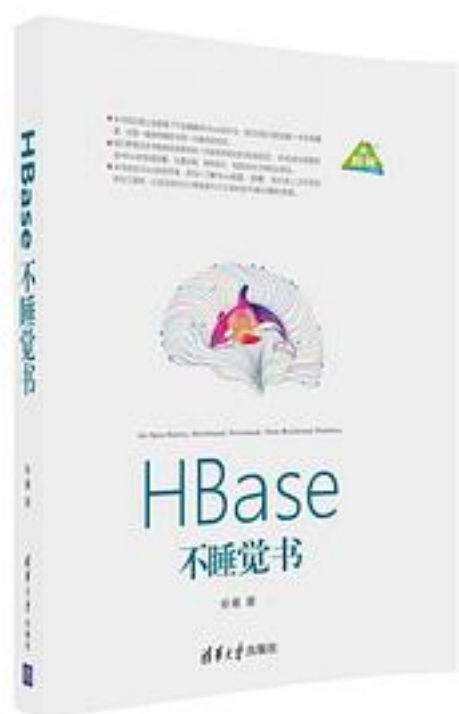


HBase不睡觉书



[HBase不睡觉书 下载链接1](#)

著者:杨曦

出版者:清华大学出版社

出版时间:2018-1-1

装帧:平装

isbn:9787302490555

HBase是Apache旗下一个高可靠性、高性能、面向列、可伸缩的分布式存储系统。利用HBase技术可在廉价PC服务器上搭建起大规模的存储化集群。使用HBase可以对数十亿级别的大数据进行实时性的高性能读写，在满足高性能的同时还保证了数据存取的原子性。

本书共分为10章，由浅入深的讲解HBase概念、安装、配置、部署。让读者对HBase先有一个感性认识，再从应用角度，介绍了高级用法、监控和性能调优。既兼顾了初学者也适用于想要深入学习HBase的读者。

本书适合于以前没有接触过HBase，或者了解HBase并希望能够深入掌握的读者，适合HBase应用开发人员和系统管理人员学习使用。

作者介绍:

杨曦，就职于硅谷某上市公司，对大数据、云计算等技术有较深研究以及项目实践经验。热衷创业、编写开源项目、撰写技术博客。

目录: 第1章 基础知识	1
1.1 Node是什么	1
1.1.1 Node与JavaScript	1
1.1.2 Runtime 和 VM	2
第1章 初识HBase	1
1.1 海量数据与NoSQL	1
1.1.1 关系型数据库的极限	1
1.1.2 CAP理论	1
1.1.3 NoSQL	2
1.2 HBase是怎么来的	3
1.3 为什么要用HBase	3
1.4 你必须懂的基本概念	4
1.4.1 部署架构	4
1.4.2 存储架构	7
1.4.3 跟关系型数据库的对比	9
第2章 让HBase跑起来	11
2.1 本书测试环境	12
2.2 配置服务器名	12
2.3 配置SSH免密登陆	13
2.4 安装Hadoop	15
2.4.1 安装Hadoop单机模式	15
2.4.2 安装Hadoop集群模式	20
2.4.3 ZooKeeper	23
2.4.4 配置Hadoop HA	27
2.4.5 让Hadoop可以开机自启动	36
2.4.6 最终配置文件	42
2.5 安装HBase	44
2.5.1 单机模式	46
2.5.2 伪分布式模式	48
2.5.3 关于ZooKeeper不得不说的	53
2.5.4 完全分布式模式	54
2.5.5 HBase Web控制台 (UI)	59
2.5.6 让HBase可以开机自启动	60
2.5.7 启用数据块编码 (可选)	62
2.5.8 启用压缩器	67
2.5.9 数据块编码还是压缩器? (可选)	71
第3章 HBase基本操作	72
3.1 HBase Shell的使用	72
3.1.1 用create命令建表	73
3.1.2 用list命令来查看库中有哪些表	74
3.1.3 用describe命令来查看表属性	74
3.1.4 用put命令来插入数据	76
3.1.5 用scan来查看表数据	77
3.1.6 用get来获取单元格数据	78

- 3.1.7 用delete来删除数据 78
- 3.1.8 deleteall来删除整行记录 80
- 3.1.9 用disable来停用表 81
- 3.1.10 用drop来删除表 81
- 3.1.11 Shell命令列表 82
- 3.2 使用Hue来查看HBase数据 123
 - 3.2.1 准备工作 123
 - 3.2.2 安装Hue 126
 - 3.2.3 配置Hue 129
 - 3.2.4 使用Hue来查看HBase 134
- 第4章 客户端API入门 136
 - 4.1. 10分钟教程 136
 - 4.2 30分钟教程 143
 - 4.3 CRUD一个也不能少 149
 - 4.3.1 HTable类和Table接口 149
 - 4.3.2 put方法 150
 - 4.3.3 append方法 157
 - 4.3.4 increment方法 159
 - 4.3.5 get方法 160
 - 4.3.6 exists 方法 164
 - 4.3.7 delete方法 164
 - 4.3.8 mutation方法 166
 - 4.4 批量操作 168
 - 4.4.1 批量put操作 169
 - 4.4.2 批量get操作 169
 - 4.4.3 批量delete操作 170
 - 4.5 BufferedMutator（可选） 170
 - 4.6 Scan扫描 172
 - 4.6.1 用法 172
 - 4.6.2 缓存 175
 - 4.7 HBase支持什么数据格式？ 176
 - 4.8 总结 177
- 第5章 HBase内部探险 178
 - 5.1 数据模型 178
 - 5.2 HBase是怎么存储数据的 180
 - 5.2.1 宏观架构 180
 - 5.2.2 预写日志 WAL 183
 - 5.2.3 MemStore 186
 - 5.2.4 HFile（StoreFile） 186
 - 5.2.5 KeyValue类 188
 - 5.2.6 增删查改的真正面目 188
 - 5.2.7 数据单元层次图 189
 - 5.3 一个KeyValue的历险 190
 - 5.3.1 写入 190
 - 5.3.2 读出 191
 - 5.4 Region的定位 192
- 第6章 客户端API的高阶用法 195
 - 6.1 过滤器 195
 - 6.1.1 过滤器快速入门 196
 - 6.1.2 比较运算快速入门 201
 - 6.1.3 分页过滤器（PageFilter） 204
 - 6.1.4 过滤器列表（FilterList） 206
 - 6.1.5 行键过滤器 211
 - 6.1.6 列过滤器 216

- 6.1.7 单元格过滤器 230
- 6.1.8 装饰过滤器 (decorating filter) 232
- 6.1.9 自定义过滤器 235
- 6.1.10 如何在hbase shell中使用过滤器 252
- 6.2 协处理器 (coprocessor) 252
 - 6.2.1 协处理器家族 253
 - 6.2.2 快速入门 255
 - 6.2.3 如何加载 258
 - 6.2.4 协处理器核心类 260
 - 6.2.5 观察者 (Observers) 262
 - 6.2.6 终端程序 (EndPoint) 279
- 第7章 客户端API的管理功能 293
 - 7.1 列族管理 293
 - 7.2 表管理 299
 - 7.3 Region管理 302
 - 7.4 快照 (snapshot) 管理 307
 - 7.5 维护工具管理 311
 - 7.5.1 均衡器 (Balancer) 311
 - 7.5.2 规整器 (Normalizer) 312
 - 7.5.3 目录管理器 (catalog janitor) 313
 - 7.6 集群状态以及负载 (ClusterStatus & ServerLoad) 314
 - 7.7 Admin的其他方法 318
 - 7.8 可见性标签管理 322
 - 7.8.1 快速入门 324
 - 7.8.2 系统标签 331
 - 7.8.3 用户标签 332
 - 7.8.4 单元格标签 332
- 第8章 再快一点! 334
 - 8.1 Master和RegionServer的JVM调优 334
 - 8.1.1 先调大堆内存 334
 - 8.1.2 可怕的Full GC 336
 - 8.1.3 Memstore的专属JVM策略MSLAB 338
 - 8.2 Region的拆分 343
 - 8.2.1 Region的自动拆分 344
 - 8.2.2 Region的预拆分 (pre-splitting) 348
 - 8.2.3 Region的强制拆分 (forced splits) 349
 - 8.2.4 推荐方案 350
 - 8.2.5 总结 350
 - 8.3 Region的合并 (merge) 350
 - 8.3.1 通过Merge类合并 (merge) Region 351
 - 8.3.2 热合并 (online_merge) 351
 - 8.4 WAL的优化 352
 - 8.5 BlockCache的优化 354
 - 8.5.1 LRUBlockCache 355
 - 8.5.2 SlabCache 356
 - 8.5.3 BucketCache 357
 - 8.5.4 组合模式 (CombinedBlockCache) 359
 - 8.5.5 总结 360
 - 8.6 Memstore的优化 360
 - 8.6.1 读写中的Memstore 360
 - 8.6.2 Memstore的刷写 (flush) 361
 - 8.6.3 总结 363
 - 8.7 HFile (StoreFile) 的合并 (Compaction) 364
 - 8.7.1 合并的策略 364

- 8.7.2 Compaction的吞吐量限制参数 377
- 8.7.3 合并的时候HBase做了什么? 380
- 8.7.4 Major Compaction 381
- 8.7.5 总结 383
- 8.8 诊断手册 383
 - 8.8.1 阻塞急救 383
 - 8.8.2 朱丽叶暂停 (Juliet Pause) 384
 - 8.8.3 读取性能调优 387
 - 8.8.4 案例分析 388
- 第9章 当HBase遇上MapReduce 392
 - 9.1 为什么要用MapReduce 392
 - 9.2 快速入门 392
 - 9.3 慢速入门: 编写自己的MapReduce 394
 - 9.3.1 准备数据 394
 - 9.3.2 新建项目 395
 - 9.3.3 建立MapReduce类 396
 - 9.3.4 建立驱动类 399
 - 9.3.5 打包、部署、运行 403
 - 9.4 相关类介绍 405
 - 9.4.1 TableMapper 405
 - 9.4.2 TableReducer 406
 - 9.4.3 TableMapReduceUtil 406
 - • • • • ([收起](#))

[HBase不睡觉书_下载链接1](#)

标签

HBase

大数据

Hadoop

计算机

数据平台

IT

Buy

技术

评论

基于 HBase 1.2，入门极好，读的时候真的没有睡觉。（就是纸质太差）

这才叫会写书的作者

原理性的讲的还行，实践的没看

不错的入门书，不睡觉。HBase的部署安装，HBase shell的使用，数据存储结构，Java API的使用，看完基本对HBase有了了解，可以撸了。

不错的入门扫盲书，循序渐进引导阅读，以后实战再回来查阅

快速查阅，看了不会睡觉，需要什么就直接查阅

看的时候确实没有睡着，挺好的一本书，非常适合入门。

入门书，写的很详实，适合多次阅读

确实是不睡觉，内容比较生动

不睡觉果然只能蜻蜓点水…

hbase科普书籍

终于找到一个版本新一点的，虽然是入门书，但还是能学到很多东西，不过HBase的精髓都在官方文档里

比较不错的入门书籍，内容不是很深入

很有意思的一本书，上班抽空一刷了

大学图书馆里有，给我的参考还是挺有用的！

非常好的入门级，许多原理剖析的清晰明确。

力荐

[HBase不睡觉书_下载链接1](#)

书评

