

数据实践 之美

Beautiful
Data
Practices

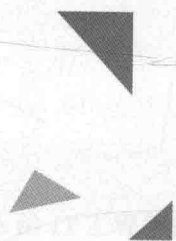
31位大数据专家的方法、技术与思想

天善智能◎著

来自全球20余家知名数据公司的31位数据专家
分享了他们在数据化思维、数据治理、
与数据可视化、数据分析与数据挖掘、
大数据应用5大主题下的最佳实践



机械工业出版社
China Machine Press



数据实践 之美

Beautiful
Data
Practices

31位大数据专家的方法、技术与思想

天善智能◎著



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

数据实践之美: 31 位大数据专家的方法、技术与思想 / 天善智能著. —北京: 机械工业出版社, 2016.12

ISBN 978-7-111-55703-6

I. 数… II. 天… III. 数据处理 IV. TP274

中国版本图书馆 CIP 数据核字 (2016) 第 305357 号

数据实践之美: 31 位大数据专家的方法、技术与思想

出版发行: 机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码: 100037)

责任编辑: 李 艺

责任校对: 董纪丽

印 刷: 北京诚信伟业印刷有限公司

版 次: 2017 年 1 月第 1 版第 1 次印刷

开 本: 186mm × 240mm 1/16

印 张: 28

书 号: ISBN 978-7-111-55703-6

定 价: 79.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88379426 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzit@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东

赞 誉

BIG DATA 给 IT 工程师们提供了新的职业发展机遇，也为各种 IT 技术提供了同台竞技的舞台，反过来因为这些人才和技术的大量涌入，“大数据文化”得以繁荣和发展！天善智能就是其中的积极贡献者。

Smartbi 副总裁 徐晶

本书不仅仅是一本案例集，还是几十位业界资深人士智慧的结晶。同一主题中从多个角度进行观点碰撞的编排方式，在市面上的数据类书籍中也是绝无仅有。向希望踏入数据领域的各位读者，强烈推荐本书。

阿里巴巴营销数据分析专家 陈丹奕

回首自己 10 年数据从业之路，从菜鸟时期看《Building the Data Warehouse》思考星型模型和雪花型模型优劣势，到现在努力践行通过 DT 解决公司降本、增效、创收、风控等业务难题，期间不知道看了多少本数据相关的书（非要量化，一整书架）。但从来没见过过一本这么多作者各自发挥各自领域特长合力完成的经验合集。作者都是各自领域经验丰富的从业人员，将自己踩过的坑、趟过的雷分享出来，看完这本书会使人了解到：BI、大数据在不同领域、不同公司的理解、定位和应用都是不一样的，会让你开阔眼界、增长见识。不管是 BI、大数据、人工智能 / 深度学习，都只是一种技术、一种生产力，目的都是为了解决问题创造价值，希望大家都能站在解决问题的角度看待这些技术。无招胜有招的高手都是阅历了百家之长！

京东集团大数据业务部架构师 李爱华

数据分析师应该是一个博学家，他们要博采众家之长，通过模仿、借鉴并加以创新，最终形成自己的一套分析路径。本书正好满足这种需求，通过 34 篇文章，几十个案例让

读者学到别人家的大数据，非常好！

数据化管理咨询顾问及培训师，《数据化管理：

洞悉零售及电子商务运营》作者 黄成明

未来，是一个数据服务的时代，未来的数据更多不是以单品的形态出现，而是融入企业和生活中的每一个环节；未来，数据将会作为一种重要的战略资产，渗透在每个行业领域。目前在数据领域，技术瓶颈、人才缺失等问题十分严重，本书在这几方面作了详尽的讲解和很多经典成功案例，值得一读。数据，正慢慢地引领着一个浪潮。

铂涛集团数据产品经理 欧锦光

这是一本链接抽象数据与现实业务的书，书中介绍的来自不同领域众多精英们对数据分析实践的宝贵经验着实令人叫绝。翻开此书，就如站在前辈们用实践堆积而成的宝藏上，弯下腰随手拾起的就是名叫“真知”的无价之宝。

电子表格应用大会主席 李奇

每一位作者都是数据行业的先行实践家，每一篇文章都是行业智慧与经验的浓缩，本书为我们带来一场使用数据技术升级各行各业的交流盛会。身处数据时代，无论是对技术、市场、产品、还是管理人员来说，都是必读的“他山之玉”。

山东财经大学数学与数量经济学院教师 郭鹏程

本书系统地介绍了数据科学在工业界的实践，把数据化思维、数据治理、BI与数据挖掘等从多个视角进行了深刻剖析。作者均为工业界的一线管理者与工程师，有深厚的数据分析和构建数据平台的功底，读者可以从书中了解到数据科学实践的前沿方法与技巧。

微软资深数据平台工程师 高彦杰

有一句话我们经常说：If you can't measure it, you can't fix it. 用数据说话，倡导用数据思维意识来推动，用数据去洞悉事物的变化，形成数据文化。本书是几十位作者在各个行业、各个技术领域数据思维的经验总结，值得大家仔细阅读。

网站分析在中国创始人，WAW中国创始人 宋星

误打误撞进入大数据投资领域3年多，经历了新奇兴奋到迷茫怯手。本书让我拨开繁杂的数据信息迷雾，从数据行业一线践行者的视角窥探行业的脉络和发展态势，更好

的把握数据项目投资的精髓和逻辑。

华澳资本合伙人 常永兵

我在发展百度慧眼业务的过程中有这样的感触：大数据的应用领域和应用价值仅仅露出了冰山之一角，然而这区区一角之地，也是一群极具科技嗅觉的数据先锋们努力打拼出来的——寻找数据、验证数据、应用数据、可持续获取和完善数据——历经层层磨难，看到一点光明，从而更加坚信大数据应用的未来。今天，很欣喜的看到本书把这些经历、经验和亮点集结成册，向全社会分享，这是一件很有价值的事情。

百度地图开放平台产品总经理 李志堂

随着数据科学的兴起，人们调侃说数据科学家已成为最抢手、最性感的职业。这凸显拥有数据分析实践能力的人才可以帮助企业进行数据资产的挖掘与利用。通过数据分析与数据挖掘推动企业的经营管理，帮助企业管理者洞察发展趋势和科学决策，这种数据价值会给企业带来商业模式上的优化，以及商业自动化上的突破，让企业保持核心竞争力。

中国传媒大学新闻学院教授，大数据挖掘与社会计算实验室主任 沈浩

数据作为一个新物种，带给人类一个全新世界，更给人类重新认识自我、改造社会的机会。数据科学作为探索数据世界奥秘的工具正在逐步被人类像材料科学一样掌握，围绕数据科学的数据思维、数据治理、数据分析与挖掘、可视化等正在迭代人类工业时代的经验思维、品质管控、材料工程、制造工艺、工业设计等知识、技能体系。领先的数据社区天善智能通过众包、众智的方式所编著的本书无疑为每位 DTer 打开一扇天窗。希望更多朋友加入天善智能社区，为 DTer 提供形式多样、营养丰富的面包。

中关村大数据产业联盟副秘书长 & TalkingData University Headmaster 陈新河

这是一本由天善社区众多大咖所共同创作的一本关于数据科学的优秀书籍。我们知道，一本技术书籍的好坏主要取决于书籍的作者质量以及体系规划情况，这本书的书籍作者都是行业内知名的专家，每一篇文章都凝聚了专业人士多年经验智慧的结晶，含金量非常高。同时，这本书采用去体系化的方式编排，围绕一个中心，从不同的视角去阐述数据科学的知识，形散而神一，是一本难得的好书，郑重推荐。

重庆韬翔网络科技有限公司（上海）董事长兼总经理、

CSDN 特邀 IT 专家、知名 IT 作家 韦玮

我们常说当下是大数据时代；我们也对各种技术字眼耳熟能详，Hadoop，Spark 等。但我们总不经意忽略，无论哪种革命，真正的源力是那些为了梦想披荆斩棘的人。本书就是由这样冲杀在数据第一线的数据先锋集体编著而成，而各位读者，你们也和他们一样，是大数据 bigbang 中的源力，因此你们在读他们的文章时一定会有所共鸣。

IBM 高级架构师 邵春宇

虽然错过了成为本书作者的荣耀，能够提前享受本书，从方法、技术、思想三省吾身，也是大美。近年来，随着移动互联网和人工智能快速发展，收集、存储和分析的数据量呈爆炸式增长、席卷神州之势，大数据时代浩浩风起波。感谢无数专家的努力，让我在大数据之巅的风波中多了一些坐标。

科大讯飞大数据专家 于俊

未来是大数据的时代。如何获取数据，分析数据和应用数据是关键！本书深入浅出，结合案例将数据实践之道娓娓道来，值得期待！

众美联商城总经理 王嘉豪

前言

朝阳门和小笼包

2016年3月5日，天善智能在北京举办“数据为王·互联网大数据沙龙”活动。在活动开始前的3月2日，我在家里收到了几本有关大数据相关的书籍，当时还没有弄清楚是怎么回事，后来打电话问梁勇（天善智能创始人之一）才知道这几本书是机械工业出版社华章分社的副总编杨福川寄过来的，代表华章支持我们的线下沙龙活动。

3月5日之后的某一天晚上，我约了杨总编出来以当面表示感谢。3月份的北京还比较冷，那两天风也比较大，我们俩约在朝阳门附近见面。见面一开口说话便发现都是湖北老乡，一个来自武汉，一个来自孝感，瞬间就感觉话题多了很多。

我们俩虽然都在北京，但是各自都有各自混吃的“地盘”，对于朝阳门附近到底哪些地方有什么好吃的一概不知。一路边走边聊，最后逛到地下一家小店。看看也到吃饭的点了，四周也没有什么合适的餐馆就只能这样将就着了。坐下之后，两人要了两份小笼包，一份粥和一份面，就坐在小饭桌上边吃边聊了。没有任何的刻意和客气，就像老朋友一样……

饭后聊了很多有关创业的话题，包括天善这个团队是如何结缘，如何辞职全职创业的。也聊到我们对大数据领域的一些认识和看法，圈内的故事，以及我们社区和社群的资源。

其间，聊到2015年我们在网上发布的19万字的商业智能BI年刊时，杨总编说：“你们社区在大数据和BI领域有不少专家资源，也积累了不少内容，为什么当时不考虑出书呢？”

“我们没有出版社的资源，也不知道出书的流程和方式”。

“需要的话，我们以后可以帮你”。

策划一本书的想法从那时就开始萌发了，就是在那个3月份略有点凉的夜晚，在朝阳门吃着小笼包结的缘。

关于本书

最初策划这本书的时候，我想过一个问题：究竟是系统性地深入讲解某个特定的与数据相关的主题（比如大数据、数据分析、商业智能等）；还是去系统化，只讲解一些与数据实践相关的经验和技巧，每篇文章相对独立。

最后我们的选择是去系统化，每一篇都是一个独特的视角，即使是同一个视角，我们也希望不同的作者能够写出不同的观点，甚至冲突。因为我们觉得这本书承载的应该更多的是启发，而不是教化。

我们期望这本书是一本数据实践的 Hacks 集合，作者由来自互联网、传统行业、大数据公司等一线的数据工作者、科学家、技术 VP、产品经理以及高级咨询顾问组成，如实地反映和总结他们在各自技术领域的真实工作经验和技巧。

2016年的4月份，我们在杨总编的帮助下，开始在天善智能社区和社群发起了写书的号召。

在写书号召发出的几天之内，我们就陆续接到来自大数据、数据分析、数据挖掘和商业智能 BI 领域 100 多位专家的报名，当然我们也力邀了一些在数据领域比较知名的朋友们加入。随后经过几轮筛选和审阅，最终有 33 篇稿件符合要求。

这 33 篇文章，我们根据主题分为了 5 个部分：

- Part 1 数据化思维，专注思维
- Part 2 数据治理，专注技术
- Part 3 BI 与数据可视化
- Part 4 数据分析与数据挖掘
- Part 5 大数据化之路，主要讲解不同行业是如何应用大数据的

所有文章独立成篇，在满足碎片化阅读的同时，也能让读者进行深入的思考和横向比较，帮助用户在碰到类似问题的时候打开思路、获得经验上的快速成长。

致谢

在此感谢各位作者艰辛的付出，同时感谢所有报名了但是因为各种原因最后未能成为本书的作者的朋友们，真诚感谢大家的努力和支持。

感谢北京铂金智慧网络科技有限公司 (Ptmind) 给本书提供的宝贵建议和支持。

我们不得不提这本书的作者们，他们是来自百度、腾讯、IBM、埃森哲、Teradata、美特斯邦威、帆软软件、永洪科技、好未来、ETHINK 象形科技、万达、乐逗游戏、况客、北京易莱信科技、奥威软件、达观数据、诸葛 IO、中科软、神策数据、融 360、奥美互动、先进数通、同程旅游等知名公司的 CEO、VP、CTO、技术总监、高级分析经理、架构师、咨询总监……都是数据领域的专家。

在本书的写作过程中，我们也得到了来自安徽象形科技 (ETHINKBI) 技术 VP 彭耀先生和北京易莱信科技的创始人 & CEO 汪尚先生的大力支持。彭耀在有关书籍的命名上给予了很多建议。汪尚在书籍文章分类上给予了我们专业的意见。

吕品 天善智能联合创始人 & 运营总监

目 录

赞誉

前言

Part 1 数据化思维 // 1

- NO.1 数据化运营的方法论体系 张子良 // 2
- NO.2 数据化营销中的“一·二·三” 叶秋萍 // 9
- NO.3 企业数据化管理之巅峰——同业对标 王卫东 // 21
- NO.4 管理数据化：柳传志 30 年管理智慧如何为我所用 袁华杰 // 33
- NO.5 数据领导力——指标体系规划与管理驾驶舱设计 刘凯 // 41

Part 2 数据治理 // 53

- NO.6 数据库开发实施工艺提升的 6 种途径 杨宝军 // 54
- NO.7 ETL 串并发数据跑批处理的理论与运用实践 何启平 // 64
- NO.8 如何高效地对复杂数据进行清洗与转换 谢佳标 // 74

Part 3 BI 与数据可视化 // 91

- NO.9 商务智能业务分析构建“5 步曲” 吕敏杰 // 92
- NO.10 构建数据体系的两个“5 步曲” 王桐 // 109
- NO.11 成功实施 BI 项目的 4 大要素 贾岩 // 119
- NO.12 Kimball 理论在 BI 项目中的应用 郭川 // 127
- NO.13 BI 数据可视化分析 SaaS 产品前瞻 吕品 // 145
- NO.14 大数据工程的系统架构设计和技术选型 韩庆安 // 153

- NO.15 数据可视化 4 步工作法 温融冰 // 164
- NO.16 如何用 R 语言对复杂数据进行可视化 谢佳标 // 169
- NO.17 新思路, 新体系: 让银行报表的 3 大痛点不再是噩梦 袁华杰 // 184
- NO.18 Cognos 在金融银行业的最佳运用 吴永帆 // 196

Part 4 数据分析与数据挖掘 // 207

- NO.19 如何做好一名商业分析师? 吴奕君 // 209
- NO.20 如何用数据驱动运营 桑文锋 // 217
- NO.21 企业增长中的精细化分析和 Growth Hacking 孔淼 // 237
- NO.22 如何基于业务实现用户行为数据产品化 吴文波 // 247
- NO.23 电商的数据化管理与运营 尚林栋 // 256
- NO.24 零售业数据分析指标的管理与应用 沈嵘 // 284
- NO.25 做好零售业数据分析必须解决的 3 个难点 邹斌 // 291
- NO.26 如何用 R 语言做量化分析 张丹 // 316
- NO.27 从 BI 到 AI, 数据分析的 4 个误区 彭耀 // 335
- NO.28 企业如何利用跨行业数据挖掘标准流程开展大数据实践 张浩彬 // 342
- NO.29 详解过程挖掘的技术和方法 汪尚 // 352
- NO.30 个性化数据挖掘的关键技术与应用实践 陈运文 // 382

Part 5 大数据化之路 // 401

- NO.31 教育行业的大数据实施路径 李宗海 // 402
- NO.32 数据科学在互联网金融中的应用 张云松 // 409
- NO.33 地理大数据驱动的智慧选址 张志成 // 419

附录 // 433

01

PART 1

数据化思维

数据化思维主要指企业在数据化运营和管理过程中运用数据的思维和方式。这部分的内容既有理论角度的阐释，也有案例角度的阐释，还有从管理、技术角度来阐释。从不同的角度我们会发现各位作者考虑问题的侧重点有所不同，打破思维僵式的最好方式莫过于此。

比如百度资深数据营销专家张子良的《数据化运营的方法论体系》和某互联网娱乐平台资深数据分析师叶秋萍的《数据化营销中的“一·二·三”》侧重于数据化运营思维的树立与拓展，而航天金穗资深数据分析师王卫东的《企业数据化管理之巅峰——同业对标》、帆软数据研究院院长兼帆软公司 CMO 袁华杰的《管理数据化：柳传志 30 年管理智慧如何为我所用》、美国注册管理会计师刘凯的《数据领导力——指标体系规划与管理驾驶舱设计》这 3 篇文章则侧重于数据化管理思维角度。

所以数据化思维首先解决的是“思维”的问题，在与数据打交道之前第一步要树立的就是如何用好数据的意识。这部分内容侧重体现了本书中“思想”这个环节。

数据化运营的方法论体系

张子良

网名胖子哥，混迹 IT 十余载，好读书，不求甚解。经史子集，诸子百家，一样不通。唯喜老庄之道，凡事随心，顺天应时，无所苛求。术业有专攻，金融和互联网领域数据方向，数据架构、数据仓库、BI 分析多有涉猎，所憾无一精通，唯有孜孜以求，继续践行。

从大数据，到互联网思维，有人迷失，有人觉醒，迷失者继续凌乱，而清醒者却开始探索其背后的本质。当喧嚣散去，山还是山，商业还是商业，本质未变，变的只是渠道和方法。互联网与大数据时代，如何回归商业的本质，数据化运营也许不是唯一的答案，但却会是一种有效的手段。

本文采用“5w+1h”方法论，试图构建数据化运营的方法论体系，带你走进数据化运营世界。全篇围绕“谁来做数据化运营？为什么需要数据化运营？数据化运营的目标和收益是什么？数据化运营如何做？何时做？从什么地方切入？”等问题一一展开，抽丝剥茧，若能给读者启示，足以快慰我心。

谁来做数据化运营 (Who)

核心观点：数据化运营的组织架构 = 顶级组织 + 自顶向下。

“名不正，则言不顺；言不顺，则事不成”，关乎数据化运营能否实现的首要问题是组织架构，作为企业的数据化运营部门，必须是顶级组织，与各业务线平级，唯有如此才能解决企业数据化运营面临的最大问题：部门沟通与协调。有了顶级的数据化运营组

组织架构，自顶向下地推动数据运营的战略和规划，才能使其变成现实。

1. 吃过螃蟹的阿里巴巴和要吃螃蟹的联想

数据化运营行业的翘楚非阿里巴巴莫属，早在2013年，阿里巴巴就通过组织架构调整，将原有的淘宝、天猫、一淘、阿里国际业务、阿里云、聚划算和阿里小企业业务7大事业群，拆分为了25个事业部，其中就包括数据平台事业部，从此数据战略正式进入阿里巴巴的公司战略。根据公开资料获取到的阿里巴巴调整后的组织架构示意如图1所示。

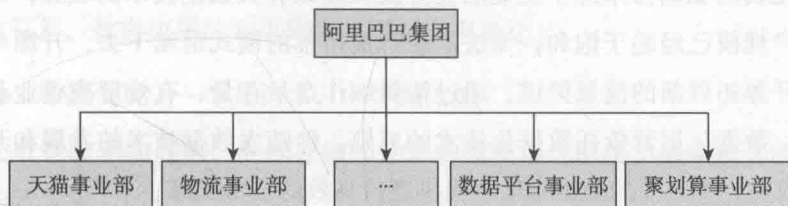


图1 阿里巴巴组织架构图 2013

2016年3月18日，联想集团宣布了新一年的组织架构调整，调整后的结构如图2所示。

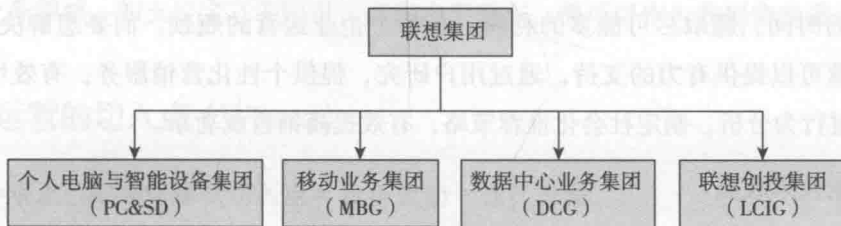


图2 联想集团组织架构 2016

图2中，笔者最为关心的是数据中心业务集团的成立，作为一个传统IT设备厂商，居然能有此决心，姑且不论其成效如何，单凭这一点，实在令人叹服，而在此背后则是数据的魅力。

2. 固若金汤的篱笆墙和拧巴的数据产品经理

业务线利益、部门利益、小群体利益是数据化运营路上必须翻过的大山，每个机构都会人为地设置各种数据互通的藩篱，好比是扎紧的篱笆墙。数据化运营最大的难题是数据打通，最浪费时间和精力的是内部关于数据互通的沟通与协调，最拧巴的莫过于数据产品经理，一方面为数据分析模型殚精竭虑，一方面巧妇难为无米之炊。

因此，执行数据化运营的团队，在企业组织架构中，必须且只能是与各业务线平级的顶级组织部门；数据化运营的推进模式必须且只能是站在整个企业级视角自顶向下的推进，而非自底向上。

数据化运营的驱动力 (Why)

核心观点：数据化运营的驱动力 = 业务运营瓶颈 + 数据技术成熟。

数据化运营的驱动力来源于企业的业务发展瓶颈和大数据技术的成熟，以互联网行业为例，用户规模已经趋于饱和，无法依靠野蛮拓荒的模式继续下去，开源与节流才是出路，通过开源拓展新的流量渠道，通过精耕细作盘活存量，有效解决企业业务发展的天花板问题。数据化运营依托数据化技术的成熟，伴随大数据技术的发展和开源分布式技术体系的成熟，为企业的数据化运营提供廉价高效的技术支撑。

1. 业务运营瓶颈

任何一个行业的发展，都有自身的规律，由盛而衰，不可避免。企业的生命周期同样要经历发展、成长、成熟、衰退等阶段。成长期抢占更多的用户和市场份额，成熟期维持更久的时间，赚取尽可能多的利润，是每个企业运营的瓶颈。而要想解决该问题，数据化运营可以提供有力的支持，通过用户研究，提供个性化营销服务，有效增加用户粘性；通过行为分析，制定社会化推荐策略，有效提高销售成功率。

2. 数据技术成熟

传统数据处理技术的瓶颈主要体现在分布式文件存储和并行计算能力上，海量数据的存储和计算，制约着企业数据应用的发展。以分布式文件系统 HDFS、并行计算框架 Map/Reduce 为代表的大数据技术的日趋成熟，企业数据化运营的技术瓶颈得以突破。同时因为 Hadoop 的开源发展模式，有效降低了企业使用大数据技术的成本。

数据化运营的目标和收益 (What)

核心观点：数据化运营的目标和收益 = 内部业务支持 + 外部数据变现。

数据化运营的目标在于改善现有业务和扩展新的领域。一方面通过数据分析，改善现有业务能力，如阿里的数据魔方；另一方面通过数据挖掘，衍生新的业务形态，如阿里的蚂蚁金服。数据化运营的收益与目标相呼应，可以体现在对内和对外两个方面，对

内输出数据服务能力，辅助改善现有业务，获取更多的收益；对外拓展新的领域，形成新的业务形态，获取数据变现收益。

1. 对内：数据让业务变得更好，却无法度量其收益

数据的价值，人人可知，却又人人不可知，原因在于当其应用于改善现有业务时，其输出为数据分析结果，而收益获取的方式却是业务推广和验证，彼此之间信息与知识单向流通，没有形成有效的数据的闭环。循环与迭代，是数据化运营的核心思想。数据团队提出分析模型，业务团队跟进验证并反馈结果；数据团队更新模型，业务团队跟进验证，如此往复，数据应用的效果则自然可以实现量化。

2. 对外：数据变现，一半是海水，一半是火焰

伴随大数据技术的成熟，数据存储和计算的瓶颈得以解决，越来越多的企业投入到数据市场的建设中，诸如大数据交易所等机构相继出现。直接的数据买卖，其实是行走在一片灰色地带，在隐私与安全的边界游走，真正能畅游者，又有几人。行业自身的差异，以及跨企业数据壁垒的牢不可破，企业要想实现数据变现很难。对外的数据变现模式有很多，但是真正能够盈利的很少。就企业而言，最好的实践则是基于现有数据，衍生新的业务领域，如支付宝之于阿里，白条之于京东，都可以作为数据变现模式的范本。

数据化运营的切入点 (Where)

核心观点：数据化运营切入点 = 业务驱动 + 迭代思维。

数据化运营的切入，可以从企业视角构建和完善企业的数据能力，即数据治理；也可以从用户视角，以业务需求为驱动，采用迭代思维，设计相关的数据产品和工具平台，满足企业的数据服务需求。数据治理是一个长周期、重投入、慢收益的模式，适合成熟稳重、有长远战略眼光的企业；业务驱动加迭代思维，强调的是快速就绪、快速响应、快速收益，比较适合创业公司和互联网企业。

1. 数据治理：徐而图之

数据治理是对企业数据资产管理行使权利和控制的活动集合（计划、监督和执行），数据治理职能指导其他数据管理职能如何执行，数据治理是在高层次上执行数据管理制度。数据治理有一套完整的知识体系，在此不赘述，有兴趣的朋友可以阅读相关专著。与数据治理相关的十大知识领域如图3所示。