Google三驾马车

1 GFS（Google File System）。一个分布式文件系统，隐藏下层负载均衡，冗余复制等细节，对上层程序提供一个统一的文件系统API接口。Google根据自己的需求对它进行了特别优化，包括：超大文件的访问，读操作比例远超过写操作，PC机极易发生故障造成节点失效等。GFS把文件分成64MB的块，分布在集群的机器上，使用Linux的文件系统存放。同时每块文件至少有3份以上的冗余。中心是一个Master节点，根据文件索引，找寻文件块。

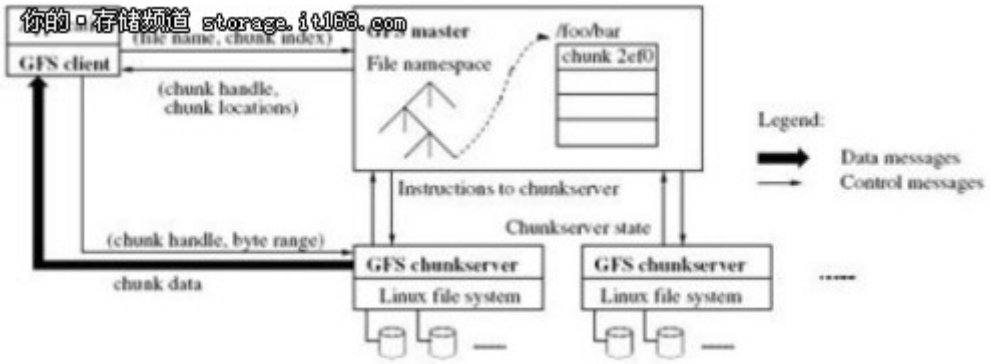
2 MapReduce。Google发现大多数分布式运算可以抽象为MapReduce操作。Map是把输入Input分解成中间的Key/Value对，Reduce把Key/Value合成最终输出Output。这两个函数由程序员提供给系统，下层设施把Map和Reduce操作分布在集群上运行，并把结果存储在GFS上。

3 BigTable。一个大型的分布式数据库，这个数据库不是关系式的数据库。像它的名字一样，就是一个巨大的表格，用来存储结构化的数据。

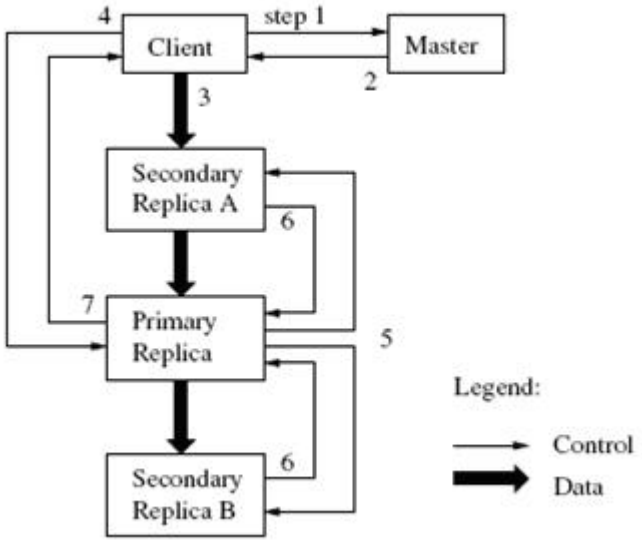
Google的伟大之处，不仅因为它建立了一个很好很强大的搜索引擎，而且还在于它创造了3项革命性技术：GFS、MapReduce和BigTable，即所谓的Google三驾马车。

《Google文件系统》，《MapReduce：超大集群的简单数据处理》，《BigTable：结构化数据的分布式存储系统》这3篇重量级论文的发表，不仅使大家理解了Google搜索引擎背后强大的技术支撑，而且论文和相关的开源技术极大地普及了云计算中非常核心的分布式技术。随后，克隆这3项技术的开源产品如雨后春笋般涌现，Hadoop就是其中的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。这样，Google以一种独特的方式，影响了大数据处理的潮流。

1. Google fs



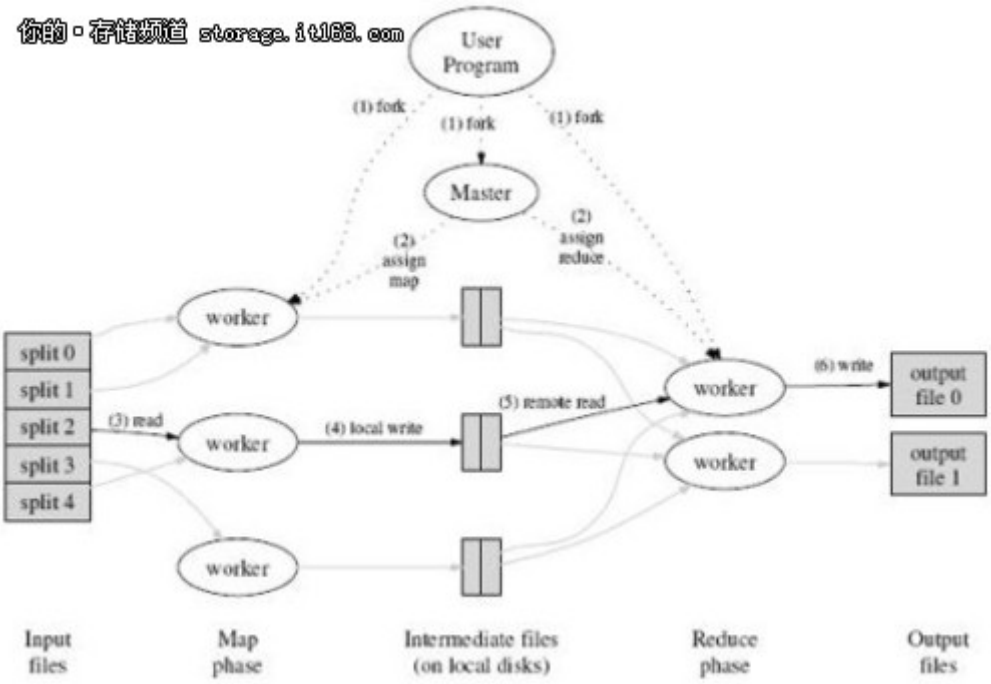
1. GFS的结构图，由一个master和大量的chunkserver构成
2. Google设置一个主来保存目录和索引信息，这是为了简化系统结果，提高性能来考虑的，但是这就会造成主成为单点故障或者瓶颈。为了消除主的单点故障Google把每个chunk设置的很大(64M)，这样，由于代码访问数据的本地性，application端和master的交互会减少，而主要数据流量都是Application和chunkserver之间的访问。
3. 另外，master所有信息都存储在内存里,启动时信息从chunkserver中获取。提高了master的性能和吞吐量，也有利于master当掉后，很容易把后备j机器切换成master。
4. 客户端和chunkserver都不对文件数据单独做缓存，只是用linux文件系统自己的缓存 。
5. GFS的复制



2.Mapreduce

Mapreduce是针对分布式并行计算的一套编程模型。

在模型的构建上，它的思路基本上基于分治，但是引入了 Map 和 Reduce 两个概念，Map 将输入的数据进行初步处理，得到一个规范化的、可被进一步处理的中间值，然后 Reduce 将这些中间值进一步进行运算，最终得出结果 。对于整个架构的实现，它其实也是运行在 Linux 上的普通用户程序，跟系统解耦合了。它采用的也是类似于 Master-Slave 的架构，有一个 master 负责分配任务：每个 Map 和 Reduce 都被切成许多小片，分配给 worker 执行。执行 Map 的 worker 将中间数据写入内存中，并且将位置回传给 master；master 再将这些位置传给 Reduce worker 执行相应的任务。Reduce 执行完任务后，将结果写入磁盘即可。



1. .Mapreduce是由Map和reduce组成,来自于Lisp，Map是影射，把指令分发到多个worker上去，reduce是规约，把Map的worker计算出来的结果合并。
2. Google的Mapreduce实现使用GFS存储数据。
3. .Mapreduce可用于Distributed Grep,Count of URL Access Frequency,ReverseWeb-Link Graph, Distributed Sort,Inverted Index 等。

3.Bigtable

GFS需要Bigtable来存储结构化数据。

1. BigTable 是建立在 GFS， Scheduler， Lock Service 和 MapReduce 之上的 。
2. 每个Table都是一个多维的稀疏图。
3. 为了管理巨大的Table，把Table根据行分割，这些分割后的数据统称为：Tablets。每个Tablets大概有 100-200 MB，每个机器存储100个左右的 Tablets。底层的架构是：GFS。由于GFS是一种分布式的文件系统，采用Tablets的机制后，可以获得很好的负载均衡。比如：可以把经常响应的表移动到其他空闲机器上，然后快速重建。