

浪潮网络助力打造面向AIGC的“三超”网络

北京2023年8月18日 /美通社/ -- 从内容生成、游戏开发、到自动化助手、机器人控制……AIGC这项“黑科技”，正在更多领域发挥着作用，并逐渐向行业渗透。例如，在数字政府场景，通过融合AIGC和数字人技术，可以提供个性化的政务服务和咨询，优化政府与公民的互动方式，提升政务服务的质量和效率。目前，AIGC主要聚焦在大模型训练、MaaS模型服务、AIGC推理三大应用场景，其中大模型训练是各新兴企业竞相角逐的主要场景。

“没有好网络，别玩大模型。”要建设大规模训练模型集群，除了需要GPU服务器、网卡等基础组件，也需要解决网络搭建的问题。网络对于大模型集群的算力释放和可靠运行至关重要，如何构建符合大模型集群计算要求的网络系统，是推进AIGC发展的关键之一。

AIGC大模型训练 对网络系统的“三超”要求

在大模型训练过程中，有三种流量模型，分别是张量并行、流水线并行、数据并行。人们熟知的ChatGPT3，采用128台A100服务器，共计1024个A100卡训练，这样单服务器节点需要4个100G网络通道；而ChatGPT4、ChatGPT5等其它大模型，对于网络的需求会更高。浪潮网络认为大模型训练对于网络的要求可用“三超”网络来概括，即：超大规模、超高带宽、超强可靠，以保障网络稳定、可靠运行，为大模型训练提供强有力的支持。

而要解决“三超”网络的挑战，就需要着重思考如何建设符合大规模训练的组网方案。从组网架构上看，当前AIGC组网一般多采用胖树架构，具有高带宽、低延迟的特性，以及较好的可拓展性。而在组网协议上，当前业界主流的是基于IB、及RoCE两种无损网络技术，两种技术都可以很好的满足大规模训练高带宽、低延迟的要求。IB的延迟足够低，而RoCE在开放性、性价比、及易维护性几方面更胜一筹。

浪潮网络融合趋势与驱动 打造基于RoCE的智能无损网络解决方案

浪潮网络作为云边协同智慧网络引领者，密切关注市场发展及变化，推出了基于RoCE的智能无损网络解决方案，助力AIGC“三超”网络的打造，其具备如下优势：

一是多协议、多场景的融合。在大规模集群中，往往存在通用计算集群、AI/HPC集群、存储等多种场景，传统方案是部署以太网、IB、FC等多套网络及多种协议，各协议之间互不兼容，大大增加了管理和维护的难度。而浪潮网络基于RoCE的智能无损网络解决方案，可以适配通用计算、AI/HPC、存储等多种场景，并实现以太/IB/FC三网融合。这样从维护多张网络到维护一张网络，大大降低了整体建设和维护成本。

二是智能弹性、动态调整。在大规模集群训练中，要求整个集群可以快速部署与交付，在节约训练时间的同时，尽可能减少宕机等故障的发生。在浪潮网络基于RoCE的智能无损网络解决方案中，通过数字化网络引擎IDE可以实现集群网络的自动化部署，加速业务上线。并实时监控设备与链路的负载和健康状态，如CRC错包，端口带宽百分比、队列缓存，CNP及Pause反压帧等，完成故障的快速定位及智能分析，实现基于业务的网络跟踪。此外，还可以提供北向标准API接口，能够与上层计算平台进行对接，实现算网联动，更好的释放集群算力。

此前，浪潮网络基于RoCE的智能无损网络解决方案，已在教科研客户项目中得以应用，方案可充分满足通用计算集群、GPU加速集群、异构计算集群、分布式存储集群、全闪存存储集群等多种场景，对于网络的高带宽、低延迟连接需求，帮助客户构建满足AIGC发展的整体网络架构。

未来，浪潮网络将持续优化基于RoCE的智能无损网络产品方案能力，同时深入研究基于UEC的网络并创新引领支持UEC的产品，帮助客户成功。

原文地址：<http://www.china-nengyuan.com/news/199297.html>