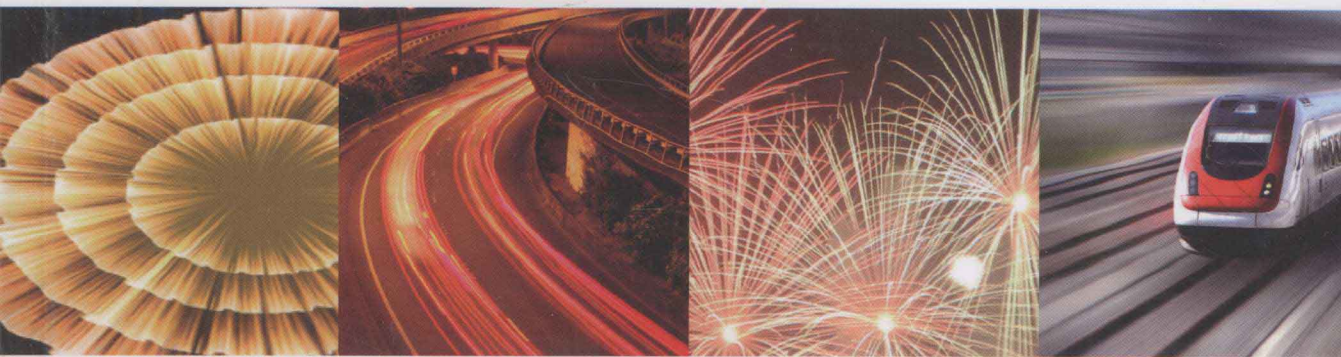


大数据挑战与 NoSQL数据库技术



陆嘉恒 编著

BIG DATA

大数据丛书

大数据挑战与 NoSQL数据库技术

陆嘉恒 编著



电子工业出版社
Publishing House of Electronics Industry
北京·BEIJING

内 容 简 介

本书共分为三部分。理论篇重点介绍大数据时代下数据处理的基本理论及相关处理技术，并引入 NoSQL 数据库；系统篇主要介绍了各种类型 NoSQL 数据库的基本知识；应用篇对国内外几家知名公司在利用 NoSQL 数据库处理海量数据方面的实践做了阐述。

本书对大数据时代面临的挑战，以及 NoSQL 数据库的基本知识做了清晰的阐述，有助于读者整理思路，了解需求，并更有针对性、有选择地深入学习相关知识。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有，侵权必究。

图书在版编目（CIP）数据

大数据挑战与 NoSQL 数据库技术 / 陆嘉恒编著. —北京：电子工业出版社，2013.4
（大数据丛书）

ISBN 978-7-121-19660-7

I. ①大… II. ①陆… III. ①数据处理②数据库系统 IV. ①TP274 ②TP311.13

中国版本图书馆 CIP 数据核字(2013)第 035612 号

责任编辑：许 艳

印 刷：三河市鑫金马印装有限公司

装 订：三河市鑫金马印装有限公司

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本：787×980 1/16 印张：27.5 字数：446.8 千字

印 次：2013 年 4 月第 1 次印刷

印 数：4000 册 定价：79.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：(010) 88254888。

质量投诉请发邮件至 zltz@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：(010) 88258888。

前 言

为什么写本书

计算机技术已经深刻地影响了我们的工作、学习和生活。大数据及 NoSQL 技术是现今 IT 领域最炙手可热的话题，其发展非常迅速，潜力巨大，悄然改变着整个行业的面貌。随着 Web 2.0 技术的发展，微博、社交网络、电子商务、生物工程等的不断发展，各领域数据呈现爆炸式的增长，传统关系型数据库显得越来越力不从心。NoSQL 数据库技术的出现为当前面临的问题提供了新的解决方案，它摒弃了传统关系型数据库 ACID 的特性，采用分布式多节点的方式，更加适合大数据的存储和管理。

政府和高校都十分重视对大数据及 NoSQL 技术的研究和投入；在产业界，各大 IT 公司也在投入大量的资源研究和开发相关的 NoSQL 产品，与之相应的新兴技术和产品正在不断涌现。这一切都极大地推动了 NoSQL 技术的发展。

大数据处理和 NoSQL 技术涉及的内容繁多，目前不同公司也有不同的 NoSQL 数据库产品，而且某一产品往往是为特定的应用而设计的，并不一定能够适用于所有的场景。很多人在学习的初始阶段需要进行大量的摸索和实践，然而目前这方面系统的参考资料却非常少。为了便于所有想了解和掌握 NoSQL 技术的朋友学习，并在学习的过程中少走弯路，笔者将自己在该领域的经验和知识的积累凝聚在本书，希望能够推动大数据处理及 NoSQL 相关技术在国内的发展。

本书面向的读者

在编写本书时，我们力图使不同背景和职业的读者都能从其中获益。

如果你是专业技术人员，本书将带领你快速地进入大数据处理及 NoSQL 的世界，全面掌握 NoSQL 及其相关技术，帮助你使用 NoSQL 技术解决面临的问题或提供必要的参考。

如果你是高等院校计算机及相关专业的学生，本书为你在课堂之外了解最新的 IT 打开一扇窗，帮助你拓宽视野，完善知识结构，为迎接未来的挑战做好知识储备。

在学习本书之前，应具有如下的基础：

- 有一定的 Linux 操作系统的基础知识。
- 有较好的编程基础和阅读代码的能力。
- 对数据库知识有一定的了解。

如何阅读本书

本书一共包括 16 章，分为三个部分。其中第一部分为理论篇，包括：大数据产生的背景、数据一致性理论、数据存储模型、数据分区与防治策略、海量数据处理方法、数据复制与容错技术、数据压缩技术和数据缓存技术。此部分重点从理论上介绍、分析大数据管理过程中遇到的各方面问题。第二部分为系统篇，包括：键值数据库、列存数据库、文档数据库、图存数据库、基于 Hadoop 的数据库管理系统、NoSQL 数据库以及分布式缓存系统。该部分以理论篇为基础，根据数据存储模型对数据库类型进行划分，每一部分以具体开源数据库为实例进行介绍，涉及系统的架构、安装以及使用等方面知识，力图使读者对 NoSQL 数据库有具体的认识。第三部分为应用篇，包括企业应用以及经验总结和对未来的展望。该部分介绍企业如何使用 NoSQL 数据库解决自身遇到的问题。

在阅读本书时，读者可以先系统地学习理论篇的知识，对海量数据处理方法有一个很好的理解，在此基础上，读者可以对后面的章节进行选择性的学习。本书涉及内容较多，从开源数据库方面讲，包括了 Dynamo、Redis、Voldemort、Cassandra、Hypertable、CouchDB、MongoDB、Neo4j、GraphDB、OrientDB、HBase、Hive、Pig、MySQL Cluster、VolteDB、MS-Velocity、Memcached 等将近 20 个数据库。因此，建议读者可以重点学习感兴趣或有一定需求的数据库系统。当然，如果时间允许，还是建议读者系统地学习本书的内容。

另外，在系统篇的学习过程中，建议读者能够一边阅读，一边根据书中的指导进行实践，亲自实践本书中所给出的编程范例。

致谢

在本书的编写过程中，还有很多 NoSQL 领域的实践者和研究者为本书做了大量的工作，他们是张林林、许翔、程明、王海涌、顾向楠、吴少辉、杨宁、杨华、吴梦迪、任乔意、於洋、张轩等，在此特别感谢。

在线资源及勘误

本书官方网站为：<http://datasearch.ruc.edu.cn/NoSQL/>。本书的勘误、讨论以及相关资料等

都会在该网站上发布和更新。

在本书的撰写和相关技术的研究中，尽管笔者投入了大量的精力，付出了艰辛的努力，然而受知识水平所限，错误和疏漏之处在所难免，恳请大家批评指正。如果有任何问题和建议，可发送邮件至 jiahenglu@gmail.com 或 jiahenglu@ruc.edu.cn。

陆嘉恒

目 录

第 1 章 概论	1
1.1 引子	2
1.2 大数据挑战	3
1.3 大数据的存储和管理	5
1.3.1 并行数据库	5
1.3.2 NoSQL 数据管理系统	6
1.3.3 NewSQL 数据管理系统	8
1.3.4 云数据管理	11
1.4 大数据的处理和分析	11
1.5 小结	13
参考文献	13

理论篇

第 2 章 数据一致性理论	16
2.1 CAP 理论	17
2.2 数据一致性模型	21
2.3 ACID 与 BASE	22
2.4 数据一致性实现技术	23
2.4.1 Quorum 系统 NRW 策略	23
2.4.2 两阶段提交协议	24
2.4.3 时间戳策略	27
2.4.4 Paxos	30
2.4.5 向量时钟	38

2.5 小结	43
参考文献	43
第3章 数据存储模型	45
3.1 总论	46
3.2 键值存储	48
3.2.1 Redis	49
3.2.2 Dynamo	49
3.3 列式存储	50
3.3.1 Bigtable	51
3.3.2 Cassandra 与 HBase	51
3.4 文档存储	52
3.4.1 MongoDB	53
3.4.2 CouchDB	53
3.5 图形存储	54
3.5.1 Neo4j	55
3.5.2 GraphDB	56
3.6 小结	56
参考文献	56
第4章 数据分区与放置策略	58
4.1 分区的意义	59
4.1.1 为什么要分区	59
4.1.2 分区的优点	60
4.2 范围分区	61
4.3 列表分区	62
4.4 哈希分区	63
4.5 三种分区的比较	64
4.6 放置策略	64
4.6.1 一致性哈希算法	65
4.6.2 容错性与可扩展性分析	66
4.6.3 虚拟节点	68
4.7 小结	69

参考文献	69
第 5 章 海量数据处理方法	70
5.1 MapReduce 简介	71
5.2 MapReduce 数据流	72
5.3 MapReduce 数据处理	75
5.3.1 提交作业	76
5.3.2 初始化作业	78
5.3.3 分配任务	78
5.3.4 执行任务	79
5.3.5 更新任务执行进度和状态	80
5.3.6 完成作业	81
5.4 Dryad 简介	81
5.4.1 DFS Cosmos 介绍	82
5.4.2 Dryad 执行引擎	84
5.4.3 DryadLINQ 解释引擎	86
5.4.4 DryadLINQ 编程	88
5.5 Dryad 数据处理步骤	90
5.6 MapReduce vs Dryad	92
5.7 小结	94
参考文献	95
第 6 章 数据复制与容错技术	96
6.1 海量数据复制的作用和代价	97
6.2 海量数据复制的策略	97
6.2.1 Dynamo 的复制策略	97
6.2.2 CouchDB 的复制策略	99
6.2.3 PNUTS 的复制策略	99
6.3 海量数据的故障发现与处理	101
6.3.1 Dynamo 的故障发现与处理	101
6.3.2 CouchDB 的故障发现与处理	103
6.3.3 PNUTS 的故障发现与处理	103
6.4 小结	104

参考文献	104
第 7 章 数据压缩技术	105
7.1 数据压缩原理	106
7.1.1 数据压缩的定义	106
7.1.2 数据为什么可以压缩	107
7.1.3 数据压缩分类	107
7.2 传统压缩技术 ^[1]	108
7.2.1 霍夫曼编码	108
7.2.2 LZ77 算法	109
7.3 海量数据带来的 3V 挑战	112
7.4 Oracle 混合列压缩	113
7.4.1 仓库压缩	114
7.4.2 存档压缩	114
7.5 Google 数据压缩技术	115
7.5.1 寻找长的重复串	115
7.5.2 压缩算法	116
7.6 Hadoop 压缩技术	118
7.6.1 LZ0 简介	118
7.6.2 LZ0 原理 ^[5]	119
7.7 小结	121
参考文献	121
第 8 章 缓存技术	122
8.1 分布式缓存简介	123
8.1.1 分布式缓存的产生	123
8.1.2 分布式缓存的应用	123
8.1.3 分布式缓存的性能	125
8.1.4 衡量可用性的标准	125
8.2 分布式缓存的内部机制	125
8.2.1 生命期机制	126
8.2.2 一致性机制	126
8.2.3 直读与直写机制	129

8.2.4 查询机制	130
8.2.5 事件触发机制	130
8.3 分布式缓存的拓扑结构	130
8.3.1 复制式拓扑	131
8.3.2 分割式拓扑	131
8.3.3 客户端缓存拓扑	131
8.4 小结	132
参考文献	132

系统篇

第 9 章 key-value 数据库	134
9.1 key-value 模型综述	134
9.2 Redis	135
9.2.1 Redis 概述	135
9.2.2 Redis 下载与安装	135
9.2.3 Redis 入门操作	136
9.2.4 Redis 在业内的应用	143
9.3 Voldemort	143
9.3.1 Voldemort 概述	143
9.3.2 Voldemort 下载与安装	144
9.3.3 Voldemort 配置	145
9.3.4 Voldemort 开发介绍 ^[3]	147
9.4 小结	149
参考文献	149
第 10 章 Column-Oriented 数据库	150
10.1 Column-Oriented 数据库简介	151
10.2 Bigtable 数据库	151
10.2.1 Bigtable 数据库简介	151
10.2.2 Bigtable 数据模型	152
10.2.3 Bigtable 基础架构	154
10.3 Hypertable 数据库	157

10.3.1	Hypertable 简介	157
10.3.2	Hypertable 安装	157
10.3.3	Hypertable 架构	163
10.3.4	Hypertable 中的基本概念和原理	164
10.3.5	Hypertable 的查询	168
10.4	Cassandra 数据库	175
10.4.1	Cassandra 简介	175
10.4.2	Cassandra 配置	175
10.4.3	Cassandra 数据库的连接	177
10.4.4	Cassandra 集群机制	180
10.4.5	Cassandra 的读/写机制	182
10.5	小结	183
	参考文献	183
第 11 章	文档数据库	185
11.1	文档数据库简介	186
11.2	CouchDB 数据库	186
11.2.1	CouchDB 简介	186
11.2.2	CouchDB 安装	188
11.2.3	CouchDB 入门	189
11.2.4	CouchDB 查询	200
11.2.5	CouchDB 的存储结构	207
11.2.6	SQL 和 CouchDB	209
11.2.7	分布式环境中的 CouchDB	210
11.3	MongoDB 数据库	211
11.3.1	MongoDB 简介	211
11.3.2	MongoDB 的安装	212
11.3.3	MongoDB 入门	215
11.3.4	MongoDB 索引	224
11.3.5	SQL 与 MongoDB	226
11.3.6	MapReduce 与 MongoDB	229
11.3.7	MongoDB 与 CouchDB 对比	234

11.4 小结	236
参考文献	237
第 12 章 图存数据库	238
12.1 图存数据库的由来及基本概念	239
12.1.1 图存数据库的由来	239
12.1.2 图存数据库的基本概念	239
12.2 Neo4j 图存数据库	240
12.2.1 Neo4j 简介	240
12.2.2 Neo4j 使用教程	241
12.2.3 分布式 Neo4j——Neo4j HA	251
12.2.4 Neo4j 工作机制及优缺点浅析	256
12.3 GraphDB	258
12.3.1 GraphDB 简介	258
12.3.2 GraphDB 的整体架构	260
12.3.3 GraphDB 的数据模型	264
12.3.4 GraphDB 的安装	266
12.3.5 GraphDB 的使用	268
12.4 OrientDB	276
12.4.1 背景	276
12.4.2 OrientDB 是什么	276
12.4.3 OrientDB 的原理及相关技术	277
12.4.4 Windows 下 OrientDB 的安装与使用	282
12.4.5 相关 Web 应用	286
12.5 三种图存数据库的比较	288
12.5.1 特征矩阵	288
12.5.2 分布式模式及应用比较	289
12.6 小结	289
参考文献	290
第 13 章 基于 Hadoop 的数据管理系统	291
13.1 Hadoop 简介	292
13.2 HBase	293

13.2.1	HBase 体系结构	293
13.2.2	HBase 数据模型	297
13.2.3	HBase 的安装和使用	298
13.2.4	HBase 与 RDBMS	303
13.3	Pig	304
13.3.1	Pigr 的安装和使用	304
13.3.2	Pig Latin 语言	306
13.3.3	Pig 实例	311
13.4	Hive	315
13.4.1	Hive 的数据存储	316
13.4.2	Hive 的元数据存储	316
13.4.3	安装 Hive	317
13.4.4	HiveQL 简介	318
13.4.5	Hive 的网络接口 (WebUI)	328
13.4.6	Hive 的 JDBC 接口	328
13.5	小结	330
	参考文献	331
第 14 章	NewSQL 数据库	332
14.1	NewSQL 数据库简介	333
14.2	MySQL Cluster	333
14.2.1	概述	334
14.2.2	MySQL Cluster 的层次结构	336
14.2.3	MySQL Cluster 的优势和应用	337
14.2.4	海量数据处理中的 sharding 技术	339
14.2.5	单机环境下 MySQL Cluster 的安装	343
14.2.6	MySQL Cluster 的分布式安装与配置指导	348
14.3	VoltDB	350
14.3.1	传统关系数据库与 VoltDB	351
14.3.2	VoltDB 的安装与配置	351
14.3.3	VoltDB 组件	354
14.3.4	Hello World	355

14.3.5	使用 Generate 脚本	361
14.3.6	Eclipse 集成开发	362
14.4	小结	365
	参考文献	365
第 15 章	分布式缓存系统	366
15.1	Memcached 缓存技术	367
15.1.1	背景介绍	367
15.1.2	Memcached 缓存技术的特点	368
15.1.3	Memcached 安装 ^[3]	374
15.1.4	Memcached 中的数据操作	375
15.1.5	Memcached 的使用	376
15.2	Microsoft Velocity 分布式缓存系统	378
15.2.1	Microsoft Velocity 简介	378
15.2.2	数据分类	379
15.2.3	Velocity 核心概念	380
15.2.4	Velocity 安装	382
15.2.5	一个简单的 Velocity 客户端应用	385
15.2.6	扩展型和可用性	387
15.3	小结	388
	参考文献	388

应用篇

第 16 章	企业应用	392
16.1	Instagram	393
16.1.1	Instagram 如何应对数据的急剧增长	395
16.1.2	Instagram 的数据分片策略	398
16.2	Facebook 对 Hadoop 以及 HBase 的应用	400
16.2.1	工作负载类型	401
16.2.2	为什么采用 Apache Hadoop 和 HBase	403
16.2.3	实时 HDFS	405
16.2.4	Hadoop HBase 的实现	409

16.3 淘宝大数据解决之道.....	411
16.3.1 淘宝数据分析.....	412
16.3.2 淘宝大数据挑战.....	413
16.3.3 淘宝 OceanBase 数据库.....	414
16.3.4 淘宝将来的工作.....	422
16.4 小结.....	423
参考文献.....	423

第1章

概 论

“这是最好的时代，这是最坏的时代；这是智慧的时代，这是愚蠢的时代；这是信仰的时期，这是怀疑的时期；这是光明的季节，这是黑暗的季节；这是希望之春，这是绝望之冬；人们面前什么都有，人们面前一无所有；人们正在直登天堂，人们正在直下地狱。”

——狄更斯《双城记》

对于数据管理界来说，这是一个充满挑战的时代。急速增长的数据让人们焦头烂额，传统关系型数据库在扩展性方面的瓶颈让人们无所适从：如何存储大数据，如何处理大数据，如何挖掘大数据，大数据已经成为数据管理界的新挑战。这又是一个充满机遇的时代，新的系统孕育而出，百花齐放，它们“标新立异”，它们“独树一帜”，它们在数据模型、事务处理等方面采取不同的策略解决海量数据带来的问题。这注定是一段不平凡的岁月。