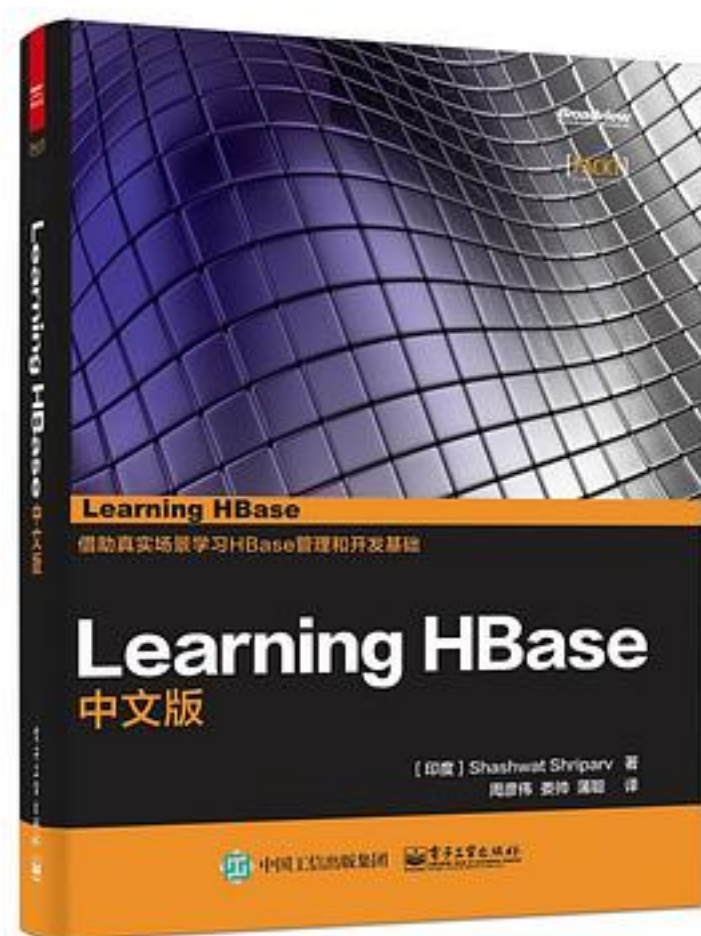


Learning Hbase中文版



[Learning Hbase中文版_下载链接1](#)

著者:周彦伟

出版者:电子工业出版社

出版时间:2015-9

装帧:

isbn:9787121270727

内容提要

《Learning Hbase中文版》是一本介绍HBase知识的专业书籍，它系统地介绍了HBase

的基本概念，与传统关系数据库的功能和特点的对比，自身的配置方法以及安装方法，同时深入介绍了HBase的运维管理和故障处理。《Learning Hbase中文版》还介绍了基于HBase的Java编程方法，以及HBase作为大数据工具的一些使用案例，这些足以帮助读者更好地理解HBase的架构，更顺利地在自己的项目中使用HBase。

《Learning Hbase中文版》不仅适合HBase初学者自学使用，也适合有HBase经验的开发人员作为工具查询之用，是一本针对HBase技术的比较完整的通用工具书，希望本书能在实际工作中对读者有所帮助。

作者介绍:

Shashwat

Shripav生于印度比哈尔邦穆扎法尔布尔县。他先后在穆扎法尔布尔和梅加拉亚邦西隆求学。他在德里英迪拉·甘地国立开放大学获得计算机应用学士（BCA）学位，在喀拉拉邦科钦科技大学（特里凡得琅 C-DAC）获得计算机应用硕士（MCA）学位。他在2010年早些时候开始研究大数据技术，当时他需要做一个用大数据技术存储和处理日志的概念验证（POC）。同时他还有另一个项目，在此项目中需要存储巨大的不同文件头的二进制文件并要处理它们。这时，他开始配置、搭建并测试 Hadoop HBase集群，并为它们写了一些代码。在做了一个成功的POC之后，他用 Java REST和 SOAP Web服务做了些开发，并搭立了一个系统，在此系统中通过 Web服务利用 Hadoop存储和处理日志，然后在HBase中通过自定义表存储这些日志，通过 HBase API和 HBase-Hive映射查询来读取数据。Shashwat成功地实现了这个项目，紧接着开始了1TB到3TB的大规模二进制文件头的处理工作，他把文件的元数据存储在HBase中，文件本身存在HDFS上。

Shashwat在特里凡得琅

C-DAC网络取证中心开始他的软件开发生涯，为取证分析开发可移动相关软件。接着，他去了 Genilok Computer Solutions公司，在那里，他的工作包括：集群计算、HPC技术和 Web技术。

在此之后，他从特里凡得琅到了班加罗尔并加入了 PointCross，在那里他开始了大数据技术工作，用

Java开发软件、Web服务和大数据平台。在

PointCross，他的很多项目都是围绕着大数据技术，例如

Hadoop、HBase、Hive、Pig、Sqoop、Flume等。从这里他又到了 HCL

Infosystems公司，开始做

UIDAI项目，这是一个在印度非常有声望的项目，它为每一个印度居民提供一个唯一身份识别号。在这里，他工作中使用的技术有：HBase、Hive、Hadoop、Pig、

Linux、脚本语言、管理 HBase

Hadoop集群、编写脚本、自动化任务和处理、为集群监控创建仪表盘。

现在，Shashwat在 Cognilytics公司工作，专注于大数据技术、

HANA以及其他高性能技术。你可以通过 <https://github.com/shripav>和

<http://helpmetocode.blogspot.com>了解更多关于他的信息。可以通过

LinkedIn, <http://www.linkedin.com/pub/shashwat-shripav/19/214/2a9>

跟他联系，也可以发邮件给他，dwivedishashwat@gmail.com。

Shashwat曾经审校过 Pig Design Pattern, Pradeep Pasupuleti, Packt Publishing一书，他还曾担任过他大学杂志 InfinityTech的编辑。

目录: 第 1 章 了解 HBase 生态系统	1
基于 Hadoop 的 HBase 架构	2
RDBMS 和 HBase 的架构对比	3
HBase 的特征	3
HBase 在 Hadoop 生态系统中的位置	4
HBase 中的数据表示	5
Hadoop	6
HBase 与关系型数据库的功能对比	8
行存储数据库的逻辑展现	9
列存储数据库的逻辑展现	9
HBase 的内部存储架构	11
开始使用 HBase	11
HBase 是何时出现的	11
HBase 的组件和功能	14
ZooKeeper	14
谁在用 HBase? 为什么要用?	19
什么时候考虑使用 HBase?	20
什么时候不使用 HBase?	21
了解一些开源的 HBase 工具	21
Hadoop 和 HBase 版本兼容性对照表	22
HBase 的应用	23
HBase 的优点和缺点	24
小结	25
第 2 章 开启 HBase 之旅	26
深入理解 HBase 的组件	27
HFile	27
Region	27
可扩展性——理解纵向扩展和横向扩展的过程	29
读写周期	32
Write-Ahead Logs	33
MemStore	33
HBase 内部管理	34
合并	34
Region 分裂	35
Region 分配	36
Region 合并	37
RegionServer 故障转移	37
HBase 的删除请求	37
读写周期	37
可用的 HBase 发行版本列表	38
HBase 的必备条件和容量规划	39
DNS 正向解析	39
DNS 反向解析	40
SSH	41
小结	45
第 3 章 搭建 HBase	46
在 Ubuntu 上下载 Java	46
主机配置	52
基于主机文件	52
基于命令	52
基于文件	52
基于 DNS	54
安装和配置 SSH	54
在 Ubuntu/Red Hat/CentOS 上安装 SSH	55

- 配置 SSH 55
- 安装和配置 NTP 56
- 容量规划 57
- 安装和配置 Hadoop 58
 - core-site.xml 63
 - hdfs-site.xml 63
 - yarn-site.xml 65
 - mapred-site.xml 66
 - hadoop-env.sh 67
 - yarn-env.sh 67
- Hadoop 的启动步骤 67
- 配置 Apache HBase 69
 - 在单机模式中配置 HBase 69
 - 在分布式模式中配置 HBase 70
- 安装和配置 ZooKeeper 74
- 安装 Cloudera 版本的 Hadoop 和 HBase 76
- 下载 RPM 包 76
- 简易安装 Cloudera 77
- 安装 Hadoop 和 MapReduce 包 77
- 在 Windows 上安装 Hadoop 78
- 小结 81
- 第 4 章 优化 HBase/Hadoop 集群 82
 - Hadoop/HBase 集群的类型 82
 - CDH 集群的推荐配置 84
 - 容量规划 85
 - 优化 Hadoop 86
 - 通用优化技巧 86
 - 优化 Java GC 86
 - 优化 Linux 操作系统 87
 - 优化 Hadoop 参数 87
 - 优化 MapReduce 88
 - 优化 HBase 91
 - Hadoop 91
 - 内存 93
 - Java 93
 - 操作系统 94
 - HBase 94
 - 优化 ZooKeeper 96
 - Hadoop 中的重要配置文件 96
 - HBase 中的重要配置文件 97
 - 小结 98
- 第 5 章 HBase 的存储、框架以及数据类型 99
 - HBase 的数据类型 100
 - HBase 中的数据存储——逻辑视图 vs. 真实物理视图 101
 - 命名空间 102
 - HBase 服务 103
 - 行键 (Row key) 104
 - 列族 (Column family) 104
 - 列 (Column) 104
 - 单元格 (Cell) 104
 - 版本 (Version) 104
 - 时间戳 (Timestamp) 105
 - 数据模型的操作 105
 - 读 (Get) 105

- 写 (Put) 106
- 扫描 (Scan) 106
- 删除 (Delete) 106
- 版本和原因 107
- 决定版本数量 108
- 版本的下界 108
- 版本的上界 108
- 模式设计 109
- 表类型的设计 113
- 短宽和高瘦设计模式的好处 114
- 复合键设计 115
- 在 HBase 中计算存储的数据大小 118
- 小结 119
- 第 6 章 HBase 集群运维与故障处理 120
- Hadoop shell 命令 121
- Hadoop shell 命令的类型 121
- HBase shell 命令 140
- HBase 管理工具 149
- hbck —— HBase 检查 149
- HBase 健康检查脚本 151
- 写 HBase shell 脚本 151
- 使用 Hadoop 工具或者 JAR 151
- 用 Hive 连接 HBase 153
- HBase region 管理 155
- 压缩 155
- 合并 155
- HBase 节点管理 155
- 服役 155
- 退役 156
- 实现安全性 157
- 安全访问 157
- Kerberos KDC 157
- 客户端的安全配置 158
- 服务器端的安全配置 159
- 简单的安全 160
- 客户端配置 161
- 标签的安全特性 162
- HBase 的访问控制 163
- 使用标签的单元格访问 168
- 配置 ZooKeeper 安全 169
- HBase 常见错误的故障排查和相关说明 170
- 集群失败的可能情况 171
- 监控 HBase 的健康状况 172
- 小结 175
- 第 7 章 HBase 脚本编程 176
- HBase 中的备份与恢复技术 176
- 离线备份/full-shutdown 备份 177
- 在线备份 178
- Windows 上的 HBase 185
- 在 HBase 中进行脚本编程 185
- .irbrc 文件 187
- 获取时间戳 188
- 开启调试 189
- 在 HBase 中开启 SQL 189

- 参与 HBase 190
- 小结 190
- 第 8 章 HBase Java 编程 191
- 准备开发环境 192
- 构建 Java 客户端程序 192
- 数据类型 196
- 数据模型的 Java 操作 196
- 读操作 196
- 写操作 204
- 修改操作 206
- HBase 过滤器 208
- 过滤器类型 209
- 客户端 API 214
- 小结 215
- 第 9 章 HBase Java 高级编程 216
- 接口、类和异常 216
- 管理任务编程 218
- 数据操作代码 224
- MapReduce 和 HBase 226
- RESTful 和 Thrift 服务接口 231
- RESTful 服务接口 231
- Thrift 服务接口 232
- HDFS 编程 233
- 高级主题简介 237
- 协处理器 237
- 布隆过滤器 238
- Lily 项目 238
- 小结 239
- 第 10 章 HBase 使用案例 240
- HBase 在当今行业中的作用 240
- HBase 和关系型数据库的未来的对比 241
- 一些现实世界中的工程使用案例 241
- HBase 在 Facebook 241
- HBase 在 Pinterest 243
- HBase 在 Groupon 244
- HBase 在 LongTail Video 246
- HBase 在 Aadhaar (UIDAI) 247
- 有用的链接和参考 248
- 小结 249
- • • • • (收起)

[Learning Hbase中文版_下载链接1](#)

标签

hbase

计算机

分布式

数据库

hadoop

Tutorial

Hbase

Hadoop

评论

非常全面的HBase书籍，从操作到大概的原理读一遍都能知晓个大概。

快速入门指南

简单快速入门

【2017.10.09】 还可以。

[Learning Hbase中文版 下载链接1](#)

书评

[Learning Hbase中文版_下载链接1](#)