

# HDFS概述

HDFS (hadoop Distributed File System) 被设计为可以运行在通用通用硬件上、提供流式数据操作、能够处理超大文件的分布式文件系统。HDFS具有高度容错、高吞吐量、容易扩展、高可靠性等特征。

## 使用场景

适合一次写入，多次读出的场景，且不支持文件的修改。适合用来做数据分析，并不适合用来做网盘

## 优点

- 高容错性

- (1) 数据自动保存为多个副本。它通过增加副本的形式，提高容错性。
- (2) 某一个副本丢失后，它可以自动恢复

- 适合处理大数据

- (1) 数据规模：能够处理数据规模达到GB、TB、甚至PB级别的数据
- (2) 文件规模：能够处理百万规模以上的文件数量，数量十分之大
- (3) 可构建在通用硬件上，通过多副本机制，提高可靠性

- 流式数据访问

- (1) 一次写入，多次读取，不能修改，只能追加
- (2) 能保持数据的一致性

## 缺点

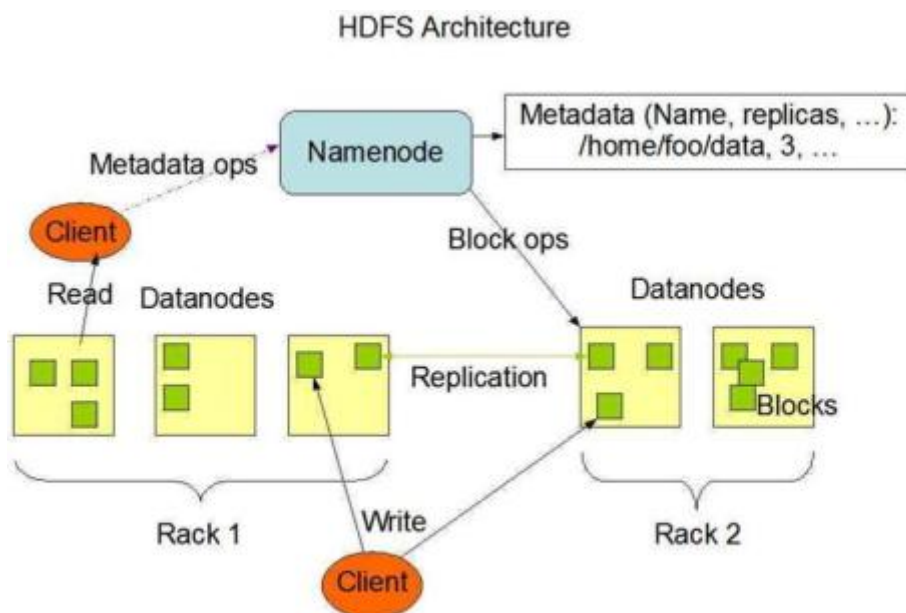
- 不适合低延迟数据访问，比如毫秒级别的存储数据
- 无法高效地对大量小文件进行存储

- (1) 存储大量小文件的话，他会占用namenode大量的内存来储存文件目录和块信息。这样是不可取的，因为一个namenode的容量是有效的
- (2) 小文件存储的寻址时间会超过读取时间，它违反了hdfs的设计目标

- 不支持并发写入、文件随机修改HDFS

- (1) 一个文件只能有一个写，不允许同时写
- (2) 仅支持数据append不允许随意位置添加

## HDFS构架图



HDFS是一个主/从 (master/slave) 体系结构的分布式系统。如图，HDFS集群拥有一个 Namenode (或两个，其中一个是备用节点) 和一些Datanode，用户可以通过HDFS客户端同 Namenode和Datanodes交互以访问文件系统。

## 1. 数据块

HDFS文件始以数据块的形式存储的，数据块是HDFS文件处理的最小单元，HDFS的数据块比 linux的数据块要大，默认128M。HDFS数据块往往会以文件的形式存储在磁盘上。

- 为什么块要设置为128M？

为了最小化寻址开销。

我们来做一个速算，如果寻址时间为10ms，而传输速率为100MB/s，为了使寻址时间仅占传输时间的1%，我们要将块设置为100M。通常情况下设置为128M。

在HDFS中，所有文件都会被分成若干个数据块分布在数据节点上存储。同时由于HDFS会将同一个数据块冗余备份到不同的数据节点上，（一个数据块默认3份）所以一个数据块丢失了并不会有太大的影响。

- 执行下面命令可以列出文件系统在各个文件由哪些块构成

```
hadoop fsck / -file -blocks
```

## 2. 名称节点 (namenode)