DRIVENETS

# 利用 Distributed Disaggregated Chassis 实现后端人工智能组网整体结构

白皮书

人工智能应用日益普及，推动了专门用于人工智能计算的大型高性能计算（HPC）集群的构建和使用。区别于其他高性能计算应用，这些应用大多数由生成式人工智能应用驱动，主要由超大规模数据中心内部使用或作为一种基于云的服务提供给企业客户。在人工智能集群的组网（或整体结构）部分中，后端网络是一种关键元件，影响着集群的整体性能及其计算资源的有效利用。

# 人工智能组网

人工智能工作负载是在服务器阵列上运行的应用，除了CPU（可以单独充当计算引擎）和部分存储容量（通常是高速SSD），这些服务器通常还托管着专用计算引擎（例如GPGPU/FPGA加速器）。在这些服务器上运行的人工智能应用，并非在特定服务器上运行，而是同时在多个服务器上运行。无论是几台服务器还是数千台机器，全都能够同步运行，运行分布在它们之间的同一个应用。

这些计算机器之间的互连被称为后端互连。相对而言，前端互连则是一种单独的组网基础设施，将这些机器连接到公共互联网以进行查询和数据传输训练。后端网络必须在运行同一应用的所有计算机之间启用任意对任意连接模式，同时还必须根据运行应用的类型和阶段，满足所需的不同流量模式。

人工智能的后端互连解决方案不同于为住宅联网或移动网络而构建的网络，也不同于为服务器阵列构建的网络，这种阵列旨在回答多个用户的查询，例如典型的数据中心。

# 人工智能集群属性

计算能力的发展推动了计算功能创建能力的产生，而这种功能能够应用于大型数据集。人工智能领域的重点是识别此类数据集及其产生的结果，以尽可能多的结论点来训练计算功能。然后，计算功能需要识别这些数据集中的模式，以预测尚未遇到的新数据集的结果。

举例而言，人工智能组网的部分包括：

- **并行计算：**人工智能工作负载是一种统一基础架构，由多台机器构成，同时运行相同的应用和相同的计算任务。
- **规模：**此类任务可以涉及数以千计的计算引擎（例如GPGPU、CPU、FPGA）。
- **作业类型：**不同的任务在规模、运行持续时间、涉及数据集的大小和数量、要生成的答案类型等方面有所不同。用于编码应用的语言以及运行应用的硬件类型也会有所不同，因此，在为运行人工智能工作负载而构建的网络内，流量模式也在不断变化。
- **带宽：**大型数据集需要高带宽流量进出服务器以供应用程序使用。在现代部署中，人工智能或其他高性能计算功能的每个计算引擎的接口速度都能达到400Gbps。
- **延迟及抖动：**某些人工智能工作负载能够产生用户预期的响应。在这种情况下，作业完成时间（JCT）是用户体验的关键因素，因而延迟成为一个重要因素。不过，鉴于这类并行工作负载在多台机器上运行，因此延迟取决于响应最慢的机器。也就是说，虽然延迟很重要，但抖动（或者说延迟变化）实际上也是实现所需JCT的一个影响因素。
- **无损行为：**继上一点之后，迟到的响应会延迟整个应用程序。在传统数据中心中，丢失的消息会导致重新传输（通常甚至不会被注意到），而在人工智能工作负载中，丢失的消息意味着整个计算要么出错，要么停止。正是由于这个原因，人工智能运行的网络需要无损行为。IP网络本质上就是有损的，因此要使IP网络表现得无损，就需要应用某些附加功能，本文稍后将对此进行探讨。

从这些属性得出的一个重要结论是，用于运行人工智能工作负载的网络与传统数据中心网络的不同之处在于，它需要同步运行。

# 人工智能的行业解决方案及其缺点

人工智能后端网络有几个著名的行业解决方案。接下来的部分描述了“同步解决方案”目前可用的几大选择，包括其主要优缺点。

## 基于机箱的解决方案

源自电信网络的基于机箱的路由器被构建为黑盒，所有内部连接都是对用户隐藏的。通常情况下，机箱所采用的架构由以Clos-3拓扑形式连接的线路卡和整体结构卡（fabric card）组成。因此，机箱行为具有可预测性和可靠性。它实际上是一种包裹在金属面板中的无损整体结构，只有网络接口面向用户。

在这种情况下，机箱的缺点在于尺寸。虽然精心编排的整体结构非常适合人工智能工作负载的网络需求，但它只有有限的数百个端口用于连接服务器，因此这种解决方案仅适合非常小型的部署。如果机箱的使用规模大于单个机箱的端口总数，则需要Clos拓扑（其实是非平衡的Clos-8拓扑），但会破坏这种模型的整体结构行为。

## 独立以太网解决方案

这种解决方案源自数据中心组网。虽然数据中心解决方案速度快并且能承载高带宽流量，但是它们基于的是以多层拓扑（通常为 Clos-5或Clos-7）连接的独立单芯片设备。只要流量仅在该拓扑中的同一设备内运行，流量的行为就会接近统一。

由于每个此类设备的平均接口数量仅限于物理上位于同一个机架中的服务器数量，因此这种单一的架顶式（ToR）设备无法满足大型基础设施的要求。将网络扩展到更高的网络层还意味着流量模式会开始改变，而应用运行完成时间也会受到影响。此外，附加机制会安装到网络上，将有损网络转变为无损网络。

人工智能工作负载流量模式的另一个属性是从数据包头部（packet header）角度来看的流量一致性。也就是说，无论网络拥塞情况如何，同一流量的不同数据包都将被数据平面识别为相同的流量，并在完全相同的路径中传输。这样会导致Clos拓扑的某些部分无法得到充分利用，而其他部分则又可能出现过载甚至是一定程度的流量丢失。

## 专有锁定解决方案

这个领域的其他解决方案是作为特定服务器阵列的专用互连来实施的。这在计算工作负载繁重的科学领域更为常见，例如研究实验室、国家研究所和大学。专有解决方案迫使客户只能选择一个互连服务提供商，为整个服务器阵列提供整套服务，从服务器本身一直到阵列中的所有其他服务器。

这个行业的本质是通过一次性预算分配来构建一台“超级计算机”。也就是说，最终构建的计算阵列预计不会再进一步扩展，而只会被更新的模型取代或超越。这就让选择专有互连解决方案的供应商锁定变得更容易忍受。然而，这在人工智能实施中则会成为问题，因为基础设施具有动态性和在线性。

从好的方面来说，此类解决方案性能表现非常出色，常见于世界最强超级计算机排行榜顶端，这些超级计算机通常使用HPE（Slingshot）、英特尔（Omni-Path）、英伟达（InfiniBand）等提供商的解决方案。

# DDC是一种人工智能组网分布式同步整体结构（DSF）

正如接下来DDC部分所说，Distributed Disaggregated Chassis (DDC) 是机箱的分布式结构。如前所述，在构建服务于人工智能工作负载的网络方面，单体机箱的主要缺点在于其尺寸，而DDC消除了机箱金属外壳的限制，从而能够解决这个问题。

从电信角度（即DDC的定义来源）来看，DDC是机箱的一种分布。从数据中心组网的角度来看，DDC是一种分布式同步整体结构（DSF）解决方案，具有人工智能工作负载后端互连所需的属性。

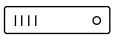
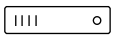
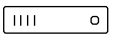
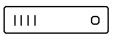
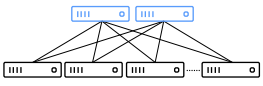
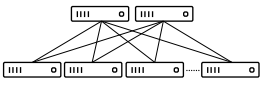
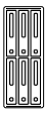
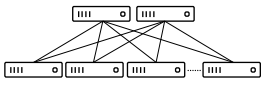
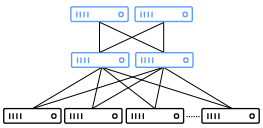
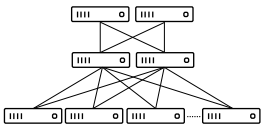
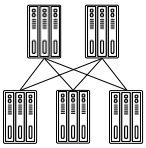
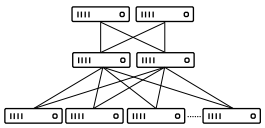
这些所需的属性包括：

**标准**——DSF的所有外部接口都是标准IP接口，可以通过标准以太网接口连接到任何服务器类型，无论网络接口卡（NIC）来自哪个供应商。

**无损**——就像黑盒机箱一样，DSF能够以线速率处理任意一组接口之间的流量，并且不会丢失数据包。

**规模**——DSF的构建方式与Clos-3类似，不受金属外壳的限制，因此能够比单体机箱具有更高的规模。不过，可以在分布式拓扑中内置额外的整体结构层，将Clos-3拓扑转换为统一的Clos-5拓扑。如此，与Clos-3拓扑相比，DSF的最大容量能够扩大一个数量级。

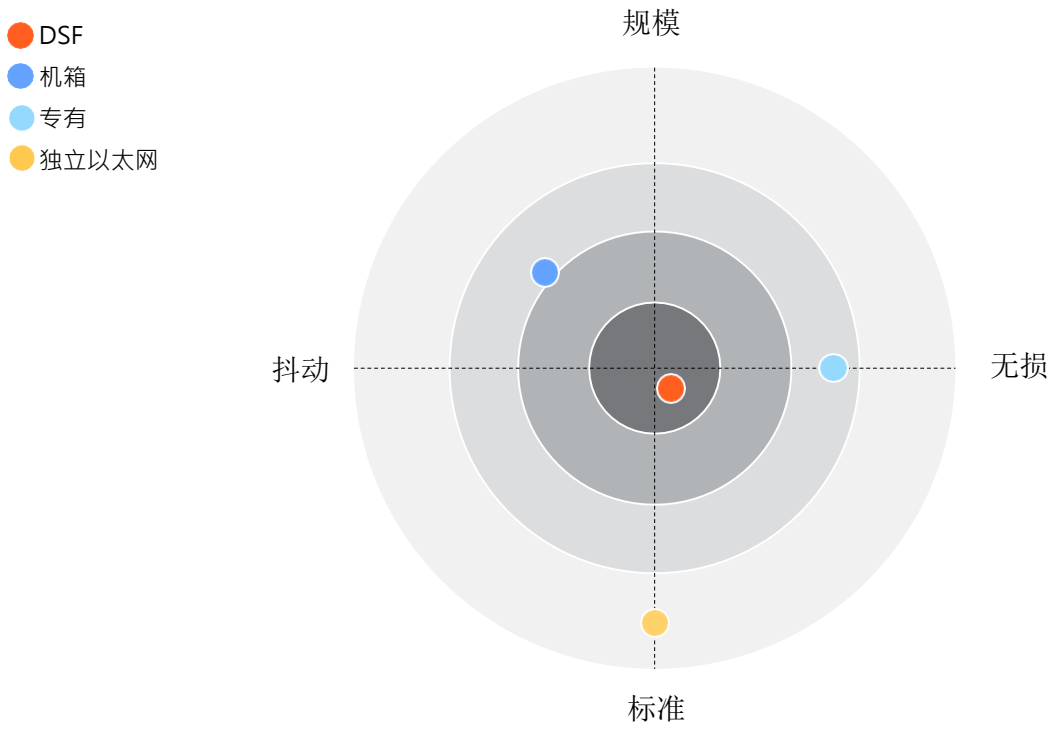
*图1：DSF/ 以太网/ 机箱/ 专有网络的规模对比*

| | DSF | 以太网 | 机箱 | 专有 |
|---|---|---|---|---|
| 独立 |  |  |  |  |
| 网络跳数 / 标准接口 | **1/数十** | **1/数十** | **1/数十** | **1/数十** |
| 第2层 |  |  |  |  |
| 网络跳数 / 标准接口 | **1/数百** | **3/数百** | **1/数百** | **0/0** |
| 第3层 |  |  |  |  |
| 网络跳数 / 标准接口 | **1/数千** | **5/数千** | **3/数千** | **0/0** |

**低抖动**——对于大多数机箱而言，任意两个线路卡之间的流量会均匀分布在所有整体结构卡之间。同样地，**DSF通**过实施调度机制来平衡所有DCF设备之间来自分布式机箱数据包转发器（DCP）设备的流量。这是通过使用在数据平面（ASIC）级别上托管的虚拟输出队列（VOQ）**来完成的**，让所有整体结构设备尽可能实现接近**100%的利用率**，**同**时不让任何设备出现拥塞情况。这种平衡的实施对数据包头部不敏感，能够避免同构和异构流量模式的影响——**即造成穿越DSF的数据包的不同延**迟。

*图2：4种解决方案类型及其人工智能工作负载相关属性的雷达图*

注：越靠近雷达中心，分数越高。




# Distributed Disaggregated Chassis (DDC)

DDC最初是由AT&T定义的架构，于2019 年9月作为开放架构贡献给开放计算项目（OCP）。DDC定义了用作运营商级网络路由器的网络元件的组件和内部连接。区别于基于单体机箱的路由器，DDC将路由器的每个组件定义为独立设备。

主要定义如下：

- 机箱的线路卡被定义为分布式机箱数据包转发器（DCP）。
- 机箱的整体结构卡被定义为分布式机整体箱结构（DCF）。
- 机箱的路由堆栈被定义为分布式机箱控制器（DCC）。
- 机箱的管理卡被定义为分布式机箱管理器（DCM）。

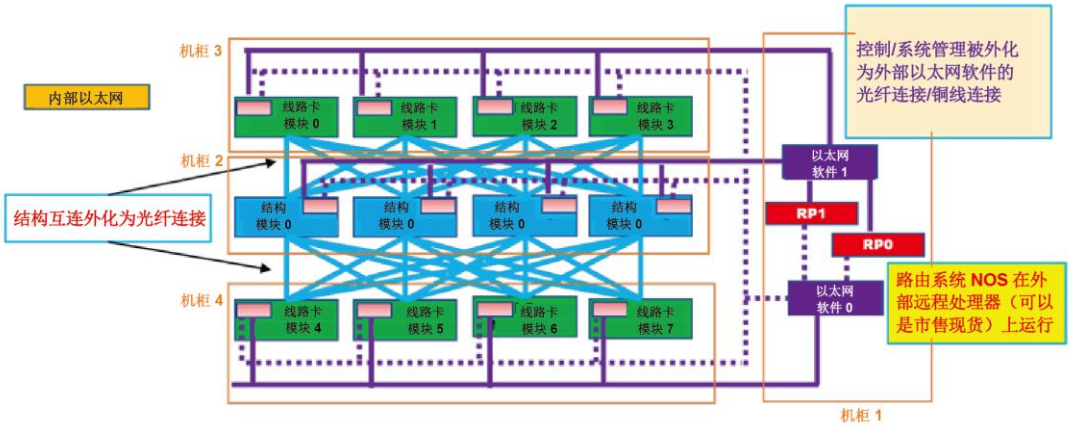
所有设备都通过标准10GbE接口物理连接到DCM，以建立控制和管理平面。所有DCP通过Clos-3拓扑中的400G整体结构接口连接到所有DCF，从而在DDC中的所有网络端口之间建立调度的、无阻塞的数据平面。DCP能够同时托管用于连接DCF的整体结构端口和用于通过标准以太网/IP协议连接到其他网络设备的网络端口。DCF不能托管任何网络端口。DCC实际上是一种服务器，用于运行定义DDC功能的主要基础操作系统。

DDC的优势包括:

- **高容量：**由于没有必须将所有这些组件容纳在一台机器中的金属机箱外壳限制，因此可以实现更高的容量。如此便能构建更广泛的Clos-3拓扑，扩展到单个机架的边界之外，从而使数千个接口可以在同一网络元件（路由器）上共存。
- **开放：**DDC是一种开放标准定义，使得多个供应商实施其组件成为可能。因此，运营商（电信公司）可以更轻松地建立多源采购方式，在扩展网络的同时也能控制自身的网络成本和供应链。
- **灵活性：**在这种分布式组件阵列中，每个组件都既可以独立存在，也可以充当DDC的一部分。因此，相较于在基于机箱的路由器上运行的服务，在基于DDC的路由器上运行的服务具有极高的灵活性。

AT&T宣布使用DDC来运行其核心MPLS、边缘和结对IP网络，同时，全球运营商也正在使用DDC实现此类功能。

*图3：DDC的高级连接结构*




注：线路卡被定义为上面的DCP，整体结构模块被定义为上面的DCF，远程处理器被定义为上面的DCC，而以太网软件被定义为上面的DCM。

来源：OCP DDC规格

DDC实现了解耦化（disaggregation）概念。控制平面与数据平面的解耦，让用户能够从不同供应商采购软件和硬件，然后在部署时再将它们组装为统一的网络元件。虽然这个概念相当新颖，但在作为DDC的一部分使用之前已经有过多次成功的部署案例。

# 让最佳性能的人工智能后端网络成为现实

人工智能后端网络带来了独特的挑战，能够影响计算性能和计算资源利用率。基于机箱的独立以太网和专有锁定解决方案都无法开放有效地解决这些挑战。基于DDC的解决方案在电信领域已被证明具有稳健性和可扩展性的解决方案，它的引入让DSF实现最佳性能的人工智能后端网络的愿景成为现实。

DRIVENETS

DriveNets是云原生网络软件和网络分解解决方案领域的领导者。DriveNets成立于2015年，总部位于以色列，为服务提供商和云提供商提供全新的网络构建方式，能够通过改变技术和经济模式来大幅提高盈利能力。DriveNets推出解决方案Network Cloud（网络云），能够将云的架构模型提升为电信级网络。网络云是一款云原生软件，可在标准白盒的共享物理基础设施上运行，从根本上简化网络运营，以更低的成本实现电信规模的性能和灵活性。欲了解更多信息，请访问 www.drivenets.com