

日期	2020年3月5日	周次	第一周	时数	2课时
教学单元	大数据概论				
教学目标	知识目标： 1. 了解大数据概念 2. 掌握大数据特点（4V） 3. 了解大数据应用场景 4. 了解大数据发展前景 5. 掌握大数据部门业务流程分析 6. 掌握大数据部门组织结构				
	技能目标： 能够掌握大数据特点（4V）				
	素质目标： 1. 提高学生逻辑思维，创新能力； 2. 树立团队协作意识； 3. 树立良好的政治素质（爱党爱国、品行端正、遵纪守法）。				
教学重点	大数据特点（4V）				
教学难点	1. 大数据部门业务流程分析 2. 大数据部门组织结构				

教学方法	讲授演示法；任务驱动法
教学手段	云班课信息化教学；多媒体教学；钉钉直播教学
教学过程及内容	
<p>1. 新课引入</p> <div data-bbox="236 875 1353 1498" style="border: 1px solid #ccc; padding: 10px;">  <p>大数据概念</p> <p>大数据 (Big Data)：指无法在一定时间范围内用常规软件工具进行捕捉、管理和处理的数据集合，是需要新处理模式才能具有更强的决策力、洞察发现力和流程、优化能力的海量、高增长率和多样化的信息资产。</p> </div> <p>2. 任务驱动</p> <p>大数据特点：</p> <p>大数据特点-Volume(大量)</p> <p>截至目前，人类生产的所有印刷材料的数据量是 200PB，而历史上全人类总共说过的话的数据量大约是 5EB。当前，典型个人计算机硬盘的容量为 TB 量级，而一些大企业的数据量已经接近 EB 量级。</p>	

大数据特点-Velocity(高速)

这是大数据区别于传统数据挖掘的最显著特征。根据 IDC 的“数字宇宙”的报告，预计到 2020 年，全球数据使用量将达到 35.2ZB。在如此海量的数据面前，处理数据的效率就是企业的生命。

天猫双十一：2017 年 3 分 01 秒，天猫交易额超过 100 亿

大数据特点-Variety(多样)

这种类型的多样性也让数据被分为结构化数据和非结构化数据。相对于以往便于存储的以数据库/文本为主的结构化数据，非结构化数据越来越多，包括网络日志、音频、视频 图片、地理位置信息等，这些多类型的数据对数据的处理能力提出了更高要求。

id	用户	日期	购买商品	购买数量
1001	canglaoshi	20170710-9:10:10	面膜	2
1002	xiaozelaoshi	20170710-9:11:20	化妆品	3
1003	beduolaoshi	20170710-9:22:50	内衣	4
1004	sslaoshi	20170710-10:12:20	海狗人参丸	100

大数据特点-Value(低价值密度)

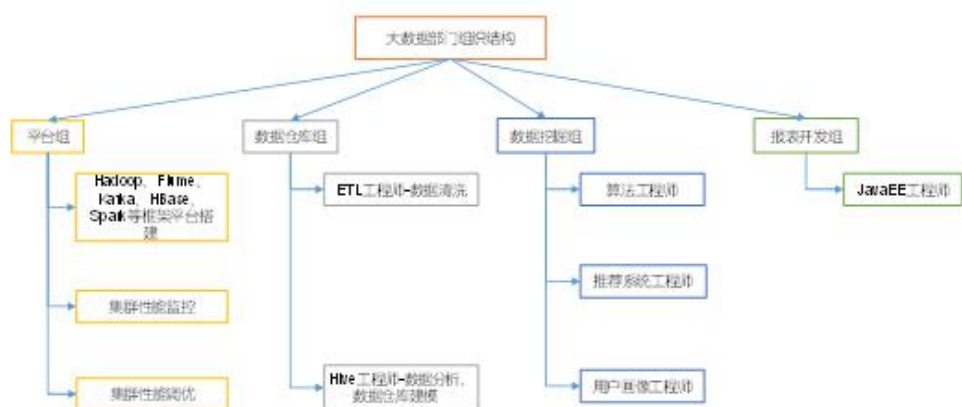
价值密度的高低与数据总量的大小成反比。比如，在一天监控视频中，我们只关心 xx 同学晚上在床上健身那一分钟，如何快速对有价值数据“提纯”成为目前大数据背景下待解决的难题。

大数据部门业务流程分析



3. 课堂总结

大数据部分组织结构



参考资料 及应用	《Hadoop 大数据分析》 主编：高永彬、钱亮宏、方志军 出版社：中国铁道出版社
作业	完成大数据概述课后练习
教学反思	部分学生没有自主学习能力，应提供层次清楚、条理分明的板书

日期	2020年3月8日	周次	第一周	时数	2课时
教学单元	Hadoop 概述				
教学目标	<p>知识目标：</p> <ol style="list-style-type: none"> 1. 了解 Hadoop 是什么？ 2. 了解 Hadoop 发展历史 3. 掌握 Hadoop 三大发行版本 4. 掌握 Hadoop 的优势（4高） 				
	<p>技能目标：能够掌握 Hadoop2.0 和 Hadoop3.0 组成</p>				
	<p>素质目标：</p> <ol style="list-style-type: none"> 4. 提高学生逻辑思维，创新能力； 5. 树立团队协作意识； 6. 树立良好的政治素质（爱党爱国、品行端正、遵纪守法）。 				
教学重点	<ol style="list-style-type: none"> 1. Hadoop 的优势（4高） 2. Hadoop 组成 				
教学难点	Hadoop 组成（HDFS 架构概述）				

教学方法	讲授演示法；任务驱动法
教学手段	云班课信息化教学；多媒体教学；钉钉直播教学
教学过程及内容	
<p>4. 新课引入</p> <p>1) Hadoop 是一个由 Apache 基金会所开发的分布式系统基础架构。</p> <p>2) 主要解决，海量数据的存储和海量数据的分析计算问题。</p> <p>3) 广义上来说，Hadoop 通常是指一个更广泛的概念——Hadoop 生态圈。</p> <div data-bbox="268 1196 1214 1675" data-label="Diagram"> </div> <p>5. 任务驱动</p> <p>Hadoop 三大发行版本：Apache、Cloudera、Hortonworks。</p> <p>Apache 版本最原始（最基础）的版本，对于入门学习最好。</p> <p>Cloudera 在大型互联网企业中用的较多。</p>	

Hortonworks 文档较好。

1. Apache Hadoop

官网地址: <http://hadoop.apache.org/releases.html>

下载地址: <https://archive.apache.org/dist/hadoop/common/>

2. Cloudera Hadoop

官 网 地 址 :

<https://www.cloudera.com/downloads/cdh/5-10-0.html>

下载地址: <http://archive-primary.cloudera.com/cdh5/cdh/5/>

(1) 2008 年成立的 Cloudera 是最早将 Hadoop 商用的公司, 为合作伙伴提供 Hadoop 的商用解决方案, 主要是包括支持、咨询服务、培训。

(2) 2009 年 Hadoop 的创始人 Doug Cutting 也加盟 Cloudera 公司。Cloudera 产品主要为 CDH, Cloudera Manager, Cloudera Support

(3) CDH 是 Cloudera 的 Hadoop 发行版, 完全开源, 比 Apache Hadoop 在兼容性, 安全性, 稳定性上有所增强。

(4) Cloudera Manager 是集群的软件分发及管理监控平台, 可以在几个小时内部署好一个 Hadoop 集群, 并对集群的节点及服务进行实时监控。Cloudera Support 即是对 Hadoop 的技术支持。

(5) Cloudera 的标价为每年每个节点 4000 美元。Cloudera 开发并贡献了可实时处理大数据的 Impala 项目。

3. Hortonworks Hadoop

官 网 地 址 :

<https://hortonworks.com/products/data-center/hdp/>

下 载 地 址 :

<https://hortonworks.com/downloads/#data-platform>

(1) 2011 年成立的 Hortonworks 是雅虎与硅谷风投公司 Benchmark Capital 合资组建。

(2) 公司成立之初就吸纳了大约 25 名至 30 名专门研究 Hadoop 的雅虎工程师，上述工程师均在 2005 年开始协助雅虎开发 Hadoop，贡献了 Hadoop 80% 的代码。

(3) 雅虎工程副总裁、雅虎 Hadoop 开发团队负责人 Eric Baldeschwieler 出任 Hortonworks 的首席执行官。

(4) Hortonworks 的主打产品是 Hortonworks Data Platform (HDP)，也同样是 100% 开源的产品，HDP 除常见的项目外还包括了 Ambari，一款开源的安装和管理系统。

(5) HCatalog，一个元数据管理系统，HCatalog 现已集成到 Facebook 开源的 Hive 中。Hortonworks 的 Stinger 开创性的极大的优化了 Hive 项目。Hortonworks 为入门提供了一个非常好的，易于使用的沙盒。

(6) Hortonworks 开发了很多增强特性并提交至核心主干，这使得 Apache Hadoop 能够在包括 Window Server 和 Windows Azure 在内的 Microsoft Windows 平台上本地运行。定价以集群为基础，每 10 个节点每年为 12500 美元。

6. 课堂总结

Hadoop 优势（4 高）

- 1) 高可靠性：Hadoop 底层维护多个数据副本，所以即使 Hadoop 某个计算元素或存储出现故障，也不会导致数据的丢失。
- 2) 高扩展性：在集群间分配任务数据，可方便的广展数以千计的节点。
- 3) 高效性：在 MapReduce 的思想下 Hadoop 是并行工作的，以加快任务处理速度。
- 4) 高容错性：能够自动将失败的任务重新分配。

参考资料 及应用	《Hadoop 大数据分析》 主编：高永彬、钱亮宏、方志军 出版社：中国铁道出版社
作业	完成 Hadoop 概述课后练习
教学反思	无

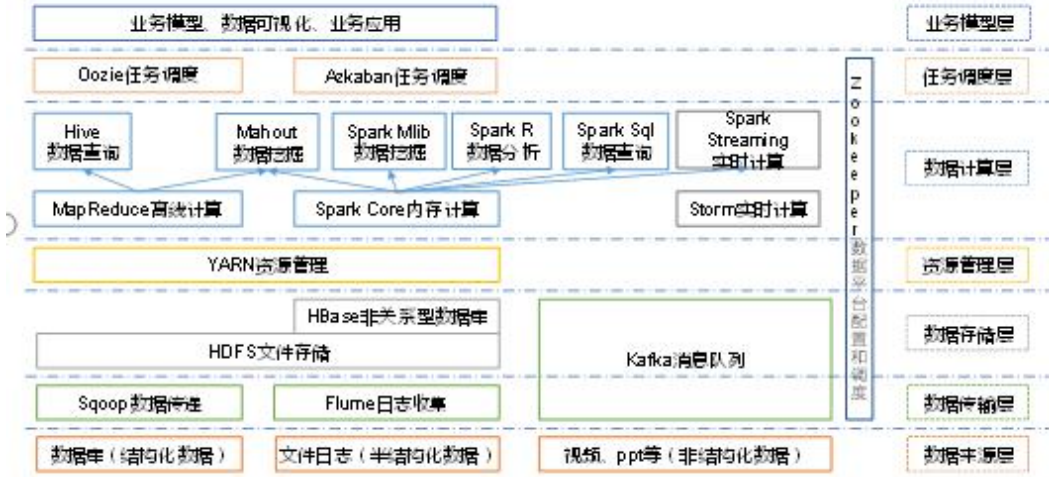
日期	2020年3月14日	周次	第二周	时数	2课时
教学单元	从 Hadoop 框架讨论大数据生态				
教学目标	知识目标： 1. 了解大数据技术生态体系 2. 理解推荐系统框架图 3. 掌握 Hadoop 生态系统组件				
	技能目标：能够理解推荐系统框架图				
	素质目标： 7. 提高学生逻辑思维，创新能力； 8. 树立团队协作意识； 9. 树立良好的政治素质（爱党爱国、品行端正、遵纪守法）。				
教学重点	1. 理解推荐系统框架图 2. 掌握 Hadoop 生态系统组件				
教学难点	理解推荐系统框架图				

教学方法	讲授演示法；任务驱动法
教学手段	云班课信息化教学；多媒体教学；钉钉直播教学

教学过程及内容

7. 新课引入

大数据技术生态体系图



任务驱动涉及的技术名词解释如下：

- 1) Sqoop: Sqoop 是一款开源的工具，主要用于在 Hadoop、Hive 与传统的数据库(MySql)间进行数据的传递，可以将一个关系型数据库（例如：MySQL，Oracle 等）中的数据导进到 Hadoop 的 HDFS 中，也可以将 HDFS 的数据导进到关系型数据库中。
- 2) Flume: Flume 是 Cloudera 提供的一个高可用的，高可靠的，分

布式的海量日志采集、聚合和传输的系统，Flume 支持在日志系统中定制各类数据发送方，用于收集数据；同时，Flume 提供对数据进行简单处理，并写到各种数据接受方（可定制）的能力。

3) Kafka: Kafka 是一种高吞吐量的分布式发布订阅消息系统，有如下特性：

(1) 通过 O(1) 的磁盘数据结构提供消息的持久化，这种结构对于即使数以 TB 的消息存储也能够保持长时间的稳定性能。

(2) 高吞吐量：即使是非常普通的硬件 Kafka 也可以支持每秒数百万的消息。

(3) 支持通过 Kafka 服务器和消费机集群来分区消息。

(4) 支持 Hadoop 并行数据加载。

4) Storm: Storm 用于“连续计算”，对数据流做连续查询，在计算时就将结果以流的形式输出给用户。

5) Spark: Spark 是当前最流行的开源大数据内存计算框架。可以基于 Hadoop 上存储的大数据进行计算。

6) Oozie: Oozie 是一个管理 Hadoop 作业（job）的工作流程调度管理系统。

7) Hbase: HBase 是一个分布式的、面向列的开源数据库。HBase 不同于一般的关系数据库，它是一个适合于非结构化数据存储的数据库。

8) Hive: Hive 是基于 Hadoop 的一个数据仓库工具，可以将结构化的数据文件映射为一张数据库表，并提供简单的 SQL 查询功能，可以

将 SQL 语句转换为 MapReduce 任务进行运行。其优点是学习成本低，可以通过类 SQL 语句快速实现简单的 MapReduce 统计，不必开发专门的 MapReduce 应用，十分适合数据仓库的统计分析。

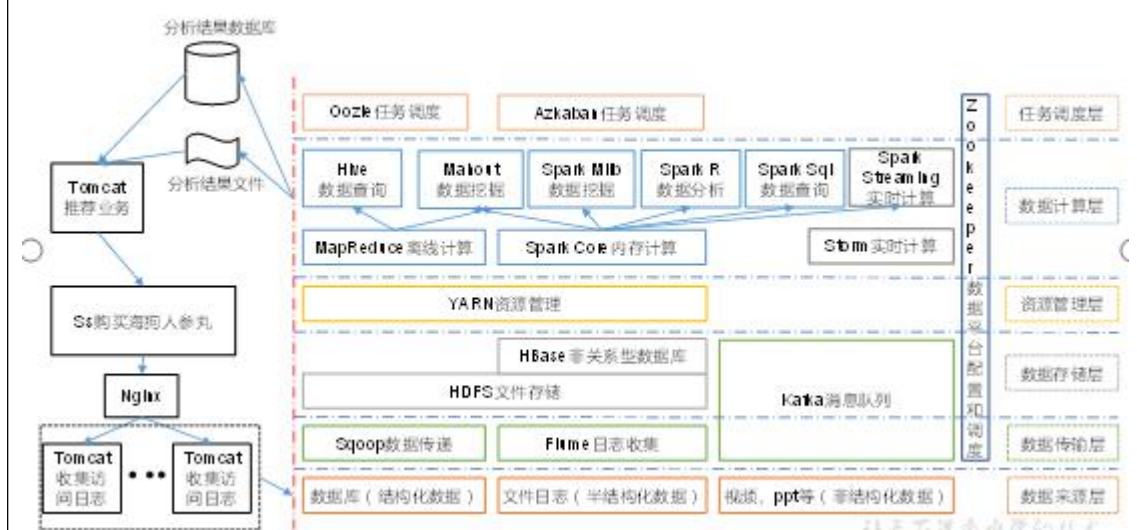
10) R 语言：R 是用于统计分析、绘图的语言和操作环境。R 是属于 GNU 系统的一个自由、免费、源代码开放的软件，它是一个用于统计计算和统计制图的优秀工具。

11) Mahout：Apache Mahout 是个可扩展的机器学习和数据挖掘库。

12) ZooKeeper：Zookeeper 是 Google 的 Chubby 一个开源的实现。它是一个针对大型分布式系统的可靠协调系统，提供的功能包括：配置维护、名字服务、分布式同步、组服务等。ZooKeeper 的目标就是封装好复杂易出错的关键服务，将简单易用的接口和性能高效、功能稳定的系统提供给用户。

8. 课堂总结

推荐系统项目框架



参考资料及应用	《Hadoop 大数据分析》 主编：高永彬、钱亮宏、方志军 出版社：中国铁道出版社
作业	完成 Hadoop 生态系统组件课后习题
教学反思	适当增加任务驱动法使用，提高学生兴趣

日期	2020年3月15日	周次	第二周	时数	2课时
教学单元	SecureCRT 工具的使用				
教学目标	知识目标： 1. Linux 远程登录及相关工具 1. 掌握 SecureCRT 工具的安装 2. 熟练使用 SecureCRT 工具				
	技能目标： 1. 能够熟练使用 SecureCRT 工具				
	素质目标： 10. 提高学生逻辑思维，创新能力； 11. 树立团队协作意识； 12. 树立良好的政治素质（爱党爱国、品行端正、遵纪守法）。				
教学重点	1. 掌握 SecureCRT 工具的安装 2. 熟练使用 SecureCRT 工具				
教学难点	熟练使用 SecureCRT 工具				
教学方法	讲授演示法；任务驱动法				

教学手段 云班课信息化教学；多媒体教学；钉钉直播教学

教学过程及内容

9. 新课引入

Linux 远程登录及相关工具介绍

Linux 一般作为服务器使用，而服务器一般放在机房，你不可能在机房操作你的 Linux 服务器。这时我们就需要远程登录到 Linux 服务器来管理维护系统。

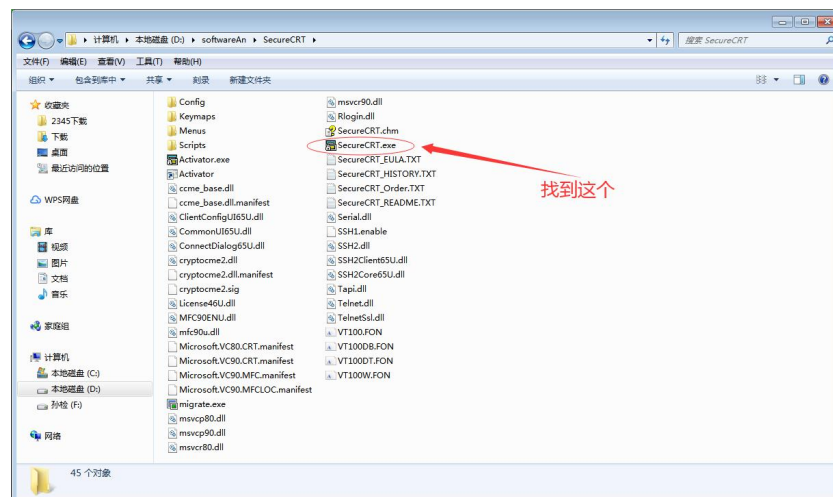
Linux 系统中是通过 SSH 服务实现的远程登录功能，默认 ssh 服务端口号为 22。Window 系统上 Linux 远程登录客户端有 SecureCRT, Putty, SSH Secure Shell, XShell 等

10. 任务示范

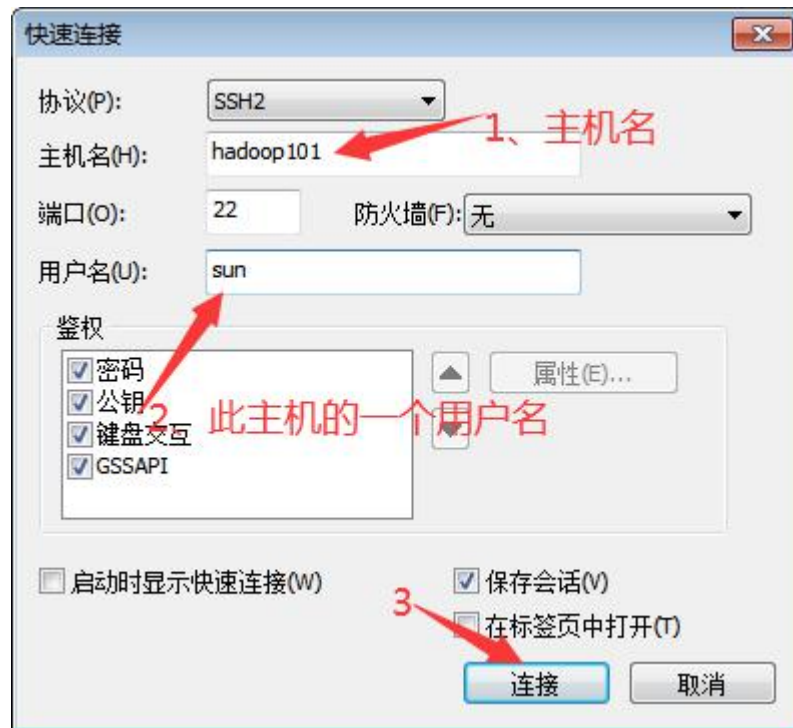
简单来说：这款工具可以帮助我们在 Windows 下面操作 Linux 系统。

安装：贼简单-我发一个安装包给你们，解压就 OK。

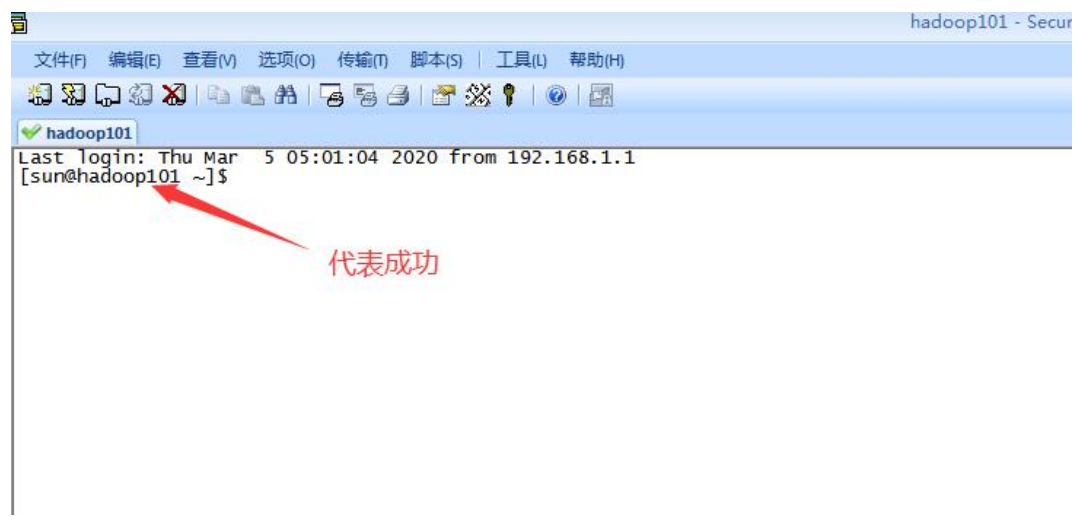
使用：解压后的目录如下图，找到 SecureCRT.exe 双击打开。



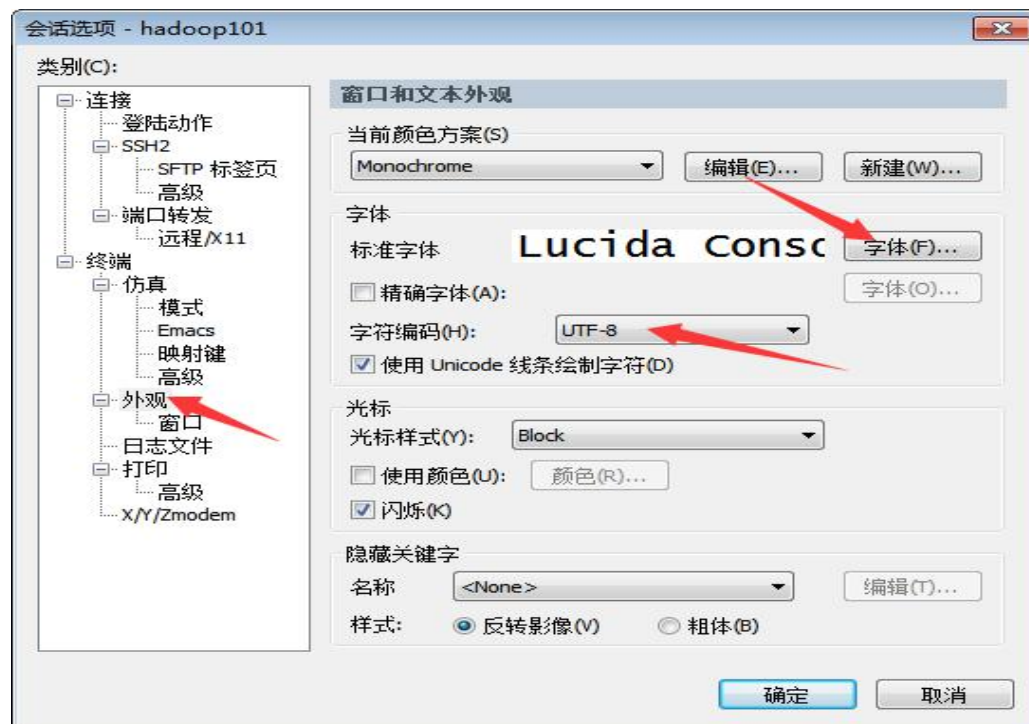
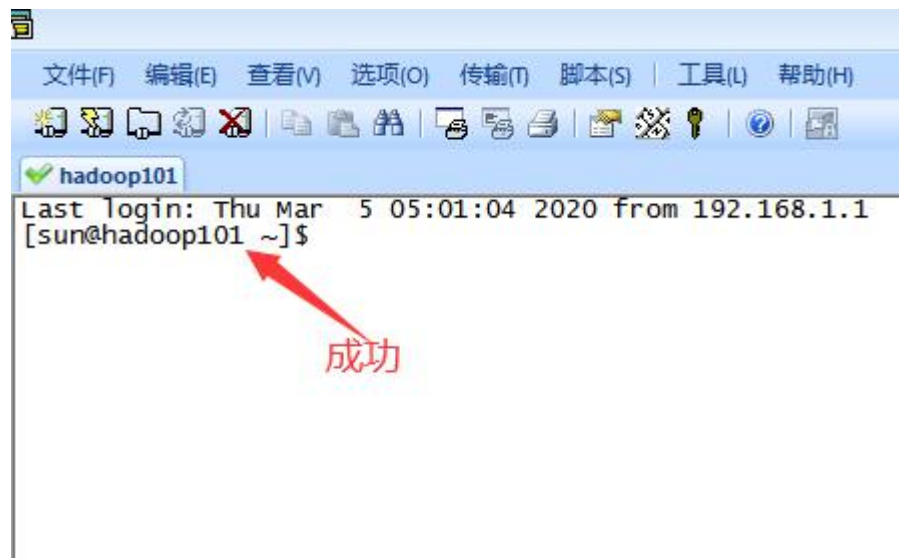
选择快速连接--这一个操作需要虚拟机打开并联网



如果弹出下面这个图:



放大字体和修改编码的操作：



11. 任务驱动

学生个人在机房安装 SecureCRT 工具

12. 总结

SecureCRT 工具是一款 Linux 连接工具，使用方便，操作简单。

参考资料 及应用	《Hadoop 大数据分析》 主编：高永彬、钱亮宏、方志军 出版社：中国铁道出版社
作业	在自己电脑上安装 SecureCRT 工具
教学反思	适当增加任务驱动法使用，提高学生兴趣

日期	2020年3月18日	周次	第三周	时数	4课时
教学单元	Hadoop 架构概述				
教学目标	知识目标： 1. HDFS 架构概述 2. YARN 架构概述 3. MapReduce 架构概述				
	技能目标：能够掌握 HDFS、YARN 架构中节点组成及各个节点的功能。				
	素质目标： 13. 提高学生逻辑思维，创新能力； 14. 树立团队协作意识； 15. 树立良好的政治素质（爱党爱国、品行端正、遵纪守法）。				
教学重点	1. HDFS 架构概述 2. YARN 架构概述 3. MapReduce 架构概述				
教学难点	MapReduce 架构概述				

教学方法	讲授演示法；任务驱动法
教学手段	云班课信息化教学；多媒体教学；钉钉直播教学

教学过程及内容

13. 课前回顾

以下是 Hadoop2.0 和 Hadoop3.0 组成



14. 知识讲解

HDFS (Hadoop Distributed File System) 的架构概述，如图 2-23 所示。

1) NameNode (nn) : 存储文件的元数据, 如文件名, 文件目录结构, 文件属性 (生成时间、副本数、文件权限), 以及每个文件的块列表和块所在的DataNode等。



2) DataNode(dn) : 在本地文件系统存储文件块数据, 以及块数据的校验和。



3) Secondary NameNode(2nn) : 用来监控HDFS状态的辅助后台程序, 每隔一段时间获取HDFS元数据的快照。

图 2-23 HDFS 架构概述

YARN 架构概述, 如图 2-24 所示。

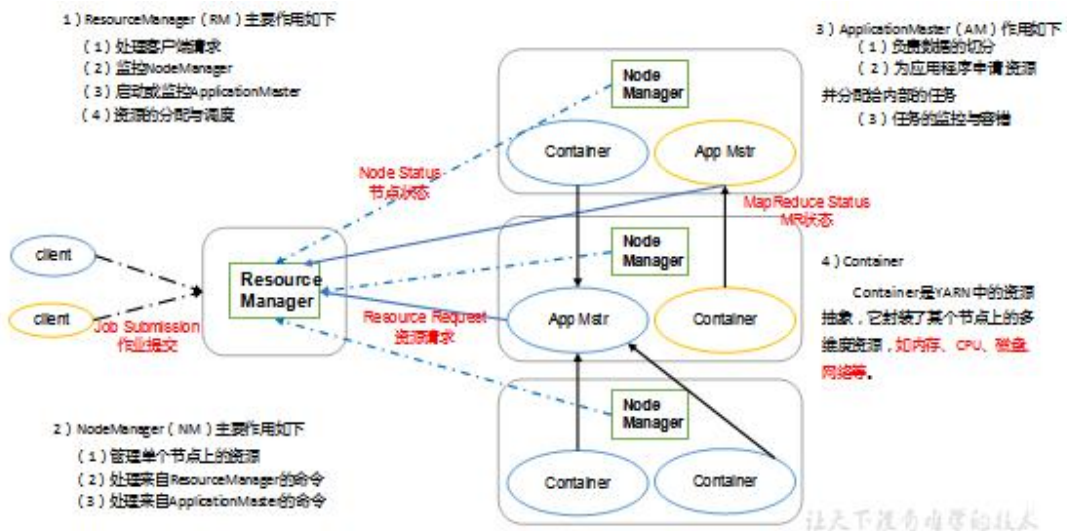


图 2-24 YARN 架构概述

MapReduce 架构概述

MapReduce 将计算过程分为两个阶段: Map 和 Reduce, 如图 2-25 所示

- 1) Map 阶段并行处理输入数据
- 2) Reduce 阶段对 Map 结果进行汇总

15. 课堂总结

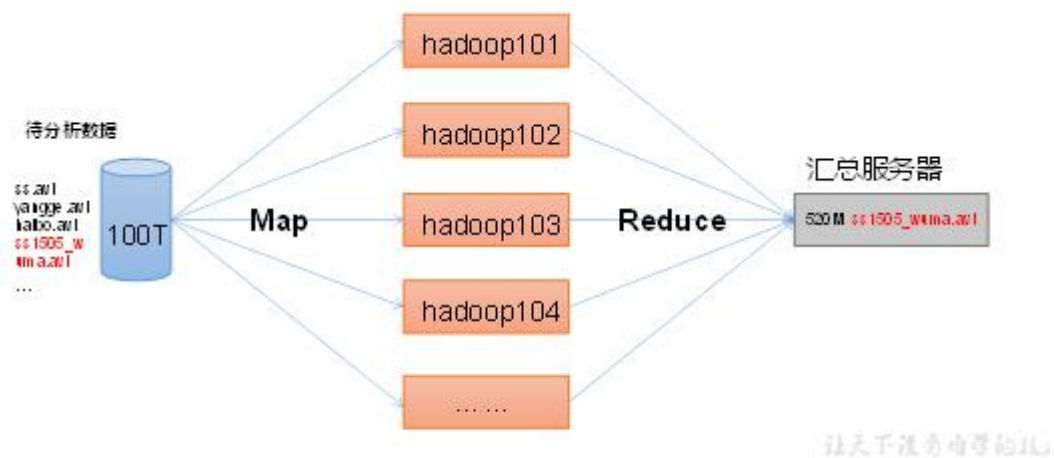


图 2-25 MapReduce 架构概述

<p>参考资料 及应用</p>	<p>《Hadoop 大数据分析》 主编：高永彬、钱亮宏、方志军 出版社：中国铁道出版社</p>
<p>作业</p>	<p>1. HDFS 架构包含哪些节点？各自有什么作用？ 2. YARN 架构包含哪些节点？各自有什么作用？</p>
<p>教学反思</p>	<p>无</p>

日期	2020年3月24日	周次	第四周	时数	4课时
教学单元	虚拟机环境准备				
教学目标	知识目标： 1. 掌握虚拟机环境准备要素 2. 了解Linux虚拟机常规操作 3. 搭建虚拟机环境				
	技能目标： 1. 能够自己克隆虚拟机 2. 能够修改虚拟机静态IP、主机名				
	素质目标： 16. 提高学生逻辑思维，创新能力； 17. 树立团队协作意识； 18. 树立良好的政治素质（爱党爱国、品行端正、遵纪守法）。				
教学重点	搭建虚拟机环境				
教学难点	搭建虚拟机环境				

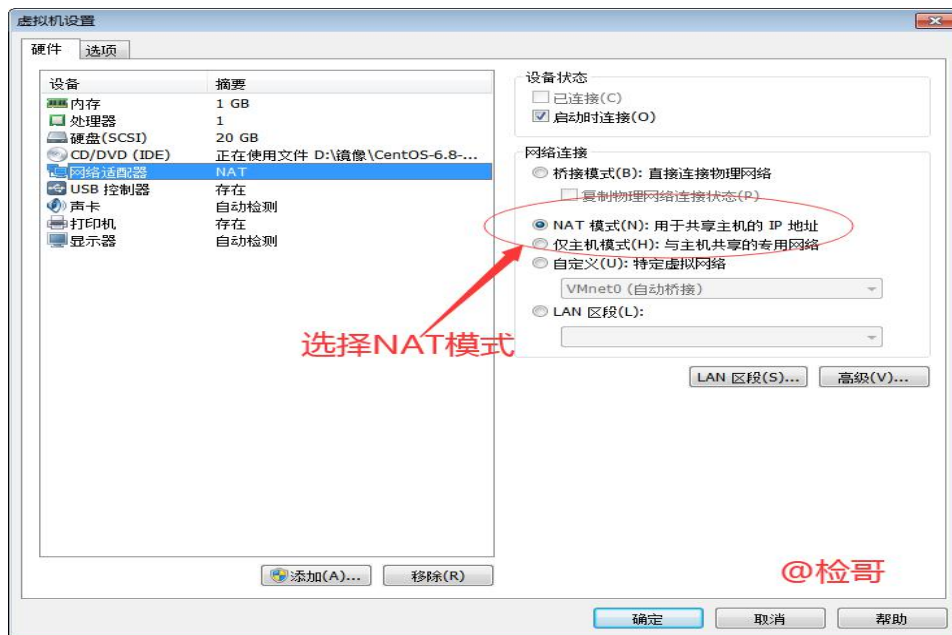
教学方法	讲授演示法；任务驱动法
------	-------------

教学手段	云班课信息化教学；多媒体教学；钉钉直播教学
------	-----------------------

教学过程及内容

课前准备

虚拟机网络模式设置为 NAT



新课导入

- 1、克隆虚拟机
- 2、详情请参考检哥 Linux 第二次课

修改克隆虚拟机的静态 IP

2.1 使用命令：`vim /etc/udev/rules.d/70-persistent-net.rules`

3、进入如下页面，删除 eth0 该行；将 eth1 修改为 eth0，同时复制物理 ip 地址

2.2 修改 IP 地址

使用命令：vim /etc/sysconfig/network-scripts/ifcfg-eth0

2.3 执行命令：service network restart 重启网络服务

```
[root@localhost 桌面]# serv
servertool  service      serviceconf
[root@localhost 桌面]# service network restart
关闭环回接口：                               [确定]
弹出环回接口：                               [确定]
弹出界面 eth0： 错误：激活连接失败：The connection is not for this device.
                                                    [失败]
[root@localhost 桌面]# service network restart
```

2.4 重启：reboot

3、修改主机名

3.1 查看主机名命令：hostname

3.2 修改主机名：vi /etc/sysconfig/network

3.3 配置主机名与 IP 同步

使用命令：vim /etc/hosts



```
sun@hadoop102:/home/sun/桌面
文件(F) 编辑(E) 查看(V) 搜索(S) 终端(T) 帮助(H)
127.0.0.1    localhost localhost.localdomain localhost4 localhost4.localdomain4
::1         localhost localhost.localdomain localhost6 localhost6.localdomain6
192.168.1.102 hadoop102
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~/etc/hosts" 4L, 183C                          3, 0-1      全部
```

3.4、重启 reboot

3.5、验证之前的操作是否成功

```
sun@hadoop102:~/桌面
文件(F) 编辑(E) 查看(V) 搜索(S) 终端(T) 帮助(H)
[sun@hadoop102 桌面]$ hostname
hadoop102
[sun@hadoop102 桌面]$ hostname -i
192.168.1.102
[sun@hadoop102 桌面]$
```

3.6 修改 window7 的 hosts 文件

- (1) 进入 C:\Windows\System32\drivers\etc 路径
- (2) 打开 hosts 文件并添加如下内容>

```
192.168.1.100 hadoop100
192.168.1.101 hadoop101
192.168.1.102 hadoop102
192.168.1.103 hadoop103
192.168.1.104 hadoop104
```

3.7 验证：打开 cmd 输入 ping hadoop101

```
C:\Users\Administrator>ping hadoop101

正在 Ping hadoop101 [192.168.1.101] 具有 32 字节的数据:
来自 192.168.1.101 的回复: 字节=32 时间=2ms TTL=64
来自 192.168.1.101 的回复: 字节=32 时间<1ms TTL=64
来自 192.168.1.101 的回复: 字节=32 时间<1ms TTL=64
来自 192.168.1.101 的回复: 字节=32 时间<1ms TTL=64

192.168.1.101 的 Ping 统计信息:
    数据包: 已发送 = 4, 已接收 = 4, 丢失 = 0 (0% 丢失),
    往返行程的估计时间(以毫秒为单位):
        最短 = 0ms, 最长 = 2ms, 平均 = 0ms
```

4. 关闭防火墙

- 1) 查看防火墙开机启动状态

```
[root@hadoop101 ~]# chkconfig iptables --list
```

2) 关闭防火墙

```
[root@hadoop101 ~]# chkconfig iptables off
```

5、创建用户

6、配置用户管理权限（暂时省略）

课堂总结

虚拟机环境准备步骤：

1、克隆虚拟机

2、修改主机名、静态 IP

3、关闭防火墙

参考资料 及应用	《Hadoop 大数据分析》 主编：高永彬、钱亮宏、方志军 出版社：中国铁道出版社
作业	完成课堂练习
教学反思	无

日期	2020年3月28日	周次	第五周	时数	4课时
教学单元	JDK 安装				
教学目标	知识目标： 1. 掌握 Linux 下 JDK 的安装与配置 2. 熟练掌握 rpm 包管理命令 3. 熟练使用 SecureCRT 工具				
	技能目标：能够在 Linux 下安装、配置 JDK				
	素质目标： 19. 提高学生逻辑思维，创新能力； 20. 树立团队协作意识； 21. 树立良好的政治素质（爱党爱国、品行端正、遵纪守法）。				
教学重点	1. 掌握 JDK 的安装与配置 2. 熟练使用 SecureCRT 工具				
教学难点	JDK 的安装与配置（配置文件修改）				
教学方法	讲授演示法；任务驱动法				

教学手段	云班课信息化教学；多媒体教学；钉钉直播教学
------	-----------------------

教学过程及内容

1、知识讲解

1.1 在/opt 目录下创建文件夹

(1) 在/opt 目录下创建 module、software 文件夹

```
[sun@hadoop101 opt]$ sudo mkdir module
```

```
[sun@hadoop101 opt]$ sudo mkdir software
```

(2) 修改 module、software 文件夹的所有者 cd

```
[sun@hadoop101 opt]$ sudo chown sun:sun module/ software/
```

```
[sun@hadoop101 opt]$ ll
```

总用量 8

```
drwxr-xr-x. 2 sun sun 4096 1月 17 14:37 module
```

```
drwxr-xr-x. 2 sun sun 4096 1月 17 14:38 software
```

1.2 安装 JDK

1. 卸载现有 JDK

(1) 查询是否安装 Java 软件：

```
[sun@hadoop101 opt]$ rpm -qa | grep java
```

(2) 如果安装的版本低于 1.7，卸载该 JDK：

```
[sun@hadoop101 opt]$ sudo rpm -e 软件包 --nodeps --force
```

查看 JDK 安装路径：

```
[sun@hadoop101 ~]$ which java
```

删除原 JDK 安装路径

```
[sun@hadoop101 ~]$ sudo rm -rf /usr/bin/java
```

2.1 用 SecureCRT 工具将 JDK 导入到 opt 目录下面的 software 文件夹下面，如图 2-28 所示

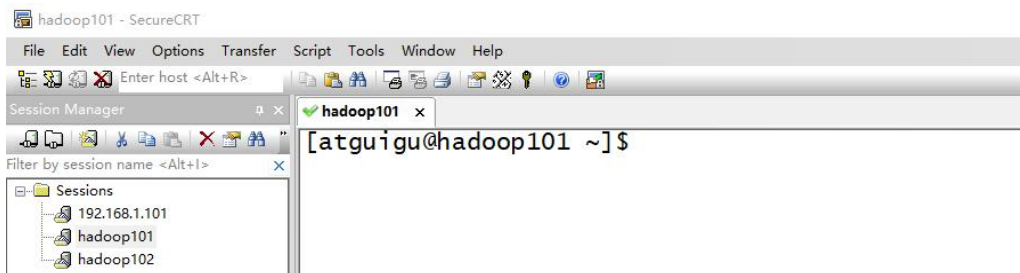


图 2-28 导入 JDK

“alt+p” 进入 sftp 模式，如图 2-29 所示

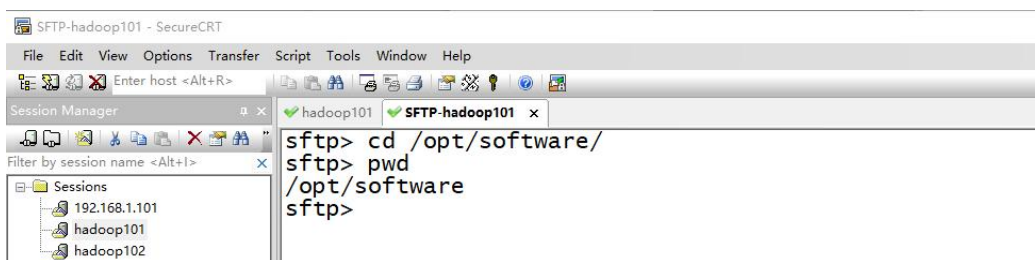


图 2-29 进入 sftp 模式

选择 jdk1.8 拖入，如图 2-30，2-31 所示

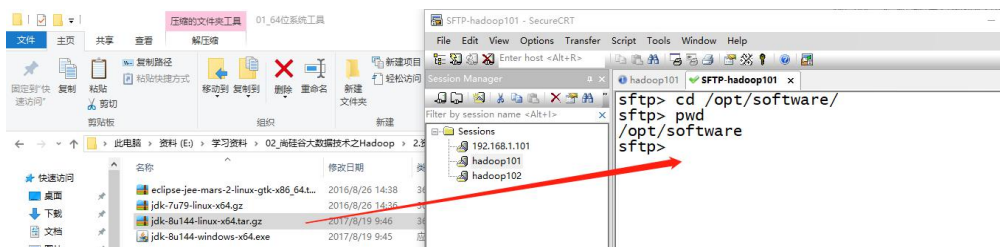


图 2-31 拖入 jdk1.8 完成

在 Linux 系统下的 opt 目录中查看软件包是否导入成功

```
[sun@hadoop101 opt]$ cd software/
```

```
[sun@hadoop101 software]$ ls
```

```
hadoop-2.7.2.tar.gz  jdk-8u144-linux-x64.tar.gz
```

解压 JDK 到/opt/module 目录下

```
[sun@hadoop101 software]$ tar -zxvf
```

```
jdk-8u144-linux-x64.tar.gz -C /opt/module/
```

2.2 配置 JDK 环境变量

(1) 先获取 JDK 路径

```
[sun@hadoop101 jdk1.8.0_144]$ pwd
```

```
/opt/module/jdk1.8.0_144
```

(2) 打开/etc/profile 文件

```
[sun@hadoop101 software]$ sudo vi /etc/profile
```

在 profile 文件末尾添加 JDK 路径

```
#JAVA_HOME
```

```
export JAVA_HOME=/opt/module/jdk1.8.0_144
```

```
export PATH=$PATH:$JAVA_HOME/bin
```

(3) 保存后退出

```
:wq
```

(4) 让修改后的文件生效

```
[sun@hadoop101 jdk1.8.0_144]$ source /etc/profile
```

2.3 测试 JDK 是否安装成功

```
[sun@hadoop101 jdk1.8.0_144]# java -version
```

```
java version "1.8.0_144"
```


注意：重启（如果 java -version 可以用就不用重启）

```
[sun@hadoop101 jdk1.8.0_144]$ sync
```

```
[sun@hadoop101 jdk1.8.0_144]$ sudo reboot
```

2、课堂总结

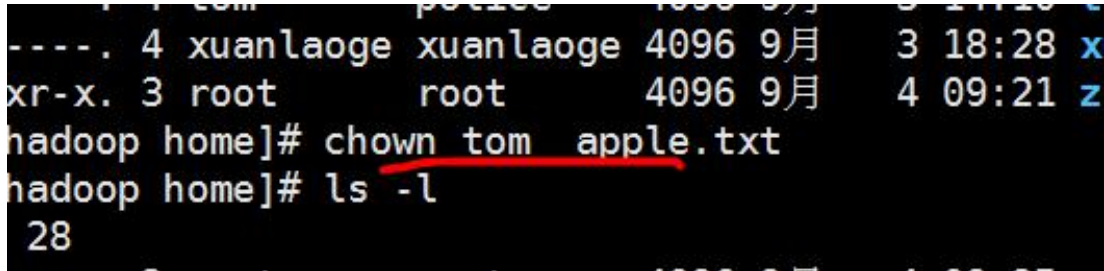
下载 JDK 安装包

修改配置文件

测试 JDK 是否安装成功

参考资料 及应用	《Hadoop 大数据分析》 主编：高永彬、钱亮宏、方志军 出版社：中国铁道出版社
作业	在自己的电脑上安装 JDK
教学反思	提高教学方法上的改革与创新

日期	2020 年 4 月 5 日	周次	第六周	时数	2 课时
教学单元	Linux 语法讲解 01				
教学目标	知识目标： 1. 掌握 sudo 命令 2. 掌握 chown 修改文件所有者 3. 熟练掌握 rpm 包的管理 4. 熟练掌握 which 命令				
	技能目标： 能够灵活运用 sudo、chown、rpm 包管理、which 等命令				
	素质目标： 22. 提高学生逻辑思维，创新能力； 23. 树立团队协作意识； 24. 树立良好的政治素质（爱党爱国、品行端正、遵纪守法）。				
教学重点	1. 掌握 sudo 命令 2. 掌握 chown 修改文件所有者 3. 熟练掌握 rpm 包的管理 4. 熟练掌握 which 命令				
教学难点	掌握 rpm 包管理命令，rpm 包各个选项的含义				

教学方法	讲授演示法；任务驱动法
教学手段	云班课信息化教学；多媒体教学；钉钉直播教学
教学过程及内容	
<p>1、课前回顾</p> <p>安装 JDK 是运用到哪些 Linux 指令？</p> <p>2、知识讲解</p> <p>1.1sudo 命令</p> <p>sudo 是 linux 系统管理指令，是允许系统管理员让普通用户执行一些或者全部的 root 命令的一个工具，如 halt, reboot, su 等等，也就是说，经由 sudo 所执行的指令就好像是 root 亲自执行。</p> <p>1.2chown 修改文件所有者</p> <p>指令：chown 用户名 文件名</p> <p>应用案例</p> <p>要求：使用 root 创建一个文件 apple.txt ，然后将其所有者修改成 tom</p>  <pre> ---. 4 xuanlaoge xuanlaoge 4096 9月 3 18:28 x xr-x. 3 root root 4096 9月 4 09:21 z hadoop home]# chown tom apple.txt hadoop home]# ls -l 28 </pre>	

1.3rpm 包的管理

介绍：一种用于互联网下载包的打包及安装工具，它包含在某些 Linux 分发版中。它生成具有.RPM 扩展名的文件。RPM 是 RedHat Package Manager (RedHat 软件包管理工具) 的缩写，类似 windows 的 setup.exe，这一文件格式名称虽然打上了 RedHat 的标志，但理念是通用的。

Linux 的分发版本都有采用 (suse, redhat, centos 等等)，可以算是公认的行业标准了。

1.3.1rpm 包查询指令

rpm包的其它查询指令:

```
rpm -qa :查询所安装的所有rpm软件包  
rpm -qa | more  
rpm -qa | grep X [rpm -qa | grep firefox ]
```

```
rpm -q 软件包名 :查询软件包是否安装  
rpm -q firefox
```

```
rpm -qi 软件包名 : 查询软件包信息  
rpm -qi file
```

```
rpm -ql 软件包名 :查询软件包中的文件  
rpm -ql firefox
```

```
rpm -qf 文件全路径名 查询文件所属的软件包  
rpm -qf /etc/passwd  
rpm -qf /root/install.log
```

1.3.2rpm 包卸载

基本语法

```
rpm -e RPM 包的名称
```

应用案例

删除 firefox 软件包

```
rpm -e firefox
```

1.3.3rpm 包安装 (JDK 解压安装—无需这么搞)

基本语法

rpm -ivh RPM 包全路径名称

参数说明

i=install 安装

v=verbose 提示

h=hash 进度条

1.4which 命令

作用：查看可执行文件的位置。

语法：which 可执行文件名称

3、课堂总结

1. sudo 命令
2. chown 修改文件所有者
3. rpm 包的管理
4. which 命令

参考资料 及应用	《Hadoop 大数据分析》 主编：高永彬、钱亮宏、方志军 出版社：中国铁道出版社
作业	1. 修改电脑用户拥有 root 权限 2. 使用 rpm 包安装软件

教学反思	设计教学环节需更加细化、合理
------	----------------

日期	2020 年 4 月 5 日	周次	第六周	时数	2 课时
教学单元	Linux 语法讲解 02				
教学目标	<p>知识目标：</p> <ol style="list-style-type: none"> 1. 熟练掌握 tar 打包指令 2. 了解 source 命令 3. 了解 sync 指令 4. 掌握 grep 指令和 管道符号 				
	<p>技能目标：</p> <ol style="list-style-type: none"> 1. 能够使用 tar 打包命令解压缩文件 2. 能够使用 grep 指令过滤文件 				
	<p>素质目标：</p> <ol style="list-style-type: none"> 25. 提高学生逻辑思维，创新能力； 26. 树立团队协作意识； 27. 树立良好的政治素质（爱党爱国、品行端正、遵纪守法）。 				
教学重点	<ol style="list-style-type: none"> 1. 掌握 tar 打包指令 2. 掌握 grep 指令和 管道符号 				
教学难点	掌握 grep 指令和 管道符号				

教学方法	讲授演示法；任务驱动法												
教学手段	云班课信息化教学；多媒体教学；钉钉直播教学												
教学过程及内容													
<p>16. 课前回顾</p> <p>在 Linux 语法讲解 01 中讲到 sudo 命、chown 修改文件所有、rpm 包的管、which 命令等</p> <p>17. 课堂示范</p> <p>2.1tar 指令</p> <p>基本说明</p> <p>tar 指令 是打包指令,最后打包后的文件是 .tar.gz 的文件。[可以压缩,和解压]</p> <p>基本语法</p> <p>tar [选项] XXX.tar.gz 打包的内容/目录 (功能描述:打包目录,压缩后的文件格式.tar.gz)</p> <p>选项说明</p> <table border="1" data-bbox="272 1780 1109 2018"> <thead> <tr> <th>选项</th> <th>功能</th> </tr> </thead> <tbody> <tr> <td>-c</td> <td>产生.tar打包文件</td> </tr> <tr> <td>-v</td> <td>显示详细信息</td> </tr> <tr> <td>-f</td> <td>指定压缩后的文件名</td> </tr> <tr> <td>-z</td> <td>打包同时压缩</td> </tr> <tr> <td>-x</td> <td>解包.tar文件</td> </tr> </tbody> </table>		选项	功能	-c	产生.tar打包文件	-v	显示详细信息	-f	指定压缩后的文件名	-z	打包同时压缩	-x	解包.tar文件
选项	功能												
-c	产生.tar打包文件												
-v	显示详细信息												
-f	指定压缩后的文件名												
-z	打包同时压缩												
-x	解包.tar文件												

通常压缩使用-zcvf 解压使用-zxvf

应用实例

1 案例 1: 压缩多个文件, 将 /home/a1.txt 和 /home/a2.txt 压缩成 a.tar.gz **【zcvf】** zxvf

```
tar -zcvf a.tar.gz a1.txt a2.txt [注意, 路径要写清楚]
```

2 案例 2: 将/home 的文件夹 压缩成 myhome.tar.gz

```
tar -zcvf myhome.tar.gz /home/ [注意, 路径写清楚]
```

3 案例 3: 将 a.tar.gz 解压到当前目录

```
tar -zxvf a.tar.gz
```

4 案例 4: 将 myhome.tar.gz 解压到 /opt/tmp2 目录下

【-C】

```
tar -zxvf myhome.tar.gz -C /opt/tmp2 [注意 ; /opt/tmp2 事先需要创建好, -C 选项的作用是: 指定需要解压到的目录。否则是在当前目录]
```

2.2source 命令

作用: 在当前 bash 环境下读取并执行 FileName 中的命令。

语法: source FileName

source (或点) 命令通常用于重新执行刚修改的初始化文档

2.3sync 指令

sync [把内存的数据, 写入磁盘], 用于数据同步, sync 命令是在关闭 Linux 系统时使用的。

2.4grep 指令和 管道符号 |

说明

grep 过滤查找，管道符，“|”，表示将前一个命令的处理结果输出传递给后面的命令处理。

基本语法

```
grep [选项] 查找内容 源文件
```

常用选项

选项	功能
-n	显示匹配行及行号。
-i	忽略字母大小写

18. 案例导入

案例 1: 请在 /etc/profile 文件中，查找 “if” 所在行，并且显示行号

```
grep -n if /etc/profile [在/etc/profile 中查找 if ,并显示行, 区别大小写]
```

```
grep -ni if /etc/profile [在/etc/profile 中查找 if ,并显示行, 不区别大小写]
```

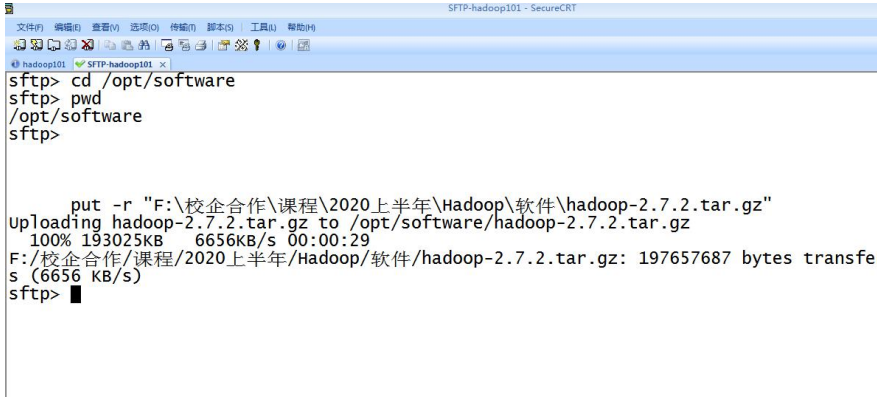
案例 2: 配合管道 | 使用案例，查看当前服务器有没有 sshd 服务在监听。

```
LISTEN 1874/sshd
[root@hadoop 桌面]# netstat -anp | grep sshd
tcp    0      0 0.0.0.0:22          0.0.0.0:*          LISTEN 1874/sshd
tcp    0      0 :::22              :::*                LISTEN 1874/sshd
[root@hadoop 桌面]# netstat -anp | grep sshd | wc -l
2
```

19. 课堂总结

<p>Linux 中命令的共同特点?</p> <p>tar 打包指令</p> <p>source 命令</p> <p>sync 指令</p> <p>grep 指令和 管道符号 </p>	
参考资料 及应用	<p>《Hadoop 大数据分析》</p> <p>主编：高永彬、钱亮宏、方志军</p> <p>出版社：中国铁道出版社</p>
作业	完成 Linux 指令讲解练习题
教学反思	实操速度较快，部分学生跟不上，适当放慢速度

日期	2020年4月11日	周次	第七周	时数	4课时
教学单元	Hadoop 安装				
教学目标	<p>知识目标：</p> <ol style="list-style-type: none"> 1. 熟练掌握 Hadoop 安装与配置 2. 掌握 Hadoop 相关配置文件 3. 了解 Hadoop 目录结构 4. 了解 Hadoop 官方文档 				
	<p>技能目标：</p> <ol style="list-style-type: none"> 1. 能够掌握 Linux 下 Hadoop 的安装与配置 2. 能够掌握 Hadoop 目录结构 				
	<p>素质目标：</p> <ol style="list-style-type: none"> 28. 提高学生逻辑思维，创新能力； 29. 树立团队协作意识； 30. 树立良好的政治素质（爱党爱国、品行端正、遵纪守法）。 				
教学重点	<ol style="list-style-type: none"> 1. Linux 下 Hadoop 的安装与配置 2. Hadoop 相关配置文件 				
教学难点	Hadoop 的安装与配置				

教学方法	讲授演示法；任务驱动法
教学手段	云班课信息化教学；多媒体教学；钉钉直播教学
教学过程及内容	
<p>20. 新课引入</p> <p>之前学过 Linux 下 JDK 的安装与配置</p> <p>21. 任务示范</p> <p>2.1 安装 Hadoop</p> <p>2.1.1 Hadoop 下载地址：</p> <p>https://archive.apache.org/dist/hadoop/common/hadoop-2.7.2/</p> <p>用 SecureCRT 工具将 hadoop-2.7.2.tar.gz 导入到 opt 目录下面的 software 文件夹下面</p> <p>切换到 sftp 连接页面，选择 Linux 下编译的 hadoop jar 包拖入</p>  <pre> sftp> cd /opt/software sftp> pwd /opt/software sftp> put -r "F:\校企合作\课程\2020上半年\Hadoop\软件\hadoop-2.7.2.tar.gz" Uploading hadoop-2.7.2.tar.gz to /opt/software/hadoop-2.7.2.tar.gz 100% 193025KB 6656KB/s 00:00:29 F:/校企合作/课程/2020上半年/Hadoop/软件/hadoop-2.7.2.tar.gz: 197657687 bytes transfere s (6656 KB/s) sftp> </pre> <p>进入到 Hadoop 安装包路径下</p>	

```
[sun@hadoop101 ~]$ cd /opt/software/
```

解压安装文件到/opt/module 下面

```
[sun@hadoop101 software]$ tar -zxvf hadoop-2.7.2.tar.gz -C  
/opt/module/
```

查看是否解压成功

```
[sun@hadoop101 software]$ ls /opt/module/  
hadoop-2.7.2
```

2.1.2 将 Hadoop 添加到环境变量

(1) 获取 Hadoop 安装路径

```
[sun@hadoop101 hadoop-2.7.2]$ pwd  
/opt/module/hadoop-2.7.2
```

(2) 打开/etc/profile 文件

```
[sun@hadoop101 hadoop-2.7.2]$ sudo vi /etc/profile
```

在 profile 文件末尾添加 JDK 路径: (shitf+g)

```
##HADOOP_HOME
```

```
export HADOOP_HOME=/opt/module/hadoop-2.7.2
```

```
export PATH=$PATH:$HADOOP_HOME/bin
```

```
export PATH=$PATH:$HADOOP_HOME/sbin
```

(3) 保存后退出

```
:wq
```

(4) 让修改后的文件生效

```
[sun@hadoop101 hadoop-2.7.2]$ source /etc/profile
```

2.1.3 测试是否安装成功

```
[sun@hadoop101 hadoop-2.7.2]$ hadoop version
```

```
Hadoop 2.7.2
```

2.1.4 重启(如果Hadoop命令不能用再重启)

```
[sun@hadoop101 hadoop-2.7.2]$ sync
```

```
[sun@hadoop101 hadoop-2.7.2]$ sudo reboot
```

2.2 Hadoop 目录结构

1、查看Hadoop目录结构

```
[sun@hadoop101 hadoop-2.7.2]$ ll
```

```
总用量 52
```

```
drwxr-xr-x. 2 sun sun 4096 5月 22 2017 bin
```

```
drwxr-xr-x. 3 sun sun 4096 5月 22 2017 etc
```

```
drwxr-xr-x. 2 sun sun 4096 5月 22 2017 include
```

```
drwxr-xr-x. 3 sun sun 4096 5月 22 2017 lib
```

```
drwxr-xr-x. 2 sun sun 4096 5月 22 2017 libexec
```

```
-rw-r--r--. 1 sun sun 15429 5月 22 2017 LICENSE.txt
```

```
-rw-r--r--. 1 sun sun 101 5月 22 2017 NOTICE.txt
```

2、重要目录

(1) bin 目录: 存放对Hadoop相关服务(HDFS, YARN)进行操作的脚本

(2) etc 目录: Hadoop的配置文件目录, 存放Hadoop的配置文件

(3) lib 目录: 存放Hadoop的本地库(对数据进行压缩解压缩功能)

(4) sbin 目录：存放启动或停止 Hadoop 相关服务的脚本

(5) share 目录：存放 Hadoop 的依赖 jar 包、文档、和官方案例

2.3 Hadoop 官方文档

官网：<https://hadoop.apache.org/>

22. 实战训练

学生个人在机房安装 Hadoop

23. 课堂总结

安装 Hadoop 的步骤：

下载、解压 Hadoop 安装包

修改 Hadoop 配置文件

测试是否安装成功

参考资料 及应用	《Hadoop 大数据分析》 主编：高永彬、钱亮宏、方志军 出版社：中国铁道出版社
作业	在自己电脑上安装 Hadoop
教学反思	实操速度较快，部分学生跟不上，适当放慢速度

日期	2020年4月15日	周次	第八周	时数	2课时
教学单元	用户管理命令 01				
教学目标	知识目标： 1. 掌握 <code>useradd</code> 添加新用户 2. 了解 <code>passwd</code> 设置用户密码 3. 了解 <code>id</code> 查看用户是否存在 4. 掌握 <code>su</code> 切换用户命令				
	技能目标： 能够使用 <code>cat /etc/passwd</code> 查看创建了哪些用户				
	素质目标： 31. 提高学生逻辑思维，创新能力； 32. 树立团队协作意识； 33. 树立良好的政治素质（爱党爱国、品行端正、遵纪守法）。				
教学重点	1. 掌握 <code>useradd</code> 添加新用户 2. 掌握 <code>su</code> 切换用户命令				
教学难点	su 切换用户				

教学方法	讲授演示法；任务驱动法
教学手段	云班课信息化教学；多媒体教学；钉钉直播教学
教学过程及内容	
<p>24. 新课倒入</p> <p>用户管理命令</p> <p>1.1 useradd 添加新用户</p> <p> 基本语法</p> <p> useradd 用户名 （功能描述：添加新用户）</p> <p> useradd -g 组名 用户名 （功能描述：添加新用户到某个组）</p> <p> 案例实操</p> <p> （1）添加一个用户</p> <p> [root@hadoop101 ~]# useradd tangseng</p> <p> [root@hadoop101 ~]#11 /home/</p> <p>1.2 passwd 设置用户密码</p> <p> 基本语法</p> <p> passwd 用户名 （功能描述：设置用户密码）</p> <p> 案例实操</p> <p> （1）设置用户的密码</p>	

```
[root@hadoop101 ~]# passwd tangseng
```

1.3 id 查看用户是否存在

基本语法

id 用户名

案例实操

(1) 查看用户是否存在

```
[root@hadoop101 ~]#id tangseng
```

1.4 cat /etc/passwd 查看创建了哪些用户

1) 基本语法

```
[root@hadoop101 ~]# cat /etc/passwd
```

1.5 su 切换用户

su: swith user 切换用户

基本语法

su 用户名称 (功能描述: 切换用户, 只能获得用户的执行权限, 不能获得环境变量)

su - 用户名称 (功能描述: 切换到用户并获得该用户的环境变量及执行权限)

案例实操

(1) 切换用户

```
[root@hadoop101 ~]#su tangseng
```

```
[root@hadoop101 ~]#echo $PATH
```

```
[root@hadoop101 ~]#exit
```

```
[root@hadoop101 ~]#su - tangseng

[root@hadoop101 ~]#echo $PATH

/usr/lib64/qt-3.3/bin:/usr/local/bin:/bin:/usr/bin:/usr
/local/sbin:/usr/sbin:/sbin:/home/tangseng/bin
```

25. 课堂总结

掌握 **useradd** 添加新用户

了解 **passwd** 设置用户密码

了解 **id** 查看用户是否存在

掌握 **su** 切换用户命令

参考资料 及应用	《Hadoop 大数据分析》 主编：高永彬、钱亮宏、方志军 出版社：中国铁道出版社
作业	完成用户管理指令练习
教学反思	部分学生指令记忆不清晰，后续组织默写

日期	2020年4月16日	周次	第八周	时数	2课时
教学单元	用户管理命令 02				
教学目标	<p>知识目标：</p> <ol style="list-style-type: none"> 1. 掌握 <code>userdel</code> 删除用户 2. 掌握 <code>who</code> 查看登录用户信息 3. 掌握 <code>sudo</code> 设置普通用户具有 <code>root</code> 权限 4. 掌握 <code>usermod</code> 修改用户 				
	<p>技能目标：</p> <ol style="list-style-type: none"> 1. 能够配置普通用户具备 <code>root</code> 权限 2. 能够对普通用户进行增加、删除、修改等操作 				
	<p>素质目标：</p> <ol style="list-style-type: none"> 34. 提高学生逻辑思维，创新能力； 35. 树立团队协作意识； 36. 树立良好的政治素质（爱党爱国、品行端正、遵纪守法）。 				
教学重点	<ol style="list-style-type: none"> 1. 使用 <code>sudo</code> 命令设置普通用户具有 <code>root</code> 权限 2. 使用 <code>usermod</code> 修改用户 				
教学难点	使用 <code>sudo</code> 命令设置普通用户具有 <code>root</code> 权限				

教学方法	讲授演示法；任务驱动法				
教学手段	云班课信息化教学；多媒体教学；钉钉直播教学				
教学过程及内容					
<p>1.1 userdel 删除用户</p> <p>1. 基本语法</p> <p> (1) userdel 用户名 (功能描述：删除用户但保存用户主目录)</p> <p> (2) userdel -r 用户名 (功能描述：用户和用户主目录，都删除)</p> <p>2. 选项说明（建议使用）</p> <p>表 1-23</p> <table border="1" data-bbox="284 1460 1308 1630"> <thead> <tr> <th data-bbox="284 1460 491 1545">选项</th> <th data-bbox="491 1460 1308 1545">功能</th> </tr> </thead> <tbody> <tr> <td data-bbox="284 1545 491 1630">-r</td> <td data-bbox="491 1545 1308 1630">删除用户的同时，删除与用户相关的所有文件。</td> </tr> </tbody> </table> <p>3. 案例实操</p> <p> (1) 删除用户但保存用户主目录</p> <pre data-bbox="242 1899 900 2024">[root@hadoop101 ~]#userdel tangseng [root@hadoop101 ~]#ll /home/</pre>		选项	功能	-r	删除用户的同时，删除与用户相关的所有文件。
选项	功能				
-r	删除用户的同时，删除与用户相关的所有文件。				

(2) 删除用户和用户主目录，都删除

```
[root@hadoop101 ~]#useradd zhubajie
```

```
[root@hadoop101 ~]#ll /home/
```

```
[root@hadoop101 ~]#userdel -r zhubajie
```

```
[root@hadoop101 ~]#ll /home/
```

1.2 who 查看登录用户信息

1. 基本语法

(1) whoami (功能描述：显示自身用户名称)

(2) who am i (功能描述：显示登录用户的用户名)

2. 案例实操

(1) 显示自身用户名称

```
[root@hadoop101 opt]# whoami
```

(2) 显示登录用户的用户名

```
[root@hadoop101 opt]# who am i
```

1.3 sudo 设置普通用户具有 root 权限

1. 添加 sun 用户，并对其设置密码。

```
[root@hadoop101 ~]#useradd sun
```

```
[root@hadoop101 ~]#passwd sun
```

2. 修改配置文件

```
[root@hadoop101 ~]#vi /etc/sudoers
```

修改 /etc/sudoers 文件，找到下面一行(91 行)，在 root 下面添加一行，如下所示：

```
## Allow root to run any commands anywhere
```

```
root    ALL=(ALL)    ALL
```

```
sun     ALL=(ALL)    ALL
```

或者配置成采用 sudo 命令时，不需要输入密码

```
## Allow root to run any commands anywhere
```

```
root    ALL=(ALL)    ALL
```

```
sun     ALL=(ALL)    NOPASSWD:ALL
```

修改完毕，现在可以用 sun 帐号登录，然后用命令 `sudo` ，即可获得 root 权限进行操作。

3. 案例实操

(1) 用普通用户在 /opt 目录下创建一个文件夹

```
[sun@hadoop101 opt]$ sudo mkdir module
```

```
[root@hadoop101 opt]# chown sun:sun module/
```

1.4 usermod 修改用户

1. 基本语法

```
usermod -g 用户组 用户名
```

2. 选项说明

表 1-24

选项	功能
-g	修改用户的初始登录组，给定的组必须存在。默认组 id 是 1。

3. 案例实操

(1) 将用户加入到用户组

```
[root@hadoop101 opt]#usermod -g root zhubajie
```

1.5 课堂总结

参考资料 及应用	《Hadoop 大数据分析》 主编：高永彬、钱亮宏、方志军 出版社：中国铁道出版社
作业	完成用户管理指令练习
教学反思	无

日期	2020年4月22日	周次	第九周	时数	4课时
教学单元	Hadoop 本地运行模式				
教学目标	<p>知识目标：</p> <ol style="list-style-type: none"> 1. 掌握 Hadoop 三种运行模式 2. 掌握 Hadoop 本地运行模式 3. 了解官方 Grep 案例 4. 了解官方 WordCount 案例 				
	<p>技能目标：</p> <ol style="list-style-type: none"> 1. 能够掌握 Hadoop 本地运行模式 2. 能够执行官方 Grep 案例 3. 能够执行官方 WordCount 案例 				
	<p>素质目标：</p> <ol style="list-style-type: none"> 37. 提高学生逻辑思维，创新能力； 38. 树立团队协作意识； 39. 树立良好的政治素质（爱党爱国、品行端正、遵纪守法）。 				
教学重点	<ol style="list-style-type: none"> 1. 掌握 Hadoop 三种运行模式 2. 掌握 Hadoop 本地运行模式 				

教学难点	执行官方 WordCount 案例
教学方法	讲授演示法；任务驱动法
教学手段	云班课信息化教学；多媒体教学；钉钉直播教学
教学过程及内容	
<p>Hadoop 运行模式</p> <p>Hadoop 运行模式包括：本地模式、伪分布式模式以及完全分布式模式。</p> <p>Hadoop 官方网站：http://hadoop.apache.org/</p> <p>本地模式（默认模式）：不需要启用单独进程，直接可以运行，测试和开发时使用。</p> <p>伪分布式模式：等同于完全分布式，只有一个节点。</p> <p>完全分布式模式：多个节点一起运行。</p> <p>1.1 本地运行模式</p> <p>1.1.1 官方 Grep 案例</p> <p>1. 创建在 hadoop-2.7.2 文件下面创建一个 input 文件夹</p> <pre>[sun@hadoop101 hadoop-2.7.2]\$ mkdir input</pre>	

2. 将 Hadoop 的 xml 配置文件复制到 input

```
[sun@hadoop101 hadoop-2.7.2]$ cp etc/hadoop/*.xml input
```

3. 执行 share 目录下的 MapReduce 程序

```
[sun@hadoop101 hadoop-2.7.2]$ bin/hadoop jar
```

```
share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.2.jar
```

```
grep input output 'dfs[a-z.]+'
```

4. 查看输出结果

```
[sun@hadoop101 hadoop-2.7.2]$ cat output/*
```

1.1.2 官方 WordCount 案例

1. 创建在 hadoop-2.7.2 文件下面创建一个 wcinput 文件夹

```
[sun@hadoop101 hadoop-2.7.2]$ mkdir wcinput
```

2. 在 wcinput 文件下创建一个 wc.input 文件

```
[sun@hadoop101 hadoop-2.7.2]$ cd wcinput
```

```
[sun@hadoop101 wcinput]$ touch wc.input
```

3. 编辑 wc.input 文件

```
[sun@hadoop101 wcinput]$ vi wc.input
```

在文件中输入如下内容

```
hadoop yarn
```

```
hadoop mapreduce
```

```
sun
```

```
sun
```

```
保存退出:: wq
```

4. 回到 Hadoop 目录/opt/module/hadoop-2.7.2

5. 执行程序

```
[sun@hadoop101 hadoop-2.7.2]$ hadoop jar  
share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.2.jar  
wordcount wcinput wcoutput
```

6. 查看结果

```
[sun@hadoop101 hadoop-2.7.2]$ cat wcoutput/part-r-00000  
sun 2  
hadoop 2  
mapreduce 1  
yarn 1
```

参考资料 及应用	《Hadoop 大数据分析》 主编：高永彬、钱亮宏、方志军 出版社：中国铁道出版社
作业	1. 本机测试官方 Grep 案例 2. 本机测试官方 WordCount 案例
教学反思	讲课太多、难度较大，适当放慢速度

日期	2020年4月25日	周次	第十周	时数	4课时
教学单元	Hadoop 伪分布式运行模式				
教学目标	知识目标： 5. 掌握 Hadoop 三种运行模式 6. 掌握 Hadoop 伪分布式运行模式				
	技能目标：能够掌握 Hadoop 伪分布式运行模式				
	素质目标： 提高学生逻辑思维，创新能力； 树立团队协作意识； 树立良好的政治素质（爱党爱国、品行端正、遵纪守法）。				
教学重点	掌握 Hadoop 伪分布式运行模式				
教学难点	Hadoop 伪分布式运行模式				
教学方法	讲授演示法；任务驱动法				
教学手段	云班课信息化教学；多媒体教学；钉钉直播教学				

教学过程及内容

1.1 伪分布式运行模式

1.1.1 启动 HDFS 并运行 MapReduce 程序

1. 分析

- (1) 配置集群
- (2) 启动、测试集群增、删、查
- (3) 执行 WordCount 案例

2. 执行步骤

(1) 配置集群

(a) 配置：hadoop-env.sh (在 etc/hadoop 目录下)

Linux 系统中获取 JDK 的安装路径：

```
[sun@ hadoop101 ~]# echo $JAVA_HOME
```

```
/opt/module/jdk1.8.0_144
```

修改 JAVA_HOME 路径：

```
export JAVA_HOME=/opt/module/jdk1.8.0_144
```

(b) 配置：core-site.xml

```
<!-- 指定 HDFS 中 NameNode 的地址 -->
```

```
<property>
```

```
<name>fs.defaultFS</name>
```

```
<value>hdfs://hadoop101:9000</value>
```

```
</property>
```

```
<!-- 指定 Hadoop 运行时产生文件的存储目录 -->
<property>
    <name>hadoop.tmp.dir</name>
    <value>/opt/module/hadoop-2.7.2/data/tmp</value>
</property>
```

(c) 配置: hdfs-site.xml

```
<!-- 指定 HDFS 副本的数量 -->
<property>
    <name>dfs.replication</name>
    <value>1</value>
</property>
```

(2) 启动集群

(a) 格式化 NameNode (第一次启动时格式化, 以后就不要总格式化)

```
[sun@hadoop101 hadoop-2.7.2]$ bin/hdfs namenode -format
```

(b) 启动 NameNode

```
[sun@hadoop101 hadoop-2.7.2]$ sbin/hadoop-daemon.sh start
namenode
```

(c) 启动 DataNode

```
[sun@hadoop101 hadoop-2.7.2]$ sbin/hadoop-daemon.sh start
datanode
```

(3) 查看集群

(a) 查看是否启动成功

```
[sun@hadoop101 hadoop-2.7.2]$ jps
```

```
13586 NameNode
```

```
13668 DataNode
```

```
13786 Jps
```

注意: jps 是 JDK 中的命令, 不是 Linux 命令。不安装 JDK 不能使用 jps

(b) web 端查看 HDFS 文件系统

<http://hadoop101:50070/dfshealth.html#tab-overview>

注意: 如果不能查看, 看如下帖子处理

<http://www.cnblogs.com/z1slch/p/6604189.html>

(c) 查看产生的 Log 日志

说明: 在企业中遇到 Bug 时, 经常根据日志提示信息去分析问题、解决 Bug。

当前目录: /opt/module/hadoop-2.7.2/logs

```
[sun@hadoop101 logs]$ ls
```

```
hadoop-sun-datanode-hadoop.sun.com.log
```

```
hadoop-sun-datanode-hadoop.sun.com.out
```

```
hadoop-sun-namenode-hadoop.sun.com.log
```

```
hadoop-sun-namenode-hadoop.sun.com.out
```

```
SecurityAuth-root.audit
```

```
[sun@hadoop101 logs]# cat hadoop-sun-datanode-hadoop101.log
```

(d) 思考：为什么不能一直格式化 NameNode，格式化 NameNode，要注意什么？

```
[sun@hadoop101 hadoop-2.7.2]$ cd data/tmp/dfs/name/current/  
[sun@hadoop101 current]$ cat VERSION
```

```
clusterID=CID-f0330a58-36fa-4a2a-a65f-2688269b5837
```

```
[sun@hadoop101 hadoop-2.7.2]$ cd data/tmp/dfs/data/current/  
clusterID=CID-f0330a58-36fa-4a2a-a65f-2688269b5837
```

注意：格式化 NameNode，会产生新的集群 id，导致 NameNode 和 DataNode 的集群 id 不一致，集群找不到已往数据。所以，格式化 NameNode 时，一定要先删除 data 数据和 log 日志，然后再格式化 NameNode。

(4) 操作集群

(a) 在 HDFS 文件系统上创建一个 input 文件夹

```
[sun@hadoop101 hadoop-2.7.2]$ bin/hdfs dfs -mkdir -p  
/user/sun/input
```

(b) 将测试文件内容上传到文件系统上

```
[sun@hadoop101 hadoop-2.7.2]$ bin/hdfs dfs -put  
wcinput/wc.input /user/sun/input/
```

(c) 查看上传的文件是否正确

```
[sun@hadoop101 hadoop-2.7.2]$ bin/hdfs dfs -ls  
/user/sun/input/
```

```
[sun@hadoop101 hadoop-2.7.2]$ bin/hdfs dfs -cat /user/sun/
```

input/wc.input

(d) 运行 MapReduce 程序

```
[sun@hadoop101 hadoop-2.7.2]$ bin/hadoop jar
share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.2.jar
wordcount /user/sun/input/ /user/sun/output
```

(e) 查看输出结果

命令行查看:

```
[sun@hadoop101 hadoop-2.7.2]$ bin/hdfs dfs -cat
/user/sun/output/*
```

浏览器查看, 如图 2-34 所示

Browse Directory

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	atguigu	supergroup	0 B	2017/12/1 上午11:05:18	1	128 MB	_SUCCESS
-rw-r--r--	atguigu	supergroup	38 B	2017/12/1 上午11:05:18	1	128 MB	part-r-00000

图 2-34 查看 output 文件

(f) 将测试文件内容下载到本地

```
[sun@hadoop101 hadoop-2.7.2]$ hdfs dfs -get
/user/sun/output/part-r-00000 ./wcoutput/
```

(g) 删除输出结果

```
[sun@hadoop101 hadoop-2.7.2]$ hdfs dfs -rm -r /user/sun/output
```

1.1.2 启动 YARN 并运行 MapReduce 程序

1. 分析

(1) 配置集群在 YARN 上运行 MR

(2) 启动、测试集群增、删、查

(3) 在 YARN 上执行 WordCount 案例

2. 执行步骤

(1) 配置集群

(a) 配置 yarn-env.sh

配置一下 JAVA_HOME

```
export JAVA_HOME=/opt/module/jdk1.8.0_144
```

(b) 配置 yarn-site.xml

```
<!-- Reducer 获取数据的方式 -->
```

```
<property>
```

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

```
  <value>mapreduce_shuffle</value>
```

```
</property>
```

(c) 配置: mapred-env.sh

配置一下 JAVA_HOME

```
export JAVA_HOME=/opt/module/jdk1.8.0_144
```

(d) 配置: (对 mapred-site.xml.template 重新命名为)

```
mapred-site.xml
```

```
[sun@hadoop101 hadoop]$ mv mapred-site.xml.template
```

```
mapred-site.xml
```

```
[sun@hadoop101 hadoop]$ vi mapred-site.xml
```

(2) 启动集群

(a) 启动前必须保证 NameNode 和 DataNode 已经启动

(b) 启动 ResourceManager

```
[sun@hadoop101 hadoop-2.7.2]$ sbin/yarn-daemon.sh start  
resourcemanager
```

(c) 启动 NodeManager

```
[sun@hadoop101 hadoop-2.7.2]$ sbin/yarn-daemon.sh start  
nodemanager
```

(3) 集群操作

(a) YARN 的浏览器页面查看, 如图 2-35 所示

<http://hadoop101:8088/cluster>

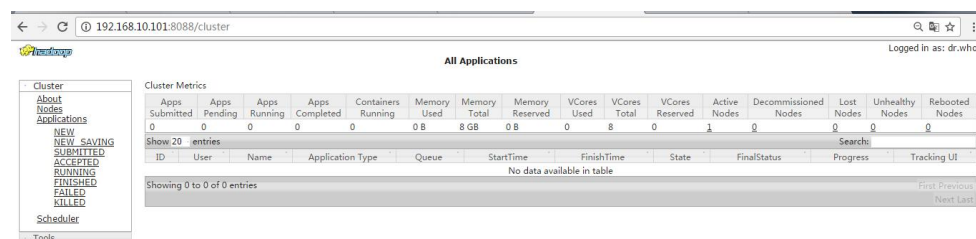


图 2-35 YARN 的浏览器页面

(b) 删除文件系统上的 output 文件

```
[sun@hadoop101 hadoop-2.7.2]$ bin/hdfs dfs -rm -r  
/user/sun/output
```

(c) 执行 MapReduce 程序

```
[sun@hadoop101 hadoop-2.7.2]$ bin/hadoop jar  
share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.2.jar  
wordcount /user/sun/input /user/sun/output
```

(d) 查看运行结果, 如图 2-36 所示

```
[sun@hadoop101 hadoop-2.7.2]$ bin/hdfs dfs -cat
/user/sun/output/*
```

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
1	0	0	1	0	0 B	8 GB	0 B	0	8	0	1	0	0	0	0

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress	Tracking UI
application_1489820373751_0001	root	word count	MAPREDUCE	default	Sat, 18 Mar 2017 07:15:25 GMT	Sat, 18 Mar 2017 07:15:42 GMT	FINISHED	SUCCEEDED		History

图 2-36 查看运行结果

<p>参考资料 及应用</p>	<p>《Hadoop 大数据分析》 主编：高永彬、钱亮宏、方志军 出版社：中国铁道出版社</p>
<p>作业</p>	<p>在自己电脑上部署 Hadoop 伪分布运行模式</p>
<p>教学反思</p>	<p>部分学生没有网络，安装软件速度较慢，搭建 Hadoop 伪分布模式未完成</p>

日期	2020年5月3日	周次	第十一周	时数	4课时
教学单元	Hadoop 集群配置				
教学目标	<p>知识目标：</p> <ol style="list-style-type: none"> 1. 了解配置历史服务器 2. 掌握配置日志的聚集 3. 了解配置文件说明 4. 掌握关闭安全模式 				
	<p>技能目标：</p> <ol style="list-style-type: none"> 1. 能够配置日志的聚集 2. 能够解决集群常见错误 				
	<p>素质目标：</p> <ol style="list-style-type: none"> 40. 提高学生逻辑思维，创新能力； 41. 树立团队协作意识； 42. 树立良好的政治素质（爱党爱国、品行端正、遵纪守法）。 				
教学重点	<ol style="list-style-type: none"> 1. 掌握配置日志的聚集 2. 掌握关闭安全模式 				
教学难点	配置日志的聚集				

教学方法	讲授演示法；任务驱动法
教学手段	云班课信息化教学；多媒体教学；钉钉直播教学
教学过程及内容	
<p>1.1.1 配置历史服务器</p> <p>为了查看程序的历史运行情况，需要配置一下历史服务器。具体配置步骤如下：</p> <p>1. 配置 mapred-site.xml</p> <pre>[sun@hadoop101 ~]\$ vi mapred-site.xml</pre> <p>在该文件里面增加如下配置。</p> <pre><!-- 历史服务器端地址 --> <property> <name>mapreduce.jobhistory.address</name> <value>hadoop101:10020</value> </property> <!-- 历史服务器 web 端地址 --> <property> <name>mapreduce.jobhistory.webapp.address</name> <value>hadoop101:19888</value></pre>	

</property>

2. 启动历史服务器

```
[sun@hadoop101 hadoop-2.7.2]$ sbin/mr-jobhistory-daemon.sh  
start historyserver
```

3. 查看历史服务器是否启动

```
[sun@hadoop101 hadoop-2.7.2]$ jps
```

4. 查看 JobHistory

<http://hadoop101:19888/jobhistory>

1.1.2 配置日志的聚集

日志聚集概念：应用运行完成以后，将程序运行日志信息上传到 HDFS 系统上。

日志聚集功能好处：可以方便的查看到程序运行详情，方便开发调试。

注意：开启日志聚集功能，需要重新启动 NodeManager 、ResourceManager 和 HistoryManager。

开启日志聚集功能具体步骤如下：

配置 yarn-site.xml

```
[sun@hadoop101 hadoop]$ vi yarn-site.xml
```

在该文件里面增加如下配置。

```
<!-- 日志聚集功能使能 -->
```

```
<property>
```

```
<name>yarn.log-aggregation-enable</name>
```

```
<value>>true</value>
```

```
</property>
```

```
<!-- 日志保留时间设置 7 天 -->
```

```
<property>
```

```
<name>yarn.log-aggregation.retain-seconds</name>
```

```
<value>604800</value>
```

```
</property>
```

关闭 NodeManager 、 ResourceManager 和 HistoryManager

```
[sun@hadoop101 hadoop-2.7.2]$ sbin/yarn-daemon.sh stop
```

```
resourcemanager
```

```
[sun@hadoop101 hadoop-2.7.2]$ sbin/yarn-daemon.sh stop
```

```
nodemanager
```

```
[sun@hadoop101 hadoop-2.7.2]$ sbin/mr-jobhistory-daemon.sh
```

```
stop historyserver
```

启动 NodeManager 、 ResourceManager 和 HistoryManager

```
[sun@hadoop101 hadoop-2.7.2]$ sbin/yarn-daemon.sh start
```

```
resourcemanager
```

```
[sun@hadoop101 hadoop-2.7.2]$ sbin/yarn-daemon.sh start
```

```
nodemanager
```

```
[sun@hadoop101 hadoop-2.7.2]$ sbin/mr-jobhistory-daemon.sh
```

```
start historyserver
```

删除 HDFS 上已经存在的输出文件

```
[sun@hadoop101 hadoop-2.7.2]$ bin/hdfs dfs -rm -R
```

/user/sun/output

执行 WordCount 程序

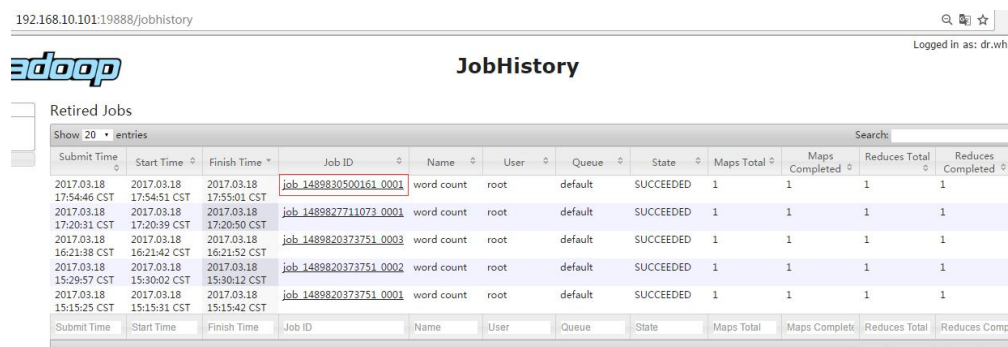
```
[sun@hadoop101 hadoop-2.7.2]$ hadoop jar
```

```
share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.2.jar
```

```
wordcount /user/sun/input /user/sun/output
```

查看日志，如图 2-37，2-38，2-39 所示

<http://hadoop101:19888/jobhistory>



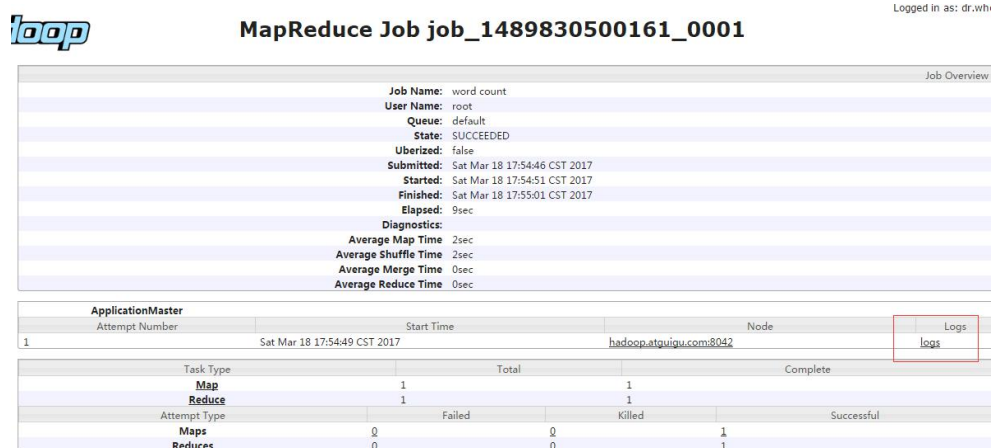
192.168.10.101:19888/jobhistory

JobHistory

Retired Jobs

Submit Time	Start Time	Finish Time	Job ID	Name	User	Queue	State	Maps Total	Maps Completed	Reduces Total	Reduces Completed
2017.03.18 17:54:46 CST	2017.03.18 17:54:51 CST	2017.03.18 17:55:01 CST	job_1489830500161_0001	word count	root	default	SUCCEEDED	1	1	1	1
2017.03.18 17:20:31 CST	2017.03.18 17:20:39 CST	2017.03.18 17:20:50 CST	job_1489827711073_0001	word count	root	default	SUCCEEDED	1	1	1	1
2017.03.18 16:21:38 CST	2017.03.18 16:21:42 CST	2017.03.18 16:21:52 CST	job_1489820373751_0003	word count	root	default	SUCCEEDED	1	1	1	1
2017.03.18 15:29:57 CST	2017.03.18 15:30:02 CST	2017.03.18 15:30:12 CST	job_1489820373751_0002	word count	root	default	SUCCEEDED	1	1	1	1
2017.03.18 15:15:25 CST	2017.03.18 15:15:31 CST	2017.03.18 15:15:42 CST	job_1489820373751_0001	word count	root	default	SUCCEEDED	1	1	1	1

图 2-37 Job History



MapReduce Job job_1489830500161_0001

Job Overview

Job Name: word count
User Name: root
Queue: default
State: SUCCEEDED
Uberized: false
Submitted: Sat Mar 18 17:54:46 CST 2017
Started: Sat Mar 18 17:54:51 CST 2017
Finished: Sat Mar 18 17:55:01 CST 2017
Elapsed: 9sec
Diagnostics:
Average Map Time: 2sec
Average Shuffle Time: 2sec
Average Merge Time: 0sec
Average Reduce Time: 0sec

Attempt Number	Start Time	Node	Logs
1	Sat Mar 18 17:54:49 CST 2017	hadoopatguigu.com:8042	logs

Task Type	Total	Complete
Map	1	1
Reduce	1	1

Attempt Type	Failed	Killed	Successful
Maps	0	0	1
Reduces	0	0	1

图 2-38 job 运行情况



图 2-39 查看日志

1.1.3 配置文件说明

Hadoop 配置文件分两类：默认配置文件和自定义配置文件，只有用户想修改某一默认配置值时，才需要修改自定义配置文件，更改相应属性值。

(1) 默认配置文件：

表 2-1

要获取的默认文件	文件存放在 Hadoop 的 jar 包中的位置
[core-default.xml]	hadoop-common-2.7.2.jar/ core-default.xml
[hdfs-default.xml]	hadoop-hdfs-2.7.2.jar/ hdfs-default.xml
[yarn-default.xml]	hadoop-yarn-common-2.7.2.jar/ yarn-default.xml
[mapred-default.x	hadoop-mapreduce-client-core-2.7.2

ml]	. jar/ mapred-default.xml
<p>(2) 自定义配置文件:</p> <p>core-site.xml 、 hdfs-site.xml 、 yarn-site.xml 、 mapred-site.xml 四个配置文件存放在\$HADOOP_HOME/etc/hadoop 这个路径上, 用户可以根据项目需求重新进行修改配置。</p> <p>1.1.4 关闭安全模式</p> <pre>[sun@hadoop101 hadoop-2.7.2]\$ bin/hadoop dfsadmin -safemode leave</pre>	
参考资料及应用	<p>《Hadoop 大数据分析》</p> <p>主编: 高永彬、钱亮宏、方志军</p> <p>出版社: 中国铁道出版社</p>
作业	完成 jdk 的安装、环境配置
教学反思	任务驱动教学学生学习效率大大提高

日期	2020年5月5日	周次	第十二周	时数	4课时
教学单元	Hadoop 完全分布式运行模式				
教学目标	知识目标： 1. 编写集群分发脚本 <code>xsync</code> 2. 掌握 Hadoop 完全分布式运行模式 3. 掌握集群单点启动				
	技能目标： 1. 能够搭建 Hadoop 完全分布式运行模式 2. 能够编写集群分发脚本 <code>xsync</code>				
	素质目标： 43. 提高学生逻辑思维，创新能力； 44. 树立团队协作意识； 45. 树立良好的政治素质（爱党爱国、品行端正、遵纪守法）。				
教学重点	掌握 Hadoop 完全分布式运行模式				
教学难点	1. 编写集群分发脚本 <code>xsync</code> 2. 掌握 Hadoop 完全分布式运行模式				

教学方法	讲授演示法；任务驱动法
教学手段	云班课信息化教学；多媒体教学；钉钉直播教学
教学过程及内容	
<p>1、 编写集群分发脚本 xsync</p> <p>1. scp (secure copy) 安全拷贝</p> <p>(1) scp 定义:</p> <p>scp 可以实现服务器与服务器之间的数据拷贝。(from server1 to server2)</p> <p>(2) 基本语法</p> <pre>scp -r \$pdir/\$fname \$user@hadoop\$host:\$pdir/\$fname</pre> <p>命令 递归 要拷贝的文件路径/名称 目的用户@主机:目的的路径/名称</p> <p>(3) 案例实操</p> <p>(a) 在 hadoop101 上, 将 hadoop101 中/opt/module 目录下的软件拷贝到 hadoop102 上。</p> <pre>[sun@hadoop101 ~]\$ scp -r /opt/module root@hadoop102:/opt/module</pre>	

(b) 在 hadoop103 上, 将 hadoop101 服务器上的/opt/module 目录下的软件拷贝到 hadoop103 上。

```
[sun@hadoop103 opt]$sudo scp -r sun@hadoop101:/opt/module  
root@hadoop103:/opt/module
```

(c) 在 hadoop103 上操作将 hadoop101 中/opt/module 目录下的软件拷贝到 hadoop104 上。

```
[sun@hadoop103 opt]$ scp -r sun@hadoop101:/opt/module  
root@hadoop104:/opt/module
```

注意: 拷贝过来的 /opt/module 目录, 别忘了在 hadoop102、hadoop103、hadoop104 上修改所有文件的, 所有者和所有者组。sudo chown sun:sun -R /opt/module

(d) 将 hadoop101 中 /etc/profile 文件拷贝到 hadoop102 的 /etc/profile 上。

```
[sun@hadoop101 ~]$ sudo scp /etc/profile  
root@hadoop102:/etc/profile
```

(e) 将 hadoop101 中 /etc/profile 文件拷贝到 hadoop103 的 /etc/profile 上。

```
[sun@hadoop101 ~]$ sudo scp /etc/profile  
root@hadoop103:/etc/profile
```

(f) 将 hadoop101 中 /etc/profile 文件拷贝到 hadoop104 的 /etc/profile 上。

```
[sun@hadoop101 ~]$ sudo scp /etc/profile
```



```
root@hadoop104:/etc/profile
```

注意：拷贝过来的配置文件别忘了 source 一下/etc/profile,。

2. rsync 远程同步工具

rsync 主要用于备份和镜像。具有速度快、避免复制相同内容和支持符号链接的优点。

rsync 和 scp 区别：用 rsync 做文件的复制要比 scp 的速度快，rsync 只对差异文件做更新。scp 是把所有文件都复制过去。

(1) 基本语法

```
rsync          -rvl          $pdir/$fname
```

```
$user@hadoop$host:$pdir/$fname
```

命令 选项参数 要拷贝的文件路径/名称 目的用户@主机:目的路径/名称

选项参数说明

表 2-2

选项	功能
-r	递归
-v	显示复制过程
-l	拷贝符号连接

(2) 案例实操

(a) 把 hadoop101 机器上的 /opt/software 目录同步到 hadoop102 服务器的 root 用户下的 /opt/ 目录

```
[sun@hadoop101 opt]$ rsync -rvl /opt/software/
```

```
root@hadoop102:/opt/software
```

3. xsync 集群分发脚本

(1) 需求：循环复制文件到所有节点的不同目录下

(2) 需求分析：

(a) rsync 命令原始拷贝：

```
rsync -rvl /opt/module root@hadoop103:/opt/
```

(b) 期望脚本：

xsync 要同步的文件名称

(c) 说明：在/home/sun/bin 这个目录下存放的脚本，sun 用户可以在系统任何地方直接执行。

(3) 脚本实现

(a) 在/home/sun 目录下创建 bin 目录，并在 bin 目录下 xsync 创建文件，文件内容如下：

```
[sun@hadoop102 ~]$ mkdir bin
```

```
[sun@hadoop102 ~]$ cd bin/
```

```
[sun@hadoop102 bin]$ touch xsync
```

```
[sun@hadoop102 bin]$ vi xsync
```

在该文件中编写如下代码

```
#!/bin/bash
```

```
#1 获取输入参数个数，如果没有参数，直接退出
```

```
pcount=$#
```

```
if((pcount==0)); then
```

```
echo no args;

exit;

fi

#2 获取文件名称

p1=$1

fname=`basename $p1`

echo fname=$fname

#3 获取上级目录到绝对路径

pdir=`cd -P $(dirname $p1); pwd`

echo pdir=$pdir

#4 获取当前用户名称

user=`whoami`

#5 循环

for((host=101; host<104; host++)); do

    echo ----- hadoop$host -----

    rsync -rvl $pdir/$fname $user@hadoop$host:$pdir

done
```

(b) 修改脚本 xsync 具有执行权限

```
[sun@hadoop102 bin]$ chmod 777 xsync
```

(c) 调用脚本形式: xsync 文件名称

```
[sun@hadoop102 bin]$ xsync /home/sun/bin
```

注意: 如果将 xsync 放到/home/sun/bin 目录下仍然不能实现全局使

用，可以将 xsync 移动到/usr/local/bin 目录下。

2、 集群配置

1. 集群部署规划

表 2-3

	hadoop102	hadoop103	hadoop104
HDFS	NameNode		SecondaryNameNode
	DataNode	DataNode	DataNode
YARN		ResourceManager	
	NodeManager	NodeManager	NodeManager

2. 配置集群

(1) 核心配置文件

配置 core-site.xml

```
[sun@hadoop102 hadoop]$ vi core-site.xml
```

在该文件中编写如下配置

```
<!-- 指定 HDFS 中 NameNode 的地址 -->
```

```
<property>
```

```
    <name>fs.defaultFS</name>
```

```
    <value>hdfs://hadoop102:9000</value>
```

```
</property>
```

```
<!-- 指定 Hadoop 运行时产生文件的存储目录 -->
```

```
<property>
```

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

```
<value>/opt/module/hadoop-2.7.2/data/tmp</value>
```

```
</property>
```

(2) HDFS 配置文件

配置 `hadoop-env.sh` --JDK 路径

```
[sun@hadoop102 hadoop]$ vi hadoop-env.sh
```

```
export JAVA_HOME=/opt/module/jdk1.8.0_144
```

配置 `hdfs-site.xml`

```
[sun@hadoop102 hadoop]$ vi hdfs-site.xml
```

```
<!-- 指定 Hadoop 辅助名称节点主机配置 -->
```

```
<property>
```

```
<name>dfs.namenode.secondary.http-address</name>
```

```
<value>hadoop104:50090</value>
```

```
</property>
```

(3) YARN 配置文件

配置 `yarn-env.sh`

```
[sun@hadoop102 hadoop]$ vi yarn-env.sh
```

```
export JAVA_HOME=/opt/module/jdk1.8.0_144
```

配置 `yarn-site.xml`

```
[sun@hadoop102 hadoop]$ vi yarn-site.xml
```

在该文件中增加如下配置

```
<!-- Reducer 获取数据的方式 -->
```

```
<property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
</property>
```

<!-- 指定 YARN 的 ResourceManager 的地址 -->

```
<property>
    <name>yarn.resourcemanager.hostname</name>
    <value>hadoop103</value>
</property>
```

(4) MapReduce 配置文件

配置 mapred-env.sh

```
[sun@hadoop102 hadoop]$ vi mapred-env.sh
```

```
export JAVA_HOME=/opt/module/jdk1.8.0_144
```

配置 mapred-site.xml

```
[sun@hadoop102 hadoop]$ cp mapred-site.xml.template
```

```
mapred-site.xml
```

```
[sun@hadoop102 hadoop]$ vi mapred-site.xml
```

在该文件中增加如下配置

<!-- 指定 MR 运行在 Yarn 上 -->

```
<property>
    <name>mapreduce.framework.name</name>
```

<pre><value>yarn</value></pre> <pre></property></pre> <p>3. 在集群上分发配置好的 Hadoop 配置文件</p> <pre>[sun@hadoop102 hadoop]\$ xsync /opt/module/hadoop-2.7.2/</pre> <p>4. 查看文件分发情况</p> <pre>[sun@hadoop103 hadoop]\$ cat /opt/module/hadoop-2.7.2/etc/hadoop/core-site.xml</pre>	
参考资料及应用	<p>《Hadoop 大数据分析》</p> <p>主编：高永彬、钱亮宏、方志军</p> <p>出版社：中国铁道出版社</p>
作业	完成课堂练习
教学反思	任务驱动教学学生学习效率大大提高

日期	2020年5月14日	周次	第十三周	时数	4课时
教学单元	HDFS 客户端操作				
教学目标	知识目标： 1. 掌握 HDFS 客户端环境准备 2. 掌握 HDFS 文件上传、下载 3. 掌握 HDFS 文件夹操作				
	技能目标： 1. 能够搭建 maven 项目访问 HDFS 文件系统 2. 能够掌握 HDFS API 操作				
	素质目标： 46. 提高学生逻辑思维，创新能力； 47. 树立团队协作意识； 48. 树立良好的政治素质（爱党爱国、品行端正、遵纪守法）。				
教学重点	掌握 HDFS 文件上传、下载				
教学难点	1. 掌握 HDFS 客户端环境准备 2. 掌握 HDFS 文件上传、下载 3. 掌握 HDFS 文件夹操作				

教学方法	讲授演示法；任务驱动法
教学手段	云班课信息化教学；多媒体教学；钉钉直播教学

教学过程及内容

1、HDFS 客户端环境准备

1. 根据自己电脑的操作系统拷贝对应的编译后的 hadoop jar 包到非中文路径（例如：D:\softwareAn\HadoopAn\hadoop-2.7.2），如图 3-4 所示。

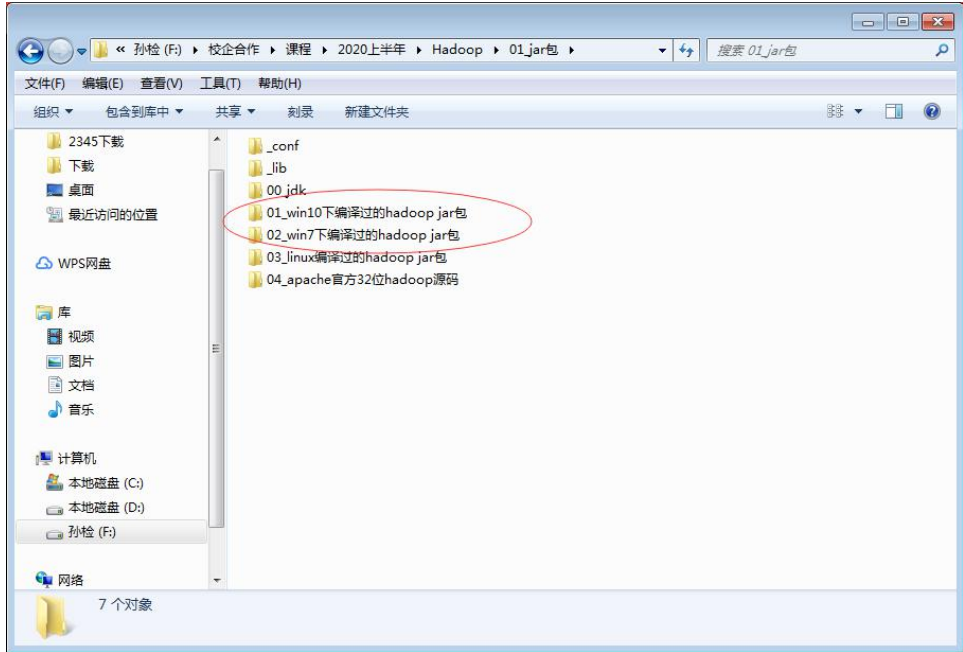


图 3-4 编译后的 hadoop jar 包

2. 配置 HADOOP_HOME 环境变量，如图 3-5 所示。

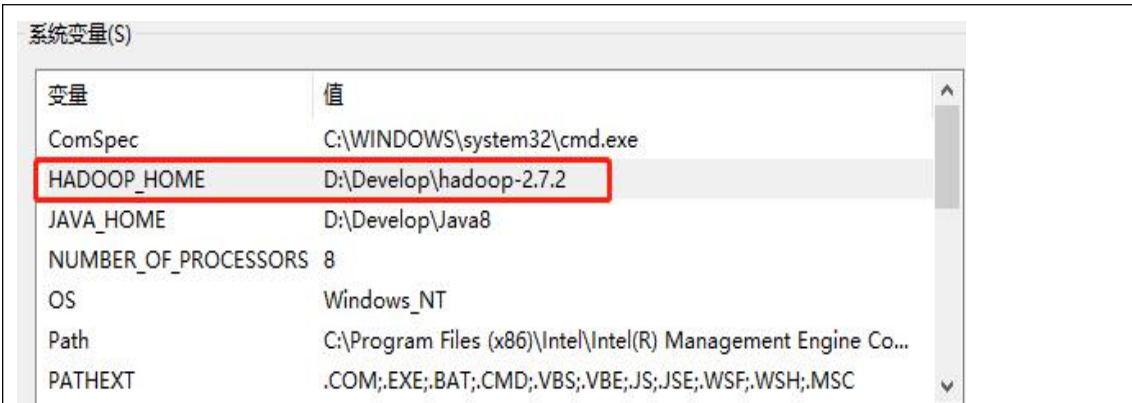


图 3-5 配置 HADOOP_HOME 环境变量

3. 配置 Path 环境变量，如图 3-6 所示。

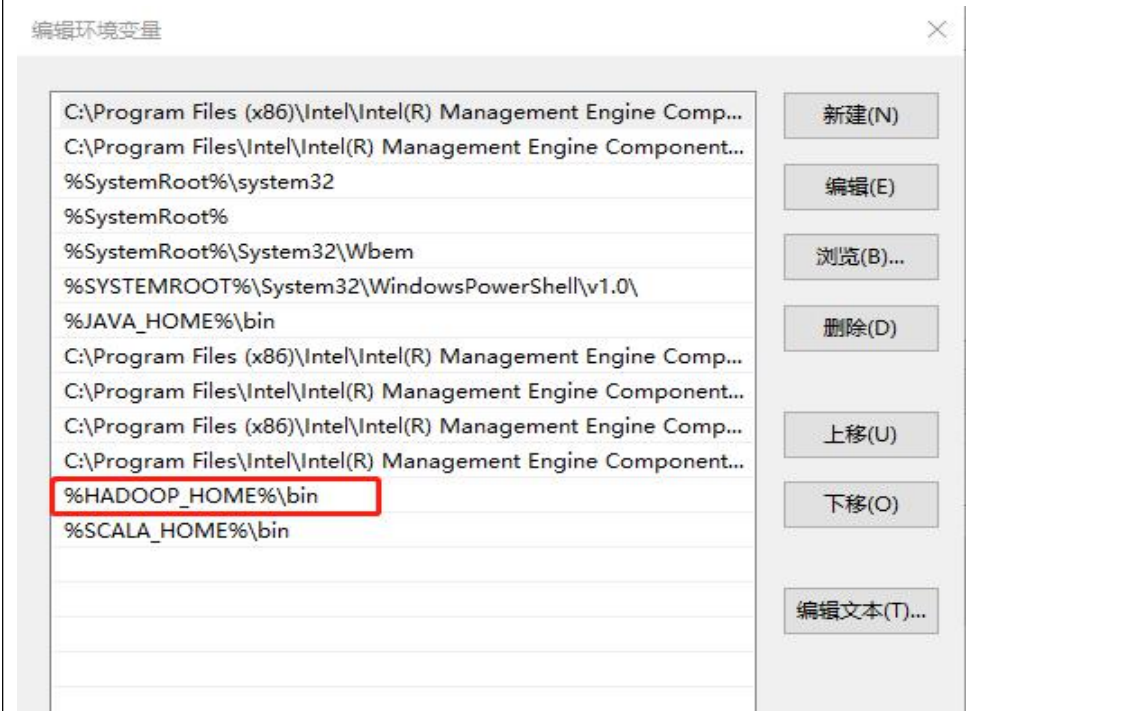


图 3-6 配置 Path 环境变量

4. 创建一个 Maven 工程 HdfsClientDemo

5. 导入相应的依赖坐标+日志添加

```
<dependencies>
```

```
  <dependency>
```

```
    <groupId>junit</groupId>
```

```
<artifactId>junit</artifactId>
  <version>RELEASE</version>
</dependency>
<dependency>
  <groupId>org. apache. logging. log4j</groupId>
  <artifactId>log4j-core</artifactId>
  <version>2. 8. 2</version>
</dependency>
<dependency>
  <groupId>org. apache. hadoop</groupId>
  <artifactId>hadoop-common</artifactId>
  <version>2. 7. 2</version>
</dependency>
<dependency>
  <groupId>org. apache. hadoop</groupId>
  <artifactId>hadoop-client</artifactId>
  <version>2. 7. 2</version>
</dependency>
<dependency>
  <groupId>org. apache. hadoop</groupId>
  <artifactId>hadoop-hdfs</artifactId>
  <version>2. 7. 2</version>
```

```
</dependency>

<dependency>

    <groupId>jdk. tools</groupId>

    <artifactId>jdk. tools</artifactId>

    <version>1.8</version>

    <scope>system</scope>

    <systemPath>${JAVA_HOME}/lib/tools.jar</systemPath>

</dependency>
```

```
</dependencies>
```

注意：如果 Eclipse/Idea 打印不出日志，在控制台上只显示

1. log4j:WARN No appenders could be found for logger (org.apache.hadoop.util.Shell).
2. log4j:WARN Please initialize the log4j system properly.
3. log4j:WARN See <http://logging.apache.org/log4j/1.2/faq.html#noconfig> for more info.

需要在项目的 src/main/resources 目录下，新建一个文件，命名为

“log4j.properties”，在文件中填入

```
log4j.rootLogger=INFO, stdout
```

```
log4j.appender.stdout=org.apache.log4j.ConsoleAppender
```

```
log4j.appender.stdout.layout=org.apache.log4j.PatternLayout
```

```
log4j.appender.stdout.layout.ConversionPattern=%d %p [%c]
```

```
- %m%n
```

```
log4j.appender.logfile=org.apache.log4j.FileAppender
```

```
log4j.appender.logfile.File=target/spring.log
```

```
log4j.appender.logfile.layout=org.apache.log4j.PatternLayout
```

```
log4j.appender.logfile.layout.ConversionPattern=%d %p [%c]
```

```
- %m%n
```

6. 创建包名: com.sun.hdfs

7. 创建 HdfsClient 类

```
public class HdfsClient {  
  
    @Test  
  
    public void testMkdirs() throws IOException,  
        InterruptedException, URISyntaxException {  
  
        // 1 获取文件系统  
  
        Configuration configuration = new Configuration();  
  
        // 配置在集群上运行  
  
        // configuration.set("fs.defaultFS",  
"hdfs://hadoop102:9000");  
  
        // FileSystem fs = FileSystem.get(configuration);  
  
        FileSystem fs = FileSystem.get(new  
URI("hdfs://hadoop102:9000"), configuration, "sun");  
  
        // 2 创建目录
```

```
fs.mkdirs(new Path("/1108/daxian/banzhang"));  
  
// 3 关闭资源  
  
fs.close();  
  
}  
  
}
```

8. 执行程序

运行时需要配置用户名称，如图 3-7 所示

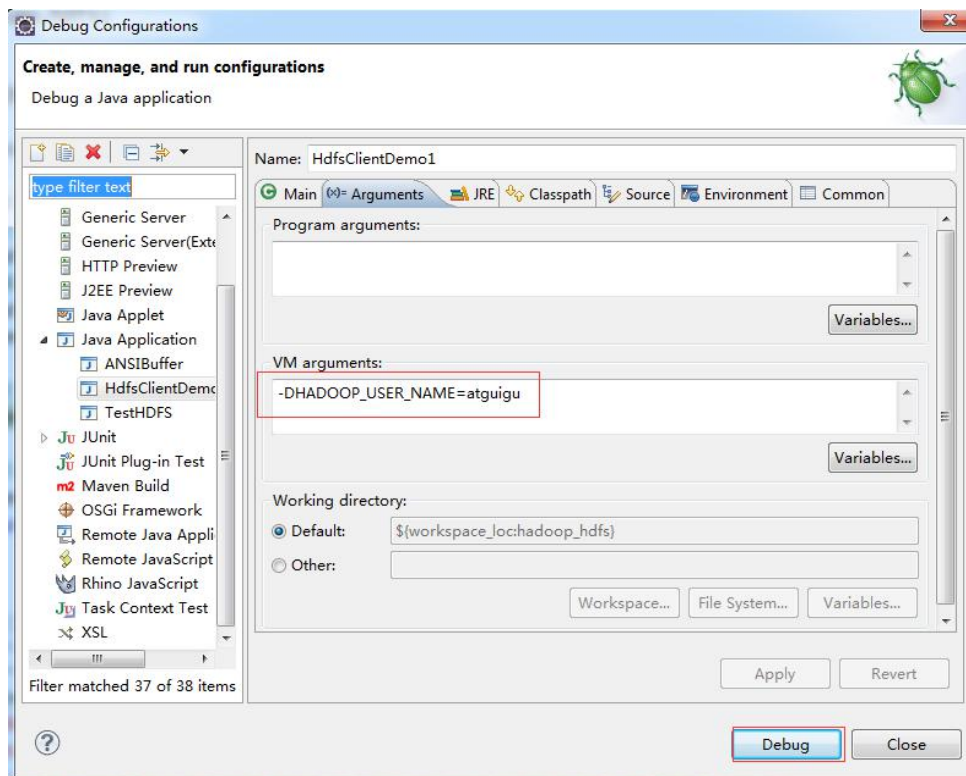


图 3-7 配置用户名称

客户端去操作 HDFS 时，是有一个用户身份的。默认情况下，HDFS 客户端 API 会从 JVM 中获取一个参数来作为自己的用户身份：
-DHADOOP_USER_NAME=sun，sun 为用户名称。

参考资料 及应用	《Hadoop 大数据分析》 主编：高永彬、钱亮宏、方志军 出版社：中国铁道出版社
作业	使用 maven 项目访问 HDFS 文件系统
教学反思	任务驱动教学学生学习效率大大提高

日期	2020年5月18日	周次	第十四周	时数	4课时
教学单元	HDFS 文件夹操作				
教学目标	知识目标： 1. 掌握 HDFS 文件夹删除 2. 掌握 HDFS 文件名更改 3. 掌握 HDFS 文件详情查看 4. 了解 HDFS 文件和文件夹判断				
	技能目标： 能够掌握 HDFS 文件夹删除、重命名等操作				
	素质目标： 49. 提高学生逻辑思维，创新能力； 50. 树立团队协作意识； 51. 树立良好的政治素质（爱党爱国、品行端正、遵纪守法）。				
教学重点	1. 掌握 HDFS 文件夹删除 2. 掌握 HDFS 文件名更改 3. 掌握 HDFS 文件详情查看				
教学难点	掌握 HDFS 文件名更改				

教学方法	讲授演示法；任务驱动法
教学手段	云班课信息化教学；多媒体教学；钉钉直播教学
教学过程及内容	
<p>1、HDFS 文件夹删除</p> <pre>@Test public void testDelete() throws IOException, InterruptedException, URISyntaxException{ // 1 获取文件系统 Configuration configuration = new Configuration(); FileSystem fs = FileSystem.get(new URI("hdfs://hadoop102:9000"), configuration, "sun"); // 2 执行删除 fs.delete(new Path("/0508/"), true); // 3 关闭资源 fs.close(); }</pre> <p>2、HDFS 文件名更改</p> <pre>@Test</pre>	

```

public void testRename() throws IOException,
InterruptedException, URISyntaxException{
    // 1 获取文件系统
    Configuration configuration = new Configuration();
    FileSystem fs = FileSystem.get(new
URI("hdfs://hadoop102:9000"), configuration, "sun");
    // 2 修改文件名称
    fs.rename(new Path("/banzhang.txt"), new
Path("/banhua.txt"));
    // 3 关闭资源
    fs.close();
}

```

3、HDFS 文件详情查看

查看文件名称、权限、长度、块信息

@Test

```

public void testListFiles() throws IOException,
InterruptedException, URISyntaxException{
    // 1 获取文件系统
    Configuration configuration = new Configuration();
    FileSystem fs = FileSystem.get(new
URI("hdfs://hadoop102:9000"), configuration, "sun");
    // 2 获取文件详情

```

```
RemoteIterator<LocatedFileStatus> listFiles =
fs.listFiles(new Path("/"), true);

while(listFiles.hasNext()) {

    LocatedFileStatus status = listFiles.next();

    // 输出详情

    // 文件名称

    System.out.println(status.getPath().getName());

    // 长度

    System.out.println(status.getLen());

    // 权限

    System.out.println(status.getPermission());

    // 分组

    System.out.println(status.getGroup());

    // 获取存储的块信息

    BlockLocation[] blockLocations =
status.getBlockLocations();

    for (BlockLocation blockLocation : blockLocations) {

        // 获取块存储的主机节点

        String[] hosts = blockLocation.getHosts();

        for (String host : hosts) {

            System.out.println(host);

        }

    }

}
```

```

    }

    System.out.println("-----班长的分割线
-----");

}

// 3 关闭资源
fs.close();
}

4、 HDFS 文件和文件夹判断

@Test
public void testListStatus() throws IOException,
InterruptedException, URISyntaxException{

    // 1 获取文件配置信息

    Configuration configuration = new Configuration();

    FileSystem fs = FileSystem.get(new
URI("hdfs://hadoop102:9000"), configuration, "sun");

    // 2 判断是文件还是文件夹

    FileStatus[] listStatus = fs.listStatus(new Path("/"));

    for (FileStatus fileStatus : listStatus) {

        // 如果是文件

        if (fileStatus.isFile()) {

            System.out.println("f:"+fileStatus.getPath().getName());

        }else {

```

```
System.out.println("d:"+fileStatus.getPath().getName());

    }

}

// 3 关闭资源

fs.close();

}
```

注意事项：检哥报错总结

The screenshot shows a CSDN article page. On the left, there is a sidebar with '15万+ 访问' (150,000+ visits), '等级' (level), and a '关注' (follow) button. The main content area has a title 'java.lang.NoClassDefFoundError:com/google/common/base/Moreobjects' and a sub-header '原创 耗子他大哥 2018-06-29 16:26:18 7249 收藏 1'. Below the title, there is a warning: '项目所属环境不同，解决方案不一定适合所有人，见谅!!!'. The main text describes a 'NoClassDefFoundError' and provides a solution: '编译可以通过，运行报错。很无奈，突然蹦出这鬼东西，经过一番猜想再结合项目环境 (JDK1.6 + maven3.1.1) 严重怀疑版本太低没有自动依赖相关jar，因为代码中确实没涉及到google的代码，搜了一下这个类所属的jar，导入依赖，完美解决。'. Below the text is a code block showing an XML dependency configuration.

```
<dependency>
```

```
<groupId>com. google. guava</groupId>
```

```
<artifactId>guava</artifactId>
```

```
<version>19.0</version>
```

```
</dependency>
```

首行报错先 `clear-->install`

如果 `RunAs--maven clear`

`RunAs--maven install` 时报错

following articles:
[ERROR] [Help 1]
<http://cwiki.apache.org/confluence/display/MAVEN/PluginResolutionException>

解决办法：
修改settings.xml文件中的节点，使用阿里的国内镜像，访问速度更快。

```
1 <mirror>
2   <id>nexus-aliyun</id>
3   <mirrorOf>central</mirrorOf>
4   <name>Nexus aliyun</name>
5   <url>http://maven.aliyun.com/nexus/content/repositories/central</url>
6 </mirror>
```

登录后复制

配置使用阿里镜像

```
<mirror>
```

```
  <id>nexus-aliyun</id>
```

```
  <mirrorOf>central</mirrorOf>
```

```
  <name>Nexus aliyun</name>
```

```
<url>http://maven.aliyun.com/nexus/content/repositories/central</url>
```

```
</mirror>
```

参考资料
及应用

《Hadoop 大数据分析》

主编：高永彬、钱亮宏、方志军

出版社：中国铁道出版社

作业	使用 maven 项目完成 HDFS 文件夹操作
教学反思	无

日期	2020年5月28日	周次	第十五周	时数	4课时
教学单元	Hadoop 编译源码				
教学目标	知识目标： 1. 了解 Hadoop 编译源码前期准备工作 2. 掌握 jar 包安装 3. 掌握 Hadoop 源码编译				
	技能目标： 能够进行 Hadoop 源码编译				
	素质目标： 52. 提高学生逻辑思维，创新能力； 53. 树立团队协作意识； 54. 树立良好的政治素质（爱党爱国、品行端正、遵纪守法）。				
教学重点	掌握 Hadoop 源码编译				
教学难点	1. 了解 Hadoop 编译源码前期准备工作 2. 掌握 jar 包安装				

教学方法	讲授演示法；任务驱动法
教学手段	云班课信息化教学；多媒体教学；钉钉直播教学
教学过程及内容	
<p>1.1 前期准备工作</p> <p>1. CentOS 联网</p> <p>配置 CentOS 能连接外网。Linux 虚拟机 ping www.baidu.com 是畅通的</p> <p>注意：采用 root 角色编译，减少文件夹权限出现问题</p> <p>2. jar 包准备 (hadoop 源码、JDK8、maven、ant 、protobuf)</p> <p>(1) hadoop-2.7.2-src.tar.gz</p> <p>(2) jdk-8u144-linux-x64.tar.gz</p> <p>(3) apache-ant-1.9.9-bin.tar.gz (build 工具，打包用的)</p> <p>(4) apache-maven-3.0.5-bin.tar.gz</p> <p>(5) protobuf-2.5.0.tar.gz (序列化的框架)</p> <p>1.2 jar 包安装</p> <p>注意：所有操作必须在 root 用户下完成</p> <p>JDK 解压、配置环境变量 JAVA_HOME 和 PATH，验证 java-version (如下都需要验证是否配置成功)</p>	

```
[root@hadoop101 software] # tar -zxvf
jdk-8u144-linux-x64.tar.gz -C /opt/module/
[root@hadoop101 software]# vi /etc/profile
#JAVA_HOME:
export JAVA_HOME=/opt/module/jdk1.8.0_144
export PATH=$PATH:$JAVA_HOME/bin
[root@hadoop101 software]#source /etc/profile
验证命令: java -version
Maven 解压、配置 MAVEN_HOME 和 PATH
[root@hadoop101 software]# tar -zxvf
apache-maven-3.0.5-bin.tar.gz -C /opt/module/
[root@hadoop101 apache-maven-3.0.5]# vi conf/settings.xml
<mirrors>
  <mirror>
    <id>nexus-aliyun</id>
    <mirrorOf>central</mirrorOf>
    <name>Nexus aliyun</name>
    <url>http://maven.aliyun.com/nexus/content/groups/public</u
rl>
  </mirror>
</mirrors>
```

```
[root@hadoop101 apache-maven-3.0.5]# vi /etc/profile
```

```
#MAVEN_HOME
```

```
export MAVEN_HOME=/opt/module/apache-maven-3.0.5
```

```
export PATH=$PATH:$MAVEN_HOME/bin
```

```
[root@hadoop101 software]#source /etc/profile
```

验证命令: `mvn -version`

ant 解压、配置 ANT_HOME 和 PATH

```
[root@hadoop101 software]# tar -zxvf
```

```
apache-ant-1.9.9-bin.tar.gz -C /opt/module/
```

```
[root@hadoop101 apache-ant-1.9.9]# vi /etc/profile
```

```
#ANT_HOME
```

```
export ANT_HOME=/opt/module/apache-ant-1.9.9
```

```
export PATH=$PATH:$ANT_HOME/bin
```

```
[root@hadoop101 software]#source /etc/profile
```

验证命令: `ant -version`

安装 `glibc-headers` 和 `g++` 命令如下

```
[root@hadoop101 apache-ant-1.9.9]# yum install glibc-headers
```

```
[root@hadoop101 apache-ant-1.9.9]# yum install gcc-c++
```

安装 `make` 和 `cmake`

```
[root@hadoop101 apache-ant-1.9.9]# yum install make
```

```
[root@hadoop101 apache-ant-1.9.9]# yum install cmake
```

解压 `protobuf` , 进入到解压后 `protobuf` 主目录,

/opt/module/protobuf-2.5.0, 然后相继执行命令

```
[root@hadoop101 software]# tar -zxvf protobuf-2.5.0.tar.gz -C /opt/module/
```

```
[root@hadoop101 hadoop-dist]# vi /etc/profile
```

```
#LD_LIBRARY_PATH
```

```
export LD_LIBRARY_PATH=/opt/module/protobuf-2.5.0
```

```
export PATH=$PATH:$LD_LIBRARY_PATH
```

```
[root@hadoop101 software]#source /etc/profile
```

验证命令: protoc --version

安装 openssl 库

```
[root@hadoop101 software]#yum install openssl-devel
```

安装 ncurses-devel 库

```
[root@hadoop101 software]#yum install ncurses-devel
```

到此, 编译工具安装基本完成。

1.3 编译源码

1. 解压源码到/opt/目录

```
[root@hadoop101 software]# tar -zxvf hadoop-2.7.2-src.tar.gz -C /opt/
```

2. 进入到 hadoop 源码主目录

```
[root@hadoop101 hadoop-2.7.2-src]# pwd
```

```
/opt/hadoop-2.7.2-src
```

3. 通过 maven 执行编译命令

```
[root@hadoop101 hadoop-2.7.2-src]#mvn package -Pdist,native
-DskipTests -Dtar
```

等待时间 30 分钟左右，最终成功是全部 SUCCESS，如图 2-42 所示。

```
[ INFO] Apache Hadoop Common ..... SUCCESS [3:35.094s]
[ INFO] Apache Hadoop NFS ..... SUCCESS [5.004s]
[ INFO] Apache Hadoop KMS ..... SUCCESS [54.027s]
[ INFO] Apache Hadoop Common Project ..... SUCCESS [0.022s]
[ INFO] Apache Hadoop HDFS ..... SUCCESS [3:58.444s]
[ INFO] Apache Hadoop HttpFS ..... SUCCESS [1:02.562s]
[ INFO] Apache Hadoop HDFS BookKeeper Journal ..... SUCCESS [33.138s]
[ INFO] Apache Hadoop HDFS-NFS ..... SUCCESS [3.993s]
[ INFO] Apache Hadoop HDFS Project ..... SUCCESS [0.022s]
[ INFO] hadoop-yarn ..... SUCCESS [0.037s]
[ INFO] hadoop-yarn-api ..... SUCCESS [1:26.119s]
[ INFO] hadoop-yarn-common ..... SUCCESS [1:20.025s]
[ INFO] hadoop-yarn-server ..... SUCCESS [0.168s]
[ INFO] hadoop-yarn-server-common ..... SUCCESS [9.107s]
[ INFO] hadoop-yarn-server-nodemanager ..... SUCCESS [19.867s]
[ INFO] hadoop-yarn-server-web-proxy ..... SUCCESS [3.397s]
[ INFO] hadoop-yarn-server-applicationhistoryservice ..... SUCCESS [7.432s]
[ INFO] hadoop-yarn-server-resourcemanager ..... SUCCESS [17.078s]
[ INFO] hadoop-yarn-server-tests ..... SUCCESS [3.998s]
[ INFO] hadoop-yarn-client ..... SUCCESS [5.962s]
[ INFO] hadoop-yarn-server-sharedcachemanager ..... SUCCESS [2.803s]
[ INFO] hadoop-yarn-applications ..... SUCCESS [0.024s]
[ INFO] hadoop-yarn-applications-distributedshell ..... SUCCESS [1.841s]
[ INFO] hadoop-yarn-applications-unmanaged-am-launcher .... SUCCESS [1.876s]
```

图 2-42 编译源码

4. 成功的 64 位 hadoop 包在

/opt/hadoop-2.7.2-src/hadoop-dist/target 下

```
[root@hadoop101 target]# pwd
```

/opt/hadoop-2.7.2-src/hadoop-dist/target

5. 编译源码过程中常见的问题及解决方案

(1) MAVEN install 时候 JVM 内存溢出

处理方式：在环境配置文件和 maven 的执行文件均可调整 MAVEN_OPT 的 heap 大小。（详情查阅 MAVEN 编译 JVM 调优问题，如：

<http://outofmemory.cn/code-snippet/12652/maven-outofmemoryerror-method>）

(2) 编译期间 maven 报错。可能网络阻塞问题导致依赖库下载不完

整导致，多次执行命令（一次通过比较难）：

```
[root@hadoop101 hadoop-2.7.2-src]#mvn package -Pdist,nativeN  
-DskipTests -Dtar
```

（3）报 ant、protobuf 等错误，插件下载不完整或者插件版本问题，
最开始链接有较多特殊情况，同时推荐

2.7.0 版本的问题汇总帖

子 <http://www.tuicool.com/articles/IBn63qf>

参考资料 及应用	《Hadoop 大数据分析》 主编：高永彬、钱亮宏、方志军 出版社：中国铁道出版社
作业	复习
教学反思	无

日期	2020年6月3日	周次	第十六周	时数	4课时
教学单元	常见错误及解决方案				
教学目标	知识目标： 1. 防火墙没关闭 2. 主机名称配置错误 3. IP地址配置错误 4. ssh没有配置好				
	技能目标： 能够解决集群过程中常见错误				
	素质目标： 55. 提高学生逻辑思维，创新能力； 56. 树立团队协作意识； 57. 树立良好的政治素质（爱党爱国、品行端正、遵纪守法）。				
教学重点	能够解决集群过程中常见错误				
教学难点	1. 主机名称配置错误 2. IP地址配置错误 3. ssh没有配置好				

教学方法	讲授演示法；任务驱动法
教学手段	云班课信息化教学；多媒体教学；钉钉直播教学
教学过程及内容	
<p>常见错误及解决方案</p> <ol style="list-style-type: none"> 1) 防火墙没关闭、或者没有启动 YARN <pre>INFO client.RMProxy: Connecting to ResourceManager at hadoop108/192.168.10.108:8032</pre> <ol style="list-style-type: none"> 2) 主机名称配置错误 3) IP 地址配置错误 4) ssh 没有配置好 5) root 用户和 sun 两个用户启动集群不统一 6) 配置文件修改不细心 7) 未编译源码 <pre>Unable to load native-hadoop library for your platform... using builtin-java classes where applicable</pre> <pre>17/05/22 15:38:58 INFO client.RMProxy: Connecting to ResourceManager at hadoop108/192.168.10.108:8032</pre> <ol style="list-style-type: none"> 8) 不识别主机名称 	


```
java.net.UnknownHostException: hadoop102: hadoop102
    at
java.net.InetAddress.getLocalHost(InetAddress.java:1475)
    at
org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(
JobSubmitter.java:146)
    at
org.apache.hadoop.mapreduce.Job$10.run(Job.java:1290)
    at
org.apache.hadoop.mapreduce.Job$10.run(Job.java:1287)
    at
java.security.AccessController.doPrivileged(Native Method)
at javax.security.auth.Subject.doAs(Subject.java:415)
```

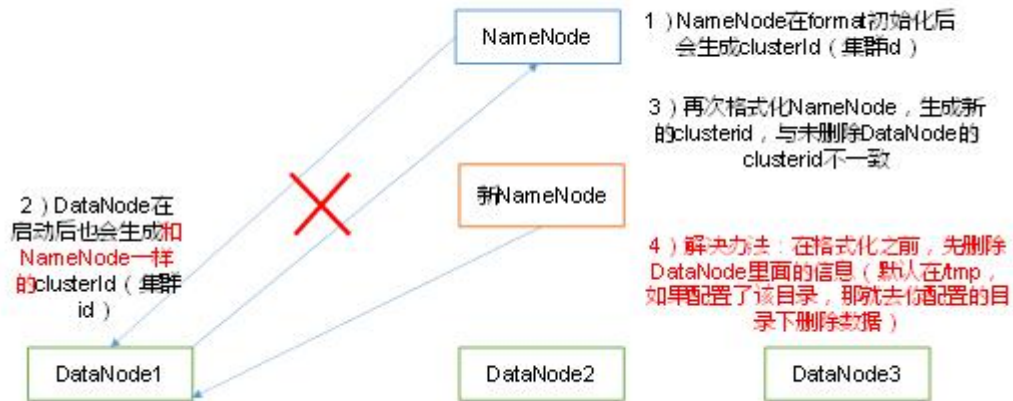
解决办法:

(1) 在/etc/hosts 文件中添加 192.168.1.102 hadoop102

(2) 主机名称不要起 hadoop hadoop000 等特殊名称

9) DataNode 和 NameNode 进程同时只能工作一个。

DataNode和NameNode进程同时只能有一个工作问题分析



10) 执行命令不生效, 粘贴 word 中命令时, 遇到-和长 - 没区分开。

导致命令失效

解决办法: 尽量不要粘贴 word 中代码。

11) jps 发现进程已经没有, 但是重新启动集群, 提示进程已经开启。

原因是在 linux 的根目录下/tmp 目录中存在启动的进程临时文件,

将集群相关进程删除掉, 再重新启动集群。

12) jps 不生效。

原因: 全局变量 `hadoop java` 没有生效。解决办法: 需要 `source /etc/profile` 文件。

13) 8088 端口连接不上

```
[sun@hadoop102 桌面]$ cat /etc/hosts
```

注释掉如下代码

```
#127.0.0.1    localhost localhost.localdomain localhost4
localhost4.localhost4

#::1        hadoop102
```

参考资料 及应用	《Hadoop 大数据分析》 主编：高永彬、钱亮宏、方志军 出版社：中国铁道出版社
作业	复习
教学反思	无