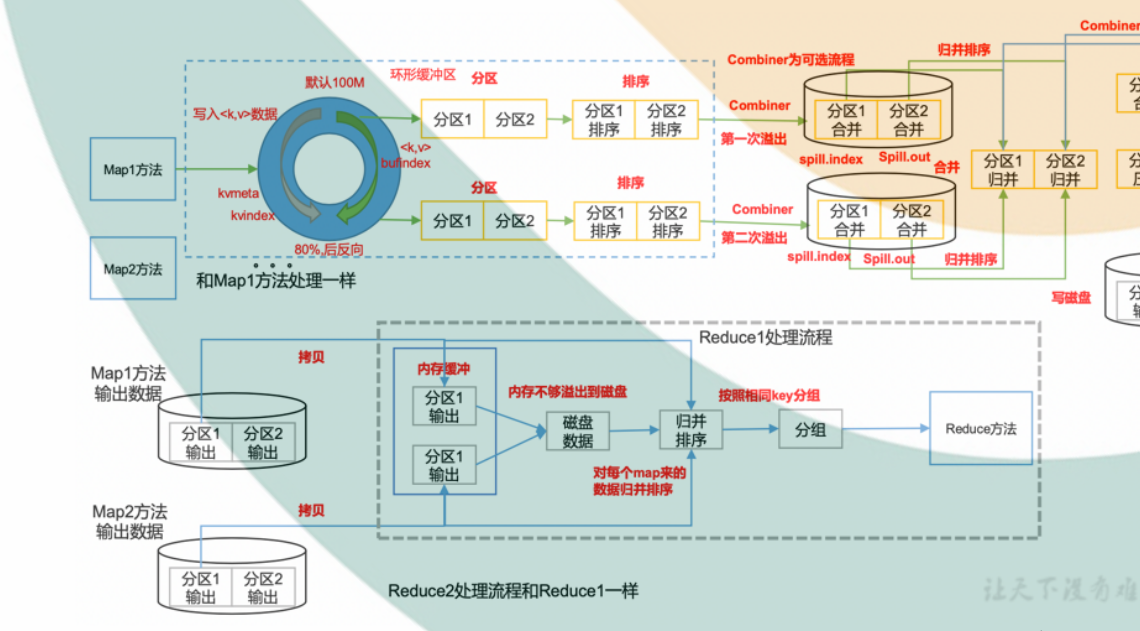
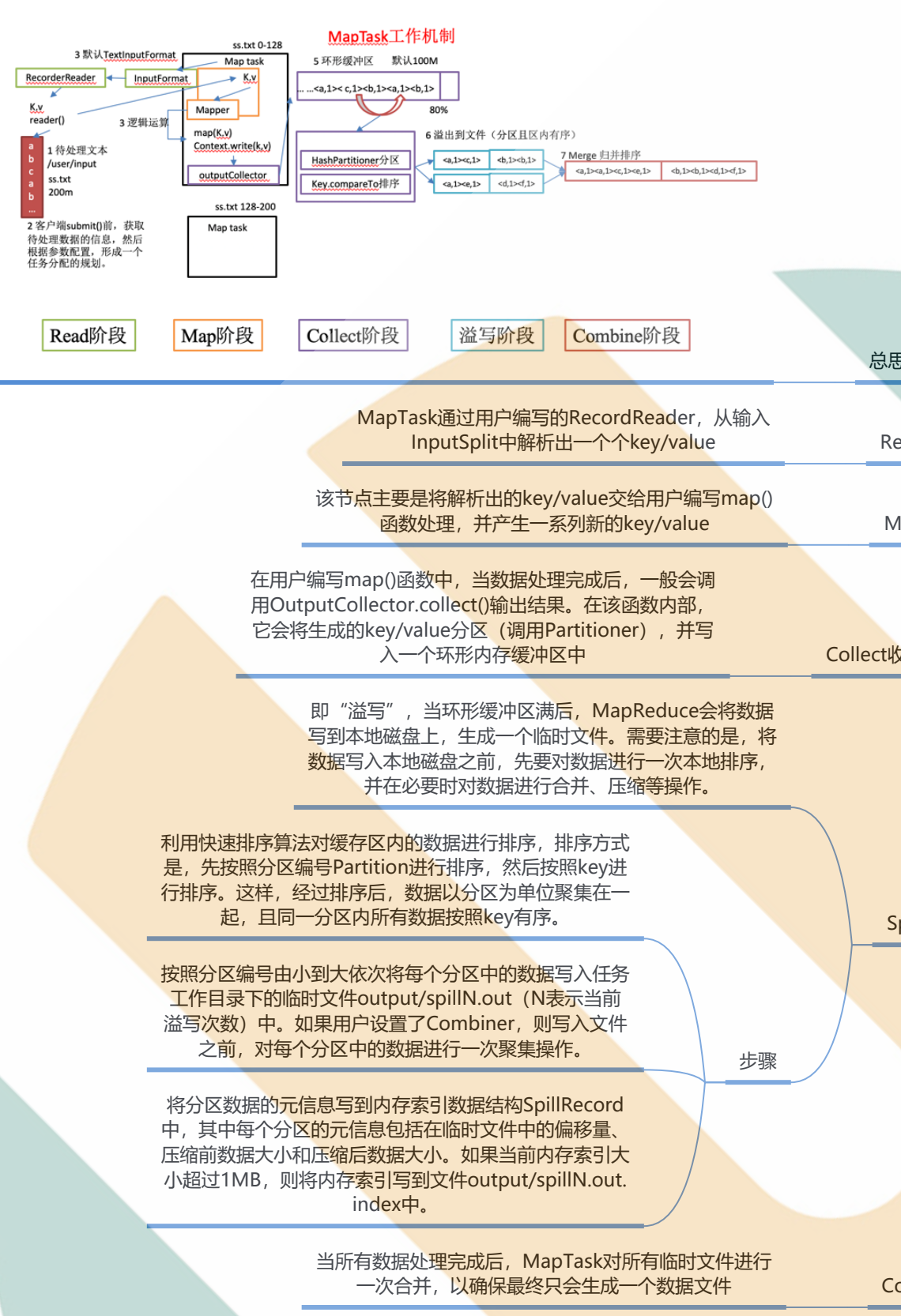
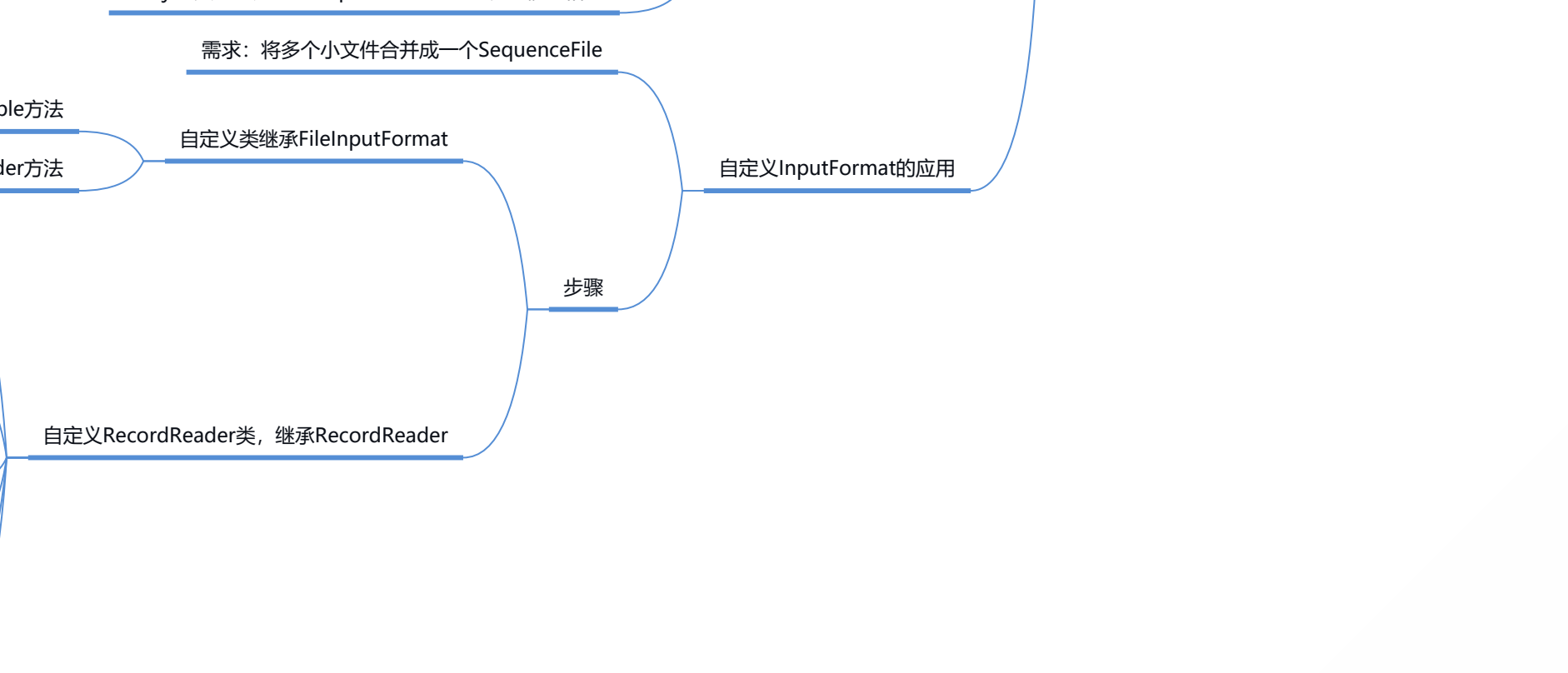
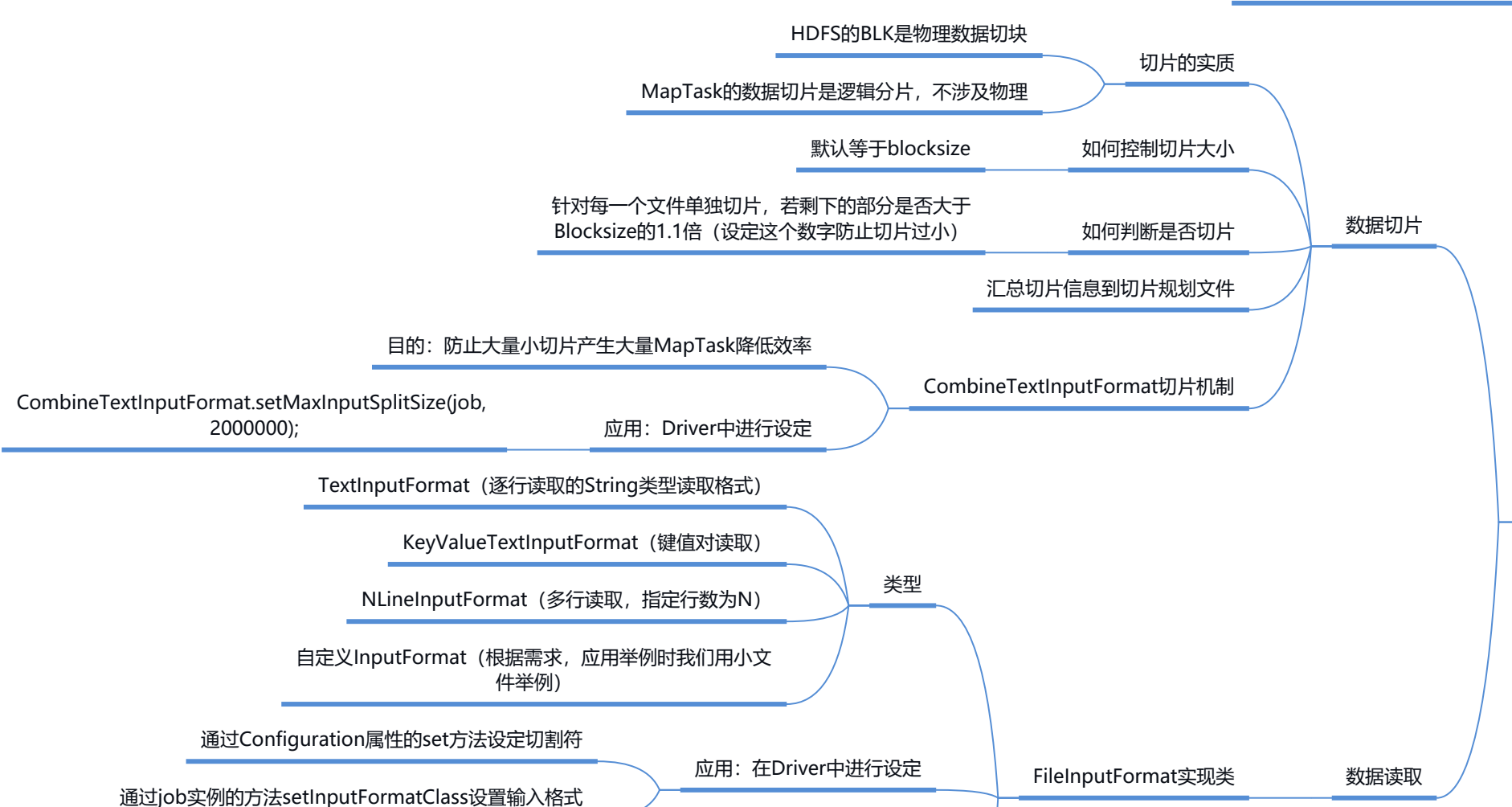
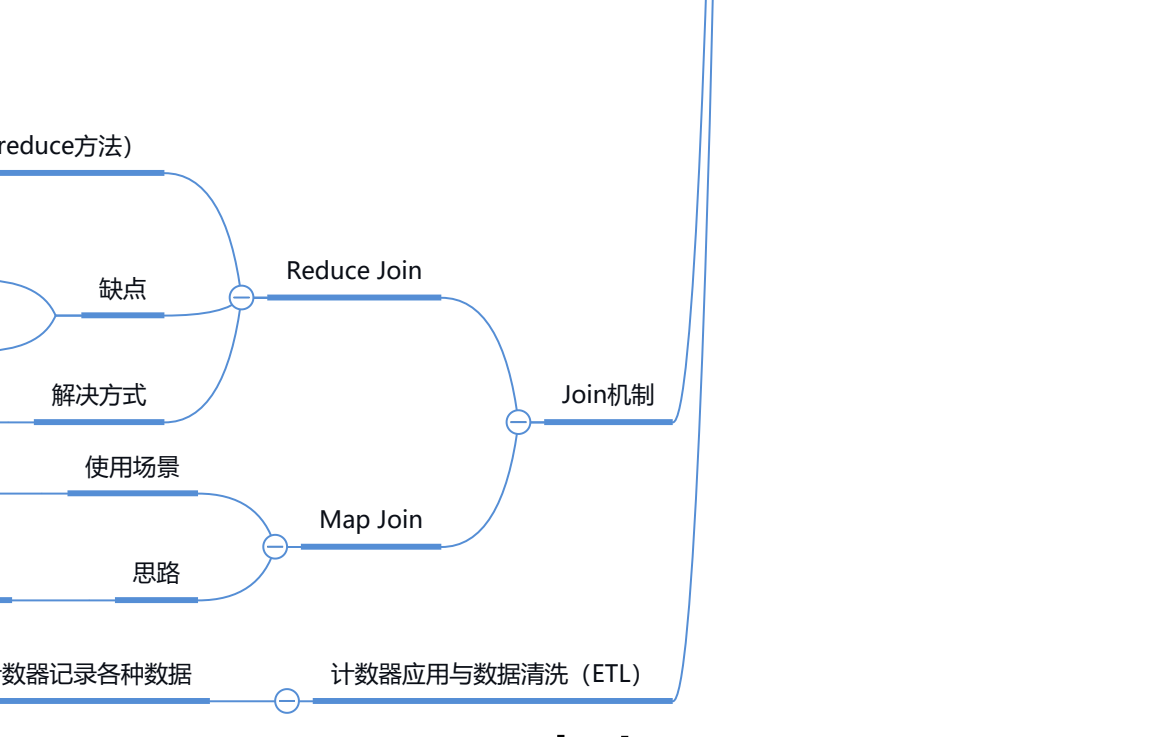
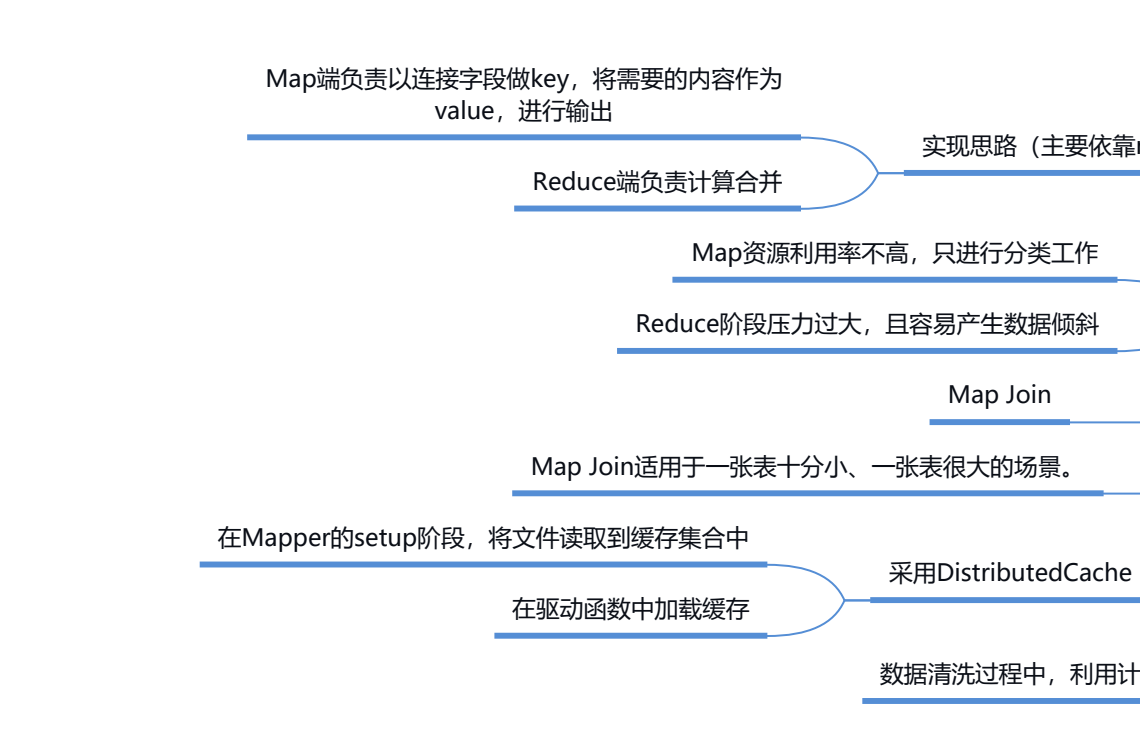
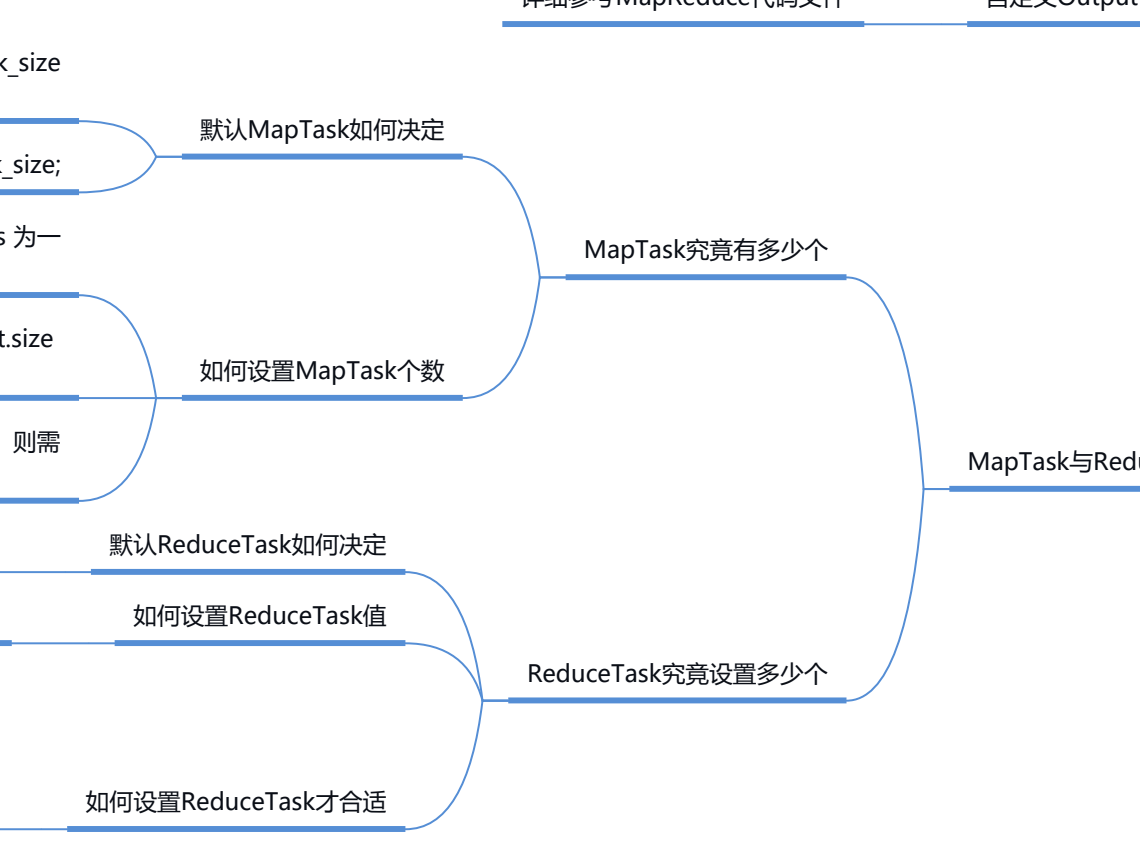
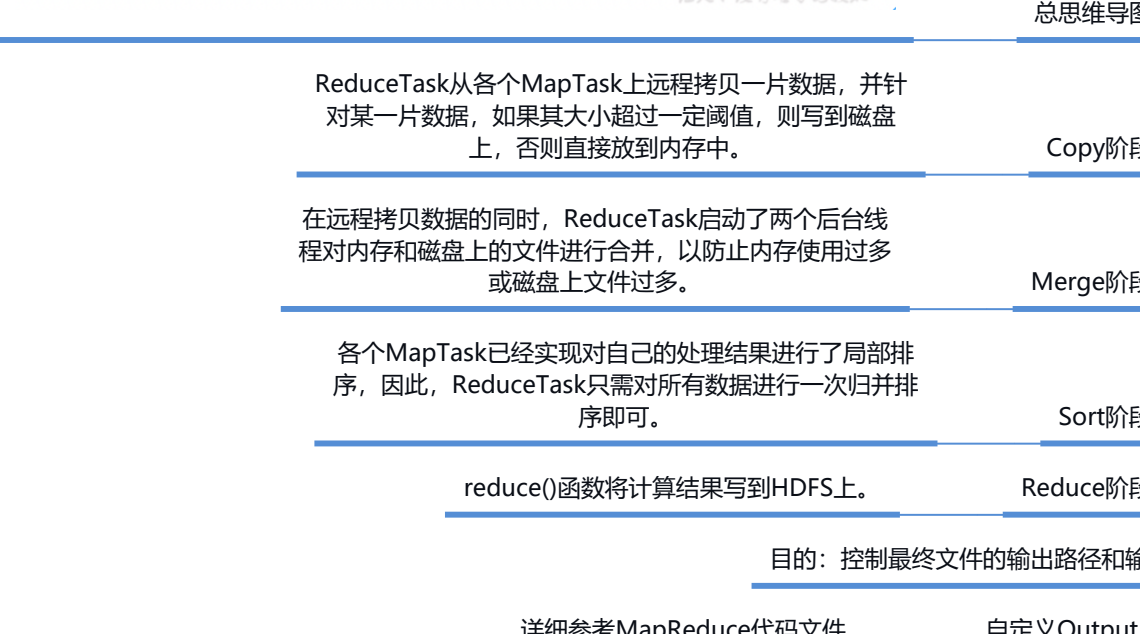
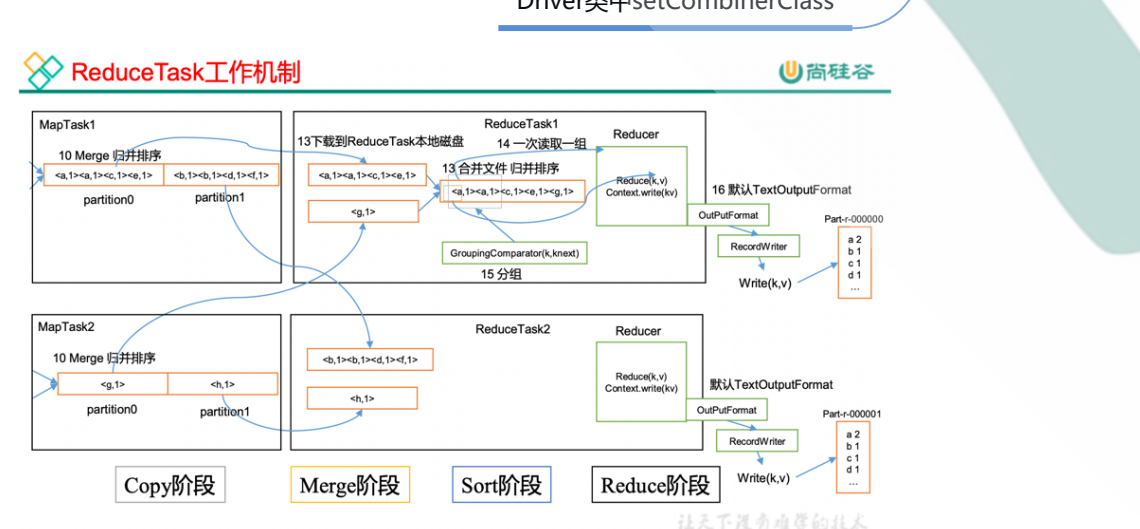
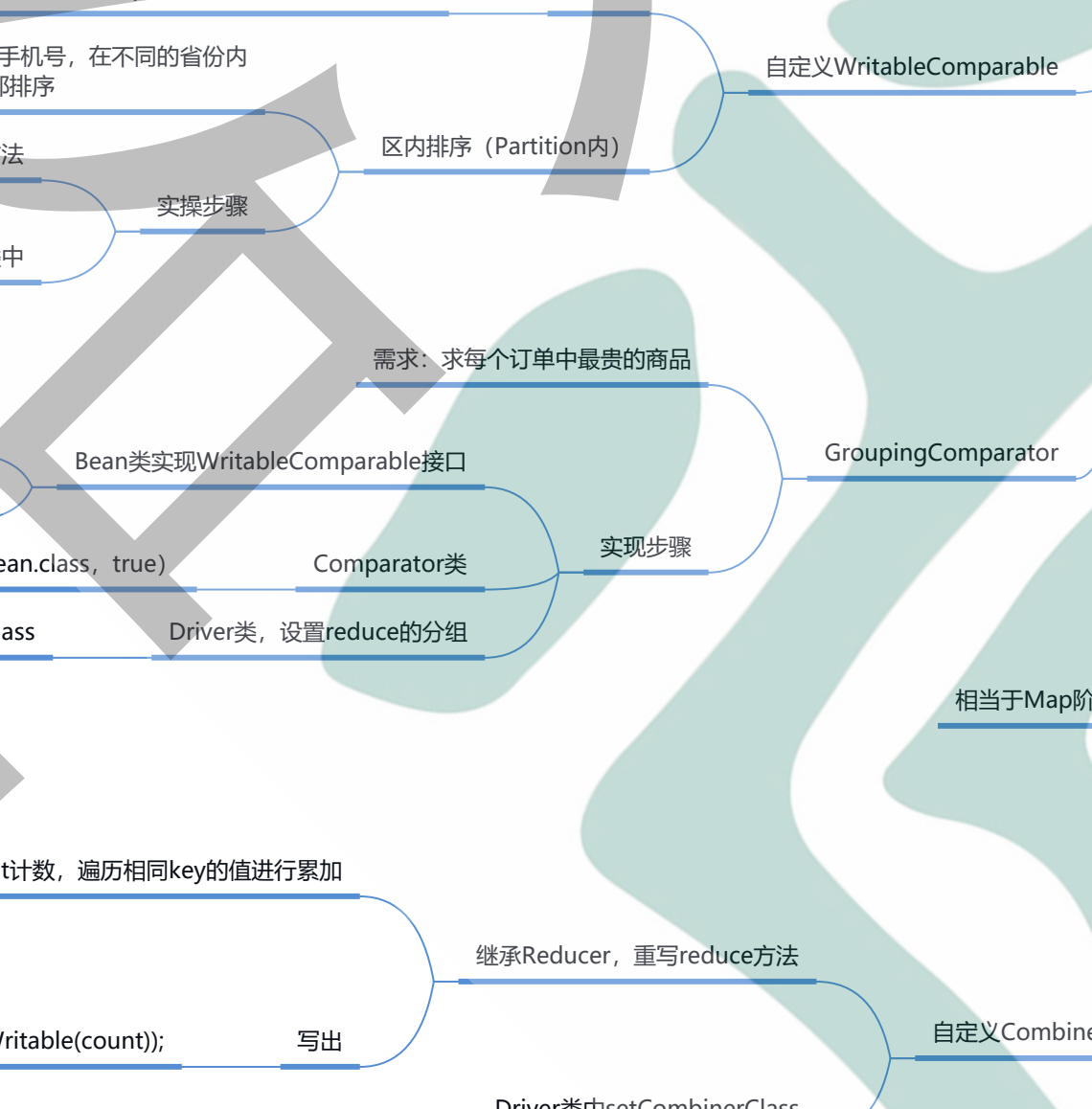
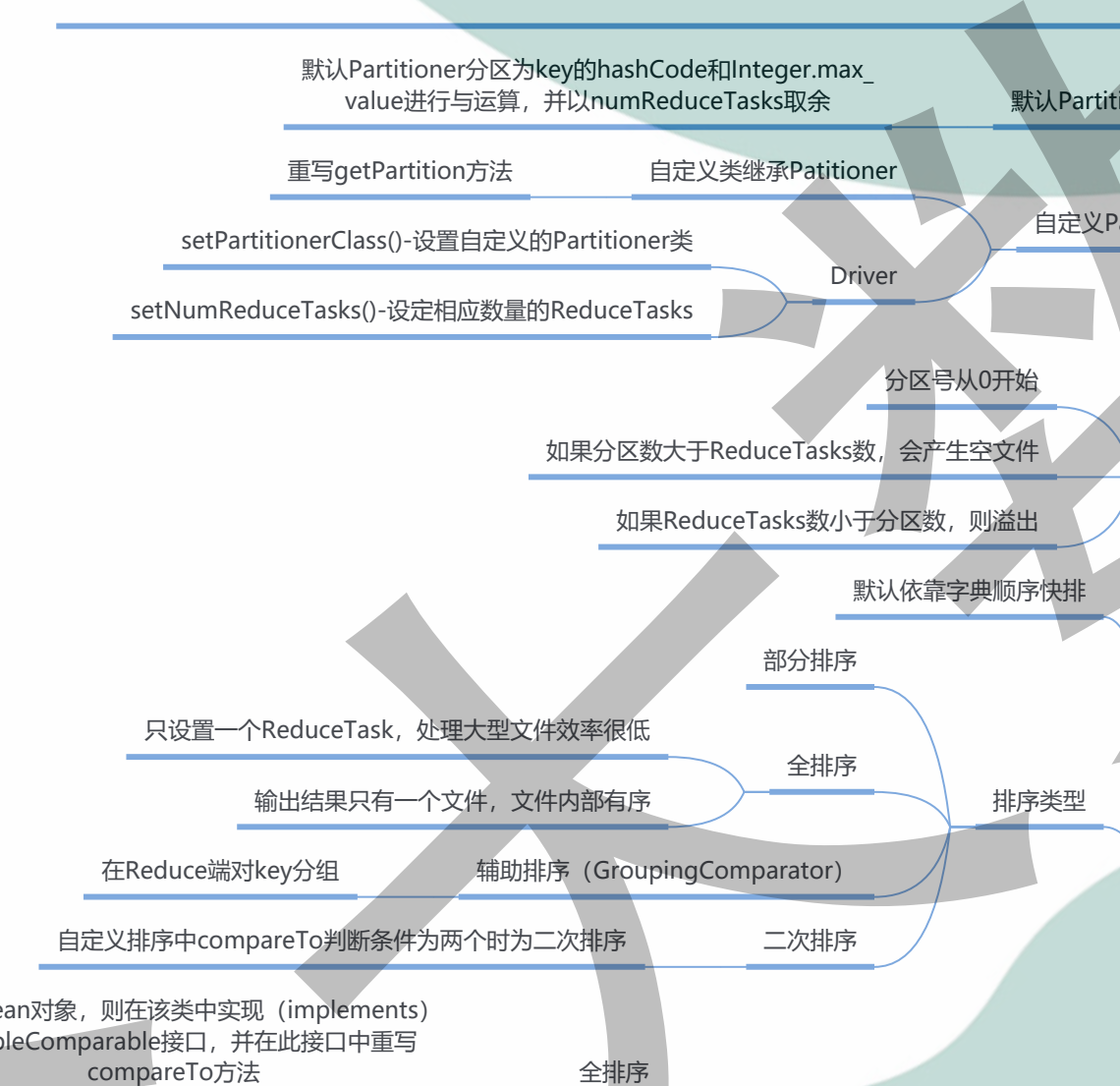


MapReduce在整个Hadoop过程中负责计算部分，它分为两个部分，MapTask和ReduceTask，其中Shuffle介于其间，它的功能运算是把HDFS中的数据进行按照RecordReader，从输入的输出Split中产生Key/Value，写入环形缓冲区。这里依靠keyhash值来计算出各个Split的Partition，并由此时期依顺序进行排序。当环形缓冲区达到80%后，开始Split（写）（此处可进行combiner），当整个文件Split结束后，进行归并（Merge）并对同样的分区（Partition）依顺序进行再排序（此处可进行combiner），最终将数据写到底盘。Reduce方法开始，以K分区，对同组K分区进行计算，利用Reduce方法结束。

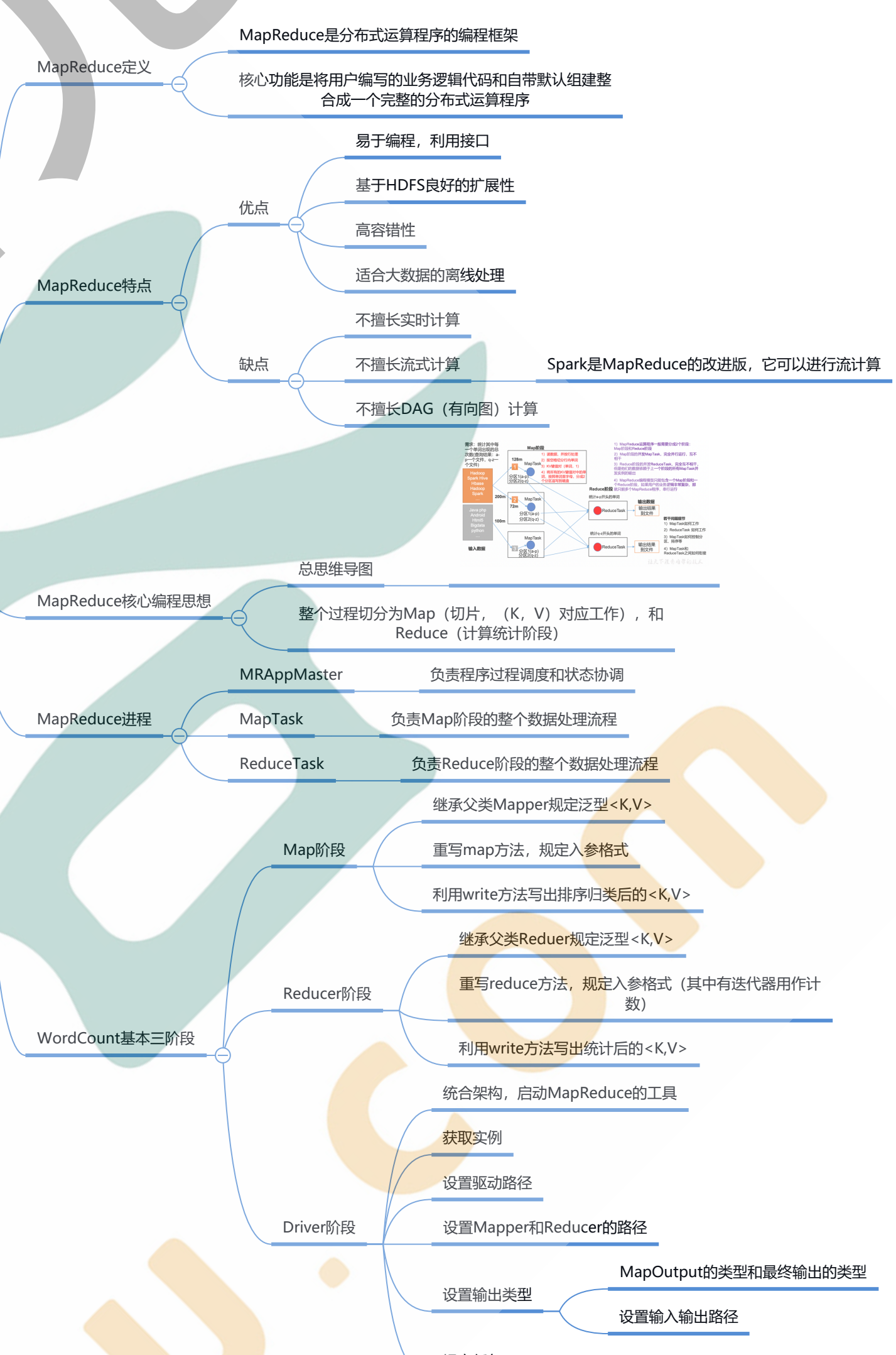
MapReduce总流程概述



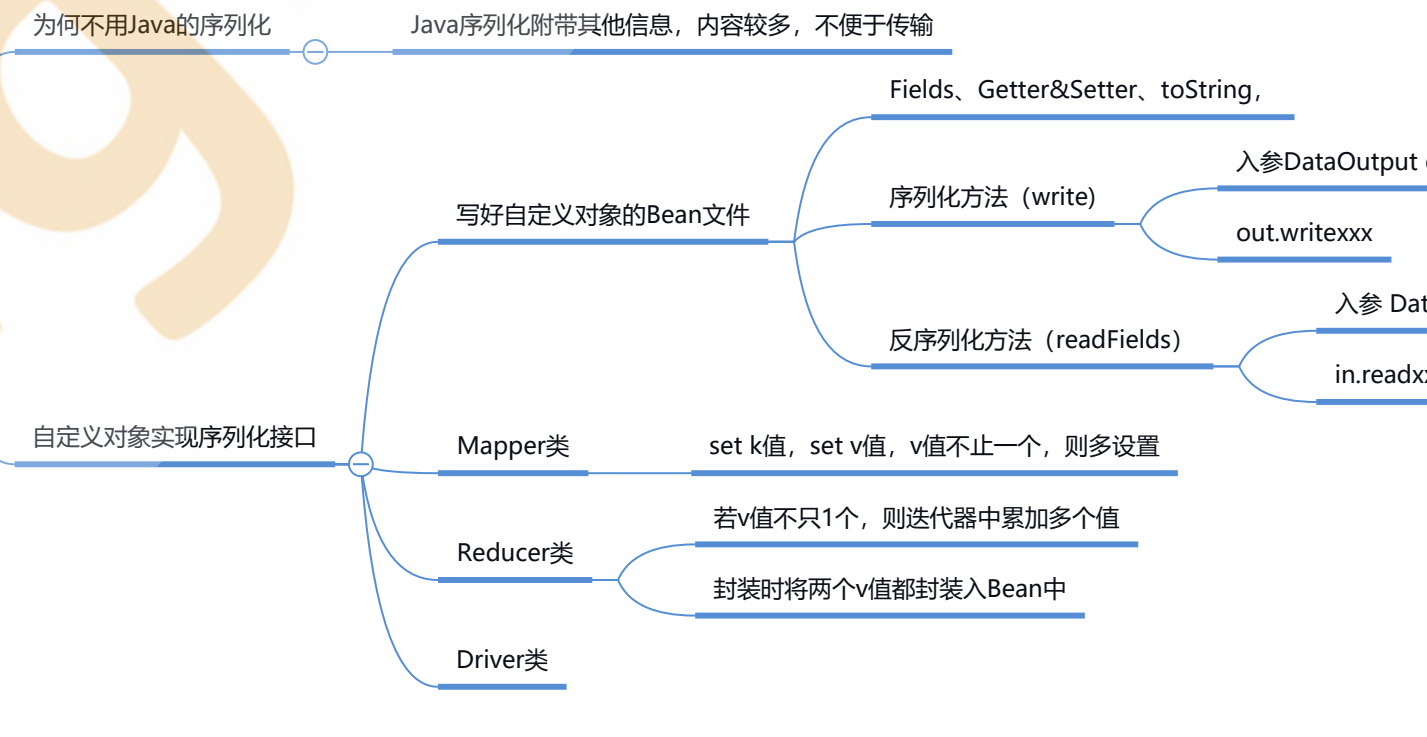
MapReduce框架原理



MapReduce



Hadoop序列化



114

FutureTask小结

1