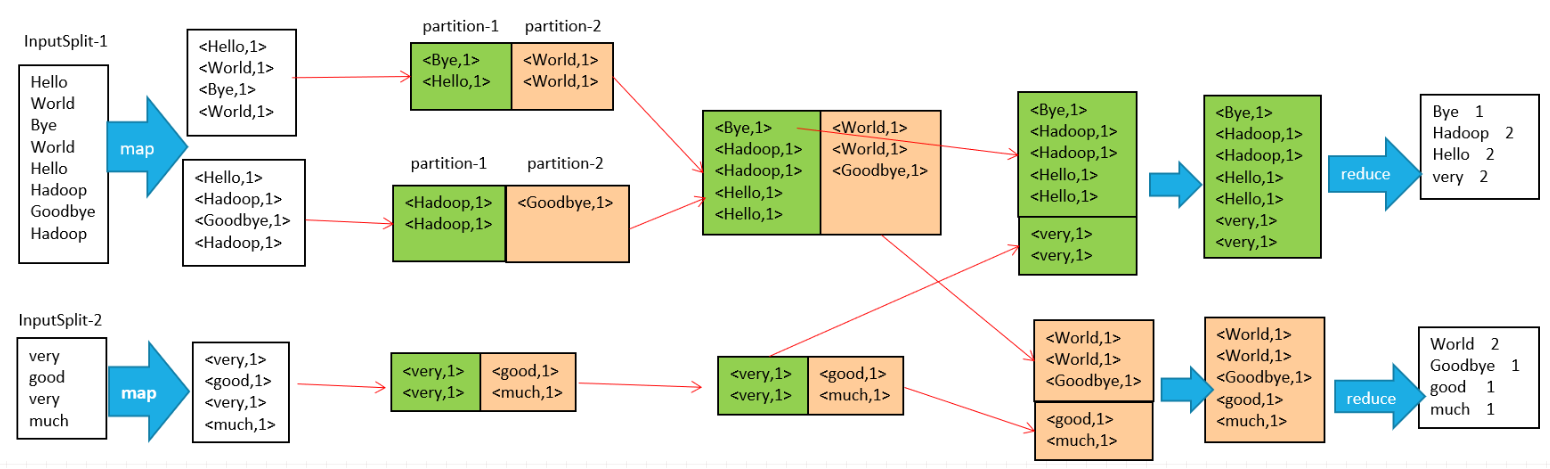
# 学习日志

1813050 张春鹏 6月25日

今日学习了MapReduce工作原理，

在MapReduce整个过程可以概括为以下过程：

输入 --> map --> shuffle --> reduce -->输出

输入文件会被切分成多个块，每一块都有一个map task

map阶段的输出结果会先写到内存缓冲区，然后由缓冲区写到磁盘上。默认的缓冲区大小是100M，溢出的百分比是0.8，也就是说当缓冲区中达到80M的时候就会往磁盘上写。如果map计算完成后的中间结果没有达到80M，最终也是要写到磁盘上的，因为它最终还是要形成文件。那么，在往磁盘上写的时候会进行分区和排序。一个map的输出可能有多个这个的文件，这些文件最终会合并成一个，这就是这个map的输出文件。对MapReduce的输入与输出进行了一定的了解。