在现在企业中，数据都特别大，单机是无法存储所有数据，所以需要把数据放在多台机器上，管理多台机器的数据的系统就叫做分布式文件系统。HDFS就是一个分布式文件系统。

HDFS适合存储非常大的文件，比如几百M级，G级，TB级等。

HDFS适合采用流式数据访问方式。数据可以分为批式数据和流式数据，把数据比作商品，批式数据就是屯下来的商品，流式数据就是刚刚进来的商品，即是实时的。流式数据处理方式可以理解为来一点处理一点。

HDFS运行于商业硬件上

对延时要求在毫秒级的应用不适合HDFS

大量的小文件，文件的元数据是保存在内存中，用nameNode管理。文件的数量受到nameNode内存的限制。

物理磁盘有block的概念，文件的读写操作是一block为单位，一般是512字节。文件系统的block是对物理block的抽象，一般为几kb。HDFS的block默认是128M,但文件小于128M时，只会占用实际大小。

NameNode负责管理文件系统树和文件和目录的元数据（数据的属性，存储位置等）

DataNode负责实际存储数据，负责读写操作，DataNode通常直接从磁盘中取数据，但如果是比较频繁的数据，它会从内存中直接存取。

NameNode的内存会限制文件的数目，HDFS Federation就是扩展NameNode的一种方式。

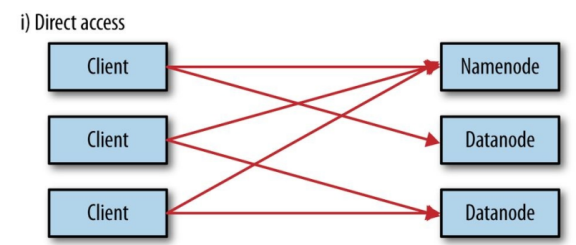
HDFS HA

配置两个NameNode,当一个挂时，另一个继续工作，用户没有明显的中段感觉。

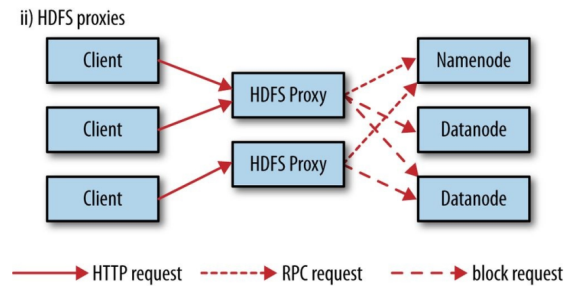
HDFS提供命令交互方式，包括把本地文件拷贝到HDFS，创建目录，列出文件列表等。HDFS的文件和目录有读、写、执行的权限。每个文件或目录都有ower（文件所有者）,group（权限组）,mode（所有人）属性，可以用来设置权限。

HDFS是Hadoop文件系统的一种实现，Local是对本地文件系统的抽象，HDFS是最常见有两种web形式，通过HTTP提供文件操作接口，这种交互方式不适合操作大文件。通过HTTP有两种访问方式，

一种是直接访问

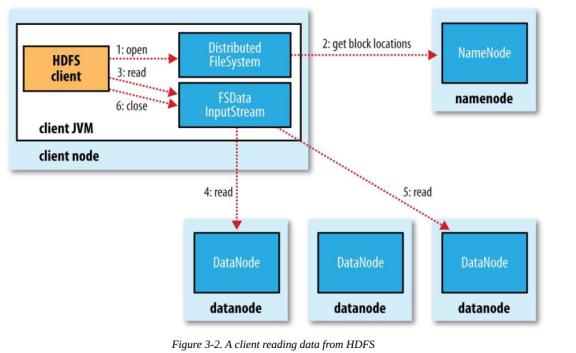


另一种是通过代理访问

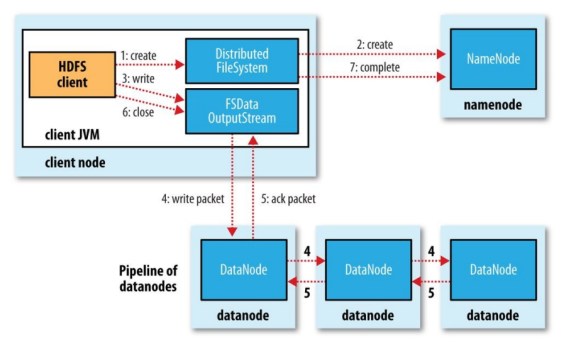


在实际应用中，对HDFS的操作实际上很多都是通过FileSystem（java接口）实现，FileSystem可以读、写数据，操作目录，删除数据等

HDFS读文件流程



HDFS写文件流程



HDFS一致性模型

描写的是文件读写操作的可见性，文件一但被创建，在文件系统的命名空间内可见，但写入文件的内容不保证可见。

Hdoop节点距离是指到达公共父节点所需要跳转数之和。

http://blog.csdn.net/bingduanlbd/article/details/51914550