



大型网站技术架构

核心原理与案例分析

作者: 李智慧

25 条标注

应用服务器需要处理大量的业务逻辑，因此需要更快更强大的CPU；数据库服务器需要快速磁盘检索和数据缓存，因此需要更快的硬盘和更大的内存；文件服务器需要存储大量用户上传的文件，因此需要更大的硬盘。

页码 5 | 位置 #360-362 | 添加于 2018-09-06 05:17

当一台服务器的处理能力、存储空间不足时，不要企图去换更强大的服务器，对大型网站而言，不管多么强大的服务器，都满足不了网站持续增长的业务需求。这种情况下，更恰当的做法是增加一台服务器分担原有服务器的访问及存储压力。

页码 6 | 位置 #379-381 | 添加于 2018-09-06 05:29

通过负载均衡调度服务器，可将来自用户浏览器的访问请求分发到应用服务器集群中的任何一台服务器上，如果有更多的用户，就在集群中加入更多的应用服务器，使应用服务器的负载压力不再成为整

页码 7 | 位置 #385-386 | 添加于 2018-09-06 05:31

个网站的瓶颈。

页码 7 | 位置 #386-387 | 添加于 2018-09-06 05:31

目前大部分的主流数据库都提供主从热备功能，通过配置两台数据库主从关系，可以将一台数据库服务器的数据更新同步到另一台服务器上。

页码 7 | 位置 #390-391 | 添加于 2018-09-06 05:32

CDN和反向代理的基本原理都是缓存，区别在于CDN部署在网络提供商的机房，使用户在请求网站服务时，可以从距离自己最近的网络提供商机房获取数据；而反向代理则部署在网站的中心机房，当用户请求到达中心机房后，首先访问的服务器是反向代理服务器，如果反向代理服务器中缓存着用户请求的资源，就将其直接返回给用户。

页码 8 | 位置 #401-404 | 添加于 2018-09-06 05:35

分布式数据库是网站数据库拆分的最后手段，只有在单表数据规模非常庞大的时候才使用。不到不得已时，网站更常用的数据库拆分手段是业务分库，将不同业务的数据库部署在不同的物理服务器上。

页码 10 | 位置 #409-411 | 添加于 2018-09-06 05:39

大型网站都是从小型网站发展而来。网站的价值在于它能为用户提供什么价值，在于网站能做什么，而不在于它是怎么做的，所以在网站还很小的时候就去追求网站的架构是舍本逐末，得不偿失的。小

型网站最需要做的就是为用户提供好的服务来创造价值，得到用户的认可，活下去，野蛮生长。

页码 13 | 位置 #437-440 | 添加于 2018-09-06 05:47

技术是用来解决业务问题的，而业务的问题，也可以通过业务的手段去解决。

页码 15 | 位置 #467-468 | 添加于 2018-09-06 05:50

但是一个前提条件是应用服务器上不能保存请求的会话信息。否则服务器宕机，会话丢失，即使将用户请求转发到其他服务器上也无法完成业务处理。

页码 28 | 位置 #684-686 | 添加于 2018-12-04 02:46

对于应用服务器集群，只要服务器上不保存数据，所有服务器都是对等的，通过使用合适的负载均衡设备就可以向集群中不断加入服务器。

页码 29 | 位置 #696-697 | 添加于 2018-12-04 02:59

对于缓存服务器集群，加入新的服务器可能会导致缓存路由失效，进而导致集群中大部分缓存数据都无法访问。

页码 29 | 位置 #697-698 | 添加于 2018-12-04 02:59

需要改进缓存路由算法保证缓存数据的可访问性。

页码 29 | 位置 #699-700 | 添加于 2018-12-04 03:00

因此关系数据库的集群伸缩性方案必须在数据库之外实现，通过路由分区等手段将部署有多个数据库的服务器组成一个集群。

页码 29 | 位置 #701-702 | 添加于 2018-12-04 03:00

网站可伸缩架构的主要手段是事件驱动架构和分布式服务。

页码 30 | 位置 #708-709 | 添加于 2018-12-04 03:08

测试程序通过多线程模拟并发用户的办法来测试系统的并发处理能力，为了真实模拟用户行为，测试程序并不是启动多线程然后不停地发送请求，而是在两次请求之间加入一个随机等待时间，这个时间被称作思考时间。

页码 37 | 位置 #769-771 | 添加于 2018-12-04 06:11

System Load即系统负载，指当前正在被CPU执行和等待被CPU执行的进程数目总和，

页码 38 | 位置 #787-788 | 添加于 2018-12-04 06:14

对系统不断地增加并发请求以增加系统压力，

页码 39 | 位置 #796-796 | 添加于 2018-12-04 06:19

根据网站分层架构，可分为Web前端性能优化、应用服务器性能优化、存储服务器性能优化3大类。

页码 42 | 位置 #829-830 | 添加于 2018-12-04 06:23

一样，反向代理服务器也具有保护网站安全的作用，

页码 44 | 位置 #866-866 | 添加于 2018-12-04 06:34

功能，代理服务器也可以通过配置缓存功能加速Web请求。当用户第一次访问静态内容的时候，静态内容

页码 45 | 位置 #868-868 | 添加于 2018-12-04 06:34

代理服务器也可以通过配置缓存功能加速Web请求。

页码 45 | 位置 #868-868 | 添加于 2018-12-04 06:35

此外，反向代理也可以实现负载均衡的功能，而通过负载均衡构建的应用集群可以提高系统总体处理能力，

页码 45 | 位置 #873-873 | 添加于 2018-12-04 06:34

网站性能优化第一定律：优先考虑使用缓存优化性能。

页码 45 | 位置 #880-881 | 添加于 2018-12-04 06:36

网站数据访问通常遵循二八定律，即80%的访问落在20%的数据上，因此利用Hash表和内存的高速访

问特性，将这20%的数据缓存起来，可很好地改善系统性能，提高数据读取速度，降低存储访问压力。