中期报告

一、项目阶段性成果：

项目组通过查阅书籍、网站博客、老师们的指导以及网络视频等资源获取关于本项目的相关信息并且不断试验，项目组取得了以下进展：

1. 在服务器上完成了Linux平台与Hadoop环境的搭建，并且安装了RPM与Hadoop生态系统中的yarn搭建过程如下

1.准备工作

1)建立虚拟机

2)选定一台机器作为 Master

2.网络配置(修改IP)

1)必要条件

(1)所有集群的节点都位于同一个局域网

(2)如果使用的是虚拟机安装的系统，那么需要更改网络连接方式为桥接（Bridge）模式，才能实现多个节点互连。

(3)虚拟机里面的CentOS是不会自动联网，手动配IP联网

(4)[CentOS7修改静态IP](http://mp.weixin.qq.com/s?__biz=MzU1OTM3ODQ0MA==&mid=2247483798&idx=7&sn=531d32452968855483687e2d139ecdbb&chksm=fc1978f1cb6ef1e730d1554efb101ee02e971a79168f0ba99f08525829d63f321fd487f2e255&scene=21#wechat_redirect)

3在 Master 节点上配置

1)关闭防火墙与SELinux

存在防火墙会导致ping的通，但是端口不通，从而导致DataNode启动，但是无法使用的问题。

2)安装 Java 环境

3)为了便于区分，可以修改各个节点的主机名

4)为了与其他机器方便通信，修改hosts

5)配置SSH免密

4.在其他 Slave 节点上配置 hadoop 用户、安装 SSH server、安装 Java 环境

5.在 Master 节点上安装 Hadoop，并完成配置

(1)官网下载下载hadoop-2.x.y.tar.gz(以下以2.7.7为例说明)

(2)解压缩

(3)为了方便记忆管理可以将文件重命名

(4)检查是否可用

(5)如果第四步成功以后将hadoop修改到环境变量里面

6.进行后续集群配置

1)关闭Hadoop(可以不执行)

2)集群/分布式模式需要修改 /usr/local/hadoop/etc/hadoop 中的5个配置文件

(1)**slaves**

(2)**core-site.xml**

(3).**hdfs-site.xml**

(4)**mapred-site.xml**

(5)**yarn-site.xml**

(6)/usr/local/hadoop/etc/hadoop/hadoop-env.sh

8)上述master配置成功以后，将 master 节点上的 /usr/local/hadoop 目录复制到其他 Slave 节点上9)slave1、slave2、slave3执行

10)首次启动需要先在 Master 节点执行 NameNode 的格式化

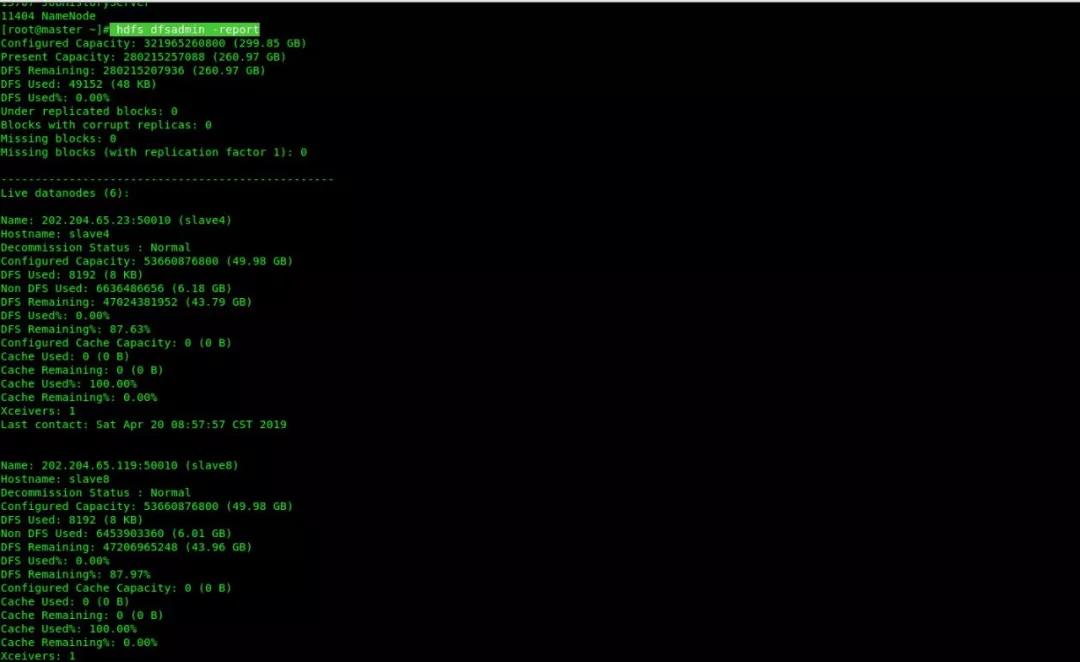
11)接着可以启动 hadoop 了，启动需要在 Master 节点上进行

12)通过jps命令查看各节点启动进程，

13)查看 DataNode 是否正常启动

(1)命令行方式

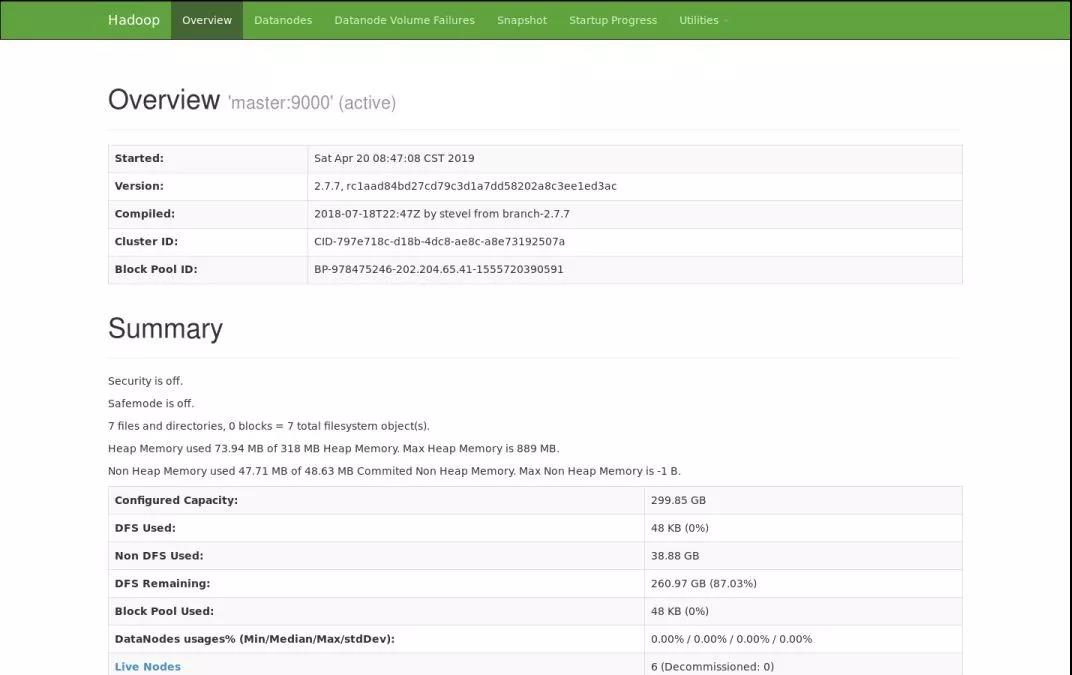
$> hdfs dfsadmin –report



如果 Live datanodes 不为 0 ，则说明集群启动成功

(2)Web方法

查询<http://master:50070/>



7.感受一下Hadoop的运行

选择运行 grep 例子，我们将 input 文件夹中的所有文件作为输入，筛选当中符合正则表达式 dfs[a-z.]+ 的单词并统计出现的次数。(执行分布式实例过程与伪分布式模式一样)

1)命令行方法

*#> hdfs dfs -mkdir -p /user/hadoop //先创建 HDFS 上的用户目录*//将 /usr/local/hadoop/etc/hadoop 中的配置文件作为输入文件复制到分布式文件系统中*#> hdfs dfs -mkdir input#> hdfs dfs -put /usr/local/hadoop/etc/hadoop/\*.xml input*//运行 MapReduce 作业（这一步比较考验硬件支持，若卡死了，可以试试重启，实在不行就加内存条）*#> hadoop jar /usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-\*.jar grep input output 'dfs[a-z.]+'*

2)图形方法：访问http://master:8088/cluster

#> stop-yarn.sh#> stop-dfs.sh#> mr-jobhistory-daemon.sh stop historyserver

8.yarn的安装

1）网络yum源

#> vi /etc/yum.repos.d/CentOS-Base.repo    *//只要是repo就是yum源*

#> yum list  *//list available*

#> yum update httpd -y *//update httpd//dangerous*

#> yum update -y  *//all rpm update*

#> yum remove -y  *//even erase, don't use yum remove（use rpm -e*

2）本地光盘yum

#> mount /dev/cdrom /mnt/cdrom

#> mv CentOS-\*.repo CentOS-\*.repo.back

#> mv CentOS-Medio.repo.back CentOS-Media.repo

#> vi CentOS-Media.repo *//改成下面的内容*

baesurl=file:*///mnt/cdrom*

#file:*///media/cdrom/*

#file:*///media/cdromcorder/*

gpgcheck=1enablle=1

#> yum list

1. 编写并调试了处理用户观看视频时的行为痕迹记录的程序