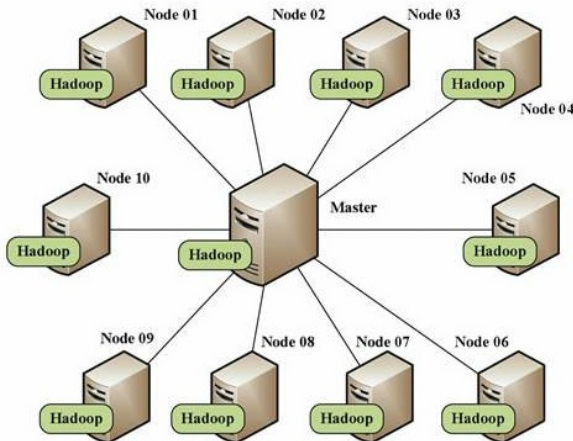


5-大数据相关的概念介绍

-----成都尚学堂-mr-zeng-----

计算机集群



大数据相关关键字

大数据：大量、海量的数据。

分布式：把数据，程序分别部署到不同的计算机设备上-通过网络连接起来共同工作（程序的协作-zookeeper）-提供统一服务。

大数据的存储（分布式存储）：把数据存在不同计算机设备上，**HDFS**，FastDFS、GFS等技术

云计算：在云端计算，指不是在客户端计算在服务器计算，通常服务器计算需要进行分布式计算-分布式计算也是云计算的一种。

大数据的计算（分布式计算）：把数据在不同计算机设备上计算

mapreduce：（数据已经准备好）基于硬盘，中间运算技术都会写入磁盘。速度一般，数据量大，成本相对低。

spark：（数据已经准备好）基于内存，中间运算直接在内存操作。速度快，数据量大，成本相对高。

storm：（数据边计算边产生）基于内存，实施处理数据，例如：双十一订单实施统计（统计各地区订单排行，各年龄订单排行，各产品订单排行）。

...

大数据技术的应用

存储：

360网盘，百度网盘，微云....

存储加计算：

- 1) 购物网站通过用户历史购买商品推荐
- 2) 广告精准推送（浏览器，应用软件）
- 3) 通过用户发送的微博分析用户喜欢的商品
- 4) qq好友推荐
- 5) 各年份各城市天气信息统计分析（数据量很大500Tb）
 -》找到每个城市所有年最高温度

- 6) 统计联通公司所有的用户的月消费额，总流量数，总通话数，总短信数。（数据量极大！！！）

.....

.....

.....

.....