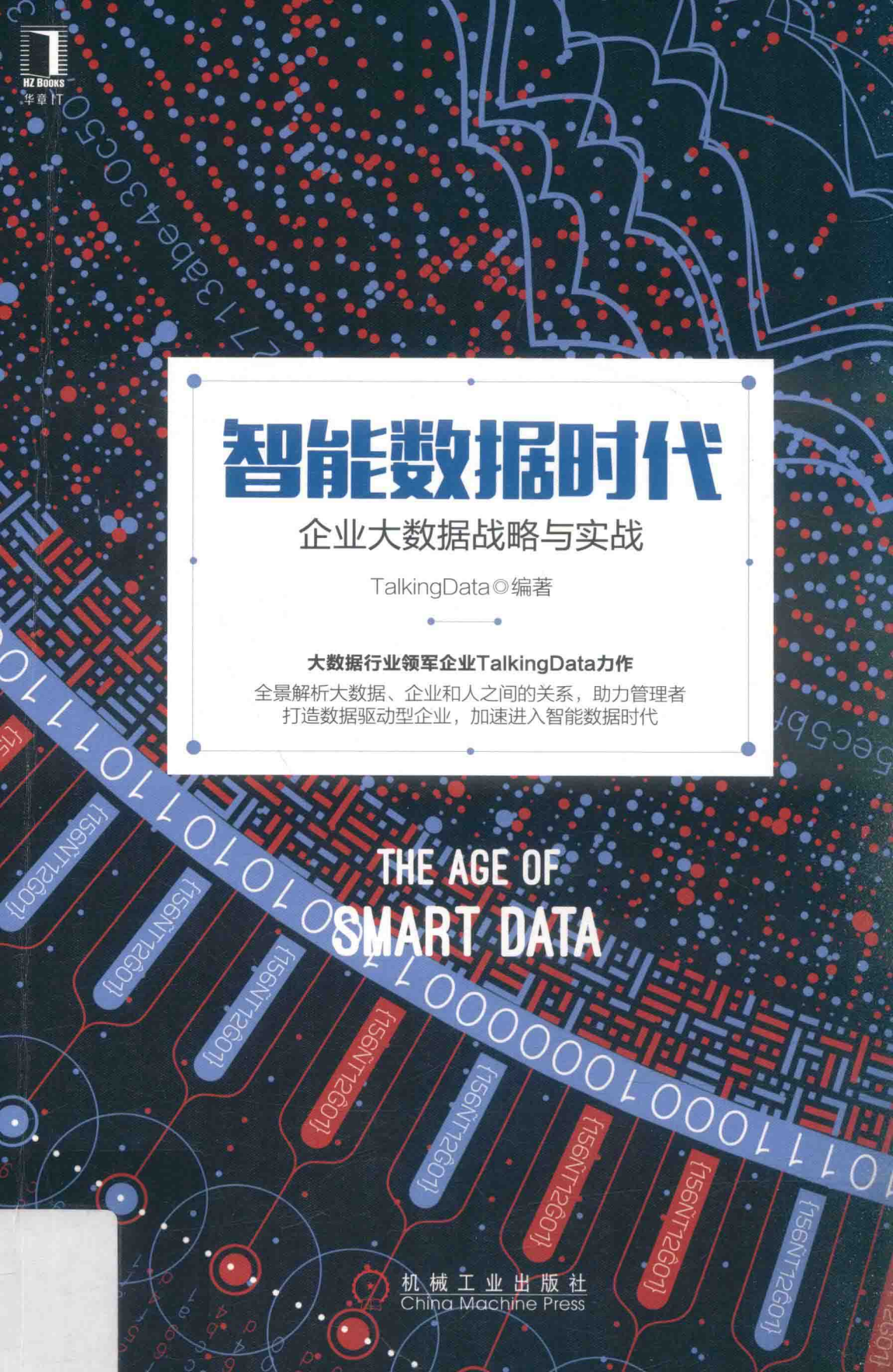
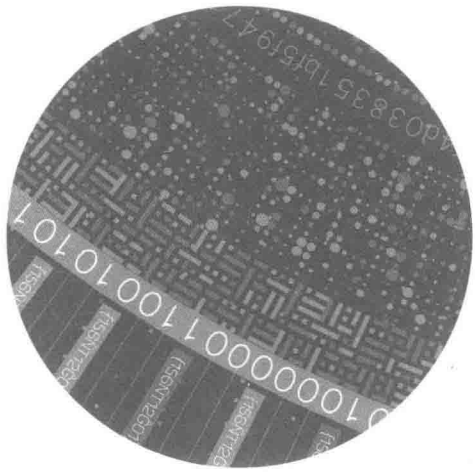

HZ BOOKS
华章IT
智能数据时代
企业大数据战略与实战
TalkingData◎编著
大数据行业领军企业TalkingData力作
全景解析大数据、企业和人之间的关系，助力管理者
打造数据驱动型企业，加速进入智能数据时代
THE AGE OF
SMART DATA
机械工业出版社
China Machine Press



智能数据时代

企业大数据战略与实战

TalkingData◎编著

THE AGE OF
SMART DATA

机械工业出版社
China Machine Press

图书在版编目（CIP）数据

智能数据时代：企业大数据战略与实战 / TalkingData 编著 . —北京：机械工业出版社，2017.5

ISBN 978-7-111-56946-6

I. 智… II. T… III. 企业管理 - 数据管理 IV. F272.7

中国版本图书馆 CIP 数据核字（2017）第 096730 号

智能数据时代：企业大数据战略与实战

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：和 静 责任校对：殷 虹

印 刷：中国电影出版社印刷厂 版 次：2017 年 6 月第 1 版第 1 次印刷

开 本：170mm × 242mm 1/16 印 张：21.75

书 号：ISBN 978-7-111-56946-6 定 价：69.00 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

前言

| PREFACE |

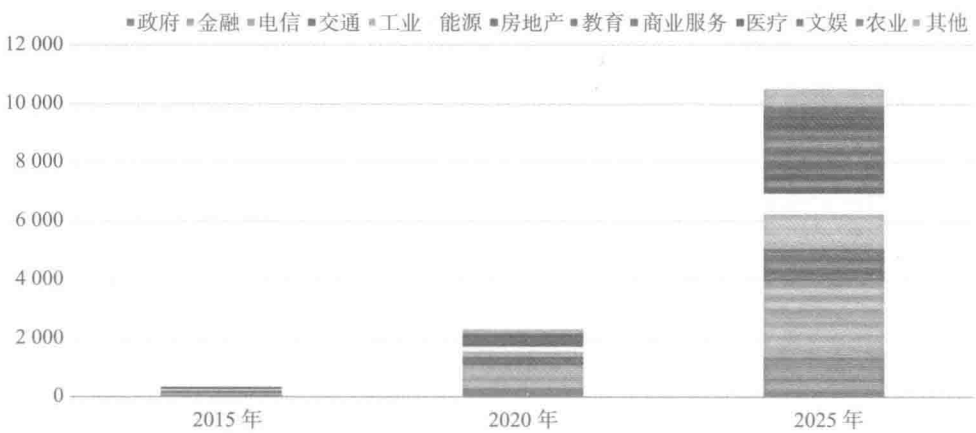
大数据这个概念自诞生以来，已经经历了几次飞跃。时至今日，大数据这个名词频繁地与人工智能、DT、预测等词汇放在一起，看上去数据的发展已经成为与科技发展甚至整个社会发展平行的存在——一切的颠覆都离不开数据。大数据是一种赋能工具，它的作用是帮助行业加速价值的流通，减少信息不对称，提高交易效率。

市面上大数据行业相关的书籍已经汗牛充栋，然而还没有这样一本书——全面地解析大数据、企业和人之间的关系，站在企业管理者的角度解答如何利用大数据加速发展、攫取更多的价值；更没有人全面告诉企业的管理者，如果想转型以适应当今智能数据时代，应该储备何种知识和人才。TalkingData 作为大数据行业的领军企业，决定写这样一本书。

竞争环境：行业快速发展，传统行业加速转型

根据 IDC 的数据显示，到 2020 年，全球大数据技术和服务市场预计将达到 589 亿美元，其中大数据基础设施占 277 亿美元，大数据软件占 159 亿美元，大数据服务占（包括专业和支持服务）153 亿美元。相比于北美等发达地区，中国大数据产业虽然年轻，但是处于快速发展期。根据 DT 大数据产业创新研究院

（DTiii）的预测，从现在到2025年，大数据产业的经济总量将呈指数级增长（如下图），覆盖的行业包括政府、金融、电信、交通、工业、能源、房地产、教育、商业服务、医疗、文娱、农业等。



2015—2025年中国大数据产业增长趋势图

注：数据来源于DT大数据产业创新研究院（DTiii，2016）。

除了飞速发展的整个行业总量之外，大数据行业本身也带有快速颠覆迭代的特征。当今社会，对传统大数据中量的需求已经很容易达到，大数据的竞争转向了数据质量。那些深入在各种行业情境中、非结构化的、与业务流程直接相关的数据，成为高价值的数据类型。只有将这部分数据挖掘出来，企业才有可能基于自己的业务进行分析甚至预测。因此，大数据时代进入了一个新的纪元——智能数据时代。

数据和人工智能是智能数据时代的鲜明特征，但是只有数据和人工智能依然不足，还需要人类智慧的参与。数据、人工智能和人类智慧，成为智能数据时代的三大要素。

数据的积累可以为人类提供更多更细的洞察分析，人类经验得以增强，人类智慧得以增长。比如，通过更多来自于手机的用户行为分析，企业可以对自己的用户有更多了解，包括他们的生活喜好、消费习惯等，以此产生更多的营销机会。人工智能本身也需要人类智慧的介入，以引导人工智能的方向，提高人工智能的

效率。比如，AlphaGo 也需要不断地与人类围棋高手对战，依靠人类智慧的辅助，才能持续提升棋力。

缺乏人类智慧的持续介入，人工智能对数据的加成作用会随着数据的变化逐步弱化甚至失效；缺乏人工智能，人类无法依靠自身处理如此复杂而且快速变化的数据；缺乏数据，人工智能无法存在，人类智慧的积累也会放缓。数据、人工智能和人类智慧互相促进，组成一个正向的循环。比如情景感知领域，基于手机上体现姿态动作的传感器数据，经过人工智能的算法，可以判断手机用户的动作和姿态（包括走路、骑车、驾驶等）。如果判断不够准确，就需要人工介入，对数据再进行整理和增强、对算法进行优化，直到结果达到可用的程度。同时，具有情景感知能力的手机，可以给应用开发者提供更多的应用场景和体验，比如运动健身、金融风控、物流管理、娱乐体验等，相应地也会产生更多的数据——这些新的数据让人类智慧更快积累，也让人工智能更加强大。比如，通过情景感知数据，发现绝大部分用户在使用 App 的时候手机都是处于手持状态，那么非手持状态的使用场景是否意味着更大的金融风险？

人工智能和人类智慧，让数据岂止“大”？智能数据时代的三大要素聚合裂变，已经产生难以想象的价值。

现实差距：人才缺口

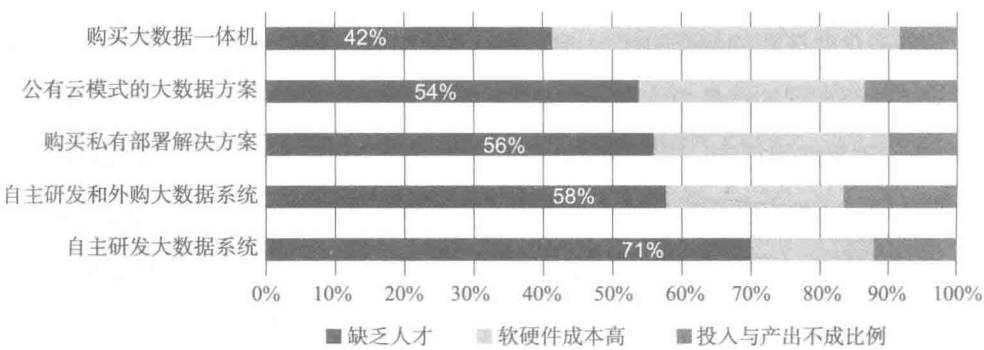
很多企业在智能数据时代举步维艰，但是也有一些新的企业脱颖而出，利用数据来增强自己的竞争力，在各个领域对传统企业形成巨大的冲击。据 A16Z 的 2016 移动互联网报告，以 GAFA（谷歌、亚马逊、Facebook、苹果）为代表的数据巨头，已经在数据和技术领域建立起牢固的优势，在年收入规模上比 Wintel 同盟要大 3 倍。反过来，它们也在通过数据和技术改变传统行业的形态和模式，包括零售、媒体发行、汽车等。

这些新的数据和技术的先锋具有一些共同的特征：实现了业务数据化和数据资产化，能够用数据来驱动场景化的应用，高效地探索和转化商业价值。这样的

企业，已经拥有数据驱动的文化，我们叫做智能企业（Smart Enterprise）：

1. 具有灵活的技术平台和数据科学能力，能支撑足够大的数据量级、足够多的数据维度、足够复杂的数据类型、足够灵活的数据格式、足够低的数据洞察延时等，提高各种数据应用场景的交付效率。
2. 具有统一的数据管理策略，以管理跨企业的、一致的数据视图，能高效地汇聚数据（包括自有数据和第三方数据），也能高效地输出数据和数据服务。
3. 具有端到端的数据工程能力，以支撑业务线的可管理的数据运营，形成数据闭环和持续的业务优化。

若要转型为智能企业，人的智慧尤为重要，因此对于无论是大数据企业还是亟待转型的传统企业来说，都提出了人才的类型、数量和知识结构的严苛挑战。但是一个严酷的现实是，现在的人才储备是远远跟不上行业需求的。从下图我们可以看出，在搭建大数据平台应用来应对转型的企业所遇到的痛点中，有一半多的原因是卡在了人才不足这个关口上。根据 DT 大数据产业创新研究院（DTiii）资料显示，到 2025 年，中国的大数据人才缺口将高达 200 万。这不仅仅是在中国，在美国问题同样严重。McKinsey 预测：至 2018 年，美国将有 60% 的组织设置首席数据官（CDO），需要 400 万名具备大数据分析能力的经理和分析师，人才缺口将达到 150 万；未来八年将有 19% 的大数据人才需求增长。



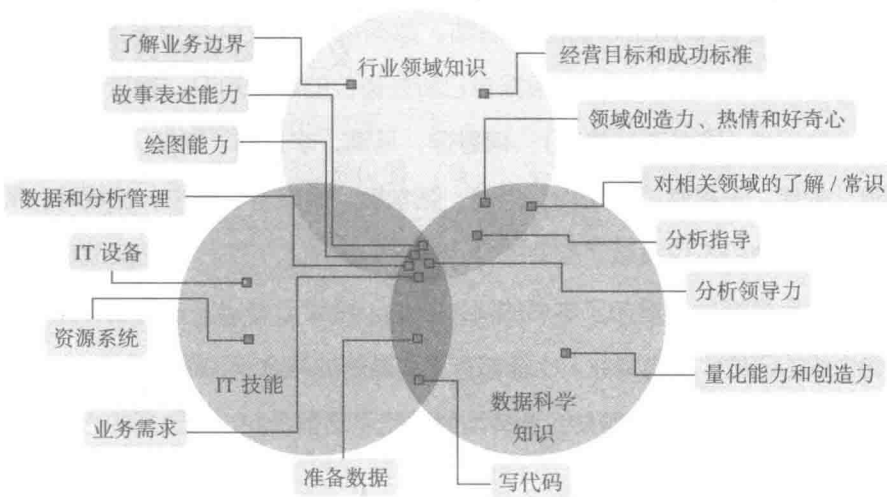
塔建大数据平台应用方式中遇到的主要困难

注：资料来源于 CSDN《2015 年中国软件开发者白皮书》。

从上面的数字我们可以看出，仅仅靠大学设立相关学位专业和社会上多开几个专家培训课程是无法弥补如此大的人才缺口的。智能数据时代大数据人才的培训，需要严谨的知识架构设计、先进的技术工具辅助以及前沿的行业最佳实践的熏陶。这个时候，仅仅靠学术界和社会培训的努力是不够的，身为一线离炮火最近的组织，大数据企业本身应当积极投入到为社会进行人才建设和储备的事业中去。

智能数据时代人才的知识架构

在智能数据时代，对于人才的知识储备的要求是综合的。如下图所示，TalkingData 认为智能企业所需的人才应当具备行业领域知识、IT 技能和数据科学知识。团队内部的人才组合必须能够合理覆盖这三个知识领域，艰巨的任务才能迎刃而解。



智能企业所需人才的知识结构

TalkingData 认为，一家企业如果要向智能企业转型，必须具有以下几种类型的人才：

1. 数据架构师。负责制定数据构架管理政策及指南，解决数据管理组织之间的争议问题；制定数据标准、应用标准、运维标准，设计模型管理流程，

整理数据需求并为其他类型的数据人员提供支持。

2. 数据工程师。大数据工程解决的是海量数据（起码在 T 级别以上）的设计、部署、存储和计算需求等方面的问题。在当今，大数据工程师要设计和部署的系统往往都是消费者和内部工作人员直接使用的应用程序。简而言之，大数据工程涉及系统的设计、部署和实施。
3. 数据分析师 / 数据科学家。大数据分析则是处理大数据工程师设计的系统上的大量数据，它涉及分析趋势、模式以及开发各种分类和预测系统。简而言之，大数据分析涉及大数据的高级计算（统计、建模预测等）。
4. 数据产品经理。能评估和洞察数据价值、分析问题并快速制定落地策略，基于数据设计商业化逻辑以及关键指标，能灵活使用各类数据工具并同时熟知项目的流程管理、体系管理、人员激励等。

本书将从一个管理者的视角，从大数据的基本概念开始，循序渐进地介绍相关工具、企业数据工程的主要活动、数据团队建设以及相关的管理支持。TalkingData 的十余位一线员工根据自己所在岗位的经验知识，参与了全书的编写工作，他们是（排名不分先后）：何香萍，马斋，李正伟，杨慧，王俊，何坤，孔元明，姜伟，王福胜，潘松柏，卢健，张学敏，曾晓春，张宁，徐岷峰，周海鹏，吕博卿。他们不仅贡献了专业的要点知识，也融入了工作中的实战经验，知识点与案例反复穿插，增加了本书的实践价值。希望通过此书，管理者能够建立起智能企业的定位，业内专业人士能够有更清晰的战略全景和逻辑脉络。

TalkingData，致力于帮助企业转型为数据驱动型企业，用数据的心智去超越未来。

杨慧

2017 年 4 月 2 日于北京

（TalkingData CEO 助理，中国人民大学商学院博士，

香港中文大学管理学系博士后）

目录

| CONTENTS |

第三篇　数据管理

第六篇　构筑数据驱动型企业

第七篇　大数据实战

第一篇

| PART 1 |

大数据基础知识