

大摩闭门会：下一步关键何在：GOOGL、AMZN 和 META 的 2027 年算力规划及后续关注要点 20260526

2027 年超大规模数据中心产能展望与 AI 芯片成本性能分析

摘要

- 2027 年四大超大规模数据中心 (Google/AWS/Azure/Meta) 合计新增产能将达 20GW，高于 2026 年的 14GW，两年累计新增约 34-35GW。
- Alphabet 产能扩张领先，预计未来两年新增 13GW；AWS 与 Azure 两年合计各增 8-10GW；Meta 2027 年预计新增 3-4GW。
- GB300 系统每 GW 成本约 330 亿美元，显著高于 TPU 系统的 250 亿美元；NVIDIA 芯片每瓦性能是定制芯片的 2-8 倍，溢价核心在于推理阶段的低代币成本。
- 亚马逊 2026 年资本支出约 50% 用于电力、机房等前瞻性建设；谷歌仅为 10%，但面对 4,600 亿美元未交付订单，预计 2027 年需大幅增支扩产。
- Trainium3/4 与 TPU 追赶 NVIDIA 的关键在于系统级优化，包括 NVLink 同类技术、内存带宽及针对 EFS 等服务的软件优化，而非单一芯片性能。
- Meta 利用 AI 助手贯通全漏斗广告，其增量广告支出回报率 (ROAS) 达搜索平台 2.3 倍，通过网红生态解决 AI 购物信任问题，驱动商业变现。

Q&A

你们是如何自下而上地构建模型，以估算超大规模数据中心运营商到 2027 年的资本支出和千兆瓦级产能目标的？

我们联合了半导体、网络与电源团队，采用自下而上的方法构建模型。该模型基于超大规模数据中心的服务器和机架交付量，涵盖了视频处理和定制 ASIC 领域。我们首先分析服务器、机架及非服务器机架部分的需求，并考虑这些设备的 IT 电力需求，最终构建出数据中心的建设模型，从而估算出各公司的千兆瓦产能目标。

在构建成本模型时，构成 Trainium、TPU v8 以及 Blackwell 系统成本的关键组件有哪些，它们之间存在怎样的差异？

主要拆解了几个核心部分来构建成本模型。首先是机柜成本，其中包含了 GPU 芯片成本，这部分根据采用 NVIDIA、TPU 还是 Trainium 的方案而有所不同。其次，我们还分别拆分了 HBM、DRAM 的成本以及 CPU 的定价。尽管超大规模厂商通常从 NVIDIA 采购完整的机柜系统而非单个组件，但鉴于当前成本的敏感性和持续上涨，我们选择将其拆分分析。除机架本身，模型还纳入了各用电源、冷却设备、电源外壳和网络收发器等非机架组件的成本。通过这种跨部门协作的自下而上成本构建，我们得出 GB300 系统每千兆瓦的成本约为 330 亿美元，而 TPU 系统的成本约为 250 亿美元。

考虑到 NVIDIA 芯片的成本远高于定制芯片，超大规模数据中心为何仍愿意为其支付溢价，其性能优势具体体现在哪些方面？

超大规模数据中心愿意为 NVIDIA 芯片支付更高的每瓦成本，核心原因在于其卓越的性能。数据显示，NVIDIA 芯片的每瓦性能是部分定制芯片的 2 到 8 倍。这种性能优势不仅能提升整体系统表现，还能有效降低代币成本，这一点随着更多工作负载从训练转向推理而变得愈 重要。

亚马逊的 Trainium 和 Alphabet 的 TPU 在未来几年内，有哪些机会可以缩小与 NVIDIA 在每瓦性能上的差距？

缩小与 NVIDIA 的差距，关键在于提升整体系统级性能，而非仅仅关注芯片本身。无论是预计在 2026 年底推出的 Trainium 3，还是 2027 年的 Trainium 4，其改进重点都应放在系统层面。这包括网络性能、内存带宽，以及在 AWS 内部针对 EFS 等服务的软件优化。能否有效利用类似 NVIDIA NVLink 的技术也将是关键。因此，无论是 Trainium 还是 TPU，系统级的整体改进是追赶差距的核心所在。

预计到 2027 年，谷歌、亚马逊、微软和 Meta 这四家超大规模数据中心运营商各自将新增多少产能，与过去相比有何变化？

预计在 2027 年，谷歌、亚马逊、微软和 Meta 四家合计将新增 20 吉瓦的产能，高于 2026 年新增的 14 吉瓦。这意味着未来两年内，这些公司总计将新增约 34 至 35 吉瓦的产能。其中，Alphabet（谷歌）通过部署 NVIDIA 和 TPU 芯片，贡献最大，预计未来两年将新增约 13 吉瓦。AWS 和 Azure 在 2026 年各自新增约 5 吉瓦，预计 2027 年将分别新增 4 至 5 吉瓦，两年合计新增产能在 8 至 10 吉瓦之间。Meta 在 2026 年将新增数吉瓦产能，预计 2027 年将达到 3 至 4 吉瓦。

通过分析各公司 2026 年的资本支出与计划落地的青瓦产能，你们在“前瞻性建设”方面有哪些发现，特别是亚马逊和谷歌的策略有何不同？

的分析显示，亚马逊 2026 年资本支出中约有一半用于前瞻性建设，主要投入于机房、电力等非芯片基础设施，为 2027 年交付的芯片做准备。相比之下，谷歌 2026 年资本支出中仅有 10% 用于前瞻性建设，大部分支出用于当年即可上线的产能。这可能与芯片交付时间安排有关，或许谷歌能更好地将芯片交付与新增电力供应同步，也可能是在利用旧有数据中心。然而，这也意味着谷歌的前瞻性建设规模不如其他厂商，考虑到其高达 4,600 亿美元的云业务未交付订单，我们预计其在 2027 年需要增加资本支出以扩充产能来满足积压的需求。

Meta 近期发布的关于其搜索和 AI 智能体愿景的演示文稿揭示了怎样的产品和广告策略，这将如何利用其新增的大量 AI 产能？

Meta 的演示文稿指出，搜索正演变成为一种更具社交属性的行为，公司正积极争取这部分广告预算。其策略是通过一个覆盖全漏斗的应用家族框架来实现：漏斗顶部由 Reels 和动态发现内容构成，中部是用户的评论和收藏行为，而底部则是通过 Meta AI 实现的购物活动。数据显示，Meta 应用家族的增量广告支出回报率是搜索平台的 2.3 倍。一个关键的表现是，约半数用户已使用 AI 助手进行产品发现，Meta 正利用其新模型抓住这一机遇，让用户在 Meta AI 上完成整个购买流程。凭借 Facebook 和 Instagram 的强大分发能力，Meta 有能力快速推动这类新产品的普及。

Meta 如何利用其平台上的内容目录和网红生态系统，构建独特的价值主张以推动增量互动和商业变现？

Meta 的独特价值主张在于其庞大的内容生态。平台拥有广告商提供的数十亿商品目录，以及围绕网红的个性化内容。调研显示，消费者不信任机器是阻碍其通过 AI 代理购物的主要原因之一。Meta 可以通过引入用户所熟悉并产生共鸣的网红来解决这一信任问题。例如，当用户想购买降噪耳机时，Meta 可以精准地展示相关领域的网红内容来辅助购买决策。这种方式能够有效推动增量的用户互动和商业变现。目前，市场对 Meta 的估值似乎并未反映其 AI 相关资本支出的任何潜在投资回报。

