单机安装Hadoop系统与WordCount程序实验

141220138 臧凤云

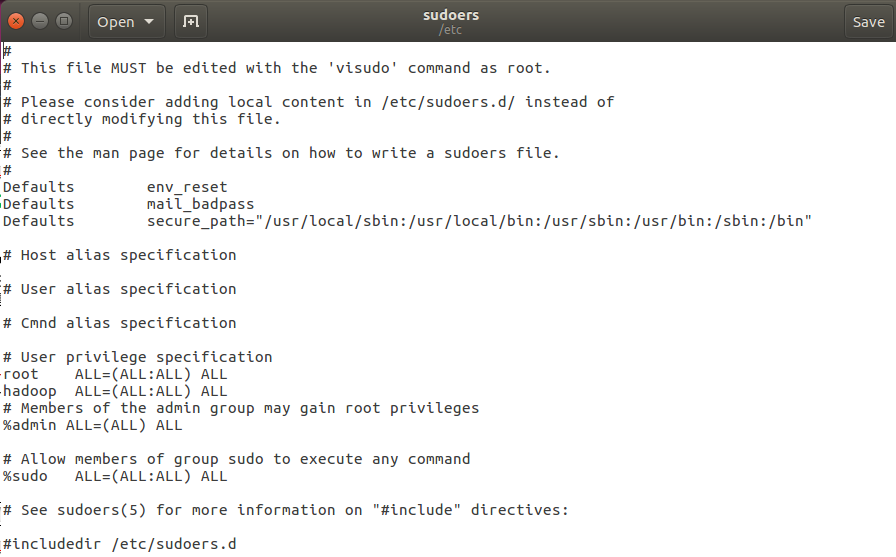
一．Hadoop系统安装运行情况

1. 创建用户

useradd Hadoop –m –s /bin/bash

passwd Hadoop

给新用户添加sudo权限：在/etc/sudoers文件中添加hadoop ALL=(ALL:ALL)ALL



新用户登陆：su - hadoop

1. 安装ssh：sudo apt-get install ssh

安装java：将Java安装在/usr/java下

安装Hadoop：建立安装目录：mkdir hadoop\_installs，将压缩包放在该文件夹下解压

1. 配置环境变量：

vim ~/.bash\_profile

在文件中添加以下

PATH=$PATH:$HOME/bin

export JAVA\_HOME=/usr/java/java-1.7

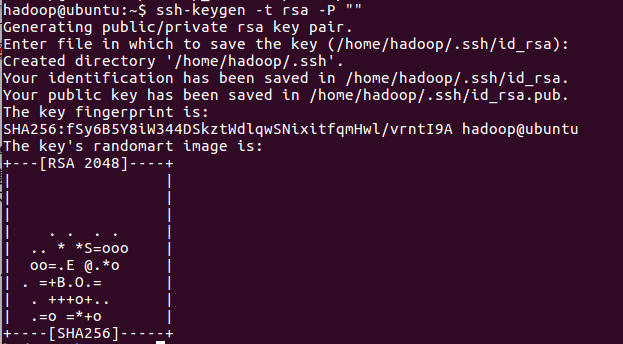
export HADOOP\_HOME=/home/hadoop/hadoop\_installs/hadoop-2.7.1

export PATH=$JAVA\_HOME/bin:$HADOOP\_HOME/bin:$PATH

export CLASSPATH=$JAVA\_HOME/lib:.

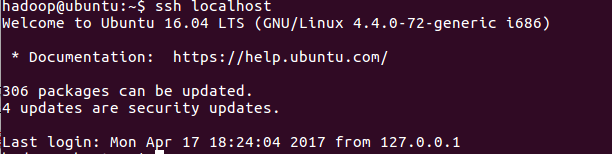
1. 免密码ssh访问配置

ssh-keygen -t rsa -P ‘’



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

sshlocalhost

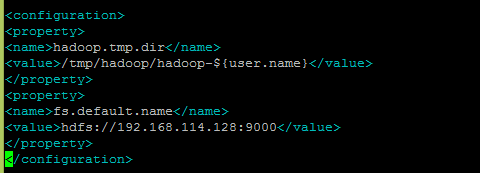


1. 修改hadoop配置文件

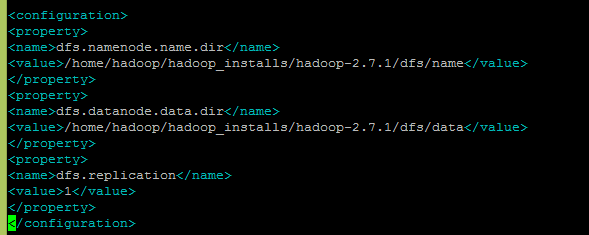
hadoop-env.sh



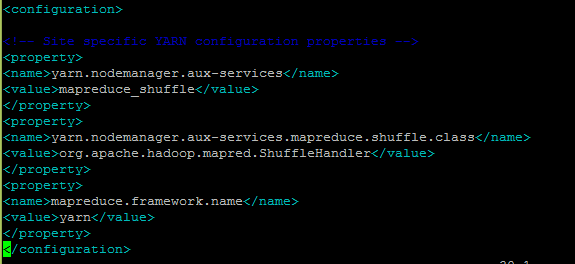
core-site.xml



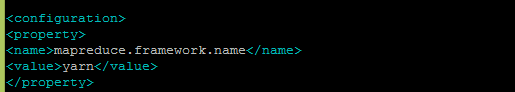
hdfs-site.xml



yarn-site.xml



mapred-site.xml

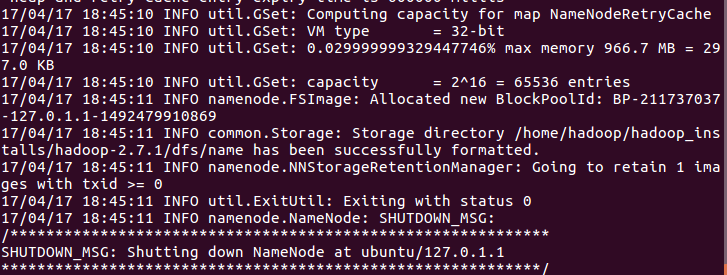


Slaves



1. 格式化NameNode

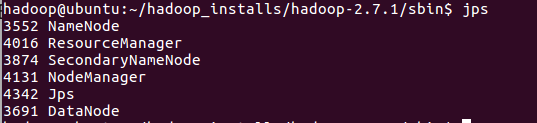
hadoop namenode –format



1. 启动HDFS和MapReduce

在hadoop\_installs/hadoop-2.7.1/sbin文件夹下执行 ./start-all.sh

JPS命令检查是否正常启动



停止HDFS和MapReduce用./stop-all.sh

二．WordCount测试

测试网页：https://en.wikipedia.org/wiki/Linear\_discriminant\_analysis

下载网页数据：wget –r –p –np –k https://en.wikipedia.org/wiki/Linear\_discriminant\_analysis

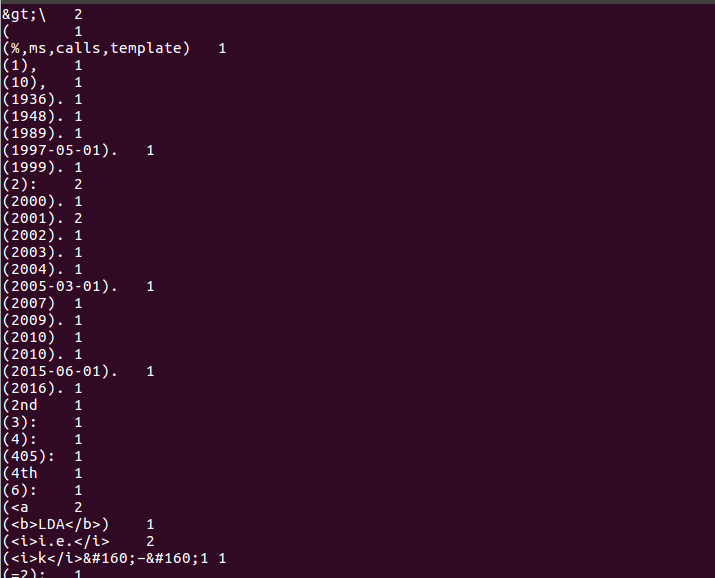
获得一个文件夹en.wikipedia.org，网页信息放置在en.wikipedia.org/wiki/ Linear\_discriminant\_analysis,将该文件拷贝到webtest.txt

建立/input 文件夹：hadoop fs –mkdir /input

将测试文件拷贝到input文件夹：hadoop dfs –copyFromLocal webtest.txt /test-in

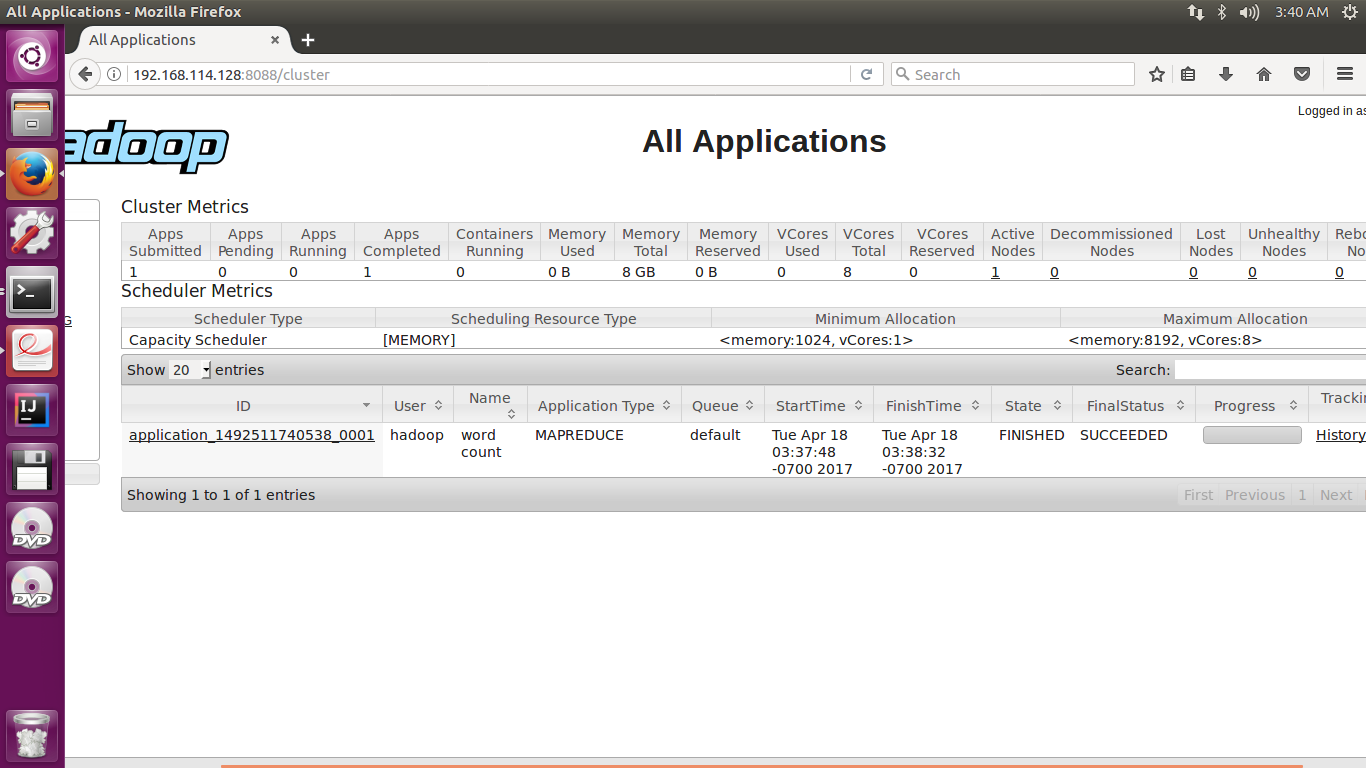
测试：hadoop jar hadoop-mapreduce-examples-2.7.1.jar wordcount /input /output

查看结果：hadoop dfs -cat /output/part-r-00000



用Hadoop的Web界面查看Hadoop集群和作业状态:

浏览器输入192.168.114.128:8088查看



三．实验体会

本次实验熟悉了如何创建新用户，如何用ssh免密码登陆，hadoop的安装配置。

实验并不难，由于课件上有些指令比较陈旧，并且我对linux下环境配置不熟悉，实验还是做了很久。