

算力需求重心从训练转向推理 全球AI基础设施建设全面加速

近日，老牌科技公司甲骨文(Oracle)以单日大涨40%的行情表现震撼全球资本市场。原因在于该公司公布的2026财年第一财季(2025年6月1日-8月31日)业绩数据显示，公司云基础设施(OCI)业务中的剩余履约义务(RPO)同比增长359%，达到4550亿美元，其中近60%来自与OpenAI签署的5年3000亿美元推理算力订单。

国内头部云计算厂商业务负责人告诉中国证券报记者，虽然这笔规模庞大的订单从签署到真正交付仍存在不确定性因素，但也向整个产业链传递了算力尤其是推理算力供不应求的明确信号，增强产业链坚定向AI和云领域投入的信心。

●本报记者 杨洁



视觉中国图片 制图/苏振

算力需求结构改变

根据甲骨文的判断，算力的需求结构已经从训练为主，向推理为主迁移。这也意味着AI已经从过去的模型研发训练进入到大规模产业应用的新阶段。

甲骨文首席执行官Safra Catz表示，AI推理市场的规模将远超AI训练市场，目前市场上的推理算力正被耗尽。甲骨文凭借过去在企业数据库领域的领先地位，使企业客户能在保障数据私密性的前提下，安全地结合私有数据与公有数据，利用大语言模型进行推理。

基于此，甲骨文给出了极其乐观的业绩指引，预计公司2026年的云基础设施收入将同比增长77%，达到180亿美元。公司还表示，之后的四年里，公司云基础设施收入将分别增至320亿、730亿、1140亿和1440亿美元。

中国信息通信研究院云计算与大数据研究所副所长栗蔚认为，大模型推理应用，正在重塑云智算技术体系，成为推动智能化转型的核心引擎。她指出，算力结构转型正在从以训练为主，转向以推理为核——当前70%以上算力用于集中式训练，未来70%以上算力将用于分布式推理。

IDC此前也发布报告认为，AI算力有望在2025年进入“训推”拐点，即从训练驱动过渡到推理驱动。

这种变化将意味着，大模型

AI基础设施全球战略角逐

在AI应用牵引之下，AI Infra(AI Infrastructure，即AI基础设施)的确定性进一步增强。AI Infra是支撑AI大模型从“概念”走向生产力的智能化底座。如今全球云厂商正全力抢占AI Infra这一战略制高点。

AI云是AI Infra的重要构成。市场调研机构Omdia发布报告显示，生成式AI(GenAI)带来AI云市场爆发，2025年上半年，中国AI云市场规模达223亿元，预计2025全年增长148%，到2030年市场规模将达到1930亿元。

统计显示，BAT(百度、阿里、腾讯)第二季度的资本开支合计达615.83亿元，同比上涨168%，重点

仍需突破不少技术瓶颈

伴随AI Infra领域“军备竞赛”愈演愈烈，万卡算力集群甚至更大规模的十万卡算力集群成为云厂商们的投资标配。

“如何减少互联的效率损失，最大限度提升整体的计算效率，对于运营大规模算力集群非常关键。”有专家指出。

同时，该专家坦言，相比美国，我国算力芯片生态比较碎片化，有几十款算力芯片，不同的芯片对应不同的开发框架、软件栈以及算子库等，也给建设和运营万卡级算力集群带来诸多挑战。

产业商业模式逐步由单纯的技术竞赛过渡到价值共享与产业落地，强调高效推理技术与垂直应用场景结合以实现商业闭环。“从算力需求的结构性转换，到技术架构的深度重构，再到大模型推理服务质量的场景化演进，云智算技术正迈向从算力供给到智能赋能的新范式，为数字经济的高效、普惠与可持续发展提供坚实技术支撑。”栗蔚表示。

阿里云智能集团弹性计算、存储产品线产品负责人告诉记者，企业端对AI大模型的需求已经不再只是尝鲜，而是进入了生产化阶段。

国家数据局局长刘烈宏此前在国新办新闻发布会上介绍，截至目前6月底，我国日均Token消耗量已突破30万亿，1年半时间增长了300多倍，反映了我国人工智能应用规模的快速增长。

沙利文最新报告数据显示，中国企业级大模型调用量大概占据了国内整体市场的1/3，日均Token的消耗量约10.2万亿，较2024年底暴涨363%。

AI应用驱动云厂商业绩增长的逻辑在国内同样得到验证。阿里巴巴最新季度业绩显示，阿里云季度收入同比增长26%至333.98亿元，增速创三年新高，其中，AI相关收入连续八个季度实现三位数同比增长，带动公共云需求持续上升。

算力需求驱动 产业链加快技术迭代

●本报记者 郑萃颖

AI算力需求高速增长，推动核心器件、封装技术快速迭代。在近日举办的第二十六届中国国际光电博览会上，AI与光通信的深度融合成为展会焦点。与此同时，上市公司积极回应光模块、共封装光学(CPO)技术进展。

财报显示，受需求驱动，光模块企业上半年业绩增长。行业专家预计，AI产业发展将直接推动光模块需求激增，1.6T模块成为竞争焦点，但行业还需经历产能与良率爬坡过程。

头部厂商引领技术变革

在中国国际光电博览会上，光通信领域头部企业集中亮相。如光迅科技展出产品生态矩阵及多款通感一体化创新产品。立讯技术带来以光电产品为核心的的数据中心互连整体方案，并展示了CPO、1.6T光模块、LPO/LRO低功耗方案等。针对下一代数据中心需求，爱德泰在展会中首次公开展示了CPO内连解决方案，把多个核心组件集成于CPO混纤模块中，推动光互联向高集成化、小型化方向发展。

随着光模块成为热门词，上市公司也于近期纷纷回应相关技术进展。中际旭创回复投资者提问称，已在CPO、线性驱动可插拔光模块(LPO)、硅光等下一代光模块技术领域积极布局。瑞斯新材表示，面对光模块持续升级的市场需求，公司持续推进光模块芯片基座/壳体项目建设，计划打造年产2000万套光模块芯片基座、1000万套光模块芯片壳体产能。

上银基金权益投研部基金经理惠军表示，在AI算力集群与超算中心建设的双重推动下，2025年全球光模块市场迎来结构性爆发，预计市场规模达121亿美元。技术路径方面，CPO技术加速渗透，预计2026年在AI数据中心的渗透率超20%，成为下一代光互连主流方案；LPO因低功耗特性受青睐，对超大规模数据中心极具吸引力。

需求增长提升公司业绩

新易盛半年报显示，今年上半年公司实现营业收入104.37亿元，同比增长282.64%；净利润39.42亿元，同比增长355.68%。高速光模块需求快速增长，今年上半年，新易盛点对点光模块销量695万只，相比上年同期的327万只同比增长112.54%。

中际旭创上半年营收147.89亿元，同比

增加36.95%；净利润39.95亿元，同比增长69.40%。中际旭创在近期投资者关系活动上表示，由于AI基础设施的建设浪潮，800G和1.6T产品的需求增长迅速。公司表示，“AI算力数据中心的需求主要是1.6T和800G，预计3.2T放量仍需要一些时间。”目前中际旭创已具备3.2T产品的生产能力。

天孚通信上半年实现营收24.56亿元，同比增长57.84%；净利润8.99亿元，同比增长37.46%。公司认为，面向AI和通用计算的数据中心光模块将成为2025年市场增长最快的领域。

纳真科技招股书显示，今年上半年公司收入同比增长124%，达到42.16亿元，其中光模块收入占总收入的77.9%。目前纳真科技已开发并批量生产基于LPO的光模块，并正在开发基于LRO的光模块，NPO、CPO技术攻关也在积极推进，旨在推出支持3.2T甚至6.4T的超高速产品。

IDC数据显示，全球AI服务器市场规模有望在2028年增长至2227亿美元，算力需求的指数级增长推动了数据中心内部和数据中心之间对更高带宽和更低延迟的光通信需求，也推动市场CPO技术的关注度提升。

面临产能爬坡

国盛证券表示，随着全球数据中心加速升级扩容，800G光模块已大规模部署，1.6T光模块也开始进入市场，行业面临的挑战已经从“需求在哪里”转变为“如何按时交付”。光模块的产能扩张，需要高端设备采购、厂房建设、人员培训，同时提高良率也需要时间。CPO等技术仍处于发展早期，800G、1.6T可插拔模块仍面临量产良率和工程化的挑战。

中际旭创在投资者关系活动上表示，目前行业光模块需求快于原材料的供给速度，一些重点原材料的供应偏紧张，对公司备货也提出了更高要求。

新易盛表示，公司泰国工厂二期已于2024年末建成并于2025年初投产，目前处于产能持续扩充阶段。天孚通信则表示，公司泰国工厂预计明年进入大规模量产。

近日光迅科技发布公告，拟定增募资不超过35亿元，用于算力中心光连接及高速光传输产品生产建设项目、高速光互联及新兴光电子技术研发项目、补充流动资金。其中，20.83亿元用于算力中心光连接及高速光传输产品生产建设项目，6.17亿元用于高速光互联及新兴光电子技术研发项目。

持续突破 国产AI芯片抢夺市场窗口期

●本报记者 王婧涵

从年初阿里宣布未来三年投入3800亿元建设云和AI硬件基础设施，到百度昆仑芯签下中国移动大额订单，市场对AI芯片的强劲需求，推动国内芯片厂商和互联网大厂利用英伟达芯片供应不稳定的窗口期全力抢占市场。

行业人士普遍认为，当前在AI推理芯片上，国内如寒武纪、海光信息、摩尔线程等公司的产品能力已经不错，未来几年国产芯片的发展前景广阔，但仍需在生态建设上持续发力。

大厂加码AI基础设施

Bernstein 7月发布的研报表示，英伟达芯片供应的不确定性为国内AI芯片供应商创造了市场机遇，预计2025年国内AI芯片需求将达到395亿美元，AI芯片市场的本土化率将从2023年的17%增长至2027年的55%。

近年来，AI需求的爆发让相关企业成为市场的焦点。事实上，国内大厂对AI芯片等基础设施的布局早已展开。2018年，阿里成立平头哥，华为发布两款昇腾AI芯片，百度启动昆仑芯研发等众多重要事件集中爆发，头部企业对AI芯片的产业化布局已初见雏形。

阿里、百度等企业一方面持续加码AI基础设施投入，保持资本开支高增长；另一方面也开始积极寻求外部订单，推进自研AI芯片的商业化。

年初，阿里宣布未来三年投入3800亿元建设云和AI硬件基础设施。在2025年半年度业绩说明会上，阿里巴巴集团首席执行官吴泳铭表示，过去四个季度，公司已在AI基础设施以及AI产品研发上累计投入超过1000亿元。

2025年云智大会上，百度宣布昆仑芯超节点在百舸AI计算平台5.0版本上可用。据悉，新一代昆仑芯P800系列产品采用昆仑芯100%自研的XPU-P架构，目前应用的最大集群规模超过3万卡。

8月底昆仑芯宣布中标中国移动2025年至2026年人工智能通用计算设备(推理型)集中采购项目的十亿元级订单，在“类CUDA生态”标段的三个标包中均拿下主要份额，成为互联网大厂自研芯片获外部大额订单的重要案例。

融资和整合推动快速发展

2025年上半年，以寒武纪、芯原股份为代表的国产AI芯片厂商业绩及订单表现格外亮眼。

9月12日，芯原股份发布公告表示，截至2025年第二季度末公司在手订单金额30.25亿元，已连续七个季度保持高位。7月1日至9月11日，公司新签订单12.05亿元，其中AI算力相关订单占比约64%。

寒武纪上半年实现营业收入28.81亿元，同比增长4347.82%。公司产品中作为云

服务器、数据中心等人工智能处理器的云端产品线营收达28.7亿元，2024年同期该项产品线收入为6105万元，云端产品线客户贡献了上半年绝大多数业绩增量。

海光信息2025年上半年实现营业收入54.64亿元，同比增长45.21%；实现归母净利润12.01亿元，同比增长40.78%。公司表示，随着多模态AI应用场景的持续拓展，国产算力基础设施正成为驱动产业智能化升级的核心引擎。报告期内国产高端芯片市场需求持续攀升，公司通过深化与整机厂商、生态伙伴在重点行业和重点领域的合作，推动公司高端处理器产品的市 场版图扩展，实现了营收增长。

行业人士表示，目前国内如寒武纪、海光信息、摩尔线程等公司在AI推理芯片上的能力已经不错，当下最重要的是抓紧时间提升产能，满足快速增长的AI算力需求。

8月以来，多家相关企业也通过新一轮融资和整合，实现快速发展。

华虹公司宣布筹划以发行股份及支付现金方式购买华力微控股权并配套募集资金。公司表示，此交易有助于公司实现规模效应，提升公司的市场占有率为盈利能力。

中芯国际公告表示，筹划以发行人民币普通股的方式购买公司控股股东中芯北方49%的股权。交易完成后，中芯国际在中芯北方的持股比例将从51%上升至100%。

寒武纪公告表示，收到证监会出具的批复，证监会同意公司向特定对象发行股票的注册申请，募集资金总额不超过39.85亿元。

重视软硬件生态协同

国产高端芯片持续突破的同时，多位行业人士也提到发展软件生态的重要性。

清华大学计算机系教授翟季冬表示，目前国内算力硬件水平已经接近甚至超过了英伟达同类芯片，但在软件生态上仍有提升空间。

“想要完善软件生态，底层系统中的调度器、内存管理、容错系统、存储系统，中层的编程语言、AI编译器、算子库，上层的编程框架、通信库和并行系统几乎缺一不可。”翟季冬表示。

吴泳铭也提到，“由于AI带来的云计算市场变革，未来云计算市场的集中度将有很大提高。开发者在开发利用时需要的技术栈很多，会倾向选择具备全方位技术产品组合的厂商，特别是在各单项产品上都具备领先地位的厂商。”

上海证券研报分析表示，完善的软件平台可以充分发挥智能芯片的硬件性能、增强智能芯片对大模型新需求的适应性，降低用户的使用门槛，使其更高效地完成人工智能算法和应用的技术创新。目前，国内头部企业正持续完善AI芯片生态，进一步提升市场竞争力。

从果园到餐桌 山东交行用金融画笔绘就乡村振兴新图景

乡村振兴作为实现中国式现代化的重要组成部分，国家提出要加大对农业农村的投入，确保农业农村优先发展。为进一步加强金融支持乡村全面振兴，做实做优金融五篇大文章，交通银行山东省分行聚焦乡村振兴重点领域和关键环节，立足农企、农户实际需求，持续强化金融供给，创新金融产品和服务模式。

为全力保障秋天丰收农户的资金需求，交通银行依托“科技+场景”双轮驱动，创新金融产品和服务模式，一是积极对接中化集团，不断拓展助力新型农业经营主体发展的方法和路径；二是深化银担合作，精准为农业经营主体提供全方位金融支持；三是深入农村农户，走上田间地头，切实了解农企的所想所需，并为其定制专属金融服务方案。

田先生是烟台蓬莱区种植大户，主要从事苹果、樱桃

的规模化种植，交通银行客户经理在走访过程中了解到，企业目前因经营周转面临300万元资金缺口，为缓解企业资金压力，帮助企业平稳度过当前经营困境，交通银行山东省分行联合山东省农业发展发展信贷担保有限公司为客户提供专项担保贷款，根据苹果种植的特点，结合田先生的基础信息、实物资产等信息，为其申请了“兴农e贷”。

“兴农e贷”是交行针对普惠贷款主体和新型农业经营主体的特色涉农业务，具有流程短、审批快、额度高、利率低的特点。客户经理优化金融服务，争取利率优惠，提高工作效率，仅用时4天完成300万涉农贷款的申请、审批和发放，该笔贷款将用于采购化肥、农药等农资，覆盖26.97万平米种植面积，有效保障农企生产经营，破解农企融资难题。

此外，交通银行山东省分行针对农业主体“缺抵押、周期长”的痛点，推出“鲁粮快贷”、“蔬菜贷”等线上产品，结合区域特色开发了“苹果贷”、“果蔬贷”、“海参贷”等场景化贷款，通过与核心企业、银担协同，降低农企、农户融资成本和融资门槛，确保贷款精准流向农业经营主体。

交通银行山东省分行积极践行金融为民初心，实实在在为乡村产业发展、新农村经营主体成长提供金融助力，以金融活水润泽田间地头，通过“产品创新+政策协同+科技提速”组合拳，为新农注入强劲动能，彰显国有大行服务乡村振兴的责任担当。下一步，交通银行山东省分行将继续积极践行金融工作的政治性、人民性，以更大决心、更多真情提升金融服务乡村振兴的质量和水平。

-CIS-