

# 第六节、Hadoop集群搭建

## 第六节、Hadoop集群搭建

### 一、JDK安装

1. 解压缩
2. 环境变量配置
3. 刷新环境变量

### 二、Hadoop安装

1. 解压缩
2. 环境变量配置
3. 刷新环境变量

### 三、Hadoop配置

1. core-site.xml
2. hdfs-site.xml
3. mapred-site.xml(重命名mapred-site.xml.template)
4. yarn-site.xml
5. hadoop-env.sh
6. slaves
7. 使用SSH拷贝文件

### 四、Hadoop启动

1. 格式化namenode
2. 启动Hadoop

### 五、Hadoop命令

## 一、JDK安装

### 1. 解压缩

```
1. tar -zxvf jdk-7u67-linux-x64.tar.gz
```

### 2. 环境变量配置

1. `export JAVA_HOME=/home/hadoopadmin/jdk1.7.0_67`
2. `export PATH=$PATH:$JAVA_HOME/bin`
3. `export CLASSPATH=.:$JAVA_HOME/lib/dt.jar:$JAVA_HOME/lib/tools.jar`

### 3. 刷新环境变量

1. `source /home/hadoopadmin/.bash_profile`

## 二、Hadoop安装

### 1. 解压缩

1. `tar -zxvf hadoop-2.7.1.tar.gz`

### 2. 环境变量配置

1. `export HADOOP_HOME=/home/hadoopadmin/hadoop-2.7.1`
2. `export PATH=$PATH:$HADOOP_HOME/bin`

### 3. 刷新环境变量

1. `source /home/hadoopadmin/.bash_profile`

## 三、Hadoop配置

配置文件目录：`$HADOOP_HOME/etc/hadoop/`

### 1. core-site.xml

```

1. <configuration>
2.     <!-- Hadoop文件系统依赖的基础配置 -->
3.     <property>
4.         <name>hadoop.tmp.dir</name>
5.         <value>/home/hadoopadmin/hadoop/data</value>
6.     </property>
7.     <!-- NameNode结点的URI (包括协议、主机名称、端口号) -->
8.     <property>
9.         <name>fs.defaultFS</name>
10.        <value>hdfs://etc01:8020</value>
11.    </property>
12.    <!-- 开启回收站机制，可以设置文件彻底删除的时间，默认为0，单位为分钟 -->
13.    <property>
14.        <name>fs.trash.interval</name>
15.        <value>60</value>
16.    </property>
17. </configuration>

```

## 2. hdfs-site.xml

```

1. <configuration>
2.     <!-- secondarynamenode的http服务器地址和端口 -->
3.     <property>
4.         <name>dfs.namenode.secondary.http-address</name>
5.         <value>etc01:50090</value>
6.     </property>
7.     <!-- 默认块复制 -->
8.     <property>
9.         <name>dfs.replication</name>
10.        <value>3</value>
11.    </property>
12.    <!-- 关闭权限校验 -->
13.    <property>
14.        <name>dfs.permissions</name>
15.        <value>false</value>
16.    </property>
17.    <!-- namenode的http服务器地址和端口 -->
18.    <property>
19.        <name>dfs.namenode.http-address</name>
20.        <value>etc01:50070</value>
21.    </property>
22.    <!-- datanode结点被指定要存储数据的本地文件系统路径 -->
23.    <property>
24.        <name>dfs.datanode.data.dir</name>
25.        <value>file:///home/hadoopadmin/hadoop/data/dfs/dn</value>
26.    </property>
27. </configuration>

```

### 3. mapred-site.xml(重命名mapred-site.xml.template)

```
1. <configuration>
2.     <!-- MapReduce JobHistory进程通信主机、端口 -->
3.     <property>
4.         <name>mapreduce.jobhistory.address</name>
5.         <value>etc01:10020</value>
6.     </property>
7.     <!-- MapReduce JobHistory的web界面主机、端口 -->
8.     <property>
9.         <name>mapreduce.jobhistory.webapp.address</name>
10.        <value>etc01:19888</value>
11.    </property>
12.    <!-- 以yarn方式运行MapReduce -->
13.    <property>
14.        <name>mapreduce.framework.name</name>
15.        <value>yarn</value>
16.    </property>
17. </configuration>
```

### 4. yarn-site.xml

---

```

1. <configuration>
2.     <!-- resourcemanager的主机名 -->
3.     <property>
4.         <name>yarn.resourcemanager.hostname</name>
5.         <value>etc01</value>
6.     </property>
7.     <!-- resourcemanager提供给nodemanager的地址 -->
8.     <property>
9.         <name>yarn.resourcemanager.resource-tracker.address</name>
10.        <value>etc01:8031</value>
11.    </property>
12.    <!-- resourcemanager中应用程序管理器界面的地址 -->
13.    <property>
14.        <name>yarn.resourcemanager.address</name>
15.        <value>etc01:8032</value>
16.    </property>
17.    <!-- 调度器接口的地址 -->
18.    <property>
19.        <name>yarn.resourcemanager.scheduler.address</name>
20.        <value>etc01:8030</value>
21.    </property>
22.    <!-- 分配给容器的物理内存量（75%） -->
23.    <property>
24.        <name>yarn.nodemanager.resource.memory-mb</name>
25.        <value>1536</value>
26.    </property>
27.    <!-- NodeManager上运行的附属服务,配置成mapreduce_shuffle才可运行MR -->
28.    <property>
29.        <name>yarn.nodemanager.aux-services</name>
30.        <value>mapreduce_shuffle</value>
31.    </property>
32. </configuration>

```

## 5. hadoop-env.sh

```

1. export JAVA_HOME=/home/hadoopadmin/jdk1.7.0_67

```

## 6. slaves

```

1. etc01
2. etc02
3. etc03

```

## 7. 使用SSH拷贝文件

可将一下命令以脚本方式执行，执行前确保SSH配置正确

```
1. scp -r /home/hadoopadmin/jdk1.7.0_67 hadoopadmin@etc02:/home/hadoopadmin
2. scp -r /home/hadoopadmin/jdk1.7.0_67 hadoopadmin@etc03:/home/hadoopadmin
3. scp -r /home/hadoopadmin/hadoop-2.7.1 hadoopadmin@etc02:/home/hadoopadmin
4. scp -r /home/hadoopadmin/hadoop-2.7.1 hadoopadmin@etc03:/home/hadoopadmin
5. scp /home/hadoopadmin/.bash_profile hadoopadmin@etc02:/home/hadoopadmin
6. scp /home/hadoopadmin/.bash_profile hadoopadmin@etc03:/home/hadoopadmin
```

## 四、Hadoop启动

### 1. 格式化namenode

```
1. cd /home/hadoopadmin/hadoop-2.7.1/bin
2. chmod 744 *
3. hdfs namenode -format
```

### 2. 启动Hadoop

```
1. cd /home/hadoopadmin/hadoop-2.7.1/sbin
2. chmod 744 *
3. ./start-all.sh
```

## 五、Hadoop命令

- cat : 查看文本文件内容

```
1. hadoop fs -cat URI
```

- chgrp : 更改文件属组

```
1. hadoop fs -chgrp [-R] GROUP URI
```

- chmod : 更改文件权限

```
1. hadoop fs -chmod [-R] <MODE[,MODE]... | OCTALMODE> URI
```

- chown : 改变属主和属组

```
1. hadoop fs -chown [-R] [OWNER][:[GROUP]] URI
```

- cp : 复制文件

```
1. hadoop fs -cp [-f] URI [URI ...]
```

- get : 拷贝文件到本地

```
1. hadoop fs -get [-ignorecrc] [-crc] <src> <localdst>
```

- ls : 查看目录

```
1. hadoop fs -ls [-d] [-h] [-R] <args>
```

- mkdir : 创建目录

```
1. hadoop fs -mkdir [-p] <paths>
```

- mv : 移动文件, 更名, 允许多个源, 一个目标

```
1. hadoop fs -mv URI [URI ...] <dest>
```

- put : 上传文件

```
1. hadoop fs -put <localsrc> ... <dst>
```

- rm : 删除文件 (不放入回收站)

```
1. hadoop fs -rm [-f] [-r | -R] [-skipTrash] URI
```

- rmdir : 删除目录
-

```
1. hadoop fs -rmkdir [--ignore-fail-on-non-empty] URI
```