

大家好，现在我们将要学习第三章分布式文件系统HDFS，本章将分为5小节，分别是分布式文件系统HDFS简介、HDFS的基本概念、HDFS的体系结构、HDFS的存储策略以及HDFS的数据读写过程。让我们一起来了解分布式文件系统HDFS吧！

(PPT2)

随着大数据时代的到来，人们可以获取的数据成指数倍的增长，需要处理分析的数据集的大小已经远远超过了单台计算机的存储能力，那么如何解决海量数据的高效存储问题呢？现在的普遍做法是把数据分块存储到网络中的不同计算机中，通过文件系统管理网络中跨多台计算机存储的数据集，这个文件系统就是分布式文件系统。

(PPT3)

分布式文件系统把文件分别存储到多个计算机节点上，成千上万的计算机节点构成计算机集群。

计算机集群的基本构架如图所示。集群中计算机节点是由普通硬件组成，这些节点存放在不同的机架（Rack）上，每个机架可以放8-64个节点，同一机架的不同节点之间通过网络互连，不同机架之间采用交换机互连。

(PPT4)

分布式文件系统采用了主/从（Master/Slave）结构，由于分布式存储的性质，集群拥有两类节点“主节点”（NameNode）和“从节点”（DataNode）。

“主节点”（NameNode）：系统中通常只有一个，充当管理员的角色，管理存储和检索多个从节点的实际数据所需的所有元数据。

“从节点”（DataNode）：系统中通常有多个，是文件系统中真正存储数据的地方，在“主节点”统一调度下进行数据块的创建、删除和复制。

客户端以特定的通信协议通过网络与“主节点”建立联系，向“主节点”提出文件访问请求，找到文件块所在的位置数据节点位置；

有了文件块所在的数据节点后，客户端以特定的通信协议通过网络与数据节点建立联系，进而读写文件块。

(PPT5)

HDFS全称为Hadoop Distributed File System，它是Hadoop平台2大核心组件之一。Hadoop平台要解决的两大核心问题：大数据的分布式存储和大数据的分布式处理，HDFS就是为了解决海量数据的分布式存储问题的，因此它的设计目标有别于其他的分布式文件系统。

1、兼容廉价的硬件设备：由于硬件出错是常态，当数据保存在计算机集群的各个计算机中，常会出现节点失效的情况，因此HDFS通过多副本机制，提高可靠性，同时还提供了快速检测硬件故障和自动回复机制，从而使得在硬件出错的情况下也能实现数据的完整性。

2、数据流读写：HDFS是为了解决批量数据处理的要求而设计的，因此它不支持随机读写，它读写文件采用流的方式进行。

3、大数据集：处理数据达到 GB、TB、甚至PB级别的数据；能够处理百万规模以上的文件数量。

4、简单的文件模型：HDFS采用了“一次写入，多次读取”的简单文件模型，文件一旦完成写入不能修改，只能追加和读出。

5、强大的跨平台兼容性：HDFS是采用java语言实现的，具有很好的跨平台兼容性，支持JVM的机器都可以运行HDFS。

(PPT6)

HDFS特殊的设计，在实现上述优良特性的同时，也是的自伤具有一些应用的局限性，主要包括以下几个方面：

1、不适合低延时数据访问

HDFS不适用于实时查询这种对延迟要求高的场景，例如：股票实盘。往往应对低延迟数据访问场景需要通过数据库访问索引的方案来解决，Hadoop生态圈中的Hbase具有这种随机读、低延迟等特点。

2、无法高效存储大量小文件

小文件通常定义为远小于HDFS的block size（默认64MB）的文件，由于每个文件都会产生各自的MetaData元数据，HDFS通过Namenode来存储这些信息，若小文件过多，容易导致Namenode存储出现瓶颈。

3、不支持并发写入、文件随机修改

为了保证并发性，HDFS需要一次写入多次读取，目前不支持多用户写入，若要修改，也是通过追加的方式添加到文件的末尾处，出现太多文件需要更新的情况，HDFS是不支持的。

针对有多人写入数据的场景，可以考虑采用Hbase的方案。