**Author:** 张小优

**Version:** 13

**Date:** 2017.05.23

**AiSchool需求文档空间**

**[AIREQ-1219]【需求】分布式文件存储-hdfs**

0

Table of ContenTs

[1 需求描述 3](#_Toc256000000)

[1.1 提出背景 3](#_Toc256000001)

[1.2 面向角色 3](#_Toc256000002)

[2 名词术语 4](#_Toc256000003)

[3 参考文档 5](#_Toc256000004)

[3.1 环境配置 5](#_Toc256000005)

[3.1.1 修改主机名 5](#_Toc256000006)

[3.1.2 安装jdk 5](#_Toc256000007)

[3.1.3 设置3台机器相互免密访问 5](#_Toc256000008)

[3.2 安装hadoop 6](#_Toc256000009)

[3.2.1 安装配置 6](#_Toc256000010)

[3.3 NFS挂载HDFS到本地 7](#_Toc256000011)

[3.3.1 配置core-site.xml 7](#_Toc256000012)

[3.3.2 修改hdfs-site.xml 8](#_Toc256000013)

[3.3.3 重启hdfs服务 8](#_Toc256000014)

[3.3.4 关闭系统自带nfs服务 8](#_Toc256000015)

[3.3.5 启动hadoop nf s 8](#_Toc256000016)

[3.3.6 挂载 8](#_Toc256000017)

[3.3.7 安装部署测试建议 8](#_Toc256000018)

[4 工作量预估 9](#_Toc256000019)

* [需求描述](#scroll-bookmark-1)
  + [提出背景](#scroll-bookmark-2)
  + [面向角色](#scroll-bookmark-3)
* [名词术语](#scroll-bookmark-4)
* [参考文档](#scroll-bookmark-5)
  + [环境配置](#scroll-bookmark-6) 
    - [修改主机名](#scroll-bookmark-7)
    - [安装jdk](#scroll-bookmark-8)
    - [设置3台机器相互免密访问](#scroll-bookmark-9)
  + [安装hadoop](#scroll-bookmark-10)
    - [安装配置](#scroll-bookmark-11)
  + [NFS挂载HDFS到本地](#scroll-bookmark-12)
    - [配置core-site.xml](#scroll-bookmark-13)
    - [修改hdfs-site.xml](#scroll-bookmark-14)
    - [重启hdfs服务](#scroll-bookmark-15)
    - [关闭系统自带nfs服务](#scroll-bookmark-16)
    - [启动hadoop nfs](#scroll-bookmark-17)
    - [挂载](#scroll-bookmark-18)
    - [安装部署测试建议](#scroll-bookmark-19)
* [工作量预估](#scroll-bookmark-20)

需求描述

HDFS的集群部署、API使用说明。

提出背景

*//todo*

面向角色

学生 　　老师 　　家长　 　校务人员 　　其它



名词术语

|  |  |  |
| --- | --- | --- |
| 标准名 | 简称 | 解释说明 |
|  |  |  |

参考文档

|  |  |  |  |
| --- | --- | --- | --- |
| 编号 | 文档、资料名称 | 来源 | 备注 |
|  |  |  |  |

环境配置

以hadoop-2.8.0版本为例，说明在3台机器上如何进行集群部署。

### 修改主机名

1. 3台示例机器分别为：192.168.102.53、192.168.50.16、192.168.102.80
2. 分别修改这3台机器的/etc/hosts文件，将3台机器的hosts文件均添加如下内容：

192.168.102.53 node1  
192.168.50.16 node2  
192.168.102.80 node3

1. 修改完后，需要重启机器，如果重启后未生效，需要查看/etc/sysconfig/network文件中是否有hostname，如有也需要一并修改

### 安装jdk

1. 在3台机器上均使用root用户将jdk1.8.0\_102解压到/usr目录下，并配置环境变量
2. 编辑/etc/profile，增加如下环境变量：  
   export JAVA\_HOME=/usr/jdk1.8.0\_102  
   PATH=$JAVA\_HOME/bin:$PATH  
   export JAVA\_HOME PATH
3. 然后执行source /etc/profile

### 设置3台机器相互免密访问

1. 3台机器的ssh端口均使用默认22端口，使用其他端口，访问时会有问题，需要额外处理
2. 登录3台机器，均使用root用户执行：ssh-keygen -t rsa 生成RSA密钥文件，执行完后，按3次回车键即可
3. 将公钥存入authorized\_keys，执行：cat /root/.ssh/id\_rsa.pub >> /root/.ssh/authorized\_keys
4. 设置本机ssh免密登录，在node1机器上执行：ssh node1 ，输入yes，即完成ssh可以免密登录本机，执行完后需要退出当前会话：exit
5. 设置node1免密登录node2、node3，在node1上分别执行：

ssh-copy-id node2  
ssh-copy-id node3

1. 即完成node1免密登录node2、node3；在node2上分别设置node2免密登录node1、node3；在node3上分别设置免密登录node1、node2
2. 执行完后，可以在任意机器上执行ssh node1、ssh node2、ssh node3免密登录需要登录的机器

安装hadoop

说明hadoop的安装配置、启动

### 安装配置

1. 将hadoop-2.8.1.tar.gz上传到node1任意目录，将压缩包解压到/opt目录下，执行命令：tar -zxvf hadoop-2.8.1.tar.gz -C /opt
2. 将hadoop安装路径添加到/etc/profile文件末尾：

export HADOOP\_HOME=/opt/hadoop-2.8.1  
export PATH=$PATH:$HADOOP\_HOME/bin:$HADOOP\_HOME/sbin

1. 然后执行source /etc/profile
2. 在/opt/hadoop-2.8.1/etc/hadoop/hadoop-env.sh中添加export JAVA\_HOME=/usr/jdk1.8.0\_102，将原来的export JAVA\_HOME注释
3. 配置hadoop，先进入到node1的/opt/hadoop-2.8.1/etc/hadoop目录，将如下内容添加到core-site.xml文件中：

<configuration>  
<property>  
<name>fs.defaultFS</name>  
<value>hdfs://node1:9000</value>  
</property>  
<property>  
<name>hadoop.tmp.dir</name>  
<value>/var/data/hadoop</value>  
</property>  
<property>  
<name>io.file.buffer.size</name>  
<value>65536</value>  
</property>  
</configuration>

1. 将如下内容添加到 hdfs-site.xml文件中：

<configuration>  
<property>  
<name>dfs.replication</name>  
<value>3</value>  
</property>  
<property>  
<name>dfs.namenode.secondary.http-address</name>  
<value>node2:50090</value>  
</property>  
<property>  
<name>dfs.namenode.secondary.https-address</name>  
<value>node2:50091</value>  
</property>  
</configuration>

1. 将如下内容添加到slaves文件中：  
   node1  
   node2  
   node3
2. 将hadoop安装文件拷贝到node2、node3集群机器上：scp -r /opt/hadoop-2.8.1/ [node2:/opt](http://node2/opt) 、scp -r /opt/hadoop-2.8.1/ [node3:/opt](http://node2/opt)  
   然后分别将hadoop的路径添加到node2、node3的/etc/profile文件中，然后执行：source /etc/profile
3. 在node1上执行： hdfs namenode -format，然后在node1上启动hdfs，命令为：start-dfs.sh
4. 启动完成后，可以在每个节点上执行jps命令查看hdfs的启动情况，通过访问node1的地址：<http://192.168.102.53:50070/> 可以查看hdfs的访问界面。

NFS挂载HDFS到本地

### 配置core-site.xml

增加如下配置：

<property> <name>hadoop.proxyuser.nfsserver.groups</name> <value>\*</value> <description>允许所有用户组用户代理</description></property><property> <name>hadoop.proxyuser.nfsserver.hosts</name> <value>\*</value> <description>允许挂载的主机域名</description></property>

### 修改hdfs-site.xml

增加如下配置：

<property> <name>nfs.dump.dir</name> <value>/tmp/.hdfs-nfs</value></property><property> <name>nfs.rtmax</name> <value>1048576</value> <description>This is the maximum size in bytes of a READ request supported by the NFS gateway. If you change this, make sure you also update the nfs mount's rsize(add rsize= # of bytes to the mount directive).</description></property><property> <name>nfs.wtmax</name> <value>65536</value> <description>This is the maximum size in bytes of a WRITE request supported by the NFS gateway. If you change this, make sure you also update the nfs mount's wsize(add wsize= # of bytes to the mount directive).</description></property><property> <name>nfs.exports.allowed.hosts</name> <value>\* rw</value> <description>允许所有主机对文件有rw权限</description></property>

### 重启hdfs服务

见hadoop相关命令

### 关闭系统自带nfs服务

service nfs stopservice rpcbind stop

### 启动hadoop nf s

hadoop-daemon.sh start portmap  
hadoop-daemon.sh start nfs3

### 挂载

mkdir -p /data/twcontentftp #创建挂载的目标点  
mount -t nfs -o vers=3,proto=tcp,nolock,noacl hdfs namenode机器IP:/ /data/twcontentftp

### 安装部署测试建议

安装部署测试上需要注意那些？例如需要多个业务系统配合使用？以及安装部署时候需要注意什么？

工作量预估

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **序号** | **需求** | **设计** | **UI** | **开发** | **测试** | **工作量合计** |
| **1** | **功能特性** |  |  |  |  |  |
| 1.1 | 子功能特性 |  |  |  |  |  |
|  | **合计** |  |  |  |  |  |