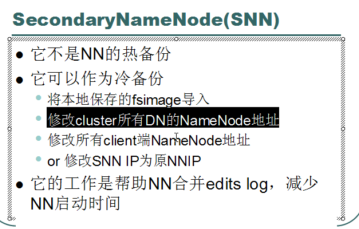
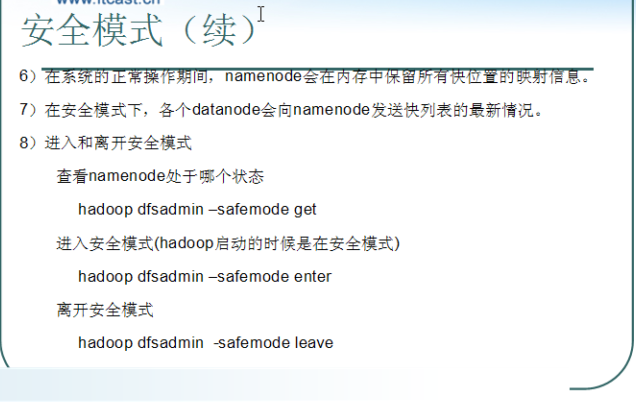
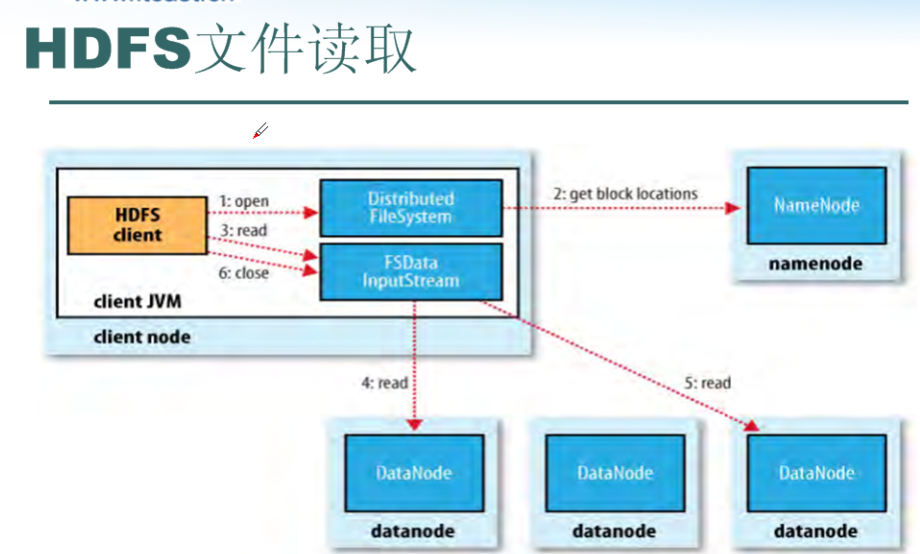
Hadoop组件依赖



SecondaryNamenODE(SNN):









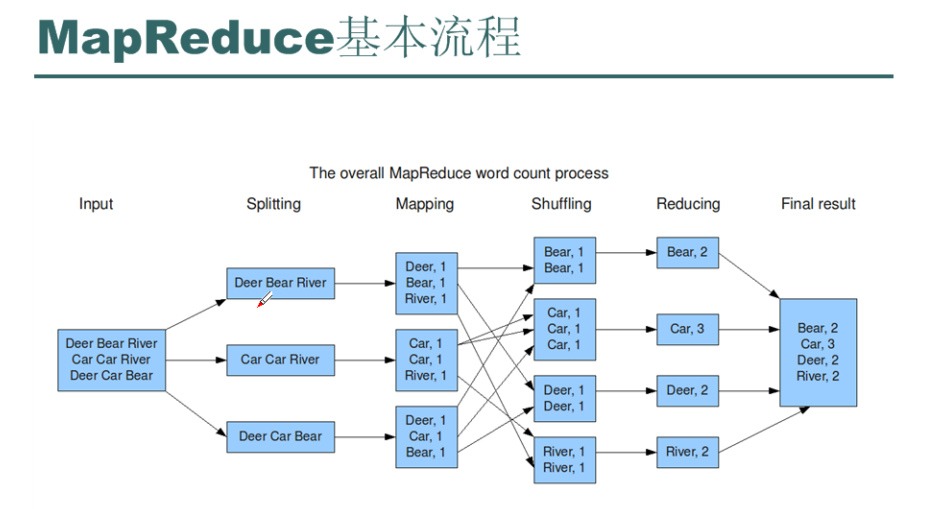


<http://www.aboutyun.com/blog-61-22.html>

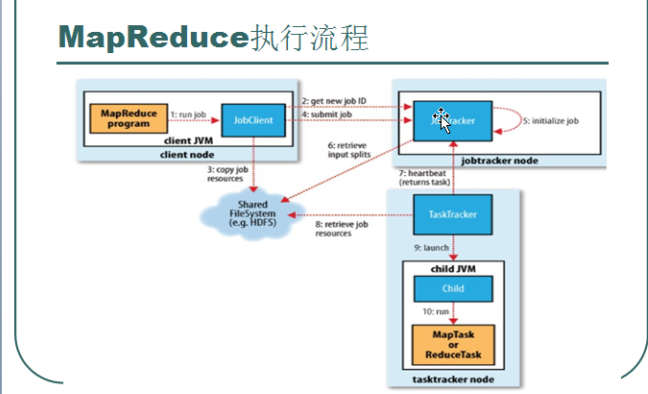


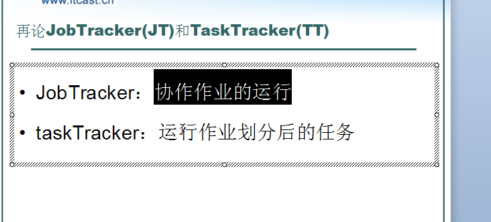


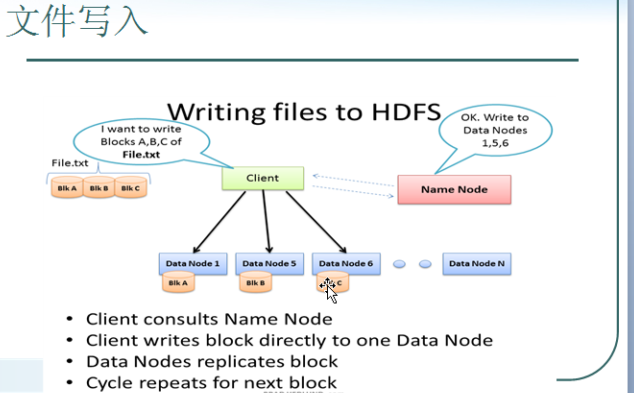




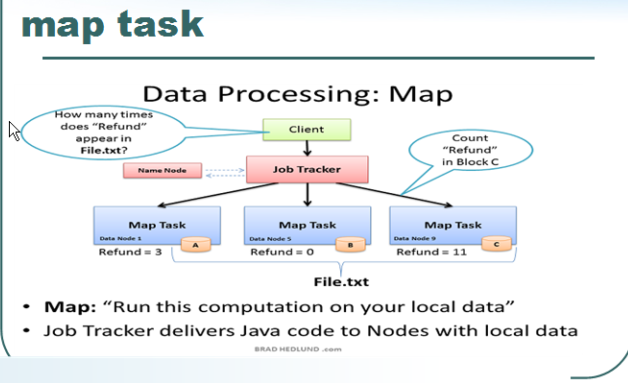
Mapreduce执行流程

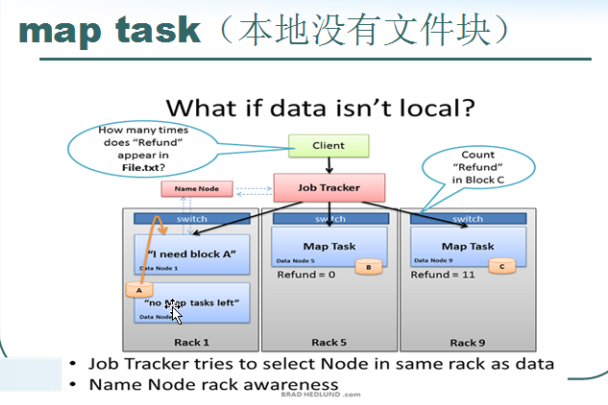




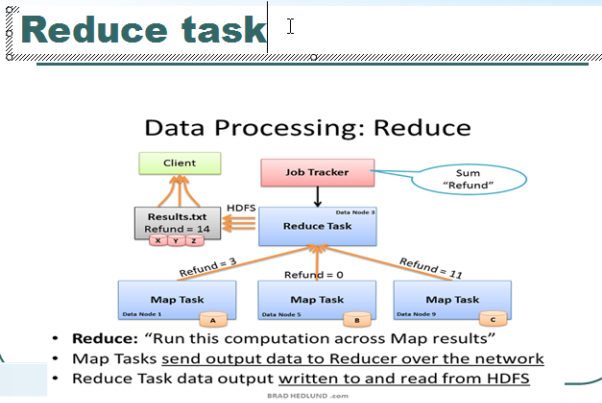


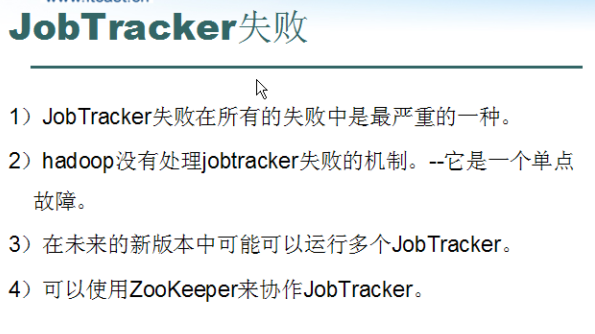
文件读取



* 

Reduce：







Hadoop能解决的问题：

海量数据需要及时分析和处理

HDFS:

文件以块方式存储

每个块带下远比多数文件系统来的大（预设64M）

NameNode主要功能提供名称查询服务，它是一个jetty服务器

Namenode保存metadate信息包括：

文件owership和permissions

文件包含哪些块

Block保存在哪个DataNode（由datanode启动时上报）

Namenode的metadate信息在启动后会加载到内存

Metadata存储到磁盘文件名为“fsimage”

Block的位置信息不会保存到fsimage

Datanode

保存block

启动dn线程的时候会向NN汇报block信息

通过向NN发送心跳保持与其联系（3秒一次），如果NN10分组没有收到DN的心跳

，则认为其已经lost，并copy其上的block到其他DN.

Block的副本放置策略

第一个副本：放置在上传文件的DN

如果是集群外提交，则随机挑选一台磁盘不太满，cpu不太忙的节点；

第二个副本：放置在第一个副本不同的机架的节点上

第三个副本：与第二个副本相同集群的节点

更多副本：随机节点

数据损坏处理:

当DN读取block的时候，他会计算checksum

如果计算后的checksum，与block创建时值不一样，说明该block已经损坏

Client读取其他DN上的block，NN标记该块已经损坏，然后复制block达到预期设置的文件备份数；

DN在其文件创建后三周验证其checksum

SecondaryNameNode(SNN)

它不是NN的热备份

它可以作为冷备份：

将本地保存的fsimage导入

修改cluster所有DN的NameNode地址

修改所有client端NameNode地址

修改SNN ip为元NNIP

帮助NN合并edits log，减少NN启动时间

打包jar测试步骤

:

Cp xxx/mul.jar /opt/modules/Hadoop/Hadoop/

Cd /opt/modules/Hadoop/Hadoop/

Chown –R Hadoop multi.jar

Chgrp –R Hadoop mul.jar

Su –Hadoop

新建文件夹 :Hadoop fs –mkdir /user/Hadoop/20160621

上传文件：

Hadoop fs –put word.txt /user/Hadoop/20160621

执行：进入hadoop-1.0.3目录下 ：

Hadoop jar mul.jar 包名.类名 /user/Hadoop/20160621