

2023 春分布式计算课程实验

一 实验目的

学习基于 MapReduce 框架的分布式计算程序设计方法。

二 实验题目

题目 1

输入文件为学生成绩信息，包含了必修课与选修课成绩，格式如下：

班级 1, 姓名 1, 科目 1, 必修, 成绩 1
 （注：
 为换行符）

班级 2, 姓名 2, 科目 1, 必修, 成绩 2

班级 1, 姓名 1, 科目 2, 选修, 成绩 3

.....,,,

编写两个 Hadoop 平台上的 MapReduce 程序，分别实现如下功能：

1. 计算每个学生必修课的平均成绩。
2. 按科目统计每个班的平均成绩。

题目 2

输入文件的每一行为具有父子/父女/母子/母女/关系的一对人名，例如：

Tim, Andy

Harry, Alice

Mark, Louis

Andy, Joseph

.....,

假定不会出现重名现象。

编写 Hadoop 平台上的 MapReduce 程序，找出所有具有 grandchild-grandparent 关系的人名组。

上面的实验在验收时必须用老师给定的实验测试数据。

三 实验报告提交

1. 实验结果验收通过之后，撰写实验报告。
2. 6 月 20 日前把源程序工程和实验报告打包，通过西电智课平台提交