

科技部原副部长吴忠泽:

“十四五”期间数据中心向算力集群演进

本报记者 刘晶

近日,科技部原副部长吴忠泽在“2021年中国IDC行业Discovery大会”上,围绕“十四五”规划的实施,提出要推动数据中心产业高质量发展。吴忠泽表示,数字经济新基建给IDC行业数字化发展带来了新机遇。“十四五”规划要求“加快构建全国一体化数据中心体系,强化算力统筹智能调度,建设若干国家枢纽节点和大数据中心的集群,建设E级和10E级的超级计算中心”等,凸显数据中心创新发展为全社会数字化转型服务的基础作用。

摆脱孤岛式计算

自去年以来,中央和地方政府相关的政策密集出台,市场已经掀起了一股新基建的热潮。伴随着“十四五”规划的实施,数字经济新基建快速推进,数据中心行业迎来了更加广阔的发展前景。

按照“十四五”规划的部署,重点推进云计算、大数据、物联网、工业互联网、区块链、人工智能、虚拟现实和增强现实等数字经济重点产业发展。

随着数字化、网络化、智能化等应用的加速落地,数据中心将成为这些数字经济重点产业技术及其应用的重要载体。同时也是工业、金融、电力、交通、医疗、教育等各个行业数字化转型和智能化升级的关键基础设施。在此背景下,工业互联网、智慧交通、智慧能源、智慧金融、智慧医疗、智慧城市等新型场景不断涌现。用户对于算力提出更多要求,推动了5G和边缘计算时代的来临。

随着云计算兴起,用户需求日趋多样化,数据中心正在从传统的孤岛式的计算和存储基础设施,逐渐向以智能为代表、以高速网络作为基础、核心与边缘形态共存的算力集群方向演变。

正由于数据中心产业是新基建的“底座”和“算力集群”,发展新基建中的任何一个领域都离不开它,因此,目前数据中心产业已经成为资本市场关注和参与的重点之一,投资者从多个角度抢滩布局。

据中国IDC圈研究统计,我国数据中心市场规模2020年已经超过了2000亿元,传统IDC业务的收入端也接近1000亿元规模,连续三

年保持了25%以上的增长,IDC产业进入了快速发展的轨道。

数据中心呈四大发展趋势

吴忠泽表示,数字科技融合创新正在完善IDC产业数字化生态体系。

近年来,全球新一轮科技革命和产业变革正在加速,特别是数字科技对于经济社会各个领域的渗透性、扩散性越来越强,以智能、绿色、泛在、融合作为特征的群体性技术革命,尤其是颠覆性创新持续涌现,正在引领第四次工业革命。

例如,以大数据为基础的人工智能,与5G、物联网等深度融合,以推进数字产业化和产业数字化转型为核心,强调数字技术从技术形态到产业形态的发展,给数字经济带来了巨大的加速度。

吴忠泽说,按照“十四五”规划部署,发展大数据这个数字经济重点产业,需要突破的方向是什么?第一,推动大数据的采集、清洗、存储、挖掘、分析、可视化算法等技术创新。第二,培育涵盖数据采集、标注、分析、存储、传输、管理、应用等全生命周期的产业体系。第三,完善大数据的标准体系。为此,我们要按照规划要求加快建立数据资源产权、交易流通、跨境传输和安全保护等一些基础制度标准规范,完善数据中心产业数字化的生态体系。

吴忠泽表示,在此背景下,数据中心呈现出几个新的发展趋势。

一是,注重打造良好的产业生态,构建跨界合作机制;加强网络供给,推动数网协同发展;加强和新基建其他领域融合创新,培育数据中心新业态和发展的新动能。

例如,按照“十四五”规划关于

实施“上云用数赋智”行动,推动数据赋能全产业链协同转型的部署。规划要求“在重点行业和区域建设若干个国际水准的工业互联网平台和数字化转型促进中心,深化研发设计、生产制造、经营管理、市场服务等环节的数字化应用,培育发展个性化定制、柔性制造等新模式,加快产业园区的数字化改造”。

为此,数据中心要和工业互联网、5G、物联网等“数网协同”发展,赋能传统产业,要通过全要素、全产业链、全价值链的全面连接,加速线上场景的变迁、全域数据的融合、智治模式的推进,推动中国经济社会“全面在线”时代的到来。

二是发展分布式、边缘式的数据中心,以提升对实时性要求较高应用的支撑,尤其是应用区块链超大规模分布式数据存储技术,扩展大量国产服务器并行计算,支撑数据库的数据存储和处理。

三是加速与人工智能的深度融合,抓住“十四五”规划部署“建设重点行业人工智能数据集”的机遇,重构数据中心的算力和能力,强调数据中心未来生产方式和服务方式的换代。基于云计算、5G、大数据的助力,数字经济正在进入到以人工智能为核心驱动力的智能经济新阶段。可以预见,未来十年,智能经济将成为中国经济的新标签,人工智能必将进一步带动包括数据中心行业在内的各行业智能化转型。

目前要重点应用人工智能技术来提升数据中心本身的智能化管理水平,强化数据的安全和隐私保护。从网络信息安全、数据资源安全、生产运营安全等方面筑牢数据中心安全发展防线,增强数据预警和溯源能力,为“新基建”各领域研发和

广泛应用提供海量数据支撑。

四是应用智慧能源技术,提升数据中心能效,注重节能环保,推动绿色数据中心的发展。按照“十四五”规划关于“开展能效在线分析、源网荷储互动、多能协同互补、用能需求智能调控”的要求,同步建设分布式高效能源互联网,形成支撑多能源协调互补、及时有效接入的新型能源网络,推广智能储能设施、智能用电设施,实现数据中心能源供需信息的实时匹配和智能化响应。

构建数据中心开放利用场景

吴忠泽认为,实施“十四五”规划,要构建IDC企业数字化应用场景优势。技术为本、场景为王。“十四五”规划要求“充分发挥海量数据和丰富应用场景的优势,促进数字技术与实体经济的深度融合,赋能传统产业转型升级,催生新产业新业态新模式,壮大经济发展新引擎”。并明确提出了数字化应用场景,如智能交通、智能能源、智能制造、智慧农业水利、智慧医疗、智慧文旅、智慧社区、智能家居、智慧政务。

例如,规划在“推进产业数字化转型”这一章节,要求“深入推进服务业数字化转型,培育众创设计、智慧物流、新零售等新增长点。加快发展智慧农业,推进农业生产经营和管理服务数字化改造”等场景。产业数字化转型衍生出来的新产业、新业态、新模式将充满着机会,需要大家主动的拥抱。

吴忠泽表示,我国是制造大国和网络大国,完全有条件、有基础利用新技术快速大规模应用和迭代升级的独特优势,坚持目标导向、问题导向,注重需求牵引,构建数据中心开放利用的场景,谋划一批“新基建”项目,在重点领域打造一批数字化应用场景的试点示范工程。

例如,“十四五”规划第十六章“加快数字社会建设步伐”中,要求“完善城市信息模型平台和运行管理服务平台,构建城市数据资源体系,推进城市数据大脑建设”等。

面对如此丰富的数字化应用场景,数据中心企业发展要建立一种升维意识,实现平台升维、产品升维、营销升维、价值升维。要在数据行业中通过联合创新、外部资源对接等方式,帮助数据中心企业加速技术和模式的创新,培育一批成长性领军企业,带动中小企业的创新活动。

通过“平台+生态”模式,提升数据中心新业态的培育能力,依托新基建不断扩大平台型数据中心的对象,创新服务模式,融入共享机制,提升平台的能级,优化创新生态。例如,规划在“加快推动数字产业化”章节,要求“鼓励企业开放搜索、电商、社交等数据,发展第三方大数据服务产业。促进共享经济、平台经济的健康发展”等。

曾经,手术室内被医治的患者与在病房外焦急等待的家属,两者虽只隔数十米,却仿佛咫尺天涯。如今,5G+沉浸式VR技术,让“零距离”陪伴成为可能。

新冠肺炎疫情爆发初期,浙江移动联合浙江大学第二附属医院,上线了“5G+VR病毒重症监护室远程诊疗监测系统”,在5G环境中,家属可通过“360° 8K全景摄像头”了解医院中患者的一举一动,并与患者实时双向互动沟通,实现家属“身临”患者床边的沉浸式探视;此外,该系统还能将医生手术治疗患者的场景分发给远在异地的学员和专家,进行远程会诊和示教。

浙江移动杭州分公司信息化运营支撑主管王亦凡在接受《中国电子报》专访时表示,未来可视诊疗行业将达到千亿元规模,前景广阔,中国移动的5G远程VR探视与示教系统未来几年将覆盖到全国各级医疗机构,以及医疗机构下属教育单位,切实为智慧医疗注入新活力。

5G+VR 远程探视 加速落地

本报记者 齐旭

成本和商业模式有待市场

进一步“消化”

确定性网络保障系统

顺畅运行

2019年,浙江移动杭州分公司利用5G大带宽、低延时和VR身临其境的技术特点,与医院共同搭建了一套可视化、数字化的“5G+VR远程查房会诊探视管理系统”,提供远程探视、病房巡查、在线学习和远程诊疗等应用服务。

新冠肺炎疫情爆发初期,浙江移动联合浙江大学第二附属医院,上线了“5G+VR病毒重症监护室远程诊疗监测系统”,在5G网络环境中,家属可通过“360° 8K全景摄像头”了解医院中患者的一举一动,并与患者实时双向互动沟通,实现家属“身临”患者床边的沉浸式探视;此外还能将医生手术治疗患者的场景,不受时间空间限制地分发给学员和专家,进行远程会诊和示教。

该系统通常在前端手术室或病房中部署3台以上高清编码终端,将手术室或病房中VR全景摄像机拍摄的三路视频或医疗信号采集汇编,传输和分发到后端示教室、远程会诊专家的VR头盔、投影机和平板电视中,以及患者的终端上。用户能够灵活切换现场手术视频画面、控制台云和摄像头,实现双向语音对讲互动。

王亦凡告诉记者,目前市面上基于5G+VR的医疗探视系统众多,中国移动创新性地将5G通信模块集成到远端的VR可穿戴设备中,做到了“5G inside”,具体来说,VR头盔等设备不需要额外叠加通信的网关,就把5G信号解耦成它所需要的这种通信信号,原则上只要有5G网络,设备就可以直接使用,使得该套设备更加轻便,也更加适用于需要快速部署网络应用场景。

4K/8K VR全景摄像机大码流视频的流畅传输、分发给网络带宽、视频编解码和终端协同带来更高挑战。在网络环境方面,王亦凡告诉记者,中国移动根据不同应用场景,给予远程VR探视与示教系统5G公网和5G专网切片两种网络环境。针对一般的探视场景、示教场景,这套系统会在公网流量环境下运营,特别是一些广域场景中,比如说医疗机构以及下属医联体所处的网络环境,不一定能够支持专网,中国移动给予医院这类特殊地点网络环境优化,将网络故障率降到最低。

考虑到医患关键医疗数据的隐私性,以及手术中的远程会诊这类需要确定性、高可靠网络的应用场景,中国移动部署了定制化的5G切片专网+边缘计算,利用IP资源预留,把数据的上行带宽设置在100兆左右,将时延控制在20毫秒内,确保视频数据直接在本地分解、下沉,提高数据上传和处理效率;同时将关键数据与公网环境隔离开,也保证了数据的隐私性和安全性。

“针对VR视频数据大码流的问题,我们会更多地通过前端对视频数据进行压缩和编解码工作,尽可能减少视频上传的网络带宽要求。”王亦凡说,“终端间的视频同步则需要不同厂商之间做好合作协同,共同解决视频同步卡顿问题。”

《2020智慧医疗发展研究报告》显示,2020年我国智慧医疗行业规模已突破千亿元大关,预计2021年规模将达1259亿元,行业将进入智能化高效化规模化的高速增长期。

与其他行业不同,智慧医疗系统不但需要提供技术的厂商“深度”垂直,还需要厂商带来人性化关怀的“温度”。VR+5G技术带来诊疗过程的相对透明可视,解决医患信息的不对称,可有效缓解医患矛盾,以及患者与家属的矛盾,增进信任。

一方面,VR用于病理讲解,为患者和家属带来比文字更易于理解的影像视频。患者或家属戴上VR头显,并打开定制的VR体验,医生或讲解员,用浅显易懂的话语为患者家属解析患者身体现状,告知医生的治疗方案是如何实施的;另一方面,利用VR手术直播,患者家属能够通过VR手术直播直观地看到医疗团队对患者的全力诊治,这样一来能够有效增强医患协同,带来信任。

据王亦凡介绍,5G远程VR探视与示教系统上线以来,仅在浙江大学妇幼保健医院,就有3000人次通过系统实现VR探视,获取收入15万元。通过门户、微信门户等点击使用超过3万次以上,带来流量收入约为20万元。

“未来可视诊疗市场将达到千亿元规模,前景广阔,中国移动将持续投入大量的资源和精力部署用于医疗探视示教的VR项目。但5G+VR应用目前仍处于发展初期,技术和商业模式较为青涩。”王亦凡说,“一是由于5G通信模组的整体布局与4G有较大差别,受制于5G核心通信模组芯片自主研发能力有待成熟,目前我们整个通信模块和终端采购成本仍然偏高,整机设备难以实现有规模的量产;二是应用配套的流量套餐对于用户来说价格仍然偏高,如果我们仍沿用传统的按G流量计费方式收取资费的话,若用户长时间使用就可能难以承受。”

随着未来更多的终端厂商和芯片厂商入局,将打通全产业链产销通路,经过市场化的消化,良性循环将逐渐形成,研发成本有望进一步削减。模组价格降下来了,应用大规模推广也就不成问题了。

据王亦凡介绍,中国移动作为电信运营商,正在寻找一种更适用于VR远程社交系统的通信服务和套餐资费方案,或将通过设备租赁和设备包年的方式,降低应用成本,更好地服务客户,也有助于进一步释放市场潜能。

对于未来行业中面临的机遇和挑战,王亦凡坦言,智慧医疗行业当前所面临的竞争异常激烈,不同类型厂商入局,不同医院、不同科室之间的医疗资料数据模式差异也非常大。对于应用集成方案提供商来说,如何打通专业和行业之间的数据壁垒,解决通信业和医疗业的数据互联,是行业正在面临的挑战。但挑战中蕴藏机遇,需要解决方案提供商、终端设备供应商等产业链厂商共同努力,共同挖掘更高质量的应用解决方案。

达摩院发布中文社区最大规模预训练语言模型

本报讯 记者宋婧报道:近日,阿里巴巴达摩院发布中文社区最大规模预训练语言模型PLUG。该模型参数规模达270亿,集语言理解与生成能力于一身。

该模型背后的操作原理是自然语言处理。这是人工智能的核心技术。它无缝弥合了复杂的人类语言和编码机器之间的通信鸿沟,让机器“听得懂”人话。而预训练语言模型正是目前自然语言处理的一种主流技术。它可以基于大规模文本预训练,得出通用的语言知识,然后经过微调,有针对性地学习后续的知识传递给不同的下游任务。

从产业发展的角度来看,基于人工智能的语言模型提供了一种基于海量非监督数据自学习的范式,减少人工干预成本,加快AI技术的商业化落地。近年来,AI领域兴起了预训练语言模型浪潮。2020

年5月,OpenAI发布的1750亿参数超大规模预训练模型GPT-3“一炮而红”。微软巨额投资跟进,获得GPT-3语言模型的独家许可。2021年1月,谷歌推出1.6万亿参数的超大规模语言模型Switch Transformer,也受到了高度关注。

“通常情况下,越多训练数据,越大参数规模,一般能够获得更好的性能。当然,到了一定程度也有一定的上限。”达摩院深度语言模型团队负责人黄松芳在接受记者采访时表示。大规模AI语言模型竞争的背后,实际上是数据量、算力和模型结构设计之间的较量。

据黄松芳介绍,PLUG的亮点主要体现在两个方面:首先,不同于GPT-3单向生成模型,PLUG是一个双向的自然语言理解和生成统一模型,集成了达摩院自研的语言理解StructBERT和语言生成PALM双模型,通过构建输入文本

双向理解能力,显著提升了输出文本的相关性。其次,得益于阿里云EFLOPS高性能AI计算集群和高效的并行训练优化算法,PLUG基于1TB的文本数据训练,参数规模达到270亿,在语言理解任务上,PLUG以80.614分刷新了CLUE分类榜单纪录;在语言生成任务上,PLUG多项应用数据较业内最优水平提升了8%以上。

然而,黄松芳也坦言:“目前PLUG是全球中文领域最大的纯文本预训练语言模型,但相比国外以英文为核心的GPT-3模型1750亿的参数规模,还只有其1/6不到。”据介绍,达摩院计划将PLUG参数规模扩大至2000亿级,并进一步提升文本生成质量。

“后续,我们在尽力继续扩大模型参数规模的同时,也会关注超大模型的落地应用实践,努力探索低碳、高效、业务可用的预训练语

言模型。另外,对于大规模训练语言模型的发展趋势,我们将从数据驱动逐步发展到知识驱动,探索数据和知识深度融合的预训练语言模型。”黄松芳说。

未来,PLUG有望广泛运用于日常生产及生活的场景中。例如,在医疗领域中,PLUG可用于电子病历的质量控制、临床医学的辅助诊疗等;在电力领域,可用于各种电力设备文档的解析,打造为电力企业员工提供设备故障诊断的AI助手,还可以帮助理解电力故障报案需求,提供高效的电力调度支持。

“超大规模预训练模型将作为一种AI基础设施,推动相关理论研究和应用更上一层,从而改变整个产业格局。”智源学术副院长、清华大学教授唐杰表示,“继基于数据的互联网时代、基于算力的云计算时代之后,接下来可能将进入基于模型的AI时代。”