

大数据平台核心技术课程设计

目标

本课程设计的目标是实现一个简单的单机可运行的MapReduce运行时库。这个运行时库至少可以支撑给定的WordCount程序运行。其中输入和输出都是csv（comma-seperated-values）格式。

可以使用第三方库，但是不能使用第三方分布式系统实现，包括Hadoop、Spark以及其他支持MapReduce编程模型的framework。

MapReduce运行时库应实现给定的sdk，sdk对应的jar包和文档可以在我们给定的附件中找到。

整个课程设计可以自由组合4人以下的小团队完成，团队中每个人的评分相同，等于团队获得的评分。

提交

我们给出一个Java语言的空项目作为起点，使用maven管理构建。

这个空项目包含一个main class：WordCount，运行的命令为：

```
java com.aliyun.odps.mapred.WordCount src.csv dest.csv
```

其中src.csv和dest.csv分别是输入和输出的表名，由一个当前目录下的csv文件表示。

例如，当前目录下有一个表src.csv，写出的表期望是dest.csv，则上述运行命令有效。

提交代码的方式是将整个项目打成zip包上传到MOOC系统中，确保它可以用 `maven clean install -DskipTests` 编译。

验证和评估

MOOC系统会运行测试集保证正确性和性能达到标准。编译失败、无法运行和结果错误都被认为不达标。结果正确的前提下按照（且仅按照）**运行速度**评分

附件列表

1. sdk-javadoc.tar.gz: sdk javadoc
2. project.zip: 空项目
3. src.csv和dest.csv: 一个WordCount的范例输入和输出。