

## 实验4 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) 编写 MapReduce 程序, 实现对输入文件中的 child-parent 关系进行挖掘, 输出 grandchild-grandparent 的对应关系表。

Input:

Output:

Steven	Jack	Jone	Jesse
Jone	Lucy	Jone	Alice
Jone	Jack	Jone	Frank
Lucy	Mary	Jone	Mary
Lucy	Frank	Mark	Alice
Jack	Alice	Mark	Jesse
Jack	Jesse	Philip	Alice
David	Alice	Philip	Jesse
David	Jesse	Steven	Jesse
Philip	David	Steven	Alice
Philip	Alma		
Mark	David		

### 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。