

《分布式编程模型与系统》期末考查作业

学号: 10195501415

姓名: 赵煜硕

成绩:

1. 实验目的
比较 Spark 和 MapReduce 执行迭代应用的性能差异
2. 实验设置
JDK1.8
IDEA
Windows
3. 实验过程
<https://github.com/zhaoyushuo123/pagerank>
使用课本上给的 mapreduce 版本的 pagerank 代码

部分片段参考

<https://wenku.baidu.com/view/8a08e207ac45b307e87101f69e3143323968f5d6.html>

配置 input 目录和 output 目录

遇到

log4j:WARN No appenders could be found for logger (org.apache.zookeeper.ZooKeeper).

log4j:WARN Please initialize the log4j system properly.

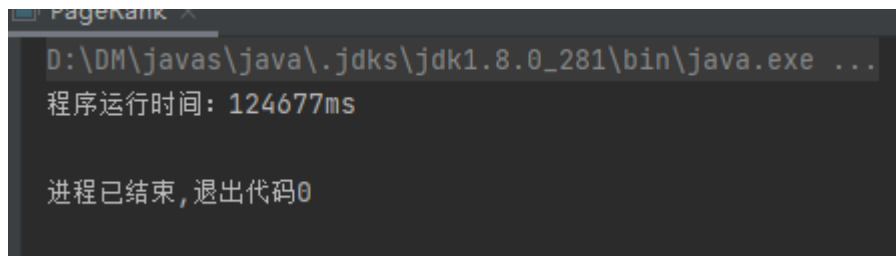
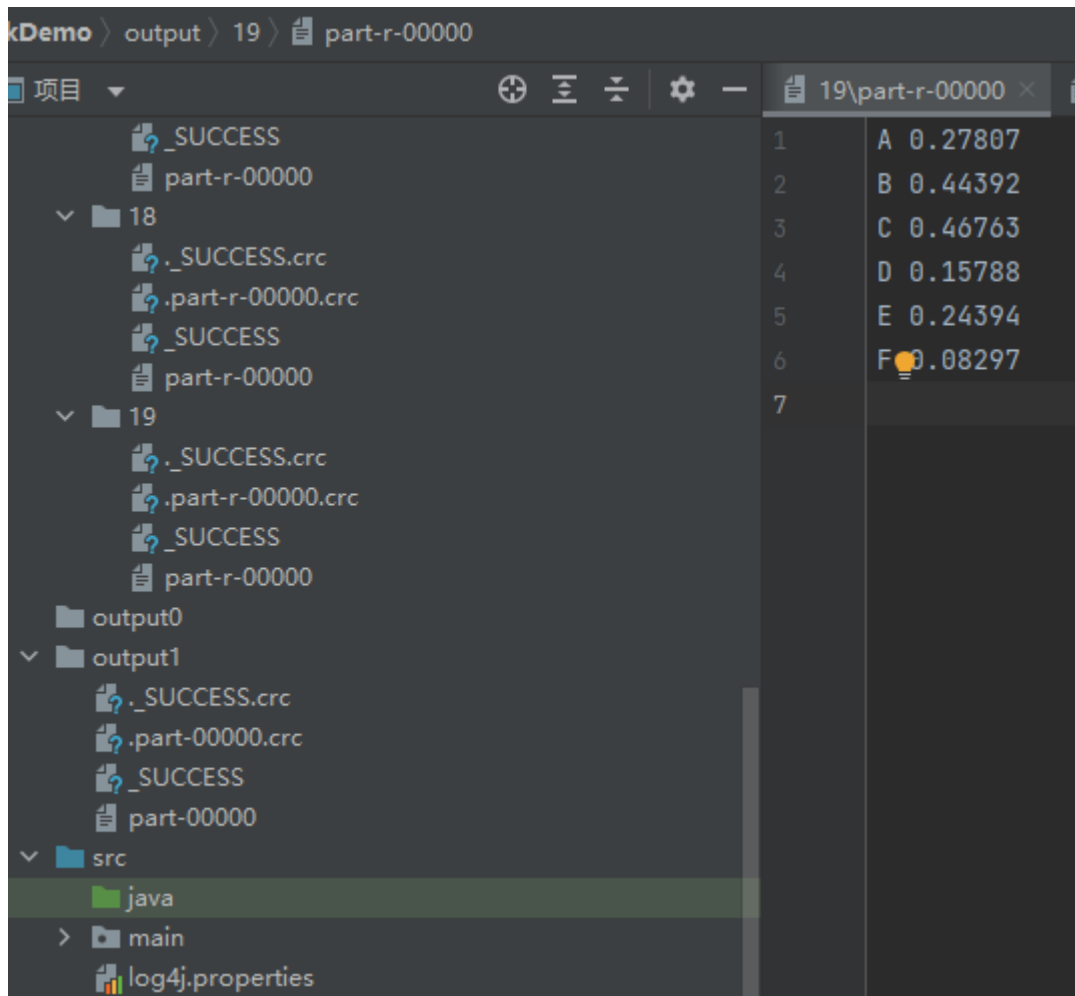
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.

参考

https://blog.csdn.net/zhengzaifeidelushang/article/details/116239936?spm=1001.2101.3001.6661.1&utm_medium=distribute.pc_relevant_t0.none-task-blog-2%7Edefault%7ECTRLIST%7EPayColumn-1-116239936-blog-123576662.pc_relevant_default&depth_1-utm_source=distribute.pc_relevant_t0.none-task-blog-2%7Edefault%7ECTRLIST%7EPayColumn-1-116239936-blog-123576662.pc_relevant_default&utm_relevant_index=1

<https://blog.csdn.net/qg125864004/article/details/109294338>

执行 MapReduce 迭代,发现 output 中会给出每轮迭代的结果



在 spark 上会快一点

