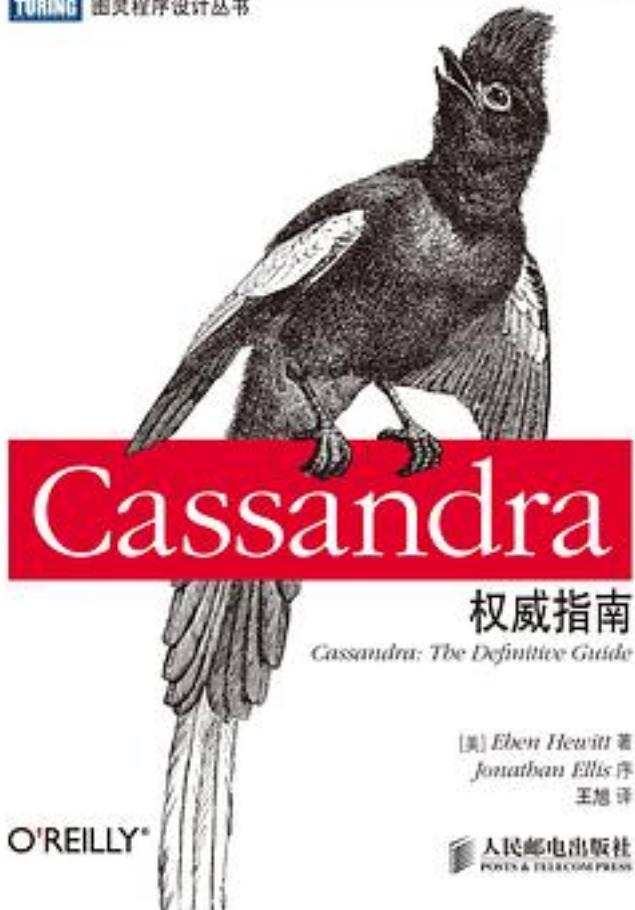


# Cassandra权威指南

TURING 图灵程序设计丛书



[Cassandra权威指南\\_下载链接1](#)

著者:Eben Hewitt

出版者:人民邮电出版社

出版时间:2011-8

装帧:平装

isbn:9787115258540

本书是一本广受好评的Cassandra图书。与传统的关系型数据库不同，Cassandra是一种开源的分布式存储系统。书中介绍了它无中心架构、高可用、无缝扩展等引人注目的特点，讲述了如何安装、配置Cassandra。

及如何在其上运行实例，还介绍了对它的监控、维护和性能调优手段，同时还涉及了Cassandra相关的集成工具Hadoop及其类似的其他NoSQL数据库。

本书适合数据库开发人员与网站开发者阅读。

作者介绍：

Eben Hewitt

跨国公司应用架构总监，负责系统战略和设计工作。他是Apache Cassandra项目的一位文档贡献者，同时也是多本技术书籍的作者，其中包括Java SOA Cookbook (O'Reilly出版)。

目录: 译者序 XIII

序 XV

前言 XVII

第1章 Cassandra 概况 1

1.1 关系型数据库有什么问题 1

1.2 关系型数据库简单回顾 5

1.2.1 RDBMS：出类拔萃与表现平平 6

1.2.2 互联网的规模 12

1.3 Cassandra 的电梯间演讲 13

1.3.1 50 个字介绍Cassandra 13

1.3.2 分布式与无中心 13

1.3.3 弹性可扩展 14

1.3.4 高可用与容错 15

1.3.5 可调节的一致性 15

1.3.6 Brewer 的CAP 理论 18

1.3.7 面向行 21

1.3.8 无schema 22

1.3.9 高性能 22

1.4 Cassandra 来自何方 22

1.5 Cassandra 的应用场景 23

1.5.1 大规模部署 23

1.5.2 写密集、统计和分析型工作 24

1.5.3 地区分布 24

1.5.4 变化的应用 24

1.6 谁在使用Cassandra 24

1.7 小结 26

第2章 安装Cassandra 27

2.1 安装二进制包 27

2.1.1 解压缩 27

2.1.2 里面有什么 27

2.2 从源码编译 28

2.2.1 其他编译目标 30

2.2.2 使用Maven编译 30

2.3 运行Cassandra 30

2.3.1 在Windows平台上运行Cassandra 31

2.3.2 在Linux下运行Cassandra 31

2.3.3 启动服务器 32

2.4 使用命令行界面的客户端 33

2.5 基本命令行命令 34

2.5.1 帮助	34
2.5.2 连接服务器	35
2.5.3 描述环境	35
2.5.4 创建keyspace和列族	36
2.5.5 读写数据	37
2.6 小结	38
第3章 Cassandra的数据模型	39
3.1 关系型数据模型	39
3.2 简介	40
3.3 集群	43
3.4 keyspace	43
3.5 列族	44
3.6 列	46
3.6.1 宽行与窄行	48
3.6.2 列的排序	49
3.7 超级列	50
3.8 Cassandra与RDBMS的设计差别	53
3.8.1 没有查询语言	53
3.8.2 没有引用完整性	53
3.8.3 第二索引	53
3.8.4 排序成为一种设计决策	54
3.8.5 反范式化	54
3.9 设计模式	55
3.9.1 具体化视图	56
3.9.2 无值列	56
3.9.3 聚合键	56
3.10 需要记住的几件事	57
3.11 小结	57
第4章 应用实例	59
4.1 数据模型设计	59
4.2 酒店应用的关系型数据库设计	60
4.3 酒店应用的Cassandra设计	61
4.4 酒店应用代码	62
4.4.1 创建数据库	63
4.4.2 数据结构	64
4.4.3 进行连接	65
4.4.4 预装填数据库	66
4.4.5 搜索应用	78
4.5 Twissandra	82
4.6 小结	82
第5章 Cassandra的架构	83
5.1 system keyspace	83
5.2 对等结构	84
5.3 gossip与故障检测	84
5.4 逆熵与读修复	86
5.5 memtable、SSTable和commit log	87
5.6 提示移交	89
5.7 压紧	89
5.8 Bloom filter	91
5.9 墓碑	91
5.10 分阶段事件驱动架构	92
5.11 管理器与服务	93
5.11.1 Cassandra守护进程	93
5.11.2 存储服务	93

5.11.3 消息服务	93
5.11.4 提示移交管理器	94
5.12 小结	94
第6章 配置Cassandra	95
6.1 keyspace	95
6.1.1 创建列族	98
6.1.2 从0.6迁移到0.7	99
6.2 副本	99
6.3 副本放置策略	100
6.3.1 简单策略	101
6.3.2 旧网络拓扑策略	102
6.3.3 网络拓扑策略	103
6.4 副本因子	103
6.5 分区器	105
6.5.1 随机分区器	106
6.5.2 有序分区器	106
6.5.3 配页有序分区器	107
6.5.4 字节序分区器	107
6.6 Snitch	107
6.6.1 Simple Snitch	107
6.6.2 PropertyFileSnitch	107
6.7 创建集群	108
6.7.1 修改集群名称	109
6.7.2 给集群增加节点	109
6.7.3 多种子节点	111
6.8 动态加入环	113
6.9 安全	114
6.9.1 使用SimpleAuthenticator	114
6.9.2 编程鉴权	117
6.9.3 使用MD5加密	118
6.9.4 提供你自己的鉴权算法	118
6.10 杂项设置	119
6.11 附加工具	120
6.11.1 查看键值	120
6.11.2 导入之前版本的配置	120
6.12 小结	122
第7章 读写数据	123
7.1 Cassandra与RDBMS查询的不同	123
7.1.1 没有Update查询	123
7.1.2 记录级的写原子性	123
7.1.3 不支持服务端事务	123
7.1.4 没有重复键值	124
7.2 写操作的基本属性	124
7.3 一致性级别	124
7.4 读操作的基本属性	126
7.5 API	126
7.6 设置与插入数据	128
7.7 使用简单的get	133
7.8 数据准备	135
7.9 切片谓词	135
7.9.1 使用get_slice读取特定列名	136
7.9.2 通过切片区间获取一组列	137
7.9.3 取出一行中的所有列	138
7.10 get_range_slices	138

7.11 multiget_slice	140
7.12 删除	142
7.13 批量变更	144
7.13.1 批量删除	144
7.13.2 区间鬼影	145
7.14 编程定义keyspace和列族	145
7.15 小结	146
第8章 客户端	147
8.1 基本的客户端API	148
8.2 Thrift	148
8.2.1 Thrift对Java的支持	151
8.2.2 异常	151
8.2.3 Thrift小结	152
8.3 Avro	152
8.3.1 Avro Ant目标	154
8.3.2 Avro规范	155
8.3.3 Avro小结	156
8.4 Git简介	156
8.5 连接客户端节点	157
8.5.1 客户端列表	157
8.5.2 循环DNS	157
8.5.3 负载均衡器	157
8.6 Cassandra Web控制台	157
8.7 Hector (Java)	161
8.7.1 特性	161
8.7.2 Hector API	162
8.8 HectorSharp (C#)	162
8.9 Chirper	167
8.10 Chiton (Python)	167
8.11 Pelops (Java)	168
8.12 Kundera (Java ORM)	169
8.13 Fauna (Ruby)	169
8.14 小结	170
第9章 监控	171
9.1 日志	171
9.1.1 跟踪查看	173
9.1.2 通用技巧	174
9.2 JMX与MBean概述	175
9.2.1 MBean	177
9.2.2 集成JMX	179
9.3 通过JMX与Cassandra交互	180
9.4 Cassandra的MBean	181
9.4.1 org.apache.cassandra.concurrent	185
9.4.2 org.apache.cassandra.db	185
9.4.3 org.apache.cassandra.gms	186
9.4.4 org.apache.cassandra.service	186
9.5 定制Cassandra的MBean	188
9.6 运行时分析工具	190
9.6.1 使用JMX和JHAT进行堆分析	191
9.6.2 发现线程问题	194
9.7 健康检查	195
9.8 小结	196
第10章 维护	197
10.1 获取环的信息	198

10.1.1 Info	198
10.1.2 Ring	198
10.2 获取统计信息	199
10.2.1 使用cfstats	199
10.2.2 使用tpstats	200
10.3 基本维护工作	201
10.3.1 修复	201
10.3.2 刷写	202
10.3.3 清理	203
10.4 快照	203
10.4.1 进行快照	203
10.4.2 清除快照	204
10.5 对集群进行负载均衡	205
10.6 退服节点	208
10.7 更新节点	210
10.7.1 删除令牌	210
10.7.2 压紧阈值	210
10.7.3 在一个工作的集群中改变列族	210
10.8 小结	211
第11章 性能调优	213
11.1 数据存储	213
11.2 回复超时	215
11.3 commit log	215
11.4 memtable	216
11.5 并发	216
11.6 缓存	217
11.7 缓冲区尺寸	218
11.8 使用Python压力测试	218
11.8.1 生成Python Thrift接口	218
11.8.2 运行Python压力测试	220
11.9 启动和JVM设置	222
11.10 小结	224
第12章 集成Hadoop	225
12.1 何为Hadoop	225
12.2 使用MapReduce	226
12.3 运行字数统计例子	227
12.3.1 将数据输出到Cassandra	229
12.3.2 Hadoop流	229
12.4 MapReduce之上的工具	229
12.4.1 Pig	230
12.4.2 Hive	231
12.5 集群配置	231
12.6 案例	233
12.6.1 Raptr.com: Keith Thornhill	233
12.6.2 Imagini: Dave Gardner	233
12.7 小结	234
附录 非关系型数据库大观	235
词汇表	261
关于作者	279
关于封面	279
• • • • • (收起)	

标签

NoSQL

Cassandra

数据库

分布式

计算机

O'Reilly

计算机科学

权威指南

评论

相对于Erlang编程那本书的300多页，这本书有点贵了——里面似乎是蜻蜓点水般地讲了一堆Java工具集。。。靠，英文版的早看过了

---

简单枯燥，不如看官方文档

---

内容略老

原来丫面向列的数据库系统是这么奇葩。。我还是去学习CouchDB吧。。总感觉JSON更舒服点。

了解一下。

工作平时经常用cassandra，对原理从来没有深究，中文版虽然省力，但是很多地方让我想去读读原版咋写的，有空还得读下原版。

没仔细看出版日期，书中版本已太老（0.7）现在已经是2.1啦。理论的部分稍微啰嗦了点，后面半本书基本没啥用途。。只能看看官方文档

很多概念很好，特别是附录里对各种NoSQL的介绍，蛮有价值的。

随便翻翻。这是个值得学习的nosql。

主要是应用为主，原理、实现讲的过少

其中的java访问cassandra的案例，切片涉及到的谓词之类（predict）概念，也真是看得不愉快。太啰嗦。案例hotel看完后，作者推荐了cassandra by example写的真好。你早说啊。你的案例也tm太不好理解了。跟别人学学。

cassandra工具书。。。

纯粹技术性书籍写得一点都不枯燥，翻译得恰如其分

版本太老

入门和基本了解还不错

用来了解一些基本概念和原理还是可以的，但是对于了解cassandra本身来说，就太过于时了。

整体不错，特别是词汇表和不同数据库的对比

[Cassandra权威指南 下载链接1](#)

## 书评

语言上略显罗嗦，例子之类的难免有凑篇幅的嫌疑。不过，就了解Cassandra和NoSQL来说，还是不错的，有助于整体了解NoSQL的起源、分类，以及Cassandra的各种设计决策。

Cassandra目前风头没有一年前那么热了，但仍不失为一个非常出色的数据存储系统，值得研究。但因为Cassandra...

Cassandra是什么？有什么特点？谁在用它？为什么用它？

这本书第一章只用了不到30页就很好地阐述了这些问题。这本书的目标是帮助开发者和数据库管理员们理解Cassandra这种新型数据库，探索它与传统的关系型数据库系统有什么异同，并且帮助读者在自己的系统中使用Cassandra。 ...

本来Cassandra的书籍就很少，这本书作为入门书籍来说还是不错的，从原理到实践都作了介绍，但如果你用它来作开发手册，信息量就太少了。同时，阅读这本书之前，最好对关系型数据库有所了解，并且有一定的编程经验，不然很难理解文字背后的含义。如果真的在做项目，那么推荐 ...

我看过了 我看过了 我看过了 我看过了 我看过了 我看过了 我看过了 我看过了  
我看过了 我看过了 我看过了 我看过了 我看过了 我看过了 我看过了 我看过了  
我看过了 我看过了 我看过了 我看过了 我看过了 我看过了 我看过了 我看过了  
我看过了 我看...  
-----

现在都2.1了，书上介绍的还是0.7，和官方文档相差太大了，买这本书不如去看官方文档。很苦恼的，工作中用到了cassandra，但是完全没有中文相关的资料可以参考，什么时候才可以出新版啊。 . . . . .

作为一个application层的开发人员，本书的例子还是太少了，hector也只讲了短短的一节。感觉更加适合DBA的角色来读。  
不过市面上好像没有什么关于cassandra和hector的书籍，网上的文档也不是很多，是不是因为cassandra还是小众的数据库？  
-----

[Cassandra权威指南 下载链接1](#)