新型智算中心组网方案研究

Research on Networking Scheme of New Intelligent Computing Center

张世华[1]，文湘江[2]，申　佳[1]，张　奎[1]，谭　蓓[1]，刘俊通[3]（1. 中讯邮电咨询设计院有限公司郑州分公司，河南 郑州 450000；2. 中国联合网络通信集团有限公司，北京 100033；3. 中国联通江西分公司，江西 南昌 330096）

Zhang Shihua[1], Wen Xiangjiang[2], Shen Jia[1], Zhang Kui[1], Tan Pei[1], Liu Juntong[3]（1. China Information Technology Designing & Consulting Institute Co., Ltd. Zhengzhou Branch, Zhengzhou 450000, China; 2. China United Network Communications Group Co., Ltd., Beijing 100033, China; 3. China Unicom Jiangxi Branch, Nanchang 330096, China）

摘　要：

当前算力需求爆发式增长，通用计算也朝着人工智能计算的方向演进，新型智算中心网络作为算力间数据交互的中心，成为影响算力性能发挥的关键。分析了智算中心对网络的需求，对网络协议、架构和运维管理等方面进行了深入研究，并结合业界发展情况，给出了智算中心组网方案的建议。

关键词：

智算中心；组网方案；无损网络；RDMA

doi：10.12045/j.issn.1007-3043.2024.06.005

文章编号：1007-3043（2024）06-0022-04

中图分类号：TN915

文献标识码：A

开放科学（资源服务）标识码（OSID）：

Abstract：

The current demand for computing power is exploding, and general computing is also evolving towards artificial intelligence. As the center for data exchange between computing power, the new network of intelligent computing centers has become a key factor affecting the performance of computing power. It analyzes the demands of intelligent computing centers on the network, conducts in-depth research on network protocols, architecture, and operation and maintenance management, and provides suggestions for the networking solution of intelligent computing centers based on industry development.

Keywords：

Intelligent computing center; Networking scheme; Lossless network; RDMA

引用格式：张世华，文湘江，申佳，等. 新型智算中心组网方案研究[J]. 邮电设计技术，2024（6）：22-25.

1 智算中心的重要性及组网挑战

算力是数字经济时代的核心生产力，是推动经济发展的新引擎。随着元宇宙、ChatGPT等业务的兴起，语言处理、机器视觉、自动驾驶等多个领域借助强大的数据计算能力，取得了长足的发展。相较于传统云计算、超算中心，智算中心更能满足日益丰富的人工智能算力需求，未来80%的场景所使用的算力资源都将由智算中心承载[1]。而参数量巨大的AI模型，对智能算力的需求飞速提升，根据IDC评估报告，2021年—2026年，中国智能算力规模年复合增长率达52.3%[2]。例如，目前L2级别的自动驾驶通常需要数百TOPS的算力，但要想真正实现L4/L5级别的自动驾驶，至少需要20 000+TOPS的算力。

受制于芯片材料、工艺、成本等因素，算力的增长速度逐渐放缓[3]，与算力需求存在极大差异，这也推动了芯片新技术以及异构算力的发展。以GPU、类脑芯片为代表的异构算力的崛起表明未来计算数据将在最合适的地方，以最合适的算力来处理。同时，当单台服务器的算力无法满足业务需求时，可使用分布式训练的智算集群，通过多台服务器以及算法优化的并行方式构建出一个计算能力和显存能力超大的集群，来应对大模型训练中算力和内存的瓶颈。大模型训练一般采用并行模式，连接集群的网络决定了智算节

收稿日期：2024-05-07

点设备间的通信效率,进而影响整个智算集群的算力性能和数据吞吐量,这对数据中心网络提出了新挑战,具体如下。

a)零丢包。智算集群对丢包十分敏感,如果网络故障不能被快速定位并传递到终端进行源端行为控制,轻则需要回退到上一个分布式训练的断点进行重训,重则可能要将整个任务从零开始重训。0.1%的丢包会使算力性能下降50%,1次训练中断会增加4 h的训练时长。因此,网络稳定性对分布式训练任务非常重要,也是当前数据中心网络的最大短板。

b)低时延。传统TCP/IP网络中,发送端给接收端发消息实际上是把发送端内存中的一段数据,通过数据中心网络传送到接收端的内存中。无论是发送端还是接收端,在报文传输过程中都需要调用CPU,复杂的报文处理流程使CPU显得力不从心,同时造成节点间通信时间变长。

c)大带宽。在并行计算模型中,单个计算节点完成计算任务后,需要快速地将计算结果同步给其他节点,以便进行下一轮计算;而在完成计算结果数据同步前,计算节点会一直处于等待状态。在大模型并行计算中,计算节点之间同步的数据量非常大,并且大部分是瞬时脉冲流量,如果网络带宽不足,数据传输就会变慢,进而影响训练效率。

2 智算中心组网技术研究

2.1 协议层——无损网络

远程直接内存访问(Remote Direct Memory Access,RDMA)可以使服务器直接高速读写其他服务器的内存数据,不需要经过操作系统/CPU/GPU的处理,成为解决智算中心组网问题的优选方案。RDMA主要流程是本端服务器RDMA网卡从内存中拷贝用户空间数据到内部存储空间,通过网卡自身进行报文封装后,使用物理链路发送到对端服务器,对端服务器RDMA网卡接收到报文后进行解封装,再将数据拷贝到内存的用户空间中,RDMA网络下服务器转发报文的路径如图1所示。RDMA的主要优势包括2点。

a)零拷贝,即不需要在内核空间和用户空间之间重复拷贝数据。

b)CPU/GPU卸载。由RDMA网卡实现报文封装和解析,CPU/GPU芯片无需参与内存读写、报文处理等工作,减少对芯片的开销。

随着AI大模型并行计算对高可靠、低时延、大带宽网络需求的增长,RDMA逐步在高性能数据中心中被推广应用。根据Uber发布的测试数据,在128块GPU和25GE网卡的配置环境下,进行VGG-16模型(网络深度为16的卷积神经网络)训练时,使用RDMA的处理性能比使用TCP高出30%,因此RDMA成为智算中心网络的最佳选择之一。RDMA的主要实现方案如下。

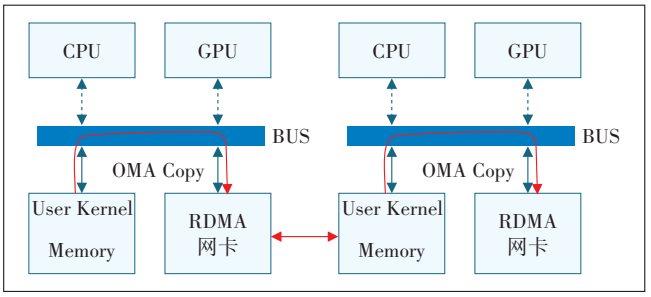

图1　RDMA网络下服务器转发报文的路径

a)InfiniBand(以下简称IB)协议。IB是一个完整的网络协议,它单独定义了1~4层的报文格式。基于credit信令机制,发送端在确认接收端有足够额度可以接收对应数量的报文后,才会进行报文发送,从根本上避免了报文在传输过程中从缓冲区溢出导致丢包的情况,实现了无损网络。IB在物理层定义了多种链路速度,例如1X、4X、12X,每种类型的链路使用四线串行差分连接,网络带宽升级到了NDR(单速率为100 Gbit/s)、XDR(单速率为200 Gbit/s)和GDR(单速率为400 Gbit/s)。此外,IB协议需使用专用的IB交换机、网卡和线缆。

b)RoCEv1。RDMA over Converged Ethernet是一种在以太网上进行RDMA的网络通信协议,而RoCEv1协议保留了IB协议的应用程序接口、传输层和网络层,将链路层和物理层替换为以太网协议。由于缺少IP路由功能,RoCEv1数据包只能在二层网络中传输。

c)RoCEv2。RoCEv2将IB的网络层、链路层和物理层替换为以太网协议,将RDMA应用数据封装到UDP报文中,再加上IP、以太网报文头,使报文可以在以太网中进行传输,并通过基于优先级的流量控制(Priority-based Flow Control,PFC)、显示拥塞通知(Explicit Congestion Notification,ECN)等流控机制,保证发送端和接收端速率匹配。RoCEv2通过普通的以太网交换机搭配支持RoCEv2的网卡实现,但对设备性能消耗较大。

d)iWARP。与RoCE协议继承IB不同,iWARP自

成一派，遵循IETF协议标准，上层包括RDMAP（为上层用户提供RDMA语义，支撑各类请求）、DDP（负责在传输层协议之上实现零拷贝）、MPA（完成与TCP适配工作，按照一定算法在TCP流中加入控制信息）。iWARP底层基于TCP/IP协议，但需要支持iWARP功能的特殊网卡设备。

目前业界比较常用的RDMA实现方案是IB和RoCEv2，而RoCEv1和iWARP存在一些技术缺陷，实际应用并不广泛。本文将重点介绍IB和RoCEv2方案。

2.2 网络架构

对于AI大模型的智算中心场景，需要特别关注数据中心网络的传输时延和可扩展性，传统的网络架构主要考虑其通用性，往往会牺牲部分性能。针对该问题，目前主流的网络架构有3种（见图2）。

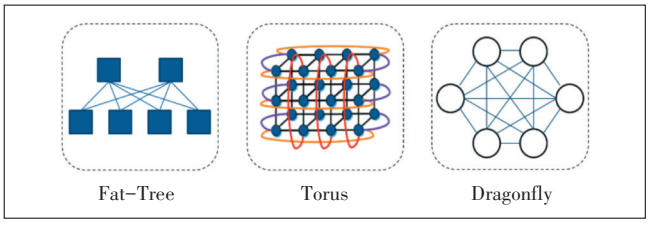

图2　智网中心网络架构示意

a）Fat-Tree。传统树形网络拓扑从叶子节点到根节点的带宽逐层收敛，整体成倒漏斗状，根节点网络带宽远小于各叶子带宽之和，容易成为转发性能的瓶颈，无法满足大规模计算的MapReduce和数据拷贝。而Fat-Tree拓扑的网络带宽是不收敛的，即每个节点的上行带宽和下行带宽相等，支持对接入带宽的线速转发，并且在横向扩展时支持增加链路带宽。Fat-Tree拓扑中所使用的网络设备均为端口能力相同的交换机，可有效降低网络建设成本。

b）Torus。Torus是一种环面拓扑，它将节点按照网格的方式排列，然后连接同行和同列的相邻节点，并连接同行和同列的最远端的2个节点，使得Torus拓扑中每行和每列都是一个环。Torus拓扑通过从二维扩展到三维、甚至更高维的方式增加新的接入节点，同时可以提高网络带宽，降低延迟。

c）Dragonfly。Dragonfly是一种分层的拓扑结构，包括Switch、Group和System 3层，其中Switch层包括一台交换机和与其相连的多个计算节点；Group层包含多个Switch，多个Switch间进行全连接；System层包含多个Group，多个Group间也进行全连接（拓扑中每个圆圈代表一个Group节点）。Dragonfly拓扑的主要优势是网络转发路径小，组网成本较低。

2.3 网络运维管理

由于RDMA的协议机制和通信方式与传统TCP/IP协议差异较大，智算中心高性能网络的运维管理方式也和IP网络存在很大差异，具体如下。

a）高精度的流量采集能力。AI大模型的流量呈现较强的突发性，常规的SNMP协议以30 s的采样周期收集流量数据，现已无法呈现网络的关键带宽指标。

b）细颗粒的流量统计能力。RDMA网络通过端口队列发送报文，因此需要将流量统计的维度从端口级别细化到队列级别。

c）自动化部署与检测能力。RDMA协议及其出色的拥塞控制机制使得网络配置复杂多样化，而智算中心的超大规模进一步增加了配置复杂度，需要自动化配置工具和可快速定位故障的检测工具来提升智算中心的运行效率。

d）流控指标采集和统计能力。若采用RoCEv2实现RDMA，需要使用PFC和ECN机制进行流量控制，运维管理系统相应地需要对PFC、ECN等关键指标进行采集和统计。

3 智算中心组网方案研究及建议

基于流控机制、网络时延、传输带宽等多个方面对IB方案和RoCEv2方案进行比较（见表1）。在性能、扩展性以及网络配置方面，IB网络占优，但成本较高，适合在高性能需求的场景中使用；而在成本、开放性、供应链方面，RoCEv2网络占优，建议在国产化算力资源池以及存储网络中使用，同时需增强RoCEv2网络的网络部署、调优及维护能力。

在组网架构方面，当网络规模较小（数千节点及以下）时，建议采用Fat-Tree。Fat-Tree拓扑具有网络直径短，端到端通信跳数少，建网成本低的优点，适用于中小规模智算中心。当网络达到一定规模后，例如上万节点时，建议采用Dragonfly和Torus。Dragonfly和Torus拓扑的建网成本更低，交换机端到端转发跳数也会明显减少，可提升网络整体吞吐和性能，适用于大规模、超大规模智算中心。

目前，OpenAI、微软、Meta、特斯拉等国外厂商选择使用IB方案组建智算中心，腾讯、阿里、字节跳动等国内厂商使用RoCEv2方案，配合自研交换机、DPU加速卡、协议优化和智能运维工具等手段来满足智算中

表1　InfiniBand和RoCEv2对比

RDMA方法	InfiniBand	RoCEv2
流控机制	基于Credit的流控机制	PFC/ECN，DCQCN（一种拥塞控制算法）等
网络时延	最低	较低
传输带宽	400 Gbit/s	可支持400 Gbit/s，主流采用200 Gbit/s
组网模式及扩展性	支持多种组网模式，最多支持单集群万卡GPU规模，且保证整体性能不下降	基本采用Fat-Tree架构，网络性能在千卡规模无明显降低
网络配置及维护	即插即用，可通过UFM实现零配置	手工配置，较复杂
成本	高	低
供应链情况	交换机、网卡、线缆、模块等均需使用IB专用产品，目前为英伟达独家供应，存在供应链风险	使用以太网交换机，主流设备厂商均可供应；网卡以英伟达的CX系列网卡为主
开放性及兼容生态	私有技术，专网专用	兼容IP网络，基于开放标准，可自主研发优化
优缺点	优点：链路层实现，性能最佳，部署简便，扩展性强；缺点：价格昂贵，被单一厂商垄断，存在供应链风险	优点：网络层实现，兼容IP网络，开放性强，自主可控，性价比较好；缺点：部署调试难度高，大规模组网性能较IB网络略差

心对高性能网络的需求，而百度、快手等厂商则选择在不同网络平面分别使用IB和RoCEv2方案。

基于以上分析，可根据方案将智算中心划成不同专区，各专区分别有5个网络平面。对于计算网，IB算力专区可采用IB NDR（400G）组网，RoCEv2算力专区采用RoCEv2（200G）组网；存储网可考虑共用，同时，考虑到训练场景对存储的即时访问带宽并没有计算网高，2个专区的存储资源池可以共用；管理/业务网仍然采用传统以太网络；带外管理网使用千兆网络连接所有硬件设备。智网中心组网建议如图3所示。

在组网架构方面，单台服务器最多支持配置8张GPU卡，对于IB网络，基于目前IB交换机能力，2层Fat-Tree架构最多支持2 048卡，如规划超出2 048卡的规模，建议使用3层架构或选用扩展性更强的Dragonfly、Torus拓扑。而RoCEv2基本采用Fat-Tree架构，可根据组网规模选择合适的交换机设备。

以往数据中心大多是计算、存储、网络资源分别由不同负责人进行维护管理，而在智算中心场景，算力的调度、性能优化与数据中心网络息息相关，其建

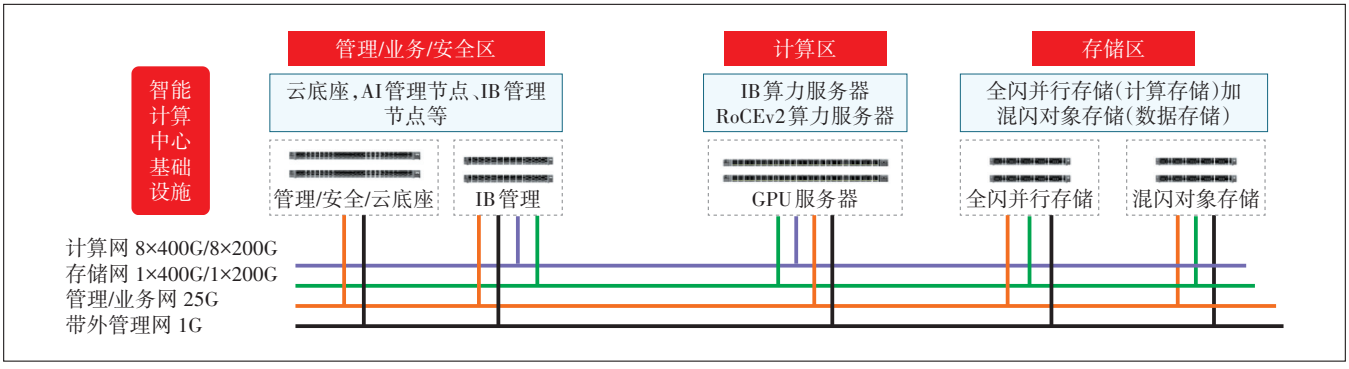

图3　智网中心组网建议

设运维过程需同步研究端到端的编排管理能力，实现算力与网络的协同优化管理。

4 总结与展望

本文对智算中心组网需求与技术演进进行了相关研究分析，以期抛砖引玉，得到同行专家的参与和讨论，共同推动网络关键技术的成熟与落地，打造大规模、低时延、高性能、大带宽以及智能化的智算中心网络。

参考文献：

[1] 国家信息中心. 智能计算中心创新发展指南[EB/OL]. [2024-01-30]. http://scdrc.sic.gov.cn/SmarterCity/445/449/0113/10715.pdf

[2] IDC，浪潮信息. 2022-2023中国人工智能计算力发展评估报告[EB/OL]. [2024-01-30]. https://www.doc88.com/p-99229765957589.html.

[3] 郝俊慧. 摩尔定律失效后，未来看“算力三定律”[N]. IT时报，2022-07-22(6).

作者简介：
张世华，工程师，硕士，主要从事核心网、通信云咨询、规划和设计工作；文湘江，高级工程师，硕士，主要从事通信云架构设计、技术选型等工作；申佳，助理工程师，学士，主要从事核心网、通信云咨询、规划和设计工作；张奎，高级工程师，硕士，主要从事核心网、通信云咨询、规划和设计工作；谭蓓，高级工程师，学士，主要从事核心网、通信云咨询、规划和设计工作；刘俊通，毕业于电子科技大学，主要从事通信网络的规划与建设工作。