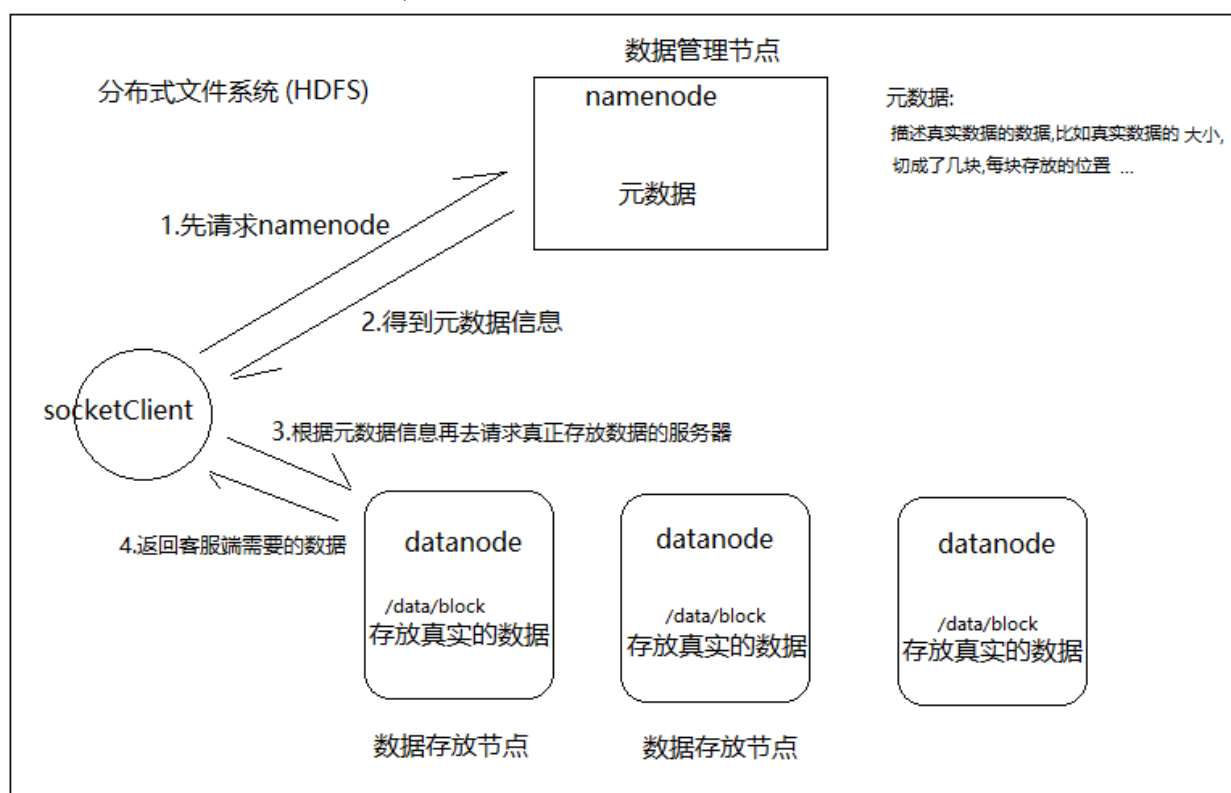


# hbasehw01

请阐述或画出Hadoop逻辑架构图，并注明主从节点所负责的任务是什么，主要是HDFS和YARN

## 1 HDFS

- 基于Google的GFS理论，用JAVA开发的分布式文件系统
- HDFS运行于文件系统之上的系统，如ext3,ext4,xfs等
- 使用标准通用硬件设备处理大量数据的随机存储
- HDFS包含NameNode, DataNode组件，是Master/Slave架构
- NN存储metadata,包含文件名，权限，block等信息
- DN 存储block数据
- 文件以block存储，每个块默认128M,每个块默认在集群中复制3份



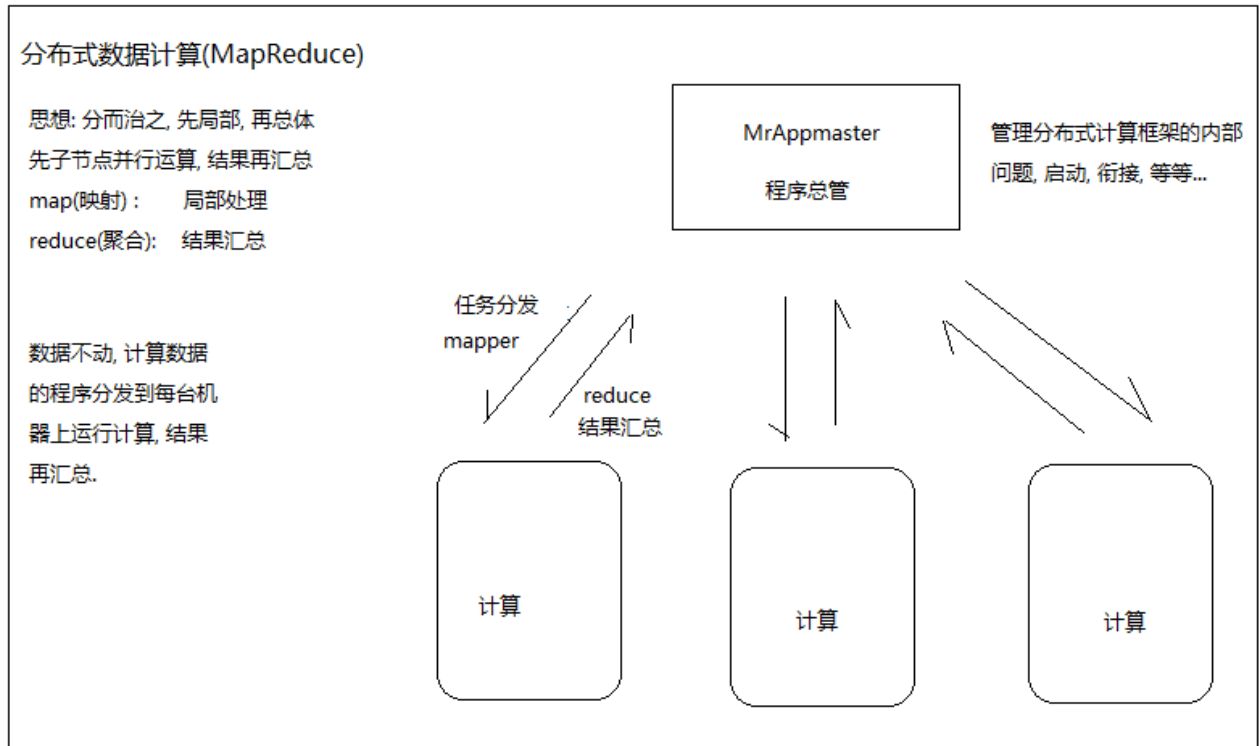
[http://blog.csdn.net/csdm\\_admin](http://blog.csdn.net/csdm_admin)

## 2 Map-Reduce

由于每个节点存储着大量数据，如果运算时数据传输必然占用大量带宽，效率低。所以采取数据不动，处理数据的程序去找存储数据的节点，在节点上进行计算。

- 不是一种语言，甚至不是一个框架，是一个处理数据计算的编程模型
- Hadoop集群上处理数据的系统
- 程序数理包含两个阶段，Map阶段和Reduce阶段
- 在Map和Reduce阶段中是Shuffle和Sort阶段
- 每个Map任务在不同机器上处理数据集的一部分

- 当所有Map任务完成后，MapReduce系统将中间数据发送给Reduce节点进行Reduce 处理



[http://blog.csdn.net/csdm\\_admin](http://blog.csdn.net/csdm_admin)

### 3 Yarn

- Yet Another Resource Negotiator
- Hadoop资源管理器, 包含ResourceManager 和NodeManager
- Resource Manager

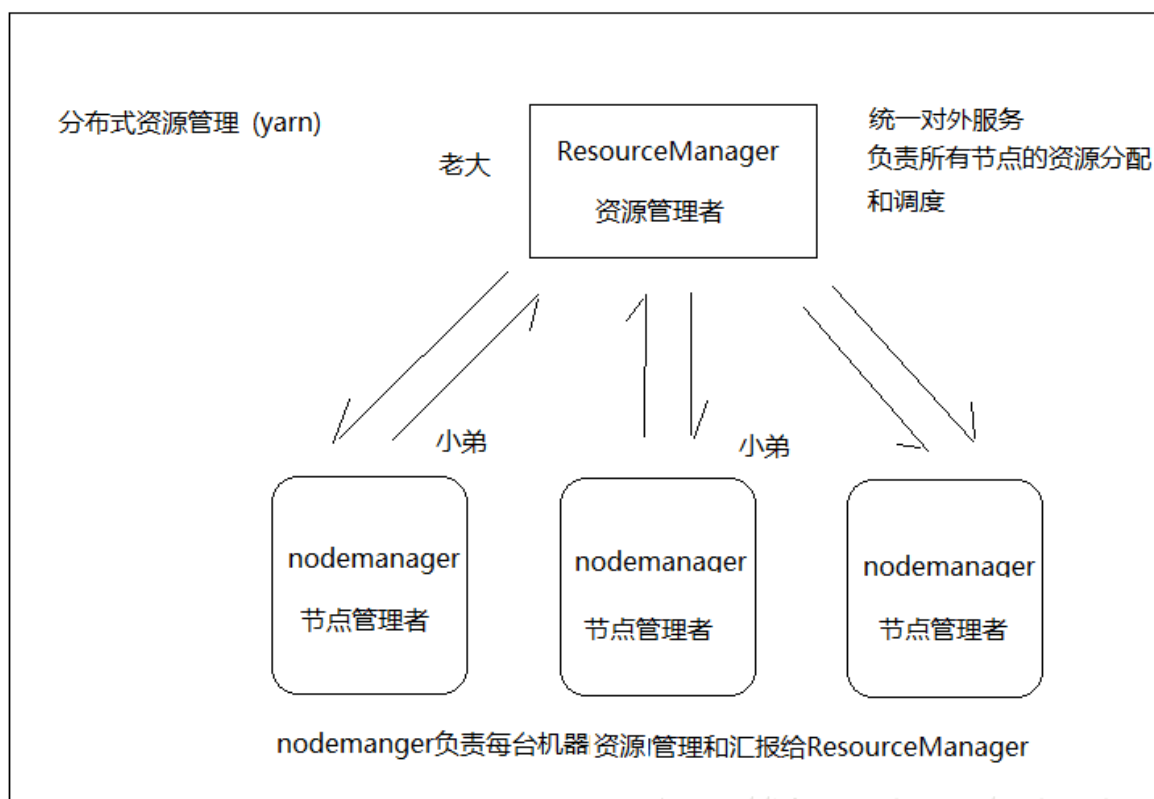
统一对外服务, 负责所有节点的资源分配和调度。

Application master; Scheduler ;

NodeManager

负责每台机器资源管理和分配。

## Contener-&gt;CPU,MEM



[http://blog.csdn.net/csdm\\_admin](http://blog.csdn.net/csdm_admin)