HAMMERSPACE

极速驱动AI工作流——破解数据孤岛难题，释放人工智能潜能

盘活现有存储资源，赋能GPU驱动的AI计算弹性扩展

订阅号 “悍宙尔空间”
发布内容偏重于
技术细节
扫码可关注

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

V. 1.4 - CN, 2025

Executive Summary - 在数据混乱中寻求秩序

想象一下，你是一位世界级交响乐团的指挥，试图在演奏家们分散在不同礼堂的情况下创作出一曲美妙的交响乐。虽然每位演奏家手中都有乐谱的一部分，但由于他们被墙壁隔开，你几乎无法让指挥将这些合成为所需的和谐旋律，从而确保演出成功。

对于所有规模化现代型IT组织来说，数字资产就如同这些演奏家手中的曲谱。尤其在尝试利用非结构化数据（如图像、音频、文本或其他不适合传统结构化数据库的文件）制定企业的AI战略时，这一问题尤为突出。

问题的核心在于，多年来非结构化数据的增长速度远超结构化数据，并且在各个行业中，非结构化数据已经占据超过80%的数字资产。更为严峻的是，这些非结构化数据（如图像、音频、文本及其他不适合传统数据库的文件）通常分散在不同供应商提供的本地和基于云的存储孤岛中，甚至跨越了多个地理位置。

具体来说，非结构化数据的增长速度已超出了其能够被有效分析和利用的能力范围。因此，这些数据不仅未能为组织带来额外的价值，反而逐渐成为了一种负担，并形成了一个不断扩大的成本中心。这些数据需要持续的管理和存储，但却几乎无法产生相应的投资回报。

> 企业无法承担放弃现有基础架构，并将非结构化数据迁移到新平台以实施AI战略的成本和风险。

使用数据分析、商业智能（BI）应用程序和数据仓库处理结构化数据已经是一个成熟的行业，从结构化数据中持续提取价值的策略也已广为人知。然而，生成式人工智能（GenAI）及相关深度学习（DL）技术的迅速发展，现在也为从非结构化数据中提取隐藏价值带来了希望。这类AI工作负载不仅能够帮助数据所有者了解他们拥有什么数据、应保留哪些数据或可以丢弃哪些数据，AI/DL应用场景还承诺帮助企业揭示以前隐藏在大量非结构化文件数据中的新价值。通过这种方式，组织最终将能够利用其结构化和非结构化数字资产来提取业务价值，并推动运营效率提升。

然而，数据孤岛、数据治理及其他与AI工作负载相关的问题

问题在于，非结构化数据孤岛的壁垒严重限制了组织快速实施AI计划的能力，同时成本和复杂性也可能失控。企业需要灵活地利用来自多个不兼容存储孤岛中的任何或全部数据，以支持GPU驱动的计算集群进行AI/DL工作流。传统上，这意味着将文件整合到全新的高性能存储设备之中，这给此类项目增加了显著的成本负担。这些高额的资本和运营成本将减缓AI计划的实施，并对预期的投资回报率提出了严峻质疑。

此外，AI应用模型正在迅速演变，可能需要不同子集的数据。这不仅造成了需要从一个孤岛复制数据到另一个孤岛的操作问题，还引发了严重的数据治理问题，包括合规性、适当的访问控制、可审计性和确保数据完整性等风险，因为数据副本正在不断增多。

为了在合理成本和时间内成功实施AI战略，组织需要具备安全地打破数据孤岛的能力，并通过适当的控制措施，直接将数据从任何供应商的存储类型中，提供给高性能的本地或基于云端GPU集群，理想情况下无需将数据复制到新的存储系统中。

如果您关注技术行业中的媒体报道，您会发现存储供应商们正掀起一股热潮，纷纷推出“万能解决方案”来解决AI工作流中的这一问题。然而，企业根本无法承受放弃现有基础设施，并将非结构化数据迁移到新平台以实施AI战略的成本和风险。此外，很少有数据环境能够整合到单一存储孤岛中，以满足数据生命周期所有阶段的需求。现有孤岛式存储架构，难以支持跨平台或高性能的GPU密集型AI数据工作流要求。

为非结构化数据打造可自由操作的架构

在本白皮书中，我们将阐述Hammerspace如何通过软件定义解决方案解决这些问题——该方案能自动化构建高性能AI数据通路，直接从现有分散的多厂商存储孤岛（包括跨多个地理位置及云平台），向GPU集群供给数据，而无需将其迁移至新存储系统之中。

我们将展示该方案在海量规模下的实践案例，重点解析Hammerspace如何支撑Meta构建全球最大规模的AI数据工作流。当前环境部署有两组各含24,000个GPU的集群，Hammerspace正被用于从1,000个存储节点以惊人的每秒12.5TB速度为Llama 2及下一代Llama 3大语言模型（LLM）训练输送数据。Hammerspace极致的线性扩展能力及对现有商用存储和服务器的兼容性，是Meta近期将AI工作流扩展至350,000个NVIDIA H100 GPU计划的关键支撑点。这属于Meta宣称"近期将建成相当于近600,000个GPU算力的AI研究与训练体系"战略布局的重要组成部分。

凭借Hammerspace**高性能并行全局文件系统**及可以桥接现有存储孤岛的自动化数据编排能力，该方案可以完美适配AI/DL工作流各阶段（无论规模大小）的多样化性能需求。

Hammerspace**高性能并行全局文件系统**

可桥接现有任意存储系统上的数据孤岛，

凭借其卓越性能，可满足任何规模的AI/DL工作流需求。

Hammerspace可以在混沌，且难分类的数据中构建秩序，其先进的、跨厂商存储孤岛文件元数据自动化索引能力，可将用户自定义标签与自动化生成的定制元数据整合至统一元数据目录。该目录可以有效简化跨平台数据的编排流程，为AI训练工作流提供高质量数据源。

全局元数据作为核心枢纽，通过精准定位三大关键维度：数据内容属性、物理/云端存储位置、关联项目与业务场景。这些元数据在Hammerspace中可被转化为可执行策略，构成AI/DL工作流，实现数据治理与数据质量管控的必要基础。

通过这种方式，Hammerspace在无需大规模数据迁移或替换现有存储基础设施的前提下，为分散的非结构化数据孤岛建立可操作的治理架构。

通过提供跨任意存储类型、任意位置数据的高性能工作流的统一访问与自动化管控，Hammerspace不仅能够帮助数据所有者充分利用现有存储资源，更能大幅加速分布式非结构化数据集在各类AI应用场景中的部署效率——**无论是本地还是云端、任何规模的GPU集群均可直接调用**。由于无需将数据复制到新存储设备单一位置之中，Hammerspace可将大语言模型训练周期及推理准备时间从数周缩短至小时级，同时显著降低构建AI工作流所需的综合成本。

Hammerspace可助力用户，显著加快使用分布式、非结构化数据集的AI工作负载处理速度，即便只是沿用现有的基础设施。

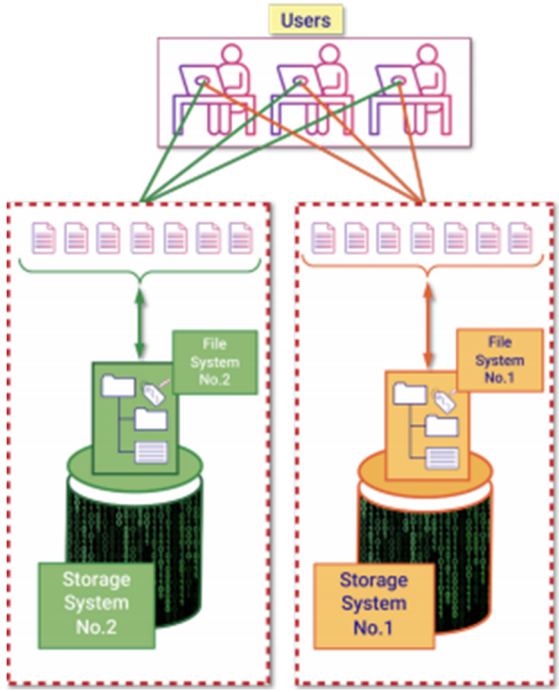
此外，我们将展示Hammerspace架构的灵活性是如何使用户摆脱受限于单一存储厂商解决方案的束缚——这种束缚会阻碍企业根据实际需求的变化，进行架构调整的能力。这一点非常关键，因为在AI应用落地的不同阶段，其性能需求存在显著差异。

例如，LoRA（低秩自适应）等新兴技术能够以远低于当前技术要求的性能指标完成模型微调。对数据所有者而言，这意味着他们可在不过度配置硬件资源的前提下，灵活调整现有基础设施。随着AI技术持续演进，这种能力将显著提升投资回报率（ROI）。

掌舵你的AI旅程

AI工作流往往包含多个阶段，不同AI/深度学习应用场景，会因行业或目标结果的不同而存在显著差异。特别是在非结构化数据处理领域，用于疾病检测的医学影像分析与视频数据中的行为识别存在本质区别；而文本数据中的情感分析（用于定向广告投放）又与前两者大相径庭；用于分析卫星图像，以预测农作物产量或指导灌溉/水资源管理的推理工作负载，与基于视频及其他传感器数据优化自动驾驶行为的预测模型也存在差异，而后者又与用于优化制造自动化流程的模型需求又截然不同

.尽管人工智能应用场景千差万别，但所有场景的共同需求，都是从多样化且分布广泛的数据源中进行数据采集。在典型的AI工作流程中，这通常意味着需要借助rsync等手动文件迁移工具，或其他单点解决方案进行大规模数据迁移，特别是为了向GPU集群输送训练和推理任务所需的数据。



上图：由多个文件系统造成的存储孤岛割裂了用户与应用程序的数据访问，迫使用户需要采用单点解决方案，在孤岛间复制数据。由于AI工作流程需要访问整合后的数据集，这种架构缺陷将导致成本增加和复杂性提升。

文件系统碎片化是核心症结

问题的本质在于，无论是人工操作还是AI/DL应用，对数据的访问最终都需要通过文件系统实现。文件系统，会将存储介质上的原始数据位，组织成人类可理解、应用程序可访问的文件和目录结构，这一过程是通过文件系统的元数据完成的——元数据充当原始数据与用户/应用所见文件结构之间的接口。

自20世纪90年代网络附加存储（NAS）技术问世以来，文件系统就被嵌入各个存储厂商的专有平台之中。尽管不同厂商通过行业标准的NFS或SMB协议呈现文件/目录结构，但其底层包含这些元数据的文件系统却形成相互隔离的厂商私有体系，彼此互不兼容。

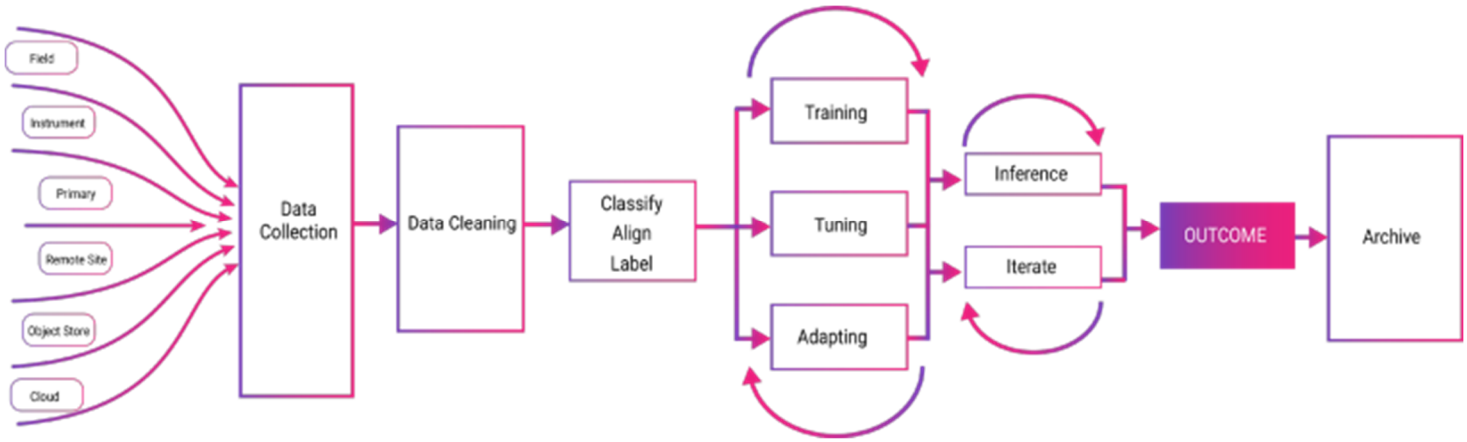
这种以存储为中心的设计方式导致：当数据量超出当前存储平台容量，或当性能需求与成本考量要求采用其他存储类型时，用户和应用不得不通过互不兼容的存储系统访问路径，来追踪迁移后的数据。**随着非结构化数据量的爆炸式增长，数据跨孤岛、跨地域、跨云平台分布的情况愈演愈烈，这种数据碎片化问题已完全失控。为弥合这些数据孤岛、跨站点和跨云环境，市场上甚至衍生出专门从事数据迁移、文件复制管理、分层存储解决方案、云网关等单点工具的细分产业，以应对由存储厂商壁垒造成的多文件系统访问与控制碎片化问题。**

数据孤岛对AI/DL工作负载的负面影响尤为突出

对于AI/DL工作负载而言，这个问题显得尤为尖锐。因为其关键的第一步就是：整合多源数据以建立统一视图。AI工作负载必须访问完整数据集，首先对文件进行分类/标记，才能确定哪些数据需要进入处理流程。然而碎片化的存储架构严重阻碍了这一基础步骤的实现。

在AI/DL的全应用周期中，数据将经历逐级精炼的过程：包括数据清洗、分类标注，直至大语言模型（LLM）训练与调优。每个处理阶段对计算和存储的性能要求均存在显著差异——从低速、低成本的海量存储与归档系统，到需要高性能GPU集群与NVMe存储设备的极速处理场景。

数据管理者面临的核心挑战在于：**如何通过统一架构同时满足差异化性能需求。即既要承载无需高性能支撑的数据分类标注环节，又要为GPU训练/调优及后续推理提供极致IO吞吐，后者通常需要NVMe存储级别的超高性能支撑。**



AI 工作流会从各种本地和远程数据源中提取数据，然后经过多个步骤，每个步骤都有不同的计算和存储要求。

企业正面临两难抉择：要么过度配置基础设施，采用全量高性能存储将所有AI周期数据集中存储，要么支付“数据搬运税”——在不同性能层级的存储孤岛间迁移数据副本，导致成本攀升、风险加剧及业务的延迟。当数据分布于多站点或云端时，这种副本损耗将更加严重，甚至可能造成昂贵的GPU集群因等待数据拷贝而空转或停机。

这种困境的痛点在于，企业已对现有基础设施进行大量投资，若直接替换为可满足全性能需求的新厂商绑定型专用存储平台，其成本将难以承受。此外，AI技术发展日新月异，若绑定适配当前AI工作流的存储方案，企业可能错失新兴技术的应用机遇。

由于AI工作流高度依赖稀缺的GPU集群，企业还需具备灵活调用云端GPU集群、或远程GPU即服务（GPU-as-a-Service）提供商的能力。

无论是在投资新的基础设施还是扩展到基于云的资源，增加“数据复制税”的复杂性和延迟都是一项显著的额外成本。这使得组织很难确定是否能够真正实现其在AI旅程中的投资回报。

破解数据孤岛难题，释放人工智能潜能

实现文件系统与基础设施层的解耦

Hammerspace通过多年的研发投入，从根本上重新设计了基于标准的高性能文件系统，成功解决了这一难题。该系统独立于专有存储基础设施，同时兼容任何厂商的现有存储系统。

与传统存储平台将文件系统嵌入专有硬件的做法不同，Hammerspace作为软件解决方案，可与任何本地或云端存储平台无缝兼容。其本质在于构建了一个凌驾于存储系统之上的高性能文件系统。通过这种方式，Hammerspace并行全局文件系统能够跨越原本互不兼容的存储孤岛——无论这些存储系统来自哪个厂商，无论其部署在本地还是云端。

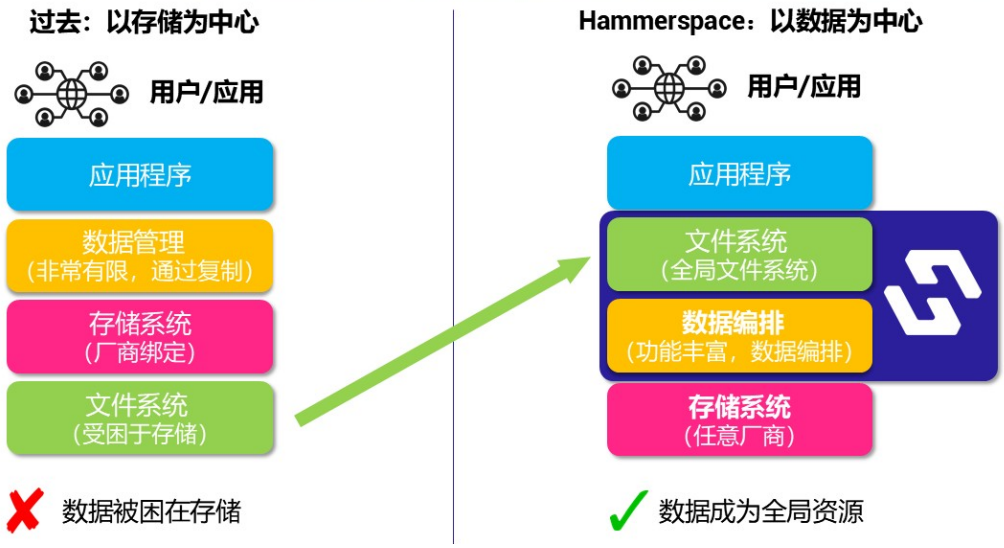
该系统的核心实现路径包括：
元数据智能同步：将现有存储系统中的元数据进行无缝整合
标准化协议接入：基于SMB、NFS、S3等标准协议呈现全局文件系统
零侵入式部署：存储端无需安装代理程序，客户端无需安装专用软件

对用户和应用而言，Hammerspace并行全局文件系统，提供与企业级存储平台完全一致的标准化挂载点。但其与传统方案的本质区别在于：

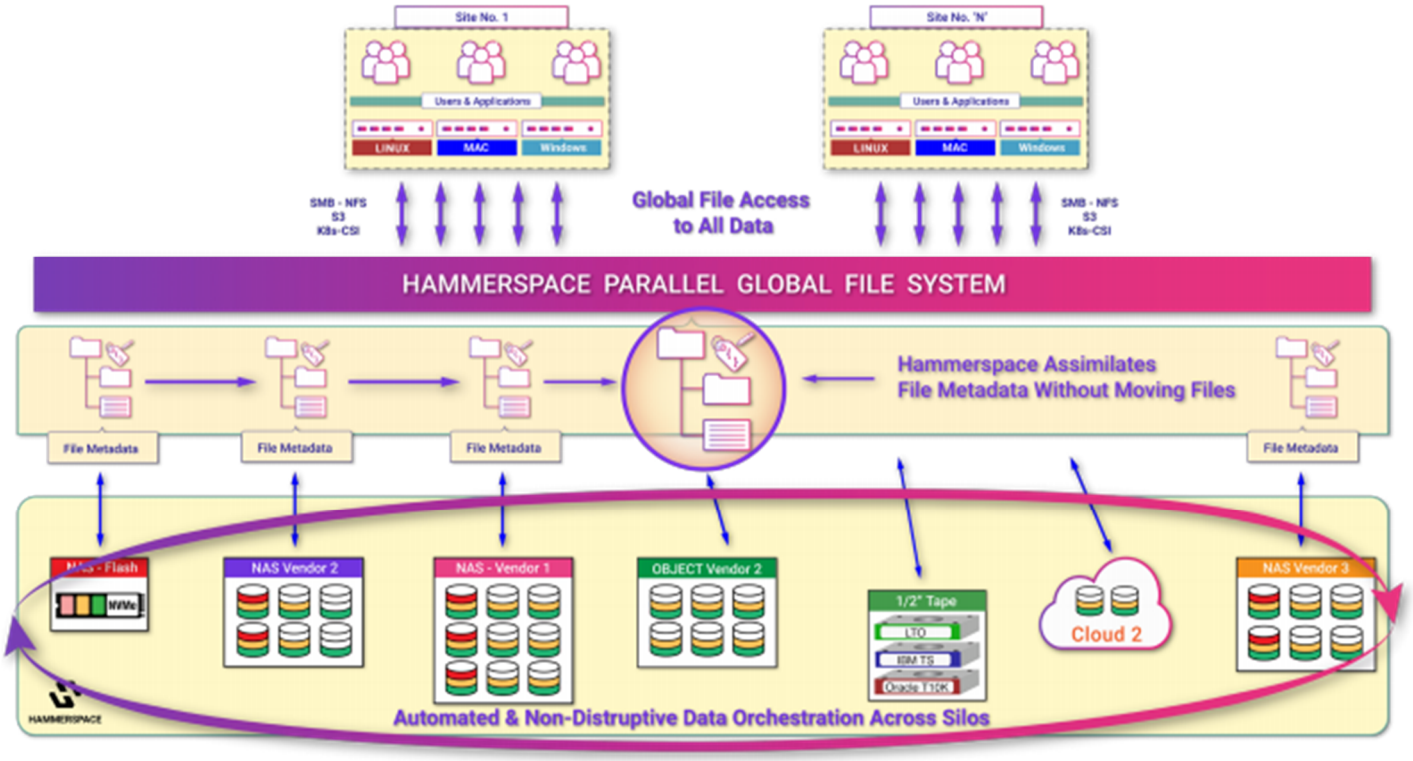
通过跨平台全局命名空间实现全域数据访问

突破多数据孤岛、多区域位置、多云平台的物理边界

保持行业标准协议兼容性的同时，提供革命性的数据流动性



上图：Hammerspace将文件系统元数据从专有基础设施中解耦，实现了全局化的文件统一访问，并支持跨任意存储系统的无干扰数据编排



上图：Hammerspace 可以整合整体文件系统的元数据，同时将数据保留在现有存储位置。通过这一方式，所有用户和应用程序均可借助标准协议，通过这一高性能文件系统在全球范围内访问所有数据。无需安装代理或专有客户端。

驾驭GPU算力巨兽：最大化GPU计算效能，驱动AI工作流提速

但仅打通数据孤岛实现企业数据的全局统一访问，并不足以解决AI难题。关键在于如何为可能分布在本地、云端或远程站点的高性能GPU集群持续输送数据，同时避免加剧数据复制与孤岛问题。

Meta的超大规模环境是一个极端案例——其AI工作流包含两个集群：24,000块GPU与1,000个存储节点。规模较小的企业环境可能只需扩展到数十块GPU，数据量甚至不足1PB。但无论规模大小，传统存储孤岛造成的算力供给瓶颈本质上是相同的。

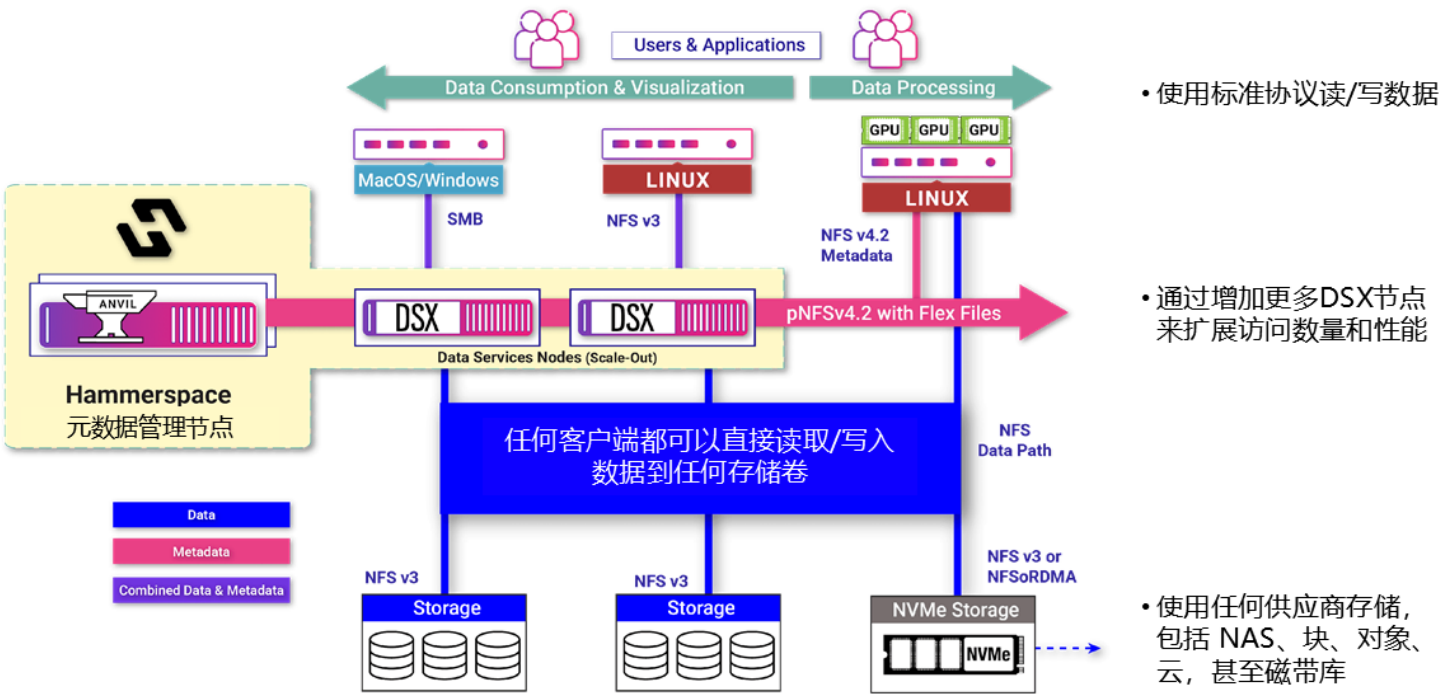
核心命题在于：如何在不强制企业将所有数据，迁移至定制化高性能（且昂贵）的新数据孤岛的前提下，满足GPU计算的数据供给需求。

Hammerspace通过构建跨孤岛的全局数据环境，解决了全局数据访问与编排的第一重挑战。如今推出的超大规模NAS架构，进一步融合HPC并行文件系统的极致性能与线性扩展能力，以及标准协议连接、简易性和企业级NAS特性，完整攻克了这一双向难题。

超大规模NAS架构可为专用存储赋予HPC级性能

超大规模NAS是一种全新架构类别的NAS存储系统，其是基于当前业界通用的标准Linux发行版中的开放标准构建的。在Linux内核中——几乎每个数据中心的现有服务器和虚拟机管理程序都内置有该内核——蕴含着关键组件，使得基于标准的HPC级并行文件系统性能得以在商用硬件乃至任何数据中心或云厂商的现有传统存储类型上实现。**这不仅意味着企业可在不改变原有数据位置的前提下，将数据供给本地或云端GPU集群运行AI工作负载，更能为现有横向扩展NAS平台注入加速动能。**

核心突破在于Hammerspace对Linux标准并行NFS v4.2与Flex Files实现的深度集成。值得注意的是，Hammerspace工程师历时多年的研发，并将其贡献至Linux开源社区，成为行业标准之一。该标准通过将元数据控制平面与数据路径分离，将并行文件系统的性能引入传统企业存储环境。**Hammerspace超大规模NAS实现了客户端与存储节点间I/O的极致并行化，其能力堪比Lustre等高性能计算（HPC）并行文件系统。关键差异在于：Hammerspace方案可基于常规商用硬件、标准化网络、开放标准接口，且无需迁移数据或在客户系统安装专有客户端软件或硬件。**而Meta的超大规模部署就是明证——该公司在现有基础设施上部署Hammerspace超大规模NAS架构，为其Llama 2和Llama 3的下一代AI工作流提供支撑，覆盖超1000个存储节点。这些系统每日处理数百万亿次AI模型运算，其规模已达任何标准都难以度量的境界。



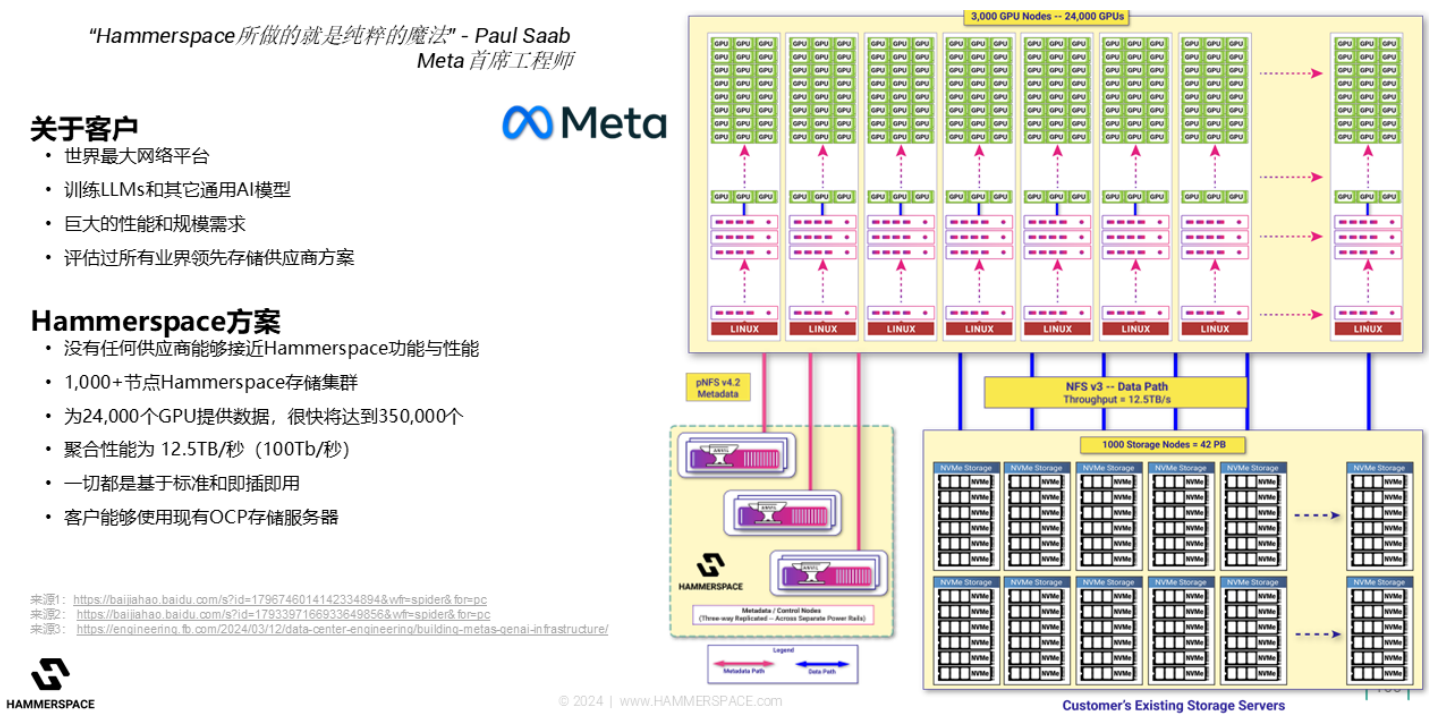
上图：Hammerspace 借助标准并行 NFSv4.2 与 Flex 文件布局技术加速通用存储以支持AI工作流。此图展示了即便是在混合环境中，用户仍旧可访问同一数据集，同时这些数据集可以是存在不同存储类型设备中，但依旧可以为GPU集群提供支持，且无需形成数据孤岛。

在项目选型过程中，Meta评估了多个横向扩展NAS供应商方案，但由于其架构性能扩展上限难以突破数十个存储节点的限制，最终予以排除。尽管能够支持超大规模扩展的HPC并行文件系统也被纳入考量范畴，但同样因以下原因被否决：系统管理复杂度过高，无法满足Meta对RAS（可靠性、可用性、可维护性）的严苛要求，更重要的是这些HPC系统必须依赖专用客户端软件、特殊硬件设备及InfiniBand等专用网络架构。

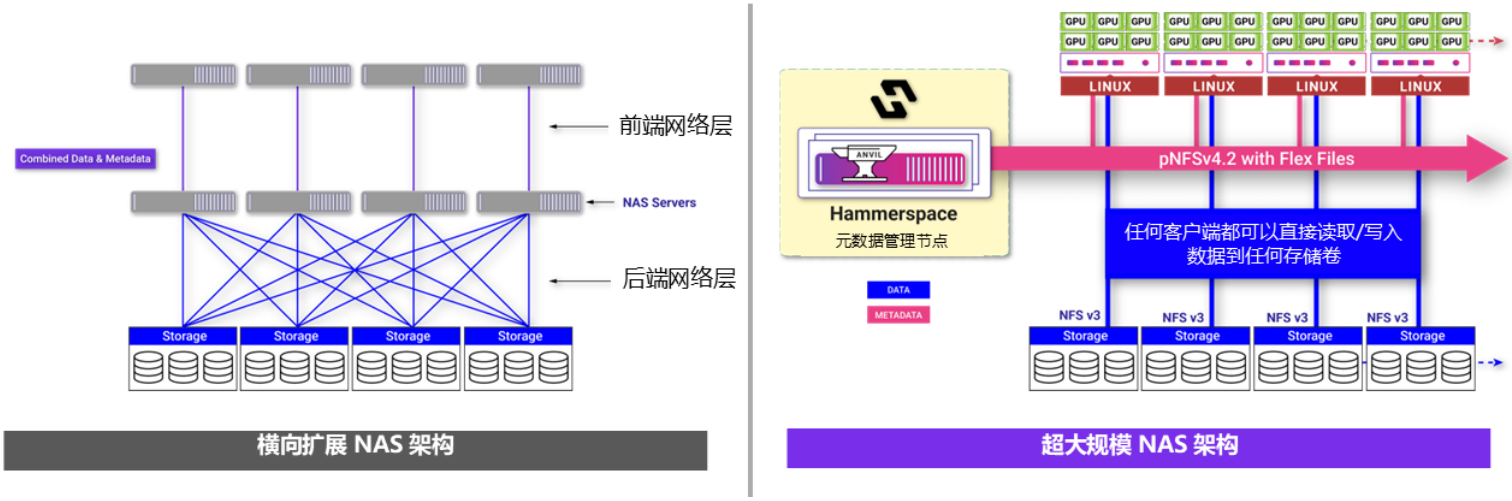
Hammerspace最终入选的关键在于,其能够无缝集成到Meta现有的环境中——无需迁移既有存储数据、无需安装客户端软件、无需改造现有存储架构。该方案成功在标准以太网上通过标准服务器实现并行文件系统性能，并通过线性扩展能力与冗余设计全面满足Meta在数据保护和业务连续性方面的严格要求。

面对Meta即有的42PB规模、由1,000个NVMe存储节点构成的存储阵列，Hammerspace展现出卓越的元数据快速同步能力，在无需数据迁移的前提下，即可为研究人员和AI训练负载提供标准NFS访问接口。

项目的另一核心要求是实现三重冗余容灾机制，以保障Meta关键AI基础设施的永续运行。整个数据中心采用三路数据副本架构，并部署在三路独立供电线路上，确保系统能够抵御逻辑故障与物理故障的双重风险。



上图: Meta 的 AI 工作流采用了 Hammerspace 的 Hyperscale NAS 架构，借助现有的通用存储、通用服务器和网络，创建了一个超大规模的、基于标准协议的可扩展 LLM（大型语言模型）训练基础设施。



上图： 横向扩展NAS架构（左）受限于专用NAS存储的瓶颈，以及额外的内部InfiniBand或基于RDMA的NFS网络瓶颈，存在扩展性限制。超大规模NAS架构（右）通过NFSv4.2的标准并行化协议突破了这一瓶颈，实现客户端到存储节点的直连并行化I/O，消除内部缓存争用与冗余网络跳转，从而可以加速现有横向扩展NAS存储性能。

中小规模环境中，采用Hammerspace的超大规模NAS架构，甚至能够加速现有横向扩展NAS平台，突破存储厂商原生性能限制，满足GPU密集型工作流需求。

自动化数据编排是AI工作流的关键

原始性能只是满足AI通路向GPU集群供给数据的必要前提。随着GPU资源日益难以获取，组织往往不得不转向云端GPU或GPU即服务提供商（GPU-as-a-Service） 。这进一步提高了技术门槛——IT团队在确保数据供给这些集群时，如何避免加剧数据孤岛问题。

而Hammerspace通过非破坏性自动化数据编排能力解决这一难题，其能在后台无感执行，实现跨孤岛、跨站点、跨云的数据自动编排，向任意位置的GPU集群输送数据。这意味着基于工作流的数据放置、数据保护等服务均可实现全自动化，且不会中断用户或应用程序访问。甚至能在应用程序或用户持续使用数据的过程中完成实时编排，全程无访问中断或工作流干扰。

传统存储架构的局限性在于，文件系统嵌入存储硬件后，当文件需要迁移至其他存储类型或位置时，必须创建并传输元数据与数据本体的副本。该操作会产生需要后期处理的分叉副本，同时消耗额外存储容量。而副本激增会增加数据治理风险，引发数据访问权限管控问题。若要在传统孤岛式系统中维护跨所有副本的审计跟踪，即便可能也是极为困难的。

Hammerspace并行全局文件系统的革新优势在于：由于该文件系统独立于存储层，彻底消除了管理分叉副本的需求。通过统一的全局文件系统，所有地点的用户和应用程序，均能获得全局数据的读写权限——访问的不再是文件副本，而是如同访问本地NAS存储般的单一文件实体。

这种方式下，如果需要使用基于云端GPU或GPU即服务（GPU-as-a-Service）解决方案，Hammerspace可以透明，且自动化地平衡不同的计算爆发式工作流，而无需创建文件副本。数据由于始终处于组织的控制之下，因此可以全局管理和治理，且满足合规性要求。

实现跨孤岛与不同地点的AI流程自动化

Hammerspace通过统一全局数据与元数据管控能力，打破异构存储系统、物理位置及云环境间的壁垒，为AI全生命周期（从数据准备到模型推理）提供端到端的自动化支持，无缝调度本地与远程资源。实现这一能力的核心能在于：自然语言策略引擎——Hammerspace数据编排系统的核心创新在于“服务级别目标（Service-Level Objectives）”策略定义：

- 使用简明自然语言制定数据访问规则、存储位置策略及保护机制
- 智能调度多厂商存储资源（NVMe/HDD/对象存储/磁带/云）
- 集成关键数据服务（版本控制、加密、跨平台容灾）

AI工作流自动化实践

1.数据清洗阶段

自动将原始数据迁移至预处理中心（配备NVMe磁盘层），业务系统仍可实时访问活跃数据

2.模型训练阶段

按策略将数据集分发至指定训练环境（本地GPU集群或云端GPU资源池），规避人工拷贝延迟

3.推理部署阶段

动态推送优化模型至边缘节点，并通过自定义元数据（算法版本、参数等）追溯迭代关系

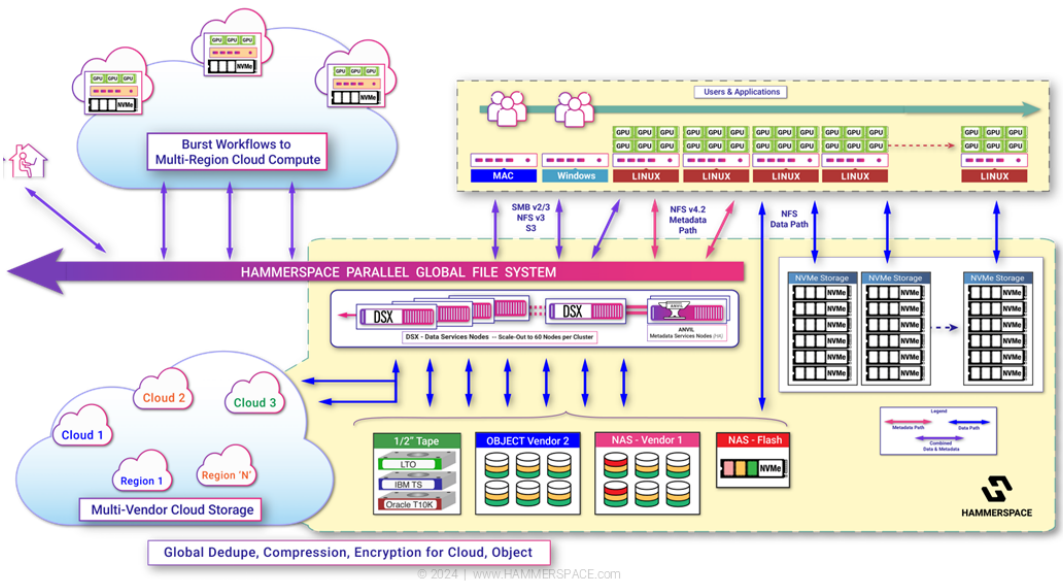
技术价值闭环

•无损协作：数据跨地域流动时维持文件锁与版本一致性（毫秒级）

•弹性架构：支持从TB级试点到EB级生产环境的平滑扩展

•审计就绪：自动化生成数据血缘图谱，满足AI可解释性监管要求

•资源优化：按需调用临时的云上HPC集群（配备GPU资源），任务完成后自动释放



上图：Hammerspace 是一个软件定义的解决方案，它可以从小规模开始部署，并能横向扩展以满足最极端的性能需求，实现跨本地、多站点以及一个或多个云的自动数据编排工作，为AI工作流提供数据支持。

赋能数据科学家，完成自助式工作流

此外，由于制药、金融服务或生物技术等行业需要出于合规要求对训练数据及生成模型进行归档，因此自动将这些数据保存到低成本存储的能力至关重要。通过自定义元数据标签追踪数据溯源、迭代细节及工作流程各环节，从归档中调用旧模型数据进行复用或应用新算法，可简化为后台自动执行的常规操作。

因此，借助Hammerspace，数据科学家能够直接以类似自助的方式，掌控AI全流程——跨多地域、存储孤岛及云端，无需向IT管理员申请数据检索，也无需介入基础设施管理。由于数据可直接从现有存储资源中无缝访问，这些工作流能够就地利用数据，无需替换传统存储系统。

总结

正如本白皮书所述，Hammerspace从设计之初便致力于解决数据中心内部及日益增长的分布式系统（跨多数据中心与云端的高性能应用场景）中的数据孤岛碎片化问题。

为适应AI/深度学习（DL）工作负载的快速转变，现有技术架构放大了IT组织长期面临的孤岛困境。这些挑战正在呈现叠加态势：

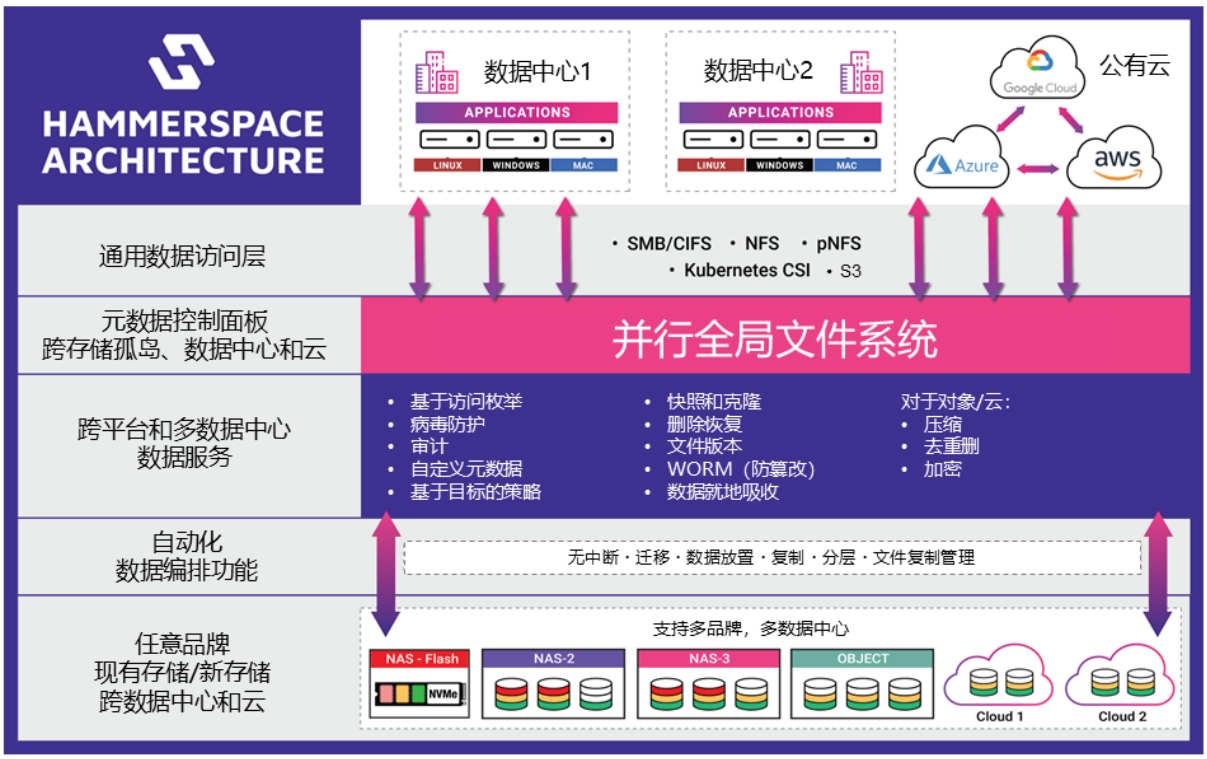
- **高性能与超大规模：AI工作流需具备无限制的纵向/横向扩展能力以满足极致性能需求。在避免基础设施过度配置的同时支持云爆发（Cloud Bursting）是核心诉求。**
- **多源数据整合：为保持竞争力并管理新型AI工作负载，需实现本地孤岛、多地域及云端数据的无缝访问。**
- **数据治理：在动态环境中需保持敏捷性，而固定基础架构常因成本或运维限制，难以得到扩展。现代企业需要的是，自动化地编排跨多厂商孤岛资源或云端/远程GPU集群的数据，同时保持可审计性及合规、安全等治理要求。**
- **标准化架构：企业需基于行业标准协议桥接现有基础设施与新型分布式资源，可避免安装专有客户端软件或代理程序。结合自动化数据编排，确保AI/DL工作负载的实施成本不会吞噬预期收益。**

- 云资源需求的无缝融合（Seamless Burst-to-Cloud）无论是由于GPU采购困难，还是因为某些小型AI场景仅需短期的计算/存储资源，快速向临时云资源进行"爆发式扩展"的能力都是关键需求。通过扩展Hammerspace文件系统并最大限度减少数据迁移来实现这一目标，正是此类计划成功所需的灵活性与适应性之关键。

Hammerspace软件能够完美解决客户启动AI流程时面临的挑战和需求，且无需通过新建存储与计算基础设施来改造数据中心的现状。同时，使用Hammerspace的客户也不再需要手动在不同厂商的存储孤岛间迁移文件副本，并为此支付"数据复制税"。

由于Hammerspace高性能并行全局文件系统（Parallel Global File System）可以真正实现数据的全局化，且数据编排可跨越所有的存储孤岛和位置，实现无缝自动化；AI工作负载如今能够优化配置，并快速适配新出现的AI应用，即使面对极端性能需求也能应对自如。

为了应对AI工作流中的众多性能变量，我们需要一种能够有效弥合本地存储孤岛与云环境之间鸿沟的新范式。这种解决方案需要新的技术和革命性的方法——**将高性能文件系统从任何存储类型的硬件中解耦。**使得AI工作流可以完全建立在任何用户的现有基础架构之上。这场变革的重要性，堪比90年代，NAS存储厂商将文件系统从操作系统中实现解耦的关键突破。而**这，就是Hammerspace的创新所在。**



上图是Hammerspace软件能力栈的逻辑视图，这些能力栈始于自身的高性能并行全局文件系统和数据编排功能。