

# Hadoop

## 什么是Hadoop?

Hadoop是一个开源的分布式计算和存储框架，其核心部件包括分布式文件系统（HDFS）和MapReduce。

## Hadoop的作用是什么?

Hadoop的作用就是将多台计算机组织成一台计算机，HDFS相当于计算机的硬盘，MapReduce相当于计算机的CPU控制器。

## HDFS怎么保证存储稳定性?

通过多存几份的方式，HDFS的存储单位是块（BLOCK），一个文件可能会被分成多个块，这些块会复制几份并存储在不同的数据节点上，如果一个数据节点发生故障也不会丢失。

## HDFS的结构是什么样的?

HDFS只有一个命名节点，有多个副命名节点，还有多个数据节点组成。

命名节点的作用是指挥其他节点存储的节点，存储映射信息也提供映射服务。

数据节点的作用是存储数据和读写数据。

## MapReduce

MapReduce的思想是什么?

MapReduce把每个节点统计到的结果汇总给他的高级节点，由高级节点来合并结果。

Map的作用是并行地映射，Reduce的作用是合并结果。