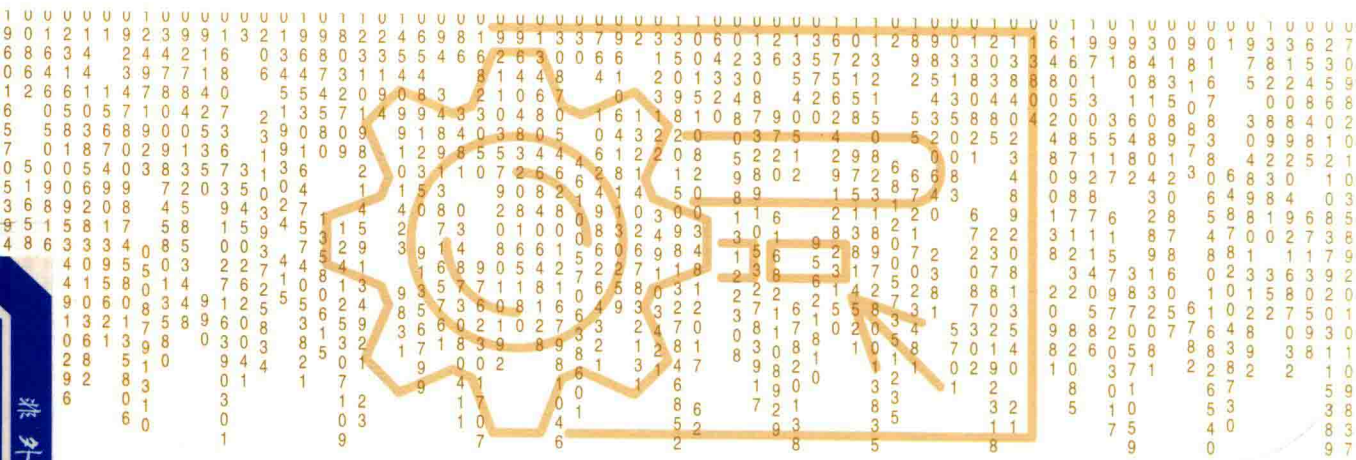




银行业信息科技风险管理高层指导委员会  
银行业信息化丛书

# 商业银行信息系统软件 测试数据治理实务

涂晓军 蔡仕志 焦伟 等编著



Software Testing Data Governance Practice  
for Commercial Bank Information Systems



机械工业出版社  
China Machine Press

海外借



银行业金融科技风险管理高层指导委员会  
银行业信息化丛书

# 商业银行信息系统软件 测试数据治理实务

涂晓军 蔡仕志 焦伟 等编著

Software Testing Data Governance Practice  
for Commercial Bank Information Systems



机械工业出版社  
China Machine Press

## 图书在版编目 (CIP) 数据

商业银行信息系统软件测试数据治理实务 / 涂晓军等编著. —北京: 机械工业出版社, 2019.1

(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)

ISBN 978-7-111-61417-3

I. 商… II. 涂… III. 商业银行 - 管理信息系统 - 系统软件 - 测试 - 数据处理 IV. F830.49

中国版本图书馆 CIP 数据核字 (2018) 第 259426 号

商业银行数据治理工作是整体数据治理工作的重要组成部分, 同时测试数据治理由于和软件测试紧密相连, 因此又具备治理目标和方法上的特殊性。本书首先回顾了商业银行数据治理和软件测试的发展概况, 然后详细介绍了测试数据治理的目标、范畴、架构、模型、生命周期、技术体系等方面的内容, 全面阐述了商业银行测试数据治理体系。全书内容分成理论篇、技术篇、扩展篇三个部分, 各部分章节的内容循序渐进、逐层深入、环环相扣, 覆盖了商业银行测试数据治理的方方面面。

本书可供从事商业银行测试数据治理相关工作的人员使用, 也可供其他行业的相关从业人员参考。

## 商业银行信息系统软件测试数据治理实务

出版发行: 机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码: 100037)

责任编辑: 鲜梦思

责任校对: 李秋荣

印刷: 北京诚信伟业印刷有限公司

版次: 2019 年 1 月第 1 版第 1 次印刷

开本: 185mm×260mm 1/16

印张: 19.5

书号: ISBN 978-7-111-61417-3

定价: 80.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 68995261 88361066

投稿热线: (010) 88379007

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzjg@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东

## 前言 |

随着商业银行业务的飞速发展，其信息系统日趋复杂，因此对软件测试工作也提出了更高的要求。在信息系统软件测试过程中，为了提高测试质量，测试者通常会采用具备一定复杂度的原始数据作为测试基础数据。原始数据作为商业银行的核心资产之一，往往涉及一定数量的敏感信息，因此，在软件测试过程中要充分考虑数据安全的因素，防范数据泄露风险。同时，为了提升测试效率和节约资源，测试者往往还需要在保留测试数据的关联关系和复杂程度的情况下，尽可能压缩数据量。如何较好地解决信息系统软件测试数据在使用过程中存在的这些矛盾，成为各家商业银行不断探索、不断研究的问题，而相关的专业书籍尚缺，亟须出版相应书籍，公开行业的先进管理经验和技术来供同业参考。

本书由中国农业银行统筹编写，是“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”的一员。本书详细介绍了商业银行测试数据治理的目标、范畴、架构、模型、生命周期、技术体系等方面的内容，为商业银行测试数据治理工作建立统一的规范，为应用先进的技术提供了书面参考。书中不仅包含了测试数据治理的基本理论和实践经验，还包含了近几年中国农业银行在测试数据治理研究方面取得的成果。

全书内容分成理论篇、技术篇、扩展篇3个部分，共10章。其中，理论篇包括前3章，简要介绍了与商业银行测试数据治理相关的基本理论。具体来说，第1章回顾了商业银行测试数据治理和软件测试的基本背景，阐述了商业银行测试数据治理的概况；第2章介绍了测试数据治理的架构和模型，并从全生命周期的角度介绍了数据治理工作的内容；第3章从组织架构、管理规范、管理流程、制度建设等方面全面阐述了

如何建立全生命周期的测试数据管理机制。技术篇包括第4~7章，详细介绍了构成商业银行测试数据治理的四项关键技术。具体来说，第4章介绍了如何建立测试数据脱敏技术体系，第5章介绍了如何建立测试数据瘦身技术体系，第6章介绍了如何建立测试数据准备技术体系，第7章介绍了如何建立测试数据复用技术体系。最后3章为扩展篇，延伸介绍了与测试数据治理相关且相对独立的3个领域。具体来说，第8章介绍了数据归档管理的相关内容，第9章阐述了测试案例标准化的相关内容，第10章阐述了数据风险评估的相关内容。

本书编写组由中国农业银行数据中心从事测试数据管理相关工作的技术人员组成，成员包括涂晓军、蔡仕志、焦伟、周期律、周晓聪、杨志刚、牛晔、谢彬、王子南、蔡立晶。

本书在编写过程中得到了郝毅、敦宏程、殷坤等行业资深专家的悉心指导，也得到相关单位的大力支持，在此，向所有指导、帮助过本书编写的领导及专家表示衷心的感谢。

尽管编写人员投入了大量精力，然而由于水平和能力有限，书中错误和疏漏之处在所难免，恳请各位读者不吝指出，批评指正。

编者

2018年10月

# 目录 |

前言

## 理论篇

### 第 1 章 商业银行测试数据治理概述 // 2

- 1.1 商业银行数据治理概述 // 2
- 1.2 商业银行的软件测试与测试数据 // 9
- 1.3 商业银行测试数据治理概况 // 13
- 1.4 本章小结 // 15

### 第 2 章 测试数据治理架构 // 17

- 2.1 测试数据治理架构概述 // 17
- 2.2 测试数据治理模型概述 // 19
- 2.3 治理模型实例：分层治理模型 // 21
- 2.4 测试数据生命周期治理 // 25
- 2.5 本章小结 // 29

### 第 3 章 测试数据管理机制 // 30

- 3.1 测试数据管理机制概述 // 30
- 3.2 组织架构 // 31
- 3.3 测试数据全生命周期管理规范 // 33
- 3.4 测试数据全生命周期管理流程 // 42

- 3.5 测试数据管理制度建设 // 48
- 3.6 本章小结 // 54

## 技术篇

### 第4章 测试数据脱敏技术 // 58

- 4.1 测试数据脱敏目标 // 58
- 4.2 确定测试数据脱敏范围 // 59
- 4.3 脱敏评价指标体系 // 60
- 4.4 脱敏技术框架 // 66
- 4.5 脱敏处理策略库建设 // 75
- 4.6 数据脱敏策略创新 // 78
- 4.7 应用案例 // 88
- 4.8 本章小结 // 92

### 第5章 测试数据瘦身技术 // 94

- 5.1 测试数据瘦身目标 // 94
- 5.2 确定测试数据瘦身范围 // 95
- 5.3 数据瘦身评价指标体系 // 99
- 5.4 主要数据瘦身技术方案 // 100
- 5.5 数据关联模型 // 102
- 5.6 基于关联聚类的关联关系梳理 // 112
- 5.7 基于关联聚类的瘦身技术 // 113
- 5.8 基于关联聚类的数据瘦身实施方案 // 117
- 5.9 应用案例 // 118
- 5.10 本章小结 // 120

### 第6章 测试数据准备技术 // 121

- 6.1 测试数据准备的意义和目标 // 121

- 6.2 测试环境准备 // 124
- 6.3 确定数据准备的内容 // 127
- 6.4 数据准备评价指标体系 // 131
- 6.5 三层结构的数据建模技术 // 133
- 6.6 基于模型的自动化技术 // 140
- 6.7 本章小结 // 147
- 第 7 章 测试数据复用技术 // 148**
  - 7.1 测试数据复用的原因 // 148
  - 7.2 复用方案设计 // 149
  - 7.3 知识领域复用方案 // 157
  - 7.4 技术领域复用方案 // 171
  - 7.5 管理领域复用方案 // 184
  - 7.6 本章小结 // 187
- 扩展篇**
- 第 8 章 数据归档管理 // 190**
  - 8.1 数据归档概述 // 190
  - 8.2 结构化数据归档方案 // 195
  - 8.3 非结构化数据归档方案 // 199
  - 8.4 本章小结 // 204
- 第 9 章 测试案例标准化 // 206**
  - 9.1 测试案例标准化的原因 // 206
  - 9.2 测试案例标准化方案 // 207
  - 9.3 测试数据标准化 // 215
  - 9.4 测试对象标准化 // 219
  - 9.5 测试逻辑标准化 // 234



9.6 测试执行标准化 // 256

9.7 测试流程标准化 // 261

9.8 本章小结 // 269

**第 10 章 数据风险评估 // 270**

10.1 数据风险概述 // 270

10.2 数据风险理论研究 // 278

10.3 数据风险要素 - 损失度评估 // 283

10.4 数据风险要素 - 风险概率评估 // 289

10.5 数据保护对策 // 294

10.6 本章小结 // 301

**参考文献 / 302**

## | 理 论 篇 |

# 第 1 章

## 商业银行测试数据治理概述

### 1.1 商业银行数据治理概述

随着现代信息技术尤其是互联网技术的发展，中国的商业银行陆续掀起了以数据大集中为主线的信息化建设浪潮。数据已经成为商业银行的重要资产之一，也成了商业银行最大的价值来源和风险来源。整合利用各类数据和信息，是现代商业银行信息化发展的重要基础。

数据治理是指对数据资产管理行使权利和进行控制的活动集合，包括规划、控制和执行等。商业银行的数据治理是安全稳定运营的需要，也是其风险管理、业务创新、合规操作等各方面的保障。

商业银行信息系统软件的测试数据治理是其数据治理工作的重要组成部分。测试数据治理虽然继承了数据治理的基本理论，但具备一定的特殊性。因此，在详细讨论商业银行信息系统软件测试数据治理实务之前，有必要对商业银行数据治理的相关内容进行简要的回顾。

#### 1.1.1 数据治理概述

##### 1.1.1.1 数据与数据管理

数据的一般性定义是指对客观事件进行记录并可以鉴别的符号，是对客观事物的性质、状态以及相互关系等进行记载的物理符号或这些物理符号的组合。它不仅指狭义上的数字，还指具有一定意义的文字、字母、数字符号的组合、图形、

图像、视频、音频等，是对客观事物的属性、数量、位置及其相互关系的抽象表示。

商业银行的数据在传统意义上是指金融产品、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。随着金融创新和信息技术的快速发展，商业银行的数据呈现几何倍数增长，传统结构化的数据逐渐扩展到以网络日志、社交媒体为代表的半结构化数据和非结构化数据中，并产生为特定业务领域服务的信息。

数据已经成为一种资产，数据资产也成为现代商业银行越来越重要的生产要素。不同于传统的财务资产，数据资产具有很多特殊性，包括搜集、存储、复制、复用等。互联网和物联网的发展，要求数据标准化，以便交换。由于数据涉及个人隐私和运行安全，所以它也是金融风险管理中的重要因素。综合来讲，数据是商业银行信息系统建设的重要资源，是业务与技术沟通的共同语言，也是商业银行价值提升的助推器。

传统的数据管理主要集中于研究备份管理和灾难恢复等数据可用性技术。随着大数据时代的到来，传统的数据管理体系体现出一些不足，包括缺少对数据管理进行专门监督和控制的组织，缺乏规范统一的数据标准和数据模型，缺少统一的主数据，缺乏统一的数据质量管理流程体系，以及数据全生命周期管理不完整等。

#### 1.1.1.2 数据治理的概念

数据治理是指对数据资产管理行使权利和进行控制的活动集合，包括规划、控制和执行等。对商业银行来说，数据治理是在明确数据责任的前提下，为促进数据有效使用和发挥业务价值而展开的一系列业务、技术和管理相结合的实践活动。通常来讲，数据是通过服务产生价值的一种资产，数据治理是在数据产生价值的过程中，治理团队对其的评价、指导、控制。

数据治理是在高层次上执行数据管理制度的活动。它是数据管理的核心职能，是规划、监督和控制数据安全、数据质量管理、元数据管理、数据生命周期管理等领域的管理。

#### 1.1.1.3 数据治理的体系

数据治理的体系包含两个方面：一是数据质量核心领域；二是数据质量保障机制。其相互关系如图 1-1 所示。

数据治理是一个系统化的工程，要有强有力的组织、合理的章程、明确的流程、完善的系统来保障。数据治理保障机制的作用是保障和规范数据治理工作，而数据治理的核心领域是保障机制得以落实的支撑。

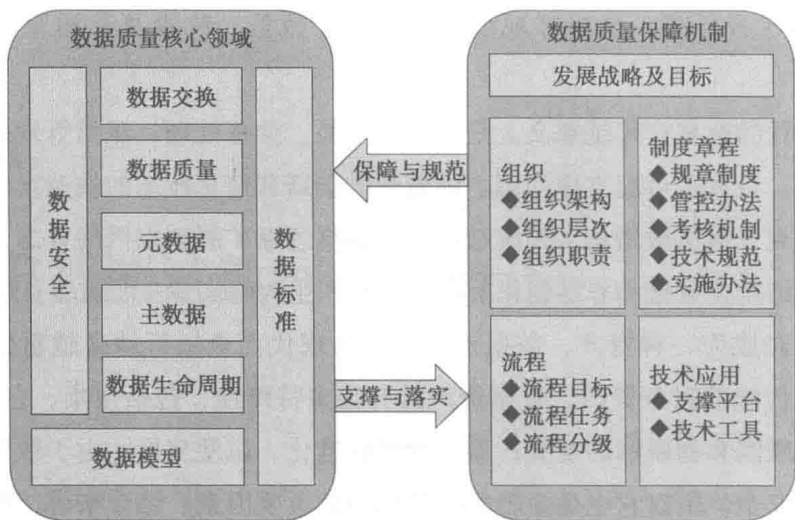


图 1-1 数据治理的两个方面

### 1.1.2 商业银行数据治理的背景

在商业银行的信息化发展过程中，数据治理的必要性和紧迫性更加凸显。由于早期系统开发都是围绕业务功能开展的，很少有与长期规划和管理相关的数据，因此大数据时代的到来为商业银行带来了许多考验。商业银行在长期经营活动中积累了大量数据，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、市场定价等领域，如果数据没有得到有效的治理，数据的问题就会导致相关业务无法正常开展，也会导致决策偏差。

全球经济一体化的深入发展和大数据时代的到来，为商业银行带来了新的机遇，也带来了严峻的考验。一方面，大数据为商业银行提供了新的沟通渠道、营销手段、新的竞争能力和更准确的决策依据；另一方面，大数据催生了新的金融形态和服务形态，考验着商业银行的数据治理能力。

综合来讲，商业银行数据治理的必要性体现在以下四个方面：数据治理是商业银行安全稳定运营的需要，数据治理是商业银行风险管理的需要，数据治理是商业银行业务及管理创新的需要，数据治理是合规的需要。

### 1.1.3 商业银行数据治理的目标

商业银行在战略决策、产品创新、客户挖掘等方面越来越倚重数据的挖掘和分析，出色的数据治理可以帮助银行加强风险控制，提升运营能力，提升关键业务的支持能力，提高合规监管和安全控制。

一般而言，数据治理的目标是提高数据的质量，保证数据的安全，实现数据和

其他信息资源在各部门的对接和共享，推进数据整合、数据应用和数据服务，从而提升机构的信息化管理和服务水平。

商业银行数据治理的内容，包括建立数据治理机制，制定数据治理制度及流程以及数据标准等。进行有效的数据治理对商业银行的主要价值和意义包括：有助于改善商业银行的决策能力，更精准地为客户服务，提高商业银行的风险管理能力，提升商业银行的运营管理水平。

#### 1.1.4 商业银行数据治理的现状

从国际视角看，越来越多的政府和企业开始进行数据治理的研究与建设，产生了丰富的理论和实践方法。近几年，大数据的发展和风险控制越来越受到重视，为了支撑大数据的发展和风险控制而进行的数据治理理论研究也就成为重中之重。

数据给现代银行工作主要带来三大方面的挑战：一是业务发展转型，经营理念向以客户为中心转变；二是竞争环境和流程日益复杂，必须强化自我约束；三是上级监管越来越细致。这些挑战给我国银行业带来进行数据治理整体战略的要求。为此，国内很多商业银行纷纷将数据治理提升到全行战略层面，在行内建立了包括管理层在内的数据治理委员会，对资金、人员等资源大力投入，在行外积极咨询专家和寻找解决方案，在业务发展模式上更注重对数据管理的提升和在内部挖掘潜力，并取得了一定的成效。

我国商业银行的数据治理还存在方法论层面和技术层面的不足。尤其是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等问题，而内部管理职能不清等因素是导致数据问题的根本原因。要持续、有序地推进银行数据治理工作，建立健全的数据管理体制，就需要完成以下建设：建立数据治理企业文化，建立完善的数据治理组织和制度，建立数据治理流程和有效的问责管理机制，建立数据治理各领域的管理体系。

#### 1.1.5 商业银行数据治理的核心领域

商业银行数据治理的核心领域包括数据模型、数据标准、元数据、数据分布和存储、数据交换、数据生命周期管理、数据质量管理、数据安全、数据服务和非结构化数据管理等。各个领域之间有机结合，相互支撑。

##### 1. 数据模型

数据模型是数据特征的抽象化，以表示实体之间的联系。数据模型是商业银行架构治理的基础领域，是数据治理的重要部分，也是数据治理的关键和重点。数据

模型一般包括概念数据模型、逻辑数据模型和物理数据模型。理想的数据模型应该具有非冗余、稳定、一致、易用等特征。

## 2. 数据标准

数据标准是商业银行建立的一套符合自身实际，且涵盖定义、操作、应用多层次数据的标准化体系。数据标准的建立是商业银行信息化、数字化建设的一项重要工作，各类数据必须遵循一个统一的标准进行组织，才能构成一个可流通、可共享的信息平台。数据治理对标准的需求可以划分为基础性标准和应用性标准两类。前者主要用于在不同系统间，形成信息的一致理解和统一的坐标参照系统，是信息汇集、交换以及应用的基础，包括数据分类与编码、数据字典、数字地图标准；后者是为平台功能发挥所涉及的各个环节，提供一定的标准规范，以保证信息的高效汇集和交换，包括元数据标准、数据交换技术规范、数据传输协议、数据质量标准等。

## 3. 元数据

元数据意指数据的数据，即对数据的描述。元数据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。元数据分为业务元数据、技术元数据和操作元数据。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。

## 4. 数据分布和存储

商业银行一般会关注总分行数据分布，主副本数据分布，核心交易系统、外围系统间的数据分布；同时，还要进一步考虑数据的存储。一般来说，商业银行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，并规范各数据存储区的数据范围，以及各存储区间的数据流向、数据服务和时效，从而用于指导全行系统的数据架构部署。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。

## 5. 数据交换

数据交换的主要目的是实现商业银行数据或信息的共享，满足商业银行在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，商业银行的数据交换包括商业银行与外部机构的数据交换和商业银行内部应用间的数据交换两大类。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。

## 6. 数据生命周期管理

数据是一种具有生命周期的资源。数据生命周期管理是一套策略、流程、实践、服务和工具，是对业务系统的数据进行贯穿其整个生命周期的管理，从数据创建之初直到最后被销毁。它使数据的业务价值和最合适、最具有成本效率的基础设施保持一致。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。

## 7. 数据质量管理

数据质量是用来描述数据对使用者有价值和有用的程度。数据质量的问题产生于设计、传输和使用、操作等方面，在不同程度上会影响商业银行的业务正常开展、用户满意度、信息可用性等。数据质量管理已经成为商业银行数据治理的有机组成部分，是商业银行分析决策、规划业务发展的重要基础。

## 8. 数据安全

数据安全是指保证数据的机密性（Confidentiality）、完整性（Integrity）和可用性（Availability），即信息安全领域的CIA安全三要素。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、账户信息安全，避免将信息泄露或遭到非法篡改。数据的安全性是商业银行各项业务有效开展的基础，也是保证银行业务连续性的关键。

## 9. 数据服务

商业银行数据治理的最终目的就是要利用数据服务于信息建设、部门管理、业务发展。更准确、更高效的服务是数据服务管理的目标。数据服务管理利用内部积累多年的数据来研究如何充分利用这些数据分析业务流程，优化客户营销策略，保障风险控制以及进行数据应用创新等。

## 10. 非结构化数据管理

在商业银行的信息化建设中，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。非结构化数据存在格式多样、来源复杂、结构不定、数据量庞大等特征，因此需要专门针对非结构化数据研究建立新的管理技术。

### 1.1.6 商业银行数据治理的保障机制

一般来讲，数据治理的保障机制包括战略、机制及各领域的技术支撑等方面。



商业银行数据治理需要建立数据治理的保障机制主要包括以下方面。

### 1. 企业发展战略

战略是选择和决策的集合，是为了实现目标而制订的行动方案。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、发展质量与发展能力的重大选择、规划。数据战略是企业发展战略中的重要组成部分，是指导数据治理的最高原则。商业银行要在企业发展战略框架下，建立数据治理的战略文化。

### 2. 数据治理组织

商业银行的数据治理需要总分行的联合，需要科技部门、业务部门等各方面的合作。数据治理组织包括制度组织和服务组织。制度组织主要负责制定数据治理和数据管理制度，通常包括数据治理委员会、数据管理制度团队等。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。从职责角度划分，商业银行数据治理架构包括决策层、管理层、执行层。

### 3. 数据治理制度和章程

健全的数据治理体系需要完善的制度和章程来保障。数据治理制度和章程类似企业的公司条例，阐明数据治理的主要目标、人员职责、决策权力和度量标准。制度和章程是确保对数据治理进行有效实施的职责问责制度，一般包括政策、制度、细则、规范四个方面。

### 4. 数据治理流程

商业银行的数据治理需要一套科学、完整的流程来保证。流程治理包括流程目标、流程任务、流程分级。全生命周期数据治理流程包括事前预防、事中检测、事后评估整改三个阶段。

### 5. 数据治理支撑系统

进行有效的数据管理和治理，需要建立丰富的数据治理工具和平台，并完善数据治理的支撑系统。其中包括数据的前台源系统、数据资产管理系统、数据质量管理体系等。相关系统之间相互协同、相互支撑，共同保障数据服务。

### 6. 数据治理成熟度模型

软件能力成熟度集成模型（CMMI）中的数据管理能力成熟度模型（DMM）将