Installing hadoop 2.4.0 on Ubuntu server

Installing hadoop 2.4.0

First of all, THESE ARE IMPORTANT FOR INDRODUCTION

*Ubuntu server 14.04
*VM in VirtualBox with 1GB RAM
*SWAP SPACE :   SELECT **more than 2 GB**

Up-to-date system.
[crayon-662f6f150ec62779640995/]
Moreover, it is advisable not to run Hadoop services through a general-purpose user, so the next step consists in adding a group hadoop and a user hadoop-user belonging to that group
[crayon-662f6f150ec68456032510/]

Installing Java
The mentioned tutorials suggest a potentially unsafe procedure in order to install the jdk through apt-get,

[crayon-662f6f150ec6b710648960/]
Finally, a couple of environment variables should be set up so that the java executables are in $PATH and hadoop knows where java has been installed: this is easily accomplished adding
[crayon-662f6f150ec6d638109301/]
at the end of /etc/profile <<< We need to edit that to set automatica**lly JAVA PATH**
[crayon-662f6f150ec6f828205205/]

Setup SSH

All communications with Hadoop are encrypted via SSH, thus the

corresponding server should be installed:
[crayon-662f6f150ec71379661590/]

and the hadoop-user must be associated to a key pair and subsequently granting its access to the local machine:
[crayon-662f6f150ec73563890622/]

Now hadoop-user should be able to access via ssh to localhost **without** (WE SET THE PASS = "null")providing a password:
[crayon-662f6f150ec75940513288/]

Disable IPV6

Hadoop and IPV6 do not agree on the meaning of 0.0.0.0 address,

LOOK AT : ??>>>>> **/etc/sysctl.conf**
[crayon-662f6f150ec78559204012/]
REBOOT SYSTEM

CONTROL
[crayon-662f6f150ec7a199616622/]

Hadoop

Download and install Hadoop

Download hadoop-2.4.0.tar.gz, unpack it and move the results in /usr/local, adding a symlink using the more friendly name hadoop and changing ownership of the directory contents to the hadoop-user user:
[crayon-662f6f150ec7c853084389/]

Setup the dedicated user environment
Switch to the hadoop-user user and add the

following lines at the end of ~/.bashrc:

[crayon-662f6f150ec7e197511321/]
To have the new environment variables in place,

1- reload .bashrc through "**source .bashrc**"

2- then open **/usr/local/hadoop/etc/hadoop/hadoop-env.sh** ,

3- uncomment the line setting JAVA_HOME and set its value to the jdk directory:
[crayon-662f6f150ec81970287357/]

Configure Hadoop

Before being able to actually use the hadoop file system it is necessary to modify some configuration files inside **/usr/local/hadoop/etc/hadoop**
All such files follow the an XML format, and the updates should concern the top-level node configuration
(likely empty after the hadoop installation). Specifically:

- in **yarn-site.xml** :
  [crayon-662f6f150ec83305468838/]

- look in **core-site.xml**:
  [crayon-662f6f150ec85288257060/]

- look in **mapred-site.xml**
  [crayon-662f6f150ec87501336244/]

- nano hdfs-site.xml:
  [crayon-662f6f150ec89412976571/]

Run these commands.
[crayon-662f6f150ec8b707337453/]

Formatting the distributed file system

USER should be **hadoop user**
[crayon-662f6f150ec8d018947841/]
Find the these 2 file and run

start-dfs.sh

start-yarn.sh
[crayon-662f6f150ec8f043895703/]
[crayon-662f6f150ec92606867723/]