# **Hadoop HDFS**

Hadoop文件系统使用分布式文件系统设计开发。它是运行在普通硬件。不像其他的分布式系统，HDFS是高度容错以及使用低成本的硬件设计。

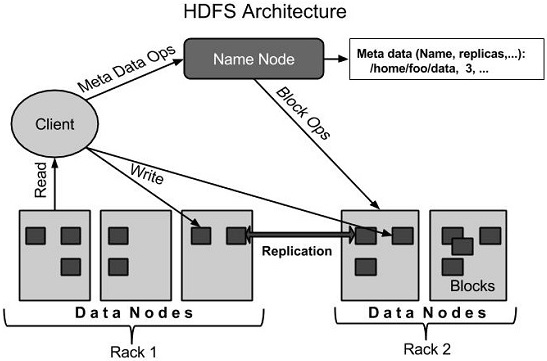
HDFS拥有超大型的数据量，并提供更轻松地访问。为了存储这些庞大的数据，这些文件都存储在多台机器。这些文件都存储以冗余的方式来拯救系统免受可能的数据损失，在发生故障时。 HDFS也使得可用于并行处理的应用程序。

## **一、HDFS的特点**

* 它适用于在分布式存储和处理。
* Hadoop提供的命令接口与HDFS进行交互。
* 主节点和数据节点的帮助用户内置的服务器能够轻松地检查集群的状态。
* 流式访问文件系统数据。
* HDFS提供了文件的权限和验证。

## **二、HDFS架构**

下面给出是Hadoop的文件系统的体系结构。



HDFS遵循主从架构，它具有以下元素。

## 三、主节点 - Namenode

主节点是包含GNU/Linux操作系统和软件名称节点的普通硬件。它是一个可以在商品硬件上运行的软件。具有名称节点系统作为主服务器，它执行以下任务：

* 管理文件系统命名空间。
* 规范客户端对文件的访问。
* 它也执行文件系统操作，如重命名，关闭和打开的文件和目录。

## 四、数据节点 - Datanode

Datanode具有GNU/Linux操作系统和软件Datanode的普通硬件。对于集群中的每个节点(普通硬件/系统)，有一个数据节点。这些节点管理数据存储在它们的系统。

* 数据节点上的文件系统执行的读写操作，根据客户的请求。
* 还根据主节点的指令执行操作，如块的创建，删除和复制。

## 五、块

一般用户数据存储在HDFS文件。在一个文件系统中的文件将被划分为一个或多个段和/或存储在个人数据的节点。这些文件段被称为块。换句话说，数据的HDFS可以读取或写入的最小量被称为一个块。缺省的块大小为128MB，但它可以增加按需要在HDFS配置来改变。

## 六、HDFS的目标

故障检测和恢复：由于HDFS包括大量的普通硬件，部件故障频繁。因此HDFS应该具有快速和自动故障检测和恢复机制。

巨大的数据集：HDFS有数百个集群节点来管理其庞大的数据集的应用程序。

数据硬件：请求的任务，当计算发生不久的数据可以高效地完成。涉及巨大的数据集特别是它减少了网络通信量，并增加了吞吐量。