

Open Storage Summit -- Day 4

主持人：

本次我们将围绕 “为人工智能 (AI) 实现企业应用现代化” 展开讨论，这也是超微开放存储峰会 (Super Micro Open Storage Summit) 的重要议题之一。我是罗布 (Rob)，现任董事总经理兼首席分析师。随着 AI 重塑企业技术格局，应用现代化已不再是选择题，而是一种竞争必需。传统的单体系统和遗留数据库正难以跟上原生 AI 工作负载 (AI native workloads)、实时决策以及弹性扩展的需求。对此，企业正加速向开放云原生架构转型，并将 PostgreSQL (简称 Postgres，需标注英文原文) 这类平台置于现代数据战略的核心。Postgres 具备可扩展性、成本效益及开放生态系统支持，能帮助企业摆脱专有系统的限制，但要充分发挥其潜力 (尤其是在 AI 驱动型应用中)，仍需克服不少挑战。

企业必须重新思考如何在技术栈的每一层实现性能交付。其中，高性能块存储 (high-performance block storage) 至关重要 —— 它能为 Postgres 的高事务负载提供低延迟、高吞吐量的访问能力。而在实际应用场景中，当 AI 推理 (AI inference) 与联机事务处理 (OLTP，需标注英文原文：Online Transaction Processing) 开始融合时，这种存储能力的重要性会进一步凸显。

在企业为提升敏捷性、可扩展性和可移植性而重构应用平台的过程中，Kubernetes (用户原文 “Cuba, netease” “Cooper netease” 为拼写误差，正确名称为 Kubernetes，需标注英文原文) 正发挥关键作用。Kubernetes 已成为部署、扩展和管理现代 OLTP 环境的基础架构或控制平面 (control plane)。当它与 AI 增强型微服务结合时，团队能够构建兼具 “云原生” “数据驱动” 特性且可实时输出洞察的智能应用。

在今天的对话中，我们将探讨基础设施层、数据层与应用层如何协同满足 AI 需求；深入分析从遗留数据库向 Postgres 转型的趋势（及高性能存储在其中的重要性）；并解读为何基于 Kubernetes 的现代化是交付下一代企业应用的关键。

非常荣幸邀请到几位嘉宾共同深入探讨这一话题：来自超微 (Super Micro) 的君霞 (Junxia, 音译)、来自 AMD 的希瓦 (Shiva)、来自光比特 (Lightbits) 的西吉 (Ciggy)，以及来自 EDB 的西蒙 (Simon)。欢迎各位参与本次讨论！

想请各位从各自的视角分享一下见解。如今“为 AI 实现企业应用现代化”已成为热点——过去几年，尤其是生成式 AI (Gen AI) 兴起后，相关讨论从未间断；如今我们又面临生成式 AI (Gen AI)、推理 (inference) 和检索增强生成 (RAG, Retrieval-Augmented Generation) 等多重需求，企业需为此做好充分准备。君霞，能否先从超微的视角为我们解读一下？

超微代表：

谢谢罗布。首先，非常荣幸能与 EDB、AMD 等合作伙伴共同参与本次会议。今天我们将探讨超微如何与行业领军企业合作推进企业级 AI 发展。

过去几年，AI 领域经历了重大变革：起初仅有十几家企业主导 AI 的研发与训练，但现在

我们看到了一个更显著的趋势 —— 企业正大规模采用 AI。从 AI 建模 (AI modeling) 向推理 (inference) 及实时应用的转型，需要全新的计算基础设施方案。超微从一开始就致力于推动这一变革，助力企业将 AI 实验转化为规模化部署。通过提供适配的规模化解决方案，超微正让各类规模的企业都能更便捷地获取 AI 基础设施。

接下来我想花点时间说明：企业部署 AI 基础设施时会面临哪些痛点，以及超微的机架级解决方案 (rack scale solution) 如何解决这些问题。我们知道，大多数企业的基础设施建设都源于 “逐个挑选离散组件” 的历史模式，这会导致 AI 基础设施的设计、部署和管理极为复杂，甚至可能出现瓶颈，限制可扩展性与性能。其次，大规模集群消耗的能源量巨大，不仅推高运营成本，还可能降低冷却效率。最后，网络复杂性已成为重要考量因素 —— 尤其是在 AI 数据中心部署中，大规模集群的网络问题更为突出。

而超微的机架级解决方案能简化部署并提升能源效率：我们首先与客户沟通咨询，明确其所需的平台与组件；随后根据客户的工作负载和电力预算设计方案；在发货前，我们会完成整个机架的组装、测试与验证。客户收到后只需接入网络和电源即可使用，大幅缩短部署时间。此外，我们还为客户提供持续的现场服务，确保集群在客户侧长期稳定运行。

机架级市场正快速增长：预计从 2023 年的 43 亿美元 (4.30 billion) 增长到 2030 年的 75.7 亿美元 (7.57 billion)，而超微是该领域的市场领导者。我们的整体解决方案能覆盖全流程，客户无需与多个供应商对接。目前，我们的解决方案已从单一系统扩展到机架集群，乃至数据中心建设层面。

值得强调的是，战略合作伙伴关系是我们提供全面前沿技术的关键 —— 例如与 EDB 的合作，以及采用 AMD 处理器的产品。光比特 (Lightbits) 能提供高性能、低延迟的块存储，支持 EDB 这类高要求的数据库应用；而 EDB 则提供企业级 Postgres 解决方案，支撑 IO 密集型企业的工作负载。

主持人：

这对我们理解主题非常有帮助。我认为，当企业计划实现应用现代化时，还需考虑如何通过这一过程推动企业应用向 AI 的下一阶段演进。希瓦，君霞已详细介绍了超微在机架级解决方案中的作用，能否再为我们解读一下 AMD 的处理器（CPU）和图形处理器（GPU）在这一体系中扮演的角色？

AMD 代表：

AMD 的产品已成为君霞刚才所提及的所有体系的核心构建块。我们拥有业内最全面的 AI 相关产品组合，今天我会重点介绍 AMD EPYC（霄龙）处理器——它是当前众多 AI 系统的核心，既可用作 GPU 的主机节点，也是推理部署的性能核心（CPU 在推理场景中同样表现出色）。

除了服务器端的高性能 CPU，我们还有面向重度训练场景的 GPU 产品线——AMD Instinct（本能）系列，该系列目前也广泛应用于各类推理场景。此外，我们还提供基础设施层面的产品：例如在网络领域，我们最新发布的 Pensando DPU（数据处理单元，Data Processing Unit），能高效实现节点间的数据传输，确保与所有计算过程协同。

此外，AMD 还积极参与各类开放标准的制定。例如，Ultra Accelerator Link（超级加速器链路）是新兴的行业标准，AMD 是该标准的主要推动者和支持者；我们还推出了 Versal 系列产品（基于 FPGA，现场可编程门阵列），能高效实现推理。

总而言之，AMD 可为所有 AI 相关场景提供最全面的技术支持，而超微则拥有业内最广泛的 AMD-based（基于 AMD）产品组合，双方的合作堪称典范。我们也与其他合作伙伴（后续幻灯片会提及）保持密切合作，但我想强调的是：AMD 在该领域的地位稳固，市场份额已大幅增长 —— 目前我们的市场份额约为 40%，众多大型企业已选择 AMD 作为其数据中心部署的合作伙伴，例如网飞（Netflix）—— 他们当前的系统均运行在 AMD 平台上，且正持续增加采购量。

这种市场认可度主要源于我们在各类工作负载中的出色表现 —— 以今天的核心议题 PostgreSQL（SQL 数据库）为例：客户关注两类基准测试，一是事务型测试（即 OLTP，衡量事务处理速度），二是查询型分析（客户的核心需求之一）。数据显示，在事务处理方面，我们的性能比竞品领先 58%；在查询分析方面，性能领先约 15%。更关键的是：我们的测试基于“单 CPU 服务器”，而竞品基于“双 CPU 服务器”—— 这本身就带来了巨大的成本优势：我们用每台服务器一半的 CPU 数量，实现了更优的性能。这正是 AMD 的核心价值所在，也是客户选择 AMD 产品能获得的核心收益：成本节省，性能无损。

主持人：

非常精彩。你提到的“单 CPU 实现高性能”这一点尤为关键——它不仅有助于降低功耗、缓解其他资源约束，还可能减少软件许可成本。西吉，能否为我们解读一下块存储（尤其是高速存储）的重要性？我知道在很多 Kubernetes 相关会议（如 KubeCon、CloudNativeCon）中，高速存储都是热点话题。光比特在这一体系中扮演着怎样的角色？

Lightbits 代表：

光比特 (Lightbits) 的核心产品是软件定义存储 (SDS, 需标注英文原文: Software-Defined Storage) 解决方案——这意味着它可在任意环境中运行。当然，我们也喜欢与超微 (Super Micro) 这类企业合作，共同打造经过验证的参考架构，确保方案能在极小的硬件占用空间内提供数百万的 IOPS（每秒输入输出操作数，需标注英文原文: Input/Output Operations Per Second）——这正是所有客户的核心需求。

客户希望在数据中心中，以尽可能小的硬件占用空间，实现尽可能多的 IO 操作。光比特的产品设计基于三大核心支柱：

第一是 NVMe over TCP 协议——这是我们多年前自主研发的协议，后捐赠给 Linux 社区及网络存储社区。如今，几乎所有存储厂商都在使用这一协议，而光比特是首个将该协议性能发挥到极致的软件厂商。

第二是目标端的智能闪存管理 (intelligent flash management)——它能延长存储数据所用 SSD（固态硬盘）的生命周期：通过减少硬件层面的垃圾回收 (garbage collection)

操作，同时提升读写性能，最终实现 SSD 寿命的延长。

第三是集群化 (clustering) —— 对于任何存储解决方案而言，高可用性 (resiliency) 都是数据中心的核心需求。光比特方案的一个关键优势是：客户可为每个卷 (volume) 或 Kubernetes 中的 PVC (持久卷声明：Persistent Volume Claim) 选择副本数量 (1 个、2 个或 3 个)。在后台，我们会自动管理这些数据副本的存储位置，确保光比特集群的存储容量得到最优利用。

回到 Kubernetes 场景：我们的许多客户都在运行各类 Kubernetes 发行版。我们与红帽 (Red Hat) 有深度合作，不少客户使用 OpenShift (红帽 OpenShift)，也有客户使用原生 Kubernetes (vanilla Kubernetes) 或其他发行版。我们开发了适用于所有 Kubernetes 发行版的 CSI (容器存储接口, Container Storage Interface) 驱动；同时也为 OpenStack (搭配 Cinder 组件)、VMware 等环境提供驱动。此外，对于追求极致性能的客户，我们还支持裸金属 (bare metal) 部署。

补充一点关于 Linux 社区的内容：NVMe over TCP 驱动已集成到 Linux 内核 4.10 及以上版本中，至今已多年。过去 5 年间，所有企业级 Linux 发行版都已内置该驱动 —— 客户在客户端无需额外安装任何组件。

最后一点：我们在超微 X14 平台 (搭载 AMD CPU) 上，基于 EDB CNPG (云原生 Postgres: Cloud Native PostgreSQL) 进行了测试。

测试结果如图所示：我们对 “3 个工作节点 + 3 个光比特目标端” 组成的集群进行压力测试，并不断增加 EDB 集群数量。数据显示，这个极小的集群 (仅 3 个工作节点、3 台运行光比特软件的服务器) 在 “只读模式” 下可实现近 100 万 TPS (每秒事务数)；在

“90% 读、10% 写” 的混合模式下，12 个 EDB 集群可实现约 70 万 TPS。这些数据非常可观 —— 这意味着客户可基于该平台灵活部署任意数量的 Postgres 数据库，甚至支持多客户共享同一环境。

主持人：

所以你们的测试验证了方案的可扩展性与效率，也展示了它在不同部署场景（如 OpenStack、VMware、各类 Kubernetes 发行版及 Linux）中的适配能力。我认为这种灵活性正是许多企业所追求的 —— 他们希望清晰了解方案的实际落地方式。刚才你提到了 Postgres，这正好引出 EDB 的话题。西蒙，能否为我们解读 EDB 在这一体系中的定位？Postgres 的优势想必十分突出。

EDB 代表：

EDB 是 PostgreSQL（开源数据库）社区代码的第一大贡献者，至今已有约 20 年历史。多年来，EDB 一直是 Postgres 的坚定倡导者。根据开发者调查，Postgres 是全球最受欢迎、也最受青睐的数据库。

EDB 的核心业务包括：提供 Postgres 的企业版（含多项增强功能）、为企业客户提供付费

支持，以及推出全新的 Postgres 数据与 AI 平台 —— 该平台具备众多创新功能，能实现数据库与 AI 的高效协同，让企业轻松构建 AI 应用。

我们的客户包括众多大型银行，如万事达卡 (Mastercard)、星展银行 (DBS)、蒙特利尔银行 (Bank of Montreal) 等。需要 EDB 提供 “关键业务级 Postgres 支持” 的客户数量正逐年大幅增长。EDB 支撑的关键业务系统涵盖多个领域：包括英国的电网、每天数十亿笔信用卡交易，以及超过 4 亿印度公民日常依赖的支付系统。这些都是绝对不能中断的关键业务系统。

今天我们齐聚于此，核心议题之一是 Postgres 与超微的合作 —— 这一合作解决了我们观察到的四大客户痛点：

关键业务级性能：确保硬件与软件的最优调优难度较大，而我们与超微合作，共同筛选出最优配置，使性能价值达到普通 Postgres 的约 6 倍。

极致高可用性：企业需要简单的解决方案 —— 确保即使数据中心的某个区域（如故障域或可用区）宕机，数据库仍能无缝运行。我们通过合作将这一需求转化为简单的解决方案。

简化部署与运维：我们希望打造 “开箱即用” 的方案 —— 无需企业全员掌握 Postgres 的每一个细节，也无需深入理解底层技术（如红帽 OpenShift，它是 Kubernetes 的一个发行版）。

AI 与数据的深度融合：不仅要简化 AI 部署，还要实现 AI 的完全主权化 (sovereign)，并推动 AI 落地生产。如今，AI 已不再是 “辅助客服人员” 的工具 —— 我们正进入 “关键业务级 AI” 时代，这类 AI 需要像时钟一样稳定运行，因此必须具备高可用性（这需要存储和硬件提供足够的容错能力），且需部署多个 AI 计算节点以实现故障转移。例如，用于应对自然灾害、实时欺诈检测的 AI 系统，已成为 “必需项”，需 24 小时不间断运行，

其重要性与 Postgres 相当。

关于超微与 EDB 合作的 “主权数据与 AI 工厂 (Sovereign Data and AI Factory)”，很多人会问 “它到底是什么” —— 本质上，它是一个 “硬件 + 软件” 集成栈：硬件来自超微（高性能硬件），软件来自 EDB 及其他管理组件，能让客户在自有数据中心内轻松部署关键业务级 Postgres 和 AI 工作负载（支持 3 个或更多节点以实现容错）。

具体落地流程如下：首先，EDB 与超微共同为客户在其数据中心内完成部署；随后进行演示与培训；之后客户即可快速展开工作 —— 只需几次点击，就能部署一个高可用的生产级 Postgres 集群；再点击几次，就能推出一个可自动创建支持工单的生成式 AI 聊天机器人。整个过程高效便捷，因为我们已提前完成所有组件的兼容性验证。

此外，客户普遍关心 “后续安全更新” 问题：我们提供全生命周期支持，包括安全更新、主版本更新 —— 不仅覆盖 Postgres 软件，还包括栈内所有组件（如你提到的 Kubernetes 发行版 OpenShift）。这意味着客户无需配备 24 小时待命的 Kubernetes 专家。

通过与超微的合作，以及 EDB AI 平台的优化，客户的数据库投资回报率 (ROI) 可提升约 6 倍 —— 这得益于超微的硬件能力，以及我们针对超微产品对 Postgres 和 EDB Postgres 产品的深度调优。性能方面，EDB Postgres 比其他数据库竞品快 150 倍，比第二大开源数据库 MySQL 快 4 倍，成本则比许多商业数据库低 6 倍。整体来看，客户能获得 90% 的价值提升。

超微与 EDB 的合作，本质是为客户简化流程：提供 “一站式” 系统，客户有任何问题（无论是硬件还是软件），只需联系一个对接方。我们的目标是让客户轻松运行关键业务级集群，从而专注于核心业务发展。

主持人：

这一切都很清晰 —— 各位通过合作共同打造了一个企业应用现代化平台，助力企业为 AI 落地做好准备。君霞，超微在帮助企业 “规模化采用 AI” 方面，有哪些具体的策略或举措？

超微代表：

AI 已从根本上改变了企业看待计算基础设施的方式。如今，各行业的企业都在快速推进 AI 的规模化部署与运营。我们观察到三大趋势：

AI 演进推动对 “优化硬件” 的高需求 —— 尤其是推理和实时应用场景，需要高性能的计算、网络和存储资源。

企业需要 “端到端的 AI 基础设施解决方案” —— 无论部署在本地、混合环境还是云上，都能简化部署流程。

随着集群规模扩大，功耗与冷却效率已成为关键考量因素 —— 功耗正成为企业的主要成本项之一。

超微具备应对这些 AI 驱动型挑战与需求的能力和专业知识：

首先，超微是一家工程驱动型企业 —— 从 AI 优化服务器到高性能存储服务器，再到液冷机架 (liquid racks)，所有产品均由我们自主设计和制造。

其次，我们提供前文提到的 “机架级即插即用解决方案”，为客户提供 “一站式” 服务，无需客户与多个供应商对接。

超微代表：

最后，我们在液冷技术领域的领先地位，不仅能应对现代 GPU 的高热量密度，还能确保高负载下的稳定性能。这些能力使我们能够提供 “计算、存储、电源管理一体化” 的即插即用解决方案，为高性能、可扩展的企业级 AI 应用提供支撑。

主持人：

没错，这对企业而言至关重要。我们昨天还在调侃，未来可能要开始计算 “每瓦代币数 (tokens per watt)” 了 —— 这或许会成为新的衡量标准。

当企业审视数据中心时，电力正成为越来越突出的约束因素，而散热与冷却则是推高电力消耗的重要原因。基于这一点，我们来问问希瓦：能源成本正如何影响现代数据中心？AMD 的能效优势能为解决这些问题提供哪些帮助？

AMD 代表：

这是个很好的问题。如你所说，电力和数据中心空间正成为现代数据中心（尤其是 AI 时代）的主要约束，而 AMD 与超微合作打造的平台，是业内能效最优的方案之一。

我之前提到过性能数据，但这些数据对客户实际意义是什么？本质上，它意味着客户能以“更小的资源占用（包括电力和数据中心空间）”实现“更多的业务价值”——这不仅能帮助客户满足能源预算要求，还能应对可能影响其利润的能源法规。如今电力成本持续上涨（尤其是在欧洲，涨幅已呈指数级），因此我们在设计产品时，始终将能效作为核心考量。我们在实验室做过一项研究：计算“现有服务器部署可被多少台 AMD 服务器替代”。结果显示，替代比例高达 7:1——这意味着采用 AMD 平台，客户能立即节省大量成本。此外，我们还发现：与现有方案相比，AMD 方案可降低高达 60% 的功耗，减少高达 87% 的服务器数量（却能实现相同性能），并最终降低 67% 的总拥有成本（TCO, Total Cost of Ownership）。

同时，服务器数量的减少还能为 AI 所需的 GPU 部署腾出空间——客户可将节省出的空间用于部署更多 GPU 节点。无论从哪个角度看，客户都能从中获益，这正是 AMD 在这类计算场景中的核心价值主张。

主持人：

核心就是 “在现有空间、电力和冷却资源下，实现更多价值”。另外，在当前环境下，“易用性” 也至关重要 —— 企业希望在获得高性能的同时，兼顾易用性与成本可控。西吉，从光比特的角度来看，你们如何在这一体系中帮助企业实现 “高性能、易用性、成本可控” 的平衡？

Lightbits 代表：

对我们和客户而言，易用性都是首要考量。光比特的解决方案可在几乎所有 Linux 发行版上安装，这为客户提供了极大的自由度 —— 客户可灵活地将更多服务器加入光比特集群，也能轻松完成操作系统或光比特软件的更新。这些操作均通过 Ansible（自动化工具）及其他自动化流程实现，无需复杂手动配置。

此外，软件定义存储（SDS）的本质意味着 “资源调整的灵活性” —— 它本质上是一组可安装的 RPM 包（Linux 软件包格式），安装后即可形成集群，后续的资源增减都非常便捷。

主持人：

这一点与 “DevOps (开发运维)” 和 “平台工程 (Platform Engineering)” 的理念高度契合 —— 这正是企业管理现代应用的目标方向。西蒙，从这个角度出发，你认为最大的机遇在哪里？尤其是考虑到 Postgres 的使用者涵盖数据科学家、开发者等不同角色，他们正构建下一代 “内置 AI” 的应用。

你如何看待 Postgres 在 “本地 AI 工厂 (on-premise AI factory)” 中为这些角色提供的支撑？

EDB 代表：

AI 领域确实涉及众多角色和专业分工：除了 AI 开发者，还有普通开发者（他们对底层硬件了解有限，更关注模型选择与优化）、数据科学家（与 AI 开发者的职责有所区分）。当前的一大挑战是：这类系统需要大量专业人才，而 AI 系统（尤其是高能耗的 AI 系统）成本高昂，因此必须确保每一个环节都经过优化且易于管理 —— 这正是 EDB 数据与 AI 平台（及与超微的合作）的价值所在：我们提供硬件支持，也配备专业团队确保系统持续维护，让不同角色无需分心于底层技术。

此外，AI 领域还出现了 “AI 硬件专家” 的新角色（可能来自软件或硬件背景）—— 正如我们在另一视频中讨论的（推荐大家观看），即使是存储层（光比特等厂商擅长的领域），也需要专家来优化 “模型存储方式” 和 “硬件使用效率”，这直接影响最终的 AI 服务交付质量。

核心目标是：让 AI 开发者和 AI 资源专注于“创造业务价值”，而非“补丁更新”或“学习底层技术”。在这一点上，Postgres 数据与 AI 平台的优势尤为突出：

首先，通过“主权数据与 AI 工厂”，我们确保客户能专注于上层业务，而由我们负责更新维护、提供现代化技术栈，并确保系统跨三个可用区运行（保障 Postgres 的高可用性）。

其次，我们通过 AI 工厂和模型构建工具（Model Builder）承担了大量“重活”——即使客户对 Postgres 不够熟悉，也能快速上手，利用已有专业技能构建交互式生成式 AI 应用（如聊天机器人），并将其用于项目协作或其他业务场景，满足企业的 AI 优化需求。

主持人：

很有道理。我们继续聚焦“AI 工厂”——这个概念对不同的人而言可能有不同含义。企业往往需要明确“起点、目标及长期演进路径”，并通过 API 或数据库直连等方式整合应用与应用数据。例如，企业可能从 RAG（检索增强生成）起步，随着数据整合逐步构建智能体（Agent）。

我想提出一个共性问题：人们常说 AI 工厂的初始投资较高，但从长期来看，它能提升效率、降低成本，最终实现总拥有成本（TCO）下降。各位认为，在 AI 工厂的建设过程中，哪些关键因素真正有助于降低 TCO？我们先从希瓦开始。

AMD 代表：

AI 工厂将运行各类 AI 工作负载，因此在基础设施投资决策中，“精准选型”至关重要。

GPU 固然是核心，但 CPU、存储（如磁盘）、网络等其他组件也需合理配置。通过“按需选型”（尤其是针对 AI 工厂模式），客户能大幅节省成本，快速实现投资回报（ROI）——这远比“盲目采购”更有效。因此，“精准匹配需求与基础设施”是降低 TCO 的关键。

主持人：

西吉，从光比特的角度来看，AI 工厂的 TCO 优化有哪些关键点？

Lightbits 代表：

光比特的核心优势在于性能，而我们深知企业为 AI 工厂的冷却和环境搭建投入巨大。因此，我们的方案设计始终围绕“极致硬件利用率”：经过长期优化，单个光比特目标端可实现约 450 万次 4K 读取（4K reads）——这意味着一个由 3 台服务器组成的小型集群，总读取性能可超过 1300 万次 / 秒。这正是客户需要的：以最小的硬件投入，获得最大的 IO 性

能和吞吐量。

主持人：

西蒙，从 EDB 的角度来看，当 Postgres 和 EDB Postgres AI 部署到 AI 工厂后，企业往往希望有一个 “数据重心”，并能灵活使用数据库内的各类功能。EDB 在这方面有哪些布局？

EDB 代表：

我认为核心是 “确保各组件协同工作”，而关键指标是 “客户从系统中获取价值的速度”—— 这通常意味着 “快速将所有组件落地生产”。这正是当前的最大挑战：构建实验性 AI 应用很容易，但对企业而言，必须有一套能 “快速落地生产并产生回报” 的系统。

在这一过程中，企业需要以简单的方式整合整个生态（如光比特的存储、所需的计算资源，并实现计算资源的弹性扩展），然后集中管理所有组件，清晰掌握数据状态、激活数据价值，最终完成部署。这正是 AI 工厂的目标，也是我们为客户创造价值的核心方式。

主持人：

君霞，作为超微的代表，你认为这些环节（基础设施、存储、数据库、AI 平台）应如何协同，共同支撑 AI 工厂的建设？

超微代表：

从超微的角度来看，正如我之前提到的，我们的整体解决方案已从单一系统扩展到机架、集群，乃至数据中心建设层面——这种灵活性使超微能提供业内最丰富的产品组合（包括存储服务）。客户可从单台服务器起步，也可部署满是存储服务器或 GPU 服务器（含必要网络组件）的机架，从而获得 AI 工厂所需的容量与性能。此外，我们还能根据企业的 IO 密集型 AI 训练、数据应用等需求，对硬件进行定制化调优。

主持人：

没错，这种“可按需扩展、灵活调整”的能力至关重要。最后一个问题，想请各位分享：对于正开启“企业应用 AI 现代化之旅”的企业，你们希望他们从今天的讨论中带走哪些

核心信息？我们先从西吉开始。

Lightbits 代表：

首先，归根结底，数据是 AI 的核心，而存储则是数据的基础。企业必须确保拥有一个 “高性能、可扩展、现代化” 的存储集群，以支撑 AI 的需求 —— 这一点至关重要。AI 场景下的吞吐量和数据量是 5 年前无法想象的，且仍在持续增长。因此，企业需认真评估自身的存储方案，并选择合适的供应商提供支持。

主持人：

接下来请西蒙分享。

EDB 代表：

好的。我认为企业客户在构建系统时，必须基于 “开放标准” —— 这是今天所有参会厂商的共识，也至关重要，因为技术始终在演进。其次，企业需 “以终为始”：思考如何快速将系统落地生产，而不仅仅是 “拥有一堆工具”。关键是：如何将这些工具部署到数据中心的 3 个节点上，并最终实现业务目标？这才是核心。

主持人：

希瓦，该你了。

AMD 代表：

我想建议企业 “保持耐心” ——AI 落地更像是一场火车旅行，而非超音速太空飞行。我们观察到，成功的客户往往具备耐心：若期望在一年内实现投资回报，大概率会失望。我的建议是：从试点项目起步，逐步将其扩展到更大规模，在这一过程中掌握 AI 部署的方法。企业无需一开始就大规模投入，可随着项目推进逐步招聘更多 AI 工程师或专家。AI 的生命周期很长，因此落地节奏应 “有机推进”。

主持人：

当然，“招聘到合适的 AI 人才” 本身就是另一个值得深入讨论的话题。君霞，你希望企业带走哪些信息？

超微代表：

AI 已改变了企业看待计算基础设施的方式，但大多数企业并不具备部署复杂基础设施（尤其是 GPU 服务器机架、网络等）的能力。而超微的机架级即插即用解决方案，既能简化部署复杂度，又能让企业根据业务增长需求，灵活扩展数据基础设施的规模。

主持人：

非常好，这是一个完美的收尾。感谢各位的参与！本次讨论非常精彩，其中包含许多企业可借鉴的实用信息，希望能助力他们在“企业应用 AI 现代化之旅”中稳步前行。再次感谢各位！