



**Mansoura University**  
**Faculty of Computers and Information**  
**Sciences**  
**Department of Computer Science**  
**First Semester- 2020-2021**



# **[CS412P] Distributed Systems**

**Grade : Fourth grade**

**By : Zeinab Awad**

# APACHE HADOOP INSTALLATION

## Apache Hadoop :

- Apache Hadoop is an open source framework used for distributed storage as well as distributed processing of big data on clusters of computers which runs on commodity hardwares. Hadoop stores data in Hadoop Distributed File System (HDFS) and the processing of these data is done using MapReduce. YARN provides API for requesting and allocating resource in the Hadoop cluster.
- The Apache Hadoop framework is composed of the following modules :
  - Hadoop Common
  - Hadoop Distributed File System (HDFS)
  - YARN
  - MapReduce

# APACHE HADOOP INSTALLATION

## Pseudo Distributed Mode :

- We will install HDFS (Name node and Data node), YARN, MapReduce on the single node cluster in Pseudo Distributed Mode which is distributed simulation on a single machine.

Each Hadoop daemon such as hdfs, yarn, map reduce etc. will run as a separate/individual java process.

- The steps to install hadoop on Ubuntu 18.04 will include :

- How to add users for Hadoop Environment ?
- How to install and configure the Oracle JDK ?
- How to configure password less SSH ?
- How to install Hadoop and configure necessary related xml files?
- How to start the Hadoop Cluster?
- How to access NameNode and ResourceManager Web UI ?

# APACHE HADOOP INSTALLATION

**-Install openjdk on ubuntu: which include both JRE and JDK by running the commands :**

Sudo apt update

Sudo apt install openjdk-8-jdk -y

**-Then verify the current java version by :**

Java -version ; javac -version

**Setup a non root -user for hadoop environment by:**

Sudo apt install openssh-server openssh-client -y

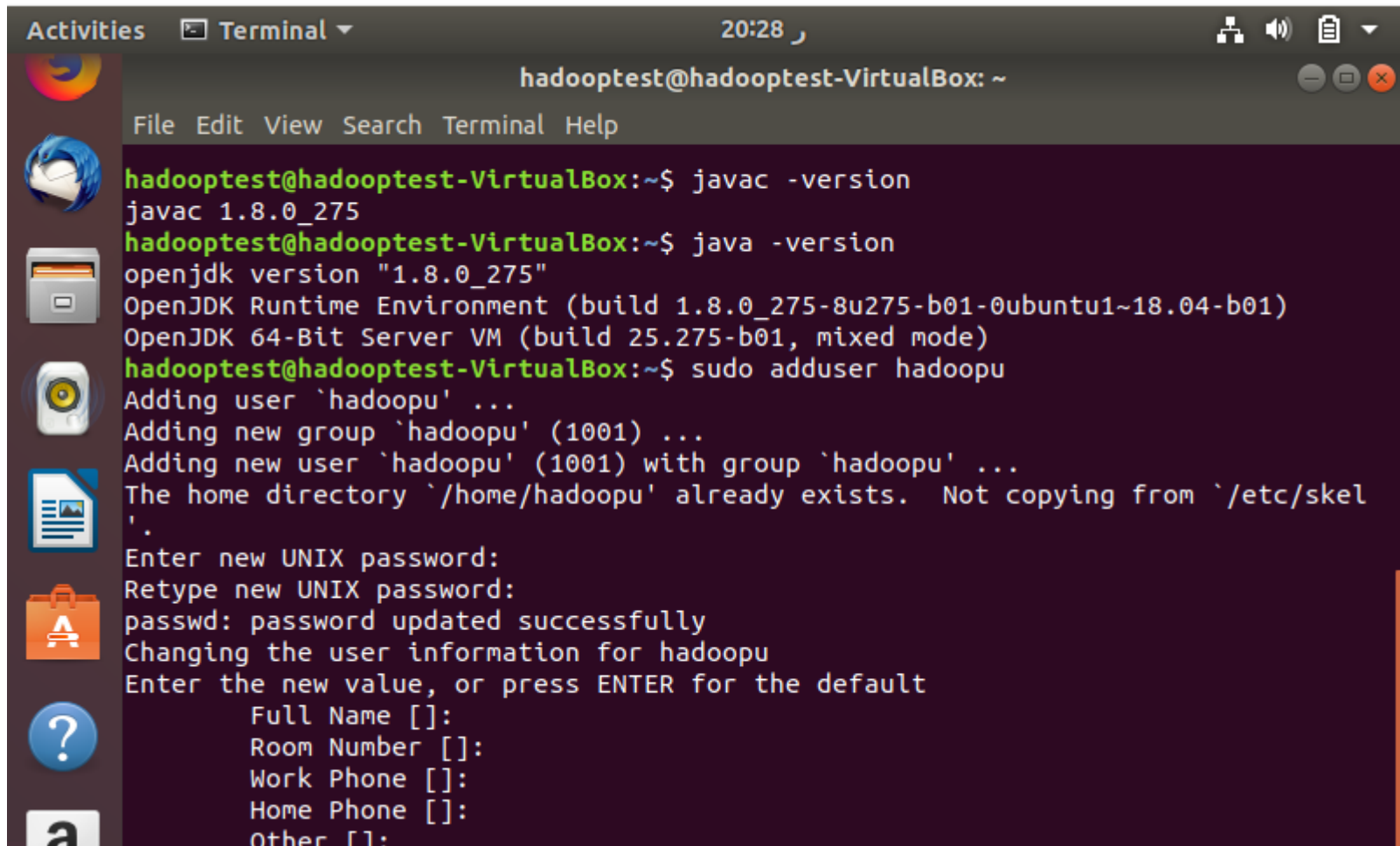
Sudo adduser hadoopu

Su - hadoopu

**- generate an ssh key pairs and define the location to store it in:**

ssh-keygen

# APACHE HADOOP INSTALLATION



The image shows a terminal window titled "hadooptest@hadooptest-VirtualBox: ~" with a menu bar (File, Edit, View, Search, Terminal, Help) and a sidebar with application icons. The terminal output shows the following commands and their results:

```
hadooptest@hadooptest-VirtualBox:~$ javac -version
javac 1.8.0_275
hadooptest@hadooptest-VirtualBox:~$ java -version
openjdk version "1.8.0_275"
OpenJDK Runtime Environment (build 1.8.0_275-8u275-b01-0ubuntu1~18.04-b01)
OpenJDK 64-Bit Server VM (build 25.275-b01, mixed mode)
hadooptest@hadooptest-VirtualBox:~$ sudo adduser hadoopu
Adding user `hadoopu' ...
Adding new group `hadoopu' (1001) ...
Adding new user `hadoopu' (1001) with group `hadoopu' ...
The home directory `/home/hadoopu' already exists. Not copying from `/etc/skel'.
Enter new UNIX password:
Retype new UNIX password:
passwd: password updated successfully
Changing the user information for hadoopu
Enter the new value, or press ENTER for the default
  Full Name []:
  Room Number []:
  Work Phone []:
  Home Phone []:
  Other []:
```

# APACHE HADOOP INSTALLATION

**-Use the cat command to store the public key as an authorized keys in ssh directory by :**

```
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
```

**-Use the chmod command to set the permission for hadoopu user:**

```
chmod 0600 ~/.ssh/authorized_keys
```

The new user is now able to SSH without needing to enter a password every time.

**Verify everything is set up correctly by using the hadoopu user to SSH to localhost :**

```
Sudo ssh localhost
```

# APACHE HADOOP INSTALLATION

## Download and install hadoop on Ubuntu :

In Firefox browser browse to apache hadoop download site

Click on hadoop version 3.2.1 binary download ,Copy the mirror link that allows you to download the hadoop tar package.

- **Use the command :**

wget <https://downloads.apache.org/hadoop/common/hadoop-3.2.1/hadoop-3.2.1.tar.gz> to download the package on Ubuntu .

-**Once the package is downloaded extract the files to initiate hadoop installation using the commands:**

Tar xzf hadoop-3.2.1.tar.gz

The Hadoop binary files are now located within the *hadoop -3.2.1* directory.

# APACHE HADOOP INSTALLATION

Thanks