**个人总结**

本学期的软工实验即将完结，我们组选择的开源软件是Hadoop，但由于完全分析Hadoop的工作量比较大（至少得掌握HDFS和MapReduce），所以我们组选择MapReduce来作为我们组本次实验的研究方向。

总的来说，本次实验让我收获很多。不论是从Hadoop软件本身还是软件项目过程中的管理和控制等，都让我学到了很多新东西。下面，我将针对这两点，对我本次的收获做一个小结。

一、 关于Hadoop的总结

现在Hadoop是一款比较热门的分布式系统，其中包含的东西也十分多，对于我们本次实验来说，完全吃透Hadoop不是很现实，所以我们选择了其中的MapReduce框架进行了研究。MapReduce是一种编程模型，用于大规模数据集的并行运算。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。

本次实验，从Hadoop集群的搭建到最后MapReduce程序的编写执行，都需要一步步地区尝试与查阅资料。不得不说，这个学习的过程虽然会碰到或大或小的问题，但每个问题的解决都是令人兴奋的，特别是当自己编写的MapReduce程序能够在集群上跑起来并产生正确结果时，都会有些小小的成就感！当然，这次实验中用到的Hadoop集群管理以及MapReduce程序的编写都还只是一些基础的知识，关于如何更好地管理Hadoop集群（如datanode多到上千台的时候集群的管理），如何编写效率更高的MapReduce程序，这些都是需要进一步去学习的。

1. 关于软件项目管理的总结

整个实验持续的时间比较长（大约十六周），这期间，对于项目的管理就是分重要了。为了使项目按照具体流程进行，我们做了一个项目的整体计划表，用来指导我们每一步应该何时做、怎么做以及事后结果等。在项目进行中，我们也会同步更新项目计划的实际完成情况，同时会记录未按照实际情况完成时的原因。

当然，仅仅根据项目计划来是不够的。项目的进行也经常会碰到许多计划之外的问题，如：任务难度问题、小组成员时间冲突问题以及任务分配问题等。不可否认，对于Hadoop的知识以及如何做需求和测试等，这些是我们所需要学习和掌握的，但对于使小组成员协调起来一起高效地完成任务也是十分重要的。本次实验，我们学会了用github来管理我们的数据和文件，知道了如何通过小组会议合理分配任务，如何调动小组成员的积极性高效地完成任务等等。

当然，个人在本次实验中也存在许多做的不够的地方。由于分工原因，实验中报告编写这些方面做的不够多；在需求和测试阶段，很多细节考虑的不够；在程序编写上还不够深入（MapReduce程序的设计还涉及到对很多调用接口进行调用来更高效地编程），这是源码阅读的还不够深入。这些做的不够的地地方都是以后需要改正或是学习的。