

Making Sense of NoSQL

解读NoSQL

[美] Dan McCreary Ann Kelly 著
范东来 滕雨瞳 译



中国工信出版集团



人民邮电出版社
POSTS & TELECOM PRESS

Making Sense of NoSQL

解读NoSQL

[美] Dan McCreary Ann Kelly 著

范东来 滕雨瞳 译

人民邮电出版社
北京

图书在版编目 (C I P) 数据

解读NoSQL / (美) 麦克雷 (McCreary, D.), (美) 凯利 (Kelly, A.) 著; 范东来, 滕雨樟译. — 北京: 人民邮电出版社, 2016. 2

书名原文: Making Sense of NoSQL

ISBN 978-7-115-41110-5

I. ①解… II. ①麦… ②凯… ③范… ④滕… III. ①数据库系统 IV. ①TP311.138

中国版本图书馆CIP数据核字 (2015) 第302980号

版权声明

Original English language edition, entitled *Making Sense of NoSQL* by Dan McCreary and Ann Kelly published by Manning Publications Co., 209 Bruce Park Avenue, Greenwich, CT 06830. Copyright ©2014 by Manning Publications Co.

Simplified Chinese-language edition copyright ©2016 by Posts & Telecom Press. All rights reserved.

本书中文简体字版由 **Manning Publications Co.** 授权人民邮电出版社独家出版。未经出版者书面许可, 不得以任何方式复制或抄袭本书内容。

版权所有, 侵权必究。

-
- ◆ 著 [美] Dan McCreary Ann Kelly
 - 译 范东来 滕雨樟
 - 责任编辑 杨海玲
 - 责任印制 张佳莹 焦志炜
 - ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路 11 号
 - 邮编 100164 电子邮件 315@ptpress.com.cn
 - 网址 <http://www.ptpress.com.cn>
 - 北京艺辉印刷有限公司印刷
 - ◆ 开本: 800×1000 1/16
 - 印张: 16.25
 - 字数: 346 千字 2016 年 2 月第 1 版
 - 印数: 1—3 000 册 2016 年 2 月北京第 1 次印刷
 - 著作权合同登记号 图字: 01-2013-8988 号
-

定价: 49.00 元

读者服务热线: (010) 81055410 印装质量热线: (010) 81055316

反盗版热线: (010) 81055315

内容提要

本书从 NoSQL 的相关理论开始，深入浅出地探讨了 NoSQL 最核心的架构模式、解决方案和一些高级主题，内容循序渐进，从理论回归于实践。

全书分为 4 个部分。第一部分介绍 NoSQL 的相关理论，如 CAP 理论、BASE 理论、一致性散列算法等；第二部分介绍 NoSQL 最核心的架构模式——键值存储、图存储、列族存储、文档存储；第三部分展现一些常用的 NoSQL 解决方案，如 HA、全文搜索等；第四部分讨论 NoSQL 的一些高级主题，如函数式编程。

全书理论与实践并重，每章后面还有通俗的案例。对于 NoSQL 的初学者来说，不失为一本了解 NoSQL 技术全貌的优秀读物。

译者序

在大数据处理领域，NoSQL 无疑是一颗璀璨的明珠。近年来，NoSQL 数据库凭借其易扩展、高性能、高可用、数据模型灵活等特色获得了大批互联网公司的青睐，百度、阿里巴巴、京东等都分别在自己的业务系统中尝试了 NoSQL 解决方案。随着传统的 RMDBS 在某些业务场景下越来越乏力，可以预见的是，越来越多的组织将会乐意尝试使用 NoSQL 解决方案来解决问题。

得益于工业界和学术界的高度关注和大力支持，许多 NoSQL 产品在短时间内大量涌现，方兴未艾，如 HBase、Cassandra、CouchDB、MongoDB、Redis、Neo4j 等。如此多的产品为架构师提供了大量的选择，但却给初学者带来一些学习的困扰。

我认为，学习一门技术，起点很重要。如果初学者将自己的起点定位在某一种 NoSQL 产品上，那么学习完这种产品，得到的知识只是 NoSQL 技术中很小的一部分。但如果将起点定位在 NoSQL 的架构、模式上，那么初学者很快就会对 NoSQL 的全貌有个清晰的认识，这样在学习具体产品时，就会事半功倍。

本书对于 NoSQL 的学习者来说是一本很好的读物，它能让初学者快速地领略到 NoSQL 技术的全貌，为后面的学习打下坚实的基础。本书的两位作者 Dan McCreary 和 Ann Kelly 是两位经验丰富的架构师，具有高屋建瓴的视角，更为难得的是，他们这本书还具有深入浅出的特点。我相信，本书作为技术人员的第一本 NoSQL 书籍来说是很合适的。

在本书翻译过程中，得到了 BBD 小伙伴们的大力支持，在此对曾途、尹康、何宏靖、刘世林、赵飞、颜如宾、王维、杨舟、陈丽竹 Ju、利晓蕾、李倩、黄艳、宋开发、丁梦希、夏阳、唐婷婷、吴松珏等表示感谢。

就像本书第 2 章提到的 CAP 理论，无论是 RMDBS 还是 NoSQL 都不可能同时满足 CAP 理论的三个方面，要学会接受缺憾。技术如此，人生亦如此。

范东来

2015 年 10 月于成都

译者简介

范东来

北京航空航天大学硕士，BBD（数联铭品）大数据技术部负责人，大数据平台架构师，极客学院布道师，著有《Hadoop 海量数据处理：技术详解与项目实战》。研究方向：图挖掘、模式分类。

滕雨瞳

清华大学苏州汽车研究院大数据处理中心高级研发工程师。常年从事基于 HBase、Redis、Cassandra 等 NoSQL 数据库的应用开发，具有丰富的 NoSQL 系统架构和性能调优经验。研究方向：大数据处理、数学机械化。

序

如果一个主题被定义为它不是什么而不是它是什么，我们该如何向别人解释它？相信我，按照一位过去 3 年在这个领域传道授业的人的说法，这是非常困难的，这也是许多专家、学者、技术厂商的共识。尽管很少人认为 NoSQL 这个名字是最好的，但似乎每个人都同意，它比其他术语更加贴切地定义了这一类产品和技术。所以我强烈建议不要由于 NoSQL 的定义方式，而放弃学习这些新的技术。要我来说，这些是值得你花费时间的。

接下来，是一些个人背景的简要介绍：作为一个信息管理方面的出版商，我听说过 NoSQL 这个词，但当时并不知道它的含义，直到 3 年前我在多伦多参加一次会议时，在走廊上碰到了 Dan McCreary。他告诉了我一些情况，使我了解了他目前的项目、使用的令人兴奋的技术和优秀的工作伙伴。他使我相信 NoSQL 技术将呈星火燎原之势蓬勃发展，像我这样的人应该尽可能地学习这方面的知识。这无疑是一个极佳的建议，于是从那时起，我们就建立了良好的伙伴关系，如共同参加学术会议、发起在线研讨会、共同撰写论文。同所有 NoSQL 社区的参与者一样，Dan 是一位才华横溢的技术高手。

就像从事一些复杂而深奥的工作的人一样，为了使那些不具备和我同等学习热情和背景的人理解，我经常试图用简单的词汇来解释复杂的事物。但就算你理解了一次完美的电梯演讲的价值，或者拼命地想要向你的母亲解释你将要做的事情，正确的解释仍然是模糊的。有时向更有学识的人解释新事物反而更加困难。这尤其体现在 NoSQL 方面，由于 RDBMS 拥有庞大的社区和无数的拥簇者，他们只需要一个查询工具就高效而愉快地工作了 30 年。

这就是《解读 NoSQL》出现的原因。如果你参与企业信息系统的建设并且正试着理解 NoSQL 的价值，那么你会庆幸这本书的出现，因为它对你来说非常直观。诚然，新技术的使用能够使技术先驱者尝到甜头，但是对于企业 IT 从业人员来说，仍然存在一些令人望而生畏的障碍。其中的一个原因是由于这些年来成熟技术的沉淀和积累，周围的人会质疑你到底是为什么想要将数据转换成一个“四不像”而不是用优美的、有序的表来进行存储。

本书的作者也意识到了这一点，并且针对你将会面对到的技术架构之间的取舍问题做了大量的研究分析。除此之外，我还欣赏他们付出了大量努力所提供的贯穿全书的案例研究。案例通常

是说服的关键，当你向你的组织介绍这些新技术时，这些源自真实线上应用的案例对于你想要表达的主题来说，其价值是难以估量的。

虽然 Dan McCreary 和 Ann Kelly 给出了第一个 NoSQL 全面的定义和为什么在生产环境中使用 NoSQL 的原因，但本书并不仅仅是一本单纯的技术书籍，它背后包含了作者与架构师、开发人员一起付出的大量努力以确保不同 NoSQL 产品的细节和特点都在本书得到了完整呈现。

本书是一本易于理解的 NoSQL 指南，对于技术经理、架构师和开发人员来说有很高的参考价值。它适合需要全方位了解数据管理解决方案以应对世界上日益增长的、复杂的、海量的、快速流动的数据所带来的严峻挑战的任何人。第 1 章的标题是“NoSQL：明智的选择”，当你选择这本书的时候，你已经做出了一个明智的选择。

Tony Shaw

Dataversity 创始人和首席执行官

前言

有时候，现实迫使我们重新审视我们认为已经了解的事物。在花费了大量的工作时间专注于以行式数据结构存储数据的数据建模任务之后，我们发现，其实建模环节并不是非做不可的。但是这些信息并不意味着我们现有的知识体系是无效的，它迫使我们去审视应该如何解决企业的技术难题。有了新的知识、技术和解决问题方式的武装之后，我们的思路才能得以扩展。

2006 年，在一个涉及房地产交易的项目中，我们花了好几个月的时间设计 XML 的语言模式和形式以存储层次结构复杂的数据。根据我的一个朋友 Kurt Cagle 的建议，我们发现，用原生 XML 数据库对数据进行存储为我们的项目节省了数月的对象建模、设计关系型数据库以及对象关系映射时间，并最终形成一个可以由非专业人员进行维护的异常简单的架构。

对进入 NoSQL 领域的人来说，能意识到企业数据可以用 RDBMS 以外的架构进行存储是重要的转折点。最初，我们可能对这些消息持怀疑态度，会带着恐惧和自我怀疑的复杂心情来看待这些信息。我们会质疑自己的技能和为我们提供培训的教育机构以及那些强调 RDBMS 和对象是解决问题唯一途径的组织。但是，我们如果要公平地对待客户和用户，就必须进行一种全方位的尝试来寻找解决企业问题的最佳方案并评估其他数据库产品架构。

2010 年，一直缺乏关注的 NoSQL 数据库技术得以在大型企业数据会议中露面，当时我们和来自 Dataversity 的 Tony Shaw 一起讨论了关于启动一个新的会议的议题。该会议旨在为所有对 NoSQL 技术感兴趣的人提供一个平台，供他们学习和展示如何使用 NoSQL 数据库技术。2011 年 8 月第一届 NoSQL Now! 会议在加州圣何塞成功举行，吸引了大约 500 名感兴趣和好奇的参会者。

在会议上，我们发现没有任何一份资料覆盖了 NoSQL 架构，或者客观地介绍了一个针对业务来选择 NoSQL 数据库产品的流程。人们想要的不仅仅是开源项目中的“Hello World”示例，而是一本能够帮助他们根据具体业务选择技术架构并像评估商业数据库系统那样评估开源技术的指南。

找到一家能够认同我们现有技术文档的出版机构是第一步，幸运的是，我们发现 Manning 出版社理解这种出版价值。

致谢

我们要感谢 Manning 出版社的每一个人，是你们帮助我们最初的思想转化为一本书。Michael Stephens 带领我们的工作步入正轨；Elizabeth Lexleigh 是我们的开发编辑，她非常耐心地阅读每一章的每一版；Nick Chase 完成了所有的技术工作；还有市场营销和生产团队以及每一个幕后工作者，我们衷心感谢你们的努力、指导和鼓励。

感谢那些回顾了案例研究和为我们提供真实 NoSQL 案例的人们，感谢你们宝贵的时间和专业知识：George Bina、Ben Brumfield、Dipti Borkar、Kurt Cagle、Richard Carlsson、Amy Friedman、Randolph Kahle、Shannon Kempe、Amir Halfon、John Hudzina、Martin Logan、Michaline Todd、Eric Merritt、Pete Palmer、Amar Shan、Christine Schwartz、Tony Shaw、Joe Wicentowski、Melinda Wilken 和 Frank Weige。

感谢那些在书稿修改过程中提出了宝贵意见和反馈的评阅者们，有了你们的参与，我们的书才会更好：Aldrich Wright、Brandon Wilhite、Craig Smith、Gabriela Jack、Ian Stirk、Ignacio Lopez Vellon、Jason Kolter、Jeff Lehn、John Guthrie、Kamesh Sampah、Michael Piscatello、Mikkel Eide Eriksen、Philipp K. Janert、Ray Lugo、Jr.、Rodrigo Abreu 和 Roland Civet。

在此，特别感谢我们的朋友 Alex Bleasdale 为我们提供 NoSQL 安全机制对应章中有关基于用户的访问控制的案例研究的代码。特别感谢 Tony Shaw 为本书撰写的序和 Leo Polovets 在出版前对最终书稿的技术审校。

关于本书

编写本书的过程中，我们抱着两个目标：第一个目标是介绍 NoSQL 数据库；第二个目标则是为读者展示将 NoSQL 系统作为独立解决方案的方法，以及扩展当前 SQL 系统解决业务问题的途径。我们欢迎任何对 NoSQL 感兴趣的读者将本书作为参考。你会发现本书包含的信息、例子以及案例的目标受众是那些有兴趣学习 NoSQL 的技术经理、解决方案架构师和数据架构师。

本书将会帮助你客观地评估 SQL 和 NoSQL 数据库系统以明白它们各自解决了哪类业务问题。如果你想要的是一本关于某个产品的编程指导，那么本书可能就不是你想要的。在本书中，你将了解到 NoSQL 兴起背后的动机以及相关的术语和概念。对于书中有些章节介绍的内容，你可能已经了解，可以根据实际情况略看或直接跳过该部分，而把精力集中在你仍然不了解的部分。

最后，我们非常重视和关注行业标准。SQL 系统相关的行业标准允许应用在使用统一语言的情况下自由切换底层数据库。但遗憾的是，NoSQL 系统还不能作出这个保证。在不久的将来，NoSQL 应用提供商会敦促 NoSQL 数据库供应商采纳一系列标准以使其能像 SQL 一样具备兼容性。

路线图

本书分为 4 部分。第一部分由 NoSQL 定义和 NoSQL 运动背后的基本概念回顾组成。

在第 1 章“**NoSQL：明智的选择**”中，我们定义了术语 NoSQL，讨论了触发 NoSQL 运动的关键性事件，并阐述了 NoSQL 系统带来的大局上的商业收益。熟悉 NoSQL 运动及其商业收益的读者可以选择略过本章。

在第 2 章“**NoSQL 概念**”中，我们引入了 NoSQL 运动中的核心概念。读者在初次通读的过程中，可以选择略过本章，但这一章对于帮助理解之后的章节非常重要。我们建议将本章作为一个参考指导，在遇到相关概念时再返回本章仔细阅读相关内容。

在第二部分“**数据库模式**”中，我们深入回顾了 SQL 和 NoSQL 数据库的架构模式。我们探讨了不同的数据库结构、如何访问它们，并通过用例展示了每种架构最适用的场景。

第 3 章介绍了“**基础数据架构模式**”。本章以回顾关系型数据库背后的驱动力以及 ERP 系统

需求如何塑造出现今关系型数据库和商业智能/数据仓库（BI/DW）系统开始，还简略地讨论了其他数据库系统，如对象数据库、修改控制系统。如果已经熟悉了这些系统，可以略过本章。

在第4章“**NoSQL 数据架构模式**”中，我们介绍了 NoSQL 相关的数据库模式。我们探讨了键值存储、图存储、列族存储（BigTable）以及文档存储。本章将用定义、示例和案例研究等内容辅助读者理解。

第5章覆盖了“**原生 XML 数据库**”的内容。因其低廉的成本和对标准的支持，该数据库常被用于政府和出版应用。我们为读者展示了两个来自金融和政府出版领域的案例。

第三部分探讨了将 NoSQL 系统应用到大数据、搜索、高可用性以及敏捷网站部署等问题上的方法。

在第6章“**用 NoSQL 管理大数据**”中，你将了解到配置 NoSQL 系统在商用硬件上高效处理大量数据的方法。我们还有一个关于分布式计算和横向可扩展性的讨论。同时，我们还为读者展示了一个不能通过商用硬件扩展完成巨型图分析的案例。

在第7章“**用 NoSQL 搜索获取信息**”中，你将学习到如何通过实现文档模型和保留文档内容的方式提升搜索质量。我们讨论了采用 MapReduce 转化技术创建倒排索引并最终获得快速搜索性能的方法，还回顾了用于文档和数据库的搜索系统，并展示了采用结构化搜索解决方案获得准确的搜索结果排序的方法。

第8章的内容覆盖了“**用 NoSQL 构建高可用的解决方案**”。我们阐述了如何使用 NoSQL 系统自带的复制和分布式特性提升系统可用性。你将了解到在采用了数据同步技术的情况下，如何使用大量低廉 CPU 来提供更长的在线时间。我们的案例研究揭示了纯点对点架构能比其他分布式模型提供更高可用性的原因。

在第9章中，我们讨论了“**用 NoSQL 提升敏捷性**”。因为移除了对象关系映射层，NoSQL 软件开发变得更简单且能快速适应业务需求的变更。你将了解到这些 NoSQL 系统是如何将富有经验的开发人员和非编程人员整合到软件开发生命周期过程中的。

在第四部分中，我们先介绍了函数式编程、安全等“高级主题”，然后，回顾了一个正式的 NoSQL 系统选型过程。

在第10章中，我们介绍了关于“**NoSQL 与函数式编程**”和类似 MapReduce 的分布式转化架构的要求这两方面内容。我们探讨了函数式编程对 NoSQL 解决方案能使用大量低廉 CPU 能力的影响，以及许多 NoSQL 数据库系统采用类似 Erlang 的基于角色的框架的原因。我们还通过 NetKernel 系统的案例介绍了结合函数式编程和面向资源编程构建性能可扩展的分布式系统的方法。

第11章覆盖了“**安全：保护 NoSQL 系统中的数据**”相关的内容。我们回顾了 NoSQL 解决方案中安全性方面的历史以及关键的安全性问题。我们还通过例子为读者展示了键值存储、列族存储和文档存储实现一个健壮的安全模型的能力。

在第12章“**选择合适的 NoSQL 解决方案**”中，我们遍历了企业为业务问题选型合适数据库的正式流程。最后，我们将以一些关于这些技术会如何影响业务系统选型的想法和信息结束本章。

代码约定和下载

清单和文本中的源代码会以等宽字体出现，以便和普通文本区分开。你可以从 Manning 出版社的网站（www.manning.com/MakingSenseofNoSQL）上下载清单中的源代码。

作者在线论坛

购买本书后，读者享有由 Manning 出版社运营的私有网页论坛的免费访问权。在该论坛上，读者可以评论本书、询问技术问题以及从作者和其他用户那获得帮助。如果想访问和订阅这个论坛，读者可以访问 www.manning.com/MakingSenseofNoSQL。这个页面提供了如何在注册后登录论坛、可以获得哪些帮助以及论坛行为规范等信息。

Manning 出版社对读者承诺，要为读者与读者、读者与作者之间有意义的交流提供一个空间。这不是一个针对任何特定数量的作者的承诺。任何在论坛中做出过贡献的作者都是志愿且不计酬劳的。因此，我们建议读者试着询问作者一些更具挑战性的问题，以免他们对问题毫无兴趣。

只要书籍仍在印刷，这个作者在线论坛和历史讨论存档就一直可以通过出版社的网站访问。

关于作者

Dan McCrearay 是一个对行业标准具有强烈兴趣的数据架构咨询师。他曾经在贝尔实验室（负责集成电路设计）、超级计算行业（负责移植 UNIX）以及史蒂夫·乔布斯（Steve Jobs）创办的 NeXT 计算机公司（软件布道师）工作，也创办了自己的咨询公司。Dan 于 2002 年开始参与制定美国联邦数据标准，该标准也在国家信息交换模型（NIEM）中获得了启用。在 2006 年面对用原生 XML 数据库存储数据时，Dan 开始了他的 NoSQL 开发历程。他还是万维网 XForms 标准制定小组的客座专家以及 NoSQL Now! 会议的联合创始人。

Ann Kelly 是 Kelly McCrearay 咨询公司的软件咨询师。在从事多年保险行业软件开发和管理项目之后，她于 2011 年开始转投 NoSQL 领域。从那时起，她就开始为客户构建能快速高效地解决业务问题的 NoSQL 解决方案，同时还提供管理应用的培训。

目录

第一部分 了解 NoSQL

第 1 章 NoSQL:明智的选择 2

- 1.1 什么是 NoSQL 3
- 1.2 NoSQL 的商业驱动 4
 - 1.2.1 容量 5
 - 1.2.2 速度 5
 - 1.2.3 敏捷性 6
- 1.3 NoSQL 案例研究 6
 - 1.3.1 案例研究: LiveJournal 的 Memcache 技术 7
 - 1.3.2 案例研究: Google 的 MapReduce——利用商用硬件生成搜索索引 7
 - 1.3.3 案例研究: Google 的 Bigtable——一个有着数十亿行和百万列的表 8
 - 1.3.4 案例研究: 亚马逊的 Dynamo——每天 24 小时接收订单 9
 - 1.3.5 案例研究: MarkLogic 9
 - 1.3.6 实践 10
- 1.4 小结 10

第 2 章 NoSQL 概念 12

- 2.1 保持组件简单以促进重用 12
- 2.2 将应用分层以简化设计 14
- 2.3 策略地使用 RAM、SSD

- 和磁盘提升性能 17
- 2.4 使用一致性散列算法维护当前的缓存 18
- 2.5 比较 ACID 和 BASE——两种可靠的数据库事务方法 19
 - 2.5.1 RDBMS 的事务控制——ACID 21
 - 2.5.2 非 RDBMS 的事务控制——BASE 22
- 2.6 通过数据库分片获得水平扩展能力 23
- 2.7 基于 Brewer 的 CAP 定理进行权衡 25
- 2.8 实践 26
- 2.9 小结 27
- 2.10 延伸阅读 27

第二部分 数据库模式

第 3 章 基础数据架构模式 30

- 3.1 什么是数据架构模式 31
- 3.2 理解应用于 RDBMS 的行存储设计模式 31
 - 3.2.1 行存储如何工作 32
 - 3.2.2 行存储的演变 33
 - 3.2.3 分析行存储模式的优点和缺点 34
- 3.3 示例: 对销售订单进行

- 连接操作 35
- 3.4 回顾 RDBMS 实现的特性 36
 - 3.4.1 RDBMS 事务 37
 - 3.4.2 固定的数据定义语言和强类型的列 38
 - 3.4.3 通过 RDBMS 视图保证安全并进行访问控制 39
 - 3.4.4 RDBMS 的复制和同步 39
- 3.5 通过 OLAP、数据仓库和商业智能系统对历史数据进行分析 41
 - 3.5.1 数据如何从操作型系统流入分析型系统 42
 - 3.5.2 熟悉 OLAP 的概念 44
 - 3.5.3 通过汇总生成专项报表 45
- 3.6 将高可用性和以读为主的系统一体化 46
- 3.7 在修改控制系统和数据同步中使用散列树 47
- 3.8 实践 49
- 3.9 小结 49
- 3.10 延伸阅读 50

第 4 章 NoSQL 数据架构模式 51

- 4.1 键值存储 52
 - 4.1.1 什么是键值存储 52
 - 4.1.2 使用键值存储的好处 54
 - 4.1.3 使用键值存储 56
 - 4.1.4 使用案例：用键值存储保存网页 59
 - 4.1.5 使用案例：亚马逊简单存储服务 (S3) 59
- 4.2 图存储 60
 - 4.2.1 图存储概述 60
 - 4.2.2 用 RDF 标准来连接外部数据 62
 - 4.2.3 图存储的使用案例 63
- 4.3 列族 (Bigtable) 存储 68
 - 4.3.1 列族存储基础 69
 - 4.3.2 理解列族存储的键 69
 - 4.3.3 列族存储的优点 70

- 4.3.4 案例研究：在 Bigtable 中存储分析信息 72
- 4.3.5 案例研究：Google 地图用 Bigtable 存储地理信息 72
- 4.3.6 案例研究：使用列族存储用户偏好信息 73
- 4.4 文档存储 73
 - 4.4.1 文档存储基础 74
 - 4.4.2 文档集合 74
 - 4.4.3 应用集合 75
 - 4.4.4 文档存储的 API 75
 - 4.4.5 文档存储的实现 76
 - 4.4.6 案例研究：MongoDB 和广告服务器 76
 - 4.4.7 案例研究：大型对象数据库 CouchDB 77
- 4.5 NoSQL 架构模式的变体 78
 - 4.5.1 定制 RAM 和 SSD 存储 78
 - 4.5.2 分布式存储 78
 - 4.5.3 分组的对象 79
- 4.6 小结 80
- 4.7 延伸阅读 81

第 5 章 原生 XML 数据库 82

- 5.1 什么是原生 XML 数据库 83
- 5.2 用原生 XML 数据库构建应用 85
 - 5.2.1 加载数据可以像拖曳那样简单 86
 - 5.2.2 使用集合来组织 XML 文档 87
 - 5.2.3 使用 XPath 运用简单的查询转换复杂的数据 89
 - 5.2.4 用 XQuery 转换数据 91
 - 5.2.5 用 XQuery 更新文档 93
 - 5.2.6 XQuery 全文搜索标准 94
- 5.3 在原生 XML 数据库中应用 XML 标准 94
- 5.4 用 XML Schema 和 Schematron 设计和验证数据 96
 - 5.4.1 XML Schema 96
 - 5.4.2 使用 Schematron 检查文档规则 97

- 5.5 用自定义模块扩展 XQuery 98
- 5.6 案例研究：在美国国务院历史学家办公室使用 NoSQL 98
- 5.7 案例研究：使用 MarkLogic 管理金融衍生品 102
 - 5.7.1 为什么 RDBMS 难以存储金融衍生品 102
 - 5.7.2 一个投资银行从 20 个 RDBMS 转换到 1 个原生 XML 数据库 102
 - 5.7.3 迁移至原生 XML 文档存储的商业好处 104
 - 5.7.4 项目成果 104
- 5.8 小结 105
- 5.9 延伸阅读 105

第三部分 NoSQL 解决方案

第 6 章 用 NoSQL 管理大数据 108

- 6.1 什么才是大数据解决方案 109
- 6.2 线性扩展数据中心 112
- 6.3 理解线性可扩展性和表现力 113
- 6.4 了解大数据问题的类型 114
- 6.5 使用无共享架构分析大数据 116
- 6.6 选择分布式模型：主从模型与对等模型 117
- 6.7 在分布式系统上使用 MapReduce 处理数据 118
 - 6.7.1 MapReduce 和分布式文件系统 120
 - 6.7.2 MapReduce 怎样做到高效处理大数据问题 121
- 6.8 NoSQL 系统处理大数据问题的 4 种方式 122

- 6.8.1 分发查询到数据，而非数据到查询 122
- 6.8.2 使用散列环在集群中均匀分发数据 122
- 6.8.3 使用复制扩展读取性能 123
- 6.8.4 使数据库将查询均衡地分发到数据节点 124
- 6.9 案例研究：使用 Apache Flume 处理事件日志 125
 - 6.9.1 事件日志数据分析的挑战 125
 - 6.9.2 Apache Flume 搜集分布式事件日志的方法 126
 - 6.9.3 延伸思考 127
- 6.10 案例研究：计算机辅助发现医疗保险欺诈 128
 - 6.10.1 什么是医疗保险欺诈检测 128
 - 6.10.2 使用图和定制的内存共享硬件检测医疗保险欺诈 129
- 6.11 小结 130
- 6.12 延伸阅读 131

第 7 章 用 NoSQL 搜索获取信息 132

- 7.1 什么是 NoSQL 搜索 132
- 7.2 搜索分类 133
 - 7.2.1 布尔搜索、全文关键字搜索和结构化搜索的比较 133
 - 7.2.2 测试常见搜索类型 134
- 7.3 提高 NoSQL 搜索效率的策略和方法 135
- 7.4 使用文档结构提升搜索质量 137
- 7.5 搜索质量量化 139
- 7.6 本地索引与远程搜索服务 139
- 7.7 案例研究：使用 MapReduce 建立倒排索引 141
- 7.8 案例研究：搜索技术文档 142

- 7.8.1 什么是技术文档
搜索 142
- 7.8.2 在 NoSQL 文档存储中
保留文档结构 143
- 7.9 案例研究: 搜索领域
语言——可检索性和
重用性 144
- 7.10 实践 145
- 7.11 小结 146
- 7.12 延伸阅读 146

第 8 章 用 NoSQL 构建高可用 的解决方案 148

- 8.1 高可用 NoSQL 数据库的
定义 148
- 8.2 量化 NoSQL 数据库的
可用性 149
 - 8.2.1 案例研究: 亚马逊 S3 的
服务级别协议 151
 - 8.2.2 预测系统可用性 151
 - 8.2.3 实践 152
- 8.3 NoSQL 系统的高可用性
策略 153
 - 8.3.1 使用负载均衡器将流量转
向到最空闲的节点 153
 - 8.3.2 结合高可用分布式文件系
统和 NoSQL 数据库 154
 - 8.3.3 案例研究: 将 HDFS 作为
一个高可用的文件系统存
储主数据 155
 - 8.3.4 使用托管的 NoSQL
服务 156
 - 8.3.5 案例研究: 使用亚马逊的
DynamoDB 作为高可用数
据存储 157
- 8.4 案例研究: 使用 Apache
Cassandra 作为高可用的
列族存储 158
- 8.5 案例研究: 使用 Couchbase
作为高可用文档
数据库 161
- 8.6 小结 163
- 8.7 延伸阅读 163

第 9 章 用 NoSQL 提升敏 捷性 165

- 9.1 软件敏捷性的定义 165
- 9.2 量化敏捷性 169
- 9.3 使用文档存储来避免对象
关系映射 171
- 9.4 案例研究: 使用 XRX 管理
复杂表单 172
 - 9.4.1 什么是复杂业务
表单 173
 - 9.4.2 用 XRX 替换客户端
JavaScript 和对象关系
映射 173
 - 9.4.3 理解 XRX 对敏捷性的
影响 176
- 9.5 小结 177
- 9.6 延伸阅读 177

第四部分 高级主题

第 10 章 NoSQL 与函数式 编程 180

- 10.1 什么是函数式编程 181
 - 10.1.1 指令式编程就是管理
程序状态 181
 - 10.1.2 函数式编程是没有副作
用的并行转化 183
 - 10.1.3 比较指令式编程和函数
式编程的扩展性 186
 - 10.1.4 使用引用透明避免重复
计算 187
- 10.2 案例研究: 用 NetKernel
优化网页内容组装 188
 - 10.2.1 组装嵌套内容, 追踪组件
依赖 188
 - 10.2.2 用 NetKernel 优化组件
再生成 189
- 10.3 函数式编程语言
示例 191
- 10.4 完成指令式编程到函数
式的编程转变 192
 - 10.4.1 使用函数作为函数的