

<<HBase权威指南>>

图书基本信息

书名：<<HBase权威指南>>

13位ISBN编号：9787115318893

10位ISBN编号：7115318891

出版时间：2013-10

出版时间：人民邮电出版社

译者：代志远,刘佳,蒋杰

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>

## <<HBase权威指南>>

### 内容概要

《hbase权威指南》探讨了如何通过使用与hbase高度集成的hadoop将hbase的可伸缩性变得简单；把大型数据集分布到相对廉价的商业服务器集群中；使用本地java客户端，或者通过提供了rest、avro和thrift应用编程接口的网关服务器来访问hbase；了解hbase架构的细节，包括存储格式、预写日志、后台进程等；在hbase中集成mapreduce框架；了解如何调节集群、设计模式、拷贝表、导入批量数据和删除节点等。

《hbase权威指南》适合使用hbase进行数据库开发的高级数据库研发人员阅读

## <<HBase权威指南>>

### 作者简介

lars george, hbase的committer, cloudera公司的解决方案架构师, 主要为hadoop和hbase提供技术支持、咨询和培训工作。

他多次在各种hadoop用户组会议和大型会议中发表演讲, 如布鲁塞尔的fosdem会议。

## &lt;&lt;HBase权威指南&gt;&gt;

## 书籍目录

- 《hbase权威指南》
- 第1章 简介 1
  - 1.1 海量数据的黎明 1
  - 1.2 关系数据库系统的问题 5
  - 1.3 非关系型数据库系统not-only-sql(简称nosql) 7
    - 1.3.1 维度 9
    - 1.3.2 可扩展性 12
    - 1.3.3 数据库的范式和反范式化 12
  - 1.4 结构 15
    - 1.4.1 背景 15
    - 1.4.2 表、行、列和单元格 16
    - 1.4.3 自动分区 20
    - 1.4.4 存储api 21
    - 1.4.5 实现 22
    - 1.4.6 小结 25
  - 1.5 hbase : hadoop数据库 25
    - 1.5.1 历史 26
    - 1.5.2 命名 27
    - 1.5.3 小结 27
- 第2章 安装 28
  - 2.1 快速启动指南 28
  - 2.2 必备条件 31
    - 2.2.1 硬件 31
    - 2.2.2 软件 37
  - 2.3 hbase使用的文件系统 47
    - 2.3.1 本地模式 48
    - 2.3.2 hdfs 49
    - 2.3.3 s3 49
    - 2.3.4 其他文件系统 50
  - 2.4 安装选项 50
    - 2.4.1 apache二进制发布包 50
    - 2.4.2 编译源码 52
  - 2.5 运行模式 53
    - 2.5.1 单机模式 53
    - 2.5.2 分布式模式 53
  - 2.6 配置 57
    - 2.6.1 hbase-site.xml与hbase-default.xml 58
    - 2.6.2 hbase-env.sh 59
    - 2.6.3 regionserver 59
    - 2.6.4 log4j.properties 59
    - 2.6.5 配置示例 59
    - 2.6.6 客户端配置 61
  - 2.7 部署 61
    - 2.7.1 基于脚本 62
    - 2.7.2 apache whirr 63

## &lt;&lt;HBase权威指南&gt;&gt;

- 2.7.3 puppet与chef 63
- 2.8 操作集群 64
  - 2.8.1 确定安装运行 64
  - 2.8.2 web ui介绍 65
  - 2.8.3 shell介绍 66
  - 2.8.4 关闭集群 66
- 第3章 客户端api：基础知识 68
  - 3.1 概述 68
  - 3.2 crud操作 69
    - 3.2.1 put方法 69
    - 3.2.2 get方法 87
    - 3.2.3 删除方法 97
  - 3.3 批量处理操作 107
  - 3.4 行锁 110
  - 3.5 扫描 114
    - 3.5.1 介绍 114
    - 3.5.2 resultscanner类 117
    - 3.5.3 缓存与批量处理 119
  - 3.6 各种特性 125
    - 3.6.1 htable的实用方法 125
    - 3.6.2 bytes类 127
- 第4章 客户端api：高级特性 129
  - 4.1 过滤器 129
    - 4.1.1 过滤器简介 129
    - 4.1.2 比较过滤器 132
    - 4.1.3 专用过滤器 139
    - 4.1.4 附加过滤器 147
    - 4.1.5 filterlist 151
    - 4.1.6 自定义过滤器 153
    - 4.1.7 过滤器总结 159
  - 4.2 计数器 160
    - 4.2.1 计数器简介 160
    - 4.2.2 单计数器 163
    - 4.2.3 多计数器 164
  - 4.3 协处理器 166
    - 4.3.1 协处理器简介 167
    - 4.3.2 coprocessor类 168
    - 4.3.3 协处理器加载 171
    - 4.3.4 regionobserver类 174
    - 4.3.5 masterobserver类 180
    - 4.3.6 endpoint 184
  - 4.4 htablepool 190
  - 4.5 连接管理 194
- 第5章 客户端api：管理功能 197
  - 5.1 模式定义 197
    - 5.1.1 表 197
    - 5.1.2 表属性 199

## &lt;&lt;HBase权威指南&gt;&gt;

- 5.1.3 列族 202
- 5.2 hbaseadmin 207
  - 5.2.1 基本操作 208
  - 5.2.2 表操作 209
  - 5.2.3 模式操作 217
  - 5.2.4 集群管理 219
  - 5.2.5 集群状态信息 222
- 第6章 可用客户端 230
  - 6.1 rest、thrift和avro的介绍 230
  - 6.2 交互客户端 233
    - 6.2.1 原生java 233
    - 6.2.2 rest 233
    - 6.2.3 thrift 240
    - 6.2.4 avro 244
    - 6.2.5 其他客户端 245
  - 6.3 批处理客户端 246
    - 6.3.1 mapreduce 246
    - 6.3.2 hive 246
    - 6.3.3 pig 252
    - 6.3.4 cascading 256
  - 6.4 shell 257
    - 6.4.1 基础 257
    - 6.4.2 命令 259
    - 6.4.3 脚本 263
  - 6.5 基于web的ui 265
    - 6.5.1 master的ui 265
    - 6.5.2 region服务器的ui 270
    - 6.5.3 共享页面 272
- 第7章 与mapreduce集成 275
  - 7.1 框架 275
    - 7.1.1 mapreduce介绍 275
    - 7.1.2 类 276
    - 7.1.3 支撑类 279
    - 7.1.4 mapreduce的执行地点 279
    - 7.1.5 表拆分 280
  - 7.2 在hbase之上的mapreduce 281
    - 7.2.1 准备 281
    - 7.2.2 数据流向 286
    - 7.2.3 数据源 291
    - 7.2.4 数据源与数据流向 293
    - 7.2.5 自定义处理 296
- 第8章 架构 299
  - 8.1 数据查找和传输 299
    - 8.1.1 b+树 299
    - 8.1.2 lsm树 300
  - 8.2 存储 302
    - 8.2.1 概览 303

## &lt;&lt;HBase权威指南&gt;&gt;

- 8.2.2 写路径 304
- 8.2.3 文件 305
- 8.2.4 hfile格式 313
- 8.2.5 keyvalue格式 316
- 8.3 wal 316
  - 8.3.1 概述 317
  - 8.3.2 hlog类 318
  - 8.3.3 hlogkey类 319
  - 8.3.4 waledit类 319
  - 8.3.5 logsyncer类 319
  - 8.3.6 logroller类 320
  - 8.3.7 回放 321
  - 8.3.8 持久性 324
- 8.4 读路径 325
- 8.5 region查找 328
- 8.6 region生命周期 330
- 8.7 zookeeper 330
- 8.8 复制 333
  - 8.8.1 log edit的生命周期 334
  - 8.8.2 内部机制 335
- 第9章 高级用法 339
  - 9.1 行键设计 339
    - 9.1.1 概念 339
    - 9.1.2 高表与宽表 341
    - 9.1.3 部分键扫描 342
    - 9.1.4 分页 343
    - 9.1.5 时间序列 344
    - 9.1.6 时间顺序关系 348
  - 9.2 高级模式 350
  - 9.3 辅助索引 350
  - 9.4 搜索集成 354
  - 9.5 事务 357
  - 9.6 布隆过滤器 358
  - 9.7 版本管理 361
    - 9.7.1 隐式版本控制 361
    - 9.7.2 自定义版本控制 364
- 第10章 集群监控 366
  - 10.1 介绍 366
  - 10.2 监控框架 367
    - 10.2.1 上下文、记录和监控指标 367
    - 10.2.2 master监控指标 372
    - 10.2.3 region服务器监控指标 373
    - 10.2.4 rpc监控指标 375
    - 10.2.5 jvm监控指标 376
    - 10.2.6 info监控指标 377
  - 10.3 ganglia 378
    - 10.3.1 安装 379

## &lt;&lt;HBase权威指南&gt;&gt;

- 10.3.2 用法 383
- 10.4 jmx 386
  - 10.4.1 jconsole 388
  - 10.4.2 jmx远程api 390
- 10.5 nagios 394
- 第11章 性能优化 395
  - 11.1 垃圾回收优化 395
  - 11.2 本地memstore分配缓冲区 398
  - 11.3 压缩 399
    - 11.3.1 可用的编解码器 400
    - 11.3.2 验证安装 401
    - 11.3.3 启用压缩 403
  - 11.4 优化拆分和合并 404
    - 11.4.1 管理拆分 404
    - 11.4.2 region热点 405
    - 11.4.3 预拆分region 406
  - 11.5 负载均衡 407
  - 11.6 合并region 408
  - 11.7 客户端api: 最佳实践 409
  - 11.8 配置 411
  - 11.9 负载测试 414
    - 11.9.1 性能评价 414
    - 11.9.2 ycsb 416
- 第12章 集群管理 421
  - 12.1 运维任务 421
    - 12.1.1 减少节点 421
    - 12.1.2 滚动重启 423
    - 12.1.3 新增服务器 424
  - 12.2 数据任务 428
    - 12.2.1 导入/导出 428
    - 12.2.2 copytable工具 433
    - 12.2.3 批量导入 435
    - 12.2.4 复制 438
  - 12.3 额外的任务 440
    - 12.3.1 集群共存 440
    - 12.3.2 端口要求 442
  - 12.4 改变日志级别 442
  - 12.5 故障处理 443
    - 12.5.1 hbase fsck 443
    - 12.5.2 日志分析 445
    - 12.5.3 常见问题 447
- 附录a hbase配置属性 451
- 附录b 计划 467
- 附录c 版本升级 469
- 附录d 分支 471
- 附录e hush sql schema 473
- 附录f 对比hbase和bigtable 475



<<HBase权威指南>>

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>