大数据存储与运算复习题

1、以下选项中，哪个程序负责HDFS数据存储。 （ B ）

A、NameNode B、DataNode

C、Secondary NameNode D、ResourceManager

2、下列哪项通常是集群的最主要的性能瓶颈？（C ）

A、CPU B、 网络 C、磁盘 D、内存

3、下面哪项是Hadoop的作者？（ B ）

A、Martin Fowler B、Doug cutting

C、Mark Elliot Zuckerberg D、Kent Beck

4、HDFS默认备份数量？（ D ）

A、0 B、1

C、2 D、3

5、下列描述说法错误的是？（ D ）

A、SecureCRT是一款支持SSH的终端仿真程序，它能够在Windows操作系统上远程连接Linux服务器执行操作。

B、Hadoop是一个用于处理大数据的分布式集群架构，支持在GNU/Linux系统以及Windows系统上进行安装使用。

C、VMware Workstation是一款虚拟计算机的软件，用户可以在单一的桌面上同时操作不同的操作系统。

D、 SSH是一个软件，专为远程登录会话和其他网络服务提供安全性功能的软件。

6、Hadoop2.x版本中的数据块大小默认是多少？（B ）

　 A、 64M B、 128M

C、 256M D、 512M

7、关于Secondary NameNode哪项是正确？（ C ）

A、它是 NameNode 的热备

B、它对内存没有要求

C、它的目的是帮助 NameNode合并编辑日志，减少NameNode启动时间

D、SecondaryNameNode 应与NameNode部署到一个节点

8、下面哪个进程负责 MapReduce 任务调度。（ B ）

A、NameNode B、Jobtracker

C、TaskTracker D、secondaryNameNode

9、 Client 端上传文件的时候下列哪项正确？（ B ）

A、数据经过 NameNode 传递给 DataNode

B、Client 端将文件切分为 Block，依次上传

C、Client 只上传数据到一台 DataNode，然后由 NameNode 负责 Block 复制工作

D、以上都不正确

10、 MapReduce编程模型，键值对<key, value>的key必须实现哪个接口？（ A ）

A、WritableComparable B、Comparable

C、Writable D、LongWritable

11、MapReduce适用于（D ）

A、任意应用程序

B、任意可以在Windows Server 2008上的应用程序

C、可以串行处理的应用程序

D、可以并行处理的应用程序

12、下面关于MapReduce模型中Map函数与Reduce函数的描述正确的是（ A ）

A、一个Map函数就是对一部分原始数据进行指定的操作。

B、一个Map操作就是对每个Reduce所产生的一部分中间结果进行合并操作。

C、Map与Map之间不是相互独立的。

D、Reducee与Reduce之间不是相互独立的。

13、MapReduce自定义排序规则需要重写下列那项方法（ B ）

A、readFields()

B、compareTo()

C、map()

D、reduce()

14、以下哪一项属于非结构化数据。（C ）

A、企业ERP数据 B、财务系统数据

C、视频监控数据 D、日志数据

15、HBase是分布式列式存储系统，记录按什么集中存放。（ A ）

A、列族 B、列

C、行 D、不确定

16、HDFS1.0 默认 Block Size大小是多少。（ B ）

A、32MB B、64MB

C、128MB D、256MB

17、Hadoop1.0默认的调度器策略是哪个。（ A ）

A、先进先出调度器 B、计算能力调度器

C、公平调度器 D、优先级调度器

18、在实验集群的master节点使用jps命令查看进程时，终端出现以下哪项能说明Hadoop主节点启动成功？（ D ）

A、Namenode,Datanode, TaskTracker

B、Namenode,Datanode, secondaryNameNode

C、Namenode,Datanode, HMaster

D、Namenode,JobTracker, secondaryNameNode

19、HBase数据库的BlockCache缓存的数据块中，哪一项不一定能提高效率。（D ）

A、–ROOT-表 B、META.表

C、HFile index D、普通的数据块

20、 MapReduce编程模型，键值对<key, value>的key必须实现哪个接口？（ A ）

A、WritableComparable B、Comparable

C、Writable D、LongWritable

1. 大数据的4V特征包含 大量、多样、高速、价值。

2、Hadoop三大组件包含 HDFS、MapReduce、Yarn。

3、Hadoop2.x版本中的HDFS是由 NameNode、DataNode、Secondary NameNode组成。

4、Hadoop集群部署方式分别是 独立模式、伪分布式模式 和 完全分布式模式。

5、加载环境变量配置文件需要使用source /etc/profile命令。

6、格式化HDFS集群命令是 hadoop namenode -format。

7、脚本一键启动Hadoop集群服务命令是 start-all.sh。

8、Hadoop默认开设HDFS文件系统端口号 50070和监控Yarn集群端口号 8088。

9、Hadoop发行版本分为 开源社区版、商业版。

1、简述大数据研究的意义。

发散型题目，略

1. 简述Hadoop版本区别。

Hadoop发行版本分为开源社区版和商业版，社区版是指由Apache软件基金会维护的版本，是官方维护的版本体系。商业版Hadoop是指由第三方商业公司在社区版Hadoop基础上进行了一些修改、整合以及各个服务组件兼容性测试而发行的版本，例如比较著名的有Cloudera公司的CDH版本。（书本P9 1.3.4第一段）

1. 简述什么是SSH以及SSH协议解决的问题。

SSH为Secure Shell的缩写，它是一种网络安全协议，专为远程登录会话和其他网络服务提供安全性的协议。通过使用SSH服务，可以把传输的数据进行加密，有效防止远程管理过程中的信息泄露问题。（书本P29 第三行）

1. 简述Hadoop集群部署方式以及各方式使用场景。

（1）独立模式：又称为单机模式，在该模式下，无需运行任何守护进程，所有的程序都在单个JVM上执行。独立模式下调试Hadoop集群的MapReduce程序非常方便，所以一般情况下，该模式在学习或者开发阶段调试使用。

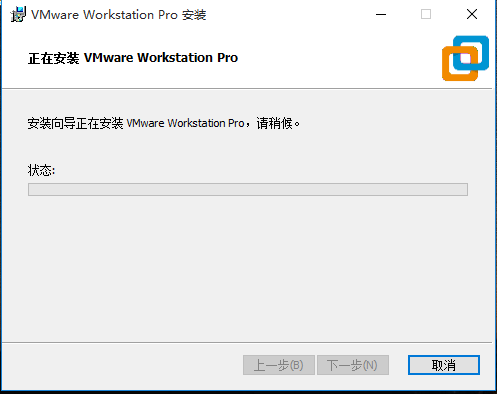
（2）伪分布式模式：Hadoop程序的守护进程运行在一台节点上，通常使用伪分布式模式用来调试Hadoop分布式程序的代码，以及程序执行是否正确，伪分布式模式是完全分布式模式的一个特例。

（3）完全分布式模式：Hadoop的守护进程分别运行在由多个主机搭建的集群上，不同节点担任不同的角色，在实际工作应用开发中，通常使用该模式构建企业级Hadoop系统。（书本 P32 第二行）

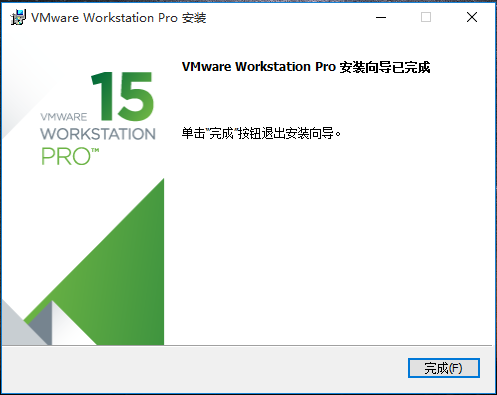
1、大数据基础环境的搭建与调试，安装测试完成请截图，至少一张安装过程图和一张结果图。

（1）虚拟机软件的安装

过程图

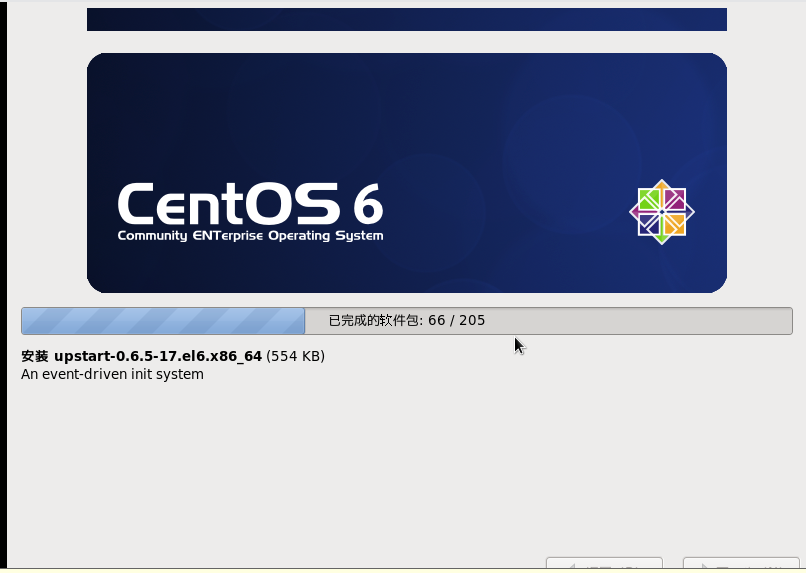


结果图



（2）linux操作系统的安装

过程图

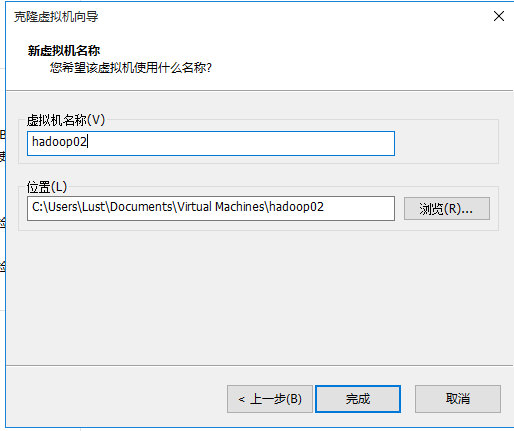


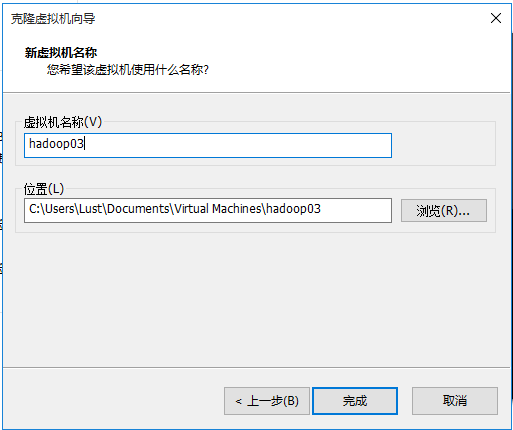
结果图



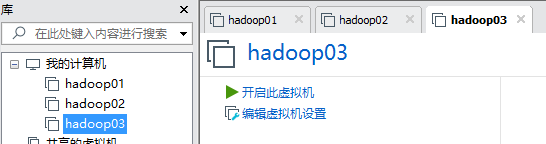
（3）linux操作系统的克隆，至少2台

过程图



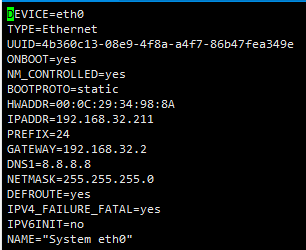


结果图

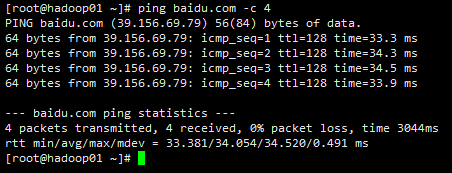


（4）网络配置与主机名修改

网络配置过程图



网络配置结果图

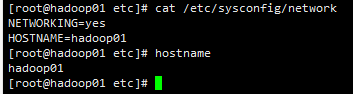


主机名修改过程图



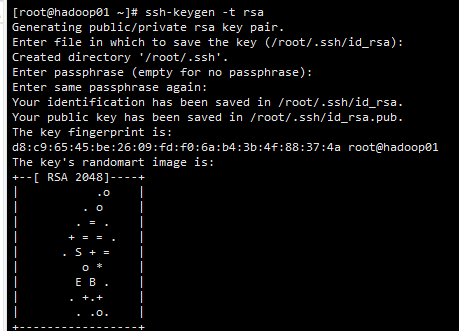


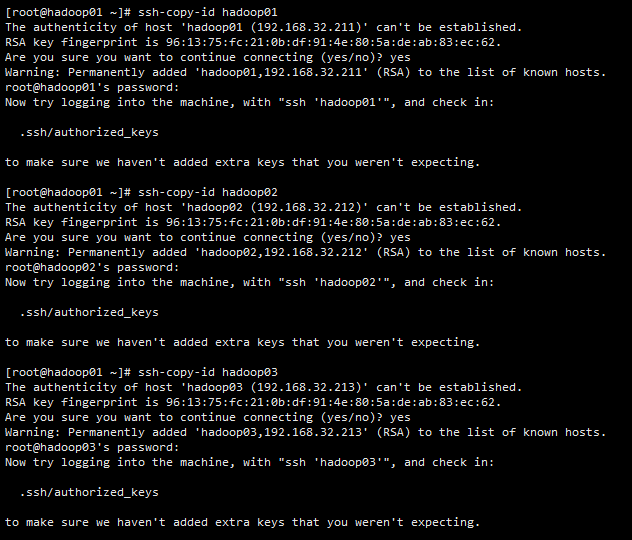
修改主机名结果图



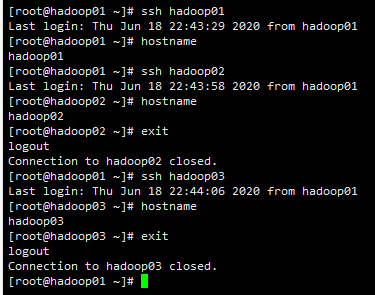
（5）SSH免密登录

过程图



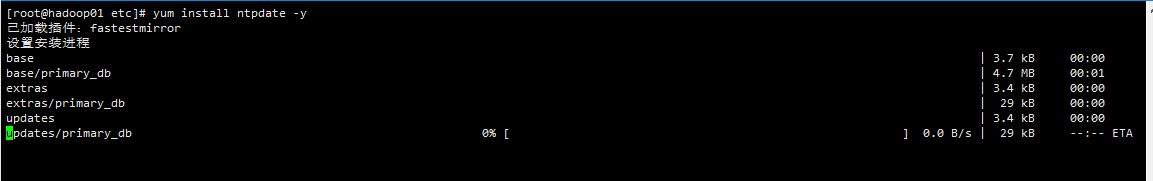


结果图

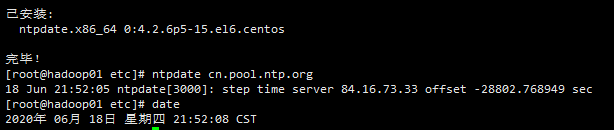


（6）时间同步

安装ntpdate过程图

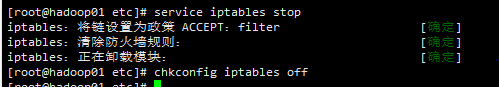


时间同步结果图



（7）防火墙服务及开机自启动服务关闭

过程图



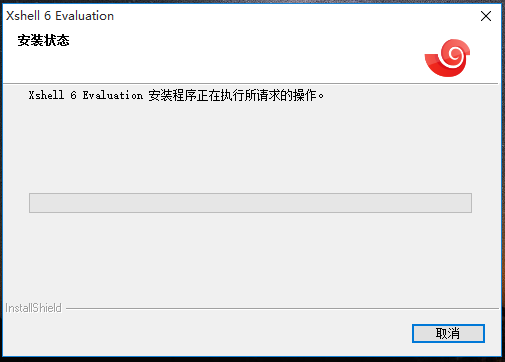
结果图



2、大数据软件的安装，安装测试完成请截图，至少一张安装过程图和一张结果图。

（1）xshell与xftp远程登录软件的安装与使用

xshell安装图

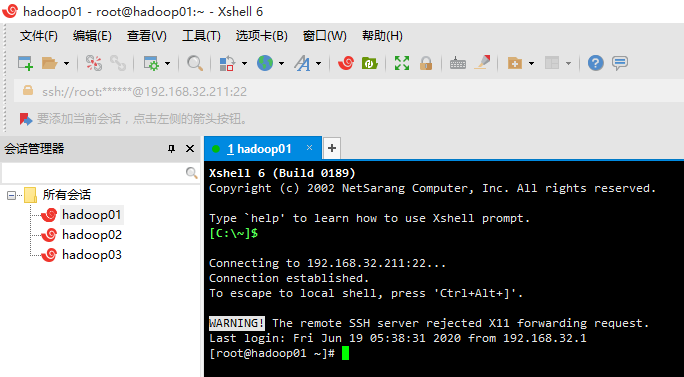


（2）3台虚拟机操作系统的jdk软件的安装与配置

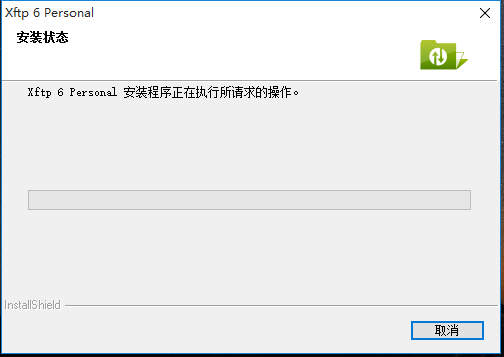
（3）3台虚拟机操作系统的hadoop软件的安装与配置

（4）hadoop框架的初始化及运行

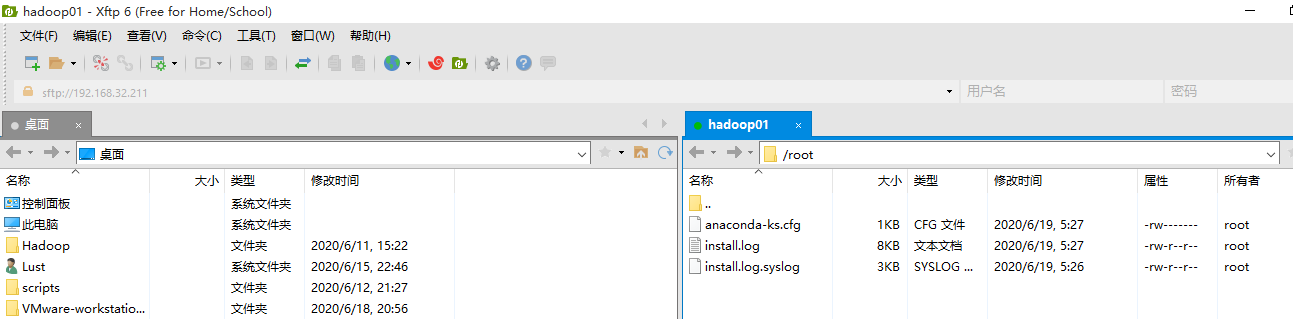
xshell使用图



xftp安装图

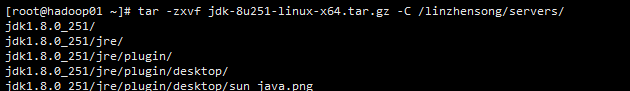


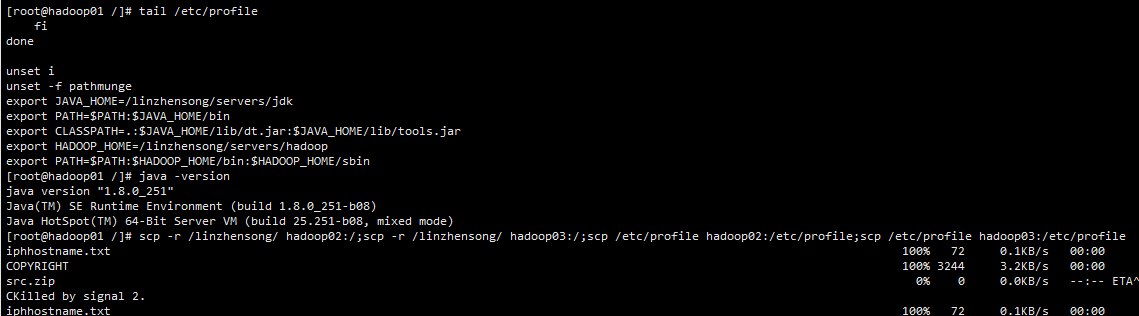
xftp使用图



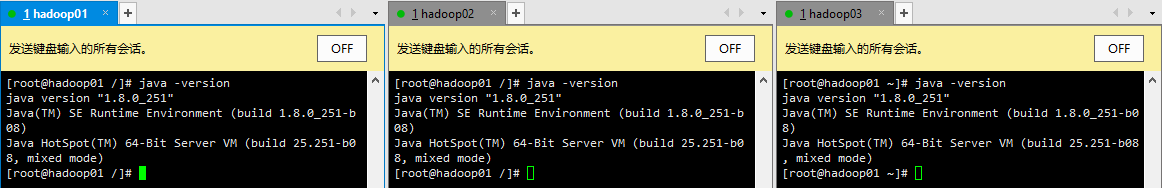
（2）3台虚拟机操作系统的jdk软件的安装与配置

过程图



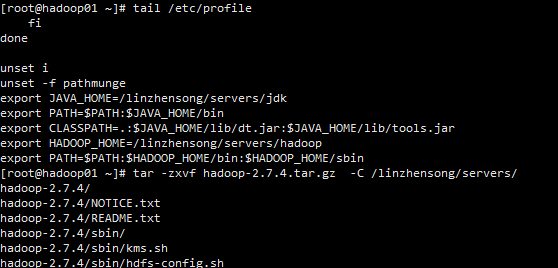


结果图

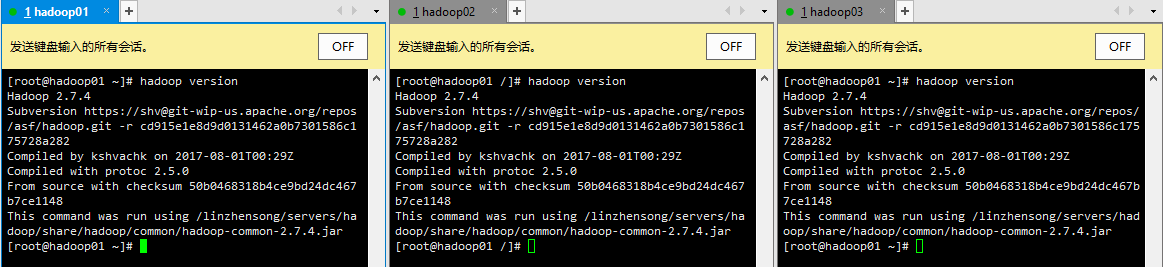


（3）3台虚拟机操作系统的hadoop软件的安装与配置

安装过程图

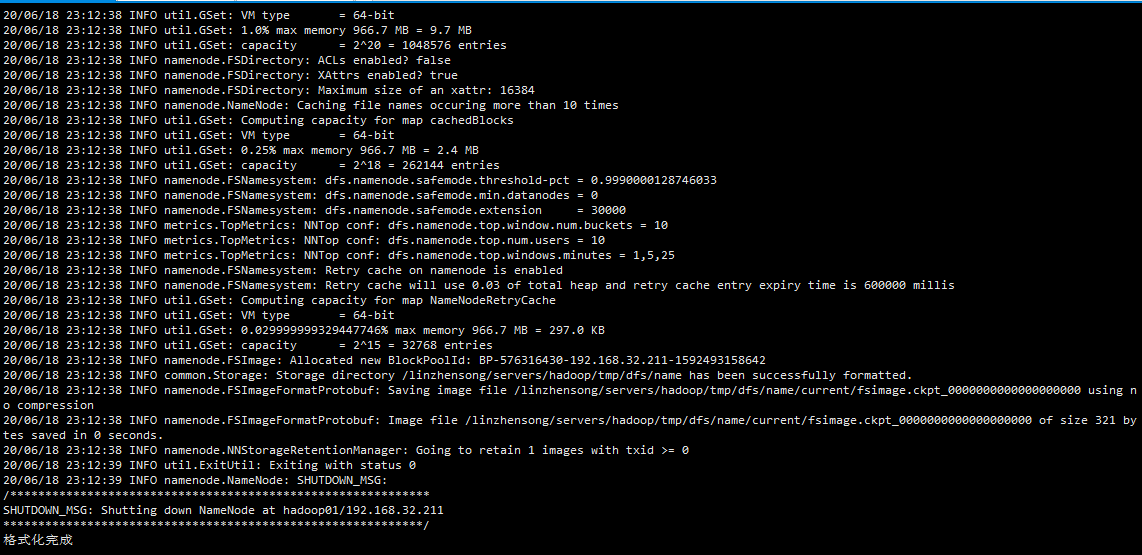


结果图

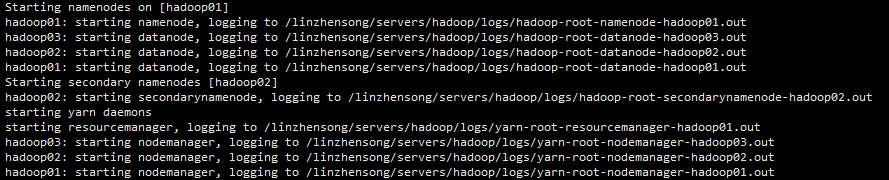


（4）hadoop框架的初始化及运行

初始化图



运行图



效果图

