



链滴

【bigdata】2.Hadoop 集群模式安装

作者: [jditlee](#)

原文链接: <https://ld246.com/article/1679037448389>

来源网站: [链滴](#)

许可协议: [署名-相同方式共享 4.0 国际 \(CC BY-SA 4.0\)](#)

上传Hadoop并解压

1.上传Hadoop2.6.5到master

2.解压

```
tar -zxvf hadoop-2.6.5.tar.gz -C /opt/module/
```

Hadoop配置

Hadoop配置文件位置: `hadoop-2.6.5/etc/hadoop`

1. `hadoop-env.sh`

```
-- 修改Javahome  
export JAVA_HOME=/opt/module/jdk1.8.0_212
```

2. `yarn-env.sh`

```
-- 增加Javahome  
export JAVA_HOME=/opt/module/jdk1.8.0_212
```

3. `core-site.xml`

```
<?xml version="1.0" encoding="UTF-8"?>  
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>  
<!--  
Licensed under the Apache License, Version 2.0 (the "License");  
you may not use this file except in compliance with the License.  
You may obtain a copy of the License at
```

```
http://www.apache.org/licenses/LICENSE-2.0
```

```
Unless required by applicable law or agreed to in writing, software  
distributed under the License is distributed on an "AS IS" BASIS,  
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.  
See the License for the specific language governing permissions and  
limitations under the License. See accompanying LICENSE file.  
-->
```

```
<!-- Put site-specific property overrides in this file. -->
```

```
<configuration>  
<property>  
  <name>fs.defaultFS</name>  
  <value>hdfs://master:9000</value>  
  <description>用来指定默认的文件系统</description>  
</property>
```

```
<property>
  <name>hadoop.tmp.dir</name>
  <value>file:/opt/module/hadoop-2.6.5/tmp</value>
  <description>hadoop临时文件存放目录</description>
</property>
</configuration>
```

4. hdfs-site.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at
```

<http://www.apache.org/licenses/LICENSE-2.0>

Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License. See accompanying LICENSE file.

```
-->
```

```
<!-- Put site-specific property overrides in this file. -->
```

```
<configuration>
  <property>
    <name>dfs.namenode.secondary.http-address</name>
    <value>master:9001</value>
    <description>secondaryNamenode地址和端口</description>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:/opt/module/hadoop-2.6.5/dfs/name</value>
    <description>保存FsImage镜像的目录，作用是存放hadoop的名称节点namenode里的
etadata</description>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:/opt/module/hadoop-2.6.5/dfs/data</value>
    <description>存放HDFS文件系统数据文件的目录，作用是存放hadoop的数
据节点datanode里的多个数据块</description>
  </property>
  <property>
    <name>dfs.replication</name>
    <value>2</value>
    <description>默认block块副本数</description>
  </property>
</configuration>
```

5. mapred-site.xml

```
<configuration>
<property>
<name>mapreduce.framework.name</name>
<value>yarn</value>
<description>指定运行mapreduce的环境是yarn</description>
</property>
<!--hadoop历史服务器-->
<property>
<name>mapreduce.jobhistory.address</name>
<value>master:10020</value>
<description>MR JobHistory服务器进程间通信地址</description>
</property>
<property>
<name>mapreduce.jobhistory.webapp.address</name>
<value>master:19888</value>
<description>MR JobHistory服务器的用户界面地址</description>
</property>
<property>
<name>mapreduce.jobhistory.done-dir</name>
<value>/mr-history/done</value>
<description>已执行完毕作业信息存储位置</description>
</property>
<property>
<name>mapreduce.jobhistory.intermediate-done-dir</name>
<value>/mr-history/tmp</value>
<description>正在运行的作业信息存储位置</description>
</property>
<property>
<name>yarn.app.mapreduce.am.staging-dir</name>
<value>/mr-history/hadoop-yarn/</value>
<description>MR作业在提交时所使用的临时目录， 是一个本地路径</description>
</property>
<property>
<name>mapreduce.map.memory.mb</name>
<value>2048</value>
<description>调度器为每个map task申请的内存数， 各Job也可以单独指定， 如果实际使用的资源超过该值， 则会被强制杀死</description>
</property>
<property>
<name>mapreduce.reduce.memory.mb</name>
<value>2048</value>
<description>调度器为每个reduce task申请的内存数， 同map task， 超出强制杀死</description>
</property>
<property>
<name>mapreduce.job.reduce.slowstart.completedmaps</name>
<value>0.8</value>
<description>当map task完成80%时， 为reduce申请资源， reduce开始进行拷贝map结果数据和做reduce shuffle操作， 默认0.05</description>
</property>
</configuration>
```

6. yarn-site.xml

```
<configuration>
<!-- 日志聚合相关-->
<property>
<name>yarn.log-aggregation-enable</name>
<value>true</value>
<description>开启日志聚合功能，开启后日志保存在hdfs上</description>
</property>
<property>
<name>yarn.log-aggregation.retain-seconds</name>
<value>86400</value>
<description>聚合后的日志在hdfs上的保存时间，单位为秒</description>
</property>
<property>
<name>yarn.log.server.url</name>
<value>http://master:19888/jobhistory/logs</value>
<description>日志聚合服务器URL</description>
</property>
<property>
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
<description>NodeManager上运行的附属服务,需配置成mapreduce_shuffle，才可运行MapReduce程序</description>
</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>master:8032</value>
<description>RM 对客户端暴露的地址。客户端通过该地址向RM提交应用程序，杀死应用程序</description>
</property>
<property>
<name>yarn.resourcemanager.scheduler.address</name>
<value>master:8030</value>
<description>RM 对ApplicationMaster暴露的访问地址。AM通过该地址向RM申请资源、释放资源</description>
</property>
<property>
<name>yarn.resourcemanager.resource-tracker.address</name>
<value>master:8035</value>
<description>RM 对NodeManager暴露的地址，NM通过该地址向RM汇报心跳，领取任务</description>
</property>
<property>
<name>yarn.resourcemanager.admin.address</name>
<value>master:8033</value>
<description>RM 对管理员暴露的访问地址。管理员通过该地址向RM发送管理命令等</description>
</property>
<property>
<name>yarn.resourcemanager.webapp.address</name>
<value>master:8088</value>
</property>
</configuration>
```

```
<description>RM对外web ui地址。用户可通过该地址在浏览器中查看集群各类信息</description>
</property>
<!-- 关闭虚拟内存检查-->
<property>
<name>yarn.nodemanager.pmem-check-enabled</name>
<value>>false</value>
<description>是否启动一个线程检查每个任务正使用的物理内存量，如果任务超出分配值，则直接
其杀掉，默认为true</description>
</property>
</property>
<name>yarn.nodemanager.vmem-check-enabled</name>
<value>>false</value>
<description>是否启动一个线程检查每个任务正使用的虚拟内存量，如果任务超出分配值，则直接
其杀掉，默认是true</description>
</property>
</configuration>
```

7. 配置slave, datanode节点

```
-- 每行末尾不能有空格
slave1
slave2
```

8. 启动集群,格式化namenode

`hdfs namenode - format` 或 `hadoop namenode - format` 因为我们namenode地址和yarn地址都master, 所以直接start-all就行了, 如果配置的yarn和namenode不是同一个地址, 那么需要在配置yarn的主机上启动yarn: `start-yarn.sh`, 在配置namenode的机器上启动hdfs: `start-hdfs.sh`

`start-all.sh`

9. 停止集群:

`stop-all.sh`

10. 重新格式化集群

配置文件hdfs.site.xml文件里

- 1.删除dfs.datanode.name.dir配置的目录
- 2.删除dfs.datanode.data.dir配置的目录
- 3.删除logs和tmp
- 4.重新执行格式化