



# Gestion de Projet Big Data & Développement d'applications Big Data

EDAH Kodjo  
Consultant Systèmes d'Information, Big-Data



# Objectifs

- Comprendre la notion et les spécificités du Big Data
- Connaître les outils de collecte, de traitement et d'exploitation des données
- Savoir utiliser les outils de visualisation des données (Dataviz)
- Piloter et maîtriser les risques des projets



## Partie 2 : Outils de collecte, de traitement et d'exploitation des données

- Les types de bases de données
- Les propriétés des bases de données
- L'exploitation des données
- TP : Hadoop MapReduce





# TP : Installation machine virtuelle cloudera

- ✓ Téléchargez la machine virtuelle (5.5Go) <https://bit.ly/38qYXUF>
- ✓ Dézippez le fichier téléchargé : cloudera-quickstart-vm-5.12.0-0-virtualbox
- ✓ Lancer Virtual Box
- ✓ Choisir le menu Fichier => Importer un appareil virtuel ou ( Ctrl + I)
- ✓ Naviguer pour aller sélectionner le fichier dezipié



Appareil virtuel à importer

Please choose the source to import appliance from. This can be a local file system to import OVF archive or one of known cloud service providers to import cloud VM from.

Source: Local File System

Please choose a file to import the virtual appliance from. VirtualBox currently supports importing appliances saved in the Open Virtualization Format (OVF). To continue, select the file to import below.

Fichier :















# TP : Installation machine virtuelle cloudera

← Importer un appareil virtuel

## Paramètres de l'appareil virtuel

Voici les machines virtuelles décrites dans l'appareil virtuel et les paramètres suggérés pour les machines importées. Vous pouvez en changer certains en les autres avec les cases à cocher.













Système virtuel 1	
 Nom	cloudera-quickstart-vm-5.12.0-0-virtualbox 2
 Système d'exploitation invité	 Red Hat (64-bit)
 Processeur	1
 Mémoire vive	4096 MB
 DVD	<input checked="" type="checkbox"/>
 Carte réseau	<input checked="" type="checkbox"/> Intel PRO/1000 MT Desktop (82540EM)
 Contrôleur de stockage (IDE)	PIIX4
 Contrôleur de stockage (IDE)	PIIX4
 Disque virtuel	cloudera-quickstart-vm-5.12.0-0-virtualbox-disk1.vmdk
 Dossier de base	C:\Users\edah\VirtualBox VMs
 Groupe primaire	/

# TP : Installation machine virtuelle cloudera

← Importer un appareil virtuel

## Paramètres de l'appareil virtuel

Voici les machines virtuelles décrites dans l'appareil virtuel et les paramètres suggérés pour les machines importées. Vous pouvez en changer certains en les autres avec les cases à cocher.

Système virtuel 1	
 Nom	cloudera-quickstart-vm-5.12.0-0-virtualbox 2
 Système d'exploitation invité	 Red Hat (64-bit)
 Processeur	1
 Mémoire vive	4096 MB
 DVD	<input checked="" type="checkbox"/>
 Carte réseau	<input checked="" type="checkbox"/> Intel PRO/1000 MT Desktop (82540EM)
 Contrôleur de stockage (IDE)	PIIX4
 Contrôleur de stockage (IDE)	PIIX4
 Disque virtuel	cloudera-quickstart-vm-5.12.0-0-virtualbox-disk1.vmdk
 Dossier de base	C:\Users\edah\VirtualBox VMs
 Groupe primaire	/













# TP : Installation machine virtuelle cloudera

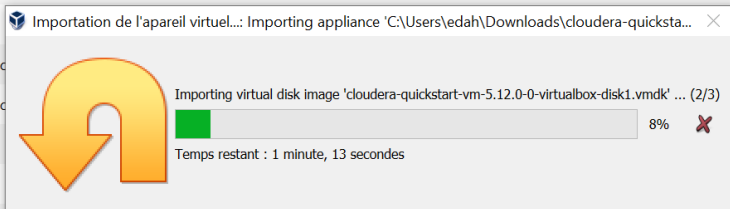
← Importer un appareil virtuel

## Paramètres de l'appareil virtuel

Voici les machines virtuelles décrites dans l'appareil virtuel et les paramètres suggérés pour les machines importées. Vous pouvez en changer certains en double-cliquant dessus et désactiver les autres avec les cases à cocher.

### Système virtuel 1

	Nom	cloudera-quickstart-vm-5.12.0-0-virtualbox 2
	Système d'exploitation invité	 Red Hat (64-bit)
	Processeur	1
	Mémoire vive	4096 MB
	DVD	<input checked="" type="checkbox"/>
	Carte réseau	
	Contrôleur de stockage	
	Contrôleur de stockage	
	Disque virtuel	
	Dossier de base	
	Groupe primaire	/

















# TP : Installation machine virtuelle cloudera

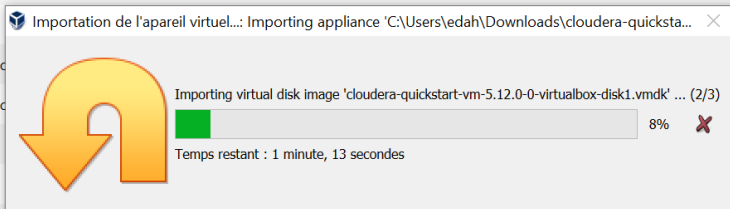
← Importer un appareil virtuel

## Paramètres de l'appareil virtuel

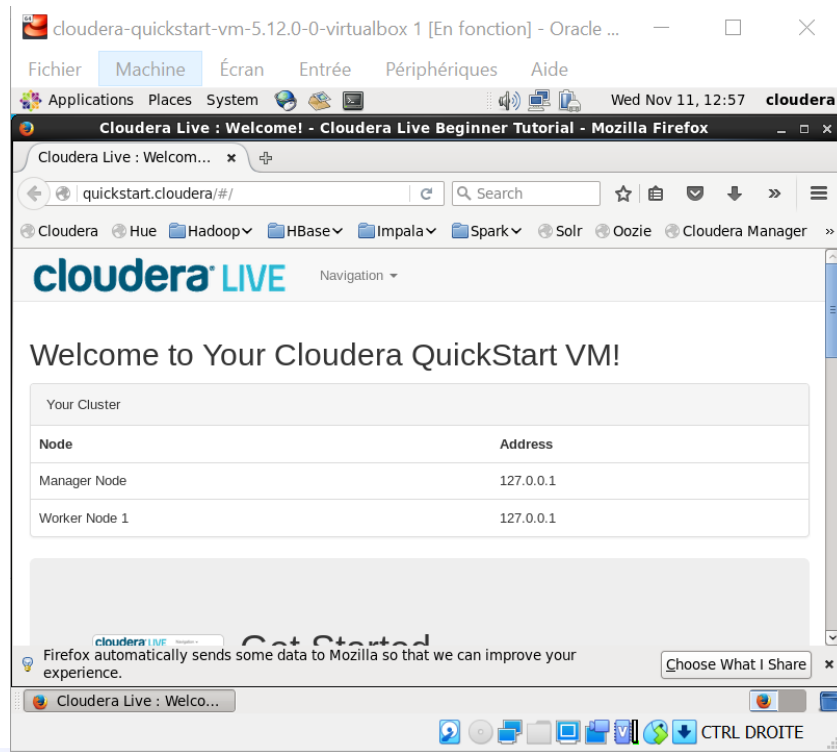
Voici les machines virtuelles décrites dans l'appareil virtuel et les paramètres suggérés pour les machines importées. Vous pouvez en changer certains en double-cliquant dessus et désactiver les autres avec les cases à cocher.

### Système virtuel 1

	Nom	cloudera-quickstart-vm-5.12.0-0-virtualbox 2
	Système d'exploitation invité	 Red Hat (64-bit)
	Processeur	1
	Mémoire vive	4096 MB
	DVD	<input checked="" type="checkbox"/>
	Carte réseau	
	Contrôleur de stockage	
	Contrôleur de stockage	
	Disque virtuel	
	Dossier de base	
	Groupe primaire	/



# TP : Installation machine virtuelle cloudera



# TP : Initiation Hadoop Map reduce

## ✓ Généralité

*Pour ces commandes, il existe 2 syntaxes possibles:*

*Avec hadoop: syntaxe du type hadoop fs <commande>*

*Avec hdfs: hdfs dfs <commande>*



# TP : Initiation Hadoop Map reduce

## ✓ Créer un dossier dans HDFS

*Commande : `hadoop fs -mkdir [-p] <paths>`*

*Exemple : `hadoop fs -mkdir /user/monDossier`*

*`hadoop fs -mkdir /user/monDossier1 /user/monDossier2 /user/monDossier3`*

*L' option -p est nécessaire si le dossier parent n' existe pas lors de la création d' un sous répertoire.*

## ✓ Lister le contenu d'un dossier

*Commande : `hadoop fs -ls <args>`*

*Exemple : `hadoop fs -ls /user`*





# TP : Initiation Hadoop Map reduce

- ✓ **Exporter un ou plusieurs fichiers de HDFS au local:**

*Commande: `hadoop fs -get [-ignorecrc] [-crc] <src> <localdst>`*

*Exemple : `hadoop fs -get /user/monDossier/monFichier.txt /home`*

- ✓ **Charger un ou plusieurs fichiers du local à HDFS:**

*Commande : `hadoop fs -put <localsrc> ... <dst>`*

*Exemple : `hadoop fs -put /home/monFichier.txt /user/monDossier`*



# TP : Initiation Hadoop Map reduce

- ✓ **Exporter un ou plusieurs fichiers de HDFS au local:**

*Commande: `hadoop fs -copyToLocal [-ignorecrc] [-crc] URI <localdst>`*

*Exemple : `hadoop fs -copyToLocal /user/monDossier/monFichier.txt /home`*

- ✓ **Charger un ou plusieurs fichiers du local à HDFS:**

*Commande : `hadoop fs -copyFromLocal <localsrc> URI`*

*Exemple : `hadoop fs -copyFromLocal /home/monFichier.txt /user/monDossier`*



# TP : Initiation Hadoop Map reduce

- ✓ **Déplacer un ou plusieurs fichiers dans HDFS:**

*Commande : `hadoop fs -mv URI [URI ...] <dest>`*

*Exemple : `hadoop fs -mv /user/monDossier1/monFichier.txt /user/monDossier2`*

- ✓ **Copier un ou plusieurs fichiers dans HDFS:**

*Commande : `hadoop fs -cp [-f] URI [URI ...] <dest>`*

*Exemple : `hadoop fs -cp /user/monDossier1/monFichier.txt /user/monDossier2`*



# TP : Initiation Hadoop Map reduce

- ✓ **Afficher le contenu d'un fichier:**

*Commande: `hadoop fs -cat URI [URI ...]`*

*Exemple : `hadoop fs -cat /user/monFichier.txt`*

- ✓ **Pour voir la description d'une commande ainsi que ses paramètres**

*Commande: `hadoop fs -help`*

*Exemple: `hadoop fs -help stat`*





Questions ?

---

Merci

