Spark Installation Guide

# 1. Scala

## 1.1 Download Scala

### 1.1.1 Case 1: The device has a normal network connection

|  |
| --- |
| wget <https://www.scala-lang.org/files/archive/scala-2.13.15.tgz> |



### 1.1.2 Case 2: The device has no network connection

Visit the following link to download the appropriate packages for Linux.

[Index of /files/archive (scala-lang.org)](https://www.scala-lang.org/files/archive/) 



A screenshot of a computer

Description automatically generated

Go to the virtual machine's Settings, enable file sharing and add the path.

|  |
| --- |
| cd /mnt/hgfs/Scala |
| ls |
| cp scala-2.13.15.tgz /root/ |
| chmod 777 /root/scala-2.13.15.tgz |

## 1.2 Configuring Scala

After downloading the package, use one of the two methods above.

|  |
| --- |
| tar xvfz scala-2.13.15.tgz |
| mv scala-2.13.15 scala |
| mv scala /usr/local/share/scala |

A screen shot of a computer

Description automatically generated





After unzipping the folder, move the folder to /usr/local/share/scala. In this path, users will share files with each other. If any user wants to use it, go to that user and declare the environment variable of that user.

|  |
| --- |
| nano ~/.bashrc |

Add the following content to the file

|  |
| --- |
| #SCALA  export SCALA\_HOME=/usr/local/share/scala  export PATH=$PATH:$SCALA\_HOME/bin |
| source ~/.bashrc |
| exec $SHELL |

A black background with red and green text

Description automatically generated

|  |
| --- |
| scala |

# 2. Spark

## 2.1 Download Spark

### 2.1.1 Case 1: The device has a normal network connection

|  |
| --- |
| wget https://dlcdn.apache.org/spark/spark-3.5.3/spark-3.5.3-bin-hadoop3.tgz |

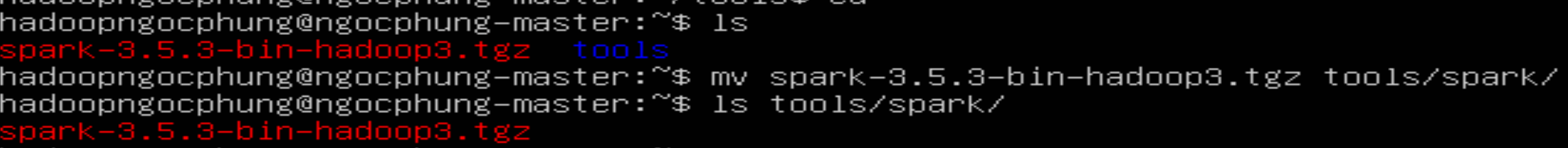
### 2.1.2 Case 2: The device has no network connection

[Index of /spark/spark-3.5.3 (apache.org)](https://dlcdn.apache.org/spark/spark-3.5.3/)

|  |
| --- |
| cd /mnt/hgfs/Spark |
| ls |
| cp spark-3.5.3-bin-hadoop3.tgz /home/hadoopngocphung/tools/spark |
| chmod 777 home/hadoopngocphung/tools/spark/spark-3.5.3-bin-hadoop3.tgz |

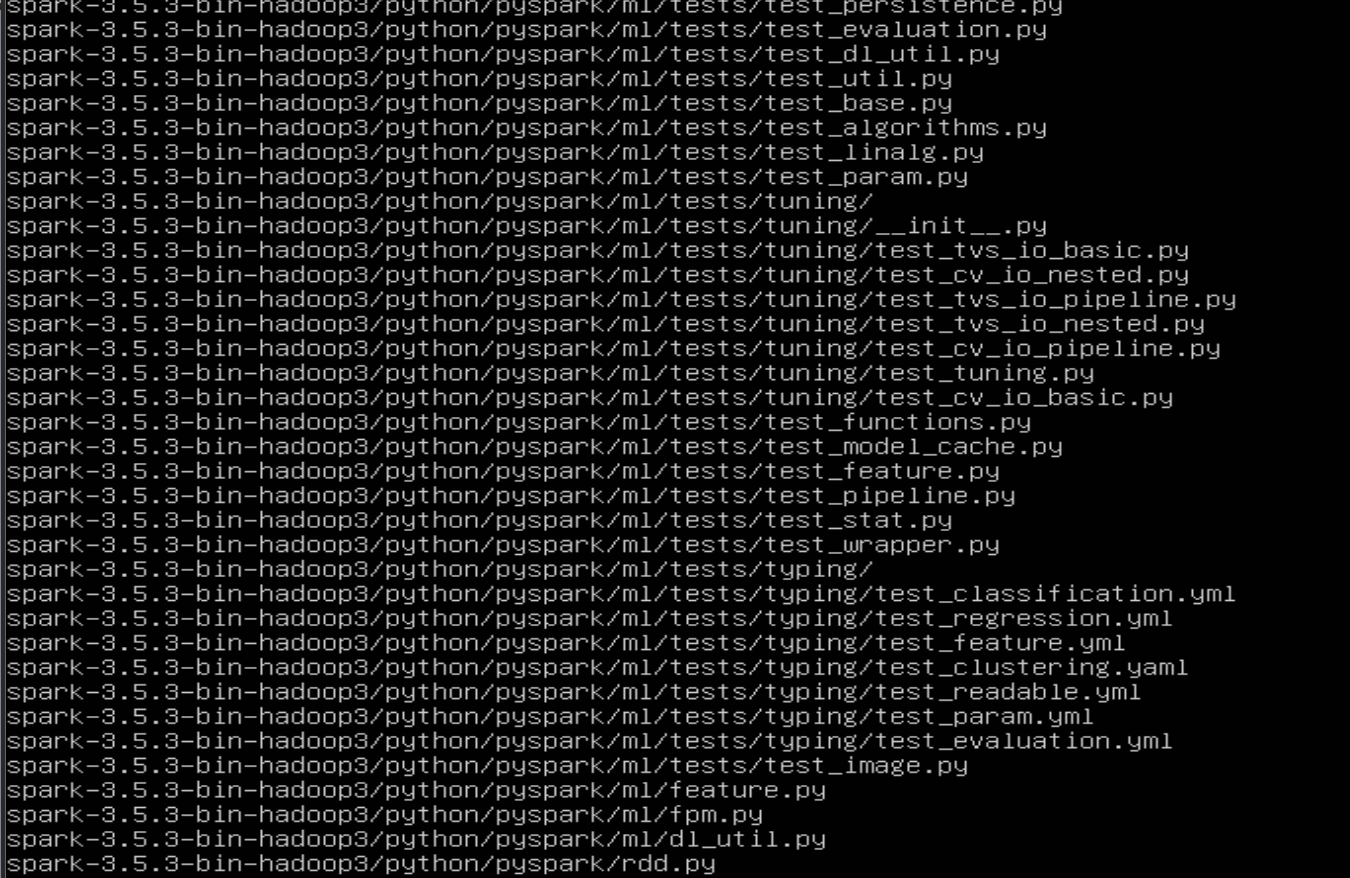
## 2.2 Configuring Spark

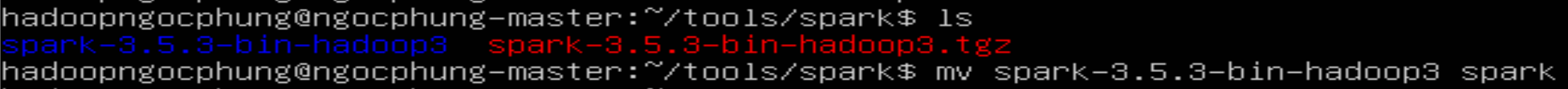
After downloading the package, use one of the two methods above.



Unzip the file and rename it for easy management

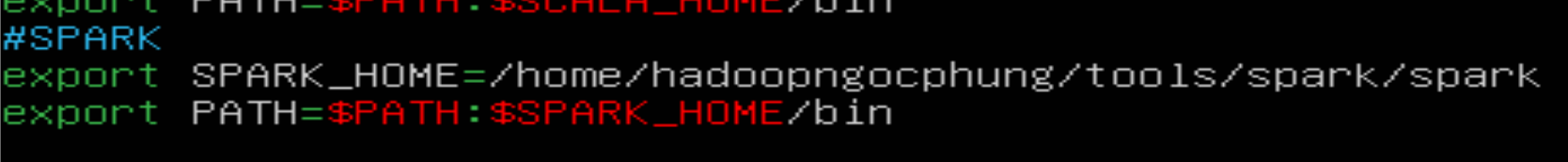
|  |
| --- |
| tar -xvzf spark-3.5.3-bin-hadoop3.tgz |
| mv spark-3.5.3-bin-hadoop3 spark |





- Go to bashrc file to declare environment for Spark

|  |
| --- |
| nano ~/.bashrc |
| #SPARK  export SPARK\_HOME=/home/hadoopngocphung/tools/spark/spark  export PATH=$PATH:$SPARK\_HOME/bin |



Restart spark

|  |
| --- |
| spark-shell |

A computer screen with white text

Description automatically generated

|  |
| --- |
| :quit |

# 3. Pyspark