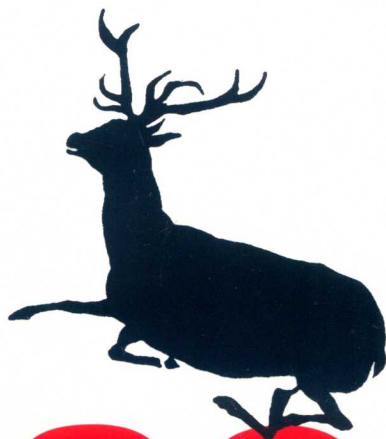


全面讲解 HBase 核心组件的
实现原理和基本操作
兼顾 HBase 应用与实战



HBase

入门与实践

—— 彭旭 著 ——

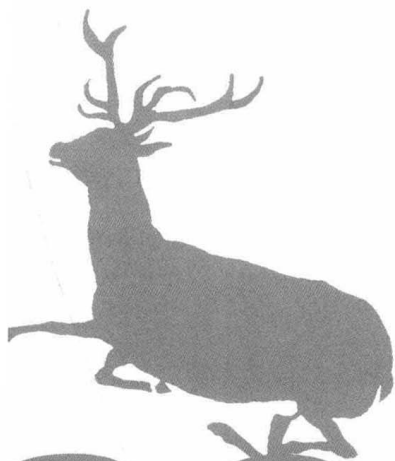
从零开始**基础入门**，深入浅出全面剖析 HBase
生产环境**调优进阶**，轻松掌握 HBase **实战知识**



中国工信出版集团



人民邮电出版社
POSTS & TELECOM PRESS



HBase

入门与实践

彭旭 著

人民邮电出版社
北京

图书在版编目 (C I P) 数据

HBase入门与实践 / 彭旭著. -- 北京 : 人民邮电出版社, 2018. 12
ISBN 978-7-115-49383-5

I. ①H… II. ①彭… III. ①计算机网络—信息存储
IV. ①TP393

中国版本图书馆CIP数据核字(2018)第214274号

内 容 提 要

本书以精练的语言介绍 HBase 的基础知识, 让初学者能够快速上手使用 HBase, 对 HBase 的核心思想(如数据读取、数据备份等)和 HBase 架构(如 LSM 树、WAL)有深入的分析, 并且让有经验的 HBase 开发人员也能够循序渐进地深入理解 HBase 源码, 以便更好地去调试和解决线上遇到的各种问题。本书更加专注于 HBase 在线实时系统的调优, 使 HBase 集群响应延迟更低。本书结合企业必备的“用户行为分析系统”, 让读者能够快速上手的同时, 也不乏企业 HBase 实际应用场景, 理论不脱离实际, 真正做到从入门到精通。

本书适合有一定 Java 基础的程序员作为 HBase 入门教程, HBase 运维人员可以将本书作为参考手册来部署和监控 HBase, 正在将 HBase 应用到在线生产环境中的软件开发人员也可以参考本书来调优 HBase 在线集群性能。

◆ 著 彭 旭

责任编辑 杨海玲

责任印制 焦志炜

◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路 11 号

邮编 100164 电子邮件 315@ptpress.com.cn

网址 <http://www.ptpress.com.cn>

三河市君旺印务有限公司印刷

◆ 开本: 800×1000 1/16

印张: 14

字数: 292 千字

印数: 1—3 000 册

2018 年 12 月第 1 版

2018 年 12 月河北第 1 次印刷

定价: 59.00 元

读者服务热线: (010)81055410 印装质量热线: (010)81055316

反盗版热线: (010)81055315

广告经营许可证: 京东工商广登字 20170147 号

对本书的赞誉

本书循序渐进地介绍了 HBase 从入门到企业实践调优，深入浅出地阐述了 HBase 架构与实现原理，以企业必备的用户行为系统为实践场景，用通俗易懂的语言带领读者走进 HBase 世界。作者在魅族云和大数据充分实践了书中架构，并实际带来了质量和效率的提升，同时也降低了云端运营成本。

——李柯辰，前魅族平台事业部总经理，现卓轩科技 CEO

彭旭是一个优秀的程序员、架构师。本书在介绍 HBase 基础知识的同时融入了他在魅族云服务团队将存储系统迁移到 HBase 的经验和教训，此外还在 HBase 源码和架构研究上有涉猎，对初学者、Java 相关开发人员和 HBase 运维人员等都是一本不错的参考书。

——何伟，前魅族 Flyme 技术委员会主席，现卓轩科技 CTO

作者将多年 HBase 的实践经验与心得体会沉淀为本书，既有经典的案例分析，也有抽丝剥茧的源码分析，对于大数据行业从业者，非常值得一读。

——张发恩，前百度云技术委员会主席，创新工场人工智能工程院首席架构师，创新奇智 CTO

作者是一个长期战斗在一线的典型程序员，本书所写的内容均贴近现实应用场景，用线上真实的案例来描述如何构建、优化 HBase 生成环境，事半功倍，不管是初学者还是入门者都可以从这些案例中吸取经验。

——唐进，前百度云产品委员会主席，现爱乐奇 CTO

与拖沓烦冗的英文著作相比，本书篇幅适中，但是以精简干练的语言完整地描述了 Hbase 从初学到熟练应用所需学习的所有 HBase 知识点。

——马杜，华云数据执行总裁

HBase 是一个历久弥坚的分布式列式存储系统，相关图书出版时间均距今已久。本书以当前稳定版本 HBase 为基础，着重介绍了 HBase 在线实时系统的调优等，很适合正在将 HBase 应用到在线生产环境的开发人员和运维人员阅读。2010 年下半年当当网在纽交所上市之前，我受当时 CTO 的委托加强当时的防盗刷/刷单系统，本质上是升级其中的高速计数器产品模块，该模块持久化部分的技术选型就是 HBase。这是当当网技术第一次在生产环境中应用 Hadoop 和 HBase。在这个过程中，最让人记忆深刻的不是技术难度，而是当时的资料很有限，HBase 的技术迭代又很快，网上大量文档在不同程度上有过时的问题，与其在网站上找过时的资料，

不如直接读代码。因此，本书全面介绍了当前稳定版本的 HBase，真是为 HBase 相关开发人员提供了很大的便利。

——傅强，前当当网技术副总裁，现九枝兰合伙人

HBase 是 Apache 下的项目，它是一个高可用、高性能、可伸缩和面向列的分布式存储系统，能实现对海量数据的高性能的读写。本书作者是前阿里天猫的工程师，有着丰富的开发经验，相信这本书一定能让读者的 HBase 理论与实践水平更上一层楼。

——杨开振，《深入浅出 Spring Boot 2.x》《深入浅出 MyBatis 技术原理与实战》
《Java EE 互联网轻量级框架整合开发：SSM 框架（Spring MVC+Spring+MyBatis）
和 Redis 实现》作者

一个有很大价值的开源项目既需要活跃的代码贡献者为其快速迭代铺路，也需要有优质的入门资料为广大初学者打下良好的基础。要学习 HBase，英文版的 *Apache HBase Book* 仍然是最佳的参考手册，而对英文不是特别好的读者来说，可以选择的资料就是《HBase 权威指南》或者是一些技术作者的专栏博客。本书对入门者来说是很好的中文读物。书中配有丰富的案例和插图，让读者可以较为轻松地理解 HBase 的常用场景和用法。希望本书对广大的 HBase 爱好者有所裨益。

——胡争，小米 HBase 工程师，HBase Committer

我几年前在维护淘宝的 HBase 集群时，时常和业务方交流如何基于 HBase 的架构来设计一个可靠和高性能的业务系统，深感对 HBase 内部原理的了解能极大改善业务系统的性能、稳定性以及成本。本书以当前 HBase 最新的稳定版本为基础，十分精准地抽离出了开发者需要重点关注的特点，并辅以几个典型的场景加以分析，堪称业务系统设计的手边助理。此外，作者拥有丰富的一线经验，因此本书具有很高的帮助开发者迅速上手生产系统的安装、部署以及 API 使用的实践价值。

——邓明鉴，前阿里巴巴高级专家，现 ZStack 首席架构师

前言

人工智能作为当前最热门的技术，其根本上离不开大数据的支持。如果把人工智能比喻成一个神经网络，那么数据则是在这个神经网络中用来传递信息的化学物质，没有信息传递的神经网络显然不名一文，因此大数据扮演着人工智能基石的角色。Hadoop 生态系统的 HDFS 和 MapReduce 分别为大数据提供了存储和分析处理能力，但是对在线实时的数据存取则无能为力，而 HBase 作为 Apache 顶级项目，弥补了 Hadoop 的这一缺陷，满足了在线实时系统低延时的需求。目前 HBase 在各大互联网公司几乎都有应用，前景广阔。

本书先介绍 HBase 基础知识，再带领读者深入研读 HBase 源码，从数据读取之 Scan 流程、HBase 架构（如 LSM 树和 WAL），到构建线上实时低延迟系统的调优，结合企业必备的“用户行为分析系统”，让读者能够快速上手 HBase 并了解 HBase 实现原理，同时让读者通过分析源码来了解 HBase 的设计思想，期望读者能够真正做到从入门到精通。

本书各章的主要内容概括如下。

- 第 1 章介绍 HBase 的诞生背景与 HBase 在国内各大公司的一些应用场景。
- 第 2 章介绍 HBase 的安装模式以及分布式部署情况下依赖的 Hadoop 和 ZooKeeper 的安装。
- 第 3 章介绍与 HBase 逻辑模型和物理模型相关的概念，如表、分区、行键、列族等，以及这些模型物理上的存储结构。
- 第 4 章介绍 HBase shell，在 shell 中使用数据定义语言与数据操纵语言，包括如何创建表、修改表、操作表的数据、使用 Scan 过滤所需要查询的数据，还介绍了一些特殊的命令，如查看集群间复制状态、管理分区等。
- 第 5 章介绍如何将 HBase 作为存储系统，包括行键设计原则、选用宽表或者高表，最后以朋友圈为例来进一步介绍如何在应用中合理地设计 HBase 表。
- 第 6 章介绍 HBase Java 客户端 API，包括在 Java API 中使用数据定义语言与数据操作语言、过滤器的使用、HBase 事务支持。
- 第 7 章介绍 HBase 架构设计，包括从 B 树开始到如何使用 LSM 来提升写效率、WAL 的作用、数据写入流程、数据读取流程。
- 第 8 章介绍两种类型的协处理器，即观察者类型协处理器和端点类型协处理器，以及如何装载使用。
- 第 9 章介绍 HBase 在线调优，包括客户端层面如何设置查询缓存、跳过 WAL 写入，服

务端层面建表语句优化、开启机柜感知、本地读、补偿重试读以及分区服务器 JVM 内存调优等。

- 第 10 章介绍数据备份与复制的几种方式，包括复制、快照、导入、导出和复制表。
- 第 11 章介绍 HBase 自带的监控系统，着重介绍了 HBase JMX 监控。

作为一个典型的内敛型程序员，我从来没有过要写书、当作者的念头，也一直觉得作家是一个很神秘的职业。由于个人性格问题，我对任何事情的描述都是直来直往，描述问题经常是直奔主题，很少会去解释背景与前因后果。机缘巧合下我遇到了几个做教育以及写书的朋友，一番纠结后兴起了写书的念头，正好最近一直在研究 HBase，加上已经到了三十而立的年龄，宝宝也将要出生，觉得自己也需要做点什么，一方面给自己留点回忆，另一方面等宝宝长大后也算是给她的一份礼物。开始写的时候进度很慢，感觉无法坚持下去，但是慢慢地写着写着竟然越来越精神，越来越投入，愿本书能够帮到你！

感谢我的父母对我个人和家庭无私的付出与无微不至的照顾，感谢我的妻子黄晶对我的包容与支持！感谢珠海市魅族科技有限公司、珠海市卓轩科技有限公司提供的平台与实践经历促成了本书！感谢李柯辰先生与何伟先生对我和本书的鼎力支持与帮助！感谢我生命中遇到的每一个人，愿大家越来越好！祝我的宝贝彭语桐开心快乐成长！

由于作者水平有限，在编写过程中，难免出现错误或者不准确的地方，但是凡是涉及实例代码之处，我都实际运行过程序，从而保证程序的准确性。如果读者在阅读过程中发现有错误之处，敬请指正，联系方式如下：

微博 17051158029

微信 17051158029

邮箱 17051158029@163.com

彭旭

2018 年 10 月

资源与支持

本书由异步社区出品，社区（<https://www.epubit.com/>）为您提供相关资源和后续服务。

配套资源

本书提供如下资源：

- 本书源代码；
- 书中彩图文件。

要获得以上配套资源，请在异步社区本书页面中单击 **配套资源**，跳转到下载界面，按提示进行操作即可。注意：为保证购书读者的权益，该操作会给出相关提示，要求输入提取码进行验证。

如果您是教师，希望获得教学配套资源，请在社区本书页面中直接联系本书的责任编辑。

提交勘误

作者和编辑尽最大努力来确保书中内容的准确性，但难免会存在疏漏。欢迎您将发现的问题反馈给我们，帮助我们提升图书的质量。

当您发现错误时，请登录异步社区，按书名搜索，进入本书页面，单击“提交勘误”，输入勘误信息，单击“提交”按钮即可。本书的作者和编辑会对您提交的勘误进行审核，确认并接受后，您将获赠异步社区的 100 积分。积分可用于在异步社区兑换优惠券、样书或奖品。

详细信息

写书评

提交勘误

页码: 页内位置 (行数): 勘误次数:

B I U

字数统计

提交

扫码关注本书

扫描下方二维码，您将会在异步社区微信服务号中看到本书信息及相关的服务提示。



与我们联系

我们的联系邮箱是 contact@epubit.com.cn。

如果您对本书有任何疑问或建议，请您发邮件给我们，并请在邮件标题中注明本书书名，以便我们更高效地做出反馈。

如果您有兴趣出版图书、录制教学视频，或者参与图书翻译、技术审校等工作，可以发邮件给我们；有意出版图书的作者也可以到异步社区在线提交投稿（直接访问 www.epubit.com/selfpublish/submission 即可）。

如果您是学校、培训机构或企业，想批量购买本书或异步社区出版的其他图书，也可以发邮件给我们。

如果您在网上发现有针对异步社区出品图书的各种形式的盗版行为，包括对图书全部或部分内容的非授权传播，请您将怀疑有侵权行为的链接发邮件给我们。您的这一举动是对作者权益的保护，也是我们持续为您提供有价值的内容的动力之源。

关于异步社区和异步图书

“异步社区”是人民邮电出版社旗下 IT 专业图书社区，致力于出版精品 IT 技术图书和相关学习产品，为作译者提供优质出版服务。异步社区创办于 2015 年 8 月，提供大量精品 IT 技术图书和电子书，以及高品质技术文章和视频课程。更多详情请访问异步社区官网 <https://www.epubit.com>。

“异步图书”是由异步社区编辑团队策划出版的精品 IT 专业图书的品牌，依托于人民邮电出版社近 30 年的计算机图书出版积累和专业编辑团队，相关图书在封面上印有异步图书的 LOGO。异步图书的出版领域包括软件开发、大数据、AI、测试、前端、网络技术等。



异步社区



微信服务号

目 录

第 1 章 HBase 简介..... 1	第 4 章 HBase shell..... 39
1.1 背景..... 1	4.1 数据定义语言..... 39
1.2 NoSQL 与传统 RDBMS..... 2	4.1.1 创建表..... 39
1.3 应用场景..... 3	4.1.2 查看所有表..... 40
1.3.1 Facebook 用户交互数据..... 3	4.1.3 查看建表..... 40
1.3.2 淘宝 TLog 等..... 3	4.1.4 修改表..... 41
1.3.3 小米云服务..... 4	4.2 数据操纵语言..... 41
1.3.4 用户行为数据存储..... 4	4.2.1 Put..... 41
第 2 章 HBase 安装..... 5	4.2.2 Get..... 42
2.1 单机部署..... 5	4.2.3 Scan..... 43
2.1.1 前置条件..... 5	4.2.4 删除数据..... 45
2.1.2 下载 HBase..... 6	4.3 其他常用 shell..... 46
2.1.3 配置 HBase..... 6	4.3.1 复制状态查看..... 46
2.1.4 启动 HBase..... 6	4.3.2 分区拆分..... 47
2.1.5 HBase 初体验..... 7	4.3.3 分区主压缩..... 47
2.2 分布式部署..... 8	4.3.4 负载均衡开关..... 48
2.2.1 环境准备..... 9	4.3.5 分区手动迁移..... 48
2.2.2 ZooKeeper 安装..... 11	第 5 章 模式设计..... 49
2.2.3 Hadoop 安装..... 12	5.1 行键设计..... 50
2.2.4 HBase 安装..... 23	5.2 规避热点区间..... 52
2.2.5 启动集群..... 28	5.3 高表与宽表..... 54
2.3 集群增删节点..... 29	5.4 微信朋友圈设计..... 55
2.3.1 增加节点..... 29	5.4.1 需求定义..... 55
2.3.2 删除节点..... 30	5.4.2 问题建模..... 55
第 3 章 HBase 数据模型..... 33	第 6 章 客户端 API..... 61
3.1 逻辑模型..... 33	6.1 Java 客户端使用..... 61
3.2 物理模型..... 35	6.2 数据定义语言..... 64

6.2.1 表管理	64	9.1.3 跳过 WAL 写入	143
6.2.2 分区管理	66	9.1.4 设置重试次数与间隔	144
6.3 数据操纵语言	68	9.1.5 选用合适的过滤器	144
6.3.1 Put	68	9.2 服务端调优	145
6.3.2 Get	70	9.2.1 建表 DDL 优化	145
6.3.3 Scan	72	9.2.2 禁止分区自动拆分与压缩	150
6.3.4 Delete	74	9.2.3 开启机柜感知	151
6.3.5 Increment	76	9.2.4 开启 Short Circuit Local Reads	153
6.4 过滤器	78	9.2.5 开启补偿重试读	154
6.4.1 过滤器简介	78	9.2.6 JVM 内存调优	155
6.4.2 过滤器使用	81	第 10 章 集群间数据复制	163
6.5 事务	94	10.1 复制	164
6.5.1 原子性	95	10.1.1 集群拓扑	165
6.5.2 隔离性	95	10.1.2 配置集群复制	166
第 7 章 架构实现	101	10.1.3 验证复制数据	169
7.1 存储	101	10.1.4 复制详解	171
7.1.1 B+树	101	10.2 快照	175
7.1.2 LSM 树	102	10.2.1 配置快照	176
7.1.3 WAL	104	10.2.2 管理快照	176
7.2 数据写入读取	107	10.3 导出和导入	178
7.2.1 定位分区服务器	107	10.3.1 导出	178
7.2.2 数据修改流程	108	10.3.2 导入	180
7.2.3 数据查询流程	113	10.4 复制表	180
第 8 章 协处理器	115	第 11 章 监控	183
8.1 观察者类型协处理器	115	11.1 Hadoop 监控	183
8.2 端点类型协处理器	132	11.1.1 Web 监控页面	184
8.3 装载/卸载协处理器	136	11.1.2 JMX 监控	185
8.3.1 静态装载/卸载	136	11.2 HBase 监控	188
8.3.2 动态装载/卸载	137	11.2.1 Web 监控页面	188
第 9 章 HBase 性能调优	141	11.2.2 JMX 监控	190
9.1 客户端调优	141	11.2.3 API 监控	192
9.1.1 设置客户端写入缓存	141	后记	195
9.1.2 设置合适的扫描缓存	143	附录 常见问题	197

第 1 章

HBase 简介

HBase 至今已有十多年的历史，其分布式弹性伸缩的特性使其在今天的大数据存储系统中占有重要的一席之地。

1.1 背景

随着计算机科技技术的发展，电子设备越来越普及，价格也越来越便宜，各种电子设备（如手机、智能家居、网络摄像头、智能汽车等）产生的数据量正在急剧增长，全球数据量的迅猛增长为大数据行业的发展提供了基础保证。市场调研机构预计未来全球每年数据量的增长率将达到 40% 以上，而且数据增长速度越来越快，对一些大公司来说，每天甚至每小时都可以产生 TB、PB 级的数据，到 2020 年，全球数据总量预计将超过 50 ZB。

这些数据可能很难存储或者分析，但是很多时候这些数据很有可能成为企业的第一生产力。例如，阿里巴巴的用户行为数据，对于推荐或者在线广告至关重要，如果你在淘宝搜索了一款手机，稍后当你浏览新闻的时候，新闻客户端就会马上给你推荐一些手机的广告，这样有针对性的广告也就是常说的千人千面，极大地提高了广告的转化率。

在分布式数据库系统发展之前，因为缺乏有效或者低成本的方式来存储大量的数据，很多企业会忽略某些数据源，这样导致当真正需要使用某些数据的时候已经找不到可以用来筛选的源数据了，或者即使使用低成本的方式（如廉价磁带或者磁盘）将数据存储下来了，也因为缺乏有效的数据筛选等处理手段而导致数据无法发挥其应有的价值。

传统系统可能很难存储、分析现在各种类型的非结构化数据（如办公文档、文本、音频、图片、用户行为等字段长度可变，每个字段又可以由可重复或者不可重复的子字段组成）与半结构化数据，这些数据存储到关系型数据库用于分析会花费更多的时间和金钱。

例如，如果使用 MySQL 存储淘宝的用户行为数据，每天产生的数据量有上百亿甚至更多字节，以 MySQL 每个表存储 4000 万行数据计算，这样每天产生的数据需要 MySQL 分表 100 个以上，可以想象这对机器、运维以及开发是多高的成本。

Google 内部使用的 GFS 以及 MapReduce 技术正好解决了大量数据存储以及分析的问题，而开源社区基于 Google 描述的基于商业硬件集群构建分布式、可扩展的存储和处理系统思想，实现了开源项目 Hadoop 的两个模块，即 HDFS 和 MapReduce。

HDFS 非常适合存储任意的非结构化或者半结构化的数据，它的分布式架构以及简单的扩展方式，使其可以对数据存储提供“无限”的支持，同时因为其可以部署在普通的商业机器集群上，使得存储成本更为低廉，而 MapReduce 则可以帮助用户在需要的时候利用集群中每台机器的能力“分而治之”，恰当地分析这些数据，提供了处理海量数据的核心能力。

GFS 以及 MapReduce 虽然提供了大量数据的存储和数据的分析处理能力，但是对于实时数据的随机存取却无能为力，而且 GFS 适合存储少量的大文件而不适合存储大量的小文件，因为大量的小文件最终会导致元数据的膨胀而可能无法放入主节点的内存。经过多年耕耘，Google 终于在 2006 年发表了一篇“Bigtable: A Distributed Storage System for Structured Data”的论文，这篇论文就是 HBase 的起源，HBase 实现了 BigTable 的架构，如压缩算法、内存操作和布隆过滤器等，HBase 弥补了 Hadoop 的上述缺陷，使得 Hadoop 生态系统更加完善。目前 HBase 在各大公司基本都有使用，如 Facebook 的消息平台、小米的云服务、阿里的 TLog 等许多服务组件。

1.2 NoSQL 与传统 RDBMS

NoSQL，是 Not only SQL 的缩写，泛指非关系型的数据库，如 HBase、MongoDB、Cassandra 等。与传统的关系型数据库（RDBMS）相比，NoSQL 最大的特征就是数据存储不需要一个特定的模式并具有强大的水平扩展能力。

随着大数据时代的到来，特别是超大规模与高并发的社交类型网站的诞生，传统的关系型数据库已经无法满足这些应用系统架构在横向扩展方面的需求了。

- 并行性：关系型数据库更依赖单机性能，对于上万的 QPS，硬盘 IO 无法支撑，并且关系型数据库除了写数据外还要写索引，而 NoSQL，以 HBase 为例，写入为顺序写入，如果能够容忍部分数据的丢失不写 WAL，HBase 写入只需写内存，速度大大提高。同时 HBase 通过分片可以将一台机器的压力均衡地转移到集群的每一台机器。
- 可扩展性：现在一般的在线系统可用性至少都要求达到 99.9%，对于淘宝天猫这种与金钱密切相关的系统，可用性要求就更高了。传统关系型数据库的升级和扩展

对于系统的可用性是一个很头疼的问题，可能需要停机迁移数据，重启加载新配置等而且需要运维人员、开发人员、DBA 的密切配合，而 HBase 集群具有线性伸缩、自动容灾和负载均衡的优势，可以很容易地增加或者替换集群节点以扩展集群的存储和计算能力。

- 数据模型：关系型数据库需要为存储的数据预先定义表结构与字段名，而 NoSQL 无须事先为需要存储的数据定义一个模式，这样可以更容易、更灵活去适配各种类型的非结构化数据。

当然，并不是说 NoSQL 已经全面优于关系型数据库了，NoSQL 相比关系型数据库也有很多缺点，例如 HBase 不支持多行事务；基于 LSM 存储模型导致需要读取多个文件来找到需要的数据，这样会牺牲一些读的性能。NoSQL 可以说是对关系型数据库的一种补充。

1.3 应用场景

当互联网概念火热的时候，很多大公司都吹嘘自己是一家互联网公司，现在，很多公司都宣称自己是一家大数据公司。确实现在很多公司都注重数据的采集，而这些数据所使用的存储系统，据我所知，国内各大互联网公司以及知名企业内部基本都使用 HBase。使用 HBase 的业务也非常多，如订单、搜索、用户画像、推荐、监控、安全风控等。HBase 集群在 Facebook 已经达到上万节点的规模，下面列举几个应用场景。

1.3.1 Facebook 用户交互数据

这是一个典型的例子，被引用的次数不计其数，Facebook 的 Like 按钮被点了多少次、有多少人浏览过某个文章、有多少人喜欢这篇文章等数据是由 HBase 的计数器来存储的，发布者能够实时地看到多少人给他点赞、多少人喜欢他的文章。

HBase 在 Facebook 的应用非常成功，也输出了一些著名的技术分享，如 Apache Hadoop Goes Realtime at Facebook 和 Facebook's New Real-Time Messaging System: HBase To Store 135+ Billion Messages A Month。

1.3.2 淘宝 TLog 等

淘宝 TLog 是一个分布式的、可靠的、对大量数据进行收集、分析和展现的系统。主要应用场景是收集大量的运行时日志，然后分析存储，最后提供数据查询和展现。淘宝赫赫有名的鹰眼系统（对请求从开始到结束整个生命周期的追踪，包括哪一步到了哪台机器、每一步花了多长时间、与多少系统有交互等）就是 TLog 的接入方，每天有上万台机器接

入 TLog，数据量多达上百 TB，其底层就是使用 HBase 作为存储层。

TLog 的设计也参照了业界著名的 openTSDB 系统——“OpenTSDB is a distributed, scalable Time Series Database (TSDB) written on top of HBase”。感兴趣的读者可以去搜索 OpenTSDB 学习一下。

1.3.3 小米云服务

小米的云服务基本是基于 HBase 存储的。这里以云端数据的同步备份功能为例，用户的照片、联系人、短信、通话记录、米聊等数据中的大部分是非结构化的，重点是数据以用户为隔离，用户只需要访问与自己相关的数据，非常契合 HBase 的分片自动负载均衡，只需使用用户 ID 来进行分区（region，或者说分片），用户量或者数据量的持续增长就可以非常容易地通过为 HBase 集群添加节点来解决。

小米对 HBase 也做了很多贡献，自研的 HBase 自动化部署与监控系统 Minos 已经开源，读者可以去 GitHub 搜索项目 Minos 了解详情。

1.3.4 用户行为数据存储

现在大部分公司都非常注重用户数据的收集（基本上手机上使用的每个 App 都有各种埋点，每一次浏览、点击事件等都会被上传到服务端存储），尤其是用户行为数据的收集，这些数据价值很高，可以用来做很多事情，例如用户画像，既可以为用户构建用户基本信息、行为特征、社交、购买力等静态标签，也可以建立短期的动态标签，例如假设用户刚买车，可能就需要购买一些相关的车载用具。

通过给用户构建一个立体的画像，系统可以近乎实时地分析用户的行为，了解用户的需求，从而实现精准化营销，这对企业尤其是电商企业至关重要。本书我们将会以电商系统的用户行为数据收集作为实战案例来一步一步学习 HBase 相关的知识。

第 2 章

HBase 安装

在生产环境中 HBase 通常与 Hadoop 一起部署分布式集群，为了方便入门体验或者测试，HBase 也提供了单机版的部署包，本章将分别介绍 HBase 的单机版部署与分布式部署。

2.1 单机部署

单机部署可以用来体验或者测试，在 Windows、Linux 以及 Mac 操作系统下都提供了启动脚本，均可一键启动 HBase。

2.1.1 前置条件

单机版的 HBase 只需一个 Java 的运行环境即可启动，表 2-1 描述了各版本 HBase 对 JDK 版本的要求。

表 2-1 Hbase 与 JDK 版本兼容关系

HBase 版本	JDK7.0	JDK8.0
HBase 2.0	不支持	支持
HBase 1.3	支持	支持
HBase 1.2	支持	支持
HBase 1.1	支持	支持但没经过严格测试

2.1.2 下载 HBase

打开 Apache 官网 HBase 下载页面 <http://www.apache.org/dyn/closer.cgi/hbase/>。

选中一个下载镜像, 进入 HBase 版本选择页面, 因为企业生产环境一般都使用当前稳定版的 HBase, 所以单击 `stable` 文件夹选择一个稳定版本的 HBase 下载, 当前稳定版是 HBase 1.2.6。注意选择 `.tar.gz` 结尾的下载包, `src.tar.gz` 结尾的下载包是 HBase 的源码。

下载完成后将文件解压, 假设我们使用 Windows 系统, 将文件解压到 D 盘的 `hbase-1.2.6` 目录, 此时我们可以定义一个环境变量 `HBASE_HOME = D:\hbase-1.2.6`。

2.1.3 配置 HBase

单机版 HBase 需要配置的唯一文件是 `%HBASE_HOME%/hbase-1.2.6/conf/hbase-site.xml`, 需要配置的内容也很少, 只需要指定 HBase 和 ZooKeeper 写入数据的路径即可。默认的情况下 HBase 会自动创建一个 `/tmp` 的目录, 然后将数据写入这个目录, 但是很多系统在重启的时候会删除 `/tmp` 目录, 因此通常需要在配置文件中指定数据写入的目录, 如代码清单 2-1 所示。

代码清单 2-1 hbase-site.xml

```
<configuration>
  <property>
    <name>hbase.rootdir</name>
    <value>file:///home/mt/hbase</value>
  </property>
  <property>
    <name>hbase.zookeeper.property.dataDir</name>
    <value>/home/mt/zookeeper</value>
  </property>
</configuration>
```

- `hbase.rootdir`: 表示 HBase 数据存储的目录, 这里 `file:/` 前缀表示目录是一个本地文件系统, 生产环境中一般将 HBase 文件存储构建在 Hadoop 文件系统 (HDFS) 上, 那么该配置项的值类似于 `hdfs://mt.com:8020/hbase`。
- `hbase.zookeeper.property.dataDir`: 表示 ZooKeeper 数据存储的目录。

2.1.4 启动 HBase

`%HBASE_HOME%/hbase-1.2.6/bin/start-hbase.cmd` 提供了一个一键启动 HBase 的脚本,