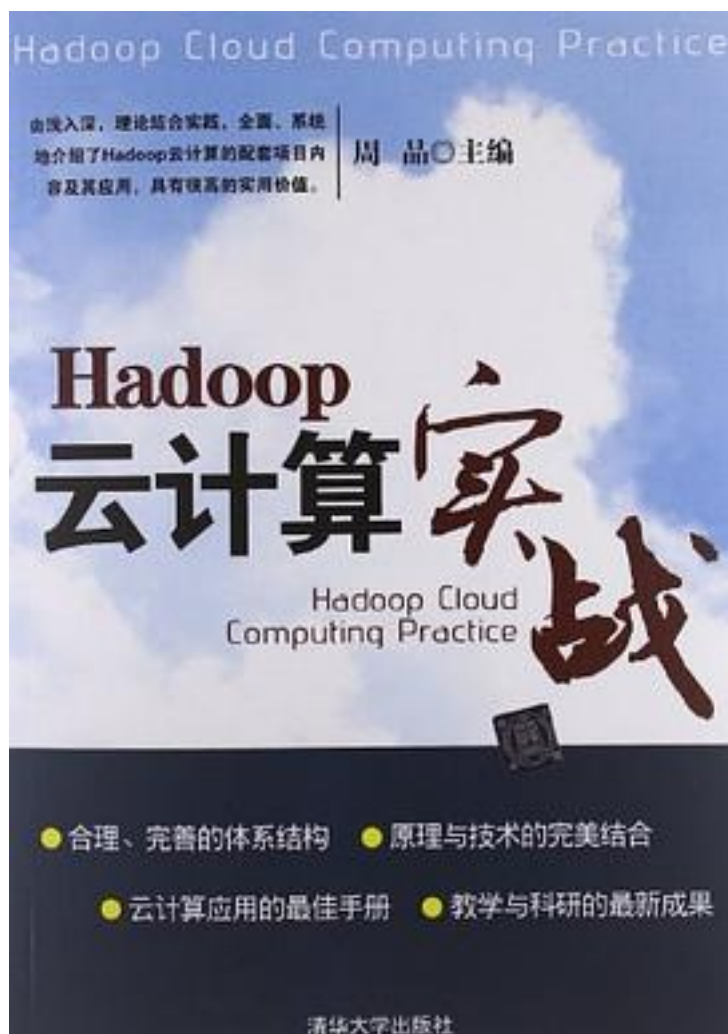


Hadoop云计算实战



[Hadoop云计算实战_下载链接1](#)

著者:周 品

出版者:清华大学出版社

出版时间:2012-10

装帧:平装

isbn:9787302296737

《Hadoop云计算实战》全面介绍了云计算的基本概念、Google（谷歌）云计算的关键

技术，以及Hadoop云计算的相关配套项目及其实战，包括Hadoop的HDFS、MapReduce、HBase、Hive、Pig、Cassandra、Chukwa及ZooKeeper等配套项目的实现机制、用法及应用。

作者介绍:

目录: 第1章 云计算概论 1

1.1 云计算概述 1

1.1.1 云计算的定义 1

1.1.2 云计算产生的背景 2

1.1.3 云时代谁是主角 3

1.1.4 云计算的特征 4

1.1.5 云计算的发展史 5

1.1.6 云计算的服务层次 7

1.1.7 云计算的服务形式 7

1.1.8 云计算的实现机制 9

1.1.9 云计算研究方向 11

1.1.10 云计算发展趋势 12

1.2 云计算关键技术研究 14

1.2.1 虚拟化技术 14

1.2.2 数据存储技术 15

1.2.3 资源管理技术 17

1.2.4 能耗管理技术 18

1.2.5 云监测技术 19

1.3 云计算应用研究 22

1.3.1 语义分析应用 22

1.3.2 IT企业应用 22

1.3.3 生物学应用 23

1.3.4 电信企业应用 24

1.3.5 数据库的应用 27

1.3.6 地理信息应用 28

1.3.7 医学应用 29

1.4 云安全 30

1.4.1 云安全发展趋势 31

1.4.2 云安全与网络安全的差别 31

1.4.3 云安全研究的方向 31

1.4.4 云安全难点问题 32

1.4.5 云安全新增及增强功能 32

1.5 云计算生命周期 33

1.6 云计算存在的问题 34

1.7 云计算的优缺点 35

第2章 Hadoop相关项目介绍 37

2.1 Hadoop简介 37

2.1.1 Hadoop的基本架构 37

2.1.2 Hadoop文件系统结构 40

2.1.3 Hadoop文件读操作 41

2.1.4 Hadoop文件写操作 42

2.2 Hadoop系统性质 42

2.2.1 可靠存储性 43

2.2.2 数据均衡 43

2.3 比较SQL数据库与Hadoop 44

2.4 MapReduce概述 45

2.4.1 MapReduce实现机制	45
2.4.2 MapReduce执行流程	46
2.4.3 MapReduce映射和化简	47
2.4.4 MapReduce输入格式	47
2.4.5 MapReduce输出格式	48
2.4.6 MapReduce运行速度	48
2.5 HBase概述	48
2.5.1 HBase的系统框架	49
2.5.2 HBase访问接口	51
2.5.3 HBase的存储格式	52
2.5.4 HBase的读写流程	52
2.5.5 Hbase的优缺点	53
2.6 ZooKeeper概述	53
2.6.1 为什么需要ZooKeeper	54
2.6.2 ZooKeeper设计目标	54
2.6.3 ZooKeeper数据模型	54
2.6.4 ZooKeeper工作原理	55
2.6.5 ZooKeeper实现机制	56
2.6.6 ZooKeeper的特性	57
2.7 Hive概述	58
2.7.1 Hive的组成	59
2.7.2 Hive结构解析	59
2.8 Pig概述	63
2.9 Cassandra概述	64
2.9.1 Cassandra主要功能	64
2.9.2 Cassandra的体系结构	65
2.9.3 Cassandra存储机制	65
2.9.4 Cassandra的写过程	66
2.9.5 Cassandra的读过程	67
2.9.6 Cassandra的删除	68
2.10 Chukwa概述	68
2.10.1 使用Chukwa的原因	68
2.10.2 Chukwa的不是	69
2.10.3 Chukwa的定义	69
2.10.4 Chukwa架构与设计	70
第3章 Hadoop配置与实战	74
3.1 Hadoop的安装	74
3.1.1 在Linux下安装Hadoop	74
3.1.2 运行模式	75
3.1.3 在Windows下安装Hadoop	80
3.2 运行Hadoop	86
3.3 Hadoop的Avatar机制	87
3.3.1 系统架构	88
3.3.2 元数据同步机制	89
3.3.3 切换故障过程	91
3.3.4 运行流程	92
3.3.5 切换故障流程	96
3.4 Hadoop实战	99
3.4.1 使用Hadoop运行wordcount实例	99
3.4.2 使用Eclipse编写Hadoop程序	101
第4章 Hadoop的分布式数据HDFS	102
4.1 HDFS的操作	102
4.1.1 文件操作	102
4.1.2 管理与更新	103

- 4.2 FS Shell使用指南 104
- 4.3 API使用 111
 - 4.3.1 文件系统的常见操作 111
 - 4.3.2 API的Java操作实例 113
- 第5章 Hadoop编程模型MapReduce 118
 - 5.1 MapReduce基础 118
 - 5.1.1 MapReduce编程模型 118
 - 5.1.2 MapReduce实现机制 119
 - 5.1.3 Java MapReduce 121
 - 5.2 MapReduce的容错性 124
 - 5.3 MapReduce实例分析 125
 - 5.4 不带map()、reduce()的MapReduce 131
 - 5.5 Shuffle过程 133
 - 5.6 新增Hadoop API 136
 - 5.7 Hadoop的Streaming 138
 - 5.7.1 通过UNIX命令使用Streaming 138
 - 5.7.2 通过Ruby版本使用Streaming 139
 - 5.7.3 通过Python版本使用Streaming 141
 - 5.8 MapReduce实战 142
 - 5.8.1 MapReduce排序 142
 - 5.8.2 MapReduce二次排序 145
 - 5.9 MapReduce作业分析 153
 - 5.10 定制MapReduce数据类型 156
 - 5.10.1 内置的数据输入格式和RecordReader 156
 - 5.10.2 定制输入数据格式与RecordReader 157
 - 5.10.3 定制数据输出格式实现多集合文件输出 160
 - 5.11 链接MapReduce作业 162
 - 5.11.1 顺序链接MapReduce作业 162
 - 5.11.2 复杂的MapReduce链接 163
 - 5.11.3 前后处理的链接 163
 - 5.11.4 链接不同的数据 166
 - 5.12 Hadoop的Pipes 172
 - 5.13 创建Bloom filter 174
 - 5.13.1 Bloom filter作用 175
 - 5.13.2 Bloom filter实现 175
- 第6章 Hadoop的数据库HBase 182
 - 6.1 HBase数据模型 182
 - 6.1.1 数据模型 182
 - 6.1.2 概念视图 183
 - 6.1.3 物理视图 184
 - 6.2 HBase与RDBMS对比 185
 - 6.3 Bigtable的应用实例 188
 - 6.4 HBase的安装与配置 189
 - 6.5 Java API 196
 - 6.6 HBase实例分析 204
 - 6.6.1 RowLock 204
 - 6.6.2 HBase的HFileOutputFormat 207
 - 6.6.3 HBase的TableOutputFormat 210
 - 6.6.4 在HBase中使用MapReduce 213
 - 6.6.5 HBase分布式模式 215
- 第7章 Hadoop的数据仓库Hive 220
 - 7.1 Hive的安装 220
 - 7.1.1 准备的软件包 220
 - 7.1.2 内嵌模式安装 220

- 7.1.3 安装独立模式 221
- 7.1.4 远程模式安装 222
- 7.1.5 查看数据信息 222
- 7.2 Hive的入口 223
 - 7.2.1 类CliDriver 225
 - 7.2.2 类CliSessionState 229
 - 7.2.3 类CommandProcessor 230
- 7.3 Hive QL详解 232
 - 7.3.1 Hive的数据类型 232
 - 7.3.2 Hive与数据库比较 233
 - 7.3.3 DDL操作 234
 - 7.3.4 join查询 241
 - 7.3.5 DML操作 243
 - 7.3.6 SQL操作 245
 - 7.3.7 Hive QL的应用实例 248
- 7.4 Hive的服务 250
 - 7.4.1 JDBC/ODBC服务 250
 - 7.4.2 Thrift服务 253
 - 7.4.3 Web接口 255
- 7.5 Hive SQL的优化 256
 - 7.5.1 Hive SQL优化选项 256
 - 7.5.2 Hive SQL优化应用实例 258
- 7.6 Hive的扩展性 261
 - 7.6.1 SerDe 262
 - 7.6.2 Map/Reduce脚本 263
 - 7.6.3 UDF 263
 - 7.6.4 UDAF 264
- 7.7 Hive实战 266
- 第8章 Hadoop的大规模数据平台Pig 274
 - 8.1 Pig的安装与运行 274
 - 8.1.1 Pig的安装 274
 - 8.1.2 Pig的运行 274
 - 8.2 Pig实现 278
 - 8.3 Pig Latin语言 279
 - 8.3.1 Pig Latin语言概述 280
 - 8.3.2 Pig Latin数据类型 282
 - 8.3.3 Pig Latin运算符 284
 - 8.3.4 Pig Latin关键字 287
 - 8.3.5 Pig内置函数 288
 - 8.4 自定义函数 291
 - 8.4.1 UDF的编写 292
 - 8.4.2 UDFS的使用 293
 - 8.5 Jaql和Pig查询语言的比较 293
 - 8.5.1 Pig和Jaql运行环境和执行形式的比较 294
 - 8.5.2 Pig和Jaql支持数据类型的比较 294
 - 8.5.3 Pig和Jaql操作符和内建函数以及自定义函数的比较 295
 - 8.5.4 其他 299
 - 8.6 Pig实战 300
- 第9章 Hadoop的非关系型数据Cassandra 308
 - 9.1 Cassandra的安装 308
 - 9.1.1 在Windows 7中安装 308
 - 9.1.2 在Linux中安装 310
 - 9.2 Cassandra的数据模型 311
 - 9.2.1 Column 311

9.2.2 SuperColumn	312
9.2.3 ColumnFamily	312
9.2.4 Row	313
9.2.5 排序	313
9.3 Cassandra的实例分析	315
9.3.1 Cassandra的数据存储结构	315
9.3.2 跟踪客户端代码	319
9.4 Cassandra常用的编程语言	324
9.4.1 Java使用Cassandra	324
9.4.2 PHP使用Cassandra	325
9.4.3 Python使用Cassandra	326
9.4.4 C#使用Cassandra	327
9.4.5 Ruby使用Cassandra	328
9.5 Cassandra与MapReduce结合	328
9.5.1 需求分析	329
9.5.2 代码分析	330
9.5.3 MapReduce代码	330
9.6 Cassandra实战	331
9.6.1 BuyerDao功能验证	331
9.6.2 SellerDao功能验证	332
9.6.3 ProductDao功能验证	333
9.6.4 新建Schema在线功能	336
9.6.5 功能验证	337
第10章 Hadoop的收集数据Chukwa	339
10.1 Chukwa的安装与配置	339
10.1.1 配置要求	339
10.1.2 Chukwa的安装	340
10.1.3 基本命令	341
10.2 Chukwa数据流处理	344
10.2.1 支持数据类型	344
10.2.2 数据处理	345
10.2.3 自定义数据模块	351
10.3 Chukwa源代码分析	352
10.3.1 Chukwa适配器	352
10.3.2 Chukwa连接器	357
10.3.3 Chukwa收集器	362
10.4 Chukwa实例分析	366
10.4.1 生成数据	366
10.4.2 收集数据	367
10.4.3 处理数据	367
10.4.4 析取数据	368
10.4.5 稀释数据	368
第11章 Hadoop的分布式系统ZooKeeper	369
11.1 ZooKeeper的安装与配置	369
11.1.1 ZooKeeper的安装	369
11.1.2 ZooKeeper的配置	371
11.1.3 ZooKeeper数据模型	373
11.1.4 ZooKeeper的API接口	373
11.1.5 ZooKeeper编程实现	375
11.2 ZooKeeper的Leader流程	378
11.3 ZooKeeper锁服务	379
11.3.1 ZooKeeper中的锁机制	379
11.3.2 ZooKeeper的写锁实现	380
11.3.3 ZooKeeper锁服务实现例子	381

11.4 创建ZooKeeper应用程序 383
11.5 ZooKeeper的应用开发 387
11.6 ZooKeeper的典型应用 395
11.6.1 统一命名服务 396
11.6.2 配置管理 396
11.6.3 集群管理 397
11.6.4 共享锁 398
11.6.5 队列管理 399
11.7 实现NameNode自动切换 402
网上参考资料 410
参考文献 412
· · · · · (收起)

[Hadoop云计算实战_下载链接1_](#)

标签

hadoop

数据挖掘

云计算

Hadoop云计算实战

评论

能把书写成这样..内容有错..汉语不同...重点不明..该有的没有...

一般般

[Hadoop云计算实战_下载链接1_](#)

书评

目前正在看这本书，正在了解其中的Zookeeper。 读书笔记：

<http://www.wangyuxiong.com/archives/51895>

<http://www.wangyuxiong.com/archives/51889>

[Hadoop云计算实战 下载链接1](#)