

TURING

한빛미디어
Hanbit Media, Inc.

BIG

走进大数据

组织如何推进大数据战略

[韩] 张东麟 著 武传海 译

DATA

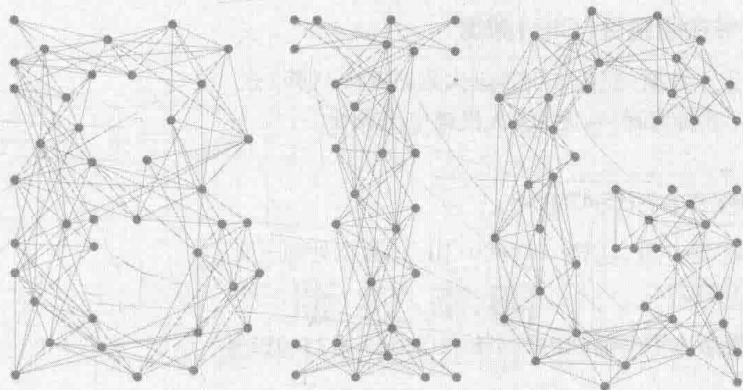


中国工信出版集团



人民邮电出版社
POSTS & TELECOM PRESS

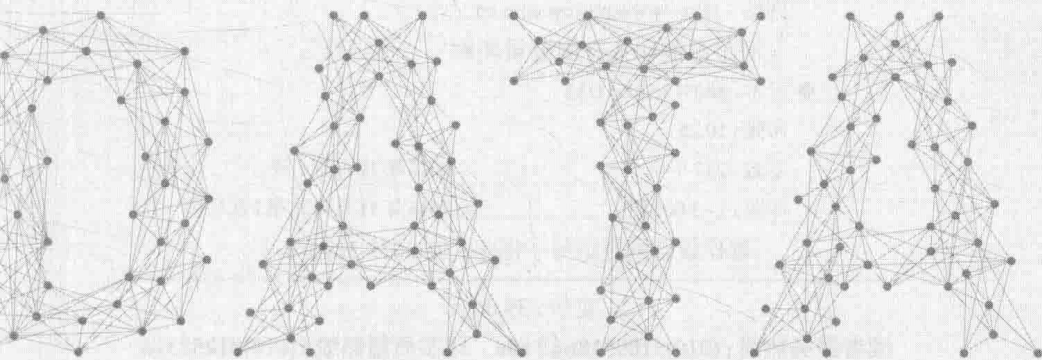
TURING



走进大数据

组织如何推进大数据战略

[韩] 张东麟 著 武传海 译



人民邮电出版社

北京

图书在版编目(CIP)数据

走进大数据：组织如何推进大数据战略 / (韩) 张东麟著；武传海译. -- 北京：人民邮电出版社，2016.11

ISBN 978-7-115-43805-8

I. ①走… II. ①张… ②武… III. ①数据处理—研究 IV. ①TP274

中国版本图书馆CIP数据核字(2016)第254059号

内 容 提 要

本书从最切实可行的大数据方案选择方法着手，面向不懂大数据相关概念、不知道如何将大数据项目引入何种领域的IT团队和实际经营团队，以自己的实际经验为基础，创立并介绍了大数据项目构建方法。帮助需要引入大数据项目的所有企业解决核心课题，以顺利开展大数据项目，并避免运行过程中出现过多失误，实现正确的策划和执行。书中收录了大数据分析与研究案例、企业引入大数据项目时发生的实际问题、策划及运营过程、实际分析时需要的技巧等，利用清晰的图表和示意图直观反映了核心法则和原理。

◆ 著 [韩] 张东麟

译 武传海

责任编辑 陈曦

责任印制 彭志环

◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路11号

邮编 100164 电子邮件 315@ptpress.com.cn

网址 <http://www.ptpress.com.cn>

三河市海波印务有限公司印刷

◆ 开本：880×1230 1/32

印张：10.25

字数：217千字

2016年11月第1版

印数：1-3 000册

2016年11月河北第1次印刷

著作权合同登记号 图字：01-2015-3696号

定价：39.00元

读者服务热线：(010)51095186转600 印装质量热线：(010)81055316

反盗版热线：(010)81055315

广告经营许可证：京东工商广字第8052号

站在巨人的肩上

Standing on Shoulders of Giants



iTuring.cn

版权声明

Original Title: 빅데이터로 일하는 기술

Big data project from A to Z by Don Chang / 张东麟

Copyright © 2014 HANBIT MEDIA, INC

All rights reserved.

Simplified Chinese translation Copyright © 2016 by POSTS &
TELECOM PRESS

Simplified Chinese translation Copyright is arranged with Proprietor
through Eric Yang Agency

如何引入大数据思维、构建大数据项目以及应用大数据战略

我在思考“企业、政府、公共机构如何引入大数据思维”这一难题的过程中写成了本书。谈论“大数据”的人很多，但一提到实际应用，大部分人都置身云里雾里，摸不着头脑。对“大数据”的赞美之词并不会为想引入、构建大数据项目的人提供任何帮助。即使阅读相关书籍并向那些具有大数据项目构建经验的人请教，也会感到一头雾水，茫然不知所措。随着学习的深入，人们会感到“大数据”整体概念越来越难。

“大数据”整体概念为什么有难度

因为“大数据”的基本概念本身就具有一定难度，它具有多种形态，是多种专业领域的大数据技术与各种 IT 技术相融合的典型代表。

由于“大数据”概念涵盖多种专业领域，即使是我们所说的“大数据专家”也只能从某个视角对其进行评价，而无法兼顾所有领域。因此，人们看到的“大数据”往往侧重某一领域。

何为“大数据”

首先，“大数据”概念始于IT行业，它发端于分布式系统（Hadoop）、NoSQL等技术。但即使如此，只从IT层面谈论“大数据”概念仍然存在着一定局限性，IT只是用于存储、处理大数据的一种技术而已。那么，融合于IT的“大数据”技术是什么呢？大数据不仅具备庞大的数据信息量，而且生成速度非常快，数据种类繁多，比如未经处理的图像、视频、位置信息和文本等。

“大数据”概念包含什么内容

根据自身包含的内容不同，“大数据”的具体概念也各不相同。依据不同行业，“大数据”可以划分为金融、物流、通信、制造和公共事业等种类，也可以依据数据来源（如智能手机、传感器和日志等）进行划分。比如，从汽车传感器获取的数据与从工厂测量设备获取的传感数据是不同的。同样，根据格式、属性、领域和来源等的不同，分析大数据的方法也多种多样。

但更重要的是引入大数据思维的目的，即我们想通过大数据分析获得什么。当然，大数据分析的目标也是非常多样的。

哪些大数据项目适合自己的企业

对于如此多样且瞬息万变的大数据概念，若只从一个领域谈论，那么无论采用何种方式，我们看到的也不过是冰山一角。若采用常规存取方法引入大数据，或者照搬他人的做法，只会招致失败，因为他人采用的数据与我们企业自身拥有的数据存在巨大差异。因此，在种类如此繁多的大数据概念中，必须找到只适合自己或自身企业的大数据。

如上所述，大数据概念的多样性也意味着获取大数据的人非常多样化。企业的情况也是一样，它们会根据自身背景与目的获取大数据。因此，想引入大数据项目的企业必须进行沟通和理解，为达成不同访问者的不同目的而努力。这也是“大数据”思维的基本常识。

如何获取大数据

目前，想要引入大数据项目的企业在成立工作小组后，通常会将其纳入 IT 部门，其他相关部门则几乎不会提供帮助。如果 IT 部门负责大数据工作组的工作，其他部门会积极配合吗？他们相互间能够很好地沟通理解吗？每个企业在沟通上都会出现这样的问题，这些问题会导致企业大数据战略的失败。

大数据战略可以应用在何处

这是实施大数据战略时一定会产生的疑问。在我访问的数百家

企业中，每当谈论“大数据”战略时，企业负责人总是会提出相同的问题：

“要将大数据战略引入我们企业的哪些部门呢？”

为什么所有企业都会问这个问题呢？其实这个问题应该由企业在职员工回答，但他们的回答都很一致：

“我对大数据战略一无所知，怎么会知道要把它应用在哪里呢？”

话虽如此，但 IT 组不懂业务，业务组不懂 IT 技术，那究竟应该如何解决这个问题呢？这是所有要引入大数据项目的企业都要面对的问题。只有解决了这个问题，才能开始实施大数据战略。这个问题也困扰了我很长时间，所以我将书中为大家展示解决这一难题的方法——大数据课题实施方法。

如何推进大数据项目的实施

即使企业历经周折成功启动了大数据项目，其中的风险也并未因此而消除。由于“大数据”概念涵盖多个专业领域，实施时必然要由多个领域的人员共同进行，那么问题在哪里呢？显而易见，不同领域的人在相互沟通和理解的过程中会产生诸多问题。大数据项目的实施要经过哪些过程与阶段才能达到预期目标呢？这也是我一直关注的问题。我曾经在大数据项目实施过程中见证了太多的错误计划与推进过程，以这些经验为基础，我总结出了书中提出的大数据项目构建方法。

如何灵活应用大数据分析结果？

开展大数据项目后还有最后一个陷阱，就是分析大数据后得到的结果没有应用于实际业务，而仅用作参考。此时，我们应该重新思考引入大数据思维的目标。为什么要耗费资源构建大数据系统，又要克服困难进行大数据分析呢？难道只是为了将大数据分析结果交给企业管理层，并听他们说“谢谢，我会参考的”吗？大数据终究只是大数据，实际决策时还是要依靠管理者的经验与直觉，而我们引入大数据思维的目的就是“依据数据进行科学决策”。如果只想将大数据当作决策时的参考，那为什么要引入大数据思维呢？从根本上讲，这种出发点就是错误的。要想解决这个问题，必须改变企业的决策文化。否则，几年之后就会意识到这样的大数据系统毫无意义。

希望可以解决大数据项目实施中的难题提供一些指导

有人将“大数据”称为“21世纪的石油”，我觉得这是非常好的比喻。但要想获得这样的效果，企业必须克服数不尽的难关。谨以此书献给那些即将踏上“大数据”这一艰难旅程的朋友，他们将要经历的艰辛于我而言仍然历历在目。作为过来人，我希望将我遇到的诸多困难提前与大家分享。希望各位在遥远而充满艰辛的旅途中不要放弃或误入歧途，顺利征服“大数据”这座高山，我在此为你们加油鼓劲。

本书读者

本书从“大数据”概念入手，内容涉及大数据项目实施中的问题及相应的解决方案，以及为了选择适合自己企业的大数据项目而进行的分析模型与数据系统的构建和设计方案等。以下几类读者通过本书将会在大数据项目实施与应用等方面得到实质性的帮助。



想引入大数据思维的企业或公共机构管理人员及在职负责人

本书是为那些想实施大数据战略的企业或公共机构管理人员而写的，所以不仅适合企业的电算部门，同样适合企业决策人及在职负责人。



策划、统计、分析、营销、宣传、商务和 IT 开发领域的业务负责人

从大数据思维的引入到实际应用，每个部门和每个级别的员工都要有一致的理解。本书有助于使参与大数据项目的所有成员达成共识，因为这种常识性的结论与大数据战略最终能否成功引入和应用存在着直接的联系。



学习大数据技术的大学生

正在学习大数据技术的大学生，通过本书可以感受到理论上的大数据技术与现实中困扰企业的大数据问题之间的差异；对于那些将来想从事大数据相关工作的大学生，本书可以为其指明今后的学习方向以及要做的准备。



想学习各种大数据系统应用案例的朋友

本书内容涉及多种大数据系统应用案例，包括最新 IT 趋势、大数据系统应用案例分析以及社交媒体分析等，可以为那些对大数据系统应用案例及应用方案感兴趣的读者提供帮助。

本书结构

第 1 章 IT 趋势与数字商务

这一章主要介绍当前 IT 趋势与新兴的数字商务。无论多么成功的跨国企业，如果跟不上 IT 发展的趋势，将很难生存。这一章通过创新成功的企业与失败企业的案例，讲解企业创新过程中如何应用大数据。

第 2 章 大数据分析趣味实例

这一章从大数据背景知识开始，为读者讲解“大数据”相关的多种概念。虽然听说过许多大数据系统应用实例，但大部分人仍然不知道“大数据”如何对自己的工作产生帮助。因此，这一章将通过使用免费的大数据系统应用工具对实际案例进行分析。

第 3 章 分析广泛应用的社交媒体

社交媒体的巨大力量使得对其的分析也变得格外重要。对社交媒体进行分析时，需要非常仔细，这一章将讲解分析社交媒体时的重要检查点、分析过程以及灵活应用分析结果的方案。

第 4 章 引入大数据项目时的常见问题

从引入大数据项目起，无论哪家企业都会遇到一些相同的根本性问题，解决这些问题是相当困难的。企业管理层、业务负责人和财务部门对大数据概念的不同理解会使大数据方案的实施陷入困境，这一章将介绍解决这些问题的方法。

第 5 章 大数据方案选择方法

所有大数据项目负责人都会问一个相同的问题：“我们企业应该在哪些领域引入大数据项目呢？”这一章将介绍如何选择符合自己企业的大数据方案。

第 6 章 具有竞争力的大数据分析

这一章展示大数据分析与传统统计分析不同点的同时，从擅长大数据分析的 Netflix 和亚马逊的具体案例入手，对大数据分析工具、数据来源以及基于云计算的大数据分析方法进行说明。

第 7 章 大数据技术与 IT 技术

大数据技术与 IT 技术之间有着不可分割的联系，这一章首先对大数据技术与传统 DW/BI 技术的不同点进行分析，针对如何将大数据技

术灵活运用于传统 DW/BI 中的方案进行探讨和思考。此外，对大数据技术的体系结构、各组成部分的作用、实际产品与解决方案以及由韩国开发并广泛应用于世界的各种大数据工具也将进行详细介绍。

第 8 章 大数据项目构建方法

这一章从大数据系统应用角度分析构建大数据项目的方法，同时结合第 5 章提到的大数据方案选择方法。

附录

我参加了 2014 年 10 月 15 日于美国纽约举办的世界最大的大数据活动 Strata Hadoop World，以及 2014 年 10 月 19 日在纳什维尔召开的 Teradata PARTNERS Conference，并对所有会议内容进行了详细记录。与本书正文内容相结合，希望这个会议记录能够帮助各位更好地把握大数据技术的全球化趋势。

大数据小问答

在各章结尾对全章内容进行总结，并以测验题形式为读者提示必须牢记的重要事项。虽然这些测验题并不难，但对学习“大数据”相关知识的朋友非常有帮助。

站在巨人的肩上
Standing on Shoulders of Giants



iTuring.cn