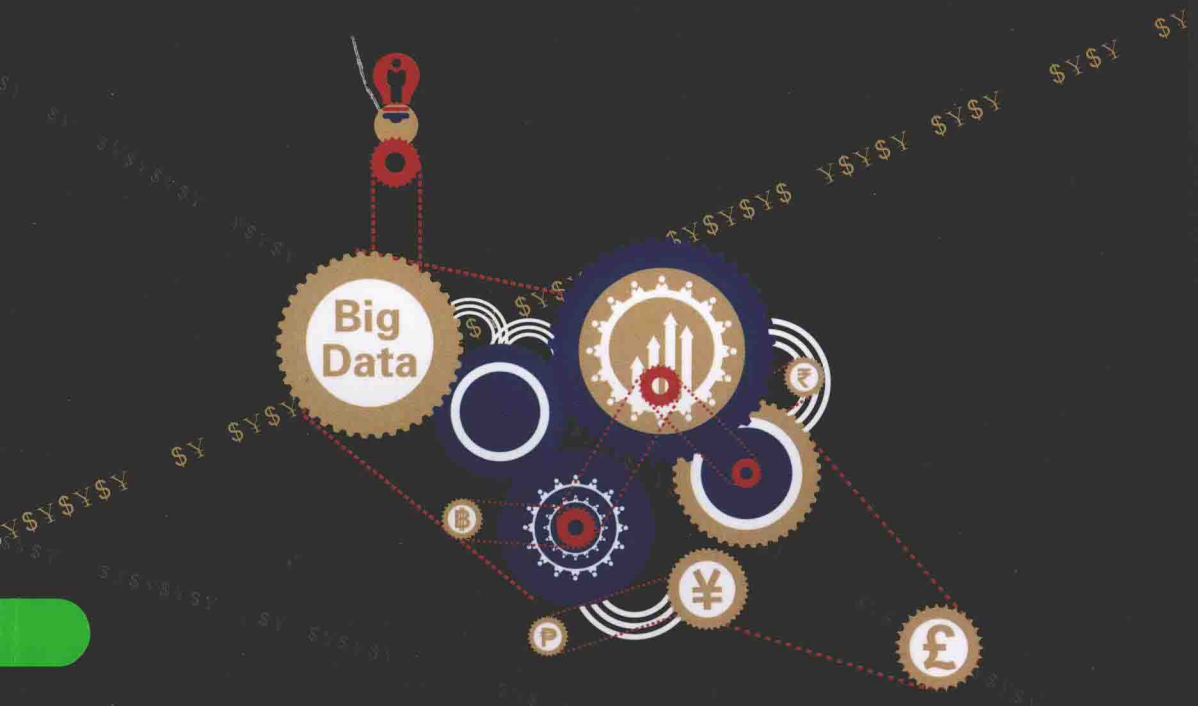
大数据丛书
十二五国家重点图书出版规划项目

Broadview
www.broadview.com.cn

金融大数据

战略规划与实践指南

陈利强 梁如见 张新宇 编著



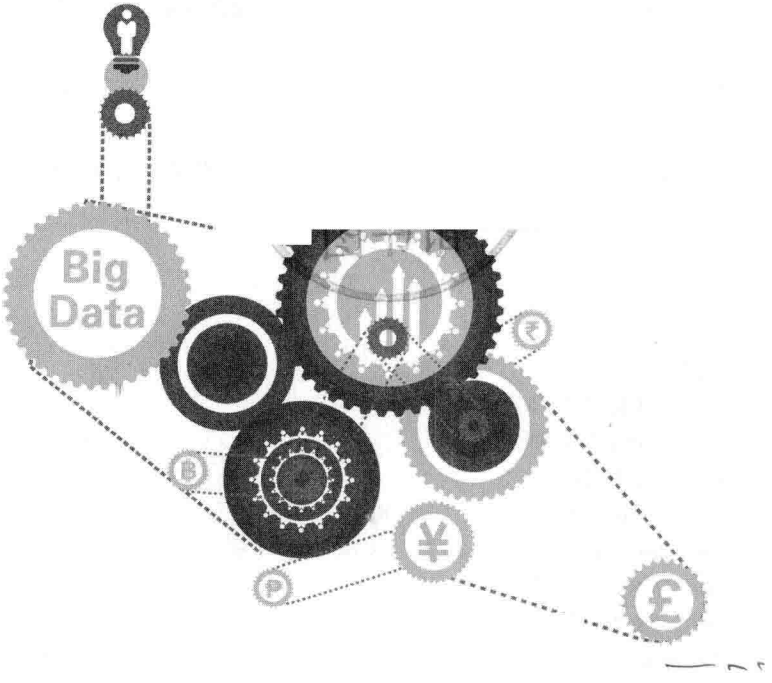
中国工信出版集团
電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

大数据丛书
十二五国家重点图书出版规划项目

金融大数据

战略规划与实践指南

陈利强 梁如见 张新宇 编著



電子工業出版社
Publishing House of Electronics Industry
北京·BEIJING

内 容 简 介

大数据与金融企业之间的关系可以用“势在必行”、“得天独厚”来形容。金融企业内部积累了丰富的用户数据和交易数据，是企业数据资产构成的核心内容；大数据技术的出现使企业可用的数据资产得到极大的扩展。本书分为三篇，分别从金融企业的大数据战略规划、场景实例以及常见问题与应对机制三个方面对金融企业在大数据时代面临的机遇与挑战做了深刻的分析，并给出了具体的实践思路。

本书提出了金融四大要素人、资金、交易和环境之间在互联网时代的笛卡尔乘积关系的重要观点，并以此作为大数据思维的重要出发点来指导金融企业的转型之路。

本书适合大数据、金融领域的相关从业者阅读，同时也是相关企业管理人员厘清思路、洞察数据和金融本质的一本不可多得的参考著作。

图书在版编目（CIP）数据

金融大数据：战略规划与实践指南 / 陈利强，梁如见，张新宇编著. —北京：电子工业出版社，2015.8
（大数据丛书）
ISBN 978-7-121-26399-6

Ⅰ. ①金… Ⅱ. ①陈… ②梁… ③张… Ⅲ. ①金融－数据处理 Ⅳ. ①F830.41

中国版本图书馆 CIP 数据核字(2015)第 138283 号

策划编辑：刘　皎
责任编辑：徐津平
印　　刷：北京天来印务有限公司
装　　订：北京天来印务有限公司
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　邮编：100036
开　　本：720×1000　1/16　印张：16.25　字数：360 千字
版　　次：2015 年 8 月第 1 版
印　　次：2015 年 8 月第 1 次印刷
定　　价：79.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：（010）88258888。

前言

深入思考金融本源

金融行业是人类有文明史以来最古老的行业之一，也是高度重视信息化建设的行业之一。

金融企业内部积累了丰富的用户数据和交易数据，是企业数据资产构成的核心内容。大数据技术的出现使企业可用的数据资产得到极大的扩展，曾经无法利用的半结构化、非结构化数据正在成为企业数据资产的重要组成部分，企业可用的数据范围、数据类型、数据时效等发生了巨大变化，数据范围从企业内部扩展到外部，数据类型从传统的结构化数据扩展为社交数据、流数据、地理空间数据、传感器数据等，数据时效从稳定的静态数据扩展为实时动态数据，这大大拓宽了金融行业数据资产的范畴和价值。

大数据时代

在大数据时代，数据成为企业的重要生产要素和战略资产，相当于农业时代的人口土地，工业时代的钢铁石油一样，拥有大量数据资产的企业将会在竞争中占有优势。维克托·迈克·舍恩伯格在《大数据时代》中提出："大数据是人们获得新的认知、创造新的价值的源泉；大数据还是改变市场、组织机构，以及政府与公民关系的方法。数据已经成为了一种商业资本，一项重要的经济投入，可以创造新的经济利益。事实上，一旦思维转变过来，数据就能被巧妙地用来激发新产品和新型服务。"数据中蕴藏的能量和价值，正在通过帮助企业了解客户需求、提供个性产品、创新优化服务、实现以产品为中心向以客

户为中心变革而显现出来。

数据资产化

数据正在成为企业的核心资产，并将深刻影响企业的业务模式，甚至重构其文化和组织。企业需要重新盘点数据资产，重新认识、发现和梳理企业的可用数据，只有准确把握自身数据状况才能发挥其应有的价值，最大化地转化为企业价值，这是制定大数据战略的前提和基础。同时，企业也需要结合业务发展需要，有意识地寻找新数据，为新数据的产生创造条件，打造支撑业务持续发展的数据生态环境。

金融大数据

选择金融与大数据作为题目，源于个人对两个行业的热爱。无论金融行业，还是当今以大数据、移动互联网为代表的信息技术，都曾经和正在很大程度地提升着我们的生活质量和幸福指数。不可否认，众多行业都受到了大数据应用带来的冲击，金融行业可以说是首当其冲。如果想要真正分析和解决问题，明确问题起源是最为重要的。我们需要思考，金融行业与大数据的结合为什么是甚至可以说一定是如此地不可阻挡？只有如此，金融才俊们才能够有明亮的眼睛，用敏捷的身手，抓住机遇，应对挑战。

金融的本源

笔者认为，造成一系列问题的根源，是我们对“金融”的本源的理解出现了偏差。金融，顾名思义，主营资金融通，在金钱由财富的象征变成了资源的载体后，可以通过交汇融合、便捷流通，产生更大的价值。也就是说，资金的价值是在流动过程中产生的。众所周知，资金的流动带有风险性，金融的诞生便是为了承载这些风险。再深入一步，为何资金流动中存在着风险呢？风险的根本成因，是资金的所有权与使用权的分离，即拥有资金所有权的甲方让渡了使用权，由乙方来使用这笔资金，乙方将利息作为收益给予甲方。横跨在甲乙双方之间的巨大鸿沟，便是“信用”。传统的金融行业，是利用经营许可牌照，在固定的银行或交易所环境中，由银行作为信用验证中介方，以抵押贷款、担保信托等金融产品形式，控制了风险，完成了资金的所有权与使用权的分离。

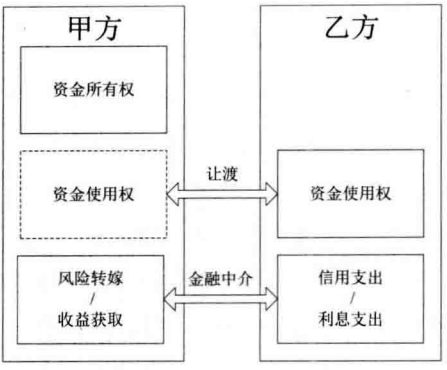


图 0-1　金融本源示意图

金融四要素

在当今社会，理解金融需要理解其四大要素：人（People）、资金（Capital）、交易（Transaction）、环境（Environment）。面对“金融”，我们在过往中往往只看到了资“金”，在互联网时代，更要看到“融”通：资金的背后是行为，行为的控制主体是人，这些人在某些场所中完成行为。近几年，金融行业在大力发展网上银行、自助银行等，走金融互联网的道路，实际上主要是解决了最后一个“环境”问题，也就是交易场所由固定的物理环境扩展到虚拟世界中。对其他三个要素的忽视，造成了诸多的问题。例如，互联网的发展，已经将“人”由法人和自然人延伸出了虚拟人，“资金”的流动速度具备了更快速更透明的基础，交易“行为”的多样化产生供应链融资等新需求。而这四个要素合在一起时，发生的并非简单的加法，而是 P(eople)*C(apital)*T(transaction)*E(nvironment)笛卡儿乘积的关系。这四者的关系是本书诸多论点的核心，也是大数据思维的重要出发点。

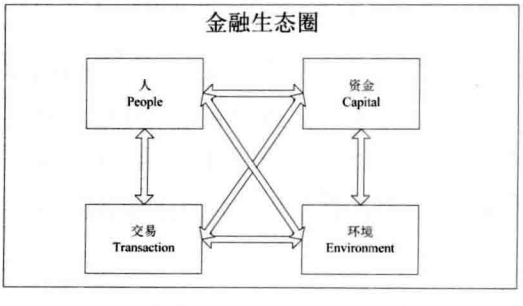


图 0-2　金融四要素示意图

互联网金融

金融进入 Web 2.0 时代以来，与互联网的特征进行了深度结合。首先，能够将“人”以社区的形式组织在一起，通过在社区里的表现建立起各类标签，积累其信用等级；其次，“资金”在已经虚拟化的基础上更进一步，产生了大量的互联网交易平台，其流转速度更加迅捷；最后，每一个业务“行为”都被完整如实地记录，供需双方虽不见面，却也已如同透明人一样地接头。不容忽视的是，每一个存活至今的互联网企业，都是在竞争中一路拼杀过来的，天生具有风险偏好和竞争力，金融行业的利润自然是互联网追逐的对象，这就是金融行业受冲击榜排名第一的原因。

解铃还需系铃人

于是，无论是金融互联网还是互联网金融，传统金融行业在大数据面前遭受的压力不可避免了。幸好，虽然互联网造成了上面的问题，它又是解决这些问题的最好途径。在信息世界里，PCTE 四个要素都分别被还原成了事物的本原，就是以“数据”的形式记录和展现出来了。只有拥抱大数据，搞清楚问题的本来面目，我们才能够准确地找到解决问题的办法。

坚信与希望

随着思考的进一步深入，我们发现，大数据技术虽新，但是，金融行业一向和数据是朋友关系，只不过原先只看到数据的一个维度，当四个维度同时相乘而来时，传统金融感受到了压力。我们坚信，在感受到压力的同时，也已经找到了动力。希望通过这本书，让更多的人加入到思考各行各业与大数据的关系中，让作者、读者和来自五湖四海的人，也能够形成一个笛卡儿乘积的效应，加快应对冲击的速度。

以上就是这本书写作时的一个心路历程。欢迎扫描与本书配套的“金融大数据”二维码。在这个二维码的背后，有一个通过智慧的不断流动而产生价值的“金融大数据”平台期待着您的来访！文字是静止的，但是记录下了流动的思想，这就如同金融是资金的高效流动一样，思想的流动必将打破个体的局限性，能够让思想有了更多转化为行动和效果的力量。

本书撰写并能付诸出版，要感谢我的伙伴们：贾旸、左妮、高明洋、初凌峰、石博慧、

郑晓勇、范铮、袁泉、郑章虎、郭太苹、周博、王煜薇、肖冰等，他们利用业余时间查资料、做实验，让每个案例都生动起来；为了让读者持续获得最新的知识，这个团队还建立微信公众号平台，身体力行地提供“闭环化的客户服务”，其专业敬业能力令我感动。感谢电子工业出版社的编辑刘皎和她背后的团队，是他们中肯的建议和忘我的付出，不断地优化这个新生儿的基因，并且给它穿上了最得体的衣裳。感谢计世资讯等为本书推广做出努力的朋友们，他们和我一样关怀照顾这本书。最后，要感谢我的家人，尤其是我的女儿，她同意放弃了许多原本属于陪伴她的时光，让我用于思考和写作。要感谢的人很多很多，在此，对所有关心、支持和帮助过我们的朋友以及正在阅读本书的读者，谨致衷心的谢意！

目　录

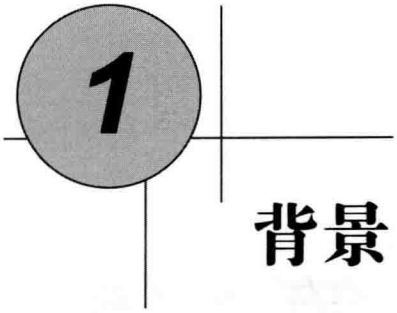
第 1 篇　大数据战略规划

第 2 篇 场景实例详解

第 3 篇 常见问题与应对机制

第 1 篇

大数据战略规划



背景

记得60年代我国地质学家在探测制造原子弹的原料铀时，采用了一个方法，就是动员各地的档案管理员，寻找在古代地方志中关于“鬼村”的描述，找到在哪些地方突然间集体性地怪异死亡、体征频繁发生异常变化等，往往在这些地方发现放射源的概率较高。在古代，人们并不知道这些异常现象的真正原因，所以归结为鬼神所为，并“如实”记录了下来。随着地质勘探的进展和科学技术的进步，人们发现这些存在铀等金属的区域往往同时有诸多鬼故事的传说，两者出现了正相关性。这帮助了那时的人们在没有优良探测仪器下，更为快速准确地找到放射源。汗牛充栋的古籍传记和轶闻怪谈可以视为当今的大数据，关联分析作为收集、储存和挖掘更多知识的一种方法，可以帮助人类提供更为科学和准确的预测工具。

1.1 大数据是什么

关于大数据的定义有很多种说法，读者可以在网上搜索获知，在本书中不做赘述。在这里特别提出以下两个观点：

其一，大数据指的是能够在合理时间内对数据资料进行撷取、管理、处理，并整理成为帮助企业经营决策达到积极目的的资讯。