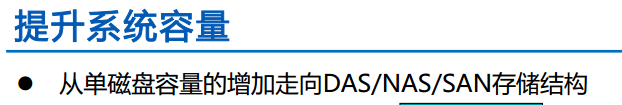




**存储：**

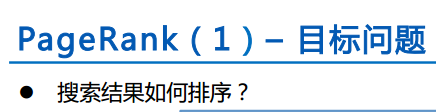


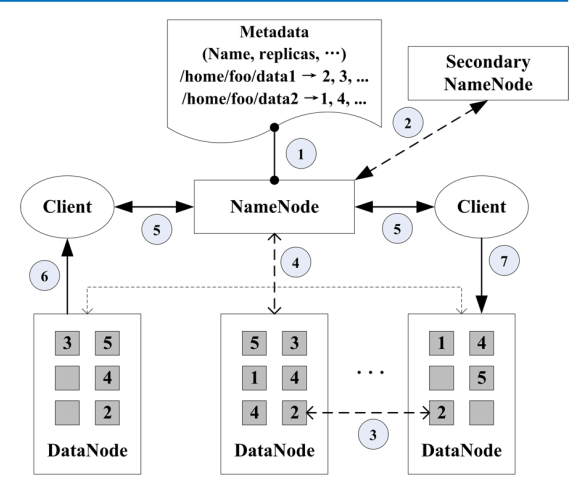


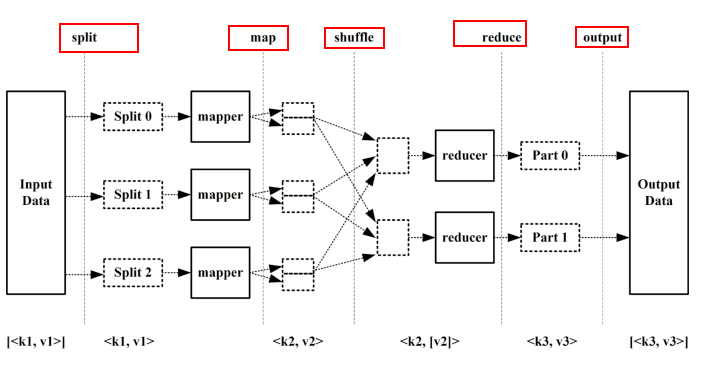
**计算：**

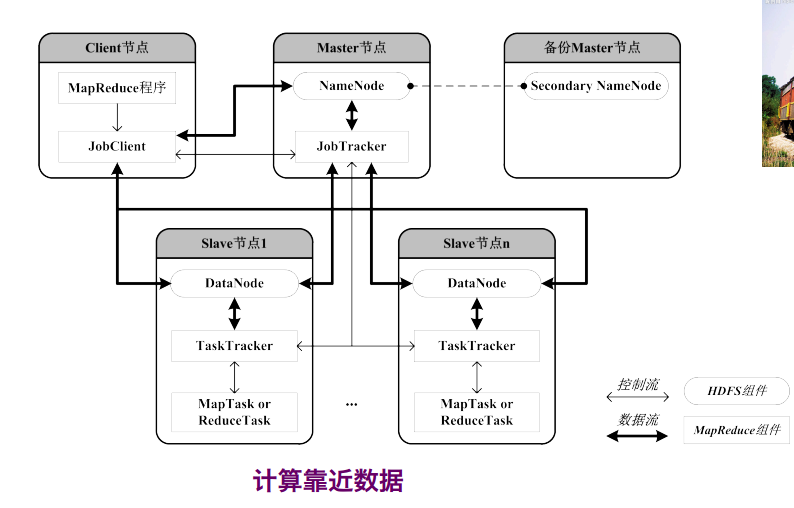


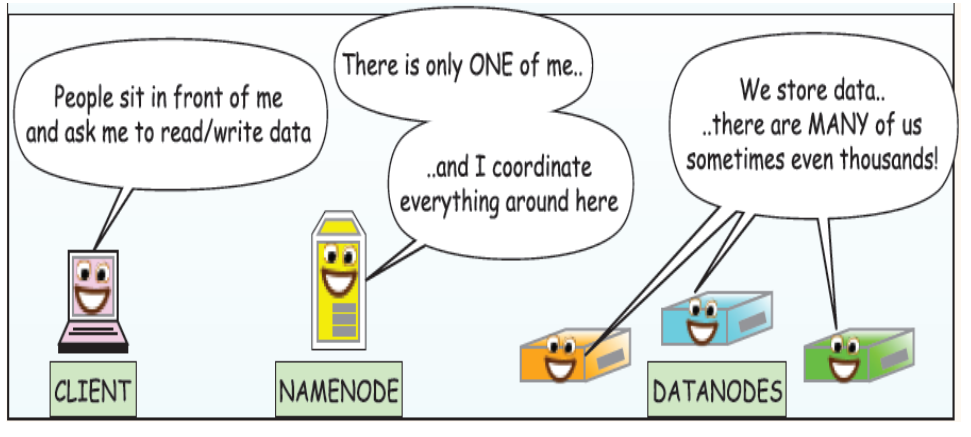


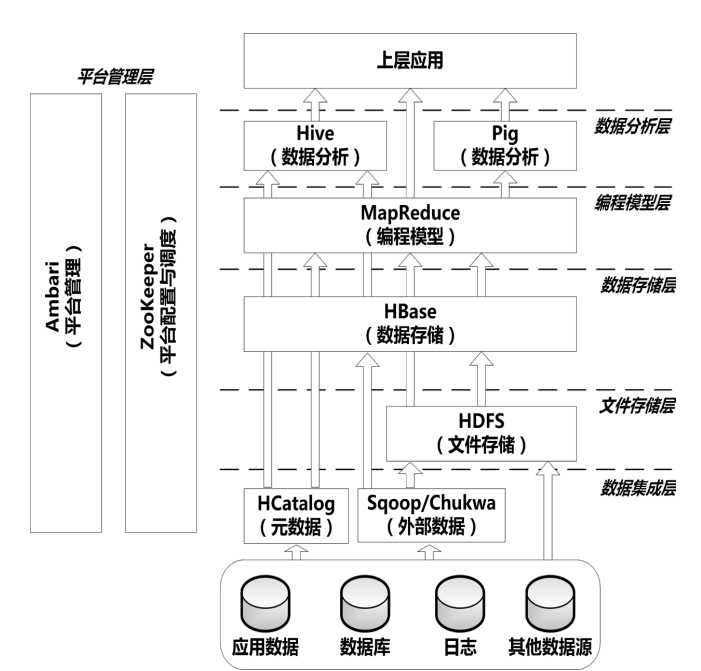














HDFS的写入策略：

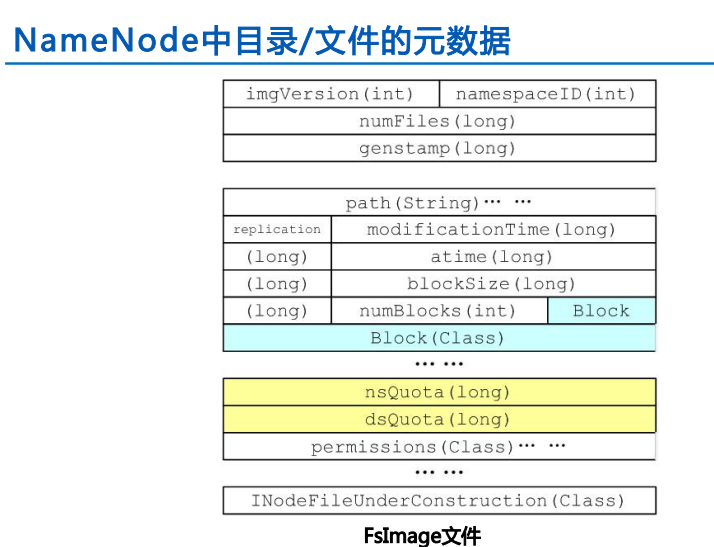
文件拆分(64MB或128MB) -> 分配DN(一般是复制到3个DN) -> 3个DN分布在两个机架，其中两个DN在一个机架上。

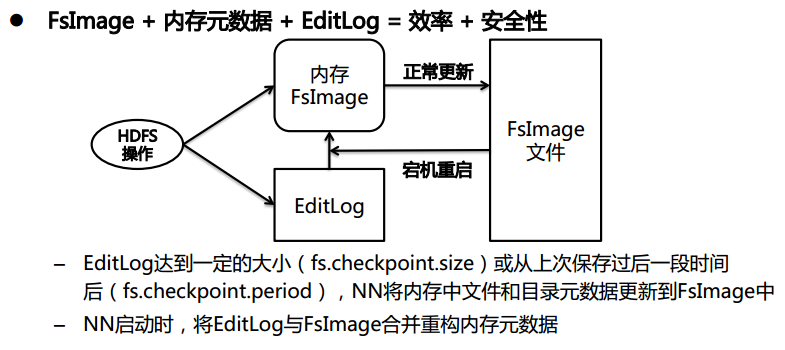


节点错误：每3秒给NN发送心跳

网络错误;Client给DN发送数据，DN需要能给Client发送ACK

存储错误:存储数据的同时，存储checksum





MR完整流程：

作业启动，作业初始化（切分数据），作业/任务调度，Map任务执行，shuffle，reduce，作业完成

