**Installation Hadoop Pseudo-distributed Mode**

1-Installations prérequis:

Java est la condition principale pour Hadoop. Tout d'abord, vérifiez l'existence de java dans votre système en utilisant la commande "java -version".

$ java -version

Si tout est en ordre, il vous donnera la sortie suivante.

java version "1.7.0\_71"

Java(TM) SE Runtime Environment (build 1.7.0\_71-b13)

Java HotSpot(TM) Client VM (build 25.0-b02, mixed mode)

Si java n'est pas installé dans votre système, suivez les étapes ci- dessous pour installer java.

Étape 1

Téléchargez java (JDK <version la plus récente> - X64.tar.gz) en visitant le lien suivant: http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html

Puis jdk-8u121-linux-x64.tar.gz sera téléchargé dans votre système.

### Étape 2

En général, vous trouverez le fichier java téléchargé dans le dossier Téléchargements. Vérifier et extraire le fichier jdk-7u71-linux-x64.gz utilisant les commandes suivantes.

$ cd Downloads/

$ ls

jdk-8u121-linux-x64.tar.gz

$ tar zxf jdk-8u121-linux-x64.tar.gz

$ ls

jdk-8u121-linux-x64 jdk-8u121-linux-x64.tar.gz

### Étape 3

Pour java disponible à tous les utilisateurs, vous devez le déplacer vers l'emplacement « / usr / local / ». Ouvrez la racine, et tapez les commandes suivantes.

$ su

password:

# mv jdk-8u121-linux-x64 /usr/local/

# exit

### Étape 4

Pour la mise en place des variables PATH et JAVA\_HOME, ajoutez les commandes suivantes à ~ / .bashrc.

export JAVA\_HOME=/usr/local/jdk-8u121-linux-x64

export PATH=$PATH:$JAVA\_HOME/bin

Maintenant, appliquez tous les changements dans le système de fonctionnement actuel.

$ source ~/.bashrc

Vérifier maintenant la commande java -version à partir du terminal comme expliqué ci-dessus.

## 2-Téléchargement Hadoop

Téléchargez et décompressez Hadoop 2.4.1 de la fondation du logiciel Apache en utilisant les commandes suivantes.

$ su

password:

# cd /usr/local

# wget http://apache.claz.org/hadoop/common/hadoop-2.4.1/

hadoop-2.4.1.tar.gz

# tar xzf hadoop-2.4.1.tar.gz

# mv hadoop-2.4.1/\* to hadoop/

# exit

## 3-Hadoop Modes de fonctionnement

Une fois que vous avez téléchargé Hadoop, vous pouvez utiliser votre cluster Hadoop dans l'un des trois modes sont supportés:

* **Mode local / autonome** : Après avoir téléchargé Hadoop dans votre système, par défaut, il est configuré en mode autonome et peut être exécuté comme un processus unique java.
* **Pseudo mode Distribuée** : Il est une simulation distribuée sur une seule machine. Chaque daemon Hadoop HDFS tels que, le fil, MapReduce etc., se déroulera comme un processus java séparé. Ce mode est utile pour le développement.
* **Entièrement mode distribué** : Ce mode est entièrement distribué avec un minimum de deux ou plusieurs machines comme un cluster.

## 4-Installation Hadoop en mode distribué Pseudo

Suivez les étapes ci-dessous pour installer Hadoop 2.4.1 en mode pseudo distribué.

### Étape 1: Configuration de Hadoop

Vous pouvez définir des variables d'environnement Hadoop en ajoutant les commandes suivantes pour **~ / .bashrc** fichier.

export HADOOP\_HOME=/usr/local/hadoop

export HADOOP\_MAPRED\_HOME=$HADOOP\_HOME

export HADOOP\_COMMON\_HOME=$HADOOP\_HOME

export HADOOP\_HDFS\_HOME=$HADOOP\_HOME

export YARN\_HOME=$HADOOP\_HOME

export HADOOP\_COMMON\_LIB\_NATIVE\_DIR=$HADOOP\_HOME/lib/native

export PATH=$PATH:$HADOOP\_HOME/sbin:$HADOOP\_HOME/bin

export HADOOP\_INSTALL=$HADOOP\_HOME

Maintenant, appliquez tous les changements dans le système de fonctionnement actuel.

$ source ~/.bashrc

### Étape 2: Configuration Hadoop

Vous pouvez trouver tous les fichiers de configuration Hadoop dans l'emplacement « $ HADOOP\_HOME / etc / Hadoop ». Il est nécessaire d'apporter des modifications dans les fichiers de configuration en fonction de votre infrastructure Hadoop.

$ cd $HADOOP\_HOME/etc/hadoop

Afin de développer des programmes en Java Hadoop, vous devez réinitialiser les variables d'environnement java dans **hadoop-env.sh** fichier en remplaçant **JAVA\_HOME** valeur à l'emplacement de Java dans votre système.

export JAVA\_HOME=/usr/local/jdk1.7.0\_71

Les éléments suivants sont la liste des fichiers que vous devez modifier pour configurer Hadoop.

**core-site.xml**

Le **core-site.xml** fichier contient des informations telles que le numéro de port utilisé par exemple Hadoop, la mémoire allouée pour le système de fichiers, la limite de mémoire pour stocker les données, et la taille des tampons de lecture / écriture.

Ouvrez le noyau site.xml et ajoutez les propriétés suivantes entre <configuration>, </ configuration> balises.

<configuration>

<property>

<name>fs.default.name </name>

<value> hdfs://localhost:9000 </value>

</property>

</configuration>

**hdfs-site.xml**

Le **hdfs-site.xml** fichier contient des informations telles que la valeur des données de réplication, chemin de NameNode et DataNode chemins de vos systèmes de fichiers locaux. Cela signifie que l'endroit où vous souhaitez stocker l'infrastructure Hadoop.

Supposons les données suivantes.

dfs.replication (data replication value) = 1

(In the below given path /hadoop/ is the user name.

hadoopinfra/hdfs/namenode is the directory created by hdfs file system.)

namenode path = //home/hadoop/hadoopinfra/hdfs/namenode

(hadoopinfra/hdfs/datanode is the directory created by hdfs file system.)

datanode path = //home/hadoop/hadoopinfra/hdfs/datanode

Ouvrez ce fichier et ajoutez les propriétés suivantes entre la <configuration> </ configuration> balises dans ce fichier.

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

<name>dfs.name.dir</name>

<value>file:///home/hadoop/hadoopinfra/hdfs/namenode </value>

</property>

<property>

<name>dfs.data.dir</name>

<value>file:///home/hadoop/hadoopinfra/hdfs/datanode </value>

</property>

</configuration>

**Remarque:** Dans le fichier ci - dessus, toutes les valeurs de propriété sont définies par l' utilisateur et vous pouvez apporter des modifications en fonction de votre infrastructure Hadoop.

**yarn-site.xml**

Ce fichier est utilisé pour configurer le fil dans Hadoop. Ouvrez le fichier yarn-site.xml et ajoutez les propriétés suivantes entre la <configuration>, </ configuration> balises dans ce fichier.

<configuration>

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

</configuration>

**mapred-site.xml**

Ce fichier est utilisé pour spécifier quel cadre MapReduce que nous utilisons. Par défaut, Hadoop contient un modèle de fil-site.xml. Tout d' abord, il est nécessaire de copier le fichier à partir **mapred site, xml.template** à **mapred-site.xml** fichier en utilisant la commande suivante.

$ cp mapred-site.xml.template mapred-site.xml

Ouvrez le fichier mapred-site.xml et ajoutez les propriétés suivantes entre la <configuration>, </ configuration> balises dans ce fichier.

<configuration>

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

</configuration>

## 5-Vérification de l'installation Hadoop

Les étapes suivantes sont utilisées pour vérifier l'installation Hadoop.

### Étape 1: Configuration du nœud Nom

Mettre en place la NameNode en utilisant la commande « hdfs NameNode -format » comme suit.

$ cd ~

$ hdfs namenode -format

Le résultat attendu est la suivante.

10/24/14 21:30:55 INFO namenode.NameNode: STARTUP\_MSG:

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

STARTUP\_MSG: Starting NameNode

STARTUP\_MSG: host = localhost/192.168.1.11

STARTUP\_MSG: args = [-format]

STARTUP\_MSG: version = 2.4.1

...

...

10/24/14 21:30:56 INFO common.Storage: Storage directory

/home/hadoop/hadoopinfra/hdfs/namenode has been successfully formatted.

10/24/14 21:30:56 INFO namenode.NNStorageRetentionManager: Going to

retain 1 images with txid >= 0

10/24/14 21:30:56 INFO util.ExitUtil: Exiting with status 0

10/24/14 21:30:56 INFO namenode.NameNode: SHUTDOWN\_MSG:

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

SHUTDOWN\_MSG: Shutting down NameNode at localhost/192.168.1.11

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

### Étape 2: Vérification DSF Hadoop

La commande suivante est utilisée pour démarrer DSF. L'exécution de cette commande lancera votre système de fichiers Hadoop.

$ start-dfs.sh

Le résultat attendu est la suivante:

10/24/14 21:37:56

Starting namenodes on [localhost]

localhost: starting namenode, logging to /home/hadoop/hadoop

2.4.1/logs/hadoop-hadoop-namenode-localhost.out

localhost: starting datanode, logging to /home/hadoop/hadoop

2.4.1/logs/hadoop-hadoop-datanode-localhost.out

Starting secondary namenodes [0.0.0.0]

### Étape 3: Vérification de script fil

La commande suivante est utilisée pour démarrer le script de fil. L'exécution de cette commande va commencer votre daemons de fil.

$ start-yarn.sh

Le résultat attendu de la manière suivante:

starting yarn daemons

starting resourcemanager, logging to /home/hadoop/hadoop

2.4.1/logs/yarn-hadoop-resourcemanager-localhost.out

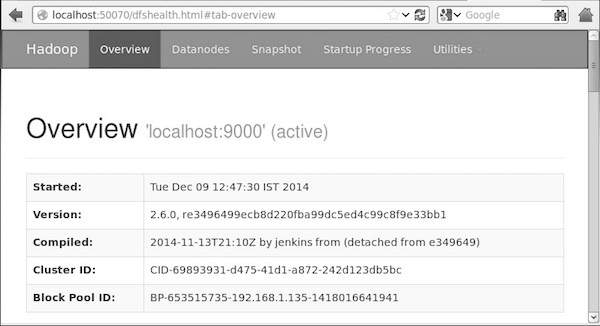
localhost: starting nodemanager, logging to /home/hadoop/hadoop

2.4.1/logs/yarn-hadoop-nodemanager-localhost.out

### Étape 4: Accès Hadoop sur navigateur

Le numéro de port par défaut pour accéder Hadoop est 50070. Utilisez l'adresse suivante pour obtenir des services Hadoop sur le navigateur.

http://localhost:50070/



### Étape 5: Vérification de toutes les applications de cluster

Le numéro de port par défaut pour accéder à toutes les applications de cluster est 8088. Utilisez l'adresse suivante pour visiter ce service.

http://localhost:8088/

### Cluster Hadoop application