io.net如何建成去中心化算力平台？

来源：Trustless Labs

## 背景

随着OpenAI推出GPT4 LLM，各种AI Text-to-Image模型的潜力被见证，基于成熟AI模型的应用的日益增加，对GPU等算力资源的需求水涨船高。

GPU Utils 2023年一篇探讨Nvidia H100 GPU供需情况的文章中指出涉足AI业务的大型企业都对GPU有着强烈的需求，Meta、Tesla和Google等科技巨头都购买了大量的Nvidia GPU用于构建面向 AI 的数据中心。Meta拥有约21000个A100 GPU，Tesla 拥有约 7,000 个 A100，Google 的数据中心也有大量 GPU 投资，尽管没有提供具体数字。受训练大型语言模型 (LLM) 和其他 AI 应用程序的需求推动，对 GPU（尤其是 H100）的需求持续增长。

同时根据Statista的数据显示，AI市场规模从2022年的134.8 Billion增长到了2023年的241.8 Billion，并预计会在2030年达到738.7 Billion，而云服务的市值也由633 Billion增长大约14%，不乏其中的一部分归因于AI市场对GPU算力的需求迅速增长。

对于快速成长并且蕴含巨大潜力的AI市场，我们可以从什么角度去解构并且挖掘与之相关的投资切入点呢？根据IBM的一篇报告，我们总结了其中提到的对于创建和部署人工智能应用程序和解决方案所需的基础设施。可以说，AI基础设施主要是为了处理并优化训练模型所依靠的大量数据集和算力资源而存在的，从硬件和软件两方面解决数据集处理效率、模型可靠性和应用可扩展性的问题。



AI训练模型和应用中需要用到的大量算力资源，偏好低延迟的云环境和GPU的算力，在软件堆栈方面也包含了分布式计算平台（Apache Spark/Hadoop）。Spark通过将需要处理的工作流分散到各个大型计算集群中，并具有内置的并行机制和容错设计。区块链天然的去中心化设计思路使得分布式节点成为常态，且BTC创立的POW共识机制确立了矿工需要通过算力（工作量）竞争来赢取出块结果，这与AI同需算力来生成模型/推理问题有着类似的工作流。于是传统的云服务器厂商开始扩展新的生意模式像出租服务器一样出租显卡，贩售算力。而模仿区块链的思路，AI算力采用分布式系统设计，可以利用闲置的GPU资源，降低创业公司的算力成本。



### IO.NET 项目简介

Io.net 是一个结合了Solana区块链的分布式算力提供商，旨在利用分布式算力资源（GPU & CPU）来解决 AI 和机器学习领域的计算需求挑战。IO 通过集成来自独立数据中心和加密货币矿工的闲置显卡，联合Filecoin/Render 等crypto项目，汇聚了 100 多万个 GPU 的资源来实现解决AI计算资源短缺的问题。

技术层面，io.net基于实现分布式计算的机器学习框架ray.io构建，为AI应用提供从强化学习，深度学习到模型调优，模型运行等需要算力环节的分布式计算资源。任何人都可以以worker或developer的角色加入io的算力网络而无需额外许可，同时网络会根据计算工作的复杂性、紧急程度和算力资源供应情况调整算力价格，根据市场动态定价。基于算力分布式的特点，io的后端也会根据 GPU 需求类型、当前可用量、请求者的位置和声誉，将 GPU 提供商与开发人员进行配对。

$IO是io.net系统的原生代币，在算力提供者和算力服务购买者之间充当交易媒介，使用$IO与$USDC相比可以减免%2的订单手续费。同时$IO对确保网络的正常运作也起到重要的激励作用：$IO代币持有人可以质押一定数量的$IO到节点，节点运行也需要在有$IO代币质押的情况下才能获得机器空闲时期对应的收益。

$IO代币当前市值约为3.6亿美元，FDV约为$30亿。

### $IO 代币经济学

$IO的最大总供应量为800 million，其中500 million枚在代币TGE时就进行了各方对应的分配，剩余300 million枚代币会在20年内逐步递减释放（释放量每月递减1.02%，约为每年递减12%）。当前IO流通量为95 million，其中由TGE时为生态研发和社区建设解锁的75 million和币安Launchpool 20 million的挖矿奖励组成。

IO测试网期间算力提供者的奖励分配如下：

第 1 季（截至 4 月 25 日） - 17,500,000 IO

第 2 季（5 月 1 日 - 5 月 31 日） - 7,500,000 IO

第 3 季（6 月 1 日 - 6 月 30 日） - 5,000,000 IO

除测试网算力奖励之外，IO还给予了参与构建社区的创作者部分空投：

（首轮）社区 / 内容创作者 / Galxe / Discord - 7,500,000 IO

第 3 季 （6 月 1 日 - 6 月 30 日）Discord 和 Galxe 参与者 - 2,500,000 IO

其中第一季测试网算力奖励和首轮社区创作/Galxe奖励在TGE时已完成空投。

根据官方文档，$IO总体分配如下：







### $IO 代币销毁机制

Io.net根据一套固定的预设程序执行$IO代币的回购与销毁，具体的回购与销毁数量取决于执行时的$IO价格。用于回购$IO的资金来自IOG(The Internet of GPUs - GPU互联网)的运营收益，从IOG中的算力购买者和算力提供商中各收取0.25%的订单预定费用，以及用$USDC进行算力购买的2%手续费。





### 竞品分析

与io.net类似的项目有Akash，Nosana，OctaSpace，Clore.AI等专注于解决AI模型计算需求的去中心化算力市场。

Akash Network 通过一个去中心化的市场模型，利用闲置的分布式计算资源，汇集和出租多余的计算能力，并通过动态折扣和激励机制应对供需失衡，基于智能合约实现高效、无需信任的资源分配，从而提供安全、经济高效且去中心化的云计算服务。它允许以太坊矿工和其他拥有未充分利用的 GPU 资源的用户将这些资源租赁出去，从而创建一个云服务市场。在这个市场中，服务的定价通过反向拍卖机制进行，买家可以出价租用这些资源，推动价格竞争性降低。

Nosana 是Solana生态中的一个去中心化算力市场项目，主要意图是借助闲置算力资源形成GPU网格，满足AI推理环节的计算需求。该项目籍由Solana上的程序定义其算力市场运行情况，并确保参与网络的GPU节点合理完成任务。目前除在第二阶段的测试网运行中，提供针对LLama 2 和 Stable Diffusion 模型推理过程的算力服务。

OctaSpace 是一个开源可扩展的分布式计算云节点基础设施，允许访问分布式计算、数据存储、服务、VPN 等。OctaSpace 包括 CPU 与 GPU 算力，服务用于 ML 任务的磁盘空间、AI 工具、图像处理和使用 Blender 渲染场景等。OctaSpace 2022 年启动，在其自己的 Layer 1 EVM 兼容区块链上运行。该区块链采用双链系统，结合了工作量证明（PoW）和权威证明（PoA）共识机制。

Clore.AI 是一个分布式 GPU 超算平台，让用户可以从全球提供算力的节点获取高端 GPU 算力资源。它支持 AI 训练、加密货币挖矿和电影渲染等多种用途。该平台提供低成本、高性能的 GPU 服务，用户可以通过租赁 GPU 获得 Clore 代币奖励。Clore.ai 注重安全性，遵守欧洲法律，并提供强大的 API 以实现无缝集成。从项目质量上来说，Clore.AI的网页较为粗糙，没有详细的技术文档验证项目自我简介的真实性和数据真实性，我们对该项目的显卡资源和真实参与程度保持存疑。

和这几个同属去中心算力市场赛道中的其他产品对比，io.net是目前唯一一个任何人都可以无准入加入提供算力资源的项目，用户可以使用最低30系消费级 GPU参与网络的算力贡献，也有诸如Macbook M2，Mac Mini等Apple芯片资源。更充足的GPU和CPU资源和丰富的API构建使IO能够支持各种 AI 计算需求，如批量推理、并行训练、超参数调优以及强化学习等。而其后端基础设施则是由一系列模块化层构成的，能够实现资源的有效管理和自动化定价。 其他的分布式算力市场项目多为面向企业的显卡资源进行合作，用户参与存在一定门槛。因此IO可能具备使用代币经济学的加密飞轮撬动更多显卡资源的能力。

以下是io.net和竞品当前的市值/FDV对比：



### 回顾与结语

$IO上线币安可谓是给这个开局就备受关注，测试网的火爆全网以及在实测延期中渐渐被众人攻击、质疑积分规则不透明的重磅项目的画上了一个配得的开局句号。代币在市场回调期间上线，低开高走，最终回归相对理性的估值区间。但对于因io.net强大的投资阵容慕名而来的测试网参与者，几家欢喜几家愁，大部分租赁了GPU但没有坚持参与每一季测试网的用户并没如愿获得理想的超额收益，反而面临“反撸”的现实。测试网期间，io.net 将每一期的奖池分为 GPU 和高性能 CPU 两个 pool 来分别计算，season 1 由于出现黑客事件积分公布推迟，但最终TGE时 GPU 池的积分兑换比例确定在近 90：1，在各大云平台厂商处租 GPU 参与的用户成本远远超过空投收益。season2 期间官方完整的实现了 PoW 验证机制，近 3w 个 GPU 设备成功参与并通过 PoW验证，最终积分兑换比例为 100：1。

在万众瞩目的开局之后，io.net是否能实现它所宣称的为AI应用提供各种环节计算需求的目标，测试网后又还存留多少真正的需求，也许只有时间能给予最好的证明。