# 构建高性能WEB站点

**知其然而不知其所以然是很多浮躁开发者的通病，也因此限制了其技术能力的提升和发展**

## 第一章 绪论

响应时间：数据在网络上的传输的时间总的来说包括两部分，即浏览器端主机发出的请求数据经过网络到达服务器的时间，以及服务器的回应数据经过网络回到浏览器端主机的时间。决定因素主要包括发送的数据量和网络带宽。

吞吐率(每秒处理请求数)：站点服务器处理请求并生成回应数据的时间主要消耗在服务器端，包括非常多的环节。

问题：

1.站点当前究竟使用了多少带宽？

2.这些带宽都用到哪里去了？

3.如何计算站点现在和可预见未来使用的带宽？

4.带宽增加后下载速度就可以加快吗？

5.使用独享带宽和共享带宽的本质区别是什么？

6.如何节省带宽？

7.什么是带宽？

缓存问题：

成千上万的缓存文件如何存储？

缓存命中率如何？

缓存的过期策略如何设计？

在拥有多台Web服务器的分布式站点上应用动态内容缓存需要考虑什么呢？

⑴减少网页中的http请求

①设计更加简单的网页，使其包含较少的图片和脚本，但是这可能牺牲了美观和用户交互。

②将多个图片合并为一个文件，利用CSS背景图片的偏移技术呈现在网页中，避免了多个图片的下载。

③合并JavaScript脚本或者CSS样式表。

④充分利用HTTP中的浏览器端Cache策略，减少重复下载。

⑵加快服务器脚本计算速度

APS.NET和JSP均有内置的优化方案，比如解释器对某个脚本程序第一次解释的时候，将中间代码缓存起来，一共下次直接使用。

⑶使用动态内容缓存

自动态内容技术产生后，聪明的工程师们为了减少动态内容的重复计算，想到了截取动态内容的胜利果实，将动态内容的HTML输出结果缓存起来，在随后的一段时间内当有用户访问时便跳过重复的动态内容计算而直接输出。

缓存带来的性能提升恰恰与有些动态数据实时交互的需求形成矛盾，这是非常尴尬的，而解决该问题的唯一途径不是技术本身，而你是如何权衡。

⑷使用数据缓存

比如对关系数据库中某字段的频繁更新和读取，这是我们为了提高缓存的灵活性和命中率，以及性能的要求，便开始考虑数据缓存。

⑸将动态内容静态化

在这种情况下缓存成为直接暴露给前段的HTML网页，而整个缓存机制也发生了根本的变化，我们普通称他为静态化，静态网页独立了，当家做主了，再也不用被脚本解释器胡来唤去。

⑹更换Web服务器软件

我们必须停止盲目的选择，停止对表面现象的崇拜，我们需要学习一些稍微底层的知识来武装自己。

⑺页面组件分离

⑻合理部署服务器

接入这些网络节点的局域网也可以相互通信，同时这些互联网之间也能通过骨干线路互通。在基于IP寻址的互联网中，IP地址相近的主机之间通信，数据经过少数的路由器即可到达。而如果通信的两端主机位于不同的运营商的互联网中，那么数据必须流经两个网联网运营商的顶级交换节点和骨干路线。显而易见，我们当然希望Web站点的用户和服务器位于同一个互联网运营商的网络内。

⑼使用负载均衡

到此为止，我们已经最大程度地发挥了单台Web服务器的处理能力，但是，当它所承受的压力达到极限时，就需要有更多的服务器来分担工作，我们需要想办法将流量合理转移到更多的服务器上。

为此，我们需要通过各种不同的方法来实现Web负载均衡，可能是简单的HTTP重定向，或者是基于DNS的轮询解析，或者通过反向代理服务器来实现负载均衡调度，还可以通过LVS来组建服务器集群。

⑽优化数据库

往往一些性能问题可能都发生在表现不佳的数据访问层面，这来源于不合理的应用程序数据访问组件设计、不合理的数据库表结构鸡舍以及对于数据库内部构造缺乏深入的了解。毫不夸张地说，也许你之前的优化全部白干了。

Web服务器与数据库服务器的数据通信一般基于标准的TCP，即便他们位于同一台物理主机也是如此。频繁的数据库连接和释放无疑将导致数据访问等待时间的加长，这段时间浪费得毫无意义。使用数据库持久连接有效地解决了这一难题，它包括不同程度上的持久化，本质的区别在于持久连接的应用范围和生命周期。还有跨进程的数据库连接池，保存了多个持久连接供应用程序重复使用。在这些采用数据库持久连接的应用设计中，同时还要注意保证数据访问的线程安全性。

索引的合理使用对于依赖数据库访问的Web应用至关重要。

随着时间的推移，你的Web站点可能逐渐被数据绑架，单台数据库服务器再也无法应付整个站点的需要，这包括存储空间以及查询时间。我们将数据散列在多台主机，包括必要的冗余数据，以此来合理地分散数据库的密集访问。

⑾考虑可扩展性

扩展对于我们来说，就像在山穷水尽的时候被指引了一条星光大道，一旦扩展度无法进行，那真是死路一条。

⑿减少视觉等待

实在不行就给用户一些提示吧！事实上，这不是什么大不了的事情，及时认识到架构的瓶颈并投入大量人力来改善，也不是一天两天就可以完成的，要意识到用户也许只是希望你不要不理他而已。

### 数据的网络传输

问题：你了解数据在网络传输中的一些原理和动机吗？比如存储转发、流量控制、带宽和响应时间等？

OSI七层网络模型或TCP四层网络模型。

电磁波的速度：铜线中电信号的传播速度大约为2.3x10^8m/s，光纤中光信号的传播速度大约为2.0x10^8m/s

45