

Configure MapReduce

1. Configure JobTracker on **elephant**

```
$ sudo nano /etc/hadoop/conf/mapred-site.xml
```

```
<?xml version="1.0"?>
<configuration>
  <property>
    <name>mapred.local.dir</name>
    <value>/disk1/mapred/local,/disk2/mapred/local</value>
  </property>
  <property>
    <name>mapred.job.tracker</name>
    <value>horse:8021</value>
  </property>
  <property>
    <name>mapred.system.dir</name>
    <value>/mapred/system</value>
  </property>
  <property>
    <name>mapreduce.jobtracker.staging.root.dir</name>
    <value>/user</value>
  </property>
</configuration>
```

2. Copy configuration to **all nodes**

- \$ cd
- \$./copy_config.sh

3. Create MapReduce paths on **HDFS**

- \$ sudo -u hdfs hadoop fs -mkdir /mapred/system
- \$ sudo -u hdfs hadoop fs -chown mapred:hadoop /mapred/system

4. Start JobTracker on **horse**

- \$ sudo service hadoop-0.20-mapreduce-jobtracker start

5. Start TaskTracker on **all nodes**

- \$ sudo service hadoop-0.20-mapreduce-tasktracker start

