



UASB03 – Projet final de la certification
Analyste de données massives
Installation de la plateforme de
développement

Daniel PONT
décembre 2019

Remerciements

Je tiens ici à exprimer ma vive gratitude à l'équipe pédagogique de STA211, NFE204 et RCP216 pour toutes les indications et les conseils fournis au cours des différentes UE qui m'ont permis de mettre en œuvre la plateforme de développement du projet UASB03.

*“You can pipe anything to anything else, and usually it'll do something.
With most of the standard Linux tools, it'll even do what you expect.”*

– Scott Simpson

Sommaire

1. Introduction.....	1
2. Installation d' Ubuntu.....	2
2.1 Récupération de la dernière version stable d'Ubuntu (19.10).....	2
2.2 Installation d'Ubuntu dans Virtual Box (6.1) sous Windows.....	2
2.3 Installation d'Ubuntu sur un disque SSD externe ou une clé USB.....	13
2.3.1 Création d'une clé USB Ubuntu live bootable.....	13
2.3.2 Exécution d'Ubuntu en mode live depuis la clé USB bootable.....	15
2.3.3 Installation d'Ubuntu sur le disque SSD externe (ou sur une 2ème clé USB).....	16
2.4 Installation des gestionnaires de paquets python pip 2 et 3.....	23
2.5 Installation d'un éditeur de texte polyvalent : Atom (1.42).....	24
2.6 Installation d'un navigateur web sécurisé (Brave).....	25
2.7 Configuration de l'environnement graphique (Gnome).....	25
3. Installation d'Anaconda3-2019.10.....	25
4. Installation de R (3.6.1) et RStudio (1.2).....	27
4.1 Installation de R (3.6.1).....	27
4.2 Installation de RStudio(1.2).....	27
5. Installation de Java 8 (JDK/JRE Oracle).....	28
5.1 Téléchargement de Java.....	28
5.2 Installation de Java.....	28
6. Installation de Scala (2.12.10).....	29
7. Installation de Hadoop (3.2.1).....	29
8. Installation de Spark (2.4.4).....	32
9. Installation de Cassandra (3.11.5).....	33
10. Installation d'Elasticsearch (7.5).....	34
11. Installation de Kibana(7.5).....	35

1. Introduction

Ce document décrit l'installation de la plateforme de développement mise en place pour l'UASB03. Il s'agit du projet final de la certification « Analyste de données massives » organisé par le CNAM Paris.

Cette plateforme de développement repose sur le système d'exploitation linux Ubuntu 19.10. Deux options d'installation d'Ubuntu sur un PC Windows sont détaillées dans ce document :

- En utilisant le logiciel de virtualisation Oracle Virtual Box 6.1 ;
- En installant Ubuntu sur un disque SSD externe ou sur une clé USB.

2. Installation d' Ubuntu

2.1 Récupération de la dernière version stable d'Ubuntu (19.10)

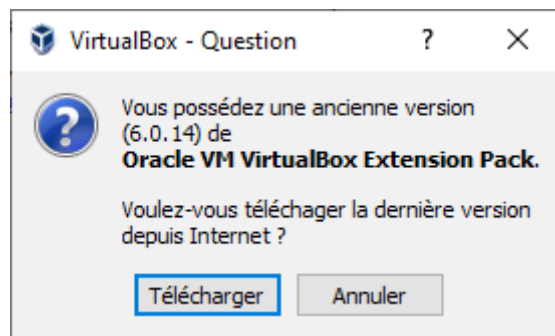
Le projet sera réalisé avec un PC utilisant le système d'exploitation [Ubuntu 19.10 \(Eoan Ermine\)](#) en version desktop.

2.2 Installation d'Ubuntu dans Virtual Box (6.1) sous Windows

La procédure est la suivante :

- [Télécharger](#) puis installer Virtual Box 6.1 avec les options par défaut.

NB : bien penser à télécharger le *Oracle VM VirtualBox Extension Pack* , en cliquant sur *Télécharger* quand la fenêtre suivante apparaît à la fin de l'installation :



Installer ensuite ce pack d'extension.

- Exécuter Virtual Box et cliquer sur le bouton « Nouvelle »



- Saisir le nom « Ubuntu 19.10 », sélectionner le type « Linux » et la version « Ubuntu (64 bits) », puis cliquer sur suivant :

? ×


← Crée une machine virtuelle

Nom et système d'exploitation

Veillez choisir un nom et un dossier pour la nouvelle machine virtuelle et sélectionner le type de système d'exploitation que vous envisagez d'y installer. Le nom que vous choisirez sera repris au travers de VirtualBox pour identifier cette machine.

Nom :

Dossier de la machine :

Type : 

Version :

Mode expert **Suivant >** Annuler

- Définir pour la mémoire vive une valeur de 8 Go (sur un PC avec 16 Go de RAM):

? ×

← Crée une machine virtuelle

Taille de la mémoire

Choisissez la quantité de mémoire vive en méga-octets alloués à la machine virtuelle.

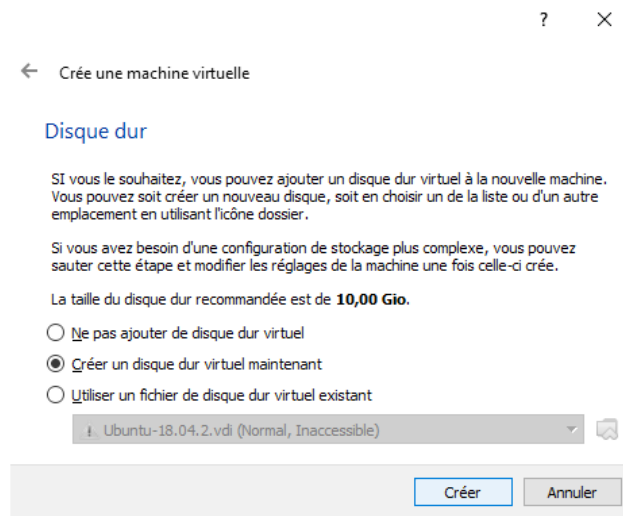
La quantité recommandée est de **1024 Mo**.

8192 Mio

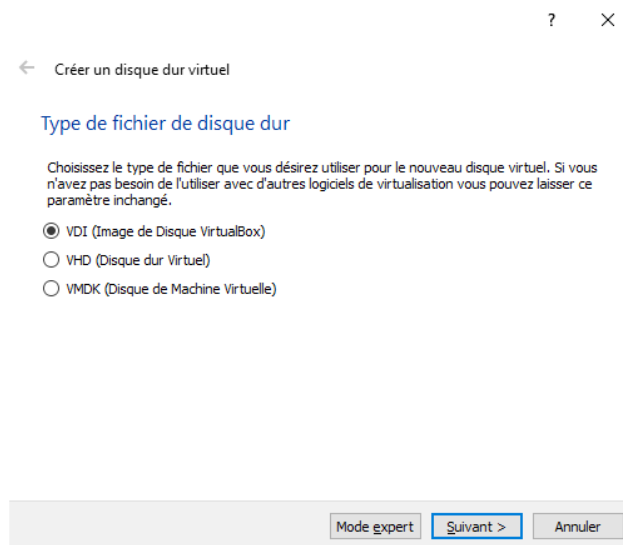
4 Mio 16384 Mio

Suivant > Annuler

- Sélectionner l'option « Créer un disque dur virtuel maintenant », puis cliquer sur « Créer » :



- Sélectionner l'option « VDI (Image Disque VirtualBox) », puis cliquer sur « Suivant » :



- Sélectionner « Taille Fixe » puis cliquer sur « Suivant » :

← Créer un disque dur virtuel

Stockage sur disque dur physique

Veillez choisir si le nouveau fichier de disque dur virtuel doit croître au fur et à mesure (allocation dynamique) ou bien s'il doit être créé à sa taille maximale (taille fixe).

Un fichier de disque dur **alloué dynamiquement** n'utilisera d'espace sur votre disque dur physique qu'au fur et à mesure qu'il se remplira (jusqu'à une **taille fixe maximale**), **cependant il ne se réduira pas automatiquement lorsque de l'espace sur celui-ci sera libéré.**

Un fichier de disque dur à **taille fixe** sera plus long à créer sur certains systèmes mais sera souvent plus rapide à utiliser.

☐ Dynamiquement alloué
☒ Taille fixe

Suivant > Annuler

NB : choisir une taille fixe plutôt qu'une taille dynamiquement allouée n'a pas grande importance. Le fait de définir une taille fixe permet simplement d'[ajouter des fichiers plus rapidement dans le disque dur virtuel](#). Si la taille initiale s'avère insuffisante, [il est possible de l'augmenter](#) avec la commande suivante :

```
VBoxManage modifyhd path_to_vdi_file.vdi --resize desired_size_in_megabytes
```

(VboxManage est un outil situé dans le répertoire *C:\Program Files\Oracle\VirtualBox*).

- Garder l'emplacement par défaut, sélectionner au moins 16Go, puis cliquer sur « Créer » :

← Créer un disque dur virtuel

Emplacement du fichier et taille

Veillez saisir un nom pour le nouveau fichier de disque dur virtuel dans la boîte ci-dessous ou cliquez sur l'icône dossier pour choisir un autre dossier dans lequel le créer.

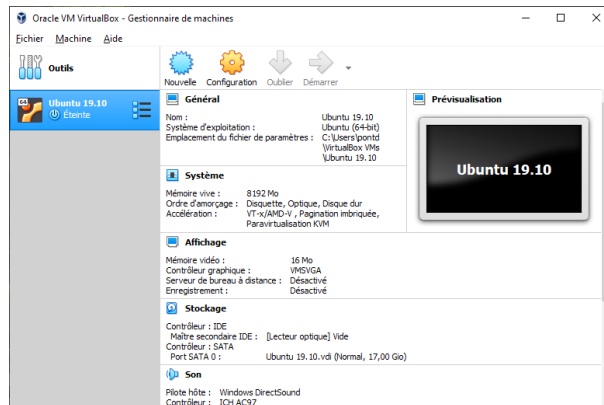
C:\Users\pontd\VirtualBox VMs\Ubuntu 19.10\Ubuntu 19.10

Choisissez la taille du disque dur virtuel en mégaoctets. Cette taille est la limite de la quantité de données de fichiers qu'une machine virtuelle sera capable de stocker sur le disque dur.

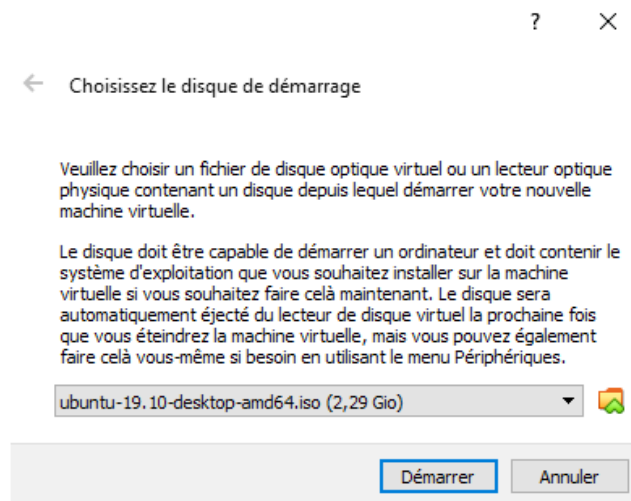
4,00 Mio 17,00 Gio 2,00 Tio

Créer Annuler

- Démarrer la VM (machine virtuelle) Ubuntu 19.10 en double cliquant sur l'entrée correspondante dans le menu de gauche :

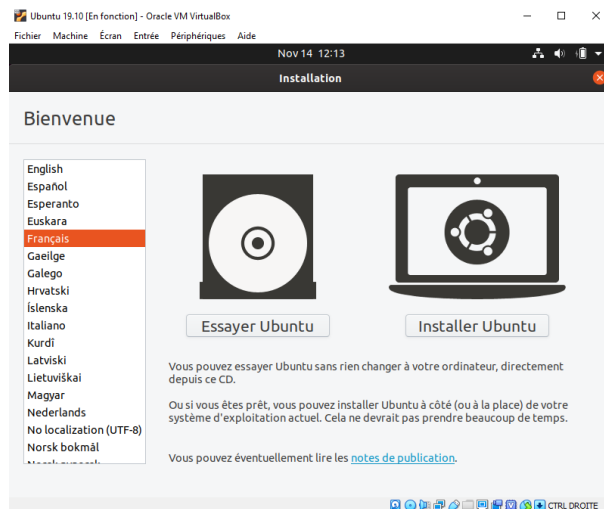


- Sélectionner l'emplacement de l'image Ubuntu(fichier .iso), puis cliquer sur « Démarrer » :

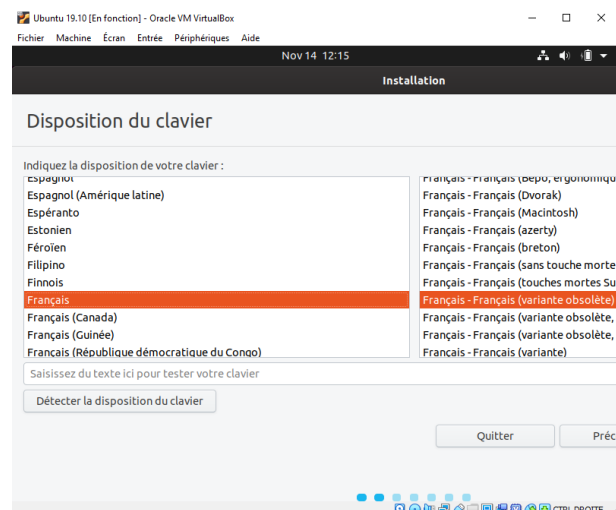


Au cours de l'installation d'Ubuntu, il est possible en fonction de la résolution graphique, que les écrans d'installation soit tronqués et que les bouton de validation situés en bas à droite ne soit pas accessibles par un clic souris. Dans ce cas utilisez la touche de tabulation pour déplacer le focus jusqu'au bouton de validation puis cliquez sur *ENTREE*.

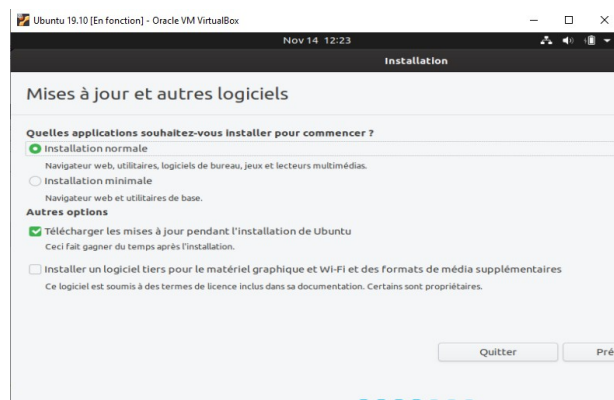
- Sélectionner la langue « Français » puis cliquer sur « Installer Ubuntu »



- Garder la disposition du clavier par défaut « Français-Français (variante obsolète) :



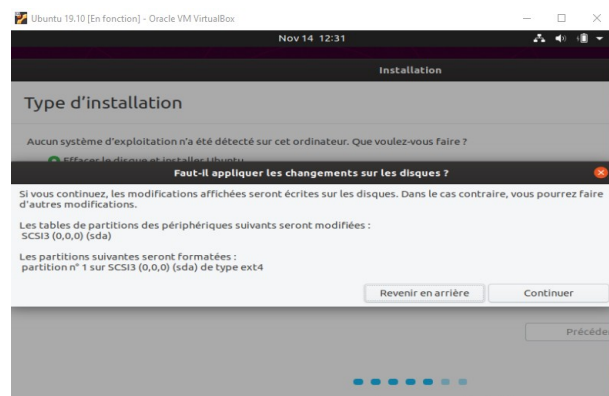
- Garder les options d'installation et de mise à jour par défaut (« Installation normale », « Télécharger les mises à jour pendant l'installation d'Ubuntu »)



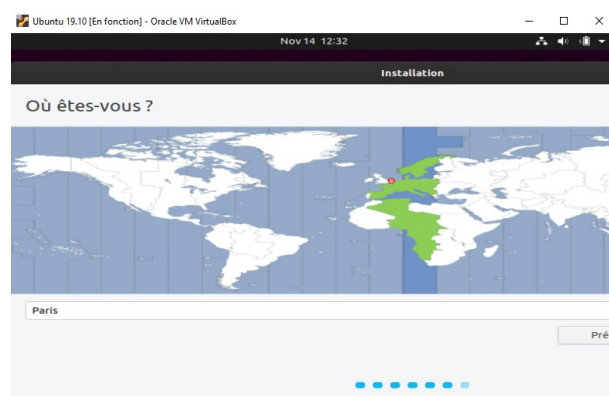
- Garder le type d'installation par défaut (« Effacer le disque et installer Ubuntu »)



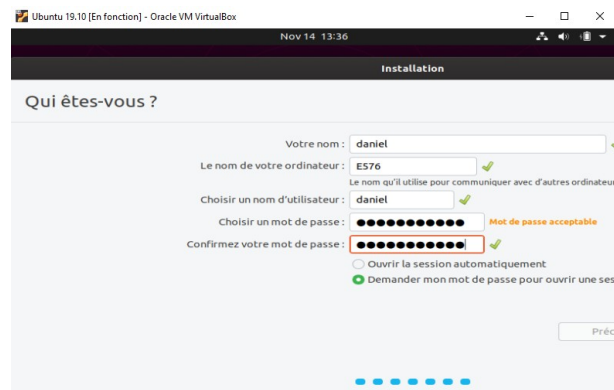
- Confirmer l'application des changements sur disque en cliquant sur « Continuer » :



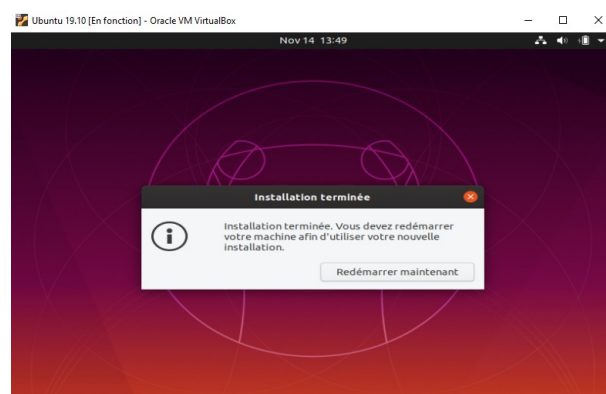
- Garder le fuseau horaire par défaut « Paris » :



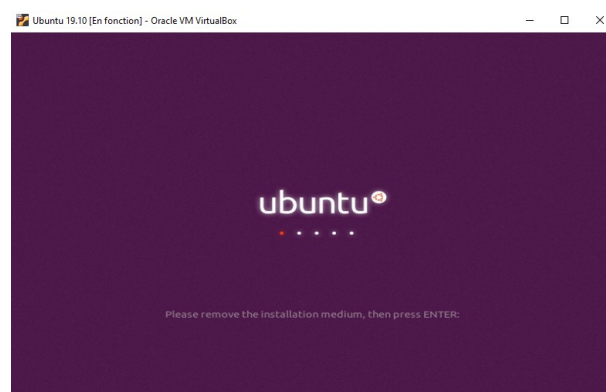
- Saisir le nom d'utilisateur, le nom de l'ordinateur, le mot de passe et garder l'option par défaut d'ouverture de session (demander le mot de passe):



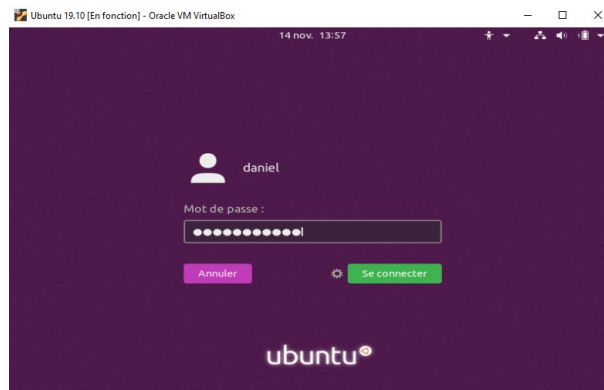
- En fin d'installation, cliquer sur le bouton « Redémarrer maintenant »:



- Appuyer sur « Entrée » :



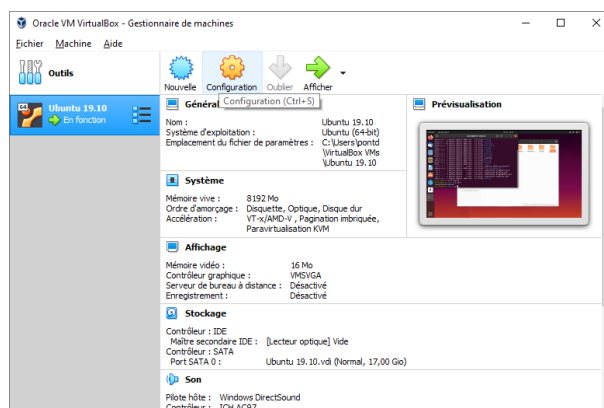
- Cliquer sur le nom d'utilisateur, saisir le mot passe et cliquer sur « Se connecter » :



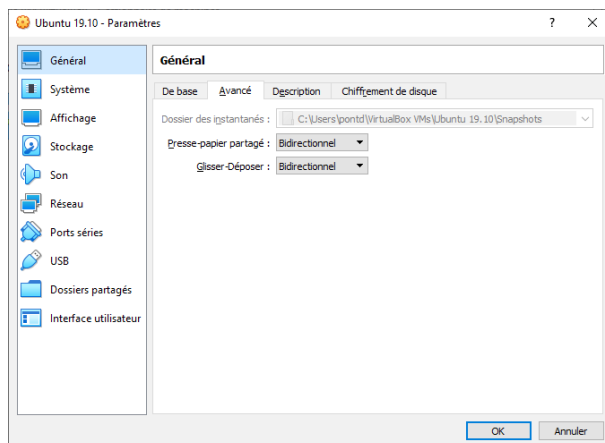
- Les écrans d'accueil à la première connexion permettent (de façon facultative) de paramétrer des services dans le cloud, d'envoyer des messages de traces en cas d'erreur Ubuntu et d'autoriser firefox à géolocaliser le PC.
- Pour entrer dans le mode plein écran presser **CTRLdroite** (bouton à droite de la barre d'espace) **F**.
- Installer ensuite « Virtual Box Guest Additions » afin de pouvoir copier-coller entre l'hôte windows et la VM ubuntu. Pour cela ouvrir une fenêtre de terminal et saisir :

```
# Installation des Guest Additions de Virtual Box
sudo apt install virtualbox-guest-x11 virtualbox-guest-utils virtualbox-guest-dkms
```

- Autoriser le copier-coller entre l'hôte Windows et la VM Ubuntu, cliquer sur le bouton « Configuration » :



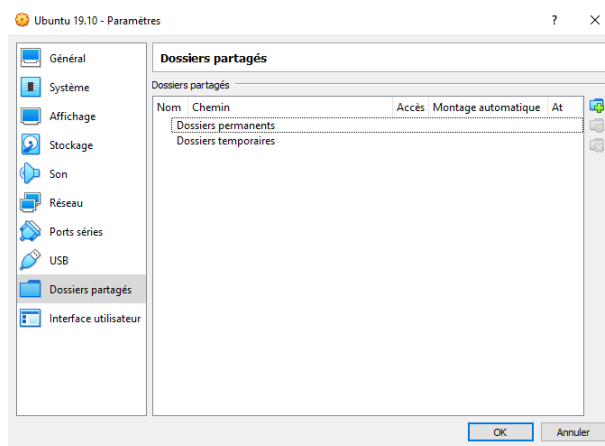
- Dans la fenêtre « Configuration » cliquer sur « Général » dans le menu de gauche, puis ouvrir l'onglet « Avancé » et sélectionner « Bidirectionnel » dans les listes déroulantes « Presse-papier partagé » et « Glisser-Déposer » :



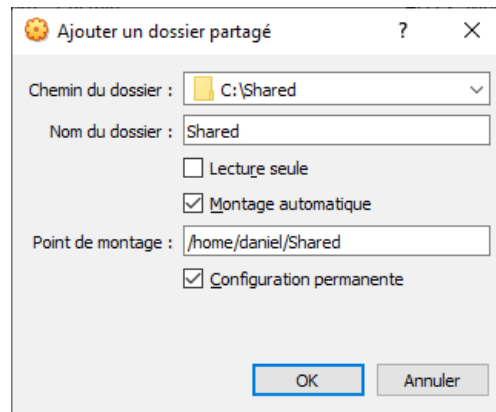
- Dans la VM Ubuntu, ouvrir une fenêtre de terminal puis saisir les commandes suivantes :

```
# Création du répertoire partagé et configuration des droits d'accès :
mkdir Shared
sudo adduser daniel vboxsf
```

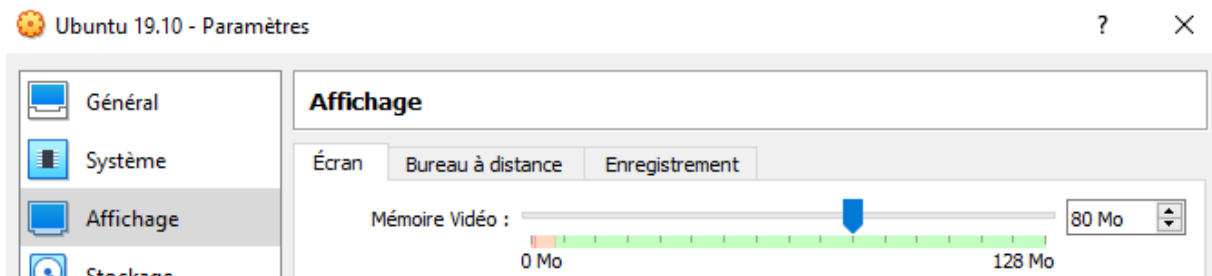
- Dans la fenêtre « Configuration » cliquer sur « Dossiers partagés » dans le menu de gauche puis sélectionner « Dossiers permanents » dans la partie droite de la fenêtre :



- Cliquer ensuite sur l'icône « + » en haut dans le menu latéral droit. Saisir ensuite le dossier Windows à partager, le point de montage dans la VM Ubuntu et cocher les cases « Montage automatique » et « Configuration permanente » :



- Pour finir, dans la fenêtre « Configuration » cliquer sur « Affichage » dans le menu de gauche et sélectionner au moins 80Mo de mémoire vidéo :



2.3 Installation d'Ubuntu sur un disque SSD externe ou une clé USB

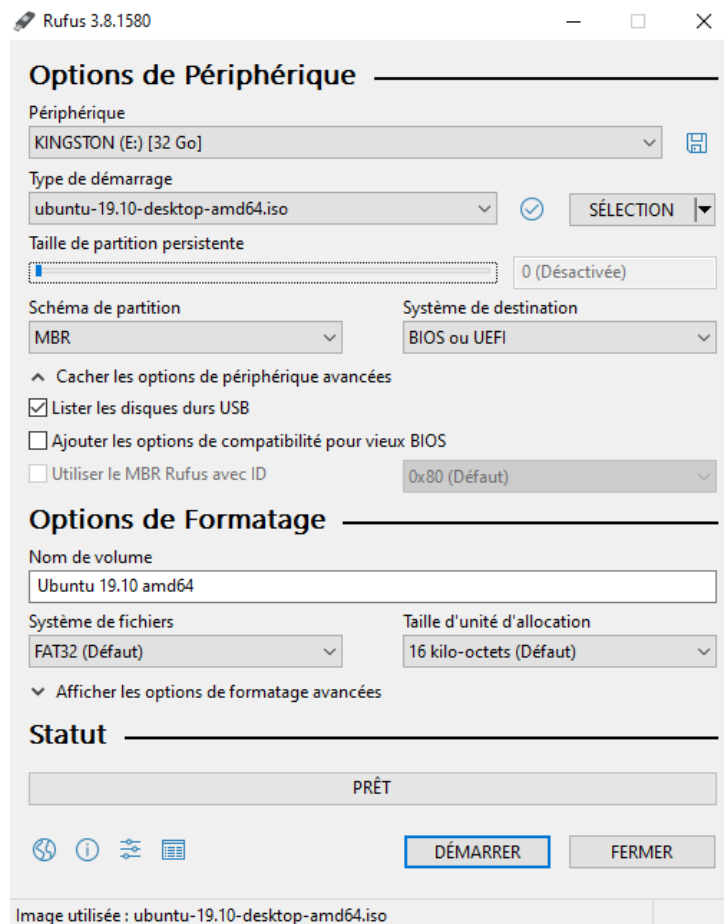
Cette procédure requiert :

- une clé USB sur lequel on va installer une version « live » d'Ubuntu ;
- une deuxième clé ou (mieux encore) un disque SSD externe sur lequel on va installer une version d'Ubuntu complète, persistante, à l'aide de la clé avec la version « live ».

2.3.1 Création d'une clé USB Ubuntu live bootable

Télécharger l'utilitaire [Rufus](https://github.com/pbatard/rufus/releases/download/v3.8/rufus-3.8.exe) : <https://github.com/pbatard/rufus/releases/download/v3.8/rufus-3.8.exe>

Exécuter Rufus :

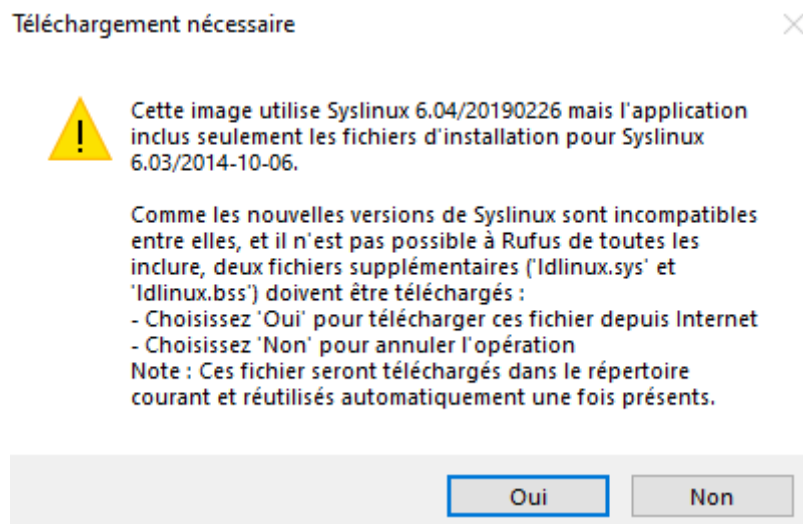


Sélectionner dans la liste *Périphérique*, la clé USB sur laquelle installer une version bootable « live » d'ubuntu.

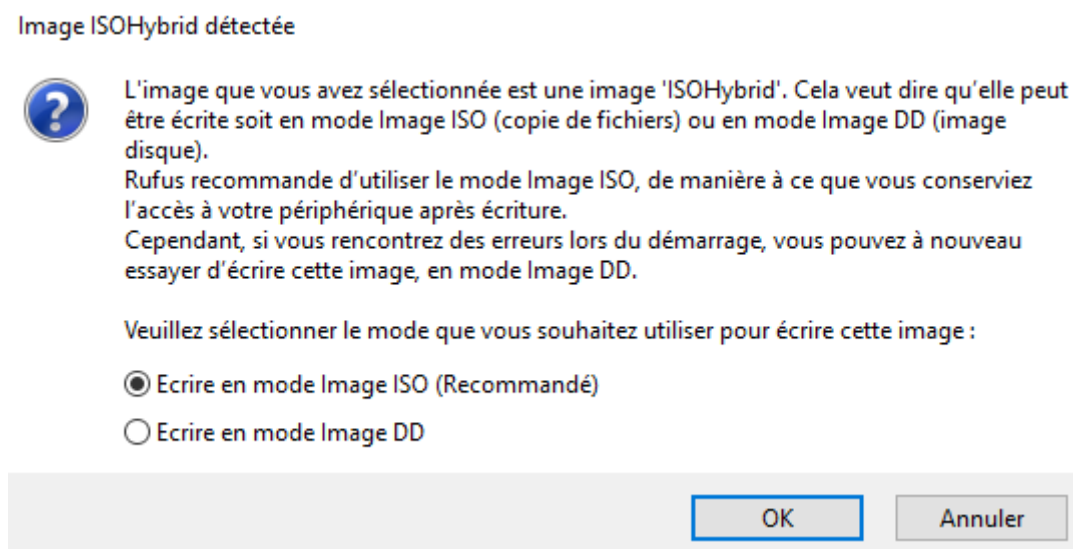
Cliquer ensuite sur le bouton *SELECTION* (associé à la liste « *Type de démarrage* ») pour sélectionner l'emplacement de l' image ISO d'ubuntu : *ubuntu-19.10-desktop-amd64.iso*

Cliquer enfin sur le bouton *DEMARRER*.

Le message d'avertissement suivant apparaît, cliquer sur *Oui* :

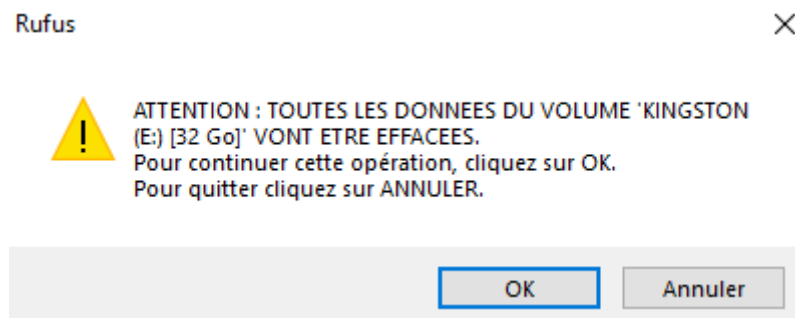


Un nouveau message d'avertissement s'affiche :



Garder l'option par défaut « *Ecrire en mode image ISO* » et cliquer sur *OK*.

Un dernier message d'avertissement apparaît :



Cliquer sur OK.

A la fin de l'installation de la clé, il faut mettre à jour manuellement les fichiers d'installation pour Syslinux :

1. Ouvrir dans l'explorateur windows le répertoire depuis lequel a été exécuté Rufus.exe
2. Ce répertoire doit contenir un sous-répertoire *rufus_files*
3. Ouvrir le répertoire *rufus_files/syslinux-6.04*
4. Copier les fichiers *ldlinux.bss* et *ldlinux.sys* à la racine de la clé USB Ubuntu bootable (en écrasant le fichier *ldlinux.sys* qui s'y trouve déjà)

2.3.2 Exécution d'Ubuntu en mode live depuis la clé USB bootable

La procédure est la suivante :

- Brancher la clé USB Ubuntu live bootable puis redémarrer le PC.
- Interrompre le processus normal de démarrage du PC et accéder aux écrans de configuration du Bios :

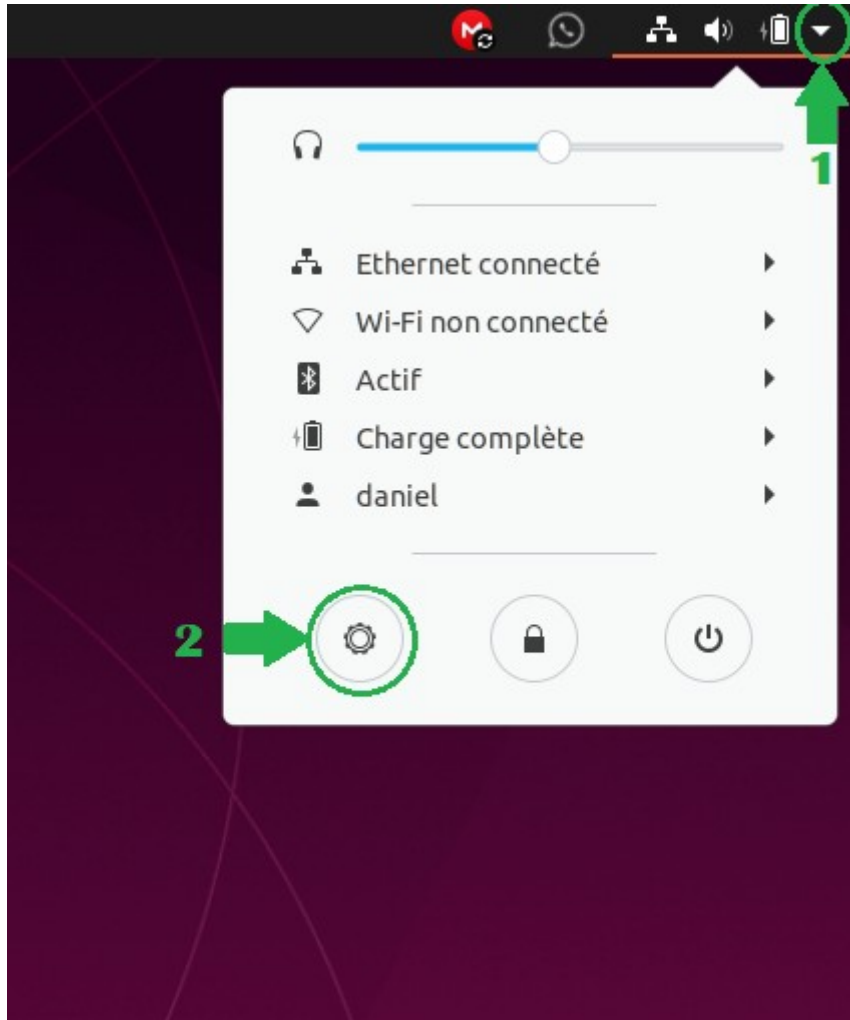
Sur les PC Acer, Asus, Dell ou Lenovo, maintenir appuyé la touche **F2** au démarrage du PC pour accéder aux écrans de configuration du Bios.

Sur les PC HP, appuyer sur la touche **F10**

- Modifier l'ordre de boot, pour que le PC boote en premier sur la clé USB Ubuntu Live plutôt que sur le disque dur interne.
- Le PC redémarre et affiche un menu proposant d'essayer ou d'installer Ubuntu. Choisir la première option : *Try Ubuntu without installing*

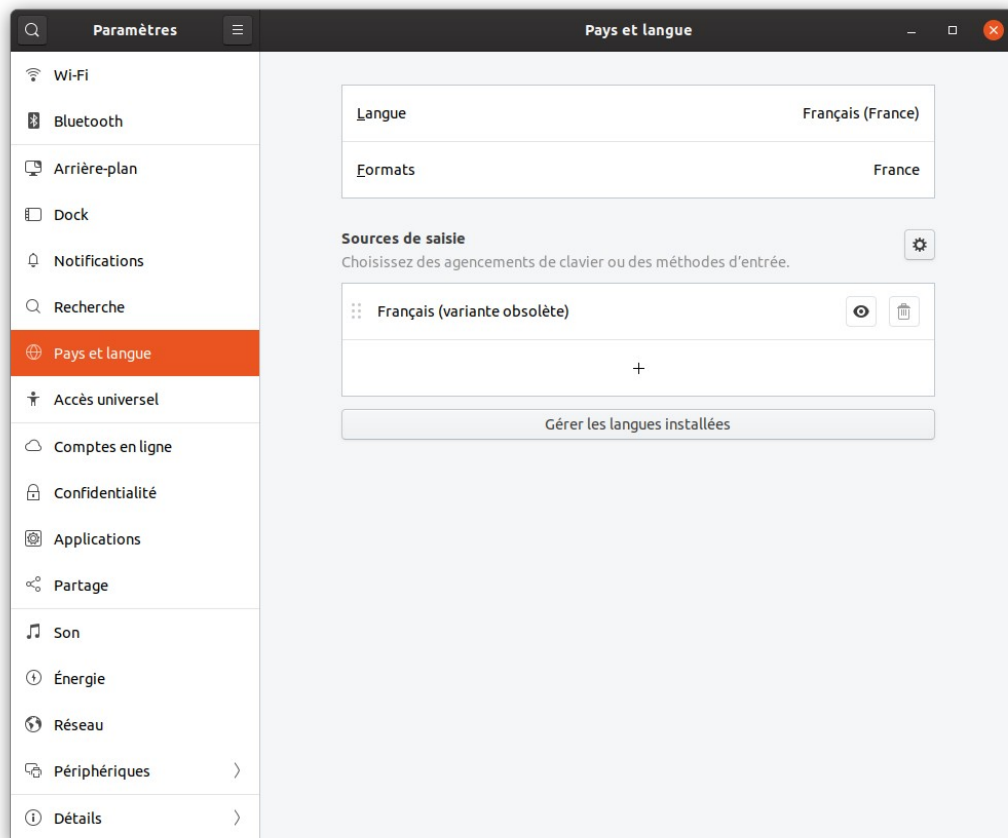
2.3.3 Installation d'Ubuntu sur le disque SSD externe (ou sur une 2ème clé USB)

Une fois démarrée la version live d'Ubuntu, la première chose à faire est de configurer le clavier (la configuration par défaut est *QWERTY*, il convient donc de passer en *AZERTY*) :

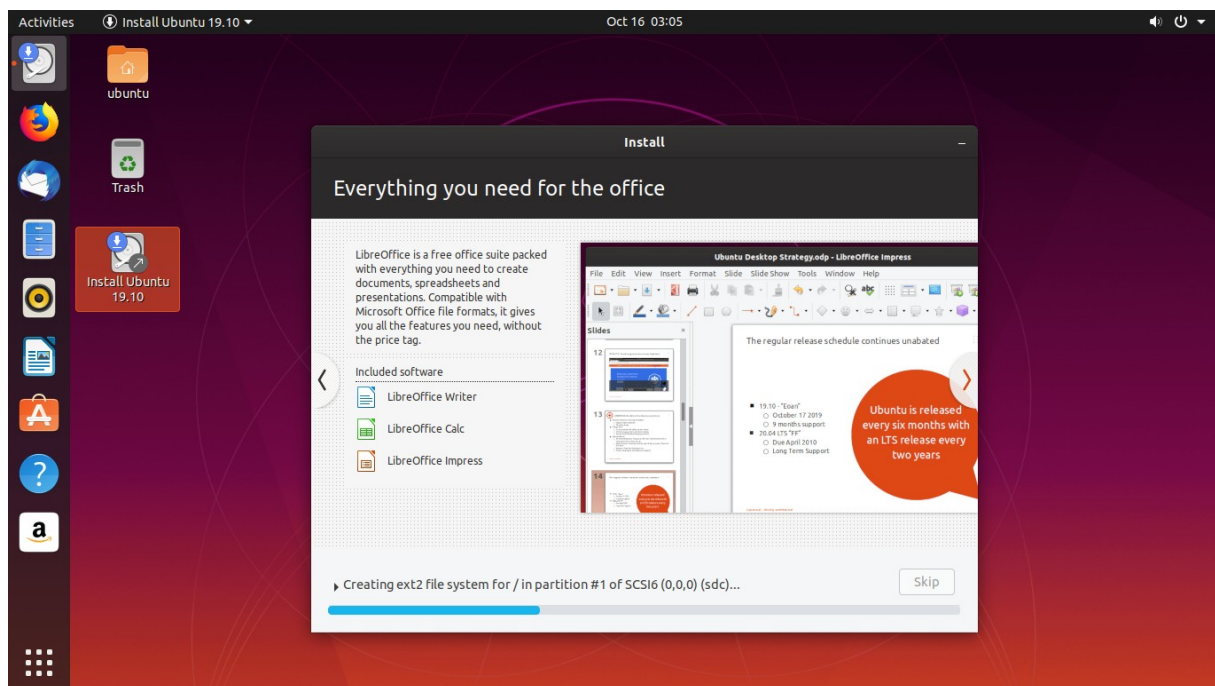


1. Cliquer sur le petit triangle dans le coin supérieur droit de l'écran
2. Dans le menu contextuel qui s'affiche, cliquer sur le bouton avec une icône représentant un rouage. Cela donne accès à l'écran de paramétrage général d'Ubuntu.

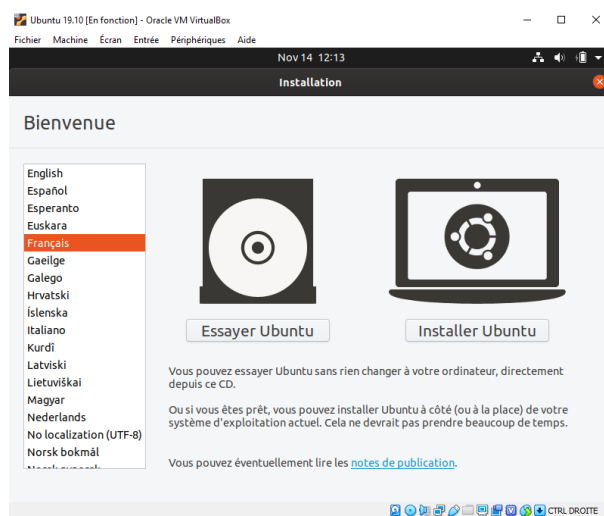
Dans le menu gauche de l'écran de paramétrage Ubuntu, cliquer sur *Pays et Langue* puis sélectionner le français dans la liste *Langue* et le panel *Sources de saisie* :



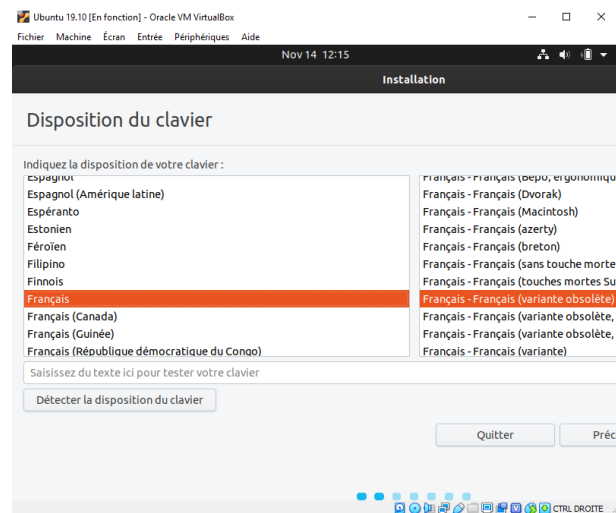
Double-cliquer ensuite dans le bureau sur l'icône *Install Ubuntu 19.10* :



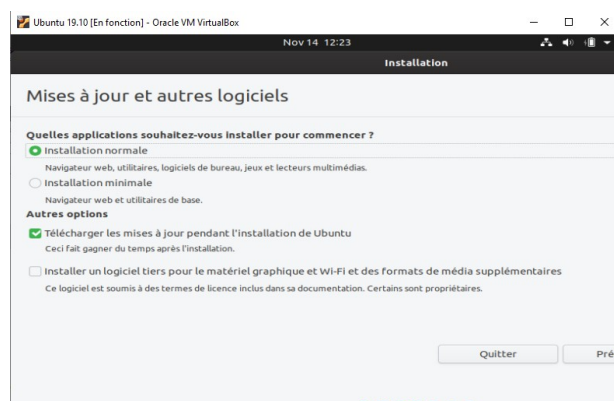
- Sélectionner la langue « Français » puis cliquer sur « Installer Ubuntu »



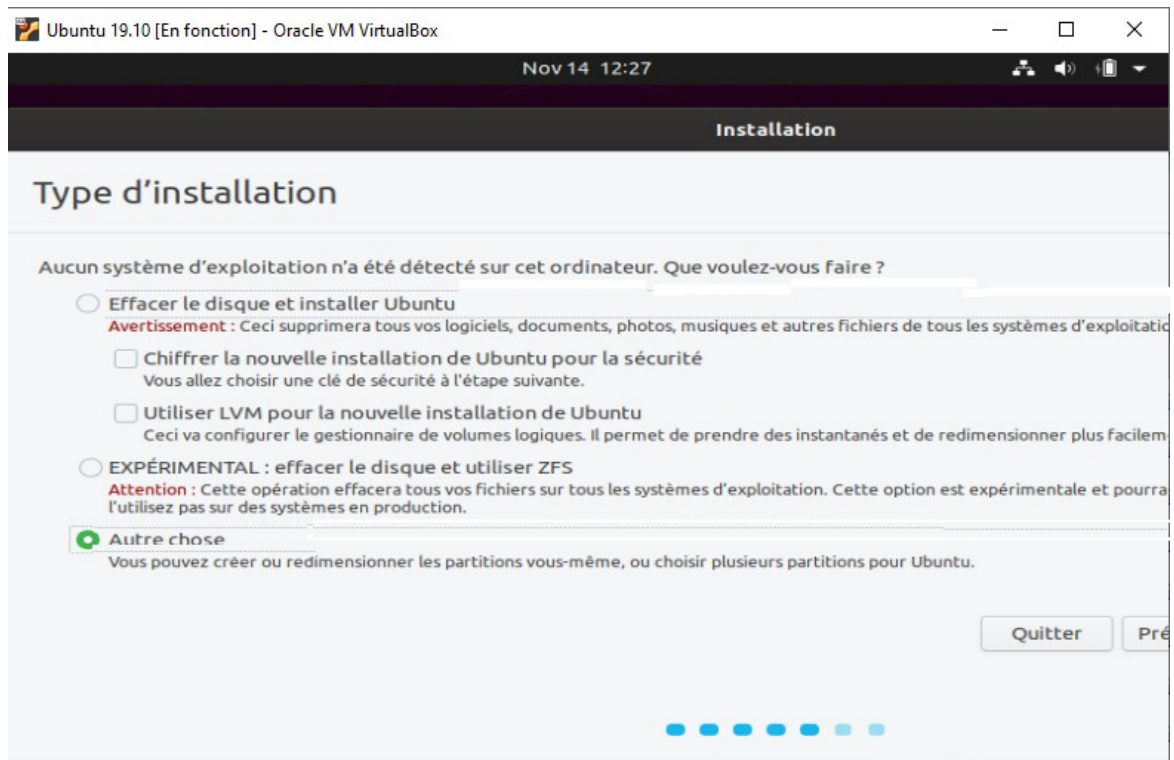
- Garder la disposition du clavier par défaut « Français-Français (variante obsolète) :



- Garder les options d'installation et de mise à jour par défaut (« Installation normale », « Télécharger les mises à jour pendant l'installation d'Ubuntu »)



- Sur l'écran du choix du type d'installation, sélectionner *Autre chose* :

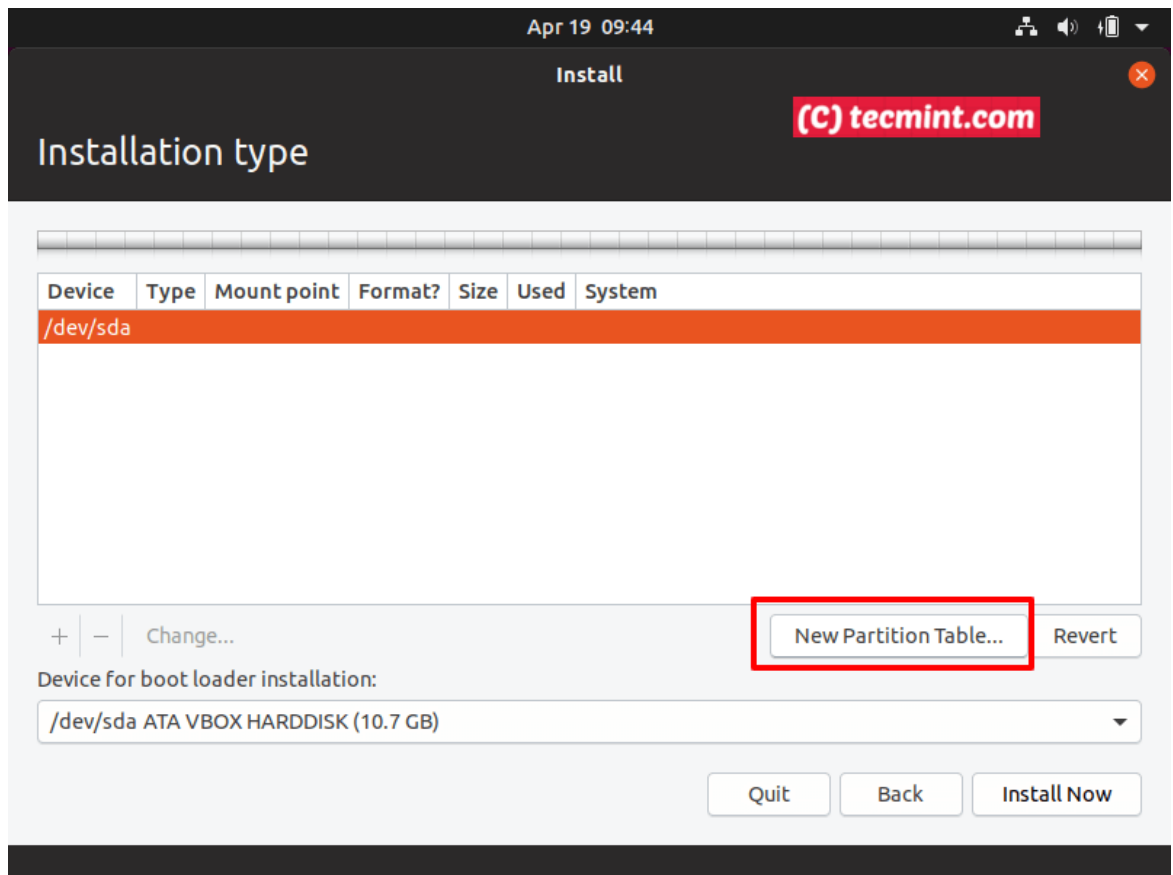


La suite de la procédure est basée sur l'article <https://www.tecmint.com/ubuntu-19-04-installation-on-uefi-firmware/> dont elle reprend les copies d'écran (en anglais)

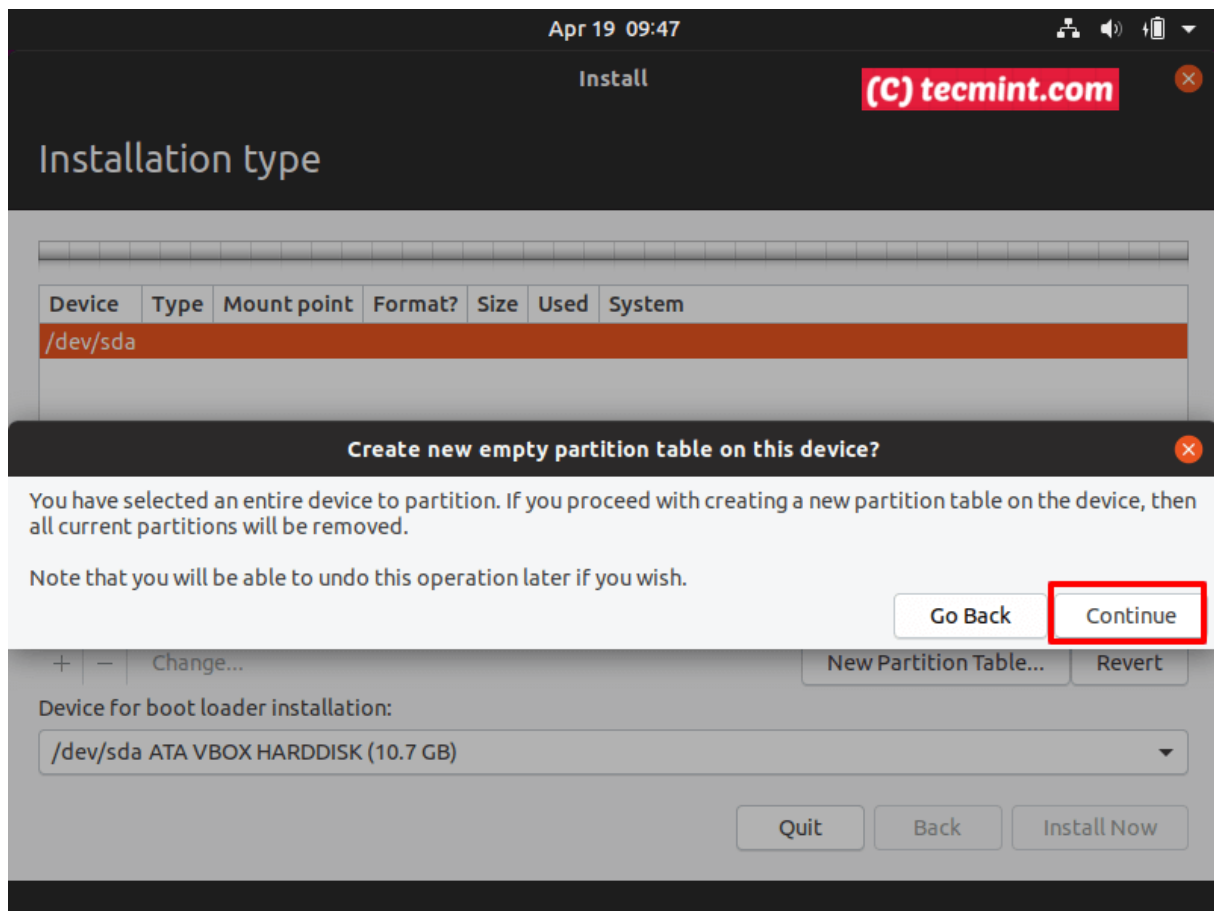
- Sélectionner dans la liste des tables de partition le disque sur lequel installer Ubuntu. Dans les copies d'écran suivante il s'agit de *sda* mais il peut tout aussi bien s'agir de *sdb* ou *sdc*.

Faire très attention dans cette étape à sélectionner le bon disque sous peine de supprimer la partition windows !

- Cliquer sur *New partition Table* :



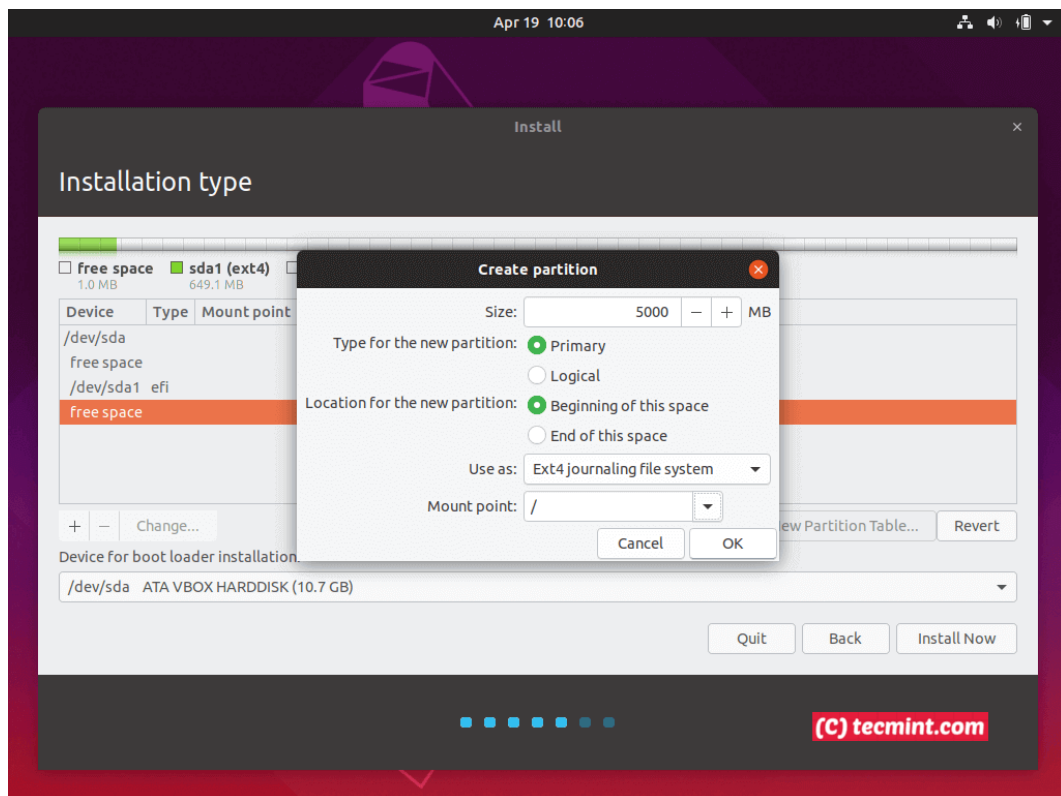
- Un message demandant la confirmation de la création de la table de partition apparaît, valider en cliquant sur *Continue* :



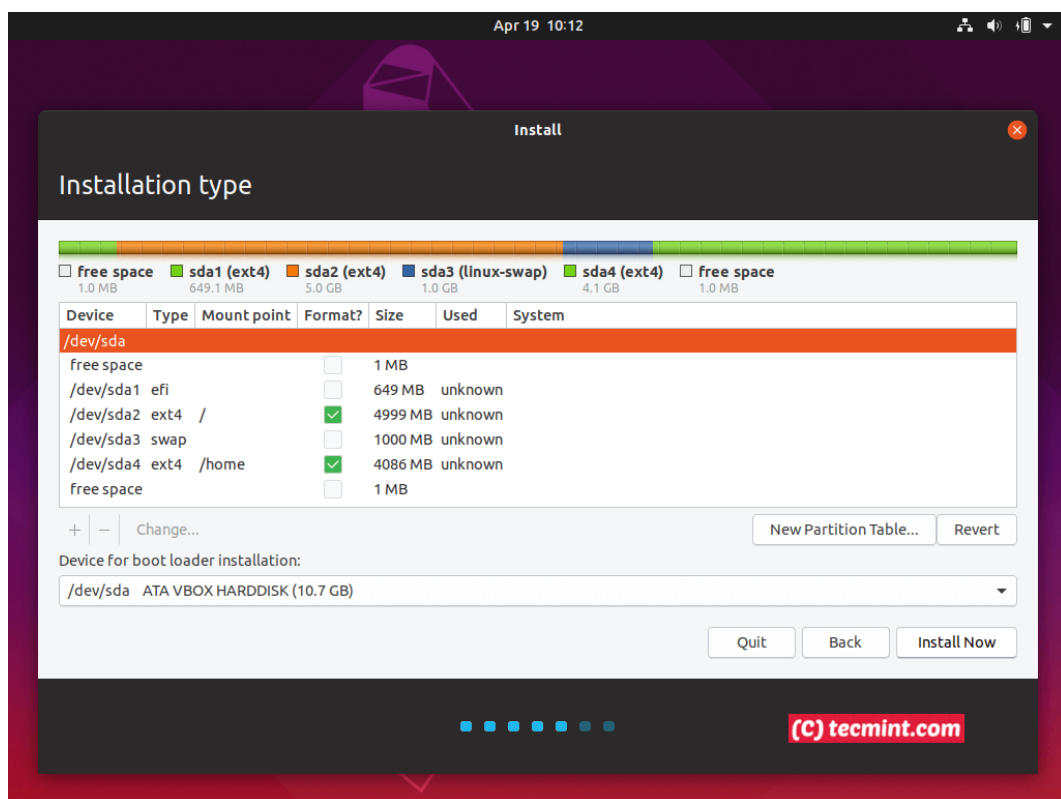
- Dans la liste *Device for boot loader installation*, sélectionner le disque sur lequel Ubuntu va être installé.
- Ajouter ensuite les partitions manuellement :
 1. Partition système EFI – 650 MB
 2. Point de montage / (partition racine) – min 5 GB – EXT4
 3. Partition de Swap – min 1GB (choisir de préférence le double de la mémoire RAM).
 4. Point de montage */home* – prendre tout l'espace disque restant – EXT4.

NB :

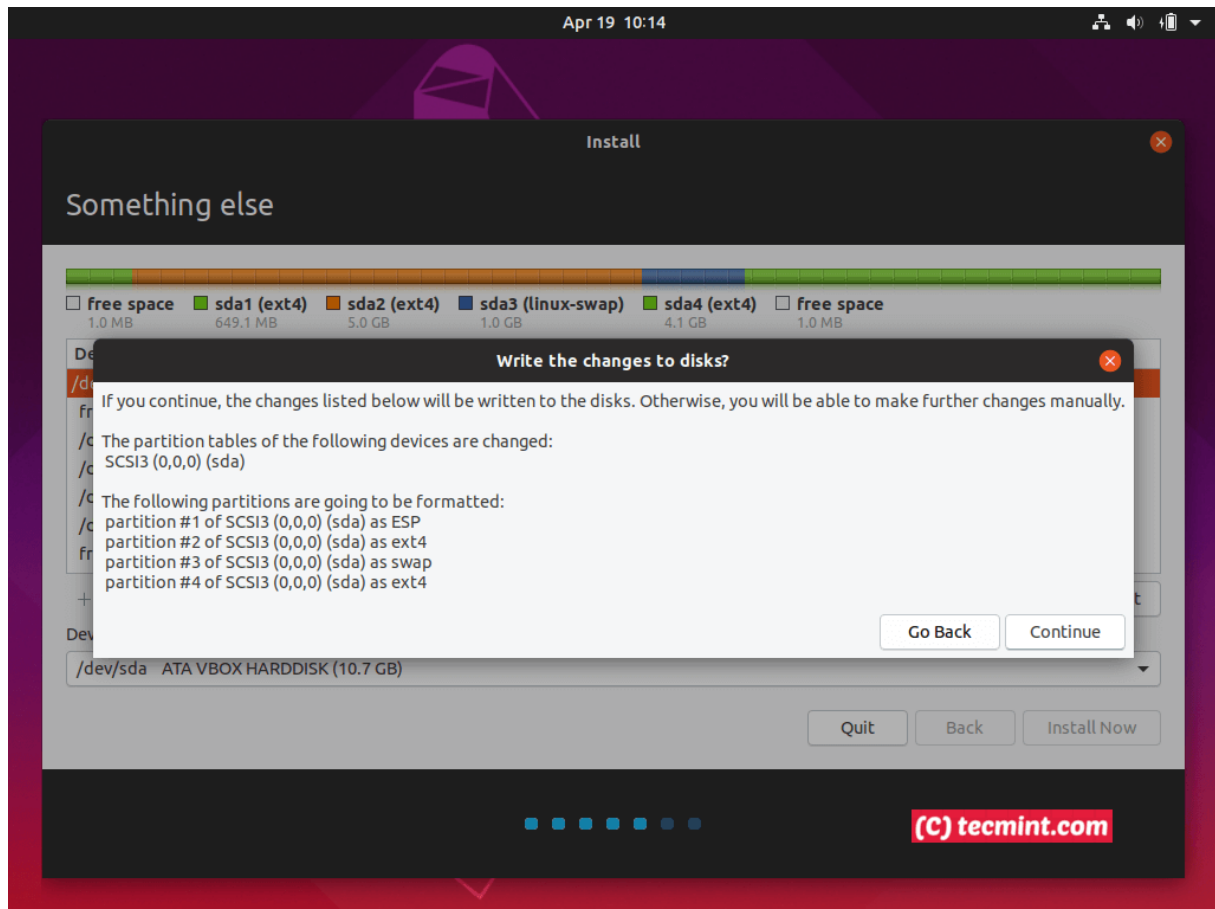
- Toutes les partitions doivent être primaires et positionnées au début de l'espace libre ;
- Si le disque cible est un disque SSD, choisir le système de fichier EXT4 pour les partitions / et */home*. Si il s'agit d'une clé USB choisir plutôt le format EXT2.
- Pour ajouter une partition cliquer simplement sur le bouton « + » et saisir les informations dans la fenêtre pop-up qui s'ouvre. Par exemple pour la partition / (racine) :



- Après avoir défini toutes les partitions, l'écran doit indiquer une table similaire à celle-ci :



- Cliquer alors sur *Install Now* :



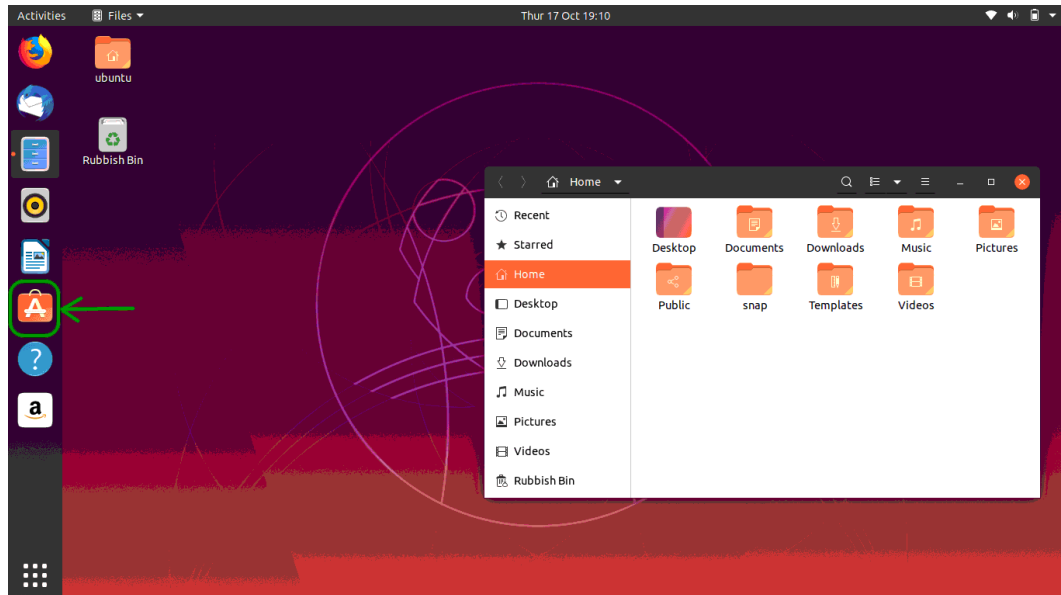
- Confirmer en cliquant sur *Continue* puis poursuivre l'installation en suivant [les instructions données dans le chapitre décrivant l'installation d'une image Ubuntu dans Virtual Box](#)
- A la fin de l'installation :
 - Arrêter le PC
 - Enlever la clé USB Ubuntu live bootable
 - Entrer dans la configuration BIOS au démarrage du PC
 - Sélectionner le disque SSD ou la clé USB contenant la nouvelle installation d'Ubuntu.

2.4 Installation des gestionnaires de paquets python pip 2 et 3

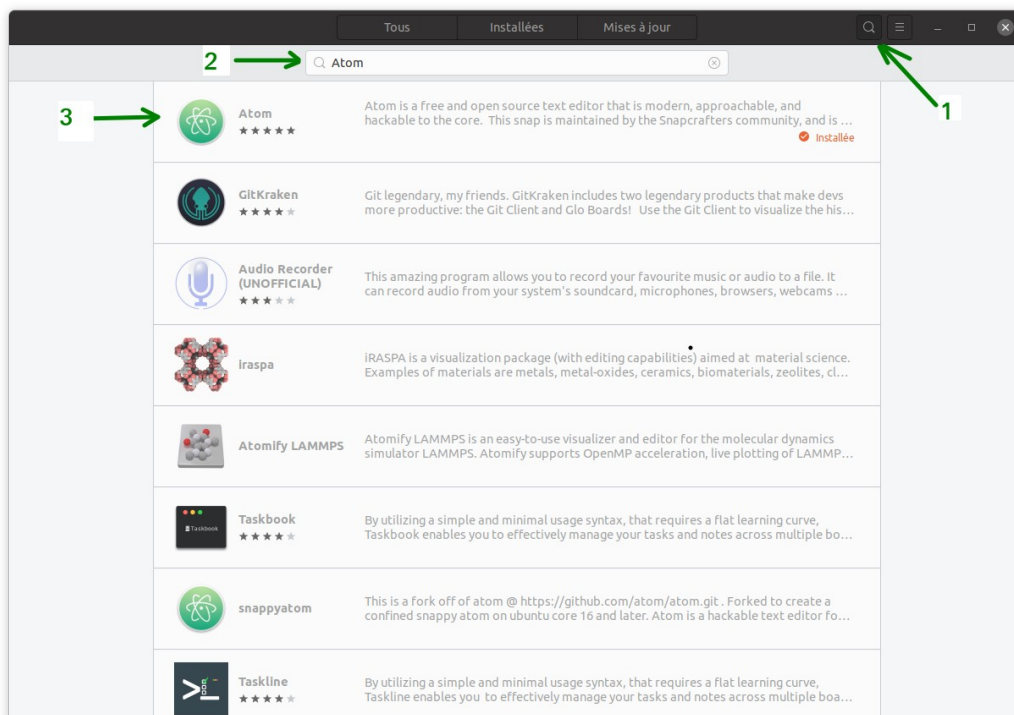
```
# Installation de pip2 :
sudo apt install python2-pip
# Installation de pip3 :
sudo apt install python3-pip
```

2.5 Installation d'un éditeur de texte polyvalent : Atom (1.42)

Cliquer sur l'icône *Logiciels Ubuntu* dans la barre latérale gauche :



Rechercher *Atom* dans la bibliothèque en ligne des logiciels Ubuntu :



1. Cliquer sur l'icône *loupe* qui donne accès à l'outil de recherche
2. Saisir « Atom » dans la zone de saisie de la recherche
3. Cliquer ensuite sur le premier résultat retourner pour démarrer l'installation

2.6 Installation d'un navigateur web sécurisé (Brave)

```
# Installation du navigateur web Brave avec le gestionnaire de paquets Snappy :  
sudo snap install brave  
# A propos du paquet snap :  
https://snapcraft.io/brave
```

2.7 Configuration de l'environnement graphique (Gnome)

Configuration de l'apparence du bureau (facultatif) :

1. Ouvrir « Logiciels Ubuntu »
2. Chercher « **gnome tweaks** » (Ajustements gnome)
3. Dans le panneau de gauche cliquez sur « Extensions »
4. Activer « Desktop icons » avec le bouton slide
5. Cliquer sur le bouton avec l'icône « rouage » associé à Desktop icons, puis :
 - Choisir taille des icônes « petite »
 - Désactiver « Montrer le dossier personnel sur le bureau »
 - Désactiver « Montrer la corbeilles ur le bureau »

La taille des icônes dans la barre latérale est configurable avec les outils standards de paramètres Ubuntu (choisir 32px)

Configurer les tailles (en pixels) des icônes sur le bureau associées aux catégories small/standard/large :

```
# Ouvrir le fichier :  
sudo gedit /usr/share/gnome-shell/extensions/desktop-icons@csoriano/prefs.js  
# Modifier les valeurs associées au paramètres const ICON_SIZE :  
const ICON_SIZE = { 'small': 32, 'standard': 48, 'large': 64 };  
# Redémarrer gnome-shell :  
killall -3 gnome-shell
```

3. Installation d'Anaconda3-2019.10

Ouvrir une fenêtre de terminal (CTRL+ALT+T) puis saisir :

```
# Installation de curl  
sudo apt install curl  
# Navigation dans le répertoire de téléchargements de l'utilisateur courant  
cd ./Téléchargements/  
# Récupération du script d'installation d'Anaconda  
curl -O https://repo.anaconda.com/archive/Anaconda3-2019.10-Linux-x86\_64.sh
```

```

# Vérification de l'intégrité de l'installateur
sha256sum Anaconda3-2019.10-Linux-x86_64.sh

46d762284d252e51cd58a8ca6c8adc9da2eadc82c342927b2f66ed011d1d8b53  Anaconda3-2019.10-Linux-
x86_64.sh

# Exécution du script d'installation :
bash Anaconda3-2019.10-Linux-x86_64.sh
# Affichage de l'accord sur la licence (taper sur ENTER pour aller au bout du texte):

Welcome to Anaconda3 2019.03

In order to continue the installation process, please review the license
agreement.

Please, press ENTER to continue

>>>
...
Do you approve the license terms? [yes|no]

# Taper « yes » afin de poursuivre la procédure d'installation :
yes
# Le script propose alors de sélectionner le répertoire d'installation
# (par défaut « anaconda3 » dans le répertoire personnel) :

Anaconda3 will now be installed into this location:

/home/daniel/anaconda3

- Press ENTER to confirm the location
- Press CTRL-C to abort the installation
- Or specify a different location below

[/home/daniel/anaconda3] >>>

# Presser sur ENTER afin de poursuivre la procédure d'installation
# Une fois l'installation terminée le message suivant s'affiche :

installation finished.

Do you wish the installer to prepend the Anaconda3 install location
to PATH in your /home/daniel/.bashrc ? [yes|no]

[no] >>>

# Taper « yes »
yes
# Activer ensuite l'installation si l'invite du terminal n'est pas préfixé par (base) :
source ~/.bashrc
# Tester l'installation :
conda list

# packages in environment at /home/daniel/anaconda3:
#
# Name                               Version           Build    Channel
_ipyw_jlab_nb_ext_conf              0.1.0             py37_0
_libgcc_mutex                        0.1               main
...

# Pour créer un nouvel environnement Anaconda (optionnel – en cas de besoin spécifique) :
conda create --name my_env python=3
# Pour activer ce nouvel environnement :
conda activate my_env

```

26 Installation d'Anaconda3-2019.10

4. Installation de R (3.6.1) et RStudio (1.2)

4.1 Installation de R (3.6.1)

La procédure est la suivante :

```
# Rechercher les mises à jour de package
sudo apt update
# Installer le package R
sudo apt install r-base
```

4.2 Installation de RStudio(1.2)

La procédure est la suivante :

1. Télécharger [Rstudio](#) ;
2. Double cliquer sur le fichier `rstudio-1.2.5033-amd64.deb` dans l'explorateur de fichiers et se laisser guider par l'installateur.

5. Installation de Java 8 (JDK/JRE Oracle)

5.1 Téléchargement de Java

Remarque préalable : le téléchargement d'un JDK Oracle nécessite de créer un compte utilisateur sur le site d'Oracle.

La procédure est la suivante :

1. Dans un navigateur web, ouvrir la page de téléchargements Java sur le site d'Oracle :
<https://www.oracle.com/technetwork/java/javase/downloads/index.html> ;
2. Sélectionner la dernière version stable du jdk 8. Au moment où ce document est rédigé, il s'agit du [Java SE Development Kit 8u231](#) ;
3. Sur la page de téléchargement cliquer sur le radio bouton « *Accept Licence Agreement* ». Le fichier d'installation pour une distribution linux générique exécutée sur un PC 64 bits est : [jdk-8u231-linux-x64.tar.gz](#).

5.2 Installation de Java

Ouvrir une fenêtre de terminal et exécuter les commandes suivantes :

```
# Se positionner dans le répertoire de téléchargements :
cd Téléchargements/
# Décompresser l'archive contenant le jdk
tar -xvf jdk-8u231-linux-x64.tar.gz
# Déplacer le jdk dans /usr/lib
sudo mkdir -p /usr/lib/jvm
sudo mv ./jdk1.8.0_231 /usr/lib/jvm/
# Assigner au jdk d'Oracle une priorité de 1, (priorité basse de manière à ce que l'installation d'une autre version
de jdk remplace cette version)
sudo update-alternatives --install "/usr/bin/java" "java" "/usr/lib/jvm/jdk1.8.0_231/bin/java" 1
sudo update-alternatives --install "/usr/bin/javac" "javac" "/usr/lib/jvm/jdk1.8.0_231/bin/javac" 1
sudo update-alternatives --install "/usr/bin/javaws" "javaws" "/usr/lib/jvm/jdk1.8.0_231/bin/javaws" 1
# Corriger les droits d'accès
sudo chmod a+x /usr/bin/java
sudo chmod a+x /usr/bin/javac
sudo chmod a+x /usr/bin/javaws
sudo chown -R root:root /usr/lib/jvm/jdk1.8.0_231
# Visualiser les alternatives java
sudo update-alternatives --config java

Il n'existe qu'une « alternative » dans le groupe de liens java (qui fournit
/usr/bin/java) : /usr/lib/jvm/jdk1.8.0_231/bin/java
Rien à configurer.
```

6. Installation de Scala (2.12.10)

Nous installerons la dernière version 2.12 de Scala, compatible avec Apache Spark 2.4.4. La procédure est la suivante :

```
# Se positionner dans le répertoire de téléchargements :
cd Téléchargements/
# Télécharger scala 2.12.10
wget http://www.scala-lang.org/files/archive/scala-2.12.10.tgz
# Créer le répertoire racine de scala
sudo mkdir /usr/local/src/scala
# Décompresser de l'archive scala dans le répertoire approprié
sudo tar xvf scala-2.12.10.tgz -C /usr/local/src/scala/
```

7. Installation de Hadoop (3.2.1)

Tout d'abord il convient de créer un utilisateur hadoop. **Le reste de la procédure d'installation est réalisée avec le compte hadoop :**

```
# Installer le serveur et le client OpenSSH
sudo apt-get install openssh-server openssh-client
# Créer l'utilisateur hadoop (password : hadoop):
sudo adduser hadoop
# Ajout de l'utilisateur hadoop au groupe pouvant exécuter un sudo
sudo usermod -aG sudo hadoop
# Se connecter en tant qu'utilisateur hadoop
su - hadoop
# Créer les clés (publique et privée)
ssh-keygen -t rsa
# Copier les clés de "id_rsa.pub" vers "authorized_keys"
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
# Vérifier la connexion ssh sans mot de passe avec la commande suivante
# (un message précisant que l'authenticité de l'hôte ne peut être confirmé apparaît- continuer en tapant yes) :
ssh localhost
```

On peut alors procéder au téléchargement et à l'installation de Hadoop proprement dite :

```
# Télécharger Hadoop 3.2.1 depuis le site web officiel d'apache
wget https://www-us.apache.org/dist/hadoop/common/hadoop-3.2.1/hadoop-3.2.1.tar.gz
# Décompresser l'archive dans le répertoire de l'utilisateur hadoop
tar xvf hadoop-3.2.1.tar.gz
```

Éditer les variables d'environnement de l'utilisateur hadoop comme suit :

```
nano /home/hadoop/.bashrc

export HADOOP_HOME=/home/hadoop/hadoop-3.2.1
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"

alias python=python3
alias pip=pip3
```

Mettre à jour les variables d'environnements dans la session courante

```
source ~/.bashrc
```


Éditer le fichier /home/hadoop/hadoop-3.2.1/etc/hadoop/**hadoop-env.sh** comme suit :

```
nano /home/hadoop/hadoop-3.2.1/etc/hadoop/hadoop-env.sh
export JAVA_HOME=/usr/lib/jvm/jdk1.8.0_231/
export HADOOP_CONF_DIR=${HADOOP_CONF_DIR:-"/home/hadoop/hadoop-3.2.1/etc/hadoop"}
```

Créer le répertoire /home/hadoop/hadooptmpdata et éditer le fichier
/home/hadoop/hadoop-3.2.1/etc/hadoop/**core-site.xml** comme suit :

```
mkdir /home/hadoop/hadooptmpdata
nano /home/hadoop/hadoop-3.2.1/etc/hadoop/core-site.xml

<configuration>
<property>
<name>fs.defaultFS</name>
<value>hdfs://localhost:9000</value>
</property>
<property>
<name>hadoop.tmp.dir</name>
<value>/home/hadoop/hadooptmpdata</value>
</property>
</configuration>
```

Créer les répertoires :

- /home/hadoop/hdfs/namenode
- /home/hadoop/hdfs/datanode

et éditer le fichier /home/hadoop/hadoop-3.2.1/etc/hadoop/**hdfs-site.xml** comme suit :

```
mkdir -p /home/hadoop/hdfs/namenode
mkdir -p /home/hadoop/hdfs/datanode
nano /home/hadoop/hadoop-3.2.1/etc/hadoop/hdfs-site.xml

<configuration>
<property>
<name>dfs.replication</name>
<value>1</value>
<name>dfs.name.dir</name>
<value>file:///home/hadoop/hdfs/namenode</value>
<name>dfs.data.dir</name>
<value>file:///home/hadoop/hdfs/datanode</value>
</property>
</configuration>
```

Editer le fichier /home/hadoop/hadoop-3.2.1/etc/hadoop/**mapred-site.xml** comme suit :

```
nano /home/hadoop/hadoop-3.2.1/etc/hadoop/mapred-site.xml

<configuration>
<property>
<name>mapreduce.framework.name</name>
<value>yarn</value>
</property>
</configuration>
```

Editer le fichier /home/hadoop/hadoop-3.2.1/etc/hadoop/**yarn-site.xml** comme suit :

```
nano /home/hadoop/hadoop-3.2.1/etc/hadoop/yarn-site.xml

<configuration>
<property>
<name>mapreduceyarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
</property>
</configuration>
```

Formater le namenode avec la commande suivante :

```
hdfs namenode -format
```

Démarrer HDFS avec la commande suivante :

```
start-dfs.sh
```

Démarrer les services YARN avec la commande suivante :

```
start-yarn.sh
```

Vérifier que tous les services / daemons Hadoop ont démarré correctement :

```
/usr/lib/jvm/jdk1.8.0_231/bin/jps

18128 ResourceManager
17526 DataNode
18279 NodeManager
17367 NameNode
17735 SecondaryNameNode
```

Vérifier la version de Hadoop :

```
hadoop version
# ou
hdfs version

Hadoop 3.2.1
Source code repository https://gitbox.apache.org/repos/asf/hadoop.git -r
b3cbbb467e22ea829b3808f4b7b01d07e0bf3842
Compiled by rohithsharmaks on 2019-09-10T15:56Z
Compiled with protoc 2.5.0
From source with checksum 776eaf9eee9c0ffc370bcbc1888737
This command was run using /home/hadoop/hadoop-3.2.1/share/hadoop/common/hadoop-common-
3.2.1.jar
```

8. Installation de Spark (2.4.4)

Ouvrir une session avec l'utilisateur hadoop :

```
su - hadoop
```

Télécharger et décompresser l'archive contenant l'installation de Spark :

```
# Télécharger Spark 2.4.4 (sans Hadoop) depuis le site web officiel d'apache
wget https://www-us.apache.org/dist/spark/spark-2.4.4/spark-2.4.4-bin-without-hadoop.tgz
# Créer le répertoire d'installation Spark
mkdir /home/hadoop/spark
# Décompresser les fichiers Spark dans ce répertoire
tar xvf spark-2.4.4-bin-without-hadoop.tgz -C /home/hadoop/spark
```

Connecter Spark à Hadoop :

```
# Se placer dans le répertoire de configuration Spark
cd /home/hadoop/spark/spark-2.4.4-bin-without-hadoop/conf
# Créer le fichier de config. des variables d'environnement à partir du fichier template
cp spark-env.sh.template spark-env.sh
# Editer spark-env.sh comme suit:
nano spark-env.sh

export SPARK_DIST_CLASSPATH=$(hadoop classpath)
```

Création des variables d'environnement SPARK_HOME et SCALA_HOME puis ajout des exécutables correspondants au PATH :

```
nano ~/.bashrc

export TERM=xterm-color
export SPARK_HOME=/home/hadoop/spark/spark-2.4.4-bin-without-hadoop
export SCALA_HOME=/usr/local/src/scala/scala-2.12.10
export PATH=$SCALA_HOME/bin:$SPARK_HOME/bin:$PATH
```

NB : la ligne `export TERM=xterm-color` permet d'éviter l'erreur suivante au lancement de spark-shell :

```
[ERROR] Failed to construct terminal; falling back to unsupported
java.lang.NumberFormatException: For input string: "0x100"
    at java.lang.NumberFormatException.forInputString(NumberFormatException.java:65)
    at java.lang.Integer.parseInt(Integer.java:580)
    ...
```

Mettre à jour les variables d'environnement dans la session courante :

```
source ~/.bashrc
```

Vérifier l'installation de Scala :

```
scala -version

Scala code runner version 2.12.10 -- Copyright 2002-2019, LAMP/EPFL and Lightbend, Inc.
```

Vérifier le bon fonctionnement de Spark :

```
spark-shell
```

La console spark-shell doit s'ouvrir avec quelques avertissements (WARNings) sans gravité ; quittez la en entrant 'exit' ou 'Ctrl-D'.

9. Installation de Cassandra (3.11.5)

La procédure est la suivante :

```
# Rechercher les mises à jour de package
sudo apt update
# Installer le package apt-transport-https
sudo apt install apt-transport-https
# Importer les clés du dépôt
wget -q -O - https://www.apache.org/dist/cassandra/KEYS | sudo apt-key add -
# Ajouter le dépôt Cassandra au système
sudo sh -c 'echo "deb http://www.apache.org/dist/cassandra/debian 311x main" > /etc/apt/sources.list.d/cassandra.list'
# Mise à jour de la liste des packages APT
sudo apt update
# Installer le package Cassandra
sudo apt install cassandra
```

Le service Cassandra démarre automatiquement après l'installation. Pour vérifier que Cassandra s'exécute normalement, taper la commande suivante :

```
nodetool status

Datacenter: datacenter1
=====
Status=Up/Down
|/ State=Normal/Leaving/Joining/Moving
-- Address  Load      Tokens   Owns (effective)  Host ID                               Rack
UN 127.0.0.1  103,72 KiB  256      100,0%            45df8f19-bdda-4f05-b988-688eec95b22c rack1
```

Pour se connecter en ligne de commandes à Cassandra, taper simplement :

```
cqlsh
```

Un bon client graphique est le DevCenter de datastax, disponible à l'URL suivante :

<https://downloads.datastax.com/#devcenter>

Pour l'installer, il suffit de décompresser l'archive.

10. Installation d'Elasticsearch (7.5)

La procédure est la suivante :

```
# Rechercher les mises à jour de package
sudo apt update
# Installer le package apt-transport-https
sudo apt install apt-transport-https
# Importer les clés du dépôt
wget -qO - https://artifacts.elastic.co/GPG-KEY-elasticsearch | sudo apt-key add -
# Ajouter le dépôt Elasticsearch au système
sudo sh -c 'echo "deb https://artifacts.elastic.co/packages/7.x/apt stable main" > /etc/apt/sources.list.d/elastic-7.x.list'
# Mise à jour de la liste des packages APT
sudo apt update
# Installer le package Elasticsearch
sudo apt install elasticsearch
```

Elasticsearch ne démarre pas automatiquement suite à l'installation. Il reste donc à activer puis à démarrer le service :

```
# Activer le service Elasticsearch
sudo systemctl enable elasticsearch.service
# Démarrer le service Elasticsearch
sudo systemctl start elasticsearch.service
```

Test de l'installation :

```
# La commande suivante permet d'obtenir des informations sur la version d'Elasticsearch installée
curl -X GET "localhost:9200/"

{
  "name" : "e576",
  "cluster_name" : "elasticsearch",
  "cluster_uuid" : "09KCh9vfSR6lWaaAlghB7A",
  "version" : {
    "number" : "7.5.0",
    "build_flavor" : "default",
    "build_type" : "deb",
    "build_hash" : "e9ccaed468e2fac2275a3761849cbee64b39519f",
    "build_date" : "2019-11-26T01:06:52.518245Z",
    "build_snapshot" : false,
    "lucene_version" : "8.3.0",
    "minimum_wire_compatibility_version" : "6.8.0",
    "minimum_index_compatibility_version" : "6.0.0-beta1"
  },
  "tagline" : "You Know, for Search"
}
```

NB : Elasticsearch est installé dans le répertoire **/usr/share/elasticsearch/**

11. Installation de Kibana(7.5)

Il convient d'installer d'Elasticsearch comme décrit au chapitre précédent avant d'installer Kibana avec la procédure suivante :

```
# Rechercher les mises à jour de package
sudo apt update
# Installer le package Kibana
sudo apt install kibana
```

Elasticsearch ne démarre pas automatiquement suite à l'installation. Il reste donc à activer puis à démarrer le service :

```
# Activer le service Kibana
sudo systemctl enable kibana
# Démarrer le service Kibana
sudo systemctl start kibana
```

Pour accéder à kibana, ouvrir un navigateur Web et taper l'URL suivante :

<http://localhost:5601/>