

HDFS

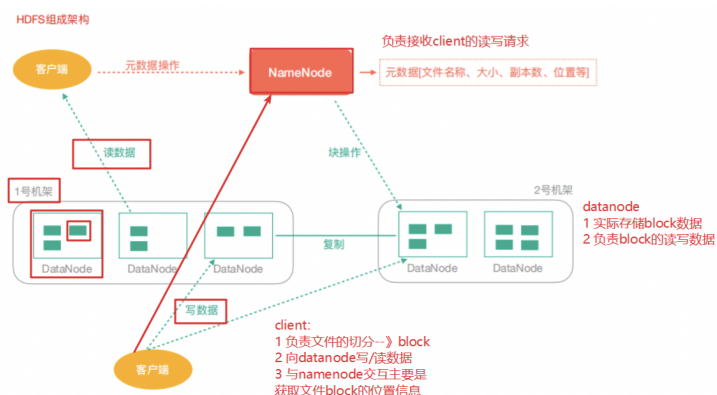
- HDFS（全称：Hadoop Distribute File System，Hadoop 分布式文件系统）是 Hadoop 核心组成，是分布式存储服务

- 重要概念

- 典型的Master/Slave架构
- 分块存储（block机制）
- 命名空间
- NameNode元数据管理
- DataNode数据存储
- 副本机制
- 一次写入，多次读出

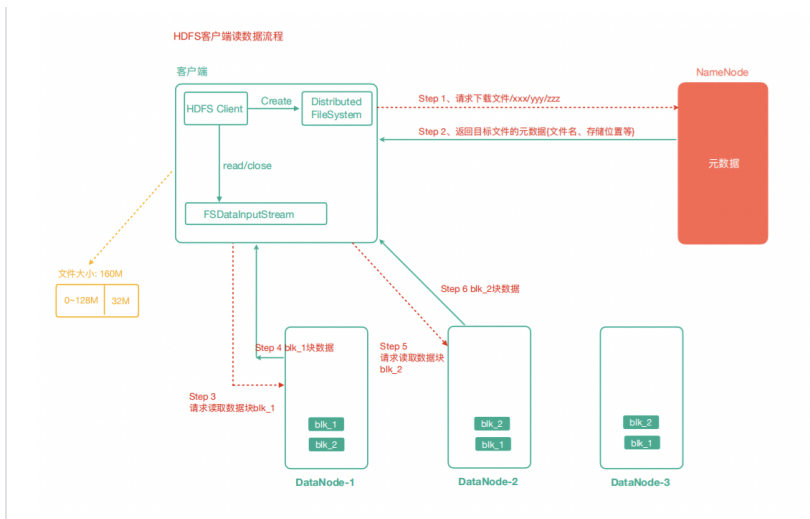
- 架构

- NameNode(NN)：HDFS集群的管理者，Master
 - 维护空间Hdfs的名称空间
 - 维护副本策略
 - 记录文件块（block)的映射信息
 - 负责处理客户端的读写信息
- DataNode(DN)：NameNode下达命令，DataNode实际执行（slave）
 - 保存实际的数据块
 - 负责数据块的读写
- Client:客户端
 - 上传文件的时候，client负责将文件切分成Block，然后进行上传
 - 请求NameNode交互，获取文件的位置信息
 - 读取或写入文件，与DataNode进行交互
 - 使用一些命令来管理或访问HDFS

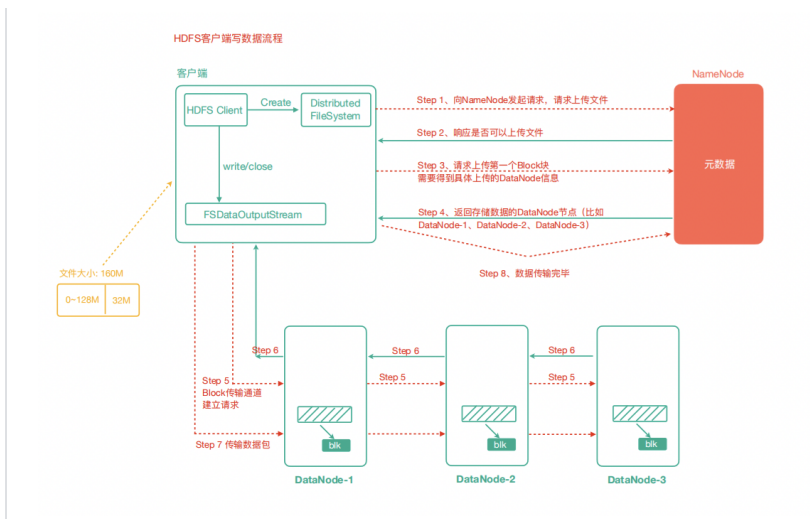


- 读写解析

- 读流程

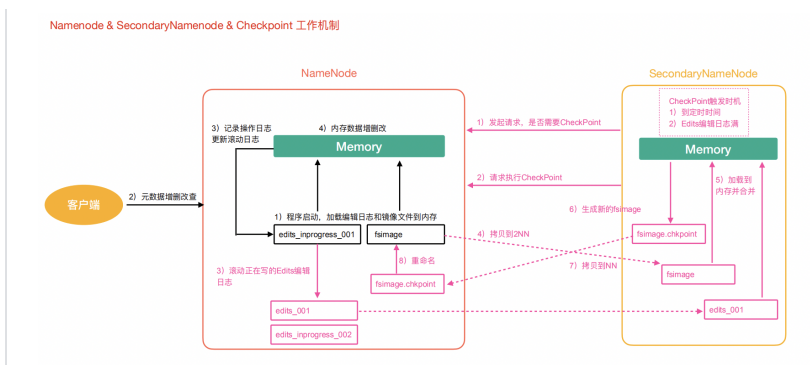


- 写流程

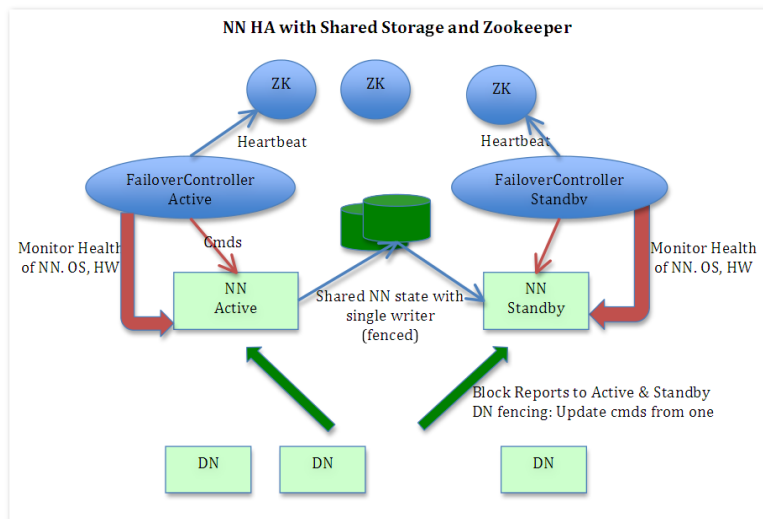


- 元数据管理流程

- 单节点:



- HA:



- checkPoint周期配置：