

实验 2 分布式存储编程

目录

0 前言.....	2
实验 2-1 MPI 编程.....	2
1. 实验目的.....	2
2. 实验要求.....	2
3. 实验内容.....	2
3.1 并行算法设计.....	2
3.2 MPI 编程.....	2
3.3 程序调优.....	3
3.4 实验提交.....	3
4. 参考资料.....	3
5. 评分标准.....	3
实验 2-2 Map/Reduce 编程.....	4
1. 实验目的.....	4
2. 实验要求.....	4
3. 实验内容.....	4
3.1 Hadoop 计算平台搭建.....	4
3.2 Map/Reduce 编程.....	4
3.3 实验提交.....	4
4. 参考资料.....	5
5. 评分标准.....	5

0 前言

实验 2 分为 MPI 编程和 MapReduce 编程两部分，学生可自由选择一部分（即 MPI 编程或 MapReduce 编程）进行实验。

实验 2-1 MPI 编程

1. 实验目的

本实验的目的是通过练习掌握分布存储并行编程的知识和技巧。

- 了解并行算法的设计方法
- 掌握 MPI 并行程序编写的基本步骤
- 掌握 MPI 编程环境和工具的使用
- 了解 MPI 程序调试和调优的技巧

2. 实验要求

- 独立完成实验内容；
- 实验报告；（简单要求如下）
 - 1) 并行算法的设计思想
 - 2) 程序设计及实现
 - 3) 回答实验中提出的问题；
 - 4) 结果分析；
- 随实验报告，附代码、程序说明以及运行结果。

3. 实验内容

3.1 并行算法设计

实现下列算法之一，说明原串行算法的原理以及并行化的方法。

- （1）矩阵乘法的并行算法
- （2）并行排序算法

3.2 MPI 编程

以 MPI 编写并行算法的计算程序。

3.3 程序调优

对程序进行调优，调优内容可包括但不限于：

(1) 自动生成不同大小的矩阵或数据，多次运行程序，记录在不同矩阵大小或数据长度下的运行时间。

(2) 设置环境变量调整处理器数目，并得出相应的性能曲线。

(3) 找出耗时最大的部分，并进行优化；针对改进进行性能比较。(可选)

(4) 对实验结果进行详细分析，将实验结果画成图表。

如果你在以上这些方面做出了比较大的努力，请在报告中做出描述。(加分项)

3.4 实验提交

完成关于本次课程设计的综合实验报告，包括对以下问题的阐述：

(1) 说明程序中所采取的并行化方式，并给出算法的详细说明。

(2) 说明如何解决 MPI 的通信以及负载均衡问题。

(3) 谈谈对 MPI 并行编程的一些认识、说明 OpenMP 或 CUDA 编程的差异。

(4) 在实验中遇到的问题以及解决方法等。

(5) 与其他同学(或本人)实现的相同算法的 MapReduce 并行实验结果进行对比和分析。(可选)(加分项)

以上问题阐述应逐点说明，并包括在实验报告中。

4. 参考资料

- MPI 编程手册：参考书籍“高性能计算之并行编程技术—MPI 并行程序设计”

5. 评分标准

- A——完成实验内容，实验结果和问题回答正确，程序运行性能较好，实验报告叙述清晰完整，有详尽的分析和总结。
- B——完成实验内容，实验结果和问题回答正确，实验报告叙述清晰完整。
- C——完成实验内容，实验叙述清晰完整。
- D——完成部分实验内容。
- E——未按时完成实验内容，或者抄袭。

实验 2-2 Map/Reduce 编程

1. 实验目的

本实验的目的是通过练习掌握 Map/Reduce 编程的知识和技巧。

- 熟悉 Hadoop 的编程环境
- 掌握用 Map/Reduce 编写并行计算程序的基本步骤
- 了解 Map/Reduce 计算程序在 hadoop 下的运行和调试机制

2. 实验要求

- 独立或合作（1~2 人）完成实验内容；
- 实验报告：（简单要求如下）
 - 5) 程序设计及实现
 - 6) 回答实验中提出的问题；
 - 7) 结果分析；
- 随实验报告，附代码、程序说明以及运行结果。

3. 实验内容

3.1 Hadoop 计算平台搭建

选择以下任一种方式来搭建 Hadoop 平台：

- （1）在 2~4 台 Linux 虚拟机上搭建分布式 Hadoop 计算平台；
- （2）在 2~4 台 Linux 计算机搭建分布式 Hadoop 计算平台；
- （3）在阿里云或腾讯云等公有云上，用不少于 3 台 Linux 虚拟机搭建一个分布式 Hadoop 计算平台。

在搭建好的平台上运行 Word Count 程序，输出运行结果。

3.2 Map/Reduce 编程

基于 MapReduce 实现以下任意一个并行算法：

- （1）矩阵乘法的并行算法
- （2）并行排序算法

3.3 实验提交

完成关于本次课程设计的综合实验报告，包括对以下问题的阐述：

- （1）概述云计算平台的设计及搭建过程。

- (2) 说明并行算法设计的基本思路，并列出参考文献。
 - (3) 说明如何进行系统配置，设置 workers 及任务调度。
 - (4) 谈谈对 Hadoop 平台以及 Map/Reduce 编程的认识。
 - (5) 在本实验中遇到的问题以及解决方法等。
 - (6) 与其他同学（或本人）实现的相同算法的 MPI 并行实验结果进行对比和分析。（可选）（加分项）
- 以上问题的回答应逐点说明，并包括在实验报告中。

4. 参考资料

- Hadoop 的 Map/Reduce Tutorial:
<http://hadoop.apache.org/docs/stable/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html>

5. 评分标准

- A——完成实验内容，实验结果和问题回答正确，程序运行性能较好，实验报告叙述清晰完整，有详尽的分析和总结。
- B——完成实验内容，实验结果和问题回答正确，实验报告叙述清晰完整。
- C——完成实验内容，实验叙述清晰完整。
- D——完成部分实验内容。
- E——未按时完成实验内容，或者抄袭。
- 完成可选部分有适当的加分。