

《splunk大数据分析》

图书基本信息

书名：《splunk大数据分析》

13位ISBN编号：978711146429X

出版时间：2014-5

作者：（美）Peter Zadrozny,（美）Raghu Kodali

译者：唐宏,陈健

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：www.tushu000.com

《splunk大数据分析》

内容概要

Splunk是一种典型的大数据处理工具，能够高效地按时序对数据进行存储、索引、访问，已广泛应用于多个领域。本书是介绍如何实时处理大数据并从中获得商业价值的一本实用指南。本书通过真实的大数据分析项目，从数据导入、访问、挖掘和可视化角度全面而系统地介绍Splunk的基本概念和使用方法，以帮助读者快速掌握Splunk。

全书共16章，分为四个部分：第一部分（第1~7章）介绍Splunk的基本操作，包括利用Splunk进行数据收集、处理、分析及结果可视化等的基本操作和命令，以及使用日志文件创建高级数据分析报表的方法；第二部分（第8~11章）使用“航空公司准点性能数据”介绍一个典型的数据分析案例，详细讲解如何运用Splunk深度挖掘现有数据仓库，并介绍一些新的Splunk命令和实用技巧；第三部分（第12~14章）详细介绍如何收集、处理、分析推文和Foursquare的签到信息等，涵盖分析社会化媒体流数据所需的知识；第四部分（第15~16章）详细介绍如何按需求扩展Splunk，以及分布式处理和高可用性的基本概念。此外，还包括两个附录，展示Splunk的性能以及各种可用的应用程序。

作者简介

Peter Zadrozny 资深软件技术专家，Opallios公司创始人兼首席技术官，专注于利用大数据和云技术为客户提供有价值产品。Peter还是圣荷西州立大学大数据相关课程的讲师。他曾先后在多家大中型公司担任行政和技术职务，并主导在欧洲开启WebLogic和在墨西哥启动Sun微操作系统。他在J2EE和性能工程领域有多部非常成功的著作，并且是流行开源项目Grinder的最初贡献者。

Raghu Kodali Solix公司产品管理和产品策略副总裁，负责产品愿景、产品管理、产品策略、用户设计和交互。在进入Solix之前，他在Oracle工作了12年，担任产品管理和产品策略高级经理。此外，Raghu是Oracle SOA传道者，著有《Beginning EJB 3 Application Development: From Novice to Profession》和《Oracle Information Integration, Migration and Consolidation》，同时发表多篇关于企业技术的文章。

书籍目录

- 第1章 大数据和Splunk / 1
 - 1.1 什么是大数据 / 1
 - 1.2 非传统的数据处理技术 / 5
 - 1.3 Splunk是什么 / 6
 - 1.4 关于本书 / 7
- 第2章 将数据导入Splunk / 9
 - 2.1 数据的多样性 / 9
 - 2.2 Splunk如何处理多样化的数据 / 10
 - 2.2.1 文件和目录 / 11
 - 2.2.2 数据生成器 / 16
 - 2.2.3 生成样本数据 / 17
 - 2.2.4 网络资源 / 21
 - 2.2.5 Windows数据 / 21
 - 2.2.6 其他资源 / 21
 - 2.3 应用程序和附加组件 / 21
 - 2.4 转发器 / 26
 - 2.5 小结 / 27
- 第3章 处理和分析数据 / 28
 - 3.1 了解组合访问日志数据 / 28
 - 3.2 搜索和分析索引数据 / 29
 - 3.3 报表 / 35
 - 3.3.1 使用最多的浏览器 / 35
 - 3.3.2 排名前五的IP地址 / 37
 - 3.3.3 浏览量来源最多的网站 / 38
 - 3.3.4 有多少404事件 / 40
 - 3.3.5 有多少事件包含购买行为 / 42
 - 3.3.6 列出购买的商品 / 42
 - 3.4 排序 / 44
 - 3.5 过滤 / 45
 - 3.6 添加和评估字段 / 47
 - 3.7 聚合 / 48
 - 3.8 小结 / 54
- 第4章 结果的可视化 / 55
 - 4.1 数据可视化 / 55
 - 4.2 Splunk是怎样处理可视化的 / 55
 - 4.3 chart / 60
 - 4.3.1 制作每一个主机的GET和POST事件数量的图表 / 61
 - 4.3.2 制作每一个产品类别的购买数和浏览数的图表 / 62
 - 4.3.3 哪个产品种类受HTTP 404错误的影响 / 63
 - 4.3.4 MyGizmoStore.com的购买趋势 / 64
 - 4.3.5 事务持续时间 / 66
 - 4.4 timechart / 67
 - 4.4.1 最高购买数量的产品 / 67
 - 4.4.2 页面浏览率和购买量 / 68
 - 4.5 使用Google Maps应用程序来可视化 / 69
 - 4.6 Globe / 71
 - 4.7 仪表盘 / 72

- 4.8 小结 / 80
- 第5章 定义警报 / 81
 - 5.1 什么是警报 / 81
 - 5.2 Splunk如何提供警报 / 81
 - 5.2.1 基于商品销售量的警报 / 82
 - 5.2.2 登录失败的警报 / 84
 - 5.2.3 日志文件中关键性错误的警报 / 87
 - 5.3 小结 / 88
- 第6章 网站监测 / 90
 - 6.1 监测网站 / 90
 - 6.2 IT运作 / 91
 - 6.2.1 主机访问量 / 91
 - 6.2.2 无内部访问的主机访问量 / 91
 - 6.2.3 HTTP请求成功的流量 / 93
 - 6.2.4 HTTP请求未成功的流量 / 93
 - 6.2.5 返回HTTP错误状态码最多的页面 / 94
 - 6.3 业务 / 96
 - 6.3.1 区域用户统计 / 96
 - 6.3.2 跳出率 / 97
 - 6.3.3 独立访问者数量 / 98
 - 6.4 小结 / 103
- 第7章 使用日志文件创建高级分析 / 104
 - 7.1 传统的分析方法 / 104
 - 7.2 范式变更 / 105
 - 7.3 语义日志 / 106
 - 7.4 日志最佳实践 / 113
 - 7.5 小结 / 115
- 第8章 航班准点率项目 / 116
 - 小结 / 118
- 第9章 将航班数据导入Splunk / 119
 - 9.1 处理CSV文件 / 119
 - 9.1.1 航班数据 / 119
 - 9.1.2 下载数据 / 120
 - 9.1.3 了解航班数据 / 121
 - 9.1.4 关于时间戳 / 123
 - 9.1.5 将字段映射成一个时间戳 / 124
 - 9.1.6 对所有航班数据建立索引 / 131
 - 9.2 从关系数据库中索引数据 / 132
 - 9.2.1 定义一个新的数据库连接 / 132
 - 9.2.2 数据库监测 / 133
 - 9.3 小结 / 136
- 第10章 分析航空公司、机场、航班和延迟 / 137
 - 10.1 分析航空公司 / 137
 - 10.1.1 计算航空公司的总数 / 138
 - 10.1.2 可视化结果 / 139
 - 10.2 分析机场 / 143
 - 10.3 分析航班 / 146
 - 10.4 分析延迟 / 151
 - 10.4.1 各航空公司航班延迟情况 / 151

- 10.4.2 各机场航班延迟的原因 / 152
- 10.4.3 冬天与夏天的航班延迟情况 / 155
- 10.5 创建和使用宏命令 / 157
- 10.6 报告加速 / 158
- 10.7 加速统计 / 161
- 10.8 小结 / 166
- 第11章 分析一个特定航班的历年数据 / 167
 - 11.1 航空公司名称 / 167
 - 11.1.1 字段查找自动化 / 172
 - 11.1.2 从搜索中创建查找表 / 173
 - 11.2 United flight 871航班 / 174
 - 11.3 小结 / 178
- 第12章 分析推文 / 179
 - 12.1 开发样本流 / 180
 - 12.2 将推文加载到Splunk中 / 183
 - 12.3 Twitter / 185
 - 12.4 最流行的单词 / 188
 - 12.5 实时的Twitter趋势 / 191
 - 12.6 小结 / 196
- 第13章 分析Foursquare签到信息 / 197
 - 13.1 签到信息格式 / 198
 - 13.2 时区注意事项 / 202
 - 13.3 装载签到数据 / 203
 - 13.4 分析签到信息 / 205
 - 13.4.1 星期日早午餐搜索 / 205
 - 13.4.2 Google地图和热门地点 / 209
 - 13.4.3 地点的签到模式 / 211
 - 13.4.4 地点的签到数量 / 212
 - 13.4.5 分析性别活动 / 214
 - 13.5 小结 / 217
- 第14章 情感分析 / 218
 - 14.1 意见、观点、信仰、信念 / 218
 - 14.2 商业用途 / 219
 - 14.3 情感分析的技术性工作 / 220
 - 14.4 情感分析应用程序 / 222
 - 14.4.1 全局性的命令 / 223
 - 14.4.2 挖掘情感 / 224
 - 14.4.3 语言的处理 / 226
 - 14.4.4 训练数据和测试数据 / 227
 - 14.5 世界情绪指数项目 / 231
 - 14.5.1 收集RSS摘要 / 232
 - 14.5.2 将新闻标题索引到Splunk中 / 234
 - 14.5.3 定义情感语料库 / 237
 - 14.5.4 对结果进行可视化 / 240
 - 14.6 小结 / 242
- 第15章 远程数据收集 / 243
 - 15.1 转发器 / 243
 - 15.1.1 流行的拓扑结构 / 244
 - 15.1.2 安装转发器 / 246

15.2	部署服务器 / 248
15.2.1	配置部署服务器 / 250
15.2.2	配置转发器 / 251
15.3	部署监控 / 252
15.4	小结 / 253
第16章	可扩展性和高可用性 / 254
16.1	扩展Splunk / 254
16.2	聚类 / 259
16.3	小结 / 264
附录A	Splunk的性能 / 265
附录B	有用的Splunk应用程序 / 281

精彩短评

1、对企业级Splunk架构已经很熟了，感觉不用再看了。

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:www.tushu000.com