算力投入巨大

高端AI服务器市场需求激增

“过去一台AI服务器装四张显卡，现在客户要的都是能装八张甚至十张显卡的高端设备。”12月18日，在2023 AI Tech Day暨首届人工智能生态发展峰会现场，服务器厂商正展示全新升级的AI大模型服务器。中国证券报记者在现场了解到，AI大模型迭代速度越来越快，厂商对智能算力投入大幅增加，支持存储和训练的高端AI服务器的需求激增。不少AI服务器厂商今年订单都集中在高端AI服务器上。

业内人士表示，随着AI大模型加速迭代，智能算力已成为稀缺资源。未来较长一段时间内，我国AI服务器市场维持供不应求状态，国产AI芯片市场规模增长迎来关键窗口期。

● 本报记者 彭思雨

2023 AI Tech Day暨首届人工智能生态发展峰会现场 本报记者 彭思雨 摄

高端AI服务器抢手

当前，大模型厂商产品迭代如火如荼。据百川智能创始人、CEO王小川介绍，目前百川大模型保持每个月一次模型数据的迭代。

AI大模型正带动AI服务器算力需求持续扩张。安擎计算机相关负责人告诉记者，2023年，由于行业投资趋于谨慎，判别式AI应用场景收缩，AI服务器市场整体销售台数同比有所下降，但订单金额同比将会呈现增长态势。这是由于AI大模型算力需求集中爆发，成为智能算力的最大需求方。

“AI大模型厂商都在加快迭代速度，客户根本等不起你去建机房，而是希望拎包入住。”鸿博股份副总裁、英博数科CEO周韡韡告诉记者，“我们现在都是找一些已经建好的机房，然后把设备放进去，快速组网，调试好后交付给客户。AI大模型的发展让人看到的不仅是效率提升，似乎全行业迭代周期都在缩短。”

高端AI服务器设备十分紧俏。“大模型训练所需数据激增，AI大模型厂商需要的是能够支持存储和训练的高端AI服务器。因此各服务器厂商目前都在升级芯片规格、扩大卡组数量，向高端AI服务器方向升级。”上述安擎计算机相关负责人称。

我国智能算力资源稀缺。艾瑞咨询发布的《2023年中国智能计算中心行业发展白皮书》显示，2022年，中国智能算力规模占全部算力的比重为22%；从服务器结构来看，2022年，我国通用服务器占服务器总量比重为93.2%，而AI服务器仅占服务器总量的6.8%。多位业内人士表示，未来较长一段时间内，我国AI服务器市场维持供不应求状态。

算力产业链迎利好

随着AI大模型算力需求强劲，今年以来，国内算力产业链公司接连斩获新订单。

12月4日，鸿博股份公告称，子公司英博数科向百川智能提供一定规模的英伟达智算服务器，及其所有的算力资源以及配套软件应用和技术服务，涉及交易总金额预计达313.82亿元。据周韡韡介绍，截至2023年12月，英博数科累计签约额超18亿元。

公开资料显示，神州数码在10月29日至11月19日期间共签订四笔销售神州鲲泰品牌昇腾AI服务器订单，合同总金额达6.24亿元。

青云科技10月15日公告称，公司向高新兴讯美科技股份有限公司采购6.84亿元GPU服务器（含配套产品），并与客户签订6.90亿元销售合同，向其销售上述GPU服务器（含配套产品）。

在AI大模型浪潮的带动下，人工智能芯片、服务器、数据中心市场规模将显著提升。IDC预计，2023年中国人工智能芯片出货量将达到133.5万片，同比增长22.5%。

人工智能服务器方面，IDC预计，2023年中国人工智能服务器市场规模将达到91亿美元，同比增长82.5%；2027年将达到134亿美元，年均复合增长率为21.8%。

智算中心建设步伐加快。据不完全统计，截至2023年8月，全国已有超过30个城市建设智算中心。

中信证券表示，AI的持续发展拉动智能算力需求和建设水平提升，为光模块、服务器等领域带来持续增长机会。

做好生态建设

AI大模型时代，进口品牌GPU等高端芯片供应面临周期波动挑战，为国产AI芯片加速推向市场打开关键窗口。不过，AI大模型的训练、推理和海量数据存储无一不需要高性能算力支撑，也对国产算力底层基础设施能力和生态建设提出考验。

艾瑞咨询产业数字化研究院负责人徐樊磊表示，在AI大模型爆发之前，国内的智能算力资源主要用于推理端。AI大模型趋势到来，使国内厂商开始推出训练端的算力硬件产品和服务，但目前相关产品能力比起全球领先算力能力仍有较大差距。

业内人士表示，在计算层面，由于芯片厂商在开发过程中使用的技术路线不同，导致芯片适配服务器等设备的开发周期普遍很长。在训练层面，单芯片算力有限，而大模型训练需要大规模的算力集群，需要算力系统具有灵活的算力扩展能力。在存储层面，多模态大模型的训练和推理对存储提出了更高要求。

近年来，我国AI算力市场高度依赖英伟达GPU硬件和相应的软件生态。周韡韡坦言，全球90%的AI工程师都在使用英伟达GPU配套的CUDA软件生态，这和英伟达GPU在全球的垄断地位完全匹配。“一些国产GPU计算性能并非绝对不够，而是若要把国产GPU和基于CUDA架构开发的设备进行适配，需要在调试和优化上花费大量精力，导致用户使用算力的效率下降。”

然而，随着英伟达芯片进口难度不断提升，国产AI芯片自主创新任重道远。

记者梳理发现，浪潮信息、海光信息、希姆计算、中科通量、瀚博半导体、墨芯人工智能、摩尔线程、天数智芯、寒武纪、燧原科技等芯片公司推出了应用于不同场景的AI推理和训练任务的芯片加速卡，涉及CPU、GPU、RISC-V等不同设计架构。

海光信息表示，海光DCU兼容CUDA生态，对文心一言等大多数国内外主流大模型适配良好。依托DCU可以实现LLaMa、GPT、Bloom、Chat-GLM、悟道、紫东太初等为代表的大模型全面应用。

如何进一步提升国产AI芯片竞争力？中国工程院院士、清华大学教授郑纬民认为，要开发基于国产AI芯片的系统，这一过程中最重要的是做好生态建设。“国产AI芯片只要达到国外芯片60%的性能，如果生态做好了，客户也会满意。”郑纬民称。

徐樊磊建议，做好自主创新要从人才、设施、科研和生态领域四方面突破，逐步提升智能算力设备软硬件功能。在硬件方面，提高国产AI芯片的稳定性和兼容性，特别是提升芯片之间、服务器集群之间的数据传输效果。在软件生态方面，降低适配门槛，让开发者逐渐使用国产芯片生态。

英博数科CEO周韡韡：

从算力集成走向AGI全栈生态服务

● 本报记者 彭思雨

由于子公司英博数科在算力热门赛道上动作频频，鸿博股份今年来备受资本市场关注。作为鸿博股份谋求业务转型的又一次尝试，英博数科如何能够在成立伊始便取得英伟达A800 GPU供应权限，并在随之而来的AI大模型浪潮中成为炙手可热的算力供应商？英博数科未来能否保障持续的算力供给？对此，中国证券报记者日前在2023 AI Tech Day暨首届人工智能生态发展峰会上对鸿博股份副总裁、英博数科CEO周韡韡进行了独家专访。周韡韡表示，英博数科未来仍将通过合规方式持续采购英伟达芯片，同时为应对“一卡难求”，公司也在积极拓展芯片合作厂商，包括国产芯片厂商。

手握算力资源

中国证券报：英博数科成为英伟达北京AI创新赋能中心指定运营主体，公司如何超前布局AI算力？

周韡韡：鸿博股份在彩票印刷行业多年来保持龙头地位，也一直在寻找新的转型发展机会，扫地机器人、元宇宙赛道都有所关注。2021年，元宇宙概念爆发，那时我们团队希望在元宇宙赛道上找到机会。通过市场调研，我们发现无论是元宇宙这一新概念，还是虚拟现实旧概念，要想把元宇宙建成良性自循环的经济体系，都离不开底层算力的支撑。

一开始接触算力领域，我们便瞄向GPU方向，GPU算力是现阶段最适合人工智能场景的芯片，而英伟达在图像识别、算法推理领域早就建立起了影响力。

恰逢当时英伟达在开展中国区生态加速计划，推进创新赋能中心建设。我们了解到，英伟达不只是愿意提供芯片硬件交易，更加愿意提供后续一系列的辅助开发服务。而我个人借助一些资源，以鸿博股份代表的身份与英伟达进行了沟通。

中国证券报：公司目前采购英伟达芯片的状况如何？

周韡韡：我们的第一个计算中心投产之后，获得了正向现金流回报，收入远超预期，因此公司进一步加速算力采购。2023年2月，公司交付首个DGX A800 AI算力集群。截至2023年10月，累计交付3000PFLOPS AI算力，持续扩张单一通讯集群超大智算单元。这里的GPU以H系列为主，从今年4月份之后，公司就没有再去订购A系列。

实际上，英伟达产品“一卡难求”主要原因仍是产能供应不足。目前英伟达GPU标准等货周期长达52周以上，意味着要等一年以上的时间才能拿到货。我们没有继续订购A系列也是基于对其产能的评估。

我们对英伟达采购的事情肯定是不会间断的，英伟达也会继续在中国区推出合规产品。公司现在也在和AMD、英特尔以及国产芯片生态进行深入沟通。市场一直在试图给公司贴上英伟达标签，但事实上，我们从底层架构到软件加速器开发，和众多品牌都有合作，而且效果不差。

提供全场景服务

中国证券报：未来英博数科业务依然会以做服务器集成为主吗？

周韡韡：公司最开始希望成为专业的AI多模态大模型训练平台，那时候我们常说，别人是AI时代的“淘金者”，我们则是那个“卖铲子的人”。

随着与客户长期沟通，我们发现他们有更多实际需求。比如，一些公司不擅长把研发模型和特定商业场景链接起来，一些创业团队缺乏资金。英博数科是重集成、轻研发的公司，在市场活动宣发、商业化场景转换以及投融资顾问服务方面非常擅长。

公司在此次活动上宣布战略升级，致力于从大模型算力服务商转变为AGI全栈生态服务平台。从最底层的算力基础设施，到大模型基座，再到最上层的商业化AI应用场景落地，公司将为客户提供全场景的服务。

中国证券报：公司此次发布了搏博云开发者平台，和百度、腾讯等大厂所提供的从IaaS到PaaS大模型生态服务平台相比有何优势？

周韡韡：我们的优势体现在我们更专注于GPU，也就是针对通用人工智能的多模态训练场景下的云平台调度。我们可以背靠原厂的支持，也可以凭借自己的先发优势，持续为客户提供更好的服务。

大厂实力确实非常强，但大厂可用于单点爆破的能力有限，他们不会割舍已有的“现金奶牛”业务，在新业务上拓展资源。然而AGI赛道正值重构整个产业架构之时，如果不“ALL IN”是很难成功的。我们虽然小，但精力更聚焦，资源可调度性更灵活。

实探北京二手房市场：

成交放大 线上看房人数增多

● 本报记者 董添

12月19日，中国证券报记者走访北京地区二手房市场发现，受北京市发布重磅楼市新政影响，成交出现较为明显的放量。最近几天，受大雪天气影响，线下看房受阻，但线上看房数量激增，VR看房再次火爆。有销售人员称，为线上客户讲解数量比前期大概增长了四成。由于普宅标准调整，部分豪宅有望变成普宅，相应贷款额度将增加，以海淀区、西城区为代表的豪宅聚集区看房人数激增。

成交有所放大

周六日作为北京二手房市场成交主力，北京地区重磅新政发布后的首个周六日成交情况受到业内高度关注。

从草签情况看，中原地产监测数据显示，12月16日（周六）、12月17日（周日），北京地区二手房合计成交1700套左右，比之前几周平均上涨了15%左右。

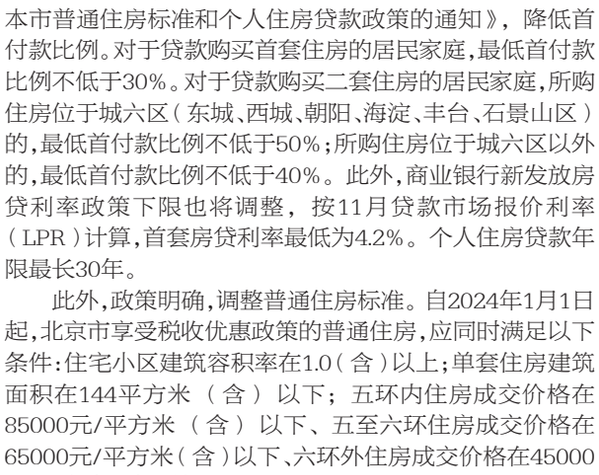
记者走访二手房市场时发现，无论是城六区还是非城六区，二手房都有成交放大的迹象。

“这几天成交的都是前期已经看了很长时间房子的客户。我们跟买房者说，趁着价格没有变动，可以尽快下手。很多购房者也感觉到现在是个阶段性的价格低谷，所以趁着周末就把合同签了。”位于大兴区一家我爱我家门店经理告诉记者，“目前，卖房子的客户很多也是连环单，也在积极看房。有些客户买了新房，手里的房子着急出售，价格也压得比较多。”

记者走访西城区一些学区房片区了解到，新政对于很多换房客户来说是非常具有吸引力的。“新政之前，贷款年限最高是25年，新政之后，贷款年限最高是30年。以贷款300万元为例，按照新政，30年还贷期限，按照等额本息还款法，每个月只需要还15649.42元。在此之前，25年还贷期限，按照等额本息还款法，每个月要还18154.86元。这样一来，每个月能少还2500元左右。如果贷款额度高，还款压力下降得更明显。”西城区金融街片区一家中介销售人士告诉记者。

15%

从草签情况看，中原地产监测数据显示，12月16日（周六）、12月17日（周日），北京地区二手房合计成交1700套左右，比之前几周平均上涨了15%左右。



视觉中国图片

线上咨询火爆

“这几天下大雪，线上看房的人特别多。VR看房又重新火爆起来了。我们店面线上接待量大概增长了40%。”丰台区一家链家门店销售人员告诉记者，“现在业主心态开始变化，一方面临近年底，很多业主希望年前回款；另一方面，很多业主认为明年3、4月份之前，楼市会有小阳春出现，所以价格还在胶着。现在看房的新客户，签约还是比较快的，一些看了很久的老客户，也慢慢开始成交。”

12月14日，北京市住建委等5部门发布《关于调整优化本市普通住房标准和个人住房贷款政策的通知》，降低首付款比例。对于贷款购买首套住房的居民家庭，最低首付款比例不低于30%。对于贷款购买二套住房的居民家庭，所购住房位于城六区（东城、西城、朝阳、海淀、丰台、石景山区）的，最低首付款比例不低于50%；所购住房位于城六区以外的，最低首付款比例不低于40%。此外，商业银行新发放房贷利率政策下限也将调整，按11月贷款市场报价利率（LPR）计算，首套房贷利率最低为4.2%。个人住房贷款年限最长30年。

此外，政策明确，调整普通住房标准。自2024年1月1日起，北京市享受税收优惠政策的普通住房，应同时满足以下条件：住宅小区建筑容积率在1.0（含）以上；单套住房建筑面积在144平方米（含）以下；五环内住房成交价格在85000元/平方米（含）以下、五至六环住房成交价格在65000元/平方米（含）以下、六环外住房成交价格在45000元/平方米（含）以下。

中原地产首席分析师张大伟对记者表示，调整普通住房标准后，预计很多“豪宅”变为“普宅”，对一些改善型客户来说，是重大利好。

这一说法也得到了很多中介的认可。万柳片区一位链家工作人员告诉记者，新政出台后，100平方米至120平方米的房子是最火爆的。之前该片区很多房子都被归为“豪宅”，如果是改善型客户，基本没有贷款的空间。政策出台后，很多客户发现可以贷款了，咨询的人非常多。