

The background features several overlapping, semi-transparent blue geometric shapes that create a sense of depth and movement. These shapes are set against a white background. Faint, vertical lines of varying lengths, resembling a digital data stream or barcode, are scattered across the blue areas, adding a technological and data-oriented aesthetic to the design.

# 全域数据治理 白皮书

# 目 录

<b>第 1 章 全域数据治理的必要性</b> .....	<b>7</b>
1.1 全域数据治理的市场动因与主要挑战 .....	7
1.1.1 数据资产安全艰巨挑战 .....	7
1.1.2 数据治理能力困局 .....	9
1.1.3 数据治理 ROI 提升 .....	10
1.2 什么是全域数据治理 .....	12
1.3 全域数据治理规划 .....	13
1.3.1 结构化数据治理规划 .....	14
1.3.2 非结构化数据治理规划 .....	15
1.3.3 机器数据治理 .....	15
<b>第 2 章 结构化数据治理</b> .....	<b>17</b>
2.1 结构化数据治理概况 .....	17
2.2 结构化数据治理的挑战 .....	17
2.2.1 多业务系统多数据源整合挑战 .....	18
2.2.2 业务迭代发展带来的持续治理挑战 .....	18
2.2.3 主流数据治理的问题总结：被动式的数据治理 .....	18
2.3 以业务为导向的即时数据治理服务 .....	19
2.3.1 数据架构与数据治理相互促进 .....	21
2.3.2 Data Fabric 架构 .....	21
2.3.3 基于 Data Fabric 架构的业务建模与治理起点 .....	23

2.3.4 以业务为导向的即时数据治理流程 .....	27
2.3.4.1 业务梳理与数据治理 .....	28
2.3.4.2 新业务开发与即时数据治理 .....	29
2.3.5 结构化数据的资产与安全合规评估 .....	29
2.4 业务导向的数据治理优势和价值 .....	30
2.4.1 结构化数据治理展望 .....	30
<b>第3章 非结构化数据治理 .....</b>	<b>32</b>
3.1 非结构化数据的特征与应用场景 .....	32
3.2 非结构化数据的安全挑战 .....	33
3.3 以中台方案为基础的即时数据治理 .....	34
3.4 非结构化数据的资产分类治理 .....	34
3.4.1 非结构化数据资产治理的业务挑战 .....	35
3.4.2 如何对非结构化数据进行分类治理 .....	36
3.4.2.1 非结构化数据的特征分析模型 .....	36
3.4.2.2 非结构化数据的分类方法与过程 .....	36
3.4.3 数据资产分类示例 .....	37
3.5 非结构化数据的安全合规治理 .....	38
3.5.1 生物医药行业的安全合规治理 .....	38
3.5.2 汽车制造行业的安全合规治理 .....	38
3.5.3 非结构化数据安全合规的治理要点 .....	39
3.5.3.1 安全合规治理的框架 .....	39
3.5.3.2 安全合规治理的组织与制度保障 .....	40

3.5.4 安全合规治理带来的价值评估 .....	40
3.6 非结构化数据的知识治理 .....	41
3.6.1 知识管理建设面临的诸多挑战 .....	41
3.6.2 组织如何规划知识管理 .....	41
3.6.2.1 知识管理的目标是改善组织业务生产力 .....	42
3.6.3 如何围绕知识战略开展知识治理 .....	43
3.6.3.1 知识治理的主要内容 .....	43
3.6.3.2 领域知识网络提升知识治理效率 .....	44
3.6.4 通过知识治理提高知识运营效率 .....	44
3.7 非结构化数据的资产保留治理 .....	45
3.7.1 非结构化数据资产保留的业务场景 .....	45
3.7.2 非结构化数据资产保留的治理 .....	45
3.8 非结构化数据治理为组织生产力赋能 .....	46
<b>第4章 机器数据治理 .....</b>	<b>47</b>
4.1 什么是机器数据 .....	47
4.2 机器数据治理概述 .....	47
4.2.1 以中台方案为基础的即时数据处理 .....	47
4.2.2 机器数据治理的价值 .....	48
4.3 加强合规与审计 .....	48
4.3.1 合规与审计面临的挑战 .....	48
4.3.2 通过机器数据治理解决合规与审计问题 .....	49
4.3.3 机器数据治理对合规与审计的价值 .....	50

4.4 提升运维效能 .....	50
4.4.1 运维管理面临的挑战.....	50
4.4.2 通过机器数据治理提升运维效能.....	51
4.4.3 机器数据治理对运维管理的价值.....	52
4.5 助力业务运营 .....	52
4.5.1 业务运营面临的挑战.....	52
4.5.2 通过机器数据治理助力业务运营.....	52
4.5.3 机器数据治理对业务运营的价值.....	53
4.6 机器数据治理赋能组织价值.....	54
<b>第 5 章 全域数据治理加速数据驱动 .....</b>	<b>56</b>
5.1 从数据治理转变为业务洞察力.....	56
5.2 为 CXO 提供数据治理的支持 .....	56
5.3 数字生态建设 .....	57
5.4 数据驱动文化打造 .....	58
<b>第 6 章 全域数据治理的十大关键.....</b>	<b>61</b>

## 文档信息

材料名称：《全域数据治理白皮书》

版本控制：V1.0 版 2022-06-18

## 关于本书

《全域数据治理白皮书》分为 6 个章节组成，将探讨：

- » 第 1 章 全域数据治理的必要性
- » 第 2 章 结构化数据治理
- » 第 3 章 非结构化数据治理
- » 第 4 章 机器数据治理
- » 第 5 章 全域数据治理加速数据驱动
- » 第 6 章 全域数据治理的十大关键

首先，假如您是组织中负责数据管理和治理的人。也许您是首席信息官（CIO）、首席合规官（CCO）、首席数据官（CDO）、数据保护官（DPO）或首席信息安全官（CISO）；又或者，您可能是一组织的一员，负责客户或用户服务，亦或是 IT 经理或架构师。

如果您发现自己读完这本书后仍在思考，“我从哪里可以学到更多？”请访问 <https://www.aishu.cn/>

# 第1章 全域数据治理的必要性

## 1.1 全域数据治理的市场动因与主要挑战

数字化转型正在颠覆全球各地的业务。大多数组织已建立自身的数字化战略，对准业务战略目标并结合最新的先进技术进行数字化转型。数据已成为组织接触客户、开展营销活动、留住客户等业务活动中最重要的战略资产。组织期望利用高质量的数据，一方面驱动优化运营效能，另一方面驱动业务的创新。数据战略已成为打造数据驱动型组织重要因素。

### 1.1.1 数据资产安全艰巨挑战

随着数据量指数级增长，严格的监管环境，网络威胁（包括勒索病毒攻击）的频繁发生，以及兼具竞争力和创新性的黑马不断颠覆市场，保护组织最具战略意义的数据资产安全并从中提取价值是业务发展的当务之急，也是一项十分艰巨的挑战。

如果说组织需要某种激励来严格管理数据，那就是法规遵从。因未能遵守严格监管要求而产生的成本势必会迅速增加，包括罚款、外部诉讼与和解费用等法律成本，以及声誉损害和业务损失。

国际上以欧盟（EU）通用数据保护条例（GDPR）为例，该条例要求处理欧盟公民数据的公司遵守严格的数据隐私法规。不遵守 GDPR 的组织将面临高达其全球年度营业额 4% 的处罚。国内《中华人民共和国个人信息保护法》、《中华人民共和国数据安全法》等法律对数据保护也做出了相关规定。由于潜在的惩罚如此严重，许多组织都在大力投资合规落地计划，有些组织甚至将安全合规能力视为一项关键的竞争优势。

然而，许多因素使得组织越来越难以遵守多个垂直行业中日益增多的法规要求。比如，全球性银行在任何时候都要遵守 100 多种不同的监管规定，其中包括证券交易委员会（SEC）法规、多德-弗兰克法案、欧盟 GDPR 和金融工具市场指令（MiFID）等。CIO 和 IT 主管正准备迎接另一项新的银行业法规：修订后的支付服务指令（PSD2），该指令允许消费者和组织使用第三方提供商（如 Google 或 Facebook）来管理财务、支付账单或分析消费。这对欧盟银行业的影响是非常显著的——它们将被要求允许第三方提供商通过应用接口（API）访问客户账户。因此，由于新的安全要求（如更强的身份检查和 API 开发），IT 成本可能会随之增加。

数据生产者和消费者越来越多，跨越不同系统的数据也越来越多，这让问题变得更加复杂。在 2017 年的一项研究《Data Age 2025: The Evolution of Data to Life-Critical》中，IDC 预测，到 2025

年，全球数据将增长到 163ZB（万亿 GB 字节）。而在“数字宇宙”研究中，IDC 估计，全球只有 0.5% 的数据实现了分析，不到 20% 的数据得到了适当保护。

立法和新法规颁布的速度也超过了现有 IT 基础设施投资的能力和新应对方案所需的预算。例如，为满足法规必须存储敏感数据的时间长度可能会超过现有架构的物理能力，因此，许多 IT 领导者过度实施数据控制流程，这不仅扼杀了创新力和生产力，还阻碍了适应监管环境变化所需内在灵活性。例如在合规及监管要求较高的金融行业，因合规及监管导致内部数据应用限制较多。

在每一个新法规或修订法规下，应用程序、人员、流程、系统和现有技术之间的差距不断扩大，暴露在风险中的可能性也将不断增加。

组织及其数据也在不断受到威胁。美国司法部统计，美国每天会发生 4000 多起勒索病毒攻击。在全球范围内，这些攻击与上一年相比增加了 60 倍。然而，正如最近的攻击所表明的那样，勒索软件是一个全球性问题。2020 年，全球因勒索软件造成的总损失高达 25 万亿美元，高额的收益让更多的犯罪者趋之若鹜。

当然，不是所有数据安全威胁都是由外部人员恶意造成。IT 管理者还必须维护数据的完整性，防止数据被意外损坏或彻底删除，或被某些不可预见的宕机事件破坏。

对于 IT 部门来说，经常进行数据备份和安装最新的杀毒软件可以最大程度降低来自勒索软件攻击、意外删除、系统损坏或宕机造成的潜在损害，但这还不够。

信息安全和身份管理解决方案可以帮助防止未经授权的访问，但数据安全必须从组织的核心扩展到每个入口。例如，金融机构可能会遵守法规，但一旦该机构允许通过客户的个人设备进行交易行为，就会引入一个新的入口，并增加遭受攻击的风险。

复杂的监管和隐私安全要求组织实现数据全生命周期管理，由于存在数据孤岛组织忽视了对非/半结构化数据的安全及治理。为了应对这些数据治理挑战，组织必须捕获、控制和利用日益增长的数据，带来具有竞争力的洞察，为新产品开发提供支撑，并能更深入地了解客户。然而，许多业务和技术领导者普遍被以下问题困扰：

- » 哪些数据对组织来说有业务价值的？
- » 这些数据是否受到保护？
- » 数据应该保留多长时间？
- » 能否快速准确的定位和获取数据？



- » 谁可以拥有访问权限？
- » 数据背后包含什么价值？
- » 数据是否可信，准备度如何？
- » 如何利用这些数据来产生竞争优势？

### 1.1.2 数据治理能力困局

组织的数据随着数字化转型的几波浪潮下，已深度覆盖组织的不同业务域。如对于政府而言数字化的业务领域主要是政务云、政务办公和智慧应用。对于企业单位而言涉及研发、生产、营销、管理等方方面面。高校主要涉及教育、学工、科研、综合管理等。医疗行业主要为科研、临床、采购、管理领域。

数字化转型逐步进入深水区，全域数据能力成为组织核心竞争力。以全域数据能力驱动组织转型、数字创新以及打造新的商业模式已经成为各地组织的共识，各领域的组织期望拥有全域的数据能力，加速数据的流通，提升数据赋能业务价值、提升安全合规能力等。在构建数据能力行动中，组织优先关注以下几个维度事项。

1. 数据文化，推动组织升级，持续行业领先
2. 全域数据应用，打造全新商业模式
3. 知识共享，基于全域数据弥补认知误区催生数字创新
4. 客观数据应用，助力组织快速使用市场变化，实现卓越运营

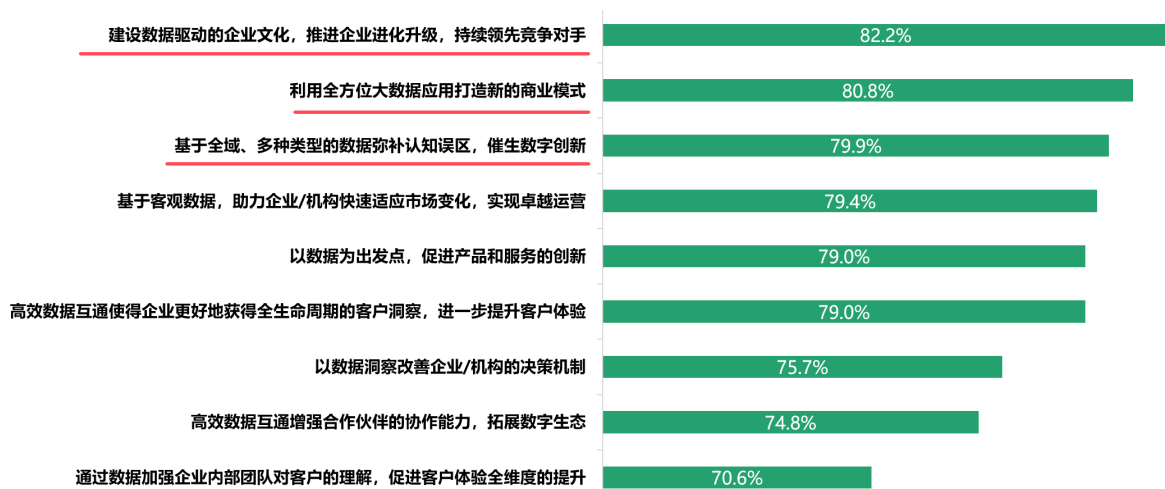


图 1 数据来源：2022 年 6 月 Forrester Consulting 进行的中国数据中台调研

通过国内外多家权威咨询机构行业洞察报告，百余家大型组织 CEO 和决策者访谈，结果显示组织经营活动实践中期望数据驱动提高运营质量、提升组织运营效率、整体精准控制组织运营成本等目标，但在业务实践应用中时常遇到三大困局：缺数据、缺能力、缺乏工具。

**缺乏全域数据：**将无从支持业务价值。特别是研发、生产、营销、客流、交易等有价值的数据都被忽略或者无法采集，也无法通过分析提升运营效率、改善营销方式。

**缺乏数据治理能力：**无法分析业务。数据孤岛阻碍了数据资产的开发。数据的孤岛基本是老大难问题，离开业务来谈数据价值是没有太多价值的，需要基于业务和数据即时治理、数据的关联能力来分析支持业务。

**缺乏工具：**无法优化业务。在数据进行分析之后，许多组织还是没有办法找到那临门一脚的方式，没有办法解决他的实际问题。真正缺乏的并不是数据，而是高质量的数据以及缺乏由数据支持的工具来解决他们的实际问题。缺乏适当的技术、平台及合作伙伴支撑分析利用全域数据。

### 1.1.3 数据治理 ROI 提升

在过去，一个组织提高效能的驱动力主要来自积极降低成本、强劲的稳定收益表现等等，但随着业务快速发展、行业更新迭代，叠加上今年突如其来的疫情影响，数字化转型的高压下，大部分组织开始积极探索新的应对策略。当下组织数字化转型趋势是“数据”引领业务变革，数据治理成为大势所趋，如何做好数据共享和数据驱动、如何发挥数据资产价值最大化，“全域数据治理”是问题的切入点。

数据治理对每个组织都是一项很大的挑战，据 Gartner 的一项调查显示，超过 90% 的数据治理项目都失败或者 ROI 不理想！为什么会有这么多数据治理项目不及预期？下面我们罗列部分导致数据治理项目 ROI 不达标的因素。

#### 一、缺乏明确的目标

缺乏清晰的战略以及对不同类型数据价值的认知影响组织转型。项目蓝图设定的目标缺少某一个业务单元近期落地的目标，多数情况下目标定得很大、很泛、不聚焦，不考虑目标可实现性和可衡量性，例如：目标就是解决组织的所有数据质量问题。

业务单元和数据治理项目团队相关人员对数据质量目标的定义和理解没有达成共识，存在分歧的情况下就开始实施治理。只从技术角度考虑怎么治，不考虑为什么治？为了治理而治理！

#### 二、分工混乱、权责不明

组织内部对谁有数据的拥有权，谁有数据的使用权，谁有数据的管理权等职责没有清晰的定义，人人都自称对数据质量负责，但实际上数据管理十分混乱、大量重复，真正出现问题后相互推诿，没有人愿意负责。

让 IT 人员去关注数据质量的定义和趋势，分析并确定数据质量问题的根本原因，这是误区认知。

### **三、高层管理者关注不足**

高层管理者对数据治理认知程度不高，将数据治理和数据管理混为一谈，认为数据治理就是 IT 部门或者是 DBA 的事，IT 团队就能搞定了，不需要高层领导过多的参与和关注。高层管理者天天高喊“数据是资产、治理很重要，要大力支持”等口号，但没有实质的行动。数据部门不能被完全赋权，或者安排一个毫无影响力的小部门去负责，这都不太可行。遇到跨部门协调，各部门嘴上答应一定好好配合，实际执行中还是我行我素，什么数据标准、数据规则，遇到强势的业务就得给业务让路、开绿灯，导致数据治理策略形同虚设。

### **四、缺乏数据治理人才**

将数据治理和系统管理混为一谈，让 IT 系统管理员对数据的质量负责。这就好比让修自来水管道的对自来水的的水质负责一样不靠谱。认为数据质量管理都是 IT 人员的事，懂算法、懂模型、懂编程就够了。事实上，数据质量团队必须具有业务分析思维、对业务流程足够了解才能做出正确决策，如果不能理解业务也可能无法理解错误数据的影响。

### **五、缺少共创型数字伙伴**

当前多数组织期望成为数据驱动型组织，但在实践中发现缺乏从顶层规划开始、自上而下的中台战略，同时缺少针对不同业务领域打造认知智能模型的能力，这要求组织既具有复杂的技术能力，也需要对行业知识和业务场景有深刻的理解。解决这些问题的最佳实践是需要由甲方业务专家和行业领先技术专家共同组织伙伴团队，通过共创的活动来实现“全域数据治理”最终的目标，但实践中发现，多数组织还需要提升或加强这样的认知。

***为提升全域数据治理 ROI，组织需要优化业务流程，加强技术探索，开展端到端的全生命周期数据管理等更多举措齐头并进。***

1. 提供即时数据处理端到端的监控能力
2. 与前台、后台实现紧密协作，实现全局洞察

3. 建立统一的数据视图及智能数据目录
4. 建构智能数据服务系统，实现对全域数据的处理以及利用
5. 推进数据资产化，通过数据清洗、制定标准规范和管理
6. 将结构化的数据转变为易于理解的标签"
7. 对各种类型数据纳入统一的数据安全管理体系
8. 建立密切贴合业务场景的标签管理体系
9. 端到端全生命周期数据质量管理
10. 通过将数据治理固化到业务流程中，实现即时治理

## 1.2 什么是全域数据治理

组织的数据大多分布在不同业务域、园区、甚至跨地域存储，数据存在的格式也是多样性的。分为结构化数据、非结构化数据、机器数据等。结构化数据指的是数据在一个记录文件里面以固定格式存在的数据。它通常包括 RDD 和表格数据。非结构化数据是数据结构不规则或不完整，没有预定义的数据模型，不方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、文本、图片、XML, HTML、各类报表、图像和音频/视频信息等等。机器数据是指由日志文件和时间序列等组成的大数据，包括日志数据 logs、指标数据 metrics 和调用链数据 traces。对于高速发展的组织而言，数据的多样性和复杂性已成为不争的事实。针对这些存储分散、类型多样、权限复杂的全域数据现状，如何有效管理，使其能够释放价值提升组织生产力和竞争力，已经成为组织数字化转型路上优先解决事项。

全域数据治理，是组织战略管理部门发起并推行的，关于如何制定和实施针对整个组织内部所有业务域数据的商业应用和技术管理的一系列政策和流程。全域数据治理是一套持续改善管理机制，通常包括了数据架构组织、数据模型、政策及体系制定、技术工具、数据标准、数据质量、影响度分析、作业流程、监督及考核等内容。

全域数据治理是以服务组织战略目标为基本原则，通过组织成员的协同努力，流程制度的制定，以及数据资产的梳理、采集、清洗、可视化管理和多维度分析，实现数据资产价值呈现、业务模式创新和经营风险控制的过程。

全域数据治理可以帮助组织更好地管理其组织数据的实用性、可用性、完整性和安全性。借助正

确的技术，全域数据治理还可以带来巨大的业务价值并支持数字化转型。基于组织数字化战略、打造数据驱动型组织。

在最基本的层面上，数据治理即控制数据并确保其安全。成功的数据治理需要清楚数据的位置、来源、访问权限以及包含的内容。有效的数据治理是维持业务合规性的先决条件，无论这种安全合规是自发的，还是行业或政府强制要求的。

合规性通常会增加数据治理的复杂性，要求能够正确搜索数据，了解其中包含的每个文字或数字，并在面对任何数据需求时快速、准确地生成正确的数据源。

对授权人员而言，还可以通过数据的质量、准确性和可用性来判定组织是否满足或违反严格的监管要求。

数据治理超越了信息管理中业务流程的支持，它包含一系列广泛的数据策略和功能，包括：

- » 数据交付和访问：与数据存储、检索和处理相关的任何操作。
- » 数据完整性：确保数据的真实性、准确性和数据质量。
- » 数据血缘关系：管理数据的来源、变更以及轨迹。
- » 数据安全防泄密 (DLP)：确保敏感数据不会发送到公司网络之外，让数据在可控范围之内。
- » 数据安全：防止未经授权的访问或数据损坏。
- » 数据同步：建立数据类型的一致性。
- » 主数据管理 (MDM)：用于定义、治理和管理数据的流程、策略、标准和工具的完整集合。

## 1.3 全域数据治理规划

如今全域数据治理作为各领域组织数字化转型的基础，成为了各地组织数字化变革中的焦点和主战场，数据治理核心内容是阻止数据孤岛蔓延，打破组织中结构化数据、非结构化数据、机器数据孤岛，实现数据的即时共享与利用。以全域数据能力，重塑组织生产力，全面提升数据质量及可访问性。

关于实施战略规划，建议以中台思路展开全域数据治理蓝图设计。在各业务域系统网络中构建多个中台方案，对组织经营活动的所有数据，进行全域数据治理，实现全业务域的数据驱动。全域数据治理框架如下图所示，分别构建结构化、非结构化、机器数据治理蓝图。



数据治理是组织对数据的可用性、完整性和安全性的整体管理，治理的目标是提高业务运营效率，提升组织决策能力和核心竞争力。目前数据治理多数由数据部门负责，以数据为核心开展数据治理工作，强调数据合规和数据落标，以治理了多少数据为 KPI 指标，当业务发生变化时需要重复开展数据治理工作。

在组织保障、体制机制、数据战略同步建设的同时，我们提出以业务为核心开展数据治理工作，从业务规划开始就进行标准前置判定，业务迭代牵引治理规则的改变。从源头开始控制，以助力业务发展为目标，将数据治理工作贯穿业务发展的全过程，实现数据的长治久安。

### 1.3.1 结构化数据治理规划

组织内各业务域输出海量异构结构化数据，应以业务为导向开展即时数据治理服务。充分应用行业领先技术通过 Data Fabric 架构实现数据资产化，以领域知识网络赋能智能应用，重塑组织生产力。结构化数据治理坚持以全域数据战略引领，设计具有高度前瞻性的全局业务框架支撑未来高速发展的业务新模式、新业态。

在全域数据战略的指引下，结构化数据治理框架应包含但不限于：

1. 安全法规遵从（国际国家法规、国内法、行业标准规范、组织内部管控）
2. 增强数据安全目录

### 3. 数据资产图谱

将其作为结构化数据治理指导方针，设计基于新一代的 Data Fabric 架构、业务认知框架为参照标准，并持续更新的组织数据资产图谱，持续通过业务梳理活动，指引数据治理和新业务开发，最终把组织之间、业务域之间、业务域与数据池、跨域数据连接起来，实现组织范围内全部数据高质量即时治理和即时业务服务。

## 1.3.2 非结构化数据治理规划

基于行业最佳实践，大部分组织采用中台方案，通过数据资产化、数据知识化及立体安全策略对海量非结构化数据科学治理使其发挥最大价值，满足业务发展的实际需要。在初始阶段对数据进行全面梳理、全局设计、甲乙双方项目团队基于行业最佳实践指导来开展治理工作，最终实现数据驱动运营和数据驱动创新，整体治理规划需坚持三个原则：

- **以中台方案为基础**：打破数据孤岛、为不同业务场景赋能。
- **即时治理**：数据在生成或导入进入组织内部时即已在治理框架下。
- **AI 驱动**：利用自然语义分析、图像识别、语音识别、领域知识网络等技术全程赋能数据治理。

非结构化数据治理规划分为如下几个维度执行落地：

- 非结构化数据的资产分类
- 非结构化数据的资产保留
- 非结构化数据的安全合规
- 非结构化数据的知识治理

## 1.3.3 机器数据治理

组织内部海量机器数据的治理要以业务需求为导向，以中台架构为基础，以领域认知智能为技术底座，将组织内部各业务域机器数据进行汇聚、整合与治理，开展可观测性能力建设，对机器数据进行分析处理，快速消除组织业务运营与管理中遇到的问题，提升业务运维效能、助力业务运营分析、加强合规与审计，以数据驱动运营，建立数据驱动型组织。

机器数据治理战略规划需按以下几个维度进行，但不限于此。

- 法律法规遵从，机器数据治理优先考虑国家法律、行业法规、组织规范的遵从。
- 业务智能运维提效，以组织战略目标为出发点，按多场景运维需求，实现统一的运维分析与管理。
- 助力业务运营，通过对机器数据的治理，可以充分挖掘数据中的业务价值，保障业务系统稳健运行，提升运营效能，提升客户体验。



## 第2章 结构化数据治理

### 2.1 结构化数据治理概况

结构化数据也称作行数据，是由二维表结构来逻辑表达和实现的数据，严格地遵循数据格式与长度规范，主要通过关系型数据库进行存储和管理。

结构化数据是现有结构再有数据，结构本身意味着业务关系和逻辑，但是每个业务系统和流程都有相关的对象、关系和逻辑，所以业务系统和流程复杂的情况下，结构化数据的管理难度就很高，数据之间往往是错综复杂的蜘蛛网。所以，结构化数据管理的分类也比较多，包括基础数据、主数据、事务数据、报告数据、观测数据、规则数据。

结构化数据的共同特点是以业务架构为基础，建立统一的数据资产目录、数据标准与模型。结构化数据治理主要涉及数据提供方（业务部门）、数据使用方（业务部门）、数据管理方（IT 部门/数据部门）、数据运营方（数据治理等服务提供方）。目前数据治理工作的开展主要以运营方为政府和企业客户基于数据治理产品或工具提供数据治理服务的方式进行。

当前主流数据治理平台主要包括元数据管理、数据标准管理、数据质量管理、数据集成管理、数据资产管理、数据安全治理、数据生命周期管理等模块。

- 元数据管理主要包括元数据采集、血缘分析、影响分析等功能；
- 数据标准管理包括标准定义、标准查询、标准发布等功能；
- 数据质量管理包括质量规则定义、质量检查、质量报告等功能；
- 数据集成管理包括数据处理、数据加工、数据汇集等功能；
- 数据资产管理：包括数据资产编目、数据资产服务、数据资产审批等功能；
- 数据安全治理包括数据权限管理、数据脱敏、数据加密等功能；
- 数据生命周期管理包括数据归档、数据销毁等功能；

### 2.2 结构化数据治理的挑战

结构化数据治理基本上涵盖了企业所有业务有关的数据，因此在数据治理过程中，应该围绕着企业数字化的战略目标进行分解，以核心业务、核心流程、核心系统进行数据的治理，形成以业务为中

心的数据资产体系。

从企业的数据战略来看，数据治理首先是管理问题，其次才是技术问题。只有保证业务战略清晰、核心业务流程清晰，业务系统建设运行良好，才能形成数据的可用性、一致性、完整性、合规性、和安全性，确保在整个数据生命周期中，都具有较高的数据质量，才能进一步通过数据运营、数据应用帮助企业实现数据资产管理、发现内部数据问题、发掘数据价值，进而实现企业数据资产的盘活和有效利用。

### 2.2.1 多业务系统多数据源整合挑战

对于中大企业或者政府，在数据治理实践过程中，由于历史的原因，之前各部门独自建设系统，形成数据孤岛、烟囱林立。近年来，设立大数据或政务服务数据管理部门后以数据大集中建设为主，但对多部门数据按照统一标准进行深度融合治理以及长期有效的贯彻执行难度大、周期长、效果不明显，数据质量仍有待进一步提升。

### 2.2.2 业务迭代发展带来的持续治理挑战

在不同的行业发展中，以互联网行业为例，业务迭代非常迅速，通常存在首发版本的数据质量最优，1.1 版本不行，2.0 版本完全不可用的说法，说明第一次做数据治理时，极重视数据质量，会有完善的流程来保证埋点的准确性，本身也没有太多的包袱；而在后续的产品迭代中，如果流程和标准的迭代相对滞后，整个数据治理的结果也会随着受影响，最终导致整个数据质量低劣，直至完全不可用，持续的业务迭代必须持续使用流程和标准的更新为基础。

### 2.2.3 主流数据治理的问题总结：被动式的数据治理

当前主流数据治理面临的主要问题是数据治理形成的数据资产质量不高、数据与业务脱节。

通过调查发现，当前主流数据治理从根本上来说都是先有业务，再有治理，也就是治理是对现有业务结构的一个反映过程，可以称之为被动式数据治理。它的典型特征就是业务开发阶段业务部门主导，数据治理阶段 IT 部门主导（或者数据部门主导）。当然，这基本上符合当前主流的业务开发和数据治理的现状。

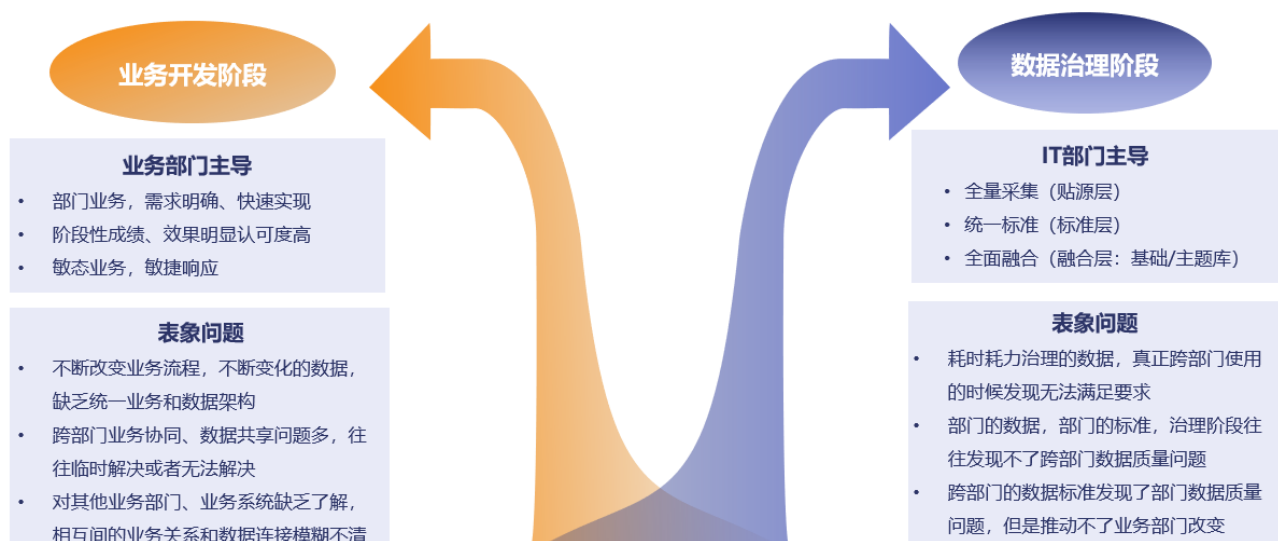


图 3 业务部门主导 VS IT 部门主导

产生这些问题的主要原因是：

- 1、现有业务是各业务部门主导的、局部性的设计，往往从源头开始就弱化或者忽视了部门业务与数据战略的连接，所以，业务部门和业务部门之间的对同一对象、同一业务术语的表达往往是不同甚至是背道而驰的，通过事后的被动式数据治理，通常很难解决这个问题。
- 2、很多组织缺乏清晰的、匹配业务战略的数据战略，各部门的业务流程、数据结构往往缺乏全局性的指引和标准约束。
- 3、在当前主流的数据治理平台中，最大的问题也是缺乏了对业务架构、业务对象、业务关系的系统性支持，从元数据管理角度，往往只见树木不见森林。

## 2.3 以业务为导向的即时数据治理服务

离开业务谈数据是没有意义的，数据治理沉淀的是业务的数据资产，最终目的是服务于数字化的运营和创新，这是数据驱动型组织的基本要义。

结构化的数据虽然随业务产生的、有结构的数据，但是实际上，如果实践中，各业务部门各自为政、水平参差不齐，仅仅在事后治理，数据与数据、数据与业务对象往往是割裂的，从单个业务看，数据都是结构化的，从整个业务的关联性看，都是孤岛。所以，大量数据往往不是有价值的业务资产，而是不断下沉的“债务”。

所以，根本点在于要建立起以业务为导向、系统性的数据治理体系，应该有数据战略统一指引，基于数据架构，在业务侧即时的治理、在应用中持续治理的数据治理体系。

整个数据治理应该是数据战略引领的，是反映全局业务框架的：

- 1、构建起一个组织的全局性业务认知模型，建立起战略级别的业务模型，从而实施业务层面的融合。
- 2、基于业务整体认知框架，指导业务协同，分布优化业务。持续更新数据资产知识网络；



图 4 数据战略引领的数据治理

与之前的业务开发和数据治理相互独立或者割裂相比，新的数据治理是即时性和协作性的模式，对于一个新的业务或者新的数据分析：

- 1、业务部门仍然主导业务，但是业务部门需要围绕核心业务流程、业务表、核心业务字段与架构团队对齐，在业务侧进行即时数据治理；
- 2、IT 部门仍然主导数据治理，但是需要在业务开发过程中即配合进行数据治理，同时提供数据服务；

新增架构团队主导业务与数据协同，这个团队是虚拟的，可以是数据治理团队中的架构师团队，负责核心业务流程、规则、核心业务对象、业务表和业务字段的约束和指导。

### 2.3.1 数据架构与数据治理相互促进

通常，主流的结构化数据治理通过大数据平台或者数据中台作为技术支撑，但是，实际上，数据治理的范围是超过数据中台的，需要有一个数据架构来把业务和业务、业务和数据、数据和数据连接起来，它是全范围的，无论是业务系统侧、结构化数据中台还是新的业务与分析应用，都会涉及到数据的即时治理和即时服务。

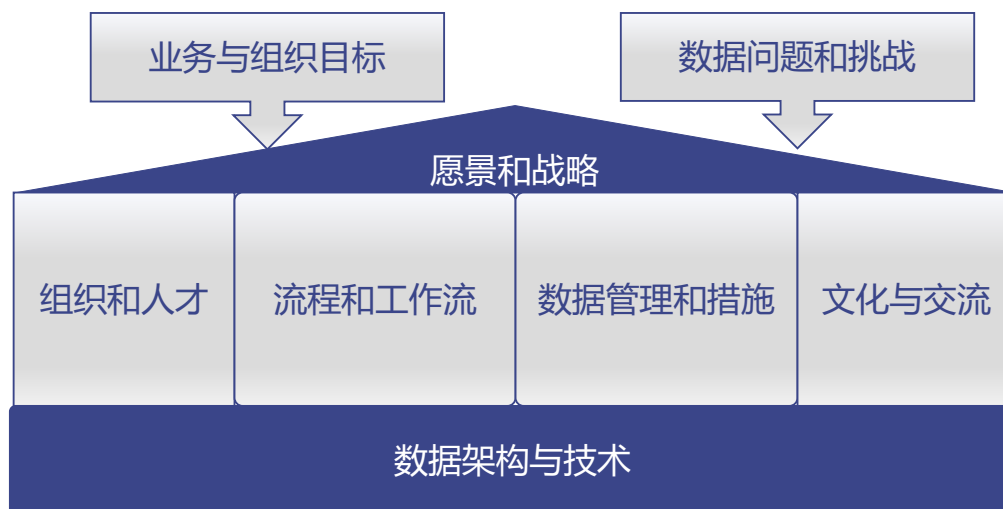


图 5 数据架构与技术

通过上图，我们看到，数据架构与数据治理是相辅相成的，协同作用非常重要。数据架构是数据治理的指引，数据治理是数据架构落地的服务化路径。数据治理不仅仅是技术，更是服务，它涉及的范围更广，涵盖了组织、人员和流程，而且——非常重要的一——创造了一种数据驱动的文化。

实施数据架构的主要目的是标准化方法和协议，以及用于在整个组织中获取、存储、管理和共享数据的系统，以改进决策制定。在现代企业中，大多数决策都是实时做出的，为了促进高效且实时友好的数据管理基础架构，数据架构为组织数据管理奠定了基础或底层蓝图。

与结构化数据中台提供的技术工具不同，以业务为导向的即时数据治理，需要的数据架构更是一种数据管理思想，是一种以业务为中心的数据资产网络。以业务认知模型，构建全局性的数据资产知识网络，为即时的数据治理提供技术支撑。

### 2.3.2 Data Fabric 架构

Data Fabric 架构作为一种全新的、囊括所有形式的数据库，被越来越多的企业用于解决数据资产多样性、分散性、规模和复杂性不断增加带来的一系列问题。

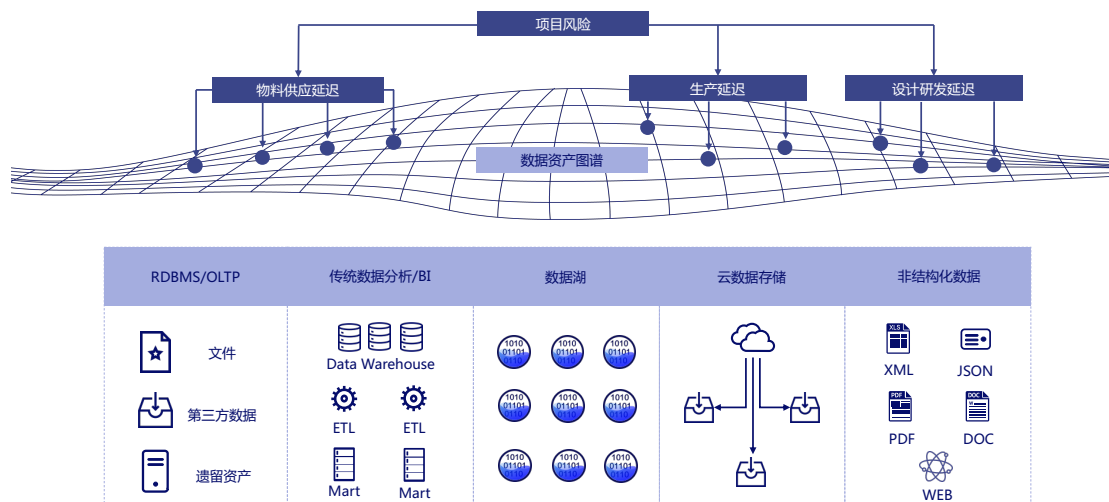


图 6 Data Fabric 架构

Data Fabric 是一种全新的数据架构，也可以叫做数据资产编织架构，它是一种面向业务、全面连接的数据架构：

- 1、它是广泛连接的数据架构，支持各类数据源，包括业务源端、数据仓库、数据湖、云存储、非结构化数据等等；
- 2、它以数据资产为中心，以业务为导向，构建数据资产知识网络，支持即时数据治理；
- 3、它是全局性、可编织的业务底图，支持业务对象建模、领域知识网络构建、业务对象关联性搜索和验证测试；

知识图谱是 Data Fabric 的重要支撑，是连接业务与业务、业务与数据、数据与数据的关系的核心表达。如下图，表达了 Data Fabric 的知识资产图谱构建、业务编织与数据运营框架。



图 7 Data Fabric 的知识资产图谱构建、业务编织与数据运营框架

第一步：构建增强的数据目录。“增强”数据目录使用 AI/ML 连接到不同的数据源和目标，抓取元数据并创建数据资产清单。这允许业务用户查找、标记和注释数据资产，并通过数据目录工具中包含的扫描器/连接器提供数据的关系。

第二步：构建数据资产知识网络，这个图谱的模型首先是以业务为导向的，围绕着业务认知模型（模型的模型）建立核心业务对象模型（核心对象、核心属性、核心系统、核心业务表、核心业务字段）。

第三步：持续的捕获元数据变化，对这些元数据进行关联分析，以便相关数据目录能够对齐和连接到核心业务对象。

第四步与第五步：以元数据变化来识别活跃的业务对象，结合业务对象的数据表、业务字段为上层提供数据准备；

第六步：业务编织和数据运营，根据业务需要，提供可组合的数据服务，包括传统的数据 API、数据集给到上层，同时也可以提供以业务对象图谱为中心的领域知识网络服务给到数据消费场景。

在 1-3 步，通常会发现数据质量问题，可以同步进行数据治理，同时，也可以围绕业务对象进行数据资产评估分析，以此评估业务流程或者业务系统的优化方向。

### 2.3.3 基于 Data Fabric 架构的业务建模与治理起点

Data Fabric 架构是也业务导向的数据架构，它形成的的是一个系统性、动态的数据资产知识网

络。基于 Data Fabric 架构的数据资产管理及治理流程是一个系统循环，如下图：



图 8 数据资产管理及治理流程是一个系统循环

第一步：为组织建立业务认知模型，这个过程也叫做业务建模过程，需要业务专家和领域建模专家参与；

第二步：对现有核心业务流程、系统和核心数据进行治理，形成基础库和核心主题库；

第三步：围绕业务认知模型，通过数据治理形成的元数据，并以业务为中心，基于数据资产知识网络的已有数据和用户交互，形成增强数据目录（信息资源目录），通过机器学习将相关元数据连接到核心业务对象，形成数据资产知识网络的底图（数据底图）。

第四步：新的业务系统、或者现有业务流程/系统升级，遵循即时数据治理的方法，围绕核心业务流程、业务表、核心业务字段与业务认知模型的核心对象对齐，在业务侧进行即时数据治理。

业务认知模型的构建，是关键步骤，也是数据治理的起步。如下所示：





图 9 业务认知模型的构建

对于业务框架，有两个关键环节：

第一个环节是组织的整体业务建模（具体的业务对象模型与整体业务模型区别开来，这个也叫做业务认知模型），基于组织的价值创造流程识别，比如对于大多数的企业组织，波特价值创造模型基本上就可以形成对企业的认知模型。如下图，是一个通用的企业认知框架。

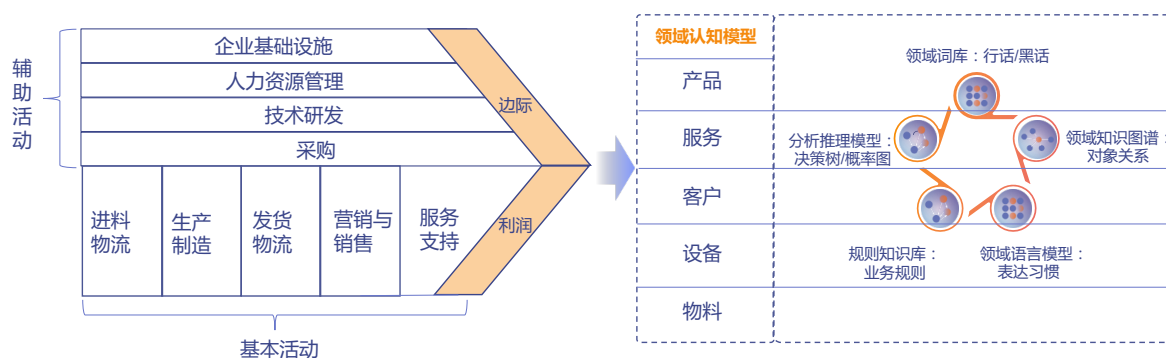


图 10 基于波特价值创造理论的企业认知框架

如下图，基于数字孪生的理念，形成智慧城市的时空认知模型：

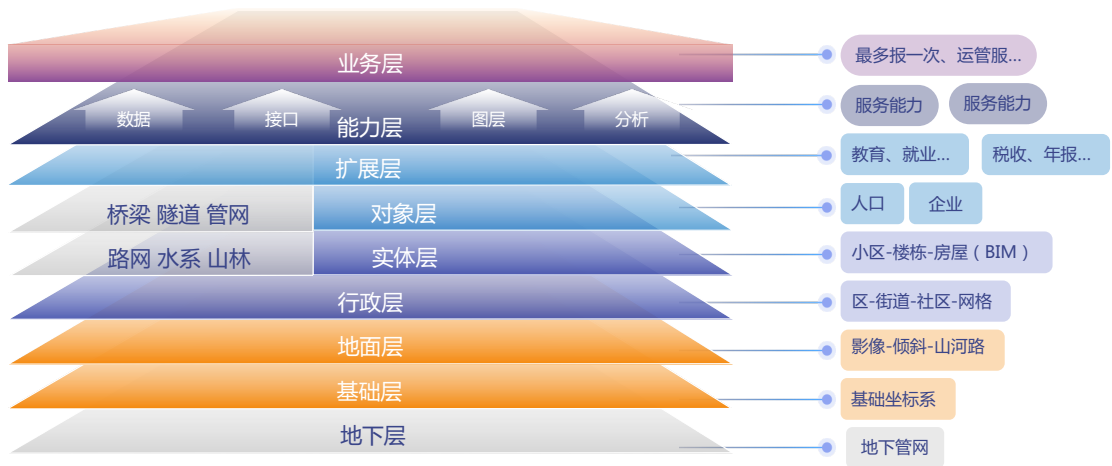


图 11 基于数字孪生的智慧城市时空认知框架

第二环节是对业务认知模型的评估，这通常是核心的数据治理起点，这个评估主要是围绕核心业务对象开展：

- 评估核心对象的核心业务流程是否稳定，是否有系统承载，系统是否成熟？
- 评估核心对象的数据采集、治理是否标准，数据质量如何？
- 评估核心对象的数据归集更新、变化是否得到有效的管理？

通过这个评估，一方面是可以知道业务职责流程/系统优化，另一方面是可以评估数据资产情况，指导数据治理。

如下图，展示了智慧城市认知框架的评估。

城市认知框架	元素	部门	业务描述	项目(系统)	重要数据(类名)	核心数据	数据质量	综合得分
对象层	人口	数据局	人口信息	网格化系统(数据局)、人口基础库、疫情防控	人员基本信息	居住地址	待校准	7
	企业	市监局	市场主体登记注册信息	XX省市场主体综合管理系统	企业基本信息	企业地址	待处理	8
实体层	房屋	市资规局	工程项目验收	一站式管理服务平台(待验收)	房屋基本信息	楼栋CAD图	数据未归集	5
		市资规局	工程项目验收	一站式管理服务平台(待验收)	楼栋基本信息	项目CAD图	数据未归集	5
	楼栋/小区	市资规局	工程项目验收(内五区)	一站式管理服务平台(待验收)	项目基本信息	项目坐落宗地(项目与道路关系)	数据情况待梳理	7
	区资规局	工程项目验收(四县)	XX县不动产登记平台房产交易系统、XX县不动产登记信息管理基础平台	无数据			2	
宗地	市资规局	国土空间规划处	时空云	宗地基本信息(无数据)	宗地坐标位置	未归集	4	
行政层	社区	市民政局	基政处(行政区划)	无系统	社区基本信息、隶属关系、边界	社区名称、行政边界道路名称	手工表数据不定期更新	7
	区县、街道	市民政局	区划地名处(行政区划)	无系统	区县、街道基本信息、隶属关系、边界(无数据)	区县、街道名称、行政边界道路名	无权威数据	2
地面层	路网水系名称	市民政局	区划地名处-地名命名	无系统	道路名称	道路名称	无权威数据	2
	农村道路	市交通局	XX市公路建设养护中心-农村公路路网发展规划(道路工程验收)	XX市公路管理局信息化管理平台	道路工程验收	道路坐标位置	数据未归集	7
	城市道路	市住建委	城建档案馆(道路工程验收)	无系统	道路工程验收	道路坐标位置	数据未归集	5
	河流水系	市资规局	测绘管理处(河流水系测绘)	时空云	水系图层	河流坐标位置	无数据	7
	倾斜	市资规局	测绘管理处(倾斜测绘)	时空云	倾斜图层	三维仿真	长期未更新	7
	影像	市资规局	测绘管理处(影像测绘)	时空云	影像图层	平面图层	长期未更新	7
基础层	基础坐标系	市资规局	测绘管理处(基础坐标测绘)	时空云	XX2000独立坐标系			10

图 12 一个智慧城市认知框架评估表

## 2.3.4 以业务为导向的即时数据治理流程

对于任何一个新的业务系统，以业务为导向的业务逻辑如下图所示：

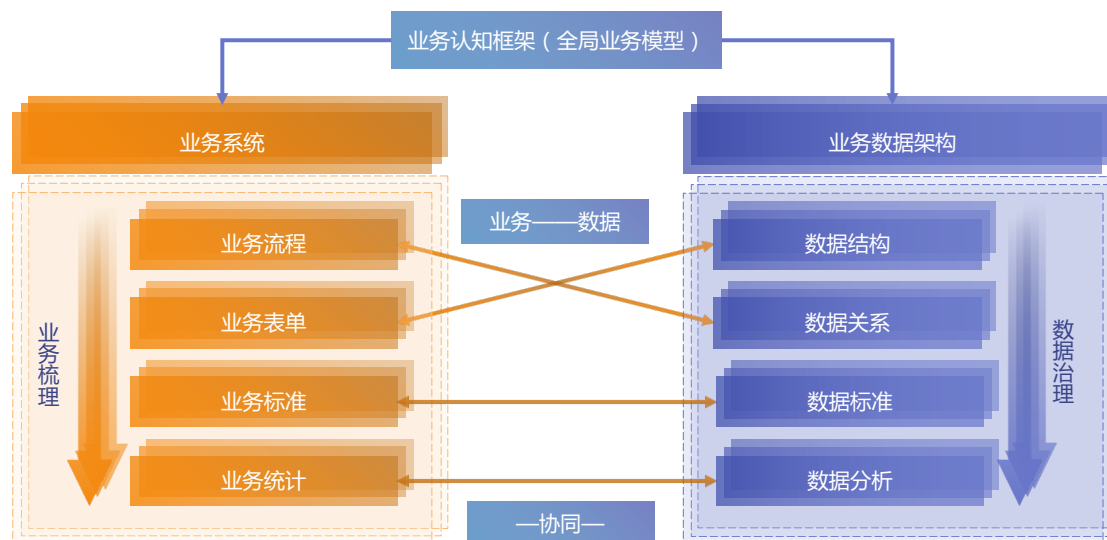


图 13 以业务为导向的业务逻辑

一方面是业务系统的业务梳理，通过业务梳理，明确业务流程、业务表单、业务标准和业务统计方式。一方面是业务的数据架构，他们之间是直接匹配关系：

- 业务表单代表业务输入和输出，对应数据结构（业务数据表）；
- 业务流程对应数据表之间的关系；
- 业务标准对应数据标准；
- 业务统计对应数据分析规则；

所以，对于业务的数据治理遵循以下基本流程：

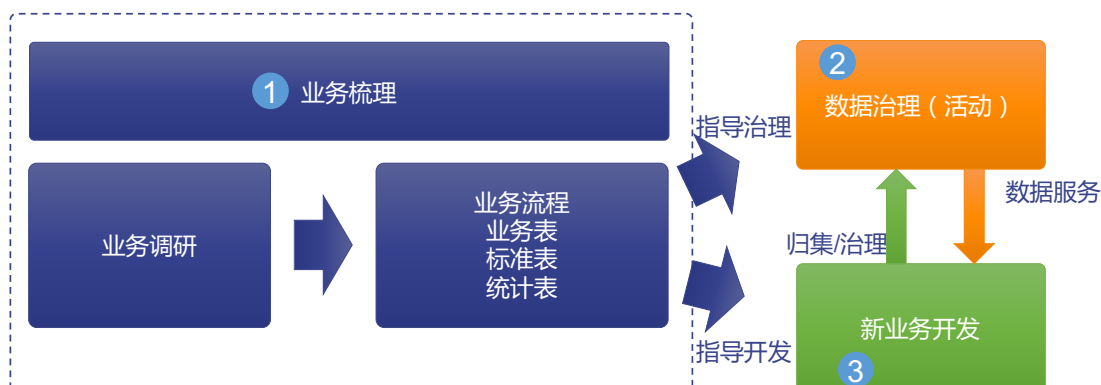


图 14 业务数据治理遵循以下基本流程

第一步，业务梳理，由业务运营团队（或者数据治理团队的架构师兼任），主要活动是业务调研，通过业务调研梳理形成业务流程、业务表、标准表和统计表；

第二步，通过业务流程、业务表等指导数据治理团队对新业务的数据归集治理，同时指导为新的业务开发提供数据服务。

第三步，通过业务流程、业务表等指导业务开发团队在新业务开发过程中遵循数据标准和规范开发，在业务过程中完成业务数据的即时治理。

### 2.3.4.1 业务梳理与数据治理

业务梳理的第一步是业务调研，通常有业务运营团队来开展，调研的颗粒度不宜太细，通常以业务流程、业务表、业务标准和统计规则清晰为标准。

无论是现有业务系统还是新的业务系统梳理，最终目标是为了明确流程，并基于流程中的节点梳理业务表：

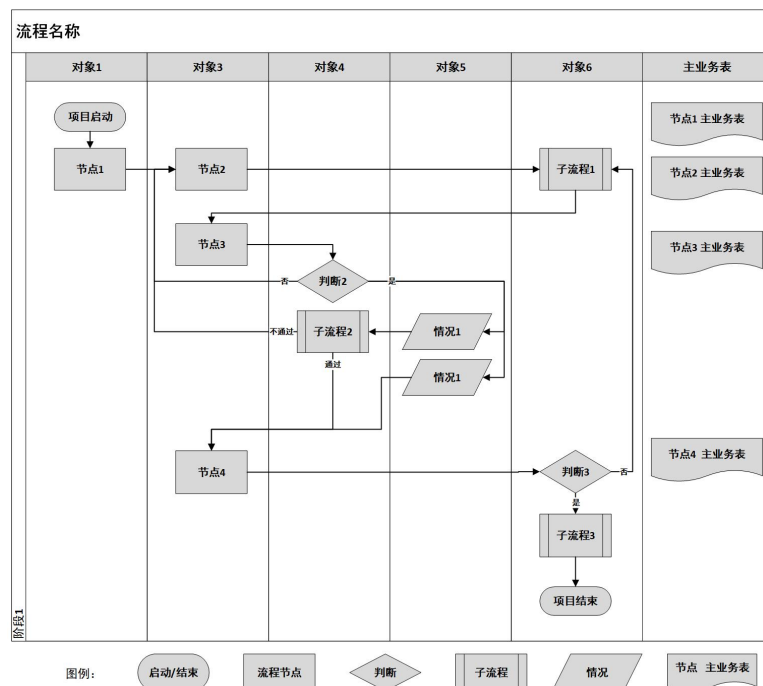


图 15 基于流程中的节点梳理业务表

并基于业务表明确其业务标准，可选形成业务标准表和业务统计表。整个业务梳理最终形成的是：业务流程、业务表以及相关业务标准表和统计表。

基于此，数据治理团队就可以面向业务系统进行数据治理了。

### 2.3.4.2 新业务开发与即时数据治理

对于新业务开发，首先也是对业务流程进行梳理，形成新业务的业务流程、业务表机器标准表和统计表，这个过程通常也是由业务运营团队来主导执行。

其次是与数据资产知识网络的核心对象进行对齐，比如产品、客户等核心对象，明确新的业务流程与核心对象直接的关联关系，以便复用相关的业务对象和数据，并约束新的业务的数据标准和规范，实现即时的数据治理，数据治理活动由数据治理团队来完成。

最后就是新的业务系统上线后，持续监测其元数据变化，以便持续跟踪业务变化、数据质量，定期形成数据质量报告，并指导业务团队持续优化业务和数据。

### 2.3.5 结构化数据的资产与安全合规评估

在整个数据治理过程中，数据资产的完整性、合规性是贯穿全局的，所以，在以业务为导向的数据治理方法中，有一个合规属性和一个整体评估：

- 合规性描述：对于增强数据目录而言，每个业务标准、业务字段都要描述其使用合规性属性，包括涉密属性、共享属性和开放属性。每个业务表的属性与字段属性是关联的，比如含有任意涉密属性字段的业务表就是涉密的等等。
- 合规性分析：基于数据资产知识网络，可以对关联的对象、业务表进行合规性分析，比如一个业务对象在不同业务表的开放性和涉密属性是否冲突。
- 数据资产评估：对于每个阶段性的治理，都可以围绕业务对象进行评估，包括其业务成熟度（基于时间、数据量及其持续性等）、核心对象关联度、业务标准、数据质量、合规性评估，其中合规性评估还包括使用的合规性，比如有条件共享的使用是否遵循相关流程规范共享。

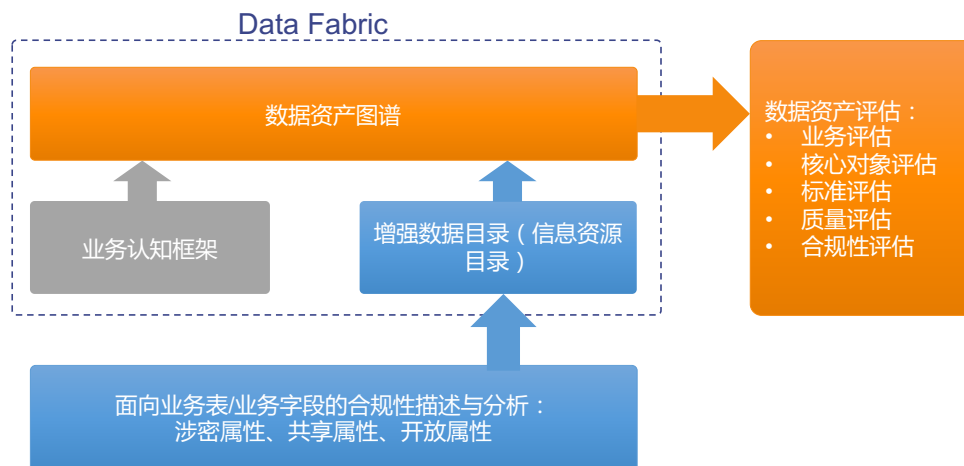


图 16 数据资产与安全合规评估

## 2.4 业务导向的数据治理优势和价值

采用新一代的 Data Fabric 数据编织架构，构建以业务认知模型为参照标准，并持续更新数据资产知识网络，并持续通过业务梳理活动，指引数据治理和新业务开发。它有几个方面的优势：

- 通过增强型数据目录和数据资产知识网络，显著降低低成本&低价值的数据集成周期；
- 通过知识图谱的对业务对象和元数据的灵活连接和扩展，降低频繁数据治理和运维带来的不断攀升的数据服务成本；
- 基于对业务对象、数据资源目录的持续 AI/ML，不断发现和优化数据资产，实现即时的数据治理。
- 为业务人员提供友好的业务编织方式，为业务对象和数据资产提供 360o 视图，并提供无缝的数据访问和共享；
- 它既可以开发常用的数据服务，也可以基于业务认知模型、数据资产知识网络，可以快速生成领域知识网络，赋能认知智能创新应用。

因此，有分析机构甚至认为 Data Fabric 架构是数据管理的未来，也是数据治理的福音。

### 2.4.1 结构化数据治理展望

从企业的数字战略来看，数据治理首先是管理问题，其次才是技术问题。只有保证业务战略清晰、核心业务流程清晰，业务系统建设运行良好，才能形成数据的可用性、一致性、完整性、合规性、和安全性，确保在整个数据生命周期中，都具有较高的数据质量，才能进一步通过数据运营、数

据应用帮助企业实现数据资产管理、发现内部数据问题、发掘数据价值，进而实现企业数据资产的盘活和有效利用。

以业务为导向的数据治理，是一个战略级、系统性的方法，并在此基础上定义组织、数据架构、和持续的多维度治理是提升业务价值的关键所在。

### 1、成立数据治理组织，高位统筹

数据治理是一个组织的全局性项目，需要 IT 部门与业务部门的倾力合作和支持，需要各个部门站在组织战略目标和组织长远发展的视角来看待数据治理。因此，数据治理项目需要得到组织高层的支持，在条件允许的情况下，成立以组织高层牵头的虚拟项目小组，会让数据治理项目事半功倍。

### 2、业务导向，架构主导，即时治理

以业务为导向，结合 Data Fabric 架构，构建业务认知模型，持续更新组织的数据资产知识网络，做到即时开发，即时治理，让业务开发围绕组织的业务认知模型来开展，形成业务与业务、业务与数据的动态知识网络，并通过元数据监控和分析，从而保障整个数据资产的质量和价值。

### 3、持续的、多维度的治理

数据治理的最终目标是提升数据价值，是一个持续漫长的运营过程，需要逐步完善、分步迭代，指望一步到位完成数据治理是不现实的，实际执行过程中，要避免以项目交付为目标，点到为止，导致治理不全面、无延续。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/506203142131010144>