hdfs的实现思想：

1）hdfs是通过分布式集群来存储文件的，为客户端提供了一个便捷的访问方式，就是一个虚拟的目录结构。

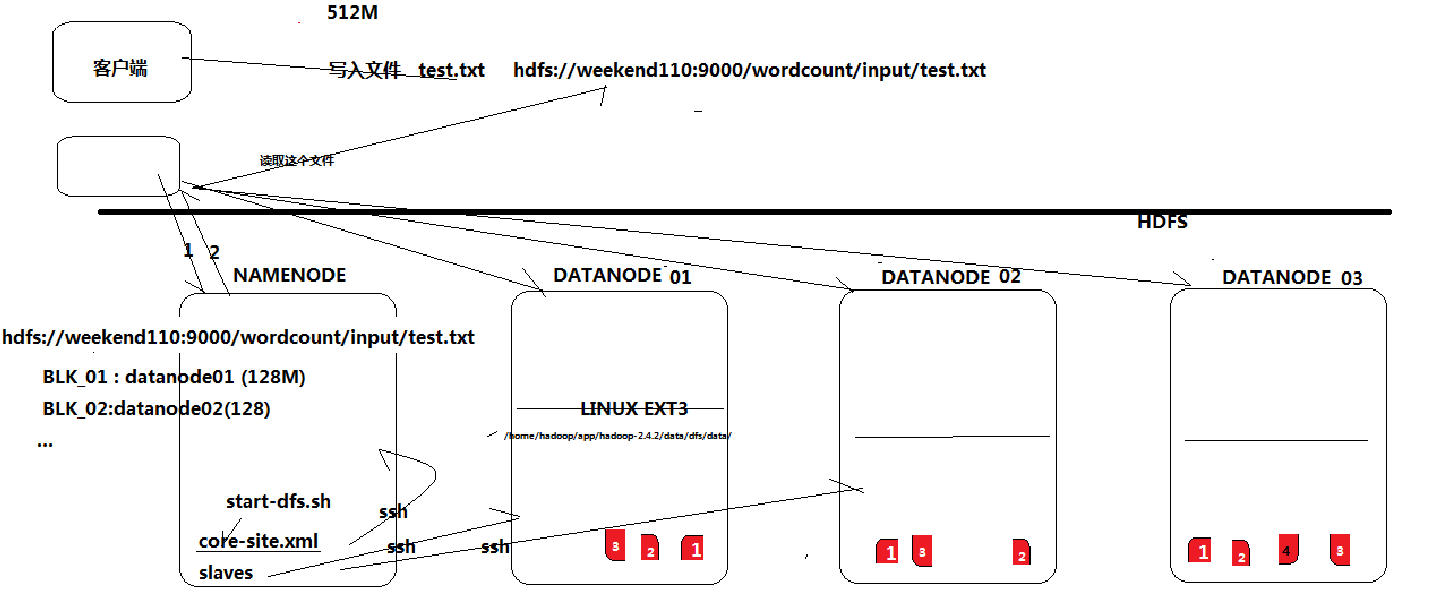
2）文件存储到hdfs集群中去的时候是被切分成block

3）文件的block存放在若干台datanode节点上

4）hdfs文件系统中的文件与真实的block之间有映射关系，由namenode管理

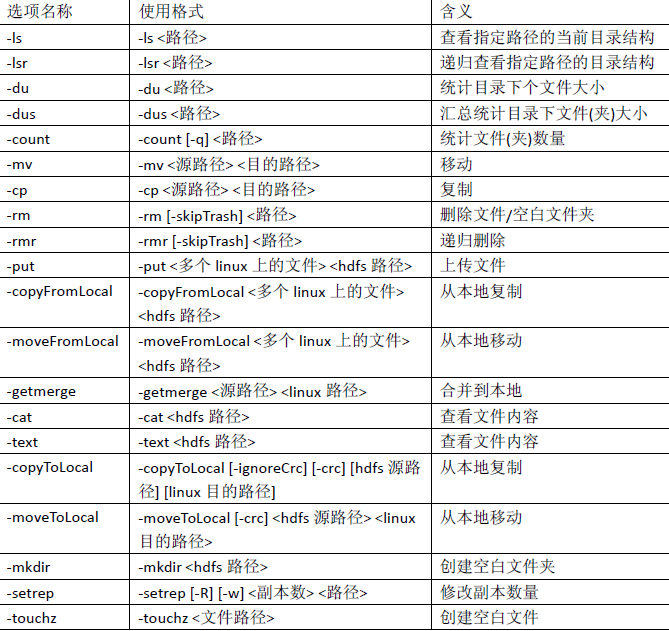
5）每一个block在集群中会存储多个副本，好处是可以提高数据的可靠性，还可以提高访问的吞吐量/并发能力

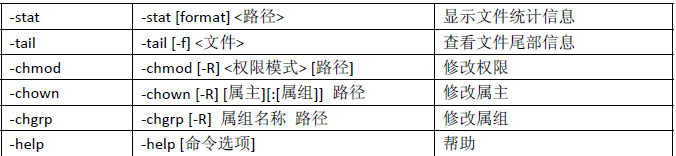
hsfs实现机制



常用HDFS Shell操作：

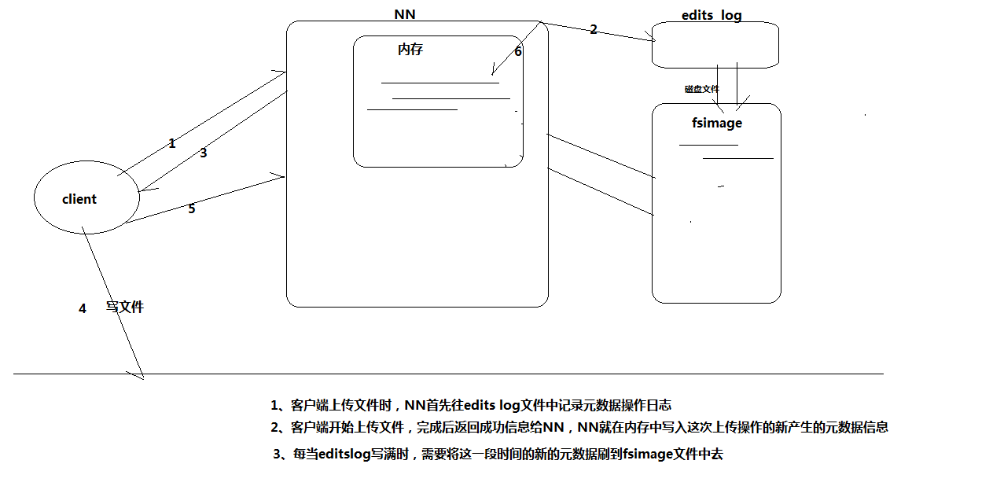
hadoop fs 操作文件（filesystem）



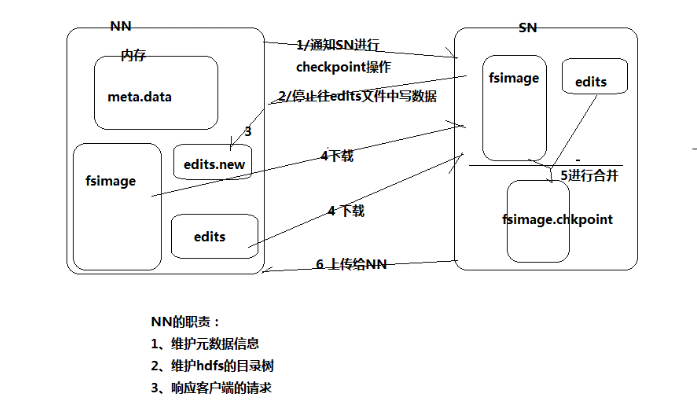


\*\*\*hadoop fs -du -s -h hdfs://hadoop-0:9000/\*

HDFS架构：NameNode，DateNode，Secondary NameNode



NN内存中是最全的。edits log是最新的。fsimage就是内存文件的镜像。



NN与SN进行交互维护数据

Chkpoint触发条件：edits超过默认64m，超过最大时间默认3600s

