



链滴

# HDFS 高可用分布式文件存储系统部署

作者: [GeekBoyDqz](#)

原文链接: <https://ld246.com/article/1571045803916>

来源网站: [链滴](#)

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

# 初识Hadoop

在大数据时代下，个人用户以及计算机系统产生的数据都飞速增长，因此大数据火了。所谓大数据，从两个方面去理解：

- 1) 数据的存储。存储的目的是用于大数据分析，所以大数据存储方案是一个研究点。
- 2) 数据的分析。如何对海量数据采取行之有效的分析方法是另一个研究点。

Hadoop的核心是HDFS和MapReduce，HDFS作为Hadoop的数据存储方案，而MapReduce作为Hadoop的数据分析方案。

## 部署Hadoop

### 1、环境准备

服务器IP	Hosts
192.168.66.30	master
192.168.66.31	slave1
192.168.66.32	slave2

此次配置的hadoop版本是hadoop-3.1.2，JDK版本jdk-8u201

#### 1.1、上传软件包

- JDK: [[Java SE Development Kit 8u201](#)]
- Hadoop: [[Apache Hadoop Download](#)]

#### 1.2、设置vim显示行号

```
[root@master ~]# vim /root/.vimrc  
[root@master ~]# set nu
```

#### 1.3、设置主机名和hosts

```
[root@master ~]# hostnamectl set-hostname master  
[root@master ~]# vim /etc/hosts
```

#### 1.4、配置免密登入

- SSH配成无密码登录，因为客户端和服务端都是本机，所以只要把ssh生成的公钥存到authorized\_keys文件（存在服务器端）即可。

```
[root@master ~]# ssh-keygen  
[root@master ~]# ssh-copy-id 192.168.66.31  
[root@master ~]# ssh-copy-id 192.168.66.32
```

## 1.5、关闭防火墙

```
[root@master ~]# systemctl stop firewalld;systemctl disable firewalld
[root@master ~]# for i in 192.168.66.3{1,2};do echo ">>> $i";ssh $i "systemctl stop firewalld;systemctl disable firewalld;systemctl status firewalld | grep Active";done
>>> 192.168.66.31
Active: inactive (dead)
>>> 192.168.66.32
Active: inactive (dead)
```

- 配置JAVA环境

把下载的hadoop、JDK解压到/usr/local

```
[root@master ~]# tar -xf hadoop-3.1.2.tar.gz -C /usr/local/
[root@master ~]# tar -xf jdk-8u201-linux-x64.tar.gz -C /usr/local/
```

添加环境变量最后的效果：

```
[root@master ~]# vim /etc/profile
#java path
export JAVA_HOME=/usr/local/jdk1.8.0_201
export JRE_HOME=$JAVA_HOME/jre
export CLASSPATH=.:$JAVA_HOME/lib:$JRE_HOME/lib
export PATH=$JAVA_HOME/bin:$PATH
#hadoop path
export PATH=$PATH:/usr/local/hadoop-3.1.2/bin:/usr/local/hadoop-3.1.2/sbin
export HDFS_NAMENODE_USER="root"
export HDFS_DATANODE_USER="root"
export HDFS_SECONDARYNAMENODE_USER="root"
export YARN_RESOURCEMANAGER_USER="root"
export YARN_NODEMANAGER_USER="root"
```

环境变量生效

```
[root@master ~]# source /etc/profile
测试JAVA是否可以使用
[root@master ~]# java -version
```

- 接下来是Hadoop配置文件的配置

```
[root@master ~]# cd /usr/local/hadoop-3.1.2/etc/hadoop/
```

### 1) 配置hadoop-env.sh

```
[root@master hadoop]# vim hadoop-env.sh
# The java implementation to use.
#export JAVA_HOME=${JAVA_HOME}
export JAVA_HOME=/usr/jdk1.8.0_121 //输入Java安装路径
```

### 2) 配置core-site.xml

注意修改第4行代码中的master，根据自身服务器主机名，修改为自己的主机名字

```
[root@master hadoop]# vim core-site.xml
<configuration>
  <property>
```

```

    <name>fs.default.name</name>
    <value>hdfs://master:9000</value>
    <description>HDFS的URI, 文件系统://namenode标识:端口号</description>
</property>
<property>
    <name>hadoop.tmp.dir</name>
    <value>/usr/hadoop/tmp</value>
    <description>namenode上本地的hadoop临时文件夹</description>
</property>
</configuration>

```

```

10 Unless required by applicable law or agreed to in writing, software
11 distributed under the License is distributed on an "AS IS" BASIS,
12 WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
13 See the license for the specific language governing permissions and
14 limitations under the License. See accompanying LICENSE file.
15 -->
16
17 <!-- Put site-specific property overrides in this file. -->
18
19 <configuration>
20 <property>
21     <name>fs.default.name</name>
22     <value>hdfs://master:9000</value>
23     <description>HDFS的URI, 文件系统://namenode标识:端口号</descript
ion>
24 </property>
25 <property>
26     <name>hadoop.tmp.dir</name>
27     <value>/usr/hadoop/tmp</value>
28     <description>namenode上本地的hadoop临时文件夹</description>
29 </property>
30 </configuration>

```

### 3) , 配置hdfs-site.xml

注意标红处为Hadoop存放数据的文件, 需要创建, 根据自身创建的目录填写;

[root@master hadoop]# mkdir /usr/hadoop //其余的目录会自动生成

[root@master hadoop]# vim hdfs-site.xml

注意修改代码中的master, 根据自身服务器主机名, 修改为自己的主机名字, 标红的IP改为自己的IP地址

```

<configuration>
  <property>
    <name>dfs.name.dir</name>
    <value>/usr/hadoop/hdfs/name</value>
    <description>namenode上存储hdfs名字空间元数据 </description>
  </property>
  <property>
    <name>dfs.data.dir</name>
    <value>/usr/hadoop/hdfs/data</value>
    <description>datanode上数据块的物理存储位置</description>
  </property>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
    <description>副本个数, 配置默认是3,应小于datanode机器数量</description>
  </property>
  <property>
    <name>dfs.namenode.secondary.http-address</name>
    <value>master:9001</value>
  </property>
  <property>
    <name>dfs.http.address</name>

```

```
    <value>192.168.66.30:50070</value>
  </property>
</configuration>
```

## 5) 配置mapred-site.xml

复制一份出来，再进行修改

```
[root@master hadoop]# cp mapred-site.xml mapred-site.xml_bak
```

注意修改代码中的master（两处）根据自身服务器主机名，修改为自己的主机名字

```
[root@master hadoop]# vim mapred-site.xml
```

```
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
  <property>
    <name>mapreduce.jobhistory.address</name>
    <value>master:10020</value>
  </property>
  <property>
    <name>mapreduce.jobhistory.webapp.address</name>
    <value>master:19888</value>
  </property>
  <property>
    <name>mapreduce.jobhistory.done-dir</name>
    <value>/history/done</value>
  </property>
  <property>
    <name>mapreduce.jobhistory.intermediate-done-dir</name>
    <value>/history/done_intermediate</value>
  </property>
</configuration>
```

## 6) 配置yarn-site.xml

注意修改代码中的master和倒数第三行的ip地址(:8099不要删除)，根据自身服务器主机名，修改为自己的主机名字，标红的IP改为自己的IP地址

```
<configuration>
  <property>
    <name>yarn.resourcemanager.hostname</name>
    <value>master</value>
  </property>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.resourcemanager.webapp.address</name>
    <value>192.168.66.30:8099</value>
  </property>
</configuration>
```

- 配置slave

```
[root@master ~]# vim /usr/local/hadoop-3.1.2/etc/hadoop/slave
```

将配置文件和二进制文件传给slave

```
[root@master ~]# vim scp_hadoop.sh
#!/bin/bash
slave_ip=("192.168.66.31" "192.168.66.32")
for i in $slave_ip;
do
  echo ">>> $i"
  rsync -av /usr/local/hadoop-3.1.2 $i:/usr/local/
  rsync -av /etc/profile $i:/etc/
  rsync -av /etc/hosts $i:/etc/
  rsync -av /usr/hadoop $i:/usr/
done
```

注意三台虚拟机的配置文件一样的, !!无须!! 将master修改为slave1, slave2;SSH连接到slave1、slave2, 生效环境变量文件

```
[root@master ~]# ssh slave1
[root@slave1 ~]# source /etc/profile
[root@slave1 ~]# exit
```

- 启动Hadoop: (以下操作在第一台master虚拟机中执行)

- 1) 格式化namenode:

```
[root@master ~]# hdfs namenode -format
```

注意:

- 1、格式化会在我们之前手动创建的目录里创建新的目录, 要保证当前用户对之前创建的目录有x权利

- 2) 启动NameNode 和 DataNode 守护进程

```
[root@master ~]# start-dfs.sh
```

- 3) 启动ResourceManager 和 NodeManager 守护进程

```
[root@master ~]# start-yarn.sh
```

- 启动验证:

- 1) 执行jps命令, 有如下进程, 说明Hadoop正常启动

```
[root@master ~]# /usr/hadoop/hadoop-2.7.3$ jps
```

```
8690 SecondaryNameNode
9300 Jps
8484 DataNode
9014 NodeManager
8874 ResourceManager
8319 NameNode
```