重构GPU算力架构：
解锁Tier 0超高速共享存储层

存储成本直降+能耗锐减+释放GPU算力周期

Hammerspace Asia
是我们的微信服务号
发布内容偏重于
公司新闻与行业方案
扫码可关注

EXECUTIVE BRIEF

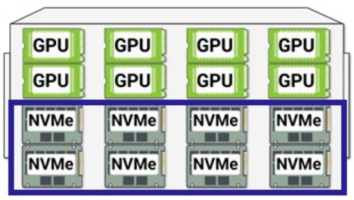
GPU服务器本地磁盘正在成为闲置资源

企业机构正在部署大规模GPU计算环境，以支持人工智能、企业级高性能计算（HPC）及其他形式的非结构化数据处理。通常，这些企业还会部署外部闪存存储系统来为GPU提供数据支持，而这类存储系统不仅增加成本与能耗，还带来了设备评估、采购及部署的时间与人力成本。

与此同时，GPU服务器中闪存存储通常处于闲置状态，主要原因包括：

1. **孤立式存储** —— 仅可供该服务器内GPU访问，集群中其他GPU无法使用
2. **缺乏数据保护** —— 若GPU服务器发生故障，则存在数据丢失风险
3. **数据迁移困难** —— 用户或应用难以在本地磁盘与外部存储间迁移数据

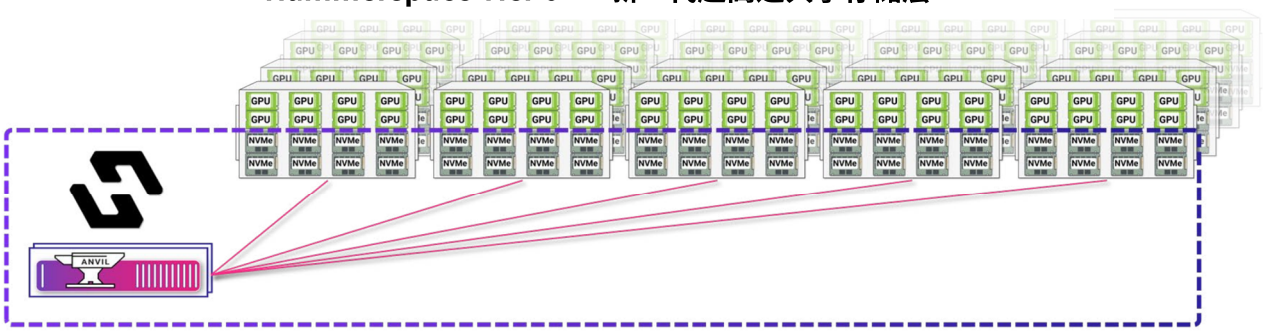
GPU服务器本地磁盘资源通常
并没有得到充分利用



单台GPU服务器的磁盘容量并不是一个小数字。当前单台GPU服务器可配置数十TB存储容量，预计明年单台服务器容量将达2PB。这意味着企业GPU集群中会出现PB级存储资源的“闲置资产”。

而通过将GPU服务器本地磁盘纳入Hammerspace全局数据平台，业界首次实现对这些高性能存储资源的新型利用模式，Hammerspace构建出了被称之为"第0级存储"（Tier 0）的超高速共享存储层。

Hammerspace Tier 0——新一代超高速共享存储层



释放闲置容量
实现成本与能效共同优化

Checkpoint训练提速
提升GPU有效算力

加速价值实现
缩短GPU基础设施回报周期

Tier 0——GPU计算基础设施架构革新

Hammerspace Tier 0技术通过将GPU服务器中闲置的“碎片化容量”转化为新型超高速共享存储层，重新定义了GPU计算基础设施架构。

存储于本地的数据不再孤立于单台GPU服务器，而是融入了一个可以跨越任意厂商本地及云端存储的统一全局命名空间。

现状： GPU 服务器本地磁盘容量处于闲置、未使用状态	**通过使用Hammerspace Tier 0技术：** 激活现有GPU服务器内置磁盘存储容量，将其作为新的超高速共享存储层
✗ **高存储成本：**外部闪存存储系统是非常昂贵的	✓ **降低存储成本：**降低对外部存储的需求，节省潜在的数百万美元成本
✗ **电源限制：**外部闪存存储系统和交换机意味着更多电源和空间的消耗	✓ **减少电力消耗：**每年可节省数百万千瓦时的电力
✗ **GPU闲置时间过长：**外部闪存存储系统性能并不够快	✓ **高GPU利用率：**checkpoint任务的处理速度提高20倍，释放GPU算力周期
✗ **价值实现需要时间：**评估、购买、部署外部存储设备都需要成本与时间	✓ **更快实现项目价值：**几分钟内即可使用存储容量；几天内上线 PB 级存储容量

Tier 0中的数据会自动获得保护，数据可以在Tier 0与外部存储系统（包括外部NAS存储、云存储乃至异地存储系统）之间实现无缝编排。

通过减少对外部存储的依赖，Tier 0可显著降低存储成本、电力消耗、能源开支及数据中心空间占用。

由于Tier 0的访问速度远超外部存储，关键的AI与HPC操作（如checkpoint保存）可实现10倍、20倍乃至100倍的性能提升。这意味着GPU服务器可大幅缩短checkpoint的操作时间，将更多资源用于计算任务。用户不仅能更高效利用现有GPU存储资源，每年还可获得额外的GPU算力增益。

最后，Hammerspace Tier 0容量可在分钟级完成部署，即使PB级容量也能在数天内上线运行，从而加速GPU计算投资的回报周期。

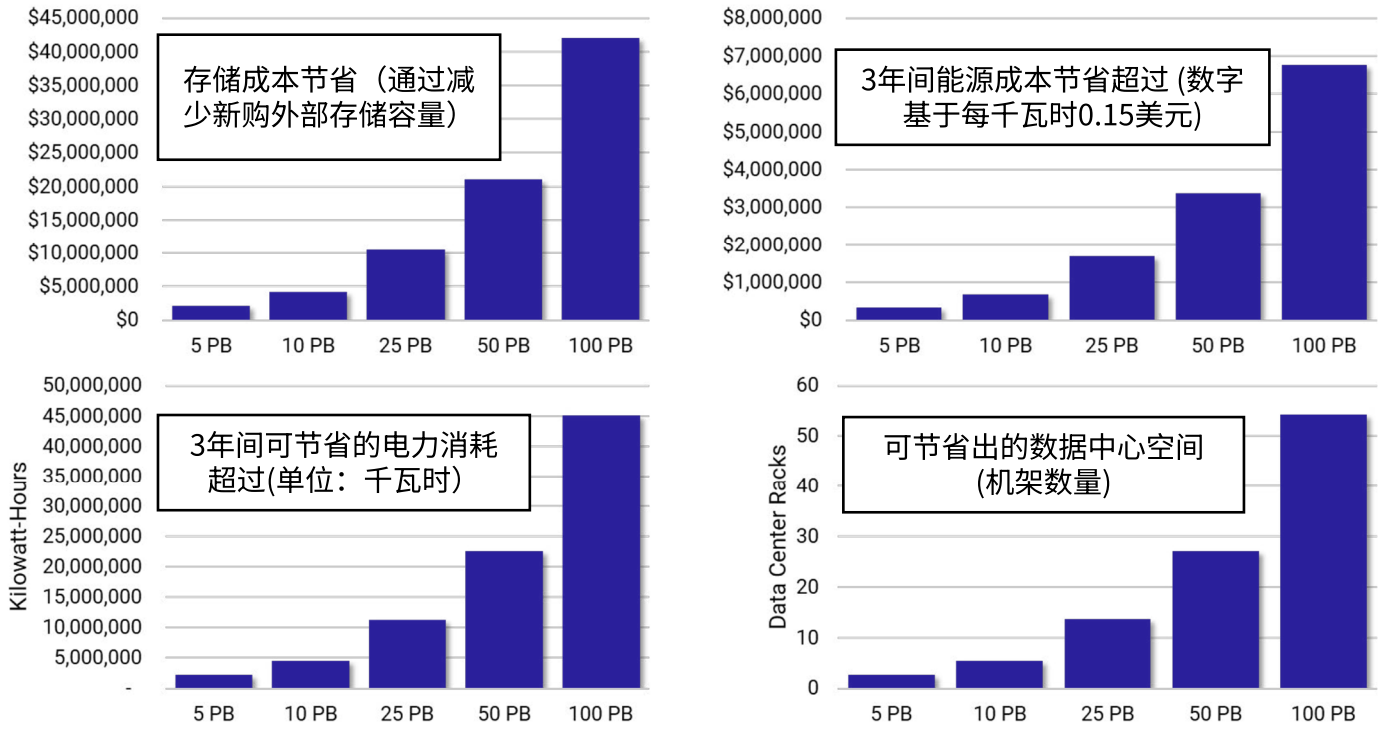
对于数百至数千台服务器规模的GPU计算集群（如NVIDIA SuperPOD架构）来说，用户可实现：

1. 节省数百万美元存储成本
2. 减少数百万千瓦时电力消耗
3. 每年提升10%~15%的GPU算力

订阅号 “悍雷尔空间”
发布内容偏重于
技术细节
扫码可关注

Tier 0 技术可显著节省成本、电力支出及机架空间

下图展示采用这种新型共享存储层可实现的成本、功耗与机架空间节省数量。需要说明的是，下图数据均为估算值，实际效果会因部署地域不同而有所差异，且具体成效取决于外部存储系统的成本与配置组合。如需评估特定环境下可实现的资源节省效果，欢迎联系Hammerspace进行定制化建模分析 - Get Started here



Tier 0技术——释放GPU算力周期，解锁GPU机会成本

除前文所述通过Tier 0容量减少外部存储需求带来的成本节省外，Tier 0技术还能通过加速checkpoint保存操作，实现提升GPU计算能力上限。

Checkpoint保存技术，被广泛应用于AI训练和HPC建模任务之中，是指周期性地将应用程序状态保存为持久化存储的过程。由于Tier 0存储的性能远高于外部存储，其在大规模环境中可将checkpoint保存速度提升10倍、20倍甚至100倍。checkpoint操作可压缩至秒级而非分钟级完成，这意味着GPU服务器每年可节省累计数百小时的checkpoint处理时间，用户可将这些时间投入数据处理任务之中。

这相当于在不增加任何GPU硬件的情况下，为现有集群提升10-15%的GPU算力！由此释放的GPU机会成本价值可达数百万美元，在大型环境中甚至可突破千万美元量级。

您可通过这份白皮书，了解将Tier 0技术用于checkpointing的详细效能提升分析。

