1. 系统安装情况

系统安装主要参考课程pdf和<http://dblab.xmu.edu.cn/blog/install-hadoop/>

(1)安装单机操作系统

使用的Linux系统为Ubuntu 18.04

(2)安装SSH

使用命令 sudo apt-get install openssh-server; Ubuntu默认安装了SSH client， 需要手动安装SSH server.

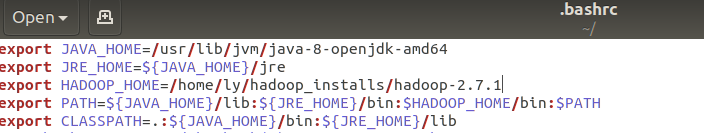
(3)安装Java

使用命令sudo apt-get install openjdk-8-jdk; 此处安装的是Java8, 并非实验 要求的Java7. 安装路径为/usr/lib/jvm/java-8-openjdk-amd64, 之后设置环境变量 会用到.

(4)解压Hadoop-2.7.1安装包, 安装后即可在单机模式下运行hadoop.

(5)配置环境变量

环境变量需要修改用户文件夹下的bashrc文件



修改完之后执行source ./bashrc指令更新环境变量.

(6)免密码SSH访问设置

执行命令ssh-keygen -t rsa产生认证文件,

将id\_rsa.pub复制到quthorized\_keys文件,

执行ssh localhost命令, 中途输入yes确认, 实现免密登录.

(7)修改hadoop配置文件

需要修改hadoop安装文件夹/etc/hadoop中的core-site.xml，hdfs-site.xml，mapred-site.xml和yarn-site.xml.



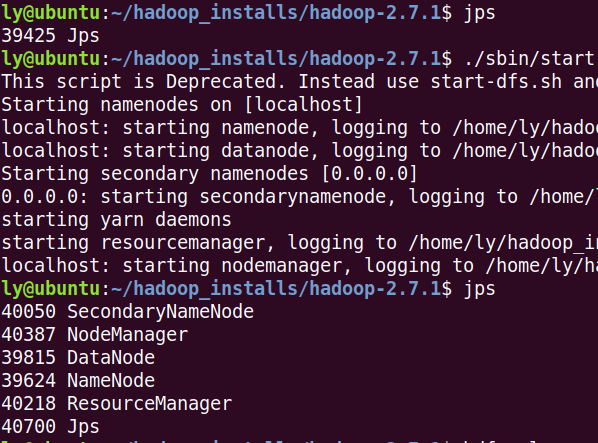


(8)格式化NameNode

执行指令hadoop namenode -format

(9)启动和停止hdfs与MapReduce

执行sbin文件夹内的脚本start-all.sh和stop-all.sh



2. 实验数据说明

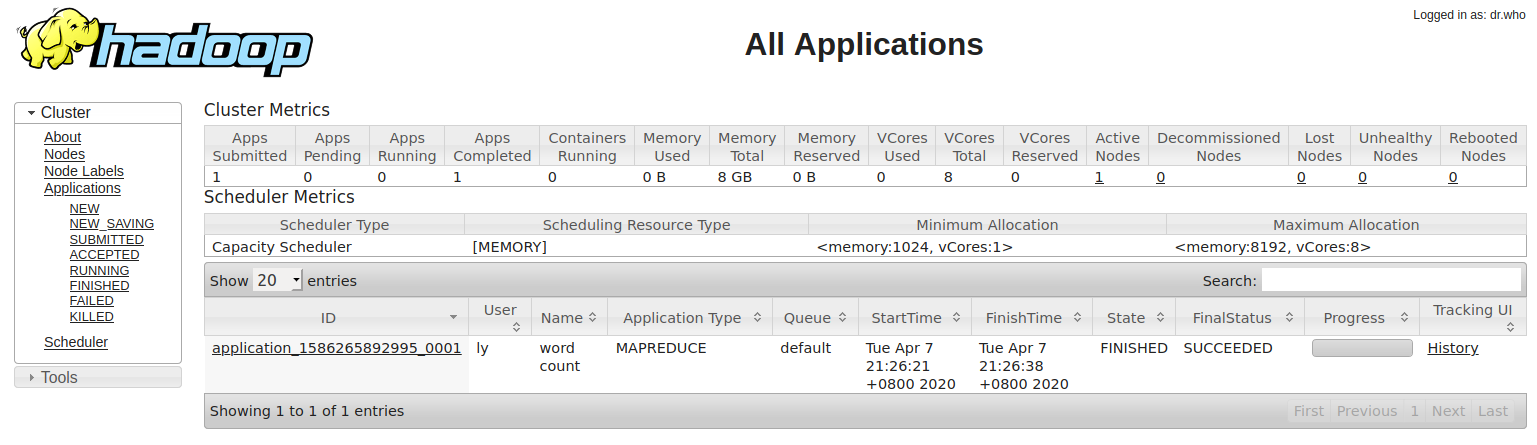
下载3个电子书对应的Plain Text UTF-8(txt文件)

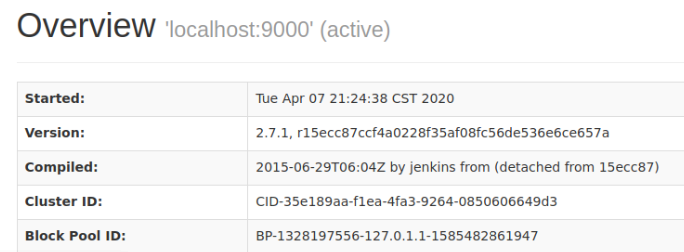
<http://www.gutenberg.org/ebooks/20417>

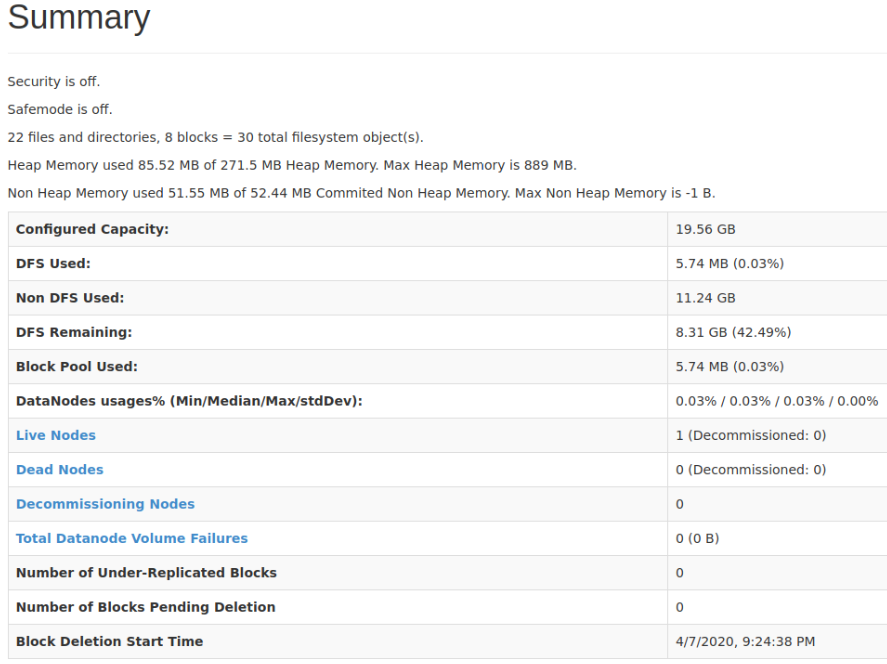
<http://www.gutenberg.org/ebooks/5000>

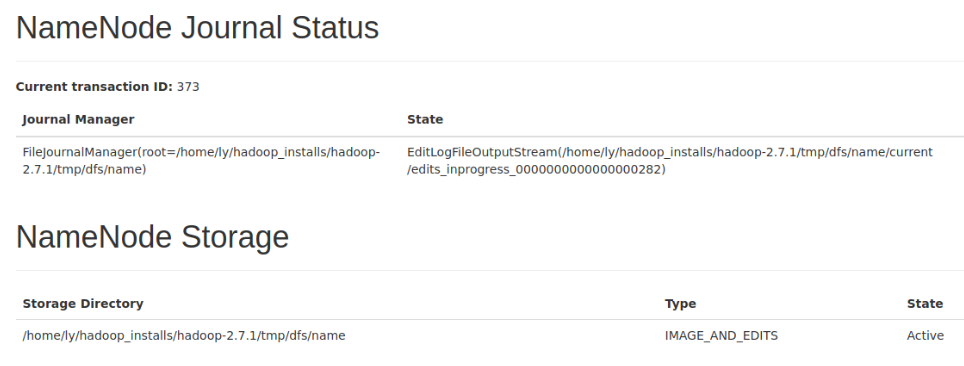
<http://www.gutenberg.org/ebooks/4300>

3. 作业运行状态

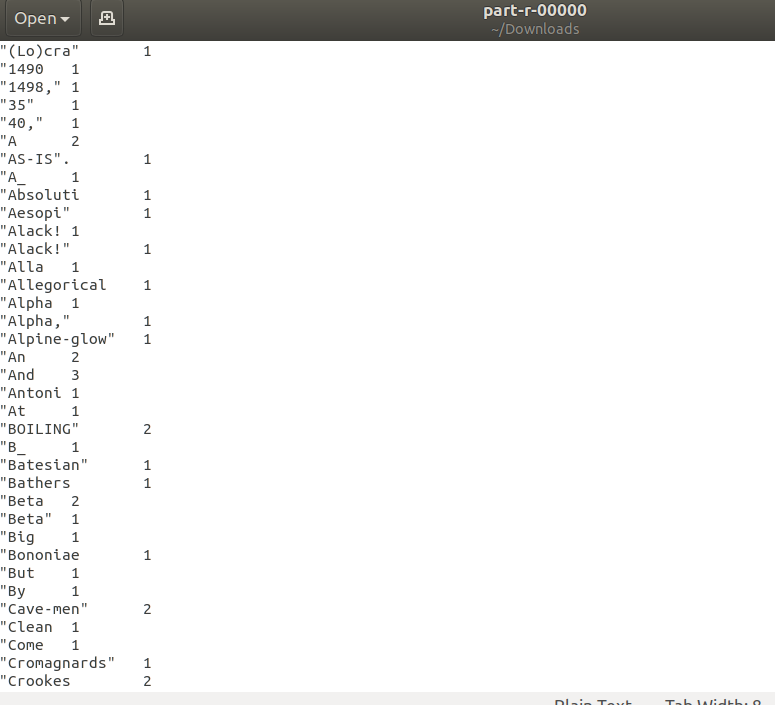








4. 实验输出结果



5. 实验体会

本次实验的主要目的是熟悉hadoop的工作流程，为之后的实验奠定基础。

以下是在本次实验中遇到的错误:

(1) ssh local失败

首次需要输入yes，按回车失败

(2) Hadoop命令无法执行

检查发现编辑bash\_profile时出现文本错误, source ~/.bash\_profile; 后 发现需要修改bashrc文件而不是bash\_profile文件。

(3) java环境变量配置失败

使用指令sudo apt-get install安装Java, 需要寻找Java的安装路径, 并非 pdf给出的示例。使用上述指令安装Java后, Java的安装路径为/usr/lib/jvm/java- 8-openjdk-amd64.