Education and Learning, 教育与学习(3)2021, 9
ISSN:2705-0408(P); 2705-0416(0)

“分布式计算与开发模式”实验教学环境构建研究

刘 扬 杨 丹
（辽宁科技大学 辽宁鞍山 114051）

【摘 要】随着大数据技术的发展，分布式计算已经越来越受到企业界、学术界的关注。针对计算机专业本科生开设的专业课“分布式计算与开发模式”的特点和大纲要求，结合对学生动手能力与创新能力培养的宗旨，研究该门课程的实验教学环境构建。在综合考虑各种可行性因素下，提出在Windows系统下安装Hadoop分布式环境，并安装java编程语言的快速开发平台Eclipse以实现Hadoop+Eclipse的分布式计算实验开发环境的解决方案。通过搭建实验环境来更好地促进和提高本门课程的理论教学效果，以达到培养创新型人才的目标。

【关键词】分布式计算；大数据；实验教学；Hadoop；Eclipse
DOI：10.18686/jyyxx.v3i9.55459

近几年来，云计算、物联网、大数据等技术迅速发展[1-4]，针对海量信息的分析以及处理，谷歌公司相继提出了分布式文件系统 GFS[5]，大数据存储结构 Big Table[6]，以及计算模型 Map Reduce[7]。从那时起，分布式系统进入了一个崭新的阶段，活跃在相关开源社区的开发人员和兴趣爱好者，实现了 Hadoop[8-10]这一开创性的、高效的、开源分布式计算架构。Hadoop 具有独特的底层设计模式、较好的容错性等长处，使得它能够稳定和精确地处理和分析大数据。Hadoop 主要用于大规模数据的存储和离线计算，目前被国内外很多大型网站采用[11-12]。

“分布式计算与开发模式”课程是计算机专业开设的一门本科专业课程，主要讲授分布式计算、大数据[13-15]的一些相关技术。对于这种理论性较强、比较枯燥的课程，理论授课内容不容易被理解和掌握，必须由实践教学来帮助学生进一步消化和吸收。由于以往的实验环境有限，教师只能讲授理论知识，不能进行具体的实验演示，学生容易对分布式的概念及技术产生枯燥、遥不可及的感觉，教学效果不理想。尽管目前大数据、云计算等在各行各业被广泛应用，综合考虑搭建成本、实验室已有的系统环境，学生的先修课程（编程语言 java）与学生的接受能力等各种可行性因素，采用在 Windows 下搭建 Hadoop+Eclipse 作为此门课程的实验教学环节，达到提高教学效果的目的。

1 Windows系统下的Hadoop环境搭建

1.1 JDK安装与系统环境变量设置

该部分主要包括以下步骤：①下载 jdk1.8.0_121，存放在某个目录下，如目录 F:\apachehadoop\Java\jdk1.8.0_121；②在计算机—属性—高级系统设置—高级选项卡—环境变量—系统变量—新建 JAVA_HOME，添加路径 F:\apachehadoop\Java\jdk1.8.0_121；③在系统变量 path 中添加%JAVA_HOME%%JAVA_HOME%\bin。

1.2 Hadoop安装与系统环境变量设置

该部分主要包括以下步骤：①首先下载 Hadoop2.8.5，然后解压到系统目录，如 D:\hadoop\hadoop-2.8.5；②在在计算机—属性—高级系统设置—高级选项卡—环境变量—系统变量—单击新建 HADOOP_HOME，添加路径 D:\hadoop\hadoop-2.8.5；③在系统变量 path 中添加 %HADOOP_HOME%\bin。

1.3 Hadoop组件下载

下载与 32 位或 64 位 Window 系统一致的组件 hadoop.dll，winutils.exe，然后覆盖到 D:\hadoop\hadoop-2.8.5\bin 目录下。若缺失这两个文件，在 Windows 系统下测试时会报错。

1.4 查看Hadoop版本信息

打开 cmd 窗口，切换到 Hadoop 下的 bin 目录，显示 Hadoop 的版本信息。

1.5 namenode及datanode缺省目录创建

在目录 D:\hadoop\hadoop-2.8.5\data 下面创建 namenode 目录以及 datanode 目录，即 D:\hadoop\hadoop-2.8.5\data\namenode 以及 D:\hadoop\hadoop-2.8.5\data\datanode，用来保存用户运行数据。

1.6 Hadoop文件配置

在目录 D:\hadoop\hadoop-2.8.5\etc\hadoop 下，主要对 core-site.xml/hdfs-site.xml/mapred-site.xml/yarn-site.xml 这四个文件进行配置（由于篇幅所限，配置文件关键代码如下）。

core-site.xml

```
<property>
<name>hadoop.tmp.dir</name>
<value>/var/log/hadoop/tmp</value>
</property>
<property>
<name>fs.defaultFS</name>
<value>hdfs://localhost:9000</value>
</property>
```

hdfs-site.xml

```
<property>
<name>dfs.replication</name>
<value>1</value>
</property>
<property>
<name>dfs.namenode.name.dir</name>
<value>/d:/hadoop/hadoop-2.8.5/data/namenode</value
```

```
>
    </property>
    <property>
    <name>dfs.dataenode.data.dir</name>
    <value>/d:/hadoop/hadoop-2.8.5/data/datanode</value>
    </property>
    mapred-site.xml（mapred-site.xml.template）
    <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
    </property>
    <property>
    <name>mapred.job.tracker</name>
    <value>localhost:9001</value>
    </property>
    <property>
    <name>dfs.permissions</name>
    <value>false</value>
    </property>
    yarn-site.xml
    <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
    </property>
```

2 Hadoop环境运行测试

2.1 格式化系统文件

在 bin 目录下，执行 hdfs namenode -format（不要重复执行）命令。

2.2 Hadoop启动

在 Hadoop 的 sbin 目录下执行 start-all.cmd 命令。

2.3 进程查看

在 bin 目录下执行 jps 命令，可看到启动的进程。

2.4 网页查看

分别访问 http://localhost:50070 网页，以及 http://localhost:8088 网页，看是否安装成功。以 Hadoop 自带的经典计算单词数量的任务 wordCount 为测试程序。具体步骤如下：

①任意位置创建一个文本文件如 D:\hadoop\hadoop_testone.txt；②用 dfs 命令创建一个文件夹，名字任意；③把 hadoop_testone.txt 上传到 dfs 目录下，并查看；④然后运行 Hadoop 提供的 demo，计算单词数；⑤运行结果在 dfs 目录中的/test/output 文件夹下，可用命令查看；⑥执行 stop-all.cmd 命令，停止 Hadoop。

3 Windows系统下的Eclipse安装与Hadoop插件配置

3.1 Eclipse安装

首先下载并解压 eclipse4.4.1 存放在 F:\eclipse4.4.1 目录下，然后安装 eclipse4.4.1。

3.2 Hadoop插件下载

下载 Eclipse 平台的 Hadoop 插件 hadoop2x-eclipse-plugin-master 并解压，将其复制到 eclipse 下的 plugins 目录下以实现 Eclipse 与 Hadoop 的关联。

3.3 Eclipse配置

主要配置步骤如下：①启动 Eclipse，在 Window->show view，打开 MapReduce tools；②选择本地 Hadoop 目录；③配置 Hadoop location；④配置后出现 DFS Locations 表示 Eclipse 与 Hadoop 关联成功。

4 结语

针对计算机专业课“分布式计算与开发模式”实验教学的特点与学生培养目标，综合考虑各种可行性因素，通过在 Windows 系统下搭建 Hadoop+Eclipse 教学实验环境。此分布式实验环境具有较好兼容性、可视化等特点，使学生可以从实践教学中加深分布式理论知识的理解和掌握，达到提高教学效果的目的。

作者简介：刘扬（1973.2—），副教授，研究方向：计算机科学与技术。

【参考文献】

[1] 贾俊婷.云计算综述及云计算在通信行业的应用[D].呼和浩特：内蒙古大学，2014.
[2] 方筠捷.物联网发展现状、趋势分析及中国的应对措施[J].江苏科技信息，2018，35（16）：61-63+80.
[3] 凌霄.国内外大数据教育应用研究的对比研究[D].深圳：深圳大学，2017.
[4] 刘钺.浅谈大数据时代背景下的数据分析[J].电脑知识与技术，2017，13（20）：5-7
[5] 付东华.基于 HDFS 的海量分布式文件系统的研究与优化[D].北京：北京邮电大学，2012.
[6] 王猛.大数据分析仓库 Hive 存储结构扩展的设计和实现[D].上海：上海交通大学，2015.
[7] 武鑫.基于 MapReduce 的协同过滤算法并行化研究[D].天津：河北工业大学，2014.
[8] 秦杰仪，曾志，孙蕾，等.基于 Hadoop 的大数据平台架设探讨[J].现代工业经济和信息化，2018，8（5）：47-49.
[9] 辛大欣，刘飞.Hadoop 集群性能优化技术研究[J].电脑知识与技术，2011，7（22）：5484-5486.
[10] 张功水.基于 Hadoop 技术的电信大数据分析平台的设计和实现[J].信息通信，2016（10）：114-115.
[11] 唐雪.基于 Hadoop 的电影推荐系统的研究与实现[D].重庆：重庆理工大学，2016.
[12] 盘隆.基于 MapReduce 的分布式编程框架的设计与实现[D]哈尔滨：哈尔滨工业大学，2011.
[13] 何志爽，万亚平，向霞.浅析大数据云计算技术及其应用[J].科学中国人，2016（23）：35.
[14] 徐立冰.腾云：云计算和大数据时代网络技术揭秘[M].北京：人民邮电出版社，2013.
[15] 梁芷梧.云计算中 MapReduce 分布式并行处理框架的研究[D].武汉：湖北工业大学，2017.