

# 分布式云平台

讲师：肖斌

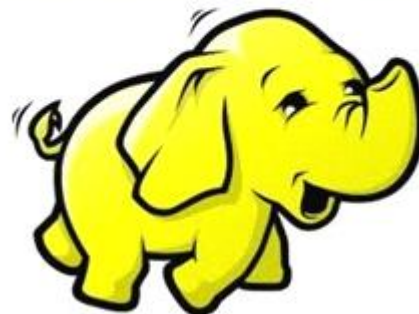
- Hadoop简介
- Hadoop分布式文件系统HDFS
- Hadoop分布式计算框架MR
- Hadoop体系架构
- Hadoop安装
- Hadoop shell
- Hadoop API
- 公司案例
- Hadoop2的特性

## Hadoop的思想之源：Google

- 面对的数据和计算难题
  - 大量的网页怎么存储
  - 搜索算法
- 带给我们的关键技术和思想
  - GFS
  - Map-Reduce
  - Bigtable ( 后面讲 )

## Hadoop创始人介绍

- Hadoop作者Doug cutting，就职Yahoo期间开发了Hadoop项目，目前在Cloudera 公司从事架构工作



- Hadoop简介
  - 名字来源于Doug Cutting儿子的玩具大象。
- 2003-2004年，Google公开了部分GFS和Mapreduce思想的细节，以此为基础Doug Cutting等人用了2年业余时间实现了DFS和Mapreduce机制，一个微缩版：Nutch
- Hadoop 于 2005 年秋天作为 Lucene的子项目 Nutch的一部分正式引入Apache基金会。2006 年 3 月份，Map-Reduce 和 Nutch Distributed File System (NDFS) 分别被纳入称为 Hadoop 的项目中

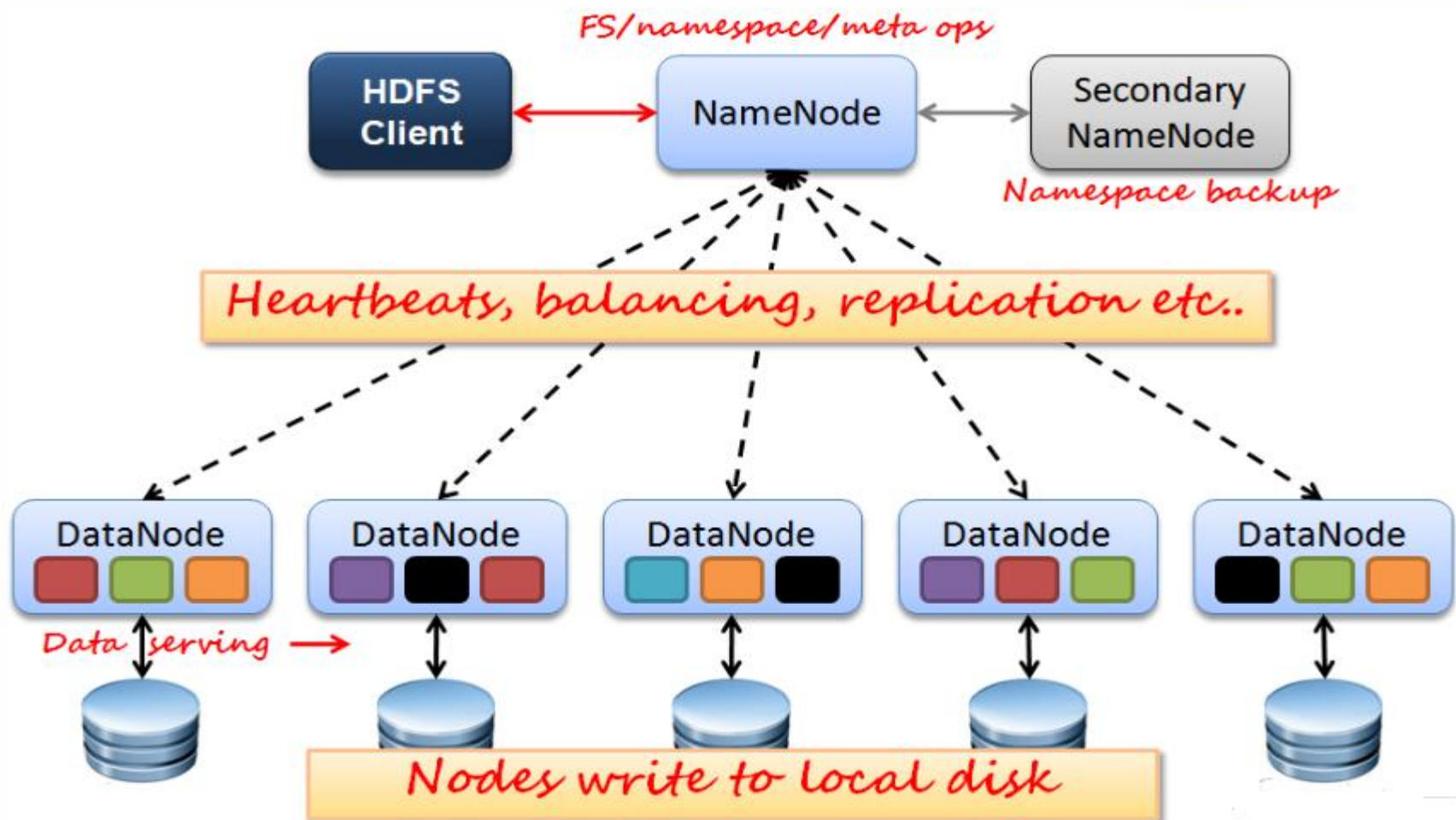
- Hadoop简介
  - 分布式存储系统HDFS（Hadoop Distributed File System）
    - 分布式存储系统
    - 提供了 高可靠性、高扩展性和高吞吐率的数据存储服务
  - 分布式计算框架MapReduce
    - 分布式计算框架
    - 具有 易于编程、高容错性和高扩展性等优点。

- HDFS优点：
  - 高容错性
    - 数据自动保存多个副本
    - 副本丢失后，自动恢复
  - 适合批处理
    - 移动计算而非数据
    - 数据位置暴露给计算框架
  - 适合大数据处理
    - GB、TB、甚至PB 级数据
    - 百万规模以上的文件数量
    - 10K+ 节点
  - 可构建在廉价机器上
    - 通过多副本提高可靠性
    - 提供了容错和恢复 机制

- HDFS缺点：
  - 低延迟数据访问
    - 比如毫秒级
    - 低延迟与高吞吐率
  - 小文件存取
    - 占用NameNode 大量内存
    - 寻道时间超过读取时间
  - 并发写入、文件随机修改
    - 一个文件只能有一个写者
    - 仅支持append

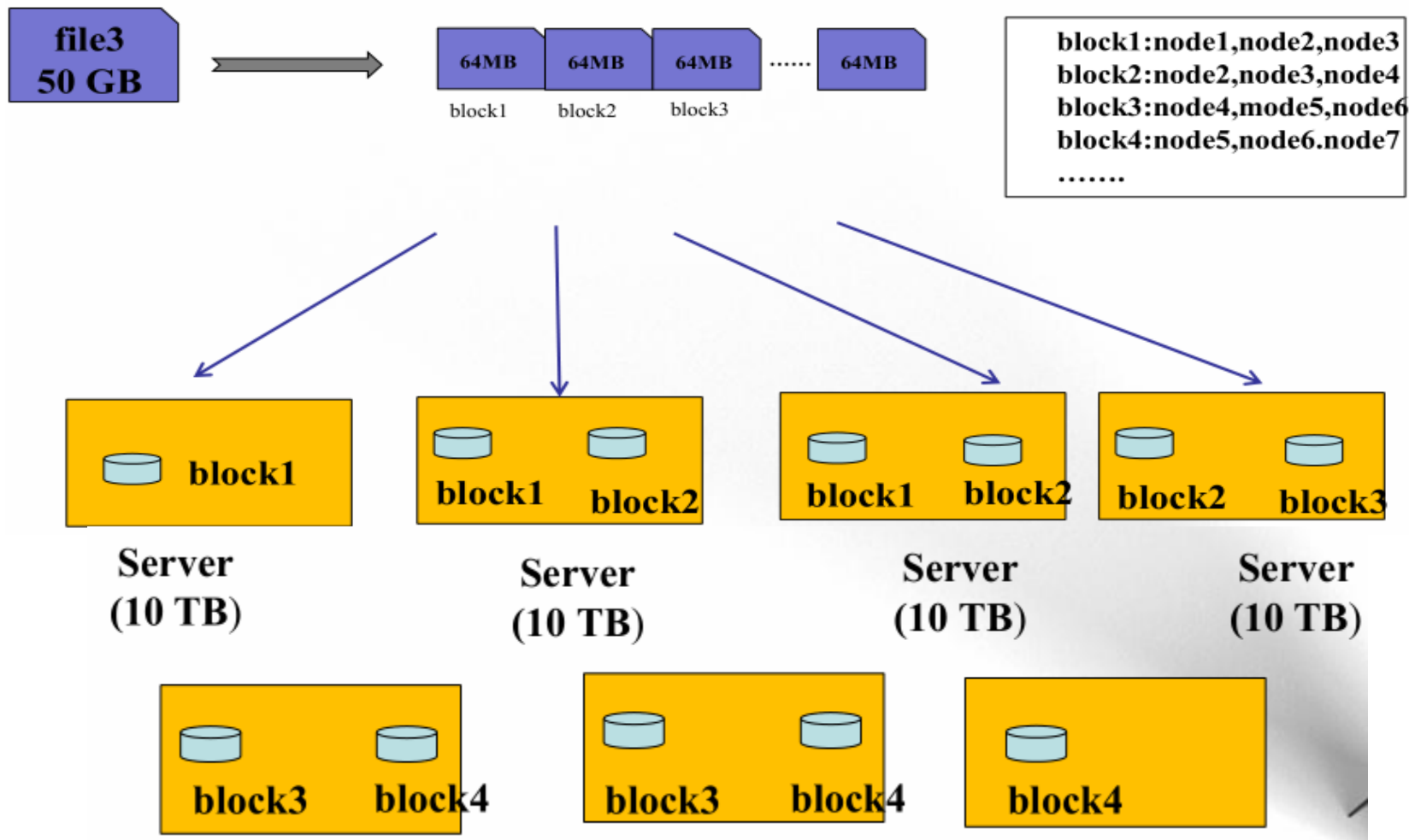


- HDFS架构：



- HDFS 数据存储单元 ( block )
  - 文件被切分成固定大小的数据块
    - 默认数据块大小为64MB , 可配置
    - 若文件大小不到64MB , 则单独存成一个block
  - 一个文件存储方式
    - 按大小被切分成若干个block , 存储到不同节点上
    - 默认情况下每个block都有三个副本
  - Block大小和副本数通过Client端上传文件时设置 , 文件上传成功后副本数可以变更 , Block Size不可变更

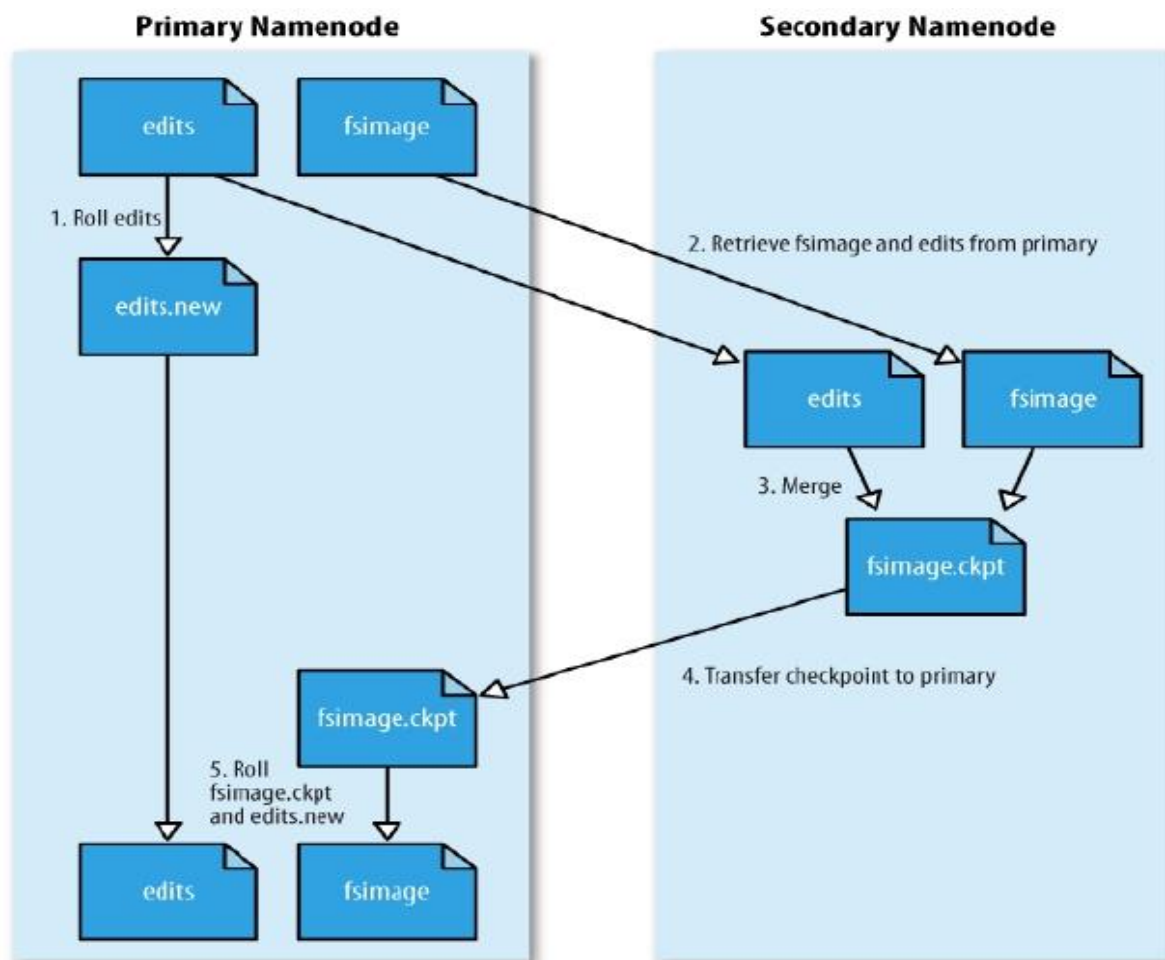
- HDFS设计思想：



- NameNode ( NN )
  - NameNode主要功能：接受客户端的读写服务
  - NameNode保存metadate信息包括
    - 文件ownership和permissions
    - 文件包含哪些块
    - Block保存在哪个DataNode ( 由DataNode启动时上报 )
  - NameNode的metadate信息在启动后会加载到内存
    - metadata存储到磁盘文件名为"fsimage"
    - Block的位置信息不会保存到fsimage
    - edits记录对metadata的操作日志

- SecondaryNameNode ( SNN )
  - 它不是NN的备份（但可以做备份），它的主要工作是帮助NN合并edits log，减少NN启动时间。
  - SNN执行合并时机
    - 根据配置文件设置的时间间隔fs.checkpoint.period 默认3600秒
    - 根据配置文件设置edits log大小 fs.checkpoint.size 规定edits文件的最大值默认是64MB

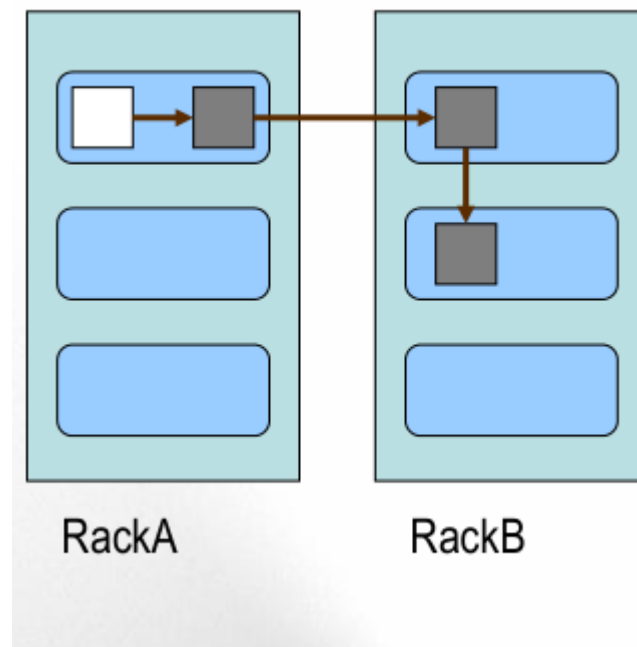
- SNN合并流程



- DataNode ( DN )
  - 存储数据 ( Block )
  - 启动DN线程的时候会向NN汇报block信息
  - 通过向NN发送心跳保持与其联系 ( 3秒一次 ) , 如果NN 10分钟没有收到DN的心跳 , 则认为其已经lost , 并copy其上的block到其它DN

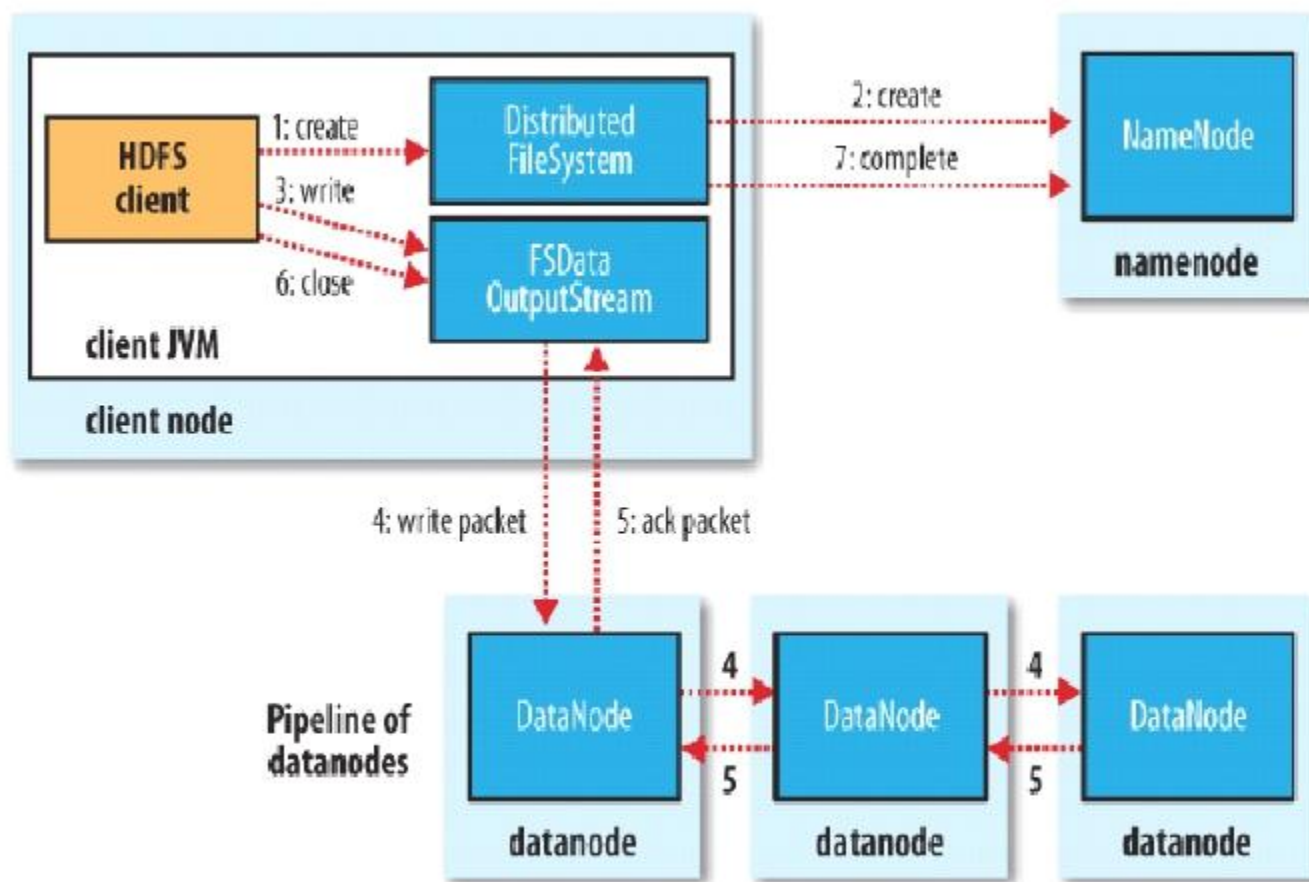


- Block的副本放置策略
  - 第一个副本：放置在上传文件的DN；如果是集群外提交，则随机挑选一台磁盘不太满，CPU不太忙的节点。
  - 第二个副本：放置在于第一个副本不同的机架的节点上。
  - 第三个副本：与第二个副本相同机架的节点。
  - 更多副本：随机节点

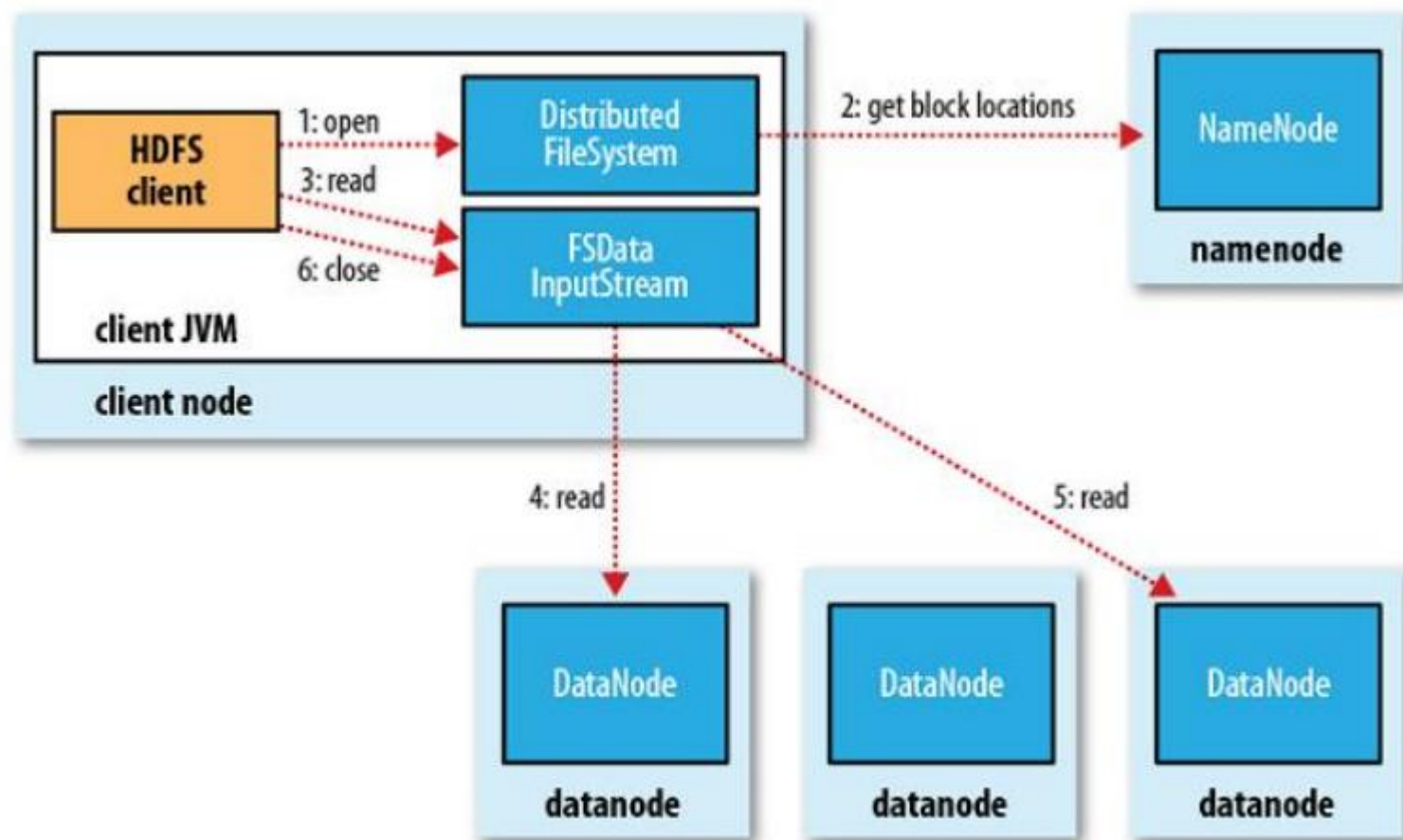




- HDFS写流程



- HDFS读流程



- HDFS文件权限
  - 与Linux文件权限类似
    - r: read; w: write; x: execute , 权限x对于文件忽略 , 对于文件夹表示是否允许访问其内容
  - 如果Linux系统用户zhangsan使用hadoop命令创建一个文件 , 那么这个文件在HDFS中owner就是zhangsan。
  - HDFS的权限目的 : 阻止好人做错事 , 而不是阻止坏人做坏事。HDFS相信 , 你告诉我你是谁 , 我就认为你是谁。

- 安全模式

- namenode启动的时候，首先将映像文件(fsimage)载入内存，并执行编辑日志(edits)中的各项操作。
- 一旦在内存中成功建立文件系统元数据的映射，则创建一个新的fsimage文件(这个操作不需要SecondaryNameNode)和一个空的编辑日志。
- 此刻namenode运行在安全模式。即namenode的文件系统对于客户端来说是只读的。(显示目录，显示文件内容等。写、删除、重命名都会失败)。
- 在此阶段Namenode收集各个datanode的报告，当数据块达到最小副本数以上时，会被认为是“安全”的，在一定比例（可设置）的数据块被确定为“安全”后，再过若干时间，安全模式结束
- 当检测到副本数不足的数据块时，该块会被复制直到达到最小副本数，系统中数据块的位置并不是由namenode维护的，而是以块列表形式存储在datanode中。

- HDFS安装
  - 伪分布式安装
  - 完全分布式安装
    - 下载
    - 解压
    - 检查java和ssh的免密码登陆
    - 修改core-site.xml
    - 修改hdfs-site.xml
    - 修改masters文件和slaves文件
    - 格式化namenode
    - Start-hdfs.sh启动

- HDFSshell

```
[root@hadoop001 bin]# ./hadoop fs
Usage: java FsShell
        [-ls <path>]
        [-lsr <path>]
        [-du <path>]
        [-dus <path>]
        [-count[-q] <path>]
        [-mv <src> <dst>]
        [-cp <src> <dst>]
        [-rm [-skipTrash] <path>]
        [-rmr [-skipTrash] <path>]
        [-expunge]
        [-put <localsrc> ... <dst>]
        [-copyFromLocal <localsrc> ... <dst>]
        [-moveFromLocal <localsrc> ... <dst>]
        [-get [-ignoreCrc] [-crc] <src> <localdst>]
        [-getmerge <src> <localdst> [addnl]]
        [-cat <src>]
        [-text <src>]
        [-copyToLocal [-ignoreCrc] [-crc] <src> <localdst>]
        [-moveToLocal [-crc] <src> <localdst>]
        [-mkdir <path>]
        [-setrep [-R] [-w] <rep> <path/file>]
        [-touchz <path>]
        [-test -[ezd] <path>]
        [-stat [format] <path>]
        [-tail [-f] <file>]
        [-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
        [-chown [-R] [OWNER] [:[GROUP]] PATH...]
```

- HDFS编程：
  - 上传文件
  - 创建新文件
  - 删除文件
  - 查看文件内容
  - ....