

实验 3 云计算编程

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）
 - 1) 程序设计及实现
 - 2) 回答实验中提出的问题；
 - 3) 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）并行串匹配算法
- （2）平行排序算法
- （3）最短路径的并行算法
- （4）矩阵乘法的并行算法（简单并行算法、cannon 算法或 DNS 算法等）

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。

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。