



链滴

# Hadoop 单机版环境搭建 - 在云服务器上

作者: [king1234](#)

原文链接: <https://ld246.com/article/1658637307573>

来源网站: [链滴](#)

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

# 一、前置条件

Hadoop 的运行依赖 JDK，需要预先安装，安装步骤见：

[Linux手动安装jdk](#)

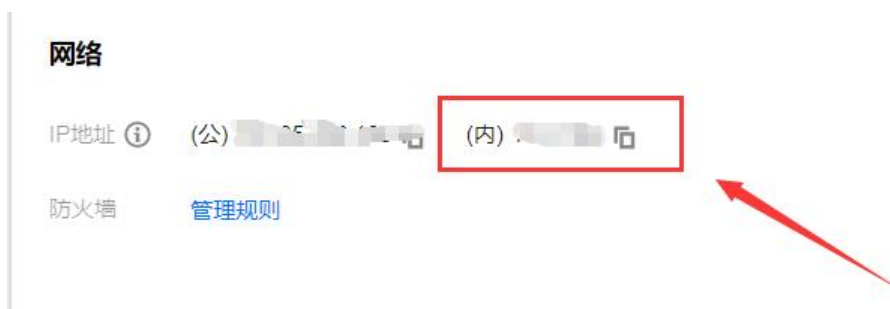
# 二、配置免密登录

Hadoop 组件之间需要基于 SSH 进行通讯。

## 2.1 配置映射

配置 ip 地址和主机名映射：

ip地址是你的内网ip 云服务器可以在服务器商的控制台看



虚拟机的一般就是ifconfig后192开头的

```
vim /etc/hosts
# 文件末尾增加
192.168.43.202 node
```

## 2.2 生成公私钥

执行下面命令行生成公匙和私匙：

```
ssh-keygen -t rsa
```

## 3.3 授权

进入 `~/.ssh` 目录下，查看生成的公匙和私匙，并将公匙写入到授权文件：

```
[root@VM-16-8-centos hadoop]# cd ~/.ssh
[root@VM-16-8-centos .ssh]# ll
total 8
-rw----- 1 root root 1679 Jul  5 16:21 id_rsa
-rw-r--r-- 1 root root 401 Jul  5 16:21 id_rsa.pub
```

写入公匙到授权文件

```
[root@VM-16-8-centos .ssh]# cat id_rsa.pub >> authorized_keys
[root@VM-16-8-centos .ssh]# chmod 600 authorized_keys
```

## 三、Hadoop(HDFS)环境搭建

### 3.1 下载并解压

下载 Hadoop 安装包，这里我下载的是 CDH 版本的，下载地址为：<https://mirrors.tuna.tsinghua.edu.cn/apache/hadoop/common/hadoop-2.10.1/hadoop-2.10.1.tar.gz>

```
# 下载
wget https://mirrors.tuna.tsinghua.edu.cn/apache/hadoop/common/hadoop-2.10.1/hadoop-2.10.1.tar.gz
# 解压
tar -zxvf hadoop-2.10.1.tar.gz
# 重命名
mv hadoop-2.10.1 hadoop2101
```

### 3.2 配置环境变量

```
vim /etc/profile
```

配置环境变量：

```
export HADOOP_HOME=/king/hadoop/hadoop2101
export PATH=${HADOOP_HOME}/bin:$PATH
```

执行 `source` 命令，使得配置的环境变量立即生效：

```
source /etc/profile
```

### 3.3 修改Hadoop配置

进入 `${HADOOP_HOME}/etc/hadoop/` 目录下，修改以下配置：

#### 1. hadoop-env.sh

```
# JDK安装路径
export JAVA_HOME=/usr/java/jdk1.8.0_201/
```

#### 2. core-site.xml

```
<configuration>
  <property>
    <!--指定 namenode 的 hdfs 协议文件系统的通信地址-->
    <name>fs.defaultFS</name>
    <value>hdfs://node:8020</value>
  </property>
  <property>
    <!--指定 hadoop 存储临时文件的目录-->
    <name>hadoop.tmp.dir</name>
    <value>/home/hadoop/tmp</value>
  </property>
</configuration>
```

### 3. hdfs-site.xml

指定副本系数和临时文件存储位置：

```
<configuration>
  <property>
    <!--由于我们这里搭建是单机版本，所以指定 dfs 的副本系数为 1-->
    <name>dfs.replication</name>
    <value>1</value>
  </property>
</configuration>
```

### 4. slaves

配置所有从属节点的主机名或 IP 地址，由于是单机版本，所以指定本机即可：

```
node
```

### 3.4 关闭防火墙

不关闭防火墙可能导致无法访问 Hadoop 的 Web UI 界面：

```
# 查看防火墙状态
sudo firewall-cmd --state
# 关闭防火墙:
sudo systemctl stop firewalld.service
```

### 3.5 初始化

第一次启动 Hadoop 时需要进行初始化，进入 `${HADOOP_HOME}/bin/` 目录下，执行以下命令：

```
./hdfs namenode -format
```

### 3.6 启动HDFS

进入 `${HADOOP_HOME}/sbin/` 目录下，启动 HDFS：

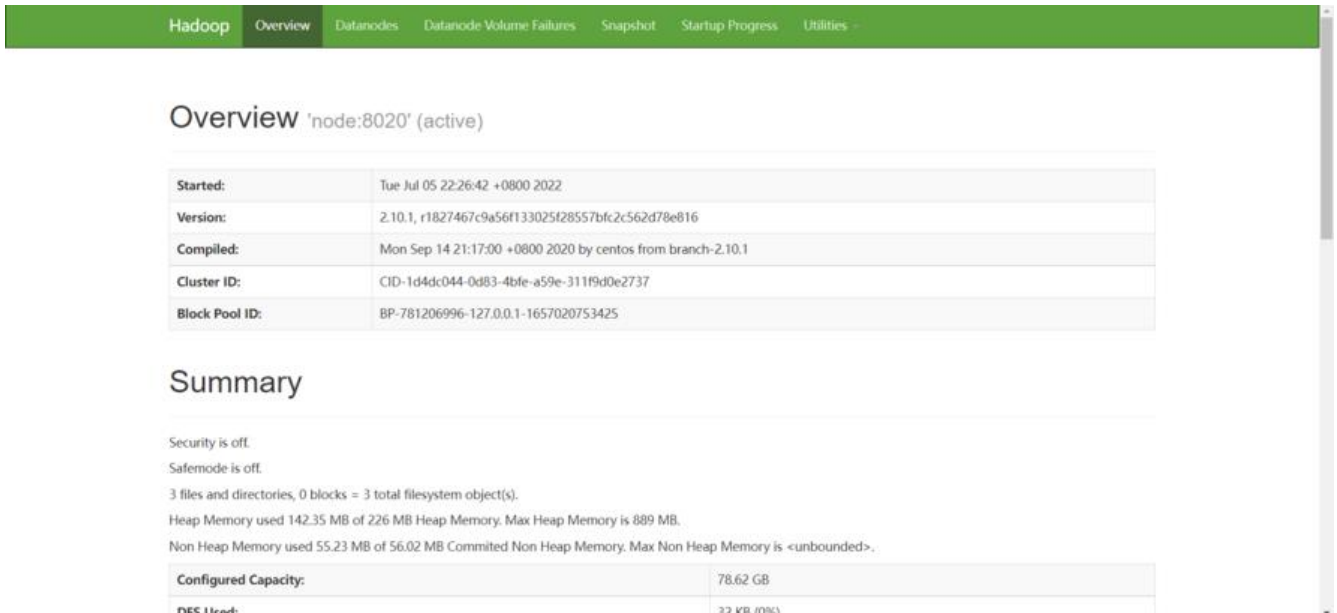
```
./start-dfs.sh
```

### 3.7 验证是否启动成功

方式一：执行 `jps` 查看 `NameNode` 和 `DataNode` 服务是否已经启动：

```
9137 DataNode
9026 NameNode
9390 SecondaryNameNode
```

方式二：查看 Web UI 界面，端口为 `50070`：



`<div align="center"></div>`

方式三：创建一个文件夹看看

```
[root@VM-16-8-centos sbin]# hdfs dfs -mkdir /winnie
[root@VM-16-8-centos sbin]# hdfs dfs -ls -R /
drwxr-xr-x - root supergroup      0 2022-07-05 22:43 /winnie
```

## 四、Hadoop(YARN)环境搭建

### 4.1 修改配置

进入 `$(HADOOP_HOME)/etc/hadoop/` 目录下，修改以下配置：

#### 1. mapred-site.xml

如果没有 `mapred-site.xml`，则拷贝一份样例文件后再修改

```
cp mapred-site.xml.template mapred-site.xml
```

```
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
</configuration>
```

#### 2. yarn-site.xml

```
<configuration>
  <property>
    <!--配置 NodeManager 上运行的附属服务。需要配置成 mapreduce_shuffle 后才可以 Yarn 上运行 MapReduce 程序。-->
```

```
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
</property>
</configuration>
```

## 4.2 启动服务

进入 `#{HADOOP_HOME}/sbin/` 目录下，启动 YARN：

```
./start-yarn.sh
```

## 4.3 验证是否启动成功

方式一：执行 `jps` 命令查看 NodeManager 和 ResourceManager 服务是否已经启动：

```
9137 DataNode
9026 NameNode
12294 NodeManager
12185 ResourceManager
9390 SecondaryNameNode
```

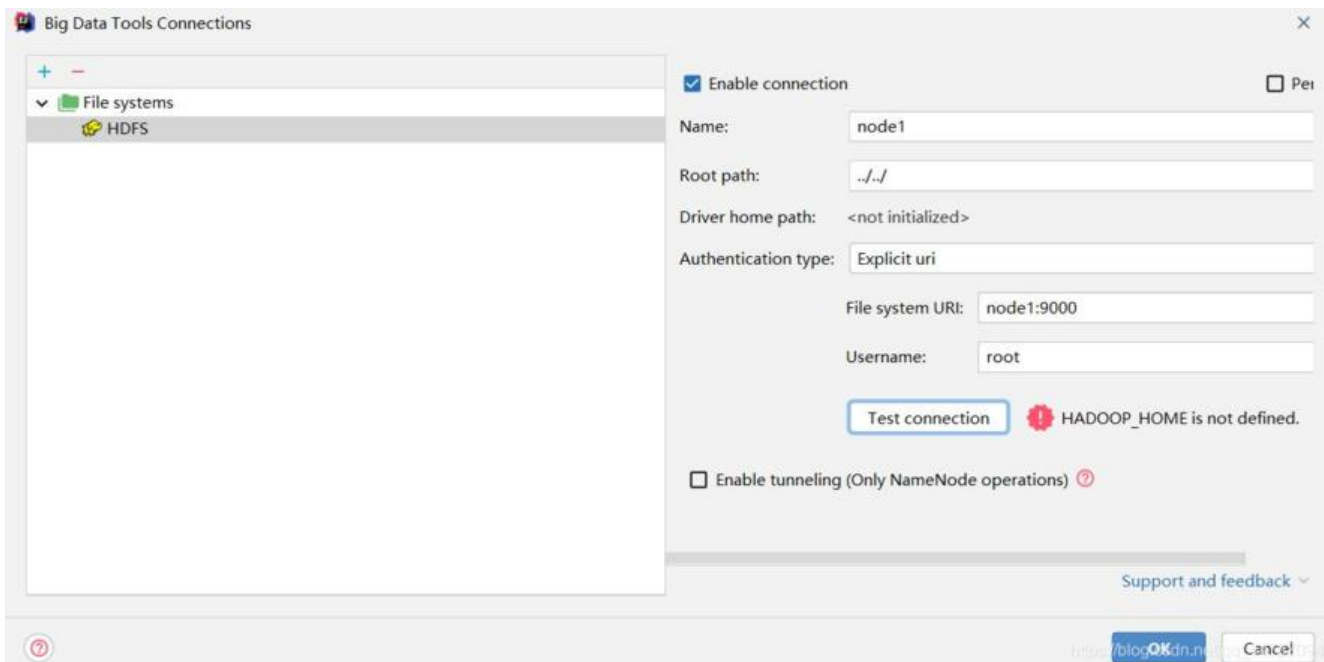
方式二：查看 Web UI 界面，端口号为 8088：

The screenshot shows the Hadoop Web UI interface. On the left is a navigation menu with options like 'Cluster', 'About Nodes', 'Node Labels', 'Applications', and 'Tools'. The main content area is titled 'All Applications' and contains several sections:

- Cluster Metrics:** A table with columns: Apps Submitted (0), Apps Pending (0), Apps Running (0), Apps Completed (0), Containers Running (0), Used Resources (<memory:0, vCores:0>), and Total Res (<memory:8192, vCo...).
- Cluster Nodes Metrics:** A table with columns: Active Nodes (1), Decommissioning Nodes (0), Decommissioned Nodes (0), Lost Nodes (0), Unhealthy Nodes (0), and Rel...
- Scheduler Metrics:** Shows Scheduler Type as 'Capacity Scheduler' and Scheduling Resource Type as '<name=memory-mb default-unit=Mi type=COUNTABLE>, <name=vcores default-unit= type=COUNTABLE>'. It also shows Minimum Allocation (<memory:1024, vCores:1>) and Maximum A (<memory:8192, vCores:1>).
- Applications Table:** A table with columns: ID, User, Name, Application Type, Queue, Application Priority, StartTime, LaunchTime, FinishTime, State, FinalStatus, Running Containers, Allocated CPU, Allocated Memory MB, Allocated GPUs, Reserved CPU, Reserved Memory MB, and Reser GPU. The table is currently empty, displaying 'No data available in table' and 'Showing 0 to 0 of 0 entries'.

## 五、idea Big Data Tools 插件连接

像这样直接连接会有错误



报错如下

On Windows you should have HADOOP\_HOME environment variable defined or Java propert hadoop.home.dir. Please, refer to [Hadoop Wiki](#) for more details

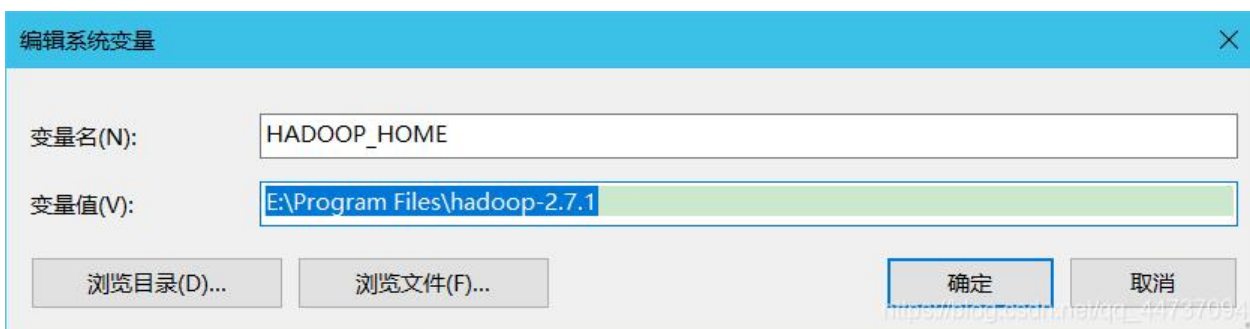
需要下载

[微云](#)

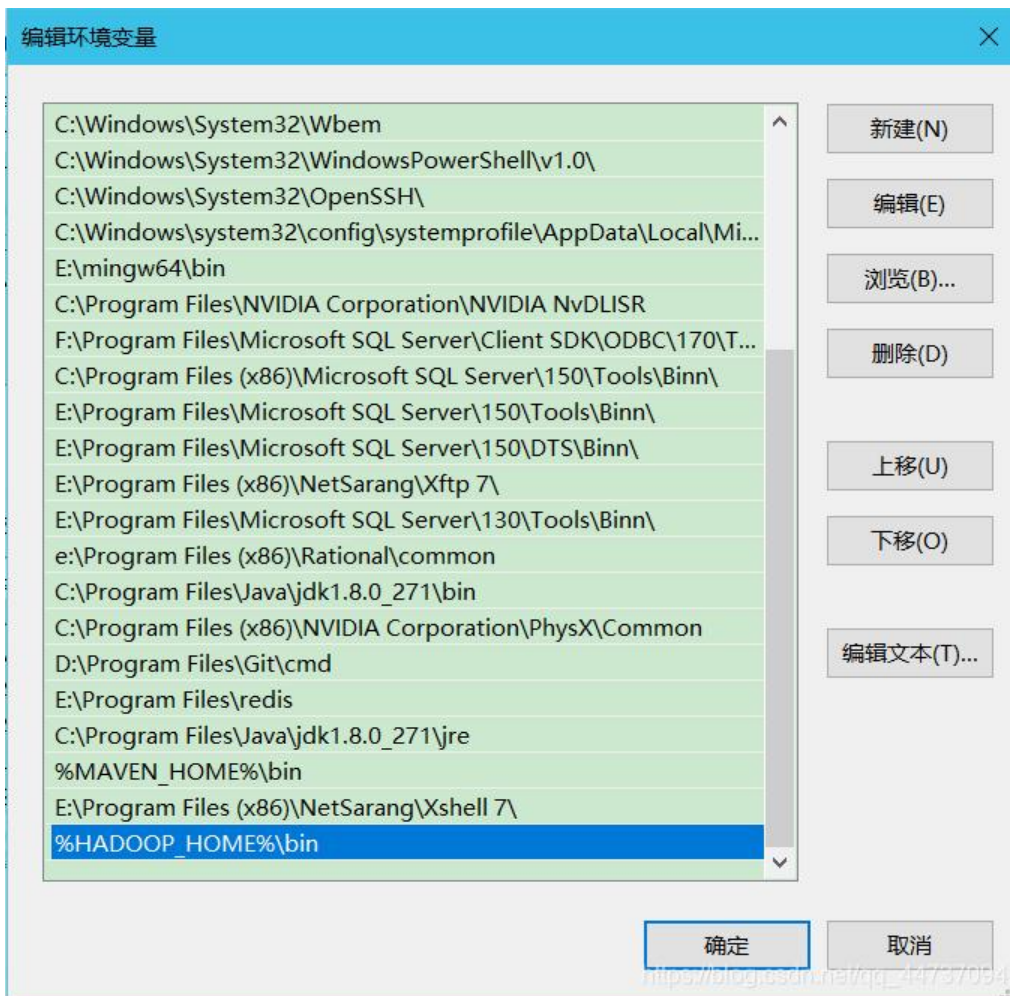
[github](#)

选择自己安装的hadoop版本对应的文件夹，解压然后配置环境

HADOOP\_HOME

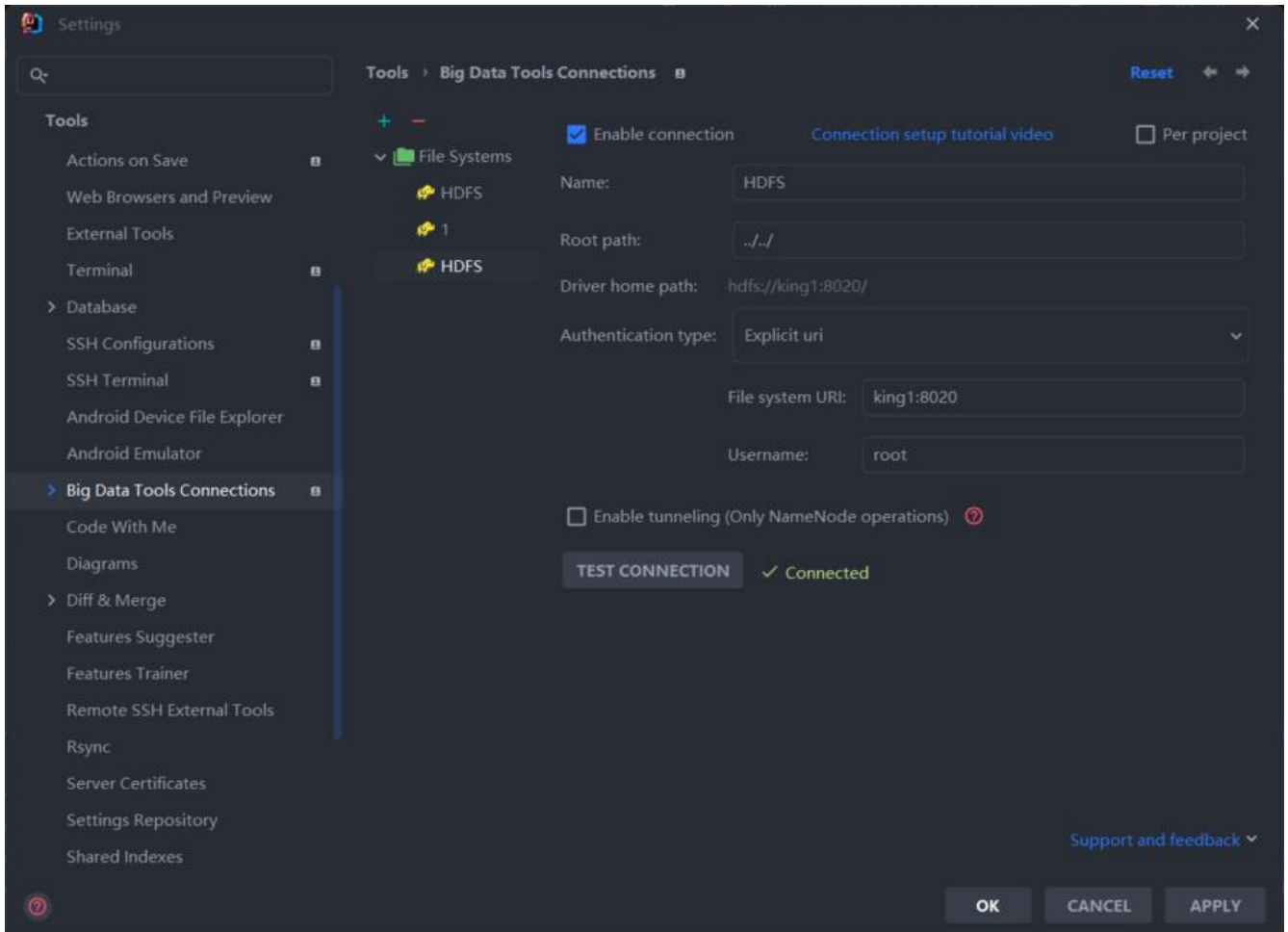


PATH



然后连接就会成功





连接不上可以试试先创建一个文件夹