

电影网站用户影评分析实训

实训1电影网站用户影评分析实训

实训目的

1. 掌握根据业务场景设计map()方法和reduce()方法的计算逻辑
2. 掌握编写MapReduce程序解决常见的数据处理问题
3. 掌握编写MapReduce程序实现电影网站用户影评分析的方法

训练要点

掌握MapReduce实现数据预处理

掌握MapReduce实现不同维度分析和统计

需求说明

通过对电影网站用户及电影评论数据进行分析，结合MapReduce编程知识，分别从评价次数、性别、年龄段、电影类型这4个维度分析用户的观影喜好，具体的统计分析需求如下。

1. 评价次数：计算评价次数最多的10部电影及评分次数。
2. 性别：计算不同性别评分最高的10部电影及评分。
3. 年龄段：计算某给定电影各年龄段的平均电影评分。
4. 电影类型：计算影评库中各种类型电影中评价最高的5部电影。

实现思路及步骤

1. 搭建Hadoop环境，启动Hadoop集群。
2. 搭建IDEA开发环境。
3. 分别编写计算评价次数最多的10部电影及评分次数的MR程序，计算不同性别评分最高的10部电影及评分MR程序，计算某给定电影各年龄段的平均电影评分MR程序，计算影评库中各种类型电影中评价最高的5部电影MR程序。
4. 调试MR程序，观察MR程序运行过程和运行结果是否正常。
5. 截图各MR程序的源码、截图IDEA上控制台的运行结果、截图HDFS生成结果文件内容。

作业要求

1. 截图各MR程序的源码，至少9个文件截图，将配置项全部放在FinalUtil类中并截图；
2. 截图IDEA上控制台的各MR程序的运行结果，至少9个文件截图；
3. 截图HDFS各MR程序生成结果文件内容，至少9个文件内容截图，截大图，图中要含有本人姓名的地址栏url；

实训参考

参考：《[第8章电影网站用户影评分析](#)》

