



Pratique de Hadoop n°06

Oozie (en Python)

Table des matières

I.	Objectif du TP	2
II.	Démarrer avec Cloudera.....	2
A.	Installation et configuration de Cloudera	2

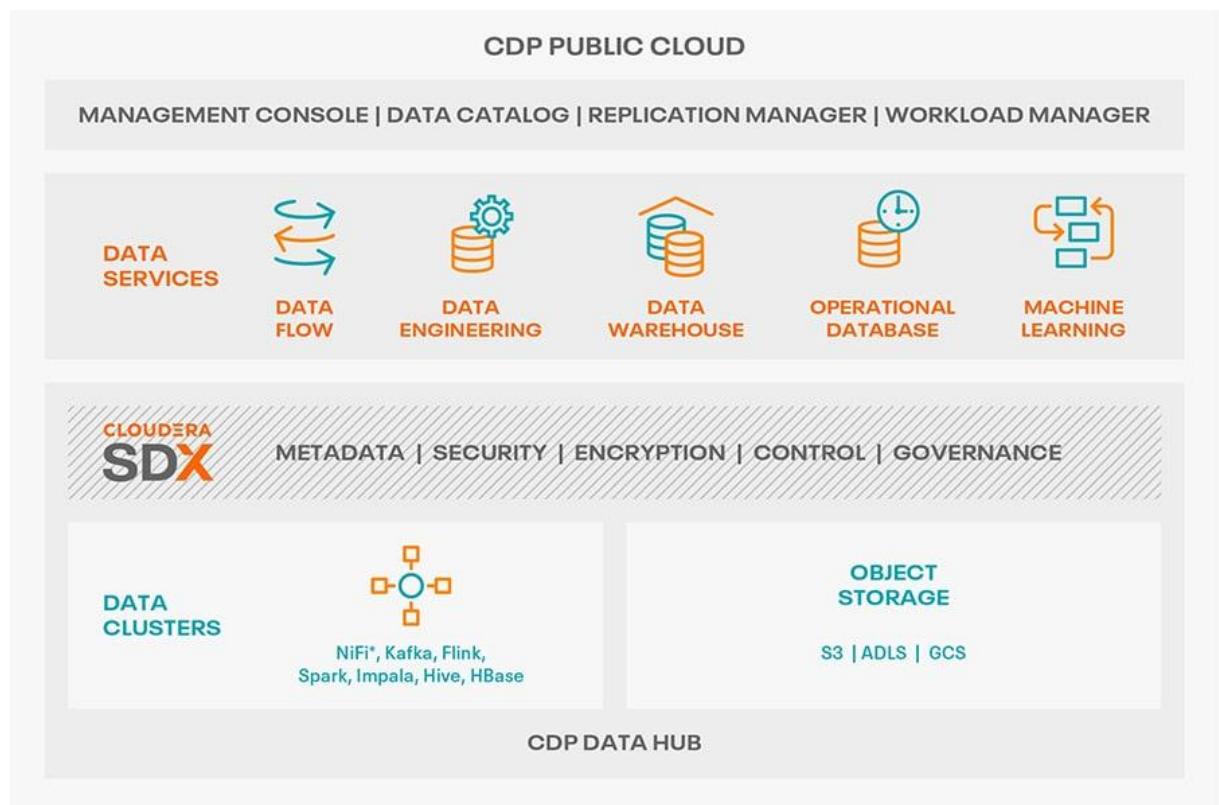
I. Objectif du TP

- Initiation à l'outil Oozie avec cloudera

A. Présentation cloudera :

Cloudera est une entreprise Américaine basée en Californie, elle se consacre au développement d'une solution Big Data basée historiquement sur le framework distribué Hadoop et qui est en train de se réorienter vers le Cloud. La Cloudera Data Plateform permet une gestion unifiée, centralisée et hybride des infrastructures Big Data pour toutes les entreprises souhaitant traiter de la donnée.

Il s'agit par ailleurs d'une solution complète intégrant les expériences de gestion Big Data et d'analyse tout au long du cycle de vie des données. Elle assure aussi la sécurité, la conformité aux normes en vigueur, la migration des données et la gestion des métadonnées sur tous les environnements.



*Flow Management rate card

II. Démarrer avec Cloudera

A. Installation et configuration de Cloudera

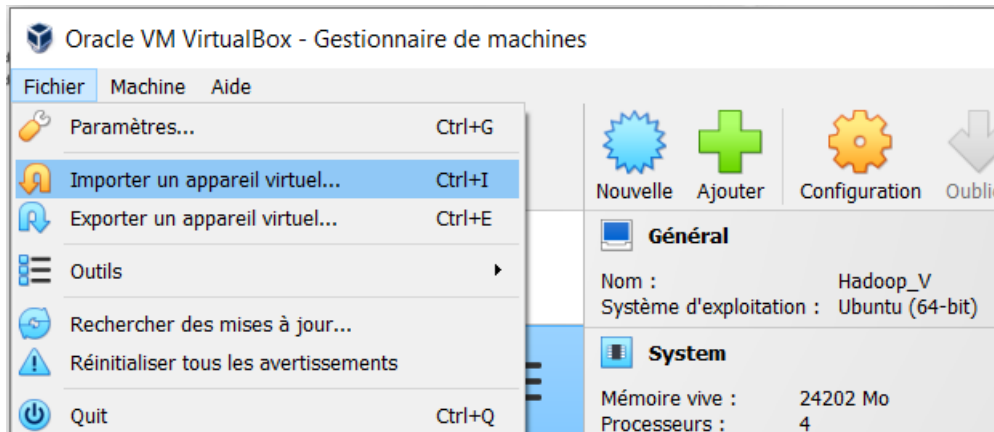
1. Télécharger et installer Virtual box :

<https://www.virtualbox.org/wiki/Downloads>

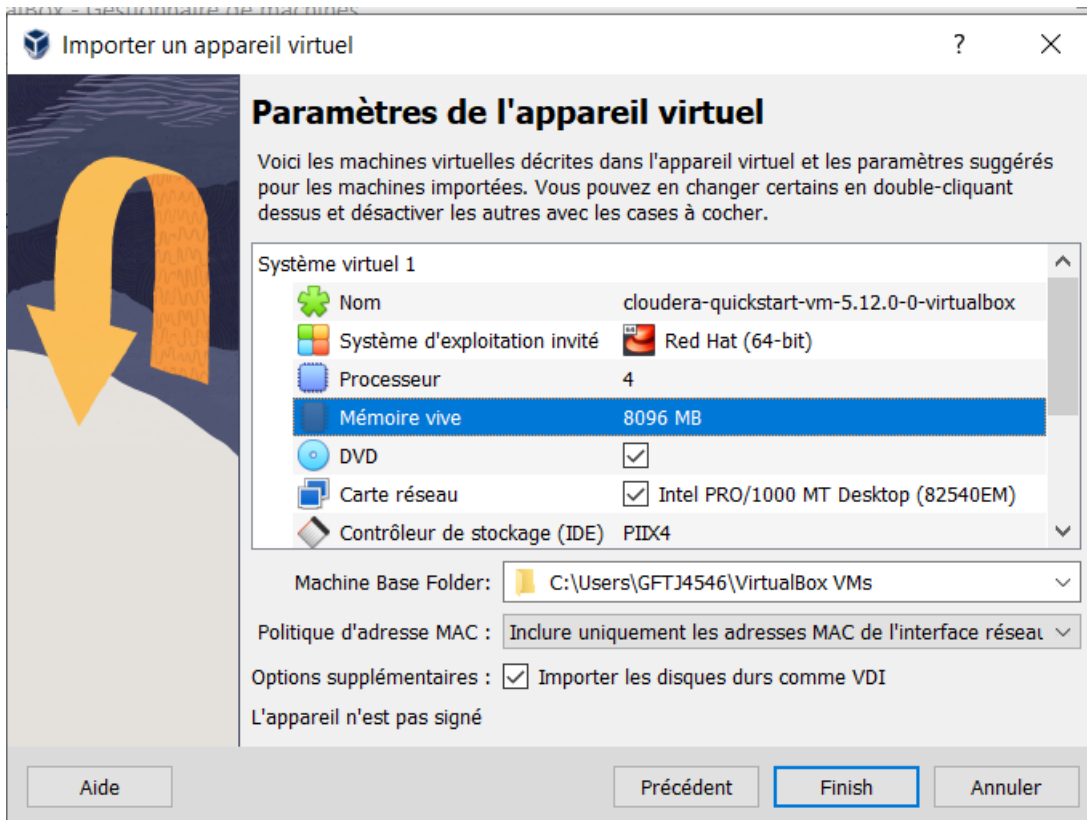
2. Puis, télécharger Cloudera quick Start VM:

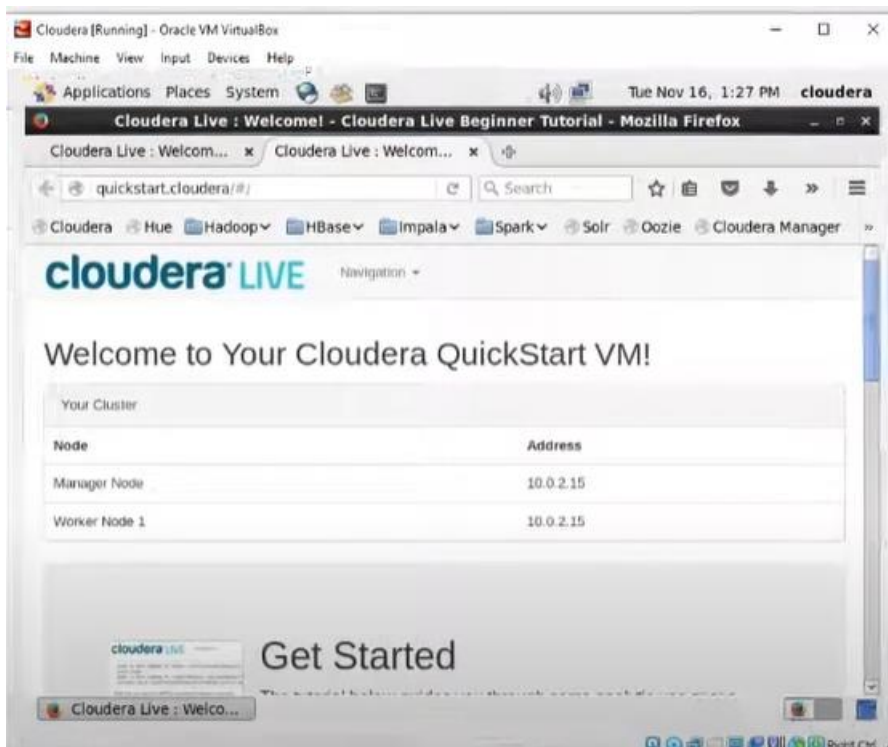
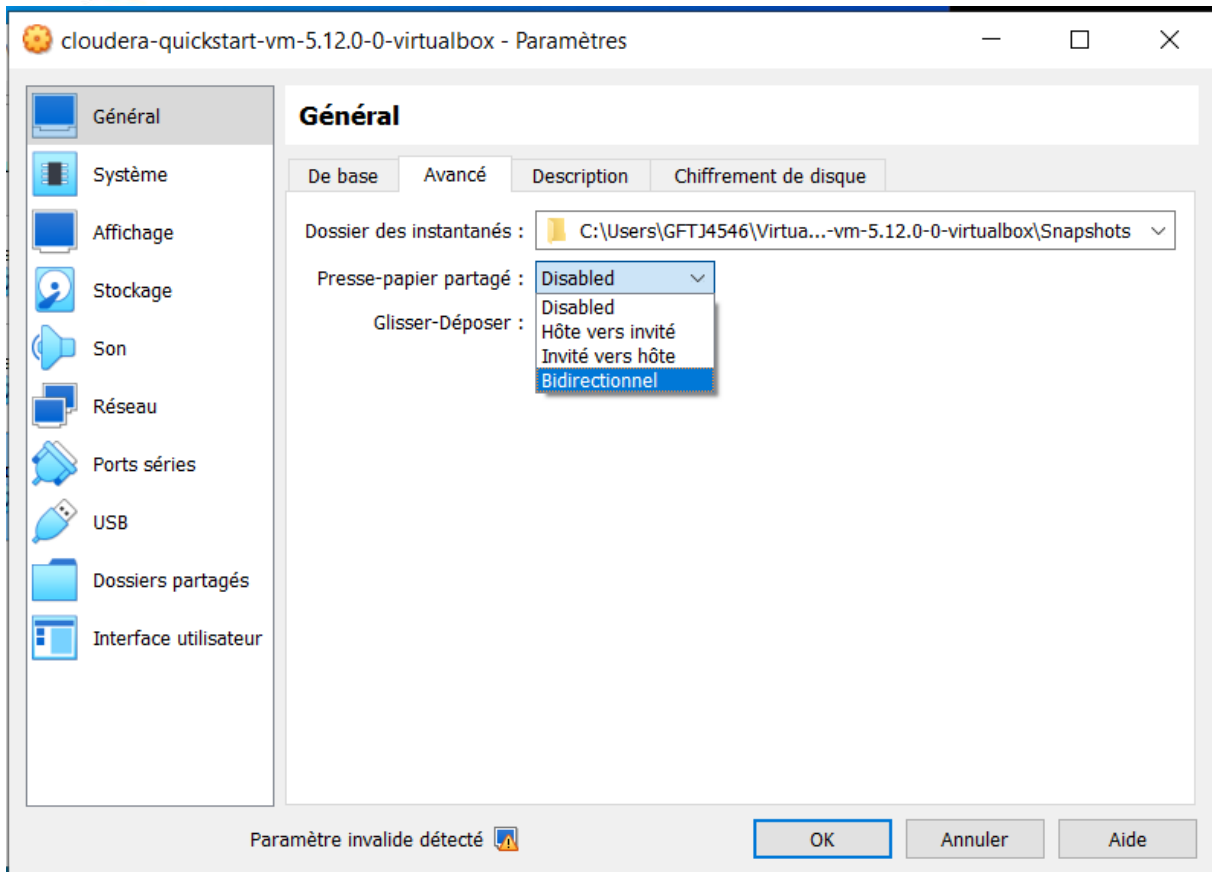
https://downloads.cloudera.com/demo_vm/virtualbox/cloudera-quickstart-vm-5.12.0-0-virtualbox.zip

3. Installer virtual box et importer Cloudera QuickStart VM:



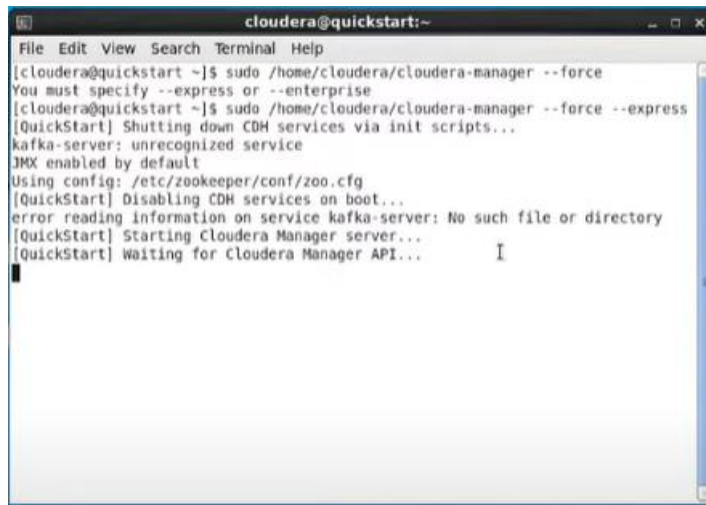
Note : 8G minimum et 2 CPU minimum





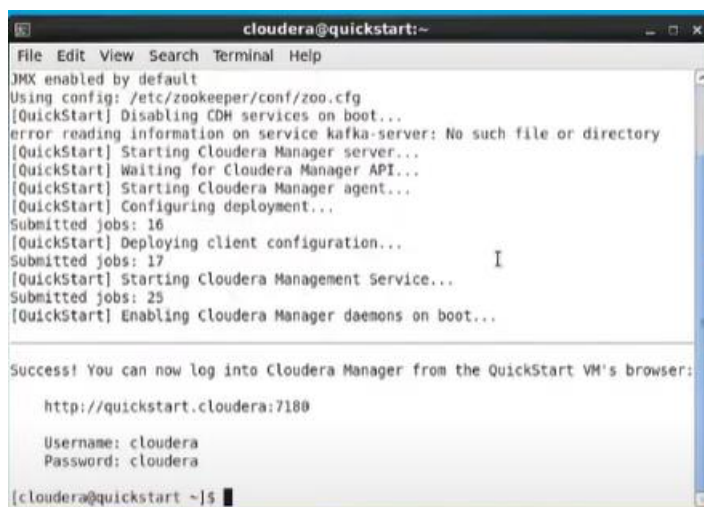
4. Après, aller sur terminal et lancer la commande :

« `sudo /home/cloudera/cloudera-manager -force -express` » :



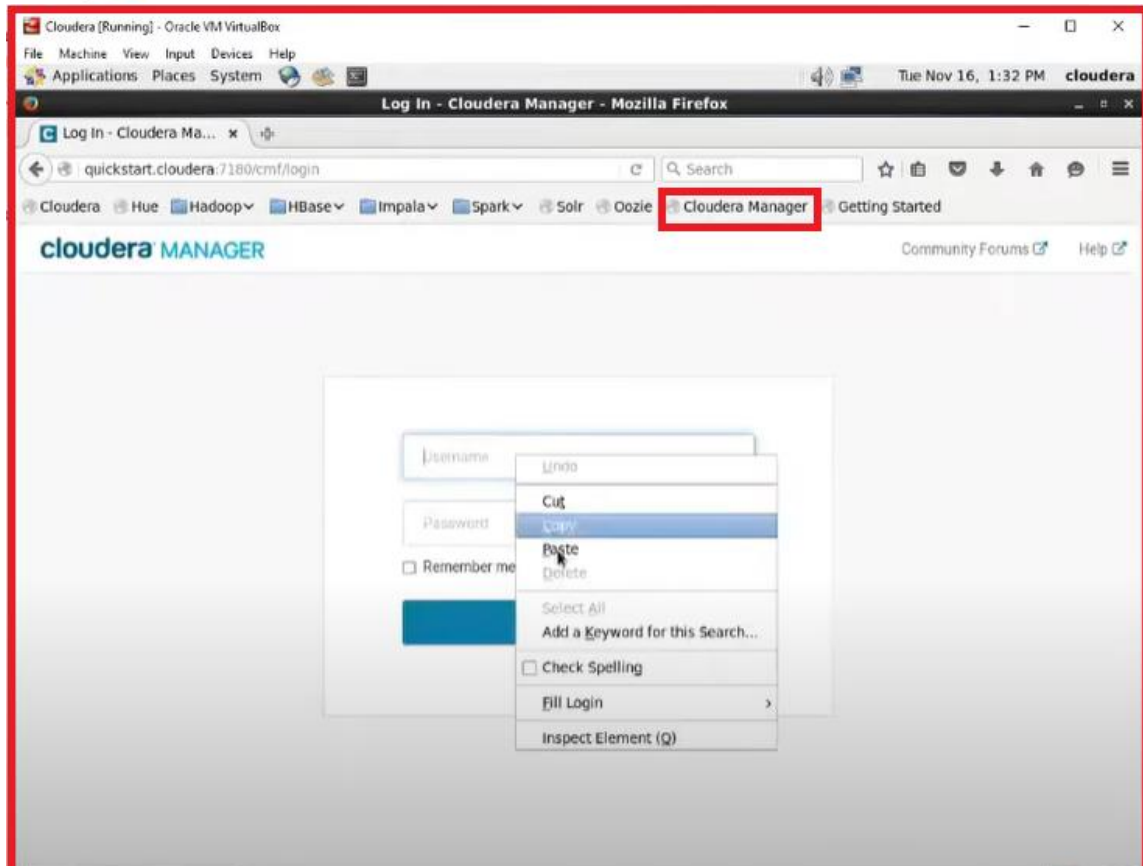
```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
[cloudera@quickstart ~]$ sudo /home/cloudera/cloudera-manager --force  
You must specify --express or --enterprise  
[cloudera@quickstart ~]$ sudo /home/cloudera/cloudera-manager --force --express  
[QuickStart] Shutting down CDH services via init scripts...  
kafka-server: unrecognized service  
JMX enabled by default  
Using config: /etc/zookeeper/conf/zoo.cfg  
[QuickStart] Disabling CDH services on boot...  
error reading information on service kafka-server: No such file or directory  
[QuickStart] Starting Cloudera Manager server...  
[QuickStart] Waiting for Cloudera Manager API...  
I
```

5. Attendre, quelques minutes jusqu'au avoir ce message :

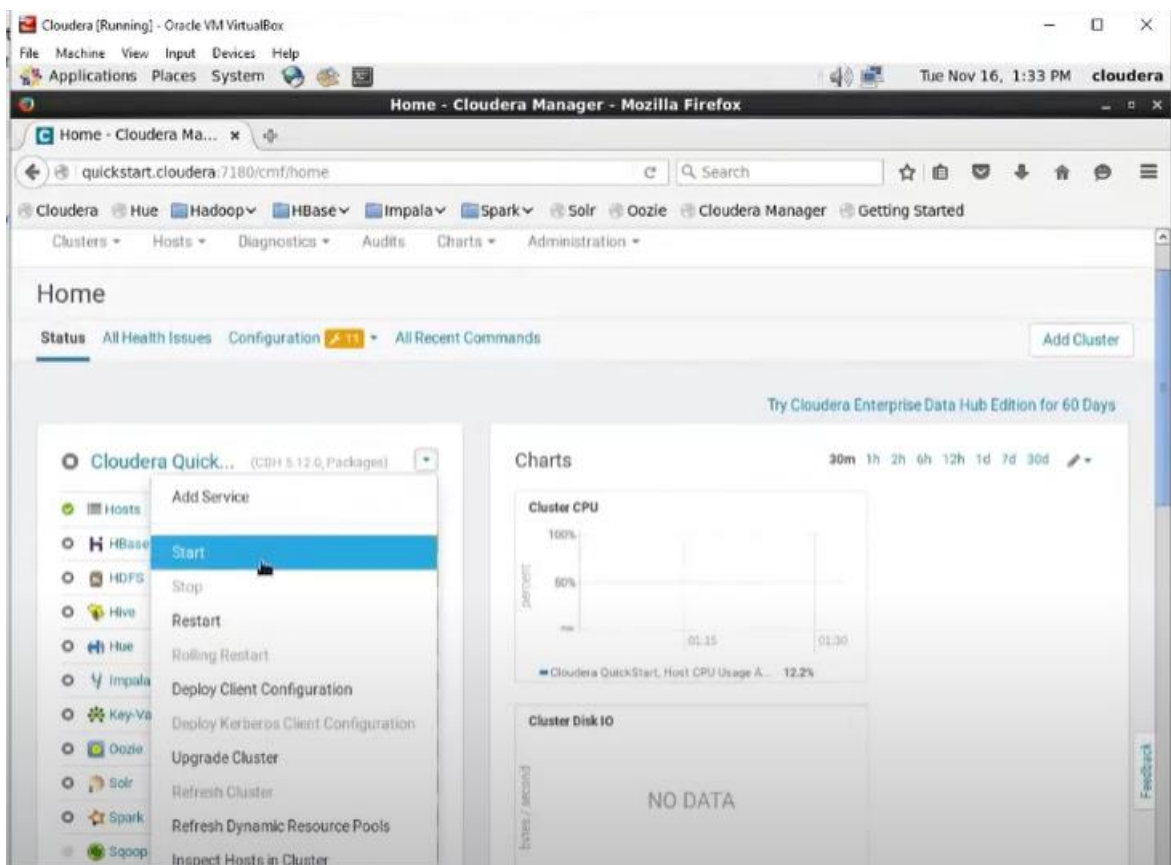


```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
JMX enabled by default  
Using config: /etc/zookeeper/conf/zoo.cfg  
[QuickStart] Disabling CDH services on boot...  
error reading information on service kafka-server: No such file or directory  
[QuickStart] Starting Cloudera Manager server...  
[QuickStart] Waiting for Cloudera Manager API...  
[QuickStart] Starting Cloudera Manager agent...  
[QuickStart] Configuring deployment...  
Submitted jobs: 16  
[QuickStart] Deploying client configuration...  
Submitted jobs: 17  
[QuickStart] Starting Cloudera Management Service...  
Submitted jobs: 25  
[QuickStart] Enabling Cloudera Manager daemons on boot...  
  
Success! You can now log into Cloudera Manager from the QuickStart VM's browser:  
  
http://quickstart.cloudera:7180  
  
Username: cloudera  
Password: cloudera  
[cloudera@quickstart ~]$
```

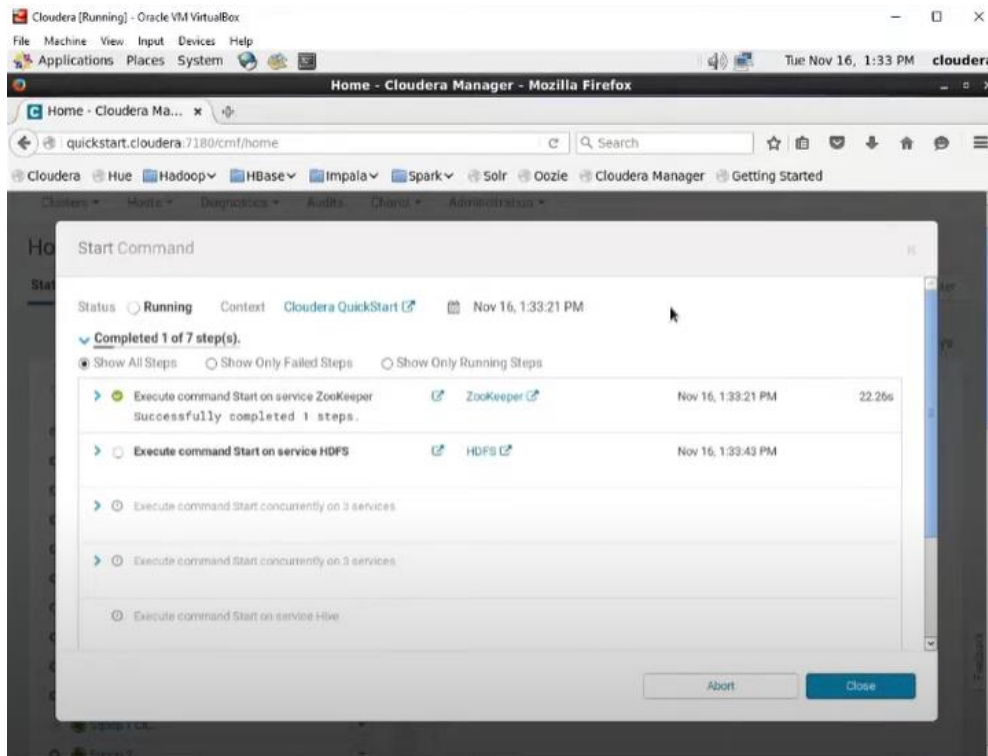
6. Puis, sur le navigateur lancer le Cloudera manager et se connecter avec les username et le mot de passe fourni avant :



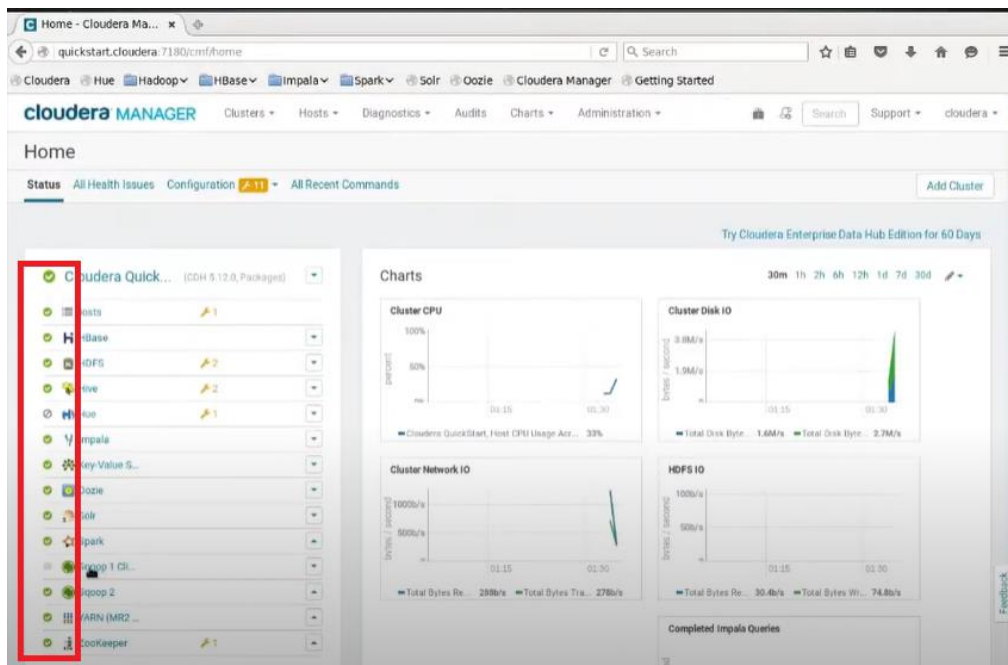
7. Lancer les différents outils big data en cliquant sur le Button Start :



8. Attendre, un moment que le lancement termine :



Tous les outils sont lancés :



III. Démarrer avec Oozie sur Cloudera :

A. Les exemples Oozie fournis par cloudera :

1. Lancer la commande suivante sur le terminal :

```
tar -xvf /usr/share/doc/oozie-4.1.0+cdh5.12.0+442/oozie-examples.tar.gz -C  
/home/cloudera/Desktop
```

2. Aller sur le dossier exemples/apps/java-main :

Les deux fichiers communs d'Oozie est trouvé « job.properties » et « workflow.xml »



3. Changer le fichier Job.Properties:

```
nameNode=hdfs://quickstart.cloudera:8020  
jobTracker=quickstart.cloudera:8032
```

A.Exercices suite au TP :

- 1) Mettre les exemples dans HDFS.
- 2) Vérifier que les fichiers sont bien transférés.
- 3) La port Oozie sur cloudera est 11000, lancer le job
« /home/cloudera/Desktop/examples/apps/java-main/job.properties » via Oozie.
- 4) Vérifier le statut de Job Oozie.
- 5) Chercher le Job Oozie sur cloudera Manager.
- 6) Lancer le job mapreduce de wordcount via oozie.