
Feuille de route des TPs Hadoop

Ceci est une assistance pour réaliser les TPs relatifs à l'écosystème Hadoop.

C'est important de télécharger les ressources à partir des sites officiels et choisir des versions stables et de suivre les étapes du TP pas à pas, une étape sauté ou ignorée déstabilise l'installation.

Commencer d'abord par télécharger tous les outils sur votre ordi.

L'idée est d'utiliser une machine virtuelle, et la configurer dans un premier temps en nœud simple.

Les étapes d'installation se résume comme suit :

- 1. Installer et configurer la machine virtuelle, il existe plusieurs solutions, pour ces TPs , j'ai choisi VirtualBox-6.1.26-145957-Win :**
<https://www.oracle.com/virtualization/technologies/vm/downloads/virtualbox-downloads.html>

- Télécharger le logiciel oracle virtual box et l'installer (virtualBox-6.0.4....exe)
- Lancer virtual box --> ajouter--> machine --> parcourir jusqu'au dossier de la machine virtuelle après décompression --> sélectionner l'icône de couleur bleu
- Sélectionner la machine et démarrer

Remarque :

- il faut avoir 6 gigas de RAM minimum pour pouvoir démarrer la machine virtuelle
- si vous avez moins de 6 giga de ram il faut sélectionner la machine virtuelle --> configuration --> système et diminuer la ram de la machine
- il faut que votre pc permet la virtualisation, en cas de message d'erreur lors du lancement, il faut redémarrer le pc -->

Accéder au bios --> mettre l'option "virtualization technologies" à "enabled"

Sur YouTube, il y a plein de vidéos pour expliquer cette étape.

- 2. Installer Ubuntu sur cette machine virtuelle (ubuntu-20.04.2.0-desktop-amd64) à télécharger à partir de :**

<https://releases.ubuntu.com/20.04/>

J'ai choisi ce système d'exploitation pour plus de facilité, vous pouvez tester Windows, seulement, il faut choisir après une version de hadoop compatible avec Windows

3. **Télécharger et Installer java** car hadoop est une application réalisée en java et nécessite pour son installation le jdk, donc, il faut choisir une version compatible avec Ubuntu : <https://www.oracle.com/java/technologies/downloads/#jdk17-windows>
4. Télécharger et installer Hadoop, choisissez la version la plus stable :

<https://hadoop.apache.org/releases.html>

Les étapes d'installation nécessitent quelques instructions en linux, donc, en cas de non compréhension, regarder un peu de syntaxe sur Google.

L'accès aux écrans de Hadoop peut se faire à travers des commandes ou des liens web, pour simplifier on utilisera les liens web.