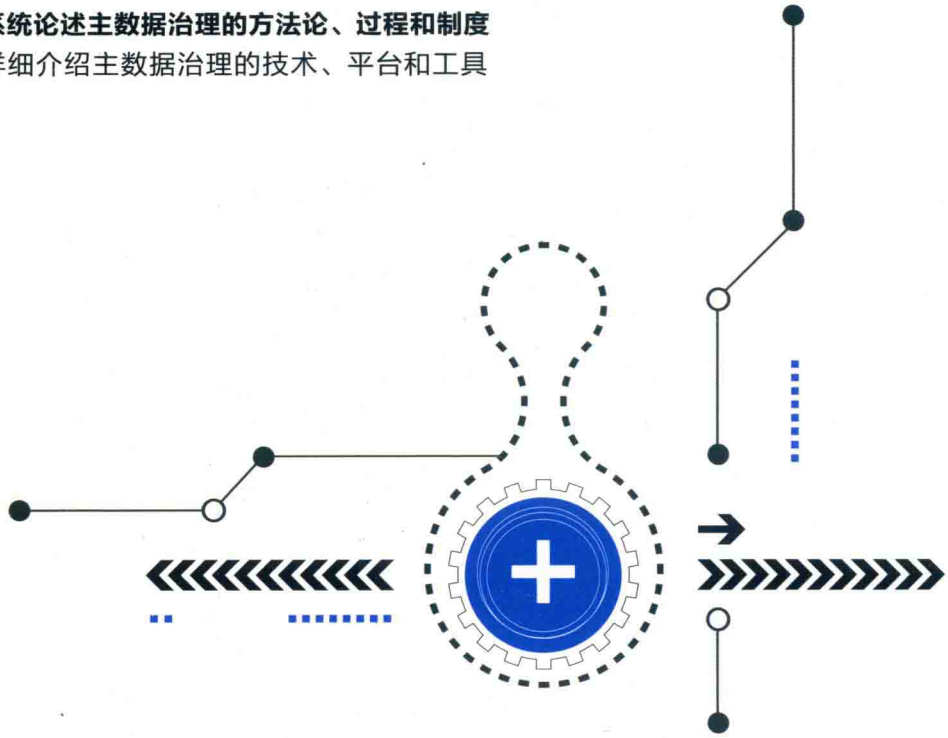
系统论述主数据治理的方法论、过程和制度
详细介绍主数据治理的技术、平台和工具

清華 开发者书库



Data Governance Driven by Master Data
Principles, Technologies and Practices

主数据驱动的数据治理
原理、技术与实践

王兆君 王 钺 曹朝辉◎编著
Wang Zhaojun Wang Yue Cao Zhaohui

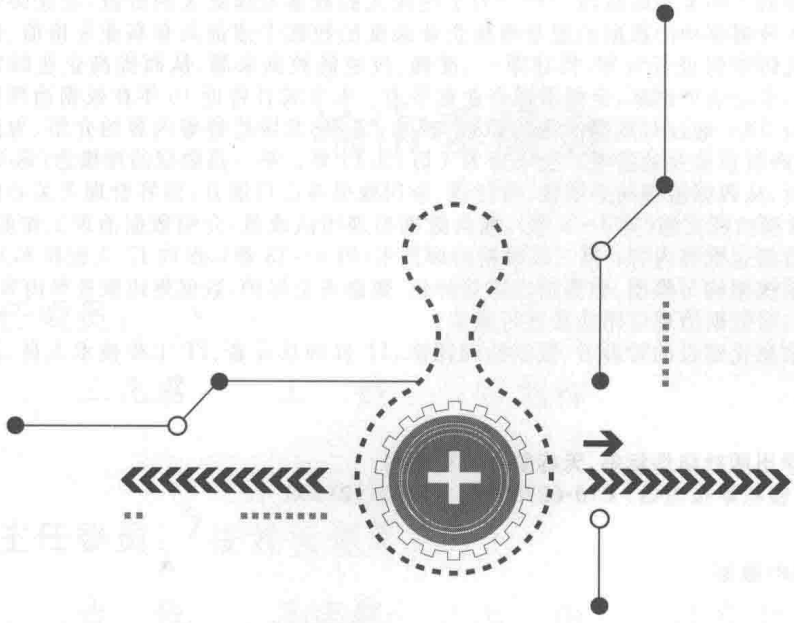
唱 伟 中国五矿集团有限公司招标采购中心主任
王红楼 丽珠医药集团股份有限公司首席信息官
梁 旭 三一重工股份有限公司董事长助理/副总经理/首席信息官
朱 亮 中国铝业集团有限公司信息管理部处长
郑 锋 特变电工集团首席信息官

联袂推荐

清華大學出版社

清華

开发者书库



Data Governance Driven by Master Data
Principles, Technologies and Practices

主数据驱动的数据治理

原理、技术与实践

王兆君 王 钺 曹朝辉◎编著
Wang Zhaojun Wang Yue Cao Zhaohui

清華大學出版社
北京

内 容 简 介

“数据”已成为企业的一项宝贵的战略资产。为了使庞大的数据发挥更大的价值，企业必须着眼于数据治理和综合利用。主数据驱动的数据治理是指从企业杂乱的数据中捕捉具有高业务价值、被企业内各业务部门重复使用的关键数据进行管理，构建单一、准确、权威的数据来源，从而提高企业的整体数据质量，提升数据资产价值，推动业务创新，全面增强企业竞争力。本书编者将近 10 年在数据治理咨询工作中积累的经验和知识进行总结，通过对数据治理的原理、技术、案例、发展趋势等内容的介绍，为读者进行数据治理、主数据管理实践提供重要的参考。全书分为 4 篇，共 14 章。第一篇数据治理概念（第 1～3 章），面向数据治理组织管理者，从数据治理的必要性、可行性、应用效果等进行展开，回答管理者关心的数据治理的核心问题；第二篇数据治理实施（第 4～8 章），面向数据治理团队成员，介绍数据治理工作的前期准备、工作步骤、治理过程、后期运维等内容；第三篇数据治理技术（第 9～13 章），面向 IT 工程技术人员，从技术视角展开数据治理的系统架构与模型、数据治理质量评估、数据安全保护、数据集成服务等内容；第四篇数据治理前景（第 14 章），对数据治理应用前景进行展望。

本书可作为从事信息化建设的管理者、数据治理团队、IT 咨询从业者、IT 工程技术人员、相关专业在校师生的参考读物。

图书在版编目（CIP）数据

主数据驱动的数据治理：原理、技术与实践/王兆君，王钺，曹朝辉编著. —北京：清华大学出版社，2019
（清华开发者书库）
ISBN 978-7-302-52295-9

Ⅰ. ①主… Ⅱ. ①王… ②王… ③曹… Ⅲ. ①数据处理 Ⅳ. ①TP274

中国版本图书馆 CIP 数据核字（2019）第 029341 号

责任编辑：盛东亮
封面设计：李召霞
责任校对：李建庄
责任印制：董　瑾

出版发行：清华大学出版社
网　　址：http://www.tup.com.cn，http://www.wqbook.com
地　　址：北京清华大学学研大厦 A 座　　邮　　编：100084
社 总 机：010-62770175　　邮　　购：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
课件下载：http://www.tup.com.cn，010-62795954
印 刷 者：北京富博印刷有限公司
装 订 者：北京市密云县京文制本装订厂
经　　销：全国新华书店
开　　本：186mm×240mm　　**印　张**：23.5　　**字　数**：528 千字
版　　次：2019 年 4 月第 1 版　　**印　次**：2019 年 4 月第1次印刷
定　　价：89.00 元

产品编号：081386-01

编审委员会

序
FOREWORD

让我们从一个简单的问题开始：IBM 有多少雇员？

这个问题看上去非常简单直接，对吗？但是请注意下面这个清单：

International Business Machines Corporation	IBM
IBM Microelectronics Division	IBM Global Services
IBM Global Financing	IBM Global Network
IBM de Columbia, S. A.	Lotus Development Corporation
Software Artistry, Inc.	Dominion Semiconductor Company
MiCRUS	Computing-Tabulating-Recording Co.

这个长长的清单中的每一项都与 IBM 有关系，有全称、缩写、别名、分支机构、全资子公司等，有的公司名称中完全没有 IBM 的字样，但它归属于 IBM，有的公司曾经归属于 IBM，后来又被卖掉了，还有的公司现在已经完全不存在了。

现在我们再来看刚才那个问题——IBM 有多少雇员？还会觉得这个问题简单吗？

事实上，我遇到过一所知名大学信息管理部门的负责人，问他这个关于雇员的问题。他告诉我，这正是让他们头疼的问题。太多不同时期建设的信息系统、不同的编号、不同的命名体系、不同的管理方式，所有信息汇总到一起之后，不知道哪些是重复的，哪些是陈旧的。用不同的方式统计，结果都不相同。

对于拥有众多部门和分支机构的大型企业，这样一种数据管理的困境随处可见。因此，不仅对于人员、物料、市场，而且对于那些与企业运营密切相关的重要信息都存在着管理的挑战。所以，我们需要数据治理，尤其是对企业中最关键的数据资产——主数据进行治理，进而提升数据质量，使数据真正成为管理和决策的可靠依据。

我们正处在历史的转折点上，数据技术在快速变革，大数据成为人们竞相议论的热点。无疑，未来的竞争就是数据的竞争。但是，在这个变革的关键时点上，更多的人将注意力的焦点放在了数据的“量”上，很少有人提及和关注数据的“质”，仿佛只要有了足够大量的数据，一切问题都可以解决。很可惜，真实情况是，海量数据如果未能经过合理的加工和组织，并确保一定的数据质量，它不仅不能解决问题，反而可能制造出更多的麻烦。也许，我们应该尽早从华而不实的喧嚣中抽身出来，通过具体而细致的数据治理工作，切实改善企业的数据环境，让大数据真正从“看”到“用”，真正活跃起来。

本书凝聚了编著者在数据治理和主数据管理领域多年的从业经验，涵盖数据治理和主

数据管理的基本概念、实施过程、关键技术等重要内容，并结合大量实际积累的案例和技术方案，系统地介绍了数据治理这一新兴领域及其应用情况，可作为工作指南为正在或准备开展数据治理工作的IT人员提供参考，更能为数据时代的企业管理者提供新的思路、新的方向。

张林　清华大学教授，清华-伯克利深圳学院院长
清华大学物联网与社会物理信息系统实验室主任

2019年1月于北京

前言

PREFACE

在过去的几十年里，对数据的计算和存储能力以及可用性的巨大进步，促成了当今数据驱动型的世界现状。数据正在对整个人类社会产生巨大的积极影响，它不仅在改变着人们生活的各个方面，而且也使得企业的运营更加高效。互联网数据中心（IDC）预测，到 2025 年，全球数据圈将扩展至 163ZB（1ZB 相当于 1 万亿 GB），是 2016 年所产生 16.1ZB 数据的 10 倍，这些数据将给个人带来全新的用户体验并且给企业带来更多的商业机会。

虽然已经有部分企业认识到数据资产的重要性，但是随着数据数量、种类以及重要性的不断增加，收集、存储和处理这些数据的难度也越来越大。如何从海量数据中挖掘出对制定决策有价值的信息，成为企业在管理和使用数据过程中面临的主要挑战。

数据治理的核心正是加强对数据资产的管控，通过深化数据服务以持续创造价值，企业领导者必须关注其中最重要的那部分数据，只有识别并充分利用这些至关重要的数据，才能发挥其巨大潜力。主数据管理就是从来源复杂的数据中捕捉关键数据，并且对这些具有高业务价值的、可以在企业内跨越各个业务部门被重复使用的数据进行管理，通过为跨构架、跨平台、跨应用的系统提供一致的、可识别的主数据对象来支持整个企业的业务需求，从而提高企业的整体数据质量，提升数据资产价值，推动业务创新，全面增强企业竞争力。主数据管理是一个全面的战略，涵盖所有需要统一定义的、企业所需的核心数据和数据标准。主数据管理的有效途径是建立一个包括主数据标准体系、主数据管控体系、主数据质量体系和主数据安全体系在内的、完整的主数据体系，建立持续长期的管理机制，这样才能构建企业数据的核心治理能力，合理利用企业数据来寻求竞争优势。

本书编者从事数据治理和主数据管理咨询工作近 10 年，亲身经历了数据治理和主数据管理在中国企业信息化浪潮中的兴起、演进和实践的过程。目前，为了配合国家信息化发展战略，很多企业把数据治理和主数据管理系统建设项目提上日程，并且开展了部分信息标准化工作。但是，从总体上看，国内企业的主数据体系建设工作仍然处在起步阶段，很多企业管理者对数据治理和主数据管理的概念理解有限，对主数据管理体系建设的重要性认识不足。编者将在数据治理和主数据管理领域的从业经验和知识积累进行总结，与大家分享和探讨，并希望能回答什么是数据治理和主数据管理、为什么需要数据治理以及如何进行主数据管理等问题。

本书坚持“贴近用户”的思路，回答用户关心的核心问题，不仅介绍主数据管理的产生背景、概念、模型和技术等理论知识，同时涵盖主数据管理项目的实施方法和过程、主数据管理

的产品和应用案例，使读者对主数据管理项目从底层技术知识到上层应用实践都能有系统的理解。同时，本书有针对性地对行业主流厂家的主数据管理产品进行了全面介绍，让读者能够更加深入地了解行业主流产品与趋势。书中案例都是近几年国内相关行业的领先企业的优秀实践，对其他企业的主数据管理和数据治理工作具有很高的参考价值。另外，书中对大数据、云计算、人工智能和区块链等新兴技术与主数据管理的结合应用也进行了探讨及趋势分析。

全书通过对主数据管理的背景、概念、模型、技术、实施、产品、案例、发展等内容的全面介绍，为读者揭开主数据管理这一新兴概念的神秘面纱，为读者进行数据治理、主数据管理实践提供重要参考。全书分为 4 篇，共 14 章。第一篇数据治理概念，包括第 1～3 章，其中第 1 章介绍数据治理的背景、意义和核心内容，并且引入数据管理的成熟度模型，使用户可以根据自评表得到成熟度评估和治理建议；第 2 章讨论主数据和主数据管理的概念和意义，为读者揭示主数据管理的必要性；第 3 章讨论主数据驱动的数据治理，系统地介绍治理框架、治理过程和数据治理工具。第二篇数据治理实施，包括第 4～8 章，其中第 4 章介绍主数据治理项目的准备工作；第 5 章讨论主数据体系规划方法；第 6 章说明主数据治理项目的具体实施步骤；第 7 章介绍主数据项目的运维和管理；第 8 章介绍目前国内主流的主数据管理解决方案和产品，并分析国内主数据管理的先进案例。第三篇数据治理技术，包括第 9～13 章，其中第 9 章介绍数据架构和模型的相关技术知识；第 10 章讨论数据集成技术及其企业应用；第 11 章介绍数据质量管理的定义、评估框架以及数据质量战略；第 12 章讨论数据生命周期管理的概念、内容和体系架构；第 13 章介绍数据安全管理和数据隐私保护。第四篇数据治理前景，包括第 14 章，主要展望主数据与大数据、云服务、人工智能和区块链应用的发展趋势。

本书既可补充从事信息化建设的 IT 部门人员的专业知识，更能为组织管理者提供信息化知识储备和工作思路，助力 IT 架构的组织优化。本书也面向咨询公司的顾问和实施人员，不仅针对主数据管理项目，而且对处理各类信息系统项目中可能出现的数据问题都具有一定参考价值。本书还可以作为企业管理软件开发人员的自学参考书，以及相关专业在校师生开阔视野、理论联系实践的参考书。

在本书的编写过程中参考和引用了国内外很多书籍和网站的相关内容，部分图片素材和个别实例的初始原型也来源于网络，部分互联网相关资源无法一一列举出处，在此向其作者一并予以感谢。众所周知，一本书难免出现不足和疏漏之处，恳请广大读者将意见和建议反馈给我们，以便在后续版本中不断改进和完善。有关数据治理的更多信息，可关注北京三维天地科技有限公司微信公众号。

编　者

2019 年 1 月

目录

CONTENTS

第一篇　数据治理概念

第三篇 数据治理技术