

<<Cassandra实战>>

图书基本信息

书名：<<Cassandra实战>>

13位ISBN编号：9787111341642

10位ISBN编号：7111341643

出版时间：2011-6

出版时间：机械工业

作者：郭鹏

页数：303

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.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能够给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《Cassandra实战》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。

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。

## &lt;&lt;Cassandra实战&gt;&gt;

## 内容概要

《Cassandra实战》是国内首本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。

《Cassandra实战》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。

《Cassandra实战》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。

《Cassandra实战》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。

## 作者简介

郭鹏，国内Cassandra领域的先驱者和实践者。

资深软件开发工程师，擅长分布式应用程序的开发和使用，时间经验极其丰富。

一直致力于Cassandra、Hadoop、Hive、MongoDB等分布式应用的底层实现原理的研究和探索，对分布式数据库和数据仓库也有深刻而独到的理解。

曾任阿里巴巴数据仓库开发工程师，参与设计和开发了多个基于Cassandra和Hadoop的大型应用。

现供职于盛大在线，参与Hadoop计算存储平台的搭建与实施。

活跃于开源技术社区，是博客园的推荐博客，也是CSDN的博客专家，热衷于与网友分享自己的心得和体会，深受社区欢迎。

博客地址：<http://gpcuster.cnblogs.com/>；新浪微博：@逖靖寒。

## &lt;&lt;Cassandra实战&gt;&gt;

## 书籍目录

前言

第1章 认识NoSQL

1.1 NoSQL的起源和发展现状

1.2 为什么要使用NoSQL

1.3 开源NoSQL产品介绍

1.4 本章小结

第2章 Cassandra快速入门

2.1 在Windows环境运行单机版Cassandra

2.2 在Linux环境运行单机版Cassandra

2.3 Cassandra的数据模型

2.4 Cassandra的数据排序规则

2.5 配置数据类型

2.6 使用命令行工具与Cassandra交互

2.7 本章小结

第3章 理解Cassandra编程接口

3.1 多语言服务开发框架Thrift

3.2 Cassandra的数据类型

3.3 Cassandra的编程接口

3.4 Cassandra 0.7.x版本新增功能

3.5 本章小结

第4章 基于Cassandra的在线交易系统

4.1 需求分析

4.2 数据模型设计

4.3 编码实现

4.4 系统功能验证

4.5 迁移到Cassandra 0.7.x

4.6 本章小结

第5章 Cassandra的集群机制

第6章 Cassandra的内部数据存储结构

第7章 Cassandra的数据更新机制

第8章 Cassandra的数据读取机制

第9章 Cassandra的数据压缩机制

第10章 Cassandra的启动流程

第11章 在分布式环境中使用的Cassandra

第12章 Cassandra与Hadoop的整合

第13章 Cassandra最佳实践

附录A 在Eclipse中修改Cassandra源代码

附录B CassSeller代码

附录C CassSeller-0.7代码

章节摘录

版权页：插图：

## &lt;&lt;Cassandra实战&gt;&gt;

## 媒体关注与评论

毫无疑问，NoSQL数据库的出现让数据库的选择性变得更加丰富。

Facebook公司推出的Cassandra无疑是开源NoSQL产品中一颗闪亮新星。

Facebook、Twitter、Digg等公司对Cassandra的使用已经证明了它不是一个玩具产品，其完全可以在复杂的生产环境下承担重要的角色。

同时，越来越多的关系数据库也开始关注NoSQL对数据库技术的影响，如最近MySQL数据库下的InnoDB存储引擎已经提供了NoSQL的访问方式。

《Cassandra实战》对Cassandra的配置、应用、编译等内容进行了详细的介绍，想对Cassandra一探究竟的读者千万不能错过。

——姜承尧 资深MySQL数据库专家，著有经典著作《MySQL技术内幕：InnoDB存储引擎》 Cassandra最初是由Facebook开发的一套开源的分布式NoSQL数据库系统，它同时具备了Google BigTable的数据模型和Amazon Dynamo的完全分布式架构，具有良好的可扩展性，目前被很多大型的Web2.0网站所使用，是一种流行的分布式结构化数据存储方案。

《Cassandra实战》作者维护Cassandra数据库已经有很长时间，具有丰富的的一线工作经验，同时本书还结合源码对一些底层的机制和原理进行了分析，值得初中级读者参考。

——杨海朝 新浪网（中国）技术有限公司首席DBA/新浪微博数据库负责人NoSQL是IT领域当下讨论最热烈的技术话题之一。

2010年，Cassandra的去中心化和无缝扩展的特性吸引了众多NoSQL粉丝的眼球，成为了NoSQL阵营中的一道亮丽风景线。

本书既能引领我们全面掌握Cassandra的基础知识，又能帮助我们深入了解Cassandra的运行机制与原理。

相信在阅读完本书后，我们将会对Cassandra有一个全面而深入的认识。

对于喜爱NoSQL的我来说，看到国内有相关的书籍出版，非常兴奋，特此向所有关注NoSQL的朋友推荐《Cassandra实战》。

——孙立 数据库专家/NoSQL先驱/去哪儿网高级系统架构师有很多技术在出现后不久就淡出了人们的视野，导致这种情况发生的原因有很多，其中一个重要的原因就是没有合适的应用环境。

在这一点上，NoSQL技术无疑是幸运的。

因为我们所熟知的关系型数据库在很多超大规模、高并发的Web 2.0网站应用中面临很多无法克服的问题，所以近几年NoSQL技术越来越受到关注和重视。

Amazon的Dynamo和Google的BigTable都是非常成功的商业NoSQL产品，很多开源的NoSQL产品也得到了长足的发展，Cassandra就是其中之一。

《Cassandra实战》由浅入深地介绍了Cassandra的基础知识和它在生产环境中的应用，无疑是一本全面了解Cassandra的好书，本书将引领你走进无比精彩的NoSQL世界。

——张勤 著有《linux服务器配置全程实录》

## &lt;&lt;Cassandra实战&gt;&gt;

## 编辑推荐

《Cassandra实战》：《Cassandra实战》内容全面，系统地讲解了Cassandra的所有功能特性及其使用方法。

《Cassandra实战》深入浅出，结合源代码分析了Cassandra的底层机制和工作原理。

《Cassandra实战》由国内Cassandra领域的驱者和实践者亲自执笔，多位数据库专家联袂推荐。

以Face book和Twitter为代表的Web 2. 应用的数据的爆炸性增长为后台存储系统带来了两个新的挑战：海量数据存储数据查询低延时为了设计出能够存储海量数据并实现低延时的数据库架构，Google和Amazon分别给出了自己的实现技术，并且能满足不同的需求。

Google开发了BigTable，Amazon开发了Dynamo。

这两种实现都具有高扩展性和高可用性，并且可以在廉价的服务器中稳定运行，节约了大量的硬件成本。

Face book基于Google和Amazon的研究成果开发出了Cassandra，它不仅具备BigTable强大的数据模型，而且还具备Dynamo的简洁性，例如端到端数据备份，以及强大的错误容灾功能等。

2008年，Face book将Cassandra贡献给了开源社区。

今天，Cassandra已经成为实现访问低延时和海量数据存储的首选架构。

大量的公司和机构都开始使用Cassandra构建实时应用程序以处理海量的数据，其中包括Netflix、TWitter和Cisco等国际知名企业。

Cassandra的开发社区非常活跃，它在0.6版中与Hadoop进行了整合，使得Cassandra中的数据能够被Hadoop处理；在0.7版中，它又新增了二级索引和在线更新Schema等功能。

更让人期待的是，Cassandra将实现HDFS的接口，成为HDFS和HBase的替换方案，能提供更好的性能。

《Cassandra实战》内容全面，实战性强，不仅系统地讲解了Cassandra的使用方法，而且还从源代码的角度对Cassandra的实现原理和工作机制进行了分析和探讨，深度、广度和实践性完美结合，相信能给广大Cassandra用户带来一定的帮助。



#### 版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>