ENTERPRISE

DATA SPACE

《数据空间探索与实践》编写组 ◎ 著

数据空间探索与实践

华为公司数据总架构师、数据首席专家**马运领衔撰写**

华为公司董事、质量与流程 IT 总裁、CIO **陶景文作序推荐**

机械工业出版社
CHINA MACHINE PRESS

内容简介

这是一部系统梳理和总结华为在数据空间领域的研究成果与实践经验的著作，全面讲解了如何构建可信、可控、可证的数据流通体系，旨在为跨主体、跨边界数据共享这一难题提供完整的解决方案。

华为致力于数据空间理论与架构的研究近8年，积极参与国际数据空间协会（IDSA）、 Gaia-X等国际组织的生态共建与标准制定，并与欧洲的相关科研机构在研究项目上开展了大量的合作。华为成功研发了自己的数据空间服务产品EDS（交换数据空间），并在鲲鹏昇腾生态率先得到验证，如今已经在华为内部的4大类20多个场景和外部的大量客户中得到广泛应用。在此基础上，华为将自己在数据空间理论框架、工程实践、落地原则等方面的认知、经验、教训凝结成了这本书。

本书将为您揭示数据流通的破局之道，从理论到实践、从架构到落地，全面解析数据空间这一创新方案。阅读本书，您将收获：

（1）一套完整的数据空间知识体系：全书共11章，系统覆盖了数据流通的时代背景与世界性难题、欧洲数据空间的最新实践、数据空间的4大设计原则和五层三维参考架构、数据连接器的6大关键能力，以及如何构建“可信、可控、可证”的数据流通保障机制。

（2）一套可落地的企业实践方法：详细阐述了企业在部署和运营数据空间时需要解答的3个关键问题（谁来用、谁来管、怎么用），并提供了4大类企业应用场景的实操指南，涵盖企业内部、集团内、产业链上下游以及与外部组织间的数据交换模式。

（3）一套面向未来的前瞻性思考：深入探讨了在AI大模型时代，数据空间如何应对数据流通的新挑战，以及如何支撑AI场景下的数据可控利用，展望了数据空间成为未来国家数据要素流通基础设施的宏伟蓝图。

本书将为您的数据战略规划和数字化转型实践提供极具价值的参考与指引，帮助您在数据浪潮中占得先机。

数据空间探索与实践

《数据空间探索与实践》编写组 ◎ 著

ENTERPRISE DATA SPACE

机械工业出版社
CHINA MACHINE PRESS

北京市版权局著作权合同登记　图字：01-2025-3625

图书在版编目（CIP）数据

数据空间探索与实践 /《数据空间探索与实践》编写组著. -- 北京　机械工业出版社，2025. 11.
ISBN 978-7-111-79619-0

Ⅰ. TP274

中国国家版本馆 CIP 数据核字第 2025ZE5306 号

机械工业出版社（北京市百万庄大街 22 号　邮政编码 100037）
策划编辑：杨福川　　责任编辑：杨福川　章承林
责任校对：张勤思　马荣华　景　飞　　责任印制：单爱军
北京联兴盛业印刷股份有限公司印刷
2025 年 11 月第 1 版第 1 次印刷
147mm × 210mm · 9.25 印张 · 3 插页 · 230 千字
标准书号：ISBN 978-7-111-79619-0
定价：99.00 元

电话服务　　网络服务
客服电话：010-88361066　机　工　官　网：www.cmpbook.com
　　010-88379833　机　工　官　博：weibo.com/cmp1952
　　010-68326294　金　　书　　网：www.golden-book.com
封底无防伪标均为盗版　机工教育服务网：www.cmpedu.com

序

以信任构筑基石，让数据释放新质生产力

我们正处在一个由数据驱动深刻变革的时代。数据作为继土地、劳动力、资本、技术之后的第五大生产要素，正成为全球数字经济发展的核心引擎。然而，与传统要素不同，数据具有“易复制、非排他、确权难”的独特属性，这使得其在流通、共享与交易的过程中，天然存在着一道“信任”的鸿沟。如何跨越这道鸿沟，让数据在安全可信的环境中自由流动，激发其“新质生产力”的倍增效应，是摆在所有数字化转型企业乃至整个社会面前的世界级难题。

华为作为一家业务遍及全球的科技公司，在自身的数字化转型历程中，深刻体验了数据从“集成”到“汇聚”，再到“交换”的演进。华为很早就意识到，要真正实现数据的价值最大化，就必须解决数据在跨主体、跨边界流通过程中的信任问题。这不仅是一项技术挑战，还是一场围绕“数据主权”的认知与实践变革。

正是基于这样的思考与探索，华为从 2016 年便开始关注和布局企业数据空间，致力于解决“数据主权保护”和“要素流转分享”这一核心矛盾。华为打造了数据空间服务，并围绕它制定了

32 项针对数据主权的保护细则，并通过有效的技术手段，将这些规则落到实处。无论是支撑鸿蒙和鲲鹏生态上下游伙伴的数据高效流转，还是应对 GDPR 要求下的跨境数据合规，抑或是解决财务外审场景中的底稿安全共享问题，这些年来，华为已在超过 20 个场景中成功应用了数据空间。这本书正是对这些实践的全面复盘与深入剖析，系统性地阐述了华为如何通过“数据连接器”这一核心组件，以及“可信、可控、可证”的技术框架，将抽象的“数据主权”理念，转化为具体、可落地、可运营的能力体系。

尤其是在当下，人工智能大模型正作为一件“超级工具”，推动着千行百业的深刻变革。我曾多次强调，数据的多样性和质量，决定了未来大模型差异化和竞争力的边界。然而，机遇与风险并存。大模型强大的学习能力，也可能在不经意间将企业的核心数字资产“卷走”。如何将企业需要保护的数据与通识数据有效结合？如何既利用大模型的能力，又保障自身的数据主权？这正是华为现在正在研究和探索的前沿方向。

与此同时，全球范围内的数据战略布局也在加速。欧盟的《数据法案》已正式生效，明确将数据空间定位为数据流通的基础设施。我国也成立了国家数据局，并发布《“数据要素 ×”三年行动计划（2024—2026 年）》，旨在将海量数据优势转化为国家竞争新优势。在这样的时代背景下，数据空间的建设已从企业的“选答题”，上升为行业乃至国家的“必答题”。

顺应这一趋势，华为的数据空间实践也在不断深化。华为持续深化与全球伙伴的合作，积极参与国际数据空间协会（IDSA）、Gaia-X 等国际组织的生态共建与标准制定，并获得了由国际数据空间协会颁发的认证。华为云发布的交换数据空间（EDS）服务，

秉持着“用我们卖的，卖我们用的”原则，已经将这些经过内部千锤百炼的能力，转化为服务千行百业的公共产品，助力更多企业和组织快速构建属于自己的数据流通信任体系。

我们乐于将这些探索与实践结集成书，并公之于众。我们希望它不仅能为正在投身于数字化、智能化建设的领导者、设计者与实施者们提供一份有价值的参考，更能抛砖引玉，引发业界对数据流通与数据主权更深层次的思考和讨论。

数据空间的未来，是一片广袤的“星辰大海”。从企业到行业，从城市到国家乃至全球，一个个独立的数据空间将通过标准化的“连接器”互联互通，最终构筑起一条安全、普惠、共赢的“数字丝绸之路”。华为愿与各界同仁携手，集众智、聚众力，共同为这个宏伟目标的实现贡献力量，打造一个真正繁荣、可信的数据新生态。

陶景文

华为公司董事、质量与流程 IT 总裁、CIO

前　言

为何写作本书

作为国内最早加入 IDSA（International Data Spaces Association，国际数据空间协会）的企业会员，华为一直致力于数据空间理论与架构的研究工作。华为参与了多个版本的 IDS 参考架构的讨论和制定。通过与欧洲相关科研机构在研究项目上的合作，例如参与 Report of Data Sovereignty with IDS and Gaia-X 和 Data Sovereignty and Privacy based on IDSA Standardization 等国际合作项目，并基于自身多年的数据治理实践，华为逐渐摸索出适合自身业务特点的数据空间应用理论、架构，并明确了落地原则和要点。

早在 2018 年 12 月，华为多个领域的数据部门就与德国弗朗霍夫研究院相关单位进行了专业研讨，讨论如何基于 IDS 架构与技术框架，为跨企业边界的数据共享和使用控制提供合适的解决方案；如何应用数据空间相关技术以实现数据共享；如何在华为内部实现对外部利益关系人数据的使用控制；以及实施数据治理的几种模式，并针对这些模式给出了场景描述。同时，双方就数据治理在

技术层面和组织层面的意义也展开了详细的讨论。

经过系统性的理论与技术可行性研究，从 2020 年开始，华为启动了 EDS（交换数据空间）原型设计并开始独立自主开发产品，参考 IDS 的架构设计，打造了数据空间的核心组件连接器 Connector, 并在 Connector 中实现了对 13 条使用控制策略的支持。2021 年 4 月，Connector 成功实现鲲鹏昇腾生态数据交换，并满足了 13 家华为伙伴的数据诉求。

经过首个业务场景的验证，我们发现华为在内外部数据交换方面存在众多类似的场景。2022 年 12 月，公司决定成立主权可控数据交换变革项目，旨在通过数据空间服务华为内部业务，定规范、立标准、建能力，并实现以下目标：保护数据主权、消除数据孤岛、促进数据流通、增加数据价值。截至 2024 年，EDS 已在华为内部的 4 大类 20 多个场景中得到应用。

随着在华为内部业务场景中的成熟应用，在公司倡导“用我们卖的，卖我们用的”理念下，EDS 开始内服外化。2023 年，华为成立了联合研发团队，并在 9 月份完成了华为云商用版本的发布。同时，为了满足客户差异化的诉求，华为云联合华为存储产品线团队预研线下一体机，提供更高级别的安全可控交换产品。此外，华为一直积极保持与欧洲的互联互通，基于 EDS 的华为数据空间方案已成为 Gaia-X 灯塔项目的基础设施和技术代表，华为云 EDS 获得了 IDSA 的测试认证，并在国内通过了信息通信研究院可信数据空间产品功能评估。

基于上述探索研究和实践历程，结合数据成为生产要素的大背景，我们认为华为面临的数据流通挑战也是业界很多组织将要面对的，于是我们将华为在数据空间领域的探索与实践总结成了本书，

希望对推动组织数字化、智能化的领导者、设计者、实施者和参与数据流通的同行有一定的启发和借鉴意义。

本书主要内容

全书共 11 章，主要内容如下。

第 1 ～ 3 章主要阐述数据空间的时代背景，说明数据成为生产要素给流通带来的挑战，以及华为如何尝试解决这些难题。第 1 章以释放数据价值为最终目标，结合数据成为生产要素的大背景，阐述数据在企业中的定位及流通模式的变化，介绍了在新形势下企业数据流通所面临的挑战。第 2 章从数据要素的特征出发，梳理各个特性所带来的数据流通挑战，进一步说明为什么数据流通是个世界级难题。第 3 章基于数据流通难题，详细阐述欧洲尝试解决这类难题，并引入数据空间理论的过程。

第 4 ～ 7 章介绍数据空间的理论，以及华为针对数据空间从底层逻辑、总体设计、数据连接器和流通体系 4 个方面做出的符合华为特色的变革。第 4 章重点介绍在引入数据空间理论后，华为结合自身特点给出的符合华为自身需求的数据空间底层逻辑。第 5 章提出华为数据空间总体设计，明确以数据可控交换为核心环节，阐述数据从接入、交换到消费的关键业务设计的过程。第 6 章深入数据空间，详细阐述数据连接器的核心作用，明确数据交换参与方可通过数据连接器实现数据交换的留痕、处理及使用控制。第 7 章 基于数据空间的管理目标，提出构建“可信、可控、可证”的数据流通体系，从而通过服务化与模板化对数据流通实现有效保障。

第 8 ～ 9 章重点介绍数据空间在企业落地中必备的流程、规范和

基于实际业务场景的实践。第 8 章基于数据空间所构建的能力，从流程和规范的维度介绍企业如何调用数据空间的能力，规范地使用和运营数据空间。第 9 章通过对众多实际使用场景的归纳，重点基于 5 个案例介绍通过数据空间进行数据交换的背景、实施方案和价值。

第 10 ～ 11 章主要涉及 AI 大模型和对数据空间未来的展望。第 10 章基于 AI 大模型对数据的需求和使用过程中的挑战，探索数据空间对 AI 数据集不同程度的控制方式，推演落地的可行性。第 11 章基于企业内的实践总结，结合数据空间作为数据流通基础设施的定位，通过对外部环境的解读，畅想数据空间在行业、城市、国家、国际层面的未来建设趋势，让数据流通在互信的基础上再无阻力，从而充分释放数据价值。

致谢

衷心感谢《数据空间探索与实践》编写组全体成员的辛勤付出，特别感谢马运（MA YUN）、杨雯、韦冬、苏本胜、田璞等同志的重要贡献。在数据空间项目的实施以及本书的编写过程中，我们得到了华为相关领域的多位领导、专家和同事的指导、支持与帮助，在这里深表谢意。

我们希望通过本书分享华为在这一领域的探索与实践经验，以帮助大家加深对数据空间的理解，并共同推动数据空间发展。但是，数据空间在国内还是一个新概念，我们虽然取得了一些成果，但这些成果主要局限于华为的业务场景，难免存在不足之处，恳请读者批评指正。

《数据空间探索与实践》编写组

目　录

序

前言

第 1 章　数据流通释放价值

1.1　释放数据要素价值的改革全面铺开　2
1.2　企业数据的价值实现　4
1.2.1　沿着业务流的数据清洁　4
1.2.2　脱离业务流的数据共享　7
1.2.3　跨越主体的数据交换　10
1.3　企业数据定位的变化　14
1.3.1　数据是 IT 的一部分　15
1.3.2　数据作为企业战略资产　15
1.3.3　数据成为生产要素　18
1.4　企业数据流通模式的演进　20
1.4.1　数据流通从集成、汇聚到交换　20
1.4.2　数据交换的关键：跨主体边界后的可控　23
1.5　小结　26

第 2 章　数据流通的世界级难题

2.1　数据要素的独特性给流通带来新难题　28

2.1.1 难以管控数据使用 29
2.1.2 增加数据使用风险 31
2.2 数据主体的利益保障和风险控制 31
2.2.1 商业利益 32
2.2.2 法律合规遵从 34
2.2.3 技术保障 35
2.3 世界主要经济体的应对战略及举措 36
2.4 小结 44

第 3 章 数据流通难题的欧洲实践

3.1 数据空间的由来 46
3.1.1 从“工业数据空间”到“国际数据空间” 46
3.1.2 数据空间的基本理念：保护“数据主权” 47
3.2 数据空间的 4 个设计原则 49
3.2.1 你决定你的数据如何被使用 49
3.2.2 以数据和服务质量为基础的公平竞争环境 50
3.2.3 去中心化、技术中立的基础架构 50
3.2.4 公共与私营组织共同治理 51
3.3 五层三维的数据空间参考架构 52
3.3.1 业务层：定义数据空间的 4 类参与方及协作关系 52
3.3.2 功能层：数据空间的 6 大能力特性 55
3.3.3 流程层：定义数据交换的 5 大步骤 56
3.3.4 信息层：数据空间的六边形模型 58
3.3.5 系统层：以连接器为核心的 6 大组件 60
3.3.6 安全维度：基于最佳实践的 6 个安全设计 62
3.3.7 认证维度：个人、组织与组件的安全合规 63
3.3.8 治理维度：以“联邦”的形式实现数据空间的安全可信交互 64
3.4 数据空间已从理论走向实践 64
3.4.1 理论体系已经相对完善 64
3.4.2 数据空间生态已经成型 65
3.4.3 领先数字企业布局数据空间配套服务 66
3.4.4 初具规模的业务应用 66

3.5 数据空间写入《欧洲数据战略》及《数据法案》 67
3.5.1 《欧洲数据战略》：建立 9 大数据空间 67
3.5.2 《数据法案》：将数据空间定位为基础设施和数据治理机制 68
3.6 小结 69

第 4 章 数据空间的底层逻辑

4.1 数据流通面临的现实诉求 72
4.1.1 组织内部门间的敏感数据获取 74
4.1.2 集团与子公司、事业群之间的可控数据共享 74
4.1.3 华为与外部企业的数据交换 74
4.1.4 华为与生态链上下游的数据协作 75
4.2 在数字世界履行数据交换合约 75
4.2.1 现实世界的数据交换合约通过法律执行有滞后性 75
4.2.2 基于数据合约的控制技术履行合同条款 77
4.3 数据空间的基本特征 79
4.3.1 参与主体通过数据连接器交换和使用数据 81
4.3.2 具备运营清算能力 81
4.3.3 以可清算范围界定数据空间边界 82
4.4 数据空间中的数据合约 83
4.4.1 可在线协商的数据合约 85
4.4.2 可标识的数据合约 86
4.4.3 基于规则进行履行验收的数据合约 87
4.4.4 覆盖数据全生命周期的数据合约 89
4.5 数据空间中的数据策略 90
4.5.1 从访问控制到使用控制 90
4.5.2 数据使用控制策略是跨边界的控制 91
4.5.3 多维可组合的“4W2H”策略描述模型 92
4.6 小结 94

第 5 章 数据空间的总体设计

5.1 数据空间的工程能力要求 95

5.1.1 多模态数据按需灵活接入 97
5.1.2 满足不同运营要求的 4 种交换模式 98
5.1.3 人、系统与设备均可按策略消费数据 99
5.1.4 兼容现有的数字化环境进行部署 101
5.2 数据空间中的参与角色 102
5.2.1 保障数据空间可持续运转的清算运营方 103
5.2.2 供应数据资源的数据提供方 104
5.2.3 获取数据价值的数据消费方 105
5.2.4 为数据增值的数据服务方 106
5.3 数据空间中面向交换的数据形态 107
5.3.1 符合交换要求的资源成为数据交换资源 109
5.3.2 数据交换单元是传输、验收、使用和清算的最小单元 111
5.3.3 数据 Offer 是由交换单元组成的数据合约标的物 112
5.3.4 供需双方共同达成的数据交换合约 113
5.4 数据空间中的交换业务流 114
5.4.1 提供与控制数据的业务流 117
5.4.2 获取和使用数据的业务流 118
5.4.3 处理与加工数据的业务流 119
5.4.4 清算与运营的业务流 120
5.5 数据空间的核心能力组件 122
5.5.1 交换数据及执行合约的连接器能力 123
5.5.2 实现组织、个人及组件认证的可信能力 124
5.5.3 决定数据如何被使用的可控能力 125
5.5.4 全链路追溯清算的可证能力 125
5.6 小结 126

第 6 章 数据空间中的数据连接器

6.1 数据连接器的关键能力特性 128
6.1.1 数据资源接入 128
6.1.2 数据资源管理 132
6.1.3 数据合约管理 136

6.1.4 数据传输和传输协议 138
6.1.5 数据消费和 App 管理 145
6.1.6 数据连接器的配置与监控运营 149
6.2 数据连接器的分层解耦架构 152
6.2.1 数据连接器的业务作业面与数据处理面 153
6.2.2 数据处理面数据集成服务的解耦架构 155
6.2.3 数据连接器南向环境资源解耦架构 158
6.2.4 数据连接器与数据空间中组件的交互关系 160
6.3 适应多场景、多环境的数据连接器 163
6.3.1 数据连接器的三种使用场景 163
6.3.2 数据连接器的三种安全级别形态 164
6.3.3 数据连接器的三种部署模式 167
6.4 小结 170

第 7 章 数据空间中的可信、可控、可证

7.1 什么是可信、可控、可证 171
7.2 基于数据空间信任框架保障可信 173
7.2.1 数据空间的信任框架 173
7.2.2 空间组织及人员可信 175
7.2.3 数据服务及组件可信 178
7.2.4 平台及环境资源可信 180
7.3 基于开放数字权利语言的可控策略实现 182
7.3.1 标准化、机器可执行的控制策略 182
7.3.2 控制策略的实现机制：定义、执行与决策 188
7.3.3 面向业务场景的服务化与模板化能力 191
7.4 基于区块链实现数据交换的可证追溯与清算 192
7.4.1 实现数据合约履行的透明可证 192
7.4.2 基于数字探针端到端的实时记录 193
7.4.3 探针数据上链防篡改 194
7.4.4 标准化的查证与清算服务 195
7.5 小结 198

第 8 章 数据空间的部署与运营

8.1 启动数据空间的 3 个关键问题 200
8.1.1 谁来用：使用场景的识别 200
8.1.2 谁来管：明确运营主体及职责 202
8.1.3 怎么用：被调用并嵌入现有业务流程 203
8.2 数据空间装载与运营 204
8.2.1 如何开通数据空间 204
8.2.2 如何在数据空间中交换数据 207
8.2.3 如何运营数据空间 213
8.3 小结 216

第 9 章 数据空间的企业应用

9.1 企业应用数据空间的四类场景 220
9.1.1 企业内各业务部门间的数据交换 222
9.1.2 企业总部与分公司间的数据交换 224
9.1.3 企业与生态链上下游间的数据交换 227
9.1.4 企业与外部组织间的数据交换 231
9.2 部署数据空间的三类典型范式 237
9.2.1 范式 1：1∶1 模式 238
9.2.2 范式 2：1∶N 模式 239
9.2.3 范式 3：N∶N 模式 240
9.3 设计业务模板以满足场景复用需求 241
9.4 小结 244

第 10 章 数据空间与 AI 大模型

10.1 AI 大模型时代数据既丰富又稀缺 246
10.1.1 数据成为驱动 AI 大模型发展的核心要素 246
10.1.2 激活非结构化数据资产的潜在价值 247
10.1.3 高质量数据集成为稀缺资源 249
10.2 AI 大模型时代数据流通的挑战 250

10.2.1 AI 大模型需要数据可用又可见 250
10.2.2 大规模高质量数据流通安全风险更高 251
10.2.3 数据形态变化导致权限失效举证更难 251
10.3 探索 AI 大模型的数据控制 254
10.3.1 控制 AI 大模型的数据读取 256
10.3.2 控制 AI 大模型的数据读取与加工 258
10.3.3 控制 AI 大模型的数据读取、加工与训练 260
10.4 数据空间支撑 AI 大模型的场景推演 262
10.4.1 自建数据空间控制 AI 数据集流转使用 262
10.4.2 参与数据空间生态获取更多的 AI 数据集 263
10.5 小结 265

第 11 章 数据空间未来展望

11.1 数据空间将成为大型组织数据要素流通的标准配置 268
11.1.1 数据空间促进大型企业跨组织协作 268
11.1.2 数据空间提升上下游生态安全水平 269
11.2 行业数据空间支撑开放协作解决行业挑战 270
11.2.1 行业数据空间提升行业整体竞争力 270
11.2.2 欧盟行业数据空间布局初见雏形 271
11.3 数据空间成为国家数据要素流通的基础设施 273
11.3.1 城市数据空间激发数据成为发展新动力 273
11.3.2 海量数据优势转化为国家竞争新优势 274
11.4 数据空间成为全球数据流通的“丝绸之路” 276
11.4.1 跨境数据流动推动全球协作共赢 277
11.4.2 数据交换标准护航全球数据流通 278
11.5 小结 279

1

第 1 章 CHAPTER

数据流通释放价值

数据已经从一个热点概念转变成全社会的共同认知。从大量企业推崇的数字化转型到智能化时代围绕数据构建 AI 大模型的竞争力，数据使全社会发生了深刻变化。数据已成为生产要素，而生产要素一定具备价值，那数据究竟如何产生价值？有价值的东西就需要得到恰当保护，那如何才能在释放数据价值的同时保护好数据权益？这些问题都会通过我们的实践总结，在本书中得到逐一解答。

本章主要围绕“数据流通”一词展开，大多数企业已经充分意识到数据的重要性。随着市场竞争的日趋激烈，在数字化和智能化蓬勃发展的情况下，每一笔数据的共享和消费都对企业异常重要。那么数据在企业中的这种定位是如何形成的？基于这样的定位，如何释放数据的价值才能让数据成为企业发展的核心竞争力？

1.1 释放数据要素价值的改革全面铺开

借助数字化转型，企业通过实现对象、过程、规则的数字化，生成了海量的数据，并打破了“数据孤岛”，加速数据流通，实现万数互联，以“数据为核心”重塑企业能力的效果正在逐渐显现。随着云计算、物联网、区块链、5G 通信技术的成熟和以大模型为基础的人工智能的进一步发展，企业对数据的应用不再局限于内部的数据分析，高质量的数据已经直接或间接地引导着企业内外部各项资源的配置，充分发挥“倍增器”的作用，深刻改变着社会和经济活动以及创造和获取知识的方式，并在社会经济中占据着与土地、劳动力、资本、技术等生产要素同等重要的位置。

在这样的背景下，以建设数字中国为目标，充分发挥数据要素作用，释放数据要素价值的改革，已从国家到地方、从政府到企业全面铺开。

2020 年 4 月 9 日，新华网受权发布《中共中央　国务院关于构建更加完善的要素市场化配置体制机制的意见》，国家首次明确地将数据作为生产要素，并提出推进政府数据开放共享、提升社会数据资源价值、加强数据资源整合和安全保护的要求。

2023 年 3 月 7 日，我国根据国务院关于提请审议国务院机构改革方案的议案，组建国家数据局，负责协调推进数据基础制度建设，统筹数据资源整合共享和开发利用，统筹推进数字中国、数字经济、数字社会规划和建设等。这表明数据将会是未来社会数字化、智能化发展的重要基础。

华为研究团队与业界 1000 多名学者、客户、伙伴耗费 3 年时间，经过 2000 多场研讨，于 2021 年 9 月 22 日发布报告《智能世

界 2030》，对 2020—2030 这个十年周期里的衣、食、住、行、城市、企业、能源、数字可信等领域进行了宏观展望，认为数字产业占比将在 2030 年提升到全球经济的 12%，全球生成的结构化数据、非结构化数据的总量将超过 1000ZB，如图 1-1 所示。

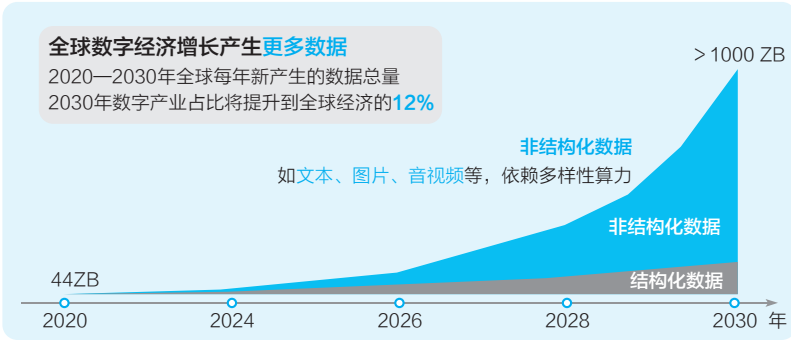


图 1-1　全球数据总量增长趋势

同时 IDC（国际数据公司）和中国国家工业信息安全发展研究中心共同预测，到 2025 年，全球数据消费量将以每年 13.9% 的速度增长，从 2022 年的 1.7 ZB 增长到 2025 年的 3.5 ZB，而“十四五”期间中国数据要素市场规模将增长 3 倍，中国将成为世界最大的数据资源国。但“有数”和“用数”是两个概念，能否靠着“金山”不挨饿，取决于如何成功地释放出数据的价值。

参照同是生产要素的“资本”，金融产业目前已经达到万亿美元规模，未来数据产业必将形成类似于金融产业的，由众多机构参与、类型不同、体量庞大的数据市场，所有的企业不管是主动前进，还是被动参与，都必将成为其中的一分子。同时，数据产业也将改变数据在企业中的定位、企业开展数据治理的目标与数据流通方式。

1.2 企业数据的价值实现

目前业界对数据的定义多种多样，如数据是指任何以电子或者其他方式对信息的记录；数据是对真实世界（包括对象、事件、概念等）的一种符号化描述，描述的方式包括文本、图像、声音、视频和数字等形式。但不管是通过哪类定义来描述数据，都改变不了数据最终都需要通过消费来产生价值，并对其他要素发挥倍增的作用。所以从数据消费的角度来看，基于数据与信息的相互关系，我们可以更加简洁地借用信息论创始人香农在《通信的数学理论》一文中描述的——“信息是用于消除不确定的东西”来描述数据，即数据经过加工处理和流通，也可用于消除不确定性。

历经信息化、数字化的企业，构建了以软件和数据平台为核心的数字世界，便捷地获取和存储了大量数据。随着云计算、人工智能技术的发展，数据在企业中除了被继续使用在传统分析报告上，还开始大规模地被应用在预测推演、模拟仿真、最优决策等活动中。企业通过对数据的推演，评估突发自然灾害对产能、运输、交付的影响；通过对全产线物料的追踪，实现质量器件精确定位和实时反馈。数据消费逐步融入企业的每一项作业、每一次决策中，深刻改变着企业内部人员的思维方式和企业的管理模式，最大限度地降低了不稳定、不确定、复杂、模糊的外部环境对企业经营运营的影响。

华为作为典型的非数字原生企业，经历过信息化、数字化的完整过程，并在逐步地转向全面智能化。华为的数据消费模式与大多数非数字原生企业相似，在各个阶段都有不同的诉求，也呈现出不同的特点。

1.2.1 沿着业务流的数据清洁

数据清洁的主要任务是通过数据清理、数据清洗和数据转换

等动作，确保数据的准确性、完整性和一致性，提高数据质量，为后续的数据应用和数据消费提供可靠的基础。通过数据清洁，可以识别和纠正数据中的错误，确保数据的准确性、完整性和一致性，同时处理无效值和缺失值，将数据转换成应用和消费需要的数据。

在信息化阶段，企业基于现代通信、网络和数据库技术，将线上作业的各个要素汇总到相应的系统数据库中。根据不同的身份和角色，员工获得相应的应用授权，他们可以通过设备和网络在授权的业务应用程序上处理各种业务事件。在操作应用时，员工可以直接对数据库进行读写，记录业务信息。尽管完成一个业务事件可能需要操作多个应用，但通过线上作业，业务的运作效率和准确性仍然得到了提升。信息化阶段的数据流通如图 1-2 所示。

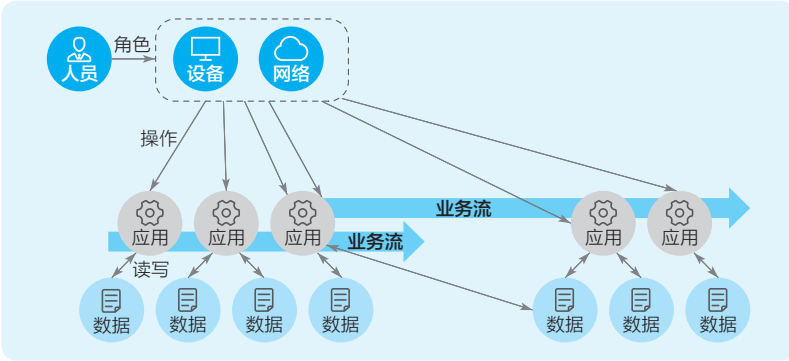


图 1-2　信息化阶段数据流通示意图

信息化阶段数据消费的需求和特点具体如下。

第一，数据消费的主体集中在各个业务流的应用调度里，数据基本不脱离业务流，主要为了支撑业务线上作业，提高业务效率。

第二，工作人员普遍通过终端设备和网络在业务应用中录入

数据或在线上作业时查询业务记录，基本不会脱离业务流去直接消费数据。

第三，对数据的要求更多的是可以被业务应用快速、高效、稳定地调用，数据结构倾向于给业务应用系统提供更高的数据可读效率。

华为在构建信息化的过程中，以自研平台和外购标准软件包相结合的方式，基本形成了从研发到销售、供应、交付、运维的完整信息链路，让企业在生产、采购、销售、客户服务、现金流动等过程中所产生的数据，在业务系统上被记录下来，从而有效服务于从电信基站、服务器、CPU 到计算机、手机、耳机等多项庞大且复杂的业务。

在信息化阶段，数据治理的目标主要是“数据清洁”，通过“公司数据工作组”的方式运作，企业业务流中的各个应用是数据消费的主体。所以在这个阶段，华为设定的数据治理目标是“3 年完成公司数据管理体系建设，5 年内实现数据清洁，用于支撑主业务流的集成、打通，确保财报准确”。在此目标基础上，华为设立数据管理专业组织，建立数据管理框架，发布数据管理政策，任命数据 Owner（责任方），形成唯一可信的数据源、开展主数据信息链治理和数据质量改进工作，以实现业务价值。具体的动作包括：

- 形成企业数据质量度量的整体方案，减少业务过程中的纠错成本，通过数据质量度量与持续改进，以数据质量综合水平得分牵引数据真实反映业务情况，降低运营风险。
- 主数据全流程贯通，提升业务运作效率，通过业务数字化、标准化，借助 IT 技术，实现“数出一孔、同源多用”，让主数据在业务上下游中快速传递、共享。

1.2.2　脱离业务流的数据共享

结合业务的持续发展和外部环境的转变，业务流中原有的数据主要面向系统应用，数据价值难以被企业高层和业务人员显性感知，且已经无法满足企业在数字化环境下整体经营决策的要求。一方面，供应、采购、制造等单一业务数据无法支撑企业完整的业务分析，跨业务的数据共享和整合分析才可以支撑更高维度的企业决策；另一方面，业务流中的数据消费重点是支持高效率的业务活动，但难以支撑企业内大范围的数据共享和分析。例如，在产品交付场景中，产品生产完成后需要及时交付，以减少存储成本，但企业又需要有一定的库存应对生产风险，这就要求企业整合涉及原料采购、生产制造周期、交付时间、运输服务等采购、制造、供应等流程中的多方数据，通过数据整合分析选择最优成本和最短交付时间，通过数据分析决策使数据在业务成本节约和效率提升方面产生价值。所以，从业务流中产生的数据需要从业务流中脱离出来，才能实现跨领域数据的共享。

2016 年，华为变革战略明确提出，要面向用户（企业客户、消费者、员工、合作伙伴、供应商）实现 ROADS（Real-time、On-demand、All-online、DIY、Social，实时、按需、全在线、服务自助、社交化）体验，持续提升效率、效益和客户满意度，用五年时间完成业务数字化转型。结合业务数字化的诉求，2017 年，华为对数据工作同步提出了“实现业务感知、互联、智能和 ROADS 体验，支撑华为数字化转型”的愿景。数字化转型阶段对数据的消费也提出了新的需求。

在这个阶段，为了满足数字化转型后企业分析与决策的消费需要，数据开始逐渐脱离业务流。大规模的数据汇聚入湖，企业开始以数据服务的方式满足不同层次的数据消费需求，以支撑业务的决策和行动，如图 1-3 所示。同时，数据消费需求也呈现出新的特点。

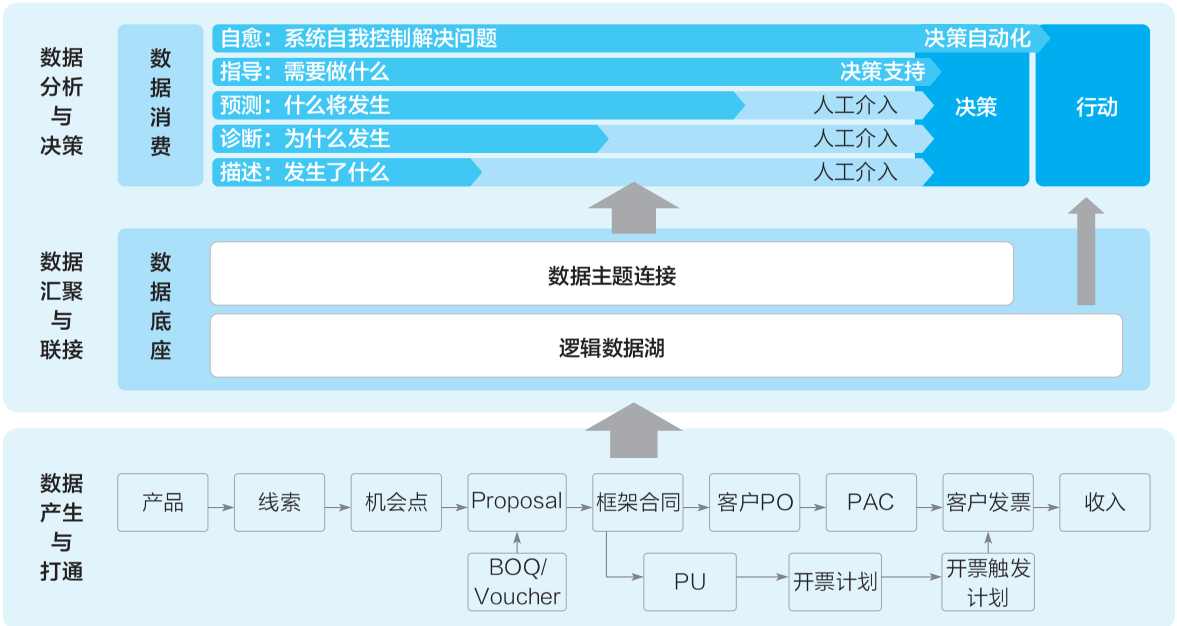


图 1-3　数据汇聚消费示意图

第一，数据的消费在响应各个业务流的应用调度和支撑业务线上作业的同时，开始跨业务流汇聚与联接长期积累下来的交易数据、用户数据、产品数据等。通过不断对数据进行整合、挖掘、分析，可以将数据融入企业的经营管理决策中，以数据发现问题 / 商机、用数据优化业务组合。

第二，业务人员除了在业务流中录入、查询作业记录以外，逐步开始直接消费数据，大规模使用跨业务流的数据进行深度业务分析。企业出现专职的数据分析人员和数据科学家，他们使用专业的数据分析工具处理数据、消费数据，数据进一步脱离业务流，直接支撑业务的决策和行动。

第三，在满足业务应用调度的基础上，数据的主要消费主体从机器转变为人，企业更关注人员对数据的消费体验，要求数据满足普通业务人员“找得到、看得懂、取得到、用得了”的需求。数据治理开始聚焦于数据本身的清洁、互联和在面向业务人员授权消费时的安全可控。

华为在数字化转型阶段的数据工作主要是，围绕信息架构、数据底座、数据服务建设，结合数据感知、数据质量和安全合规能力，打造出覆盖数据全生命周期的数据治理体系，汇聚企业跨业务流的全域数据，通过数据服务、数据地图、数据安全防护与隐私保护，实现随需共享、敏捷自助、安全透明的数据消费。

在这个阶段，由于业务战略规划和数字化转型的影响，企业内跨业务流的数据消费成为主流，所以华为通过“数据资产管理变革项目”的运作，将数据治理目标调整为“建设数据底座、打破数据孤岛，实现企业的数据安全共享、敏捷自助，支撑企业数字化转型”。通过对数据服务、数据地图、数据安全防护与隐私保护等能

力的构建，实现数据价值。具体的动作包括：

- 通过数据汇聚，实现业务状态透明可视，提供基于“事实”的决策支持依据。
- 基于业务规则的数字化、算法化，将决策式 AI 嵌入业务流，逐步替代人工判断。
- 提供数据自助分析服务，让业务人员“找得到、看得懂、取得到、用得了”数据，扩大数据的应用范围，寻找新的业务应用机会点，打造差异化竞争优势。

1.2.3 跨越主体的数据交换

不管是信息化阶段，还是数字化转型阶段，企业数据的价值实现普遍还是集中于企业内部对业务产生的数据的消费，如脱离业务流的数据共享，或通过对企业业务流中的数据进行汇聚、共享，解决企业内跨领域数据连接分析的问题，以固定报表、自助分析等手段满足企业降低经营风险、提升运营效果的需求。但随着智能化的全面推进，对不稳定、不确定、复杂、模糊的外部环境的担忧，使企业进一步希望通过对上游供应商、生态伙伴、下游渠道商和第三方数据的掌握，将这些跨主体的数据与企业内部数据相结合，并将知识提炼成数据、将智慧转化为模型，充分利用 AI 大模型等技术在全连接数字世界中进行探索、模拟和仿真，以增强认知和消除不确定性。如企业内最优决策或模拟仿真等深度应用，除了需要企业内部业务活动产生的数据外，还需要上下游和生态的业务数据，同时知识推理、深度学习和 AI 大模型等也需要大量的通用数据与行业的专业知识，如图 1-4 所示。在全连接数字世界中，无形中被消费的数据的总量已经超越了单一企业主体所能提供的数据总量。

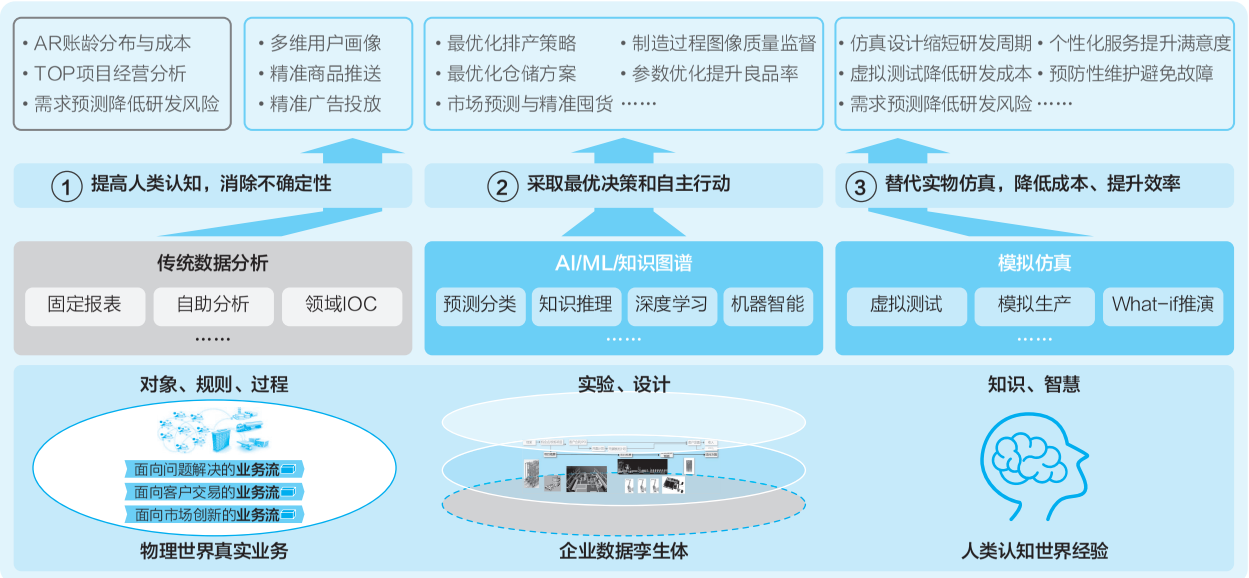


图 1-4　企业数据深度应用示意图

华为也在积极应对全面智能化的到来，识别出相对于信息化、数字化，智能化可能带来的数据消费新变化。单一企业所产生的数据无论如何汇聚都难以满足智能化时代的需求，企业需要从外部生态、数据孪生体、通用知识和智慧中获取更准确、更丰富的多源数据，并通过机器学习、人工智能等技术实现数据消费的智能化。数据的消费模式、数据的消费人员、被消费数据的类型都发生了新的变化。

第一，数据的消费不再拘泥于企业内部，跨企业主体的数据开始被大规模消费，包括供应商、渠道商、生态伙伴、第三方组织等在内的外部数据。通过将全方位的内、外部数据与 AI 大模型等技术手段结合，企业可以实现智能决策和自主行动，支撑智能业务落地，将数据构建成高质量发展的核心竞争力。

第二，业务人员不再直接处理和分析数据，专职数据分析人员和数据科学家更多的工作是将积累的知识和智慧固化到大模型中，通过算力、算法实现对数据的大规模处理和分析。业务人员对数据的消费更依赖于在和 AI 大模型对话时提供准确的提示词，从而高效地获取准确的结果，从“人找数”演进到“数找人”。

第三，除了关注原有的数据质量外，因为涉及跨企业主体的数据，数据消费对外部数据的管理权责、跨主体消费的安全可控等提出了全新的要求。同时，因为生成式 AI 的出现，数据消费的主体又回到了机器，所以也要求数据结构可以更好地适用于生成式 AI 的处理和消费。

华为在智能化阶段的数据工作处于起步阶段。目前，华为在继承信息化阶段、数字化阶段构建的数据能力的基础上，结合信息技术的发展，积极探索生态数据管理、数据要素量化、跨主体可控交换，以及打造面向云化、SAAS 化和人工智能的数据治理等新能力。通过对数据要素和新技术的研究，结合全面智能化（All

Intelligence）战略，让数据在算力中不断地释放潜力。

结合数据成为生产要素的契机，华为在当前提出了“数据成为企业高质量发展的核心竞争力”愿景来牵引未来的数据工作，在数据保护和价值释放之间寻求最优解。虽然这个阶段的数据工作目标还没有最终成形，但如下几项数据工作，必然是支撑全面智能化战略不可或缺的保障。

- 更高质量的数据：面向不同数据消费场景提供高质量的数据，是数据工作永恒的话题，包括如何结合人工智能等新技术更高效地逼近这个目标；如何在 AI 大模型的环境下给生成式 AI 提供更高质量的语料。
- 更安全的数据：专注于企业专属的高质量数据，探索如何将数据与人工智能掌握的通用知识有效结合。在生成式 AI 的应用中，须确保数据安全得到充分保障，防止数据被投毒、非法转移或被大模型滥用。
- 更大范围的数据：数据成为生产要素后，需通过更大范围的流通和消费来释放价值，这就促使数据跨企业边界、跨国家边界的流转更加频繁，如何实现合规、安全、可控的数据消费成为一项重要数据工作。

我们始终认为智能时代的核心就是数据，智能时代的代表——人工智能有一个飞轮，飞轮里包括算力、算法、数据、产品，数据在这个飞轮里面会扮演越来越重要的角色。管理得好，数据就是飞轮的驱动力；管理得不好，数据可能就是飞轮的反作用力，将影响企业现有的优势。

上述企业数据价值实现在信息化、数字化和智能化阶段有着不同的诉求和特点等内容，是基于华为在业务发展过程中对数据工作的认知和总结。各企业由于拥有不同的业务、处于不同的发展阶段，数

据消费模式肯定也处于不同的阶段。但对于非数字原生企业而言，数据驱动业务的发展及创新，通过数据解决企业成本和效率等业务问题的核心目标是不变的，在业务消费中实现数据价值的宗旨始终如一。

在过去 30 多年的信息化、数字化浪潮中，可以看到华为对数据治理的目标随着消费的升级、价值的实现，也在不断调整：定位从早期的数据清洁到“数据共享”，再到目前的愿景“数据成为企业高质量发展的核心竞争力”，充分反映着数据对华为业务发展过程的支撑和影响。伴随着数据消费模式和目标的变化，数据在华为中的定位和流通方式在各个阶段也在不断迭代和优化。

1.3 企业数据定位的变化

结合企业的信息化、数字化和未来全面智能化的进程，数据消费的广度和深度都在不断增强，数据通过充分融入业务来释放价值。通过对企业中数据组织和人员、管理范围和权责、专业能力和影响力等多方面变化进行解读，可以反映出数据在企业中的定位在不断提升。这个定位的变化过程大致可以分为三个阶段，如图 1-5 所示。

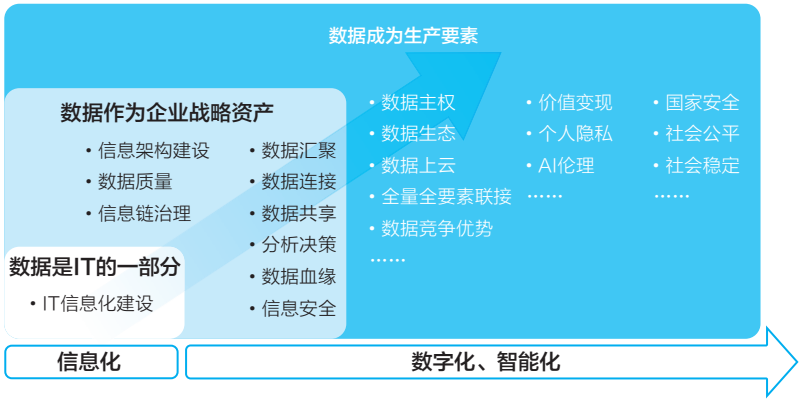


图 1-5　企业数据定位的变化

1.3.1　数据是 IT 的一部分

在企业信息化阶段，数据工作主要围绕建设信息系统和辅助各项工作来开展。这些系统多属于按专项工作需求建设的“烟囱”系统，功能以信息填报采集、业务线上办理为主，提升了工作效率，但没有改变业务管理模式，只是基于简单的数据表记录进行统计和呈现。直到数字化阶段，基于数据的管理模式创新和决策科学化才真正展开，所以在信息化阶段，数据基本被 IT 部门管理，数据工作就是 IT 工作的一部分，数据管理就是数据库管理。IT 部门主要负责业务数据库从设计、测试到部署交付的全生命周期管理，保证数据库管理系统的稳定性、安全性、完整性和高性能，满足各业务应用高效率调度的消费需求。

- 组织和人员层面：没有独立的数据管理组织和岗位，普遍挂靠在业务应用开发部门或平台运营部门。其他部门与 IT 部门合用 IT 开发人员或数据库管理人员，相关工作由 CTO 或 CIO 等技术侧负责人直接领导。
- 管理范围和职责层面：不设立专门的数据管理范围和职责，没有单独针对数据管理工作的规划和需求，信息架构设计主要体现在对数据库和物理表的设计上。企业通过数据库开发评审和运营来统一数据设计语言。
- 专业能力和影响力层面：数据专业能力和影响力基本可以忽略不计。数据专业能力基本指向数据库设计能力或 IT 开发能力，业务和数据、应用和技术脱节，业务人员难以独立消费数据，强依赖 IT 人员的支撑。

1.3.2　数据作为企业战略资产

2014 年，华为通过“数据管理总纲”，首次明确提出数据是公

司的核心资产，要求数据的价值和风险应被有效管理，以支撑内部管理简化、业务流集成、运营效率提升和对经营结果的真实呈现，并在 2017 年将数据提升为企业的战略资产。在这个阶段，华为发布“华为数据治理体系框架”，围绕公司战略方向和业务需求展开，在上层通过政策指引提出治理的明确要求，制定数据架构标准管理和数据质量管理的规范，确保数据治理的政策规范在流程中得到保障，并且不同层次的规范都有对应的组织负责，如图 1-6 所示。

华为通过 7 项举措构建数据体系和各项专业能力，落地“华为数据治理体系框架”，全面开展企业数据治理工作，让数据发挥价值，使数据真正符合企业战略资产的定位。

举措 1：发布公司层面的政策指引，在公司全局层面明确信息架构、数据产生到应用的各项基本管理要求，并要求统一遵从。

举措 2：明确数据架构管理标准，通过架构统一公司的数据语言，打通业务流、消除信息孤岛、提升业务流集成效率。

举措 3：开展全面的数据质量管理工作，明确数据在创建、维护、应用过程中的规则及质量要求，将数据质量与业务负责人绑定，确保业务数据真实可靠。

举措 4：建立数据管理流程，明确数据管理的关键活动和角色，以及各一级领域与周边组织的协作关系，覆盖数据的全生命周期。

举措 5：构建业务负责制的管理责任体系，分层分级任命数据 Owner，并在公司层面设置公司数据 Owner，确保公司数据工作可以统筹规划。

举措 6：设立独立的数据管理组织，任命公司级数据管理部门和领域级数据管理部门，通过专业的数据组织和人员规划落实公司制定的数据管理目标。

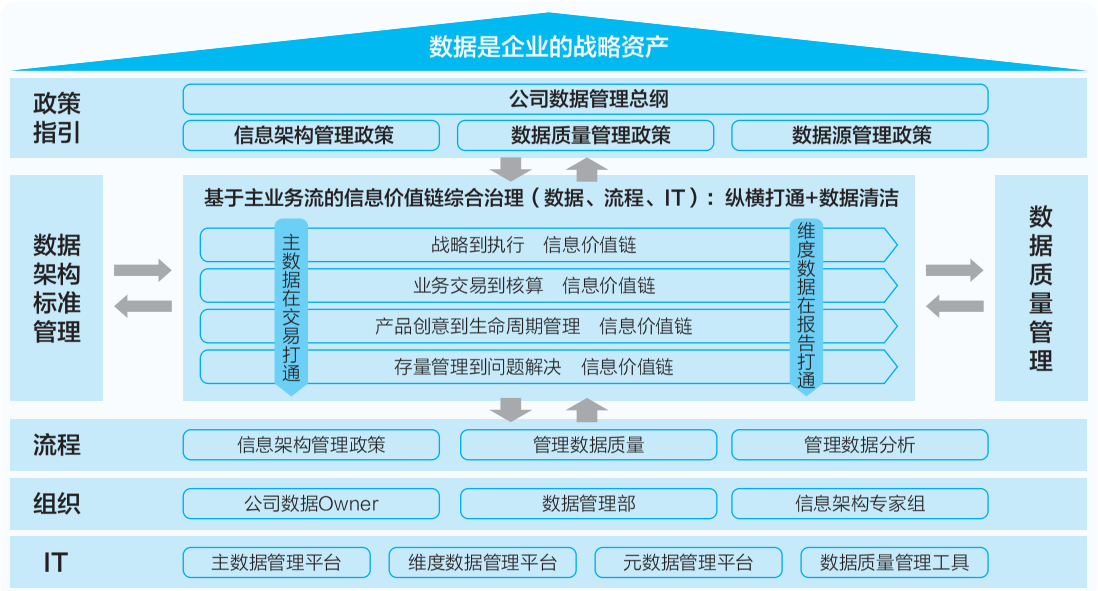


图 1-6　华为数据治理体系框架

举措 7：开发数据治理平台，将数据治理固化到平台中、融入 IT 工作流程里，确保各项政策、流程、规范和方法可以得到有效执行，并全面监控。

其中，数据底座在数据作为企业战略资产阶段的重要成果，形成了支撑华为数字化转型的关键能力，承担着打破数据孤岛、确保源头数据准确、促进数据共享、保障数据隐私与安全等责任。从 2017 年开始建设，到 2022 年年底，数据底座最终达到了如下规模，为华为实现数字化转型和应对业务连续性提供了充分保障。

- 数据采集：采集并管理 360 万张物理表，包含亿级规模的物理字段，华为首次在企业层面掌握了准确的物理表和字段的数量与分布。
- 数据入湖：将 21 个数据库纳入管理，实现了 27 个业务领域、1400 个以上业务对象的日常管理和 800 亿次的每日调度。
- 数据连接：对维度数据、图模型、指标、标签、算法等实现初步管理，其中基于图模型的连接达到 220 亿个数据实例，有效满足了业务连续性的需求。
- 数据消费：发布和管理 5000 个数据服务，支撑 220 个业务租户、9.5 万张分析报告的日常消费。

1.3.3 数据成为生产要素

2020 年，国家将数据定位为生产要素，而华为正处在数字化转型和应对业务连续性的关键时期。面对这样的内外部环境变化，华为进行了统一规划，并分层次地开展了对这些变化的探索和研究。最终，华为在四个主要方向上建立了初步的理论基础和专业实践。这些努力结合了企业业务发展的转变和技术发展趋势，以满足

企业全面智能化对数据的需求。

1. 数据要素在企业中的增值逻辑和价值量化

迎合数据成为生产要素的契机，显性衡量数据对企业内部业务的贡献度和价值，让基于数据消费形成的业务价值可以回归成数据价值，并通过对数据价值的量化和分配，让数据提供者以及数据生命周期中的各方参与者都能体现出工作价值，从而激发治理的积极性，形成数据流通与消费的内生动力，促进企业数据共享。

2. 数据可控交换

数据具有可量化的价值，需要满足外部生态系统和内部多产业的数据消费需求。然而，如果这些有价值的数据得不到妥善保护，它们易于扩散且难以确立所有权的特性可能会损害数据提供方的权益。在缺乏信任的环境中，数据交换双方如何确保数据在交换和消费过程中全程可控、透明可查，是数据提供方愿意持续分享数据的关键所在。

3. 云化、SAAS 化背景下的数据治理

随着企业竞争环境的变化和 IT 技术的快速发展，全面云化、SAAS 化成为企业必然的选择，大部分企业的 IT 基础设施和系统应用将逐步交予专业的云厂商和软件厂商进行管理。企业间最大的差异将是经营运营活动中产生的业务数据的差异，如何治理并消费好这些差异化的数据是影响企业竞争优势的关键。

4. AI 数据集的治理

为了应对 AI 对企业数据消费的冲击，大量的非结构化数据和之前企业难以被消费的数据将被纳入大模型的训练中以支撑业务的

使用。如何实现该类数据集在大模型全生命周期中的合规使用和有效管理，确保数据真实不被篡改以及全流程可追溯，成为企业是否愿意相信大模型、使用大模型的关键。

数据成为生产要素对于所有的组织来说都是一个全新的课题，华为在这个阶段的数据工作也处于探索研究和逐步实践的过程中。目前在数据的定位上，华为已明确提出下一阶段数据工作的愿景为“数据成为企业高质量发展的核心竞争力”。

1.4 企业数据流通模式的演进

不同阶段的数据消费模式，都在释放着不断增加的数据价值，逐步提升了数据在企业中的重要性，同时这也意味着数据流通模式需要不断发展，结合计算机相关技术进步，满足不同阶段的数据消费需求，成就数据在企业中的定位。

随着企业从信息化到数字化的转变，数据的内涵和外延发生了深刻的变化，从 IT 的一部分转变为生产要素。数据定位的不断提升推动着企业数据流通模式的演变，督促着企业围绕数字化转型主线进一步深化协作、融合，发挥数据要素的倍增作用，形成新质生产力。流通的数据才能产生价值，企业就是基于不断扩大的数据流通、汇聚、共享的边界和范围，通过持续挖掘、分析来实现数据价值。华为经过多年的探索和实践，在信息化、数字化阶段形成了较为成熟的流通机制和模式，积累了一些较好的经验，体现了公司管理层对数据流通重要性的统一认知。

1.4.1 数据流通从集成、汇聚到交换

基于多年实践的总结，华为将企业数据流通归纳成三种模式，

包括基于业务流的数据集成、基于数据湖的数据汇聚、基于数据合约的数据交换，如图 1-7 所示。三种模式均有各自的特点，企业可以基于不同的流通需求选择最适合的数据流通模式。

1. 数据集成模式

在企业信息化阶段，数据主要沿着业务流流通，不管是在企业内或企业间。华为发布了《主数据集成管理规范》《主干交易平台产品集成接口规范》等政策规范，通过 TIBICO、DataStage、PowerCenter、ODI 等 ETL 工具，围绕业务流进行应用系统间的数据集成，通过应用系统间的权限管控实现业务流中的数据流通。

2. 数据汇聚模式

在数字化转型过程中，企业数据消费开始更多地通过跨业务数据分析来支撑经营 / 运营活动，数据流通突破业务流的约束。华为发布了《公司数据湖建设规范》《数据服务管理流程》《数据底座共享与安全管理规定》等政策规范，通过元数据采集以及数据底座、数据服务的建设，将跨域的数据汇聚，实现数据在企业内基于权限控制的跨业务流汇聚和安全共享。

3. 数据交换模式

随着数据成为生产要素，数据的共享、开放程度史无前例。促进数据的流通不仅是企业创新发展的必然要求，也是社会经济发展的必然要求。因此，企业数据的流通开始在数据集成和汇聚这两种模式的基础上，逐步以独立生产要素的身份参与数据交换。特别是跨主体的数据流通交换，需要在实现数据跨企业边界流通后，可以继续基于数据合约实现使用控制，保护数据提供方的权益。

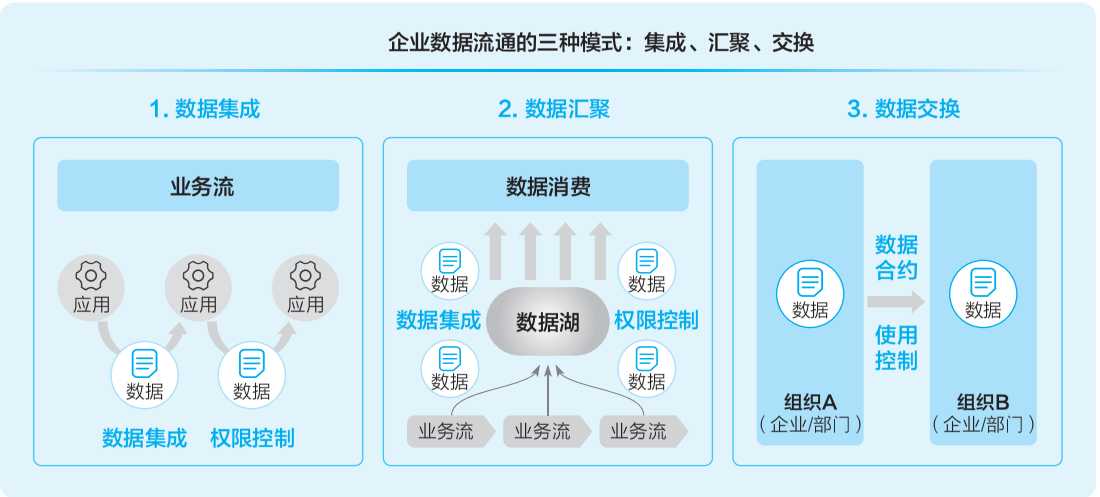


图 1-7　企业数据流通的三种模式

在企业数据集成、汇聚和交换的三种流通模式中，数据集成依附于业务流，并主要面向 IT 系统应用；数据汇聚则将数据汇聚到逻辑数据湖，面向企业内部安全共享，主要服务企业内部，且流通范围明确。数据在这两种流通模式中的风险基本可以通过权限进行控制。

随着企业开始数字化转型后，数据感知能力加强，应对业务连续性的复杂推演和分析需求增多，急需更多样、更大范围的上下游关联企业的数据用于消费。在此情形下，数据的易复制、非排他和传播迅速等在原有流通模式中的优点，却在数据交换的流通模式中成为弊端。

1.4.2　数据交换的关键：跨主体边界后的可控

当数据交换需要作为一种模式参与数据流通时，数据流通的目的、控制边界和数据消费群体都发生了显著变化，特别是数据在越过企业边界进行数据交换时，通过原来的企业内权限控制已经难以对该流通模式下的数据进行有效管控。因为权限控制时，数据提供方只能根据消费方申报的数据消费场景的合理性进行判断。授权审批完成就意味着数据提供方对消费方是“责任转移、无限信任”。消费方是否只按申报的用途消费数据，或是否会将数据共享给其他未获授权的消费方，数据提供方完全不知情，也无法控制。

企业在经营过程中不可能完全独立于其他企业而存在。在业务活动中，企业必然会与外部客户、供应商以及合作伙伴建立联系，这种联系在跨企业边界或跨国界的情况下尤为明显，如图 1-8 所示。一旦建立了联系，数据流通就成为必然，许多数据需要跨越企业边界，传递给国内的合作伙伴或客户。

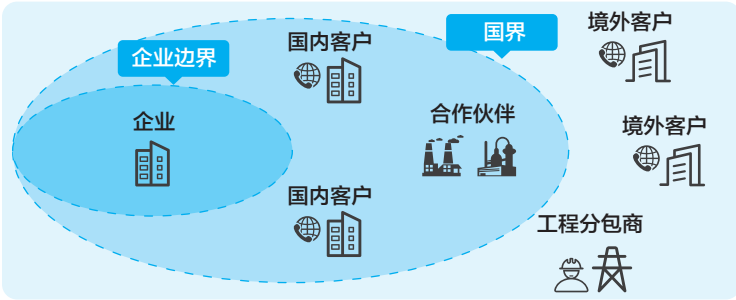


图 1-8　跨企业边界或跨国界合作示意图

当数据跨过企业边界后，原有的数据授权体系和防护模式已经改变，企业无法确保提供的数据是否按签署的合同条款被使用，或者当数据要流通到不同国家的境外客户时，需要受到客户所在国家数据合规条款的约束，但跨过企业边界和国界后，消费方又无法证明是否在合规地消费数据。在数据交换模式下，数据流通目的、数据可控边界和数据消费群体都发生了变化。

1. 数据流通目的的变化

数据流通的目的已经从仅满足业务交易或企业内部决策的需要扩展到了更广泛的领域，即利用多元化的外部数据来减少业务中的不确定性。例如，金融评级公司可能会利用电网消费数据来评估企业的开工情况。这些电网消费数据既不属于金融评级公司，也不是被评价企业的直接业务数据。

2. 数据可控边界的变化

数据从企业内流通至边界外，消费方式由权限可控变为不可控。例如，企业需要跨边界获取外部供应商的数据来预判供应风险，但获取到数据后，外部供应商会担忧企业是否真实将数据用于

预判供应风险，数据申请的用途和真实消费过程是否一致。

3. 数据消费群体的变化

数据消费群体已脱离企业熟悉的认知范围，提供方对消费方能否持续保护数据缺乏信任。例如，第三方机构需要企业提供数据做一次营商资质评估，但第三方评级机构往往是招标入选，不是企业多年的业务合作方。

这三方面的变化导致数据需要跨边界存储，以及数据消费方处于难以被数据提供方充分信任的环境中，再加上数据具有的易复制、非排他、传播迅速等特性，更加剧了数据提供方对数据跨边界流通后的扩散担忧，使得供需双方难以达成数据交换。

构建数据交换企业间基本的信任是实现数据交换的前提。在数据跨边界满足消费需求的同时，如果数据提供方想要实现对消费方的持续可控，至少需要具备以下两个条件。

一方面，通过在消费侧构建可控环境建立信任。结合对应的交换协议，在消费侧构建一个数据提供方信任的可控环境，将数据提供方对数据的管理控制延伸到消费侧，在利用数据消费非排他性优势的同时，保证数据提供方可以充分行使排他性控制，从而保护数据提供方的权益。

另一方面，通过技术手段强化对合约履行的信任。数据的传播迅速和易复制特性，使得依靠合同约定和事后监管来防止侵权变得更加困难。为了应对这一挑战，数据交换需要采取“以快制快”的策略，利用技术手段确保纸质合同中的管理条款能够实时且准确地执行。通过记录、追溯和证明数据的整个消费过程，可以确保提供数据的初衷与实际使用目的保持一致。

目前，如何确保数据跨过企业边界流通后在消费侧可被控制是一个难题，华为自 2007 年开始首次建立数据管理实体组织以来，走过了十几年的数据治理历程。从最初的数据管理体系搭建，到主干业务流全场景覆盖，确保了各个业务流程产生的数据准确。2017 年启动数字化转型后，华为提升了数据治理能力，在实践中形成了数据全生命周期治理规范与方案。

现在，我们沿着数据流通释放价值的主线，总结经验，探索未来。早期的数据治理经验为数据要素化和全面智能化奠定了坚实的基础。随着对产业数字化转型的深入理解，我们认识到数据作为生产要素的价值释放和权益保护的重要性，以及实现数据可控交换的必要性。这不仅要求华为在理论、方法、流程规范和技术能力上进行创新，还需要在数据治理实践中不断验证和探索，以构建一个综合性的系统工程。

1.5 小结

本章对企业信息化、数字化和即将面对的智能化实践进行了总结，结合业务和新技术的发展，详细介绍华为对未来数据工作的部分前瞻性探索和实践。后续将通过十章内容，从企业中各阶段数据消费模式的演进和数据定位的变化开始，围绕数据消费新场景和新技术发展，抓住数据通过流通释放价值的核心观点，详细阐述华为跨主体、跨边界开展数据交换的理论、方法、治理思考和实践过程。

2

第 2 章 CHAPTER

数据流通的世界级难题

如前所述，经过几轮数字化浪潮，数据总量已经呈现爆发式增长趋势。同时，基于数据驱动获得巨大商业成功的组织也数不胜数。数据越来越被重视，当前在国内已经成为第五生产要素。生产要素要通过流通体现价值，与企业通过购买、出售、投资等方式获取资本要素的流通收益一样，数据要素通过流通也可以释放出巨大的价值潜力。通过收集和融合原始数据，企业可基于数据做出决策或创造价值。流通使数据从产生方流向使用方，使用方通过对数据的分析和应用，提取有价值的信息，以此发现新的商业模式、优化产品服务或提高运营效率，从而释放数据的价值。总的来说，数据流通释放价值主要体现在增值和消除信息不确定性方面。

- 数据通过流通实现增值：原始数据面对不同的场景可以反馈不同的信息，不同类型的原始数据汇聚融合后可以产生更多的信息。这一过程中，数据的流通不仅促进了信息的提炼和优化，也使数据能够被更广泛地应用，从而在不同领域产生价值。
- 数据流通后消除信息不确定性：数据流通有助于消除信息的不确定性。通过共享和交换数据，各方都能够获得更全面、准确的信息，从而降低决策风险，提高预测和规划的准确性。信息的透明性和可获取性是实现数据价值的重要保障。

然而，因自身特点、拥有主体与使用主体之间潜在的利益冲突、流通过程的安全性等，数据面临着与土地、资产等要素不同的流通问题。

2.1 数据要素的独特性给流通带来新难题

数据作为一种特殊的生产要素，其流通面临着一系列独特的挑战和难点。数据作为不可实物化的要素，与土地、劳动力等要素不同。土地、劳动力能依附于实物，交换就意味着要素的产权可以发生实质性转移。数据与技术也存在差别，技术通过登记明确产权后，需要体现在产品上，可以通过产品追溯合法性。而数据作为要素，目前对数据特征的识别并不完整，数据的部分特征在流通、确权和定值等方面存在矛盾。

结合企业在信息化、数字化过程中的数据流通、数据消费经验，以及业界对数据特征的共识，我们总结了数据的现实特征（部分），如图 2-1 所示。

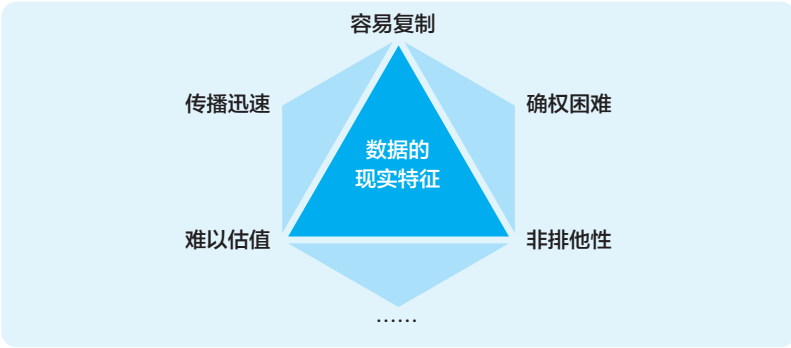


图 2-1　数据的现实特征（部分）

- 容易复制：复制速度快，成本低，难度小。
- 确权困难：数据所有权、控制权、处置权的识别与其他要素有显著差异。
- 非排他性：复制给 A 并不妨碍复制给 B；A 在使用时并不妨碍 B 也同时使用。
- 传播迅速：基于信息化、数字化、云化等手段，无距离限制，可无限供给。
- 难以估值：价值不确定性和“阿罗信息悖论”。
- 异质性：A 对数据价值的定义无法直接用于 B。
- 协同性：单一数据或单一模型的价值有限。

从数据流通的角度看，数据的非排他性和容易复制是关键难题。

2.1.1　难以管控数据使用

数据的非排他性是指数据可以同时被多个不同的主体并行使用。由于数据存在非排他性，数据的价值无法锁定，随着复制和传

播可能会不断降低。数据可以轻易地被复制和共享，明确数据的所有权和控制权变得复杂，这可能产生关于数据来源和使用权的争议。同时，数据经过多次复制和修改后，数据加工产物的收益及风险给数据提供者带来了不对等的地位，权责的归属和分割也变得具有挑战性。

数据的非排他性使数据的持有和使用难以用统一的管理规则来界定。数据种类繁多，不同类型的数据需要用不同的元数据模型和标准来描述。即使是同一类数据（例如天气预报），由于描述的详细程度可以根据不同的应用场景而变化，因此很难通过简单直接的方法来识别它们是否为同一类数据。此外，不同类型和用途的数据通常需要不同的管理规则，因此制定一套适用于所有数据的统一规则是非常困难的。例如，个人隐私数据需要严格的保护措施，而公共数据可能需要更大程度的开放和共享。这也意味着在法律或治理层面，仅通过管理规则来实施非排他性的管理是非常具有挑战性的。

数据的非排他性让数据加工后的权责和后续流转的追踪变得更加复杂。数据的分析和加工，既要尊重数据处理者对数据增值的正向作用，又要保障数据持有者的原始权益。但数据的分析和加工既可能保留数据的原始形态（如部分原始数据的片段），又可能因叠加相关规则完全改变数据的结构和信息内涵。当数据加工产物产生明确的收益时，从数据提供方的角度看，缺乏价值分配规则，例如个人数据与平台加工产物之间，个人是否有权享有相关收益。同时，在面临风险时，数据提供方往往不可避免地需要承担相应的责任。数据经过多层转换和加工后，溯源变得更加困难，这使得数据加工产物的收益和风险分配更加难以确定。

2.1.2　增加数据使用风险

数据目前普遍以电子化的方式存储，这在使数据流通便捷的同时，也意味着容易复制，几乎可以零成本、便捷地将数据复制到各类存储载体中。数据容易复制还体现在其传播迅速：海底电缆目前最快可以实现每秒 80 万兆比特的传输，5.5G 技术的下行速率可达到每秒 1.1 万兆比特，这意味着 TB 级容量的数据通过网络，可以在秒级时间内传输到全球。

数据被多次复制和传播后，数据的原始来源和版本管理变得复杂。数据在复制和传播过程可能会被修改或损坏，质量的可靠性和一致性也难以保障。数据的易复制性增加了数据泄露和未授权访问的风险，甚至可能导致数据被滥用，如用于欺诈、误导或不当目的。当数据被复制和使用时，确定数据的责任归属会变得困难，特别是在数据出现问题时，还可能会降低其独特价值，使数据价值评估变得更加复杂。

2.2　数据主体的利益保障和风险控制

由于自身的特点，数据相比于其他要素面临更大的挑战。从数据主体的视角看，收益与风险不对等是阻碍流通的核心原因。虽然有很多商业组织通过数据进行创新获得成功，但相比于所有行业，在开展数字化转型的大企业中，实现数据复用增值的仅有 8.3%（参考《全国数据资源调查报告（2023 年）》）。在商业利益、法律合规遵从和技术保障等方面，数据流通都面临不同程度的新挑战，如图 2-2 所示。

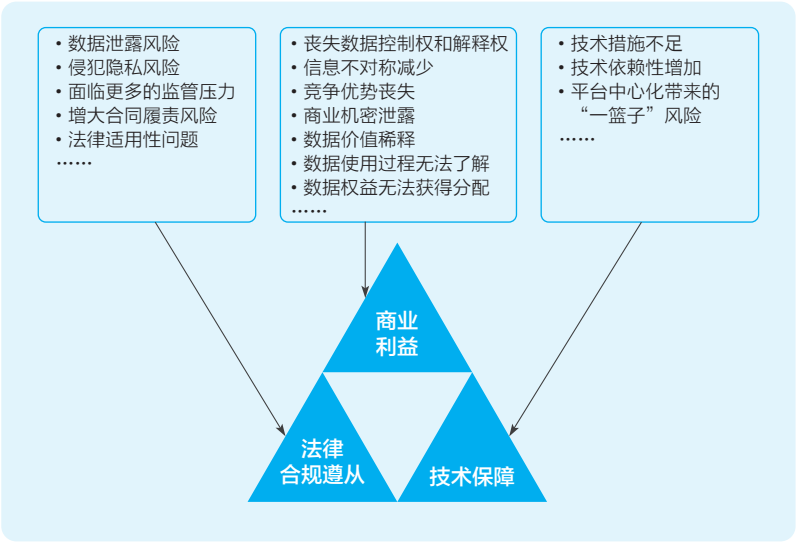


图 2-2　从数据主体视角看数据流通面临的挑战

从主体的收益和风险角度看，商业利益的收益与丧失优势的风险不对等，相关法律法规在不断地变化和发展，可能使主体开放数据后疲于应付新的合规要求，既想避免技术上过多投入，又担心技术依赖风险。因此，数据主体在数据流通中处于相对矛盾的地位。

2.2.1　商业利益

1. 丧失数据控制权和解释权

数据主体担心失去对自己数据的控制权。在数据共享过程中，数据主体可能无法决定自己的数据如何被使用和分发。一般的数据集成协议和机制，只能依赖事后稽查，通常是在问题已经发生且造成一定的法律后果后才能采取一些补充措施，但此时对数据主体而

言已经丧失对数据的控制权。同时，数据在流通过程中可能面临被篡改、滥用或未授权访问的风险，从而导致数据主体对数据质量产生无限责任。

2. 信息不对称减少

数据共享减少了信息不对称，使原本拥有独特数据资源的企业或个人失去了信息优势。其他竞争者通过共享获得同等的数据资源，可能会迅速缩小与数据拥有者之间的差距。

3. 竞争优势丧失

数据拥有者可能依赖其独特的数据集来开发新产品、服务或优化业务流程，从而获得市场竞争优势。数据共享后，其他竞争者也可能利用相同的数据来模仿或改进，从而削弱数据拥有者的市场地位，可能改变市场的竞争格局。原本的市场领导者可能因为数据共享而面临新兴竞争者的挑战，需要重新制定自己的市场策略。

4. 商业机密泄露

在数据共享过程中，如果保护不当，可能会泄露商业机密，如客户列表、定价策略、研发成果等，这些信息一旦被竞争对手获取，可能会直接影响数据拥有者的商业利益。

5. 数据价值稀释

当数据被广泛共享时，其独特性和稀缺性可能会降低，导致数据的市场价值下降。数据拥有者原本可以通过出售或授权使用数据来获得收益，但共享后，这种收益模式可能不再可行。

6. 数据使用过程无法了解

数据主体担心数据的使用过程缺乏透明度，无法了解数据如何被处理和分析。例如，一些企业在使用共享数据时，未明确告知数据主体数据的具体使用目的和范围，导致数据主体对具体的数据使用情况一无所知，极易产生后续的风险。

7. 数据权益无法获得分配

数据主体担心自己在数据产生价值的过程中的权益得不到合理分配。数据的收集者和使用者可能获得巨大利益，而数据主体却没有得到相应的回报。例如，一些互联网公司通过收集用户数据进行精准营销获得广告收入，但用户作为数据提供者，往往没有获得任何形式的回报。

2.2.2 法律合规遵从

1. 数据泄露风险

在数据共享过程中，如果未能落实法律规定的保护措施，可能导致数据泄露，给数据主体带来安全风险。尤其是云服务提供商在提供由面向客户的技术服务产生的与客户关联的数据时，数据持有主体及其关注运营平台应如何管理和使用数据。

2. 侵犯隐私风险

数据共享可能涉及个人隐私信息，违反法律规定的隐私保护要求可能会侵犯数据主体的隐私权。

3. 面临更多的监管压力

数据共享活动会给企业增加监管压力，导致企业面临更多的

监管审查和合规要求。

4. 增大合同履责风险

数据共享可能涉及违反与数据提供方签订的合同条款或法律规定，如违反保密协议，这可能导致法律诉讼、罚款或承担其他法律责任。

5. 法律适用性问题

随着技术的发展，现有的法律规定可能无法完全适应新的数据共享场景，导致法律适用上的不确定性增加。新兴的数据共享模式，如区块链、云计算等，可能需要新的法律解释，以确保数据共享的合法性和安全性，尤其是数据跨境场景。在全球化背景下，数据可能跨境流动，一国的法律规定可能与其他国家的法律要求发生冲突。如欧盟的 GDPR 对数据保护有严格的规定，其他国家的企业在与欧盟成员国进行数据共享时，必须遵守 GDPR，否则可能面临法律风险。在当前缺乏成熟惯例或者成文法的背景下，企业即使拥有海量数据，也不敢共享与业务直接合作无关的数据。

2.2.3　技术保障

1. 技术措施不足

数据共享涉及数据的采集、存储、传输等多个环节，依赖多种不同的技术保障，任何一环出现纰漏都将带来安全问题。当前企业缺乏标准、规范、可靠的技术以支撑其在数据共享中的安全保障。由于在数据集成、存储、采集等方面的技术点众多，如何找到更匹配本企业当前技术栈的共享技术，避免在进行技术集成时出现技术漏洞，是技术方面的难点。

2. 技术依赖性增加

为了在数据共享中保持竞争优势，数据拥有者可能需要依赖更先进的技术来保护数据安全和隐私，这可能会增加对技术的依赖和技术成本。平台型或者技术驱动型的数据交换服务商，依赖自身的技术积累和优势，可能会用技术“绑架”数据主体。面向数据共享交换的系统性集成技术，目前还未形成普遍的技术共识，缺乏具备公信力的执行标准。

3. 平台中心化带来的“一篮子”风险

目前很多数据服务都采用中心化平台的架构设计。相比于去中心化的架构，中心化平台更容易被单点故障影响整个系统，一旦安全防护措施被突破，可能导致大量敏感数据泄露。此外，数据更容易被集中恶意删除，完整恢复的可能性更低。因此，选择中心化的平台型服务，从技术角度看更像把“数据放在一个篮子里”，当数据因各种原因被删除或者丢失时，企业可能会面临业务中断甚至无法恢复运营的风险，造成巨大的经济损失和品牌声誉损害。

2.3 世界主要经济体的应对战略及举措

在当今数字化时代，数据已成为国家和全球竞争的关键资源，它不仅是推动经济增长、促进科技创新、优化决策过程的重要资产，也是衡量一个国家综合竞争力和软实力的重要指标，以数据为基础的数字经济已经成为全球经济增长的重要引擎。全球主要经济体都在积极探索并解决数据流通共享的难题。本节简要介绍美国、欧盟、中国、日本四个主要经济体在促进数据流通上的战略举措。

1. 美国

美国一直试图重新领导数据共享开放相关规则的制定，持续增强自身的影响力。

美国持续制定数据管理与利用方面的战略和行动计划。2009 年的《透明和开放政府备忘录》体现了美国对公共数据开放的高度重视。2018 年的《开放政府数据法案》强调了政务公共数据具体的开放指导原则，促进公共和私营部门基于数据融合的创新，对全球政府数据的开放产生了影响，并为其他国家提供了借鉴。2019 年发布的《联邦数据战略》，进一步描述了未来 10 年联邦政府数据开放和共享的方向。2020 年修订了《国防部数据战略》，进一步明确了数据作为战略资源的角色，并提出了数据治理、数据质量和数据共享的指导原则。2023 年发布了《数据、分析和人工智能采用战略》，这是对之前战略的迭代整合，期望实现更进一步的数据开放，推动人工智能、先进模式识别和自主技术的投资，依托数据打造先进能力。

通过上述的战略、法案和行动，美国积极调动政府部门、产业界、学术界、公众等创新主体并促进互动，让数据“活”于流动之中。其中，美国政府承担了统筹打通、汇聚、共享等环节的工作。具体来看，美国政府积极建立了 Data.gov 等，用于提供各级政府和非营利机构的数据，并与谷歌、微软、亚马逊等科技巨头合作，签订“零成本合同”，由使用数字政府服务的企业和个人缴纳服务费作为合作企业的收益。为顺利推进共享开放，美国采用军方标准建立安全保护方案，并联合科研机构和行业基金会探索数据共享新模式，以推动数据共享。美国信息技术与创新基金会（ITIF）在报告《探索数据共享模式以最大限度发挥数据优势》中提出六种数据共享模式（数据共享伙伴关系、数据联盟、数据托管、数据合

作社、分布式联合数据分析、合作研发协议），帮助政策制定者选择最适合特定需求的数据共享模式，增强数据共享的优势，并推动数据效益的最大化。

美国基于强大的技术领先优势，倾向于自由主义导向，维持数据隐私监管和创新之间的平衡；积极协调数据主体，促进数据流通；积极推动公共数据开放，并通过一系列政策来推动数据开放。

参考资料：

[1] CAICT 互联网法律研究中心．美国白宫《联邦数据战略和 2020 年行动计划》概述［EB/OL］.（2020-01-13）［2024-04-07］. https://www.secrss.com/articles/16595.

[2] 大数据技术标准推进委员会．美政府发布《联邦数据战略》，提出 16 项行动计划［EB/OL］.（2019-06-13）［2024-04-07］. https://www.secrss.com/articles/11352.

2. 欧盟

欧盟在数据上的整体策略是以数据治理为核心，充分激发数据价值，从而释放数字经济的潜力。在推进数据共享和开放上，欧盟采取战略顶层设计引领、深化全球合作、强化新兴技术治理等多头推进策略。

欧盟的数据流通开放采取了如下四大关键措施。

一是发布统一的治理框架，如《欧洲数据战略》和《欧洲人工智能白皮书》。欧盟为构建健康共通的数据空间提供蓝图，并制定了新的《欧盟数字十年网络安全战略》。《数据治理方案》为数据共享流通奠定体制机制的基础，并对数据合规和隐私保护提出要

求。《数据方案》旨在促进企业之间以及企业与政府之间的数据共享，使消费者和企业对其拥有的数据有更多的控制权，并以此构建欧盟内部统一的数据获取和利用的治理规则框架，改变欧盟长期以来存在的内部市场分散的问题。

二是统一管理机构。欧盟构建了核心行业和公共利益领域的统一数据空间，并计划成立一家中立的数据管理公司，承担数据管理机构的职责。

三是加大数据领域投资。欧盟计划在 2021—2027 年投资有关数据空间和联合云基础设施的高影响力项目，投资总金额将达到 60 亿欧元，以此增强欧洲在数据储存、处理、利用和兼容方面的技术能力与基础设施建设。

四是加强数据专业人才建设，侧重提升个体数据权利和技能，推动中小企业的数据能力建设，增强中小企业数据创新能力。

欧盟在数据共享和利用上强调跨部门和成员国之间的共享，同时在欧盟框架下实现与其他经济体的互通。

参考资料：

[1] European Commission. European Data Strategy [EB/OL].（2020-02-19）[2024-03-01]. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en.

[2] International Data Spaces. IDS Reference Architecture Model（IDS-RAM）4 [EB/OL]. [2024-03-02]. https://docs.internationaldataspaces.org/ids-knowledgebase/ids-ram-4.

[3] Gaia-X AISBL. Home-Gaia-X: A Federated Secure Data Infrastructure [EB/OL]. [2024-03-01]. https://gaia-x.eu/.

3. 中国

我国将数据作为第五生产要素，提出数据要素发展与数据安全并重，发展和安全一体两面，在博弈中寻求平衡点。

政策层面，我国加快数据立法及政策出台，数据基础制度建设初见雏形，数据战略从 2014 年大数据首次被写入政府工作报告开始。2016 年工业和信息化部发布的《大数据产业发展规划（2016—2020 年）》明确了具体的落地方向。2019 年持续深化，大数据已连续 6 年被写入政府工作报告，党的十九届四中全会首次将“数据”列为生产要素。自此之后，国家出台了一系列推进政策以完善数据基础制度（如 2020 年《关于构建更加完善的要素市场化配置体制机制的意见》、2022 年《中共中央 国务院关于构建数据基础制度更好发挥数据要素作用的意见》、2023 年《数字中国建设整体布局规划》等）。2022 年 12 月 19 日，《中共中央 国务院关于构建数据基础制度更好发挥数据要素作用的意见》，即“数据二十条”发布，旨在从数据产权、流通交易、收益分配、安全治理等方面构建数据基础制度，标志着中国在数据要素市场培育和数据基础制度建设方面迈出了重要一步，对于促进数据合规高效流通使用、赋能实体经济具有重要意义。2023 年 10 月 25 日，国家数据局正式挂牌成立，负责协调推进数据基础制度建设，统筹数据资源整合共享和开发利用，统筹推进数字中国、数字经济、数字社会规划和建设等。地方政府也在通过政策制定与试点落地促进数据要素流通，积极完善制度流程、健全交易体系、创新多元机制并落实监管。

除了政策、制度和组织保障外，我国在具体行动方面提出建设新型的“数据基础设施”，包括以下几个方面。

- 发布《“数据要素 ×”三年行动计划（2024—2026 年）》，推动数据与其他生产要素协同，以数据流引领资源优化配置。
- 发布《数字中国建设整体布局规划》，提出了 2025 年和 2035 年的数字化发展目标，强调要夯实数字基础设施和数据资源体系“两大基础”，并推进数字技术与经济、政治、文化、社会、生态文明建设“五位一体”的深度融合。
- 组织建设数据资源库，以推动公共数据汇聚利用，释放商业数据价值潜能。
- 持续建设规范高效的数据交易场所，出台数据交易场所管理办法，建立健全数据交易规则，制定全国统一的数据交易、数据安全等标准体系。
- 建立数据流通准入标准规则，强化市场主体数据全流程合规治理，确保流通数据来源合法、隐私保护到位。
- 积极进行技术体系布局，在数据基础设施建设中采取跟随型和技术超车型布局策略，以确保在技术应用上取得领先优势，并在将技术转化为新兴产业、未来产业方面取得领先。
- 积极培育数据要素流通和交易服务生态，提供数据产品开发、发布、承销和数据资产的合规化、标准化、增值化服务，提高数据交易效率。

我国在数据战略和行动上全面布局，旨在通过数据驱动促进经济社会的高质量发展，并确保数据的安全和合规使用。

参考资料：

[1] 工业和信息化部.《“十四五”大数据产业发展规划》解读[EB/OL].（2021-12-01）[2024-05-07]. https://www.gov.cn/zhengce/2021-12/01/content_5655197.htm.

［2］ 邢萌，郭冀川．中国数据要素政策大事记［EB/OL］．（2023-03-10）［2024-04-07］．https://36kr.com/p/2496301063083906.

［3］ 新华社．中共中央 国务院关于构建数据基础制度更好发挥数据要素作用的意见［EB/OL］．（2022-12-19）［2024-06-07］．https://www.gov.cn/zhengce/2022-12/19/content_5732695.htm.

［4］ “国家数据局”微信公众号．十七部门关于印发《“数据要素 ×”三年行动计划（2024—2026 年）》的通知［EB/OL］．（2024-01-05）［2024-03-07］．https://www.cac.gov.cn/2024-01/05/c_1706119078060945.htm.

4. 日本

日本积极融入数字变革时代，并提出数字社会愿景，每一位国民都可以通过数字技术选择满足自身需求的服务，大力推动“数字化惠民，不让一个人掉队”，强化国际竞争力，实现经济健康可持续发展。

在具体的战略制定上，日本政府在 2021 年发布了三大数字战略，包括《综合数据战略》《科学技术创新基本计划 2021—2025 年》和《综合创新战略 2021》，旨在建设世界顶级数字国家所需的数字基础，明确数据战略的基本思路，并制定社会愿景及实现该愿景的基本行动指南。日本积极构建数据生态架构，从基础设施到价值体系设置 7 个分层（基建层、数据库层、连接层、应用层、规则层、社会层和价值层）。分层的核心脉络是从数据采集、存储、加工、分析再到应用服务、价值再生产的完整产业链。在数据的流通利用治理政策上，日本侧重于数据在国内市场、产业中的流通利用，同时也在探索如何打通国内和国际的双数据流通，成为“自己的数据能在世界范围内被放心利用”“世界的数据能在日本放心存放”的社会。日本还积极推动并倡导数据的国际化流通，提出了

“可信数据自由流动”（Data Free Flow with Trust，DFFT）倡议。

为推进数据战略落地、促进数据流通利用，日本采取了一系列行动，包括：成立日本数字厅，负责推进数字社会进程和数字化转型；为推进数字制度建设，开发了工具“数字经营改革的评价指标”（DX 推进指标），并成立了数字化转型（DX）加速研究会；政府部门牵头促进数据流通利用，通过内阁各部门协同推动、链主企业引领带动、行业协会推广示范等举措，不断完善本国数据流通网络；鼓励多层次数据交易市场发展，创新构建了“个人数据商店”和“数据银行”模式，探索个人数据收益分享制度，并建立了综合性数据交易机构，以促进各行业的数据开放共享；积极倡导数据自由流动，与欧盟建立数据流通合作关系，实现了欧日之间的双向互认；推动建立数据空间，如政务数据 GBiz 平台、农业数据 WAGRI 平台、防灾信息 SIP4D 平台等，以实现企业间数据安全、可信、高效的流通共享。

参考资料：

［1］ 国际网络观察员．日本官方发布《2022 先进数字技术制度政策动向报告》战略政策全扫描［EB/OL］.（2022-07-21）［2024-04-07］. https://mp.weixin.qq.com/s?__biz=MzI4OTkyNDgxNA==&mid=2247718463&idx=8&sn=fddd3f957b68b34c8faaaf2792f12bd6&chksm=ec2a659edb5dec88f28fe4158f609497ebbfae8a6bb9c331770aadfabceee0dd1aa841e86db9&scene=27.

［2］ 村井纯．日本数字厅所推进的数字化转型：日本的数字化政策并非失败［EB/OL］.（2022-01-08）［2024-04-07］. https://cn.japanpolicyforum.jp/society/pt20220108210103878l.html.

［3］ 走出去智库（CGGT）战略研究部译．日本观察丨日本数据战略及

双边、多边数字政策分析［EB/OL］.（2023-04-14）［2024-02-09］. https://baijiahao.baidu.com/s?id=1763118008937881362&wfr=spider&for=pc.

［4］ 数据信任与治理. 欧盟、日本、韩国的数据相关政府机构简要对比［EB/OL］.（2023-08-24）［2024-03-07］. https://www.secrss.com/articles/58103.

2.4 小结

数据通过流通才能充分释放价值。由于自身的非排他性和无损复制性，数据的流通与其他要素相比面临诸多不同方面的要求。站在数据主体的视角，相比于资本、土地等要素，既要担忧流通的安全与风险，又要考虑提供数据后的权益与责任。世界不同经济体也在积极设计并推进有利于数据流通和促进发展的各项措施。由此可见，需要从不同层面考虑数据流通的可靠实现，既要安全与主权，又要有效利用。这不是一个简单的技术问题，也不是一个靠机制和规则就能解决的问题，而是需要进行系统性的工程设计。

3

第 3 章 CHAPTER

数据流通难题的欧洲实践

数据要通过流通才能充分发挥其潜在价值，但如前文所述，因数据自身的特点，数据流通会导致利益主体面临其他要素所不需要应对的挑战。究其本质，数据流通时最需要应对的难点是数据的非排他性和易复制性。为了有效应对这些难点，最关键是实现可控的数据交换，欧洲以数据空间构建了数据可控交换的系统性架构。

本章根据国际数据空间协会及数据空间支持中心（Data Space Support Center，DSSC）、大数据价值协议（Big Data Value Alliance，BDVA）等欧洲数据空间研究及支持机构公开的内容，尽可能地对数据空间的来源、关键设计、在欧洲的定位、当前应用及推进情况等方面进行“原味”的呈现。数据空间在欧洲正进行得如火如荼，当前欧洲已经构建了相对成熟的运作机制，并在越来越多的行业中

形成数据交换空间。

3.1 数据空间的由来

3.1.1 从“工业数据空间”到“国际数据空间”

主权保护的“数据空间”理念主要起源于欧洲。随着技术和商业经济的发展，欧洲商业结构均进行着广泛的转型，数字化、自动化、个性化和去中心化趋势明显。参与者的倍增以及公司基础设施的日益分散，导致了协作和协调的难度增大，对个性化、按需提供产品和服务以及混合产品产生了新的需求。这些物理产品与数字服务的结合影响着商业模式、生产和供应链。

在此过程中，随着数据成为业务的焦点，其角色也从过程结果和产品使能者转变为独立的驱动力。为支撑工业 4.0 的战略发展，2014 年年底，德国联邦教育与研究部提出了“工业数据空间行动”（图 3-1），在工业 4.0 项目下启动相关研究，旨在构建一个安全可信的“工业数据交换空间”，帮助消除企业在数据流通和交易中的信任问题，并由德国弗劳恩霍夫协会牵头负责。

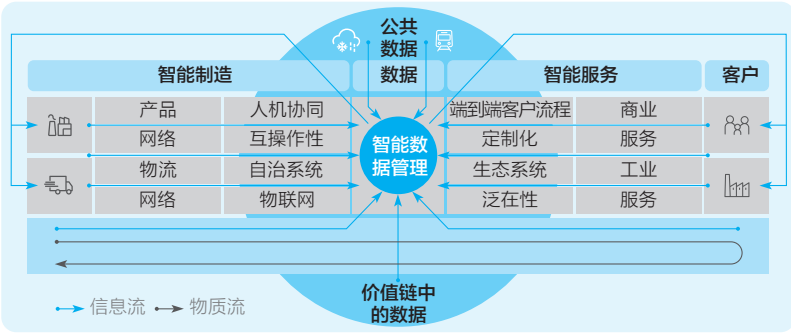


图 3-1　工业数据空间的定位和价值

为了落实“工业数据空间行动”，加强企业间的合作，汇集更加强大的科研力量，在德国各界专家人士的推动下，弗劳恩霍夫协会联合安联、拜尔与大众等 18 家企业于 2016 年 1 月成立“工业数据空间协会”（Industrial Data Space Association，IDSA），共同助力构建“工业数据空间”。

2017 年，工业数据空间协会正式发布《工业数据空间参考架构模型 1.0》，而此时距离“工业 4.0”理念的提出已过去几年，工业领域对信息化的需求远超预期。协会认为，数据共享不应局限在工业领域，而是要辐射到整个产业链，并将其推广至欧洲层面。2019 年，“工业数据空间协会”正式变更为“国际数据空间协会”（International Data Spaces Association，IDSA），覆盖范围也从工业扩展到多个领域。之后，IDSA 持续吸引更多国际成员加入，影响力也逐渐从欧洲扩展到全球。

3.1.2　数据空间的基本理念：保护“数据主权”

从数据空间的推进过程来看，IDSA 从一开始就专注于促进商业生态系统中可信和有主权的数据共享。数据空间是一种数据流通共享的新理念。

- 从计算机科学视角看，数据空间以“分散自治 + 协同互操作”的联邦数据体系架构为核心技术基础，无需核心数据的中央存储，而是通过分布式技术为数据主权保护与跨主体信任构建提供功能支撑。
- 从商业的角度来看，因为数据空间是数据合作的多边形式，所以欧洲交通出行、制造领域形成了社区，并在《欧洲数据战略》的要求下拥有了特定领域的数据空间。

- 从应用实践层面看，数据空间依托并适配现有的标准与技术，并结合数据经济领域已获广泛认可的治理模型，在可信的商业生态语境下，促成安全、合规且标准化的数据交换与共享，从而为创建智能服务场景和促进跨公司业务流程创新提供基础，同时保证数据所有者的数据主权。

数据空间提升了合作效果，降低了创新壁垒并更易实现创新，加速了数据价值链上的数据交换和数据处理，从而释放了数据的价值。

如图 3-2 所示，数据空间旨在为数据提供方和数据消费方基于双方的合作约定达成数据交换，且始终保护数据主权。数据所有者和数据提供方通过数据连接器将数据传输到数据空间，并允许其他人使用数据，同时保留一定的控制权，如谁能使用、如何使用、何时使用、何地使用以及以什么价格使用。这是数据主权的体现，也是释放数据价值的基础。数据消费方接收数据提供方提供的数据，并根据其使用策略使用数据。

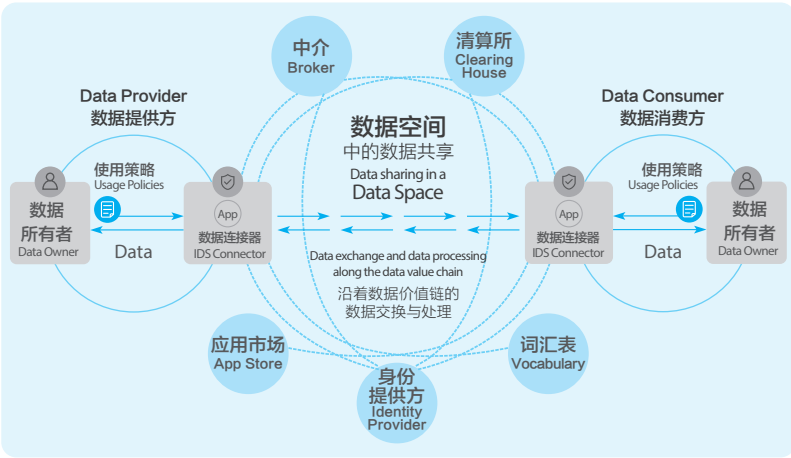


图 3-2　数据空间示意图

通过数据空间，数据消费方对数据的质量和可靠性充满信心，这也是数据主权的体现。从交换逻辑的角度看，数据连接器是一个专用的软件组件，允许参与者将使用策略附加到数据空间中的数据上，执行使用策略并无缝跟踪数据来源。数据连接器充当数据和服务的网关，并作为应用程序和软件的可信执行环境。

3.2　数据空间的 4 个设计原则

数据空间的出发点在于增强数据流通互信，它在设计上始终围绕在保障数据流通交换的各方权益的基础上增强互信。从设计原则上看，相比于平台型或者中心化的数据共享交换平台，数据空间更强调基于多方共识的合约治理机制，以及通过技术手段保障数据所有者对流通过程的自决能力。图 3-3 简要概括了数据空间的设计原则。



图 3-3　数据空间的设计原则

四个设计原则，从流通信任的基本点（即使用可控）出发，强调数据及质量保障的环境和技术架构要求，并对持续治理和发展的模式给出指导。以下将展开介绍各个设计原则。

3.2.1　你决定你的数据如何被使用

原则 1 是数据主权，即基于增强透明度和保护数据主权的原

则，为用户提供全新服务。

GDPR 授予个人决定数据收集者可以对其个人数据做什么和不能做什么的权利，同时欧洲数据空间也将提供用于行使这些权利并保持对这些数据的控制的工具。然而，欧洲数据空间不仅对个人如此，对公司 / 组织及其数据也是如此。在特定部门需求的驱动下，数据空间将促进适用于它的开发权的行使。数据主体可以使用这些工具给予或撤销其同意，并更改访问权限，以及指定访问和使用其数据的新条件。

此外，数据主体可以选择将数据权利的管理外包给第三方（例如数据中介机构），就像今天的用户（个人和组织）将其财务余额的管理外包给金融机构一样。这些工具将使数据主体能够始终要求其数据存储在哪里和要求访问权限方面的透明度。

3.2.2 以数据和服务质量为基础的公平竞争环境

原则 2 是数据的公平竞争环境，因为数据共享和交换需要平等的环境。

所有数据空间参与者都有一个公平的竞争环境，意味着新进入者在寻求进入数据空间时不会面临不可逾越的障碍（例如数据生态系统的准垄断结构）。在公平的竞争环境中，参与者竞争的是它们的数据和服务质量，而不是它们控制的数据量。如果生态系统由合作而不是竞争的理念主导，就可以出现一个公平的数据共享和交换环境。这可以通过对底层数据空间软基础设施的健全设计和彻底维护来实现。

3.2.3 去中心化、技术中立的基础架构

原则 3 要求数据空间是去中心化、技术中立的基础架构，并

可在现有技术标准上提出进一步的要求。

数据空间的基础架构将不是一个单一的集中式 IT 基础架构。它将由符合一系列功能、技术、操作和法律方面协议的数据空间的整体可互操作组件组成。从技术角度来看，软基础架构可以被视为可互操作的基于 API 的 IT 平台的集合，用户通过高级身份和统一管理机制控制数据流。软基础设施的设计将包括数据共享和交换交易（即数据货币化）机制。数据空间的软基础架构将是技术中立的，所有行为者都可以自由地根据其工程能力做出自己的选择。数据空间的参与者也必须能够控制它们与谁共享哪些数据。

3.2.4　公共与私营组织共同治理

原则 4 要求数据空间需要公共与私营组织共同治理。

为了设计、创建和维护数据领域的竞争环境，良好的治理至关重要，所有利益相关者都需要有代表和参与的感觉，包括数据服务的用户（个人、企业）或提供商以及它们的技术合作伙伴和专业人员。单一的政府或市场机制往往无法有效解决公共问题。公共干预意味着首先需要指明正确方向，然后调动公共和私人力量来实现相应的目标。

政府、非政府组织、私营企业等各方共同参与和合作，共同制定政策、实施项目和提供服务，以推动社会经济发展和满足公众需求。《数据治理法案》确认了战略、战术和运营三级机构的公私合治模式：每个数据空间都有一个治理机构，并由一个中央治理机构监督它与 N 个数据空间的互操作性相关工作；欧洲数据创新委员会为战略层面的权威机构；在战术和行动层面引入一个权威机构“数据交换委员会”（DEB）。

3.3 五层三维的数据空间参考架构

由于数据交换是以数据为核心的使用和处置权的交换，因此非常有必要考虑不同利益相关方的诉求。IDS 参考架构模型符合通用系统架构模型和标准（例如，4+1 视图模型，ISO 42010），采用五层结构，以不同的视角表达不同利益相关方的关切点和核心观点。IDS 参考架构模型（Reference Architecture Model，RAM）定义了业务、功能、流程、信息、系统五个层级模型，以及安全、认证、治理三个维度，构建了一套完整的数据合约及数据主权保护架构，如图 3-4 所示。

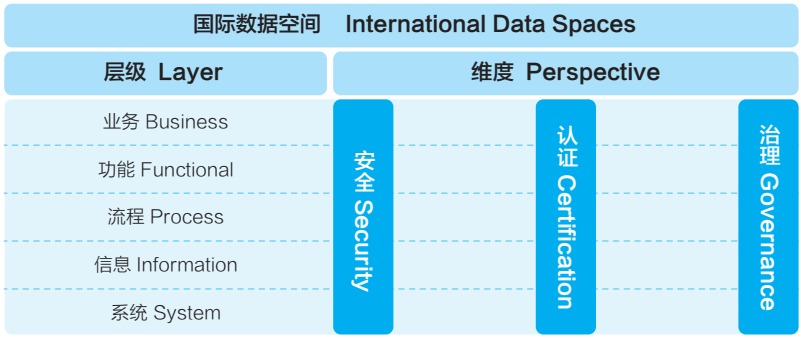


图 3-4 IDS 参考架构模型分层图

IDS 参考架构模型面向复杂信息系统，通过层次化的架构促进数据空间内各种组件的标准化。同时从安全、认证和治理的角度，考虑数据空间运行的设计要求，整合业务、功能、流程、信息和系统层面的衔接点，以提高未来数据空间的互操作性。

3.3.1 业务层：定义数据空间的 4 类参与方及协作关系

IDS 参考架构模型的业务层定义了数据空间的参与角色及它们之间的基本连接和互动关系。数据的流通交换是一项系统性工程，包括数据的准备、传输、使用、增值及最终的回收或处理，整个过

程包含了不同类型的角色。IDS RAM 业务层从达成数据流通交换的结果出发，尽可能地保障不同角色的利益。

IDS 定义了四类 15 种角色，主要有核心参与方、中介方、服务提供方及治理机构，如图 3-5 所示。

- 核心参与方是“数据提供者”和“数据消费者”，核心参与方的角色可以由任何拥有、想要提供和 / 或想要消费或使用数据的组织承担，所以还包括应用提供商。
- 中介方包括中间服务提供商、应用商店提供商、清算所等。
- 服务提供方是提供数据空间组件服务的角色，如数据连接器开发组织。
- 治理机构具有设置和执行标准化的数据交换、信任创建及最终实现 IDS 可持续操作的权力和任务。治理机构包括认证机构、评估机构、标准化组织和国际数据空间协会等。

不同角色之间的交换关系可以三条流区分，即如图 3-5 所示的数据流、元数据流、软件流。

- 围绕数据流的交互：可以直接接触到数据内容的参与方。直接的交互主要有两类：一类是数据提供者与数据消费者的内容交互；另一类是服务提供方对数据进行增值时进行的数据内容交互。
- 围绕元数据流的交互：主要目的是使数据提供者与数据消费者更好地达成交换，如数据寻源或探查需要市场中介方（如通过公共或者特定的数据市场寻找数据、引用第三方发布的数据标准确认交换数据的格式等）提供数据资源或产品信息，数据交换过程或结果需要清算运营方结合交换记录提供保障。这一过程仅需要元数据即可。

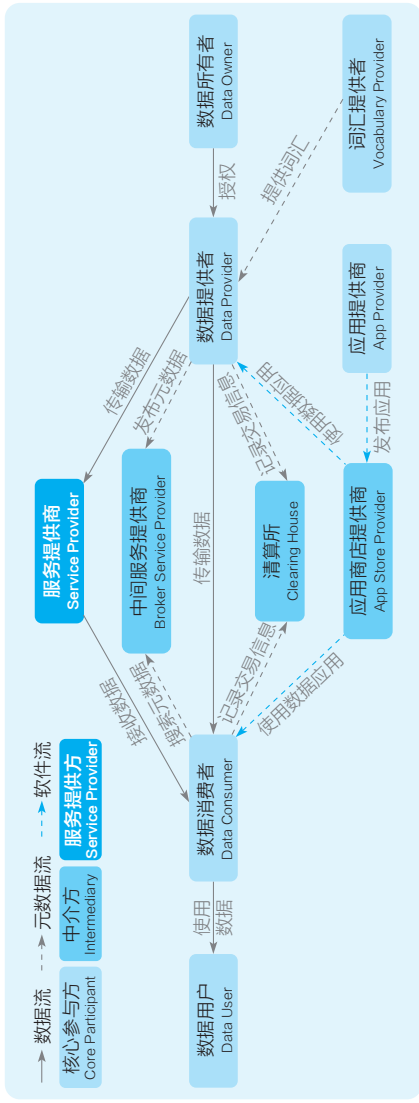


图 3-5　IDS 参与方及交互示意图

- 围绕软件流的交互：主要在数据增值方面，数据提供者或者数据消费者需要购买或下载对应的数据应用程序，以消费数据。

3.3.2　功能层：数据空间的 6 大能力特性

IDS 参考架构模型的功能层定义了数据空间的功能需求以及由此产生的待实施的功能，如图 3-6 所示。

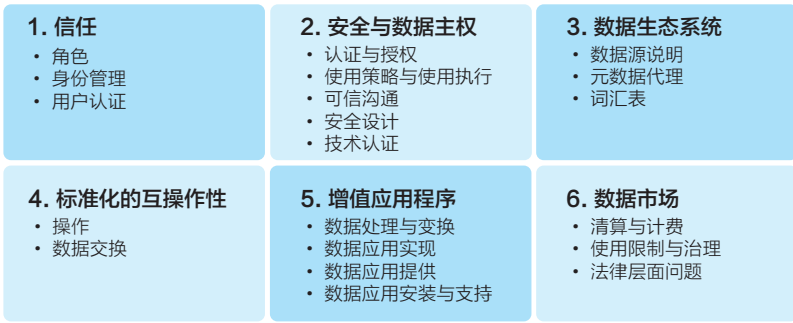


图 3-6　IDS 参考架构模型的功能层

- 信任：关键目标是保证身份是真实的、可信的。信任特性包括三个主要方面（角色、身份管理和用户认证），并通过治理视角进行加强。角色相关的能力和权限与数据交换的流程和规范是一致的。数据空间所有对象的身份遵循欧洲信任框架生成并进行管理。参与数据空间的用户需要通过认证规范要求，并实现与现实世界相关的信任锚（Trust Anchor）。
- 安全与数据主权：和安全与数据主权相关的需求代表了数据空间的基本特征，包含认证与授权、使用策略与使用执行、可信沟通、安全设计和技术认证。安全设计需要提供认证与授权的流程和功能特性。使用策略需要能够被有效

实施，以体现数据主权。数据空间中的信息通信需要符合相关的可信标准，且具有基于安全视角的设计。所有的实施技术应该通过相关的技术认证。

- 数据生态系统：需具备描述、查找和正确解释数据是数据空间生态系统的关键的能力，主要包括数据源说明、元数据代理和词汇表。数据源说明是数据的来源元数据，主要用于定向不同数据接入点。元数据代理是对要参与交换的数据的描述，方便交换参与方理解数据。词汇表是数据交换的标准术语，为了满足不同参与方的交互需求，在保障数据交换主干业务流的基础上，可以有不同的标准。
- 标准化的互操作性：参与者之间的标准化数据交换通过数据连接器组件承载。数据连接器应兼容不同的交换标准协议以实现互操作。
- 增值应用程序：提供数据的加工、处理及数据应用程序的特性，要求对应的数据应用程序可以处理和转换数据。数据空间应该具备数据应用程序的上架、安装及持续的运营支撑服务。
- 数据市场：在数据空间中需要统一数据市场概念。数据市场能够对数据交换进行清算和结算，对数据交换中的策略进行验证，并能在法律层面对市场的交易 / 交换规则进行落实。

3.3.3 流程层：定义数据交换的 5 大步骤

IDS 参考架构模型的流程层明确了不同组件之间的互动方式，包括 5 个关键的业务流程。

1. 登入流程

该流程明确了数据提供者或数据消费者如何进入数据空间。

首先，数据提供者或者消费者需要完成组织的认证和注册，并获取经过 IDS 认证的数据连接器以进行数据交换。在此基础上，参与方需要进行连接器的可用性配置，将身份提供程序所需的信任锚（例如 CA 的根证书）调配到连接器上，以启用对通信合作伙伴身份信息的验证。每个连接器应提供自述，供其他 IDS 参与者访问。

2. 数据供给流程

该流程描述了如何提供数据或寻找合适的数据。数据提供者首先登入数据连接器创建数据资产自描述信息，然后发布数据资产自描述信息，可以通过数据连接器开放端口发布，也可以发布到元数据代理节点（如数据市场）。数据消费者选择合适的元数据代理即可搜索数据资产。

3. 数据合同协商流程

供需双方通过协商使用控制策略来实现数据交换，合同要约描述了数据提供者在什么条件下愿意将数据提供给数据消费者。供需双方通过数据连接器完成使用控制策略的协商。数据提供者将数据策略添加到合同要约中，并通过数据连接器通知数据消费者。数据消费者接收后可以更新数据合同内容，并发给数据提供者。经过此过程的反复，最终达成双方均可接受的数据合约。

4. 交换数据流程

该流程定义了如何在数据空间参与者之间传输数据。数据合约达成后，供需双方将根据合约履行数据交换。数据连接器执行数据合约，可以采取异步或者同步的方式。在订阅的情况下，数据消费者可以要求数据提供者更新其所请求的数据，连接器也根据响应的更新条款进行更新（如按一定的时间周期）。

5. 发布和使用数据应用程序流程

该流程明确了提供应用程序到数据空间或者使用数据空间中的数据应用程序的操作要求。通常情况下，应用提供商将 App 发布到相应的中介节点（如 App Store）。App 发布需要获得认证机构的认证，App 提供方需要将 App 镜像推送到 App Store 进行注册，并发布 App 元数据。对于成功发布的应用，相应的元数据和应用图像均存储在应用商店中，参与者可以通过应用商店提供的搜索接口进行检索。

3.3.4 信息层：数据空间的六边形模型

IDS 参考架构模型的信息层定义了一个利用 Linked-data 原则来描述 IDS 组件动态、静态特征的概念模型——信息模型，它是与领域无关的通用语言。信息模型是数据空间参与者和组件共享的基本协议，实现了兼容性和互操作性，支持对数据资源的描述、发布和标识。一旦识别出相关资源，参与者即可通过发现服务来交换和消费它们。此外，信息模型还描述了数据空间的基本组成部分、参与者、基础设施组成部分等。

信息模型也被称为六边形模型，如图 3-7 所示。

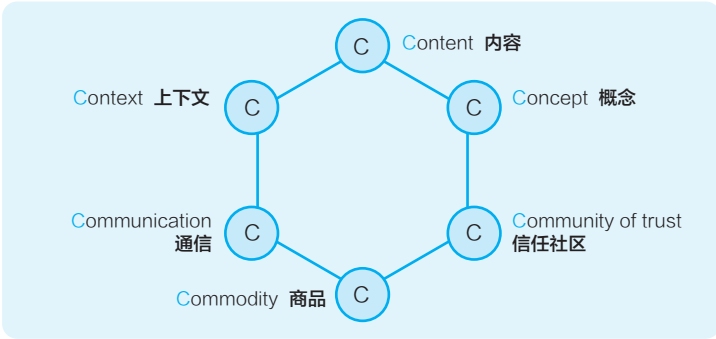


图 3-7　信息模型

Content 用于回答以下问题：该资源提供什么类型的内容（文本、音频、视频、图片、软件、Opaque、Container）、内容是什么格式（JPEG、SVG、PDF 等）、有没有内容样例、文档的大小和创建日期。

Concept 用于解释如何理解其他维度中使用的术语：Keyword 是使用非正式语言标签解释一个实体；Term 是可以复用的标准词汇，可在不同领域中通用，如 URI；Type 是使用概念模式给实体分类。

Community of trust 是 IDS 为维护数据主权和安全、可信的数据交换所特有的属性：数据交换双方的组织描述（包括数字身份 X.509、描述信息 URL、公司地点、业务、认证等级等）；双方数据连接器的部署信息（本地或云上）、安全档案、资源目录、服务器地址等；认证机构是什么，认证等级是什么（认证等级在参与者元数据描述文件以及 X.509 中）；数据的使用控制策略是什么。

Commodity 用于评估资源的价值和价格：资源是否来自可信源，采集的背景是什么，资源质量如何，资源的使用有什么限制，资源如何定价。

Communication 考虑用何种方式来交换资源内容，回答的问题有：客户提取信息有什么前提，通信使用哪种协议，一个合法的请求是什么样子，处理请求的端点地址是什么。

Context 包含与资源相关的时间、地理（经纬度）、现实世界实体（例如，文艺复兴时期、某建筑等）等信息，回答的问题有：内容覆盖的时间段是什么，什么时间在哪里收集的数据，数据是大实体中的哪个子实体。

3.3.5 系统层：以连接器为核心的 6 大组件

IDS 参考架构模型的系统层把业务层中的角色映射到数据架构和服务架构上，从而保证用户在功能层提出的要求得以实现，这也是 IDS 的技术核心。系统层包括逻辑软件组件的集成、配置、部署、扩展性等方面。核心组件包括以下六个。

1. 身份提供服务

参与者的身份识别和访问管理（IAM）包括定义标识（即声明身份）、身份验证（即验证身份）和授权（即基于身份做出访问决策）。IDS 中的身份提供者由三个互补的组件组成：证书颁发机构（CA）负责颁发和管理技术身份声明；动态属性提供服务（DAPS）提供短期令牌，其中包含有关连接器的最新信息；参与者信息服务（Participant Information Service）则以机器和人类可读的方式提供 IDS 参与者的业务相关信息。

2. IDS 连接器

IDS 连接器允许参与者通过它发布的数据端点交换数据，不需要通过中央实例来存储数据。IDS 连接器必须能够被来自其他组织的 IDS 连接器访问。一个参与者可以在本地或云环境中操作多个 IDS 连接器。IDS App 是 IDS 连接器内部用于实现业务逻辑的服务，可用于处理数据、连接外部系统或控制 IDS 连接器。IDS App Store、元数据代理和 Clearing House 均基于 IDS 连接器架构，以支持安全可信的数据交换。

3. 应用市场及数据应用程序

IDS App 是一个独立的、功能性的、可重用的软件资产，可在 IDS 连接器上部署、执行、管理。IDS App 包括三种不同类型的

App——数据 App（以某种方式操作可用的数据）、适配器 App（访问企业信息系统）和控制 App（允许从外部系统控制连接器，需要针对特定版本连接器的相应 API 进行编程），每种 App 负责在 IDS 系统中执行不同的任务。所有类型的应用程序都可以通过 IDS 连接器进行下载和完全管理。

4. 元数据代理（Metadata Broker）

元数据代理用于数据资产自描述信息的注册、发布、维护和查询。连接器可以将其自描述发送到元数据代理中，每个参与方都能通过元数据代理找到对应的数据资产。在一个数据空间内，可以有多个元数据代理，且允许分发元数据代理功能。元数据代理还必须具有有效的 IDS 标识，并在通信中使用有效的 DAT。元数据代理的大多数用例都需要面向人的自描述接口，同时必须确保网站用户和连接器的权限。

5. 清算中心（Clearing House）

清算中心的功能是基于记录进行清算、计费，以及提供与使用控制策略相关的日志服务。

6. 词汇中心（Vocabulary Hub）

IDS 中的互操作性要求使用常用的标准化术语来描述数据、服务、合约等，这些标准化标识符的集合构成了数据空间的词汇表。词汇表中的术语必须是机器可读的，IDS 信息模型是所有参与者共享的词汇中心，也是所有 IDS 组件必须理解的最小术语集。在特定领域中，需要使用额外的词汇表扩展基本信息模型，并应提供与核心词汇表相同的实现方式。因此，IDS 词汇中心是负责托管、维护、发布和记录额外的词汇表的节点。

数据空间依赖不同参与者的连接建立分布式网络。IDS 连接器负责数据交换，并按照自描述中的规定向元数据代理提供元数据，例如数据 Offer 的使用控制策略。合约通过连接器传输到清算所，清算所记录数据传输记录。连接器通过词汇中心获取词汇表。用户可以将其他 IDS 应用程序下载到 IDS 连接器中，以对数据进行操作。IDS 参考架构中的组件说明如图 3-8 所示。

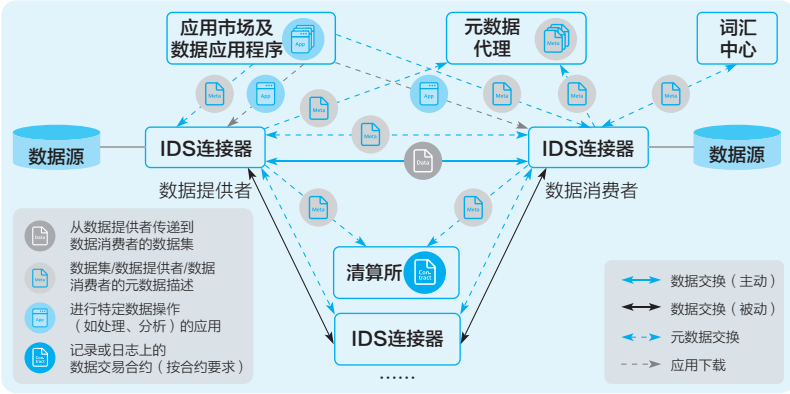


图 3-8　IDS 参考架构中的组件说明

3.3.6　安全维度：基于最佳实践的 6 个安全设计

IDS 安全架构可提供识别参与者的方法，保护通信和数据交换，并控制数据的使用，最终提供可信任的数据供应链。基于安全视角的设计及实现原则是“尽可能使用现有标准和最佳实践”和“支持安全级别的可扩展性”。IDS 安全维度包含了 6 个设计。

- 通信安全：使用连接器时，有两层安全措施，即使用加密通道进行点对点通信与基于实际交流端点进行端到端的鉴定授权。
- 身份管理：声明身份、鉴权验证身份、基于身份允许访问。

- 信任管理：IDS 利用密码学方法建立参与方之间的信任。公钥基础设施（PKI）的一个中心原则是每个实体都被分配密钥，让每个实体都能鉴定其他的参与者。
- 信任平台：连接器之间要能互相理解，并验证对方的安全画像；连接器上的所有 App 用容器进行隔离；连接器要能够进行远程完整性验证。
- 数据使用控制：数据使用控制作为访问控制的延伸，需通过附加数据使用策略，保证数据的使用符合要求。
- 数据溯源：数据供给记录跟踪了数据在何时、以何种方式、被谁修改等新数据产生的过程；核心目的是保证透明和可追溯。

3.3.7　认证维度：个人、组织与组件的安全合规

认证维度要求所有参与 IDS 的个人组织，以及核心组件都要经过认证，包含三个阶段和两个方面。

认证维度包括的三个阶段：

1）申请阶段：启动 IDS 评估和认证流程。

2）评估阶段：根据确定的评价标准对申请人或核心组件进行评价。

3）认证机构审查评价报告：如果评价的结果是积极的，认证机构就会颁发证书。

认证维度包括的两个方面：

1）参与方认证：评估每个参与者是否实现了规定的安全级别，包括其引用的基础设施的可靠性和流程的合规性。

2）核心组件认证：核心组件必须提供所需的功能和安全级

别，认证重点是互操作性和安全性，且需要遵循 IDS 定义的开发和维护流程。

3.3.8 治理维度：以“联邦”的形式实现数据空间的安全可信交互

IDS 以“联邦”的形式，将数据的相关权力分配给不同的角色，以实现整个生态的安全可信交互。IDS 参考架构的治理维度从治理和合规性方面定义了国际数据空间的角色、功能和流程。因此，它重点表明如何在 IDS 参考架构的每一层上定义治理的核心问题。当前主要定义了如下几类治理问题：

- 为数据交换、互操作性和新的数字化业务模式的使用提供基础设施。
- 在数据所有者、数据提供者和数据消费者之间建立可信赖的关系。
- 担任参与者之间调解的受托人。
- 促进协议和合同的谈判。
- 提高数据交换和数据使用的透明性与可追溯性。
- 允许私人和公共数据交换。
- 考虑参与者的个别要求。
- 确保加入的数据空间服务满足去中心化架构。

3.4 数据空间已从理论走向实践

3.4.1 理论体系已经相对完善

IDS RAM 从最开始的白皮书到完整的参考架构，截至 2024 年已经迭代了 4 个版本，如图 3-9 所示。

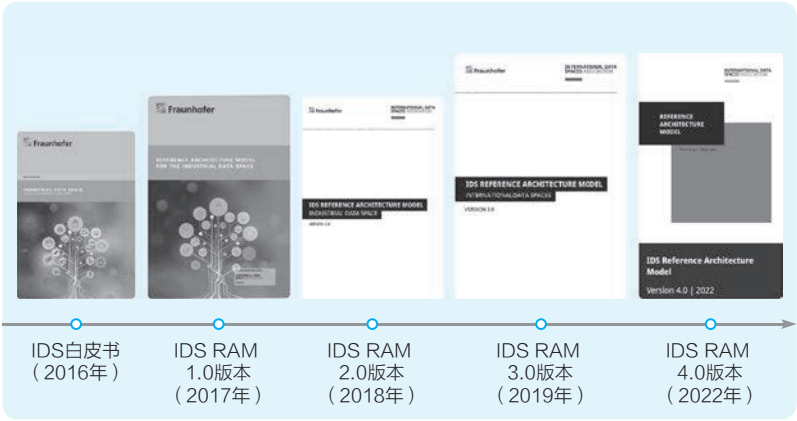


图 3-9　IDS RAM 演进时间图

IDS RAM 4.0 相较于前面的版本，重点在于将数据交换的管理控制逻辑与数据集成逻辑解耦，从而实现对现有集成协议的开放兼容。

2024 年，IDSA 规划了 IDS RAM 5.0 版本的更新，5.0 版本将从互操作性、多样化的系统资源架构、更加标准和模块化的设计等方面，如数据空间数据交换互操作上的推进，对连接器的对接协议、身份与安全方面的共享协议输出标准协议，在信任框架中支持去中心化、联邦架构等多种形态等。

3.4.2　数据空间生态已经成型

国际数据空间协会（IDSA）是一个致力于创建数据空间标准的非营利组织，旨在确保在特定治理框架内参与者之间安全的数据共享，以促进信任并维护数据主权。IDSA 的使命是创造全球数字经济的未来，通过国际数据空间（IDS）实现安全、有主权的数据共享系统，使所有参与者都能实现其数据的全部价值。截至 2024

年上半年，国际数据空间已吸引全球 150 多个组织、850 多名行业专家投入研究并推进数据空间建设，数据空间相关的研究和分支机构增长至 15 家。

在 2023 年 7 月 5 日的 2023 全球数字经济大会上，IDSA 中国能力中心正式成立，这是 IDSA 在中国的代表机构，由下一代互联网国家工程中心（CFIEC）负责运营。该中心的使命是成为国际数据流通的“连接器”，致力于为全球数据空间技术和标准贡献中国力量，并构建国际数据共享流通的生态体系。

3.4.3 领先数字企业布局数据空间配套服务

在 2024 年 3 月数据空间全球年会上，谷歌、亚马逊、T-Systems、英伟达等厂商积极布局数据空间配套服务。

- 谷歌基于欧洲 EDC 开源项目开发能与自己的数据产品集成和协同的配套服务，为数据空间提供市场、审计和算法服务。
- 亚马逊将数据空间的数据连接器纳入其服务的提供范畴，并与其数据分析应用、AI 和机器学习服务完成集成。
- 作为欧洲领先的 IT 服务提供商的 T-Systems，是最早加入 IDSA 并提供全套数据空间服务的提供商，积极参与了数据空间产品的构建。
- 英伟达将数据空间与 AI 紧密结合，它认为未来在 AI 的驱动下，数据空间是高效、标准化地获取高质量语料的有力支撑方案。

3.4.4 初具规模的业务应用

欧盟数据空间的发展及应用分为 5 个阶段，分别是探索阶段、准备阶段、实施阶段、运营阶段、规模使用阶段。

- 探索阶段重点考虑数据空间带来的利益、潜力和变化点，包括识别利益相关者、收集数据交换需求及场景、识别现有模式的差距等。
- 准备阶段是当一个数据空间倡议有了足够数量的合作伙伴，并且一致同意推进该倡议时，由合作伙伴共同开发用例并实施数据空间。
- 实施阶段是数据空间计划有足够详细的项目计划、里程碑和资源。
- 运营阶段是数据空间计划有了经过测试的基础设施和治理框架，并且第一个用例开始运行。
- 规模使用阶段是数据空间计划已被证明能够获得新的参与者并始终如一地、有机地接受新的用例。

目前，欧盟数据空间的发展可以在 IDSA 的数据空间雷达中得到体现，数据空间雷达是欧盟数据空间的中心存储库，收录了全球各种数据空间，目前包含 18 个不同行业的数据空间，每个数据空间均可以动态查阅（参见 https://www.dataspaces-radar.org/radar/）。

从推进方式上看，欧盟数据空间正在进行三大类数据空间的布局：第一类是通用数据空间，由《欧盟数据战略》或欧洲国家政府支持建设，数据来源于多个行业主管部门；第二类是行业数据空间，在《欧盟数据战略》的指导下，由行业龙头主体基于发展探索而形成；第三类是点对点的场景化测试床应用。

3.5　数据空间写入《欧洲数据战略》及《数据法案》

3.5.1　《欧洲数据战略》：建立 9 大数据空间

2020 年 2 月 19 日，欧盟委员会发布《欧洲数据战略》（*A European*

Strategy for Data），包括背景介绍、关键点、愿景、问题、战略、国际路径、结论以及附录八个部分。

《欧洲数据战略》概述了欧盟未来五年实现数据经济所需的政策措施和投资策略，明确提出要在欧洲战略部门和公共利益领域建立数据空间，并明确了 9 大公共数据空间（工业、绿色可持续发展、出行、医疗卫生、金融、能源、农业、公共行政、技能）的建设计划。

《欧洲数据战略》提出，数据空间将促进数据在欧盟境内跨行业自由流动，落实欧盟的规范和价值观（如个人数据保护法、消费者保护法和竞争法），确保数据访问和使用规范公平、切实可行且清晰，并拥有明确和可靠的数据治理机制。欧洲则通过数据空间实现积极开放的国际数据流动。

3.5.2 《数据法案》：将数据空间定位为基础设施和数据治理机制

2023 年 11 月 27 日，欧盟理事会正式通过了《数据法案》（*Data Act*），并于 2024 年 1 月 11 日正式生效。

《数据法案》的主要目的是促进生成数据的重复利用，提高欧盟行业通过联网产品和关联服务产生的数据的价值，为数据驱动的创新提供更加开放的市场环境。《数据法案》明确提出，通过数据空间提升欧洲云服务市场的竞争水平，提升数据的可用性并建立一个值得信赖的环境。

《数据法案》将数据空间定位为基础设施与数据治理机制，并提出数据空间要通过统一的规则和标准，确保数据的安全和隐私保护，同时提升数据的可访问性和利用率。

3.6　小结

本章从数据空间在欧洲的发起、设计、参考架构、应用推广及数据空间在欧洲的战略定位出发，厘清数据空间的来龙去脉。欧洲数据空间的设定符合数据价值释放的大逻辑，基于数据的特点，从数据利益相关方的角度出发，系统性地设计开放、可演进、分布式、去中心化的参考架构。在推进策略和治理模式方面，欧洲数据空间也进行了长期的演进，从理论到实践，再从实践到研究，迭代推进，具有相当大的影响力。数据的流通交换受很多因素的约束，我们在看到其先进性的同时，也需要考虑我们实际所处的环境，进一步理解和分析 IDS 参考架构与实际业务场景的结合方法。

4

第 4 章 CHAPTER

数据空间的底层逻辑

数据空间是用于解决数据流通问题的系统化思考成果，欧盟 IDS 参考架构给出了较为详细的设计说明，欧洲推进数据空间也有接近 8 年的历程。从商业组织或者治理层面看，是否需要完全照搬或者是否可以在欧盟实践的基础上加速这一进程，是需要进一步思考的。同时，从参考架构和实践层面看，数据空间涉及的要素非常广泛，非常有必要解构数据空间的底层逻辑，避免数据空间被泛化。本章进一步结合华为实践，从数据流通面临的现实诉求入手，阐述数据空间的底层逻辑，给出数据空间的本质和关键特征，并详细描述数据空间的关键要素。

4.1 数据流通面临的现实诉求

作为一家业务形态多样的公司，华为形成了“集团 + 事业群 + 区域子公司”共同治理形态，并坚定实施生态战略，协同各行各业伙伴实现行业成功。在此过程中，华为各功能域、产业、区域子公司之间存在大量的数据交换。面向外部，华为对客户、供应商、生态伙伴及第三方组织也存在大量的数据交换诉求。如图 4-1 所示，华为存在多种跨不同主体的数据交换诉求。

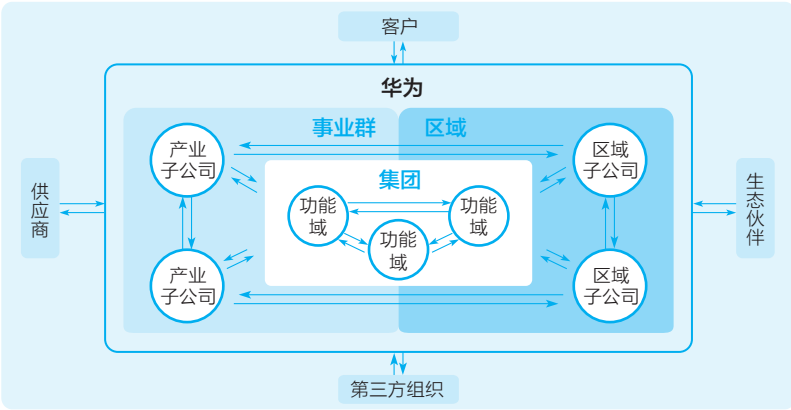


图 4-1　华为数据流通的类型

华为内部：集团功能域与功能域（如供应域需要获取财经域加工后的财务数据等）、功能域与事业群（如华为制造域与华为终端）、功能域与区域子公司（如财经域与德国代表处）等均基于业务流产生数据交换，以改进对应主体的整体能力。

华为与外部主体：主要包括华为与客户、华为与生态伙伴、华为与供应商、华为与第三方组织四大类跨主体交换数据的场景。

经过诸多业务领域的反馈和调研，在内外部数据流通中，有如图 4-2 所示的业务场景案例。

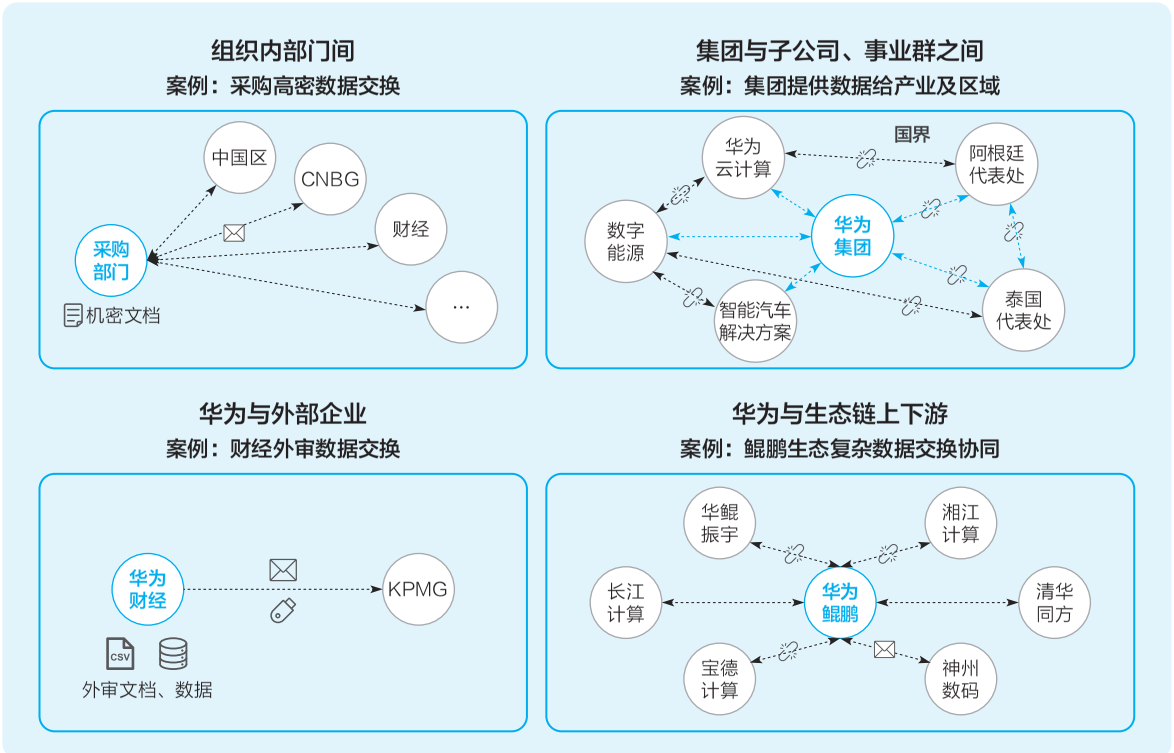


图 4-2　华为数据可控交换的业务场景案例

这些跨主体的业务场景，都是基于业务流已产生的数据，通过数据融合分析实现新的业务治理模式，但当前因为数据的使用风险，需要依赖数据可控交换的能力。

4.1.1 组织内部门间的敏感数据获取

内部功能域之间主要通过数据底座进行数据共享，但存在一些敏感、高密数据不适合通过数据湖共享，且使用场景也比较低频。针对此类数据，需要构筑安全且使用清晰的能力。如采购文档的获取，为推进业务合作，不同产业的业务人员需要获取历史或当期采购合同，但需要控制文档不被扩散以规避风险。当前的方式是，消费者申请文档权限，并从系统下载到本地查看，提供方难以追溯。作为管理文档的业务部门，迫切需要对数据授权后可视化的使用控制和全过程的追溯能力。

4.1.2 集团与子公司、事业群之间的可控数据共享

集团各功能域作为使能和支撑部门，提供各类原始记录和报告等数据给各业务组织。在此过程中，各功能域需要遵从内部监管，满足外部合规要求。同时数据供给需要满足业务相关性要求，仅提供与产业 / 区域相关的数据，按需切片，实现使用可控、去向可溯。另外，重要、机密或跨境数据需要遵循使用控制要求。产业与区域之间，也需要确保在业务范围内进行数据交换，如仅交换属于产业 / 区域的数据，并且从集团获取的数据不能再转给其他产业，各交换主体间须按需协商、分类管控（部分数据需要可控制、可追溯）。

4.1.3 华为与外部企业的数据交换

华为与行业外部组织存在业务合作，如外部审计、高校学

研合作等。当前的数据流通也亟须构建安全可控的通道，确保各方的业务信息或研究成果不被泄露。以财务外部审计为例，因业务复杂，华为每年需要 300 多名账务人员，通过邮件、U 盘等方式发送外审资料。数据给到审计公司后，华为依托 NDA 和行业自律约束来保障数据安全。外发材料的统计查证需要投入大量的成本。

4.1.4　华为与生态链上下游的数据协作

华为生态链上下游的数据复杂多样，交换价值高。双方均需要更多的数据来促进产品改进和提升客户满意度，但均有类似的安全考虑。因双边数据交换需要跨企业网络边界，双方达成的交换协议需要依托一定的技术保护手段来降低数据消费方泄密和失控的风险。在数据的安全存储、合规合理使用方面，双方都具备对等的能力。完成数据交换和使用后，双方均可反馈具体的使用情况，并由统一的规则将记录上报存证。

综上，在内外部数据流通交换过程中，数据在跨主体、跨边界交换时，双方需要在现有线下协议的基础上，采取各项措施强化数据交换的互信机制。此外，供需双方在按照协议完成数据交换并使用数据后，均可提供数据交换过程和使用过程的记录凭证，以进一步降低对数据使用风险的担忧。

4.2　在数字世界履行数据交换合约

4.2.1　现实世界的数据交换合约通过法律执行有滞后性

结合华为的业务场景，数据流通的各方参与者基于各自的数字化建设产生了相应的数据，双方基于业务合作需要进一步融合更

多数据以获取进一步的商业成功。但当前缺乏可控数据交换的能力，双方均存在诸多顾虑。关于数据的可控交换，需要厘清关键的难点和业务逻辑。可控数据交换的关键在于应对好数据流通过程中数据的非排他性和可复制性。在此之前，可以看看现有的数据流通过程是如何实现的。如图 4-3 所示，数据使用合同条款的有效实施通常通过法律合同和技术实施两个方面来保障。但当前的数据交换合约更多还是延续传统的数据集成模式，线下交换双方基于合作意向签署纸质合同，并约定相关的数据条款。

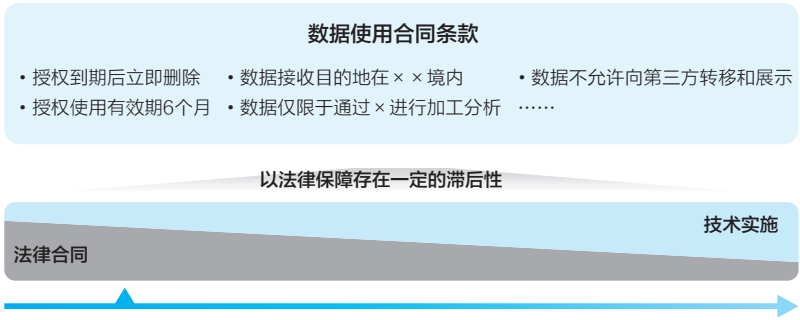


图 4-3　数据使用合同条款相比于技术保障更多依赖法律

现有的数据交换过程通常包括以下四个方面。

第一，现实世界的任何主体组织之间需要进行数据交换时，双方会基于法律要求、内部合规要求、业务使用场景等约定数据的范围、数据如何传输、数据的使用场景等，并签署一份具备法律效力的协议或者合同，以确保后续的数据集成。

第二，在现有的框架下，数据交换双方依据协议、合同由双方的 IT 团队进行数据集成，并明确采用的数据集成工具、对接流程、最终数据的存储和使用。数据一旦到了消费方以后，数据交

换合同的履行基本就依靠合同的威慑力、行业自律和事后的法律追责。

第三，在此过程中，真实的数据接收者与合同中的消费方是否一致、合约中的描述条款与实际的数据集成配置是否匹配、数据的存储与使用要求是否得到真实履行，数据提供方依靠纸质合同很难验证。

第四，影响数据交换的关键条款过多地将法律作为底线，通常发生问题时可能已经造成一定影响或损失，因此存在一定的滞后性。

上述过程可能导致数据合约的权益保护存在一些问题，如时间成本（提供方等待法律程序需要时间）、资源消耗（相关方需要投入相应的诉讼资源）、执行难度（需要根据结果协同相关方执行）、威慑效果有限等。因此，需要增强技术方面的合约履行能力。

4.2.2　基于数据合约的控制技术履行合同条款

仅仅依靠现有的法律框架定义商业数据交换合同，并基于零散的技术点实现数据交换，无法实现可信、高效的数据交换。数据空间的核心是将主体之间的数据交换活动标准化、数字化，把数据交换视作企业的标准业务活动（如同企业与企业之间的业务集成会参考 EDI 标准并指定相互认可的集成技术），围绕主体的共同关注点构建一套用于数据交换的数字化作业标准。其中，更为关键的是将现实世界主体之间的数据交换合同转化为数字世界可履约的数字化能力，并通过技术履行合同条款。主要思路如图 4-4 所示。

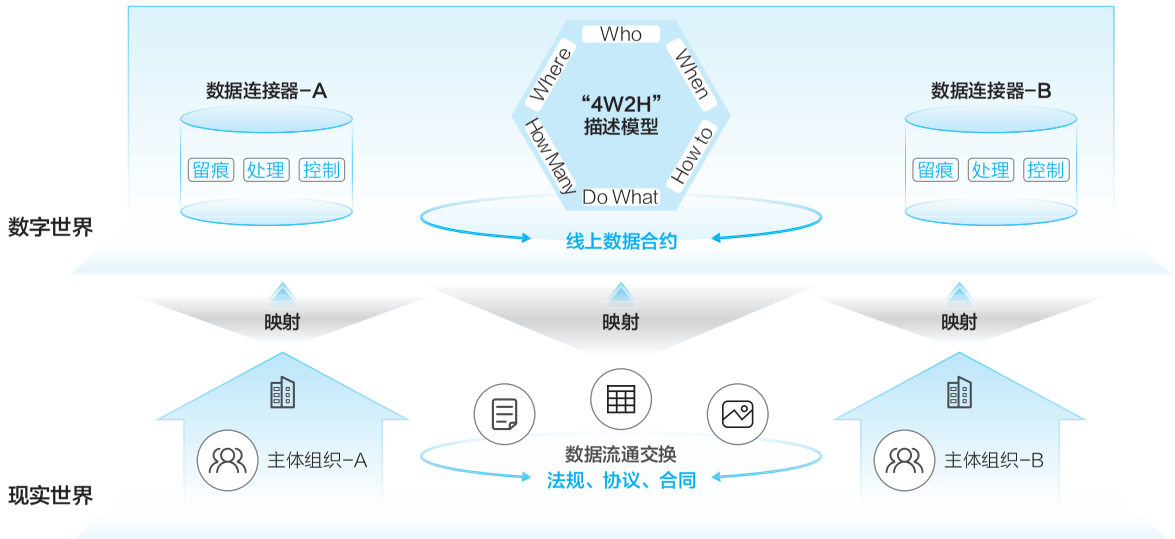


图 4-4　现实世界的数据交换映射到数字世界

1）为实现可控数据交换，需要向现实世界的主体组织、数据交换合同、数据的流动存储及使用提供可控的数字化能力，系统性提升参与方之间的信任。

2）在数字世界中，主体组织通过数据连接器实现数据交换的留痕、处理和控制。数据连接器是数据交换主体发出或者接收数据的载体，数据在连接器中被留痕、处理并受控使用。

3）数据连接器对线下的数据合同，如合同的参与主体、交换数据的元数据、数据范围、质量要求、使用条款等进行结构化描述，并按照此标准格式发送、接收、处理和使用数据。其中，最关键的使用条款使用“4W2H”模型进行描述，展现数据被什么条件的消费方使用（Who）、数据在哪里被存储和使用（Where）、数据使用时间（When）、数据将被如何处理（Do What）、数据通过什么方式被使用（How to）、数据使用次数（How Many）等。数据连接器负责承载和处理被数字化的数据合约。

综上，在数据空间中实现可控数据交换的关键是，通过技术保障数据交换协议的有效实施，应对数据的非排他性和可复制性，确保数据流通过程中的合约遵从。

4.3　数据空间的基本特征

数据空间是数字世界履行数据交换合约的平台。结合实际的业务场景来看，数据空间为数据主体提供了完成数据交换的数字化作业标准。如图 4-5 所示，数据空间具备三个基本特征，不同的数据空间之间通过数据连接器互联互通，构成数据空间的生态系统。

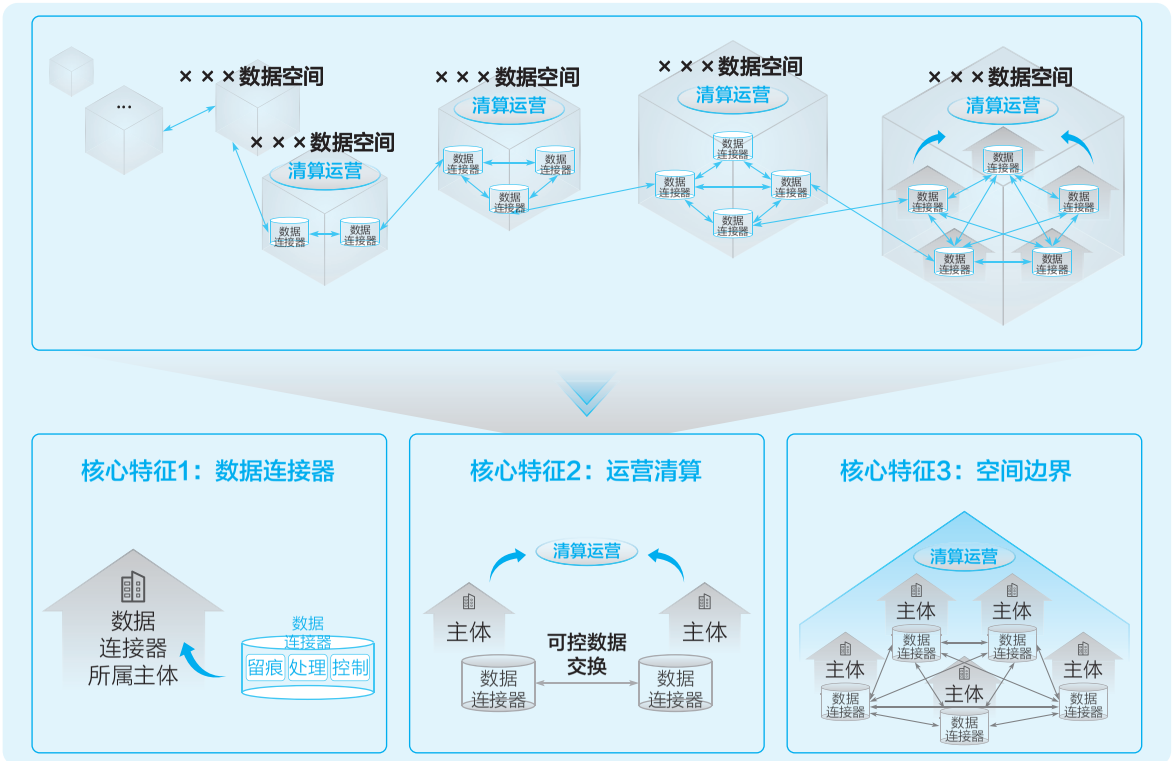


图 4-5　数据空间的核心特征及生态系统

数据空间生态以数据连接器为基础，所有的数据交换通过数据连接器执行，且数据交换一定是可清算的。此外，数据空间是有边界的，该边界本质上是数据流通的管理边界。

4.3.1　参与主体通过数据连接器交换和使用数据

所有的参与主体均通过数据连接器交换和使用数据，这其中包含三种含义。

第一，所有需要交换的数据均通过数据连接器进行交换，数据的传输、集成均根据数据连接器中的合约被执行和管理。

第二，数据连接器为数据供需双方提供加工处理服务，数据按照合约定义的策略在数据连接器中进行加工处理。加工后形成的新数据资源也将按照合约中的策略进行进一步的使用和流转。数据空间为数据供需双方提供了对数据加工产物进行流转和处理的能力。

第三，数据消费方在收到数据后，按照合约中的条款使用数据。数据连接器提供实施和评估数据使用条款的能力，确保各方达成合约要求。

数据空间为数据主体交换数据提供留痕、处理及使用控制能力，这是数据连接器最基本的能力特性，也是数据空间的最关键特点。数据连接器为数据主体在数字环境中提供了数据交换的抓手，融合了不同的数据集成技术，并实施了数据交换的数字化作业标准。

4.3.2　具备运营清算能力

数据空间的运营清算特征是指，数据交换的全过程被记录存证，运营主体在数据空间中以交换参与主体的视角，对所有被交换的数据和使用情况提供可视可管能力，具体包括以下几点。

- 数据空间有唯一的运营主体，负责保障数据空间中的交换主体能够如约进行数据交换并使用数据。运营主体需要对参与主体及所使用的数字资源的可信负责，包括身份管理、认证等。
- 数据空间可以在数据交换过程中定义数据采集的探针。数据交换日志的探针主要围绕数据合约展开，且只能获取数据交换的操作元数据，不能采集具体的数据内容。不同的数据空间运营主体可能定义不同的数据探针标准，在数据空间互联互通的导向下，未来应该会逐渐形成一套标准的数据空间清算探针。
- 数据空间的运营清算可以为数据主体提供运营报告和服务能力，并基于数据连接器提供一套数据交换的运营指标，如有多少数据进入数据空间，交换出去哪些数据，哪些是被高频交换的数据，从其他主体处获取了哪些数据，哪些数据被高频使用等。

数据空间具备对空间内各主体、各类数据的运营清算能力，可以促进高价值、高质量的数据最大限度地被使用，从而促进供需匹配，充分释放数据价值。

4.3.3 以可清算范围界定数据空间边界

除了数据连接器及明确的运营主体，数据空间还需要有清晰的边界概念。数据空间是参与主体依托数据连接器进行数据交换的管理环境，数据空间的边界是通过数据连接器进行连接的数据的流动边界。一方面，数据空间以参与主体的数据连接器为管理节点，参与主体必须通过数据连接器进入数据空间。另一方面，不同的参与主体申请数据连接器并加入某个数据空间后，即建立了与该数据空间内所有其他主体进行数据交换的能力，所有的交换活动也被统一记录和上报。数据空间的边界是个管理的逻辑边界，主要体现在如下两点。

- 适用清算规则的数据主体及数据连接器在数据空间边界内：数据空间以可管理的数据连接器为边界点，形成数据空间的逻辑边界。同一主体可以加入不同的数据空间，并在不同的数据空间中分别申请连接器以交换数据。加入数据空间并遵循数据空间清算规则进行清算的数据连接器，属于同一数据空间的范畴。
- 数据空间独立管理数据交换的控制策略和治理规则：数据空间内不同的运营规则、主权策略等，在一定程度上也代表了数据空间边界。如欧盟 Catena-X 数据空间在欧盟云设施规则的基础上，要求交换的数据要符合汽车工业相关数据标准，而出行数据空间在隐私方面会有更严格的规则。因此，运营规则也可以体现数据空间的差异性。

数据空间需要运营主体在清晰的管理规则下持续地运营，定义并执行数据交换的规则，才能更好地促成更多的主体参与到数据交换生态中。运营方通过运营清算能力实施上层的交换规则及运行数据连接器，以构建明确的数据空间边界。

4.4　数据空间中的数据合约

数据空间可提供数据可控交换的标准数字化能力，在现有的数据集成技术基础上将数据交换合约数字化定义、条款数字化执行，实现供需双方基于数据的价值获取。如何定义并设计数据合约？数据合约的关键要素是什么？本节将从对象数字化的视角描述数据空间的合约构成。数据交换合约可以理解为明确的业务主体（企业 / 组织等）之间，就数据资源的交换以及使用所达成的在技术上具有一定约束力的一系列要约和承诺，其全流程如图 4-6 所示。

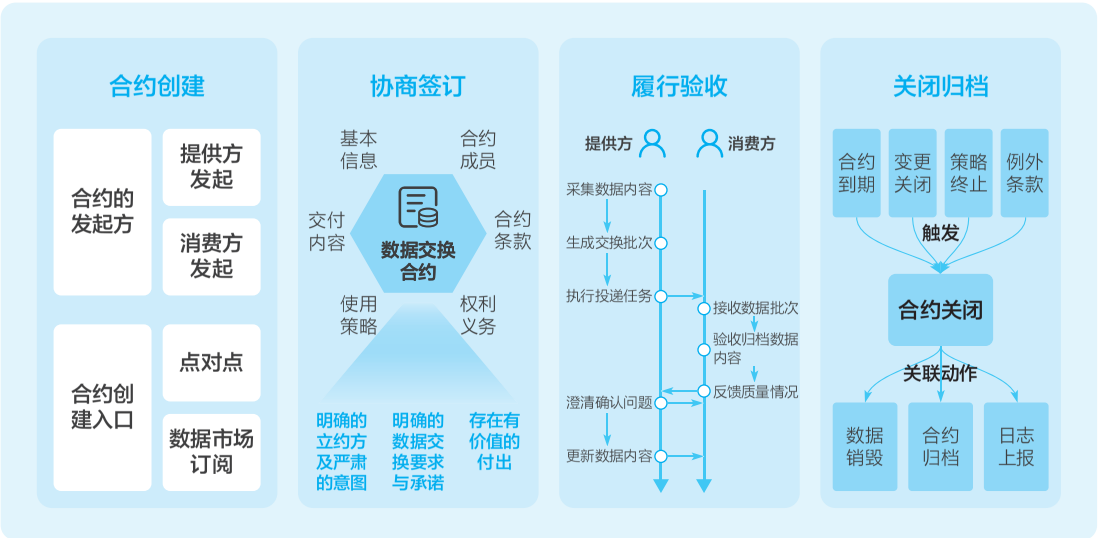


图 4-6　数据交换合约全流程示意图

数据合约具备一系列独特且关键的特征，支撑其在数据交换场景中有效运作。以下从四个维度展开阐述：可在线协商，让参与方便捷地对数据合约内容进行沟通确定；可标识，利于精准识别与管理合约；基于规则进行履行验收，保障合约执行规范有序；覆盖数据全生命周期，从创建到关闭归档全面把控。下面详细介绍这些数据合约的特征。

4.4.1　可在线协商的数据合约

在传统的电子合约中，双方通过在线平台或电子方式进行合同的声明和签署。而数据空间支持多种合约发起方式，并提供灵活的合约创建入口。数据合约的发起主要有如下两种形式。

1. 点对点协商

数据提供方与数据消费方通过线下沟通，对要交换的数据和使用策略达成一致意见后，由数据提供方生成数据合约，并将目标数据资源、数据策略等添加到数据合约中，该场景属于“线下清晰，线上执行”。数据消费方收到合约后确认即可签订。数据即由提供方传输至消费方，快速达成数据交换。如在华为敏感数据跨内部组织交换的场景中，敏感数据的交换规则和使用策略（如对个人属性进行脱敏、使用后销毁等）由公司行业管理部门发布，内部组织间进行数据交换时，通过策略模板预置即可直接发起合约并交换数据。

2. 数据市场订阅协商

数据供需双方通过公开市场达成数据交换。数据提供方将自有资源封装为数据 Offer，并在数据 Offer 中增加默认的数据使用策略。数据消费方基于数据市场上数据商品的描述信息和使用要求，可以选

择接受默认的使用要求或者叠加新的诉求。消费方发起订阅后，数据提供方会收到该订阅，如果需要修正，提供方会将修正后的使用要求再发回消费方。该过程可以反复，在最终达成一致后，生成数据合约。如在华为鲲鹏昇腾生态数据空间中，华为数据提供方发布部件的共性问题解决方案，供特定的数据消费方使用。该数据 Offer 的使用策略中明确了可消费数据的主体范围、数据被消费的方式、时间等。数据消费方可以订阅该数据 Offer，并结合具体的合作项目需求更改数据使用策略，经过华为数据提供方认可后即可达成合约。

4.4.2 可标识的数据合约

与现实的商务合约类似，数据合约除了要具备法律所认可的电子合同必备要素外，每一种构成要素都有数据合约自己独特的要求。数据合约基于明确的立约方，有明确的数据交换要求与承诺，且数据在供需两侧的价值需要可显性化。以下简要描述数据合约中最关键的要素。

1. 数据合约主体

数据合约主体除了具备现实世界主体在法律层面的履约信息外，还需要具备数字身份 ID，而且此数字身份 ID 是经过验证且具备合法效力的。这也要求当数据空间部署在不同的环境和平台时，除了要结合平台或者环境自有的身份认证程序外，数据合约还需要建立数据空间内的唯一识别 ID（可以与所处环境的身份 ID 一致）。

2. 合约标的

通常情况下的合约以有形商品、服务、软件等作为标的，而数据合约中的标的，我们先以“数据商品”这个词来代替。当前业界对数据商品的定义正在逐步发展和统一，此处仅从数据空间可控

数据交换的角度定义，数据商品通常是指对可信来源的数据资源的整合，并以可交付和可履行的形式交付。数据商品包含分类信息、元数据信息、数据质量说明、数据来源说明、数据范围说明、数据依附条件（提供方侧）、数据的期望服务水平等信息。

3. 数据使用策略

数据面向不同的使用场景可释放不同的价值，数据如何被使用的条款（即使用策略）是数据合约最关键的组成部分。数据使用策略详细描述了数据如何交付、传输、存储及消费方接收后以何种方式使用。相较于传统合同的条款，数据合约中的使用策略条款需要约定策略如何被定义、如何在技术或操作层面被执行，通常情况下采用已发布或者具备一定权威性的标准作为实施参考。华为数据空间采用了 W3C 的 ODRL（开放数字权利语言）作为数据使用策略的参考标准。

4.4.3　基于规则进行履行验收的数据合约

数据合约生成后，数据将根据合约规定从数据提供方到数据消费方。合约的履行与合约约定的方式有关，主要有以下几种。

1. 数据的副本复制

数据连接器可以配置不同的级别（参见 6.3 节），数据连接器的级别既代表了数据连接器的数据存储处理能力，也代表了可信等级能力。在数据空间内，数据在不同的连接器之间交换，数据可以作为副本被复制，也可以基于虚拟链接技术在消费方的计算环境内用于计算。是否需要复制副本，取决于交换双方所申请的数据连接器的资质，但最终取决于数据合约中的约定方式。这意味着对于规模较大的数据计算，如果需要使用虚拟链接技术，并在消费方的计算环境中进行计算，对消费方所申请的数据连接器的等级和资源配置

要求较高。如在华为鲲鹏昇腾数据空间内，对于在一定期限内只需读取文件的合作伙伴，不需要复制数据副本。

2. 数据的更新方式

数据有一定的时效性。通常情况下，使用同一套元数据描述的数据是在不断更新的。对应的文件和结构化的数据记录，在数据空间内都通过数据交换单元（参见 5.3 节）进行交换。对于合约中同步更新数据版本的要求，数据连接器提供了主动更新和被动更新两种方式：主动更新由消费方自己触发，避免更新过多非必要的过程版本；被动更新是数据一旦有版本更新就推送到数据消费方。

3. 数据交付验收

根据合约将数据从提供方传递到消费方，需要提供相应的机制确保供需双方已交换的数据是满足合约约定的。在数据连接器内，双方可以通过合约的使用策略约定验收方式。一种是“到货即验收”，即通过“通知”策略，数据首次从提供方传输（或虚拟访问）后，供需双方均会在该合约下收到相应的通知（类似于邮件的回执功能）。另一种是“质量反馈即验收”，即数据首次从提供方传输（或虚拟访问）后，数据消费方需要在合约约定的时限内主动反馈数据验收单（主要是数据质量，包括准确度、完整度等指标），双方以验收消息为验收依据。数据的交付验收，保障了供需双方对数据合约的有效履行，也提供了数据交付进度的数字化管理方式。

数据的交付验收相比于实物资产的交换交易更加复杂，很多时候是在最终的使用过程中，消费方才会得到详细的数据质量分析，数据空间能否提供丰富的质量检查功能也是很重要的一点。数据质量工具可以在数据投递之前帮助提供方进行质量检查和异常数

据过滤。数据消费方接收数据后，对于数据质量的相关疑问，可以通过数据连接器与提供方分析澄清。数据提供方再基于合约的质量改进要求提供数据内容更新服务。

4.4.4　覆盖数据全生命周期的数据合约

相比于传统的商务合同，数据合约的关闭依赖技术环境及策略评估的结果，需要考虑关闭的触发条件及触发之后数据的留存处理，数据合约的生命周期管理包含了对所交换数据的管理（合约中数据的元数据通过数据交换单元呈现，是合约的关键组成部分，通常不删除）。数据合约的生命周期包含生效、提前终止、到期关闭，合约不同生命周期阶段对应的交换数据和基于数据已产生的加工产物如何处理，是合约最终关闭前的重要考虑因素。数据的留存处理方式首先是基于合约中的使用策略决定，包括如下几种。

- 合约到期，数据内容同步删除：合约中约定的策略失效条件达成，合约失效后触发数据删除操作。
- 合约到期，数据内容在度过缓冲期后删除：在合约策略协商中，双方约定合约策略到期后数据留存的缓冲期，度过缓冲期后触发数据删除操作。
- 合约到期，数据内容加密归档后长久保留：在合约策略协商中，双方约定合约到期后数据加密保留。

对于加工产物，因对应的数据来源及加工的形式比较复杂，加工产物的数据主权问题主要通过再协商的方式解决。一种是加工产物策略与原始数据策略一致，加工产物的归属取决于合约约定。另一种是加工产物的数据主权经过供需双方再协商，由数据消费方享有。

数据空间中围绕合约的所有交换日志会在供需双方及运营侧

进行上链检测，数据清算中心保留完备的合约，但不包括数据内容。另外，值得强调的一点是，数据空间的数据合约与基于区块链的数据智能合约有一定的区别和联系。数据智能合约作为一种运行在区块链上的程序，具备自动执行、条件触发、不可篡改等特性，符合数据空间协议的一些基本要求，如数据的安全、可追溯等。但智能合约在链下数据获取、敏感和高价值资产上链、全面的数据互操作性、代码的安全和可靠性方面，与数据空间协议的要求存在差异，无法满足数据空间定义的所有场景要求。

4.5 数据空间中的数据策略

4.5.1 从访问控制到使用控制

数据主权是指，自然人或公司实体拥有对其数据进行排他性自决的能力，更多地体现在数据主体对数据如何被使用的控制权利。在具体实现上，数据主权的很大一部分体现在数据合约中的交换和使用控制策略将如何被定义和执行。从实施使用控制策略的角度看，数据的使用控制是对传统数据访问控制的扩展，可以对数据的使用提供精准保护，如图 4-7 所示。

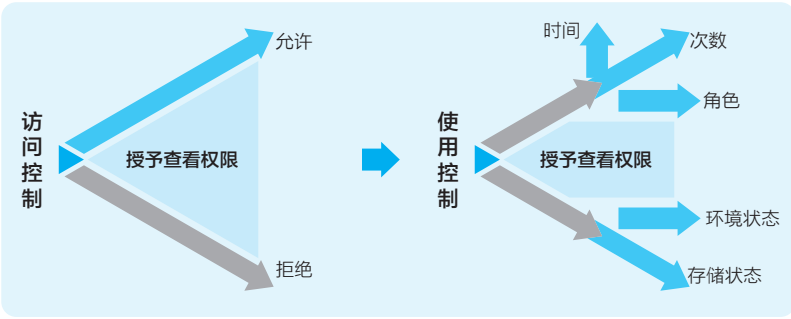


图 4-7　数据权限从访问控制扩展到使用控制

在 IT 系统中，一般通过访问控制实现用户对各类资源的管理。访问控制指的是用户或者程序对于某个资源是否具备访问权限，想要获得权限就需要授权，授权就是授予资源权限的过程。权限控制存在多种访问控制模型，如自主访问控制（DAC）、强制访问控制（MAC）、基于角色的访问控制（RBAC）、基于属性的访问控制（ABAC）等。

数据资源的访问及权利管理理论一直在发展，业界是从数字权限管理（DRM）开始深入研究的，但早期 DRM 的实现受限于专有文件类型，无法支持数据的重用、组合和扩展。在此背景下，数据使用控制的概念在 2004 年被提出并不断发展。数据使用控制是指，在使用数据之前和使用期间，对使用者拥有的权限、应尽的义务和所处访问环境的条件进行评估。

由此可见，数据使用控制是对访问控制的进一步延伸和细化，控制规则在数据被授予访问权限后才会被实施和评估，它强调的是用户或者程序在什么条件下可以对数据进行何种类型的操作。如图 4-7 所示，用户被授予查看权限后，使用控制可以进一步指定查看的条件，如查看的时间、查看的次数、软件版本、数据存储地点等。

4.5.2　数据使用控制策略是跨边界的控制

如果仅限于单一封闭系统内的使用控制，通过现有的 ABAC 等控制模型就可以较好地实现，但在分布式、多实例的环境系统中，现有的控制模型又面临着新的挑战。数据空间需要完整地记录数据被跨环境使用的证据。使用控制不仅指数据所有者在其封闭系统中能够决定数据如何被使用，还强调数据被发布到外界后，所有者仍能决定数据被如何访问和使用。也就是说，使用控制策略需要

具备控制数据流的能力。

如图 4-8 所示，主体 A 期望数据发布到主体 B 之后，数据的使用控制策略依然能被有效执行。

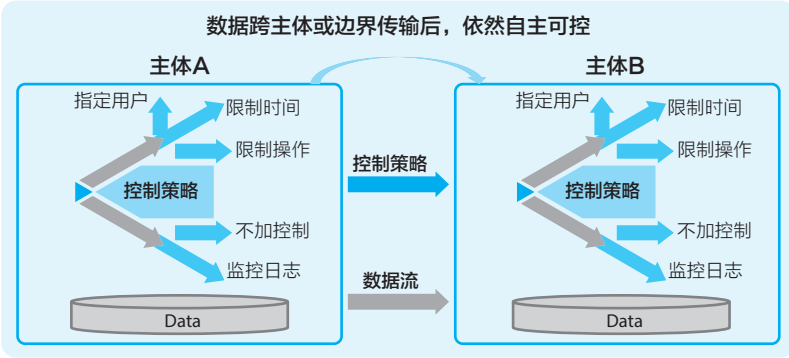


图 4-8　跨主体或边界的数据使用控制策略

使用控制包含了三层含义：

- 数据在传递过程中受使用控制策略的保护（如不被篡改等）。
- 数据在传递后仍然适用使用控制策略。
- 数据使用后不会被扩散，并能按照策略被销毁。

即使主体 A 和主体 B 所处的系统环境不同，该使用控制策略也对数据的流向始终有效。其中的数据使用策略，既包括预防性策略（不满足条件不可使用），也包括监测性策略（如主动记录日志）。

4.5.3　多维可组合的“4W2H”策略描述模型

数据使用控制策略是一种数字权利，对数字权利的描述有多种标准语言可参考。当前各类标准语言均偏向于实现层，如 ODRL 规则，是机器可执行的语言描述，难以支持业务用户的可读和编

辑。在数据合约的定义过程中，业务层面的使用控制条款定义基本上由非结构化文档承载。为实现数据交换相关策略的高效定义和生成，需要在数据连接器内构建一套描述模型，将业务使用条款结构化，同时与 ODRL 建立映射关系，连接业务自然语言与实现语言。

如前所述，数据使用控制策略倾向于对数据流的控制，主要包括数据交付和使用要求，策略要保证数据可被指定主体使用的同时不被转移出去。如图 4-9 所示，“4W2H”策略描述模型为了更好地连接业务语义与策略描述语言，选择从交付和使用数据的关键要素出发。

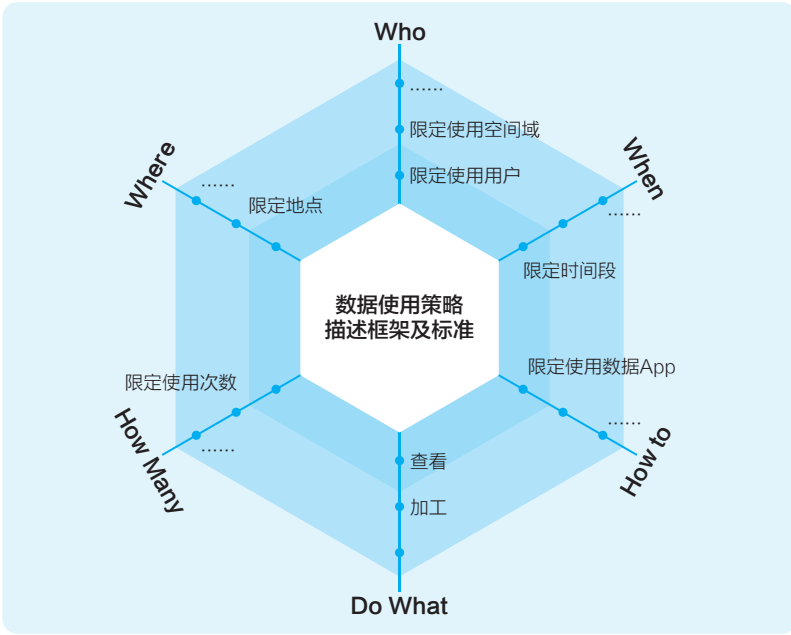


图 4-9　“4W2H”策略描述模型

下面是“4W2H”策略描述模型的介绍。

- Who：主要与参与主体相关，即根据用户、组织或数据连

接器的属性值选择对应的 ID 范围，但不包含数据工具和系统等服务类资源，如限定特定的组织使用。

- Where：主要与数据的存储、传输环境的策略相关，通过对应基础设施资源的认证注册信息进行约定。
- How to：通过什么方式，与数据使用工具 / 系统等相关，如限定只能通过已注册的 BI 工具、AI 大模型等使用数据。
- How Many：操作的次数限制，如用户读取次数、数据 App 的访问次数等。
- When：以时间为基础的控制条件，如指定的时间周期、指定的时间段、限定的有效期等。
- Do What：对数据进行何种操作，包括查看（读取）、加工、下载等。

“4W2H”策略描述模型，以结构化的形式承载了自然语言描述的条款信息。“Do What”中包含的对数据的操作可转化为 ODRL 的操作类策略，其他的可以转化为 ODRL 中的约束类策略，最终组合成数字权利规则。

4.6 小结

数据空间为现实主体提供了数据可控交换流通的新模式，核心主旨是确保数据供需双方关于数据交换的过程及结果能够标准化及规范化落地，充分利用技术手段保障数据主体对数据的持续控制，强化信任机制，促进各方基于数据实现共赢。数据空间以数据合约和数据使用控制策略为主要对象，通过数据连接器承载合约及策略的技术实现。数据空间为数据主体提供了数据流通交换的全生命周期管理能力。这也意味着数据空间需要具备系统性的数据工程能力，设计要涵盖数据的采集、存储、处理、分析和保护等多个方面。

5

第 5 章 CHAPTER

数据空间的总体设计

数据空间涉及数据采集、存储、处理等多个环节，需要系统化的数据工程设计。数据空间以数据可控交换为核心环节，同时数据的接入、消费，以及从接入、交换到消费能力落地需要兼容现存多样化的数字化环境和系统。

本章在 IDS 参考架构模型的基础上，更进一步阐述数据从接入、交换到消费的关键业务设计过程，并明确数据空间的关键能力。

5.1 数据空间的工程能力要求

从工程视角来看，数据交换包括数据接入、数据可控交换、数据消费三个阶段，具体对应四个部分，如图 5-1 所示。

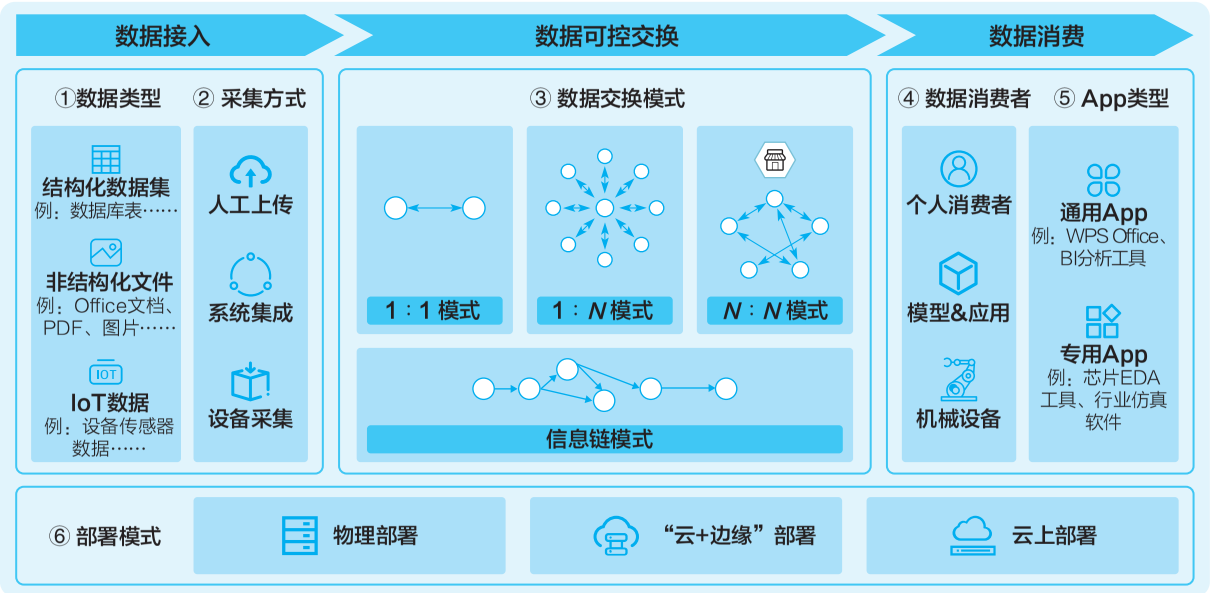


图 5-1　数据交换的工程设计考虑点

- 数据接入：不同类型的数据选择不同的方式接入数据连接器中。
- 数据可控交换：根据数据空间运营及数据交换策略的影响范围选择不同的交换模式，并根据现有的数字化资源选择合适的部署模式。
- 数据消费：数据空间支撑数据价值闭环，提供满足不同类型需求的数据消费能力，包括消费对象和消费 App。
- 数据连接器的部署：数据空间架构兼容现有的数字设施资源，工程层面需要考虑云上云下已有的资源。部署需要提供物理部署、“云 + 边缘”部署及云上部署多种方式。

5.1.1　多模态数据按需灵活接入

在数据的接入和采集环节，数据空间可以利用现有的各类技术，采集多种不同格式以不同方式产生的数据。系统性的数据采集能力需要重点考虑数据的类型和采集方式，确保数据的准确性、完整性和可利用性。

采集数据之前，需要做两件事情。

1. 分析数据类型

结构化数据通常以关系数据库二维表为主，如 CSV 文件、SQL 数据库等，具有固定格式和模式。非结构化数据无固定格式，如文本、图像、音频和视频等，通常需要特定的处理方法来加强安全保护（如脱敏等）。

2. 选取适当的数据采集方式

数据连接器可以提供多种入口：

- 人工录入：适用于数据量不大、易于人工输入的情况，如调查问卷、在线表单。
- 传感器上报：适用于需要实时监控的场景，如环境监测。
- 系统集成：如通过 API 从外部系统或服务获取数据，适用于结构化数据的批量采集。
- 网络爬取：自动从互联网上采集数据，如新闻、社交媒体内容。

在数据的接入和采集环节，数据连接器还可以提供数据清洗和预处理能力，以提高数据质量。这要求为数据连接器设计自动化的数据清洗流程，包括去除重复、纠正错误和标准化数据格式。数据接入后，对于不同格式、不同增量的场景需要选择差异化的数据存储和组织方案，考虑数据采集及存储的可扩展性和灵活性。

5.1.2 满足不同运营要求的 4 种交换模式

数据可控交换模式可在一定程度上代表数据价值释放模式，其作用是为数据提供方在可控范围内提供更多的数据交换方式。不同的数据交换模式，在数据空间中的角色、流程、共享方式、数据开放程度、数据对象的管理方式等方面均存在差异。通过支持不同的模式，数据主体可以将同一份数据以不同的形态投入不同类型的数据空间，在数据“无损复制”下实现数据价值最大化。当前有以下四种模式。

1. 1∶1 模式

1∶1 模式下，供需双方对需要交换的数据内容及范围非常清楚，数据对象通常比较集中和单一。数据交换局限在两点之间，不需要第三方角色干预且不适于向第三方披露（如机密合作等），数

据交换的流程简短。数据空间中只存在两个主体，数据通常由提供方主动推送到消费方。

2. 1 : *N* 模式

1 : *N* 模式适用于由一方主导、多个并列协作方存在竞合状态的业务协作场景，如为了获取多家不同类型渠道的销售数据，但又需要保护渠道之间的独立性，主导方与并列协作方存在同类型数据的交换，并列协作方互不相连、相互隔离。

3. *N* : *N* 模式

N : *N* 模式适用于生态构建场景。在生态链主的运营下，各参与方作为对等角色互相交换数据。该模式下，由于生态成员的互联互通，数据通常是具备行业属性的，且随着连接度的提升，生态所对应的行业数据标准将越来越成熟，有利于提高成员之间数据对接的标准化程度，提升生态数据流通效率，释放行业数据价值。

4. 信息链模式

相较于以上几种模式，数据链模式适用于数据价值链条清晰、高价值数据的“大股东”需要全面跟踪并控制数据加工处理权限流转过程的场景。该模式通常具备相对成熟的数据供应活动，拥有不同数据的数据处理方将上游数据和自有数据处理后再按要求传递到下游。在该模式下，数据空间需要提供基于数据二次转移的数据流监测和控制能力，数据合约需要多方磋商。

5.1.3　人、系统与设备均可按策略消费数据

从消费数据的对象类型看，不同对象在处理过程中对策略和权限的管理方式存在较大的差异。

- 面向用户的权限控制关注用户对数据的直接操作权限，如访问权限、修改权限、消息知会等。
- 面向模型或应用的权限控制通常更为复杂，如同样是数据访问限制，指定模型或应用只能访问特定的数据处理逻辑和过程。
- 面向设备的权限控制需要关注设备如何收集、处理和传输数据，以及如何与上层应用控制策略联动协同。

针对不同类型的数据消费对象在权限控制上的差异，数据空间需要提供不同控制层级和深度的策略下发、实施与评估能力，这对数据空间的策略服务标准化程度要求比较高，且需要适配差异化的对象权限鉴权机制。

如果进一步把所有数据消费系统视为数据应用程序，那面向不同数据类型、处理逻辑会有两大类数据应用程序。

- 第一类是通用数据应用程序，即不包含特定业务逻辑的App，如文档处理类中的办公软件、通用数据分析工具等。
- 第二类是面向特定业务场景、用于处理特定格式或者业务逻辑的专用数据应用程序。

通用数据应用程序的原子功能及接口的标准化程度一般较高，数据空间可以通过与 API 集成并管控 API，实现数据应用与数据流通策略的融合。

在专用数据应用程序中，数据的编辑和处理可能需要遵循特定的业务规则和标准，应用程序需要在遵守这些规则的基础上再叠加控制策略，数据空间可能更容易支撑数据存储控制、数据访问控制等颗粒度比较大的策略。

5.1.4　兼容现有的数字化环境进行部署

如第 3 章所述，数据空间是开放可演进的架构，数据空间的部署和应用是在现有 IT 资源的基础上实现数据高效、可控流通。常见的数据空间部署方式有物理部署、“云 + 边缘”部署、云上部署等。

1. 物理部署

物理部署是将数据空间中的多个数据连接器部署在不同的地理位置或者网络节点上，将数据连接器所依赖的服务器、存储设备等部署在多个数据中心中，并连接已有的 IT 资源。该模式适用于供需双方需要将各自的资源纳入数据空间，但同时又希望能自主管理数据基础设施的场景。在华为与各被投资公司的尽调空间中，双方将自有的数字基础设施资源注册到数据空间内独立管理（如资源的运维、扩容），而与数据空间相关的资源操作记录需要完整地上报到数据空间的运营中心，但数据连接器各自部署在对应的数据中心的机房中。

2.“云 + 边缘”部署

“云 + 边缘”部署方式下，数据空间的运营管理、数据的处理和分析等均在数据空间服务提供方的云上进行，各个主体独立存储交换数据，并可在端侧直接消费数据。端侧需要提供更高安全性、高可靠性的防护能力，且需要保持与云端互联。该模式强调对数据内容的相对独立管理，为参与方提供了更加灵活的选择。

3. 云上部署

云上部署模式下，数据空间完全部署在云上，数据交换的参与方像使用 SaaS 服务一样使用数据连接器服务。交换主体无须关

注数据空间的集成逻辑和资源设施状态。这种模式属于便利模式，当前华为的内部实践均通过该模式部署数据空间。

不同的部署方式对数据空间的开放集成能力要求不同，选择哪种部署方式取决于业务需求、成本效益、性能要求和安全考虑。

5.2 数据空间中的参与角色

数据空间的接入、可控交换、数据消费及部署均存在多种不同的影响因素。在不同的工程实现环节需要有哪些角色参与并执行相应的活动以及不同类型的角色分别做什么，都需要进行明确的定义，从而保证数据空间的各个部分可以高效运转和运行。本节将在 IDS 参与角色的基础上，描述各类不同角色的核心职责。

由于数据空间涉及多种不同类型的参与主体，同时涉及数据接入、数据交换、数据消费及空间部署等多个环节，因此必须明确数据空间的关键角色及其核心职责，保证数据空间持续使用并逐步构建起主体间的高效流通通道。数据空间的四类角色如图 5-2 所示。

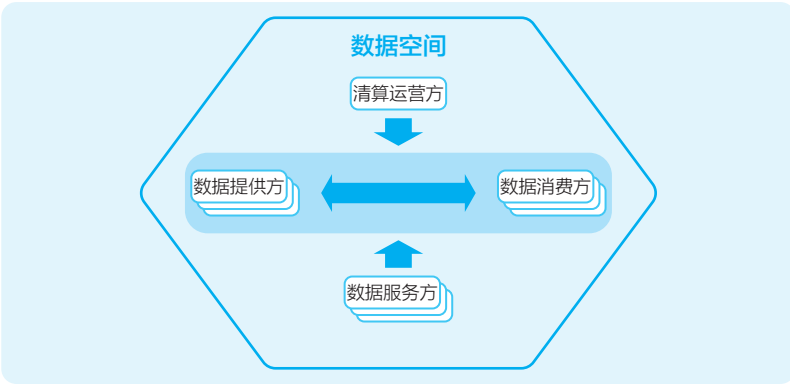


图 5-2　数据空间的四类角色

- 清算运营方：数据空间中必不可少的角色，负责保障数据空间的持续运转。
- 数据提供方：数据交换的核心参与方，负责准备和提供数据资源，是数据价值的获得方之一。
- 数据消费方：数据价值的直接获得者。
- 数据服务方：实现数据空间中数据价值增值的主要主体。

从数据交换的视角来看，数据空间各类角色的关键职责可以从数据交换管理、交换数据描述和数据生命周期管理等方面考虑。

5.2.1　保障数据空间可持续运转的清算运营方

清算运营方是数据空间运行的责任主体，负责数据空间的申请、建设和持续运营，具体包括如下几类职责。

1. 数据空间管理与规则制定

清算运营方负责搭建和管理数据空间环境。首先需要构建数据空间，包括定义和制定各类数据空间服务的准入规则，如面向基础设施的安全服务、加密、可信存储服务等。其次是建立各类参与方的注册及认证规则，确保数据空间的参与用户、组织及其关系是可信的，并对引进的各类可信数据锚点（如权威第三方的组织机构身份校核服务）进行管理。另外，清算运营方还负责定义和管理数据交换与交易的规则，比如发布数据空间的策略标准和提供相应的模板。最后，清算运营方还需基于数据消费生态定义准入规则，促进数据消费的闭环，使数据空间被广泛应用。

2. 监督记录数据交换 / 交易

清算运营方需要监督数据合约的有效履行，虽然数据连接器

可以通过技术服务监督合约履行，但运营层面的监督管理通常也是必要的，尤其是需要向法务或者公共监督机构提供交换 / 交易凭证的场景。监督记录数据交换 / 交易，需要清算运营方定义相关的探针，并采集、整合相应的交换 / 交易日志。

3. 提供清算服务

为了更好地促进数据主体参与数据空间并交换数据，清算运营方需要提供相应的数据清算服务，以便让数据交换主体更加清晰地知道高价值的数据。数据清算是从合约、数据资源、交换 / 交易主体等维度，统计及分析数据交换的总量、频率、使用量等。数据清算的内容和结果也是组织内衡量数据价值的重要参考内容。

5.2.2 供应数据资源的数据提供方

数据提供方的主要职责是准备和提供数据资源，并保障数据质量，以便达成数据交换。下面从数据交换管理、交换数据描述和数据生命周期管理方面分别说明数据提供方的关键职责。

1. 数据交换管理

数据提供方负责定义数据交换的策略，明确合约中的策略要求，确保数据质量满足合约的描述。另外，数据提供方需要支撑数据运营方，向其提供数据交换过程记录、选择或应用数据交换服务，以及提出数据清算服务的需求并使用清算服务。

2. 交换数据描述

数据提供方负责描述和发布数据资源的元数据，具体包括定义或引用数据空间中已明确的数据交换词汇表，也包括引用与交换内容相关的行业数据标准。数据提供方需要将这些元数据发

布到数据市场，或者将相应数据资源的元数据封装到对应的数据产品中。

3. 数据生命周期管理

数据提供方在数据的采集、获取、存储、汇聚、增强、融合计算方面均可以使用数据空间对应的服务。同时，数据提供方还可以对数据进行处理，并通过与数据处理方的协同，将数据加工成数据成品。

数据提供方需要对数据进行全生命周期管理，从数据的定义、创建、连接、存储到销毁。同时，数据提供方还需要往前延伸到各种利益相关方，包括数据 Owner、数据产生者等。数据提供方必须在数据 Owner 或者数据产生者等法律界定的利益相关方的授权下制定数据策略（很多时候数据 Owner 就是数据提供方），确保对数据质量负责。在数据权益清晰的情况下，数据提供方可以在数据共享交换时决定自己的数据如何被使用，从而实现数据价值的最大化。

5.2.3　获取数据价值的数据消费方

数据消费方根据自身的诉求，与数据提供方或数据服务方达成数据合约，合规地获取和使用高价值数据，并依托数据空间的清算能力提供数据交换和使用的过程信息，确保在数据交换过程中可自证清白。下面同样从数据交换管理、交换数据描述和数据生命周期管理方面分别说明数据消费方的关键职责。

1. 数据交换管理

与数据提供方相对，数据消费方负责协商数据合约并执行合

约中的数据使用策略。另外，数据消费方也需要支撑数据运营方，向其提供数据交换过程记录、选择或应用数据交换服务，以及提出数据清算需求，并使用清算服务。

2. 交换数据描述

数据消费方是订阅和使用数据资源元数据的主体，其行为在数据交换中主要表现为查询和获取数据资源的元数据。

3. 数据生命周期管理

相比于数据提供方，数据消费方是数据价值的获得者，通过丰富、汇聚不同数据提供方的数据来达成数据诉求。在数据的存储和连接处理上，数据消费方需要执行数据空间设定的规则。

此外，数据消费方还需要进一步根据合约明确数据用户与数据资源的关系，确保所获取的数据可以满足用户的需求。

5.2.4 为数据增值的数据服务方

数据服务方是负责数据商品流转的关键角色，下面从数据交换管理、交换数据描述和数据生命周期管理方面分别说明其职责。

1. 数据交换管理

数据服务方可以接受数据提供方的委托，按数据合约转发 / 转售数据产品；可以在一定程度上定义数据策略，协助数据提供方定义数据商品，落实数据提供方的质量要求。同时，数据服务方在接受提供方 / 消费方的数据处理委托时，也需要执行供需双方制定的数据处理相关策略。

2. 交换数据描述

数据服务方在接受提供方的委托时，需要查询或获取被处理数据的元数据。按照提供方的合约要求处理完数据后，数据服务方需要描述并发布数据加工产物的元数据，并反馈给数据提供方。

3. 数据生命周期管理

数据服务方主要负责连接和融合处理数据，并在一定程度上支撑数据的丰富和汇聚，以及为数据供需双方描述和供应新数据资源。

综上所述，数据服务方负责基于数据合约为数据提供方或者消费方提供专业数据服务，如数据清洗、脱敏、合规、计算、整合、信息提取等。需要强调的是，数据主权仍归属于数据提供方。

5.3　数据空间中面向交换的数据形态

数据空间为数据流通提供了一种新的方式。相比于业务流的对接集成、跨业务流的数据湖共享，由于数据空间的数据对象随着使用权被传递后转变为要素化、商品化的有价（值）交换，因此需要一套可以描述交换过程的业务对象。

类比电子商务的交换，如工厂未贴标的货物、经过包装的成品、可组合销售的商品、供需达成的订单等，数据在数据空间中被不同主体所用时也是以不同的形态进行流转的。从数据产生到数据交换完成的过程看，数据对象主要有如图 5-3 所示的形态。

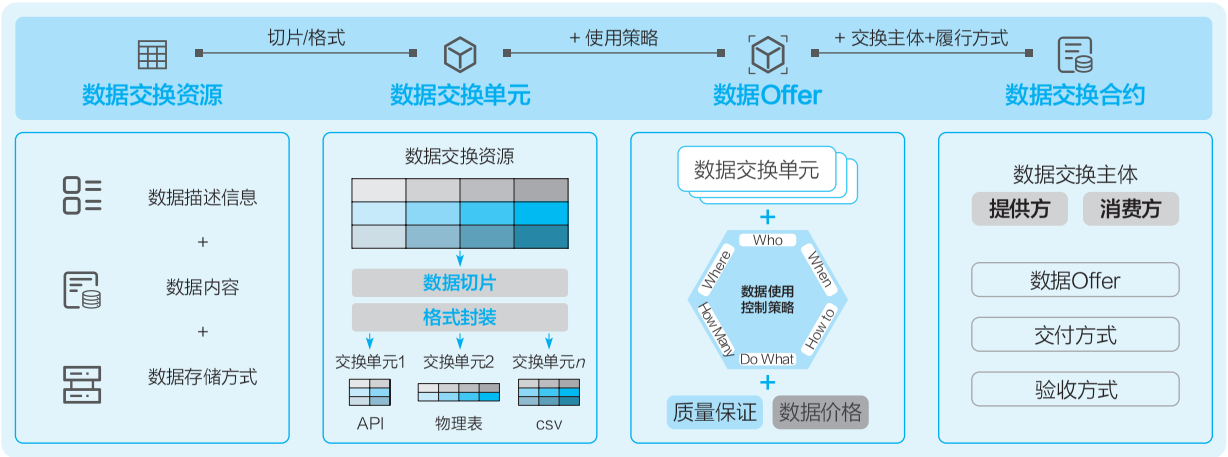


图 5-3　数据空间的数据对象形态

1. 数据交换资源

将组织内的原始数据（记录）转化为符合交换要求的资源，这是数据进入数据空间进行交换的前提条件。“符合交换要求”中最关键的是数据主权的确定，即数据交换资源必须有清晰明确的主权方。

2. 数据交换单元

数据交换资源属于原始的“物料”，还需要进行切片、封装和组合，将其转化为适合交换的单元，以确保传输、验收、使用和清算可以有合适的承载对象。

3. 数据 Offer

数据交换单元相对独立，业务场景所需要的数据通常是多个交换单元的组合。数据 Offer 可以满足场景化的要求，将相关联的多个（也可以是单个）数据交换单元进行组合，并叠加交换或交易相关的策略要求，丰富质量和价格 / 价值方面的描述。

4. 数据交换合约

数据交换需要供需双方就交换的数据、如何交换、如何验收等达成一致，并形成数字化的合约，对应的数据对象就是数据交换合约。

接下来具体介绍各个数据对象。

5.3.1　符合交换要求的资源成为数据交换资源

数据资源是组织或者个人拥有或控制的数据集合，是通过一定的业务活动产生的电子化记录。通常情况下，数据资源按照一定

的规范在组织内部使用。当以价值量化或者要素化的方式进行跨主体交换时，需要在原有规则的基础上增加可交换的处理要求，将数据资源转化为数据交换资源。

数据交换资源是面向未来合约的要求、符合数据连接器的描述规格、需要放入数据空间内并按照数据空间的管理规则进行管理的可交换的数据资源。企业内有众多的数据资源，符合交换要求的资源即可成为可交换资源。数据以物理表、虚拟链接和非结构化文件等形式，从主体原始的业务系统中分离出来，并形成可对外交换的资源。

1. 什么是数据交换资源

数据交换资源包括数据描述信息、数据内容、数据存储格式等。数据描述信息主要以元数据为主，包括数据的定义、数据来源、数据产生的上下文背景、数据所遵循的数据标准和术语等偏向业务元数据方面的信息，也包括数据结构、数据格式、数据表达形式、数据更新频率等偏向技术元数据方面的内容。此外，数据交换资源是带数据内容的资源，有清晰、明确的数据大小、数据量等。

2. 数据交换资源的颗粒度设计

数据交换资源应该避免成为数据资源目录，这两个对象还是有明显的差异的。其中最关键的一点是，数据交换资源代表着对应数据所能提供的信息量的大小，这里更强调对应元数据的范围。有些场景下，元数据本身过于抽象，可以选择叠加数据内容范围，也可以切分数据交换资源。设计数据交换资源的颗粒度，还需要考虑消费市场的诉求。以结构化数据为例，如合同基本信息，从提供方的视角来看，通常包括合同号、合同类型、合同主体、合同要素、

合同文档等。但对于业务模式复杂的场景，不同类型的合同对应的内容差别很大，此时应按照合同类型切分不同的合同基本信息，最后形成不同的数据交换资源。对于非结构化文档，可以将一份报告定义为一个数据交换资源，但是可以有不同的格式。

数据交换资源不同于数据资源目录，数据资源目录是对数据资源的分类，同一个分类下可能有多个不同资源的元数据。数据资源想要成为可交换的资源，需要经过利益相关方的权责确认，即有确定的管理责任主体对其权利负责，避免数据交换后出现负面事件。

5.3.2　数据交换单元是传输、验收、使用和清算的最小单元

数据交换单元是数据交换资源的子集，是传输、验收、使用和清算的最小单元，按使用方式封装成不同的数据格式后提供给消费方。数据资源通过数据切片、格式封装后，形成有明确内容范围的、可重复交换和共享的数据交换单元。无论是数据交换资源还是数据交换单元，在向消费方交付时，都需要对原始数据进行分析、加工，确保达到可交付的状态。

数据交换单元与数据交换资源有明确的血缘关系：数据交换资源是站在数据提供方的视角对所要交换的数据在数据空间内进行初步治理所形成的产物；数据交换单元是满足一定数据标准，从数据自身的特点和独立性出发，对数据交换资源进行切片和封装后形成的相对原子化的数据对象。以非结构化的文件数据交换资源为例，一份具体的咨询报告可以成为数据交换资源，但如果咨询报告本身包含不同方向、不同领域的分析洞察，就可以按照内容分类的角度

将其切割成不同的数据交换单元。对于非结构化的数据，如查询天气信息的 API 包含了不同区县的监测点，可以基于关键属性的可选值将其拆分不同的数据交换单元。

数据交换单元可以结合数据交换资源的元数据描述属性值进行拆分组合，对于有多个描述属性值的情况，虽然可以将数据交换资源拆分成大量的交换单元，但还是需要通过消费侧的引导，逐渐沉淀成可重复的交换单元。

5.3.3 数据 Offer 是由交换单元组成的数据合约标的物

数据 Offer 是由多个数据交换单元按照使用场景、数据策略、数据分类、数据质量要求、数据的价值和价格等多个维度，组合而成的面向数据合约的标的物。封装完成后，数据 Offer 成为具有特定信息的数据或者服务。

相比于数据交换资源治理视角，数据交换单元交互视角的划分，数据 Offer 是基于数据消费方的场景化需求对数据交换单元的整合。数据 Offer 有以下几个特征。

- 满足反复出现的数据需求：数据 Offer 需要有一定的市场需求洞察和设计，且符合市场批量需求。如果无此重复需求，数据交换单元可以直接用于交换，以简化交换路径。
- 有明确的消费群体：数据 Offer 的消费群体要相对清晰，在设计和上架时，可以明确具体的消费群体。
- 相对清晰的使用场景：这意味着数据 Offer 的业务价值可以相对明显。
- 对数据交换单元的进一步组合：数据 Offer 可以是不同数据交换资源的数据交换单元的组合。比如，在华为云面向司

法协查的场景中，华为云作为客户委托的数据平台方，需要向司法机关提供指定的数据，也需要向客户告知具体反馈了哪些数据（符合司法协查规定可告知的场景下）。在该特定场景下，需要提供多个数据交换资源的组合。

除此之外，数据 Offer 还需要进一步考虑市场目录的设计、定价和如何满足监管要求，但这几方面的内容还在持续地探索中。

5.3.4　供需双方共同达成的数据交换合约

如前所述，数据合约是明确的业务主体间就交换数据以及使用数据所达成的一系列要约和承诺。数据合约是主体间合同的数字化承载对象，数据空间可采用技术化手段确保合约中的条款被执行。

数据交换合约是站在数字化技术的视角，为履行数据提供方的主权要求而定义的数据对象。但数据交换合约与现实世界的商业合同也具有绑定和映射关系，数据空间需要体现这层映射关系。数据交换合约与线下合约是一一对应的，线下合约是正式协议（书面的、对齐的、签署的），供需双方除了需要对数据主权方面进行说明外，还需要明确范围、职责及角色，以及对交换数据质量的期望。

此外，数据交换合约也有几点基本要求。第一，数据交换合约是可寻址的，有唯一的名称和版本号。第二，所有合约都可以通过词汇表和行业数据标准进行自我描述，且无歧义。第三，数据交换合约的条款未必可以穷尽技术方面的标准，但必须是所有相关方可以预见的，可以明确条款中所描述的效果。第四，线下数据合同是可计算的，如果无法与线上数据交换合约建立匹配的映射关系，

则需要在数据交换合约中有明确的体现（即有些策略可能通过治理方式来保障，而非技术，但也是策略的一部分）。

数据空间中数据形态的定义和转化非常重要，各种数据对象及其转化关系，本质上是数据交换对象和过程的数字化，是数据交换业务标准化的关键成果。

5.4 数据空间中的交换业务流

明确了数据空间的参与角色和关键对象后，本节将在 IDS 参考架构的基础上，进一步细化数据空间各参与角色的典型用户旅程和关键业务流。业务流是一组端到端的活动集合，能够为利益相关者创造一个有价值的结果，这些活动在能力建模时用业务流阶段来描述。如前文所述，数据空间包含四类参与角色，下文将围绕四类角色的用户旅程设计其对应的业务流。

供需双方的业务流是数据交换的主业务流，一切业务设计都围绕这两个角色对应的业务流展开。数据提供方的业务流重点在于供应数据并获取价值，消费方的业务流重点在于获取并使用数据，如图 5-4 所示。

清算运营方是数据空间得以开展和持续运行的核心角色，是数据空间的使能方，负责建立合规、互信机制，驱动数据空间及其生态有效运行。数据服务方是数据空间的有机组成部分，负责在数据空间中提供数据专业服务，通过接受供需双方的委托增加数据价值。二者的业务流如图 5-5 所示。

下面将分别介绍上述业务流。

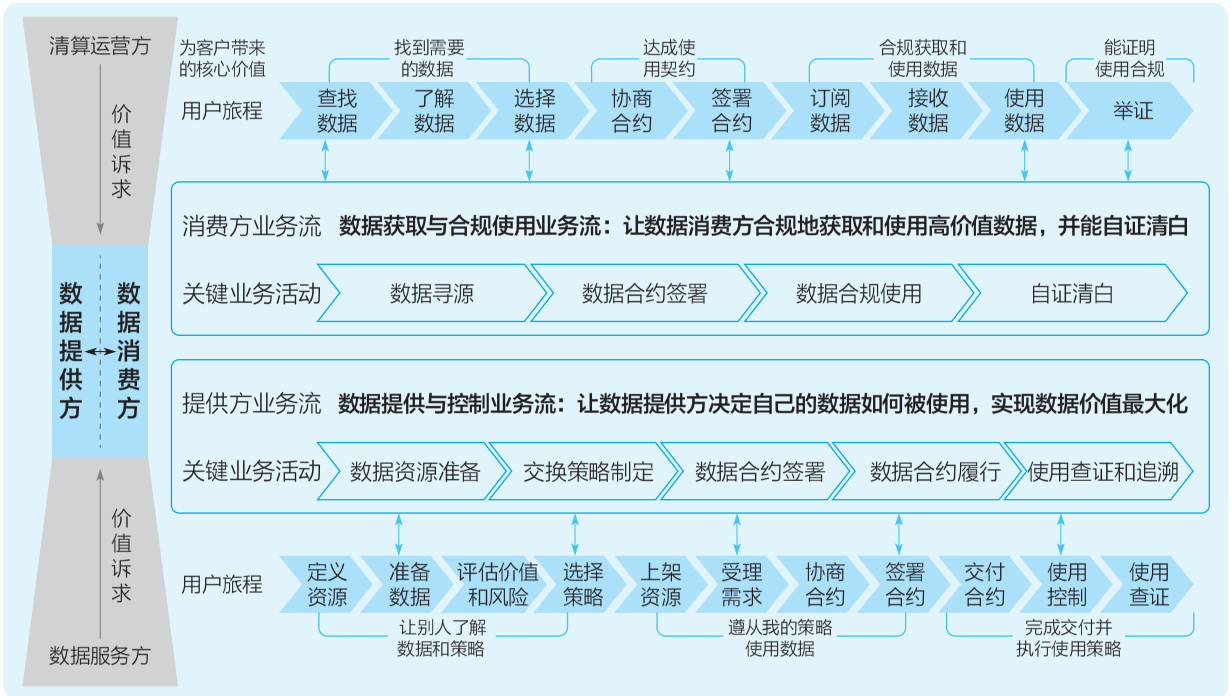


图 5-4　数据提供方和数据消费方的业务流

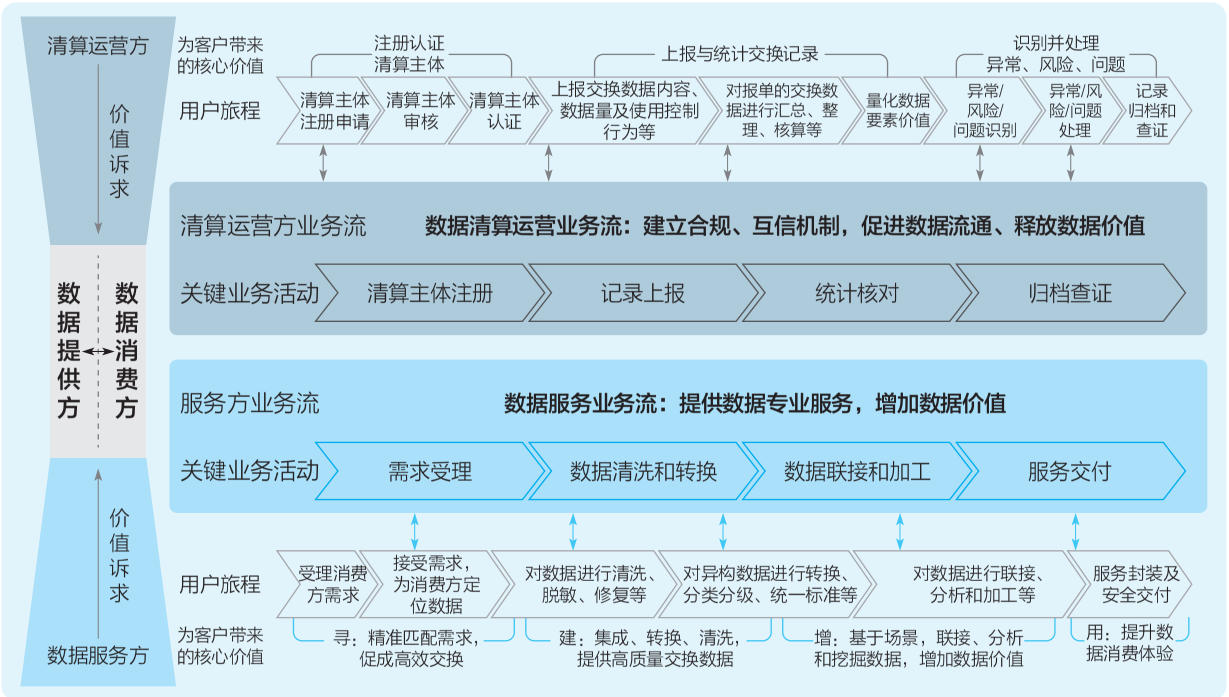


图 5-5　数据服务方和清算运营方的业务流

5.4.1　提供与控制数据的业务流

数据提供方业务流是帮助数据提供方在数据空间内提供数据后获得最大化数据价值的活动集合。数据提供方用户旅程的关键点是让数据消费方可查询并理解数据，在消费方订阅数据之时可以按照其策略签订合约，然后根据已签订的合约交付数据，并能在达成数据交换后获取相关利益。数据提供方通过数据资源准备、数据交换策略制定、数据合约签署、数据合约履行、数据使用查证和追溯等关键业务活动实现该业务流。关键业务活动的具体过程如下。

1. 数据资源准备

数据提供方首先将业务数字化过程中产生的各类数据和分析结果，通过数据治理方法清晰定义数据资源的元数据、内容范围，然后按照数据本身的价值潜力和所包含的信息量大小将其转化为数据资源，并在获得内部数据利益相关方的同意和授权后发布。

2. 数据交换策略制定

数据成为资源后，数据提供方对其进行质量评估、风险评估、价值评估，然后切分成数据交换单元，并从数据策略定义库中选择合适的策略附加到相应的数据交换单元中。策略明确后，数据提供方可将潜在的、可重复的、信息关联度高的数据交换单元组合成数据商品，并发布到市场。

3. 数据合约签署

数据商品发布后，提供方主动发起或者消费方发起数据订阅请求，均可生成合约初稿，双方就数据交付形式、验收形式、使用控制策略等进行磋商，达成一致后签署数据合约。此业务活动能为数据供需双方就数据交换形成有效力的协议，以便于进入合约履行环节。

4. 数据合约履行

根据合约约定的数据交付条件，数据提供方发起数据投递或者数据消费方发起数据获取请求，提供方传输数据商品所涵盖的数据或提供远程访问链接，从而履行数据合约。一旦数据消费方完成数据接收确认，意味着数据提供方已完成合约履行。

5. 数据使用查证和追溯

数据使用查证为数据提供方提供了数据流向及数据使用记录的详细过程，详细程度取决于合约约定的情况（如合约约定提供方可获知消费方何时开始使用数据，最后一次使用数据是什么时候），提供方可根据该策略查阅消费方的使用记录。

5.4.2 获取和使用数据的业务流

与数据提供方对等，站在数据消费方的视角帮助数据消费方在数据空间内获取和使用数据的业务流就是数据消费方业务流。拥有数据市场或者数据空间的数据提供方开放了相应的数据商品，即意味着数据消费方可进入该业务流。数据消费方业务流的目的是让消费方合规地获取和使用高价值数据，并在使用数据之后能够自证清白，确保按约使用数据。数据消费方通过数据寻源、数据合约签署、数据合规使用、自证清白 4 个关键业务活动实现该业务流。

1. 数据寻源

数据消费方通过数据市场或者点对点的方式查询数据商品，并基于查找结果确认数据内容、范围和质量要求。符合需求的数据商品将被选择并加入预选数据清单中。数据寻源主要通过已发布的数据目录或者数据市场来完成。如在点对点模式中，数据消费方与数据提供方已达成共识，数据提供方将自有的特定目录开放给数据

消费方，数据消费方在指定目录下获取所需的数据交换单元。数据消费方也可以按照数据市场目录、数据 Offer 的名称及标签等信息进行模糊搜索，获取相应的数据商品。

2. 数据合约签署

数据消费方选择完数据 Offer 后，如果需要对数据的内容、范围、策略进行调整，则发起数据合约磋商，请求数据提供方进行调整，待双方达成一致意见后，再签署数据合约。在点对点模式下，通常跳过线上的磋商环节，合约生成即确认。在公开市场订阅模式中，因供需双方还需要撮合，所以数据合约的反复磋商比较常见。

3. 数据合规使用

数据消费方验收收到的数据后，按照合约中的策略使用数据。这其中包括几个步骤：根据合约要求配置用户权限；对收到的数据进行归档；如果对收到的数据进行加工，不仅需要对加工产物进一步获取相应的主权，还需要发起协商，完成加工产物的主权转移。

4. 自证清白

数据交换的全过程通过区块链技术存证登记，数据消费方可以通过本地日志或者清算中心报告，提供客观的使用凭证，以满足合规和监管方面的要求。

5.4.3　处理与加工数据的业务流

数据服务方在数据空间中承接数据提供方或消费方的处理请求时，会经历一系列标准化的流程，以提供准确可靠的数据服务，增加数据价值。

1. 需求受理及数据获取

数据服务方会与数据提供方或消费方明确数据处理需求，如

数据的类型、格式、预期用途、处理的复杂度等，并在此之上形成数据处理合约，明确数据处理的范围、时间、费用、保密等法律条款。根据数据处理合约，数据服务方会从数据提供方处接收原始数据，并对数据进行初步验证。

2. 数据清洗与转换

数据预处理包括对数据进行清洗，如去除重复记录、纠正错误和异常值、填补缺失值等；还可包括数据标准化，即将不同来源的数据转换成统一的格式。数据转换包括数据映射和格式转换等。如果将数据字段映射到统一的模型或架构中，需要对数据进行必要的转换，如日期格式转换、数值单位转换等。

3. 数据联接加工

数据联接加工包括将不同颗粒度的数据整合，如将合同与订单整合成一份数据；也包括对数据进行模型计算，如数据聚合、分类、排序等；还可进一步包括统计分析（如运用统计学方法对数据进行分析，提取有用信息）和数据挖掘（使用数据挖掘技术发现数据中的模式和关联）。数据联接加工不限于技术形式，数据空间支持的一切数据处理服务均可以被数据服务方调用。

4. 服务交付

数据服务方将加工好的数据以数据商品的形式交付给数据提供方或者消费方。数据服务交付需要保障整个数据处理过程中每一步的输出都符合规定标准。在此过程中，数据服务方可以通过收集数据消费方的反馈，不断优化数据处理结果，但也需要提供必要的服务支持和咨询。

5.4.4 清算与运营的业务流

站在清算运营方的视角，数据清算运营业务的业务流是建立

合规、互信机制，促进数据流通、释放数据价值。数据交换的清算运营是一个复杂的过程，涉及多个环节，用于确保数据交换的合规性、安全性和互信性，包括清算主体注册、数据交换记录的采集与上报、数据交换及履行的统计核对、交换结果归档查证。

1. 清算主体注册

清算运营方对参与数据交换的各方主体（包括用户、企业和其他组织）进行资质审核。审核通过后，主体在清算平台进行注册，获取参与数据交换的资质和数据连接器。注册通过后，由清算运营方发放相应的资质信息，如数字证书、认证等级证书等，确保数据交换主体的可信。

2. 数据交换记录的采集与上报

在数据交换过程中，清算运营方通过数据采集探针实时追踪和记录数据的流向与状态，并建立日志管理系统，用于记录数据传输、使用的详细信息（如时间、参与主体、数据交换单元等信息）。

3. 数据交换及履行的统计核对

该环节包括对交换的数据进行数据量的核对，确保数据的完整性和准确性；也包括对数据交换的频率、规模、类型等进行统计分析，为数据供需双方提供数据价值量化的基础。在交换和使用数据的过程中，该环节还需要具备异常检测能力，及时发现异常的数据交换和使用行为，并中止潜在风险。

4. 交换结果归档查证

清算运营方在法律法规的基本要求下，叠加数据供需双方的合约要求，提供数据交换的归档和历史查证能力。归档是指对数据合约的

生成和履行过程进行归档，包括交换过程的所有状态信息。建立查证机制是为了支持参与主体对交换结果进行验证和查证。同时，清算运营方还需保留详细的审计跟踪记录，以便在必要时进行审计和调查。

5.5 数据空间的核心能力组件

从数据空间的能力要求、参与角色、数据形态、角色业务流可知，数据空间以各方达成的规则和数据合约为准绳，为数据提供方和数据消费方建立可信机制，确保数据能按照合约策略被使用，数据交换和使用的全过程可被证明及追溯，从而促进数据流通。经过 20 多次的场景验证和版本演进，华为构建了数据空间“1+3+2”的能力框架，如图 5-6 所示。

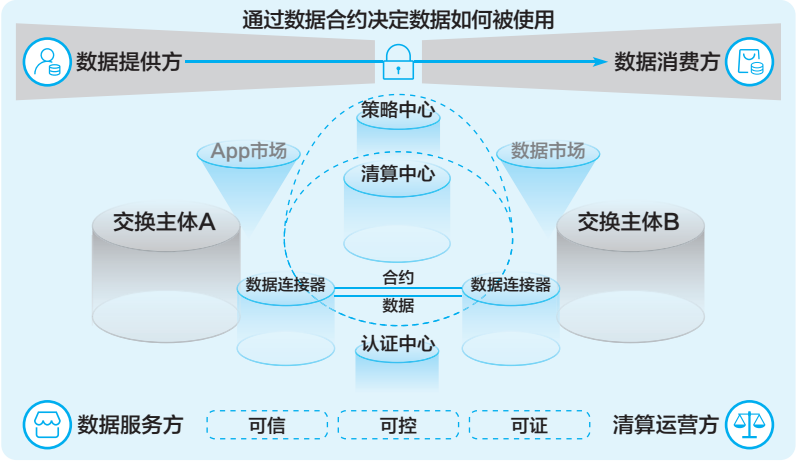


图 5-6　华为数据空间的能力框架

华为数据空间的能力框架是“1+3+2”的组合，具体如下。

- 1 个数据连接器：数据连接器是数据空间最关键的组件，负责交换数据及执行合约。

- 3 个交换履行保障中心：数据空间围绕可控交换需要构建可信、可控、可证的能力。认证中心保障数据空间的参与方及数字资源是可信的。策略中心负责统一定义并下发“数据主权”的具体说明、达成效果、实施策略的评估要求。清算中心确保数据交换的全过程可追溯，供需双方的数据交换有第三方凭据。
- 2 个市场：数据市场是数据 Offer 的集散地，能够促进供需匹配。App 市场是实现数据增值的工具和应用库，通过 App 市场可以实现数据价值的闭环。

数据空间中的数据市场和 App 市场是开放的，且目前已经有相对成熟的技术实现。以下将重点介绍“1+3”组件的能力。

5.5.1　交换数据及执行合约的连接器能力

数据连接器允许参与者将使用策略附加到数据空间的数据中，执行使用策略并无缝跟踪数据来源。数据连接器组件的能力设计有以下几个特点。

- 数据连接器采用分布式架构，能够兼容多种协议；支持以开放节点进行数据交换；可以在本地或云环境中运行，不需要一个中心实例来存储数据。
- 数据连接器具备安全隔离能力，使用应用容器等管理技术提供安全隔离环境。应用系统通过标准 API 来存储、访问或处理数据，并通过数据连接器交换其他参与者的数据。
- 数据连接器能够控制数据使用，确保数据随策略流通，以及数据消费者按策略使用数据。
- 数据连接器实现了数据交换的业务过程与数据集成传输的解耦。在数据交换的业务过程中，数据目录制定及资源发布、数据合约磋商、数据传输等关键步骤包含一套标准化

的业务动作，而数据连接器负责执行这些动作。数据集成传输是一项相对成熟的技术，目前也有现成的成熟协议或者技术组件。数据连接器对数据集成传输技术是开放兼容的，比如各类结构化数据库、物联网传感器的消息流等。

5.5.2 实现组织、个人及组件认证的可信能力

认证中心负责对组织、个人及技术组件进行可信认证，实现参与方身份与平台环境的可信。通过从身份、环境、应用、App 等各个层级考虑可信设计，构建一个增强信任的数据交换环境，可以满足不同业务场景下的差异需求。这是数据空间交换数据的信任基础，具体能力要求如下。

- 身份认证：在数据交换或交易中，首先需要确认参与方的身份。这包括个人用户、企业实体以及自动化的软件代理。身份认证可以通过多种方式实现，例如使用数字证书、生物识别技术或基于区块链的数字身份验证，也可以引用第三方可信数据源，从而确保身份的真实性和合法性。
- 环境安全：数字环境必须是可信的，包括软件和硬件，目的是保障数据交换操作和数据传输过程的安全，如采用高安全的加密方法等。
- 应用层级的可信设计：在应用层面，需要考虑应用程序和服务的可信，包括应用程序的设计过程、代码等是安全的；还需要进行完备的恶意软件或者漏洞检测，确保服务的运行环境是已隔离和安全的。
- 技术组件的认证：技术组件（如服务器、数据库和网络设备）需要有安全性和可靠性设计，并须进行定期的安全审计和安全测试。

除了上述围绕数据交换的可信关键点，数字基础设施的安全也是可信的关键，数据空间需要有一套明确的可信安全标准服务接口要求。

5.5.3　决定数据如何被使用的可控能力

通过数据使用控制策略，可以在数据主权可控的基础上确保数据消费方合规使用数据。策略中心负责实现数据使用控制策略的模型化、标准化，并通过技术驱动保障数据交换条款和条件的有效实施。数据空间的可控能力是相比于其他数据平台、网络空间等最为关键的特征，具体能力要求如下。

- 具备数据使用控制策略全生命周期的管理能力：涉及数据使用控制策略的定义、发布、管理、实施、评估等标准化对接服务。
- 分布式的可控能力：不仅需要在数据连接器端支持决策，还需要支持通过认证的第三方服务决策，并在达成控制目标的基础上，兼容不同的决策实现标准。
- 服务化和模板化的决策标准：数据使用控制策略通常涉及数据从交换合约生成、数据投递、交付到验收和最终回收的全生命周期，对应的策略会影响到数据交换的不同业务阶段。策略中心需要提供场景化、服务化的能力，提升用户的使用体验，确保策略的定义和选择能够明确表达合同条款。更进一步地，合同条款可以通过语义转化模型自动匹配相应的模板并生成策略实例。

5.5.4　全链路追溯清算的可证能力

数据交换的可证能力是指数据交换和使用的全流程操作日志上链，清算中心通过探针在数据交换全链路中采集日志，构建去中

心化的上链存证能力，并为不同参与方提供查证、清算服务。具体能力要求包括如下几点。

- 以合约为中心的过程记录：数据交换的过程记录以合约生效为起点，以合约关闭为终点，可证是合约全过程的可证。在合约条件范围内的所有操作过程都要进行完整的记录，并形成以合约为中心的完备日志。
- 基于合约的事中响应：数据交换合约中存在一些无法通过技术来实施的策略，这时就需要通过治理及运营规则来保障合约实施。无论是因为技术，还是因为运营方式，在合约生效期内出现关于数据的争议时，清算中心都应该提供交换 / 交易事件协商的服务。
- 以数据价值量化为导向的清算能力：清算中心根据合约履行的全过程记录，为数据交换主体提供高频数据、高频交易对象等多维度的统计分析结果，并从数据价值量化的角度提供对应的清算能力。

5.6 小结

华为数据空间在 IDS 参考架构模型的基础上，从数据接入、交换、数据使用等方面，利用现有的数字化系统丰富了部署视角，细化了能力关键点。数据空间的能力建设需要从不同的交换模式、关键的业务角色、核心业务对象的流转过程等角度来考虑。从数据供需双方的业务流视角来看，重点是保证数据交换合约的可靠定义和如约履行，其中的关键组件为数据连接器。在数据连接器执行数据合约之前，首先需要保证主体可信；在执行数据合约时则需要保证数据使用可控、数据交换全过程可证。上述内容只是对数据连接器、可信、可证和可控能力的简要说明，后续将论述各核心能力组件的具体设计要点。

6

第 6 章 CHAPTER

数据空间中的数据连接器

在讨论数据空间的基本特征时，我们强调了参与主体需要通过数据连接器来交换数据。作为数据空间的核心组件之一，数据连接器扮演着将现有系统及其数据接入数据空间生态系统的桥梁角色。它也是一种软件架构，旨在通过安全的架构设计来确保数据主权在受信任方之间的数据交换过程中得到有效保护。数据连接器可以部署在本地服务器或云服务器上，以实现数据空间的灵活部署。

华为数据空间中的数据连接器主要应用于以下三种场景：数据留痕、数据加工处理，以及数据使用控制。一个开放、可扩展且兼容的数据连接器对于数据空间适应不断变化的技术环境和业务需求至关重要。本章将介绍华为在数据连接器的架构设计、功能特性和应用场景方面的探索与实践。

6.1 数据连接器的关键能力特性

在数据连接器的应用实践中，确保数据可信交换顺利进行至关重要。为此，数据连接器必须具备一系列关键特性，包括但不限于数据资源管理、数据 Offer 管理、数据合约管理、数据传输以及数据消费等，如图 6-1 所示。在华为的实践中，这些特性已经得到了详尽的设计和实施，本节将对这些特性进行详细的说明。

数据连接器的特性基本包含了数据交换的完整流程，下面将按照业务流的顺序对各业务环节及功能特性做详细说明。

6.1.1 数据资源接入

在企业环境中，数据的存储方式多种多样：结构化数据可能被保存在各业务系统的数据库内，或是集中存储于数据湖中；而非结构化数据则更加多元，它们可能被存储于企业文档管理平台、对象存储服务（OBS）、各类数据库或本地。至于数据的提供方式，则更是五花八门：有的通过数据 API 提供，有的通过数据库直接连接，有的借助 FTP 传输，有的依托企业文档服务，有的通过电子流系统实现无缝对接，还有的通过电子邮件或本地传输等方式提供。

华为的数据连接器针对上述种种情形进行了全面的适配，特别是在系统对接方面，已经建立了标准化的接口和模板。这使得与企业电子流系统的对接工作能够更加迅速、顺畅地进行。华为 EDS 支持的数据资源接入方式如图 6-2 所示。

如图 6-2 所示，数据连接器主要支持结构化与非结构化两大类数据资源的接入，在功能实现过程中，针对两类数据资源接入的处理方式具体如下。

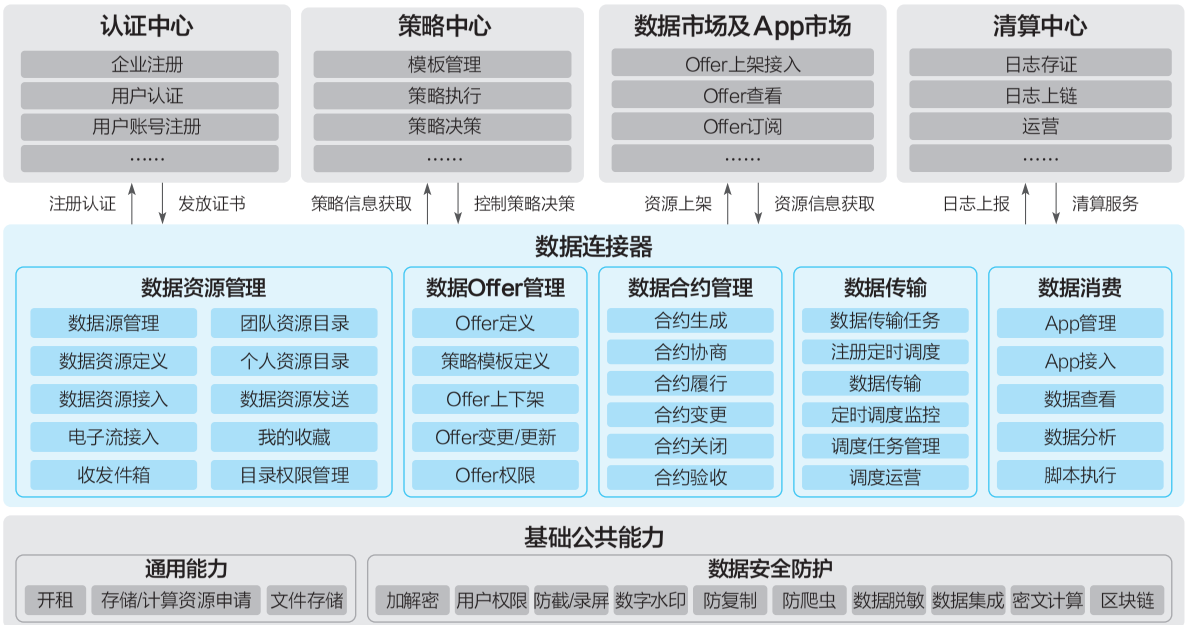


图 6-1　数据空间中的数据连接器

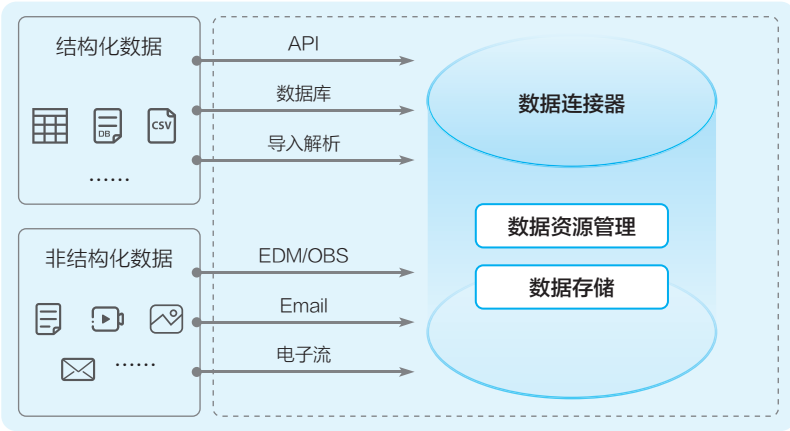


图 6-2　华为 EDS 支持的数据资源接入方式

1. 结构化数据接入

（1）API 接入

API 接入时，鉴权和数据切片是两个重要的考虑因素。鉴权用于确保只有合法用户才能访问 API，企业中 API 的鉴权方式有多种，如 AKSK、JWT 等。在设计 API 接入时，要考虑如何去适配 API 的鉴权，同时在接入时通过对 API 数据的参数和输出内容进行定制实现数据切片。

（2）数据库接入

数据库接入时，需要考虑接入数据库类型的可扩展性。不同类型数据库的驱动程序不一样，其 SQL 语句在语法上也可能存在差异，因此在接入过程中，一般需要先对 SQL 语句的生成进行抽象，并提炼出接口，再使用这些抽象接口。当增加新类型的数据库时，只需要为新数据库适配接口即可。另外，数据库接入需要考虑

海量数据的接入，一般使用流式处理数据记录，以防止内存溢出（Out Of Memory，OOM）。从安全性来说，数据库接入尤其需要考虑 SQL 注入的风险。由于数据表、数据字段以及查询过滤条件均由外部输入，这些数据会参与 SQL 语句的生成，因此在接入时必须要对输入的内容进行过滤处理，防止 SQL 注入。

2. 非结构化数据接入

（1）EDM/OBS 接入

EDM/OBS 接入时，需要解决大文件上传的问题。当网络不稳定或受到网络防火墙的限制时，用户上传或下载大文件可能会失败，这种问题一般可使用文件分片传输方案解决，即当某个文件片段传输失败的时候，对该片段进行自动重试。另外，EDM 接入也需要考虑传输的效率。在某些情况下（例如逻辑多租连接器），数据来源与传输目的地都是同一个 EDM，此时可以使用 EDM 自身的复制机制完成传输。对于非同源的 EDM，仍使用传统的“源端下载 + 目标端上传”的方式。

（2）Email 接入

Email 接入一般用于数据投递。在该场景下，Email 可视为一种外部数据源，由于涉及数据投递，因此需要对 Email 的目标邮箱地址进行严格管控。华为 EDS 将 Email 视为一种外部数据源进行管理，Email 数据源的创建需要进行审批，审批通过后，传输合约中才会显示并允许选择该 Email 数据源。另外，Email 接入也需要考虑多个数据文件为单一附件的需求，某些场景还会要求对附件进行加密。

（3）电子流接入

电子流接入就是公司内部的业务电子流系统与数据连接器的

对接。在企业内部，每个电子流的接入场景可能都不一样，有些场景将数据接入即可，有些场景在数据接入后还需要按照业务要求生成数据目录并将文件放入其中，还有些场景需要对文件的访问权限进行控制。这就需要数据连接器提供标准的对接接口，包括资源、合约、策略、权限等；同时，根据对接的场景形成标准模板并提供给电子流系统，这样才能快速地完成电子流对接。目前，华为内部已对接了超过 10 个电子流系统，并逐步收敛形成对接模板，供各类电子流系统选择。

6.1.2 数据资源管理

数据接入后，数据连接器就可以将接入的数据创建为数据资源或数据交换单元，并通过精心设计的数据目录对二者进行高效管理。更进一步地，数据提供方可以对数据资源施加使用控制策略，将其转化为策略驱动的数据 Offer。这些数据 Offer 随后被发布到数据市场中，供数据消费方订阅，从而实现数据的价值流动。数据资源的管理方式如图 6-3 所示。

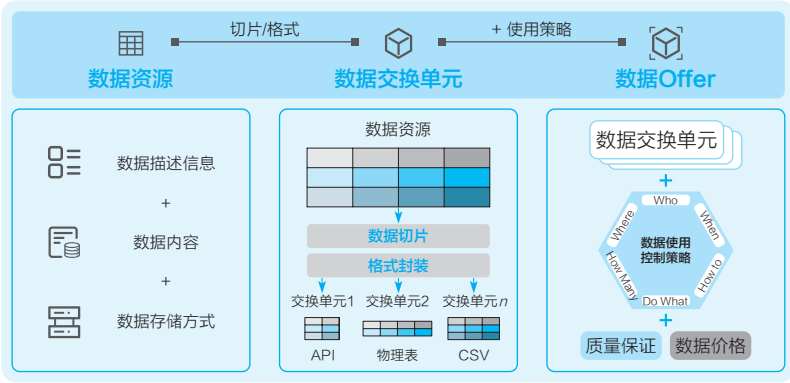


图 6-3　数据资源的管理方式

1. 数据资源

数据通过物理表、虚拟链接和非结构化文件等形式，从主体原始的业务系统中被分离出来，形成可对外交换的数据资源。数据资源通常包含数据描述信息（主要以元数据为主）、数据的技术存在形式（如数据存储的格式、数据的大小、数据的读写方式等）、数据内容、清晰的利益相关方等关键信息。通常情况下，数据提供方可根据数据本身信息量的大小设计和划分数据资源。

2. 数据交换单元

数据交换单元是数据资源的子集，是传输、验收、清算的最小单元。数据提供方通过数据切片、格式封装，能将数据资源转换为内容范围明确、可重复交换和共享的数据交换单元。在将数据资源和数据交换单元交付给数据消费方时，数据提供方需要对原始数据进行分析、加工，使其达到可交付的状态。

3. 数据 Offer

数据 Offer 是多个数据交换单元的组合，即将不同的数据交换单元按照使用场景、数据策略、数据分类、数据质量要求、数据价值等多个维度，组合成面向数据合约的标的物，通常是具有特定信息的数据或者服务。数据 Offer 可上架到数据市场，供多个数据消费方搜索和订阅。数据 Offer 的样例如图 6-4 所示。

数据 Offer 由数据提供方定义，主要内容包括：

- 可见范围：所有数据连接器可见、指定数据连接器可见等。
- 订阅类型：受控、公开。
- 数据清单：数据 Offer 中的数据明细。
- 控制策略：数据 Offer 中预设的数据交换策略和数据使用策略。



图 6-4　数据 Offer 样例

数据提供方完成数据 Offer 的定义后，可以将其上架到对应的数据市场，供数据消费方订阅。对于不再需要的数据 Offer，数据提供方可以进行下架处理。

数据 Offer 上架到数据市场后，元数据会在数据市场上进行展示，供数据消费者快速理解该数据 Offer，从而匹配需求，并按需进行订阅。数据 Offer 在数据市场中的展现形式如图 6-5 所示。



图 6-5　数据 Offer 在数据市场中的展现形式

各类数据接入数据连接器之后，需要进行规范治理，其中关键的一项能力就是资源的编目管理。根据业务，数据空间默认有 3 个一层目录，包括本地接入、外部接收、加工生成，用户可以根据实际情况在一层目录下建立自己的文件夹，对数据进行管理。用户在收件箱中收到文件时，也可以通过归档功能将数据资源归档到目录中。根据实际的业务场景，用户可以将数据资源目录分为个人目录（我的数据）和团队目录（团队数据），个人目录强调对个人数据的管理，团队目录强调多人协同及权限管理。数据资源目录如图 6-6 所示。



图 6-6　数据资源目录

数据资源也可以直接发布到数据连接器的联邦目录中，数据消费方通过数据连接器对外提供的自描述信息中的联邦目录地址进行数据的搜索和订阅。随着数据空间技术的不断发展，数据连接器在技术上经历了从集中式到去中心化的变化，数据市场也呈现出多样化的形态，包括集中式数据市场和联邦式数据目录。集中式数据市场以其统一的管理优势，为用户提供了便捷的数据访问和交换体验；而联邦式数据目录则以其分布式特性，保证了数据的多样性和访问的灵活性。

6.1.3 数据合约管理

数据合约是数据连接器之间进行数据交换的基础，所有的数据交换必须基于合约上的条款和策略进行。数据合约生成后，数据提供方可以基于该合约持续向数据消费方发送数据。这些数据受统一的合约策略控制，策略到期或者提供方提前终止后，消费方对应的合约数据也会被销毁。

1. 数据合约生成方式

- 点对点发送：提供方明确知道消费方是谁、要发送哪些数据、控制策略是什么，并主动创建合约，审核通过后即可生成合约。合约生成后，消费方可直接接收和使用数据。
- 数据市场订阅：提供方把数据和使用控制策略一起封装成数据 Offer 并上架到数据市场。不同消费方使用数据连接器在数据市场上根据需要进行订阅，并和提供方协商生成合约。

2. 数据合约的组成

- 基本信息：提供方数据连接器和消费方数据连接器，以及合约名称等。数据合约基本信息样例如图 6-7 所示。



图 6-7　数据合约的基本信息

- 合约协同成员：分为个人和团队两种。个人合约只能由创建合约的人自己使用，且数据内容只能是创建人的数据。团队合约，即团队成员都可以用这个合约去发送数据，数据内容只能是对应团队的数据。
- 控制策略：包括数据交换策略和数据使用策略，用于明确消费方的使用用户、使用时间、使用次数以及合约的终止策略等。合约中的控制策略样例如图 6-8 所示。

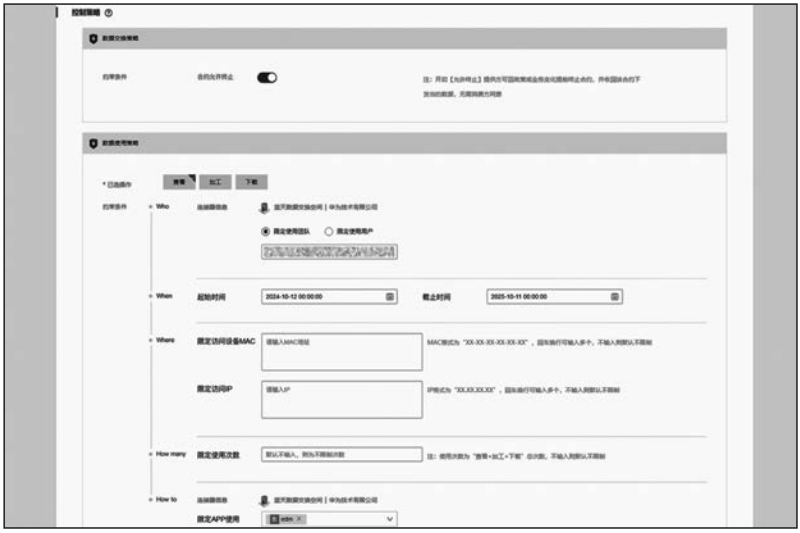


图 6-8　合约中的控制策略

- 数据内容：合约生成后，提供方可以通过批次的形式不断新增数据内容。如果合约中有终止策略，数据提供方也可以对单个批次的数据进行终止撤回。
- 数据批次：合约中多次发送数据会生成数据发送批次，数据消费方根据批次号可快速找到发送的数据并接收。

3. 数据合约策略模板

创建合约时，系统会根据策略模板加载配置好的策略内容，华为 EDS 支持数据连接器用户根据需要定制自己的策略模板。配置策略模板时，用户可以自定义设置控制策略有哪些、是否必填、是否有默认值等，从而满足更多个性化的控制需求。策略模板的样例如图 6-9 所示。

6.1.4 数据传输和传输协议

数据连接器最核心的能力之一就是数据传输。数据接入数据连接器、用户本地上传、通过电子流采集数据、数据连接器之间交换数据、数据连接器与 App 之间交换数据等场景，都需要用到数据传输能力，而且存在场景复杂、调用点多的特性。

1. 数据传输的使用场景

- 数据上传：用户通过数据连接器提供的 UI 界面，从本地上传数据资源。
- 数据采集：提供方数据连接器从提供方的后端系统中采集数据。
- 数据交换：数据通过提供方数据连接器被推送到消费方数据连接器，或者消费方数据连接器从提供方数据连接器中拉取数据。

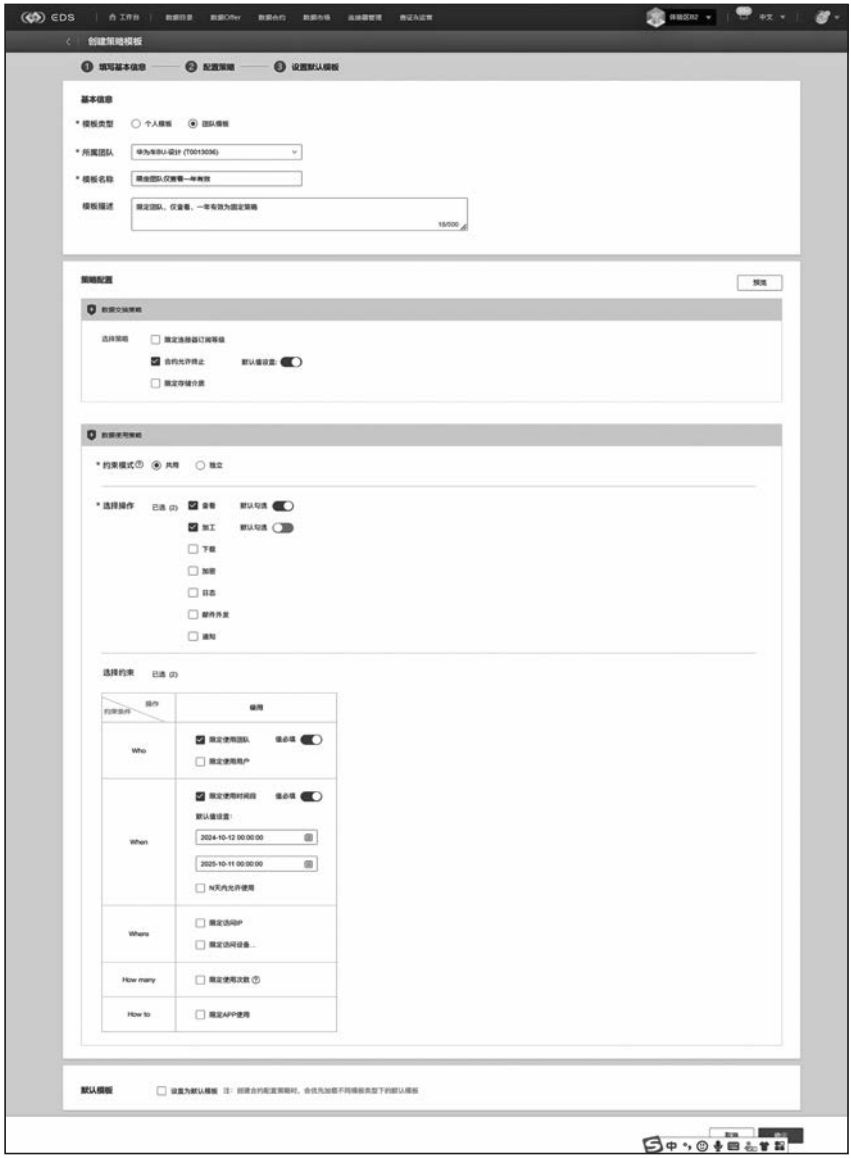

EDS
创建策略模板
1 填写基本信息
2 配置策略
3 设置默认模板
基本信息
* 模板类型
个人模板
团队模板
* 所属团队
* 模板名称
模板描述
策略配置
预览
选择策略
* 约束模式
共用
独立
* 选择操作
已选 (2)
查看
默认勾选
加工
默认勾选
下载
加密
日志
邮件外发
通知
选择约束
已选 (2)
操作
约束条件
值域
Who
限定使用团队
值必填
限定使用用户
When
限定使用时间段
值必填
默认值设置：
2024-10-12 00:00:00
2025-10-11 00:00:00
N天内允许使用
Where
限定访问IP
限定访问设备...
How many
限定使用次数
How to
限定APP使用
默认模板
设置为默认模板

图 6-9　策略模板样例

- 数据使用：将数据从数据连接器传输到数据 App，进行预览、查看、加工等操作。
- 数据下载：用户通过数据连接器提供的 UI 界面，将数据下载到本地。

2. 数据传输的切片处理

在数据集成过程中，对数据进行切片可以提高数据处理的效率和灵活性，优点如下。

- 增加传输速度：对数据和文件进行切片，可以增强文件传输的并发性能，加快数据的传输速度。
- 提升数据的安全性：对结构化数据进行切片，可以实现只传输必要的数据和字段，非必要的数据内容通过切片进行隐藏。
- 支持断点续传：对非结构化数据进行切片，能准确找到当前传输的进度，从而实现断点续传。
- 提高系统资源利用率：数据切片允许更精细地分配资源，从而提高资源利用率和系统性能。
- 容错性增强：在数据传输过程中，如果某个切片发生错误，只需重新传输该切片，而不是整个数据集，这有助于提高传输的可靠性。
- 增加数据传输的灵活性和用户体验：用户可以自定义数据切片的数量和切片的方式，根据数据量选择不同的切片方式，从而提升数据传输的灵活性和用户体验。

EDS 中的分片处理如图 6-10 所示：用户可以选择过滤字段，并在字段上增加条件；也可以选择分片方式，选择后系统会自动对数据进行分片，并按设置的并发线程进行数据传输。



图 6-10　分片处理样例

数据传输的任务管理如图 6-11 所示：用户需要对数据传输的过程进行监控，查看任务的运行情况、系统的资源利用情况，以保障数据传输的效率和稳定性。同时，数据传输任务管理还可以对传输失败的数据进行日志分析以及重新发送。

任务详情

序号	中文字段名	英文字段名	字段值
1	调度作业ID	uid	bcfcbe30-71f2-4a0d-be7e-7eea716b968e
2	请求ID	requestId	79994328-18a3-4a05-885f-081ef639ccdf
3	调度作业类型	scheJobType	
4	优先级	priority	8
5	定时器任务ID	timerJobId	abf94bff-a5db-46aa-8ff2-d57ecc541119
6	作业对象类型	jobObjectType	ARTIFACT
7	调度作业对象ID	jobObjectId	d861895d-ab83-4ca6-a705-063b7cf64444
8	数据源ID	dataSourceId	2c978233-88dc-469d-81af-2992c2ea1b93
9	合约ID	agreementId	
10	数据批次ID	dataBatchId	

图 6-11　任务管理详情

以下是对数据传输关键特性的说明。

- 实时监控：对数据传输过程进行持续监控，以实时了解数据流动的状态和进度。
- 任务运行状态：检查每个传输任务的运行情况，确保它们按照既定的时间表和性能指标进行。
- 资源监控：监控系统资源的利用情况，包括 CPU、内存、存储空间和网络带宽，以确保传输过程不会因资源瓶颈而受到影响。
- 日志记录：详细记录数据传输日志，包括成功和失败的传输事件，为后续分析提供数据支持。
- 失败分析：对传输失败的数据进行深入的日志分析，以确定失败的原因，如网络问题、系统错误或数据问题。
- 重试机制：为传输失败的数据设置自动重发机制，减少人工干预，提高传输的可靠性。
- 告警通知：建立告警系统，在检测到异常或失败的传输时，立即通知运维团队，以便快速响应和处理。
- 性能优化：根据监控和日志分析的结果，对数据传输进行性能优化，包括调整传输参数、优化网络配置或升级系统资源。

3. 数据集成线程池

数据集成服务负责 EDS 中的数据采集与交换。为提升数据操作的性能，数据集成服务会在数据上传、采集、交换等多个环节中使用线程池，在不颠覆原有线程池使用方式的基础之上，从降低线程池参数修改的成本以及多维度监控这两个方面降低故障发生的概率，并提升性能。数据集成线程池如图 6-12 所示。

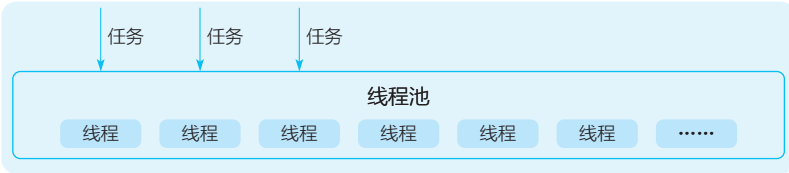


图 6-12　数据集成线程池

通过数据集成线程池，可以做到对任务和线程的状态监控、异常告警以及动态调整，包括如下几个主要能力。

- 监控健康状态：用户可以在线实时了解线程池的实时状态，比如线程池活跃度、任务的执行频率、耗时、异常、线程池内部统计信息等。
- 线程池异常报警：线程池负载关注的核心问题是基于当前线程池参数分配的资源够不够。在运行过程中，如果线程池出现了异常或者队列中等待任务的数量过多便进行报警（支持定制阈值）。
- 动态调整线程池：发生故障的时候可以快速调整，从而缩短故障恢复时间。用户可以在线通过线程池的名字找到指定的线程池，然后对其参数进行修改，保存后会实时生效。目前支持的动态参数包括核心数、最大值、队列长度等，还支持调整是否开启报警、队列等待任务告警阈值。

4. 数据传输的安全

在数据传输过程中，有些场景需要对数据进行处理，以确保数据的安全。

- 在上传、采集数据时，对数据进行安检，防止数据中携带不允许交换的机密信息。
- 在上传、采集、交换数据时，按照业务需求对数据进行脱敏。

- 在传输或存储数据时，对数据进行加密，在使用时对数据进行解密。

5. 数据连接器的传输协议

数据连接器的传输协议是数据传输的标准，目前 EDS 有一套自身的标准，同时也正在做和欧洲 DSP 互通的适配，DSP（Data Space Protocol）是用于数据传输的关键技术，提供了一种标准化的方式来定义如何在数据空间内共享和使用数据，其中包含目录、合约、传输协议。

6. 数据连接器协议适配器服务设计

如图 6-13 所示，数据连接器作为数据交换的主要执行模块，具备对各类协议的适配能力，还支持与其他协议的转换与适配。

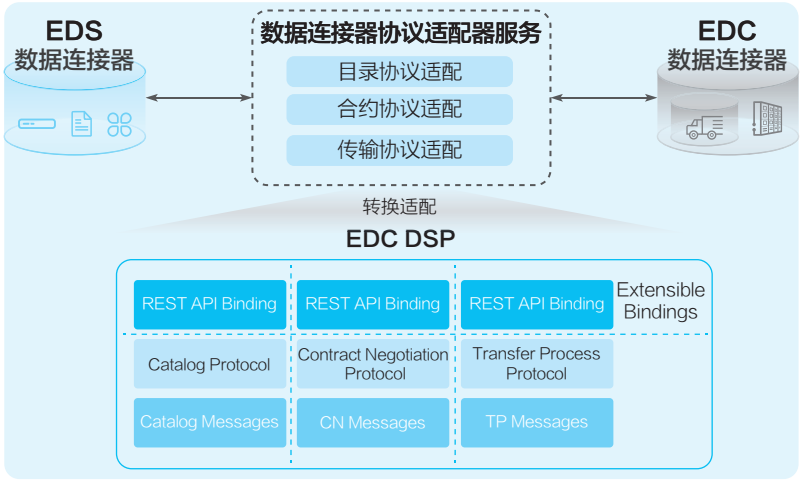


图 6-13　数据连接器协议适配器服务设计

华为 EDS 数据连接器协议适配器服务支持以下主要特性：

- 可作为独立的微服务集群化部署。

- 支持逻辑多租（多连接器适配）。
- 支持与提供方与消费方的双向适配。
- 实现了与欧洲 EDC 数据连接器的互通。
- 未来可按需开发新的协议适配器，支持与其他协议的数据连接器互通。

6.1.5　数据消费和 App 管理

为了实现数据的可控消费，消费方需要通过受控的 App 来访问和使用数据。在数据连接器中构建一套针对 App 接入的认证机制与控制技术，可以确保 App 的可信和可控，以及数据能被受控地访问和使用。图 6-14 展示了四类数据消费 App 接入数据连接器的方式。

下面将针对四类 App 接入数据连接器的方式做详细说明。

1. 原生 App 接入

对于很多新的数据消费场景，如碳数据计算与分析，因为没有传统存量应用的迁移包袱，所以可以基于数据空间开发对应的内置 App，通过注册后安装到数据连接器中即可使用。原生 App 可以很好地契合数据空间的架构，符合数据空间安全、可信、可控等各个方面的技术要求，不需要额外进行改造适配。

2. 工具组件类 App 接入

对于工具组件类 App，如文档处理、算法模型、代码编辑工具等，这些成熟的 App 直接对数据进行加工处理，没有自带的数据存储功能。数据空间支持这类 App 通过注册后部署在数据连接器中的 App 容器内。用户可通过数据连接器启动该类 App，并受控访问数据连接器中的数据。工具组件类 App 仅需进行少量的适配改造即可满足接入要求。

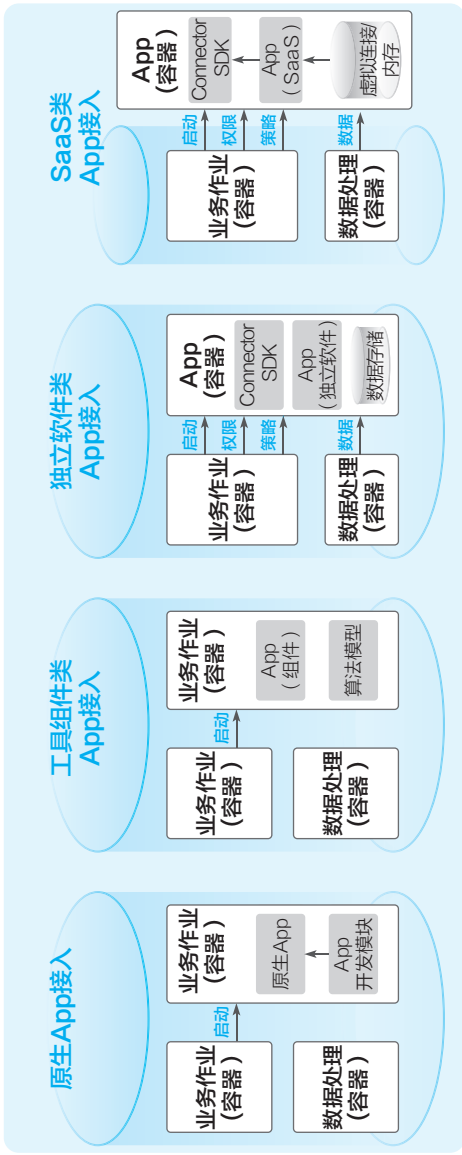


图 6-14　四类数据消费 App 接入数据连接器的方式

3. 独立软件类 App 接入

很多有存量的业务 App 是相对独立的软件，通常自带数据存储能力。用户可以通过在数据连接器中安装 SDK 的方式将 App 接入数据连接器，并通过 SDK 实现独立软件的启用、策略融合、数据存储管控。经过改造，独立软件类 App 可以接入数据空间，同时实现数据的接入、交换和可控消费。

4. SaaS 类 App 接入

对于上云的 SaaS 服务，如大型行业分析平台，这些 App 依赖云平台的基础设施，控制难度比较大，适配改造成本也相对较高，可以考虑通过虚拟连接、安全内存等方式实现数据的可控访问和消费。

针对各类 App 的接入，数据连接器提供了通用的接口标准。App 接口的对接支持两种模式：一种是 App 按照标准 SDK 进行对接；另一种是在数据连接器内注册并部署 App，并与数据连接器通过标准接口进行通信。图 6-15 展示了 App 对接的标准 SDK。

目前华为对外发布了 12 个标准接口，已覆盖账号授权、业务生成、策略对接、数据处理等关键业务活动，支持各类 App 的调用和接入。在制定接口标准的过程中，确保不同 App 间的兼容性和互操作性至关重要。华为确立了通用的 API 规范，并提供了详尽的文档说明，包括接口功能、请求方法、参数以及数据格式。通过引入版本控制和统一的认证授权机制，进一步增强了接口的可用性和安全性。同时，华为提供了 SDK 以简化对接开发工作量，所有接口标准旨在实现数据连接器与不同 App 间的无缝集成和互操作性。

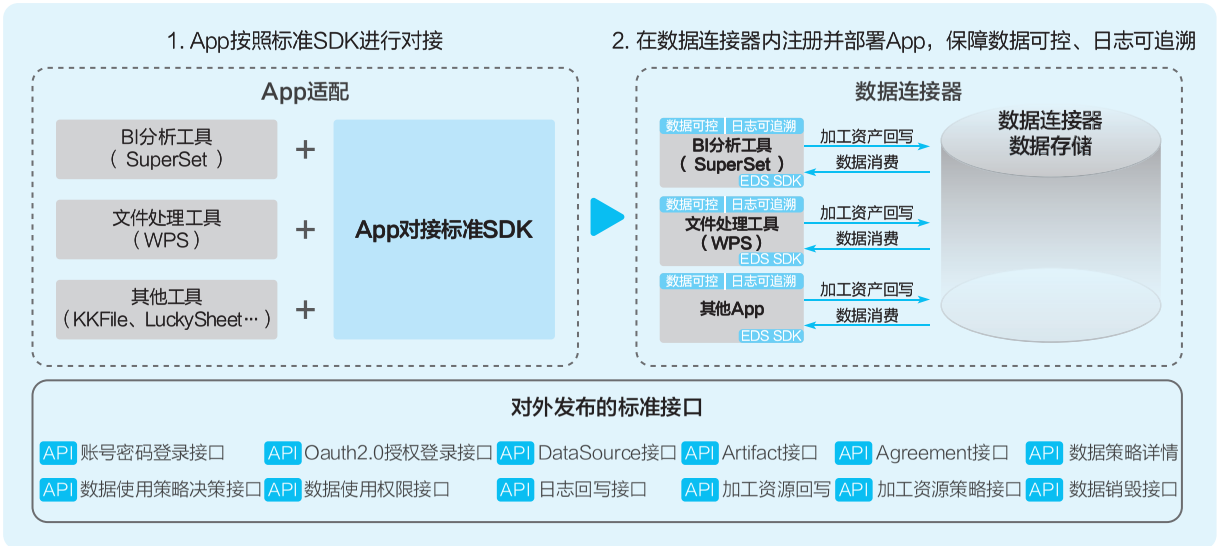


图 6-15　App 对接的标准 SDK

在 EDS 数据连接器中使用 App 需要先经过严格的注册管理。获得许可证书之后，App 才能接入数据连接器供消费方使用。下面将针对 App 的注册与消费过程进行详细说明。

- 第一步：注册 App。在数据连接器的应用管理模块可以进行 App 注册，注册信息主要包括：App 关联的数据类型，比如是文件还是结构化数据；App 的使用场景，比如查看场景、加工场景、下载场景等；App 的部署信息，注册时会要求填写注册 App 的部署信息，以便在一体机场景下对 App 的数据方案做硬件级的策略防护。
- 第二步：在数据使用策略中限定可以使用数据的 App，如图 6-16 所示。

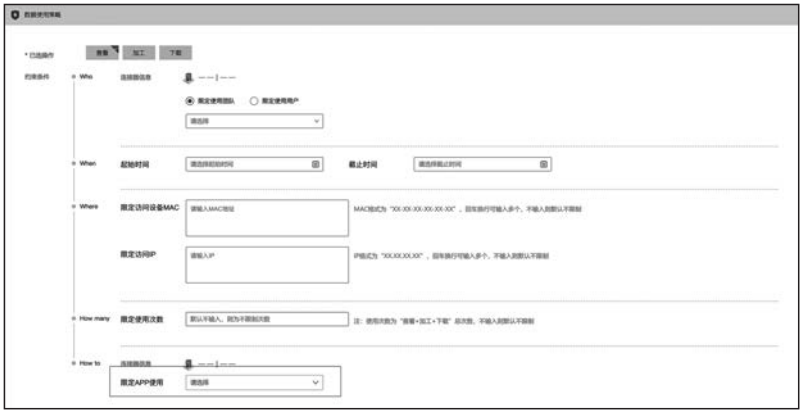


图 6-16　在数据使用策略中限定 App 使用

- 第三步：消费方在消费数据时选择对应的 App 进行数据消费。

6.1.6　数据连接器的配置与监控运营

为了更好地服务业务对数据连接器的管理需求，数据连接器

提供面向管理方的配置与监控运营能力，主要分为配置、监控与运营两部分。

1. 数据连接器配置

在使用过程中，用户要求对业务功能进行配置，以快速适配不同场景需求。EDS 数据连接器实现了业务功能项的配置化，比如可以在配置中心配置数据的接入方式、接入类型、操作主体、交换方式、合约审批、水印等。数据连接器功能配置样例如图 6-17 所示。



图 6-17　数据连接器功能配置样例

根据业务应用的案例场景进行分析，华为 EDS 从谁来建、谁来管、交换流程、资源管理、交换合约、使用策略几个维度将有相似业务的场景总结为场景模板，用户在创建时可以选择场景模板对数据连接器进行创建。数据连接器创建后，这些场景模板中的配置会默认填写在数据连接器的配置模块中，用户可以根据实际业务需求对数据连接器的配置进行调整。

2. 数据连接器监控与运营

为了对数据连接器进行有效的运营管理，我们可以从数据、合约和用户三个维度进行细致的监控与记录，同时对所有相关行为进行详尽的授权日志记录，以确保数据交换的透明度和可追溯性。

（1）数据

从数据维度来看，数据连接器需要记录所有交换数据的详情，包括发送的数据、接收的数据以及经过加工处理的数据。这些记录有助于我们了解数据流动的具体情况，评估数据交换的效率，并在必要时进行问题追踪。同时，还需要记录与数据连接器发生数据交换的其他数据连接器的信息，这有助于我们构建一个完整的数据交换网络视图，从而更好地管理和优化数据流。

（2）合约

数据交换合约的管理涉及对发出的合约和收到的合约的统计分析，这有助于我们评估合约的签署情况。此外，还需要记录与各数据连接器签订的具体合约内容以及执行的数据投递次数等信息，这些信息对于分析合约的履行情况至关重要，同时也为合约的调整和优化提供了依据。

（3）用户

用户管理涉及对用户操作数据的记录，包括操作的时间、地点以及具体的行为。通过对用户行为的分析，我们可以识别出异常的数据操作行为，并及时对这些异常行为进行审视和处理。这种细致的用户行为记录，为用户行为模式分析提供了数据支持，有助于优化用户体验和提高数据交换的效率。

（4）授权日志

授权日志的记录对于保障系统的安全性至关重要。管理员对人员的授权记录，包括添加用户、更改或删除权限等操作，都需要被详细记录。这些记录不仅有助于追踪权限变更的历史，还可以在发生安全事件时提供调查的依据。

6.2 数据连接器的分层解耦架构

IDS 提倡开放与兼容，华为在进行数据连接器这一核心模块的设计时，也遵循这个理念。数据连接器从业务作业和数据处理两个层面考虑解耦，它在设计时被分解为独立运作的模块，实现了高度的解耦和灵活性。这种设计允许每个模块专注于特定的功能，简化了维护工作，因为更新或修复单个模块不会影响其他部分。数据连接器分层解耦架构如图 6-18 所示。

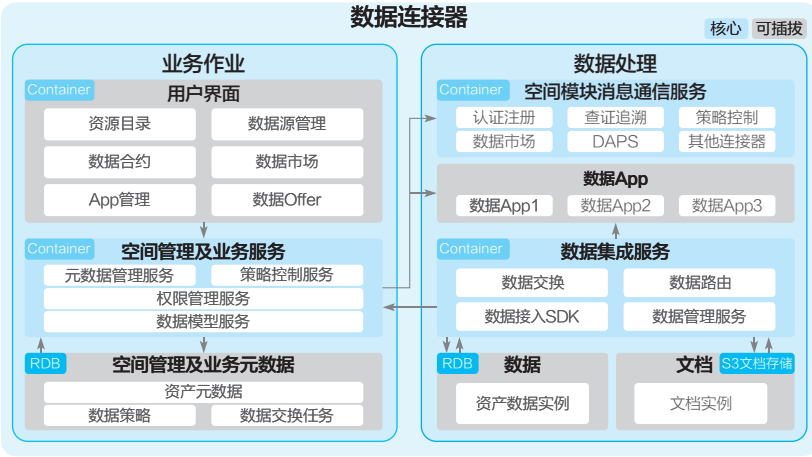


图 6-18　数据连接器分层解耦架构

数据连接器的主要功能包含了业务作业与数据处理，我们在设计上考虑将这两个功能进行解耦以获得更多的灵活性。同时，空间管理及业务服务、数据集成服务、空间模块消息通信服务等作为核心关联组件需要在数据连接器中内置，其余的用户界面、存储等组件支持插拔式扩展。下文针对数据连接器的分层解耦与插拔式设计做进一步阐述。

6.2.1　数据连接器的业务作业面与数据处理面

在华为数据连接器中，空间管理及业务服务、数据集成服务、空间模块消息通信服务三个组件是核心，其他组件为可插拔设计，可以替换为其他同功能的通用组件。核心组件中，空间管理及业务服务组件负责管理数据连接器的启动配置信息、内部服务的部署配置信息、组件交互相关的配置信息，以及数据连接器的内部业务功能和元数据。数据集成服务组件负责数据实例的管理，由空间管理及业务服务组件发起指令对数据集成服务组件进行调用，实现数据实例的存取更新、文档的存取更新、数据交换的任务发起和数据传输等功能。空间模块消息通信服务主要负责与周边组件交互的协议标准，比如数据连接器与认证中心，数据连接器与清算中心，数据连接器与数据连接器。除了数据连接器内部的协议外，华为现在也在加入一些对标准协议的适配，比如欧洲数据空间的数据交换协议（DSP），该组件会提供内部协议到 DSP 互转的适配。数据连接器的业务作业面与数据处理面如图 6-19 所示。

华为数据连接器将业务作业和数据处理进行了分层，包括业务作业面和数据处理面。

- 业务作业面负责业务的操作，这些操作产生的业务数据（非实际交换的数据）由业务作业面进行管理，比如用户请

求的合法验证、合约的创建、合约中使用控制策略的选择、合约的协商、用户管理等。同时业务作业面会调度数据处理面的接口对业务操作的结果进行处理。

- 数据处理面负责需要交换数据的处理，比如数据连接器数据源的管理、交换数据的接入、数据的传输、数据交换任务切片，还支持实时数据流的解耦和传输、大文件的稳定传输等。

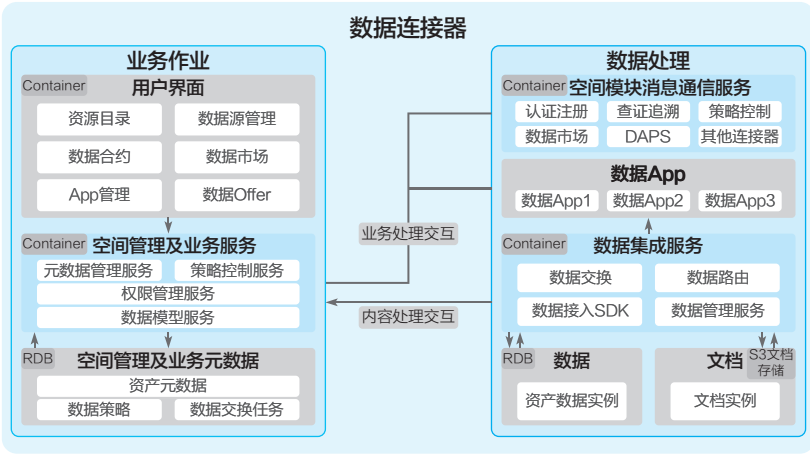


图 6-19　数据连接器的业务作业面与数据处理面

业务作业面和数据处理面之间的明确分工确保了业务作业的高效、数据传输和管理的安全性与合规性。业务作业面的结果和策略影响数据处理面的数据传输，而数据处理面的传输状态也会反馈给控制平面，以便于进行进一步的控制和管理。业务作业面的数据和数据处理面的数据分开管理会带来如下优势。

- 业务作业性能提升：业务作业数据的量级比交换数据的量级要小很多，分不同的库存储可以提升业务作业和数据传输的效率，特别是资源目录的查看、搜索、新增等。

- 数据连接器元数据统一管理：业务作业与数据处理分离，可以进一步推动服务间标准接口的设计，使数据连接器具备较好的可扩展性。
- 提升数据连接器的安全性：业务作业数据与交换数据分离后，数据的安全策略和数据实例的安全策略可以分开实施，用户作业在非必要情况下不接触实际的数据。数据连接器也支持将数据处理部署到业务本地，从而使数据所有者能够保持对其数据的完全控制，防止数据滥用和泄露。

6.2.2　数据处理面数据集成服务的解耦架构

数据处理面对数据的管理方式多种多样。例如，对于非结构化数据，可以使用文件、FTP、SFTP、OBS 协议、S3 协议进行传输；对于结构化数据，可以使用 CSV/Excel 文件、API、云服务（例如华为云的云数据迁移 CDM 服务）以及 DB 方式进行传输。当使用 DB 方式时，还需要支持不同的数据库（例如 MYSQL、PostgreSQL、DWS、openGauss 等）。

如果对每个能力都单独开发，不仅工作量大，还不能快速部署以响应用户，而且没有统一的标准也容易出错，导致难以维护。为了解决上述问题，华为将数据连接器的数据集成能力封装成独立的数据集成服务，并且使用可插拔的方式对数据传输与处理能力进行扩展，其架构如图 6-20 所示。

如图 6-20 所示，数据集成服务使用了分层的架构设计，分为接口层、核心层与插件层。

- 接口层：提供标准的集成接口给数据连接器进行调用，数据连接器无须知道实现是 OBS 协议、集成服务还是数据库等，只需要按业务需求调用接口即可。

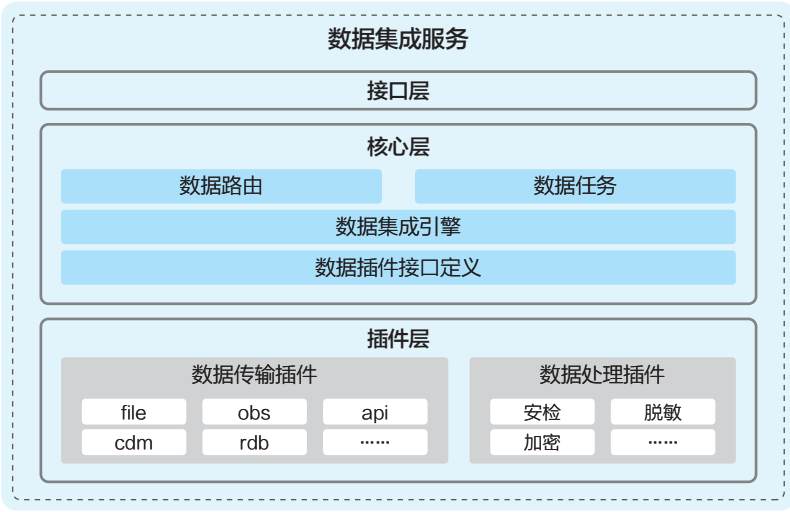


图 6-20　华为数据集成服务的解耦架构

- 核心层：包括数据路由、数据任务、数据集成引擎、数据插件接口定义。数据路由主要用于设置数据传输的节点和传输顺序；数据任务负责对数据传输的任务进行管理；数据集成引擎用于实现具体的数据集成逻辑；数据插件接口定义负责定义集成接口的实现标准，并提供给下层的插件层。
- 插件层：按核心层定义的集成接口实现标准对各项插件能力进行实现，也可以直接使用已有的集成能力进行封装，比如图 6-20 中的 cdm 是一个集成云服务，通过将云服务的能力按标准进行封装并与核心层对接，即可使用该能力。

此外，该架构还具有插件的热插拔设计。插件的热插拔是指在不重启应用的情况下，能够动态地添加、更新或删除插件。可以将插件做成微服务并部署后，与数据集成服务的核心层进行对接，以完成插件的热插拔。由于是通过微服务间的接口进行调用的，因

此插件的性能会比默认在集成服务内实现的性能差一些。

上述这种可扩展的插拔式设计使数据连接器能够根据业务需求轻松扩展数据传输与数据处理能力，提高了系统的适应性和竞争力。此外，各个数据插件之间互相隔离，即使某个插件出现问题，也不会对整个系统造成影响。

在业务实际应用中，不同的业务用户和使用场景对数据连接器有很多差异化的需求，主要体现在 2 个方面。

（1）满足灵活扩展的需求

数据连接器被用于接入各类不同的数据类型，需要支持不同的数据来源，传输的数据也从简单的文件到复杂、海量的结构化数据，这对采集、交换、存储的组件能力要求很高。数据连接器需要按照不同场景下的客户要求进行灵活扩展，同时满足性能要求。

（2）兼容复杂环境设施

部署数据连接器时需要考虑如何兼容各个参与方现有的 IT 软硬件设施。例如，在一个广泛开放的数据空间生态中，数据连接器可能由多个软硬件提供方来提供和定制。这种情况下，数据连接器的架构设计需要考虑兼容问题，如用户界面支持定制、元数据管理兼容企业现有的工具等。这就要求数据连接器预留兼容外部组件的接口，以在实际应用中实现与不同 IT 环境与软硬件设施的兼容。

从数据连接器的架构与业务实际需求考虑，本书在设计上划分出了数据连接器的核心组件和可插拔的非核心组件。数据连接器的空间管理及业务服务、数据集成服务、空间模块消息通信服务三个组件为核心组件，在设计上要求严格遵循标准规范，并内置到数据连接器中，避免差异化定制；用户界面、数据存储、数据消费应

用等组件设计为可插拔的非核心组件，用户可以替换为其他同功能的通用组件。兼容可扩展的插拔式设计使得系统能够根据业务需求轻松扩展，无须重写整个系统。这种灵活性还允许开发者根据特定任务选择最合适的技术和工具，提高了系统的适应性和竞争力。此外，核心组件与非核心组件的分离增强了系统的安全性，即使非核心组件出现问题，也不会对整个系统造成影响。

6.2.3 数据连接器南向环境资源解耦架构

在数据连接器中，南向的环境资源与平台进行对接，北向的环境资源与三中心和市场进行对接。由于数据连接器的部署环境多样，可以是公有云、私有云或一体机，因此将平台资源的能力实现和数据连接器解耦，数据连接器负责提供统一的南向标准定义。对数据连接器来说，不管什么平台，能力标准都相同，差异的部分由各类环境资源平台按标准进行适配。比如华为内部文件的存储使用 EDM，在公有云上使用的是 OBS，但对数据连接器来说都是文件存储，使用标准的南向接口就可以。内部对 EDM 的实现由内部团队负责，公有云上的 OBS 由公有云团队实现，数据连接器可以不做修改，这样可以提高数据连接器的部署效率和通用性。

华为的数据连接器提供了一套统一的多云适配标准，涵盖了计量、订单、租户中心、用户登录、区块链、水印、组件认证、证书管理以及基础数据等多个方面。这些标准已经成功应用于华为私有云、华为公有云以及存储一体机的实现中。部署到私有云时，会启用私有云的组件实现和配置；部署到公有云时，会使用公有云的组件实现和配置；而部署到一体机中时，则采用存储一体机的组件实现和配置。这样的策略确保了数据连接器在不同环境中的灵活性和一致性。数据连接器南向环境资源解耦架构如图 6-21 所示。

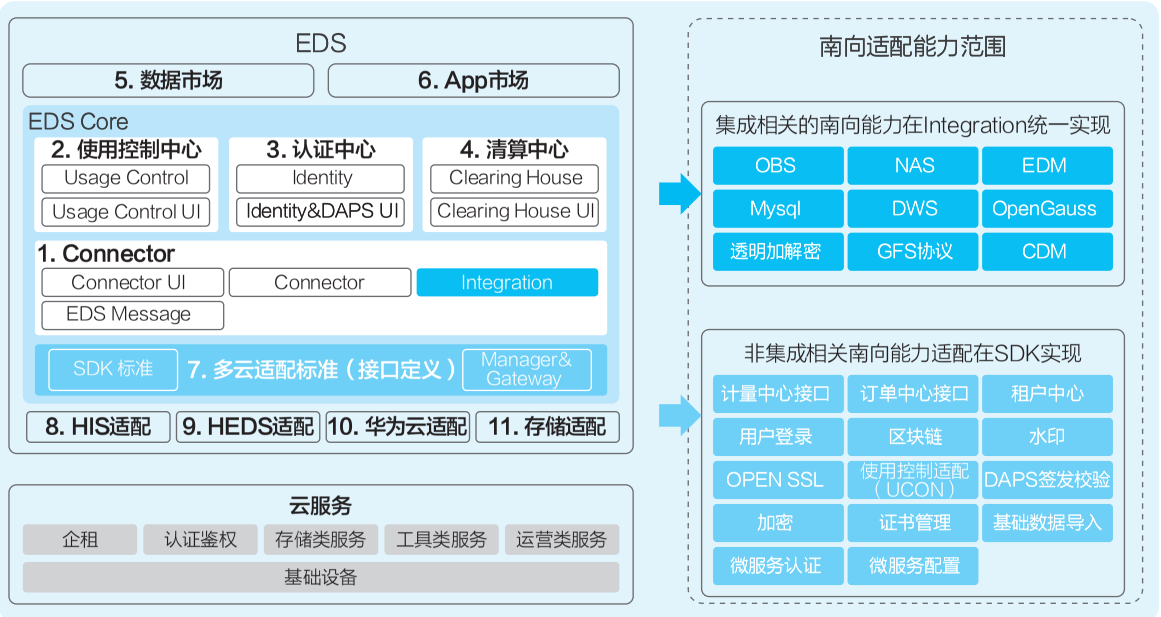


图 6-21　数据连接器南向环境资源解耦架构

在南向资源接口的设计中，需要我们遵循以下关键规则。

- 定义清晰的接口契约：通过定义标准化的接口，南向资源可以与上层软件系统保持松耦合，使系统能够灵活地适应变化。
- 版本控制和兼容性：维护接口的版本，确保新旧版本之间的兼容性，使南向资源的更新不会影响上层应用。
- 配置管理：提供灵活的配置管理机制，使数据连接器可以根据不同环境自动调整其行为和配置。
- 模块化设计：将南向资源接口设计为独立的模块，便于单独开发、测试和维护，以提高系统的可维护性。
- 微服务架构：采用微服务架构，将南向资源作为独立的服务运行，通过轻量级的通信协议进行交互，实现高度解耦。
- 安全性设计：通过使用 HTTPS、OAuth 等安全协议，确保南向资源与软件系统之间的通信安全。

6.2.4 数据连接器与数据空间中组件的交互关系

数据空间中，数据连接器是最重要的功能模块，同时数据连接器间的互相交换也需要与数据空间的其他能力与服务进行集成，以确保数据空间有效运行。需要与数据连接器进行集成协同的组件包括认证中心、策略中心、数据市场及 App 市场、清算中心等，主要组件间的交互关系如图 6-22 所示。

对各组件间交互关系的具体说明如下。

1. 数据连接器与数据空间管理组件

从业务流程的角度来说，数据连接器是在开通数据空间之后由业务按需创建和分配产生的，其生命周期由数据空间管理组件管理。



图 6-22　数据空间中主要组件间的交互关系

2. 数据连接器与认证中心组件

数据连接器软硬件、数据连接器内的参与主体、用户等需要一套身份管理机制。数据连接器依赖认证中心提供的基础注册认证服务。获得认证后的各类身份信息，如数据连接器的数字认证证书，企业、用户的认证信息等，用于支撑数据连接器的启用与初始化。

3. 数据连接器与策略中心组件

数据空间内数据的跨主体、跨边界可控交换，以及现实合约在数字世界中的有效实施，依赖策略中心组件提供的策略服务。策略中心组件定义并下发策略到数据连接器，数据连接器中策略的评估与决策也很大程度上依赖策略中心组件，这是数据连接器进行数据可控交换的基础。

4. 数据连接器与数据市场及 App 市场组件

有了数据连接器，用户就具备了交换数据的基本条件。但在实际业务中，依然有两个重要的问题。一是如何高效地检索、分享有哪些可供交换的数据资源，这依赖于数据市场组件。数据市场接收数据连接器上架的数据 Offer，通过发布的方式实现更大范围的数据开放，促进数据有效流通。二是如何受控地使用数据。数据连接器作为一个安全可信的容器，可以部署各类 App，不同的 App 具备不同的业务属性，业务属性取决于业务的具体场景，类似于安卓市场。App 市场可以构建一个接入数据连接器的通道，经过 App 市场注册认证的 App，可以部署到数据连接器中，实现受控的数据消费。

5. 数据连接器与清算中心组件

数据空间内的各类交换过程记录，通过数据连接器中的各类

探针采集后上报到清算中心进行存证。基于清算中心的数据，数据空间管理者和各数据连接器的管理者可以获得查证追溯和统计清算等服务。

6.3　适应多场景、多环境的数据连接器

6.3.1　数据连接器的三种使用场景

数据空间中的数据连接器主要承担了数据交换的工作，并且具备数据留痕、数据处理与使用控制三个核心特性。这三个特性在不同的使用场景中发挥对应的作用，如图 6-23 所示。

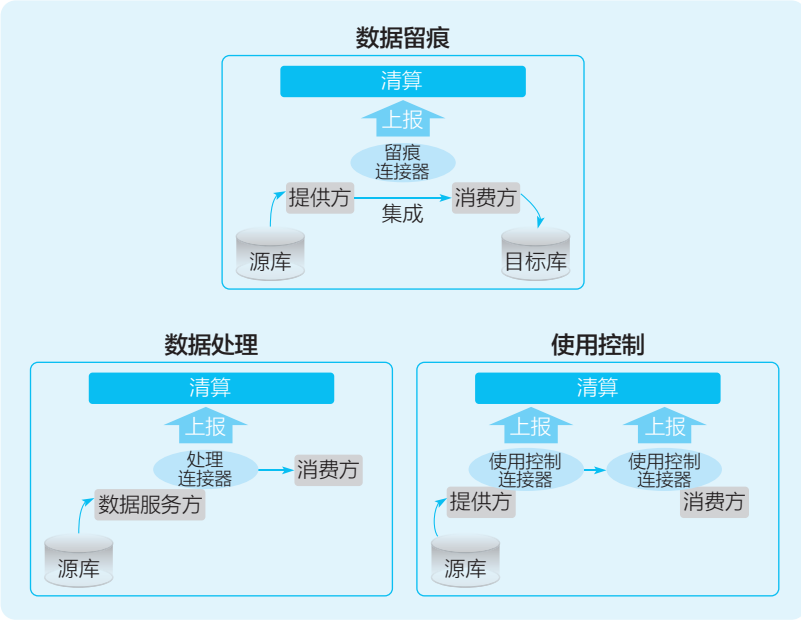


图 6-23　数据连接器的核心特性

数据连接器的三种使用场景说明如下。

1. 数据留痕

数据连接器中的所有数据交换过程与行为记录都通过非破坏式的方式被记录和保存。除了在数据连接器中进行分布式保存外，相关记录还会同时上报给清算中心。

2. 数据处理

对于数据服务方需要通过数据连接器进行专业的数据加工处理，数据连接器提供了接入各类应用并进行数据处理的能力，数据服务方基于数据连接器可以实现在合规安全的前提下为其他参与方提供专业的增值服务，同时记录并上报处理过程。

3. 使用控制

数据提供方通过数据连接器对数据设置使用控制策略，数据消费方按照要求使用数据，受到使用控制策略的强制约束。数据连接器的使用控制能力可以确保使用与控制的过程被记录并上报，实现了数据抵达消费方之后，数据提供方还能通过技术对数据如何被使用进行有效管控，从而保证了数据交换合约的有效履行。

6.3.2 数据连接器的三种安全级别形态

数据连接器的目标是提供一个安全可信可控的集成环境。结合工程技术实现，我们发现数据连接器可以在应用层、环境层，以及物理硬件层实施使用控制策略。

在业务使用数据连接器进行数据交换时，通常对数据连接器的安全可信能力有三种不同程度的要求，具体如下。

- 在企业内部基础环境可信与管控程度较好的情况下，仅需要在应用层面叠加使用控制策略。
- 在企业间进行交换时，数据连接器的部署环境相对复杂，除了需要在应用层叠加使用控制策略外，还需要对操作系统与存储资源进行有效的管理。
- 重要数据流通场景对数据的完全可控有更高的要求，需要在应用、环境、硬件等层面都应用使用控制策略，实现最强的管控力度。

华为数据连接器通过工程技术的实现，能够在应用层、环境层以及物理硬件层进行细致的控制。应用层控制涉及软件和数据处理流程，环境层关注操作系统和网络环境的安全，而物理硬件层则确保服务器和存储设备的安全。通过在每一层采用适当的技术控制手段，能够实现信任和控制力度的逐层增强。通过这种方式，确保了信任体系的持续性，同时也使数据连接器具备了面对不同场景的适应性和弹性，如图 6-24 所示。

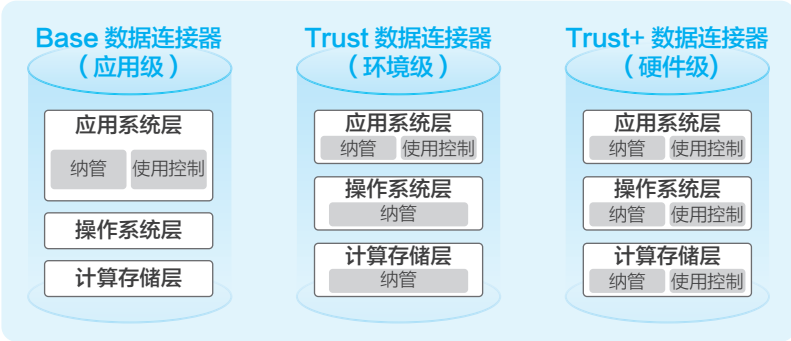


图 6-24　三种不同场景下采用的数据连接器

图 6-24 中三种数据连接器适用不同的应用场景，它们的部署成本与控制力度各有不同，下文针对三种数据连接器的差异做详细说明。

1. 应用级的控制（Base 数据连接器）

应用级的控制是指通过对应用程序和数据应用进行加强，确保数据在应用层面上的安全。通过安全认证技术，向经过认证的各类 App 开放数据连接器的各层访问和使用权限，可以有效地防范针对应用程序的漏洞攻击，同时通过使用控制、数据加密、安全审计等手段，对数据交换和消费过程进行有效的管控和记录。由于只需要在应用层面做改造，因此成本较低，可以快速实现部署应用。

2. 环境级的控制（Trust 数据连接器）

环境级的控制是指通过对操作系统和网络环境进行加强，确保数据在系统上的安全。将数据连接器部署在可控的环境中，数据连接器采用的操作系统、存储设备都可以被统一纳管，从而减少违规访问的暴露面，有效降低操作系统和存储设备带来的安全威胁。相比于 Base 数据连接器，Trust 数据连接器可以提供增强的控制和安全等级，同时因为需要对环境进行纳管，所以会带来额外的环境设施统一管控成本。

3. 硬件级的控制（Trust+ 数据连接器）

硬件级的控制是指通过对硬件设备进行加固，确保数据在物理上的安全。例如采用加密芯片、安全存储技术、物理隔离等手段，限制数据的访问范围，可以在硬件层面实现数据的安全管控与防护，有效防止数据被非法获取和篡改，从而实现最高等级的保护，这就要求存储设备的底层硬件完全可控。因为需要将使用控制技术叠加到硬件层级，所以相应的改造成本也最高。华为硬件级控制的数据连接器设计如图 6-25 所示。

硬件级控制实现了数据使用控制、应用安全执行环境、硬件级的安全存储等软硬件能力的结合，提升了数据安全可信的防护能力，其主要特性如下。

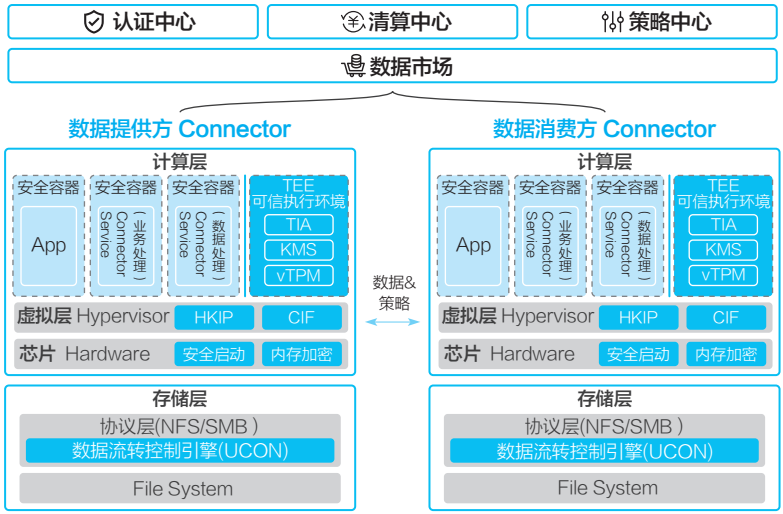


图 6-25　华为硬件级控制的数据连接器设计

- 提供数据可信执行环境（TEE）能力，包括硬件级的认证、秘钥管理，以及敏感及高密数据的存储，对接后可提供更高的安全能力。
- 支持流转数据文件透明加解密，且加解密支持硬件加速，并提供数据安全存储和彻底销毁的能力。
- 在存储层提供了数据流转控制引擎（UCON），将其与数据连接器的使用控制打通即可提供硬件层面的数据控制，例如在存储层实现数据的彻底删除。

6.3.3　数据连接器的三种部署模式

EDS 数据连接器的组件可以分开部署或集中部署，例如根据数据连接器的配置将数据连接器的各个组件连接，实现协同运作，从而完成数据连接器的完整功能部署。数据连接器在不同环境下的部署模式如图 6-26 所示。

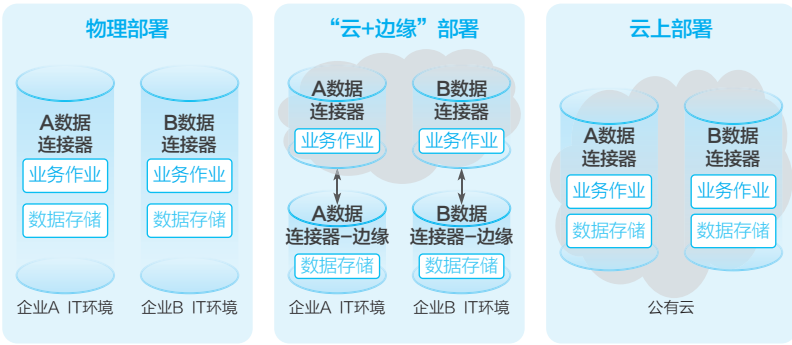


图 6-26　数据连接器在不同环境下的部署模式

数据连接器三种部署模式的具体特点与适用场景如下。

1. 物理部署

数据连接器的所有组件部署在一个区域内，一般是在企业内部。所有的组件和资源由部署企业独享并拥有最强的控制程度。

物理部署可以带来很多优势，诸如控制程度的增强，可以实现有效控制硬件、软件与数据、网络等，减少暴露风险。同样地，物理部署因为需要更多的软硬件资源和维护人员而带来成本的增加，数据连接器的扩展和维护更新也会变得更加复杂。这种模式适用于对数据控制和安全性有高要求、对成本不敏感的场景，在实际业务中一般适用于企业内外重要、涉密等高防严管控的场景。

2. “云 + 边缘”部署

数据连接器的业务作业面部署在云上，共用云上资源；而数据处理面的大部分组件部署在企业的边缘一体机中，数据流由用户自己管控。元数据的存储和管理在云上完成，信息流由云服务商管控。

“云 + 边缘”部署模式带来的好处诸如数据在本地完成存储和处理，符合数据本地驻留的管理要求；边缘设备对云上依赖小，可以在无网络连接时继续运行；本地使用数据时会获得更好的性能。同时，“云 + 边缘”部署模式也会增加管理和互联的复杂性。这种模式适用于需要实时处理数据以及数据需要本地驻留的场景，在实际业务中一般适用于不同企业之间进行数据交换的场景，各方对数据的本地存储有较强诉求，同时也有各自的本地部署资源。

3. 云上部署

数据连接器的所有组件都部署到云上，数据和元数据的管理与存储都在云上完成，并通过云上多租的方式实现隔离。

云上部署模式显而易见的优势是总成本最低、可扩展性好、可用性高以及访问更加便捷。同时，因为需要访问云上资源，在性能保证、安全与隐私保护、网络质量等方面会存在较多的挑战。这种模式适用于关注部署便利性及成本的场景，但可能在性能和安全方面有所妥协，在实际业务中一般适用于需要大量复杂的数据交换但数据保密等级要求较低的场景，例如集团内部组织间的资料共享。

针对如何选择部署模式，下面提供 2 个华为应用实践。一个是企业内部各部门间重要数据的共享场景。考虑到各部门共用一个云平台，而且数据交换发生在企业内部网络中，组织间的信任基础比较高，所以选择部署成本最低、访问最快捷的云上部署模式。另一个是华为与生态伙伴协同的场景。考虑到伙伴的 IT 能力各不相同，有实力的伙伴都有自己的 IT 基础设施，综合考虑伙伴现有环境、部署能力以及接受度后，华为选择混合部署模式。有实力的伙伴可以将数据连接器部署到其指定的 IT 环境中，其他伙伴可以选择在华为云上通过开租获取数据连接器。详细的部署说明可参见第 9 章实践案例部分。

6.4 小结

在数据空间中，数据连接器作为核心模块，承担着数据留痕、处理与控制的关键职能。它通过确保记录过程不破坏数据，保持数据的完整性，同时对数据处理和使用控制策略进行精细化管理，以确保所有操作的透明性和可追溯性。

首先，在工程落地设计上，数据连接器采用分层解耦和兼容可扩展的架构，将业务作业面和数据处理面分离，形成独立运作的模块。这种设计不仅提升了系统的灵活性和可维护性，也强化了其安全性。数据连接器的互操作性通过与其他数据空间能力与服务的集成得到加强，如空间管理与业务服务、认证中心服务、策略中心服务和清算中心服务等，共同确保数据空间的高效运行。

其次，为了实现数据的可控消费，数据连接器为应用程序提供了四种接入方式：原生 App、工具组件类 App、独立软件类 App 和 SaaS 类 App。华为提供相应的认证机制、控制技术和通用接口标准与 SDK，确保数据访问的安全性和便捷性。

最后，针对不同业务场景，华为开发了三种类型的数据连接器：Base、Trust 和 Trust+。这些数据连接器在应用层、环境层、物理硬件层提供不同级别的控制力度，构建起一个多层次的信任保障体系。为了适应不同环境下的部署需求，EDS 数据连接器支持物理部署、“云 + 边缘”部署和云上部署等多种模式。组织可以根据自身需求、资源、预算和技术能力选择最合适的部署模式，甚至可以采用混合部署模式，以兼顾灵活性和效率。

7

第 7 章 CHAPTER

数据空间中的可信、可控、可证

数据连接器是在数据空间中实现数据可控交换的核心组件，为数据流通提供了基础条件。然而，要支撑数据空间这样一个复杂体系的运行，还需要在系统层面进一步构建能力。华为提出，数据空间系统应围绕三个关键方向构建体系化能力：可信、可控和可证。本章将详细探讨华为如何在数据空间中设计并实现这些能力。

7.1 什么是可信、可控、可证

数据所有者在数据流通中存在顾虑，主要表现为“不敢、不愿、不能”充分交换数据。“不敢”是因为数据流通会增加泄露风险，可能导致利益和商誉受损；“不愿”是因为数据包含企业的技

术和商业秘密，泄露意味着竞争力受损；“不能”是因为企业间的数据保护机制和技术手段不一，难以保证数据在流通过程中得到有效保护和管控。面对广泛的数据流通需求，技术难度成为当前存在的主要问题。

数据流通中“不敢、不愿、不能”这些受到普遍关注的问题，其实在实物的流通中也存在，比如高价值、高保密性、易损物品的流通。例如高价值的货币的运送，高保密等级的考试试卷的运送，易损的生鲜的运送等。在这些实物的运送中采取了更多的保障手段：银行的货币押送需要对押送企业、人员进行严格的审查监管，以确保身份可信；生鲜的运送需要使用冷链专用设备，以实现全程温度可控；考试试卷的运送则配有全程的监控和定位手段，以实现每份试卷去向的可追溯。参考类似的设计，在数据的流通中，也需要设计一套完善的数据流通体系，这套体系应该围绕“可信、可控、可证”的目标去构建，从而对数据这种特殊物品的流通实现有效保障。

- 可信是指数据空间中的企业、人员和平台组件的可信程度。这意味着在数据交换和处理过程中，确保所有参与要素的可信性是至关重要的。也就是说，人员和平台组件需要被认为是可信的，以确保数据的安全和对数据的保护。这不仅是一种信任，也是一种责任，是构建数据空间的基础能力。
- 可控是指在进行跨主体和边界的数据交换时，通过技术手段实现精准控制。这意味着数据的采集、存储和使用都应在严格的控制下进行，以防止数据的滥用或误用。只有这样，才能确保数据在流通过程中不被破坏、篡改或泄露。通过可控保证各方的权益，从而使数据提供方愿意供应高价值的数据。

- 可证是指数据提供方可查证追溯，数据消费方可自证清白，数据运营方可统计清算。这意味着在数据交换和处理过程中，所有的操作和结果都应该是可查证的，无论是提供数据的一方，还是消费数据的一方，都有权知道数据的来源、处理过程和结果。这种透明性和可追溯性是维护数据空间公正和公平的重要手段。

这三个目标是数据空间体系的价值目标，是实现数据安全、保护数据隐私的重要保障。在具体的工程实现过程中，需要充分利用技术创新，以实现更高程度的可信、可控和可证。

7.2　基于数据空间信任框架保障可信

企业间的数据交换涉及敏感和高价值资源，因此构建一套用于数据交换的信任框架至关重要，这样的机制能够确保数据的安全性、完整性和准确性。通过加密技术、访问控制和数据脱敏等手段降低数据泄露、滥用和篡改的风险。通过确立透明的数据使用规则和强有力的访问控制机制，参与企业能够建立起相互信任，从而营造更加开放和高效的数据共享环境。此外，信任框架还包括对数据交换过程的监控、记录和审计，这有助于提升数据管理的合规性，并在发生争议时提供清晰的责任追溯。本章就华为如何围绕可信、可控、可证三个目标完善数据空间的核心能力做进一步阐述。

7.2.1　数据空间的信任框架

数据空间的信任框架为数据空间内的信任和安全提供了一套定义明确的方法与措施。信任框架是数据空间的基础层，确保了数据空间的参与者能够在一个安全、可靠的环境中进行数据的交换和共享。华为数据空间的信任框架如图 7-1 所示。



图 7-1　华为数据空间的信任框架

在信任框架中需要重点定义出以下标准和规则。

- 建立身份验证规则：通过认证和评估机制，信任框架可确保只有合规的组织和个人才能参与数据空间，从而建立一个可信的生态系统。华为数据空间身份验证规则将需要认证的对象分为三层：第一层是对用户、组织的认证；第二层是对组件的认证，包括连接器、App 等；第三层是对环境资源的认证，包括容器、操作系统、硬件 / 虚拟机等。
- 保障数据安全：信任框架包括一系列安全措施，如加密、访问控制和数据使用监控，以保护数据不会被未授权访问或滥用。数据安全规则主要包含数据存储的加解密、数据传输的加解密、数据的访问控制以及数据的使用控制。数据的使用控制将在后面的章节中进行详细说明。

7.2.2　空间组织及人员可信

在华为的数据空间中，对组织的认证包括如下方面。

1）基于全国组织机构统一社会信用代码库标准接口提供的信息提交系统认证申请，审批通过后即可通过认证。统一社会信用代码是用于唯一标识中华人民共和国境内依法注册、依法登记的机关、企业、事业单位、社会团体、民办非企业单位的识别标识码，类似于居民身份证，是组织机构在社会经济活动中的统一身份证明。

2）通过华为 IAM 对组织进行认证，审批通过后即可通过组织认证。华为云对企业的认证是基于企业所提供的相关信息并进行身份核验来完成的，包括法人认证和被授权人认证两种方式。法人认证需要提供企业证件、法定代表人姓名与身份证号；被授权人认证则需要提供企业证件、法定代表人姓名与身份证号、加盖单位公章的授权书、被授权人姓名与身份证号。

在华为的数据空间中，对人的认证具体通过如下手段实现。

1）华为数据空间使用 IDaaS 服务进行认证，华为云则使用云上 IAM 进行认证。

2）数据空间中的认证中心提供对标准认证协议的支持，用户可以通过标准认证协议与企业内部认证系统进行对接，比如 OIDC、OAuth、DID 协议。认证通过后，数据空间的认证中心会给用户确定一个唯一的 ID，并且建立用户和组织的关系，以便在后续访问中证明用户所属的组织。

在数据空间身份认证中，分布式数字身份（DID）也是需要考虑的认证方式之一，如图 7-2 所示。它基于区块链技术，允许用户

拥有和控制自己的身份信息。DID 的核心优势在于去中心化、身份自主可控、可信的数据交换、完全自主研发可控、技术设施标准化、高性能 DID 操作、异构通用性、数据安全性、身份可控性以及开箱即用。

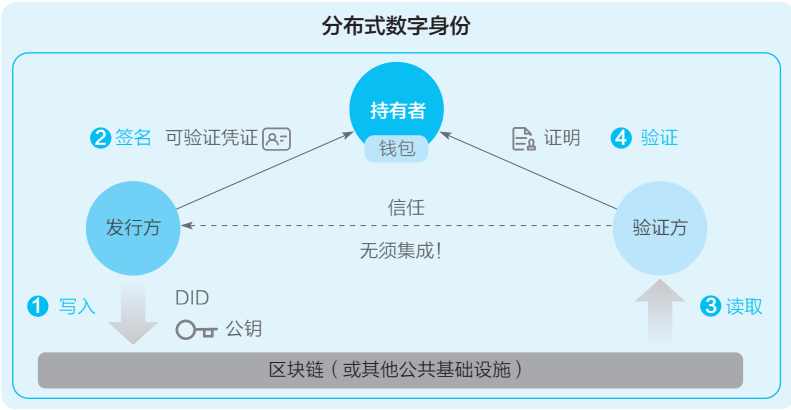


图 7-2　分布式数字身份认证示例

当不同数据空间的身份需要互认或数据空间需要与企业认证中心对接时，会出现两种或多种的认证方式或认证协议，这样会导致身份认证不能互通。基于这种场景，项目组开发了一个认证凭据互认组件：凭据桥。凭据桥允许不同系统和网络之间安全地交换与验证用户的身份凭证。目前华为慕尼黑实验室实现了一个凭据桥来解决跨域的身份认证问题，同时也可以用来解决数据空间与企业认证的快速对接问题。在凭据桥中，会将不同鉴权协议转换为一个标准协议，再转换成目标认证的鉴权协议。目前标准协议使用 W3C 的 DID 标准和 Verifiable Credentials 标准。这些标准确保了凭证的互操作性和可移植性，使用户可以在不同的服务和平台之间无缝地使用自己的身份凭证，具体设计如图 7-3 所示。

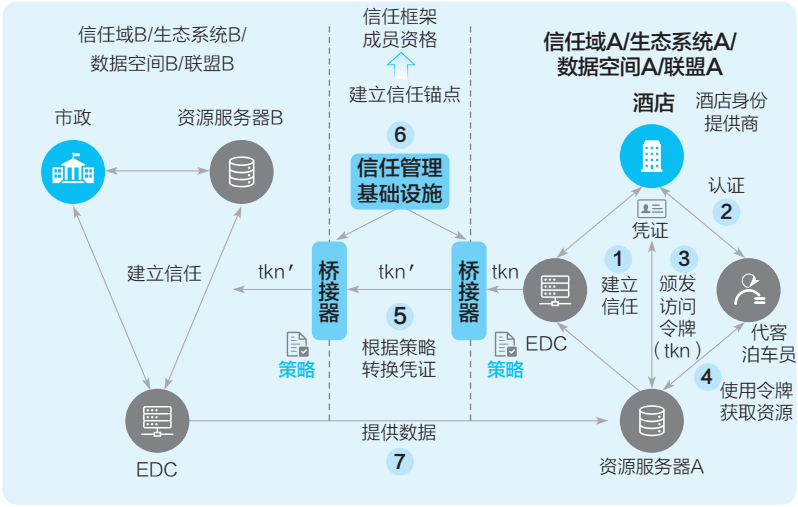


图 7-3　跨域身份认证设计示例

图 7-3 以酒店代客泊车访问市政停车系统为例，展示了基于凭证桥接层的跨域身份认证方案：作为前置条件，酒店（信任域 A）先与自身部署的 EDC 数据连接器建立信任（步骤①），确保 EDC 可安全代表酒店参与跨域交互），市政（信任域 B）也与自身的 EDC 数据连接器建立信任（确保 EDC 能代表市政参与跨域交互）。在此基础上，酒店代客泊车员先通过酒店身份提供商完成认证（步骤②），获取访问令牌（tkn，步骤③），再使用该令牌向资源服务器 A（酒店管理系统）获取停车数据（步骤④）；资源服务器 A 收到请求后，将请求转给酒店 EDC 处理，酒店 EDC 将该令牌传入酒店侧的桥接器，桥接层依据策略（Policy），结合 TRAIN（TRust mAnagement INfrastructure，信任管理基础设施）验证双方信任锚（如联盟资质，步骤⑤和步骤⑥），将酒店令牌转换为市政信任域认可的凭证（tkn'）；转换后的凭证经市政侧的桥接器传递至市政 EDC 与资源服务器，市政 EDC 基于与市政的信任关系校验凭证的有效

性后，向资源服务器 B（市政停车系统）请求并获取停车数据（步骤⑦）。桥接层通过策略化凭证转换与去中心化信任校验，打通酒店（信任域 A）与市政（信任域 B）的身份互认，解决跨域信任割裂、流程烦琐的痛点，实现身份自动互通，降低集成成本。

7.2.3 数据服务及组件可信

华为数据空间对空间内的组件使用 DAPS（Dynamic Attribute Provisioning Service）进行交换时的认证和管理。所有的组件都要在 DAPS 中进行注册，注册组件需要包含组件的基本信息、证书信息、信任库以及动态认证属性，如图 7-4 所示。



图 7-4　华为数据空间使用 DAPS 对组件进行认证和管理

DAPS 也会对数据空间中的证书进行统一管理，所有 EDS 组件共用相同的根证书。通过将根证书加入信任库，可以实现 EDS 组件间的 HTTPS 通信由 DAPS 统一对 EDS 组件的证书进行登记注册。使用白名单机制，可对 EDS 组件进行认证。EDS 组件在通信时会携带 DAT（动态属性 Token），用于进一步确定身份。该 DAT

通过 DAPS 颁发与校验，DAPS 负责维护 EDS 组件的关键属性（例如 Connector 的安全级别），并将这些属性附加到 DAT 中。

DAPS 的动态属性管理功能允许在不直接修改 X.509 证书的情况下，动态地更新和管理与参与者身份相关的属性。这些属性可能包括安全状态、成员身份、诚信等级或组织地址等。动态属性的更新可以通过 DAPS 进行，无须重新颁发证书，这减少了证书撤销的需求，并为 EDS 中的参与者提供了更灵活的属性处理方式。DAPS 认证过程如图 7-5 所示。

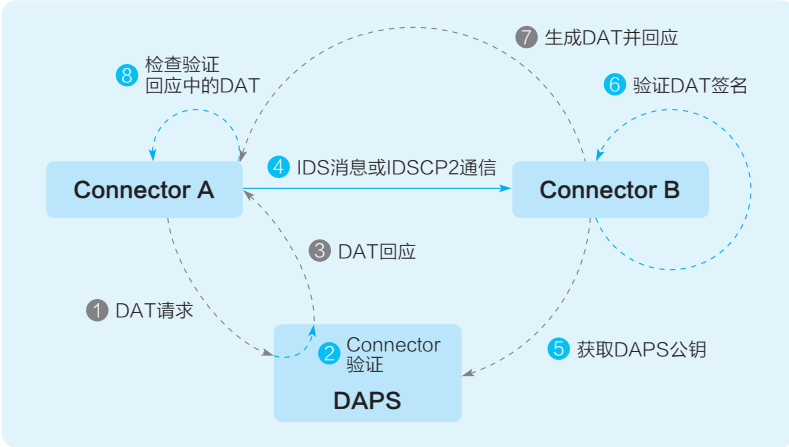


图 7-5　DAPS 认证过程

DAPS 认证的具体过程如下。

1）Connector A 生成 DAT 请求（包含发起方标志 iss），使用自身私钥签名，并在请求时指定动态属性 scope。

2）DAPS 根据 iss 从白名单中查找 Connector A，并用相应的公钥进行签名验证。

3）DAPS 生成 DAT（包含 kid 与动态属性），使用自身私钥签

名，并发回给 Connector A。

4）Connector A 将 DAT 缓存，在传输 IDS 消息或进行 IDSCP2 通信时发送给 Connector B。

5）Connector B 从 DAPS 中获取 DAPS 的公钥列表，并根据 DAT 中的 kid 从返回的列表中得到公钥。

6）Connector B 使用从上个步骤中得到的公钥对 DAT 进行签名验证，验证通过后才进行后续处理。

7）Connector B 与 DAPS 通信得到 DAT（类似 Connector A 的步骤 1 ～ 3），然后将该 DAT 随同响应数据一起发给 Connector A。

8）Connector A 收到 Connector B 的响应后，对响应中的 DAT 进行签名验证（类似 Connector B 的步骤 5 ～ 6）。

DAPS 的认证结果日志记录和客户端认证过程会在 DAPS 中详细记录，以支持后续问题的定位和回溯。

7.2.4 平台及环境资源可信

在数据空间中，平台及环境的可信可采用可信计算来保障。在 EDS 一体机中，华为使用鲲鹏一体机中的可信计算体系来保障硬件及操作系统的可信。

在云部署的情况下，一般要求云平台需通过专业的国际和本地的云合规认证，比如 ISO/IEC 27001 认证标准和网络安全等级保护标准。

ISO/IEC 27001 是目前国际上被广泛接受和应用的信息安全管理体系认证标准。该标准以风险管理为核心，通过定期评估风险和对应的控制措施来有效保障组织信息安全管理体系的持续运行。公有云获得 ISO/IEC 27001 认证证书，表明其能够提供一个符合国际公认信息安全标准的云环境。此外，针对不断变化的云环境和服务，

公有云还会按照认证要求定期邀请第三方机构进行复审，以保障云服务的信息安全管理能力一直处于业界领先水平。这包括但不限于 DDoS 高防、Web 应用防火墙、数据加密，数据库安全等服务。

相关控制要求主要分为组织、人员、物理、技术控制四个方面。组织方面包括但不限于信息安全策略、访问控制、身份管理、认证管理、遵守信息安全政策、规则和标准等具体要求。人员方面包括但不限于信息安全意识、教育和培训等具体要求。物理方面包括但不限于物理和环境威胁防护、设备选址和保护、设备安全处理或重复利用等具体要求。技术方面包括但不限于源代码访问、技术漏洞管理、数据脱敏、时钟同步、网络隔离、应用程序安全需求、安全编码、变更管理、审计或测试期间的信息系统保护等具体要求。

同时，作为中国信息安全的基本制度，网络安全等级保护（简称等保）是公安部用于指导国内各组织单位进行网络安全建设的依据，目前已成为包括云计算在内的各行业广泛遵循的通用安全标准。

保护等级由低到高分为五级，划分依据是系统受破坏后的危害范围和严重程度。目前 1 级系统因影响小，基本不需要备案；5 级系统目前还不存在，只是理想状态。这里主要介绍 2 ～ 4 级。

- 2 级：一旦受到破坏，会对相关公民、法人和其他组织的合法权益造成严重损害，或者对社会秩序和公共利益造成危害，但不危害国家安全。需国家信息安全监管部门对该级信息系统的信息安全等保工作进行指导。
- 3 级：一旦受到破坏，会对相关公民、法人和其他组织的合法权益造成特别严重损害，或者会对社会秩序和公共利益造成严重危害，或者对国家安全造成危害。需国家信息安全监管部门进行监督、检查。

- 4 级：一旦受到破坏，会对社会秩序和公共利益造成特别严重危害，或者对国家安全造成严重危害。需国家信息安全监管部门进行强制监督、检查。

通过等保 4 级测评，意味着云厂商能够提供安全、可靠、稳定的云平台和云服务。

此外，针对不断变化的云环境和服务，公有云应按照认证要求定期进行复审，以保障云的安全性一直处于业界领先水平。

数据空间的安全可信是全栈全链条的，在云部署场景下，云平台作为关键基础设施，是保证数据空间安全可信至关重要的一环。通过诸如 ISO/IEC 27001 认证和网络安全等级保护认证，并在此基础上提供安全可靠的云服务，可以为数据空间的安全可信构筑坚实的基础。

7.3 基于开放数字权利语言的可控策略实现

7.3.1 标准化、机器可执行的控制策略

数据交换合约中的一项重要内容就是双方约定的条款，例如数据如何使用、如何保存、如何处理等。传统的手段是通过具有法律效力的合同来跟踪，如果违反其中的约定，则通过法律手段来维护各方的权益。在数据流通中，因为数据具有无损复制、难以确权等特性，所以数据约束条款的执行存在很大的现实困难，依靠人去严格遵守的方式已被无数的法律案件证明难以让人信服。

为此，华为探索一种技术驱动的保障手段，将合约中用自然语言描述的条款约束转换成标准化的约束模型，再转换成机器可强制执行的语言，从而确保条款和约束的有效实施，具体设计如图 7-6 所示。

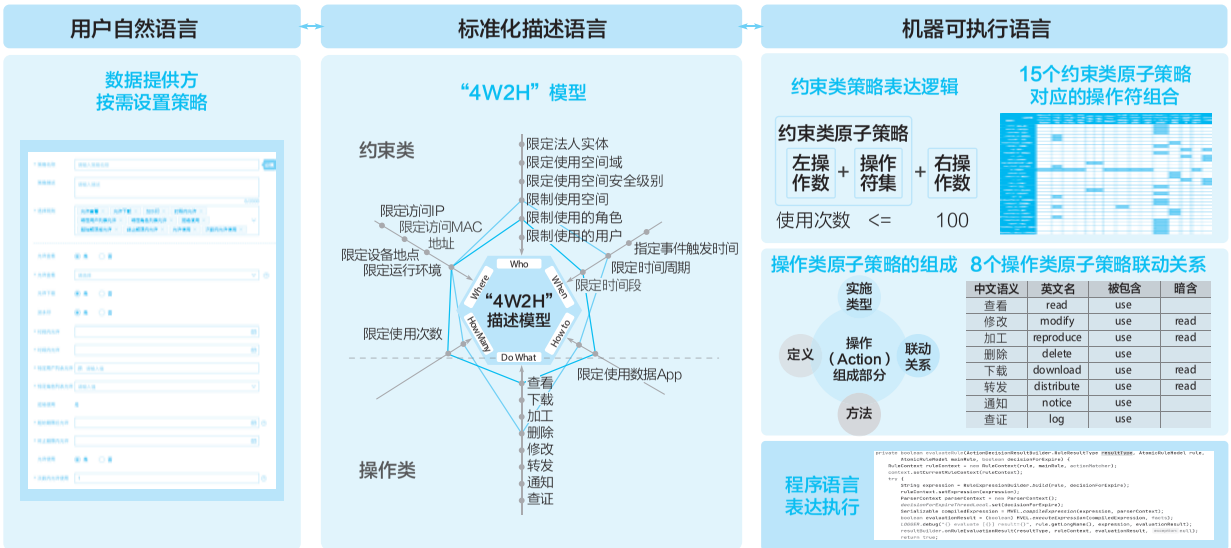


图 7-6　用户自然语言到标准化描述语言再到机器可执行语言

基于图 7-6 的逻辑设计，华为在实施数据空间的可控能力时，将数据空间中的数据策略从访问控制扩展到使用控制，然后通过使用控制实现跨边界的控制能力延伸。为此，华为基于用户合约、条款中自然语言所包含的策略要求，提出了“4W2H”模型来对策略进行标准化描述。在此基础之上，华为通过构建一套策略描述到机器可执行语言的映射机制，实现了策略在机器执行层面的充分表达与准确执行。从机器可执行的角度考虑，华为将策略分为约束类和操作类，并通过原子策略的组合实现对业务控制诉求的精准表述与执行。

1. 标准化描述语言模型

在“4W2H”模型的基础上，使用控制描述模型支持从可控对象、交换阶段两个方面继续延伸扩展，从而提供面向不同对象、阶段的分层控制能力，如图 7-7 所示。

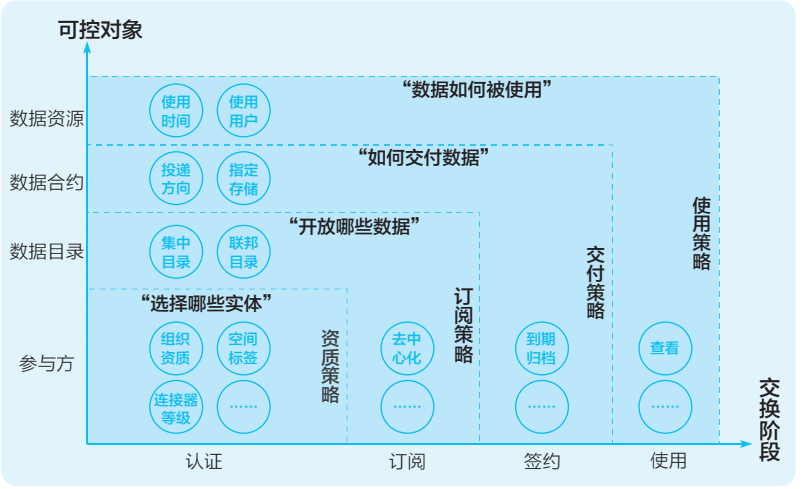


图 7-7　标准化的使用控制描述模型

使用控制描述模型包含四类策略，具体说明如下。

1）使用策略：约束消费方如何使用数据的控制策略，如限定时间、次数等。

2）交付策略：在数据合约中约定数据如何被传输和留存，如指定存储区域等。

3）订阅策略：决定数据商品如何被寻源的控制策略，如允许目录集中式开放等。

4）资质策略：决定哪些主体可以参与数据交换的控制策略，如限定企业信用等级为 3A 等。

2. 面向最终用户的策略项

华为在规划实现使用控制策略的过程中，充分结合实际业务场景的需求，考虑到数据策略覆盖数据交换全过程，将策略约束从消费方扩展到供需双方。32 个原子控制策略项如图 7-8 所示。

目前华为已累计设计并落地了 32 个原子控制策略项，满足了不同业务场景的应用需求。在使用的过程中，用户可以基于实际管控需求，灵活选择和组合原子控制策略项。

3. 约束类策略

约束类策略主要用于量化描述用户对数据的操作，比如“允许执行 ×× 操作 100 次”，图 7-9 展示了约束类策略的表达逻辑。

针对约束类原子策略，每一个原子策略的约束条件都可以通过“左操作数 + 操作符集”的格式来表达，再通过“右操作数”对约束条件进行量化。

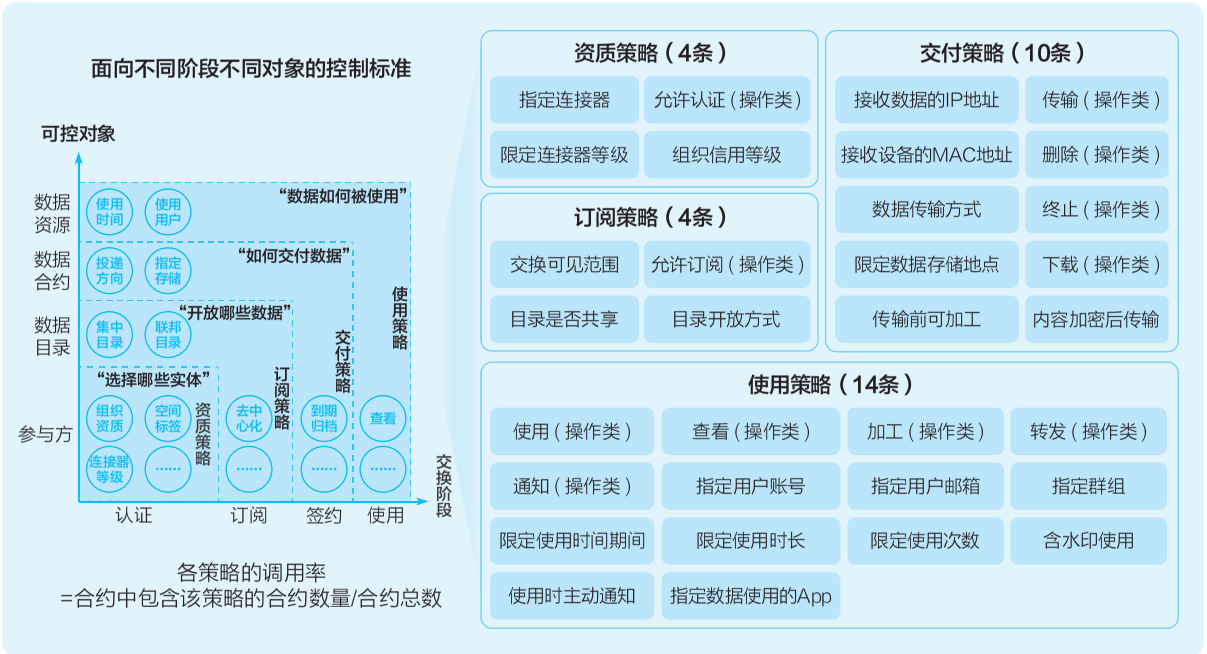


图 7-8　32 个原子控制策略项

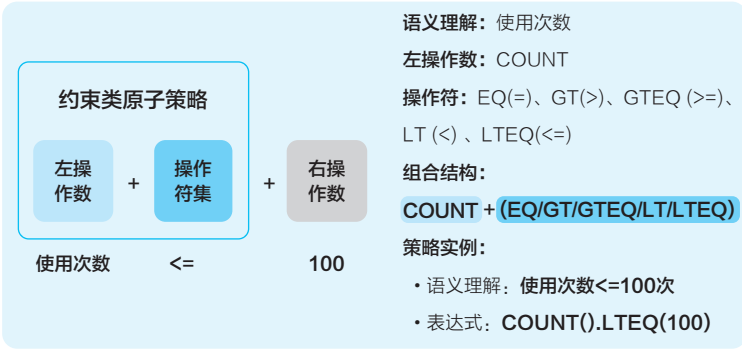
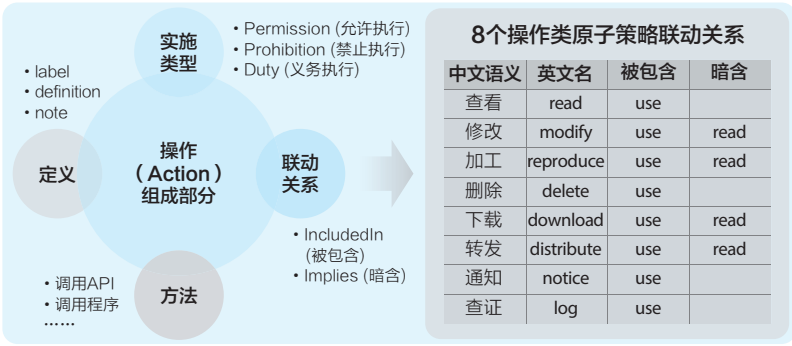


图 7-9　约束类策略的表达逻辑

4. 操作类策略

操作类策略主要用于描述用户可以执行哪些基本的、不可再分的操作，例如“允许查看”。每一个操作类策略包括四个核心组成部分：策略定义、策略实施类型、策略间的联动关系以及策略的执行方法。基于对数据的基础操作，华为一共定义了 8 个操作类原子策略，用于控制具体的操作行为。操作类原子策略涵盖了数据的查看、修改、加工、删除等基本操作，每个策略都旨在满足特定的安全和业务需求，如图 7-10 所示。



中文语义	英文名	被包含	暗含
查看	read	use	
修改	modify	use	read
加工	reproduce	use	read
删除	delete	use	
下载	download	use	read
转发	distribute	use	read
通知	notice	use	
查证	log	use	

图 7-10　操作类原子策略

图 7-10 中这 8 个操作类原子策略，通过定义描述操作的含义；实施类型包括允许执行、禁止执行，以及义务执行三类；操作之间的联动关系包括被包含、暗含；方法包括调用 API、调用程序等。以“转发”操作为例，华为的操作描述规则如下。

- 实施类型：允许执行。
- 定义：允许数据消费方附加使用策略后转发数据给第三方。
- 联动关系：暗含 read 操作，被 use 操作包含。
- 方法：{com.huawei.cabg.ids.handler.PolicyPxpComponent.distribute}。

7.3.2 控制策略的实现机制：定义、执行与决策

使用控制策略的实现机制是通过对策略进行统一定义并分发给所有的数据连接器，提供方基于原子策略项生成具体的策略实例并关联到具体的数据内容。在数据传输过程中，策略跟随数据内容一起传递给下游的消费方，消费方解析策略并进行判断决策，通过之后再允许用户执行对应的操作。

1. 使用控制策略分发与使用流程

关于使用控制策略的实现，华为参考并扩展了 ODRL（Open Digital Rights Language），这是一个由 W3C 制定的开放标准，用于表示数字内容的权限和义务。ODRL 提供了一种灵活且可互操作的方式来表达数字内容的使用和约束要求，支持在 Web 上共享的数字媒体的透明和创新使用。ODRL 策略用于表示对某一资产（如文件、图片、音乐等）允许或禁止的行动，以及利益相关方应履行的义务。

虽然 ODRL 为我们描述了可执行策略，提供了基本规范，但我们仍需要解决自然语言与机器可解析和可执行语言之间的转换问题。如自然语言指定“数据可以使用 14 天，并在到期后删除”，如何将此业务规范级别的策略转化为与 ODRL 对等的可执行策略。在实际业务场景中，很多时候，策略的初始定义来源是企业的业务系统，它充当了 PAP 的角色（如华为业务电子流中数据提取的相关规则）。业务规范级别策略与实施级别策略的转换能力，是数据使用控制策略竞争力的关键来源。使用控制策略处理机制如图 7-11 所示。

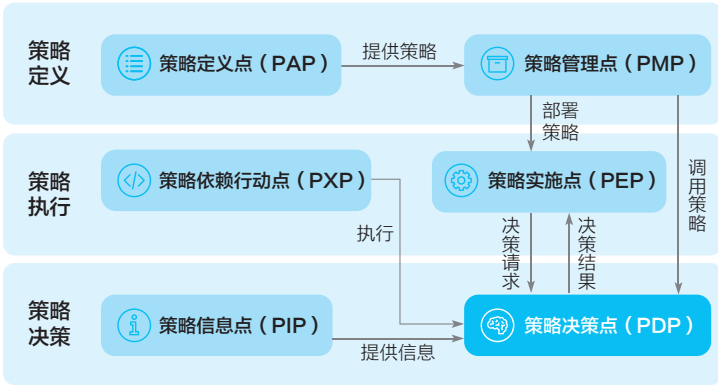


图 7-11　使用控制策略处理机制

数据使用控制策略的主要执行步骤如下。

1）PEP 拦截数据流。

2）PEP 将数据流转换为决策请求并发送给 PDP。

3）PDP 启动策略评估并调用 PIP 以检索其他信息。

4）PIP 将请求的信息提供给 PDP。

5）PDP 在 PXP 触发附加操作（例如发送验证码）。

6）PXP 向 PDP 确认操作成功。

7）PDP 向 PEP 发送决策结果。

8）PEP 对拦截的数据流执行决策。

2. 策略的三种决策方式

在实际应用过程中，基于提供方和消费方数据交换所处的业务场景不同，对策略的决策处理机制也不一样。对此，华为设计了提供方决策、消费方决策、第三方决策三种策略决策方式，以满足不同场景下的业务需求。三种策略决策方式如图 7-12 所示。

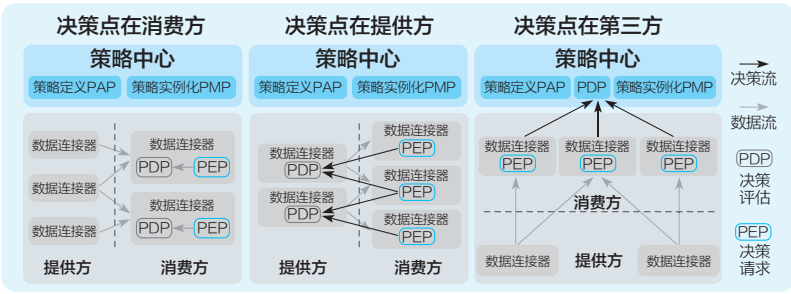


图 7-12　三种策略决策方式

图 7-18 直观地展示了消费方、提供方和第三方决策的基本概念与相互关系。接下来将深入探讨这些决策类型的特点、影响因素及其在实际应用中的具体表现。

（1）消费方决策

如果交换双方存在较高的信任基础，策略的决策能力可以下放到消费方数据连接器进行判断和决策，即使消费方数据连接器

与提供方数据连接器断开，仍然可以执行策略决策，正常使用数据。这样带来的好处就是决策的效率很高，同时因为提供方没有参与决策的过程，所以需要有事后通知和其他机制确保决策的合理性。

（2）提供方决策

如果提供方对数据的管控诉求较高，策略的决策能力可以要求必须经过提供方进行决策，消费方将执行信息传递给提供方进行决策，由提供方决策后下发执行指令，这样提供方可以随时控制消费方的数据消费行为，实现更严格的管控。提供方决策对于消费方的控制力度最强，不过同步会为提供方带来额外的决策工作量，同时由于每次决策都要提交到提供方，因此需要保持提供方数据连接器和消费方数据连接器之间的稳定连接，对模块间通信技术的要求更高。

（3）第三方决策

对于存在大量数据连接器的情况，决策活动有专门的运营组织负责制定统一规则，供需双方可以将策略决策都上报到第三方的决策机构，统一决策后再下发执行指令。第三方决策提升了决策的中立性，但需要有配套的决策机构与平台能力。

7.3.3　面向业务场景的服务化与模板化能力

可控能力在不同的业务场景下可以基于业务管控诉求进行灵活组合，用户在使用过程中通常基于数据密级、隐私标签、业务标签等对策略项进行组合使用，这样就逐步形成了不同场景下的推荐策略组合。在构建丰富原子策略的基础上，华为对原子策略进行分

类分级管理，建立数据治理规则与策略的动态配置关系，形成了标准化模板，如图 7-13 所示。

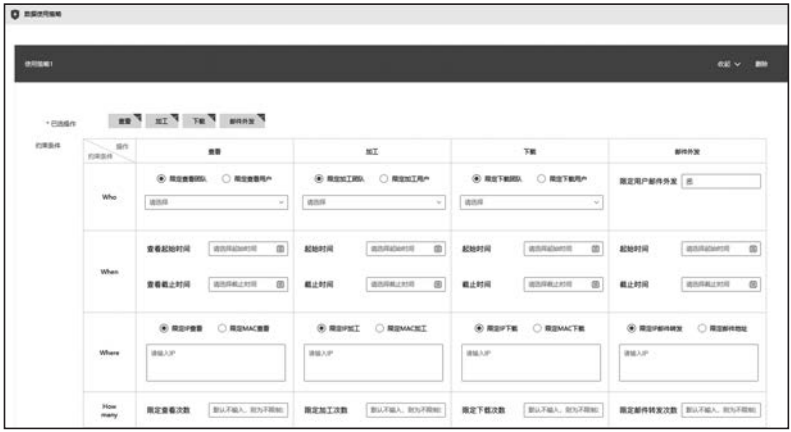


图 7-13　模板化策略应用样例

首先将策略原子化，从而实现与 ODRL 等标准的兼容转换，对策略分类分级以适用不同对象和交换的不同阶段。其次，进一步将原子策略模板化，面向用户屏蔽原子策略的复杂配置，基于原子策略兼容治理规则形成标准模板，对齐业务使用场景。再次，用户选择数据使用策略时可根据资产标签自动匹配适用模板，数据根据资产标签自动匹配合适的策略模板。数据资产控制策略与模板动态联动，标签变化与原子策略联动。最后，基于策略模板，可以沉淀用户使用策略时的经验，降低使用策略的复杂度。

7.4　基于区块链实现数据交换的可证追溯与清算

7.4.1　实现数据合约履行的透明可证

在现实的数据交换过程中，对各参与方而言，需要考虑数据

合约履行的透明性、可追溯性以及合规性，这就要求数据提供方能查证追溯、数据消费方能自证清白、清算运营方能清算监管等。在构建可证体系的过程中，华为尝试沿着数据交换端到端流程进行记录、汇聚和分析，构建过程记录、上报备案、第三方监管、存证防篡改等能力，从而进一步面向不同类型的用户实现查证追溯、自证清白与清算运营等价值目标，具体设计如图 7-14 所示。

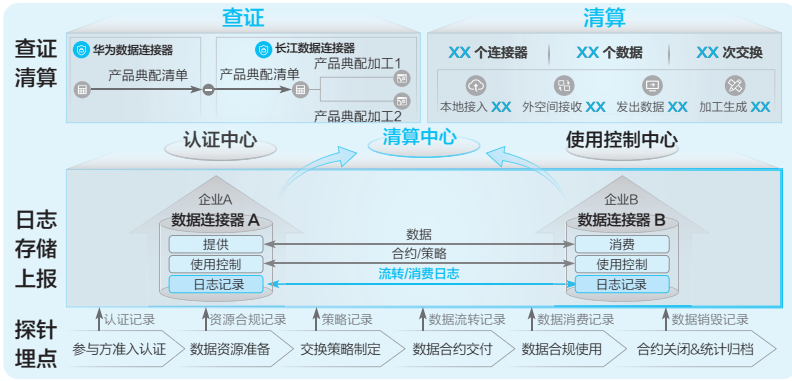


图 7-14　查证与清算的处理逻辑设计

在工程实现过程中，华为基于自身软硬件平台能力，构建了 EDS 的可证能力框架：通过数字探针实现端到端的行为记录，将记录的数据保存在各方的数据连接器中，同时上报给清算中心，最后面向不同的业务用户提供查证与清算的服务能力。另外，还采用区块链、数字探针等技术，实现了技术的增强保障与自动处理。

7.4.2　基于数字探针端到端的实时记录

追溯与清算的基础数据依赖于业务流上的数据探针，通过探针可采集全流程的过程记录，包括元数据的采集、内容的采集、日志数据的采集等。将这些数据采集、存证、关联，即可实现数据全

生命周期链路的查证，知道数据的完整流向。同时，也可以通过采集数据使用策略的执行情况，实现数据消费行为的可视化和量化。

通过在数据流通过程中插入探针，可以实现数据的全面监控和追踪。这些探针可以记录数据的来源、去向、传输路径等信息，从而实现数据流通过程的透明可视。数字探针技术通过对数据传输和处理过程的监控与分析，为数据的合规性和安全性提供了保障。通过数字探针技术，可以实时监控数据的使用情况，发现和处理可能的违规行为，从而为消费方自证清白提供手段。

7.4.3 探针数据上链防篡改

区块链技术以其去中心化的分布式账本特性，为确保数据的安全性和真实性提供了坚实基础。通过将数据分布式存储于网络中的各个节点，不仅实现了数据的去中心化，还实现了分布式管理。这种分布式存储机制显著提升了数据的安全性，即使某个节点遭受攻击或发生故障，数据的完整性和可用性也不会受到影响。

区块链的去中心化和不可篡改特性，为数据流通提供了一个可信的存证平台。数据空间结合区块链网络，实现数据的传输和存储，并允许完整地追溯数据的来源、流向和用途。这种透明度为数据提供方进行查证和追溯提供了强有力的支持，确保了数据流通的透明度和可追溯性，还为消费方自证清白等价值诉求提供了技术支持。

综上所述，区块链技术通过其独特的分布式账本、不可篡改的数据记录等特性，为数据空间提供了更多的安全保障，为数据的安全存储、可信流通和有效管理提供了全面的解决方案。

7.4.4　标准化的查证与清算服务

查证服务的目的是确保透明性、增强信任，让参与方了解数据的来源、流向和处理过程。清算服务的目的是衡量数据交换的价值，评估各参与方的贡献，推动数据流通形成互利互惠的正向循环。基于这两个业务诉求，下面解释华为对查证服务与清算服务的详细设计逻辑。

1. 查证服务的设计

查证服务主要面向提供方、消费方和运营方：通过查证，提供方证明数据的合法性和准确性；消费方验证数据的可靠性和适用性；运营方监控数据流通过程，确保安全性和合规性。查证内容应覆盖数据从产生到销毁的全流程，关注数据变化、操作和访问权限，以确保数据的完整性和一致性。通过全面查证，有利于构建健康可持续的数据流通环境。标准化的查证服务设计如图 7-15 所示。

2. 清算服务的设计

清算涉及提供方、消费方和清算方，旨在确认数据价值、收益和成本，并通过评估、记录和分配确保流通过程的公正性和透明性。清算内容包括由数据提供、接收和加工创造的新价值，要求详细记录和评估以确保准确性。标准化的清算服务设计如图 7-16 所示。

在清算服务中，首先需要确定清算主体和对象，为参与方和交换合约提供统一 ID 管理。其次，针对不同场景记录数据流通过程，并通过统一接口上报至清算中心。最后，基于上报记录进行统计核对，按主体维度归结与统计，形成清算账单，支撑价值贡献评估。

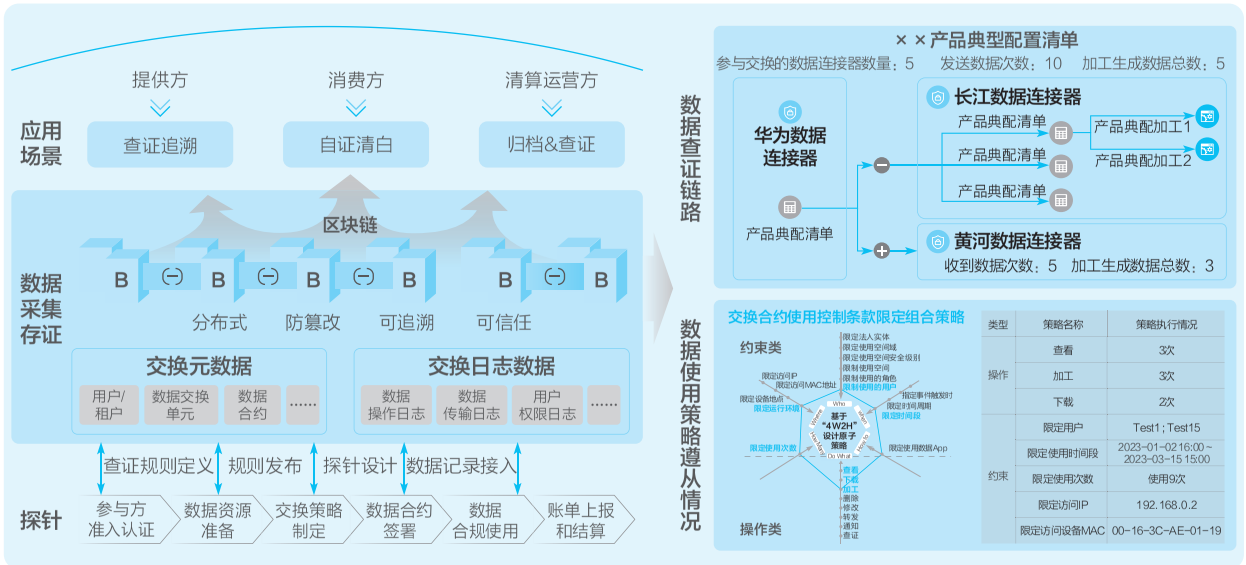


图 7-15　标准化的查证服务设计



图 7-16　标准化的清算服务设计

7.5 小结

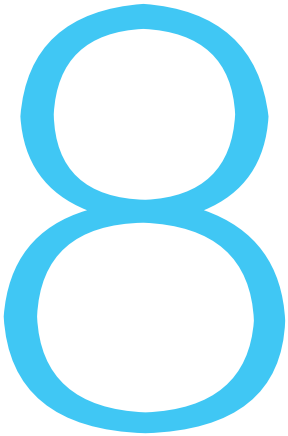
本章深入探讨了数据所有者在数据流通过程中面临的问题，以及如何通过技术手段来解决这些问题。数据所有者普遍面临“不敢（担心数据安全）、不愿（缺乏激励机制）、不能（技术障碍）”交换数据的问题，这与实物流通中的高价值、高保密性、易损物品的流通有着相似之处。华为认为，数据空间需要实现数据流通的“可信、可控、可证”。

首先，我们构建了一个信任框架来保障数据空间的可信性。该框架涵盖用户、组织、组件、环境资源等多个层级，实现认证、发证、监管等端到端环节的信任管理，从而消除数据所有者对安全性的顾虑。

其次，我们利用开放数字权利语言（ODRL）实现了数据流通的可控性。ODRL 提供了标准化、机器可执行的控制策略，以及面向不同阶段、对象的分层控制能力。通过设计并落地 32 条原子策略，满足了不同业务场景的应用需求。机器对策略的表达与执行，以及控制策略的实现机制，为数据交换提供了强有力的保障，解决了数据所有者的技术障碍问题。

最后，基于区块链技术，我们实现了数据交换的可证性。这包括实现数据合约履行的透明可证，基于数字探针端到端的实时记录，以及探针数据上链防篡改。此外，我们还提供了标准化的查证与清算服务，确保了数据交换的透明性、可追溯性以及合规性，消除了数据所有者对合规性的顾虑。

数据空间的“可信、可控、可证”不仅是数据空间管理的价值目标，也是构建数据空间能力的三个牵引方向。通过实现这三个目标，能够构建一个既安全又高效的数据流通环境。



第 8 章 CHAPTER

数据空间的部署与运营

在部署和运营数据空间的过程中，华为不仅关注数据空间的技术实现，还重视与之配套的管理措施建设。华为不断在各个场景中进行业务实践的目标是构建既能满足技术需求又符合管理规范的数据空间，同时积累的经验也能为其他企业提供参考。

本章将深入介绍华为在数据空间部署和运营中遇到的挑战，并提供一系列基于华为实践的指南和规范。根据华为的实践经验，新业务应用数据空间的首要步骤是明确使用主体、管理责任和使用方式。这一步至关重要，因为它为数据空间的顺利启动、数据的高效交换和有效运营奠定了基础。在数据空间的实施和运营阶段，必须制定符合业务自身特点的流程规则。华为的指南和规范覆盖了数据空间部署和运营的全过程，旨在协助企业顺利部署和运营符合自身需求的数据空间。

8.1 启动数据空间的 3 个关键问题

数据空间作为一种创新的数据管理理念，其成功实施需要企业相应管理规则的支撑。华为在推进数据空间的过程中，首先与业务部门共同探讨并明确几个核心问题：谁是数据空间的使用主体（谁来用）、谁负责运营数据空间（谁来管），以及如何将数据空间整合到现有业务中（怎么用）。为了确保数据空间的顺利实施，业务部门必须先确定这些问题的答案，并据此制定相应的管理规则。基于业务管理的实际需求，华为开展了一系列实践探索。

8.1.1 谁来用：使用场景的识别

在探讨数据空间的使用主体时，华为结合业务需求从两个关键维度进行考量：数据交换的价值高低以及交换双方之间的信任程度。基于实际应用实践，华为提出了两种场景建议，分别用“Must”和“Should”来指导识别适用数据空间的场景，如图 8-1 所示。

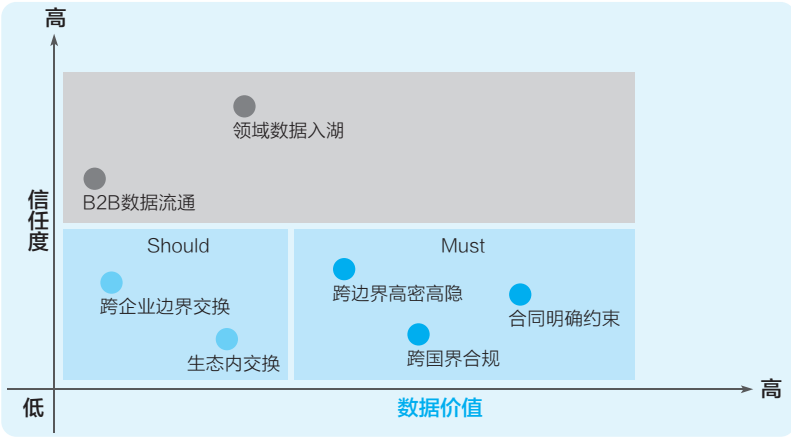


图 8-1　谁来用数据空间的识别规则

图 8-1 的右下部分为“ Must”场景，针对的是那些数据价值高且双方信任度不高的业务。在这种情况下，数据交换对于业务至关重要，因此需要确保数据的安全和合规性。华为建议采用数据空间来加强数据管理和保护。根据华为企业实践，“ Must”场景的识别可参考如下具体规则。

1）数据交换需要跨主体边界时。由于主体间的信任度较低或一般，因此数据的安全性和保密性成为主要关注点，需要通过技术手段确保数据交换的安全性。例如，华为与外部投资机构进行尽职调查数据交换、财经部门与外部审计事务所交换审计底稿等场景就适用这一规则。

2）交换的数据涉及高密、高隐、高价值的信息。企业对这些敏感数据具有很强的保护需求，因此在交换过程中需要增强防护措施。例如，企业内部采购部门高密合同文档的借阅、从业务系统中提取使用敏感数据等场景就适用这一规则。

3）数据交换主体对数据的使用有明确约定。合约要求数据的使用必须符合约定，并且能够进行有效的管控和监督，以确保合规性。例如，华为与供应商签署共享协议交换库存数据、华为与智选车伙伴签署合作合同交换车型研发资料等场景就适用这一规则。

图 8-1 中左下部分为“Should”场景，该场景适用于数据价值相对较低或双方信任度有待提升的情况。这类场景下，数据交换虽然重要，但可能不需要最严格的控制措施。华为建议采取更为灵活的管理策略，同时鼓励交换双方建立信任，逐步提高数据交换的安全性和效率，“Should”场景的识别可参考如下具体规则。

1）数据交换具有明确的主体和边界，且主体间的信任度较高，但需要通过更加透明的交换机制来增强互信，以共同应对外部

风险。例如，华为与鲲鹏 / 昇腾整机战略合作伙伴间的数据交换场景就适用这一规则。

2）数据提供方希望维护自身数据权益，保障数据的稀缺性和价值。这要求在数据交换过程中实现有效的管控，过程需要透明可证，从而确保数据交换的公正性和合理性，并增强各方信任。例如，华为与科研机构互相交换最新的科研技术成果的场景就适用这一规则。

8.1.2 谁来管：明确运营主体及职责

在解答“谁来管”的问题时，华为通过调研和访谈众多数据空间业务责任方，达成了一个共识：数据空间应由一个明确的责任主体全面负责运营管理。该责任主体将根据业务需求，规划和执行数据空间的运营工作。华为依据在数据空间运营方面的经验，建议数据空间的运营管理遵循以下规则。

1）申请开通的数据空间必须指定一个明确的运营方，该运营方应由公司正式任命的业务部门主管或项目经理担任，确保数据空间的运营管理具有权威性和明确的责任归属。通常，数据空间的管理运营职责由申请方或其下属组织承担。例如，鲲鹏 / 昇腾生态数据空间由计算产品线下属的硬件生态发展部负责运营管理，而财经外审数据空间则由公司账务管理部负责。

2）数据空间运营方可授权数据专业组织履行数据内容治理职责，确保交换数据来源于已认证的可信数据源，保障数据的质量和可靠性。在华为的实践中，数据专业组织也在数据空间中承担管理职责，如业务领域的数据管家负责支持业务在数据空间中接入和管理公司数据湖中已认证过的数据源。

3）数据空间是管理概念，参与方通过数据连接器进行数据交换。运营方的职责范围也应基于该数据空间认证和部署的数据连接器来确

定。例如，华为鲲鹏 / 昇腾生态数据空间包含华为和整机伙伴共 20 多个数据连接器。作为运营方，华为的运营管理应覆盖所有这些连接器。

8.1.3　怎么用：被调用并嵌入现有业务流程

在探讨“怎么用”数据空间的问题时，华为意识到，作为一项公共能力，数据空间需要与企业现有的业务流程和平台有效协同。为此，华为制定了一套数据可控交换流程，并将其作为公共能力流程发布，旨在将数据空间无缝集成到日常业务流程中，从而实现业务对数据空间能力的灵活、高效调用，具体流程如图 8-2 所示。

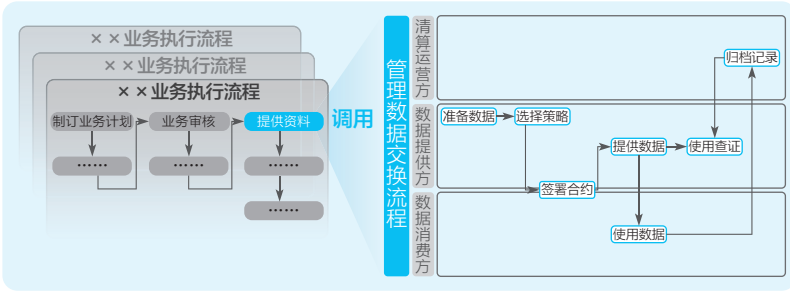


图 8-2　数据空间能力被调用嵌入现有业务流程

数据空间的设计理念旨在使其能够平滑且自然地融入现有业务流程中。在数据空间中，可控交换流程作为一种基础服务，可供其他业务流程调用。在业务流程调用可控交换流程之前，业务按照自身的流程和管理要求执行业务流程活动。当涉及需要进行数据共享交换的环节时，业务可以按需调用可控交换流程，执行数据共享交换的相关动作。与流程活动相对应的是，数据空间提供的功能服务也可以灵活地被其他业务系统调用，使用户能够在现有平台上无缝地使用数据空间服务，同时不降低现有业务流程的效率。在华为的应用实践中，如财经外审流程，华为就在对外提供审计资料的业

务活动中调用了可控交换流程；在法务对外司法协查流程中，华为在向律师提供协查资料的业务活动中也调用了可控交换流程。

8.2 数据空间装载与运营

数据空间的业务装载和运营管理是实现数据空间应用的基础。通过实际业务实践，华为将数据空间业务分为三个主要阶段：开通、数据交换、运营管理。

在数据空间的开通阶段，华为确立了明确的申请流程和资源分配机制，确保数据空间的快速部署和初始化。

在数据交换阶段，华为制定了详细的规则指南，指导数据的接入、发送和消费。

在运营管理阶段，华为明确了数据空间的运营机制、组织及运营内容。用户基于自身所涉及的不同阶段遵循相应的规则和规范，以确保数据空间的高效运作和业务目标的顺利实现。

本节将详细探讨华为在数据空间的业务装载与运营过程中所遵循的规则和规范。

8.2.1 如何开通数据空间

数据空间应该有一个明确的责任主体，由该责任主体向数据空间的平台服务提供方申请开通数据空间，平台服务提供方将按需分配数据空间的软硬件资源。开通数据空间后，数据空间责任主体应该进一步分配和初始化数据连接器，对组织与人员进行注册认证，以便进一步装载具体业务。华为基于业务实践制定了数据空间业务开通的标准流程，如图 8-3 所示。

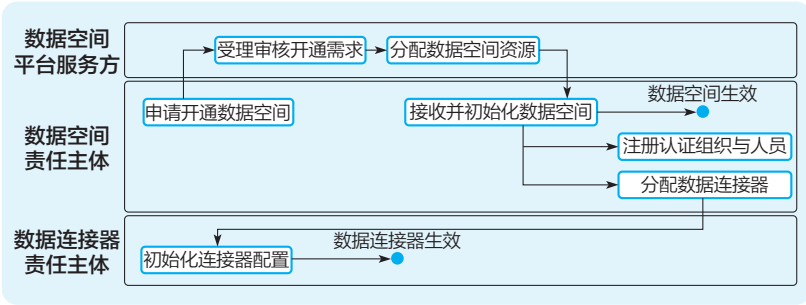


图 8-3　数据空间业务开通流程

下面将基于华为实际应用实践，对数据空间开通业务流程中的关键活动和要点做详细介绍。

1. 申请开通数据空间

为了确保数据空间的高效运作和数据交换的安全性，对申请开通数据空间的主体进行责任主体认证是必要的第一步。这一过程的目标是确保开通数据空间的责任主体不仅可信，还具备相应的管理能力和资源配置。华为针对开通数据空间设计的关键业务活动如下。

（1）主体注册与审核

开通责任主体需要在平台上进行注册，并提供必要的组织信息，包括但不限于机构名称、联系方式和法律身份等。此外，责任主体还需提交相关的法律文件、资质证明和管理团队信息等，以证明其合法性和管理能力。这些资料将由平台或第三方审核机构进行严格审核。审核通过后，责任主体将获得认证标志，这是其具备开通数据空间资格的明确证明。在华为的应用实践中，由业务和平台执行审核动作。例如，在鲲鹏 / 昇腾数据空间中，业务负责提供伙伴的组织信息并进行第一轮审核；伙伴导入平台后，由公司认证平

台校验伙伴信息的真实性和完整性；通过审核后发放数据空间责任主体的认证证书。

（2）需求评估与分配资源

根据数据空间的特性和需求，责任主体需要分析所需的资源类型和数量，例如选择合适的连接器类型、选择数据的存储容量等。同时，分配的数据空间资源也需要指定责任人，以确保对数据空间的有效管理和责任的明确归属。在华为的应用实践中，主要由平台针对实际业务诉求按需分配合适的资源。例如，投资尽调业务对数据空间有很强的隔离部署需求，平台按需在私有环境中部署数据空间，实现管控增强。

（3）空间业务初始化

在这一阶段，责任主体需要配置数据空间的使用控制策略，确保在数据空间内进行数据交换所采用策略的一致性。例如，外审业务对使用控制策略有很强的精细化管理要求，平台在常用的策略基础上，额外适配了限定 IP、限定 MAC 地址等策略。此外，责任主体还需要设置统计清算规则，以便于对数据交换活动进行监控和审计。例如，华为数字能源数据空间需要基于分公司和集团总部对数据交换进行统计记账的需求输出清算规则。

2. 启用数据连接器

为确保数据连接器能够高效且安全地满足业务需求，在启用前必须根据业务需求完成一系列通用规则的设置，主要包括以下要点。

（1）数据连接器的部署设置

根据业务的实际需求，确定每个数据连接器的具体部署要求。

这涉及选择适合的部署模式，如公有云、私有云或硬件一体机等。每种模式都有其优势和限制，用户应该基于业务实际需求选择正确的部署模式。例如，在华为内部组织间的数据共享实践中，考虑到快速部署和成本效益，通常选择公有云部署；而企业的投资尽调业务由于保密要求较高，因此更倾向于私有云部署。

（2）数据连接器的功能配置

启用数据连接器前，必须收集并分析业务需求，以确定所需的功能支持。这包括考虑将要处理的数据类型（结构化数据、非结构化数据或两者的组合）、数据存储方式（如数据库、数据湖或分布式文件系统），以及数据交换方式（实时数据流、批量数据传输或事件驱动的数据交换）。在系统配置时，还需考虑业务流程的多个方面，如审批流程、数据访问权限、数据加密和备份策略等。这些配置旨在确保数据连接器的操作既符合组织的业务和法规要求，又支持业务的灵活性和扩展性。在华为的应用实践中，根据业务需求，我们对数据连接器配置进行了分类，以适应结构化数据交换和非结构化数据交换的不同需求。例如，为投资尽调场景配置了少量文件的点对点快速传递通道；为财经外审场景配置了海量文件的多人协同传递通道；为与合作伙伴共享物流信息等场景配置了结构化数据的即时查询分析通道。这些典型的通道配置使华为能够迅速匹配业务所需的主要功能，提高业务的易用性。

8.2.2　如何在数据空间中交换数据

在完成数据空间和数据连接器的开通及初始化之后，具体的业务用户就可以在数据空间中按需进行数据交换。在执行交换之前，各参与方应该先梳理业务现状，确定数据交换的基本业务规则，如是否建立团队、合约内容、数据内容、发送方式、消费 App

等，确定规则之后，可以按照分配的角色执行发送与消费数据的业务动作，具体流程如图 8-4 所示。

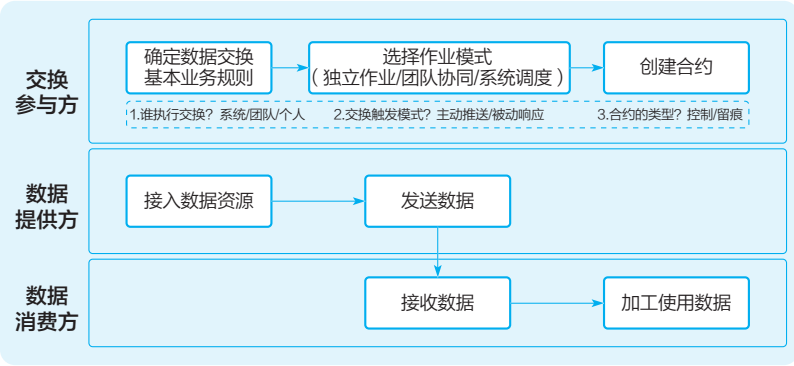


图 8-4　数据交换标准业务流程

基于图 8-4 提供的标准业务流程，下面针对数据交换业务适配、数据接入与发送、数据接收与消费三个关键活动做详细介绍。

1. 数据交换业务适配

在数据交换的业务适配过程中，业务需要确定交换双方的具体作业模式、合约种类以及使用策略等关键内容。

（1）数据交换作业模式

在数据空间中，数据交换的作业模式应根据实际业务需求来确定，以适应不同的业务场景。具体而言，可选择以下三种模式。

- 独立作业模式：适用于个人代表组织处理少量文件的场景，例如企业投资尽调资料的交换。此模式允许个人独立完成数据接入和发送的全过程，其优点在于流程简洁、操作高效，但缺点是对于复杂的交换场景，扩展性和兼容性可能不足。

- 团队协同作业模式：适用于多人协作处理大量文件的场景，如在鲲鹏 / 昇腾业务中华为与众多伙伴进行能力资质举证与评估。在此模式下，数据接入与发送流程由团队成员分工合作完成，优点包括强大的大批量文件处理能力、清晰的权限和职责分离，以及有效的数据管控，但这也要求业务分工和数据管理必须更为精细、高效。
- 系统调度作业模式：适用于业务系统具备完善的前置处理动作，数据空间仅作为后台服务被调用的场景，例如采购合同文档借阅，电子流系统调用数据空间执行最终的可控投递，或供应链结构化物流状态信息共享。此模式提高了数据交换过程的自动化水平，有效减少了人工操作，但同时也要求业务规则必须明确，且业务系统需足够成熟。

（2）选择数据交换的合约种类

在数据空间中，合约定义了数据交换的规则和条件，生成适配业务的合约是确保数据交换顺利进行的关键。华为经过实践，在数据空间中提供两种合约生成模式：一种是点对点合约，由数据提供方主动发起，适用于双方直接交换数据的场景，如企业投资尽调场景中华为与外部具体投资机构的数据交换。在这种模式下，数据提供方制定合约条款，而数据消费方只需接受合约即可开始履行。另一种是由消费方发起订阅的合约，适用于消费方需求不明确的场景，提供方上架数据到市场后，由消费方根据实际需求发起订阅并协商生成合约。

此外，数据空间还提供了两种不同的合约内容范围选项：一种是标准合约，适用于数据交换内容固定且可预测的场景，例如采购合同文档借阅，每次借阅事件都会生成一个内容明确的标准合约。

另一种是框架合约，适用于数据交换内容可能变化且需要灵活性的情况，如智选车生态中，框架合约定义了数据交换的目录范围，并允许消费方在该范围内按需持续获取数据。

（3）选择数据交换的使用策略

在数据交换使用策略方面，数据空间提供了丰富的策略选项，以适应多样化的业务控制需求。这些策略覆盖了数据保密级别、个人隐私保护、业务类型等多个关键领域，目前有 4 大类 32 种策略可供选择。企业可以根据自身需求，挑选合适的策略来构建策略集。例如，对于投资尽调这种高保密级别的数据交换，可采取更为严格的使用控制策略，如限定消费方的访问 IP 和 MAC 地址、限定仅具体用户可以查看等。

根据不同数据的特性和业务的具体需求，企业可以定制化地选择和组合不同的策略，形成策略集。通过业务的持续应用和优化，可以为不同的业务场景创建差异化的策略模板，以便于快速复用和部署，目前华为已经形成如财经外审、投资尽调、生态协同等策略模板。这样的策略选择机制也增强了业务操作的灵活性和效率。

2. 数据接入与发送

为了保证数据的交换质量，华为在实践之后提出可供交换的数据需要选择可信的数据来源，按需进行必要的处理之后再发送给消费方。

（1）数据的接入

不同公司对数据治理的要求不同，建议结合现有的数据治理

平台，选择可信的数据来源进行接入，如公司数据湖的资产，从而保证接入数据的质量。

（2）数据的预处理

预处理是数据接入后的重要环节。原始数据在接入之后，往往需要经过一系列的预处理操作，以满足业务需求和提高数据的可用性。这包括但不限于为数据打上交换标签，进行数据切片、清洗和格式化等操作。通过这些预处理步骤，能够实现按需交付，确保消费方接收到的数据是它们所需要的，并且是符合预期格式和质量的。在应用实践中，当涉及结构化数据的交换（如供应链库存数据的交换）时，华为需要针对不同的伙伴数据打上对应的业务标签进行区分，同时对数据内容进行行列切片，确保数据交换内容的合理合规。

（3）数据的发送

不同类型数据的发送也会存在区别，如投资尽调等文件的发送通常是单一接收主体、一次性发送。鲲鹏生态等结构化数据的发送需要定期多次发送。在发送时，不同数据需要选择不同的发送规则。

3. 数据接收与消费

华为通过众多的实践发现，用户要求数据空间具有数据管理功能，用于支撑用户进行复杂的数据管理活动，包括数据分类、归档、检索和安全保护，确保数据在整个生命周期中的完整性和可用性。关于数据接收与消费的具体要点如下。

（1）数据的接收与归档管理

主体间如果存在大量的高频数据交换，就要求业务不仅要能

够接收数据，还要能够对其进行有效的分类和管理。为此，数据空间提供了一套完善的接收与归档管理机制，旨在满足消费方的使用和管理需求，确保数据的有序存储和便捷检索。例如，指定存储资源，消费方可以在策略允许的情况下将接收的数据存储到指定内容中，并设置存储、转换的规则；供应链物流数据可以设置对结构化数据进行定期存储、处理的接收规则；外审的业务可以基于审计主体与审计科目进行数据目录编制，将数据放在对应的目录中进行精细化管理。

（2）数据的消费

数据消费存在两种主要场景：简单消费和复杂消费。简单消费通常涉及对单一文件对象的直接操作，如文档的查看和编辑。这种类型的数据消费对于用户而言直观且易于操作，通常不需要复杂的数据处理技能。在实际应用中，对于外审、尽调这类以非结构化文件交换为主的场景，华为数据空间内置了文件处理 App，可以满足对文档、图片等常用文件的查看和编辑需求，App 与策略联动可实现防截屏、数字水印等安全防护。复杂消费则涉及对多个数据对象的综合处理，如执行 ETL（提取、转换、加载）操作、生成数据报告等。这类消费要求用户具备一定的数据处理能力，并且需要使用专业的工具和应用程序。在华为与鲲鹏 / 昇腾伙伴交换物流数据的案例中，华为接入并认证了数据加工分析工具永洪 BI，实现多方数据的汇聚、关联与报表制作，满足了多方进行数据联合分析的业务需求。

为了适应这两种不同的数据消费场景，数据空间提供了相应的流程和工具支持。用户可以根据自己的具体需求，选择合适的数据消费流程和工具，从而实现数据的有效利用和价值最大化。

8.2.3　如何运营数据空间

在数据空间成功开通并生效之后，其运营活动的管理显得尤为关键，这需要由专业的管理方或运营团队来负责。华为根据自身在数据空间领域的业务实践，提出了一套全面的管理运营体系构建建议。这套建议包括了运营组织的任命、运营机制的建立以及运营内容的制定等核心要素。数据空间运营体系应该围绕具体的业务场景来构建，并据业务变化进行不断调整和优化，通过运营体系提升数据空间的运营效率，同时保障数据交换的安全性和合规性。

1. 任命运营组织

在数据空间中，数据空间所有者（Owner）不仅是数据空间的总责任人，也是数据空间运营的主要参与方。数据空间所有者负责整个数据空间的运营，包括但不限于数据连接器的申请、审批，数据交换的查证追溯，以及数据空间的日常运营管理。它的职责是确保数据空间的运作符合公司的政策和法规要求，同时保障数据的安全性和合规性。在华为的实践中，由业务领域主导构建各自的数据空间，不同的业务领域、部门或项目可能会基于特定的业务需求和跨主体、跨边界的数据管理要求，申请建立专属的数据空间。针对这种情况，公司将正式指定业务部门主管或项目经理担任数据空间所有者。这些被任命的所有者将同时承担运营方的职责，全面监督和管理数据空间内的活动，保障数据流通的效率与质量。例如，财经外审数据空间由负责外审业务的部门主管管理，而鲲鹏 / 昇腾生态数据空间则由华为硬件生态发展部门主导。这样的安排确保了数据空间的运营与业务活动和需求紧密对接。

数据空间的所有者通常会授权并批准管理员作为数据空间的实际管理执行者。这些管理员负责执行数据空间的具体运营管理任

务，包括但不限于对数据空间内各数据连接器交换、消费的数据和行为进行采集、分析与运营。管理员的职责是确保数据空间的日常运作顺畅，及时处理数据交换过程中出现的问题，并提供必要的技术支持和咨询服务。

2. 建立运营机制

数据空间运营重点服务于通过数据空间进行的跨主体、跨边界的数据流通，需要针对数据空间内各数据连接器交换、消费的数据和行为开展统计、清算、存证等运营活动。运营活动必须全面覆盖主权可控数据交换的整个流程，包括对数据连接器交换的数据、持有的数据以及加工过程中的数据进行详尽的统计和清算。同时，还需要确保所有行为均符合基于合约配置的可控策略，以及维护这些策略的一致性。

1）主权可控数据空间运营重点服务于通过数据空间进行的跨主体、跨边界的数据流通，对数据空间内各数据连接器交换、消费的数据和行为开展统计、清算、存证等运营活动。在华为应用实践中，针对数据空间的统计、清算和存证等活动，由数据空间业务运营方发布数据空间中交换过程的记录与统计清算报告。

2）运营应覆盖主权可控数据交换的完整流程，除了统计、清算、存证各数据连接器交换、持有和加工的数据，还应涵盖基于合约配置的可控策略的一致性及对应的行为。在华为实践中，数据提供方基于合约与策略的一致性进行异常分析和预警。

3）各数据空间统计、清算、存证的交换记录，在满足数据空间运营管理需求的基础上，同时应响应集团整体运营管理的要求，并在必要的时候向业务提供查证追溯和自证清白的支持。例如，公司数据管理部作为数据空间所涉流程的责任方，代表集团参与数据空间的运营管理。

3. 制定运营内容

数据空间的运营依托于数据连接器，并覆盖数据交换的整个流程。为了适应不同主体的需求，运营体系提供了一系列可组合的服务，如统计和清算等。运营责任主体需明确发布运营管理内容，确保数据交换遵循统一标准，包括参与者生命周期、合约与策略执行、交换数据资源、用户行为记录、异常与争议等关键信息。具体运营内容的制定要点可以参考如下规则。

（1）参与者生命周期运营

管理数据空间内各数据连接器对应的交换主体的基本信息和认证信息，为统计和清算提供基础。华为会管理各类参与主体的加入、认证和退出，定期更新信息，并由数据空间运营方负责连接器的认证、配置及策略定义。例如，华为会针对各类参与主体的加入、发证、退出等进行管理，定期发布各参与主体的最新信息，并将伙伴的加入和退出信息及时同步到数据空间的各参与方。

（2）合约与策略执行运营

管理数据空间内各交换合约及其控制策略的信息，包括合约名称、策略组合及审批流程。华为会定期分析合约的制定情况以及策略调用情况，优化配置，形成适用于不同场景的策略模板。例如，合约与策略的制定及调用情况会例行进行统计分析，针对不同使用场景下的策略反馈，调整和优化策略配置，形成不同场景下最合理的策略模板。

（3）交换数据资源运营

管理跨主体、跨边界交换数据的元数据信息，为清算提供数

据保障。华为会统计分析产业业务单元的数据交换记录，确保数据流通全景图的清晰。例如，数字能源产业业务单元对数据交换记录的统计分析有强烈需求，为了跟集团和其他产业开展合作，产业业务单元维度的管理运营者需要清楚地统计出数据流通的全景图，确保数据空间的所有交换明细记录都可以围绕管理要求进行分层统计分析。

（4）用户行为记录运营

记录数据交换过程及交换后消费数据的行为信息，为合约遵从提供服务支撑。华为会制定行为规范，定期统计分析用户行为，确保合规性。例如，司法协助场景需要对用户的行为进行强管控，确保所有用户行为合规。为此，华为针对所有行为记录制定了对应的行为规范，并基于规范定期进行统计分析，如按照人员、地域、时间等维度对操作行为进行统计和筛选。

（5）异常与争议的运营

清理、归档并公正处理数据空间内的交换异常问题及争议。华为会指定专人负责统计分析和汇报，同时协调处理异常和争议事件，并向数据空间责任方进行汇报。

8.3 小结

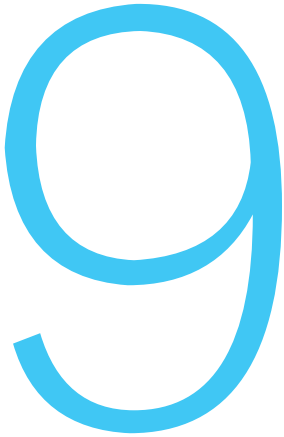
本章深入分析了企业在部署和运营数据空间时所面临的挑战，并基于华为自身的业务实践总结，提供了一套运营规则以供参考。

首先，我们指出了启动数据空间时必须回答的三个关键问题：谁来用、谁来管，以及怎么用。这些问题的答案将直接影响数据空

间的管理和使用效率。对于“谁来用”的问题，我们建议从数据价值和信任度两个维度来识别使用场景，并提出了 Must 和 Should 两种场景建议。在“谁来管”的问题上，我们强调了明确运营主体的重要性，并根据业务需求和管理颗粒度提供了管理规则的建议。至于“怎么用”，我们认为数据空间应嵌入现有业务流程中，作为企业业务流程中的一个能力流程。

其次，我们探讨了数据空间的业务装载与运营。数据空间的业务装载、运行、管理、运营是应用数据空间的前提。我们详细描述了数据空间的完整业务流程，在如何开通数据空间的部分，说明了由责任主体向平台服务提供方申请开通数据空间，并进行资源分配和初始化的过程。而在数据交换方面，我们强调了在执行数据交换前，各参与方需确定基本业务规则，并按分配的角色执行数据发送与消费。

最后，我们讨论了如何运营数据空间。运营机制的建立、运营组织的任命以及运营内容的描述是运营数据空间的关键。我们提出了主权可控数据空间运营能力的概念，强调了运营应覆盖完整的数据交换流程，并基于合约配置的可控策略进行行为管理。同时，指出了数据空间 Owner 的角色和责任，以及公共数据空间的运营管理。关键运营内容的采集、存储、统计、清算等是支撑数据空间运营的基础。



第 9 章 CHAPTER

数据空间的企业应用

华为作为一家全球化的企业，其业务分布在 170 多个国家和地区，需要与外部的客户、供应商、生态伙伴以及内部的产业、区域、职能组织进行大量的业务协同。这种复杂的全球协同模式使数据的高效连接和安全流通变得尤为重要。在全球化的背景下，数据不仅需要在不同国家和地区之间进行高效流通，还需要得到安全保障。这就要求华为具备强大的数据保护能力和风险管理能力。

在与外部的客户、供应商和生态伙伴进行业务协同时，华为需要确保数据的安全流通，防止数据泄露和滥用。同时，在与内部的产业、区域和职能组织进行协同时，华为也需要保证数据的准确性和完整性，以确保业务的顺利进行。通过与多方沟通和对多个领

域进行调研，我们收集到了广泛的数据流通（交换）需求，考虑到交换主体、交换方式、边界类型、管理方式等维度上的相似性，我们将这些需求分为四大类进行统一适配和探索。本章将对数据空间在企业中的主要应用场景与实践进行介绍。

9.1 企业应用数据空间的四类场景

华为在 2022 年正式启动数据空间产品设计和实际业务应用，并以鲲鹏 / 昇腾生态为业务试点进行技术与业务验证。在业务实践过程中，根据用户反馈对产品能力进行打磨和完善，以满足不断变化的业务需求。华为根据四条主要的需求线索进行了场景拓展，旨在让更多的业务领域了解并应用数据空间技术，以解决不同场景下的问题。

截至 2024 年 10 月，华为数据空间研究与应用工作已取得较明显的成果：累计构建了 21 个以上的业务数据空间，部署并开通了 100 多个数据连接器。此外，已有超过 1 万注册用户，签署了超过 1 万份线上数据交换合约，上架交换的数据资源超过 5.5 万个。华为作为复杂企业，应用数据空间的解决方案已经基本成熟，并积累了较广泛的企业应用实践经验，整体情况如图 9-1 所示。

沿着企业内各业务部门间、企业总部与分公司间、企业与生态链上下游间、企业与外部组织间四条主要线索，经过持续的探索实践，华为逐步积累了较多的实际案例。每个实际案例围绕理解和对准业务痛点与诉求，数据空间方案如何适配实施，以及如何实现业务目标等问题进行探索。本节选取了各类业务应用数据空间的典型案例进行详细介绍，让读者有一个直观的感受。

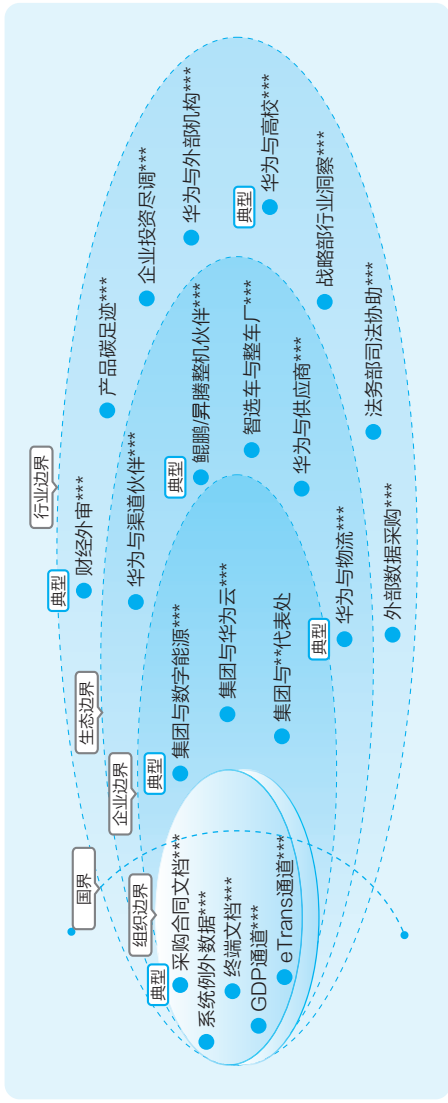


图 9-1　华为应用数据空间的四类场景

9.1.1 企业内各业务部门间的数据交换

在华为集团内，不同部门负责管理各自的数据，这些责任部门称为数据 Owner。例如，采购管理部门负责管理合同文档，而人力资源管理部则负责管理员工资料。数据 Owner 负责统一管理这些数据，并确保其在各类业务活动中得到充分利用。为了支持业务活动的有效开展，数据 Owner 会受理和审核其他部门对数据的获取需求。

在这一过程中，数据 Owner 承担着保护重要数据的责任，它们对数据滥用和扩散的风险非常警惕，因此，即使申请方是集团内部的其他组织，也有必要明确说明数据的使用范围、目的和方式。此外，对于重要数据的场景，申请方还需签署承诺书，保证不会滥用或扩散数据，并承诺在使用完毕后负责数据的销毁。数据 Owner 会监督这一过程，确保数据的安全和合规性得到维护。

华为在采购重要文档借阅、IT 系统数据例外提取等业务中进行了该场景的实践验证，形成了如下的应用案例。

案例：采购重要文档借阅

在企业集团内部，重要数据共享常因安全顾虑而受限。采购部门的关键合同文档需在集团内流通，但数据保护责任重大，导致对内部使用范围和期限有严格要求。为应对这一挑战，采购部门引入交换数据空间（Exchange Data Space，EDS），通过提供使用控制策略和查证追溯等功能，实现了文档的精准权限管理和安全共享。EDS 的实施不仅增强了安全保障，解决了数据提供方的顾虑，也简化了消费方的获取流程，促进了集团内部的数据流通和业务效率，同时确保了数据使用的合规性和可追溯性。采购重要文档借阅案例说明如图 9-2 所示。

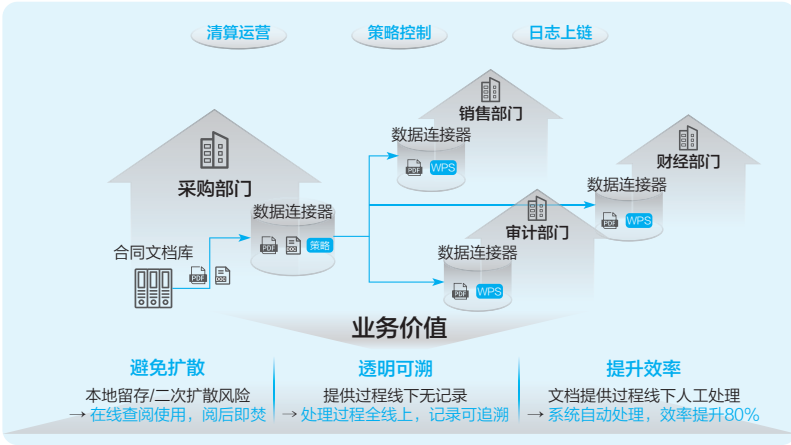


图 9-2　采购重要文档借阅案例说明

采购部门负责保管大量重要的合同文档，这些文档需要在集团内部向不同组织进行共享。然而，数据的敏感性和保密性要求提供方对数据的流通进行严格的控制，以防止滥用和未经授权的扩散。尽管消费方是集团内部组织，但提供方仍然需要确保数据的安全，要求消费方遵守严格的使用规则，包括限制使用范围、禁止滥用和到期后的数据销毁。以前由于缺乏有效的技术手段来实施这些控制措施，只能通过严格审核消费需求、让消费方签署承诺书等方式来进行管控，数据一旦提供就只能无条件相信消费方能够履行保护职责，出现违规滥用和泄露也只能进行事后追责。

该场景下，业务的关键诉求是实现对文档的控制精准到人，过程有迹可循。华为将 EDS 作为服务被高密文档的管理系统调用，方案的实施要点如下。

- 开通采购文档借阅数据空间，给采购部门部署数据连接器用于接入和提供文档，给销售、财经等部门部署数据连接器用于接收和使用文档。

- 消费方向采购部门提交文档使用需求，通常通过线上的电子流进行申请与审批。申请与审批的过程中，调用 EDS 的控制策略服务，可以根据管控要求选择文档使用策略，如限定用户、限定使用时长等。
- 审批通过的借阅需求，由数据连接器从文档库中采集对应的文档，并按照审批过程中生成的策略生成数据交换合约，将文件和策略一起投递到消费方所在的数据连接器中。
- 消费方收到通知，进入消费方数据连接器中，按需使用文档，使用过程受到控制策略的约束。策略到期后，文件权限收回，过程操作记录上报归档。

采用 EDS 后，企业集团在采购合同文档共享方面取得了显著的收益，实现了每年数千次借阅申请中数万份重要合同文档的安全共享。这种增强的安全保障不仅有效解决了提供方对数据共享的顾虑，也解决了消费方获取文档的困难。通过 EDS，企业提高了重要数据的共享效率，同时降低了数据滥用和泄露的风险。此外，EDS 还支持对全流程进行详细查证追溯，确保了数据共享与消费全流程的透明可溯。

9.1.2 企业总部与分公司间的数据交换

大型集团企业的组织结构往往较为复杂，由多个单元组成，包括职能部门、产业部门和销售部门等。这些企业通常采用“集团总部 + 分公司”的组织架构。在此模式下，集团总部负责统一管理和控制公共职能平台，如财务和供应链。而产业、销售部门则以独立分公司的形式运作，拥有一定程度的自主权，如特定的产业分公司或国家销售分公司。这种架构使得集团与分公司间的业务往来在许多情况下类似于两个独立实体之间的交易，要求双方明确权责、

基于商业合同进行业务往来。

在集团总部与分公司基于业务进行数据交换时，集团管理者普遍希望对数据流通进行分层记录和管理，旨在输出基于各个主体的线上统计账本，从而促进数据的充分共享。通过在集团总部和分公司构建数据空间，给各个主体部署数据连接器，有效记录在集团总部和各分公司之间的数据流转情况，从而自动形成按主体分类的数据流通账本。这个账本将展示每个主体拥有的数据量、提供的数据量、接收的数据量以及数据加工情况。通过优化数据记录和管理流程，集团能够更有效地监控数据流向，确保数据安全，并促进数据在集团内部的高效利用。这不仅能提升业务效率，还能加强对整个集团数据资产的有效管理。

华为在集团与数字能源业务单元、集团与华为云业务单元等业务中进行了该场景的实践验证。

案例：集团与数字能源业务单元

华为数字能源业务单元专注于数字技术和电力电子技术的融合，致力于推动清洁能源和能源数字化的发展。数字能源作为一项新的业务，虽然相对独立运作，但是业务开展依然离不开集团职能组织的支持，如供应链、财经等业务开展需要在集团职能组织基础上进行扩展，与集团进行充分的协同。为了更好地支撑数字能源业务单元，华为开通数字能源数据空间，并按需部署数据连接器，用于数字能源与集团各职能组织进行充分高效的数据交换共享。通过数据空间，可以提升集团与数字能源之间的数据交换共享效率，支撑数字能源业务对数据需求的自闭环，同时也能满足集团管控的要求，对数据跨主体的流转进行有效统计，生成清算报告。集团与数字能源业务单元交换应用案例说明如图 9-3 所示。

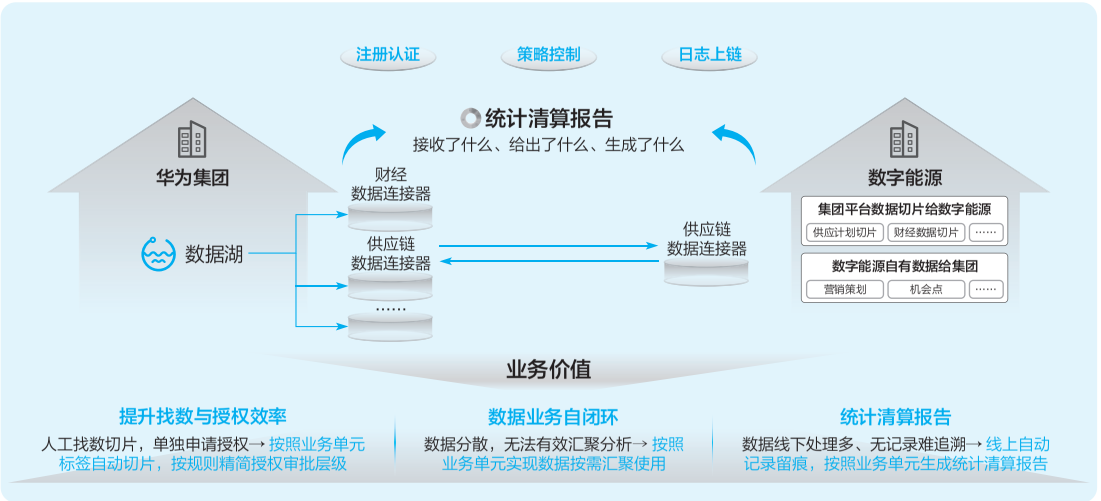


图 9-3　集团与数字能源业务单元交换应用案例说明

华为数字能源业务单元需要与集团内的众多职能部门进行数据交换共享。集团内部的数据分散在不同的部门和系统中，没有按照能源产业维度进行区分，这导致需要从全量数据中获取能源相关的数据，会变得复杂和低效。同时，没有基于产业维度的数据授权渠道，授权的沟通审批复杂，这不仅增加了管理成本，也延长了数据获取的时间。由于缺乏有效的数据共享交换渠道，集团与数字能源业务单元之间的协作和沟通成本增加，影响了整体的工作效率和业务创新能力。从集团统一管控的需求出发，也无法实现有效的线上记录统计。

该场景下业务的关键诉求是构建一套基于业务单元的专用数据流通渠道，实现数据的高效获取与授权，同时数据跨主体的流通过程能实时记录以实现统计清算。为此华为构建了以数字能源业务单元作为责任方的数据空间，该空间可以与其他业务数据空间进行各类数据的交换，通过签署用于留痕的数据交换合约，数据空间可以满足主体对统计清算运营的需求。

9.1.3　企业与生态链上下游间的数据交换

在企业发展过程中，构建以核心业务为中心的生态圈至关重要，这有助于吸引和整合多样化的合作伙伴，从而提升企业的竞争力。以华为为例，公司从产业线角度出发，不断围绕关键业务构建生态系统，包括数字基础设施的鲲鹏计算生态和昇腾 AI 生态、智能终端的鸿蒙 OS 和 HMS 生态、智能驾驶的 MDC（移动数据中心）生态，以及智能世界云底座的华为云生态等。华为不仅在产业链生态建设上发力，在供应链和采购等职能领域也注重与供应商和渠道合作伙伴建立紧密的生态网络，实现产业链各环节的高效协同和资源优化配置。

在生态链中，数据的充分交换是一个基础性问题。作为生态链的主导者，华为希望激发上下游合作伙伴的积极性，鼓励它们共享数据资源，并充分利用这些数据。所有生态链参与方都希望通过数据整合与共享深度挖掘生态数据价值，支撑业务创新。为此，华为联合各类合作伙伴打造生态数据空间，通过数据连接器实现数据的安全可控接入和获取，促进生态数据的充分共享与融合，解决生态中的复杂和共性问题，为整个生态链及其参与方带来收益。其一，业务通过数据空间增强了生态链的竞争力，因为数据共享促进了信息流通和知识共享，使各参与方能更快响应市场变化，共同开发新产品或服务。其二，效率提升是另一个重要收益，数据共享减少了信息孤岛，优化了资源配置和决策过程。其三，数据融合有助于创新，通过整合不同来源的数据，企业能发现新的商业模式和增长机会。其四，数据共享还帮助企业更好地理解客户需求，提供个性化产品和服务，增强客户满意度和忠诚度。其五，在供应链管理方面，数据共享提高了透明度和协同效率，减少了库存成本，优化了物流流程。其六，数据共享有助于风险管理，通过实时监控和分析数据，企业能及时发现并应对潜在风险。华为通过构建生态数据空间，旨在建立一个开放、协作、共赢的商业生态系统，推动整个产业链的持续健康发展。

华为在鲲鹏 / 昇腾生态、供应链与渠道伙伴、智选车与整车厂等业务中进行了该场景的实践验证，形成了成熟的应用案例。

案例：鲲鹏 / 昇腾与生态伙伴

在华为鲲鹏 / 昇腾生态中，华为作为主导企业，致力于推动上下游众多整机合作伙伴进行深度合作。华为鼓励各方共享数据，以充分利用这些资源，并确保所有参与方都能从生态链数据整合中获

得收益。企业间以前通常采用系统集成和线下协同的方式进行数据交换，这些方式面临断点较多和数据分散的问题，导致各方数据难以实现有效整合和充分利用。为解决这些问题，华为主导构建了鲲鹏 / 昇腾生态数据空间，与合作伙伴通过数据连接器安全可控地交换数据，有效支持解决生态中的复杂和共同关注的问题，如研发深度协同、多方协同营销、端到端质量追溯等。鲲鹏 / 昇腾生态应用案例说明如图 9-4 所示。

在鲲鹏 / 昇腾生态场景中，数十家合作伙伴与华为在研发、销售、供应、服务等各个业务环节有往来。现有的数据交换方式，如系统对接、邮件、电话等，无法有效整合与分析数据，导致在解决端到端问题时面临诸多困难，如产品质量的端到端联合追溯和研发过程的深度协同。为此，鲲鹏 / 昇腾数据空间围绕华为构建，所有参与方与华为建立直接的数据交换通道。基于业务需求，伙伴们可以交换各类复杂数据。具体的实施要点如下。

- 开通鲲鹏 / 昇腾数据空间，为华为和所有生态伙伴部署数据连接器。数据连接器可根据参与方实际情况，选择部署在华为公有云或生态伙伴指定的 IT 环境中，实现分布式部署。
- 根据业务领域需求，适配和叠加结构化和非结构化数据交换的接入能力，如支持 API、CSV、数据库、文件等类型接入。
- 主体间签署数据交换合约，根据不同场景选择合适的合约进行数据交换，如单次的文件交换合约、长期更新的结构化数据的交换合约等。
- 接入业务分析工具对数据进行融合分析，如接入数据加工与报表制作工具，对各方数据进行融合，并生成端到端的质量追溯报告，共享加工结果。

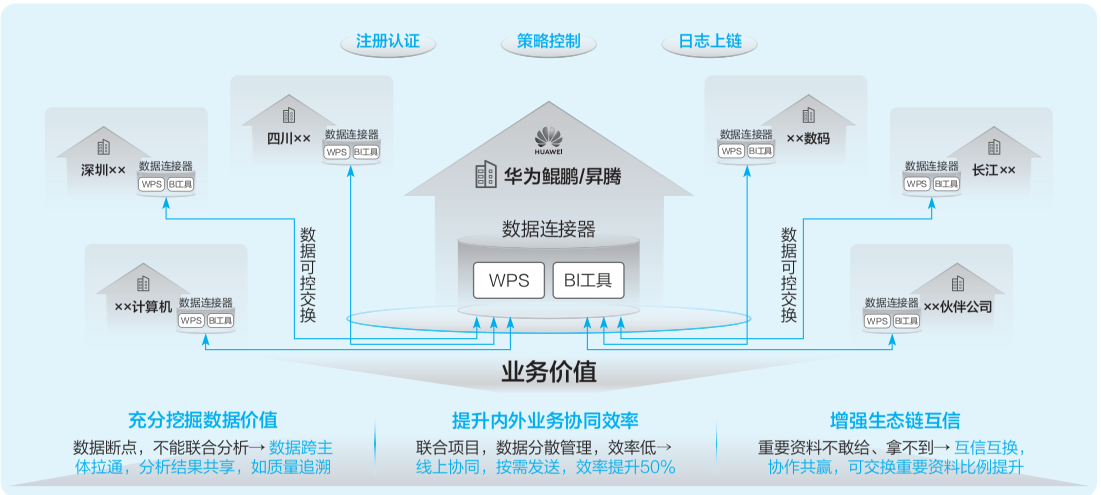


图 9-4　鲲鹏 / 昇腾生态应用案例说明

通过鲲鹏 / 昇腾数据空间，华为与生态伙伴之间的数据交换与业务协同得到促进。业务收益主要体现在以下方面：

- 通过生态数据的充分共享与融合，支持复杂问题处理，如通过质量数据融合与分析，实现质量问题端到端联合追溯。
- 提升生态各方的业务协同效率，数据空间中的数据连接器可以快速接入各类数据，进行批量、往返的数据交换，替代原有线下手工作业，提升业务作业效率。
- 增强生态链的互信，以往华为与生态伙伴间存在数据共享的顾虑，随着数据共享收益的显现，各参与方对生态链的信任度增强，愿意主动共享高价值数据。

9.1.4　企业与外部组织间的数据交换

在企业运营过程中，与外部组织的业务互动和数据交换是常态。例如，与审计机构合作进行外部审计，与政府机关协作以提供司法协助，或与高校联合开展科研项目。这些合作通常聚焦于特定业务活动，与长期全面的生态伙伴协同有所区别。在这些情况下，各方对数据交换的管理要求通常更为严格，以保护各自权益并确保活动的合法合规。对此，数据的提供方普遍强调在数据交换过程中遵循最小授权原则。这意味着企业在共享数据时，应仅授予完成特定任务所必需的最低权限，以保障数据安全和保密。同时，企业也必须确保接收数据的使用符合法律法规，防止不当使用带来的法律风险。这种做法有助于有效保护各方利益，并确保业务活动的顺畅进行。

企业与外部组织之间的数据交换变得越来越频繁，但这种跨企业的数据流通常常面临信任基础薄弱和数据敏感性高的挑战。通过构建数据空间、引入技术手段，增强了各方之间的信任，确保了数据的安全流通和合规使用。数据空间使提供数据的企业能够对外部合作伙

伴进行数据的最小授权，这种精细的权限控制不仅保护了数据提供方的利益，也满足了其对数据保护的严格要求。同时，数据空间还允许数据消费方通过透明的数据使用记录来自证清白，确保其行为符合既定的数据使用政策和法规要求，同时规避潜在的法律和商业风险。

华为在校企联合科研、企业财经外审、企业投资尽调、对外司法协助等业务中进行了该场景的实践验证，形成了成熟的应用案例。

案例：校企联合科研

在企业与行业内外组织的交流合作，如与高校的联合科研或与行业机构的成果交流中，存在着一种普遍现象：参与方众多，各方都拥有大量高价值的研究成果，它们渴望通过共享成果实现合作共赢，同时也极为重视对自身知识产权的保护。与垂直生态链中通常由一个“链主”主导并基于特定业务目的的数据交换不同，行业内的合作更倾向于平等的网状关系，各方都主张平等合作。这种合作模式带来了较大的不确定性，因为通常在消费方在按需搜索和匹配后才能确定所需的数据。

为了构建一个适用于多方平等合作类型的数据空间，华为不仅为各方部署了数据连接器，还引入了多方共享的目录系统。通过这一目录系统，各方能够上架其元数据，而消费方则可以根据需求进行搜索和匹配，一旦匹配成功，便可以申请订阅。这种方法能够最大限度地开放各方的高价值数据，同时，由于仅展示元数据而非原始数据内容，也能有效保护数据的安全性。对于消费方而言，在需求不明确的情况下，它们可以进行广泛的搜索和匹配，从而找到最符合需求的数据资源。这样的数据空间不仅促进了数据的开放和共享，也确保了数据的安全和各方的利益，为行业内的合作提供了一个高效、安全的平台。校企联合科研应用案例说明如图 9-5 所示。

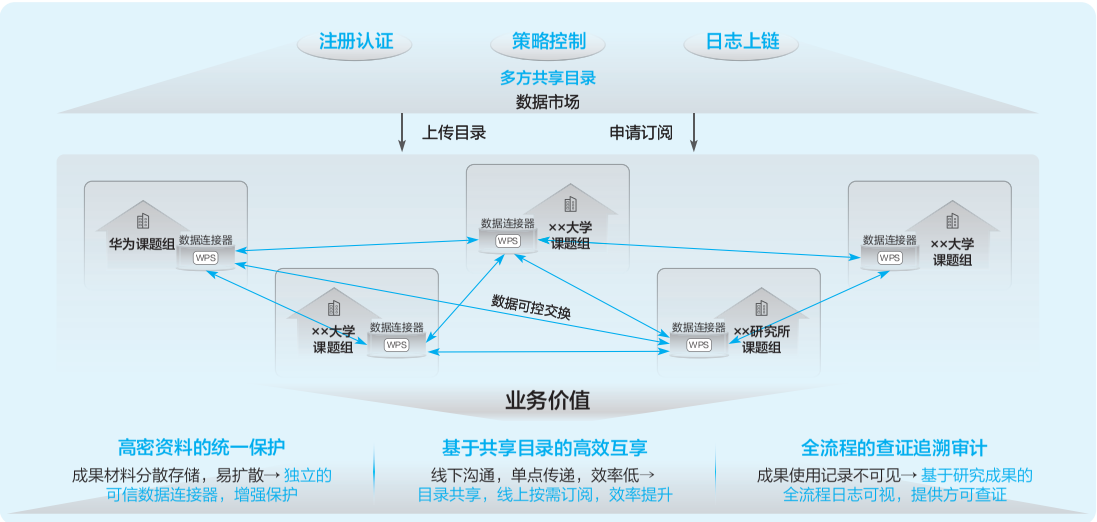


图 9-5　校企联合科研应用案例说明

在当前的科研合作环境中，华为等企业与外部高校和研究机构共同承担复杂科研课题的研究任务。这种合作模式下，各方在分工承接子课题的同时，也期望能够按需共享研究成果，以促进研究进展。然而，由于对知识产权保护的考量，如何在保障研究成果权益的基础上实现数据的充分共享，成为一大挑战。传统的数据共享方式可能无法满足各方对数据安全性和使用控制的需求，导致合作效率受限，研究成果的流通和应用速度减缓。

在构建联合科研的数据空间时，核心是实现多方平等参与，并部署一个公共的数据目录实现充分的共享。具体的实施要点如下。

- 开通联合科研数据空间，为华为和所有科研机构部署数据连接器。数据连接器可根据参与方实际情况，选择部署在华为公有云或机构指定的 IT 环境中，实现分布式部署。
- 部署多方可自由访问的共享目录，各方可以将自己的研究成果以数据 Offer 的形式上架，空间内其他参与方可在共享目录中搜索和匹配自身需要的数据，并发起订阅和合约协商。
- 数据提供方通过共享目录受理其他参与方的数据获取及合约协商申请，经过协商达成一致后签署数据交换合约。
- 数据提供方基于合约将数据发送至数据消费方的数据连接器中，实现受控使用。
- 各方可以进行数据追溯，也能访问清算中心获取详细的记录，保证交换过程的透明。

实施联合科研数据空间的收益很多，具体体现在以下方面。

- 数据空间有效保护各方的资料，实现统一水平的有效保护，确保各方产出成果的权益不受侵犯。

- 通过数据空间，华为与外部高校、研究机构能够形成互享互利的循环机制，促进内部每年众多校企联合科研项目的顺利进行。
- 数据在数据空间中的自由流通，配合全流程的查证追溯能力，确保了科研活动的合规性，为各参与方提供了一个安全、高效、透明的合作平台。这样的数据空间不仅加速了科研进展，还为未来的科研合作模式提供了新的可能性。

案例：财经外部审计

财经外部审计（外审）是财务部门每年的一项重要例行工作。华为的财经部门与审计事务所需要反复交换大量的资料，资料需要从各个领域整理收集并发送给对应的专业审计师，需要传递的数据量大而且分类规则复杂，其中还包含大量的高密数据，数据传递涉及全球上百个国家。为此，华为构建财经外审数据空间并进行全球分布式部署，外审数据接入财经的数据连接器并按照规则自动分配，匹配对应的合约，发送到对应国家的审计事务所，限定指定的审计师使用。通过这样的环节，实现了外审资料访问的精准到人，复杂的过程可以进行查证追溯，同时也大幅提升了财经人员和审计师的工作效率。财经外审应用案例说明如图 9-6 所示。

财经外审通常分为中期审阅、预审与终审。在各审计阶段，华为的财经部门与审计事务所需要反复交换大量的资料，资料需要从各个领域整理收集并发送给对应的专业审计师。其中存在两个关键痛点：一个是数据量大，需要人工仔细分发给对应的专业审计师，传递效率低且难以跟踪进度，一旦分发错误难以撤销和追溯；另一个就是很多重要的原始资料底稿，在查看浏览后需要销毁，审计师通过签署承诺书的方式获取使用，但实际资料的流向和使用无法管控，存在泄露风险。

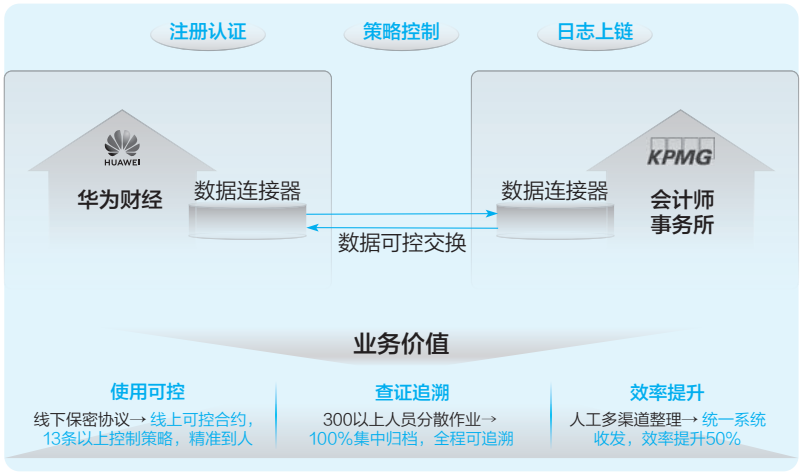


图 9-6　财经外审应用案例说明

针对业务提出的提升效率、精准分发和使用可控等诉求，在部署外审数据空间时涉及的实施要点如下。

- 开通外审数据空间，数据连接器的部署支持在全球进行部署与存储。
- 支持海量资料接入数据连接器，提供批量集成功能，并基于审计主体、PBC（由客户提供）科目等业务标签自动分类归集，通过团队和文件夹实现清晰有效的管理。
- 支持按照资料分类设置不同的使用合约，将资料批量发送给对应的外部审计师，实现资料分发精准到人，审计师按策略在线或者下载使用资料，到期自动销毁，避免扩散。

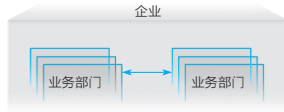
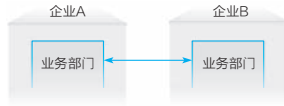
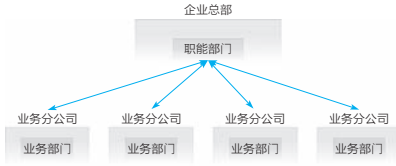
华为外审通过数据空间实现华为数十个被审计主体与全球数十个审计事务所之间的每年超过 100 个审计科目数万份资料的传递。外审数据空间为业务带来的收益如下。

- 为线下签署的保密协议提供了一层技术保障。通过线上合约实现保密条款在数据空间中强制实施，避免出现违规操作和数据泄露。
- 清晰记录海量复杂数据的传递与使用。有效解决了以前超过 100 个审计科目数万份资料的分类、传递、使用过程记录的难题，线上自动记录留痕，实现清晰的统计分析。
- 提升作业效率。财经作业人员可以高效地传递海量资料、实现资料访问的精准到人。外部审计师可以随时看到资料获取进度，减少了反复沟通的成本。

9.2　部署数据空间的三类典型范式

基于华为自身的业务实践，我们发现不同的业务形态对应不同的数据空间形态。华为从运营管理的主要维度考虑，形成了 1∶1、1∶*N*、*N*∶*N* 三种典型的业务范式，对各范式的比较分析见表 9-1。

表 9-1　部署数据空间的三类典型范式

业务范式	运营管理方	典型应用场景	业务形态
1∶1	供需双方自运营	企业内各业务部门间	
		不同企业的业务部门间	
1∶*N*	主导主体管理运营	企业总部与分公司间	

（续）

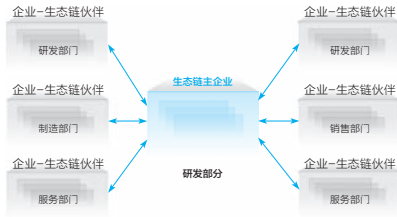
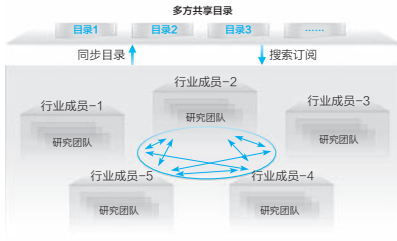
业务范式	运营管理方	典型应用场景	业务形态
1∶N	主导主体管理运营	企业与生态链上下游间	
N∶N	第三方运营 / 共同运营	企业与外部组织	

表 9-1 给出了企业应用数据空间的三种典型范式，下文将针对 1∶1、1∶*N*、*N*∶*N* 三种范式的特点、适用场景和优缺点做进一步的说明。

9.2.1 范式 1：1∶1 模式

华为基于采购文档借阅、投资尽调等案例，总结出 1∶1 模式。该模式既包含企业内各业务部门间的点对点交换，也包含企业间的点对点交换。该模式有如下特征。

（1）交换关系

- 直接连接：两个参与方之间直接建立连接，不依赖中介。
- 简单高效：只涉及两个节点，直接进行协商和交换，过程简单。

（2）运营管理

- 自我管理：每个参与方自行管理自己的数据和交换流程。
- 无须协调：不存在多方协调和运营管理的需求，降低了管理复杂性。

（3）优缺点

- 优点：简单快捷、成本低廉，适合小规模或私密性强的数据交换。
- 缺点：可扩展性和灵活性不足，难以扩展到更多参与方，安全性依赖双方的信任和技术水平。

1∶1 模式由于点对点，简单直接，适合于私密性强、参与方固定的数据交换，其优势在于简单高效，但缺点是扩展性有限。具体应用在诸如点对点项目合作、敏感数据交换等业务中。

9.2.2　范式 2：1∶*N* 模式

华为基于集团与分公司、生态链、外审等案例，总结出 1∶*N* 模式。在该模式下，数据空间由一个业务上的主导方来牵引构建，按责任主体开通多个数据连接器，由主导方来负责整个数据空间的运营管理。该模式在企业内部与分公司、生态链中都有适用场景，有如下特征。

（1）交换关系

中心化结构：一个主导参与方作为中心节点，其他分支参与方围绕中心节点构建网络。

（2）运营管理

- 中心化管理：由主导参与方负责空间的运营管理，包括参与方认证、交换标准制定等。

- 标准化流程：中心节点制定统一的交换标准和流程，确保数据交换的一致性和合规性。

（3）优缺点

- 优点：便于集中管理和控制，适合大规模数据交换和管理需求，提高了数据交换的标准化和效率。
- 缺点：中心节点可能成为瓶颈，对中心节点的依赖可能导致单点故障，可能存在中心化带来的权力集中问题。

1∶*N* 链主主导模式适合于需要集中管理和控制的场景，中心节点应具备强大的数据处理和协调能力，同时需要建立标准化的数据交换流程。在企业中具体应用在诸如企业总部和分公司间的数据交换、企业与生态上下游间的数据交换等业务中。

9.2.3 范式 3：*N*∶*N* 模式

华为基于联合科研等案例，总结出 *N*∶*N* 模式。在该模式下，成员是 *N*∶*N* 网状关系，各方都希望看到其他行业成员开放出来的数据，从而实现按需搜索和获取需要的数据，这样就需要部署一个多方的数据市场支撑按需订阅交换。由于各方的平等交换关系，数据空间由第三方或者参与方共同进行运营管理。

（1）交换关系

分布式交换：所有参与方平等连接，形成一个多对多的网络结构，没有明显的中心。

（2）运营管理

分布式管理：参与方之间自主协商数据交换的规则和条件，共同参与运营管理或委托给第三方。

（3）优缺点

- 优点：增强了数据交换的灵活性和扩展性，没有单点故障的风险，可充分建立连接。
- 缺点：管理复杂性增加，需要有效的共识机制来协调多方，可能存在协调成本高和效率低的问题。

N：*N* 模式适合于多方参与、平等交换的场景，需要建立有效的共识机制和数据治理框架，以确保数据交换的有序性和合规性。该模式具体应用在诸如构建行业生态解决共同问题、多方联合创新分工合作等业务中。

9.3　设计业务模板以满足场景复用需求

业务范式是从产品功能收敛的视角考虑了通用性，结合业务具体的应用实践经验沉淀而来的。对于每个场景下如何将业务流程与管理规则融入数据空间中，不同的业务在使用和运营过程中也积累了实际的业务经验。华为在不断拓展数据空间应用范围的时候，进一步提出不仅要从产品能力的视角提供应用数据空间的范式参考，还应该将已经运行的业务数据空间的实践经验沉淀到数据空间中，形成数据空间的业务模板，这样就可以实现新业务通过学习成熟业务，快速复制经验的效果。

从业务视角来看，在华为的这些业务实践中业务参与方既有内部组织部门，也有各分公司，还有众多的外部伙伴和机构；交换的数据内容既有各类非结构化文档，也有 API 和 DB 形式的结构化数据；交换的形式既有 1：1 的单向传递，也有以华为为中心的 1：*N* 星型交换，还有各方按需订阅的 *N*：*N* 网状自由交换；消费场景既

有个人用户的文件浏览编辑、数据加工分析，也有特定业务 App 的使用等；控制层级既有严格约束消费行为的使用控制策略，也有记录上报日志实现留痕追溯的控制等。基于数据空间的运行经验，从业务视角来看，业务规则主要落地到管理和执行两个层面，为此我们提出了业务模板模型，如图 9-7 所示。

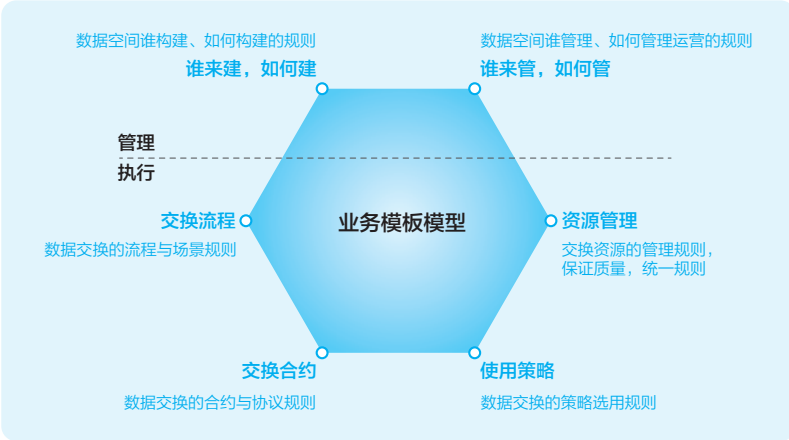


图 9-7　业务模板模型

业务模板模型主要回答业务在管理和执行过程中所涉及的必要环节和关键问题。华为以外审为例输出了企业财经外审 PBC 资料传递业务模板，模板包含外审业务中涉及的从管理到操作层面的业务实践经验沉淀，可以用于指导数据空间的各层管理者，实现空间的快速搭建，落地管理规则。如管理规则就包含数据空间的管理要求，在外审数据空间中采用 1∶*N* 的形态，对参与的主体设置企业和个人身份的注册管理等细则。同时也能指导具体的业务执行者，实现操作经验的传递，让业务快速启用。如外审业务中针对文档密级设置 3 类管控层级的交换合约，资源按照审计主体与审计科目设置资源编目等规则，由业务保存并落地到对应的业务模板中。应用案例如图 9-8 所示。

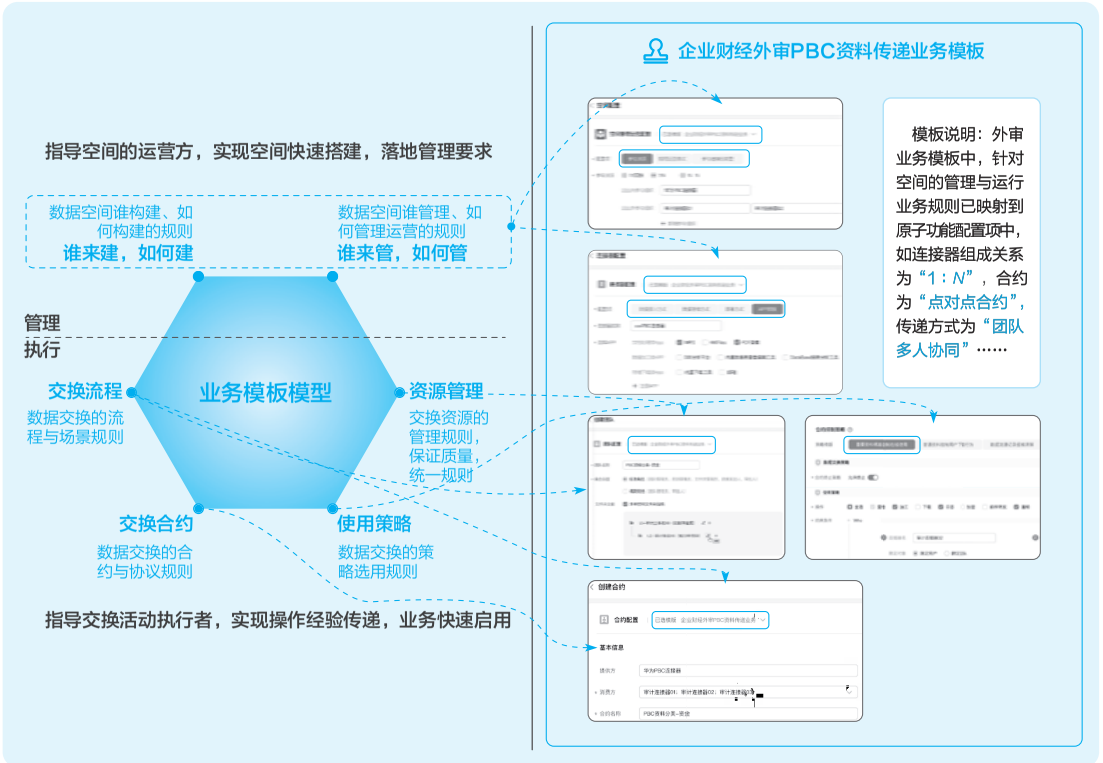


图 9-8　业务场景模板的应用案例

目前基于华为的业务实践已形成 14 类业务场景模板，业务模板通过模板库进行共享。业务在运行和运营过程中，管理和执行的规则可以进行持续更新，在业务拓展和复用的过程中起到支撑作用。

9.4 小结

本章综合分析了企业在应用数据空间时遇到的四类主要场景，并探讨了部署数据空间的三类典型范式，旨在为企业数据流通交换提供指导和参考。

首先，根据交换主体、方式、边界和管理等维度，将企业的数据流通需求归纳为四大类场景：企业内各业务部门间的数据交换、企业总部与分公司间的数据交换、企业与生态链上下游间的数据交换以及企业与外部组织间的数据交换。这些场景覆盖了从内部管理到外部合作的多个层面，每个场景都有其特定的业务痛点和需求。通过对这些场景的深入分析和对相关实践案例的介绍，本章展示了数据空间方案如何适配实施，以及如何助力实现业务目标。

其次，根据华为的业务实践，提出了 1∶1 模式、1∶N 模式和 N∶N 模式三种典型的数据空间部署范式。这些范式分别对应不同的业务形态和运营管理需求，从点对点的简单交换到多方参与的复杂交换，体现了数据空间在不同业务场景下的灵活性和适应性。

最后，强调了设计业务模板的重要性，以满足不同场景下数据空间的复用需求。通过将成熟的业务经验沉淀为数据空间的业务模板，新业务可以快速学习和复制成功经验，从而提高数据空间的应用效率和效果。

10

第 10 章 CHAPTER

数据空间与 AI 大模型

在智能时代，得益于高质量数据的推动，AI 大模型成为释放数据价值的新引擎。AI 大模型的训练和应用依赖大规模、多样化的数据集，这些数据集不仅蕴含企业的核心知识，也具有高价值和保密性。智能时代为数据共享带来新的机遇，同时也带来了保障数据主权的挑战。数据作为第五生产要素，其流通不仅促进知识共享和创新，还推动 AI 生态的繁荣发展。然而，数据的易复制性、确权难、非排他性等特性，使得构建基于价值的数据流通体系面临挑战。数据空间的创新解决方案，通过实现数据交易的“可信、可控、可证”，为数据的高效、安全流通提供了保障，解决了流通中的难题，为 AI 数据集的买卖双方创造了一个安全、高效的交易环境。

10.1 AI 大模型时代数据既丰富又稀缺

2023 年是 AI 大模型爆发的元年。2022 年 11 月 30 日，ChatGPT 横空出世，上线后仅 2 个月就突破一亿用户数，打破 TikTok 达到一亿用户用时九个月的纪录，成为史上增速最快的应用。ChatGPT 的火爆标志着 AI 时代的来临，开启了 AIGC 的全新时代，它是一场颠覆性的变革，将会重构我们的生活及工作。

10.1.1 数据成为驱动 AI 大模型发展的核心要素

数据、算力、算法是构筑 AIGC 体系的三大要素，其中数据是影响 AIGC 的核心要素，数据的高度决定了 AIGC 的高度。业界普遍认为以数据驱动的 AI 时代已经来临。

在 Gartner 2022 年发布的人工智能技术成熟度曲线中，“以数据为中心的人工智能”（Data-centric AI）被列为人工智能技术和应用的四大创新类别之一，高质量的训练数据集、完备的数据应用策略将会更好地服务于模型的开发与应用。

知名人工智能专家吴恩达教授提出“二八定律”：80% 的数据 + 20% 的模型 = 更好的机器学习，模型的能力取决于给它投喂了多少高质量的数据，一个模型的好坏 20% 由算法决定，80% 由数据决定。以 GPT 为例，GPT 模型架构从第 1 代到第 4 代均较为相似，而用来训练模型的数据规模和质量却有很大的不同。GPT-1 是由 4.8GB 未过滤原始数据训练的，GPT-2 是由经人类过滤后的 40GB 数据训练的，GPT-3 是由从 45TB 原始数据中过滤出的 570GB 数据训练的，而 GPT-4 则在此基础上又加入了高质量的人类标注数据，使得 GPT-4 在理解和生成自然语言方面达到了前所未有的水平。

此外，OpenAI 的缩放定律（Scaling laws）指出，模型性能随着计算量、数据量和参数量的提升而提升。DeepMind 的 Chinchilla 缩放定律指出：模型大小和训练词元（Token）的数量应等比缩放，即模型大小每增加一倍，训练词元的数量也应增加一倍；大模型要经过其参数量 20 倍的数据训练后才达到饱和。OpenAI 的阿尔特曼定律指出，“智能”的增长通过计算参数数量和词元数量来衡量。通过以上三个定律不难发现，数据在 AI 大模型中扮演着至关重要的角色。

与此同时，AI 已成为释放数据价值的新引擎，AI 发展的突破得益于高质量数据的发展，AI 大模型的训练和推理应用过程都离不开高质量、大规模、多样性的数据集。这些数据集可能会涉及企业的高价值、高密级的数据。如何使数据提供方更愿意共享数据集给 AI 大模型训练，并保障其数据主权，成为智能时代数据流通的新挑战。

10.1.2　激活非结构化数据资产的潜在价值

AI 技术的发展，尤其是 AIGC 的出现，大幅提升了计算机处理与利用数据的能力，之前无法处理的数据能够被处理，之前低价值的数据也可以被有效利用。以往企业应用使用的主要是规模较小、模态单一的结构化数据，而大模型使用的主要是大规模、多模态的非结构化数据，如图 10-1 所示。

IDC 调研显示，目前企业的数据中有高达 80% 是非结构化数据，这些非结构化数据，如文本、图像、音频和视频等，蕴含着丰富的信息和潜在价值，通过大模型的深度学习算法对其进行分析和处理，可以激活其价值。例如，自然语言处理技术可以从文本中提取关键信息，图像识别技术可以识别图像内容，语音识别技

术可以将语音转换为文本。大模型通过训练学习数据中的模式和关联，能够洞察非结构化数据中的隐含趋势、情感倾向和行为模式，为决策提供支持，优化服务，推动创新。这种能力使得非结构化数据成为企业和研究者挖掘洞察、提升效率和创造价值的重要资源。

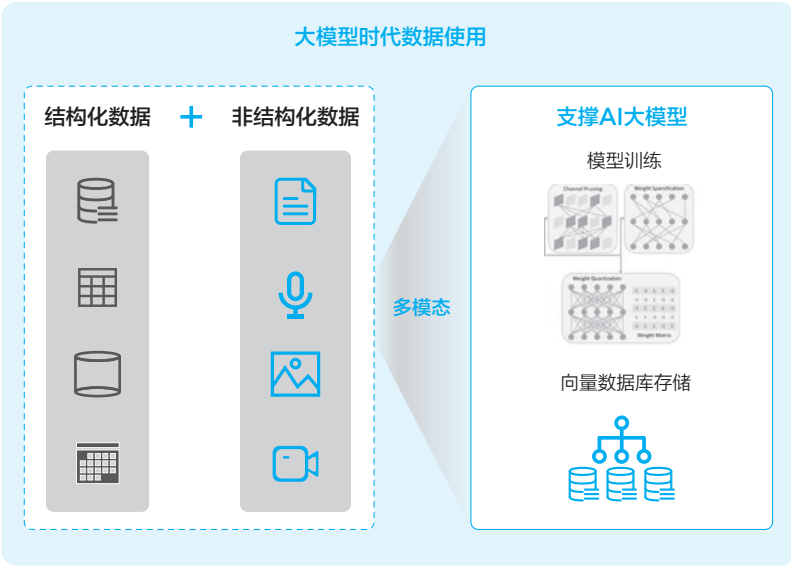


图 10-1　大模型使用的主要是大规模、多模态的非结构化数据

此外，随着数字化转型的推进，企业作业的交互方式也发生了改变，大量业务是通过图片、视频、语音来实现运作的，如通过拍照来实现货物的签收。这些作业流中产生的非结构化数据也是蕴含着重要信息的资产。

大模型可以对各种非结构化数据进行转换，变成多维的向量，将企业长期以来积累的所有信息与知识投入模型，为业务服务。以

华为为例，在业务流交互过程中产生的图片、视频、文件以及公司积累的大量政策发文、流程文件、专家案例、会议纪要、培训课程等高价值资产，都将被大模型进行创新、优化和重塑，从而使更多数据释放价值。

10.1.3　高质量数据集成为稀缺资源

在人工智能和机器学习飞速发展的背景下，数据的质量直接决定了算法模型的性能和输出的准确性。然而，大模型训练所需数据集总量的增速远大于高质量数据生成的速度，导致高质量数据稀缺。

然而，获取高质量数据集并非易事。现实世界中的数据往往是杂乱无章的，这就需要数据集提供方花大量的时间与成本进行清洗、标注和验证。企业在模型训练过程中有 80% 的时间耗费在数据准备与数据处理上，构建数据集是模型训练中最耗时耗力的部分。

此外，基于数据的非排他性特性，一个高质量的数据集能够被多个 AI 大模型共享和使用，而不会降低其价值和可用性。随着数据集在不同模型中的应用与流动，它们被赋予了更高的价值。

高质量数据集不仅能够提升模型的性能，还能够加快模型训练的进程，从而加速 AI 技术的发展和应用。掌握高质量数据集意味着拥有竞争优势和市场先机。随着 AI 技术的不断发展，高质量数据集的需求将会持续增长，它们的价值也会得到比以往更大的认可。

10.2 AI 大模型时代数据流通的挑战

在 AI 大模型时代，数据流通的挑战日益凸显。大模型训练需依赖大量数据，这些数据不仅要可用，还要具有可见性。但随着数据量的激增，大规模高价值数据的流通带来了更高的安全风险，包括数据泄露、滥用等威胁。同时，数据形态的不断演变导致传统的数据权限管理方法可能失效，使数据的举证和权限验证变得更加困难。这给数据流通带来了新的挑战，因此需要依靠更有效的技术路线和方法来实现数据的共享和保护。

10.2.1 AI 大模型需要数据可用又可见

在小模型时代，隐私计算和联邦学习等技术的核心理念是“数据可用不可见”，即在不暴露原始数据的前提下，允许对数据进行分析和学习，从而保护数据的隐私和安全。隐私计算技术，如安全多方计算（MPC）、同态加密（HE）和零知识证明（ZKP），能够在数据加密状态下进行计算，确保数据在处理过程中不会被泄露。这些技术允许多个参与方在不共享各自数据的情况下，共同完成计算任务。联邦学习则是一种分布式机器学习方法，它允许多个参与方在本地训练模型，并通过交换模型参数或梯度信息来共同优化模型，而无须直接共享原始数据。这种方式不仅保护了数据的隐私，还促进了跨机构的协作和知识共享。

然而，随着大模型时代的到来，数据量剧增和模型训练方式不断演进，这些技术面临着新的挑战。例如在进行大模型训练时，针对不同训练阶段需要把数据集转换成不同的模式，比如在监督微调（SFT）阶段，需要通过领域专家把原始数据集转换成高质量的问答对，数据不仅要可用也要可见。而且 AIGC 有“记忆”数据的

能力，数据一旦被训练进模型，通过提示工程等方式都有可能复现训练进模型的原始数据，从而使数据泄露。

隐私计算是在不暴露数据的情况下进行计算；联邦学习是数据在本地，多个参与方共同训练模型。但在大模型时代，数据的可用性和可见性要求对隐私计算和联邦学习技术提出了挑战，需要依靠新的技术路线和方法来实现数据的共享和保护。

10.2.2　大规模高质量数据流通安全风险更高

大模型的泛化能力往往与其训练数据的广度和深度密切相关。为了使模型能够更好地理解和处理各种复杂情况，需要引入更多的数据，这使得数据流通规模扩大，从而支持大模型的复杂计算和学习。同时，因为高质量数据的稀缺性，数据的价值变得更加显著。

随着数据流通规模的增加与数据价值的提升，数据泄露、滥用或未授权访问的风险也随之增加。面对这些变化，数据集提供方需要更有效的保障措施来确保其权益。它们需要确保与模型消费方签订的协议得到实际执行，防止数据滥用和侵权行为。数据空间提供了一种创新的解决方案，它通过建立一个可控、可信、可证的数据交换环境，保障了数据持有者的权益。在这个空间内，数据流通的过程可以被实时监控和记录，确保数据的使用符合协议规定，从而降低数据滥用的风险。数据空间作为促进数据流通和保障数据权益的有效平台，为数据持有者提供了一个可靠的解决方案。

10.2.3　数据形态变化导致权限失效举证更难

在大模型时代，企业对数据集的使用呈现出显著的变化，数据除了被用于支持传统的业务流程，还是训练和优化大型机器学习模型的关键资源。数据主要被用于模型微调和 RAG（Retrieval-

Augmented Generation，检索增强生成）向量知识库的构建，如图 10-2 所示。

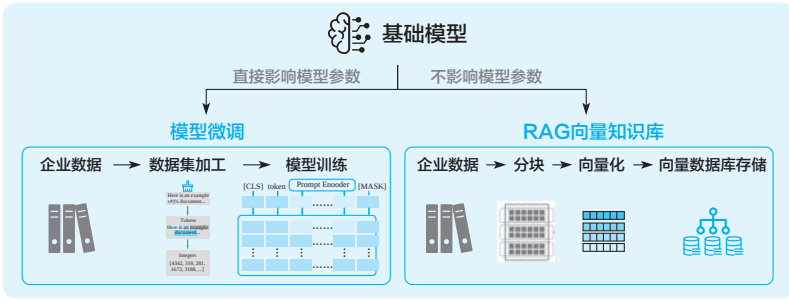


图 10-2　企业数据在大模型中的两种主要应用方式

模型微调是指企业在引用基础大模型的基础上，使用企业数据经过清洗、加工、处理后得到的数据集进行二次训练，以适应特定的业务场景或任务。RAG 向量知识库则是将企业数据进行分块、向量化后，以向量的形式存储，使模型能够更有效地检索和利用知识，从而提高生成内容的准确性和相关性。

大模型时代，数据经过向量化后使用的形态发生改变，数据要么被转换成高维向量存储在向量数据库，用户通过外挂向量知识库进行使用，要么经过加工处理后通过改变模型参数参与模型训练。

在传统的数据处理中，结构化数据的权限管理相对直接，通常通过数据库管理系统实现，利用表和字段级别的访问控制来确保数据安全。例如，特定的用户或用户组可能被授权访问特定的数据库表或字段，而对其他数据则没有访问权限。这种方法能够明确地界定数据访问的边界。对于非结构化数据，如文本、图片或视

频，权限管理通常通过文件权限来实现。文件权限可以控制谁可以读取、写入或执行文件，从而间接地管理对数据的访问。但是数据被转换成高维向量或模型参数后，数据的形态和存储方式会发生改变。在向量化过程中，原始数据的结构和语义信息被编码到连续的向量中，这使数据的边界变得模糊，难以用传统方法进行权限控制，如图 10-3 所示。此外，模型参数可能涉及多个数据源，难以追溯到具体的数据项或数据集，数据原有的权限管控将会失效。

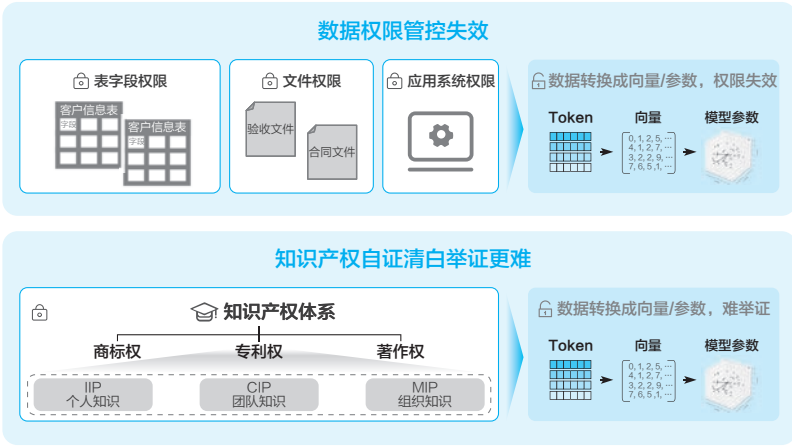


图 10-3　数据转换成向量与参数，权限失效，举证更难

另外，当前 AIGC 领域版权诉讼频发，存在许多使用未经授权的版权作品进行大模型训练的行为。而以往通过知识产权体系对知识进行保护与控制的手段，在大模型时代可能会失效。生成式 AI 本身运行机制较为复杂，想要证明其侵权存在很大的困难。首先，生成式 AI 能够生成大量新颖的内容，这些内容可能与现有作品相似，但并不完全复制，使界定侵权与否变得更加复杂。其次，大模型训练时使用的数据集可能包含来自不同来源的文本、图像、音频

等，这使追踪特定内容的原始数据版权归属变得更加不易。最后，数据变成了模型参数与向量后，在侵权行为发生后想要保全证据、自证清白的技术难度也有所增加。

因此，大模型时代，数据的使用由原来的表、字段、文件等转化成向量与参数的形态后，传统的权限失效，知识产权侵权举证更加困难。

10.3 探索 AI 大模型的数据控制

华为数据空间在大模型场景下的探索是通过数据空间的认证中心、数据合约、使用控制策略、App 及标准 SDK 服务、数据清算等能力，帮助 AI 数据集提供方实现数据全链路可视与使用控制，促进 AI 数据集的交换及共享。数据的可控交换涉及数据的存储、传输、使用的安全与控制，控制程度不同，投入的成本不同。AI 数据集的可控交换方案可以概括为三个阶段，分别是线下签订协议、创建空间与注册、AI 数据集流通与控制。此外，基于数据集提供方对数据控制需求的强弱，AI 数据集流通与控制又可分为控制 AI 大模型读取的授权管控方案，控制 AI 大模型读取与加工的双因子管控方案，控制 AI 大模型读取、加工与计算的 GPU 管控方案。AI 数据集的可控交换方案如图 10-4 所示。

1. 第一阶段：线下签订协议

AI 数据集提供方与大模型消费方基于数据集交换达成一致意见，并签订协议。协议内容需包括被访问的数据集、允许访问数据集的大模型、数据集的访问期限等。

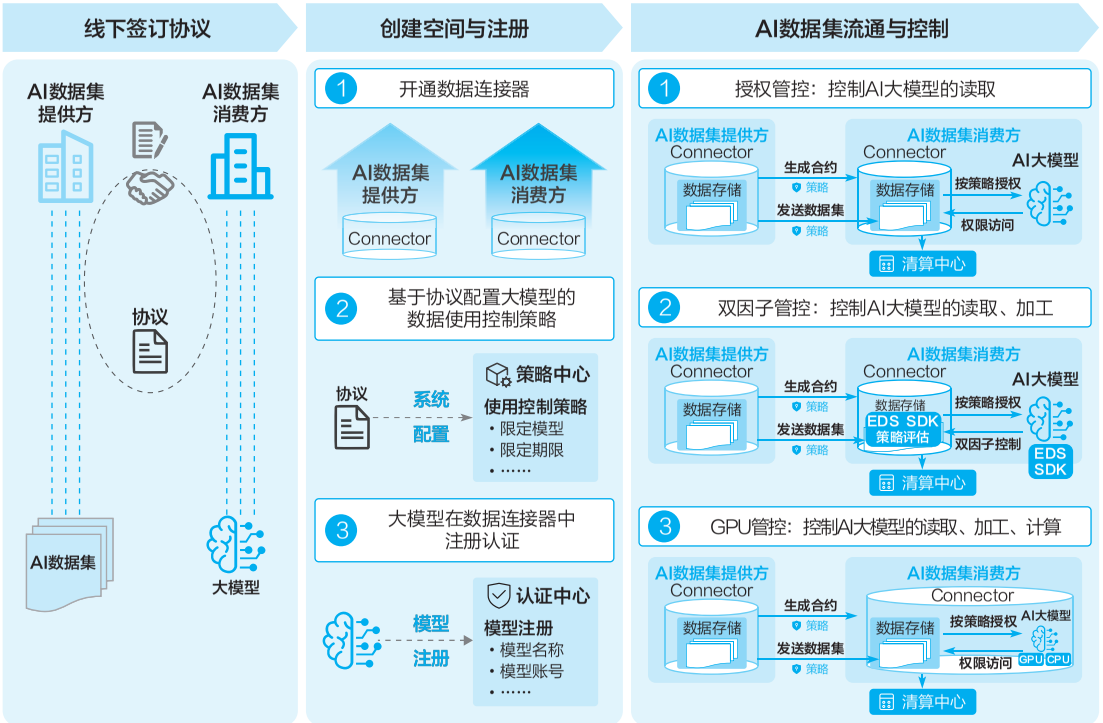


图 10-4　AI 数据集的可控交换方案

2. 第二阶段：创建空间与注册

1）开通数据连接器：给 AI 数据集提供方与大模型消费方分别开通数据连接器，用于管理、交换和使用 AI 数据集。

2）基于协议配置大模型的数据使用控制策略，主要包括以下内容：限定可以访问数据集的模型；限定可以访问数据集的接口；限定可以访问数据集的账号；限定可以访问数据集的期限等。

3）大模型在数据连接器中注册认证：在认证中心注册大模型的名称、账号、模型运行环境等相关信息；认证中心对大模型的基本信息进行认证。

3. 第三阶段：AI 数据集流通与控制

1）AI 数据集提供方把需要交换的 AI 数据集上传到数据连接器；基于策略中心配置的“大模型数据使用控制策略”生成合约；发送 AI 数据集进行数据集流通。

2）AI 数据集消费方通过数据连接器对接收到的 AI 数据集进行存储；进行 AI 数据集流通控制，基于提供方数据控制需求的强弱，选择授权管控、双因子管控、GPU 管控三种方案中的一种；将 AI 数据集传输日志、大模型使用日志等上报清算中心进行留痕。

10.3.1 控制 AI 大模型的数据读取

将 AI 数据集放入数据连接器中，通过数据连接器基于授权管控方案实现控制 AI 大模型的数据读取，如图 10-5 所示。

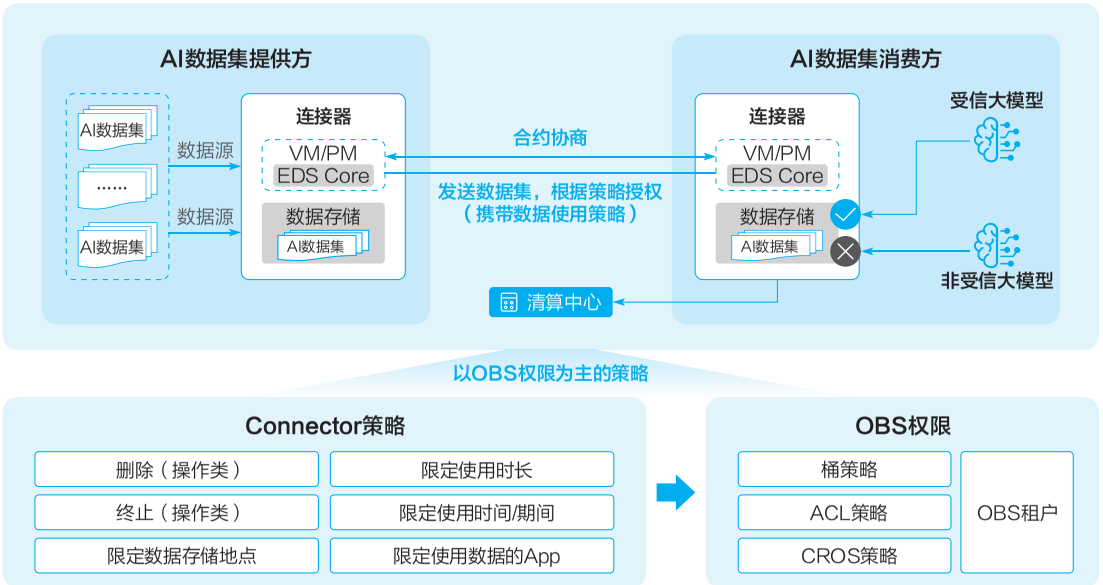


图 10-5　控制 AI 大模型数据读取的授权管控方案

首先，AI 数据集提供方把需要交换的 AI 数据集上传到数据连接器，基于策略中心配置的“大模型数据使用控制策略”生成合约，并发送 AI 数据集进行数据集流通。

其次，AI 数据集消费方通过数据连接器对接收到的 AI 数据集进行存储。

然后，数据存储服务（如 OBS）根据所属连接器中合约的数据使用控制策略，生成 AI 大模型访问数据集的权限。授权方式可以是按账号授权、按接口授权、按 Token 授权等。

此外，大模型基于授权访问 AI 数据集时，数据存储服务根据已授予的权限进行鉴权，鉴权通过即可访问 AI 数据集，鉴权不通过则无法访问数据集。

目前的使用控制策略基于 OBS 权限，主要支持限定使用时间 / 期间、限定使用数据的 App、限定数据存储地点、终止、删除等。

10.3.2 控制 AI 大模型的数据读取与加工

将 AI 数据集放入数据连接器中，通过数据连接器基于双因子管控方案实现控制 AI 大模型的数据读取与加工。双因子管控方案在授权管控方案的基础上，在大模型的数据存储服务中增加“策略评估模块”（EDS SDK），从而实现具备授权 + 策略评估模块的双因子管控，如图 10-6 所示。

策略评估模块基于使用场景可以选择在数据消费端决策、在提供端决策、在第三方策略评估服务中决策。目前，该模块增加了账号密码登录、加工资源回写、数据销毁、日志回写等 12 种标准 SDK 接口。

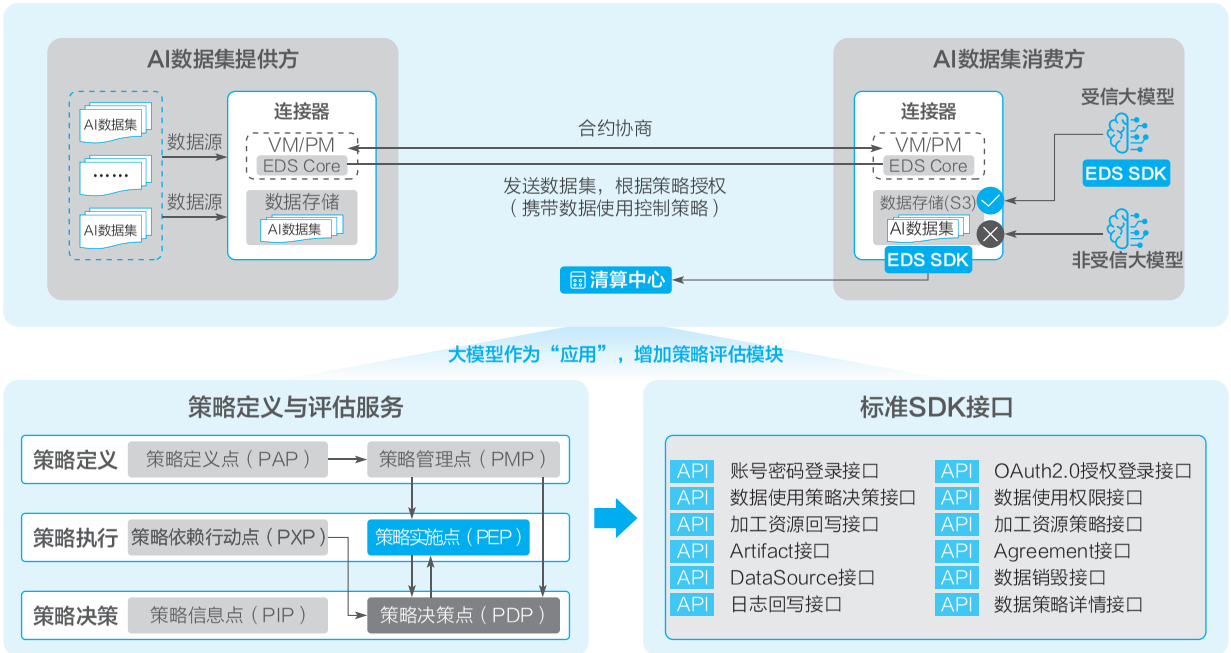


图 10-6　控制 AI 大模型读取与加工的双因子管控方案

授权管控与双因子管控这两种方案都是采用将 AI 数据集放入数据空间，将大模型放在数据空间外的方式。授权管控方案的优点在于不需要对大模型进行改造，能快速适配；缺点在于不能匹配所有的使用控制策略，且日志需要解析后才能回写。而双因子管控方案的优点在于可匹配 EDS 的所有使用控制策略，可对数据进行销毁，且支持实时日志回写，格式清晰；缺点在于需要对大模型进行改造。在大模型场景下，需基于 AI 数据集流通的具体需求，如成本、控制程度、两种方案的优缺点等选择适合的方案。

10.3.3 控制 AI 大模型的数据读取、加工与训练

授权管控方案属于应用级控制，即通过对应用程序或数据应用进行加强，确保数据在应用层面上的安全。双因子管控方案属于存储级控制。它们都是采用 AI 数据集在数据连接器内，AI 大模型在数据连接器外的方式。而当要对模型的数据读取、加工与训练的整个过程进行流通管控时，就需要把模型也放到数据连接器中，即前文提到的第三种方案“GPU 管控”。它是硬件级的控制，即通过对底层系统设备的安全性能进行强化，确保数据从操作系统到处理器的全栈安全。此外，还可以通过物理隔离和访问控制等手段，限制数据的访问范围，防止数据被非法获取和篡改。

多年以来，传统机密计算解决方案都基于 CPU 设计，基于硬件可信执行环境的“Trust+”级别的 EDS 解决方案，为使用中的数据提供机密性和完整性保护。而一直以来，对于 AI 等计算密集型场景中存在的性能问题，如何将机密计算从 CPU 扩展到 GPU，使其可在高带宽内存中全速处理数据，突破 AI 训练和推理工作的负载性能瓶颈，是解决问题的关键。

“Trust+”解决方案的可信根从软件下沉到硬件，构建了可

信计算的底座。结合虚拟机和虚拟化技术，信任边界从 CPU 扩展到 CPU+GPU，AI 模型 / 数据的机密性和完整性得到保护，如图 10-7 所示。

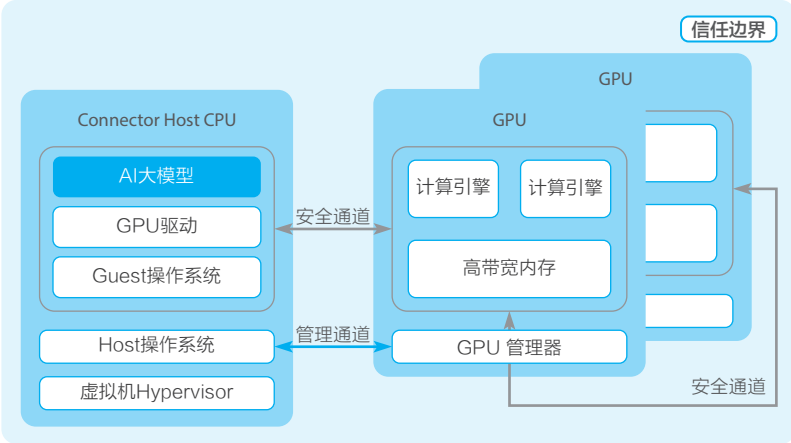


图 10-7　控制 AI 大模型数据读取、加工与计算的 GPU 管控方案

GPU 管控方案有三种不同级别的运行模式：

- 整个 GPU 独占单个虚拟机。
- 单个虚拟机同时分配多个 GPU。
- 利用虚拟化技术将单个 GPU 分割成多个实例，每个实例单独对应一个虚拟机，实现多租户计算。

在算力中心或云平台上进行大模型训练时，通过构建硬件级安全围栏，并采用集群化并行方式，可提升企业数据和大模型在训练过程中的安全性与可用性。将信任边界扩展到 GPU 后，在 CPU 中运行的代码可以安全地将计算和数据卸载到 GPU 中，有利于在最低级别的 GPU 体系结构中构建对数据和代码的安全性保护。

在垂直领域，当企业使用云服务训练行业大模型时，云厂商需要保障企业上传的基础模型和行业数据在微调训练时的数据安全，防止平台管理员超级用户、云平台其他租户或者恶意用户窃取数据。另外，对于大模型企业，在其私有化部署模型到客户环境中时，可通过构建安全围栏，确保其模型在客户环境中的安全和隐私，防止用户窥探和窃取模型信息。

在模型推理阶段，平台提供加密的数据输入和推理结果输出，可实现模型推理全流程的数据安全保护，使用户的提示词（Prompt）不会被推理服务提供方直接轻易获取。

10.4 数据空间支撑 AI 大模型的场景推演

由于高质量数据集的稀缺性，AI 数据集消费方需要通过 AI 数据集的流通，获取更多的优质数据来构建模型，而 AI 数据集提供方在这一过程中扮演着至关重要的角色。在数据流通的过程中，AI 数据集提供方的核心诉求是保护 AI 数据集的价值。保护 AI 数据集的价值就是保障 AI 数据集不被滥用、不被扩散、不被泄露。因此可从提供方与消费方的视角，探索数据空间支撑 AI 大模型的场景。场景一：AI 数据集提供方自建空间，自己控制数据被谁使用。场景二：AI 数据集消费方通过参与数据空间生态构建，获取更为广泛的 AI 数据集。

10.4.1 自建数据空间控制 AI 数据集流转使用

AI 数据集提供方自己建立一个数据空间，自主运营并控制自己的数据集给谁使用，从而在数据集可控交换的基础上实现价值最大化。AI 数据集提供方自建空间模式如图 10-8 所示。

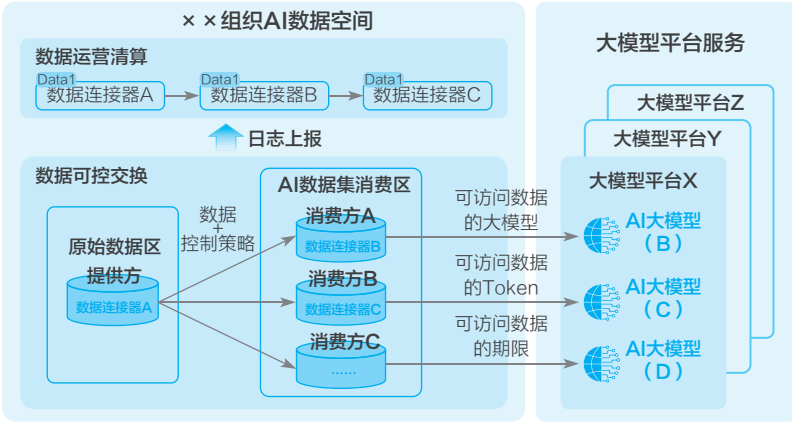


图 10-8　AI 数据集提供方自建空间模式

AI 数据集提供方首先自己建立一个数据空间，把需要交换的 AI 数据集上传到数据连接器中；然后基于策略中心配置的“大模型数据使用控制策略”生成合约；最后发送 AI 数据集进行数据集流通。

AI 数据集消费方首先向 AI 数据集提供方申请加入数据空间；然后 AI 数据集提供方给消费方设立一个数据连接器，并根据所属数据连接器中合约的数据使用控制策略，生成 AI 大模型访问数据集的权限，授权方式可以是按账号授权、按接口授权、按 Token 授权等；最后 AI 数据集消费方基于授权访问 AI 数据集。

AI 数据集提供方通过数据空间的三中心和数据连接器即可控制 AI 数据集的使用控制策略和安全防护。

10.4.2　参与数据空间生态获取更多的 AI 数据集

数据空间生态汇聚了各行各业的 AI 数据集，并通过数据市场

进行数据交换，使 AI 数据集充分地共享。AI 数据空间生态模式如图 10-9 所示。

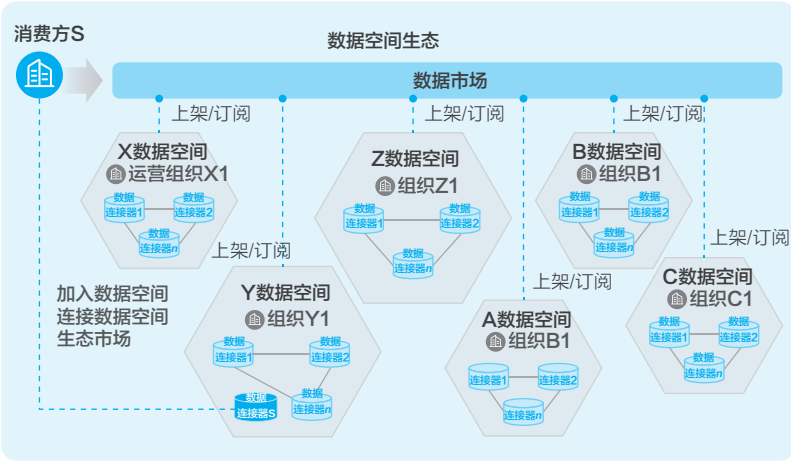


图 10-9　AI 数据空间生态模式

AI 数据集提供方通过数据市场创建独立的数据空间，并将需要交换的 AI 数据集上架到自己的数据空间中。此外，AI 数据集提供方也可以从统一的数据市场中获取所需数据集，并将其与自己的数据集进行汇聚。

AI 数据集消费方可以选择加入所需 AI 数据集的数据空间，也可以选择自主建立一个独立的数据空间，然后开通数据连接器，并注册需要进行 AI 数据集交换的大模型。从数据市场内获取到所需的 AI 数据集，AI 数据集消费方通过数据连接器与数据使用合约即可进行数据交换。AI 数据集提供方通过数据空间的三中心和数据连接器控制 AI 数据集的使用控制策略和安全防护。

10.5　小结

随着智能时代的到来，数据作为第五生产要素，其重要性日益凸显。在基于价值的流通体系内，AI 数据集被视为一种资产，其价值随着流通而增长。数据的流通促进了知识的共享和创新，不同企业之间可以通过数据交换获得新的视角和解决方案，推动整个 AI 生态的繁荣发展。但从数据自身的易复制、确权困难、非排他性、传播迅速、难以估值、异质性、协同性等特点看，数据的基于价值的流通体系构建面临着一系列独特的挑战。为了应对这些挑战，数据空间作为一种创新的解决方案被提出，它通过提供“可信、可控、可证”的环境，让 AI 数据集的买卖双方能够高效、安全地进行交易，从而对数据这种特殊的物品流通实现有效保障。

参考资料：

[1] 无忧知识星球．2024 年 AI 大模型训练数据白皮书：AI 时代的数据驱动力［EB/OL］.（2024-05-28）［2024-03-02］. https://www.aigcdaily.cn/news/a24e0soi4ms1rca/.

[2] JARED K，SAM M C，TOM H，et al. Scaling laws for neural language models［J］. arXiv，2020，arXiv：2001.08361.

[3] JORDAN H，SEBASTIAN B，ARTHUR M，et al.Training compute-optimal large language models［J］. arXiv，2022，arXiv：2203.15556.

[4] OpenAI. Revealing OpenAI’s plan to create AGI by 2027［EB/OL］.（2024-03-03）［2024-05-09］. https://yqli.tech/pdf/llm/q_star_info.pdf.

11

第 11 章 CHAPTER

数据空间未来展望

随着数据空间架构的不断完善，数据空间理念日臻成熟，近年来数据空间概念逐渐在国内普及，国内的数据空间建设工作正处于积极探索和启动试验的阶段。数据空间方案具备解决数据流通难题的潜力，引起了各方的广泛关注和高度重视。数据要素市场建设和公共数据开放的有序推进，对数据空间作为数据基础设施提出了迫切要求。各级政府高度重视数据空间的建设与发展，已经出台了一系列政策和规划，旨在推动数据空间在各行业的广泛应用，促进数据要素价值的释放。相关数据空间系统架构国际标准的发布，也为数据空间建设提供了标准的参考方案和互联互通的标准基础。

畅想未来，数据空间的建设，将形成一批有影响力的数据空间解决方案和最佳实践。参考欧盟在这方面的经验和进展，在政府

的规划和指导下，企业数据空间将成为企业实现跨组织协作和上下游协同的标准配置，成为衡量企业数字化水平的重要标志。各个行业则会通过数据空间的建设打破企业之间的信息壁垒，促进整个行业的数字化协同发展，从而提升其行业的整体竞争力。从城市到国家，从地区到全球，数据空间作为数据跨域流通的基础设施，将逐渐从点到面，互联互通，为全球经济合作和交流打开新的局面，推动全球经济的合作共赢。

11.1 数据空间将成为大型组织数据要素流通的标准配置

大型组织的上下游生态是一个复杂的系统。通常，大型集团企业通过构建高效的协作组织，推进业务流程的高效协同，促进数据等要素资源的共享共用，协同创新，提升制造水平。在上下游的多维协同和数字化转型中，依托新一代信息技术，特别是数据的交换和共享，提高数据分析预测、决策支撑、风险管控能力，提升供应链协同的效率和水平。

11.1.1 数据空间促进大型企业跨组织协作

以华为为例，其自身的业务分布在 170 多个国家和地区，需要全天候通过数据的高效连接、安全流通，实现复杂业务全球协同。在华为内部需要连接多个财务中心、供应中心、数百个国家中心仓、终端 EMS（电子制造服务）厂商、数千家体验店，以及多个研究所 / 院和上千个实验室；对外则需要连接更多数量的合作伙伴，如数以千计的合作高校 / 机构、人才联盟伙伴、解决方案开发伙伴以及数以万计的渠道销售伙伴等。

在大中小企业相互依存、相互促进的发展生态中，大型企业

往往是发挥引领作用的，它们需要构建以自身为中心的上下游生态。数据空间作为一个组织内部及其与外部环境进行数据交换和协作的基础设施，将成为大型组织数据要素流通的标准配置。它基于通用协议和安全机制，确保数据的可信共享和高效流通。

数据空间在大型组织中扮演着至关重要的角色，它提供安全、可控的数据共享机制，促进组织与合作伙伴间的数据交换。数据空间强化数据主权的概念，使数据提供者能够更精细地控制数据的使用范围和条件。数据空间将作为一种服务提供给上下游组织。上下游组织，通常是缺乏软件开发能力的中小型组织，可以根据实际需求获取数据空间服务，免去自建数据空间的成本和复杂性。

跨组织数据空间协作将是大型组织与上下游生态间实现数据共享和协同工作的关键，其具体应用场景包括但不限于供应链协同、合作伙伴关系管理、市场与客户洞察以及风险管理与合规性管理等。

11.1.2　数据空间提升上下游生态安全水平

大多数现代大型企业、政府等组织的组织架构非常复杂，既包括众多相互连接的区域、云服务以及其他基础设施，还包括与远程和移动环境的连接、物联网设备等非常规的 IT 连接。传统的方法要么通过防火墙等设施构建名义上的组织安全信任边界，要么让设备通过 VPN 进行安全连接，传统的 B2B 业务数据对接模式已经不能满足大型组织网络的环境复杂性带来的安全需求。木桶效应导致供应链上下游中的最薄弱环节很容易成为整个生态中的安全短板。

基于零信任理念的数据空间对人、终端和应用组件进行统一身份化管理，以访问主体的身份、网络环境、终端和组件状态等作

为认证的动态考量因素，确保接入网络的用户和组件持续可信。通过数据空间，组织将安全信任边界扩展到了整个供应链上下游。数据空间基本架构不要求所有参与者使用单一安全等级，但所有参与者必须满足最低安全要求，这样有利于保证同一数据能够在不同域中得到等级一致的保护。

不难想见，未来，每个大型组织都将有各自上下游生态的数据空间，数据空间将成为大型组织数据要素流通的标准配置。落实访问策略的数据空间将促进上下游供应链数据治理水平的一体化，大型组织的生态建设将得到技术的加持和制度的保障。

11.2 行业数据空间支撑开放协作解决行业挑战

与单一大型组织的上下游生态不同，行业数据空间将构建数据驱动的价值链，解决行业共同面对的挑战。行业数据空间以开放协作为原则，通过数据空间基础设施在一个行业内部促进数据流通、强化技术合作、提供数据与分析服务以及构建数据交换市场，共同解决行业所面临的共同问题，推动行业的数字化转型和创新发展，从而发挥“数据要素 ×”的增值力，并实现数据价值增长。

11.2.1 行业数据空间提升行业整体竞争力

行业数据空间的兴起有助于构建良好的数据生态，并平衡数据垄断与有效竞争的关系，促进数据流通和创新。同一行业的多家企业之间往往存在着复杂的竞合关系，行业数据空间的构建和运行将打破行业内的数据垄断态势，通过数据共享，将竞争转化为合作，提升行业的整体竞争力和服务水平。

行业数据空间鼓励价值链上的所有参与者共享数据，通过协同合作来解决行业面临的复杂问题，如供应链的脆弱性和日益严格的合规要求。通过制定统一的标准和规则，确保不同系统与组件之间的兼容性和互操作性，降低技术整合的复杂性和成本。通过开放数据生态系统，提高数据交换的透明度，建立参与者之间的信任关系，这对于数据驱动的决策和创新至关重要。行业数据空间特别强调数据主权和互操作性，这是构建可信赖和高效数据生态系统的基石。

行业内部成员间原本存在的不信任感，将通过数据空间的数据合约和使用控制等技术打破。行业数据空间的构建将推动数据的标准化、接口的兼容性以及参与方之间的信任机制建设，确保数据能够在不同组织间安全、高效地流通。得益于技术的发展和政策的支持，行业数据空间将成为推动产业数字化转型和提升产业链协同效率的重要力量。

11.2.2　欧盟行业数据空间布局初见雏形

汽车行业正经历着数字化转型，需要面对供应链管理、数据安全、环境法规遵从等行业挑战，行业内部企业之间存在合作的必要性和紧迫性，行业价值链上所有参与者之间的合作有利于共同找到解决方案，增强价值链。整个行业生态都认识到，面对这些挑战，没有单一实体能够独立应对并期望生存。

欧洲的行业数据空间 Catena-X 项目正是在汽车行业深刻变革的背景下发起的，项目的参与伙伴包括整车制造商、技术供应商和软件公司、汽车零部件供应商和研究机构等。Catena-X 项目的目标是创建一个开放的数据生态系统，通过数据链来提升汽车行业的竞争力、效率和可持续性。Catena-X 项目鼓励所有参与者在遵守

数据主权和互操作性原则的基础上共享数据，以促进创新和业务流程的优化。Catena-X 项目致力于制定统一的数据交换标准，以减少实施成本并缩短上市时间。Catena-X 项目还旨在通过数字化手段支持汽车行业的绿色转型，包括减少碳排放和促进循环经济。

欧盟已经规划了 17 类行业数据空间，这些是由《欧盟数据战略》或欧洲国家政府支持建设的，并且 IDS 产业生态主体基于发展需求自发探索形成了 46 个行业数据空间项目，部分行业数据空间如图 11-1 所示。此外，欧盟数据空间建设覆盖的行业领域在不断扩展，特别是在跨行业、制造、能源、供应链和交通等领域的数据空间数量较多。



图 11-1　欧盟规划多个行业数据空间

这些数据空间不仅是技术成果，更是协同成果，建设努力需要所有参与方就共同的标准和设计原则达成一致。通过这样的规划，欧盟旨在实现数据的有序流通和应用，推动单一数据市场的发展，并确保企业数据、公共数据、个人数据的安全和隐私保护。

11.3　数据空间成为国家数据要素流通的基础设施

城市已经发展成为国民经济和社会发展的核心载体。数百个城市作为一个个板块，共同拼出了国家级的大版图。城市作为数据要素资源流动的关键枢纽，对建设国家数据要素流通基础设施的大网络具有战略意义。而基于已有国家级网络和算力基础设施构建的四通八达的国家数据要素流通基础设施，则将我国的海量数据优势进一步转化为国家竞争的新优势。

11.3.1　城市数据空间激发数据成为发展新动力

随着城镇化驱动城市规模变大，需要提高城市数字化能力以提升管理、治理和服务效率。数字经济的发展更有利于提升营商环境，加快数字产业化发展，驱动城市数字化发展。经历多年数字化转型进程和智慧城市建设浪潮，城市已经沉淀了海量的数据，从政务数据扩展到公共数据，从传统的以文本、证照为主的政务办公扩展到涵盖物联、视频、遥感等多种模态的城市感知数据等等，它们分散在不同的主体中。

数据空间基础设施通过开放能力，能够打通城市全空间多要素多模态数据，推动智慧城市生态圈的融合发展，为城市提供后数字化时代的智慧城市解决方案。作为城市数据要素流通不可或缺的基础架构，数据空间是智能时代城市资源共享和应用的重要设施，将推动城市数据公共服务深入社会生产生活的各个方面。应围绕应用需求使数据有序流动起来，建设具有数据交换流通共享能力的智慧城市基础设施。

智慧城市的建设进入深水区，亟须数据空间基础设施赋能市、区、街以及产业园区的多级协同：市辖区作为城市主体的核心组成

部分，建设强调市 / 区协同配合，注重集约共享；县域聚焦本区域特点和发展诉求，强调自身的集约一体化建设；社区街道作为基层用户，既是城市治理最后一公里的使用者，也是很多城市治理数据产生的源头、智慧城市业务体验的构建点。未来基于数据空间的多级协同智慧城市将通过数据高效连接市、区、街，全面提升城市治理、服务、安全、效率的水平。产业园区综合服务于管理方、运营方、园区企业和园区公众，将得益于数据空间基础设施的数据连接和协同能力。

“数字中国”战略立足于全国各地数百城市的数字化，聚焦数据在城市治理、服务、安全、效率等领域的经济和社会价值。建设城市数据空间是践行顶层战略设计的重要举措。数字化和科技创新作为先进生产力，打破了城市空间的时空限制，城市智能化、共享化、信息化重构了时空关系，在传统地理空间的基础上形成了要素流动空间。

通过建设城市数据空间，最终释放城市数据价值，激活城市数据要素潜能，提升城市居民的服务能级，提高城市治理精细化水平，打造城市数字经济创新发展高地，为赋能整体城市数字化转型奠定坚实基础。

11.3.2 海量数据优势转化为国家竞争新优势

目前针对各类生产要素，国家基本上已经构建了全国范围的流通机制，并建设了相应的基础设施，以承载其在全国范围内的流通，发挥生产要素的经济和社会价值。比如：可以以转包、出租、互换、转让、入股等方式转移农村土地的承包经营权；通过在人才市场上以签约、外包、服务等方式实现人力资源流动，从而为企业等组织提供自雇员工之外的多种劳动力资源；在专利事务所等机构

签订的技术转让合同中，通过明确是否允许将该项技术转让给第三方，就可以实现专利等技术的流通和转让；作为一般等价物的资金，有非常健全而庞大的金融机构来实现价值流通，甚至通过汇兑来跨越币种边界。

数据要素的流通同样离不开基础设施。数据要素基础设施的建设对于推动数字经济的发展具有决定性作用。它不仅促进了数据资源的有效整合和利用，还为数据的流通和交易提供了基础支撑，保障了数据的安全和隐私。随着数字经济的快速发展，数据要素基础设施已成为国家竞争力的重要标志。

全国连通的可信数据流通基础设施自下而上依赖已有的国家级基础设施：互联网络基础设施层、算力与存储资源层。前者提供网络的连接和数据传输的基础服务，构成了数据流通的物理基础和通信支撑；后者负责数据的计算和存储，确保数据资源的可访问，为数据的初步处理提供支持。在这两层之上的，是对于数据的交换流通来说不可或缺并且发挥关键作用的数据空间基础设施。数据空间与其他结构层共同构成了一个跨区域、跨行业、跨主体的数据流通基础设施，通过各层的协同运作，实现数据的安全、高效和合规流通，促进数据要素市场的全面发展。

数据空间作为国家数据要素基础设施的重要组成部分，是在网络、算力等设施的支持下，面向用户提供数据流通交换服务的新型基础设施。它是覆盖硬件、软件、开源协议、标准规范、机制设计等的有机整体，是数据要素价值释放的关键。

随着数据要素从市区县级到省级，再到国家级，数据流通大联网形态将会应运而生。搭建省际数据高速、区域数据高速、国家数据高速“大动脉”，为数据要素供给、流通和交易提供跨城市、

跨区域的四通八达的路网基础，超越行政区划边界，加快内外融合大循环运转。

数据空间不仅是技术层面的需求，更是推动数字经济发展的战略性资源。它为数据要素市场化配置改革提供了重要支撑，有利于推进数据的全国市场化流通交易，有利于挖掘和释放数据要素的潜能，有利于构建以数据为关键要素的数字经济，将我国以市场优势积累的海量数据转化为国家竞争的新优势。

11.4 数据空间成为全球数据流通的“丝绸之路”

两千多年前，亚欧大陆上勤劳勇敢的人民，探索出多条连接亚、欧、非几大文明的贸易和人文交流的通路，历史学家将其统称为“丝绸之路”。千百年来，“和平合作、开放包容、互学互鉴、互利共赢”的丝绸之路精神薪火相传，推动了人类文明的进步，是促进沿线各国繁荣发展的重要纽带，是东西方交流合作的象征，是世界各国共有的历史文化遗产。进入新世纪，在以“和平、发展、合作、共赢”为主题的新时代里，传承和弘扬丝绸之路精神显得更为重要和珍贵。

如何促进经济要素有序自由流动、资源高效配置和市场深度融合，如何建立和加强信息世界的互联互通关系，构建全方位、多层次、复合型的数据互联互通网络，是我们建设全球数据流通“丝绸之路”时需要关注的大背景和需要深入思考的时代课题。

根据联合国《2021 年数字经济报告》，随着跨境数据流动量的增加，参与数字经济的各大经济和地缘政治主体都试图将数据治理纳入有利于自身的国际贸易规则。我们有充分的理由对数据和跨境

数据流动构建具有全球视野的治理框架，以便在不断扩大的数据生态系统中始终发挥领先者地位。

11.4.1　跨境数据流动推动全球协作共赢

随着新一代信息技术革命的持续推进，全球数据领域正从消费互联网时代向价值互联网时代转变，探其本质，在于数据连接规则的变革。泛在连接能力极大拓宽了数据连接范围，提升了资源量级，价值创造来源从个人数据演变为产业数据。根据世界贸易组织（WTO）公布的数据，2023 年全球数字服务贸易规模达 4.25 万亿美元，其中数据跨境流动是核心驱动力。根据上海数据交易所的研究报告，预计 2030 年全球数据交易市场规模将达到 3011 亿美元，年复合增长率为 13.2% 。

各国陆续达成各种双边、多边形式的新数据流通协议，各国法律和主要多边协议在数据跨境方面体现不同司法管辖权的要求。

2023 年 7 月 10 日，欧盟与美国达成了新的数据隐私框架协议，以促进跨大西洋数据流动。2023 年 10 月 21 日，英国和美国建立了数据桥。2023 年 10 月 28 日，欧盟和日本达成了一项跨境数据流动协议。

中国提出了《全球数据安全倡议》，强调各国应尊重他国主权、司法管辖权及数据安全管理权，共同维护全球数据流动的安全与秩序，不仅关注保障本国数据安全，也尊重其他国家在数据跨境流动方面的相关规则。2024 年 6 月 26 日，中国与德国签署了《关于中德数据跨境流动合作的谅解备忘录》。

除多边协议外，全球各国在数据本地化、跨境转移方面陆续

出台不同的法律法规。随着 AI 大模型的迅猛发展，未来数据全球竞争必将越发激烈。

正如在《超级版图：全球供应链、超级城市与新商业文明的崛起》一书中，作者帕拉格·康纳所言，在能源、交通和通信等方面实现互联互通后的世界图景，将使得基础设施的重要程度超过传统的国境线。客观反映世界的地图强调的是行政区划的区隔，油气管道、高速铁路、通信光缆正象征着全球化的网络文明，数据空间所带来的全球互联互通，正是这一范式在数据领域的体现。

传统地缘政治较量的着眼点都是独占之争，主要表现为在全球供应链和市场上对土地、劳动力、资本等生产要素资源的独占式争夺，而随着数据空间在全球数据供应链上的普遍展开，各国将更多着眼于制订有效的数据使用策略，享受数据共享带来的帕累托效应，在合作共赢的数据丝绸之路上越走越远。

11.4.2 数据交换标准护航全球数据流通

建立在统一的数据治理、共享的规则和标准基础之上的信任机制是数据交换的基础。通过全面的身份管理，确定参与者的身份，并对所有参与者进行组织评估和认证，以确保他们彼此之间的信任。

分布式的数据空间一定要有一个标准化的框架。清算的标准、使用控制的标准、资源的描述、接入认证等都需要互认、支持互操作，在底层是能够互认互通的。

数据空间除了基于现有安全措施的架构规范之外，还需要对各个参与方组成部分进行评估和认证。根据确保数据主权的核心要求，数据所有者在将其数据传输给数据消费者之前，会为其数据附

上使用限制信息。只有完全接受数据所有者的使用策略，数据消费者才能使用这些数据。

数据空间要求将数据源和数据作为资产进行全面描述，为数据加载特定领域的词汇表，并实现全面实时的数据搜索。

未来在产业标准层面将广泛实现：

- 就数字和数据相关权利及原则达成共识。
- 就数据交换的重要概念的定义达成共识。
- 规定实体间相互访问数据的允许条件。
- 加强对数据价值和跨实体数据流量的测度。
- 将开放数据视为一种公共品。
- 探讨新兴的数据治理方式的标准化。
- 制定数据交换方面的国际 / 国家标准。

11.5　小结

在这个数据驱动的时代，数据空间如同宇宙星系，无边无际。每一个数据连接器，就像是星系中的一颗星星，它们汇聚成银河，再组成星团，形成庞大的数据星系。在这个过程中，数据空间的维度不断扩展，从单一到多维。

在多级数据空间所构成的瀚海星云中，大型企业等组织就如同一颗颗行星，它们的供应链上下游就像卫星一样环绕在它们的周围，形成了一个个组织生态，在深远的广袤宇宙空间中，不断交换着信息和能量。每一个数据连接器，犹如一粒粒星星，它们虽小，却蕴含着无限的应用可能性。这些星星是构成数据宇宙的基本单元，它们独自存在时，或许显得微不足道，但当它们以某种规律

成对运动形成双星结构，组合成星状结构或者网状结构时，便能形成行业或产业级的璀璨星座，信息流动的光芒将照亮整个数据宇宙。

企业数据空间如同自带卫星的一颗颗行星，通过无形的引力相互作用，汇聚成太阳系一般的行业数据空间；城市数据空间，如同银河系中的一颗颗恒星，通过能量交换相互连接，构成了更为复杂的星系结构，形成国家数据空间。国家级数据空间的融合发展，更是孕育形成国际化的数据空间，如同星系到超星系团的演化。这些组合遵循着某种数据流动的能量规律，等待着更大规模的星际运动，数据空间中存在的未知数据融合模式将带来智慧的大爆发。

作者代表简介



马运

（MA YUN）

华为公司数据总架构师、数据首席专家、信息架构专家委员会主任、数据管理教研室主任。

2015年至2022年期间，在华为公司担任数据管理部部长、质量与流程IT首席数据官，负责公司信息架构、数据底座和数据治理体系建设，主持公司数据资产管理、元数据驱动产品信息管理、主权可控数据交换等变革项目，深度参与数字化转型工作。畅销书《华为数据之道》第一作者，《华为数字化转型之道》主要作者。

曾在瑞典爱立信总部工作多年，担任产品线业务总监、集团信息管理部总监。曾任北京交通大学副教授、统计教研室主任、经济系主任，在瑞典国家公路与运输研究院任客座研究员。

当前研究领域：企业AI数据管理、AI+区块链、数据要素价值量化。

华为在数据空间领域的探索与实践，为破除数据孤岛、保障数据主权可控、推动数据跨业务跨机构跨行业共享提供了宝贵经验。深圳正积极探索基于可信数据空间的新型数据共享开放模式，推动公共数据在服务政府数字化转型的同时，也能更好地服务企业和社会。此书呈现的“可信、可控、可证”体系，对政府与企业突破数据流通壁垒、建立内外贯通协同的数据资源开发利用体系具有非常现实的借鉴意义，值得企业和政府部门研读。

——王耀文 深圳市政务服务和数据管理局副局长、二级巡视员

《数据空间探索与实践》是一本不可多得的实用指南，为数据治理与流通提供了宝贵经验和系统方法。我们与华为携手建设的“南通家纺可信数据空间”，正是书中理论在区域特色经济中的生动实践，有力推动了南通家纺产业集群的数字化转型，让数据成为驱动产业升级的新动能。本书为探索数据要素化的企业与政府提供了清晰指引，我们从中受益匪浅，也期待更多伙伴投身数据空间建设，共创数字经济新篇章。

——郭攀 南通高新控股集团董事长

作为数据管理领域的实践者，华为的数据空间探索展现了从技术到生态的完整闭环。书中提到的“可信、可控、可证” 体系，与 DAMA 数据管理框架中数据安全与治理的核心理念一脉相承。该体系在跨境数据转移（如满足欧盟 GDPR 要求）、财务外审等场景的落地，以及数据连接器的分层解耦架构设计，为企业数据管理从合规走向价值创造提供了可复用的方法论，值得数据管理从业者深入学习。

——汪广盛 国际数据管理协会（DAMA）大中华区主席

华为在数据空间领域的系统性探索兼具国际前沿的理论深度与国内落地的实践价值。书中系统阐述的数据治理三阶段（连接集成、汇聚、交换智能化），与学术研究中数据要素流通的演进逻辑高度契合。华为提出的跨主体数据流通主权保护理论方案及落地实践，为破解数据确权难、跨主体流动信任机制缺失等全球性学术瓶颈问题提供了方法论级借鉴方案。对于研究数据空间架构与数据价值释放机制的学者而言，此书既是对“知行合一”认识论的工程化复现，更是理论联系实践的绝佳范本。

——彭国超 中山大学信息管理学院教授、大数据研究院副院长

可信数据空间为数字世界中的各类参与者提供了精准管控信息分享边界的能力，它让数据蕴含的价值在有序流转中得到更充分的释放。这一作用，在数字经济加速发展的当下，显得尤为关键。作为数据空间的贡献者、建设者与实践者，华为将自身在该领域的经验与思考凝结成书，用平实易懂的语言系统解答了关于数据空间“为什么、是什么、怎么建、怎么用”的问题，为读者提供了清晰而务实的指引。

——赵志海 中国信息通信研究院信息化与工业化融合研究所大数据高级总监

上架指导：计算机/大数据
ISBN 978-7-111-79619-0
9 787111 796190
定价：99.00元

客服电话：(010) 88361066 88379833 68326294