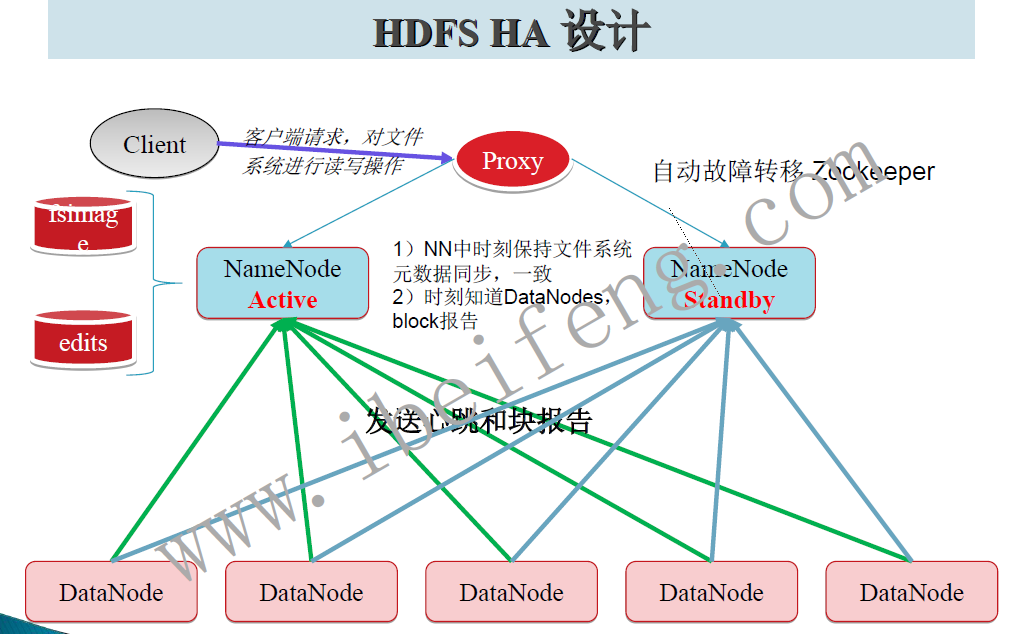
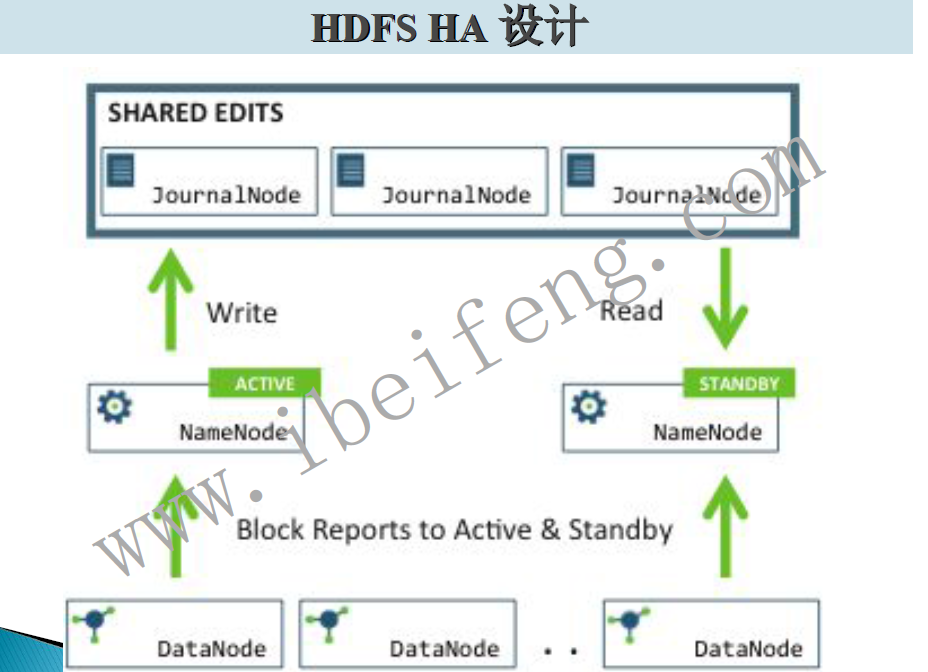
NameNode主要在一下两个方面影响HDFS集群

* NameNode机器发生意外,如宕机,集群将无法使用,知道管理员重启
* NameNode机器需要升级,包括软件,硬件升级,此时集群也将无法使用





在只要一个NameNode的情况下,容易出现单点故障,首先NameNode在启动的时候会读取fsimage和edits的文件,然后载入内存中,SecondaryName是辅助NameNode工作的,他定期的合并fsimage和edits,这个样子可以防止下一次NameNode重启启动的时候合并fsimage和edits花费时间过长。如果想做HA，即我们可以增加一个NameNode，当然可以更多，这里就只增加一个，NameNode在启动以后，如果文件系统的内容发生变化的时候，他会在内存中进行更新元数据信息，并且把相关操作记录在edits上面，edits在磁盘上，所以内存中保存的元数据信息等于磁盘上面fsimage和edits的信息和，想要做HA，就是要不断的copy NameNode的元数据信息，怎么拷贝呢？那就是首先获得fsimage信息，但是要想实现和NameNode的实时同步，那么最重要的就是要同步edits文件，edits就是Journal（日志），通过这个想法，如果想要做HA，就可以把active的NameNode的edits文件放到一个公共的地方，这个样子另外一个standy的NameNode就可以不断的获取这些edits来更新自己的元数据信息。所以在hadoop的ha中，edits信息放在了JournalNode进程中进行保存，JournalNode最少有3个，他是轻量级的，可以在NameNode上或者ResourceManager等，这个样子就实现了共享edits，当集群包含了两个NameNode，那么客户端怎么连接上呢？如上图所示，提供一个Proxy，这个代理用来决定用户连接的哪个NameNode（Standby），还有就是在同一个时刻，只能有一个NameNode对外提供服务，不能同时提供服务。所以需要隔离（fence）他们。

在fence方面，有许多中方法，例如sshfence，zookeeper等，本次只介绍这两种，其中sshfence是冷备，他在active的namenode出现故障的时候，需要手动启动另外的standby的namenode使它变成active。所以很不方便。还有一种就是使用第三方zookeeper，他可以做到在active的namenode出现故障的时候，自动的切换的standby的namenode。

所以配置HA需要考虑这几个方面：

\* share edits

JournalNode

\* NameNode

Active，Standby

\* Client

Proxy

\* fence

同一时刻仅仅有一个NameNode对外提供服务

使用的方式sshfence

两个NameNode之间能够ssh无密码登录

131(NameNode) ssh -> 132

132(NameNode) ssh -> 131

根据上面的四个要点。我们来配置HA