Indian Institute Of Technology, Delhi

COL733: Cloud Computing Technology Fundamentals

Instructor: S. C. Gupta

Assignment 3: Fault Recovery in HDFS

Report

August 31, 2019

Submitted To:
S. C. Gupta
Professor
Computer Science Department

Submitted By: (Group 4)

Shantanu Verma	2016CS10373
Pradyumna Meena	2016CS10375
Manav Rao	2016CS10523
Shubham	2016CS10371

Index

1. Installation

Execute the following steps to install **Hadoop** on Baadal.

1.1. Installation of Java:

➔ Update and upgrade the OS.

```
sudo apt-get update && apt-get upgrade
```

➔ Install the default JDK.

```
sudo apt-get install default-jdk
```

1.2. Install and operate a hadoop file system with a new user account.

➔ Using the following command, add a Hadoop system user

```
sudo addgroup hadoop
```

```
sudo adduser --ingroup hadoop hadoopuser
```

```
sudo adduser hadoopuser sudo
```

1.3. Configure SSH:

➔ Hadoop needs first **SSH** access to handle nodes in a cluster. Log into the new user using following command.

```
su - hadoopuser
```

➔ Access the local machine via **SSH** using keys on the **master node.**

```
ssh-keygen -t rsa -P ""
cat  ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
```

1.4. Attach the IP addresses of the master and the slaves in Hadoop file system.

➔ Open the hosts file stored at **PATH:** /etc/hosts on all the nodes.

```
sudo vi /etc/hosts
```

➔ Add the following line to the file on all the nodes.

```
10.17.50.113        hadoop-master
10.17.6.80          hadoop-slave-1
10.17.6.36          hadoop-slave-2
10.17.5.93          hadoop-slave-3
```

1.5. Grant SSH access across VMs using SSH key from master VM.

➔ Copy SSH key from master to slave VMs.

```
ssh-copy-id -i ~/.ssh/id_rsa.pub hadoopuser@hadoop-master
ssh-copy-id -i ~/.ssh/id_rsa.pub hadoopuser@hadoop-slave-1
ssh-copy-id -i ~/.ssh/id_rsa.pub hadoopuser@hadoop-slave-2
ssh-copy-id -i ~/.ssh/id_rsa.pub hadoopuser@hadoop-slave-3
```

1.6. Download Hadoop on all the VM.

➔ Run following commands to download and install **Hadoop** on all the VMs.

```
scp ./Downloads/hadoop-1.2.0.tar.gz baadalvm@10.17.50.113:./
```

```
sudo mkdir /opt/hadoop
mv ./hadoop-1.2.0.tar.gz /opt/hadoop/
cd /opt/hadoop
tar -xzf hadoop-1.2.0.tar.gz
sudo mv hadoop-1.2.0 hadoop
```

1.7. Recursively change the possession of Hadoop folders to "*hadoopuser*"

➔ Run the following command in all the VMs.

```
sudo chown -R hadoopuser/opt/hadoop
```

1.8. Hadoop configuration.

➔ Add the following lines to "*core-site.xml*" stored at **PATH:** /opt/hadoop/hadoop/conf.

```
<configuration>
      <property>
           <name>fs.default.name</name>
           <value>hdfs://hadoop-master:9000/</value>
      </property>
      <property>
           <name>dfs.permissions</name>
           <value>false</value>
      </property>
</configuration>
```

➔ Add the following lines to "*hdfs-site.xml*" stored at **PATH:** /opt/hadoop/hadoop/conf.

```
<configuration>
      <property>
           <name>dfs.data.dir</name>
           <value>/opt/hadoop/dfs/name/data</value>
           <final>true</final>
      </property>
      <property>
           <name>dfs.name.dir</name>
           <value>/opt/hadoop/dfs/name</value>
           <final>true</final>
      </property>
      <property>
           <name>dfs.replication</name>
           <value>2</value>
      </property>
</configuration>
```

➔ Add the following lines to "*mapred-site.xml*" stored at **PATH:** /opt/hadoop/hadoop/conf on the master node.

```
<configuration>
      <property>
           <name>mapred.job.tracker</name>
           <value>hadoop-master:9001</value>
      </property>
</configuration>
```

➔ Add the following lines to "*hadoop-env.sh*" stored at **PATH:** /opt/hadoop/hadoop/conf.

```
export JAVA_HOME=/usr/lib/jvm/java-7-openjdk-amd64
export HADOOP_OPTS=-Djava.net.preferIPv4Stack=true
export HADOOP_CONF_DIR=/opt/hadoop/conf
```

➔ Add the following lines to file stored at **PATH:** /opt/hadoop/hadoop/conf/masters on the master node.

```
hadoop-master
```

➔ Add the following lines to file stored at **PATH:** /opt/hadoop/hadoop/conf/slaves on the master node.

```
hadoop-slave-1
hadoop-slave-2
hadoop-slave-3
```

➔ Format master node and start all nodes by running following commands on the master node.

```
/opt/hadoop/bin/hadoop  namenode  -format
/opt/hadoop/bin  start-all.sh
```