

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-6684c56a526be598808432/]

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-6684c56a526c9932057746/]

Installing Java

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

[crayon-6684c56a526cc454957507/]

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-6684c56a526cf716054521/]

at the end of /etc/profile <<< We need to edit that to set automatically **JAVA PATH**

[crayon-6684c56a526d2255274987/]

Setup SSH

All communications with Hadoop are encrypted via SSH, thus the

corresponding server should be installed:

```
[crayon-6684c56a526d5512786184/]
```

and the `hadoop-user` must be associated to a key pair and subsequently granting its access to the local machine:

```
[crayon-6684c56a526d8294235181/]
```

Now `hadoop-user` should be able to access via `ssh` to `localhost` **without** (WE SET THE PASS = "null")providing a password:

```
[crayon-6684c56a526db527303117/]
```

Disable IPV6

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

LOOK AT : `??>>>> /etc/sysctl.conf`

```
[crayon-6684c56a526de832441761/]
```

REBOOT SYSTEM

CONTROL

```
[crayon-6684c56a526e1154686430/]
```

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-6684c56a526e4008844193/]
```

Setup the dedicated user environment

Switch to the `hadoop-user` user and add the

following lines at the end of ~/.bashrc:

```
[crayon-6684c56a526e8441647678/]
```

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-6684c56a526eb077352525/]
```

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-6684c56a526ee627481746/]
- look in **core-site.xml**:
[crayon-6684c56a526f0709600948/]
- look in **mapred-site.xml**
[crayon-6684c56a526f3469094996/]
- nano **hdfs-site.xml**:
[crayon-6684c56a526f6014957159/]

Run these commands.

```
[crayon-6684c56a526f9186710023/]
```

Formatting the distributed file system

USER should be **hadoop user**

```
[crayon-6684c56a526fc151872413/]
```

Find the these 2 file and run

start-dfs.sh

start-yarn.sh

[crayon-6684c56a526ff205639953/]

[crayon-6684c56a52703889199274/]