

5

…

…

…

9 5 1

…

6 8 5 3

…

6 9 3

…

485

… …

… …

…

3565

… …

…

5

5

6 9 0 3

…

…

…

…

…

…

…

…

…

…

…

4

…

…

…

…

…

…

…

5 9

… 9 8 7

…

…

…

7 1 5 4 2 3 8 7 5

3 2

5 …

0 8

…

…

2 1 3 … 1 5 2 1 3 9 的

51320881

9 9 5 1

5 1 6

9 5 2 1 3

…

…

…

… …

…

…

…

…

3 0 5

… 9 5

2

3 5 0

…

…

…

…

…

7 1 5 1 2

5

3 2

… 5

0 8

…

…

…

…

…

8

9 5 3 1 6

…

…

…

…

…

6

5

1

5

…

…

8

5 2 1 3

…

3

6 5 1 3

5

6

9

6

3

3 B 0 5 1 8 2

6

7 1 5 …

2 3 8 1 5 1 3 2

1 0 8

8 6 5

9 6 5

8 5

…

5

1

…

3 5 1 8

…

8 5 5 2 1

…

1695

… …

… …

**银行业信息科技风险管理高层指导委员会**

**银行业信息化丛书**

**银行数据治理**

陆顾新 陈石军 王 立 等编著

|  |  |
| --- | --- |
| …  …  …  …  …  85  …  …  …  …  654  8  3  5 | 5 …  …  …  3  4  7  …  i  0 5 |

|  |  |  |
| --- | --- | --- |
| …  …  0  8 | …  …  …  …  6  1  3  5 | …  …  …  3 |

… 9

1

.0

8

|  |  |
| --- | --- |
| 3  5  6  5 | …  …  …  3 |

3

9

0

8

5

5

3

**Data Governance of**

**Commercial Bank**

**机械工业出版社**

**CHINA MACHINE PRESS**



1…9…………5

2l0eg1

1 1 8 9 5 1 3 9 0 8 9 5 1

… 6 8 9 3

9 8 5 1 3 6 9 8 0 5 3 1 6 8 3

…

3

9

3

8 9 5 3 1 8

…

…

…

9 5 1

…

…

…

…

…

5

1

8

0 8

…

9 6 3

1

5

7

4

…

…

8

1

…

…

9

6

5 3 0 8

3 5

8 9 8 7

2 7 9 5 1 0 7 5 3 1 9 0 8 5 1 5 3 2 1

0 1

………………口

…

6 5 2 1 3

…

…

6 5 1 3

…

…

5

…

…

…

0 5 3

9 6 5

…

5

3 5 …

6

…

…

… …

…

3

5

6

0 9

9 5 3 1 6

…

…

…

…

…

…

…

…

3 6 0 5

1 3 2

…

2

3

8

7

7 g 5 1

… 1

5

2

7 5 3 1 g

…

5

…

…

5

3

0

8

9 8 7

6 0 8 9 5

8 9 6 3

…

…

…

2965139351935

71542

… …

…

i

7 …

…

… …

…

4

… … 2

… 3 3

… … 8

3595

875132

7

1

5

2

5

1

7

…

…

…

|  |  |
| --- | --- |
| 3 | 0 5  … 1 |

05316051

5 … … 8

…

… …

04517

5

…

…

…

…

…

6

9

51532

9903

…

…

…

521309

895213

6313

**银行业信息科技风险管理高层指导委员会**

**银行业信息化丛书**

**银行数据治理**

**陆顾新** **陈石军** **王** **立** **等编著**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 5 3 …  …  … …  … …  … …  … 8 1 5  … 多  … 9 5 | …  …  …  053  367  …  985  1695  4951  …  i  …  69451  …  …  …  …  …  …  95  …  13805  …  65816  …  845  …  …  …  …  …  …  …  …  …  …  585  …  …  …  …  …  5  …  05  53  …  …  5  …  …  8  …  …  …  …  …  …  …  …  …  …  …  …  …  5  …  …  5  85  …  …  …  95  …  35  2350  36356903  …  …  …  …  …  …  …  95  3835  85  …  …  …  …  993  9635  …  …  …  9  …  …  …  9435  893  …  …  …  …  65213  …  …  513  …  6513  6  563  …  09 | 5 1 5  …  …  …  …  i 5  0  7  5  3  …  1  1  …  …  …  …  5 2 1 3 0 6  … …  5 1 2 2  …  …  …  …  …  …  … 5  5  3  … 6 5 1 3 8 3 5 上  … …  … … | 5 4 2  …  …  …  …  …  7  … …  7 5  2  5  0 8 | …  …  8 9 5 3 1 9  2 3 1  …  …  …  …  8 9 5 1  9 6 3 | 1 6 9 5  … …  … …  … …  … …  8 …  … 9 常 | …  9 6 5 1 3 6  6  5  …  6  9  …  5 1 9  6 5 2 1 3  …  …  1  3  3  0  985  … 6  …  …  …  56  …  …  9  …  1  8  …  16  0  5  90  … 3 | … 3  …  …  …  …  …  3 8 7  9 6 3 5 6 9 0 3  i  0 5  …  …  …  …  …  5  9 5 1 3 | 3 5 6 5 3 1 9 5 1  …  …  3 5 1  6  0 5 3 1 2  …  0  5  3  … |

**Data Governance of**

**Commercial Bank**

**机械工业出版社**

**CHINA MACHINE PRESS**

本书通过回顾银行业信息化的历史和我国银行业数据治理的历程，阐 明了数据对于银行发展的重要作用。同时，本书结合当前商业银行面临的 数据治理现状，提出了银行数据治理的概念和范围，全面阐述了如何建立 商业银行数据治理体系，客观提出了涵盖战略、机制、领域、技术支撑的 从上至下指导、从下而上推进的多层次、多维度、多视角的全方位商业银 行数据治理体系框架，即以企业战略对数据治理的目标及规划为指导，以 配套的数据治理流程为推力，积极利用系统和技术手段保障数据治理的有 效开展与落实。

本书可供商业银行从事数据管理的人员使用，也可供其他行业的相关 从业人员参考。

**图书在版编目** **(CIP)** **数** **据**

银行数据治理/陆顾新等编著.—北京：机械工业出版社，2016.5 (2016.8重印)

(银行业信息化丛书)

ISBN 978-7-111-53485-3

I.① 银 … Ⅱ.①陆 … Ⅲ.①银行一数据管理—研究 IV.①F830.49

中国版本图书馆 CIP数据核字(2016)第073294号

机械工业出版社(北京市百万庄大街22号 邮政编码100037)

策划编辑：王振国 责任编辑：王振国 责任校对：陈延翔

责任印制：李 洋

三河市国英印务有限公司印刷

2016年8月第1版第2次印刷

184mm×260mm·13印张·321 千字

3001—4900册

标准书号：ISBN 978-7-111-53485-3

定价：69.80元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

电话服务

服务咨询热线：010-88361066 读者购书热线：010-68326294 010-88379203

**封面无防伪标均为盗版**

网络服务

机 工 官 网 ：www.cmpbook.com

机 工 官 博 ：weibo.com/cmpl952

金 书 网：www.golden-book.com

教育服务网： www.cmpedu.com

**“银行业信息化丛书”编委会**

主 编：尚福林

副主编：郭利根

编 委：(按姓氏拼音排序)

陈天晴 陈文雄 方合英 甘 煜 谷 澍 侯维栋 李 丹 李 浩 李丽芳 李 翔 李振江 林晓轩 林治洪 潘卫东 庞秀生 曲家文 单继进 童 建 王 兵 王 健 王用生

谢肿达 许 文 薛鹤峰 于富海 张华宇 张依丽 朱鹤新

编 辑：(按姓氏拼音排序)

傅晓阳 龚伟华 何 禹 焦大光 金磐石 李 璠 李海宁 李建军 梁 峰 刘国建 刘秋万 刘子瑞 鲁 森 骆絮飞 吕仲涛 牛新庄 谭 波 汪 航 王 燕 吴永飞 奚力铭

徐 徽 于慧龙 余宣杰 周黎明 周天虹

工作组： ( 按姓氏拼音排序)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 曹文中 | 陈宇能 | 黄登玺 | 黄绍儒 | 霍宝东 | 贾俊刚 金建新 |
| 李洪伟 | 李 燕 | 林长乐 | 刘文波 | 孙 莉 | 唐 宗 卫剑钒 |
| 夏建伟 | 闫晓鹤 | 张 健 | 张立书 | 钟 亮 | 朱学良 |

 **总** **序**

信息化是推动经济社会变革的重要力量。坚持走中国特色的新型工业化、信息化、城镇 化、农业现代化道路，是党中央立足全局、放眼未来、与时俱进的战略决策。2014年2月 27日，中央网络安全和信息化领导小组的成立，更加体现了中央保障网络安全、推动信息 化发展、维护国家利益的决心。银行业作为国家经济体系的重要行业之一，是信息化的重要 推动主体、参与主体和受益主体。银行业持之以恒地贯彻落实国家信息化战略，不仅是推动 加快我国信息化进程的必然要求，也是银行业改革发展、转型升级和更好服务实体经济的内 在需求。

近年来，我国银行业审时度势、积极作为，坚持基础建设与科技创新并重、提升服务与 保障安全并举的科学发展导向，以推进信息化为契机，调整经营理念、优化经营机制、完善 服务模式，在服务手段信息化、管理模式信息化、信息安全保障等方面取得积极进展，推动 了银行业的核心竞争力、市场适应力和贴身服务能力的进一步提升。 一是服务手段信息化发 展迅速。电子银行、自助银行、智能支付终端等信息化服务渠道日渐普及，使得金融服务覆 盖面更加广泛、服务方式更加便捷、服务产品更加丰富。二是管理模式信息化迈出实质性步 伐。注重依托核心数据库、运用先进数据挖掘分析工具，推进银行经营决策逐步智能化，风 险管理日趋精细化，产品创新逐渐体现个性化，银行业经营管理信息化水平不断提升。三是 信息安全保障取得积极进展。银行业信息安全越来越受重视，相关科技基础设施建设步伐加 快，多层次、立体化、全方位的信息安全保障体系正在逐步形成。

当然，我们也应该清醒地认识到，银行业信息化面临着复杂的内外部环境，核心技术受 限、网络安全威胁、隐私保护和信息保密等挑战将长期存在，银行业自身认识不尽到位、技 术储备不够充分、资源投入相对不足、过度依赖外包等问题仍较为突出，针对银行业特殊需 求的信息化产品、工具和方法还比较单一，缺乏应对复杂需求的灵活创新能力。总的看来， 银行业信息化还有很长的路要走，信息科技风险将成为当前和未来较长时期银行业的重要风 险领域之一。

银行业信息化既不能因为成绩而骄傲自满，也不可因为差距而妄自菲薄，更不可因为困 难而畏首畏尾。各银行业金融机构要勇于直面困难、主动迎接挑战，坚决按照国家信息化总 体战略部署，切实坚持“自主可控、持续发展、科技创新”的基本方向，紧紧抓住信息化 发展机遇，推动信息服务和信息安全再上新台阶。 一是借助信息化推动银行业金融机构治理

**总序**

能力现代化。积极引入先进的信息科技治理和管理理念，运用现代信息技术缓解治理中的信 息不对称问题，推动流程银行建设，提高治理有效性。同时，理顺信息化建设的体制机制， 加快信息化建设进程，为银行业转型发展提供有力保障。二是依托信息化推动金融服务智慧 化。要充分利用互联网、移动计算蓬勃发展的大环境，积极应用大数据等新兴技术，创新思 维模式，充分发挥金融数据和信息的价值，研发智能化、个性化、便捷化的产品和服务，灵 活响应客户诉求，努力改善客户体验，尽力发掘潜在客户需求，增加产品和服务的吸引力， 培育更为坚实的客户基础，形成新的业务和利润增长点。三是以自主创新增进安全可控能 力。要坚持市场起决定作用的基本方针，探索形成以研发创新支持应用推广、以市场应用激 发创新动力的良性正反馈机制。推动应用自主创新信息技术，建立自主创新信息技术落地银 行业的配套机制，力争金融领域关键信息技术自主创新占比逐步提高，不断提升信息系统的 开放性、灵活性和整体集约化水平。四是利用信息技术强化行业协作。要加强银行业信息化 建设的统筹规划，促进信息化资源的集约共享，提升数据(灾备)中心布局的合理性，增 强同业协同协作，共同应对外包集中度等风险。

为更好地推进落实银行业信息化战略，由银行业信息科技风险管理高层指导委员会指导 推动，编著了“银行业信息化丛书”(简称“丛书”)。这套丛书致力于挖掘、研究、总结、 提炼和传播国内外信息化最佳实践、宝贵经验和最新成果，内容涵盖银行业信息科技治理与 管理、信息系统开发与应用创新、信息安全、基础设施与运行维护、信息科技监管等主要领 域，可为银行业信息科技人才培养提供一些基础性、前瞻性、实用性的知识和信息。

展望未来，银行业信息化任务艰巨、时间紧迫。希望银行业在有关各方支持下，推动信 息化工作更加积极主动、规范有效、科学前瞻，为我国银行业持续健康发展、提升服务水平 提供坚实的支撑，为增强国家网络安全保障能力、提升信息化建设水平提供有力支持，为贯 彻落实创新驱动发展战略、实现中华民族伟大复兴的中国梦做出积极贡献。



 序

我国银行业信息化已发展了30 多年，银行早期的数据基本上是交易的副产品，很少 用于分析决策、风险控制等。到了21世纪，越来越多的商业银行开始利用数据进行更为 精准的客户营销、提高客户的服务水平、进行实时或准实时的风险分析和反欺诈分析， 数据价值将逐步得以体现。但是，数据使用过程中暴露出来的问题也越来越突出，例如： 企业层级的通用业务规则缺乏，导致不同业务部门之间的标准常有矛盾或是混淆；缺乏 企业整体的宏观设计，仍然停留在面向功能为主的架构设计；数据的交换和共享困难， 存在严重的数据不同步、不一致；没有明确的角色、职责、职权分工，缺乏跨部门的数据 管理体系等。

事实上， 一旦缺乏正确的数据管理和高质量的数据，那么数据价值就无法得到体现，进 而银行业信息化建设就成了一纸空谈。因此，数据治理工作成为银行业信息化面临的重要挑 战，也是银行数据价值充分发挥的前提。幸运的是，各大商业银行逐步意识到这一问题，越 来越重视数据治理工作，纷纷将数据治理提上了建设议程，开始建立适应本行的数据治理 体系。

数据治理是一个长期而繁琐的过程，而且面临各个方面的挑战。例如：在数据组织的一 致性方面，需要建立数据标准，用于指导数据格式、数据结构、数据字典、编码规则等，这 样才能保证系统间数据的有效交互；在数据质量的保证方面，不仅要从技术层面建立相应的 事前、事中和事后的校验、监控、矫正等机制，还要在业务层面提出合理的、 一致的业务规 则。当然，在技术层面建立辅助系统和工具是行之有效的手段，可以使数据变得有序、合 理、规整，对银行业的风险管理、营销支持、决策分析等经营管理起到关键性作用，但最重 要的还是要从战略层面上予以支持和指导，从战术层面建立具体的章程，并且作为一种重要 的企业文化来执行。

在银行业信息科技风险管理高层指导委员会的牵头组织下，课题组编写了《银行数据 治理》这本书，就如何建立商业银行数据治理体系进行了探讨，从战略、文化、组织、制 度章程、系统工具等方面进行了详细阐述，涵盖了数据标准、元数据、数据分布、数据质 量、数据安全等数据治理领域。希望这本书能对我国商业银行数据治理工作带来一定启发， 从而起到重要的引导作用。

作为拥有30多年工作经验的银行从业人员，目睹了中国工商银行一个又一个系统的建

序

设过程，见证了各个系统间数据综合利用的难度，也深知数据治理在各个商业银行都是一项 艰巨而具有挑战性的工作。很高兴中国工商银行能有这个机会，集中国银行、中国邮政储蓄 银行等业界同人的鼎力相助编著此书，并将商业银行数据治理工作进行系统化阐述，希望这 本书能够起到抛砖引玉的效果，逐步将商业银行的数据治理工作推向成熟化。

**吕仲涛**

 **前** **言**

全球已进入信息高速发展的时代，信息化是当今世界发展的根本趋势，是推动经济社会 变革的重要力量。我国在2006年发布了《2006—2020年国家信息化发展战略》,对我国的 信息化发展进行了战略指导。

为贯彻国家信息化发展战略要求，推进落实银行业信息化战略，在银行业信息科技风险 管理高层指导委员会指导下，我们着手编写了“银行业信息化丛书”之《银行数据治理》。 《银行数据治理》的核心是阐述如何对银行数据进行全面管理，创造价值，为银行业务发展 提供全面、高效的数据支撑。

商业银行在业务开展过程中积累了海量的数据，如客户数据、业务交易数据、内部管理 数据、外部数据、系统日志等。这些数据是银行最为重要的资产之一，是银行支持精细化管 理、实现差异化服务、提升风险分析能力的基础。然而，海量的数据不仅意味着海量的机 遇，也意味着无限的风险。目前，我国商业银行普遍面临着数据质量不高、数据支持决策能 力不足等问题，导致数据远未发挥其应有的价值，并已经成为银行提高自身竞争力的巨大障 碍，主要表现为：缺乏整体数据治理战略和文化、数据治理管理流程和职责划分不清、数据 标准不统一、数据质量不高、缺乏配套的技术和系统支撑等。

如何开展数据治理，实现数据对业务的价值创造，是银行信息化建设中要破解的一道难 题。《中国银行业信息科技“十二五”发展规划监管指导意见》指出：“商业银行要重点加 强对数据治理的制度建设和流程建设，建立和完善数据治理制度体系，规范工作流程，理顺 内部协作关系，提升数据质量和数据应用水平，提高数据价值创造能力。”数据治理需要一 个完整的治理体系，并从战略、机制、领域、技术等方面予以支撑。

在信息化快速发展的今天，银行的发展要靠数据，数据要靠治理，治理需出成果，成果 需要在实际应用中体现，而应用应与银行的整体发展战略契合，总结起来即为“银行靠数 据、数据靠治理、治理出成果、成果靠应用、应用看前景”。

本书通过阐述数据治理的概念和范围，回顾我国银行业数据治理历程，结合分析各商业 银行当前面临数据治理现状，全面阐述了商业银行数据治理体系，提出以战略、机制、领 域、技术为支撑的从上至下指导、从下而上推进的实施原则，着重阐述建立多层次、多维 度、多视角的全方位商业银行数据治理体系框架，包括但不限于：以企业战略对数据治理的 目标及规划为指导，以配套的数据治理流程为推力，涵盖完整的数据治理领域，积极利用系

**前言**

统和技术手段保障数据治理的有效开展与落实。

全书共有13个章节，总体内容结构如下：

第1章：银行的数据。阐述什么是数据以及数据对银行的作用，还介绍了银行的数据包 括什么、有什么特点、管理上存在什么不足，从而使读者对银行的数据有一个大概的了解。

第2章：银行数据治理概述。阐述银行数据治理的背景、目标及价值，从数据治理的发 展和现状出发，思考如何对银行的数据进行治理。

第3章：银行数据治理保障机制。介绍确保银行数据治理工作能够得到有效落实的保障 措施，包括战略层面、组织机构、制度章程、支撑系统等各个方面。

第4章至第12章，具体介绍了银行数据治理各个核心领域(包括数据模型、元数据管 理、数据标准、数据质量管理、数据生命周期管理、数据分布与存储、数据交换、数据安 全、数据服务)目前存在的问题、治理难点、应对措施，使读者了解到每个领域的数据治 理应该涉及什么样的工作、应该注意哪些细节，从而为实际的数据治理实施提供具体的指导 方法。

第13章：数据治理展望。数据治理的成效要靠数据发挥的价值来体现。本章介绍了在 大数据时代如何进一步加强数据治理、如何有效地利用数据，并以大数据在金融界的实际应 用案例生动地说明了只有高质量的数据才能发挥数据的最大价值。

本书在编写过程中得到了银监会各级领导的大力指导，得到了银行业同行的有力支持， 编写组所有人员也付出了辛勤汗水，在此对所有编写与指导人员表示衷心的感谢，包括中国 工商银行的张艳、张颖、罗毅、李伟良、王涛、李甲长、王鹏、乔敏、杨明伟、肖睿、郭 斯、巩垒、王超、袁如怡、高继荣，中国银行的张建游、王玮、曾玉、武晓利、刘晓波，中 国邮政储蓄银行的汪航、因亚南、张放、朱昌堆、陈元星，大连银行的石立军、于琦等。

尽管编写人员都投入了大量的精力，付出了艰辛的劳动，然而由于水平和能力所限，书 中错误和疏漏之处在所难免，恳请大家批评指正。

**编著者**

**目** **录**

**总序**

序

前 言

**第** **1** **章** **银** **行** **的** **数** **据** [**1**](#bookmark1)

1.1 数据、信息和知识 [1](#bookmark2)

1.2 银行的数据及作用 [3](#bookmark3)

1.2.1 银行数据的来源及特征 [3](#bookmark4)

1.2.2 数据对银行发展的重要作用 [7](#bookmark5)

1.3 银行数据管理存在的不足 [10](#bookmark6)

**第** **2** **章** **银** **行** **数** **据** **治** **理** **概** **述** [**15**](#bookmark7)

2.1 银行数据治理的背景 [16](#bookmark8)

2.2 银行数据治理的目标与价值 [19](#bookmark9)

2.3 银行数据治理的历史与现状 [21](#bookmark10)

2.3.1 国际社会数据治理发展历程 [21](#bookmark11)

2.3.2 我国银行业数据治理前进之路 [23](#bookmark12)

2.3.3 我国银行数据治理的现状 [26](#bookmark13)

2.3.4 银行数据治理的思考 [28](#bookmark14)

2.4 建立银行数据治理体系 [30](#bookmark15)

2.4.1 银行数据治理保障机制 [32](#bookmark16)

2.4.2 银行数据治理核心领域 [33](#bookmark17)

**第** **3** **章** **银** **行** **数** **据** **治** **理** **保** **障** **机** **制** [**36**](#bookmark18)

3.1 企业发展战略 [36](#bookmark19)

3.1.1 数据治理文化 [36](#bookmark20)

3.1.2 认知与行动 [37](#bookmark21)

3.2 数据治理组织 [37](#bookmark22)

3.3 数据治理制度章程 [39](#bookmark23)

3.4 数据治理流程 [41](#bookmark24)

**目录**

3.5 数据治理支撑系统 [42](#bookmark25)

3.6 数据治理成熟度模型 [42](#bookmark26)

**第** **4** **章** **数** **据** **模** **型** [**45**](#bookmark27)

4.1 数据模型体系 [45](#bookmark28)

4.2 数据建模的方法、原则及意义 [47](#bookmark29)

4.3 数据模型建模要求 [48](#bookmark30)

4.4 数据模型设计实例 [50](#bookmark31)

4.5 非结构化数据模型 [60](#bookmark32)

4.6 数据模型主题 [61](#bookmark33)

**第** **5** **章** **元** **数** **据** **管** **理** [**63**](#bookmark34)

5.1 元数据管理概述 [65](#bookmark35)

5.2 搭建元数据管理体系 [67](#bookmark36)

5.3 主要元数据对象的管理 [70](#bookmark37)

5.4 非结构化数据的元数据管理 [74](#bookmark38)

5.5 案例 [75](#bookmark39)

**第** **6** **章** **数** **据** **标** **准** [**76**](#bookmark40)

6.1 银行数据标准的现状 [76](#bookmark41)

6.2 银行数据标准的定义与分类 [78](#bookmark42)

6.3 银行数据标准建设 [80](#bookmark43)

6.4 非结构化数据的数据标准管理 [86](#bookmark44)

6.5 案例 [87](#bookmark45)

**第** **7** **章** **数** **据** **质** **量** **管** **理** [**90**](#bookmark46)

7.1 商业银行数据质量存在的问题及影响 [90](#bookmark47)

7.2 建立全面管控的数据质量管理体系 [95](#bookmark48)

7.3 非结构化数据的数据质量管理 [102](#bookmark49)

7.4 案例 [103](#bookmark50)

**第** **8** **章** **数** **据** **生** **命** **周** **期** **管** **理** [**105**](#bookmark51)

8.1 数据生命周期管理及意义 [106](#bookmark52)

8.2 银行业数据生命周期管理的现状 [107](#bookmark53)

8.3 建立数据生命周期管理体系 [108](#bookmark54)

8.4 案 例 [118](#bookmark55)

**第** **9** **章** **数** **据** **分** **布** **与** **存** **储** [**120**](#bookmark56)

9.1 商业银行系统数据存储管理 [121](#bookmark57)

9.2 商业银行系统数据分布 [126](#bookmark58)

9.2.1 总分行数据分布 [126](#bookmark59)

9.2.2 商业银行系统主副本数据源分布 [126](#bookmark60)

9.2.3 商业核心银行系统与外围系统间的数据分布 [128](#bookmark61)

9.3 案 例 [128](#bookmark62)

**第** **1** **0** **章** **数** **据** **交** **换** [**131**](#bookmark63)

10.1 制定数据交换的标准 [132](#bookmark64)

10.2 搭建数据交换体系和数据交换系统应用 [135](#bookmark65)

10.3 全流程数据流管理 [138](#bookmark66)

**银行数据治理**

10.4 外部数据交换 [139](#bookmark67)

**第** **1** **1** **章** **数** **据** **安** **全** [**141**](#bookmark68)

11.1 数据安全及重要性 [141](#bookmark69)

11.1.1 数据安全的概念 [141](#bookmark70)

11.1.2 数据安全的重要性 [143](#bookmark71)

11.1.3 数据安全的发展历程 [144](#bookmark72)

11.2 数据安全领域及管理 [145](#bookmark73)

11.2.1 数据存储安全 [145](#bookmark74)

11.2.2 数据传输安全 [149](#bookmark75)

11.2.3 数据使用安全 [151](#bookmark76)

11.3 案例 [154](#bookmark77)

**第** **1** **2** **章** **数** **据** **服** **务** [**158**](#bookmark78)

12.1 银行在大数据时代的战略应对措施 [158](#bookmark79)

12.2 构建大数据服务体系 [161](#bookmark80)

12.2.1 结构化数据处理系统 [162](#bookmark81)

12.2.2 非结构化数据处理系统 [164](#bookmark82)

12.2.3 流数据处理系统 [165](#bookmark83)

12.3 数据资产视图 [167](#bookmark84)

12.4 大数据在银行业的应用 [170](#bookmark85)

12.4.1 国内外银行业大数据应用案例 [170](#bookmark86)

12.4.2 银行业大数据应用场景 [171](#bookmark87)

**第** **1** **3** **章** **数** **据** **治** **理** **展** **望** [**175**](#bookmark88)

13.1 构建银行数据治理的长效机制 [175](#bookmark89)

13.2 大数据时代的银行数据治理 [176](#bookmark90)

13.3 大数据时代的银行数据利用 [176](#bookmark91)

**附** **录** [**180**](#bookmark92)

附录 A 《中国银行业信息科技“十二五”发展规划监管指导意见之大型商业银行和股份制

商业银行篇》(摘录) [180](#bookmark93)

附录 B 《贷款统计分类及编码标准(试行)》(摘录) [190](#bookmark94)

附录 C 《银行监管统计数据质量管理良好标准(试行)》(摘录) [192](#bookmark95)

附录 D 推荐读物 [195](#bookmark96)

**参** **考** **文** **献** [**196**](#bookmark97)



**第** **1** **章**

**银行的数据**

从20世纪90 年代末开始，以中国工商银行、中国农业银行、中国银行、中国建设银行 四大国有商业银行为代表的中国银行业陆续掀起了以数据大集中为主线的信息化建设浪潮。 数据大集中缩小了我国银行与国外银行在信息技术上的差距，使得利用集中的数据进行分析 和挖掘成为可能。

银行业信息化发展的重要基础就是各类数据和信息的整合利用，在未来一定时期内，银 行的数据作用将更为凸显。我国银行业已越过数据大集中的信息整合的初期阶段，后续银行 将会深入进行数据整合，基于统一的数据标准和数据模型，加强数据管控和治理，深入地对 银行数据信息进行挖掘、分析、处理和快速利用，充分发挥数据仓库等系统在加强客户和市 场分析能力，提升银行管理决策水平，提高银行核心竞争力等业务领域的推动作用。

特别是随着银行与互联网、合作方等相关领域的合作日趋密切，各类数据也以几何级数 增长方式而迅速膨胀。当前，我国大型商业银行的数据量已经达到 TB 甚 至PB 以上级别， 并且非结构化数据量也在快速膨胀。我国金融业正在步入大数据时代的初级阶段，并呈现快 速发展的势头。数据爆炸性地增长，对银行进行数据分析和处理，以及源源不断的产品和业 务创新等都带来了极大的挑战。未来银行将会是一个集数据分析、数据解读为一体的数据大 行。因此，我们必须了解什么是数据，什么是银行的数据，银行数据到底有什么样的作用和 存在哪些不足。

**1.1** **数据、信息和知识**

数据、信息和知识都是社会生产活动中的一种基础性资源，都可以采用数字、文字、符 号、图形、声音、影视等多媒体来表示，同时具有客观性、真实性、正确性、价值性、共享 性、结构性等特点。三者之间既存在密切的内在联系，又有着显著的区别。

数据、信息和知识是知识工作者对客观事物感知和认识的三个连贯的阶段，主要表现在 以下方面：

**1.** **数据的组织阶段**

数据是一种将客观事物按照某种测度感知而获取的原始记录。它既可以直接来自测量仪

**银行数据治理**

的组织过程。知识工作者应用信息系 统进行信息的捕获、导入、存储、加 工、挖掘、传输、使用和支持，这是 信息的创造过程。知识工作者在智能 化的信息系统支持下，运用大脑进行 知识的获取、提炼、分析、研究和发 现，这是知识的发现过程。

在已有知识的基础上，对信息进 行分析、对比、演绎并找出有价值的

器的实时记录，也可以来自人的认识，但是大量的数据多是借助于数据处理系统自动地从数 据源进行采集和组织的，比如银行借助计算机产生的客户数据。数据是最原始的记录，未被 加工解释，不能回答特定问题。它仅仅反映客观事物的某种运动状态，除此以外没有其他意 义。它与其他数据之间没有建立相互联系，是分散和孤立的。

根据结构的不同，数据可划分为结构化数据和非结构化数据两种。

**(1)结构化数据** 简单来说，就是存储在关系型数据库中的数据，可以用二维逻辑表 达实现。结构化的数据可以方便地通过计算机和数据库技术进行管理，信息量值出现在固定 的位置，它能够忠实、详细地记录交易活动，是一种显性的表示。

**(2)非结构化数据** 指信息的形式相对不固定，不方便使用数据库二维逻辑表达的数 据，包括报表、账单、多媒体视频、网页、电子邮件、语音、图片、图像、办公文档等。银 行的非结构数据中超过1/2来源于与客户之间的互动，都是以客户为中心产生的数据。非结 构化数据信息量值的位置不固定，价值相对隐匿。

**2.信息的创造阶段**

信息是对数据进行加工处理，使数据之间建立互相联系，形成回答某个特定问题的文 本，以及被解释具有某些意义的数字、事实、图像等形式的信息。简单地说，信息是加工后 的数据。

**3.** **知识的发现阶段**

知识是知识工作者运用大脑对获取或积累的信息进行系统化的提炼、研究和分析的结 果。知识能够精确地反映事物的本质。

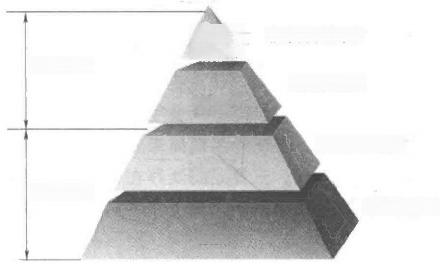
数据、信息、知识所处的三个阶段形成一个螺旋上升式循环周期。人们运用相关的信息 系统，对信息和相关的知识进行规律性、本质性和系统性的思维活动并创造出新的知识。之 后，新的知识又开辟了需要进一步认识的对象领域，然后使人们补充获取新的数据和信息， 进入新一轮的上升式循环周期。但三者又有一定区别：

**(1)具有各自突出的特性** 数据具有真实性、客观性的特点；信息具有针对性、时效 性、较少不确定性的特点；知识具有规律性、本质性、系统性的特点。

**(2)获取的途径不同** 数据通过采集获得；信息通过加工获得；知识是实践经验的总 结，是信息的结晶，是通过实践和学习获得的。

**(3)对知识工作者的支持过程不同** 知识工作者运用数据处理系统进行数据的编码，

采集、录入、存储和处理，这是数据



洞察力、创造力

主观意识

提炼的信息

知识

有组织的数据

信息

客观存在

事实和数字

数据

智意

部分，并将其深化到已有知识架构 图1 - 1 数据、信息、知识及智慧示意图

**银行的数据** **第1章**|

中，则上升为智慧。本书主要对数据以及数据产生信息进行探讨。图1-1 为数据、信息、知 识及智慧示意图。

**1.2** **银行的数据及作用**

银行的迅速发展是建立在其庞大的数据群及基于其抽象出的信息基础之上的。如果没有 数据，没有数据提炼的信息，那么银行业的发展只能成为一纸空谈。

管理大师汤姆·彼得斯说过：“一个组织如果没有认识到管理数据和信息如同管理有形 资产一样极其重要，那么它在新经济时代将无法生存。”数据给商业银行创造了能够更好地 理解客户、服务客户的机会。

世界经济论坛报告认为：“大数据为新财富，价值堪比石油。”大数据之父维克托曾在 2013年预测，数据列入企业资产负债表只是时间问题。

那么,数据是什么?银行又有什么样的数据?银行的数据有什么样的特征?银行的数据 到底能够发挥什么样的作用?这些都是我们关心的问题。

**1.2.1** **银行数据的来源及特征**

我国商业银行通过广泛的渠道及丰富的产品，在业务开展过程中积累了大量的数据，如 客户数据、业务处理数据、内部管理数据、外部数据、系统日志等。特别是近年来，随着社 交网络、微博、微信、论坛的兴起，数据来源也已从传统的结构化数据逐渐扩展到以网络日 志、社交媒体为代表的半结构化和非结构化数据，银行数据量呈现几何级数的爆发式增长， 商业银行逐步进入大数据时代。

大数据时代已经悄然而至，认同这一观点的人越来越多。那么银行的数据到底从哪里来?银 行有哪些数据?下面将从技术角度加以解惑，充分介绍它作为新时代变革力量的根本原因。

**1.** **银行的数据是什么**

银行业数据原来是金融产品、业务处理、运营管理以及客户服务的副产品，只在后台起 支撑作用。随着信息科学技术进步和数据处理能力的提高，大量的数据以结构化或非结构化 的形式在 IT 信息系统中传输、储存、转换、分析和挖掘，并产生能够为特定业务领域服务 的信息。

图1-2所示为银行的信息与数据。

随着网络技术的发展，特别是Internet (因特网)和 Intranet (内部网)技术的飞速发 展，商业银行在使用互联网技术、不断提升客户服务水平和加强风险管理的同时，存款、贷 款及汇款(简称存贷汇)等核心业务的结构化数据与日俱增，而电子商务、移动应用、社 交网络日益活跃，导致电话语音、网上银行交易记录、电子影像、网点视频、扫描文件、微 博、即时通信等非结构化数据(图1-3)、半结构化数据也日趋增大，非结构化数据的增长 幅度远远大于结构化数据。

**2.银行数据的来源和分类**

银行的数据主要来源于商业银行自身业务运营过程中积累的大量客户基本资料、客户交

银行数据治理

**(** **1** **)** **内** **部** **数** **据**

1)客户数据：即当事人基本信

息，主要用于描述客户的自身特点，

属于基础数据范畴。客户(当事人)

是指与银行有联络或与银行有利害关

系，以及银行希望保留其信息的所有

相关参与者。其中也包括银行本身，

如个人、外部机构、内部机构。

① 个人客户数据包括：客户姓

名、性别、国籍、证件类型、证件号

码、联系地址、通信电话、职业状

态、工作单位、职务、宗教信仰、婚

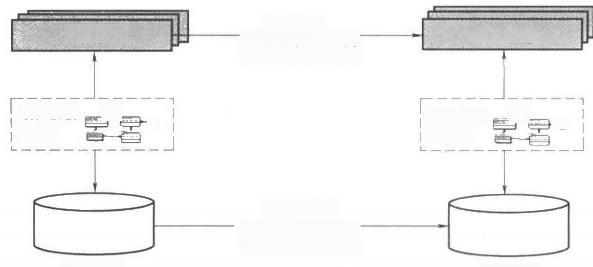
姻状况、文化程度、子女数量、父母

姓名、语言偏好，以及客户与账户的

关系等信息。

②公司客户数据包括：机构名称、 图1 - 3 非结构化数据

机构代码、经营范围、经营状态、所属



人脑思维

识别、转换、扩充-

统计、分析、挖掘

数据模型 

元数据

计算机系统

识别、转换、扩充-

统计、分析、挖掘

原始数据

数据模型 元数据

衍生信息

衍生数据

原始信息

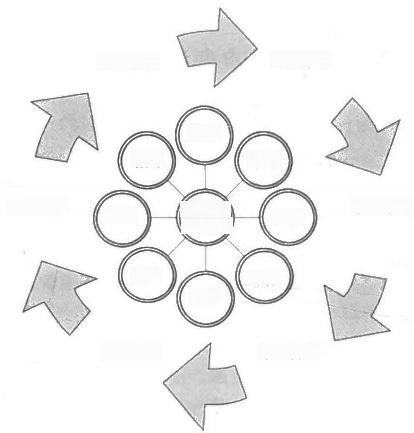


日

5

图1 - 2 银行的信息与数据

易和产品服务数据、运营管理数据，当然也包括外部宏观经济环境的运行指标等数据。根据 来源不同，银行信息系统的数据可分为内部数据和外部数据两大类，见表1-1。



数据收集

报表

微博

扫描

网页

语音

数据挖掘

非结构) 化数据》

数据利用

数据分类

数据评估

数据管理

视频

账单

邮件

国家、所属省份、公司网站、公司地址、公司类别、法人电话、财务电话、注册资金、行业代 码、经济性质、企业规模、资产规模、建筑物产权、办公面积、营业面积、客户来源等。

银行的客户数据来源于银行自身的多个系统，如客户管理系统、核心银行系统、交易系 统、中间业务系统、资产负债管理系统等。银行通过客户数据，除了提供基本的金融服务 外，还可以根据客户特性提供个性化的金融产品服务，同时也可以同步检验产品的占有率和

推 广 效 果 等 。

2)交易及运营数据：指客户通过银行各渠道完成金融服务的相关数据，包括协议数据、 介质数据、产品数据、渠道数据及事件数据。银行利用上述数据进行客户消费行为分析、金

**银行的数据** **第1章**

**表1-1** **银行信息系统的数据**

|  |  |  |
| --- | --- | --- |
| 数据来源 | 数据分类 | 说 明 |
| 内部数据 | 客户数据 | 客户基本资料 |
| 交易及运营数据 | 包括产品、介质、渠道、事件、资源等相关数据 |
| 资产数据 | 资产和负债等 |
|  |  |
| 外部数据 | 商业数据 | 宏观经济数据、金融市场价格数据、行业分析报告等 |
| 用户行为数据 | 互联网上点击数据、社交网络的朋友圈、电商平台的消费数据等 |
| 地理位置数据 | 客户所处的城市、地理空间等 |
| 供应链数据 | 银行合作方数据，如电商等合作方提供的数据 |
|  |  |

融产品创新并提供个性化的金融产品服务。

① 协议数据，即两个或两个以上当事人之间潜在的或实际的约定，协议正式提供并确 认与协议目的相关的规则和义务。

② 介质数据，即当事人获取访问权限所必须提供的访问标识 (Access ID) 及相关的认 证信息。银行通过对介质的鉴定，确定来者是谁，能替他提供什么访问权限。介质的载体称 为介质资源项，如存折、磁卡、IC 卡、条码卡等。有载体的介质称为实物介质，无载体的 介质称为数字化介质。介质的鉴定是通过渠道完成的， 一个介质可以通过多个渠道被鉴定。 认证方式不同，鉴定时使用到的认证信息也不同，同时对应的权限也不同，如信用卡介质、 存折介质、网银登录介质、柜员卡介质。

③ 产品数据，是指银行及其关联的当事人提供给市场，能满足客户(包括内部客户) 的某种需求，银行可从中赚取各种实际或潜在收益的货物(有形)与服务(无形),包括产 品代码、产品名称、产品描述、利率、费用、认购条件、额度、有效期等。

④ 渠道数据，即银行希望保留其信息，银行或第三方机构拥有的，直接或间接进行产 品服务、客户沟通、市场营销、产品销售的途径，如柜面渠道、ATM 渠道、第三方直联 渠道。

⑤ 事件数据，即银行为业务目标的实现或业务的执行而希望保留的将发生或已发生的 事情，或描述金融机构欲将保存信息当作执行任务与传导业务信息发生的事，如交易事件、 簿记事件、项目事件、沟通事件。

3)资产数据：主要指客户在银行的资产和负债数据，同时也包括银行自身的资产负债 数据，主要包括活期存款、定期存款、票据、债券、固定资产等。该类数据可用于银行内部 风险管理和资产效率提升。

**(2)外部数据**

1)商业数据：主要是指经过分析整理后的研究数据，包括宏观经济数据、行业分析报 告、竞争与市场价格数据、特殊定制数据、政府或上级部门的文件等。银行主要通过从外部 将其获取后以手工或系统的方式加载到银行系统，主要用于银行风险管理、产品研发和大数 据分析提供基础。

2)用户行为数据：主要是指用户在互联网上的行为数据，包括 APP 应用上的点击数

 **银行数据治理**

据、社交媒体和社交网络数据、电商平台的消费数据、社交网络朋友圈等。用户行为数据可 以通过电商或社交媒体等渠道获取，主要是为银行数据营销、产品设计、数据反馈、风险管 理等提供数据支持。

3)地理位置数据：主要包括银行客户使用的移动设备位置以及客户所处的地理空间数 据等。目前，该类数据的典型应用场景包括银行结合商家推出的优惠服务实施精准营销，或 者针对特殊目标人群的位置，为网点新增或撤并提供决策依据。常见的位置信息主要包括银 行客户的移动定位信息、客户邮政地址、网址、国家、地区等。当前，银行客户地理位置数 据主要可以通过通信运营商、地图服务提供商等渠道获取。

4)供应链数据：主要是指银行同合作方之间与客户交易的相关数据。由于银行自身的 供应链数据不全，无法支撑客户的供应链金融服务，因此银行通过与具有客户交易数据提供 和接收方的相关合作方(如电商等)合作，根据完整的供应链数据为客户提供全面的金融 服务。

**3.** **银行数据的特性**

目前，虽然各大商业银行基本上实现了数据集中，但是较多的仅停留在物理层面，数据 的逻辑集中程度不高，更多的还是以功能为核心，如果要很好地支持未来银行信息化建设， 必须要对支撑银行战略的数据需求及特性有一个全面的了解。

**(1)银行数据的全局特性** 从数据涉及的范围来看，数据要服务于企业战略，必须具有 全局性，即需要从企业级的视角理解数据需求，按照面向主题的管理需求对数据进行全局性 的规划。银行数据主要来源于以下多个层面的系统：银行核心业务系统、独立于核心的专业 交易系统、与业务相关度较高的后台管理系统(如信贷管理系统等)、渠道系统、其他信息 系统(如同业信息、宏观经济数据等)。由此可见，数据范围基本覆盖了银行所有业务和管 理系统，并不仅仅局限于核心业务系统。

**(2)银行数据的多维特性** 由于定位和目标的不同，因此作为管理需求的数据与业务系 统的交易数据有很大的区别。前者是要求具有多维特性，而后者往往是单维度的。为了保障 交易系统的效率，业务系统往往只提供最基础的交易流水和基本账务数据，可以说是最小化 定制的信息。而管理数据往往要求提供更多的观察视角，以适应不同管理者的差异化需求， 在账务数据的基础上，还整合了各个相关的外围系统的数据，相当于又加载了产品、客户等 更多的标识，从而形成了多维属性。例如，对于一笔贷款业务，在管理数据中既可以按照不 同的管理条线(如个人业务、国际业务等)进行展示，又可以按照所属的行业、期限、产 品、客户、地区等其他视角进行展示。

**(3)银行数据的关联特性** 业务系统一般不可能直接提供管理所需的多维度信息，往往 需要通过特殊的数据加工平台和工具对相关数据进行关联、转换和加工。管理数据的组织以 满足“管理主题”为特征，通常需要借助基础的数据平台(如数据仓库),经过复杂的数据 映射、筛选、连接、聚合等加工，横跨多个处理系统(包括其他管理信息系统)的数据。 可以说，管理数据是经过整合、提炼、派生后的多维度信息。

数据的复杂性决定了需要从多方面进行数据的管理。在数据的组织上，要选用科学的方 式，称为“数据模型”;在保证数据的正确性、有效性等方面，称为数据质量；在数据的部 署方面，需要从数据分布和主数据管理着手；在如何保证数据能够提供高时效的服务方面， 则需要考虑数据生命周期和数据交换。因此，可以看出，银行数据的复杂性决定了数据管理

**银行的数据** **第1章**

的复杂性和多面性。

**1.2.2** **数据对银行发展的重要作用**

近些年来，伴随着计算机技术和移动互联网技术的迅速发展，互联网(社交、搜索、 电商)、移动互联网(微博、微信)、物联网(传感器、智慧地球)、金融(银行、股市、 保险)、电信(通话、短信、彩信)、安全和监控等领域都在疯狂地产生着海量的数据，特 别是来自于社交媒体、电话、图片、视频等数据源的数据量已蔚为壮观。IDC (互联网数据 中心)的数据显示，2014年，数字世界的信息量增幅就超过了4ZB, 并以每两年翻一番的 速度激增。其中，非结构化数据或者半结构化数据的超大规模和快速增长，占据了数据量总 数的80%～90%,比结构化数据(如数据仓库)的增加速度快了10～50倍。

存储在银行信息系统中的数据(包括银行内部系统自身产生的和外部获取的海量数据) 隐藏着大量有价值的信息，为创新业务产品、重塑业务流程、创造业务价值等奠定了坚实的 数据基础。数据已经成为与金融资产和人力资本同等重要的商业资本，成为能为银行带来经 济利益的重大资产之一，是银行持续发展的坚实基础，是银行独有的“金矿”。

一般来说，银行日常业务处理和经营管理决策过程的背后，实质往往是数据的生产、传 递和利用过程。在信息化发展的背景下，银行的数据价值日益体现，不仅是系统建设的利 器，同时也是银行价值提升的强大基础，是银行精细化管理、实现差异化服务、提升风险分 析能力的有力保障。

**1.数据是银行信息系统建设的重要资源**

溯本求源，信息系统实质上是一个数据的输入 (I) 、 加工 (P) 和输出(O) 的工具， 而“程序=数据结构+算法”就像一个数据工厂的流水线。其核心就是数据的加工和流转， 最有价值的其实就是数据。如果没有数据，就像设备没有原料，银行的信息系统建设就是一 副空架子，产生不了任何价值。

**(1)数据是规划系统架构的“路线图”** 从企业级的高度来看，数据其实为IT 系统规 划、设计与建设提供了另外一种视角。要建立多少个系统?系统间的边界与关系如何划分? 新老系统间如何对照衔接?这一系列问题可以通过对数据的统一规划来布局，即通过合理的 数据分类来指导系统边界的划分，按照数据的流转关系来确定系统间的关联参数，并以此为 参考，构建耦合度合理、适度和优雅的系统架构。

**(2)数据是实现系统间无缝衔接的“黏合剂”** 信息系统的集中最大的难点在于逻辑集 中。要实现真正的逻辑集中，最根本的就是实现不同系统间数据的互联互通，确保数据的一 致、准确和完整。要达到这个标准，系统间的数据交流和共享无疑是至关重要的。通过研究 数据标准和规范，统一数据接口标准，可以跨越“异构”“自治”等多种复杂应用系统，实 现系统间的无缝衔接。

**2.** **数据是业务与技术沟通的共同语言**

银行信息系统建设的过程中出现的大多数问题一般不属于单纯的技术问题，而是业务需 求与技术实现出现了脱节，往往体现为双方缺乏深层次的理解与沟通，其原因就在于没有找 到可以直接进行交流的共同语言。这种共同语言其实就是“数据”,数据是对业务管理过程 的真实记录。在信息系统建设过程中，通过分析数据基本上就可以完成对业务需求向技术实

 **银行数据治理**

现的转化和解析。同时，数据还可以成为系统功能模块及流程设计的依据，通过“数据” 媒介能够促进业务与技术的有效融合，这也是保障业务意图准确转化为技术实现的一种有效 方式。

在银行经营管理的转型过程中，业务流程注定是不断变化和发展的，而其中的数据一般 则是相对稳定的。理论上来讲，任何业务需求都可以解析为实际需要的数据项，以及数据之 间的关联规则和流转过程。这些细粒度的元数据很少因业务流程的改变而发生实质性的变 化。这其中需要依靠“数据模型”对各类数据以及数据之间的关系进行高度的抽象，并将 业务运营管理中的各种规则描述清楚。

**3.** **数据是银行价值提升的“助推器”**

数据作为一种资源，最终将发挥其重要作用。通过对数据的全面分析，可以加强对银行 价值创造节点的识别与分析，并通过信息化手段推动业务流程的再造和产品的创新，进一步 促进科技与业务的融合，提升应用系统的效率、服务和管理价值，主要体现在以下方面：

**(1)充分了解客户需求，提升客户服务满意度** 洞察和理解客户，是商业银行发现商 机、营销产品的必要条件。完善的客户数据，有助于商业银行对客户进一步细分，更好地了 解和把握客户的具体需求，从而提供个性化、差异化的金融产品和服务，最大限度地提升客 户满意度。

传统商业银行进行客户细分的主要依据是客户的年龄、性别、职业、财富水平等维度。 基于此类维度的细分进行产品营销和设计很容易造成“误伤一片”,导致交易成功率较低， 造成不必要的资源浪费。我国有关机构通过对美国商业银行发展过程的研究，发现目前美国 商业银行信息技术的竞争主要体现在两个方面，分别是银行的业务需求同信息技术的结合程 度以及客户关系的管理水平。我国正在经历信息系统的大量应用和开发阶段，随着时间的不 断推移，各大商业银行在开发层面和设备技术投入方面的差距将逐渐缩小，而通过数据充分 地了解客户个性化、差异化的需求，从而提供智能化的客户服务，是加强客户关系管理和业 务需求提炼的有效途径，最终也必然成为商业银行信息科技的核心竞争力。

同时，数据挖掘和分析也有助于银行保留并防止丢失老客户，挖掘和拓展新的客户，增 强客户对银行的黏性。例如，澳大利亚某大型商业银行通过对客户财富结构、购买偏好、与 竞争对手客户结构差异等数据进行分析，为银行的小微企业客户提供了一项免费的增值服 务，从而提高了客户黏性。

**(2)实现精准营销，降低营销费用** 收集并分析客户交易数据，可以推动商业银行有 针对性地将金融产品和服务推荐给特定客户群，收集客户反馈意见，提高营销成功率，降低 营销成本。在往常的银行柜面营销理财产品过程中，理财顾问或者客户经理需要不断地询问 客户的需求，然后根据客户的描述，推荐合适的理财产品。这种营销方式缺乏针对性、效率 不高。如果理财顾问或客户经理掌握了客户相关的分析数据，如客户往常的理财产品购买情 况、资产负债情况、客户行为分析等，则可以根据客户实际情况有针对性地开展精准营销、 交叉营销，提升产品销售的成功率，从而降低营销费用。

例如，某国外银行通过对客户交易行为数据进行分析，发现了一个占比不大(约7%) 却很有意思的客户群体，这类客户群的财富水平不高，达不到银行的贵宾门槛，所以常常被 银行忽略。但是，这类客户有个特点，就是交易行为非常活跃，他们的消费习惯能够为银行 带来可观的价值。银行成功根据客户交易数据分析，精准定位这一类客户群体，并提供了针

**银行的数据** |**第1章**|

对性的金融产品和服务，实现了精准营销。

又如，某国外大型银行通过对客户行为数据的分析为自己的企业客户提供营销支持。他 们在为一个手机零售商户分析了其客户在购买手机前后的其他购买行为。发现客户在购买之 前出现频率最高的地方是交通枢纽，而购买之后则最可能出现在食品杂货店中。这样的分析 帮助银行明确营销的最佳地点，从而对客户资源配置产生积极的影响。

**(3)加快金融产品和服务创新** 经过多年的信息技术大发展，当前我国商业银行的管 理水平得到了极大改善，提供的金融产品和服务也越来越丰富，在便利并满足客户各类需求 的同时，也大大提高了银行的竞争力。例如，电子商务的发展使网上银行业务能够提供个性 化的金融服务，从而克服时间、地域上的限制，高质量地服务社会，促进社会进步。信息技 术的发展一方面使数据积累成倍增长，数据资产的价值作用也更加突出，另一方面则使得信 息的传递更加高效和快捷，对数据的快速反应和使用提出了更高的要求。我国的商业银行应 该充分利用信息技术积累的海量客户数据，通过加强对数据的治理、分析和挖掘，以数据来 驱动金融产品的创新，提高产品与服务的质量；同时通过数据分析来增强对市场客户的需 求，特别是高价值客户需求变化的敏感性、洞察力和预见性，从而强化数据对各项业务创新 的支撑作用，坚持有所为、有所不为，争取在特色产品和局部业务领域确立商业银行的优势 地位。另外，我国商业银行应该加强与社交平台、媒体、电商等机构的大力合作，通过对客 户在银行外部的数据进行挖掘和分析，充分了解客户的特点和个性化需求，从而提供有针对 性、个性化和智能化的金融产品和服务。

随着市场竞争的日趋激烈，要求银行更贴近客户的实际需求，快速推出满足客户实际需 要的金融产品和服务。不管是银行内部产生的数据还是从外部渠道获取的数据，都提供了一 个了解、挖掘客户真实需求的窗口，这个数据窗口为银行提供了一个满足客户需求、快速实 现金融产品和服务创新的广阔天地。

**(4)实现业务流程整合，降低管理和运营成本** 商业银行可以利用数据有效整合分割 的金融业务流程，实现整个业务链条的整合，进而降低管理运营成本。有报告数据显示， 2013年各银行金融机构就开始根据本行的实际情况，大力推进前台、中台和后台的流程再 造，不断优化服务流程，使服务更加快捷与流畅。我国商业银行已经通过对数据的整合、挖 掘和分析，在充分了解客户需求的基础上，依托蓬勃发展的互联网平台和现代信息科技手 段，大力进行流程整合和流程再造工作。据不完全统计，2013年全年涉及主要业务流程优 化的项目达到120个以上，柜面业务平均办理时间明显缩短，在客户的金融服务体验显著改 善的同时，银行的管理和运营成本却在持续下降。

**(5)助力风险管理** 风险表现为损失的不确定性，风险表现出来的只有损失而没有获利 的可能性。商业银行的数据是助力银行进行风险管理的基础，而数据的不正确、不及时、不 完整都是造成风险的潜在隐患。银行只有尽可能多地了解、掌握、挖掘和分析客户的各种数 据，才有可能减少不确定因素，才有可能发现客户交易过程中的一些隐性风险，对潜在风险 进行识别，对虚假、欺诈行为进行识别。银行对客户数据的采集、整合、挖掘与分析的能 力，决定了银行风险管理能力，也是银行竞争力的重要组成部分。

商业银行需要充分发挥数据在银行风险管理中的作用，整合客户交易数据、行为数据和 位置数据等信息，用于客户信用评级、分析风险事件、预防欺诈等。新巴塞尔协议对银行数 据的深度、广度及完整性提出了更为严格的要求，通过长年数据积累和掌握全面、权威、合

 银行数据治理

规的风险基础数据，并通过数据仓库以及风险数据集市为风险偏好计算评估提供高效、准 确、整合的信用风险、市场风险、操作风险以及计划财务的数据支持。

从外部环境来看，近年来经济不断下行、互联网金融遭受冲击，为银行信誉风险管控带 来了更大的挑战；从内部来看，银行信贷规模不断扩张，客户结构也在发生着深刻变化。外 部和内部的各种挑战要求银行以先进信息技术架构为前提，以大数据挖掘和分析为基础，加 快推进全行信用风险管理体系由经验管理向数据分析管理、由分散管控向集中管控进行转 型，以有效提高风险感知水平，降低不对称风险。

**(6)优化资产配置，提升银行竞争力** 数据是银行定量分析和精细化管理不可或缺的 基础，通过了解客户及自身的资产负债数据，调节产品配置，合理利用资源，降低流动性风 险，提高资本利用率。同时，银行发展模式转型、金融创新和管理升级等都需要充分利用数 据进行支持。

**(7)有利于保障宏观经济稳定性** 金融业是我国宏观经济稳定的基石，建立全面严格、 监管到位的金融监管体系离不开各类商业银行。而严格全面的银行业监管措施和信息披露要 求都必须基于数据的有效管理，涉及内容广泛，信息量大，如资产负债表、利润表、经营管 理资料等，囊括客户、资金及分润等多领域数据，时间要求紧，质量要求高，必须通过维护 并梳理数据才能达到报送要求。

中国银行业监督管理委员会成立以来，逐步建立和健全了涵盖信用风险、市场风险、操 作风险、流动性风险的全面动态监管体系，商业银行按照非现场监管的数据报送要求，对底 层交易数据进行分类、加工、汇总，形成各项监管数据上报，并从体制、机制、流程、系统 等方面保障数据质量。但是，当前银行监管报送数据普遍面临“数出多门”、数据源质量不 高等问题。因此，更需要银行加强数据治理，加强统一标准的建设，保障报送监管数据的 质量。

**1.3** **银行数据管理存在的不足**

当前，我国的商业银行正处于战略转型的关键时期，银行需要探寻新的发展模式，逐步 实现从“银行信息化”向“信息化银行”的转变，银行业务的电子化、产品的多元化、服 务的个性化、渠道的网络化、经营管理的信息化以及核心竞争力的提升都离不开“数据” 的强力支持。信息化银行建设、银行精细化管理及业务的创新决策，都对银行数据的准确性 和时效性提出了更高的要求，亟待加强数据治理，深化数据应用，有效支撑银行业务发展， 有效提升银行管理水平，实现数据对业务的价值创造。

当前激烈的利率市场化竞争、金融监管改革的深化、互联网金融的冲击、经济结构的调 整以及大数据的快速增长，都给商业银行带来了巨大挑战和机遇。从国外的 ING Direct ( 美 国最大的网上银行)、Zest Finance (美国互联网金融公司)到我国的蚂蚁金服、前海微众银 行、百度金融、网易宝、人人贷等企业对数据的应用，正深刻地改变着当前的金融生态和金 融格局。

正是在这种内部变更需求和外部竞争压力的双重驱动下，大型商业银行和股份制商业银 行、城市商业银行以及农村商业银行纷纷借助 IT 信息系统推动数据治理体系的建设，提升

银行的数据 **第1章**

企业数据质量。虽然已经有了一定的数据治理成果，但是从整体上看，我国商业银行的数据 管理仍存在以下问题。

**1.** **数据多头管理，缺少专门对数据进行监督和控制的组织**

多数商业银行的系统建设和管理职能分散在各个不同的部门，致使出现数据管理职责分 散、权责不明确的情况。由于各个部门关注数据的角度不尽相同，因此缺少一个组织从全局 视角对数据进行管理，导致无法建立一个统一的数据管理规程与标准等，相应的数据管理监 督措施也就无法得到落实，并且由于数据考核体系不完善，导致数据管理标准和规程无法有 效执行。同时，各部门为应对迅速变化的市场和社会需求，逐步建立了各自的信息系统，各 部门站在各自的立场生产、使用和管理数据，使数据分散在不同的部门和信息系统之中，缺 乏统一的数据规划、可信的数据来源和数据标准，导致数据不规范、不一致、冗余、无法共 享等问题频繁出现。

**2.** **数据不够完整，具有一定片面性**

尽管银行经过多年的信息系统建设累积了海量数据，但是受业务场景、技术水平及组织 管理等限制，银行数据仍具有一定的片面性，主要体现在以下几方面：

1)银行所获取的数据是客户想展现的数据，而不一定是全部的数据，是有附加条件的、 具有一定片面性的。以银行贷款和信用卡审批为例，客户展现给银行的都是有利于审批的信 息，而客户在此过程中很可能隐藏了诸如家庭变故、自身恶习、信用状况等负面信息。

2)未形成全面有效的数据模型管理，导致系统设计过程缺乏一定的前瞻性。信息系统 在设计时，前瞻性考虑不足或系统性能考虑不周全等原因会造成表结构设计的规范化、灵活 性不足，信息字段登记缺失等，也是数据不完整的主要原因之一。

3)数据采集缺少沟通与协作。数据收集工作并非某个部门的少数人员能妥善完成，而 需要各部门之间、各层级之间相互支持与协作，互相认可，才能最终高质高效地完成。

4)外部数据获取不足。当前银行存储的数据主要是由银行系统自身产生的，银行外部 数据相对缺乏，导致客户数据不完整，不能全视角去衡量客户特征。现今社交网络、论坛、 电商等包含着大量的客户数据，而目前银行则缺乏有效的手段或渠道来获取这些客户的信 息，难以建立完整的客户信息视图。

例如：传统的信用卡审批一般依赖于线下数据，依靠人工进行审核，但线下数据成本较 高而且信息量有限，央行的征信系统也仅仅能查到个人的违约信息，没有违约记录就没有其 他行为信息，缺乏对互联网客户的特征把握，往往导致客户审核通过率比较低或者审核额度 不能够真实反映客户价值和需求。互联网金融的发展使得线上、线下信息能够互相结合，利 用互联网上的社交信息、电商平台上客户的消费信息等，可以识别不同客户的交易行为，提 炼和分析客户的价值，从而采用多元化的信用评估手段，真实地还原客户信用等级，建立起 完整的客户信息视图。

商业银行与互联网公司逐渐意识到自身数据的不够全面，纷纷结合自身优势，迈出跨 界、转型的步伐，充分利用多方位数据进行产品营销、风险管理等业务活动。2014年1月 中国工商银行电商平台“融e 购”正式上线，定位于打造消费和采购平台、销售和推广平 台、支付融资一体化的金融服务平台，以及用户流、信息流、资金流一体化的数据管理平 台，是工商银行获取完整的客户数据的最新尝试。2014年7月，中国银行、招商银行、中 国建设银行、平安银行、中国邮政储蓄银行、上海银行、兴业银行7家银行与阿里巴巴集团

 **银行数据治理**

携手合作，银行提供资金，阿里巴巴集团提供贷款企业的订单、交易、通关、物流、退税、 外汇等数据，通过对企业客户上下游交易链的完整数据分析，推出全新 B2B互联网金融产 品，为中小企业提供无抵押信用贷款，使银行获取外部互联网社交数据并在风险管理中进行 应用。

**3.** **数据不一致，缺乏统一标准**

实际工作中经常会遇到这样的现象：不同部门的领导做工作汇报时，同一指标的数据相 去甚远，甚至截然相反，到底哪个数据准确?无从确认。造成这种现象的原因很多，数据标 准不统一就是一个重要原因。

银行经常会面临部门之间、系统之间数据不一致、统计口径和加工方法不一致而导致数 据可信度大幅降低的问题，归根结底就是缺乏有效的数据标准而造成的。数据标准化是数据 环境建设中的一个重要环节，通过数据标准可以保障基础数据的一致性和严密性。只有合理 制定标准并严格执行数据标准，才能确保各应用系统标准的一致性。数据标准作为数据的规 范化定义和统一解释，能够增强业务部门和技术实施部门对数据定义和使用的一致性；通过 促进系统集成，充分实现信息资源共享，并作为信息系统开发时进行数据定义的依据，使得 企业能够从提升管理能力的数据需求出发，建立起数据共享和信息交换的平台。

**4.** **数据分布零散，部署和同步原则需要进一步完善**

随着业务发展及应用系统的建设，各大银行依靠科技手段实现的数据大集中工作已初见 成效，通过将分布在各个分支机构和营业网点的数据进行集中管理，为提高银行核心竞争 力、应对激烈的市场竞争提供了广阔的平台，也为银行改革管理、产品创新、服务提升和业 务发展提供了基础。

由于专业化管理的需要，商业银行具有为数众多的应用系统，每个应用系统都存储有不 同的业务数据，数据分散在不同的应用系统中。源数据(也称为基准数据或主数据)分布 原则不够完善，缺乏源数据的生成、管理、维护机制来保证数据的正确性和权威性，缺乏数 据同步原则，导致难以及时、准确地提供完整、可靠和一致的企业级信息，是数据大集中后 我国银行普遍面临的问题。

例如，生产上有些客户的账户和协议状态显示为正常，但是客户ID 信息已经作废了， 原因就是有些账户和协议在开立时没有同步在客户信息的系统中进行登记，作废客户ID 信 息的交易在作废时检查到没有存在正常的账户和协议，从而允许作废，导致客户ID 信息虽 然作废了，但是账户和协议仍然正常的现象。

因此，只有制定长效管理机制，正确维护源数据，才能保证交易系统的参照完整性，只 有制定源数据和数据副本(也称为辅数据)的数据技术分布原则和数据同步原则并加以指 导，才能确保数据提供方和接收方的各系统间源数据和数据副本的一致性。需要对相关应用 展开全面检查，针对存在的问题制定整改方案，并对整改计划进行跟踪。

**5.数据共享与整合程度不足**

面对日益庞大的数据量，如果缺乏有效的数据共享方法，系统之间势必造成一个个的 “信息孤岛”,因而数据无法得到充分利用。只有通过有效的数据交换共享机制，对共享信 息进行科学的管理与合理的利用，才能为客户服务、统计分析等提供有效的整合数据。

部分银行已经意识到，如果没有信息的整合共享，就无法满足多方位的业务活动的开 展，特别是近年来我国商业银行越来越多地面临实时或者准实时的数据共享需求，如准实时

**银行的数据** **第1章**

营销等。因此，有些银行开始着手制定统一的数据交换传输体系，实现数据的及时传输与整 合，从而为各类业务开展提供及时完整的数据。

**6.** **数据正确性有待提升**

数据正确性是银行数据分析结论有效性和准确性的基础，也是数据可靠性的重要前提和 保证。数据正确性存在问题是一种信息失真的表现，它的产生是多方面因素综合作用的 结果。

错误数据的产生可能源于多个环节，如业务需求不断迭代和更新，会导致原先采集的数 据信息不能够满足后续的需求；又如在系统使用过程中，柜员输入数据时输错客户姓名等客 户信息。同样的问题也会出现在技术实施方面，如在进行系统设计时，对元数据描述出现错 误的理解，数据处理的技术环节出现设计不当，就很容易造成数据不正确，并且会随着数据 的传输造成错误数据的不断扩散。

数据正确性问题在不同程度上会对银行日常业务的有效开展造成较大影响，从而导致客 户满意度下降，同时也会降低统计和分析数据信息的可用性。因此，必须从数据产生的各个 阶段入手进行治理，防范信息出现失真，有效提高数据质量。银行的数据质量管理涉及预 防、识别、度量、分析、监控、清洗等多个方面，是个循环管理的过程，必须通过改善和提 高组织管理水平，提升人员技能，制定相匹配、完善的管理流程来实现。目前，我国不少商 业银行缺乏专业的质量管理团队和相配套的规章制度，缺乏统一的包括问题提交、跟踪、治 理、关闭在内的闭环管理流程，缺乏明确的数据认责关系和跟踪系统，无法保证数据质量， 无法做到在数据问题根源的地方解决问题。

**7.缺乏数据生命周期管理**

世间万物都有生命周期，数据也不例外。即使数据不会消亡，其也会拥有有效性和有效 周期。随着信息科技系统的运行和时间的积累，数据不断增长，规模越来越大，部分数据对 业务处理及系统当前、将来的运行、监控、经营分析等所起的作用会越来越小，如果仍然保 留在系统内，反而会徒增数据库压力，增加系统性能开销，降低访问效率，使系统反应变 慢。因此，银行需要对数据开展生命周期管理，落实清理策略及清理工具的规划、部署，将 应用系统中各种数据存放在适合的存储设备上，利用适当技术手段对这些数据进行处理、分 析、管理。

通过对数据进行生命周期管理，商业银行可以优化应用系统的存储结构，有效控制在线 数据的规模，提高生产数据访问效率，减少系统资源浪费，提高应用系统运行的整体效率和 效果，保证应用系统健康高效运行。

**8.数据挖掘与应用力度不足**

数据来源分散和冗余，数据的产生、捕获、整合、存储、访问等缺乏完善的系统支撑和 技术手段等，使得难以深入挖掘埋藏于数据深处的规律和趋势，无法实现数据价值的最大 化。特别是非结构化数据，目前多以简单查询为主，未对非结构化数据资源进行统一管理、 挖掘及综合利用，非结构化数据在营销、风险管理、产品创新等方面的价值未得到充分 发挥。

例如，美国富国银行通过挖掘客户需求，采用交叉销售的模式为企业利润的创造做出了 巨大的贡献。富国银行通过满足客户在财务上的全部需求，与客户共同成长，利用交叉销 售，向一个公司客户推销5.3个金融产品，向个人客户平均推销4.6个产品，大大提高了银

 **银行数据治理**

行对客户资源的利用能力，实现了效益的最大化。

又如，全球最大的信用卡公司VISA, 通过使用Hadoop (一个对大量数据进行分布式处 理的开源式基础框架，能充分利用集群威力进行高速运算和存储，非常适用于半结构化和非 结构化数据)分析数据，大规模缩减了分析处理的时间，提升了数据使用效率。VISA将自 己拥有的全球最大的支付网络系统 VisaNet作为信用卡付款验证。为降低信用卡诈骗、盗领 事件带来的损失，VISA公司需要分析每一笔事务数据，找出可疑交易。虽然每笔交易的数 据记录只有短短200位，但每天要处理上亿笔交易，而通过对Hadoop 的使用，成功地将分 析时间从1个月缩短到13min, 更快速地找出可疑交易，更迅速地对银行提出预警，及时阻 止诈骗交易的产生。

再如，苹果公司通过与移动运营商合作，收集用户体验数据，不断完善产品功能，提升 用户满意度。数据从被交易的对象到重要的金融资产，在金融领域发挥越来越重要的作用。 通过新技术的不断利用，深入挖掘数据背后的潜在价值，是银行发展、创新的必经之路。

**9.数据安全管理亟待加强**

近年来，各国银行信息安全事件频发，大量用户信息被泄露，引起了社会公众的广泛关 注。金融机构在个人信息数据安全保护方面亟待加强。

2013年末，在美国第二大零售巨头塔吉特 (Target) 发生的数据泄露事件中，约有 4000万 Target 超市的用户数据，包括用户名、信用卡号、信用卡使用期限甚至信用卡确认 码丢失，导致1/3的美国人受到波及，大量用户信用卡被盗刷。

2014年初，韩国金融机构爆发的个人信息泄露事件涉及金融信息多达1亿多条，导致 2000 多万的信用卡用户受到牵涉，从国家政要、企业高管、社会名流到普通民众都未能幸 免。但金钱损失还不是最让人担忧的，因为外泄的个人信息非常详细，包括户名、居民身份 证号码、电话号码、银行账号、年收入、居住情况等大量敏感信息，即使重新办卡也只能更 换密码和安全账号，无法改变证件号码等其他信息，民众对随之而来的诈骗担忧在韩国引发 了社会恐慌。

银行客户数据泄露，不仅会给银行带来经济上的损失，而且银行可能面临因数据泄露而 带来的巨大信誉风险和法律风险，更为严重的情况是还可能会影响国家经济正常秩序，造成 巨大的损失和混乱。如何通过制度、流程和技术手段保护数据已经成为现代商业银行必须面 对的一个重要课题。



**第** **2** **章**

**银行数据治理概述**

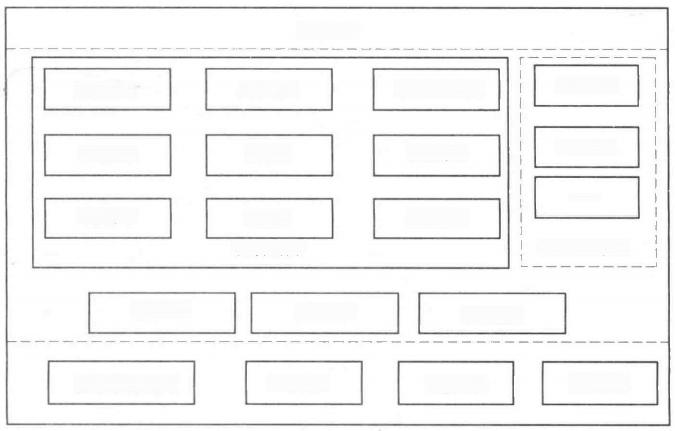
国际上对数据治理比较通用的定义是：数据治理是对数据资产管理行使权力和控制的活 动集合(规划、监控和执行)。数据治理的职能是指导其他数据管理职能的执行。因此，数 据治理是在高层次上执行数据管理制度。

从这点来说，数据治理是数据管理的 一 大职能(甚至是核心职能),是规划、监督和控

制数据安全管理、数据质量管理、元数据管理、数据生命周期管理等领域的数据管理。

我国一般将数据治理和数据管理进行综合考虑，认为数据治理是将数据作为组织资产而 展开的一系列的具体化工作，包括从组织架构、管理制度、操作规范、信息技术应用、绩效 考核支持等多个维度对组织的数据模型、数据架构、数据质量、数据安全、数据生命周期等 方面进行全面的梳理、建设以及持续改进的过程。也就是说，数据治理贯穿于各个数据管理 方面的具体工作之中。

图2 - 1 所示为银行数据治理框架。



数据治理

数据生命周期

数据安全

… ·

主数据

结构化数据

技术规范

战略目标和规划 组织架构 制度章程 流程管理

非结构化数据

数据交换

技术平台

数据模型

数据模型

技术工具

数据分布

数据安全

数据标准

数据质量

元数据

图2 - 1 银行数据治理框架

 **银行数据治理**

**2.1** **银行数据治理的背景**

**1.数据治理是银行自身发展的迫切要求**

毋庸置疑，数据已经成为商业银行参与竞争的重要武器。在从银行信息化到信息化银行 的转变过程中，商业银行的发展模式从规模驱动的模式转变为更加注重集约式、以效益和质 量为核心的新模式。这个转变必须依靠全面的管理提升和内部挖潜，而数据是定量分析和细 化管理不可或缺的重要基础。

随着全球一体化经济的不断深入发展，商业银行作为经营风险的信用中介正在发挥着日 益重要的作用，同时，也面临着信用、市场、操作等各类风险的严峻挑战，而银行应对这些 风险的能力直接影响着金融市场和金融秩序的稳定性，从客观上要求银行不断提升自身的风 险管理水平。随着利率市场化、互联网金融、金融脱媒的不断推进，金融市场的有序开放成 为金融体制改革的主流，倒逼银行必须进行自身转型，走特色化、差异化的发展道路，要求 银行充分利用数据这一利器提升自身精细化管理水平。

银行在长期的经营活动中积累了大量的数据，这些数据除了支持银行业务处理流程之 外，越来越多地被用于风险控制、决策分析、绩效考核、市场定价等管理领域，如果数据存 在问题，并且没有有效的治理机制，将会导致错误数据如雪球般越滚越大，导致相关领域业 务无法正常开展，从而导致决策出现偏差。因此，银行加强数据治理工作势在必行，只有切 实落实好数据治理工作，才能实现银行数据向银行知识的价值升华，真正成为银行提高经营 管理水平、应对金融市场挑战的助推力。

如何保证从海量增长的数据中、从各种问题缠绕的数据中为业务或决策提供及时、准 确、完整、可靠的数据?数据治理即为解决之道。数据治理将使银行的业务和战略实现 “内外兼修”,增强数据洞察力，实现以“数据驱动”为战略的增长模式。

**(1)数据治理是银行运营安全的需要** 数据是银行经营的重要组成部分，银行需要安 全地保管自身及客户的交易信息，避免将其泄露或被他人非法篡改，给银行带来不可估计的 损失。同时，数据一致性、完整性和可用性对于保障银行各项业务有效、连续地开展也非常 关键。

**(2)数据治理是银行风险管控的需要** 在银行的风险管理方面，需要密切关注贷款 客户等数据的变动，需要通过对数据进行有序集成、合理分析，确保金融市场交易的正常进 行，还需要依赖数据的治理措施和成效，为业务部门设定业务限额，确保发挥业务价值的最 大化，提高银行资产质量，减少非预期损失。

**(3)数据治理是银行业务创新的需要** 随着金融全球化的不断深入，银行之间的竞争 也越来越激烈，银行需要对历史的业务数据进行挖掘、分析、比对，对数据质量、数据分 布、数据服务进行相关治理工作的落实，通过创新业务模式的发展思维，在传统的业务运营 基础上推出各种中间业务、理财服务等新型业务，进而提高客户体验，提升银行竞争力。

**2.数据治理是政策和监管的要求**

随着社会信息化建设的不断深入，数据作为政府及企业最重要的资产之一，是准确识别 客户，洞察市场需求，强化业务创新，掌握客户风险的基础和前提。随着全球经济一体化和

**银行数据治理概述** | **第2章**

金融自由化不断快速地推进，完备、持续的数据体系建设已经成为银行业可持续发展的生 命线。

我国政府一直非常重视数据的发展和风险控制，通过一系列法律法规、方针政策的制 定，促进了数据治理的研究和进步。近几年，国家安全生产监督管理总局制定并下发了关于 安全生产监督管理信息的《隐患排查治理数据规范(试行)》《工贸企业安全生产标准化数 据规范(试行)》《生产经营单位基础数据规范(试行)》《全国安全生产监管监察机构代码 编制规则(试行)》等4项指导性技术文件，用于指导和规范企业隐患自查自报信息在国 家、省、市(地)三级安全监管部门间的共享和交换，保障了各级各部门业务信息系统间 数据互联互通的有效性和及时性。

同时，对数据的管理和应用能力已经被各家银行视为核心竞争力。随着互联网金融建设 规模的不断扩大，在关注系统功能的同时，如何确保数据的完整、准确、及时已经越来越多 地受到重视，如何为经营管理营造健康的数据环境已经成为各家银行关心的问题。银行业未 来的发展必将由经验式决策模式转型为数据式决策模式，实现精细化经营管理和可持续发 展。积极推进银行数据治理，提升数据质量，逐步拓展数据来源范围，充分利用内外部数据 资源，是不断提升银行经营管理水平的源泉。

银行信息系统的建设大都经历了一个从分散到集中、从局部到整体的过程。通过多年的 整合、梳理，目前银行已经形成了较为统一的数据管理体系，但仍存在着不少的数据质量问 题。例如，无法满足数据的一致性和完整性要求，无法满足合规性、唯一性、有效性要求， 无法满足及时性、可获取性、准确性要求等。随着对数据价值认识的不断深化，几乎所有的 银行都在数据集中、信息系统建设、内部制度建设等方面加大了设备和人力投入，把数据治 理提升到了一个新的战略高度。同时，中国人民银行、中国银行业监督管理委员会等金融管 理部门也对银行的数据管理提出了更高的要求，在银行的数据基础、数据标准、数据服务等 方面都出台了详细的标准和规范，要求各家银行切实加强数据管控，提升数据质量。

2007年，中国银行业监督管理委员会下发《中国银行业实施新资本协议指导意见》,明 确了实施新资本协议的重大战略意义、目标、原则、范围、方法及时间表，并指出推进新资 本协议实施的关键是做好各项基础工作，要求商业银行要加快数据清洗和补录工作，建立符 合新资本协议要求的完整、严格、 一致的数据标准和相应的数据处理平台，制定数据质量管 理规章，确保数据的及时性、准确性和全面性，并特别指出强化数据基础工作。

2011年，中国银行业监督管理委员会在《银行监管统计数据质量管理良好标准》中明 确了15项原则、61项标准，要求银行从基础数据的管理抓起，持续从组织、制度、系统、 流程等方面加强数据质量管理，并通过检查和评估等措施来监督银行数据治理工作。其中， 中国银行业监督管理委员会着重指出，随着非现场监管信息系统和客户风险统计信息系统不 断完善，银行数据及时性和全面性基本可以得到保证，但在数据准确性上存在较大差距。要 求各家银行加强数据标准化建设，银行数据标准化规划要统筹考虑监管统计数据标准化，全 面覆盖监管统计数据及指标。数据标准要符合国家标准化政策及监管统计规定，银行与监管 统计相关的业务领域及系统应遵循统一的业务规范和技术标准，建立全行统一的监管统计数 据结构，客户、产品、机构、账户等基础信息在跨系统时要保持名称、定义、口径、来源等 方面的一致性。各系统应该建立统一的数据交换标准，实现系统间信息共享。

2011年，中国银行业监督管理委员会在《中国银行业信息科技“十二五”发展规划监

**银行数据治理**

管指导意见》中明确指出，目前商业银行数据应用工作较为滞后，缺乏统一的数据标准， 大部分银行尚未建立适应经营管理、外部监管和信息披露要求的企业级数据平台，风险 管理数据分散在各个应用系统之中，缺乏统一的规划、部署与集中管理，报表可审计、 可追溯性不足，风险分析、计量工作主要依赖于人工处理，缺乏有效的技术支撑，降低 了风险分析的有效性。在此，各家银行需要重点加强数据治理工作，继续深化数据应用， 整合分布在各应用系统中的客户信息，逐步形成全行统一的客户信息视图，建设操作型 客户信息管理系统；逐步探索建立全行统一的客户营销和客户关系分析管理平台，深度 挖掘客户数据资源，实现差异化客户营销，提升金融服务效率和服务水平。同时，需要 重视数据治理的制度建设和流程建设，从建立和完善数据治理制度体系，规范工作流程， 理顺内部协作关系等方面进一步加强，通过制定数据标准，统筹数据管理，梳理、整合 数据资源，提升数据的准确性、完整性和一致性，提升数据共享水平，通过建立元数据 管理、数据安全管理、数据传输与分发等管理流程，涵盖数据生命周期各个阶段，提高 数据管控能力，进而提升全行数据质量和数据应用水平，提高数据价值创造能力。

中国银行业监督管理委员会下发《商业银行资本管理办法(试行)》,明确了商业银行 资本充足率计算及计量要求，并明确要求商业银行建立数据仓库、风险数据集市和数据管理 系统，以获取、清洗、转换和存储数据，并建立数据质量控制政策和程序，确保数据的完整 性、全面性、准确性和一致性，为资本计量和内部资本充足评估等工作做好支撑。

中国银行业监督管理委员会在《中国银行业实施新监管标准的指导意见》中强调了数 据基础的重要性，提出通过新监管标准的实施，强化数据基础，切实解决我国银行业金融机 构长期存在的数据缺失、质量不高等问题，以持续改进银行业风险管理。

2012年，中国银行业监督管理委员会制定了《中国银监会监管数据标准化规范(中 小银行及农村金融机构)(试行)》,实现其现场检查的标准化、规范化、体系化、制度化 和流程化，使得现场检查能够对风险和问题“准确定位”“精确制导”;推动银行业金融 机构加强自身信息科技建设，提升数据管理和数据治理水平，优化业务流程和风险管理 架构。

中国人民银行先后下发了《金融业机构信息管理规定》 《贷款统计分类及编码标准》 《存款统计分类及编码标准》等多项标准，牵头组建了由多家银行组成的金融统计标准化工 作小组，加大数据标准化的工作力度，提高数据治理的水平，并于2014年发布了《银行数 据标准定义规范》。该规范着重指出，随着我国银行业的迅速发展，数据标准已经成为各银 行业金融机构普遍关注的问题，其中，规定了银行业务金融机构对数据标准的统一定义与规 范，提出了银行数据标准定义框架以及各属性规范。该规范的发布和实施，对提高银行业数 据管控水平，规范银行业数据标准的定义，统一银行业数据属性的描述，促进银行间数据共 享等方面具有重要的指导意见。

2014年，中国人民银行还发布了《统计数据与元数据交换 (SDMX) 标准》,主要介绍 了统计人员在采集、处理和交换统计信息时所使用的统计概念和方法，规定了对外披露统计 信息时统计数据的机构范围、地理区域、存流量性质、时间属性、频度以及文件格式等内 容，主要适用于金融统计中数据和元数据的交换和共享，同时，规范了我国金融统计信息的

**银行数据治理概述** **第2章**

处理、交换和对外发布流程，促进了监管部门、金融机构等相关部门间的互联互通、信息共 享，有助于我国银行业数据治理整体工作的开展。

**2.2** **银行数据治理的目标与价值**

随着全球信息化的迅猛发展和我国商业银行多态化经营程度的日趋加深，银行在客户挖 掘、客户营销、产品创新方面越来越倚重数据挖掘和分析，良好的数据治理可以帮助银行加 强风险控制、提升运营能力，及时为关键业务能力的提升以及跨区域经营提供有效的支持。 从某种意义上来看，银行的资产其实已经不是现金等实物，而是“数字”。与此同时，监管 部门也对数据治理提出了更高要求，数据治理越来越成为我国银行业的一个热点话题，并且 愈加受到人们的关注。

近年来，银行逐渐意识到数据不仅可以帮助银行进行日常业务运作和管理，更重要的是 数据可以作为战略资产，与业务运作和管理决策相结合，帮助银行提高业务流程效率、财务 绩效并可以促进产品创新。伴随着这种理念的深入，银行信息化正呈现出新的趋势，即信息 化建设重点从业务流程的自动化和信息化，转向基于数据及分析的数量化和精细化管理，其 特征表现为银行在大量积累的业务数据基础上，思考如何通过数据分析来优化业务流程，提 升管理决策水平。

随着银行不断深化以客户为中心的转型，银行在客户、产品、渠道、创新、营销等方面 的竞争压力越来越大。银行要应对这种挑战，就必须充分挖掘客户、合理营销、随客户需求 创新。同时，银行业务转型的一个重要机制是银行内部有功能完善的业务流程，具备高度差 异化的产品功能与定价功能，而合理的业务流程需要各应用系统之间良好的互联互通和数据 信息共享。银行对现有客户的业务发掘和拓展新客户，都离不开数据的支持。

巴塞尔新资本协议对数据治理的内在要求，使银行的数据体系能够明确解决“需要什 么数据”“为什么需要”“如何获取”“怎么应用”等一系列问题，真正实现基于数据基础 的经营决策分析和风险管控。

2008—2009年，某国有大型银行针对其在信息大集中趋势下亟待解决的问题，逐步构 建了数据治理架构：成立专门机构，建立数据中心，负责制订制度化管理规范和办法；建立 统一数据标准，明确数据标准提出、修改、退出、冲突的流程，使得数据遵循全行统一语 言；建立强制性管理制度，通过行级决策，规定银行内部各系统建设过程中必须融入数据标 准，必须对其是否符合数据标准进行评审，不符合标准可一票否决；推动数据仓库建设，整 合规划各系统数据，对数据标准按主题等维度进行区分；进行数据质量管理，定期检查数 据；在数据仓库基础上进行分析应用，并在多方面取得了显著的成效。银行数据治理的价值 主要体现在以下几个方面：

**1.** **更好地服务于银行决策**

在良好的数据治理环境下，未来管理层需要借助数据分析，实现敏锐的业务洞察能力， 更好地适应经营过程中的不确定性因素，提升银行竞争力。目前，有些银行开始通过数据挖 掘与分析形成预测性结果，为高级管理人员把握战略方向、提升业绩水平提供更准确和富有

**银行数据治理**

前瞻性的决策依据。

**2.更为精准地为客户服务**

随着产品同质化的加剧，许多银行在客户管理方面已经超越产品设计，把关注点集中在 了通过丰富客户个性化的服务体验来提升自身差异化的竞争力。

1)通过对客户进行智能化的细分，可以将现有和潜在的各类客户细分为不同的差 异化子群，并根据其不同子群的特征使用差异化产品组合和营销策略，提高营销的精 准度。

2)通过详细和准确地分析客户行为，可以帮助银行更有效、更及时地给客户提供相应 服务，并在客户发生关键变化后及时进行干预。

3)通过对产品定价和客户贡献度进行分析，能够帮助银行更好地了解利润在不同客户 群体、不同地区和不同业务条线的分布情况，可以据此重新评估战略重点；对客户关系进行 分析，可以帮助银行识别其客户“下一个最合适的”产品或服务。

4)通过分析客户贡献的潜力，以预测客户的未来利润增长潜力，以便银行更有效地选 择目标对象，合理配置资源。

5)通过分析客户的特征及行为，可以提高银行产品营销的针对性和有效性，改善客户 满意度和忠诚度，最终提升营业收入和市场份额。

**3.** **提高银行风险管理能力**

银行在经营过程中可能会面对各种风险，比如对手可能违约的信用风险、误操作或流程 管控缺失带来的操作风险、市场波动带来的市场风险以及政策合规性等方面的风险。通过对 这些风险因素进行分类、识别、计量与分析，可以有效地管理和降低风险，为银行持续稳健 经营提供保障。

1)在内部审计方面，通过自动化方式和对定量数据的使用，可以更有效地分配审计资 源，从而使审计流程变得更加高效。

2)通过压力测试，可以将极端不利环境以模拟的方式体现到银行的业务运营中，从而 测试其对包括盈利能力、偿债能力和资本能力在内的各种影响，以便提前预估风险。

3)通过对贷款信用风险进行分析，评估每笔贷款的违约倾向和概率，从而确定有效的 回款策略，恰当地分配资源。

4)在反欺诈方面，可以针对海量的历史行为数据探知隐藏在深处的欺诈模式，对可疑 交易进行实时阻断或人工介入调查，降低内外部欺诈行为带来的损失。

**4.提升银行运营管理水平**

分析技术可以应用在银行的网点运营、费用报销、人力资源以及营销资源投放等领域， 帮助银行识别冗余、低效的流程环节并加以改进，最终达到提高运营效率、降低成本的 目的。

1)通过进行营销资源组合分析，分析银行营销资源在各个渠道、时间段的投放效果， 进而优化营销资源的投放规模和投放策略。

2)通过对费用进行分析，有效地分析各项支出，从而降低已知成本，以及发现和识别 未知的异常支出状况。

**银行数据治理概述** | **第2章**|

3)对网点进行优化，通过将网点的贡献、业务量等与网点周边的环境、竞争状况以及 客户类型等进行结合分析，从而为网点定位、开设、撤并、资源(人/机具)调配等提供决 策依据。

4)在人力资源分析及绩效管理方面，可以帮助银行及时地掌握人员流动变化状况及工 作情况，提高员工工作效率，控制人力成本等。

正是由于数据价值不断被挖掘，使得数据治理已经成为银行的核心竞争力，这无疑是未 来几年银行的建设和管理重点。而数据治理是一个长期而且完全体系化的过程，需要银行各 个层级人员有正确的认识并积极参与，既要有良好的规划，又要有适当的方法。

**2.3** **银行数据治理的历史与现状**

**2.3.1** **国际社会数据治理发展历程**

随着信息技术的发展和数据的不断累积，无论是政府还是企业，数据治理的发展势头都 非常迅猛，各种理论和实践方法层出不穷。近几年，随着大数据时代的来临，各国政府都非 常重视大数据的发展和风险控制，同时也注重为了支撑大数据发展和风险控制而进行的数据 治理理论研究。

2009年，美国总统奥巴马签署了开放和透明政府备忘录，要求建立更加开放透明的政 府、参与合作的政府，体现了美国政府对开放数据的重视。2012年，美国政府投资推出了 “大数据研究与发展先导计划”,开启了美国的大数据时代。与此同时，美国很多行业开始 通过大数据创新思维，围绕大数据的采集和利用能力开展各种工作，如气候变化、交通模 式、健康和疾病数据、购买行为以及通过媒体互动表现出的社会行为。

20世纪90 年代之前，美国银行业信息化的特征主要是以运用信息技术来支持业务运 作，主要功能是进行数据的保存与处理，很少有直接面对顾客的服务功能。20世纪90年代 以后，随着美国银行业零售业务的发展，直接面对顾客的服务型信息处理系统在美国银行业 信息化过程中日渐突出。目前，美国银行业通行的信息处理系统主要有大型通用型信息处理 系统和顾客服务型信息处理系统两种类别，但是后者已呈现出逐渐取代前者的势头。为支持 海量数据而产生的数据治理成为美国银行业的重要工作。以美洲银行为例，美洲银行建立了 数据治理的框架，明确信息技术部门和各个业务部门各自角色的职责和分工。数据管理、数 据访问、数据质量、元数据等方面的角色多达29种，通过数据治理为数据用户提供一站式 信息服务。

2010年，欧盟委员会公布了“2020 战略”,这是继里斯本战略之后欧洲联盟(以下简 称欧盟)的第二个十年经济发展规划。欧盟认为，为加强创新潜力，应尽可能地以最好的 方式使用资源，这些创新资源就是数据，开放数据将成为新的就业和经济增长的重要工 具。为了应对这一变革，2010年11月，欧盟通信委员会向欧洲议会提交了“开放数据： 创新、增长和透明治理的引擎”的报告。该报告以开放数据为核心，制定了应对大数据

 **银行数据治理**

挑战的战略。2011年11月，该报告被欧盟数字议程采纳，12 月12 日正式推进这一战 略。2013年，法国政府发布的数字化路线图，将大数据列为国家创新战略重点实施领域 之一。

2013年，日本政府正式公布了新IT 战略——“创建最尖端 IT 国家宣言”。该宣言全 面阐述了2013—2020年期间以发展开放公共数据和大数据为核心的日本新信息技术国家 战略，提出要把日本建设成为一个具有“世界最高水准的广泛运用信息产业技术的社 会”。

从国际环境来看，数据已成为当今社会发展的原材料，数据治理是一项具有战略意义、 势在必行的必要工作。数据存在的机遇不会等人，同样，数据出现的危机也不会等人。寻求 竞争优势的国内外企业也在积极主动地利用数据治理为自身业务增加价值，使整个企业都获 得新的机遇，而不是单单以一种被动的方式来管理数据。

如今，互联网用户超过了几十亿，我们已经迎来数据大爆炸时代，要管理及控制位于公 司不同部门的TB 级数据变得越来越困难。许多公司使用了“堡垒”方法，竖起一堵厚厚的 边界墙来阻止不信任数据的进入。不过这种方法问题重重，原因在于不是所有数据都有同样 的价值，不是所有风险都在边界外，也不是所有的控制措施都能够有效地防止欺诈。数据安 全的堡垒方法形成了一应俱全的方法，从而使企业在低质量数据方面保护力度过大，而在高 价值信息，如客户账户资料或者员工社会保险号码等，其保护力度过小，没有考虑具体的业 务环境或者应用。

数据治理是企业的责任，需要统一的解决方案和治理模型来保护及共享在不同层面 的数据。越来越多的公司通过数据治理提升企业运作效率以及业务洞察力，数据治理也 正在被越来越多的中国企业关注。“进行数据治理，提升业务价值”成为信息技术领域最 近的流行话题。由于“数据治理”被热门关注，很多公司都将自己的解决方案贴上了 “数据治理 (Data Governance)” 的标签，包括数据存储的解决方案、安全解决方案、数 据分析解决方案等。数据治理和众多的新兴学科一样，目前也有很多种定义。IBM公司认 为，数据治理是根据企业的数据管控政策，利用组织人员、流程和技术的相互协作，使 企业能将数据作为企业的核心资产进行管理和应用。数据治理既包括行之有效的数据治 理组织架构，也包括正确的方法论、流程，还包括相关的工具。IBM 公司通过自身实践 并结合大量的行业客户案例实施，总结出包含4大领域11个要素的数据治理框架和方法 来指导数据治理工作的开展。数据治理最开始的驱动力是为满足企业符合外部法规的需 求以及内部进行风险管控的要求。

从理论和国外的实践来看，大型企业一般会建立企业级的数据治理委员会，由业务部门 领导、信息技术部门领导共同参与，让业务与业务之间、业务与技术之间能够有更充分的讨 论与沟通，从而使各方对宏观的数据战略、制度达成共识。在企业级之下，通常还可以有部 门级、项目级的委员会，负责某些局部的数据治理，在最基层面向某一个业务领域应该有相 应的数据管理专员。数据治理委员会、数据管理专员会制定出一系列与数据相关的标准和制 度，由数据管理服务组织执行。其实际上是信息化建设团队，负责数据仓库、数据集成等技 术平台和工具的建设。

**银行数据治理概述** **第2** **章**

**2.3.2** **我国银行业数据治理前进之路**

近20年来，我国银行业发展的历程也是现代信息技术在银行业逐步渗透且影响不断加 深的过程。信息技术不但给银行带来新的业务，增添运营渠道，而且改变了银行的经营模 式。在此过程中银行拥有越来越多的数据，有力促进了银行业务的发展，另一方面，这些数 据的质量普遍不高，导致银行信息变得不那么“可信”。为此，我国银行纷纷将数据治理提 升到战略层面，进行高规模的咨询并在决策层面亲自领导该工作，但由于只是针对数据的一 般性而缺乏从银行特殊性尤其是银行的数据是为业务服务的这一根本立脚点考虑，导致执行 力低下，数据治理效果并不明显。

回顾历史，我国各家银行大都经历了一个从分散到集中、从局部到整体的过程，比如国 有商业银行早期的系统建设大部分是以二级分行为单位开展的，到20世纪90年代才逐渐走 向统一与集中。初始系统中的数据格式、数据名称、数据标准都很不统一，通过近10年的 整合与梳理，逐渐实现了数据集中和统一管理。

数据给商业银行工作主要带来了三大方面的挑战。 一是业务发展转型，经营理念向以客 户为中心进行转变，这就需要银行收集全面信息，以进行领导和管理层面的综合决策及精细 化管理、资产管理为核心向精准客户营销转变。二是竞争环境和流程日益复杂，必须强化自 我约束。这方面需要按照巴塞尔协议清晰构筑风险资本，进行预测性财务管理，制订精准的 绩效考核办法。三是上级监管越来越细致。作为国家命脉的银行业，包括中国人民银行、中 国银行业监督管理委员会、国家外汇管理局等部门对其进行多口监管，管理者需要各银行准 确报送大量数据的同时还要求这些数据口径标准基本一致，以利于行业统计的方便性和及 时性。

面对挑战，银行从业者对解决该问题的必要性也达成了共识，整个中国银行业需要制定 和实施一个总的、战略性的数据治理工作。为此，我国相当多的银行已经开展数据治理工 作，在行内建立了包括管理层在内的数据治理委员会，对资金和人员等资源大力投入，在行 外积极寻找咨询专家和解决方案。近年来，随着金融全球化的不断深入和我国市场经济的逐 步发展，我国金融体系也在不断地变革、发展，金融创新和新型金融机构不断涌现，我国银 行在信息系统建设上纷纷实现了数据大集中，并启动了数据仓库及各类数据分析应用系统的 建设。为满足各应用方向的数据需要，各银行陆续开始进行数据治理规划。在数据治理规划 的指导下，进行了如数据标准、数据质量管理、元数据管理等各个领域建设，部分银行还进 行了数据生命周期、数据模型、数据分布和存储等领域建设。

数据资产逐渐成为银行越来越重要的生产要素。管理数据、应用数据、挖掘数据价值等 一系列与数据相关的工作，也成为一个现代银行加快业务创新、提高精细化管理和科学决策 水平的最重要基础工作之一。

“十一五”期间，我国银行数据管理工作的重点着眼于数据仓库建设，多数国有或股份 制银行实现了企业级数据的逻辑大集中。数据仓库的建成上线，在一定程度上解决了内部报 表和外部报送的问题，但仍然存在手工报送率高、数据质量较差的问题，对管理分析的提升 作用也没有达到管理层的预期目标。因此，在“十一五”末期，多数银行将数据管理工作 的重点转向深层次的数据治理工作，如加强数据标准化，拥有强有力的数据管控组织和更加

 **银行数据治理**

有效的技术工具，全面提升数据质量，挖掘数据资产价值，以更好地满足监管要求和内部经 营管理的需要。

在《中国银行业信息科技“十二五”发展规划监管指导意见》中，明确要求银行建立 数据治理机制，推进数据标准化和质量建设。充分认识数据在支持经营决策、内部管理与金 融服务中的核心价值和战略意义，切实提高对数据价值和数据治理的认识，大力开展数据治 理工作。推进行业数据基础标准框架规划与标准建设。针对新资本协议的实施，以及银行 间、行业间等基础性强、关注度高的重点领域关键数据项和统计指标项，积极推进数据标准 建设。加快建立数据质量管理机制，进一步完善管理流程，明确数据管理、定义、录入、使 用部门和信息科技部门等各方的质量管控职责，逐步建立日常监督管理的长效机制，保障数 据质量的持续提升。积极开展数据架构规划，对信息模型、主数据以及数据集成架构等内容 进行前瞻性设计，制定数据生命周期管理规范，提升数据服务能力。

对于大型商业银行来说，由于数据量庞大，应用系统多且关联复杂，基本上在数据治理 各个领域都会建设，而中小型银行由于数据量少、应用系统少的原因，在数据分布、数据标 准、数据清理方面要求较低。

**1.在数据集中方面**

各大型商业银行在20世纪末21 世纪初基本上都进行了数据大集中工程。数据大集中使 得银行业发展初期数据分散、客户资源割裂的状态得到改变，进一步提升了银行的管理水 平，增强了业务开发能力和服务创新能力。比如2000年，中国工商银行启动了数据大集中 工程，并在2002年完成了全部工程建设，将分布在全国各地的40多个数据中心整合为互相 连接、互为备份的北京、上海两大数据中心，从而由过去的成千上万家“小银行”的集合 体，整合成真正意义上的一家大银行；2006年，中国农业银行完成全国数据大集中，实现 了“全国一网， 一网打尽”的目标，形成了比较完整的应用体系，统一了全行的业务应用， 有效支持了各部门的稳健运营。

**2.在组织架构方面**

2008年，中国农业银行成立信息中心，将其作为数据管理归口部门，结合“新一代核 心银行系统建设工程”启动了数据综合治理工作，从数据问题入手，解析背后的业务管理 问题和系统问题， 一方面完善业务制度，加强业务源头管控，另一方面优化系统功能，加强 逻辑整合，从根本上保障数据质量，促进数据共享，提高数据价值。

2003 年，中国建设银行成立信息中心，承担全行信息资源综合管理职能，负责内外部 信息报送与提供，牵头建立数据管控体系，组织制定数据标准，负责企业级管理信息系统的 业务运行。

大连银行在2007年以“数据治理、信息共享、应用推广”为核心任务，组织全行数据 治理文化培训，制定了完备的数据治理组织架构，建立了数据治理的“三道防线”,明确了 “数据管理办公室”作为数据治理的执行机构，厘清了相对清晰的报告路线、各岗位职责， 明确了数据标准管理流程和数据质量问题发现、分析、改进的闭环控制流程。

光大银行建立了三个层次的数据治理的组织架构。第一层为决策管理层，即董事会及其 下的风险管理委员会、银行管理层及其下的信息科技与数据管理委员会；第二层为组织协调 层，即数据管理小组；第三层为执行层，即总分行相关部门。董事会是最高决策机构，下设 风险管理委员会，在数据方面负责监督数据战略和数据政策的实施和执行情况；信息科技与

**银行数据治理概述** **第2章**

数据管理委员会是最高决策机构。数据管理小组是数据相关工作的直接领导与组织，负责数 据相关各领域、各环节的决策支持、监督执行和组织落实。数据管理小组是由各部门指派人 员组成的虚拟组织。总分行相关部门是数据相关工作的执行者。

**3.在元数据管理方面**

2006年，为解决元数据的手工管理难题，中国工商银行启动了元数据管理系统的初步 建设，把数据库表结构、公共代码等技术元数据在应用开发阶段纳入元数据管理系统进行管 理。2009年，中国工商银行又开始构建企业级元数据管理系统。通过构建企业级元数据管 理体系和建设元数据管理系统，实现了全行“一致性、规范性、开放性、共享性”的管理 要求。

中国邮政储蓄银行在建设数据仓库的同时，研究建立了企业级的元数据管理体系，保障 了全行数据结构的一致性；建立了全行基础数据和业务数据标准体系，规范银行数据从产生 到应用的各个环节，从而使数据发挥价值的过程更加顺畅。

**4.** **在生命周期管理方面**

2005年，中国工商银行启动了数据生命周期管理规范研究项目，对数据的产生、使用、 清理提出了相应的规范和要求，开始建立中国工商银行的数据生命周期管理体系。

**5.** **在数据质量管理方面**

2008年，中国工商银行以数据仓库数据质量问题发现机制为突破口，建立起全行性数据 质量管理系统，并以此平台为基础开展了有效的数据质量治理工作。

大连银行基于基础数据平台构建了“数据质量监测平台”,按照 IDC (互联网数据中 心)标准将数据质量度量指标分为及时性、 一致性、唯一性、完备性、合规性六类，建立 检核指标500多个。各分支行可直接下载所辖机构数据质量问题，同时，数据质量报告通过 邮件形式发送给相关干系人。这样就保证了基础数据平台数据整合的质量和效率。

**6.在数据标准方面**

2011 年，中国工商银行依托元数据管理系统建设集团数据标准管理服务系统，随后分 批次建立了1500多个数据标准。2012年，中国工商银行实施全行渠道类型、产品协议相关 数据标准的落地改造。

2009年，中国农业银行开展了数据治理的规划工作，在全面梳理现状和问题的基础上 形成了数据综合治理整体方案。2011年完成基础数据平台二期上线，满足了各业务领域的 多项重点基础数据需求，并对多个分析应用系统提供数据支持，初步实现了数据的“统一 采集、集中管理、共享使用”,实现了常用指标库的数据统一加工和发布，初步解决指标数 据缺口大、口径不一致、准确性不高的问题。

2003年，中国建设银行开始成立专职的部门推进数据标准化工作，实施了资产负债管 理、客户关系管理、信贷风险管理、财务管理等八大业务管理与应用系统的建设，启动了数 据标准建设工作。为适应业务、信息系统建设的需要，中国建设银行制定并发布了一系列数 据标准，形成了基本覆盖全行业务的数据标准体系，并应用于业务需求描述、信息管理及信 息系统开发，增强了数据定义与使用的一致性，降低了系统整合难度，为促进信息资源共享 提供了保证。中国建设银行参加了中国人民银行理财、资金信托统计制度和数据标准的研究 与论证，配合中国人民银行完成存、贷款数据标准的制定，将建设银行产品数据标准与金融 行业数据标准相结合，建立匹配的产品数据标准体系。

 **银行数据治理**

大连银行在2012年4月建立了《大连银行监管数据标准化工作小组》,统筹数据标准 建设工作，依据中国银行业监督管理委员会《中小银行数据标准化规范》及中国人民银行 《贷款统计分类及编码标准》等数据标准建立了监管标准数据集市，实现了监管数据的标准 化报送。同时，大连银行从技术规范角度建立了机构、柜员及公共代码标准规范，并在基础 数据平台实现了部分标准落地。通过基础数据平台诊断与规划咨询项目，建立了企业客户信 息主数据标准规范，并建立了客户信息整合平台，实现了核心系统、个人贷款、信用卡、信 贷服务、国际结算、网上银行系统的客户信息整合，并为反洗钱、对公和个人客户关系管理 提供客户信息服务。

**7.** **在数据服务方面**

2007年，中国工商银行引进了企业级数据仓库，启动了全行统一的数据仓库建设，完 成境内外主要业务系统数据入库工作。数据仓库支持的应用系统遍布各个业务条线，实现了 客户信息、账户信息、产品信息、交易信息、管理信息等内容的集中管理。

2010年，中国邮政储蓄银行建立数据分析团队，使数据分析工作进入了常态化的工作 模式，有效地支撑和推动了银行业务的发展，并积极推动数据分析成果的业务应用，把数据 价值真正转化为商业价值。

数据治理不仅仅是一个技术问题，更是一个长期的管理课题。在数据治理过程中，各大 商业银行基本上都制定了相应的管理规范，建立了配套的数据治理管理平台，并在继续做好 数据支撑业务发展的同时，在贴近业务、服务业务等方面将进一步加强，从而实现“信息 创造价值”的目标。

**2.3.3** **我国银行数据治理的现状**

经过多年的信息化银行建设，我国各家银行积累了海量的、丰富的数据资源。在当今大 数据时代，数据本身非常重要，其潜在资产价值对于银行而言更为关键，数据作为资源的再 利用价值为银行业金融业务创新提供了支撑力和驱动力。同时，要避免大而无当的数据，让 有价值的银行信息变得更“可信”,使数据获得恰当的应用，发挥潜在的价值。未来那些最 善于利用数据分析来引导决策、控制风险、进行产品创新的银行将获得更多新的竞争优势。 如何提升数据质量，在复杂数据中寻找规律，发现价值，赢得客户，造福社会，已经成为银 行业最为关注的问题。

当前，各家银行已经越来越重视数据的作用，纷纷将数据治理提升到全行战略层面，开 展了一系列持续的治理工作，在业务发展模式上更注重对数据的管理提升和内部挖掘潜力， 并取得了一定的成效，但在推行精细化管理和建立分析挖掘平台方面，还存在数据基础薄 弱，数据不完整和不一致问题。进一步来看，数据管理体制不健全、内部管理职能不清等因 素是导致数据问题的根本原因。因此，要持续、有序地推进银行的数据治理工作，仍需要解 决下列问题：

**1.缺少数据治理企业文化**

银行的数据治理工作是一项庞大而复杂的工程，不是个别部门或少数人员能够妥善完成 的，而是需要各部门之间、各层级之间的相互支持与协作，互相认可共同的数据治理目标， 尤其需要加强科技部门与业务部门之间的合作，才能最终高质量、高成效地完成数据治理工

**银行数据治理概述** |**第** **2** **章**|

作。因此，在数据资产被高度认可和广泛利用的今天，数据治理不仅需要作为银行的一项职 能工作在企业内贯彻执行，而且应该成为一种企业文化在员工及工作之间进行融合。企业文 化的建设必须落到实处，应该从战略角度启动、开展和推进数据治理工作，建立一种以数据 资产为导向的企业文化，将数据治理与信息科技治理、公司治理有机地结合起来。

**2.** **没有完善的组织和制度**

随着数据治理工作逐渐被重视，银行内部已离不开一个企业级的数据治理职能组织，应 该由专业的业务和技术人员组成专门的组织，来承担数据管理者的职责，牵头负责落实全行 数据治理工作。同时，数据治理组织承担的数据治理工作均需要依据行内完善的规范和制度 开展，这也要求建立相关的制度以保障数据治理组织顺利推进各项治理工作。

目前，多数商业银行的数据治理组织和制度由科技部门牵头，整个组织体系业务参与度 欠佳，但从数据问题分析结果来看，业务部门是产生数据问题的最初、最大来源，数据治理 的目的是使银行业务更好地被经营和管理，所以也需要业务部门对数据治理工作进行重视， 深入参与和主动负责，保证数据治理的真正落地。

**3.** **未建立数据治理流程和有效的认责管理机制**

当前，使用数据的部门因具有明确的、迫切的数据需求，而成为处理问题的主要推动 者。但是，在一般情况下，由于没有完整的流程管理机制，在推动工作中会出现以下问题：

1)不同领域的数据分布在不同的系统中，并由不同的部门进行管理，当需要进行跨领 域、跨系统的数据治理工作时，会出现沟通成本较高，协调难度较大的问题，而且对责任部 门的工作驱动力也较低。

2)解决数据问题时仅从局部考虑，为解决当前紧急问题，较少考虑数据复用和共享机 制，不利于后续的系统整合工作。

3)数据问题解决后，较少进行跟踪管理和规范治理，后续的认责机制不健全，不利于 体现数据治理的重要性和必要性。

因此，建立健全的数据治理流程是全面和有效开展数据治理工作的基础，同时还需 要建立决策、沟通、监控、考核的完整有效机制，创建数据治理决策协调机构，创造全 行数据治理理念，有效地解决银行数据的责、权、利问题，这样才能使数据治理工作顺 利推进。

**4.缺少数据治理各领域的管理体系**

**(1)缺少企业级数据标准管理体系** 虽然银行拥有了大量的数据资源，但是，也经常 面临重要数据缺失，系统间数据不一致，统计口径和加工方法不一致，导致数据可信度降低 的问题。深入分析后发现，出现这种现象的原因为缺乏有效的数据标准化。建立数据标准体 系是数据环境建设中的重要环节，对数据进行标准化，是保障基础数据的完整性、 一致性和 严密性的基础。因此，商业银行有必要制定合理的数据标准并严格执行，确保各系统之间的 数据一致性。同时，为了保障业务系统能够采集真实、完整、有效的数据，在建设系统时必 须通过数据标准对数据的定义、描述等予以规范和约束。

**(2)缺少企业级元数据管理体系** 目前，我国大多数银行的元数据管理仅限于少数系 统和少数用户，尚未达到体系化程度，也存在完备性不足的情况。企业级元数据管理是银行 数据治理的重要载体，它以全行元数据管理为基础，将科学的方法和管理策略通过企业级的 元数据管理体系实施与执行，再配以完善的组织流程和制度规范，将有效提升数据治理的成

 **银行数据治理**

效。因此，建立企业级元数据管理体系势在必行。

**(3)缺少企业级数据质量管理体系** 数据的质量对银行业发展尤为重要，并且银行对 数据质量的治理重视程度会直接影响数据治理的成效。银行的数据质量管理应当涵盖数据质 量问题的预防、识别、度量、分析、监控、清洗等管理活动，以满足对数据质量的要求。为 此，银行需要建立专业的质量管理团队，依靠质量管理系统持续地管理数据质量，利用数据 剖析工具持续地监控数据的流动，通过明确数据认责关系，尽可能在数据问题根源处解决 问题。

**(4)缺少完备的数据生命周期管理体系** 通过数据生命周期管理体系建设，银行可优 化应用存储结构，有效控制在线数据规模，提高生产数据访问效率，减少系统资源浪费，提 高应用系统运行的整体效率和效果，保证应用系统健康高效运行。但是，当前我国银行大部 分在系统无法支撑时才考虑数据清理备份工作，难以做到对数据生命周期的统筹管理，并且 对支持数据生命周期管理的系统和工具建设力度不够，不能很好地支撑全行数据生命周期管 理工作。

**(5)缺乏完善的系统支撑和技术手段** 现代的银行系统数据量庞大，各式系统多种多 样，如果不依赖技术手段，没有相应的支撑平台和工具，就不可能理解如此大的数据量和看 到其潜在的价值，利用简单图形分布寻找规律或异常情况等传统方法，在当今海量数据面 前，已毫无作用。同样，没有充分的技术手段，仅依靠手工处理，显然也无法把数据治理工 作做好。因此，必然需要先进的技术手段、配套的系统支撑数据治理工作高效有序地开展。

**2.3.4** **银行数据治理的思考**

中国银行业监督管理委员会在《中国银行业信息科技“十二五”发展规划监管指导意 见》中明确指出：“十二五”期间，大中型银行要把数据治理作为重要的制度性建设与基础 性工作，加强组织保障、制度保障与流程保障，有序推进、重点强化；统一数据标准，提高 数据质量，深化数据应用，有效支撑银行业务发展，有效提升银行管理水平。因此，对银行 数据治理的思考可以从以下三个方面进行：

**1.提高认识，建立数据治理体系**

数据治理体系建设是指建立数据拥有者、使用者、数据以及支撑系统之间的和谐互补关 系，从全企业视角协调、统领各个层面的数据管理工作，确保内部各类人员能够得到及时、 准确的数据支持和服务。

商业银行需要积极开展数据架构规划，对信息模型、主数据以及数据集成架构等内容进行 前瞻性设计，制定数据生命周期管理规范，提升数据服务能力，形成包括客户、产品、机构等 核心基础数据的统一视图；加强元数据管理和应用，制定元数据管理策略、标准、流程，建立 统一的元数据管理与传输机制，将元数据管理与系统开发、运维与数据架构管控等流程结合起 来，保证元数据的有效性和准确性；不断丰富前台源系统的基础数据，持续加大数据积累与整 合的广度和深度，建设统一的数据平台，满足前台营销、统计分析和决策支持、风险管理和新 资本协议实施等多种需求，持续提升对数据的挖掘、分析与深度运用能力。

**2.加强数据治理水平，确保数据安全**

数据安全或者说数据隐私的重要性比以往有显著提升，这也需要在数据治理中加强对数据

**银行数据治理概述** | **第2章**

安全的重视。在传统应用场景中，数据由企业收集，在企业内部应用，数据所有权的问题并不 突出。在大数据时代，数据要更多地进行跨界整合，外部应用的商业模式不断创新，这其中就 涉及更多数据所有权、数据隐私的话题。用户信息究竟属于企业还是属于用户?在什么条件下 企业可以拿来用于商业应用?这些问题的答案目前还在探讨当中，但毋庸置疑的是，企业需要 在数据治理过程中，更加注意与数据安全、数据隐私相关的制度和政策并加以落实。

随着互联网、物联网、云计算等技术的快速发展，以及智能终端、网络社会、数字地球 等信息体的普及和建设，全球数据量出现爆炸式增长，仅在2011年就达到1.8×10⁴亿 GB。 IDC 预计，到2020年全球数据量将增加50倍，大数据时代已经到来。 一方面，云计算为这 些海量的、多样化的数据提供了存储和运算平台，同时数据挖掘和人工智能从大数据中发现 知识、规律和趋势，为决策提供信息参考。但是，大数据的发展将进一步扩大信息的开放程 度，具有数据量大、类型多样、运算高效、产生价值等特点，同时，随之而来的隐私数据或 敏感信息的泄露事件时有发生。面对大数据发展的新特点、新挑战，如何保障数据安全是需 要认真研究的课题。

**3.** **树立大数据治理意识，顺应大数据时代发展趋势**

首先，在大数据时代，数据质量更加重要，从精准营销、风险识别等应用场景来看，因 为数据与运营结合程度较以往更紧密，要求的数据粒度更细，任何一点错误都可能直接带来 业务上的损失。而传统的指标应用反而对运营环节没有如此直接的影响。因此，在大数据环 境下对数据质量的需求是显著提升了而不是降低了。

其次， Hadoop 、Spark 等大数据技术的应用，对数据治理的技术手段提出了许多新的要 求。传统模式下基于 RDBMS 进行管理， SQL 是通用的数据访问方式。而在大数据环境中， Hadoop 、MPP 、RDBMS 、Spark 并存，如何在混搭的异构环境中实现对数据资产的可视化统 一管控，避免大数据系统成为不可管理的黑盒子，是传统行业应用大数据技术需要面对的关 键问题之一。特别是大数据技术人才目前更多地流向互联网企业，进入传统行业的少之又 少，在人才可得性短期不能快速解决的情况下，需要依靠技术手段来确保传统企业信息技术 人员对数据资产的可视与可控。

2011年6月，美国麦肯锡全球研究院发布题为《大数据：下一个创新、竞争和生产力 的前沿》的研究报告，首次正式提出“大数据时代已经到来”的观点，数据正成为与物质 资产和人力资本相提并论的重要生产要素，大数据的使用将成为未来提高竞争力和生产力的 关键要素。大数据时代不仅将改变每个人的日常生活和工作方式，而且将从根本上改变国家 治理与社会治理状况。大数据时代的社会治理要求我们要积极顺应大数据时代的发展趋势和 本质要求，树立大数据治理意识。

长期以来，我国的数字管理水平非常落后，社会理性化程度始终处于较低水平。正如历 史学家黄仁宇所指出的，传统中国社会缺乏“数目字管理”的传统，常常以道德替代法律， 国家与社会管理是模糊和杂乱的，这是传统中国社会难以飞跃发展的根本原因所在。在传统 社会中，由于人类自身生物能力以及信息技术水平的限制，人们能够获得的数据信息十分有 限，社会管理者始终面临着“信息不对称”的严重制约，始终面临着较高的决策风险，存 在着诸多“不可治理”的社会公共事务，无法根本性地提高社会治理绩效与水平。当前， 我国的社会治理依然大体延续着传统中国社会的治理模式，主要还是依靠感觉、经验来管理 社会，社会领域的治理缺乏应有的数据支撑，某些社会领域的公共事务由于基础数据的缺失

 **银行数据治理**

而长期处于“无法治理”状态。从总体上看，整个社会治理的理性化、科学化、现代化程 度依然比较低。大数据技术的发展，为改变传统经验型社会治理模式带来重大机遇，“理念 是行动的先导”,大数据时代的社会治理在理念上要高度重视大数据治理意识的培育，大力 推动社会计算和智慧治理。

大数据时代使海量、异构数据的动态采集、实时存储、即时分析成为现实，多元化、个 性化、移动化的大数据应用也得以实现，社会治理也可以更多地依托大数据技术得以突破性 地创新和提升。这就要求在理念上要率先树立大数据治理意识，注重社会治理领域的大数据 建设和整合，特别是要尽快构建社会治理的基础数据体系，如公民信用、不动产、家庭人 口、收入和税收等基础性大数据，彻底改变诸多社会领域的“不可治理”状态，极大降低 “信息不对称”带来的决策风险。同时，为推进社会治理数据化，还要树立社会计算和智慧 治理意识。所谓“社会计算”是指将系统科学、人工智能、大数据挖掘等科学计算理论作 为研究方法，将社会科学理论与计算理论相结合，帮助人类更加深入地认识社会、改造社 会，解决政治、经济、文化等领域复杂性社会问题的一种理论和方法论体系。

在大数据时代，社会是可计算的，恰如全球复杂网络研究权威专家巴拉巴西在《爆发》 一书中的判断：在大数据时代，人类行为中有93%是可以预测的。社会计算的出现，可以 对社会成员的行为逻辑、社会事件的发展态势做出判断、预测和模拟，使得社会治理也成为 可计算、可预测和可控制的领域，从而可以推动社会治理由传统的经验治理向智能治理转 型，推动社会治理数据化、科学化、理性化。

随着银行进入后数据集中时代，加强数据治理已经成为各商业银行的当务之急。在这个 过程中如何借鉴国外银行的先进经验，如何使用日新月异的技术，如何有效结合我国银行的 实际，建立具有中国特色的银行数据治理机制，是值得探讨的重要课题。

**2.4** **建立银行数据治理体系**

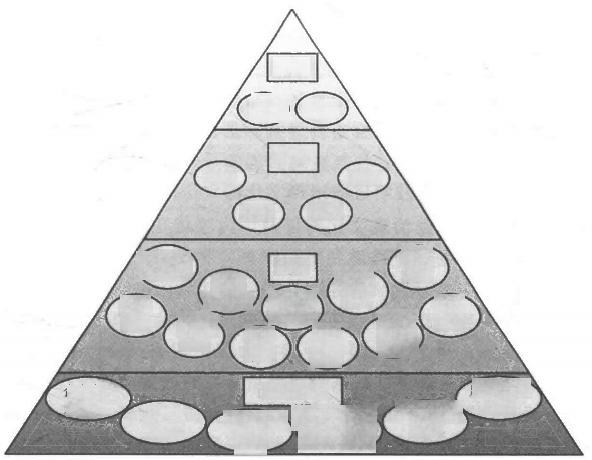
我国银行在过去十几年的信息化发展中，都或多或少有些重生产轻分析的现象，这导致 了大量以账户为中心的传统交易系统的存在。随着我国金融改革不断深化，尤其是当互联网 金融崛起并对传统银行业造成冲击时，银行自身面临更严格的监管要求和更严峻的市场挑 战，开始对信息整合、数据挖掘提出更高的要求。但与此同时，在对数据利用提出迫切需求 的过程中许多深层次的数据问题开始逐步暴露。例如，数据认责不明导致数据源头录入质量 不高，数据标准缺失导致统计口径混乱，整体数据质量缺乏有效的监控和管理。这些长期存 在的问题如同连环锁一般，使得银行在进一步挖掘数据的深层次应用时障碍重重。因此，建 立数据治理体系成为当务之急，也是银行应对未来发展的战略之举。

数据治理是一项复杂、长期、系统性的工程，涉及思维、方法、组织、系统工具等多方 面要素的综合运用。为了满足企业内部的信息使用需要， 一般会通过建立专门的数据治理体 系来保证数据的可用性、可获取性、高质量、 一致性以及安全性。下面将对银行的数据治理 体系架构进行介绍，以使读者对数据治理的文化、方法、组织、支撑系统有更好的了解。

金字塔对大家来说并不陌生，古埃及宏伟壮观的金字塔堪称历史上的奇迹，其结构使大 家印象非常深刻。基于信息化银行给数据治理带来的诸多机遇，以及面临的一系列问题，我

**银行数据治理概述** | **第** **2** **章**

们 对 商 业 银 行 数 据 治 理 体 系 进 行 研 究 分 析 ， 发 现 银 行 的 数 据 治 理 体 系 也 是 一 个 金 字 塔 结 构 ， 依 次 为 战 略 、 机 制 、 领 域 、 技 术 支 撑 ， 从 上 至 下 指 导 ， 从 下 而 上 推 进 ， 形 成 一 个 多 层 次 、 多 维 度 、 多 视 角 的 全 方 位 框 架 ， 如 图 2 - 2 所 示 。



战略

目 标 规划

机制

组织

制度 角色

元 数 据 数据 数据

数据 标准 数据生 模型 数据

交换 质量

技术支撑

数据治理

工具

管理系统

分布 数据 命周期 数 据 安全

领域

生命周期

管理

企业级 元数据

数据清理

数据 模型

质量检查

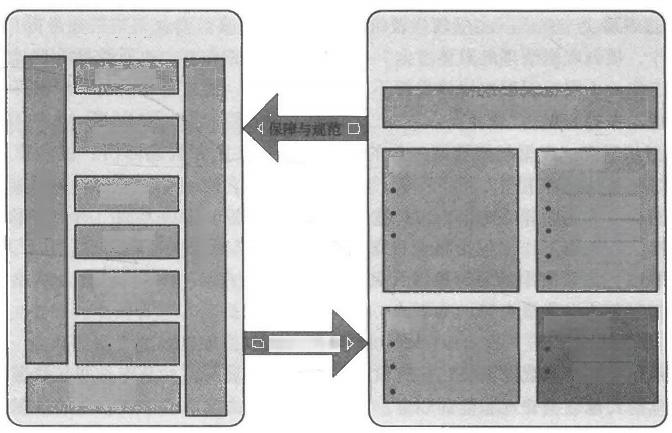
流程

工具

图2 - 2 商业银行数据治理体系

从 商 业 银 行 数 据 治 理 体 现 的 金 字 塔 结 构 可 以 看 出 ， 实 质 上 商 业 银 行 的 数 据 治 理 体 系 包 含 两 个 层 面 ： 一 是 数 据 治 理 的 核 心 领 域 ， 二 是 数 据 治 理 的 保 障 机 制 。 战 略 、 机 制 及 各 领 域 的 技

术 支 撑 是 商 业 银 行 进 行 数 据 治 理 的 保 障 机 制 。 其 相 互 关 系 如 图 2 - 3 所 示 。



核心领域

数据交换

发展战略及目标

数据质量

制度章程

元数据

主数据

数据生命

周期

技术应用

\*

技术工具

数据模型

规章剩度 管控办法 考核机制 技术规范 实施办法

流程目标

流程任务 流程分级

组织架构 组织层次 组织职责

支撑与落实

保障机制

*立掉平合*

数据安全

数据标准

组织

流程

图2 - 3 商业银行数据治理的保障机制与核心领域

 **银行数据治理**

**2.4.1** **银行数据治理保障机制**

**1.** **发展战略及目标**

战略是选择和决策的集合，共同绘制出一个高层次的行动方案，以实现高层次目标。数 据战略是企业发展战略中的重要组成部分，是数据管理计划的战略，是保持和提高数据质 量、完整性、安全性和存取的计划，是指导数据治理的最高原则。数据治理是否与企业发展 战略相吻合也是衡量数据治理体系实施是否成熟、是否成功的重要标准。要在企业发展战略 框架下，建立数据治理的战略文化，包括企业高层领导对数据治理的重视程度、所能提供的 资源、重大问题的协调能力，以及对数据治理文化的宣传推广、培训教育等一系列措施。数 据治理与企业发展战略以及数据治理文化的普及等内容将在“数据治理保障机制”的“企 业发展战略”中进一步进行阐述。

**2.数据治理组织**

数据治理的组织包括制度组织和服务组织。制度组织主要负责数据治理和数据管理制 度。这些组织是跨职能的，通常商业银行会建立数据治理委员会、数据管理制度团队等组 织，负责整体数据战略、数据政策、数据管理度量指标等数据治理规程问题。与政府部门和 机构一样，制度组织执行类似于行政部门的职责。数据服务组织主要是由数据管理的专业人 员组成，包括数据架构师、数据质量分析师、元数据管理员等，主要执行数据治理各个领域 的具体实施工作。这里只简单介绍数据治理组织包括哪些内容，具体介绍详见“数据治理 保障机制”的“数据治理组织”相关章节。

(1)组织架构 有效的组织机构是项目成功的有力保证，为了达到项目预期目标，在 项目开始之前对于组织机构及其责任分工做出规划是非常必要的。数据治理涉及的范围很 广，牵扯到不同的业务部门和科技部门，是商业银行全局大事，如何成立和成立什么样的组 织应该依据商业银行本身的发展战略和目标来确定。

(2)组织层次 建立由高层领导者组成的数据治理委员会，组织跨业务部门和 IT 部门 的协调工作，规划数据治理的总体方向，并在其下设立工作组，执行数据治理计划和监督数 据管理等工作。不同组织层次应该发挥不同的职责职能，建立合理的组织层次有利于快速推 动数据治理工作的开展。

(3)组织职责 根据数据管理工作的实际需要，在业务管理部门、技术管理部门和业 务应用部门间要确定各工作人员的职责。不同的组织负责的职责不尽相同，例如不同的业务 部门应该明确各自业务开展对数据的具体要求和相关规则，而技术部门则会根据业务部门的 需求负责具体的实施工作，包括将业务部门提出的要求转化成技术语言，用于事前的控制 (例如字段的约束)、事中的逻辑控制(例如控制不能为空)、事后的核查，以及具体的技术 操作和编制定期的治理报告等。

**3.** **制度章程**

制度章程是确保对数据治理进行有效实施的职责问责制度，其中一些是数据治理职能的 职责，也包括其他数据管理职能的职责。数据治理是最高层次的、规划性的数据管理制度活 动。换句话说，数据治理是主要由数据管理人员和协调人员共同制定的高层次的数据管理制 度决策。这里只简单介绍包括哪些制度章程，具体描述详见“数据治理保障机制”的“数

**银行数据治理概述** **第2章**

据治理制度章程”相关章节。

**(1)规章制度** 数据治理规章制度类似于企业的公司条例，它会阐明数据治理的主要 目标、相关工作人员、职责、决策权利和度量标准。

**(2)管控办法** 管控办法是基于规章制度与工具的结合，是可落地、可操作的办法。

**(3)考核机制** 考核是保障制度落实的根本，建立明确的考核制度，实际操作中可根 据不同商业银行的具体情况，建立相应的针对数据治理方面的考核办法，并与个人绩效相 关联。

**(4)技术规范** 技术规范是保障数据治理平台可持续管理的基础，随着数据量的增长、 技术水平的发展，为更好地、可持续地实现数据资产的管理与应用，需要建立明确的技术 规范。

**4.** **流程管理**

流程管理包括流程目标、流程任务、流程分级。根据数据治理的内容，建立相应流程， 且遵循本单位数据治理的规章制度。实际操作中可结合所使用的数据治理工具，与数据治理 工具供应商进行协商，建立符合商业银行的流程管理。具体描述详见“数据治理保障机制” 的“数据治理流程”相关章节。

**5.技术应用**

技术应用包括支撑核心领域的工具和平台，例如数据质量管理系统、元数据管理系统 等。它们是数据治理能够顺利开展的技术保障。只有建立丰富的数据治理工具和平台，才能 从各个领域有效地进行数据的管理和治理，才能有效提高商业银行的数据价值。具体描述详 见“数据治理保障机制”的“数据治理支撑系统”相关章节。

**6.** **成熟度模型**

CMMI 协会在2014 年发布了数据管理能力成熟度模型 Data Management Maturity (DMM), 可以用来评估和提升企业的数据管理水平，帮助企业跨越业务与IT 之间的鸿沟。

DMM 模型可以帮助企业在管理数据资产上达成共识。它包括了5个连续能力和25个提 升的过程域，可以反映所有数据管理的内容。DMM 可以促进企业建立自己的数据管理成熟 度路线图。具体描述详见“数据治理保障机制”的“成熟度模型”相关章节。

**2.4.2** **银行数据治理核心领域**

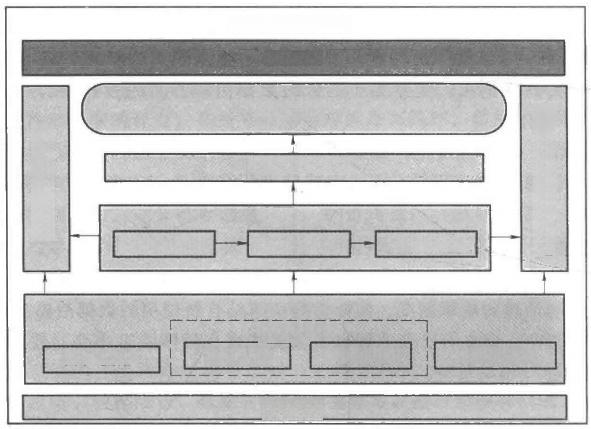
每个数据治理的领域都可作为一个独立方向进行研究治理，目前总结的数据治理领域包 括但不限于以下内容：数据标准、元数据、数据模型、数据分布、数据存储、数据交换、数 据生命周期、数据质量、数据安全以及数据共享与服务。

同时，各领域之间需要有机结合，如数据标准、元数据、数据质量等几个领域相互协同 和依赖。通过数据标准的管理，可提升数据合法性、合规性，进一步提升数据质量，减少数 据生产问题；在元数据管理的基础上，可进行数据生命周期管理，有效控制在线数据规模， 提高生产数据访问效率，减少系统资源浪费；通过元数据和数据模型管理，将表、文件等数 据资源按主题进行分类，可明确当事人、产品、协议等相关数据的主数据源归属、数据分布 情况，有效实施数据分布的规划和治理。

数据治理领域是随银行业务发展而不断变化的，领域之间的关系也需要不断深入挖掘和

**银行数据治理**

分析，最终将形成一个互相协同与验证的领域网，全方位地提升数据治理成效。 图2-4所示为商业银行数据治理的核心领域。



数据治理核心领域

数据服务

数据交换

数据分布和存储

数据模型

逻辑数据模型

元数据管理

业务元数据

数据标准 指标元数据

数据安全

数据生命周期管理

数 据 质 量 管 理

概念数据模型

物理数据模型

技术元数据

操作元数据

图2-4 商业银行数据治理的核心领域

**1.** **数据模型**

数据模型是数据治理中的重要部分，合适、合理、合规的数据模型，能够有效提高数据 的合理分布和使用，它包括概念数据模型、逻辑数据模型和物理数据模型，是数据治理的关 键、重点。具体描述详见“数据模型”相关章节。

**2.** **元数据管理**

元数据分为业务元数据、技术元数据和操作元数据，三者之间关系紧密。业务元数据指 导技术元数据，技术元数据以业务元数据为参考进行设计，操作元数据为两者的管理提供支 撑。具体描述详见“元数数据管理”相关章节。

**3.** **数据标准**

数据标准是银行建立的一套符合自身实际，涵盖定义、操作、应用多层次数据的标准化 体系。它包括基础标准和指标标准(或称为应用标准)。与数据治理其他核心领域具有一定 的交叉，比如元数据标准、数据交换和传输标准、数据质量标准等。商业银行的数据标准一 般以业界的标准为基础，如国家标准、监管机构(如国家统计局、中国人民银行、工信部) 制定的标准，结合商业银行本身的实际情况对数据进行规范化， 一般会包括格式、编码规 则、字典值等内容。良好的数据标准体系有助于商业银行数据的共享、交互和利用，可以减 少不同系统间数据转换的工作。具体内容详见“数据标准”相关章节。

**4.数据质量管理**

数据质量管理已经成为银行数据治理的有机组成部分。高质量的数据是商业银行进行分 析决策、业务发展规划的重要基础，只有建立完整的数据质量管理体系，才能有效提升银行 数据整体质量，从而更好地为客户服务，提供更为精确的决策分析数据。具体内容详见

**银行数据治理概述** **第2章**

“数据质量管理”相关章节。

**5.** **数据生命周期管理**

任何事物都具有一定的生命周期，数据也不外如是。从数据的产生、加工、使用乃至消 亡都应该有一个科学的管理办法，将极少或者不再使用的数据从系统中剥离出来，并通过合 适的存储设备进行保留，不仅能够提高系统的运行效率，更好地服务客户，还能大幅较少因 为数据长期保存带来的存储成本。数据生命周期一般包括在线阶段、归档阶段(有时还会 进一步划分为在线归档阶段和离线归档阶段)、销毁阶段三个大阶段，管理内容包括建立合 理的数据类别，针对不同类别的数据制定各个阶段的保留时间、存储介质、清理规则和方 式、注意事项等。具体内容详见“数据生命周期”相关章节。

**6.数据分布和存储**

数据分布和存储主要涵盖了数据如何划分和存储，总行系统以及总分行数据如何分布， 主数据及参考数据(也称为副本数据或辅数据)如何管理。只有对数据进行合理的分布和 存储，才能有效地提高数据的共享程度，才能尽可能地减少数据冗余带来的存储成本。具体 内容详见“数据分布和存储”相关章节。

**7.数据交换**

数据交换是银行进行数据交互和共享的基础，合理的数据交换体系有助于银行提高数据 共享程度和数据流转时效。 一般商业银行会对系统间数据的交换制定一些原则，比如对接 口、文件的命名、内容进行明确，规范系统间、银行系统与外部机构间的数据交换规则，指 导数据交换工作有序进行。建立统一的数据交换系统， 一方面可以提高数据共享的时效性， 另一方面也可以精确掌握数据的流向。具体内容详见“数据交换”相关章节。

**8.数据安全**

商业银行的重要且敏感数据大部分集中在应用系统中，例如客户的联络信息、资产信息 等，如果不慎泄露，不仅给客户带来损失，也会给商业银行带来不利的声誉影响，因此数据 安全在数据管理和治理的工作是至关重要的。如何保障数据不被泄露和非法访问，是非常关 键的问题。数据安全管理主要解决的就是数据在保存、使用和交换过程中的安全问题，通常 通过各种渠道、各种手段进行全方位的防控。具体内容详见“数据安全”相关章节。

**9.** **数据服务**

数据的管理和治理是为了更好地利用数据，是数据应用的基础，如何发挥数据的有效价 值，如何在客户营销、风险控制、绩效考核等方面充分利用行内外数据，进行数据的创新应 用是银行需要考虑的问题。具体内容详见“数据服务”相关章节。

**10.** **非结构化数据管理**

银行存在大量的非结构化数据，并且随着大数据时代的到来，银行也会从外部获取数据 (例如社交网站、新闻等),非结构化数据存在格式多样、来源复杂、结构不定、数据量庞 大等特性，因此对于这些数据如何管理和使用也是各商业银行面临的重大难题。具体内容详 见非结构化数据的管理相关章节。



**第** **3** **章**

**银行数据治理保障机制**

数据治理是一个系统化的工程，是一个从上至下指导，从下而上推进的工作。因此，在 指导方面必须要得到大家的共识，要有一个强有力的组织、合理的章程、明确的流程、健壮 的系统，这样才能使数据治理工作得到有效的保障。

**3.1** **企业发展战略**

企业发展战略是关于企业如何发展的理论体系，是一定时期内对企业发展方向、发展速 度与质量、发展能力的重大选择、规划和策略。企业发展战略包括企业发展目标、业务发展 规划及与之配套的 IT发展规划等政策性原则和行动纲领，其中数据治理相关的发展规划是 IT 发展规划的重要组成部分，任何企业的数据治理工作均应在此之下进行设计和执行。企 业发展战略是指导数据治理的最高原则，要在企业发展战略框架下，建立数据治理的战略 文化。

**3.1.1** **数据治理文化**

在开展数据治理之时，首先要从战略上将数据放在银行重要资产的高度。现代银行业对 信息化的依赖程度极高，银行生产经营的过程和结果最终都体现为信息，而信息的载体即数 据。银监会《中国银行业信息科技“十二五”发展规划监管指导意见》要求银行应将数据 治理作为一项重要的资产管理工作，通过建立适应企业自身需求和特点的数据治理体系，推 动数据管控各环节的有效运转，实现银行数据资产的高效利用。

数据是银行的战略资产，在大数据时代，如何管好数据，挖掘其中的价值，构建信息化 银行，已经成为各商业银行的重要战略课题。我们面对的数据越来越多，特别是随着电子商 务的蓬勃发展，线上数据急剧增长。为保证“资产”的增值，我们需要结合企业发展战略 整体要求，建立清晰的数据治理目标，用于指导、约束后续一系列的数据治理行为。

数据治理战略的组成部分可能包括：数据治理的愿景、商业案例摘要(包括例子)、指 导原则、长远目标、管理措施、实施路线等。

**银行数据治理保障机制** **第3章** |

**3.1.2** **认知与行动**

信息化领域有一个词语——“Garbage in,garbage out”, 翻译成中文就是“垃圾进、垃 圾出”,意指用垃圾数据作样本，产生的研究结果也毫无价值。在大数据时代的今天，银行 要用数据说话，要采用先进的数据挖掘技术和全面的数据分析来指导企业的经营和决策，需 要一个健全的数据环境做支撑，这就要求把数据治理工作做扎实、做细致。不仅要严格确保 数据源的可信性、完整性和合理性，还要缩短基础数据获取和加工环节，极大地提高数据分 析应用的实施效率。

数据的污染可能发生在数据产生、采集、传输、流转、加工、存储、提取、交换等各个 环节，因此要保证数据治理目标的实现就必须对数据进行全流程管控，并使员工在实际工作 中坚决贯彻执行。管理的问题、操作的问题、系统的问题等最终都有可能导致数据质量问 题，要使数据治理人人有责的理念深入人心，避免因单个环节数据管控失灵而导致的蝴蝶效 应，就要在全行范围梳理和建立数据认责机制，使每个人都能够清楚地知道自己在数据管控 流程中扮演的角色，认清自己对数据质量所承担的职责与义务，并明确具体的工作要求，才 能在全行范围内对数据质量高度关注的认识达成统一，并实现最终的治理效果。

总而言之，数据治理不是一个短期和简单的项目，而是一个长期且复杂的工程，作为一 项持久的工作，银行要将数据治理各领域的工作内容进行常规化，将数据作为银行重要资产 的理念融入其企业文化和经营战略中，才能逐步享受到数据治理为银行带来的红利。

**3.2** **数据治理组织**

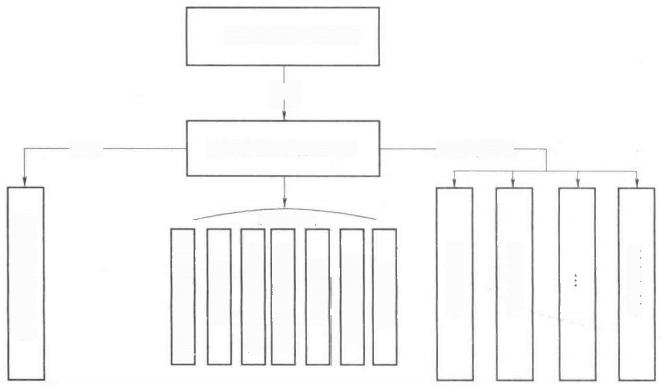
数据治理工作是一项全行性的工作，需要科技部门、业务部门等各方面的通力合作，需 要总行、各分支机构的上下联动和一起努力。 一般商业银行均会成立由行领导牵头的数据治 理管理委员会，由数据治理管理委员会统一领导数据治理工作；成立数据治理综合部门作为 数据治理管理委员会的办事机构，负责企业的数据治理综合管理。在数据治理管理委员会的 统一领导下，各业务管理部门、经营机构负责本业务线、本机构的数据治理，通过数据治理 管理委员会协调解决重大数据治理相关的管理事项。

图3-1所示为商业银行数据治理组织架构。

从职责角度，数据治理在全行的组织架构可划分为决策层、管理层、执行层。

**(1)决策层** 数据治理管理委员会即为数据治理的决策层。作为全行数据治理工作的 决策机构，组长可由负责数据治理的主管行长担任，成员由相关部门总经理担任。有些银行 没有设置单独的数据治理委员会，而是在已有的如信息化管理委员会下设置了数据治理办公 室。总之，数据治理管理委员会建立的目的是确立数据治理的目标、资源，保证涉及部门的 负责人涵盖其中，尽快在全行搭建起数据治理体系，为后续的工作形成共识并进而共同朝着 一致的目标推进。数据治理决策层负责重大事项的协调和决策，指导审批数据治理工作规划 和年度目标，审批数据治理相关的政策、组织建设、管理流程以及重要的工作成果文档，听

**银行数据治理**



业 务 专 业1

业 务 专 业2

业 务 专 业1

數 据 服 务

數 据 生 命 周 期

数 据 质 量

数 据 交 换 和 集 成

数 据 分 布 和 存 储

数 据 模 型

元 数 据 及 數 据 标 准

信 息 科 技 相 关 单 位

指导- 数据治理综合管理部门 组织与协调-

数据治理管理委员会

工作小组

领导

图3 - 1 商业银行数据治理组织架构

取并评价数据治理工作情况汇报。

(2)管理层 数据治理综合管理部门为数据治理管理层，是全行数据治理工作的管理 机构，向数据治理决策层负责，可以由一个专门设置的独立部门担任，也可以由分别负责数 据标准、数据质量、数据模型等各领域的业务部门和信息技术部门共同承担。数据治理管理 层负责制定数据治理相关的政策、流程、组织、角色职责，负责数据治理相关的日常管理工 作，组织和牵头数据治理建设和推广工作，管理、评估各部门和信息技术系统的数据治理工 作情况。

**(3)执行层** 数据治理执行层是全行数据治理日常工作的执行机构，向数据治理决策 层负责，由全行全体成员(包括总行、分行、支行各机构成员),包括业务部门和科技部门 共同组成。因此， 一般情况下这是一个虚拟组织。数据治理执行层参与、配合数据治理日常 管理工作，根据数据治理管理组要求，制定数据标准、数据质量等内容，评审数据治理成 果，执行数据治理要求。各业务部门和科技部门属于数据治理执行层，但两者之间有不同侧 重点。各业务部门的重点是在数据治理管理层的指导和协调下，依据数据治理整体目标制定 业务标准，完善涉及数据治理的业务需求(如产品系统输入项的值域控制要求等)等，负 责各自业务条线数据治理相关内容的推动、治理，并定期进行后评估。科技部门负责发现数 据治理问题，确定数据问题的原因(如业务需求不完善或科技系统问题),并将问题归类汇 总后发送相关执行部门进行数据治理。

举个例子来说，中国农业银行的数据治理架构包括董事会、高级管理层、归口管理部 门、业务部门、科技部门、内控合规部门以及各级分支机构。董事会和高级管理层对全行数 据管理负领导职责，信息中心作为归口管理部门，统筹数据管理的组织实施，业务部门和科 技部门分别负责业务和系统层面的数据管理工作，内控合规部门负责数据管理的内部控制监 督，各级分支机构负责辖内的数据管理相关工作。

又 如 ， 中 国 工 商 银 行 的 数 据 治 理 整 体 工 作 是 在 操 作 风 险 管 理 委 员 会 的 统 一 领 导 下 ， 实 行 “统一领导、条线管理、分工协助”的管理模式。数据治理综合管理部门负责数据治理工作

**银行数据治理保障机制** **第3章** |

的牵头管理，各部门、各级机构负责本专业线、本级机构的数据治理工作。

再如，大连银行成立“数据管理办公室”,将其作为数据治理的执行机构。

还有巴克莱银行建立了数据标准管理的专业化组织和队伍，形成了三层组织体系，包括 决策层、管理层及执行层；设有数据标准委员会，负责统筹全行范围数据标准管控方案及各 项工作，指导数据标准办公室开展日常工作。

**3.3** **数据治理制度章程**

一种文化的形成从来不是一蹴而就的，需要思想及行动上的指导。数据治理的战略文化 亦是如此，要在全行上下树立“数据即资产”的企业文化，不是一朝一夕的事情。

在数据治理体系搭建之初，要对企业的业务、信息技术、管理等环节进行充分的现状调 研，结合实际情况制定数据治理蓝图规划，明确企业在数据治理方面的目标原则和实施策 略。在目标原则和实施策略形成后，需要出台相关的政策，并在全行范围内对数据治理的目 标和规划进行宣传、贯彻，通过对执行层的培训使当事人明确并掌握数据治理相关工作的具 体要求，并以会议、简报等形式开展持续的宣传，使决策层和管理层能够及时了解数据治理 的阶段性成果和可能存在的困难，增强银行开展数据治理的信心，及时取得必要的支持。我 国银行的数据现状普遍都是一个先污染、后治理的过程，数据治理必然带来新标准的确立和 旧系统的改造，是一个有破有立、无破不立的过程。这一过程涉及大量跨部门、跨条线、跨 系统的沟通协调，同时也涉及不小的投资。因此，数据治理这项工作从起步之初就需要得到 银行高层的高度重视，并在涉及系统改造、奖惩机制等重大决策时获得足够的支持，那些能 够取得突破并实现数据质量真正改观的数据治理项目往往都得到了领导的大力支持。

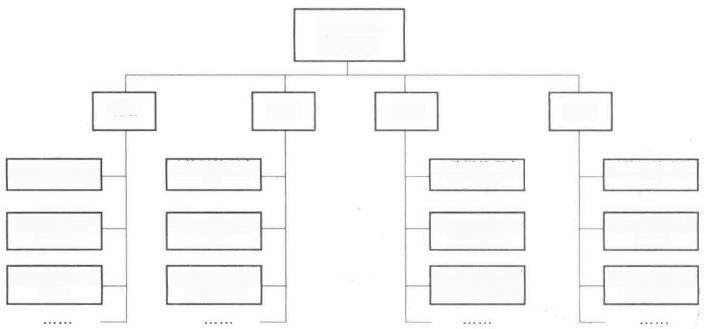
在数据治理体系日臻完善并进入良性运转后，还需要在数据认责的框架下建立考核机 制，制定奖惩措施。比如，对前台录入的数据质量进行定期考核和及时监控，在源头堵住不 良数据的入口；同时，面对呈指数级增长的数据量，数据管理部门和业务责任部门要制定数 据生命周期管理策略，通过计算数据回报率(数据回报率=数据价值/数据成本)使数据的 管理和应用逐步精细化。

健全的数据治理体系需要完善的制度保障，数据治理规章制度包括政策、制度、细则、 规范4个方面的内容。商业银行数据治理制度体系如图3-2所示。

**(1)数据治理政策** 它是数据治理工作的指导总纲，它的主要目的是明确数据治理的 工作意义及工作目标，明确数据治理整体的主要工作内容与工作要求，明确各部门、机构在 数据治理工作中的职责分工，为数据治理各领域的管理工作提供参考依据。数据治理政策适 用于总行和各分支机构，其主要内容包括：数据治理战略规划、数据治理范围、数据治理要 求和工作目标，以及数据治理工作的职责分工。

**(2)数据治理制度** 包括元数据管理、数据模型、数据标准、数据质量、数据生命周 期管理、数据使用等各领域的管理办法。其主要目的是明确数据治理各领域的主要工作内容 与工作要求，明确各部门、机构在各个数据治理领域工作中的职责分工，为总行和分支机构 开展数据治理领域的管理工作提供参考依据。数据治理制度适用于总行各部门、各分支机构 和各应用系统，其主要内容为各数据治理领域的管理办法，具体包括：各数据治理领域的管

**银行数据治理**



数据治理

制度体系

规范

元数据管理实

施细则

数据质量管理

实施细则

数据生命周期

管理实施细则

元数据管理办 法

数据质量管理 办法

数据生命周期 管理办法

元数据技术规 范

数据质量技术 规范

数据生命周期 技术规范

数据治理规划

信息标准管理

政策

数据质量管理 政策

制度

政策

细则

图3 - 2 商业银行数据治理制度体系

理范围，数据治理工作流程、职责分工、工作方法和工作要求等。

**(3)数据治理细则** 包括数据模型、数据标准、数据质量、元数据管理、数据生命周 期管理等各领域的管理实施细则。其主要目的是明确数据治理领域的主要工作环节，明确每 个工作环节中的主要活动，为分支机构开展数据治理领域工作提供参考依据。数据治理细则 适用于总行各部门、各分支机构和各应用系统。其主要内容包括：数据治理领域工作环节的 主要活动内容，数据治理领域工作环节主要活动的责任人、产出物。

**(4)数据治理规范** 包括数据模型、数据标准、数据质量、元数据管理、数据生命周 期管理、数据分布和存储等各领域的技术和操作规范。其主要目的是明确数据治理领域工作 关键概念，明确数据治理领域架构、内容、技术要求，为数据治理领域内容建设和执行等提 供参考依据和评判标准。数据治理规范适用于总行各部门、各分支机构和各应用系统。其主 要内容包括各类专项技术或操作规范。

一些典型的制度或规范领域见表3-1。

**表3-** **1** **典型的制度或规范领域**

|  |  |
| --- | --- |
| 制度或规范领域 | 主 要 内 容 |
| 数据治理工作 | 数据治理架构与职责分工，对数据治理的各领域都有一个总体的要求 |
| 数据标准 | 数据标准体系建设的职责分工、标准管理和标准应用 |
| 数据质量 | 数据质量术语定义、评估维度和规则、技术要求、管理的职责与流程 |
| 数据文件交换 | 数据交换文件的命名规范、交换原则、安全管理等 |
| 元数据 | 元数据管理原则、范围和管理方法等 |
| 数据生命周期 | 数据生命周期管理的原则，包括清理、归档、销毁等各阶段的具体技术要求 |
| 主辅数据源 | 主数据分布原则、辅数据分布原则 |
| 数据安全 | 数据安全性的原则、机制、算法等，以及生产数据使用过程中的数据变形等 |
| 数据认责 | 将数据的各项责任划分到不同的部门，明确各类数据的所有者、管理者等以及其管理职责 |

在数据治理制度上，各商业银行都在不同领域制定了相应的章程、办法、规范等。例如， 招 商 银 行 制 定 数 据 管 理 办 法 、 数 据 质 量 管 理 办 法 、 数 据 标 准 管 理 办 法 等 指 导 业 务 和 技 术 实 施

**银行数据治理保障机制**| **第3章**|

数据治理制度；光大银行在数据认则、数据标准、数据质量等方面制定了指导性制度；中国 工商银行针对业务操作制定了数据质量、数据标准的管理办法，针对技术开发制定元数据、 数据生命周期、数据质量等一系列的技术规范。

**3.4** **数据治理流程**

数据治理是一项长期而复杂的工作，除了制度的指导与约束外，还需要一套完整的、科 学的、严谨的流程来保证，以使数据治理的组织架构可以高效地运作起来，最终取得良好的 数据治理效果。综合一些商业银行的成功经验， 一个完整的全生命周期数据治理流程由事前 预防、事中监测、事后评估与整改三个阶段组成。

**1.要做到事前预防**

例如，良好的生活习惯是身体健康、预防疾病最有效的措施，完善的制度、严谨的工作 方式也是防止数据发生问题的最重要保证。上述谈到的战略文化、规章制度的贯彻落实，就 是最好的事前预防举措。此外，可以将相关举措系统化，并与日常工作有效结合，改进流 程，以便更有效地防止数据问题的发生。我国一些大型商业银行已经研发了专门的数据治理 平台，如企业级元数据管理系统、数据质量管理系统、数据生命周期管理平台等，将相关制 度规范对各数据治理领域的要求在管理平台中实现硬控制和约束，并在流转的各个环节由相 应的组织和角色负责，实施认责机制，通过这些措施，有效地预防了数据问题的产生。

**2.要加强事中监测**

在日常业务开展及信息技术运维过程中，应组织分析各领域的数据质量问题，监测报告 本系统的数据结构变化情况、数据分布情况、数据对业务服务的满足情况、在线数据增长情 况等。要做好事中监测，就要完善、细致的数据治理办法，并需要数据治理组织架构的管理 层、执行层能够认真地贯彻落实。某些大型商业银行，为确保核心系统的稳定性，专门组建 了包括应用架构师、数据模型专家、数据库管理员 (DBA) 、 专业岗位领导在内的数据结构 设计管理团队，实现全面的数据结构变更审批，并参与重点项目的数据模型设计，确保数据 结构的设计具备前瞻性、稳定性，以减少对生产运维的影响。事中监测还可以通过建立定期 例会制度(如按月),针对当前项目中涉及的数据标准、数据模型、文件、接口等基础性资 源的设计、变更情况进行研讨，确保各项基础性资源良性发展。

**3.** **要进行事后评估与整改**

定期对系统展开全面的数据治理状况评估，从问题率、解决率、解决时效等方面建立评 价指标，通过系统记录并跟踪需要整改的数据问题，要求按期整改优化，必要时进行一定的 考核。 一般事后评估与整改可以安排在项目投产一周后进行，并按年进行回顾。

**4.** **要综合各种方式进行治理**

目前，我国商业银行数据治理基本上通过系统或工具发现问题，然后由技术人员进行批 量修改，或者由业务人员进行修订或补录。其工作基本上由银行内部人员完成。对于涉及客 户的相关数据，则很难进行治理。从这一点来说，可以借鉴当前互联网企业的部分做法，发 动客户积极参与到银行的数据治理工作中，比如：

1)通过短信、电话、邮件通知客户确认相关信息。

 **银行数据治理**

2)对于重要的企业客户，通过走访确认和补充缺失与错误的信息。

3)通过活动、游戏等形式，并给予一定的奖励让客户主动完善相应信息。

**3.5** **数据治理支撑系统**

数据治理不能停留在手工管理层次，而应将相关规划、制度、规范、流程等通过系统支 撑来实现，以促进数据治理高效执行。

针对数据治理的关键领域，需要搭建配套的系统进行管理，同时由于各领域间存在着相 互协同与验证的关系，相关配套系统也需要互相引用支撑，才能有效提升数据治理的效率， 优化数据治理的效果。

1)建立数据资产管理系统，统一管理全行数据资产，包括元数据、数据模型、数据标 准，以及其他重要的企业数据资产，并提供可视化的数据查询及展现功能，支持通过功能嵌 入等方式实现数据资产的快速与便捷访问。

2)搭建数据质量管理系统，落实数据质量问题治理工作，实现问题发现、跟踪、治 理、评价的全流程闭环管理；搭建数据生命周期管理平台，落实数据生命周期管理等。

3)不断丰富前台源系统的基础数据，持续加大数据积累与整合的广度和深度，建设统 一的数据仓库平台，满足前台营销、统计分析、决策支持、风险管理和新资本协议实施等多 种需求，持续提升对数据的挖掘、分析与深度运用能力。

4)在建立上述配套支撑系统的基础上，还需要实现各系统间的互联互通，相互协同与 验证，提供数据统一、形式多样的数据服务。

例如，中国工商银行针对元数据的管理开发了元数据及软件资源管理系统，针对数 据标准建立了集团信息标准管理服务系统；针对数据质量建立了数据质量管理平台；针 对数据交换有了统一交换系统；针对数据生命周期管理开发了统一清理工具和清理展现 平台。

又如，中国农业银行建设了元数据管理系统，搭建了业务元数据和技术元数据的集中管 理和发布平台。

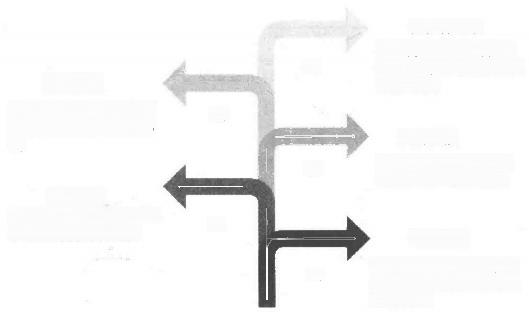
基本上各商业银行都会针对本行具体情况建立不同的应用系统和工具，用以推动数据治 理工作的全面开展。

**3.6** **数据治理成熟度模型**

CMMI 协会发布的数据管理能力成熟度模型 (DMM) 可以根据企业的数据管理能力分 为5个层次，如图3-3所示。

根据数据管理能力成熟度模型提供的分类方法，成熟度可以分为5个等级。其中，1级 为初始级(又称为可执行级),此时流程通常是临时性的，整体环境不够稳定；2级为可管 理级，流程是可重复发生的，但可能无法针对组织中所有项目重复流程，存在基本的项目管 理和流程规则，但仍有超出预期成本和时间的风险；3级为已定义级，建立了标准流程集，

**银行数据治理保障机制** **第** **3** **章**



优化管理级

在可持续基础上管理， 在反馈的基础上创新和 改进流程

已定义级

与业务战略保持一致， 标准化和一致性实施

可管理级

可执行级

被动、非正式，经常临 时性的，数据孤岛式的 数据管理

可量度级

在整个企业中可量度可 跟踪，跨职能的依赖协 调

已定义和文档化的流程， 在业务部门层面执行

01

04

02

05

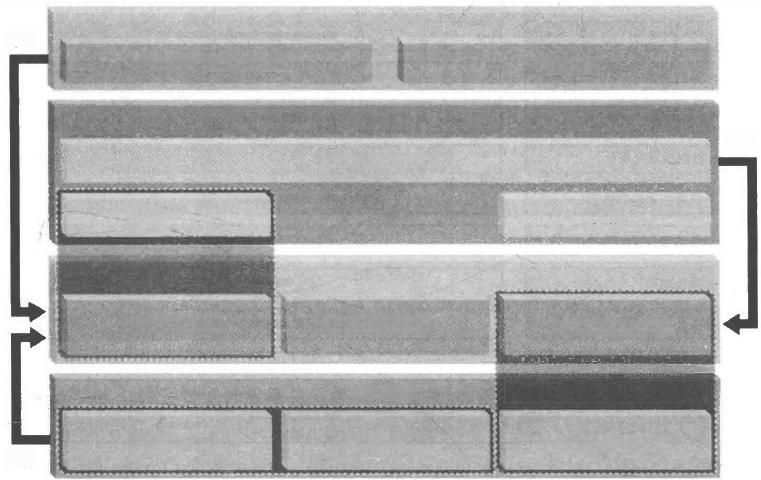
03

00 未执行

图3 - 3 数据管理能力成熟度模型

通过组织的标准流程集定制标准、流程描述和项目过程，以适应特定项目或组织单位；4级 为定量管理级(又称为可量度级),对流程进行定量度量和控制，所选的子流程大大提高了 整体流程绩效；5级为优化管理级，在该级明确了组织的定量流程改进目标，并不断优化以 适应变化的业务目标。

IBM也提出了数据管理能力成熟度模型(见图3-4)。其共包含10个类别，用于度量数 据治理能力，分别隶属于四个相互关联的组，见表3-2。



结果

需求

**数据风险管理和合规**

使能器

增强

组织结构和认知度

负责人

核心领域

信息生命周期管理

支持领域

分类和元数据

支持

信息安全和私密性

审计、日志和报告

数据质最管理

数据架构

价值创造

政策

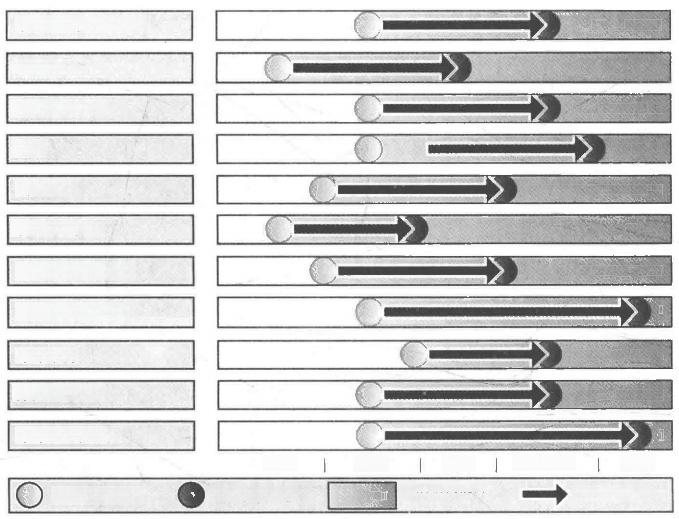
图3 - 4 IBM 数据管理能力成熟度模型

IBM 数据管理能力成熟度模型框架提供了衡量当前状态和未来状态之间差距的参考，比 如 某 用 户 其 数 据 管 理 能 力 成 熟 度 评 估 结 果 如 图 3 - 5 所 示 ， 其 成 熟 度 级 别 与 能 力 成 熟 度 模 型 一 一对应。

 **银行数据治理**

**表3-2** **数据关联组**

|  |  |  |
| --- | --- | --- |
| 关联组 | 类 别 | 内 容 |
| 结果 | 数据风险管理和合规 | 确定数据治理与风险管理关联度，用于量化、跟踪、避免或转移风险等 |
| 价值创造 | 确定数据资产是否帮助企业创造更大价值 |
| 组织结构 和认知度 | 负责人 | 用于评估企业针对数据治理是否拥有合适的数据治理委员会、数据治理工 作组和全职的数据治理人员，是否建立了数据治理章程以及高级主管对数据 的重视程度等 |
| 政策 | 为企业管理数据在高级别层面指明方向，制定质量控制规程，用于管理数据 以实现资产增值和风险控制等 |
| 核心领域 | 数据质量管理 | 主要是指用于来提高数据质量，保证数据准确性、 一致性和完整性的各种方法 |
| 信息生命周期管理 | 主要是指对结构化、半结构化以及非结构数据全生命周期管理相关的策略、 流程和分类等 |
| 信息安全和私密性 | 主要是指保护数据资产、降低风险的各种策略、实践和控制方法 |
| 支持领域 | 数据架构 | 主要是指系统的体系架构设计，支持向适当用户提供和分配数据 |
| 分类和元数据 | 主要是指用于业务元数据和技术元数据以及元模型、存储库创建通用语义 定义的方法和工具 |
| 审计日志和报告 | 主要是指与数据审计、内部控制、合规和监控超级用户等有关的管理流程 |

数据风险管理和合规

价值创造

组织结构与意识

策略

管理工作

数据质量管理

信息生命周期管理

信息安全和私密性

数据架构

分类和元数据

审计日志和报告

评估当前状态

初始

确定未来状态

受管 定义

识别所需能力

及规划

定量管理 优化

开发路线图

图3 - 5 数据管理能力成熟度评估结果



**第** **4** **章**

**数据模型**

数据模型 (Data Model) 是抽象描述现实世界的一种工具和方法，是表示实体及实体之 间联系的形式。韦伯斯特字典中对模型的定义是“对不能直接观察的事物进行形象的描述 和模拟”,即模型是对客观世界中复杂事物的抽象描述。在信息技术系统用于描述信息时， 必须抽取局部范围的主要特征，模拟和抽象出一个能反映局部世界中实体和实体之间联系的 模型，即数据模型。

数据模型是银行数据架构管理的基础领域，如：基础数据标准体系建设需要依赖数据模 型，数据分布需要依赖企业级逻辑模型，以明确主要实体的主辅应用；数据生命周期管理需 要依据数据所属主题和功能判断数据的在线、离线和归档周期。

完整、高质、稳定的数据模型有利于统一业务概念，促进信息整合和知识积累，维持数 据库稳定，提高应用开发质量，同时也可以作为技术与业务之间沟通的工具。

数据模型包括了三个部分：数据结构、数据操作、数据约束。

**1.** **数据结构**

数据模型中的数据结构主要用来描述数据的类型、内容、性质以及数据间的联系等。数 据结构是数据模型的基础，数据操作和数据约束都基本上是建立在数据结构之上的。不同的 数据结构具有不同的操作和约束。

**2.** **数据操作**

数据模型中的数据操作主要用来描述在相应数据结构上的操作类型和操作方式。

**3.** **数据约束**

数据模型中的数据约束主要用来描述数据结构内数据间的语法、词义联系、它们之间的 制约和依存关系，以及数据动态变化的规则，以保证数据的正确、有效和相容。

**4.1** **数据模型体系**

数据库设计是一项庞大的工程项目，数据库设计质量的好坏直接影响系统中各个处理过 程的性能和质量。为了更好地进行数据库设计，业界提出各种数据库设计方法，并运用软件 工程的思想和方法，提出各种设计准则和规范。其中关键一点就是数据模型体系的设计。商

 **银行数据治理**

业银行的数据模型体系一般分为4层，包括：企业级概念数据模型、企业级逻辑数据模型、 应用级逻辑数据模型、应用级物理数据模型。

一个典型的数据模型项目或新系统开发项目中的数据库设计工作可包括以下阶段：需求 收集阶段、分析与设计阶段、实现阶段。

**1.需求收集阶段**

概念数据模型 (Conceptual Data Model,CDM) 是需求收集阶段的阶段性结果，它包含 了重要的业务概念以及需求，并且通过相应的方法论将这些高级的业务概念关联起来。

**2.分析与设计阶段**

逻辑数据模型 (Logic Data Model,LDM) 是分析与设计阶段的产物，将高级的业务概 念以数据实体或属性及其关系的形态在逻辑层面更详细地表达出来，主要的表现形式是 ERD (实体关系图)。

**3.** **实现阶段**

物理数据模型 (Physical Data Model,PDM) 是考虑到物理实现上效率的实现以及很多 限制而获得的数据模型，是最终结合数据库系统实现的产物。

企业级概念数据模型指导企业级逻辑数据模型，企业级逻辑数据模型指导应用级逻辑数 据模型，应用级逻辑数据模型反向指导与完善企业级逻辑数据模型，并指导生成应用级物理 数据模型。

设计一个完善的数据库应用系统不是一蹴而就的，它往往是上述几个阶段的循环反复。 如果没有科学的数据库设计方法，数据库设计质量难以保证，往往使数据库运行一段时间后 出现各种不同程度的功能或性能问题，无形之中增加了系统维护的代价。为保证数据模型的 设计质量，数据模型的设计应遵循先概念数据模型，再逻辑数据模型，最后物理数据模型的

顺序。

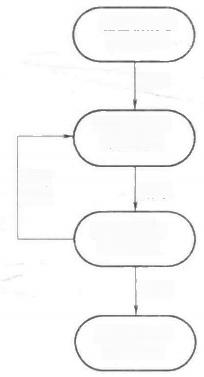
图4-1 所示为数据模型体系。

**(1)概念数据模型** **(CDM)** 概念数据模型用于表示数据的 逻辑特性，即只是在概念上表示数据库中将存储什么信息，而忽略 这些信息的实现细节。同时，它也是对系统主要实体的高层次业务 见解，比如识别关键主题领域、定义核心实体的主键。

**(2)逻辑数据模型** **(LD** **M)** 逻辑数据模型即实体关系模型 (Entity Relation Model),是一种描述数据的模型。它利用实体和它 们之间的关系描述包含在现实世界中的数据。这是一个比较详细的 数据业务见解，比如细化实体间关系，详细的属性定义(主外键、 索引等重要属性),添加了关联的、特色的以及子类的实体，尽可 能详细化的范式(遵循第三范式),实体间的约束关系等。

**(3)物理数据模型** **(PDM)** 物理数据模型是针对具体的关系

型数据库系统 (Retational Database Management System,RDBMS), 对逻辑数据模型进行物理模型设计，形成可供具体的关系型数据库



企业级概念

数据模型

指导

企业级逻辑 数据模型

指导

应用级逻辑 数据模型

指导

应用级物理 数据模型

反向 指导

图4- 1 数据模型体系

系统中实现的表结构。它需要考虑具体数据库系统的实现，比如：考虑具体数据库系统实现 方式(用什么数据库，怎么分区等);考虑运行性能(如反范式化，数据部署);考虑使用 索引的类型 ， 是否建立物理主键 ； 考虑数据安全控制 ； 考虑数据恢复和重新开始的方法 ； 考

数据模型 **第4章**

虑记录数，记录增长的估计。

像 IBM、Teradata 等信息技术公司针对金融机构也制定了具体的数据模型体系，比如 IBM的金融服务数据模型 (Financial Service Data Model,FSDM),见表4-1。

**表4-1** **IBM的金融服务数据模型**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 模型类别 | | IBM方法论中的数据模型 | |
| 企业级 应用级 | 业务需  求要点  企业级的概念数 | | A级 ●可以用九大数据概念来描述所有金融服务信息  ·给B级的结构和内容定义了约束 | |
| ·用层次化的概念术语来组织业务概念  B级 ·不关心信息之间在具休应用中进行交互的业务规则 ·采用业务人员熟悉的语言 | |
| 巡掘数据模型 | | C级 | ·用ER图描述业务信息，包含分类实体、关系及属  性定义  ·不考虚实施层面的约束，只维护逻辑视图  ·对各方面利益相关者的特有视图进行统筹整合 |
|  |  | C'级 | ·用ER图描述应用数据  ·考虑具体实施层面的约束，引入派生数据  ·允许可控的数据冗余 |
| 物理数 | 据模型 | D级 | ·描述物理数据结构  ·考虑系统部署层面的约束 |

**4.2** **数据建模的方法、原则及意义**

**1.数据建模考虑因素**

数据建模时需要考虑以下三个方面：

1)是什么?

通过子类来精确分类事物；概念层面的分类将概念分成一系列的子类；子类还可以继续 划分子类。

2)做什么?

通过事物间的关系来体现事物的动态特征，每个关系通过关系类型的不同加以区别。 3)有什么?

通过各种属性描述来体现事物的静态特征；描述层面的分类表达概念有一系列的相关属 性；每个属性通过描述类型的不同来区别。

图4-2所示为数据建模时应考虑的因素。

**2.** **数据建模的原则**

**(1)继承性原则** 遵循已有概念主题，并在已有主题基础上进一步细化，不提出新的 概念。

**(2)稳定性原则** 对业务功能进行抽象概括，确保核心实体相对稳定，且提供系统性 视图。

 **银行数据治理**

**(3)前瞻性原则** 为保证模型具有前瞻性，模 型适当超前设计，这样能够适应业务的发展变化。

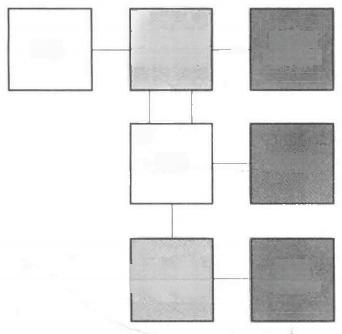
**(4)保鲜性原则** 物理表与应用逻辑模型建立 对应关系，并及时更新与维护逻辑模型，以保持模 型的新鲜性。

根据上述建模体系以及建模方法，整理完成的 各级数据模型概貌从上到下依次为概念数据模型、 逻辑数据模型、物理数据模型。

**3.数据模型的意义**

数据模型治理意义体现在：

(1)统一业务概念 使业务和信息技术之间的 沟通更准确、更有效，对新系统的需求和范围的理



②

事物与事物

之间的关系

①

事物 事物类型

③

事物的描述

事物间 关系类型

事物 描述类型

事物

图4-2 数据建模时应考虑的因素

解更准确。

(2)加强数据标准 基于企业概念数据模型，对关键数据进行标准化编码工作，以企 业级数据模型内容为框架进行标准的制定。

(3)促进信息整合 包括信息入口的整合和信息的事后整合。这样有利于控制表结构 粗犷式增长，信息整合能最大化发挥价值作用。

(4)发挥信息价值 促进企业精细化管理；统一客户视图，提供更好的客户服务。

(5)提速应用开发 整体视图可以加快新应用的开发，更快地响应客户服务。新应用 的开发更符合企业级的数据规范，避免出现新的信息孤岛。

(6)提高应用开发质量 通过深入学习与了解数据模型，有助于提高开发人员在开发 过程中程序设计的完整性和系统性。

商业银行数据模型应作为一种模型资产，全部纳入系统进行统一管理，并与企业级元数 据管理系统相关联，通过两者建立关联关系，实现高阶思想的落地实现。

通过实体所属概念主题分类，为数据生命周期管理、数据分布管理、数据安全奠定基 础。每一个表都可以依据它归属的主题，以及银监会和巴塞尔协议要求，明确其在线、离 线、归档保存周期。每一个表都可以依据它归属的主题，明确其主数据来源及辅数据分布原 则。同时，通过将上述信息纳入企业级元数据管理系统，可实现实时跟踪和记录数据生命周 期管理及数据分布的符合情况。

**4.3** **数据模型建模要求**

**1.** **数据模型包含的内容**

数据模型建模标准主要针对企业级逻辑数据模型和应用级逻辑数据模型。企业级逻辑数 据模型和应用级逻辑数据模型应包含逻辑数据模型 ER 图和逻辑数据模型说明文档两部分 内容。

**(1)逻辑数据模型** **ER 图** 逻辑数据模型 ER 图应至少包含以下内容：主题域视图 (Stored Displays)、实体、关系。

**数据模型** |**第4章**

1)主题域视图应至少包含主题域视图的名称、定义。主题域视图的定义，应用业务语 言描述主题域视图所表达的主要内容。

2)除与业务部门无关的应用(例如纯粹的工具)无法按照主题画出逻辑数据模型 ER 图外，所有应用的主题域视图划分应与主题的子类、小类保持一致，并至少细化到主题子 类，可补充按业务场景划分的主题域视图；判定一个应用是否与业务部门无关，需经过相关 审议通过后进行登记管理。

3)主题域视图里面应包含本主题的实体，以及与其具有第一层关联关系的其他主题的 实体和关系。

4)实体应至少包含实体名称、实体定义、实体主键、普通属性。

5)实体间的关系分为标识性关系、非标识性关系、多对多的关系，需要选择其中一种 作为两个实体之间的关系。

6)各应用数据分布应遵循主题的分布规则。

**(2)逻辑数据模型说明文档** 逻辑数据模型说明文档应包含以下内容：

1)文档说明、术语定义、本次版本与上次版本的变更内容描述。

2)数据模型内容说明、实体清单，可适当补充场景分析。

**2.模型设计应与项目结合**

模型设计流程一般包括概念数据模型(需求收集阶段)、逻辑数据模型(分析与设计阶 段)、物理数据模型(实现阶段)三个阶段，与项目各阶段的对应关系见表4-2。

**表4-2** **模型设计各阶段项目及对应关系**

|  |  |  |  |
| --- | --- | --- | --- |
| 模型设计阶段 | 对应项目生命  周期各阶段 | 与数据库设计相 关的输出文档 | 描 述 |
| 概念数据模型设 计阶段 | 项目定义阶段、需 求分析阶段 | 项目方案、业务 需求 | 将需求分析了解到的数据库相关内容体现到方 案中，并将相关数据模块体现在业务需求分析中 |
| 逻辑数据模型设 计阶段 | 功能设计阶段 | 各应用的逻辑数 据模型ER图、逻辑 数据模型说明文档 | 以业务需求分析为依据进行分析，完善项目所 涉及应用的逻辑数据模型ER图及相应的说明  文档 |
| 物理数据模型设 计阶段 | 系统设计阶段 | 系统设计说明 | 完成物理数据模型，完成设计说明 |

**(1)概念数据模型设计阶段** 本阶段是数据库设计的起点，面向需求收集，主要任务 是基于系统的设计思路、开发原则、功能需求和数据需求，充分了解系统包含的业务概念以 及需求，并且通过相应的方法论将这些高级的业务概念关联起来。

**(2)逻辑数据模型设计阶段** 逻辑数据模型设计阶段的主要工作是应用级逻辑数据模 型建模，在跟随项目中加以完成。以业务需求为依据进行分析，形成和维护该项目所涉及各 个应用的应用级逻辑数据模型的 ER 图及相应的说明文档。逻辑数据模型设计独立于特定类 型的关系型数据库系统及物理实现，是用来描述实体关系的模型。

**(3)物理数据模型设计阶段** 物理数据模型设计阶段包括对逻辑数据模型的反范式化 和表结构物理实现两部分，主要面向性能需求，目的是提高数据库系统运行与维护效率。本 阶段的主要任务是通过事务分析确定主要逻辑实体的访问特性，最终确定数据信息的存储模 式、物理表结构设计原则等。

 **银行数据治理**

**3.** **数据模型设计要求**

为了更好地设计数据模型，应该明确一些设计原则来规范数据模型的设计，以下为推荐 的一些模型设计要求。

**(1)概念模型** 进行概念模型设计时应充分收集与了解系统的功能需求和数据需求， 整理数据库设计思路、开发原则，并完成以下几点内容：

1)分析并形成简要的ER 图(仅含主键、外键等重要属性),在项目方案中完成数据结 构的概要设计。

2)业务需求需要体现业务数据量和数据量变化趋势分析、需要保留的业务数据(或账 务)和处理原则、业务数据的可变性分析。

3)从数据质量管理角度，充分了解系统功能对数据信息在各个数据质量维度的业务规 则需求。

**(2)逻辑数据模型**

1)应用级逻辑数据模型的设计工作应接受企业级概念数据模型及企业级逻辑数据模型 的指导。可以直接在企业级逻辑数据模型中选择业务数据所涉及的内容，作为逻辑数据模型 所包含的实体、关系和属性。

2)根据业务需求来设计应用级逻辑数据模型，同时根据业务需求明确功能模块需要存 储的各类数据信息，分析数据相互间的关系，明确信息实体、实体间关系、属性及组合等性 质，形成应用级逻辑数据模型。

3)逻辑数据模型不低于第三范式。

**(3)物理数据模型**

1)物理数据模型设计阶段需要明确数据各维度的业务规则和技术规则，对于数据质量 管理的要求可以通过数据定义语言 (DDL) 设置实现。

2)物理数据模型设计阶段需要明确各表的数据生命周期要求及策略。

3)表结构需登记维护信息，比如：创建人、创建日期、创建时间，最后更新人、最后 更新日期和最后更新时间。

4)主键、外键的建立以逻辑数据模型 ER 图为依据。主键、外键根据是否通过 DDL 在 数据库中物理实现，区分为逻辑主键(或外键)和物理主键(或外键)两类。表结构的物 理主键(或外键)是指在数据库中真正物理实现的主键(或外键),逻辑主键(或外键) 是指逻辑上存在主外键关系但实际不需要物理实现的主键(或外键)。

5)表结构字段设计应遵循数据标准，统一制定各类数据的字段类型及长度等。如果不 存在相关数据标准，则还需要提出相关数据标准的新增申请。

**4.4** **数据模型设计实例**

结合某商业银行具体案例，对逻辑数据模型、物理数据模型设计展开详细描述。

**1.** **围绕功能模块构建局部逻辑模型**

相对于需求说明和项目方案，系统功能说明能够从功能上对系统进行进一步细分，更大 程度上体现和具体化数据的关联和用户对数据的处理要求，方便建立更为科学的局部逻辑模

**数据模型** **第4章**

型——围绕子功能模块建模。若要开发的系统较小，涉及的数据项较少，则可直接建立全局 逻辑模型，即一开始就考察分析全部的数据。具体步骤如下：

**(1)数据提取、局部数据分析** 系统功能说明要包含系统的每个功能模块，以及每个 模块的流程图、流程说明、输入和输出等。研究每个功能模块，结合需求说明、项目方案以

及数据结构设计，收集相关数据项。

1)该功能模块要求输入以及从其他系统采集或从其他文件导入的数据。 2)该功能模块输出的数据。

3)与程序员实现该功能模块编码有关的数据，如参数表等。

4)其他数据，考虑是否有不包含在以上三类数据中的数据(以上基本包括需要数据库 存放的绝大部分数据。可依据系统实际情况与用户讨论是否存在未包含在这三类数据中的其 他数据)。

**(2)确定数据信息**

1)每个数据项的名称、别名、类型、长度、取值范围、数据量的大小和代码等基本 特征。

2)数据项之间的联系。确定数据项之间的联系(一对一、 一对多、多对多),考虑数 据项之间是否有层次结构，例如银行机构有总行、 一级分行、二级分行等层次。

3)数据的来源。注意是否有冗余的来源，以及来源不同的数据是否有矛盾。

关于输出数据，若只是简单的查询结果，则将其排除；若是复杂判断、计算的结果，则 需要收集。

考察项目方案中实体与所收集到的数据项的关系，对收集到的数据项进行分析和归类 (可能会产生新的实体，即在项目方案未定义的实体)。 一般这些数据项之间没有绝对的分 界，但可以先根据这些数据项的语义进行初步归类。

下面以该商业银行的机构管理系统为例，介绍收集、分析、归类数据项的方法。该系统 由多个模块组成，在此以其中的“手工录入”模块为例。假设在项目方案中已定义了实体 “机构基本情况”。系统功能说明对该模块的介绍如下：

步骤1:手工录入模块

该功能模块提供给用户录入、修改某些指标值和经营数据。具体手工录入项见表4-3。

**表4-3** **手工录入数据**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 实体 | 属性 | 是否必须输入 | 实体 | 属性 | 是否必须输入 |
| 机构基本情况 | 机构名称 | 是 | 领导班子 | 姓名 |  |
| 机构基本情况 | 机构地址 | 是 | 领导班子 | 员工编号 | 是 |
| 机构基本情况 | 机构类别 |  | 领导班子 | 出生日期 |  |
| 机构基本情况 | 联系电话 |  | 领导班子 | 性别 |  |
| 机构基本情况 | 机构标识 | 是 | 领导班子 | 民族 |  |
| 机构基本情况 | 直接上级机构 | 是 | 领导班子 | 参加工作时间 |  |

考察机构信息录入模块、领导班子信息录入模块的输入数据，同时结合一些系统需求说 明等资料，选取以下数据项：机构标识、机构名称、机构类别、机构地址、联系电话、直接 上级机构，以及员工编号、姓名、出生日期、性别、民族、参加工作时间。

 **银** **行** **数** **据** **治** **理**

输出数据是操作成功或失败信息， 一般这类信息不放入数据库存储；模块所需的控制类 数据在对模块进行详细设计、编码时确定。关于此类数据的建模处理与输入输出数据类似， 在此忽略该模块的此类信息。

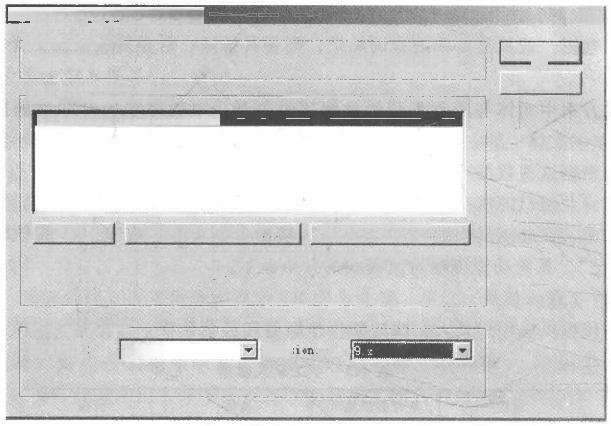
下面确定这些数据项的类型和取值范围。考察收集到的数据项，依据数据项的语义以及 语义间的关系，将数据项“机构标识、机构名称、机构类别、机构地址、联系电话、直接 上级机构”归入项目方案所定义的实体——“机构基本情况”;将“员工编号、姓名、出生 日期、性别、民族、参加工作时间”归类为新的实体——“领导班子”。

这一步收集到的数据并不一定能做到完整，有些系统可能比较特殊，所需数据没有体现 在模块的输入、输出以及编码当中。这些数据的确定可能需要在以后数据库设计步骤中或相 应数据库的前台应用系统设计过程中才能确定，需要数据库设计者随时关注数据的完整性， 并做好沟通。

步骤2:标识实体 (Entity)

局部逻辑模型主要由实体、属性和关系组成，构建局部逻辑模型的第一步就是标识模型 的实体。

上面已对数据项进行了初步分类，每一类就是一个实体，而类中的每个数据项就是实体 的属性。在逻辑数据模型 ER 图中创建实体，标识实体的名称和定义，如图4-3所示。



Create Model-Select Ienplate

-New Model Type-

C Logical C Physical

Cancel

Create Vsing Template:

Bank Loeicrl.Phvsicsl Model

Remove Browse Eile System, pfowse Modelhart

Creates a new model with both logical and physical levels(ERwin classic)and defanlt settings.

Tar get Database

Database: ORACLR Yersi

C Logigal/Physical

刘

DK

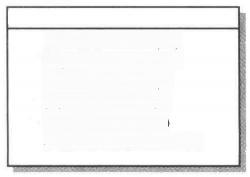
图 4 - 3 创 建 实 体

可以建立单独的逻辑或物理模型，也可以在开始同时选择两者。本文介绍的是逻辑、物 理设计的方法，所以在此处选择 “Logical/Physical”, 这样在进行从逻辑到物理的转换时可 以直接转换。此外，物理设计与特定 DBMS相关，因此在物理设计部分，本文将以 Oracle 数 据库设计为例介绍物理设计的方法步骤。

在上面的例子中，标识出了两个实体，即机构和领导班子，如图4-4所示。

步骤3:标识属性域 (Domain)

对于在多个分类(实体)中出现的数据项，可以标记这些项的属性域，这样对于这些 项可以直接在域中维护，无须到每一实体中去修改。

**数据模型** **第** **4** **章**

机构

邻导班子

领导班子

员工编号：CHAR(16)

姓名： VARCHAR(30) 出生日期： CHAR(8)

性别： NUMBER(1)

民族： NUMBER(2)

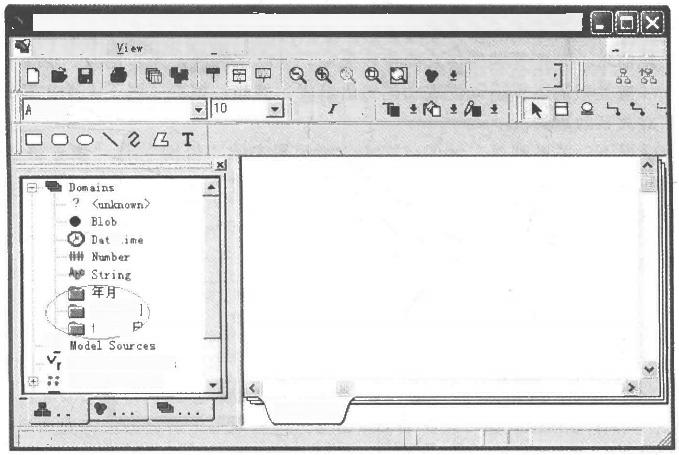
参加工作时间：CHAR(8)

|  |  |
| --- | --- |
| 保存所有机构的基本信息 |  |

|  |
| --- |
| 保存机构领导班子的信息 |

图4-4 实体标识

属性域是一组值的集合，用来指定属性的允许值集合、大小和格式，多个属性可以共用 一个属性域。不是所有属性都需要使用属性域，具体是否使用要根据用户的需求决定。如图 4-5所示，假设“年月”“修改时间”“修改用户”在多个实体中出现，则可以为它们定义 属性域。



**ERvin- [敷据库模型Y1.11.ER1\*:**

Model MedelMart Tools Findow Help

|Logical D-0

Arial

eti

修改时间

修改用F

Validation Rules

Default Values

Display1

ORACLE

**Associates**

Format

**Computer**

Eile Edit

**系统指标/** **.** **.** **.**

B U

Ready

明

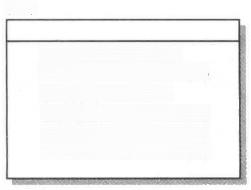
×

图4-5 属性域的定义

步 骤 4 : 标 识 属 性(Attribute)

标识完实体后，下一步就是标识这些实体的属性。根据步骤1对数据项的分类，将每个 数据项加入到相应的实体中，定义属性的数据类型、长度、默认值和取值范围。

在上面的例子中，为两个实体“机构”和“领导班子”标识属性，如图4-6所示。



机构

机构标识： CHAR(10)

机构名称： VARCHAR(30) 机构类别： CHAR(2)

机构地址： VARCHAR(60) 联系电话： VARCHAR(30) 直接上级机构： CHAR(10)

图4-6 实体属性的标识

可以为逻辑数据模型 ER 图中的属性创建新的数据类型。如果用户自己创建新的数据

 **银行数据治理**

类型， 一定要为这个数据类型设置对应于物理图的数据类型，否则会在以后的使用中

报错。

在标识属性时，需要区分以下几种属性：

**(1)单值或多值属性** 大多数属性都是单值的，但有时也存在具有多个值的属性。例 如：联系电话， 一个机构可能存在多个联系电话。

处理多值属性的方法有两种： 一是将多个值连接组合成一个单值属性(如多个联系电 话可以看成一个字符串);二是将其规范化成一个新的实体。

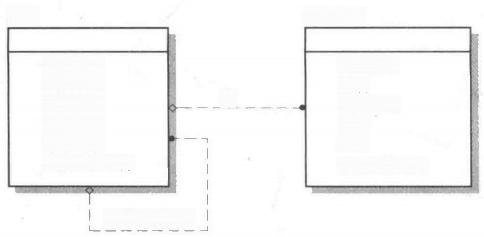
**(2)派生属性** 派生属性是指可以通过别的属性得到的属性。例如：若Z 的值是X 列 的值与 Y 列值的和，则称Z 的值为派生属性。

记录下多值属性和派生属性，以备在后续步骤规范化中使用。

步骤5:标识关系 (Relationship)

根据系统需求，判断哪些实体之间存在关系，并添加对应关系。同时，还要为每个关系 定义名称。 一般关系的名称由动词或动词短语组成，例如：拥有、属于、存在等。关系的名 称应有利于对关系的理解。

大多数情况下，关系都是二元的。图4-7所示为实体“机构”和“领导班子”之间的 关系，但有时也会存在递归关系或三元关系，图中实体“机构”就存在一个递归关系(直 接上级机构也是一个机构)。



机构

机构标识

机构名称

机构类别

机构地址

联系电话

直接上级机构

存在上级机构

员工编号

姓名

出生日期

性别

民族

参加工作时间

领导班子

有

图4 - 7 实体关系的标识

标识完关系后，检查每个实体，检查实体中是否存在隐式的关系没有被标识，检查每个 关系是否是必需的，并检查是否存在陷阱(包括扇形陷阱和深坑陷阱)。

通常每个实体都存在关系，如果发现存在没有关系的实体，则应该检查这个实体是否在 模型其他地方已经出现过，然后重新检查需求，以确认没有漏掉某些关系。如果关系均已被 标识，再与用户讨论这个特殊实体的使用。

步骤6:确定候选键 (Candidate Key)、主键 (Primary Key)

这一步的目标就是为每个实体标识候选键，然后从中选择一个作为主键。

从 候 选 键 中 选 择 主 键 应 遵 循 以 下 原 则 ：

1)相对简单的候选键。例如，若 “A” 及 “ (B,C)” 都是候选键，则 “A” 比较 简单。

2)值很少变化的候选键。

3)在未来一直都具有唯一性的候选键。

4)关于键值，对文本来说选取字符最少，对数字来说选取最大值最小的候选键。因为

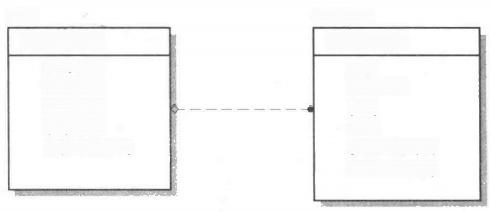
**数据模型** **第** **4** **章**

子表的行数一般比较多，字表的外键所占空间越少越好。

5)若同时存在字符、数字型候选键，则优先选择数字型候选键作为主键。 6)从用户的观点选择最易于使用的候选键。

当从候选键中确定了主键之后，其余的候选键则称为替换键 (Alternate Key)。

在上面的例子中，“机构”实体中只有“机构标识”是候选键，因此也就是该实体的 主键；同理， “领导班子”的“员工编号”是该实体的主键。实体主键标识如图4-8 所示。

机构

机构标识

机构名称

机构类别

机构地址

联系电话

直接上级机构

领导班子

员工编号

姓名

出生日期

性别

民族

参加工作时间

机构标识(FK)

图4-8 实体主键标识

因为已经在步骤5定义了实体间的关系，所以在添加或选定实体的主键时，会自动依据 实体间的关系产生有联系的实体的外键。有时为了体现逻辑模型的可读性，可将实体里的外 键改名(改名后称为外键的别名)。另外，可以自动维持外键别名与关联的实体的主键之间 的参照关系。

步骤7:特化/泛化实体(可选)

在逻辑模型中，可能会出现以下情况：两个或多个实体存在共同的属性，而这些属性又 是实体必须的部分，这就造成逻辑模型的冗余。

这种情况可以利用超类和子类来解决。超类是一个实体，包含若干个实体中出现的公共 属性；而子类是存放非公共属性的实体，同时具有相应超类中的属性特征。子类可以区分不 同的角色，比如员工是一个实体，而经理、普通工人都是员工，同时经理和普通工人之间有 差异，此时员工实体是经理实体、普通工人实体的超类，而经理实体和普通工人实体是员工 实体的子类。

构造超类和子类的过程叫作特化/泛化。特化是一个自顶向下的过程，它先定义超类集 合，然后定义有关的子类，这些子类集合的定义基于超类实体的一些特别特征。泛化是一个 自下而上的过程，它从初始的子类中产生泛化的超类，是特化的逆过程。

超类和子类之间的关系根据子类是否包含超类的所有对象，可分为完全子类 (Complete Subtype) 和不完全子类 (Incomplete Subtype)。完全子类是指子类包含了超类的所有实体； 不完全子类是指不能确定子类是否包含了超类的所有实体。

此外，超类和子类的关系可分为包含关系 (Inclusive Subtype Relationship) 和排外关系 (Exclusive Subtype Relationship)。包含关系指的是超类的一个实体可以属于一个或多个子 类；而排外关系的超类中的一个实体最多只能属于一个子类。

下面是一个总账科目子类的例子，这是不完全子类(见图4-9),超类与子类的关系属

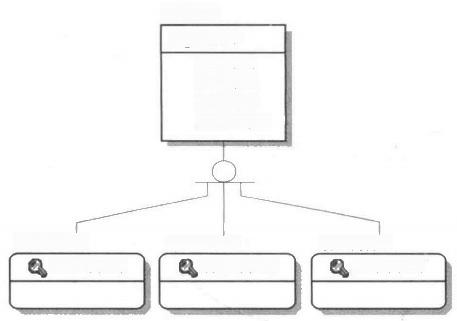
于排外关系。

 **银行数据治理**

因为总账科目还包括了权益类、费 用类等科目，所以图4-9所示关系是个 不完全子类； 一个科目不能既属于资产 类又属于负债类科目，所以这种关系也 是排外关系。

步骤8:删除与关系模型不相容的 特性

在已创建的局部数据模型中，可能 存在一些与传统的关系型数据库系统不 兼容或关系型数据库系统不便于处理的 结构，如：多对多关系、多对多递归关 系、复杂关系等。在这一步，要将这些 数据结构转换成能被或便于关系型数据 库系统处理的形式。



总账科目

② 科目号

科目级别

科目分类

科目名称

上级科目号

收入类科目

科目号(FK)

科目号(FK)

科目号(FK)

资产类科且

负债类科且

图4-9 不完全子类

**(1)删除多对多关系** 对于多对多关系，要进行分解，并标识一个中间实体，即用两 个一对多关系来代替这个多对多关系，并标识出它们之间的新实体，与多对多关系相关的属 性应该成为这个中间实体的属性。

例如有两个实体，用户组和用户，每个用户组可以拥有多个用户，每个用户可以属于多 个用户组，它们之间是多对多关系，可以通过插入一个实体“用户组的用户”将这个多对 多关系分解为两个一对多关系。

**(2)删除多对多递归关系** 对于存在的多对多递归关系，要进行分解，并标识一个新 实体，方法与删除多对多关系类似。

首先，将多对多递归关系的一个实体描述成两个实体，使递归关系和普通多对多关系一 样，然后引入一个中间实体，分解多对多关系，然后重新组合被分开的实体，注意不要丢失 任何关系。

**(3)删除复杂关系** 复杂关系是三个或更多实体之间的关系。如果逻辑模型中存在这 种关系，应该将这种关系分解成多个一对多关系，标识一个中间实体。

对局部逻辑模型进行以上的处理之后，最后还要检查所有的一对一关系，看这些一对一 关系的实体是否表示一个对象，如果是的话，可以将两个实体合并成一个实体。

通过上面的步骤，可以删除数据模型中关系数据库难以实现的数据结构，简化数据 模型。

步骤9:列出事务、检查模型是否支持用户事务

这一步的目标就是检查局部逻辑数据模型，确保模型支持用户所需要的事务。

事务是指一系列用来完成某一功能的与数据库相关的动作。

事务代表了现实世界的事件，例如：查询机构的当月等级得分，添加新的机构，审批机 构变更等。

根据对数据库操作类型的不同，事务可分为检索事务、更新事务和混合事务。检索事务 主要是用来查询信息的事务，更新事务用于数据库记录的增删改，混合事务是既查询数据又 增删改数据的事务。例如：查询机构的当月等级得分就是检索事务。

**数据模型** 第4章

前提是要列出用户所需事务，见表4-4。

**表4-4** **用户所需事务**

|  |  |
| --- | --- |
| 序号 |  |
| 事务名称 |  |
| 事务类型 |  |
| 访问的实体 |  |
| 备注 |  |

检查模型是否支持用户事务的方法：在逻辑数学模型 ER 图中，通过描述每个事务的需 求，检查模型是否提供了事务所需的所有信息，如实体、关系和属性，以及这些信息是否已 经通过主键/外键连接起来。

步骤10:用规范化方法检查数据模型

用规范化方法检查每个实体，从实体中去除不需要的属性或集合，以避免不必要的数据 冗余和数据操控语言 (DML) 异常。

范式理论建立在函数依赖和多值依赖基础上，要具有良好的数学基础。对于每一实体 (或关系——用主键代表实体得到关系间的函数依赖),可以先依据属性间的语义关系获得 一个初始函数依赖集，然后利用规范化函数依赖理论中“函数依赖集最小覆盖算法”对函 数依赖集进行极小化处理，去除冗余数据。

利用此方法可以使逻辑模型获得较好的结构，但是这是非常烦琐且费时的工作，因为对 每一实体都需要重复这样的过程，并且该方法依赖于初始函数依赖集，而该依赖集是由设计 者根据业务、属性语义确定的，没有方法确定这样的一个依赖集是否完全、合理。

基于这些原因，上述方法在实际操作中往往不被设计者采用。在此建议使用人工分析的 方法，不再利用“函数依赖集最小覆盖算法”来去除冗余。具体方法介绍如下：

对每一实体，考察其属性之间的语义关系，建立该实体的函数依赖集，然后根据各范式 定义，通过人工观察与分析，依次去除冗余，使之满足各种范式要求。 一般模型要求达到第 三范式。

对于关系，用主键代表实体，建立关系的函数依赖集，然后根据范式定义，通过人工观 察与分析，依次去除冗余关系。

下面说明每一范式的具体规范化方法，操作在逻辑数据模型 ER 图中进行。

第一范式(1NF): 实体的属性必须是原子的。必须去除实体中的多值属性，并将它们 放入一个新的实体。例如上面例子中实体“机构”中的属性“联系电话”,每个机构可能登 记有多个电话，这是个多值属性，根据第一范式，必须为这个属性创建一个新的实体。

第二范式(2NF): 遵循第二范式的实体首先必须遵循第一范式，而且它的所有非键属 性要完全依赖于主键中的每个组成键(满足1NF, 消除部分依赖)。

第二范式只针对具有复合主键的实体。当实体的主键为复合主键时，非键属性不能只依 赖于主键的一个部分。

对于下面的例子，实体“机构资产质量”用来表示某机构在某个月的资产质量情况， 它的主键是“机构编号”和“年月”的复合主键。但是，非键属性“机构名称”却只依赖 于机构编号，所以违反了第二范式，应该将这个属性拆分出来，如图4-10所示。

 **银行数据治理**

第三范式(3NF): 遵循第三范式的实体首先必须遵循第二范式，而且它的非键属性必 须直接依赖于主键，而不能依赖于其他非键属性(满足2NF, 消除传递依赖)。

检查每个实体的非键属性，看是否有非键属性依赖于其他非键属性。

下面的例子中，实体“机构”的非键属性“负责人姓名”依赖于非键属性“负责人编 号”(虽然“负责人姓名”依赖于“机构编号”,但这是传递依赖而不是直接依赖),所以 这个实体违反了第三范式，应该将这个属性拆分，如图4-11所示。

机构资产质量

|  |
| --- |
| 机构编号  年月 |
| 机构名称  次级贷款  可疑贷款  损失贷款 |

图4-10 机构资产质量实体

机 构

机构编号

机构名称

机构类别

负责人编号

负责人姓名

图4-11 机构实体

步骤11:检查、设置完整性约束

完整性约束是为了防止数据模型不一致而强加的限制，标识完完整性约束之后，就可以 得到一个完整而准确的描述视图的局部逻辑数据模型。在此主要考虑下面五种类型的完整性 约束：

**(1)必需的数据** 某些属性必须包含某些值，即不允许为空。例如：每个机构都必须 有机构名称。

**(2)值域约束** 每个属性都有一个值域。已经在步骤4设置了属性的值域。

**(3)实体完整性** 实体完整性是指实体的主键不能为空。

**(4)参照完整性** 参照完整性是指子实体的外键值必须与父实体的主键值相同或为空。 要实现参照完整性，需要定义一些约束，定义主键和外键在什么条件下可以插入、更新或删 除，以及这些操作的结果。

一般有以下三种关于参照完整性的约束 (Oracle 支持这三种约束): Restrict 、Cascade、 None(Oracle 用 NULL)。

**(5)业务规则** 最后，还要考虑业务规则的约束。数据库要被现实世界的业务要 求约束。例如：对机构的领导干部编制有限制， 一级分行机构最多只能有100个干部 编制。

通常业务规则有两种实现方法： 一是通过属性的检查约束 (Check) 实现，如上面 的例子就可以通过检查约束实现；二是通过触发器来实现，这一步需要在物理设计时 完成。

这五种约束已经在步骤3到步骤5中设置，此处加以检查与修改。

步骤12:与用户讨论局部逻辑模型

与用户讨论逻辑模型，对于不符合要求的地方要做相应的修改，以确保模型能体现企业 的需求。

**2.** **构** **建** **并** **检** **查** **全** **局** **逻** **辑** **模** **型**

在这个步骤中，要将各个局部逻辑数据模型组合成全局逻辑数据模型，并检查合并过程

**数据模型** **第4章**

中有变化的区域，以确保得到符合要求的全局逻辑数据模型。

对于一些实体比较少、比较简单的模型，可能没有划分局部逻辑模型，可以略过这 一步。

步骤1:将局部逻辑模型合并为全局模型

至此，已经完成了对局部逻辑数据模型 ER 图的设计，在这一步，要将每个局部逻辑数 据模型的 ER 图合并成全局逻辑模型。

对于实体较少的简单数据模型，将局部模型合并起来相对简单，但对于大型系统，必须 采取系统化的方法才能合并局部模型。合并局部模型时需要完成的任务包括：

1)检查实体的名称和主键。检查每一局部逻辑模型，去除重复的实体，保证实体与实 体名字的一一对应关系。

2)可能有些实体名字相同，但实际上不是同一个实体，此时需要对实体进行改名以区 分实体；或者实体相同，但有不同的名字，此时要改名使实体名字统一。检查时需要比较模 型中每个实体的内容。

3)检查关系的名称。检查局部模型中的关系，看是否存在重复关系。 一般只有重复的 实体间才会存在重复关系。

4)合并来自局部模型的实体。合并的情形有：合并名称相同、主键相同的实体；合并 名称相同、主键不同，但候选键相同的实体；合并名称不同的实体。

通常，具有相同主键的实体代表相同的对象，应该合并。合并后的实体包括初始的实体 属性，删除重复的属性。有些实体具有相同的名称，相同的候选键，但主键不同，也应该合 并成一个实体，选择一个主键。对于名称不同的实体，主要根据实体的内容来决定是否要进 行合并。

5)包含(不合并)每个局部模型的唯一的实体。前面已经合并了相同的实体，剩余的 实体应该都是模型中唯一的实体，不用进行任何改动就可以直接加入全局模型。

6)合并局部模型的关系。检查所有关系的名称和目标，先合并名称相同且目标相同的 关系，再合并名称不同但目标相同的关系。

7)包含(不合并)每个局部模型的唯一的关系。对于剩下的关系，不需进行任何改动 就可以加入到全局模型中。

8)检查是否有丢失的实体和关系。需要注意是否存在不同局部模型中的实体之间的关 系以及它们之间的主、外键关系。

9)检查外键。在构造全局逻辑模型的过程中，可能合并了实体和关系，改变了主键， 或者标识了新的关系，需要在这一步检查子实体中的外键是否正确。

10)检查完整性约束。检查全局逻辑模型的完整性约束是否与每个局部模型的完整性 约束冲突，必要时，可以通过与用户讨论来解决问题。

最后可得全局逻辑模型的 ER 图。

步骤2:检查全局逻辑数据模型

用规范化的方法检查全局逻辑数据模型的实体，并检查这些实体是否能支持所有的用户 事务。

步骤3:与用户讨论全局逻辑数据模型

与用户进行讨论，确保全局逻辑数据模型确实能体现用户的需求。

 **银行数据治理**

**4.5** **非结构化数据模型**

大多数的信息是非结构化的，包括纸质的文件、报告、视频和音频文件、照片、传真 件、信件等。在信息化银行建设过程中，所处理的数据除了常见的结构化数据外，还有大量 的非结构化数据。

非结构化数据包含复杂的内容，并具有不同的结构特点，传统关系数据库无论从描述能 力上还是从管理数据的规模上，都无法应对非结构化数据管理的要求。因此，需要专门针对 非结构化数据特点，研究建立一种新的管理技术。

数据模型是数据管理的核心。数据模型定义了数据描述结构、数据操作方法以及数据完 整性约束条件。基于数据模型，可以建立可扩展的数据存储模型，使数据能够以某种结构和 方式进行存储与读取；可以建立面向上层应用的灵活多样的数据操作模型，支持用户对数据 的高效访问。非结构化数据管理与传统结构化数据管理的目标是相同的，即支持数据的永久 保存、快速有效的数据读取操作，从而有效地支持上层应用和用户的数据处理逻辑。因此， 在非结构化数据管理中，建立有效的数据模型，仍然是达到数据管理目标的核心和关键。现 有的非结构化数据模型主要有关系模型、扩展关系模型、面向对象模型、E-R 模型以及分 层式数据模型等。这些模型存在扩展性与复杂性方面的问题，而在面向大数据时这些问题更 加突出，因此无法胜任大数据的管理。

非结构化数据例如文本、图形、图像、音频和视频等，从内容上没有统一的结构，数据 是以原生态形式保存的，因此计算机无法直接理解和处理。为了对不同类型的非结构化数据 进行处理，所采用的基本思路是对这些非结构化数据进行描述，基于描述性信息实现对非结 构化数据内容的管理和操作。对于非结构化数据的描述，目前有三类：基于关键字的语义描 述；基于底层特征的描述；基于概念的语义描述。因此， 一个非结构化数据可以由基本属

性、语义特征、底层特征以及原始数据四 个部分构成，而且四个部分的数据之间存 在各种联系。

基于上述对非结构化数据的分析，可 以使用学术界提出的四面体模型。四面体 模型由一个顶点，四个刻面和刻面之间的 交线组成，如图4-12所示。

**1.** **语义特征**

语义特征是以文字表达的非结构化数 据特有的语义属性。对于书籍、电影类， 语义特征一般指作者的创作意图、主题说 明等。该属性是非必备属性。

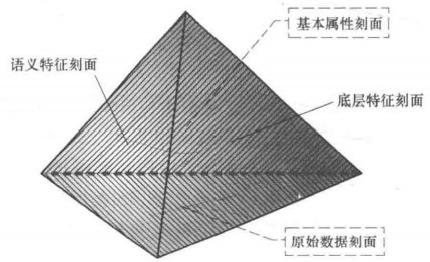


图4-12 非结构化数据四面体模型

**2.** **基本属性**

基本属性是非结构化数据都具有的一般属性，这些属性不涉及数据的语义。即基本属性 包括名称、类型、创建者、创建时间等。

**数据模型** 第4章

**3.** **底层特征**

底层特征是通过各种专用处理技术(如图像、语音、视频等处理技术)获得的非结构 化数据特性，如电话录音的格式、时长要求、比特率等相关技术属性。

**4.** **原始数据**

原始数据是非结构化数据的原生态文件。

四面体模型的特点在于语义特征与底层特征进行一体化表达的集成性；图像、文本、视 频、音频等多种异构数据表达的统一性；支持语义特征、底层特征动态变化的可扩展性和简 单性。基于四面体模型构建的非结构化数据管理模型，将能够实现异构数据进行统一存储与 关联操作，从而更好地支持大数据的深度处理。

**4.6** **数据模型主题**

数据模型主题又称为业务概念，是从业务角度对企业信息进行的分类，它具有通用性和 概括性的特点。针对金融业，业界也有多种主题模型分类方法，如 IBM的 FSDM 主题模型、 Teradata的 FS-LDM 主题模型。数据主题模型为复杂的数据库系统实施提供了规范和基础结 构，促进了业务部门用户和信息技术分析人员之间的有效沟通，形成对重要业务定义和术语 的统一认识。

业界较为通用的模型主题分类包括：当事人、产品(条件)、协议、账户、事件、渠 道、介质、地理位置、资源项。以上分类适用于可以用该类主题表达和涵盖的业务概念类的 实体。

**1.** **当事人**

当事人代表与银行有联络或与银行有利害关系，以及银行希望保留其信息的所有相关参 与者，其中也包括银行本身。当事人有四个重要的类型，包括个人、外部机构、内部机构、 岗位。

除了类型外，在企业概念数据模型中，当事人用角色来描述其所实现的功能，角色包括 客户、员工、合作方等。

**2.** **地理位置**

地理位置是银行希望保存的地理位置信息，代表可找到某人某物之处、信息所在位置或 界定的区域。

**3.** **产品**

产品是银行及其关联的当事人提供给市场，能够满足客户(包括内部客户)的某种需 求，银行可从中赚取各种实际或潜在收益的货物(有形)与服务(无形)。产品按结构类型 划分，可分为附加产品、单一产品和组合产品三类。

**4.** **介质**

介质是当事人获取访问权限所必须提供的访问标识 (Access ID) 及相关的认证信息。 银行通过对介质的鉴定确定来者是谁，能替他提供什么访问权限。介质的载体称为介质资源 项，如存折、磁卡、IC 卡、条码卡等。有载体的介质称为实物介质，无载体的介质称为数 字化介质。介质的鉴定是通过渠道来完成的。 一个介质可以通过多个渠道被鉴定。认证方式

 **银行数据治理**

不同，鉴定时使用到的认证信息也不同，同时对应的权限也不同。

**5.** **协议**

产品协议是识别两个当事人之间的协议。该协议明确规定其中一方从另一方取得产品使 用权。产品协议是产品的实例。 一个产品可对应零到多个产品协议； 一个产品协议只可对应 一个产品。根据产品协议和账户的关系，产品协议可分为以下三类：

**(1)账户协议** 需要开通一个客户账户记录交易情况的产品协议，如存款协议、贷款 协议、信用卡协议(对应单一产品)。

**(2)账户功能协议** 依赖于已存在的账户协议才能发挥作用的协议，是账户协议的部 分特征条件。因为其在结构上可重复使用于多个账户协议而被独立出来，比如定活通产品对 应的产品协议(对应附加产品)。

**(3)非账户协议** 不需要开通一个客户账户记录交易情况的产品协议，如旅行支票、 银行汇票(对应无账户的单一产品以及组合产品)。资源协议是指识别当事人间使用、销 售、交换资源项的协议。抵质押协议是指当事人抵质押资源项目当作无法清偿债务时的保证 的协议。当事人协议是指识别当事人以何种方式组成或互动的协议。保证协议是指保证人和 债权人达成的明确相互权利义务，当债务人不履行债务时，由保证人承担代为履行或连带责 任的协议。雇佣协议是指个人和组织间的一种协议，组织同意雇佣个人。合作方协议是指一 方当事人承诺以某种方式与另一方当事人一起合作或一起从事某项工作的协议，如商户合作 协议。

**6.** **账户**

账户是记录和监控货币及非货币数量变化的储存分类，协助银行达成各种核算需求。账 户既可以是独立运作的账户，也可以是汇总账户；既可以是客户账户，也可以是内部户；既 可以是货币账户，也可以是非货币账户。例如：存款账户、基金账户、黄金账户、总账账 户、公积金账户和第三方存款账户。

**7.** **事件**

事件是银行为业务目标的实现或业务的执行而希望保留的将发生或已发生的事情。 事件主要有以下类别：

**(1)业务活动** 在履行企业章程过程中由银行执行或为其执行行动相关的事件。

**(2)沟通事件** 在客户或潜在客户与银行之间以交换信息为目的的事件，如客户咨询、 客户投诉、主动营销、客户申请等。

(3)银行关注的外部事件 包括：动乱事件、生态事件、经济事件、司法事件、政治 事件、社会事件。

**8.** **资源项**

资源项是银行拥有、管理、使用的，或银行关心的其他当事人拥有的，有形或无形的有 价值的东西。资源项不包含其他8个数据概念的内容，如客户在某银行的资产作为协议进行 管理，不列入资源项范围。

**9.** **渠道**

渠道是银行希望保留其信息，银行或第三方机构拥有的，直接或间接进行产品服务、客 户沟通、市场营销、产品销售的途径。



**第5章**

**元数据管理**

近年来，大数据技术已经开始驶入快速发展的轨道，国内外从事大数据技术研究和应用 的企业越来越多，应用领域也越来越广，应用行业也越来越细。但是，在行业应用中，由于 缺乏统一的数据描述，使得数据在识别、传输、转换、共享等过程中常常出现歧义，数据的 拥有者不能及时准确地按照业务要求提供数据的情况时有发生。在大数据时代，借助于元数 据了解数据的业务含义、上下文等信息的需求越来越强烈。据了解，ISO/IEC 、JTC1 、SC32 下设的多个工作组从事不同领域的标准研制工作，其中就包括元数据。对于我国大数据时代 元数据标准的制定，有国内专家指出，我国迫切需要加强元数据标准或元数据模型的研究和 应用，健全与完善元数据标准规范和元数据模型，使大数据向着标准化、条理化、脉络化方 向发展，实现无歧义沟通、理解和共享使用，更好地对数据进行管控，挖掘大数据，发挥数 据价值。

在传统金融体系外，大数据技术及应用风生水起，而作为拥有大量客户信息及客户交易 数据的银行却越发显示出在数据标准体系建设、信息共享、信息挖掘和利用等方面的短板。 挖掘银行数据价值、发挥元数据作为银行数据信息地图的重要作用，将分散在不同系统、不 同工具、不同人员中的元数据信息进行统一管理，实现从业务层面到技术层面的全面贯通， 对银行业务的发展具有重大意义，是银行数据规范和高效管理的重要支撑。

元数据，英文名是 Metadata, 意指关于数据的数据，即对数据的描述。早在20世纪末， 元数据管理体系的概念和相关工具就已经广泛出现，但由于当时的数据规模还不够大，而且 元数据管理本身又包含了太多的内容，以至于它一直未得到充分的重视。而在今天看来，元 数据正在成为解决诸多数据问题时必须要抓住的一个“精髓”要素。

元数据描述了数据的结构、内容等多项内容，提供了对数据对象的描述、定位、管理、 检索、评估、选择和交互等功能，是数据治理的重要基础。元数据是数据对象的信息地图， 通过元数据管理，能够准确勾勒出银行数据资产的整体视图，支持科学地制定信息数据管理 政策。通过元数据管理，也能够建立统一的数据表达形式、元数据标准，使数据可视化，方 便数据的灵活交互和扩展。

当前，我国银行业在元数据管理及体系建设方面起点不一，发展程度也不一样。根据 IBM公司对元数据管理6个阶段的划分，结合我国银行元数据管理发展情况来看，目前我国 部分元数据管理较为成熟的银行已发展到元数据集中管理和元模型驱动阶段。这6个阶段主

 **银行数据治理**

要包括：

**1.** **初始阶段**

元数据分散在日常的业务和管理职能中， 一般由个人或某部分人进行管理，供局部使 用，跨范围使用通常需要沟通以获取元数据信息，面临较多的沟通工作。

**2.** **从属于业务系统阶段**

随着各业务系统自动化构建的完成，相应的元数据也随着需求整理、设计、开发、实施 和维护等过程被各业务系统孤立地全部或局部管理起来，元数据处于一个个竖井的业务系统 中。元数据直接互联互通困难，调阅各种文档以及跨系统的桥梁常常是实现元数据互联互通 的主要手段。

**3.** **元数据集中存储**

元数据虽然在局部产生和获取，但是会进行集中存储。可支持企业层面获取和查找元数 据，但是由于元数据依然在局部产生和维护，因此缺乏统一的元数据标准，也缺乏有效的管 理方法和管理流程。

**4.** **元数据集中管理**

基于元数据的集中存储，增强了对元数据的集中管控，主要加强了元数据标准及元数据 间的交互管控，也增强了上下游各应用间的协作水平。

**5.** **元数据模型驱动**

通过构建元数据模型，基于业务元数据、技术元数据、操作元数据间的关联关系，建立 基于主题域等层次结构，增强元数据的可读性，使之有利于用户充分理解元数据对应的业务 含义。

**6.** **元数据管理自动化**

元数据管理自动化主要是遵循统一的元数据模型和元数据规范，建立逻辑层次元数据和 物理层次元数据的关联，实现逻辑层次元数据和物理层次元数据的自动更新，实现各个应用 系统间数据格式的映射和自动产生。

元数据管理体系的构建是一个复杂和长期的过程。银行在企业级元数据管理体系建设方 面主要面临以下几大难点：

**(1)构建有效的企业级元数据管理体系** 如何高效构建企业级元数据管理体系并使之 成功落地，并有效发挥元数据管理在银行数据治理中的价值，是银行开展元数据管理体系建 设面临的首要问题。结合业界应用的成功实践和经验，银行元数据管理体系建设需要以元数 据系统建设为核心，采取集中式的元数据管理模式，并辅以相应的元数据技术规范制度和指 引，制定合适的元数据采集、维护和管理流程，将元数据管理融入银行信息系统日常的需求 编写和管理、项目设计、开发编码、系统测试和生产运维等研发过程中，实现与应用系统开 发、测试以及版本制作等流程紧密结合，发挥元数据管理系统在银行应用系统全生命周期开 发平台中的重要支撑作用。本书后续章节将对此进行详细论述。

**(2)确保元数据质量** 元数据质量直接影响着元数据价值的有效发挥，是元数据管理 体系建设是否成功的重要标准之一。如何确保元数据质量，充分发挥元数据作为银行数据信 息地图的重要价值，是元数据管理的重要问题。元数据质量的管控，需要结合银行自身应用 数据特点和元数据管理方式，将元数据与银行应用系统研发等流程紧密结合，通过将元数据 纳入开发、测试以及版本制作流程等方式，结合一系列的辅助技术工具，综合利用版本硬控

**元数据管理** | **第5章**

制、信息采集和数据核对校验等方式，来确保元数据质量。元数据质量及治理是一个持续不 断改进和提升的长期过程。

**(3)充分挖掘元数据价值** 元数据的价值不仅仅是对元数据进行统一管理，提供一个 统一的台账视图，在数据标准统一、信息共享，以及提升银行快速应对变更、基于数据信息 分析决策等方面都具有重要价值。例如，通过各类元数据间关联关系的管理，构建数据资产 关联关系地图，从而可以迅速定位软件变更带来的影响；构建面向业务和技术等不同用户视 角的银行数据资产视图，可以有效提升银行基于数据决策的可靠性和权威性。

5**.1** **元数据管理概述**

**1.** **元数据**

人类大脑思维非常神奇，往往可以习惯地利用数据模型将数据和信息加以结构化，还可 以利用元数据进行释义并形成数据；同时，通过大脑思维的识别、转化、扩充、统计、分析 和挖掘等，可以获取衍生信息，并同样利用数据模型和元数据进行结构化和释义。

元数据最早于1969年由Jack E.Myers 提出。长期以来，在银行领域，元数据主要应用 在数据仓库。在数据仓库中，元数据是用以描述数据仓库内数据的结构和建立方法的数据。 对于数据仓库这类集成类数据系统来说，存储的表、视图等数据对象数以万计，再加上内部 复杂的数据抽取、加载和转换，联机分析处理 (On Line Analytical Processing,OLAP) 和数 据挖掘，以及众多的数据集市，很容易形成“蜘蛛网”现象。而解决“蜘蛛网”问题的关 键就在于元数据管理，它可以帮助最终用户理解和使用数据仓库中的数据，也是保证数据质 量的关键。同样，它也应用在图书馆日常管理中。图书馆中的书籍多如繁星，要想在浩瀚的 书海中快速准确定位到某种想要选择的具体书籍，书签必不可少，而书签就是书籍的元 数据。

根据银行数据特点，结合业界理论和实践经验，元数据一般可以分为业务元数据、技术 元数据、操作元数据三个类别，内容覆盖结构化数据、半结构化数据以及非结构化数据。

**(1)业务元数据** 业务元数据是定义和业务相关数据的信息，用于辅助定位、理解及 访问业务信息。业务元数据的范围主要包括：业务指标、业务规则、数据质量规则、专业术 语、数据标准、概念数据模型、实体/属性、逻辑数据模型等。

**(2)技术元数据** 它又可以分成结构性技术元数据和关联性技术元数据。结构性技术 元数据提供了在信息技术的基础架构中对数据的说明，如数据的存放位置、数据的存储类 型、数据的血缘关系等。关联性技术元数据描述了数据之间的关联和数据在信息技术环境之 中的流转情况。技术元数据的范围主要包括：技术规则(计算/统计/转换/汇总)、数据质 量规则技术描述、字段/列、衍生字段/列、事实/维度、统计指标、表/视图/文件/接口、报 表/多维分析、数据库/视图组/文件组/接口组、源程序、系统、软件、硬件等。技术元数据 一般以已有的业务元数据为参考进行设计。

**(3)操作元数据** 操作元数据主要是指与元数据管理相关的组织、岗位、职责、流程， 以及系统日常运行产生的操作数据。操作元数据管理的内容主要包括：与元数据管理相关的 组织、岗位、职责、流程、项目、版本，以及系统生产运行产生的操作记录，如运行记录、

 **银行数据治理**

应用程序、运行作业。

技术元数据、业务元数据、操作元数据三者关系紧密：业务元数据指导技术元数据；技 术元数据以已有的业务元数据为参考进行设计；操作元数据作为技术元数据和业务元数据管 理的要素之一，为两者的管理提供支撑。

**2.** **元数据管理及重要意义**

伴随着我国商业银行信息化建设的发展，银行形成了包括核心系统、数据仓库、风险管 理、客户关系管理等在内的多种业务和管理系统。特别是近年来，随着大数据技术的飞速发 展，银行数据量成倍增长，数据形式也越来越多样化。大数据给银行数据处理带来的大容 量、快速处理和多样性，使得银行在数据标准建设、信息共享、信息挖掘利用等方面面临挑 战，也越来越重视元数据的管理和体系的建设工作。

元数据管理是指元数据的定义、收集、管理和发布的方法、工具及流程的集合。它涵盖 元数据定义，元数据的管理原则、管理模式和方法，元数据相关制度、规范、手册，元数据 管理系统，元数据管理相关的日常处理流程等。元数据管理是一个以相关元数据规范、指引 为基础，以元数据管理系统作为技术支撑，与应用系统的开发、设计和版本制作流程紧密结 合的完整体系。

商业银行通过构建元数据管理系统，可以实现将不同系统、不同工具、不同人员中的元 数据信息进行统一集中管理，实现从业务层到技术层的全面技术贯通，为银行科技系统更高 效、规范地运作提供系统支撑，对银行业务发展具有重大意义。商业银行通过元数据进行管 理，可以在以下方面进行提升：

1)统一表达形式，建立统一标准，使数据更易读，更好地实现信息共享，最大程度地 发挥信息的价值作用，降低沟通成本，提升沟通效率，增强上下游各应用与分行间的协作 水平。

2)用户更清晰地理解数据含义及数据间的关联关系，迅速定位软件设计变更带来的影 响，及时对相关系统设计做出必要的调整，如数据定义、接口，提升快速应对变更的能力。

3)实现规范、标准落地，确保元数据设计/登记质量，可以更好地支撑数据分布、数 据交换、数据集成、数据生命周期管理、数据标准等数据治理相关的工作内容。

4)实现公共资源的统一分配和登记，从而确保有效管理，不遗漏、不冲突。

5)实现对元数据资产的统计、分析和挖掘，例如血缘分析、孤儿分析、影响性分析以 及各类统计功能等，提升基于数据所做决策的准确性和可信性。

**3.** **元数据管理体系简介**

商业银行元数据管理对于数据治理工作起着重要的作用，是数据标准和数据质量管理的 重要基础。

商业银行应该建立一个体系清晰、内容全面、标准明确的元数据管理体系，紧密围绕国 家相关标准和行业规范，充分结合商业银行自身的整体系统发展战略，全面考虑相关规范和 管理要求的可持续发展和良好的可操作性。

商业银行元数据管理应围绕“一个中心、 一套制度、 一套流程”建设符合银行自身特 色的元数据管理体系。

**(1)以建立企业级元数据管理系统为核心** “一个中心”就是指建立一个企业级的元数 据管理系统，并以此作为元数据管理核心。企业级元数据管理系统是银行元数据管理和应用

**元数据管理** **第5章**|

元数据设计的统一平台。通过元数据管理系统，可以为银行元数据提供统一的命名规则、业 务规则、数据标准和管控流程，有利于降低上下游沟通成本，实现信息的共享和集成，提升 银行部门间、总分行间的协作水平，可以快速应对应用系统和业务变更，提升银行迅速决 策、快速创新和业务流程改进的能力，从而提高银行市场竞争力。

银行元数据管理体系的建设宜以元数据管理系统元数据为主，以各应用系统元数据为 辅，把分散的元数据信息进行统一的存储并采用集中式的管理模式。元数据管理系统应至少 涵盖银行当前主要的元数据管理种类，明确各类元数据管理内容和管理方法，制定相匹配的 元数据变更、通知、审批等日常管控流程；建立职责清晰的应用系统角色，做好元数据管理 系统的权限管理；灵活采用信息登记、扫描等元数据获取方式，充分考虑系统易用性；充分 参与应用系统的设计、开发、测试、版本制作等流程，将元数据采集、管理和应用纳入银行 应用系统日常研发过程中。

**(2)以制定企业级元数据技术规范和指引为保障** “ 一套制度”就是指建立企业级元数 据技术、实施等相关规范和指引，为元数据管理系统的建设和管理保驾护航。元数据技术、 实施规范和指引是银行进行元数据管理的依据和约束，是全行各部门、各分行在应用系统研 发过程中应遵循的元数据管理指针。相关的元数据技术规范和指引应以银行元数据模型为指 导，为企业级元数据管理系统的建设和管理提供重要的制度支撑；需要明确元数据具体含义 和分类，从全行层面明确银行元数据的管理模式和管理方法；对于主要元数据管理对象，还 需要明确具体的管理内容和技术要求，以及涉及相关应遵循的规范等内容。

**(3)与银行日常应用系统的研发、测试、版本制作等流程紧密结合** “ 一套流程”就是 指针对元数据信息的采集、维护和日常管理等制定一套符合银行自身特色，与银行应用系统 日常研发过程紧密结合的元数据管理流程。元数据管理流程需要与应用系统版本制作流程相 结合，将元数据管理作为应用系统研发过程中必不可少的一个环节。在系统上落地相关的制 度规范对元数据信息的约束控制，努力实现以系统版本硬性控制为主，以信息收集方式补充 元数据、定期核对的管理手段为辅的元数据管理方法，综合利用各种技术手段，确保元数据 登记质量。

**5.2** **搭建元数据管理体系**

**1.建立企业级元数据管理系统**

对于银行来说，建立元数据管理系统是其元数据管理体系建设的核心。企业级元数据管 理系统建设需要在充分考虑自身实际情况的基础上，实现以下6个建设目标：

**(1)企业级管理** 主要是指制定元数据管理相关规范和制度时，应基本上涵盖银行所 有应用系统使用的各个方面的元数据，确保规范完整性强、覆盖面广，实现企业级元数据管 理，结束原来面向个别应用系统的元数据管理状况。

**(2)版本化管理** 在元数据管理模式上，做到元数据管理与各版本的应用开发设计相 结合，与版本制作流程相结合，通过硬性控制手段，保证数据质量，提高数据可靠性，为银 行各业务系统的方案设计、程序开发，版本的系统测试、制作、投产组织等项工作，提供权 威的数据支持，成为银行科技工作各项环节中必不可少的资源库，保证银行应用版本有序推

**银行数据治理**

出，为业务发展提供服务。

**(3)标准化管理** 商业银行所有应用系统的元数据， 一般要求按照统一的规范和数据 标准要求进行集中管理，各应用系统使用的数据信息均会以企业级元数据管理系统中提供的 信息为准；数据信息的集中管理，确保银行各应用系统使用的数据信息准确、 一致。

**(4)系统化管理** 企业级元数据管理系统需要为全行的数据架构管理提供支撑平台， 在元数据资源分配上，进行统一分配，保证了元数据资源的有效与合理利用，避免重复、冲 突导致的各类生产事故；同时，为各类标准规范(如应用设计开发规范、数据生命周期管 理规范等)管理的落地，提供系统化支撑手段。

**(5)自动化管理** 在数据集中管理的前提下，根据各项制度要求，各类数据信息的变 化都应通过系统加以体现和更新；下游使用上游的数据，要在系统中予以登记；上游数据变 化后，系统自动根据登记的使用情况进行通知，以保证银行各系统间上、下游沟通顺畅，提 升沟通配合效率，使各项开发工作有条不紊地按照项目要求开展，这样就会对银行应用系统 建设具有积极的推动作用。

**(6)易用性** 系统应以人为核心，即“以用户为中心”是系统建设的主要目标之一， 提升系统易用性，对提高工作效率，确保元数据质量，提高用户体验性等有着重要意义。例 如：针对不同用户进行相关信息的展现；增加记忆功能，如我的收藏、我的常用、我的最近 浏览、我的关注等；增加搜索引擎，实现对系统所有信息的全文搜索；为用户提供方便、快 捷的批量导入导出功能或工具，增强用户的体验性，便于用户操作。

**(7)信息价值利用** 通过元数据管理系统，对集中管理的数据对象建立内在的关联关 系，在此基础上实现银行各系统间数据流向的分析，实现元数据的影响分析和血缘分析，清 晰展现银行各应用系统的全局框架，为系统架构、应用设计、开发工作提供有效的数据 支撑。

企业级元数据管理系统作为银行全行级的元数据管理平台，至少需要包含以下主要功 能：元数据日常管理和维护、版本管理、权限管理、提交与审批管理等功能，以及孤儿分 析、影响性分析、关联度(活力)分析等信息统计分析功能。

**(1)元数据日常管理和维护** 支持开发人员通过页面联机、批量方式对技术元数据、 业务元数据和操作元数据进行新增、修改、删除，提供各类元数据间的调用关系管理、变更 通知以及配套的审批管控流程。

**(2)版本管理** 支持元数据维护时录入版本、项目等版本制作相关信息，通过制定相 关的应用系统开发流程规范，将元数据维护和变更管理纳入应用系统开发流程，提供元数据 基线管理、版本制作、版本提交等功能，并实现与相关版本发布系统或平台的有效对接。以 版本发布为基准，做好元数据的基线管理。

**(3)权限管理** 信息安全是元数据管理系统的重要环节。元数据管理系统采用分类授 权、分角色授权等原则，对访问元数据管理系统进行权限控制。设置用户角色与权限，建立 与应用系统、各类元数据的关联关系；针对用户角色权限的变更，提供配套的角色与权限申 请审批管理流程，加强事中的审核控制、事后的监督核查。

**(4)提交与审批管理** 元数据管理系统在管理各类元数据资产时，需要根据不同元数 据的特性，并考虑实际应用场景和管理的需求，设定相应的审批控制流程，从而加强对元数 据的管理，保证元数据平台的准确性、时效性和稳定性。对于涉及上下游应用调用的接口、

**元数据管理** | **第5章**

文件等元数据，应将下游应用纳入审批流程，确保可以及时评估和应对上游元数据变更带来 的影响；对于重要的元数据对象，设置特殊的高层级审批流程，确保系统核心元数据的稳定 性。另外，在审批流程设计过程中，需要兼顾审批流程的时效性设计，确保审批时效不影响 银行应用系统正常的研发等活动。

**(5)信息统计分析** 主要提供元数据管理系统质量评估、各类资产多维度统计和报表 功能、资产变更统计及排名分析等。通过元数据管理系统提供的信息统计分析功能，能够及 时了解元数据管理系统中元数据资产整体和变更情况，为开展孤儿分析、影响性分析、关联 度(活力)分析、元数据治理等工作提供依据。

**2.** **建立企业级元数据技术规范和实施指引**

商业银行元数据管理体系，需要以各类元数据技术规范和实施指引为指导，规范元数据 管理。由于元数据管理与数据质量管理、数据生命周期管理、数据分布、模型管理以及数据 交换等领域关系密切，商业银行需要基于上述领域制定元数据管理的规范，以明确元数据管 理的技术规则、技术要求、管理方法、实施原则、实施方法和手段，指导全行元数据日常管 理及元数据系统建设。

其中，元数据管理的相关规范和指引作为银行元数据管理的指导和依据，需要明确以下 内容：

**(1)元数据定义、类别和管理范畴等内容** 元数据定义、类别和管理范畴的划分一方 面需要充分借鉴业界的有益实践，同时也要充分考虑银行自身的实际情况，以银行自身的元 数据模型为指导，明确元数据的定义、具体类别、每一类别的主要管理对象及管理内容。元 数据定义不清、类别划分不合理、管理范畴不明确将制约企业规范化、标准化元数据管理体 系的建设，在日常的元数据管理中也容易出现管理不到位、出现问题后互相扯皮等情况。

**(2)元数据管理模式和管理方法** 主要明确内容至少应包括：银行元数据管理采用的 集中式管理模式要求，元数据的收集方式，以及元数据管理系统在元数据管理体系的定 位等。

当前，元数据管理模式主要分为集中式和分散式两种。对于分散式元数据管理模式，元 数据分散在各个应用系统中分别进行管理。由于元数据分散部署的模式不利于跨系统间元数 据的互联互通，因此结合我国银行业实际情况，对于目前商业银行来说，建立企业级元数据 管理系统进行集中式管理，是符合当前我国银行元数据管理的实际情况的。集中式元数据管 理模式在元数据的统一管理、统一标准、关联关系分析等方面具有较多的优势。

**(3)对主要的元数据管理对象的组件属性信息、技术管理要求** 主要内容包括：主要 元数据对象组件属性信息基本组成、是否必须控制、必要性说明、约束要求等；还包括主要 元数据对象的类别和管理内容、收集要求、应遵循的相关规范等技术要求等。

**3.** **制定合适的元数据管理流程**

企业级元数据管理系统的集中式元数据管理模式需要将企业所有的元数据统一在元数据 管理系统中进行管理。进行系统设计时需要充分重视元数据质量，不能脱离日常的应用系统 研发管理流程， 一旦脱离，久而久之，其管理的元数据就会变成食之无味、弃之可惜的 “鸡肋”。同时，需要利用一系列工具和技术手段，确保元数据的数据质量。关于元数据的 质量问题，对经历过元数据管理体系建设的银行来说是有着切肤之痛的。以某商业银行平台 表结构技术元数据管理为例，该行早在2008年以前就将平台表结构统一纳入了元数据管理

 **银行数据治理**

系统进行管理，然而由于历史和流程改造成本等原因，始终未下定决心将平台表的日常维护 和管理纳入企业应用系统的研发流程中，并作为必不可少的一个环节，导致其管理的元数据 质量一直受到使用方的诟病，从而出现上游不及时登记，下游不敢使用的现象，严重影响了 其价值的发挥。

银行建设元数据管理体系时，需要结合自身数据和数据管理的特点，制定符合本行特色 的元数据管理流程。在银行应用系统的全生命研发周期管理模型中，银行应用系统日常的研 发流程涵盖需求编写、需求管理、项目设计、开发编码、系统测试、版本发布和生产运行。 将元数据管理与应用系统的日常研发流程紧密结合，目的就是将元数据的收集、维护、应用 等日常流程融入银行应用系统的全生命研发周期管理模型中。在需求编写、需求管理，以及 项目设计、开发、测试和生产运维阶段以元数据管理系统为准，既可以避免歧义又能辅助确 保元数据质量；同时，在项目研发过程中产生新的元数据，应及时按照流程将其纳入元数据 管理平台。

与研发流程紧密结合，其核心是采用以系统硬性控制的管理手段为主，以信息收集方式 补充元数据、定期核对的管理手段为辅的管理方法。硬性控制手段主要是指将元数据的应用 纳入应用系统日常研发关键节点，以元数据管理平台中的元数据为准，作为主数据源，作为 系统开发、测试和版本制作的依据，以及上下游各应用间沟通的依据。特别是在测试或版本 制作时，将元数据管理平台管理的元数据嵌入测试或版本制作流程中，通过系统硬性控制的 方式，来确保主要元数据对象的质量。当然，不是所有的元数据都可以适用硬性控制的方 式。对于无法与版本制作流程相结合的元数据，以信息收集的方式补充至元数据管理系统， 并定期进行核对，以保障元数据的质量。在此过程中，注意辅以各种技术支撑手段或工具， 以减少收集补充工作量。

元数据管理系统日常管理维护流程还需要制定元数据日常维护审批流程、元数据稳定性 管控流程、变更时效性的管控流程等，将研发过程中各阶段相关人员纳入流程(如将下游 应用相关人员纳入上游接口审批会签流程),以利于上下游联动，提升快速应对变更的 能力。

**5.3** **主要元数据对象的管理**

这里主要列举几种常见的元数据，并在集中管理模式下，就元数据管理内容、方法、流 程等内容进行介绍。

**1.** **表结构**

**(1)管理信息** 主要管理内容包括：表基本信息、字段信息、约束/索引信息、管理信 息(维护人、维护时间等)、审批信息(审批人等)。

**(2)管理方法** 表结构属于技术元数据范畴，商业银行所有的表结构都需要在元数据 管理系统中进行管理，并与生产保持一致。表结构管理应该和应用系统的开发、测试、版本 制作等研发流程紧密结合起来，包括由元数据管理系统的表结构信息来驱动开发环境、测试 环境中数据库表的建立。通过这种方式可以确保元数据管理系统表结构信息的准确性和及 时性。

**元数据管理** **第5章**

表结构的创建会有一定的技术要求，主要包括：

1)表结构在创建时参考和引用了逻辑数据模型，在达到一定程度的基础上，甚至可以 做到直接通过逻辑模型设计表结构。

2)表名、字段名、表空间、索引信息、约束信息、分区信息符合各类型数据库的开发 要求，如数据模型设计、数据库应用开发、数据生命周期管理、数据分布等。

3)已建立数据标准的表字段设计会以数据标准为指导，表结构中字段的业务含义、数 据类型、长度、精度与相应的数据标准保持一致，数据字典可以选择标准的全集或子集。

4)不适合建立数据标准或者暂时未定义标准的字段， 一般按照既有设计模式来设计， 但在同一系统内部要在设计上保持一致，其数据字典也应该在字段里加以保存。

**(3)管控流程** 根据不同的应用场景，元数据管理系统需要建立相应的控制审批流程， 通过流程控制和审核监督，保障核心应用系统数据结构的稳定性，加强变更的时效性管理。 同时，通过将元数据管理系统纳入版本发布流程管理，以元数据管理系统登记的表结构信息 驱动开发环境、测试环境中数据库表的建立，保障元数据系统的数据质量。简单列举部分管 控流程如下：

1)正常审批控制流程。在应用系统开发阶段，应用系统开发人员会将表结构相关信息 在元数据管理系统中加以登记和体现，通过审批方式确保其变更合理性和正确性。审核通过 后，可以通过元数据管理系统中的表结构信息及动态生成的数据定义语言 (DDL) 脚本来 驱动开发环境、测试环境中数据库表的建立。

2)特殊审批控制流程。为了维护核心系统数据结构的稳定性、及时性，商业银行一般 会对表的频繁变化、应用变化比例比较大，属于重要应用系统核心的表进行严格控制，设定 特殊的审批流程。例如，某商业银行对这种变更和审批进行重点跟踪，对变更提交的合理性 进行评估和审批。

**2.** **文件**

**(1)管理信息** 文件基本信息、文件结构信息、文件传输信息，以及文件的提供方、 文件的使用方等管理信息。

**(2)管理方法** 文件及文件传输信息属于技术元数据范畴。在集中式元数据管理模式 下，银行应用的文件及文件传输信息可以全部在元数据管理系统中登记管理，并与生产保持 一致。文件及文件传输信息在开发阶段就应该明确并体现到元数据管理系统中。使用方在需 要使用文件时，需要使用部门在元数据管理系统中明确使用情况。当提供方出现文件结构变 化、逻辑变化、数据字典变化时，需要元数据管理系统通知接收方保持同步变化。生产环境 中的数据传输交换系统也会依据元数据管理系统的内容进行相关调整，确保元数据管理系统 中的文件及文件传输信息的准确性和及时性。

文件及文件传输信息应遵循以下要求：

1)文件及文件传输信息，包括文件命名、字段信息登记、文件使用信息等，要有一定 的规范标准，遵循我国或行业的相关设计和传输规范，结合银行数据传输的实际情况，制定 符合银行自身特色的传输文件及传输信息规范。

2)文件结构字段的设计以数据标准为指导，字段的业务含义、数据类型、长度、小数 需要与相应的数据标准保持一致，取值字典应是数据标准技术定义中数据字典的全集或 子集。

 银行数据治理

**(3)管控流程** 一般来说，文件相关审批流程主要涉及文件登记时效性、文件使用等 方面，确保文件信息能够及时记录在案，从而能够让接收方及时获知相应信息。

**3.** **接口**

接口是指相互联系的应用系统之间，为完成特定功能而进行的相互之间传送数据的约 定，这些约定包括传送数据的方向、数量、每个字段的类型和长度等内容。数据的传输是依 据两者之间特定的通信协议完成的。

**(1)管理信息** 主要是对接口程序的管理。管理信息包括接口的基本信息、接口说明 的文档、接口的使用信息。

**(2)管理方法** 接口程序属于结构性技术元数据范畴，应用接口及接口调用关系均在 元数据管理系统统一登记管理。所有接口新增或者某应用在调用一个接口(或不再调用一 个接口)时，都要求于应用开发阶段在元数据管理系统中登记。接口的修改、作废等，由 元数据管理系统自动通知使用接口的联系人，从而对接口的使用方进行相应的改造。

接口管理要有一定的技术规范，对接口编码、交易代码、接口调用关系等在命名、规则 等方面进行约束。

1)元数据管理系统依据接口规范的要求，可以制定相匹配的规则，支持自动分配接口 编码、公共代码等功能。

2)接口相关的信息定义要有一定的标准。特别是系统间通信信息中字段的业务含义、 数据类型、长度应该与相应的数据标准保持一致。其数据字典应来源于标准中字典的全集或 子集。

**(3)管控流程** 接口管控主要是对接口变更管理和接口调用的审批管理，设置相应的 流程和岗位，并提供接口涉及的系统间通知等功能。

**4.** **公共代码**

公共代码是指银行为各个应用系统较为常用的、需统一编码的特定功能而封装的统一模 块，可减少重复开发，实现资产复用， 一般使用一定标准的字符串或者数字串，例如错误 代码。

**(1)管理信息** 公共代码基本信息(包括中英文名称、公共代码含义、服务范畴、业 务种类等)、启用或废弃时间、维护信息、使用信息等。

**(2)管理方法** 公共代码是需要通过字典描述，并且随各应用系统的变化不断进行扩 充和调整，各相关应用系统需要及时了解最新情况的数据资源。依据公共代码的业务管理范 畴，可以将公共代码资源的管理方式分为统一分配或者统一登记两种。

对于统一分配的管理方式，元数据管理系统制定相匹配的分配、命名等管理规则；应用 系统应在系统开发阶段，在元数据管理系统上进行申请登记，并需要以元数据管理系统分配 为准，实现硬性控制管理。

对于统一登记管理的管理方式，需要由相关应用系统部门根据公共代码情况在元数据管 理系统中进行登记。对于此类公共代码，需要通过比对方式保证其登记的质量，如定期通过 将生产系统使用的公共代码值导出与元数据管理系统登记进行比对，并根据差异情况进行组 织治理。

公共代码的种类众多，各类公共代码需要遵循相关的技术要求。对于重要的公共代码， 如交易代码、错误代码等，商业银行需要制定相应的管理规则和管理方法；同时，元数据管

**元数据管理** **第5章**

理系统制定相匹配的各类管理规则，实现对公共代码的规范、标准管理，保证公共代码的 质量。

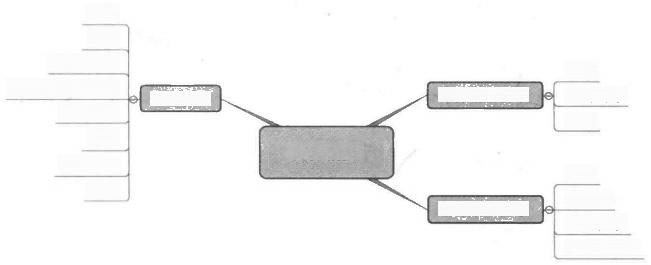
**(** **3** **)** **管** **控** **流** **程** 公共代码根据种类差异，存在不同的管控流程。对于重要的管理公共 代码，如错误代码、授权代码等，需要制定严格的审批管理流程；对于一般公共代码，仅需 进行相应通知即可。

**5.指标**

指标是指基于银行各应用系统产生的基础数据，按产品、机构、客户、渠道、交易、会 计科目等统计维度，根据一定的运算规则及统计频度进行统计的量化结果，用于银行内部管 理、监管报送等场景。根据指标的共享使用程度，可以分为共享指标和特色指标；根据指标 的加工数据源，分为基础指标和衍生指标。

**(1)管理信息** 指标的管理信息包含指标技术元数据及指标业务元数据。指标业务元 数据描述指标的业务定义，管理要素包括指标基础信息(包括指标名称、指标编码、指标 分类、业务口径等)、结构信息(包括维度、统计对象、度量单位等)及版本信息等。

图 5 - 1 所示为某商业银行指标业务元数据管理信息。



编码

名称

基础/衍生

自定义/开发生成

业务部门

口径

业务元数据

制定依据

标签

指标版本信息

维度

统计对象

属性

版本

维护人 项目信息

生效时间段

指标结构信息

指标信号

图5-1 某商业银行指标业务元数据管理信息

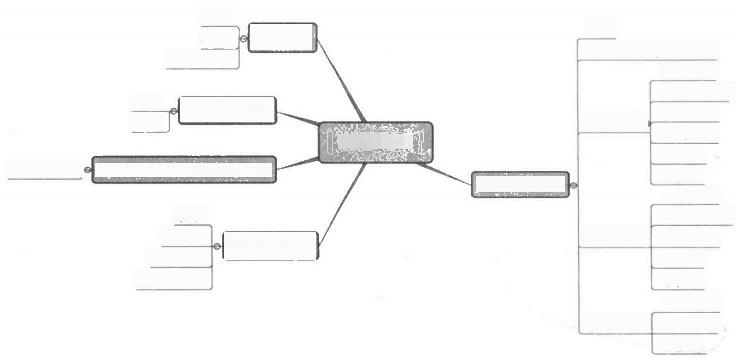
指标技术元数据从技术层面进行指标定义，管理要素包括指标基础信息(包括编码、 名 称 ) 、 指 标 结 构 信 息 ( 包 括 维 度 技 术 定 义 ) 、 与 业 务 指 标 对 照 关 系 、 指 标 加 工 信 息 ( 包 含 指标加工、存储、引用等信息)、指标版本信息等。

图5 - 2所示为某商业银行指标技术元数据管理信息。

**(** **2** **)** **管** **理** **方** **法** 指标元数据信息都应该在元数据管理系统中进行统 一 登记管理。为了 避 免 指 标 重 复 开 发 或 同 名 不 同 义 ( 是 指 指 标 名 称 相 同 ， 但 口 径 存 在 差 异 ) 指 标 的 产 生 ， 对 于新增的指标，需要在指标加工前就明确指标元数据信息。通过与系统中存量指标元数据比 对，进行相似度分析，形成“先有指标元数据，后有指标加工”的管理机制，严格控制同 名不同义指标的新增准入。 一 般来说，指标业务元数据由指标需求提出方明确，指标技术元 数据由指标加工处理的技术部门明确并进行登记。对于存量指标，则需要定期梳理回顾，建 立相应的指标退出机制，及时清退低效指标，从而形成指标“新增有控制、存量有管理” 的良性发展态势。

**(** **3** **)** **管** **控** **流** **程** 指标元数据管控主要是针对因指标变更而引起的指标业务元数据及技 术元数据的新增、修改、删除等的操作、审核、仲裁管理，设置相应的流程和岗位，并提供 指标涉及的系统间通知等功能。在此以新增指标准入管控流程为例。在指标统计需求提出阶

**银行数据治理**



编码

名称

基础/衍生

输入数据源

指标结构信息

技术元数据

程序名

业务元数据 技术与业务元数据对照关系

指标加工信息

系统名称 输入数据源

存储应用。[取数逻辑 输出表

维护人

系统名称

使用应用。展现功能 维护人

版本

维护人

项目信标版本信息 生效时间段

频度

加工、存储、使用路径 系统名称

加工规则 加工系统输出结果

维度

属性

指标信息

维护人

图5 - 2 某商业银行指标技术元数据管理信息

段需要准确录入指标元数据信息，并建立严格的审核流程，由相关审批人员负责对指标信息 完整性、指标的使用范围等进行审查，并结合系统分析与存量指标相似程度情况，判断指标 是否允许新增准入，以确保审批通过后，指标元数据信息的准确性，并为后续指标开发等工 作提供依据。

**5.4** **非结构化数据的元数据管理**

由于非结构化数据的类型和表现形式多样，其元数据管理办法很难统一，并且非结构化 数据的内容不确定、不易获取(如影音、视频),因此其元数据的提取和管理在业界还是一 个新领域。在此，仅就几类非结构化数据进行元数据管理方法的初步探讨。

1)对于影音、视频、图片等形式的非结构化数据，其内容一般较难获取，因此根据前 面介绍的四面体模型来管理这类非结构化数据的元数据。数据生成或获取时明确其语义特 征、基本属性、底层特征三方面属性，可以将其抽离出来作为结构化数据存储，那么就可以 按照结构化数据的元数据管理方法进行管理。

2)对于网页、文档等，不仅可以获取语义特征、基本属性、底层特征等属性作为元数 据，还可以从其中对其内容进行读取解析，抽取出关键字，作为非结构化数据的标签。主要 分为三种类别：

第一种是提取文章里的关键词，作为主题词标签元数据。

第二种是对文章进行多个维度的分类，打上分类标签，如在舆情分析过程中，会打上文 章所涉及的银行、产品及业务、风险点、正负面情感等。

第三种是文章本身的结构化标签元数据，如时间、作者、来源、大小等。

在实现过程中，只有通过分词、文本去重、各类分析模型(如主题分析模型、分类模 型、情感分析模型)处理，并辅以各种自动化训练手段，才能获取网页、文档等各类文本

的 非 结 构 化 数 据 的 标 签 元 数 据 。

3)对于具有加密方式的非结构化数据，要获取其内容， 一般需要比较高级的权限，因

**元数据管理** **第5章**

此在对其进行管理时只能获取基本的语义特征、基本属性，即可以知道文档的名称、时间 等，而对于其内容则无法获知， 一般和影音、视频、图片的相关元数据管理方法类似。

**5.5** **案例**

**1.中国工商银行元数据管理**

中国工商银行在经过多年的元数据管理信息系统建设后，最终形成了全行级的元数据管 理系统。其建设历程主要经历了三个主要阶段：

第一阶段：2004年启动“摸家底”工作，对软件资源进行全面梳理，建立数据管理信 息系统，对表结构、文件、接口、公共代码等进行系统化管理，于2008年完成系统建设和 数据的充实。

第二阶段：2008—2011年，基于元数据管理对象模型，对系统关键模块进行重构，对 功能进行优化，从而提升了整体的数据质量。

第三阶段：从2011年开始，进一步完善了系统功能，把全行数据标准、主机参数表、 软件资产、业务指标、主机资源等纳入元数据管理平台；同时，为了支撑业务全球化发展及 提升分行内部资产管理水平，提供多语言版本和分行版特色资源管理功能。

通过近10年的建设， 一是实现了软件资源和软件资产的统一管理，保证了各类数据资 源的准确性和数据展现的及时性，基于各种数据资源统计、分析和挖掘，提升了基于数据所 做决策的准确性和可信性；二是提供了软件系统元数据的设计平台，将业务规则、数据标准 进行了统一，增强了部门间、分行间的协作，较大地降低了沟通成本，提升了研发效率；三 是提升了快速应对系统变更的能力，通过迅速定位和分析软件设计变更带来的影响，从而及 时对相关系统设计(如数据定义、接口等)做出调整。

目前，该行元数据管理体系采用典型的集中式元数据管理模式，构建了以元数据管理系 统为核心的以相关的技术规范指引流程相配套的元数据管理体系。其元数据信息系统以硬性 控制的管理手段为主，以信息收集方式补充元数据的管理手段为辅，保证了资产的准确性、 完整性、时效性，通过各类元数据关联分析，提升元数据深层价值。由于历史存量因素，系 统管理的部分元数据存在一定的质量问题。中国工商银行元数据管理和数据治理团队相关人 员在对当前应用系统研发的全流程进行深入分析和研究后，实现了通过元数据管理系统动态 生成脚本，实现了增量的版本发布硬性控制，通过各种技术工具将生产环境和元数据管理系 统的元数据信息核对辅助进行存量提升，在元数据管理系统建立主要元数据对象的质量评价 体系，从准确性、完整性、时效性等多维度建立计分卡，形成全面的元数据质量评价体系， 有效提升了银行元数据管理系统中各主要元数据对象的质量，准确率基本达到100%。

**2.** **中国农业银行元数管理**

中国农业银行建设了元数据管理系统，搭建了业务元数据和技术元数据的集中管理和发 布平台。系统中涵盖了15000余张表，近30万个字段，超过4000个数据字典。



**第** **6** **章**

**数据标准**

标准对我们的生活有很大的影响，如度量的标准、安全的标准等。如果标准不对或没有 遵循标准，则容易引起混乱。例如，2007年国家颁布了手机充电器标准，并强制加以实施， 在降低消费成本、节约社会资源、减少电子废弃物方面成效显著。同样对于银行来说，数据 的标准化以及标准化过程中的治理显得尤为重要。

数据标准建设的难点在于以下几个方面：

**(1)历史问题比较严重** 我国商业银行基本上都是建设完多个系统后才发现系统之间 数据无法共享，比较简单的例子就是同样的信息存在不同的字典项。

**(2)标准涉及面比较广** 银行众多的业务系统存储了各种信息，需要参考国际、国内 等标准，同时会涉及金融、工业、农业等不同产业，并且一些标准也在不断地演化中，对银 行数据标准体系的建设都提出了较大的挑战。

**(3)不同部门的认知存在差异** 较为明显的是不同部门在制定数据标准或者进行数据 统计时，会用到不同统计口径和加工方法等指标。

因此，要做好银行的数据标准， 一是要从上至下认识到标准的重要性，制定从标准制 定、使用至废止的完整流程，严格按照流程执行；二是在系统建设时就要按照标准设计，避 免边治理边出现新的问题；三是要有配套的系统工具，以支持数据标准的有力落地。只有将 数据标准贯彻到位，才能在以下方面起到积极作用。

首先，统一的数据标准可以提升业务效率。数据标准统一了业务语言，明确了业务规则，规 范了业务处理过程，从而可以提升银行的整体业务效率，满足管理决策对信息及时性的要求。

其次，统一的数据标准促进了数据共享。数据标准统一了各类系统的数据定义，降低了 系统间集成的复杂度，提高了系统间交互效率，并为管理分析类系统提供了一致的分析指标 和分析维度定义。

再次，统一的数据标准提升了数据质量。数据标准明确了数据填写及数据处理的相关要求， 规范了数据源的格式，同时提供了管控方面的保障，因此数据标准将直接提升银行的数据质量。

**6.1** **银行数据标准的现状**

标准的历史极其悠久，早在公元前7000年，古埃及人就用石圆柱作为重量单位，那个

**数据标准**|**第6章**

时候标准就已具备雏形。已知的西方国家第一次标准化尝试发生在1120年，英格兰诺曼底 王朝国王亨利一世规定长度单位厄尔(旧时量布的长度)必须等于他的前臂的长度，并将 厄尔规定为他的王国里的标准长度单位。在我国，以前1斤等于16两，所以有半斤八两之 说，这个是我国以前的重量标准。由此可以看到，以前在交流不通的年代，各地各时都有一 定范围内的标准，对于统治者在自己辖地的管理起着关键的作用。

但随着科学技术的进步和交通运输业的不断发展，交流与沟通越来越顺畅，原先一地一 标准的情况已不再适用当今的形势。因此，现在需要制定统一标准，尤其是银行的数据标 准，在营销、分析、统计、挖掘方面可以通过统一的标准实现数据形式的一致性，从而共享 数据资源。

国外数据标准化工作起步较早，随着市场竞争的日益激烈及政府规范要求的日趋严格， 不少国际大型企业已将信息当作企业的重要资产来对待。其中大部分企业均以数据标准为核 心，建立相关的配套措施和协作机制，确保数据标准能够融入企业的每个业务环节中。同 时，业界也自发性地建立起相关的信息建设经验交流团体、信息管理框架及成熟度模型，以 持续完善企业自身的信息管理及应用能力。

目前，国内不少大型银行，如：中国工商银行、中国建设银行、光大银行、国家开发银 行、中国农业银行等均已经意识到数据标准对企业发展的重要性，先后聘请咨询公司，开展 了标准体系建设的探索，不同程度地开展了数据标准化建设相关项目。但银行一般会在标准 化方面存在以下问题：

**1.对数据标准化建设的认识有一定基础，但执行力度不够**

1)管理和执行力度不够。

2)对数据标准化建设内涵认识不足。

3)缺少针对数据标准知识的学习体系。

4)业务实质性参与数据标准化建设的积极性有待提高。

**2.** **现有管理制度流程不能满足信息大集中的数据标准管理要求**

1)以业务制度代替专业化的数据标准管理制度，存在大量局部标准。 2)借用其他工作流程来制定和维护数据标准。

3)缺少指导性的数据标准管理制度总则。

4)跨部门协调解决问题困难，周期长。

5)缺少针对数据标准的维护发布流程，产生问题数据。

6)分行有根据自己特色业务制定数据标准化工作的要求。

**3.** **现行数据标准管理体系不能满足信息大集中的管理要求**

1)现行数据标准管理组织结构、管理职能不完整。

2)没有常设性信息管理委员会或领导小组。

3)没有数据标准管理的牵头部门。

4)标准管理维护职能不明确，临时指定负责人员。

5)缺少管理数据标准应用情况的职能。

6)缺乏统一协调指挥的层次化组织体系。

7)无法形成数据标准统一视图。

8)数据标准的权威性和连续性不足。

 **银行数据治理**

**4.** **现有数据标准质量的认可程度较低**

1)数据标准内容冲突或不一致。

2)数据标准定义不准确，容易引起歧义。

3)数据标准定义要素不完整。

4)多机构树、多产品树影响经营管理分析。

**5.** **业务数据标准管理还停留在文档管理阶段**

1)数据标准的全行共享能力很弱。

2)没有数据标准变更痕迹。

3)使用困难，查找周期长。

正因为我国商业银行在标准认识和建设方面存在各种不足，才导致了因标准贯彻而出现 的各种各样的问题。举几个简单例子：

首先，国民经济的行业分类标准，针对2002年国家标准中谷物种植这一分类，在2011 年进行了进一步的细分，区分了稻谷、小麦等小类。但是，这些标准只是作为国家标准发 布，在很多商业银行中的实际应用中并未跟上国家标准的变化，主要原因是数据标准的管理 停留在文档阶段，没有相应的系统进行支持。这种情况不仅造成了标准的共享能力很弱，而 且导致了实际贯彻标准过程中无迹可寻。通过查阅文档来寻找标准，不仅对人员提出了较高 的要求，而且查找效率低下。

其次，早期商业银行在各项业务系统建设过程中基本上是各自为政的，只要满足本业务 系统的业务开展即可。但是，随着数据挖掘、分析等应有领域的兴起，各业务系统的数据标 准不一致对分析工作造成很大的困扰。例如：在个人存贷系统中用0表示男，用1表示女 而在信用卡系统中则正好相反。对于单个系统来说，不管是业务处理，还是统计分析，都不 会造成影响。但是，现在商业银行基本上以数据仓库作为基础进行数据挖掘、分析和决策， 那么,在进入数据仓库前必须了解各个业务系统相关数据到底表示什么含义，若存在不一致 的情况，则必须进行映射。由于银行尤其是大型商业银行业务系统众多，每个业务系统存在 大量的数据表，具有不同数据字典和编码规则的字段。 一般情况下，大型商业银行有将近 30万张表，即使以1%的比率对相关存储规则进行调查和映射，几千张表的工作量也是惊 人的。

再有，最近某商业银行的两个不同部门就第三方支付的总体业务量的统计数据产生了分 歧，最终发现一个部门对日志中标明为第三方支付渠道的交易进行了统计，而另一个部门则 认为只有具有第三方支付牌照的机构所涉及的交易才需要进行统计。由此可以看出，不同部 门对同一件事物的标准以及统计口径是存在差异的。

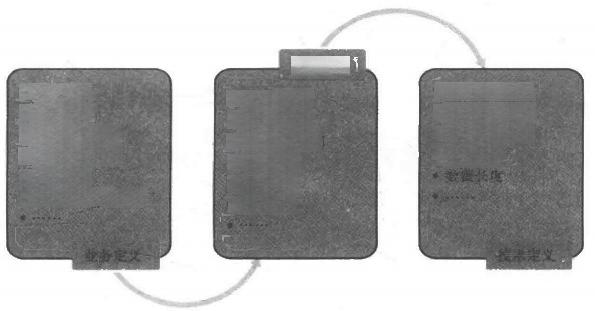
从上面的例子可以看出，数据标准是一个基础性的工作，需要各方面达成共识。

**6.2** **银行数据标准的定义与分类**

**1.** **数据标准的主体构成**

数据标准是对银行业务经营和管理所涉及的各项业务信息的统一定义与解释。它包含业 务定义、技术定义和管理信息三个方面的内容，如图6-1 所示。

**数据标准** **第** **6** **章**

管理信息

标准所有者

·数据产出系统 。致据来源系统 数极使用系统 收调类烈

·枢准各称

业务定义

扩展业务定义

亚务使用规则

\*标准来源

相关业青倍息编号

标准管理者

标准使用部门

标通状态

自动目期

失效日别

图6-1 数据标准的主体构成

**(** **1** **)** **业** **务** **定** **义** 业务定义主要是明确标准所属的业务主题以及标准的业务概念，包括 业务使用上的规则以及标准的相关来源等。对于代码类标准，还会进 一 步明确编码规则以及 相关的代码内容，以达到定义统 一 、口径统 一 、名称统 一 、参照统 一 以及来源统 一 的目的， 进而形成 一 套 一 致、规范、开放和共享的业务数据标准。

**(** **2** **)** **技** **术** **定** **义** 技术定义是指描述数据类型、数据格式、数据长度以及来源系统等技 术属性，从而能够对信息系统的建设和使用提供指导和约束。

**(** **3** **)** **管** **理** **信** **息** 管理信息是指明确标准的所有者、管理人员、使用部门等内容，从而 使数据标准的管理和维护工作有明确的责任主体，以保障数据标准能够持续地进行更新和 改进。

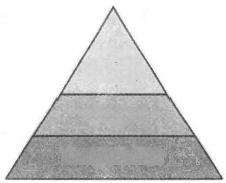
**2.** **数据标准体系的层次结构**

对于统 一 、共享的数据标准来说，灵活、可扩展是非常重要的特性。因此，数据标准体 系的建设就需要将基础性的业务元素和特定的业务规则进行分离，建立 一 套完整的基础业务 信息，并能够根据不同的业务规则对这些基础业务信息进行灵活的组合，从而满足业务运营 和管理的需要，进 一 步明确全行关键性指标。

根据以上方法，数据标准体系建设主要包含三方面内容：基础业务信息的标准化、业务

规则的标准化和关键绩效指标的标准化。

因此，数据标准体系的层次结构可用图

关键 关键绩效指标的标准化

6 - 2加以表示。 绩效指标

**(** **1** **)** **关** **键** **业** **务** **信** **息** **项** 关 键 业 务 信 息 项

业务指标 业务规则的标准化

包含的是银行运营和管理所产生和使用的具

有共同特征的基础性数据，如客户名称、贷 关键业务信息项 基础业务信息的标准化

款种类、交易机构等。

图6-2 数据标准体系的层次结构

**(** **2** **)** **业** **务** **指** **标** 业务指标包含的是为了

满足银行内部管理以及外部监管的需要，在基础性业务信息基础上按照 一 定的计算、统计规

则进行组合的业务信息，如风险贷款客户数、睡眠账户数等。

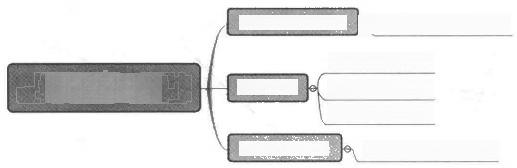
**(** **3** **)** **关** **键** **绩** **效** **指** **标** 关键绩效指标是在业务指标的基础上，根据银行的管理和考核要

求识别出来的关键性业务指标。

 **银行数据治理**

**3.** **数据标准的分类框架**

根据数据标准体系的层次结构，参照业界领先实践，数据标准的分类框架如图6-3 所示。



按九大信息领域分类

按业务部门分类

业务指标 按业务领域分类

按业务目的分类

关键绩效指标 参照平衡计分卡分类

信意械流分测辑架

关键业务信息项

图6-3 数据标准的分类框架

为了清晰地表述从各个不同的业务视角查看业务指标的需要，数据标准分类框架对于业 务指标综合了三种分类方式进行展示。其中，关键业务信息项唯一属于九大信息领域中的某 一个，关键绩效指标则属于业务指标的一部分。

**6.3** **银行数据标准建设**

数据标准化建设主要是处理好现状与理想状态的关系、当前成果与远期目标之间的关 系，设计出可以实施的数据标准化的短期、中期、长期实施方案和实现路径，逐步建立起系 统性的数据标准管理体系，确保数据标准化的工作可持续健康发展。

数据标准化建设规划的目标是确保经营管理信息的完整性、有效性、 一致性、规范性、 开放性以及共享性，从而能够畅通无阻地交互信息，进一步提升企业核心竞争能力，助力战 略发展目标的实现。

**1.** **建设原则**

**(1)数据标准化建设规划需要遵循一定的原则** 在进行数据标准化建设时，要重点考 虑代价、效益和风险三个主要因素，确定代价最小、效益最好、风险最低的方案。数据标准 化建设时考虑的六条评价基本原则如下：

1)规划在内部的统一性。即方案总体的各个部分必须与银行整体战略、规划相互配套 和衔接。

2)规划与环境的一致性。即必须保持对银行现有环境的适应性。

3)规划中存在的风险性。 一方面要认真考虑数据标准可能发生的偏差，另一方面要提 高实施方案预见能力，制定周密应变措施。

4)规划推进时间序列性。要对数据标准化行动的具体步骤做出稳妥安排，避免个体主 观随意性。

5)规划与资源的合理配套性。要掌握配套资源的合理使用。

6)规划的客观可行性。要立足于现实，具有可行性和可操作性。

**(2)数据标准化建设需要按照成熟度模型发展过程逐步推进** 数据标准化建设不是一 蹴而就的，需要遵循成熟度模型发展过程来逐步推进。目前，大部分商业银行数据标准化的 成熟度主要介于2级和3级之间。具体成熟度介绍详见3.6节。

**数据标准** **第6章**

**(3)短期规划的关键成功要素是初步建立组织制度、管理体系，形成标准内容框** **架** 数据标准化建设的难点是在不改变现有业务环境的条件下提升数据标准化程度。两者本 质上是存在相互博弈和制约的。建立集中的管理组织形成协调机制，设定管理体系标准化操 作流程，建立标准内容框架，将成为较小程度增加管理成本、较大程度获益的必然选择。

在短期规划中需要考虑如何快速获取业务价值，同时降低管理成本，并对业务操作现有 环境的影响较小。建立第三方集中管控机制，将大大降低内部沟通壁垒。

标准化管理组织需要形成完善的管理流程和角色分配，在短期规划中，首先要提升快速 解决业务难题的能力，降低业务操作复杂程度。因此，简明实用的管理流程将是标准化快速 见效的重要手段。

标准化操作管理的重要部分是标准内容的本身。短期规划中重点需要解决的是标准化的 需求在哪里，哪些重点信息需要进行标准化，这些标准化的基本内涵是什么。因此，在短期 规划中需要完成标准内容框架及标准内容包含要素。

**(4)中期规划的关键成功要素是形成业务标准与技术标准的对接机制** 数据标准化中 期规划重点需要考虑业务标准与技术标准相衔接的问题，特别是如何处理存量业务标准的问 题。例如，现有“客户地址”信息项有多种不同的技术标准，如何形成统一的标准，需要 循序渐进，初期需要协调好对接机制，将多套变成相对较少数量的标准，未来再通过系统改 造或者系统转换变成一套业务标准。

**(5)长期规划的关键成功要素是文化和变革** 数据标准化工作的长期规划重点是要考 虑到标准化推广和执行落地工作，其实就是“广而告之”。要让大多数人形成意识层面的统 一认识，需要经过不断的文化宣传和引导。企业标准化的最高境界是认知层面的统一，在这 个过程中需要经过不断的业务创新和变革，通过政策制定、标准化战略落地、标准化推广、 知识培训的手段，最终达到意识层面的统一。

**2.建设方法**

数据标准化建设规划是指导企业长期数据标准化工作不断提升和发展的行动纲领，是形 成未来企业核心竞争力的关键指标。因此，数据标准化建设规划制定过程的实质性内容，均 包括对下述问题的分析研究以及寻找答案。

1)目前数据标准化建设的态势如何?

2)希望未来的态势如何?

3)阻碍发展的可能因素是什么?

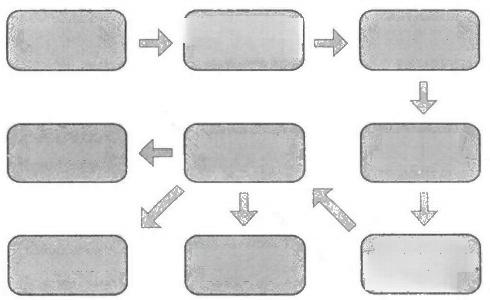
4)为实现总目标应采取什么行动?

对这些问题做出适当的回答，既需要分析范围广泛的环境信息，又需要对管理人员、技 术人员的知识和经验加以整理和利用。因此，必须进行战略性思考，运用整体战略规划的方 法，通过寻求整个企业的整体有效性来制定数据标准化建设规划。规划的制定是数据标准化 建设战略管理的起点和首要环节，是数据标准化管理工作和组织职能的一个十分重要的组成 部分。数据标准化建设规划描述了未来企业所能达到的数据标准化成熟度级别，并制定时间 表。因此，只有真正重视数据标准化建设规划的制定这种软投入，信息化工作才能获得最大 的回报。

以下介绍数据标准化建设规划的总体设计方法(图6-4):

第一步，参考通用的数据标准化管理功能组件。

**银行数据治理**



[2.设计符合本行的 管理框架和组件

6.成熟度评估 结果和分析

7.发现关键问题 及改进建议

5.成熟度评估： 能力分级、优先级、 整合性标准化

3.结合组件 设计调研问卷

4.调研问题 的收集与分析

1.参考业界 最佳实践

9.中长期规划

8.短期规划

图6-4 数据标准化建设规划的总体设计方法

第二步，结合通用的数据标准化管理功能组件设计银行特有的数据标准化管理框架及组 件活动 。

第三步，结合组件设计调研问卷。

第四步 ， 通过调研问卷反馈的相关信息进行问题的收集与分析 。

第五步，结合成熟度模型对业务优先级进行分析。

第六步 ， 参考成熟度模型评估 ， 设定未来短期 、 中期 、 长期规划目标 。

第七步 ， 寻找关键问题并提出改进建议 。

第八步，设计短期规划，提升业务价值。

第九步 ， 设计中期规划 ， 进行战略部署 ； 设计长期规划 ， 评估及持续改进 。

**(1)通用的数据标准化管理功能框架** 通用的数据标准化管理功能框架应该能够满足 银行数据标准化长期可持续性发展需要 。 一般来说 ， 通用的数据标准化管理功能框架的主要 内 容 包 含 6 个 部 分 ， 见 表 6 - 1 。

**表6-1** **通用的数据标准化管理功能框架的主要内容**

|  |  |
| --- | --- |
| 框架领域 | 描 述 |
| 文化和变革 | 企业文化是由一个组织的价值观、信念、仪式、符号、处事方式等组成的特有文化形象。企业文化的变 革目的是不断促进企业的变革和业绩的增长。企业文化的变革主要涉及政策流程的变革、管理的变革  以及学习方法的变革 |
| 组织结构 | 描述了在业务和技术领域组织和进行人员分配的方法。其中，定义了关键的流程、组织角色、职责以 及与其他外部组织的交互方式 |
| 治 理 | 治理就是通过统一的解决方案和治理模型来保护以及共享企业内不同层面的数据。通过数据治理能 够把企业的独立系统结合起来，并且重新定义数据的价值和保护机制 |
| 信息架构 | 信息架构包含单个或多套应用逻辑设计说明，以及业务模式、业务规则、数据模型、分层管理、组织模 式和目标模式 |
| 商务智能和 主数据架构 | 技术架构降低复杂程度、成本和方案实施风险；描述了主要的功能组件或者端到端的架构解决方案； 提供通用语言；提供框架或范围 |
| 商务智能和 主数据服务 | 开发、部署和运行商务智能工具所需要的支持服务 |

( 2 ) 结 合 通 用 的 数 据 标 准 化 管 理 功 能 组 件 设 计 银 行 特 有 的 数 据 标 准 化 管 理 框 架 及 组 件

**数据标准**| **第6章**

**活动** 以某商业银行为例，该行结合通用的信息管理框架进行了本土化，从而更加适合我国 国情，适合数据标准化管理的实际情况，更有利于解决现在所遇到的问题，更符合长远规 划、任务设计的要求。

该商业银行的业务管理模式及组织管控模式，不适用于部分商务智能相关工作，而且该 银行数据标准化建设更加关注基础信息、指标信息、关键绩效指标信息的定义标准化。因 此，去除了商务智能管理的相关内容，同时把工具的管理与应用结合为一个主要内容。该商 业银行数据标准化管理框架及组件活动见表6-2。

**表6-2** **某商业银行数据标准化管理框架及组件活动**

|  |  |
| --- | --- |
| 框架领域 | 描 述 |
| 文化和变革 | 数据标准化建设需要在全行范围内建立标准化管理、在沟通、共识的文化理念和制度。全行数据标准管 理组织共同遵循一致的管理制度和流程，在沟通层面建立权威、顺畅的渠道，同时通过培训将文化理念进 行传播、宣传和贯彻 |
| 数据标准 | 对业务经营和管理所涉及的各项业务信息进行统一定义与解释，包含基础业务数据标准、经营管理指标 数据标准以及关键绩效指标三方面内容。梳理和分类后的数据标准可以支持全行范围内的共享，提供灵  活、扩展的信息使用能力 |
| 制度和流程 | 数据标准化管理制度的目标是建立标准的完整生命周期管理框架，制定申请、审批和发布的制度办法， 规定标准管理组织的职能范围、汇报体系，建立人力资源调配、沟通合作的机制。管理流程基于管控组织 建立标准操作步骤，保证标准管理制度落实到具体流程中 |
| 组织和角色 | 建立数据标准管控运营模式，定义组织架构、职责范围、关键角色，定义组织和角色之间的管控机制，结 合数据标准管理现状，设计审核、监督、运营等工作岗位 |
| 管理平台 | 对全行数据标准进行内容管理。融入数据标准的完整管理流程，从信息的生命周期进行全程管理，提供 良好的用户体验。同时，将数据标准化管理的组织、角色职责、考核等关键因素在该管理平台中予以实现 |

**(3)调研问卷** 根据数据标准化管理框架设计相应的题库，覆盖数据标准化建设现状5 个方面所需的问题，即文化和变革、数据标准、制度和流程、组织和角色、管理平台。通过 这些问题理解标准化建设规划需要解决的问题并分析未来的主要任务。

**(4)问题收集与分析诊断** 根据调研问题的反馈，结合相关建设规划考虑的要求，设 计问题收集与分析诊断模板。

通过调研，可以分析目前遇到的主要问题，这些问题将成为建设规划的重要输入来源， 为后续规划环境的一致性、客观可行性提供参照。

一般来说，在标准化建设初期都会遇到相同的问题，包括对数据标准化建设的认识、管 理制度流程、组织结构、管理职能等都存在不足。

**(5)参考先进实践经验** 成功的国内外项目经验值得各大银行在数据标准化建设规划 中重点考虑，具体包含以下三方面的内容：

1)建立数据标准化管理机构与相应管理流程。多家银行均设立了数据标准化管理机 构，作为研究确定数据标准化建设与应用有关的重大决策唯一机构。同时设定了相关管理制 度和流程，作为数据标准化各项工作的操作指引。

2)建立详细的标准体系。首先将商业银行信息按照其产生方式、加工程度以及包含的 业务规则复杂程度，可以划分为基础信息、公用信息和专用信息三个层次。基础信息是通过 各种业务处理产生或各种渠道采集的基础性数据，在全行范围内必须是唯一定义的，包括客

 **银行数据治理**

零 售

银行卡

户信息、产品定义、账户定义、组织机构定义等。公共信息是在基础信息上按照一定业务加 工规则汇总，被多业务主体所共用的信息。专用信息是从部门视角和独立业务主体出发，经 过特定专用业务智能产生的数据。各个数据标准化项目的首要工作是将全球通行的金融业信 息分类与自身情况相结合进行定制化，形成各自的标准内容分类体系框架，明确哪些信息项 需要标准化，哪些信息项不需要标准化，为今后银行的数据标准制定设定方向。

大部分银行解决了全行范围的唯一识别代码体系建设的问题。数据标准化的一项重要工 作是编制全行数据标准化代码体系，逐步建立起全行统一的信息指标代码库，规范信息指标 代码的名称、唯一编号、取值范围、数据格式、长度等业务和技术属性。

短期内建立数据标准业务与技术的对接机制，长期形成一套有效的事前控制、事中监 控、事后考核的业务标准与技术标准的“硬性控制”。

3)建设标准化服务平台，形成企业级别统一的数据标准开发、评审、发布、失效的全 生命周期管理工具。标准化服务平台的定位是重在“服务”,辅在“管控”。因此，标准化 工具使用的范围越广泛，应用的业务条线越多，规范的系统越多，标准化工具的效果就 越好。

然而，通过对各大企业相关建设成果的调研发现，大部分企业还处于技术数据标准建设 的初始阶段，并没有成体系地从业务数据标准化的高度着手开展企业级的数据标准化工作。 未来数据标准不能停留在纸面，应从统一的业务标准转化为技术标准，强化对业务的软性服 务，不能因增加业务操作成本，降低工作效率，而形成为了控制而控制、为了管理而管理的 恶性循环。

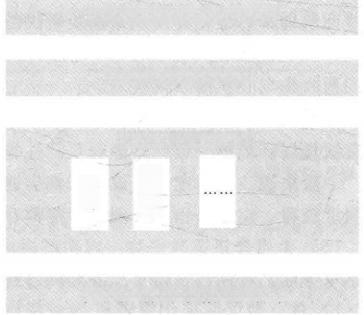
**3.** **建设内容**

**(1)建立数据标准管理组织** 建立数据标准管理专职机构，负责数据标准的统一管理， 承担数据标准的制定、维护、应用和监督等工作；建立数据标准管理团队，承担数据标准管 理与应用推动工作。以某商业银行为例，其建立的银行数据标准化组织机构如图6-5所示。

数据标准化建设工作领导小组是数据标准工 作的领导和决策组织，负责审定数据标准化工作 的发展方针和政策，审议数据标准管理相关重大 事项，审批重大数据标准。同时，需要建立各级 机构数据标准管理岗位，明确数据标准管理的各 组织职责。

**(2)建立数据标准管理流程** 数据标准管理 流程应至少覆盖以下6个方面，满足数据标准日 常管理需要：

1)数据标准建立流程。数据标准建立的基 本原则是遵循国际标准、国家标准和金融行业标 准(简称外部标准),同时还要考虑满足银行业 务需求。当外部标准不能完全满足行内业务需求

信息标准化建设工作领导小组

信息标准管理专职机构

信息标准管理专业线

信息标准管理团队(各业务部门)

图6-5 某商业银行数据标准化组织机构

时，在遵守相应外部标准框架的基础上，根据业务需求扩充形成新增数据标准。

数据标准的建立流程包括：需求收集、起草与申请、评审、形成送审稿、审批、发布。 数据标准需求部门向标准管理岗位提交标准建立申请，标准管理岗位对标准进行评审并

**数据标准** **第** **6** **章**

审批，包括：审查数据标准的属性完备性、数据标准制定相关文档的统一和完备性，形成最 终审批结果并进行发布。

2)数据标准修改流程。数据标准修改是指数据标准建立后根据业务需要的发展变化对 具体的数据标准项进行变更，并通过核准程序在全行实施。数据标准修改应平衡业务变化与 数据标准稳定之间的矛盾，需适时反映数据标准业务的含义和业务规则的变化，同时保持数 据标准的相对稳定，避免由于数据标准的频繁变动对业务应用和系统建设造成冲击。

数据标准修改的前提条件包括：

① 发现因相关业务变化，导致需要重新定义数据标准内容时，应及时提交数据标准修 改申请。

② 为解决数据标准冲突确需修改或者完善某个数据标准内容时，由该标准负责部门提 交数据标准修改申请。

3)数据标准停用流程。数据标准停用是指废止无法适应业务发展要求的标准，或者依 据数据标准冲突解决方案对具体的数据标准进行废止处理，并通过核准流程结束该标准的生 命周期，实现标准的退出机制。数据标准停用遵循谨慎对待的原则，对于满足条件的标准才 发起停用流程。

4)新建标准例外流程。数据标准新建例外流程为紧迫程度高的标准创建提供绿色通 道，只有获得标准管理专职机构批准的新建标准，才可以进入绿色通道。在绿色通道新建流 程中，赋予标准评审、审批、发布等环节较高优先级，优先得到处理。

5)数据标准应用流程。数据标准应用是指业务部门在日常经营管理中使用信息项、业 务指标的标准化业务属性，采用相应的数据标准。

6)数据标准冲突协调流程。数据标准冲突协调流程是指当发现两个及两个以上标准存 在定义冲突时，统一向标准管理专职机构提交冲突协调申请。标准管理专职机构负责组织相 关管理专业就冲突问题进行讨论，制定冲突解决方案。

**(3)建立数据标准管理系统** 企业级数据标准管理系统旨在解决线下手工标准管理模 式造成的低效且缺乏统一发布、同步更新和实时响应的问题。它实现了对数据标准的全流程 统一管理，集中涵盖了标准建立、变更、查询、映射管理和流程管理等功能。

基于数据标准管理的实际需求，数据标准管理系统采用统一的框架，包括数据存储层、 基础服务层、标准管理层等逻辑架构，如图6-6所示。

1)数据存储层：包括数据标准、标准映射、用户操作日志、系统参数配置、用户权限 管理等信息。

2)基础服务层：提供数据标准系统的基础服务，具体包括：系统管理、数据标准对外 服务接口、权限接口。

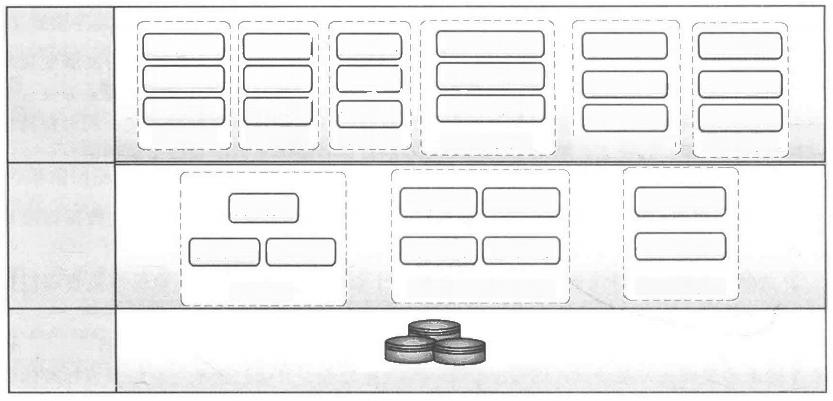
3)标准管理层：对数据标准进行统一管理、映射情况监控。

数据标准管理系统的功能主要由数据标准的建立、变更、查询、映射管理、流程管理和 知识库管理等模块组成，每个模块再细分为若干子模块，以满足标准管理的各项需求，并规 范数据标准的操作管理。其功能架构如图6-7所示。

① 标准建立。通过标准录入或者标准导入两种方式建立新的标准信息，并通过发布功 能发布最新的数据标准信息。

② 标准变更。标准变更包括编辑已发布的标准信息，废止已发布的标准信息，以及新

**银行数据治理**



标准映射新增

标准映射查询

标准映射删除

标准整改流程

标准映射管理

标准导入 标准查询

基础服务层

标准映射 落地查询

数据标准对外服务接口

数据存储层

数据标准存储库

角色管理

权限配置 日志管理

系统管理

UA权限验证

UA账号同步

统一认证接口

知识库发布

知识库更新

知识库下载

知识库管理

体系浏览 关键字检索 分类查询 标准查询

标准修改 标准废止]

标准再发布

标准变更

标准录入

标准导入

标准发布

标准建立

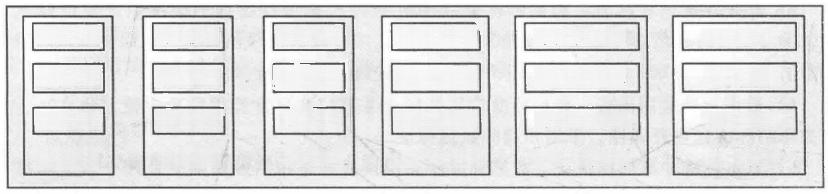
标准创建流程

标准变更流程

流程管理(EOA)

标准管理层

图6-6 数据标准管理系统逻辑架构



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 标准录入 | 标准修改 | 体系浏览 | 标准映射新增 | 标准创建流程 | 知识库发布 |
| 标准导入 | 标准废止 | 关键字检米 | 标准映射查询 | 标准变更流程 | 知识库更新 |
| 标准发布 | 标准再发有 | 分类查询 | 标准映射删除 | 标准整改流程 | 知识库下载 |
| 标准建立 | 标准变更 | 标准查询 | 标准映射管理 | 流程管理(EOA) | 知识库管理 |

图6-7 数据标准管理系统功能架构

标准的再发布。

③ 标准查询。实现按主题、按业务系统两个不同的视图对数据标准信息进行展现查询， 并支持对标准信息的关键字检索过滤。

④ 标准映射管理。实现标准在源系统的映射管理功能，包括映射新增、映射删除、映 射查询。

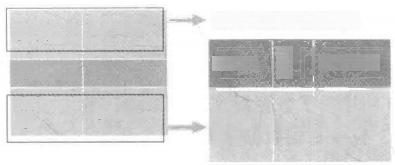
⑤ 流程管理 (EOA) 。 实现标准创建流程、标准变更流程、标准整改流程，方便用户对 数据标准问题的跟踪和处理。

⑥ 知识库管理。包括知识库发布、知识库更新、知识库下载，对数据标准建立、发布、 映射过程中的相关文档资料进行统一管理和归档。

**6.4** **非结构化数据的数据标准管理**

对于从非结构化数据中抽离出来的结构化信息，可以按照数据标准进行管理。例如，对 于前面所说的基本属性、底层特征可以建立数据标准，并与结构化数据标准实现共享。图 6-8所示为非结构化数据标准示例。

**数据标准** 第 6 章



2013-03-

15

|  |  |  |
| --- | --- | --- |
| 标准名  客户端 型号 | 长 度 | 标准代码  1 iphone 2三星 3 … |

发布日期

评论次数

客户端型 号

(共享结构化数据标准)

lphone

2

1

图6 - 8 非结构化数据标准示例

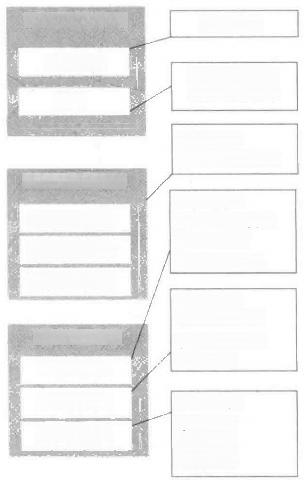
**6.5** **案例**

**1.巴克莱银行数据标准化管理**

图6-9 所示为巴克莱银行数据标准化管理组织。

巴克莱银行建立了数据标准管理的专业化组织和 汇报关系

队伍，形成了三层组织体系，包括决策层、管理层及

执行层。 信息标准委员会 财务总监

目前设有信息标准委员会 (Information Standardi- 分管领导

业务部门领导

zation Committee), 由财务总监 (CFO) 和各业务及技 成员 科技部门领导

术部门负责人组成，负责统筹全行范围内数据标准管

汇报对象：

控方案及各项工作，指导数据标准办公室开展日常工 信息标准委员会

作；委员会下辖信息标准办公室 (Information Stand- 信息标准办公室

ardization Office) 作为全行级长久性机构，配有8～10 标准质量专家 业务部门负责人

汇报对象：

办公室负责人

名工作人员，负责数据标准实施、报告以及综合管控。 标准合规审计 流程所有者

此外，作为各项标准日常管理的信息标准社区 (Infor- 信息架构专家

mation Community),则由大量的业务负责人、数据负 人

责人和系统负责人组成，分别从业务规则、数据管理 信息标准社区 信息架构专家

以及系统等不同角度共同管理维护数据标准。 业务负责人

技术总监

与组织角色相配套，巴克莱银行制定了关键角色 ：

的职能分工，保证决策层、管理层和执行层各司其职、 开发经理

运营经理

分工明确，同时与行内其他管理委员会形成良性互动。

**2.花旗集团数据标准化管理** 图6 - 9 巴克莱银行数据标准化管理组织

花旗集团建立了一套数据标准框架，通过组织协

调集团范围内的信息定义、数据标准等信息管理，实现风险管理改进、绩效提升和业务增长 的目标。其在以下方面得到了提升：

1)改善了集团范围内数据报表的一致性和完整性。

2 ) 在 集 团 范 围 内 提 供 对 相 关 数 据 内 容 的 统 一 认 知 。

3)实现信息的协同互用(在不同的系统、部门和业务之间)。

4)实现并购后的报表整合。

5)减少了没有必要的数据映射和转换。

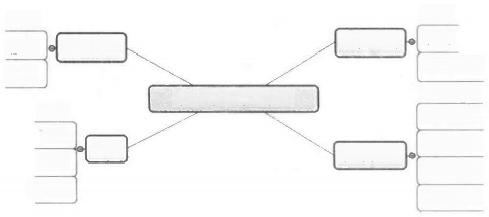
 **银行数据治理**

6)减少了因数据质量差引发的问题和错误。

7)减少了因报表错误引发的成本和风险。

**3.** **中国工商银行数据标准化管理**

中国工商银行在2010年推出了数据标准管理系统，管理全行数据标准，累计颁布的数 据标准总量达到3万多项，包括基础数据标准和指标数据标准。数据标准管理系统作为我国 金融行业首个全行级数据标准管理服务平台，为标准发布、标准学习交流、标准贯彻提供了 有力支撑。图6-10所示为中国工商银行数据标准化功能框架。



收藏 任务

日常管理

统计分析

系统参数 用户权限 帮助管理 备份恢复 日志管理

查询 下载 匹配 讨论区

新增 修改

删除

数据标准管理系统

扩展应用

系统管理

应用

图6- 10 中国工商银行数据标准化功能框架

**4.** **招商银行数据标准化管理**

招商银行在2010年9月启动了数据标准建设项目，2010年年底完成了第一阶段客户标 准、机构员工标准、公共代码标准的编制，以及数据标准编制规范和管理框架的咨询工作， 为数据标准的建设、管理和执行打下了良好基础。2011 年3月开展第二阶段工作，完成了 产品数据标准、业务数据标准和指标标准的编制，形成了完整的数据标准体系，包括几百个 业务对象、2000多项数据标准、400多个基础产品、600多项指标标准，并且建立了数据标 准管理系统和全流程数据标准管理机制，营造了用数据说话的文化与氛围，将其经营管理中 产生的数据、依赖的数据以统一的口径展现，促进了数据分析在客户经营、绩效管理、资源 配置、定价管理、风险管理等诸多方面应用与管理水平的提升。

**5.** **中国光大银行数据标准化管理**

中国光大银行从2008 年开始建设数据标准体系，并取得了丰硕的成果，形成了以科技 战略委员会领导下的数据标准化小组为管理组织，涵盖标准定义、执行、监督评审等各个环 节的良性数据标准闭环工作机制与流程，并配套管理制度与工具的建设。

中国光大银行的数据标准体系建设以应用执行为工作重心，与应用系统建设紧密结合， 在促进数据共享、提升数据质量、支持业务发展等方面发挥着重要作用，为实现“高度信 息化及最具创新力银行”的业务发展愿景奠定了坚实的基础。

**(1)持续开展数据标准体系定义与维护** 《中国光大银行IT 战略规划(2007—2010)》 中指出“由于银行不断发展的业务和监管需求，如客户获取、服务和挽留水平的提高、管 理数据质量的高要求以及严格的监管合规要求等，光大银行需要借鉴数据仓库研究成果，启 动数据标准体系建设，定义全行统一的数据模型，建立全行数据管理及相关流程，来进一步 提高对数据的要求。”2007年年底，中国光大银行启动数据标准体系建设工作，在引进具有 行业先进水平的金融逻辑数据模型 (FS-LDM) 和外部专家的行业经验，并结合巴塞尔Ⅱ监 管要求、新会计准则以及中国人民银行、中国银行业监督管理委员会监管报表规定等的基础

**数据标准** **第6章**

上，构建了中国光大银行数据标准体系框架。3年时间内完成了6个基础主题以及关键统计 指标、新资本协议风险等方面的标准，并建立起完善的工作机制，持续更新，以确保其与业 务发展的一致性。

**(2)形成了六位一体的数据标准体系应用方法**

1)规划先行。引进“实施优先级评价体系”,从迫切性、难易度、业务关注度等多个 层面进行考量，确定了各标准主题的实施路径，并结合业务部门需求与系统建设情况，适时 调整实施路径，保证标准“有人用”才是有用的标准。

2)业务驱动。始终贯彻以业务需求为驱动这一核心原则，使业务部门从标准应用中获 得益处，更有主动性地去推动标准的不断应用。例如，证件类型标准的应用，规范了个人客 户识别有效证件类型，明确了“台湾居民来往大陆通行证”为有效证件，解决了台湾居民 无法在中国光大银行办理第三方存管业务的难题。

3)项目推进。在标准体系应用策略上，通过选取业务价值高、涉及面广的项目或专题 进行专项改造、主数据管理等方式，有计划、有组织地推进，解决其涉及业务部门多、系统 多、关联复杂、风险度高的难题。同时，结合系统开发流程，开展数据标准评审，保证数据 标准在新建系统中有效执行。

4)阶段实施与区别对待。在标准体系应用的具体方案设计中，本着阶段实施与区别对 待的原则，从业务影响、技术影响、系统改造难度等多方面充分论证应用实施方案，明确系 统实施范围、标准落地方式以及分步实施策略，有效控制改造影响面，降低系统改造风险。

5)完善管理。颁布《中国光大银行数据标准管理办法》,对数据标准的组织结构、工 作职责，以及制定、复审、变更、执行的工作流程与制度要求进行了详细规定；重点明确了 数据标准化小组统筹计划，组织协调数据标准体系专项改造及系统应用，这一机制的建立对 落实数据标准体系应用起到了关键性作用。

6)平台支撑。2008 年，建设了数据标准管理工具，实现了数据标准体系的维护、更 新、发布等功能，向数据标准管理组织以及全行各部门、各分支机构全面开放，为大家学 习、使用、管理数据标准提供了有效工具。

中国光大银行依托数据管理工作，持续推动数据标准体系的应用，建立了以数据质量问 题改进为主要业务驱动力、数据标准为数据质量评价基础、元数据管理为支撑的三位一体数 据管理机制，侧重以外部监管数据质量要求为业务驱动力，以企业元数据管理为基础，推动 数据标准体系在各应用系统中的执行，逐步形成常态化的、闭环的数据标准体系应用良性机 制，从整体上提升了中国光大银行的数据质量及数据管理水平。

**6.** **中国农业银行数据标准化管理**

中国农业银行针对监管报表涉及的重点数据，制定并下发了《个人客户信息项标准》 《信贷类信息项标准》等多批数据标准，并在2010年在全行范围内发布了常用指标库，建 设了包括9大类925个标准，还启动了数据标准管理应用建设，为数据标准管理提供了强有 力的技术支撑。



**第** **7** **章**

**数据质量管理**

数据质量用以描述数据对使用者有价值和有用的程度。从本质上来说，数据质量问题是 信息失真的表现，不同程度上影响了银行相关业务的开展，导致客户满意度下降，同时也降 低了统计信息的可用性。因此，必须从各阶段入手进行数据治理，防范信息失真，有效提高 数据质量。

数据质量问题产生的因素有多个方面，主要有设计问题、传输和使用问题、操作问题等 方面。正因为数据质量产生于多个方面，因此其治理的难度不言而喻，例如：

1)由于历史原因造成大量缺失数据和错误数据，比较典型的就是早期银行开户时对客 户信息控制不严格，从而造成客户信息缺失或错误。

2)需求考虑不周导致数据质量问题的产生。需求人员在拟定需求时，往往从当前使用 场景出发，对后续使用场景以及与其他系统之间的关联关系考虑较少，出现系统间数据不一 致，当前数据无法满足新业务场景使用等情况。

3)操作不合规形成数据问题。操作人员在操作时错误或不严谨，也会导致问题产生， 比如输入随意的11位数字用作手机号码等。

只有数据质量提高了，才能更好地服务客户，提升银行经营管理水平。例如，完整的客 户信息可以有效地分析客户行为，给客户提供更为全面的服务；系统间一致的数据可以提升 数据的共享程度等。

**7.1** **商业银行数据质量存在的问题及影响**

随着信息化银行时代的到来，银行掌握的各类客户基本信息和客户交易数据已不是单纯 的信息技术资产，而是逐渐成为银行的重要战略资源。 一方面这些数据是银行正常办理各类 业务的基础，另一方面通过对已有数据的挖掘和分析，还可以服务于银行的精准营销、风险 管控、经营决策等。但是，随着银行数据的海量增长和数据源的不断扩充，在数据应用过程 中往往会发现数据存在各种各样的质量问题，包括数据缺失、数据不准确和数据不一致等。 这些问题直接影响到银行基本业务的开展和数据进一步的挖掘和分析。

这些问题产生的原因很多，从整个数据产生的流程上来看，包括需求不明确导致的设计

**数据质量管理** **|** **第7章**|

不周全，数据在传输和使用过程中出现偏差，人员在操作系统时直接输入错误等。例如，业 务需求不断变化，出现新增的客户信息要素，而已开户客户的信息需要重新联系客户进行采 集，导致部分客户信息要素不全；又如，柜员输入时操作不当，在输入时输错了客户姓名、 联系方式、证件号码等。

数据质量管理是指对数据从产生、获取、存储、共享、维护、应用等各个阶段可能引发 的各类数据质量问题，进行识别、度量、监控、预警等一系列管理活动，并通过改善和提高 组织的管理水平使数据质量获得进一步提高。数据质量管理是一个循环渐进的管理过程，其 终极目标是通过可靠的数据提升数据的使用价值，并最终为企业赢得经济效益。其主要目的 是解决企业内部数据使用过程中遇到的数据质量问题，提升数据的完整性、准确性和真实 性，为银行的日常经营、精准营销、管理决策、风险管控等提供坚实、可靠的数据基础。

对银行而言，数据质量管理工作的目标是在数据治理管理规定框架下，依据数据质量的 特点、规模、复杂程度和变化情况，建立和实施长效数据质量管理机制，营造数据质量管理 氛围，全面提升数据质量意识，实现信息资源的效益最大化。

很多人对数据质量重视程度较低，这不仅仅表现在使用数据的业务人员身上，科技人员 也有此表现，认为得不偿失，为了1%的收益付出100%的努力。果真如此吗?从大的方面 来说，民众对统计数据存在一定怀疑，这就是典型的统计上的数据质量问题；从小的方面来 说，由于手机号码错误导致真实客户收不到自身的交易短信，造成了客户流失。试问，谁还 敢说数据质量不重要?这里以两个真实的例子来说明数据质量的重要性。

实例一：某商业银行最近在做客户信息完整性、真实性的梳理及质量工作中发现，某分 行有一个客户10年前申请了一笔助学贷款，但是客户姓名、客户地址、客户联系方式居然 全记录为“0”,而后这个客户近两年又申请了一笔住房贷款，居然又发生了这种情况。幸 运的是，此客户助学贷款如期还完，住房贷款也在按期还款。抛开贷款金额不说，这种数据 质量问题对银行将造成极大的风险，如果客户不按期还款，那么银行就无法联系到客户，也 就没有办法采取任何挽救措施，将对银行造成一定的资金损失。

实例二：最近王先生为其爱人办理了一张金卡副卡，但是当王太太打电话要求开通时， 要求其输入办卡时预留的电话号码，结果王太太输入自己及爱人的电话号码时都提示不正 确，导致银行卡无法开通。王先生专门到银行柜台进行了咨询和办理，发现办卡时柜员忘记 录入电话号码，导致不仅浪费了王先生的时间和精力，而且给柜员的工作增加了负担。

从以上两个简单的例子以及从评价数据质量的角度来看，商业银行数据质量存在的问题 主要有以下几个方面：

**1.无法满足一致性和完整性要求**

一致性是指相同的指标在不同的系统中应该保持一致。但是，在各银行的现行系统中， 很多数据是不一致的。比如，标识个人客户的“性别”,有的系统用“01”“02”标识，有 的则用 “A”“B” 标识；还有，标识企业经济性质的字段有的系统叫作“企业性质”,有的 系统叫作“公司性质”或“公司类型”。由于各系统间广泛存在的指标、字段不一致现象 使系统间的数据交换需要建立复杂的映射关系，导致数据交换、加工非常困难。

随着银行业务复杂性和监管要求的提高，对客户信息、业务信息的要求也越来越高。比 如，银行在开展同业业务时，不仅需要业务本身的期限、金额、利率、起始日、到期日等信 息，而且需要交易对手的信息。还有，在判断业务风险时，对集团客户、关联客户信息的需

 **银行数据治理**

求越来越多。但很多银行并没有建立起相应的信息系统，对很多信息都不进行采集，甚至对 一些客户的通信地址、邮箱、手机等重要沟通信息也存在缺失的现象，这就导致客户或业务 信息的不完整。

**2.** **无法满足合规性、唯一性、有效性要求**

银行作为受监管最为严格的金融部门，几乎所有的业务都有明确而且严格的监管要求， 即使是一些创新业务，也会在很短的时间里出现监管规则。这就对银行信息系统的功能和适 应性提出很高要求。比如“理财业务”,刚开始有的银行将其放在表外核算，后来有的银行 只将“非保本理财”放在表外核算。但是，随着银行监管要求的不断提升，关于“理财产 品”核算和统计也出现了新规范，要求银行必须及时适应，否则就会出现不合规的问题。

在银行对客户信息及交易行为进行记录时，某个特定信息在数据集中时不应该有重复 值。但是，由于过去业务操作不规范，而且内控体系不严密，数据的唯一性无法得到保证。 比如“客户号”作为客户的重要标识，在核心系统中应该是唯一的。但在实际业务中，有 的客户既用身份证开户，又用户口本开户，有时还用护照开户，这就导致出现多个客户号。 如果从跨系统的角度考虑这一问题，问题将更为突出。

随着银行业务品种的不断增加，银行业务系统的数量也在不断增加。为了确保业务数据 能够在系统间进行有效处理，业务数据都应该遵守统一的数据格式。而在实际工作中，各系 统建设的时间早晚有差异，有的系统是行内自主开发，而有的系统则从行外采购，遵循的数 据格式各不相同。这就导致系统内数据无法进行有效的加工，从而无法满足分析和统计的 要求。

**3.** **无法满足及时性、可获取性、准确性要求**

在银行经营管理架构中，业务系统建设通常由业务和信息技术部门负责，其重点在于业 务运行正常。当国家相关部门，如标准、统计、监管等对数据标准、数据口径有新要求时， 通常又由负责统计分析的部门来落实。工作目的和要求不同，导致新口径、新标准在落实中 常常滞后于相关政策的要求，这就使数据的及时性要求无法得到满足。

在日常经营决策中，银行综合管理部门应该可以方便、简易地获得行内各业务系统的数 据，各业务部门也应该可以方便地得到本地业务数据。但是，在实际工作中，由于系统数量 较多，系统功能差异明显，以及人员流动等原因，行内各系统“有什么数据”和“数据在 哪里”常常成为困扰管理人员的难题。另外，由于部门数据需要经过多次加工，需要整合 为多个系统的数据，也使数据的可得性大大下降。例如，某一时间点客户的“授信使用率” 就需要整合授信管理、信贷管理、债券投资、票据贴现、国际业务等多个系统的数据，还要 考虑多个币种的折算。正因为存在上述问题，对数据质量最重要的准确性要求就很难得到保 证。数据准确性不高，必然导致分析的偏差和决策的错误。

从近10年全球银行业发展的趋势来看，在整个经营过程中，强化数据整合能力，通过 数据分析确定发展战略，全方位地推行数字化管理成为先进银行的共同选择。从最近风起云 涌的互联网金融发展路径看，更是将数据积累、数据处理、数据分析的优势发挥到了极致。 例如，通过对海量数据的分析，阿里巴巴小额贷款股份有限公司可以达到日放贷10000笔的 量级。可以预期的是，随着经验的积累、模型的优化、信用环境的改善、风险控制能力的提 高，互联网金融公司必将对传统银行业形成重大冲击。在利率市场化不断加快、行业竞争日 益激烈的背景下，数据质量问题已经成为决定商业银行生死存亡的关键因素，数据质量问题

**数据质量管理** **第7章**|

对商业银行经营决策将会带来各方面的影响，比如：

**1.达不到监管要求**

在全球经济一体化趋势持续加快的推动下，影响一个国家经济金融稳定的因素越来越 多，为了做好系统性风险的识别和防范，监管部门对监管统计数据质量的要求不断提高，比 如对风险敞口的监测，对银行交易对手风险程度的判断，对影响银行流动性风险、信用风险 相关因素的分析，还有对银行创新业务规模的要求等。这些监管要求对银行数据的完整性、 及时性等提出了更高要求，数据质量管理工作薄弱的银行越来越难以达到监管部门的要求。

**2.** **找不到目标客户**

在利率市场化稳步加快的背景下，各银行受保护的利差收益将逐步消失，细化市场分 工、突出自身优势、实施差异化经营成为银行参与竞争的必然选择。但是，在客户信息和交 易数据积累不足的情况下，银行无法对客户风险度、贡献度进行度量，难以确定适合自身能 力的目标客户群，也难以制订有针对性的价格体系服务客户。

**3.** **绩效考核不到位**

随着现代商业银行经营管理水平的提高，精细化管理要求各银行要通过实施管理会计来 实现按产品、机构、部门、渠道、员工等进行核算和评价，这就要求相关部门能够精细、准 确、及时地得到银行的经营数据。这对许多银行的数据颗粒化程度、会计科目的设置等都提 出了更高的要求。

**4.** **风险管理难以细化**

风险管理是银行经营的永恒话题，商业银行在开展业务的过程中，必须对客户、行业、 区域，甚至国家的风险进行量化评估，做出是否进入，以及如何定价的经营决策。此外，对 风险承受度评估的压力测试也对数据质量有很高的要求。在数据质量存在问题的情况下，风 险管理不可能实现细化和量化。

**5.** **影响产品创新**

现代商业银行的竞争归根结底是产品的竞争，是服务能力的竞争。银行的产品创新总体 上可以分成两种： 一是填平补齐式的，主要解决人有我无的问题；二是独创领先式的，主要 确立人无我有的优势。要提高这两种创新的针对性、有效性，都需要对产品本身的成本、风 险，以及客户的需求、偏好进行分析，这就需要有大量高质量的数据作为支撑。

近10年来，各商业银行对数据价值的认识不断深化，几乎所有的银行都在数据集中、 信息系统建设、内部制度建设等方面加大了投入，中国人民银行、中国银行业监督管理委员 会等金融管理部门也对银行的数据质量管理提出了更高要求。从各行的实践看，全面提升数 据质量需要多管齐下，建立全方位的提升数据质量管理体系。

**1.** **要树立“数据是资产”的理念，构建数据质量管理工作的组织体系**

在传统的银行经营管理中，对贷款、投资、同业拆放等营利性资产比较重视，后来逐渐 对品牌、软件等无形资产较为重视，部分银行对抵债资产、自用及非自用不动产也非常重 视。随着信息技术的发展，“数据也是资产”的理念日益得到银行的认可。从现在的发展方 向来看，“客户黏性”已经成为影响银行经营的一个重要因素。比如，有的客户只在某家银 行有一笔存款或贷款，而没有其他业务，银行就无法获得其他经营信息，无从对其风险度、 经营偏好进行分析，也无法持续开展新业务、增加中间业务等活动。银行的持续经营和发 展，需要的是大量有“黏性”的客户，有“黏性”的客户才是银行忠实的客户，才是可以

 **银行数据治理**

为银行带来持续收入和持续合作的客户。要提高客户的“黏性”,就要为客户创造价值，与 客户同步成长。由于竞争的加剧，银行和客户的互动关系正朝着“给客户带来价值——客 户黏性增加——产生大量交易数据——对数据进行分析和决策——有针对性地研发产品—— 进一步提升客户价值”的逻辑关系发展。数据质量管理工作的专业性强、涉及面广，与行 内所有的业务部门都有关系，与风险管理、营运、计划财务等综合管理部门以及信息技术等 部门也密切相关。要做好数据质量管理工作，必须建立归口管理明确、分工细致、团结合作 的组织体系。目前，有的银行成立了专门的管理信息部，或由信息中心牵头，相关部门配 合，有的银行则由计划财务部或信息技术部牵头，其他部门配合。无论是哪种模式， 一个完 善的组织体系是做好数据质量管理工作的基础条件。

**2.建设数据质量检核工具，完善全流程的数据质量管理制度**

评价数据质量高低，必然要从一致性、唯一性、完整性等几个角度对数据进行分析。由 于目前银行业务高度复杂， 一些重要数据通常涉及多个系统。比如对公客户的信息，通常在 柜面业务系统、信贷系统、对公客户信息系统、国际业务系统等中都存在，部分数据项出现 重叠，有些数据项会不一致。由于银行系统中无论是数据项还是数据量都庞杂无比，靠人工 对其质量进行筛查及检核是不可能的，必须建设功能强大的数据检核工具。数据检核工具的 主体内容是数据检核规则，这些规则要体现不同的数据质量要求，比如根据合规性要求，数 据口径必须符合监管要求，就要将监管要求的数据口径转化为检核规则，再运用这些规则对 各系统中的数据进行检核。另外，根据唯一性原则，需要对一些重要信息进行跨系统检核， 判断其是否唯一。从各银行的具体情况来看，部门林立、系统众多、规模庞大、业务繁杂是 当前中资银行的基本特点。在这种格局中，开展数据质量管理工作必须有完善的制度作为保 证。从全行管理的角度讲，要有各部门的分工、职责等制度；从基础制度讲，要有关于数据 标准、数据口径、名词解释的制度；从具体操作看，要有查询、上报、纠正、考核等制度； 从系统开发与运作看，要有需求讨论、系统开发、标准落地、维护升级等制度。这些制度有 些偏业务，有些偏技术，有些需要两者紧密结合。总之，只有通过建立健全这些制度，才能 形成一个有利于进行数据质量管理的制度环境。

**3.** **建立成体系的数据质量监控、考核体系，制订数据质量提升规划**

由于数据质量管理工作涉及银行的前台、中台甚至后台，既涉及总行部门，也与支行的 柜员、客户经理有紧密联系。仅仅有操作手册、业务制度还不能保证数据质量，必须建立多 层次的数据质量监控、考核体系。比如，董事会应该将数据质量纳入内部审计工作范围，把 数据质量作为评价经营层绩效的重要方面；银行绩效考核部门应该将数据质量纳入部门、机 构的绩效考核中；归口管理部门应该通过日常监管和例行检查，对相关部门和支行的数据质 量管理工作进行监测和评价；支行则应将数据质量考核纳入本支行的考核方案。只有对数据 质量开展有效监测，将其纳入绩效考核方案，所有制度才有可能得到执行，数据质量才有可 能得到持续提升。各商业银行要在互联网时代、大数据时代争得发展先机，必然要将提升数 据质量作为管理工作的重点内容。由于受到各行信息系统建设的历程、路径各不相同，数据 标准化程度也存在较大差异，制度建设和管理体制千差万别等条件的约束，提升数据质量必 然是一个长期的过程。各行需要通过全面评估本行的数据质量，根据不同数据对本行业务的 不同影响，制定系统的提升数据质量规划，分步骤地解决数据质量问题。比如，对于重要客 户重要标识的唯一性问题，可能需要尽快予以解决；对于客户信息完整性问题，可能要考虑

**数据质量管理** |**第7章**|

业务规模、管理模式逐步予以解决；对于标准化问题，需要按“新系统新要求，老系统分 步走”的模式予以解决。总之，数据质量的提升不可能一蹴而就，需要通过较长时间的不 懈努力，需要通过有目标、有秩序的规划，逐步使银行内的海量数据变成经营决策的基础， 变成贡献银行利润的“金矿”。

**7.2** **建立全面管控的数据质量管理体系**

数据质量管理已经成为银行数据治理的有机组成部分。目前，我国主要的商业银行也基 本形成了制度、流程和系统三位一体的数据质量管理体系。

只有建立完整的数据质量管理体系，以规范和制度为约束、以工具为辅助、以系统为支 撑，才能够从设计、开发、生产等各个环节发现数据质量问题，及时进行治理并提炼相应的 数据质量检查规则，防止同类问题的重复发生，有效提升银行数据整体数据质量，从而更好 地为客户服务，提供更为精确的决策分析数据。

**1.规范与制度体系**

“没有规矩不成方圆”,要做好数据质量管理工作，首先要有明确的理论指导和约束， 因此规范和制度先行是极其有必要的。

规范制度应该是普遍适用的，可以有效指导数据质量管理工作，包括明确不同成员的职 责与工作方法，从而有效推动数据质量问题的主动规避、发现和治理。

**(1)在技术层面制定相应的规范和手册**

1)制定相应的管理规范。技术层面上，应完整全面地定义数据质量评估维度，包括完 整性、时效性、唯一性、正确性等。按照已定义的维度，在系统需求阶段、设计阶段、开发 阶段、测试阶段、运行阶段都应根据实际情况设计数据质量检查规则，用于发现各阶段的数 据质量问题，以及时进行治理。

根据业界成果理论，数据质量评估维度的具体内容见表7-1。

**表7-1** **数据质量评估维度的具体内容**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 维度 | 描 述 | 衡 量 标 准 | 是否自 动核查 |
| 1 | 完整性 | 必须的数据项已经被 记录 | 业务指定必需的数据是否缺失，不允许为空字符或空值 等。例如，数据源是否完整、维度取值是否完整、数据取值 是否完整等 | 是 |
| 2 | 时效性 | 数据被及时更新以体现 当前事实 | 当需要使用时，数据能否反映当前事实。即数据必须及 时，能够满足系统对数据时间的要求。例如，数据处理(获 取、整理、加载等)的及时性 | 是 |
| 3 | 唯一性 | 该数据在特定数据集中 不存在重复值 | 在指定的数据集中是否存在重复值 | 是 |
| 4 | 参照完整性 | 数据项在被引用的父表 中有定义 | 数据项是否在父表中有定义 | 是 |
| 5 | 依赖一致性 | 数据项取值满足与其他 数据项之间的依赖关系 | 数据项取值是否满足与其他数据项之间的依赖关系 | 是 |

 **银行数据治理**

(续)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 维度 | 描 述 | 衡 量 标 准 | 是否自 动核查 |
| 6 | 基数一致性 | 数据项在子表中出现的 次数符合标准 | 数据项在子表中可出现的次数是否符合标准。例如， 一 个账户每年的计息次数为4次，即要符合账户计息次数为  1:4的规定 | 是 |
| 7 | 正确性 | 数据正确体现了真实 情况 | 数据内容和定义是否一致 | 是 |
| 8 | 精确性 | 数据精度满足业务要求 的程度 | 数据精度是否达到业务规则要求的位数 | 是 |
| 9 | 技术有效性 | 数据符合已定义的格式 规范 | 数据项是否按已定义的格式标准组织 | 是 |
| 10 | 业务有效性 | 数据符合已定义的业务 规则 | 数据项是否符合已定义的业务要求 | 是 |
| 11 | 可信度 | 数据的可信赖程度 | 根据客户调查或客户主动提供获得 | 否 |
| 12 | 可用性 | 数据在需要时是可用的 | 数据可用的时间和数据需要被访问时间的比例 | 否 |
| 13 | 可访问性 | 数据易于访问 | 数据是否便于自动化读取 | 否 |
| 14 | 适用性 | 数据格式和展现满足用 户需要 | 根据客户调查或客户主动提供获得 | 否 |

注：“是否自动核查”是指能否通过系统自动检查的手段监测数据质量问题。

基于以上维度，在规范中应明确以下内容：

>明确涉及数据质量的设计要求，从事前控制问题的发生，包括在系统设计、功能设计 时充分考虑可能发生的问题。

>明确相应的逻辑原则，通过应用系统控制，从事件过程中矫正可能发生的数据质量， 包括技术上的逻辑和业务上的逻辑。

>明确如何检查数据质量问题，从事后及时发现问题并予以解决。

2)制定相应的管理手册。商业银行需要制定数据质量管理工作的相关手册，用于具体 指导数据质量管理工作的开展。手册需要明确各数据质量管理团队成员的职责，以及数据质 量管理的工作方法，从而有效推动数据质量问题的主动规避、发现和治理。

**(2)明确相应的管理流程** 数据质量问题会发生在各个阶段，因此需要明确各个阶段 的数据质量管理流程。例如，在需求和设计阶段就需要明确数据质量的规则定义，从而指导 数据结构和程序逻辑的设计；在开发和测试阶段则需要对前面所提的规则进行验证，确保相 应的规则能够生效；最后在投产后要有相应的检查。每个阶段的工作应该有一定的指导流 程，从而将数据质量问题尽可能消灭在萌芽阶段。

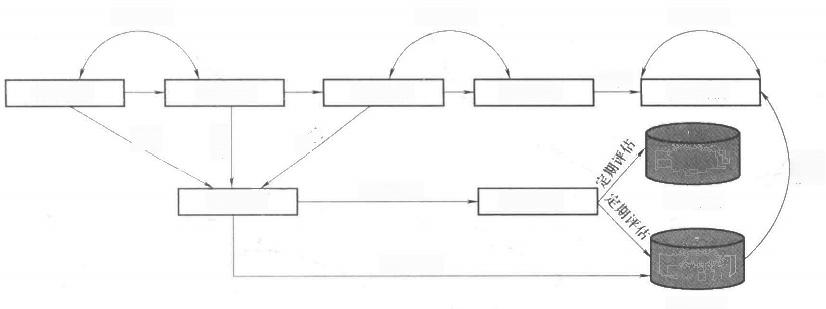
图7-1 所示为某商业银行数据质量管理流程。

**2.** **建设数据质量管理系统与工具体系**

在业务系统建设过程中，往往存在重功能轻数据的情况，致使系统中遗留了大量有问题 的历史数据，并且错误数据不断出现。

银行的数据质量管理措施，宜采用控制增量、消灭存量的策略，有效控制数据质量问题 的产生。结合数据标准检查，依托业务系统上线的机会，解决增量数据的规范性问题，通过

**数据质量管理** **第7章**



规则验证

投产

开发阶段

妙

问题 修正

规则导入

检查规则

规则录入

管理平台

测试阶段

检查工具

规则定义

规则使用

生产系统

投产阶段

设计阶段

需求阶段



设 计

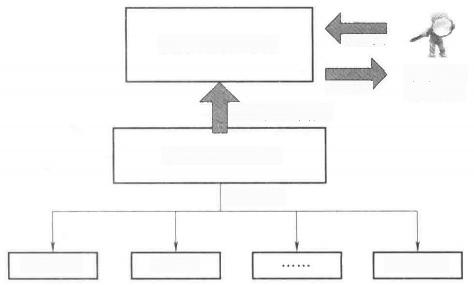
图7-1 某商业银行数据质量管理流程

建设数据质量管理系统，实现数据质量问题发现、分析、分发、监测、考核、指标发布等 功能。

**(1)建立数据质量管理系统** 结合数据仓库等信息集成作用，对数据制定相应的数据 质量检查规则，及时有效地发现数据质量问题，并予以解决。数据质量管理系统应该包含多 个维度的功能，包括评价指标体系，不同类型数据质量问题的解决模块，问题的分发、下 载、考核等模块。要实现源头的数据质量评价、分析和跟踪功能，实现境内外数据治理的一 体化，解决集团客户信息、统计分析、客户评价、营销等一系列境内外业务发展中遇到的数 据质量问题。

**(2)要有数据质量检查工具** 对数据进行及时监测，通过检查规则对数据进行检查， 找出存在的数据质量问题，以便能及时对数据或程序进行整改，进而提高银行应用系统的数 据质量。银行数据质量管理系统与工具体系如图7-2所示。

数据质量管理系统的基本原理是

将检查规则部署于数据来源系统，通 数据质量管理系统 检查规则

过筛选和过滤问题数据，将具有质量 数据质量

问题的数据通过平台统一储存，并实

间题迫踪 管理人员

错误数据的纳入

现以下功能：

**(1)错误数据自动捕获和存储功** 数据质量检查工具

能 通过部署检查规则，数据质量检 质量检查

查系统可以对每天新增或存量的数据

进行扫描，符合特定条件的数据会被 个人存款 银行卡 投资理财

筛选出来，按照不同维度进行划分并

加以存储，并可以提供用户下载和

图7-2 银行数据质量管理系统与工具体系

查看。

**(2)错误数据增量与存量统计功能** 可以为用户提供查询统计功能，使用户了解到当 期新增错误数据量以及还有多少待修改错误数据量。

**(3)考核通报功能** 考核通报是数据治理的重要手段之一。数据质量管理系统应支持 考核报表的自动生成和展现，用于反映数据质量情况和数据治理工作进度。所有考核通报都

 **银行数据治理**

应以数据质量检查规则为依托，例如个人客户信息完整率考核报表，用户在查询通报的同 时，可以通过通报相对应的检查规则将不完整客户清单直接下载到本地，用于后续补录。

**(4)质量监控功能** 对于时效性较高的问题，要求具有监控和通知功能， 一旦发现问 题，则应立即通知用户(包括邮件、短信等方式),保证数据的时效性。

**3.建立检查与跟踪体系**

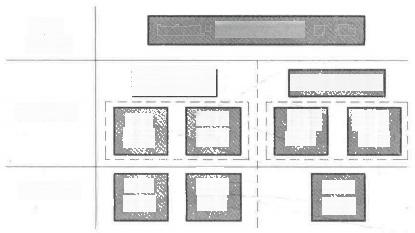
发现问题后还必须对问题的治理进度进行跟踪，要确保问题得到有效解决，因此有必要 建立数据质量问题的检查和跟踪体系，利用系统、人工等多种手段从多个渠道发现数据质量 问题，并有效进行跟踪和治理。

一般来说，可从事前、事中、事后等不同阶段进行处理，包括定期对数据表进行的设计 检查、生产问题提炼、收集数据接收方系统设计和开发过程中发现的数据质量问题等方面发 现数据质量问题，通过数据质量管理的系统来进行综合管理。系统还应包括问题提交模块、 问题分析与治理模块等，这样才能实现数据质量问题提交、跟踪、治理和关闭的流程化 管理。

**4.** **开展专项治理**

除日常的例行数据质量问题检查和治理外，还必须有针对性地进行专项数据治理，有计 划地提升某一领域的数据质量，例如配合各机构进行客户信息的九要素完善、可信手机号治 理等，才能为各类及各方向的数据质量提升奠定良好的基础。

**5.** **建立考核体系**

银行对各层级数据质量治理都有明确 度 高阶管理办法

的考核制度和办法。以某商业银行为例，

其数据质量相关办法与指标体系如图7-3、

陵心营理办法 支持性臂理办法

图7-4所示。 管理办法 团管

**6.** **建立数据质量管理流程**

通常情况下，商业银行数据质量管理 实施细则 

流程主要分为5个步骤，包括问题收集、

问题分析、制定方案、实施治理和考核监 图7-3 某商业银行数据质量相关办法

测。每个步骤都规定了具体的工作任务和

质量管理事项。

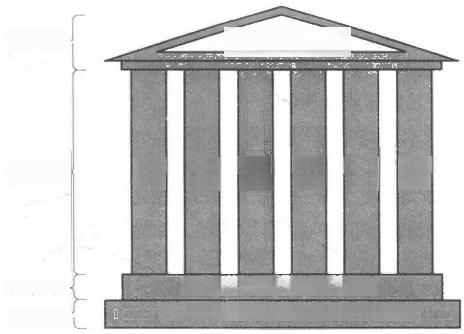
**(1)问题收集** 问题收集是银行数据质量管理流程的第一步。 一般来说，问题收集的 方法主要有以下4种：

1)根据监管要求与相关业务部门开展专题治理行动。

中国人民银行、中国银行业监督管理委员会等金融监管机构对各大商业银行报送的信息 质量都会定期考核，并且也会根据我国银行业发展中遇到的问题对商业银行内部各系统、各 业务的数据质量提出相关要求。为了达到监管部门对数据质量的要求，避免被监管部门通报 或罚款，涉及部门会通过召开专题会议的形式，向数据质量管理牵头部门介绍当前监管中发 现的数据质量问题，并共同分析问题产生原因，制定治理方案。

例如，中国人民银行要求的个人客户存款账户信息真实性核实工作，要求各商业银行在 规定期限内完成本行个人客户存款账户对应的姓名、开户证件类型和开户证件号码的核实工 作。这种治理活动由于有监管部门的明确要求，各部门开展治理的积极性较高，数据质量问

**数据质量管理** **第7章**



数据质量指标体系

对公

专业

数据查询、数据下载，数据屏蔽

用户管理、用户统计、参数管理、日志查询

基础功能

系统管理

运管 专业

个 金 专业

内控

专 业

指标体系

财会 专业

专业治理

*科技*

专 业

图7-4 某商业银行数据质量指标体系

题分析和纠改的进展较为顺利，并且由于监管要求是具有持续性的，在存量错误数据纠改完 成后，增量数据通过持续的监测与考核可以有效防止问题的反弹，达到持续性治理的效果。 但是，监管往往具有滞后性，已经发现的数据质量问题可能已经积累了较长时间，调整起来 对当前业务会产生较大的影响，也会有较高的治理成本。

2)根据规章制度开展数据质量问题筛查。

业务部门会通过制度规定对目前主要业务操作流程进行规范， 一般都会规定系统录入过 程中哪些字段是必须输入的，其格式、数据标准及有何校验关系等。有些系统的使用说明中 还包括字段的数据字典值等信息，由数据质量牵头部门通过专业的数据挖掘和探索工具，在 数据交换系统或数据仓库系统中根据业务制度和系统使用说明进行探索性检查。

根据规章制度开展的问题筛查，其优势在于行业的业务条线和系统都是确定且有具体范 围的，所以能够全面地对全行各类业务和系统进行检查。同时，由于这种检查是由数据质量 牵头部门进行的主动探索性检查，所以具有预防性，可以在数据质量问题发生的早期及时发 现并予以纠正。但其也有不足之处，如由于是数据质量牵头部门主动探索性发现的数据质量 问题，有些问题暂时对业务和监管没有重大影响，因此业务部门缺乏治理的积极性，后续治 理进展可能比较缓慢；另外，由于业务流程和业务系统在不断变化，对应的制度和手册可能 没有同步更新，因此有些发现的问题可能并非是真正的数据质量问题，或者有些问题不能被 及时发现。

3)配合重要项目或系统开展专项数据质量问题检查。

在一些重要项目或系统启动前，数据质量牵头部门会提前介入，对项目或系统要使用的 数据源开展检查，确保系统所使用的上游数据的准确性，同时提前审查系统业务需求中是否 设置了数据质量检查功能，对一些输入字段的完整性、准确性、有效性、 一致性等进行控 制。例如：某商业银行在建立法人和个人客户统一评价体系时，分别实施了法人客户和个人 客户统一评价专项治理项目，梳理出80类对重点项目有重要影响的数据质量问题，有力推 动了相关系统数据治理工作的开展。

专项检查方式的优势在于其通过在系统设计阶段加入数据质量管理环节，可以有效预防

未来数据质量问题的发生，在系统投产前期将上游数据源提前进行清理，也可大大降低系统

 **银行数据治理**

投产后再进行数据治理的工作成本。但是，限于数据质量管理资源的有限性，无法做到所有 的系统和项目在启动前都有数据质量管理人员提前介入，只能根据项目或系统的紧要程度来 分级，大多数小的项目或系统的需求暂时无法实现数据质量管理人员的提前介入。

4)由用户报告发现的数据质量问题。

很多数据质量问题是在数据的使用过程中发现的。例如，柜员是接触业务数据最为频繁 的人员，因此业务用户也是数据质量问题的重要来源之一。可以通过数据问题上报，不同用 户可随时将遇到的数据质量问题上报至数据质量主管部门。

人员报告数据质量问题的优势在于可以系统、全面地收集到各系统的数据质量问题，在 第一时间发现通过相关业务人员报送发现的问题。同时，由于问题直接来源于业务人员，因 此可快速定位问题数据的业务源系统，方便后续的原因分析和治理效果的评估。但是，由于 问题是由业务人员直接提出的，是否是数据质量引起的问题最终还需要数据质量主管业务部 门组织相关业务和技术部门分析后才能确定，因此可能很多资源被用在判断是否是数据质量 问题上面。同时，由于业务人员在报告问题时描述现象的详略程度不一致，因此可能有些问 题需要经过多次邮件、电话沟通才能补充完分析所需要的资料。

**(2)问题分析** 问题分析阶段的主要任务包括确定是否属于数据质量问题、定位业务 源系统和找出问题产生原因三项。

1)排除非数据质量问题。排除非数据质量问题贯穿于整个问题分析阶段， 一些比较明 确的问题在问题分析步骤初期可以直接排除。另外， 一些问题需要经过分析后才能最终排 除，如业务源系统内数据没有问题，后续业务在使用数据过程中因加工逻辑问题或口径问题 导致的报表数据错误。

2)定位业务源系统。定位业务源系统主要依靠业务专家和技术专家的支持，通过集中 办公、专题会议、电话和邮件等多种方式沟通并确认。对于业务用户直接反馈的数据质量问 题，可以找相关系统主管部门的业务专家进行确认；对于通过仓库系统发现的数据质量问 题，首先需要熟悉仓库架构的技术专家找到仓库内数据的来源，确定大致业务源系统的范 围，然后由相关系统业务主管部门的业务专家来进行最终确认。

3)确定问题产生的原因。在定位到业务源系统后，需要对问题产生的原因进行确认 这个过程需要业务专家和问题提出人员的大量参与，通常需要反复提取错误数据的样本数据 进行协助分析。业务源系统的数据可以通过业务部门的业务专家直接从前台提取，但是，对 于一些历史数据或涉及多业务源系统的数据，都需要依赖于数据仓库的技术专家通过数据仓 库进行提取。

**(3)制定方案** 在确定问题产生的原因后，由数据质量管理部门牵头，系统主管部门 和科技部门参与，制定数据质量问题解决方案。在制定方案时，要综合考虑各种方案的优点 和缺点，考虑其对现有业务的影响，以及存量错误的纠改和增量错误的防控，如果涉及前台 业务人员的人工干预，还要充分评估工作量，防止前台业务人员工作压力过大。

目前数据质量的改进方案主要有以下四种：

1)优化前台系统或者交易，即通过在业务系统中增加相应的控制条件，防止产生新的 错误数据。该方案一般适用于因业务人员录入或操作错误而导致的数据质量问题，由业务部 门负责提出具体优化需求，科技部门评估其技术可行性，评估通过后交由研发部门负责具体 实施。

**数据质量管理|第7章** |

2)通过数据变更解决，即通过直接修改后台数据库的方式，大规模更新问题数据。该 方案适用于数据量较大，但纠改规则明确，并且根据已有其他信息作为参照可以推定确定待 修改信息的正确值。 一般需要业务部门给出明确的纠改标准，由科技部门在系统后台数据库 中进行批量刷新。

例如：个人客户的性别和出生日期在个人账号实名制前不是必填项，因此很多客户的性 别和出生日期是空的或者错误的，在实施实名制后，这两类信息可以通过身份证号来统一进 行判断和刷新。

3)通过手工修改解决，即通过前台业务人员或柜员在业务系统中直接对某些问题字段 进行核实和修改。该方案适用于无法通过系统大批量更正的信息或者需要与客户进行确认的 信息。

例如：个人客户的联系方式(手机、固定电话),仅通过校验工具无法确认其是否真 实、有效，需要业务人员在客户来银行办理业务时与客户沟通确认，或通过电话、短信等形 式与客户联系确认。

4)通过制度、通报等手段约束业务人员。通过发布业务操作规范来指导业务人员在办 理业务时完整、准确地选择或填写各项信息，可以有效预防因操作原因导致的数据质量问 题，同时需要配套通报、考核等奖惩制度，督促业务人员严格遵照规范办理业务。

例如：近年来随着银行精细化管理的要求越来越高，各种中间业务收入都需要还原到单 一客户，但是由于柜员录入信息不完整或不准确，造成部分中间业务数据无法还原，以及银 行风险控制的原因，对已经发生的交易严格控制进行反交易，所以无法对历史数据进行修 改。同时，由于部分特殊交易确实无法还原到单一客户，所以无法通过系统控制方式完全禁 止，只能通过加强对业务办理人员的指导和约束来预防此类错误的发生。

**(4)实施治理** 实施治理包括两部分内容：

1)预防性措施。防止未来新的错误数据产生，防止出现边治理边污染的现象，找到并 实施从根源上可以解决数据质量问题的合适方案。比如，可优化前台系统或者交易，在前台 系统中增加对字段的校验程序，或单独开发一项数据清理程序嵌入到前台系统之中，用于对 进入系统的数据进行统一清理和规范；或者通过制度、通报等手段约束业务人员，如要求柜 员在使用会计科目时严格遵循行内会计科目使用规范，并定期对全行会计科目错误或滥用的 情况进行通报考核，考核结果作为年终绩效的参照依据。

2)清理性措施。用于纠正现有的存量错误数据，具体的清理性措施有数据变更和对错 误进行手工修改等。

**(5)考核监测** 考核监测作为一种监督、检查的手段，可以确保治理的时效性和最终 效果，督促业务人员根据已确定的治理方案按计划有序开展治理工作。通常，目前我国主要 的商业银行已建立了一套较为成熟的综合与专业相结合的多层次通报体系。

1)全行综合性考核通报。全行综合性考核通报是银行内部层级最高、综合度最大的通 报，由数据质量综合管理部门负责起草，按月通报全行数据质量情况，起到对不同机构相关 专业的督促作用。

全行性通报包括多个专业的综合性指标， 一般以年为周期， 一个周期内为保持得分的可 比性， 一般不会对指标进行较大的调整。

2)分行考核通报。各分行根据总行下发的考核办法制定本行的数据质量考核办法，并

 **银行数据治理**

对所辖二级行和支行的数据质量进行考核，督促分行数据质量管理工作的开展。

分行考核通报的特点是可以将考核粒度做到最细级别，弥补总行考核仅到一级行的不 足。由于最终绝大部分产生数据质量问题的业务或交易是在支行或网点层级产生的，因此直 接考核支行或网点对数据质量问题纠改和预防的意义更大。

3)全行专业考核通报。在全行综合性考核通报的基础上，总行各专业可依据本专业的 业务特点开展专业性数据质量考核通报，通报本专业内部全行数据质量情况，尤其是一些对 专业业务的办理有较大影响的事项。

上述三种通报手段都是从业务角度对数据质量进行考核通报的。其优点是便于业务部门 理解，并且指标的业务含义明显，提升后的价值更容易衡量；其缺点是这类通报往往以百分 比显示最终考核的结果，而无法展示到单条错误数据的粒度。

4)数据质量错误数据统计表。数据质量错误数据统计表从技术角度对发现的数据质量 问题逐条进行统计，分为增量错误数据统计和存量错误数据统计两类，并且可以支持对发现 的错误数据明细进行下载。其存量统计功能更方便了解当前系统内的数据质量整体情况，对 于已经完成纠改的错误数据，其可以自动在下个统计周期的存量错误数据报表中剔除，这样 每期的存量错误数据统计表则直接展示当期状态下系统的错误数据总量。

数据质量管理已经成为银行数据治理的有机组成部分。目前，我国主要商业银行也基本 形成了制度、流程和系统三位一体的数据质量管理体系。

“源头负责制”是一般商业银行数据质量管理的基本原则，各部门依据该原则明确分 工，确保数据质量管理工作的顺利开展。

从战略到具体办法三个层面的数据质量管理制度为质量管理工作不但提供了原则性意 见，而且提供了具体措施和方法。以制度为依据，通过将所有数据质量问题的处理流程制度 化、标准化、系统化，明确从问题收集到最终的考核监测每个步骤，以及相应流程的可选择 的处理方法。

数据质量管理系统的建成则为制度的落地和流程的执行提供了技术保障，通过系统方式 将数据质量管理制度的规定和形成治理流程固化到系统中，为数据质量管理工作的参与者提 供了有力抓手。

**7.3** **非结构化数据的数据质量管理**

非结构化数据由于格式和内容的多样性， 一般不会直接用于分析，即单个非结构化数据 难以产生价值。只有将非结构化数据的相关元数据(或标签)、主要内容抽取出来，并以类 似结构化数据存储后，才能发挥价值。因此，非结构化数据的质量管理主要在于提升数据可 用度， 一般通过对数据的清洗、映射、关联等机制，把繁杂的大数据变成能应对的、有效的 “小数据”,构建干净、完备的数据集，从而提高数据质量。

1)通过清洗规则库识别并标识出非结构数据的可疑或无效信息，必要时可结合数据生 命周期管理策略及时清理。

2)通过与相关的数据标准进行映射，将非结构化数据的“基本属性”转换为商业银行 的标准信息进行存储(如将“建行” “建设银行” “中国建设银行”等转换成标准代码

**数据质量管理**| **第7章**|

JG0001), 并作为结构化数据存储在系统中。

3)通过关键信息与商业银行其他结构化数据进行关联(如通过客户姓名、手机号等辨 识其是否为本行客户),提高非结构化数据入库质量和数据价值。

**7.4** **案例**

**1.中国工商银行数据质量管理**

中国工商银行在2008年建立了数据质量管理规范，并搭建了数据质量管理系统，以 “把好数据质量入口关”和“数据质量源头负责制”为指导原则，将数据质量管理作为一项 重要工作来抓，积极开展数据质量检查和数据质量问题治理工作。先后在个人业务、公司业 务、机构业务等13个专业条线结合业务重点开展了形式多样的数据治理工作，并取得了大 量丰硕的成果。据不完全统计，依托数据质量管理系统，部署了数据质量检查规则近万余 条，覆盖了上游各业务应用，检查范围涉及数据字段完整性、数据字段业务有效性、数据字 段业务关联性、数据字段业务唯一性等各个维度，从而实现对数据质量的全面监控。从数据 质量管理系统建立至今，通过平台完成的数据质量问题纠改已达几百万条，成果十分显著。

**(1)客户信息质量治理** 客户信息的完整性和有效性是提升精准营销成功率、线上线 下联动服务水平、客户行为分析能力的前提条件，也是满足反洗钱和反恐怖融资等监管和报 送要求的基础。为持续提高客户信息数据质量，中国工商银行在全行开展了个人及法人客户 信息的数据治理工作，在提升个人和法人客户信息数据质量上取得了良好的效果。通过部署 客户信息相关数据质量检查规则，累计发现、分析、定位问题数据几千万条，有效提升了客 户信息完整率。

**(2)支持对公业务收费数据质量治理工作** 从2010年开始开展对公业务收费数据质量 通报工作，部署相关考核报表和检查规则，通过加强培训和持续考核通报等手段提升对公中 间业务收入可分摊给单一客户的比例，将对公中间业务收入金额中可分摊给单一客户的比例 提升了1倍多。

**(3)持续开展主机总分不平监控** 按日对主机总账、分户账余额不一致情况进行监控， 发现问题后分行将相关原因反馈至总行，对人工误操作造成的不平在考核通报中给予扣分， 将系统原因导致的不平转给相关业务部门做系统优化和改造。

**(4)元数据质量专项治理** 随着中国工商银行业务的不断发展，信息量不断增大，元 数据信息也日趋增多，数据的准确性和时效性日趋重要。科技队伍通过优化系统功能、加强 流程管控等措施解决数据质量问题，建立全方位质量评价体系，完善监控措施。通过一系列 措施，提升了表及文件的准确性。

除规则部署外，中国工商银行还通过平台实现了“问题上报模块”,提供数据质量问题 的提交、跟踪、解决和关闭的全闭环流程管理。目前，用户主要包括总行、 一级行和二级行 各业务部门用户以及相关科技部门用户。用户在发现数据质量问题时，可通过此功能提交数 据质量问题，及时对数据质量问题进行转发、跟踪和解决，避免当前人工管理的各种弊端。

**2.** **中国建设银行数据质量管理**

中国建设银行根据业界普遍认可的数据质量最佳实践及数据质量管理的研究成果，结合

 **银行数据治理**

自身数据现状，制定了一套完善的数据质量管理实施方法论，为数据质量监控体系建设提供 了理论指导和行动指南。该方法论是一个包括定义和验证、影响分析和共性分析、追踪根本 原因、预防和修复数据质量问题、趋势监控、识别研究偏差的闭环过程。该方法论承认数据 很少具有理想质量的事实，为了达到质量提升目标，需要对流程持续进行改进与提高。

中国建设银行还在数据质量管理框架中引入了六西格玛(西格玛，英文 Sigma, 是用于 衡量标准误差的统计单位。六西格玛引申为100万个数据里只有3、4个问题)作为数据质 量的衡量标准。采用界定、衡量、分析、改善、控制的六西格玛项目管理方法，经过两代项 目的实施，中国建设银行在全行建立了数据质量管控的组织框架和基本流程，包括数据采集 的规范与标准、量化的数据质量监测流程，以及相配套的组织形式与部门职责分工等，构建 了一个具有动态、有效、发展的数据仓库的数据质量治理体系。

**3.** **中国农业银行数据质量管理**

中国农业银行以中国银行业监督管理委员会《银行监管统计数据质量管理良好标准 (试行)》为标杆，采取多项措施来全面加强数据质量管理。

**(1)加强组织机构建设** 中国农业银行将数据质量管理纳入内审稽核体系和内控评价 体系，开展了监管统计数据质量管理专项审计，并督促相关部门落实整改。

**(2)完善报表业务及管理制度，落实部门数据质量管控职责** 中国农业银行共完成了 1104报表6000余项指标加工规则的整理工作，制定了客户风险监测基础报表实施细则，逐 条细化每项客户风险数据的业务规则、部门分工与报送流程。

**(3)深入加强系统支撑** 中国农业银行通过建设大数据平台(数据仓库)及信息报告 系统，有力支撑了数据质量的提高。

**(4)不断完善数据质量监控、检查与考核评价体系** 中国农业银行不断完善数据质量 监控、检查与考核评价体系，形成了统分结合、常态化的数据质量监控体系，并定期开展监 管统计现场检查和数据质量考核。



**第8章**

**数据生命周期管理**

商业银行积累了海量数据信息，但不是所有的数据都具有相同的使用价值，随着时间的 推移，有些数据被频繁访问或更新，而有些数据则很少被访问。快速产生的、大量的、低价 值的数据不仅严重消耗了系统资源，拖累了系统处理性能，还可能影响系统的稳定性而导致 未知的系统风险。事实上，数据同万物一样，是一种具有生命周期的资源。对数据进行生命 周期管理，研究银行客户行为规律和金融业务需求，已经成为金融发展的趋势和方向。

由于银行拥有海量的数据，因此对数据进行生命周期管理也成了银行的一大挑战， 例如：

1)众多数据的生命周期管理策略如何管理?

2)清理后的数据如何提供访问?如何存储?

3)数据保留时间如何确定?哪些数据可以长时间保留?

数据生命周期管理的出发点与目标是优化应用存储结构，有效控制在线数据规模，提高 生产数据访问效率，减少系统资源浪费，提高应用系统运行的整体效率和效果，保证系统健 康高效运行。其管理内容是应用系统的数据在生产应用系统中的在线阶段、归档阶段、销毁 阶段的单一流向过程。为完成数据生命周期的全流程管理，必须从以下方面设计整体的架构 体系。

1)健全相关技术规范体系，包括规范、手册、指南等一系列规章制度。这是指导数据 生命周期管理的基本依据。 一般来说，在规范中明确数据生命周期管理过程中在线、归档等 各个环节的保留周期、清理策略、存储方式等具体要求，在手册或者指南中明确相应的操作 方法等。另外，有些商业银行也会制定管理办法等规章制度，用于具体明确相关职位的职 责，以及在数据生命周期管理方面的具体考核办法等。

2)对数据生命周期管理策略信息实施统一管理，实现各数据表管理策略的登记，包含 但不限于清理方式、保留周期、清理周期、存储介质、归档方式、销毁方式等。

3)建设实施数据生命周期管理策略的程序或工具，负责按照策略对数据进行清理、转 移、归档等动作。

4)制定评价数据生命周期策略及其实施效果的评价体系，对数据生命周期管理策略合 理性及实施情况进行合理评价。

 银行数据治理

数据访问频率

**8.1** **数据生命周期管理及意义**

数据的生命周期管理最初是由美国的存储技术公司 StorageTek 提出来的，经过这些年的 发展和宣传，目前逐渐开始进入市场化阶段。数据的生命周期管理作为一项企业信息化战 略，是一项相对长期的工作。

早期银行的金融产品较少，主要以存贷汇为主，而且数据以省(市)为单位分布存放， 数据量相对较少。此外，由于还没有数据集中挖掘与分析的需求，日常处理对象主要是短期 数据，时间跨度小，处理的数据规模相对可控。但是，随着各家银行的数据大集中以及业务 的快速发展，客户群体越来越大，客户发生的交易行为也越来越频繁，尤其是在大数据、互 联网金融时代，银行业集中存放和处理的数据量急剧增加，有些银行的数据容量甚至达到了 PB 级。有种说法是：“今天一个现代人一天所吸收的信息，比莎士比亚一生所获得的信息 还要多。”由此可见，当今社会信息量是呈爆发式增长的。对银行来说， 一方面数据成为银 行的重要资产，对于分析客户行为、评价客户信用、提升营销精准度等方面都起到了重要的 作用；但是，另一方面日益增大的数据量对系统存储、访问效率造成越来越大的压力。因 此，必须权衡效率和需求之间的关系，合理保留数据。

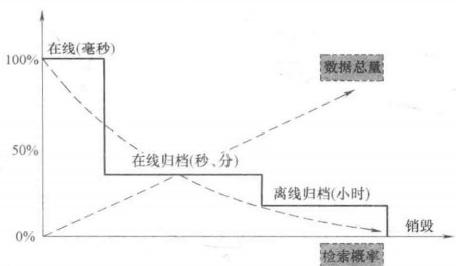
正常情况下，数据生成时间越长，其获得的访问频率则越低。但是，数据总量越来越 大，如果将不活跃的数据进行迁移，则可以降低企业的拥有成本。

数据生命周期中数据生成时间、数据访问频率、数据总量等参数间的关系如图8-1

所示。

倘若不对数据进行有效的数据 生命周期管理，则会出现系统问 题。例如，数据增长过快导致的系 统资源浪费；又如，表数据量过大 导致访问效率过慢等情况。以某商 业银行为例，早期因为没有进行数 据生命周期管理，出现了各种问 题，极大地影响了系统服务效率和 客户服务水平。从以下两个例子可 以看出数据生命周期管理的重 要性。

案例一：某商业银行电子影像



数据生成时间

图8-1 数据生命周期中各参数间的关系

平台采用了外部产品，基本上有史以来的影像数据都存储在系统内，造成系统内数据量庞 大，某些分行的数据量达到几十太字节 (TB), 平时的数据备份时效非常缓慢。2013年某 天系统出现问题，数据恢复效率很低，其间发生的业务数据出现了丢失现象，只能重新扫 描。从此案例可以看出，没有好的数据生命周期管理机制，会造成系统中无用或者使用价值 很低的数据不能及时清理，不仅会对系统资源造成浪费，而且会使得在出现紧急情况时无法 应对。

**数据生命周期管理**| **第8章**

案例二：某商业银行信贷系统自投产以来，由于监管部门和业务部门要求，在合同未到 期之前，必须能够提供相关借据的查询，因此合同、借据等数据一直存在系统中的同一张表 中，随着时间的推移，该表中数据量越来越大，造成了查询瓶颈。为此，此银行将该表拆分 为当前表和历史表，这样不仅提高了当前表的查询效率，而且使历史数据也同样满足了监管 和业务要求。从这个例子可以看出，合理的数据生命周期管理策略，可以提升内外部的服务 水平。

数据生命周期是以成本与数据不断变化的价值相对应的方法，从数据的创建到最终处置 对数据进行全程管理，从而可以使企业根据业务需求、监管、法律等方面的要求对数据进行 分类、存档、恢复、删除。因此，在满足业务部门、客户的合理需求、监管部门的监管要求 的前提下，能够有效地控制在线数据规模，不断提高生产数据的访问效率，减少系统资源浪 费并合理解决成本，保证应用系统健康高效地运行，成了大型商业银行的挑战，同时也是整 个企业数据架构的重要组成部分。

**8.2** **银行业数据生命周期管理的现状**

商业银行普遍拥有众多的业务系统，数据种类也很复杂，数据规模急剧膨胀，业务系统 生产数据库存放了大量的历史数据，这些数据从系统上线以来基本保存在数据库里面，因而 其规模呈线性增长，容量越来越庞大，面临很多问题。

1)数据库的体积太大，最直接的体现就是业务数据库表容量较大，从而导致系统访问 性能、响应速度下降。根据性能容量测试情况，当主机数据库的配置、基础环境相同，而基 础数据量增加到3倍时，响应时间就会增加到4倍，个别业务预期指标已无法满足。部分小 型商业银行目前尚未意识到这种问题，但大型商业银行已经开始采取各种措施进行应对。

2)数据库表容量太大，给后台的维护带来困难，导致效率低下，索引优化、统计分析 等后台维护时需要较长的维护窗口。例如，银行的核心系统每天需要备份大量的生产数据， 随着数据量的不断增大，备份时间和日结工作时间也越来越长，数据库锁表风险也随之 变大。

3)一旦发生系统故障或误操作，则数据库的恢复将非常耗时，对整个业务支撑系统来 说可能是致命的。从前面的例子可以看出，当数据恢复时间窗口变得越来越大时，对于业务 的进行会造成较大的影响，甚至有可能出现业务无法操作的情况。

4)所有数据都集中存储在高端存储设备中，这样随着数据的不断增多，扩容带来的投 资成本也会不断上升。虽然存储设备的价格呈现不断下降的趋势，但是在大数据时代，数据 量爆发式增长的速度已经超出了价格的下降速度，扩容成本也成为银行不断面临的难题。

业界不同厂商为解决不同时期、不同利用价值的数据管理的难题，提出了各种数据生命 周期管理策略和解决方案。这里可以简单介绍一下不同厂商的数据生命周期管理策略和相应 的解决方案。

IBM公司在很早之前便提出了“信息生命周期管理”的概念，并随之推出了各种集成 与存储的解决方案，从数据创建到最终处置的生命周期进行管理。例如，1974年， IBM公 司推出的大型机分层存储管理 (HSM) 软件就是数据生命周期管理的最早形态。

 **银行数据治理**

IBM公司认为，每个企业在系统中至少有30%以上的数据是过期或者非活动状态的， 但是至少有80%的数据要保留很长一段时间，用于长期数据的可访问性，因此信息保留的 需求与归档架构应该相对应。IBM公司通过建立分层的存储硬件结构和数据保护，根据数据 的类型，自动将不同生命周期阶段的数据存放在最合适的存储设备上，实现了数据分类、备 份、发布、迁移、归档、删除的全面自动化管理。在存储介质方面，由于SATA 磁盘比 PC 磁盘更加便宜，可以作为近线(在线归档)数据的保留介质，而磁带、光盘则比磁盘便宜， 更加节约能源消耗，但不适宜用于在线或近线， 一般考虑将其作为离线归档的存储介质，从 而延长数据可获取的时间。

ORACLE 公司认为生命周期管理是“在信息的整个生命过程中，将信息的商业价值和 最适当、最有成本效益的 IT 基础设施相匹配的策略、步骤、实践方法和工具”。从这方面来 看，其实大家对数据生命周期管理的认知基本大同小异。同样，ORACLE、日立数据等公司 都以不同的存储介质实现数据层次化迁移、归档、保护等，以实现高效、低成本、不同访问 边界的数据生命周期管理体系，从而解决了成本问题，满足了法律法规和审计的要求，减少 了信息的管理风险，更为重要的是提高了业务的连续性和服务水平。

我国各银行或多或少地借鉴了厂商的一些经验，使用或自行探索银行数据的生命周期管 理体系，并已取得了一定的成果。

例如，中国工商银行在2006年即开始探索数据生命周期管理办法，至2009年基本建立 了数据生命周期管理体系，从策略管理、清理方式统一、清理信息展现等方面均进行了积极 的探讨，并在2008年获得了由中国人民银行开设的软科学技术一等奖。

**8.3** **建立数据生命周期管理体系**

**1.** **生命周期管理的定义及阶段**

在此先以一个例子看看银行的数据是如何流动的。

一笔交易数据从客户填单开始就诞生了(在当今大数据时代，甚至客户在打开网上银 行，或者踏入网点时数据就开始诞生了),此时的数据拥有的价值较高，许多相关部门的人 员都要对数据进行存取和处理(例如用于风险分析)。在交易完成以后，该笔数据的价值开 始逐渐下降，此时将它转存到低成本的存储介质中以节约成本。而当该笔业务发生后续服务 问题时，譬如质量、咨询和改进等，此时又需要该条信息的内容，把它重新激活，提取到高 效设备中。随着质量保证期期满，这一信息的价值又重新下降，直到一定时间以后，退出它 的生命周期。

除了关注数据在其生命周期各阶段中的不同价值，数据本身的内容同样十分重要。特别 是在当前的经营环境下，数据需要以更加认真的处理方式来对待。内容可以说是数据存储位 置和方式的又一决定性的因素。举例来说，信息技术人员在一段时间结束以后，是不能随便 清除所有电子邮件文档的。其中的垃圾邮件，当然可以马上删除，但是那些属于业务交易、 人力资源和财务等重要文档却应当被保留，用于进一步的分类、储存和保护，以便在更长时 间内进行使用。同时，由于这些数据可能在一定时期内会有人查阅，所以它也不应该被简单 地移动到备份系统中。在这里，关键的问题是识别的流程、制度规定以及方便实用的技术支

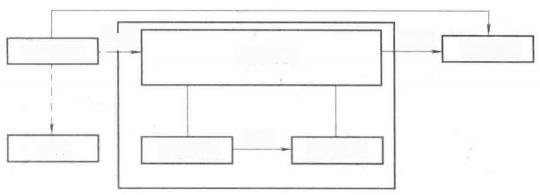
**数据生命周期管理**| **第8章**

持能力。

从成本考虑，存储设备可以分为高等、中等、低等各不相同的硬件平台，对应重要程度 不同的、结构不同的数据，然后还要选择相应的软件。

由此可以得出数据生命周期管理的简单定义，即数据生命周期管理 (Data Life Cycle Management,DLM) 是对不同的业务数据进行贯穿其整个生命周期的管理，通过完整的数据 生命周期管理解决方案，让不同类型的数据存放在适合的存储设备上，利用适当的技术手段 对这些数据进行处理和分析。这样，用户将可以提高现有存储设备的利用率，利用自动化的 数据管理技术实现自动的数据管理，减少银行的信息技术管理成本，满足监管和业务的数据 保管和管理的要求。从这点来说，数据生命周期管理不是一个具体的产品，而是一个管理理 念，更强调从企业级别的视角，对数据进行更有效的管理、更有价值的开发和利用的全 过程。

一个完整的数据生命周期管理阶段及解决方案如图8-2所示。



销毁 销毁阶段

在线阶段

离线归档

归档阶段

在线归档



转移

备份

图8-2 数据生命周期管理阶段及解决方案

具体内容如下：

**(1)数据的在线阶段** 此阶段为数据产生与使用的阶段，其保留价值、存储设备、用 户访问频率、安全保护等要求都较高，因此解决方案应综合考虑用户访问、业务查询、法律 法规等各种因素，其数据的保存时间要能维持应用系统正常运转以及维系业务活动正常开 展。简单地说，这个阶段数据要能够提供业务开展和查询要求。比如客户开立的账户信息、 留给银行的客户信息等，只要这个客户在银行具有资产，就认为其数据处于在线阶段。又 如，客户进行的交易明细， 一般来说，短期内(如3个月内)查询概率较高，而较长时间 内(如一年前)查询概率较低，因此可以认为短期的明细处于在线节点。在线阶段要考虑 的因素比较多，包括数据的保留时间、存储位置、存储设备要求、访问机制、数据安全和高 可用措施、是否需要下一阶段的生命周期管理等。具体来讲，可以包括以下内容：

1)保留时间。数据在这个阶段保留的时间长度。在实际情况下，部分数据是可以直接 确定保留时间的。例如，销户的客户数据一般就不会提供实时的在线查询了，因此可以考虑 销户后保留一段时间即可进入下一阶段。再如客户的交易明细，由于数据量很大，很多银行 考虑将明细分成两部分存储： 一部分在核心的交易系统上，这部分时间较短，例如中国农业 银行的个人存款明细在主机上仅保留一个月，中国工商银行各类明细的保留时间则从3个月 到2年不等；另一部分特殊数据需要根据数据状态、关联数据要求等条件才能确定具体的保 留时间，例如某银行对于借据的保留时间是根据合同来确定的，只要合同没到期借据就要一 直保留。 一般来说，在线阶段数据的保留时间主要是出于保证系统高效运行、节省系统资源 的目的而提出的时间。此时在设计数据时就应该考虑以什么样的形式来确定数据的保留时

 **银行数据治理**

间， 一般都是以数据的时间(包括生成时间、访问时间等),辅以状态(如销户、废止等状 态)来进行确定的。因此，数据表都会包括一种或多种这样的字段来支撑。

2)存储位置。 一般来说，这个阶段数据的存储位置位于数据产生的系统内。但是，也 会出现一些比较特殊的情况，比如某商业银行将所有的业务影像进行统一管理，虽然客户开 户时的证件影像是通过个人交易系统产生的，但是，实际存储在影像管理系统内。因此，这 一要素，各商业银行是根据各自业务和系统特点来自行决定的。

3)存储设备要求。这个阶段因为要维系业务高效运转，因此存储设备性能好、吞吐量 大、访问速度快， 一般会采用高端磁盘阵列设备。但是，如果不是针对客户的业务系统，此 要求一般会降低，可以考虑不同类型的存储设备，比如低端磁盘阵列设备。

4)数据保护措施。同样，因本阶段的重要性，为了防止数据被破坏，各商业银行会从 硬件和软件、系统级和应用级两方面对系统数据进行保护，包括磁盘阵列 (RAID) 、 镜像、 系统或应用备份、灾备等措施。

5)考虑下一阶段的策略。当在线阶段时间到了之后，则需要根据数据类型、查询需 求、系统容量等因素，决定是否进行归档或者销毁。不同银行策略不一，部分城市商业银行 由于数据量少，对系统压力不大，目前还未考虑进行归档或者销毁数据，而大型商业银行出 于成本考虑，会将还有查询要求的数据从核心系统迁移出来进行归档处理，供后续查询。

**(2)数据的归档阶段** 此阶段是指在系统资源受限的情况下，为保持应用系统的健康 安全运行，将访问频率较低且应用需求不高的数据从应用系统中分离出来，保存在另外的磁 盘或应用系统上，以备后续查询。

未必所有的数据都会有归档阶段， 一是对于系统压力不大的银行，若不对数据进行清 理，则不会产生数据的归档阶段；二是对于一直有效的数据，也不会具有归档阶段，例如一 直未销户的客户信息；三是对于到期后无用的数据，也不会具有归档阶段，例如临时加载使 用的信息，确认使用完成后即失去使用价值，那么就可以直接跨过这个阶段进行销毁了。

归档阶段是为了能够备查，因此要考虑到监管和客户需求，必要时要延长归档的保留时 间。其管理要素基本与在线阶段一致，但由于其查询频率较低，可以利用比较低廉的存储介 质来支撑这个阶段。但还要考虑以下内容：

1)归档方式。可以根据数据的特性选择不同的归档方式。例如，对于客户的历史交易 明细，由于还需要提供给客户进行服务，基本上所有银行选择以应用方式进行归档，包括在 原系统内建立历史表，或者新建一个专门用于查询的系统。前者称为历史表归档，后者称为 应用归档，有些银行将两种方式都称为近线阶段。如果数据基本上无查询要求，较多银行则 采用磁带、光盘、移动磁盘等相对独立的存储介质，以备不时之需。

2)归档要素与存储结构。对于不同的归档方式，其归档要素和存储结构也不一样。例 如，如果采用历史表或者应用来进行归档，其元数据已涵盖在应用系统内，并且采用数据为 结构化的形式，可以提供快速查询功能；但是，如果采用磁带、光盘等介质，极有可能采用 文本方式，则在归档数据时要将结构定义、数据字典等信息一并归档，否则容易陷入无法查 询的困境。

**(3)数据的销毁阶段** 此阶段是指数据到期后(包括在线阶段、归档阶段)进行销毁 的过程。以人工方式销毁的数据，为保证后续审计需要，应对销毁内容、时间、方式、核准 部门及人员等信息进行登记。

数据生命周期管理|第8章

2.数据生命周期管理模型

前文已经划分了生命周期的各个阶段，并明确了相关的注意事项，下面将介绍数据生命 周期的具体管理模型。

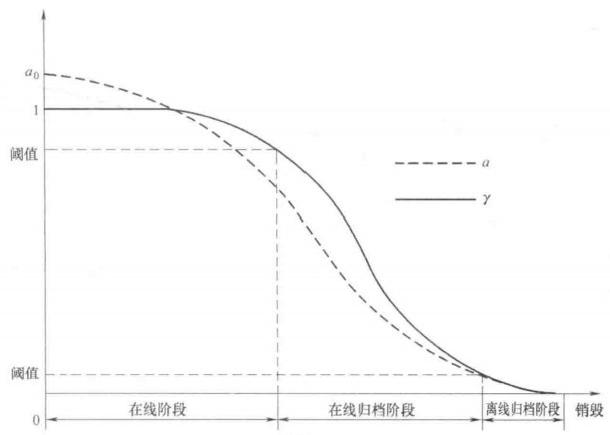
(1)数据在线度 随着时间的推移，新数据不断产生，历史数据使用频率不断下降， 使用价值也相应衰减到较低水平，但这个阶段的划分是相对的，在线的定义也是定性的。为 了更好地体现数据的生命周期，需要定量描述数据所处的各阶段，使用数据在线度 (Data On-line Level,DOL)γ准确反应数据的在线程度和使用衰减情况。数据在线度与用户对数 据的访问频度 a 相关。 a 为单位时间内的访问次数，与数据的迁移安全性、代价、必要性有

关。假设迁移因子为m, 则关联函数为γ,=f(a,m₁), 其 中t 为时间。

在新生数据被频繁访问期间，即在线阶段，此时γ=1。随着时间的推移，访问频率发 生变化，数据在线度随之加以修正，直至达到某阈值，数据即处于归档阶段(包括在线归 档和离线归档)。此后受资源所限或其他因素所致，可考虑进行数据销毁。当γ趋近于0或 达到某阈值时，可对数据进行离线归档处理。当数据生命周期进入归档阶段时，需要对数据 进行迁移或从系统中将数据分离出来。但也有部分业务系统，因设备老化、技术限制或迁移 难度大等原因，无法实现对访问频率下降的数据进行迁移。同时也存在一些数据虽然访问频 率下降，但是在一段时间内仍需维持在线或在线归档状态，或者借助新技术在满足业务系统 正常运转的前提下提前对数据进行迁移的情况。迁移因子影响了数据在线度，对数据生命周 期阶段划分具有阻碍或促进作用。综上所述，在线度关联函数如下：



其 中 ，a₀ 为初始上线频繁访问时平均访问次数，且O≤m,≤1,t 为时间，可根据数据及 系统需求，将单位设为年(或月、日)等。图8-3反映了数据在线度与数据生命周期之间的 关系。



时间 t

图8-3 数据生命周期管理阶段模型

 **银行数据治理**

**(2)数据生命周期的管理驱动力** 系统的建设是为了实现业务需求，商业银行要成功 实施数据生命周期管理，必须关注业务系统的特征属性，制定有效的管理策略。 一般而言， 数据生命周期管理策略应遵循以下原则：

1)根据业务活动响应要求与连续性，确保与关键业务流程相匹配，与系统运行目标保 持一致。

2)满足管理制度和行业标准的要求。

3)确保关键数据的安全。

4)依据数据的价值高低分配适当的存储设备，使重要的数据得到足够的存储资源，同 时兼顾一般数据的存储成本。

数据管理驱动力为 DMDF(Data Management Driving Force), 业务因子s 为不同业务活 动所需响应时间、业务连续性要求等，制度因子r 为各项规范标准，价值因子v 为数据价值 及安全性需求。数据管理驱动力随数据生命周期数据在线度γ的变化而变化，则相互关 系如下：

DMDF=f(s,v,r,y)=min{1,svy+g(r)}

参数s 、r 、v 及目标函数与对应的业务系统相关，且O≤s,v,y≤1,g(r) 标识行业标 准要求。根据不同大小的数据管理驱动力，可以有针对性地构建合理的分层存储硬件环境， 提高现有系统处理效率和存储资源利用率，满足不同类型的业务数据在不同生命周期阶段的 存储要求。同时，将有效的业务数据分离出来，加以分类并制定不同的管理策略；将无用的 数据及时加以清理，提高数据的安全性和系统的可靠性与高效性。

**3.** **数据生命周期管理实施方案与流程**

如何将数据的生命周期管理落到实处?如何兼顾监管和业务需求，制定合理的生命周期 管理策略?又如何实时掌握真实的执行情况?这些都属于整个实现流程所需要考虑的。以某 商业银行为例， 一般从以下流程实现数据生命周期的管理和落实。

**(1)数据生命周期管理需求** 商业银行对业务系统实施数据生命周期管理的驱动力来 自两方面：

1)现有系统已受到或已预测到数据激增对系统性能、维护成本、数据安全等产生的负 面影响。

2)新建系统时提前对数据的生命周期进行考虑。

无论改造还是新建系统，数据生命周期管理的实施需综合考虑业务人员的业务需求、维 护人员的维护经验、法律法规的规定要求、数据增长的定量分析和项目实施的软硬件成本 等，要在数据生命周期各阶段明确数据管理的内在需求，实施有效的数据管理方案。

**(2)计算业务系统数据管理驱动力**

1)预测系统数据生命周期。根据数据生命周期管理模型，预测分析相关业务系统数据 访问频率、迁移情况，确定数据在线度γ。

2)确定业务活动类别。商业银行业务系统按照业务类别、响应时间要求、重要程度， 可以将数据分为不同类别。例如，某商业银行按照应用分层分组的原则对数据进行分类，其 中的度量是指业务因子s, 取行业经验值，如对于核心系统，其度量值取最大值1,而办公 系统的度量值相对较小，可以取值为0.1或0.2,见表8-1。

**数据生命周期管理** **第8章**

**表8-1** **某商业银行数据分类情况(1)**

|  |  |  |  |
| --- | --- | --- | --- |
| 业务类别 | | 说 明 | 度量值 |
| 业务处理类 | 产品服务 | 包括面向客户、涉及账务处理且时效要求极高的业务处理系统，如存贷系统 | 1 |
| 市场营销 | 为营销人员提供营销策略、营销对象信息的系统 | 0.6 |
| … | …… |  |
| 决策支持 | 办公管理 | 满足日常办公需求，时效较低的业务系统，如公文、邮件等 | 0.2 |
|  |  |  |
|  |  |  |  |

3)确定业务系统数据价值。不同业务系统包含的数据种类各不相同，并且具有不同的 价值和安全性需求。例如，某商业银行根据主题和功能特点对数据进行二维分类,见表 8-2。业务数据的价值因子 v 按照最高类别确定，取行业经验值。

**表8-2** **某商业银行数据分类情况(2)**

|  |  |  |  |
| --- | --- | --- | --- |
| 数据类别 | | 说 明 | 度量值 |
| 当事人 | 基础数据 | 用于记录当事人，包括法人、机构、客户的相关信息，如客户信息等 | 1 |
| 日志 | 系统或用户操作过程中产生的日志信息 | 0.2 |
|  |  |  |
| 事件 | 基础数据 | 用于记录事件发生的主体信息，如客户的交易明细等 | 1 |
| 参数 | 系统内的字典表、公共代码等，如操作码 | 0.5 |
|  |  |  |
|  |  |  |  |

4)分析制度管理需求。根据行业规范(不同业务对数据保存期限要求不同)确定存储 及备份要求，有些重要数据即使已处于归档状态，也需较高的资源配置来确保可用性，如采 用在线归档。

5)确定数据管理驱动力。根据模型分析，确定数据管理驱动力 DMDF为f(s,v,r,

y) 。 例如，商业银行核心系统有较高的时效性和业务连续性要求，客户账户数据属于敏感 信息，除销户外，数据在线度为1,该系统具有最高的数据管理驱动力，值为1。此外，核 心系统中包括了不同的数据类别，应分别计算各分类数据管理驱动力，为精细化数据管理提 供依据，如交易明细数据，随着时间增加，该类历史数据访问频率降低，业务响应时间要求 较低，可参照分类数据管理驱动力从系统中分离出来进行分表或分库处理。

**(3)数据生命周期管理方案** 数据生命周期管理方案一般包括数据存储方案、数据迁 移策略、数据访问策略等。数据存储方案中应至少明确数据存储期限、数据存储容量、数据 存储介质要求。数据存储策略为分层保存策略。

1)新生成的数据和使用频率较高的数据(数据在线度较高的数据)应在线存储，而且 使用高端的存储设备，以保障系统的性能。

2)对于生成时间较久而且访问频率较低的数据，则应将其迁移到成本相对较低的中高 端存储设备中。

3)对于使用较少但仍有保存价值的数据，则应将其迁移到成本较低的中低端存储设

**银行数据治理**

备中。

在确定数据存储策略之后，还需要明确数据的迁移策略：明确需要迁移的数据、迁移的 目标位置、迁移频率、迁移时间点以及迁移方式(增量或全量)。数据迁移应在不影响系统 正常运行的前提下进行，迁移前需要进行可行性认证和演练，迁移后的数据应支持反向迁移 至上一阶段。数据迁移应保障数据的完整性、可用性和一致性，整个数据迁移过程的操作应 被记录并可查询。

在数据迁移完成以后，数据访问方式应明确对在线和归档阶段数据访问的技术实现方 式。 一般情况下，业务系统应为用户提供对在线和在线归档阶段数据的透明访问。

**(4)优化调整** 随着业务的发展，系统的业务数据类型和业务规范不断变化。因此， 一个完备的数据生命周期管理方案，还必须包括对系统的优化调整，即根据业务系统实际运 行情况、新的业务类型和规范、技术进步等调整原有的存储迁移策略和访问方式，以满足业 务系统发展及新的业务规范需求。在商业银行核心业务系统中存储的客户相关交易数据随着 时间的迁移，客户可能不再要求或很少要求获取较早的历史交易记录，但此类数据有较高的 价值以及保密和监管要求，对此类数据，各阶段生命周期管理方案要进行优化调整。

**(5)工具支撑** 在实施阶段，必须要有统一的工具或者执行程序，在工具或程序中会 定义每张表的具体数据生命周期管理策略，包括进行清理的依据、清理什么样的数据、清理 后是该归档还是直接销毁等，否则各种各样的清理格局是会失控的，尤其在动辄几万甚至几 十万个表的情况下(银行尤是如此)。

**(6)信息展现** 在后续检查阶段，必须要有一个系统来体现具体的情况效果，可以定 义以下要素，即什么时间、清理了什么数据、清理效果如何。

当前大型商业银行出于成本考虑，会对数据进行清理，从而考虑以上因素。例如，中国 工商银行编制了统一的清理程序、统一的清理策略管理系统、统一的清理信息展现系统，而 部分中小银行在目前系统承压可控的情况下，尚未考虑数据清理工作。例如，大连银行总体 数据量尚未达到TB 级，目前数据是一直保留的。但随着业务的发展、历史数据的沉淀和积 累，这项工作必须提上日程。

**4.** **数据生命周期管理的对象**

数据生命周期管理的对象比较多。 一般来讲，针对数据对象或使用方式的不同，将数据 生命周期管理对象分为结构化数据、非结构化数据、备份数据。

**(1)结构化数据** 结构化数据具有一定结构，可以划分出固定的基本组成要素，能以 表格形式进行表达，可以用数据库的表、视图等进行表示， 一般是数据库表中的一行，例如 一条条的交易明细。

**(2)非结构化数据** 非结构化数据一般定义为无法使用数据库二维逻辑表来表现的数 据，包括所有格式的办公文档、图片、可扩展标记语言 (XML) 、 超文本标记语言 (HT- ML) 、各类报表、图像和音频/视频信息等。随着大数据时代的来临，非结构化数据越来越 庞大，在银行表现的有客户的影像(如身份证复印件)、内部的资料(如规章制度)、用于 培训和宣传的影音等，都在非结构化数据范围之内。

**(3)备份数据** 备份数据是指为确保应用系统连续运行，防止数据被破坏，并考虑应 用系统的恢复要求，满足法律法规及监管需求，定期对应用系统在线阶段数据进行的快照。 一般来说，它分为两种情况： 一种是为保障系统而做的备份；另一种是指满足内部与外部审

**数据生命周期管理**|**第8章**|

计检查、用户与业务特殊要求等对历史数据提取、查询的需要，以及满足法律法规及监管需 求所做的备份。

**5.明确生命周期管理细节**

在明确了什么是数据生命周期管理，以及具有什么样的阶段后，需明确在数据生命周期 管理的方方面面应该有什么样的具体措施和规定来有条不紊地推动执行。

**(1)应用和数据分类** 银行开展的业务繁多，同时内部管理、外部监管的要求也是多 种多样的。对于银行来说，不同业务场景、不同监管需求、不同决策支持要求会区分或多或 少的应用层级，且金融业务数据模型 (FSDM) 在我国银行业基本进行了采用。因此，可先 将应用、数据划分成不同层级、不同类别，针对不同层级和不同类别的数据制定相应的原 则，化整为零后可以得到有效执行。纵观各大银行，事件、当事人、协议、账户等主题一般 通用，而对外客户服务的如个人存款、贷款、理财等业务种类也大同小异。因此，可以先按 业务类型(如个人存款、对公存款等)划分，再按主题划分。如果可以将主题再进行细分， 或在此之下再根据数据的特点进行细分，相当于将数据的类型做更小类的聚类，其共同特征 将会更为明显，则更容易制定每类数据的具体策略。例如，某商业银行按照应用层级、应用 组级、主题、特点四个维度将其数据划分为1万多类。

**(2)数据清理的方式** 根据数据的特点、实际应用场景，对不同数据进行生命周期转 移时采取的手段是不一样的。参照某银行， 一般可以分为以下几种：

1)数据更新方式。这种方式一般是对数据直接进行更新，并没有数据迁移的动作，例 如利率表的变化，又如某个地区名称的变化，即指随时或不定期保持最新数据，可以理解为 表中删除了一条数据，又在相同位置新增了一条数据。按这种理解方式，部分仅用于数据加 载的临时表也属于这种清理方式。

2)前台操作界面删除方式。这种方式一般由客户或系统用户操作，在前台页面将无效 的数据直接删除。例如，客户原住址房产卖了，若已确认这条联系地址作废，则可以主动在 客户联系信息里删除，也可以通过客服，让系统用户进行删除。虽然在前台页面做了删除动 作，但是在系统内部有可能执行不同的处理方式。 一种是真实删除了，例如机构撤并后某个 机构不存在了；另一种是通过逻辑方式删除，即改变这条信息的状态，比如客户的联络信 息，可能会分析客户住址变动历史，将此条信息仅进行逻辑删除，后续再通过批量程序进行 清理迁移。

3)批量程序清理迁移方式。此种方式较为常用。银行一般会考虑联机时刻的性能，将 数据清理迁移这种消耗资源较大的动作放在系统闲暇时段(如凌晨两点),通过批量操作将 已成为历史的数据从在线阶段迁移到归档阶段或者直接予以销毁。

**(3)确定数据清理的频率** 确定好数据可以保留多长时间后，还必须明确另外一件事 情，即多长时间清理一次，是不是可以任意确定清理周期?

举个例子，客户的交易明细，如果和业务部门协商后，明确提供给客户查询两年内的明 细，超过两年的话，对不起，客服帮你查，或者亲自跑趟柜台查询。这样就可以知道交易明细 保留两年，那多久清理一次呢?如果能够确定一个时间，即清理周期和保留时间相同，也是两 年，那么对不起，其实在系统中真实的保留时间有可能达到四年。请详细看一看下面的解释。

2012年1月1日开始清理两年前的数据，那么系统中保留了从2010年1月1 日开始的 数据，按照两年的清理周期，下次清理日期则是2014年1 月1日。这样，系统中保留的数

 **银行数据治理**

据已经达到4年了。因此，可以看出，真实的数据保留时间为既定的数据保留时间加上清理 周期。因此，清理周期不宜过长，否则就脱离初衷了。可以制定一个规则，比如保留时间在 一年之内的数据，最少一月清理一次；如果保留周期超过2年，可以实行半年清理一次。如 果条件允许，可以一年清理一次。这需要根据不同银行、不同业务性质等来综合考虑。

**(4)选择合适的归档方式** 前面介绍了几种归档方式，那么在实际应用中应该用什么 样的归档方式呢?当然可以任性一些，可以将数据一直安稳地保存在系统内，但是在数据量 爆发、成本高居不下的情况下，大多还是选择使用更为低廉成本的归档系统和设备。

1)老化表或者称为历史表归档。这其实不算是一种正式的归档方式。 一般是为了当前 表的高效，会将部分数据拆分出来，以同构的表继续保留在系统中。这种情形主要用在当前 表的访问和处理效率受到影响但历史数据还有较多查询需求的情况下。

2)使用专门用于数据归档的系统。为了确保源系统的高效运行，减轻存储压力，优化 系统结构，多数银行会采用将清理后仍有较高频率查询要求的数据用归档系统来保留，并提 供一些基本服务。目前，商业银行主要实施的对象为交易明细，例如个人的活期明细、基金 交易明细等。有时也会将作废的账户信息、协议信息等也使用这种方法。

3)将离线的介质用于归档。如果数据清理后没有查询要求，或者查询要求较低，可以 考虑利用磁带、光盘等离线介质存储清理数据。

**(5)数据清理注意事项** 数据生命周期管理中清理动作不是简单删除就可以了，必须 要注意以下两点：

1)部分数据不能简单地物理删除，例如涉及数据接收方使用分析的数据，要通过相关 标志位、最后更新时间等字段实现逻辑删除，然后再按数据生命周期管理策略进行物理删 除。只有这样才能避免数据接收方数据不遗漏，否则上游数据删除之后，数据提供方却不知 晓，这样极易出现垃圾数据，造成统计数据错误、业务操作不成功等现象。

2)清理数据时，应该考虑所清理的数据对系统的影响，比如考虑重建索引、重组等相 关事项。尤其像删除分区这种动作，必须考虑分区的重分布。

**(6)境内外数据的分别对待** 由于境内外具有不同的监管措施，因此对于有境外业务 和数据的商业银行，对境外数据需要区别对待。在制定数据生命周期管理策略上，要根据不 同国家的政策和管理要求按地区进行区分，在工具的支持上，也需要能按地区进行设置。

**6.** **建立数据生命周期管理的技术支撑系统**

数据生命周期管理是一项长期而繁杂的工作，如果仅使用台账式登记和人工管理的方 式，其耗费成本将很大，而且容易出现错漏现象，有必要借助工具和系统手段进行管理，包 括清理策略的统一管理、清理信息的统一展现、统一的清理程序或工具等。

**(1)清理策略的统一管理** 每个表在设计初期，即应与需求部门协商此类数据的具体 保留时间，并将其登记在册，这样不仅容易管理，而且可以作为资产加以传承。

登记的做法可以是多样的，简单原始的办法是通过台账管理。较好的做法是通过系统管 理，并辅以审批、核对、审计等手段，可以比较好地保证数据生命周期管理策略的合理性、 正确性。

举个简单例子，在设计账户表时，在与业务部门沟通后，如果认为账户销户后可以保留 半年，那么,当销户状态生效时间为当前时间减去半年时，即可以启动清理程序，对销户的 账户进行清理和归档处理。在进行登记时，要明确此表依据哪个字段信息进行清理，依据条

**数据生命周期管理**| **第8章**|

件是什么(依据账户状态字段和销户时间字段，依据条件为状态为销户，时间为半年前) 用什么样的工具清理(是自己开发，还是使用统一的工具),清理的方式是什么,清理后是 否归档，多长时间清理一次。只有信息完整和准确，才能保证可以与清理工具进行联动实 现，真实展现信息。

**(2)统一的清理程序或工具** 银行数据在初期清理时，各应用系统各自为政，当系统 存储量大了，出现效率低下的情况时，才会开始考虑是不是数据量太大了?是不是该清理 了?一般是哪个大了清理哪个，编制一个个特定的程序开始对某个数表进行生命周期管理， 根本没有一个统一的管理方式，到后来发现到处都是各类程序，业务运行程序、系统优化程 序、数据清理程序混杂在一起，给程序的管理工作带来很大困扰。因此，有必要建立一个统 一的管理程序，并且对这些清理程序进行有效管理。当然，最好的方法是有一个统一的清理 程序或工具。

一般做法分为以下两种：

1)清理程序的统一管理程序。即将各应用系统的清理程序登记在此管理程序内，统一 调度，这样在一定程度上做到了有序可控。

2)采用统一的清理程序或工具。即由一个程序或工具来负责所有数据的清理工作。 一 般地，可以将清理策略统一管理的内容导入或直接发布到此程序或工具中。该程序或工具根 据不同的条件清理不同的数据。此方法对工具的权限要求较高，基本上要有涉及应用的最高 权限。由于涉及部分敏感系统，因此大部分银行是这两种方式混用的状态。

**(3)清理信息的统一展现** 数据清理对减轻系统压力是一项有力的措施。如何评估系 统的数据生命周期管理是否成熟?又如何知道为科技投入降低了多少成本?这些必须用事实 说话，拿数据说话。因此，有必要将每个系统，甚至每张表的清理信息进行获取，包括什么 系统、什么表、什么时候清理了多少数据，必要时还需要加上数据生命周期的策略，以便更 好地评估清理策略的合理性以及实施效果。

**7.非结构化数据的生命周期管理**

对于结构化数据，主要是针对表中的记录进行清理、归档或者销毁，即每条数据记录都 有一定的生命周期。而非结构化数据则类似于影像、文档等，均是一个个独立对象，因此其 生命周期相比于结构化数据存在较大的差异。 一般来说，商业银行会根据数据的类别、特 点、属性等相关信息来确定其生命周期。例如，商业银行将数据进行分类，不同类型的数据 根据版本号、状态、访问时间来确定不同类型数据的生命周期管理策略。商业银行可以将数 据分成资讯数据、规章制度、办公文档、培训视频、宣传影像、论坛数据等各种类型，然后 根据不同类型的数据特性，例如规章制度的版本、论坛帖子的最后访问日期等进行生命周期 的制定。

举个例子，表8-3 为某商业银行非结构化数据的数据类型与保存期限。

**表8-3** **某商业银行非结构化数据的数据类型与保存期限**

|  |  |  |
| --- | --- | --- |
| 数据类型 | 数据场景及范围 | 保存期限 |
| 资讯数据 | 各类资讯信息、新闻、动态等有效时间较短的非结构化数据 | 5 年 |
| 规章制度 | 行内办法、规范、制度、规程、手册、指南等 | 3 年 |
| 行外法律、法规、行业政策等 |

**银行数据治理**

(续)

|  |  |  |
| --- | --- | --- |
| 数据类型 | 数据场景及范围 | 保存期限 |
| 专业知识 | 指能够通过其获得技能、技巧的非结构化数据，主要指案例总结、经验分享等 | 2 年 |
| 办公文档 | 指公文、邮件等用于办公的各类非结构化数据 | 2年 |
| 业务档案 | 指业务办理过程中产生的文档、扫描件等，主要在电子影像及文档服务平台 | 3 年 |
| 语音视频 | 业务办理过程中的语音视频 | 半年 |
| 教育培训使用的课件 | 1年 |
| 宣传或会议保留的音视频 | 半年 |
| 社交数据 | 指各论坛、园地中用户自由提交的非结构化数据 | 3 年 |
| 系统日志 | 指系统运行时记录相关运行信息，不以表的形式存储的数据 | 半年 |

1)对于规章制度等具有版本属性的非结构化数据，当前版本一直保留，上一版本超过 保存期限后需要归档并清理。

2)对于具有状态属性的非结构化数据，无效或超过保存期限后需要归档并清理。

3)对于具有最后访问日期属性的非结构化数据，超过保存期限无访问的数据需要归档 并清理。

4)无以上属性的非结构化数据，从数据生成之日起开始计算保留时间，超过保存期限 后需要归档并清理。

当然，这里仅仅是举例说明，对于不同的商业银行，其做法可能稍有差异，具体的非结 构化数据的数据生命周期管理策略还应综合考虑业务对非结构化数据的在线时间需求、应用 服务器的存储压力、业务系统的活动情况等来综合制定。

**8.4** **案例**

以某商业银行数据生命周期管理为例，该银行建立了完整的数据生命周期管理体系。

1. 规范建设

该商业银行建立了数据生命周期管理的技术规范，并作为全行性规范下发到各直辖机 构。规范对各类术语的定义、管理细节、具体管理措施等进行了规定和明确，每隔一至两年 修订一次，以保持能够跟上业务和系统的发展。每年组织全行培训，宣传数据生命周期管理 的重要性、具体注意细节等，以深入贯彻数据生命周期管理的理念。

此商业银行基于数据的主题、数据的功能以及数据所在的不同业务系统详细规范不同类 别数据的生命周期管理策略，共涉及12000 多类数据，对具体策略的制定和实施提供了很好 的依据。

**2.** **系统与支撑**

此商业银行针对数据生命周期管理的每个环节建立了相关系统和工具，来支撑其落地实 施工作。

**(1)设计工作相关系统** 此商业银行建立了表的元数据管理系统，在进行表结构设计 时，就需要明确其数据生命周期管理策略，并在系统中进行登记。此系统明确了数据生命周

**数据生命周期管理** **第8章**|

期管理的各类要素，包括保留时间、清理周期、清理后的归档方法、清理的条件、清理的方 式等，同时系统还会根据规范的实际要求进行逻辑的检查和控制，若不符合相关逻辑，则在 表设计提交时予以提示，在很大程度上保证策略的合理性。在此商业银行的相关管理制度和 手册中，明确了在对表结构进行设计审批时，要对生命周期管理策略进行关注，使其能够兼 顾到系统性能、业务的连续以及监管的要求，从不同角度找到恰当的融合点。

**(2)数据清理相关工具** 此商业银行针对不同数据库类型的系统开发了不同的清理工 具，如 ORACLE 、DB2 、MySql 等，其原理基本一致。通过将明确的数据生命周期策略形成 清单直接发布到工具中，工具监控每张表当天是否需要清理，如果达到清理条件，则工具会 根据清理条件自动执行清理动作，并生成相应的清理信息。

**(3)版本发布相关工具** 将数据生命周期管理策略自动形成清理程序和脚本，通过版 本工具直接发布到生产部署的清理工具中去，所有的工作都做到自动执行，减少了人工 参与。

**(4)数据归档相关系统** 此商业银行对于清理后的数据，有两个归档系统可以存储： 一个是用于结构化数据的历史数据存储系统，另一个是用于非结构化数据的影像和文档的存 储系统。

两个系统采用分布式数据库、云存储架构等较新的技术，使用廉价的存储设备，并提供 可查询功能。由于归档系统要满足数据在需要的时候能够提供查询，因此归档系统的数据保 留时间将会远远大于在线的保留时间。 一般情况下，为满足监管要求，会在归档系统中将数 据保留10～15年，以满足不同场景的数据备查要求。

**(5)清理信息展现系统** 此商业银行为了准确评估数据生命周期的管理效果，有效推 动数据生命周期管理工作，建设了数据清理信息展现平台。各应用系统会将清理后的清理信 息集中传输给此平台，并进行集中展现，从而实现信息的统一集中管控。

**3.评价与跟踪**

此商业银行将对数据生命周期管理情况进行有效跟踪并做评价。具体来说，首先比对设 计系统中登记的数据生命周期管理策略与生产上是否一致；其次将实际的数据生命周期策略 与规范进行比较，检查是否合理，并将其作为一个计分项记入系统的评价之中。

**4.** **数据生命周期管理执行效果**

通过以上措施，此商业银行数据生命周期管理已形成成熟体系，将全行上百个系统纳入 了统一管理范畴，年清理数据量达2PB 左右。



**第** **9** **章**

**数据分布与存储**

系统内，特别是分布式系统内的数据分布与存储，各银行关注得比较多，而对于企业范 围内的数据分布与存储，各银行一般考虑得比较多的是主副本数据分布，对于其他方面，如 总分行、核心与外围系统的数据分布与存储布局则考虑得较少。

在银行业务快速发展的今天，特别是随着大数据和互联网金融的兴起，如何合理分布和 存储数据，以通过最低的成本投入获得最高的效能，是各商业银行普遍面临的问题。例如， 数据集中了，但不能满足分行的特色需要，分行经常在本地构建自己特色的系统，并且每天 从总行申请获取大量的数据，这些数据的使用从全行角度来看，缺乏有效的管理视图，并存 在系统功能冗余建设的问题(如总行系统功能和分行系统功能的重复);各系统数据冗余量 大，管理成本高且数据经常不一致；核心系统数据量越来越大，核心系统功能日益臃肿，不 能适应产品的灵活创新需求，管理成本越来越高；数据存储缺乏统一指导，各类数据依据自 身需要四处存放，技术产品缺乏统一规划等。

从数据架构角度来看，银行系统布局的核心是数据的布局。数据信息为 IT 系统规划建 设提供了另外一种视角，系统间边界与关系的划分，新老系统间的对照衔接，可以通过对数 据分布与存储的统一规划来布局，通过合理的数据分布与存储来指导系统边界的划分，按照 数据的流转关系来确定系统间的关联关系，构建耦合度合理的系统架构。

对于银行， 一般会关注以下几类数据分布：

**1.** **总分行数据分布**

纵观商业银行 IT 系统的发展史，特别是大型银行， 一般都经历了数据从分布到集中的 过程，通过数据中心的构建，统一存储数据，降低运维管理、设备投入成本。但数据完全集 中也会带来一些问题。我国地域广阔，各省份经济发展水平有高低之分，也导致了各分行会 存在自身的一些特色需求，如当地的监管、当地的大客户营销、当地的内部经营管理策略 等。数据的完全集中不利于这些特色需求灵活、快速实现，应适当地将一些数据部署于分 行，以满足分行特色需要。因此，有必要对数据的总分行分布进行合理规划。

**2.** **主副本数据分布**

在数据中心内部，为满足各类银行业务需要，系统呈多样化，分门别类，如核心账务系 统、后台经营分析和风险管理系统等。这些系统之间存在数据交互和冗余部署，以期达到系 统最优化、用户体验最佳的目的。因此，也需要综合考虑系统间如何进行数据部署，特别是

**数据分布与存储** **第9章**

主副本数据的分布。

**3.** **核心交易系统、外围系统间的数据分布**

银行的核心交易系统主要是指账务处理系统，外围系统一般包括渠道、产品管理、内部 管理、业务辅助支撑、增值类服务等不涉及账务处理的系统。许多银行的核心交易系统部署 在主机上(这里的主机一般为IBM大型机),以求获得相对高效稳定的交易性能；外围系统 则部署在开放平台，以充分利用开放平台的开放性、扩展性以及相对廉价的成本。随着云计 算、分布式技术的飞速发展，如何从数据架构角度将数据在主机和平台间合理部署，目前正 是各个商业银行正在重点突破的课题。

在考虑数据如何合理分布的同时，还需要进一步考虑数据的存储，两者是紧密关联的。 对于银行，依据数据使用特性， 一般将数据划分为交易型、集成型、分析型、历史型等不同 的数据存储区，并规范了各数据存储区的数据范围，各存储区间的数据流向、数据服务及时 效，从而用于指导全行系统的数据架构部署。

做好数据的分布和存储规划， 一方面可以提高数据的使用效能，提升应用系统的服务能 力，另一方面还可以减少银行系统运维成本。现在已经进入大数据时代，随着互联网金融的 兴起，数据呈现出爆发式增长态势，有效地合理存储及布局这些 PB级数据，对银行信息的 挖掘利用和系统建设成本的把控具有重要的意义。

**9.1** **商业银行系统数据存储管理**

**1.结构化数据存储**

**(1)数据存储区的划分** 银行有许多产品，如传统的存贷款业务，需要与客户进行交 易，各银行均有许多的账务、协议、流水等交易数据。这些数据随着时间的推移，逐渐变 “旧”,成为历史数据，而且数据规模逐步增大，但是，由于客户查询及监管等方面的需要， 不能丢，需要采用更为合理的存储方式。另外，为了对整体经营管理进行把控、决策，需要 对数据进行统计分析。因此，可以看到，银行的运营行为不是单纯的一种模式，由此对于产 生的数据，在数据规模、使用频度、使用特性、服务时效等方面，也不尽相同，需要采用不 同的存储策略。

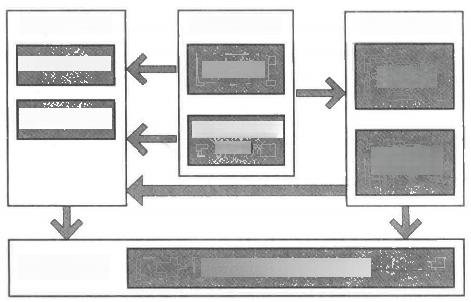
通常情况下，综合数据规模、使用频率、使用特性、服务时效等因素，从存储体系角 度，可以将商业银行的数据存储划分为四类存储区域，即交易型数据区、集成型数据区、分 析型数据区、历史型数据区，如图9-1 所示。虽然银行有大有小，业务也不完全相同，但总 体上其数据存储体系可以被上述四类数据存储区域划分方式所覆盖。

1)交易型数据区。交易型数据区包括渠道接入、交互控制、业务处理、决策支持与管 理等各类联机应用数据；存储客户自助或与银行操作人员在业务交互办理过程中产生的原始 数据的存储，包括业务处理数据、内部管理数据和一些外部数据，其存储的是当前状态 数据。

2)集成型数据区。集成型数据区包括集成的操作型数据和数据仓库模型数据。

① 集成的操作型数据，汇聚了多个操作型业务应用原始数据形成的统一数据存储，是 当前或接近当前的不断变化的数据，数据的详细程度和存储结构接近于操作型业务应用，支

银行数据治理



集成型数据区

数据仓库

操作型数据存侧

(0DS)

报表系统

历史型数据区 结构化、非结构化历史数据

交易型数据区

业务处理数据

内部管理数据

分析型数据区

数据集市

图9- 1 数据存储布局

持跨产品时效性高的汇总统计、周期较短的业务分析、跨应用的数据查询/筛选等功能。

② 数据仓库模型数据，按照数据库三范式格式并按照一定的业务主题(如当事人、协 议、事件等)汇聚保存各操作型应用的历史信息和保留较长时间的明细数据，为后续的决 策支持性应用提供公共数据服务以及汇总计算。

3)分析型数据区。分析型数据主要是用于决策支持与管理的各类集市应用的数据。为 了对业务执行情况进行深入分析，需要对原始数据进行进一步汇总统计和分析，统计分析结 果用于最终的决策展示，因此分析型数据区存储了这些统计、分析模型结构的指标数据。

4)历史型数据区。历史型数据区包括近线应用、归档应用、外部审计数据平台应用等 的数据，主要用于存储管理交易型数据区近线数据、按数据生命周期规范清理的离线数据以 及外部审计平台的历史数据，按操作型数据格式存储，同时，还存储少部分分析型应用的衍 生统计指标，用于满足业务查询、审计、公检法等对离线长期数据查询需要。历史型数据区 的主要作用是为各种历史数据提供归档后的数据保管和查询服务。物理上，可以部署独立的 历史型数据区(一些大型商业银行由于数据规模较大， 一般独立部署),也可以与交易型数 据区整合部署，即相关历史数据继续保留在交易系统中。

**(2)数据存储区的特点** 商业银行各数据存储区的特点见表9-1。

**表9** **-** **1** **商业银行各数据存储区的特点**

|  |  |  |  |
| --- | --- | --- | --- |
| 数据区 | 数据类型 | 数据周期 | 数据加工程度 |
| 交易型  数据区 | 业务处理数据 | 根据业务需要 | 原始数据 |
| 内部管理数据 | 根据业务需要 | 原始数据 |
| 外部数据 | 根据业务需要 | 原始数据 |
| 集成型  数据区 | 集成的操作型数据 | 短期 | 源系统模型、数据清洗、汇总加工 |
| 数据仓库模型数据 | 较长期 | 数据仓库模型、数据整合、数据清 洗、汇总加工 |
| 分析型  数据区 | 集市数据 | 根据业务需要 | 指标、模型化 |
| 挖掘数据 | 根据模型训练需要 | 模型化、抽样 |
| 报表数据 | 根据业务需要 | 指标、模型化 |
| 历史型  数据区 | 结构化归档数据 | 长期 | 原始数据 |
| 非结构化归档数据 | 长期 | 图片等 |

**数据分布与存储** **第9章**

**(3)数据存储支撑的服务**

1)交易型数据区。交易型数据区处于信息收集的最前端。该区的数据旨在对联机交易 进行服务，是满足客户办理业务需要的重要区域。通过应用的渠道接入、交互控制与客户进 行信息交互；通过业务处理的业务逻辑控制与操作，完成对各种业务的交易事务，最终体现 为交易型数据区的具体业务数据。服务主要通过联机接口进行，少数服务通过批量的交易处 理，并通过文件或接口的形式对集成型数据区、分析型数据区进行服务；服务等级高，需要 实时响应，需要高可用性。

2)集成型数据区。集成型数据区处于数据流向的中端。该区会形成信息的全貌，其数 据是对交易型数据区的整合与集成，形成全数据视图的数据平台。服务既包括通过联机接口 查询(如操作型数据存储、客户公共信息平台),也包括通过文件接口的形式批量对分析型 数据区进行服务(如 ODS、数据仓库),少数服务通过动态查询进行(数据仓库)。服务等 级低，不需要实时响应，对高可用性要求较低。

3)分析型数据区。分析型数据区处于数据流向的末端。对信息的统计、分析、挖掘进 行支撑，数据源主要来自集成型数据区，少部分来自交易型数据区。分析型数据区所产生的 数据可回传到交易型数据区。服务主要通过动态查询的形式对分析型应用进行服务；服务等 级中等，不需要实时响应，对高可用性要求中等。

4)历史型数据区。历史型数据区处于数据生命周期的末端。该区主要包括归档数据， 是交易型数据区的过期数据，并根据数据生命周期管理原则加以归档。该区还包括审计数 据，作为备查数据存放。该区主要对操作型数据存储的过期数据提供归档查询服务，并对审 计备查数据进行历史数据存储。服务等级中等，需要准实时响应，对高可用性要求中等。

**(4)数据存储技术** 数据存储技术平台的选择一般要结合数据的重要等级、服务等级、 安全等级、数据类型、数据访问特性(增/删/改/查)等多种因素综合考虑。

1)对于交易型数据，由于属于交易系统，重要且存在大量用户的高并发访问，要求实 时响应并提供7×24h 的服务，因此一般采用面向联机事务处理系统 (OLTP) 的关系型数据 库来实现。从实际情况看，绝大多数银行采用 DB2 及 Oracle 这两种数据库实现交易型数据 的存储。

2)对于集成型数据，该类数据主要服务于银行的决策支持，其重要程度、服务等级相 对较低，但其数据规模相对较大(一些大型银行的集成型数据可达 PB 级),以后台批量运 算为主，所以一般采用大规模并行处理系统 (MPP) 架构的技术平台进行数据的存储管理， 如 IBM DBF、Teradata、Greenplum 等。而随着大数据时代的到来，数据呈现爆发式增长，而 且出现了大量的文本、网页、图像等半结构化和非结构化类型的有分析价值的数据，与之相 应各类技术平台也雨后春笋般地涌现。目前， Hadoop、Spark 是互联网企业运用较多的技 术，各银行也在进行探索和引入，并在不同业务场景中获得了初步的成功。

3)对于分析型数据，主要面向客户关系管理、风险管理、绩效管理等业务领域服务， 侧重于各类业务指标的计算和展现服务，具有批量计算、在线分析挖掘与即时响应等特点 一般采用Oracle 、SAS类产品存储，并与分析挖掘工具、可视化工具结合满足用户的数据分 析挖掘需要。

4)对于历史型数据， 一般区分结构化数据与非结构化数据进行存储技术平台考虑。银 行的结构化历史型数据主要以客户的交易明细为主，需要满足大量客户的并发访问需求，传

 **银行数据治理**

统是采用Oracle 等关系型数据库，采用分库形式(如按历史时间段)提供服务，随着 MPP 类系统、Hadoop 的查询并发响应能力的提高， 一些银行已经采用此类系统进行历史数据存 储，部分银行正在进行技术可行性评估，逐步考虑迁移至该类系统提供服务，减少因分库而 带来的架构及操作复杂性。银行的非结构化历史型数据包括电子影像文档等。此类数据规模 巨大， 一般采用磁盘文件、光盘、数据库(存放索引)等多种技术平台实现该类数据的存 储与查询访问。

**2.** **非结构化数据的采集和存储**

**(1)非结构化数据的采集** 商业银行的非结构化数据主要来源于行内和行外两类途径。 其采集方法主要有以下几种：

1)行内文档信息的采集。对于行内业务系统信息的采集，商业银行一般会有统一的工 具或系统从各类业务系统中获取信息。对于被采集业务系统会提供数据库、文件存储等模块 的只读用户，提供统一的工具或数据采集模块来获取数据，而且采集工具或系统一般仅存储 索引信息，而不存储业务系统信息实体。

2)行内系统日志的采集。很多商业银行都有自己的海量数据采集工具，多用于系统日 志采集，如 Hadoop的 Chukwa,Cloudera 的 Flume,Facebook 的 Seribe 等，这些工具一般都 采用分布式架构，能满足每秒数百兆字节的日志数据采集和传输需求。

3)网络数据的采集。随着网络信息的爆炸式增长，网络已成为全球最大的、最重要的 信息资源共享平台，在网络上每时每刻都在动态刷新的页面中，蕴藏着大量的可用的数据。 未来，数据将被认为是与自然资源、人力资源一样重要的战略资源，隐含着巨大的经济价 值。目前，从网络上采集到的数据多是非结构化和半结构化的数据，如何快速、准确地从中 找到并提取商业银行所需要的信息，是当前最热门的研究课题之一。

无论是通用搜索引擎，还是主题搜索引擎，“网络爬虫”在其中都扮演着重要的角色。 网络爬虫是搜索引擎获取网页的主要工具，即使在企业内部，也是获取企业中各类网页中 信息的重要手段。网络爬虫的搜索策略是网络爬虫的实现关键，是其核心技术。另外，还有 部分网站会公开 API, 通过这种方式也可以从网站上获取数据信息。通过网络爬虫或者 API 方式可以将非结构化数据从网页中抽取出来，将其存储为统一的本地数据文件，并以结构化 的方式存储。它支持图片、音频、视频等文件或附件的采集，附件与正文可以自动关联。除 了网络中包含的内容之外，对于网络流量的采集还可以使用DPI 或 DFI 等带宽管理技术进行 处理。

4)其他数据采集方法。对于银行生产经营数据或科学研究数据等保密性要求比较高的 数据，可以通过与企业或研究机构进行合作，使用特定系统接口等相关方式来采集数据。例 如，商业银行与中国人民银行征信中心关于个人征信相关数据的采集。

**(2)非结构化数据的存储** 商业银行的非结构化数据在早期只是业务发展或项目开展 过程的信息记录，因此一般只要能够提供检索即可，从而对其存储和管理的要求并不是很 高。但是，近年来非结构化数据作为大数据分析的重要支撑数据，对其管理提出了越来越高 的要求，因此有必要针对不同情况的数据或应用场景来制定不同的存储策略，并提供具有针 对性的管理办法。

早前商业银行非结构化数据的存储有点类似于传统档案的馆藏资源，表现为存储容量 大、增长速度快、媒体形式多样等特点。其主要存储方式主要有以下几种：

**数据分布与存储** | **第** **9** **章**

1)直接存储在结构化数据库的二进制大对象 (BLOB) 字段中。

以前，商业银行大部分应用系统中的非结构化数据，如报告、报表、图片等都是以二进 制格式保存在结构化数据库的 BLOB字段中。保存在 BLOB字段中的优点是调用文件的速度 很快，维护和管理简单，与其他应用系统没有任何关联；缺点一是由于非结构化数据文件 大，在数据量不断增大的情况下，会使结构化数据库迅速膨胀，导致数据库性能下降，进而 影响整个应用系统的性能；缺点二是由于各应用系统相对封闭和独立，其他应用无法共享相 关文档资料。

2 ) 以 FTP等上传方式保存到文件服务器中。

以这种方式保存非结构化数据的应用较少，比较典型的有数字档案馆、知识管理和网站 等应用系统。

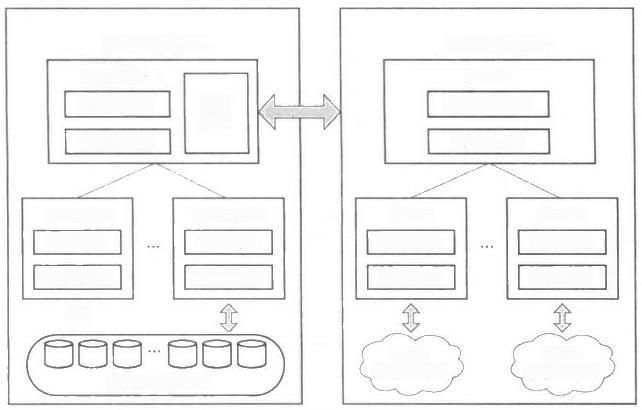
3)通过文件系统直接存储到文件服务器中。

对于大多数没有应用系统的非结构化数据，如常用的工具软件，开发的应用系统软件、 源代码、开发过程文档、技术研究资料等，以及新闻的素材、资料等，通常都将文件直接存 储到文件服务器中。

面对各业务系统各自为战的存储系统，许多企业发现这种方式存在诸多的弊端，目前都 正在评估和选择运用新的技术以提高这些信息资源的利用价值。较为常用的是分布式存储与 并行处理架构。

数据的分布式存储和并行化操作处理是大数据管理平台的核心技术之一。MapReduce 计算模型体现了分治算法的分而治之思想。Map 把一个复杂的问题分解成多个子问题， 每个子问题都相对简单并且可以独立处理； Reduce 把各个子问题的处理结果进行合并。 MapReduce 是适用于大数据存储与并行化处理的简单而有效的一种计算模型。基于 Ma- pReduce 模型可以构建一种非结构化数据分布式存储与并行处理框架，该框架由存储引擎 和检索引擎组成。

图 9 - 2 所 示 为 数 据 分 布 式 存 储 与 并 行 处 理 框 架 。



并行化检索引擎

Master

任务分配引擎

负载均衡器

Datanode 1

特征提取

数据持久化

金

Cache

、索引、特征

Namenode

任务跟踪器

数据分割器

Datanode N

特征提取

数据持久化

Slave N

实例检索

语义检索

Slave 1

实例检索

语义检索

Cache

索引、特征

分布式存储引擎

HBASE Q HDFS

访问 接口

图9 - 2 数据分布式存储与并行处理框架

 **银行数据治理**

分布式存储引擎采用 MapReduce 模型，并基于 Hadoop 的 HBase 和 HDFS 实现。它的主 要功能是基于四面体数据模型提取非结构化数据的基本属性、语义特征和底层特征信息，并 将这些信息连同原始数据存储在 HDFS和 HBase 中。

存储引擎由Namenode 和 Datanode 构成。Namenode 通过 Map 操作将存储数据分成若 干子块，并将这些子块分配给各个 Datanode 进行特征提取和数据永久保存的操作。同时， Namenode 会开启一项任务来追踪处理过程，并在各子块存储完成后执行 Reduce 操作对结 果进行合并。Namenode 和 Datanode 的这种协同工作方式，实现了大量数据的并行与分布 式存储。

并行化检索引擎主要负责数据的快速查询，是基于 MapReduce 模型建立的 Master-Slave 并行化框架。Master负责管理所有 Slave 可分解和分配具体检索任务，并对检索结果进行有 效整合； Slave 模块负责对本地数据实行具体的检索任务并将结果返回给 Master。通过 Master 的有效调度，多个 Slave可并行化执行检索任务，从而实现检索并行化的处理架构。

由于 Master可以实现任务的灵活分发和结果的有效整合，多个 Slave 可以并行化执行任 务且可灵活扩展，因此该引擎具有很好的可扩展性和高效性。

**9.2** **商业银行系统数据分布**

**9.2.1** **总分行数据分布**

银行在机构设置上， 一般有总分行的划分。总分行数据分布规划，主要考虑在总行、分 行模式下的数据分布策略。

银行目前的硬件相比20世纪80年代时的处理能力已经有了极大的飞跃，这为数据的集 中处理提供了可能。目前，商业银行基本构建了数据中心，用于银行 IT 系统的集中部署与 运营。但是，为了满足分行的特色需要，建议可以适当地将数据在分行进行部署，具体可参 考以下指导原则：

1)对于总行统一开发部署到分行的应用所需要的文件继续保留。

2)除满足中国银行业监督管理委员会、中国人民银行等外部监管机构特殊需要的及分 行对少量客户特别服务需要的总行统一部署文件及特色文件外，分行不应部署其他数据。

3)优化总行系统，满足全行统一的经营分析、营销管理、风险管理、业务支撑的 需求。

4)分行特色业务数据分析需要充分利用总行提供的各种功能，满足分行特色业务数据 分析的目标。

**9.2.2** **商业银行系统主副本数据源分布**

**1.** **主副本数据的定义**

主本数据源是指每一项数据都必须唯一确定一个应用系统负责管理。该应用系统称为数

数据分布与存储|**第9章**

据的拥有者，拥有者管理的数据称为主本数据源。

副本数据源是指从系统访问效率考虑，允许将部分数据复制、冗余存放到其他应用系统 中，该冗余数据称为副本数据源。

**2.** **主副本数据的确定**

关于主本数据的确定， 一般有以下三种方式：

**(1)由产生数据的应用系统作为主本数据源** 一般情况下，产生某项数据的应用系统 作为该数据的主本数据源应用。比如，客户刷卡时在信用卡系统中产生了信用卡交易明细数 据，则信用卡系统即为这些明细数据的主本数据应用。

对于仅负责数据简单传输的应用系统(例如用于文件传输的系统等)中的数据，不应 定义为主本数据源。

**(2)由共享数据的应用系统作为主本数据源** 在构建银行系统时， 一般会将一些各专 业条线使用到的数据单独建立系统，例如客户信息系统、产品定义系统、机构管理系统等， 这些数据在全行运营中具有共享性。 一般将负责共享某项数据的应用系统作为该数据的主本 数据源应用。

**(3)由频繁使用数据的应用系统作为主本数据源** 如果两个主本数据源分属于不同的 两个应用系统，两者之间存在多对多关系数据，考虑到业务需要，该关系数据需要在两个应 用中都进行存储。这时，应当选取上述两个应用系统中最频繁使用关系数据的应用系统作为 该关系数据的主本数据源。

每一个企业级最细粒度的数据都应有主本数据源，由其保证数据的正确性和权威性。主 本数据源负责该数据的生成(或采集)、管理维护，并将储存的数据提供给副本数据源应用 系统使用。所有的主本数据源，在业务数据逻辑上应有单一的信息视图，而且还要有完整的 信息描述。

副本数据是相对主本数据而言的，若主本数据需要额外的冗余存储，则冗余部分即为副 本数据。应有可靠的机制确保数据同步，对冗余数据有清晰的使用约束，并确保不因冗余数 据而对业务正确性产生影响。

对于以下情况，可以考虑分布副本数据：

1)主本数据涉及大量查询，使得主本数据所在系统压力及效率下降，而所需数据更新 不频繁，因此可分布副本数据至其他系统提供查询服务。

2)面向整合、分析、挖掘的应用系统可分布副本数据。

3)出于灾备考虑，应用系统可分布副本数据。

此外，在进行主副本数据管理时，应确保对主副本数据的可追溯性，即主副本数据的创 建、维护、使用、流转、加工、销毁整个生命周期，都可进行追踪和关联分析，这有利于保 证数据的一致性。

**3.** **主副本数据的同步方式**

为保证主副本数据的一致性，需要对主副本数据源之间的数据进行同步，可选择实时、 准实时、非实时三种方式。

**(1)实时** 通过调用联机接口形式，同步更新副本数据源，实现主副本数据源同步。 通常情况下，商业银行将该方式用于账务处理。

**(2)准实时** 通过系统工具或调用联机接口形式，异步延迟更新副本数据源，实现主

 **银行数据治理**

副本数据源同步。通常情况下，商业银行将该方式用于时效要求较高的精准营销、短信推送 等场景。

**(3)非实时** 通过批量接口形式传输数据文件，更新副本数据源，实现主副本数据源 同步。通常情况下，商业银行将该方式用于经营分析场景。

**9.2.3** **商业核心银行系统与外围系统间的数据分布**

随着业务的发展，特别是随着互联网金融的崛起，客户个性化、差异化需求越来越多， 金融产品创新需求层出不穷，需要银行的核心交易系统具有灵活架构支持，同时应兼顾金融 服务成本的合理投入，金融服务成本控制水平将影响市场竞争力。

从历史发展来看，银行各类系统的部署与技术发展密切相关。银行的核心交易系统承担着 银行存贷款及结算等核心业务，必须部署在安全性、稳定性、性能更高的技术平台上。虽然这 样的技术平台价格比较高，但是为了满足上述要求，这类技术平台仍是各大商业核心银行系统 功能和数据部署的首选。从实际情况看，IBM主机在核心银行系统中占比非常高。对于非银行 核心交易系统，则大多采用开放平台技术，比如内部运营、客户评价、绩效管理等系统。

随着银行业务发展和监管需要，各核心银行系统的功能越来越丰富。出于各方面因素考 虑，许多风险控制、增值类服务也基于核心交易系统增长，这导致核心银行系统的功能耦合 程度加剧，系统灵活性降低，投入成本增加。与此同时，通用、开放平台技术的发展为银行 核心交易系统的解耦合带来了机遇， 一些更为开放、低成本的数据快速同步技术、内存处理 技术越发完善和成熟，使得原本需要耦合在核心银行系统才能实现的功能在外围系统实现成 为可能。因此，在数据分布方面，银行有必要对核心银行系统、外围系统的数据部署进行统 筹规划及治理。

核心银行系统、外围系统的数据分布，需要结合业务目标、技术目标进行统筹考虑。

一是业务上，可以考虑将与账务处理无关的业务功能从主机核心交易系统中加以剥离， 特别是一些需要风险控制功能、增值服务功能、数据分析功能，应剥离至外围系统去实现。 在数据层面，可以将客户、账户及与其紧密相关的介质、协议等信息部署在主机中；产品、 服务类协议、非货币账户、托管账户及其相关信息部署在平台上；同时，为了降低主机业务 系统的查询压力，可以将客户、明细等信息在平台冗余部署，提供数据查询及相关增值 服务。

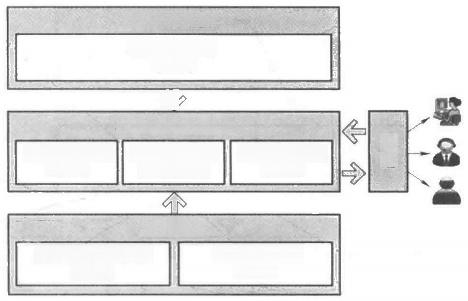
二是技术上，应提升系统灵活度，支持业务快速创新；保证核心银行系统的稳定性，支 持安全运营；合理调配昂贵设备和低成本设备的应用功能和数据部署，实现有效的成本控 制；针对联机事务处理、联机分析处理和混合型处理多种应用场景，充分利用分布式技术和 平台高可用技术提升开放平台技术架构，有效承接开放平台各类应用业务量的快速增长。

**9.3** **案例**

**1.** **案例一：主副本数据的分布**

某商业银行客户信息主副本数据分布如图9-3所示。

**数据分布与存储**| **第9章**



核心银行系统

客户基本信息

(主本)



客户增值服务信息

客户基本信息 客户持有产品 客户偏好信息

(副本) 信息(副本) (副本)

数据仓库

客户持有产品信息 客户偏好信息

(正本) (正本)

渠 道

图9-3 客户信息主副本数据分布

客户信息系统在任何银行都是非常重要的业务系统。从功能来看，客户信息系统主要承 担着银行客户信息维护、客户信息查询等功能，与存贷款等业务办理流程紧密相关。

随着银行金融产品趋同，增值服务多元化和便捷性成为客户选择金融产品的重要关 注点。为提升产品竞争力，商业银行正在大力发展客户增值服务。客户增值服务提升的 重点是信息整合、共享与价值创造，相关的信息包括客户的基本信息、行为偏好、持有 产品信息等。目前，关于这些增值服务信息的查询访问量非常高，若在传统客户信息系 统中实现这些功能，则对客户信息系统的设备投入、运维带来非常大的挑战，而且会给 传统的与业务办理流程紧密相关的客户信息维护、查询功能带来冲击，进而影响业务的 办理。

因此，此商业银行将传统个人金融业务、对公业务、信用卡业务等多个系统的客户基本 信息用副本数据分布，并与来自数据仓库等集成分析系统的客户行为偏好信息、产品持有信 息结合，建立对应的增值服务信息系统，进而提供多渠道、可快速响应的客户增值信息服 务。这样， 一方面减轻了传统个人金融业务、对公业务、信用卡业务、客户信息系统的压 力，另一方面也更好地满足了业务需求，带来了良好的用户体验。

在通常情况下，为了保证信息进入的范围，需要制定进入客户增值服务信息系统的数据 进入原则。此商业银行制定了以下原则：

**(1)满足多渠道信息共享的原则** 存储多个渠道共享的信息，不存储单个渠道使用的 信息，在这种业务场景下允许数据进入客户增值服务信息系统。

**(2)以客户为中心确认数据进入范围** 客户增值服务信息系统存储的数据都是客户维 度结构化信息，具体包括客户基础信息、客户服务信息和客户衍生信息等。

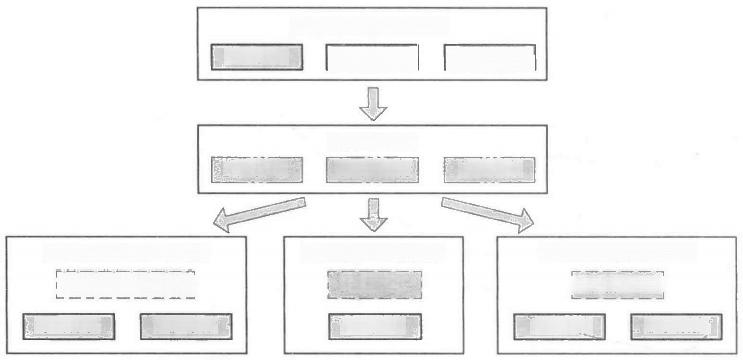
**(3)进入数据非关键路径访问的场景原则** 客户增值服务信息系统与产品处理、账务 核算关联度必须要很低，这样才能在出现服务故障或者性能波动时不至于影响关键业务路径 的处理。

因此，进入客户增值服务信息系统的副本数据具有满足可延时的特点，主要表现为数据 同步存在一定延时，业务对数据一致性可以有一定的容忍度。

**2.** **案例二：核心银行系统与外围系统的数据分布**

某商业银行客户主机与平台数据分布如图9-4所示。

**银行数据治理**



主机核心银行系统(DB2)

账户事件信息 协议介质信息

数据交换系统

准实时复制

风控系统：Oracle、大内存

事中干预、事后监控

布控名单 风险事件

流数据处理系统：大内存

增值信息推送

历史明细：MPP数据库

在线查询

协议数据

历史明细

实时复制

客户信息

批量传输

交易事件

图9 - 4 客户主机与平台数据分布

为提升核心银行对金融产品创新的灵活、快速支撑能力，满足客户个性化、差异化需 求，该商业银行从全局应用功能和数据布局角度对全行核心系统架构进行了整体规划，制定 了核心银行系统主要进行账务处理相关业务，并将非账务处理功能剥离至开放平台实现的整 体战略。

功能解耦的关键是数据的解耦。根据此战略思想，该商业银行成立团队进行课题攻关， 通过流技术、数据同步技术、内存技术、分布式计算技术的理论研究、原型验证及生产试 点，解决了主机核心银行系统与开放平台外围系统的数据实时、准实时同步问题，使得原先 必须在主机核心银行系统实现的高时效风险控制、增值功能在低成本开放平台上实现成为现 实 ， 有 效 支 撑 了 日 益 增 长 的 风 险 控 制 、 增 值 服 务 功 能 需 求 。



**第10章**

**数据交换**

数据交换的主要目的是实现对银行数据或信息的共享，满足银行日常经营管理和业务运 营过程中的数据使用需求，以及满足外部监管等需要。

一般来说，银行数据交换包括两大类：商业银行与外部机构的数据交换，以及商业银行 内部应用间的数据交换。前者一般表现为从外部数据服务提供商获取信息，以及和监管机构 进行数据交换(提供监管报送数据和接收反馈结果);后者表现为应用间的信息交换，其目 的是保持主副本数据源的一致性，促进数据共享，满足外部监管等需要。商业银行数据交换 的主要形式是批量文件， 一般由数据提供方提供数据文件，并提供对应的核对文件(包含 数据文件名称和大小等信息),供接收方使用前核对。

各银行均构建了自身的数据交换体系，有的做得比较简单，如通过文件传输协议 (File Transfer Protocol,FTP) 工具或自主研发的文件传输工具进行简单传输；也有的相对复杂， 如通过统一的系统将各类抽取、转换、加载 (Extract-Transform-Load,ETL) 工具以及传输 工具进行整合，为银行的数据交换提供统一服务。 一般情况下， 一个相对完善的银行数据交 换体系一般包含数据交换标准、数据交换体系架构和数据交换时效性管理等方面的内容。

1)为了加强数据交换的标准管理，商业银行需要制定企业级数据交换规范，规范银行 内部应用间以及银行内部应用与外部机构间的数据交换，内容一般涵盖文件交换原则、文件 及文件内容命名规范、文件质量检查等内容。

2)根据银行实际的管理需要和技术条件，建立企业级的数据交换体系，建立集中式或 非集中式的数据交换系统、对外数据报送平台，作为银行数据交换的体系支撑，承担起企业 数据总线的作用，为全行应用间以及应用与外部机构间数据交换提供7×24h 服务。

3)为了加强数据交换的时效性管理，通过对数据交换中数据流的管理，建立企业数据 流向的宏观视图，实现对数据交换的联动管理和交换时效管理，保证数据交换的准确性和时 效性。

随着银行业务的发展，各行面临的数据交换问题越来越严峻。比较典型的情况包括：各 类系统需要交换的文件数及数据规模越来越大，对交换体系形成压力；数据交换体系需要足 够灵活，可满足银行内部的总分行间、数据中心内部、境内外机构以及与外部监管机构等不 同场景的数据交换需要；需要保证大规模数据交换的准确性，并满足不同场景的数据交换时 效要求。这里将结合上述问题，从数据交换的标准、体系、流程管理等方面介绍构建数据交

 **银行数据治理**

换体系并进行管理的方法。

**10.1** **制定数据交换的标准**

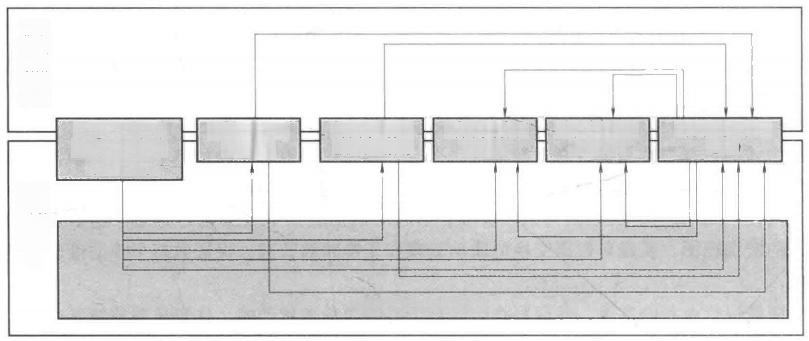
在商业银行的两类数据交换中，商业银行与外部机构的数据交换标准往往由外部机构提 供， 一般只需要遵从即可。这里重点针对商业银行内部应用间的数据交换标准进行介绍。

商业银行制定企业级数据文件交换技术规范，需要对数据交换原则、接口和文件命名规 范、接口内容规则、检查规范，以及日志、安全管理等内容进行明确，全面规范全行应用 间、应用与外部机构间的数据交换，指导数据交换工作顺利进行。

**1.合理规划数据交换模式，满足各类数据交换的业务需求**

一般来说，商业银行存在多种数据交换途径、多种技术手段，也存在多种不同的应 用需求场景。既有复杂的企业级批量数据文件交换，又有简单的应用间的直接交换；既 存在实时、准实时的高时效数据交换场景，也存在 T+2 这样的一般时效的数据交换场 景。数据交换应充分考虑数据交换系统的技术特性和数据交换的应用需求，规划数据交 换模式，针对每一种模式建立合理的数据交换准入条件，即每种交换模式的适用场景， 避免数据交换过于集中或依赖，降低单一系统或交换路径的数据交换压力，满足各类数 据交换的需求。

当前，业界各种交换模式主要分为集中式数据交换和非集中式数据交换两类。所谓集中 式数据交换，可以理解为一个大 HUB 集线器，在所有数据提供方和接收方之间传输的文件 均经过该 HUB 进行传输；非集中式数据交换一般为网状结构，数据提供方和接收方系统可 根据需要进行任意点对点传输，不需要经过中间环节的系统进行路由。 一般来说，大多数银 行，特别是大型商业银行倾向于采用“集中式数据交换模式(如中国工商银行、中国建设 银 行 ) 为 主 ， 各 种 非 集 中 的 点 对 点 交 换 模 式 为 辅 ” 的 交 换 体 系 架 构 ， 如 图 1 0 - 1 所 示 。



非集

中式

数据

交换

模式

核心银行系统

(实时金融业务

数据处理)

集中

式数

据交

换模

式

集中式数据交换系统

外围系统(数据

仓库)

外围系统(报 表统计)

外围系统(金 融产品销售)

-外围系统(风 险监控)

外围系统(信 贷审批)

图10-1 大型银行数据交换体系架构(数据中心)

该体系架构具有较好的体系灵活性，可在交换效果和实现成本间取得较好平衡，主要原 因如下：

**数据交换** |**第10章**

**(1)集中式交换模式** 体系中“集中式交换模式”负责主体交换，满足大部分数据交 换需求。这种集中处理可达到以下效果：

‘1)便于数据交换架构进行集中规划，通过为各应用提供统一的交换处理功能，能节省 数据交换的成本。

2)便于集中进行高可用性容错机制建设，以较低的成本为整个交换系统提供可用性 保证。

3)便于进行集中监控，能及时发现问题并进行针对性治理，相对于分散运维更能大幅 度降低运维成本。

4)可充分利用网络、服务器和存储等硬件设备资源，节省设备投入。

**(2)非集中的点对点交换模式** 体系中“非集中的点对点交换模式”作为有效补充机 制，可用于简单、高时效场景的数据交换。引入该模式可避免数据交换过于依赖“集中式 交换模式”,降低后者的数据交换压力。

在此需要强调一点，商业银行不仅可以在数据中心部署“集中式+非集中式”的交换 体系架构，而且在企业发展到一定规模后，可进一步在各分行内部或者区域中心内部部署小 规模的“集中式+非集中式”交换体系架构，满足本地的交换需求。但是，不同的交换体 系架构间仍需要建立传输通道，满足相互数据交换的需求。

**2.** **制定文件交换规范**

银行每天都有大量的数据需要进行内外部、内部系统间交换， 一些大型商业银行每天交 换的文件个数达到10万数量级。为保证数据交换的顺利进行，必须制定规则进行约束，而 文件作为数据交换的主要载体，对文件交换进行统一规范就显得尤为重要。

各商业银行可根据自身特色确定规范，但宜遵循一定的原则，主要包括：

**(1)规范对象** 规范对象应包括但不限于文件命名、文件内容、文件检查、文件管理。

**(2)规范场景**

1)对于数据中心、各分行内部或者区域中心内部，如使用集中式数据交换模式，宜分 别就“供数方 →集中式数据交换系统”以及“集中式数据交换系统 →使用方”两种传输场 景制定规范。

2)对于数据中心、各分行内部或者区域中心内部，若使用非集中的点对点交换模式， 需对“供数方→接收方”的传输场景制定规范。

3)对于数据中心、各分行内部或者区域中心之间，若有传输通道，则还需对该传输通 道内“供数方→接收方”传输场景制定规范。

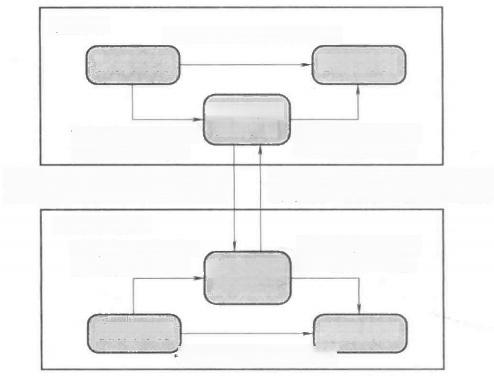
图10-2所示为某银行规范涵盖场景。

4)对于每种传输场景(可定义为一种处理功能),每批需要交换的数据文件都应由数 据提供方提供一个核对文件，用于接收方校验该批数据文件的完整性。

5)数据文件和核对文件名中应包含足够的元素用于标识一次交换过程。元素对象可包 括以下内容：提供方标识、接收方标识、日期、每天交换的序号、文件功能描述和地区号标 识等。相关元素宜使用完整的单词或通用的缩写，避免产生误解。相关元素不仅是文件名要 素，也是交换系统中每条交换路径及其作业的区分标识，宜在元数据管理系统登记数据交换 任务时指定，供集中式数据交换模式的数据交换系统应用分配作业和监控时使用。

6)数据交换规范中对每种传输场景，可约定文件传输形式(文件或压缩包),以及提

**银行数据治理**



“供数方”

供数方系统

“供数方→集中式数

据交换系统”规范

向数据中心“分行/本地中心”的传 输通道中“供数方→接收方”规范

分行/本地中心

“供数方→集中式数

据交换系统”规范

接收方系统

供数方一接收方”规范

数据中心向“分行/本地中心”的传 输通道中“供数方→接收方”规范

“集中式数据交换系 统→使用方”规范

“集中式数据交换系 统一使用方”规范

集中式数据 交换系统

集中式数据 交换系统

“接收方”规范

接收方系统

供数方系统

数据中心

图10- 2 某银行规范涵盖场景

供方和接收方文件的存放路径。

7)数据文件和核对文件的内容结构也需要做好约定，包括但不限于以下内容：文件码 制、记录长度、记录中字段定义及其位置(可通过字节位置定位或使用分隔符定位)等。 其中，核对文件宜同时包含数据文件统计信息和明细信息， 一般第1 行可用于统计信息 (如提供方名称、文件传输日期、传输文件个数等),从第2行开始各行可用于描述每个数 据文件的明细(如文件名、文件总大小等)。

**3.** **规范文件的检查机制**

数据交换过程需要保证数据文件的质量，避免错误数据加载。数据文件质量检查内容可 依据各行实际情况制定， 一般包括文件合法性检查、文件完备性检查、文件并发处理检查、 文件重复性检查等。

**(1)应用系统设置检查** 接收方应用系统应设置文件异常处理机制、延误报警机制和 延误时间间隔参数——当超过“延误时间间隔”还未收到对应文件时，则启动延误报警机 制提示延误预警信息；当发现对文件无法通过各种相关检查时，则启动异常处理机制(包 括报警提示和实际处理等)。

**(2)文件合法性检查** 提供方应用系统应确保文件、核对文件(也称为清单文件或控 制文件)的文件命名以及文件内容等符合规范要求；接收方应用系统在收到数据文件后、 文件加载前，也应对上游下传的数据文件名称、核对文件名称以及格式内容进行检查，判断 是否符合相关规范要求。

**(3)文件完备性检查** 主要是指接收方应用系统应根据数据文件所对应的核对文件中 的字段信息(如文件名、文件大小、文件个数等),对数据文件的实际信息与核对文件中的 内容进行验证，校验数据文件的完整性，可避免发生文件仅传输了一部分的问题。该问题若 不能被及时发现，则将导致接收方应用系统遗漏加载部分信息，进而引发业务问题(如统

计 指 标 出 错 等 ) 。

**(4)文件重复性检查** 主要是指数据文件重复，接收方应用系统应进行数据重复控制 检查，以保证在一定业务周期内避免数据重复处理(重加载)。接收方应用系统可通过增加

**数据交换**|**第10章**

相关业务控制信息实现数据文件重复控制检查。例如：可通过数据字段组合“提供方标 识+地区号标识+日期+每天交换的序号”来保证数据文件加载作业的唯一性，避免重复 处理。

**(5)文件内容检查** 接收方应用系统在文件加载后、实际使用前可通过各种措施进行 一定程度的文件内容检查，检查内容可包括：数据类型检查、业务逻辑检查等。数据类型检 查主要是针对文件中各字段数据类型的正确性进行检查；业务逻辑检查主要是根据产品业务 逻辑规则逐条对文件记录的合法性进行检查。错误记录可写入日志供后序排查。

**4.** **数据文件安全及生命周期管理**

数据文件安全需要遵循数据安全相关规范，包括：敏感数据需要考虑进行加密传输；文 件存取权限等需要遵循应用系统数据文件目录相关安全处理原则等。

数据交换系统中，应规划文件生命周期管理策略以指导清理机制设计，在满足数据交换 需求的同时，减少存储容量压力。 一般在数据交换系统中的数据文件可在线保留若干天，之 后可进一步采用压缩技术进行压缩，存放一定时间，存储介质可以是磁盘或磁带，在确保数 据提供方和接收方应用系统针对该文件的处理没有问题后可删除。

**10.2** **搭建数据交换体系和数据交换系统应用**

当商业银行规模较小时，数据交换可采用非集中式点对点传输交换模式，由数据提供方 应用系统直接传输文件给接收方应用系统。该方案主要依赖点对点传输工具的选择。相关工 具可以在FTP 协议上自行开发，也可选购业界典型工具，如 IBM Connect-Direct(CD) 工具 或东方通公司GTP 工具等。相关工具宜具备以下特点：能同时完成发送和接收；具备失败 自动重传功能(若能实现断点续传，则效果更好);易于配置任务；能实现集中监控查询 (如传输节点状态和任务传输日志)。相关工具可部署在需要进行数据交换的各应用的服务 器上，形成点对点传输网络。

在商业银行规模逐步发展壮大后，需要建设“集中式+非集中式”的数据交换体系架 构。其中，非集中式交换结构可继承使用原先点对点传输网络，集中式交换结构需要重新规 划搭建。其实质就是搭建一个数据交换系统，实现提供方应用系统(上游)、数据交换系统 (中游)和接收方应用系统(下游)的有机结合，以统一的规范实现数据的获取、转换、校 验、发送等功能，实现信息在银行内部的快速共享流动。

集中式交换结构的建设有助于商业银行以较低的成本完成高时效和高可用建设规划，其 额外提供的文件转换处理功能和一对多复制分发功能，可有效降低提供方和接收方应用系统 的数据交换程序复杂度，通过将专业的服务集中到“数据交换系统”,可显著降低商业银行 的总体数据交换成本。

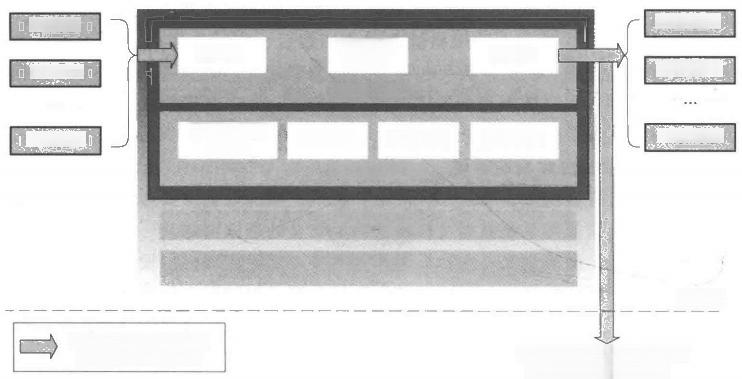
下面对数据交换系统架构进行简单介绍。

一个数据交换系统由专用服务器(集群)和物理存储构成，并通过传输工具和提供方 应用系统、接收方应用系统进行连接。数据提供方应用系统负责将文件(源文件)传到数 据交换系统的物理存储上，由数据交换系统检测源文件是否到达并进行内部处理(转换、 校验),在物理存储上生成结果文件，最后由数据交换系统将结果文件传给接收方应用

 **银行数据治理**

系统。

集中式数据交换系统逻辑架构如图10-3所示。



目标应用1

接出区

目标应用2

目标应用n

本环境

其他环境中数据交换系统/ 数据交换存储/接入区

|  |  |
| --- | --- |
| 源应用1  源应用2  …  源应用m | 接入区  作业触发信 号检测 |

传输工具(负责传输数

据文件和核对文件)

中间结果

存储区

数据交换存储

文件检查 文件处理

数据交换处理

集中式监控管理

集中式配置管理

作业调度

图10-3 集中式数据交换系统逻辑架构

集中式数据交换系统一般包含以下功能模块(数据交换系统的功能性需求):

**(1)数据交换存储** 通过接入区目录与接出区目录，向外提供整个集中式数据交换系 统对外的数据接口，分别用于接收方和提供方应用系统发送的待处理文件，以及发送处理结 果文件给接收方应用系统。处理中间结果可存放在结果存放区。当集中交换规模较小时，数 据交换存储可以使用对应服务器的本地硬盘，由本地文件系统管理；当集中交换规模较大 时，数据交换存储可以使用逻辑共享存储(如分布式文件系统存储节点服务器上的存储) 或物理共享存储(共享的SAN 存储盘机，需要使用共享文件系统管理，典型产品包括IBM GPFS、昆腾 Stornext 等)。

**(2)数据交换处理** 对数据文件进行处理。处理功能可由商业银行根据自身需求统一 设定， 一般可包括以下类型：字段格式转化、字段码制转化、文件按行列筛选、复制分发、 压缩/解压、文件合并等。在处理前，宜通过文件检查模块进行检查，具体的检查要求包括 文件合法性和文件完备性。当文件数较大时，合理的调度尤为重要，此时需要部署作业调度 模块，分为调度服务器端和调度客户端两部分。调度服务器端一方面通过队列存储待处理作 业，另一方面根据调度客户端所在服务器的进程空闲状态，将作业分配给对应的调度客户端 所在服务器进行处理。调度客户端一方面负责调度工作，即接收调度服务器端指令启动对应 处理作业；另一方面登记日志，供集中式监控管理模块查询使用。

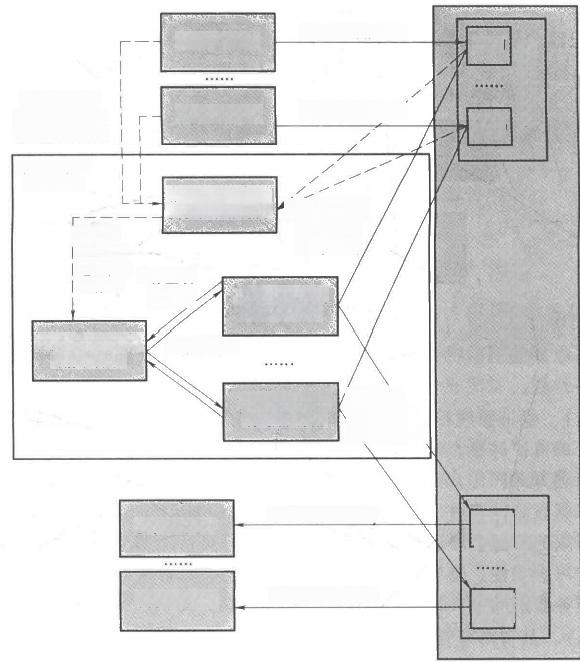
图10-4所示为数据交换处理流程。

**(3)集中式监控管理** 使用统一的数据库记录处理日志、文件传输日志、模块或服务器 的状态等信息；使用远程控制命令实现远程控制机制(如服务器/模块进程启停以及作业重做 等)。两者可通过前台查询控制界面相结合，为运维人员提供统一的监视查询和控制干预视图。

**(4)集中式配置管理** 主要包括系统参数和作业参数的配置管理、用户权限管理、系 统管理等。

**(5)传输工具** 负责传输数据文件和核对文件。其工具可以复用原先非集中式点对点

数据交换 第 1 0 章



1.文件传输/写入

目录1

2.2检测方式

(提供方文件传

输后报文通知)

集中式数据

交换系统

|3.文件

检查 →

调度模块(客户

端1)/服务器

调度模块(服务器

端)/服务器

/5.文件处

、理：读取 → 处理一写入

7.文件传输/读取

目喇

7.文件传输/读取

目剥

21检测方式 (目录文件 轮询扫描),

作业触发信号检 测模块/服务器

调度模块(客户 端m)/服务器

6.作业状 态反馈

通知/登 记作业

1.文件传输/写入

数据交换存储

提供方系统1

接收方系统1

接出区目录

提供方系统n

接入区目录

接收方系统k

4.作业发布

目录n

图10-4 数据交换处理流程

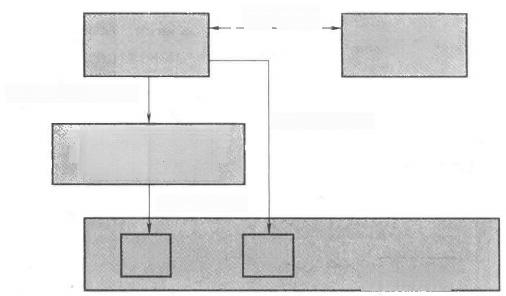
传输网络中的传输工具，只需要将发送方或接收方调整为集中式数据交换系统应用所属的服 务器。此外，“源应用”或“集中式数据交换系统应用所属的服务器”宜在应用层面实现 “发送成功数据文件后，才发送核对文件”的机制。某些情况下，传输工具也可以是分布式 文件系统或共享文件系统的客户端软件，它们负责提供远程接口，供提供方和接收方应用系 统直接把数据交换存储当成本地盘进行文件读写，实现高效传输。但是，该场景一般不宜过 多，应有针对性地选择交换时效性要求高、传输数据规模大、技术上符合要求(若与“数 据交换系统应用”距离较近，则属于同一个网段内)的应用使用。事实上，每个应用可以 同时包括文件系统客户端和传输工具两种模式，以满足不同的需求，如图10-5所示。

另外，集中式数据交换系统还应针对以下特性进行特别设计(非功能需求):

1)高性能。作为企业级的基础系统，首先需要具备高性能，以满足服务基线时间窗口 的需求。当数据交换规模较大时，“专用服务器集群+分布式/共享文件系统+基于调度的 负载均衡”是实现高性能的主要方法，其核心是在数据交换系统应用内部，各服务器(特

别 是 作 为 调 度 客 户 端 的 文 件 处 理 服 务 器 ) 共 享 数 据 交 换 存 储 ， 每 台 服 务 器 都 能 看 到 存 储 上 任何一个文件，在调度的“指挥”下不停地获取文件进行处理和分发，保持服务器上所有 进程尽可能处于满负荷工作状态，使处理和分发速度均达到最大值。由于各服务器对共享存 储的读写顺序需要协调，因此需要在各服务器上安装分布式/共享文件系统的客户端软件。

 银行数据治理



元数据交换

数据提供方系统1

2.文件传输工具方式

1.文件系统方式

集中式数据交换系统一文

件传输服务器(挂载管理

共享存储的文件系统)

文件系统写入

目录1 目录2

数据交换存储

文件系统元数据 服务器

图10-5 应用的文件系统客户端和传输工具两种模式

此外，梳理各接收方应用系统的文件处理要求，通过“合并处理方式”实现一次处理多次 使用，也可提高性能。

2)可扩展性。保证系统根据业务要求合理配置，稳步发展，从业务架构和物理架构两 个方面考虑系统的可扩展性。业务架构是指应实现基于调度的配置化开发，即：作业执行命 令不应采用批处理方式固化，而应通过抽象形成一系列公共函数、文件名等作为参数传入； 实际使用时应根据收到的文件名，读取配置，执行命令的灵活选择和组装(单个函数和参 数，以及不同函数的前后步序运行关系)。上述模式可灵活地适应功能的增加和修改，实现 业务功能层面的可扩展性。物理架构指文件处理/分发应具备横向扩展能力，可通过调度结 构增加客户端服务器的实现。

3)可管理性。具备高可管理能力，保证系统的生产运行，在出现异常的情况下，可以 快速定位并恢复。可管理性是集中式监控管理的基础，其核心是：作业的日志信息内容规 范，且能实现重做(能选择和调节本作业以及前后步序作业);模块进程或服务器有一套完 整的“心跳”登记机制和一键式启停命令。

4)高可用性。作为7×24h 工作的企业级数据交换总线，数据交换系统应用中单个模 块进程或服务器故障/维护停机时，应尽可能不影响整个数据交换系统应用的正常运行，其 实质是可在其他服务器上调度故障/维护停机的模块进程，以及可由其他服务器接替故障/维 护停机服务器的职责。数据交换处理架构主要通过集群和负载均衡的角度充分考虑系统的健 壮性，其他核心服务模块，例如传输工具和调度服务，可通过HA 双机热备的方式来保障高 可用性。此外，高可用性的重点是“高”。因此，发生故障时需要能快速检测和恢复，这要 求模块进程或服务器的“心跳”登记和检查频率需足够高(如10s 一次),恢复命令需要一 键式实现。维护(停机)的方案宜采用分批操作原则，以保证总有设备能正常工作。

5)安全性。作为数据线的心脏，需要保证系统、数据安全加工，安全接入和安全 分发。

**10.3** **全流程数据流管理**

为了提升数据交换时效，商业银行需要对数据交换过程中的数据交换时效进行管理。数

**数据交换** **|第10章**

据流管理就是实现对数据交换时效管理的有效途径。

对于采用集中式元数据管理方法的银行，建议将统一的数据交换系统的数据与元数据管 理系统相结合，将元数据纳入元数据管理系统进行管理，包括元数据的登记、审批和管理 等。同时，为了保证元数据登记质量，建议将元数据管理系统与应用系统的研发过程相结 合，以元数据管理系统登记文件信息为准进行版本发布，通过系统实现数据交换的数据流 管理。

在元数据管理系统上，可以通过对数据交换文件结构、码制转换信息、调用关系以及输 入输出信息的管理，实现批量数据流的管理目标。

全流程数据流管理的主要作用包括：

**1.** **实现数据提供方和接收方联动修改的管理**

宏观上，对于各个应用间的批量数据流交互情况进行管理；细节上，数据提供方和接收 方登记文件结构及调用关系，上游信息修改自动通知调用方联动进行修改，建立与程序、作 业步、作业、应用的关联关系，从而实现影响性分析和血缘分析功能。

**2.对各环节输出文件的承诺时间进行管理**

主要通过对提供方和接收方数据文件交换设计阶段的承诺时效，与系统投产后的文件实 际传输时效进行管理，对不符合承诺时效的数据交换文件进行治理，以提升数据交换时效。

**3.** **实现应用间数据流向的宏观视图展现**

主要建立基于交换数据文件的应用级数据流向视图，整体展现银行内部数据流向，把控 和关注关键的交换路径，维护银行数据交换的稳定性。

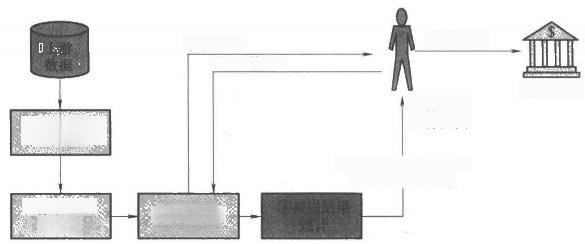
**10.4** **外部数据交换**

随着我国银行监管体制的日益成熟，在日常的经营管理中，银行面临各种监管的数据交 换需要。为了加强对外交换数据的统一管理，商业银行宜建立统一的对外数据报送平台，整 合各类数据报送需求，实现对各外部机构的数据统一管理、统一报送、统一接收，降低分行 单独向监管机构报送所产生的数据冗余和质量问题。同时，建立外部金融信息库系统，与外 部数据服务商建立金融信息交换，并对各类金融信息进行统一管理，实现全行共享。

一般来说，商业银行与外部数据交换主要包括监管信息报送和外部金融信息交换等。其 中，监管报送是商业银行应金融监管机构要求，为满足监管机构收集金融数据、进行相关合 规性检查需要，按监管部门要求向其报送相关数据的行为。报送对象包括：中国人民银行、 中国银行业监督管理委员会、中国证券监督管理委员会、国家外汇管理局等金融监管机构。 目前，从报送方式上，监管报送可以分为业务部门调整手工报送和业务部门调整系统自动报 送两类。业务部门调整手工报送包括由系统自动出报表，由业务人员进行手工调整后，进行 报送。业务部门调整系统自动报送包括业务人员参与报送流程、业务人员未参与报送流程两 类。具体报送流程如图10-6所示。

由于手工报送容易产生一系列的潜在业务风险，如核对遗漏、修订有误、明细与汇总不 一致等问题，因此，商业银行需要建立一套完整的、由业务主导的监管自动报送系统，并实 现事前校验、事中预警、事后信息反馈等全流程管理。同时，商业银行还需要建立统一的对

**银行数据治理**

①数据加载

②数据校验

(规则)

报表结果

订正数据

③数据处理

手工报送

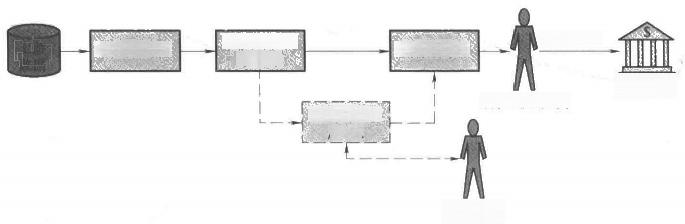
监管报送

业务人员

报送结果统计，与上

期情况的比较等

监管机构



②数据校验

(规则)

科技运维人员

③数据订正

/确认

监管报送 业务人员

①数据加载

④数据处理

系统报送

监管机构

图10-6 手工报送与系统报送流程

外数据报送平台，整合外管报送、反假货币数据报送、理财产品报送、银税数据共享、个人 /企业征信等系统，利用各类数据报送软件和报送系统，实现对中国人民银行、国家外汇管 理局、工商税务等机构的数据统一管理、统一报送、统一接收，降低分行单独向监管机构报 送所产生的数据冗余和质量问题。

同时，建立外部信息系统，与外部数据服务商建立信息交换平台进行统一整合，并为行 内各应用系统提供文本、XML等访问接口，以达到统一外部金融信息渠道，统一管理各类 金融信息，实现全行金融信息共享的目的。



**第11** **章**

**数据安全**

数据是银行的生命线，银行数据安全管理的目的就是保障客户的基本信息、交易信息、 账户信息的安全，避免将其泄露或遭到非法篡改。同时，数据安全对于保障银行各系统之间 源数据和数据副本的一致性非常重要，是各项业务有效开展的基础。当银行面临各种人为或 自然的突发事件时，数据的安全性(包括数据的完整性、可用性、共享性等)可保障银行 关键业务系统的及时恢复，是实现银行业务连续性的关键。

**11.1** **数据安全及重要性**

**11.1.1** **数据安全的概念**

数据安全是指保证数据的机密性 (Confidentiality) 、 完整性 (Integrity) 和可用性 (A- vailability) 。数据安全的实质是采取措施保护银行数据资产，使之不因偶然或者恶意侵犯而 遭受破坏、篡改及泄露；保证银行信息系统能够连续、可靠、正常地运行，使安全事件对银 行业务造成的影响降到最小，确保银行业务运行的连续性；同时最大限度地获得利润回报和 利用商业机遇。机密性、完整性和可用性是银行数据安全的目标，也是基本原则。

**1.机密性**

机密性要求银行数据不能被泄露给非授权的个人、集体或其他实体，不能为其所用。简 而言之，所有人员都可以访问的信息属于公开信息，需要限制访问的信息一般属于敏感信息 或秘密。对于银行来说，敏感信息是指机密性或完整性保护要求较高的信息， 一旦泄露将给 银行或用户造成损失。 一般来说，商业银行敏感信息主要包括以下几类：

**(1)客户基本信息** 主要是指客户基本信息、客户交易信息、客户鉴别信息等。

1)客户基本信息，主要是指跟客户身份相关，能够反映客户真实基本情况的信息，如 客户姓名、客户地址、客户联系方式以及账户信息等。

2)客户交易信息，主要是指客户在参与各种银行日常业务时形成的交易记录等，如存 取款交易记录、转账交易记录等。

 **银行数据治理**

3)客户鉴别信息，主要是指客户在参与各种银行日常业务时，登录系统、执行交易、 办理业务等活动中需要进行控制访问或者授权执行的信息，如网银登录密码、银行卡交易密 码等。

**(2)银行信息系统关键配置类信息** 主要是指日常运行过程中的关键配置类信息，如 系统 IP 和用户等。银行数据安全离不开应用系统的安全稳定运行，而系统相关的关键配置 类信息对保障系统安全和数据安全尤为重要。

**(3)银行内部经营管理信息** 主要是指银行内部经营管理过程中形成的大量敏感信息， 如财务信息、人力资源信息等。

银行数据信息的机密性导致数据容易被攻击者威胁，常见的途径包括：网络监控、肩窥 (Shoulder Surfing)、黑客攻击、病毒攻击。攻击较轻时会给银行造成资产损失、社会声誉受 损，重则影响国家正常经济秩序，影响国家安全稳定。2014年，韩国金融监督院证实：韩 国国民银行和农协银行等多家大型商业银行的大量用户信息被泄露，内容涉及手机号码、个 人地址、信用卡账号乃至部分银行的交易记录。由于此次泄露的个人信息十分详细，可能引 发各种金融诈骗的担忧致使韩国民众蜂拥销户或变更相关信息，大约有超过150万用户办理 了信用卡的停用、注销或重办业务，极大地影响了银行的声誉。同样，信息安全事件也为我 国敲响了警钟。2014年，部分用户的12306账户信息在互联网遭疯传，包括用户账号、明 文密码、身份证信息、邮箱等信息，引起了公安部的高度重视。初步怀疑该问题和“用户 在不同网站使用相同用户名和密码”有关，黑客很可能采取“收集各种网站已泄露的用户 和密码信息+使用‘撞库攻击’的方式”非法登录12306 网站，获取了对应用户的注册相 关信息。

**2.** **完整性**

完整性是指保证数据和信息系统的准确性和可靠性，并禁止对重要数据和关键数据进行 非授权的修改。硬件、软件和通信机制必须协同工作才能保证数据被正确处理，并保证在没 有意外或恶意篡改的情况下将数据安全完整地传递到目的地。银行数据及信息系统的完整性 可能因为两方面的原因遭受破坏： 一方面，数据在使用、传输、存储等过程中可能被攻击， 导致数据完整性被破坏，病毒、逻辑炸弹攻击等均可能破坏数据及信息系统的完整性；另一 方面，信息处理方法的不正确和不规范也可能导致信息完整性遭到破坏，如由于不正当的操 作导致误删除文件，可能造成重要数据文件的丢失。银行可通过研发、购买等方式提供具有 完整性保护功能的系统，数据正确的使用方法和使用流程也能够保证攻击者或者用户错误操 作不会对系统或数据的完整性造成损害。

**3.** **可用性**

可用性是指银行系统和网络应该提供适当的能力，从而能够在可以接受的性能级别下， 以一种可以预见的方式运行，保证经过认证的用户能够对银行数据和资源进行适时和可靠的 访问。保障可用性是指银行系统和网络能够在崩溃后以一种安全并且快速的方式恢复，使银 行日常经营活动不受到太大的负面影响。

系统可用性经常受多方面因素的影响，如设备或软件故障、人为失误、黑客和病毒攻 击、环境因素和自然灾害等。银行信息系统建设过程中，应充分考虑环境及系统的高可用 性。建立完备的应用系统灾备和恢复体系，制定完备的系统灾备恢复预案，以及日常及特殊 情况下的演练机制等，是保障系统可用性的重要手段。

**数据安全** **第11** **章**

**11.1.2** **数据安全的重要性**

保障银行数据的安全性，其核心就是保证银行数据的机密性、完整性及可用性。银行在 国家、社会及民众生活中具有极其重要的作用，因此银行应时刻关注银行数据及信息系统的 安全。在计算机及信息技术高度发达的今天，银行的数据安全尤其具有重要的价值及意义。

**1.** **安全攻击形势多样化要求银行重视数据安全**

随着信息技术发展和广泛应用， 一方面，银行数据作为银行敏感信息的重要载体，已成 为维持社会经济活动和生产活动的重要基础资源，成为政治、经济、文化、军事乃至社会任 何领域的基础。随着银行对信息系统的依赖性不断增强，信息技术在提高银行日常经营水平 和工作效率的同时，也增大了银行的风险——当重要数据信息受到严重侵扰和破坏时，银行 将面临资产损失、业务中断、社会声誉受损等一系列风险。

当今社会，银行信息系统既面临着计算机欺诈、恶意破坏、自然灾害(火灾、水灾) 等大范围安全威胁，又面临着计算机病毒、计算机攻击和黑客非法入侵等破坏，同时还面临 着内部人员误操作或有意攻击造成的内部安全风险的威胁。此外，随着计算机技术的发展和 深入应用，各种威胁变得越来越错综复杂，相互交织，彼此间的界限也变得越来越模糊。在 世纪之交时，全球大范围爆发的“爱虫”病毒给全球用户造成了数百亿美元的损失。美国 每年因信息与网络安全问题所造成的损失高达上百亿美元。国外某权威机构调查显示：计算 机攻击事件正在以每年64%的速度增加。

然而，从世界范围已知的银行数据安全事故来说，银行普遍面临且能造成巨大损失的风 险主要来源于银行的内部。国外统计结果表明，在企业信息受到的损失中，70%是由于内部 员工的疏忽或有意造成的。

相信大家都听说过“温水煮青蛙”的故事。青蛙能逃离沸水，却会死于温水文火!这 个故事告诉我们，身处危险之中却浑然不知是非常可怕的事。对银行而言，在业务领域，造 成危机的许多诱因就潜伏在日常的经营管理当中，缺乏完备的管理机制、规范的流程或者管 理人员的麻痹大意均可能将各种微小事故累积成无法挽回的严重损失；同样，在数据安全领 域，缺乏数据安全意识和没有充分重视，也可能导致与银行内部数据安全相关的一些小事情 形成“连锁反应”、“滚雪球效应”和“恶性循环”,最终演变成从内部摧毁银行的危机。

当今银行的业务及技术使用者很少受过全面、严格的信息安全培训，却依然每天都在以 各种不安全的方式处理大量的银行数据信息。此外，银行的合作方或外包公司的外部人员也 都可能以不同的方式使用银行数据，或者访问银行的信息系统。他们都对银行数据信息的安 全构成了潜在的威胁。比如，粘贴在桌面或计算机屏幕边便于员工本人记忆系统登录口令的 一张便条，或者过于简单的登陆口令，就足以毁掉花费了大量人力物力建立起来的银行信息 安全体系和系统；心存不满的银行内部员工偷窃、散布和买卖银行或客户敏感数据信息，为 竞争对手提供机密的技术与商业数据等行为和案例，也层出不穷。2005年1月28日，某银 行职工冯某，作为银行一名颇有前途的青年员工，利用职务之便，挪用公款800余万元，导 致210余万元损失，最终被一审以挪用公款罪、贪污罪判处有期徒刑16年。

**2.** **银行数据安全关系国家经济命脉、事关国家安全稳定**

数据安全性、可靠性和有效性不仅是商业银行赖以生存和发展的重要基础，还事关整个

 **银行数据治理**

银行业的安全和国家金融体系的稳定，是保护国家、组织、个人信息安全的基石。在信息和 互联网时代，无论个人、企业还是国家，数据安全的重要性都是前所未有的。它既是个人和 企业行使合法权益的基本手段，也是保障整个信息社会正常运行的先决条件。我国著名的计 算机专家沈昌祥院士指出：“信息安全保障能力是21世纪综合国力、经济竞争实力和生存 能力的重要组成部分，是世界各国在世纪之交奋力攀登的制高点。”如果没有数据安全，大 至国家的经济体制与秩序安全、金融与货币安全、产业与市场安全、战略物资与能源安全 等，小至民众的日常生活、人际交往、金融交易等，均得不到有效的安全保障。

我国对银行信息科技风险管理日益重视，各监管机构均已对银行信息科技风险管理提出 明确要求。2009年3月，中国银行业监督管理委员会颁布了《商业银行信息科技风险管理 指引》(以下简称《指引》),从信息科技治理、信息科技风险管理、信息安全、信息系统开 发测试和维护、信息科技系统的运行和业务的连续性管理、外包、内部审计、外部审计等方 面，对商业银行信息科技风险管理工作提出了全面要求。监管部门提出的这些管理措施对各 银行不断改进和完善信息科技风险管理工作具有十分重要的指导意义，充分体现了我国政府 对银行业信息科技风险管理的高度重视。

数据安全是银行可持续发展的基础，商业银行的正常运作离不开数据安全，同时数据也 是银行业务创新的源泉。银行重要敏感数据信息的泄露或信息系统日常业务的中断，小则会 使银行失去核心竞争力、失去市场、失去客户、失去银行资产与社会声誉，大则会造成国家 日常金融秩序混乱，影响社会的正常和稳定。

同样，人们日常生活也与银行数据安全息息相关。保存在银行系统中的个人客户身份基 本信息、交易信息、身份鉴别信息等敏感信息，如果被泄露或非法利用，就会对客户个人隐 私、个人资产、个人名誉等造成损害。

一个真正意义上的银行信息安全防护体系，需要包括有效的信息安全管理体系、健全的 信息安全治理结构、不断创新的信息安全管理策略和信息安全技术产品。这个体系为银行数 据和系统安全构筑了一道坚固的“防火墙”,是银行保障自身的可持续发展，保障国家经济 命脉和社会安全稳定的重要屏障。

**11.1.3** **数据安全的发展历程**

**1.** **我国信息系统安全的发展**

早在1981年，我国政府就对计算机信息系统安全给予了极大关注，并于1987年8月推 出了《电子计算机系统安全规范》。

20世纪90年代初期，类似数据或信息安全法规和规定还是非常少的，这一局面到1994 年2月18日有了根本转变，这一天国务院颁布了《中华人民共和国计算机信息系统安全保 护条例》,该条例规定了计算机信息系统安全保护的主管机关、安全保护制度、安全监管 等。从1994年起，我国信息安全法律法规体系进入了初步建设阶段， 一大批相关法律法规 先后出台，包括《计算机信息网络国际联网安全保护管理办法》《金融机构计算机信息安全 保护工作暂行规定》等。

2000年12月28日，《全国人民代表大会常务委员会关于维护互联网安全的决定》规定 了一系列条款，禁止相关人员利用互联网危害国家、单位和个人合法的权益。该规定更加重

**数据安全**|**第11** **章**

视网络及互联网的安全，也更加重视信息内容的安全。以其为标志，我国信息安全建设进入 了一个新的阶段。

2003年7月22日，国家信息化领导小组第三次会议通过了《国家信息化领导小组关 于加强信息安全保障工作的意见》,标志着我国信息安全法律体系的建设进入一个更高的 阶段。该意见明确了加强数据信息安全保障工作的总体要求和主要原则，确定了后续工 作重点，包括实行信息安全等级保护，加强以密码技术为基础的信息保护和网络信任体 系建设，建设和完善信息安全监控体系，重视信息安全应急处理工作，加强信息安全技 术研究开发，推进信息安全产业发展，加强信息安全法制建设和标准化建设，加快信息 安全人才培养，增强全民信息安全意识等，使得我国信息安全法律体系的建设进入了目 标明确的新阶段。

**2.** **国外信息系统安全的发展**

发达国家关注计算机安全是从20世纪60年代开始的。瑞典早在1973 年就颁布了《数 据法》。这是世界上首部直接涉及计算机数据安全信息问题的法规。随后，丹麦等西欧国家 都先后颁布了数据法或数据保护法。

美国早在20世纪80年代就对计算机安全保密问题开展了一系列有影响的工作。针对窃 取计算机数据和危害计算机信息系统的种种行为，1981 年美国成立了国家计算机安全中心。 该机构1983年公布了可信计算机系统评价准则，1986年制定了计算机诈骗条例，1987 年制 定了计算机安全条例，1999年制定了信息技术安全评价通用准则 (CC)。2003 年，美国国家 安全局又发布了信息保障技术框架 IATF4.0 版本。

在亚洲，日本政府对计算机信息系统安全也相当重视，1984 年日本金融界成立了金融 工业信息系统中心，1985年制定了计算机安全规范，1986 年成立了计算机安全管理协会， 1989 年日本警视厅又公布了《计算机病毒等非法程序的对策指南》。

**11.2** **数据安全领域及管理**

银行数据安全基本覆盖数据存储、数据传输、数据使用等领域，是数据安全治理的重点 领域。

**11.2.1** **数据存储安全**

银行信息系统作为银行数据存储的重要载体，是保障银行数据存储安全的重要基石。银 行数据存储安全主要是保障银行信息系统的物理安全、系统安全以及存储数据的安全，加强 对生产环境的管理。

**1.** **物理安全**

银行信息系统物理安全保护的目的主要是使存放计算机、网络设备的机房、信息系统设 备和存储数据介质等免受物理环境、自然灾难、人为操作失误和恶意操作等各种威胁的攻 击。银行信息系统的物理安全是银行数据安全体系的“基石”。没有物理安全，其他任何安 全措施都毫无意义。

 **银行数据治理**

物理安全主要包括：物理位置的选择、防火、防雷、保障电力供应、电磁防护、机房访 问控制、防盗窃和防破坏等方面。

**(1)物理位置的选择** 主要是在初步选择系统物理运行环境时进行考虑。银行系统机 房对安全性具有极高的要求。即使发生地震、水灾、雷击等自然灾害和设备故障，机房也必 须具备很高的可靠性，以保证业务不会停止。

**(2)防雷** 主要是指防止雷电对机房系统电子设备造成损害，包含建筑物防雷、电源 系统防雷和信号系统防雷等。

**(3)防火** 主要考虑采取各种措施防止火灾的发生，以及火灾发生后能够及时灭火。 做好防火工作需要从设备灭火、建筑材料防火和区域隔离防火等多方面考虑。

**(4)保障电力供应** 稳定、充足的电力供应是维持信息系统持续正常工作的重要条件。 许多因素威胁到电力系统，最常见的是电力中断和波动。重要系统可配备备份供电系统，以 备不时之需。

(5)电磁防护 主要是对信息系统设备的电磁信号进行保护，确保用户信息在使用和 传输过程中安全。

(6)机房访问控制 主要是对内部授权人员和临时外部人员进出银行开发、测试机房 生产环境实施控制，是防护物理安全的第一道关口。

(7)防盗窃和防破坏 主要考虑了系统运行的设备、介质以及通信线缆的安全性。防 盗窃和防破坏的控制点主要侧重在机房内部对设备、介质和通信线缆方面的保护。监视和检 测是防盗窃和防破坏的有效手段。

2. 系统安全

系统安全主要是指银行数据存储信息系统的安全性，主要包括用户身份鉴别、资源授权 访问及检查、服务器安全管理、网络通信安全管理、日志及监控管理等。

(1)用户身份鉴别 用户身份鉴别是指计算机及网络系统确认操作者身份的过程，是 信息系统安全中的一个重要环节，是解决信息系统安全问题时首先遇到的问题。身份鉴别对 保证数据只被合法授权用户获取和访问起着重要作用，因而建立强有力的身份鉴别体系成为 保障各类信息系统安全的关键。

常用的身份鉴别技术包括：用户名/口令鉴别技术、物理介质鉴别技术、 一次性口令鉴 别技术、基于PKI 机制的鉴别技术、生物特征鉴别技术等。

1)用户名/口令鉴别技术。用户名/口令鉴别技术是通过输入用户 ID 和已设定口令进 行身份有效性验证的技术，是最简单、目前应用最普遍的身份识别技术。

2)物理介质鉴别技术。这是通过识别用户持有的物理介质(如磁卡、IC 卡)的有效性 进行身份验证的技术。通常物理介质中记录有固定的静态信息，由合法用户随身携带，登录 时必须将物理介质插入专用的读取设备以读取其中的信息，验证用户的身份。

3)一次性口令鉴别技术。这是一种让用户口令按照时间或使用次数不断变化、每个口 令只能使用一次的身份鉴别技术，是一种与时间同步的动态口令系统。

4)基于 PKI 机制的鉴别技术，即基于数字证书认证方式的认证鉴别技术，可以有效保 证用户的身份安全和数据安全。

5)生物特征鉴别技术。生物特征鉴别技术是用计算机将人体所固有的生理或行为特征 收集起来并进行处理，由此进行个人身份鉴定的技术。目前，研究和使用的生物特征包括脸

**数据安全**| **第11** **章**

部、虹膜、视网膜、指纹、掌纹、手形等与生俱来的生理特征，以及语音、签名、步态等行 为特征。

**(2)资源授权访问及检查** 资源授权访问及检查主要是指：在身份认证成功后，银行 给用户授予相应的访问权限，银行需要通过一定的技术手段，控制用户对数据资源的访问内 容及访问方式等，并对用户访问轨迹进行记录和检查。银行信息系统应通过一定的技术手 段，根据访问用户的角色、权限及相关的安全管理制度规范，对用户访问系统资源的内容及 访问方式进行控制。例如：通过赋予只读、读写或任意权限控制用户访问资源。

由于访问控制属于防范“计算机系统、网络和信息资源”被非授权访问的第一道防线， 所以其地位非常重要。因此，访问控制的实现机制涉及面很广，它涵盖了几种不同的对计算 机系统、网络和信息资源进行访问控制的机制。目前，访问控制方法和技术主要有以下 几种：

1)基于规则的访问控制：这是一种基于特定的规则规定的访问控制方法，属于强制性 的访问控制。

2)限制性的用户接口：这种方法通过不允许用户提交某些功能、信息或访问某些系统 资源的请求，从而限制用户的访问能力。常见的有三种主要的限制性接口：菜单和命令、数 据库视图、物理限制接口。

3)访问控制矩阵：访问控制矩阵 (Access Control Matrix,ACM) 是通过矩阵形式表示 访问控制规则和授权用户权限的方法。

4)访问控制能力表：能力是访问控制中的一个重要概念，它是指请求访问的发起者所 拥有的一个有效标签 (Ticket), 它授权标签表明的持有者可以按照何种访问方式访问特定 的客体。与访问控制列表 (Access Control List,ACL) 以客体为中心不同，访问控制能力表 以主体为中心建立访问权限表。

5)访问控制列表 (ACL): 它是以文件为中心建立的访问权限表。其主要优点在于实 现简单，对系统性能影响小。它是目前大多数操作系统(如 Windows 、Linux 等)采用的访 问控制方式。同时，它也是信息安全管理系统中经常采用的访问控制方式。

6)访问控制安全标签/标记列表：安全标签/标记是限制和附属在主体或客体上的一组 安全属性信息。安全标签/标记的含义比能力更为广泛和严格，因为它实际上还建立了一个 严格的安全等级集合。访问控制标签/标记列表 (Access Control Security Labels List,ACSLL) 是限定用户对客体目标进行访问的安全属性集合。访问控制标签/标记列表最常见的用途是 支持强制访问控制模型。

7)基于内容的访问控制：基于内容的访问控制中，对客体的访问主要取决于客体的 内容。

**(3)服务器安全管理** 服务器安全管理是指银行应用系统服务器在使用过程中自身的 安全配置管理。 一般在系统建设时，要求在登录警告设置、退出系统和远程服务器管理安 全、服务器软件安装更新控制等方面采取相应的管理措施，进行安全配置，提高系统安 全性。

**(4)网络通信安全管理** 网络通信安全管理指的是银行应用系统服务器在使用网络提 供服务时的安全配置管理。 一般在系统建设时，要求在网络服务管理、网络协议管理、网络 链接限制、网络监控等方面采取相应的管理措施，进行安全配置，提高系统安全性。

 **银行数据治理**

**(5)日志及监控管理** 日志及监控管理指的是银行应用系统服务器在使用网络服务时 的安全配置管理。 一般要求在日志设置、警告设置、日志审计等方面采取相应的管理措施， 进行安全配置，提高系统安全性。

加强内部人员对银行应用系统的操作管理已是银行亟待解决的问题，需要引入安全审计 管理实现对操作行为和操作内容的监控，这需要引入行之有效的技术手段，防范内部安全隐 患和风险。

**3.存储数据的安全**

银行数据存储的形式很多，如磁盘、磁带等各种介质。如何保障存储数据的安全是银行 在进行数据存储安全设计时需要重点考虑的内容。其中，数据备份是一种常见的数据存储安 全保障方式。

**(1)数据备份的方式** 一般来说，银行数据备份方式常见的有如下几种：

1)LAN/LAN-FREE 备份，是以网络为基础进行数据传输的备份模式。通常配置一台独 立的主机作为备份服务器，并连接备份磁带库，由它负责整个系统的备份操作。在数据备份 时，备份对象把数据通过网络传输到备份服务器并写入磁带库中实现备份。

2)LAN-FREE 备份，是以 SAN (存储区域网络)为基础进行备份数据传输的备份架构 模式。在 LAN-FREE 备份中，多台主机通过 SAN 共享磁带库，数据流不再经过网络而直接 从磁盘阵列传输到磁带库内，无须占用网络带宽 (LAN-FREE), 从而彻底解决传统 LAN 备 份方式需要占用大量 LAN 资源的问题。

3)数据库冷备份、热备份。冷备份又称为“脱机备份”,是指在数据库保持关闭的状 态下，把数据库的配置文件、数据文件、控制文件、重做日志文件和归档日志文件复制到其 他地方保存起来。热备份又称为联机备份或归档备份，是在数据库保持开放状态下对数据库 进行的备份。热备份的好处是不必关闭数据库，不影响数据库系统的正常运行。

4)全量备份、增量备份。全量备份 (Full Backup), 即对设定的整个目录或数据进行 全部备份，而不考虑它在上次备份后是否被改变。增量备份，又分为累积增量备份和差分增 量备份。累积增量备份只备份自上一次全备份以来变化过的数据。差分增量备份只备份自上 一次备份以来增加、改动过的数据(不管上一次备份是全备份、差分增量备份，还是累积 增量备份)。

**(2)数据备份的基本要素和要求** 数据备份的基本要素包括备份主体、备份目的、备 份通路、备份执行引擎、备份策略等。备份主体是指需要对其进行备份的备份源。备份目的 是指将备份主体的数据备份到何处。备份通路是指数据传输的途径，如 LAN 备份或 LAN- FREE备份。备份执行引擎是指控制数据从源流到备份目的地的控件，通常为备份软件。备 份策略是指备份引擎的工作原则，如备份频率、数据保留周期等。

银行数据备份的要求是，备份主体清楚，备份目的明确，备份通路多样畅通，具有强大 的数据备份执行引擎和完整的备份策略。

**(3)数据备份的一般架构** 一般来说，数据备份的系统架构至少包括：备份主服务器、 介质服务器、客户端。其中，备份主服务器是备份系统的核心，主要承担了备份资源的配置 和管理、备份策略配置和管理、备份作业调度以及备份信息的存储和查询等功能。介质服务 器接受备份主服务器的管理，在备份作业执行过程中，它是备份数据读写的实际执行者，它 连接备份设备(如磁带库),并最终将数据写入其中。客户端接受备份主服务器的管理，在

**数据安全** |**第11** **章**

备份过程中，它将数据以 LAN 或其他方式传送给备份主服务器/介质服务器，并由主服务 器/介质服务器将数据写入备份设备。

**11.2.2** **数据传输安全**

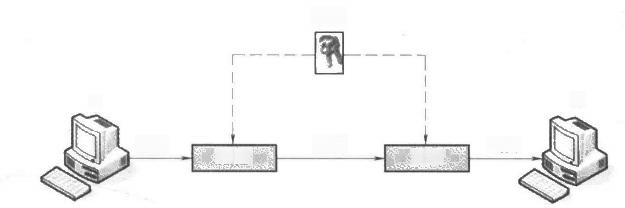
数据传输过程中的安全管理一般涉及数据的加密和数据网络安全控制等。

**1.** **数据的加密**

加密技术是一种重要的安全保密措施，是最常用的安全保密手段。数据加密就是对明文 (未经加密的数据)按照某种密码算法(数据的变换算法)进行处理，从而形成难以理解的 密文(经过加密的数据)。即使密文被截获，入侵者(或窃听者)也无法理解其真正的含 义，从而防止数据信息泄露。加密技术在银行应用广泛，最常见的方式是将客户明文密码加 密后以密文形式存储。

加密技术包括算法和密钥这两个要素。算法是将消息(普通的文本或者可以理解的信 息)与秘密数字(密钥)结合，产生不可理解的密文的变换步骤；密钥用于对数据进行编 码和解码。在数据安全保密中，可通过适当的加密技术和管理机制来保证网络的通信安全， 从而保证传输数据的安全。数据加密的技术分为两类，即对称加密和非对称加密。

**(1)对称加密** 对称加密采用了对称密码编码技术。它的特点是加密和解密时使用相 同的密钥。发送方使用密钥和加密算法加密数据，接收方用同样的密钥和相应的解密算法来 解密。加密和解密算法是互逆的。发送方 A 与接收方 B 利用对称加密技术进行保密通信的 过程如图11-1所示。



密钥

B

明文

解密

加密

明文

密文

A

图11-1 利用对称加密技术进行保密通信的过程

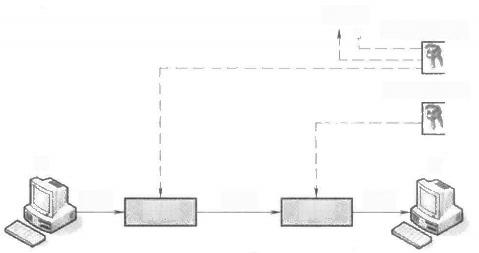
对称加密的典型代表算法包括：数据加密标准 (Data Encryption Standard,DES) 和国 际数据加密算法 (International Data Encryption Algorithm,IDEA)。使用对称密码算法的效率 比较高，因此在对长消息进行加解密时是很好的选择。它的缺点是要求事前提供一条安全的 渠道，在通信双方首次通信时协商一个共同的密钥。

(2)非对称加密 与对称密码算法不同，非对称密码算法需要有两个密钥：公开密钥 (Public Key) 和私有密钥 (Private Key)。公开密钥与私有密钥是一对，如果用公开密钥对 数据进行加密；则只能用对应的私有密钥解密；如果用私有密钥对数据进行加密，那么只有 用对应的公开密钥才能解密。因为加密和解密使用的是两个不同的密钥，所以称为非对称加

密。非对称加密的典型代表算法有 RSA 算 法 ( 以 Ron Rivest,Adishamir,Len Adleman 三 位

学者名字命名的算法)和椭圆曲线密码编码学 (Elliptic Curves Cryptography,ECC) 算法。 发送方A 与接收方B 利用非对称加密技术进行保密通信的过程如图11-2所示。

 **银行数据治理**



给公众

 B的公开密钥

B的私有密钥

A

明文

解密

加密

密文

明文

B

图11-2 利用非对称加密技术进行保密通信的过程

非对称密码算法的保密性比较好，要管理的密钥数量比较少，它消除了最终用户交换密 钥的需要，但复杂度却很高，加密和解密花费时间长、速度慢，不适于对文件加密而只适用 于对少量数据进行加密。

**(3)散列算法** 也称为单向散列函数、哈希算法或消息摘要算法。它通过把一个单向 数学函数应用于数据，将任意长度的一块数据转换为一个定长的、不可逆转的数据。这段数 据通常叫作消息摘要。消息摘要代表了原始数据的特征，当原始数据发生改变时，重新生成 的消息摘要也会随之变化，即使原始数据的变化非常小，也可以引起消息摘要的很大变化。 因此，消息摘要算法可以敏感地检测到数据是否被篡改。消息摘要算法再结合其他算法就可 以用来保护数据的完整性。

散列算法主要应用在一些只需加密而不需解密的场景，如验证数据的完整性、口令表的 加密、数字签名等。目前常用的散列算法有 MD5 算法和SHA-1 算法等。

**(4)数字签名** 数字签名 (Digital Signature) 就是附加在数据单元上的一些数据，或是 对数据单元所做的密码变换。这种数据或变换允许数据单元的接收者用以确认数据单元的来 源和数据单元的完整性并保护数据，防止他人伪造。数字签名技术是非对称密码算法和散列 算法的典型应用，其主要功能是保证数据传输的完整性，确认数据来源身份，防止交易中的 抵赖和欺诈。

**(5)数字证书** 数字证书引入的主要目的是保证互联网上电子交易及支付的安全性， 防范交易及支付过程中的欺诈行为。数字证书属于网上建立的一种信任机制，它由CA (Certificate Authority) 证书授权，是网络中唯一的身份识别标志，类似于个人身份证号码。

数字证书颁发的过程一般为：用户首先产生自己的密钥对，并将公共密钥及部分个人身 份信息传送给认证中心，认证中心在核实身份后，将执行一些必要的步骤，以确信请求确实 由用户发送而来；然后，认证中心将发给用户一个数字证书，该证书内包含用户的个人信息 和他的公共密钥信息，同时还附有认证中心的签名信息，用户就可以使用自己的数字证书进 行相关的活动了。数字证书由独立的证书发行机构发布，提供不同级别的可信度。

除上述技术外，常见的还包括数字信封、消息识别码等。

**2.** **数据网络安全控制**

银行网络安全为信息系统在网络环境的安全运行提供支持。 一方面，确保网络设备的安 全运行，提供有效的网络服务；另一方面，确保通过网络所传输数据的机密性、完整性和可 用性等。网络环境是银行抵御外部攻击，保障银行数据安全的主要防线，需要加强各方面的

**数据安全**| **第11** **章**

防护。对网络安全的保护， 一般来说，要关注两个方面：共享和安全。开放的网络环境实现 了银行数据的内外共享，但是也提供了各种外部攻击的窗口。因此，银行需要在充分保障数 据安全的情况下实现最大程度的数据资源共享。 一般来说，数据传输的网络安全包括网络结 构安全、网络访问控制、入侵防范和网络设备防护等方面。

**(1)网络结构安全** 网络安全的全方位实现需要关注整个银行网络的资源分布、架构 合理性等。只有银行网络结构安全了，才能在其上实现各种技术功能，达到网络安全保护的 目的。例如：银行机构是由多个业务部门组成的，各部门的地位、重要性不同，部门所要处 理的信息重要性也不同，需要对整个网络进行子网划分。

**(2)网络访问控制** 对于网络而言，最重要的一道安全防线就是边界，边界上汇聚了 所有流经网络的数据流，必须对其进行有效的监视和控制。目前进行网络访问控制的方法主 要有： MAC 地址过滤、VLAN 隔离、访问控制列表和防火墙控制等。同时，还需要加强对 边界的完整性检查，做好边界的防御工作。

**(3)入侵防范** 网络访问控制可以认为是网络安全的第一道闸门，起着大门警卫的作 用，但是它对网络内部发生的事件却无能为力。基于网络的入侵检测是防火墙之后的第二道 安全闸门。

一般来说，网络入侵的对象主要包括安全漏洞、口令攻击、协议漏洞、缓冲区溢出、拒 绝服务、被滥用的合法工具、不正确的系统维护措施、低效的系统设计和检测能力等。

网络入侵防范的方法或措施主要包括：入侵检测、控制互联互接、恶意代码防范、内部 安全管控等。

1)入侵检测主要对网络或计算机系统的活动进行实时监测，为网络安全管理提供有价 值的信息，主要包括审计追踪、网包分析、实时活动监控等方面。

2)控制互联互接策略的重点是：在满足服务要求的前提下，通过在网络架构设置障碍 点并把收集信息的相关模块集中部署，来减少数据信息的移动。

3)恶意代码防范主要是对诸如僵尸程序、病毒等恶意代码的防控，防范工作重点主要 在于网络边界。

4)内部安全管控主要关注源自于内部的故意泄露、误操作或恶意攻击风险，建立有效 的认证机制和风险防控机制，例如电子证书、智能卡以及加强对可疑活动的监控措施；建立 完善的安全审计体系，利用支撑的安全监控技术手段，切实落实完整的安全责任查证和认定 机制。

**(4)网络设备防护** 对网络安全的防护，除了对网络结构、网络访问控制等采取相应 的安全措施外，另外一个重要的方面就是对实现这些控制要求的网络设备进行保护。由于通 过登录网络设备对各种参数进行配置、修改等，都会直接影响网络安全功能的发挥，因此， 网络设备防护主要是对用户登录前后的行为进行控制。

**11.2.3** **数据使用安全**

根据我国有关监管部门相关定义，银行数据中包含的客户信息主要包括客户的身份信 息、财产信息、账户信息、金融交易信息、信用信息、鉴别信息、衍生信息等。银行在办理 业务过程中，可采集获得大量客户信息，相关信息一般在以下几类典型应用场景中得到处理

 **银行数据治理**

和使用。

1)业务人员根据业务办理需要，在业务应用系统中按权限访问和使用客户数据。比 如，客户征信信息查询、营销过程中个人目标客户筛选和客户活动名单查询、信用卡风险管 理及反洗钱过程中的客户信息查询/账户信息查询等。

2)银行业务部门将部分客户数据通过系统对接或人工方式提供给合作机构开展业务外 包或其他业务合作等。典型场景包括银行卡、个人金融等业务部门提供相关客户信息给合作 机构，以便开展相关的业务营销；此外，还包括银企对账、信用卡违约透支合作催收、与合 作方开展的积分兑换活动等。

3)在日常生产运维中，根据业务部门的申请，信息科技部门在生产环境中查询及提取 部分客户信息并反馈给业务部门；或者为解决相关生产问题，需要使用生产数据进行分析研 究的场景等。

4)为了满足应用系统研发测试的需要，银行信息科技部门获取脱密后的生产数据，进 行相关的研发、测试工作等。

针对上述典型数据使用场景，目前，商业银行可以通过以下方式加强对数据使用的安全 管理。

**1.** **加强从业务系统层面进行控制，防范非授权访问和下载打印客户数据信息**

通过严格的用户分级授权，实现只有获得授权的业务人员才能使用业务系统中涉及客户 数据处理的功能，并且对授权业务人员的操作记录日志作为事后审计依据。在此基础上，针 对提供批量查询/打印/下载客户数据功能，而且信息泄露风险相对较大的业务系统，通过科 技部门与业务部门携手制定客户数据分级保护策略；进一步采取措施，对于展示查询结果的 页面禁止复制、打印和保存页面等操作；对于查询生成的电子文件，实施加密授权控制，避 免这些信息在下载后未经授权传播，降低信息泄露的风险。

2. 部署客户端安全控制工具，建立完整的客户端信息防泄露机制，防范将客户端上存 储的个人客户信息非授权传播

1)利用客户端安全控制工具，对包括客户敏感数据信息在内的各类敏感信息进行加密 存储和严格授权访问，并在后台实现对打印、还原、复制等行为的记录和审计检查功能。

2)利用客户端安全控制工具，实现对U 盘等移动存储介质的硬性控制。通过控制客户 端上只能使用专用U 盘读写，以及专用U 盘上存储的信息无法被行外终端读取，避免信息 随意复制到行外。

3)利用客户端安全控制工具，定期对客户端存储文件进行扫描，对于可能涉及客户敏 感数据等各类敏感信息，提示用户及时进行加密授权，并通过通报考核等手段加以督促。

4)部署专用监控工具，实现了对外发送邮件以及打印、刻录内容的自动检查，对于涉 及敏感数据的行为自动进行阻断或提示。

**3.** **建立完善的数据安全管理体系**

完善的数据或信息安全管理体系一般涵盖完备的数据安全规范制度体系、职责明确的数 据安全管理架构和有效的数据安全审查机制。

关于数据或信息安全管理体系， 一般商业银行借鉴国内外的信息安全标准，如信息安全 管理体系 (ISMS) 系列标准(即27000系列标准)和业界最佳实践，构建一套具有银行自 身特色的数据安全规范体系，以指导和推动银行数据安全体系建设，保障数据及信息系统

**数据安全**| **第11** **章**

安全。

**(1)建立数据安全规范制度体系** 一般来说，商业银行具有完备的数据或信息安全制 度体系，包括基础安全规范、安全管理制度和安全技术规范三个层面。数据或信息安全管理 制度体系分为三层结构：总体安全策略、具体管理制度、各类操作规程。

1)总体安全策略应当阐明管理层的承诺，提出管理数据信息安全的方法。

2)具体管理制度是在数据信息总体安全策略的框架内，为保证安全管理活动中各项管 理内容的有效执行，专门制定的数据或信息安全实施规则，以规范安全管理活动，约束人员 的行为方式。

3)各类操作规程是为进行某项活动所规定的途径或方法，是有效实施信息安全政策、 安全目标与要求的具体措施，即技术规范。

**(2)组建数据安全管理组织机构** 数据安全管理组织机构是指在组织中负责数据或信 息安全管理职能的机构。安全管理首先要建立一个健全、务实、有效、完善的数据安全管理 组织机构，明确机构成员的安全职责，这是数据安全管理得以实施和推广的基础。安全管理 机构必须首先得到管理层的支持，其职责应该在政策中明确定义，并直接向管理层报告。

安全管理机构的工作职责通常包括根据组织情况，制定与组织发展相符合的安全管理要 求；就安全管理方面的工作与机构内外部的相关组织进行沟通与协调；在组织内落实安全管 理制度所涉及的工作职责；针对安全措施的落实情况进行检查；对发现的问题督促整改等。

安全管理机构应对各种岗位的职责进行明确的定义。为达到对机构数据信息安全工作进 行有组织、有目的的管理，还需要设置专门的安全管理职能部门或岗位，其主要工作职责是 负责具体工作的落实。同时，由于上层信息安全战略或方针的确定需要机构领导层全面把握 和决策，因此，还需要设立数据或信息安全领导小组来负责信息数据安全工作的总体走向和 未来发展。

此外，还需要组建一定数量的安全管理团队。各种数据安全工作需要具体的人员来负 责，因此，首先要保证具有一定数量的基本岗位工作人员，而且为了加强对关键事务的管 理，还应设置专职的安全管理人员。另外，数据或信息安全岗位的人员配备，应当遵循职责 分离的原则。职责分离是威慑和预防欺诈或恶意行为的一种手段。

**(3)建立有效的数据安全审查机制** 为保证信息安全方针、制度能够正确得到贯彻与 执行，及时发现现有安全措施的漏洞和脆弱性，管理职能部门应定期组织相关部门人员按照 数据安全审核和检查程序进行安全核查。检查的主要内容包括：现有数据安全措施的有效 性、数据安全配置与数据安全策略的一致性以及数据安全管理制度的落实情况等。

**4.** **对于生产及研发测试过程中使用的各类敏感数据进行严格管理**

1)银行各类业务系统的数据均存储在生产环境中，为了保证生产环境中数据的安全 性，生产环境中各类存储介质均应由专人保管，故障磁盘更换后要集中进行消磁或物理损毁 处理，从而确保生产数据安全。

2)对于根据业务需求需要在生产环境查询客户信息的情况，应规范数据审批流程，必 须由需求部门提出申请并审批后才能实施查询，查询操作由技术人员通过专用工具实施，查 询结果通过专用FTP 服务器或专用邮箱反馈给业务部门指定联系人，完成传输后数据要进 行清理；同时，对于查询日志等进行记录和事后审计。

3)针对银行研发测试过程需要使用的生产数据情况，生产环境数据只有脱密后才能使

 **银行数据治理**

用。生产数据脱密方法一般包括：制定规范的数据脱密相关规范或操作手册，开发统一的脱 密执行工具等。

数据脱密需要建立完善的规范或流程。商业银行应从总行层面制定数据脱密相关规范和 操作手册。规范或手册应针对各类数据使用场景，制定规范的数据脱密审批流程和可操作的 执行流程，并明确相关的管理职责。

一般来说，数据脱密过程包括数据申请使用、数据使用审批、数据脱密、脱密后数据的 交换和使用等过程。

① 数据申请使用是指生产数据的需求方根据实际业务的需要，向银行的生产中心申请 使用生产数据的过程。数据申请应有规范的流程，并在相关规范中予以明确。

② 数据使用审批是指数据相关提供或管理人员，针对数据提出方的数据使用需求、相 关规范或制度，以及根据应用实际的数据情况，评估申请的数据是否需要进行脱密。为了便 于数据提供或管理人员评估，宜建立客户敏感数据信息的清单。

③ 数据脱密是指当数据提供或管理人员评估认为申请的数据含有敏感数据时，应依据 相关流程，使用数据脱密工具对生产数据执行脱密过程。

④ 脱密后数据的交换和使用是指数据完成脱密后，由数据提供或管理人员按照一定的 方式提供脱密后的数据的过程，以及数据后续的使用过程。

研发统一的数据脱密工具是实现数据脱密的关键。数据脱密工具应以统一的数据脱密策 略为指导，在满足业务或研发测试需要的情况下，根据各类系统的特点，以自行研发或购买 等方式制定覆盖各类系统平台的、统一的数据脱密工具。统一的数据脱密策略应充分考虑满 足业务需求，同时保证脱密数据的不可逆和脱密效率等因素，覆盖境内外主机和平台应用。

**5.** **严格与外单位合作中的个人客户信息安全管理**

对于与外单位业务合作中提供银行数据敏感信息的情况，应依据相关的对外数据提供和 使用规范，进行严格的审批管理，业务主管部门必须提交相关的申请，供管理或科技部门对 提供数据的内容及技术保护措施进行审批与审核；对于提供的数据信息， 一般通过系统对接 或者使用工具对文件加密后进行传输，并通过保密协议等约束合作机构在使用银行敏感数据 信息过程中的保护责任。

**11.3** **案例**

案例一：敏感数据的处理

近年来，信息技术的飞跃发展，市场竞争越来越激烈，客户习惯越来越差异化，导致应 用系统更新升级和新应用功能推出的频率越来越快。以某商业银行为例，除每季度、月度有 新的应用系统版本投产外，还有各类特殊或紧急的版本投产。在如此频繁的应用开发和版本 投产过程中，系统测试，特别是模拟生产环境的测试，能够提早发现问题，有效保障版本投 产的质量。

模拟生产环境的测试，最重要的一个部分就是使用真实的生产数据来进行真实场景测 试。然而，上至国家，下至各行业协会，对数据安全均有严格的监管要求。2009年3月， 中国银行业监督管理委员会颁布了《商业银行信息科技风险管理指引》 (以下简称《指

**数据安全**| **第11** **章**

引》),从信息科技治理、信息科技风险管理、信息安全、信息系统开发测试和维护、信息 科技系统的运行和业务的连续性管理、外包、内部审计、外部审计等方面，对商业银行信息 科技风险管理工作提出了全面要求。那么,银行应如何加强对敏感信息数据的管理?如何对 敏感数据进行脱密?

首先，应建立全行级的信息安全相关规范。规范应明确敏感数据的定义与范畴，明确在 银行各类场景的敏感数据管理方法和流程，如研发、测试、生产问题处理等；建立由总行至 分行，横跨各信息科技管理部门的信息安全管理和技术支持架构，以及明确各自的职责等。

其次，为了满足应用版本投产的质量，需要利用有效手段，对敏感数据进行脱密，将脱 密后的生产数据纳入各类测试流程。而数据变形作为数据脱密的有效手段， 一方面确保了敏 感信息不被泄露，有效防范信息安全风险；另一方面，以其高效的执行效率，满足了测试数 据要求，也确保了测试数据准备不影响测试周期及相关安排。

数据变形需要以数据变形策略为指导，以变形工具为实际手段，确保数据变形的易用 性。数据变形策略至少应明确数据变形所涵盖的敏感信息范畴，对敏感信息类型进行明确的 划分；要有明确的数据变形算法，针对各类敏感信息类型，在满足业务测试需要的前提下， 以简单、不可逆等为原则，注重变形效率。变形工具开发应基于变形策略，充分考虑各类平 台差异，可有效参考业界成熟产品，当然也可以自行进行研发。数据变形往往需要配置具体 的变形参数配置表，用于制定具体表、字段含有的敏感信息类型，以便准确定位需要变形的 敏感信息。数据变形需要在工具执行操作、变形参数配置表等方面考虑易用性和系统化管 理，减少开发人员梳理的工作量。

当前，数据变形在测试数据准备过程中的应用也较为广泛，是数据脱密较为常用的手 段。以某国有商业银行为例，还建立了全行级的数据变形管理手册，用于指导数据变形工作 的开展；还建立了跨总分行、各机构的数据变形管理团队，并组织自行研发了适合各类不同 平台的变形工具。目前，经过变形工具脱密后的大量数据广泛应用于各类季度或月度版本的 各类测试中，在满足测试需要的同时，有效保障了信息安全。

案例二：个人客户信息的保护

个人客户信息包括了身份信息、财产信息、账户信息、金融交易信息、信用信息、鉴别 信息、衍生信息等，对其进行保护是商业银行信息安全工作的重要组成部分。某商业银行已 经建立了相应的制度和措施，以有效保证客户信息的安全，确保不会被泄露和窃取。

**1.** **技术硬控制**

技术硬控制措施主要包括两个方面： 一是在业务应用系统层面进行控制，在应用系统开 发的技术规范中针对业务系统处理的个人客户信息，从用户访问授权、敏感信息加密保存和 传输、页面展示信息及电子文件保护、与合作方互联的信息保护等方面制定了详细的技术控 制措施要求，并在系统中实现了相应的控制功能；二是为了防范将业务应用系统的查询结果 下载到行内用户本地计算机上后未经授权扩散到行外，在客户端上部署了一系列硬控制措 施，具体情况如下：

**(1)在业务系统安全控制方面** 通过严格的权限控制、加强业务系统页面展示和下载、 加强日志记录和审计等措施避免业务系统在未经授权的情况下传播个人客户信息，降低信息 泄露的风险。其中包括：

1)通过严格的用户分级授权，实现只有授权业务人员才能访问业务系统中所在辖区

 **银行数据治理**

(如： 一级分行、二级分行、支行或网点)的涉及个人客户信息处理的功能，并且将用户的 操作记录日志作为事后审计依据。

2)针对提供批量查询、打印、下载个人客户信息功能，而且信息泄露风险相对较大的 总分行业务系统，会同相关业务部门进一步采取严格的控制措施，按照分级保护的思路，对 展示查询结果的页面禁止复制、打印和页面保存等操作，对查询生成的电子文件实施加密授 权控制。

**(2)在客户端安全控制上** 通过部署客户端安全控制工具(包括文档加密、邮件拦截、 网站访问限制等)建立起完整的客户端信息防泄露机制，防范将客户端上存储的个人客户 信息非授权传播。其中包括：

1)部署电子安全控制工具，对包括个人客户信息在内的各类敏感信息进行加密存储和 严格授权，并对打印、复制等行为的记录和审计进行检查和核实。

2 ) 对U 盘等移动存储介质进行硬控制(包括：行内客户端上只能使用本行专用U 盘读 写，专用 U 盘上存储的信息无法在行外终端上读取),避免信息被随意复制到行外。

3)部署信息防泄露专用工具，每月定期对客户端存储文件进行扫描，对于可能涉及个 人客户信息的各类敏感信息，提示用户及时进行加密授权，并通过通报考核等手段督促。

4)部署专用监控工具，实现对外发送邮件以及打印、刻录内容的自动检查，对于涉及 敏感信息的行为自动进行阻断或提示。

**2.** **加强管理**

**(1)在管理制度建设方面** 在保密管理以及信息安全管理两方面制定相应的制度和 要求。

1)在制定的《商业秘密保护办法》中，明确了个人客户信息属于商业秘密信息(商密 二级)以及具体的保护要求。

2)在制定的《信息及信息系统安全管理办法》及其实施细则中，从业务系统用户安 全、业务系统信息分级保护、信息防泄露与终端安全、对外合作客户信息保护等方面制定了 涉及个人客户信息的安全管理流程和管理策略。

3)各专业部门结合本专业工作内容和特点，分别制定了本专业涉及个人客户信息保护 的管理办法和工作规程，包括《个人客户信息管理办法》 《私人银行业务保密管理规定》 《个人客户营销管理系统应用管理办法》《信用卡中心信息安全相关管理制度》《信用卡违约 透支合作催收管理办法》等，并下发通知，从客户授权、客户信息适用业务范围、对外合 作中的客户信息保护等方面对个人客户信息的信息安全管理提出了明确要求。

**(2)在日常管理方面** 针对风险相对集中的领域，分别采取以下管理措施：

1)对个人客户信息异常操作进行监控。部署操作风险监控系统对个人客户信息异常操 作行为的监控，包括同一柜员单日多次查询同一账户、柜员N 日累计查询同一账户、柜员 无正式依据查询客户账户余额、柜员频繁查询不活跃账户、柜员频繁查询未启用卡片、 一定 时期内柜员频繁无依据查询客户、个人不活跃账户频繁查询后异常支付等模型。 一旦监控到 此类准风险事件，将进行核查和处理。该商业银行某年内柜员频繁查询不活跃账户、个人不 活跃账户频繁查询后异常支付两类事件共报警300 多次，核查后的实际异常操作行为有几十 次，有效保护了个人客户信息。

2)对向第三方机构提供的个人客户信息进行严格审批管理。对于提供给监管机构和合

**数据安全**| **第11** **章**|

作单位等第三方机构的个人客户信息，根据相关要求，业务部门均需向同级信息安全审核部 门提交申请。同级信息安全审核部门对提供信息的内容及技术保护措施进行审批审核，并根 据制度要求和审批意见，采取“限制提供信息的范围、使用通过安全检测或符合规范的设 备、信息在加密或变形后提供、与合作方签订保密协议”等措施保护本银行个人客户信息 安全性。

3)对测试研发过程中使用个人客户信息数据以及日常生产运维进行严格管理。制定 《数据管理办法》《数据管理办法实施细则》等规章制度，明确相应的管理要求，对于测试 研发过程中需要使用的涉及个人客户信息的各种数据，都需要进行申请和审批，并通过专用 工具对信息中的关键字段进行脱密处理后供测试研发使用。

4)在生产运维过程中，通过实施生产环境与互联网/办公/测试环境的网络区域隔离、 访问权限控制、对外传输限制等措施，确保生产区域的数据安全。对于磁带库中的数据， 一 方面数据恢复都要通过提交变更待审批通过后才能执行，另一方面，磁带在机房中需要由专 人保管，出库有严格流程控制，磁带外借有专人运输，避免发生磁带丢失事件。对于生产环 境所有查询下载客户信息的情况，必须由业务部门提出申请并审批后才能实施。运维人员使 用专门查询平台进行数据查询，查询结果都经过加密后通过敏感数据专用FTP 服务器或安 全产品邮箱提供给业务部门。



**第12章**

**数据服务**

银行通过数据治理，保证了数据作为重要资产的价值。数据是重要资产的观念已经在金 融行业成为共识，但要发挥数据资产的作用，还需要一个对数据进行加工处理、分析挖掘的 过程。价值并不是简单地浮现在数据的表面，数据内在的规律、知识才是其最为核心的价 值。此外，在大数据时代下，银行业随着信息化建设以及互联网金融冲击，数据也呈现出了 爆发性增长、多样化演变的趋势，原有的小系统、结构化数据的抽样、小规模分析已经不能 满足发展的需求，急需借助大数据战略打破数据边界，囊括声音、影像等非结构化数据的加 工处理，来构建更为全面的银行全景视图。

与其他行业相比，大数据对银行更具有潜在价值。麦肯锡公司的研究显示，金融业在大 数据价值潜力指数中排名第一。这主要是因为， 一方面，大数据决策模式对银行更具有针对 性。发展模式转型、金融创新和管理升级等都需要充分利用大数据技术、践行大数据思维。 另一方面，银行具备实施大数据的基本条件： 一是数据众多，银行不仅拥有所有客户的账户 和资金收付交易等结构化数据，而且拥有客服音频、网点视频、网上银行记录、电子商城记 录等非结构化数据；二是银行拥有处理传统数据的经验；三是银行业较高的薪酬水平能够吸 引实施大数据的人才；四是充分的预算可以利用多项大数据新技术。

因此，银行应从大数据就是大资产的高度，尽早制定大数据战略，研究、应用大数据，分 享大数据所带来的利润。业务发展对大数据在数据服务时效、分析挖掘能力，以及对行内行 外、线上线下多样化海量数据的融合与综合应用等方面提出更高的要求。因此，夯实大数据基 础是利用大数据推动业务发展的源泉和基石。大数据分析基础性建设工作，是在一定的数据规 范要求下，通过数据质量和数据标准管控，不断丰富基础数据的来源，扩展处理数据的类型， 将银行数据整合并纳入大数据基础平台，依托云服务，提供集成型的数据服务和各类分析挖掘 工具，实现业务数据的集成与共享，满足不同时效性的分析需求。为适应大数据时代海量数据 分析挖掘的需求，银行应从优化数据布局、提升数据集成处理能力、强化数据管控等方面提升 对结构化数据和非结构化数据的集成和管理能力，为分析挖掘工作奠定坚实的数据基础。

**12.1** **银行在大数据时代的战略应对措施**

发展模式转型、金融创新和管理升级是未来5～10年内我国大多数银行的战略目标。围

**数据服务** **第12章**

绕这一战略目标，银行大数据应用应该在以下方面取得突破性进展。

**1.利用大数据，推动发展模式的战略转型**

随着利率市场化和民营银行设立预期的加剧以及互联网金融的兴起，银行业竞争日 趋激烈，利差进一步缩窄，银行纷纷进行发展模式的战略转型。经营模式从“以产品为 中心”向“以客户为中心”转型，营销模式从“粗放营销”向“精准营销”转型，服务 模式从“标准化服务”向“个性化服务”转型。实现战略转型目标要求银行必须可靠、 实时掌握客户的真实需求，“全面完整描述客户的真实面貌”,离开这一前提，转型目标 就成了空中楼阁，虽然好看，但是很难落地实施。银行员工想在客户买房时推荐住房贷 款，想在客户留学时推荐留学贷款，想在客户一走进银行时就对客户说：“您的孩子已经 成年，可以办一张信用卡吗?”但这一切必须建立在银行充分掌握客户信息的基础上。而 银行目前了解客户的渠道非常有限，信息量低、时效性差，这是目前银行转型并不成功 的重要原因。

大数据技术的出现，正逢其时。通过广泛收集各渠道、各类型的数据，利用大数据技术 整合各类信息，还原客户的真实面貌，可以帮助企业切实掌握客户的真实需求，并根据客户 需求快速做出应对，实现“精准营销”和“个性化服务”。现在已经有了大量成功案例。美 国沃尔玛公司充分利用天气数据，研究天气与商品数量增减的关系，根据飓风移动的线路， 准确预测哪些地方要增加或减少何种商品，并据此进行仓储部署，确保产品能够及时满足消 费者需求。花旗银行新加坡有限公司根据客户的刷卡时间、地点，结合客户的购物、餐饮习 惯等个人信息，可以精准地向客户推荐商场及餐厅优惠信息。西班牙对外银行 (BBVA) 推 出的 ATM 机 ABIL, 不仅比传统的 ATM 机更安全、更便利，还具有记忆功能，不但能记住 客户习惯的取款金额、取款频率，还能够根据客户的账户情况给出取款建议。

**2.** **利用大数据，推进风险决策模式的创新**

银行也是具有经营风险的行业，风险管理是银行的生命线，风险管理的成败决定着银行 的存亡。目前，银行在进行信用风险决策时，主要依据客户的会计信息、客户经理的调查、 客户的信用记录以及客户抵质押担保情况等，通过专家判断进行决策。这种决策模式带来了 以下严重后果：

一是这种模式只适用于经营管理规范、会计信息可靠、信用记录良好的大公司或有充分 抵质押物并经营良好的中小公司，而对于占企业总数97.3%的4200 万小微企业并不适用， 这导致银行失去了大量的有价值客户；

二是决策基本上取决于信贷审批人员的主观判断，缺乏足够的客观证据，信息不对称、 标准不统一，引发了前台客户经理和后台信贷审批人员的激烈矛盾，致使业务流程复杂、效 率低下；

三是决策依据的主要是企业过去的静态信息，而不是实时的动态信息，这种信息的时效 性、相关性和可靠性不足，风险不能够得到有效控制。

事实上，大数据技术可以有效地解决上述问题。 一方面，通过多种传感器、多个渠 道采集数据，可以帮助银行更全面、更真实、更准确、更实时地掌握借款人信息，有效 降低信息不对称带来的风险。另一方面，利用大数据技术可以找到不同变量间新的相关 关系，形成新的决策模型，使决策更准确、更统一、更公正。因此，银行完全可以利用 大数据创新出新的风险决策模式，并由此赢得新的客户，形成新的利润增长点。Zest Cash

 **银行数据治理**

就是一个成功的案例。Zest Cash 是一家小额贷款公司，由Google 的前首席信息官 Douglas Merrill 创办。Merrill 发现，银行在进行贷款决策时，决策的依据往往非常有限，主要依靠 客户的信用记录。信用记录有瑕疵或根本没有信用记录的客户往往无法获得银行贷款。 Merrill 说：“为什么要仅仅依赖一个数据来决定人的信用等级呢?为什么不用网络上那成 百上千的数据，再设计一个算法，来判断一个人是否会按照约定来还钱呢?”经过努力， Merrill 通过大数据技术成功地解决了这一问题。他把从网络上搜集的几千条碎片化的数据 整合成完整的客户拼图，较为准确地还原了客户的真实状况和实际信用状况，并因此获 得了丰厚的利润。我国的阿里金融也是一个典型的案例。阿里金融利用阿里巴巴 B2B、 淘宝、支付宝等电子商务平台上客户积累的信用数据及行为数据，引入网络数据模型和 在线视频资信调查模式，通过交叉检验技术辅以第三方验证来确认客户信息的真实性， 向这些通常无法在传统金融渠道获得贷款的弱势群体批量发放“金额小、期限短、随借 随还”的小额贷款。重视数据，而不是依赖担保或者抵押的模式，使阿里金融获得了向 银行发起强有力挑战的核心竞争力。

**3.** **利用大数据，促进管理升级**

近20年来，银行的管理升级行动层出不穷，全面质量管理 (TQM)、 业务流程管理 (BPM) 、业务流程再造 (BPR) 、ISO 质量管理、平衡计分卡、关键绩效指标 (KPI) 、 经济 增加值 (EVA) 、6S 、 六西格玛、核心竞争力等各种管理理念和管理模式在银行轮番应用。 就像一位管理大师所说：“一些管理者像15岁的女孩子一样，流行什么就追求什么。”在很 多管理升级行动实施后，企业的绩效却并未因此得到提升。根据1993年的调查，世界500 强企业中65%的企业宣称已经进行或者将要进行业务流程再造。然而，2001年的一份统计 研究显示，70%的业务流程再造项目以失败而告终。管理升级失败的一个重要原因是：很多 管理理念或管理模式是在缺乏有效数据支持的情况下，仅凭管理咨询公司或企业领导人 “拍脑袋”决策推行的。 一些看起来很漂亮、在其他企业也成功的理念和方法却并不适用于 本企业。

大数据的本质特征之一是在决策模式上与传统模式不一样。大数据强调决策应建立在牢 固的数据证据基础上，应通过对比、实验等方法对现有的认知规律进行检验，去伪存真。银 行应利用大数据技术和思维，对相应的管理理念和管理模式进行对比、检验，以找到真正能 够解决本企业问题的管理利器，实现管理升级，提升企业利润。国际连锁便利店7-Eleven 就 是一个很好的案例。7-Eleven在进行管理升级时，管理者一开始认为，对于服务行业来说， 提升店员的礼貌服务就会提升企业的利润，但他们并未随意地“拍脑袋”决策，而是进行 了一次实验。他们设立了百万美元的大奖，重奖那些对顾客礼貌服务的店员，并通过培训提 高店员的礼貌程度。这些措施实施后，原来只有33%的顾客在进门时受到了招呼，而实验 中这个比例提升到了58%;原来只有32%的顾客受到了微笑服务，而实验中这个比例提升 到了49%。但是，实验结果却表明，让店员更加礼貌地服务反而降低了店面的销售量。经 过仔细分析和调查后，他们找到了原因——对于顾客来说，最在意的是能否快速地找到自己 想要的商品并快速离开，而对于店员是否对他微笑并不在意。过于重视店员的礼貌程度，反 而会降低店员的工作效率，引起客户的不满。

要实现以上目标，准确地进行战略决策、持续地推进和鼓励创新、技术的创新利用是必 不可少的。银行将会从下面几个方面来实现大数据的价值发现。

**数据服务** **第12章**

**1.制定与银行整体战略相融合的大数据战略**

银行将制定符合自身实际的大数据战略，明确大数据应用的方向，制定大数据实施路线 图，落实大数据预算。银行大数据战略应融入银行的整体战略，而非独立于整体战略之外。 只有将大数据战略与银行整体战略进行有机统一，才能保证大数据成为银行发展的核心 要素。

**2.** **以持续改进的方式应用大数据，避免颠覆性创新**

银行的创新模式与互联网企业不同，属于持续改进型的创新模式。银行创新往往由监管 政策变化、新竞争者出现、科技进步等外部原因引发，并以较为缓慢的步骤实施，很难像互 联网企业那样进行自发的颠覆性创新。银行人的通常思维定式是：通过传统的方式就可以挣 钱，为什么要冒险进行创新呢。在美国商业周刊2010年4月的创新公司排名中，金融企业 在全球最具有创新能力的前50家企业中只占据4家，表现最好的摩根大通集团也只列第39 名。因此，银行在应用大数据时，要充分考虑到银行自身的特点和银行人的思维定式，以 “小步快跑”的方式进行持续改进，避免采取颠覆性的创新方式。

**3.** **设立专门的大数据实验室或创新实验室**

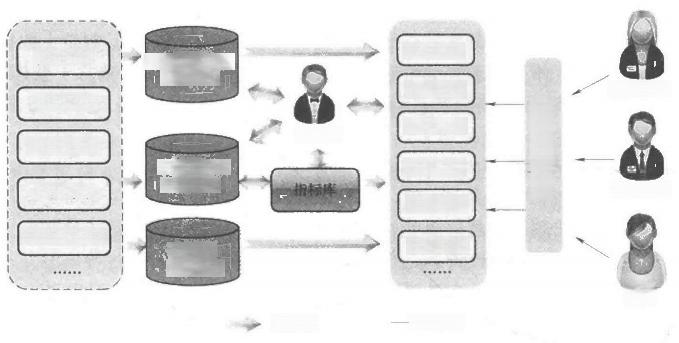
我国银行业将借鉴美国银行的经验，设立创新实验室或大数据实验室，实验室应包括业 务、管理、科技、统计等各方面的人才。实验室统一负责大数据方案的制定、实验和推广。 在推行大数据方案前，实验室应先进行实验，实验完成后，要对项目的风险和收益做出有大 数据证据支持的评价。不要以非大数据的思维方式进行大数据方案推广。

互联网、物联网、即时通信工具、社交网络等的兴起和普及，特别是大数据技术的应 用，正深刻改变着当前的金融生态和金融格局。从国外的 Wonga 、Lending Stream 、Zest Cash 、Klarna 、Pawn Go 到国内的阿里金融、百度金融， 一股“银行如果不改变，我们就改 变银行”的飓风席卷全球。2012年1月，达沃斯世界经济论坛发布的《大数据，大影响： 国际发展的新可能》的报告宣称，大数据已成为与货币和黄金一样的一种新的经济资产类 别。2012年3月29日，美国总统办事室 (EOP) 公布了《大数据研究和发展计划》,把大 数据研发应用从商业行为提升到国家战略层面。在这种新形势下，商业银行何去何从，如何 “应势而变”,怎么以未来5～10年为界限从战略高度研究应用大数据，已成为当前的迫切 任务。

**12.2** **构建大数据服务体系**

数据创造价值，信息引领未来。作为信息革命的新高潮，互联网和大数据技术改变了传 统商业模式，也改变了消费者的习惯，甚至带来了商业理念和金融文化上的一系列变化。未 来，优秀的银行首先应该是数据银行，更是数据分析、数据解读的优秀银行。经过多年的信 息化建设，银行已经积累了大量的客户信息、交易记录、信用数据、财务信息等金融数据， 初步形成了大数据的基础。目前，大型商业银行基本都已建立了统一的处理结构化数据、非 结构化数据、流数据的存储和加工平台，利用数据分析挖掘技术，在客户服务、风险管理等 领域开展了大量卓有成效的运用，从而满足银行发展方式和经营机制上深层次变革的需求。 银行大数据服务体系典型架构如图12-1所示。

**银行数据治理**



客户营销

市场分析

贷款业务

价值发现

互联网信息

风险评估

社交媒体

决策支持

电子商务平台

流数据

存储加工

*客户*

操作

非结构化数据 存储加工

结构化数据 存储加工

多种渠道接入

管理人员

信息中介

客户经理

存款业务

分析师

信息流

图12-1 银行大数据服务体系典型架构

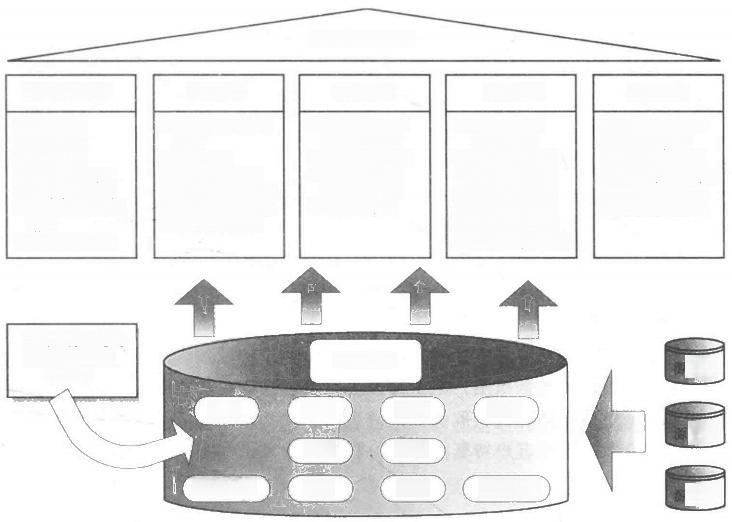
大数据服务体系的战略目标是：以数据为根本，以业务为导向，通过对大数据的集中、 整合、挖掘和共享，实现对多样化、海量数据的快速处理及价值挖掘，利用大数据技术支持 产品快速创新，提升以客户为中心的精准营销和差异化客户服务能力，增强风险防控的实时 性、前瞻性和系统性，推动业务管理向信息化、精细化转型，全面支持信息化银行建设。

**12.2.1** **结构化数据处理系统**

当前银行业数据处理与应用主要针对结构化数据，其过程就是数据仓库的发展过程。数 据仓库的概念大约在20世纪70年代到80年代形成，在理论层面上基本认同了以下观点： 应当把那些新出现的、不可以预测的、大量存在的分析型负载从业务处理系统中剥离出来， 采用专门的体系架构和设计来进行处理。数据分析和业务处理存在相当大的差异，以至于需 要用具有不同体系结构的系统来分别处理它们，这就导致了数据仓库的出现。 一个典型的数 据仓库服务体系如图12-2所示。

数据仓库平台的建设实现了企业异构数据的集成，企业按照分析主题重组数据，建立了 面向整个企业的一致的信息视图。在此基础上，结合数据挖掘、联机分析等技术，为实施精 准化营销、有效控制企业风险。实现经营资源的优化配置等提供了数据基础，全面提升了数 据的共享水平，深入挖掘了数据的业务价值，大大增强了决策的科学性。数据仓库中存储的 数据，需要进一步转化为能实际应用的知识，用于指导后续工作，这样才能充分发挥其数据 的价值。数据挖掘利用了来自统计学的抽样、估计和假设检验，人工智能、模式识别、机器 学习的搜索算法、建模技术和学习理论，通过数理模式来分析企业内储存的大量资料，以找 出不同的客户或市场划分，分析出消费者的喜好和行为。数据挖掘需要数据库系统提供有效 的存储、索引和查询处理支持，源于高性能(并行)计算的技术在处理海量数据方面常常 是重要的。因此，数据仓库往往和数据挖掘技术密不可分，而数据挖掘技术在数据仓库海量 数据的支撑下，更能发挥其重要作用。通过数据仓库中的海量数据，建立一定的数据挖掘模 型，用于预测客户的产品需求和行为特征，作为后续营销、决策的依据，这将是数据仓库后 续主要的发展方向之一。

**数据服务** 第 1 2 章

业务远景与目标

客户关系管理

√ 客户细分

√ 客户获取

√ 客户挽留

√ 客户净值提高 √ 目标市场营销 √交叉销售、向上

销售

√ 营销活动管理

财务管理

√监管报表、信息

披露

√ 总账分析

√ 财务盈利分析

√ 财务预算与计划 √ 预测与场景分析 √ 动态财务报告

绩效管理

√产品、机构(部

门)绩效考核

√ 渠道绩效考核

√ 员工(客户经

理)绩效考核

√ 业务量考核

风险管理

√ 合规与披露

(Basel Ⅱ)

√ 信用风险管理

√ 市场风险管理

√ 操作风险管理

√ 全面风险管理

√ 资产负债管理

√ 反欺诈与反洗钱

信息管理

√ 企业数据整合

√ 元数据管理

√ 数据质量管理

√ 数据有效期管理

√ 数据安全管理

√ 数据模型管理

面向业务主题的

逻辑数据模型

(LDM)

协 议

营销活动

数据仓库

|  |  |
| --- | --- |
| 渠 道  *客户*  事件 | 地域 产 品  财务 |

机 构

客户资产

数据

激据

数据

图12-2 典型的数据仓库服务体系

最早将数据分析应用于银行业务的是美国一家银行的信用卡部门，他们在20世纪70年 代借助计算机技术将统计分析用于申请审核和风险控制。到20世纪80年代末期，数据仓库 开始成为向银行决策人员和分析人员提供商业智能的工具。美洲银行是较早进行尝试并获得 成功的先行者。20世纪90年代，随着财务分析在银行战略地位的不断演变，财务分析型数 据仓库开始被接受和采用。美国的银行业20世纪90年代中后期基本上确立了将集中的数据 管理系统，即数据仓库，纳入银行信息系统体系架构中。进入21 世纪后，银行与客户之间 发生了直接交互的前端业务，银行操作人员越来越需要能收到决策模型和客户轮廓信息，用 于帮助他们处理日常工作，这极大地推动了同时分析历史数据和当前数据的复合型需求的 发展。

近年来，我国大多数商业银行 一般按照“数据集中化、业务综合化、管理扁平化、服 务网络化、决策科学化”的理念进行银行信息系统体系建设。 一般包含两大数据平台： 一 个是基于数据大集中的策略，面向金融业务数据处理，构建高效、统一 的核心业务处理平 台；另一个是面向分析处理，构建完整、 一致、反映时间变化的数据仓库。我国银行最早从 2000年起着手开展数据仓库的研究，逐步建立了以企业级数据仓库为基础的结构化数据的 采集、存储和应用体系，实现了客户信息、账户信息、产品信息、交易信息、管理信息等内 容的集中管理，用于海量数据挖掘和分析的数据仓库平台数据总量达几百太字节 (TB), 包 含客户、柜员、各有关机构的交易行为、关联性和交易习惯等大量有价值的信息。依托数据 仓库平台，提供通用查询等数据查询类工具、灵活查询等数据探索挖掘类工具和搜索引擎等 信息定制类工具，并搭建分析师工作台提供工具集、模型管理的功能，基本满足银行进行用 户数据分析与挖掘的需求。

 **银行数据治理**

在大数据时代下，要实施有效的精准营销、风险管控和经营分析，银行不仅要知道客户 “做了什么”,而且要知悉“客户为什么这么做”,“客户的真实意向是什么”。因此，数据 的全面性就显得尤为重要，只有获取全面的数据，才有可能从这些数据中解读出最完整、最 准确的信息。传统银行已经实现了大量客户信息和交易数据的积累，可进一步丰富完善新型 业务数据、行外征信数据等结构化数据。

数据仓库不仅要具备数据挖掘、分析的能力，而且需要具备将这些数据挖掘、分析结果 贯穿到业务流程中实际支持决策的能力。因此，未来的发展上，数据仓库应能够提升数据服 务的时效性，可以基于对实时数据和历史数据的分析，跨越核心业务数据平台和数据仓库平 台，形成两个平台之间的实时互动，指导业务操作。随着银行自助渠道的发展，非结构化数 据也将逐步增大。据统计，至2012年，非结构化数据占有比例将达到互联网整个数据量的 80%左右。如果企业结合利用非结构化数据，将能大大扩展企业所掌握的信息和知识，效果 可想而知。在这样的发展形势之下，对非结构化数据利用的处理需求也越来越多。对于数据 仓库，必须扩展其处理的数据范围，至少可以支持基于结构化和非结构化两类数据融合的数 据分析。例如，通过解析从外界获取的资讯信息(一般为非结构化数据),根据资讯信息的 关键字分析，结合数据仓库客户特征的分析，将合适的资讯信息推荐给需要的目标客户，为 客户提供更多的增值服务。

**12.2.2** **非结构化数据处理系统**

随着互联网的逐步发展，数据每天都在以惊人的速度增长，信息的种类也在不断地扩 展。IBM公司的研究部门调查发现，全球大多数的信息是非结构化的，包括纸上的文件、报 告、视频文件、音频文件、照片、传真件、信件等。由于内容的表现形式存在多样性，人们 在对其进行管理和利用方面耗费了大量时间，所以迫切需要有效地将非结构化数据管理 起来。

传统数据仓库以银行结构化数据为主，进行异构数据的整合、加工，进而提供服务。在 大数据时代下，新兴的数据类型不断涌现，更多地呈现出一种非结构化的状态，传统关系型 的数据仓库难以满足此类数据的存储及分析需求，因此涌现出了Hadoop 、Spark 等新型分布 式数据库，相应地，单纯的数据仓库服务体系架构也逐渐向大数据服务体系架构演变。各银 行纷纷启动面向海量非结构化数据的数据加工处理与分析系统的建设，并注重传统银行数据 与新兴互联网金融数据的综合运用。

银行经营过程中产生和使用的数据主要分布于数据源系统、数据集成系统、分析型系统 三类系统中。随着网络技术的发展、商业银行的业务发展以及客户接触渠道的多样化，银行 的数据种类和数据量剧增，银行存储的客户数据、交易数据等结构化数据，以及电话语音、 微博、即时通信等非结构化、半结构化数据都在飞速膨胀，尤其后者的增长速度远远大于前 者。在强调数据即是价值的今天，任何一份数据都是银行珍贵的资产。为了一点金子，需要 保存全部沙子，又如何从沙子里找出这一点金子?这些都是摆在商业银行前面的问题。大数 据加工处理技术则很好地解决了这些问题。

通过非结构化数据存储加工系统，可以做到：

1)利用其提供 PB 级的数据存储，把商业银行业务产生的业务数据都存储下来，实现

**数据服务**| **第12章**

海量数据存储。

…2)利用其对非结构化数据加工处理的优势，实现大数据时代结构化和非结构化数据的 有效融合，发挥数据价值。

3)承担核心系统某些消耗性交易(例如：账号历史数据打印查询功能的交易),让核 心系统更好地处理实时交易业务，充分发挥传统数据库的优势，做到优势互补，保证金融银 行业 IT 信息系统的持续健康发展。

为实现信息价值向经营效益的有效转化，银行需要获取除客户交易信息之外的客户 购买行为、社交行为、外部资讯等信息。 一方面，可通过互联网金融等新业务形态的发 展，进一步采集除客户银行传统金融业务之外的客户物流、资金流信息以及各种客户行 为信息；另一方面可以通过跨行业合作、购买外部数据等方式，引入来自互联网、物联 网、电商等渠道的资讯、社交行为、物流、客户电子交易行为信息等丰富的数据资源。 这些新引入的信息，大多是非结构化数据，可以作为对银行传统数据的补充，使得数据 基础更为全面。

按照信息整合共享的原则，整合各专业数据平台，建立整合共享的数据平台，并通过大 数据技术对结构化、非结构化、内外部数据进行融合处理，实现客户物流、资金流、信息流 的“三流合一”,形成更加完善的客户视图，进而奠定坚实的大数据分析挖掘基础。

**12.2.3** **流数据处理系统**

随着互联网时代的不断发展，银行业务发生了巨大的改变，对大型银行现有的主机+开 放平台系统架构的处理能力也提出了进一步的挑战。

**1.** **随着业务的快速发展，对系统处理的吞吐量提出更高要求**

除了传统的自助渠道、柜面渠道外，近年来新兴的各种电子渠道(如网上银行页面、 平板电脑 APP、手机银行 APP等)的交易量也在飞速上涨。

**2.** **基于主机实现的非核心账务类增值服务消耗主机资源较大**

通常主机分别通过准实时推送模式、定时批量模式实现风险监控、客户营销和客户服务 领域的高时效增值服务功能。

**3.** **随着业务的快速发展，主机账务查询压力也在不断增大**

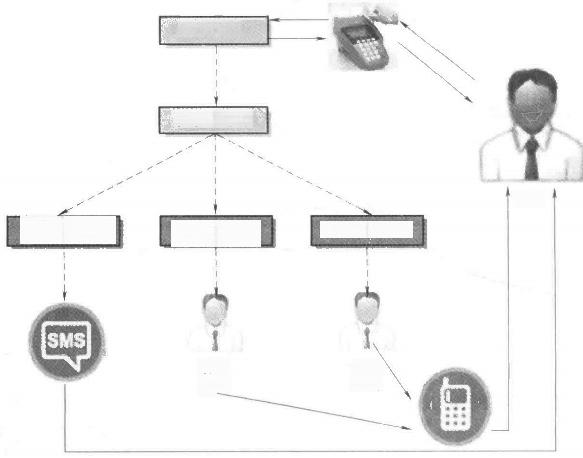
商业银行对部分业务场景的数据处理时效要求越来越高，例如如何在第一时间分析 出用户的需求?如何对欺诈交易进行识别?生产系统如何进行及时有效的监控?这些问 题在流式数据处理之前，商业银行一般短则几分钟，多则大半天才能得知答案。目前， 基于流式处理的技术很好地为商业银行时效要求高、分析简单的应用场景提供了很好的 解决方案。

所谓流式处理技术，是针对流式数据的一种分布式、高可用、低延时、具有自身容错性 的实时计算技术。它是根据一组处理规则来工作的工具。通过提取和分析来自各个系统的信 息，解决企业监控和管理的各种问题。

目前，主流的流数据处理技术主要有Storm 、Spark 、Streaming等，它们的基本原理是类 似的，均对来自各个系统的高频度事件流进行分析，根据既定的一些分析规则，对每个独立 事件进行处理分析，计算得出所需要的服务信息，用于时效要求的服务场景。

**银行数据治理**

典型的流数据处理系统设计思路如图12-3所示。



交易系统

流数据处理

客户

操作风险监控

客户

经理

通知消意平合

个人客户营销

业务 人员

图12-3 典型的流数据处理系统设计思路

现对此典型的流数据处理系统设计思路进行简要介绍。

1)客户通过 POS 等渠道进行交易。

2)流数据处理系统引入大内存、分布式计算等技术实现交易数据的快速处理。

3)流数据处理系统准实时推送服务，广泛应用于余额提醒、事件式营销、风险防控等 交易场景。

4)客户经理、业务人员及时对客户进行营销、提醒等服务。

流数据处理系统主要作为交易系统的延伸，通过对交易数据的采集、存储、计算、分 发、持久化等环节的快速处理，承担交易过程中涉及的风险控制、增值类数据加工处理服 务，为客户营销、风险监控、客户服务等业务系统提供异步准实时交易信息处理和推送 服务。

通常流数据处理系统提供两种不同的服务方式， 一种是逐笔对交易信息进行处理的 “流式”服务模式，另一种是一次性对一批交易信息进行处理的准实时批量服务模式。两种 服务模式具体的处理模式、技术特点、时效、设备资源和服务领域见表12-1。

**(1)“流式”服务模式** 这种模式需要逐笔对交易报文进行处理，经过规则计算，并以 联机报文的方式对外提供服务。此类服务模式适用于对时效要求非常高(秒级响应)的业 务功能，典型的应用场景包括：余额提醒、商户圈信息推送、信用卡交易监控(反欺诈) 信息筛选、柜面客户大额资金变动后的触点式营销等。

**(2)准实时批量服务模式** 这种模式是对一段时间内发生的一批交易报文进行统一的 处理，通过 Spark 将存储在 HDFS 中的信息抽取出来进行统一计算，主要以批量文件的方式 对外提供服务。此类服务模式适用于分钟级响应的业务功能，典型的场景包括：客户购买理 财产品后营销信息推送、大额资金流转至第三方存管的营销、风险监控(分钟级时效)、交

**数据服务** **第12章**

易监控统计、分行特色销售统计等。

**表12-1** **两种服务模式的比较**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 处理模式 | 技术特点 | 时效 | 设备资源 | 服务领域 |
| “流式”处理 | 逐笔交易处理， 时效较高 | 秒级 | 集群设备多， 配置较高，管理 复杂度较高 | 对客服务(余额提醒)、触点式营销(高时效、需要营销人 员当面服务客户的营销功能)、风险监控(反欺诈) |
| 准实时 批量处理 | 分钟级批量处理 | 分钟级 | 集群设备少， 管 理 复 杂 度 较 低 | 客户营销(非高时效)、风险监控(发卡、商户收单、逸贷 等)、交易监控统计、分行特色功能 |

**12.3** **数据资产视图**

银行陆续建立了数据仓库等大数据基础平台，并依托大数据基础平台提供分析挖掘服 务，支持业务人员在决策分析、精准营销、风险管理、客户服务等领域开展工作。随着大数 据技术应用的不断深入，建立基于用户视角的数据资产统一视图，实现准确高效的数据展 现、导航、应用，进一步提升数据分析能力的需求也日益凸显。

如何构建元数据查询与管控服务，在银行大数据技术应用过程中回答各类用户“数据 是什么”、“数据在哪里”的问题，成为银行数据资产管理工作面临的一项新课题。

**1.** **数据地图**

为回答在银行大数据技术应用过程中，各类用户“数据是什么”“数据在哪里”“数据 怎么用”的问题，银行应考虑针对数据信息构建面向用户视角、可视化、高效、智能、易 用的“数据地图”服务，帮助用户对所关心的数据信息进行快速定位以及便捷快速地获取。

数据地图可通过建立标准化接口，从银行数据管理系统或者各业务系统获取数据信息， 实现数据信息的自动采集，并应提供目录导航、数据搜索、数据画像、血缘分析等功能。

**(1)目录导航** 支持多维度的数据分类展示方式，并可基于用户视角进行个性化定制， 方便用户理清数据。

**(2)数据搜索** 对采集的数据信息进行加工与管理，形成语义信息，提供模糊搜索、 高级搜索、结果筛选、关键词推荐等智能化搜索功能，帮助用户快速定位数据。

**(3)数据画像** 提供数据名称、存储周期、数据结构、数据字典、关联标准等数据信 息，为用户形成完整的数据画像信息。

**(4)血缘分析** 支持上下游血缘分析，并以可视化的方式把数据从源系统到最终接口 指标的整个路径给刻画出来，辅助用户定位数据问题。

**(5)构件服务** 以构件形式提供给其他系统调用，供各类用户(例如信息管理人员、 数据分析人员、技术开发人员)在工作中使用。

**2.** **指标库**

银行经营中要使用大量的统计指标，针对指标元数据信息，如果缺乏完整的业务管理流 程，将会在应用中出现指标虽然多，但是可用性不高的问题。例如：缺乏指标的一致性管 理，将出现同名异义指标、指标元数据信息不完整、指标积累越来越多、退出困难等问题。

 **银行数据治理**

对此，银行应考虑建立全行统一指标库，提供指标元数据管理视图，完善指标管理 功能。

1)在指标实现上，采用先建立标准再开发指标的流程，使指标标准与指标库形成对 应，实现指标元数据对指标加工的硬性控制。

2)建立指标统一管理视图，完善指标库信息服务功能，建立指标、数据标准以及相关 指标信息的映射或关联关系，实现指标的快捷查询和检索。

3)按照指标新增、应用、维护和退出的管理流程，建立指标生命周期管理机制，在不 同阶段制定相应的管理措施，特别是指标的退出策略，让指标“有始有终”。

**3.** **客户信息统一视图**

近年来随着信息化银行建设的推动，部分银行通过引入外部数据源、深化推进客户信息 质量治理，积极推进个人客户信息在各业务系统中的整合与共享，建立统一的客户信息视 图，并基于该视图衍生出各专业系统的专业客户信息视图，推进客户信息视图在营销、信 贷、风险控制、经营管理等领域的深化与应用，提升客户服务和经营管理信息化水平。

银行新系统建设应该坚持以客户为中心进行系统架构设计，建设以客户信息系统为中 心，客户关系管理、统一客户授信、统一评价等信息集成与加工，推进业务由基本信息直接 应用转型到定性分析、定量分析、综合评价信息的集成视图相结合应用，不断提升信息服务 支持能力。

客户信息一般指从客户、合作机构(外部信息源)直接采集的客户静态基础类信息 广义上包含为客户服务时产生的动态业务类信息及对客户的分析评价关联类信息。客户评价 指标用于发现价值客户，推进客户细分，根据客户评价分析模型对客户贡献、产品、资产负 债信息与行为信息进行加工分析，形成各维度客户动态评价指标结果。客户信息视图是各类 客户信息的集成，主要包含客户静态信息和动态评价指标信息，能让信息使用者更加全面地 了解和分析客户。

这里以个人客户信息视图集成为例进行简要介绍。银行依托对数据仓库的分析、挖 掘与加工，推进个人客户统一评价，生成客户评价指标，与客户基本信息组合形成全行 客户统一视图，供前中后多个业务系统共同使用。随着客户信息标准统一和系统信息集 成的实现，各业务部门也加快了信息应用，推进个人客户统一评价，并逐步形成各专业 领域的信息视图，通常来讲主要包括个人客户信息统一视图、个人客户信用风险视图、 网点业绩视图。

**(1)个人客户信息统一视图** 个人客户信息统一视图包括客户基本信息、客户评价信 息、关系人信息、客户财务信息、账户协议信息、营销活动信息、事件信息、持有产品信 息、信用及风险评估信息等模块，展示个人客户的全方位信息，如图12-4所示。

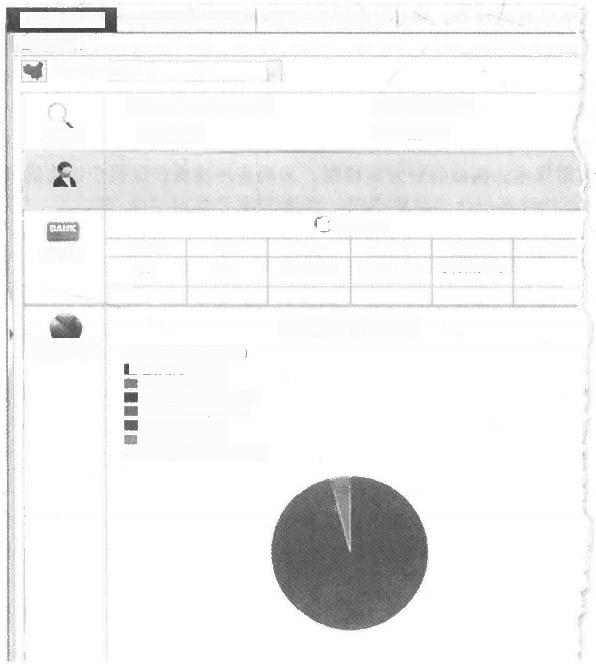
**(2)个人客户信用风险视图** 个人客户信用风险视图包括客户基本信息、融资信息、 关联关系信息、对外担保信息、结算信息、综合贡献等模块，展示个人客户信用方面的全方 位信息。视图样式可参考图12-5。

**(3)网点业绩视图** 通过网点业绩视图展现的个人客户数据，主要从辖内个人客户、 客户经理管的客户及代发工资个人客户角度出发，展现个人客户的资产负债、存贷款、重点 产品持有情况。该视图需要展现的信息如图12-6所示。

**数据服务** **第12章**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 客户概况 | 客户统一视图查询 | | 客户统一视图维护 | | |
| D客户统一视图查湖  客户身份信息  客户联绪信息  客户基本信息 客户偏好信息  客户证件信息 | | | | 客户账号信息  服务协议信息  账户协议信息 卡信息查询  卡交易明细  PP卡信息 | |
| 综合指标  渠道信息  客户评价信息 活跃度信息  信用卡价值信息 | | | 星级信息  要客信息 | 营销活动倍息 | 营销活动倍息  产品推荐信息 |
| 关系人信息 | | 家庭状况信息  客户关系信息 | | 事件信息 | 产品到期信息  提醒信息  服务日志  信用卡办卡进度  卡片自动还款 |

图12-4 个人客户信息统一视图

**客户信息概况**

客户统 一 视图查询

**☑客户极况**

所在地区：北京

**客户统** **一** **视图维护**



评价信思

管户吴属值胞

产品信息

全行总资产：123,456,789.00

星级要富：否

服务网点号：12345

客户经理性别：男

惜记卡

贷款 理财

存款

全行总负债：789.00

信用卡要客：否

服务网点名称：中山公园支行

联系电话：12345678

|  |  |
| --- | --- |
| ☑ 已持有产品  贷记卡 网上银行  第三方存管 余额变动提醒 | 电话银行  手机行)(动 |

网银证书

客户资产结构(万元)

全行资产负债表 存款余额：12,000.00

■基金余额：0.00

焦券余额：0.00

本外币理财余额：0.00

灵通快线余额：0.00

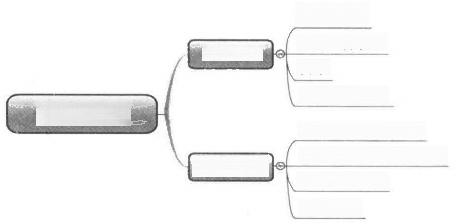
黄金余额：0.00

保险余额：0.00

第三方存管余额：345.79

图12-5 个人客户信息风险视图

银行数据治理



存款、贷款

中间业务收入 贡献

产品渗透情况

个人客户数据

基本信息、客户等级

资产、负债、存款、贷款

产品持有情况 推荐产品

单体分析

群体分析

图12-6 网点业绩视图需要展现的信息

**12.4** **大数据在银行业的应用**

大数据技术正在为每个领域带来变革性影响，成为各行各业颠覆性创新的原动力和助推 器。随着银行服务边界的扩展、客户接触渠道的多样化，银行的数据种类和数据量剧增，银 行业因势而动，利用大数据技术，通过对数据的研究和应用推动自身发展和创新。

银行依托大数据平台可以进行客户行为跟踪与分析，进而获取用户的消费习惯、风险收 益偏好等。针对用户这些特性，银行等金融部门能够准确实施风险管理、精准营销、经营分 析、绩效考核等。

**12.4.1** **国内外银行业大数据应用案例**

**1.** **应用大数据技术，构建客户分析模型，实现客户挽留，提供个性化的客户服务**

招商银行通过构建客户流失预警模型，对流失率等级前20%的客户发售高收益率理财 产品予以挽留，并通过数据分析识别出信用卡高价值客户经常出现在星巴克、DQ、麦当劳 等场所后，通过多倍积分累计、积分店面兑换等活动吸引优质客户。美国富国银行 (WELLS FARGO) 基于大数据定位有流失倾向的客户群，进行流失前的客户挽留。中国银 联利用大数据技术分析持卡人交易行为特点，挖掘分析客户属性，例如性别、职业等，形成 客户“画像”,为客户提供个性化服务。

**2.** **利用丰富的大数据资源和大数据挖掘分析技术，进行风险评估，实现高效的风险管控**

美国 ZestFinance 公司针对信用记录不好或者没有信用卡历史的人，采用大数据技术进 行分析，考察贷款人的数千个信息线索以甄别客户是否具有更高的还款意愿，降低贷款损失 风险。花旗银行通过社交网络、公共网页获得客户的信用记录以及信用历史，应用在信用风 险评估、精准营销等场景。美国富国银行采集客户交易数据和行为数据，建立风险评估模 型，采用批发模式处理小额贷款，超过2/3的贷款决策由系统自动完成，并获得超越同业约 1%的息差水平。

**3.** **基** **于** **各** **类** **结** **构** **化** **和** **非** **结** **构** **化** **数** **据** **，** **利** **用** **大** **数** **据** **采** **集** **挖** **掘** **分** **析** **技** **术** **，** **服** **务** **于** **银** **行** **的** **经** **营** **管** **理**

美国银行利用大数据技术分析该银行某呼叫中心员工姓名牌中置入感应器的信息，分析 发现一起享受工间休息并相互交流的员工工作效率更高，他们可以在日常交流中分享应付

**数据服务** | **第12章** |

“难缠”顾客的小窍门。据此，美国银行推行集体工间休息，此后员工表现提升了23%。中 国光大银行通过数据仓库和大数据分析平台，为分行提供网点预选地址周边的商户情况、客 户结构以及同业结构信息等，通过综合分析评估选址的科学性与合理性。

**4.** **银行业借助大数据在互联网金融方面发挥更大的作用**

互联网与金融的结合越来越密切，互联网企业利用其庞大的用户群，借助大数据、云计 算、社交网络和搜索引擎等信息技术优势，纷纷涉足金融领域，支付宝、财富通等第三方支 付工具成为新的支付手段。到2013年，我国互联网支付规模达到5.37万亿元；小贷、P2P 等网络融资模式，依托云计算等技术促进了审批效率的提升，阿里小额贷款每天可审批贷款 8000笔，京东最短3min 可完成从申请到放款的全流程。

而银行业也在积极通过各种方式利用互联网和大数据技术拓展渠道、提升服务。各大商 业银行纷纷试水自己的电子商务平台，如中国工商银行的“融e 购”、中国建设银行的“善 融商城”、交通银行的“交博汇”,中国银行的“中银易商”;中信银行推出供应链网络金融 平台，基于大数据对客户进行分析，从而降低风险，提供高效的线上服务；兴业银行与百度 公司就互联网金融创新、大数据、产品营销等方面开展战略合作，利用大数据技术布局互联 网金融。

**12.4.2** **银行业大数据应用场景**

**1.在客户营销和服务方面**

银行依托数据仓库和信息库，加强线上线下联动及全渠道信息的整合共享，全面融合并 综合利用行内外各类信息，包括结构化、非结构化的信息，形成全行客户的统一视图；通过 数据深入分析客户的消费偏好，准确发现并预测客户的金融需求，设计有针对性的产品，提 供个性化的服务。

**(1)精准营销** 银行可基于数据仓库实现客户特征的分析，可以根据客户基本信息、持 有产品、交易行为等进行目标客户的分析和筛选，并实现精准化营销的闭环管理流程，为真 正向以客户为中心的转型发展奠定良好的技术实现基础。

某银行依托数据仓库，精确定位1亿多目标客户，成功营销800 多万户，营销效率得到 大幅度提升，有力支持了客户拓展和重点产品交叉销售工作；主动向980万目标客户提供金 融服务关怀活动，有效提高了840万目标客户的忠诚度和归属感。

某银行以理财产品营销为切入点，建立了保本理财客户预测模型。通过提取某分行辖内 2013年7月购买保本理财产品的四星级以上客户，观察其在购买前半年的表现，选取与客 户购买产品紧密相关的信息，通过 SAS EM 中的决策树、评定模型 (Logit 模型)、神经网络 模型分别建立购买预测模型，并从中选择最优模型来计算客户在未来一个月内购买保本理财 的可能性，精准定位目标客户。2013年年底，选取了两家支行作为试点，通过模型，短信 营销客户反馈率较全行平均水平提升了11倍；客户经理营销客户购买率较以往提升4倍以 上，营销成功率显著提升，在预测模型推荐并最终购买理财产品的客户中，有24.4%的客 户以往从未购买过理财产品，有效地实现了客户拓展。

**(2)客户服务** 商业银行每天都会从不同渠道接收到海量的客户心声，包括客户的投 诉、抱怨、建议等，而这些正是银行提高客户服务效率和质量的基础。以前都是通过监测分

 **银行数据治理**

析人员随机抽查来分析客户到底具有什么样的需求，但这种处理方式不仅容易遗漏一些重要 的建议和投诉，而且其庞大的工作量也难以保证这些建议和投诉的合理分类和正确处理。如 何处理这些非结构化数据，才能获得客户的真正需求，才能提高服务质量和降低成本，是银 行使用非结构化数据价值的一个挑战。 一般来说，可以通过下面的步骤和方式来实现客户需 求的深度洞察。

1)对客户意见进行分类：对客户意见进行分类可以分析各类别意见的发生趋势，提升 对应部门处理效率。但是，在分类时要结合具体业务处理流程和实际场景。简单来说，其实 就是要给各类意见打上标签，比如这类意见反映的是哪类产品，如银行卡、自助服务等。

2)对热词进行分析：热词分析就是要从客户反馈中提取最频繁使用且意义较明确的词 汇，从而形成可视化的意见词云，但是有时候无法精准表达客户意见，存在一词多义和多词 一义的情况，例如柜员、工作人员等，因此需要将纷繁杂乱的描述语句进行归并、抽象、提 炼，从而形成统一、概括的表示方式，例如对象-属性-评价。

3)选择贴合业务场景的分析模型：文本分析模型有很多种，商业银行要选择适合自己 的模型进行文本分析和挖掘，如是用朴素贝叶斯模型，还是 LDA 模型?或者抽象出本体模 型。选择出合适的模型后再根据实际情况进行训练。 一般来说，只有文本数量较多时才会使 用合适的模型进行自动化训练。

**2.** **在风险管理方面**

提升行内外结构化与非结构化数据的综合运用能力，深入分析客户行为特征和关联关 系，在客户特征的基础之上，进一步完善客户信用评分评级模型，侦测客户信用风险，对客 户风险等级进行动态调整，实现对客户授信的精细化管理，并将大数据挖掘分析的结果快速 应用于信贷业务处理的前台、中台和后台，实现从大数据中提炼出的有价值风险信息的全面 共享，形成分析和业务办理的联动效应。

商业银行在向以客户为中心的转型过程中，在风险管理领域也带来了一场变更，即：由 被动应对向主动防控转变；由应对监管的强制要求进行风险管理，向真正利用风险管控合理 运作银行资本的战略转变；由局部风险管理，向全球风险监控、全面风险覆盖、全新工具计 量、全程风险管理转变。同时，伴随科技进步，客户获取银行服务的渠道日益多样，银行内 部管理所涉及的信息量将急剧增大，客户和银行可能遭受损失的风险也越来越大。因此，要 求风险管理系统能有新的技术架构帮助应对这样的发展形势。

**(1)信用风险管理** 对于风险管理领域，数据仓库可以支持风险管理领域对客户、债项 层面的内部评级和风险综合评价，支持对风险资产的综合管理，同时将评价结果应用于信贷 审批、风险监测、风险报告、政策制定、拨备计提、资本计量等各领域。中国工商银行基于 数据仓库开展了内部评级、资本管理、不良客户管理等工作，取得了较好的效果。在内部评 级方面，通过客户的基本信息、客户评价信息、客户持有产品等数据，对客户进行风险评 级，细分客户群体，统一了全行卡、贷款等信贷业务的授信策略，实施了针对客户个体的差 异化综合授信，真正实现了以客户为中心的服务理念，提升了信贷服务水平。

例如，某银行基于神经网络实现了信用卡欺诈交易的实时识别和干预，进一步提升了反 欺诈识别准确度，堵截信用卡盗刷、网银转账、手机银行转账和 B2C 支付等业务的风险交 易；某银行运用决策树算法，通过对客户资金流向、履约情况、关联关系等特征进行分析， 及时预测和量化客户信用风险，实现及时对不同的风险资产采取差异化的处置措施。

**数据服务**| **第12章**|

**(2)舆情分析** 在大数据时代的背景下，网络上存在大量以文本为代表的非结构化数 据，特别是互联网上的新闻、社交媒体数据，其中更不乏与商业银行有关的舆情信息。对这 些信息进行有效的挖掘和利用，将为商业银行及时掌握在互联网上传播的潜在风险事件提供 一个全新的视角。

实现舆情风险管理的文本挖掘技术过程一般分为文本获取、预处理、分析挖掘、可视化 展现等步骤。

1)文本获取：商业银行获取文本的方式包括已采购的财经新闻、行业动态、研究报告 等外部资讯信息，还可以通过开发采集工具来进一步扩大新闻媒体的采集范围，比如对以微 博、论坛为代表的社交媒体信息进行采集。采集方式包括搜索引擎(通过搜索引擎进行关 键词搜索，之后将所有结果通过爬网程序进行采集)、新闻网站、论坛页面适配(通过 Web 爬虫程序抓取微博或论坛页面，并从页面结构中解析出正文和评论数据)、微博页面适配 (通过模拟实际用户登录后对微博信息进行采集)。

2)预处理：包括中文分词和文本去重等步骤，从而实现文本的预处理。前者是在获取 到文本数据之后，将文本切分成词汇的集合，使得机器能够更好地理解词汇组成的文本。后 者则通过相似哈希算法快速对海量文本相似程度进行计算：将文档看成特征词的集合，为每 个特征词分配唯一编码；根据特征词的编码以及在文档中的权重，通过相似哈希算法生成文 档的信息指纹(可比较的64位二进制编码);文档指纹完全相同的文档，则认定为内容相 同；指纹间不同的位数越少，则说明文档内容越相似。文本虚词的增减、语句位置的变换将 不会影响近似文本的发现。

3)分析挖掘：通过主题分析、关联分析、情感分析、分类分析等方法实现文本的挖 掘。主题分析主要对分析所涉文本中主要的关键词及相关的主题词进行提取；关联分析主要 对文本进行算法处理，得到文本的数字特征格式；情感分析主要对文本情感进行分析，从而 得出文本的情感基调，判断主观语义；分类分析主要通过分类器进行训练，从而得到不同维 度下分类的概率特性。

4)可视化展示：通过“词云”“词汇网络”“主题河流”“传播途径”等展现方式，展 示不同的数据结果。

一般来说，内控合规所涉及的事件往往是商业银行不可逾越的“高压线”,其中绝大多 数员工违规事件的发生，都会对商业银行产生极大的负面影响和不可估量的损失，因此对于 合规的实时监测就显得尤为重要。合规风险事件监测可以对商业银行内控合规中涉及的受 贿、非法集资、飞单等关注点进行实时监测，自动识别信息正负面以及相关的业务类别，从 而尽早发现，甚至预先判断风险，以及时做出相应的对策。

舆情分析还可以有更多、更广泛的实际应用领域，例如银行产品和服务的口碑监测、危 机事件的预警等。

**3.在经营管理方面**

商业银行需要综合利用自身的大数据信息资产，加强运用大数据进行分析挖掘，从而用 于指导经营分析，实现精细化经营管理，真正“以数据说话”。通过大数据挖掘分析方法进 行数据探索，呈现全行网点运营基本情况，并建立网点运营分析模型，指导全行做好网点运 营标准化实施工作。通过网点运营分析模型的持续监控，掌握网点运营过程中的资源配置、 业务发展、人员管理等运营情况，及时发现问题，做出经营决策，提升和优化运营效率。

 **银行数据治理**

某银行近期实现产品和网点五级分类的常态化展现，指导资源配置，帮助网点进行调整 优化，并将重点关注应用大数据支持网点经营。在了解基层行一线网点信息需求的前提下， 制定信息支持解决方案，加大上下信息交流，并通过主动的信息推送，让一线人员能够及时 掌握市场变化、客户变动情况，灵活采取针对性措施，切实通过信息支持来提高一线的服务 营销能力和市场竞争能力。

某商业银行利用数据仓库中的数据，对已经推出的各种产品进行业务量、市场规模、盈 利能力的统计和排序比较。通过对这些数据进行分析，相应的产品管理人员可以及时知晓银 行产品的市场销售实际情况，适时调配产品管理资源；实现各类产品关键销售数据的排序比 较和分析，使产品管理人员根据分析结果进行绩效管理；帮助业务人员对产品推广中的问题 进行深入分析，寻求改善产品本身及产品推广工作的方法和途径，为有效提升企业产品管理 能力奠定基础。

**4.在绩效考核方面**

从我国商业银行意识到绩效考核评价的重要性开始，银行绩效评价经历了一个由简单到 复杂、由低级到高级、连续演变、梯次推进的发展阶段，但是其根本的宗旨并没有改变，就 是完善考核机制，彻底改变以前由于思想观念陈旧和技能严重匮乏造成的银行绩效评价体系 简单的问题，这样可以客观评价各级机构、员工所创造的效益。通过数据仓库的数据整合能 力和统一的指标加工口径，可以保证绩效考核标准的一致性、客观性，有利于推进绩效考核 的实际推广与应用。

某银行从2010年开始，构建了基于数据仓库的统一的绩效考核平台，并在全行范围内 进行了大力的推广。目前，绩效考核系统已完成了在全行的推广工作，有望借助于该系统使 该银行的绩效管理上升到一个新的台阶。

以上只是简单说明了几个大数据的典型应用场景，在实际应用中，大数据的价值还远远 不止上述这些。外资银行在非结构化数据利用方面已经走在了前面。我国商业银行不仅可以 将非结构化数据应用于上述的业务场景，而且可以在内部管理方面取得突破，例如可以利用 各类开发文档、运行报告发现设计开发普遍存在的问题，从而提高开发效率和质量，还可以 对信贷审批文档进行相似度分析，发现那些信贷报告中存在的问题等。



**第13** **章**

**数据治理展望**

数据治理是一个长期的持续过程，需要有一套长效保障机制。数据治理的方法和手段也 不是一成不变的，应结合实际情况推陈出新。在大数据蓬勃发展的今天， 一方面，将数据治 理理念与新兴技术结合方可获得更大的治理效果；另一方面，通过深化治理效果可以促进大 数据的发展与应用，进而创造出更大的业务价值。

**13.1** **构建银行数据治理的长效机制**

数据治理不是一个临时性的运动。从银行业务发展、数据治理意识形成、数据治理体系 运行角度，需要一个银行数据治理的长效机制进行保证。

数据治理是银行业务开展及管理创新的需要，而银行业务开展及管理创新是一个永恒且 持续向前的过程。银行需要推出各种中间业务、理财服务，即使传统的存贷汇业务也需要创 新业务模式，改善客户体验；银行机构的信贷管理部门需要密切关注贷款分类及客户信息的 变动，以保障资产分类的准确性；市场部门日常的交易更是依赖大量的数据分析；资产负债 管理部门需要通过数据分析，为各部门设定业务限额，在保障安全的情况下实现全行效益最 大化。上述的创新都需要银行机构利用商业智能工具对客户信息和业务数据进行挖掘。

数据治理意识的形成不是一蹴而就的。银行数据治理要从机构上层意识转变着手，强化 数据治理的理念，通过营造数据治理的文化，确保数据治理在全行的落实与执行。这是一个 长期过程，银行机构各层级的数据管理人员必须不断地沟通、培训和推广数据资产价值的重 要性以及数据治理职能的业务贡献，提升数据使用者对数据治理的意识及对数据治理效益的 认可程度，持续改进银行数据管理机制，充分挖掘数据价值，提升银行核心竞争力。

数据治理是一项长期且体系化的过程。银行数据治理要从数据的长远规划与规范管理着 手，逐步制定数据治理的相关管理规范及“先治理，后建设”的数据治理战略，并落实在数 据标准和数据质量的绩效考核指标上。只有合理制定数据治理相关考核指标并切实推进到业务 基层，才有可能从数据产生开始实现对数据的真正治理，避免数据环境出现“先污染后治理， 边污染边治理”的局面。“以用带治”,以数据应用为抓手，通过将数据充分利用于业务管理 和业务经营中，使业务部门切身体会到数据的作用，以及数据问题对业务分析与决策的影响，

 **银行数据治理**

促进数据质量的改造，推动诸如数据标准、元数据管理等数据治理和管理工作的开展。

**13.2** **大数据时代的银行数据治理**

大数据、云计算、移动互联网等技术，将人类社会带入了一个以PB 为单位的大规模生 产、分享和应用数据的新时代。面对如此巨大规模、数据种类多样化的大数据，银行的数据 治理应如何面对?——当治理的对象发生变化时，治理体系也应进行改进以适应大数据的发 展变化!

**1.** **完善数据管控相关标准，提升相关系统控制能力**

对于银行，大数据在数据范围方面，除了涵盖原有传统的客户、协议、账户、事件等结 构化数据外，已经逐步扩展到非结构化数据的存储管理及利用。因此，需要从数据模型、元 数据、数据标准、数据质量、数据生命周期等方面依据非结构化数据的特点，补充相关治理 管控标准，并通过相应的管理系统(如元数据管理系统)实现控制，确保非结构化数据得 到有效管理和使用。

**2.** **利用大数据技术提升数据集成及共享能力**

海量数据给银行的数据治理带来挑战，但是，大数据技术的快速发展，对银行的数据治 理来说也是一种机遇，可使数据治理的方法和手段更加丰富，使数据治理的效果更为显著, 数据的价值可以获得更大的发挥。

信息化建设已成为金融服务创新的重要动力。随着银行经营转型，信息化建设发展也面 临着新的使命和挑战。新时期，大型商业银行只有掌握基于海量数据处理的高效分析体系和 技术，才能在与国内外同业日益激烈的竞争中赢得胜利。建议重点从以下两方面提升大数据 平台服务能力，整合存储各类大数据，实现对多样化、海量数据的快速处理及价值挖掘。

**(1)建立高效、开放、异构、弹性的大数据平台** 以通用、开放、高性价比软硬件设备 为基础设施，采用分布式计算等大数据技术，基于云服务平台的理念，构建面向多种数据类 型的高容量、弹性可扩展、资源可灵活调配、数据高时效的大数据平台，为总分行用户提供 开放式数据分析云服务，实现“全渠道、全客户、全产品”信息的综合分析与快速共享， 提升客户拓展、风险管控和产品创新与营销能力等。

**(2)实现“全、易、准”的数据资产管理和数据基础服务** 进一步丰富完善基础信息， 重点补充行外政府部门、第三方合作机构等各类有价值数据，并按照统一标准进行客户信息 整合，实现客户物流、资金流、信息流的“三流合一”,形成更加完善全面的客户视图；通 过大数据技术实现“数据地图”等可视化数据服务，提升数据资产的易用性，充分发挥信 息价值；通过元数据的统一管理和分析，提供信息检索、指标灵活定制及查询等数据服务， 提升数据资产的一致性和可用性。

**13.3** **大数据时代的银行数据利用**

随着大数据时代的到来和互联网金融的兴起，银行的业务挖掘分析需求出现了爆发式的

**数据治理展望**| **第13章**|

增长，而且呈现出处理时效高、数据多样性及海量数据等特点。银行应结合大数据时代的需 求特点，融合行内外、线上线下数据，以分布式技术、云计算技术、可视化技术、文本挖掘 等大数据技术作为分析手段，提升银行数据服务能力，增强服务时效，充分挖掘银行大数据 信息价值，全面提升银行盈利能力。

从大数据技术兴起并发展至今，各银行通过探索与尝试，已经完成了大数据基本体系架 构的构建，并在客户营销、客户服务、风险管控等领域获得了初步成功。虽然大数据在各银 行已不再神秘，但大数据带来的机遇和蕴含的价值仍待进一步挖掘和利用。从长远来看，进 一步加强业务和技术融合，提升大数据应用服务能力是银行进一步努力的目标。对此，结合 国内外银行业大数据的应用情况，针对银行后续如何通过深化利用大数据来创造业务价值， 提出以下建议。

**1.** **利用大数据技术挖掘分析各类海量信息，发现市场热点与需求，实现产品创新服务**

随着互联网的发展，用户的线上行为越来越多，相关的诉求、爱好散落在各类社交媒体 网站，这些信息是银行原有资产数据的有效补充。而在信息获取方面，爬网、FLUME 等日 志采集技术，为银行获取相关信息资源提供了技术手段，信息的获取逐渐变得简单。另外， Hadoop 、Spark 等各类成本低且具备高容量、高效计算能力的软硬件技术为这些海量数据的 存储、计算、挖掘分析提供了可能。因此，银行可以通过大数据技术，减少金融脱媒带来的 冲击，发现市场热点与需求，实现产品创新服务。例如：可以将大数据应用到产品生命周 期，深入挖掘客户需求，把握客户“痛点”,推动产品创新。利用大数据技术对社交网站信 息、在线客户评论、博客、呼叫中心服务工单、用户体验反馈等信息进行深度挖掘和分析， 充分洞察客户，分析客户的情绪，了解客户对产品的想法，获知客户需求的变化趋势，从而 对现有产品进行及时调整和创新，使其贴近客户的生活场景和使用习惯，并向客户提供最新 的服务，更好地利用大数据的优势，在同业中取得竞争优势。

可以基于大数据创新产品评价方法，为产品创新提供数据支撑。通过大数据分析，改变 目前以规模、总量为主的业务评价方式，建立一套完整的以质量、结构为主的全新评价方 式，以引导全行真正追求有质量、有效益的发展。基于大数据，进行产品评价方法创新，对 客户结构进行深入分析，从行业集中度、潜力客户挖掘等入手，提高产品效益。还可以面向 不同客户群推出各类针对性的网络融资产品，丰富银行的产品体系，满足各群体用户的特性 需求。

**2.加强内外部信息联动，重点利用外部信息提升银行风险防控能力**

进一步加强与税务部门、海关、法院、电力部门、水务部门、房产交易登记中心、环保 部门以及第三方合作机构的数据互联共享，有效拓宽信息来源渠道，深度挖掘整合系统内外 客户信用、关联关系、交易行为、交易习惯、上下游交易对手、资金周转频率等数据信息， 利用大数据技术查找与分析不同数据变量间的关联关系并建立相应的决策模型，提升银行风 险防控能力。

在信用风险方面，可以结合外部信息，完善信用风险防范体系，基于可视化分析有效控 制信用风险的传导。引入和应用大数据理念和技术，实现信息收集整合，统一信用风险模型 管理，构建覆盖信用风险模型训练、模型管理、日常预警、评分评级、客户信用视图以及业 务联动控制的信贷大数据平台，建立多维度、全方位的风险预警体系，为供应链分析、担保 圈分析提供信息基础。同时，利用大数据可视化分析方法，有效控制信用风险的传导。例

 **银行数据治理**

如，在供应链分析方面，对供应链中的单个企业或上下游多个企业进行关联关系、行业分 布、资金流向等挖掘分析，推进供应链金融业务扩展；在担保圈分析方面，对企业通过互相 担保或连环担保而形成的担保圈进行分析，协助控制供应链信贷风险，防止集群性信贷风 险，间接帮助降低管理供应链金融的成本。

在市场风险方面，基于市场信息有效预测市场变动，基于大数据处理技术提升海量金融 交易的定价能力。在既有历史市场数据报价的基础上，进一步引入宏观经济参数、公开数据 披露信息，挖掘分析市场价格的变动规律，为市场风险的预测以及交易的开展提供指导。利 用大规模分布式并行计算技术，构建定价估值引擎批量网格计算服务模式，支持对海量交易 的实时定价，有效提升银行的风险管控与定价能力，为金融市场业务的发展提供有力支撑。

在操作风险方面，依托大数据信息整合优势，有效防控操作风险。在整合我行各专业结 构化信息的基础上，引入社交网络等和客户相关的非结构化信息，构建客户、商户行为的全 景画像，及时捕捉各类异常行为，有效防控操作风险。而在交易反欺诈领域，利用流数据、 分布式内存等大数据处理技术，基于神经网络模型、决策树和规则模型等，及时拦截异常交 易，有效防范犯罪分子盗取客户账户信息、窃取客户资金，建立起客户资金的安全网，并通 过运用可视化展现技术，支持从业务数据中识别风险线索，实现由“风险监控”向“业务 监控”模式的转变，提升风险的提前预警能力。加强跨专业风险监控模型的研发，通过由 点带线、由线及面的矩阵式关联监控，提前识别风险交织趋势，防范风险传染。

**3.** **利用大数据技术提升经营管理水平，优化业务流程，实现精细化经营决策**

通过进一步融合结构化和非结构化数据，内部经营数据和外部市场信息，有效拓展分析 外延，丰富分析内涵，提升分析时效，支持业务和管理创新。

**(1)在经营决策方面** 纳入外部资讯信息和市场动态，与内部经营信息相结合，制定 更有针对性、更高效的经营策略，并充分利用可视化技术，为经营决策提供分析结果的清晰 直观展现。通过补充与完善外部数据的采集和处理，进行结构化信息和非结构化信息的整 合，实现经营分析外延的拓展，从市场和经营环境的高度分析各级机构的发展方向、竞争压 力，制定更合理、更有效的经营策略。同时，应用大数据可视化技术，实现复杂分析过程和 分析要素向用户的有效传递，增强分析结果说服力和指导性，向经营人员提供有力的信息 支撑。

**(2)在资源配置方面** 依托大数据采集和计算能力，实现精细化、科学化的动态管理。 引入外部市场信息，依托大数据计算能力，更好地模拟市场变动情况，提升测算的敏感性和 有效性，加强财务预测的可靠性和有效性，为总体资源配置提供更好的信息支撑，实现对具 体资源配置的动态管理。同时，利用分布式环境下低成本高扩展优势，进一步加强经营计划 分解的精细化和科学化，为更低层级、更细粒度的经营计划的制定提供更精细化的信息辅助 支撑。

**(3)在过程改进方面** 优化业务流程，充分利用大数据信息资产，通过对交易、日志 的专题挖掘，探索当前业务处理流程中的节点和瓶颈，寻求优化解决方案。通过采集客户在 柜面、自助设备、排队机上业务操作及等候时间等数据，分析客户从排队等候到完成全部交 易的流程合理性，提出过程改进优化方法，提升网点整体运营效率和客户体验。

**(4)在运维保障方面** 基于流数据处理技术，搭建准实时的应用交易级监控平台，实 现交易运行情况的及时监控，保障业务运行稳定高效。数据中心集中监控系统通过流数据的

**数据治理展望**| **第13章**

方式，实时获取交易状态、交易时间、发起方等要素，并实时解析，准确测算交易量、响应 时间、成功率等关键指标，实时、准确、全面地掌握全行各业务系统运行情况，结合可视化 分析技术实现事件的智能分析与实时干预。

**4.** **利用大数据信息快速处理技术，重点推进高时效信息增值服务**

随着市场竞争的加剧，信息快速捕获与应用能力需求在不断增加。客户资产视图、事件 式营销、智能推荐、生活资讯服务等均需高时效信息的支持，而流数据信息截取、大内存计 算技术的成熟与运用，使得这些高时效需求的实现成为可能。银行可以利用大数据信息的快 速处理技术，推进银行高时效信息服务，第一时间创造营销机会并提升客户满意度。

 **附录**

附录 A《 中国银行业信息科技“十二五”发展规划

监管指导意见之大型商业银行和股份制商业银行篇》(摘录)

**第一章** **信息科技发展状况**

“十一五”期间，大型商业银行和股份制商业银行(以下简称大中型银行)信息化建设 取得了巨大进步，各银行机构围绕自身整体发展战略，推进信息科技规划与信息科技架构设 计，开展应用系统与基础设施建设，逐步建立安全高效的信息系统运行平台，开发种类多 样、功能丰富的产品体系，为加快业务发展、加强内部管理和推进全面风险管理提供了有力 支持，信息科技已成为银行核心竞争力的重要组成。同时，信息科技治理机制逐步建立，组 织体系进一步健全，信息科技队伍结构与人员素质得到优化和提高，信息科技管理水平得到 有效提升。

**(1)信息科技战略规划与架构管理能力明显提高，信息化建设有序推进** 对信息科技 战略规划重视程度不断提高，逐步建立信息科技战略蓝图，明确规划的具体实施路线，规划 方法趋向科学；部分银行机构设置专门的规划管理机构负责蓝图设计并督导实施。重视信息 科技架构优化，开展基础架构、应用架构、数据架构规划，推进架构评审管控机制，制定架 构管控策略和程序，推动架构规划的贯彻落实。

**(2)信息科技治理体系逐步建立，为信息科技可持续发展奠定基础** 探索信息科技治 理模式，确立信息科技战略定位，提升信息科技服务价值。完善决策体系，明确董事会、高 级管理层、信息科技部门以及相关业务部门职责；设立信息科技专业委员会，推进首席信息 官制度；优化信息科技组织架构与岗位设置，加强内控建设，推进专业化、流程化管理。提 高人力资源规划与管理水平，推进素质能力模型应用和员工职业发展。建立信息科技制度管 理体系，推进制度持续性建设。提高成本意识，推进信息科技成本管理。

**(3)基础设施建设持续加强，信息科技服务能力不断提高** 基本完成基础架构建设， 加大投入巩固基础设施建设，重点保障核心账务以及关键业务、关键渠道类系统的资源需 求。初步建立基础设施相关标准和规范，提升基础设施资源的管理水平和使用效率；实施基 础设施资源整合和扩容，提升基础设施对业务运行的承载能力；积极引入新技术，提高资源 利用率，推动信息科技资源集约化管理。

**附录**

在提高基础设施使用效率的同时，深入排查运营风险隐患，积极采用设备冗余、集群等 多种手段提升系统高可用性，持续提升信息科技服务能力。开展机房达标工程，提升数据中 心机房建设标准，改造机房环境；加强网络规划，做到高可用性与高可靠性并重，为业务发 展提供安全的基础保障环境。

**(4)信息安全技术保障体系初步建立，信息科技风险管理水平不断提升** 初步建立信 息安全制度规范和技术标准，强化信息安全风险评估和持续改进机制。初步建立等级化的信 息系统安全技术保障体系，积极运用先进、成熟的安全技术和产品，加强信息系统生命周期 安全管理。采取有效措施防范信息安全风险，运用加密技术和安全产品，规范和加强身份认 证、授权管理、责任认定等，建设网络信任体系，确保交易防抵赖。电子银行分层次、差异 化的安全认证方式和服务体系基本建立。广泛采用数字证书和动态口令等手段实现用户身份 鉴别，数字证书载体由文件、智能卡发展到 USBkey, 语音识别技术在电话银行中得到应用， 部分银行机构在ATM 机中开始运用指纹识别和视频技术。在信息保护方面，广泛运用防抵 赖、抗攻击和加密技术保护系统和信息安全，传统的电话银行和自助设备的安全性大为改 善，进一步推动了服务范围的拓展。

研发、运维管理组织架构日趋完善，基本形成了规范化的研发体系和适应大规模数据集 中处理的运维体系。积极开展灾备体系建设，强化突发事件应急管理，积极开展应急演练， 加强与国家相关部门以及行业间的联防协作，有效提高了应急处理能力。

**(5)应用系统建设初具规模，有力支持了业务发展** 基本完成数据集中，形成全行 “一本账”的账务核算体系，支持由“以账户为中心”向“以客户为中心”转变，基本实 现本外币一体化、全功能柜员管理，整合业务功能和处理流程，推进会计记账集中处理，支 持业务品种多样化发展。核心应用系统基本保持稳定持续运行，实现了7×24h 不间断服务。 建立内部资源管理系统，推进人、财、物的有效管理。完善客户关系管理系统，逐步实现全 行统一客户评价标准体系。建立多层次管理信息报表体系，建设集中的报表系统，提高量化 管理水平。推进风险管理系统建设，初步建立信用风险内部评级体系，逐步推进市场风险内 部模型法建设；开展自我评估与内部损失事件收集等工作，探索实施高级计量法，推进操作 风险管理。

**(6)电子银行渠道与创新应用不断丰富，金融服务水平持续提高** 初步建成电子银行 运营管理体系，建立涵盖网上银行、电话银行、手机银行、自助银行、电子商务的一体化、 立体交互式的电子银行产品交易平台，和以客户中心、门户网站、消息服务为支撑的信息服 务网络，电子银行交易替代率显著提高，电子银行渠道多样化服务体系基本形成。

加快创新，形成面向企业和个人客户的系列化产品体系。在网上银行方面，普遍提供在 线查询、转账、汇款、银证转账、代客外汇买卖等服务，部分银行机构试办网上小额质押贷 款、住房按揭贷款等授信业务。在手机银行方面，智能手机应用不断涌现。在自助银行方 面，出现存取款一体机、自助缴费设备、自助发票打印设备等新型设备以及基于指纹识别技 术的无卡交易。在电子商务方面，为电子商务应用提供全套在线支付解决方案，并进一步丰 富了手机支付、电话回呼支付等支付手段。

在充分肯定成绩的同时，也需要认识到存在的问题和不足，主要表现在以下方面：

**(1)信息科技与业务融合不足，信息科技治理水平有待提升** 信息科技战略与企业战 略结合不够紧密，部分银行机构尚未从企业战略层面开展信息科技整体规划；业务部门对信

 **银行数据治理**

息科技战略规划的制定参与不够，业务架构不够清晰；信息科技战略规划后评价机制有待完 善。信息科技组织结构、决策机制有待加强，决策职责分工和关键决策范围不够清晰，缺乏 科学的决策分析方法和决策工具，缺乏有效的监督问责机制。首席信息官机制有待进一步推 进。在信息科技架构与标准管理方面，尚未形成专业化的组织管理体系，对信息科技建设与 运行的整体指导作用有待加强。制度框架有待完善，制度执行和监督力度相对薄弱；技术标 准亟待健全。信息科技人力资源管理体系有待健全，部分银行机构信息科技人力资源不足， 存在科技人员总量占比少、关键岗位配比低、核心技术领域高端人才缺乏等问题。绩效考核 体系不够完善，激励与约束机制有待加强。信息科技成本管理有待加强，资源配置效率有待 提高，信息系统建设投入产生的业务价值有待于量化评估。

**(2)应用系统对经营管理和风险防控的支撑力度有待加强** 应用系统架构扩展性和灵 活性不足，应用系统间耦合性高，核心应用系统运行传导性风险较大。核心应用系统缺乏统 一的产品管理机制，产品配置与产品组合灵活度不够，产品开发周期长。应用系统建设整体 规划不足，“部门银行”现象、重复建设与资源浪费现象较为突出。管理信息系统对经营决 策与前台营销支持不够；分析型客户关系管理系统所覆盖的业务范围不全面，与操作型客户 关系管理系统的整合程度不高；指标流程化、规范化管理程度不高，应用系统对报表灵活定 制的支持程度有待提高；对并表的自动化处理能力亟待加强。风险量化模型建设与系统整合 度有待加强；各类风险模型的准确性、稳定性、规范性与透明度有待提高；计量结果在风险 管理流程与风险监控等方面的应用有待提高；风险管理系统建设的整体规划前瞻性与系统性 有待加强。

**(3)基础设施的灵活性和完备性不足** 信息科技基础设施复杂度高、部署时间长，基 础设施布局规划不足，支持业务发展的灵活性和完备性不足，难以有效适应业务快速发展需 要。数据中心规划选址过于集中，区域性风险较为突出；建设标准和管理规范有待完善；内 部能耗偏高，资源循环利用能力有待提高，运营成本控制有待加强。部分银行机构数据中心 安全保障和运维管理能力不足。

**(4)数据标准规范不够完善，数据质量和共享水平亟待提升** 数据治理体系建设较为 滞后，缺乏专业化的数据治理组织机构以及明确的责任人体系，尚未形成良好的决策机制和 管理机制。数据标准体系不够完整、清晰，数据标准贯彻执行力度亟待加强。数据质量亟待 提高，对于风险管控和精细化管理的支持亟待加强。应用系统建设存在“重功能、轻数据” 现象，难以提供准确、完整、统一的基础数据。风险数据同业共享机制有待建立。数据质量 的长效管理机制与持续改进机制亟待加强。

**(5)信息科技风险管理本身存在薄弱环节** 信息科技风险管理尚未纳入全面风险管理 体系；对信息科技风险的认识不充分，管理资源投入不足；对国外技术及设备的依赖度过 高；信息科技外包缺乏整体规划与战略性设计。研发和运维体系有待加强，制度、标准与流 程建设有待完善，研发质量保证体系、运维自动化与一体化以及绩效考核评价体系有待进一 步健全。业务连续性管理处于起步阶段，业务应急机制不足，应急预案体系不够完整，信息 系统灾备有效性不足，灾备切换演练未能真正贴近实战。电子银行的经营环境有待完善，制 度、法规建设有待加强。外部攻击威胁以及银行卡和电子银行犯罪问题日益突出，安全防护 能力建设有待加强；银行间信息共享和追踪配合机制有待建立。

**附录**

**第三章** **信息科技发展的目标、原则、重点与实施策略**

**第一节** **总** **体** **目** **标**

以满足业务发展战略要求为目标，全面优化信息化建设发展环境，努力提高信息科技自 主创新水平，强化信息科技风险防范和信息安全保障能力，加强全面风险管理基础设施建 设，推进向信息化银行转变；树立“科技引领”理念，增强核心竞争力，促进信息科技与 业务发展的深度融合，推动业务和产品创新、流程创新、管理创新，增强可持续发展能力 为社会公众提供丰富、安全和便捷的多样化金融服务。

主要目标是：

**(1)深化信息科技治理** 建立信息科技治理机制和治理模式，完善信息科技治理决策 机制，优化信息科技人才队伍总量、素质、结构，完善制度体系，提高信息科技资源的使用 效率与效果。

**(2)提高自主创新能力** 加大应用研发和产品创新力度，以科技创新促进客户服务、 产品创新、渠道拓展、风险防范、持续运营、管理决策能力的大幅提升。

**(3)加快发展电子银行** 加大电子银行发展力度，拓展应用领域，发挥电子银行引领 金融创新作用，推进电子银行成为银行交易主渠道和服务主平台，形成强大的产品与服务竞 争力。

**(4)不断夯实基础设施建设** 持续优化基础设施规范和标准，进一步加强重要基础设 施建设，逐步提升基础设施管理水平，不断增强服务保障和可持续发展能力。

**(5)深化数据治理，夯实数据基础** 建立数据治理体系与数据标准，提升数据质量和 数据应用水平，提高数据价值创造能力。

**(6)大幅提升信息科技风险防控能力** 建立风险防控体系，推进信息科技风险管理与 国际先进水平接轨，提升信息安全保障能力。

**第二节** **指** **导** **原** **则**

坚持科学发展，注重战略思维、国际视野与创新理念，紧密跟踪银行业务发展趋势，把 握信息技术应用方向，不断提升信息科技工作的集约化、规范化、自动化、精细化管理水 平，有效平衡业务发展与信息科技能力建设、信息科技风险管理之间的关系，促进协调发 展，全面提升信息科技核心竞争力，使信息科技成为推动业务转型、管理变革、产品创新和 战略落实的重要力量。

**(1)坚持围绕银行发展战略原则** 要树立以信息科技发展进步提升银行核心竞争力、 提升客户服务水平、促进业务发展的思想。要始终围绕银行总体发展战略和目标，以客户为 中心，与业务紧密融合、协同发展，实现业务发展和信息科技水平的同步跃升。

**(2)坚持自主创新原则** 要坚持自主创新的科学发展道路，树立创新精神，提高研发 能力，牢牢掌握信息化建设主动权。要充分发挥信息科技的第一生产力作用，将自主创新能 力建设作为银行信息科技发展的战略重点，坚持自主研发与适度引进相结合，注重知识转 移，重视培养创新型专业人才队伍，推动银行核心竞争力的有效提升。

**(3)坚持全面风险控制原则** 随着信息科技日益深入运用到银行经营管理全过程，要

 **银行数据治理**

将信息科技风险管理纳入银行全面风险管理体系，贯穿于银行各个经营领域，覆盖信息系统 建设的全生命周期，完善信息科技风险管理体系，提升信息安全管理水平，保障信息系统安 全、稳定运行。

**(4)坚持科技引领原则** 要紧密围绕银行发展战略，把握大中型银行业发展方向，切 实加强信息科技核心能力建设，以科技进步和创新引领银行业务创新与发展，推进银行发展 方式的战略性转变。

**第三节** **战** **略** **重** **点**

**(1)树立“科技引领”发展战略，推动转变发展模式** 充分认识信息科技在银行现代 化建设中的核心价值，树立“科技引领”发展战略，把信息技术作为推动银行转变发展模 式、实施差异化战略的关键源动力。以科技为依托，以创新为手段，形成特色服务与市场品 牌，提升产品研发能力、差异化服务能力与支持保障能力。

**(2)深化治理，着力提高信息科技核心竞争力** 着重信息科技体制机制创新，加强激 励与约束机制建设。完善信息科技决策机制，提高决策的科学性、有效性。加强信息科技队 伍建设，科学规划信息科技人才成长路径，加快创新型、专业型、复合型人才培养，形成具 有自主创新能力、具有国际化视野的高端信息科技人才核心梯队。加大信息科技投入，夯实 基础设施建设，优化资源配置，探索精细化管理的有效方法，提高银行信息科技发展的核心 竞争力。

**(3)提高信息科技自主创新能力，提升金融服务创新和服务水平** 创新是银行可持续 发展的源动力，应紧紧围绕业务发展战略重点，坚持业务与信息科技协调发展。深入研究业 务发展格局，跟踪全球金融服务信息化发展趋势，关注新技术发展方向，加强基础性、前瞻 性、关键性技术研发，提高自主创新能力。重点突破信息科技服务的薄弱环节，从银行与客 户的价值链入手，深度发掘客户需求，增强创新产品附加值，提高创新的系统性和针对性 通过信息科技创新带动管理创新、服务创新和产品创新。

**(4)构建立体化全面风险管控体系，提高信息科技风险防控水平** 要制定全面风险管 理系统建设规划，按照新监管标准实施要求，加强风险管理数据基础设施的建设和风险计量 模型的应用，建立健全具有统一性的、集成化的，覆盖信用风险、市场风险、操作风险及其 他实质性风险的全面风险管理系统；建立内生式的主动风险管理模式，使风险管理成为银行 健康、稳健经营的重要保障力量。要把信息科技风险管理纳入银行全面风险管理体系中，逐 步完善规章和制度，贯穿于业务流程，覆盖信息系统生命周期各个阶段，逐步建立起信息科 技风险管控的日常化、流程化、持续化机制。

**(5)紧密结合新技术，构筑现代化电子银行** 充分发挥电子银行对金融创新的引领作 用，以电子银行创新推动金融现代化发展，加快推动电子银行在社会各领域的全面运用，进 一步提升公众金融服务的质量和效率，使电子银行成为银行交易主渠道和服务主平台。电子 银行建设应始终贯彻发展创新与风险可控并举，业务模式创新与技术创新并重。要充分运用 电子银行的渠道优势、运营优势、服务优势和营销优势，加快电子银行发展和传统金融服务 提升的相互促进，实现电子银行的持续、稳定、快速、健康发展。

**(6)全面提升数据共享水平，深入挖掘数据的业务价值** 数据是银行最为重要的资产， 是银行业务经营与信息科技应用的核心，是支持精细化管理、提升风险分析能力的基础。在 推进应用系统逻辑集中过程中，核心是建立统一的数据标准，实现各类应用之间的互联互

**附录**

通，保证数据的真实性、完整性、 一致性，全面提升数据应用价值。

**(7)加强环境建设，促进行业协调发展** 积极配合国家相关部门，完善法律法规体系 建设，有效防范金融信息犯罪和网络犯罪；加强协作，积极开展银行业交流与合作；加强金 融信息标准化建设，提升公共信息共享能力，促进银行机构与非银行金融机构以及公共事业 机构的合作，建立公共接入平台，推进跨机构、跨行业的高效互连互通。

**第四章** **推进信息科技治理能力建设，加强信息科技战略与企业战略协同**

逐步建立长效的信息科技治理机制，确立信息科技治理模式，强化决策机制，完善制度 建设，优化组织架构，加强成本精细化管理。

**第** **一** **节** **深入开展信息科技治理体系建设**

**(1)构建长效治理机制** 推进在现代公司治理框架下、与业务运营模式相适应的信息 科技治理模式与治理机制。完善信息科技治理结构，发挥董事会在信息科技治理中的核心作 用，提高监事会监督效果，建立信息科技治理履职评价制度；加强信息科技委员会建设，推 进首席信息官制度；完善信息科技决策规范和流程，提高信息科技决策专业性，把握信息科 技发展方向与战略；分阶段、有步骤推进信息科技治理建设，逐步提高信息科技治理成 熟度。

**(2)加强信息科技与业务的协作融合** 进一步建立新的环境下信息科技与业务部门间 关系定位与责权利制度框架，推进业务和信息技术相融合的创新机制，倡导建立战略性合作 关系。进一步明确在信息科技管理活动中业务部门的职责、流程，强调并发挥业务部门在业 务需求分析与整合、应用项目业务价值分析、项目可行性研究与用户验收测试等过程中的主 导、推动作用。

**(3)加强内部管理** 提高信息科技资源使用效率，提高成本管理水平，逐步建立可量 化管理的信息科技服务需求与交付管理体系、信息科技项目评估体系，从业务价值、信息科 技成本收益、信息科技服务水平等多维度对信息科技资源利用效果进行全过程的跟踪评价。

**(4)加强信息科技决策体系科学化、规范化建设** 加强决策体系建设，明确决策流程、 授权机制、决策方法，加强制度和规范建设，着力提高决策的透明度和科学性。要建立科学 决策的组织架构，明确董事会、高级管理层、专业委员会、各部门在决策过程中的职责分 工。决策领域要覆盖信息科技关键决策，应包括战略决策、架构决策、重大项目建设决策等 重要领域。要加强对决策过程与执行效果的审计监督，加强约束，落实责任。

**(5)加强成本管理，推进信息科技财务管理体系建设** 应进一步增强成本意识，完善 信息科技财务管理体系，逐步建立信息科技预算、成本控制、成本分摊一体化的财务管理体 系，提高信息科技资源使用的效率与效果。建立健全精细化的信息科技预算管理体系，加强 成本控制。加强战略与规划对信息科技预算的指导性，提高信息科技预算的精确程度，提升 资源配置决策质量和配置效率。建立完善信息科技预算考核机制，明确项目开发与系统运维 服务预算相关主体职责权限，并将预算考核纳入到相关责任主体绩效考核中，提高信息科技 资源使用效率与效果。

强化全面成本管理理念，逐步建立成本分摊模型和内部计价机制。逐步完善数据基础与 管理工具，合理归集与分配信息科技成本，为精细化的信息科技财务管理奠定基础。建立对

 **银行数据治理**

信息科技投入的评价与考核机制，逐步强化需求部门对于信息科技服务的成本观念，基于投 入产出分析或成本收益分析进行信息科技投资决策，加强成本控制。基于信息科技资源使用 量与服务水平等要素，建立精细化的信息科技成本分摊框架与分摊方法，逐步提高信息科技 部门成本管理水平以及成本分摊的准确性，促进资源合理利用。探索信息科技服务水平协议 管理方法，探索实施信息科技服务内部计价。

**第二节** **优化信息科技组织架构**

基于战略规划与治理要求，持续优化信息科技组织架构，重点解决管理职能交叉或缺 失、运行效率低下等问题，并建立信息科技组织结构评估与优化机制，持续提升组织管理能 力。应定义关键信息科技能力，明确研发策略。应依据内控与合规建设要求，建立包含信息 科技专业委员会、信息科技管理部门、信息科技风险管理职能部门与信息科技审计职能部门 的组织体系，明确职责和人员岗位要求；重点加强在信息科技规划与架构、研发与项目管理 以及制度与标准建设等重点领域的组织建设。

增强信息科技服务能力，加强服务流程管理。在业务需求管理、项目管理等领域，建立 面向内部用户的、稳定的服务界面，提高信息科技与业务协作效应；在信息科技运行维护等 领域，建立基于服务流程的、专业化、集约化信息科技服务组织体系。

**第三节** **加强信息科技队伍建设，提升信息科技核心竞争力**

牢固树立人才资源是科学发展第一资源的观念，系统规划科技人才队伍建设，并纳入长 期战略，统筹抓好各类人才队伍建设；要加大培养力度，引进优秀人才，建立健全科学的培 训体系，完善职业生涯发展规划，健全绩效考核机制，打造梯次合理、素质优良、专业化程 度高、适应战略发展要求的科技人才队伍。

**(1)继续加大信息科技人力资源投入，提高关键岗位信息科技人员比例** 科学制定信 息科技人力资源规划，加大信息科技人力资源投入，提高关键岗位信息科技人员比例；原则 上，信息科技人员占比应不低于3%。以自主开发为主的银行机构，信息科技人员占比应不 低于4%。

基于人力资源规划目标，建立人员补充机制；董事会和高级管理层应督促、指导制定有 关人力资源政策和措施。应进一步明确总、分行信息科技职能定位，优化总、分行科技人员 配置结构。

**(2)拓宽信息科技人员职业发展空间，完善激励约束机制** 应完善岗位职责说明书制 度，细化人员岗位职责和技能要求。进一步建立并持续完善与行政序列并列、与薪酬挂钩的 信息科技专业职级体系，完善晋升机制，拓展信息科技人员职业发展空间。应进一步完善激 励约束机制，制定精细、量化的绩效考核指标，全面、科学评价信息科技部门与人员工作 绩效。

**(3)明确核心能力发展要求，完善人才培养机制** 明确信息科技核心能力，制定适合 银行发展战略的专业人才培养计划。基于信息科技战略与规划、信息科技人力资源规划、外 包策略与未来技术方向明确核心能力要求，制定信息科技人员专业化培养方向，在需求分析 管理、架构管理、测试管理、系统与数据库管理、服务水平管理等关键领域加强专家级人才 的培养与引进。

进一步提高信息科技培训的专业化水平，针对不同关键岗位和关键能力，制定细化、差

**附录**|

异化培训目标，推进信息科技培训与专业岗位技能要求的有机结合；应紧跟信息科技与业务 发展动态，建立覆盖入职培训、在职岗位培训、专业培训的全方位培训课程体系。加强培训 效果评估和跟踪，提高培训成果转化率。

**第四节** **构建信息科技制度框架，有效提高管理水平**

制定、完善信息科技制度体系建设策略，明确制度框架、职责分工和授权原则、重点建 设领域及其优先级。制度框架要基本覆盖信息科技各主要领域，包括战略规划、架构、运 行、开发、测试、安全、灾备管理等。要不断完善各项制度，结合实际情况，重点加强对架 构管理、数据治理、外包管理、信息科技成本管理和业务连续性管理的制度建设和流程 建设。

应设立专业化团队或岗位负责制度体系建设的总体管理，承担组织协调以及框架体系建 设、制定建设计划、统一建设标准等各项工作。应明确各级管理人员在制度管理领域角色和 职责，明确各项制度、流程和标准建设的责任人。应建立制度建设持续改进机制，加强制度 变更管理。积极推进多渠道的制度宣传和培训，提高制度执行力。

充分认识信息科技标准化工作意义，统筹规划总体框架、积极构建全行统一的技术与管 理标准体系，贯穿于企业架构各个层次，覆盖系统、应用、安全、数据等各领域；要建立健 全标准管理流程，建立专家评审机制，把好标准质量关；要加强对标准实施情况的监督、跟 踪，开展标准后评估和修订工作，保持标准的连续性和一致性；要积极借鉴国际标准和最佳 实践，注重与国际标准的兼容，加强我国银行业自身标准的研究、制定；要树立标准的权威 性，切实增强标准的执行力、影响力。

**第五节** **建立健全架构管控体系，贯彻落实信息科技战略**

充分认识信息科技架构规划的重要性，在业务架构基础上，建立覆盖应用架构、数据架 构和基础架构的信息科技架构规划。要制定并落实应用架构、数据架构、基础架构建设原 则，建立架构设计规范标准，指导并约束项目和项目群实施。

建立信息科技架构管控流程，完善战略规划、架构设计、项目群实施的全流程管理体 系，加强协同与融合，逐步从项目驱动、业务条线驱动方式转变为架构驱动的信息科技建设 模式，建立企业级架构管控组织和流程，建立架构管控平台，确保架构贯彻落实。

建立架构评价体系，持续跟踪并监测架构规划的执行情况，在保持架构相对稳定的基础 上，适时调整架构规划以适应发展要求。探索架构成果资产化管理，将企业架构成果纳入到 可重用资产管理范畴，建立企业架构资产库。

**第六章** **加强信息科技风险管理，完善研发运维体系**

“十二五”时期，大中型银行要把加强信息科技风险管理作为一项重要任务，培育信息 科技风险管理文化，优化组织架构，制定信息科技风险管理战略、政策、制度、流程、方 法，提高组合风险管理水平，将信息科技风险管理纳入到全面风险管理体系中；把风险管控 贯穿于信息系统生命周期，建立起信息科技风险管控的日常化、流程化、持续化机制。

**第一节** **建立全面的信息科技风险管理体系与长效管理机制**

**(1)构建全面的信息科技风险管理体系** 构建全面的信息科技风险管理体系，根据风

 **银行数据治理**

险战略与信息科技战略规划，制定信息科技风险管理战略、政策；进一步完善信息科技风险 管理组织架构，完善决策流程，推进董事会和高级管理层履职评价；推动将信息科技风险管 理纳入全面风险管理框架，探索信息科技风险与其他各类风险组合管理。建立信息科技风险 识别、计量、监测和控制流程，涵盖信息系统生命周期各个环节。建立信息科技风险监测指 标体系，明确关键风险指标，建立常态化的信息科技风险监测、预警与处置机制，开展日常 评估、检查和审计等工作，全面识别风险，及时制定和实施控制措施，建立评估控制有效性 的程序。建立清晰的信息科技风险报告机制，定义报告内容、频率、对象及路线，董事会、 高级管理层、信息科技部门、风险管理部门以及审计等各相关部门应及时掌握风险状况与趋 势。建立信息科技风险信息采集、评估与监控管理平台，建立自动化管理流程。积极探索信 息科技风险损失计量方法和计量标准，逐步建立信息科技风险损失数据库，收集信息科技风 险损失数据，直观计量信息科技风险损失，实施信息科技风险定量分析及趋势分析，支持信 息科技风险管理决策。

**(2)加强协作，提升信息科技风险管理协同效应** 加强与国家相关部门和机构协作 建立信息科技风险协防机制，提高主动防御能力。加强同立法机构、公安机关、电信、电 力、交通、消防等部门协作，保障数据中心等重要基础运营环境安全，促进电子交易法律保 障体系建设，打击电子交易违法行为和金融网络犯罪；建立跨行业、跨机构、跨区域的应急 协调机制，积极应对各类突发事件，保障业务经营活动持续性。

**第二节** **加强信息安全管理，全面提升信息安全保障能力**

**(1)夯实信息安全基础，推进信息资产分类分级管理** 深入开展等级保护，建立重要 系统安全防护体系和技术管理体系，建立主动防御机制。逐步推进信息资产识别和分类、分 级工作，建立信息资产分级标准、规范，明确安全策略和保护要求。落实管理责任，统筹推 进信息安全管理工作，确保信息安全管理范围的全面覆盖。加强敏感信息保护，防止信息资 产违规泄露，加强向外部提供信息的统一管理，严格审核，归口发布。综合运用技术和管理 手段，加强终端设备的安全管理。

**(2)强化系统建设各环节的安全管控，加强安全质量管理** 建立和完善信息系统生命 周期安全管理机制，加强信息安全管理制度体系建设，覆盖信息安全方针、策略、管理要 求、操作流程、应急计划和联动机制。加强信息系统建设全过程安全管理，制定信息安全规 划，建立信息安全架构，统一部署信息安全功能，提高安全措施效率。加强需求与设计阶段 安全功能需求分析与设计，抓好安全测试工作并贯穿系统研发过程，检测安全漏洞，评估安 全需求的符合性。加强上线前安全评估，评价系统安全性以及对整体安全保障体系影响。加 强系统运行安全评审，及时发现并处置安全隐患。

**(3)优化信息安全技术防控手段，实现纵深防御** 优化信息安全技术防御体系，在物 理安全、网络安全、系统安全、应用安全、数据安全、操作安全等不同层面，完善身份认 证、访问控制、资源管理、日志分析、操作审计等安全功能，主动应对安全威胁，重点防范 外部攻击，提升信息安全保障体系的健壮性和有效性。强化基础设施安全保障，深化应用系 统安全建设，增强应用产品安全性。建立用户认证和访问控制管理流程，加强数据访问权限 管理，有效保护信息资产。

**(4)建立信息安全保障能力评价机制，保障信息安全管理有效性** 建立信息安全保障 体系评价机制，建立量化指标，及时开展评估、审计，建立持续改进机制，保障信息安全管

**附录**

理的持续性、有效性。

**第十章** **建立数据治理机制，推进数据标准化和质量建设**

“十二五”期间，大中型银行要把数据治理作为重要的制度性建设与基础性工作，加强 组织保障、制度保障与流程保障，有序推进、重点强化；统一数据标准，提高数据质量，深 化数据应用，有效支撑银行业务发展，有效提升银行管理水平。

**第** **一** **节** **提高认识，建立数据治理体系**

充分认识数据在支持经营决策、内部管理与金融服务中的核心价值和战略意义，切实提 高对数据价值和数据治理的认识，大力开展数据治理工作。建立健全“决策、管理、执行” 三层数据治理组织架构，设立由银行高级管理层负责、有关部门负责人参与的数据治理决策 机构，负责数据治理日常决策，协调跨部门协作。明确职能部门统筹全行数据治理工作，加 强对各部门数据治理工作监督考核。要建立数据责任人体系，明确数据定义、录入、使用部 门和信息科技部门工作职责分工，在各主要部门设置数据治理工作岗位。要在全行层面统筹 管理数据需求、数据标准规范，加强数据采集与使用过程的规范化、标准化。加强采集源头 的数据质量管理；在分支机构建立相应组织体系，贯彻落实数据治理政策和制度。

建立和完善数据治理制度体系，规范工作流程，理顺内部协作关系，建立和完善数据治 理决策程序，逐步形成透明高效的决策机制。逐步建立数据治理的评价、考核机制。

**第二节** **加快数据标准建设，统** **一** **数据规范**

推进行业数据基础标准框架规划与标准建设。针对新资本协议实施，以及银行间、行业 间等基础性强、关注度高的重点领域关键数据项和统计指标项，积极推进数据标准建设。加 强行业数据标准在各行信息系统建设中的落地实施，进一步提高银行信息系统与行业监管信 息系统之间数据对接的准确性、 一致性。积极制定和完善企业级数据标准，指导、规范业务 操作和系统建设，建立完善数据标准的需求响应、制定、评审、发布、应用等管理程序，建 立数据标准评审与评价反馈机制，保障数据标准应用效果。

**第三节** **加强数据全生命周期管理，提高数据质量**

加快建立数据质量管理机制，进一步完善管理流程，明确数据管理、定义、录入、使用 部门和信息科技部门等各方的质量管控职责，逐步建立日常监督管理的长效机制，保障数据 质量的持续提升。要着力加强数据需求统筹管理，建立数据需求管理程序，进一步完善和优 化数据应用和系统开发流程，严格数据需求管理，从源头解决数据“孤岛”问题。大力开 展数据综合治理工作，全面梳理现有数据情况，分析质量问题，制订改进计划，先行推进新 资本协议实施等重点业务专题与基础数据领域工作；改善基础数据质量，提高数据一致性、 准确性及自动采集率。积极利用技术手段，探索建立集中、统一的数据质量管理平台，加强 数据质量管理支持系统建设，提高自动化管理水平，提高质量管理效率。

**第四节** **优化数据架构，推动数据信息逻辑整合**

积极开展数据架构规划，对信息模型、主辅数据源以及数据集成架构等内容进行前瞻性 设计，制定数据生命周期管理规范，提升数据服务能力。加强主数据管理，形成包括客户、 产品、机构等核心基础数据的统一视图。加强元数据管理和应用，制定元数据管理策略、标

 **银行数据治理**

准、流程，建立统一的元数据管理与传输机制，将元数据管理与系统开发、运维与数据架构 管控等流程结合起来，保证元数据的有效性和准确性。

不断丰富前台源系统的基础数据，持续加大数据积累与整合的广度和深度，建设统一的 数据平台，满足前台营销、统计分析和决策支持、风险管理和新资本协议实施等多种需求， 持续提升对数据的分析、挖掘能力与深度运用能力。

**附录B 《** **贷款统计分类及编码标准(试行)》(摘录)**

**4** **代码结构与编写格式**

**4.1** **贷款统计分类代码结构**

本标准采用线分类法和分层次编码方法，代码结构分为三层，代码结构图如下所示。 O ×× ××××××

—(数字或字母数字组合)统计分类标志值代码 -(数字)统计分类标志编码

—(字母)金融工具大类编码

代码第一层为金融工具大类编码，采用《金融工具统计分类及编码标准(试行)》中的 金融工具大类编码，采用1位大写拉丁字母表示，贷款用F 表示。

代码第二层为统计分类标志编码，采用2位阿拉伯数字表示，贷款分类标志见“5 贷 款分类标志及编码”。

代码第三层为统计分类标志值代码，采用6位阿拉伯数字或阿拉伯数字与大写拉丁字母 组合表示，该段代码不足6位则右补0,各统计分类标志值代码表见“6 贷款分类标志值 及代码表”。

例如： F02 C01000 表示借款人为公司的贷款； F03 CNY000 表示交易币种为人民币的 贷款。

**4.2** **分类标志值代码表编写格式**

本标准中按照分类标志编排顺序依次进行分类标志值代码编码，其基本格式如下所示： 编码+分类标志名称

表示： ……

编码方法： ……

代码表： ……

**5** **贷款分类标志及编码**

贷款分类标志、分类标志说明及编码见下表。

|  |  |  |
| --- | --- | --- |
| 编码 | 分类标志 | 说 明 |
| 01 | 资产负债类型 | 指贷款资金的债权、债务属性 |
| 02 | 国民经济部门 | 指对借款人所属的国民经济部门分类 |

**附录**|

(续)

|  |  |  |
| --- | --- | --- |
| 编码 | 分类标志 | 说 明 |
| 03 | 币种 | 指贷款的交易币种 |
| 04 | 地区 | 指借款人登记注册地或常住地 |
| 21 | 行业 | 指借款人在有关部门登记注册的或主要从事的行业 |
| 22 | 企业出资人经济成分 | 指借款企业实际控制人经济成分 |
| 23 | 企业规模 | 指借款企业的经营规模 |
| 24 | 合同期限 | 指贷款合同初始约定的还款期限 |
| 25 | 剩余期限 | 指贷款业务由报告日至合同约定到期日的剩余期限 |
| 51 | 产品类别 | 对归属于贷款工具的金融合约按特定契约特性进行分类 |
| 52 | 贷款逾期期限 | 指贷款的本金或利息超过贷款合同规定期限的时间 |
| 53 | 贷款实际投向 | 指贷款合同上载明的实际用途 |
| 54 | 个人经营贷款用途 | 指个人经营性贷款的实际用途 |
| 55 | 贷款质量 | 指贷款质量的五级标准分类 |
| 56 | 贷款迁徙方式 | 指贷款从年初到报告期末两个时点的质量迁徙情况 |
| 57 | 贷款经营类型 | 指依据贷款资金来源及贷款发放对象的确定方式不同进行的贷款分类 |
| 58 | 贷款还款方式 | 指贷款合约中载明的借款人偿还金融机构贷款的方式 |
| 59 | 贷款财政扶持方式 | 指国家为了实现特定的政治经济目标，由财政安排专项基金向国有企业或劳 动者个人提供资助的方式 |
| 60 | 贷款担保方式 | 指借款人根据要求提供的第三方保证的方式 |
| 61 | 利率是否固定 | 指贷款合同存续期内利率水平是否变动 |
| 62 | 利率期限单位 | 指贷款利率适用的时间单位 |
| 63 | 贷款利率重新定价期限 | 指贷款业务自报告日到最近可重定价日或到期日的期限 |
| 64 | 贷款状态 | 指贷款处于正常、展期、逾期或核销状态 |
| 65 | 抵减存款准备金方式 | 指贷款是否作为存款缴纳存款准备金的轧差项 |

**6** **贷款分类标志值及代码表**

**6.3** **币** **种**

**03** **币种**

表示： CurrencyCode

编码方法：三位大写字母

**代码表**

|  |  |  |
| --- | --- | --- |
| 代码 | 名称 | 说明 |
| USD | 美元 | 采用GB/T12406中的三位字母型代码 |
| HKD | 港元 |
| CNY | 人民币 |
| \* | \* |

 **银行数据治理**

**附录** **C** **《银行监管统计数据质量管理良好标准(试行)》(摘录)**

**(二)制度建设**

**原则5** **管理制度**

银行建立全面、科学和有效的监管统计管理制度，保证监管统计工作的规范性和严 肃性。

具体标准：

5.1 监管统计归口管理部门牵头其他相关业务部门建立符合银行监管要求的统计管理 制度。统计管理制度的内容包括但不限于组织管理、部门职责、协调机制、安全保密、系统 保障、监督检查、数据质量控制以及监管统计数据源管理等方面，并以文件的形式下发 执行。

5.2 监管统计归口管理部门根据监管要求和管理实际，持续评价和及时更新统计管理 制度。

5.3 监管统计归口管理部门将本行的统计管理制度报银行业监督管理机构(以下简称 监管机构)备案，管理制度若有重大调整或变化，及时向监管机构报告。

**原则6** **业务制度**

银行建立适应本行业务实际的监管统计业务制度，做好监管统计数据源和统计项目的归 属管理。

具体标准：

6.1 监管统计归口管理部门组织其他相关业务部门制定符合本行实际的统计业务制度。 统计业务制度覆盖所有的监管统计报表和数据要求，并正式发布。

6.2 统计业务制度对监管统计的内容、方法和口径等方面做出统一规定，保证监管统 计报表和数据中每一个统计项目的归属关系及取数路径清晰、准确。

6.3 监管统计归口管理部门根据监管要求和管理实际，组织其他相关业务部门持续评 价和及时更新统计业务制度。

6.4 监管统计归口管理部门将本行的统计业务制度报监管机构备案，业务制度若有重 大调整或变化，及时向监管机构报告。

**(三)系统保障和数据标准**

**原则7** **基础系统**

银行业务及管理基础系统实现对监管统计涉及的各项基础业务和管理信息的全面覆盖， 数据维护管理制度和流程完备，具有较强的数据质量控制功能，保证系统信息全面、及时、 准确反映业务实际。

具体标准：

7.1 业务及管理基础系统的设计、建设和升级改造征询并充分考虑监管统计业务需求， 系统功能支持监管统计信息自动采集。

7.2 业务及管理基础系统涵盖监管统计涉及的全行各业务领域及各类相关机构。如由 于新的监管要求变化而暂时无法实现全覆盖，要有明确的系统完善时间安排。

**附录**|

7.3 业务及管理基础系统有完备的数据维护管理制度和流程，尤其是要有对质疑信息 明确核查和纠正的相关规定。

7.4 业务及管理基础系统信息及时录入，经过复核，并动态更新，保证全面、及时、 准确地反映业务实际。对监管统计人员及系统管理人员反馈的存在完整性、准确性等缺陷的 质疑信息，及时予以核查并纠正。

7.5 业务及管理基础系统具有良好的可拓展性，满足新业务、新产品和监管新要求的 需要，并为监管统计管理系统预留能够灵活加载的数据接口。

**原则9** **数据标准**

银行数据标准化规划统筹考虑监管统计数据标准化，全面覆盖监管统计数据及指标。数 据标准符合国家标准化政策及监管统计规定，在监管统计相关的业务领域及系统中规范 执行。

具体标准：

9.1 银行充分考虑监管统计要求，将监管统计所涉及的数据标准纳入本行信息标准化 规划。

9.2 监管统计相关的银行数据标准符合国家标准化政策及监管统计规定。数据标准明 确，文档规范。

9.3 银行与监管统计相关的业务领域及系统遵循统一的业务规范和技术标准，建立全 行统一的监管统计数据架构，客户、产品、机构、账户等基础信息跨系统保持名称、定义、 口径、来源等方面的一致性。各系统建立统一的数据交换标准，实现系统间信息共享。

**(四)数据质量的监控、检查与评价**

**原则10** **日常监控**

银行建立有效的数据质量监控体系，对监管统计数据质量实施持续监测，采取有力措 施，落实各环节责任，确保数据质量。

具体标准：

10.1 银行建立有效的覆盖监管统计数据生产报送全流程的数据质量监控机制。

10.2 监管统计相关业务部门对本部门监管统计的数据源信息质量适时监控，对监管统 计归口管理部门提出的数据源质量问题及时整改，保证数据源信息的准确性和完整性。

10.3 监管统计归口管理部门对监管统计数据实施有效的跟踪监控，强化总体复核，发 现数据质量问题，反馈到相关责任部门并督促其及时纠正。

10.4 监管统计归口管理部门建立与报表频度相适应的数据质量监测台账，对监管统计 数据质量问题实时记录，定期分析。

10.5 监管统计归口管理部门定期通报相关业务部门和分支机构提供的监管统计数据质 量情况，总结交流经验，并提请实施问责。

**原则11** **监督检查**

银行将数据质量检查纳入内控合规检查范围，建立监管统计现场检查制度，定期组织 实施。

具体标准：

11.1 银行将数据质量纳入内控合规检查范围，内控合规管理部门在制定年度检查计划 时充分考虑监管统计归口管理部门的意见和建议。

 **银行数据治理**

11.2 银行建立监管统计现场检查制度。归口管理部门定期牵头组织行内监管统计现场 检查，检查内容包括监管统计法规制度的执行及其数据质量等情况。

11.3 归口管理部门将统计现场检查报告报主管领导，对影响数据质量的重大问题，提 交内控合规部门跟踪检查，并报监管机构。

11.4 银行督促被查单位根据统计现场检查发现的问题以及归口管理部门的要求实施整 改，及时对整改情况跟踪评价，确保整改落实到位。

11.5 监管统计相关业务部门定期对监管统计的数据源信息质量进行检查，及时组织整 改，并将检查整改结果抄送监管统计归口管理部门。

11.6 银行对监督检查发现的重大问题，依据有关规定对相关部门及责任人予以问责和 处理。

**原则12** **考核评价**

银行建立有效的考核评价体系和奖惩机制，保障全系统监管统计工作的有效落实。 具体标准：

12.1 监管统计归口管理部门定期对相关业务部门和分支机构提供数据的质量情况进行 考核和评价，考评结果在行内公布。

12.2 监管统计相关业务部门按照业务范围对监管统计数据源管理工作进行考核评价， 落实相应的激励约束措施。

12.3 银行将监管统计数据质量的考核结果纳入本机构绩效考核体系。

**(五)数据的报送、应用和存储**

**原则13** **数据报送**

银行加强向监管机构报送数据环节的管理，积极落实各项报送要求，认真审核，减少迟 报、漏报和差错。

具体标准；

13.1 银行建立数据报送流程管理制度和报送联系人制度，保障与监管机构联系畅通。

13.2 银行与监管机构建立网络连接，实现监管统计报表和数据的快速、安全传输，按 监管统计制度规定的频度和时间要求及时报送数据信息，并建立监管统计报送应急预案。

13.3 数据报送符合监管统计制度要求，不存在漏报的数据信息，保证填报地区、机构 范围和统计数据的完整性。

13.4 数据报送的文件格式正确，系统能正常读取数据，符合逻辑校验关系要求。对数 据异常变化的质询能做出合理解释；对历史数据进行修改，说明缘由，并报送监管机构 备案。

**原则14** **分析应用**

银行在保密安全的前提下充分共享和使用监管统计信息，为本机构经营决策和风险管理 服务，同时通过扩大共享、加强分析、适时披露，促进数据质量提升。

具体标准：

14.1 监管统计归口管理部门对监管统计信息合理分类，对相关业务部门设定合理的用 户权限，充分实现机构内的统计信息共享，扩大应用范围。

14.2 监管统计归口管理部门牵头建立全面、科学的监管统计信息分析框架，做好监管 统计信息的监测、分析和风险预警工作。

**附录** |

14.3 相关业务部门有效分析使用监管统计信息，为本行合规审慎经营提供决策参考， 并以此推动数据质量不断提高。

14.4 银行应关注报送监管机构的监管统计信息与对外披露的相关信息间的差异，并将 其中重大差异主动向监管机构做出说明。

**附录** **D 推** **荐** **读** **物**

[1] 维克托·迈尔-舍恩伯格，肯尼思·库克耶.大数据时代(生活、工作与思维的大变 革) [M]. 盛杨燕，周涛，译.杭州：浙江人民出版社，2013.

[2] 桑尼尔·索雷斯.大数据治理 [M]. 匡斌，译.北京：清华大学出版社，2015. [3] 王飞.数据架构与商业智能 [M]. 北京：机械工业出版社，2015.

[4] 赵飞.基于全生命周期的主数据管理 MDM 详解与实践 [M]. 北京：清华大学出版

社，2015.

[5] 包磊，黄亮，罗兵，等.作战数据管理 [M]. 北京：国防工业出版社，2015.

[6] 王飞.商业智能深入浅出——大数据时代下的架构规划与案例 [M]. 北京：机械工

业出版社，2014.

[7] 麦吉利夫雷.数据治理工程实践——获取高质量数据和可信信息的十大步骤 [M].

刁兴春，曹建军，张健美，译.北京：电子工业出版社，2010

[8] 李春田.标准化概论 [M]. 北京：中国人民大学出版社，2010

[9] 刘伟毅.互联网金融 [M]. 北京：中国经济出版社，2014.

**参** **考** **文** **献**

[1] 刘松林.商业银行数据治理探析 [J]. 中国金融电脑，2010(7):70-71.

[2] 程普升.面向数据与数据治理：商业银行信息化建设的转型方向 [J]. 中国金融电脑，2011(12):

36-39.

[3] 宋敏，覃正.国外数据质量管理研究综述 [J]. 情报杂志，2007(2):7-9.

[4] 约翰·卡尔洛·柏妥.美国大数据治理模式也面临不少新问题 [N]. 东方早报，2014-5-20.

[5] 张颖.理论与实践相结合：构建企业级元数据管理体系 [J]. 中国金融电脑，2013(7):19-21. [6] 马泽君.农业银行数据管理的探索与实践 [J]. 中国金融电脑，2012(2):15-18.

[7] 王永利.在信息科技变革中谋求长远发展 [J]. 中国金融，2012(8):36-38.

[8] 汪航.大数据时代的数据管理战略 [J]. 中国金融电脑，2013 (7):25-28.

[9] 张义祯.树立大数据治理意识 [N]. 学习时报，2014-12-15.

[10] 李沫，郎波.一种非结构化数据库的四面体数据模型 [J]. 中国科学： F 辑信息科学，2010,40

(8):1039-1053.

[11] 李恒.元数据管理系统的研究与实现 [J]. 电脑知识与技术，2011,07(13):2997-3000.

[12] DAMA International.DAMA 数据知识管理体系指南 [M]. 马欢，刘晨，等译.北京：清华大学出

版社，2012.

[13] 王玮.大数据时代的商业银行数据治理研究 [J]. 中国金融电脑，2013(7):36-38.

[14] 张功臣，赵汉体，高翔.城市商业银行数据治理之路 [J]. 金融电子化，2015(3):76-78.

[15] 许文，潘明道，徐明圣，王昭.结合银行业务特点开展数据治理工作 [J]. 银行家，2012(12)

125-128.

[16] 于一超，何琳.商业银行信息标准化管理的理论与实践 [J]. 国际金融，2012(10):33-36.

[17] 姜才康，孙小林.当前中国银行间市场数据标准化探索与实践 [J]. 中国货币市场，2011(1):

40-44.

[18] 洪正华.银行业数据管理探索 [J]. 银行家，2012(4):123-125.

[19] 姚远.中小型商业银行数据质量管理探索 [J]. 财经界，2014(36):132-133.

[20] 田江，刘晨.银行业数据质量管理方法研究与实践 [J]. 电子科学技术，2015(3):379-383.

[21] 吴永飞.商业银行数据生命周期管理探讨 [J]. 中国金融电脑，2014(6):24-30.

[22] 刘晓.大数据环境下数据中心的数据生命周期管理研究 [J]. 中国金融电脑，2014(10):71-75.

[23] 刘锦森，刘巍.银行数据管理实践之核心对象 [J]. 金融电子化，2015(2):56-58.

[24] 刘怡.浅析数据标准在银行实务工作中的应用 [J]. 中国经贸，2014(11):162-164.

[25] 孙中东.企业级数据治理框架下的数据质量管理 [J]. 金融电子化，2011(06):57-60.

[26] 付娆.驾驭大数据，治理为本 [J]. 金融电子化，2014(4):49-50.

[27] 刘锦淼，李海丽.数据标准化业务发展的坚实地基 [J]. 金融电子化，2014(4):53-55.

**银行业信息科技风险管理高层指导委员会**

金书网：www.golden-book.com

**银行业信息化丛十**

**银行数据治理**

商业银行信息系统研发风险管控

全球化时代的银行信息系统建设

商业银行私有云设计与实现

银行信息系统架构

银行信息安全技术与管理体系

商业银行业务连续性管理

金融数据挖掘与分析

银行数据中心基础设施建设与运维管理

银行业信息科技监管

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1 6 9 5  …  …  …  …  …  …  …  8 9 5  9 8 6 | … 0 9 1  …  2 0 5 3  …  5 1  …  605312  5653  …  95  …  350  … …  9 | 6 0 5 1 3 2  … | 9 0 8 5 1 5 3 2  …  …  …  …  …  0 7 5  2 7 9 5  0 | 1 0 8 6 |

地址：北京市百万庄大街22号 邮政编码：100037

**电话服务** 服务咨询热线：010-8**8361066** 读者购书热线：010-68326294

010-88379203

机工官网：www.cmpbook.com

**网络服务**

机工官博：weibo.com/cmp1952

教育服务网：www.cmpedu.com

**封面无防伪标均为盗版**



机械工业出版社微信公众号



上架指导金融

ISBN 978-7-111-53485-3



定价：69.80元