<<深入NoSQL>>

图书基本信息

书名：<<深入NoSQL>>

13位ISBN编号：9787115296382

10位ISBN编号：7115296383

出版时间：2012-11

出版时间：人民邮电出版社

作者：Shashank Tiwari

页数：294

字数：455000

译者：巨成

前言

随着用户内容的增长，所生成、处理、分析和归档的数据的规模快速增大，类型也快速增多。此外，一些新数据源也在生成大量数据，比如传感器、全球定位系统（GPS）、自动追踪器和监控系统。
这些大数据集通常被称为大数据，它们给存储、分析和归档带来了新的机遇与挑战。
数据不仅仅快速增长，而且半结构化和稀疏的趋势也很明显。
这样一来，预定义好schema和利用关系型引用的传统数据管理技术就受到了挑战。
在探索海量数据和半结构化数据相关问题的过程中，诞生了一系列新型数据库产品，其中包括列族数据库（column-oriented data store）、键/值数据库和文档数据库，这些数据库统称NoSQL。
NoSQL产品千变万化，特性和价值主张各有不同，因此常常难以选择。
本书能帮助你理解整个NoSQL领域。
书中展示了构建许多NoSQL产品的基本概念，覆盖了相对较多的NoSQL产品，而非单单深入介绍某一种产品。
本书主要关注广度和基本概念，而不是全面覆盖每一种产品的API。
因为要介绍不少NoSQL产品，所以会涉及大量的比较分析。
如果不确定如何开始用NoSQL以及如何学习管理和分析大数据，那么你会发现本书是一本很好的入门指南和参考用书。
读者对象　　本书的主要目标读者是开发者、架构师、数据库管理员以及技术项目经理，但任何理解数据库技术的人可能都会觉得本书很有帮助。
计算机专业的许多学生和研究员也会对大数据和NoSQL这一主题感兴趣，他们会因阅读本书而获益良多。
任何开始进行大数据分析和使用NoSQL的人也都能从本书中受益。
本书内容　　本书首先介绍NoSQL的基础知识，然后逐步过渡到围绕性能调优和架构性指引的高阶概念上。
我们主要关注与NoSQL相关的基本概念，并以许多不同的NoSQL产品为例来解释它们。
书中包括了有关MongoDB、CouchDB、HBase、Hypertable、Cassandra、Redis和BerkeleyDB的演示和样例，此外还包括其他一些NoSQL产品。
NoSQL很重要的一部分就是大数据处理。
本书将介绍基于MapReduce的可伸缩处理，演示Hadoop用例，还有Hive和Pig这样的高层抽象。
第10章关注云NoSQL，介绍Amazon Web Service和Google App Engine提供的平台。
本书包含许多用例演示，同时也会讨论Google、Amazon、Facebook、Twitter和LinkedIn的可伸缩数据架构。
在最后一部分中，本书会比较NoSQL产品，讨论不同产品在应用程序栈中共存的话题。
本书结构　　本书分为四大部分：　　NoSQL入门　　NoSQL基础　　熟悉NoSQL　　掌握NoSQL　　每部分的内容都是以前一部分为基础的。
第一部分为NoSQL入门，定义了NoSQL产品的类型，并初次介绍了用NoSQL存储和访问数据的几个例子。
第1章定义NoSQL。
第2章以超经典的Hello World程序开头，介绍了几个使用NoSQL的例子。
第3章介绍NoSQL产品交互与接口。
第二部分介绍了各种NoSQL产品的一些基本概念。
第4章解释存储架构。
第5章和第6章介绍基本的数据管理，演示CRUD操作和查询机制。
数据集随时间和使用情况而演变。
第7章探讨数据演变相关的问题。
传统的关系型数据库关注利用索引来优化查询。

第8章介绍NoSQL下的索引。
对NoSQL产品缺少事务支持的批评往往过头。
第9章陈清事务相关概念，以及分布式系统所面临的事务完整性挑战。
第三、四部分介绍高阶话题。
第10章介绍Google App Engine数据存储和Amazon SimpleDB。
很多大数据处理有赖于MapReduce风格的处理方式。
第11章介绍MapReduce的基本知识。
第12章扩展了MapReduce的覆盖范围，以演示Hive为Hadoop MapReduce任务提供的类SQL抽象。
第13章回顾了数据库架构及内部结构。
第五部分是本书的最后一部分。
第14章比较NoSQL产品；第15章提出了共存的想法，以及按需选择数据库的观点；第16章谈论可伸缩应用程序的优化。
本部分的话题看似驳杂，却为实际应用NoSQL打下了基础。
第17章展示一些工具和实用程序，部署NoSQL时用得上。
阅读必备　　请参照各处代码示例安装相应的软件，安装步骤和设置说明参见附录A。
本书约定　　为了描述得更加清楚明了，在本书中我们有如下约定。
本格式表示针对目前讨论内容的注释、小技巧、提示等。
段落样式如下。
楷体表示新词及重点词。
文件名、URL和书中的代码使用如下这种字体：persistence.properties。
代码有两种展示方法：大部分代码样例使用monofont字体，无高亮；当前上下文中重要的代码以及与之前代码段不同的代码加粗显示。
源代码　　对于本书所有例子中的代码，你可以手工输入，也可以使用本书附带的源码文件。
所有源码可下载。
在网站上通过本书书名（使用检索框或书名列表）进入本书的详情页，点击下载代码的链接即可获取源码。
网站上包含的代码都附有下面的图标：　　代码示例的标题里包含源码文件名。
如果只是代码片段，则文件名如是：　　Code snippet flename　　因为很多书名很相似，所以按ISBN号可能查起来更容易。
本书ISBN号是978-0-470-94224-6。
下载好的代码可以用解压缩工具打开。
除此之外，在下载页面也能看到本书及Wrox其他图书的代码。
勘误　　虽然我们尽全力消除文本和代码中的所有错误，但错误总是难以避免的。
如果你发现了本书的错误，比如拼写错误或问题代码，请反馈给我们，非常感谢！
如果你能指出一个问题，也许就能避免另一位读者的困惑，同时也能帮助我们提高本书的质量。
要前往本书的勘误，通过书名进入本书的详情页，然后点击勘误（Book Errata）链接。
在该页面中，你可以看到由读者提交并由Wrox编者发布的本书的所有勘误。
要想获得完整的图书列表及每本书的勘误链接，也可以访问。
如果在勘误页上没找到你发现的错误，填写表单提交你发现的错误。
我们会尽快确认信息，如果勘误正确，会在勘误页上发布，并在本书的后续版本中修正问题。
若想与包括本书作者在内的同行们讨论，请加入P2P论坛。
这个Web论坛主要用于发布Wrox图书信息及相关技术消息，以及与其他读者及技术用户互动。
论坛提供主题订阅功能，选定了自己感兴趣的主题后，每当有相应主题的新帖发布时，论坛会以电子邮件的形式通知你。
Wrox作者、编者、其他行业专家和读者都会在这里活动。
你会发现一系列不同的论坛，它们不仅对你阅读本书有所帮助，同样也能在你开发应用时提供帮助。

加入论坛请参照下面的步骤。

(1) 点击注册链接。
(2) 阅读使用守则并点击“ 同意 ”。
(3) 输入必填信息及其他你乐于提供的可选信息，并点击“ 提交 ”。
(4) 你会收到一封邮件，其中描述了如何确认你的账户和完成加入论坛的流程。

阅读消息无需在P2P上注册，但如果要发布消息，就必须加入论坛。

加入论坛后，可以发布新消息和回复他人发布的消息。

任何时候你都可以通过Web浏览论坛。

如果希望通过电子邮件接收特定论坛的新消息，请在论坛列表中点击论坛名称旁边的“ 订阅本论坛 ”图标。

要了解更多有关Wrox P2P的信息，请阅读P2P FAQ（点击P2P页面上的FAQ链接）了解论坛使用方法，以及P2P和Wrox图书常见问题。

<<深入NoSQL>>

内容概要

《深入NoSQL》是一本全面的NoSQL实践指南。
书中主要关注NoSQL的基本概念，以及使用NoSQL数据库的切实可行的解决方案。
书中介绍了基于MapReduce的可伸缩处理，演示Hadoop用例，还有Hive和Pig这样的高层抽象。
《深入NoSQL》包含许多用例演示，同时也会讨论Google、Amazon、Facebook、Twitter和LinkedIn的可伸缩数据架构。

《深入NoSQL》适合NoSQL数据库管理人员和开发人员阅读。

作者简介

Shashank Tiwari，创业者、开发者、技术作家、演讲者和导师，技术性创业公司Treasury of Ideas的创始人。
他是一位经验丰富的软件开发者和企业家，长期关注高性能应用、分析、Web应用以及移动平台，对数据可视化和统计机器学习有着浓厚的兴趣，喜欢喝咖啡、吃甜点、骑自行车。
他撰写了许多技术文章和著作，并且应邀在全球各地的技术会议上进行演讲。

<<深入NoSQL>>

书籍目录

章节摘录

版权页： 插图： 现在，随着隔离级别设置为可重复读取，步骤1和步骤4返回给事务1的数据集并不一样。
在步骤4中，除原来的三条记录外，还能看到id为4的记录。
为了避免幻象读，需要为读施加范围锁并使用最高级别的隔离：可序列化。
可序列化意味着顺序处理，或者说串行的事务处理，但事实并非总是如此。
在一些数据库里，可序列化隔离是通过快照来实现的。
这些数据库在每个事务开始时为事务提供一个快照，然后只允许那些自创建快照以来没有发生任何改变的事务进行提交。
使用更高的隔离级别会增大饿死（starvation）和死锁的可能性。
一个事务锁住了其他事务要使用的资源时会发生饿死，而两个并发事务相互等待对方释放资源时会发生死锁。
回顾完ACID事务和隔离级别的概念，现在可以开始探讨如何在高度分布式系统里运用这些想法了。
9.2 分布式ACID系统 要理解ACID是否适用于分布式系统，首先要了解分布式系统的各种属性，看看ACID对它们有哪些影响。
分布式系统有各种不同的形状、大小和形式，但是它们都具备几个典型特征，并且也面临相似的复杂性问题。
随着分布式系统逐渐增大并伸展出去，复杂性的挑战更为突出。
不止如此，如果系统需要提供高可用性，那么挑战会成倍增加。

<<深入NoSQL>>

编辑推荐

云计算时代关键数据库技术全面展示NoSQL基础概念和实践方案 理解大数据的各种技术架构和思想 全面的NoSQL实践指南

版权说明