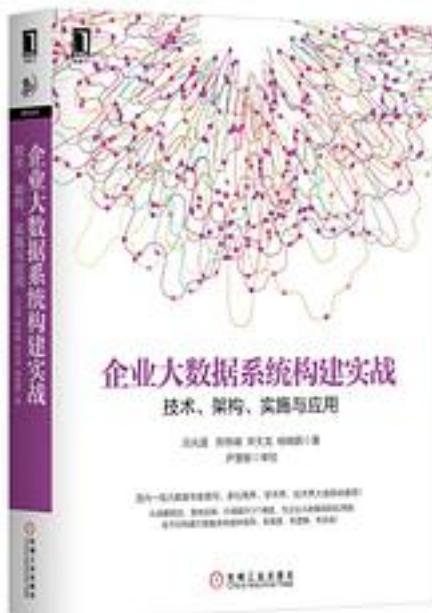


企业大数据系统构建实战：技术、架构、实施与应用



[企业大数据系统构建实战：技术、架构、实施与应用 下载链接1](#)

著者:吕兆星 郑传峰 宋天龙 杨晓鹏 著

出版者:机械工业出版社

出版时间:2017-5-26

装帧:平装

isbn:9787111568766

对于很多企业而言，大数据的重要性不言而喻，但是如何构建、实施和应用大数据系统却是一个复杂工程。本书让读者认识到大数据不仅仅是数据、技术、架构、应用，更是结合了商业模式、战略定位、信息安全、单位协同、组织保障、实施选型的完整体系。

本书内容从大数据的规划定位、组织实施和价值提升三个维度展开，兼顾从整体性、全局性、安全性、价值性、技术性、体系性等方面的考虑。

第一部分：企业大数据战略规划

主要从宏观的角度介绍大数据的定位、组织保障、解决方案选择和自主实施思路，目的是从全局角度引导建立大数据工作的整体思维。

第二部分：企业大数据落地实施

主要从执行层面介绍了大数据落地相关的技术、架构、开发、大数据工作流、应用和价值评估，直接以落地视角解读大数据工作中每个环节涉及到的流程、知识和方法，这也是本书的核心章节。

第三部分：大数据价值、变革和挑战

主要涉及大数据的社会价值、当前问题和挑战以及大数据的未来趋势，这是对现有大数据工作的延展以及未来趋势的探索。

作者介绍：

吕兆星 (EthanLv)

资深大数据技术专家，精通基于大数据的分布式数据挖掘、存储与计算技术，及其生态体系架构；精通垂直搜索技术、机器学习、文本情感倾向性挖掘、网络爬虫、全文索引体系架构。曾任软通动力集团大数据研究院总架构师、HiveCloud创始人，萝卜网CTO，国美在线大数据中心高级架构师等。

主导研发的大数据和文本挖掘平台包括：DMP、DSP、推荐系统、决策运营系统、iCreations系列产品、蜂棱系列产品、军犬舆情系列产品等。成功应用到能源、电力、电商、电信、金融、政府、食品、医疗保健等行业，超过500个政府和企业用户。《基于机器学习的数据挖掘模型》获得国家级技术创新基金，萝卜课堂特邀高级讲师。

郑传峰 (PeterZheng)

大数据业务应用领域专家，主导大数据方向战略规划，包含数据产品、数据应用、数据价值变现等方向。曾任软通动力数据科技公司资深数据应用专家，HiveCloud首席战略官。

阶段性负责国美电器、国美在线、库巴网会员营销、网站运营和产品设计工作，在CRM系统、DMP数据平台、精准营销系统、广告精投、能源大数据拥有多年的操盘经验，包含大数据上层应用服务产品设计、咨询和实施。参与多个大型企业大数据战略规划和实施，行业覆盖零售、电商、电信、政府、交通、能源和电力等。

宋天龙 (TonySong)

大数据领域的资深数据分析、挖掘和建模专家，精通端到端数据价值场景设计、业务需求转换、数据结构梳理、数据建模与学习，以及数据工程交付。曾任软通动力集团大数据研究院数据总监，Webtrekk（德国最大的网站数据分析服务提供商）中国区技术和咨询负责人，国美大数据中心经理。

拥有丰富的大数据项目工作经验，参与过集团和企业级大数据存储平台、大数据开发和集成平台、数据体系规划、大数据产品开发、网站流量系统建设、网站智能推荐、企业大数据智能等大型数据工作项目。参与实施客户案例包括Webpower、德国OTTO集团电子商务（中国），Esprit中国、猪八戒网、顺丰优选、乐视商城、泰康人寿、酒仙网，国美在线、迪信通等。合作培训及沙龙单位包括人民大学、数盟、萝卜网、Netconc

epts、触脉、中商联数据分析委等。萝卜课堂、天善学院特邀讲师，百度文库认证作家，36大数据、站长之家、互联网分析沙龙专栏作家。著有《网站数据挖掘与分析：系统方法与商业实践》一书。

杨晓鹏 (KelvinYang)

大数据及BI技术领域资深架构师，精通传统数据模式及大数据分布模式的数据存储、计算与应用架构，以及大数据量的数据迁移、存储、索引、计算、分析与挖掘等相关环节的设计、实现与优化。曾任软通动力集团大数据研究院高级架构师，HiveCloud总架构师，主导大数据存储平台、计算平台和应用服务平台的设计与研发。曾任居然之家O2O大数据平台总负责人、中国银联大数据报文分析项目高级技术顾问、国美在线大数据中心高级技术工程师。

曾参与企业级项目包括大型电商网站的BI系统、数据仓库、大数据系统等设计和研发项目，金融银行类企业风险及异常交易分析项目。实施大中型企业数据项目包括居然之家、中国银联、华农保险、中国电信等超过50家客户案例。精通大数据Hadoop, Hive, HBase, Impala, Spark等组件架构与实施，精通数学模型，自主开发实现分治/覆盖的C4.5决策树、马尔科夫预测、KMeans、Apriori等模型算法程序，成功应用到电商、金融等行业。

目录: Contents 目录

前言

第1章 企业大数据战略定位 1

1.1 宏观 1

1.2 微观 4

1.2.1 资源协同 5

1.2.2 战略定位 6

1.2.3 启动契机 7

1.2.4 大数据历程 9

1.3 本章小结 12

第2章 企业大数据职能规划 13

2.1 大数据组织架构体系 13

2.1.1 大数据部门在企业中的角色 13

2.1.2 常见的大数据职能及职责 17

2.2 大数据职位构建体系 24

2.2.1 基础平台类 24

2.2.2 数据管理类 26

2.2.3 技术研发类 27

2.2.4 产品设计类 30

2.2.5 数据挖掘类 32

2.2.6 数据分析类 33

2.3 大数据制度和流程规范 35

2.3.1 制度和流程规范意义 35

2.3.2 制度和流程规范内容 35

2.3.3 制度和流程规范模板 42

2.4 本章小结 44

第3章 企业大数据解决方案 45

3.1 企业大数据解决方案实现方式 45

3.1.1 独立研发 45

3.1.2 第三方解决方案 46

3.1.3 联合开发 57

3.2 如何选择解决方案 58

3.2.1 外部环境分析 58

3.2.2 内部环境分析	59
3.2.3 需求规划分析	62
3.2.4 解决方案特性分析	63
3.2.5 解决方案费用评估	67
3.3 本章小结	70
第4章 企业大数据自主实施思路	71
4.1 制定规划原则	71
4.1.1 价值性	71
4.1.2 实时性	72
4.1.3 高效性	72
4.1.4 安全性	72
4.1.5 延展性	73
4.1.6 全局性	74
4.2 制定目标蓝图	75
4.3 制定建设目标	76
4.4 明确组织规划	78
4.4.1 组织结构设计的作用	79
4.4.2 组织结构设立的导向	79
4.4.3 组织结构的最终设立	81
4.5 设计技术方案	85
4.5.1 大数据系统建设方案	85
4.5.2 大数据系统与传统BI的融合方案	91
4.6 制定人才规划	94
4.6.1 指导思想	94
4.6.2 规划原则	94
4.6.3 核心内容	95
4.7 投入产出评估	97
4.7.1 数据投入与产出的内涵	97
4.7.2 数据投入与产出的特征	98
4.7.3 数据投入与产出的管理	99
4.8 数据风险管理	105
4.8.1 数据风险管理的概念	105
4.8.2 数据风险管理的类型	106
4.8.3 数据风险管理的原则	109
4.8.4 数据风险管理与控制	110
4.9 本章小结	114
第5章 大数据技术介绍	115
5.1 核心技术	115
5.1.1 Hadoop生态	115
5.1.2 NoSQL	142
5.1.3 实时计算	150
5.1.4 全文检索	160
5.2 相关技术	204
5.2.1 数据可视化	204
5.2.2 数据缓存	220
5.2.3 中间件	227
5.2.4 关系型数据库	236
5.2.5 数据ETL	245
5.3 大数据算法库	250
5.4 本章小结	276
第6章 大数据架构设计	277
6.1 大数据架构设计原则	277
6.2 大数据核心架构要素	279
6.3 大数据架构设计模式	284

6.4 本章小结	289
第7章 大数据技术开发	290
7.1 数据采集	290
7.1.1 批量采集	291
7.1.2 增量采集	292
7.2 数据存储	293
7.2.1 HDFS文件存储引擎	294
7.2.2 Hive数据存储引擎	295
7.2.3 HBase列式存储引擎	295
7.2.4 MySQL关系型数据存储引擎	296
7.3 多维计算	296
7.4 功能服务	299
7.5 平台管理	301
7.5.1 监控管理	301
7.5.2 调度管理	302
7.5.3 权限管理	304
7.6 应用域	307
7.7 本章小结	308
第8章 大数据工作流	309
8.1 数据源	310
8.1.1 日志/文件	310
8.1.2 数据库	310
8.1.3 网络爬虫	311
8.1.4 第三方API/合作	311
8.2 数据处理	312
8.2.1 数据质量校验	312
8.2.2 清洗转换	316
8.2.3 质量提升	320
8.2.4 数据脱敏	321
8.2.5 集成整合	323
8.3 数据存储	324
8.3.1 关系型数据库	324
8.3.2 分布式文件系统	325
8.4 数据计算	325
8.4.1 三种数据计算时效性	325
8.4.2 结构化数据计算	327
8.4.3 半/非结构化数据计算	333
8.4.4 深度挖掘学习	360
8.5 数据应用	376
8.5.1 辅助决策	376
8.5.2 数据驱动	377
8.6 数据质量管理	379
8.6.1 数据质量建设的内涵	379
8.6.2 影响数据质量的常见因素	380
8.6.3 数据质量建设的框架	381
8.7 本章小结	392
第9章 企业大数据业务应用	393
9.1 大数据应用场景概述	393
9.1.1 场景商业目的分析	394
9.1.2 场景数据来源分析	394
9.1.3 场景数据难易分析	397
9.1.4 场景应用举例	397
9.2 用户画像	407
9.2.1 业务应用背景	407

9.2.2 主要实现过程	408
9.2.3 关键应用场景	414
9.2.4 应用价值提炼	415
9.2.5 场景总结回顾	417
9.3 个性化营销	419
9.3.1 业务应用背景	419
9.3.2 主要实现过程	421
9.3.3 关键应用场景	424
9.3.4 应用价值提炼	425
9.3.5 场景总结回顾	426
9.4 精准广告	427
9.4.1 业务应用背景	427
9.4.2 主要实现过程	429
9.4.3 关键应用场景	438
9.4.4 应用价值提炼	439
9.4.5 场景总结回顾	440
9.5 征信	441
9.5.1 应用场景背景	441
9.5.2 主要实现过程	442
9.5.3 主要应用场景	447
9.5.4 应用价值提炼	449
9.5.5 场景总结回顾	449
9.6 本章小结	450
第10章 企业大数据价值评估	451
10.1 资产价值	451
10.1.1 数据规模	451
10.1.2 数据价值度	452
10.1.3 数据鲜活度	454
10.1.4 数据关联维度	454
10.1.5 数据粒度	455
10.2 业务价值	455
10.2.1 用户体验提升	455
10.2.2 运营优化	457
10.2.3 销售贡献	460
10.2.4 供应链优化	461
10.3 本章小结	462
第11章 大数据的社会价值	463
11.1 民生价值	463
11.2 政务价值	465
11.3 产业价值	468
11.4 本章小结	470
第12章 大数据当前问题及挑战	471
12.1 数据挑战	471
12.2 安全挑战	472
12.3 价值挑战	474
12.4 认知挑战	475
12.5 技术挑战	478
12.6 人才挑战	480
12.7 本章小结	481
第13章 大数据未来趋势	482
13.1 价值资产化	482
13.2 产业生态化	487
13.3 主体社会化	490
13.4 应用智能化	491

[企业大数据系统构建实战：技术、架构、实施与应用](#) [下载链接1](#)

标签

大数据

企业

实用落地

系统

佳作

计算机

数据分析

作者阵容强大

评论

这是一本真正能指导企业实现复杂的大数据系统落地的书，而不只是关注某一项技术。几位作者都是在大数据各个环节都有深厚积累和丰富实践经验的，非常互补。

工具书

比较全，但止于皮毛，有拼凑之嫌

就是一锅大杂烩

整体还算推荐：1-4章试读后觉得不错就买了；5-8章干货满满，读大数据架构及相关技术介绍很全面，不过具体的技术还是要单独去全面了解；9-13章没什么干货，大致翻一遍。

罗列齐全的工具书

这本书关于企业大数据实施方面内容真的很不错，确实是少有的技术佳作。

通俗易懂，涵盖内容广泛且实用！

内容丰富全面，既可以作为具体技术学习，也能用于工具查询。从架构、业务、技术等维度深入的介绍了大数据生态系统相关技术。本书唯一不足的地方是有个别组件介绍程度不够深入，当然，也不能要求那么完美。组件的具体应用还要根据实际业务场景来定。总之，是一本不错的的数据技术书籍。值得一看。

13个章节，492页内容，一看作者们花了许多心血，值得一读。

[企业大数据系统构建实战：技术、架构、实施与应用_下载链接1](#)

书评

如果你是学习软件开发的学生，如果你初入计算机行业，如果你是互联网大数据从业人

员，你一定要看这一本书。作者以多年经验，作为一名资深专业人士，从企业管理者的角度，对行业内部进行总结，细节方面，不仅涉及具体技术，更从整体层面给你讲述架构。文章涉及行业内具体事例，...

[企业大数据系统构建实战：技术、架构、实施与应用](#) [下载链接1](#)