Install Hadoop on Linux based OS

Mac OS/Redhat/Ubuntu
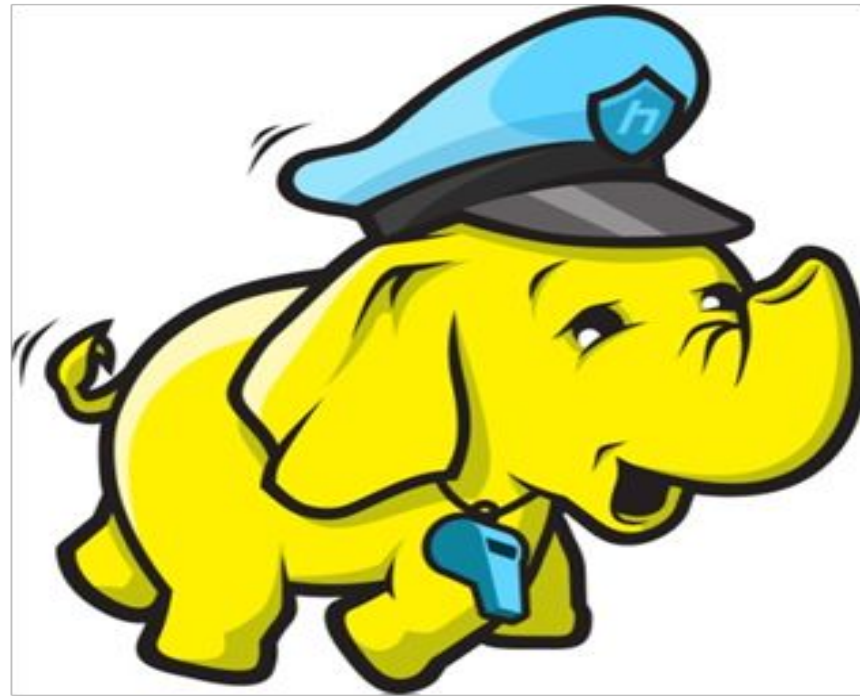
Agenda

Install Java & ssh on your Ubuntu
Visit Apache Hadoop website
Download Hadoop tar.gz file
Untar the Hadoop tar.gz file
Directory structure of Hadoop
What is ssh (Secure Shell)
How to configure password less ssh
Configuration files in Hadoop
Modify configuration files
Format Namenode
Start the services of Hadoop
Verify Hadoop services
Stop the services of Hadoop

Install Java & ssh on your Ubuntu

Run the below commands in terminal to install Java

sudo add-apt-repository ppa:webupd8team/java
sudo apt-get update
sudo apt-get install oracle-java7-installer

Run the below command in terminal to install ssh

sudo apt-get install openssh-server openssh-client

Make sure that your Ubuntu is connected to internet during execution of these commands

# Download Hadoop tar.gz file

http://www.motorlogy.com/apache/hadoop/common/hadoop-1.2.1/

Index of /apache/hadoop/co...

www.motorlogy.com/apache/hadoop/common/hadoop-1.2.1/

Index of /apache/hadoop/common/hadoop-1.2.1

Name	Last modified	Size	Description
Parent Directory		-	
hadoop-1.2.1-1.i386.rpm	22-Jul-2013 15:20	39M	
hadoop-1.2.1-1.i386.rpm.mds	22-Jul-2013 15:45	1.1K	
hadoop-1.2.1-1.x86_64.rpm	22-Jul-2013 15:30	35M	
hadoop-1.2.1-1.x86_64.rpm.mds	22-Jul-2013 15:45	1.1K	
hadoop-1.2.1-bin.tar.gz	22-Jul-2013 15:27	36M	
hadoop-1.2.1-bin.tar.gz.mds	22-Jul-2013 15:45	1.1K	
hadoop-1.2.1.tar.gz	22-Jul-2013 15:27	61M	
hadoop-1.2.1.tar.gz.mds	22-Jul-2013 15:45	958	
hadoop_1.2.1-1_i386.deb	22-Jul-2013 15:23	39M	
hadoop_1.2.1-1_i386.deb.mds	22-Jul-2013 15:45	1.1K	
hadoop_1.2.1-1_x86_64.deb	22-Jul-2013 15:33	35M	
hadoop_1.2.1-1_x86_64.deb.mds	22-Jul-2013 15:45	1.1K	
releasenotes-1.2.1.html	22-Jul-2013 15:15	519K	

Apache/2.2.22 Server at www.motorlogy.com Port 80

Click on Save File

Untar the Hadoop tar.gz file
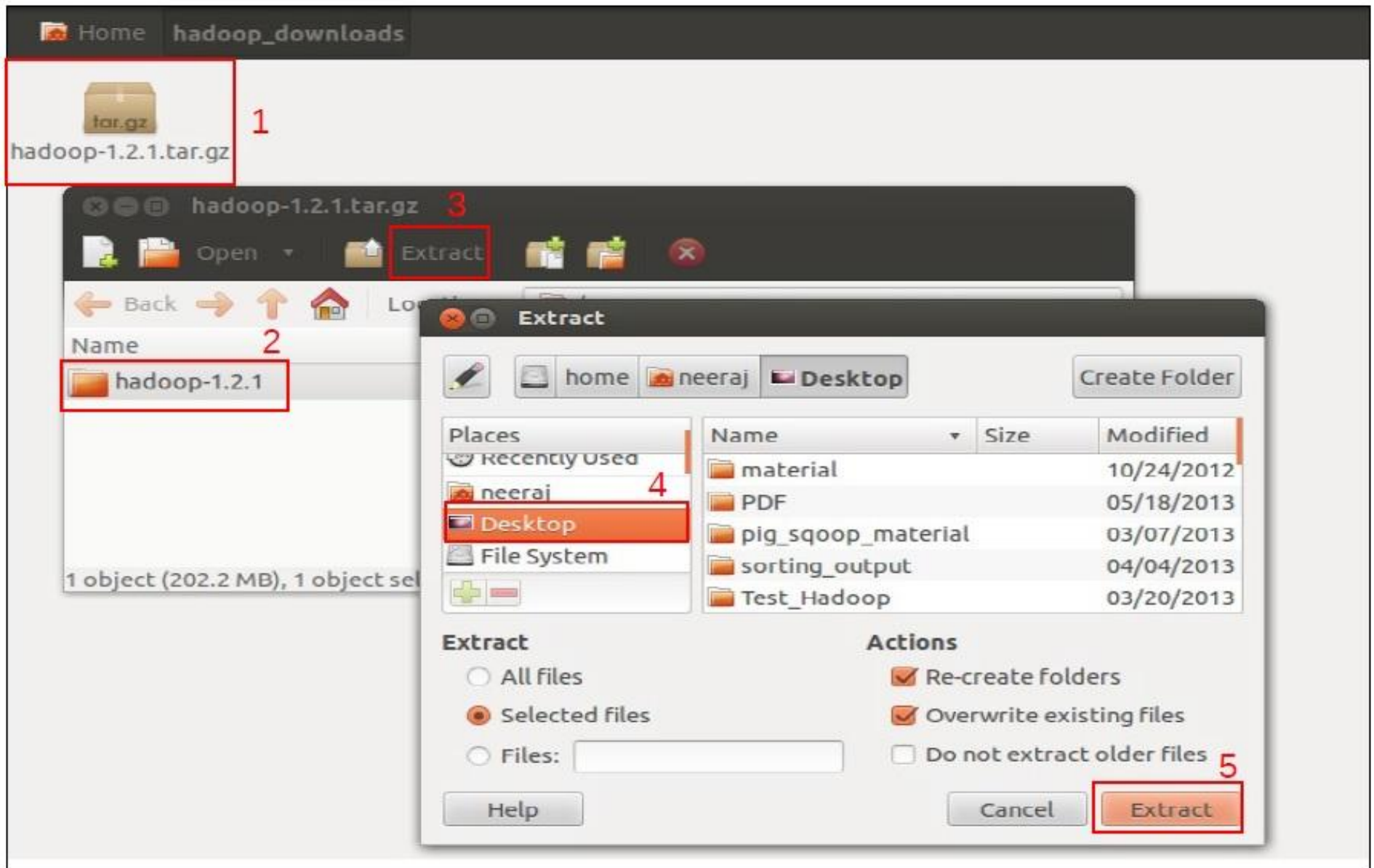
Directory structure of Hadoop

HADOOP_HOME

bin   [contains scripts to start/stop Hadoop]

conf [contains configuration files of Hadoop]

lib   [contains libraries to run Hadoop]

src   [contains the source code of Hadoop]

logs [contains log files]

Configuration files in Hadoop

There are two types of important configuration files in Hadoop...

Default configuration files (Read-only) :

       src/core/core-default.xml

       src/hdfs/hdfs-default.xml

       src/mapred/mapred-default.xml

Site-specific configuration files (can be modified) :

       conf/core-site.xml

       conf/hdfs-site.xml

       conf/mapred-site.xml

What is SSH (Secure Shell)

Tool, used to remotely connect to other machine.

Namenode uses SSH for communicating with Datanodes.

Jobtracker uses SSH for communicating with Tasktrackers.

SSH must be installed, even if you are installing Hadoop on a single machine.
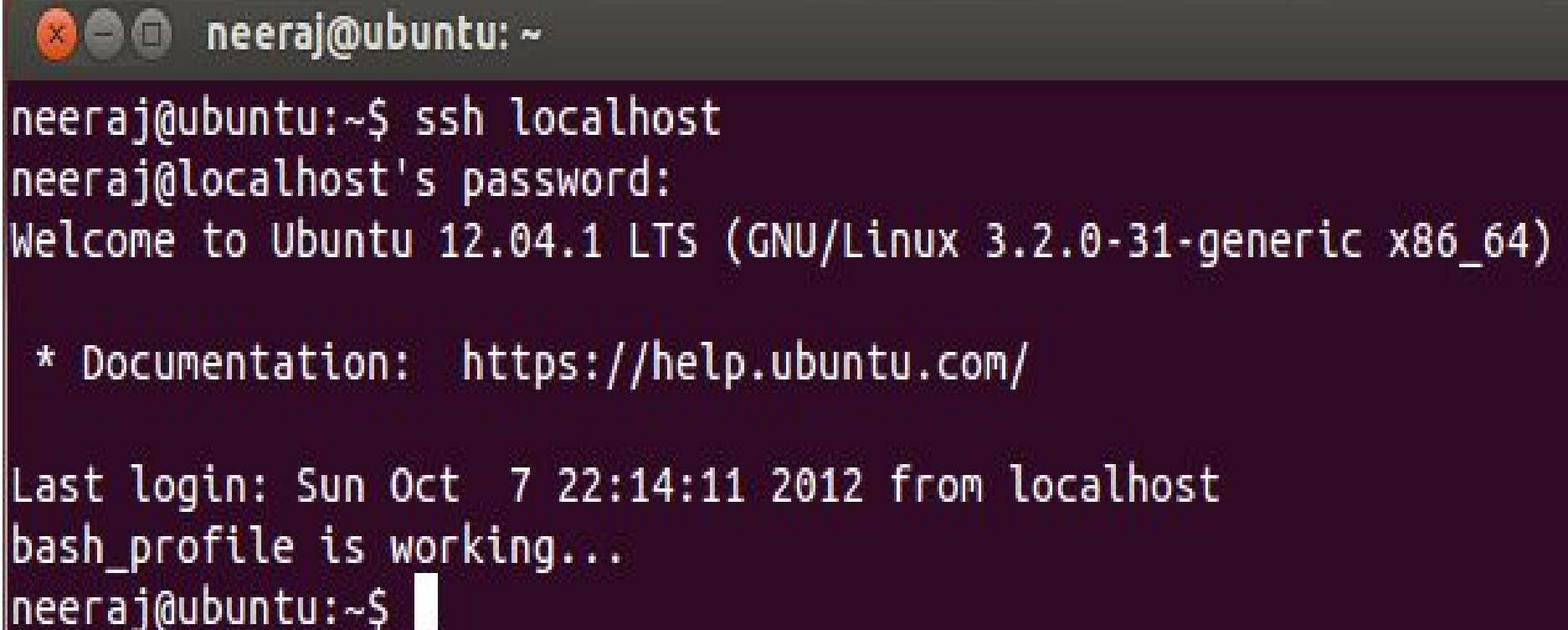
ssh < IP of another machine >

For example...   ssh 192.168.342.187
                        ssh 127.0.0.1
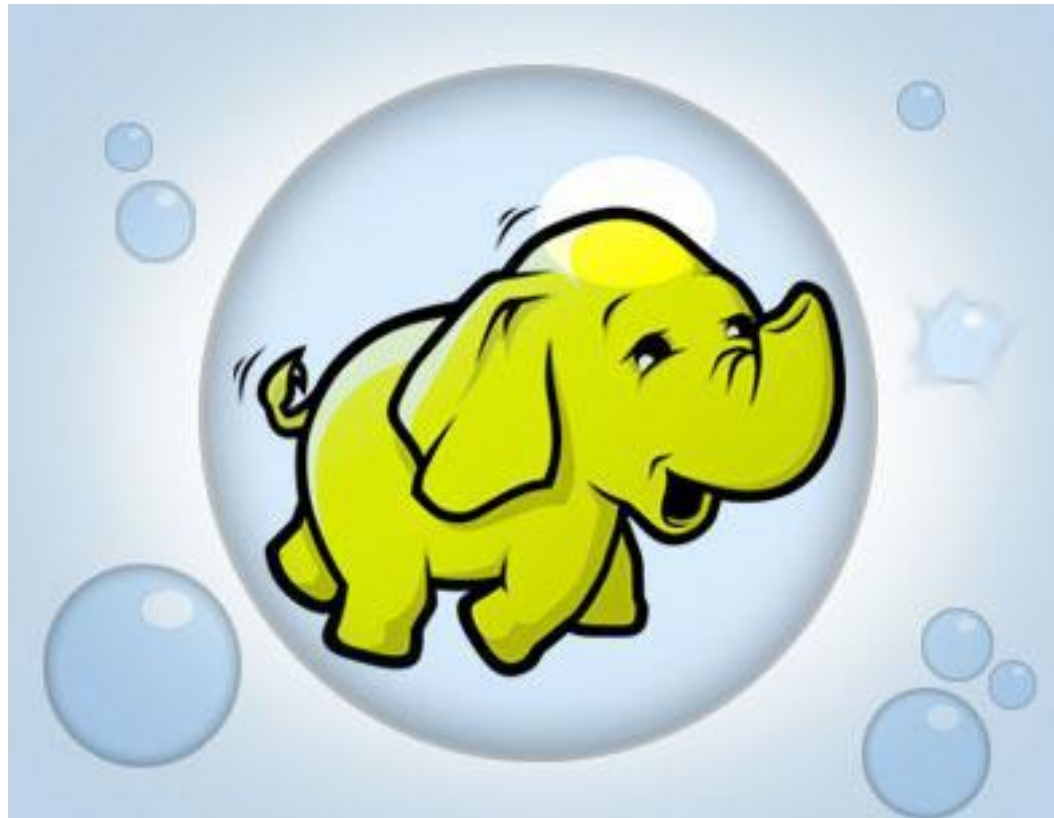   ssh localhost

Run the following command on your terminal

ssh localhost

…Thanks…

For online Hadoop training, send mail to neeraj.ymca.2k6@gmail.com