HZ BOOKS
华章科技

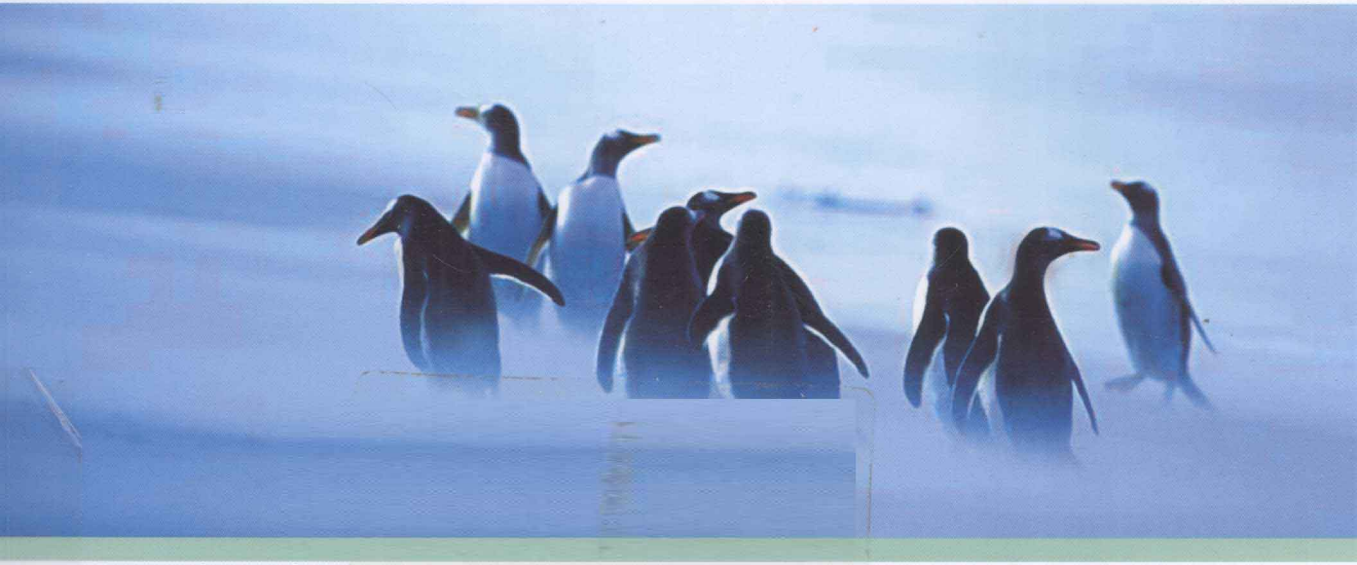
内容全面，系统地讲解了Cassandra的所有功能特性及其使用方法
深入浅出，结合源代码分析了Cassandra的底层机制和工作原理

国内Cassandra领域的先驱者和实践者亲自执笔，多位数据库专家联袂推荐

数据库技术丛书

Cassandra in Action

Cassandra实战



郭 鹏◎著

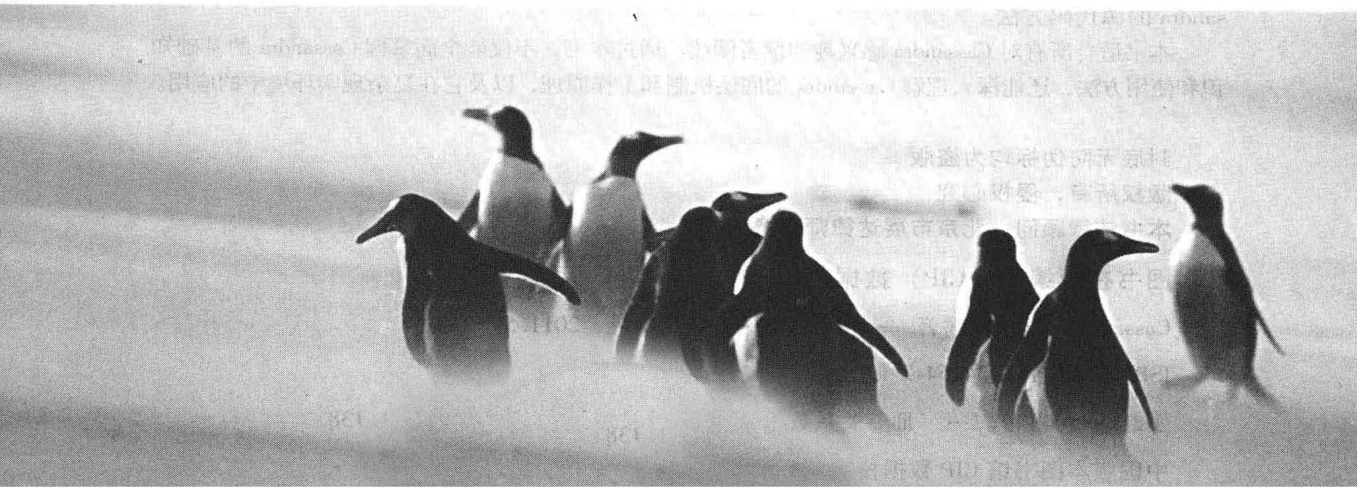
机械工业出版社
China Machine Press

数据库 技术丛书

Cassandra in Action

Cassandra实战

郭 鹏◎著



机械工业出版社
China Machine Press

国内首本 Cassandra 专著，由 Cassandra 领域的先驱者和实践者亲自执笔，多位数据库专家联袂推荐，权威性毋庸置疑。

本书内容全面，基于 Cassandra 最新版撰写，系统地讲解了 Cassandra 的所有功能特性和使用方法；实战性强，不仅包含大量示例代码，而且还设计了一个完整的在线交易系统实例；有一定的深度，不仅结合源代码分析了 Cassandra 的底层机制和工作原理，而且还精心总结了一些关于 Cassandra 的最佳实践。

全书一共分为 13 章，首先简单介绍了 NoSQL 的优势，以及几种具有代表性的 NoSQL 数据库的功能特性；其次详细讲解了 Cassandra 的安装和配置、数据模型和排序规则、编程接口等基础知识；接着以迭代的方式演示了一个基于 Cassandra 的在线交易系统的完整开发过程，很好地将基础理论融入到了实践中；再接着结合源代码分析了 Cassandra 的集群机制、内部数据存储结构、数据更新机制、数据读取机制、数据压缩机制、启动流程等与 Cassandra 的底层机制和工作原理相关的内容；最后讲解了 Cassandra 在分布式环境中的应用、与 Hadoop 的整合，以及相关的最佳实践。附录中包含了本书示例的源代码以及在 Eclipse 环境中编辑和修改 Cassandra 的源代码方法。

本书适合所有对 Cassandra 感兴趣的读者阅读。通过本书，不仅能全面掌握 Cassandra 的基础知识和使用方法，还能深入理解 Cassandra 的底层机制和工作原理，以及它在复杂现实环境中的应用。

图书在版编目（CIP）数据

Cassandra 实战/郭鹏著. —北京：机械工业出版社，2011.5

ISBN 978-7-111-34164-2

Ⅰ. C…　Ⅱ. 郭…　Ⅲ. 关系数据库－数据库管理系统　Ⅳ. TP311.138

中国版本图书馆 CIP 数据核字（2011）第 067289 号

机械工业出版社（北京市西城区百万庄大街 22 号　邮政编码　100037）
责任编辑：陈佳媛
北京市荣盛彩色印刷有限公司印刷
2011 年 6 月第 1 版第 1 次印刷
186mm×240mm · 20 印张
标准书号：ISBN 978-7-111-34164-2
定价：59.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换
客服热线：（010）88378991；88361066
购书热线：（010）68326294；88379649；68995259
投稿热线：（010）88379604
读者信箱：hzjsj@hzbook.com

前　言

为什么写本书

最早开始接触 Cassandra 是在 2010 年年初，那时 Cassandra 才刚刚进入人们的视线。由于传统的数据库已经无法满足项目业务的需求，公司便派我去寻找和尝试其他的数据库解决方案。当时有两种备选方案：Hadoop 项目中的 HBase 和 Facebook 开源的 Cassandra。

在对比了各方面的资料和分析了项目的业务需求之后，我们最终选择了配置简单、部署方便，使用高效的 Cassandra。与 HBase 不同，Cassandra 是一套独立的系统，不需要搭建和了解 HDFS 及 Zookeeper 集群即可开始使用，并且可以在 Windows 系统中直接进行测试。所以，我们在很短的时间之内就完成了功能和性能的评估，为后期的项目实施做好了技术储备。

到 2010 年 6 月份，部门开始正式搭建实时数据中心，需要将公司前台数据库产生的业务数据定期抽取到实时数据中心，提供给业务部门的相关人员使用。于是我们开始设计 Cassandra 的业务数据存储结构，并开始分析 Cassandra 内部的实现机制和源代码，从 Cassandra 的工作原理上指导我们的业务数据存储结构设计和测试方案。在整个设计和测试的过程中，我们遇到的一些问题，都可以通过活跃的 Cassandra 开发社区和开放的源代码找到答案。而且，我们还可以通过修改少量的源代码来修复一些 BUG 和增强系统管理的特性。最终，这些修改也都在开源社区提供的后续版本中得到了体现。得益于活跃的开源社区，

Cassandra 的功能和版本也在不断地更新。尤其是在最新的 Cassandra 0.7.x 版本中，提供了大家都非常期待的二级索引与在线更新 Schema 的功能。

Datastax 公司作为 Cassandra 的主要开发成员，开始为 Cassandra 的使用者提供相应的服务支持，并开发出了一系列强大的工具，进一步提高了 Cassandra 的实用性。另外，在 2011 年的 5 月，Datastax 将提供一款振奋人心的工具——Brisk。通过 Brisk，Cassandra 将取代 Hadoop 项目中的 HDFS 和 HBase 的角色，直接与 Hadoop 项目中的 MapReduce 和 Hive 项目进行集成，从而提高整个海量数据分析系统的性能和易用性。

相信未来 Cassandra 能够我们带来更多的惊喜。

本书面向的读者

初中级程序员。他们可以从本书中的实际例子了解 Cassandra 的基本概念，以及如何通过各种编程接口在 Cassandra 中写入和读取数据，如何建立二级索引，如何动态修改 Schema 信息，如何通过 MapReduce 做海量的数据分析，等等。

中高级程序员。他们可以通过本书对源代码的分析和讲解来了解 Cassandra 中的内部实现原理，从而能选用更加高效的实现方式，调整最适合的 Cassandra 集群运行配置，增强程序的整体运行性能和稳定性。

DBA 和系统运维人员。他们可以通过本书中的运维管理等内容，了解如何在 Cassandra 集群中安全高效地增加或减少服务器，以及如何恢复集群的错误，等等。

系统架构师。他们可以根据本书所讲的内容掌握 Cassandra 能做什么，不能做什么，适合做什么，从而为不同项目做出最合适的架构选型。

联系作者

正如大家所知，开源技术的更新速度非常之快，Cassandra 亦是如此。尽管我们已经很努力地将最新的内容融入其中，但仍难免会遗留部分内容。所以，如果你发现本书存在任何错误和问题，或者是你对书中的某些内容不理解需要与我们进一步探讨，再或者是你有好的意见或建议，都欢迎随时与我取得联系，我的邮箱是：gpcuster@ gmail. com。

另外，我的个人博客也会经常更新关于 Cassandra 的最新动态和使用心得等内容，也欢迎大家来这里与我交流。博客的地址是：http://gpcuster. cnblogs. com。

致谢

感谢老张、许玉勤和姜迅，是你们的信任与支持给了我接触和学习 Cassandra 的机会和应用 Cassandra 的平台。

感谢童家旺、何勇、任振中、王海和陈晓峰，在学习和应用 Cassandra 的过程中，你们

给予了我很多帮助与支持。

感谢我的家人，一年来，我的大部分原来属于你们的业余时间都给了这本书，感谢你们的理解。

感谢杨福川和曾珊两位编辑为本书付出的耐心与努力，你们为本书提出了很多宝贵的建议，你们的敬业精神令我钦佩。

感谢本书的所有读者，希望你们能从本书中有所收获，大家的认可是我不断前进的动力。

郭　鹏

2011 年 4 月

目　　录

第 1 章

认识 NoSQL

本章内容

- NoSQL 的起源和发展现状
- 为什么要使用 NoSQL
- 开源 NoSQL 产品介绍
- 本章小结

1.1 NoSQL 的起源和发展现状

对于 NoSQL 这个新兴的名词，大家的理解不尽相同。在网站 http://nosql-database.org/ 上对 NoSQL 有一个较为全面的解释："下一代的数据库产品应该具备这几个特点：非关系型的、分布式的、开源的、可以线性扩展的。"这类数据库最初的目的在于提供现代网站可扩展的数据库解决方案，它开始于 2009 年年初，目前正在快速发展。这种类型的数据库具有以下特点：自由的 Schema，数据多处备份，简单的编程 API，数据的最终一致性等。所以，将这种类型的数据库称为 NoSQL（不仅仅是 SQL，全称为"not only sql"）数据库。

NoSQL 数据库并不是要取代现在广泛应用的传统数据库，而是采用一种非关系型的方式解决数据的存储和计算的问题。

由于 NoSQL 中没有像传统数据库那样定义数据的组织方式为关系型的，所以只要内部的数据组织采用了非关系型的方式，就可以称之为 NoSQL 数据库。

目前，可以将众多的 NoSQL 数据库按照内部的数据组织形式进行如下分类：

- Key/Value 的 NoSQL 数据库
- 面向文档的 NoSQL 数据库
- 面向列的 NoSQL 数据库
- 面向图的 NoSQL 数据库

不同的数据组织适合于不同的应用场景，后面将进行介绍。

1.2 为什么要使用 NoSQL

SQL 语言和关系型数据库（MySQL、PostgreSQL、Oracle 等）是通用的数据解决方案，占有绝大多数的市场。不过在最近兴起的 NoSQL 运动中，涌现出一批具备高可用性、支持线性扩展、支持 Map/Reduce 操作等特性的数据产品，它们具有如下特性：

- 频繁的写入操作、相对较少的读取统计信息的操作（如网站访问计数器），应该使用基于内存的 Key/Value（键/值）存储系统（如 Redis）或者是具备本地更新特性的文档存储系统（如 MongoDB）。
- 海量数据（如数据仓库中需要分析的数据）适合存储在一个结构松散、分布式的文件存储系统中，如 Hadoop。
- 存储二进制文件（如 mp3 或者 pdf 文档）并且能够直接为用户的浏览器提供下载功能，可以使用 Amazon S3。
- 临时性的数据（如网站的 session、缓存 HTML 页面信息等）适合存储在 Memcache 中。
- 如果希望数据具备高可用性，并且能够将数据丢失的风险降到最低，同时整个系统具备线性扩展的能力，可以考虑使用 Cassandra 和 HBase。

使用这些数据产品并不是要取代原有的数据产品，而是为不同的应用场景提供更多的选择。NoSQL 代表着：**选择合适的方案处理合适的业务场景**。上面介绍的几种 NoSQL 应用场景也许能够帮助我们选择合适的数据存储方案，网上也有不少值得参考的资源。和其他的技术方案一样，选择适合的业务场景才是最重要的。

绝大多数的应用都有非常复杂的应用场景，怎样才能找出一款 NoSQL 产品能够适用于所有的需求？答案是搭配使用多款 NoSQL 产品。传统数据库中的通用（One For All）的情况在 NoSQL 中是不存在的。

如图 1-1 所示，可以在一个网站中使用 4 款数据产品来提供服务。

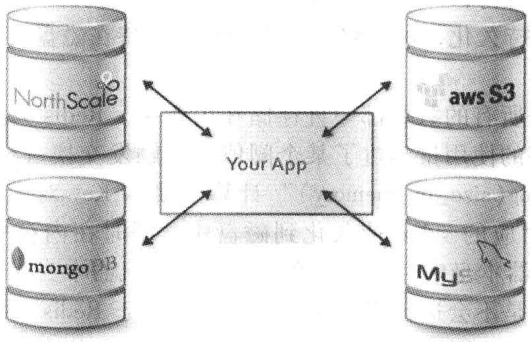


图 1-1　NoSQL 产品组合

- MySQL 用于存储敏感的数据，比如用户的资料、交易的信息等。
- MongoDB 用于存储大量的、相对不敏感的数据，比如博客文章的内容、文章访问次数等。
- Amazon S3 用于存储用户上传的文档、图片、音乐等数据。
- Memcached 用于存储临时性的信息，比如缓存 HTML 页面等。

选择多样的数据存储方案同样有利于提升我们对 NoSQL 的数据产品的理解，帮助我们从大量的解决方案中选择最适用的产品，而不是把眼光仅仅放在某一款产品上。

核心的思想是：**最适用的才是最好的**。

1.3　开源 NoSQL 产品介绍

1.3.1　Key/Value 的 NoSQL 数据库

1. Memcached

Memcached 是国外社区网站 LiveJournal 开发的高性能的内存 Key/Value 缓存服务器，目的是通过缓存数据库查询结果，减少数据库访问次数，以提高动态 Web 应用的速度，从而提高系统的可扩展性。