

数据库工程师面试题

前言

1. 程序从CSV 格式的文件中读取数据(数据文件请自行准备)，每行包含英文逗号分隔的两个整数，其中含有上万行数据，可以看做一个 `foo(a int, b int)` 的数据库关系表，样例如下：

```
123,456
222,789
100,456
...
...
10000,12345
```

题目

用任意编程语言实现 `SELECT AVG(a), b from foo group by b;`
并将结果输出到标准输出。

要求

1. 不要使用任何第三方软件如(数据库，Kafka等)，csv解析库除外
2. Query的执行速度越快越好，尽可能的使用内存，CPU的资源(可以认为内存有能力完全装载数据文件)
3. 完成测试，单元测试和回归测试
4. 在项目文件夹下有完整的README，至少要包含程序编译环境，编译命令以及程序使用的算法以及整体思路
5. 完成程序的性能分析文档，指出性能瓶颈以及未来优化方向，以及有没有其他方式完成这道题。并指出适用场景
6. 将程序放在源代码托管服务平台中(如GitHub，Gitlab等)
7. 作业没有时间限制，请尽可能保证作业的代码风格