1. Hadoop
2. 是分布式系统基础架构，一个能够对大量数据进行分布式处理的软件框架。Hadoop以一种可靠、高效、可伸缩的方式进行数据处理；用户可以在不了解分布式底层细节的情况下，开发分布式程序。
3. Hadoop的框架最核心的设计就是：HDFS和MapReduce。HDFS为海量的数据提供了存储，而MapReduce则为海量的数据提供了计算。

* NameNode中记录了文件是如何被拆分成block以及这些block都存储到了那些DateNode节点.NameNode同时保存了文件系统运行的状态信息.
* DataNode中存储的是被拆分的blocks.
* Secondary NameNode帮助NameNode收集文件系统运行的状态信息.
* JobTracker当有任务提交到Hadoop集群的时候负责Job的运行,负责调度多个TaskTracker.
* TaskTracker负责某一个map或者reduce任务.
* Hadoop YARN：作业调度和集群资源管理的框架。
* Hadoop MapReduce：一种用于并行处理大型数据集的基于YARN的系统。
* ZOOKEEPER：分布式协调服务基础组件
* Sqoop：数据导入导出工具

MapReduce、JobTracker、TaskTracker

NameNode：大领导（NameNode HA）

Second Namenode：快照，没有NameNode更新那么快，但是NameNode可以作为备份的NameNode使用

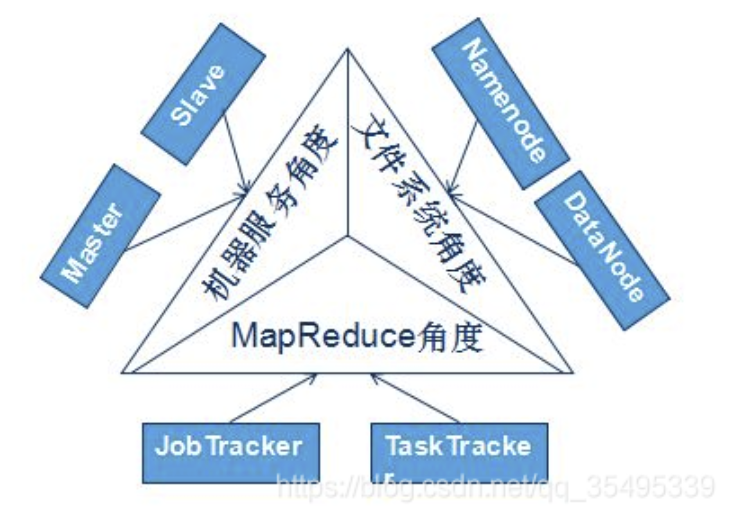
DataNode：小领导，负责每一个节点

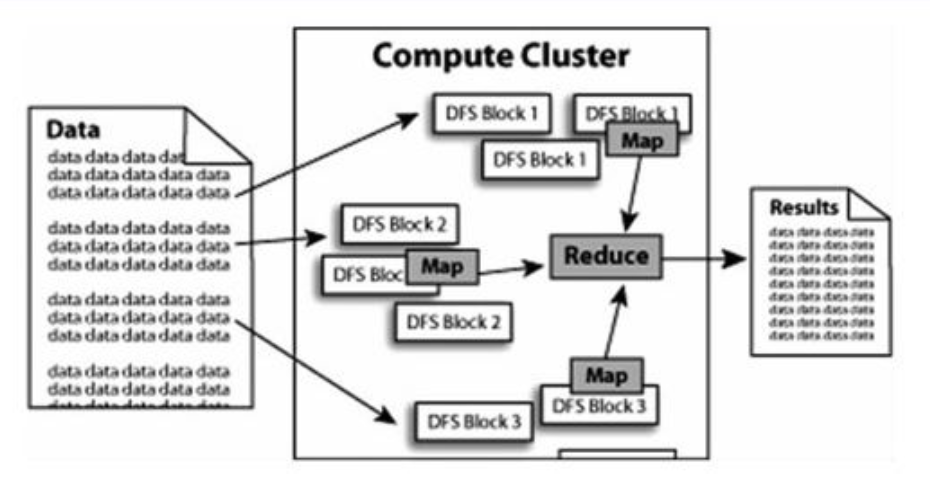
Hadoop

1. 集群：一个Master和若干个Slave

* Master：管理集群的资源、接收客户端提交的请求
* Slave：管理当前节点资源

需要一个master节点





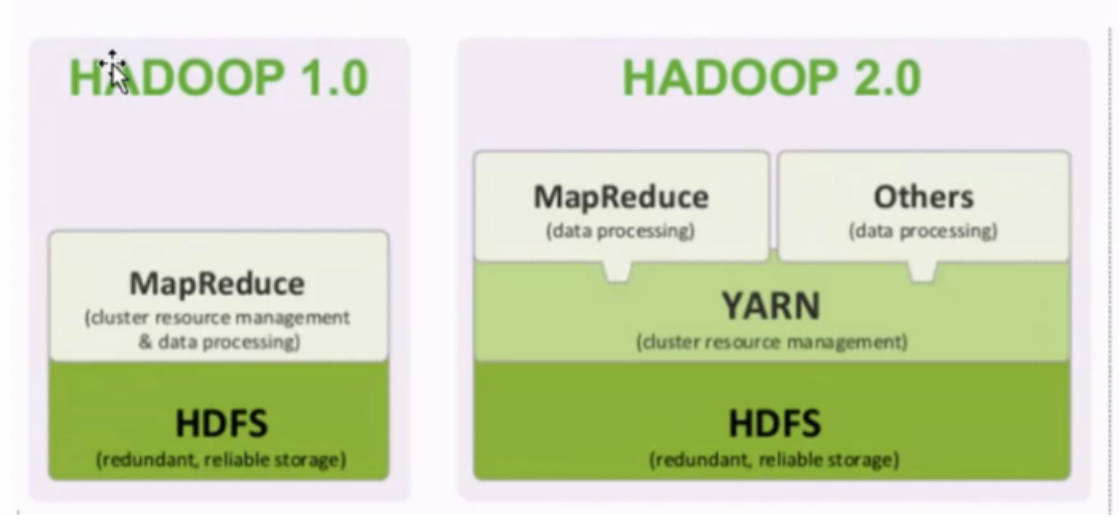
1. MapReduce两种主要的后台程序：jobtracker和tasktracker。

* Jobtracker：是主线程，它负责接收客户作业提交，调度任务到工作节点上运行，并提供诸如监控工作节点状态及任务进度等管理功能。
* Tasktracker：由jobtracker指派任务，实例化用户程序，在本地执行任务并周期性地向jobtracker汇报状态。

Tracker是跟踪者，跟踪器。JobTracker是项目经理。

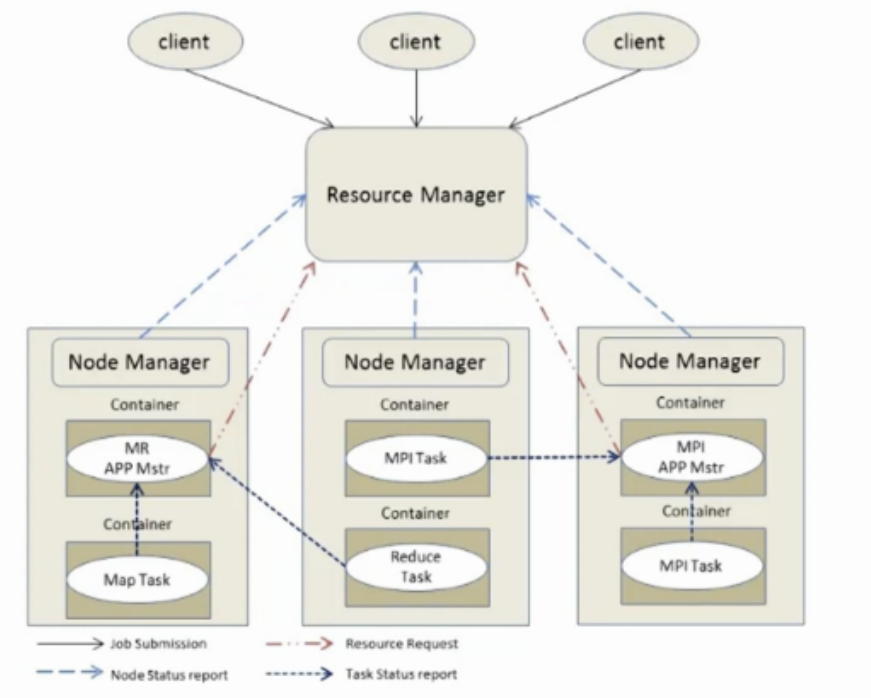
1. 在hadoop2的0.23版本之后，

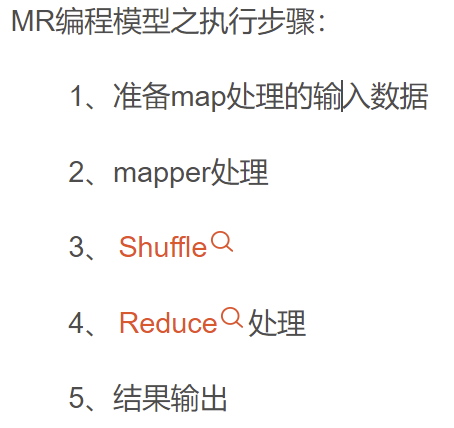
* HDFS主要是Hadoop的存储，用于海量数据的存储；
* MapReduce主要运用于分布式计算；
* YARN：资源管理系统。



通过YARN实现资源的调度与管理，总体上仍然是Master/Slave结构。ResourceManager为Master，NodeManager为Slave，并通过HA方案实现了ResourceManager的高可用。

* ResourceManager：负责对各个NodeManager上的资源进行统一管理和调度。
* NodeManager：它是每个节点上的资源和任务管理器，它不仅定时向ResourceManager汇报本节点上的资源使用情况和各个Container的运行状态，还接收并处理来自ApplicationMaster的Container启动/停止等各种请求。
* ApplicationMaster：当用户提交一个应用程序时，需要提供一个用以跟踪和管理这个程序的ApplicationMaster，它负责向ResourceManager申请资源，并要求NodeManger启动可以占用一定资源的任务。





提供思路：根据Hadoop，有一个管理节点，由管理节点进行资源的分配

区别之处：client找mapreduce和摄像机找边缘计算节点

P2P可以做到既从别的节点获取资源，又给别的节点提供资源。

1. 不行就丢弃
2. 分布式管理会遇到什么问题

软件

管理节点可以做HA

计算能力方面

计算核心故障处理

物理安全

视频的压缩和安全传输

计算网络优先CFN

算力感知网络

<http://news.sohu.com/a/501724147_121036026>

