[**大型网站架构系列：20本技术书籍推荐**](http://www.cnblogs.com/itfly8/p/5128025.html)

<https://www.cnblogs.com/itfly8/p/5128025.html>

## ****1.硬件升级****

普通的P4服务器一般最多能支持每天10万独立IP，如果访问量比这个还要大， 那么必须首先配置一台更高性能的专用服务器才能解决问题 ，否则怎么优化都不可能彻底解决性能问题。

## ****2.负载均衡****

它是根据某种负载策略把请求分发到集群中的每一台服务器上，让整个服务器群来处理网站的请求。   
公司比较有钱的，可以购买专门负责负载均衡的硬件（如：F5）,效果肯定会很好。对于大部分公司，会选择廉价有效的方法扩展整个系统的架构，来增加服务器的吞吐量和处理能力，以及承载能力。

## ****3.服务器集群****

服务器集群就是指将N台服务器集中起来一起进行同一种服务，它们之间通过网络实现通信。让N台服务器之间相互协作，共同承载一个网站的请求压力。   
在客户端看来就像是只有一个服务器。集群可以利用多个计算机进行并行计算从而获得很高的计算速度，也可以用多个计算机做备份，从而使得任何一个机器坏了整个系统还是能正常运行。

## ****4.数据库读写分离****

基本的原理是让主数据库处理事务性增、改、删操作（INSERT、UPDATE、DELETE），而从数据库处理SELECT查询操作。

## ****5.数据库分表技术（垂直分割，水平分割）****

当一张的数据达到几百万时，你查询一次所花的时间会变多，如果有联合查询的话，很有可能会死在那儿了。分表的目的就在于此，减小数据库的负担，缩短查询时间。

## ****6.表建立相应的索引****

使用索引可快速访问数据库表中的特定信息。

## ****7.页面静态化****

前台实现完全的静态化最好，可以完全不用访问数据库。

## ****8.缓存技术（MemCache、Redis）****

缓存技术就是另一个解决方案，就是将动态数据存储到缓存文件中，动态网页直接调用 这些文件，而不必再访问数据库

## ****9.禁止外部盗链****

外部网站的图片或者文件盗链往往会带来大量的负载压力，因此应该严格限制外部对于自身的图片或者文件盗链

## ****10.控制大文件的下载****

大文件的下载会占用很大的流量，并且对于非SCSI硬盘来说，大量文件下载会消耗 CPU，使得网站响应能力下降。因此，尽量不要提供超过2M的大文件下载，如果需要提供，建议将大文件放在另外一台服务器上。

当一个网站发展为知名网站的时候（如新浪，腾讯，网易，雅虎），网站的访问量通常都会非常大，如果使用虚拟主机的话，网站就会因为访问量过大而引起 服务器性能问题，这是很多人的烦恼，有人使用取消RSS等错误的方法来解决问题，显然是下错药

对于当今大流量的网站，每天几千万甚至上亿的流量,是如何解决访问量问题的呢?以下是一些总结的方法:   
第一，确认服务器硬件是否足够支持当前的流量。   
普通的P4服务器一般最多能支持每天10万独立IP，如果访问量比这个还要大，那么必须首先配置一台更高性能的专用服务器才能解决问题，否则怎么优化都不可能彻底解决性能问题。   
  
第二，优化数据库访问。   
服务器的负载过大，一个重要的原因是CPU负荷过大，降低服务器CPU的负荷，才能够有效打破瓶颈。而使用静态页面可以使得CPU的负荷最小化。前台实现完全的静态化当然最好，可以完全不用访问数据库，不过对于频繁更新的网站，静态化往往不能满足某些功能。   
缓存技术就是另一个解决方案，就是将动态数据存储到缓存文件中，动态网页直接调用这些文件，而不必再访问数据库，WordPress和Z-Blog都大量使用这种缓存技术。我自己也写过一个Z-Blog的计数器插件，也是基于这样的原理。   
如果确实无法避免对数据库的访问，那么可以尝试优化数据库的查询SQL.避免使用Select \*from这样的语句，每次查询只返回自己需要的结果，避免短时间内的大量SQL查询。   
  
第三，禁止外部的盗链。   
外部网站的图片或者文件盗链往往会带来大量的负载压力，因此应该严格限制外部对于自身的图片或者文件盗链，好在目前可以简单地通过refer来控制盗链，Apache自己就可以通过配置来禁止盗链，IIS也有一些第三方的ISAPI可以实现同样的功能。当然，伪造refer也可以通过代码来实现盗链，不过目前蓄意伪造refer盗链的还不多，可以先不去考虑，或者使用非技术手段来解决，比如在图片上增加水印。   
  
第四，控制大文件的下载。   
大文件的下载会占用很大的流量，并且对于非SCSI硬盘来说，大量文件下载会消耗CPU，使得网站响应能力下降。因此，尽量不要提供超过2M的大文件下载，如果需要提供，建议将大文件放在另外一台服务器上。目前有不少免费的Web2.0网站提供图片分享和文件分享功能，因此可以尽量将图片和文件上传到这些分享网站。   
  
第五，使用不同主机分流主要流量   
将文件放在不同的主机上，提供不同的镜像供用户下载。比如如果觉得RSS文件占用流量大，那么使用FeedBurner或者FeedSky等服务将RSS输出放在其他主机上，这样别人访问的流量压力就大多集中在FeedBurner的主机上，RSS就不占用太多资源了。   
  
第六，使用流量分析统计软件。   
在网站上安装一个流量分析统计软件，可以即时知道哪些地方耗费了大量流量，哪些页面需要再进行优化，因此，解决流量问题还需要进行精确的统计分析才可以。我推荐使用的流量分析统计软件是GoogleAnalytics（Google分析）。我使用过程中感觉其效果非常不错，稍后我将详细介绍一下GoogleAnalytics的一些使用常识和技巧。

1、服务器环境的选择，建议采用Linux服务器

2、页面纯静态化，不要使用伪静态，大家都知道如果是非静态页面，都需要服务器端利用一定的CPU资源进行编译之后才能显示出来，而静态页面则直接显示，另外千万不能使用伪静态，伪静态实际上只是为了迎合搜索引擎而产生的一门技术，在使用过程中，会加大服务器的资源消耗，当访问量大的时候要慎用。

3、负载均衡技术，通俗点讲就是将网页上的元素分开放置到不同的服务器上来分担单台服务器的压力，如：数据库单独一台服务器，静态页面一台服务器，图片单独一台服务器等等，具体应该视情况而定，流量越大，就越应该分得越细，像国内的门户网站就是这样做的，看看他们的图片就能知道，图片都是用子域名或者其他域名引用在其他图片服务器的。

4、数据库优化，很多大型网站应用的时候大部分世界都花在这个上面，当网站数据变大，访问量上升，数据库首当其冲成为第一个需要解决的瓶颈，

　　　　--数据库性能调优

　　　　　　--1.聚集索引、主键

　　　　　　--2.尽量不要用临时表

　　　　　　--3.多多使用事务

　　　　　　--4.表设计要规范

　　　　　　--5.不要使用游标

　　　　　　--6.避免死锁

　　　　　　--7.不要打开大数据集

　　　　　　--8.最好不要select \*

　　　　　　--9.不要使用text数据类型，用varchar

　　　　　　--10.不要给诸如“性别”列创建索引

　　　　　　--11.不要使用Insert插入大量的数据

　　　　　　--12.尽量用join代替where,因为where进行全表搜索