

<<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

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

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。

<<深入NoSQL>>

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章澄清事务相关概念，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时用得着。

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 filename` 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体后，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。

<<深入NoSQL>>

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

第一部分 NoSQL入门

第1章 NoSQL的概念及适用范围

1.1 定义和介绍

1.1.1 背景与历史

1.1.2 大数据

1.1.3 可扩展性

1.1.4 MapReduce

1.2 面向列的有序存储

1.3 键/值存储

1.4 文档数据库

1.5 图形数据库

1.6 小结

第2章 NoSQL上手初体验

2.1 第一印象——两个简单的例子

2.1.1 简单的位置偏好数据集

2.1.2 存储汽车品牌 and 型号数据

2.2 使用多种语言

2.2.1 MongoDB驱动

2.2.2 初识Thrift

2.3 小结

第3章 NoSQL接口与交互

3.1 没了SQL还剩什么

3.1.1 存储和访问数据

3.1.2 MongoDB数据存储与访问

3.1.3 MongoDB数据查询

3.1.4 Redis数据存储与访问

3.1.5 Redis数据查询

3.1.6 HBase数据存储与访问

3.1.7 HBase数据查询

3.1.8 Apache Cassandra数据存储与访问

3.1.9 Apache Cassandra数据查询

3.2 NoSQL数据存储的语言绑定

3.2.1 Thrift

3.2.2 Java

3.2.3 Python

3.2.4 Ruby

3.2.5 PHP

3.3 小结

第二部分 NoSQL基础

第4章 理解存储架构

4.1 使用面向列的数据库

4.1.1 使用关系型数据库中的表格和列

4.1.2 列数据库对比RDBMS

4.1.3 列数据库当做键/值对的嵌套映射表

4.1.4 Wehtable布局

<<深入NoSQL>>

- 4.2 HBase分布式存储架构
- 4.3 文档存储内部机制
 - 4.3.1 用内存映射文件存储数据
 - 4.3.2 MongoDB集合和索引使用指南
 - 4.3.3 MongoDB的可靠性和耐久性
 - 4.3.4 水平扩展
- 4.4 键/值存储Memcached和Redis
 - 4.4.1 Memcached的内部结构
 - 4.4.2 Redis的内部结构
- 4.5 最终一致性非关系型数据库
 - 4.5.1 一致性哈希
 - 4.5.2 对象版本
 - 4.5.3 闲话协议和提示移交
- 4.6 小结
- 第5章 执行CRUD操作
 - 5.1 创建记录
 - 5.1.1 在以文档为中心的数据库中创建记录
 - 5.1.2 面向列数据库的创建操作
 - 5.1.3 键/值映射表的创建操作
 - 5.2 访问数据
 - 5.2.1 用MongoDB访问文档
 - 5.2.2 用HBase访问数据
 - 5.2.3 查询Redis
 - 5.3 更新和删除数据
 - 5.3.1 使用MongoDB、HBase和Redis更新及修改数据
 - 5.3.2 有限原子性和事务完整性
 - 5.4 小结
- 第6章 查询NoSQL存储
 - 6.1 SQL与MongoDB查询功能的相似点
 - 6.1.1 加载MovieLens数据
 - 6.1.2 MongoDB中的MapReduce
 - 6.2 访问HBase等面向列数据库中的数据
 - 6.3 查询Redis数据存储
 - 6.4 小结
- 第7章 修改数据存储及管理演进
 - 7.1 修改文档数据库
 - 7.1.1 弱schema的灵活性
 - 7.1.2 MongoDB的数据导入与导出
 - 7.2 面向列数据库中数据schema的演进
 - 7.3 HBase数据导入与导出
 - 7.4 键/值存储中的数据演变
 - 7.5 小结
- 第8章 数据索引与排序
 - 8.1 数据库索引的基本概念
 - 8.2 MongoDB的索引与排序
 - 8.3 MongoDB里创建和使用索引
 - 8.3.1 组合与嵌套键

<<深入NoSQL>>

- 8.3.2 创建唯一索引和稀疏索引
- 8.3.3 基于关键字的搜索和多重键
- 8.4 CouchDB的索引与排序
- 8.5 Apache Cassandra的索引与排序
- 8.6 小结

第9章 事务和数据完整性的管理

- 9.1 RDBMS和ACID
- 9.2 分布式ACID系统
 - 9.2.1 一致性
 - 9.2.2 可用性
 - 9.2.3 分区容忍性
- 9.3 维持CAP
 - 9.3.1 妥协可用性
 - 9.3.2 妥协分区容忍性
 - 9.3.3 妥协一致性
- 9.4 NoSQL产品的一致性实现
 - 9.4.1 MongoDB的分布一致性
 - 9.4.2 CouchDB的最终一致性
 - 9.4.3 Apache Cassandra的最终一致性
 - 9.4.4 Membase的一致性
- 9.5 小结

第三部分 熟悉NoSQL

第10章 使用云中的NoSQL

- 10.1 Google App Engine
 - 10.1.1 GAE Python SDK : 安装、设置和起步
 - 10.1.2 使用Python进行基本的GAE数据建模
 - 10.1.3 查询与索引
 - 10.1.4 过滤和结果排序
 - 10.1.5 Java App Engine SDK
- 10.2 Amazon SimpleDB
 - 10.2.1 SimpleDB入门
 - 10.2.2 使用REST API
 - 10.2.3 使用Java访问SimpleDB
 - 10.2.4 通过Ruby和Python使用SimpleDB
- 10.3 小结

第11章 MapReduce可扩展并行处理

- 11.1 理解MapReduce
 - 11.1.1 找出每股最高价
 - 11.1.2 加载历史NYSE市场数据到CouchDB
- 11.2 MapReduce和HBase
- 11.3 MapReduce和Apache Mahout
- 11.4 小结

第12章 使用Hive分析大数据

- 12.1 Hive基础
- 12.2 回到电影评分
- 12.3 亲切的SQL
- 12.4 HiveQL连接

<<深入NoSQL>>

- 12.4.1 计划解释
- 12.4.2 分区表
- 12.5 小结
- 第13章 综览数据库内部
 - 13.1 MongoDB内部
 - 13.1.1 MongoDB传输协议
 - 13.1.2 插入文档
 - 13.1.3 查询集合
 - 13.1.4 MongoDB数据库文件
 - 13.2 Membase架构
 - 13.3 Hypertable底层
 - 13.3.1 正则表达式支持
 - 13.3.2 布隆过滤器
 - 13.4 Apache Cassandra
 - 13.4.1 点对点模型
 - 13.4.2 基于Gossip和Antientropy
 - 13.4.3 快速写
 - 13.4.4 提示移交
 - 13.5 Berkeley DB
 - 13.6 小结
- 第四部分 掌握NoSQL
- 第14章 选择NoSQL
 - 14.1 比较NoSQL产品
 - 14.1.1 可扩展性
 - 14.1.2 事务完整性和一致性
 - 14.1.3 数据模型
 - 14.1.4 查询支持
 - 14.1.5 接口可用性
 - 14.2 性能测试
 - 14.2.1 50/50的读和更新
 - 14.2.2 95/5的读和更新
 - 14.2.3 扫描
 - 14.2.4 可扩展性测试
 - 14.2.5 Hypertable测试
 - 14.3 背景比较
 - 14.4 小结
- 第15章 共存
 - 15.1 MySQL用作NoSQL
 - 15.2 静态数据存储
 - 15.2.1 存储多元化在Facebook中的应用
 - 15.2.2 数据仓库和商业智能
 - 15.3 Web框架和NoSQL
 - 15.3.1 Rails和NoSQL
 - 15.3.2 Django和NoSQL
 - 15.3.3 使用Spring Data
 - 15.4 从RDBMS迁移到NoSQL
 - 15.5 小结

<<深入NoSQL>>

第16章 性能调校

16.1 并行算法的目标

16.1.1 减少延迟的含义

16.1.2 如何增加吞吐

16.1.3 线性扩展

16.2 公式与模型

16.2.1 Amdahl法则

16.2.2 Little法则

16.2.3 消息成本模型

16.3 分区

16.4 规划异构环境

16.5 其他MapReduce调校

16.5.1 通信成本

16.5.2 压缩

16.5.3 文件块大小

16.5.4 并行复制

16.6 HBase Coprocessor

16.7 布隆过滤器

16.8 小结

第17章 工具和实用程序

17.1 RRDTool

17.2 Nagios

17.3 Scribe

17.4 Flume

17.5 Chukwa

17.6 Pig

17.6.1 使用Pig

17.6.2 Pig Latin基础

17.7 Nodetool

17.8 OpenTSDB

17.9 SOLANDRA

17.10 Hummingbird和C5T

17.11 GeoCouch

17.12 Alchemy Database

17.13 Webdis

17.14 小结

附录A 安装与配置

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。

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

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>