2021春分布式计算课程实验

# 一 实验目的

1. 学习基于MapReduce框架的分布式计算程序设计方法。
2. 学习基于Spark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 <br> （注：<br> 为换行符）

班级2, 姓名2, 科目1, 必修, 成绩2 <br>

班级1, 姓名1, 科目2, 选修，成绩3 <br>

………., ………, ………, ……… <br>

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 <br>

Harry, Alice <br>

Mark, Louis <br>

Andy, Joseph <br>

……….., ………… <br>

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。

## 题目3

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 <br> （注：<br> 为换行符）

班级2, 姓名2, 科目1, 必修, 成绩2 <br>

班级1, 姓名1, 科目2, 选修，成绩3 <br>

………., ………, ………, ……… <br>

编写一个Spark程序，同时实现如下功能：

1. 计算每个学生必修课的平均成绩。

2. 统计学生必修课平均成绩在：90~100,80~89,70~79,60~69和60分以下这5个分数段的人数。

**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. 7月1日前把源程序工程和实验报告打包，发送至邮箱：xddistcomcourse@163.com
3. 邮件标题风格：实验报告+学号+姓名
4. 附件文件命名方式：实验报告+学号+姓名