

How to import the virtual machine:

- 1) Install VirtualBox (<https://www.virtualbox.org/wiki/Downloads>)
- 2) Download BigData_LocalHadoop.ova by using the link available on the web site of the course
- 3) Launch VirtualBox
- 4) Import the Virtual Machine by selecting: File -> Import appliance
 - Select BigData_LocalHadoop.ova
- 5) Change the path of the shared folder
 - Select the name of the imported Virtual Machine and then click Settings
 - Select the Shared Folders tab
 - Click on the "data" folder and edit it
 - Set the value of "Folder path" to the local folder of your PC that you want to share with the Virtual Machine (Do not change the folder name). Pay attention that the shared folder will be mounted in read/write mode. Hence, I suggest to do not select c:\, d:\ or / as shared folder
 - To access the shared folder in the Virtual Machine use the folder /media/sf_data/

Operating system of the Virtual Machine: Ubuntu 14.04

Software:

- Hadoop and Spark
- Eclipse

username: bigdata

password: BigData

When the virtual machine starts the HDFS file system is in the SAFE MODE status (i.e., read-only status)

Use the following command to turn of SAFE MODE:

```
hadoop dfsadmin -safemode leave
```

The folder /home/bigdata/Programs contains two example Java programs based on the MapReduce programming paradigm:

- Exercise1 : Word coount
- Exercise1WitCombiner: Word count with combiner