MASI Lab hadoop-hbase cluster setup Quick checklistA

# Warning

* DO NOT INCLUDE kyuss, mastadon and masigate as Hadoop cluster node.

# STOp all running hadoop, hbase and yarn service

## Stop Hadoop Distributed File System

ssh to master node as ‘hadoop’ user -> cd $HADOOP\_HOME-> sbin/stop-dfs.sh

## Stop Distributed HBase cluster

ssh to master node as ‘hadoop’ user -> cd $HBASE\_HOME-> sbin/stop-hbase.sh

## Stop YARN

ssh to master node as ‘hadoop’ user -> cd $HADOOP\_HOME-> sbin/stop-yarn.sh

## Check if service stop on all slaves

## Check if service stop on all master

* Check if port 2181 is occupied for all of machines. If not, please manually free the port according ‘MASI Lab Hadoop HBase cluster setup’ section 12.D

# Create hadoop user

* Login to a new server node as admin ‘masi’, create a new user ‘hadoop’ with password ‘hadoop’
* Grant sudo privileges to user ‘hadoop’
* Logout the new server, and login the new server as user ‘hadoop’

# Install vim

* Check if new server install VIM text editor or not
* If not, install VIM

# Edit hosts file.

* Edit /etc/hosts, comment two line of private host names for new server
* Edit /etc/hosts.local, comment two line of private host names for new server

# Setup passphraseless ssh

* Accept all incoming request by command ‘sudo iptables –P INPUT ACCEPT’
* Generate default public/private rsa key pair by command ‘ssh-keygen’, simply keep pressing Enter.
* Repeat public key of a new server to all existing cluster nodes.
* Similarly, login to all existing cluster nodes, copy their existence public key to new server.
* Test if you can do ssh as user ‘hadoop’ without password, if not, please refer to Section 4.D of documentation ‘MASI Lab Hadoop HBase cluster setup’

# Setup hadoop distributed cluster

* Login to master node ‘masi-10’ as ‘hadoop’ account, go into directory /usr/local/hadoop/etc/hadoop
* Edit file slaves, add a new line and input new server’s hostname/alias.
* Edit file yarn-site.xml. Change the value of ‘yarn.scheduler.capacity.node-locality-delay’ to the total number (masters+slaves) of new cluster

# Setup hbase distributed cluster

## Login to master node ‘masi-10’ as ‘hadoop’ account, go into directory /usr/local/hbase/conf

## Edit file hbase-site.xml. Edit the value of ‘hbase.zookeeper.quorum’, add new server’s hostname/alias.

## Edit file regionserves, add a new line and input new server’s hostname/alias.

# Deploy and install hadoop & hbase on a new server

## Login to master node ‘masi-10’ and prepare compressed Hadoop, HBase package.

## Secure copy (scp) the Hadoop and HBase package to new server

## Login to new server as account ‘hadoop’, extract Hadoop and HBase package to /usr/local

## Create folder for Zookeeper, mkdir ‘/usr/local/zookeeper/’ and mkdir ‘/usr/local/zookeeper/data’

## Check if folder ‘/drive2’ exists, if not create it.

## Grant ‘hadoop’ account permission to edit related folders.

#### 

**sudo chown hadoop:hadoop /drive2**

**sudo chown hadoop:hadoop /usr/local/hadoop**

**sudo chown hadoop:hadoop /usr/local/hbase**

**sudo chown hadoop:hadoop /usr/local/zookeeper**

# Setup yarn configruation

* Login to new server as account ‘hadoop’, go into directory /usr/local/hadoop/etc/hadoop
* Edit file yarn-site.xml. Edit value of ‘yarn.nodemanager.resource.cpu-vcores’ according to new server’s total cpu.
* Edit file yarn-site.xml. Edit value of ‘yarn.nodemanager.resource.memory-mb’ according to new server’s total memory can be allocated to YARN.
* Make sure your value setting meets for equation, where you can find value of ‘mapreduce.map.memroy.mb’ in file ‘mapred-site.xml’

# Start hadoop, hbase and yarn service

## Start Hadoop Distributed File System

ssh to master node as ‘hadoop’ user -> cd $HADOOP\_HOME-> sbin/start-dfs.sh

## Start Distributed HBase cluster

ssh to master node as ‘hadoop’ user -> cd $HBASE\_HOME-> sbin/start-hbase.sh

## Start YARN

ssh to master node as ‘hadoop’ user -> cd $HADOOP\_HOME-> sbin/start-yarn.sh

## Check if service start correctly on each slaves.

## Check if service start correctly on each masters.

# Maintain hadoop, hbase cluster

## Login to master node, using bash to check Hadoop HDFS

## Using bash to check HBase cluster

## Using bash to check YARN nodes