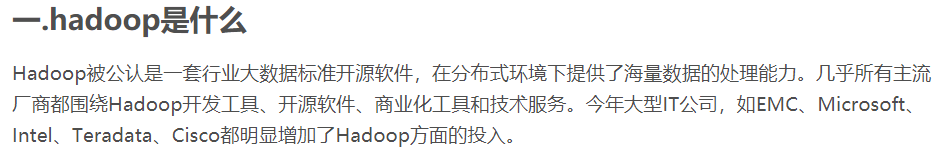
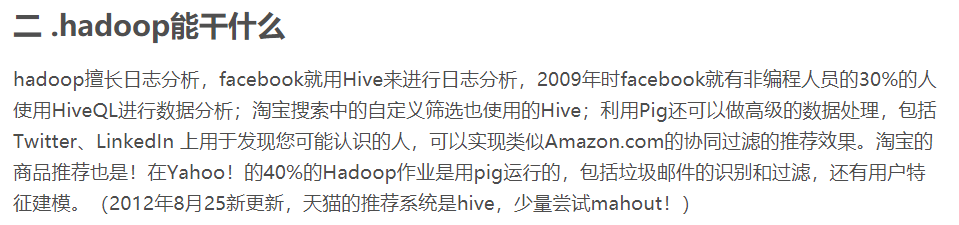
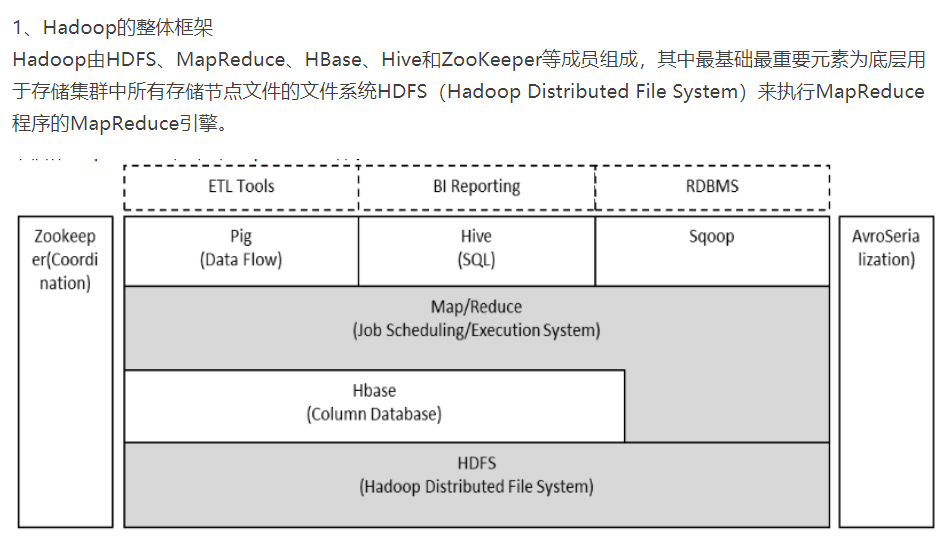
一、单机

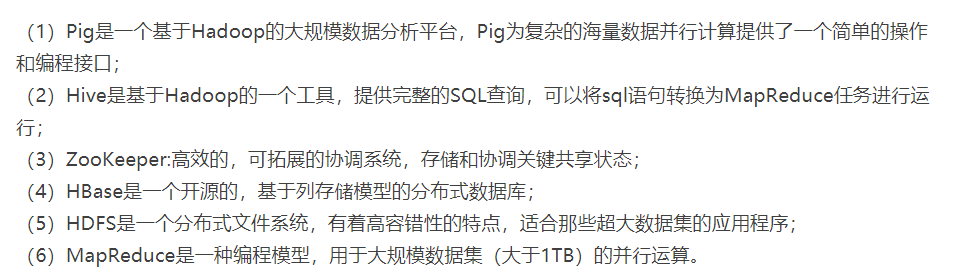




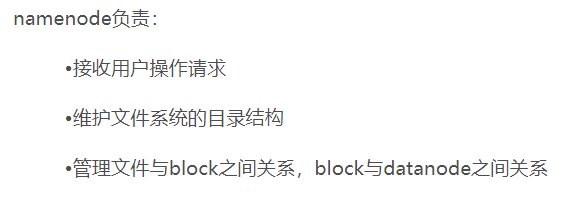
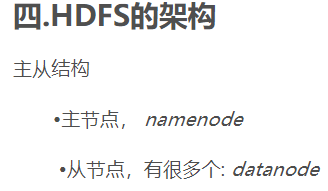


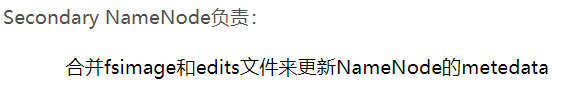
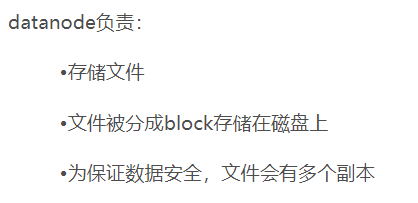


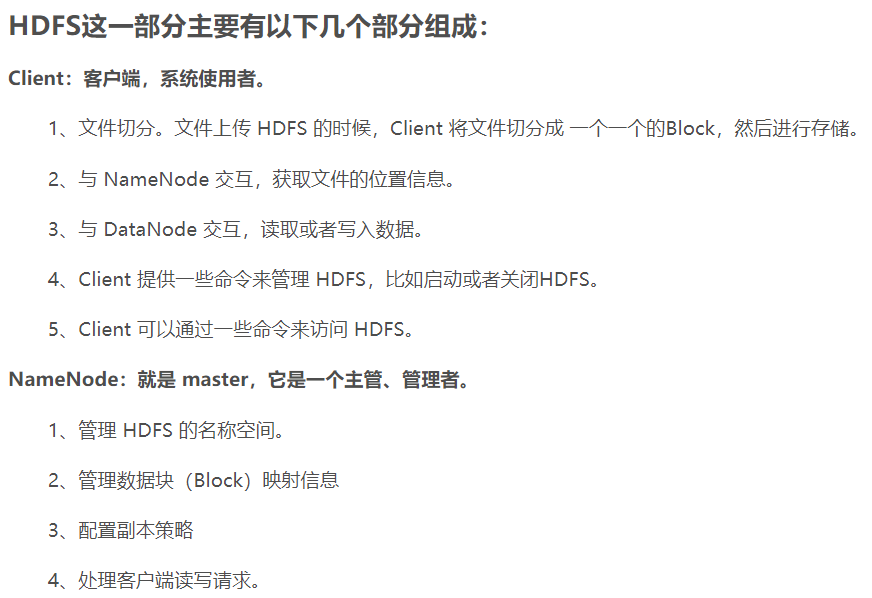


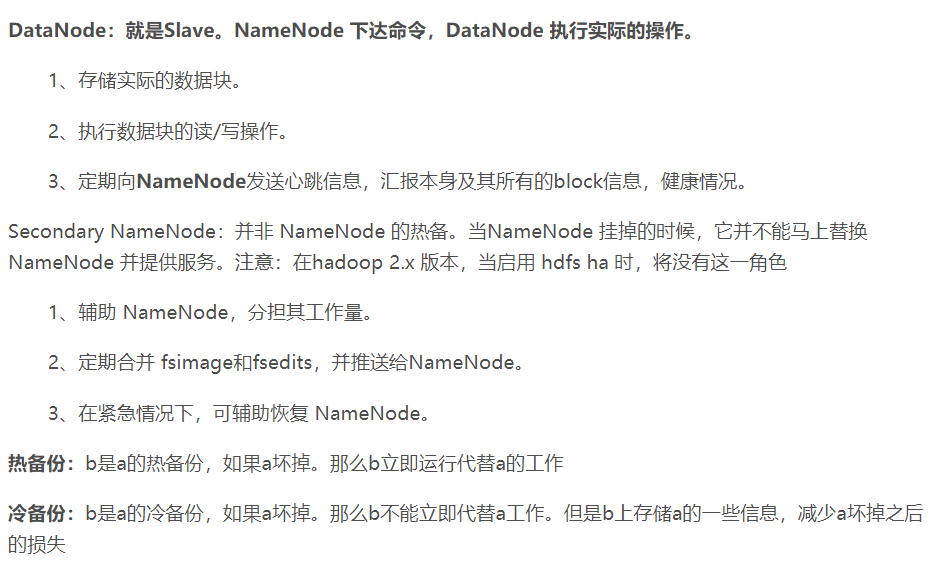


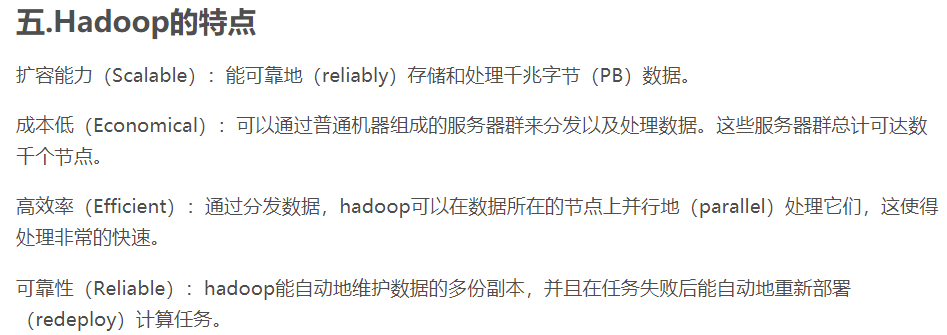




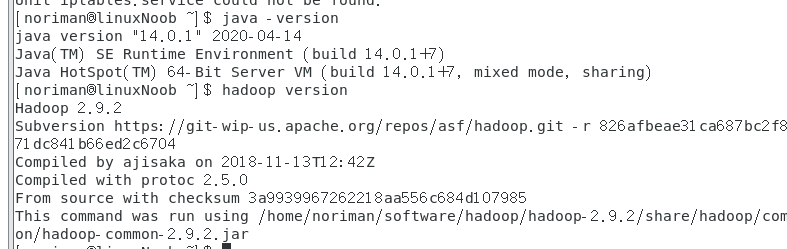


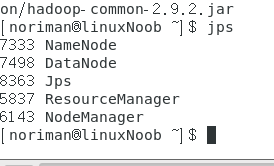






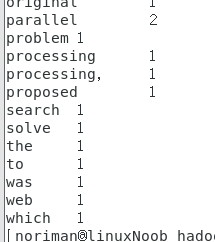
1.在linux虚拟机安装jdk，hadoop。结果如下



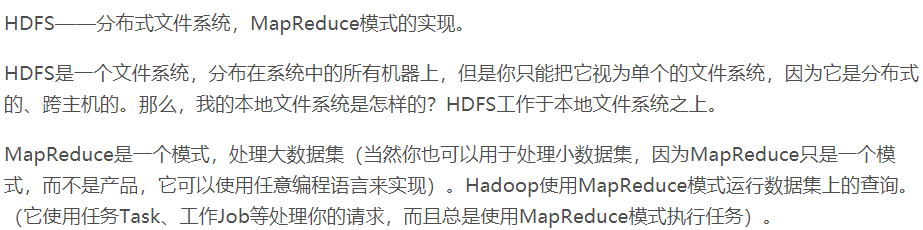


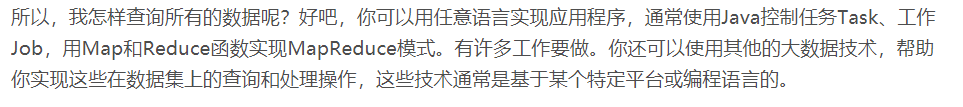


2.Wordcount

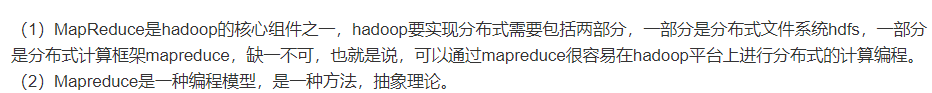


二、mapreduce 分布式

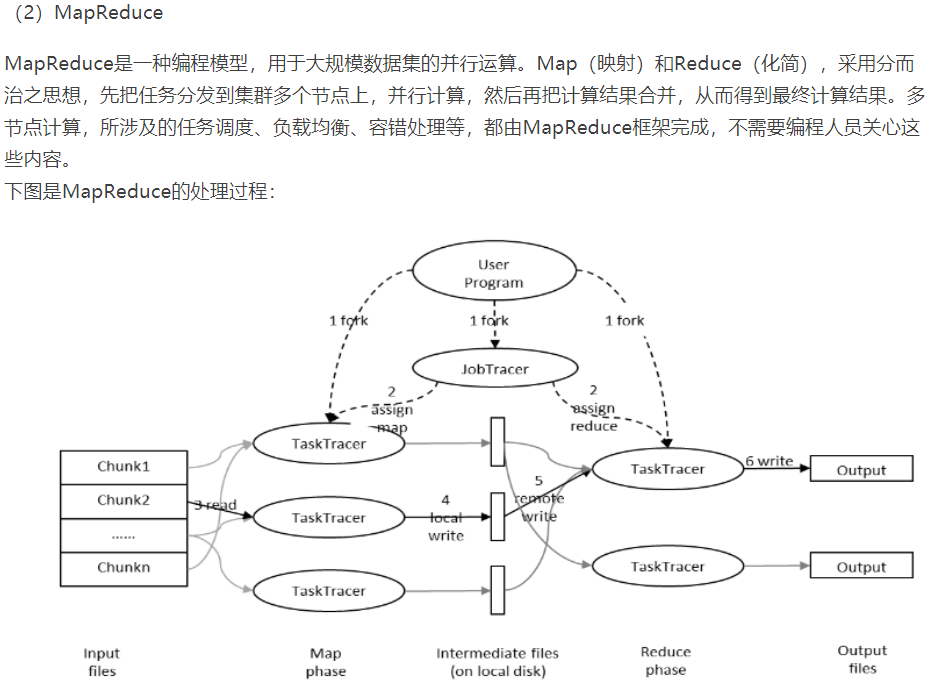


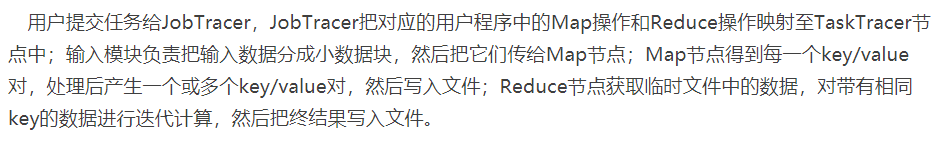


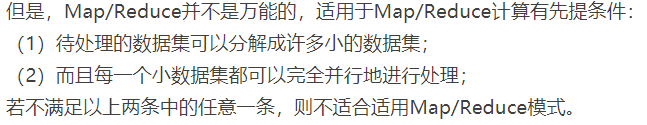
1.MapReduce

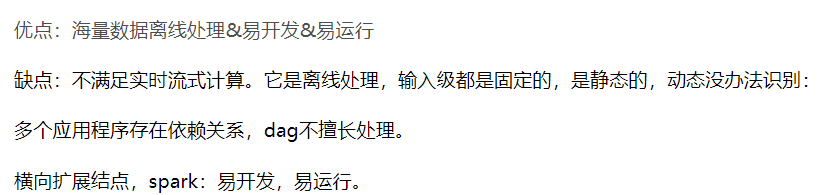


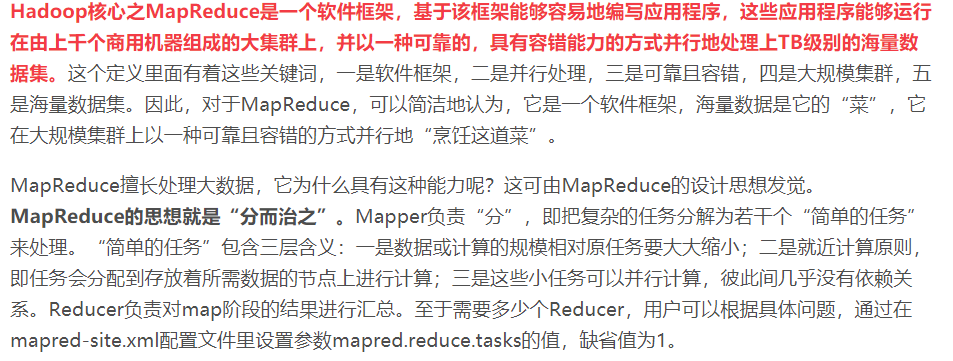
关于怎样解释MapReduce 详见 奋斗的小炎 MapReduce的通俗理解与入门

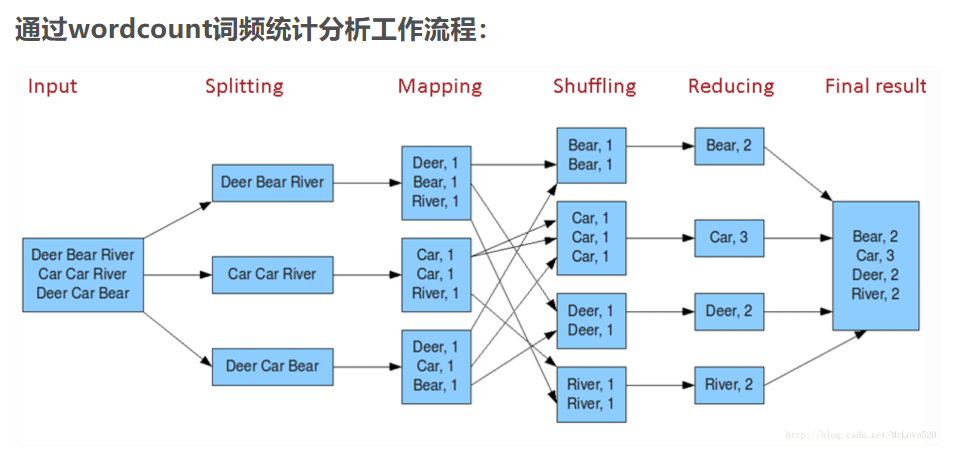






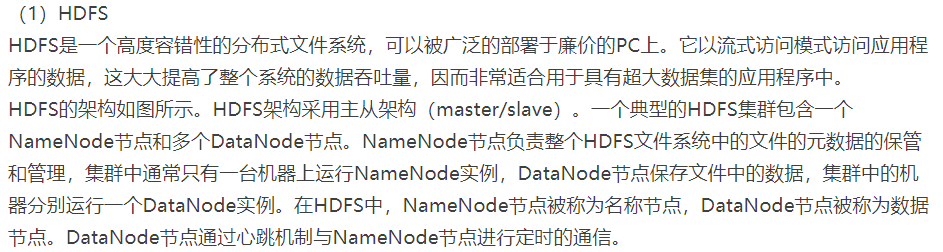


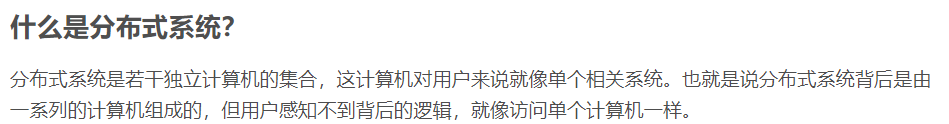


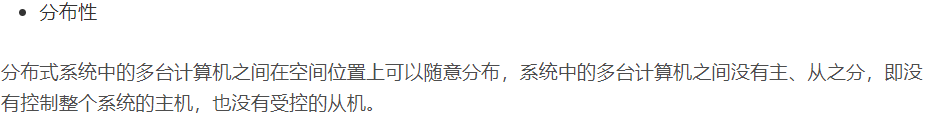
关于工作过程详见

<https://blog.csdn.net/qq_24140237/article/details/81206963?ops_request_misc=&request_id=&biz_id=102&utm_term=hadoop&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduweb~default-1-81206963>

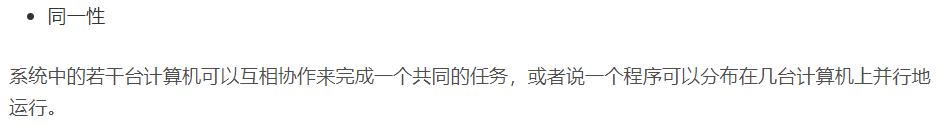
2.分布式

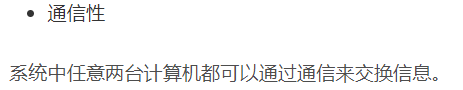


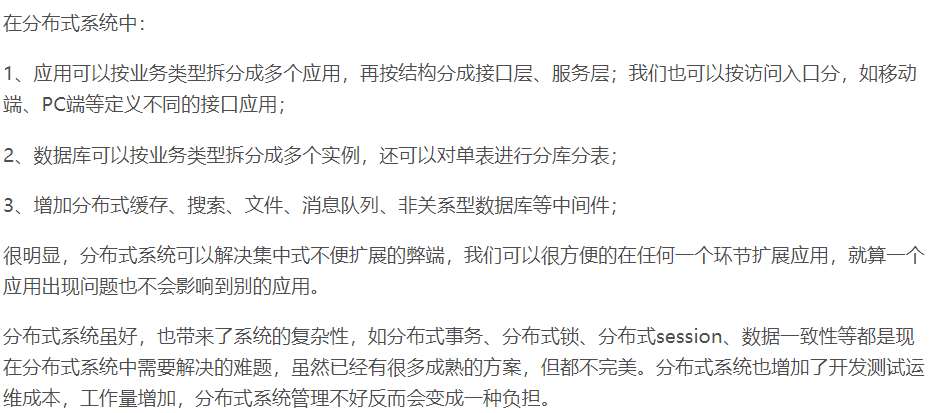


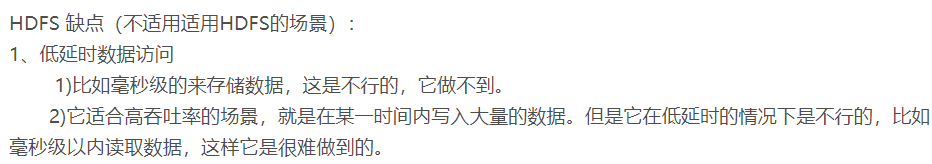


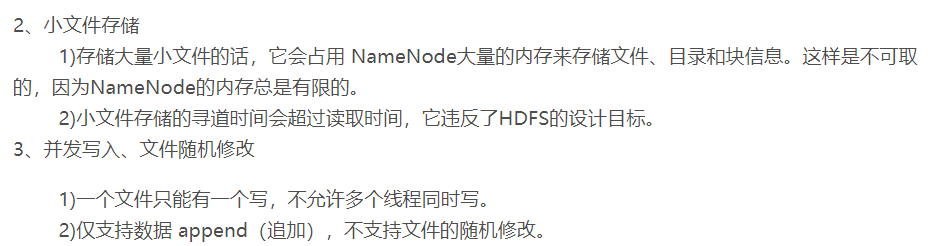


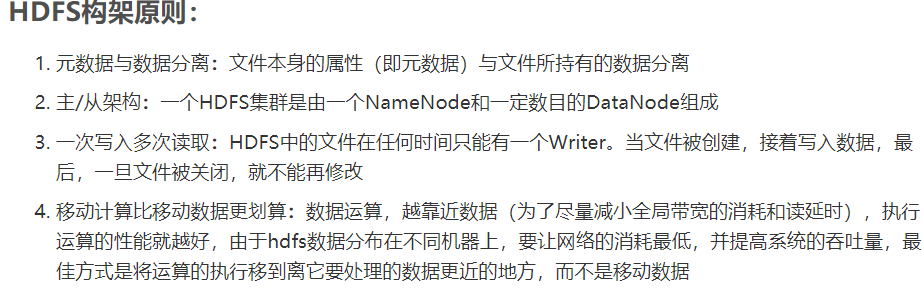




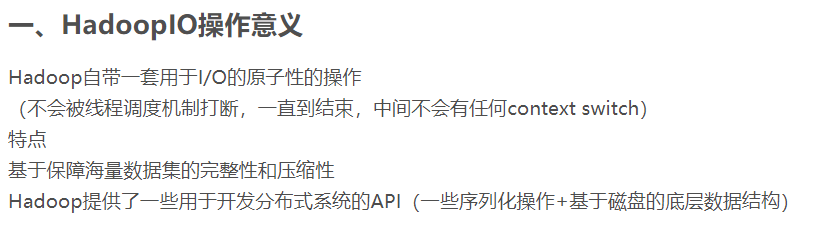




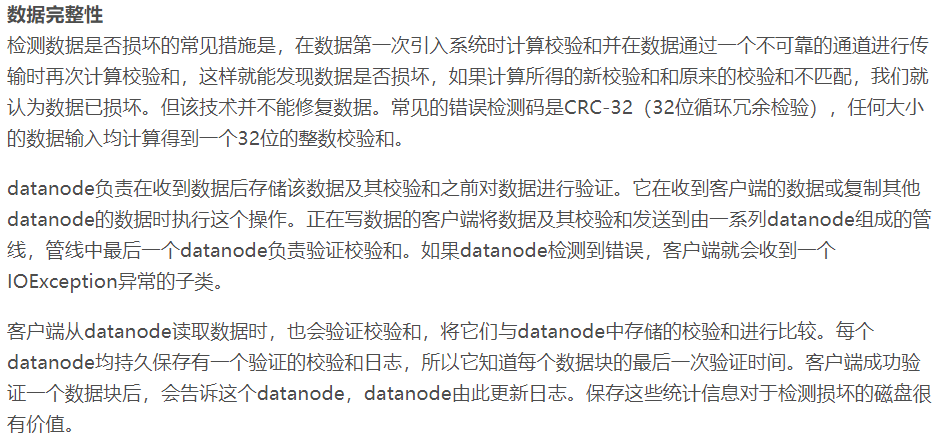




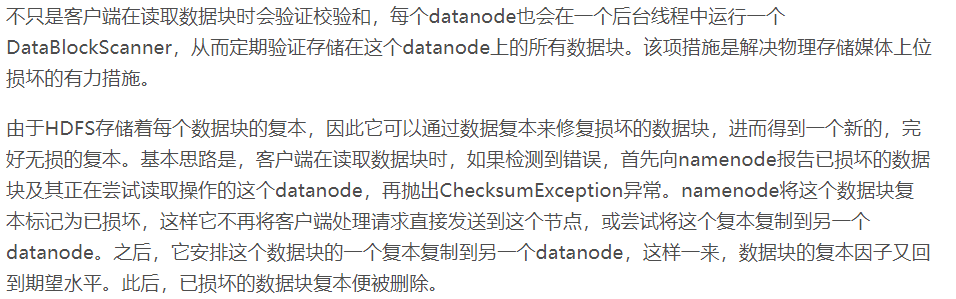
三、IO

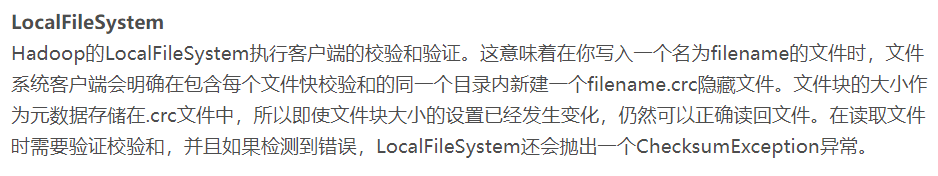


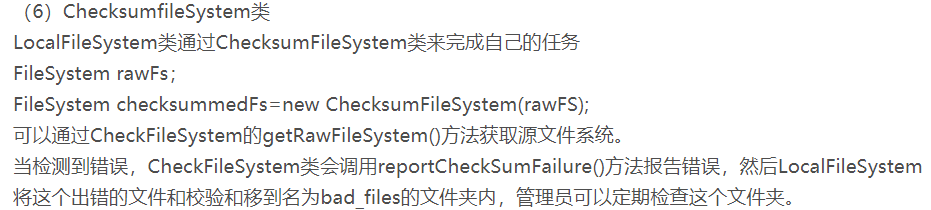


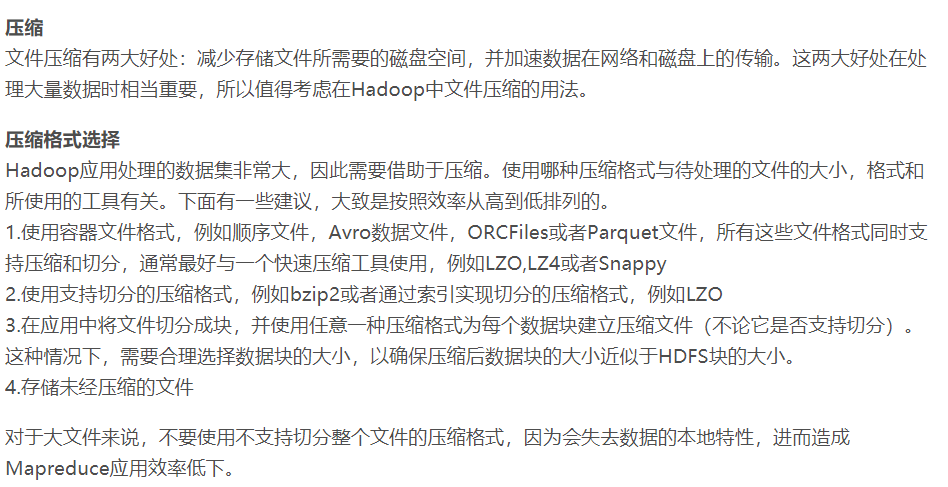


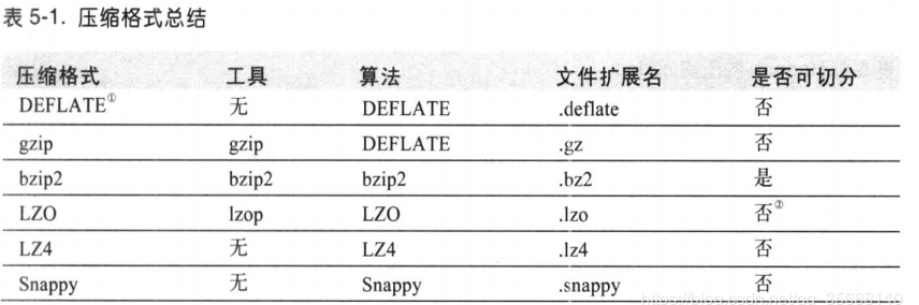
后台

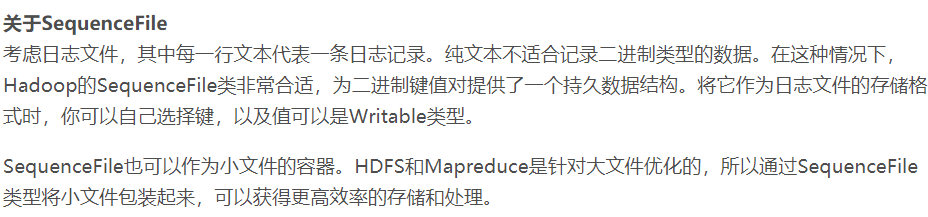


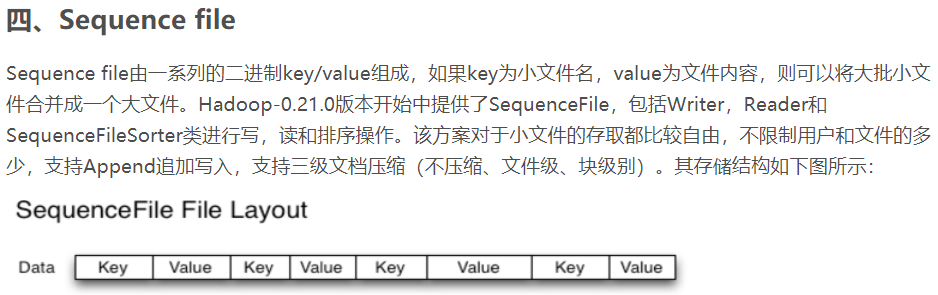


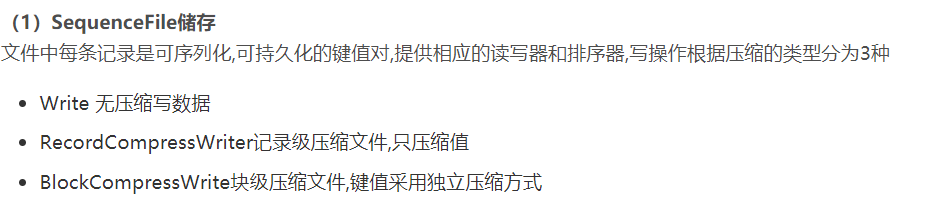


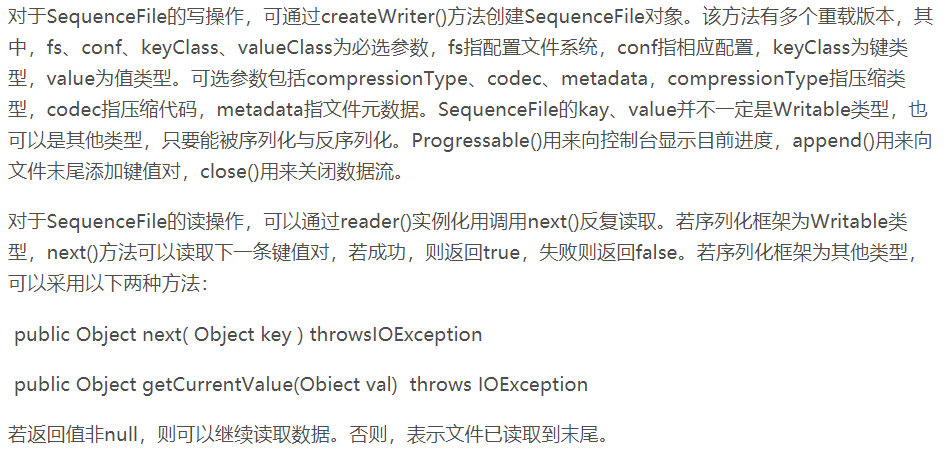


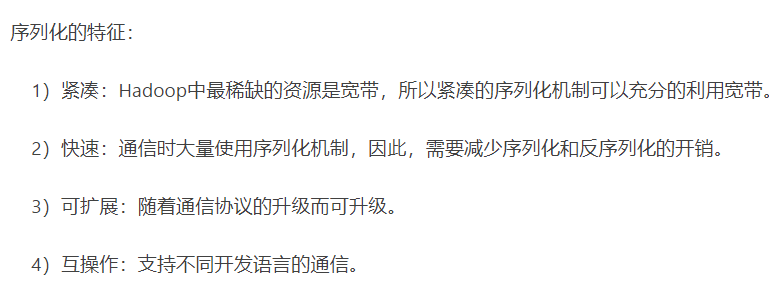
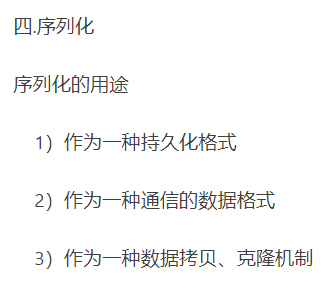






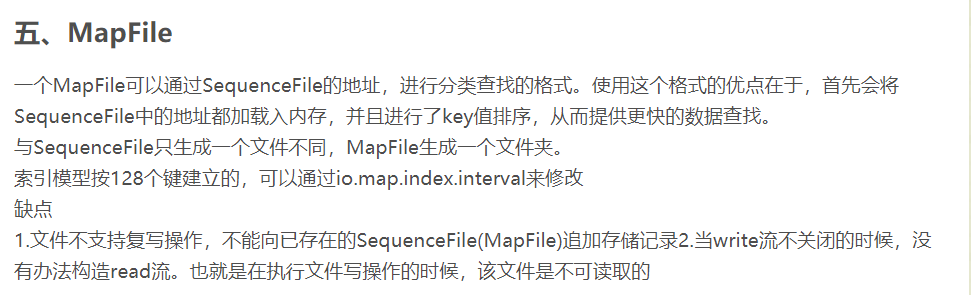


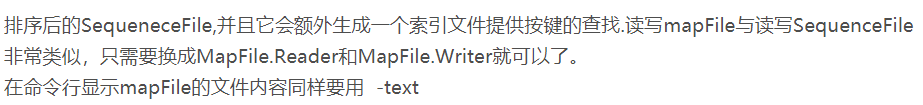


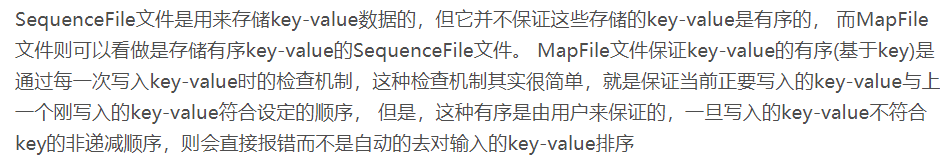




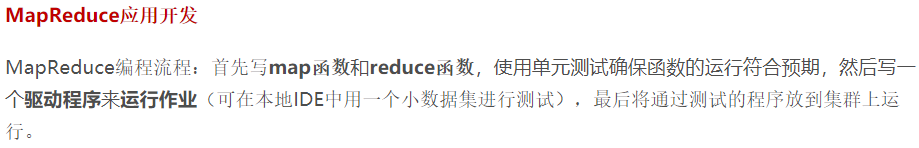


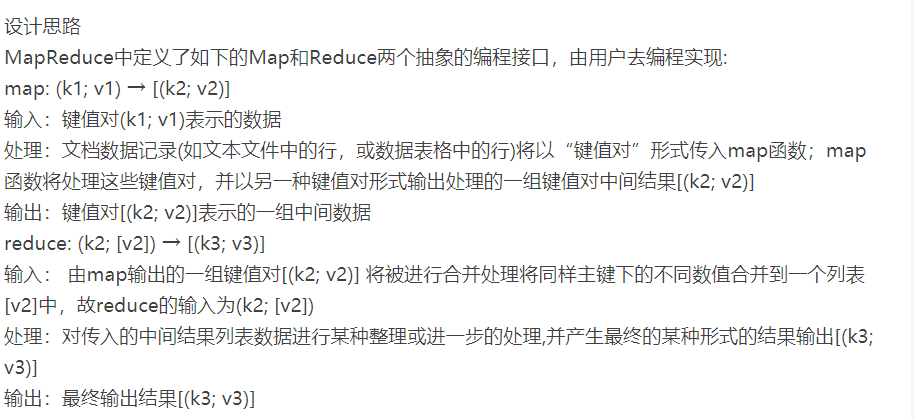


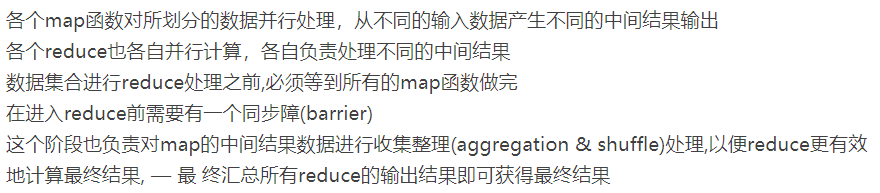


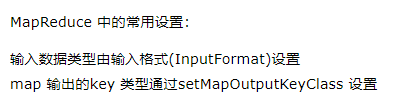


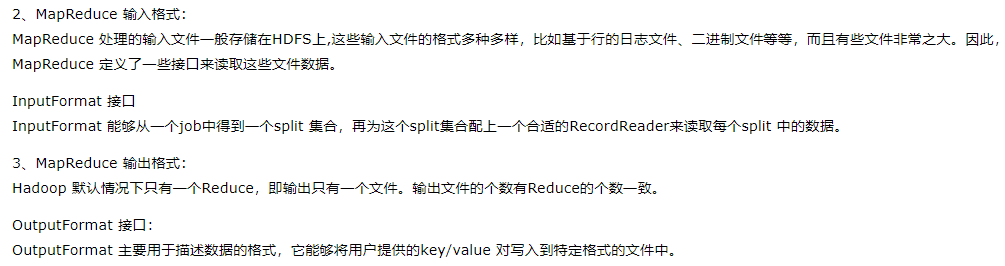
四、MapReduce 应用开发

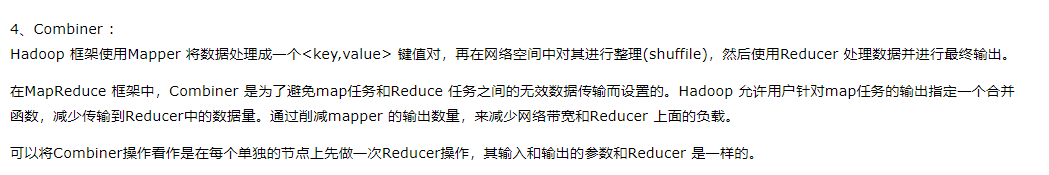


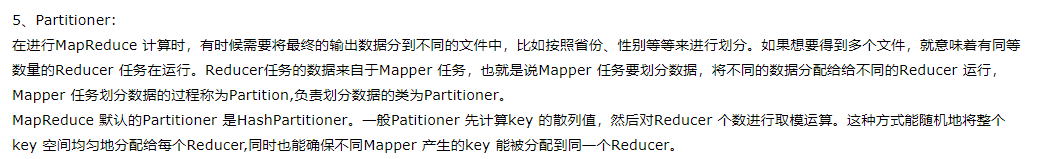


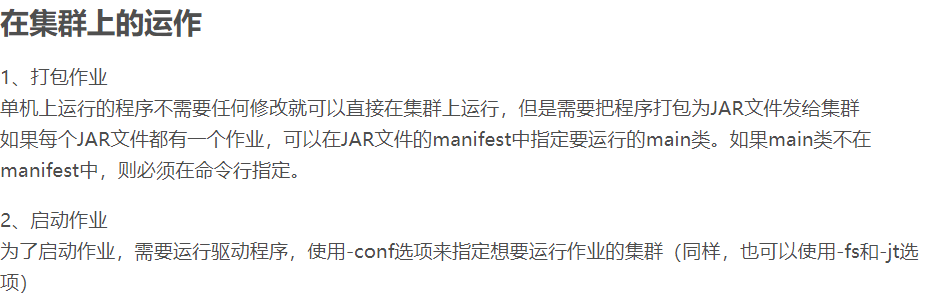






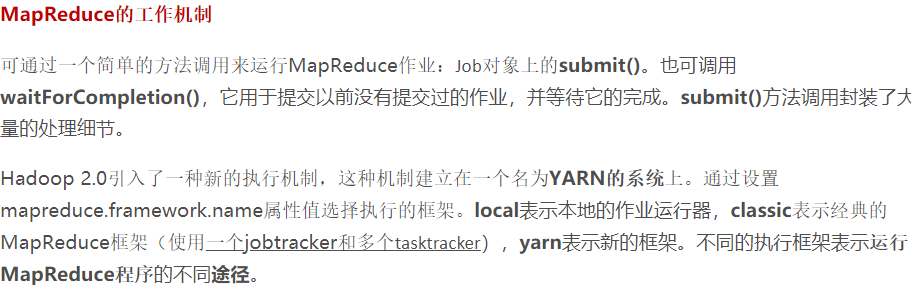


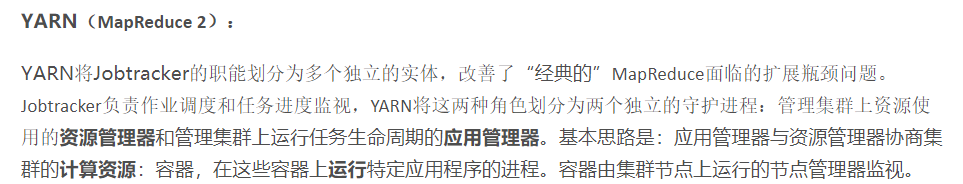




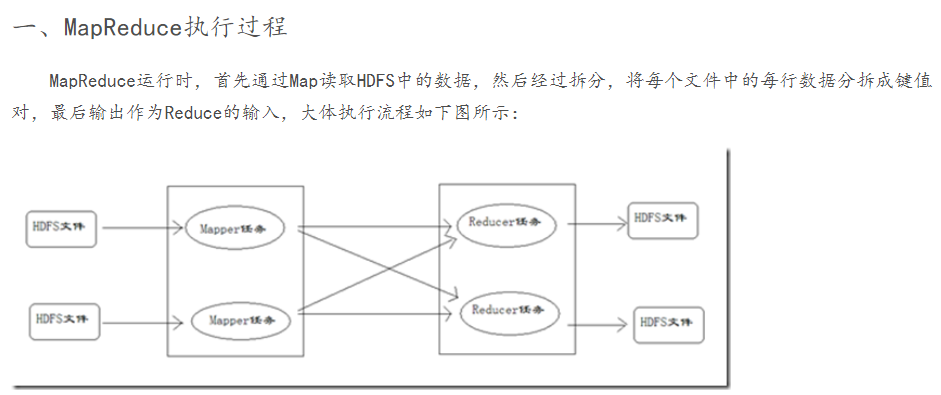
五、工作原理（ [MapReduce工作原理图文详解](https://blog.csdn.net/luzhensmart/article/details/90202313?ops_request_misc=%257B%2522request%255Fid%2522%253A%2522159309517419724846433023%2522%252C%2522scm%2522%253A%252220140713.130102334.pc%255Fall.%2522%257D&request_id=159309517419724846433023&biz_id=0&utm_medium=distribute.pc_search_result.none-task-blog-2~all~first_rank_ecpm_v3~pc_rank_v2-3-90202313.first_rank_ecpm_v3_pc_rank_v2&utm_term=MapReduce+%E5%9F%BA%E6%9C%AC%E5%B7%A5%E4%BD%9C)）

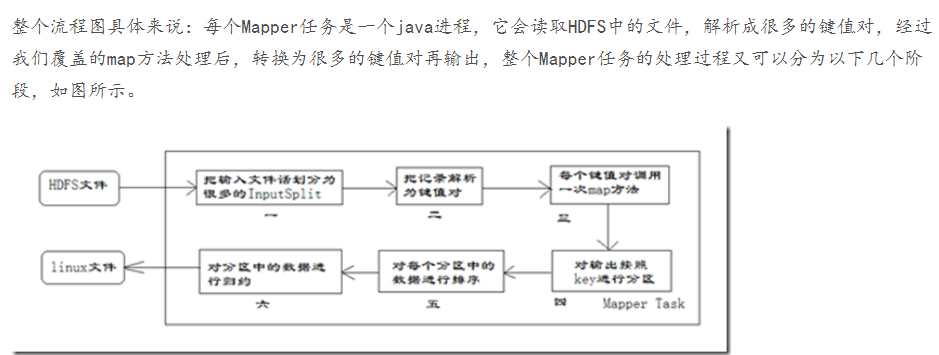
（<https://blog.csdn.net/u010185220/article/details/80933892> ）



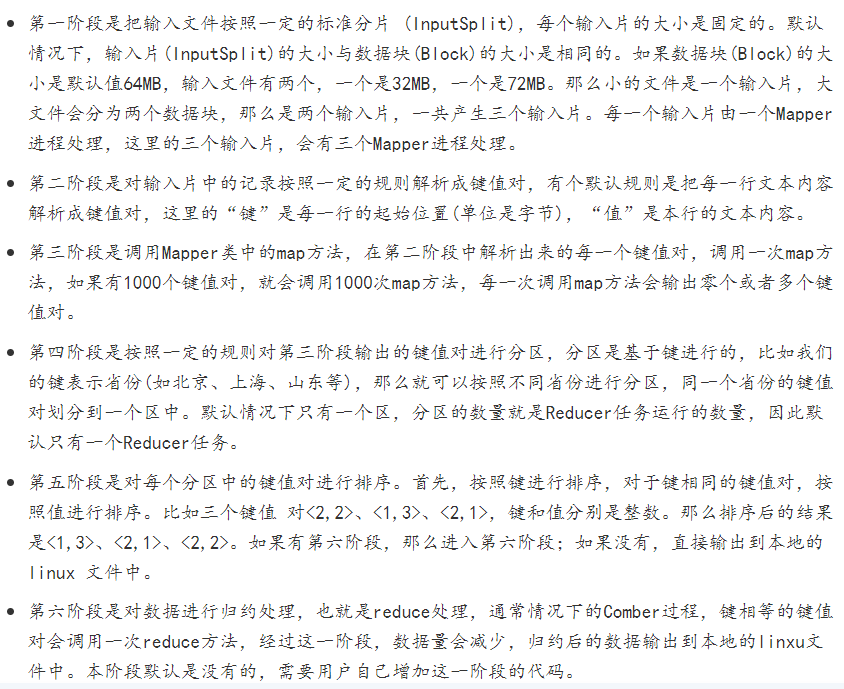


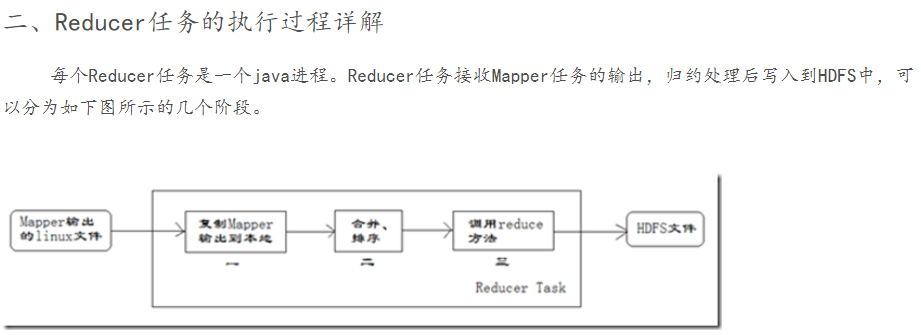
（weixin\_30652491  [Hadoop学习之Mapreduce执行过程详解](https://blog.csdn.net/weixin_30566063/article/details/96089911?ops_request_misc=%257B%2522request%255Fid%2522%253A%2522159309522819195188460340%2522%252C%2522scm%2522%253A%252220140713.130102334..%2522%257D&request_id=159309522819195188460340&biz_id=0&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduend~default-1-96089911.first_rank_ecpm_v3_pc_rank_v2&utm_term=+Hadoop%E5%AD%A6%E4%B9%A0%E4%B9%8BMapreduce%E6%89%A7%E8%A1%8C%E8%BF%87%E7%A8%8B%E8%AF%A6%E8%A7%A3)）

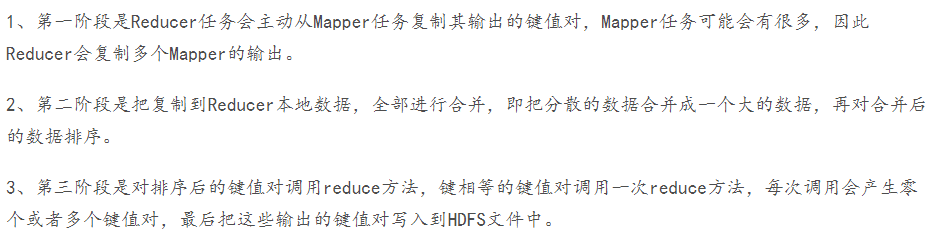




Mapper







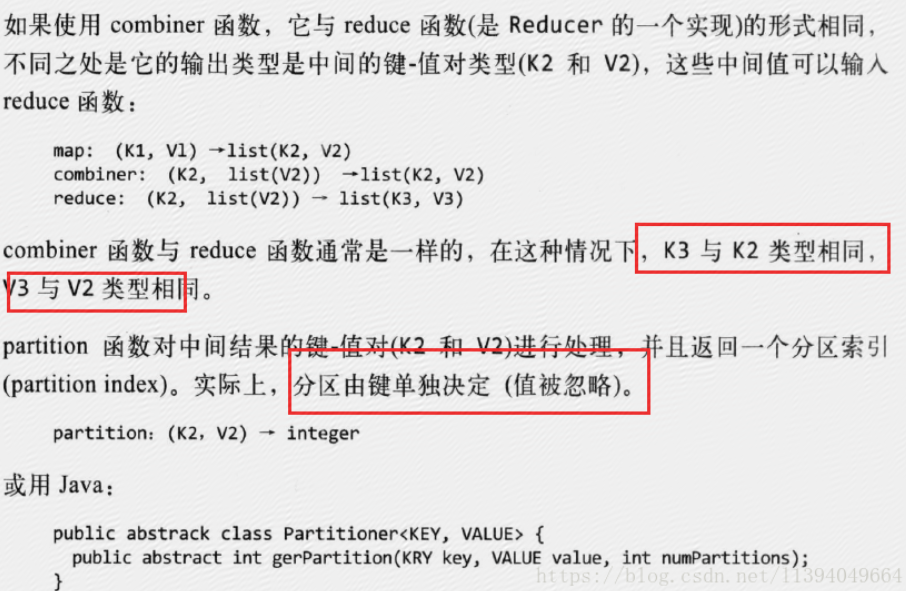
（fanxin\_i [MapReduce的基本工作原理](https://blog.csdn.net/fanxin_i/article/details/80388221?ops_request_misc=%257B%2522request%255Fid%2522%253A%2522159309525619725247654579%2522%252C%2522scm%2522%253A%252220140713.130102334..%2522%257D&request_id=159309525619725247654579&biz_id=0&utm_medium=distribute.pc_search_result.none-task-blog-2~all~first_rank_ecpm_v3~pc_rank_v2-5-80388221.first_rank_ecpm_v3_pc_rank_v2&utm_term=MapReduce%E7%9A%84%E5%9F%BA%E6%9C%AC%E5%B7%A5%E4%BD%9C%E5%8E%9F%E7%90%86)）

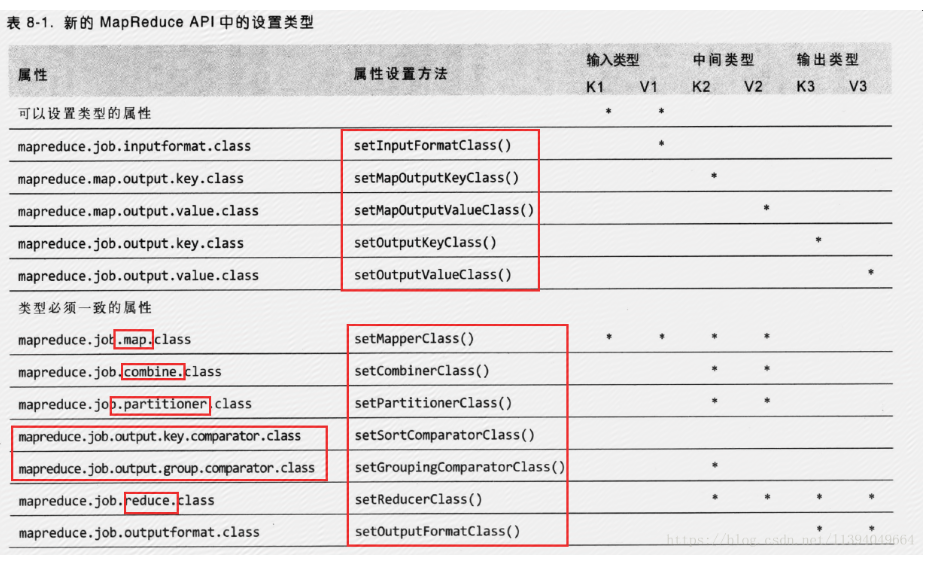
MapReduce的基本模型和处理思想： 1.如果对付大数据处理：分而治之，对相互之间不具有计算依赖关系的大数据，实现并行最自然的办法就是采取分而治之的策略。  2.上升到抽象模型：Mapper与Reduce，MPI等并行计算方法缺少高层并行编程模型，程序员需要自行指定存储，计算，分发等任务，为了克服这一缺陷，MapReduce借鉴了Lisp函数式语言中的思想，用Map和Reduce两个函数提供了高层的并发编程模型抽象。  3.上升到架构：统一架构，为程序员隐藏系统层细节，       MPI等并行计算方法缺少统一的计算框架支持，程序员需要考虑数据存储、划分、分发、结果收集、错误恢复等诸多细节；为此,MapReduce设计并提供了同意的计算框架，为程序员隐藏了绝大多数系统层面的处理系统。

六、类型格式

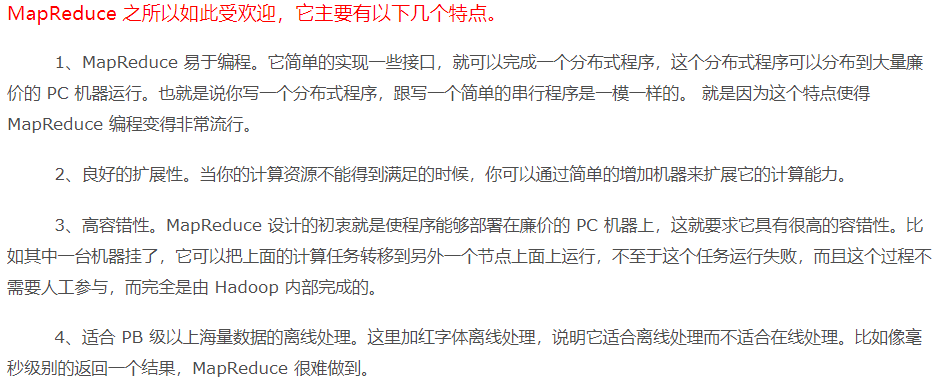
（[哎呦、不错哦 Hadoop权威指南---MapReduce的类型与格式](https://blog.csdn.net/l1394049664/article/details/82719648?utm_medium=distribute.pc_relevant.none-task-blog-BlogCommendFromBaidu-8.nonecase&depth_1-utm_source=distribute.pc_relevant.none-task-blog-BlogCommendFromBaidu-8.nonecase)）





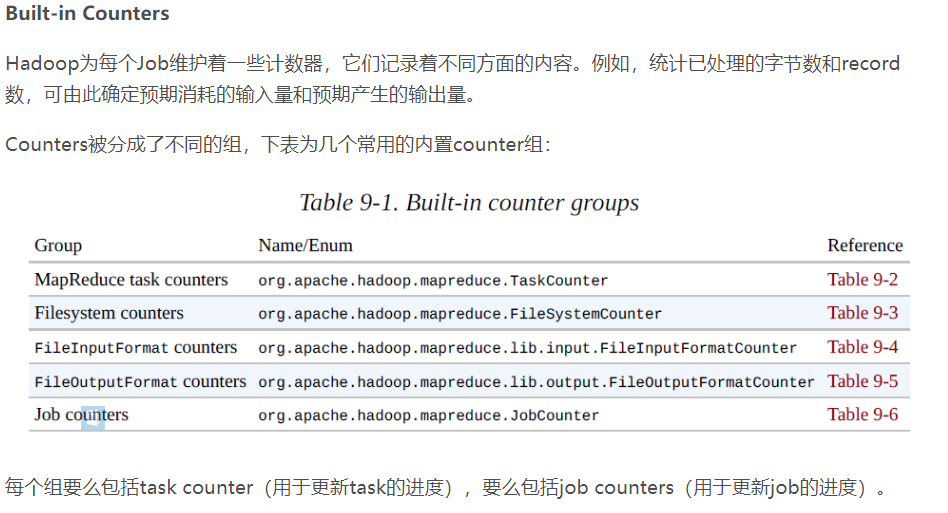


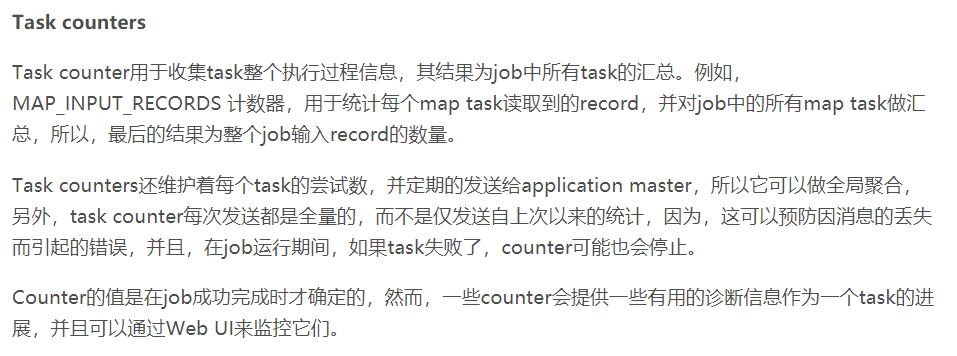
七、特性——计数 排序



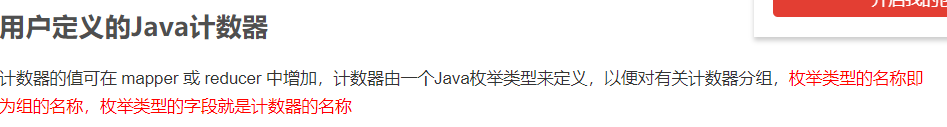


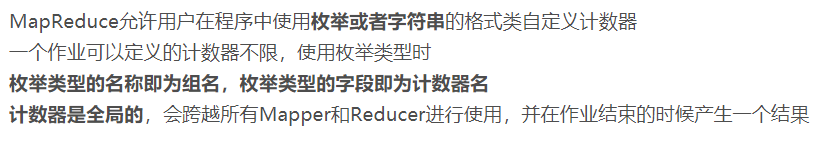


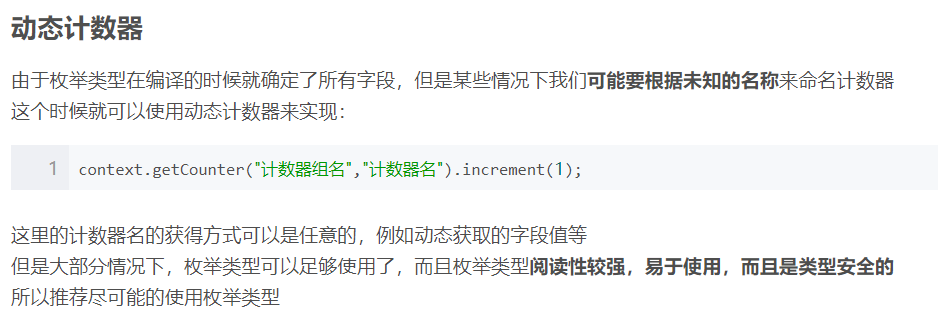


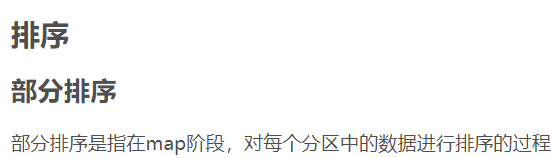




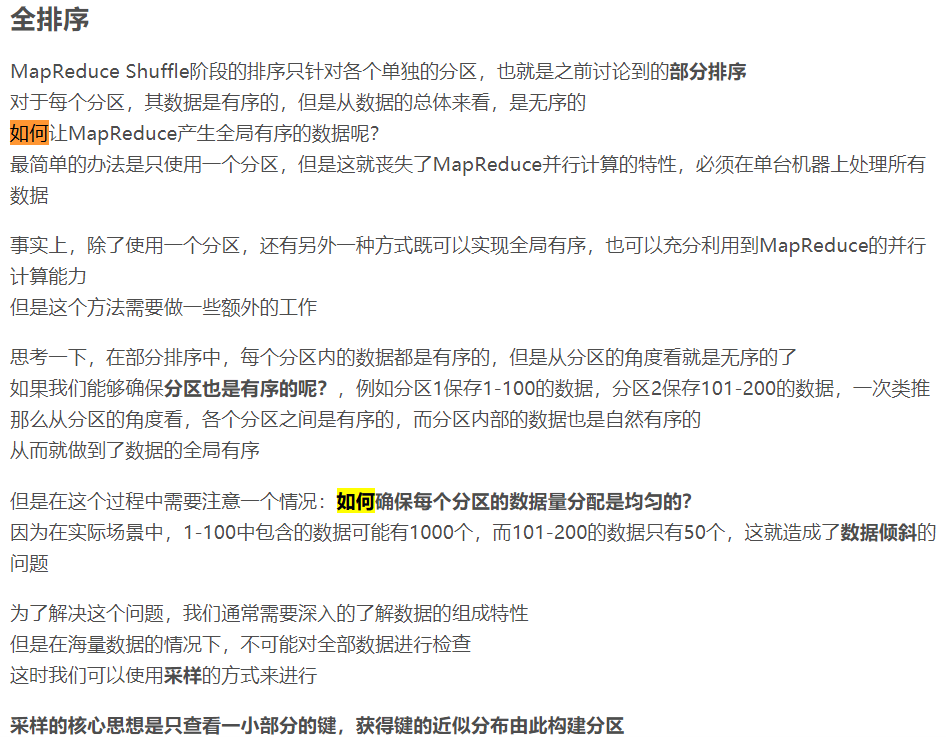


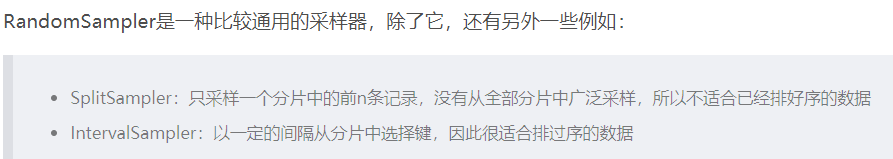




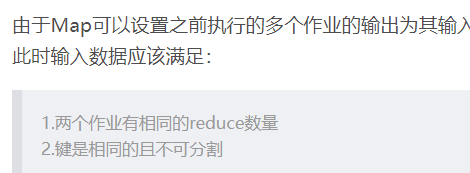


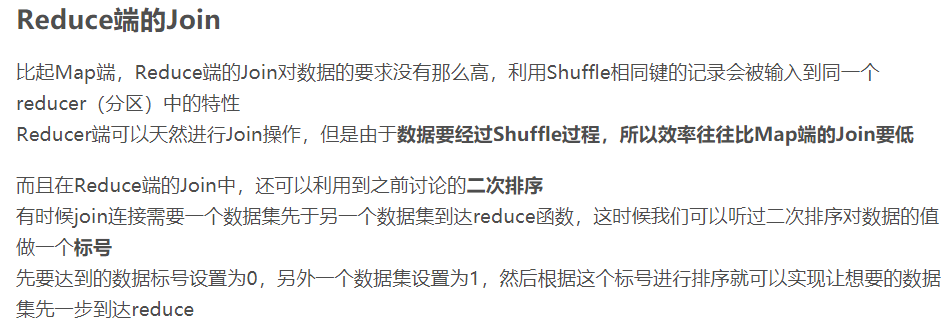


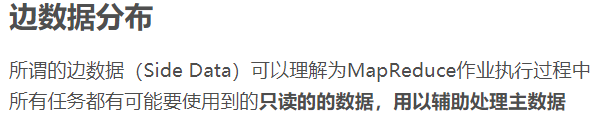












（示例代码详见 [MapReduce高级特性](https://blog.csdn.net/qq1010885678/article/details/51366342?utm_source=blogxgwz7?utm_medium=distribute.pc_relevant.none-task-blog-baidujs-5) 及 [MapReduce 的特性](https://blog.csdn.net/liu1390910/article/details/79199339)）

八、集群

（三种模式 <https://blog.csdn.net/qq_33890819/article/details/80945016> ）

问题：secondarynamenode未启动，因为hdfs-site.xml配置错误。

Clone 2个虚拟机 修改主机名，映射，设置无密码ssh登录，结果如下

