技术白皮书 · TECHNICAL WHITEPAPER

Aelion Twelve（A12）

超级纳管云台技术白皮书

Version 3.3 RC（定稿候选版 · 非公开发布版本）· 2026 年 9 月

异构 GPU 算力的可信接入 · 智能调度 · 分层价值链

CaaS 算力 → MaaS 模型服务 → AaaS 智能体应用

A12 Project | Version 3.3 RC（定稿候选版 · 非公开发布版本） | 2026 年 9 月

系统架构 · 可信核验 · 智能调度 · 模型服务 · 智能体平台 · Token 经济与路线图

目录

摘要

§0 版本效力声明：本文件为 Version 3.3 RC（定稿候选版 · 非公开发布版本），仅作为技术评审与尽调参考材料流转，不构成公开发布版本；封面与页眉的版本号标识版本谱系，不代表发布效力。全部披露事项已统一编号为「披露闭环追踪表（D-01 至 D-30）」，完整表格见附录 C，逐项列明编号 / 事项 / 所属章节 / 要求形式 / 当前状态 / 承诺日期 / 公示链接；全部编号闭环并经官方渠道公示后，本文件方升格为正式发布版，并同步公示发布记录。本声明为版本效力的唯一完整表述；封面与页眉的版本标识为本声明的简写，两者如有出入，以本声明为准。

A12 是一个面向异构 GPU 算力资源的**调度与服务平台**。它将云端 GPU 集群、专业 IDC 与闲置工作站/边缘 GPU 统一接入、核验、度量与调度，按任务特征把高通信强度的大模型训练留在高速互联的集中式集群，把低通信强度的推理、检索增强（RAG）、参数高效微调（LoRA/QLoRA）与批处理任务分配到成本更优的分

布式资源上，并在统一的任务接口之上提供模型服务（MaaS）、智能体应用（AaaS）与完整的开发者工具链。

本白皮书只描述当前已验证或具备明确工程路径的设计，并遵守三条写作纪律：**所有性能指标均为设计目标而非实测结论**；**所有安全机制均同时说明其边界与失效条件**；**所有涉及项目主体与历史的信息以披露清单形式呈现，未披露前不作为可信度依据。**

关键词：异构算力调度；可信资源核验；混合算力供给；模型服务；智能体应用

实例冷启动	首字延迟 TTFT	GPU 有效负载	验证通过率
≤ 10 秒 容器化实例调度 · 设计目标	**≤ 400 ms** 公测期中位数 · 设计目标	**≥ 75 %** 连续批处理 · 设计目标	**≥ 99 %** 内测退出判据 · Go/No-Go

第一章产业背景与问题定义

1.1 AI 算力供给的结构性现状

生成式 AI 的工作负载正在快速分化：基础模型预训练与全参数微调需要高频梯度同步，天然依赖 NVLink/InfiniBand 级别的集群互联；而推理、Embedding、Rerank、RAG 数据处理与 LoRA/QLoRA 微调等任务，通信强度低一个到多个数量级，可以在单机或小规模集群内闭环完成。两类负载对基础设施的要求差异显著，却长期被"同一种供给方式"服务。

与此同时，算力供给端呈现三层结构：超大规模云平台提供标准化但价格较高的托管算力；专业 AI IDC 与企业自有集群具备高速互联但利用率波动大；全球还存在数以千万计的消费级与工作站级 GPU，因缺乏可信的接入与调度机制而处于闲置状态（量级为公开行业口径的估算参考，非精确统计；本章行业数据均用于说明结构判断）。**问题不是"算力不够"，而是不同物理条件的算力没有被组织到合适的工作负载上。**

1.2 混合供给的组织效率问题

将上述三层供给统一编排，需要解决四个工程问题：

1. **异构适配**：NVIDIA、AMD 及其他加速卡的驱动、运行时与算子库差异，使上层任务难以无成本迁移。任何宣称"完全无感统一"的方案都回避了真实的适配工作量。
2. **资源可信**：在开放接入的环境中，存在资源伪装、显存超售与重复申报的可能。仅依赖节点自报信息不足以建立调度所需的信任。
3. **SLA 与弹性的矛盾**：闲置的边缘 GPU 在线率不稳定，直接承载企业级在线服务会导致可用性塌陷；而完全使用集中式资源则放弃了对闲置算力的利用。

4. **经济可持续性**：仅靠代币排放补贴供给的模式，在真实需求不足时会陷入"排放—抛压—激励失效"的循环。

1.3 A12 的解法（概述）

针对上述问题，A12 的方案可以概括为四点，其可行性边界在对应章节中逐一说明：

1. **按通信强度分流的调度架构**：高通信任务固定在集中式集群，低通信任务下沉分布式资源（第 3、5 章）。
2. **分层的资源可信核验体系**：以"声明核验 + 实测挑战 + 运行时遥测 + 任务级验证 + 历史声誉"组合建立信任，不依赖任何单一"绝对防伪"机制（第 4 章）。
3. **双轨供给模型**：弹性尽力模式承载可中断任务，专属保障模式承载 SLA 业务，二者物理隔离（第 4、5 章）。
4. **以真实服务收入为锚的经济模型**：平台收入优先，代币作为激励与结算凭证，回购销毁与收入挂钩且有硬上限（第 8 章）。

第二章项目定位与治理透明化

2.1 平台定位声明

A12 采用"**中心化运营 + 分布式执行**"的架构定位：平台由运营主体集中负责资源准入核验、任务调度、计费结算、争议处理与安全风控；算力资源本身分布全球（个人、机构、IDC 与边缘节点），任务执行在资源所在位置分布式完成，调度决策基于各节点的实时状态作出。用户之所以选择 A12，是因为平台以统一接口与可信核验降低了使用分布式异构算力的组织成本。

平台承诺：所有调度与计费规则公开可查；任务计量记录对租户可导出、可审计；未来若引入链上存证与结算，仅作为增强透明度的可选协议层（见 §8.6），不改变平台的运营责任主体。

2.2 项目主体与团队披露（强制披露清单）

> **披露状态：下表信息依据运营主体对外披露材料与公开注册记录整理。全部待补事项已统一编号为「披露闭环追踪表（D-01 至 D-30）」，完整表格见附录 C，逐项列明编号 / 事项 / 所属章节 / 要求形式 / 当前状态 / 承诺日期 / 公示链接；下表「编号」列与附录 C 逐行对应，读者可按编号检索。清单只允许收敛、不允许扩大——每新增一项披露承诺须先登记编号再写入正文。编号口径：上一版声明 21 个编号，而全文实际存在 30 处待补标记（登记遗漏 9 处）；本版按 30 处标记逐条登记为 D-01 至 D-30（完整表格见附录 C），属补登记而非新增承诺。本版未新增任何待补事项：图 5-1（D-21）与图 6-1（D-22）的参数表已随附录 D 直接给出，两项状态由「待补充」转为「已补充待佐证」，正文待补标记因此由 30 处收敛至 26 处，本轮随 D-02 / D-06 / D-07 三项信息补充进一步收敛至 24 处。此后任何新增承诺须先登记编号再写入正文。披露完整度是公开发布的硬性前置条件：任一必填项未补齐，本白皮书不进入公开发布。版本状态声明：本文件为定稿候选版（Version 3.3 RC，定稿候选版・非公开发布版本）。本文件所列【发布前补充**

/ 启用】项补齐并经官方渠道公示之前，本文件仅作为技术评审与尽调参考材料流转，不构成公开发布版本；全部前置项补齐后升格为正式发布版，并在官方渠道公示发布记录。支出侧口径见 §2.5《资金与资源透明度安排》。公示落点承诺（双落点 + 第三方时间戳）：公示事项先在当前内测预览域名（aeliontwelve.vercel.app/disclosure）建立过渡公示页，作为过渡期的有效公示落点；官网正式主域名启用后 72 小时内完成全部公示事项向主域名的迁移并发布迁移公告。为使自存证具备独立证据力，过渡期同步引入第三方时间戳：每次公示内容更新均生成内容 SHA-256 哈希，并至少采用下列方式之一固化时点——RFC 3161 时间戳服务签发的令牌、公开 Git 仓库的提交哈希、或写入 §8.6 存证合约的链上哈希；三者的验证材料与公示内容一并归档，任何人可离线核验。快照存档规则：每周一次全量快照，且每次内容变更即时增量快照；快照与时间戳令牌、内容哈希一并存放于公示页 /archive 归档目录，全部历史版本永久保留、可回溯比对。过渡页在主域名启用前持续维护（至少按月更新），直至迁移完成并发布迁移公告后满 30 日方可下线；迁移公告须在公示页与官方社交账号同步发布，载明迁移前后 URL、内容哈希与生效时点。公示节奏与启动时点：过渡公示页首版——含披露闭环追踪表 30 项的初始状态与首张许可状态表——不晚于 2026 年 9 月 21 日上线（本版发布后 5 个工作日内，与八项表中零成本事项的承诺日期同日）；此后追踪表每周一（UTC）更新一次、许可状态表每月第一个工作日更新一次，上线首周即为第一个更新周期，使「按周 / 按月更新」自首版起即可被外部核对；全部历史版本随 /archive 归档保留，任何一次更新未按节奏执行即视为逾期并适用逾期处理规则。

编号	披露项	内容	状态
—（已提供）	运营主体	Aelion Twelve Limited，经加拿大 Corporations Canada （加拿大公司管理局）注册，注册编号 1819483-1，已取得《公司注册证书》	已提供（注册信息可公开核验）
D-01	核心团队	Adrian Lin （林哲远）· 首席分布式计算科学家，硅谷华人， 12+ 年分布式计算与资源调度经验； Claire Zhang （张清妍）· 首席 AI 基础设施架构师， 10+ 年 AI Infrastructure 与云原生架构经验； Ethan Carter ·AI 平台工程负责人， 10+ 年大型互联网与 AI 平台工程经验； Tunde Akinwale （通德 · 阿金瓦莱）· 分布式系统负责人， 8+ 年分布式系统与高并发后端经验； Amara Okoye （阿玛拉 · 奥科耶）·GPU 基础设施与可观测性负责人， 7+ 年 GPU/Cloud Infrastructure 经验； Noah Bennett ·ML Systems 与 Agent 研究负责人	名单已提供；可验证履历链接（LinkedIn / 学术档案）将以「姓名 + 档案编号」形式在公示页逐人列示【发布前补充 · D-01 · 闭环关键路径】
D-02 / D-03	联系方式	**官方邮箱、办公地址、法务联络渠道**	**官方邮箱：aelion12ai@gmail.com；办公地址：2482 Yonge St,**

			Suite 185, Toronto ON, Canada（D-02 已公示：官网法律页与页脚，2026-09-15）；法务联络渠道：aelion12ai@gmail.com（与官方邮箱共用）；文书收件地址：2482 Yonge St, Suite 185, Toronto ON, Canada（D-03 已公示，2026-09-15）
D-04	投资方	Axiur Twelve（2021 年注资并设立专项开发小组）；Bluechip 及其他资本机构（2025 年跟投；机构全称、背景简介与投资轮次【发布前补充 · D-04】）	待公开佐证【发布前补充 · D-04：工商变更 / 公开新闻 / 领投方确认函，任一即可；Bluechip 全称与简介】
D-05	知识产权	核心算力调度技术专利三地推进：中国申请已通过；尼日利亚申请进度约 90%；美国预计 2026 年底。上述进度表述以申请号 / 专利号公示为准，公示前不作为核验依据	专利号 / 申请号【发布前补充并公示 · D-05】
D-06	代码开放	**核心调度与验证模块的代码仓库地址（GitHub 公示）或第三方审计报告（全文公开，核验标准与 §8.6 合约审计一致）**	**代码仓库地址：https://github.com/aelion-tweive（D-06 已补充待佐证，随公示页首版发布；仓库公开范围与第三方审计安排以公示页披露为准）**
D-07 / D-08	官方渠道	官网正式主域名【发布前启用 · D-08】；当前 aeliontwelve.vercel.app/disclosure 仅作内测预览地址，不作为对外官方渠道；官方社交账号（D-07 已公示，2026-09-15；防仿冒——仅以下账号为官方账号，其余均为仿冒）：X（Twitter）https://x.com/AELION_TWELVE；YouTube https://youtu.be/pIQGQeiPdDw；Telegram https://t.me/AelionTwelvee；Discord https://discord.gg/b5zy5typXN	部分提供；主域名为公开发布硬性前置项

> 披露闭环关键路径（八项主体信息）：以下八项主体信息类披露，是本文件升格为正式发布版的前提，也是客户准入、节点招募、牌照申请与融资尽调的共用前置，列为最高优先级——① 团队 6 人可验证履历链

接；② 官方邮箱与办公地址；③ 法务联络渠道；④ 投资方全称、投资轮次与任一公开佐证；⑤ 专利号 / 申请号；⑥ 代码仓库地址或第三方审计报告；⑦ 官方社交账号清单（防仿冒）；⑧ 官网正式主域名。八项的「当前状态 + 承诺日期」见下表逐项列示，并与 §2.2 披露清单表各行、附录 C 追踪编号（D-01 至 D-08）三者一一对应；状态在公示页按周更新，白皮书内同步保留本表，使尽调对象本身可被核对。截至本版，② 官方邮箱与办公地址、⑥ 代码仓库地址、⑦ 官方社交账号清单三项已由运营主体提供，状态更新为「已补充待佐证」，随公示页首版（2026-09-21）发布。

序号	八项主体信息披露事项	对应 §2.2 清单行（追踪编号）	责任岗位	当前状态	承诺日期
①	团队 6 人可验证履历链接（LinkedIn / 学术档案）	核心团队（D-01）	人力资源负责人	待补充	2026-10-15
②	官方邮箱与办公地址（aelion12ai@gmail.com；2482 Yonge St, Suite 185, Toronto ON, Canada）	联系方式（D-02）	运营主体负责人	已公示（官网法律页与页脚，2026-09-15）	2026-09-21
③	法务联络渠道（法务邮箱 aelion12ai@gmail.com（与官方邮箱共用）；文书收件地址 2482 Yonge St, Suite 185, Toronto ON, Canada）	联系方式（D-03）	法务与合规负责人	已公示（与官方邮箱共用，2026-09-15）	2026-09-21
④	投资方全称、投资轮次与任一公开佐证	投资方（D-04）	首席财务官（CFO）	待补充	2026-10-31
⑤	专利号 / 申请号（中国 / 尼日利亚 / 美国）	知识产权（D-05）	法务与合规负责人（知识产权）	待补充	2026-11-30
⑥	代码仓库地址或第三方审计报告（全文公开；GitHub：https://github.com/aelion-tweive）	代码开放（D-06）	首席技术官（CTO）	已补充待佐证	2026-11-30
⑦	官方社交账号清单（防仿冒；X / YouTube / Telegram / Discord，完整链接见 §2.2 官方渠道行）	官方渠道（D-07）	市场与品牌负责人	已公示（2026-09-15）	2026-10-15

⑧	官网正式主域名	官方渠道（D-08）	首席技术官（CTO）	待补充	2026-10-31

注：责任岗位为运营主体内部的职能分工，由运营主体指定履行该职能的人员、外部顾问或代理机构承担；在对应人员到岗前由运营主体负责人代行，具体承担主体随公示页列示（顾问与外包配置见 §2.5 人力保障，D-20）。承诺日期由运营主体负责人批准后生效。节奏口径：承诺日期为考核节点，逾期即触发下述逾期处理规则；附录 C 的按周更新为过程披露，记录进展但不构成新的考核时点——两者以承诺日期为准。②③（官方邮箱与办公地址、法务联络渠道）为零成本项——不依赖任何第三方配合、不需要审批或外部核验——故承诺日期定为本版发布后 5 个工作日内（2026-09-21），与过渡公示页首版同日；①⑦ 需团队成员与平台后台配合，承诺日期 2026-10-15；④⑤⑥⑧ 需外部机构或监管流程配合，承诺日期如上。硬性观察点：第一个为 2026-09-21（②③ 零成本项与公示页首版，检验执行力）；第二个为 2026-10-31（D-04 投资方佐证与 D-08 主域名，主体可核验性的实质变化点）；第三个为 2026-11-30（D-05 专利号与 D-06 代码仓库，技术资产真实性）。逾期处理：任一事项在承诺日期仍未闭环的，运营主体须在该日期后 5 个工作日内于公示页说明逾期原因、当前进展与修订后的承诺日期，并同步更新本表与附录 C；同一事项连续两次逾期的，须另行公示整改说明；修订后的日期不得晚于原日期后 30 日（需外部机构配合的事项除外，此类事项须同时说明机构反馈周期）。

2.3 命名沿革与技术属地布局

命名沿革：Aelion Twelve 为项目及其运营主体名称。Axiur Twelve 系 2021 年向本项目注资并同步设立专项开发小组的战略投资方，与 Aelion Twelve 构成**投资方与被投项目**的关系。两个名称中的"Twelve"同源于投资达成时所使用的项目代号，属于投资沿革自然留存的命名痕迹。投资协议、注资凭证等书面文件纳入主体信息披露范围（见 §2.2）。

技术属地布局：核心算力调度技术的专利布局与研发团队构成保持一致。分布式系统负责人（Tunde Akinwale）与 GPU 基础设施负责人（Amara Okoye）常驻尼日利亚，相关研发工作在尼日利亚完成，专利因此首先在发明完成地提交申请，再按目标市场重要性依次推进中国（申请已通过）与美国（预计 2026 年底）。这是跨国研发团队的常规属地布局策略，具体进展以专利号 / 申请号公示为准（见 §2.2）。

命名与符号可用性：代币符号 A12 与项目名称 Aelion Twelve 的商标及主流行情平台唯一性检索于发布前完成，检索范围与结果【发布前补充并公示 • D-09】，确保不存在与既有项目 / 商标的冲突。

2.4 合规承诺

平台在运营中遵守：用户身份识别（KYC）与制裁名单筛查；算力租户用途审查与内容合规规则（见 §4.7）；数据保护与跨境传输合规评估；在涉及数字资产的司法辖区，Token 相关活动仅在取得必要许可后开展。运

营主体企业注册已完成（加拿大）；MSB 与 SEC 相关牌照申请正在推进中——涉及上述许可的业务在取得许可前不开展。

许可进展观测点：为使「取得许可前不开展」可被外部观测，运营主体在公示页维护一张许可状态表，逐项列明 MSB 注册、SEC 相关许可等事项的申请阶段、受理 / 回执编号的披露方式、预计完成时点区间与当前状态；首张许可状态表的发布日期不晚于过渡公示页上线日（见 §2.2），此后按月更新，更新责任岗位为运营主体的法务与合规负责人（外部律所参与事项在表中一并标注责任律所）；任何涉及许可的业务上线公告，须同时附对应许可状态的查询链接，链接失效时以「监管机构名称 + 受理 / 回执编号 + 该监管机构官方查询入口」作为替代核验方式，并在 5 个工作日内修复或替换链接。

2.5 资金与资源透明度安排

公测激励、第三方会计鉴证、合约安全审计、算力采购与节点分成、团队薪酬等均依赖运营主体的自有资金。为使"自有资金承担"的表述可被核验，作出如下安排：

支出科目	预算口径	资金来源	披露状态（追踪编号）
公测激励	一次性总额上限：30,000,000 A12 + 5 BTC + 900,000 USDT（见 §8.5）；法币与稳定币部分存入专门奖励托管账户，按公示的释放规则拨付	运营主体自有资金	上限已确定；托管机构与账户安排【发布前补充 · D-10】
第三方会计鉴证	按季度执行；费用区间【发布前补充 · D-11】	运营主体自有资金	甄选标准已定（§8.3）；机构名称【发布前补充并公示 · D-12】
智能合约独立审计	两份合约独立审计；费用区间【发布前补充 · D-13】	运营主体自有资金	审计标准已定（§8.6）；机构名称与报告全文【发布前补充并公示 · D-14】
算力采购与节点分成	随业务量浮动；年度上限【发布前补充 · D-15】	运营主体自有资金	口径随结算规则公示；年度上限随 D-15 一并公示

团队薪酬	核心团队 6 人编制；年度总额区间【发布前补充 · D-16】	运营主体自有资金	以资金状况披露清单形式留存可核验记录
资金状况披露	运营主体自有资金规模区间、已发生投入的主要科目、未来 12 个月资金缺口与补足方式（经营收入 / 股权融资 / 其他融资及各自合规前提）【发布前补充 · D-17 / D-18 / D-19】	运营主体自有资金及外部融资	以披露清单形式提供并留存可核验记录；D-19 列为最优先闭环项
人力保障	核心团队 6 人为当前编制；阶段 0 启动前公布招聘计划、顾问配置（法务 / 财务 / 合规）与外包边界【发布前补充 · D-20】	运营主体自有资金	随阶段 0 启动一并公示；路线图 Go/No-Go 判据同时是产能止损机制：任一判据未达成即推迟扩张，不以透支人力为代价跳过阶段

上表是「已承诺支出与资源保障」的唯一口径（本版已删除原并行的列表口径，避免同一节出现两套数字）：每一项已承诺支出均对应明确的资金来源、上限口径与追踪编号（D-10 至 D-20），读者可按编号在附录 C 逐项核对。本表确立「科目—上限—来源」的可核验框架，已确立可评估的框架，数值随追踪表（附录 C）公示后具备可评估性——在运营主体自有资金规模区间（D-17）与未来 12 个月资金缺口（D-19）公示之前，本文不主张支付能力已可被外部评估。闭环优先级：D-19（未来 12 个月资金缺口与补足方式）列为本表最优先项；若资金绝对规模因商业保密不宜公开，D-17 可以区间口径或「已向委托方单独提供并留存可核验记录」的确认函形式满足。

第三章系统架构总览

3.1 设计原则

系统设计遵循四条原则，每条原则都对应可验收的工程含义：

1. **物理拓扑亲和**：任务的调度目的地由其通信—计算特征决定，而不是由"去中心化程度"决定。训练类任务留在高带宽集群，推理类任务下沉边缘。
2. **异构硬件抽象**：统一的是任务提交接口与计量口径，而非底层执行环境；各厂商差异通过明确的适配矩阵披露，而非隐藏。

3. **分层可验证**：信任来自多层机制的叠加（声明、实测、遥测、任务验证、声誉），任何单层被绕过都不构成系统性风险。
4. **纵向价值链**：CaaS（算力）→ MaaS（模型服务）→ AaaS（智能体应用）逐层增值，平台收入随价值链上移而提高单价与黏性。

3.2 四层架构

层	名称	核心职责
Layer 1	资源接入与可信核验层	节点接入、资源描述、能力实测、状态监测、声誉记录
Layer 2	调度与资源池化层	任务画像、拓扑感知、双路径调度、资源租约、生命周期管理、租户控制台与沙箱运行时
Layer 3	模型服务层（MaaS）	模型接入与版本管理、垂直模型流水线、API 网关、推理性能工程、服务计量
Layer 4	智能体与应用层（AaaS）	Agent 编排画布、开发者工具链、预置行业 Agent、开放平台与统一收银台、商业结算

层与层之间以标准接口解耦。控制平面（调度、计量、结算）与数据平面（模型执行、数据传输）分离：数据面传输优先在任务所在集群内部完成，控制面不承担大规模张量数据的转发。

图 3-1 A12 四层系统架构：纵向价值链逐层增值，控制平面与数据平面分离

3.3 生态角色

角色	输入	获得回报	风险与义务
算力提供方（Worker）	GPU 资源与在线时长	计算服务收入（法币 / 稳定币计价， Token 激励额外发放）	遵守任务生命周期规则；滥用赔偿条款见 §4.7
验证节点	核验算力与保证金	验证服务报酬	须质押保证金；作恶罚没
模型开发者	模型与适配能力	调用收入分成	模型许可合规
Agent/SaaS 开发者	编排应用与工具	应用收入分成	应用合规与内容责任
租户 / 企业	真实业务需求	模型与 Agent 服务	数据授权合规

3.4 设计参照系（以太坊与 Filecoin 的工程启示）

A12 的机制设计大量借鉴了两个经过大规模验证的分布式系统白皮书的工程范式，但**不照搬其共识机制**——以太坊验证的对象是账本状态，Filecoin 验证的对象是存储正确性，而 A12 验证的对象是计算过程与性能，验证对象的不同决定了机制形态的差异。

> **特别声明：下表仅为机制设计层面的概念类比，用于借助成熟系统的语言帮助读者理解 A12；它不代表 A12 具有与以太坊 / Filecoin 同等的去中心化程度、信任模型或安全假设。** A12 是中心化运营的平台（见 §2.1），当前不发行公链、不构成 Layer2 或任何区块链网络；区块链仅作为远期可选的协议层（见 §8.6），其引入以监管评估为前提。

映射关系如下：

参照系统机制	A12 对应设计	所在章节
以太坊账户模型与 nonce 序列	运营者账户体系：声誉、结算、惩罚按账户聚合，多设备不可隔离责任	4.5
以太坊 Gas 计量	任务计量采用 "CU· 时长 + Token 数 " 双口径，逐笔记录、可导出审计	5.3 / 6.5
EIP-1559 基础费 + 小费的价格发现	调度定价 = 公示基础价（随供需按公开规则浮动） + 质量加成，杜绝暗调	5.3 / 7.5

以太坊智能合约结算	协议层在成熟公链部署结算存证合约，回购销毁链上公示可核验	8.6
Filecoin 抵押与罚没（Collateral / Slashing）	收益延期释放池 + 违约罚没，罚没上限不超过托管保证金	4.4 / 4.5
Filecoin 存储交易契约（Storage Deal）	资源租约：以 " 设备序列号 + 运营者 " 为键的账本级互斥锁定	4.5
Filecoin 挑战式证明（PoSt）	挑战对象从 " 存储正确性 " 迁移到 " 计算正确性 "，即 L2 随机基准挑战	4.2

4.1 设计原则与边界（重要声明）

开放环境中的资源核验不存在"绝对防伪"。A12 的设计目标是：**使作弊的综合成本（被检出概率 × 损失）显著高于作弊收益**，并使异常节点能被持续识别与降权。为此，本平台明确"三不做"：

1. 不使用任何"硅片物理指纹"作为设备身份，也不做基于指纹的永久拉黑（物理特征随温度、电压与老化漂移，既不可稳定复现，也无法支撑永久惩罚）。
2. 不使用零知识证明认证物理世界的模拟量（证明系统能证明"某段计算正确"，不能证明"某物理事件发生过"）。
3. 不以访问租户模型权重或数据为验证前提（保护租户隐私）。

4.2 五层可信核验体系

层级	机制	说明
L1 声明核验	采集 GPU 型号、显存、驱动、 PCIe/NUMA 拓扑，与已知规格库交叉比对	通过 NVML/ROCm 等官方接口采集；明确其在用户态运行，可被深度篡改，故仅作第一道过滤
L2 能力实测	平台下发的随机基准挑战： GEMM 矩阵乘、显存读写、张量核心压测，多种规模与精度组合	挑战种子由平台验证集群以随机信标（VRF）生成，节点不可预知；返回结果摘要、耗时与遥测
L3 运行时遥测	持续采集利用率、显存、温度、功耗、任务执行指标	用于发现 " 申报与实际不符 " 及超售迹象

L4 任务级验证	按任务类型选择：抽样重执行（平台验证集群）、冗余计算（多节点比对）、数值容差判定、争议时交互式重放定位	生成式输出采用结构化校验与人工抽检结合，不以数值相似度作唯一标准
L5 历史声誉	按运营者（见 §4.5 ）聚合：在线率、性能偏差、任务成功率、验证通过率	声誉影响调度优先级、任务准入与结算周期

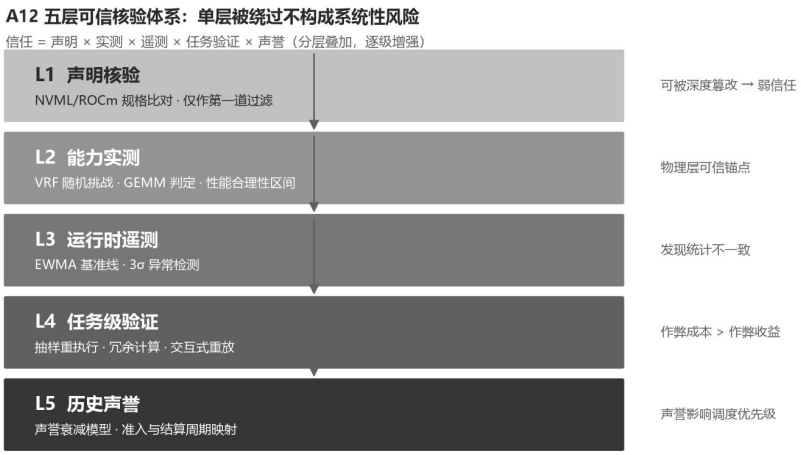

图 4-1 A12 五层可信核验体系：分层叠加，单层被绕过不构成系统性风险

五层核验体系中的 L2 与 L3 层采用可量化的统计判定方法，说明如下。

L2 能力实测的判定模型。 平台验证集群生成随机基准挑战：随机矩阵 A ∈ ℝ ^(m×k)、B ∈ ℝ ^(k×n)（元素服从均匀分布，种子由 VRF 信标派生），节点在目标 GPU 上以指定精度（FP16/BF16）执行 GEMM 计算 C = A × B，返回结果摘要（SHA-256(C)）与实际耗时 T。判定采用两组统计量：

- **正确性**：验证集群以参考实现复核抽样摘要，并要求相对 Frobenius 范数误差 ||C − C_ref||_F / ||C_ref||_F ≤ ε（ε 按精度档位设定，FP16 典型取 10^{-2} 量级，覆盖浮点非结合性与不同算子实现带来的合法偏差）；
- **性能合理性**：T_hat = 2·m·k·n / η_peak 为基于实测吞吐基线 η_peak 推导的预期耗时（m、k、n 为挑战 GEMM 的矩阵规模），(1−δ)·T_hat 为其合理下界。实测耗时低于该下界（T < (1−δ)·T_hat）即判定为"虚报算力"或"预计算缓存"攻击；显著偏慢的任务不影响挑战通过，由 SLA 与声誉机制处理（见 §4.5 与 §5.3）。
- **基线标定与污染防护**：η_peak 由平台验证集群对官方参考硬件档位定期实测标定，基线表公开并版本化；不以未上链节点历史实测的直接聚合作为基线来源，防止低质样本拉低基线使检测失效、或高质样本抬高基线误伤真实节点。

- **高性能节点误判防护**：实测优于基线属合法情形（能力升级，见 §4.6），仅当实测低于物理下界时触发虚报流程；被判定运营者可申请远程复测与人工复核（提交实测记录与运行环境说明），复核期间相关收益冻结于 Pending 状态，误判即全额解冻且不计入声誉负面记录。

多种矩阵规模（从 $1K^2$ 到 $16K^3$）与精度组合随机抽选，使节点无法仅对单一规模做针对性优化。

L3 运行时遥测的异常检测。 对利用率、显存占用、温度、功耗等时序指标采用指数加权移动平均（EWMA）维护基准线：

$$s_t = \alpha \cdot x_t + (1 - \alpha) \cdot s_{t-1}$$

其中 α 为平滑系数（典型 0.05–0.2）。当瞬时值偏离基准线超过 3σ（σ 由历史残差的 EWMA 方差估计）并持续 K 个采样窗口时，标记为遥测异常事件；异常事件累计计入运营者声誉（见 §4.5），并触发对应任务的加密度抽检。该方法的边界同样明确：遥测发现的是"统计不一致"，不构成对篡改手段的穷举证明——这正是需要五层叠加而非单点判定的原因。

挑战协议的形式化。 L2 挑战的生成与披露遵循承诺–揭示（Commit–Reveal）范式（该范式与以太坊随机数信标、Filecoin 窗口化证明同源）：验证集群每轮先公示挑战承诺，后揭示种子，节点可事后独立审计验证集群"未选择性出题"：

$$(y, \pi) \leftarrow VRF.Eval(sk_V, epoch); \quad Verify(vk_V, epoch, y, \pi) = 1$$

$$Com_q = SHA\text{-}256(seed_q \,\|\, \varepsilon \,\|\, \delta) \quad \text{（承诺仅绑定种子与公共参数）——挑战执行后揭示 } seed_q$$

其中 (y, π) 为 VRF 输出与证明，sk_V / vk_V 为验证集群签名密钥对，任何第三方可用 vk_V 离线核验种子合法性。注意承诺值必须只包含 $seed_q$ 与公共参数（ε、δ）：A_q、B_q 由 $seed_q$ 派生，若纳入承诺则在承诺阶段即锁定矩阵内容，"先承诺后揭示"的防选择性出题设计将失效。挑战验证的参考判定流程如下（示意伪代码，非最终实现）：

```
def verify_challenge(node, seed, params):
    # 1. 由 VRF 种子派生不可预知的挑战矩阵（规模与精度由 seed 决定）
    A, B = derive_challenge_matrices(seed, params.sizes, params.precisions)
    C_ref = reference_gemm(A, B, precision=params.precision)   # 验证集群参考实现
    # 2. 正确性判定：相对 Frobenius 范数误差 ≤ ε
    ok_math = frobenius_rel_err(node.reply.C, C_ref) <= params.eps
    # 3. 性能合理性：实测耗时不低于物理下界的 (1 − δ) 倍
    m, k, n = A.shape[0], A.shape[1], B.shape[1]
    T_hat = 2 * m * k * n / throughput_baseline(node.tier)
    ok_time = node.reply.T >= (1 - params.delta) * T_hat
    if not (ok_math and ok_time):
        telemetry.flag(node.operator, "L2_FAIL",
                       magnitude=abs(node.reply.T / T_hat - 1))
    return ok_math and ok_time
```

4.3 验证者网络（明确权责与成本）

- **验证集群**：平台自建并运营的核心验证集群，承担全部抽样重执行与基准挑战的生成、下发与判定；其成本计入平台服务费，向租户透明列示。
- **志愿验证节点**：高声誉节点可申请加入验证网络，须质押保证金（金额与拟承担的验证工作量挂钩），按完成的有效验证获得报酬；验证错误或与被验证节点合谋，罚没保证金并永久取消验证资格。
- **随机信标**：挑战种子由验证集群以可验证随机函数（VRF）生成并公示种子承诺，节点无法预知挑战内容，验证集群事后可证明未选择性出题。

4.4 双轨节点接入

模式	适用对象	承载任务	质押要求	SLA
弹性尽力模式	个人工作站、闲置 GPU 、边缘设备	可中断、可重试的批处理、Embedding 、 RAG 数据处理、离线推理	无前置质押	不承诺；任务失败自动重试
专属保障模式	企业 IDC 、专业集群、高可用工作站	在线推理、模型服务、企业 API	收益托管或保证金（按协议）	承诺可用性目标，冗余部署，违约赔付

弹性节点**不承载**任何 SLA 业务——企业级可用性由专属资源池的冗余部署保障，从架构上消除"键鼠一活动就掉线"与"在线服务"的矛盾。

4.5 防女巫与防超售机制

1. **运营者账户体系**：所有声誉、结算与惩罚以运营者（自然人或法人）为单位聚合。同一运营者名下多台设备的表现合并计算，单设备违约影响运营者整体声誉与结算周期，使"注册大量新身份各作恶一次"无法隔离损失。
2. **全局资源租约与跨身份显存互斥：每块 GPU 以「设备序列号 + 运营者账户」为键在平台账本中登记，资源被任务锁定期间，同一键不允许创建第二份可售资源。更关键的是设备序列号在账本层面全局唯一——同一序列号在同一时刻只允许绑定一个运营者账户、至多存在一份有效租约，不论以哪个账户申报；变更序列号的账户绑定须双方复核并设 30 天冷却期。这使「一块物理卡换三个身份、各申报一次全部显存」在账本层面即互斥，而非依赖事后检测才被发现。重复申报同一物理设备（含虚拟化切分后多身份申报）触发账户级审查；L1 声明核验采集的硬件组合（GPU 型号、显存总量、PCIe/NUMA 拓扑、VBIOS 版本与序列号）另用于跨账户查重——仅作查重与复核信号，不作为设备身份证明或永久拉**

黑依据（与 §4.1「三不做」一致）。为使该互斥可被外部观测，跨身份去重统计（申报设备数 / 去重后设备数 / 差额）纳入各阶段验收报告「节点规模」字段一并公开（见 §9.2.1）。

3. **拓扑聚类疑似检测**：对申报资源的 PCIe 拓扑、NUMA 结构、驱动与固件组合做聚类分析，发现"疑似同一物理机的多个身份"时限流观察并触发人工复核。拓扑特征仅用于聚类检测与调度参考，**不作为设备身份证明。误判防护与申诉通道**：合规多机运营者（网吧、院校机房、小型工作室、代理 IDC）天然具备相同硬件与固件组合，可提交设备清单、采购凭证等材料申请**运营者身份复核**（处理时限 7 个工作日，结果公示）；声誉模型区分"判定作弊"（罚没）与"聚簇怀疑"（仅限流观察，不罚没、不设永久降权，复核通过即恢复权重）。
4. **新节点收益成长曲线**：新注册运营者前 30 天实行收益上限与渐进释放，使批量注册新账号的一次性作弊期望收益趋近于零；随诚信记录积累逐步提高上限。
5. **收益延期释放池**：节点收入按 Earned → Pending → Released 三态流转。Pending 观察期内（按任务价值 1–72 小时分档）接受抽检与争议处理；检出违约则罚没相应 Pending 收益，用于赔付受损租户。对高价值 SLA 业务另设保证金，见 §4.4。

声誉的量化模型（设计目标）。 运营者声誉值 $R \in [0, 100]$ 按"指数遗忘 + 任务反馈"更新：

$$R_{t+1} = \lambda \cdot R_t + (1 - \lambda) \cdot r_{task}$$

其中 $r_{task} \in [0, 100]$ 为单任务评价（综合验证通过情况、性能偏差、SLA 达成），$\lambda \in (0, 1)$ 为遗忘因子——λ 越接近 1，历史权重越大、偶发抖动被平滑得越多，但恢复越慢；λ 越接近 0，新评分影响越大、对近期变化越敏感，恢复也越快。该设计使声誉在"稳定性"与"响应性"之间取得平衡：长期良好记录可平滑偶发抖动，持续劣化则使 R 单调收敛到近期表现水平。违约额外叠加惩罚倍数：每次确认的违约事件使 R 直接乘以衰减系数 κ（取值 0.8，可由风控规则调整），并被记入运营者级审查流程。声誉 R 的调度权重与准入映射关系（如 $R < 60$ 禁入 SLA 资源池）作为设计目标在阶段 0 内测中校准后公示。

罚没函数借鉴 Filecoin 的抵押罚没（Collateral / Slashing）模型，但罚没上限严格不超过运营者托管保证金，避免出现"负资产"账户：

$$Slash(o) = min(m\_violation \times V\_disputed, D\_escrow(o)); \quad R\_o \leftarrow \kappa \cdot R\_o$$

其中 $V_{disputed}$ 为争议任务价值，$m_{violation}$ 为责任认定倍数（初值 1.0，重复违约累进），$D_{escrow}(o)$ 为运营者托管保证金——罚没额以保证金为硬上界，超出部分由平台滥用赔付基金承担，运营者不承担无限责任。

（收益托管与罚没的合约级示意见 §8.6 协议层——该能力属阶段 4 交付物，放在协议层章节统一呈现，避免被误读为当前已交付。）

4.6 动态能力评级（CU）

节点综合能力评分定义为：

$$CU(i) = w\_P \cdot P(i) + w\_B \cdot B(i) + w\_M \cdot M(i) + w\_N \cdot N(i) + w\_R \cdot R(i)$$

其中 P 为计算基准得分，B 为显存带宽得分，M 为显存容量得分，N 为网络质量得分，R 为历史可靠性得分；权重 w_P、w_B、w_M、w_N、w_R 随任务类型调整（训练任务提高 P/N 权重，在线 API 提高 N/R 权重），满足 w_P+w_B+w_M+w_N+w_R=1。各分项均来自 L2 实测与 L3 遥测，而非厂商标称值。

为保证异构硬件间的可比性，各分项先做 min–max 归一化到 [0, 1]：

$$X'(i) = (X(i) - X\_min) / (X\_max - X\_min)$$

其中 X_min / X_max 取自同代硬件档位的实测分布分位数（P5 / P95），以抑制离群值拉伸。归一化得分经上式加权求和得到 CU(i) ∈ [0, 100]，再按校准后的切点划分 Tier-S/A/B/C：切点选取遵循"梯队内 CU 离散度最小、跨梯队任务通过率差异最大"的原则，并在阶段 0 内测中以实际任务通过率回归校准后公示（校准前以硬件规格表作为临时划分依据）。

节点按实测 CU 分为 Tier-S/A/B/C 四级梯队，梯队与可承载任务范围在 §5.4 的调度约束表中给出，任一梯队的任务表与硬件规格保持可验证的一致性。

4.7 节点主（供给方）保护与内容合规

平台对节点主承担的义务与租户义务对等：

- **最小权限执行**：租户任务运行在容器隔离环境（cgroups v2 + Seccomp + 只读根文件系统），Node Daemon 以非特权用户运行；隔离可显著缩小攻击面但不能绝对消除风险，平台为此投保并设立滥用赔付基金。
- **用途审查**：租户注册须完成 KYC；任务镜像经白名单仓库分发，禁止节点拉取任意外部镜像；加密货币挖矿、端口扫描、垃圾流量等用途列入禁运清单并由遥测检测。
- **滥用赔偿**：因平台风控疏漏导致节点主损失的，按协议从滥用赔付基金中补偿。
- **退出自由**：节点可随时退出，退出不影响已进入 Released 状态收益的提取。
- **机密计算前瞻（研究项）**：针对"租户代码与权重在节点内存中的隐私保护"，跟踪基于可信执行环境（TEE）的机密计算路线——即支持 CC（Confidential Computing）模式的数据中心 GPU，可在硬件层面加密显存与计算过程，使节点主自身也无法读取租户数据。该能力依赖特定硬件世代与厂商支持，列入研究地平线（见 §9.3），在生态内达标设备形成规模前不作为承诺项；在此之前，容器隔离 + 用途审查 + 赔付基金仍为主要防护组合。

5.1 硬件适配层（HAL）——诚实的异构策略

A12 通过硬件抽象层统一**任务提交接口**，而非声称消除底层差异。明确表述：

- **接口统一**：租户以统一任务描述（模型、数据、算力需求、时延与预算约束）提交，无需选择具体 GPU 型号。
- **后端异构**：底层执行依赖各厂商的驱动、编译工具链与算子实现。**路线图承诺 NVIDIA CUDA 栈生产可用；AMD ROCm 栈以实验性支持交付**；国产芯片适配列为研究项，按实际工程进度另行公告。
- 平台不使用任何与第三方商标近似的产品名。

5.2 双路径调度

调度器依据任务的通信—计算比 α_comm 做路径选择：

$$\alpha\_comm = T\_comm / T\_compute$$
$$= （跨节点通信数据量 / 链路有效带宽）\div （计算量 / GPU 实际吞吐）$$

- α_comm ≥ 1（通信受限）：调度至**集中化算力中心**（云平台 GPU 集群、专业 AI IDC、HPC）——A12 通过标准适配器接入这些资源并提供统一编排，替代租户自行比价与拼接。
- α_comm < 1（计算受限）：调度至 **A12 分布式资源池**（经核验的 IDC、工作站与边缘节点），执行本地闭环任务（推理、Embedding、LoRA/QLoRA、批处理），避免广域网高频同步。

α_comm 的两个时间分量采用经典 Hockney 通信模型与 Roofline 吞吐估计建模：

- **通信时间**：T_comm ≈ T_startup + m / BW + o(m)，其中 m 为单步同步需传输的数据量（如全参训练中约等于待同步参数字节数 × 同步次数），BW 为链路有效带宽（实测，非标称），T_startup 为固定启动延迟（广域网典型 10–100ms，NVLink/IB 典型 < 10μs），o(m) 为协议栈附加项；
- **计算时间**：T_compute = FLOPs / η_eff，其中 FLOPs 为单步浮点运算量（全参训练近似 6 × 参数量 × token 数），η_eff 为实测有效算力吞吐（roofline 模型下受算术强度约束，通常为峰值算力的 30%–60%）。

以 7B 参数模型全参微调为例：数据并行下每步需同步约 28GB（FP16 权重 + 优化器状态按实现而异），在 25Gbps 广域网链路（有效带宽 ≈ 3GB/s）上单步通信 ≈ 9.3 秒；同样的同步量在 InfiniBand 400Gbps（有效 ≈ 45GB/s）集群内 < 1 秒——通信时间相差一个数量级，而该规模下单步计算仅数秒。这正是"高通信任务必须留在集中式集群"的量化依据。需要强调的是：25Gbps 属专线 / 数据中心互联量级；按家庭上行 1Gbps（有效 ≈ 110 MB/s）重算，同一 28GB 同步量约需 **265 秒/步**，较上述示例高近 30 倍——该结论在真实边缘带宽分布下远比示例更极端。调度器的 α_comm 参数库按目标供给端带宽分布（上行

P50/P90、跨境时延）构建，并对估算误差保留容错：阈值两侧的边界任务以实测复核决定归属。调度器在任务提交时估算 α_comm 并落库，与任务实际执行的观测值回归比对，持续修正带宽与吞吐参数库。五类典型任务的分流估算如图 5-1 所示：

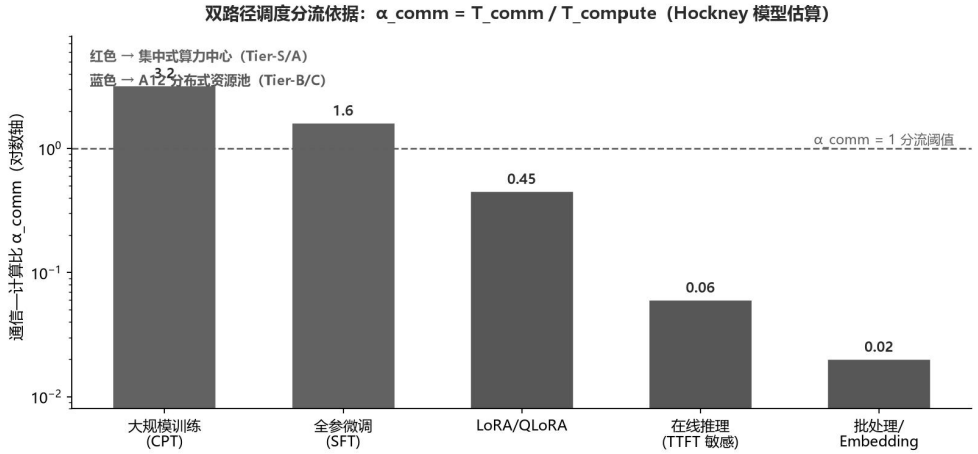

图 5-1 双路径调度分流依据：α_comm 阈值两侧分别为集中式集群与分布式资源池

图 5-1 数据复算说明：本图的五类任务分流结论由统一参数表支撑，参数表已随本版在附录 D 表 D-1 给出（追踪编号 D-21，状态：已补充待佐证），含各任务的通信—计算比 α_comm 估算区间（逐带宽档列示）、T_comm 与 T_compute 取值及硬件档位假设、分流阈值与边界任务实测点，读者可按 α_comm = T_comm ÷ T_compute 逐档独立复算。α_comm 区间为按 §5.2 Hockney 通信模型与 Roofline 吞吐估计得到的设计估算值，随阶段 0 实测样本扩充按季度复核更新，更新记录随公示页留档。

5.3 调度决策与 SLA 模型

调度评分为多目标加权：

$$Score(n,t) = w1 \cdot P\_compute + w2 \cdot P\_memory + w3 \cdot P\_network + w4 \cdot P\_reliability + w5 \cdot P\_SLA - w6 \cdot C\_compute - w7 \cdot C\_network$$

权重按任务类型动态调整。调度求解在工程上简化为**约束过滤 + 贪心评分**两步：先按硬约束（梯队准入、SLA 准入、显存/拓扑匹配、声誉阈值）过滤候选集，再对剩余候选按 Score 排序取最优，整体复杂度 O(N log N)，在千级节点规模下单次决策毫秒级完成；多目标权重的长期优化列为研究项，不作为首版承诺。

SLA 可用性的冗余数学。 设单节点月度可用性为 A_0 ，主备双节点独立部署、独立供电与网络出口，则服务整体可用性为：

$$A = 1 - (1 - A_0)^2$$

以 A_0 = 99.5% 为例，双机冗余后 $A = 1 - 0.005^2 = 99.9975\%$，即月度累计不可用时间从约 3.6 小时压缩到约 1.1 分钟。平台据此做容量规划：目标可用性 A* 反解出所需单节点可用性 $A_0^* = 1 - \sqrt{(1 - A^*)}$，

仅当专属资源池中满足 A_0 ≥ A_0 * 的主备对数量 ≥ 已售 SLA 订单数 + 安全余量时才接受新订单——余量不足时拒绝新 SLA 订单而不是超卖，这是 §4.4 双轨隔离在容量层面的落实。需强调：上述 99.5% 与 99.9975% 均为设计基线口径的示例计算，同样属于设计目标而非实测结论；示例用于说明冗余合成的数量级关系，其中的 99.9975% 不是对外承诺值（对外承诺下限见下文 SLA 目标值框架）。

控制平面可用性（A_platform）。 上式仅覆盖资源节点侧冗余；调度、计量、结算与验证集群等控制平面组件由平台自建自营，其可用性单独建模：控制平面按多可用区部署，WSS 网关多活；验证集群不可用时，挑战下发自动降级为抽检模式，Pending 收益释放顺延且不计节点违约。对外承诺的服务级可用性 = A_platform × A_node，其设计目标值见下文《SLA 目标值框架》，正式承诺值随正式发布版一并公示并登记为追踪编号 D-23。控制平面故障期间的计费规则：故障时段暂停计费，已达 SLA 订单按赔付条款执行——租户不为平台故障付费。

SLA 目标值框架（阶段 2 对外销售的前置条件）。 对外承诺的服务级可用性目标值随正式发布版一并公示；本文先行给出设计目标：专属保障模式服务级可用性 A_service ≥ 99.5%，其中控制平面 A_platform ≥ 99.9%、节点侧合成可用性 ≥ 99.6%（按 A = A_platform × A_node 合成）。三项口径明确如下：其一，99.5% 是对外承诺下限，而非预期运行值——容量规划按上式的 A* 反解 A_0 * 逻辑执行；它与前文 99.9975% 的示例并不冲突，后者是设计基线 A_0 = 99.5%（月度口径）下的理论合成值，下限与理论合成值之间的差额即为对外承诺的安全边际。其二，A_0 的设计基线取月度 99.5%（专属保障模式主备节点）；按 $A = 1 - (1 - A_0)^2$ 反解，对外承诺的节点侧合成可用性 ≥ 99.6% 对应 A_0 * ≥ 93.68%，该 93.68% 是对外承诺下限对应的反解值，不是单节点的设计基线，实际容量规划以 99.5% 基线执行。其三，统计窗口：A_service、A_platform 与 A_node 均按自然月统计（月度窗口），年度可用性取 12 个自然月的算术平均，SLA 赔付以自然月为核算周期。上述全部数值均为设计目标而非实测结论；在目标值正式公示之前，平台不对外签署含 SLA 的服务合同——阶段 2 的 SLA 商业化以该公示为硬性前置；平台故障期间「故障不计费」的计费与赔付规则随目标值同步公示。

供给端经济性与专属池余量处置。 弹性模式下单台 RTX 4090 级设备的月度收入区间、电费与折旧占比、与「闲置 / 转售」机会成本的对比测算表【发布前公示 · D-24】——该测算是 §9.2 节点留存与规模目标的第一性依据。专属池达标主备对长期高于已售订单时，闲置容量转入弹性池承接可中断任务，或按容量预留协议与企业租户结算，避免单侧闲置。

供给端经济性测算框架（先结构、后数值）：测算表按统一字段结构公示，字段的单位与取值口径约定如下——① 设备型号与档位：以实测 CU 分档列示（如 RTX 4090 级、24 GB 显存），并标注样本节点数；② 电价区间：人民币元 / 千瓦时（含税），按供给方所在国别与居民 / 商业电价分档列示，跨境样本同时标注币种与汇率口径；③ 硬件折旧口径：直线法，折旧年限 36 个月，残值率 5%（与设备档位一并公示，变

更时同步说明理由）；④ 月度可承载任务量：CU·小时 / 月，取同档位节点实测 P50，并列 P90 供对照；⑤ 任务单价区间：美元 / CU·小时，采用 §5.3 公示基础价口径，不含质量加成；⑥ 节点月度收入区间：美元 / 月，分别列示平台分成前与分成后两个数值；⑦ 机会成本对比：闲置情形计 0；转售情形按二手市场均价折算的月度等效收益（美元 / 月）列示。数值随正式发布版一并公示。§9.2 中阶段 1 的节点留存与规模判据，以该测算表公示版为生效复核依据；测算表未公示前，相关判据仅作为设计目标。

调度决策的参考实现如下（约束过滤 + 贪心评分，示意伪代码）：

```
def schedule(task, pool):
    # 硬约束过滤（梯队准入 / SLA 模式 / 显存 / 声誉阈值），复杂度 O(N)
    cand = [n for n in pool
            if n.tier in TIERS_FOR(task.kind)
            and (not task.sla or n.mode == "DEDICATED")
            and n.free_vram >= task.vram
            and n.reputation >= R_MIN(task.sla)]
    if task.sla and sla_headroom(pool) < SLA_MARGIN:
        return Rejected("SLA 余量不足：拒单不超卖（见 §4.4）")
    # 贪心评分取最优，复杂度 O(|cand| · log |cand|)
    return max(cand, key=lambda n: score(n, task), default=Requeued(task))
```

调度定价机制：任务定价采用"公示基础价 + 质量加成"结构——基础价随供需按公开规则浮动（参照 EIP-1559 的价格发现思路，见 §3.4），质量加成按资源梯队（Tier-S/A/B/C）与 SLA 档位加收。全部价格参数公示可查，杜绝暗调。

5.4 资源梯队调度约束表

梯队与可承载任务范围的对应关系如下（与 §4.6 的 CU 分级一致）：

梯队	典型硬件（实测达标）	可承载任务	承载 SLA
Tier-S	高端数据中心 GPU + 高速互联（NVLink/IB）	分布式训练、CPT、大规模全参微调	专属保障模式
Tier-A	数据中心级单机（A100/H100 级）	大参数在线推理、长上下文服务、多任务 SFT	专属保障模式
Tier-B	专业工作站（RTX x090 级及以上）	常规在线推理、LoRA/QLoRA、Embedding	弹性尽力为主，达标可申请专属
Tier-C	消费级 GPU、边缘设备	批处理、RAG 数据处理、离线推理、轻量推理	仅弹性尽力模式

说明：梯队以 L2 实测 CU 得分划分而非型号准入；同一物理节点实测不达标即降级，实测超标可升级；"能力升级"与 §4.2 的"虚报判定"互斥——虚报仅在实测低于物理下界 $(1-\delta)\cdot T\_hat$ 时触发，超基线性能经 CU 分档正常体现（防误判机制见 §4.2）。

5.5 任务生命周期与争议处理

任务状态机：Pending → Scheduled → Running →（Suspended）→ Completed / Failed / Disputed。争议任务冻结对应 Pending 收益，进入交互式重放定位：在验证集群上重放任务切片，定位责任方（节点、平台或租户输入），按责任划分赔付。

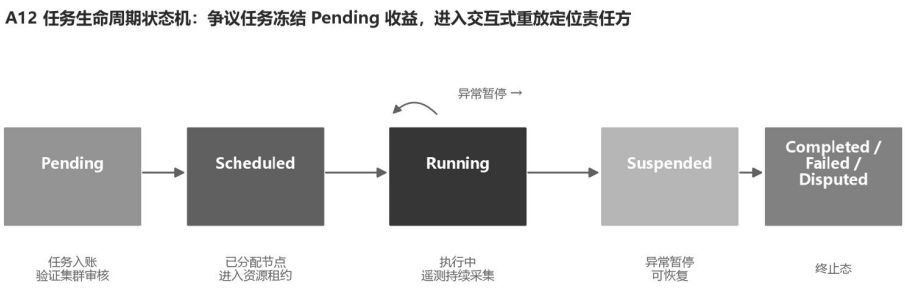

图 5-2 A12 任务生命周期状态机：争议任务冻结 Pending 收益，进入交互式重放定位责任方

5.6 面向租户的控制台与安全沙箱运行时

为提供不逊于传统云平台的租户体验，Layer 2 封装了租户交互层与安全执行环境。以下均为设计目标，验收指标随阶段 0 内测公开：

1. **Web 终端与文件通道**：节点注册后主动向调度网关建立出站加密 WebSocket（WSS）连接，形成反向代理隧道；租户控制台以 Web 终端仿真组件直连节点，免除公网 IP 端口映射与密钥分发流程，兼容任意 NAT/防火墙环境。控制台同时代理在线 Notebook 服务与文件存储服务，上传采用 HTTP 分块与断点续传，保证大容量数据集与训练脚本的可靠传输。
2. **容器化秒级实例调度**：不采用传统虚拟机模式，统一以轻量级容器运行时作为沙箱。平台预置精简的基础镜像（Ubuntu LTS + CUDA + PyTorch 推理栈 + 常用依赖），对镜像只读层在闲置时段预加载；任务启动时仅实时拉取数兆字节的定制层。**设计目标：单实例冷启动时间 ≤ 10 秒**（同地域、镜像缓存命中档位；跨区或缓存未命中场景按节点上行带宽分档另列目标，实测数据以公开验收报告为准）。
3. **物理级配额隔离**：以 cgroups v2 与 Seccomp 对容器的 CPU、内存与 IO 带宽做硬限制；通过 NVIDIA Container Runtime 保证容器内仅可见被分配的 GPU 设备。该隔离可显著缩小攻击面并防止旁侧进程读取租户内存，但不承诺绝对杜绝容器逃逸（残余风险见第 10 章），平台以滥用赔付基金兜底。

第六章 Layer 3：模型服务（MaaS）

6.1 模型接入与版本管理

统一 Model Registry 记录模型来源、版本、权重摘要、运行环境与评测结果；企业私有模型仅托管、不改变知识产权归属。该承诺以 §4.7 安全机制与 §5.6 隔离措施为边界：弹性节点模式下 IP 保护依赖容器隔离

（不承诺绝对杜绝容器逃逸，残余风险由 §4.7 投保与滥用赔付基金兜底）；对 IP 隔离有强需求的租户建议使用专属保障模式，或采用 §4.7 前瞻的 TEE 机密计算档位（研究项）。模型上架须提交模型卡（Model Card），记录预期用途、已知局限与评测基准，与平台计量系统关联。

6.2 垂直行业模型流水线

平台规划以"一个垂直方向打穿"的方式建立模型层竞争力，**首个垂直方向为 Web3 代码与安全工程**（与已披露的团队背景一致；本节全部指标为规划目标，非已完成事实）：

1. **语料治理：构建结构化数据采集管道，对开源合约仓库、公开漏洞披露社区与链上已验证合约源码进行采集与许可筛查；经语法树过滤、近似去重与人工抽检后形成预训练语料。规模目标：首批预训练语料 ≥ 10B tokens。许可证治理：建立许可证白 / 黑名单——宽松许可（MIT / Apache / BSD）优先收录；GPL / AGPL 等 copyleft 许可证语料的收录边界、以及"训练权重是否构成衍生作品"的法律口径，以专项法务意见为准【法务意见：发布前补充并公示 • D-25】；去重与许可筛查规则随语料管线文档公开。10B tokens 为首批规模，模型规模按 Chinchilla 标度律与语料量同步规划（效率优先的中小参数档位）；后续批次的数据预算与时间表【发布前补充并公示 • D-26】。**
2. **持续预训练（CPT）与监督微调（SFT）**：以开源代码基座模型为起点做领域持续预训练，再叠加"缺陷定位 → 攻击路径构造 → 后果预测 → 修补建议 → 回归测试"多步推理指令集做监督微调。训练在集中化算力中心（Tier-S 梯队）执行，遵循第 5 章双路径规则；训练规模（实例数、并行策略、上下文长度）随阶段预算公示，不做固定承诺。
3. **工具调用对齐（DPO）**：模型输出接入外部工具沙箱（编译测试框架、静态分析工具），以工具执行结果构建偏好数据对做直接偏好优化，降低"幽灵漏洞"式误报。该机制为确定性改进项：**模型结论必须可被工具复核。**

两个训练阶段的目标函数如下（设计目标口径）。

SFT 阶段采用掩码交叉熵损失：仅对助手回复部分计算损失，提示词部分掩去：

$$L_{SFT}(\theta) = -\frac{1}{|y|}\cdot\sum_{t=1}^{|y|}\log p_\theta(y_t \mid x, y_{<t})$$

其中 x 为输入（含多步推理指令模板），y 为目标输出序列，$|y|$ 为目标序列长度，p_θ 为参数 θ 下的模型条件概率。多步推理指令集（缺陷定位 → 攻击路径构造 → 后果预测 → 修补建议 → 回归测试）以结构化模板组织，使损失集中在推理链与结论的生成质量上。

DPO 阶段以工具执行结果构造偏好对 (y_w, y_l)——y_w 为通过工具复核的输出（编译通过、测试通过、静态分析确认），y_l 为被工具否定的输出——在参考模型 p_{ref} 约束下优化：

$$L_{DPO}(\theta) = -\mathbb{E}_{(x, y_w, y_l)}\left[\log\sigma\left(\beta\cdot\log\frac{p_\theta(y_w|x)}{p_{ref}(y_w|x)} - \beta\cdot\log\frac{p_\theta(y_l|x)}{p_{ref}(y_l|x)}\right)\right]$$

其中 σ 为 sigmoid 函数，β 为偏离参考模型的强度系数（典型 0.1），隐式替代了独立的奖励模型训练。该路线的工程意义在于：偏好信号**不依赖人工标注的主观判断，而以工具执行的客观结果为锚**，使"降低幽灵漏洞误报"成为可复现、可回归测试的优化目标，而非一次性调参结果。

训练规模规划：Chinchilla 标度律。 垂直模型训练预算由目标损失 L 与模型参数量 N、训练 token 数 D 共同决定，遵循 Chinchilla 类经验形式：

$$L(N, D) = E + A / N^{\alpha} + B / D^{\beta}$$

其中 E 为不可约误差，A、B、α、β 为与任务相关的经验系数。该式说明：单纯扩大模型而不增加数据，或单纯增加数据而不扩大模型，都会进入边际收益递减区；平台据此做"模型规模 ↔ 语料规模 ↔ 算力预算"的三元规划，并随阶段预算公开训练规模（实例数、并行策略、上下文长度），不做固定承诺。

合规边界：模型能力定位为**辅助参考工具**，输出用于开发辅助与内部安全评估；不对外出具带 CVSS 评级的正式审计报告，该类服务须由持资质合作方出具。接洽流程已启动，目标**公测启动前完成至少一家持资质合作方的签约并披露名称**；签约前该产品线不对外正式商用，合作进展按阶段在官网更新（受第 10 章前瞻性声明约束）。

6.3 OpenAI 兼容 API 网关

网关严格遵循 OpenAI 接口标准，实现存量应用零代码迁移：

- **标准端点**：提供向后兼容的 /v1/chat/completions、/v1/models 与 /v1/embeddings，支持服务端事件（SSE）流式返回；已集成 OpenAI SDK、LangChain 或 LlamaIndex 的应用仅需替换网关地址与密钥。
- **动态负载均衡**：网关按各资源梯队推理队列深度分流——长上下文高负载任务路由至 Tier-A 及以上资源，高频短任务路由至边缘资源池。**设计目标：公测期首字延迟（TTFT）中位数 ≤ 400ms（≤ 8K 上下文档位；32K 长上下文档位随公测实测数据单独设定目标）。**
- **负载均衡算法**：同梯队内采用"二次选择（Power of Two Choices）"策略——随机抽取两个候选实例，将请求派往队列深度较小者。该策略无需全局最小值检索，通信开销 O(1)，却能把队列长度从全局随机的指数分布压到多项式衰减：经典排队论结论是，当系统负载接近饱和时，二次选择使平均排队长度从 O(n) 量级降到 O(log log n) 量级。相比"始终选最空实例"的全局扫描方案，在千级实例规模下以极小的长尾代价换取了调度器自身的低延迟与水平扩展性，与网关 TTFT 目标匹配。

排队论容量规划。 请求到达近似泊松过程、单实例服务率 μ、并发实例数 c 时，排队等待时间由 Erlang C 公式给出（M/M/c 模型）：

$$TTFT \approx T\_prefill(s) + W\_q, \quad W\_q = C(c, \rho) / (c\mu - \lambda), \quad \rho = \lambda / (c\mu)$$

其中 T_prefill(s) 为预填充阶段计算耗时（随上下文长度 s 线性增长），C(c, ρ) 为 Erlang C 等待概率。该模型用于两个工程决策：其一，给定 TTFT 目标（400ms）反解各梯队所需最小并发实例数 c；其二，在队列深度逼近饱和（ρ → 1）前触发扩容——这正是负载均衡器以队列深度为路由信号的理论依据。公测期以实测到达率 λ 校准 μ 后公示容量规划表。

6.4 推理性能工程：连续批处理与冷启动优化

- **迭代级连续批处理**：以迭代为粒度动态插入与退出请求，请求结束后即时回收其显存页，新请求进入就绪队列等待下一迭代插入，消除静态批处理的"木桶效应"。**设计目标：GPU 有效负载率 ≥ 75%**（以公开验收报告实测为准）。
- **显存分页管理**：采用分页式键值缓存（PagedAttention 类方案），将 KV 缓存切分为固定长度显存页并以页表维护逻辑映射，支持不连续物理分配，降低碎片率；支持多轮对话分支共享前缀显存块。
- **冷启动缓存**：在各资源池节点预加载基础模型的只读权重；用户自定义 LoRA 权重（通常仅数十 MB）经分布式缓存就近拉取与热切换，支持同一 GPU 上多租户、多模型的无缝切换。

KV 缓存容量规划（设计目标）。 分页式管理的容量需求按下式估算：

$$M\_KV = 2 \times L \times H\_kv \times d\_head \times s \times b \times bytes\_per\_elem$$

其中 2 为 K/V 两份缓存，L 为层数，H_kv 为键值头数（采用 GQA 分组查询注意力时 H_kv ≪ 注意力头数，可显著压缩缓存），d_head 为每头维度，s 为序列长度，b 为并发批大小，bytes_per_elem 为精度字节数。以 80 层、H_kv = 8、d_head = 128、FP16 的模型承载 32K 上下文为例：单序列 KV 缓存 ≈ 2 × 80 × 8 × 128 × 32768 × 2 B ≈ 10.7 GB——这一量级说明 KV 缓存而非模型权重本身，常常是长上下文服务的显存瓶颈。分页机制将碎片率上界压到"每序列不足一个页"（典型页大小 16 tokens，理论浪费 ≤ 页大小 × 序列数），并以前缀共享使多轮对话与相同系统提示词的请求复用同一批物理页。

连续批处理的吞吐模型。 GPU 有效负载率定义为目标算子实际执行时间占比：

$$Util = T\_busy / (T\_busy + T\_idle + T\_overhead)$$

静态批处理中 T_idle 来自"快请求等待批内最慢请求"（木桶效应），批越大浪费越重；连续批处理以迭代为粒度重组批次，使 T_idle 趋近于零，残余开销只剩调度本身（T_overhead，毫秒级）。结合 KV 缓存按需分配，调度器可在"批大小上限 × 序列长度上限"构成的显存包络内动态最大化并发序列数——这是 75% 有效负载率设计目标（以公开验收报告实测为准）的机制来源，而非拍脑袋数字。

Roofline 性能模型。 单卡推理峰值吞吐受算术强度 I（FLOPs/byte）约束：

$$T\_eff(I) = min(\pi\_peak, \beta\_mem \cdot I)$$

其中 π_peak 为峰值算力，β_mem 为显存带宽。低算术强度任务（如解码阶段多为 memory-bound）的实际吞吐由 β_mem·I 决定，而非峰值算力——这正是 KV 缓存压缩、低精度推理、PagedAttention 与连续批处理等技术追求"降低每 token 显存访问量"的理论依据。 **性能工程前瞻（研究项，见 §9.3）。** 推理优化社区的前沿方向将作为 MaaS 层的持续演进路线跟踪，包括：

1. **投机解码（Speculative Decoding）**：以小模型起草、大模型并行验证的生成方式，在输出质量不变的前提下将解码延迟压缩数倍；其草稿—验证结构与 A12 的验证集群架构天然契合；
2. **低精度推理（FP8 / FP4）**：以硬件原生低精度格式承载 KV 缓存与矩阵乘，等比压缩显存占用与带宽需求——按 §6.4 的 M_KV 公式，FP8 化可直接使同卡可承载的并发长上下文序列数翻倍；
3. **Prefill/Decode 分离架构（Disaggregated Serving）**：把首字生成（计算密集）与后续解码（带宽密集）拆分到不同资源池，与 A12"按通信—计算比分流"的调度哲学同构；
4. **KV 缓存分级卸载**：HBM → DDR → NVMe 的三级缓存层级，以温度感知的页面迁移延长可服务上下文长度；
5. **语义缓存（Semantic Caching）**：对语义重复的请求直接复用缓存答案，在知识库问答类负载上显著降低算力消耗；
6. **MoE 专家并行**：混合专家模型的专家权重跨节点分布与动态路由，使"百亿参数级模型在分布式资源池上服务"成为可能——这正是分布式算力网络区别于单机方案的差异化机会。

其中，GQA 与低精度推理两项机制对显存占用的量化效果如图 6-1 所示（其余机制的收益以容量上限与吞吐提升形式体现，见第 5 章）。

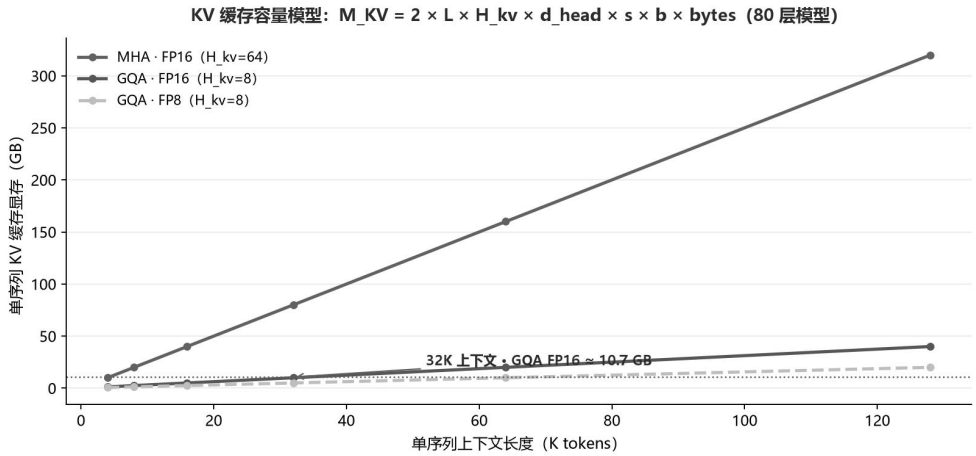

图 6-1 KV 缓存容量模型：GQA 与低精度对显存占用的压缩效果（80 层模型，单序列）

图 6-1 数据复算说明：本图的显存压缩对比由统一参数表支撑，参数表已随本版在附录 D 表 D-2 给出（追踪编号 D-22，状态：已补充待佐证），含模型档位（7B / 70B 级两档，各自参数自洽）与上下文长度、

GQA 分组配置、分页式 KV 缓存块大小与命中率口径、压缩前后的单序列 KV 显存占用与可承载并发序列数（两档分列），读者可据此独立复算；MHA 行为无 GQA 压缩的对比基线行。命中率采用块级口径：在一个自然日（UTC+0）统计窗口内，被复用的 KV 物理页读取次数 ÷ KV 物理页总读取次数；前缀共享复用计入命中，请求级与 Token 级统计口径不采用。表中数值为按 §6.4 M_KV 公式得到的设计估算值，随阶段 0 实测样本扩充按季度复核更新。

6.5 模型市场与计量结算

- 按调用次数、Token 数、算力时长或订阅计费；结算默认以法币或稳定币计价，A12 Token 作为手续费折扣与激励凭证（见第 8 章）。
- 模型开发者按调用收入分成，分成比例在模型上架协议中约定并公示；计量记录可导出审计。

第七章 Layer 4：智能体与应用（AaaS）

7.1 可视化 Agent 编排画布

面向不具备工程化开发能力的用户，提供低代码拖拽式智能体编排系统：

- **可配置节点**：触发类（Webhook、定时、事件监听、对话框）、推理类（调用 Layer 3 模型或外部模型，参数可实时调节）、工具类（代码执行、SQL 查询、外部 API、静态分析）、控制类（分支、循环、解析、失败恢复与补偿）。
- **工作流 DSL 编译**：前端画布渲染为有向图，编译器校验连通性与循环依赖后转换为声明式工作流定义（JSON/YAML），作为可版本化资产保存。
- **DAG 执行引擎**：后台无状态执行引擎按拓扑排序调度节点，无依赖节点并发执行，有依赖节点经上下文状态树传递中间结果；执行记录全程留存，供争议回放。

执行调度的数学性质如下：工作流编译为有向无环图 G = (V, E) 后，引擎以 Kahn 算法做拓扑排序——维护各节点入度表，入度为零的节点进入就绪队列并发执行，每完成一个节点将其后继入度减一，整体复杂度 O(|V| + |E|)；图含环时存在入度永不为零的节点，编译器据此在**提交期**拒绝循环依赖，而非在执行期失败。调度并行度受限于图的最大反链宽度（宽度 w 的图任何时刻至多 w 个节点可并发），而**整条工作流的完成时间存在下界：关键路径长度**，即从源点到汇点的最长加权路径——由各节点预估耗时加权求得。编排画布据此向用户展示关键路径高亮与预估完成区间，把"哪些节点拖慢了整体"从经验判断变为可计算结论。

DAG 拓扑排序的参考实现（Kahn 算法）：

```
def kahn_toposort(graph: DAG):
     in_deg = {v: 0 for v in graph.nodes}
     for u in graph.nodes:
         for v in graph.edges[u]:
             in_deg[v] += 1
     queue = [v for v, d in in_deg.items() if d == 0]
```

```
    topo = []
    while queue:
        v = queue.pop(0)
        topo.append(v)
        for w in graph.edges[v]:
            in_deg[w] -= 1
            if in_deg[w] == 0:
                queue.append(w)
    if len(topo) == len(graph.nodes):
        return topo
    raise CycleError
```

- **工具协议标准化（前瞻）**：工具类节点将逐步对齐 MCP（Model Context Protocol）等开放工具接入协议，使外部工具厂商"一次适配、多平台可用"；多 Agent 工作流之间探索基于 A2A（Agent-to-Agent）式协议的任务委托与状态互认，为跨组织的智能体协作预留标准接口。上述均为设计目标，以实际协议版本成熟度为准。

7.2 开发者工具链：VS Code 插件与 Web IDE

- **代码辅助插件**：基于语言服务器协议（LSP）的编辑器扩展，连接 Layer 3 推理网关，在编写代码时按上下文（Fill-in-the-Middle）提供补全、文档注释与单元测试生成建议。
- **内嵌静态检查**：插件内嵌轻量 AST 解析器做实时扫描，对常见危险模式给出内嵌告警与一键修复建议；告警规则与模型输出一样定位为**辅助参考**，不替代人工审计与正式审计报告。
- **云端 Web IDE**：控制台内嵌浏览器版 IDE，可一键启动隔离容器环境执行编译与测试，与§5.6 的沙箱运行时复用同一套隔离与配额机制。

7.3 预置行业 Agent 矩阵

平台规划预置三类可开箱部署的商业 Agent 模板（上线计划见第 9 章阶段 3），均以§7.1 的编排系统实现、以第 6 章模型服务为推理后端：

1. **代码与安全评估助手**：多角色协作工作流——协议结构分析、基于公开漏洞数据的用例生成、沙箱内动态验证、结构化结果汇总。工作流底层支持 ReAct、CoT 等主流推理模式与模型路由，按输入复杂度自动平衡成本与准确率。**输出为内部参考报告，不含 CVSS 正式评级**；正式审计由持资质合作方出具（暂未确定，确认后披露）。
2. **链上数据研报 Agent：监测去中心化交易所交易量、借贷协议锁仓量异动与巨鲸地址变化，结合时间序列分析定期产出定量研报；数据源与口径公开可查。研报类 Agent 仅接入具备再分发授权的公开数据源；商业数据供应商接入前完成许可协议签署，授权状态随产品文档公示【数据源清单与授权状态：发布前补充并公示 • D-27】。**
3. **企业知识库问答 Agent**：向量数据库 + 混合检索 + 多租户隔离，面向企业与社群的 FAQ 与知识问答；答案附来源溯源，溯源失败的问题按"拒答"处理。

7.4 开放平台与统一收银台

- **开放 API 网关**：向第三方 SaaS 供应商（IDE 厂商、安全工具厂商、数据终端等）提供加密 API 接入，支持按调用次数、并发量与席位的权限与配额管理，提供账单明细导出。
- **混合收单：统一收银台当前的受理范围为稳定币（USDT/USDC）；法币支付通道须在取得必要许可、且收单主体与支付服务商确定并公示之后方才开通，在此之前统一收银台不并行提供法币入金通道。通道开通后，租户可按需要在法币通道与稳定币通道之间选择；所有交易经平台统一计量与对账。法币收单涉及支付服务与资金存管安排，与 §2.4 同口径：在取得必要许可前不开展相应业务【收单主体与支付服务商：确定后公示 • D-28】。链上自动兑换与 drip 分红等表述不再使用——链上结算仅作为 §8.6 协议层的可选能力，在监管评估通过后另行公告。**

法币收单可执行性安排：在收单主体与支付服务商确定并公示之前，平台不开通法币入金通道，统一收银台仅受理稳定币（USDT/USDC）结算；企业客户的收款、退款与对账路径，以公示页披露的持牌收单主体及其服务条款为准。收单主体公示已列入披露闭环追踪表（§2.2、附录 C 编号 D-28），是面向企业客户开通法币通道的唯一前置。

7.5 AaaS 结算与分账

结算链路：用户任务 → Agent 服务 → 模型与工具调用 → 算力执行 → 统一计量 → 收入按贡献在 Agent 开发者、模型方、算力方之间分账。分账规则公开，记录可导出审计。

第八章 Token 经济模型

8.1 设计原则

1. **真实收入优先**：平台价值的锚是服务收入，不是 Token 流通市值。
2. **固定总量**：A12 初始发行 100,000,000 枚，永不增发。
3. **激励与结算分离**：服务结算以法币/稳定币计价；Token 用于激励发放、手续费折扣与治理建议，租户不被迫持币。
4. **全部数学可核算**：本章每个机制附约束条件与示例，不存在依赖"未公布参数"的隐藏矛盾。

8.2 代币分配模型

分配对象	比例	数量	释放规则
生态建设	60%	60,000,000	按有效工作量发放；年度释放上限为总量的 10% （≤ 10,000,000 枚 / 年），避免集中流通

社区空投与早期激励	30%	30,000,000	上线首期解锁 20%（6,000,000 枚）；剩余 80%（24,000,000 枚）自上线起分 2 年线性释放（每年 12,000,000 枚）
流动性与做市	10%	10,000,000	上线时注入，按流动性治理规则补充
团队与运营	0%	0	不预留；团队与运营支出由运营主体以自有资金承担

说明：本分配方案已最终确定。空投与生态池均设释放上限以管理上线流通抛压；团队不预留份额系平台既定决策，代之以运营主体自有资金承担长期运营。生态池冻结期间的处理口径（与 §8.7 闸门一致）：治理章程公示之前，60,000,000 枚生态池处于冻结状态，不计入流通量，不参与本表的释放曲线（生态池发放按「有效工作量 + 年度释放上限 ≤ 总量 10%」单独计量），并在供应量披露中单独列示为「冻结未流通」，避免与释放曲线重复计算。

关于团队 0% 预留的激励机制说明：团队不持有 Token 份额，意味着团队收益与二级市场 Token 价格**完全解耦**——不存在团队持币抛压，也不存在任何代持或暗池（团队名下没有任何可变现的 Token 头寸，这一点可由链上透明度与第三方审计共同核验）。团队与运营支出的真实回报来自运营主体的股权增值与服务收入，而非 Token 变现。该设计的代价是团队激励与币价关联减弱，因此通过收入挂钩的回购销毁（§8.3）把服务收入增长转化为全体持有人的共同利益，形成间接一致。

空投流通量的解析式与微分形式（t 为上线后月数）：

$$S_{airdrop}(t) = 6{,}000{,}000 + 24{,}000{,}000 \times \min(t/24,\ 1) \quad (\text{枚})$$

$$dS_{airdrop}/dt = 1{,}000{,}000 \text{ 枚/月} \quad (0 < t \le 24；\ t > 24 \text{ 后导数为零})$$

叠加生态池年度上限后的流通总量曲线如图 8-1 所示——供应扩张速度在数学上被封顶，回购销毁进一步使实际流通低于上限：

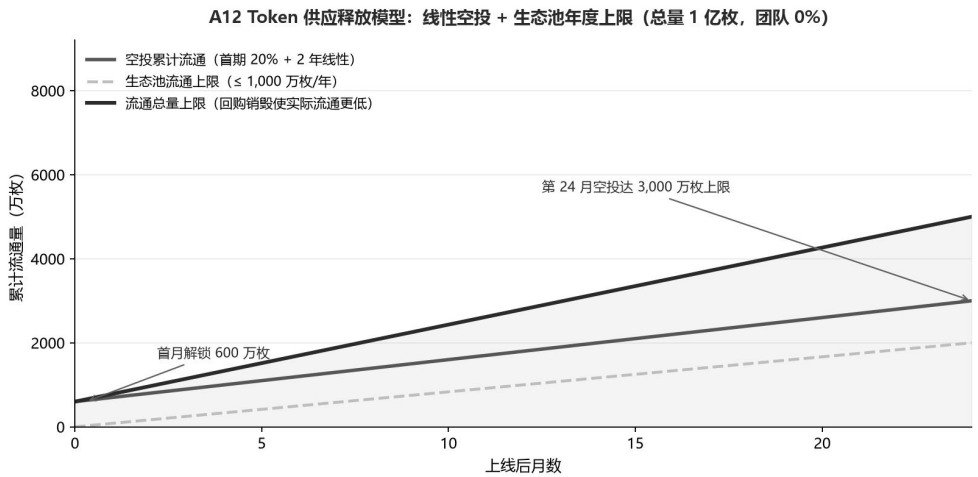

图 8-1 Token 供应释放模型：线性空投 + 生态池年度上限（回购销毁使实际流通更低）

8.3 回购销毁机制

- **销毁来源**：平台将上一季度**真实服务净收入**的 20%（初始比例，可由治理建议调整，上限 30%）用于从二级市场回购 A12 并销毁。

净收入的审计口径（消除口径操纵空间）：

$$NetRevenue\_q = GrossRevenue\_q - C\_compute\_q - C\_infra\_q$$

其中 GrossRevenue_q 为当季平台服务总收入（CaaS/MaaS/AaaS 全部到账收入）；C_compute_q 为算力采购与节点分成成本；C_infra_q 为验证集群、带宽等直接基础设施成本。人力、市场与行政费用不计入扣减项（防止把运营亏损转嫁为"净收入偏低"）。每季度三项数据由第三方会计师事务所出具鉴证报告后，随销毁交易一并公示。鉴证机构按以下标准甄选：具备国际四大网络成员所或同等规模资质、具有 Web3 或科技企业审计经验、与运营主体及其关联方无独立性问题；目标在公测启动前完成委托并披露机构名称【鉴证机构：确定后于官方渠道公示 · D-12】。

- **硬上限**：季度销毁量不得超过当季回购预算的 100%（即仅可销毁当季实际回购所得，不允许动用储备或借贷资金执行销毁）；且当生态共建池剩余量低于 5,000,000 枚时，暂停销毁，改为将回购部分注入生态池，保证长期激励池不被抽干（该阈值为长期保护性条款：按年释放上限 10,000,000 枚测算，预计运营五年以上方可能触发；公测与早期运营阶段以"季度销毁 ≤ 当季回购预算"为实际约束）。
- **示例核算**：假设平台某年服务净收入 2,000,000 美元，则当年回购预算 400,000 美元；按季度执行、市价买入、链上公示销毁交易。销毁量完全由真实收入决定——收入为零则销毁为零，不存在"为维持通缩叙事而抛售储备"的动机。

回购与流通供应的核算公式。 季度回购预算与销毁数量定义为：

$$B\_q = p\_burn \times NetRevenue\_q \quad (p\_burn\ 初始\ 20\%，上限\ 30\%)$$
$$\Delta S\_q = B\_q / VWAP\_q$$

其中 VWAP_q 为该季度成交加权均价，ΔS_q 为当季销毁数量——按成交量加权而非单一时点价格执行，抑制择时与操纵空间。销毁交易在链上公示，任何人可按"净收入凭证 → 回购交易 → 销毁交易"三段链路独立核验。

流通供应曲线。 空投池上线后的累计流通量遵循 §8.2 给出的释放公式（首月 600 万枚，此后每月线性增加 100 万枚，第 24 个月达到 3,000 万枚上限）；叠加生态池"年度释放 ≤ 总量 10%"的硬上限（§8.2），流通盘的扩张速度在数学上被封顶，与回购销毁的紧缩效应方向一致。上述公式使"释放节奏—抛压管理—收入回购"三者可被独立核算，避免任何依赖未公布参数的隐藏矛盾。

8.4 空投与积分

公测与早期贡献通过积分记录，积分按公开规则折算为 Token 空投。积分获取与有效工作量（完成任务量、验证通过率、在线时长）挂钩，与单纯账户数量脱钩；上线首期解锁空投池的 20%，剩余 80% 分 2 年线性释放（见 §8.2），避免上线集中抛压。公测期口径承诺：积分计量口径与上线后 CU·时长的结算口径同源——折算系数可调、计量口径不变，折算原则在公测开始前公示。公测期结算币种、兑付主体与结算周期【发布前补充并公示 · D-29】（§9.2 "结算准时率"即指按公示周期足额到账的比例）。高价值任务可选 USDT 信用保证金（可随时赎回，赎回导致声誉重置），在不依赖 Token 质押的公测窗口提供作恶成本下限。

折算系数调整约束：积分折算系数的调整受三重约束——单次调整幅度不超过 ±20%；两次调整间隔不少于 90 天；任何下调须提前 30 天在公示页公告并说明理由。全部调整记录留档可查，调整程序与上限纳入治理章程（§8.7）的公示范围，确保「计量口径不变」不等于「兑换水平可被单向下调」。

首次折算系数的确定：上述三重约束（±20% / 90 天 / 提前 30 天）规范的是「已存在系数的调整」，首次系数另按以下规则确定——确定原则：以公测期积分计量口径与上线后 CU·时长结算口径的换算基准为锚，即首次系数取「单位有效工作量在两种口径下的等价 CU·小时数」的倒数，使积分持有人的相对位置不因口径切换而改变；确定主体：由运营主体（平台方）在听取治理建议后确定，重大分歧提交治理章程规定的程序处理；公示时点：不晚于公测（含激励活动）开始之前，与 §8.5 激励活动规则同步公示，并纳入治理章程（§8.7）的公示范围。首次系数一经确定即受三重约束管辖，其确定依据与测算记录随披露闭环追踪表一并留档。

8.5 公测激励的说明

公测与空投以 **§2.4 所述许可状态与证券法律意见为前置条件**，满足后另行公告具体日期：**无激励的开放式网络公测**（不涉及任何 Token 发放，仅验证网络与产品能力）可在许可取得前先行开展；**含 Token 空投的激励活动**在取得必要许可并完成法律意见评审后启动。激励预算总额（于激励活动启动时执行）：**30,000,000 A12（即空投池全额，发放严格遵循 §8.2 释放节奏：首期解锁 6,000,000，其余 24,000,000 分 2 年线性释放）+ 5 BTC + 900,000 USDT**，全部由运营主体（平台方）自有资金承担；活动规则、参与资格、奖励总额与领取条件在**平台官方网站**公示。Token 奖励受 §8.2 释放规则约束，法币与 BTC 奖励计入运营预算且设总额上限，不与"激励与结算分离"原则冲突。

地域合规边界：公测激励与空投活动的参与以通过 KYC 与制裁名单筛查为前提（见 §2.4）。以下地区的用户不参与 Token 空投与激励分配：联合国安理会制裁名单及 OFAC 制裁地区；中国大陆地区（依据境内监管要求，不向境内居民发放任何形式的 Token 激励）；平台风控认定的其他高风险司法辖区。美国用户的参与资格待证券法律意见出具后另行公告。欧盟（MiCA 法规）、英国（FCA）、新加坡（MAS）、韩国、

日本等对加密资产有明确监管要求的法域，其准入状态与合规路径以官方公告为准。**完整地域准入条款在官网活动规则中公示**，并在领取环节二次校验。

8.6 协议层

平台不在路线图核心期自研公链。结算与存证的协议层采用在成熟公链/L2 上部署智能合约的方式实现：合约承担结算凭证存证、销毁公示与（如监管允许）稳定币结算通道。自研共识链不在任何阶段的承诺范围内。DEX 相关规划取消，原因有二：其一，自建链上兑换与做市功能会分散核心研发资源，与 §9.1 "资源聚焦于调度平台、控制台、MaaS、AaaS 四条产品线"的原则相悖；其二，DEX 涉及多地司法辖区对去中心化交易的许可不确定性，在合规路径清晰前不布局。Token 流动性改由专业做市机构以传统方式提供，遵循公开透明的市场惯例：上线初始流动性由 §8.2 流动性与做市池（10%）注入；具体做市机构名称与合作条款在上线前 1 周于官网公布【做市机构：确定后于上线前一周公布 · D-30】。

销毁公示合约的参考接口如下（阶段 4 交付物示意，非最终实现；借鉴以太坊 EIP 标准接口与 OpenZeppelin 审计惯例）：

```
contract A12BurnLedger {
     using SafeERC20 for IERC20;
     IERC20  public immutable A12;
     address public constant BURN_SINK = 0x000000000000000000000000000000000000dEaD;
     uint256 public totalBurned;

     event Burned(uint256 indexed quarter, uint256 amount, bytes32 revenueDigest);

     function burn(uint256 amount, bytes32 revenueDigest, bytes calldata sig) external {
         require(_verifyAuditorSig(revenueDigest, amount, sig), "invalid attestation");
         totalBurned += amount;
         A12.safeTransfer(BURN_SINK, amount); // SafeERC20 安全转账；链上公示：任何人可独立核验销毁
         emit Burned(currentQuarter(), amount, revenueDigest);
     }
 }
```

收益托管与罚没的参考合约接口（与销毁合约同属协议层阶段 4 交付物示意，非最终实现；其"收益延迟释放 + 条件罚没"结构与 Filecoin 的 miner vesting / slashing 合约同源）：

```
contract A12EscrowAndSlash {
     struct Operator { uint256 locked; uint256 released; uint256 reputation; }
     mapping(address => Operator) public operators;

     function lockEarnings(address op, uint256 amount) external onlySettlement {
         operators[op].locked += amount;          // Earned → Pending 状态上链
     }

     function slash(address op, uint256 amount, bytes32 evidenceDigest) external onlyArbiter {
         Operator storage o = operators[op];
         uint256 penalty = amount < o.locked ? amount : o.locked; // 内联 min：罚没上限 = 托管额
         o.locked -= penalty;
         o.reputation = (o.reputation * 8) / 10;  // 违约后声誉直接衰减
         emit Slashed(op, penalty, evidenceDigest);
     }
 }
```

安全审计计划：上述合约在进入阶段 4 部署前，须由独立第三方安全机构完成完整审计——甄选标准与 §8.3 鉴证机构一致（国际网络成员所或具备 Web3 合约审计资质、与运营主体无独立性问题）；审计报告**全文公开**；审计通过前合约不部署主网、不接入真实资金托管与罚没流程。

8.7 治理

平台运营决策（调度策略、定价、风控规则）由运营主体负责；生态治理通过 **A12 全球治理 DAO** 展开。

DAO 面向社区贡献者、行业创作者、技术贡献者与优秀 Worker，全球治理席位共 **300 席**，首期开放 **40 席**（以正式招募公告为准）。

治理机制（设计目标口径，参数以治理章程公示为准）：成员经"自荐申请 + 社区评审公示"遴选，任期 12 个月、可连任一次；提案须获得不低于成员数 10% 的联署方可进入表决；一般事项采用简单多数通过，法定参与率不低于 25%；分配参数调整、罚没规则变更等重大事项须不低于三分之二多数；决议由运营主体限期执行并在官网公示执行结果；连续三次无故缺席表决或违反利益回避规则的成员予以除名。技术载体采用链下快照投票 + 链上公示存证（与 §8.6 协议层一致，不依赖自建公链）。

DAO 成员权益：

1. 参与生态提案、投票与重大事项治理；
2. 参与平台规则及相关治理决策；
3. 根据生态贡献获得对应激励；
4. 享受专项积分及相关生态权益；
5. 优先参与 A12 全球生态建设。

DAO 与运营主体的权责边界（消除"中心化运营 vs DAO"的张力）：DAO 是**生态治理层**，不是平台运营层。生态事项由 DAO 实质决策，平台运营事项由运营主体负责并接受 DAO 建议，边界如下：

事项类别	DAO 权限	运营主体角色
生态激励分配、生态活动、社区提案	**提案权 + 投票权**，决议对生态池使用具有约束力	执行决议并公示；未采纳须在 14 天内书面说明理由
调度策略、定价、风控与合规规则	建议权（无投票约束力）	最终决策权；对 DAO 建议的采纳情况定期公示
Token 经济参数调整（如回购比例在 20%–30% 区间内移动）	提案权 + 投票权，决议提交运营主体执行	执行须受 §8.2 与 §8.3 硬上限约束，不得突破

该划分与 §2.1 "中心化运营 + 分布式执行"定位一致：平台运营责任主体唯一且明确（避免"说了不算"的装饰性治理），生态资源的分配权则实质交由 DAO（避免"伪去中心化"观感）。

治理反捕获机制（设计目标口径，细则以治理章程公示为准）：① **提案双重门槛**——提案须同时满足席位联署 ≥ 10% 与提案人最低持币 / 声誉门槛，防止单低成本席位串联；② **单次提案额度上限**——任一生态池使用提案可动用金额 ≤ 生态池未释放余额的 5%，超大事项拆分表决；③ **利益回避**——提案人及其关联方不得对自身受益提案投票，违反者提案无效并触发除名程序（"利益回避"定义写入治理章程）；④ **紧急暂停**——运营主体发现违规或恶意提案可暂停执行最长 14 天并公示理由，超期未恢复须提交 DAO 全体表决；⑤ **同标准鉴证**——生态池发放与 §8.3 回购销毁适用同一套第三方会计鉴证与公示标准；⑥ **首期名单先行**——首期 40 席的遴选标准与当选名单在 DAO 启动前公示。

> 治理生效文件清单：① 治理章程（含遴选细则、表决规则、利益回避与紧急暂停程序）；② 首期 40 席名单与遴选标准（与反捕获机制⑥「首期名单先行」为同一事项，统一时点：DAO 启动前公示，不再单列重复条目）；③ 生态池单次提案额度上限的解释规则。「公示即生效」的认定方式：以公示页发布时间戳（UTC）+ 内容 SHA-256 哈希共同认定生效时点，两者随内容一并归档；验证主体为任何可独立取证的第三方——时间戳由 §2.2 所述第三方时间戳机制（RFC 3161 令牌 / 公开 Git 提交哈希 / §8.6 存证合约链上哈希）提供，与该机制共用同一套存证与归档流程；版本归档规则：全部历史版本永久保留于 /archive 归档目录，争议时以归档中时间戳最早且哈希匹配的完整版本为准。治理章程公示前，60,000,000 枚生态池不动用任何额度；治理章程公示的目标时点为 2026 年 10 月 31 日前（目标时点，非承诺日期；与阶段 0 目标启动窗口 2026 年 10 月上旬相衔接），责任岗位为法务与合规负责人；若超过该时点仍未公示，生态池继续全额冻结，运营主体须在超期后 5 个工作日内公示延期原因与修订后的时点；连续两次延期的，须将「章程长期未公示情形下的替代授权机制」（如按已公示的年度释放上限的固定比例分期解冻、或授权 DAO 就生态池使用先行表决）提交治理建议程序审议——替代授权机制获通过前，生态池不动用任何额度。反捕获机制①–⑥的全部参数以章程公示版为准。

第九章技术路线图

9.1 资源聚焦原则

核心资源集中于四条产品线：**调度平台（Layer 1/2）、租户控制台与沙箱运行时、模型服务（MaaS，含首个垂直模型流水线）、AaaS 工具链（编排画布 + 开发者工具 + 预置 Agent）**。其余方向（跨厂商深度适配、协议层深化、更多垂直行业模型）为远期实验项，进入路线图前须通过上一阶段的退出判据。

9.2 分阶段路线图（含退出判据）

阶段	时间	交付目标	退出判据（Go/No-Go）
阶段 0：封	2026 年	邀请制 50–100 个 NVIDIA 栈节	验证通过率 ≥ 99%；调度成功率 ≥

闭内测	10 月上旬	点；核心调度与五层核验上线；租户控制台（Web 终端 + 文件通道）；真实结算（法币 / 稳定币）	95%；实例冷启动 ≤ 10 秒；零重大安全事故
阶段 1：全球公测 · Worker 招募 · 空投一期	2026 年 10 月下旬起	弹性尽力模式开放注册（目标 1,000 节点）；收益延期释放池上线；OpenAI 兼容网关 + 连续批处理；空投一期与公测激励以 §2.4 许可状态为前置、按官网公示执行	节点月留存 ≥ 60%；任务纠纷率 < 1%；结算准时率 100%（按公示结算周期足额到账的比例）；TTFT 中位数 ≤ 400ms（≤ 8K 上下文档位；32K 长上下文档位单独设定目标）
阶段 2：SLA 商用	2027 H1	专属保障模式 + 冗余部署；MaaS 上线；首个垂直模型流水线交付 Beta；首批企业租户	达成 SLA 目标可用性；至少 3 家付费企业租户；正毛利（毛利构成按算力服务与模型收入分项核算并公示）；GPU 有效负载率 ≥ 75%（连续批处理档位）
阶段 3：AaaS 与扩展	2027 H2	编排画布 + Web IDE 公测；预置 Agent 矩阵上线；ROCm 实验支持；回购销毁启动	AaaS 月收入 > 0 且增长；ROCm 任务通过率 ≥ 90%
阶段 4：协议层	2028+	成熟公链上的结算存证合约；跨厂商适配研究	监管评估通过；审计报告公开

> 说明：阶段 0 封闭内测（不涉及 Token 发放）不以日历日期为承诺——其启动以 §2.4 许可状态确认与 §2.2 披露闭环关键路径的完成度为前置；在前置条件满足且团队产能就绪的前提下，目标启动窗口为 2026 年 10 月上旬（目标窗口，非承诺日期），实际启动以验收公示为准。空投与激励活动以 §2.4 许可状态与证券法律意见为前置条件（见 §8.5），满足后另行公告日期。时间线以工程验收为驱动，宁可推迟、不跳过阶段：任一 Go/No-Go 判据未达成即推迟扩张，不以透支人力为代价跳过阶段（判据的第一性依据见 §5.3 供给端经济性测算框架）。

9.2.1 验收报告规范（固定附件）

每一阶段结束均公开发布验收报告。报告由「九项固定字段 + 阶段附加字段」两部分构成，固定字段不随阶段调整：① 节点规模（在册 / 活跃 / 达标节点数，按资源梯队分列；含跨身份去重统计——申报设备数 / 去重后设备数 / 差额，与 §4.5 的序列号全局唯一互斥对应）；② 任务量（当期总任务数与按任务类型的分布）；③ 验证通过率（按 §4.2 的 L2 与 L4 复核口径统计）；④ 调度成功率（首次调度成功占比，区分弹性尽力与专属保障模式）；⑤ 实例冷启动 P50/P95；⑥ TTFT P50/P95（≤ 8K 上下文档位；32K 长上下文档

位单列）；⑦ GPU 有效负载率（连续批处理档位，取 P50）；⑧ 事故清单（事故等级、影响时长、根因与整改项）；⑨ 未达标项与原因（逐条比对本阶段 Go/No-Go 判据，列明偏离值与归因）。阶段附加字段由各阶段在启动公告中列明，只增不减——固定字段不得删除或替换。首份验收报告于阶段 0 结束后 7 日内发布，此后每一阶段结束后 7 日内发布；进入下一阶段的唯一条件是上一阶段判据全部达成并公示验收报告（不以日历日期为承诺）。该排他表述不因融资进度、市场窗口或人力安排而例外。

9.3 前沿技术展望（研究地平线）

定位声明：本节描绘 A12 的中长期技术愿景，用于说明平台的演进方向与技术品味。**以下所有条目均为研究项或远期展望，不属于任何阶段的交付承诺**；任何条目进入正式路线图前，须通过其前置阶段的量化退出判据（见 §9.2），并另行公告。

A12 的长期蓝图可以概括为一句话：**让全球每一块闲置 GPU 通过统一的可信接口参与 AI 价值链，并让智能体成为算力的原生消费者。**沿着"CaaS → MaaS → AaaS"的纵向价值链，技术演进分为三个地平线。

聚焦说明：本章有意只保留与平台核心能力（调度、推理、Agent 执行）直接相关的方向；联邦学习、差分隐私、同态加密、模型水印、碳感知调度、WASM 沙箱、后量子密码等更远期或跨领域方向不在本文展开，将随垂直行业扩展在后续技术博客中另行探讨——避免"什么都想做、什么都没做深"的观感。

地平线一：算力层（当前 → 2027）——把分布式算力做到"可信、可用、可调度"。

前沿方向	概念要点	与平台的关系
eBPF 内核级遥测	在 Linux 内核中安全运行沙箱化探针，以近零开销采集系统调用、网络与 GPU 驱动事件	使 L3 运行时遥测从用户态采样升级为内核级观测，压缩伪装空间
意图驱动调度（Intent-based）	租户只声明目标（SLO、预算、时延），调度器自动求解资源配置与部署拓扑	从"任务匹配资源"演进为"目标驱动编排"，是 §5.3 调度模型的自然延伸
集群数字孪生	用离散事件仿真为整个资源池建数字镜像，调度策略先在孪生体上回放验证再上线	降低新调度策略的生产风险，支撑 SLA 容量规划

地平线二：模型与执行层（2027 → 2028）——让前沿模型效率与分布式执行深度耦合。

前沿方向	概念要点	与平台的关系
FSDP / ZeRO 分片训练	将优化器状态、梯度与参数分片到多设备，把全参微调的显存门槛降到单卡可承受范围	扩大分布式资源池可承载的训练任务谱系

FP8 混合精度训练	前向 / 反向以 FP8 执行、关键累加以高精度保留	与低精度推理共同构成平台 " 精度 — 成本 " 产品曲线
投机解码 / FlashAttention-3 / Disaggregated Serving	见 §6.4 性能工程前瞻	MaaS 层推理效率的持续演进
TEE 机密计算	CC 模式 GPU 硬件加密显存与计算过程	见 §4.7；打通 " 高敏感企业负载上分布式资源 " 的信任瓶颈

地平线三：智能体与经济层（2028+）——智能体经济（Agentic Economy）的结算基础设施。

当智能体成为算力的主要消费者，调度平台将从"给人用的云"演进为"给 Agent 用的算力市场"：

1. **MCP / A2A 工具与协作协议**（见 §7.1）：Agent 以标准协议发现工具、委托任务、互认状态；
2. **推理时计算缩放（Test-time Compute）**：通过 CoT / ToT / 多采样投票等策略，让模型在困难问题上"多想一会"，以质量换成本；调度平台据此提供"快速 / 标准 / 深度"三档推理服务；
3. **Agent 自主结算**：智能体以预设预算与策略自主采购算力（竞价、锁价、套餐），结算记录可审计——这与 §8.6 协议层的智能合约结算能力衔接；
4. **算力要素市场化**：声誉、梯队、SLA 与实时供需共同形成公开的算力价格信号，让资源定价从"挂牌制"演进为"市场化撮合"。

上述地平线一、二为工程演进路径，地平线三为方向性愿景；三者的关系是**层层依赖、逐级验证**——没有可信的算力层就没有高效的执行层，没有执行层的规模化就没有智能体经济层的结算需求。本节全部内容受第 10 章前瞻性声明约束。

第十章风险因素与免责声明

1. **执行风险：平台交付依赖核心工程团队到位与稳定；团队披露信息见 §2.2，未完整披露前相关风险由读者自行评估。缓解：分阶段路线图设量化 Go/No-Go 退出判据，每阶段按 §9.2.1《验收报告规范》公开验收报告（九项固定字段 + 阶段附加字段，阶段结束后 7 日内发布）。**
2. **技术风险**：容器隔离不能消除全部逃逸风险；验证体系不能绝对阻止超售，仅能通过经济与运营手段抑制；跨厂商适配存在不确定工期；垂直模型流水线的语料建设与训练效果存在不确定性。**缓解**：容器隔离 + 用途审查 + 滥用赔付基金三层防护（见 §4.7）；SLA 主备冗余与声誉准入抑制超售（见 §5.3、§4.5）。
3. **市场风险**：算力价格波动、云厂商竞争、AI 负载结构变化都可能影响平台收入假设。**缓解**：双路径调度按任务特征择优路由，收入结构覆盖 CaaS/MaaS/AaaS 三层，降低单一市场依赖（见 §5.2、§3.1）。

4. **监管风险**：数字资产相关活动在不同司法辖区面临不同要求，Token 的可用功能可能随监管调整；混合收单与协议层能力以监管评估为前提。**缓解**：分法域准入矩阵与领取环节二次校验（见 §8.5）；涉及许可的业务在取得许可前不开展（见 §2.4）。
5. **智能合约风险**：智能合约可能存在未发现的漏洞（如重入、整数溢出、权限配置错误）；在第三方安全审计完成前，合约交互存在资金损失风险。**缓解**：合约仅作结算凭证存证与销毁公示的协议层可选能力（见 §8.6）；部署前须通过独立第三方审计并全文公开审计报告，审计前不上主网、不接入真实资金。
6. **前瞻性声明**：本白皮书所载全部指标为**设计目标**，全部路线图为**规划**，不构成任何收益承诺。实际结果可能与规划存在重大差异。
7. **免责**：本白皮书不构成投资建议、证券发行文件或法律意见。涉及第三方商标的权利归各自权利人所有。

附录 A 术语表

术语	定义
运营者（Operator）	平台账户主体（自然人或法人），声誉、结算与惩罚的聚合单位
资源租约	平台账本中 " 设备序列号 + 运营者 " 维度上的互斥资源锁定记录
随机信标（VRF）	可验证随机函数，用于生成不可预知、事后可验证的挑战种子
收益延期释放池	Earned → Pending → Released 的收益三态流转与违约罚没机制
通信 — 计算比 α_comm	任务跨节点通信时间与纯计算时间之比，双路径调度依据
CU	节点综合能力评分，由实测分项加权而成
连续批处理（Continuous Batching）	以迭代为粒度动态插入 / 退出请求的推理调度技术
HAL	硬件适配层，统一任务提交接口、披露后端适配矩阵
CPT / SFT	持续预训练 / 监督微调，垂直模型流水线的两个训练阶段
DPO	直接偏好优化，以工具执行结果构建偏好对的模型对齐方法
TTFT	首字延迟（Time To First Token），在线推理体验核心指标
Model Card	模型卡，记录模型预期用途、局限与评测结果的标准化文档

EWMA	指数加权移动平均，遥测基准线的统计估计方法（见 §4.2 ）
Hockney 模型	经典通信时间模型： T ≈ 固定延迟 + 数据量 / 带宽（见 §5.2 ）
GQA	分组查询注意力，通过共享键值头压缩 KV 缓存显存占用的注意力结构（见 §6.4 ）
关键路径	DAG 中从起点到终点的最长加权路径，工作流完成时间的下界（见 §7.1 ）
Power of Two Choices	二次选择负载均衡策略：随机取两候选、派往较空者（见 §6.3 ）
投机解码（ Speculative Decoding ）	小模型起草、大模型验证的加速生成方法（见 §6.4 、 §9.3 ）
Prefill/Decode 分离	将首字生成与后续解码拆分至不同资源池的推理服务架构（见 §6.4 、 §9.3 ）
TEE / 机密计算	可信执行环境，硬件级加密显存与计算过程（见 §4.7 、 §9.3 ）
eBPF	Linux 内核可编程观测技术，用于内核级低开销遥测（见 §9.3 ）
MCP / A2A	面向 Agent 的开放工具接入协议 / 智能体间协作协议（见 §7.1 、 §9.3 ）
PQC （后量子密码）	抵抗量子计算攻击的下一代密码算法体系。本文不展开，仅作术语索引（见 §9.3 ）
数字孪生	为物理资源池建立可回放仿真的数字镜像（见 §9.3 ）
承诺 – 揭示（ Commit–Reveal ）	先公示哈希承诺、后揭示原值的防操纵协议范式（见 §4.2 ）
Erlang C / M/M/c	多服务窗排队模型，用于并发容量规划与时延反解（见 §6.3 ）
Gas 计量	以太坊式资源计量口径， A12 用于任务级计费与审计（见 §3.4 ）
Collateral / Slashing	Filecoin 的抵押罚没机制， A12 罚没函数的参照来源（见 §3.4 、 §4.5 ）
Chinchilla 标度律	模型损失关于参数量 N 与训练 token 数 D 的经验缩放关系（见 §6.2 ）
Roofline 模型	算术强度约束下的峰值性能模型（见 §6.4 ）

ReAct / CoT / ToT	Agent 的推理 + 行动 / 思维链 / 思维树决策模式（见 §7.3 、 §9.3 ）
模型路由（ Model Routing ）	根据输入复杂度自动选择不同规模模型的推理策略（见 §7.3 、 §9.3 ）
联邦学习 / 差分隐私	隐私保护式分布式训练与加噪梯度聚合。本文不展开，仅作术语索引（见 §9.3 ）
CaaS	Compute as a Service ， 算力即服务 —— 本平台资源层（ Layer 1/2 ）的对外商业形态
MaaS	Model as a Service ， 模型即服务 —— 本平台模型层（ Layer 3 ）的对外商业形态
AaaS	Agent as a Service ， 智能体即服务 —— 本平台应用层（ Layer 4 ）的对外商业形态
SLA	Service Level Agreement ， 服务等级协议 —— 对可用性、性能与赔付的量化承诺
DAO	Decentralized Autonomous Organization ， 去中心化自治组织（见 §8.7 治理机制）
IDC	Internet Data Center ， 互联网数据中心
RAG	Retrieval-Augmented Generation ， 检索增强生成
LoRA / QLoRA	Low-Rank Adaptation 及其量化版本，参数高效微调方法
MSB	Money Services Business ， 美国 FinCEN 货币服务业务牌照
SEC	U.S. Securities and Exchange Commission ， 美国证券交易委员会
OFAC	Office of Foreign Assets Control ， 美国财政部海外资产控制办公室
DEX	Decentralized Exchange ， 去中心化交易所（本平台不做 DEX ， 见 §8.6 ）
NVML	NVIDIA Management Library ， GPU 状态遥测库
WSS	WebSocket Secure ， 加密 WebSocket 通道
SLO	Service Level Objective ， 服务等级目标（ SLA 的量化前置指标）

EIP-1559	以太坊改进提案 1559，基础费 + 小费的费用定价机制（见 §3.4）
Go/No-Go	路线图各阶段的量化退出判据决策机制（见 §9.2）
PoSt	Proof of Spacetime， Filecoin 时空证明（见 §3.4）

附录 B 图表索引

图：图 3-1 四层架构层次图（§3.2）；图 4-1 五层核验体系（§4.2）；图 5-1 双路径分流（§5.2）；图 5-2 任务生命周期状态机（§5.5）；图 6-1 KV 缓存容量模型（§6.4）；图 8-1 供应释放模型（§8.2）。

主要表：强制披露清单（§2.2）；披露闭环关键路径（§2.2）；八项主体信息披露状态表（§2.2）；已承诺支出与资源保障表（§2.5）；设计参照系（§3.4）；资源梯队调度约束（§5.4）；验收报告固定字段（§9.2.1）；DAO 与运营主体权责边界（§8.7）；代币分配模型（§8.2）；回购销毁参数（§8.3）；风险因素（第十章）；术语表（附录 A）；披露闭环追踪表（附录 C）；图表参数表（附录 D）。

附录 C 披露闭环追踪表（D-01 至 D-30）

本表是「披露闭环追踪表」的全文，共 30 项，与 §2.2 披露清单表的「编号」列逐行对应，并覆盖本版全文全部待补事项；运营主体注册信息（§2.2 第一行）已提供且可公开核验，不占用待补编号。编号口径：上一版声明 21 个编号，而全文实际存在 30 处待补标记（登记遗漏 9 处）；本版按 30 处标记逐条登记为 D-01 至 D-30，属补登记而非新增承诺；正文标记随 D-02 / D-06 / D-07 三项信息补充由 26 处进一步收敛至 24 处；同一事项在不同章节的重复登记合并为同一编号，拆分、合并与信息补充均不改变承诺范围。编号颗粒度：可独立闭环、外部可分别核验的数据点拆为独立编号——例如「资金状况披露」因规模区间（D-17）、投入科目（D-18）与 12 个月资金缺口（D-19）三者的数据来源与商业敏感度不同而拆分，三者可分别闭环；同一披露行为的多处标记不重复编号。字段口径：要求形式指该事项对外呈现的方式；当前状态分「待补充 / 已补充待佐证 / 已公示」三档；承诺日期为目标承诺日期（非合同承诺），由运营主体负责人批准后生效。节奏口径：承诺日期为考核节点，逾期即触发逾期处理规则；按周更新为过程披露，两者以承诺日期为准。公示链接：本列给出公示页锚点，根地址过渡期为 aeliontwelve.vercel.app/disclosure（见 §2.2 双落点），完整地址 = 根地址 + 锚点；官网主域名启用后 72 小时内，本列与公示页同步更新为正式域名，旧链接保留跳转与归档。公示页首版（含本表 30 项初始状态）不晚于 2026-09-21 上线，此后每周一（UTC）更新并保留全部历史版本。责任岗位：主体信息八项见 §2.2 八项状态表，其余事项由运营主体对应职能负责人承担。逾期处理：任一事项逾期未闭环的，按 §2.2 八项表下方规则执行（5 个工作日内说明原因与修订日期；连续两次逾期须公示整改说明），本表同步更新并保留全部历史版本。

编号	事项	所属章节	要求形式	当前状态	承诺日期	公示链接锚点（根地址见 §2.2 与本表导语）
D-01	核心团队 6 人可验证履历链接（LinkedIn / 学术档案）	§2.2 ①	逐人列示「姓名 + 档案链接 / 档案编号」	待补充	2026-10-15	disclosure#D-01
D-02	官方邮箱与办公地址	§2.2 ②	公示页列示官方邮箱、办公地址	已公示（2026-09-15）	2026-09-21	disclosure#D-02
D-03	法务联络渠道（法务邮箱与文书收件地址）	**§2.2 ③**	**公示页列示法务联络邮箱与收件地址**	**已公示（与官方邮箱共用，2026-09-15）**	**2026-09-21**	**disclosure#D-03**
D-04	投资方全称、背景简介、投资轮次与任一公开佐证（含 Bluechip 全称与简介）	§2.2 ④	全称 + 轮次 + 工商变更 / 公开新闻 / 领投方确认函（任一即可）	待补充	2026-10-31	disclosure#D-04
D-05	专利号 / 申请号（中国 / 尼日利亚 / 美国）	§2.2 ⑤	逐项列示受理号或专利号与受理局	待补充	2026-11-30	disclosure#D-05
D-06	代码仓库地址或第三方审计报告（全文公开）	**§2.2 ⑥**	**GitHub 仓库地址，或审计报告全文**	**已补充待佐证**	**2026-11-30**	**disclosure#D-06**
D-07	官方社交账号清单（防仿冒）	§2.2 ⑦	逐平台列示官方账号唯一链接	已公示（2026-09-15）	2026-10-15	disclosure#D-07
D-08	官网正式主域名	§2.2 ⑧	主域名启用公告 + 全站迁移完成	待补充	2026-10-31	disclosure#D-08
D-09	命名与符号（A12 / Aelion Twelve）唯一性检索范围与结果	§2.3	商标与行情平台检索报告	待补充	2026-10-31	disclosure#D-09
D-10	激励托管账户安排（托管机构与账户）	§2.5 / §8.5	机构名称与账户安排说明	待补充	2026-10-31	disclosure#D-10
D-11	第三方会计鉴证费用区间	§2.5	区间数值 + 计费口径	待补充	2026-09-21（随公示页首版）	disclosure#D-11
D-12	第三方会计鉴证机构名称	§2.5 / §8.3	机构全称与资质说明	待补充	2026-10-31	disclosure#D-12
D-13	智能合约独立审计费用区间	§2.5	区间数值 + 计费口径	待补充	2026-11-30	disclosure#D-13

D-14	智能合约审计机构名称与审计报告全文	§2.5 / §8.6	机构全称 + 报告全文公开	待补充	2026-11-30	disclosure#D-14
D-15	算力采购与节点分成年度上限	§2.5	年度上限数值 + 口径说明	待补充	2026-10-31	disclosure#D-15
D-16	团队薪酬年度总额区间	§2.5	区间数值 + 编制口径	待补充	2026-10-31	disclosure#D-16
D-17	运营主体自有资金规模区间	§2.5	区间数值；可代以「已向委托方单独提供并留存可核验记录」确认函	待补充	2026-10-31	disclosure#D-17
D-18	已发生投入的主要科目	§2.5	科目清单与金额区间	待补充	2026-10-31	disclosure#D-18
D-19	未来 12 个月资金缺口与补足方式（最优先闭环项）	§2.5	缺口区间 + 补足方式（经营收入 / 股权融资 / 其他）与合规前提	待补充	2026-09-30	disclosure#D-19
D-20	人力保障：阶段 0 招聘计划、顾问配置与外包边界	§2.5	招聘计划 + 法务 / 财务 / 合规顾问配置 + 外包边界	待补充	2026-10-31	disclosure#D-20
D-21	图 5-1 参数表（α_comm 区间、带宽档位、分流阈值与边界实测点）	§5.2	已随本版给出：附录 D 表 D-1	已补充待佐证	已给出（2026-09-14）	disclosure#D-21
D-22	图 6-1 参数表（模型档位、GQA 配置、分页块与命中率口径、显存与并发）	§6.4	已随本版给出：附录 D 表 D-2	已补充待佐证	已给出（2026-09-14）	disclosure#D-22
D-23	SLA 目标值（A_service / A_platform / A_node）与故障赔付规则	§5.3	目标值 + 统计窗口 + 赔付规则一并公示	待补充	2026-10-31	disclosure#D-23
D-24	供给端经济性测算表（含字段单位与取值口径）	§5.3	测算表全文 + 样本节点数	待补充	2026-10-31	disclosure#D-24
D-25	语料许可证治理的专项法务意见	§6.2	法务意见公开版（含 copyleft 收录边界口径）	待补充	2026-11-30	disclosure#D-25
D-26	后续批次语料的数据预算与时间表	§6.2	批次规模、预算与时间表	待补充	2026-12-31	disclosure#D-26
D-27	数据源清单与再分发授权状态	§7.3	数据源清单 + 授权状态	待补充	2026-11-30	disclosure#D-27
D-28	收单主体与支付服务商（法币通道）	§7.4	持牌收单主体全称与服务条款	待补充	2026-11-30	disclosure#D-28
D-29	公测期结算币种、兑付主体与结算周期	§8.4	结算币种 + 兑付主体 + 结算周期	待补充	2026-10-31	disclosure#D-29
D-30	做市机构名称与合作条款	§8.6	机构全称与条款摘要（上线前 1 周公布）	待补充	待定（依赖阶段 4 Token 上线决策）；	disclosure#D-30

					中间节点：做市机构接洽启动时点随阶段 3 启动公告一并公示，短名单与接洽进展按周更新	

附录 D 图表参数表（图 5-1 / 图 6-1）

本附录给出图 5-1 与图 6-1 的复算参数表，对应追踪编号 D-21、D-22（状态：已补充待佐证）。佐证路径：表中数值为按 §5.2 Hockney 通信模型、Roofline 吞吐估计与 §6.4 M_KV 公式得到的设计估算值；其佐证方式为平台验证集群在阶段 0 内测（邀请制 50–100 个 NVIDIA 栈节点，见 §9.2）中，对五类任务与两档模型档位实测复算——实测 T_comm / T_compute / 显存占用与本表逐项比对，实测记录随公示页留档；任何第三方亦可按本附录参数独立复算并比对。版本化标识：两张参数表按版本化维护，每次更新标注版本号（如 D-1.1 / D-2.1）与日期戳，全部历史版本随 /archive 归档保留，便于引用与比对；参数区间随阶段 0 实测样本扩充按季度复核更新。表 D-1 另给出 T_comm（按带宽档）与 T_compute（含硬件档位假设）两列，使 α_comm = T_comm ÷ T_compute 可逐档复算，不再只给比值而不给分母。

表 D-1　图 5-1 参数表：α_comm 估算区间（逐带宽档）、T_comm / T_compute 取值与分流归属（D-21）

任务类型	单步同步量（锚定）	T_comm（秒/步，按带宽档）	T_compute（秒/步，硬件档位）	α_comm 区间（对应带宽档）	分流归属	边界任务实测点
全参数微调（7B 级，数据并行）	≈ 28 GB/步	2,607 @100M；261 @1G；26 @10G；9.5 @25G；0.6 @IB400G	3–9（RTX 4090 / H100 级单卡，η_eff 取峰值 30–60%，§5.2 口径）	0.1–870（逐档：290–870 @100M；29–87 @1G；2.9–8.7 @10G；1.1–3.2 @25G；0.1–0.3 @IB400G）	集中式集群——除 IB 级互联外，全部带宽档位下 α_comm > 1	7B 全参微调（28 GB/步同步量）
分布式预训练 / 大规模 CPT（70B 级）	≈ 280 GB/步（随参数量线性增长）	255 @10G；25.5 @100G 专线；6.2 @IB400G	20–60（H100 × 8 集群，η_eff 30–60%）	4–13 @10G；0.1–0.3 @IB400G	集中式集群——分布式资源池最高档（IDC 10 Gbps）下 α_comm 已达 4–13，远超阈值 1；且张量 / 流水并行要求节点内 NVLink/IB（Tier-S，§5.4）	70B 级 CPT 的分片梯度同步
LoRA / QLoRA 微调	≈ 0.1–1 GB/步（仅同步	0.9–9 @1G；0.09–0.9 @10G	3–9（RTX 4090 级单卡）	0.1–3 @1G；0.01–0.3 @10G	边界区：以实测复核决定归属	LoRA 微调（< 1 GB/步）

	适配器参数）					
在线推理（≤ 8K 上下文）	KB–MB 级（请求与响应）	10^{-3}–10^{-2}	0.05–2（单请求 prefill + decode）	10^{-3}–10^{-2}	分布式资源池	—（远离阈值）
Embedding / Rerank / RAG 与批处理	KB–MB 级	10^{-4}–10^{-2}	10^{-2}–10^{1}	10^{-4}–10^{-2}	分布式资源池	—（远离阈值）
分流阈值	—	—	—	$\alpha\_comm = 1$；阈值两侧以实测复核决定归属	—	边界任务按实测 α_comm 与实测带宽定档

表 D-2　图 6-1 参数表：模型档位（7B / 70B 级）、GQA 配置与 KV 显存占用（D-22）

模型档位	上下文长度	GQA 分组配置	分页块大小与命中率口径	单序列 KV 显存占用（压缩前 → 分页后）	可承载并发 32K 序列数
7B 级（32 层，d_model 4096，d_head 128）· 基线行	32,768 tokens	MHA（H_kv = 32，无 GQA 压缩）	块大小 16 tokens；命中率（块级，自然日窗口）设计估算 P50 ≈ 0.35	FP16：≈ 17.2 GB → ≈ 17.2 GB（基线行：无 GQA 压缩，分页仅消除碎片）	≈ 3 条（80 GB 单卡 − 7B 权重 14 GB = 66 GB 可用）
7B 级 + GQA	32,768 tokens	GQA 8 组（H_kv = 8）	块大小 16 tokens；命中率（块级，自然日窗口）设计估算 P50 ≈ 0.35	FP16：≈ 4.3 GB → ≈ 4.3 GB	≈ 15 条（66 GB 可用）
7B 级 + GQA + FP8	32,768 tokens	GQA 8 组（H_kv = 8）	块大小 16 tokens；命中率（块级，自然日窗口）设计估算 P50 ≈ 0.35	FP8：≈ 2.1 GB → ≈ 2.1 GB	≈ 30 条（66 GB 可用；FP8 使并发数翻倍）
70B 级（80 层，d_model 8192，d_head 128）· 基线行	32,768 tokens	MHA（H_kv = 64，无 GQA 压缩）	块大小 16 tokens；命中率（块级，自然日窗口）设计估算 P50 ≈ 0.35	FP16：≈ 85.9 GB → ≈ 85.9 GB（基线行：无 GQA 压缩）	0（单卡不可承载）
70B 级 + GQA	32,768 tokens	GQA 8 组（H_kv = 8）	块大小 16 tokens；命中率（块级，自然日窗口）设计估算 P50 ≈ 0.35	FP16：≈ 10.7 GB → ≈ 10.8 GB（与 §6.4 正文示例同配置）	≈ 1 条（单机双卡 160 GB − 70B 权重 140 GB = 20 GB 可用）
70B 级 + GQA + FP8	32,768 tokens	GQA 8 组（H_kv = 8）	块大小 16 tokens；命中率（块级，自然日窗口）设计估算 P50 ≈ 0.35	FP8：≈ 5.4 GB → ≈ 5.4 GB	≈ 3 条（20 GB 可用）

注：两档 MHA 行均为「无 GQA 压缩」的对比基线行，其压缩前后数值相同系预期行为——分页机制仅消除显存碎片，不做压缩，不应误读为分页无效果；压缩效果体现为 GQA 行相对基线行的显存下降（7B 级

17.2 → 4.3 GB，70B 级 85.9 → 10.7 GB）。70B 级档位与 §6.4 正文示例（80 层、H_kv = 8、FP16 单序列 ≈ 10.7 GB）为同一配置；7B 级为单卡并发场景的典型档位，两档分别对应「压缩效果对比」与「单卡可承载并发」两个复算目标。

参考文献

以下为本文所引用的关键学术论文、技术标准与官方文档。正文中的机制命名与公式口径均以上述来源为准；标注"设计目标"的指标为本文自行定义，不引用任何未经验证的第三方实测数据。线上来源的访问日期统一为 2026-09-14。

学术论文

1. Hoffmann et al., *Training Compute-Optimal Large Language Models*（Chinchilla 标度律，见 §6.2），2022，arxiv.org/abs/2203.15556
2. Rafailov et al., *Direct Preference Optimization: Your Language Model is Secretly a Reward Model*（DPO，见 §6.2），NeurIPS 2023，arxiv.org/abs/2305.18290
3. Hu et al., *LoRA: Low-Rank Adaptation of Large Language Models*，2021，arxiv.org/abs/2106.09685
4. Dettmers et al., *QLoRA: Efficient Finetuning of Quantized LLMs*，2023，arxiv.org/abs/2305.14314
5. Leviathan et al., *Fast Inference from Transformers via Speculative Decoding*（投机解码，见 §6.4），ICML 2023，arxiv.org/abs/2211.17192
6. Ainslie et al., *GQA: Training Generalized Multi-Query Transformer Language Models from Multi-Head Checkpoints*（见 §6.4），2023，arxiv.org/abs/2305.13245
7. Rajbhandari et al., *ZeRO: Memory Optimizations Toward Training Trillion Parameter Models*，SC 2020，arxiv.org/abs/1910.02054
8. Shazeer et al., *Outrageously Large Neural Networks: The Sparsely-Gated Mixture-of-Experts Layer*（MoE，见 §6.4、§9.3），ICLR 2017，arxiv.org/abs/1701.06538
9. Micikevicius et al., *FP8 Formats for Deep Learning*（见 §9.3），NVIDIA，2022，arxiv.org/abs/2209.05433
10. Dao et al., *FlashAttention: Fast and Memory-Efficient Exact Attention with IO-Awareness*（见 §9.3），NeurIPS 2022，arxiv.org/abs/2205.14135
11. Shah, Bikshandi, Zhang et al., *FlashAttention-3: Fast and Accurate Attention with Asynchrony and Low-precision*（FlashAttention-3，见 §9.3），2024，arxiv.org/abs/2407.08608

12. Williams, Waterman & Patterson, *Roofline: An Insightful Visual Performance Model for Multicore Architectures*（见 §5.2、§6.4），Communications of the ACM, 2009，doi.org/10.1145/1506609.1506623
13. Kahn, A.B., *Topological Sorting of Large Networks*（见 §7.1），Communications of the ACM, 5(11): 558–562, 1962，doi.org/10.1145/368996.369025
14. Erlang, A.K., *Solution of some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges*（M/M/c 与 Erlang C 排队模型，见 §6.3），Elektroteknikeren, 13, 1917

技术标准与官方文档

1. ERC-20 代币标准，Ethereum Improvement Proposals（见 §8.6）
2. OpenZeppelin Contracts 安全惯例（SafeERC20 / Math）（见 §8.6）
3. Solidity 官方文档（见 §8.6）
4. Model Context Protocol（MCP）规范（见 §7.1、§9.3）
5. NIST 后量子密码（PQC）标准化项目（见 §9.3）
6. WebAssembly（WASM）与 eBPF（见 §4.7、§9.3）

设计参照系

1. Ethereum Whitepaper——账户体系与 Gas 计量参照（见 §3.4）
2. Filecoin Whitepaper——抵押罚没与存储证明参照（见 §3.4、§4.5）