一、HDFS的优点和缺点

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

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

伸缩性

添加节点可以处理更大的数据集

HDFS缺点：

无法毫秒级数据访问

小文件存取

占用NameNode 大量内存

寻道时间超过读取时间

一个文件只能有一个写者

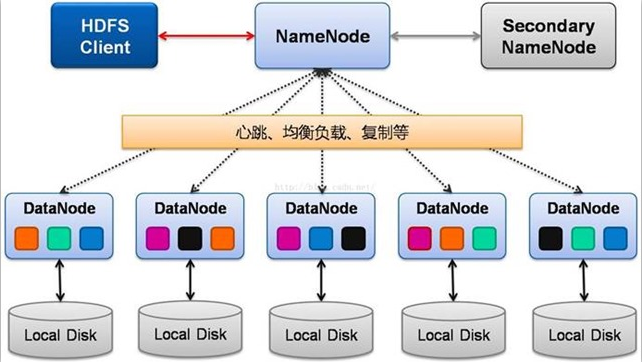
仅支持append

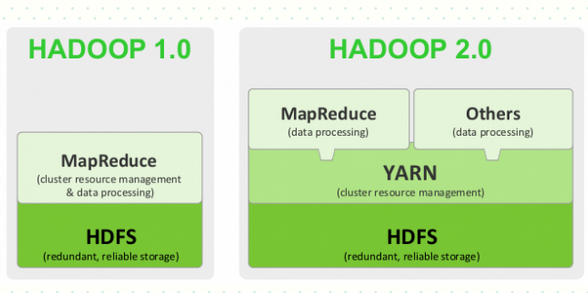
二、概述

1. HDFS集群分为两大角色：NameNode、DataNode (Secondary Namenode)
2. NameNode负责管理整个文件系统的元数据
3. DataNode 负责管理用户的文件数据块
4. 文件会按照固定的大小（blocksize）切成若干块后分布式存储在若干台datanode上
5. 每一个文件块可以有多个副本，并存放在不同的datanode上
6. Datanode会定期向Namenode汇报自身所保存的文件block信息，而namenode则会负责保持文件的副本数量
7. HDFS的内部工作机制对客户端保持透明，客户端请求访问HDFS都是通过向namenode申请来进行

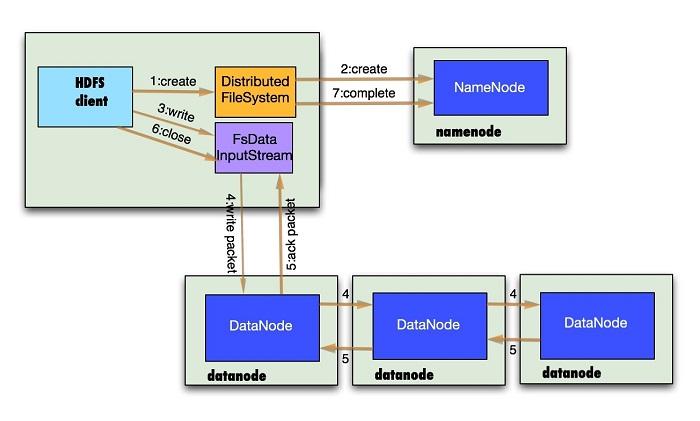
三、架构图

1.HDFS的架构图





三、HDFS写数据流程

IMG_256

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请求第一个 block该传输到哪些datanode服务器上

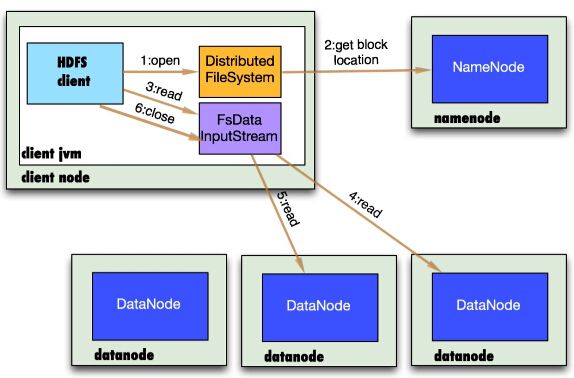
4、namenode返回3个datanode服务器ABC

5、client请求3台dn中的一台A上传数据（本质上是一个RPC调用，建立pipeline），A收到请求会继续调用B，然后B调用C，将真个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，A收到一个packet就会传给B，B传给C；A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。

三、读数据流程



1、跟namenode通信查询元数据，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，现在本地缓存，然后写入目标文件