



用数据创造价值 以智能激发增长

数据科学基础平台白皮书

 | **PERCENT 百分点**

目 录

IDC观点	01
政企数字化转型进入纵深阶段,数据科学加速数字化能力进阶	01
基于数据科学的全流程能力演变和生态协同是数据智能发展进步的必然趋势	01
构建数据科学基础平台体系,推动一体化数据智能研究与实践	01
第一章	
数据科学驱动行业数字化进阶,打造政企核心竞争力	02
1.1 政企数字化转型进入纵深阶段	02
1.2 打造为政企全面赋能的数据智能核心	03
1.3 以数据科学加速政企数字化能力进阶	04
第二章	
行业数据/信息转型现状与演变趋势	06
2.1需求侧 - 行业数据应用持续向深度和广度延伸	06
2.2供应侧 - 基于数据科学的能力演变和生态协同	10
第三章	
构建数据科学基础平台体系,推动一体化数据智能实践	12
3.1 建设一体化的治理平台底座,打造高质量数据资产	13
3.2 兼容高速演进的先进工具,提供高效洞察与决策支撑	13
3.3 促进应用生态繁荣发展,与行业实现深度融合	14
3.4 形成可持续迭代的数据技术框架,综合提升全流程响应能力	15
3.5 小结:以数据科学基础平台体系应对数字化转型多维度难题	17

第四章

IDC建议:政企机构数据智能全流程实践关键举措 **19**

- 4.1 将数据科学能力纳入未来发展战略 19
- 4.2 规避单点方案和开源技术的成本陷阱, 选用可持续迭代的全栈能力平台 19
- 4.3 明确数据价值转化流程, 集结跨领域人才团队, 激发应用创新活力 20
- 4.4 量化跟踪数据驱动决策和业务的收益 21
- 4.5以全面数据价值转化为目标, 寻求敏捷高效的技术路径 21

第五章

行业实践:百分点科技的数据科学平台 **22**

- 5.1 数据科学基础平台的愿景和价值主张 22
- 5.2数据科学基础平台的含义和创新 24
- 5.3 平台应用实践 26

第六章

未来展望 **32**

- 6.1 机遇与挑战 32
- 6.2 未来展望 33

结语 **35**

IDC观点



政企数字化转型进入纵深阶段, 数据科学加速数字化能力进阶

在数字经济持续高速增长的大背景下, 一大批政企机构开始进入数字化转型的纵深阶段, 努力实现由数据/信息驱动的全面业务整合和创新。IDC认为, 各机构应全面提升数据科学理论水平, 深度掌握数据应用技术、产品、方案和交付方法, 围绕更大范围、更高效率、更快响应和更深融合四个维度目标, 构建完善的数据价值转化全流程能力体系。即以数据科学为指导建立涵盖“治理平台+分析工具+应用运营”全栈能力的数据智能核心, 打通“感知 - 数据 - 信息 - 知识/洞察 - 决策 - 行动/应用 - 再感知”的价值链条, 形成内部认知能力和外部生态应变能力的闭环。



基于数据科学的全流程能力演变和生态协同是数据智能发展进步的必然趋势

基于数据科学的全流程效能提升, 将围绕能力演变、架构演变和生态发展等维度深入展开。其中, 能力体系将趋向全栈化, 构建“倒三角”形态的全栈支撑体系, 包括以强调“统一治理”的平台化能力逐步消除数据孤岛, 以趋向智能化的工具能力为业务端提供可灵活定制的洞察和决策服务, 以高效的运营能力开发/部署与业务场景深度融合的数据应用; 架构体系则力求持续迈向云原生化和全栈智能化。此外, 各厂商也将围绕数据价值转化周期推动专业化分工, 建立生态协同机制, 创新商业模式, 与政企用户实现共赢。



构建数据科学基础平台体系, 推动一体化数据智能研究与实践

各政企机构有必要构建一个拥有数据全流程能力的重要载体, 即一个一体化的数据科学基础平台体系, 通过广泛、开放的数据协同, 为各业务方提供端到端的完备数据智能体验。要通过全栈化能力体系下的治理平台、分析工具和智能应用建设, 以及整体技术架构的转型升级, 构建端到端的一体化解决方案, 与行业发展实现深度融合, 形成可持续迭代的技术框架, 强化关键环节技术能力, 有效解决各机构的战略、组织、流程以及成本投入方面的诸多问题。

01

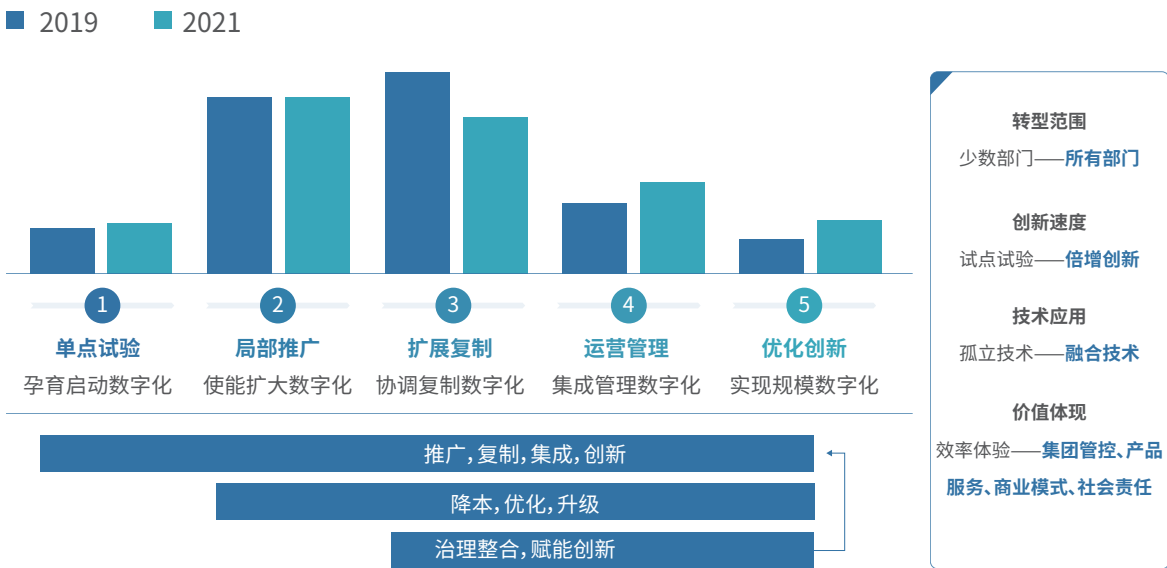
第一章

数据科学驱动行业数字化进阶, 打造政企核心竞争力

1.1 政企数字化转型进入纵深阶段

近年来, 数字经济持续高速增长。IDC数据显示, 到2022年, 全球65%的GDP将由数字化推动; 2022年到2023年, 数字化转型的直接投资将超过6.8万亿美元。IDC中国企业数字化转型成熟度研究显示, 一大批企业已经进入到数字化转型的纵深阶段, 即利用集成管理数字化推动常态化运营管理, 利用规模数字化实现企业优化创新。

图1 中国企业数字化转型成熟度 2019 vs 2021



来源: IDC, 2022

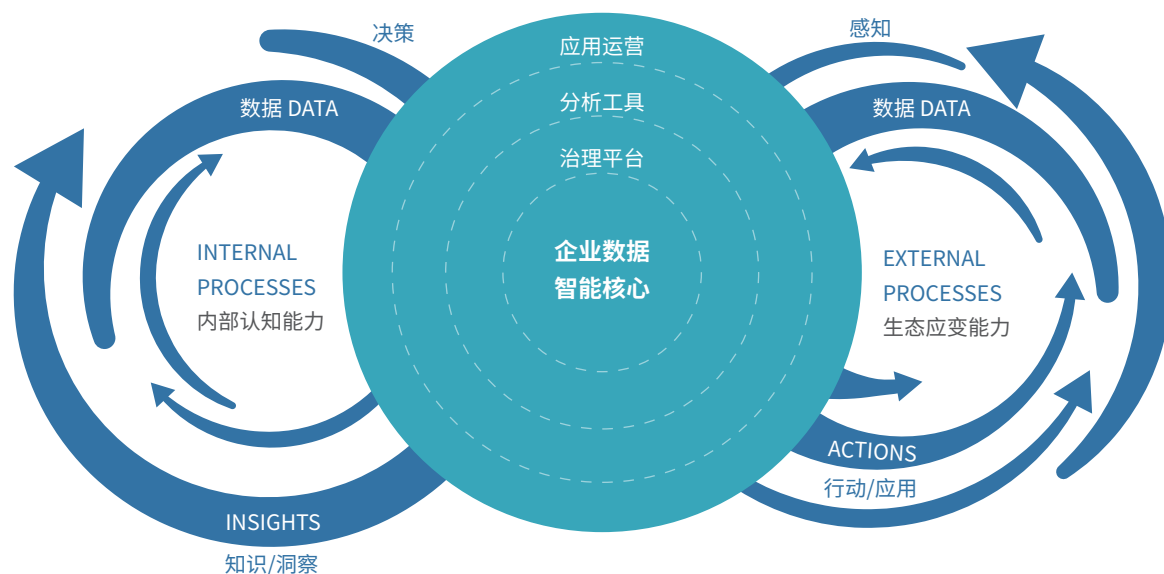
数字化转型涵盖三个基本组成: 基础设施转型、业务应用转型、数据/信息转型。其中, 业务应用转型为主动性项目, 由运营目标驱动; 基础设施转型为被动性项目, 由成本驱动, 是支撑业务应用创新和数据资产优化的底座; 而数据/信息转型则兼有主动性和被动性特征。对于进入数字化转型的较高阶段(第4和第5阶段)的组织来说, 进一步推进数字化能力的关键在于实现由数据/信息驱动的全面业务整合和创新。

- 数据/信息转型的被动性:**随着政企数字化转型的深入,日趋活跃的创新型业务引发了数据的爆发式增长,例如数字化原生企业在运营过程中通过全流程数字化体系产生的大量交易数据、用户数据、经营数据和业务运行监测数据,以及传统行业在大力推进物联网、工业互联网、智能制造过程中新增的传感器监测数据、数字化生产过程数据、产品运行数据等。政企组织必须正视不断增加的数据管理难度——包括安全、合规、存储成本、调用性能等。
- 数据/信息转型的主动性:**政企运行数据蕴含巨大的价值,承载丰富的业务属性,是保障组织动态发展的血液。对于已具备足够的数字化基础,并进入转型更高阶段的政企机构来说,明确的数据战略必不可少。管理者需要着眼于“data-driven-everything”即“全面数据赋能”,将数据战略与整体发展战略、组织、业务、技术、运营管理等多领域加强关联,以数据价值转化技术推动组织业务的全方位进步,以构建精细化的业务洞察能力和高效精准的决策能力,大幅提升组织运营的自动化和协同化水平,创造出与行业知识、行业实践融合创新的更多可能。

1.2 打造为政企全面赋能的数据智能核心

IDC认为,涵盖治理平台+分析工具+应用运营全栈能力的企业数据智能核心,是打通“感知 - 数据 - 信息 - 知识/洞察 - 决策 - 行动/应用 - 再感知”的价值链条并形成左右互联闭环的关键。

图2 IDC 企业数据智能研究模型



来源:IDC, 2022

数据智能核心涵盖以下三个层面的能力：

- **以强调“统一治理”的平台能力**，逐步消除数据孤岛，按照业务需求对数据进行主题化以及场景化建模，形成高质量、易调用的数据资产，实现“感知 - 数据 - 信息”的转化；
- **以趋向智能化的工具能力**，为业务端用户提供可灵活定制的洞察和决策服务，提升团队效率和创新能力，实现“信息 - 知识/洞察”的转化；
- **以高效的运营能力**，开发/部署与业务场景深度融合的数据应用，建立趋向数字原生化的业务体系，实现“知识/洞察 - 决策-行动/应用-再感知”的转化。

强化数据智能核心的目的，是持续提升数据价值转化能力，而与数据价值转化全流程有关的理念、目标、方法论和实践过程，共同构成了数据科学的研究框架：

数据科学是融合了数学、统计学、计算机科学、人工智能等多个领域的交叉学科，旨在研究如何从数据中发现规律，丰富认知，挖掘价值；

数据科学的研究范围涵盖了数据智能核心的全部三个能力层面，有效贯通“感知 - 数据 - 信息 - 知识/洞察 - 决策 - 行动/应用 - 再感知”的闭环价值转化链条：

- 数据治理能力包含数据标注、数据清洗、数据整合、数据目录等，是数据科学的基石；
- 数据分析能力包含描述性分析、预测性分析、诊断性分析、规范性分析等，是数据科学的内核；
- 数据应用能力包含嵌入式应用、驱动式应用等，是数据科学的服务性呈现。

1.3 以数据科学加速政企数字化能力进阶

随着数字经济的蓬勃发展，政企现有的数字化业务手段将逐渐在竞争中失去差异性。从面向未来发展的视角看，各类机构的数据智能能力将成为其数据/信息转型的重中之重，也是其实现可持续发展的关键竞争力。而强化数据智能核心的关键举措，是遵循数据科学理论，围绕更大范围、更高效率、更快响应和更深融合四个目标维度，持续提升数据价值转化能力。

- **更大范围:**整合更多有效数据源,支持更多业务场景,部署更多创新型应用。

根据IDC Global DataSphere(专门记录全球数据增长的产品)显示,2020年全球创造了59ZB(Zettabyte, 泽字节。1ZB=1万亿GB)的数据;2025年,全球新产生数据量将高达180ZB,数据已经无处不在。

- **更高效率:**借助更高级的数据挖掘手段,实现自适应/自学习的价值传递闭环。
- **更快响应:**采用更先进的架构,在“感知-数据-信息-知识/洞察-决策-行动/应用-再感知”价值转化周期中实现流程化、管道化、规模化。
- **更深融合:**与领域知识不断融合,重新定义技能和业务边界,不断发现创新场景。

良好的数据智能核心可赋予政企自我优化和自我进化的能力优势,帮助政企更加清晰地洞察内外部形势,不断完善知识体系,高效协同供应链生态,在激烈的市场竞争环境中保持活力和韧性。

02 第二章

行业数据/信息转型现状与演变趋势

在当前全社会广泛拥抱大数据的时代背景下,围绕数据科学领域的技术进步与业务创新活动异常活跃,各类政企对全面提升数据理论水平,深度掌握数据应用技术、产品、方案和交付方法都有十分强烈的诉求:

- **全方位提升数据应用能力:**重点关注数据资源的一体化管理和敏捷化使用,建立统一的数据开放与全时空服务体系,持续实现数据实践领域的的能力沉淀。在此基础上,各机构期望能够加强组织与人才建设,统筹数据治理与运营过程,合理配置资源,在保证安全合规的前提下,加速数据价值的挖掘与创新。
- **长期、稳定地服务全量用户:**政企希望在控制总体投入规模的前提下,通过合理地规划设计,建立统一、稳定、可持续迭代的数据服务平台,不断沉淀数据资产,为自身开发、测试、交付、运营以及各类型客户、生态、社会公众等提供广域、多维的数据服务,通过数据应用创新实现充分的内、外部协同。
- **深化数据智能与业务的融合:**政企在深化数据智能的过程中,不断汲取行业发展经验,赋予数据以更深的业务属性,将大数据和人工智能领域的最新发展成果引入业务创新的过程中。通过长期地迭代与提升,形成稳定的智能化洞察与决策能力,全方位提升综合竞争能力。

2.1 需求侧 - 行业数据应用持续向深度和广度延伸

政企力求利用数据破除条线壁垒,依托统一的数据资源体系构建面向全域业务的一体化数据服务能力。在这个过程中,政企需要持续深化数据智能的应用能力与范围,将更多的监测、洞察和决策需求通过智能化过程加以实现。随着数据智能的进一步推广,围绕数据价值挖掘与释放过程的专业化分工趋势也将越来越显著,更多的深度治理和算法创新工作有望全面生态化,由更加专业化的团体通过分工与协作完成。

各行业在利用数据赋能业务发展的过程中,普遍经历从单点技术、孤立场景向一体化、融合化迈进的必由路径。IDC的一项调查显示,伴随数据应用的逐渐深入,政企机构普遍遇到多个维度的棘手问题。

战略与愿景：方法论缺失，对数据价值的认知不到位。深化数据应用的过程与政企数字化转型战略、组织建设、技术发展以及人才培育等因素息息相关，需要通过科学的方法指引，建立体系化的目标和发展路径。许多政企机构缺乏对数据价值的深入认知，也没有建立体系化的发展路径和方法论，数据应用仅停留在IT部门的日常工作计划中，甚至成为运维工作的沉重负担，这无疑会使其在数字化时代与行业领先者产生新的发展鸿沟，进而加大与行业竞争对手的差距。

在IDC的调研中，“我们缺少方法论指导来开展并落实数字化和智能化转型的相关工作”是受访企业在智能化方面面临的第二大难题，81%以上的受访者将这一项的难度评分评为8分以上。（1-10分代表难度由低到高）

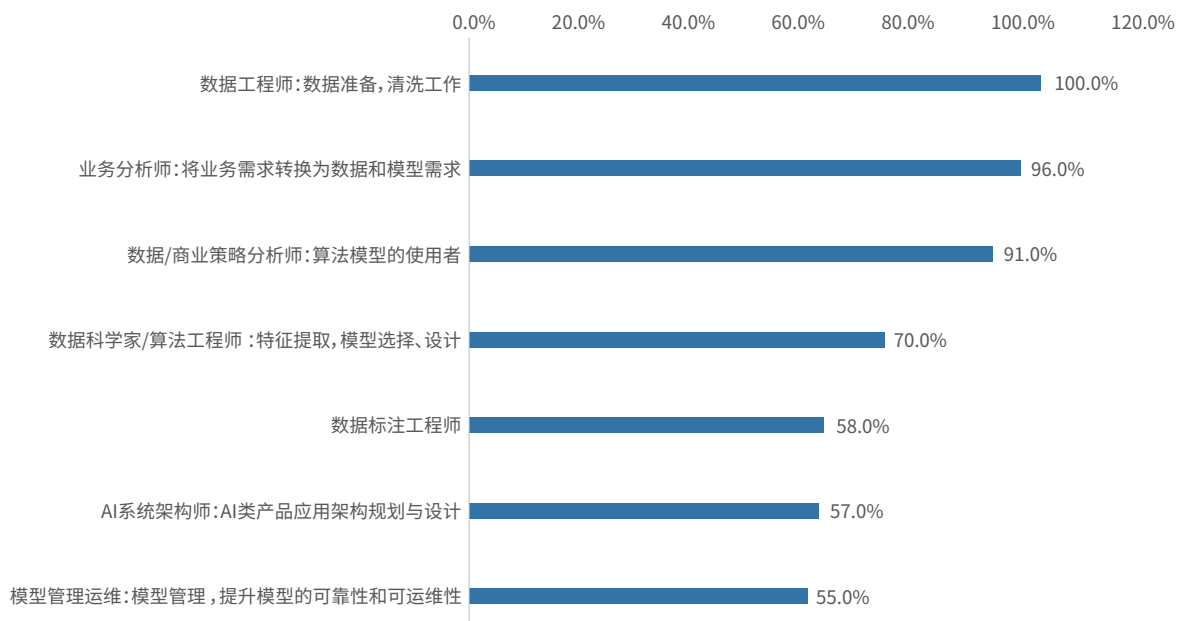
过程：缺少端到端的一体化能力，业务侧收效不明显。一些政企的数据应用始终无法深入，仅实现了基本的统计分析、展示以及通用化的结果呈报，数据应用与业务价值相去甚远，更无法满足大量创新业务的需要，这是受多方因素制约的结果。例如，数据平台建设缺乏规划，投入不足，导致基础能力薄弱，资源总量、稳定性和灵活性均严重不足；数据治理能力低下，相关的大数据系统功能简单、部署分散、状态孤立，无法面向业务形成高价值的服务，这将严重影响机构的创新能力与效率。

在IDC的调研中，“缺乏端到端的工具，数据应用和业务价值离得太远”成为第一大难题，63%以上的受访者将这一项的难度评分评为8分以上。（1-10分代表难度由低到高）

技术：工具平台复杂纷乱，难于整合与升级迭代。当前的数据服务市场呈现出百花齐放的发展格局，不同技术平台各具特色，产品水平和能力不一。多数传统政企机构囿于自身有限的技术水平，无法将外部先进的大数据技术与经验进行有效整合，进而应用于整体发展。一些企业还经历了弯路，不仅造成严重的资源浪费，还耽误了宝贵的升级时机。

组织与人才：基础现状薄弱，数据质量差，人才匮乏。很多机构的信息化和数字化基础仍十分薄弱，表现为应用系统老旧、数据质量差、投入无序等。更重要的是，一些传统政企单位缺少数据科学的人才，也难以组织起专业化的团队力量，尤其是业务与技术交叉的算法工程师、数据科学家、AI系统架构师等岗位，呈现出能力的真空状态，无法适应当前快速发展的数字化趋势。IDC对一些数据应用较为深入的机构进行了调研，发现与基础治理相关的数据工程师、业务分析师等岗位相对健全，但与智能化相关的数据科学、算法、模型等岗位依然有较多欠缺。

图3 中国企业在组织与人才方面准备概况



来源:IDC 中国2022年数据科学调研, n=100

IDC持续研究了不同行业的数据/信息转型状态,并围绕现状、愿景和挑战做了较为深入的分析,以下重点呈现政府、金融、制造、零售四个典型行业的相关情况。

政府行业:已经具备一定的数据资产管理能力

政务信息化已具备深厚基础,政务大数据正迈向全面融合与智能,城市大脑、一网统管、一网通办等一批跨部门、跨行业、跨领域的创新应用体现了充分的数据融合与成果共享,疫情防控等突发场景也进一步推升了数据智能的应用成效。政府行业在“数据-信息”和“知识/洞察-决策-行动/应用”等环节仍有较大进步空间,其对数据价值认知度较高,但深化数据应用面临一定的管理条线壁垒。政府应用领域的外部生态多样,内部跨部门协作的复杂性、合规性需求较高。同时,受内部流程的谨慎性制约,大部分应用的智能化程度不高,发展较为缓慢。政府行业应在扩展数据驱动应用的广度以及改善数据挖掘效率的同时,更加重视数据价值链条各环节的响应度,深化数据与业务领域知识的融合度。

金融行业:具备领先的数字化基础和数据资产管理能力

在安全性和合规性的要求下，金融行业依托深厚的数字化基础，在互联网业务拓展、风控、监管、智能检索等领域对数据智能进行了深入实践，实现了一大批场景化应用，也切实取得了丰富的业务成果。金融行业新业务拓展迅速，竞争激烈，数据价值有待进一步深入挖掘。金融业务创新面临严格的监管限制，合规性需求较高。金融IT基础架构仍处在向云化转型的阶段，需要持续关注基础架构与业务的协同转型与升级。

2022年IDC金融行业预测：为了提高全球弹性和规模，到2025年，30%的全球顶级银行将在多行业生态系统环境中共享数据、应用程序和运营。到2027年，由于更复杂的人工智能/机器学习(AI/ML)和深度学习算法的成熟，通过开放银行资源利用现有数据，10%的身份欺诈企图将被减少。



制造行业：高度异构的应用和数据环境

当前制造行业对数据资产管理能力的诉求，主要体现在利用物联网IoT技术加强过程数据采集，通过实时和离线数据分析提升生产和供应链过程的观测能力等方面，努力实现预测性维护、流程优化、节能减排、供应链优化等业务成效。制造行业数字化转型理念相对保守，数字化应用具有较高的生态联动性，有高度异构的应用和数据环境，需要扩展数据驱动应用的广度，并更多地兼顾“体验”类和“决策”类应用场景。

2022年IDC制造业预测：由于客户的期望值不断提高，到2026年，65%的G2000原始设备制造商将实时挖掘服务互动和工单数据，为产品开发决策提供信息并实现质量目标。到2025年，40%的制造商将部署基于人工智能的企业工具，以支持决策过程和最大化数据价值，从而使收入/利润提高5%。



零售行业：积极推进数字资产转型

传统零售企业正在积极推进数字化转型，大批线下商超已开始利用数字化手段推进线上线下一体化建设，数字化运营平台对品类、价格、消费行为和库存管理等数据进行持续积累，并加强分析和挖掘力度，增加个性化营销等创新模式，提升整体运营效率，拓展业务空间，推动决策优化。零售行业对数据价值的总体认知仍需进一步提升，在推进线上线下一体化的过程中，努力应对快速演变的应用和数据环境，并克服组织和人员技能相对稀缺等客观问题。

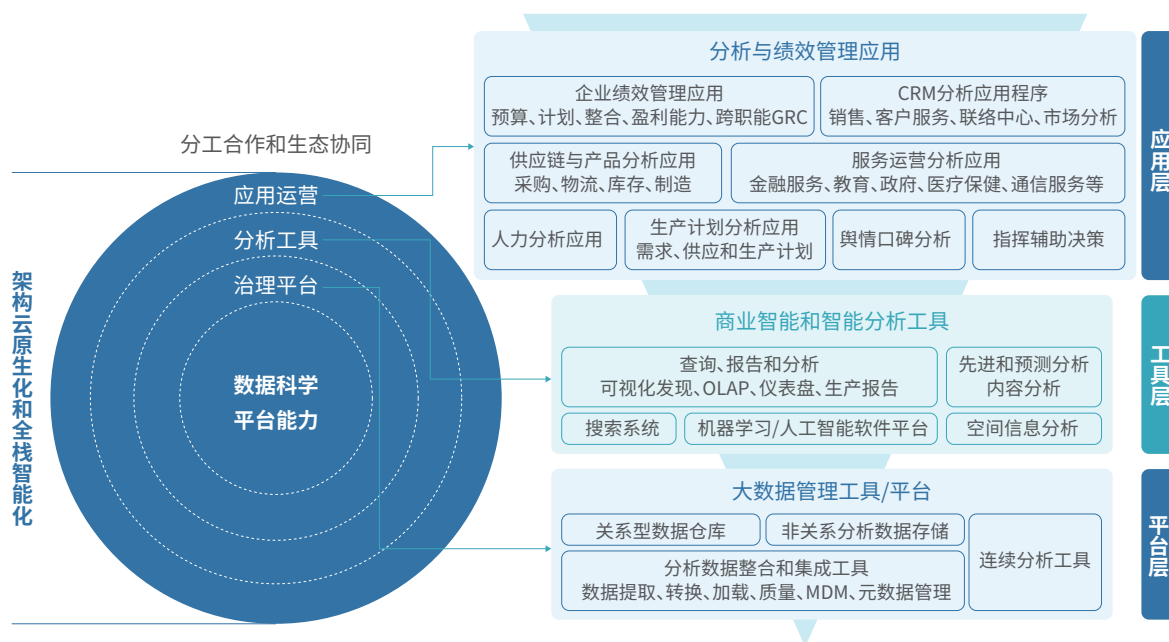
2.2 供应侧 – 基于数据科学的能力演变和生态协同

IDC观察到,随着政企对数据智能需求的不断增加,数据应用与行业纵深领域的联系越来越紧密,相关技术的专业化分工趋势也日渐明显。国内外的主要数据服务厂商都期望构建起以用户为中心的数据全栈化能力体系和应用生态,切实为客户发展提供高价值的服务。各厂商应紧密关注全流程下的分工合作与模式创新,包括:

- **围绕数据价值转化周期推动专业化分工:**将理论与实践相结合,围绕用户的行业特性推进数据能力体系化建设,将理念、方法论、工具平台、持续运营机制进行结合,形成咨询设计、基础设施建设、技术平台、交付以及数据运营等不同业态。
- **建立生态协同机制,创新商业模式:**以共建基础能力平台+行业专业化服务的模式为基础,拓展生态厂商之间的合作机制,特别是在依法依规的前提下,通过商业模式的创新促进不同实体之间的数据合作,帮助各行业实现更广泛的数据创新。

在全行业分工协同的大背景下,基于数据科学的全流程效能提升,将围绕能力演变和架构演变的多个维度深入展开。其中,能力体系趋向全栈化,围绕数据资产管理和价值释放所需要的技术和创新要求,紧密结合数据科学的三层级能力,构建“倒三角”形态的全栈支撑体系。架构体系则力求持续实现数据智能应用的云原生化和全栈智能化。

图4 数据科学平台模型概览



来源: IDC, 2022

- **治理平台——夯实一体化治理平台基础：**以数据科学领域最新成果为依托，构建开放、互联、高度治理的一体化数据科学平台，帮助不同特性的服务商和政企客户固化行业能力，打造领域知识库，实现智能化创新，为各角色用户提供可灵活定制的多维数据服务。
- **分析工具——兼容百花齐放的开源分析工具：**兼容各类高速演进的先进分析工具，涵盖与描述性、预测性、诊断性、规范性分析软件以及与机器学习/人工智能平台有关的各类查询、搜索、报告、可视化等功能。
- **应用运营——构建更加繁荣的应用生态：**由创新社区和政企用户群体共同驱动的更加丰富的数据应用，为各行业提供丰富的选择，例如舆情口碑分析、指挥辅助决策、营收预测分析、服务运营分析、供应链与产品分析、企业绩效管理应用等。
- **总体架构——向云原生化和全栈智能化发展：**全面引入云原生发展成果，利用强大的数据基础引擎和中间件能力，提升数据应用的性能、可用性和稳定性，保障数据智能体系的可持续发展。同时，强化全栈智能能力，打通数据应用与人工智能间的壁垒，将人工智能与机器学习手段引入到数据价值转化流程中，既通过人工智能方法显著提升数据应用的智能化水平，也进一步将更多的基础数据引入人工智能的优化与迭代过程。

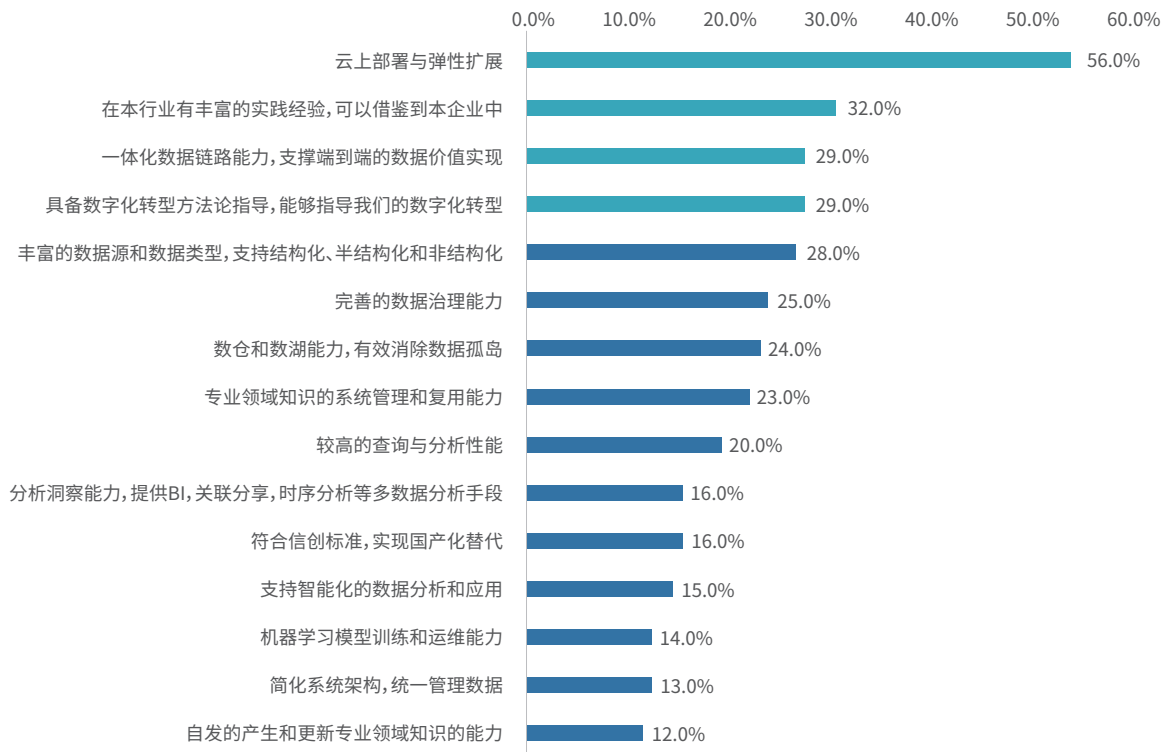
03 第三章

构建数据科学基础平台体系, 推动一体化数据智能实践

依照IDC的数据智能概念模型, 在深入推进数据应用实践的过程中, 政企机构可以进一步打通各环节间的关联, 实现数据服务与业务需求的紧密连接, 进而推动数据供需两端的一体化管理, 建立以数据为核心的长期运营机制。

IDC对政企数据应用所关注的核心需求进行了调研, 结果显示受访对象普遍对基础设施承载能力、行业经验成果、全流程一体化数据应用能力以及数据方法论等较为关注。因此, 各政企机构有必要创建一个拥有数据全流程能力的重要载体, 即一个一体化的数据科学基础平台, 通过广泛、开放的数据协同, 为各业务方提供端到端的完备数据智能体验。

图5 政企机构对数据科学服务商的关注维度



来源: IDC, 2022

3.1 建设一体化的治理平台底座, 打造高质量数据资产

以强调“统一治理”的平台能力逐步消除数据孤岛, 并按照业务需求对数据进行主题化以及场景化建模, 打造高质量和易调用的数据资产, 实现“感知 - 数据 - 信息”的转化。即打造完备的数据治理全流程能力, 覆盖包含采集、集成、治理、开发、运营等环节在内的数据价值转化全流程。

- **规范的采集与汇聚:** 保证源头质量, 提升数据的就绪度。能够支持更加丰富的数据源, 支持互联网数据抓取、业务系统对接、日志文件收集、流式数据推送、人工填报等多种数据采集和集成手段, 涵盖结构化、半结构化和非结构化数据等不同数据类型。
- **高效的治理:** 确保数据质量与安全, 赋予数据以业务属性, 将数据转化为有效的行业知识。即具备完善的数据治理能力, 通过数据标准管理、元数据管理、数据生命周期管理、数据质量管理等标准化过程, 全面支撑数据治理体系, 规范处置文本、图像、音视频等类型丰富的数据。
- **体系化的组织与管理:** 进一步将业务化的数据转化为有价值的资产。政企机构一方面要做好需求端管理, 理清数据与业务的发展关系, 统筹把控好开发与沉淀活动; 另一方面也要规范数据治理与开发流程, 建立数据分层与分级管理机制, 使数据资源尽可能地满足不同业务对象的多样化需求。
- **供需一体的开发与运营:** 当数据转化为机构资产, 并开始持续产生价值时, 以数据服务为核心的高效运营便成为关键, 其有助于在数据供需两方间建立持久、高效的连接, 推动前台业务创新和提质增效。
- **关键环节的AI加成:** 在数据治理侧, 可以采用AI自然语言处理的方法, 将原始数据快速地与数据治理形成对标, 检查数据质量, 降低数据治理以及定制化开发成本, 提升整体效率。同时, 利用自然语言处理和OCR等手段, 融合多源异构的数据, 例如从PDF中提取表格、图片, 利用语音识别、图像识别进行“转文字”操作等, 以提高数据质量和可用度。在数据服务侧, 基于自然语言处理等能力提升服务的智能化和可交互性, 增强服务成效。

3.2 兼容高速演进的先进工具, 提供高效洞察与决策支撑

以趋向智能化的工具能力, 为业务端用户提供可灵活定制的洞察和决策服务, 提升团队效率和创新力, 实现“信息 - 知识/洞察”的转化。

在利用数据智能开发新场景的过程中, 政企开始越来越多地面对创新领域里一些更加复杂的问题, 有必要重点关注两类能力的提升:

- **丰富的分析洞察能力:**具备敏捷BI、搜索、问答、推荐、关联分析、时序分析等多种数据分析手段;
- **高效的知识生产能力:**能够支持指标体系、标签体系、知识图谱三大类知识表示形式,并提供可视化的建模、开发和调度工具,实现便捷的知识生产。

各政企机构需要适时引入智能化领域的发展成果,将其与数据价值转化周期做深度融合,用智能化手段增强数据的分析能力,包括描述性分析、预测性分析、诊断性分析、规范性分析以及人工智能/机器学习分析能力,加强各类模型的开发、匹配和优化工作。

- **描述性分析:**建立科学有效的数据指标体系,寻找事务的内在和外在发展规律,加强对事务的洞察能力。在数据智能相关的重要分析领域中,描述性分析是最基础也是应用最为广泛的分析场景,例如通过商业智能(BI)实现对组织发展状态的全面呈现。
- **预测性分析:**以数据挖掘、机器学习等技术为基础,通过预测模型分析当前和历史数据,识别风险和机会,提出趋势、方向、事件发生可能性等方面的预测性参考。例如在工业设备故障预测、交通态势管控、零售营销等领域实现交互式分析、态势预测、精准化推荐等应用形态。
- **诊断性分析:**以业务逻辑为导向,通过数据分析引起结果的原因,并寻找影响结果的方法。例如在产品合格率下降、商品销售额发生剧烈变化、异常事件突然增多等场景下,利用数据分析定位原因,对相关过程提出改进建议。
- **规范性分析:**深入探索事件起因和结果的关系,帮助决策者在决策过程中减少人为因素干扰,细化决策内容,加大智能化决策过程的占比。例如在政府治理中,提升应急预案的自动化、智能化水平;在金融风控中,提前采取预防和阻断措施等。此外,一些行业数字化先行者积累了大量的行业决策模型,这些模型在日常实践过程中可以通过人工智能手段实现更加自动化的适配和优化管理。
- **机器学习/人工智能增强:**数据基础能力的持续加强,也将进一步推动AI能力的自我优化进步。如何为人工智能和机器学习平台提供可用、易用的数据,是大数据与AI深度融合的一项关键问题,也对底层的技术架构和数据架构提出了更高的要求。良好的数据基础服务有利于打造更具想象力的感知智能、认知智能和决策智能,提升组织创新的水平与活力。

3.3 促进应用生态繁荣发展,与行业实现深度融合

以高效的运营能力,开发和部署与业务场景深度融合的数据应用,建立趋向数字原生化的业务体系,实现“知识/洞察 - 决策 - 行动/应用-再感知”的转化。

政企机构坚持提升自身的业务创新能力对于保持持久的生命力至关重要。不断进步的大数据技术与行业需求深入结合,在发掘新场景、解决新问题的过程中,将行业知识与行业实践经验进行深度融合和数字化沉淀,实现数据智能能力的持续性迭代,这有助于机构的长期、稳定、可持续发展。

- **应用牵引,盘活政企数据资源:**通过平台化基础能力+专业化定制+算法/模型沉淀+长期服务的模式,形成数据价值转化全流程的可持续演进。
- **多方协作,推动政企场景创新:**政企依托一体化的平台建立起广泛的数据服务能力,在面向行业领域的特定需求时,平台企业、生态企业、业务方共同协作,灵活检索和调配数据资源,创新模型与算法,针对场景建立定制化资源供给和流程编排,从而实现对业务洞察和业务决策能力的服务化输出。
- **长期迭代,沉淀专业化场景应用:**长期的行业数据应用实践有利于形成一批成熟的数据模型、算法模型和实用化流程,涉及到政企业务中的用户洞察、经营分析、智能营销、运营管理等一系列重要环节,关系到政企数字化运营的能力和效率问题。这些工程化成果具有极高的行业推广价值,也是组织的高附加值资产。通过合理有效的动态沉淀过程和规范的管理手段,政企机构可以将这些成果固化为标准化的引擎组件,并利用灵活的商业模式向内、外部广泛推广。
- **AI算法模型共创,提升创新成效:**政企机构在不断挖掘新应用场景的同时,还应该联合包括业务部门、IT部门、平台服务商以及算法团队在内的各方力量,共同开展算法孵化和迭代工作。由于行业特性上的一些局限,机构可考虑更多引入专业化的外部力量,利用先进的体系化平台和外部专业化团队解决算法创新问题。

3.4 形成可持续迭代的数据技术框架,综合提升全流程响应能力

面向全流程构建可持续迭代的数据技术框架,以大数据全栈技术实现数据治理,以认知智能技术驱动知识沉淀和人机交互,以场景化分析洞察技术挖掘数据价值,以数据智能应用赋能客户业务。

技术框架将包含基础计算引擎、分布式存储、软件定义网络、数据仓库、数据湖、离线与实时计算、数据同步等一系列基础技术和产品,同时还可能涉及到算法模型管理、数据全链路监控、血缘关系分析、数据标签管理、数据画像等高阶技术能力。在此基础上,数据应用还将与人工智能进一步结合,面向场景提供深度的定制化服务。面对多层次的组合式需求,政企机构需要联合内、外部资源,搭建出开放、灵活的数据技术框架,使其能承载和连接更多的专业化工具与产品,实现不同层面能力的协同与迭代。

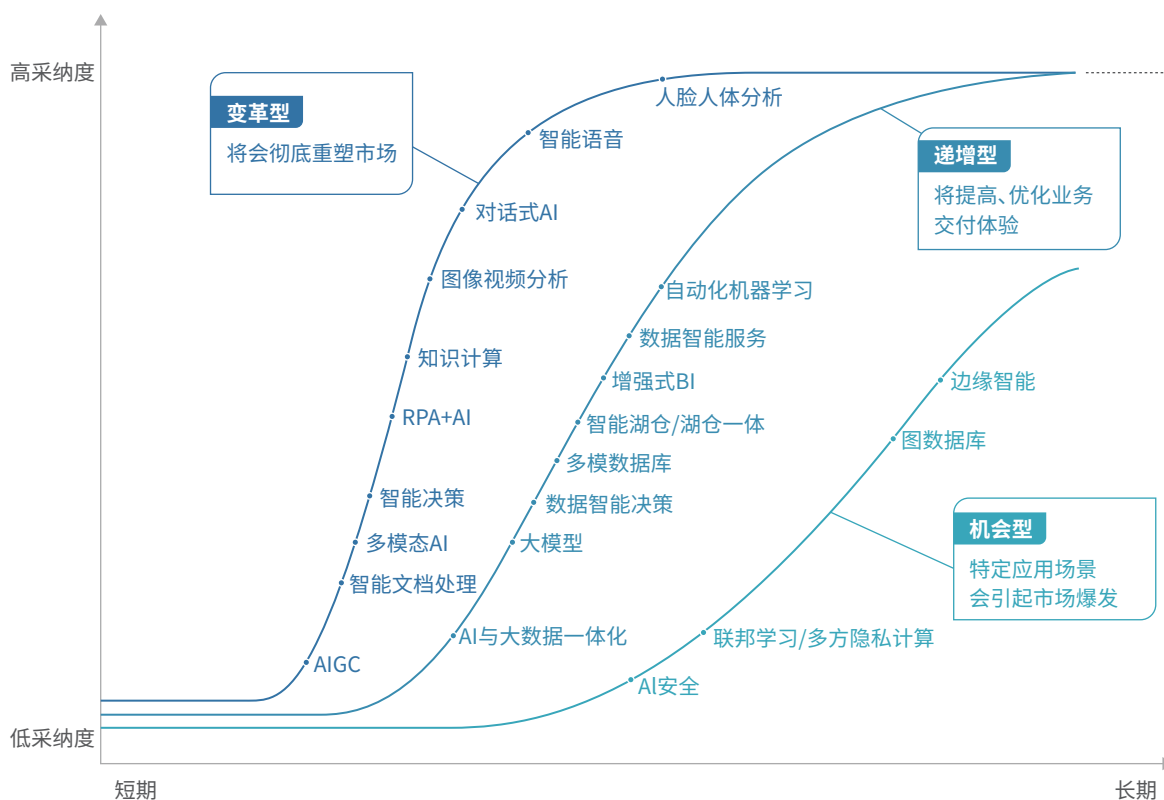
各机构在打造适用于自身发展的数据技术框架时,应根据实际发展现状,侧重选择不同类型的场景和

技术路径。在中国的大数据和人工智能发展现状中,领先者与初入局者差异巨大,例如,领先者已经开始探索一些前沿场景,包括数字人/元宇宙、数字孪生等,而初入局者仍在利用有限的技术尝试碎片化的场景,例如质量检测、舆情分析、营收分析等。

IDC将大数据和人工智能相关的技术按照其特性和市场前景进行了阶段性分类。根据IDC的市场分析方法, IDC技术图谱代表了各种技术采用生命周期在当前的快照,随着时间的推移,这些技术所处的阶段以及对市场的影响都可能会发生变化。

- **关注递增型技术, 持续提高和优化业务交付体验:**利用新一代技术显著改进现有的技术类别,用以交付更好的业务成果。从业务流程的角度看,这些技术提供了超越当前最佳实践的渐进式但可测量的改进。有代表性的递增型技术包括数据智能服务、机器学习、增强型BI、智能湖仓、多模数据库、AI与大数据一体等。
- **重视机会型技术, 重点观察特定应用场景的市场爆发:**这些技术基于特定的用例而发展,具有不确定的能力和有限的可能性来改进现有的技术/过程。此外,它们的潜在变化也缺乏对当前最佳实践的确定性影响。有代表性的机会型技术包括边缘智能、图数据库、联邦学习/隐私计算、AI安全等。
- **战略性关注变革型技术, 获得重塑市场的先机:**这些技术从根本上改变了当前的最佳实践,并且将彻底重塑市场和投资策略,可能创造新的商业和/或市场机会,并带来新的企业和消费者能力。它们可能与当前的技术有很大的不同,并且可能带来很多未被认识到的市场影响/机会。有代表性的变革型技术包括多模态AI、图像视频分析、智能文档处理、智能决策、RPA+AI、知识计算、智能语音、人脸人体分析、对话式AI等。

图6 中国人工智能与大数据技术图谱, 2022



来源: IDC, 2022

3.5 小结: 以数据科学基础平台体系应对数字化转型多维度难题

数据科学基础平台体系能够切实帮助政企机构解决数据/信息转型中的多维度难题与挑战, 这些挑战主要涉及决策信息不明确、人才配置不合理、数据文化缺失、业务流程滞后、单点技术孤立等普遍性问题。

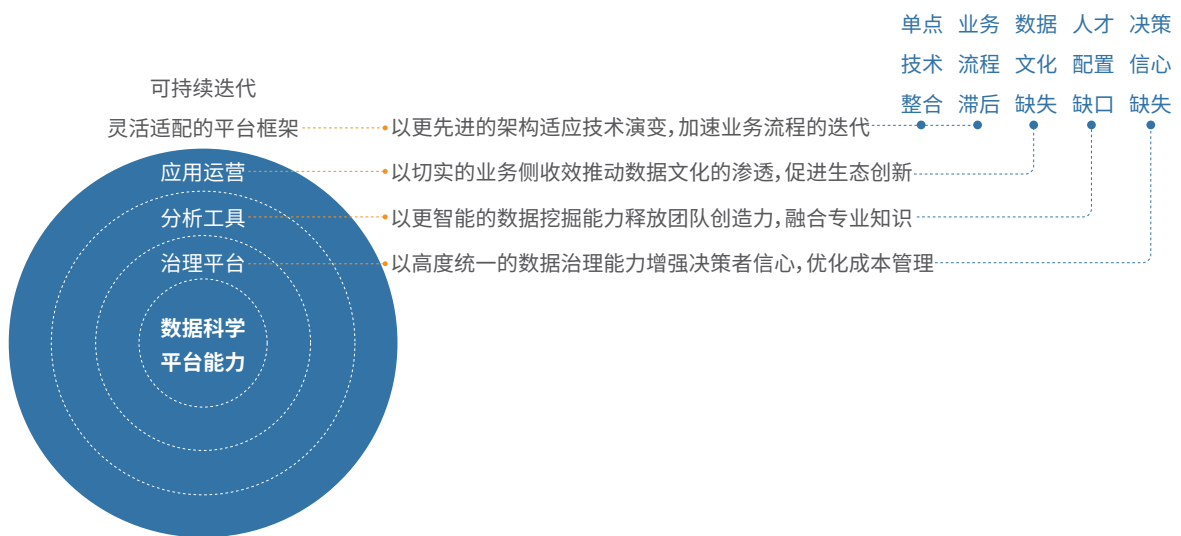
以数据科学基础平台为依托, 通过将统一治理平台、先进分析工具、高效应用运营和可持续迭代的技术平台框架相结合, 能够有效帮助各类机构在数字化转型过程获得能力体系的全方位提升。

- 以高度统一的数据治理能力增强决策者信息, 优化成本管理。
- 以更加智能的数据分析挖掘能力释放团队创造力, 推动数据与专业知识的结合, 形成深度行业洞

察力。

- 以实际的业务侧成效, 推动数据文化的深入普及, 促进行业生态创新。
- 以更先进的技术架构适应技术演变趋势, 加速业务流程的升级迭代, 实现不同技术的整合与融合。

图7 数据科学基础平台助力数字化转型



来源: IDC, 2022

04 第四章

IDC建议:政企机构数据智能全流程实践关键举措

4.1 将数据科学能力纳入未来发展战略

在当前数据智能向行业纵深发展的时代背景下,政企机构对于数据产品和技术所能产生的具体场景价值越来越关注,期望能在端到端的业务应用上体现出更丰富、更具象化的数据智能价值。

各机构应将数据科学能力纳入未来企业发展战略,基于数据价值链全流程的理论基础,在顶层设计的引导下,打通数据集成、数据治理、数据建模、数据分析、数据应用的全流程,形成端到端的一站式解决方案,这也要求相关的团队能够掌握数据科学的理论、方法和技术,具备业务分析、数据建模、数据应用、数据工程实践等全方位能力,构建融合技术、数据、场景的智能化应用,为自身乃至整个行业的数字化转型赋能。

完备的端到端一体化数据价值转化体系,既具备数据智能领域通用的场景、流程、工具、交付方法,又能积累足够的行业场景化模型和成熟引擎,进而在组织内部的不同行业项目中,根据客户具体数据、业务场景进行适配。各政企机构还应注重利用技术迭代,强化采集、存储、计算、智能增强和服务等关键环节能力并使其更易于相互连接,从而以体系性的能力进步,提升全行业数字化转型成效。

4.2 规避单点方案和开源技术的成本陷阱,选用可持续迭代的全栈能力平台

政企机构应注意减少割裂式的单点系统部署活动,避免造成不必要的能力孤岛和重复建设现象。同时谨慎对待开源技术的无序引入,防止因自身团队能力不足而造成长期的发展停滞,并遭遇大量运维问题的困扰。建议各类组织适时引入先进的数据智能发展成果,建设科学、融合的数据治理和资产管理平台,将自身的数据能力打造成标准化服务,这既包括增强机构传统数据(小数据)的价值呈现能力,也包括在深化数字化转型的过程中,新打造的海量数据应用能力。

新一代数据智能平台的发展目标和重点包括:让数据资产更加标准清晰,让开发过程更加敏捷化,让数据服务更加全面多维,以及让数据运营更加高效。

平台自下而上需要囊括数据、方法、产品、场景4个不同层面的能力、组件和要素,这些重要的组成内容相互作用,缺一不可,共同构成了向上支撑和向下反馈的优化迭代闭环,为政企机构实现可持续的数据

科学能力进步提供了体系化保障。

图8 可持续迭代的全栈能力平台架构



来源：IDC, 2022

4.3 明确数据价值转化流程，集结跨领域人才团队，激发应用创新活力

以数据智能为基础的应用创新涉及到算法/模型、知识图谱、智能搜索、融合分析、业务洞察等一系列关键能力的提升。政企机构在利用大数据推动业务创新的实践活动中，应明确数据价值转化全流程的重要环节，根据自身发展情况补齐短板，建立起完善的能力体系。更重要的是，要集结跨领域人才团队，联合相关的行业生态，发掘更多的行业深度应用场景，建立数据驱动业务发展的创新业务体系，并将相关的业务过程在平台工具中固化下来，形成领域性的知识沉淀，建立面向领域的深度洞察和专业化智能服务，使行业数据智能的能力和水平获得不断提升。

在知识积累的作用下,当用户需要支撑一个具象化的场景时,可以先通过检索行业类目、标签等识别出已有的场景化知识沉淀,获取到在此场景下可能需要用到的数据、方法、展示方式、分析报告模板等。各企业也可以采用创新商业模式的方式,将一些场景化知识内容向行业内推广,从而实现创新性的数字化开放与合作局面。

4.4 量化跟踪数据驱动决策和业务的收益

各机构在构建数据智能体系化能力的过程中,应根据自身的业务特性和应用成效,逐步梳理出一套用于对数据驱动决策和应用的收益进行量化评价的指标体系,以推动自身变革路径的优化管理,以及对未来投资方向的精准把控。

在这个过程中,各机构应以数据科学为依托,进一步加强对流程、方法、模型的深入认知,全方位理解其产生业务价值的过程,明确所部署的模型或用例如何增强业务成效、赋能业务创新。上述领域的量化指标也是政企机构数字化评价体系的重要组成部分,有利于推动其整体数字化能力的持续提升。

4.5 以全面数据价值转化为目标,寻求敏捷高效的技术路径

政企机构应努力寻求敏捷高效的技术路径,特别是充分引入当前的人工智能发展成果,不断增强数据智能引擎的自动化、智能化水平,提升数据价值挖掘的成效。在此过程中,值得深入实践的关键技术领域包括:

- **多模态数据融合:**利用创新性的多模态预训练模型,通过大规模的数据预训练,掌握不同模态之间的语义对应关系,进而以统一的方式对文本、语音、图像和视频进行建模和学习,可以显著提升各种数据训练和模型优化任务的效果。
- **语义理解与知识构建:**可利用自然语义分析技术对接入的各类数据信息进行实时动态分析处理,标注关键信息并关联相关的领域知识图谱。同时,依托一体化平台可以对各行业领域内的指标、标签、图谱以及典型的数据应用场景进行梳理和固化,形成领域知识。
- **分层数据建模和统一数据服务:**按照业务需求建设合理的分层数据模型,通过结构化治理实现清晰的数据架构,创建相关的数据主题和专题,逐层提升数据价值密度,并最终定制出丰富的场景化服务。

05 第五章

行业实践:百分点科技的数据科学平台

5.1 数据科学基础平台的愿景和价值主张

百分点数据科学基础平台是一站式的数据价值实现平台,它涵盖了数据应用构建的全栈工具集,并内置行业数据标准、算法模型、知识图谱等领域知识,赋能数据价值实现的各个环节,帮助客户提高构建数据应用的效率及深度广度。

百分点科技成立伊始即专注于数据智能应用的研发和产业化,相继推出了个性化推荐、用户行为分析、舆情洞察等SaaS服务。在此过程中,百分点科技积累了海量数据采集、存储、建模、分析、机器学习、可视化和应用的端到端技术能力,形成了独特的数据科学方法论——“数据决策力”模型,并将此思想落实到百分点科技的数据科学系列产品中,为客户的数字化转型赋能。

图9 数据决策力模型



来源:百分点科技,2022

数据决策力,是基于数据进行科学决策并产生价值的能力。“数据决策力”模型描述了建设数据决策力的过程和释放数据决策力的过程。

建设数据决策力:它服务于政企机构长周期的数智化转型。数据决策力的建设和迭代由五大步骤形成的闭环构成,即核心信念(Belief)、架构设计(Architecture)、专业团队(Staff)、基础设施(Infrastructure)和机构能力(Capability),简称为数据决策力建设的“BASIC理论”。

释放数据决策力:它服务于政企机构短周期的数据智能应用建设。数据决策力的释放和验证由三大步骤形成的闭环构成,即业务数据化(Datafy),发现洞见(Discover)、设计重构(Design),简称为数据决策力释放的“3D理论”。

“BASIC理论”和“3D理论”是一个完整的整体,正如“恒星-行星”之间的关系。

经过十余年发展,百分点科技孜孜不倦地践行和完善“数据决策力”模型,服务的客户从企业拓展到政府,从国内跨越到海外。在历经10,000多家客户业务场景的实践和锤炼后,百分点科技逐步演进出专业化、一体化、智能化的数据科学基础平台。

百分点数据科学基础平台,围绕数据价值转化过程,将数据科学中的关键工作统一到一个平台上,打通数据采集/汇聚、数据存储/治理、数据处理/计算、数据分析、数据应用各阶段,让数据科学团队中的每个成员都只需专注于核心业务问题,免于陷入复杂的技术环境:

- **数据工程师:**可通过平台完成数据采集/汇聚、数据存储/治理、数据ETL等工作。自研集成引擎,支持多种数据源接入,通过可视化方式完成多源异构数据集成;提供数据目录、数据质量、数据血缘等治理能力;支持流批一体计算、低代码开发模式,支持 workflow 编排。处理后的数据可接入到知识生产、知识应用工具中。
- **数据分析师:**可通过平台完成数据分析、数据应用等工作。通过利用数据工程师加工后的数据,采用数据探查、数据可视化、行业模型、知识应用等工具展开数据分析,各工具之间数据互联互通,并支持以编码方式进行数据分析。
- **数据科学家:**平台提供了NLP、智能语音、计算机视觉等方面预训练模型,同时支持自助式机器学习、AutoML、MLOps等能力供数据科学家构建机器学习模型,可将训练得到的模型注册为服务供数据分析使用。

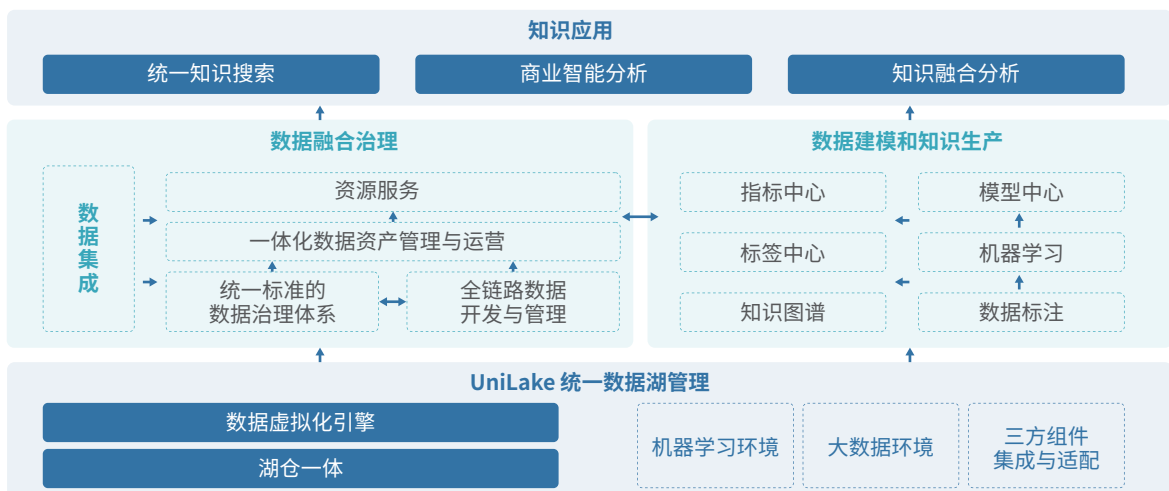
百分点数据科学基础平台是数据科学领域的实践创新成果,它基于先进的数据科学方法论以及行业应用实践,打通了从数据接入到数据分析、数据应用的完整链条,形成如下核心优势:

- **完善的数据科学工具集:**覆盖数据集成、数据治理、数据建模、数据分析、数据应用的全流程,产品可灵活组合搭配,具备为客户构建端到端场景化解决方案的坚实基础。
- **端到端的解决方案:**产品内涵了数字化转型方法论、路径规划、产品工具到项目交付、数据运营服务一体化的能力,能帮助客户打通“感知 - 数据 - 信息 - 知识/洞察 - 决策 - 行动/应用 - 再感知”的闭环,高效支撑客户现有业务的强化升级和新业务的探索创新。
- **丰富的行业经验:**百分点科技已服务超过10,000家企业和政府客户,通过行业的实践,积累了扎实的行业知识、业务理解和行业上下游资源,形成了大量行业数据标准、行业知识库、领域算法模型等,具备为客户提供科学可落地的实践经验。

5.2 数据科学基础平台的含义和创新

百分点数据科学基础平台致力于解决数据价值转化过程中的共性需求,为数据工程师、数据分析师和数据科学家群体提供能力全面、交互自然、知识驱动的通用工具,高效构建数据应用。

图10 百分点数据科学基础平台功能架构



来源:百分点科技,2022

为了让用户专注于数据价值的实现,从各种技术组件的复杂配置和相互适配中解脱出来,百分点数据科学基础平台提供了统一的湖仓一体架构,屏蔽了底层存储和计算组件的技术细节,并基于此提供了数据融合治理、数据建模与知识生产、知识应用三类专业工具集。平台的主要功能有:

- **湖仓一体架构:**支持外部数据源物理入湖和虚拟入湖,支持结构化、半结构化和非结构化数据,并提供统一的元数据、数据权限和安全管理。支持Hive、Spark、Flink等大数据处理引擎,以及R、TensorFlow、PyTorch等机器学习工具,支持批流一体计算,支持联邦查询。支持通过标准SQL访问湖仓数据。
- **数据融合治理:**高效集成网络开源数据、业务系统、日志文件、IoT数据、人工填报信息等多种数据源。支持关系型数据库的探查和CDC实时捕获。支持多种数据处理任务的开发和统一调度。支持完善的数据治理体系,包括数据标准管理、元数据管理、数据生命周期管理、数据质量管理、主数据管理等。支持数据资产的管理和运营。通过数据资源目录提供数据库、文件、API等多种数据服务方式。
- **数据建模与知识生产:**平台提供了统一的多模态数据建模能力,包括数据标注、机器学习、模型管理等,提供了在线Notebook和低代码的模型构建能力,并且内置了多种机器学习算法和开箱即用的AI模型。这些模型可以用于构建业务知识,平台支持标签、指标和知识图谱三种常见的知识表示形式,并提供了对应的知识生产和管理能力。
- **知识应用:**平台提供了统一知识搜索,可以快速检索数据湖、数据仓库、数据资源、标签、指标、知识图谱,支持全文检索、图文搜索、以图搜图、内容推荐等能力。平台提供敏捷BI,可以快速分析数据仓库、标签和指标数据。平台还提供了知识图谱分析和推理工具,支持实体分析、关联分析、时空分析、图谱挖掘等多种数据分析手段。平台统一管理领域内的数据标准、数据加工算子、数据仓库模型、知识图谱本体模型、分析算法等知识,并运用这些知识增强数据治理、数据建模和知识生产、知识应用的全过程。

百分点数据科学基础平台具有如下特点:

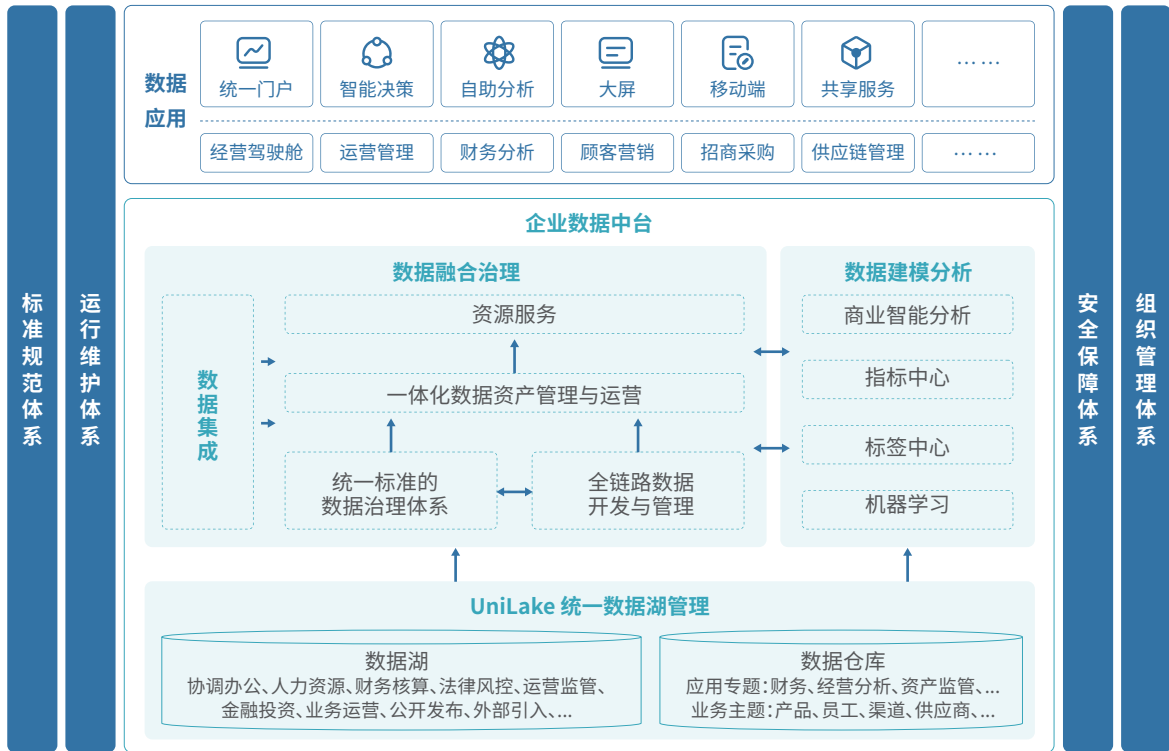
- **专业化:**平台包含了数据价值链必需的工具和算法模型,实现从数据到信息、从信息到知识、从知识到决策的转换。
- **一体化:**平台中的工具、数据、模型、知识都遵循统一的协议、标准和规范,可以无缝对接和互操作。
- **智能化:**平台能够复用领域知识,并以此增强用户开发数据产品的能力,标准化用户进行数据价值链实现的转换过程。

- **低代码:**支持用户以低代码/无代码方式开展各项工作,降低技术门槛,提高实施效率。
- **兼容性:**兼容多种信创软硬件,覆盖芯片、整机、操作系统、中间件、数据库和应用等;适配多个厂商的存储与计算平台。

5.3 平台应用实践

5.3.1 企业数据中台解决方案

图11 百分点企业数据中台解决方案架构图



来源:百分点科技,2022

方案简介：

基于数据科学基础平台，构建以主数据管理和数据服务为核心的一站式企业级数据中台，形成完善的多源异构数据采集能力、多模态数据存储和治理能力、数据全生命周期管控能力、数据分析及服务能力，支持企业建立数据驱动运营模式。

主要用户：

- 1.面向基层执行者，提供数据资产查询、共享和分析能力，提升执行效率。
- 2.面向中层管理者，提供运营分析报告和业务指标监测预警，辅助业务决策。
- 3.面向高层管理层，提供应用算法模型指导决策，辅助战略决策。

方案价值：

- 1.融合企业内外部数据，构建数据资产。通过物理入湖或逻辑入湖方式集成企业内外部数据，经过融合治理形成企业自有数据资产以及服务能力，消除数据孤岛。
- 2.数据驱动业务执行，提升运营效率。统筹企业数字化运营建设，提供各类业务分析报告和场景化数据API，以智能化应用替换简单重复劳动，提升企业经营管理效率。
- 3.数据指导业务决策，降低经营风险。基于业务全链条数据，构建场景化分析和决策模型，辅助决策者准确把握形势并科学决策。

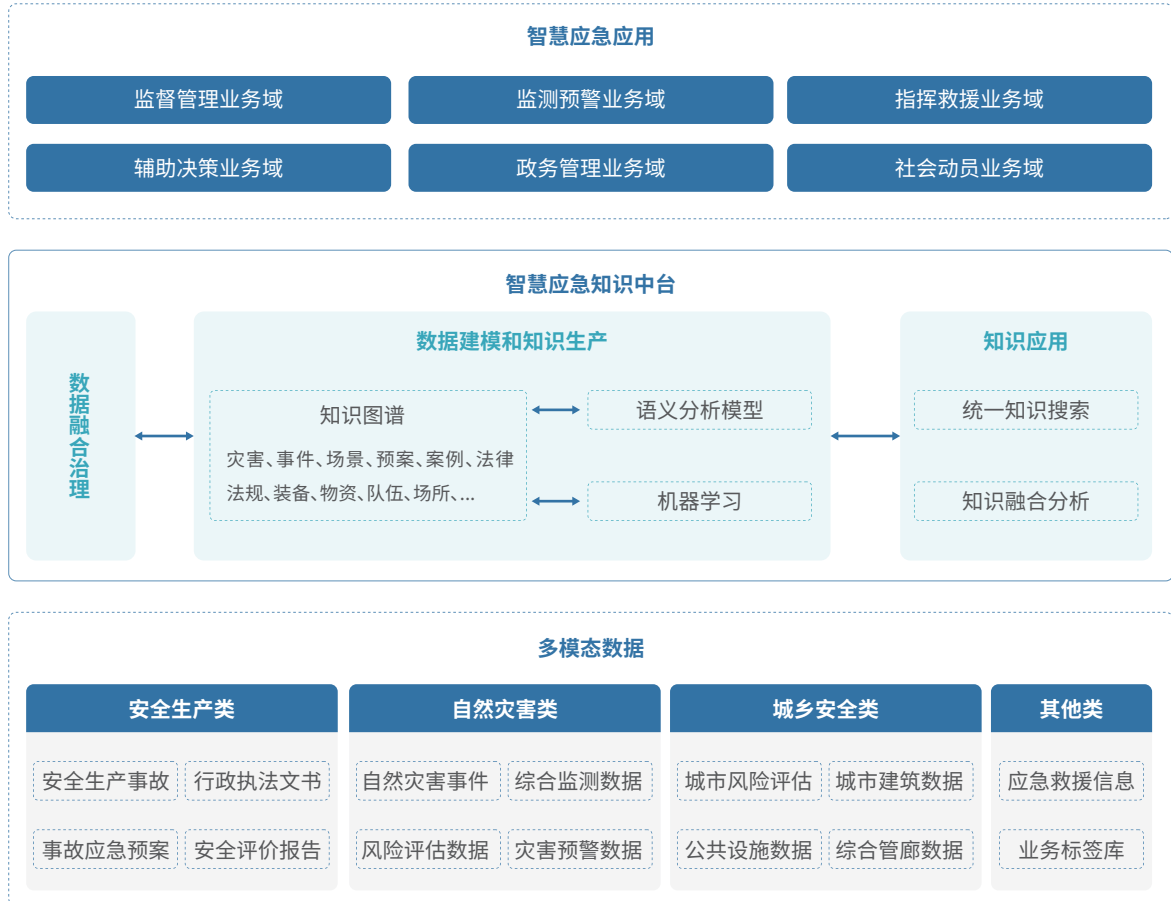
经典案例

中国中免大数据管理与决策分析平台

建设中免数据中台，打造招标采购、运营管理、市场营销、物流配送四个方面的数字化运营体系，形成端到端的数据服务体系，有效促进业务的集成和协作。业务部门分析流程实现线上化，规范门店营销流程。形成数据决策分析体系，人均数据分析效率大幅提升。融合“人货场”与“进销存”双要素，建立销售归因、客群甄别等18个场景化决策模型，实现市场营销、采购补货等方面的智能决策，成功落实集团会员战略，会员销售占比由52%提升至87%，显著拉动消费额度和复购率。

5.3.2 智慧应急知识中台解决方案

图12 百分点智慧应急知识中台解决方案架构图



来源:百分点科技, 2022

方案简介:

利用数据科学基础平台,融合应急业务中的多模态数据并形成应急知识图谱,促进监督管理、监测预警、指挥救援、辅助决策、政务管理、社会动员等业务领域的智能化、精细化、科学化、专业化,助力应急管理部门能力提升和治理体系现代化。

主要用户：

- 1.面向应急处置人员,实现应急指挥中的智能辅助决策,提高应急处置效率。
- 2.面向预案编制人员,实现数字化预案的管理、编制和审核等,提高预案质量。
- 3.面向公众,实现应急知识的智能搜索和问答,提升公众的应急意识和技能。

方案价值：

- 1.打造智慧应急知识库,包含应急管理领域的法律法规、标准规范、事故案例、灾害链、致灾因素、处置手段等内容,为预案编制、应急指挥、防灾减灾、安全生产、数字化预案等应用系统提供智能推荐、搜索、问答和关联分析等智能服务,实现业务与知识深度融合、相互促进,为应急管理智能化、精细化赋能。
- 2.基于知识库开展分析挖掘,研究灾害之间的影响关系,提升对灾害的认识,指导相应的预防、预警、预测和处置能力建设。

经典案例

灾害链分析和预警应用系统

应急灾害链分析和预警应用系统是为应急管理部信息研究院研发的领域知识图谱应用系统,系统构建了隐患、灾害、事故、灾害事故损失、人员伤亡、时间和地点这七类实体及它们之间的关系。系统面向应急专家和研究员,为他们撰写报告提供数据和系统服务能力,包括灾害链知识搜索、灾变孕育预警和灾害事故分析模块,实现灾害事故新闻的图谱搜索、分析以及灾害的预警提示等智能应用场景。基于灾害链图谱的预警辅助,有助于各地区各部门在灾害发生时及时了解灾害事故演变情况,采取“断链”处置及早斩断灾害演变,控制损失。

5.3.3 网络信息融合分析解决方案

图13 百分点网络信息融合分析解决方案构图



来源：百分点科技，2022

方案简介：

利用数据科学基础平台，融合网络开源数据、网络日志数据和人工导入数据，构建包含人、地、事、物、组织等要素的统一、灵活、动态可扩展的安全领域知识图谱，并基于此构建实体分析、关联分析、时空分析、图谱挖掘等高级研判工具，帮助公安管理部门提高预测、预警、预防能力，打击网络犯罪活动。

主要用户：

- 1.面向网信部门，识别网络安全风险，辅助网络空间治理。
- 2.面向公安部门，辅助案件综合研判，打击网络犯罪活动。

方案价值：

- 1.提高国家网络安全防范能力。结合网络威胁信息库和敏感内容库,识别网络攻击行为和风险内容,阻止有害网络信息的扩散,全面提升网信管理部门对网络空间危险事件的识别能力,加强防护措施提升突发事件应对处置能力,保障国家网络空间的安全。
- 2.提高网络犯罪案件侦破能力。综合利用多源数据融合优势,深挖网络数据的潜在价值,对可疑目标进行挖掘,助力公安管理部门有效打击各类利用网络开展的非法活动,提升网络案件侦办效率,维护社会安全稳定。

经典案例

国家级网络空间安全分析系统

为某海外国家建设网络空间海量数据分析项目,实现网络日志数据的实时接入和处理,以及准实时的数据加工。实现日志数量2,000亿条/日、网络数据存储量200TB/日、峰值数据流量达20TB/秒的海量数据实时处理,实时构建千亿级别的关系图谱,展现数据对象之间的关联关系、时间与空间等维度的分布。采用人工智能技术对多媒体资料进行语义分析,并提供秒级的数据搜索能力,实现海量数据的深度分析和挖掘。系统有效协助安全部门识别和控制有害网络信息,协助调查网络犯罪,从而净化网络空间,保障国家网络安全。

06 第六章

未来展望

6.1 机遇与挑战

在全球化发展面临严峻挑战的今天，以政企数字化转型为代表的数字经济仍在高速发展，推动各行业不断突破自我局限，实现持续增值与加速。全面的数据价值转化能力为政企运营坚韧性保驾护航，更好的对抗宏观经济的不稳定性，领先开拓新市场。同时，各机构也需要适时引入风险管理体系，在控制风险和鼓励创新文化之间寻求信任平衡。

面对未来经济领域的诸多不确定因素，数据科学领域仍需要直面以下机遇与挑战：

- **与实体经济深度融合带来更广阔发展空间，但需努力保持发展的可持续性：**数据科学赋能传统产业转型升级，催生新产业、新业态和新模式，为数字经济发展带来新的推动力。在此背景下，政企数据科学发展的可持续性是当前最大挑战，尤其是在能源紧缺的大背景下，如何利用大数据创造新的产品和服务，激活政企创新活力，实现增值效应，是值得全社会深入思考的问题。
- **组织数据文化快速进步，但需解决好决策体系的优化问题：**在实现数据科学创新的过程中，政企机构自上而下的数据文化认知非常重要。管理层应主动优化和重构自身的决策体系，更多地依靠客观数据产生洞察力，进而指导业务发展，避免在整体发展问题上的主观和盲目。对外部的生态协同工作同样应建立起良好的数据共享与交换机制。
- **人工智能成果百花齐放，但须高度重视风险管理问题：**AI在各行业的应用依然面临较大的风险管理和合规性挑战。在推进立法和管理流程建设的同时，相关的技术体系仍需要进一步突破和完善，例如AI领域的图神经网络、概率编程以及因果推断可解释等。尤其在金融、交通、制造等一些重要领域，应力求利用大数据与AI的融合平台实现全面风险管理，增强AI技术的规范性和可解释性，推动全流程合规性建设。
- **技术领域发展态势迅猛，但关键环节能力需取得进一步突破：**从中期看，诸如多模态数据理解和融合等仍未达成理想状态，其实际应用效果还不够突出。OCR、NLP、视觉挖掘等领域仍需要投入持续的研究和关注，其有助于对整个数据融合治理过程有更显著的能力增强。同时，AI增强分析、自主分析等主动从数据中发现问题规律的能力也仍需持续增强。除此之外，数据编织、数据虚拟化、数据多方安全计算等技术的发展以及相关的商业模式创新均值得关注。

- **用户体验持续改善,但需更加重视用户层面的深度交互需求:**在用户一侧,数据可视化、自然交互、XR等更贴近用户体验的技术有待继续发展完善,许多用户特别是决策层的用户并不关心底层的数据汇聚,他们只关注报告的形态和大屏呈现效果,希望新的想法能够快速反映到数据呈现形式上,能通过 AI 等方式与数据进行深度交互。
- **数据基础治理环节日趋完善,但提升智慧化水平仍需多环节加速进步:**当前的成果仍更多集中在与数据处理相关的各项环节,但未来数据科学的基础理论将更加聚焦在从数据、信息到知识的转换过程,进而推动向全行业智慧化的本质性转变。因此,围绕数据科学和AI的进化,仍存在巨大的发展空间。

6.2 未来展望

随着数字经济的发展,尤其是产业数字化进程的加速推进,数据科学技术将越来越关注端到端数据应用的落地,这将推动数据科学产业、业务和技术领域的一系列发展变化。

数据驱动的产业重塑:随着数字经济时代的深入发展,全社会数字化进程势不可挡,这将催生以数据科学探索和数据智能实践为基础的产业新格局、新形态、新模式。例如,电商领域的大规模发展引发了传统零售模式的重塑,推动传统零售商构建以数据为驱动力的线上线下一体化新零售体系。在我国特有的海量数据应用前景和巨大的市场应用规模优势下,数据智能产业有望迎来空前的发展机遇。围绕用户的数据资源体系顶层设计、深层数据治理、一体化平台建设、数据科学开发与应用、数据运维与运营等需求,有望形成创新型企业分工合作和深度协同的良好生态发展格局。此外,随着“一带一路”战略推进带来的国际市场机遇,相关机构也应进一步做好人才和技术储备,保持足够的竞争优势,迎接更广阔的国际市场挑战。

数据驱动的业务流程重塑:数据创新应用将逐渐成为大多数组织的核心业务组成,在组织发展分析、洞察、决策过程中发挥不可替代的作用。例如,以手机应用和智能识别引发的新贷款征信业务流程,推动了银行业务的大范围拓展。未来的数据创新业务将持续强化人工智能/机器学习能力的应用,通过人工智能提升数据模型的创新效率,不断增强洞察、决策过程的智能化水平,实现数据价值释能过程的自我优化。值得一提的是,在业务创新过程中,各行业领域知识的生产和应用能力有望进一步得到加强。传统行业拥抱数字化时面临着严重的冷启动问题,必须借助行业内现存的专业知识破局。因此,深入研究如何更好地生产和应用领域知识将具有重要的现实意义。

数据科学技术的深化发展:随着产业分工和协作的加强,数据科学相关的技术框架将得到进一步发展完善。同时,围绕强化数据价值释放各环节效能的理论、技术研究将不断深化,包括但不限于以下细分

领域：

- **数据空间的结构和转换。**数据科学需要从理论角度进一步阐明数据空间(包括数据、信息、知识和决策等)的层次结构,以及不同结构、不同操作、不同价值之间如何转换,最终落实到数据价值链的相关技术上。
- **多模态数据的理解和融合。**数据科学面对越来越多的非结构化数据,如何从这些数据中提取信息并进行融合,是应用落地的重要问题。
- **可解释性和因果分析。**可解释性对于业务决策过程十分重要,它是模型更加可靠、稳定和可复制的重要保障,而强化因果分析可以带来更强的可解释性。目前,关于可解释性机器学习、因果推断模型等领域的研究热度在持续提升,并有望很快将成果投入应用实践。
- **端到端的模型和算法。**目前的模型和算法,都是应用于数据价值链的某个阶段,解决局部问题。随着相关技术的发展,有望构建从数据到决策的端到端模型和算法,使数据科学的应用方案更加成熟。
- **数据安全和隐私保护。**全社会对数据安全和隐私保护的关注度都在快速提升,数据科学也必须融合相关技术使自身的应用符合法律和伦理要求。



结语

数据科学领域的技术、方法、模型、流程将持续获得进步，数据科学基础平台将有效帮助各类政企机构夯实数据应用基础，依托平台长期开展与数据智能相关的能力积累、业务创新与技术升级迭代。从市场发展的角度看，数据科学基础平台具备广泛的行业推广基础和巨大的可预期市场规模，综合发展前景值得期待。



关于IDC

国际数据公司 (IDC) 是在信息技术、电信行业和消费科技领域, 全球领先的专业的市场调查、咨询服务及会展活动提供商。IDC 帮助 IT 专业人士、业务主管和投资机构制定以事实为基础的技术采购决策和业务发展战略。IDC 在全球拥有超过 1100 名分析师, 他们针对 110 多个国家的技术和行业发展机遇和趋势, 提供全球化、区域性和本地化的专业意见。在IDC 超过 50 年的发展历史中, 众多企业客户借助 IDC 的战略分析实现了其关键业务目标。IDC 是 IDG 旗下子公司, IDG 是全球领先的媒体出版, 会展服务及研究咨询公司。

IDC China

IDC中国(北京): 中国北京市东城区北三环东路36号环球贸易中心E座901室

邮编: 100013

+86.10.5889.1666

Twitter: @IDC

idc-community.com

www.idc.com

版权声明

凡是在广告、新闻发布稿或促销材料中使用 IDC信息或提及IDC都需要预先获得IDC的书面许可。如需获取许可, 请致信gms@idc.com。翻译或本地化本文档需要IDC额外的许可。获取更多信息请访问www.idc.com, 获取更多有关IDC GMS信息, 请访问<https://www.idc.com/prodserv/custom-solutions>。

版权所有 2022 IDC。未经许可, 不得复制。IDC保留所有权利。