

FUNDA REPORT

Deep|CRWV: 容量扩张，每GW收入仍有上行空间

Channel checks point to over 2GW by 2026 and stronger 2027-28 demand, with pricing and mix adding upside.

摘要

我们首次覆盖CRWV，核心看多来自AI数据中心容量扩张与大客户长约兑现。渠道调研显示2026年底Active Power有望超过2GW，高于管理层>1.85GW指引；我们预测2026-2028年底为2.10/3.60/5.40GW，每GW年化收入升至105/120/130亿美元。2027-2028年需求可见性较高、take-or-pay提供下行缓冲，但对每GW 300-400亿美元高价情景保持谨慎。

目录

- 一、摘要
- 二、为何是现在
- 三、核心争议与我们的观点
- 四、业务概览与详细分析
- 五、Funda 预测与市场一致预期
- 六、风险

一、摘要

我们首次覆盖CoreWeave（CRWV）。公司主要向客户提供用于 AI 训练和推理的 GPU 集群，以大客户长约为基础融资建设，并在集群通过验收后开始确认收入。结合多轮渠道调研，我们对 CRWV 未来几年的收入增长保持建设性的观点。

- 结合渠道调研，我们认为 2026 年末 CRWV 的 Active Power 有望达到 2GW 以上；对比管理层在二季度业绩会上给出的指引为超过 1.85GW。对于 2027 年和 2028 年的新增容量，渠道反馈也让我们保持较为积极的预期。
- 渠道反馈显示，大型客户已在安排 2028 年的外部算力采购，部分客户开始讨论 2029 年的容量。在公司按约交付的情况下，大客户合同中的take-or-pay安排也能为收入提供一定下行缓冲。
- 我们对每 GW 每年300-400亿美元的高收入情景保持谨慎，但新签合同和到期续约有望逐步带动单位容量收入提高。大客户换用新一代GPU，或旧卡转向短约和 GPU 云服务时，也有重新议价的机会。Managed Inference 及更多通用云服务的额外贡献，暂未单独计入预测，但未来可能带来上行空间。

二、为何是现在

市场对 CRWV 的分歧，集中在容量能否按时交付、AI 需求能否延续，以及每 GW 收入能提高到什么水平。公司需要先投入大量资金，再通过集群交付和客户付款收回成本。如果项目延期，收入确认就会推迟，部分租金和融资支出仍会发生。

根据我们的渠道调研，CRWV的交付情况较前期有所好转。随着下半年新增集群开始计费，较高的新签合同价格有望逐步体现在收入中。大型客户已开始安排 2028 年的外部算力采购，这让我们对 CRWV 未来几年的需求更有信心。

三、核心争议与我们的观点

1. CRWV能否按时交付新增容量？

a. 市场争议

市场对数据中心交付瓶颈的讨论仍在持续，对社区和政府反对新建项目的担忧也在增加。对于CRWV，新增容量能否按期上线，直接影响未来几年的收入增长。若项目延期，CRWV的收入确认会相应推迟，部分租金和融资成本仍需承担。

b. 我们的观点

结合渠道调研，我们认为 2026 年末 Active Power 有望达到 2GW 以上；对比管理层在二季度业绩会上给出的指引为超过 1.85GW 。

c. 判断依据

根据我们的渠道调研（9月上旬），CRWV的交付情况较前期有所好转，8月以来未见新的重大延期。我们在渠道调研中了解到两组容量估计。一组侧重整体部署节奏，另一组按主要站点拆分，给出的预测更详细，也更乐观。后一组对 2027 年部分新一代 GPU 部署量的估计，与硬件调研了解到的采购需求在数量级上大致相符。机房通电后，还需完成 GPU 安装、调试和客户验收，才会开始计费。渠道估计，一般来说当季新增 Active Power 对应的集群约有一半可在当季开始计费，其余多在下一季度开始计费。若容量集中在季末上线，当季收入贡献有限，部分租金和调试费用却已发生，利润率及按期末 Active Power 折算的单位容量收入可能暂时偏低。

2. CRWV在 2027 年和 2028 年是否仍有足够的业务可见性？

a. 市场争议

市场一直在争论 AI 需求的持续性。对于 CRWV，投资者关心的是，2027 年和 2028 年是否仍有足够的业务可见性，还是一门只受益于当前算力紧缺的短期生意。

b. 我们的观点

根据渠道调研，2027 年较有吸引力的大规模外部算力，大部分已获得客户承诺或进入谈判，2028 年也有很大一部分容量进入讨论。这些远期采购安排，让我们对 CRWV 未来几年的需求更有信心。

CRWV 的大客户合同通常采用 take-or-pay 结构。在公司按约交付的情况下，客户即使没有充分使用已承诺的容量，仍需按合同付款。我们认为，这类长约为未来几年的收入提供了较好的可见性，也能在需求下行时提供一定缓冲。

c. 判断依据

渠道反馈显示，大型客户仍在提前约 30 个月安排外部 AI 算力采购。在该渠道关注的 2027 年具备吸引力的大容量区块中，约 75%-90% 已获得客户承诺或进入谈判；2028 年的外部容量需求仍然较高，可见性约为 60%。采购团队正在讨论 2027 年第二季度至 2028 年的交付，部分大型客户已开始规划 2029 年的电力区块。

大客户合同通常为 3-5 年，部分战略项目更长。在我们了解的合同安排中，客户主动取消已承诺订单，可能需要支付剩余合同款项或承担赔偿责任，较高的退出成本使其主动违约的动力相对有限。部分合同另有可选容量，待客户确认后才转为正式采购承诺。

3. CRWV 每 GW 每年能够产生多少收入？

a. 市场争议

近来市场一直在讨论 CRWV 每 GW 可以产生的收入。我们听到了各种各样的数字，最高达到每 GW 每年 300-400 亿美元。

b. 我们的观点

我们对每 GW 每年 300-400 亿美元这个口径保持谨慎，但依然对未来 CRWV 每 GW 的收入增长保持建设性的观点。现有长约通常按原合同价格计费，涨价主要通过新签合同和到期续约逐步体现。我们的预测未采用上述高价情景，Managed Inference 及更多通用云服务的额外贡献也未单独计入。

c. 判断依据

i. 新签合同和到期续约逐步反映更高价格

根据渠道反馈，新签合同报价上调后，较高价格有望随着新增集群交付和旧卡续约逐步体现在收入中。今年部分 H100 合同进入续约期，2025 年的低价合同也将陆续到期，早期低价合同对应的收入占比预计逐步下降。

ii. 新一代 GPU 交付，以及大客户换卡时的重新议价

结合渠道反馈，我们预计新增集群中的 GB200、GB300 占比将提高，后续 Rubin 交付也有望带动单位容量收入上升。大客户换用新一代 GPU 时，需要与 CRWV 重新协商；公司通常会争取提高单价、延长合同期限或增加最低采购承诺，换卡后的合同条件往往更有利于 CRWV。

iii. 旧卡转向短约和 GPU 云服务，提供重新定价的机会

大客户完成迁移后，释放的旧卡可以继续出租给 AI 原生公司、中小型模型实验室和企业客户。这些客户采购的规模较小、合同期限较短。对同一型号的 GPU 而言，短约和 GPU 云服务通常对应更高的单位报价。不过，合同到期后的续租情况和实际利用率，仍会影响设备最终能够带来的收入。

iv. Managed Inference和更多配套服务带来额外收入

按 Token 收费时，CRWV承担更多模型服务和资源调度工作，折算到 GPU 小时的收入可能更高。但实际收入取决于客户调用量，资源池利用率较低时，闲置算力会抵消部分高报价带来的收益。存储、软件及通用云服务长期也可能带来增量。相关收入目前尚难单独预测，我们暂未额外建模。

我们了解到的部分高价订单，可能来自第三、第四梯队 AI 实验室等中小客户。这些客户愿意为紧缺算力支付更高价格，但能否长期维持采购仍取决于后续融资。这也是我们未将上述高价情景纳入基础预测的原因。

四、业务概览与详细分析

1. CRWV的基础业务与需求来源

a. 商业模式

CRWV主要向客户提供可直接用于 AI 训练和推理的完整 GPU 集群，服务范围涵盖 GPU 算力、电力、液冷、高性能网络、存储、调度软件和集群运维。公司提前锁定数据中心和电力资源，采购并部署 GPU 及配套设备，再将完成调试和验收的集群交付给客户。

面向大型客户的核心合同通常采用3-5年的照付不议（take-or-pay）模式。客户提前锁定特定 GPU 及容量，CRWV再以客户合同和 GPU 资产为基础获得融资，用于采购设备及建设或租赁数据中心。项目完成通电、装机、集群调试及客户验收后开始确认收入。CRWV 的收入主要来自集群租赁和相关服务，主要支出包括 GPU 及配套设备、数据中心租赁、电力、融资和运维支出。

CRWV主要的商业模式可以概括为“先锁定需求，再融资建设，最后交付并确认收入”。长约提高了未来收入的确定性，也使公司能够通过债务融资快速扩大资产规模；相应地，容量能否按时投产直接决定收入和现金流的兑现节奏。如果项目延期，收入确认会相应推迟，部分融资及场地租赁支出仍会继续发生。

b. 客户为什么需要CRWV

根据渠道调研，客户选择CRWV，主要看重能否更早获得可用算力。

i. 算力瓶颈正从 GPU 供应转向可用的数据中心容量。

2025年以前，算力扩张的主要制约来自 GPU 供应；进入2026-2028年，瓶颈逐步转向已经通电，并同时具备液冷、网络和运维条件的数据中心容量。大型客户需要的往往是100-300MW级别的完整集群，仅仅采购到 GPU 并不能直接形成可用于训练和推理的算力。芯片到位后，客户仍然需要解决电力、冷却、高性能网络、存储、集群部署及后续运维等问题。

'Up until 2025, the shortage was GPUs. But from 2026 to 2028, the shortage is energized AI data center capacity. Nvidia is shipping more GPUs than two years ago, but finding 100 to 300 MW of ready power with liquid cooling, networking and operational staff is becoming the harder problem.'

对大型云厂商而言，内部自建和外部租赁会长期并行。自建数据中心有利于控制长期成本，也便于统一管理基础设施；但从锁定电力、完成建设到集群通过验收，需要较长的建设周期。实际 AI 需求可能在此期间继续上修，导致原有建设计划无法满足新增需求。外部租赁可以补足内部容量与实际需求之间的缺口，并降低单个自建项目延期对整体算力计划的影响。

大型训练集群如果推迟数月交付，影响会继续传导至模型训练、产品迭代和商业化进度。相较于节约部分单位算力

成本，客户通常更加重视能否及时取得一整块可以直接投入使用的容量。因此，即使 CRWV 的单位价格高于大型云厂商的内部成本，客户仍然可能选择外部租赁。

'The largest hyperscalers are not primarily leasing because it is cheaper; they are leasing because time is more valuable than capital.'

1. CRWV以更高的前置风险换取更早的容量。

CRWV能够更早交付，部分原因在于其愿意更早承担场地、电力和长期租赁承诺。大型云厂商同样具备锁定这些资源的技术和资金能力，但其投资纪律和内部审批更加严格，通常需要在具体工作负载、业务归属及预算更加明确后，再作出长期资本承诺。CRWV 的决策链条相对更短，也更愿意在最终客户需求尚未完全确定时，提前与场地开发商和数据中心运营商锁定容量。

以我们在渠道调研中了解到的一个项目为例，该项目计划于 2027 年投产，CRWV在 2025 年便锁定相关场地并承担长期付款义务；当时下游大客户尚未确定相关 GPU 将具体用于哪个模型或业务部门。

CRWV提前锁定的电力和场地，预计能带来约2-3年的交付时间差。电力通常需要提前数年预订，需求出现后再行锁定，时间上往往已经来不及。在算力供不应求的阶段，更早锁定容量使 CRWV 能够比其他供应商更早交付，但这种领先并非永久存在。

'This existing portfolio creates a 2 to 3 year commercial advantage because you cannot retroactively create power reservations made several years earlier. This is not permanent.'

相应地，CRWV也承担了更高的前置风险。如果实际需求落地慢于预期，公司仍需承担已经锁定的场地租赁、设备采购及相关融资支出。

1. CRWV有大规模 GPU 集群的软件与运维能力

CRWV的 SUNK 将大模型训练常用的 Slurm 与 Kubernetes 结合，方便客户沿用原有训练流程部署大型 GPU 集群；集群监测和故障处置则由配套的运维系统支持。结合渠道反馈，客户采购时首先看重能否及时获得性能稳定的大集群，软件的作用是让交付后的集群更容易投入训练。

'The customers are not signing CoreWeave because they want Kubernetes, they're signing because they want a large GPU cluster quickly that performs well. And CoreWeave's Kubernetes-based stack helps them make that cluster usable.'

Kubernetes本身是开源软件，单项功能也可以由大型云厂商开发。更难在短期内复制的是长期运行大规模 GPU 集群积累的故障记录和处理经验：例如识别可能出问题的节点，及时将其移出训练任务，减少大型训练作业中途被打断的情况。结合渠道反馈，GPU优化软件的优势预计在 2029 年前仍较明显，此后可能减弱；长期运行大规模集群积累的运维能力，到 2030 年以后仍可能保持优势。

'The individual features can be copied, but replicating the complete operating system around them is different.'

'The GPU optimized Kubernetes software will be strong until 2029 in which it will start to become moderate. ... the AI cluster operations will stay strong even beyond 2030.'

1. 需求可见性已经延伸至 2028 年

渠道反馈显示，大型客户仍在提前约 30 个月安排外部 AI 算力采购。在该渠道关注的 2027 年具备吸引力的大容量区块中，约 75%–90% 已获得客户承诺或进入谈判；2028 年的外部容量需求仍然较高，可见性约为 60%。采购团队正在讨论 2027 年第二季度至 2028 年的交付，部分大型客户已开始规划 2029 年的电力区块。

'Hyperscalers are still securing external AI capacity 30 months forward. For 2027, 75% to 90% of attractive large blocks are committed or under negotiation. For 2028, demand for external capacity is still high, with around 60% visibility. Procurement teams are discussing Q2 2027 through 2028 delivery right now, and the biggest customers are already discussing 2029 power blocks.'

渠道反馈中的采购安排已延伸至 2028–2029 年，并涉及具体交付时间和电力区块。我们认为，这些远期采购需求有助于支持 CRWV 后续的容量扩张。

1. 现有算力租赁业务

a. 大客户长约构成现有业务基础

CRWV 现有算力租赁业务仍以大型云厂商和头部模型公司为主。这类客户通常采购大规模专属裸金属集群，由 CRWV 负责 GPU

、网络、存储及底层集群运维，客户在其上部署自己的云平台、训练框架和软件体系。

大型云厂商会将采购的容量纳入自身算力池。结合渠道调研，我们推测，其中约 25% 用于前沿模型训练和后训练，约 25% 用于自有 AI 产品推理，约 20% 用于向外部客户提供 GPU 云服务和托管

AI 服务；其余主要用于模型研发与评测、企业专属部署、多模态及视频生成等工作负载。大型云厂商也会根据自建数据中心的交付进度，在自建容量和外部容量之间调度负载。

大客户通常先与 CRWV 签署主服务协议，约定安全、服务水平和违约责任等条款，再通过具体订单确定 GPU 型号、集群规模、交付时间、价格及验收条件。合同期限一般为 3–5 年，部分战略项目更长。

已承诺的容量通常采用照付不议（take-or-pay）安排，客户即使没有充分使用，仍需按约定金额付款。部分合同另有可选容量，待客户确认后才转为正式采购承诺。

结合渠道反馈，大客户合同通常先约定一定金额的算力采购，GPU 数量再根据具体型号和价格确定。对于尚未交付的容量，客户调整集群配置或换用新卡时，需要与 CRWV 重新协商；在采购金额不变的情况下，单价提高意味着对应的 GPU 数量减少。已经投入运行的集群仍按原合同计费。

2024–2025 年签订的部分项目仍在按原合同逐批交付，2026 年以来的新签订单和部分到期续签订单则开始反映更高的报价。随着这些容量完成验收并开始计费，以及新一代 GPU 在新增集群中的占比提高，我们预计单位容量收入逐步上升。

在我们了解的部分大型项目中，大客户预付款比例约为合同金额的 40%–60%。这部分预付款用于数据中心基础设施和 GPU 采购。集群投入使用后，客户仍按合同约定支付算力费用，预付款不作为后续账单的充值余额逐步抵扣。在项目建设阶段，这部分资金可以减少 CRWV 需要补充的外部融资。

客户预付款比例还会随合同期限、客户信用状况和基础设施的定制程度而变化。预付款较低的项目，可能通过最低采购承诺、分阶段付款和担保条款，保障 CRWV 前期建设与设备投入的回款。

CRWV 拿到客户合同后，可以结合合同现金流、GPU 资产和相关机房权益安排项目融资，再完成设备采购和集群建设。集群交付后，客户付款逐步流入，用于覆盖运营支出、后续设备投入及相关债务本息。大型客户合同多为 3–5 年，部分项目的融资期限和 GPU 折旧期长于首份客户合同。合同到期时，如果 GPU 仍可使用，CRWV 需要通过续约或重新出租维持后续现金流。

大客户项目通常分批交付。同一份长期合同下，不同集群可能在不同季度上线，每批容量完成部署并通过客户验收后才开始计费。长期合同提前锁定了客户需求，但具体收入确认时间仍取决于数据中心通电、GPU 安装、集群调试和客户验收的进度。

结合行业调研，我们认为在 CRWV 按约交付的情况下，大客户主动取消已承诺订单的损失可能不止于预付款，还可能需支付剩余合同款项或承担赔偿责任。较高的退出成本，使大客户违约的动力相对有限。

b. 大客户换用新卡与旧卡再利用

大客户在合同期内可能会经历 GPU 换代。新一代 GPU 推出后，客户可以要求调整尚未交付的 GPU 型号，也可以将已经运行的工作负载迁移到新集群。渠道估计，约一半的大客户合同包含 GPU 更新或迁移安排。

对于尚未安装的容量，双方可以在交付前调整订单中的 GPU 型号。已经投入运行的集群则需要先准备好新集群，再将客户负载迁移过去。新一代 GPU 的功率、液冷和网络要求通常更高，新集群可能部署在其他数据中心，原有 GPU 则继续留在原来的机房。

客户如果希望换用新一代GPU，需要与 CRWV 重新协商。在这类谈判中，CRWV通常会要求提高单价、延长合同期限、增加最低采购承诺，或者由客户继续保留一部分旧卡负载。结合渠道反馈，换卡后的合同条件通常更有利于CRWV；客户对新卡需求较强时，单价或整体采购承诺往往也会随之提高。

原客户完成迁移前，旧集群仍由其占用，CRWV无法重新出租。迁移完成后，释放的旧卡可以转向推理、企业及中小型 AI 客户。

c. 中小客户承接旧卡，合同更短但单位价格更高

不同代际的 GPU 可以承接不同的工作负载。最新一代 GPU 通常用于前沿模型训练，H200和 B200 也可承担大规模推理；H100和 A100 仍可用于模型微调及中小模型推理。

大客户释放的旧卡可以继续出租给 AI 原生公司、中小型模型实验室和企业客户。这些客户需要的 GPU 数量较少，合同期限也更短。部分客户没有自己的调度和运维体系，因此会采购 CRWV 的 GPU 云服务，而不只是裸金属集群。

在我们了解的部分项目中，中小客户合同约为1-3年，预付款约为合同金额的10%-20%，部分合同还约定最低使用量。相比大客户长约，这类合同规模较小、期限较短，客户也可以较快获得容量。

短约和 GPU 云服务通常对应更高的单位报价。以我们了解的 GB200 报价为例，大客户长期裸金属合同约为3.6-4.0美元/GPU小时，一项 GPU 云服务报价约为 6 美元/GPU小时。旧卡也可以采用短约或 GPU 云服务的形式重新出租，具体价格仍取决于型号和客户需求。

旧卡重新出租后，原有设备可以继续产生现金流，相关合同和预付款也可以用于支持新集群融资。不过，GPU的会计折旧期约为 6

年，而中小客户合同通常只有1-3年。首轮短约到期后，CRWV仍需继续寻找客户，旧卡的续租情况和实际利用率会影响设备最终能够带来的收入。

除传统算力租赁外，CRWV还可以将部分 GPU 投入托管推理业务，按 token 向客户提供推理服务。

表：渠道反馈的部分合同安排对比

条款	大客户长约	中小客户及旧卡租赁
客户类型	大型云厂商及头部模型公司	AI原生公司、中小型模型实验室及企业客户
服务形式	大规模专属裸金属集群	GPU云服务、按需租赁或较小规模专属集群
合同期限	通常为3-5年，部分战略项目更长	通常为1-3年，其中以2-3年较为常见
预付款	部分大型项目约为合同金额的40%-60%	部分项目约为合同金额的10%-20%
采购承诺	已承诺容量通常采用照付不议安排，部分合同另含可选容量	部分合同约定最低使用量

GPU安排	具体订单约定 GPU 型号、容量和交付时间；尚未交付的容量可以在交付前调整 GPU 型号	主要承接大客户释放的上一代GPU，也可以租用较小规模的新卡容量
价格	一个 GB200 报价案例中，长期裸金属约为 3.6–4.0 美元/GPU 小时	同一 GB200 报价案例中，按需或 GPU 云服务报价可以接近 6 美元/GPU 小时；旧卡按具体型号定价

来源：渠道调研，Funda整理

1. Managed Inference：从 GPU 租赁延伸至 Token 收费

a. CRWV开始向推理服务延伸

CRWV正从 GPU 租赁延伸至模型服务，逐步接手模型部署、流量路由、扩缩容和集群运维。客户仍可以通过 CKS 自行管理推理服务，也可以将更多工作交给CRWV，并在 Dedicated Inference 与 Serverless Inference 之间选择。

产品	客户与 CRWV 的分工	计费方式	主要适用场景
CKS	CRWV提供 GPU 集群及底层基础设施，客户自行部署和管理推理框架	按 GPU 小时收费	具备较强工程能力、需要完全控制软件栈的客户
Dedicated Inference	客户选择模型、GPU 及运行环境，CRWV负责部署、路由、扩缩容和日常运维	按 GPU 小时收费	负载较稳定、对性能和定制要求较高的生产场景
Serverless Inference	客户通过 API 调用模型，底层资源池及模型服务由 CRWV 统一管理	按 Token 收费	模型试用、短时突发及需求波动较大的场景

来源：公开数据，渠道调研，Funda整理

三类产品中，CRWV承担的运营工作有所不同。CKS 仍然接近传统 GPU 租赁，客户自己管理模型服务和 GPU - 利用率；Dedicated Inference增加了模型服务和集群运维，但容量仍由单一客户使用；Serverless Inference则由 CRWV 统一管理资源池，客户只按照实际使用的 Token 付费。CRWV 承担的服务环节越多，单位算力对应的收入空间越高，同时也需要自行管理资源池利用率。

公司将 Serverless 和 Dedicated 共同纳入 Managed Inference 。2026 年第二季度，Managed Inference的已签约年化经常性收入已经超过 1 亿美元，公司预计 2026 年末至少达到 2.5 亿美元。相对于公司现有收入规模，这项业务目前仍处于早期。

1. Serverless放量取决于共享资源池利用率

按 Token 收费时，折算的单位 GPU 收入可能高于长约裸金属，但

CRWV 需要提前准备算力、应对突发调用，而收入只在客户实际使用时产生。如果请求不足，或高峰集中在相同时段，GPU在其余时间仍会闲置，较高报价带来的收益也会被低利用率部分抵消。

CRWV公开资料按照客户负载的稳定程度举例：对于客户来说，模型探索阶段的持续利用率通常只有10%-20%，按 Token 付费较合适；波动较大的生产负载约为30%-50%，需要结合具体使用情况比较两种方式；持续利用率达到70%-90%时，采用 Dedicated Inference 并按 GPU 小时付费通常更经济。

结合渠道反馈，纯按需集群若只能用到约六成的容量，闲置 GPU 仍会抵消部分高报价带来的收益。现阶段，- Managed Inference预计仍以 Dedicated 及带有最低消费或预留容量的 Token 合同为主；纯按需 Serverless 则先承接模型试用、短时突发等使用量不稳定的需求。

不同客户的调用时间如果能够错开，CRWV便可以用共享资源池承接更多请求，减少 GPU 闲置。负载稳定的客户通常会选择 Dedicated 以控制成本；对于留在 Serverless 的流量，客户和模型种类越多，空闲时段越有机会被其他调用填上。但如果新增请求集中在相同高峰，CRWV仍需为高峰保留容量。

开源模型应用增多，有望为资源池带来更多不同来源的调用。批处理、KV Cache 和推理引擎优化则主要提高单张 GPU 处理 Token 的效率。

2. 从裸金属到 Token 收费，单位 GPU 收入逐级提高

结合近期渠道反馈，以 GB200 为例，不同产品形态的报价大致如下：

产品形态	价格口径	说明
长约裸金属算力	约3.6–4.0美元/GPU小时	客户提前锁定容量，并自行管理主要软件及利用率
按需 GPU 云	约 6 美元/GPU小时	合同更加灵活，CRWV承担更多容量准备和即时交付责任
带预留或最低消费的 Token 服务	约 12 美元/等效 GPU 小时	客户按 Token 付费，CRWV负责模型服务及资源调度
高价值按需 Token 场景	约20–40美元/等效 GPU 小时	来自部分渠道较为乐观的反馈，不作为基础情景

来源：渠道调研，Funda整理

从裸金属长约转向 GPU 云服务，客户获得了更灵活的容量，报价也相应提高。按 Token 收费时，CRWV进一步负责模型服务和流量调度，折算到 GPU 小时的收入可能更高；与此同时，公司需要提前准备算力，实际收入会受到客户调用量和资源池利用率影响。

渠道提到的一类 GB200 预留 Token 合同，折算后约为 12 美元/等效 GPU 小时；部分 GB200 推理场景有机会达到 20 美元，也有渠道对未来提出30–40美元的更乐观预期。即便是 20 美元，也更接近较好的使用情形，未必能代表 Managed Inference 整体：

‘\$20 per GB200 equivalent GPU hour of gross inference revenue is possible, but don’t use it as a base case for an entire managed inference fleet.’

1. GPU周边产品已有收入基础，通用云功能仍是远期机会

除 Managed Inference 外，CRWV已经提供存储、CPU 、网络 and 软件等配套产品。公司在 2026 年第二季度披露，GPU之外的 AI 产品与服务 ARR 已经超过 4 亿美元。

渠道估计，GPU计算仍占客户在 CRWV 上支出的95%以上。存储、网络 and 软件等产品主要围绕 GPU 集群一并销售，相关收入规模仍明显小于算力业务。

根据渠道反馈，CRWV在技术上具备继续扩展数据平台、数据库等企业云功能的能力，可以自行开发，也可以集成第三方产品。但这些产品目前并非公司的战略重点。传统云服务涉及的产品较为分散，每增加一项功能都需要投入产品研发、客户适配、销售支持和后续运维，开发周期较长，早期能够带来的收入也很难与大型 GPU 合同相比。相较之下，GPU容量交付和 Managed Inference 更容易形成规模较大的订单，我们预计公司现阶段仍会优先将资源投入这两类业务。

随着客户在 CRWV 上运行更多 AI 应用，存储及配套软件的采购可能随之增加。公司后续也可能增加自有功能、接入更多第三方服务，让现有 GPU 客户在算力之外采购更多产品，带来额外的收入。

2. 从锁定容量到开始计费：容量交付与资金投入

a.

场地和电力通常提前一至三年或更久锁定

项目签约初期，CRWV通常会支付保证金和开发费用，并就电力接入支付相关款项或提供信用支持；机房租金和实际用电费此时尚未全面起算。进入施工阶段后，部分项目还需按建设进度支付供电、液冷设备的预付款和进度款。随着数据大厅建成、交付并通电，相关容量逐步转为 Active Power。

结合渠道调研，预计 CRWV 目前约七成容量采用第三方机房。对于这部分容量，土地、建筑、基础供电、备用电源、基础冷却和数据大厅通常由开发商投资；电力接入和变电站主要由开发商及公用事业公司承担。高密度液冷改造可能由开发商与 CRWV 分担，CRWV主要负责 GPU 服务器及机柜、高性能网络、存储和集群软件。

开发商可以依靠 CRWV 的长期租赁承诺筹集建设资金，CRWV无需在项目初期支付完整的数据中心土建成本。不同数据大厅交付后，机房租金和容量费逐步起算；实际用电费用则随设备安装和运行增加。

b. Active Power到开始计费仍需完成装机和验收

机房通电后，部分容量仍在等待 GPU 安装，另一些已经完成装卡，尚处于烧机、集群调试或客户验收阶段。完成验收并进入正式服务期后，相关容量才开始计费。

交付环节	通常所需时间	较快情况	主要工作
通电至 GPU 安装完成	4-8周	约 4 周	GPU服务器、机柜、网络及存储设备进场和安装
GPU安装完成至集群调试完成	2-5周	约1.5-2周	烧机测试、网络与存储联调、故障排查
集群调试完成至客户验收	1-4周	最快约 5 天	性能测试、机柜验证及客户验收
客户验收至开始计费	1-2周	最快可在验收后立即开始	集群进入正式服务期

来源：渠道调研，Funda整理

设备安装、网络与存储联调及验收准备通常会交叉推进。进展较快的项目从通电到开始计费约需六周；设备到货、液冷调试或客户验收出现延误时，交付周期会相应拉长。每季度新增 Active Power 中，可能有一半左右能够在当季开始贡献收入，其余部分则更多在下一季度陆续进入计费。

1. 客户计费前的三至九个月，资金投入明显增加

根据渠道调研，单个项目的资金压力通常在客户开始计费前的三至九个月最为明显。这段时间，GPU及配套设备陆续到货，机房开始起租，设备安装和调试也带来用电及运维支出。

机房通电后，即使 GPU 尚未安装，CRWV也已承担租金、容量费、电力预留费及部分融资利息，实际用电量则仍然较低。GPU 安装后，用电、冷却和运维费用随着集群测试逐步增加；设备达到可使用状态后，折旧也可能先于客户计费开始。

GPU订单通常在集群交付前数月确定。CRWV 先向供应商支付定金和进度款，设备到货及安装后仍需支付尾款；客户则在集群完成验收后开始计费。

客户预付款可以缓解这段时间的资金压力，但款项通常随建设节点分批到账。CRWV 还会随项目建设提取贷款，并通过设备融资支付部分硬件款项；公司债和可转债则为多个项目同时建设提供资金周转。

2. 近期延期以电力接入和交付验收为主

根据我们的渠道调研，CRWV的交付情况较前期有所好转，8月以来未见新的重大延期，但个别项目仍可能跨季度交付。GPU 和网络设备的供应压力有所缓解，目前更容易影响交付进度的是电力接入、变电站送电，以及最后的液冷调试和客户验收。

3. 容量交付与收入预测

a. 下半年新增容量的项目分布

CRWV在 2026 年第二季度末的 Active Power 为 1.5GW 。公司在二季度业绩会上预计，年底 Active Power 将超过 1.85GW 。我们与两位不同背景的行业专家沟通，了解到以下两组容量预估。

- i. 第一组反馈来自一位 NCP 行业专家。他预计，CRWV在 2026 年 Q3 能够新增约200-300MW的-Active Power，对应季度末容量约1.7-1.8GW。对于全年，该渠道预计，随着下半年项目陆续交付，公司年底 Active Power 能够超过 2GW 。
- ii. 第二组反馈来自 CRWV 某大型客户的一位数据中心专家。该渠道根据主要站点及地区的交付安排，对 Q3 、 Q4 新增容量进行了更详细的拆分，给出的整体预估也更为乐观。下表按其反馈整理，列示各项目或地区预计新增的 IT 容量。

26Q3 主要新增容量来源项目或地区	预计新增 IT 容量（兆瓦）
得克萨斯州 Denton	60
俄克拉何马州 Muskogee	18
佐治亚州 Dalton	27
北达科他州 Polaris Forge 1	62
现有托管机房及其他美国项目	120
英国及欧洲项目	50
其他分期交付项目	50

来源：渠道调研，Funda整理

26Q4 主要新增容量来源项目或地区	预计新增 IT 容量（兆瓦）
得克萨斯州 Denton	10
俄克拉何马州 Muskogee	5
佐治亚州 Dalton	45
北达科他州 Polaris Forge 1	37
现有托管机房及其他美国项目	170
英国及欧洲项目	75
其他分期交付项目	95

来源：渠道调研，Funda整理

该渠道对 2027 年新增容量及不同 GPU 型号部署分配的预估，与我们通过硬件渠道了解到的 CRWV 2027 年采购需求在数量级上大致相符，两组调研对未来部署规模提供了一定交叉验证。该渠道对 2026 年 Q3 、 Q4 交付节奏的判断则更为乐观；若按其预计的规模交付，年底容量有望明显超出当前预期。因此，我们未将其对 2026 年剩余时间的容量预估纳入模型，仅将这组反馈作为下半年容量交付的乐观情景参考。

容量交付后，GPU 安装、集群调试和客户验收仍需时间。结合行业反馈，季度新增的 Active Power 中，约有一半对应的集群可能在当季开始计费，其余更多在下一季度体现。若新增容量集中在季末交付，当季收入贡献有限，而部分机房租金和调试费用已经发生，利润率及按期末 Active Power 折算的单位容量收入都可能暂时偏低。

从成本构成看，电力和机房租金等支出大致随容量增长，集群运维、软件、研发和管理费用则不一定需要同比例增加。随着已计费容量扩大，后几项费用占收入的比例有望逐步下降。

1. 新签合同与设备更新逐步带动单位容量收入

根据渠道反馈，新签合同报价上调后，较高价格有望随着新增集群交付和旧卡续约逐步体现在收入中。今年部分 H100 合同进入续约期，2025年的低价合同也将陆续到期，早期低价合同对应的收入占比预计逐步下降。

与此同时，新增集群中的 GB200、GB300 占比提高，后续 Rubin 也将进入交付。根据行业调研及我们的初步推算，裸金属情形下，每吉瓦 Rubin 对应的收入可能比 Blackwell 高约20%。合同价格的变化和新一代 GPU 的交付，有望逐步提高单位可计费容量对应的收入。

前文讨论的Managed Inference，以及未来可能扩展的云服务，也可能带来额外收入。相关增量目前尚难单独预测，我们暂未额外建模。若按 Token 收费的业务逐步放量，每吉瓦收入可能进一步上升。

调研中部分更高的每 GW 报价，可能来自依赖风险投资融资的 NeoLabs。相关客户的持续采购能力取决于后续融资，涉及容量也较小，我们未将其纳入基础预测。

2. 新增集群带来的单 GW 设备投入逐渐增长

CRWV 多数项目采用租赁机房，但 GPU 机柜、网络和存储设备仍需由公司采购。根据渠道反馈，Rubin 机柜采购报价约为每柜 1,000 万美元。在这一价格水平下，Rubin 集群的机柜部分资本开支可能达到每吉瓦 400 亿美元以上，高于目前约 300 亿美元的水平。

随着越来越多的新增容量用来部署 Rubin，每吉瓦对应的资本开支预计也会逐渐上升。设备采购早于客户开始计费，因此新卡占比提高，还会增加项目交付前的资金投入。

3. 机房运营成本以租金和电力为主

我们也针对 GPU 集群的运营成本进行了调研，两个渠道反馈的费用拆分如下。每 GW 每年的费用大致在 22 亿-29亿美元，以机房租金和电费为主，设备折旧和融资利息另计。

费用项目（单位 \$B/GW）	渠道一	渠道二
机房租金	1.8	1.5-2
电费	0.35-1.05	0.6
维护、现场人员及其他	N.A.	0.3
合计	2.2-2.9	2.4-2.9

来源：渠道调研，Funda整理

五、Funda 预测与市场一致预期

1. 关键指标预测

我们的预测主要由新增容量、计费进度和单位容量收入驱动。新增集群交付及低价合同陆续续约后，单位可计费容量收入预计逐步提高；Managed Inference和更多通用云服务的额外贡献，暂未单独计入预测。

来源：公开数据，Bloomberg，渠道调研，Funda测算

1. 主要预测依据

我们在模型中预计，2026—2028年末 Active Power 分别为 2.10GW、3.60GW 和 5.40GW。新增容量通电后，还需完成 GPU 安装、调试和客户验收，才能开始计费。因此，季度内较晚上线的容量，收入贡献更多体现在后续季度。

单位收入方面，我们假设2026—2028年末每 GW 可计费容量对应的年化收入分别为 105 亿、120 亿和 130 亿美元。新签和续约价格上调，以及 GB200、GB300 和后续 Rubin 在新增集群中的占比提高，有望逐步带动单位容量收入上升。不过，新一代 GPU 的设备投入也更高，项目交付前的资金需求会相应增加。

从成本构成看，电力和机房租金等支出大致随容量增长，集群运维、软件、研发和管理费用则不一定需要同比例增加。随着已计费容量扩大，后几项费用占收入的比例有望逐步下降。

六、风险

1. 下行风险

a. 算力需求仍是首要风险

当前的算力紧缺，并不能让我们准确判断未来几年的需求。CRWV 在主动拓展长约之外的业务，希望在供不应求时获得更高收入。公司在争取更强收入弹性的同时，也把自己更多地暴露在需求波动中。一旦需求增长不及预期，这部分业务的价格和利用率可能更快承压。

b. 部分高价合同的可持续性

我们注意到市场上存在类似“每GW ARR

30B-40B美元”的讨论，我们了解到部分高价订单可能来自第三、第四梯队 AI 实验室等中小客户。这些客户愿意为紧缺的算力支付更高价格，但未必有足够资金长期维持这样的采购。若这些客户后续融资跟不上，相关容量在续租或重新出租时，价格可能回落。即使只涉及少部分容量，也可能影响市场对 CRWV 每 GW 收入持续上升的预期或叙事。

c. 后续扩张仍离不开外部融资

我们在渠道调研中也询问了融资情况，目前尚未听到融资渠道明显受阻的反馈。但按我们的预测，CRWV到 2028

年自由现金流仍无法转正，后续建设依然需要持续融资。能否持续获得足够资金，是实现 2030 年 8GW 容量目标的关键；融资一旦受阻，容量交付和收入增长都可能受到影响。

2. 上行风险

a. 每 GW 收入的上行空间及提升速度可能被低估

若算力紧缺持续，新签和续约价格可能进一步上涨，带动每 GW 收入高于我们的假设。

Managed Inference 及更多通用云服务的额外贡献尚未单独计入预测，若这些业务加快放量，每 GW 收入的提升也可能快于我们预期。

b. 容量上线速度可能快于预期

若主要项目交付更顺利，新增容量可能高于我们的假设。新增集群从通电到开始计费的时间若有所缩短，也会使更多容量提前贡献收入，带来收入预测的上修空间。

免责声明

免责声明与使用条款

分析师独立性声明

本报告所载观点均真实反映分析师对相关证券及发行人的个人独立观点。报告酬劳之任何部分，均不会直接或间接与本报告所载具体建议或观点相关联。本公司恪守严格合规流程，以确保所采用数据来源及信息之完整性、安全性及合法性。

关于 FUNDA

Consensus AI Technology Pte. Ltd. (“本公司”或

“FUNDA”)系一家注册于新加坡的独立研究机构。本公司不隶属于任何证券交易商、投资银行或资产管理公司，亦未在任何司法管辖区注册为投资顾问或经纪交易商。本公司与本报告所涉公司之间不存在任何投资银行或承销业务关系，亦不接受上市公司的赞助或资助。

非投资建议与报告中立性

本报告内容及其所载信息仅供参考，不构成任何投资建议或商业决策依据。报告中观点仅代表相关观点持有者个人在特定时间点的看法，不代表 FUNDA 的立场。作为独立运营的平台，FUNDA

保持中立第三方地位，不对上述观点的准确性、完整性或适用性作出任何明示或暗示的保证。

- 非邀约：** 本报告不构成买卖或持有任何证券或金融工具之邀约、招揽、推荐或建议，亦不构成任何法律、税务或会计建议。
- 第三方数据：** 本报告中引用的第三方数据均来自公开或授权渠道，本公司不对其准确性或完整性作出任何明示或默示保证。
- 不参与决策：** 本公司不以任何直接或间接形式参与客户或读者之投资决策过程。请读者根据自身投资目标及财务状况做出独立投资决策，并在进行任何交易前咨询专业财务或法律顾问意见。

时效性声明

鉴于市场环境与技术趋势的快速变化，本报告反映的是截至发布日之前的市场情况，相关信息可能随时间推移而失效。FUNDA 无义务对报告内容进行实时更新。

知识产权、保密与商业使用限制

本公司提供的一切报告、数据、模型及内容均构成本公司独家专有知识产权，受版权法和知识产权法保护。

- 权利归属：** 除非书面协议另有明确规定，因本报告产生或与之相关之一切知识产权（包括著作权）均独家归属于本公司。
- 授权使用：** 机构客户可在其内部合规政策允许之范围内使用本报告，但严禁向任何第三方披露或共享。
- 禁止商业使用：** 未经本公司事先明确书面同意，任何个人或实体不得复制、转载、分发、转售、重新包装或以其他方式将本报告或其任何部分用于任何商业或盈利目的。
- 法律追索：** 任何未经授权之商业使用、再分发或侵犯本公司知识产权之行为，本公司将依法追究其一切法律责任及追索权。

前瞻性陈述与免责声明

本报告可能包含前瞻性陈述，包括预测、估算及行业趋势分析。该等陈述受制于可能导致实际结果与预期产生重大差异之风险及不确定性。对于因使用或依赖本报告内容而产生之任何直接、间接或附带损失，本公司概不承担任何责任。