

## Spark 安装

- Hadoop 配置

修改/home/bigdata/hadoop-2.7.3/etc/hadoop/yarn-site.xml, 增加如下内容

```
<property>
  <name>yarn.log-aggregation-enable</name>
  <value>true</value>
</property>
<property>
  <name>yarn.log.server.url</name>
  <value>http://bigdata:19888/jobhistory/logs</value>
</property>
<property>
  <name>yarn.nodemanager.pmem-check-enabled</name>
  <value>>false</value>
</property>
<property>
  <name>yarn.nodemanager.vmem-check-enabled</name>
  <value>>false</value>
</property>
```

在所有节点上应用同样的修改, 重启整个 Yarn 服务

```
cd /home/bigdata/hadoop-2.7.3
sbin/stop-yarn.sh
sbin/start-yarn.sh
```

启动 MapReduce History Server

```
sbin/mr-jobhistory-daemon.sh start historyserver
```

在浏览器中打开 mapreduce history server 地址

<http://bigdata:19888>

- Spark 配置

解压缩 spark-2.1.0-bin-hadoop2.7.tgz, 进入到 conf 目录

```
tar zxvf spark-2.1.0-bin-hadoop2.7.tgz
cd spark-2.1.0-bin-hadoop2.7/conf
```

创建文件 spark-env.sh, 内容是

```
export HADOOP_CONF_DIR=/home/bigdata/hadoop-2.7.3/etc/hadoop
```

同样在 conf 目录下创建文件 spark-defaults.conf,内容是

```
spark.master=local
spark.yarn.historyServer.address=bigdata:18080
spark.history.ui.port=18080
spark.eventLog.enabled=true
spark.eventLog.dir=hdfs:///tmp/spark/events
spark.history.fs.logDirectory=hdfs:///tmp/spark/events
```

在 hdfs 上创建目录/tmp/spark/events

```
hadoop fs -mkdir -p /tmp/spark/events
```

启动 Spark History Server

```
cd spark-2.1.0-bin-hadoop2.7
sbin/start-history-server.sh
```

在浏览器中打开 spark history server 地址

```
http://bigdata:18080
```