

Installation d'Hadoop

(Ubuntu 22.04.3, 24/12/23)

Version : **3.3.6**

OS : **Ubuntu 22.04.3**

1. Installation de Java
2. Installation d'Hadoop
 - a. Configuration SSH
 - b. Ajout PATH
 - c. Hadoop Streaming
3. HDFS

Sur une machine fraîchement installée, pensez à :

```
xana@xana-VirtualBox:~$ sudo apt update && sudo apt upgrade -y
```

Puis à :

```
xana@xana-VirtualBox:~$ sudo snap refresh
```

Ces opérations prennent chacune quelques minutes

1.Installation de Java

Commencer par voir si java est installé , sur ubuntu il n'est pas installé par défaut:

```
xana@xana-VirtualBox:~$ java -version
La commande « java » n'a pas été trouvée, mais peut être installée avec :
sudo apt install openjdk-11-jre-headless # version 11.0.20-1+1.0ubuntu1-22.04.05
```

Pour ce faire on va devoir installer 2 packages default-jre et default-jdk :

JRE : Java Runtime Environment

```
xana@xana-VirtualBox:~$ sudo apt install default-jre
```

A la suite de ça on peut vérifier si java est en effet installé :

```
xana@xana-VirtualBox:~$ java -version
openjdk version "11.0.21" 2023-10-17
OpenJDK Runtime Environment (build 11.0.21+9-post-Ubuntu-0ubuntu122.04)
OpenJDK 64-Bit Server VM (build 11.0.21+9-post-Ubuntu-0ubuntu122.04, mixed mode, sharing)
```

JDK : Java Development Toolkit

```
xana@xana-VirtualBox:~$ sudo apt install default-jdk
```

On peut alors vérifier si le compilateur java (javac) est installé :

```
xana@xana-VirtualBox:~$ javac -version
javac 11.0.21
```

2.Installer la configuration pour communiquer en SSH avec le cluster

Curieusement...ssh n'est pas installé de base sur ubuntu , il faut donc installer les packages ssh & pdsh :

SSH (Secure Shell) : installe le client et le serveur OpenSSH

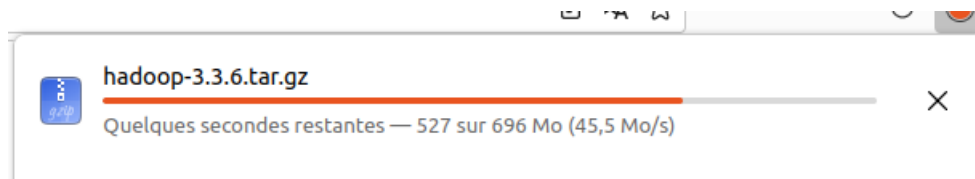
```
xana@xana-VirtualBox:~$ sudo apt-get install ssh
```

PDSH (??Shell) : d'une machine client permet de maintenir la co aux machines distantes ?

```
xana@xana-VirtualBox:~$ sudo apt-get install pdsh
```

3.Télécharger Hadoop

Par ligne de commande ou sur le site , télécharger la version recommandée :



Télécharger aussi le fichier qui contient la shasum :

```
wget https://downloads.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz.sha512
```

Vérifier que les hash correspondent bien (manuellement):

```
xana@xana-VirtualBox:~/Téléchargements$ shasum -a 512 hadoop-3.3.6-src.tar.gz
9a919288db245a897aec6794fdd80e1a21f10c5e46dabdbc61280f89a78f27ab9bf8ad9f14329f61f974a3e6d8f54a3e52cd8dbe0c42c34bb04bff773d03f77a hadoop-3.3.6-src.tar.gz
```

Ou par script :

```

import sys
import re
import subprocess

tar_folder = sys.argv[1]
key_file = sys.argv[2]

print( tar_folder,key_file)

source_text = subprocess.run(f"cat {key_file}", capture_output=True, shell=True).stdout
source_utf = source_text.decode("utf-8")
source_sum = re.findall(r"= (.*)\n",source_utf)[0]

tar_512 = subprocess.run(f"shasum -a 512 {tar_folder}", capture_output=True, shell=True).stdout
tar_utf = tar_512.decode("utf-8")
tar_sum = tar_utf.split()[0]

if tar_sum == source_sum :
    print(f"[SUCCESS] Matching SHA512SUM of files : \n {tar_folder} \n {key_file} \n")
else :
    print("!!! Corrupted file, please review download or check throughly any security danger ")

```

Unzip et le copier vers l'emplacement de son choix :

```
xana@xana-VirtualBox:~/Téléchargements$ tar -xvzf hadoop-3.3.6.tar.gz -C /home/xana/Documents/hadoop test
```

Vous devriez avoir :

```
xana@xana-VirtualBox:~$ ls Documents/hadoop_test/hadoop-3.3.6
bin  etc  include  lib  libexec  LICENSE-binary  licenses-binary  LICENSE.txt  NOTICE-binary  NOTICE.txt  README.txt  sbin  share
```

4.Installation

Associer l'installation Java courante avec l'application Hadoop :

```
xana@xana-VirtualBox:~/Documents/hadoop_test/hadoop-3.3.6$ nano etc/hadoop/hadoop-env.sh
```

```

# The java implementation to use. By default, this environment
# variable is REQUIRED on ALL platforms except OS X!
export JAVA_HOME= /usr/lib/jvm/java-11-openjdk-amd64

```

```
54 export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64 ( pas d'espaces ...)
```

Il faut retrouver dans son répertoire où se situe son installation

Ensuite vous pouvez tester dans le repertoire Hadoop-3.3.6:

```

xana@xana-VirtualBox:~/Documents/hadoop_test/hadoop-3.3.6$ bin/hadoop
Usage: hadoop [OPTIONS] SUBCOMMAND [SUBCOMMAND OPTIONS]
or      hadoop [OPTIONS] CLASSNAME [CLASSNAME OPTIONS]
where CLASSNAME is a user-provided Java class

OPTIONS is none or any of:

buildpaths          attempt to add class files from build
--config dir        Hadoop config directory

```

