Prepear

ubuntu-14.04-server-amd64.iso
http://releases.ubuntu.com/14.04/

hadoop 2.7.3
https://archive.apache.org/dist/hadoop/common/hadoop-2.7.3/hadoop-2.7.3.tar.gz

hbase 1.3.0
https://archive.apache.org/dist/hbase/1.3.0/hbase-1.3.0-bin.tar.gz

hive 2.1.1
https://archive.apache.org/dist/hive/hive-2.1.1/apache-hive-2.1.1-bin.tar.gz

zookeeper 3.4.9
https://archive.apache.org/dist/zookeeper/zookeeper-3.4.9/zookeeper-3.4.9.tar.gz

Download Resource (All Nodes)

```
cd ~

cd /opt/

wget https://archive.apache.org/dist/hadoop/common/hadoop-2.7.3/hadoop-2.7.3.tar.gz
wget https://archive.apache.org/dist/hbase/1.3.0/hbase-1.3.0-bin.tar.gz
wget https://archive.apache.org/dist/hive/hive-2.1.1/apache-hive-2.1.1-bin.tar.gz
wget https://archive.apache.org/dist/zookeeper/zookeeper-3.4.9/zookeeper-3.4.9.tar.gz
```

```
root@hadoop-master:~# wget https://archive.apache.org/dist/hadoop/common/hadoop-2.7.3/hadoop-2.7.3.tar.gz
--2017-02-14 08:08:54--  https://archive.apache.org/dist/hadoop/common/hadoop-2.7.3/hadoop-2.7.3.tar.gz
Resolving archive.apache.org (archive.apache.org)... 163.172.17.199
Connecting to archive.apache.org (archive.apache.org)|163.172.17.199|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 214092195 (204M) [application/x-gzip]
Saving to: 'hadoop-2.7.3.tar.gz'

100%[======================================================================>] 214,092,195 3.43MB/s   in 1m 58s

2017-02-14 08:10:53 (1.73 MB/s) - 'hadoop-2.7.3.tar.gz' saved [214092195/214092195]

root@hadoop-master:~# wget https://archive.apache.org/dist/hive/hive-2.1.1/apache-hive-2.1.1-bin.tar.gz
--2017-02-14 08:27:31--  https://archive.apache.org/dist/hive/hive-2.1.1/apache-hive-2.1.1-bin.tar.gz
Resolving archive.apache.org (archive.apache.org)... 163.172.17.199
Connecting to archive.apache.org (archive.apache.org)|163.172.17.199|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 149756462 (143M) [application/x-gzip]
Saving to: 'apache-hive-2.1.1-bin.tar.gz'

100%[======================================================================>] 149,756,462 4.99MB/s   in 38s

2017-02-14 08:28:10 (3.74 MB/s) - 'apache-hive-2.1.1-bin.tar.gz' saved [149756462/149756462]

root@hadoop-master:~#
```

Unzip（All Nodes）

```
cd /opt/

tar -xvf hadoop-2.7.3.tar.gz
tar -xvf hbase-1.3.0-bin.tar.gz
tar -xvf apache-hive-2.1.1-bin.tar.gz
tar -xvf zookeeper-3.4.9.tar.gz
```

profile

modify profile（All Nodes）

```
vim ~/.bashrc
```

append env settings （All Nodes）

```
# Hadoop Cluster
export HADOOP_BASE_PATH=/opt
export JAVA_HOME=/usr/lib/jvm/java-8-oracle/
export CLASSPATH=.:$JAVA_HOME/lib/dt.jar:$JAVA_HOME/lib/tools.jar

export HADOOP_HOME=$HADOOP_BASE_PATH/hadoop-2.7.3
export HADOOP_CONF_DIR=$HADOOP_BASE_PATH/hadoop-2.7.3/etc/hadoop
export PATH=$PATH:$JAVA_HOME/bin:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
```

load profile （All Nodes）

```
source ~/.bashrc
```

```
java -version
```

修改 hadoop-env.sh 文件（All Nodes），查找JAVA_HOME，并修改

```
cp $HADOOP_HOME/etc/hadoop/hadoop-env.sh $HADOOP_HOME/etc/hadoop/hadoop-env.sh.bak

vim $HADOOP_HOME/etc/hadoop/hadoop-env.sh
```

```
export JAVA_HOME=/usr/lib/jvm/java-8-oracle/
```

修改 yarn-env.sh 文件（All Nodes），查找JAVA_HOME，并修改

```
cp $HADOOP_HOME/etc/hadoop/yarn-env.sh $HADOOP_HOME/etc/hadoop/yarn-env.sh.bak
vim $HADOOP_HOME/etc/hadoop/yarn-env.sh
```

```
export JAVA_HOME=/usr/lib/jvm/java-8-oracle/
```

修改 core-site.xml 文件（All Nodes）

```
cp $HADOOP_HOME/etc/hadoop/core-site.xml $HADOOP_HOME/etc/hadoop/core-site.xml.bak
vim $HADOOP_HOME/etc/hadoop/core-site.xml
```

```
<configuration>
    <property>
        <name>hadoop.tmp.dir</name>
        <value>file:/tmp/hadoop</value>
        <description>A base for other temporary directories.</description>
    </property>
    <property>
        <name>fs.defaultFS</name>
        <value>hdfs://Hadoop-NameNode:9000</value>
    </property>
    <!-- file system properties -->
    <property>
        <name>fs.default.name</name>
        <value>hdfs://Hadoop-NameNode:9000</value>
    </property>
    <property>
        <name>io.file.buffer.size</name>
        <value>131072</value>
    </property>
    <property>
        <name>hadoop.proxyuser.hduser.hosts</name>
        <value>*</value>
    </property>
    <property>
        <name>hadoop.proxyuser.hduser.groups</name>
        <value>*</value>
    </property>
</configuration>
```

修改 hdfs-site.xml 文件（All Nodes）

```
cp $HADOOP_HOME/etc/hadoop/hdfs-site.xml $HADOOP_HOME/etc/hadoop/hdfs-site.xml.bak

vim $HADOOP_HOME/etc/hadoop/hdfs-site.xml
```

```
<configuration>
    <property>
        <name>dfs.namenode.secondary.http-address</name>
        <value>Hadoop-NameNode:9001</value>
    </property>
    <property>
        <name>dfs.namenode.name.dir</name>
        <value>file:/tmp/hadoop/name</value>
    </property>
    <property>
        <name>dfs.namenode.data.dir</name>
        <value>file:/tmp/hadoop/data</value>
    </property>
    <property>
        <name>dfs.replication</name>
        <value>3</value>
    </property>
    <property>
        <name>dfs.webhdfs.enabled</name>
        <value>true</value>
    </property>
    <property>
        <name>dfs.permissions.enabled</name>
        <value>false</value>
    </property>
</configuration>
```

修改 mapred-site.xml 文件（All Nodes）

```
cp $HADOOP_HOME/etc/hadoop/mapred-site.xml.template $HADOOP_HOME/etc/hadoop/mapred-site.xml

vim $HADOOP_HOME/etc/hadoop/mapred-site.xml
```

```
<configuration>
    <property>
        <name>mapreduce.framework.name</name>
        <value>yarn</value>
    </property>
    <property>
        <name>mapreduce.jobhistory.address</name>
        <value>Hadoop-NameNode:10020</value>
    </property>
    <property>
        <name>mapreduce.jobhistory.webapp.address</name>
        <value>Hadoop-NameNode:19888</value>
    </property>
</configuration>
```

修改 yarn-site.xml 文件（All Nodes）

```
cp $HADOOP_HOME/etc/hadoop/yarn-site.xml $HADOOP_HOME/etc/hadoop/yarn-site.xml.bak

vim $HADOOP_HOME/etc/hadoop/yarn-site.xml
```

```
<configuration>
    <property>
        <name>yarn.nodemanager.aux-services</name>
        <value>mapreduce_shuffle</value>
    </property>
    <property>
        <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
        <value>org.apache.hadoop.mapred.ShuffleHandler</value>
    </property>
    <property>
        <name>yarn.resourcemanager.address</name>
        <value>Hadoop-NameNode:8032</value>
    </property>
    <property>
        <name>yarn.resourcemanager.scheduler.address</name>
        <value>Hadoop-NameNode:8030</value>
    </property>
    <property>
        <name>yarn.resourcemanager.resource-tracker.address</name>
        <value>Hadoop-NameNode:8031</value>
    </property>
    <property>
        <name>yarn.resourcemanager.admin.address</name>
        <value>Hadoop-NameNode:8033</value>
    </property>
    <property>
        <name>yarn.resourcemanager.webapp.address</name>
        <value>Hadoop-NameNode:8088</value>
    </property>
</configuration>
```

修改 hadoop/etc/hadoop/masters 文件（All Nodes），默认如果没有那么就得新建一个，在里面加上一行：

```
cp $HADOOP_HOME/etc/hadoop/masters $HADOOP_HOME/etc/hadoop/masters.bak

vim $HADOOP_HOME/etc/hadoop/masters
```

```
Hadoop-NameNode
```

修改 slaves 文件（All Nodes）

```
cp $HADOOP_HOME/etc/hadoop/slaves $HADOOP_HOME/etc/hadoop/slaves.bak

vim $HADOOP_HOME/etc/hadoop/slaves
```

```
Hadoop-DataNode-1
Hadoop-DataNode-2
```

格式化HDFS系统（Master Node）

```
$HADOOP_HOME/bin/hadoop namenode -format
```

```
scp -r /opt/hadoop root@Hadoop-DataNode-1:/opt/
scp -r /opt/hadoop root@Hadoop-DataNode-2:/opt/
```

启动 dfs，yarn（Master Node）

```
$HADOOP_HOME/sbin/start-dfs.sh
# Master: NameNode, SecondaryNameNode
# Slave: DataNode
jps

$HADOOP_HOME/sbin/start-yarn.sh
# Master: ResourceManager
# Slave: NodeManager
jps
```

停止 dfs，yarn（Master Node）

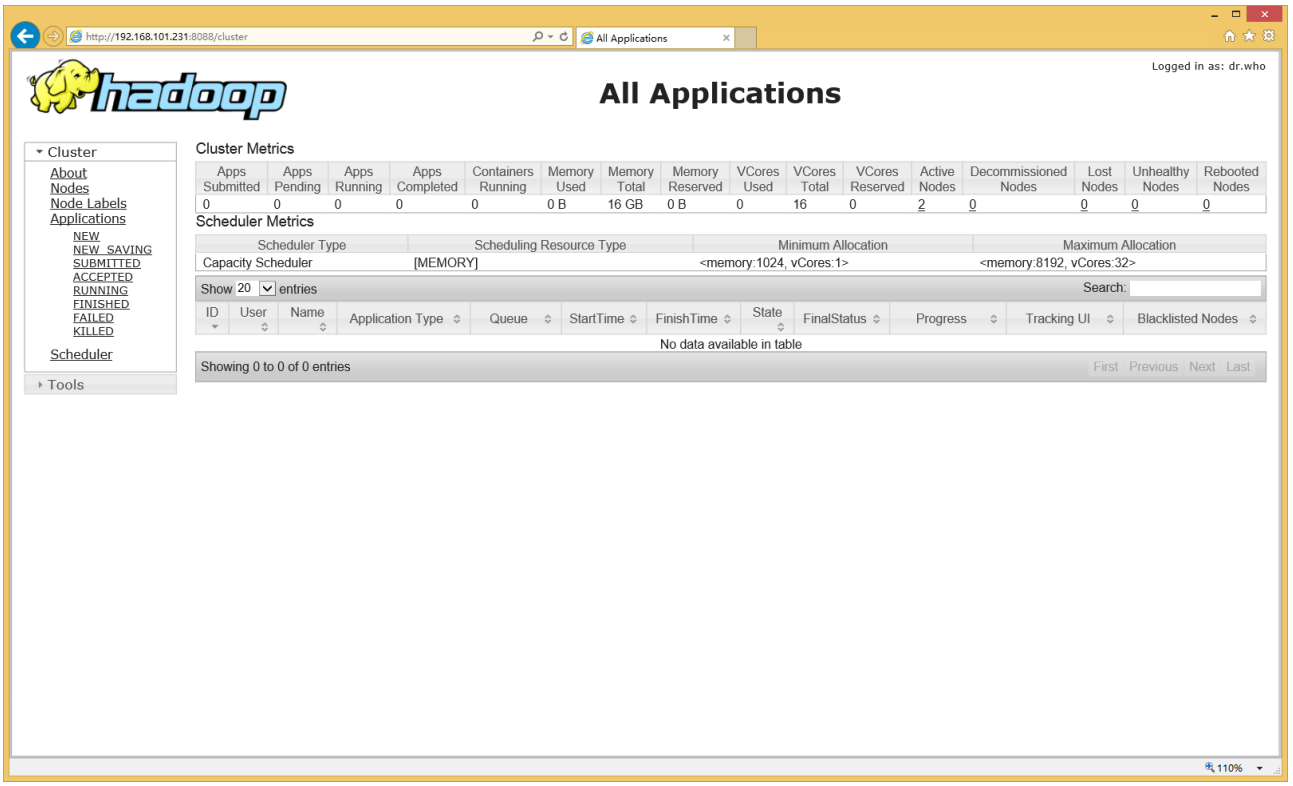
```
$HADOOP_HOME/sbin/stop-yarn.sh
$HADOOP_HOME/sbin/stop-dfs.sh
```

查看hdfs的运行状态（Master Node）

```
$HADOOP_HOME/bin/hdfs dfsadmin -report
netstat -tnulp | grep java
```

ResourceManager web UI

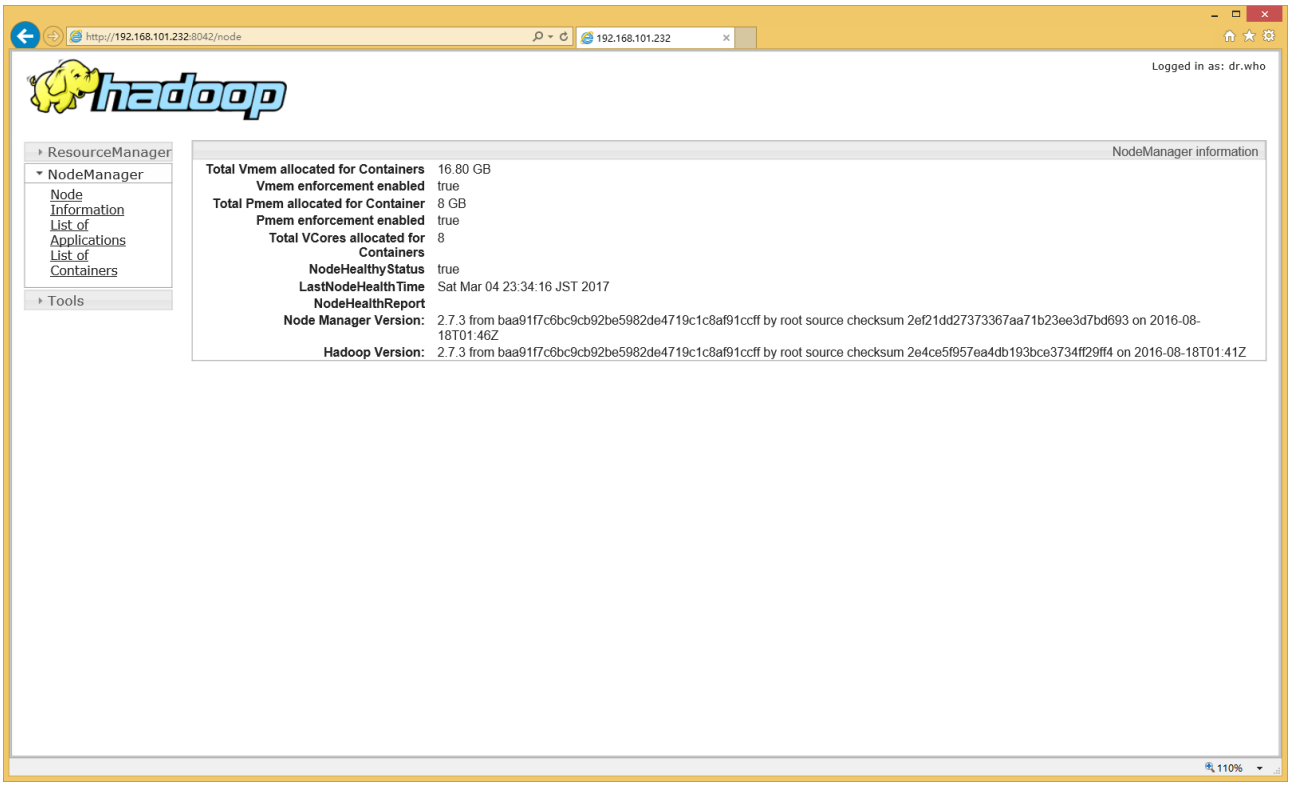
http://192.168.101.231:8088/



NodeManager Web UI

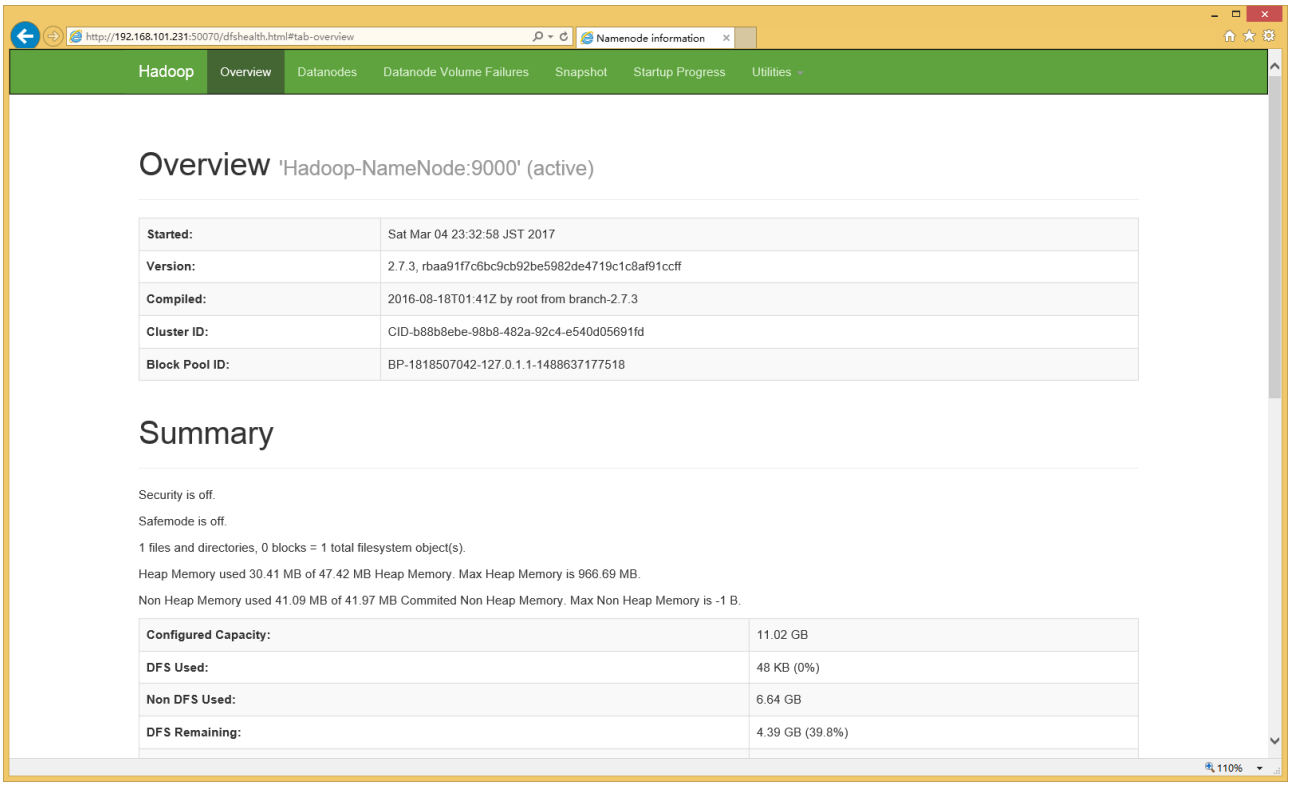
http://192.168.101.232:8042/

http://192.168.101.233:8042/

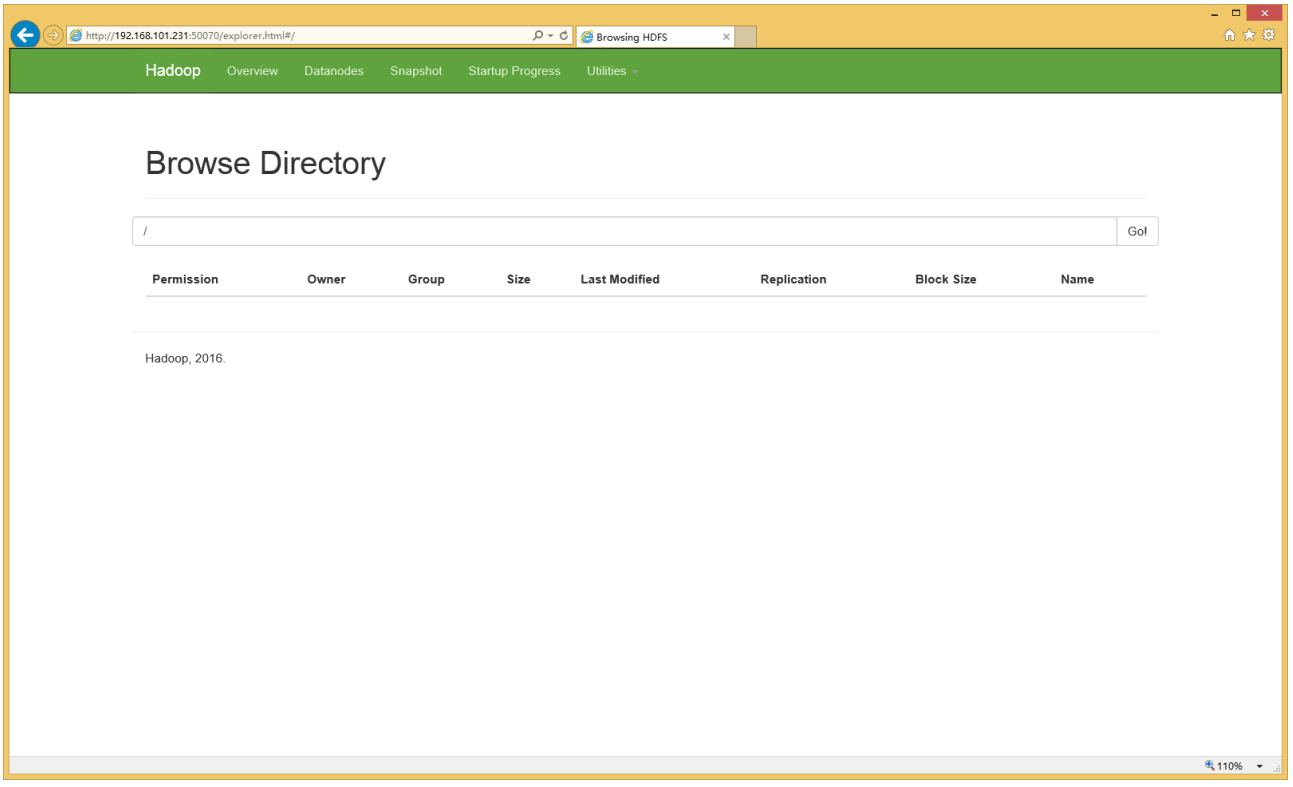


NameNode Web UI

http://192.168.101.231:50070/



http://192.168.101.231:50070/explorer.html#/



```
1.
2.  $HADOOP_HOME/bin/hadoop fs -mkdir /input
3.  $HADOOP_HOME/bin/hadoop fs -ls /
4.
5.  $HADOOP_HOME/bin/hadoop fs -put $HADOOP_HOME/LICENSE.txt /input
6.  $HADOOP_HOME/bin/hadoop fs -put $HADOOP_HOME/NOTICE.txt /input
7.  $HADOOP_HOME/bin/hadoop fs -put $HADOOP_HOME/README.txt /input
8.  $HADOOP_HOME/bin/hadoop fs -ls /input
9.
10. $HADOOP_HOME/bin/hadoop jar $HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapre
    duce-examples-2.7.3.jar wordcount /input /output
11.
12. $HADOOP_HOME/bin/hadoop fs -ls /output
13. $HADOOP_HOME/bin/hadoop fs -cat /output/part-r-00000
```

```
root@master:~# $HADOOP_HOME/bin/hadoop fs -mkdir /input
root@master:~# $HADOOP_HOME/bin/hadoop fs -ls /
Found 1 items
drwxr-xr-x   - root supergroup          0 2017-03-09 00:02 /input
root@master:~# $HADOOP_HOME/bin/hadoop fs -put $HADOOP_HOME/LICENSE.txt /input
root@master:~# $HADOOP_HOME/bin/hadoop fs -put $HADOOP_HOME/NOTICE.txt /input
root@master:~# $HADOOP_HOME/bin/hadoop fs -put $HADOOP_HOME/README.txt /input
root@master:~# $HADOOP_HOME/bin/hadoop fs -ls /input

Found 3 items
-rw-r--r--   3 root supergroup      84854 2017-03-09 00:02 /input/LICENSE.txt
-rw-r--r--   3 root supergroup      14978 2017-03-09 00:02 /input/NOTICE.txt
-rw-r--r--   3 root supergroup       1366 2017-03-09 00:02 /input/README.txt
root@master:~#
root@master:~# hadoop jar $HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.3.jar wordcount /input /output
17/03/09 00:02:57 INFO client.RMProxy: Connecting to ResourceManager at master.trex.com/172.17.0.2:8040
17/03/09 00:02:58 INFO input.FileInputFormat: Total input paths to process : 3
17/03/09 00:02:58 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1249)
        at java.lang.Thread.join(Thread.java:1323)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:609)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:370)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:546)
17/03/09 00:02:58 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1249)
        at java.lang.Thread.join(Thread.java:1323)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:609)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:370)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:546)
17/03/09 00:02:58 INFO mapreduce.JobSubmitter: number of splits:3
17/03/09 00:02:58 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1489017721151_0001
17/03/09 00:02:59 INFO impl.YarnClientImpl: Submitted application application_1489017721151_0001
17/03/09 00:02:59 INFO mapreduce.Job: The url to track the job: http://master.trex.com:8088/proxy/application_1489017721151_0001/
17/03/09 00:02:59 INFO mapreduce.Job: Running job: job_1489017721151_0001
17/03/09 00:03:11 INFO mapreduce.Job: Job job_1489017721151_0001 running in uber mode : false
17/03/09 00:03:11 INFO mapreduce.Job:  map 0% reduce 0%
17/03/09 00:03:30 INFO mapreduce.Job:  map 67% reduce 0%
17/03/09 00:03:31 INFO mapreduce.Job:  map 100% reduce 0%
17/03/09 00:03:38 INFO mapreduce.Job:  map 100% reduce 100%
17/03/09 00:03:39 INFO mapreduce.Job: Job job_1489017721151_0001 completed successfully
17/03/09 00:03:39 INFO mapreduce.Job: Counters: 50
        File System Counters
                FILE: Number of bytes read=42582
                FILE: Number of bytes written=559213
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=101526
                HDFS: Number of bytes written=30052
                HDFS: Number of read operations=12
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Killed map tasks=1
```

```
1.
2.  $HADOOP_HOME/bin/hadoop fs -mkdir /user
3.  $HADOOP_HOME/bin/hadoop fs -mkdir /user/root
4.  $HADOOP_HOME/bin/hadoop fs -mkdir /user/root/input
5.  $HADOOP_HOME/bin/hadoop fs -ls /
6.
7.  $HADOOP_HOME/bin/hadoop fs -put $HADOOP_HOME/LICENSE.txt /user/root/input
8.  $HADOOP_HOME/bin/hadoop fs -put $HADOOP_HOME/NOTICE.txt /user/root/input
9.  $HADOOP_HOME/bin/hadoop fs -put $HADOOP_HOME/README.txt /user/root/input
10. $HADOOP_HOME/bin/hadoop fs -ls /user/root/input
11.
12. $HADOOP_HOME/bin/hadoop jar $HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapre
    duce-examples-2.7.3.jar wordcount /user/root/input /user/root/output
13.
14. $HADOOP_HOME/bin/hadoop fs -ls /user/root/output
15. $HADOOP_HOME/bin/hadoop fs -cat /user/root/output/part-r-00000
```