

实验 3 MapReduce 编程实践

一. 实验目的

- 1) 掌握 MapReduce 基本编程方法
- 2) 掌握用 MapReduce 解决一些常见的数据处理问题

二. 实验平台

操作系统: Linux

JDK 版本: 1.7 或以上

Hadoop 版本: 建议 Hadoop 2.7.x

Java IDE: 建议 Eclipse

三. 实验内容和要求

(一) 编程实现文件合并和去重操作

编写 MapReduce 程序, 实现对多个文件的合并, 并剔除其中重复的内容, 去重后的内容输出到一个新的文件。

(二) 编程实现倒排索引

编写 MapReduce 程序, 实现对多个输入文件的内容建立倒排索引, 输出单词到文档的映射关系及单词在该文档中的出现次数。

四. 实验报告

将实验过程、代码、执行结果及实验中发现的问题列出, 包含无法解决的问题。