

几乎所有的服务器和代理都会记录下它们所处理的 HTTP 事务摘要。这么做出于一系列的原因：跟踪使用情况、安全性、计费、错误检测，等等。本章简要介绍了日志记录，研究了通常会记录 HTTP 事务哪些方面的信息以及一些常见日志格式中所包含的内容。

21.1 记录内容

大多数情况下，日志的记录出于两种原因：查找服务器或代理中存在的问题（比如，哪些请求失败了），或者是生成 Web 站点访问方式的统计信息。统计数据对市场营销、计费和容量规划（比如，决定是否需要增加服务器或带宽）都非常有用。

可以把一个 HTTP 事务中的所有首部都记录下来，但对每天要处理数百万个事务的服务器和代理来说，这些数据的体积超大，很快就会失控。不应该记录实际上你并不感兴趣，甚至从来都不会去看一眼的数据。

通常，只记录事务的基本信息就行了。通常会记录下来的几个字段示例为：

- HTTP 方法；
- 客户端和服务器的 HTTP 版本；
- 所请求资源的 URL；
- 响应的 HTTP 状态码；
- 请求和响应报文的尺寸（包含所有的实体主体部分）；
- 事务开始时的时间戳；
- Referer 首部和 User-Agent 首部的值。

483

HTTP 方法和 URL 说明了请求试图做些什么——比如，GET 某个资源或 POST 某个定单。可以用 URL 来记录 Web 站点上页面的受欢迎程度。

版本字符串给出了与客户端和服务器的有关的一些提示，在客户端和服务器的出现一些比较奇怪或非预期的交互动作时，它会非常有用。比如，如果请求的失败率高于预期，那版本信息指向的可能是一个无法与服务器进行交互的新版浏览器。

HTTP 状态码说明了请求的执行状况：是否成功执行，认证请求是否失败，资源是否找到等（HTTP 状态码列表参见 3.2.2 节）。

请求 / 响应的大小和时间戳主要用于记账；就是记录流入、流出或流经应用程序的字节有多少。还可用时间戳将观察到的问题与当时发起的一些请求关联起来。