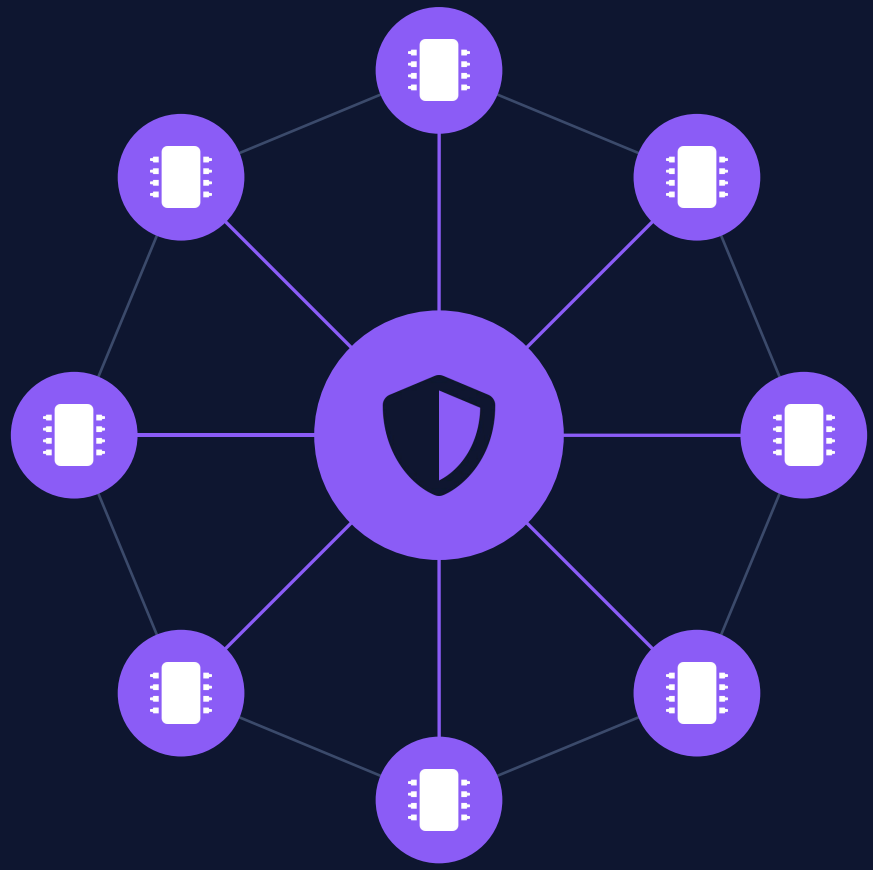
A12 技术优势

分布式算力接入 · 用户数据隐私



CTO 技术简报 | 面向市场团队与重点客户 | 2026年10月

一句话读懂 A12

我们将全球散落的 GPU 接入网络，以可验证的信任加以智能利用，同时每个租户的数据始终只属于他们自己。

A12 连接云 GPU 集群、专业数据中心以及闲置的工作站/边缘 GPU，对其进行验证和测量，按任务特征调度工作，并在此之上提供模型服务（MaaS）和 AI 智能体应用（AaaS）。

≤ 10 秒

实例冷启动

基于容器的调度

≤ 400 毫秒

首 token 延迟

公测期间中位数

≥ 99%

验证通过率

封闭测试退出标准

注：右侧数字为白皮书设计目标，实测结果将在各阶段验收报告中发布。

问题不在于算力太少，而在于组织不够好

当前算力供给的三个层级

超大规模云平台

标准化、易用，但价格昂贵

AI 数据中心 / 企业集群

互联速度快，但利用率波动大

消费级与工作站 GPU

数量庞大，因缺乏可信接入方式而闲置

需要解决的四个工程问题

异构性

不同厂商的驱动和内核库，导致工作负载难以迁移

信任

开放接入会引来伪造、显存超卖和重复挂单

SLA 与弹性的矛盾

闲置 GPU 会不可预期地掉线；在线服务会因此崩溃

可持续性

补贴驱动的供给在真实需求不足时难以为继

四层架构：从算力到 AI 智能体

第 4 层 **智能体与应用（AaaS）**

智能体画布 · 开发者工具 · 行业智能体 · 开放平台

第 3 层 **模型服务（MaaS）**

模型接入与版本管理 · OpenAI 兼容 API · 推理调优

第 2 层 **调度与资源池化** 今日重点

任务画像 · 双路径调度 · 租约 · 租户控制台与沙箱

第 1 层 **接入与信任验证**

节点接入 · 基准测试 · 运行时遥测 · 信誉

CaaS 算力 → MaaS 模型 → AaaS 应用：价值逐层累积



控制平面与数据平面分离

- 控制平面：调度、计量、计费
- 数据平面：模型执行、数据传输
- 数据保留在执行任务的集群内部；控制平面不中转模型或批量数据

01

分布式算力接入

将闲置 GPU 变成真实、可信、可调度的资源



一块 GPU 如何接入 A12

1



安装节点代理

节点守护进程以非特权用户运行，无需管理员权限

2



建立加密连接

节点主动向平台拨出，无需公网 IP 或端口映射

3



硬件体检

读取型号、显存、驱动和拓扑结构，与官方规格比对

4



随机突击测试

随机基准测试测量真实算力，无法提前准备

5



评级入池

按实测得分（CU）归入 S / A / B / C 等级



入池不是终点：持续遥测和信誉追踪会让变强的节点升级、变弱的降级，并持续标记异常。

五层验证：信任靠层层堆叠，而非一招制敌

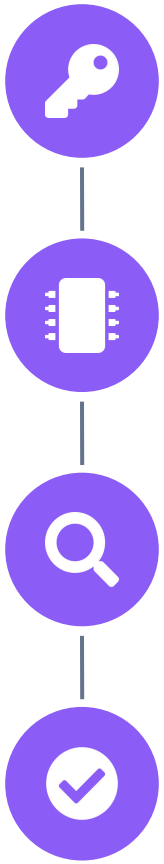
L1	声明规格	*“查身份证”*	读取自报硬件信息并与规格库比对，仅作第一道筛选
L2	基准测试	*“突击测验”*	平台下发随机测试，测量真实算力和显存性能
L3	运行时遥测	*“运动手环”*	追踪利用率、显存、温度和功耗，捕捉与声明不符的情况
L4	任务验证	*“抽查作业”*	抽样重跑、跨节点比对、争议时回放任务以定责
L5	信誉	*“信用分”*	在线时长、成功率、通过率随时间累积，影响任务分配和收益

设计目标

作弊成本 ≫ 作弊收益

- 没有开放系统拥有“完美反作弊”，我们也不声称有
- 绕过一层不等于系统性风险
- 异常节点会被持续标记、降级，最终被移除

随机挑战：无法提前背答案



平台先锁定题目

可验证随机函数（VRF）生成种子并先发布承诺，节点无法预测测试内容

节点当场作答

在 GPU 上运行随机矩阵运算，规模 1K 到 16K 随机，精度组合随机

平台检查两件事

答案对不对？速度合不合理？

事后公开种子

任何人都可以独立验证平台没有挑题



答案必须正确

与平台参考结果的误差必须在允许范围内（已考虑不同硬件间合法的浮点差异）。



速度必须合理

快到超出物理可能 = 虚标或预缓存答案，立即标记。

被标记的运营方可申请重测；如属误判，冻结收益全额返还，不留记录。

三道反作弊关卡：账本、身份与收益



一块卡挂多个单？

账本级排他

- 每块 GPU 按序列号 + 运营方注册，序列号全球唯一
- 同一时间只能有一个有效租约，一块卡不能用三个身份卖三次
- 换绑需要双方审核和 30 天冷却期



批量假账号？

按运营方归集

- 信誉、收益和惩罚都归集到运营方（个人或公司）
- 新运营方前 30 天有收益上限且逐步释放
- 批量注册的一次性收益趋近于零



假交付？

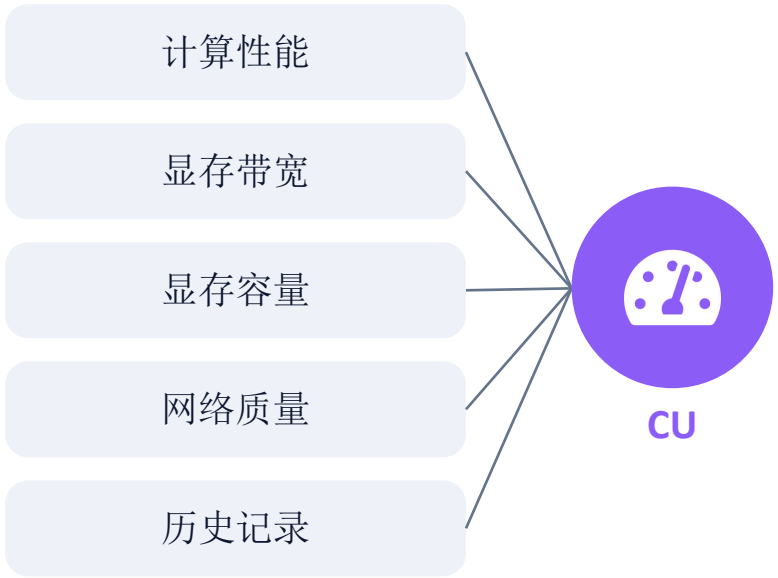
延迟释放收益

- 收益流程：已赚取 → 待定 → 已释放
- 在 1–72 小时的待定窗口内处理抽查和争议
- 违规者没收待定收益，用于补偿受影响租户

合法多机运营方（网吧、校园实验室、小型工作室）可在 7 个工作日内提交设备清单审核；仅被怀疑只会限流，绝不没收。

实测评级：按真实能力分级，而非型号

能力得分（CU）是五个实测因素的加权组合



每个因素都来自 L2 基准测试和 L3 遥测，而非厂商规格表；权重因任务类型而异。

等级	典型硬件（已验证）	适用工作负载	供给模式
S 级	高端数据中心 GPU + 高速互联	分布式训练、大模型全参数微调	专属
A 级	数据中心单节点（A100/H100 级）	大模型在线推理、长上下文服务	专属
B 级	专业工作站（RTX x090 级及以上）	标准在线推理、LoRA 微调、嵌入	主要弹性；可申请专属
C 级	消费级 GPU、边缘设备	批量任务、RAG 数据处理、离线推理	仅弹性

测试不达标的机器降级，超过标准的可以升级。

智能调度：每个任务去最适合的地方

调度依据：通信与计算之比

α = 通信时间 ÷ 计算时间

α ≥ 1 通信密集

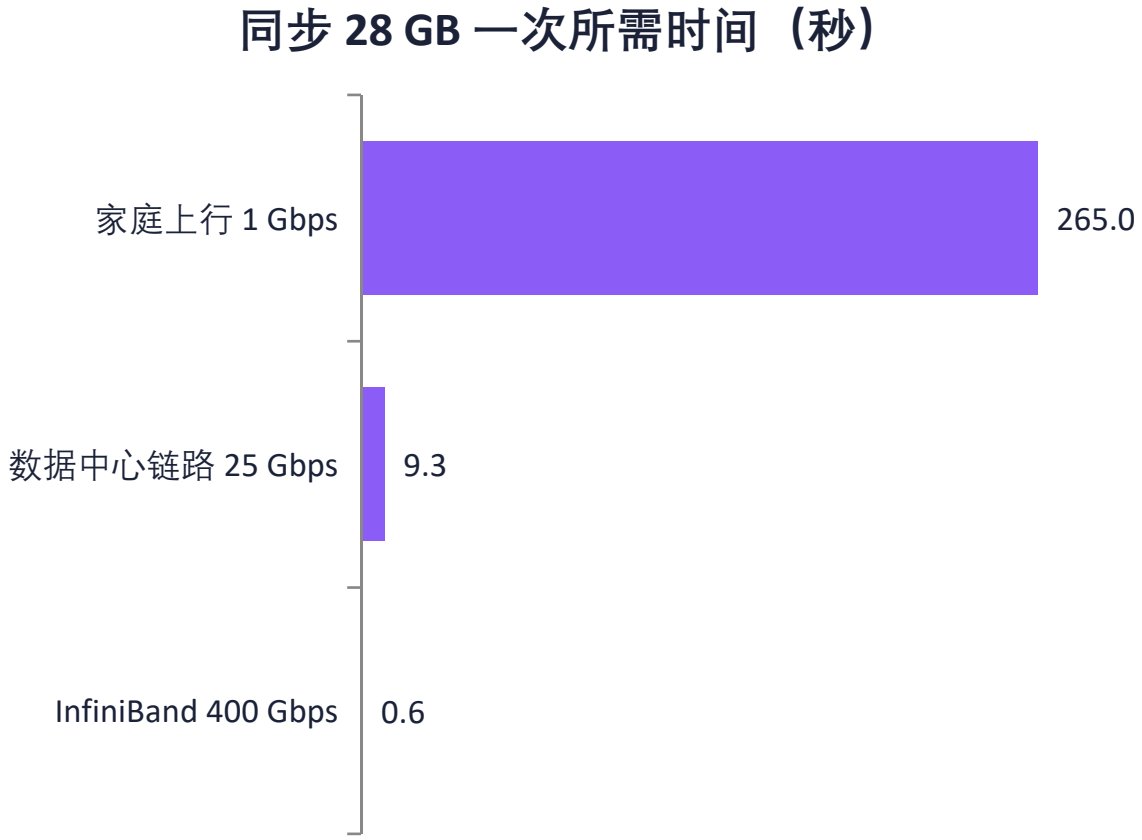
集中式集群

大模型训练、全量微调、大规模预训练

α < 1 计算密集

A12 分布式资源池

在线推理、嵌入、RAG、LoRA 微调、批量任务



示例：7B 模型全量微调（每步同步约 28 GB）。设计估算来自白皮书 §5.2 和附录 D，非实测。

两条供给通道：弹性与可靠，物理隔离

弹性尽力而为

供给	个人工作站、闲置 GPU、边缘设备
任务	可中断、可重试的批量任务、嵌入、RAG、离线推理
SLA	无；失败任务自动重试
准入	无需预付押金

专属保障

供给	企业级数据中心、专业集群、高可用工作站
任务	在线推理、模型服务、企业 API
SLA	可用性目标、冗余部署、违约赔偿
准入	托管收益或保证金

冗余示例：主节点 + 备节点

99.5%

单节点：每月约 3.6 小时宕机

99.9975%

双节点：每月约 1.1 分钟宕机

专属容量不足时，我们宁可拒单也不超卖。
承诺底线为 ≥ 99.5% 服务可用性（设计目标）。

02

用户数据隐私

你的模型和数据只属于你



第一原则：验证算力绝不要求看你的数据

A12 的明确承诺：

验证绝不依赖于访问租户的模型权重或数据

作为信任系统的硬约束写入白皮书第 4 章“设计原则与边界”。

测试由平台出题

基准测试使用平台生成的随机矩阵，与租户任务无关；任务检查采用抽样重跑和结果比对。

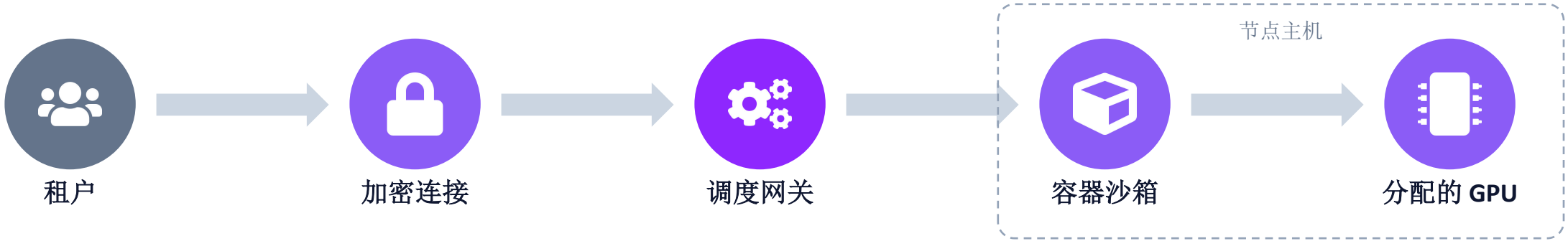
控制平面绝不触碰业务数据

调度、计量、计费是控制平面功能；模型运行和数据传输都在任务集群内部，绝不经过控制平面中转。

透明、可审计的计量

每个任务的用量都被记录并可导出审计；调度和定价规则公开。

数据旅程上的五道防线



1

加密传输

节点主动建立出站加密 WebSocket 隧道；无需开放公网端口，无需分发密钥

2

身份与用途审核

租户完成 KYC；挖矿、端口扫描等滥用行为被禁止并通过遥测检测

3

镜像白名单

任务镜像只能来自经批准的仓库；节点不能随意拉取外部镜像

4

容器隔离

cgroups v2 + Seccomp + 只读根文件系统，配有严格的 CPU、内存和 IO 限制

5

GPU 可见性限制

容器只能看到分配给它的 GPU；节点代理以非特权身份运行

选择与数据敏感度匹配的保护等级

模型 IP：企业私有模型仅托管，归属权永不改变。模型注册表记录来源、版本和权重摘要。

标准

弹性节点

容器隔离 + 用途审核，并有滥用赔偿基金兜底

适用：基于公共数据的批量任务、嵌入、离线推理

现已可用

高敏感

专属保障模式

运行于企业级数据中心/专业集群，带冗余部署，与弹性资源池物理隔离

适用：企业私有模型、在线 API、需要强 IP 隔离的工作负载

现已可用

前瞻性

TEE 机密计算

支持机密计算模式的数据中心 GPU 在硬件层加密内存和计算，连节点所有者都无法读取租户数据

适用：在分布式资源上运行高敏感企业工作负载

研究方向，非当前承诺

我们坦诚说明边界，并为风险兜底

我们不做的承诺

- 容器隔离大幅缩小了攻击面，但我们不承诺能杜绝每一次容器逃逸
- 没有开放系统拥有完美反作弊；验证通过经济学规律威慑作弊
- 所有性能数字都是设计目标；实测结果以公开验收报告为准
- TEE 机密计算在合格硬件规模落地之前不算承诺

我们如何兜底

- 平台保险 + 滥用赔偿基金
- 违规节点的待定收益被没收，用于补偿受影响租户
- 平台故障期间计费暂停，租户绝不为我们的失败买单
- 计量记录可导出审计；调度和定价规则公开
- 每个阶段结束后 7 天内发布验收报告

路线图：每个阶段都有可验证的退出标准

0 1 2 3 4

阶段 0

封闭测试

2026年10月初

邀请制，50–100 个节点；核心调度和五层验证上线

退出标准
通过率 ≥ 99%；零重大安全事故

阶段 1

全球公测

2026年10月下旬起

开放弹性注册；延迟释放收益；OpenAI 兼容网关

退出标准
月度节点留存 ≥ 60%；争议率 < 1%

阶段 2

SLA 商业上线

2027年上半年

专属模式 + 冗余部署；MaaS 上线；首批企业租户

退出标准
SLA 目标达成；≥ 3 家付费企业

阶段 3

AaaS 与扩张

2027年下半年

智能体画布和 Web IDE；实验性支持 AMD ROCm

退出标准
AaaS 月度收入持续增长

阶段 4

协议层

2028年及以后

链上结算记录；跨厂商研究

退出标准
通过监管审查；审计公开

时间节点以工程验收为准：任何标准未达成就延期，绝不跳步。日期为目标窗口，以发布的验收报告确认。

这对客户意味着什么

更低成本

按任务特征自动路由：低通信量工作跑在高性价比的分布式算力上，训练则留在高速集群里。

更可靠

五层验证、实测评级、物理隔离的供给通道和冗余设计，意味着算力真实、服务可控。

更安心

验证不看数据、多层隔离、透明保障，数据和模型始终掌握在客户手中。

谢谢 · 问答

基于《A12 技术白皮书》v3.3 RC