秀湖会议集萃

算力网：新基建背景下的分布式系统
——第八期CCF秀湖会议报告

关键词：算力网　分布式系统

整理：郭得科　彭晓晖　王一帆　廖小飞

背景与意义

在数字经济时代，数据成为新的核心生产资料，算力成为继热力、电力之后新的关键生产力。2018年12月19日，中央经济工作会议将5G、人工智能、工业互联网、物联网等定义为“新型基础设施”，并在2019年的政府工作报告中明确要求“加强新一代信息基础设施建设”。近年来，我国的数字化转型和智能化推进产生了爆发式增长的算力需求，如何通过算力网聚集全国各地的智算、超算和数据中心算力，实现跨地域的算力统一供给、管理和使用，推动中国算力经济进入引领阶段，仍然是有待解决的问题。

针对上述问题，2023年12月1~3日，第八期CCF秀湖会议邀请了来自学术界和产业界的20余位专家围绕“算力网：新基建背景下的分布式系统”这一主题开展观点报告和集中讨论。按照算力网的组成部分，会议组织了四个专题讨论。“多样性算力与并网”专题从CPU、GPU、FPGA等多种算力入手讨论了算力的多样性和异构性，分享了网格计算、云计算等算力设施的发展历史，以及超算、智算和数据中心的发展现状。专家们指出，算力并网是算力网建设的第一步，需要从计算、存储、传输等多个维度实现算力资源的并网，形成统一的算力资源池。“分布式存储设施”专题介绍了数据中心存储、分布式内存等多个方向的发展现状和趋势，专家们认为分布式存储要围绕国家数据基础设施展开，提供具有一致性、容错保障、统一访问接口的分布式文件系统是算力网建设的关键。“新型计算机网络”专题分别从互联网体系结构、算网融合架构、在网计算、多维标识网络以及新型智算中心网络等方面介绍了网络的发展趋势以及算力网对网络的新需求。“分布式系统软件”专题分享了算力网的三个基础学术问题、统一的编程模型、面向新场景的定制化抽象和算力网的应用需求，专家们认为分布式系统软件是算力网的操作系统，核心是提供新的抽象机制来满足新旧应用在算力网上有序执行的需求。

出席本次秀湖会议的专家主要来自计算机领域的算力、存储、网络和系统软件四个方向，同时有大量的产业界代表参会。会议讨论了算力网相关技术的发展现状和趋势，形成了关于算力网的初步共识，并发出了相应倡议。

多样性算力与并网

算力网将异构、异地、异属的算力资源通过物理网络连接在一起，通过抽象和池化形成跨全国甚至全球的计算资源池，并根据计算任务的需求统一管理和分配算力，从而为各行各业的应用提供更加强大和丰富的算力服务。

实践分享

算力网的本质是算力的基础设施化、服务化，

但算力的多样性和跨地域特性给并网交易和使用带来诸多挑战。

首先，算力的多样性使算力资源的共享使用变得十分困难。随着大规模科学计算的发展和人工智能（AI）大模型的广泛应用，单一算力集群已经难以满足日益增长的算力需求。整合多样化算力，构建新型算力网是解决算力瓶颈的有效途径。由于体系结构的不同，用户的程序往往需要重新编译以适应不同类型的算力。从服务类型角度来看，算力网包括以CPU为代表的基础算力、以众核处理器为代表的超算算力，以及以GPU、FPGA为代表的智能算力三类。基础算力通常用于云计算、边缘计算等场景，对计算的精度无特定要求；超算算力则主要服务于药物分子设计、基因组分析、天文气象数据分析等科学计算领域，这类计算属于双精度浮点运算，对计算精度要求高；智能算力主要服务于近年兴起的人工智能模型的训练和推理。

其次，算力的多样性给算力的计量和交易带来挑战。电网使用“度”作为电能的基本计量单位，发电厂输送给电网的电能和用户从电网获取的电能也均以“度”为计量单位进行交易结算。而算力的计量方式比较多，早期人们通常使用CPU的频率衡量其处理能力，超算使用每秒完成浮点运算次数（FLOPS）衡量一个集群的浮点运算能力，近些年出现的GPU则使用每秒完成操作次数（OPS）评估其性能。这些算力的计量方法与应用密切相关，算力网急需类似“度”这样的基本计量单位，提高算力交易的便捷性。信息高铁算力网提出了基本操作数（Basic Operations，BOPs）[1]作为算力的基本度量单位，这是一次重要的探索性尝试。

最后，算力的跨地域特性给并网管理带来困难。算力并网是整合汇聚多样化算力的主要手段，需要从计算、存储、传输等多个维度实现算网资源的统一接口、分配和调度。云计算数据中心通常使用虚拟化技术，将大规模集群算力资源池化为虚拟机或容器。算力网的算力通常分布在全国乃至全球各地，这些算力状态的存储、查询和统计远比在数据中心困难。当前，多样算力并网研究还处于起步阶段，国际上典型的代表是天空计算（Sky Computing），国内的电信运营商和云计算厂商均根据各自的业务优势采用不同的并网技术，但算力资源的寻址和命名、时间强一致性等分布式系统基本学术问题尚未引起各界的重视。

网格计算是计算技术历史上将算力并网形成基础设施的第一次尝试，虽然其已经逐渐成为历史，但仍然留下了DataGrid等至今仍在运行的网格系统，网格资源并网的经验和技术也将为算力并网的实施提供有价值的参考。

观点争鸣

华中科技大学教授金海强调算力网的构建需要充分吸取网格计算和云计算的经验，特别是网格计算失败的经验，包括缺乏有效的商业服务模式、缺乏市场化激励机制、低估了多网格中心协同管理难度、跨域分布式计算性能理想化和用户使用复杂度高等。此外，算力网应该继承网格计算和云计算优秀的技术沉淀，涵盖海量数据传输协议、虚拟化和服务封装思想以及服务屏蔽分布式系统异构性理念。他认为从需求侧和供给侧两方面考虑算力网能够帮助用户解决的真实痛点问题和需要克服的瓶颈问题，才能避免算力网重蹈网格计算的覆辙。

湖南大学教授唐卓指出，大规模科学工程计算所需的硬件和软件资源需要多个算力中心协同提供，单一算力中心的计算处理能力和应用资源有限，并且成本越来越高，在数据、算力、算法、软件多样性等方面都越来越难以满足数字政府、智慧城市、科学计算、智能制造、工业设计等复杂场景应用的需要。他认为，超算互联需要构建超算中心间资源访问的全局视图，设计存算一体化的体系架构，打造超算互联网资源共享与解算服务。他以湖南省区域算力网为例，认为应搭建“云网边端”协同的泛在化算力网络，建立云资源接入和一体化算力协同机制，实现资源弹性灵活调用。

中国科学院计算技术研究所研究员胡杏认为大模型时代的智能计算设计需要符合大模型的发展趋势。她指出当前大模型具有3个发展趋势：（1）大模型的

计算规模呈现大、多、长的特点；（2）大模型的服务规模快速扩展；（3）大模型的能力边界快速扩展。针对上述发展趋势，她提出了智能计算系统的3个展望：（1）发展更强单节点算力、大容量内存、高速互联网络，支持更大模型的训练和推理；（2）构建开放且高可扩展的智能系统栈，充分挖掘系统效率；（3）构建与人类知识对齐的决策计算库，自动化地完成人类知识解构和对齐，降低计算复杂度。

华为技术有限公司资深技术专家孙宏伟分析了多样性计算的应用与挑战，介绍了核心数据结构及数据处理模式的演进过程：从关系型数据库到分布式数据库，再到推广半结构化数据，最后演化至大模型非结构化数据。他提出，数据应用的变化推动了计算系统中计算、存储、互联各部分的变革，而器件创新则推动了计算系统影响因素的转换。在此基础上，他还总结了未来计算的十个核心问题，涵盖计算架构、内存池化、互联、编译等方面。

此外，与会专家还对多样算力和并网方案进行了集中研讨，形成如下3条结论：（1）明确需求是算力网建设要解决的首要问题，需要考虑算力网的易用性，单纯从技术领域看待问题是管中窥豹；（2）算力网只有通过抽象屏蔽计算的异构性才能释放算力能力，算力网要解决跨平台、多样化的难题；（3）构建算力网需要模式创新，算力网的构建靠国家需求推动，未来能助力产业升级和数字经济发展，降低算力网的使用成本和复杂性是决定算力网是否能够取得成功的关键因素，算力应通过网络服务更多的人。

分布式存储设施

算力网的应用在形态和场景上都有多样性，具有多种数据存储访问模式，以满足不同场景下的性能和功能需求。例如，一些应用为保证实时性需求，可能需要低延迟的数据存储访问，而另一些应用可能更注重数据存储的安全性和可靠性。算力网的分布式存储架构需要统一化和智能化，从而满足不同应用的数据存储访问需求。这给算力网的分布式存储系统的设计和实现带来了很大的挑战。

实践分享

在高性能计算和云计算等传统的分布式系统中，用户需要依次完成数据存储部署和计算任务部署。例如，在大数据应用处理中，用户需要先将切分（sharding）后的数据存储在不同计算节点，之后再部署大数据计算任务，即分布式存储系统对用户而言并不透明。算力网应提倡像电力网和电话网那样的算力资源按需使用模式，用户只需关心计算任务的执行结果，而无须关心数据的存储和访问细节。在算力资源按需使用模式下，算力网需要针对计算任务的处理数据和算力资源的状态数据这两类主要数据构建分布式存储系统。算力资源的状态数据是为计算任务提供准确部署和调度策略的重要依据，支撑计算任务高效完成数据处理和计算过程。面向计算任务处理数据和算力资源状态数据的读写访问，算力网分布式存储系统的架构设计需要考虑两个问题：（1）如何构建全网统一的存储接口和抽象；（2）如何优化系统架构充分发挥各种异构存储硬件的性能。

针对统一存储接口的问题，传统的分布式系统通过提供分布式文件存储、分布式对象存储、分布式键值对存储、分布式数据库、分布式块存储等多层级抽象来满足不同场景的应用需求，用户可以根据自身的应用需求，部署或调用不同层级的分布式存储系统完成应用构建。例如，云计算中的Alluxio系统桥接了应用框架（Spark、Flink和TensorFlow等）与分布式存储（Amazon S3、GlusterFS和HDFS等），统一了存储在不同分布式存储系统中的数据接口，为上层应用框架提供统一的客户端API和全局命名空间。Alluxio系统本质上是通过中间层提供统一数据接口，而算力网的分布式存储架构应该通过构建一种统一的存储抽象，原生地支持数据统一存储和访问。

针对系统架构优化问题，传统的分布式存储方案存在系统调用、锁机制、通信协议等开销，无法充分发挥存储硬件本身的读写访问性能，应在架构上优化分布式存储系统，尽可能缩短存储路径，减少多余的系统开销。例如，可使用SPDK和DPDK

等面向存储读写访问的用户态协议栈，绕过内核态进行数据的读写操作，避免用户态和内核态频繁切换产生的开销；或使用 libfabric 和 UCX 等面向 RDMA 的高性能通信框架，缩短数据在分布式节点之间的传输链路，加速数据在分布式节点之间的读写过程；或基于 CXL 总线构建大内存架构，建立分布式的内存池化系统，提升算力网存储系统的读写性能和容量可扩展能力。

观点争鸣

中山大学教授肖侬指出，新一代计算基础设施的架构变化和挑战主要集中在大内存并行计算系统和跨域数据共享方面。针对大内存并行计算系统，统一的大内存总线结构和 CXL 协议提供了更高的性能和灵活性，同时也面临着资源分配、容错容忍、缓存一致性和软件适配等挑战。对于跨域数据共享，构建跨域统一的全局命名空间和实现计算与数据访问的协同成为了关键问题。为应对这些挑战，面向新一代计算基础设施的存储系统应考虑单节点内的异构器件和内存管理、超节点内的一致性内存管理、大规模计算系统中的运行时环境以及算力网上的数据共享访问。同时，技术驱动和应用驱动也是推动新一代计算基础设施架构变化的两个重要因素。

华中科技大学教授周可从智能存储角度分析了分布式存储的背景和发展，他指出智能存储是一种结合了存储系统和人工智能技术的新型存储架构。从存储的角度，智能存储关注系统结构、数据管理和运维管理方面的优化；从 AI 的角度，智能存储具备感知决策能力和数据认知能力。感知决策能力包括智能缓存、资源分配、自动调参和故障预测，通过学习和分析存储系统的工作负载和运行状态，实现智能化的决策和优化；数据认知能力涵盖内容查询、知识推理和智慧生成等能力，通过深度学习和自然语言处理等技术，实现对存储数据的理解和利用。新型存储架构需要通过并行和分布式技术，充分利用存储系统的多个 I/O 路径，以提高存储的吞吐量和并发性能，应统筹多任务 AI 使能，实现多任务的智能化处理，提升存储系统的整体智能水平。

厦门大学教授张一鸣指出，在异构混合云存储中使用缓存技术存在诸多问题，包括 I/O 模式不具有局部性、单次读取操作过多、缓存一致性问题。相比之下，多级混合存储是一个更好的解决方案。在随机存取存储器（RAM）和硬盘（HDD）的混合存储中，可以使用 primary-recovery-backup 的方式，将数据的一份存在 RAM 中，其余备份存在 HDD 中，通过“一跳原则”控制故障检测和恢复的范围，并使用 CubicRing 结构让所有节点同时充当主节点和备份节点。他认为，面向算力网，跨网络的文件同步也是一种重要的多级混合存储技术，可以基于固定分块和变长分块的增量同步过程，减少传输和校验的开销。

清华大学研究员陈康认为存储系统之所以难以实现，主要是因为需要同时满足高性能、高可靠性和高可用性的要求，并且需要保证存储的可维护性。传统的存储架构已经无法跟上硬件性能的发展，需要尽可能缩短存储路径，减少拷贝和内核操作。同时，人工智能应用中的海量小文件访问也对存储系统提出了新的挑战，单一的元数据服务器难以承受如此重负，并且会对其他应用造成冲突干扰。为了应对这些挑战，下一代分布式存储系统需要充分利用硬件性能，重视元数据和小文件的读写性能，处理大量的元数据，并具备深层次的层级结构；还应具备统一的存储基座，支持不同的存储接口，并针对不同系统进行特别优化，避免过多地依赖内核。在分布式存储方面，工业界和学术界的专家均认为分布式存储与数据管理相关，而数据管理则与应用发展联系紧密。在国家新基建的背景下，分布式存储需要围绕国家数据基础设施这一大方针展开，为用户提供统一服务。专家们还指出存储最本质的需求是在时间和空间层面保证数据的持久性，当前的云存储对用户来说仍然是有感的，在算力网中，分布式存储应该像电力网和电话网一样做到用户无感。因此，构建算力网时，应该将科学研究、工程技术、应用等统一起来，不要相互隔离，在模式上进行创新，做出里程碑式的存储系统。

新型计算机网络

实践分享

网络是算力网的根基，是将广泛分布的算力资源连接在一起的物理介质。因此，网络的功能和性能是决定算力网能否成功的关键因素之一。目前在业界已有一些相关研究和技术成果，包括算网融合架构、在网计算、多维标识网络以及新型智算中心网络等。

算网融合的发展历经了从数据中心内算网融合到云网融合，再到云－边－端算网融合的三个重要阶段。第一阶段旨在实现数据中心内大规模服务器间的高性能网络连接；第二阶段旨在实现分布式云数据中心之间、用户接入点之间的高性能网络连接，并确保任意用户对任意云服务的高效访问；第三阶段旨在实现云－边－端泛在联接与泛在计算紧密结合，并通过统一弹性调配确保各类算力应用的高效执行。

在网计算理念为算网融合探索了一种非常独特的发展途径，其核心思想是在网络设备中执行原本在终端主机上运行的程序，显著减少网络中的流量传输，加速分布式计算业务的迭代执行。在网计算理念符合机器学习、大数据处理等典型分布式系统的两大发展趋势：（1）可编程数据平面使网络本身正在变得具有计算能力；（2）很多分布式计算应用的性能瓶颈从计算环节转变到网络传输环节。在网计算侧重使用现有的可编程网络设备，尤其是可编程的流量转发设备，实现典型集合操作或通用计算功能。在网计算改变了网络数据传输及处理机制，也突破了现有的分布式计算模式，对于分布式应用的加速赋能具有重要意义。

标识命名是支撑网络运行的重要基石，传统互联网标识体系的不足在于网络层标识IP地址具有身份与位置的双重属性、应用层的统一资源标识符（URI）解析面向主机位置、传输层缺少独立标识以及网络层标识承载单一等。算网融合中的多元化服务对IP标识寻址机制提出了新的需求。多维标识网络通过身份与位置分离机制，构造多维统一用户标识（UID）和网络连接标识（NID），分别用于标识网络用户身份要素和网络用户位置要素。为了提升路由转发效率，网络层转发方式采用UID和NID混合寻址方式。多维标识网络技术实现了多元异构网络对象的类型、归属、特性、位置等多语义特征要素的融合标识与表征。现有的算力寻址存在算力和IP未能深度融合、缺乏多维属性表征能力和智慧服务能力等挑战。

在传统算力中心的网络架构下，存在应用和网络解耦、拥塞控制粒度粗以及服务器侧无法感知细粒度网络信息等不足。为了满足人工智能大模型时代下计算和网络的融合需求，新型智算中心的网络设计具有重要的意义。数据流的路径选择和拥塞控制是影响网络性能的核心因素，新型智算中心通过设计新型网络协议栈，结合网络精准信息与应用诉求，对这两方面进行了优化，从盲发的被动拥塞控制转变为基于感知的主动流量控制，从“局部”的决策转发变为“全局”最优调度。此外，该技术通过通信库实时感知应用诉求并输入到网络协议栈以优化应用性能，并且引入了可编程交换芯片，通过对数据面编程的技术使网内信息可精准定制。在算网融合的背景下，新型智算中心的网络需要在芯片、数据集、架构和交付性能等方面进行优化。这些技术为构建无阻塞、高带宽、低时延的新型智算中心网络以及在AI产业中形成标准开放的技术体系打下了良好基础。

观点争鸣

中国移动研究院首席技术专家孙涛认为，算力网络是一种以算为中心、网为根基的新型信息基础设施。为实现算网一体化，中国移动研究院从算力原生、全调度以太网、算力路由、在网计算和数据快递五个方面展开了技术探索。首先，通过为异构算力构建统一抽象机制，打破“编译－链接－执行”的紧耦合生态模式，追求更普惠的计算模式。其次，采用基于报文的转发及调度机制，构建了无阻塞、高带宽、低时延的新型智算中心网络。在网络路由系统中引入计算因子，将算力请求动态引导到最佳

的算力服务节点，实现了网络和计算的联合调度优化。再次，在网计算突破了现有分布式计算模式的通信瓶颈和性能拓展瓶颈，将网络数据传输及处理机制从“端到端”改变为“端网端”，对可靠传输提出了更高的要求。最后，基于UDP协议设计新型传输协议，充分利用高带宽网络，实现了超长距广域网环境下的超高吞吐数据传输。结合中国移动的创新试验网CFITI的进展，孙涛进一步阐明算网一体已积累较好的基础，面临巨大的机遇。此外，中国移动在国际互联网工程任务组（IETF）发起了算力路由工作组。

中国科学院计算机网络信息中心研究员谢高岗认为算力互联包含计算机内部互联、数据中心网络、数据中心间网络等形态。算力网对互联网体系结构的变革影响微乎其微，算力应用产生的网络流量在互联网流量中的占比非常低，对于互联网而言，算力网更适合在overlay层创新。但是，互联网对网内的算力资源存在巨大需求，尤其在端边云协同、网络功能虚拟化、网络协议动态部署和基础设施持续演进等场景中需求更加鲜明。中国科学院开展了算力网基础设施的初步探索，主要面临网络多维资源的测量感知、灵活高通量数据包计算、大规模多维度资源分配控制和智能运维（AIOps）等方面的挑战。他认为，测量感知是提供网络功能与业务功能服务化的基础，随着资源维度增加，系统复杂度指数也在增加，如何设计既能保证测量精度，又能降低测量开销的高效数据结构和测量算法是难点。

国防科技大学教授郭得科指出，随着分布式计算系统的发展，很多应用的整体性能瓶颈已经从计算环节转向网络通信环节，在网计算成为提升算力效用的一种独特且高效的方法。在网计算具有打破算网边界、加速传输计算过程和提升系统整体效能的优势。他还介绍了在网计算在加速向量数据的同步聚合和键－值（key-value）数据的异步聚合等典型分布式应用场景的显著优势。他认为，在网计算改变了网络数据传输及处理机制，对于分布式应用的加速赋能有着重要意义。在网计算在发展过程中首先需要解决应用场景的差异化、设计实现封闭化以及编程语言门槛高等挑战性问题。最后，他认为在网计算的发展趋势是编程范式统一、通信原语统一、多资源联合调度、面向分布式应用的通用网内聚合服务、面向边缘侧联邦学习加速的在网计算服务等。

北京交通大学教授郜帅指出，解决互联网标识承载单一等基础性问题的关键在于设计出具备多种语义的网络标识，使网络具备面向多种网络对象寻址的能力。目前，应用和网络之间的通道具有唯一性，这不利于网络空间新技术的发展。多维标识融合体系通过身份与位置分离机制，构造了多维统一用户标识和网络连接标识，实现了多语义特征要素的融合标识与表征，从而具有智慧路由和决策的优势。他强调了多维标识对算力网的助力作用，传统IP标识寻址机制难以满足算网融合需求，因此如何为IP标识增加算力与网络融合寻址能力成为关键。针对这一问题，他指出算网融合应具有灵活性、可扩展性、兼容性和多维属性的特征。由于天然连通服务与网络，多维标识具备多维扩展属性，能在网络层增加算力感知与寻址能力。最后，他认为可利用多维标识融合网络提供多元化新型网络寻址和路由能力，增强网络服务效能，以扩展适应未来新型网络的应用。

阿里云网络研究负责人翟恩南分析了面向AI的新型数据中心智算网络体系的机遇和挑战，认为网络已成为大模型训练的瓶颈。大模型的训练需要庞大的算力支持，而算力的线性扩展依赖数据中心的网络互联架构。他认为传统数据中心不能很好地支持大模型训练业务，因为大模型训练的通信具有数据流量少、带宽高、周期性等特点，传统等价多路径（ECMP）/流量控制/网络监控方法不再适用，大模型训练过程的长尾时延非常关键。因此他认为，影响网络性能的核心因素是路径选择和拥塞控制。然而，传统网络存在应用和网络解耦的问题，网络被视为黑盒，拥塞控制粒度粗，路径选择基于ECMP无法适应大象流（elephant flow）。因此，需要设计新型网络协议栈，结合网络的精准信息和应用需求，进行拥塞控制和路径选择。他还介绍了阿里云在拥塞控制和路径选择方面的最新进展。最后，

他指出算力网需要满足应用模式创新或低成本的基础设施供给，才能实现规模化应用和推广。

分布式系统软件

分布式系统软件是算力网的灵魂，系统架构有松耦合与紧耦合两个类型。松耦合的分布式系统软件的代表是万维网，程序员通过一套共同遵循的架构风格（REST）和应用协议（HTTP）协同开发应用与服务。紧耦合的代表是超级计算机的系统软件，它将集群视作一台并行计算机，并为其开发操作系统来管理超算作业在众多计算机节点上的执行。目前，算力网的系统软件研发也存在这种分类。网络领域从业者认为计算应该融入传输过程，代表技术是在网计算。计算机从业者多数认为算力网应是一台分布式计算机，系统软件是一个相对紧耦合的分布式操作系统，负责屏蔽算力资源的异构、异属、异地属性，以及管理多样化的算力网应用。本节从计算机从业者的视角探讨算力网的资源管理、运行时抽象、编程方法等基础学术问题。

实践分享

计算机系统软件研究的核心问题是抽象。单机操作系统的核心抽象是进程，超级计算机系统的核心抽象是作业，云计算系统的核心抽象是容器。算力网面临的核心学术问题是需要什么样的新抽象来桥接大规模算力网应用和分布式算力资源，并为应用的编程提供基础支持。从系统的角度来讲，这个抽象涉及以下四个有待产学研各界共同努力解决的基本问题。

1. 如何统一命名空间（universal name space）。命名空间是任何一个计算机系统的基本问题，系统必须对其拥有的硬件资源进行统一编址和抽象，以便上层应用方便地使用资源。编址的目的是快速准确地查询和定位资源，抽象的目的是通过统一的方式使用资源。单机系统的虚拟地址、文件标识符，以及互联网的IP地址和万维网的URL均是命名空间的例子。算力网的资源来自不同机构、地域，且具有很强的异构特性，这使它命名空间的统一比其他任何一个计算系统都更具挑战性，该问题也被称为算力资源的池化。当前国内外的研究工作尚未系统地考虑这个问题。算力资源的池化可以学习电话网、互联网和万维网的经验，在技术上提供统一坐标系，以支持各种治理需求，支持用一个算力账户无缝访问全网资源。在使用上，全国各地的用户可向一个算力网提交计算任务，算力网就像一个“大队列”，统一调度任务和即时分配资源，从而有效避免云计算中“占而不用”（stranding）的问题。在运营方面，应鼓励多种资源合作共享，形成价格优势。

2. 如何提供一个像万维网网页一样的运行时抽象。万维网本质上是围绕网页建立的一套结构化信息传输、共享和显示的分布式系统，其核心价值在网页。程序员通过开发和部署网页为终端用户提供服务，终端用户通过访问网页间接使用万维网上的数据和算力等资源。类似的抽象还有互联网的IP数据包、超级计算机的作业和云计算系统的容器与微服务。作为一个新事物，算力网应该有一个与网页等类似的核心抽象，并围绕该抽象构建其传输、执行、结果存储的分布式系统。如果不存在这样一个抽象，那么算力网本质上依然是云计算、超算或互联网。

3. 如何提供统一的编程方法。算力网包含来自端边云的资源，具有高度异构特性。如何像万维网的HTML与JavaScript一样提供一套编程语言和框架，实现同一套程序代码在不同体系结构的算力上无须修改即可直接编译运行，是算力网在编程方面需要解决的核心问题。超算是算力网计算资源的一个子集，即使在这个领域，统一编程模型与框架依旧是一个基本挑战。目前，美国、欧洲等正在研制统一的超算并行编程模型，以实现一套程序代码在不同异构系统上直接编译运行的目标，但尚未实现一套程序代码在不同超算系统上达到一致的性能，即存在“代码可移植但性能难移植”的问题。国产异构超算系统统一并行编程模型面临的一大挑战在于国内异构处理器架构复杂多样，无法直接应用国际上已有的统一并行编程模型。学术界亦有通过研究高层次图优化和低层次算子编写来统一深度学习

领域编程模型的工作。超算领域尚且如此困难，算力网的统一编程方法研究将更具挑战性。

4. 如何系统性地评价算力网的性能。这个问题涉及研究和建设算力网的必要性：现有的云计算和超算是否无法满足新场景和应用的计算性能需求，以至于需要建设一个新的信息基础设施？答案是肯定的。据公开的文献资料显示，即使是全球最先进的谷歌和 Mate 的数据中心 CPU 资源利用率，也无法超过 40%[2]。Mate 在最新的论文中称，使用一系列优化技术的 xFaaS 平台可以将整体资源利用率提升至 60%[3]。另外，使用 Docker 启动容器通常需要十几秒才能执行一个原本运行时间只需要几毫秒的程序。这是因为容器启动需要下载镜像、容器化、运行时初始化等多个费时的步骤。虽然可以通过复用初始化好的内存状态、remote fork 等方法加快容器启动，但容器等需要提前预分配资源的部署与执行模式阻碍了系统性能上限的提升。这种水平的利用率和服务质量无法满足数字经济时代算力作为生产力基本要素而基础设施化的要求。因此，需要一套体现算力网系统价值和用户价值（应用开发者、资源整合者、运营商和消费者）的评价体系。信息高铁算力网提出了通量（goodput）和良率两个核心指标，分别体现其系统价值和用户价值。通量指单位时间内系统能够保质完成的任务数，良率则指保质完成的任务数占用户提交的任务数的比例。

观点争鸣

中国科学院计算技术研究所孙凝晖院士强调，系统软件最重要的部分是资源管理，如命名问题、并网调度问题等。另外，如何支持新的编程框架和部署方式也是重点问题。对于算力网来说，其部署一定不是超算的部署模式，而是如同互联网般的软件开发和部署模式。他提出了算力网部署的“三异”：异构、异属、异地，分别表示架构、所属服务商以及地理位置的差异性。他希望算力网能够形成“小云协同”“小云联邦”的模式，使用小云解决专项问题，以此突破大云，形成反垄断。

国防科技大学教授王怀民院士对未来战略问题进行了阐述，他认为现阶段国家的发展战略利好华为等企业。在美国的限制下，要注重国家科技力量的有效运用，企业各有专攻，相互协作才能使中国产品获得全世界市场的认可。在科技领域，我们要做领航者，并利用好群智的力量。他希望算力网能够像高铁、新能源一样，未来成为国家发展的重要引擎之一。

中国科学院计算技术研究所研究员徐志伟提出，算力网将是一个多样性的产业生态及学术方向，需要形成“算力网基本学术辞典”，同时算力网需要站在巨人肩膀上进行创新。他提出了算力网的 3 个基础学术问题：（1）是否需要统一命名空间；（2）如何前瞻且客观地评价算力网的用户价值；（3）如何前瞻且客观地评价算力网的性能。基于上述基础学术问题，他提出算力网研究需要 3 个方面的学术抽象：（1）算力网为资源空间赋名；（2）算力网页；（3）算力网本征性能公式。

清华大学教授翟季冬分析了面向国产异构算力的编程模型及编译优化的研究挑战及研究目标。针对国产异构算力构成的超算系统，他指出在编程及编译方面目前存在两个研究挑战：（1）国产架构复杂多样，实现国产异构系统的统一并行编程模型的难度大；（2）性能移植难，一套程序代码在不同的超算系统难以获得一致的性能。为构建面向新一代国产异构系统的统一并行编程模型，他提出应该基于国际统一并行编程模型标准 SYCL，增加对国产硬件的支持；设计异构处理器的统一中间表示，挖掘复杂多样化架构的共性优化；设计多维度感知的异构处理器体系结构，建立程序与处理器间的最佳匹配；兼顾功能通用性和性能可扩展需求，实现多层次感知的高效运行时系统。

中国移动研究院网络与 IT 技术研究所副所长刘景磊指出，中国移动十分关注民营和国资的战略配合，将会采用对等并网、吸纳算力、带流量帮销售等方式与阿里云进行合作。他强调，要在算力网体系结构层面实现反垄断，国内的开源项目应保持各层的融通，全力扶持华为等企业形成单体技术栈；希望华为能够建立起像 NVIDIA CUDA 一样的生态优势，希望寒武纪等保持关键技术栈的竞争力。这就要求在每一层做

好抽象，并支持跨不同硬件架构的应用迁移。

华为技术有限公司资深技术专家孙宏伟认为，我们需要利用软件生态尤其是应用软件生态的强大力量，在国内建设一个大的算力网软件体系，并和国际对接。另外，他建议算力网硬件厂商能够做到硬件开放、软件开源、积极拥抱生态，例如昇腾对接 PyTorch、百度飞桨（PaddlePaddle）等计算框架，Mindspore 对接寒武纪等硬件，以开放的态度应对算力网的技术挑战。

共识与倡议

共识

1. 算力已成为数字经济时代社会基本生产力要素之一，科学计算、AI 大模型应用催生了巨大的算力需求，算力网的研究和建设是解决需求的有效途径。

2. 数据已成为数字经济时代社会的重要生产资料之一，分布式存储领域应连通科研、学术和产业，建设国家数据基础设施，为数据的存储和流通提供统一服务。

3. 算力网的灵魂是其分布式操作系统，应设计可移植的统一并行编程模型，研究新的操作系统抽象和实现，以满足多种类型应用的功能和性能需求。

4. 网络的功能和性能是决定算力网能否成功的关键因素之一，应研究新型智算中心网络、广域算力节点的网络互联体系、算网联合路由等，以满足算力网应用的高性能传输需求。

5. 算力网的建设需要借助群智的力量，利用生态的力量团结产学研各界，建立大的技术体系和标准，积极拥抱开源开放生态，应对技术挑战。

倡议

1. 计算机发明以来，美国凭借先发优势和信息高速公路等计划引领信息技术 70 余年。算力网有可能是中国原创的重大发明与工程，中国学者应敢于提出颠覆式创新技术。

2. 分布式计算系统是算力网的学科基座，有很多基础科学问题亟待梳理，分布式计算系统领域的学者应该带头提炼相关科学问题，从学术上为算力网的研究和建设打好基础。

3. 算力网的作用远不止满足 AI 大模型的算力需求，它的原创应用不应直接由做底层技术的人提出，而应从创新、培育、实验、演化而来，呼吁政府、企业、学界建设相应的算力实验平台，培养算力网的杀手级应用。 ■

致谢：感谢参加本次秀湖会议的全体人员。感谢 CCF 业务总部工作团队为本次会议提供的优质服务。感谢并行科技及秀湖会议年度合作单位腾讯、华为对本次会议的大力支持。

整理：郭得科　彭晓晖　王一帆　廖小飞

附：与会专家名单

特邀代表：孙凝晖　王怀民　孙　滔　王亚晨

正式代表：陈　康　郜　帅　郭得科　胡　杏
金　海　刘景磊　罗瑞丽　石宣化
孙宏伟　唐　卓　王晓虹　王宝川
魏星达　肖　侬　谢高岗　徐志伟
余　跃　翟恩南　翟季冬　张一鸣
甄亚楠　周　可

论坛组织：廖小飞　彭晓晖　王　雄

会议秘书：俞子舒　李奉治

参考文献

[1] 王磊, 孙凝晖. BOPs：一种算力度量指标[J]. 中国计算机学会通讯, 2024, 20(1): 44-49.

[2] Delimitrou C, Kozyrakis C. Quasar: Resource-efficient and QoS-aware cluster management[C] // *Proceedings of the 19th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*. 2014: 127-144.

[3] Sahraei A, Demetriou S, Sobhgol A, et al. XFaaS: Hyperscale and Low Cost Serverless Functions at Meta[C]// *Proceedings of the 29th Symposium on Operating Systems Principles (SOSP)*. 2023: 231-246.

（本文责任编委：**郭得科**）