

# 分布式系统第六次实验思路说明

---

西安电子科技大学 17030140014 张笑天

---

[代码地址](#)

## 基础设计

本程序参考MapReduce中WordCount程序，完成分布式作业。

## 问题与解决

### 1. JAR打包问题

使用IDEA的功能打包，利用MAVEN配置文件获取依赖，从而实现可以在Windows IDE下生成JAR包。

### 2. 主键问题

对于按照科目统计每个班成绩，使用科目和班合并作为主键。

### 3. yarn命令使用

注意使用yarn命令的格式，应当指明jar包中main的完整路径。

以下是实验使用的命令：

```
# 每个学生必修课平均成绩
hadoop fs -put ~/Documents/StudentScore.txt /StudentIn

yarn jar ~/Documents/ScoreCalculation.jar
tech.enigma.hadoop.scorecalculation.ScoreCalculation /StudentIn /StudentOut

hadoop fs -copyToLocal /StudentOut ~/Documents/

# 按科目统计每个班平均成绩

hadoop fs -rm -r /StudentOut

yarn jar ~/Documents/ScoreCalculation.jar
tech.enigma.hadoop.scorecalculation.ClassCalculation /StudentIn /StudentOut

hadoop fs -copyToLocal /StudentOut ~/Documents/

# 寻找孙辈-祖辈关系

hadoop fs -put ~/Documents/ChildParent.txt /CPIn

yarn jar ~/Documents/ChildParent.jar
tech.enigma.hadoop.childparent.ChildParent /CPIn /CPOut
```

```
hadoop fs -copyToLocal /CPOut ~/Documents/
```

## 实验结果展示

- 每个学生必修课平均成绩 [Exp\\_1\\_1.txt](#)
- 按科目统计每个班平均成绩 [Exp\\_1\\_2.txt](#)
- 寻找孙辈-祖辈关系 [Exp\\_2.txt](#)