

Web容量规划之美

利用云端扩展Web容量

The Art of Capacity Planning

Arun Kejariwal John Allspaw 著

王林平 张丽媛 译

第二版

中国电力出版社

Web容量规划之美

第二版

Arun Kejariwal 和 John Allspaw 著
王林平 张丽媛 译

Beijing • Boston • Farnham • Sebastopol • Tokyo

O'REILLY®

y Media, Inc. 授权中国电力出版社出版

Copyright © 2018 Arun Kejariwal, John Allspaw. All rights reserved.

Simplified Chinese Edition, jointly published by O'Reilly Media, Inc. and China Electric Power Press, 2019.
Authorized translation of the English edition, 2018 O'Reilly Media, Inc., the owner of all rights to publish and sell the same.

All rights reserved including the rights of reproduction in whole or in part in any form.

英文原版由 O'Reilly Media, Inc. 出版 2018。

简体中文版由中国电力出版社出版 2019。英文原版的翻译得到 O'Reilly Media, Inc. 的授权。此简体中文版的出版和销售得到出版权和销售权的所有者——O'Reilly Media, Inc. 的许可。

版权所有，未得书面许可，本书的任何部分和全部不得以任何形式重制。

图书在版编目 (CIP) 数据

Web容量规划之美：第二版 / (美)阿伦·科加里瓦尔 (Arun Kejariwal), (美)约翰·奥斯帕 (John Allspaw) 著; 王林平 张丽媛译. — 北京: 中国电力出版社, 2019.8

书名原文: The Art of Capacity Planning, 2nd Edition

ISBN 978-7-5198-3107-3

I. ①W… II. ①阿… ②约… ③王… III. ①网络服务器—容量—规划 IV. ①TP368.5

中国版本图书馆CIP数据核字(2019)第076249号

北京市版权局著作权合同登记 图字: 01-2018-7458号

出版发行: 中国电力出版社

地 址: 北京市东城区北京站西街19号 (邮政编码100005)

网 址: <http://www.cepp.sgcc.com.cn>

责任编辑: 刘炽 (liuchi1030@163.com)

责任校对: 黄蓓 太兴华

装帧设计: Karen Montgomery, 张 健

责任印制: 杨晓东

印 刷: 北京天宇星印刷厂

版 次: 2019年8月第一版

印 次: 2019年8月北京第一次印刷

开 本: 750毫米×980毫米 16开本

印 张: 14.5

字 数: 268千字

印 数: 0001—3000册

定 价: 58.00元

版 权 专 有 侵 权 必 究

本书如有印装质量问题, 我社营销中心负责退换

O'Reilly Media, Inc.介绍

O'Reilly Media通过图书、杂志、在线服务、调查研究和会议等方式传播创新知识。自1978年开始，O'Reilly一直都是前沿发展的见证者和推动者。超级极客们正在开创着未来，而我们关注真正重要的技术趋势——通过放大那些“细微的信号”来刺激社会对新科技的应用。作为技术社区中活跃的参与者，O'Reilly的发展充满了对创新的倡导、创造和发扬光大。

O'Reilly为软件开发人员带来革命性的“动物书”；创建第一个商业网站（GNN）；组织了影响深远的开放源代码峰会，以至于开源软件运动以此命名；创立了《Make》杂志，从而成为DIY革命的主要先锋；公司一如既往地通过多种形式缔结信息与人的纽带。O'Reilly的会议和峰会集聚了众多超级极客和高瞻远瞩的商业领袖，共同描绘出开创创新产业的革命性思想。作为技术人士获取信息的选择，O'Reilly现在还将先锋专家的知识传递给普通的计算机用户。无论是通过书籍出版，在线服务或者面授课程，每一项O'Reilly的产品都反映了公司不可动摇的理念——信息是激发创新的力量。

业界评论

“O'Reilly Radar博客有口皆碑。”

——Wired

“O'Reilly凭借一系列（真希望当初我也想到了）非凡想法建立了数百万美元的业务。”

——Business 2.0

“O'Reilly Conference是聚集关键思想领袖的绝对典范。”

——CRN

“一本O'Reilly的书就代表一个有用、有前途、需要学习的主题。”

——Irish Times

“Tim是位特立独行的商人，他不光放眼于最长远、最广阔的视野并且切实地按照Yogi Berra的建议去做了：‘如果你在路上遇到岔路口，走小路（岔路）。’回顾过去Tim似乎每一次都选择了小路，而且有几次都是一闪即逝的机会，尽管大路也不错。”

——Linux Journal

目录

前言	1
第1章 容量规划的目标、问题和过程	11
背景	11
准备	12
快捷但不太好的数学	15
预测你的系统何时会崩溃	16
用系统统计表呈现问题	18
买东西，采购是一个过程	25
性能与容量，两种不同的概念	27
社交网站与开放式API的影响	29
延伸阅读	31
关键章节	31
参考资料	32
第2章 设定容量目标	35
不同种类的需求和测量方法	36
解读正式测量	36
服务等级协议	40
业务容量需求	42
用户期望	43

架构决策	45
提供测量点	46
资源瓶颈	48
硬件决策（垂直、水平和对角扩展）	49
灾难恢复	53
延展阅读	54
参考资料	54

第3章 测量：容量的单位 57

容量跟踪工具	63
度量指标采集系统基础	65
轮询数据库和轮询数据库工具	66
Ganglia	66
SNMP	67
将日志视为过去的度量指标	67
监控可作为识别紧急问题的工具	68
网络测量和规划	69
负载均衡	71
应用程序监控	73
应用层的测量	73
存储容量	75
数据库容量	87
找到数据库上限	91
缓存系统	93
缓存效率：工作集和动态数据	93
确认缓存系统上限	96
一个真实的示例：缓存测量	97
特殊用途和多用途服务器	102
API的使用率及其对容量的影响	106
示例和现实	109
小结	109

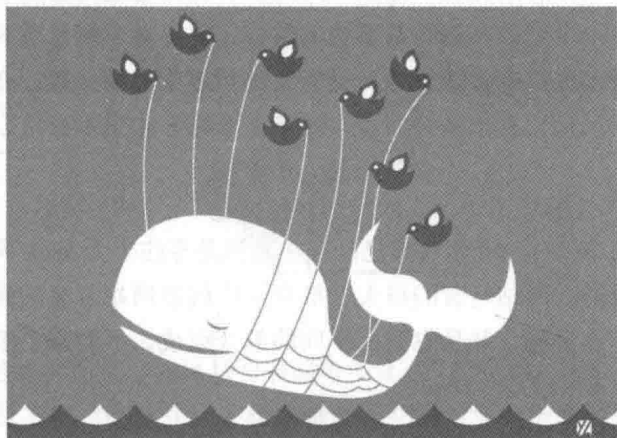
延展阅读	110
性能	110
网络	111
负载均衡	112
存储	112
数据库及缓存	112
参考资料	113
第4章 趋势预测	115
曲线拟合	117
趋势、曲线、时间	118
将应用程序级度量指标和系统统计进行关联： 数据库实例	123
预测峰值驱动的资源使用量：Web服务器示例	126
小数据集的附加说明	129
自动预测	130
安全因素	133
采购	135
采购时间：关键性度量指标	136
适时库存	136
增加容量后的影响	138
长期趋势	139
流量模式变更	139
应用程序使用量变更和产品规划	142
迭代和校准	143
最好的猜测	145
对角线扩展时机	146
小结	146
延展阅读	147
采购或租赁	147
时间-序列预测	148

曲线拟合	148
测量	148
参考资料	148
第5章 部署	149
自动化部署基本原理	150
目标1：最小化提供新容量的时间	150
目标2：变化集中分布	150
目标3：决不登录单个服务器（针对管理）	151
目标4：使新的服务器自动工作	152
目标5：在简单的故障检测中保持一致性	152
自动化安装工具	153
准备操作系统镜像	157
安装过程	158
自动化部署	159
定义角色和服务	160
示例1：分离静态网页内容	161
用户管理和访问控制	162
决策	163
示例2：多数据中心	164
小结	165
延伸阅读	165
参考资料	166
第6章 自动扩缩容	167
挑战	168
使用Amazon EC2进行自动扩缩容	172
设计指南	174
要求	175
固定数量扩缩容	176
扩缩容百分比	179

百花齐放	188
先进的方法	190
小结	190
延展阅读	192
参考资料	194
 附录A 虚拟化	 197
附录B 处理瞬时增长	205
附录C 容量工具	213

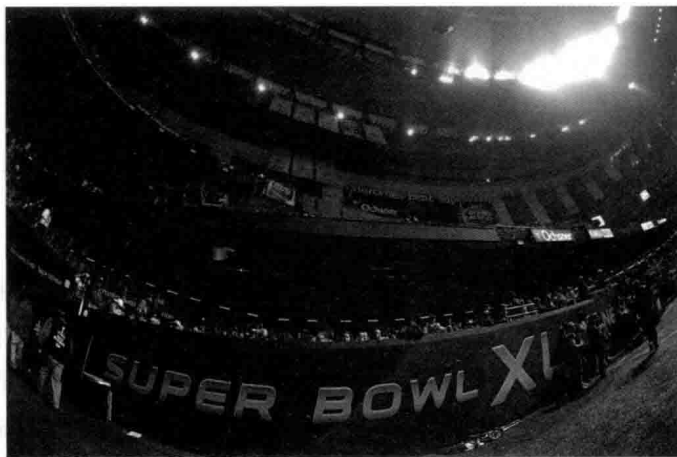
前言

在某年国际足联世界杯之前，Twitter 上最常见的讨论话题是，在上届国际足联世界杯期间，Twitter 服务总是不可用。特别是每当巴西或日本在比赛中打入一球时，Twitter 服务的访问高峰会导致服务不可用。失败鲸^{译注 1} 在 Twitter 早期的可用性问题上总是会出现。因此，某年国际足联世界杯期间 Twitter 服务的目标之一是无故障、无停机持续服务。此外，另一个关键目标是确保 Twitter 移动应用程序在进行 Twitter 照片分享或其他对性能要求较高的操作能够保证高性能。如何去实现这个目标呢？



译注 1：失败鲸（Fail Whale）是微博客网站 Twitter 宕机时的标志画面，由华裔设计师 Yiyang Lu 设计，目前在全世界正火热流传。

类似前面的情况，如像超级碗这样的大型赛事会大幅度增加推特的使用，另一个重点在于 Twitter 微博转发、收藏、DMS 高峰等流量拥堵的情况下保证 Twitter 的高可用性。可以想象，我们可以分析过去的高峰流量相对于正常流量的比例，然后提出削减幅度的预估峰值持续运行。话虽如此，我们应该部署这样的容量能力来应对这种的一次性事件吗？特别是考虑到这种容量很可能在一年中的大部分时间是比较空闲的。你如何处理意外事件，假设在某年超级碗期间发生断电？确保在此类事件中服务的高可用性需要设计针对体系结构和容量规划的系统方法。



容量规划自古就有，从经济学到工程学等领域都有应用。通俗地讲，容量规划就是资源管理，当资源有限且有一定成本的时候，就需要进行容量规划。当一家工程建设公司设计一个新的高速公路系统的时候，需要对车辆承载容量进行规划，正如一个为大城市提供电力的能源公司需要进行容量规划一样。在某些方面他们关注的重点和网站运营有很多共同点，许多基本概念和重点都可以应用于这三门科学。

虽然系统管理在 20 世纪 60 年代就已经存在，但是专注于为 Web 网站提供服务的分支还是新出现的。网站运营的很大一部分工作就是网站容量的规划和管理。这些只是过程而不是目标，并且它们由不同的部分组成。尽管每个组织采用的方式各不相同，但基本原理是一样的：

- 确保有合理的资源（服务器、存储设备、网络等）能用于处理预期或者突发的流量高峰。

- 有明确的采购和审批机制。
- 准备好调整资金支出，以支持业务需求。
- 有适当的部署和管理系统用于管理部署后的资源。

为什么编写和修订本书

运维工程师和软件开发工程师经常会面临一个难题，那就是无法通过任何途径获取到关于计算保持网站或移动应用程序正常运行所需的容量的帮助信息。现有书籍中关于计算机容量规划的内容更偏重于数学理论上的自研规划，而不是整个过程的实际执行（参阅附录 C）。此外，当今敏捷开发环境的现状是，容量规划是一个连续的过程，应该灵活运用且适应当前的业务情况。基于静态理论模型的容量规划将无法达到预期的效果。

许多著作只讲解了 Web 网站用例的基本模型，缺乏具体的信息或建议。相反，它们倾向于只提供设计出来用于阐述排队理论原理的数学模型，这些数学模型是传统容量计划的基础。这种方法也许在数学上有趣的“优雅”（决定何种程度的流量峰值是有用的、可“吸收”的各种服务，由于内置的队列，而不影响网站 / 移动 App 的可用性），但是当运维工程师或开发人员得知只有一周时间来为不可预知的额外流量做准备（可能是由于某一项超棒的新功能发布带来的）时，或者当他看到自己的网站因为 Facebook、纽约时报、Reddit、Digg 等网站首页链接带来的流量负荷而即将崩溃的时候，这种方法对于他们并没有太多帮助。



我们发现大部分有关网站容量规划的书都暗含着这样的假设：在非互联网环境（比如制造业或工业工程）中建立的过程和概念在互联网环境中同样适用。尽管这类与规划等相关的理论可能确实相似，但是这些概念的实际应用并不能很好地适应网站的短期发展。据我们观察，大多数 Web 开发环境的迭代很快、很频繁，以至于不能像其他领域一样可以进行详细且严格的容量规划工作。当运维工程师构建

了系统的排序模型的时候，系统的新版本已经部署完毕，并且新版本的使用方式可能已经发生了巨大的变化。在 2016 年计算机协会（ACM）一篇题为“为什么谷歌在一个代码仓库中存储了数十亿行代码？”（作者 R. Potvin 和 J. Levenberg，都是谷歌员工）中提道：

开发工程师正常情况下每天提交 16000 行代码变化到代码库，另有 24000 行代码更新是由自动化系统提交的。

又或者，一些在科技领域、社会方面或现实世界发生的其他事情使得所有建模和仿真变得过时了。

与工业界人士的探讨让我们受益良多，因为他们也同样遇到了很多关于可扩展性和容量方面的问题。随着时间的推移，我接触过许多不同的公司，每家公司采用不同的架构，每一种架构都遇到不同的问题。但是，他们采用的解决方案非常相似，我们希望能在本书中介绍其中一些解决方案。自从本书第一版出版以来，计算机领域发生了翻天覆地的变化。2009 年云计算开始出现，如今像 AWS 和 Azure 这样的公共云都已经发展到超过百亿级别。因此我们在第一版没有太关注云计算领域。同理，公有云今天能提供一个更大的各种实例的类型，包括图形处理单元（GPU）和现场可编程门阵列（FPGA），基于实例虽然有利于驱动更高的运营效率和任务，但对给定服务的实例类型的“最佳”选择任务变得更具挑战性。跨区域 IDC 部署公有云的使用不断增长，尽管从容灾的角度来看很好，但也会直接影响你对容量规划的关注程度，数据复制与负载均衡相关处理会受跨区域 IDC 部署影响。

早在 2009 年，面向服务的架构（SOA）应用还不是很广泛，同时微服务还只是在萌芽阶段。在编写本书时，无服务架构（Serverless，开发者再也不用过多考虑服务器的问题，计算资源作为服务而不是服务器的概念出现）还只是新生事物。不同的微服务和第三方服务之间的相互作用影响了容量规划的过程，除此之外，服务模块化、接口化也已经有了更多应用。有鉴于此，我们认为是时候编写本书的第二版，并纳入这些内容了。

最后但是同样重要，随着新网站和新移动 App 每天不断涌现，在网络操作的世界中总会有一些新的东西。因此，像“O'Reilly Velocity and Fluent”这样的会议作为互联网从业人士的社区和交流分享论坛，已经蓬勃发展。此外，博客是一个非

常棒的学习他人经验的资源。有关系统、工具、基准优化和方法学等这类主题的论文已经发表。为此，第二版中的两个新的关键亮点是每章结尾的“延展阅读”和“参考资料”部分。如果你想深入讨论某个特定的主题，这些章节提供了丰富的信息来源。

本书的重点和主题

本书不是关于如何构建复杂的模型和仿真的，也不是关于耗费大量时间来反复运行基准测试的。本书也不是关于利特尔法则^{译注2}、马尔可夫链^{注1}或泊松到达率这样的数学原理的。



这本书介绍如何在实际工作中切实有效地开展容量规划和管理。本书还会介绍实际工具的使用，以及如何调整这些工具的功能以适应网站的发展。当在高速路上车辆爆胎时，你可能会花大量时间努力找出爆胎的原因，你也可能会直接安装备胎，然后重新上路。这就是我们正在应用的容量计划方法——自适应法，而非理论上的。请记住，本书中有大量信息看起来很像常识，这是一件好事。通常，解决问题最简单的方法往往是最好的方法，容量规划也不例外。本书涵盖了不断增长的网站的容量规划过程，包括测量、采购和部署。我们将介绍一些久经考验且受欢迎的测量工具和技术。为此，我们会确保工具和技术的平台无关性。

当然，数据库、Web 服务器、缓存服务器和存储解决方案等方面的细节内容已经超出了本书范围。我们将使用实例来说明每个相关过程和概念。这样做是为了能

注 1: 利特尔法则 (Little's law) 由麻省理工大学斯隆商学院 (MIT Sloan School of Management) 的教授 John Little 于 1961 年提出并加以证明。它是一个有关提前期与在制品关系的简单数学公式: $L = \lambda W$, 这一法则为精益生产的改善方向指明了道路。

译注 2: 马尔可夫链 (Markov chains) 因 A.A. Markov (1856—1922) 得名, 是数学中具有马尔可夫性质的离散时间随机过程。该过程中, 在给定当前知识或信息的情况下 (即当期以前的历史状态) 对于预测将来 (即当期以后的未来状态) 是无关的。

够用尽可能通用的方式来产出资源管理，资源管理过程本身也是我们希望重点强调的。例如，数据库是用来存储数据并提供数据访问的。大多数比较流行的数据库都支持将数据复制到其他服务器，从而提高冗余性和架构扩展能力。本书也涵盖像 Postgres、Oracle 或 MySQL 数据库复制的技术实现（其他书的主题），主要包括在容量规划能力和部署方面复制方式的使用。

实际上，这本书关于如何测量、规划和管理正在发展的网站应用程序，与你所选择的底层技术无关。

本书的读者对象

本书适合系统、存储、数据库和网络等领域的管理员、项目经理和容量规划师。

本书适合那些希望（也许是担心）自己的网站将来成长为像 Facebook、Flickr、Myspace、Twitter 和其他一些经历过访问流量激增而进行大规模扩容的公司。本书讲述的方法都是来自真实网站的运营经验，这些网站都经历了从流量大增到迅速壮大这一过程。如果你也期望网站流量大增，那么请阅读本书。

本书主要内容

第 1 章，容量规划的目标、问题和过程。本章介绍了一些大流量网站的常见问题。

第 2 章，设定容量目标。本章阐述了对一个快速增长的 Web 应用程序进行容量规划时会涉及的各种问题，以及规划的容量如何适用整个网站的可用性和性能。

第 3 章，测量：容量的单位。本章讨论了容量的测量和监控。

第 4 章，趋势预测。本章介绍如何将测量数据转化成预测，以及趋势分析是如何适用整个容量规划过程的。

第 5 章，部署。本章讨论与部署相关的概念：自动化安装、配置和管理。

第 6 章，自动扩缩容。本章讨论利用云服务进行自动扩容的相关概念。

附录 A，虚拟化。本章讨论怎样在容量规划中利用虚拟化和云服务。

11

附录 B，处理瞬时增长。本章为陷入容量危机的站点提供建议，同时也为处理站点故障提供了最佳实践的经验。

附录 C，容量工具。这是一份关于测量、安装、配置和管理工具的说明清单，这些工具贯穿全书并高亮显示。

排版约定

本书使用了下述排版约定。

斜体 (*Italic*)

表示 URL、主机名、目录和文件名、UNIX 命令和选项、程序，偶尔用于强调。

等宽字体 (*Constant Width*)

表示程序列表，以及在段落内部引用程序元素，例如变量或函数名、数据库、数据类型、环境变量、语句和关键字。

等宽粗体字 (*Constant Width Bold*)

表示运行命令时你需要键入的文本。

斜体等宽字体 (*Constant Width Italic*)

表示变量输入，应该替换成用户提供的值。

O'Reilly Safari



Safari Books Online (<http://safaribooksonline.com>) 是一个应需而变的数字图书馆，它同时以图书和视频的形式出

版世界顶级技术和商务作家的专业作品。Safari Books Online 为企业、政府部门、教育机构和个人提供了多种套餐和价格。

订阅者可以在一个完全可搜索的全文数据库中访问上千种图书、培训视频和正式出版之前的书稿。这些内容由以下出版社提供：O'Reilly Media、Prentice Hall Professional、Addison-Wesley Professional、Microsoft Press、Sams、Que、Peachpit Press、Focal Press、Cisco Press、John Wiley&Sons、Syngress、Morgan Kaufmann、IBM Redbooks、Packt、Adobe Press、FT Press、Apress、Manning、New Riders、McGraw-Hill、Jones& Bartlett、Course Technology 等。

关于 Safari Books Online 的更多信息，请访问我们的网站 (<https://www.safaribooksonline.com/>)。

使用示例代码

本书旨在帮助你完成工作。一般来说，你可以在自己的程序或者文档中使用本书附带的示例代码。你无需联系我们获得使用许可，除非你要复制大量的代码。例如，使用本书中的多个代码片段编写程序就无需获得许可。但以 CD-ROM 的形式销售或者分发 O'Reilly 书中的示例代码则需要获得许可。回答问题时援引本书内容及书中示例代码，无需获得许可。在你自己的项目文档中使用本书大量的示例代码时，则需要获得许可。

我们不强制要求署名，但如果你这么做，我们深表感激。署名一般包括书名、作者、出版社和国际标准图书编号。例如：“The Art of Capacity Planning, Second Edition by Arun Kejariwal and John Allspaw. Copyright Arun Kejariwal, John Allspaw, 978-0-596-51857-8”。

如果你觉得自身情况不在合理使用或上述允许的范围内，请通过邮件和我们联系：permissions@oreilly.com。

联系我们

请把对本书的评价和问题发给出版社。

美国：

O'Reilly Media, Inc.
1005 Gravenstein Highway North
Sebastopol, CA 95472

中国：

北京市西城区西直门南大街 2 号成铭大厦 C 座 807 室 (100035)
奥莱利技术咨询（北京）有限公司