

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)

```
1. cd ~
2.
3. cd /opt/
4.
5. wget https://archive.apache.org/dist/hadoop/common/hadoop-2.7.3/hadoop-2.7.3
  .tar.gz
6. wget https://archive.apache.org/dist/hbase/1.3.0/hbase-1.3.0-bin.tar.gz
7. wget https://archive.apache.org/dist/hive/hive-2.1.1/apache-hive-2.1.1-bin.t
  ar.gz
8. 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)

```
1. cd /opt/
2.
3. tar -xvf hadoop-2.7.3.tar.gz
4. tar -xvf hbase-1.3.0-bin.tar.gz
5. tar -xvf apache-hive-2.1.1-bin.tar.gz
6. tar -xvf zookeeper-3.4.9.tar.gz
```

profile

modify profile (All Nodes)

```
1. vim ~/.bashrc
```

append env settings (All Nodes)

```
1. # Hadoop Cluster
2. export HADOOP_BASE_PATH=/opt
3. export JAVA_HOME=/usr/lib/jvm/java-8-oracle/
4. export CLASSPATH=.:$JAVA_HOME/lib/dt.jar:$JAVA_HOME/lib/tools.jar
5.
6. export HADOOP_HOME=$HADOOP_BASE_PATH/hadoop-2.7.3
7. export HADOOP_CONF_DIR=$HADOOP_BASE_PATH/hadoop-2.7.3/etc/hadoop
8. export PATH=$PATH:$JAVA_HOME/bin:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
```

load profile (All Nodes)

```
1. source ~/.bashrc
```

```
1. java -version
```

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

```
1. cp $HADOOP_HOME/etc/hadoop/hadoop-env.sh $HADOOP_HOME/etc/hadoop/hadoop-env.sh.bak
2.
3. vim $HADOOP_HOME/etc/hadoop/hadoop-env.sh
```

```
1. 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

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

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

修改 core-site.xml 文件 (All Nodes)

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

```
1. <configuration>
2.   <property>
3.     <name>hadoop.tmp.dir</name>
4.     <value>file:/tmp/hadoop</value>
5.     <description>A base for other temporary directories.</description>
6.   </property>
7.   <property>
8.     <name>fs.defaultFS</name>
9.     <value>hdfs://Hadoop-NameNode:9000</value>
10.  </property>
11.  <!-- file system properties -->
12.  <property>
13.    <name>fs.default.name</name>
14.    <value>hdfs://Hadoop-NameNode:9000</value>
15.  </property>
16.  <property>
17.    <name>io.file.buffer.size</name>
18.    <value>131072</value>
19.  </property>
20.  <property>
21.    <name>hadoop.proxyuser.hduser.hosts</name>
22.    <value>*</value>
23.  </property>
24.  <property>
25.    <name>hadoop.proxyuser.hduser.groups</name>
26.    <value>*</value>
27.  </property>
28. </configuration>
```

修改 hdfs-site.xml 文件 (All Nodes)

```
1. cp $HADOOP_HOME/etc/hadoop/hdfs-site.xml $HADOOP_HOME/etc/hadoop/hdfs-site.xml.bak
2.
3. vim $HADOOP_HOME/etc/hadoop/hdfs-site.xml
```

```

1. <configuration>
2.   <property>
3.     <name>dfs.namenode.secondary.http-address</name>
4.     <value>Hadoop-NameNode:9001</value>
5.   </property>
6.   <property>
7.     <name>dfs.namenode.name.dir</name>
8.     <value>file:/tmp/hadoop/name</value>
9.   </property>
10.  <property>
11.    <name>dfs.namenode.data.dir</name>
12.    <value>file:/tmp/hadoop/data</value>
13.  </property>
14.  <property>
15.    <name>dfs.replication</name>
16.    <value>3</value>
17.  </property>
18.  <property>
19.    <name>dfs.webhdfs.enabled</name>
20.    <value>true</value>
21.  </property>
22.  <property>
23.    <name>dfs.permissions.enabled</name>
24.    <value>false</value>
25.  </property>
26. </configuration>

```

修改 mapred-site.xml 文件 (All Nodes)

```

1. cp $HADOOP_HOME/etc/hadoop/mapred-site.xml.template $HADOOP_HOME/etc/hadoop/
   mapred-site.xml
2.
3. vim $HADOOP_HOME/etc/hadoop/mapred-site.xml

```

```

1. <configuration>
2.   <property>
3.     <name>mapreduce.framework.name</name>
4.     <value>yarn</value>
5.   </property>
6.   <property>
7.     <name>mapreduce.jobhistory.address</name>
8.     <value>Hadoop-NameNode:10020</value>
9.   </property>
10.  <property>
11.    <name>mapreduce.jobhistory.webapp.address</name>
12.    <value>Hadoop-NameNode:19888</value>
13.  </property>
14. </configuration>

```

修改 yarn-site.xml 文件 (All Nodes)

```
1. cp $HADOOP_HOME/etc/hadoop/yarn-site.xml $HADOOP_HOME/etc/hadoop/yarn-site.xml.bak
2.
3. vim $HADOOP_HOME/etc/hadoop/yarn-site.xml
```

```
1. <configuration>
2.   <property>
3.     <name>yarn.nodemanager.aux-services</name>
4.     <value>mapreduce_shuffle</value>
5.   </property>
6.   <property>
7.     <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
8.     <value>org.apache.hadoop.mapred.ShuffleHandler</value>
9.   </property>
10.  <property>
11.    <name>yarn.resourcemanager.address</name>
12.    <value>Hadoop-NameNode:8032</value>
13.  </property>
14.  <property>
15.    <name>yarn.resourcemanager.scheduler.address</name>
16.    <value>Hadoop-NameNode:8030</value>
17.  </property>
18.  <property>
19.    <name>yarn.resourcemanager.resource-tracker.address</name>
20.    <value>Hadoop-NameNode:8031</value>
21.  </property>
22.  <property>
23.    <name>yarn.resourcemanager.admin.address</name>
24.    <value>Hadoop-NameNode:8033</value>
25.  </property>
26.  <property>
27.    <name>yarn.resourcemanager.webapp.address</name>
28.    <value>Hadoop-NameNode:8088</value>
29.  </property>
30. </configuration>
```

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

```
1. cp $HADOOP_HOME/etc/hadoop/masters $HADOOP_HOME/etc/hadoop/masters.bak
2.
3. vim $HADOOP_HOME/etc/hadoop/masters
```

```
1. Hadoop-NameNode
```

修改 `slaves` 文件（All Nodes）

```
1. cp $HADOOP_HOME/etc/hadoop/slaves $HADOOP_HOME/etc/hadoop/slaves.bak
2.
3. vim $HADOOP_HOME/etc/hadoop/slaves
```

1. Hadoop-DataNode-1
2. Hadoop-DataNode-2

格式化HDFS系统 (Master Node)

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

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

启动 dfs , yarn (Master Node)

1. `$HADOOP_HOME/sbin/start-dfs.sh`
2. `# Master: NameNode, SecondaryNameNode`
3. `# Slave: DataNode`
4. `jps`
- 5.
6. `$HADOOP_HOME/sbin/start-yarn.sh`
7. `# Master: ResourceManager`
8. `# Slave: NodeManager`
9. `jps`

停止 dfs , yarn (Master Node)

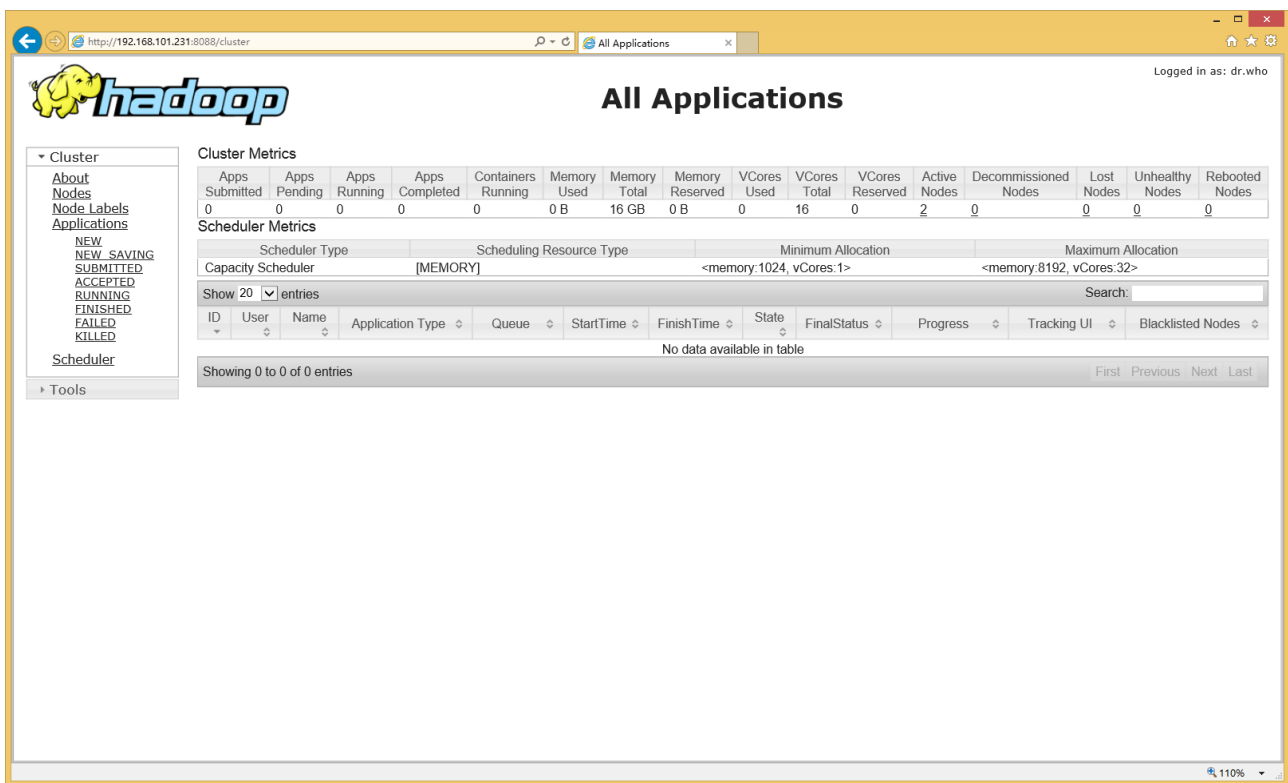
1. `$HADOOP_HOME/sbin/stop-yarn.sh`
2. `$HADOOP_HOME/sbin/stop-dfs.sh`

查看hdfs的运行状态 (Master Node)

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

ResourceManager web UI

<http://192.168.101.231:8088/>



The screenshot shows the Hadoop All Applications web interface. The top navigation bar includes the Hadoop logo, the title "All Applications", and the user "dr.who". The left sidebar contains a "Cluster" menu with options like "About", "Nodes", "Node Labels", "Applications", and "Scheduler". The main content area displays "Cluster Metrics" and "Scheduler Metrics".

Cluster Metrics

| Apps Submitted | Apps Pending | Apps Running | Apps Completed | Containers Running | Memory Used | Memory Total | Memory Reserved | VCores Used | VCores Total | VCores Reserved | Active Nodes | Decommissioned Nodes | Lost Nodes | Unhealthy Nodes | Rebooted Nodes |
|----------------|--------------|--------------|----------------|--------------------|-------------|--------------|-----------------|-------------|--------------|-----------------|--------------|----------------------|------------|-----------------|----------------|
| 0 | 0 | 0 | 0 | 0 | 0 B | 16 GB | 0 B | 0 | 16 | 0 | 2 | 0 | 0 | 0 | 0 |

Scheduler Metrics

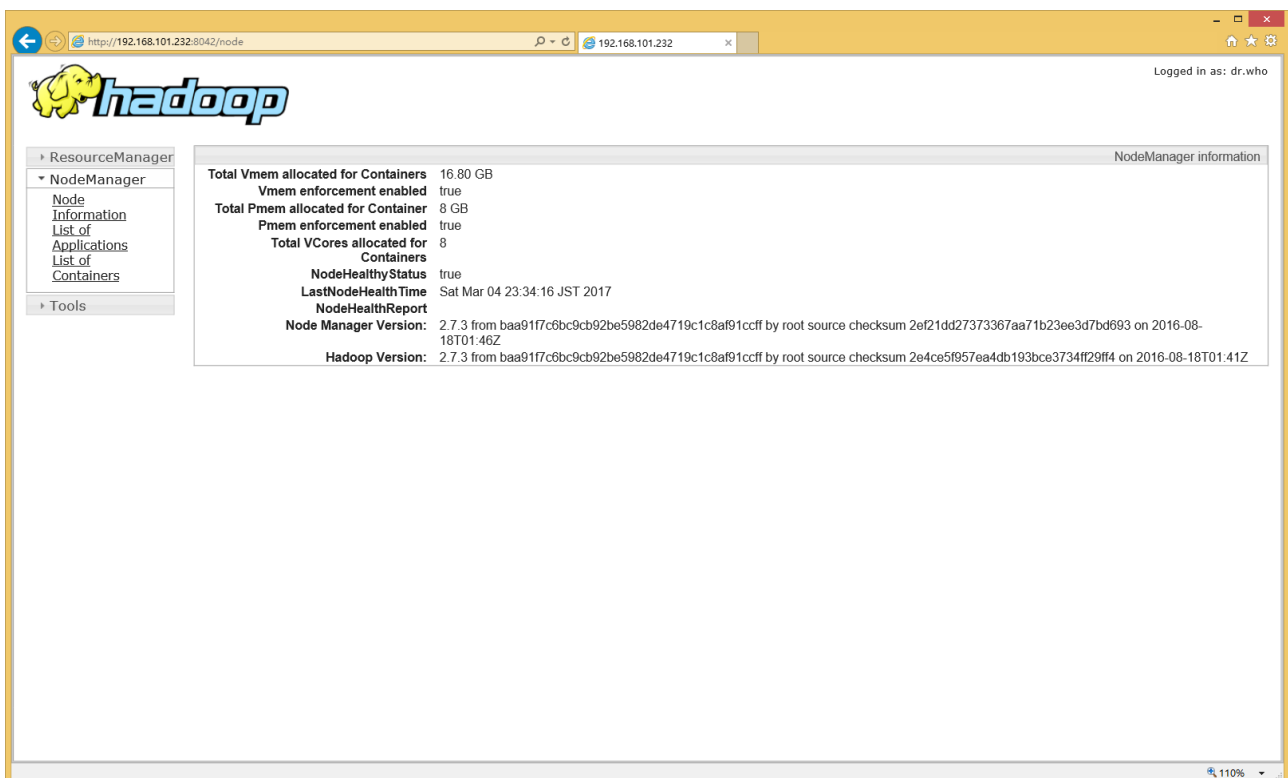
| Scheduler Type | Scheduling Resource Type | Minimum Allocation | Maximum Allocation |
|--------------------|--------------------------|-------------------------|--------------------------|
| Capacity Scheduler | [MEMORY] | <memory:1024, vCores:1> | <memory:8192, vCores:32> |

Below the scheduler metrics is a table with columns: ID, User, Name, Application Type, Queue, StartTime, FinishTime, State, FinalStatus, Progress, Tracking UI, and Blacklisted Nodes. The table is currently empty, displaying "Showing 0 to 0 of 0 entries".

NodeManager Web UI

<http://192.168.101.232:8042/>

<http://192.168.101.233:8042/>



The screenshot shows the Hadoop NodeManager web interface. The top navigation bar includes the Hadoop logo, the title "NodeManager", and the user "dr.who". The left sidebar contains a "ResourceManager" menu with options like "NodeManager", "Node Information", "List of Applications", and "List of Containers". The main content area displays "NodeManager information".

NodeManager information

| | |
|---------------------------------------|--|
| Total Vmem allocated for Containers | 16.80 GB |
| Vmem enforcement enabled | true |
| Total Pmem allocated for Container | 8 GB |
| Pmem enforcement enabled | true |
| Total VCores allocated for Containers | 8 |
| NodeHealthyStatus | true |
| LastNodeHealthTime | Sat Mar 04 23:34:16 JST 2017 |
| NodeHealthReport | |
| Node Manager Version: | 2.7.3 from baa91f7c6bc9cb92be5982de4719c1c8af91cff by root source checksum 2ef21dd27373367aa71b23ee3d7bd693 on 2016-08-18T01:46Z |
| Hadoop Version: | 2.7.3 from baa91f7c6bc9cb92be5982de4719c1c8af91cff by root source checksum 2e4ce5f957ea4db193bce3734f129ff4 on 2016-08-18T01:41Z |

NameNode Web UI

<http://192.168.101.231:50070/>

←

http://192.168.101.231:50070/dfshealth.html#tab-overview

×

Namenode information

⌵

⌵

⌵

⌵

⌵

⌵

Hadoop

Overview

Datanodes

Datanode Volume Failures

Snapshot

Startup Progress

Utilities

Overview 'Hadoop-NameNode:9000' (active)

| | |
|----------------|--|
| Started: | Sat Mar 04 23:32:58 JST 2017 |
| Version: | 2.7.3, rbaa91f7c6bc9cb92be5982de4719c1c8af91ccff |
| Compiled: | 2016-08-18T01:41Z by root from branch-2.7.3 |
| Cluster ID: | CID-b88b8ebe-98b8-482a-92c4-e540d05691fd |
| Block Pool ID: | BP-1818507042-127.0.1.1-1488637177518 |

Summary

Security is off.

Safemode is off.

1 files and directories, 0 blocks = 1 total filesystem object(s).

Heap Memory used 30.41 MB of 47.42 MB Heap Memory. Max Heap Memory is 966.69 MB.

Non Heap Memory used 41.09 MB of 41.97 MB Committed Non Heap Memory. Max Non Heap Memory is -1 B.

| | |
|----------------------|-----------------|
| Configured Capacity: | 11.02 GB |
| DFS Used: | 48 KB (0%) |
| Non DFS Used: | 6.64 GB |
| DFS Remaining: | 4.39 GB (39.8%) |

110%

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

←

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

×

Browsing HDFS

⌵

⌵

⌵

⌵

⌵

⌵

Hadoop

Overview

Datanodes

Snapshot

Startup Progress

Utilities

Browse Directory

/

Go!

| Permission | Owner | Group | Size | Last Modified | Replication | Block Size | Name |
|---------------|-------|-------|------|---------------|-------------|------------|------|
| Hadoop, 2016. | | | | | | | |

110%


```

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 $HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.3.jar wordcount /input /output
17/03/09 00:02:57 INFO client.RMProxy: Connecting to ResourceManager at master.trex.com/172.17.0.2:8040
17/03/09 00:02:58 INFO input.FileInputFormat: Total input paths to process : 3
17/03/09 00:02:58 WARN hdfs.DFSCClient: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1249)
    at java.lang.Thread.join(Thread.java:1323)
    at org.apache.hadoop.hdfs.DFSOutputStream$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.DFSCClient: 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
```