


| | | |
|--|---------------------------------|---|
| Matière : Big Data et architectures associées Enseignante responsable : Leila Baccour Enseignantes de TP : Hana MALLEK, Maysam Chaari, Samar Akremi | Section : D-LSI ADBBD | AU : 2023-2024  |
|--|---------------------------------|---|

TP1 : Mise en place l'environnement de travail

I. Objectifs :

L'objectif principal de ce TP est de familiariser les étudiants avec le Framework Apache Hadoop et de les guider dans la mise en place d'un environnement de développement adapté. Pour ce faire, nous allons suivre les étapes suivantes :

- **Découverte de la notion du Big Data et l'importance du Framework Apache Hadoop :** Nous commencerons par explorer les concepts fondamentaux d'Apache Hadoop, ses composants clés, ainsi que ses applications dans le domaine du traitement de données distribuées.
- **Préparation de la machine virtuelle :** Nous aborderons la préparation d'une machine virtuelle compatible avec le système d'exploitation Linux. Cette étape inclura **la mise à jour du système**, l'installation du kit de développement Java (**JDK**), et la configuration **d'une clé SSH**. Ces opérations sont cruciales pour garantir un environnement stable et performant pour l'installation de Hadoop.

II. Définitions

Une machine virtuelle : un fichier informatique, couramment désigné sous le nom d'image, qui simule le comportement d'un véritable ordinateur. Cette machine virtuelle consomme une partie des ressources de l'ordinateur sur lequel elle est déployée.

Un hyperviseur : une couche logicielle installée sur un ordinateur pour superviser et contrôler le fonctionnement des machines virtuelles.

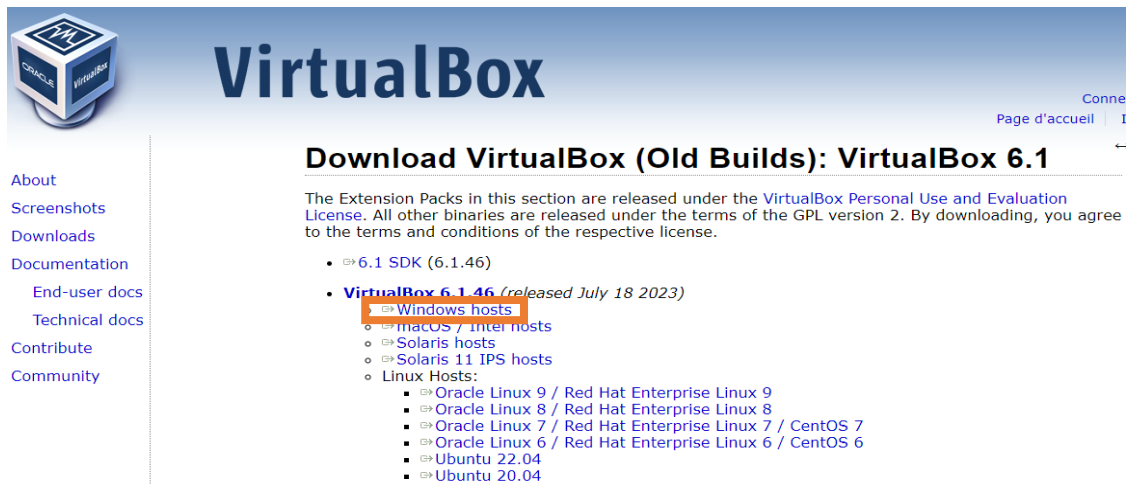
III. Recommandation

Afin d'assurer le bon fonctionnement de notre environnement de travail il nous faut avoir les caractéristiques suivantes :

- ✓ Système d'exploitation 64-bit
- ✓ RAM avec 8 GO ou plus
- ✓ 20 GB espace de disque disponible ou plus
- ✓ Intel VT-x or AMD-V virtualisation supporté (enabled) sinon il faut le rendre (enabled) en accédant au BIOS.

IV. Mise en place de l'environnement de travail

a) Télécharger VirtualBox 6.1 en cliquant sur "Windows hosts" à travers ce lien : https://www.virtualbox.org/wiki/Download_Old_Builds_6_1



VirtualBox

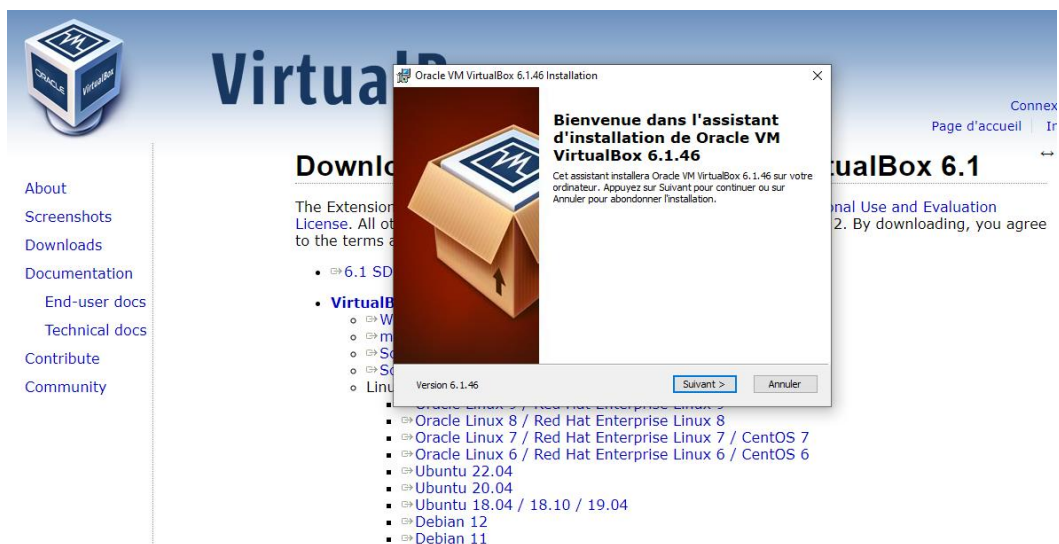
Page d'accueil | I

Download VirtualBox (Old Builds): VirtualBox 6.1

The Extension Packs in this section are released under the [VirtualBox Personal Use and Evaluation License](#). All other binaries are released under the terms of the GPL version 2. By downloading, you agree to the terms and conditions of the respective license.

- 6.1 SDK (6.1.46)
- VirtualBox 6.1.46** (released July 18 2023)
 - Windows hosts**
 - macOS / Intel hosts
 - Solaris hosts
 - Solaris 11 IPS hosts
 - Linux Hosts:
 - Oracle Linux 9 / Red Hat Enterprise Linux 9
 - Oracle Linux 8 / Red Hat Enterprise Linux 8
 - Oracle Linux 7 / Red Hat Enterprise Linux 7 / CentOS 7
 - Oracle Linux 6 / Red Hat Enterprise Linux 6 / CentOS 6
 - Ubuntu 22.04
 - Ubuntu 20.04

b) Suivre les étapes d'installation de VirtualBox.



VirtualBox

Page d'accueil | I

Download VirtualBox (Old Builds): VirtualBox 6.1

The Extension Packs in this section are released under the [VirtualBox Personal Use and Evaluation License](#). All other binaries are released under the terms of the GPL version 2. By downloading, you agree to the terms and conditions of the respective license.

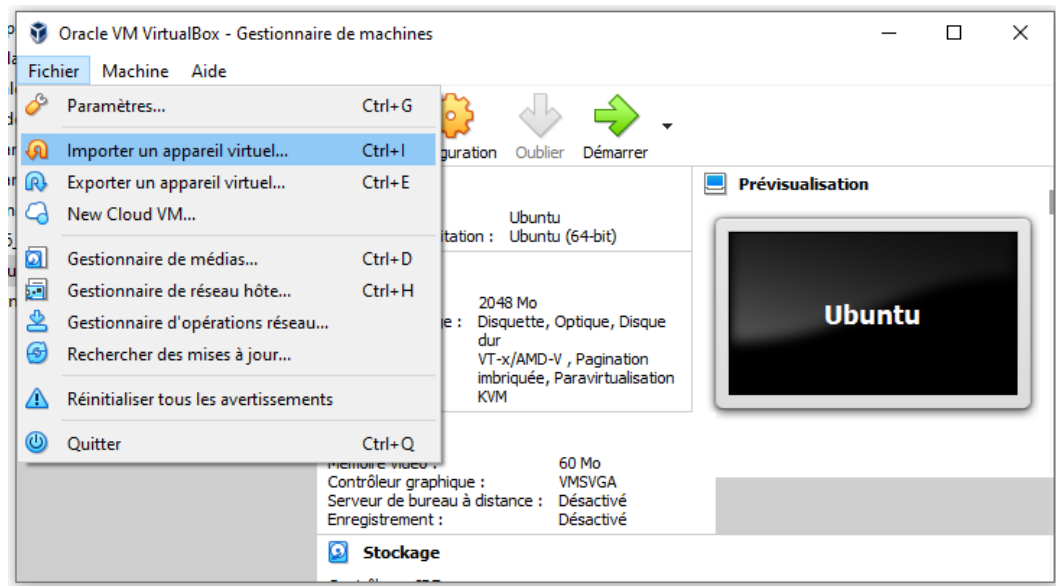
- 6.1 SDK (6.1.46)
- VirtualBox 6.1.46** (released July 18 2023)
 - Windows hosts**
 - macOS / Intel hosts
 - Solaris hosts
 - Solaris 11 IPS hosts
 - Linux Hosts:
 - Oracle Linux 9 / Red Hat Enterprise Linux 9
 - Oracle Linux 8 / Red Hat Enterprise Linux 8
 - Oracle Linux 7 / Red Hat Enterprise Linux 7 / CentOS 7
 - Oracle Linux 6 / Red Hat Enterprise Linux 6 / CentOS 6
 - Ubuntu 22.04
 - Ubuntu 20.04

c) Télécharger le fichier au format compressé OVA (ubuntu.ova) qui est disponible dans le Classroom.

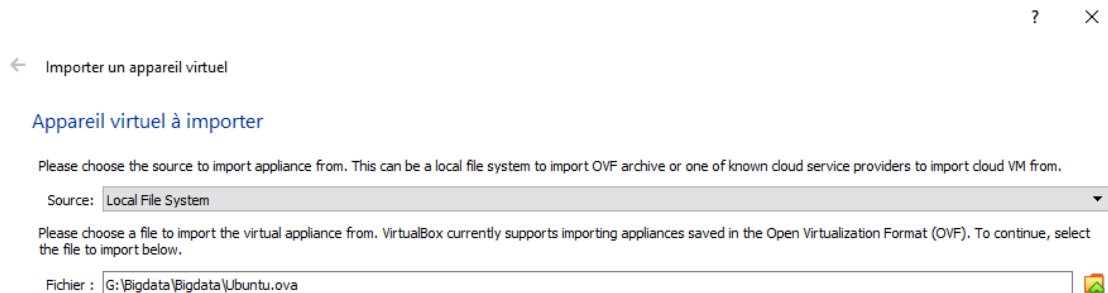
Le fichier OVA encapsule une machine virtuelle préconfiguré et qui comporte des applications pré-téléchargé qui sont nécessaires pour les travaux ultérieurs.

La machine virtuelle ubuntu.ova prend en charge le système d'exploitation Linux. Dans le dossier "Téléchargements", vous trouverez Apache Hadoop-2.10.2 et Apache-Hive 2.1.0, que nous utiliserons lors des prochaines séances.

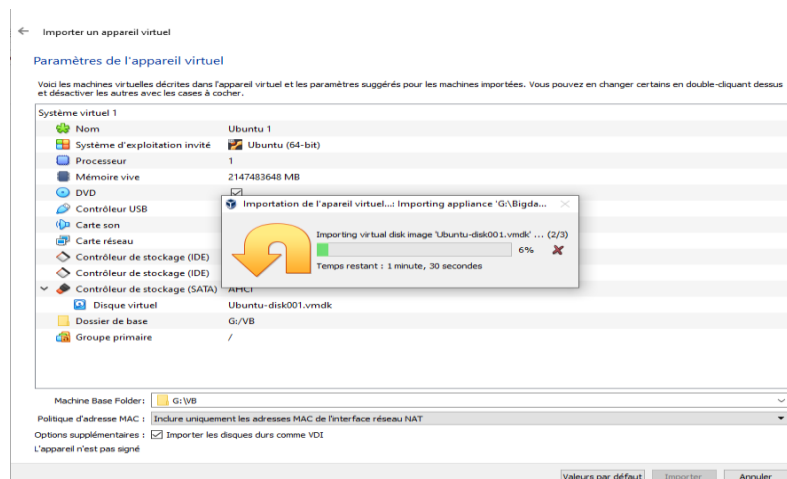
d) Dérouler le menu Fichier puis cliquer sur Importer un appareil virtuel (une machine virtuelle).



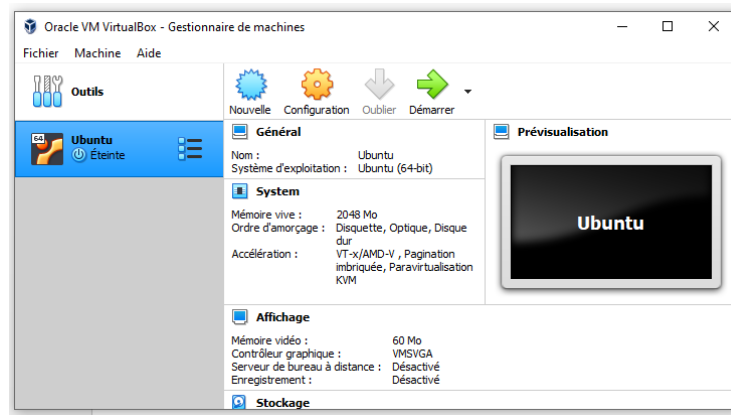
- e) Parcourir les fichiers pour importer l'appareil virtuel « ubuntu.ova », qui a déjà été téléchargé dans la question précédente, puis cliquer sur suivant.



- f) Changer l'emplacement d'installation de votre machine virtuelle « D:\MV », puis cliquez sur « importer »



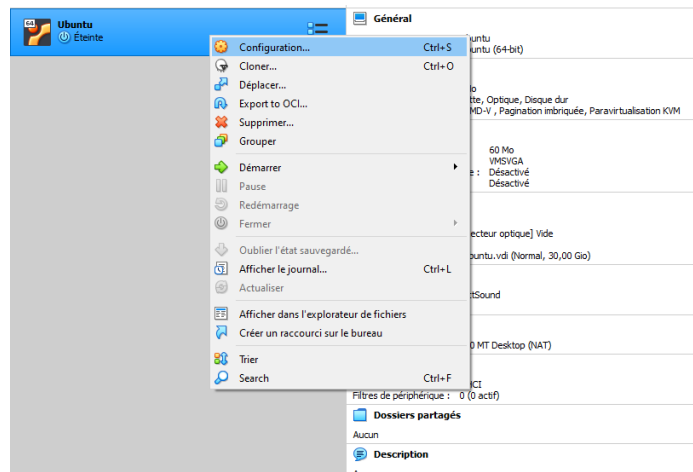
- g) Attendre un peu et voilà la machine virtuel « Ubuntu » est bien importée.



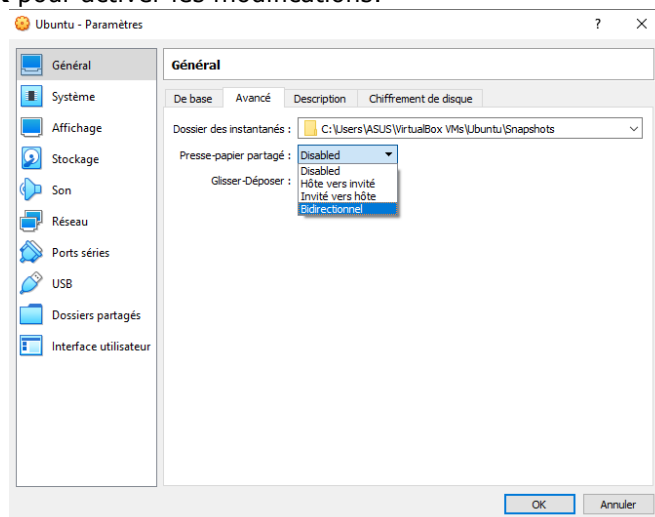
- Cliquer sur démarrer puis taper le mot de passe « hadoop » pour ouvrir le système

Remarque : Il est essentiel de mémoriser le mot de passe 'hadoop', car vous en aurez besoin pour mener à bien les procédures d'installation ultérieures en utilisant la commande 'sudo'.

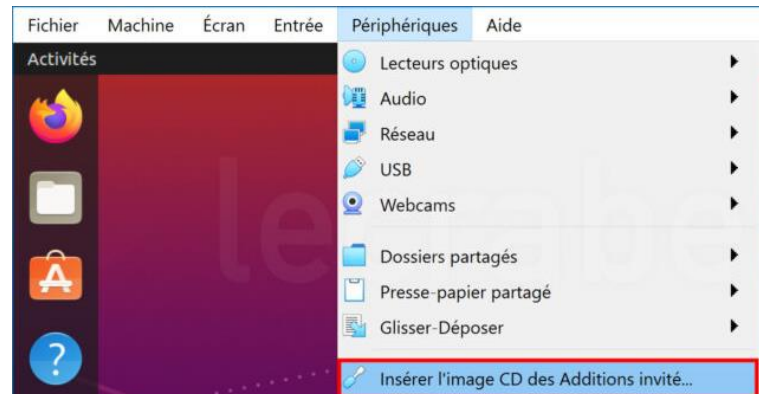
- Activer le copier/coller entre une machine virtuelle et son hôte tout en suivant les étapes suivantes :
 - Cliquer avec le bouton droit de la souris sur la machine virtuelle et cliquer sur **Configuration**



- Dans la rubrique **Général**, ouvrir l'onglet **Avancé**.
- Dérouler la liste **Presse-papier partagé** et sélectionner l'option **Bidirectionnel**.
- Cliquer sur **OK** pour activer les modifications.



- Dérouler le menu "Périphériques" et sélectionner "Insérer l'image CD des Additions Invité". Afin d'ajouter la fonctionnalité du presse-papiers partager (copier/coller) dans la machine virtuelle.



- Suivre les instructions pour terminer l'installation.
- **Redémarrer** la machine virtuelle pour appliquer les modifications.

➔ Vous pourrez ainsi copier/coller des éléments de votre ordinateur à la machine virtuelle et vice versa.

V. Installation de Java sur le système d'exploitation Linux

a. Mis à jour les fichiers de dépôt

Ouvrir le terminal et Mettre à jour les fichiers de dépôt en tapant la commande :

sudo apt-get update

- **apt-get install** : est une commande qui permet l'installation de paquets.
- **apt-get update** : l'option « update » met à jour la liste des fichiers disponibles dans les dépôts APT présents dans le fichier de configuration /etc/apt/sources.list.

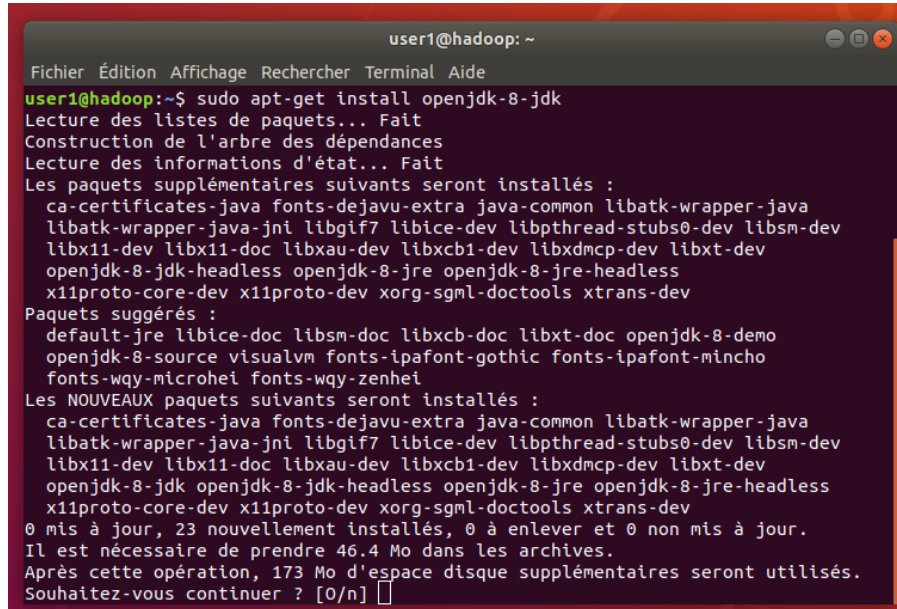
```

user1@hadoop: ~
Fichier Édition Affichage Rechercher Terminal Aide
To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.

user1@hadoop:~$ sudo apt-get update
[sudo] Mot de passe de user1 :
Atteint :1 http://security.ubuntu.com/ubuntu bionic-security InRelease
Atteint :2 http://tn.archive.ubuntu.com/ubuntu bionic InRelease
Atteint :3 http://tn.archive.ubuntu.com/ubuntu bionic-updates InRelease
Atteint :4 http://tn.archive.ubuntu.com/ubuntu bionic-backports InRelease
Lecture des listes de paquets... 38%
  
```

b. Installer java

- Taper la commande: **sudo apt-get install openjdk-8-jdk**
 - ✚ La commande **sudo** vous nécessite d'écrire le mot de passe (hadoop dans notre cas)
- Taper O (pour dire oui et continuer l'installation)



```
user1@hadoop: ~  
Fichier Édition Affichage Rechercher Terminal Aide  
user1@hadoop:~$ sudo apt-get install openjdk-8-jdk  
Lecture des listes de paquets... Fait  
Construction de l'arbre des dépendances  
Lecture des informations d'état... Fait  
Les paquets supplémentaires suivants seront installés :  
  ca-certificates-java fonts-dejavu-extra java-common libatk-wrapper-java  
  libatk-wrapper-java-jni libgif7 libice-dev libpthread-stubs0-dev libsm-dev  
  libx11-dev libx11-doc libxau-dev libxcb1-dev libxdmcp-dev libxt-dev  
  openjdk-8-jdk-headless openjdk-8-jre openjdk-8-jre-headless  
  x11proto-core-dev x11proto-dev xorg-sgml-doctools xtrans-dev  
Paquets suggérés :  
  default-jre libice-doc libsm-doc libxcb-doc libxt-doc openjdk-8-demo  
  openjdk-8-source visualvm fonts-ipafont-gothic fonts-ipafont-mincho  
  fonts-wqy-microhei fonts-wqy-zenhei  
Les NOUVEAUX paquets suivants seront installés :  
  ca-certificates-java fonts-dejavu-extra java-common libatk-wrapper-java  
  libatk-wrapper-java-jni libgif7 libice-dev libpthread-stubs0-dev libsm-dev  
  libx11-dev libx11-doc libxau-dev libxcb1-dev libxdmcp-dev libxt-dev  
  openjdk-8-jdk openjdk-8-jdk-headless openjdk-8-jre openjdk-8-jre-headless  
  x11proto-core-dev x11proto-dev xorg-sgml-doctools xtrans-dev  
0 mis à jour, 23 nouvellement installés, 0 à enlever et 0 non mis à jour.  
Il est nécessaire de prendre 46.4 Mo dans les archives.  
Après cette opération, 173 Mo d'espace disque supplémentaires seront utilisés.  
Souhaitez-vous continuer ? [O/n] ☐
```

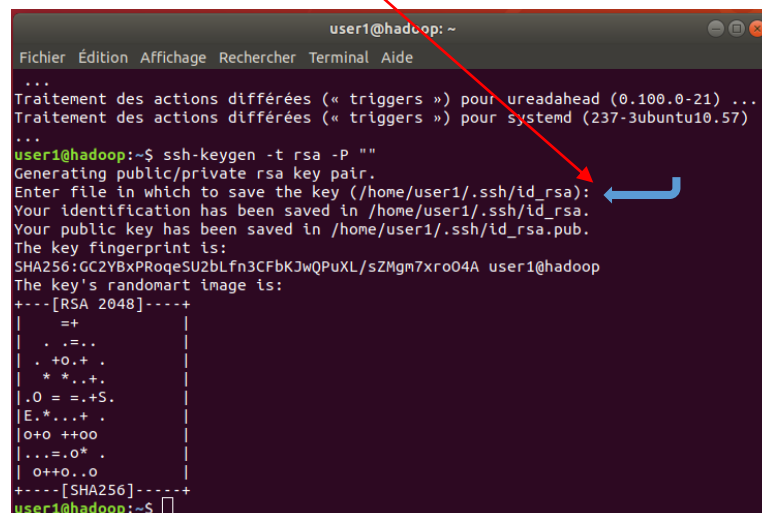
- Vérifier l'installation de java en tapant la commande : **java -version**

c. Installer ssh

- Taper la commande: **sudo apt-get install ssh**

d. Générer une clé de cryptage

- Taper la commande : **ssh-keygen -t rsa -P ""**
 - RSA désigne un algorithme de cryptographie asymétrique, utilisé pour assurer l'échange des données confidentielles sur Internet.
 - -P "" signifie sans mot de passe.
- Puis taper sur le bouton **entrer** pour générer une clé cryptée



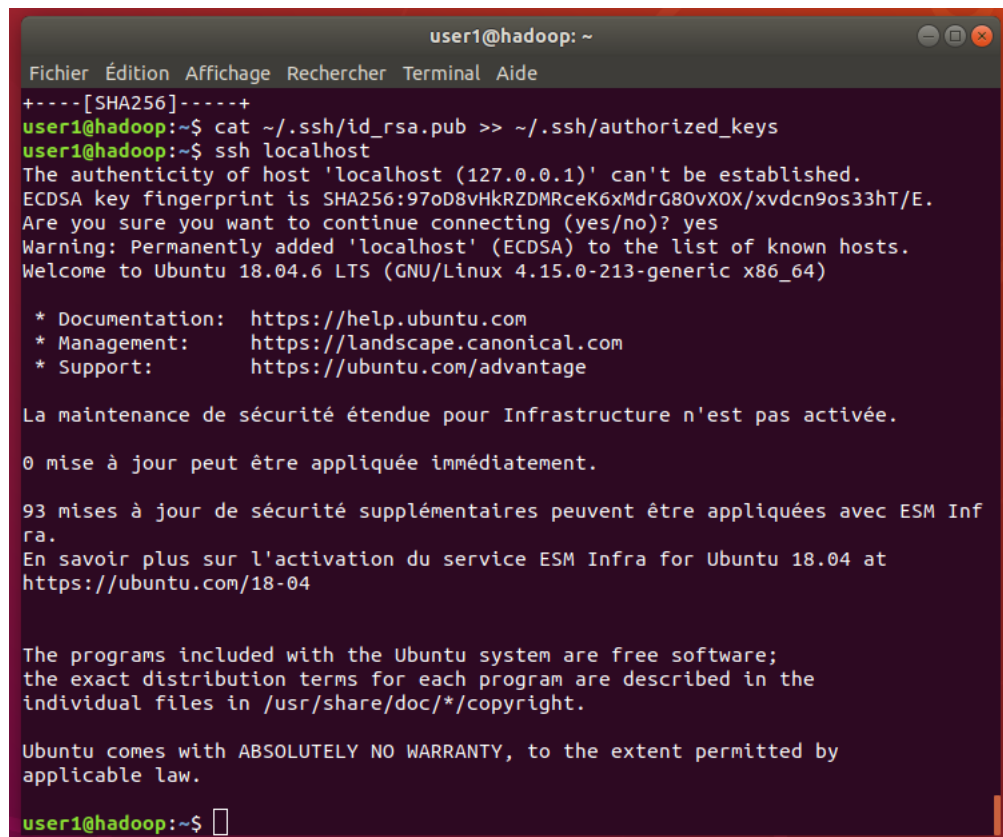
```
user1@hadoop: ~  
Fichier Édition Affichage Rechercher Terminal Aide  
...  
Traitement des actions différées (« triggers ») pour ureadahead (0.100.0-21) ...  
Traitement des actions différées (« triggers ») pour systemd (237-3ubuntu10.57)  
...  
user1@hadoop:~$ ssh-keygen -t rsa -P ""  
Generating public/private rsa key pair.  
Enter file in which to save the key (/home/user1/.ssh/id_rsa):   
Your identification has been saved in /home/user1/.ssh/id_rsa.  
Your public key has been saved in /home/user1/.ssh/id_rsa.pub.  
The key fingerprint is:  
SHA256:GC2YBxPROqeSU2bLfn3CFbKJwQPuXL/sZMgm7xro04A user1@hadoop  
The key's randomart image is:  
+---[RSA 2048]---+  
|  =+  |  
|  .+.  |  
|  +O.+  |  
|  * * .+  |  
|.O = .+.S. |  
|E.*...+  |  
|o+o ++oo  |  
|...=.O*  |  
| o+o..o  |  
+---[SHA256]-----+  
user1@hadoop:~$
```

- Copier la clé dans « authorized_keys » à travers la commande suivante

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

e. Vérifier l'installation

- Taper la commande : **ssh localhost**



```
user1@hadoop: ~
Fichier  Édition  Affichage  Rechercher  Terminal  Aide
+----[SHA256]-----+
user1@hadoop:~$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
user1@hadoop:~$ ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ECDSA key fingerprint is SHA256:97oD8vHkRZDMRceK6xMdrG80vX0X/xvdcn9os33hT/E.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.
Welcome to Ubuntu 18.04.6 LTS (GNU/Linux 4.15.0-213-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

La maintenance de sécurité étendue pour Infrastructure n'est pas activée.
0 mise à jour peut être appliquée immédiatement.

93 mises à jour de sécurité supplémentaires peuvent être appliquées avec ESM Infra.
En savoir plus sur l'activation du service ESM Infra for Ubuntu 18.04 at
https://ubuntu.com/18-04

The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

user1@hadoop:~$
```