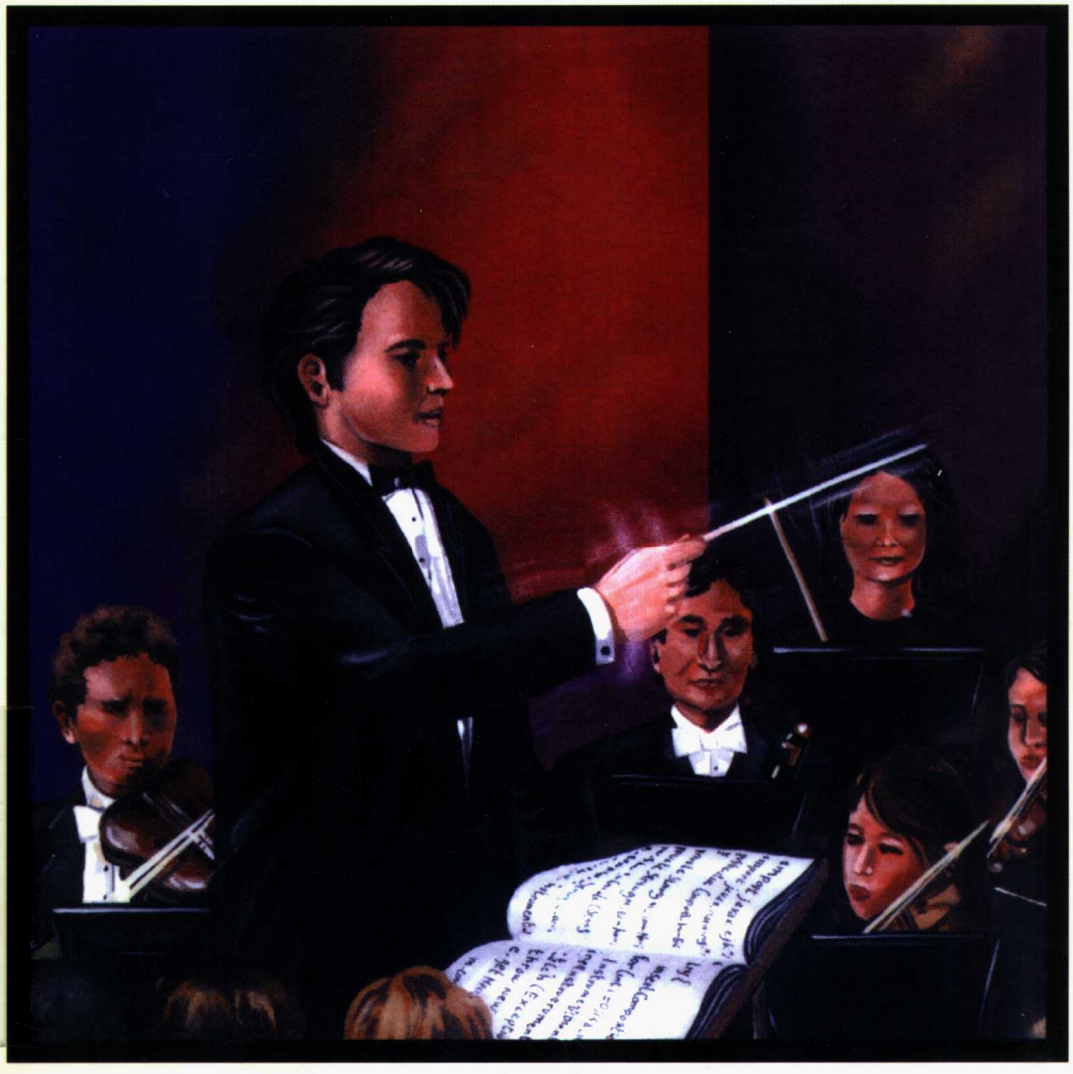
Building Scalable and High-Performance Java Web Applications Using J2EE Technology

J2EE Web 应用高级编程

Greg Barish 著　林琪 英宇　译



清华大学出版社
http://www.tup.com.cn
http://www.tup.tsinghua.edu.cn

J2EE Web 应用高级编程

Greg Barish 著

林 琪 英 宇 译

清 华 大 学 出 版 社

(京) 新登字 158 号
北京市版权局著作权合同登记号：01-2002-3007

内 容 简 介

本书介绍的是用 J2EE 技术构建可扩展和高性能的 Java Web 应用的知识,重点阐述了用 J2EE 规范构建高效的端到端解决方案。本书首先介绍了与可扩展和高性能的 Web 应用有关的基础性知识；然后讲解了用 J2EE 构建 Web 应用的方法和技术，涵盖了客户/服务器通信、客户请求处理、应用服务器的构建和企业应用集成以及数据库设计和操作等知识，对于涉及到的每一种相关技术，均从其可扩展性和性能角度进行了总结和分析；最后本书还对 Web 服务作了总体性的介绍。

本书内容浅显易懂、示例生动、代码丰富，适合于 Web 开发人员和 Java 程序员阅读。

图书在版编目(CIP)数据

J2EE Web 应用高级编程/(美)巴里希著;林琪,英宇译.—北京：清华大学出版社，2002
书名原文：Building Scalable and High-Performance Java Web Applications Using J2EE Technology
ISBN 7-302-05887-3
Ⅰ.J…Ⅱ. ①巴…②林…③英…Ⅲ. JAVA 语言程序设计 Ⅳ. TP312

中国版本图书馆 CIP 数据核字(2002)第 071698 号

出 版 者：清华大学出版社(北京清华大学学研大厦，邮政编码：100084)
http://www.tup.com.cn http://www.tup.tsinghua.edu.cn
责任编辑：陈宗斌
印 刷 者：北京昌平环球印刷厂
发 行 者：新华书店总店北京发行所
开　　本：787×1092 1/16 **印张**：18.5 **字数**：473 千字
版　　次：2002 年 10 月第 1 版 2002 年 10 月第 1 次印刷
书　　号：ISBN 7-302-05887-3/TP・3493
印　　数：0001～4000
定　　价：38.00 元

前　言

构建 Web 应用时，有两个最基本的问题。首先是开发应用的选择实在太多，“条条大路通罗马”。目前，可以在数以百计的语言和技术中任意挑选，它们都声称能够使 Web 应用的开发更为容易。另一个问题是：如何在部署应用时使之具有高性能和可扩展性没有达成共识，而这却是任何可通过 Internet 访问的 Web 应用所必须具有的共性。

幸运的是，现在已经有了一些统一应用的基础架构，它们不仅可以简化 Web 应用的开发，而且还能够支持有效的部署。其中最为常用的方法之一当属 Java 2 企业版(Java 2 Enterprise Edition，J2EE)规范。J2EE 规范分析并扩展了 Java 技术现有的基础，从而使技术人员可以构建支持高性能和可扩展的 Web 应用。J2EE 最为诱人之处在于它对于事务管理、命名和安全等主要的低级平台服务提供了内置的支持，而以往这些服务通常都需要花费大量的设计时间，而且还需要专门的技术方能实现。

尽管 J2EE 将大多数 Web 应用开发的部署进行了统一，但是理解起来却是一件会让人头疼的事情。为了理解 J2EE，要么需要阅读规范本身，要么必须博览众多关于各种特定 J2EE 技术的书(如 Servlet、EJB 和 JDBC 等)。规范本身相当枯燥，利用这一资源往往不具备实践性。如果要把关于每一种技术的所有书都读完，所带来的问题就是花去的时间将会长得使人难以接受，要知道，一些仅仅是介绍 EJB 的书甚至就超过了 1000 页。而且这些书中往往充斥着一些关系不甚紧密的细节问题，还常常会有一些不需要的附录，或者是附录中的内容可以在网上轻松地找到。最后，这些书中一般都没有提及 Web 应用设计的一些不太受关注的领域，例如，数据库设计和网络效率等。

本书目标

对于如何利用 J2EE 技术构建高性能和可扩展的 Web 应用，需要有一个实用的概括性总结，而编写本书的目的就是要满足这一需求。在此需要达到一种巧妙的平衡，即一方面要介绍一系列全新的技术以及它们之间的相互关系，还要提供足够的示例以使读者确实了解到 J2EE 究竟如何工作。另一方面又不能过于宽泛，以致于变成一本既没有重点而又厚重的大书(比如，重得甚至于有人会因为举这本书而受伤)。因此，我们的目标很明确，就是要提供一个言简意赅而又实用的概要性介绍。

书中将覆盖 J2EE 的所有主要元素——其规范本身、Servlet、JSP、EJB、消息、JDBC，还有其他一些内容。全书中会提供大量的示例。示例表述得相当简洁，这一点是故意所为，我们的想法是如此精简可以使读者更好地理解如何使用一种技术，而不会因为过于深入到繁杂的细节中去从而把读者搞得头昏眼花。应该知道，关于这些详细的内容，大量的书中都已经有所介绍。如果对 J2EE 中某一部分特别感兴趣，可以参考这些书作为本书的补充材料。

除了要提供一个面面俱到的概括之外，本书还有一个目标，就是想填补有关 Web 应用设计

的空白，它在 J2EE 规范里是没有提及的。尽管规范中说明了如何将不同 Java 技术和 API 加以连接从而构造一个企业应用基础架构，但是并没有谈到相关的一些问题，如网络和数据库设计等。例如，虽然规范中描述了 HTTP 可用于与一个 J2EE 系统通信，但是关于 HTTP 和关系数据库的细节却只字未提。作为一个经验丰富的 Web 应用设计者来说，应该知道设计系统时若要保证高性能和可扩展性，这两者都是至关重要的。

总而言之，本书有如下目标：

- 定义并指出构建可扩展的和高性能 Web 应用时相关的诸多问题。
- 提供一个 J2EE 技术“导航图”，以便按图索骥来设计 Web 应用。
- 简要地描述主要的 J2EE 技术，并重点强调与高性能和可扩展性相关的细节。
- 填补 J2EE 规范所遗漏的 Web 应用设计的空白，如与 HTTP 和数据库设计相关的重要细节内容，这也是与 J2EE 技术有关的最为常用的两个方面。
- 指出不同特定设计方法的优点，并利用实际的性能图表来描述它们之间存在的差异。

最后一点是为了使本书内容更加受到关注。例如，如果说“连接池很好”，这种说法当然也不错，而这也正是很多书所惯用的描述方式。但是如果能够为这一论述提供实际的性能图表作为“证据”，那么将会更具有说服力，同时也更为清晰。本书就是要达到这个目的。

本书面向的读者

对于熟悉 Java 的技术人员和设计人员，如果希望构建基于 Java 的 Web 应用从而得到高性能和可扩展性，但同时又对 J2EE 如何用于这些目标以及其底层技术的工作原理不甚了解，那么本书正是为此而编写的。

对于希望了解 J2EE 规范未涉及到的一些内容的读者，这本书同样适用，在此特别强调了目前已经存在的、高效的网络和数据库设计等问题，还讨论了与 XML 和 SOAP 等 Web 服务技术相关的问题，它们在将来可能会出现。

最后，这本书还面向已经对部分 J2EE 技术(如 Java Servlet)比较熟悉，但不了解其他技术(如 Java 消息服务)的读者。

有关性能测量的说明

本书中贯穿着多种性能测量和比较的描述。尽管对于每一种体系结构来说，几乎都有着同样的一般趋势(这是因为有关性能趋势的描述是与体系结构无关的)，不过在此把用于测试系统的详细情况列出还是很有帮助的。

所有测试都是在一台带有单个 CPU833 MHz Pentium III，并有 256KB RAM 的 Dell Latitude 计算机上完成的。操作系统和应用软件包括：

- Windows 2000，Professional 版本。
- Apache Web 服务器，1.3.14 版本。
- Java 运行时环境和开发工具包(Java Runtime Environment and Development Kit)，1.3 版本。

- J2EE SDK 及参考实现，1.3 版本(Beta 版)。
- Apache Jakarta/Tomcat Servlet 容器(Apache Jakarta/Tomcat Servlet Container)，3.2.1 版本。
- Oracle 数据库系统，8.1.6 版本。

开始吧

我希望这本书能够在进行 J2EE 应用开发时提供一个全程的参考。对于如何向您的客户提供应用功能(这其中既包括个人也包括其他企业)，可能需要不断地做出选择。您也许会提出这样的问题，如："业务逻辑是放在数据库服务器上好呢，还是放在应用服务器上？"，或者是"我们的批数据需要通过一个 Web 服务器传输呢，还是通过一个消息服务器传输？"。暂不考虑与这些选择有关的详细内容，您肯定希望以某种方式利用这些特性从而提高性能和可扩展性。本书就是要帮助您理解如何在 Web 应用设计中做出权衡，这包括一般性的和专用的方法，从而有助于您达到目标。

目　录

第1章　可扩展和高性能Web应用

1.1　Web应用的出现

仅仅是在几年之间，在世界范围内，无论是信息的提供方式还是使用方式都因 Internet 而发生了改变。其硬件和软件技术使得每个人不仅能够成为信息的使用者，而且几乎所有人都能够作为信息的提供者。Internet—— 特别是 World Wide Web(Web)—— 在非常短的时间内就已经被公众认为是重要的信息共享的平台，许多组织都在尽力创建有用的 Web 应用，从而为使用者提供更大的价值。

这些 Web 应用允许使用者在线地购买书和光盘。它们使得企业可以使用 Internet 来实施安全的事务处理。工人利用 Web 应用寻找工作；老板利用 Web 应用寻找雇员；使用由经纪人提供的在线应用可以购进和抛售股票；旅行者则可以利用 Web 应用来预订机票和宾馆。这样的例子还有很多很多。很明显，目前无论是在公共的 Internet 上，还是在数不胜数的公司内部网(Intranet)中，都存在着大量有用的 Web 应用。

本书介绍构建具有高性能和可扩展性的企业 Web 应用所需的通用技术。一般来说，这表示所构造的应用不仅要有合理、稳定快速的运行速度，而且对于不断增长的用户和请求需求还应具有较强而渐增的容许性。尽管我们会花很多时间来考虑这个一般性的问题，但是我们要讨论的核心却是围绕着 Java 2 企业版(Java 2 Enterprise Edition，J2EE)规范来建立一个解决方案。在对建立这一类应用的具体细节深入分析之前，就目前而言，很重要的一点是需要明确和理解整个问题。更准确地说，就是要定义 Web 应用和可扩展性，这是非常重要的。

1.1.1　基本定义

本书中，Web 应用有一个非常通用的定义—— 这是一种通过 Internet 技术加以连接的客户/服务器软件，可以传输其处理的数据。通过“Internet 技术”，我所指的是在信息的使用者和提供者之间，组成相应网络基础架构的硬件和软件的集合。Web 应用可以通过专门的客户端软件来访问，也可能利用一个或多个有关的 Web 页面访问，这些页面基于某种特定的用途可进行逻辑分组。这里所说的用途可以是任意一件事情，例如，可以是买书、处理股票订单，也可能仅仅是作为让使用者阅读的内容。

注意，我们所讨论的是 Web 应用，而不只是“Web 网站”。实际上，这二者之间的区别对于理解本书的一个关键主题相当重要。大多数非技术人员往往不会区别 Web 网站和 Web 应用。除了术语的提法不同之外，对他们而言，其作用都是一样的，都可以使之实现在线购书、在线预订机票、在线购票等操作。

不过，如果您是一名技术人员，那么这两者之间还是存在差别的。对您来说，如果有人谈

到类似于一个 Web 网站的性能问题时，您可能会开始想到其后台的具体细节。我也不例外。您会考虑所运行的是 Apache 还是 IIS，还有它使用的是 Java servlet、PHP 还是 CGI-bin Perl 脚本。技术人员和非技术人员在思路上所存在的不同可能会造成一定的误解。技术人员，往往习惯性地把“Web 网站”与服务器端联系起来。而与此同时，我们都知道，Web 应用并不仅仅包括服务器端；它还需要有网络和客户端。因此，从这个意义上说，一个 Web 网站(服务器)与 Web 应用(客户机、网络和服务器)并不是一回事。

虽然这本书所强调的是服务器端解决方案，在此还会讨论客户端和联网的有关内容，这是因为对于终端用户如何理解 Web 应用，它们有着很重要的影响。也就是说，我们会讨论 Web 网站内端对端的交互，它意味着由客户到服务器，再返回到客户。这自然是一个重点。毕竟，大多数使用 Web 的人所关心的就是它的端到端的操作。如果在网上购买音乐会的门票需要花费他们较长的时间，则造成这一现象的原因可能有多种，可能是因为一个速度较慢的调制解调器(MODEM)所致，或者原因在于一个负载过重的服务器，也许还可能是由于网络阻塞所造成。无论原因如何，结果都是一样的，即“慢吞吞”的应用会耗费大量的时间。作为技术人员，我们所关注的将不仅仅是对于一个用户应用可能比较慢，更重要的则是随着访问量的增加，系统也许会越来越慢。

既然我们对于 Web 应用的范围已经有了更为明确的认识，下面再来看它的核心组件。任何一个在线应用都有一些主要的组成部件，每个部件即代表着一种可能性——也许是一个问题，也许是一个挑战，这要取决于您如何看待它。可能您对这些组件已经比较熟悉了，不过在此确保每个人都明白是一件有益而无害的事情，特别是这些术语将贯穿本书始终。下面就先从客户端开始见图 1-1。

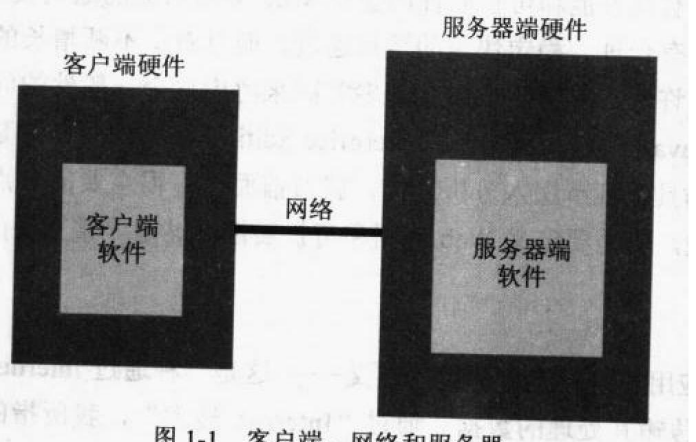


图 1-1 客户端、网络和服务器

我们可以说使用者是通过客户端软件(即使用 Web 来检索和处理数据的 Web 浏览器或应用)来使用 Web 应用的，这些客户端软件运行在客户端硬件(即 PC、PDA 等等)之上，而应用数据的提供以及生产者对具体处理的控制则均通过服务器端软件(即 Web 服务器、服务器端组件软件、数据库等)来实现，这些服务器端软件运行于服务器端硬件(即高端多处理器系统、集群等等)。将客户端与服务器进行连接(从调制解调器或客户端设备的网络端口到服务器端的网络设备)就形成联网基础架构。图 1-1 用图的形式描述了客户端/服务器的关系。注意服务器端要大一些，通常情况下，我们假设服务器端有更多可支配的资源。

在此，需要强调的很重要的一点，即作为服务器软件的一种，要把 Web 服务器特别区别出

来，因为它在调度客户端与服务器之间的通信(HTTP 流量)时总是起着一个核心的作用。在本书中，当提到“服务器端”时，一般都包括 Web 服务器。如果有必要将它与服务器端的其他软件相区别时，我们还会专门指出。

1.1.2　Web 的本质特性及其挑战

尽管由于 Web 应用的存在使得 Internet 迅速成为一个应用广泛的媒体，不过，不理解 Internet 的本质特性，也会造成一些技术上的难题。即使是最基本的问题，例如，使信息的提供者快速而可靠地为所有需要的人发送相应的信息，这个问题就既不简单也不好理解。与其他挑战一样，这个问题的复杂性在于必须处理 Internet 媒体的本质特性。基于许多原因，Internet 与诸如广播、电视和报纸等其他信息共享模式存在着显著的区别。这其中最为重要的两个原因可能是，一方面它所面向的群体实在过于广泛 (客户的数量不可估计)，而另一方面，所面向的客户群体在任何时刻都可能对任意的提供者提出信息请求(工作需求量不可估计)。

与其他媒体不同，Internet 信息提供者不具有预先了解其客户的能力。例如，报纸能够在印刷每一期之前知道它的发行量。另外，这些媒体还有一个优势，就是可以控制其增长速度，以确保每天有足够的员工来发放报纸，而且在最后期限之前的头一天晚上可以有足够的资源和时间来准备第二天早上的报纸发放工作。除此之外，报纸也不会遇到发行量突然跳跃式增长的情况。与 Internet 相比，即使是大城市的大型报纸，其增长也不过是渐进式的。例如，《华盛顿邮报》始创于 1877 年，当时其发行量为 10000。到了 1998 年，每天的发行量已经接近了 800000 份，而且周日版的发行量还要多。*其平均增长速度为每年不少于 6500 份，即每天 17 份。

Web 应用的开发人员对于 Web 应用的增长速度却是既爱又恨。一方面，如果每天都有 17 位新的用户，他们当然希望这样的增长速度。要是能够为增长过快而考虑适当扩展，该是多么幸福！这样您就不用每天工作到晚上 9:30，而是在下午 5 点就轻松地回到家里。同时，如此高的增长率也正是人们对 Web 应用这样着迷的原因之一，因为利用它只在几秒钟之内就可以达到世界上的几乎每一个角落。Web 应用的增长率可能达到数百万之多。尽管在企业方面这是一个很好的预兆，但同时也带来一个严重的问题，该如何处理这么多的需求呢？

在 Internet 上，发行量就相当于“页面的点击率”、也就是说，对于某个文档的请求数目。页面点击率可能会在一夜之间出现“疯长”。 Starr 报告的在线公布应该算是 Web 领域中对此最适当的一个例子。因为大多数美国人都知道，这个报告是克林顿在任期间由独立顾问处所整理的。我们可以毫不夸张地说，无论是美国公众还是国际报业集团都迫不及待地想对其有所了解。

Starr 报告于 1998 年夏天在政府 Web 网站上首次在线发布时，数以万计的人都试图进行下载。作为 Internet 主要 ISP 之一的 Sprint 公司，其代表报告称带宽需求高峰值较之平常要高出 10 到 20 个百分点；AOL 的一个代表则表示“突然出现了 30 个百分点的增长”；NetRatings，这是一个像 Nielsen 一样的 Internet 上知名度很高的公司，据它估计，某些时候，每 5 个 Web 用户中就有一个在请求此报告，或者是相关的新闻。关于这一事件以及 1998 年秋天有关 Internet 可扩展性的林林总总，CNET.COM 上有大量相应的报道。

---

* 资料来源：http://www.thewashingtonpost.com