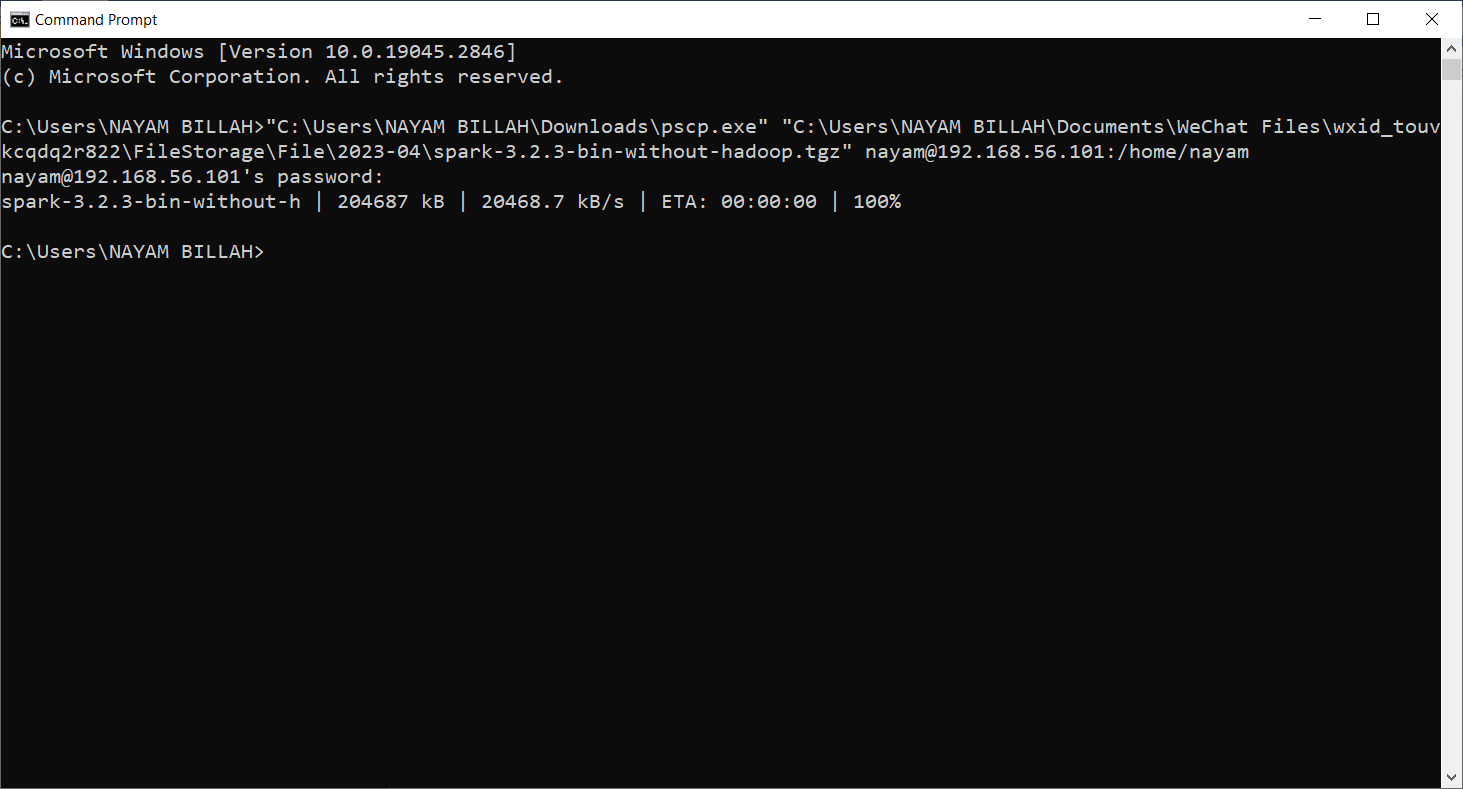
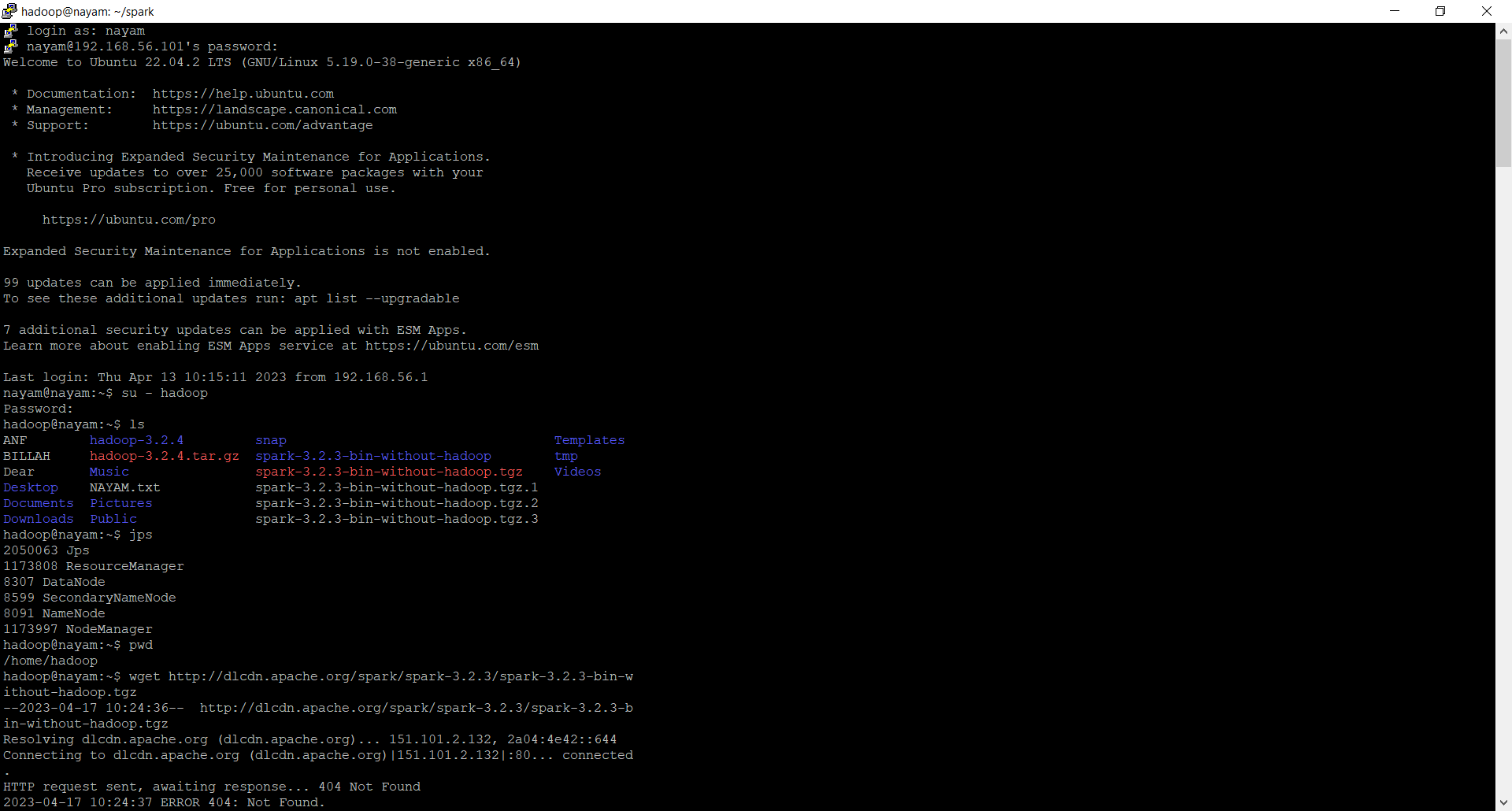
LAB-5

In this session, we will work Apache Spark which is an open-source unified analytics engine for large-scale data processing. Spark provides an interface for programming clusters with implicit parallelism and fault tolerance.

First of all, we have to download binary package. We use Spark 3.2.3 without Hadoop version.

download pscp.exe 64-bit x86 (an SFTP client, i.e., general file transfer sessions much like FTP) and binary package by using the command then, Open the command prompt of our windows operating system. then, drag and drop from your local directory the pscp.exe as well as binary package to the command prompt. Finally, put your user name, the IP address of your Hadoop, home directory and path name and press enter and give the password then we will able to open the Spark on the Hadoop system. 

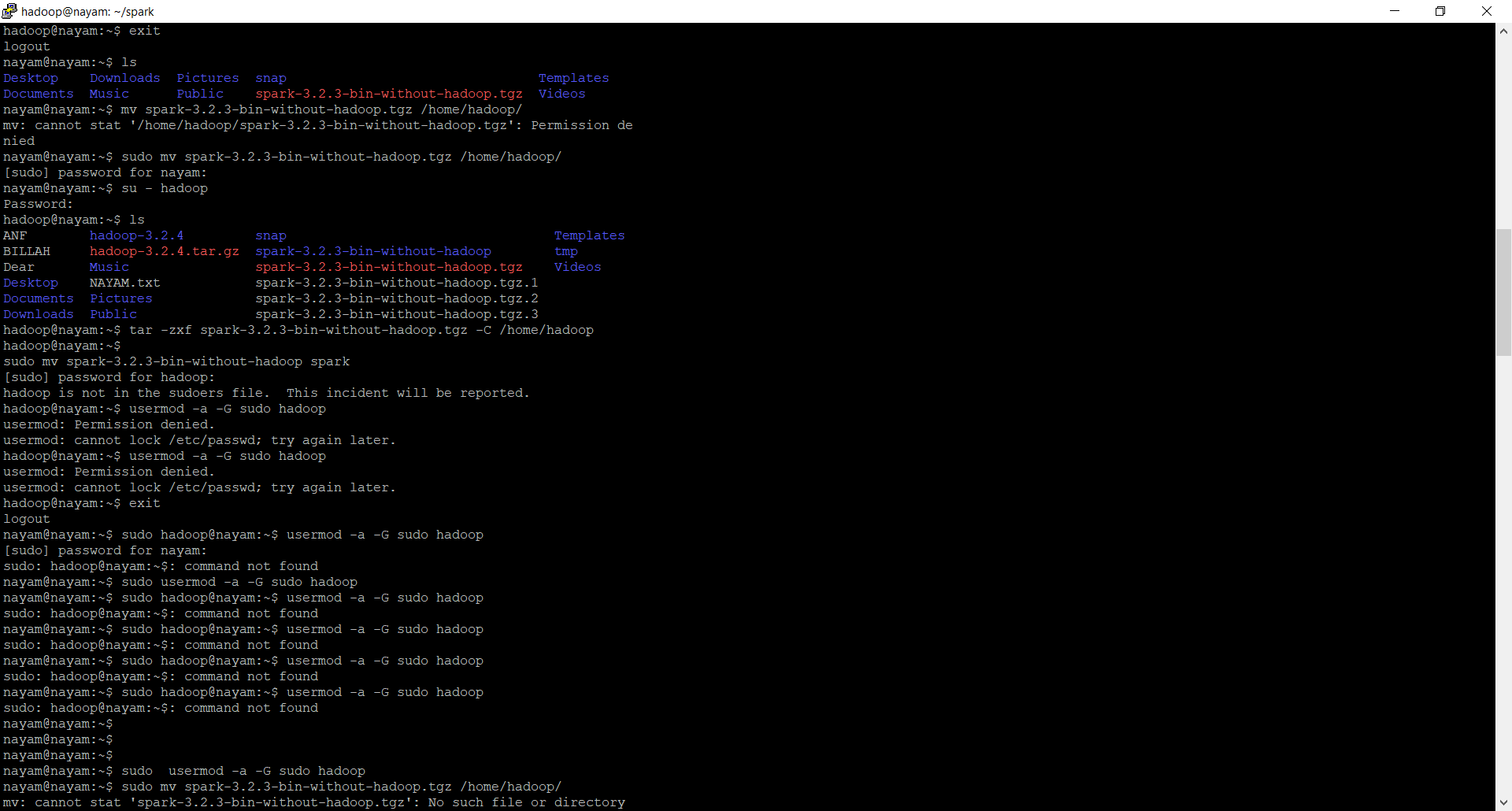
open Putty terminal to implement our caommands to execute

we have to be sure spark-3.2.3-bin-without-hadoop.tgz is in /home/hadoop. 

by using the user name and the password of 192.168.56.101’s open the system to run and execute the commands and enter the Ubuntu operating system. Then, type some commands to check your current status such as write su – hadoop and give the passowrd to enter your hadoop distributed file system. There are some basic commands like ls, jps, pwd and so on.

in addition, use the command to move the file to hadoop

sudo mv spark-3.2.3-bin-without-hadoop.tgz /home/hadoop/



Now, Unzip the Spark package suing the following command:

tar -zxf spark-3.2.3-bin-without-hadoop.tgz -C /home/hadoop

sudo mv spark-3.2.3-bin-without-hadoop spark

. There are also some commands to remove the file from Hadoop and run and execute the file again.

rm -r spark-3.2.3-bin-without-hadoop

tar -zxf spark-3.2.3-bin-without-hadoop.tgz -C /home/hadoop

sudo mv spark-3.2.3-bin-without-hadoop spark

we will move to next step- setup environment variables. for setting the environment look at the commands,

hadoop@nayam:~$ cd /home/hadoop/spark

hadoop@nayam:~/spark$ cp ./conf/spark-env.sh.template ./conf/spark-env.sh

hadoop@nayam:~/spark$ ./conf/spark-env.sh

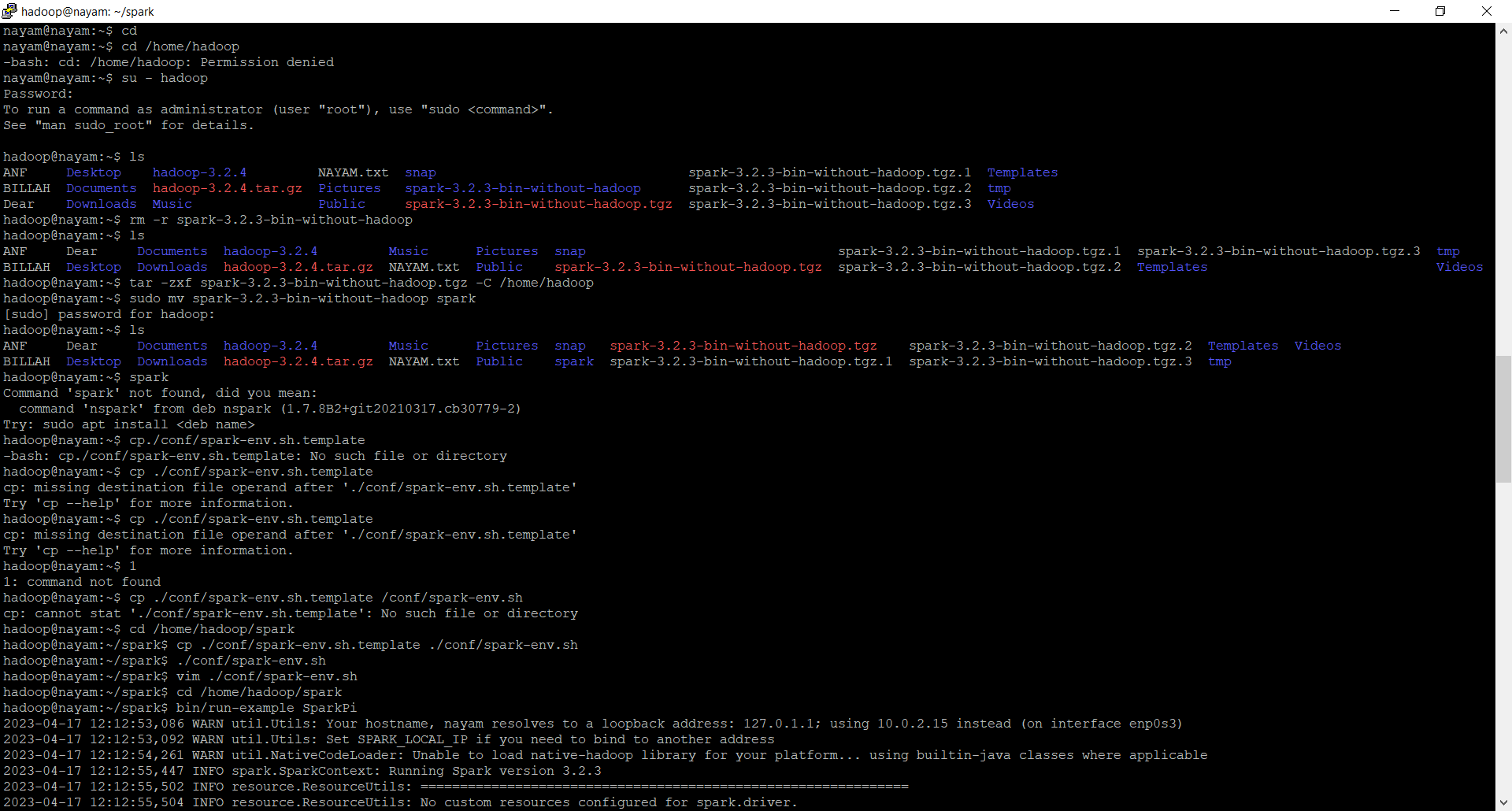
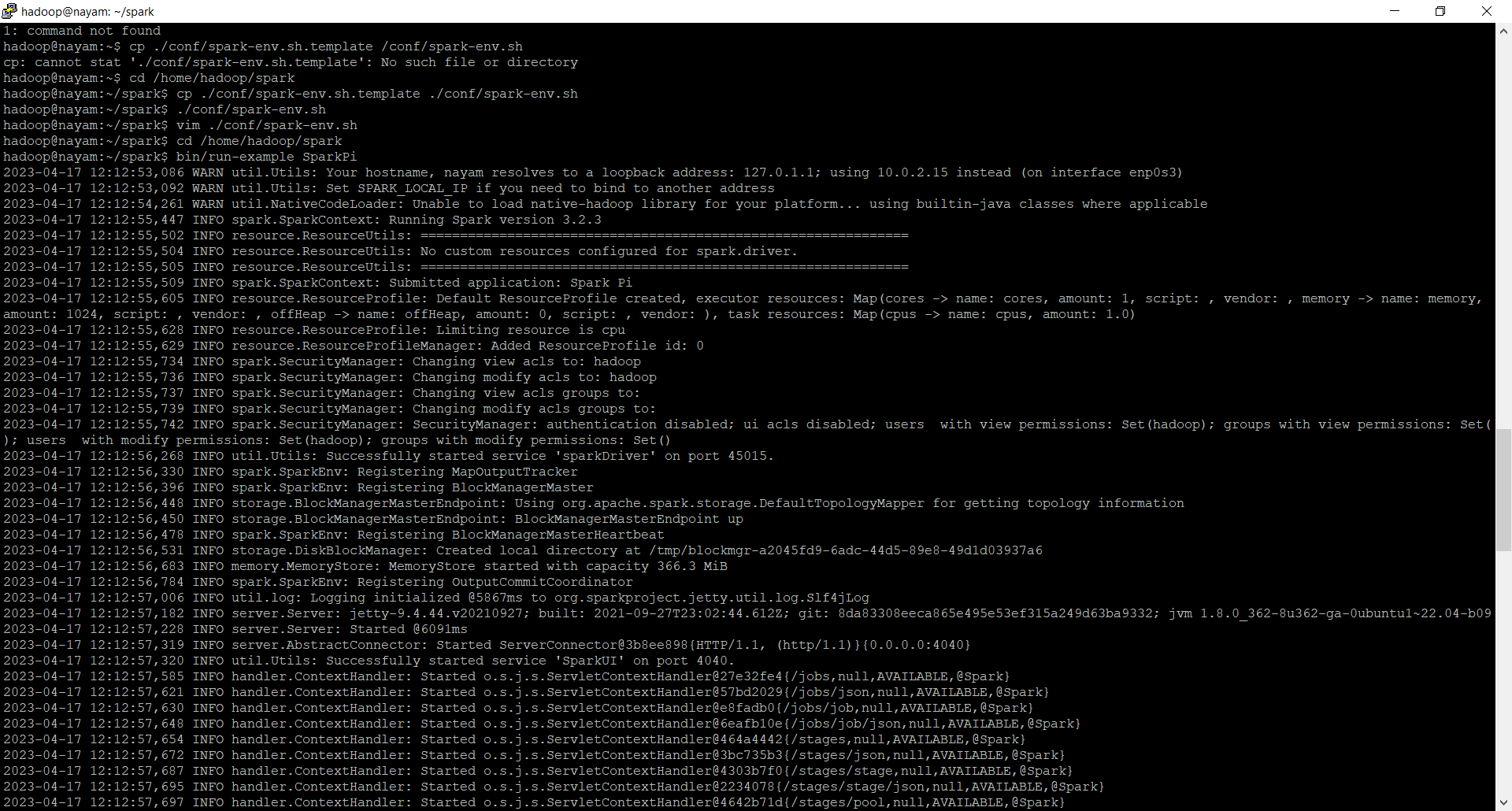
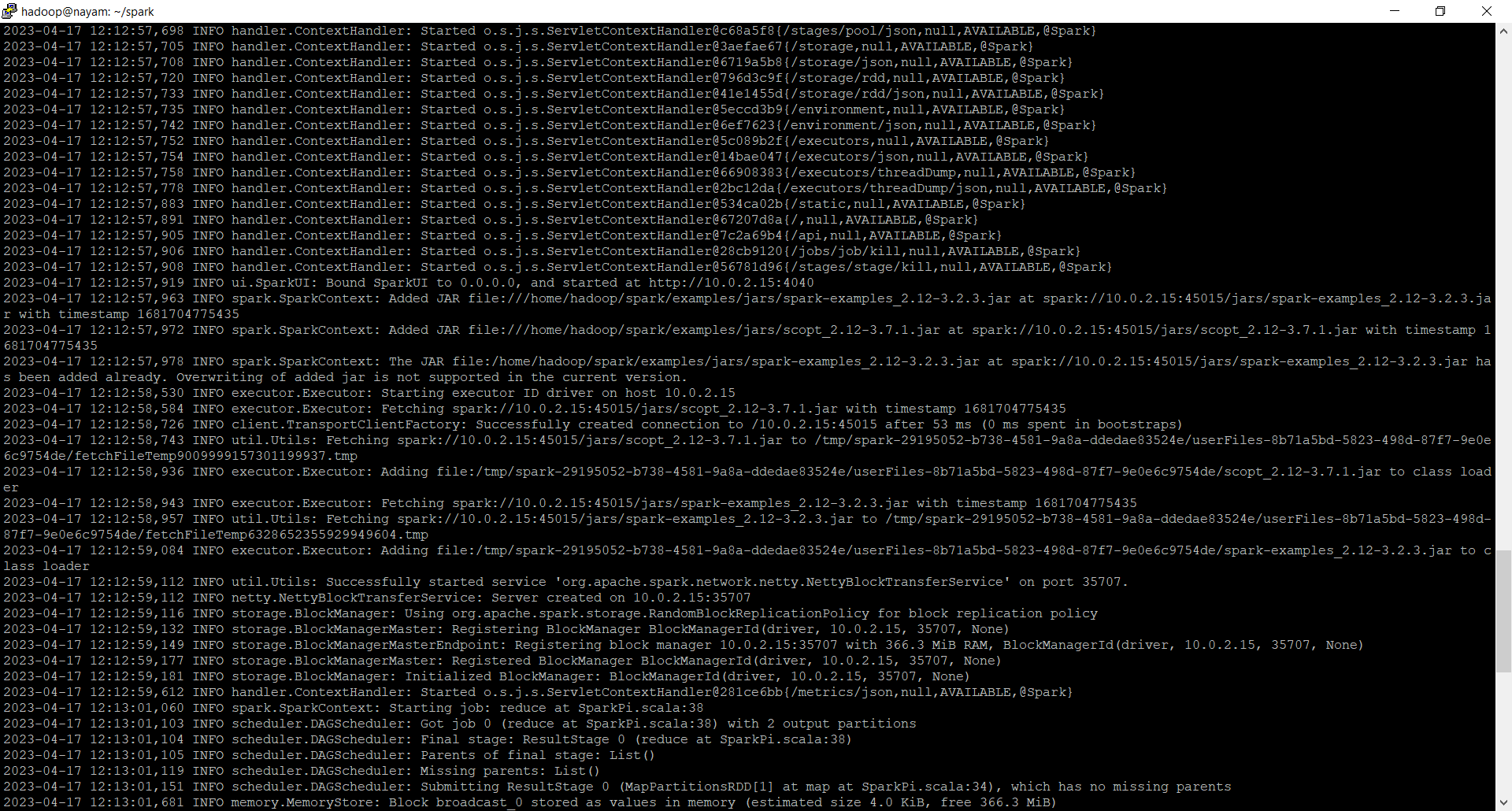
hadoop@nayam:~/spark$ vim ./conf/spark-env.sh

hadoop@nayam:~/spark$ cd /home/hadoop/spark

hadoop@nayam:~/spark$ bin/run-example SparkPi

there will be appended a line to the file : spark-env.sh

export SPARK\_DIST\_CLASSPATH=$(/home/hadoop/hadoop-3.2.4/bin/hadoop classpath)

Finally, by using bin/tun-example SparkPi , then, we got the test of Spark.