

城市智能体参考架构与技术实现

5.1 城市智能体参考架构

5.1.1 全面发展的城市智能体新理念

现有城市智能中枢以使用者为本,即以城市管理者和市民等用户为中心,注重对用户需求关注,在业务活动与需求的分析、早期的测试和评估,以及迭代式的设计方面更强调用户体验。技术框架以 IT 应用数据为驱动,强调各类信息的数字化和融通汇总,业务上大多局限于政务活动。从整体思想领导力上看,现有智能中枢更贴近于人的“大脑”,汇聚四肢、躯体信息,形成思想决策,但缺乏反向输出,未对四肢、躯体形成有效指挥。

对此,我们提出从局部智能中枢走向全面发展的城市智能体的新理念,如图 5-1 所示。城市智能体将以人为本,在以使用者为本的基础上,不仅关注用户与系统的交互,更加注重系统如何影响人的能力和特征;在技术上,加速鸿蒙感知、算力智能、网络使能、泛在安全等方面的建设,构建智能感知数字底座;在业务上,拓展政务服务、政务办公、城市治理、城市管理、城市实时性问题快速响应和高效处置等业务。城市智能体将从“大脑”扩展到整个“人体”,包含中枢“大脑”、感知“五官”、云网“经脉”、数据“血液”和应用“手脚”,形成全身联动机制,打造全流程、更智能的城市智能体。

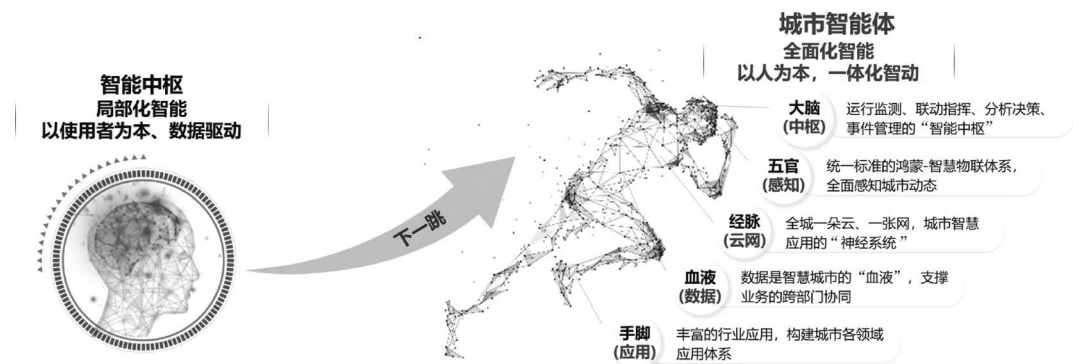


图 5-1 城市智能体理念

智慧的大脑,即会学习、会思考,能判断、能决策的城市和政府运行指挥中枢。旨在形成跨城管、市监、水务、城运中心、环保、应急、人社等部门应用协同联通机制;算随数建,本地政务数据本地用,挖掘数据价值孵化适配本地场景的 AI 算法;训推一体,持续迭代促进自

主演进,快速响应业务需求变化及新增场景需求。

灵敏的感官,即全域灵敏感知。利用视频感知、卫星遥感、噪声检测、环境感知、震动感知、水质监测和物联感知等技术,形成地上、地下一体全模态感知;运用鸿蒙操作系统、CIM(City Information Modeling,城市信息模型)、视频分析、人工智能等搭建数字孪生平台,实现全域态势精准感知,为城市和公众安全提供保障。

顺畅的经脉,即城市的云和网。打造连接市、县、乡、村的专网,以网带云,达到市区一朵云,包含政务云、赋能云、产业云、公有云等,以云带网,形成市区一张无损低时延网络。同时,云网协同,打造全域一体化安全。

流动的血液,即鲜活的、实时的数据。“聚”实现包含供水、电力、燃气等公共数据和交易、新闻等社会数据汇聚,达到多源异构亿级数据分钟级入湖;“通”实现数据高效流通,从分钟级跨越到秒级;“享”实现数据可信共享,利用区块链技术,实现数据在各部门之间可用不可见,跨域流通可追溯。

灵活的手脚,即跨领域的应用协同。例如,通过“一网通办”,实现一件事一次办、“无感申办”和“一码通办”的“好办”,以及零跑腿、“秒批秒办”“免审即享”的“智办”;通过“一网统管”,实现 AI 慧眼识事、智能调度、联动处置和每日效能分析,让政务服务数字化,城市管理精细化。

5.1.2 城市智能体参考架构

城市智能体是运用大数据、云计算、物联网、人工智能、区块链、数字孪生等技术,集数据、智算与业务三位一体的新型信息基础设施。城市智能体通过对城市全域运行数据进行实时汇聚和大数据分析,实现全面感知城市生命体征;通过辅助宏观决策指挥,预测预警重大事件,配置优化公共资源,保障城市安全有序运行。城市智能体在政务服务、政务办公、城市感知、城市运行、城市治理和产业赋能等方面可以提供综合应用能力,实现整体智治、高效协同、科学决策,推进城市智能化和智慧化。

如图 5-2 所示,城市智能体参考架构是系统化的架构,由智能感知、智能联接、智能中枢、智能应用 4 层组成。城市智能体参考架构通过分层分级建设,选取合适的技术和产品,提升城市的智能化水平。

城市智能体参考架构各层之间是相互协同的,形成一个有机整体,就像人体一样,“能感知”“会思考”“有温度”“可进化”,共同服务于城市的智能化发展。城市智能体具有以下 4 个特点。

一是能感知。城市智能感知体系是城市的“神经末梢”,是城市的“视觉、听觉、嗅觉、触觉”的有机组成,是构建城市全域智能化的数字底版,通过数字孪生技术,将物理世界 1:1 还原为数字世界,使能城市的精准化感知、精细化治理、科学化决策,不断增强城市的监测预警和风险防范的综合能力。

二是会思考。城市大模型汇聚各种领域的知识、经验,具有跨领域的知识和语言理解能力,能够基于自然对话方式理解与执行任务。城市大模型的泛化能力和“智能涌现”赋能城

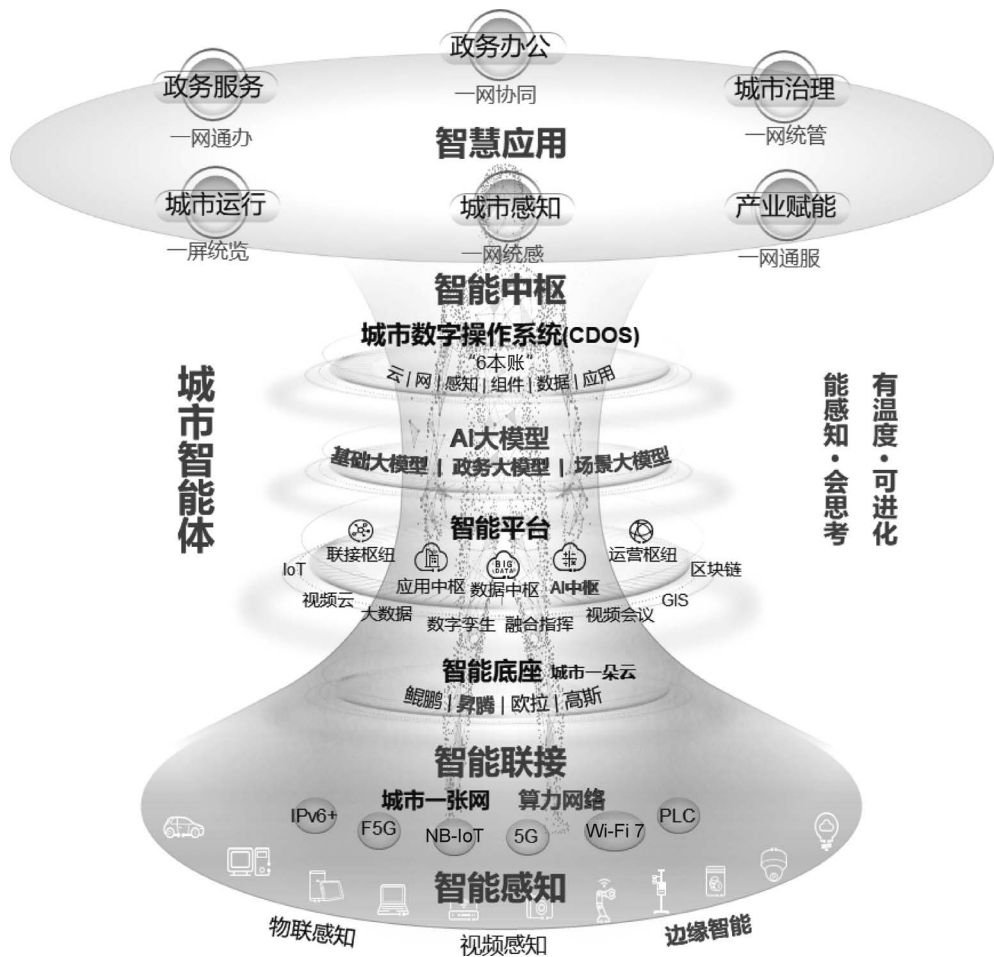


图 5-2 城市智能体参考架构

市大脑智能化升级,使得其会学习、会理解、会思考,实现从“感知智能”到“认知智能”,使城市管理逐步摆脱对人工的依赖,提高城市管理水平与服务效能,从而驱动城市激发创新活力。

三是有温度。基于 AI 大模型精准的语义理解和分析能力打造 12345 热线助手、政务服务助手等,围绕优政、利民、惠企,为企业和民众提供 7×24 小时全方位服务,保障城市居民的获得感、幸福感、安全感。

四是可进化。城市智能体将整个城市比作一台机器,训练其去完成种类繁多的任务。考虑整个城市运转的不可逆、不可停止的特性,自进化成为城市智能体核心且显著的特征。随着大模型自进化技术的发展,城市智能体自进化成为一种可能。在无人监督情况下,以大模型的思维链和自洽性特性,不断完善城市智能体模型,拓展其能力边界,实现面向场景的业务闭环和生态价值闭环,最终实现城市智能体自进化。

1. 智能感知

以城市智能感知体系建设为基础,使能城市万物智能互联、业务智慧联动。通过布设覆盖城市范围的多种类型传感器,建立全域全时段的城市智能感知体系,对城市运行状态进行多维度、多层次精准监测,全面获取影像、视频、各类运行监测指标等海量城市数据,实现对城市环境、设备设施运行、人员流动、交通运输、事件进展等的全方位感知,实时获取城市全域全量运行数据,为城市智能化转型提供数据基础。城市运行时刻处于发展变化中,必须时刻掌握物理城市的全局发展与精细变化,让城市可感知、能执行,实现孪生环境下的数字城市与物理城市同步运行。

通过城市全域的泛感知建设,实现动态的感知、精准的控制,随时随地感知城市运行动态,研判城市运行的趋势和规律,提前发现城市潜在运行风险,精准发出预警信息,为科学决策提供有效的技术支撑保障。万物互联是城市感知体系建设的前提和基础。全场景的智慧化将唤醒和千亿联接的升级,伴随着感知、联接能力全面提升,人与物将在数据构筑的智能城市环境中进行交互,以感知塑造智能、智能提升认知、认知锐化感知,推动城市智能化转型深度融合,实现可持续发展。

2. 智能联接

城市智能联接是城市公共基础设施网络,向上联接政务云,向下联接各委办局、企事业单位、学校、医院、社区等机构,统一承载了城市中千行百业的各类业务,起到了非常关键的作用。万丈高楼平地起,打牢基础是关键,只有建设一张强健可靠的网络,才能更好地支撑城市的智能化转型和高效治理,有效提升百姓的满意度和幸福感。

政务外网。它是政府办公、政务服务和城市治理的重要支撑平台,当前主要覆盖到区县一级,乡村覆盖率不足;同时各委办局专网未整合、业务未打通,数据流转不通畅,老百姓办事往往需跑到多个部门。通过 IPv6+智能网络,实现乡村简单便捷的快速覆盖以及各委办局的专网整合,让数据多跑路、群众少跑腿。

移动政务网。传统移动政务由于数据通过 Internet 公网承载,带宽受限,存在 APP 打开慢,语音/视频会议易卡顿等问题。通过打通政务外网与运营商网络间的壁垒,实现敏感政务数据高质量、高安全承载。以移动优先的理念,将政务应用逐渐转移到“指尖”,实现政府部门随时随地办文、办事、办会的需求。

城市物联网。随着城市基础设施建设的加快,物联网正在由万物互联逐步演进到万物智联,对网络提出了更高的要求。基于 OpenHarmony 操作系统构建的智能物联网,统一了各类繁杂的标准协议,实现了千万级物联终端的安全接入。

社会服务网。以城市居民为主要服务对象,面向各类教育机构、医院诊所和社区等,提供统一的公共接入和算力服务,打通各机构间藩篱,实现数据的共治共享,进一步方便人民的工作和生活。

AI 算力直连网。智能无损算力网络,通过 1ms 低时延、零丢包推动算力进一步普及,让城市 AI 算力成为像水电一样唾手可得,随时随地、即取即用。

3. 智能中枢

智能中枢是城市智能体参考架构的核心引擎,智能中枢以软硬件基础设施、人工智能基础算力资源为底层支撑,以人工智能平台、AI大模型和场景化应用算法为核心要素,通过统一的智能调度,高效支撑AI与政务办公、政务服务、城市治理等业务深度融合,协助各委办局业务智能化改造,是助力城市高质量发展的核心平台和提升智能化水平的核心抓手。城市智能中枢包括智能底座、智能平台、AI大模型和城市数字操作系统(City Digital Operating System,CDOS)4部分。

1) 智能底座

智能底座提供大规模AI算力、海量存储及并行计算框架,支撑大模型训练,提升训练效率,提供高性能的存、算、网协同。根据不同场景需求提供系列化的算力能力,适应不同场景提供系列化、分层、友好的开放能力。另外,智能底座还包含品类多样的边缘计算设备,支撑边缘推理和数据分析等业务场景。

2) 智能平台

海量的数据从智能感知层生成,经过智能联接层传输汇聚到智能平台,通过数据治理与开发、模型开发与训练,积累行业经验,最终构建具备智简创新、敏捷高效、极致体验等特点的智能平台。智能平台理解数据、驱动AI,支撑基于AI大模型的智慧应用的快速开发和部署,使能城市智能化。

3) AI大模型

AI大模型分为3层,即基础大模型、城市大模型、场景大模型。

(1) 基础大模型(L0):提供通用基础能力,主要在海量数据上抽取知识学习通用表达,一般由业界的L0大模型供应商提供。

(2) 城市大模型(L1):基于L0基础大模型,利用城市各领域数据和知识构建,面向城市管理和服务的预训练大模型,可以无监督自主学习城市领域的海量知识。城市AI大模型有以下3个特点。

① 行业重塑:AI大模型叠加行业场景,赋予行业场景更智能的处理能力,提升业务效率,降低企业成本,促进行业创新,为行业的发展注入新的生命力,重塑行业的智能化进程。

② 持续演进:在城市场景使用大模型提升业务效率的同时,也会产生大量的业务数据,这些数据再对大模型进行训练,让大模型的能力越来越强大,推理越来越准确,成为城市智能化的有力支撑。

③ 统建共享:城市各部门统一规划、统一建设,共同打造多样化多层次的大模型,构筑满足各类场景各种需要的大模型,为不同部门业务场景提供多样化的选择,服务城市智能化发展。

(3) 场景大模型(L2):基于L1模型生产出来,面向更加细分场景,满足实际场景部署的模型。场景大模型在发展过程中呈现出了行业重塑、持续演进、开放共建等特点。场景大模型包含以下4类。

① 政务办公场景大模型:如辅助公文生成、辅助签批督办等场景。

- ② 政务服务场景大模型：如政务办事、营商惠企、市民热线等场景。
- ③ 城市治理场景大模型：如事件感知、事件分拨、智慧调度等场景。
- ④ 城市安全场景大模型：如城市安全风险感知等场景。

4) CDOS

CDOS是站在城市CEO视角构建的,具有开放、多元、包容特征的城市数字操作系统。CDOS通过建设4大平台能力,即南向资源与连接能力、北向业务数字化使能能力、西向安全可靠能力、东向智能与运营能力,打造“应用一本账、组件一本账、数管一本账、云管一本账、网管一本账和感知一本账”6个一本账,构建全量、全要素的连接和多云融合统一的资源池,为城市智能化升级提供强大的底座,实现各业务场景灵活调用和共享,使能数字城市从“建设态”走向“运行态”。

4. 智慧应用

智慧应用是城市智能体参考架构的价值呈现,通过ICT和城市场景AI大模型的结合,快速创造价值,逐步实现城市全场景的智慧。

通过整合城市各类信息资源,对AI应用能力进行全面升级,调用计算机视觉、语音识别、自然语言处理等能力,实现对城市人、车、物、事、环境的全动态感知,支撑市政管理、交通管理、园林绿化、应急管理、校园管理、特种设备管理、工地管理等各领域的智慧化应用升级。

通过AI赋能城市各部门业务应用,升级城市管理模式。一方面,革新城市各类要素全域覆盖、动态感知能力;另一方面,利用大模型对政府业务系统进行改造,提升业务智慧化、人性化水平。借助大模型,切实解决城市治理和政务管理中的堵点,实现城市问题的发现、治理、监督的智能化流程改造,创新城市精准治理模式。

通过建设统一应用使能平台,推动城市智慧应用集约高效建设,普惠政府、社会的业务创新。一方面,为城市各委办局提供共性能力平台,避免各业务重复建设AI平台、独立开发AI算法造成资金浪费以及部门信息化能力不足造成的业务智能化应用创新受限等问题,根据业务部门需求,快速孵化各类AI能力,有效降低AI使用难度和建设成本;另一方面,通过城市大模型牵引政务服务伙伴,构建基于大模型的服务生态,聚合行业优秀AI创新合作伙伴,构建技术合作平台、行业AI创新孵化平台、成果展示平台、人才培养平台,共同做大产业空间。

5.2 智能化技术实现

5.2.1 智能感知

城市智能感知体系建设是以安全可控的OpenHarmony操作系统及统一的标准规范为基础,实现感知终端的互认互信,通过构建城市互联互通的感知网络、分层协同的感知平台、统一汇聚的感知大脑、纵深防护的感知安全,以及持续运营的感知中心,解决传统城市感知的底账不清、烟囱林立、数据孤岛、感知盲区等问题,实现城市动态精准感知、终端互联互通、

协议标准统一、业务分级协同、场景持续创新、数据持续运营,支撑城市治理精准感知、快速反应、科学决策。

1. 城市鸿蒙物联操作系统

1) 城市感知体系概述

城市感知体系(见图 5-3)是基于 OpenHarmony 操作系统打造城市的全场景、分布式的感知设备操作系统,通过构建统一的城市感知标准规范,保证终端设备联接协议统一、设备标识统一、接口及数据模型标准统一,近端设备互联互通,降低终端设备数字化门槛,加速城市感知终端数字化。

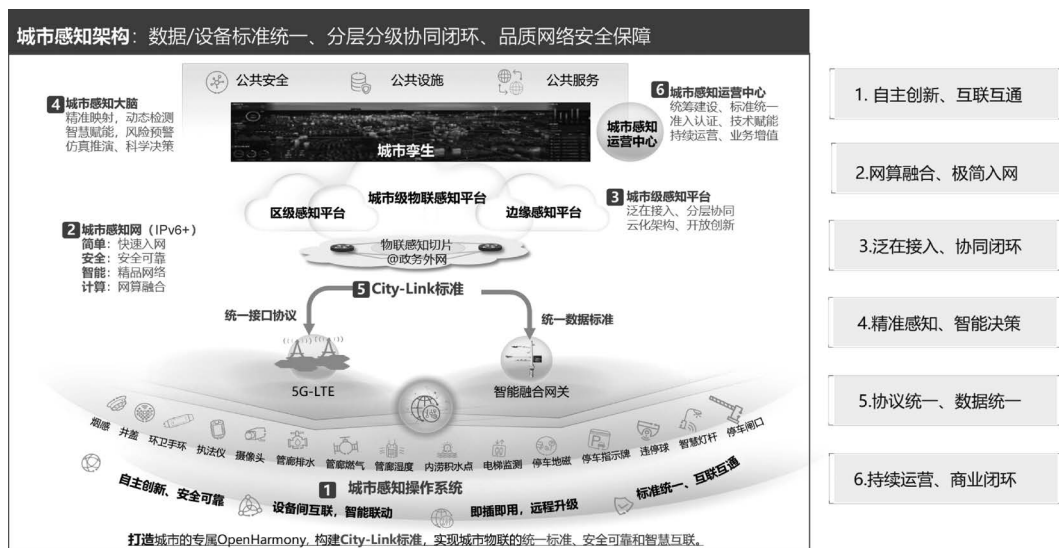


图 5-3 城市感知体系

2) 鸿蒙感知终端

针对城市感知设备种类繁多、标准不统一、设备难协同、数据难互通等问题,打造安全可控、可信的全新感知体系,让数据在设备与设备之间、设备与云端之间都能打通并自由流动,实现跨设备、跨系统、跨业务的协同服务。

统一系统。城市感知终端操作系统可以覆盖大大小小的设备设施,并且支持功能解耦,可随产品自由裁剪,支持小到传感器,大到智能控制器、巡检仪器等城市全场景各种不同类型的设备互联互通。基于 OpenHarmony 的城市感知设备操作系统,是实现这一切的最核心技术要素之一。

统一标准。城市感知数据将设备实体抽象化,统一物模型标准、设备接口标准、数据格式标准,构建横向纵向数据互通的基础。以物模型为底座,统一协议,统一管理,统一授信,有效屏蔽不同厂商之间的差异,减少开发适配工作量,降低安全风险,使数据可靠、高效的流动。

设备协同。通过分布式软总线,实现多设备联接、智能判断、自动协同来代替人工操作,使城市感知体系中的感知终端、控制器与手持终端之间可以感知靠近,并自连接、自组网,实

现配置的自动同步和传感数据的自动上报,达到节省人力、缩短周期、降低成本的目标。

安全增强。城市多元设备具备自动连接、设备间数据自由流动等特性,终端操作系统构建了一套新的协同安全、数据安全的生态秩序,实现“正确的人,通过正确的设备,正确的访问正确的数据”,满足对设备中的数据隐私与网络安全保护要求。

多端部署。城市感知体系中,不仅需要增加设备种类来繁荣产业生态,更需要不断地丰富行业服务,吸引更多合作者投入到行业应用的开发中。利用基于 OpenHarmony 操作系统,为终端提供统一编程框架、自动适配多终端硬件能力的 UI 控件,以及为不同屏幕的终端提供自适应的响应式布局,开发者可以“一次开发,多端部署”,保证了完整、多样和便携的分布式体验。

3) 鸿蒙操作系统功能

为了实现城市感知终端在不同的场景下将多种设备自动协同成一个逻辑终端,提供多端协同的使用体验,充分参考现代交互式操作系统的分层架构,终端操作系统架构从下到上依次分为内核层、系统服务层、应用框架层和应用层,如图 5-4 所示。

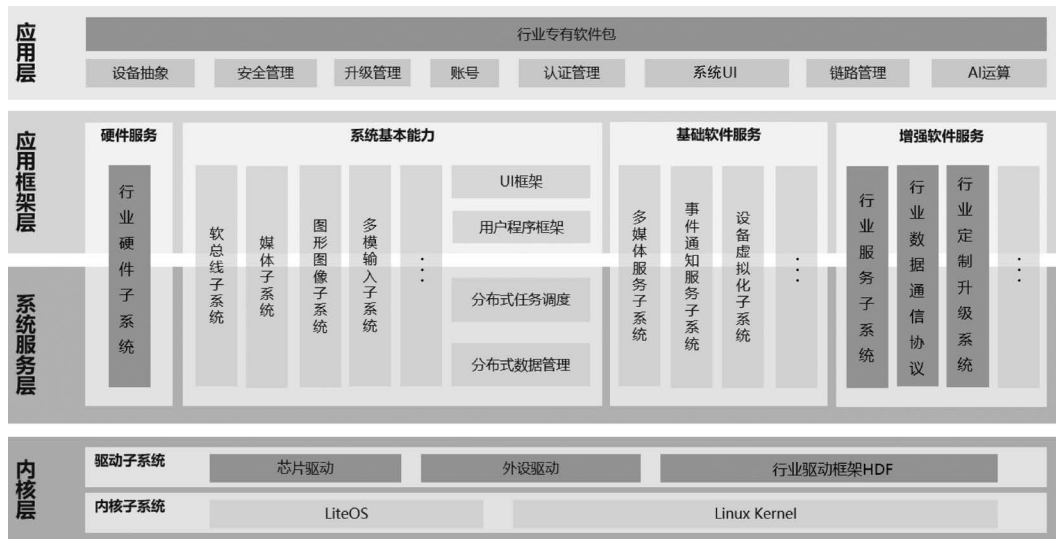


图 5-4 城市感知终端操作系统架构

内核层：主要提供对设备硬件的抽象及管理和控制能力,如进程管理、内存管理、文件系统和外设管理。

系统服务层：行业硬件子系统实现对城市感知设备硬件规格属性信息、配置属性信息的统一采集和一致性描述,全局同步,全局可见,为软件功能实现提供统一的硬件接入、查询和使能等能力;系统基本能力和基础软件服务分别为终端设备提供分布式操作的基础能力和公共、通用的软件服务;行业服务子系统为终端提供针对行业专有设备的差异化服务;行业数据通信协议针对城市感知设备定制专有数据通信协议,包括设备与设备之间、设备与云平台之间的通信协议;行业定制升级系统针对城市感知设备定制专有服务,包含带有行

业属性的系统升级包制作、下载、升级,以及针对可靠性和安全性的强化升级。

应用框架层:应用和系统交互的桥梁,为应用开发提供 JavaScript/C/C++ 等多语言的用户程序开发框架和原子化服务开发框架,能够为系统服务层的软硬件服务提供对外开放的多语言框架 API。

应用层:包括系统应用和行业应用。系统应用作为系统的一部分,向用户和行业应用提供通用的系统服务能力;行业应用是城市感知特定业务功能的各种应用,支持跨设备调度与分发,为用户提供一致、高效的应用体验。

4) 城市鸿蒙关键技术

全场景,统一内核:终端变得多样化、碎片化,操作系统如何解决终端割裂,达成跨平台、跨终端协作变得尤为关键。鸿蒙作为面向全场景的统一型操作系统,都采用了微内核架构,微内核拥有可扩展性强和安全性高的优点,可适配不同的硬件终端,如图 5-5 所示,灵活性和安全性更高,能更好地适应物联网时代的需求,有效解决物联网时代终端碎片化、安全性低的痛点。



图 5-5 城市鸿蒙操作系统设备分类

分布式软总线:实现不同领域感知设备之间近场感知、自发现、自组网,完成无感连接,多个感知设备自动协同宛如一个物理设备,可以提供任务在多个感知设备上的一致体验感,如图 5-6 所示。为了提供感知设备间协同的良好体验,分布式软总线突破了高带宽、低时延、高可靠的技术难点,在各种复杂环境里,最大限度地提升空口利用率,减少网络包头开销,优化网络包确认流程,智能感知网络变化,自适应流量控制和拥塞控制。



图 5-6 分布式软总线示例

得益于分布式软总线技术的加持,鸿蒙网关与鸿蒙感知设备的自组网,可以实现感知设备快速上线、统一管理,具备以下关键能力。

- (1) “发现”,自动搜索周围是否有相关设备。
- (2) “连接”,自动与所发现的设备建立连接。
- (3) “组网/拓扑管理”,对所有发现的设备进行网络拓扑管理,例如组成星状网络拓扑、网状网络拓扑等。

原子化服务:终端设备除了支持需要安装的应用,也需要支持免安装的应用,即原子化服务。在城市感知体系中,设备和场景都具备多样性,应用开发就变得更加复杂,原子化服务方式使开发更简单,服务更便捷。

一次开发,多端部署:原子化服务只需要开发一次,便可以部署在多种终端设备上,极大降低了开发成本。仅需为不同形态的设备配置不同参数,IDE 就能够自动生成支持多设备分发的 APP 包。APP 包上架应用市场后,应用市场会自动按照设备类型进行 APP 包的拆分、组装和分发,实现一次开发,多端部署,如图 5-7 所示。

免安装,秒级打开:使用者不感知应用安装和卸载过程,体验全新升级。原子化服务提供了全新的服务和交互方式。传统的应用需要使用者跳转到应用市场,搜索下载目标应用,而原子化服务可通过意图识别(例如扫一扫、碰一碰、语音等)、事件触发免安装能力完成应用的部署和运行,实现服务直达的业务体验,减少安装过程对用户的干扰。

多设备协同:原子化服务提供应用程序跨多个设备运行的能力,使得城市感知体系提供的运行环境不再受单一设备的能力局限,多端设备可协同。

分布式安全框架:全新分布式安全框架构建系统完整性保护、隔离和访问控制的分布式安全基础平台,如图 5-8 所示。其中包括分布式设备互信认证,支撑软总线可信连接与安

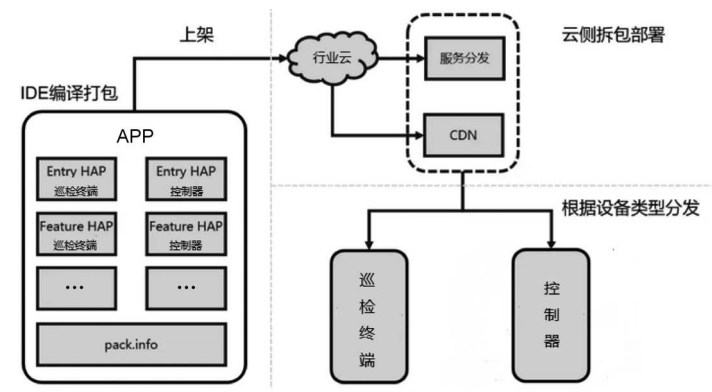


图 5-7 一次开发多次部署示例

全通信,确保设备连接可信,数据传输不泄露,防止恶意偷听;分布式用户身份认证,构建了统一的用户身份认证框架;跨设备数据安全,通过设备安全等级管理和数据分级保护能力,实现数据处理、流转时的安全管控;程序访问控制,为基于 Access Token(访问令牌)的权限管理、隔离与访问控制架构;可信安全环境(TEE-client),支撑 TEE(Trusted Execution Environment,可信执行环境)操作系统南向跨芯片平台快速部署和北向开发。

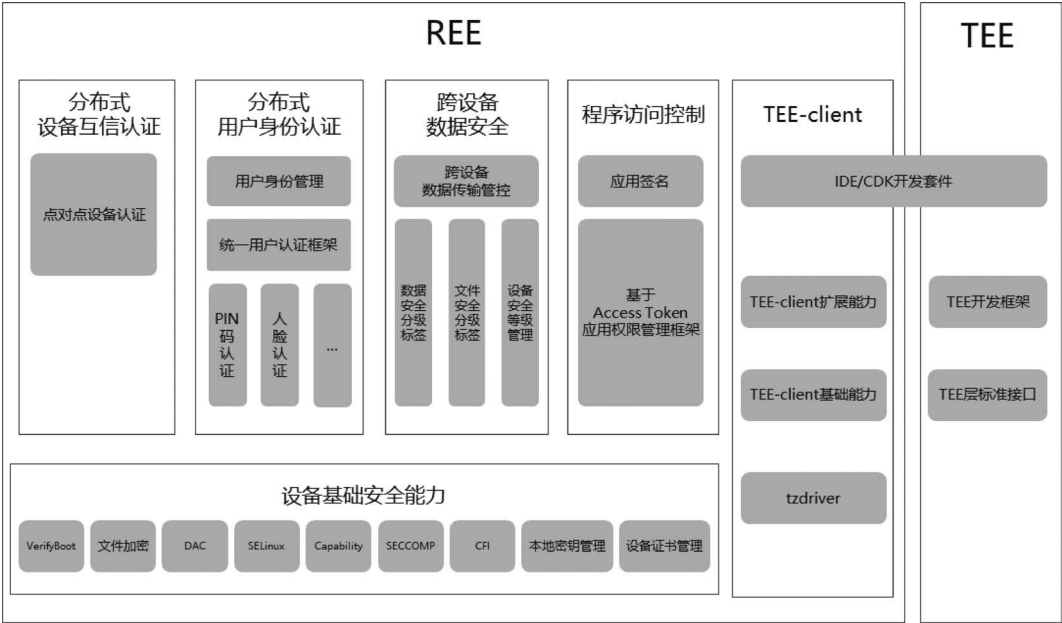


图 5-8 分布式安全框架

2. 感知终端

1) 多协议支持

感知终端分布在社会内的多个领域,有各自的归属单位、实现技术、部署时期,未来保障

感知设备的长期可使用、可管理、可演进,需要支持多种物联感知协议进行北向联接,如图 5-9 所示,最大范围地保证感知设备的信息上报。

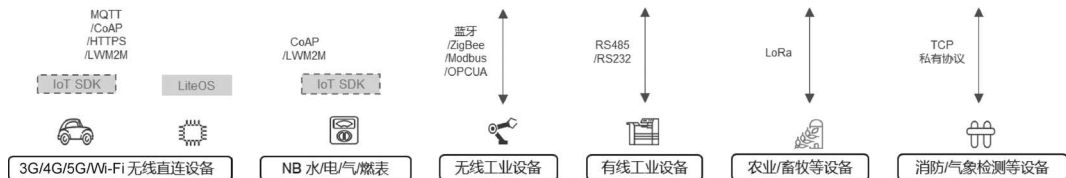


图 5-9 多协议支持

支持内容包括：

- (1) 支持通用原生协议接入：MQTT(S)、CoAP、LWM2M、HTTP/2、HTTPS 等。
- (2) 支持行业协议接入：OPC-UA、Modbus、JT808、BACnet、ONVIF 等。
- (3) 支持私有协议：通过协议插件方式实现基于 TCP 的私有协议接入。

2) 多接入模式支持

感知终端根据自身的设备硬件特点、部署位置特点、应用领域特点会有不同的接入场景,需要对多种接入模式进行支持,如图 5-10 所示。

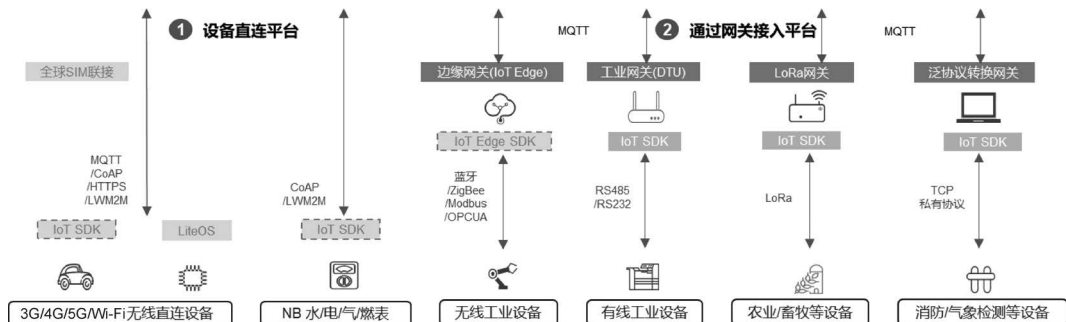


图 5-10 多接入模式支持

支持内容包括设备直连平台、通过网关接入平台。

3) 智能运维支持

随着社会的发展,感知覆盖范围越来越广、设备部署数量越来越大,传统的专用系统专人现场维护已经在经济上、效率上不具备可行性,无法保障感知设备的长期稳定运行,感知终端需要支持远程设备运维。

支持内容包括远程参数配置、远程设备诊断、远程软固件升级。

5.2.2 智能联接

1. 城市光网

城市光网是全市的数字化传输高速公路,其具备基础性、专用性、高品质 3 大特征,如图 5-11 所示。作为基础传输层网络,城市光网具备架构的稳定性,业务接入的多样性,承载能力的统一性,可统一承载敏感、非敏感的各类智慧城市业务,带宽具备无限扩展能力。同

时,不同于城市公共通信网及各类专线网络,城市光网是一张智慧城市专网,服务于泛政府各类传输需求,包括委办局政务办公上传下达、城市各类感知数据的无损回传、城市管廊生命线信息的实时获取、医疗影像的快速调阅、教育考试信息的安全传递、企业服务信息的有效获取等,该类数据的传输事关国计民生,需要一张高安全的传输专网来承载。城市光网还具备高品质的运力能力,可对如城市公共安全视频等大流量业务提供端到端的直达硬管道,对视频会议等重保业务真正实现“零丢包、零卡顿”。

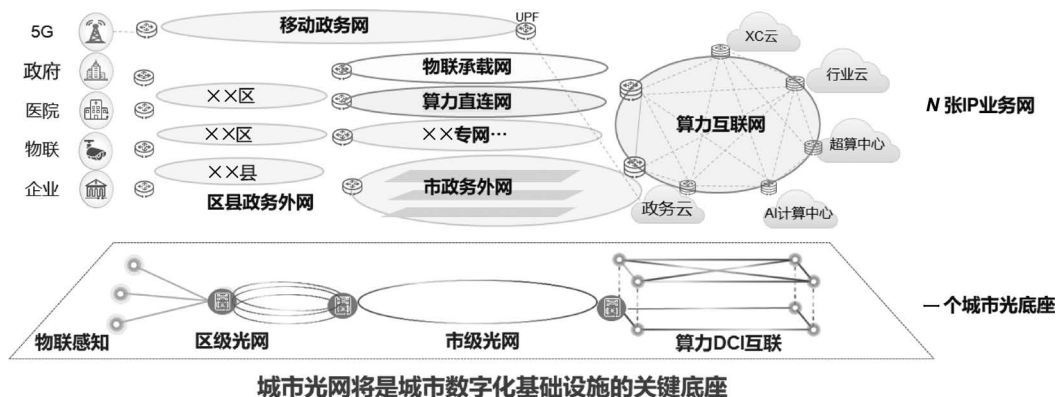


图 5-11 城市光网示意图

根据智慧城市网络的发展趋势,也对其传输网提出了明确的需求。其一是统筹集约,需要光传输网支撑多网合一,带宽按需扩展,节省财政支出。其二具备安全隔离能力,需要一张高可靠、强隔离的统一专网,独立于公用网络。其三是具备品质运力服务的能力,具备极致低时延,可自动驾驶,底层带宽可视可管。

而城市光网作为城市的基础光传输骨干网,架构稳定,具备多网合一、随需扩容的能力。基于物理隔离的硬管道技术,具备高安全的特点。同样利用刚性管道技术实现了无损传输、超低时延的运力品质。利用其打造一张百吉带宽、无限扩展的网络底座,实现全城 1ms 时延圈,成为智慧城市传输网的建设方向。

2. 政务外网

城市业务内容呈现爆炸式增长,如数字孪生 CIM(City Information Modeling,城市信息建模)、视频会议、专网整合、一网通办、一网统管、市区协同等,对政务外网的诉求多样化、定制化,网络不仅统管集约化建设实现建好网,更需要通过服务化运营用好网,让政务外网从可用能用到好用爱用;服务化运营需要显性化网络服务,明确运营主体,提升网络价值。

随着政府关键业务上云,业务内容出现了爆炸式增长,复杂的业务对网络提出多样化和定制化的诉求,包括业务体验相关的,业务相互隔离互不影响的,还有网络能够横向和纵向全覆盖,全域随时随地灵活实时接入的,这些诉求导致网络成本指数级上升。要建好网络,需要从以前的离散化烟囱式建设转向集约化建设,从顶层统一规划,实现资源共享;除了要建好网,为了满足差异化的业务体验,还要以用户/业务为中心考虑如何用好网,通过提供更多更好地网络服务和体验来不断提升网络价值,保障网络可用能用及好用爱用,更好地发挥

政务外网作为基础设施的乘数效应,并从政务外网走向政务一张网,使其具备以下特性。

架构灵活:网络基础设施和业务进行分层解耦,IP 和光网络分层,满足短期和长期业务目标,电子政务外网、视频大联网、物联承载网实现物理和逻辑分层,专网业务隔离;网络模块化和标准化,通过乐高式拼接,设备分档,网络设备新增和扩容不影响现有业务。

网络智能化:高质量满足上层服务的 SLA 和体验,实时进行流量、质量、中断的秒级感知,针对拥塞、质量劣化进行智能调优,实现 1min 发现、3min 定位、5min 恢复,同时进行 AI 主动性预测故障,保障业务无感调优恢复。

网络开放:通过存量、告警、性能、发放、测试、优化 6 大类接口能力开放,满足上层服务快速全方位感知,通过敏捷开放可编程,自定义应用实现生态开放,保障上层服务随意编排,满足不同用户业务诉求,并且做到新服务快速上线,网络即插即用,对不同服务进行 SLA 承诺。

3. 算力网络

算力网络主要解决的是低碳排放、多部门重复建设、资源利用率不高的问题,形成城市算力一张网。首先,要将政务算力、产学研算力和边缘算力等通过高速、安全、弹性的网络连接起来,形成一个城市的算力资源池;其次,提供用户接入 POP 节点,用户可以更方便地接入算力资源池,由于用户存在差异化需求,因此对于特别重要和重度算力用户,建议光纤直达,对于政务用户,可以通过政务外网接入算力资源池,对于工业用户,可以考虑通过工业互联网接入算力资源池等。算力一张网最终将使算力成为与水电一样,“一点接入、即取即用”。

1) 算力网络的边界

算力一张网既要负责算力资源的互联互通,又要完成算据的上传和结果的下载,如图 5-12 所示。因此,它包含数据中心互联(Data Center Interconnect,DCI)和用户接入数据中心(Data Center Access,DCA)。算力资源包括核心算力和边缘算力等;用户接入包括政务、科研、企业、个人等用户;对于政务业务来说,算力中心主要包括市级和区级政务云,以及城市级的智算中心和超算中心,以深圳为例,包括鹏城实验室以及国家超算深圳中心等。通过算力一张网把这些算力连接起来,形成算力资源池,供政府各委办局需求方调用;各委

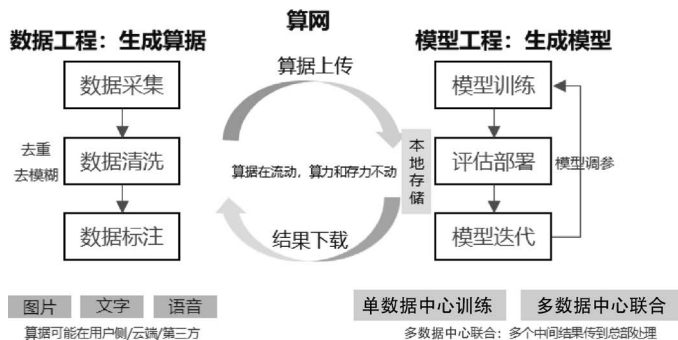


图 5-12 算力一张网

办局的用户直接通过政务外网即可接入城市所有的政务算力和 AI/HPC (High-Performance Computing, 高性能计算) 算力资源池。

2) 算力网络的特征

当前政务和科研单位已经在使用超算中心或智算中心的算力,但是由于经费有限,带宽很低,100Mb/s 带宽传输 30TB 的数据需要 1 个月时间,而如果用 100Gb/s,1 小时内就可以完成。然而,弹性大带宽至今无人可以提供,虽然有科研单位使用硬盘快递方式传输数据,但会面临硬盘丢失、被复制等信息安全风险,而且硬盘数据需要在算力中心进行杀毒、复制等也要消耗大量人力,因此亟须一张支持弹性大带宽的专网,既能快速传输数据又能确保数据安全。

首先,网络是安全的,数据传输过程中,业务是隔离的,不与互联网联通;数据是加密的,即便被看到也是密文;数据的密钥尽可能是动态的,而且无法被窃听,量子加密可以满足这一点,一旦被窃听,量子态就会发生变化,传输可以中断并查找到窃听器。

其次,网络带宽是弹性的,因为并不是所有用户都需要 7×24 小时使用网络,大部分用户都只在算据上传及结果下载时需要申请带宽资源,因此,带宽必须是弹性的,对于算据比较大的用户,需要申请弹性大带宽。

再次,网络是敏捷拆建的,可以分钟级建立一个连接,使用完后也可以分钟级拆除一个连接。

最后,网络作为一个整体对外提供服务,服务 SLA(服务等级协议)可承诺,用户体验可改进,真正做到网络易用好用。

4. 物联承载网

随着智慧城市建设进入深水区,物联感知体系相关的行业千差万别,业务需求多种多样,这就导致了各行业、单位的物联感知都是分散性建设,物联承载网也采用了不同的方式,如 NB-IoT/4G/5G 无线方式、自建专网、租用运营商专线等,分散的数据承载方式在初期可以满足城市发展的需求,但是随着各行业物联业务的增加,物联感知终端越来越多,也逐渐暴露出了资源浪费、成本高、数据共享难、安全问题频发、缺乏统一责任主体等问题,这些问题都影响了智慧城市的进一步深化和演进。

统筹规划建设的物联承载一张网是深化智慧城市建设发展的重要基础设施。城市物联承载网业务单位涉及交通、城管、燃气、水务、气象等多个部门领域,物联感知终端也包括各类摄像头、气体传感器、温湿度传感器、应力传感器、控制设备、红外雷达等多种类型的终端,呈现泛在部署、多终端、多业务承载的特点,感知终端的数据在边缘汇聚后,通过物联承载网上传到云平台或者业务平台,物联承载网除了在满足基本的时延和带宽的基础上,还需要重点考虑布局规划、多业务承载能力、成本、可靠性、可扩展性、可维护性等维度。

布局规划:根据各个城市物联感知对象、业务功能、接入用户数、数据量和网络流量等特征规划“数字网格”,将城市各区县划分成城市、小镇、社区、水域、森林等不同类型、不同层次、不同疏密度的数字网格(如覆盖 $2 \sim 3\text{km}^2$ 、 $3 \sim 5\text{km}^2$ 、 $5 \sim 10\text{km}^2$),在每个网格内,设置 POP 节点,为网格内的物联、视频等业务提供接入服务和安全防护服务,物联承载网边缘即

是每个网格内的 POP 节点。多个(10~15 个)网格对应规划一个汇聚节点,多个汇聚节点对应规划一个核心节点,整体形成三层的网络架构模式。各层设备尽量利用当地已有的机房资源,以减少网络建设部署难度。

多业务承载能力: 物联承载网需要满足多终端、多业务接入承载隔离的需求,尤其是当终端属于不同的政府、企业单位时,客户会要求网络运营方提供终端业务隔离方案,包括基于 VLAN(Virtual Local Area Network,虚拟局域网)、VPN(Virtual Private Network,虚拟专用网)技术的软隔离方案和基于网络切片技术的硬隔离方案。

成本: 物联承载网要在集约化的原则规划建设,避免建设多张物联网络产生投资浪费的问题,且可以根据城市的实际需要按业务需求分阶段建设,如优先针对重点场所、重点单位的物联需求进行建设,或者针对重点业务场景(如综合管廊、城市内涝监测、燃气监测等生命线类)进行覆盖建设,并逐渐扩展到全市其他业务场景。

可靠性: 物联承载网作为智慧城市感知底座和基础设施,同时提供数据上传、信息发布、紧急求助、公共 WLAN(Wireless Local Area Network,无线局域网)等服务,要实现感知数据的可靠传输、管理平台和终端的可靠通信,需要从网络传输、部署环境、设备功能等多方面考虑可靠性问题。技术上需要提供网络冗余保护方案,具备故障倒换时间在 50ms 以内的能力,确保业务无损。

可维护性: 网络建成后,设备部署调整、业务配置、日常监测、排障等工作量很大,需要具备可视化、自动化、智能化的维护方案,包括新设备节点即插即用、网络状态感知、故障诊断等关键能力。

可维护性: 物联承载网接入多种业务,不同的终端容易出现仿冒等问题,针对物联承载网要具备安全接入认证及安全态势感知能力。

5.2.3 智能底座

1. 高性能计算

计算能力简称算力,实现的核心是 GPU/NPU、CPU、FPGA、ASIC 等各类计算芯片,以及对应的计算架构。AI 算力以 GPU/NPU 服务器为主。算力由计算机、服务器、高性能计算集群和各类智能终端等承载。算力需要支持系列化部署,训练需要支撑不同规格(万卡、千卡、百卡等)的训练集群和边缘训练服务器;推理需要支持云上推理、边缘推理、高性价比板卡、模组和套件。并行计算架构需要北向支持业界主流 AI 框架,南向支持系列化芯片的硬件差异,通过软硬协同,充分释放硬件的澎湃算力。

2. 大规模数据存储

随着 AI 技术的兴起,尤其是大模型时代的到来,大模型驱动了大算力、大数据(训练数据、推理数据)需求的同步剧增,为数据存储带来了存算协同、存内计算、异构存储、多模态存储、海量存储、小文件存储、高性能存储、多格式存储、增量存储、指纹存储、向量存储等存储的全系列新问题,需要考虑多模态、高可扩展的、AI 原生的大规模数据存储,以应对和引领未来 AI 技术日新月异的发展。

而传统的存储容器,例如 DAS(Direct Attached Storage,直连式存储)、SAN(Storage Area Network,存储区域网络)和 NAS(Network Attached Storage,网络附加存储),存在一些缺陷:它是静态的,其设计不具备可扩展性。

(1) DAS 不能提供数据共享能力,如果多个应用需要共用同一份数据,往往需要花费大量的时间进行数据迁移,导致环境中存在多份相同的数据,并且多份数据之间同步困难,而且 DAS 不易扩展。

(2) SAN 相比 DAS 更具灵活性和可扩展性,但是 SAN 也不具备数据共享能力。

(3) NAS 系统能够给应用服务器提供统一的文件系统空间,满足多台应用服务器之间共享数据的需求。非分布式集群的 NAS 设备一般使用双控制器节点提供服务,每个节点支持特定的业务负载,当容量不够时通过扩展硬盘框的方式增加存储容量。这种方式并不完美:首先业务和节点的绑定,意味着一个业务及其关联的文件系统只在一个节点上工作,容易造成系统整体的负载不均;其次,这种系统本质上是 Scale-up(纵向扩展)的扩容方式,追求单机性能,无法做到系统性能随容量增加而线性增加。

为了解决传统存储体系结构存在的容量不易扩展、性能不易扩展的难题,智能化大数据文件存储子系统将 3 个传统的存储体系结构层组合为一个统一的软件层,创建一个跨越存储系统中所有节点的单一智能文件系统。采用全 Active(主动)的 Share Nothing 方式,系统的数据和管理数据(元数据)分布在各个节点上,避免了系统资源争用,消除了系统瓶颈;即使出现整节点故障,系统也能够自动识别故障节点,自动恢复故障节点涉及的数据和元数据,使故障对业务透明,完全不影响业务连续性。整系统采用全互联全冗余的组网机制,全对称分布式集群设计,实现存储系统节点的全局统一命名空间,从而允许系统中任何节点并发访问整系统的任何文件;并且支持文件内的细粒度的全局锁,提供从多个节点并发访问相同文件的不同区域,实现高并发高性能读写。对外提供标准的 NFS(Network File System,网络文件系统)/CIFS(Common Internet File System,通用 Internet 文件系统)接口。另外,针对大规模数据存储的冷热访问规律,提供灵活策略的数据透明迁移,自动将冷数据迁移到大容量节点,当数据访问时,自动回迁到高性能节点。

因此,智能化大数据文件存储系统在适应各种形态数据内容、大规模数据存储上,具有更强的适应性、性能和扩展性,成为智能化大数据存储的首选。

3. 操作系统和数据库

1) 操作系统

操作系统对上层应用,要屏蔽不同硬件的差异,提供统一的接口,要完成不同硬件的兼容适配,提供良好的兼容性,为应用程序的部署提供尽可能的便利;针对不同的硬件特征,操作系统进行针对性的优化,确保能充分发挥硬件的能力。通过集群通信库和作业调度平台,整合 PCIe 4.0、100G RoCE 等多种高速接口,实现多 CPU、CPU 和 GPU、NPU 的协同调度,充分释放训练处理器的强大性能,支持更快地进行图像、语音、文本等场景的 AI 模型训练。

2) 数据库

海量、格式多样的数据,追求极致的业务性能,对数据库也带来了新的挑战。为了适应业务的变化,数据库需要高性能、海量数据管理,并提供大规模并发访问能力;同时具备高可扩展性、高可靠性、高可用性、高安全性、极速备份与恢复能力。

4. 云服务

在 AI 大模型智能化时代,云与 AI 紧密结合,AI 对云原生的需求包括 IaaS 服务及相关 PaaS 等云服务,以更好地支撑智能平台的快速构建。

1) IaaS 服务

为了合理利用智能底座的硬件资源,智能底座通过虚拟化、容器化、弹性伸缩、SDN (Software-Defined Networking, 软件定义网络) 等技术,对外提供计算算力、数据存储、网络、安全等云基础服务能力,提升资源的利用效率。

AI 计算服务支持多个开源框架,包括 PyTorch、TensorFlow、MindSpore 等;支持 1400 多算子、900 多主流模型,方便用户快速迁移使用原有 GPU 模型;提供完整的端到端完整迁移工具链,以及支撑用户迁移过程中在语法开发、调试、训练等各阶段所需的工具。

智能底座提供多种数据存储服务,包括块存储服务、文件存储服务和对象存储服务。

智能底座提供云原生网络服务,面向多元、复杂的业务,一方面提供与物理网络无差别的网络性能与体验;另一方面提供安全、隔离的网络环境,具备与传统网络无差别的虚拟网络服务。典型网络服务包括虚拟私有网络 VPC (Virtual Private Cloud, 虚拟私有云) 服务、EIP (Elastic IP, 弹性 IP) 服务、VFW (Virtual Firewall, 虚拟防火墙) 服务、ELB (Elastic Load Balance, 弹性负载均衡) 服务和 VPN (Virtual Private Network, 虚拟专用网) 服务。

智能底座提供云平台+租户安全服务能力,平台安全满足等保三级认证,自身安全防护措施包括网络、边界区域、云平台自身、主机、应用层等的安全防护,主要包括防火墙、堡垒机、日志审计等安全防护措施。租户安全用于保证人工智能创新应用环境的日常运行安全,满足创新应用的安全等保要求,提供业务所需的相关安全资源的快速调度能力和安全防护体系支撑。

2) PaaS 服务

PaaS 服务为人工智能应用提供容器、微服务开发框架与微服务、中间件、数据库及大数据服务等云原生应用服务,满足人工智能创新应用快速迭代和能力构建。云平台安全服务提供平台安全和租户安全。

5. 算力开放

城市智能体致力于打造领先的、坚实的 AI 算力底座,提供多种算力供给模式来满足行业客户的差异化需求,使能“百模千态”,加速千行万业走向智能化。

采用多种算力开放模式,如图 5-13 所示。

(1) 裸算力模式:提供算力底座,包括智能感知、智能联接和智能底座。通过直接提供领先的昇腾 AI 算力,使能客户和伙伴灵活打造差异化的算力平台和 AI 服务。

(2) 多租户模式:提供智能感知、智能联接、智能底座和基础云平台,通过基础算力+



图 5-13 算力开放模式

HCS/HCSO 基础云平台能力,方便客户面向多租户提供 AI 算力。

(3) 云算力模式: 为不同租户提供叠加的 DataArts 数据治理平台、ModelArts 一站式 AI 开发平台,使能客户快速进行大模型开发。

(4) MaaS(Model as a Service,模型即服务)模式: 面向行业提供开箱即用的模型即服务,即 MaaS 模式,加速业务应用上线。

5.2.4 智能平台

智能平台提供数据治理与开发、模型开发与训练等城市级 AI 平台能力,服务行业应用构建,是海量数据的汇聚点,是城市智能体大脑和决策系统的基础。智能平台对各式各样的数据(数字、文字、图像、符号等)进行筛选、梳理、分析,并加入基于常识、行业知识及上下文所做的判断,形成支撑大模型智能分析、决策和辅助行动,助力大模型实现各行业的全场景智慧。智能平台对下发挥智能底座的能力,对上服务行业应用构建,起到承上启下的中枢作用,按技术领域和作用域进一步细分可以分为应用中枢、数据中枢和 AI 中枢。

1. 应用中枢

应用中枢是一种高效、灵活、可扩展的业务平台,能够满足对于快速创新和应对不确定性的需求,对城市实现智能化转型和提升运营效率起到重大的作用。在智能化大模型加持下,重塑应用中枢的体验,包括在数字孪生城市、CIM 等。

1) 大模型+数字孪生城市

通过数字孪生技术,将城市实体与数字模型进行紧密结合,实现城市状态的实时监测、预测和优化。大模型+数字孪生城市提供新的应用中枢,提供的能力包括:

(1) 城市数据采集与处理。通过采集城市的各种数据,如交通流量、空气质量、建筑信息等,并通过大模型处理和分析,提取出有价值的信息,用于支持城市决策和管理。

(2) 城市状态监测与预测。大模型可以对城市状态进行实时监测,及时发现异常情况,同时还可以利用历史数据和实时数据,对城市未来状态进行预测,为城市管理提供科学依据。

(3) 城市规划和设计。大模型可以用于城市规划和设计,通过对城市数据的分析和模拟,可以对城市规划方案进行评估和优化,提高城市规划和设计的科学性和准确性。

(4) 城市能源管理和优化。大模型可以用于城市能源管理和优化,通过对能源数据的分析和模拟,可以对能源消耗进行预测和优化,提高能源利用效率和管理水平。

(5) 城市交通管理和优化。大模型可以用于城市交通管理和优化,通过对交通数据的分析和模拟,可以对交通拥堵进行预测和优化,提高城市交通运行效率和管理水平。

2) 大模型+ CIM

首先,大模型具备强大的数据处理能力,能够处理海量的城市数据。通过对这些数据的处理和分析,可以提取出有价值的信息,用于支持城市决策和管理。

其次,大模型还具备强大的预测和模拟能力。通过对历史数据和实时数据的分析,可以预测城市未来的状态,从而为城市管理提供科学依据。同时,大模型还可以用于模拟城市规划和设计的效果,评估不同方案的科学性和可行性。

此外,大模型还具备强大的空间分析和可视化能力。通过将城市数据与地理信息系统(GIS)相结合,可以生成三维的城市模型,为城市管理和规划提供更加直观和科学的数据支持。

大模型+ CIM 提供能力包括:

(1) 城市交通规划。大模型可以通过分析历史交通数据和实时交通数据,预测交通流量和交通拥堵情况,为城市交通规划提供科学依据。

(2) 城市环境监测。大模型可以通过分析空气质量、噪声等环境数据,预测城市环境状况,为城市环境监测和管理提供科学支持。

(3) 城市灾害防控。大模型可以通过分析历史灾害数据和实时灾害数据,预测城市灾害风险和防控重点,为城市灾害防控提供科学指导。

(4) 城市能源管理。大模型可以通过分析能源消耗数据,预测城市能源需求和供应情况,为城市能源管理提供科学依据。

2. 数据中枢

数据中枢收集、清洗、整合和管理来自不同来源的数据,包括但不限于政务数据、公共数据、互联网数据等。这些数据在大模型的训练过程中被用作输入,因此数据中枢的有效管理和预处理对于提高模型训练的效率和准确性至关重要。其中,数据汇集、数据治理、数据预处理在大模型的训练之前是必要的过程。

1) 数据汇聚

实现政务数据、公共数据和社会数据的汇聚,在大数据中心汇聚和落地,建立大数据信息资源池,提高数据资源集中化程度和有序化水平,实现多源异构海量数据的全承载,为充分发挥大数据价值创造条件。构建“物理集中为主、逻辑集中为辅”的数据中枢。

数据来源主要包括政务数据资源、公共数据资源和社会数据资源。大数据中心的建设将汇聚并融合来自政府、公共服务管理单位与互联网等异构多源、分散多样的数据,服务于上层的大数据应用与智能化应用。

(1) 政务数据。充分利用全国一体化大数据体系的政务数据共享交换的机制,广泛汇集政务生产数据,为智能化 AI 大模型的训练提供高价值的政务数据。

(2) 公共数据。公共数据是指国家机关、法律法规规章授权的具有管理公共事务职能的组织以及供水、供电、供气、公共交通等公共服务运营单位(统称公共管理和服务机构)在依法履行职责或者提供公共服务过程中收集、产生的数据。这些数据具有海量、高价值,但这类数据往往涉及个人、企业敏感的隐私,宜采用脱敏、隐私计算等处理后进行归集。

(3) 互联网数据。中国互联网、移动互联网用户规模居全球第一,拥有丰富的数据资源和应用市场优势,大数据部分关键技术研发取得突破,涌现出一批互联网创新企业和创新应用。政府加快大数据部署,深化大数据应用,已成为稳增长、促改革、调结构、惠民生和推动政府治理能力现代化的内在需要和必然选择,但互联网数据庞大、复杂,涉及社会各个领域,不宜全面归集,建议根据业务发展需要,针对业务场景进行数据归集。

2) 数据治理

数据来自各个领域,归集的数据很难直接用于智能化的模型训练,如何使平台持续有效地运作起来,确保数据可信赖(准确、完整、一致)、可共享(方便、及时)、可应用(内容为用户所需),就必须要有科学的数据管理规范 and 机制。通过政策规范、组织、流程等手段,规范人和系统在数据产生、使用、质量等方面的管理行为,促进数据共享、提高数据质量、保证数据安全、实现数据价值。

数据中枢提供数据治理生产线,核心是从数据的集成、开发、治理到数据应用消费的全生命周期智能管理。一站式实现从数据入湖、数据准备、数据质量到数据应用等全流程的数据治理,同时融合智能化治理能力,帮助数据开发者大幅提升效率。

3) 数据预处理

大模型训练需要进行数据预处理,以提高数据的质量和可靠性、帮助模型更好地理解 and 处理数据、提升模型的泛化能力、减少模型的计算量和训练时间等。数据预处理提供可视化预处理工具平台,使能大模型的训练。数据预处理方法包括数据分组和标注、特征选择、特征转换、数据关联等。

(1) 数据分组和标签标注。根据项目需求将数据进行分组和标签标注,以便于模型训练和评估。

(2) 特征选择。根据项目需求选择相关的特征,提取出对模型训练有帮助的特征。

(3) 特征转换。将数据转换为适合模型训练的格式,如将文本数据转换为数值型数据、将图像数据转换为向量形式等。

(4) 数据关联。数据关联是连接不同数据集或不同特征之间关系的过程,它有助于增强模型的理解能力和预测能力。通过关联分析,可以发现数据之间的相关性、因果关系或其他隐藏的关系模式。

3. AI 中枢

AI 中枢作为 AI 能力的生产和集中化管理平台,沉淀 AI 资产和能力,实现数据、算法、模型等要素的复用和流程化管理。通过这种方式,实现多层次可复用,算法、模型的标准化管理,以及可复用服务封装能力,提高开发效率,避免重复开发,同时保证数据和模型的质量。

AI 中枢实现资源管控,包括计算资源、存储资源等,支持资源弹性调度;实现数据统一管理,并支持版本管理和一键部署上线,加快应用上线周期;AI 中枢提供 AI 算法训练、推理的必要基础平台支撑。

1) AI 训练平台

随着城市治理、政务服务以及各委办局传统业务的提质增效,AI 融入传统业务的场景越来越多,开启普惠 AI 的应用,但是 AI 需要在不同的应用场景下,通过场景化数据训练、开发才能比较好适应对应场景的应用,具有开发成本高,开发门槛高等难点,作为城市智能中枢的人工智能开发平台,需要进一步降低人工智能应用的开发门槛。AI 训练平台包括 AI 计算框架、数据管理、模型训练、训练可视化等核心能力。

(1) AI 计算框架。AI 计算框架使用算力,利用数据集进行算法模型的开发和训练。AI 计算框架的主要目的是把程序员从烦琐细致的具体编程工作中解放出来,从而可以将主要精力集中在人工智能算法的调优和改进上。在深度学习的发展过程中,涌现出了形形色色的 AI 框架,MindSpore、PyTorch、TensorFlow 就是其中的代表,并且各自在短时间内拥有了一批簇拥者,呈现出百花齐放、百家争鸣的局面。AI 计算框架是 AI 的根技术,当前市场上除国外主流的 PyTorch、TensorFlow 等开源框架外,国内自研的 AI 计算框架的能力和生态已快速发展,例如华为自主研发的全场景 AI 计算框架 MindSpore、百度的计算框架 PaddlePaddle 在市场知名度较高,并已成熟商用。

(2) 数据管理。AI 开发过程中经常需要处理海量数据,数据准备与标注往往耗费整体开发一半以上时间。数据管理提供了一套高效便捷的管理和标注数据框架,支持数据集版本的可视化管理,提供创建、删除、修改、发布数据集功能;在线标注支持图片、文本、语音、视频等多种数据类型,涵盖图像分类、目标检测、音频分割、文本分类等多个标注场景;团队标注可提供专业的团队管理、人员管理、数据管理,实现从项目创建、分配、管理、标注、验收全流程可视化管理。

(3) 模型训练。模型训练包括开发环境和训练管理两部分。开发环境部分通过集成开源的 Jupyter Notebook,提供在线的交互式的 AI 开发环境。一方面可以对数据处理、传统机器学习、深度学习算法进行 AI 代码研发;另一方面可以对整个 AI 研发流程进行管理和定制,使用数据处理环节产生的训练数据集、测试集,进行训练,产出模型。训练管理部分提供清晰的向导式训练过程管理,支持训练作业管理、作业参数管理和算法管理等功能,支持主流 AI 计算框架,同时支持高阶深度学习框架,具备高性能分布式、自动超参选择、自动模型架构设计的特点,能够提供更加高效、易用的 AI 开发能力。

(4) 训练可视化。为了更好地支持用户优化模型,观察 AI 训练的运行过程,需要提供对应的可视化工具,通过解析训练运行过程中产生的模型指标数据,生成对应的图表,展示作业在运行过程中的计算图、各种指标随着时间的变化趋势以及训练中使用到的数据信息。同时,提供训练日志可视化、整体资源可视化以及相应的可视化学习,支持训练状态、模型结构可视化。

2) AI 推理平台

AI 推理主要原理是基于训练好的算法模型,输入需要 AI 识别的数据流(视频、图像、语音、文本等数据),通过对应算法模型的 AI 分析、处理,给出识别的结果的过程。AI 推理平台实现对推理资源的统一管理,通过部署调用已经训练好的模型,基于输入数据实现模型结果的输出。如图 5-14 所示,AI 推理平台包括模型管理、模型部署和资源管理 3 个模块,可满足多样化、多场景的 AI 模型管理和推理需求。

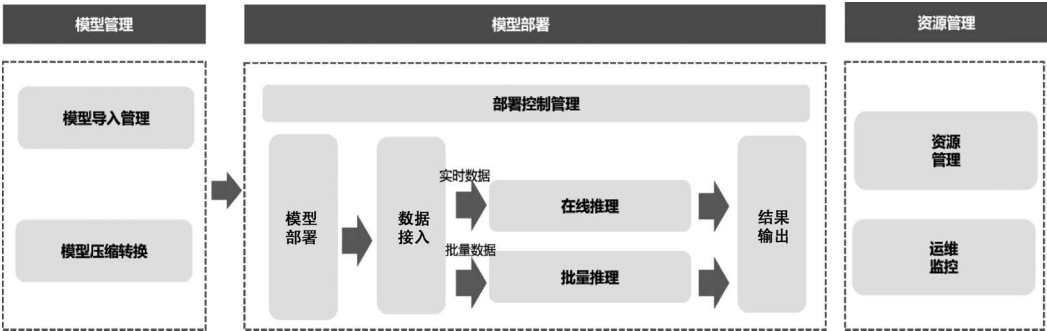


图 5-14 AI 推理平台

(1) 模型管理。AI 模型的开发和调优往往需要大量的迭代和调试,数据集、训练代码或参数的变化都可能会影响模型的质量,如果不能统一管理开发流程元数据,可能会出现无法重现最优模型的现象。模型管理可导入所有训练版本生成的模型、查询模型信息和状态、管理模型的版本等。

(2) 模型部署。在完成训练作业并生成模型,可在“部署上线”页面对模型进行部署,也可以将从 OBS 导入的模型进行部署,推理平台支持在线推理,在线推理将模型部署为一个 Web Service,并且提供在线的测试 UI 与监测能力,算法模型部署完毕后即可进行数据源的接入(如视频图像),并进行 AI 推理服务。

(3) 资源管理。智能底座的云服务为 AI 推理平台提供 GPU、NPU、CPU 等算力资源的支撑,AI 推理平台匹配 AI 用户使用的模型进行精细化的资源管理和分配。AI 推理平台提供如下几类资源模式。

① 专属资源池：提供独享的计算资源。专属资源池为某个租户独占,不与其他用户共享。

② 共享资源池：提供共享的计算资源。共享资源池提供多用户的 AI 推理作业共享运行环境,按用户 AI 推理作业创建的顺序,先到先分配资源,资源释放后重新在共享资源池

中再分配。

③ 边缘资源池。AI推理需要部署到业务的边缘,但需要考虑对前端边缘算力资源各种异常复杂管理、AI算法能力不断更新迭代,而边缘的位置决定了其AI规模不可能做得很大很复杂,因此,部署边缘设备,通过中心云的接入管理形成边缘资源池,相对地,AI平台可以创建边缘的资源池并分配给用户,把算法从中心通过云边协同下发到边缘部署。

5.2.5 AI大模型

模型的本质是对现实世界中数据和规律的一种抽象和描述。模型的目的是从数据中找出一些规律和模式,并用这些规律和模式来预测未来的结果。在机器学习中,模型是用来进行学习和预测的核心部分,通常使用训练数据来不断优化和调整模型的参数,使得模型的预测结果尽可能接近实际结果。

模型的本质可以理解为是对数据的一个函数映射,将输入数据映射到输出数据。这个函数映射可以是线性的、非线性的、复杂的或简单的。模型的本质就是对这个函数映射的描述和抽象,通过对模型进行训练和优化,可以得到更加准确和有效的函数映射。

在机器学习中,模型的本质还包括模型的复杂度和泛化能力。模型的复杂度可以理解为模型所包含的参数数量和复杂度,复杂度越高,模型越容易过拟合,即在训练数据上表现很好,但在新数据上表现很差。泛化能力是指模型在新数据上的表现能力,泛化能力越强,模型对未知数据的预测能力越好。

大模型指网络规模巨大的深度学习模型,具体表现为模型的参数量规模较大,其规模通常在百亿以上级别。研究发现,模型的性能(指精度)通常与模型的参数规模息息相关。模型参数规模越大,模型的学习能力越强,最终的精度也将更高。

盘古AI大模型是一个完全面向行业的大模型系列,围绕“行业重塑,技术扎根,开放同飞”³大方向,持续打造核心竞争力,包括5+N+X的3层架构,如图5-15所示。



图 5-15 盘古 AI 大模型框架

(1) L0 层是 5 个基础大模型,包括盘古自然语言大模型、盘古多模态大模型、盘古视觉大模型、盘古预测大模型、盘古科学计算大模型,它们提供满足行业场景的多种技能。

(2) L1 层是 N 个行业大模型,既可以提供使用行业公开数据训练的行业通用大模型,包括政务、金融、制造、矿山、气象等;也可以基于行业客户的自有数据,在盘古 AI 大模型的 L0 和 L1 上,为客户训练自己的专有大模型。

(3) L2 层是为客户提供更多细化场景的模型,它更加专注于某个具体的应用场景或特定业务,为客户提供开箱即用的模型服务。

大模型是迈向 AGI(Artificial General Intelligence,人工通用智能)的关键,每个行业、应用需要围绕大模型的能力进行重构。未来大模型演进主要有以下几点趋势。

(1) 模型参数将覆盖从十亿级至千亿级各个规模级别。未来对大模型能力的需求将越来越高,如对图片、视频,多算法统一需求、多模态问题处理能力等,千亿级参数规模大模型将会是主流。

(2) 多模态多能力大模型是未来。研究发现多模态多任务训练的大模型在各项任务中表现均比单模态单任务独立训练大模型更优,未来单个大模型会逐步统一图片、文本、视频、音频等模态,支撑更多技能,以满足生产环境中各种复杂的需求。

(3) 大模型系统的重要性越来越突出。千亿级大模型是世界上最复杂的系统,大模型构建需要依赖大量高质量数据、数千块训练芯片、人工反馈强化学习训练、数据回流等,是一个高度复杂的系统工程。未来随着模型规模及能力的不断增加,对大模型系统要求会不断提升。

1. 盘古基础大模型

1) 盘古视觉大模型

盘古视觉大模型是指使用深度学习技术训练的大型神经网络模型,用于解决计算机视觉领域的各种问题。这些模型通常由数百万个参数组成,可以对图像、视频等视觉数据进行高级别的理解和分析,例如图像分类、目标检测、语义分割等任务。

计算机视觉从传统机器学习发展到大模型阶段的历程分为 5 个阶段:传统机器学习阶段、深度学习阶段、手工模型阶段、预训练模型阶段和自监督大模型阶段。

(1) 传统机器学习阶段。在计算机视觉领域,传统机器学习算法如 SVM(Support Vector Machine,支持向量机)、决策树等被广泛应用。这些算法需要手工提取特征,因此在处理复杂的视觉任务时效果不佳。

(2) 深度学习阶段。随着深度学习技术的发展,卷积神经网络(Convolutional Neural Network,CNN)成为计算机视觉领域的主流模型。2012 年,AlexNet 模型在 ImageNet 比赛中取得了突破性的成果,使得深度学习技术在计算机视觉领域得到广泛应用。

(3) 手工模型阶段。为了进一步提高模型的性能,研究人员开始构建更大的神经网络模型。2014 年,Google 提出了 Inception 模型,该模型在 ImageNet 比赛中取得了优异的成绩。此后,研究人员不断提出更大、更复杂的模型,如 ResNet、DenseNet 等。

(4) 预训练模型阶段。为了解决数据不足的问题,研究人员开始使用预训练模型。预训练模型是指在大规模数据集上训练好的模型,可以用于解决各种计算机视觉任务。2018 年,Google 提出了 BERT 模型,该模型在自然语言处理领域取得了突破性的成果,也被应用于

计算机视觉领域。

(5) 自监督大模型阶段。自监督学习是指利用无标注数据进行训练的一种方法。在计算机视觉领域,自监督学习可以用于解决数据不足的问题。2020年,Facebook提出了DINO模型,该模型使用自监督学习方法,在ImageNet比赛中取得了优异的成绩。

盘古视觉大模型整体架构如图5-16所示。盘古视觉大模型是从2018年开始研发的全球最大视觉预训练模型,预训练大模型是解决AI应用开发定制化和碎片化的重要方法,从一个场景、一个模型的“作坊式”开发走向大模型+开发工作流的“生产线”式开发模式。

盘古视觉大模型中针对政务、工业等场景使用Transform+CNN架构开发了通用视觉模型,包括L0预训练大模型、万物检测模型、万物分割模型等,并提供L2场景模型功能包,具体细分领域有独有的工作流,可以在此基础上根据行业场景构建相应的L1/L2行业场景模型。

对于常见的视觉处理任务,例如图像分类、物体检测等场景,基于盘古视觉大模型通过自动化模型抽取、参数自动化调优等模块实现视觉大模型的训练和部署。

(1) 物体检测:基于盘古视觉大模型的物体检测工作流,实现模型自动化抽取和调参,适合任意物体检测场景。

(2) 姿态估计:基于盘古视觉大模型的姿态估计工作流,实现对人体关键点检测和姿态的估计。

(3) 视频分类:基于盘古视觉大模型的视频分类工作流,实现视频分类场景的高精度模型构建和发布。

(4) 图像分类:基于盘古视觉大模型的图像分类工作流,实现常见分类场景的高精度模型构建和发布。

(5) 异常检测:基于盘古视觉大模型的异常检测工作流,非正常即异常。

(6) 目标跟踪:基于盘古视觉大模型的目标跟踪工作流,实现对物体的位置进行检测以及跟踪。

(7) 语义分割:基于盘古视觉大模型的语义分割工作流,根据图像语义将不同区域分割成不同类别。

(8) 实例分割:基于盘古视觉大模型的实例分割工作流,对物体的位置进行检测并且识别出物体的轮廓信息。

(9) 万物分割:在大量的视觉任务中,需要对物体进行分割任务的标注,传统标注方法耗费大量的人力。此外,在图像修复、AIGC(Artificial Intelligence Generated Content, AI生成内容)等领域需要对特定的物体进行擦除和生成,因此首先需要对特定的物体进行分割,用于下游的任务。采用分割大模型可以对任意感兴趣的物体进行分割,支持自动和交互的方式对图片进行分割。

(10) 万物检测:相比于传统的目标检测,扩展了目标检测的检测类别,针对未标注训练的目标实现检测。OVD(Open-Vocabulary Object Detection, 开放目标检测)包含基于标签的检测和基于复杂语义推理的目标检测。基于标签的目标检测,用户输入待检测物体的

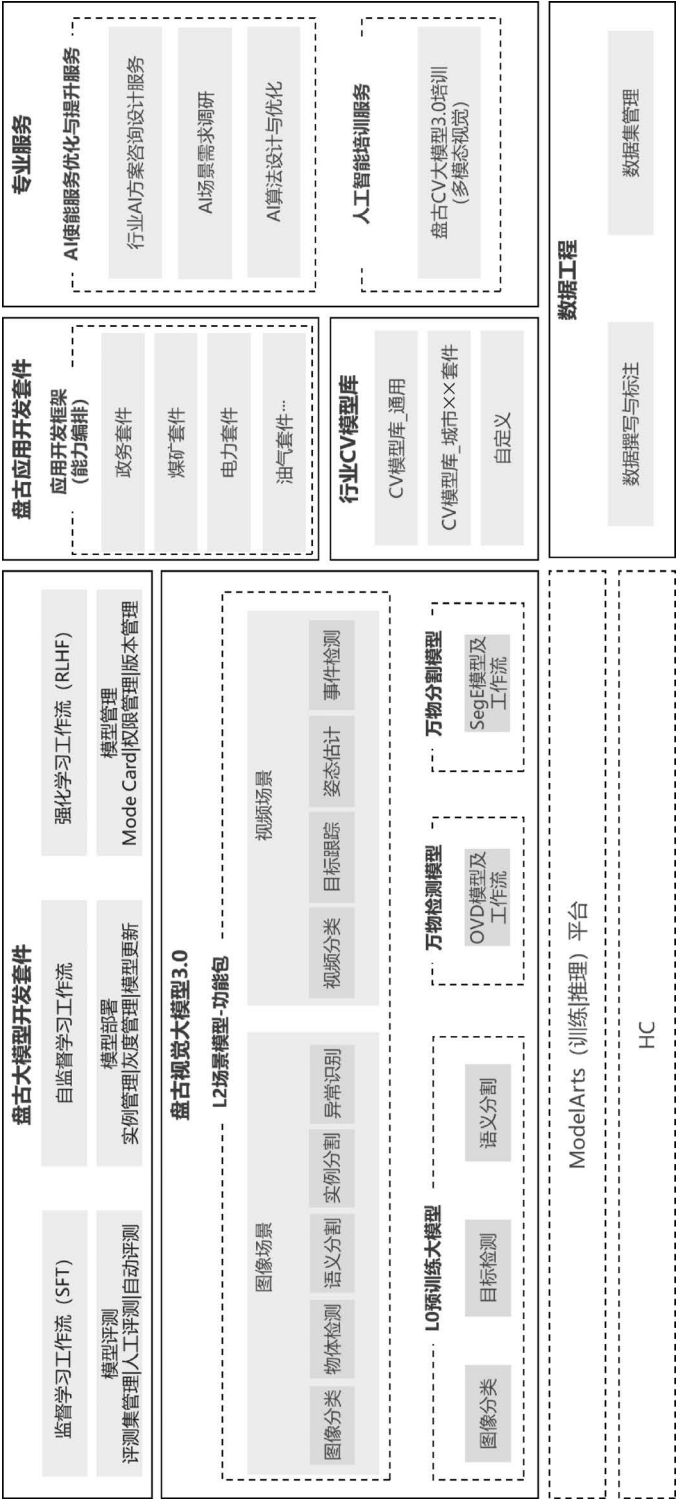


图 5-16 盘古视觉大模型整体架构

标签,算法根据标签内容自动识别图中所有待检测的物体,标签可以具有属性信息,例如方位、颜色等信息。基于复杂语义推理的目标检测,用户输入问题或者描述,算法对问题进行推理,根据推理的结果检测图中的物体。

此外还提供盘古视觉大模型开发套件、盘古应用开发套件、行业计算机视觉模型库、数据工程以及专业服务。

2) 盘古自然语言大模型

人类语言具有无处不在的歧义性、高度的抽象性、近乎无穷的语义组合性和持续的进化性,理解语言往往需要具有一定的知识和推理等认知能力,这些都为计算机处理自然语言带来了巨大的挑战,使其长期以来都是机器难以逾越的鸿沟。因此,自然语言处理被认为是目前制约人工智能取得更大突破和更广泛应用的瓶颈之一,又被誉为“人工智能皇冠上的明珠”。

自然语言处理(Natural Language Processing, NLP)大模型通常用于解决复杂的自然语言处理任务,这些任务通常需要处理大量的输入数据,并从中提取复杂的特征和模式。通过使用大模型,深度学习算法可以更好地处理这些任务,提高模型的准确性和性能。

自然语言处理发展到大型语言模型(Large Language Model, LLM)的历程分为5个阶段:规则、统计机器学习、深度学习、预训练、大型语言模型。

(1) 规则阶段。大致从1956年到1992年,基于规则的机器翻译系统是在内部把各种功能的模块串到一起,由人先从数据中获取知识,归纳出规则,写出来教给机器,然后机器来执行这套规则,从而完成特定任务。

(2) 统计机器学习阶段。大致从1993年到2012年,机器翻译系统可拆成语言模型和翻译模型,这里的语言模型与现在的GPT-3/3.5的技术手段一模一样。该阶段相比上一阶段突变性较高,由人转述知识变成机器自动从数据中学习知识,主流技术包括SVM、HMM(Hidden Markov Model, 隐马尔可夫模型)、最大熵模型(Maximum Entropy Model, MaxEnt)、条件随机场(Conditional Random Field, CRF)、LM(Levenberg-Marquardt, 莱文伯格-马夸特)算法等,当时人工标注数据量在百万级左右。

(3) 深度学习阶段。大致从2013年到2018年,相对上一阶段突变性较低,从离散匹配发展到Embedding连续匹配,模型变得更大。该阶段典型技术栈包括Encoder-Decoder(编码器-译码器)、LSTM(长短期存储)、Attention、Embedding等,标注数据量提升到千万级。

(4) 预训练阶段。从2018年到2022年,相比之前的最大变化是加入自监督学习,张俊林认为这是NLP领域最杰出的贡献,将可利用数据从标注数据拓展到非标注数据。该阶段系统可分为预训练和微调两个阶段,将预训练数据量扩大3~5倍,典型技术栈包括Encoder-Decoder、Transformer、Attention等。

(5) 大型语言模型阶段。从2023年起,目的是让机器能听懂人的命令,遵循人的价值观。其特性是在第一个阶段把过去的两个阶段缩成一个预训练阶段,第二阶段转换为与人的价值观对齐,而不是向领域迁移。这个阶段的突变性是很高的,已经从专用任务转向通用任务,或是以自然语言人机接口的方式呈现。

盘古自然语言大模型整体架构如图5-17所示,是2021年研发的千亿参数、40TB训练

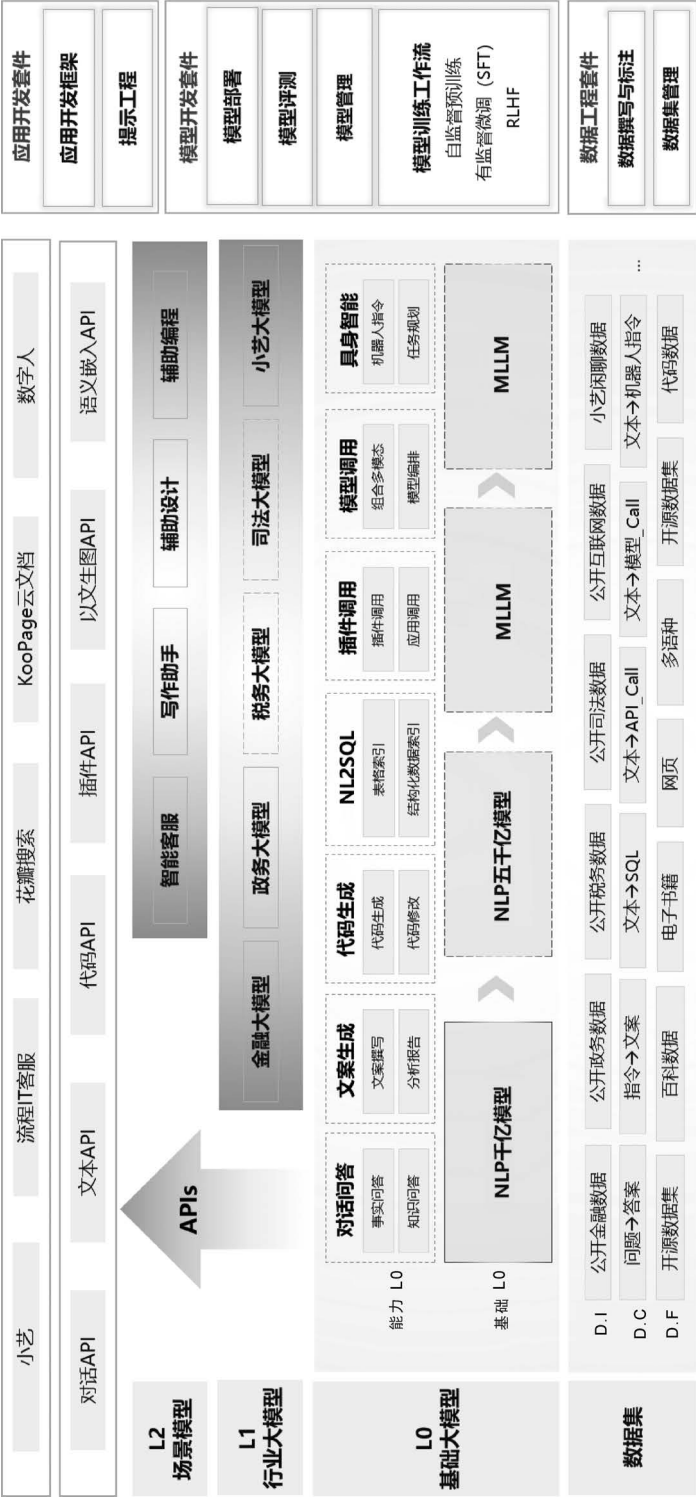


图 5-17 盘古自然语言大模型整体架构

数据的 AI 生成式自然语言大模型,也是全球最大中文语言预训练模型,支持自然语言内容的理解和生成。盘古自然语言大模型提供一整套的平台能力,支持用户开发行业大模型和场景化大模型的工具链以及配套的专业支持服务,能够帮助用户使用自己的数据高效开发,适用于本行业场景的自然语言大模型。依托于大数据和人工智能,大模型在专业行业领域的应用,除了利用文本问答、文案生成、理解生成、代码生成、插件对接等通用能力辅助办公,同时还结合行业业务特点在客服、营销、代码、应用等领域提供更专业的辅助能力。

盘古自然语言大模型主要包含如下能力:盘古自然语言大模型、模型开发套件、应用开发套件、数据工程。

盘古自然语言大模型具备的主要特性如表 5-1 所示。

表 5-1 盘古自然语言大模型具备的主要特性

能力项	场 景	场 景 描 述	
内容生成	文案生成	通用文案生成	针对一些通用场景,根据具体需求生成一篇文章。如具体主题、文字数量要求等
		公关稿件生成	面对相关部门、个人的负面舆论,生成公关稿件的草稿文案
		政府公文生成	根据具体事项,生成政府公文的草稿文案
		写作提纲生成	根据一个主题,生成写作提纲
		邮件撰写	根据办公事项场景,写一封邮件
		邮件回复生成	根据邮件上文,生成邮件回复
	建议生成	会议讨论点生成	针对会议议题,生成会议的重点讨论点建议
	表格生成	通用能力	根据一系列字段内容,生成表格
问答能力	实体识别/信息抽取	通用实体识别/信息抽取	基于一篇文章,根据需要,可以对文章中的关键信息如时间、地点、人物、国家、组织机构等进行抽取和罗列
	单轮闭卷问答	通用单轮闭卷问答	根据一些具体的通用常识问题,给出具体的一次性回答
	一定的拒绝能力	通用能力	对于包含敏感信息的问题拒绝回答
	数学能力	通用数据能力	具备基础的四则运算能力,提供常规计算问答
	多轮闲聊	通用能力	对于普通对话场景,提供多轮闲聊问答能力
	多轮知识问答	通用能力	提供多轮常识类问答以及对问答中意图和对上文问题的修改能力
	带人设的多轮问答	以职位为基础的问答	以具体角色如客服人员的口吻对问题进行回答

续表

能力项	场 景	场 景 描 述	
归纳总结	阅读理解	通用阅读理解	基于输入的文字内容,对该段内容进行针对性问题的回答
		政策、发文阅读理解与问答	基于政府官方网站的发文,进行阅读理解和针对性问题的回答
		法律法规阅读理解与问答	基于法律法规,进行阅读理解和针对性问题的回答
		案件卷宗阅读理解与问答	基于案卷卷宗,进行阅读理解和针对性问题的回答
		会议内容问答	基于会议内容(语音识别转文字后的口水稿),进行针对性问题的回答
	文本理解	通用文本摘要	基于一篇文章,输出一两句话的摘要
		新闻快讯摘要	基于一篇新闻,输出一两句话的摘要
		会议纪要生成	基于会议内容(语音识别转文字后的口水稿)生成摘要
代码生成	单轮代码生成	通用单轮代码生成	给定题目生成代码(C、C++、Java、Python、JavaScript、Go)
	代码补全	代码补全	给定目标及一段代码,往下继续写作(C、C++、Java、Python、JavaScript、Go)
	NL2SQL	通用 NL2SQL	通过大模型生成具体场景指令,自动调用相关数据库进行数据计算、统计等操作,例如经济数据的分析等

3) 多模态大模型

多模态大模型是指模型中涉及两种及其以上的模态信息。目前,大多数的多模态模型只涉及文本和图像两种模态。根据两种模态在模型上的位置可以将这些多模态模型分为文生图大模型、图生文大模型以及图图文大模型。

对于文生图大模型,即模型根据用户提供的文本生成对应的图像。在扩散模型(Diffusion Model)被提出来之前,GAN-Based 模型一直是图像生成的主要模型。由于扩散模型在图像生成方面的优秀表现,已经成为目前主流的图像生成框架。

对于图生文大模型,与文生图相反,即模型根据用户提供的图像生成对应的文本,其广泛应用在 OCR(Optical Character Recognition,光学字符识别)领域。除了简单地从图像中提取文字,更具有挑战的是理解输入的图片,并提供相对应的文本解释。

对于图图文大模型,一般是指将图像和文本一起输入模型中,其广泛应用于视觉-语言预训练任务中。与单一的语言、视觉预训练不同,多模态联合的预训练可以为模型提供更加丰富的信息,从而提升模型的学习能力。

多模态大模型技术发展主要分为以下几个阶段。

(1) 视觉特征提取阶段。在 2018—2019 年,多模态预训练开始成为最重要的一个研究

方向时,大家主要是基于目标检测的视觉特征抽取,做单/双流的图文特征融合,其中代表性工作包括单流 UNITER 和双流 LXMERT。

(2) 端到端方法阶段。进入 2020 年,开始尝试端到端的方法,因为之前的两阶段方法存在效率不高的问题,以及领域迁移的问题,其中代表性工作包括基于 ResNet 的 Pixel-BERT、E2E-VLP,以及 Transformer 的 VILT。

(3) Scaling-up 方法阶段。2021 年,开始了数据以及模型规模的 Scaling-up 方法,其中代表性工作包括 ALBEF、SimVLM、mPLUG。

(4) 大一统方法阶段。2022 年之后,大家开始基于大一统的方法,可以做单/多模态,其中代表性工作包括 Coca、Flamingo 以及 mPLUG-2。

盘古多模态大模型整体架构如图 5-18 所示。盘古多模态大模型是融合语言和视觉跨模态信息,实现图像生成、图像理解、3D 生成和视频生成等应用,面向产业智能化转型提供跨模态能力底座。具备原生支持中文,通过亿级中文图文对训、百万中文关键词训练,拥有更佳中文理解能力。具备精准语义理解精准图文描述,对齐语义理解,智能语境识别,更具自然美感。多模态多尺度训练,逼近自然美感生成内容。具备更强泛化性,强大泛化能力,适应各种复杂的应用场景和用户需求。支持行业客户二次训练专属模型,打造大模型体验。



图 5-18 盘古多模态大模型整体架构

盘古多模态大模型在 L0 层提供以文生图、以图生图、图像编辑、可控生成等能力,客户可以基于 L0 的以文生图和可控生成能力(Capabilities),以及 ModelArts 内置的便捷微调 workflow,直接构建像素风格的游戏原画场景的生成模型,并部署为 AI 应用。

此外,在 L0 和 L1 大模型的基础上,提供了大模型行业开发套件,通过对客户自有数据的二次训练,客户就可以拥有自己的专属场景模型,如宫崎骏画风模型等。

盘古多模态大模型关键价值特性如表 5-2 所示。

表 5-2 盘古多模态大模型关键价值特性

关键价值特性	子 特 性	说 明
图像生成	以图生图	自动理解输入图像的语义、构图和风格与文本输入的图像属性要求,进行图像重构
	以文生图	超 5 亿高质量中文图文训练数据。通过自然语言生成对应语义的图像。支持泛化的语言概念和丰富的风格
	可控生图	根据线稿、深度图、人物姿态等控制信号生成相对于原图的图像
	概念植入	根据少量图像的快速训练,复制相关视觉概念并根据指定概念生图
	图像编辑	可对图像进展拓展、空洞补全、根据姿态或者线稿等信号进行图像生成
图像理解	图像描述	秒级识别,精准输出图像主题,支持情景、行为等深层次语义理解。给出短语描述或者一句话描述
	图像问答	基于给定的图像进行内容和人物识别、情景和行为理解,结合问题给出精准答复
3D 生成	3D 生成-3D 空间新视角生成	根据原始视频+雷达点云(可选),基于 NeRF 生成大空间新视角视频,比传统 3D 重建+人工修模+渲染,效率提高 2 倍多,视觉真实度大于传统渲染方案;可同时生成视频、点云以及对应的标签信息
	3D 编辑-3D 空间内容增删	根据原始视频+雷达点云(可选),基于 NeRF 删除、增加物体,物体支持静态、动态物体,如静态障碍物和移动的车辆,比传统 3D 重建+专业 3D 软件人工设计+渲染,效率提高 2 倍多,视觉真实度大于传统 3D 软件方案;可同时生成视频、点云以及对应的标签信息
	3D 可控生成-设计图对齐的 3D 漫游生成	基于轻量化全景相机,在建筑项目采集视频,云端自动对全景视频进行 3D 建模、语义对齐,生成和设计图对齐的 3D 全景漫游,跨日期对比 3D 全景漫游
	3D 动作生成	根据自然语言描述,自动生成人物模型的运动序列,用于驱动 3D 数字人

说明：盘古科学计算大模型、盘古预测大模型在政务场景中应用不多,本书不作为重点说明。

2. 盘古城市大模型

城市的数字化正在经历从数治到智治的转变,由传统的统计、归纳技术向理解、预测技术演进;由传统的数据、信息向未来的知识和经验升级;由机器辅助人向机器替代人,甚至

超越人进化。盘古基础大模型提供了包括盘古视觉大模型、盘古自然语言大模型和盘古多模态大模型的基础能力,在此基础上,针对城市领域,我们为每个城市构建了专属大模型,提供政务服务、政务办公、城市治理和城市安全感知的业务智能化能力,让每个城市拥有自己的智慧源泉。盘古城市大模型的架构是开放的,支持伙伴开发丰富智慧助手,使能客户业务创新。我们在城市治理的4大核心业务领域,联合生态伙伴,构建了政务服务大模型、政务办公大模型、城市治理大模型和城市安全感知大模型,通过大模型赋能,让城市更温暖、更安全、更智能。

1) 政务服务大模型

随着公众对政务服务需求的日益多样化,个性化服务的提供变得尤为重要,需要业务系统能帮助业务部门理解其背后的真实需求,从而为公众提供更加贴心、更加个性化的服务,增强公众对政府的满意度和信任感。另外,政务服务往往涉及复杂的社会问题,这些问题通常没有简单的答案,需要综合考虑各种因素,对这些问题进行多维度、深层次的分析,从而形成准确有效的闭环。

(1) 政务热线服务差。传统热线电话端主要依靠转人工与公众进行沟通,群众诉求密集时,座席负荷高,效率低,每天人均100多通电话,接线负荷较高,工单填写烦琐,咨询事项(4000多项)涉及面广,全凭经验。

(2) 群众办事咨询难。面向个人和企业的办事咨询了解,常常出现相应事项办事入口找不到、办事指南看不懂、在线客服回答不准确等问题。在线引导差,耗时长,单事项平均在线交互达10次,机器人机械应答,交互体验差。

(3) 惠企政策落地难。企业在面对政府发布的大量惠企政策时,经常会发生政策“找不到”“看不懂”“申报难”,政府为了评估惠企成效,往往需要企业填写大量问卷调查和企业数据,耗时长,操作烦琐,成效低。

政务服务大模型以NLP技术为基础,面向这些挑战在多方面构筑能力,成为推动政务服务数字化、智能化的关键力量。

(1) 深化语义理解与文本生成。传统的文本处理方法往往基于规则或简单的统计模型,难以对复杂的自然语言文本进行深层次的语义理解。NLP大模型技术通过深度学习的方法,可以自动学习到自然语言的语法和语义结构,实现对文本更深层次的理解和分析。这一技术在政策文本分析、公众意见挖掘等方面发挥着重要作用。

(2) 提高智能问答系统的性能。智能问答系统是政务服务中的重要组成部分,能够自动回答公众的问题和疑惑。基于NLP大模型技术的智能问答系统不仅能够理解和分析问题的语义,还能从海量的信息中检索出相关的答案,并以自然语言的方式呈现给公众,大大提高了问答的准确性和效率,为公众提供了更好的服务体验。

(3) 实现文本信息的自动摘要和推荐。政务服务中经常需要处理大量的文本信息,如政策文件、公众反馈等。大模型技术可以实现对这些文本信息的自动摘要和推荐,帮助工作人员快速了解重要内容和关键观点。不仅可以提高工作效率,还能确保工作人员对信息的全面和准确掌握。

(4) 强化多模态数据处理能力。随着多媒体技术的发展,政务服务中涉及的数据类型也越来越丰富,包括文字、图像、语音等。大模型技术结合计算机视觉、语音识别等技术,可以实现多模态数据的联合处理和分析,为政务服务提供更全面的数据支持。

基于政务服务大模型在这些方面构筑的能力,面向政务服务场景打造了12345热线智能助手、政务服务智能助手和营商惠企智能助手。

(1) 12345热线智能助手。针对政务热线的咨询应答、受理填单、派发处置、回访办结、质检归档、数据治理6个业务场景构建了智能客服、智能座席助手、智能回访、智能质检、智能数据分析5项关键智能化应用。

(2) 政务服务智能助手。针对政务办事过程的咨询、申报、受理、审批4个业务场景构建了智能引导、智能辅助申报、智能辅助审批3项关键智能化应用。

(3) 营商惠企智能助手。针对惠企政策“找不到”“看不懂”“申报难”,构建了政策智答、政策精准匹配、惠企成效分析,以及惠企政策推演的4个关键场景应用。

随着智能化技术在政务服务应用场景的不断应用,政务服务大模型将在政务服务中发挥更大的作用,推动政府数字化转型和服务质量的持续提升,具体业务价值体现在以下3方面。

(1) 优政。热线服务更有“效率”,咨询问答准确率提升到95%,工单整理时间缩短80%(平均填单时长降低2min)。

(2) 利民。群众服务更有“温度”,咨询交互次数从10次降到2次,首办成功率为85%。

(3) 惠企。政策服务更有“成效”,政策解读从月级缩短到1~2周,单篇平均耗时5min以内,精准匹配企业,政策主动推送,政策、企业匹配准确率达99%。

2) 政务办公大模型

当前,虽然很多政府办公流程都已通过数字化进行管理,但因智能化的缺失,各类政务办公业务依然面临着很多痛点和难点,政务办事效率低,低端重复劳动高,系统难熟练应用与掌握,办公体验差,主要体现在:

(1) 办文过程中。传统公文人工拟稿耗时1~2周,耗时长,周期长,重复劳动多,公文量大,阅读、处理时间长。

(2) 办会过程中。传统政府部门办公会前准备时间长,会后需要大量的时间整理领导发言、会议要点、待办任务等,人工纪要效率低。

(3) 办事过程中。传统公文任务靠人摘录和督办跟踪,效率低,易遗漏,体验差。

在人工智能蓬勃发展的浪潮中,通过人工智能实现政务办公智能化。政务办公大模型是为政务办公场景设计的大型人工智能模型,具备处理和分析大量政务数据的能力,从多方面优化办公流程、提高办公效率并提供更智能化的决策支持。通过大规模的海量数据进行预训练后,大模型可以按需产生原创内容,不仅可以分析或分类现有数据,还能创建全新内容;在进一步提供政府行业领域专业数据的训练和调优后,可让大模型技术快速适应或微调以贴合政府办公领域的应用场景。目前已经结合政府行业文事会办公特点构建政务办公智能助手,基于大模型实现“一句话办文”“一句话办会”“一句话办事”等智能化应用,加速政

府行业的智能化升级。

(1) 一句话办文。借助大模型的信息精准检索、文稿自动生成、内容智能抽取能力,实现全流程的辅助办文,根据关键词一键生成完整公文草稿,同时对政府公文格式、文本、敏感词等常见错误进行快速校对,还可以从公文阅读、摘要提取、资料推荐到草拟意见等审批流程中辅助领导批示。

(2) 一句话开会。借助语音自动识别并转写文字的能力及语义上下文内容的理解能力,可自动生成会议符合公文范式纪要或摘要,并自动识别会中提及的待办任务等,极大提高会议效率。

(3) 一句话办事。大模型支持文件到事项的辅助分解和分拨,智能抽取事务流程并生成应用,从而提升机关办事数字化水平,为政府办公带来了极大的便利和效益。

具体业务价值体现在以下3方面。

(1) 辅助快速拟文。拟文周期可由原来的1~2周缩短为1~2天,5000多字的公文1min内生成,极大提升拟文效率。

(2) 提升会议效率。语音转写与自动纪要助力会议效率提升,要点秒级提取,纪要自动生成,会议纪要撰写时间由原来1~2天,降低至分钟级,极大提升办会效率。

(3) 提升办事效率。对话式办公及智能生成应用重构办公模式,公文任务自动提取,并一键式导入督办系统,方便后续事项跟进,缩短事项处理时间,提升事项跟踪效率,避免遗漏,提升办事体验。

3) 城市治理大模型

城市治理面向区域内的问题和事件,对于采集、分拨、处置、结案各环节通常采用人工经验判断的方式进行处理,往往被动响应,处置力量到达现场时间较长,事件处置效率低,群众满意度差,主要体现以下方面。

一是人工巡查时效差、强度高。城市事件发现依靠人工巡查,工作强度大。固定点位摄像机无法覆盖城市全场景。

二是工单分拨效率低、易出错。城市治理工单量大,月均数万件。工单分拨工作繁重,易出错,效率低。

三是疲于处置,被动治理。民意信息杂,人工统计难,原因分析难。群众“急难愁盼”问题处置效率低。

城市治理大模型的关键技术主要包括深度学习、自然语言处理、计算机视觉、多源数据融合等技术,这些技术在城市治理大模型中发挥着重要作用。通过使用人工智能大模型技术,促进城市治理业务与人工智能的不断融合,实现域内事件处理全流程智能化,包括事件智能发现、城运工单自动分拨、事件问题智能分类,充分提升事件的处置效率,提升城市治理智能化水平。

从事前、事中、事后3个阶段分别构建事件智能感知助手、工单智能分拨助手、效能智能分析助手和智能问数助手,赋能和改善传统治理模式,强化事件及时发现、问题及时处置、事后科学评价的闭环管理。

(1) 事件发现阶段,事件智能发现(事件智能感知助手)。利用视频 AI 技术自动发现事件,不仅能够帮助基层工作人员实现市政、环卫、违停等多个场景事件的识别,还能够自动上报至事件中心完成派遣,从而有效提高事件的处置效率,减轻基层工作人员的工作压力。

(2) 事件处置阶段,工单智能分拨(工单智能分拨助手)。运用 AI 技术赋能工单自动分拨,能够快速将事件与权责部门进行匹配,同时为复杂事件提供相似事件的解决建议,大幅缩短人工派单时间,提升事件处置精确度和处置效率。

(3) 事件处置完成,效能智能分析(效能智能分析助手 & 智能问数助手)。运用 AI 技术对处置效果进行跟踪、挖掘热点问题、挖掘责任主体和问题根因,持续提升政府服务效率,增强人民群众的获得感和满意度。

具体业务价值体现在以下方面。

一是巡查永久在线,问题一网打尽。视觉大模型盘活城市视频资源,一图多识,7×24 小时在线巡查。无人机巡查,及时发现楼顶、工地等盲区的问题,覆盖率超 90%。

二是工单自动分拨,问题快速处置。工单秒级分拨,自动分拨率、分拨准确率均达 91%。民意速办,紧急工单 24 小时内闭环。

三是效能分析,支撑城市精细化治理。大模型工单分析,自动生成日报、周报、专报,辅助决策。大模型问数,识别管理者查询意图,数据、图表“一呼即出”。

4) 城市安全感知大模型

随着城市的不断发展、城市治理的复杂度也在不断增加,近年来,城市安全事件的频繁发生也说明了传统治理模式面临着越来越大的困难和挑战,主要体现在:

一是城市安全疲于事后应急,缺乏事前预警。传统城市安全事故感知能力弱,成本高,往往亡羊补牢、事后应急,造成损失大。

二是安全专项检查全靠“人拉肩扛”。执行城市安全专项任务,依靠人工巡查,工作量大,难以全面覆盖。

城市安全感知大模型是专为解决城市安全领域的问题而设计构建的一种大型人工智能模型,结合自然语言处理、计算机视觉、多源数据融合等多种先进技术,打造城市安全感知智能助手,为城市安全提供全方位的解决方案,确保城市居民的生活安全和城市的稳定发展。

从数字化、智能化角度来看,需要从安全风险预防、应急响应、安全数据分析等多个维度为城市安全构筑智能底座。

(1) 安全风险预防。如何预测和识别城市中的潜在安全风险,如交通事故、火灾、犯罪活动等。

(2) 应急响应。在突发事件发生时,如何快速响应并进行有效的资源调配。

(3) 安全数据分析。如何从海量的数据中分析出与安全事件相关的关键信息。

具体业务价值体现在以下两方面。

(1) 盘活视频资产,打造安全卫士。视觉大模型智能泛化分析,安全事件一图多识,泛化识别算法支持 100+安全事件,实现城市安全事件“弱信号”识别,负一秒预警,极大减少城市安全隐患。

(2) 安全专项任务, AI 减负提效。安全巡查任务自动执行, 减少人工投入, 周期由原来 1 周缩短到几分钟, 实现安全隐患精准秒级检索, 极大地提升了安全巡查效率。

5.2.6 城市数字操作系统

1. 概述

1) 建设背景

当前, 数字化改革逐步进入深水区, 统筹集约化成为趋势, 为进一步助力政府、企业、产业数字化转型, 加快构建数字城市新发展格局, 推动高质量发展, 从国家到地方做出了一系列战略部署、总体要求和相关规划。

(1) 国家战略部署: 《国务院关于加强数字政府建设的指导意见》(国办发〔2022〕14 号) 提出构建结构合理、智能集约的平台支撑体系, 构建开放共享的数据资源体系, 构建协同高效的政府数字化履职能力体系; 《全国一体化政务大数据体系建设指南》(国办发〔2022〕102 号) 提出数据资源一盘棋、一体化融合、一体化共享和开发利用趋势。

(2) 广东省总体要求: 《广东省数字政府“十四五”发展规划》提出建立全省统一政务应用超市, 建立全省业务中台服务“一张清单”; 《广东省人民政府关于进一步深化数字政府改革建设的实施意见》要求构建政务信息化项目一体化管理工作机制。

(3) 深圳市总体规划: 《深圳市数字政府和智慧城市“十四五”发展规划》围绕打造数据驱动、一体协同、智能高效、安全可控的鹏城自进化智能体, 构建标准统一的数字底座、集约高效的智能中枢和泛在连接的统一门户, 高效赋能全场景智慧应用, 实现全域感知、全网协同、全业务融合、全场景智慧, 全面支撑“四位一体”的数字深圳建设。

2) 现状与挑战

随着数字化平台的逐渐成熟, 千行百业的数字化转型逐步进入深水区。政府、企业、园区等部门和组织都认识到平台和技术能力要扎扎实实地落在应用场景中才能发挥最大的价值, 因而推动数字技术与业务场景的融合变得尤为重要。

智慧城市的建设主体负责构建一体化数字平台和统一技术底座, 为进一步发挥城市数字底座在赋能场景应用创新方面的作用, 要持续开放数字平台, 沉淀共性能力, 并通过数字资源运营满足持续变化的业务需求, 赋能基层治理效能, 打造高价值的场景化应用解决方案, 但由于资源底数不清、技术能力碎片化调用难、共性能力沉淀不足, 导致在数字资源赋能应用创新方面存在诸多挑战, 如:

(1) 资源治理, 家底不清晰。数字资源分部门、分层级多头管理, 缺乏整合, 缺少数字资源管理及申请平台, 难以做到数字资源“可看、可用”, 数字资源家底不清晰。

(2) 资源共享, 配置调度难。市区两级平台一体化程度不够, 资源配置不均衡, 跨区共享不畅通, 配置效率不高, 难以统筹调度; 资源共享交换平台不完善, 数据资源缺少深度分析, 资源价值未充分释放。

(3) 应用敏捷, 分散构建周期长。应用低水平重复开发, 各地各部门独自建设、分散管理, 共性轻应用缺乏共享, 难以全市统筹协调, 业务应用智能化程度不高, 缺乏通用服务沉

淀,场景化应用全码开发构建周期长。

(4) 业务提效,效果评价难。业务应用建设效果评价难量化,数字化项目建设、管理存在“黑匣子”,建设效果评估缺乏客观数据支撑。

(5) 持续运营,重建设轻运营。缺乏运营,成效难以保障,建设周期长,成效显露慢,且难以持续。

(6) 机制保障,创新不足。机制保守,支撑保障不足,难以支撑业务优化、整合和创新;组织保障不足,难以快速协调解决分歧与争议。

3) 业务全景

CDOS 主要有两条业务主线：一条是传统的信息化项目建设业务线；另一条是数字化供给侧结构改革带来的数字资源供给业务线。两条业务主线相互作用,形成正向循环相互促进,保障城市数字化建设运营效果。CDOS 业务全景如图 5-19 所示。

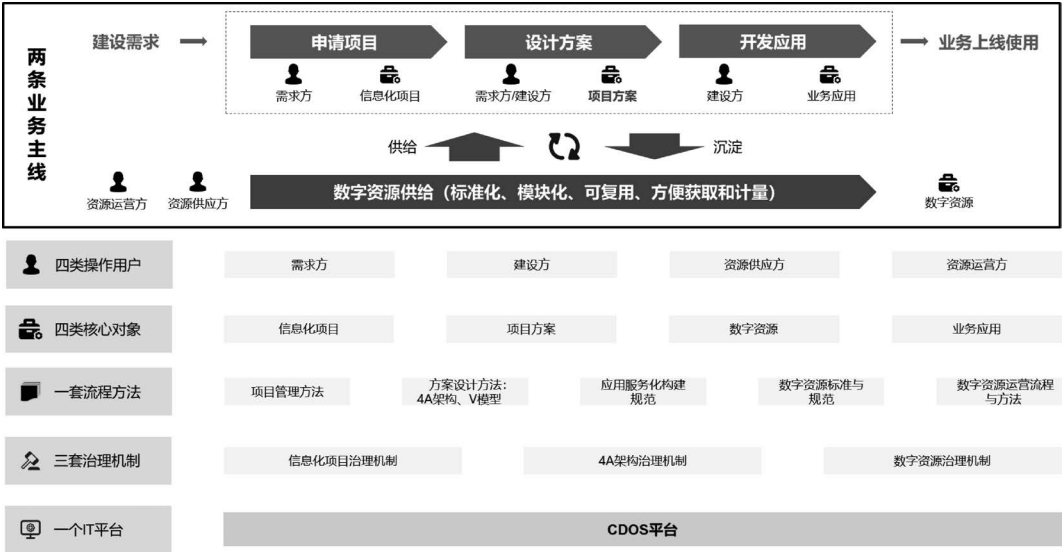


图 5-19 CDOS 业务全景

业务围绕四类操作用户和四类核心对象展开。

(1) 四类操作用户：需求方、建设方、资源供应方、资源运营方。

(2) 四类核心对象：信息化项目、项目方案、数字资源、业务应用。

构建一套流程方法来端到端的指导作业过程。

(1) 项目管理方法。

(2) 方案设计方法：4A 架构、V 模型。

(3) 应用服务化构建规范。

(4) 数字资源标准与规范。

(5) 数字资源运营流程与方法。

建立三套治理机制来消除运作过程中出现的分歧和风险,有效保障业务顺利流转。

(1) 信息化项目治理机制。

(2) 4A 架构治理机制。

(3) 数字资源治理机制。

所有城市数字操作业务承载在一个 IT 平台,即 CDOS 上,保障业务可落地。

4) 定位及愿景目标

CDOS 以赋能城市数字化转型为愿景,以数字资源共建、共治、共享为目标,以持续运营为手段,以专业方法为指导,以组织机制为保障,以 IT 平台为承载的数字化业务治理框架。CDOS 定位及愿景如图 5-20 所示。

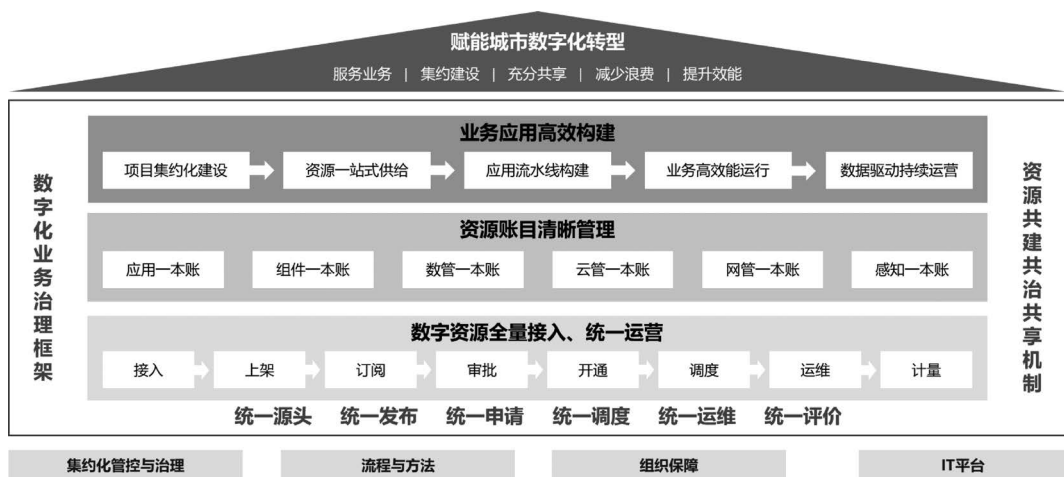


图 5-20 CDOS 定位及愿景

2. CDOS 方案架构

CDOS 方案以“共建共享,创新赋能”为核心理念,链接政数局、委办街镇用户及开发伙伴,通过数字化赋能推动数字及技术开放共享,创新应用敏捷构建,实现技术与业务的融合,提升多跨融合场景协同管理与服务,让智慧城市建设运营成果真正发挥价值。

如图 5-21 所示,CDOS 总体架构包括三部分:变革咨询/方法导入、平台建设及持续运营。

1) 变革咨询/方法导入

数字化转型要想成功实现,前期变革咨询和方法导入尤为重要,这是帮助和客户统一思想,转变观念,改变和再造组织形态、业务模式,优化业务战略和运作机制的前提,也是后续变革顺利实施的基础,变革咨询/方法导入包含 CDOS 落地组织阵型及运作模式、方法论咨询和导入,CDOS 集约化管控治理体系。

一是 CDOS 落地组织阵型及运作模式:设计和构建数字化变革相适应的组织阵型和运作模式,包含业务应用的规划、设计、开发、运营和运维,数字资源的规划、设计、开发、运营和运维,核心是打破业务提需求,IT 负责实现的传统模式,实现业务和 IT 一体化运作。



图 5-21 CDOS 总体架构

二是方法论咨询和导入,包括以下内容:

(1) 4A 架构设计方法论: 从价值流到 BA、IA、AA、TA。

(2) V 模型服务化设计方法论: 业务能力化、能力场景化、场景活动化、活动对象化、对象服务化、服务系统化。

(3) 服务化交付方法论: 应用服务化构建规范、DevOps。

(4) 资源运营方法论: 数字资源标准与规范,数字资源统一注册、上架、申请、开通、计量、运营、运维。

(5) 生态运营方法论: 生态分类与认证标准,生态准入、注册、培训、结算、评价、退出。

三是 CDOS 集约化管控治理体系,包括以下内容。

(1) 信息化项目治理体系: 数字化转型/智慧城市专班、信息化项目管理委员会。

(2) 4A 架构治理体系: 智慧城市专班、架构管理委员会、BA/IA/AA/TA 架构分委会。

(3) 平台治理体系: CDOS 委员会。

2) 平台建设

CDOS 为城市数字化转型提供强大的底座,实现各业务场景灵活调用和共享,使能数字城市从“建设态”走向“运行态”。CDOS 包含四大平台能力,即南向资源与连接能力、北向业务数字化使能能力、西向安全可靠能力、东向智能与运营能力以及应用服务市场,实现应用一本账、组件一本账、数管一本账、云管一本账、网管一本账和感知一本账的全量全要素的连接和多云融合统一的资源池。

CDOS 的特点如下。

(1) 开箱即用的 IT SaaS 服务,预集成的可 IT(认证、办公、客服等)服务,可 SaaS 服务。

(2) 降低复杂基础设施的使用难度,联接多云、IoT 等,存量应用整体上云,资源弹性高效使用,全球等距体验一张网。

(3) 稳定、可靠、安全的运行服务,高可用、高可靠的韧性服务,分级、分域弹性安全防护体系和智能安全运营。

(4) 复用数字化实践资源,应用/数据/AI/资源开发流水线,IT 中央集权和 ITGC/AC 200+治理规则。

(5) 企业价值场景的用数和赋智能力,企业数据资源共享、治理和开发的平台,AI 场景化资源和算法模型。

(6) 可分可合,异构可信的架构和能力,灵活编排适配不同场景。

(7) 使能构建开放数字化生态,开发者的生态,价值链生态。

3) 持续运营

(1) 业务/应用运营:围绕应用规划、设计、开发、上线、运维、运营等全生命周期的运营服务,支撑应用业务目标达成。

(2) 平台运营:围绕做厚 CDOS 平台,提供数字资源规划、管理、效能分析等运营服务,持续沉淀数字资源。

(3) 生态运营:为入驻平台的厂商提供准入资质评估、培训认证、服务结算、退出等运营服务,建立良性的生态圈,激发市场活力,激励厂商贡献能力和服务。

3. CDOS 平台能力

1) 南向资源与连接平台

南向资源与连接平台是数字资源治理和运营的统一平台,实现城市数字资源全生命周期可视、可溯、可管、可用、可评,并对外提供一致性的 IT 服务体验(一键申请、一键订阅、一键开通、一键部署),如图 5-22 所示。

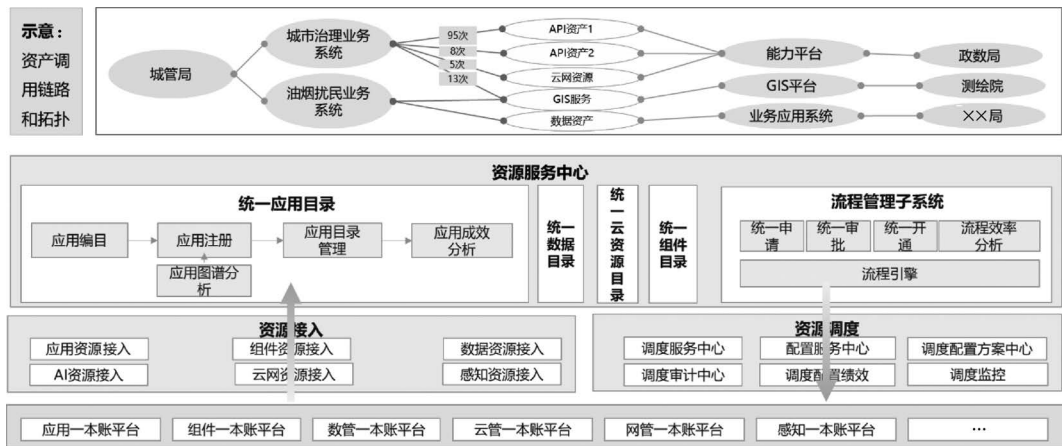


图 5-22 CDOS 南向资源与连接平台

资源与连接平台包括资源接入、资源调度和资源服务中心 3 部分。

(1) 资源接入:实现应用资源、组件资源、数据资源、AI 资源、云网资源、感知资源等各类资源的接入,并与 CDOS 可以进行无缝连接和互操作。

(2) 资源调度：实现的资源服务的统一调度和配置,包含调度服务中心、配置服务中心、调度配置方案中心、调度审计中心、调度配置绩效、调度监控等。

(3) 资源服务中心：实现各类资源从编目、注册、管理等统一服务,包括统一应用目录、统一数据目录、统一云资源目录、统一组件目录以及流程管理子系统。

2) 北向业务数字化使能平台

业务数字化使能平台是一站式、集成式的应用开发平台,使能业务重构和应用现代化,如图 5-23 所示。

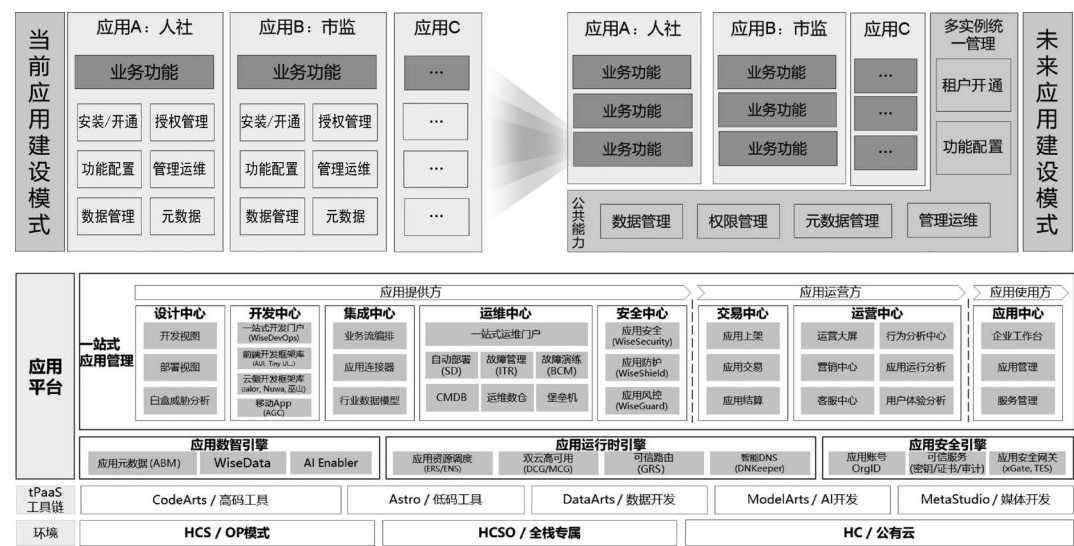


图 5-23 CDOS 北向业务数字化使能平台

传统应用建设模式从各业务部门应用独立建设,从业务功能安装、开通、配置、管理、授权、运维等都单独开发建设,既造成重复的投入和资源浪费,又不利于资源的共享和业务协同,而基于一站式应用开发平台,打通应用提供方、应用运营方、应用使用方全链条,通过打造相应的能力中心,沉淀共性的需求、服务和能力,使能业务应用敏捷开发,实现集约化的应用开发和部署,节省资源投入,缩短了开发周期,加快了业务系统上线效率。

3) 西向安全可信平台

安全可信平台构建零信任环境下安全可信,守护云、网、应用和数据安全底线。平台提供 9 大安全可信“盔甲”：可信用户管理、可信应用构建、可信生产环境、可信运维、可信联接与通信、可信数据所有权、可信安全防护、可信安全作战、可信智能大脑,如图 5-24 所示。

4) 东向智能与运营平台

智能与运营平台提供全局运营、业务应用运营、CDOS 平台运营、IT 运维服务/IT 安全运营、空间运营等,打造实时、可视、可管、可控的运营/运维指挥中心,如图 5-25 所示。

5) 云管一本账

基于统一多云管理平台,打造“管理一朵云、运营一朵云”的体验,打造全市一朵云,

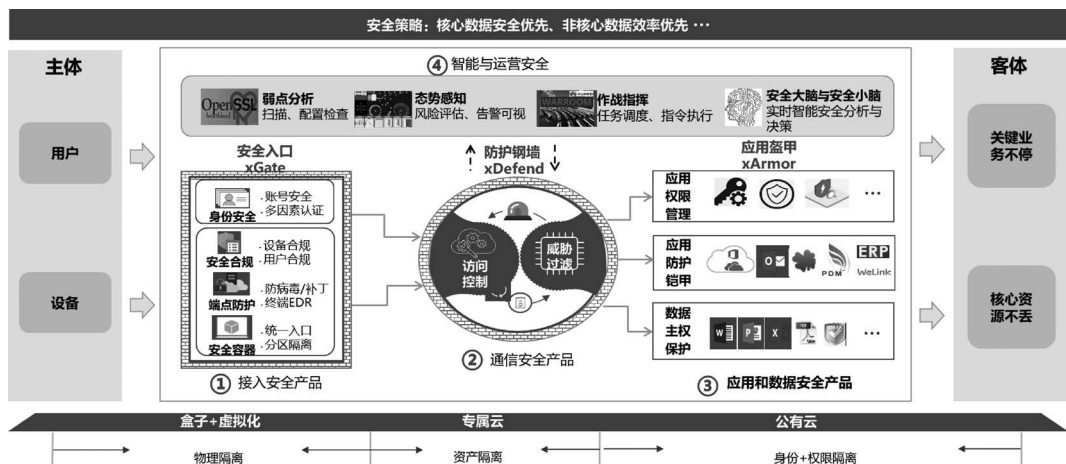


图 5-24 CDOS 西向安全可信平台

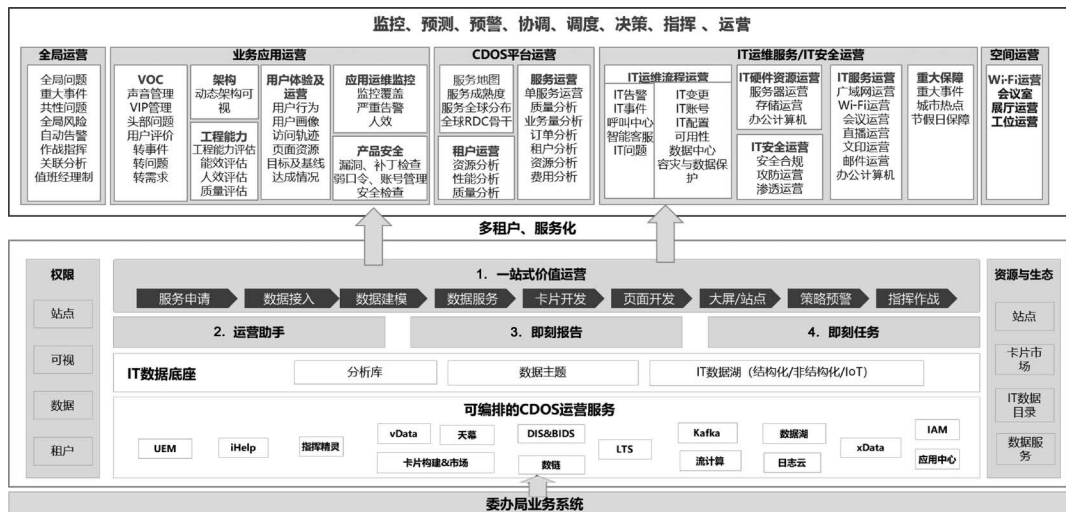


图 5-25 CDOS 东向智能与运营平台

如图 5-26 所示。

建设统一云管平台,将市级、区级及委办局自建云统一管理,实现市区两级有效管理与协同。

(1) 资源协同、共建共享: 实现云资源、灾备资源、运维专家、管理方法、知识经验等能力全市共享。

(2) 一级运营、两级运维,统一监测: 云服务在统一云管平台上完成申请、审批、发放、销毁,分级运维,重要指标实时汇聚至统一运营中心,实现全市政务云安全运行。

(3) 数据协同: 基于市统一大数据平台,实现多维数据融合共享,助力城市精准治理。

(4) 应用协同: 基于市统一 PaaS 平台,实现应用灵活组合创新,市区按需要部署。

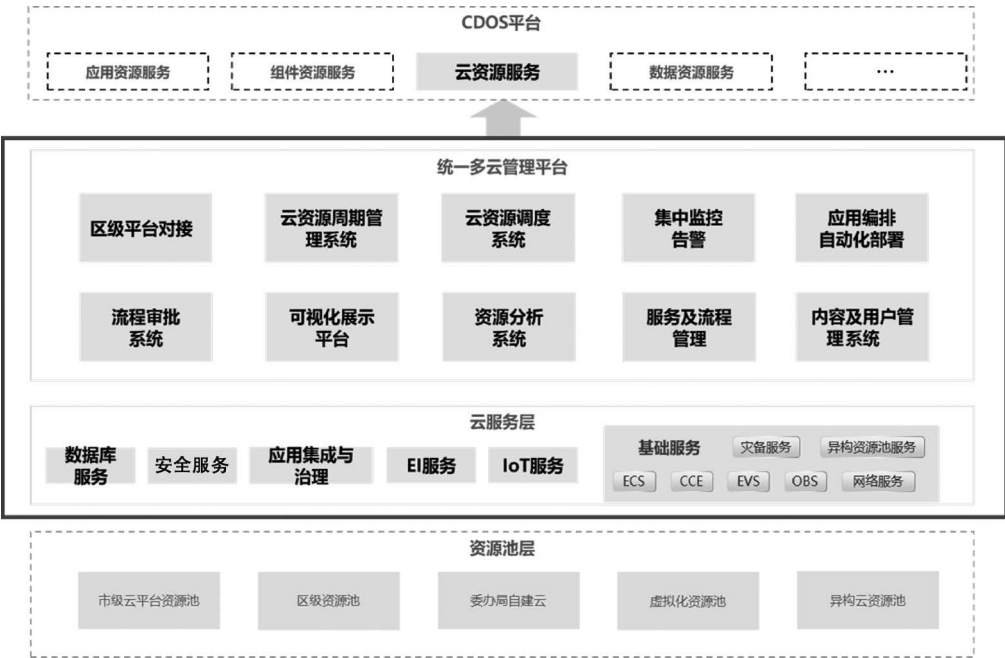


图 5-26 CDOS 云管一本账

6) 网管一本账

传统城市网络分建分管，网络缺乏统筹，可视化和管理能力弱，业务端到端保障难。

(1) 没有端到端的业务可视能力：网络分建分管，纵向业务没有端到端网络质量可视化能力，出现问题缺少定界手段。

(2) 级联协同端到端会议保障：会议系统在市里，区县会议体验差，难以界定区县网络还是市级网络问题。

《国家电子政务外网运维管理系统对接规范》要求完成部省(强制)及省市(建议)运维系统对接，提升管理工作水平和运维效率。

网管一本账通过业务应用牵引，按需配置，构建可信安全的全市一张网管理机制，实现全市各类网络跨域管理、统一运维运营，如图 5-27 所示。

方案价值：

(1) 网络一本账提供场景化网络服务能力，全网可视，网络家底一览无余。

(2) 结合网络服务化，提升业务端到端开通速度，重保业务有效保障业务质量，SLA 可承诺。

(3) 统一网管平台提供跨域模块，同时完成 IP+光管理，质差级联协同 IFIT 实现故障快速定位。

7) 数管一本账

数管一本账通过构建一体化大数据治理体系，赋能城市治理与数据要素流通，如图 5-28 所示。

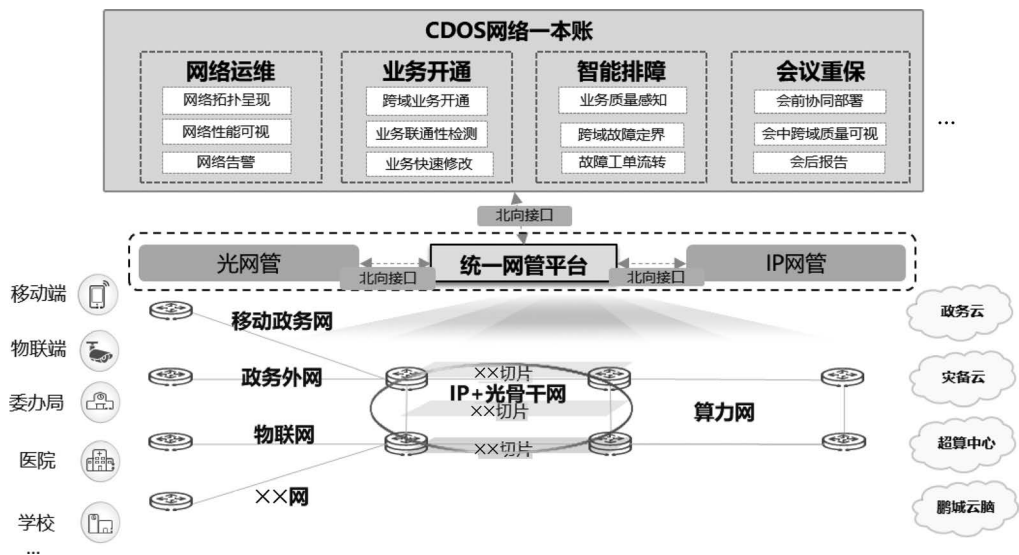


图 5-27 CDOS 网络一本账平台

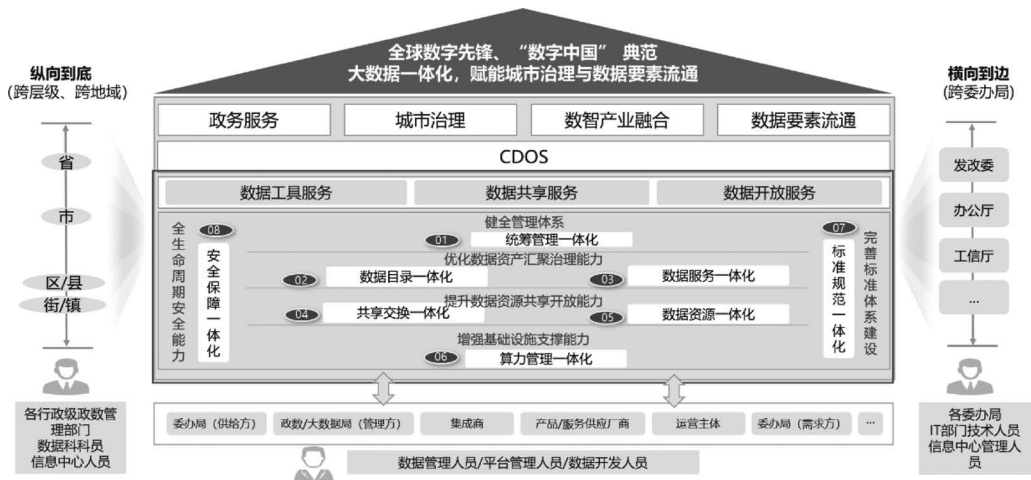


图 5-28 CDOS 数管一本账

4. CDOS 持续运营

1) 管理流程及治理体系

政府数字化转型是多维度、全方位的系统工程，路阻且长，难以一蹴而就，需要构建配套的管理流程及治理体系，如图 5-29 所示，以保障其转型过程得以稳步推进。

数字化转型运行模式(建议)：

- (1) 数字化转型顶层规划和设计，识别愿景、目标、关键举措及路标等。
- (2) 不涉及业务模式变化，仅对原业务流程进行优化(如各委办局自身业务流程简化)。
- (3) 业务模式发生变化，涉及政府跨部门间流程、IT、组织等全方位改革，通过信息化项

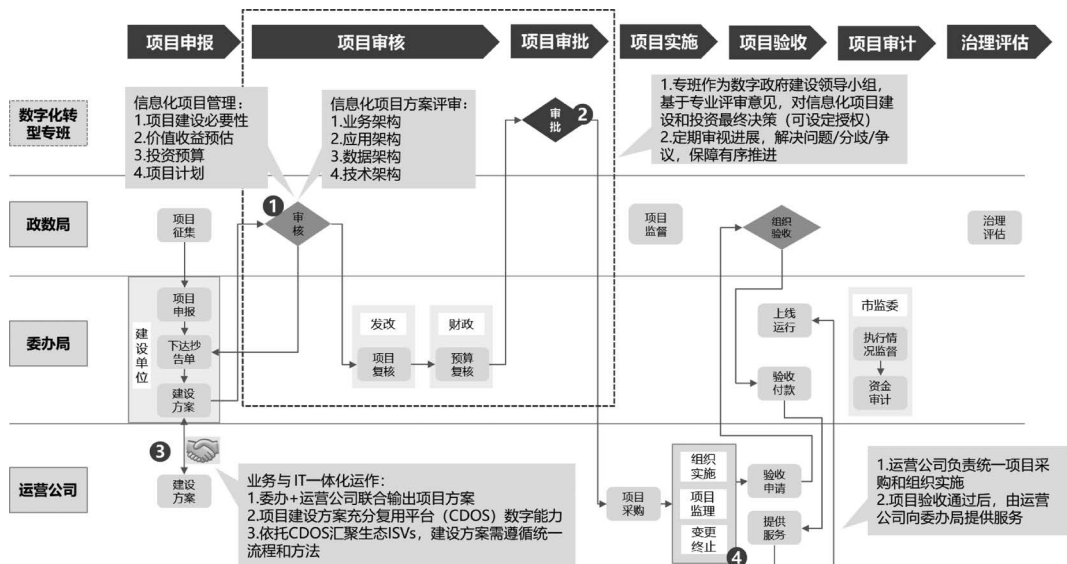


图 5-30 项目流程职责和机制

(2) 信息化项目方案评审：业务架构、应用架构、数据架构、技术架构。

项目审批：

(1) 专班作为数字政府建设领导小组，基于专业评审意见，对信息化项目建设和投资最终决策(可设定授权)。

(2) 定期审视进展，解决问题/分歧/争议，保障有序推进。

项目实施和验收：

(1) 运营公司负责统一项目采购和组织实施。

(2) 项目验收通过后，由运营公司向委办局提供服务。

3) 信息化项目建设和运营协同机制

信息化项目建设和运营阶段，核心是在统一的治理框架下，分层有序运作并有效协同，如图 5-31 所示。

业务流程优化、项目建设方案及业务运营过程产生的业务应用需求，由业务委办/运营公司/ISV 基于统一的方法和 CDOS 平台进行构建。

政数局/运营公司负责云、网、感知、数据、组件服务等平台 IT 的统筹规划以及 CDOS 平台的建设和运营。

(1) 规划建设：规划服务地图及路标，逐步构建

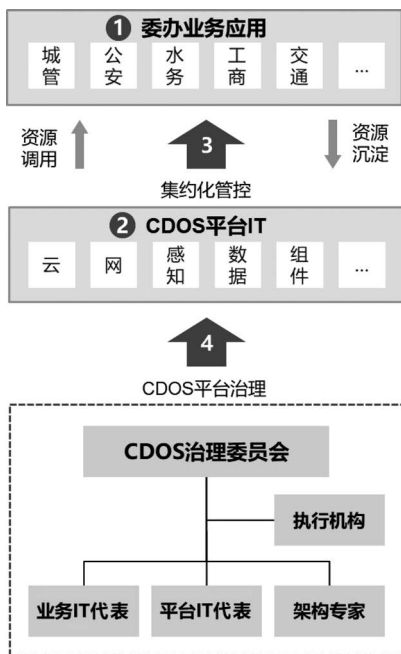


图 5-31 建设和运营协同机制

公共应用服务、平台服务、基础设施服务、安全服务、数据服务等 IT 能力。

(2) 提供服务：把 IT 的能力形成平台标准服务、组合服务、场景服务等，发布服务标准，明确服务所有者，承诺 SLA，保障稳定运行。

(3) 共建生态：引进 ISV 优秀能力，构建开放的平台生态，建立良性竞争机制，支持共建，形成统一的平台服务能力。

政数局/运营公司负责在业务应用建设过程中落实集约化管控要求，从软件架构治理到数据治理，再到业务治理。

CDOS 委员会履行 CDOS 平台治理职责，对 CDOS 平台架构、平台需求、资源规划、资源地图、资源所有者、实施主体、资源冲突、生态引入等重大事项进行管理和决策，确保服务的有序建设及争议的有效解决。

4) 持续运营

充分发挥平台价值，聚合生态优势能力，打通资源供给端和应用场景端的生态共创通道，全面赋能，加速政府数字化转型，如图 5-32 所示。



图 5-32 持续运营平台

基于制度、方法、流程和组织的建设，在业务效能、指标驱动、考核评价的牵引下，通过业务运营、应用/数据运营、数字资源运营、平台运营、生态运营，实现信息化项目建设标准化、规范化，实现运营组织高效有序运作。

业务运营：

(1) 通过业务流程分析、度量指标定义、指标可视与监测、问题分析与改进，实现业务效率与结果的不断提升。

(2) 针对识别到的低效，使用 4A 方法，进行业务及流程方案设计。

应用/数据运营：

(1) 应用规划、设计、开发、上线、运维、运营等全生命周期的管理。

(2) IT 应用方案设计。

(3) 应用/数据服务化设计。

数字资源运营：

(1) 通过对数字资源规划和管理等，使能数字资源全生命周期的可视、可管、可用、可控、可溯，实现数字资源的高效运转。

(2) 持续沉淀数字资源。

平台运营:

(1) 运营流程适配,包含对数字资源服务的注册、上架、订阅、开通、计量、计费、运营、运维等进行全生命周期管理。

(2) 运营指标设计、运行和监测。

生态运营:为入驻平台的厂商提供准入资质评估、培训认证、技术支持、服务结算、退出等运营服务,建立良性的生态圈,激发市场活力,激励厂商贡献。

5.2.7 智能应用

基于在城市智能化转型的实践总结,目前城市智能体主要在面向政务业务、城市业务和产业经济三大方面构建了六大业务应用。

1. 智能化使能政务业务

1) 政务服务“一网通办”

构建全流程一体化在线政务服务平台,推动群众和企业办事线上一个总门户、一次登录、全网通办,线下只进一扇门、最多跑一次。

2) 政务办公“一网协同”

面向政府一体化协同的办文、办事、办会场景,依托数字底座,使能多端联动,打造跨层级、跨地域、跨系统、跨部门、跨业务的“协同办公平台”。

2. 智能化使能城市业务

1) 城市感知“一网统感”

围绕城市融合感知系统建设行动计划的总体规划,逐步建成统筹规范、泛在有序的新型智慧城市运行感知体系。

2) 城市运行“一屏统览”

通过对城市全域数据进行实时汇聚、监测、治理和分析,全面感知城市生命体征,辅助宏观决策,预测预警重大事件,配置优化公共资源,处置指挥重大事件,保障城市有序运行。

3) 城市治理“一网统管”

围绕“高效处置一件事”构建高效管理、协调、拉通的治理体系。

3. 智能化使能产业经济

产业赋能“一网通服”:助力政府精准服务中小企业实现数字化转型升级,促进企业场景创新及专业化水平提升,不断推进产业集群高质量发展。