

## <<HBase实战>>

### 图书基本信息

书名：<<HBase实战>>

13位ISBN编号：9787115324467

10位ISBN编号：7115324468

出版时间：2013-9-1

出版时间：人民邮电出版社

作者：Nick Dimiduk,Amandeep Khurana

译者：谢磊

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实战>>

### 内容概要

《HBase实战》是一本基于经验提炼而成的指南，它教给读者如何运用HBase设计、搭建及运行大数据应用系统。

全书共分为4个部分。

前两个部分分别介绍了分布式系统和大规模数据处理的发展历史，讲解HBase的基本原理模式设计以及如何使用HBase的高级特性；第三部分通过真实的应用和代码示例以及支持这些实践技巧的理论知识，进一步探索HBase的一些实用技术；第四部分讲解如何把原型开发系统升级为羽翼丰满的生产系统。

《HBase实战》适合所有对云计算、大数据处理技术和NoSQL数据库感兴趣的技术人员阅读，尤其适合对Hadoop及HBase感兴趣的技术人员参考。

阅读《HBase实战》不要求之前具备HBase、Hadoop或者MapReduce方面的知识。

## <<HBase实战>>

### 作者简介

1. 本书两位作者Nick Dimiduk和Amandeep Khurana都是长期的HBase实践者，非常有真材实料。
2. 本书译者谢磊是ChinaHadoop技术社区发起人，资深IT人士，从事软件行业十多年，现在从事企业级大数据解决方案的工作，致力于大数据技术的普及和推广。

## <<HBase实战>>

### 书籍目录

#### 第一部分 HBase基础

##### 第1章 HBase介绍

###### 1.1 数据管理系统：速成

###### 1.1.1 你好，大数据

###### 1.1.2 数据创新

###### 1.1.3 HBase的崛起

###### 1.2 HBase使用场景和成功案例

###### 1.2.1 典型互联网搜索问题：BigTable发明的原因

###### 1.2.2 抓取增量数据

###### 1.2.3 内容服务

###### 1.2.4 信息交换

###### 1.3 你好HBase

###### 1.3.1 快速安装

###### 1.3.2 HBase Shell命令行交互

###### 1.3.3 存储数据

###### 1.4 小结

##### 第2章 入门

###### 2.1 从头开始

###### 2.1.1 创建表

###### 2.1.2 检查表模式

###### 2.1.3 建立连接

###### 2.1.4 连接管理

###### 2.2 数据操作

###### 2.2.1 存储数据

###### 2.2.2 修改数据

###### 2.2.3 工作机制：HBase写路径

###### 2.2.4 读数据

###### 2.2.5 工作机制：HBase读路径

###### 2.2.6 删除数据

###### 2.2.7 合并：HBase的后台工作

###### 2.2.8 有时间版本的数据

###### 2.2.9 数据模型概括

###### 2.3 数据坐标

###### 2.4 小结

###### 2.5 数据模型

###### 2.5.1 逻辑模型：有序映射的映射集合

###### 2.5.2 物理模型：面向列族

###### 2.6 表扫描

###### 2.6.1 设计用于扫描的表

###### 2.6.2 执行扫描

###### 2.6.3 扫描器缓存

###### 2.6.4 使用过滤器

###### 2.7 原子操作

###### 2.8 ACID语义

###### 2.9 小结

## &lt;&lt;HBase实战&gt;&gt;

## 第3章 分布式的HBase、HDFS和MapReduce

## 3.1 一个MapReduce的例子

## 3.1.1 延迟与吞吐量

## 3.1.2 串行计算吞吐量有限

## 3.1.3 并行计算提高吞吐量

## 3.1.4 MapReduce：用布式计算最大化吞吐量

## 3.2 Hadoop MapReduce概览

## 3.2.1 MapReduce数据流介绍

## 3.2.2 MapReduce内部机制

## 3.3 分布式模式的HBase

## 3.3.1 切分和分配大表

## 3.3.2 如何找到region

## 3.3.3 如何找到-ROOT-表

## 3.4 HBase和MapReduce

## 3.4.1 使用HBase作为数据源

## 3.4.2 使用HBase接收数据

## 3.4.3 使用HBase共享资源

## 3.5 信息汇总

## 3.5.1 编写MapReduce应用

## 3.5.2 运行MapReduce应用

## 3.6 大规模条件下的可用性和可靠性

## 3.6.1 HDFS作为底层存储

## 3.7 小结

## 第二部分 高级概念

## 第4章 HBase表设计

## 4.1 如何开始模式设计

## 4.1.1 问题建模

## 4.1.2 需求定义：提前多做准备工作总是有好处的

## 4.1.3 均衡分布数据和负载的建模方法

## 4.1.4 目标数据访问

## 4.2 反规范化是HBase世界里的词语

## 4.3 相同表里的混杂数据

## 4.4 行键设计策略

## 4.5 IO考虑

## 4.5.1 为写优化

## 4.5.2 为读优化

## 4.5.3 基数和行键结构

## 4.6 从关系型到非关系型

## 4.6.1 一些基本概念

## 4.6.2 嵌套实体

## 4.6.3 没有映射到的一些东西

## 4.7 列族高级配置

## 4.7.1 可配置的数据块大小

## 4.7.2 数据块缓存

## 4.7.3 激进缓存

## 4.7.4 布隆过滤器

## 4.7.5 生存时间 (TTL)

## &lt;&lt;HBase实战&gt;&gt;

## 4.7.6 压缩

## 4.7.7 单元时间版本

## 4.8 过滤数据

## 4.8.1 实现一个过滤器

## 4.8.2 预装过滤器

## 4.9 小结

## 第5章 使用协处理器扩展HBase

## 5.1 两种协处理器

## 5.1.1 Observer协处理器

## 5.1.2 endpoint协处理器

## 5.2 实现一个observer

## 5.2.1 修改模式

## 5.2.2 从HBase开始

## 5.2.3 安装observer

## 5.2.4 其他安装选项

## 5.3 实现一个endpoint

## 5.3.1 为endpoint定义接口

## 5.3.2 实现endpoint服务器

## 5.3.3 实现endpoint客户端

## 5.3.4 部署endpoint服务器

## 5.3.5 试运行

## 5.4 小结

## 第6章 其他的HBase客户端选择

## 6.1 在UNIX里使用HBase Shell脚本

## 6.1.1 准备HBase Shell

## 6.1.2 使用UNIX Shell脚本创建表模式

## 6.2 使用JRuby进行HBase Shell编程

## 6.2.1 准备HBase Shell

## 6.2.2 访问TwitBase的users表

## 6.3 通过REST访问HBase

## 6.3.1 启动HBase REST服务

## 6.3.2 访问TwitBase的users表

## 6.4 通过Python使用HBase Thrift网关

## 6.4.1 生成Python语言的HBase Thrift客户端库

## 6.4.2 启动HBase Thrift服务

## 6.4.3 扫描TwitBaseuser表

## 6.5 asynchbase：另外一种HBase Java客户端

## 6.5.1 创建一个asynchbase项目

## 6.5.2 改变TwitBase的密码策略

## 6.5.3 试运行

## 6.6 小结

## 第三部分 应用系统实例

## 第7章 通过实例学习HBase：OpenTSDB

## 7.1 OpenTSDB概述

## 7.1.1 挑战：基础设施监控

## 7.1.2 数据：时间序列

## 7.1.3 存储：HBase

## &lt;&lt;HBase实战&gt;&gt;

## 7.2 设计一个HBase应用系统

## 7.2.1 模式设计

## 7.2.2 应用架构

## 7.3 实现一个HBase应用系统

## 7.3.1 存储数据

## 7.3.2 查询数据

## 7.4 小结

## 第8章 在HBase上查询地理信息系统

## 8.1 运用地理数据

## 8.2 设计一个空间索引

## 8.2.1 从复合行键开始

## 8.2.2 介绍geohash

## 8.2.3 理解geohash

## 8.2.4 在有空间感知特性的行键里使用geohash

## 8.3 实现最近邻居查询

## 8.4 把计算工作推往服务器端

## 8.4.1 基于查询多边形创建一次geohash扫描

## 8.4.2 区域内查询第一幕：客户端

## 8.4.3 区域内查询第二幕：WithinFilter

## 8.5 小结

## 第四部分 让HBase运转起来

## 第9章 部署HBase

## 9.1 规划集群

## 9.1.1 原型集群

## 9.1.2 小型生产集群（10~20台服务器）

## 9.1.3 中型生产集群（50台以下服务器）

## 9.1.4 大型生产集群（超过50台服务器）

## 9.1.5 Hadoop Master节点

## 9.1.6 HBase Master

## 9.1.7 Hadoop DataNode和HBase RegionServer

## 9.1.8 ZooKeeper

## 9.1.9 采用云服务怎么样？

## 9.2 部署软件

## 9.2.1 Whirr：在云端部署

## 9.3 发行版本

## 9.3.1 使用原生Apache发行版本

## 9.3.2 使用Cloudera的CDH发行版本

## 9.4 配置

## 9.4.1 HBase配置

## 9.4.2 与HBase有关的Hadoop配置参数

## 9.4.3 操作系统配置

## 9.5 管理守护进程

## 9.6 小结

## 第10章 运维

## 10.1 监控你的集群

## 10.1.1 HBase如何输出监控指标

## <<HBase实战>>

- 10.1.2 收集监控指标和图形展示
- 10.1.3 HBase输出的监控指标
- 10.1.4 应用端监控
- 10.2 HBase集群的性能
  - 10.2.1 性能测试
  - 10.2.2 什么影响了HBase的性能
  - 10.2.3 优化支撑系统
  - 10.2.4 优化HBase
- 10.3 集群管理
  - 10.3.1 启动和停止HBase
  - 10.3.2 优雅停止和让节点退役
  - 10.3.3 增加节点
  - 10.3.4 滚动重启和升级
  - 10.3.5 bin/hbase和HbaseShell
  - 10.3.6 维护一致性——hbck
  - 10.3.7 查看HFile和HLog
  - 10.3.8 预先拆分表
- 10.4 备份和复制
  - 10.4.1 集群间复制
  - 10.4.2 使用MapReduce作业进行备份
  - 10.4.3 备份根目录
- 10.5 小结
- 附录A 探索HBase系统
- 附录B 更多关于HDFS的工作原理

## <<HBase实战>>

### 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>