

专家洞察

# 数据之巅 制胜未来

## 金融行业数据中台 五大关键成功要素

IBM 商业价值研究院



## 主题专家



**王保育**  
IBM 金融行业事业部  
资深解决方案专家  
baoyuw@cn.ibm.com



**吴大维**  
IBM GBS CBDS 团队  
副合伙人  
wudavid@cn.ibm.com



**张玉明**  
IBM GBS CBDS 团队  
资深数据架构师  
zhangyum@cn.ibm.com



**张宁**  
IBM GBS CBDS 团队  
资深数据治理专家  
znznbj@cn.ibm.com



**方杰**  
IBM GBS CBDS 团队  
资深人工智能专家  
fangfj@cn.ibm.com



**王莉**  
IBM 商业价值研究院  
高级咨询经理  
gbswangl@cn.ibm.com

扫码关注 IBM 商业价值研究院



官网



微博



微信



微信小程序

## 谈话要点

### 数据中台建设是一个系统化工程

数据治理重在执行，数据资产化事半功倍。

### 将数据转换为生产力

开放服务，建立生态，资产服务化数据变现。

### 践行数据驱动业务

平台化运营适者生存，数据业务化与众不同。

—

## 知易行难

1996 年，当 Bill Gates 说出 “Banking Is Essential, Bank Is Not” (银行业务是必需的而银行不是) 时，银行家们不会料到，仅仅过了 16 年，金融科技的大潮开始拍击金融业的基石，这句话一语成谶。

新的技术 A (AI) B (Blockchain) C (Cloud) D (Big Data) 被每一个银行家津津乐道。不管是否理解，在不得不拥抱新技术时，银行家们心里是恐慌的，包括对未来的恐慌。哪怕银行家们拥抱了新技术，他们还是忽视了 EFGHI……E 既是体验 (Experience) 也是生态 (Echo System)，F 是 (Frame) 重构，G 是 (Go) 试错，H 是 (Hypothesis) 大胆假设，而 I 是 (IOT) 万物互联，IOT 代表着所有技术和业务模式的最终趋势。

如果说，ABCDI 是风口，EFGH 是风向，那么风眼只有一个，那就是数据。在数据爆炸时代，全球如今每天创建 491EB 字节的数据。<sup>1</sup> 然而，在利用数据方面，一个企业所产生的数据中大约有 55% 没有得到使用。<sup>2</sup> 过去十年中，《财富》1000 强公

司中有四分之三已经被替换，领先的 79.4% 的高管担心被数据驱动的初创公司所颠覆，只有 7.3% 的企业对其充未来数据策略满信心。<sup>3</sup>

为了快速响应市场新一代客户需求，在前台，金融机构发展了新渠道来提高客户体验；在后台，增强了新核心来优化业务流程；而在中台领域，金融机构希望通过数据中台来实现数据资产化智能管理与共享，从而满足业务灵活性需求。

那么，什么是数据中台？数据中台的概念脱胎于中国互联网市场，最早由阿里巴巴提出。最初基于当时流行的 SOA 架构理念，将零售核心业务划分为多个应用组件，其中枢系统构成业务中台。随着阿里跨界多个混合业态的开展，在复杂的业态中实现数据的互联互通并产生洞察成为新难题。于是，与业务中台相对应的数据中台理念随之兴起。

包括银行业在内，业界对数据中台的认知尚未统一，存在多种不同观点：一部分观点侧重于业务视角，定义为云上数据中台业务模式；其他观点则认为，数据中台是聚合和治理跨域数据，将数据抽象封装成服务，提供给前台以业务价值的逻辑概念，或数据中台是数据服务 (Data Service) 工厂；也有观点认为企业中台是数字化平台战略的中枢系统，用以协调前台的快变与后台的稳定。我们认为，数据是金融机构的战略资源；只有得到有效地运用，数据才能转化为资产；而实现这种转化的关键就是数据中台。金融行业的数据中台是以数据为中心，智能化、全链路地开展管理、应用和服务的平台化体系。数据中台使得金融机构的产品、客户、渠道、流程、风险不再割裂。它以数据为生产要素，建立了新型的生产力和生产关系。

企业为什么需要数据中台呢？结合银行业 IT 系统现状及数据应用中的问题，数据中台建设的特殊使命主要在于解决以下四类核心问题：

- 1) 数据信任：因为数据质量不齐、获取难度大，数据消费者包括经营决策者对数据失去信心；
- 2) 需求响应：开发周期长、效率低、服务响应慢、计算资源紧张、数据时效性不强；
- 3) 协作效率：架构平台化、组织模式去中心化的趋势下，数据复用与协作越发重要；
- 4) 创新乏力：全局规划、确保共识，统筹数据积累，才能更好地开展创新。

知易行难，火爆的数据中台虽是行业共识，但我们到底需要什么样的数据中台？先看一些典型挑战：

- “盲人摸大象”：有多少数据理不明白，投入产出比说不清。往往只顾低头走路，却忘了抬头看路。如何实现“数以治用”？数据资产如何盘点和管理？如何衡量数据资产的价值？数据资产如何变现？
- “大炮打蚊子”：业务一线要的是“雪中送炭”型的数据服务，但数据分析团队却往往陷入“闭门造车”，难以成为一线的合作伙​​伴。结果是，大炮造好了，却只能用来打蚊子。数据如何真正赋能业务？数据资产如何沉淀和复用？
- “小马拉大车”：敏态业务往往采取临时开发的做法，缺乏全局性考虑，难以响应数据服务的千变万化，就像小马拉大车，根本拉不动。如何真正从全生命周期角度对数据进行管理？如何实现平台化运营？

以典型股份银行为例，数据全日采集的系统就超过二十个，总表数超过两万，总字段数超过五十万，如何找到和用好这些数据资产？确实是一个难题。用高薪打造的数据团队，却有 50-80% 的时间花在寻找多个来源的数据。单单是数据治理工具、资产管理应用或 AI 平台已经力不从心。

数据中台建设，本质上是一个复杂的系统工程，就像有些银行虽然已经构建了“一湖两库”（大数据湖、数据仓库和集团信息库），但还在继续借助外脑不断完善大数据服务云数据资产管理，并力争成为真正的数据驱动型企业。

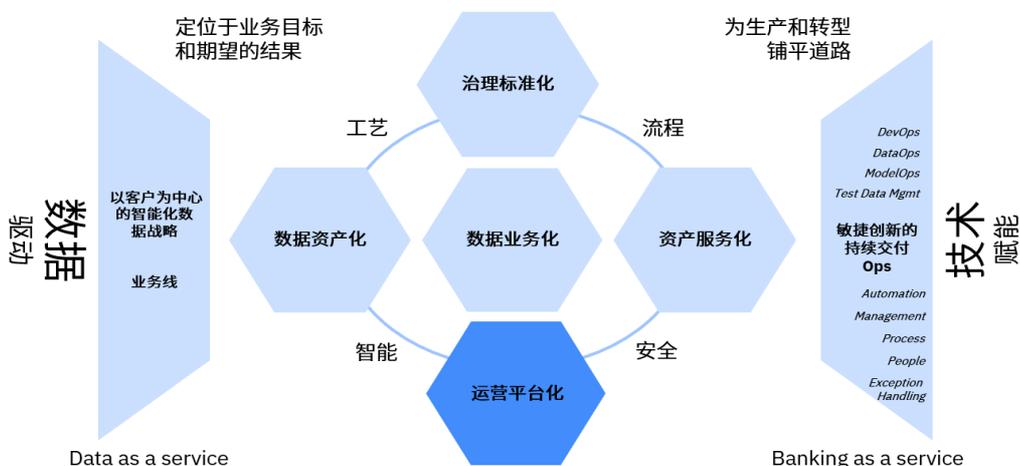
基于多年实践经验和本地化项目积累，我们总结了金融机构建设好数据中台的五大关键成功要素，分析了行动领域中的重点、难点、典型场景、参考案例和能力建议，以及相关的监管规定和行业标准。我们将这五大要素归纳为**四个现代化**（治理标准化、数据资产化、资产服务化、数据业务化）加**一个平台化**（运营平台化），简称为“4+1”（见图 1）：

- 1) **治理标准化**：数据为王，治理先行，数据管控制不容缓，贯标执行，行稳致远；
- 2) **数据资产化**：盘点数据，分条析理，数据资产化事半功倍，让数据资产成为企业的第四张“报表”；
- 3) **资产服务化**：开放服务，构建生态，资产服务化让数据高效变现，敏捷精准、快速制导；
- 4) **数据业务化**：沉淀智库，积微致著，数据化业务与众不同，高屋建瓴，反哺业务，驱动创新；
- 5) **运营平台化**：平台经济，统一运营，按需应变，平台化运营适者生存，衔接前后台，开放新新业态。

最后我们还提出了金融行业数据中台行动指南，把宏观战略咨询规划与微观落地具体执行相结合，将有助于把数据中台转型的愿景快速变为现实。通过业务协作以及平台化运营，实现数据产品化的开放服务，推动数据中台快速变现，打通数据大动脉，疏通服务微循环。从而更有效地面对金融科技创新和监管合规持续强化带来的冲击，并进一步提高金融机构敏捷性和灵活性以及运营效益和资本收益率。

图 1

数据中台的五大关键成功要素

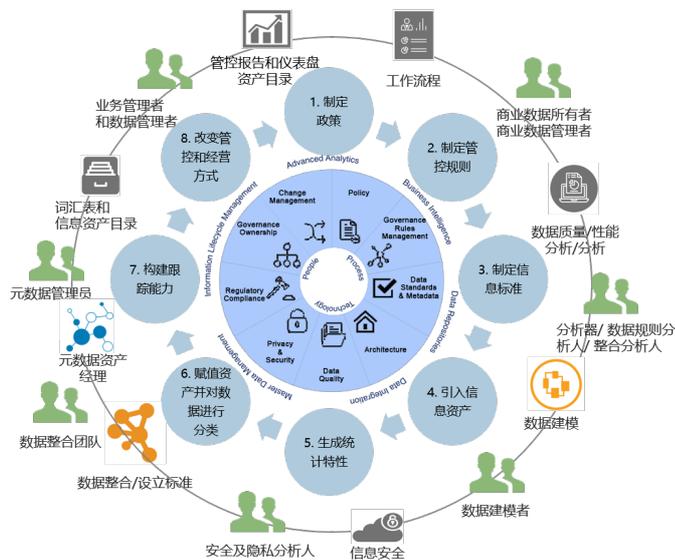


# 1. 数据为王，治理先行

## 数据管控任重道远

2018 年，银保监会发布了《银行业金融机构数据治理指引》，要求金融机构明确数据治理架构、落实数据管理和质量控制、全面实现数据价值。大多数金融机构花费了大量的时间和精力，处理杂乱无章、疏于整合的数据，但是依然存在诸多挑战：客户信息、产品信息、机构信息、员工信息、地址信息等数据缺失、重复或者不一致；客户数据分散且难以建立关联（如存贷款和卡客户之间），无法形成单一视图；没有准确识别产品的数据定义，无法准确统计和报表展现；业务部门之间较少分享数据，没有统一的数据模型标准和统一客户标识等。

图 2  
数据治理参考框架



成功的数据治理，是对企业中数据可用性、相关性、易用性、完整性和安全性的整体管理。它可以帮助企业管理其信息知识并回答各种问题，例如：我们对信息了解多少？这些数据来自何处？这些数据是否符合公司政策和规则？为了全面提升数据质量，保证数据的准确性、一致性、时效性、完整性和科学性，解决当前信息系统缺乏完整性、数据逻辑性不强的问题，金融机构需要将统筹规划与敏捷执行高度结合，采用业界成熟数据治理框架，加强治理专项整治，实现智能化数据治理，强化外部数据管理，从而循序渐进、持续优化，快速达到“数以治用”的目标。

### 1.1 数据治理框架

数据治理提供了一种全面的方法，帮助企业提升整体数据管理效率。在成熟的数据治理框架中，数据治理是协调人员、流程和技术工具，将数据视为企业的重要资产进行妥善管理，并能够加以良好利用（见图 2）。

### 1.2 专项数据治理

数据治理应当采取什么样的策略呢？我们建议，应当在全局规划的基础上，重视以需求为导向，以专项整改的形式快速提升重点数据质量，实现局部速赢。既要呈现业务部门的现实痛点，也要充分反映业务部门对深度应用数据的方向建议。数据治理专项整治过程中应注重下列内容：

- 业务主导，需求驱动：将业务价值和业务需求紧迫度作为选择数据专题的核心考虑因素之一；
- 基础先行，快速见效：优先实现共享性强、影响面广的基础应用，释放业务价值，快速推进业务发展；
- 横向联动，纵向贯通：善于运用各类数据专题之间的成果，使得大数据应用功效倍增；将数据专题与具体业务流程的结合，例如通过客户标签画像将客群战略、客群分析、营销策略、活动执行、渠道协同等环节贯通起来，实现端到端的应用。

### 1.3 AI+ 数据治理

金融行业的海量数据增长，给数据治理工作带来的极大的压力。针对这种情形，如何才能提高数据治理的效率？业界已经开始尝试 AI+ 数据治理的途径。

例如，IBM 提出“认知型数据治理”的理念并且应用于实践，使用机器学习技术，帮助数据治理顾问构建数据目录。利用机器学习技术，通过反馈学习，自动匹配相同语义的数据，如果匹配的置信度得分低于某个阈值，系统就会将候选数据记录提交给人类专家做出判断。这时专家只需处理整个数据集中具有弱匹配特征数据的子集(对比原有数据集合，这个数据子集通常占比非常小)，从而大大提高工作效率。

智能化的数据治理还可以帮助数据使用者寻找和发现所需的数据集。例如，一位新入职的数据科学家接受了开发机器学习模型的任务，用于检测特定产品或服务的客户流失情况。他可能并不知道可以使用哪些数据集来开始任务。借助基于 AI 的 IBM 数据治理技术，这位数据科学家可以通过 IBM 的智能化数据管理平台，轻松搜索“客户挽留”等业务术语，从而获得所有相关实体的图形视图。然后，就可通过深入分析来了解数据的质量和真实性。

### 1.4 外部数据治理

在大数据环境中，每个企业都不是孤单的个体，而是和其他的数据主体共生于复杂的大数据生态中，频繁地和其他的数据主体之间进行数据的往来交互。因此，企业不但需要考虑妥善管理企业自身由于业务发展所产生的数据，也要把眼光拓展到如何管理从外部环境攫取的数据。外部数据治理包括四个要素：识别、引入、应用和综合管理。

- 识别：企业在引入外部数据时，要充分调研企业的内部需求，再评估外部环境中是否有数据主体可以提供相应的数据、获取数据的成本，从投入产出比的角度分析，是否值得投资；

## 借助人工智能提升数据治理效率

为提升整体数据管理效率，全面发挥数据价值，某金融客户与 IBM 合作开展了数据治理项目。该项目首先完成了数据分类、数据分级和安全保护等设计。接下来需要实现客户重要信息的定位，即，将客户重要信息映射到各个系统数据库表的具体字段。

客户当前的数据字典和元数据描述存储于 Excel 中，包括 10 个系统、1 万余张表、20 多万个字段。其中字典数据质量欠佳，有一半字段没有中文名，而且一部分字段的英文名是由拼音首字母组成（例如，资金流向和证件类型都是 ZJLX）。如果按照英文到中文的关键字进行匹配和定位具体字段就很容易出错。

如果采用人工方式锁定和分析个人客户重要信息的分布情况，至少需要一名有经验的数据分析师 15 个工作日以上的时间。针对相关法律规定、监管文档及附件进行个人客户相关的重要信息进行搜集得到的大约 50 种信息项，IBM 使用词向量技术，泛化为 12 类客户重要信息，包括涉及联系电话的手机号、座机号在内的词库扩展，并使用泛化后的信息和业务系统数据库进行匹配。

借助人工智能技术，IBM 帮助客户在 3 个工作日内完成了客户重要信息定位的全部工作，而且准确性优于人工分析的结果。

- 引入：一旦决定引入外部数据，则需验证数据提供主体是否可以“安全可靠”地提供外部数据，即数据的获取应符合相关法律法规要求，且外部数据的提供者可以按照约定的时间和方式，稳定地提供高质量的数据。企业应该设计统一的管理流程来确保外部数据的安全接入；
- 应用：数据引入后，企业应该建立外部数据整合平台，为所有内部用户提供数据的整合和加工服务，包括外部数据之间的整合加工，以及外部数据和内部数据的关联和融合处理；
- 管理：应建立覆盖外部数据引入、加工、应用的完整监控体系，持续监控数据流转的各个环节，一旦外部数据的提供质量、应用成效等发生巨大变化，就需要重新调整该项外部数据的采集策略。

### 1.5 打造 OneID 体系

在中台的数据环境中，我们从各个外部和内部数据源获取了纷繁多样的数据，期待能够链接这些数据，并赋能业务，得出信服的商业决策。在这些从不同的终端、以不同的数据格式采集出来的数据“大杂烩”融通的过程当中，我们也会遇到把包含不同信息、覆盖差异化数据的样本进行关联和对接的棘手问题。

如何才能有效解决这个问题？我们需要紧紧抓住最重要的业务对象进行唯一性识别，建立一整套 ID 系统，包括个人客户 ID、商户 ID、对公客户 ID、设备 ID、地址 ID、社交媒介 ID 等等。

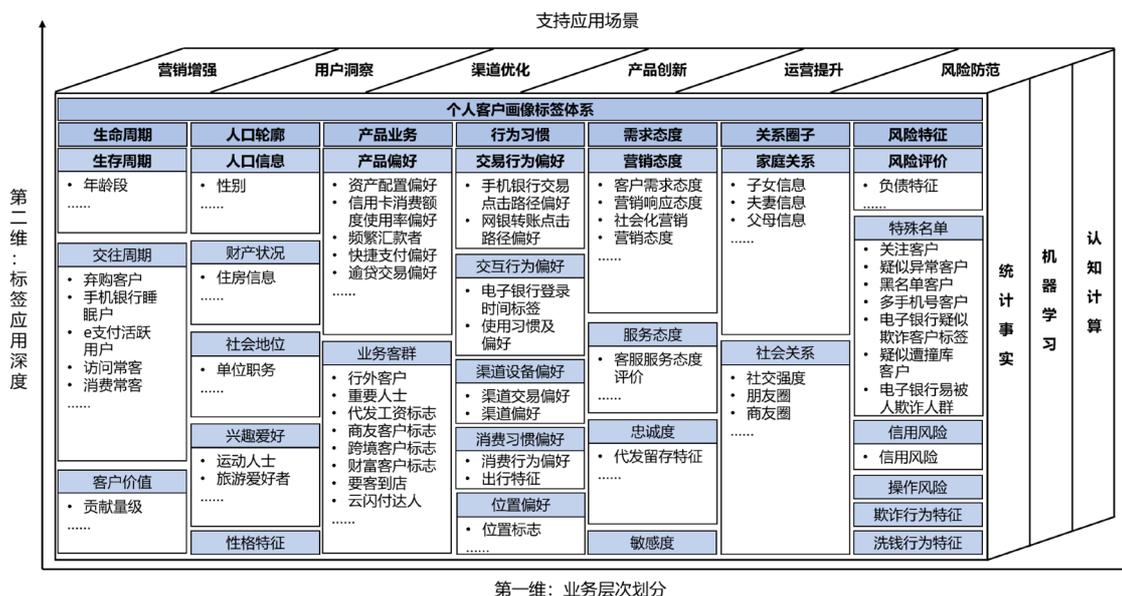
基于这些核心实体，将来源于不同数据源、不同形态的数据进行打通和整合。通过这种融合，建立丰富的标签信息，并且从业务视角串联关键实体之间的关联关系，构建全面的信息视图，刻画实体的业务活动和互动行为。这个过程，即为构建 OneID 体系的过程。

OneID 体系将极大地帮助我们理解客户以及企业内部的运营情况。并基于这种理解，进一步提升和改善业务效能和客户体验。打造 OneID 体系的重点有三个（见图 3）：

- 技术基础：使用 AI 技术处理指纹、人脸、语音特征等生物识别信息，从结构化和非结构化数据中提取社交信息、手机号、证件号码、邮箱、地址和网址等个人信息，关联并识别客户身份。提升客户识别的准确性，并实现外部、内部数据的打通和对接；
- 甄别筛选：对于重复的数据要素，应选取可信数据作为参考来源，并就数据的规范性进行验证。对于不符合规范性要求的数据，要在 One ID 体系建设的过程中明确规范要求物理落实，实现关键数据质量的提升；
- 衍生增值：OneID 的数据融合结果，不仅限于数据源的整合加工。综合运用数据挖掘以及 AI 技术，围绕业务核心实体，进行多样化的分析，衍生和加工丰富的标签数据，推行出新的高业务价值数据，并构建完整的标签体系，从而支持多样化的应用场景。

图 3

打造 OneID 数据体系示意



## 2. 盘点数据，分条析理

### 数据资产化事半功倍

虽然企业越来越认识到数据作为资产的重要性，但是，许多机构还不能很好地解决数据不全、不准、不通等难题。数据资产化强调充分融合业务、技术和管理，目的在于帮助企业摆脱数据孤岛和数据难找、难用以及数据一本糊涂账的困境，从而对数据进行有效控制、共享、保护、交付并提高数据资产的价值。这涉及到一系列不同的数据管理功能，包括企业数据架构、数据模型与设计、数据存储与操作、数据安全、数据集成与互操作性、文件和内容、参考数据和主数据、数据仓库和商务智能、元数据、数据质量等内容。

#### 2.1 数据资产的“采”、“析”、“用”、“治”、“运”

如何才能最大化地发挥数据资产的价值？一个行之有效的企业级数据资产管理框架必不可少。该框架需要涵盖数据资产从采集到应用及运营管理的“采”、“析”、“用”、“治”、“运”全过程，同时通过平台和保障机制为数据资产管理提供平台支撑和制度规范保障(见图4)。在数据资产化进程中，有五项工作至关重要，包括：全面认识数据资产、数据资产盘点、数据资产管控、数据资产分析和数据资产价值评估。

图 4

企业级数据资产管理框架



#### 2.2 数据资产的三维立体视角

只有客观、准确地认识数据资产，才能对其进行有效地管理和应用，我们可以从业务模型、应用架构和数据湖三维立体视角来全面认识数据资产的范围和形态。

- **业务架构视角：**业务架构中的实体模型将流程模型及产品模型中产生的所有数据需求进行逻辑化和抽象化表示，反映企业的战略目标在能力举措实施过程中所用到的所有数据实体及实体与实体间的关系，是对企业数据资产的一种抽象化表达。例如，IBM 金融行业数据模型 (FSDM) 对金融机构实体模型进行表示，包括客户、机构、产品、条件等九大主题等；
- **应用架构视角：**数字化转型中，交易型应用、交互型应用、分析型应用、服务集成型应用架构相互融合，每个应用域产生和关注的数据类型和形态有所不同。例如，分析型应用关注的资产数据侧重在标签类数据、多维分析数据、机器学习模型、数据产品等基于交易型应用系统产生的原始数据所进行深加工后的洞察数据，描述数据资产要充分加以考虑；
- **数据湖视角：**数据湖所涵盖的数据类型不单纯是对数据产生层数据的镜像。从数据集中管理、分析挖掘甚至审计的视角对数据资产类型有更广泛的定义，例如，典型类型包括描述性数据、存放性数据、历史数据、汇集性数据、上下文数据、发布型数据等。

预览已结束，完整报告链接和二维码如下：

[https://www.yunbaogao.cn/report/index/report?reportId=1\\_38330](https://www.yunbaogao.cn/report/index/report?reportId=1_38330)



云报告  
https://www.yunbaogao.cn

云报告  
https://www.yunbaogao.cn