

全域数据治理 白皮书

目 录

第 1 章 全域数据治理的必要性	7
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
第 2 章 结构化数据治理	17
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
第3章 非结构化数据治理	32
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
第4章 机器数据治理	47
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
第 5 章 全域数据治理加速数据驱动	56
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
第 6 章 全域数据治理的十大关键.....	61

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。也许你是首席信息官（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

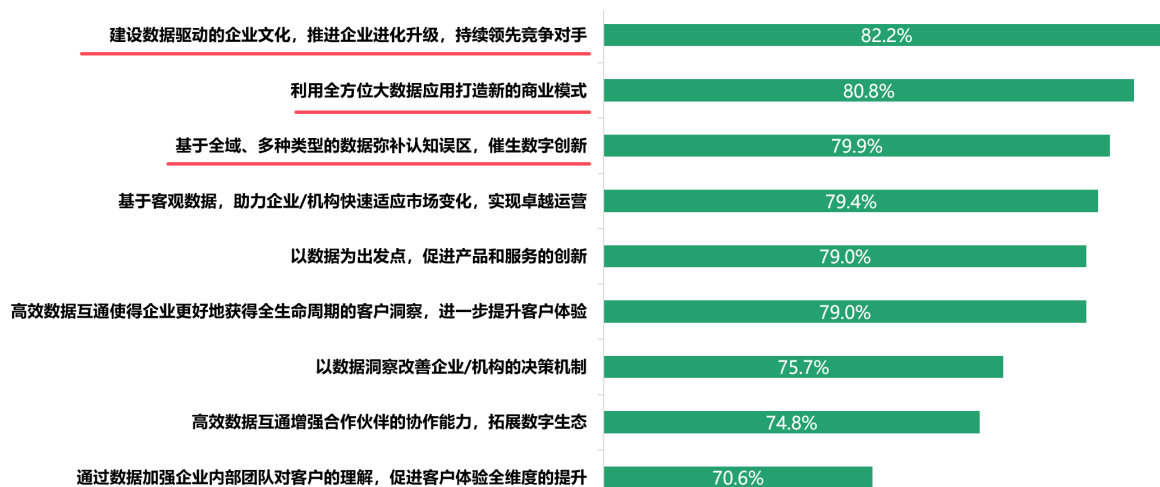


图 1 数据来源：2022 年 6 月 Forrester Consulting 进行的中国数据中台调研

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，不方便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。
- » 数据安全防护（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。



图 2 全域数据治理框架

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。

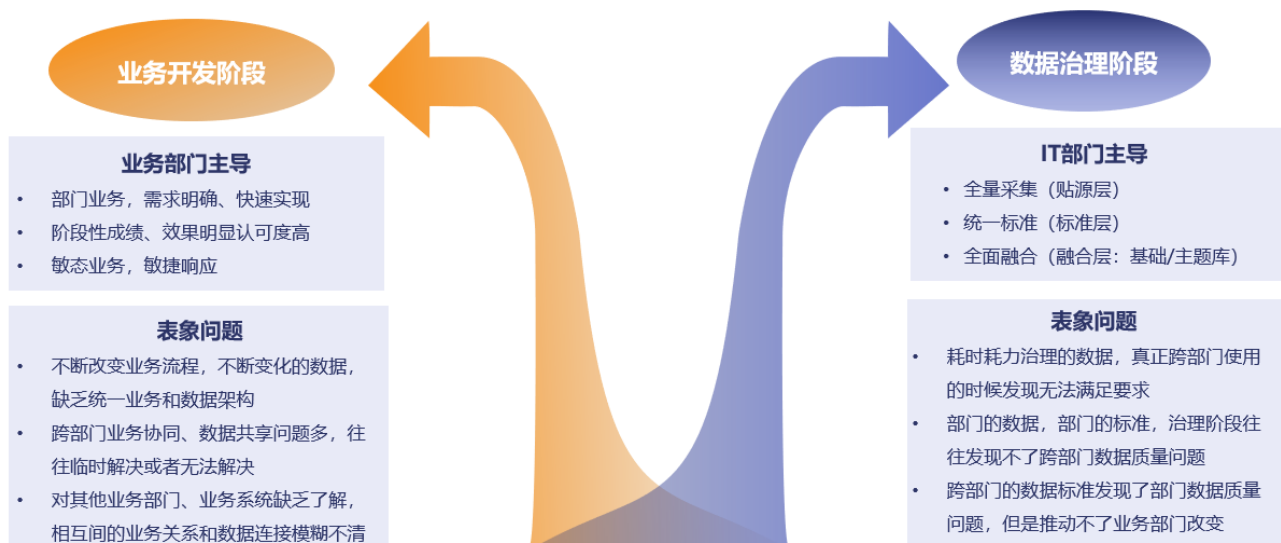


图 3 业务部门主导 VS IT 部门主导

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。

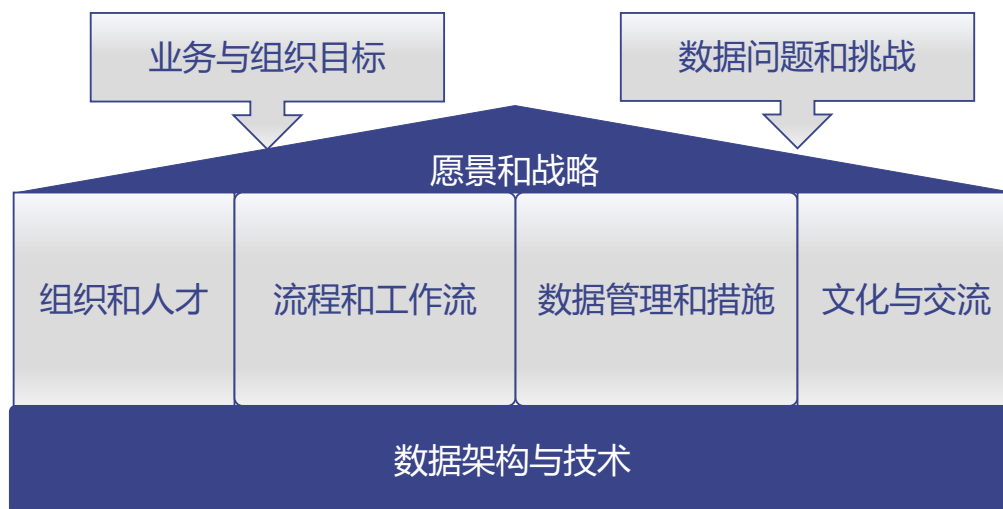


图 5 数据架构与技术

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——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、囊括所有形式的架构，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。

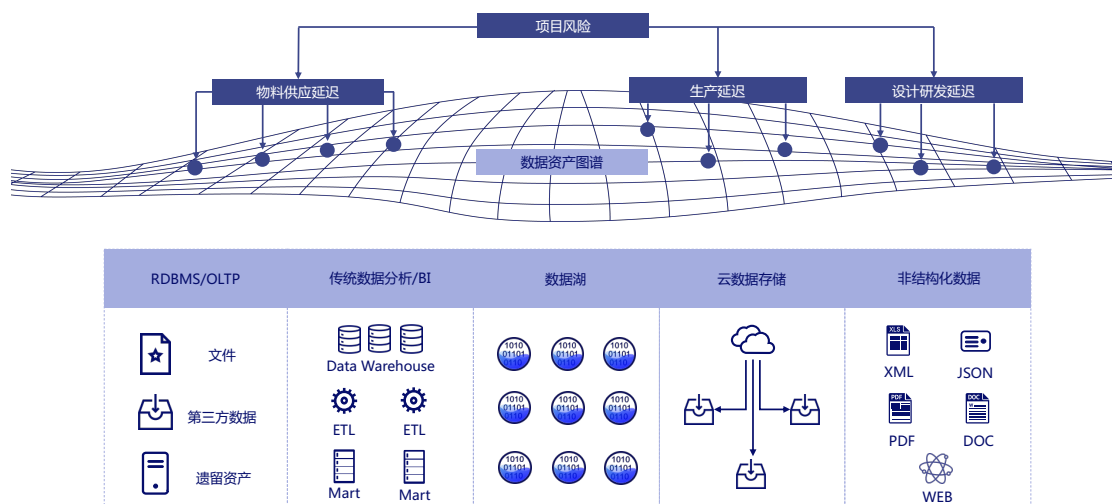


图 6 Data Fabric 架构

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 架构是业务导向的数据架构，它形成的是一个系统性、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，是关键的步骤，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。

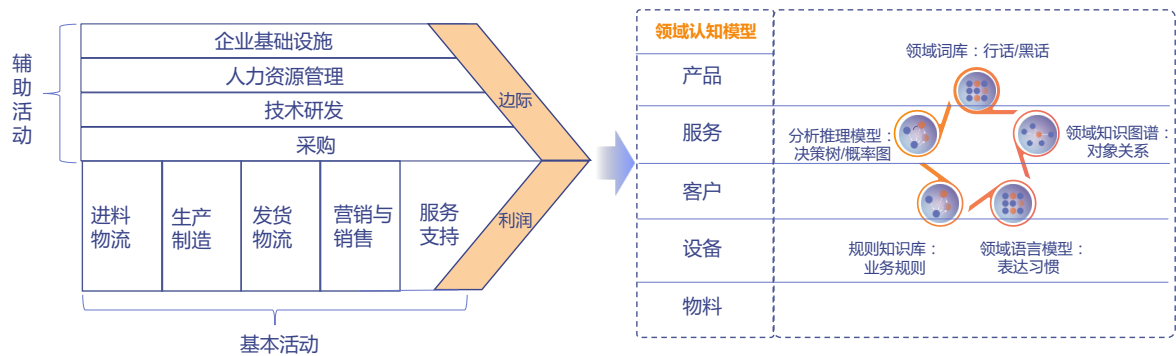


图 10 基于波特价值创造理论的企业认知框架

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、行政边界道路名称	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（无数据）	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：

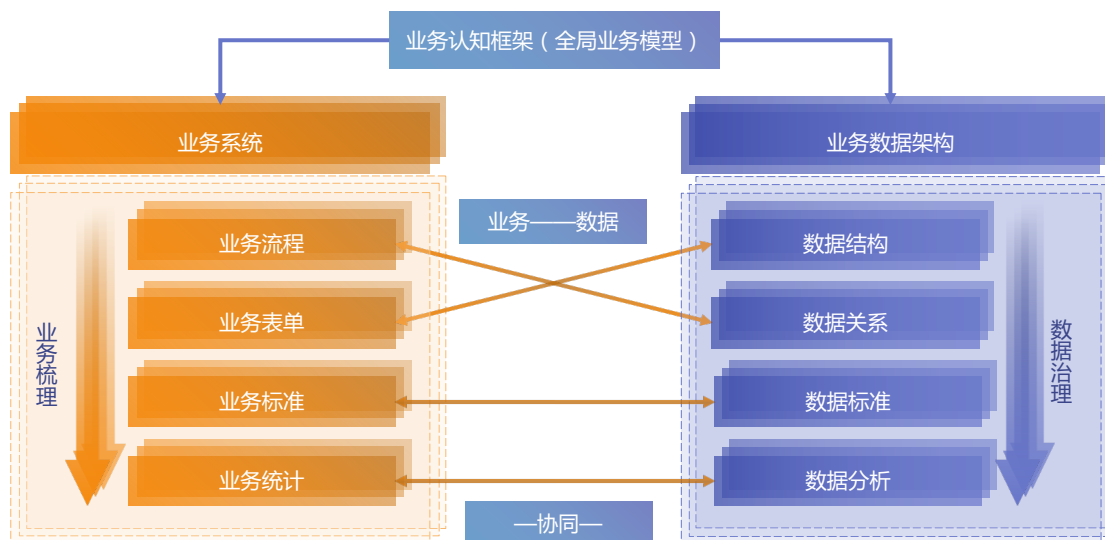


图 13 以业务为导向的业务逻辑

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：

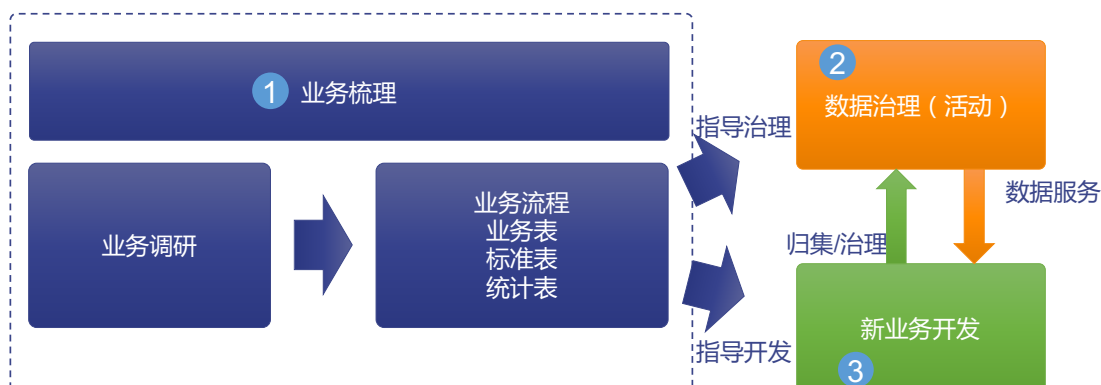


图 14 业务数据治理遵循以下基本流程

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：

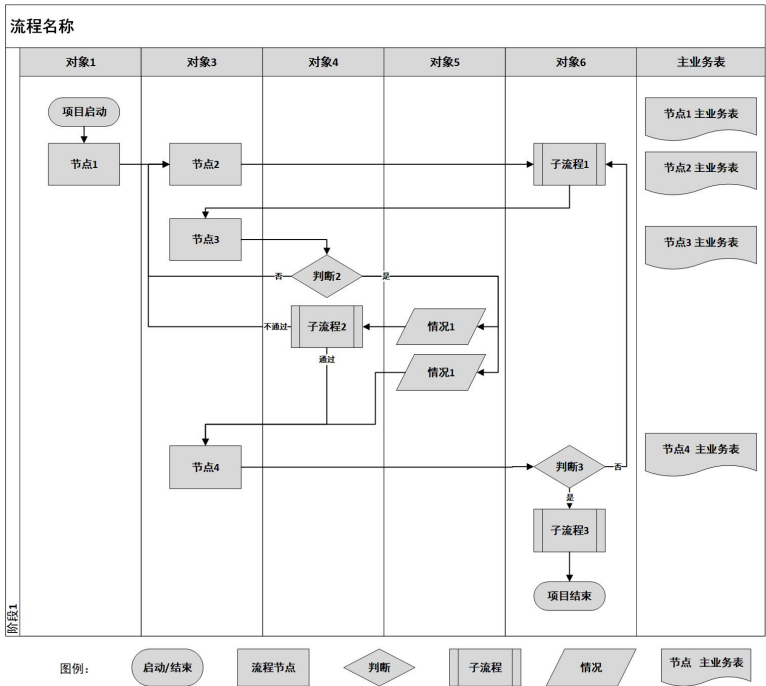


图 15 基于流程中的节点梳理业务表

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 360° 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。

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。

第3章 非结构化数据治理

3.1 非结构化数据的特征与应用场景

非结构化数据和结构化数据是一组相对的概念。

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，不方便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、HTML、各类报表、图像和音频/视频信息等。

非结构化数据具有海量、分散、多样、异构的特点，因而在非结构化数据管理方面主要存在几类挑战：

- 海量数据规模，且数据存储占比高。根据 Gartner 分析，超过 80% 的商业相关信息都是以非结构化格式存在的。
- 数据来源丰富，分散各个系统。组织在信息化建设过程中，建设了大量业务系统，众多业务系统都包含结构化数据，形成数据孤岛。
- 数据格式多样，且复杂。对于非结构化数据而言，形式多样关系复杂，基于常规单一算法技术很难识别、分析。
- 具有异构性，结构不标准。非结构化数据，结构形式复杂，难以通过洞察为业务决策提供有效指引。
- 难以保障内容安全合规。内容泄密风险始终高居不下，同时，内容审查对许多组织来说也是很大的挑战。

非结构化数据在组织的业务过程中无所不在，承担了业务数据留存以及再利用的主要责任。其应用场景与价值主要体现在以下层面：

- 业务过程数据的承载
- 工作成果的承载
- 业务规则的承载与留痕
- 组织显性知识的主要载体

3.2 非结构化数据的安全挑战

在数字化时代，数据已经成为与物质资产和人力资本同样重要的基础生产要素。随着数据价值的凸显和不断挖掘，数据面临的信息安全风险也与日俱增，给个人隐私和国家安全带来了严重的安全隐患。为应对日益严重的安全形势，我国颁布了《数据安全法》和《个人信息保护法》，在法律层面为数据安全和个人隐私保护提供法律保障。

随着日益完善的法律环境，组织的业务合规性面临严峻的挑战，从合规及内控体系建设、到行业层面的行业监管合规以及国家层面法律法规的遵从，需要通过一系列的数据治理方法来逐渐改善和提升。

在组织内部，需要确保组织经营合规，保障组织发展和创新的数据安全体系；在组织外部，需要基于行业特殊性的行业法规保障行业监管合规。如生物医疗行业、汽车制造行业出台相应的行业法规，《关于全面加强药品监管能力建设的实施意见》，GMP（生产质量管理规范）等等。此外还需要遵守各国的法律法规，例如欧盟《(GDPR)通用数据保护条例》，保护个人隐私和数据，并通过法案约束来建立组织和公民之间的信任关系；《国家网络安全战略》明确了个人信息保护的法律法规要求，战略要求落到实处，并建立实施关键信息基础设施保护制度。

多维度的业务合规性的基础是数据、访问、信息合规，组织需要搭建包括数据安全、访问安全、信息安全等在内的立体安全体系。



图 17 立体安全

3.3 以中台方案为基础的即时数据治理

经过近几年的非结构化数据治理工作，我们发现一些在治理过程中的突出的问题，主要体现在以下两点：

各类组织经过长期的信息化建设，已经形成了大量的非结构化数据，这些非结构化数据分布在各个位置、个人电脑、众多业务系统。数据需要根据业务发展会在各个业务系统流转，但是因为这些业务系统的建设由于历史原因，并没有统一的规则，不同业务系统之间数据割裂，形成众多数据孤岛。同时巨量的沉淀数据分散在这些数据孤岛之中，难以发现和再利用。

随着数字化转型在各个行业逐渐推开，各类组织也意识到这些沉淀数据的价值。但是治理过程往往是痛苦的，需要组织专门的人力对海量的沉淀数据进行治理。同时，随着组织内外环境的变化和业务本身的不断发展，造成治理效果的严重滞后。

另一方面，由于海量非结构化数据的天然缺陷，对它的检索、内容的发现与分析、数据的利用、价值的挖掘等造成极大的困难。如果没有自动化的工具和技术，单靠大量人力投入治理工作显然是不现实的。

综上，我们认为，为了保证非结构化数据发挥最大价值，满足业务发展的实际需要，必须通过对数据进行有效的治理来实现，来提升数据质量和可访问性，最终实现数据驱动运营和数据驱动创新。这个治理过程必须满足以下三个特征：

- **以中台方案为基础**：打破数据孤岛、实现非结构化数据的统一管理、全面提升非结构化数据质量和可访问性、为不同业务场景赋能。
- **即时治理**：数据在生成时已在治理框架下，避免后续投入大量人力治理造成的隐患及滞后性。
- **AI 驱动**：人工智能领域正在取得巨大进步。通过对自然语义分析、图像识别、语音识别、领域知识网络等技术的利用，可以极大减少人工的投入，同时 AI 的工作效率已经远超人类。因此，通过 AI 驱动数据治理的时代已经到来。

在以中台方案为基础，通过即时治理和 AI 驱动这个非结构化数据治理的指导思想下，非结构化数据治理从以下四个维度展开：数据资产分类、安全合规、知识治理、数据资产保留。

3.4 非结构化数据的资产分类治理

不同类型的非结构化数据拥有不同的生命周期，决定了它的业务特征和行为。在查询、权限控

制、格式、版本、审核流程、合规管控等都可能不同。因此可以说非结构化数据治理的基础，是数据资产的分类，它决定了数据治理的方向和成效。

3.4.1 非结构化数据资产治理的业务挑战

在工作过程中会产生大量的文档，按照这些文档各自的特征来分，可以是研发资料、实验数据、会议纪要、测试报告、项目周报、项目方案、项目蓝图、流程制度、公文公告、项目总结、标准体系、最佳实践等等。

在不同的业务场景中，需要对这些不同的文档进行分类的管理，针对不同类型的文档，可以定义不同的目录结构、权限体系、安全管控、流程审批、留存管理等。

同时在利用层面，对不同类型的文档在查询和洞察时往往有不同的关注点。如产品资料关注产品的类型、版本、适用的场景等等；合同资料关注甲方乙方、合同金额、生效日期等等。

但是非结构化数据的最大特征是难以描述，没有预定义的数据模型，不方便用数据库二维逻辑表来表示。传统的方式是通过文件夹和文件名的形式来承载文档的分类信息，即线性分类，针对一个类别只选取单一分类维度进行分类的场景，同一分支的同层级类别之间构成并列关系，不同层级类别之间构成隶属关系。同层级类别互不重复，互不交叉。但是这种线性分类会带来以下一些困难：

- 1) 分类时难以多维度呈现数据：同一个文档可以有多种角度进行描述，除了文件对象的名称、格式、所有者、创建时间等基本特征和定义外，还需对数据内容的客观理解进行描述，如标签、摘要、相关性等，但是文件夹结构只能从一个维度对文档进行归类，难以从多个视角归类文档；
- 2) 难以精确定位非结构化数据：通过文件名和内容关键字检索出大量相似内容，查找文档花费大量时间进行筛选；
- 3) 非结构化数据由于包括各类格式的文档、图片、音频、视频等多种异构格式的文件，较之结构化数据，其更难标准化描述和理解，如果由人工进行添加标签等属性特征来描述，则效率极其低下，面对海量的非结构化数据几乎不可实现；
- 4) 难以对非结构化数据进行洞察利用：由于非结构化数据的关键业务属性没有固定格式来承载，因此难以对非结构化数据的内容进行深入的洞察，以及结合相关联的其他非结构化数据一起分析、从而对业务决策和价值发现提供有效支撑；

3.4.2 如何对非结构化数据进行分类治理

3.4.2.1 非结构化数据的特征分析模型

每种不同类型的文档都有其各自的特征，这些特征可以用属性来描述。比如项目资料的属性可能会包含：项目名称、项目时间、项目行业、涉及的产品等等。因此在对非结构化数据资产进行分类之前，关键工作是对其基本特征与内容特征进行提取分析，并通过多样化的元数据来实现表达。这些高质量的元数据实现了结构化的描述非结构化数据资产，极大地提高了理解和分类的准确度及效率。因此，我们建立一个非结构化数据的特征分析模型，如下图所示：

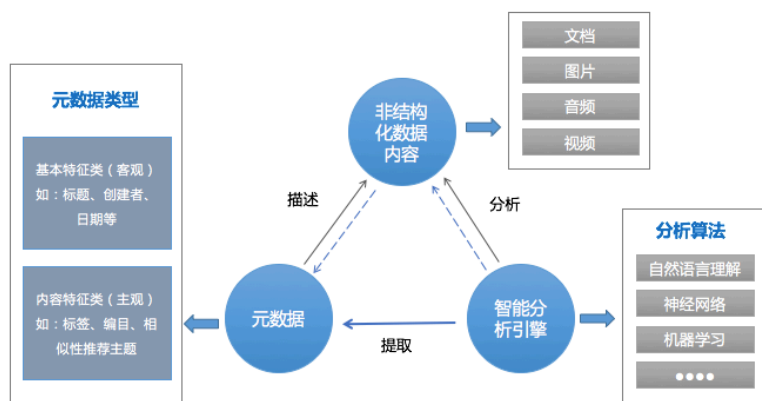


图 18 非结构化数据的特征分析模型

非结构化数据的元数据可以分为基本特征类（客观）和内容特征类（主观）两个类别。

- 基本特征类：参考文件的基本属性，实现对非结构化数据对象的规范化定义，如标题、格式、来源、大小、创建者、日期等。
- 内容特征类：基于对非结构化数据的内容的理解和分析，归纳其内容包含的特征，加深对目标文件的客观理解，如标签、编目、相似性推荐主题（如行业名称、作者单位）等。

内容特征类元数据，一般首先由人主观定义出关键词、编目，并借助于数据资产管理平台的人工智能算法（语义理解、机器学习、正则表达式等），实现自动化提取和标注。自动化工具避免了人工标注带来的繁复和低效，AI 驱动是非结构化数据分类取得成功的关键点。

3.4.2.2 非结构化数据的分类方法与过程

非结构化数据的分类方法包括人工分类和自动分类两种，这两种方式是互为补充的，可以在文档

生命周期的不同阶段即时进行。

人工分类由用户根据数据主题、业务类别、适用对象、保密级别等因素，进行主动识别和命名类别。

自动分类一般由数据资产管理平台根据特征分析模型进行自动识别分析数据内容，提取属性信息（元数据）并自动标注完成。智能化的机器分类覆盖存储、检索以及消费使用等多个阶段。

非结构化数据的分类首先必须基于不同的业务场景，在业务场景下，按照业务活动的过程来展开，在业务活动中去识别和定义非结构化数据的输入、输出及其相关要求与特征，最后经过归纳总结，来对非结构化数据进行分类治理，这个过程主要以人工为主，需要各个业务领域的专家资源的配合。

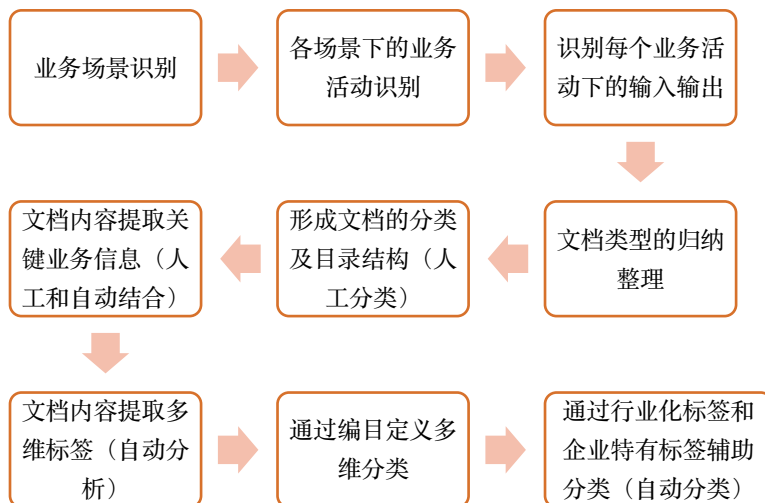


图 19 非结构化数据进行分类治理

治理规则制定以后，在规则的执行落地过程中，可以通过 AI 技术辅助人工进行分类治理，同时保证治理的即时性。

3.4.3 数据资产分类示例

- 1) 通过业务活动识别分类：通过人工梳理业务活动及流程，按照管理主体 - 管理范围，逐步向下进行拆解业务主题，形成树形分层结构的数据资产目录。
- 2) 通过编目分类：对某些特定的文档类型（如发票），人工预定义出该类文档的一组描述属性（即编目），将该编目输入到数据资产管理平台，当用户上传此类文档时，通过平台智能分析引擎提取出编目属性值并标记该文档。
- 3) 通过“密级”分类：按照不同行业的数据分级保护的要求，需要根据数据重要程度和敏感程

度由高到低划分不同的秘密等级，因此通过“密级”标签实现辅助分类。进行数据分级时，应考虑其安全属性（机密性、完整性、可用性）遭到破坏时造成的影响的客体。

- 4) 通过“敏感信息”分类：通过对非结构化数据内容的扫描和基于敏感词规则库的自动分析，自动将敏感文档标注区分后归类到对应目录位置（如隔离库）。

3.5 非结构化数据的安全合规治理

在前面的章节中已经介绍，非结构化数据的安全与合规面临着诸多挑战。从国家层面、行业层面对非结构化数据的安全合规都提出了不同的要求，下面通过两个典型行业分别加以介绍。

3.5.1 生物医药行业的安全合规治理

GMP（生产质量管理规范）：要求制药、食品等生产企业应具备良好的生产设备，合理的生产过程，完善的质量管理和严格的检测系统，确保最终产品质量（包括食品安全卫生等）符合法规要求。

以客户的业务为中心，围绕安全管理与合规管理开展安全合规治理工作。安全管理主要有四个方面，包括业务应用安全，数据安全，访问安全，信息安全。合规管理主要有六个部分，包括短缺药品合规，说明书合规，数据合规，标签合规，患者安全报告防泄密合规，流程合规。

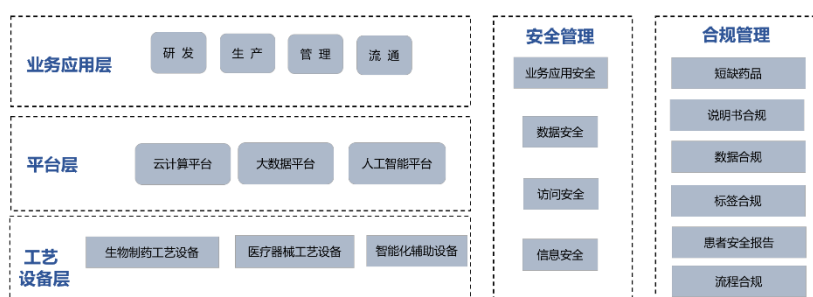


图 20 生物医药行业的安全合规治理

3.5.2 汽车制造行业的安全合规治理

汽车制造企业在文档管理方面需遵守的法律法规及标准如下：

- 《商业秘密保护规定》
- 《GB/Z 26822-2011 文档管理电子信息存储真实性可靠性建议》
- 《ISO27001 信息安全体系》

安全合规场景包括：

- 通过业务合规治理的体系，跨国车企，及外资车企的业务数据符合出境合规
- 通过合同等形式，以及建立第三方服务商的数据合规管理制度，深入管控第三方的数据合规
- 企业敏感文件通过内容安全服务工具与技术，防止信息外泄；

3.5.3 非结构化数据安全合规的治理要点

以健全的安全合规制度和运营方法，及强有力的组织保障为基础，通过立体安全体系在内的数据安全，访问安全，信息安全等管理方法，结合个人隐私数据保护，企业敏感信息识别，非法内容管控，内容访问审计，等等治理工具，达到安全合规的治理目标。

针对不同的行业，有各类法律规范，在做安全合规治理时，必须和这些行业规范进行深入的结合。

3.5.3.1 安全合规治理的框架

安全合规的治理需要以国家法律法规与行业标准为指引，结合组织业务实际情况，对非结构化数据资产进行分类分级、并在非结构化数据的整个生命周期进行管控。其治理的基本框架如下所示。

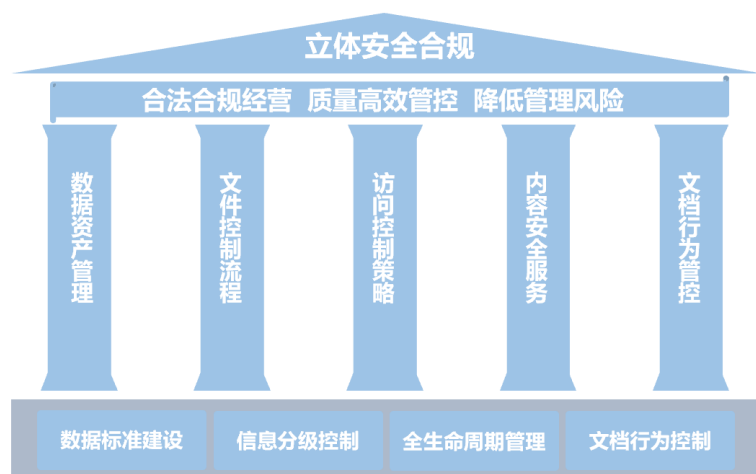


图 21 安全合规治理框架

针对不同的安全场景，可以采用的技术手段有以下方式：

➤ 文件泄密管控：

- 对内：文件上传解密，明文存储，内部透明访问；文件下载落地加密，保障数据不外泄；
- 对外：通过外发包管理工具，为解决外发文档易泄露、难管控的问题，比如控制外发文件的读取权限、有效期、打开次数等。

➤ **隐私数据管控：**

把业务流程中涉及到的各种隐私数据梳理出来进行分类管理，并实现可视化，解决隐私数据到处分散、易泄露的问题。

➤ **文档内容安全管控：**

企业客户使用文档过程中，进行文档库检查与杀毒，比如文档下载、预览、编辑等操作时，检查文档的安全合规性。

➤ **非法内容管控：**

检测并识别到非法内容和敏感词内容的文档，比如用户上传含有非法内容的文件将被隔离。

3.5.3.2 安全合规治理的组织与制度保障

安全合规的治理，涉及到各个业务领域，不同的业务领域有不同的诉求，必须精准的理解安全合规的需求才能做好安全合规的治理工作。这就要求在安全合规治理过程中，治理团队必须将各个领域的专业人员纳入进来。同时，安全合规治理是公司的战略分解，需要从高层、中层管理层都纳入进来，由他们来保障治理工作的持续推进。同时还要有数字化支持团队、持续运营的团队等来做支撑。

以客户业务为中心，围绕安全合规开展全域数据治理工作，可以从制度制定、人员配备、体制运营等方面建立适合业务实际的数据合规体系，以做到数据合规有人管、有制度保障、有体系运营管理。比如通过安全运营，持续优化事前安全策略、制定安全规范引导用户事中行为，借助数字化工具实时预测和报警安全事件，定期开展事后审计等。

3.5.4 安全合规治理带来的价值评估

通过安全合规的治理，可以为组织带来如下价值：

- 提高组织合规审查效率；
- 通过一系列的业务合规性措施，改善组织内部的业务合规的管理能力和流程；
- 帮助组织的相关业务流程符合法律法规的要求，助力组织数字化转型；
- 提高组织客户的安全信息保护意识，降低文件泄露的风险；
- 完善组织内部的安全合规机制和流程；

- 帮助组织员工充分理解国家数据安全相关法律法规，减低因员工自己的行为所带来的风险；

3.6 非结构化数据的知识治理

3.6.1 知识管理建设面临的诸多挑战

现代经济理论已经把知识列为重要的生产要素，知识已广泛运用于现代组织的生产过程。知识管理的建设有助于提高组织核心竞争力和经济效益，是现代组织管理的重要组成部分。数字经济时代，知识生产力是组织竞争力的关键。

在组织经营过程中，业务活动与知识之间的相互渗透程度日益提高，传统管理模式里知识的管理相对粗放，往往依赖于员工自发或者组织突击式管理，知识管理建设面临诸多挑战：

- **缺乏知识战略规划**：知识管理战略不清晰，对企业经营战略目标承接缺失，缺乏科学的方法体系支持，导致知识管理没有重点，缺乏动力，导致治理失败；
- **知识壁垒林立**：过多的知识壁垒，导致人才流动带来知识资产流失，团队间缺乏知识共享的文化，也没有专门的管理部门或者制度，知识管理难以发挥价值；
- **知识管理成本过高**：知识抽取、关联、传递等工作耗费大量人力成本，历史知识资产梳理工作量巨大，缺乏智能高效的治理方法和工具，投入产出比较低；
- **知识管理效果不明显**：治理成效滞后，投入决策难度较高，无法准确预估或者量化价值，同时没有有效的运营手段，导致治理效果不可持续。

3.6.2 组织如何规划知识管理

企业知识多具有碎片化、数量大、来源多等特征，因此知识管理很难标准化。传统的知识管理往往花费大量的人力投入，却很难带来业务经营改善的成效。

随着科技的进步以及数字化转型的逐步深入，在知识内容生命周期中的知识采集、知识管理、知识关联、知识应用四个领域，自动化与人工智能技术的快速发展，大量抽象而复杂的知识场景，可以不再依赖人力，通过信息技术的融合能够形成巨大的知识赋能。

因此，在数字化发展的宏观趋势推动下，企业知识管理的数字化转型潜力很大，通过科技赋能，大大提升知识管理的治理效率：空间上能覆盖企业各类业务场景，客观全面的归集和管理知识内容；时间上能持续积累和传承知识，避免因人才流动或主观原因而造成知识内容的流失。两者相辅相成，

共同提高组织核心竞争力和经济效益。

3.6.2.1 知识管理的目标是改善组织业务生产力

通过知识治理借助智能分析手段来改善投资回报率应该作为知识治理的指导理念，将组织战略、业务场景、知识融合相结合，提供端到端一体化的知识管理解决方案，并分步骤全程协助组织进行知识管理数字化转型及知识运营，完成通过知识管理改善业务效益的建设目标。

知识管理工作的开展，需要围绕三个维度来展开：知识战略、知识治理、知识运营。

首先是制定知识管理的战略。知识战略的制定，必须在组织经营战略的指引下开展。

知识管理战略的制定，通常需要考虑的因素有：组织经营目标的年度规划、影响业务目标的关键领域和业务环节、在这些领域或者环节中确定最有价值的知识内容与知识活动、经营目标对不同岗位角色的能力要求等。



图 22 知识管理战略框架图

基于以上这些因素来定义整个组织一到三年的知识战略规划，为接下来的知识治理具体工作的开展框定范围和确定指引方向，确保知识治理措施合理得当、保证知识治理的投入产出比以及治理效果的可持续性。

知识管理的内容对象和范围并不一定要包括广义上的所有知识内容，一般需要综合考虑管理价值、需求强度、成本接受度三个方面，最终选择最适合进行知识管理的知识对象进行圈定。选取的知识范围不同，将直接影响后续知识管理的需求、方案、实施要求等。

理论上全面的知识覆盖有利于发挥知识统一管理的赋能效果，但在一些特定情况下，比如在资源有限，或者组织经济活动非常聚焦的情况下，需要针对性圈定高优的知识对象进行管理。

主要从两个方面考虑：

- **业务价值**：指该知识对于组织经营活动的价值
- **管理成本**：指实现对该知识的管理需要投入的资源及人力成本

优先选定的范围应该是业务价值高、管理成本低业务领域；其次是业务价值高、管理成本高的领域。不推荐选择业务价值低的领域。

3.6.3 如何围绕知识战略开展知识治理

在制定清晰的知识战略、确定关键的业务领域之后，就要进入有序的知识治理阶段。首先需要根据知识战略进行分解，进一步明确知识治理的关键范围，核心知识资产，并对当前的业务活动、知识治理现状进行梳理分析。包括每个业务环节对知识的要求、知识的创建、管理、利用的场景及相关角色的关系。识别每个环节的知识的输入、输出，知识的特征、应用场景、带来的价值。

3.6.3.1 知识治理的主要内容

知识治理的主要内容可以从以下几个方面进行：

围绕核心业务场景梳理知识主题：在重点业务范围，进一步梳理业务活动的知识需求、对历史知识内容的盘点，形成知识体系，知识体系不仅包含知识的内容样例、分类、结构层次、功能等，还应该解释知识资源在组织内外部的相关链接及特征。这其中，比较重要的是知识标签体系。知识体系可以知识主题的形式来承载，知识主题的梳理需要符合知识战略，并保持长期相对稳定。

知识主题的梳理，需要组织内部的业务专家来一起协作，专家在梳理过程中起到绝对的主导作用。同时，通过 AI 技术和自动标签技术，可以实现将符合主题特征的知识自动关联、归集到对应的主题下。

高效管理的知识库：在知识体系下，将知识内容进行目录级管理，并承载知识内容的创建、审核、发布的过程管理，同时也可以实现对知识内容的精准定位。

通过专家促进知识沉淀与交流：每个组织内部都有一些核心的业务专家，从某种意义上说，这些专家决定了这个组织的知识能力、从而对提高业务生产力有着举足轻重的作用。根据组织不同的业务领域、结合公司的人力资源管理战略指定相关的业务专家，最大化发挥专家在知识治理、创建、分享上的价值，促进公司全员的知识储备和能力的提升。

围绕知识兴趣形成知识圈：组织业务发展的主体是广大的员工，他们是组织价值的直接创造者。

激发广大员工的知识创造力，是知识治理的重要途径。通过知识圈的建设，可以实现知识按照兴趣和不同业务需求进行汇聚、酝酿沉淀、互动交流。通过将共同领域、兴趣的人汇聚在一起，进一步促进知识的产生、改善、分享、利用的效能，使知识生态的以不断完善。

3.6.3.2 领域知识网络提升知识治理效率

在知识的应用场景，需求可以分为管理类与赋能类。管理类需求特指基于知识内容本身的增删改查需求，比如知识的存储、分享、检索等；赋能类需求指应用数字化技术或 AI 技术“理解”内容语义，进而主动通过各类知识工具，面向业务活动进行赋能的需求。

管理类需求可以通过知识生命周期理论，进行覆盖式挖掘，一方面从下而上，对组织现状进行问题分析和需求挖掘上报，另一方面从上而下，承接战略需求向下进行的目标拆解。

通过 AI 技术、一方面可以自动化实现知识内容的分离、提取、关联，从而更叫高效的发现知识、汇聚知识；另一方面，通过对人员的特征分析、行为分析，自动发现知识的优化场景、利用场景：将知识内容精准推荐给需要的人，将知识的互动信息发给知识贡献者来改善知识内容，将知识需求信息发给知识创建者来提供知识内容。

通过这些自动化的手段，不但保证了知识治理的及时性，极大减少人工投入，还可以帮助知识管理者发现尚未发现的问题，为知识治理的改进、决策提供不可替代的帮助，从而不断提升治理效能。

机器学习需要大量的训练（如知识图谱标注等），训练以海量的数据为基础，并需要投入大量的人力标注投入。针对不同行业、不同领域的知识网络建设，可以在相同行业和领域复用，从而大大提高机器学习的效率。

3.6.4 通过知识治理提高知识运营效率

知识管理没有终点，他是伴随着组织的经营过程不断发展的，需要根据企业经营战略不断进行迭代、推广、价值评估等。知识主题、知识库、专家领域、知识圈的需要围绕知识战略去展开工作，这个过程需要持续不断的运营工作，从而为组织的经营发展和业务创新持续提供动能。

知识运营与知识治理相辅相成，知识治理效果的达成与可持续性，需要通过知识运营来实现，反过来讲，通过知识治理也可以提高知识运营的效率。

3.7 非结构化数据的资产保留治理

3.7.1 非结构化数据资产保留的业务场景

非结构化数据作为工作成果重要承载物，对国家、行业以及企业自身都有重大价值，同时在法律法规、行业规范和企业自身管理角度，对此类资产的保留都做了很多管理要求。由管理对象不同，对非结构化数据资产的保留可以分为如下三种情况：

- **国家法规要求**：国家对政府机构、国有企业中的各级组织的不同类型的文件，保留时长不同，同时对内容的完整性也有一定的要求。
- **行业规范要求**：制造业、生物医药、金融、房地产等等不同行业对行业特征的文档内容也有按照不同类型进行保留的要求。如质检报告需要长期保留，药物试验资料需要保留至上市后五年，房屋设计图纸需要长期保留等等。
- **企业职能部门要求**：财务、人事、采购、法务、市场、行政等企业的职能部门也管理了大量的文档资料，除了需要满足国家法律法规以及行业规范要求之外，还需要按照供公司内部要求，做好文档保留工作，以便于更高效的利用当前资料和历史资料。

3.7.2 非结构化数据资产保留的治理

对于不同场景下的各类非结构化数据的保留要求，如何进行治理呢？一般的治理过程如下(有先后顺序)：

- 1) **识别业务场景**：按照业务场景收集相关文档资料作为治理对象；
- 2) **识别管理规范要求**：该业务场景下有哪些法律、法规要求，以及组织内部管理要求；
- 3) **文档分类**：在不同场景以及管理要求下，将文档进行分类管理；
- 4) **针对分类定义管理策略**：针对不同类型的文档，定义详细的保留策略。

对于具体的治理策略，主要有以下几个方面：

- **保留年限**：区分不同类型的文件需要的保留期限，通常会有 3 年、5 年、10 年、永久保留等不同期限，可通过元数据进行记录。
- **归档形式、位置**：不同类文档或不同部门的文档的归档形式、位置要求不同，可于各部门的目录层级中创建归档目录，或者单独统一的位置创建归档目录用于不同场景的文档保留。

- **归档后的权限要求**：归档后的文档权限需要调整，便于长期保留和后续利用。
- **归档时的元数据要求**：在不同的要求下，对文档归档时的历史痕迹信息以及相关元数据需要进行不同的保留，如文件的创建时间，密级，审核的历史信息等。
- **纸质文档归档与电子文档的匹配管理**：当前情况下，可能还有部分电子文档同时保留了纸质内容。在电子文档保留时需要将纸质文件的物理位置添加至文件标签便于查看，以及实际文件的查找。
- **对归档文档的查询要求**：归档后的文件为了方便后续利用，需要对元数据（标签、文件标题、文件内容、相关编写人等信息）进行设置，以方便查找定位。
- **文档审核要求**：文档的归档，往往需要人工介入进行审核与管理、需要定义不同角色的权责、审核的流程等。

3.8 非结构化数据治理为组织生产力赋能

通过非结构化数据中台的建设，打破数据孤岛、实现了非结构化数据的统一管理、全面提升非结构化数据的数据质量和可访问性，同时也极大减少了重复能力建设带来的成本浪费和管理的复杂性。

利用非结构化数据中台，结合对非结构化数据的即时治理，提高企业经营的安全与合规，规避和减少因违规或者泄密造成的经营损失；提高非结构化数据资产的管理效率、利用效率，并通过知识治理，实现数据资产知识化，知识不断深入到业务场景，提升组织的业务能力、激发组织的创造力，为组织经营目标的达成以及保持竞争力持续赋能。

同时，随着非结构化数据处理的技术不断完善和发展，尤其是 AI、机器学习、领域知识网络等技术能力的利用进一步加速了治理效率的提升，不断促进非结构化数据治理效果带来的生产能力提升。

第4章 机器数据治理

4.1 什么是机器数据

机器数据，顾名思义指的就是由机器产生的数据，包括软件和硬件。具体到 IT 领域，主要指服务器、网络设备等硬件或虚拟硬件运行过程中产生的状态数据，应用系统运行过程中产生的日志数据、性能数据。通过机器数据可以准确的掌握业务及业务承载平台的基本运行状态，例如 CPU、内存、磁盘等资源的使用情况、网络流量情况以及应用运行情况，是运维监控领域重要的数据来源。

这里我们谈到的机器数据，主要包括以下几种类型：

- 指标数据：指软/硬件系统运行过程中的各项性能指标数据，具有实时、海量等特点
- 日志数据：包括机器运行日志、系统日志、应用日志等，格式差异大，具有实时、海量等特点
- 链路数据：指与用户体验相关的端到端的应用跟踪数据，具有实时、海量等特点

4.2 机器数据治理概述

伴随着中国数字经济的快速发展，产业数字化占数字经济规模的比重也呈逐年升高的趋势，这也有力地推动了数字化转型的步伐。组织日渐认识到数字化对管理流程和商业模式的作用，越来越多的组织启动数字化转型计划。

数字化转型的核心和关键就是对数据价值的充分挖掘与释放，通过数据驱动创新和运营，推动数字化的变革与发展。机器数据作为大数据应用最重要的组成部分，目前仍然散落在各个应用系统和管理系统中，孤立存在，尚未有效形成数据资产，价值利用较低。

4.2.1 以中台方案为基础的即时数据处理

在多云与云原生技术的大趋势下，外部环境发生质的变化，交付越来越敏捷。面向机器数据的治理要以业务需求为导向，以运维中台方案为基础，以数据虚拟化为架构，以领域认知智能为技术底座，将这些分散的各类机器数据进行有效的汇聚、整合与治理，通过可观测性能力建设，对数据进行分析应用，解决组织业务运营与管理中遇到的问题，提升运维效能、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。



图 23 以中台方案为基础的即时数据处理

4.2.2 机器数据治理的价值

机器数据治理，加速了组织的数字化转型步伐，提升了组织的数字化服务能力。首先，以中台思想为基础的机器数据治理，打破数据隔阂，有效解决组织面临的数据孤岛、数据标准不一致的问题，促进数据流动，充分释放了机器数据的价值。

其次，通过对多种数据源数据的汇聚与集中，在安全管控方面，进一步满足了对法律法规及行业要求的合规性，并显著提升对信息管理审计能力，加强对安全风险的管理与控制。

第三，通过对机器数据的整合与关联分析，能充分释放数据在运维管理及业务运营方面的价值，提升故障的识别及根因定位能力，保障 SLA；同时，通过对数据业务特性的分析，提升业务决策分析水平，持续优化用户体验及业务服务，驱动业务的健康发展。

4.3 加强合规与审计

4.3.1 合规与审计面临的挑战

进入数字化时代，人工智能、云计算、社交网站、大数据处理等新技术逐渐普及，国家对信息安全与个人隐私保护越来越重视，相继出台了《网络安全法》、《网络安全等级保护制度 2.0》、《数据安全法》、《个人信息保护法》等多项法律法规，各行业也陆续出台了行业法规要求，如《商业银行内部控

制指引》等，监管机构及主管部门对合规审计的内容要求越来越细，频率要求越来越高。对于组织来说，在合规与审计方面面临着诸多挑战，主要体现在：

1. 海量日志数据合规留存难。日志数据由各类系统产生，日志数据分布散、数据量大、结构复杂、存储环境多样化，因此传统的日志管理技术存在兼容性差、查询效率慢，存储成本高等问题。
2. 海量日志数据的实时性、一致性难以保证。日志每时每刻都在产生，数据量巨大，流量高，对日志系统的吞吐量要求极高。
3. 日志数据管理时，缺乏有效的技术手段来避免个人隐私数据泄露风险。
4. 审计工作量大、效率低。审计工作依赖于多个云平台提供的日志，审计的方位涉及云主机、系统、数据库、服务到租户等，缺少统一的审计平台，需要使用多种工具才可完成。

4.3.2 通过机器数据治理解决合规与审计问题

日志数据采集：采用无代理采集以及代理采集 2 种方式，支持多种标准采集协议（Syslog/SNMP-Trap/SNMP/JDBC/SSH/HTTP），通过主动推送或主动拉取方式将异源异构的海量日志数据接入至日志处理平台，从而实现日志统一采集接入管理。系统的日志采集策略包括支持实时、定时采集日志、日志采集任务可压缩传输日志、可配置限制日志采集任务对宿主机网络、CPU、内存等资源的使用等，以满足不同业务场景下对日志采集的灵活使用。

日志数据解析：日志解析服务需要支持水平扩展，来满足数据吞吐量的高要求；增加缓存应对日志流量峰值；支持多线程处理，时延低，Schema on read 等技术来保证日志解析的实时性；支持分布式部署，来满足可靠性要求；通过缓存+偏移量标识技术，保证日志一致性。

日志数据存储：对于海量数据应该考虑数据的分级存储方案以节约存储成本。对日志数据分级处理，采用分级存储的架构，将机器数据按照热数据、温数据、冷数据进行划分，充分利用发挥服务器本地盘（SSD 硬盘+SAS 机械盘）和 NAS 存储的多层次存储结构的性能优势。每类数据保存周期不同，实现机器数据的全生命周期管理，低成本满足安全合规需求。同时，系统存储具备安全机制，保护日志记录不受未授权的读取、删除或者修改，日志记录要持久化存储，要存储在非易失性介质中。

日志数据查询：提供丰富的搜索方式，配合智能搜索引擎，以应对多源异构复杂搜索场景。同时可提供实时搜索、全文检索、秒级响应的搜索体验。

通过对机器数据治理，可以为组织有效解决以下问题：

1. 海量日志数据合规留存，满足各类法律法规的要求，保证数据的实时性、一致性。

2. 日志数据管理时，有效避免个人隐私数据泄露风险。
3. 在统一平台上开展审计工作，提升审计效率与审计结果的准确性。

4.3.3 机器数据治理对合规与审计的价值

通过机器数据治理，帮助组织加强安全管理，加强合规与审计能力，降低运营风险。

4.4 提升运维效能

4.4.1 运维管理面临的挑战

数字化技术的发展及云原生应用的兴起，业务系统变得愈加灵活与复杂，分散独立的监控运维工具，导致运维工程师无法有效对故障及风险进行管理，同时难以在海量的告警和复杂的链路中快速找到问题的根本原因，难以保证业务的 SLA，运维成本难控制。具体面临的挑战有以下几个方面：

1. 发现问题难

由于整体的业务数字化转型带来大量的新业务投放，而底层的架构在不断的被云化、虚拟化、容器化和微服务化，在这个过程中复杂的基础架构和上层业务应用的关系，各类组件会产生大量的机器数据，并且机器数据之间的关系错综复杂，如果没有进行数据治理的话，很难发现问题。

2. 告警风暴抑制难

告警风暴是指系统在短时间内发出海量告警消息的现象，这通常是由于系统出现了某种故障，导致产生的告警消息数远超运维人员所能处理的最大极限。如何对海量告警进行合理的分析、去重、合并和压缩也就成了运维过程的一大难点。

3. 根因定位难

随着业务不断发展，微服务架构越来越受到各大组织的青睐，随之也给传统运维带来更大的挑战，多维 KPI 指标数量繁多、相互之间关系复杂，要想在故障发生后，真正解决问题其实是一个复杂的根因定位的过程，目前缺乏一种综合根因定位的分析手段，使得根因定位难，问题解决困难。

4. 资源分析与评估难

在数字化转型的背景下，伴随着虚拟化、云计算和大数据技术的不断发展，各类业务系统上云、容器化后，如何保证计算资源、存储资源的合理利用；如何预知各类资源的使用趋势；如何进行各类

资源的容量规划等问题，对运维人员来说缺少分析判断的依据，难以进行分析和评估。

4.4.2 通过机器数据治理提升运维效能

在组织数字化转型的过程中，IT 运维管理面临的难点和挑战，需要通过机器数据的治理，满足多场景运维的要求，实现统一的运维分析与管理。面向机器数据的整合与治理，是以运维领域机器数据为基础，通过对多源的、分散的、海量的机器数据进行采集、整合、加工、清洗等操作，再通过机器学习、人工智能算法等技术对数据进行分析应用，以提高在 IT 运维管理的效能。

1. 将各类软、硬件系统的指标数据、调用链数据和日志数据进行全面的采集和存储，建立起统一的机器数据平台；
2. 通过缓存、多线程、schema on read、分布式部署等技术，更高效、更可靠的进行数据解析，使得数据可以很好应对各种场景下的运维需求，为后续数据利用打好基础；
3. 按照 IT 基础设施或者业务系统进行模型的建立，将各类机器数据标准化，打通指标数据、调用链数据和日志数据，使各类系统的机器数据形成关联关系；
4. 通过人工智能算法的应用，将各类机器数据进行加工和分析，实现告警降噪、故障定位、资源分析与评估、问题及时发现等能力。

通过将各类软、硬件系统的机器数据的统一采集、加工和治理，应用机器学习和人工智能算法建立起机器数据治理平台，有效提升运维效率，解决运维工作中的各类问题，主要包括：

- 及时发现问题，通过对多源的、分散的机器数据进行统一全面的采集、整合、关联、加工、治理等操作，再通过机器学习和人工智能技术对数据进行分析应用，及时的发现各种架构和环境下的故障，并且能够在故障发生前预测问题隐患。
- 告警降噪，分析告警的相似性、相关性和因果性，对海量告警进行去重、合并、抑制、压缩，只保留和显示核心的、有价值的告警信息。
- 根因定位，根据告警的时间点，以及各类软、硬件设备之间的关联性关系，通过算法将数据训练出不同的告警关系拓扑图，以辅助进行根因定位。
- 容量预测，在全面数据采集的基础上，采用人工智能、大数据技术，通过对历史数据的学习训练，对容量当前和将来使用、需求自动完成全面的分析。对于各类资源建设、管理、规划辅助实现全面的智能化。

4.4.3 机器数据治理对运维管理的价值

随着云计算、大数据等相关技术的广泛应用，对于数据中心建设、管理、规划等方面面临的运维问题也越来越多。通过机器数据的治理，打破数据孤岛、建立数据之间的关联关系，以统一的视角建立模型，利用机器学习和人工智能算法，帮助组织充分识别和挖掘机器数据的价值，提升组织运维效能，保障业务 SLA。其主要价值体现在：

- 通过降低故障处理时间，保障业务 SLA，降低业务风险。
- 通过资源分析及趋势预测，降低基础设施成本。
- 提升单位人力效能，降本增效。

4.5 助力业务运营

4.5.1 业务运营面临的挑战

数字化转型是技术与商业模式的深度融合，通过技术的创新应用驱动商业模式、业务模式的变革；商业模式、业务模式的变革又不断促进技术的创新与发展。伴随数字化转型进程的推进，业务呈现出移动化、敏捷化、规模化等特点，并更加注重客户体验。

随着商业模式、业务模式的变革，业务运营也面临新的挑战：

- 数据获取难：数据分散，存在于不同的系统，且数据格式不统一，难以有效利用
- 运营能力弱：对业务状态及趋势、用户体验等缺乏有效分析，难以对业务进行持续优化
- 业务驱动不足：缺乏数据支持，难以形成持续的业务跟踪与反馈，对业务发展的驱动力不足

4.5.2 通过机器数据治理助力业务运营

在组织数字化转型过程中，业务模式也在悄然发生变化，业务形态呈现移动化、敏捷化的特征，并越来越关注用户体验，这就要求组织能通过对数据的治理，加强业务运营分析能力，准确掌握业务的现状及对未来趋势的预判。面向机器数据的整合与治理，通过对多源、分散、海量的机器数据进行采集、整合与加工处理，通过即时数据处理能力，提炼出业务运营相关数据，通过人工智能技术对数据进行分析，助力业务运营的决策分析，持续提升用户体验。

1. 将与应用系统相关的各类软、硬件系统的指标数据、调用链数据和日志数据进行全面的采集和

存储，建立起统一的机器数据平台；

2. 采用缓存、多线程、schema on read、分布式部署等技术，通过即时、高效的数据解析能力，实现对数据的汇聚、存储及解析；
3. 建立数据分析模型，提取业务运营相关指标数据信息，如业务类型、业务量、客户分布、客户来源、业务渠道、用户操作等数据，为业务运营分析形成数据基础；
4. 通过人工智能技术应用，对提取的业务运营指标数据进行智能化分析，在客户营销、渠道优化、应用改善、提升用户体验等方面提供运营分析，实现业务洞察，驱动业务发展。

首先，通过对机器数据的治理，将分散的各类数据进行采集和接收，实现对数据统一、集中的治理，并运用技术对异构数据进行格式化、标准化；

其次，通过对机器数据的治理，可以充分挖掘数据中的业务价值，主要体现在：

- 客户维度，通过对操作终端的识别，用户操作行为记录，用户属地，基本特征等信息的综合分析，可辅助生成客户画像，为客户提供个性化推荐和服务，提升客户体验。
- 业务维度，通过追踪用户访问渠道，对用户操作行为、业务流量、业务结果反馈、客户体验过程等进行综合分析，可帮助业务运营人员进行渠道优化、业务引流、产品规划布局、提升内容交互体验等，加强对业务的价值输出。
- 应用维度，通过对应用系统的运行状态及健康度进行实时的分析和评估，从业务视角围绕用户体验对应用系统进行持续的优化，不断提升产品竞争力。

4.5.3 机器数据治理对业务运营的价值

数字化转型期间，大量应用了云、大数据等相关技术，并实现业务的快速迭代与上线，发展过程中，业务应用的访问状态、业务健康状态、业务资源的度量等均需要可观测，可分析。通过机器数据治理，帮助组织充分识别和挖掘机器数据业务价值，助力数字化变革。其价值主要体现在：

- 提升运营分析能力
 - 打破数据孤岛，实现对机器数据的价值链关联
 - 业务运营数据提取与分析，实现对业务的精细化管理，为业务规划与决策提供数据支撑；
 - 持续的数据治理，以数据驱动运营

- 提升用户体验
 - 通过提升应用的健康度和操作的便利性，提升用户体验；
 - 辅助客户画像，通过个性化推荐和服务，提升用户体验；
 - 持续跟踪用户服务过程和服务反馈，改进用户体验；
- 驱动业务健康发展
 - 围绕用户体验，对业务应用进行持续优化和完善，提高用户留存
 - 精细化的运营管理，通过数据支持决策，确保业务方向正确
 - 持续的业务健康体检，促进业务蓬勃发展

4.6 机器数据治理赋能组织价值

随着数字化转型的深入，管理者越来越重视对数据资产的开发和利用，希望通过对机器数据的治理，将海量、分散、异构的数据进行整合治理，实现数据价值开发。

- 在故障管理上，通过对数据的智能分析，准确识别故障与根因定位，降低 MTTR，减少系统故障对业务带来的经济损失；
- 在资源使用上，通过对容量的监测及分析，实现资源的合理化分配，避免资源浪费，有效控制并降低成本支出；
- 在业务运营上，通过对机器数据的洞察与分析，实现业务运营分析，辅助业务运营决策，推动业务持续健康发展；
- 在合规与审计上，遵从法规要求，加强数据管理合规性，提升数据审计的效率、质量、频率；

机器数据每时每刻都在大量生成，然而机器数据的价值确未能被充分识别和利用，究其原因主要包括以下几个方面：

- 数据量大且分散，独立存在于各个系统中；
- 缺乏有效的工具对各类机器数据进行采集；
- 各类机器数据格式差异大，缺乏标准化，难以对数据进行解析；
- 缺乏有效的技术手段对机器数据进行分析应用；

为提升机器数据治理 ROI，释放机器数据价值，结合当前机器数据现状，组织需结合业务战略及目标，加强技术探索及应用，在机器数据的采集、治理及应用上进行体系及能力建设，实现对机器数据端到端的全生命周期管理，打造数据驱动型组织。

- 明确机器数据管理目标，制定机器数据管理制度和流程
- 优化、完善监控手段，加强对机器数据的采集能力
- 建设机器数据治理平台，具备对海量实时数据的即时数据处理能力
- 通过对数据的整合、清洗及标准化，推进数据资产化
- 对数据进行统一规划，建立数据的分类管理及指标体系
- 遵从组织数据安全规范，加强数据安全管控
- 加强智能化能力建设，促进在合规与审计、运维管理、业务运营等方面的数据消费场景应用

第 5 章 全域数据治理加速数据驱动

5.1 从数据治理转变为业务洞察力

经过全域数据治理后，提升了数据质量和可访问性。让数据产生更大的价值，还需要重视从数据治理转变为业务洞察力。从历史上看，数据治理一直关注数据安全、保护、管理和控制，以及支持合规性要求所需的流程。与今天的其他一切类似，数据治理正在发生转变。当然，它将继续做好原来的数据治理需求并推行数据质量标准，但其主要重点将转向致力于确保用于分析功能的数据优化，在数据优化的实践中为组织提供价值。

随着组织继续高速发展，拥有前所未有的速度获取和存储大量数据的能力，数据增长和复杂性的上升逐年加速。对于所有这些数据，确保其准确性、可发现和可访问性是数据质量流程的核心，也是分析流程的成功关键，旨在向组织股东和高级管理者提供业务洞察力。

庞大的数据量以及如何驱动数据价值可能会令人乏力，所以业务洞察力最好从高质量的数据中创建，这些数据可以很容易地被授权用户及时引用。然后，它会变得像承诺、支持和专注于实现关键业务功能的最高质量数据一样容易。提高这些关键数据资产的质量首先要对业务具有最大影响的数据进行优先排序，并制定一个和内容一致的参考数据标准，并在组织内的不同业务线中使用。

只要关键系统是井然有序的，数据质量问题就必须跨越多个组织中的数据源。但要注意并非所有数据都是平等对待——数字化环境中缺乏拥有者的数据、处于休眠状态多年或已被其他记录系统取代的数据是从被移除的主要选择，余下的部分都应该与业务有切实的相关性，并符合数据质量标准。

分公司、部门和业务线独立运营已逐渐为既成事实——这是支持敏捷性和快速响应不断变化的业务需求的必要条件。通过跨部门和组织边界管理实现治理高质量数据的一致性，同时实现从数据治理到业务洞察力的转变。

仅靠技术无法解决数据质量问题，再多的机器学习也无法在半夜神奇地清理人工生成的数据。人们必须参与数据治理流程——他们必须参与其中并牢记管理责任。全域数据的治理只有结合先进的技术，支持正确的流程，由个人及业务驱动，才能真正将受治理的数据转变为具体的业务洞察力。

5.2 为 CXO 提供数据治理的支持

为 CXO 提供数据治理支持始于 CXO 为数据治理提供支持和保障。数据治理首先需要公司管理层支持，并注重组织建设，即一把手挂帅，保证组织有执行力。

组织数据治理策略必须在 CXO 级别建立和驱动，具有领导支持自上而下的治理方法，这种支持方式可说服不同意见利益相关者，可成为在组织范围内让多方达成共识的必要力量。是否具有强大的 CXO 支持，这种差别可在一个混乱的，不可调和的数据治理组织与一个在所有领域都一致性的实施平顺数据治理的组织之间清晰可见。

避免使用最高管理层的行政权力来要求每个人都必须通过试图扩大规模或在所有领域同时实施治理计划，这样有可能造成混乱。最好的方法是采用迭代，并为管理者提供数据的更多控制权，以便对当前的挑战或机遇采取明智的行动。

数据治理计划离不开背后的管理者支持，同时也能为管理者带来价值。数据治理为执行管理层提供的一些价值包括：

- **更好的决策**：在组织过程和决策中，管理良好的数据更容易被发现，使相关方更容易获得有用的见解。意味着决策将基于正确的数据，确保更高的准确性和整体信任。
- **更高的运营效率**：作为业务的一项资产，数据治理流程不断验证数据的适用性和功能及其对业务的参考价值，帮助这些管理者所负责的部门做出决策。那些与部门特定相关决策制定也可依赖同类完整和准确的数据。
- **提高对数据的理解和关系**：可以向管理者承诺，治理过程可以指出他们拥有哪些数据、存储在哪里、如何使用、谁在使用它以及谁对它负责。这种理解将提供转化为对审计的及时响应、更有效的业务复盘活动以及防止数据损坏和泄露而采用更积极主动的方法。
- **监管合规性**：无论是内部监管、行业或垂直监管，还是由政府监管，数据治理都可确保和证明组织与监管要求中规定的规则保持一致。
- **商业价值**：提高运营效率和全域数据治理所有价值都会带来另一个好的效果，通过更快的决策和更确定的行动来帮助改善业务，提高收入。

全域数据治理为 CXO 提供较多价值，真正执行落地时需要管理层达成统一认识全域数据治理是成为真正的数据驱动型组织必经之路。

5.3 数字生态建设

全域数据治理还需要重视数字生态建设。大部分组织倾向于选择生态丰富、能力领先的外部供应商，在数据治理方面建立长期战略合作伙伴关系。数字化生态系统，是指包含一系列基于标准，规模可变的硬件、软件、数字设备和服务，可系统地实现组织数字化，数据交流，以帮助组织提高运营效

率。它是一组互连的信息技术资源，可以作为一个整体来工作。例如，数字化生态系统可以由供应商、客户、贸易伙伴和第三方数据服务商组成。

数字化生态可重构产业链关系，优化资源配置，形成网络结构下的短链模式，以高效的业务协同、数据协同、要素协同，实现价值共创、利益共享。对于产业链上游，数字化生态助力组织实现智能化供给，保证产业链供给安全；借助数字化生态下各组织的知识共享、技术共享、产能共享、订单共享、员工共享等多种数字化模式创新，实现优化资源配置，提升竞争优势；对于产业链下游，数字化生态打破生产与终端用户相互割裂的状态，将生产、商业、消费、社交有机融合，借助数据分析和治理优化产品设计、产能投放，精准匹配用户需求，提高有效供给。

要实现数字化生态必须要充分发挥海量数据和丰富应用场景优势，促进数字技术与组织业务模式深度融合，赋能组织的传统业务模式转型升级，催生新产业、新业态和新模式。在组织进行海量数据和丰富的应用场景进行有效优化提升和最大价值挖掘的过程中，需要整合专业技术与专业服务能力的生态伙伴深度参与到数据治理活动中来，并将汲取各行业的最佳实践、先进技术为组织所用。数据治理是长期且系统工程，通过建立长期的伙伴关系，可以彼此深入业务，共同面对挑战，更有效的实现数字化生态的建设。

5.4 数据驱动文化打造

我们认为数据驱动文化是组织全域数据治理能否持续的关键因素。数据驱动文化（Data-driven Culture）的定义非常简单，其实就是通过数据驱动决策（Data-driven Decision Making，简称 DDDM），即凭借数据事实或数据模型来制定决策或战略计划。但正是这看似简单的理论，却能给组织带来巨大的收益：

- Forrester 的调研则表明，那些通过分析数据来获取洞见并且做决策的组织，其增长速度比那些拍脑袋做决定的公司要快整整 8 倍。
- MIT 的教授 Erik Brynjolfsson 发现着重于数据驱动文化发展的公司，其盈利情况比那些拍脑袋做决定的公司要高 5-6%。

嗅觉灵敏的顶级组织已经清晰地意识到，数据治理不再仅仅是一项成本支出，而是能带来巨额利润的生产力，不仅是业务和利润的增长，IDC 的研究调查表明，拥有数据驱动文化的公司，在客户满意度，产品上市时间，工作效率和利润上都遥遥领先。很多组织已经意识到数据中蕴含的价值，并愿意为了挖掘这些数据价值而做出各种努力和尝试。但由于不知道搭建“数据驱动文化”的完整路径，

难免走入弯路，时常陷入迷茫。付出大量金钱得到的经验教训让人们意识到真正打造一个拥有数据驱动文化的组织确实不是一件容易的事。事实上，只有 8% 的组织成功完成了向数据驱动文化的转型。

打造了优秀的数据驱动文化的组织都具有以下特征：

- 组织所有员工都重视数据的重要性；
- 各部门的数据透明，无缝协作；
- 组织员工都能在任何需要时无障碍地获得数据支持；
- 组织数据处理响应速度快（数据的即时治理能力强）；
- 会通过数据分析和治理探索商机，优化公司现状；
- 通过深度挖掘数据，获取创新 idea；
- 数据是组织战略的核心，各部门基于数据制定决策；
- 注重数据隐私条款和使用条款。

如何打造一家拥有优秀数据驱动文化的组织，以下是关键的必经之路：

» 数据成熟

规范组织的数据质量，有成熟的数据管理体系和统一的数据库，这是数据驱动文化的基石。

» 打造数据驱动领导力

数据驱动领导力是一个非常重要的部分，因为组织的改变以及文化的转型，不是一两个部门或者仅靠基层员工的努力就能推动的。组织的决策层和管理层才是推动组织文化发展的核心。只有当决策层在制定组织的战略时，坚定地将建设数据驱动文化列入其中，并且管理层把数据视为公司的重要资产，并且鼓励、推动员工依赖数据做决策，才可能真正迎来组织层面的变革。

» 培养组织内的数字化人才

组织需要既懂业务又懂数据的人才，这种类型的复合型人才非常急缺，在无法招聘到合适人才的情况下只能自己培养。拥有数字化人才，意味着组织内关键人员都具有了一定的数据理解和分析能力。该阶段的组织，会通过下面的形式进行业务决策：

- 当某个团队遇到问题、面临抉择的时候，大家不会急着拍脑门做出决定或发表见解，而是有意识地想要去找寻可以支持决定的数据。
- 团队内所有人都有“在极短的时间内”就能获取这些数据的方法。

» 基于数据的决策流程

组织的数据底层建设很难一步到位，覆盖到所有方面。当决策流程中的每一步都需要依赖数据分析的时候，就需要更加多元的数据来支持。可能原有的数据维度已经不能满足所有的业务需求，需要回到数据成熟的基石中进一步去扩充、完善数据体系。

所以打造数据驱动文化不是一蹴而就，而是螺旋上升的过程。随着发展进程，业务的需求会反向促进数据底层的建设。

第6章 全域数据治理的十大关键

组织出于各方面原因采用全域数据治理方式，选取一种有效且适用于组织的治理模型，可以使数据在访问性、可信度、可理解性及保持活跃度得到提升，将会带来以下十大关键价值：

» 对于监管和法规遵从，全域数据治理能力非常关键

全域数据治理通过建立和遵循数据质量标准，降低了风险以及不正确决策（基于错误数据）所带来的意外成本，同时也有效确保了严格遵守监管和法规遵从的规范。

» 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。

» 领域认知智能保障 IT 与业务对治理的理解保持一致

通过领域认知智能驱动构建各领域知识图谱，让 IT 部门快速理解业务，主动发现业务问题，并与业务部门保持在同一个频道，彻底解决数据治理痛点难点问题。

» 即时数据治理保持数据一致性确保了完整性和精确性

这是可信赖的数据基础，也是不断改进流程模型、数据类别定义、数据分类和组织定义的基础，所有业务决策都将基于这些定义。

» 分析和决策基于定义明确且准确的数据

全域数据治理指导数据通过数据链条的结构和流动，尤其是在分析过程中。治理确保组织数据获取机制设置为收集需要的数据，并在业务线的策略和更大组织的战略目标之间达成一致性。

» 基于事实的实时决策贯穿整个组织

全域数据治理是确保数据真实性的关键，这反过来又为数据用户建立了实现决策实时目标所需的信心。

» 数据信心鼓励分享见解

拥有以数据驱动为中心的组织文化只有在数据完整准确的情况下才能鼓励共享信息和见解。全域数据治理可对业务带来很多帮助，因为治理过程需要整个组织分享来自不同地方数据和见解。

» 全域数据治理促进协作和建立问责制

根据标准化规则或策略来管理和控制数据的使用和适当维护，可以消除系统中的低效率，促进业务部门之间的协作，并更大程度建立数据所有者问责制。

» 数据基于其参考价值保持干净和相关性

存储的数据质量与业务相关性非常重要。全域数据治理方法使每个人都是数据的管理者，负责保持数据的良好状态。

» 全域治理的数据为组织提供竞争优势

整个组织中，无论数据是集中还是分散，当它得到有效的管理和控制时，获得有价值的赋能、驱动运营和创新的过程都更容易实现。



上海爱数信息技术股份有限公司

AISHU Technology Corp.

总部地址：上海市联航路 1188 号浦江智谷 8 号楼 2 层

邮 编：201112

咨询热线：021-54222601

服务热线：400-880-1569

传 真：021-54222601-8800

客服邮箱：support@aishu.cn

www.aishu.cn



爱数微信公众号