

聚焦2026世界人工智能大会

算力竞赛迈入新阶段： 从“单卡比拼”走向“系统级较量”

◎ 记者 宋薇萍

2026世界人工智能大会（WAIC 2026）世博展馆内，一组组形庞大黑色机柜前排队起长队。参观者等待的并非某个消费电子产品，而是“超节点”——今年大会最受关注的算力基础设施。

现场，华为首次公开展出昇腾950超节点（Atlas 950 SuperPoD）真机。这款以单柜64卡为基本单元、最高可扩展至1024张NPU卡互联的产品，提出一个目标：“像一台计算机一样工作”。不远处，沐曦股份发布曦景S600超节点，主打“零线缆”设计；昆仑芯集中展出从芯片、加速卡、整机到超节点的完整产品矩阵，其32/64卡超节点是国内率先实现量产交付的超节点产品之一；曦智科技则展示了基于光电混合路线的光跃超节点，已实现2000卡规模落地。

从芯片厂商到云服务商，从光电技术到太空计算，AI基础设施的竞争主线已经清晰——不再是单一芯片算力的比拼，而是系统级能力的全面较量。

算力需求生变

“过去大家问的是‘你们有多少张卡’，现在问的是‘Token成本多少，响应时延多长’。”大会现场，一位算力服务商负责人这样描述市场变化。

这种转变源于AI应用重心的迁移。大模型训练的高峰期过后，行业焦点正加速转向推理落地和智能体应用。算力的性质随之改变——从一次性采购的固定资产，转变为持续性的运营支出。单位Token的推理成本、毫秒级的响应时延，成为决定商业模型能否跑通的关键指标。

“AI Agent等智能推理应用快速爆发，推理算力需求将达到训练算力的10倍甚至100倍。”燧原科技董事长、CEO赵立东对记者表示，训练需要的是极致的算力，而推理更看重性价比。目前，行业对芯片的评判标准已发生转变，业界不再单纯追逐芯片峰值算力，单位Token成本、性价比成为关键。

华为公布的案例显示，昇腾超节点在低时推理场景中，单卡性能相比传统服务器提升3倍，推理成本大幅下降。其核心支撑是灵衢互联协议，本次亮相的1024卡规模昇腾950超节点是目前业界规模最大的超节点，可实现256TB跨物理节点统一内存编址，RTT时延控制在3微秒以内。

什么样的节点才算“超”

“超节点”概念并不新鲜，但今年大会上，产业界试图给出更严格的定义。

根据现场多家企业的共识性表述，真正的超节点必须具备三个技术特征：超大带宽、超低时延、统一内存编址。其中“统一内存编址”被视作灵魂——它意味着不同节点的内存被纳入同一个地址空间，任意处理器可直接通过Load/Store指令读写远端内存，无须经过额外的编解码流程。

按照这个标准，可以明确区分“真超节点”与“伪超节点”。后者本质上仍是传统服务器堆叠架构，依赖PCIe或RoCE协议互联，跨服务器带宽、时延受限，不支持内存语义的直接访问。

“好比一个城市，如果每个区域之间通行都要经过收费站和翻译，效率必然大打折扣；真正的超节点就像把整个城市的路网统一编码，车可以直接开到任何地方。”现场技术人员打了个比方。

昇腾384超节点是这一理念的先行者。华为透露，该产品自2025年发布以来已规模商用750多套，覆盖互联网、运营商、金融、医疗、交通、制造等行业，是国内唯一训练出SOTA（状态最优）模型的超节点产品，也是目前中国唯一规模商用的超节点产品。

国产方案密集落地

华为之外，多家国产算力企业在本届大会上拿出了系统级产品。

沐曦股份发布的曦景S600超节点，采用OEX架构设计，通过正交连接器和两级交换拓扑，实现单机柜64卡高速互连，并可横向堆叠扩容至万卡集群。其“C3O设计”——计算节点零线缆、交换节点零线缆、节点之间零线缆——直接回应了大规模集群部署中布线复杂、信号损耗、故障率高企的工程痛点。产品采用解耦模块化设计，计算、交换、电源、液冷模块均支持快速部署和维护，电源与液冷管路支持盲插热插拔，单节点故障也不会影响整集群运行。

昆仑芯在此次大会上集中展出了从芯片、加速卡、整机到超节点的完整产品矩阵。展区中，昆仑芯32/64卡超节点吸引了众多观众驻足。据现场工作人员介绍，该产品是国内率先实现量产交付的超节点之一，采用自主研发的超高密度集成架构，单柜可集成32或64张加速卡，并支持扩展至512张卡。通过柜内全互联设计，卡间通信



➡ 真正的超节点必须具备三个技术特征：超大带宽、超低时延、统一内存编址

➡ 华为昇腾 950 超节点提出一个目标：“像一台计算机一样工作”

➡ 沐曦股份发布曦景 S600 超节点

➡ 燧原科技首发新一代燧原超节点系统——云燧 ESL64-O 超节点



2026世界人工智能大会世博展馆内，企业纷纷展示超节点产品。本版图片 记者 宋薇萍 摄

从WAIC看国产算力“新”布局

◎ 记者 赛世平

当万卡集群逐渐成为AI算力基础设施的重要形态，行业关注的重点也在发生变化。算力规模仍在扩张，但比“卡数”更受关注的，是大规模集群能否稳定运行，国产软硬件能否形成系统能力，以及建成的算力如何真正支撑模型、科研和产业应用。

在7月17日举行的2026世界人工智能大会上，曙光8000（登峰）十万卡AI超集群真机首次亮相，并获评大会“场馆之宝”。中科曙光副总裁王圣勇在大会现场接受上海证券报记者采访时表示，十万卡集群的落成和展出，代表国产算力基础设施迈上了一个新的台阶。除了规模扩大，这套系统中的核心技术体系和核心部件均实现国产自研，很多已达到国际先进水平。

“在7月17日举行的2026世界人工智能大会上，曙光8000（登峰）十万卡AI超集群真机首次亮相，并获评大会“场馆之宝”。中科曙光副总裁王圣勇在大会现场接受上海证券报记者采访时表示，十万卡集群的落成和展出，代表国产算力基础设施迈上了一个新的台阶。除了规模扩大，这套系统中的核心技术体系和核心部件均实现国产自研，很多已达到国际先进水平。

“在7月17日举行的2026世界人工智能大会上，曙光8000（登峰）十万卡AI超集群真机首次亮相，并获评大会“场馆之宝”。中科曙光副总裁王圣勇在大会现场接受上海证券报记者采访时表示，十万卡集群的落成和展出，代表国产算力基础设施迈上了一个新的台阶。除了规模扩大，这套系统中的核心技术体系和核心部件均实现国产自研，很多已达到国际先进水平。

“在7月17日举行的2026世界人工智能大会上，曙光8000（登峰）十万卡AI超集群真机首次亮相，并获评大会“场馆之宝”。中科曙光副总裁王圣勇在大会现场接受上海证券报记者采访时表示，十万卡集群的落成和展出，代表国产算力基础设施迈上了一个新的台阶。除了规模扩大，这套系统中的核心技术体系和核心部件均实现国产自研，很多已达到国际先进水平。

“在7月17日举行的2026世界人工智能大会上，曙光8000（登峰）十万卡AI超集群真机首次亮相，并获评大会“场馆之宝”。中科曙光副总裁王圣勇在大会现场接受上海证券报记者采访时表示，十万卡集群的落成和展出，代表国产算力基础设施迈上了一个新的台阶。除了规模扩大，这套系统中的核心技术体系和核心部件均实现国产自研，很多已达到国际先进水平。

“在7月17日举行的2026世界人工智能大会上，曙光8000（登峰）十万卡AI超集群真机首次亮相，并获评大会“场馆之宝”。中科曙光副总裁王圣勇在大会现场接受上海证券报记者采访时表示，十万卡集群的落成和展出，代表国产算力基础设施迈上了一个新的台阶。除了规模扩大，这套系统中的核心技术体系和核心部件均实现国产自研，很多已达到国际先进水平。

“在7月17日举行的2026世界人工智能大会上，曙光8000（登峰）十万卡AI超集群真机首次亮相，并获评大会“场馆之宝”。中科曙光副总裁王圣勇在大会现场接受上海证券报记者采访时表示，十万卡集群的落成和展出，代表国产算力基础设施迈上了一个新的台阶。除了规模扩大，这套系统中的核心技术体系和核心部件均实现国产自研，很多已达到国际先进水平。

“在7月17日举行的2026世界人工智能大会上，曙光8000（登峰）十万卡AI超集群真机首次亮相，并获评大会“场馆之宝”。中科曙光副总裁王圣勇在大会现场接受上海证券报记者采访时表示，十万卡集群的落成和展出，代表国产算力基础设施迈上了一个新的台阶。除了规模扩大，这套系统中的核心技术体系和核心部件均实现国产自研，很多已达到国际先进水平。

“在7月17日举行的2026世界人工智能大会上，曙光8000（登峰）十万卡AI超集群真机首次亮相，并获评大会“场馆之宝”。中科曙光副总裁王圣勇在大会现场接受上海证券报记者采访时表示，十万卡集群的落成和展出，代表国产算力基础设施迈上了一个新的台阶。除了规模扩大，这套系统中的核心技术体系和核心部件均实现国产自研，很多已达到国际先进水平。

企业及个人用户全面开放普惠、高效、便捷的算力服务。

大集群背后的体系协同

随着集群规模和算力密度提高，AI基础设施越来越难以依靠某一个产品或单一环节完成。

在系统建设上，曙光8000具有“芯片、计算、存储、网络、散热、应用、服务”全链路全自研AI基础设施能力。其中：海光等国产芯片为系统建设提供底层支撑；scaleFabric类IB原生RDMA高速网络实现十万卡集群高可靠连接；ParaS-tor分布式存储支撑大模型训练和科学计算中的海量数据读写。

“十万卡集群这样一台机器，实际上是由多个团队协同完成的。”王圣勇说，曙光8000背后包括计算、存储、网络、整机集成和液冷等不同团队，是中科曙光多个产品线和技术能力共同作用的结果。

芯片是这一协同链条的起点。曙光8000搭载海光等国产芯片，也为国产芯片提供了更大规模的工程验证场景。本次大会上，海光完整展出CPU、DCU全系列自研芯片、高速互联整机集群、DTK/DAS/DAP全栈软件平台与全域Token化运营解决方案，目前已经形成从芯片到集群级算力协同的完整能力图谱。

海光信息总裁助理兼智能计算产品部总经理杜夏威接受上海证券报记者采访时表示，国产芯片进入十万卡集群后，评价标准正在由单芯片峰值性能和跑分，转向有效算力利用率、线性加速能力和故障恢复效率等系统级指标。

对于国产算力而言，硬件进入集群只是第一步。行业用户从原有平台迁移时，仍然面临技术路线差异、软硬件适配和性能调优等问题。杜夏威表示，海光CPU和DCU不仅具备异构融合能力，同时也完整兼容了主流软件应用生态，通过软件栈无缝兼容、模型广泛适配、生态开放协同等方式，降低用户迁移调优成本。

多位受访人士透露，未来，随着大集群由建设转向应用，国产算力产业协同也将继续向软件适配、模型运行和行业场景延伸。

从“凭税放款”到“现场看厂” “银税互动”告别粗放时代

◎ 记者 黄坤

10年前，一份纳税记录，曾让无数小微企业第一次敲开银行信贷的大门；10年后，这把曾经的“万能钥匙”，仍是银行放贷的重要参考，但正在失去“一键授信”的魔力。

近日，上海证券报记者从业内采访获悉，今年以来，多家股份制银行、城商行陆续调整税务业务，由过去“全线上自动审批”转向“线上申请+线下复核”，部分银行还提高了新客户准入门槛，并切断中介渠道。

从“凭纳税记录就能贷款”，到客户经理走进企业厂房、仓库和生产线——银行税务业务正变得审慎，这折射出“银税互动”业务的进一步规范和升级，银行普惠金融业务进一步强调质量优先。

从一键授信到调整优化：纯线上税贷“退潮”

“都在缩减纯线上税贷。”某城商行普惠金融业务部人士李梦（化名）告诉记者，其所在银行已对税务业务进行调整：小额优质客户仍可保留全线上办理，大额或经营波动客户，必须由客户经理开展线下尽职调查。

华东一家银行小微信贷人士也向记者证实，该行对线上税贷进行了调整，已切断中介渠道，并同步税务动态贷后预警，搭配差异化利率与政府担保分险，投放策略从过去追求规模增长转向“精细化，低不良”路线。

还有银行直接下架这一类产品。“我们目前没有线上税贷相关产品了。”某股份行人士告诉记者。

据李梦观察，今年行业调整节奏明显加快。目前税贷变化主要体现在三个方面：一是产品模式变化，由过去全线上模式，转向线上申请、线下复核，核实企业实际经营场景；二是产品数量减少，部分银行对存量产品进行升级更新；三是新客户准入门槛抬高，更加关注企业经营稳定性。

线上税贷，曾是银行普惠金融领域的“明星产品”。2015年前后，随着税务数据电子化快速推进，税务数据能够较为真实地反映企业经营情况，加之审批流程标准化、线上化程度高，线上税贷让大量此前难以获得贷款支持的小微企业，第一次拥有了信用额度。

纯信用、无需抵押、审批快、成本低，这是线上税贷当年能够迅速推广的主要原因。李梦表示，线上税贷让普惠金融业务真正实现“增量扩面”，但也易滋生虚开发票、多头共债、中介包装造假等风险，如今真正的纯线上税贷已经越来越少。

税贷从兴盛到收缩，在业内看来，意味着普惠金融业务从规模扩张阶段，步入质量优先与风险出清阶段。

苏商银行特约研究员薛洪言在接受记者采访时表示，银行对税贷业务的调整，在于纠正过去过度依赖单一税务数据模型而忽视实质性风险识别的偏差，通过引入线下复核能更真实地验证企业经营与纳税数据的匹配度，有效拦截中介包装的虚假申请，在宏观经济增长放缓背景下筑牢资产质量防线。

从饱受青睐到风险暴露，单一数据难辨经营能力

曾经受到银行青睐的线上税贷，何以步步收缩？核心原因在于风险收益平衡发生变化。

“最大的压力来自资产质量，不良率大幅攀升。”李梦向记者表示，现阶段小微企业经营基本面走弱，市场需求不足、原材料与人力成本持续上涨，大量小微企业现金流薄弱，第一还款来源不足，仅凭纳税数据已难以判断其真实偿债能力。

在实操中，单一依靠静态纳税数据，容易产生风控盲区。上述小微信贷人士表示：企业过去纳税情况良好，并不代表未来现金流一定稳定；企业账面经营数据正常，也不代表不存在隐性负债、资金挪用等风险。

税务数据有重要价值，但绝非唯一信贷审批依据。过去，市场对税务存在一个认知误区，即认为税贷就是依靠税务数据完成放款。实际上，税贷本质是一种数字化信贷产品，税务数据只是其中的重要参考，需要结合更多数据信息判断企业还款能力和还款意愿。

围绕税贷的数据包装和欺诈风险也逐渐出现。“部分经营主体会通过调整纳税申报数据、虚增营业收入的方式骗取更高额度贷款。”中国城市发展研究院投资部副主任袁帅向记者说，甚至有中介机构批量包装空壳企业的税务数据，帮助不符合资质的主体套取银行信贷资金。

单纯依靠税务数据授信，也提高了银行风险处置难度。李梦表示，税贷大多无抵押、无第三方担保，出现不良后处置难度极大、回收比例极低，相比担保类的经营贷风险损失倍数更高。

与此同时，监管导向变化也促使银行更加重视税贷的贷款质量，而非单纯追求规模。

从凭税放款到多维风控：税贷探索发展新路径

今年上半年，税务总局、金融监管总局联合印发《关于进一步规范和规范“银税互动”工作的通知》，进一步优化银税合作机制等。截至2025年底，全国银行业金融机构已累计通过“银税互动”机制为诚信纳税的小微企业发放贷款4517.72万笔、贷款金额15.7万亿元。

谈及税务业务具体的发展方向和路径，南京银行上海分行普惠小微信贷业务相关负责人在接受记者采访时表示，银行不是要放弃税务数据，而是要从“单维税务数据”升级为“多维交叉验证+线上线下双风控”。

香港中文大学（深圳）高等金融研究院客座教授郑磊也向记者表示，银行应放弃对纯线上模式的过度依赖，构建“数据+人工”的双轮驱动风控体系，同时严守合规底线，防范中介黑产渗透，如此才能在税贷业务的下半场竞争中赢得可持续发展空间。

有银行在针对不同行业，探索税贷差异化授信阈值。上述南京银行相关负责人表示：例如制造业企业可能存在进项抵扣较多、账面利润较低的情况；商贸企业流水较大但纳税波动明显；科创企业前期可能亏损、无大额纳税。

