问题：1：搭建hdfs集群的时候，NN和DN这两个进程会挂掉：nn没有格式化，需要format一下；假如是dn的话，前后两次启动的时候namespace不一样（具体查看log文件，查看异常信息），相关步骤：1：rm –rf 本地存储目录（/tmp/hadoop-root）；2hadoop namenode –format； 3 执行start-dfs.sh。

:2：Dfsadmin –setQuota 限制文件数量；

Dfsadmin –setSpaceQuota 限制磁盘空间

:3：什么样的文件算是小文件？

数据块默认是64M，如果一个文件大小小于64M那么它也要占用一个数据块，使用Archive的方式归并小文件，数据块的大小使用dfs.block.size这个属性进行配置

:

4：Start-dfs.sh的警告信息：unable to load native-hadoop library for you platform, using built-java class 没有找到native库，使用内置的java code

5：重复运行命令行的wordcount会提示目录存在，可以先删除在使用；

6默认的hadoop conf路径变成了/etc/hadoop，在执行start-dfs.sh前会source环境变量hadoop-config.sh，确定hadoop-env.sh存在，然后执行hadoop-deamon.Sh

1：hdfs架构

2：secondarynamedoe

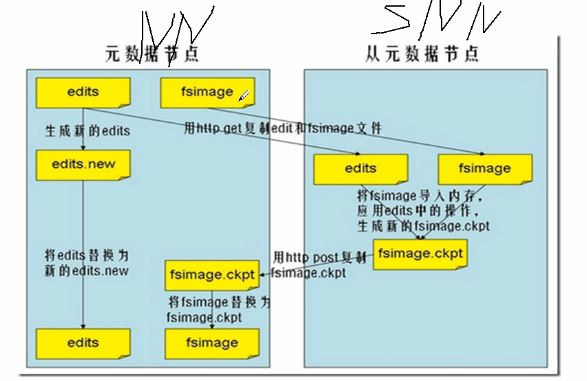
3：

常见的分布式文件系统：

FastDFS（以文件作存储）：难实现并行化处理，难实现负载均衡；Moosefs支持fuse；

Hdfs架构：namenode，datanode，适合大数据处理，批处理，高容错性，不适合小文件存储，不支持并发写入；

Secondarynamenode：不是namenode的备份，他只是周期性地合并fsimage，edits文件，推送给namenode，他可以帮助恢复namenode，随着hadoop2.0，现在出现为checkpoint node，backup node：



在设置secondarynamenode的时候我们只需要在core-site.xml中设置fs.checkpoint.period, fs.checkpoint.dir, fs.check.size，