# HDFS读写文件流程

## 1、HDFS写流程：

客户端要向HDFS写数据，首先要跟namenode通信以确认可以写文件并获得接收文件block的datanode，然后，客户端按顺序将文件逐个block传递给相应datanode，并由接收到block的datanode负责向其他datanode复制block的副本

****如图：****   


****写详细步骤：****

1、根namenode通信请求上传文件，namenode检查目标文件是否已存在，父目录是否存在   
2、namenode返回是否可以上传   
3、client会先对文件进行切分，比如一个blok块128m，文件有300m就会被切分成3个块，一个128M、一个128M、一个44M请求第一个 block该传输到哪些datanode服务器上   
4、namenode返回datanode的服务器   
5、client请求一台datanode上传数据（本质上是一个RPC调用，建立pipeline），第一个datanode收到请求会继续调用第二个datanode，然后第二个调用第三个datanode，将整个pipeline建立完成，逐级返回客户端   
6、client开始往A上传第一个block（先从磁盘读取数据放到一个本地内存缓存），以packet为单位（一个packet为64kb），当然在写入的时候datanode会进行数据校验，它并不是通过一个packet进行一次校验而是以chunk为单位进行校验（512byte），第一台datanode收到一个packet就会传给第二台，第二台传给第三台；第一台每传一个packet会放入一个应答队列等待应答   
7、当一个block传输完成之后，client再次请求namenode上传第二个block的服务器。

## 2、HDFS读流程:

客户端将要读取的文件路径发送给namenode，namenode获取文件的元信息（主要是block的存放位置信息）返回给客户端，客户端根据返回的信息找到相应datanode逐个获取文件的block并在客户端本地进行数据追加合并从而获得整个文件

****如图****



****读详细步骤：****

1、跟namenode通信查询元数据（block所在的datanode节点），找到文件块所在的datanode服务器   
2、挑选一台datanode（就近原则，然后随机）服务器，请求建立socket流   
3、datanode开始发送数据（从磁盘里面读取数据放入流，以packet为单位来做校验）   
4、客户端以packet为单位接收，先在本地缓存，然后写入目标文件，后面的block块就相当于是append到前面的block块最后合成最终需要的文件。