

算力包，为算力普惠提速

在算力大发展的今天，算力行业亟须普惠性、低价格、低门槛的公共基础设施服务。九章云极以“1度算力”概念推出“算力包”服务，降低了门槛，实现了普惠，满足了长尾需求，实现了成本收入比的最优化。

文 | 方磊 九章云极 DataCanvas 董事长

算力作为新时代的基础设施之一，正逐渐展现出其重要性和热度。回顾历史，基础设施从道路、铁路到电力、互联网，均经历了从物质能量运输到计算智能获取的演变。这些基础设施的发展遵循着投资收益周期，转折点在于转化为公共服务，实现商业价值的显著提升。算力领域亦遵循这一逻辑，今年可视为算力向公共服务转变的元年。这一转变背后隐含的核心逻辑是：作为普惠性质的基础设施，算力是推动应用创新的强大动力，其转化为公共服务正是商业化潜力爆发或进入新阶段的显著标志。

在 AI 时代，大模型的飞速发展天文数字般的算力投入已成为众人瞩目的

焦点。推动应用创新的真正动力是什么呢？显然，它并非凭空想象而来。我们都明白，优秀的商业应用往往源自需求的反向推动。在这一过程中，各行各业的专业人士通过实践和商业历练，最终挖掘出真实的需求。同时，在这一过程中，一个至关重要的角色便是能够提供普惠公共基础服务的基础设施。

在算力时代，我们将这种基础设施称为“普惠算力”，它是创新应用的强大引擎。正如电力的普及、道路的畅通、高铁的便捷，当这些公共服务成本降低时，它们便成为应用创新的最大推动力，激发人们的智慧与行业的创造力，从而孕育出我们梦寐以求的杀手级应用，以



赛迪网官方微信



数字经济官方微信

及各行各业中能够提升效率的多样化落地场景。这是从以往基础设施商业化进程中总结出的规律，我们相信这一规律同样适用于算力这一新型基础设施。

普惠算力是应用创新的核心驱动力。基于这一逻辑，当前我们的行业是怎么样的？

算力行业的供需特点鲜明：供给稳定而需求灵活多变。如同波音公司飞机制造，高性能硬件如 GPU 的生产能力相对固定，但用户算力需求却灵活多样，既体现在数量上也体现在质量上。为解决这一矛盾，中间层运营机构应运而生，它们如同航空公司，通过租赁、购买等方式拥有硬件和软件系统，提供定制化的算力服务，满足用户灵活多变的需求。

算力行业现状呈现结构性失衡。一方面，科技巨头竞相投入算力资源，AI 应用热度高涨，显示整体算力资源不足；另一方面，某些特定品类算力可能过剩，原因在于定制化服务市场规模有限，如同私人飞机销量有限。当定制化需求过剩或算力服务提供商无法满足更广泛需求时，结构失衡便会产生。

因此，算力行业亟须普惠性、低价格、低门槛的公共基础设施服务，以满足旺盛的算力需求。这类服务能让更多人轻松获取算力资源，推动算力服务的普惠化、大众化。当前，这类服务供给不足，亟须行业共同努力，促进算力资源的优化配置和均衡发展。

破局之道

当前，裸金属销售模式占据主导地

位，并伴随着一定的市场竞争。然而，我们渴望的是更加便捷、灵活的算力消费方式，类似于购买一张算力的“机票”，但在市场中这样的选择却十分有限。那么，如何解决这一问题呢？

第一，依靠技术驱动。依靠技术来改变一些事情，这是整个行业底层的推动力。比如依靠九章云极推出的 DataCanvas Alaya NeW 智算操作系统所支持的 Serverless 技术，可以提升 GPU 的利用率，这可能是其中的一种方式，把算力变得更灵活、更好用。

Serverless 并非复杂的技术名词，其本质在于按需分配资源。在 CPU 的世界里把资源分片地卖给别人，核心技术叫做“虚拟机”。但与 CPU 世界的虚拟机不同，在 GPU 世界中，我们更关注工作载荷的具体需求，通过 Serverless 技术，我们能够为每个任务寻求最佳的调度和运行策略，确保在给定任务上获得最优结果。这种技术推动使算力应用更加灵活多变。

第二，依靠专业商业创新。正如航空公司不仅依赖于技术，更在于商业运营的创新，我们也致力于提供规模化、标准化的算力产品，让各行各业能够轻松消费算力。这其中，精确度量是关键。在此，我们引入了“1 度算力”概念，通过标准化的度量方式，将庞大的算力基础设施变成可分割、可分配的普惠产品。

第三，“1 度算力”也正是破局之道。正如电的单位“度”一样，“1 度算力”也为我们提供了一个清晰的计量标准。我们定义了“312 TFLOPS（每秒浮点运

算次数)乘以1小时等于1度算力”，使算力成为可度量、可交易的商品。

有了“1度算力”的概念后，我们能够精确地将各种异构计算能力转化为可分割的算力包，就像电卡或流量包一样，用户可以根据自己的需求购买不同额度的算力。这些算力包只有在用户使用时才会被扣除和计费，大大降低了用户接触和使用先进算力基础设施的门槛。

全新产品“算力包”

全新算力产品“算力包”，是从用户角度出发，以创新性的“按需购买、即买即用”的服务模式，以“度”为算力计量单位，为广大用户提供高质量、高性能、灵活“度”数的算力一体化资源包。算力包在三个关键维度上展现出其深远的价值与意义：

从社会层面来看，算力包成功地将庞大的基础设施转化为一种公共服务。我们不再仅仅将算力定向提供给单一客户并进行定制化服务，而是将其打包成算力包，供众多中长尾客户自主选择购买，如100度、1000度等，这就极大地提升了基础设施的利用效率，对社会整体而言无疑是一大利好。

对于运营者来说，算力包带来了算力经济的高附加值。传统上，将智算中心整体打包给大型企业时，主要附加值体现在建设投资上。但当我们以算力包的形式提供给各行各业时，其中的软件工具、行业应用、应用模板等便成为了新的高附加值来源，为运营者创造了更多的收入来源。这是因为行业客户在利用

算力实现AI落地时，往往需要低门槛的入门、丰富的范例及多样化的帮助，而算力包正好满足了这些需求。

对于使用者而言，算力包降低了门槛，实现了普惠。一方面，价格大幅降低，如同我们不再需要花费高昂的价格租用私人飞机，而只需购买一张机票即可轻松抵达目的地一样，算力包也让我们能够以更低廉的价格享受到强大的算力。另一方面，技术门槛也显著降低，因为算力包中包含了众多高附加值的服务，使得大模型的开发和应用变得更加容易。

对于行业客户来说，算力包满足了长尾需求，实现了成本收入比的最优化。同时，通过低门槛获取AI工具来开发应用，将进一步促进AI应用的繁荣与发展。

因此，九章云极DataCanvas将秉持观察到的算力基础设施规律，全力以赴推进建设“按需消费，按度计量”的算力生态。力争在三年内纳管10万P算力，并以算力包的形式提供给各行各业及所有需要低门槛普惠算力的用户。他们不仅可以享受到这些算力，还能使用当前最先进的AI工具来充分发挥其效用，真正实现好用、可用且经济的算力。

我们眼中的未来

在这个量级上看整个行业的发展前景，我们不禁要问：推动行业前行的核心动力何在？对此，我将其精髓概括为以下三点：

第一，本质是需求推动算力商业产品的构建。

算力包正是这一理念的初步实践，它



来源：九章云极 DataCanvas

图 1 普惠算力是应用创新的推动力

并非终点，而是起点。我们基于用户对普惠算力的迫切需求，成功推动了算力包的诞生与发展，这一创新无疑将广泛惠及客户与行业。随着需求的不断演变，未来必将催生出更多新型算力产品，不仅九章云极 DataCanvas，整个行业都将迎来这一变革，为客户提供更多价值。

第二，商业产品构建的同时，我们以效率倒逼算力核心技术的突破。

以裸金属租用为例，若 GPU 使用率仅为 20%，则大部分资源被闲置，如同租车却未使用，造成资源浪费。算力包则如同按需付费的打车服务，更加经济高效。当我们将公共基础设施以算力包的形式提供给客户，满足其碎片化需求时，基础设施的运营效率便获得了提升。技术的提升，哪怕只是 5% 或 10% 的效率增长，都能带来显著的社会效益与经济效益。因此，我们致力于通过效率提升来推动算力核心技术的突破，如采用加速技术提升推理效率，以及应对 OpenAI o1 模型发布后推理预算的大幅增加，这

些都要求我们不断优化技术，以满足日益增长的算力需求。

第三，普惠算力将激发更多创新业务场景的出现。正如我们观察到的，只有普惠的服务与算力，才能促进应用的广泛普及与生态的繁荣发展。

我们坚信，自 2024 年踏入算力领域以来，通过引入“1 度算力”的度量标准，我们正步入一个可度量、普惠化的新时代。九章云极 DataCanvas 愿成为这一变革的引领者，与行业伙伴携手共进，共同构建这一繁荣生态。

责任编辑：杜玢翰 dubh@staff.ccidnet.com