

图书基本信息

书名：<<大数据技术全解：基础、设计、开发与实践>>

13位ISBN编号：9787121223433

10位ISBN编号：7121223430

出版时间：2014-1

出版时间：电子工业出版社

作者：杨巨龙

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>

## 内容概要

“大数据”一词已成为当下最热门的词汇之一。

人们热切期待大数据能给工作、学习和生活等方面带来前所未有的变化，而大数据的采集、存储、计算、分析挖掘、呈现和安全等技术正是实现这一变化的关键。

《大数据技术全解：基础、设计、开发与实践》的重点不在于对大数据技术的原理性介绍，而是从实战角度出发，系统地阐述如何基于Hadoop开源软件为客户设计、安装和开发一个大数据系统。

《大数据技术全解：基础、设计、开发与实践》还着重分享了具体实践中的一些案例，不仅对于从事大数据系统开发的IT技术人员具有珍贵的参考价值，而且对准备实施大数据系统项目的企业级客户也具有指导作用。

《大数据技术全解：基础、设计、开发与实践》可以作为计算机软件专业的本科生和研究生的大数据技术教材，也可作为大数据技术开发培训、大数据系统咨询规划培训的教材，更是大数据领域内一线的需求分析人员、系统分析人员、高级开发人员和开发管理人员的必备参考书。

书籍目录

第1篇 基础篇	
第1章 大数据的三把利剑.....	2
1.1 豌豆杂交实验.....	2
1.2 曹冲称象启示.....	2
1.3 谷歌的三把利剑.....	4
1.4 智慧改变世界.....	4
第2章 企业的大数据观.....	5
2.1 企业面临的挑战.....	5
2.1.1 数据能力是核心竞争力.....	5
2.1.2 从粗放经营到智慧经营.....	6
2.1.3 技术与商业的双重挑战.....	6
2.2 企业大数据从哪来.....	7
2.2.1 来自于主体的产生.....	7
2.2.2 来自于客体的产生.....	8
2.2.3 来自于社会的产生.....	9
2.3 企业大数据如何存储.....	9
2.3.1 非结构化数据存储.....	9
2.3.2 结构化数据存储.....	10
2.3.3 半结构化数据存储.....	11
2.3.4 大数据存储的问题.....	12
2.4 企业大数据如何加工.....	13
2.4.1 分析或挖掘模型设计.....	14
2.4.2 并行处理程序编码.....	15
2.4.3 结果在全局中呈现.....	17
2.5 企业的大数据到哪里去.....	17
2.5.1 大数据对企业影响深远.....	17
2.5.2 大数据是一种新商品.....	18
2.5.3 精准营销需要大数据.....	18
2.6 企业大数据观总结.....	19
第3章 大数据和大数据系统.....	20
3.1 大数据.....	20
3.1.1 大数据概念.....	20
3.1.2 大数据的特征.....	21
3.1.3 数据计量单位.....	21
3.1.4 大数据来源.....	22
3.1.5 大数据类型.....	22
3.2 大数据系统.....	22
3.2.1 设计目标和原则.....	23
3.2.2 系统的设计思想.....	23
3.2.3 系统的逻辑架构.....	25
3.2.4 与现有系统的关系.....	26
3.2.5 当前的大数据系统.....	28
第2篇 技术篇	
第4章 分布、键值对与族.....	36
4.1 分布与MapReduce .....	36

4.2 键值对的奥妙所在.....	38
4.3 动态数据库表原理.....	39
第5章 HDFS (分布式文件系统) .....	40
5.1 设计目标.....	40
5.2 基本概念.....	40
5.2.1 块.....	41
5.2.2 名称节点与数据节点.....	41
5.3 系统架构.....	42
5.3.1 逻辑架构.....	42
5.3.2 物理架构.....	43
5.4 运行机制.....	43
5.4.1 文件读取.....	43
5.4.2 文件写入.....	44
5.4.3 特别关注.....	46
5.5 系统功能.....	49
5.5.1 多文件系统.....	49
5.5.2 目录管理.....	50
5.5.3 文件管理.....	50
5.5.4 文件归档.....	51
5.5.5 并行复制.....	52
5.6 系统I/O 特性.....	52
5.6.1 完整性校验.....	52
5.6.2 压缩与编码解码.....	54
5.6.3 序列化.....	55
5.6.4 特殊文件结构.....	56
5.7 非Java 访问接口.....	57
5.8 系统性能.....	58
5.8.1 可靠性措施.....	58
5.8.2 性能优化.....	59
第6章 MapReduce 分布式编程模式....	60
6.1 不同于传统.....	60
6.2 设计思想.....	61
6.3 基本概念.....	61
6.3.1 map()函数.....	62
6.3.2 reduce()函数.....	62
6.3.3 键值对.....	63
6.3.4 中间结果.....	63
6.3.5 移动代码.....	64
6.3.6 作业和任务节点.....	65
6.4 系统架构.....	65
6.4.1 逻辑架构.....	65
6.4.2 物理架构.....	66
6.5 运行机制.....	67
6.5.1 作业运行.....	67
6.5.2 作业调度.....	69
6.5.3 任务执行.....	70
6.5.4 状态更新.....	70

<<大数据技术全解：基础、设计、开>>

6.5.5 作业完成.....	71
6.5.6 故障处理.....	71
6.6 关键技术.....	72
6.6.1 计数器.....	72
6.6.2 排序.....	73
6.6.3 连接.....	73
6.6.4 shuffle.....	74
6.6.5 内存处理.....	75
6.6.6 分布式缓存.....	76
6.7 类型与格式.....	77
6.7.1 MR 的类型.....	77
6.7.2 输入格式.....	78
6.7.3 输出格式.....	80
6.8 MR 的开发.....	81
6.8.1 开发端环境的建立.....	82
6.8.2 开发及单元测试.....	84
6.8.3 本地运行测试.....	87
6.8.4 集群运行.....	89
6.8.5 作业调试.....	92
6.8.6 远程调试.....	93
6.8.7 作业调优.....	94
6.9 MR  workflow.....	95
第7章 HBase 分布式数据库.....	96
7.1 设计目标.....	96
7.2 基本概念.....	97
7.2.1 逻辑模型.....	97
7.2.2 物理模型.....	98
7.2.3 区域.....	99
7.2.4 基本单元.....	99
7.2.5 Region 服务器.....	100
7.2.6 Master 主服务器.....	101
7.2.7 .META.元数据表.....	101
7.2.8 -ROOT-元数据表.....	102
7.3 系统架构.....	103
7.3.1 逻辑架构.....	104
7.3.2 物理架构.....	105
7.4 运行机制.....	105
7.5 系统功能.....	107
7.5.1 用户界面.....	107
7.5.2 shell 操作.....	109
7.6 库表设计.....	112
7.7 访问接口.....	112
第3篇 设计篇	
第8章 系统设计背景和目标.....	116
8.1 系统设计背景.....	116
8.2 系统设计目标.....	117
8.2.1 存在问题.....	117

8.2.2 设计目标.....	118
第9章 系统架构设计.....	121
9.1 逻辑架构设计.....	121
9.1.1 系统逻辑架构.....	121
9.1.2 系统运行逻辑.....	122
9.2 功能架构设计.....	123
9.2.1 大数据管理系统的功能.....	124
9.2.2 ZooKeeper 系统的功能.....	126
9.2.3 Chukwa 采集系统的功能.....	127
9.2.4 Pig 系统功能.....	127
9.2.5 Hive 系统功能.....	128
9.3 数据架构设计.....	128
9.3.1 数据总体架构.....	129
9.3.2 分布式文件数据结构.....	130
9.3.3 分布式数据库数据结构.....	130
9.3.4 关系型数据库数据构成.....	131
第10章 运行架构设计.....	132
10.1 物理架构设计.....	133
10.1.1 网络拓扑.....	133
10.1.2 软/硬件选型.....	133
10.2 集成架构设计.....	135
10.2.1 总体集成设计.....	135
10.2.2 专项集成设计.....	137
10.3 安全架构设计.....	140
10.3.1 用户层安全.....	140
10.3.2 应用层安全.....	141
10.3.3 数据层安全.....	141
10.4 开发架构设计.....	142
第4篇 安装篇	
第11章 安装规划.....	144
11.1 安装目标.....	144
11.2 安装步骤.....	145
第12章 环境准备.....	147
12.1 主机准备.....	147
12.2 介质准备.....	147
12.3 基础安装.....	148
12.3.1 JDK 安装.....	148
12.3.2 用户创建.....	148
12.3.3 SSH 配置.....	149
第13章 集群安装.....	150
13.1 HDFS 集群.....	150
13.1.1 解析配置.....	150
13.1.2 模板创建.....	151
13.1.3 复制分发.....	153
13.1.4 运行启动.....	153
13.1.5 测试验证.....	154
13.2 HBase 集群.....	156

13.2.1 解析配置.....	156
13.2.2 模板创建.....	157
13.2.3 复制分发.....	158
13.2.4 运行启动.....	158
13.2.5 测试验证.....	158
13.3 ZooKeeper 集群.....	159
13.3.1 解析配置.....	159
13.3.2 模板创建.....	160
13.3.3 复制分发.....	160
13.3.4 运行启动.....	161
13.3.5 测试验证.....	161
第14章 分布式应用安装.....	163
14.1 Pig 安装.....	163
14.1.1 本地安装.....	163
14.1.2 本地验证.....	164
14.1.3 集成配置.....	164
14.1.4 集成验证.....	165
14.2 Hive 安装.....	165
14.2.1 内嵌安装与验证.....	166
14.2.2 从内嵌改为独立.....	166
14.2.3 从独立改为远程.....	169
14.3 Chukwa 安装.....	171
14.3.1 基础系统安装.....	171
14.3.2 代理系统安装.....	172
14.3.3 收集系统的安装.....	174
14.3.4 作业系统的安装.....	175
14.3.5 HICC 系统的安装.....	176
第15章 集成联调.....	177
15.1 集群间的集成联调.....	177
15.1.1 HBase 与HDFS集成联调.....	177
15.1.2 HBase 与ZooKeeper集成联调.....	178
15.2 分布式应用与集群间的集成联调.....	179
15.2.1 Pig 与HDFS 的集成.....	179
15.2.2 Hive 与HDFS 的集成.....	180
15.2.3 Hive 与Hbase的集成.....	181
15.2.4 Chukwa 与HDFS的集成.....	183
15.3 客户端与分布式系统间的集成联调.....	184
15.3.1 与分布式集群系统的集成.....	184
15.3.2 与分布式应用系统的集成.....	185
第5篇 开发篇	
第16章 大数据系统应用开发思路和环境.....	188
16.1 总体思路.....	188
16.1.1 大数据读写应用的开发.....	188
16.1.2 大数据分析应用的开发.....	188
16.2 开发环境.....	189
16.2.1 Plugin 插件的安装.....	189
16.2.2 Hadoop 开发环境的配置.....	190

16.2.3 示例程序验证.....	191
第17章 HDFS 文件读/写应用开发.....	196
17.1 文件列表.....	196
17.2 文件读取.....	197
17.3 文件上传.....	198
17.4 文件创建.....	199
17.5 文件写入.....	200
17.6 文件压缩与解压.....	201
17.6.1 压缩写入.....	201
17.6.2 解压后读取.....	203
17.7 目录创建.....	205
17.8 文件重命名.....	206
17.9 删除文件.....	207
17.10 查看文件时间.....	208
17.11 查看文件是否存在.....	209
17.12 查找文件位置.....	210
17.13 查找集群所有的节点.....	211
17.14 SequenceFile 文件格式转换.....	212
17.14.1 创建SequenceFile 格式的文件.....	213
17.14.2 读取SequenceFile 格式的文件.....	215
17.15 MapFile 文件格式转换.....	217
17.15.1 创建MapFile 格式的文件.....	217
17.15.2 读取MapFile 格式的文件.....	218
17.15.3 SequenceFile 格式转换成MapFile 格式.....	220
第18章 HBase 数据库读/写应用开发.....	222
18.1 创建表.....	222
18.2 删除表.....	224
18.3 查询数据库中的表.....	225
18.4 插入记录.....	226
18.5 查询记录.....	228
18.5.1 列族的查询.....	228
18.5.2 查询所有记录.....	229
18.5.3 基于行键查询.....	231
18.5.4 基于标签值查询.....	232
18.5.5 组合条件查询.....	234
18.6 修改记录.....	236
18.7 删除记录.....	237
第19章 ZooKeeper 开发.....	238
19.1 创建节点.....	238
19.2 删除节点.....	240
19.3 加入子节点.....	242
19.4 列出节点成员.....	242
19.5 获取节点内容.....	244
第20章 MapReduce 开发.....	247
20.1 定制数据类型.....	247
20.2 定制输入格式.....	250
20.3 定制输出格式.....	254



20.4 将整个文件作为输入.....	259
20.5 小文件聚合成一个文件.....	264
20.6 多集合文件输出.....	266
20.7 对压缩数据处理.....	268
20.8 定制partitioner .....	271
20.9 定制combiner .....	274
20.10 MapReduce 组合.....	278
20.10.1 迭代组合.....	278
20.10.2 线性组合.....	283
20.10.3 依赖组合.....	284
20.10.4 前后链式组合.....	288
20.11 多数据源连接.....	292
20.11.1 Reduce 端连接.....	292
20.11.2 Map 端连接.....	297
20.11.3 Map 端过滤的Reduce端连接.....	300
20.12 全局参数应用.....	305
20.13 全局文件应用.....	309
20.14 关系数据库访问.....	311
20.14.1 关系数据库读取.....	312
20.14.2 关系数据库的写入.....	315
第21章 Pig 开发.....	319
21.1 脚本编程.....	319
21.1.1 脚本语言.....	319
21.1.2 脚本编程.....	322
21.1.3 脚本运行.....	323
21.2 自定义函数.....	327
21.2.1 编译打包.....	327
21.2.2 测试运行.....	328
第22章 Hive 开发.....	329
22.1 HiveQL 语言.....	329
22.1.1 HiveQL 的数据类型.....	329
22.1.2 HiveQL 的常用操作.....	330
22.2 UDF 编码.....	339
22.3 UDAF 编码.....	341
22.4 客户端编码.....	343
22.4.1 与Hive 服务器端连接建立.....	343
22.4.2 与Hive 进行指令交互....	344
22.4.3 客户端命令组织.....	345
22.4.4 程序运行结果.....	346
第6篇 实践篇	
第23章 企业大数据盘系统.....	348
23.1 系统开发背景.....	348
23.2 系统架构设计.....	348
23.3 系统功能设计.....	349
23.4 系统代码实现.....	350
第24章 Hadoop 的日志分析.....	352
24.1 系统开发背景.....	352

24.2 系统架构设计.....	352
24.3 系统功能设计.....	353
24.4 系统代码实现.....	353
24.4.1 系统连接代码.....	353
24.4.2 Hive 模块代码.....	355
24.4.3 系统组织代码.....	356
24.5 系统实现效果.....	357
参考文献.....	358

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>