

# MapReduce 编程规范及示例编写

## 编程规范

mapReduce 编程模型的总结：

MapReduce 的开发一共有八个步骤其中 map 阶段分为 2 个步骤，shuffle 阶段 4 个步骤，reduce 阶段分为 2 个步骤

### Map 阶段 2 个步骤

第一步：设置 inputFormat 类，将我们的数据切分成 key，value 对，输入到第二步

第二步：自定义 map 逻辑，处理我们第一步的输入数据，然后转换成新的 key，value 对进行输出

### shuffle 阶段 4 个步骤（可以全部不用管）

第三步：对输出的 key，value 对进行分区

第四步：对不同分区的数据按照相同的 key 进行排序

第五步：对分组后的数据进行规约(combine 操作)，降低数据的网络拷贝（可选步骤）

第六步：对排序后的数据数据进行分组，分组的过程中，将相同 key 的 value 放到一个集合当中

### reduce 阶段 2 个步骤

第七步：对多个 map 的任务进行合并，排序，写 reduce 函数自己的逻辑，对输入的 key，value 对进行处理，转换成新的 key，value 对进行输出

第八步：设置 outputformat 将输出的 key，value 对数据进行保存到文件中