakebite_py3-3.0.6-py3-none-any.whl (101 kB)
    |-----| 101.5/101.5 KB 1.3 MB/s eta 0:00:00
Collecting protobuf<4,>2.4.1
  Downloading protobuf-3.20.3-cp310-cp310-manylinux_2_12_x86_64.manylinux2010_x86_64.whl (1.1 MB)
    |-----| 1.1/1.1 MB 6.8 MB/s eta 0:00:00
Collecting argparse
  Downloading argparse-1.4.0-py2.py3-none-any.whl (23 kB)
Installing collected packages: argparse, protobuf, snakebite-py3
Successfully installed argparse-1.4.0 protobuf-3.20.3 snakebite-py3-3.0.6
hadoop@hadoop-2274802010592:~/iDragonCloud/71ITDS40303/Thực hành/HadoopWithPython/python/HDFS$
```

Chuẩn bị môi trường cài đặt gói ubuntu update cho server devOps và tải python3

sudo apt update

sudo apt install python3-pip

Chạy các services để start cổng Hadoop

sudo service ssh start

hdfs namenode -format

start-dfs.sh

start-yarn.sh

start-all.sh

nmap -p 9000 localhost

nmap -p 8080 localhost

Forward port 'ifconfig':

```
hadoop@hadoop-2274802010592:~/iDragonCloud/71ITDS40303/Thực hành/HadoopWithPython/python/HDFS$ ifconfig
eth0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST> mtu 1500
    inet 172.23.0.104 netmask 255.255.0.0 broadcast 172.23.255.255
    ether 02:42:ac:17:00:68 txqueuelen 0 (Ethernet)
    RX packets 54064 bytes 88609118 (88.6 MB)
    RX errors 0 dropped 0 overruns 0 frame 0
    TX packets 35968 bytes 2763748 (2.7 MB)
    TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS 2

^

×

Port	Forwarded Address	Running Process	Origin
○ 8000	localhost:8000		Auto Forwarded
○ 9000	localhost:9000		Auto Forwarded
<div>Add Port</div>			