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

Reza Naidji COLLÈGE DE BOIS-DE-BOULOGNE

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 minimun :

- 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	Crade VM VirtualPere Costionnaire de machin			
Fichier Machine Écran Entrée Périphériques Aide	Fichier Machine Aide	162		
	Nouvelle Configuration Oublier Afficher		设 Détails 💿 Instantanés	
	cdh5-0_TEMPLATE	📃 Général	Prévisualisation	
	Eteinte	Nom : cdh5-0		
	cdh5-0 ⇒ En fonction	Systeme d'exploitation : Ubuntu (64-bit)		
		Mémoire vive : 8192 Mo Processeurs : 4 Ordre d'amorçage : Disquette, Optique, Disque dur Accélération : VT=XAND-V , Pagination imbriquée, Paravirtualisation KVM	ubuntu*	
		I Affichage		
		Mémoire vidéo : 60 Mo Serveur bureau distant : Désactivé Capture vidéo: Désactivée		
		Stockage		
		Contrôleur : IDE Maître secondaire IDE : [Lecteur optique] VBoxGuestA Contrôleur : SATA Port SATA 0 : cdh5-0-disk1.vdi (Normal, 25/	Additions.iso (56,28 Mio) 5,00 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	é hôte, 'VirtualBox Host-Only Ethernet Adapter')	
		🖉 USB		
		Contrôleur USB: OHCI, EHCI Filtres : 0 (0 actif)		
		Dossiers partagés		
		Aucun		
		Description		
		Aucune		
				📴 💿 🖶 🖉 🚍 📇 🔟 🚳 💽 CTRL DROITE
🚳 🥝 📜 🖸 🚺 🛓 💵				🛓 🌮 💿 📴 N 🍓 🍽 🛱 🌵 12:44 2016-02-12

2. Caractéristiques du cluster



DATANODE

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 = desabled (/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
 - o 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 installa 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
- Demarrer les datanodes
- Telecharger et installer le package cloudera manager 5

Wget<u>http://archive.cloudera.com/cm5/installer/</u> latest/cloudera-manager-installer.bin

chmod u+x cloudera-manager-installer.bin

./cloudera-manager-installer.bin



Firefox Web	Browser										Fr Fr	·•• 9:27
	reza@cdhS-0:-5 su Password: root@cdhS-0:-# nager-installer -2016-02-08 21 ra-manager-inst Resolving archi	Co - y - root Cluster installation - Cloo Cluster installation × Cluster installation × Cluster installation ×	udera Manager - Mozilla Firefox wizard/wizard#step=installStep		El ▼ C) Q. Search	☆	6	a †	Ĥ	0	=	
	Connecting to a C onnected. HTTP request se Length: 517878 Saving to: 'clo 100%[Cloudera manager Cluster Installatio	n successfully.					Sug	port+	admi	£.₩.	
	root@cdh5-0:-# root@cdh5-0:-#	4 of 4 host(s) completed succes Hostname cdh5-0	IP Address 192.168.56.102	Progress	Status Installation completed successfully.				Detail	P		
		cdh5-1 cdh5-2	10.0.3.111		Installation completed successfully.				Detail	0		
a v	© © reza@cdh5-0:- tagcdh5-0:-\$ []	cdh5-3	10.0.3.113		Installation completed successfully.				Detail	0		
•												
		M Back		0000		-		H C	ontinu	e		



Setup - Clouder	a Manager - Mozilla Firefox tup - Clou 🗙 💠		_	_			14	Fr	0) 1	8:37	
🔶 🖲 localhos	at 7180/cm//clusters/1/express-add-services/index#step=showDbTe	estConnStep	C) •	C Search	1	1	۵	+	ŧ	ę	
	Database Setup				Configuring an External Database section of Anager will create this database in a later step. Username: Password: Poe voluce20ca Successful Username: Password: rman 0 0p0DrOv/2r Successful Username: Password: rman						
	Configure and test database connections. If using of the <u>Installation Guide</u> 0° . ○ Use Custom Databases ● Use Embedded Database When using the embedded database, passwords an	ustom databases, create the databases first e automatically generated. Please copy the	according to the Installing a	and Configuring an E	ixternal Database section of						
	Hive		🗸 Skipped. Clouder	ra Manager will create	e this database in a later step.		t [2] ≪).				
	Database Host Name:	Database Type:	Database Name :	Usemame:	Password:						
	cdh5-0:7432	PostgreSQL -	hive	hive	xzłjUCZCca						
	Reports Manager				🥪 Successful	tep. sstul sstul					
	Currently assigned to run on cdh5-0. Database Host Name:	Database Type:	Database Name :	Username:	Password:						
	odh5-0:7432	PostgreSQL -	iman	iman	0p0hOvr2vr						
	Navigator Audit Server				g an External Database section of g an External Database section of I create this database in a later step. e: Password: @ Successful e: Password: @ Succesful e: Password: @ Succes						
	Currently assigned to run on cdh5-0. Database Host Name:	Database Type:	Database Name :	Username:	Password:						
	cdh5-0.7432	PostgreSQL -	nay	nav	ftHMnTNdXd				2 4000		
	Navigator Metadata Server				🖌 Successful						
	Currently assigned to run on cdh5-0.	Patabase Torres	Detabase Norma	Uniment	Bernard						
	refb5.0.7432	PostoreSOI +	Database Name :	Osemane.	E58D9e1bAQ						
		, mgrowe	(mine)	(international data in the second se	[]						
		00000	- Phinand Planda	•• H-••••••••••••••••••••••••••••••••••	. Ikia datahasa in a latas atau						
	N Back				H Continue						

Cluster	Setup - Cloudera Manager - Mozilla Firefox			1	4 12	40	9:44 Pł	мф
0	♦ @ localhost.7180/cml/clusters/1/express-add-services/index#step=commandDetailsStep □ + C Q, search	i.	û	ė (t ń	ø	≡
>_	cloudera manager				s	upport-	adm	in-
	Cluster Setup Image: First Run Command Status: Finished Start Time: Feb 8, 9.3728 PM Duration: 6.8m Finished First Run of all services successfully. Details Context Start Time > Upply: Glent Configuration Context Start Time > V Start Cloudent Management Service, Zookeeper Successfully completed 1 steps. Feb 8, 9.372 > V Start Cloudent Management Service, Zookeeper Successfully completed 1 steps. Feb 8, 9.372 > V Start MANN (MEZ Included) Successfully completed 1 steps. Feb 8, 9.382 > V Start MANN (MEZ Included) Successfully completed 1 steps. Feb 8, 9.405 > V Start Hee Successfully completed 1 steps. Feb 8, 9.405 > V Start Hee Successfully completed 1 steps. Feb 8, 9.405 > V Start Hee Successfully completed 1 steps. Feb 8, 9.425 > V Start Hee Successfully completed 1 steps. Feb 8, 9.435	Al C Failed Ord Duration 56 PM 15.545 58 PM 2 PM 26.545 8 PM 71.24s 80 PM 95.72s 86 PM 75.18s 11 PM 101.06s 12 PM 22.3s	r ⊖ Running On Actions	y				
٦	N Back		H Continue					



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

			and the second
😻 🗇 💷 reza@cdh5-0: -	🔘 🔘 🙂 Ubuntu Start Page - Mozilia Firefox		
HDFS: Number of large read operations=0	Ubuntu Start Page 🛛 🗙 🙀		
HDFS: NUMBER OF WRIte operations=3			1
Launched man tacks=18	Search or enter address VC A Search Q II V	• » =	
Launched reduce tasks=1	Mort Michael + Cotting Started		
Data-local map tasks=10	Manak Andre Control Mantee		
Total time spent by all maps in occupied slots (ms)=311502			
Total time spent by all reduces in occupied slots (ms)=9122			
Total time spent by all map tasks (ms)=311502			
Total time spent by all reduce tasks (ns)=9122	ubush 9		
Total vcore-seconds taken by all map tasks=311502	ubuncu		
Total vcore-seconds taken by all reduce tasks=9122			
Total megabyte-seconds taken by all reduce tasks=916976048			
Han-Deduce Framework			
Map input records=10	Canala		
Map output records=20	Google		
Map output bytes=180			
Map output materialized bytes=340			
Input split bytes=1430			
Combine input records=0			
Combine output records=8			
Reduce input groups=2			
Reduce Shull be by tes=340			
Reduce output records 8			
Spilled Records=40			
Shuffled Maps =10	O Uhushu hala . O Uhushu shaa . O Uhushu commusiliya		
Failed Shuffles=0	Gound anopy Gound anopy Gound community		
Merged Map outputs=10			
GC time elapsed (ms)=3591			
CPU time spent (ms)=15740			
Physical memory (bytes) snapshot=4559884288			
Total completed base urace (buter)=4845468686			
Shuffle Errors			
BAD ID=0			
CONNECTION=0			
IO_ERROR=0			
WRONG_LENGTH=0			
WRONG_MAP=0			
WRONG_REDUCE=0			
File Input Format Counters			
Bytes Read=1180			
Butes Written-07			
Job Finished in 68.73 seconds			
Estimated value of PL 15 3.14800000000000000000		/	
			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).