**Learning Objective 2**

1, install the tools

1. Virtualisation application

[VMWare](http://vmware.com/), [VirtualBox](http://virtualbox.org/), and other virtualization environments.

2. Hortonworks vm installed virtual box <http://hortonworks.com/hdp/downloads/>

3. Windows: git bash shell <http://git-scm.com/downloads>

4. Windows: a ssh directory browser<http://www.swish-sftp.org/>

5. Filezilla<https://filezilla-project.org/>

6. 7 zip

2, install the virtual machine

instructions here

<http://hortonworks.com/wp-content/uploads/unversioned/pdfs/InstallingHortonworksSandbox2onWindowsusingVB.pdf>

3. Download spark binary. please note over time the versions and download links change.

change

this is the current compatible jar for the Hortonworks VM

wget <http://public-repo-1.hortonworks.com/HDP-LABS/Projects/spark/1.2.0/spark-1.2.0.2.2.0.0-82-bin-2.6.0.2.2.0.0-2041.tgz>

4. Set up ssh with HortonWorks VM

1. connect with filezilla
2. host 127.0.0.1

port 2222

username root

password hadoop

5 Set the SPARK\_HADOOP\_VERSION flag in the Hortonworks vm environment variable

ssh root@127.0.0.1 -p 2222

hadoop

cd /etc/

vi bashrc

then append

export SPARK\_HADOOP\_VERSION=2.2.0

then :wq then init 6

then in the Hortonworks vm ’s Bash shell check with printenv you should should see

6.Install the Apache Spark Binaries and Example Source Code

chmod -R 4777 spark-1.2.0.2.2.0.0-82-bin-2.6.0.2.2.0.0-2041.tgz

tar xvfz spark-1.2.0.2.2.0.0-82-bin-2.6.0.2.2.0.0-2041.tgz

once unpacked rename the directory to spark1.2.0 or something similar

7.

cd sp\*

./bin/spark-shell

cd sp\*

sbt/sbt assembly

if the spark build fails or hangs on a resolving …….. or fails

then delete the following directories in the root folder on the vm

.m2 .ivy .sbt

see screenshot below in the ssh directory browser with show hidden folders enabled in the windows control panel

8. Running the spark api from the scala interface