

Web 缓存是可以自动保存常见文档副本的 HTTP 设备。当 Web 请求抵达缓存时，如果本地有“已缓存的”副本，就可以从本地存储设备而不是原始服务器中提取这个文档。使用缓存有下列优点。

- 缓存减少了冗余的数据传输，节省了你的网络费用。
- 缓存缓解了网络瓶颈的问题。不需要更多的带宽就能够更快地加载页面。
- 缓存降低了对原始服务器的要求。服务器可以更快地响应，避免过载的出现。
- 缓存降低了距离时延，因为从较远的地方加载页面会更慢一些。

本章解释了缓存是怎样提高性能降低费用的、如何去衡量其有效性以及将缓存置于何处可以发挥它的最大作用。我们还会解释 HTTP 如何保持已缓存副本的新鲜度，缓存如何与其他缓存和服务器通信等问题。

## 7.1 冗余的数据传输

有很多客户端访问一个流行的原始服务器页面时，服务器会多次传输同一份文档，每次传送给一个客户端。一些相同的字节会在网络中一遍遍地传输。这些冗余的数据传输会耗尽昂贵的网络带宽，降低传输速度，加重 Web 服务器的负载。有了缓存，就可以保留第一条服务器响应的副本，后继请求就可以由缓存的副本来应对了，这样可以减少那些流入 / 流出原始服务器的、被浪费掉了的重复流量。

## 7.2 带宽瓶颈

缓存还可以缓解网络的瓶颈问题。很多网络为本地网络客户端提供的带宽比为远程服务器提供的带宽要宽（参见图 7-1）。客户端会以路径上最慢的网速访问服务器。如果客户端从一个快速局域网的缓存中得到了一份副本，那么缓存就可以提高性能——尤其是要传输比较大的文件时。

[161]

在图 7-1 中，Joe 的五金商店旧金山分店的用户通过 1.4Mbit/s 的 T1 因特网连接，从亚特兰大总店下载一个 5MB 的库存文件要花 30 秒的时间。如果在旧金山分店里缓存了这个文档，本地用户通过以太网连接只要花费不到 1 秒的时间就可以获得同一份文档了。

表 7-1 说明了在几种不同的网速下，传输几种不同大小的文档时，带宽会对传输速度产生什么样的影响。带宽会给较大的文档带来显而易见的时延，不同类型网络的速度差异会非常明显。<sup>1</sup> 一个 54kbit/s 的 Modem 传输一个 5MB 的文件需要 749 秒（超过 12 分钟），而在快速以太网 LAN 中，只要不到一秒的时间。

注 1：这张表只列出了网络带宽对传输时间的影响。它假定网络效率为 100%，而且不存在网络或应用程序的处理时延。通过这种方式给出的时延是下限值。实际的时延要大一些，而小型对象的时延则主要是由非带宽开销造成的。