

中国工程院院士郑纬民：算力联网统一调度是大势所趋



超算基础软件是提升HPC算力应用效率的关键

HPC 算力又称超级计算机系统算力(简称“超算算力”),用于应对最富有挑战性的计算问题,作为国际科技竞争战略的制高点,是名副其实的“国之重器”。当下,我国算力指数已经进入了E级,这意味着中国已经有能力制造出每秒能做100 亿亿次数值运算的超级计算机。郑纬民表示,近十年,我国在顶尖超算系统研制和超算系统部署数量上都处于国际领先行列;从质量上看,我国在过去十年内曾经制造出3 台世界范围内最快的计算机;从数量上看,在全世界 500 台最快的计算机统计中,我国占其中的162 台,数量超过美国,居世界领先地位。“目前,我国有13 个国家超算中心,在超算算力的提供方面发挥了重要作用。”郑纬民介绍道。

然而,在响应速度上限不断拓展的同时,我们仍然面临着实际应

用层面的挑战。郑纬民表示,将领先算力高效转化为解决尖端科学与工程难题的能力是世界范围内的难题,且当前异构架构路线的选择更加大了鸿沟。“在实机运行软件时,我们计算机的实际运行速度只能达到理想反射速度的10%~20%。因此,提高这个比例就是我们要努力的方向。”

谈到目前的解决方向,郑纬民强调,超算基础软件是提升转化能力的关键之一,是实现并行应用开发、优化、部署、运行高效的基础和核心。当下,国产超算系统已部署部分基础软件,而更多的完善方向,则亟待研究人员去探索。

人工智能应用场景遍布各大领域

2020 年4 月,国家发改委明确将人工智能纳入新基建范围,人工智能成为新兴热门产业之一,受到社会各界的重视。根据不同的使用场景,郑纬民将人工智能归纳成三

大类应用:一是以卷积网络为核心应用方向的图像检测类;二是以AlphaGo 为代表的,以强化学习为核心的博弈决策类;三是以ChatG-PT 为代表的,以Transformer 为核心的自然语言处理类。据了解,Transformer 模型是一种基于自注意力机制的神经网络模型,它的核心思想是利用自注意力机制来捕捉输入序列中的上下文信息,从而实现更加准确的语言建模和翻译。郑纬民表示,在这几个应用类别中,只有自然语言类人工智能需要大机器人的硬件支撑。

参数是大模型发展的一项重要指标,参数越多,就意味着计算机的算力越大。当前,人工智能正快速向更大的模型发展,参数呈指数级上升;人们耳熟能详的GPT-3 的参数高达1700 亿,是GPT-2 的113 倍。在我国,人工智能产业快速发展,包括西安、武汉、成都、沈阳等在内的20 多个城市陆续启动了人工智能计算中心建设,应用场景遍布公安、金融、政府、运营商等各大领域。郑纬民表示,预计到2025 年,中

国人工智能产业规模将超过4500 亿元,带动产业规模超1.6 万亿元。

互联互通是数据中心算力发展的必由之路

在演讲中,郑纬民重点强调了算力互联的概念。他表示,算力的互联互通是实现“东数西算”的一项基本条件,更是数据中心算力发展的必由之路。郑纬民表示,在当前的应用场景中,带宽不足导致的低延迟是算力互联无从避免的局限性。他提出,要对现状做出改善,一方面要提高算力输送效率,通过并网实现高带宽、低延迟的算力互联;另一方面,可以屏蔽异构基础设施,通过统一的编程框架和编译的资源管理与调度软件实现算力的互通。

最后,郑纬民表示,目前还没有类似的研究机构来解决算力的互联互通问题。他建议,成立东数西算研究院,团结领域内各大企业,实现算力的统一调度,稳步推进“东数西算”的发展与预后工作。

中国科学院院士梅宏：数据作为资源和要素应独立于应用

本报记者 赵晨 实习生 蒋安昊

7 月5 日,中国科学院院士梅宏出席2023 全球数字经济大会主论坛并发表主旨演讲。梅宏认为,数字经济时代正在开启,这是在农业经济、工业经济之后的一种新经济形态。数据技术体系、数字治理体系以及数据要素市场构成了数字经济发展的三大基石,数据成为生产要素将是这个时代的重要特征之一。

梅宏从五个方面提出了大数据技术发展面临的一系列挑战,从管理、处理、分析、治理四个维度分析了大数据技术未来的发展趋势。

第一,如何构建以数据为中心的计算体系?梅宏表示,线性增长的数据处理能力与指数膨胀的数据处理需求存在剪刀差,迫切需要寻找新的性能提升驱动力,新硬件技术、人工智能技术为大数据计算体系带来了新的机遇和挑战。同时,超大规模数据的能耗占比大、增速快,现有的计算体系偏重性能,能效不高,可持续发展能力有限,带来了能效方面的挑战。

第二,如何管理超大规模的数据要素?在梅宏看来,越来越多的数据资源正以数据要素的形态独立存在于不同空间、组织和信任域,并参与数字经济活动的全过程,需要对大规模、跨域的数据进行高效的组织和管理。数据价值最大化要求数据组织管理的范围从传统的单域模式到跨域模式,这就带来了跨域管理的挑战。数据管理是大数据产业的基础,可用性也是一大挑战。大数据规模呈指数级高速增长,大数据管理系统往往规模极大,因此各类错误发生概率不断提升。另外,还可能存在恶意节点引发虚假信息或数据篡改等严重问题。

第三,如何满足大数据高效处理的需求?梅宏认为,新型数据处理需求对现有的大数据处理系统与架构提出了多种技术挑战。具体来看,实时数据的特征是动态倾斜、稀疏关联、超大容量,需要应对资源效率低、时空开销大、扩展困难等技术挑战;图数据的特征是不规则访问,计算访存比小、依赖关系复杂,需要应对并行流水效率低、访存局部性低、内外存通道利用率低等问题。

第四,如何实现多源异构大

数据的多模态融合分析与可解释性分析?梅宏提出了两方面挑战:一方面,异构多模态数据的融合分析从小规模、单模态转变到大规模、多模态;另一方面,模型在可解释性、稳定性、公平性、可回溯性这四个方面也需要更高的性能和要求。

第五,如何形成系统化大数据治理框架与关键技术?梅宏指出,数据要素化需要有高效的大数据治理技术体系支撑,涉及数据汇聚融合、质量保证、安全与隐私保护、开放流通、标准化与生态系统建设等多个方面。

关于大数据技术的未来发展趋势,梅宏认为,从总体来看,数据与应用将进一步分离,以促进数据要素化。尽管历史上数据库技术的出现带来了数据与应用的第一次分离,但这种分离仍不够,数据要作为资源和要素进一步独立于应用,真正成为一个独立存在的实体。高效大数据技术栈的构成是可持续发展的关键,特别是在“双碳”背景之下,要形成低碳高效、可持续发展的路线。要在资源共享、节能高效的云数据中心基础上,布局全国一体化高效大数据中心,形成低碳发展新格局。数据相关技术需要形成大数据标准规范,积极推进开源社区的建设。

在数据管理上,从单域到跨域的数据管理需要促进数据要素的共享与协同。数据管理正在从面向和限定于单域的孤立服务,发展到跨越空间域、信任域和管辖域的数据共享与协同服务的新阶段,要打破“数据孤岛”,促进数据价值的最大化。

在数据处理上,近数处理成为突破性能瓶颈的重要途径,包括存储上移、算力下移及分布式计算模式转变。系统设计要从扩展性优先走向性能优先,需要拥有性能优先的系统设计、领域专用硬件和新兴处理方法。

在数据分析上,从单域单模态分析到多域多模态融合,要实现广谱关联计算;从聚焦关联到探究因果,要实现分析结果可解释。

在数据治理上,数联网将成为数字化时代的新型信息基础设施,需要形成一套完整的数联网基础软件理论、系统软件架构、关键技术体系,以数据驱动为手段,研究在数联网环境下保障服务质量与保护质量的原理、机制与方法。

坚持纾困与培优两手抓 推动中小企业平稳健康发展

