

09/02/2016 retard  
Modification le 2013-02-12

# Installation d'un cluster avec Cloudera Manager

Programme : A.E.C. BigData et BI

Cours : 420-BD3-BB gr.000214

Instructeur : M. Hafed Benteftifa

Travaux Pratique 1

## Table des matières

1. Introduction .....	3
2. Caractéristiques du cluster.....	5
3. Étapes principales d'installation .....	6
6. Comment tester l'application.....	11
7. Conclusion .....	12

# 1. Introduction

L'objectif de ce TP est d'installer un cluster avec une distribution hadoop grace au principe de virtualisation par conteneur.

Le namenode/datanode sera une version linux distribution Ubuntu version 14.04 LTS s'exécutant comme une machine virtuelle sous un système d'exploitation hôte Windows ayant comme configuration :

- Processeur : intel I7 8 cœurs à 3.4 Ghtz.
- RAM : 16 GO.
- Windows 7 Entreprise.
- Virtual Box version 5.
- Configuration d'un seconde carte ethernet avec une adresse IP Statique.

Ensuite, nous installerons un serveur de virtualisation légère comme LXC pour créer nous datanode.

Les datanodes seront des conteneurs avec une configuration comme ce qui suit :

- Un conteneur Ubuntu.
- La configuration d'une adresse IP Statique.
- L'installation du JDK version 7 ou +.

La distribution Hadoop sera une version de Cloudera Manager version 5 avec au minimum :

- HDFS.
- MapReduce v1.
- YARN (MapReduce v2).
- Oozie
- HUE
- Cloudera management service
- Zookeeper
- HIVE

Une fois que notre cluster sera « up and running », nous allons lancer une tâche YARN et consulter les paramètres d'exécution et le résultat.

cdh5-0 [En fonction] - Oracle VM VirtualBox

Fichier Machine Écran Entrée Périphériques Aide

Oracle VM VirtualBox - Gestionnaire de machines

Fichier Machine Aide

Nouvelle Configuration Oublier Afficher

Détails Instantanés

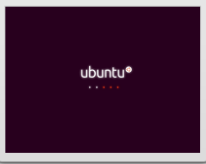
cdh5-0\_TEMPLATE  
Éteinte

cdh5-0  
En fonction

**Général**  
Nom : cdh5-0  
Système d'exploitation : Ubuntu (64-bit)

**Système**  
Mémoire vive : 8192 Mo  
Processeurs : 4  
Ordre d'amorçage : Disquette, Optique, Disque dur  
Accélération : VT-x/AMD-V, Pagination imbriquée, Paravirtualisation KVM

**Prévisualisation**



**Affichage**  
Mémoire vidéo : 60 Mo  
Serveur bureau distant : Désactivé  
Capture vidéo : Désactivée

**Stockage**  
Contrôleur : IDE  
Maître secondaire IDE : [Lecteur optique] VBoxGuestAdditions.iso (56,28 Mo)  
Contrôleur : SATA  
Port SATA 0 : cdh5-0-disk1.vdi (Normal, 256,00 Gio)

**Son**  
Pilote hôte : Windows DirectSound  
Contrôleur : ICH AC97

**Réseau**  
Interface 1 : Intel PRO/1000 MT Desktop (NAT)  
Interface 2 : Intel PRO/1000 MT Desktop (Réseau privé hôte, VirtualBox Host-Only Ethernet Adapter)

**USB**  
Contrôleur USB: OHCI, EHCI  
Filtres : 0 (0 actif)

**Dossiers partagés**  
Aucun

**Description**  
Aucune

CTRL DROITE

12:44  
2016-02-12

## 2. Caractéristiques du cluster

### NAMÉ NODE

CDHS-3  
• UBUNTU 14.04  
• jdk 7 ou +  
• IP statique  
• 192.168.56.102

LXC *server*

CLOUDERA MANAGER v5

SSH

CDHS-1

• UBUNTU  
• jdk 7 ou +  
• IP statique  
• 10.0.3.111

COMPT GENER

CDHS-2

• UBUNTU  
• jdk 7 ou +  
• IP statique  
• 10.0.3.112

COMPT GENER

CDHS-3

• UBUNTU  
• jdk 7 ou +  
• IP statique  
• 10.0.3.113

COMPT GENER

### DATA NODE

### 3. Étapes principales d'installation

#### 1. Préparation de la machine virtuelle UBUNTU avec virtual box.

- Installer une version de Ubuntu 14.04 LTS avec une second carte ethernet
- Installer les composants additionnels de virtual Box
- Activer le root
- Installer l'éditeur VIM
- Configurer une adresse IP statique par exemple 192.168.56.102
- Configurer les fichier /etc/hosts et /etc/hostname
- Faire un point de sauvegarde en clanant la machine

#### 2. Configurer les pré-requis pour le namenode

- Fixer le SELinux = disabled (/etc/default/config)
- Desactiver le firewall (sudo service ufw stop)
- Fixer la valeur VM.swappiness = 0 (sudo sysctl vm.swappiness=0)
- Désactiver le IPV6 (/etc/sysctl.conf)
- Installer, configurer, et tester le SSH sur le namenode

#### 3. Installer le serveur LXC et configurer le mini-dhcp de lxc

- Installer le serveur LXC (sudo apt-get install lxc)
- Vérifier que le serveur LXC « Up and Running » (sudo lxc-checkconfig)
- Définir les adresses IP Statiques pour les datanodes
  - Dans le fichier /etc/default/lxc-net décommenter la valeur LXC\_DHCP\_CONFIGFILE=/etc/lxc/dnsmasq.conf
  - Editer le fichier /etc/lxc/dnsmasq.conf avec les valeurs suivantes :

---

```
dhcp-host=cdh5-1,10.0.3.111
```

```
dhcp-host=cdh5-2,10.0.3.112
```

```
dhcp-host=cdh5-3,10.0.3.113
```

---

- Éditer le fichier /etc/hosts avec les valeurs suivantes :

---

```
192.168.56.102 cdh5-0
```

```
10.0.3.111 cdh5-1
```

```
10.0.3.112 cdh5-2
```

```
10.0.3.113 cdh5-3
```

---

#### 4. Créer nos datanode et les préparer pour cloudera

- Créer le conteneur (sudo lxc-create -t ubuntu -n cdh5-1)
- Démarrer, entrer dans le conteneur, et configurer le mot de passe root
  - i. sudo lxc-start -d -n cdh5.1
  - ii. sudo lxc-attach -n cdh5.1
  - iii. passwd root
- Éditer le fichier /etc/hosts avec les valeurs suivantes :

---

```
192.168.56.102 cdh5-0
10.0.3.111 cdh5-1
10.0.3.112 cdh5-2
10.0.3.113 cdh5-3
```

---

- Ajouter JVM 7 ou +
- Ajouter un composant pour cloudera (apt-get install pt-transport-https)
- Installer, configurer, et tester le SSH
- Quitter le datanode, le stopper et le clone.

---

```
exit
sudo lxc-stop -n cdh5-1
sudo lxc-clone -o cdh5-1 -n cdh5-2
sudo lxc-clone -o cdh5-1 -n cdh5-3
```

---

#### 5. Créer nos datanode et les préparer pour cloudera

- Se positionner en root
- Démarrer les datanodes
- Télécharger et installer le package cloudera manager 5

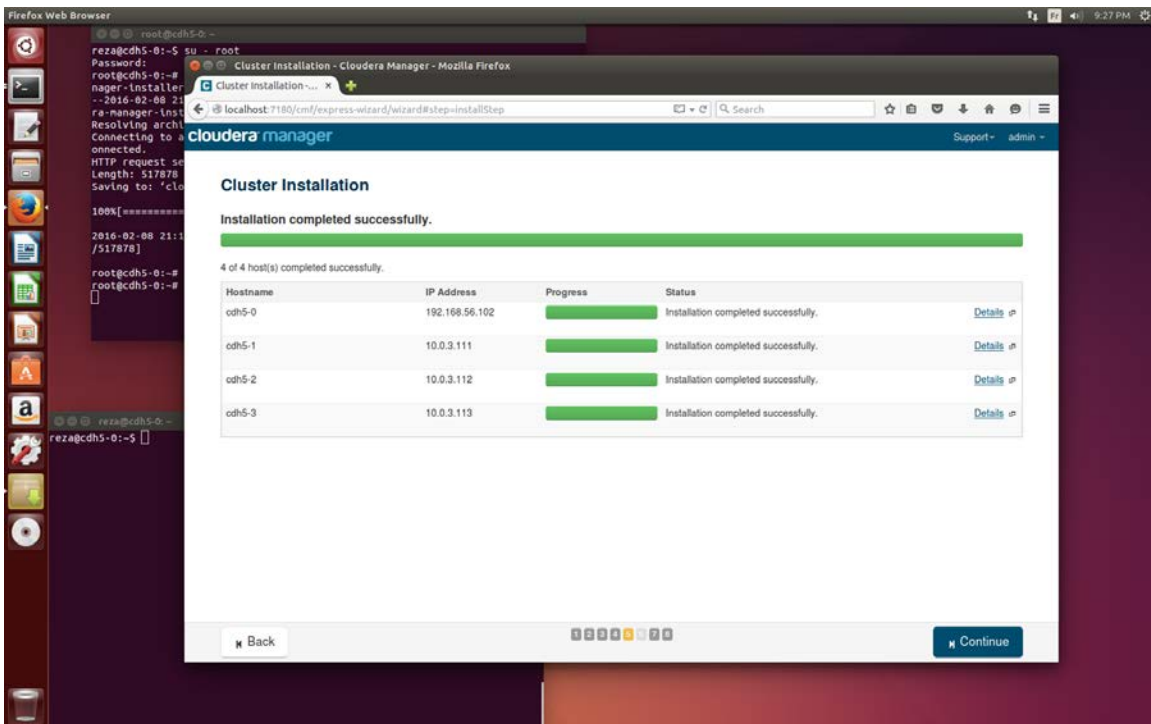
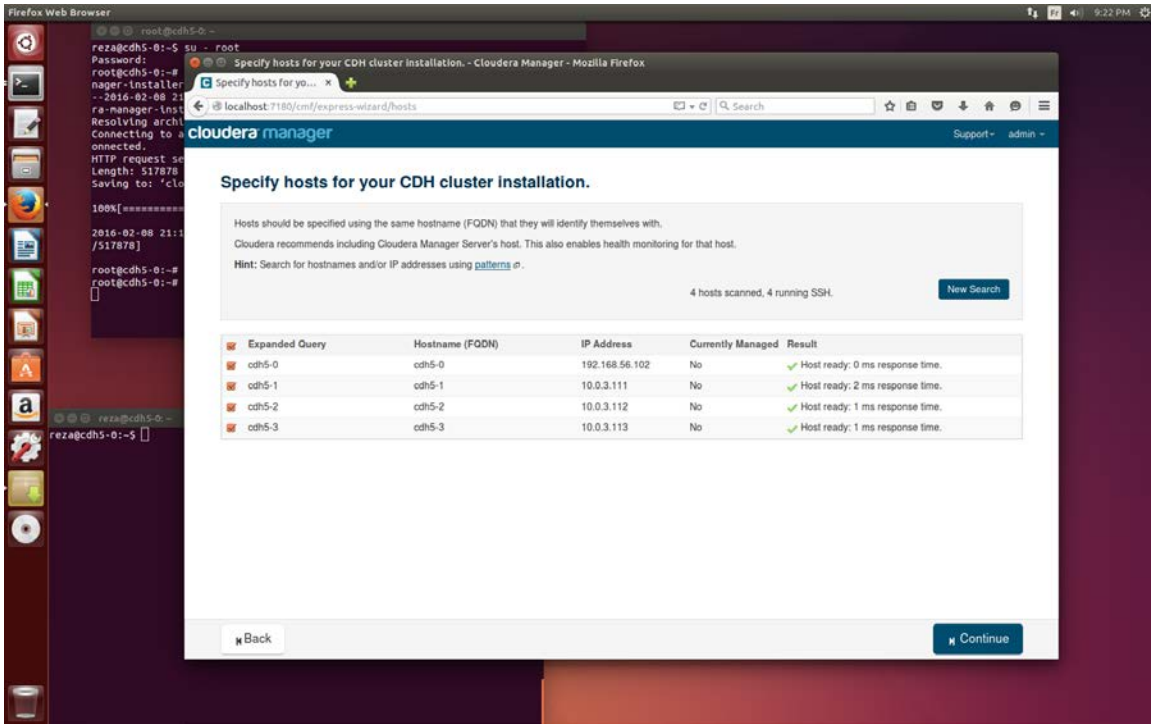
---

```
Wget http://archive.cloudera.com/cm5/installer/latest/cloudera-manager-installer.bin
```

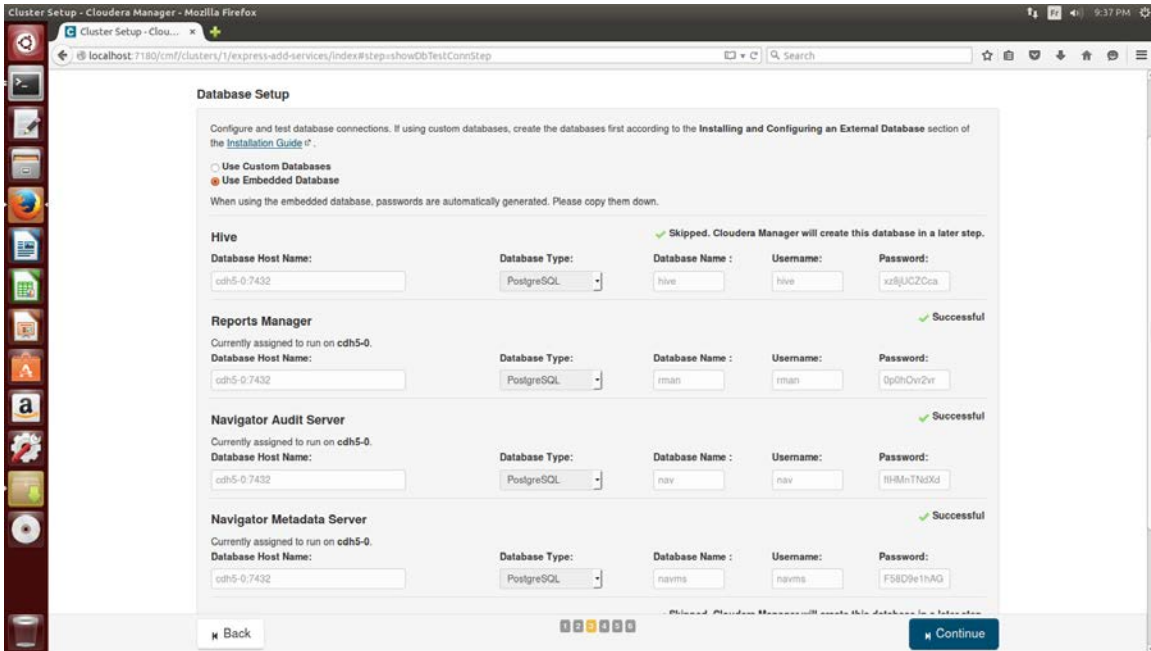
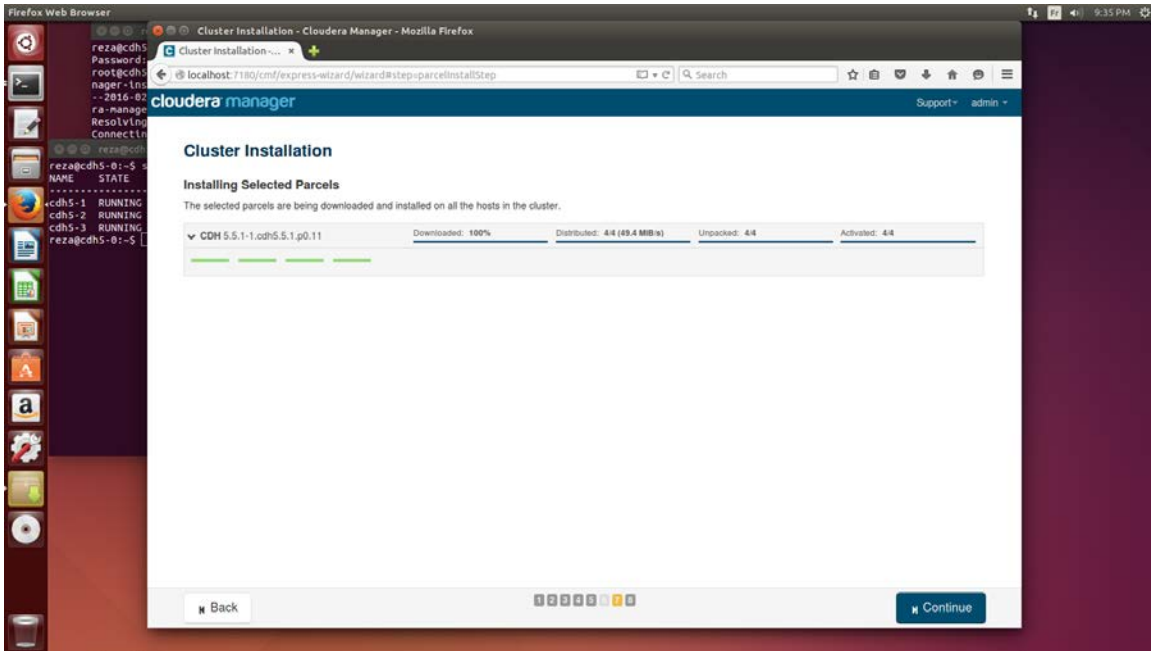
```
chmod u+x cloudera-manager-installer.bin
```

```
./cloudera-manager-installer.bin
```

---







**Cluster Setup**

✓ **First Run Command**

Status: **Finished** Start Time: Feb 8, 9:37:26 PM Duration: 6.8m

Finished First Run of all services successfully.

**Details** Completed 7 of 7 steps.

Step	Context	Start Time	Duration	Actions
✓ <b>Deploy Client Configuration</b> Successfully deployed all client configurations..	Cluster 1	Feb 8, 9:37:26 PM	15.54s	
✓ <b>Start Cloudera Management Service, ZooKeeper</b> Successfully completed 2 steps.		Feb 8, 9:37:42 PM	26.54s	
✓ <b>Start HDFS</b> Successfully completed 1 steps.		Feb 8, 9:38:08 PM	71.24s	
✓ <b>Start YARN (MR2 Included)</b> Successfully completed 1 steps.		Feb 8, 9:39:20 PM	95.72s	
✓ <b>Start Hive</b> Successfully completed 1 steps.		Feb 8, 9:40:56 PM	75.18s	
✓ <b>Start Oozie</b> Successfully completed 1 steps.		Feb 8, 9:42:11 PM	101.06s	
✓ <b>Start Hue</b> Successfully completed 1 steps.		Feb 8, 9:43:52 PM	22.3s	

Buttons: Back, Continue

**Cluster Setup**

**Congratulations!**

The services are installed, configured, and running on your cluster.

```

reza@cdh5:~$ ssh root@cdh5-1
root@cdh5-1:~# ssh root@cdh5-2
root@cdh5-2:~# ssh root@cdh5-3
root@cdh5-3:~# ssh root@cdh5-0
root@cdh5-0:~$

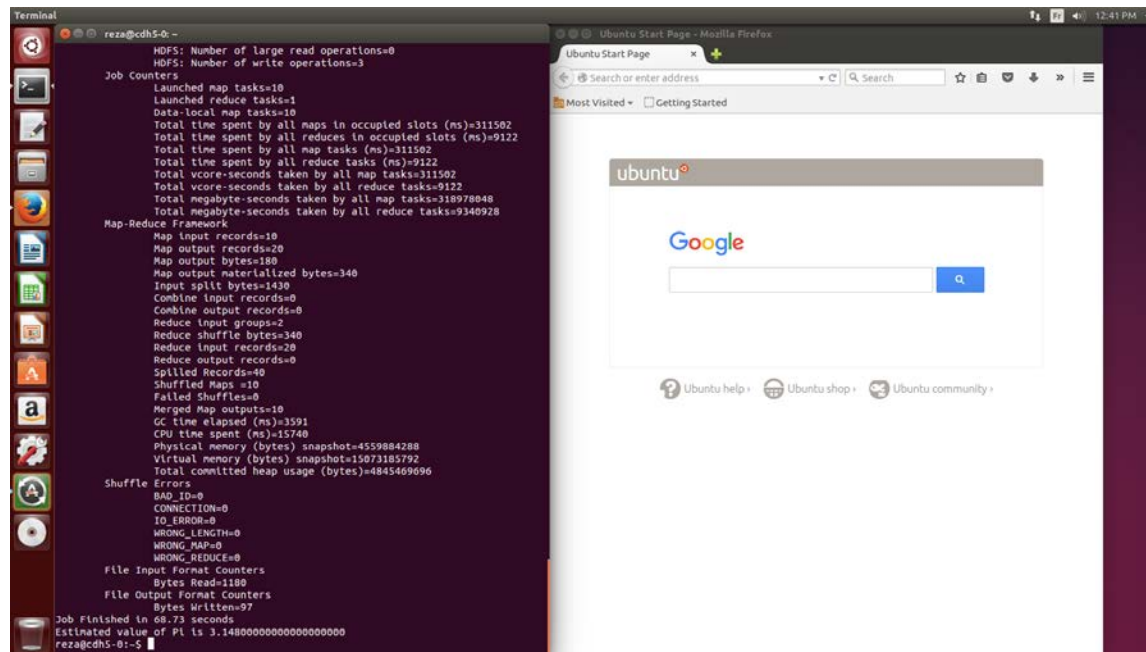
reza@cdh5:~$ cat /etc/passwd | grep cdh5
cdh5-1 RUNNING
cdh5-2 RUNNING
cdh5-3 RUNNING
reza@cdh5-0:~$

```

Buttons: Back, Finish

## 6. Comment tester l'application

reza@cdh5-0:~\$ sudo -u hdfs hadoop jar /opt/cloudera/parcels/CDH/lib/hadoop-mapreduce/hadoop-mapreduce-examples.jar pi 10 100



## 7. Conclusion

La réelle difficulté n'était pas l'installation et la configuration du cluster mais plutôt les points suivants:

- Trouver une machine hôte assez puissante
- Et d'apprivoiser le serveur LXC

J'ai du recommencer à plusieurs reprises mon installation parce qu'il y avait une erreur de timeout pendant l'installation de cloudera. Mais une fois que j'ai trouvé une machine très performante, l'installation de cloudera ne m'a pris que 40 min sur 4 nœuds (dont 3 conteneurs).