

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



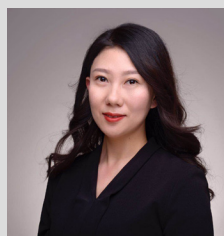
王保育
IBM 金融行业事业部
资深解决方案专家
baoyuw@cn.ibm.com



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。¹ 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。² 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。³

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：

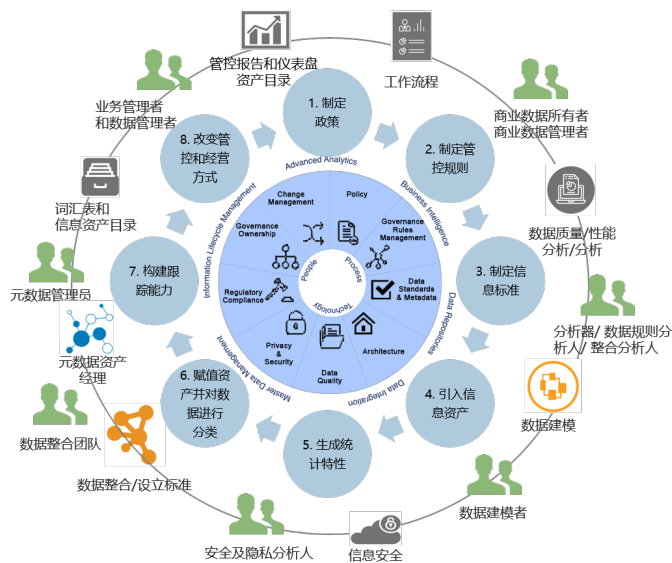
- 1) 数据信任：因为数据质量不齐、获取难度大，数据消费者包括经营决策者对数据失去信心；
- 2) 需求响应：开发周期长、效率低、服务响应慢、计算资源紧张、数据时效性不强；
- 3) 协作效率：架构平台化、组织模式去中心化的趋势下，数据复用与协作越发重要；
- 4) 创新乏力：全局规划、确保共识，统筹数据积累，才能更好地开展创新。

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。

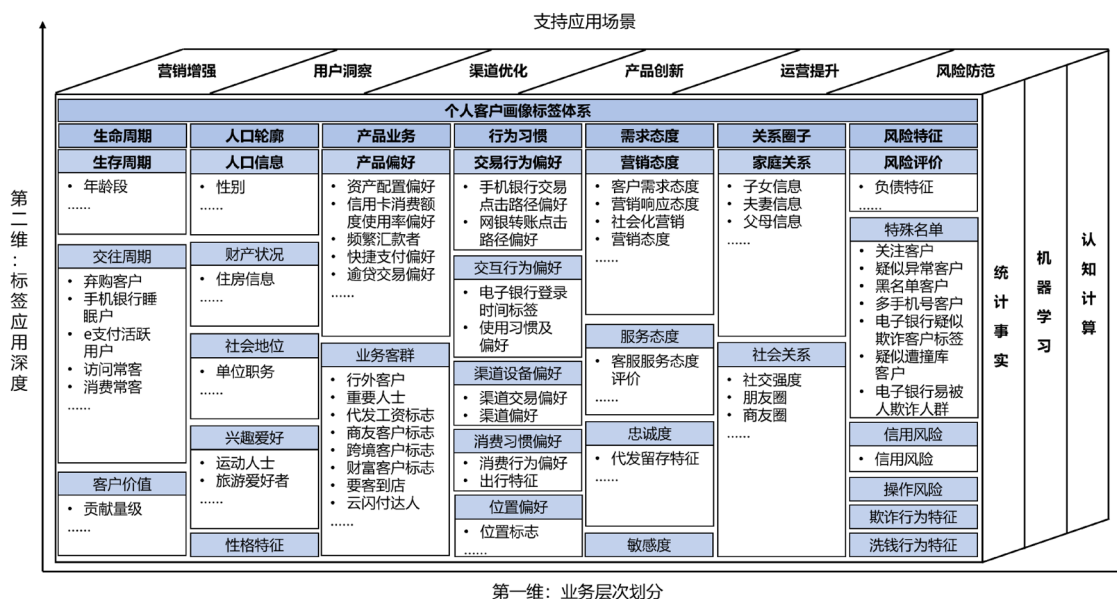
基于这些核心实体，将来源于不同数据源、不同形态的数据进行打通和整合。通过这种融合，建立丰富的标签信息，并且从业务视角串联关键实体之间的关联关系，构建全面的信息视图，刻画实体的业务活动和互动行为。这个过程，即为构建 OneID 体系的过程。

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

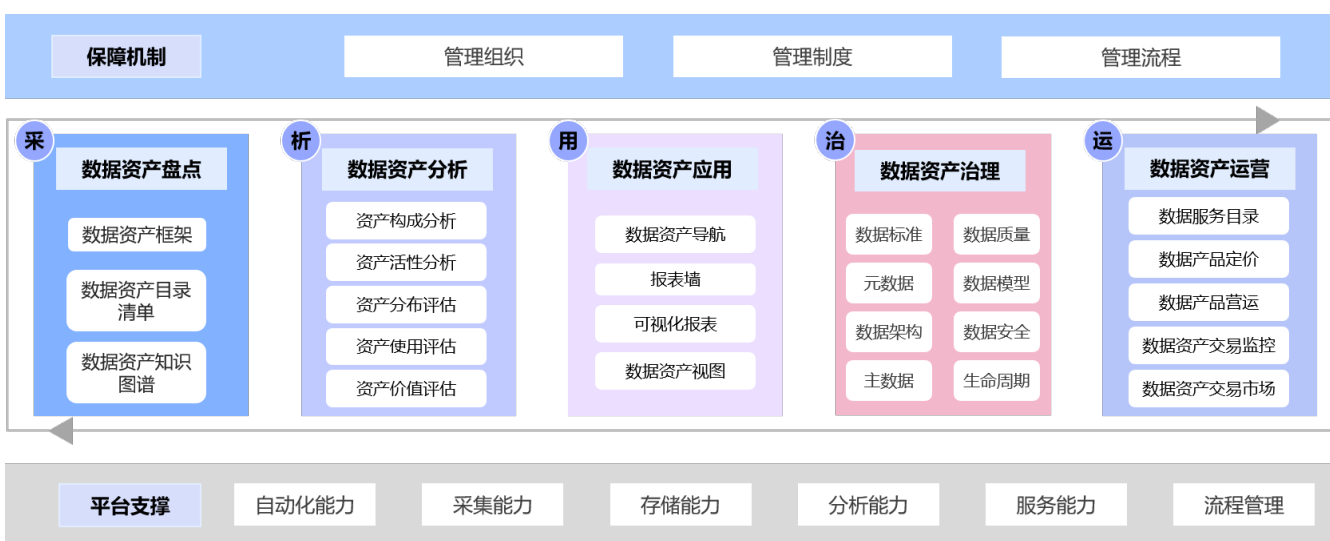
虽然企业越来越认识到数据作为资产的重要性，但是，许多机构还不能很好地解决数据不全、不准、不通等难题。数据资产化强调充分融合业务、技术和管理，目的在于帮助企业摆脱数据孤岛和数据难找、难用以及数据一本糊涂账的困境，从而对数据进行有效控制、共享、保护、交付并提高数据资产的价值。这涉及到一系列不同的数据管理功能，包括企业数据架构、数据模型与设计、数据存储与操作、数据安全、数据集成与互操作性、文件和内容、参考数据和主数据、数据仓库和商务智能、元数据、数据质量等内容。

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，有五个工作至关重要，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为广泛的定义，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。

2.3 数据资产盘点

为什么需要做数据资产盘点呢？主要是为了掌握整个企业的数
据有什么、是什么、在哪里、谁能用。通过盘点数据资产，有效支
持业务用户与分析师团队从海量数据中更快地、准确地获取所需
数据，实现数据供给与数据需求之间的平衡。在数据资产盘点核
心过程和步骤中（见图 5），行业数据模型资产和数据资产目录
总账对盘点结果影响重大。

- **行业数据模型资产：**借助成熟的数据模型资产，对企业开展数
据资产盘点工作将起到事半功倍的效果，也可以少走很多弯路。
例如，某银行在新一代核心系统建设中利用 IBM 金融服务数
据模型 (FSDM) 构建起一个覆盖领域最为全面的企业级数据模
型资产；某证券公司基于 IBM 证券期货行业数据模型 (SDOM)
模型对经营及客户数据按照重要性和敏感性进行分类分级，从
而满足监管单位的严格要求；某寿险公司基于 IBM 保险业信
息模型 (IIW) 模型中的 21 个主体域构建整个企业级的数据资
产目录分类框架，在 6 个月时间内完成对全公司核心业务数
据资产的盘点。

- **数据资产目录总账：**数据资产目录总账是指数据资产目录
的视图框架和结构，由代表不同类型的数据资产分录账组
成。它定义了整个企业级数据资产分类的全貌，是数据资
产分类的顶级框架。数据资产目录总账可以帮助企业对数
据资产概况了如执掌，游刃有余的驾驭复杂的数据问题。
例如，IBM 帮助某银行立足于大数据云平台建设契机，
全面定义数据资产内容，梳理数据资产定义信息，构建起
一套符合该行历史数据现状同时又满足大数据云平台发展
要求的数据资产信息目录。

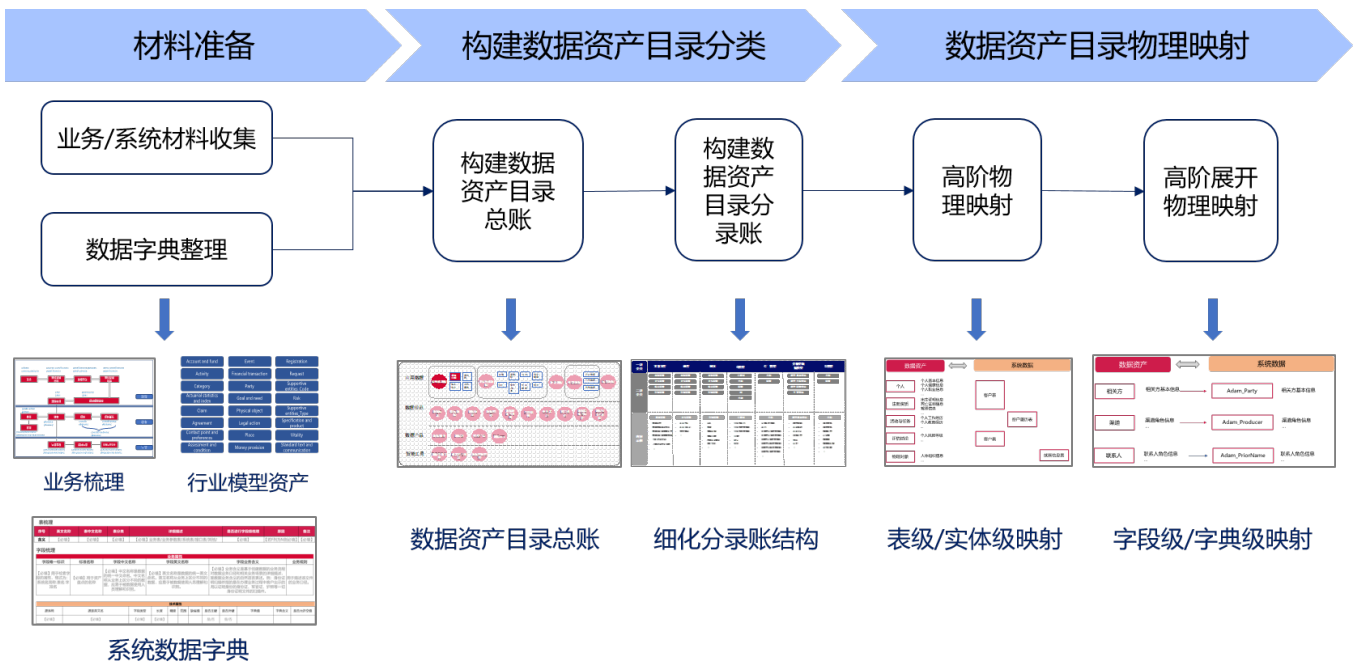
2.4 数据资产管控

数据资产盘点历经千辛万苦，但如果没有建立相应的运营管
控流程，一段时间后数据资产可能又会沦为数据沼泽。因此，
数据资产运营团队必须及早建立数据资产运营管控流程，对
数据资产的属主、使用者、加工者进行责任认定；建立基线
管理机制，有条不紊地对增量数据资产进行科学合理的把关
和登记。

数据资产管控的关键环节包括：年度计划制定、变更申请 -
受理 - 核准 - 修订、数据资产发布和版本维护、数据资产引
用和执行、数据资产培训等。

图 5

数据资产盘点的核心步骤



2.5 数据资产分析

对数据资产进行量化分析，可以帮助数据资产管理者制定更为科学的数据资产管理政策，合理调配资源，识别优先级最高的问题。典型的数据资产分析工作包括四个方面（见图 6）：

- 数据地图：支持用户在视图中查看全部的数据资产，包括数据概览、数据资产容量与数据资产分布；
- 数据资产质量评价：一是自动化监测入湖数据的数据质量情况，例如监测数据表的空值率；二是用户评价，支持数据消费者对数据资产使用的准确性和及时性评价，倡导“数据民主”；
- 合标分析：支持按多种维度统计入湖数据的合标情况，包括部门、应用系统、分区等；按照湖内部署的规则，统计已贯标和未贯标的字段数量、贯标率等；
- 数据热度分析：支持按照多种维度展示数据资产使用频率，通过数据消费活跃度体现出数据供给者对数据应用做出的贡献和价值。

2.6 数据资产价值评估

数据资产价值包括内在价值和外延价值。内在价值，涉及数据规模、数据质量、数据失效；外延价值，涵盖成本价值、市场价值和经济价值。数据资产价值评估近年来在数据资产管理界比较热门，涉及对数据资产的确权 and 定价。具体而言是对数据资产的成本定价、经济价值、效能价值、业务价值进行评估和管理的相关活动。

图 6

数据资产分析典型场景示例



数据资产化的典型场景

场景 1:

为了更好地开展创新研发和数字化转型，某银行亟需加强数据体系建设，弥补数据规划的缺失和大数据能力的不足。IBM 帮助该银行开展了全面数据治理、构建了数据资产管理平台，并设计了智慧营销数据模型。

通过数据资产盘点，该银行实现了对全行数据资产的全面管控，包括 500 多项数据标准，100 余项数据质量规则，涉及客户、机构、产品、新资本协议等。数据资产管理平台集成了工商、万得、路透、电商贷等外部数据，并连接了数据仓库、大数据平台、风险集市等系统，有效管理了基础数据标准的 7 大类数据。IBM 还从客群属性、交易行为等维度出发，运用大数据建模方法，对小微、理财、个贷等 7 大客群开展了价值贡献分析，建立了 18 个智慧营销价值提升子系统模型并完成了试点。

数据体系的建设，促进了全行数据质量的提升，有效支持了该银行的数字化转型和经营发展。

场景 2:

某金融控股集团，希望利用数字化手段发挥集团管控的作用，提升子公司的数据质量。

IBM 帮助该集团设计了数据资产价值评估的框架，从 7 个维度对各个子公司的数据资产情况进行了评估和排名，用以指导子公司改进数据资产质量。

该项目为集团公司未来开展数据整合分析、打通客户壁垒、实现整合营销奠定了基础。量化的数据资产评估框架，为集团的数据资产管控提供了决策支持。通过对数据资产开展穿透式管理，为集团管控提供了有力的数字化抓手。

3. 开放服务，构建生态 资产服务化数据变现

数据资产是数据中台的核心资源和引擎，只有将数据资产进行更为彻底的服务化，才能实现赋能业务的终极目标。在提供数据资产化服务时，金融机构首先要遵守行业监管等相关规定，例如，国家网信办于 2018 年颁布的《金融信息服务管理规定》、央行于 2020 年发布的《个人金融信息保护技术规范》等条例。数字资产服务化的成功之道在于构建并交付面向合作伙伴、提供商、供应商或消费者开放的新型金融生态并促进其互操作。

3.1 如何开放数据服务？

在数据资产变现中，开放接口 API 已经成为企业扩展产品、获取客户、帮助合作伙伴提供高价值服务以及扩张生态系统的关键步骤。API 作为一种敏捷的、可伸缩的、可消费的业务即服务模式，就像云计算改变 IT 交付模式一样，正在显著地改变数据服务消费和应用程序开发市场。

基于开放平台和互联网模式，开放数据服务通过应用编程接口 (API) 这一业务协作语言，快速实现 API 共享、保护、分发、

控制和盈利，并以安全、合规、可控的方式实现数据访问，助力实现 API 盈利。例如，访问基础数据或服务时免费，访问高价值数据或服务时则需支付相应费用，收取平台访问费用。开放数据服务有四项重要工作，包括：建立数字服务生态、资产服务化启航、规范开放标准接口、构建 API 平台能力（见图 7）。

3.2 建立数字服务生态

数字资产服务化的关键是明确 API 业务目标和绩效指标。具体举措包括：通过业务目标确定 API 对象是内部客户、合作伙伴还是外部客户；通过行业趋势、竞争分析确定数字资产服务业务模式和盈利模式；通过绩效指标，梳理系统的规模和部署需求，包括系统特色价值，评估现有资产的竞争力以及目标消费群体的需求，定义将要提供 API 数字服务的领域和场景；评估将要开放的 API 是否符合监管要求，是否符合保护用户隐私的要求；制定 API 平台总体发展和运营的评价体系以及评价指标；制定 API 平台的规章制度、业务条款、用户协议，对 API 文档、知识产权相关工作及其他对外发布的网站内容进行审核，包括 API 版本发布流程、发布方式和发布计划，同时要对 API 使用者进行大力支持。在数字资产服务化方面，已经有一些行业做了积极尝试，参见下页边栏案例。

图 7

开放数据服务的重要工作



资产服务化实现数据变现

某电信公司基于其海量的用户数据，通过资产服务化实现了数据变现。该公司归纳了 3700 余个用户标签，覆盖 4000 余个手机品牌、9 万多个终端型号数据。

基于这些数据，该公司提供了多元化的数据洞察服务，以满足不同的客户需求。例如，可以根据用户需求，提供定制化报告服务、个性化模型训练，包括基于居民区、工作区与商圈的人群分析、基于终端的人群分析等。

该公司开发的风控产品，融合了三大运营商的脱敏数据，在充分保障用户隐私安全的前提下，能够提供反欺诈、信息核验和风险评估服务。例如，面向外部合作伙伴提供标准 API 接口服务的产品就有 60 多个，并予以明码标价。该公司提供的个人风控 - 信息核验服务和个人风控 - 用户评级服务，能够帮助金融机构为贷前用户的信息核验提供支撑，包括证件、手机号、姓名两两一致性验证、职住核验、号码风险检查等。

此外，该公司还开发了银联合作产品，例如银联_消费评分（包括消费自由度、消费能力、消费趋势）、银联_消费画像（包括月交易水平、交易频度）等，同时提供标准化 API 网关服务，支持多租户 API 的注册、发布、订阅、在线测试、访问验权、监控分析等功能，从而使数据发挥更大的社会价值。

3.3 资产服务化启航

数据资产服务化起步时，要从两个方面着手做好准备（见图 8）：

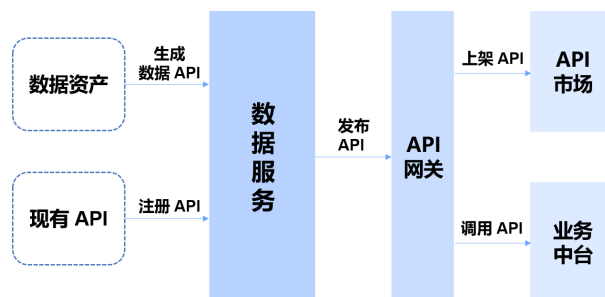
- 建立数据产品服务目录，树立新思维：整合和优化现有的各类数据服务，形成多层次、可感知的数据产品服务目录，包括数据资产查询服务、统计分析指标服务、公共建模指标服务、标签数据服务、决策分析模型 / 报告服务等。树立“业务数据化，数据业务化”新思维，依托数据产品服务，开辟数据中台赋能业务一线的“新战场”；
- 采用微服务架构，统一管理 API 服务：采用基于微服务框架的技术架构，将数据产品服务目录进行微服务的分拆和封装，形成数据服务 API，建立与业务中台服务交互，为业务中台提供“火力”支援。统一管理 API 数据服务，提供核心数据能力跨系统、跨平台的复用和共享。快速生成、注册、统一管理、发布、促销数据服务 API。

3.4 规范开放标准接口

银行业架构网络 BIAN（Banking Industry Architecture Network）是一个全球性协作的非营利性生态系统和开放、独立的社区。它由全球的领先银行、技术提供商、顾问和学者组成。它定义了一个通用、标准、灵活的业务体系结构，包括：银行业常见的服务域、业务场景、服务操作、业务对象、业务功能、API 接口，从而促进了全球银行业协作和服务开放。

图 8

支持数据资产变现的微服务架构



OpenAPI 规范 (OAS) 是由 Linux 基金会社区驱动支持的与语言无关的全球 RESTful APIs 标准接口, 使人类和计算机可以发现和理解服务的功能, 而无需访问源代码、文档或通过网络流量检查。通过有效映射与之关联的所有资源和操作来轻松开发和使用 API。

Swagger 是一套围绕 OpenAPI 规范 (以前称为 Swagger 规范) 构建的开源工具, 帮助支持 OpenAPI 规范的实现, 包括设计, 构建, 编写和使用 REST API 以及客户端和服务端代码生成、部署、监控、文档, 可视化 API 操作, 并促进内、外部使用者快速采用 API。

3.5 构建 API 平台能力

API 平台为用户提供了一个完整的生态系统, 涵盖了从 API 开发、部署、再到维护的整个生命周期, 包括 API 网关、API 管理平台、API 开发者门户。

API 网关作为 API 调用的进出口, 负责 API 的安全管控、策略执行、数据采集、路由调度等以保证 API 调用符合企业安全和规范, 通常还具有负载均衡和缓存的功能。

API 管理平台负责从创建、定义、组装、版本到发布的 API 生命周期的管理, 并保存和发布 API 定义到 API 网关, 提供 API 测试, 同时支持 API 提供者对 API 组合、转换和标准化, 从而能快速地把资产和服务暴露成 API。

API 开发者门户通常是一个自助服务的门户, 也有开发者社区的功能, 包括 API 论坛、API Blog 等互动功能, 适用于开发人员集中发现 API、学习 API、在线测试 API 和注册应用, 同时通过收集 API 使用者的反馈, 有助于持续改进和增强 API。

API 平台环境应支持多租户架构、应用隔离和集中 API 管理、发现和测试的需求, 同时提供 OpenShift 集成和混合云支持, 在安全方面, 支持各种加密、身份验证和授权协议, 包括 OAuth2.0 授权方式、LDAP 集成、基于角色授权; 在管理方面, 应提供 API 使用情况分析和面板功能, 支持实时监控 API 各结点的运行情况, 配置系统资源, 进行系统资源扩展。

4. 积微致著, 高屋建瓴

数据业务化助力与众不同

在激烈的市场竞争中, 金融机构希望数据管理部门改变以往的被动供数模式, 更加主动地参与全行业务创新, 使数据服务嵌入到企业的业务中, 并成为其中重要部分, 即做到真正的数据业务化, 实现数据驱动业务。

通常, 原始数据并不能直接使用, 单一模型也不能直接发挥作用, 这就需要数据中台提供有价值的、可复用的、标准化的数据服务模块, 我们称为智库资产。它将有效地提高数据资产回报率, 同时改善后续项目质量、降低风险、提高全行生产效率, 并避免将来重复制造同一个轮子。

4.1 可复用智库资产

可复用智库资产, 是提供在给定条件下解决一类问题的可复用工件, 包括软件开发生命周期中的数据模型、AI 模型、Graph 模型、代码库、测试脚本等等。工件描述应包括业务驱动、IT 驱动、业务收益、使用条件、适用范围等内容, 同时应描述如何使用, 包括相关规则和指导及参考案例, 以便于用户定制和裁减。

数据业务化的关键是要静心沉淀数据资产化、资产服务化过程中的知识经验, 同时正确地把握可复用工件的**颗粒度** (资产解决问题范围的程度)、**可变度** (资产可被修改程度)、**实现度** (资产提供的完备程度), 并形成真正的可复用智库资产。

可复用智库资产的描述可借鉴下面的参考属性:

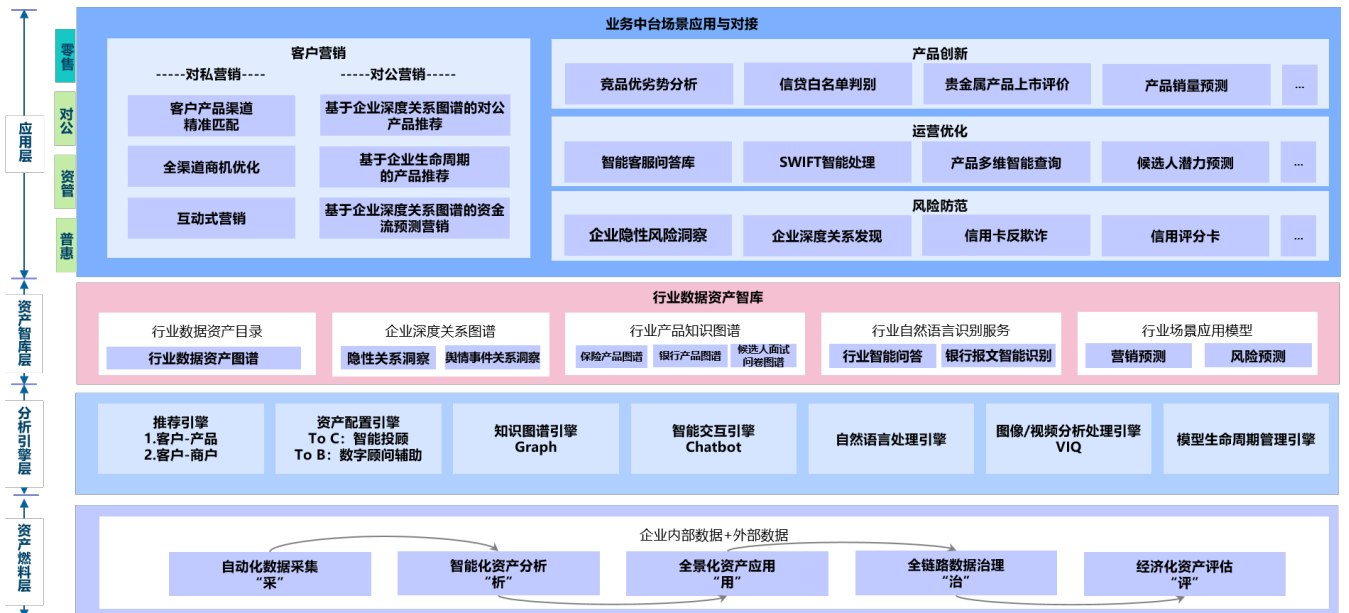
- 资产描述: 名称、摘要、作者、社区、创建日期、类型、用途、行业、支持范围;
- 资产治理: 版本、状态、上次更新时间、人员、审核日期、人员、审核截至日期;
- 资产安全: 保密级别、安全分类、访问控制、使用条款、限制范围;
- 资产度量: 复用次数、下载次数、反馈、评分、修改次数、按月、季度、年统计。

4.2 行业资产智库示例

在数据中台的资产智库中，以 AI 为例，在人工智能算力、算法、数据三要素中，任何企业都可以通过云服务来购买或利用高速和可扩展的算力，同样大家都可以访问开源社区中的模型和算法，并且可以基于数据进行实验和调整，因此企业无论在算力还是算法上都已经无法形成竞争优势。只有在内外部数据结合基础上通过机器学习、深度学习、复杂网络分析等技术开发出的适合企业自身的、经过生产验证的模型资产，在经过标准化封装并且上升到智库层面，才能真正形成有别于竞争对手的数据中台独特优势。

实践出真知。例如，IBM 在金融行业的长期积淀，形成了一批可在金融行业快速复用，并能立即产生业务价值的资产智库（见图 9）。这些资产智库由 IBM 金融行业的业务专家，在专业知识的基础上对数据进行深度解读，并由数据科学家团队基于大数据、人工智能、复杂网络分析等新技术构建而成，包含基于可服务于不同场景的行业数据资产目录、企业深度关系图谱、行业产品知识图谱、行业资产应用模型等。

图 9
金融行业资产智库服务场景示例



• 行业数据资产目录

行业数据资产目录主要指行业数据资产图谱。行业数据资产图谱是基于银行、保险企业的数据现状形成的数据资产分类目录。图谱中的数据资产项分类标准明晰，金融机构仅需要利用该图谱进行快速对标即可对企业数据资产进行查漏补缺。同时，IBM 提供了与数据资产图谱匹配的人工智能模型，企业可以将其数据资产（包括结构化数据资产和非结构化数据资产）通过该模型进行自动识别，即可分类到数据资产图谱的相应目录中去，大大节省了企业对数据资产进行盘点整理的工作难度和工作量。

例如，IBM 帮助某证券企业利用行业数据资产图谱进行数据治理，使得数据资产的整理工作效率提升了 500% 以上，在降低了工作量的同时保证了数据质量和数据资产覆盖率，实现了增效增质的双重目标。

• 企业深度关系图谱

企业深度关系图谱不同于仅基于外部数据即可看到的企业股权关系，董监高关系等通用图谱，更多地融合了行业知识的深度关系。例如基于隐性关系算法发现的潜在关联企业，基于深度关系挖掘的疑似企业实控人等，为企业进行风险探查提供了深层次的分析工具和手段，助力智能风险防控。

例如，IBM 帮助某金融集团利用外部数据和企业内部数据搭建的企业深度关系图谱，实现了对企业，法人等实体的深度关系挖掘，尤其对隐性关系和复杂关系的识别帮助企业信贷风险等领域发现了以往不能发现的风险，有效提升了其风控能力。

• 行业产品知识图谱

针对银行、保险公司的产品条件关系形成的行业产品知识图谱，企业可以利用这些图谱，结合 IBM 的自然语言处理工具，将自身的产品信息快速导入到图谱中，使其服务于产品智能查询、智能问答等多个场景。

以保险产品为例，该图谱已经将保险产品中常见的关系，属性，条件等内置到图谱中，企业仅需要将自有的数据导入即可，避免了搭建知识图谱过程中对图谱的重复设计，支持了产品知识的快速上线，服务业务场景。

例如，IBM 利用行业产品知识图谱帮助多家金融行业企业实现了多个复杂的业务场景应用，如基于自然语言识别的复杂条件产品筛选，基于问题查询的金融资产报表查询等，使得以往需要大量定制化查询的工作变成了通用的后端组件，支撑多个业务场景。

• 行业资产应用模型

IBM 在多家金融行业企业结合企业自身的独特内部数据（包含多种非结构化的异构数据源），通过迁移学习等方式利用外部数据和内部数据，提炼出包括精准营销（重点产品响应预测模型、事件式场景式营销、智能推荐、产品交叉销售、智能投顾、财富顾问、潜客挖掘、客户唤醒与提升等）和智能风控（电子渠道反欺诈、伪卡交易侦测与防控、借记卡 / 信用卡反欺诈、客户违约预测、个人消费信贷预期、信贷风险传导预警等）等多个智库资产，这些智库模型已经在客户实际项目中获得良好效果。

行业资产应用，助力效率提升

某国有四大行之一，为了提升国际汇款的报文解析效率，与 IBM 合作开展了 SWIFT 报文智能解析项目。

按照银行国际汇款流程，汇出行与汇入行之间通过报文沟通汇款信息。当汇款流程报错时，汇出行会给汇入行发一个 SWIFT 查询报文，以描述报错的现象。以往的报文排查费时费力，需要耗费大量人力排查原因，排查流程长，严重影响客户满意度。

IBM 帮助客户设计了整体的智能分析架构，对报文体系进行了业务分拆和梳理，形成智库资产，并引入 AI 自然语言处理技术，对 SWIFT 报文开展自动识别分类。例如，通过 AI 开展报文预读，将无头寸、尽职调查、欺诈等需优先处理的报文筛选出来，置顶优先处理，并对关键字进行高亮展示，便于业务人员及时处理。通过开展单证类报文自动退电、安慰电智能归档、一查自动转往报等，减少经办处理环节，提高了业务处理的整体效率。

该项目优化了报文处理的业务流程，有效提升了报文解析的效率。与原流程中具备 10 年工作经验的业务人员相比，每份报文的标准作业效率提升了 2.5 ~ 3 倍，预计每月能有效减少 2600 分钟的工单处理时间。同时，将原本的双人检测变成了单人检测，节省了 50% 的员工人数，大大提升了交易报文处理效率，增强了银行核心竞争力。

5. 夯实基础， 按需应变

平台化运营适者生存

金融行业的许多敏态业务，例如互联网贷款和促销等创新应用，需要根据市场动态调整，对中台数据数据请求多而庞杂。过去针对这类敏态需求，临时开发上线的应用导向的数据供应模式往往缺乏全局性考虑，尤其在数据加工等多个环节，造成系统重复建设和效率低下，即便后期进行系统重构，代价也十分高昂。

5.1 平台化模式

相对于以往应用导向的竖井式结构，平台化是一种系统组件功能整合、资源共享和多方互动的处理模式。

数据中台的平台化运营是大运营的概念，既包括数据治理，也包括数据资产化和资产服务化运行所必需的支撑。它将数据访问、数据存储、数据搜索等公共职能平台化，将全行信息进行有效整合和组织分类，把管理、决策、分析等作为一种数据服务对外提供，并围绕平台与流程开展营销活动，根据治理策略控制信息库存，保证信息供给及正确流向，从而响应数据服务的千变万化（见图 10）。下面我们将分析平台化模式中的安全合规、数据操作（DataOps）、模型纠偏、参考实现四个部分。

图 10

数据中台的平台化运营



5.2 安全合规

数据中台涉及的监管合规内容广泛，以安全隐私保护为例，中国人民银行 2020 年 2 月发布的《个人信息金融信息保护技术规范》，规定了个人金融信息在收集、传输、存储、使用、删除、销毁等生命周期各环节的规范性安全防护要求。

数据中台应严格遵守央行规定，其安全框架应涵盖开展评估、设计规划、强化建设、落实运营、监控报告五个环节；范围应涵盖外部数据管理，包括外部数据识别、分类、引入、应用等环节，要提供完整的数据保护闭环流程，保障企业数据安全；通过诊断、定义、设计、开发、交付、验收步骤，全面落实个人隐私保护要求，自动发现、分类、监控、分析中台数据，明确数据风险管理职责，进一步加强建立和落实数据风险监测和管理。

5.3 DataOps

DataOps（数据操作）是人员、流程和技术的编排和一种协作式数据管理实践，旨在改善组织中数据管理者与数据使用者之间数据流的通信、集成和自动化，用于将可靠的、业务就绪的数据快速交付给数据使用者操作、应用程序和人工智能（AI），并使用元数据来提高动态环境中数据的可用性和价值。

DataOps 使敏捷的数据协作能够在整个数据生命周期内提高操作和分析的速度和规模，它与所支持 AI 的自动化、内在治理、数据保护和强大的知识目录一起工作，以在整个企业中运营连续的高质量可信数据。

数据中台作为企业统一的现代化数据收集、组织和分析中心，其数据操作在运作机制上与现代化工厂的流水线有类似之处。它可以对数据进行不落地的加工处理，也允许将需要保留的数据保存到指定的存储（数据库、数据仓库、Hadoop 等）。DataOps 以企业数据资产目录为中心，提供数据虚拟化、数据湖和仓库集成、数据共享、数据管控、数据科学 / 机器学习和嵌入式仪表板等服务（见图 11）。

DataOps 的一个重要能力是数据虚拟化。企业面临异构数据环境，在数据虚拟化模式中，数据不再被复制，并且仅存在于源中。数据虚拟化将所有数据源连接到一个单一的虚拟数据集中来自自动创建虚拟资产，通过结构化查询语言对虚拟表进行统一查询。通过使用每个数据源的处理能力并访问每个数据源已物理存储的数据，可以避免因移动和复制数据而引起的延迟。所有存储库数据都可实时访问，无需提取，转换和加载以及重复的数据存储。

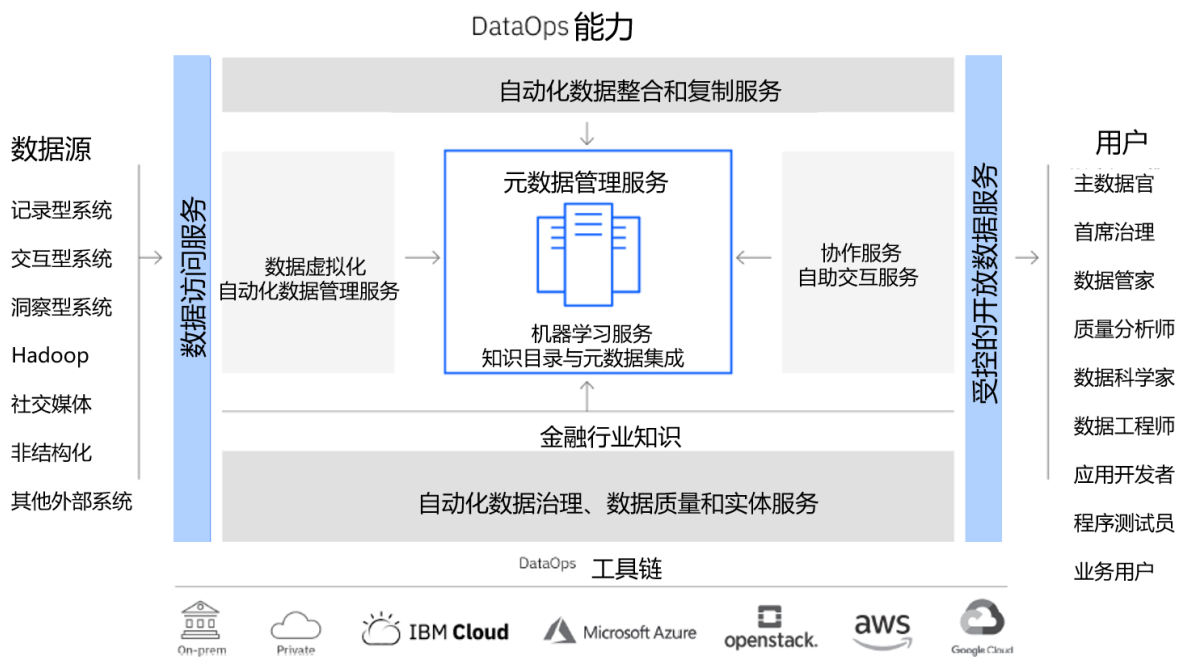
5.4 AI 模型纠偏

金融机构每天借助 AI 决策，但模型可信吗？如果模型有偏差或不公正，那结果又将如何？AI 模型的问题在于，当在生产中遇到新数据时，它们会发生变化，并且其准确性可能会下降，称之为 AI 模型漂移。如何保持 AI 模型的准确性并解决 AI 常见的黑匣子问题，帮助 AI 模型结果公平、可解释和合规？

数据中台的运营 AI 需要具备模型纠偏能力，以不断监测 AI 程序中的偏见决策。通过纠偏“de-biasing”技术，努力减轻这种影响，并对 AI 算法给出的建议提供解释。模型纠偏跟踪和度量来自 AI 模型的结果。当超过自定义漂移阈值时生成漂移警报，从而确保 AI 模型无偏差，同时便于业务用户轻松理解和解释，并可在业务事务中审计。

例如，在批准超过特定风险特征的借款人之前，房屋贷款模型偏差可以被及时发现和矫正，可以对信贷审批模型进行解释，包括账户余额、账户状态等否定因子的量化分析以及申请者年龄、国家等批准因子的量化分析，从而可监控风险模型的绩效、偏差并对信贷审批的结果有一个更清晰的解释并为客户创造更公平和透明的结果。

图 11
数据中台 DataOps 能力



5.5 平台参考实现

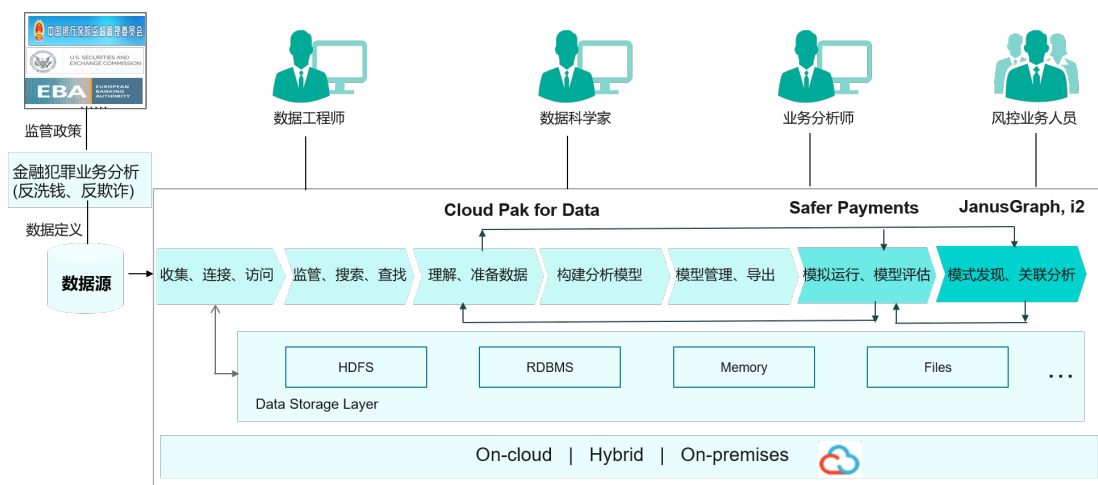
数据中台的技术支撑可以有不同的实现途径，以 IBM Cloud Pak for Data 软件为例，它是专为数据科学、数据工程和应用程序构建而设计。它具有高度集成的数据和分析服务以及简单的协作式任务驱动体验，可以进行包括模型分析、构建和部署。它内置机器学习模型进行自动数据分类和术语分配，其中的虚拟化引擎在 100TB 数据集查询场景下，与联盟方式相比性能最高快 430%。同时支持多个实时系统聚合和复杂查询。

在数据中台的 AI 模型解释、纠偏方面，开源技术能力还比较有限，IBM Watson OpenScale 针对业务用户进行了优化，以在运行时评估其可解释性和模型性能，并帮助实现 AI 模型的可信和透明。

在 API 平台实现方面，IBM Cloud Pak for Integration 能够帮助支持所有集成和数字化转型计划所需的速度、灵活性、安全性和规模，并且预先集成了一组功能，包括 API 生命周期、应用和数据集成、消息传递和事件、高速传输和集成安全性，其中的 IBM API Connect 拥有强大的工具和直观的工作流，覆盖了从创建到管理的整个 API 生命周期，支持自动创建 API、轻松发现资产、为开发人员提供自助服务访问，以及内置安全性和监管。

图 12

基于数据中台的监管科技沙箱示例



5.6 数据中台典型应用

我们以金融风控为例说明，看一下如何基于数据中台构建监管科技沙箱，从而真正实现数据驱动业务的典型应用。

在反洗钱领域，银保监会《银行业金融机构反洗钱和反恐怖融资管理办法》规定，应当全面识别、评估、处理自身面临的洗钱和恐怖融资风险。当前商业银行每日反洗交易筛查总量为几百万到上千万笔，主要在夜间进行可疑交易过滤，大约有成千上万笔反洗钱线索需要几十到几百位业务人员检验和核查，误报率较高，人工验证工作量大，并且无法及时、准确了解当前反洗钱模型的准确率和命中率，反洗钱模型开发、测试和部署调整到上线通常需要花费 4-6 周以上时间，而且缺乏有效的反馈评估机制。

IBM 监管科技沙箱基于数据中台实现数据操作、模型开发，基于认知技术实现仿真模拟、模型评估，基于图计算实现可视化关联分析，具体来讲，通过 IBM Cloud Pak for Data 实现数据访问、连接搜索、治理管控、加工处理、模型构建，基于 IBM Safer Payments 对模型命中率、误报率给出科学反馈分析，自动生成业务人员易懂的业务规则，同时允许参数调整和试错运行，通过开源 JanusGraph 和 IBM i2 进行社交网络分析、对资金流向等深度探索和关联发现。沙箱途径将有助于将金融风控的数据准备、模型开发、部署时间从数周减少到数天并可获得即时的模型评估反馈和调整，帮助提升模型质量和快速找到有价值线索，同时使金融创新业务在可控的试验环境中得以低成本地快速安全实现（见图 12）。

知行合一，思考与行动

数据中台行动指南

为了帮助金融机构更好地建设数据中台，结合 IBM 的实践经验，我们提出三条行动指南，供金融机构参考。

1) 践行数据驱动业务

将数据中台建设做为一个系统化工程，包括数据的连接、处理、应用的全生命周期管理，促进内外数据发挥最大的业务价值，将数据转换为生产力。

- 选择可信赖的战略合作伙伴，科学规划数据中台总体蓝图，借鉴行业最佳实践，制定适合自身发展的数据中台发展路线；
- 盘点数据资产，建立全行统一的数据标准，贯彻执行数据治理的规章、制度、流程，严格数据质量问责，实现跨不同业务线及与 IT 的高效数据协作；
- 完善数据资产价值量化评估，建立数据生态和上下游数据价值链，基于行业标准封装、开放数据服务接口，加速数据资产变现过程并促进业务创新。

2) 平台化数据服务

整合数据治理、数据资产化、资产服务化功能，通过数据工厂流水线方式解决数据不全、不通、不准的难题，实现智能化数据管理与共享。

- 选择统一平台来支持数据字典、数据资产目录、元数据管理、AI 模型训练、测试、部署、运行，包括微服务多云架构和团队协作；
- 满足监管合规，严格遵循个人金融信息保护等监管规定，全面实现数据分类、数据安全、数据隐私保护、数据脱敏、数据审计，外部数据采集和应用管理；
- 实现数据虚拟化，支持跨异构数据源等信息系统的统一连接和访问，改善用户数据体验，而无需了解数据层的复杂性或位置存储，并支持未来扩展。

3) 提高数据服务品质

基于行业开放标准构建数据中台系统和应用，关注模型质量，在开放服务过程中，注重用户体验和第三方开发者应用反馈，通过资产沉淀不断提高数据服务品质。

- 开展数据模型纠偏，及时发现数据质量缺陷，不断监测 AI 模型的偏见决策，通过纠偏技术减轻影响，对 AI 算法结果提供解释，并帮助实现模型的可信和透明；
- 采用敏捷设计思维，使业务人员和科技人员共同参与创新讨论，提高数据工艺品质，归纳、总结和标准化、系列化企业级数据资产智库，加强优质数据资产复用；
- 开放 OpenAPI，在资产服务化过程中，参照 BIAN 的服务蓝图、业务场景、API 标准以及标准规格 Swagger 接口进行数据服务开放和变现，并实现差异化运营。

结语

Brett King 在《银行 4.0》谈到，银行正在被重新定义。要应对变化，银行“要么适应，要么灭亡”。银行制胜未来的法宝已不再是一个产品，而是一种数字化能力。

世界上的数据只有 20% 能被公开搜索到，其余的 80% 都掌握在企业内部。金融机构坐拥数据“金矿”，具备得天独厚的优势。然而，随着数据的“半衰期”持续缩短，如果不能及时利用已经生成和即将生成的数据，那么数据的价值将不断衰减直至消失。

数据中台以数据为生产要素，建立了新型的生产力和生产关系。实现统一平台化之上的治理标准化、数据资产化、资产服务化、数据业务化是数据驱动业务的成功之道。得数据者得天下，数据之巅，致胜未来！

—

需要思考的重要问题

- 您在数据领域遇到哪些问题？
- 您打算如何构建数据中台？
- 您打算从哪里起步？

备注和参考资料

- 1 张珂. “不可思议的数字: 互联网每天到底能产生多少数据?”. 资本实验室. 20190415. http://www.coinsay.com/article/coinsay_9983.html
- 2 “Companies Collect a Lot of Data, But How Much Do They Actually Use?”. Priceonomics Data Studio. 20190807. <https://priceonomics.com/companies-collect-a-lot-of-data-but-how-much-do/>
- 3 “Big data executive survey 2018”. NVP. <http://newvantage.com/wp-content/uploads/2018/02/Big-Data-Executive-Survey-2018-Findings.pdf>

选对合作伙伴，驾驭多变的世界

在 IBM，我们积极与客户协作，运用业务洞察和先进的研究方法与技术，帮助他们在瞬息万变的商业环境中保持独特的竞争优势。

IBM 商业价值研究院

IBM 商业价值研究院 (IBV) 隶属于 IBM Services，致力于为全球高级商业主管就公共和私营领域的关键问题提供基于事实的战略洞察。

了解更多信息

欲获取 IBM 研究报告的完整目录，或者订阅我们的每月新闻稿，请访问：ibm.com/iibv

访问 IBM 商业价值研究院中国网站，免费下载研究报告：

<https://www.ibm.com/ibv/cn>

© Copyright IBM Corporation 2020 IBM Corporation
New Orchard Road
Armonk, NY 10504
美国出品
2020 年 4 月

IBM、IBM 徽标、ibm.com 和 Watson 是 International Business Machines Corp. 在世界各地司法辖区的注册商标。其他产品和服务名称可能是 IBM 或其他公司的商标。以下 Web 站点上的“Copyright and trademark information”部分中包含了 IBM 商标的最新列表：ibm.com/legal/copytrade.shtml。

本文档为自最初公布日期起的最新版本，IBM 可随时对其进行修改。IBM 并不一定在开展业务的所有国家或地区提供所有产品或服务。

本文档内的信息“按现状”提供，不附有任何种类（无论是明示还是默示）的保证，包括不附有关于适销性、适用于某种特定用途的任何保证以及非侵权的任何保证或条件。IBM 产品根据其提供时所依据协议条款和条件获得保证。

本报告的目的仅为提供通用指南。它并不旨在代替详尽的研究或专业判断依据。由于使用本出版物对任何组织或个人所造成的损失，IBM 概不负责。

本报告中使用的数据可能源自第三方，IBM 并不独立核实、验证或审计此类数据。此类数据使用的结果均为“按现状”提供，IBM 不作出任何明示或默示的声明或保证。

国际商业机器中国有限公司
北京市朝阳区北四环中路 27 号
盘古大观写字楼 25 层
邮编：100101

