# HDFS

问：这里是直接在hdfs里面对文件追加内容，如果是客户端那边对文件追加内容，再重新上传文件，那么整个文件内容都会存储一遍吧

答：不会的；

你想如果一个文件有1000个block，我如果要追加1kb的内容的话，还需要将1000个block重新传输一遍，hadoop作为一款成熟的框架，肯定不会这么做的

client先获得这个文件最后一个块blk的信息，返回这个blk存储在哪三个dn上，然后建立管道pipeline，接下来，客户端将追加的内容沿着管道依次追加到这三个副本中

至于文件的其他block，不会受到影响

问：@老游～讲师 他是怎么知道是追加的还是修改的修改的话也只替换修改部分的block吗

答：我们说，hdfs不支持文件的随机修改，它只支持一种修改，那就是向文件末尾追加内容

问：老师。snn和nn 为什么要保存 旧的 edit和fsimage

答：应该也是出于数据安全、容错考虑；万一集群出了问题，我还能通过过去的edit、fsimage恢复部分的数据，减少损失；要知道企业中的数据，那就是它的最宝贵的资源；

问：最新的fsimage是所有的元数据,还是最新一个小时或1000命令的,它的后缀编号是代表这一段事务的数据吗@老游～讲师

答：最新的fsiamge，其中包含的除了如正在写入日志文件（edits\_inprogress\_0000000000000000332）外的的所有的元数据

fsimage\_0000000000000000331 比如最新的fsimage名字是它，那么表示它里边包含了1到331所有事务下，所产生的元数据

咱们的集群fsimage存储目录是

/kkb/install/hadoop-2.6.0-cdh5.14.2/hadoopDatas/namenodeDatas/current

是在hdfs-site.xml中的如下属性决定的



Edits目录是在

/kkb/install/hadoop-2.6.0-cdh5.14.2/hadoopDatas/dfs/nn/edits/current

是由hdfs-site.xml中另外一个属性，如下决定的

