

大数据技术实验报告册

原创

848698119 于 2021-10-30 23:19:49 发布 3238 收藏 2

分类专栏: [大数据技术](#) 文章标签: [big data](#) [hdfs](#) [hadoop](#)

版权声明: 本文为博主原创文章, 遵循 [CC 4.0 BY-SA](#) 版权协议, 转载请附上原文出处链接和本声明。

本文链接: https://blog.csdn.net/qq_45896330/article/details/121058156

版权



[大数据技术](#) 专栏收录该内容

3 篇文章 0 订阅

订阅专栏

实验报告册, 拿来吧~

大数据技术实验报告册

实验一: HDFS读写文件

实验二: 应用MapReduce排序数据

实验三: 应用Hive统计学生人数

实验四: 应用Zookeeper实现进程协作

开心地微笑



CSDN @848698119

大数据技术实验报告册

大数据实验时间:

第一次实验	HDFS读写文件	9月24日
第二次实验	应用MapReduce排序数据	10月8日
第三次实验	应用Hive统计学生人数	10月15日
第四次实验	应用zookeeper实现进程协作	10月22日

实验一：HDFS读写文件

实验名称	HDFS 读写文件			成绩	
实验类型	验证性	实验学时	2	日期	9月24日
实验目的	1. 会在Linux环境下编写读写HDFS文件的代码; 2. 会使用Jar命令打包代码; 3. 会在master服务器上运行HDFS读写程序 3. 会在Windows上安装Eclipse Hadoop插件; 5. 会在Eclipse环境编写HDFS文件代码, 6. 会使用Eclipse打包代码				
实验设备	PC				

一、实验原理(要求、任务等)

实验要求: 实验结束时, 每位学生均已搭建HDFS开发环境; 编写HDFS写, 读代码; 在master机上执行了读写, 读程序。通过实验了解HDFS读写文件的调用流程, 理解HDFS读写文件的原理

实验原理:

1. Java Classpath。Classpath设置的目的, 在于告诉Java执行环境, 在哪些目录下可以找到你所要执行的Java程序所需要的类或者包。

2. Eclipse Hadoop 插件

Eclipse是一个跨平台的自由集成开发环境(IDE)。通过安装不同的插件Eclipse可以支持不同的计算机语言, 比如C++和Python等开发工具, 亦可以通过hadoop插件来扩展开发Hadoop相关程序。

二、实验内容和步骤

1. 配置master服务器classpath.

2. 在master服务器编写HDFS写程序

3. 编译并打包HDFS写程序

4. 执行HDFS写程序

5. 在master服务器编写HDFS读程序

6. 编译并打包HDFS读程序

7. 执行HDFS读程序

8. 安装与配置Eclipse Hadoop 插件

9. 使用Eclipse开发并打包HDFS写文件程序

10. 上传HDFS写文件程序Jar包并执行

11. 使用Eclipse开发并打包HDFS读文件程序

12. 上传HDFS读文件程序Jar包并执行

实验名称	应用 MapReduce 排序数据			成绩	
实验类型	验证性	实验学时	2	日期	10月8日
实验目的	基于 MapReduce 思想，编写 Sort 程序				
实验设备	PC				
一、实验原理(要求、任务等)					
<p>要求：更能理解 MapReduce 编程思想，会编写 MapReduce 版本排序程序，然后将其执行分析执行过程。</p> <p>原理：MR 默认会对键进行排序，然后有的时候我们也有对值进行排序的需求。满足这种需求一是可以在 reduce 阶段排序收集过来的 Values，但是数量巨大的 Values 可能会导致内存溢出等问题。</p>					
二、实验内容和步骤					
<ol style="list-style-type: none"> 1. 启动 Hadoop 2. 验证 HDFS 上的 sort1.txt, sort2.txt, sort3.txt 是否存在。 3. 上传数据文件到 HDFS 的 /input/sort 文件夹下。 4. 编写 MapReduce 程序。 5. 使用 Eclipse 开发工具将其代码打包 6. 上传 HDFS sort 程序 jar 包并执行 					

实验名称	应用 Hive 统计学生人数			成绩	
实验类型	验证性	实验学时	2	日期	10月 15日
实验目的	1. 学会创建 Hive 表 2. 学会使用 HQL 实现对表的创建, 修改, 删除, 显示等。				
实验设备	PC				

一、实验原理(要求、任务等)

使用 Hive Shell 完成以下操作:

- (1) 进入 Hive 命令行接口
- (2) 在 Hive 默认数据库 default 下新建 student 表, 并将下表中的数据载入 Hive 里的 student 表中。
- (3) 编写 HiveQL SELECT 语句, 完成以下查询: 查询 student 表中所有记录, 查询 student 表中所有女生记录, 统计 student 中男女生人数。

二、实验内容和步骤

- (1) 使用命令 "hive" 进入 Hive 命令行。
- (2) 使用 "create table" 命令在 Hive 默认数据库中创建表 student
- (3) 首先, 查询 member 表中所有记录
- (4) 其次, 查询 student 中所有属性记录
- (5) 最后, 统计 student 中男女生人数。
- (6) 使用命令 "quit" 退出 Hive CLI

实验名称	应用 Zookeeper 实现进程协作			成绩	
实验类型	验证性	实验学时	2	日期	10月22日
实验目的	1. 掌握 Java 代码如何连接 Zookeeper 集群及通过代码读写 Zookeeper 集群的目录下的数据。 2. 掌握 Zookeeper 如何实现多个线程间的协作。				
实验设备	PC				

一、实验原理(要求、任务等)

要求: 用 Java 代码实现两个线程, 一个向 Zookeeper 中某一目录中写入数据, 另一个线程监听此目录, 若目录下数据有更新则将目录中数据读取并显示出来。

原理: 通过 Zookeeper 实现不同物理机器上的进程间通信。

场景使用: 客户端 A 需要向客户端 B 发送一条消息 msg1。

实现方法: 客户端 A 把 msg1 发送给 Zookeeper 集群, 然后由客户端 B 自行去 Zookeeper 集群读取 msg1。

二、实验内容和步骤

本实验主要完成多线程通过 Zookeeper 完成彼此间的协作问题, 实验过程包含启动集群, 编写代码, 客户端提交代码三个步骤

步骤: 1. 启动集群

2. 导入 jar 包

3. 编写 Java 代码

4. 做成 jar 包, 将上述代码打成 jar 包, 用工具上传到客户端节点, 并执行代码。

实验心得：通过本次实验，掌握了，，，，理解了，，，，明白了，，，，，通过实验让我们对于理论的学习产生了浓厚的兴趣，为后边的学习打下了坚实的基础。



