

分布式系统作业

第 4 次作业

姓名：唐晨轩

班级：人工智能与大数据

学号：19335182

一、问题描述

Homework-4

利用 Culler 并行程序设计方法计算 1000×1000 的矩阵与 1000×1 的向量之间的乘积, 要求清晰地呈现 Culler 并行程序设计的四个步骤, 并比较程序在不同阶段具有不同配置时如不同的子任务数量、不同的线程数量、不同的映射方案的性能差别。

二、解决方案

Culler 的设计方法

分解(decomposition): 将原问题分解为多个能被并行的子问题

- 分解不一定是静态的, 也可以在程序执行时动态生成任务
- 目标: 使机器一直有活干
- 核心: 确定依赖关系

分派(assignment): 将线程/工人分配到每一个子问题上

- 负载均衡(load-balanced)
- 减少通信开销
- 可静态或动态, 即调度
- 程序员对分解背锅, 语言/运行时系统对指派背锅

协调(orchestration): 不同线程间的交流

- 涉及结构化通信、添加维持依赖关系的同步、内存安排数据结构、任务调度
- 目标: 减少通信/同步开销, 保留数据间的局部性

映射(mapping): 将并行执行逻辑 (线程/工人) 对应到硬件资源上

- OS: pthread、编译器: ISPC、硬件: CUDA 线程块
- 将相关的线程放在同个处理器上

我的思路是:

分解: 将原问题 1000×1000 的矩阵乘以 1000×1 的向量分解为 1×1000 矩阵与 1000×1 向量的计算

分派: 将每个子任务分派到各自的线程上计算

协调: 通过共享内存, 每个子任务的运算结果都存在于 res 数组中, 在需要访问内存时加锁保护

映射: 这个是操作系统的工作, 我无法干涉

1. 使用 4 个线程，每个线程分配 250 个 1×1000 矩阵与 1000×1 向量的计算
2. 使用 8 个线程，每个线程分配 125 个 1×1000 矩阵与 1000×1 向量的计算

由于我的 perf 在 hw2 之后出问题了，一直没有成功解决，所以这次作业对性能的评判我选择使用运算时间 time。

在 1 思路下的性能：

```
hadoop@ubuntu: ~/hw4$ g++ hw.cpp -lpthread
hadoop@ubuntu: ~/hw4$ ./a.out
time = 0.428333
```

在 2 思路下的性能：

```
hadoop@ubuntu: ~/hw4$ g++ hw.cpp -lpthread
hadoop@ubuntu: ~/hw4$ ./a.out
time = 0.211556
hadoop@ubuntu: ~/hw4$
```

三、实验中遇到的问题

还是 perf 的问题，hw2 使用过后再次打开虚拟机使用就显示未安装了，而且按之前的安装步骤又走不通，确实是个问题，还需多方搜集建议去修复。