**Hadoop安装实验报告**

141070010 杜文钦

一、系统安装运行情况

1. 创建hadoop用户

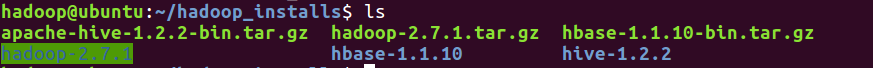
$sudo groupadd hadoop-user

$sudo useradd -m -g hadoop-user -d /home/hadoop hadoop

$sudo passwd hadoop

1. 解压安装Hadoop

下载解压hadoop-2.7.1到/home/hadoop\_installs/中

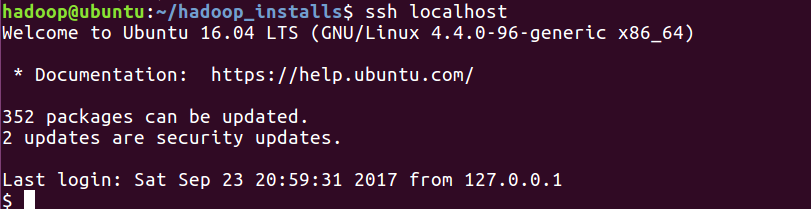


3.设置SSH免密码访问，登录Hadoop用户

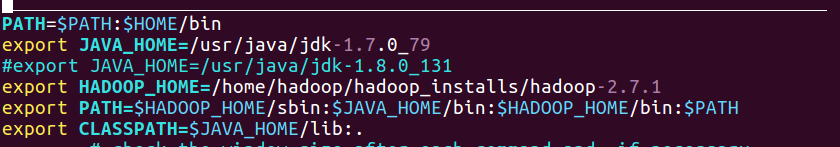
ssh-keygen -t rsa -P ""

cat ~/.ssh/id\_rsa.pub >> authorized\_keys

ssh localhost



1. 在~/.bashrc下设置环境变量



1. Hadoop伪分布式配置

配置 core-site.xml ,hdfs-site.xml,yarn-site.xml,mapred-site.xml

(1)core-site.xml配置:

<configuration

<property>

<name>hadoop.tmp.dir</name>

<value>file:/home/hadoop/hadoop\_installs/hadoop-2.7.1</value>

<description>Abase for other temporary directories.</description>

</property>

<property>

<name>fs.defaultFS</name>

<value>hdfs://localhost:9000</value>

</property>

</configuration>

(2)hdfs-site.xml配置:

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

<name>dfs.namenode.name.dir</name>

<value>file:/home/hadoop/hadoop\_installs/hadoop-2.7.3/tmp/dfs/name</value>

</property>

<property>

<name>dfs.datanode.data.dir</name>

<value>file:/home/hadoop/hadoop\_installs/hadoop-2.7.3/tmp/dfs/data</value>

</property>

</configuration>

1. yarn-site.xml配置

<configuration>

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

</configuration>

1. mapred-site.xml配置

需要先进行重命名，将mapred-site.xml.template重命名为mapred-site.xml。

<configuration>

<property>

<name>mapreduce.framework.name</name>

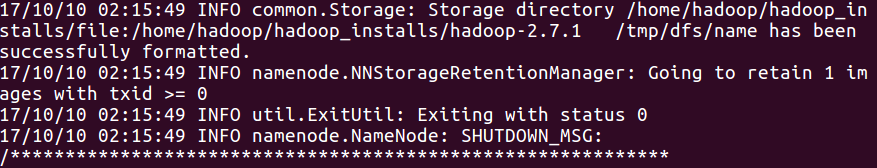
<value>yarn</value>

</property>

</configuration>

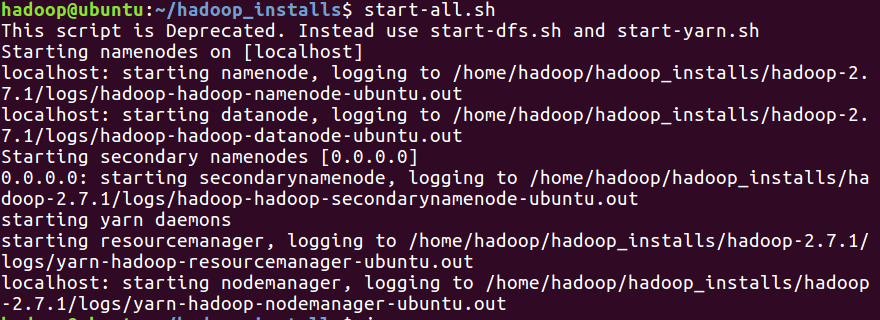
1. NameNode格式化

$hdfs namenode –format

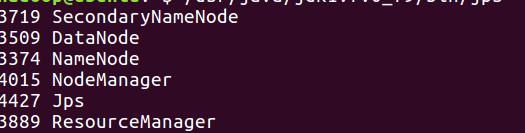


1. Hadoop启动

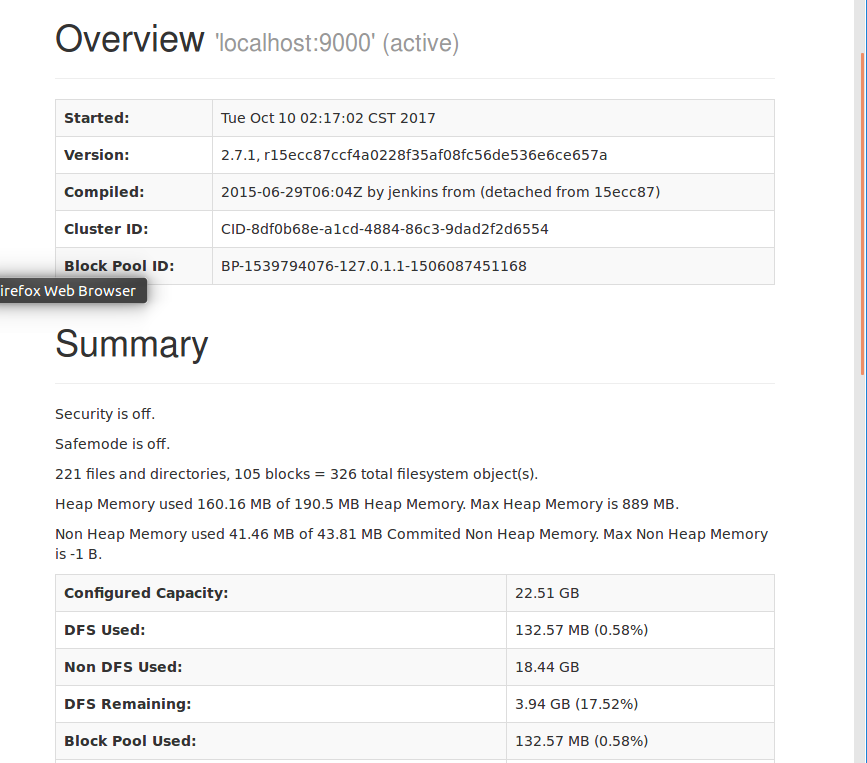
$ Start-all.sh



Jps查看进程



成功启动后，可以访问 Web 界面 [http://localhost:50070](http://localhost:50070/) 查看 NameNode 和 Datanode 信息，还可以在线查看 HDFS 中的文件。

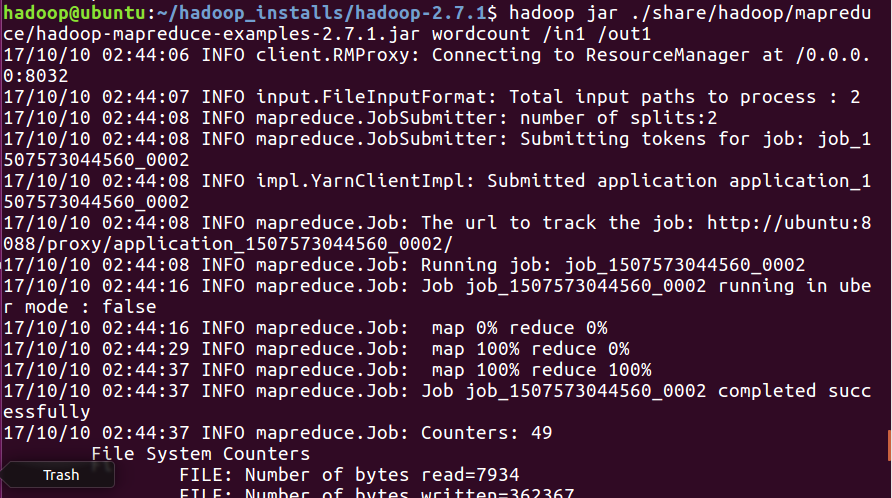


随便找了两个英文文档，上传到本地hdfs上：

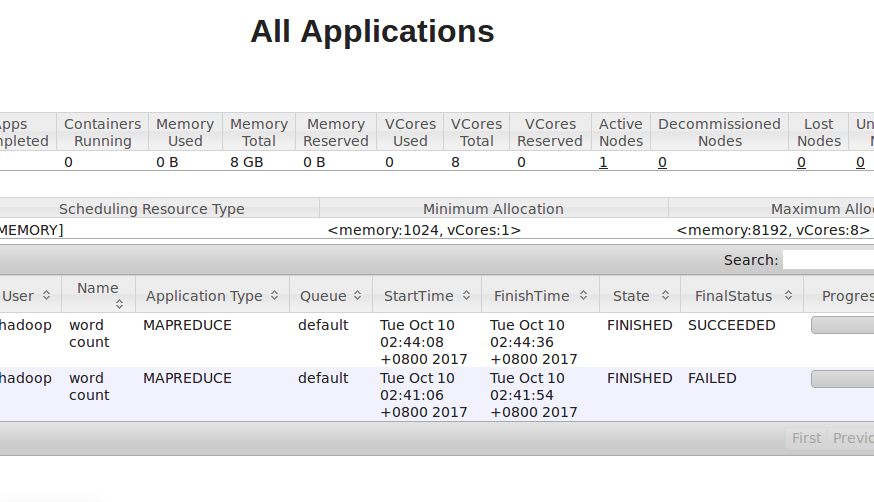




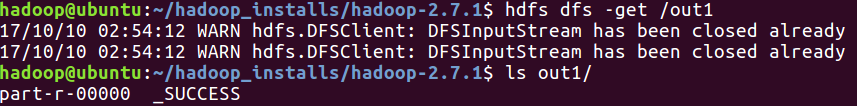
进行测试：

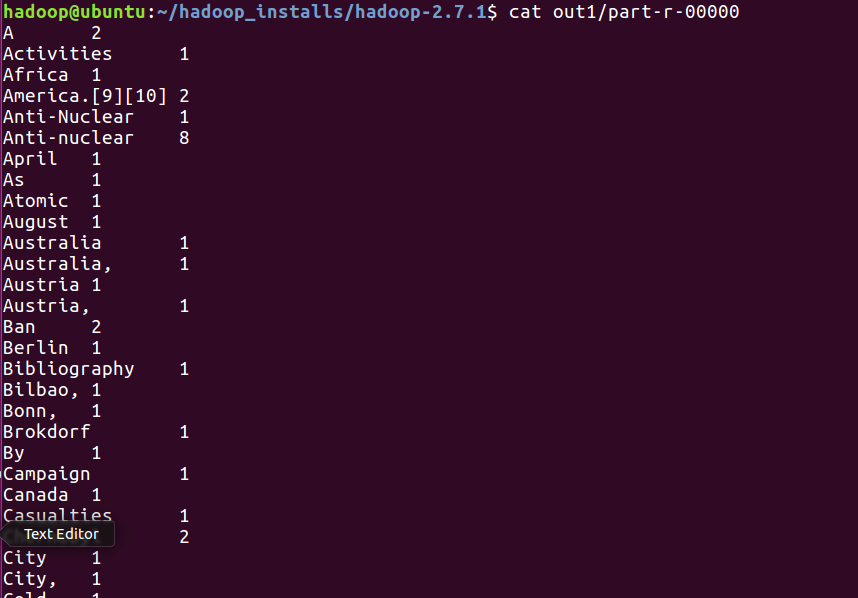


通过<http://localhost:8088/>查看运行情况：



复制运行结果到本地中，并查看结果：





1. 实验体会

本次实验掌握了Hadoop的安装配置，中间遇到了很多问题，新版本的hadoop安装指令与老师给的ppt中有差异，因此我参照了网上的博客和官方文档，解决了问题。这说明只有官方文档最靠谱。同时熟悉linux系统这点也很重要。