Introduction to Big Data

ECAP456

Dr. Rajni Bhalla

Associate Professor

Learning Outcomes

After this lecture, you will be able to

- Starting HDFS
- Creating User Account
- Configuring Key Based Login
- Configuring Hadoop on Master Server

Starting HDFS

```
hadoop namenode -format
```

Starting HDFS

```
start-dfs.sh
```

HDFS Operations to Read & Write the Data

- Execute almost all operations on Hadoop Distributed File Systems.
- You can execute various reading, writing operations .
- You can add access rights and browse the file system to get the cluster information.

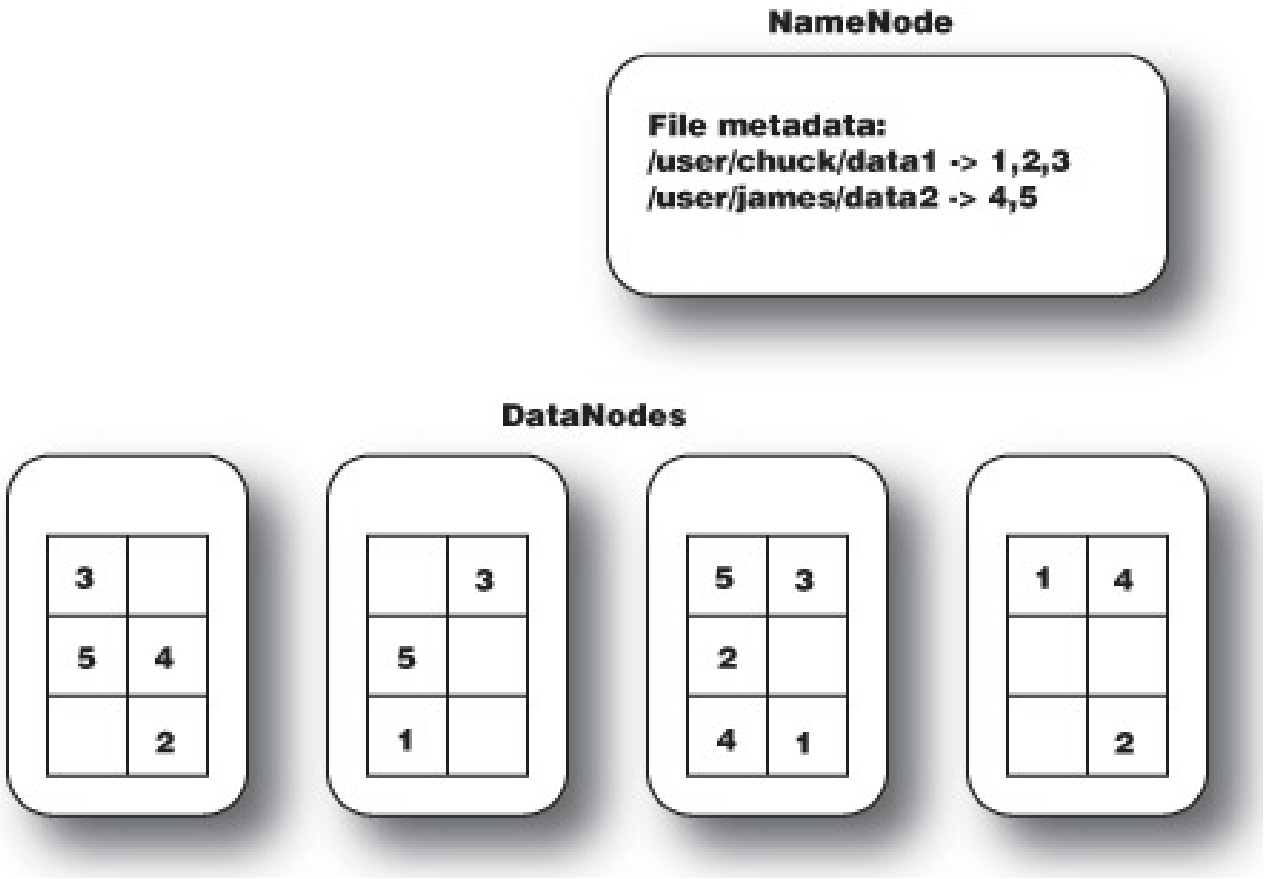
HDFS Operations to Read & Write the Data

- HDFS Operations to Read the file

NameNode

HDFS Operations to Read & Write the Data

- HDFS Operations to Read the file



HDFS Operations to Read & Write the Data

- HDFS Operations to Read the file

You can put a read request to NameNode for a particular block location through distributed file systems. The NameNode will then check your privilege to access the DataNode and allows you to read the address block if the access is valid.

```
bin/hdfs dfs -ls  <path>
```

HDFS Operations to Read & Write the Data

- HDFS Operations to Read the file

Similar to the read operation, the HDFS Write operation is used to write the file on a particular address through the NameNode.

```
bin/hdfs dfs -ls  <path>
```

Listing Files in HDFS

```
$HADOOP_HOME/bin/hadoop fs -ls <args>
```

Inserting Data into HDFS

Step1: $HADOOP_HOME/bin/hadoop fs -mkdir /user/input

Inserting Data into HDFS

Step1: $HADOOP_HOME/bin/hadoop fs -mkdir /user/input

Step2: $HADOOP_HOME/bin/hadoop fs -put /home/abc.txt /user/input

Inserting Data into HDFS

Step1: $HADOOP_HOME/bin/hadoop fs -mkdir /user/input

Step2: $HADOOP_HOME/bin/hadoop fs -put /home/abc.txt /user/input

Step3: $HADOOP_HOME/bin/hadoop fs -ls /user/input

Retrieving Data from HDFS

Step1: $HADOOP_HOME/bin/hadoop fs -cat /user/output

Retrieving Data from HDFS

Step1: $HADOOP_HOME/bin/hadoop fs -cat /user/output

Step2: $HADOOP_HOME/bin/hadoop fs -get

/user/output/ /home/hadoop_tp/

Shutting Down the HDFS

Shut down the HDFS files by following the below command

```
$ stop-dfs.sh
```

Shutting Down the HDFS

```
$ stop-dfs.sh
```

Multi-Node Cluster

- Installing Java
- Syntax of java version command

```
$ java -version
```

Creating User Account

System user account is used on both master and slave systems for the Hadoop installation.

Creating User Account

useradd hadoop

passwd hadoop

Creating User Account

- Mapping the nodes

vi /etc/hosts

Creating User Account

- Mapping the nodes

vi /etc/hosts

Enter the following lines in the /etc/hosts file.

192.168.1.109 hadoop-master
192.168.1.145 hadoop-slave-1
192.168.56.1 hadoop-slave-2

Configuring Key Based Login

```
# su hadoop

$ ssh-keygen -t rsa

$ ssh-copy-id -i ~/.ssh/id_rsa.pub tutorialspoint@hadoop-master

$ ssh-copy-id -i ~/.ssh/id_rsa.pub hadoop_tp1@hadoop-slave-1

$ ssh-copy-id -i ~/.ssh/id_rsa.pub hadoop_tp2@hadoop-slave-2

$ chmod 0600 ~/.ssh/authorized_keys

$ exit
```

Installation of Hadoop

```
# mkdir /opt/hadoop

# cd /opt/hadoop/

# wget http://apache.mesi.com.ar/hadoop/common/hadoop-1.2.1/hadoop-1.2.0.tar.gz

# tar -xzf hadoop-1.2.0.tar.gz

# mv hadoop-1.2.0 hadoop

# chown -R hadoop /opt/hadoop

# cd /opt/hadoop/hadoop/
```

Configuring Hadoop

- Hadoop server must be configured in core-site.xml and should be edited where ever required.

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

Configuring Hadoop

- Hadoop server must be configured in hdfs-site.xml and should be edited where ever required.

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

```
<property>
<name>dfs.name.dir</name>
<value>/opt/hadoop/hadoop/dfs/name</value>
<final>true</final>
</property>
<property>
<name>dfs.replication</name>
<value>1</value>
</property>
</configuration>
```

Configuring Hadoop

- Hadoop server must be configured in mapred-site.xml and should be edited where ever required.

```
<configuration>

<property>

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

</property>

</configuration>
```

Configuring Hadoop

JAVA_HOME, HADOOP_CONF_DIR, and HADOOP_OPTS should be edited as follows:

- export JAVA_HOME=/opt/jdk1.7.0_17
- export HADOOP_OPTS=-Djava.net.preferIPv4Stack=true
- export HADOOP_CONF_DIR=/opt/hadoop/hadoop/conf

Installing Hadoop on Slave Servers

```
# su hadoop

$ cd /opt/hadoop

$ scp -r hadoop hadoop-slave-1:/opt/hadoop

$ scp -r hadoop hadoop-slave-2:/opt/hadoop
```

Configuring Hadoop on Master Server

- Master server configuration

```
# su hadoop

$ cd /opt/hadoop/hadoop

Master Node Configuration

$ vi etc/hadoop/masters

hadoop-master
```

Configuring Hadoop on Master Server

- Slave Node Configuration

```
$ vi etc/hadoop/slaves

hadoop-slave-1

hadoop-slave-2
```

Configuring Hadoop on Master Server

- Name Node format on Hadoop Master

```
# su hadoop

$ cd /opt/hadoop/hadoop

$ bin/hadoop namenode –format
```

Hadoop Services

Starting Hadoop services on the Hadoop-Master procedure explains its setup.

```
$ cd $HADOOP_HOME/sbin

$ start-all.sh
```


That's all for now...