算力网络场景下的超算互联网建设

陈进雄
广东南方电信规划咨询设计院有限公司　广东　深圳　518000

摘　要：在当今信息技术飞速发展的背景下，大数据管理、智慧城市建设、智能制造、自动化及远程驾驶技术以及区块链等先进技术正呈现迅猛发展的趋势，其进一步丰富了新型计算模式的应用场景，从而引发对算力资源的强烈需求，推动数据中心的快速扩展与进化。因此，本文将结合算力网络，讨论算力网络场景下的超算互联网建设策略。

关键词：算力网络；超算互联网；建设方法

前言：截止到2022年初期，中国已成功建立起500万个标准机架，总算力达到惊人的130EFLOPS。但面对如此庞大的体系，数据中心平均利用率大约只有55%，这主要是由于算力需求的差异性、网络传输的性能局限以及算力成本之间的矛盾所导致的。为了有效解决这些问题，世界顶级科技公司已经开始寻求创新方案。例如，微软和Facebook选择在海底或北极地区建设数据中心，以利用自然环境降低能耗；国内大企业如阿里巴巴、腾讯、百度和华为也纷纷把数据中心设置在资源丰富的西部地区。这些做法不仅是对能源供需不均衡状态的直接回应，也是迈向“东数西算”战略的重要一步。

1 算力网络

2021年5月，我国迈入一个新的技术革新时代，国家发改委联合其他三个部门共同发布《全国一体化大数据中心协同创新体系算力枢纽实施方案》。这一关键文件首次将“算力网络”纳入了国家级的讨论与规划之中，突显了构建集数据中心、云计算及大数据为一体的创新型算力网络的重要性。在这份方案中，不仅明确了建设全国范围内统一的算力网络和国家枢纽节点的宏伟蓝图，也启动高效能的“东数西算”项目，旨在推动我国在全球数据计算领域的领先地位。

随着“东数西算”工程的逐步实施，我国的数据中心网络逐渐呈现出云化的趋势，这标志着我国在技术进步的道路上又迈出坚实的一步。这一进步不仅仅体现在云计算和网络的深度融合上，更在于算力与网络的一体化，使算力的可达性和泛在性成为可能。如此高度的协同意味着将计算能力深度集成至网络之中，实现了云网络、边缘计算与设备端的无缝对接和高效运作。

IETF和华为提出的计算优先网络（CFN1）强调在网络层之上实现计算任务的动态路由，这一机制能根据实时的计算资源性能和网络状况灵活地调整计算任务的派遣。类似地，中国电信定义的算力网络（CPN）侧重于一种新型信息基础设施，该基础设施在云、网、边之间按需分配和灵活调度计算、存储及网络资源。另一方面，中国联通和中国移动各自提出的算力网络定义则更注重于算力与网络的深度一体化。中国联通认为算力网络是由云网协同发展演变而来，不仅包括了AI网络连接服务和用户数据算力连接网络，还涵盖城域光网、5G URLLC网络等，强调了算力服务的新型网络设备和超融合设备的重要性。而中国移动将算力网络视为一个以算力为中心，将网、云、数、智、安、边、端、链（ABCDNETS）等多要素融合的新型信息基础设施，强调了其泛在化的社会级服务特性。

2 算力网络基础设施架构思路

2.1 资源纳管

2.1.1 算网资源智能化感知

在超算中心和边缘数据中心接入及资源自动感知的问题上，首要任务是定义算力网络资源池的结点类型和采用的接入技术。其中，对于计划加入网络的国家级、区域级甚至边缘级的算力集群，需要选定结点的具体类型，并为这些算力资源，包括计算资源、存储设施、网络带宽、软件支持及数据内容等多个层面的内容，以此来确立接入标准。之后，将对这些算力资源进行彻底的审核、抽象化建模和封装处理，确保这些资源可以被顺利地接入并纳入到网络之中。接着，通过算力资源的注册与发布流程，创建一份详尽的服务目录，以满足终端用户的各样需求。接下来的步骤，涉及到研发代理组件及配套的南北向接口，这些技术手段能够通过网络联接、安全策略的配置、实时的监控行为以及消息的订阅和发布等机制，实现算力资源的自动感知功能[1]。

2.1.2 算例网络多维资源协同调度

在算力网络资源成功接入之后，确保资源得到统一高效的调度与编排，就需要深入探究具体的调度策略和技术手段。首先，需要深入了解所接入的异质算力

集群所采用的调度器类型，比如常见的Slurm、PBS和Volcano。紧接着，对这些调度器的任务执行模式、调用参数及关联过程进行详尽分析，从而提炼出它们的共性与个性，从而构建一个统一的算力集成调度模型。

其次，算力资源的混合调度场景成为重点。当协同计算任务提交时，系统会分析调度参数并确定不同算力集群可提供的计算资源、存储容量和网络带宽。依据细化的调度策略和手法，结合算力性能和网络路由的优选，选取最适合的算力集群进行任务计算。在资源调度方面，存在如基于优先级排序、负载平衡、成本效益和任务及资源的亲和性等四种策略。调度时应全方面考虑计算、存储、网络和软件等资源的配合使用，同时根据调度策略的具体影响因素和目标，实现资源利用的多维度协同[2]。

2.1.3 数据统一存储

以超级计算的应用场景为例，目前的趋势展现出从传统的计算密集型超级计算向数据密集型超级计算的转变，即向以数据为核心的高性能数据分析平台的发展。其通过整合存储能力中心节点，联合提供多方面的数据服务。面对数据的跨域分散和自治隔离，现有体系结构未能有效地聚合这些数据，难以实施有力的管理与共享。这不仅导致数据在多个存储中心重复存储，还降低了数据访问效率，进而影响了数据处理的整体性能。面对算力网络环境下数据处理与流转的需求，攻克数据统一存储与高效流转的技术障碍显得尤为关键。以下三个问题需要着重考虑：

第一，面对存储资源和数据资源在广阔区域内分散且各自独立运行的问题，需制定全局数据空间描述的方法，并采用恰当的数据索引技术，以实现对异构存储资源的统一管理及高效访问。

第二，考虑到算力网络环境下可能面临的带宽限制和较高的延迟，采用基于数据血缘关系的智能数据流动选路技术和加速技术变得必要。这种技术能够支持智能网络路由选择、数据压缩、数据合并与拆分，从而加快网络内大规模文件或小文件数据的高速流动或迁移，提升算力网络中不同节点之间的数据传输效率。

第三，构建一个数据流通总线，以联接不同的数据存储系统，支持异构存储

系统中多样数据的统一跨域传输[3]。通过实时考量网络带宽、数据所在位置等多重因素进行路由选择、数据压缩、数据合并与拆分，解决数据高速智能调度的问题。例如，在云计算平台的基础设施中，设计专门的存储流转和数据管理模块，为常见的存储系统（如并行文件系统Lustre、文件存储系统NFS和对象存储OSS）提供管理能力，其中的代理组件可被部署至各个存储能力中心节点，以优化整体操作流程。

2.2 网络能力结构

2.2.1 组网拓扑

为迎合不同业务需求，在覆盖“省市”两级的超算算力网络中，可以引入SRv6网络分片技术（Segment Routing over IPv6）。这种技术可以使物理光网络，划分为若干网络业务平面，形成一个“多业务面共存”的网络结构。在这样的结构之下，根据业务应用的具体SLA（服务级别协议）需求，选择将业务部署在适合的网络分片中。网络层面，通过物理层的技术手段将网络一分为多，为每个分片网络分配专属的队列资源，从而确保数据能够实时无阻塞地转发，实现从源头到终端的带宽确保及业务间的严格隔离。为增强这种隔离，并依据业务需求进一步精细化管理，每个网络分片内部可以基于业务种类分配不同的虚拟私网（VPN）来实现软隔离。这种软隔离方法允许在同一网络分片内部划分出更细致的业务子网络。网络的运维管理方面，通过采用软件定义网络（SDN）技术，可以实现对业务的快速部署和流量的灵活调整。

2.2.2 地址规划

在超级计算（超算）算力网络的构建过程中，IPv6地址族被选为基础网络地址的核心，以支持通过EVPN+SRv6技术同时承载IPv4和IPv6的服务以及网络管理任务。IPv6地址由128个二进制位组成，并使用十六进制的表示方法，提供了几乎无限的地址空间。为了高效及规范地分配这些地址，超算算力网络采取了一种“申请再使用”的策略：由负责网络管理的单位为每个接入点分配一个96位的IPv6地址前缀，并确保每个节点获得一个32位的充足的地址空间[4]。A市建立的超算网络中，IP地址根据其使用目的分为三大类：服务地址、终端地址和管理地址，分别用于不同类型的设备和需求，确保网络的有序运行和安全。

服务地址主要分配给需提供外部服务的服务器和存储设备等，这些地址既包括IPv4也包括IPv6地址，保证了服务的高可用性和广泛的可访问性。终端地址则分配给内部使用但不直接面向服务提供的设备，如工作站、笔记本电脑和移动设备等，同样支持IPv4和IPv6地址，以应对多样的终端设备需求。管理地址专用于网络内部的设备和系统，例如网络设备的环回地址、互联接口地址以及网络管理系统、安全管理系统和DNS服务器等的地址。

3 算力网络场景下的超算互联网建设实践分析

3.1 整体部署

该计划选定A市、B市、C市作为三大核心节点，利用100 Gbit/s的光纤环网专线实现这些核心节点之间的直接连接。周边的13个地（市）则通过与最邻近的核心节点的10 Gbit/s的网络连接，实现广域网络的整体布局。从服务部署层面来看，各个地节点均安装了运营商路由器（PE）、SD-WAN网关、资源纳管和调度服务组件，旨在优化网络性能和资源分配。

3.2 应用实现

3.2.1 需求分析

根据国家对地观测科学数据中心的统计信息，该中心每年聚集逾10PB的对地观测信息。为了最大限度地利用这些庞大的数据资源，不仅需配备大容量且高效的存储系统，以满足海量数据保存的需求，还必须依靠PB级的高性能、智能与云计算等多样化计算资源共同作用，完成数据处理与分析工作。此外，为确保数据顺畅传输，10 Gbit/s的高速专属网络亦是不可或缺的条件。值得注意的是，上述数据仅仅包括了遥感的初级数据层面。若进一步对更高阶的数据产品进行生成、分析及探索，则对计算力、存储空间及网络资源的需求将更为巨大[5]。

3.2.2 遥感数据

遥感数据产品生产过程作为一个例子，演示了算力网络在实际操作中的关键作用。首先，整个生产过程依赖于从Landsat8原始数据开始，依次经历地表反射率数据的生成、地形校正（TC产品）以及F mask云掩膜产品的生产，接着对TC产品应用F mask云掩膜，最后进行数据的拼接和图像的制作输出。这个生产线中，各步骤所需的计算资源支持各不相同：地表反射率数据的生成由于其庞大的计算需求，采用高性能计算集群来完成计算任务和数据的产出。

在进行数据产品的拼接及图像输出时，需处理大量图形拼接与绘制工作，因此需部署具备GPU资源的云主机来执行这些任务，最终产出JPG格式的图像。此过程中还体现了算力网络在管理跨域异构存储系统和实现数据智能流动方面的功能。

3.3 应用场景

在具体的应用支持领域内，算力网络适配多种场景，包括分布式数据处理、高通量计算、弱耦合任务以及工作流管理等。这些领域能够从表1中详细了解。通过对数据的处理角度来看，算力网络对于那些要求在跨域分布式存储系统中进行数据处理的场景表现出色。从计算任务的角度观察，算力网络为那些结构松散且依赖工作流模式的计算任务提供强大支持。它能实现任务的并行执行、拆分以及最终结果的整合，并且特别强调以应用任务流为中心的算力资源调度及以数据流动性为核心的智能管理。

结语：本文基于超级计算网络（超算网络）的互联网建设，提出一种分级的设计方法，探讨算力基础设施的架构需求。这个平台遵循省级一体化大数据中心的空间布局要求，同时利用国家超算中心的网络和算力优势，建立以三大核心节点为基础的省级一体化算力中心，实现三个低延时的核心算力区的形成。该平台支持扩展至其他16个地区，促进低延时边缘算力中心的接入，其标志着我国在全面推进数字经济发展道路上再次取得重要进展。

参考文献

[1]尹林,王富,王小龙,等.面向算力互联的快速光交换技术研究[J/OL].光通信研究,1-11[2024-06-18].

[2]王继彬,张虎,陈静,等.算力网络场景下的超算互联网建设探索与实践[J].邮电设计技术,2024,(02):14-21.

[3]刘扬.浅析算力网络基础设施及我省发展路径思考[J].网络安全和信息化,2024,(02):6-8.

[4]罗鉴,雷波,郑秀丽,等.网络5.0应用场景分析与总体技术要求[J].信息通信技术与政策,2023,49(12):12-20.

[5]聂秀英,金伟,张杰.基于网络5.0的重叠网形态算力网络[J].信息通信技术与政策,2023,49(12):81-88.