

## 1. install java

```
sudo apt-get install default-jdk -y
```

# Download Apache Spark

## 1) Create folder spark in home directory

open terminal and execute following commands

```
cd spark
```

## 2) run following command on terminal

```
apt-get install wget https://dlcdn.apache.org/spark/spark-3.2.0/spark-3.2.0-bin-hadoop3.2.tgz
```

or

download from url( after download copy tar file spark folder)

<https://spark.apache.org/downloads.html>.

## 3) run following command on terminal

```
tar -xvf spark-3.2.1-bin-hadoop3.2.tgz
```

## 4) set-up spark environment

run following command on terminal

```
gedit ~/.bashrc
```


This command will open the bashrc file in the gedit text editor, scroll down to the bottom of the bashrc file, and in the end add these lines

```
SPARK_HOME=/home/username/spark/spark-3.2.1-bin-hadoop3.2 export  
PATH=$PATH:$SPARK_HOME/bin:$SPARK_HOME/sbin
```

5) Finish the installation by using the source command –

## 6) Verify the Installation of Apache Spark

```
spark-3.2.1-bin-hadoop3.2/sbin/stop-worker.sh
spark-3.2.1-bin-hadoop3.2/sbin/stop-workers.sh
spark-3.2.1-bin-hadoop3.2/sbin/workers.sh
spark-3.2.1-bin-hadoop3.2/yarn/
spark-3.2.1-bin-hadoop3.2/yarn/spark-3.2.1-yarn-shuffle.jar
pccoe@pccoe-HP-280-G2-MT:~/spark$ gedit ~/.bashrc
pccoe@pccoe-HP-280-G2-MT:~/spark$ ssource ~/.bashrc
ssource: command not found
pccoe@pccoe-HP-280-G2-MT:~/spark$ source ~/.bashrc
pccoe@pccoe-HP-280-G2-MT:~/spark$ spark-shell
22/04/18 14:18:33 WARN Utils: Your hostname, pccoe-HP-280-G2-MT resolves to a loopback address: 127.0.1.1; using 16
face enp4s0)
22/04/18 14:18:33 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.spark.unsafe.Platform (file:/home/pccoe/spark/spark-3.2.1-bin-hado
2-3.2.1.jar) to constructor java.nio.DirectByteBuffer(long,int)
WARNING: Please consider reporting this to the maintainers of org.apache.spark.unsafe.Platform
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
22/04/18 14:18:38 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jar
Spark context Web UI available at http://10.12.0.101:4040
Spark context available as 'sc' (master = local[*], app id = local-1650271718876).
Spark session available as 'spark'.
Welcome to
```



```
version 3.2.1
```

8) After running above command find url (<http://10.12.0.101:4040> – ip may change) on terminal. Paste this url in browser and see spark cluster.

Detail steps available at : <https://netmiko.com/tutorials/data-science/install-apache-spark-on-ubuntu/>

## Scala Program :

```
val num=Array(1,2,3,56,78)
```

```
val data=sc.parallelize(num)
```

```
data.foreach(println)
```

or

declare string and print it .

Execute following commands on apache spark scala prompt.

```
var Var1 : String = "Ankit"
```

```
println(Var1)
```

or

create filename first.scala in *home*/spark directory

write following code to it and save.

```
// first.scala
```

Scala program to find a number is positive, negative or positive

```
object Ex{
```

```
    def main(args: Array[String]) {
```

```
    /**declare a variable*/  
  
    var number= -100;  
  
    if(number==0){  
  
        println("number is zero");  
  
    }  
  
    else if(number>0){  
  
        println("number is positive");  
  
    }  
  
    else{  
  
        println("number is negative");  
  
    }  
  
    }  
  
    }
```

Go to terminal and compile program by

scalac first.scala

Run using

scala Ex( object name given )

<https://youtu.be/jsIfdh0Mxo0>