Installation of Spark, scikit-learn TensorFlow, Keras

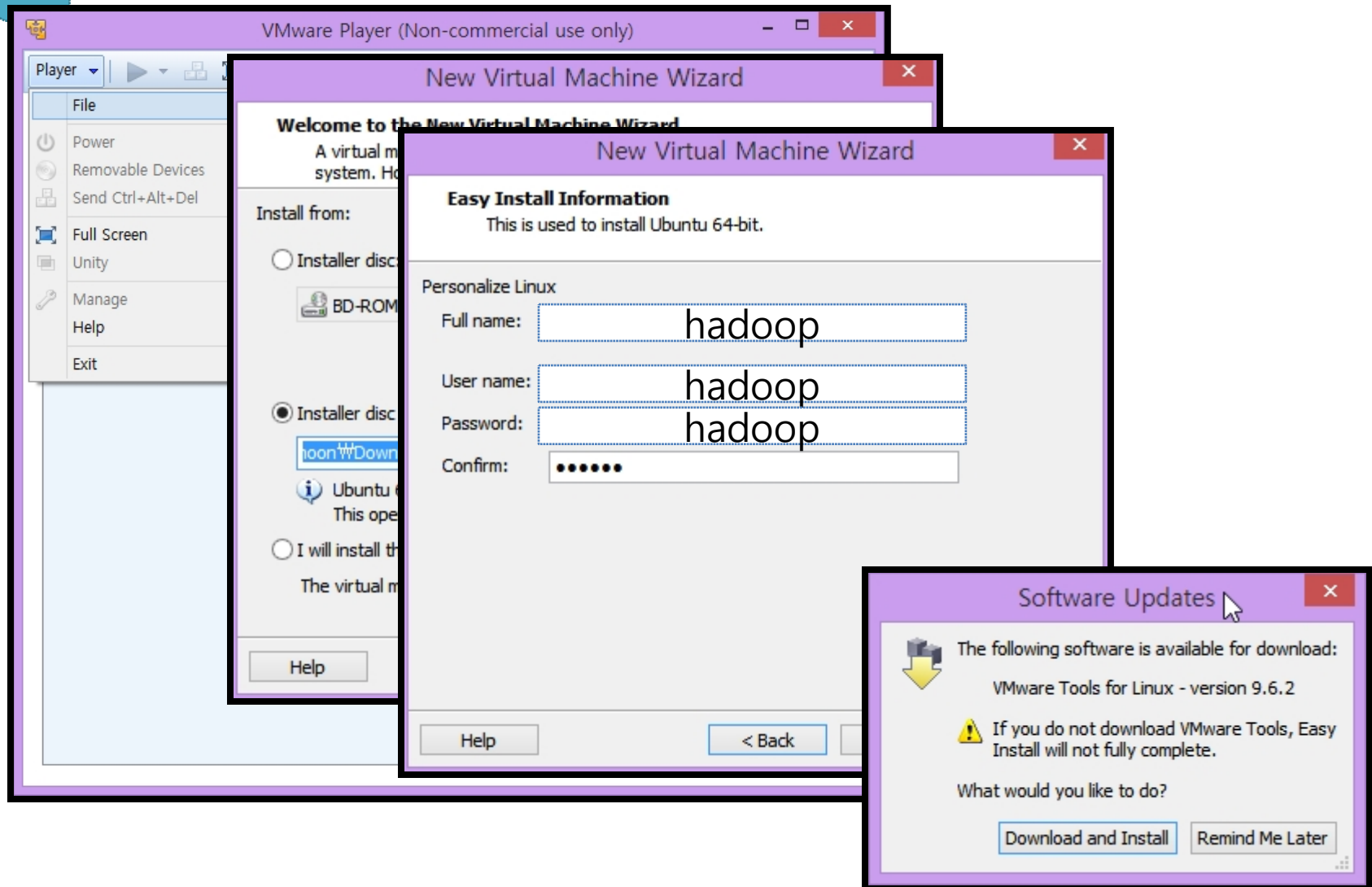## Younghoon Kim

(nongaussian@hanyang.ac.kr)

Requirements

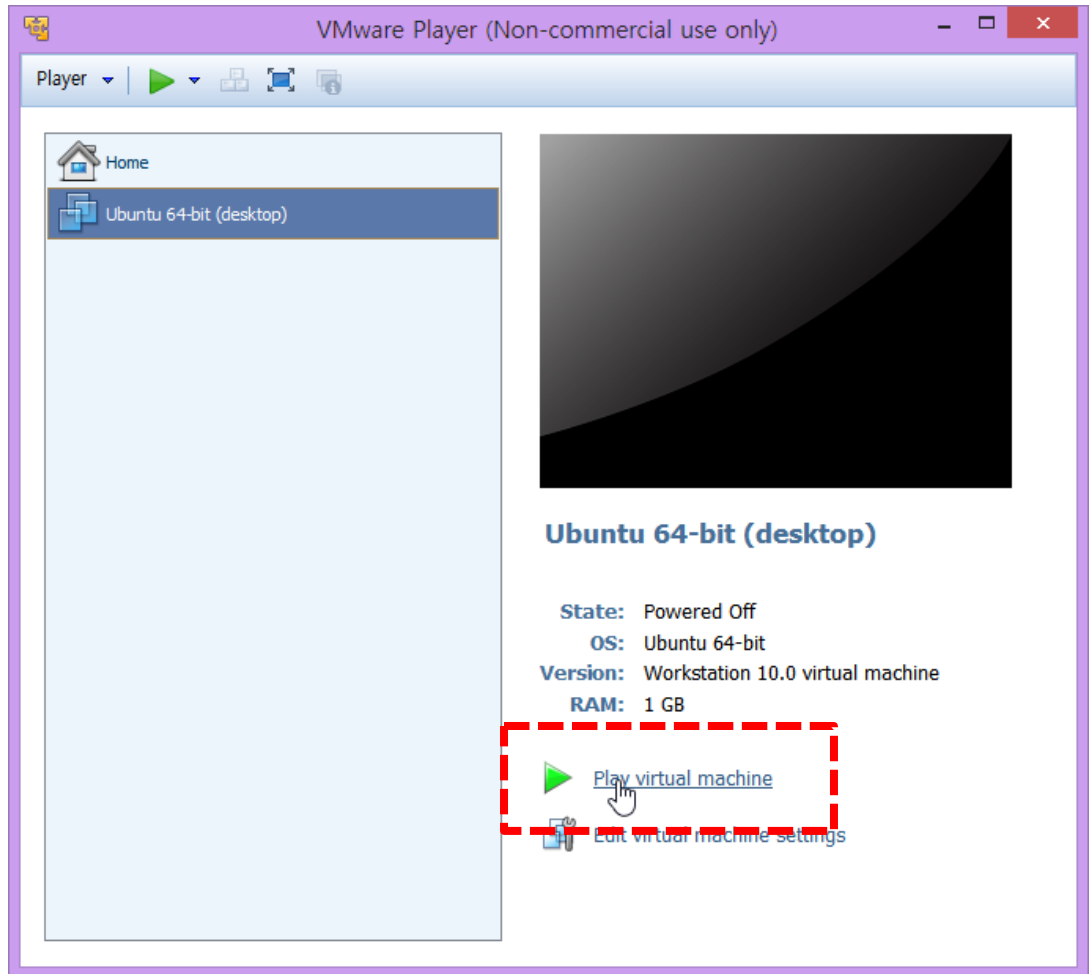VMWare + Ubuntu 16.0
(Mac) VirtualBox + Ubuntu 16.0

Vmware (VIrtualBox) and Ubuntu
16.0 will be distributed in the
class.

Install Ubuntu On VMware

Start Ubuntu!

- Select Ubuntu & play the virtual machine

Install Program

- Install vim & ssh
  - $ sudo apt-get install vim
  - $ sudo apt-get install ssh

- Execute the following commands:
  - $ ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
  - $ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys

Install Java 1.8

- $ sudo add-apt-repository ppa:webupd8team/java
- $ sudo apt-get update
- $ sudo apt-get install oracle-java8-installer

Download Spark 2.2.0

- Download spark 2.2.0 package
  - Open the web page on ubuntu firefox:
  https://spark.apache.org/downloads.html
  - spark-2.2.0-bin-hadoop2.7.tgz file will be appeared in $HOME/Downloads

  - Or $ wget https://d3kbcqa49mib13.cloudfront.net/spark-2.2.0-bin-hadoop2.7.tgz -P $HOME/Downloads

"wget" is recommended.

Install Spark 2.2.0

There is space between "tgz" and ".".

■Install spark
- –$ cd $HOME
- –$ mv Downloads/spark-2.2.0-bin-hadoop2.7.tgz .
- –$ tar -zxvf spark-2.2.0-bin-hadoop2.7.tgz
- –$ mkdir spark
- –$ mv spark-2.2.0-bin-hadoop2.7/* spark/
- –$ vi ~/.bashrc

  Add the next line on the bottom of ~/.bashrc
      export SPARK_HOME="$HOME/spark"
- –$ source ~/.bashrc

Configuration

- Configuration
  - $ cp $SPARK_HOME/conf/slaves.template $SPARK_HOME/conf/slaves

Execute Spark 2.2.0

■Execute spark
–$ $SPARK_HOME/sbin/start-all.sh

■Check
–Open the browser
–Go to "localhost:8080"

Spark 2.2.0 ████████████████████████████

URL: ████████████████
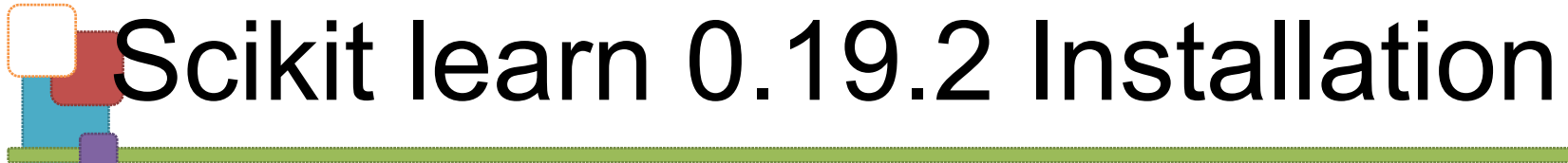
REST URL: ████████████████

Alive Workers: 0

Cores in use: 0 Total, 0 Used

Memory in use: 0.0 B Total, 0.0 B Used

Applications: 0 Running, 0 Completed

Drivers: 0 Running, 0 Completed

Status: ALIVE

Workers

Worker Id	Address	State	Cores	Memory

Running Applications

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration

Scikit learn 0.19.2 Installation

- Scikit-learn requires:
  - Python (>= 2.7 or >= 3.3) : this class requires python 2, and ubuntu has python 2 already.
  - NumPy (>= 1.8.2)
  - SciPy (>= 0.13.3)

- Pip installation
  - $ sudo apt-get install python-pip
  - $ sudo pip install numpy
  - $ sudo pip install scipy

- Scikit-learn installation
  - $ sudo pip install –U scikit-learn

- Installation Checking:
  - $ python
  - >>> import sklearn

TensorFlow 1.8.0 Installation

▪Installation for python 2 without GPU support (for ubuntu on VM ware)
  ▪$ sudo pip install tensorflow==1.8.0

▪Version checking
  ▪$ python
  ▪>>> import tensorflow as tf
  ▪>>> tf.__version__

```
In [1]: import tensorflow as tf

In [2]: tf.__version__
Out[2]: '1.3.0'
```

▪(Optional but highly recommended) IPython installation
  ▪$ sudo pip install ipython

Hello, Tensorflow

```
$ python

>>> import tensorflow as tf
>>> hello = tf.constant('Hello, TensorFlow!')
>>> sess = tf.Session()
>>> print(sess.run(hello))
Hello, TensorFlow!
>>> a = tf.constant(10)
>>> b = tf.constant(32)
>>> print(sess.run(a + b))
42
>>>
```

Keras 2.0 Installation

- Installation
    - $ sudo pip install keras

- Installation Checking
    - $ python
    - >>> import keras