

编程作业 2: Java 文件 IO 操作及 HDFS 编程 (共计 10 分 : 3+3+3+1)

1. 数据生成 : 从 1 至 2014*512 之间的每个整数, 重复生成 256 次, 例如整数 1, 则生成 256 个 1, 依次类推, 将所有整数依次写入单机环境的本地文件系统之中。要求如下 :
 - a) 编写一个 Java 单线程程序, 作为上述数据生成 (i.e., Producer.java) 代码
 - b) 编写一个 Java 多线程程序, 读取 Producer.java 的生成数据, 将该数据写入文件 (i.e., MTWriter.java)
 - c) 变化 MTWriter.java 的线程数量从 1、2、4、8、16、32, 尝试至少 3 个不同的文件 IO 写入方法或者策略, 计算其从开始生产第一个整数到完成所有数据写入之间的时间差, 绘制相应的(1)时间曲线图 : x 轴为线程数、y 轴为时间 (单位 : 秒)、曲线为不同文件 IO 写入方法; (2)文件空间大小的直方图 x 轴为不同文件 IO 写入方法 ; y 轴为空间大小 (单位 : MB)
2. 针对上述任务, 将这些整数依次写入 HDFS 文件系统之中
 - a) 利用上述 Producer.java 代码生成数据
 - b) 编写一个 Java 多线程程序, 读取 Producer.java 的生成数据, 将该数据写入 HDFS (i.e., HDWriter.java)
 - c) 变化 HDWriter.java 的线程数量从 1、2、4、8、16、32, 尝试至少 2 个不同的 HDFS 写入方法或者策略, 计算其从开始生产第一个整数到完成所有数据写入之间的时间差, 绘制相应的(1)时间曲线图 : x 轴为线程数、y 轴为时间 (单位 : 秒)、曲线为不同文件 IO 写入方法; (2)文件空间大小的直方图 : x 轴为不同文件 IO 写入方法 ; y 轴为空间大小 (单位 : MB)
3. 针对上述任务, 将这些整数依次写入 Apache Cassandra KVS 系统之中
 - a) 利用上述 Producer.java 代码生成数据
 - b) 编写一个 Java 多线程程序, 读取 Producer.java 的生成数据, 将该数据写入 Cassandra (i.e., CAWriter.java)
 - c) 变化 CAWriter.java 的线程数量从 1、2、4、8、16、32, 尝试至少 2 个不同的 HDFS 写入方法或者策略, 计算其从开始生产第一个整数到完成所有数据写入之间的时间差, 绘制相应的(1)时间曲线图 : x 轴为线程数、y 轴为时间 (单位 : 秒)、曲线为不同文件 IO 写入方法; (2)文件空间大小的直方图 : x 轴为不同文件 IO 写入方法 ; y 轴为空间大小 (单位 : MB)
4. 完成一个 word 文件的作业报告, 包括上述 3 个小题目的对应图表, 并加以讨论和分析。

作业提交截止日期: 2012/12/08 日 23:59:59, 晚交者计零分。

作业提交内容: 学号-hw2.zip, 文件命名格式和文件输出内容格式不对着, 均计零分。

在该 zip 文件中包括 4 个 JAVA 文件和一个 Word 文件。

针对所有提交的作业代码, 在公共平台进行统一验证和排名, 效果佳者评分优。