为了配置一致性，HDFS使用NWR模型来保证数据一致性。其中N是保存数据副本的节点数量,W是在更新完成之前需要确认的收到更新的副本数量，R是通过一次读操作访问一个数据对象时要联系的副本数量。只需W+R>N，分布式系统就可以提供强一致性。（W，R自己都是可控制的， 所以我们可以控制HDFS的一致性策略）

整个系统的命名空间（数据块到文件映射+文件的属性）都存在fsimage中，对于元数据的更新都存在日志 editlog中

HDFS读取文件.emf

1. HDFS客户端与NN取得通信，获取开头块数据的地址
2. 客户端通过FSDataInputStream读取Datanode上的块数据
3. 数据返回到客户端
4. 联系NN读取下一个块数据
5. 读取完所有的块数据，关闭读取流

HDFS写文件.emf

1. 与NN取得通信
2. 在NN上建立命名空间
3. NN生成新文件，并返回文件输出流
4. FSDataOuputStream将文件分成packet，在DN间进行管道传输
5. 返回所有ACK，完成一个块的存取
6. 完成所有数据
7. 关闭写数据流