



Hadoop-MapReduce

个人总结

郑思文

2015 年 6 月



1. 说明

1.1. 概述

MapReduce 是一种编程模型，用于大规模数据集（大于 1TB）的并行运算。概念"Map（映射）"和"Reduce（归约）"，和它们的主要思想，都是从函数式编程语言里借来的，还有从矢量编程语言里借来的特性。它极大地方便了编程人员在不会分布式并行编程的情况下，将自己的程序运行在分布式系统上。当前的软件实现是指定一个 Map（映射）函数，用来把一组键值对映射成一组新的键值对，指定并发的 Reduce（归约）函数，用来保证所有映射的键值对中的每一个共享相同的键组。

1.2. 说明

本文主要是介绍我在本学期的软件工程综合实验中所做的工作和总结。包括从开始的任务书到测试分析，再到现在的总结报告。我在小组内的主要任务是撰写各个阶段的文档，与小组同学进行各个阶段的讨论并承担其他必要的任务。

2. 任务书

在任务书的撰写阶段，本小组先后多次讨论，进行了三次修改，最终确定了本次综合实验的任务，并负责撰写任务书。

3. 软件需求规格说明书

软件需求规格说明书从初稿到终稿经历了 5 个版本的更新，更新内容包括用例图、用例需求规格的更新。在每周课程的评审中，根据评审意见进行不断修改。拟定评审单，并根据评审单对其他组的说明书进行评论。

在这个阶段，整个小组为讨论用例图，连续熬夜并最终完成初稿，在这个过



程中，大家同心协力，认真读书籍及源码，积极提意见，为软件需求规格说明书的完成奠定了坚实的基础。

4. 测试需求规格说明书

测试需求规格说明书从初稿到终稿经历了 3 个版本的更新。与需求规格说明书对照，主要进行的测试用例设计包括：配置测试、身份验证测试、作业提交测试、修改优先级测试、杀死任务测试、杀死作业测试。经过第一次评审之后，增加了处理任务测试、分配 Mapper 任务测试、分配 Reducer 任务测试。在第二次评审之后，根据评审意见，再次修改测试规格说明书，得到最终的测试规格说明书。

本次测试需求规格说明书在整个撰写阶段，历时 3 周，共产生 3 个版本。

5. 实验 6-8 方案

实验 6-8 主要包括工作日志、进度控制分析报告、变更与管理分析报告、工作量分析报告。

我根据小组各个成员的报告，汇总完成了实验 6-8 方案。并在评审前后进行修改，得到了 2 个版本的方案说明书。

我负责更新的工作日志记录着本学期本实验的主要过程，包括所做工作的时间、地点及分工。进度控制分析报告，主要由 Microsoft Project 控制，记录着实验的计划和实际完成情况。变更与管理分析报告主要记录着各个阶段所有文档的变更记录。工作量分析报告主要记录组内各个成员的工作量。

6. 测试报告

根据测试用例，我们抓紧时间做了测试分析，并追加了并行计算能力测试，以测试 MapReduce 的并行能力。



测试报告经过修改，共产生 2 个版本。在测试过程中，小组编写脚本，进行测试。基本完成测试用例中所写的用例。

7. 个人总结

本学期的整个实验课即将结束，我的总结如下：

1.我在组内的主要任务是文档编写，共计 15 篇，包括任务书 1 个版本，软件需求规格说明书 5 个版本，测试需求规格说明书 3 个版本，实验 6-8 说明书 2 个版本，测试报告 2 个版本。以及工作日志 1 篇、软件需求规格说明书复评审问题处理报告 1 篇。

2.我在总结报告中，负责实验方法总结报告。另有其他组员负责数据分析报告和问卷设计。

3.在整个实验过程中，我的工作并不仅仅限制于文档编写。在需求规格说明书的用例分析部分、测试分析部分，我也深度参与并作出自己的贡献，比如用例图设计时，功能的分析，角色的选择等。

4.在整个实验课中，我的思想有了很大转变，最初觉得任务很艰难，老师这有点为难我们，慢慢地按照老师的安排，一步一步去完成每一个小目标，如软件需求规格说明书、测试需求规格说明书等，直到现在，完成既定的测试目标之后，整个实验课接近尾声，回过头来看，做的这些事还是很有意义的。

通过本次实验课，我学会了如何去分析一个软件，如何去撰写规格说明书，如何去做测试用例，以及如何进行测试。更为重要的是，通过合理分工、团队合作来完成一项比较庞大的任务，而这一工作，单靠个人很难完成。

5.给老师的建议：这门课刚开始给人的感觉是，很大、很难、为难人。我会觉得自己能力不足，难以下手。后期慢慢熟悉起来，才觉得是可以完成的。希望老师在课程开始的时候加强引导，增强同学们的信心。另外，老师的要求非常细，对细节要求严格，诸如 Github、Microsoft Project 的使用，也让我们学习到了额外的知识。