**学习日志**

姓名：李奥 学号： 1812998 日期： 6.23

**学习内容安排：**

了解MapReduce以及Hadoop分布式文件系统（HDFS）

**学习反馈：**

MapReduce简介：

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）"，是它们的主要思想，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。

Hadoop分布式文件系统（HDFS）简介：

Hadoop是一个分布式的存储计算框架,其中底层的HDFS是构建面向应用的高层次模块的基础(当然也可以使用其API直接访问HDFS), 比如HBase可以是基于HDFS来存储其HFile文件, 计算框架MapReduce可以使用HDFS中存储的数据.

未掌握知识：

学习心得：简单的了解了MapReduce以及Hadoop分布式文件系统（HDFS）