

Configuring a Hadoop multi-node test cluster

1. Log in to the slave node on Oracle cloud.
2. Do steps 1 to 15 (from the single node tutorial).
3. In Step 16, create only datanode directory since we want namenode to be running on the master node.
4. Do Step 17 as it is.
5. In step 18, do the same for all nodes in the cluster (both master and slaves).
 - a. Edit `hadoop-env.sh` as in single node set up.
 - b. Specify the namenode to be `hadoopmaster` in `core-site.xml`.
Change *localhost* to *hadoopmaster*.
 - c. Change the replication from 1 to 3 in `hdfs-site.xml`. Also, keep only the property for data node directory.
 - d. Add the following to `yarn-site.xml`

```
<property>
  <name>yarn.resourcemanager.resource-tracker.address</name>
  <value>hadoopmaster:8025</value>
</property>
<property>
  <name>yarn.resourcemanager.scheduler.address</name>
  <value>hadoopmaster:8030</value>
</property>
<property>
  <name>yarn.resourcemanager.address</name>
  <value>hadoopmaster:8050</value>
</property>
```

- e. Add the following to `mapred-site.xml`.

```
<property>
  <name>mapreduce.framework.name</name>
  <value>yarn</value>
</property>
<property>
  <name>mapred.job.tracker</name>
  <value>hadoopmaster:54311</value>
</property>
```

- f. Additionally, add the following to `/etc/hosts` file:

```
168.138.4.55 hadoopmaster
168.138.13.9 hadoopslave
```

- g. Update `/etc/hostname` file with the hostname of the machine, if needed.
(If `hostname` command on the terminal prints the name you want, then not needed.)
- 6. Reboot all nodes

```
hduser@hadoopmaster:~$ sudo reboot
hduser@hadoopslave:~$ sudo reboot
```

- 7. Make the following changes only to master node: (to specify master and worker nodes)
 - a. Add `hadoopmaster` to `/usr/local/hadoop/etc/Hadoop/master` file.
 - b. Add
`168.138.13.9 hadoopslave`
to `workers` file in the same location.
Note: If we want the namenode also to act as a datanode, please add namenode's details to the workers file.
 - c. For password less SSH communication, we need to copy the public key of master to `authorized_keys` file of all slaves.

```
hduser@hadoopmaster:~$ sudo ssh-copy-id -i ~/.ssh/id_rsa.pub
hduser@hadoopslave
```

- 8. Now, namenode needs to be formatted. (steps 19 and 20 of single node tutorial)
- 9. Then all daemons can be started.

```
hduser@hadoopmaster:~$ jps
2883 ResourceManager
2340 NameNode
2645 SecondaryNameNode
3178 Jps
hduser@hadoopslave:~$ jps
2340 NodeManager
2148 DataNode
2442 Jps
```

We have a hadoop cluster configured with one master node and one slave node. Repeat the above steps for each additional worker node to be added to the cluster.