释放超融合的全部潜能

提供采用下一代数据平台超融合解决方案

采用 Intel® Xeon® 处理器的思科 HyperFlexTM 系统

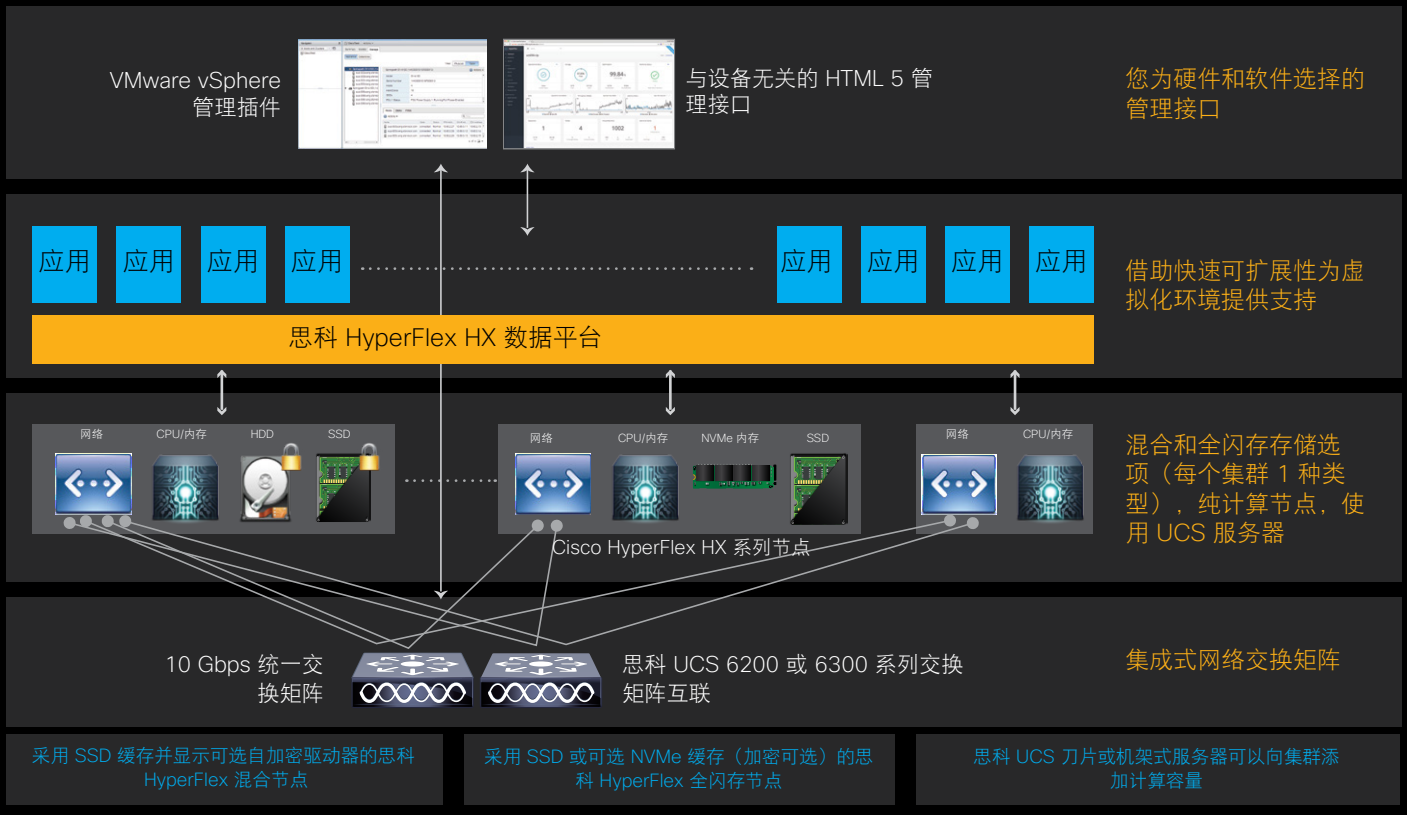
适用于自适应基础设施的平台

应用需求的不断变动导致服务器、存储系统和网络交换矩阵之间的关系不断变化。虽然虚拟环境和第一代超融合系统解决了某些问题，但已经跟不上应用和业务的发展速度。这就是当今新的 IT 运营模式需要新的 IT 消费模式的原因。基于思科统一计算系统™（思科 UCS®）设计的思科 HyperFlex 系统可释放超融合解决方案的全部潜能，提供您需要的灵活性、可扩展性、安全性和生命周期管理功能，真正实现操作简便性。通过部署思科 HyperFlex 系统，您可以同时利用由云带来的“随增长随投资”消费模式和本地基础设施的优势。

快速灵活的超融合系统

思科 HyperFlex 系统采用专门设计的端到端软件定义的基础设施，可消除第一代产品中存在的不足。思科 HyperFlex 系统将各种软件定义的功能集于一身：通过思科 UCS® 服务器实现软件定义的计算；通过强大的思科 HyperFlex HX 数据平台软件实现软件定义的存储；通过能够与思科以应用为中心的基础设施（思科 ACI™）解决方案轻松集成的思科® 统一交换矩阵实现软件定义网络 (SDN)。凭借混合或全闪存存储配置以及管理工具的选择，思科 HyperFlex 系统提供了预集成集群，该集群在一小时或更短时间内即可启动并正常运行，并且能独立扩展资源，最大程度地满足应用资源需求（图 1）。

图 1：思科 HyperFlex 系统提供下一代超融合解决方案



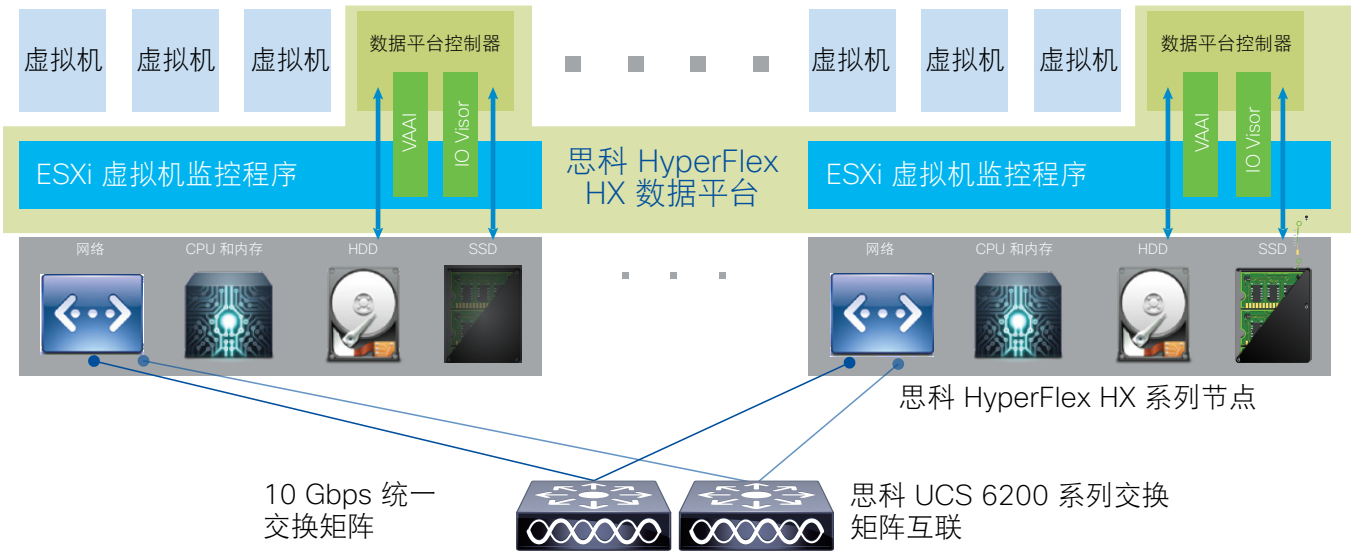
思科 HyperFlex HX 数据平台：存储优化的新高度

各种应用（特别是虚拟服务器上托管的应用）独特的数据需求产生了许多存储孤岛。HX 数据平台是思科 HyperFlex 系统的基础，是专为超融合环境构建的高性能、结构化日志、横向扩展文件系统。该数据平台的创新重新定义了横向扩展和分布式存储技术，超越了第一代超融合基础设施的范围，提供了广泛的企业级数据管理服务。

HX 数据平台包括以下功能：

- **企业级数据管理**功能可在分布式存储环境中提供完整的生命周期管理和增强的数据保护。这些功能包括复制、无间断线内重复数据删除、无间断线内压缩、精简调配、节省空间的即时克隆以及快照。
- **支持混合和全闪存模式**，能让您根据容量、应用、性能和预算要求选择合适的平台配置。
- **简化的数据管理**将存储功能集成到现有管理工具中，能够对应用执行即时调配、克隆和基于指针的快照操作，大幅简化日常运营。
- 通过高级自动化和协调功能**提高了可控性**，同时通过强大的报告和分析功能提高了对 IT 运营的可视性和洞察力。
- **独立扩展**计算、缓存和容量层，使您能够根据不断变化的业务需求灵活地横向扩展环境，有效实现可预测的“随成长随投资”。在添加资源时，数据会无中断地在集群间自动再平衡，充分利用新的资源。
- 采用线内重复数据删除和压缩**持续优化数据**，提高了资源利用率，为数据扩展提供更多空间。
- **动态数据分布**支持所有集群资源参与 I/O 响应，从而优化了性能和恢复能力。混合节点组合使用固态磁盘 (HDD)（用于缓存）和普通硬盘驱动器 (HDD)（用于容量层）。全闪存节点在缓存层使用 SSD 或非易失性存储器 (NVMe) 存储，在容量层使用 SSD。此方法有助于消除存储热点并使集群的性能可用于所有虚拟机。如果某个驱动器出现故障，系统可以快速执行恢复，因为系统可以使用集群内剩余组件的汇聚带宽访问数据。
- 采用可自行恢复的高可用性架构实现**企业数据保护**，支持无中断滚动升级，同时提供全年全天候 Call Home 和现场支持选项。
- **基于 API 的数据平台架构**提供数据虚拟化灵活性，支持现有和全新的云原生数据类型。

图 2：分布式思科 HyperFlex 系统



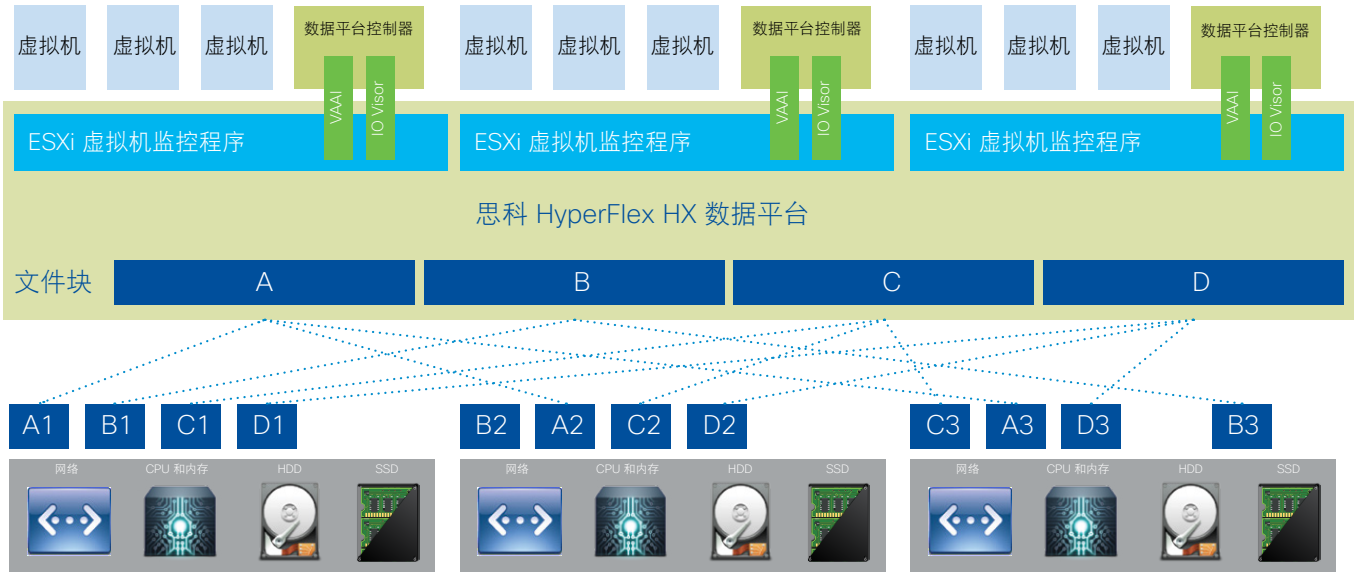
架构

在 Cisco HyperFlex 系统中，数据平台可跨三个或多个 Cisco HyperFlex HX 系列节点创建高度可用的集群。每个节点包含一个 HX 数据平台控制器，该控制器使用基于闪存的内部 SSD 或结合使用基于闪存的 SSD 和高容量 HDD 来存储数据，从而实现横向扩展和分布式文件系统。控制器之间通过万兆以太网通信，可提供跨集群节点的单一存储池（图 2）。节点通过使用文件、块、对象和 API 插件的数据层访问数据。随着节点的增加，集群可进行线性扩展，以提供计算、存储容量和 I/O 性能。

在 VMware vSphere 环境中，控制器通过固定的处理器核心数和内存容量占用虚拟机，使其能够提供一致的性能，且不会影响集群中其他虚拟机的性能。控制器可在没有虚拟机监控程序干预的情况下通过 VMware VM_DIRECT_PATH 功能访问所有存储。在全闪存内存配置中，控制器将节点的内存（专用于写入日志记录的 SSD 驱动器）和其他 SSD 用于分布式容量存储。在混合配置中，控制器将节点的内存和 SSD 用作分布式缓存层的一部分，并将节点的 HDD 用于分布式容量存储。控制器通过使用两个预安装的 VMware ESXi vSphere 安装捆绑包 (VIB) 将数据平台集成到 VMware 软件中：

- **IO Visor**：此 VIB 提供一个网络文件系统 (NFS) 安装点，从而使 ESXi 虚拟机监控程序可以访问与单个虚拟机连接的虚拟磁盘驱动器。从虚拟机监控程序的角度来看，它仅仅与 NFS 连接。
- **用于阵列集成的 VMware vStorage API (VAAI)**：此存储卸载 API 允许 vSphere 请求快照和克隆等高级文件系统操作。控制器通过操作元数据而不是实际复制数据来完成上述操作，从而实现快速响应，并快速部署新应用环境。

图 3：数据跨集群节点条带化分布



工作原理

HX 数据平台控制器处理虚拟机监控程序所访问卷的所有读写请求，从而协调虚拟机的所有 I/O。（虚拟机监控程序有一个独立于数据平台的专用启动盘。）此数据平台实施的分布式结构化日志文件系统始终使用 SSD 中的缓存层加速写入响应；在混合配置中使用 SSD 中的文件系统缓存层加速读取请求；使用 SSD 或 HDD 实施持久层。

数据分布

传入数据通过缓存层分布到集群中的所有节点，从而实现性能优化（图 3）。将传入数据映射到均匀存储于所有节点的条带单元，从而有效分布数据；数据副本数量由您设置的策略决定。应用写入数据时，系统将数据发送至基于条带单元的适当节点，条带单元包含相关信息块。此数据分布方法通过结合同时进行多个流写入的功能，可防止出现网络和存储热点；无论虚拟机位于何处，均可提供相同的 I/O 性能；还可为您的工作负载分布提供更高的灵活性。其他架构使用的位置方法不能充分利用可用的网络和 I/O 资源。

- **数据写入操作：**在进行写入操作时，系统会将数据写入本地 SSD 缓存，并在确认写入操作前将数据副本同时写入远程 SSD。
- **数据读取操作：**对于全闪存内存配置中的读取操作，系统会直接从分布式容量层中的 SSD 读取本地和远程数据。对于混合配置中的读取操作，通常直接从本地 SSD 缓存读取在本地的数据。如果数据不在本地，系统将从远程节点上的 SSD 缓存检索数据。通过此过程，平台能够在进行读取操作时利用所有 SSD，从而减少性能瓶颈，并提供出色的性能。

当使用 VMware 动态资源调度 (DRS) 等工具将虚拟机移动至新位置时，思科 HX 数据平台不需要移动数据。此方法可显著减少在系统间移动虚拟机的影响和成本。

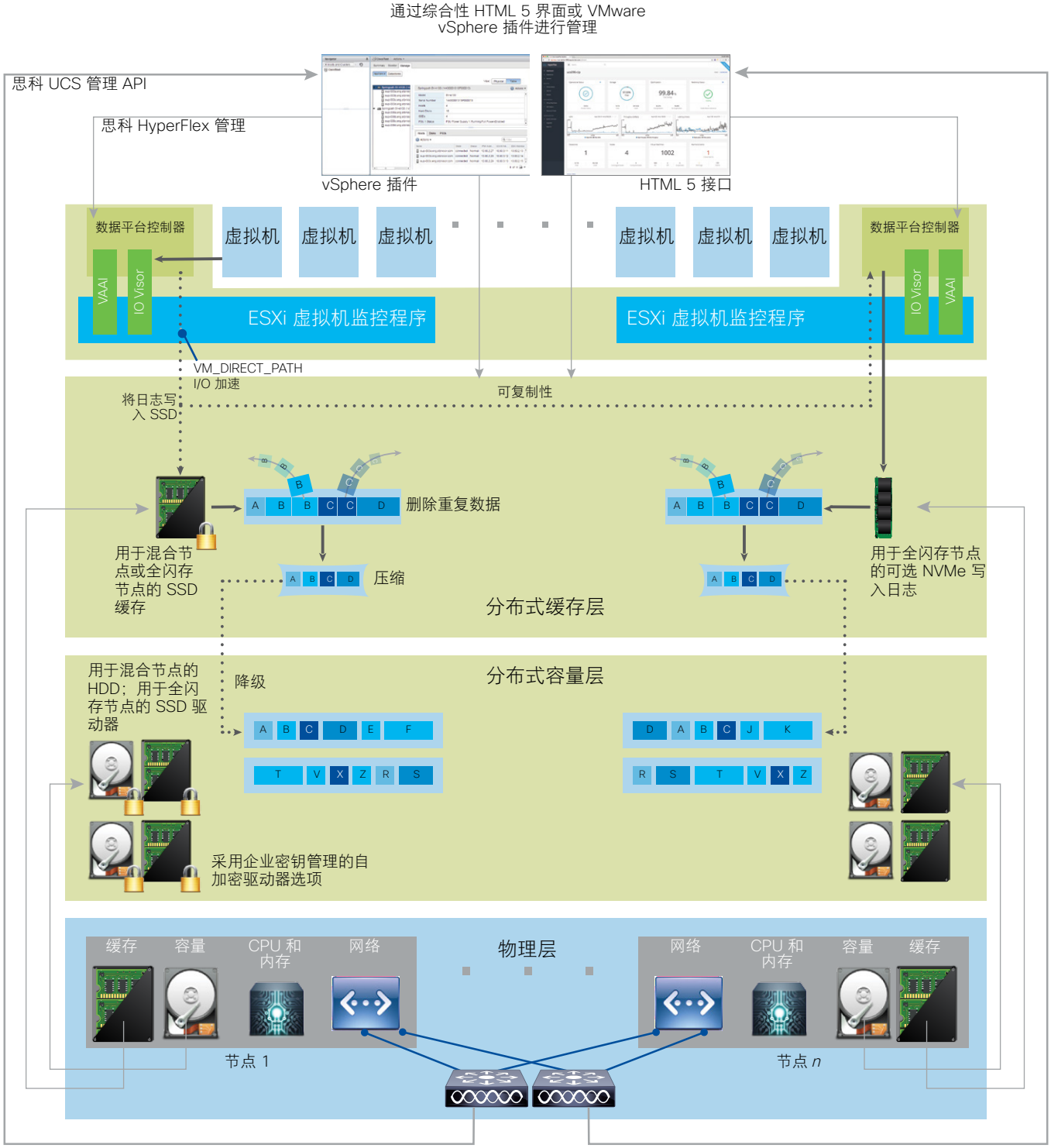


图 4：通过思科 HyperFlex HX 数据平台执行的数据流写入操作

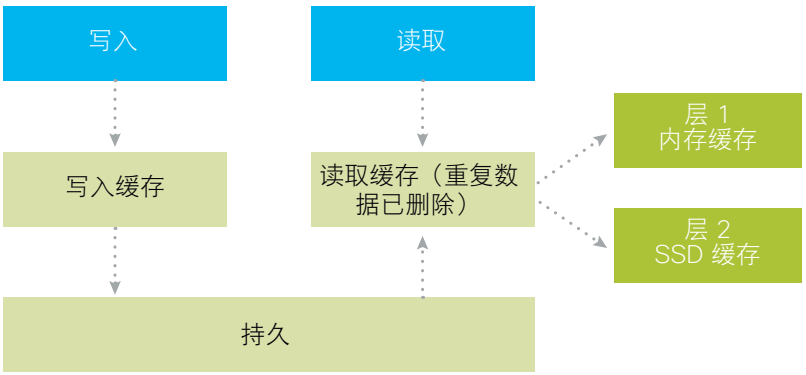


图 5：分离的数据缓存层与数据持久层

数据操作

该数据平台实施分布式结构化日志文件系统，此系统根据节点配置更改缓存和存储容量的处理方式。

- 在混合配置中，该数据平台使用 SSD 中的缓存层加速读取请求和写入响应，在 HDD 中实施容量层。
- 在全闪存内存配置中，该数据平台使用 SSD 缓存层加速写入响应，通过 SSD 实施容量层。通过从容量层的 SSD 获取的数据直接履行读取请求。不需要专用的读缓存来加速读取操作。

在以上两种配置中，传入数据可在满足可用性要求所需数量的节点之间进行条带化分布：通常需要两个或三个节点。根据所设置的策略，在传入的写入操作数据复制到集群中其他节点的 SSD 后，系统会将其确认为持久性数据。此方法能降低因 SSD 或节点故障丢失数据的可能性。然后，写入操作将转出存储至全闪存配置下容量层的 SSD 或混合配置下低廉的高密度 HDD 进行长期存储。您可以选择只使用 SSD 来提高性能，增加密度并降低延迟，也可以同时使用高性能的 SSD 和低成本高容量的 HDD 来优化数据的存储成本。

结构化日志文件系统会不断整合要写入到缓存的数据块，直到可配置大小的写入日志容量已满或工作负载条件指示系统将块转出存储至 SSD 或旋转磁盘。当现有数据（逻辑上）被覆盖时，结构化日志方法只需附加一个新块并更新相关元数据。数据转出存储至 HDD 时，写入操作只需单次寻道操作，即可写入大量序列数据。传统的读取-修改-写入模式需要在 HDD 上进行大量的寻道操作，且一次只能写入少量数据，与其相比，此方法可显著提高性能。此布局还有益于寻道操作不耗时的 SSD 配置。此布局会对数据执行传入写入操作和随机覆盖操作，从而降低 SSD 的写入放大级别以及闪存介质经历的写入总数。

当各节点的数据转出存储至磁盘时，系统会对数据执行重复数据删除和压缩操作。由于此过程发生在系统确认写入操作之后，因此这些操作不会造成性能下降。当重复数据删除块的大小比较小时，有助于提高重复数据删除率。通过数据压缩可进一步减少数据占用空间。此后，在写入缓存段获得释放可供重用时，系统会将数据移至 SSD 或 HDD 存储（图 4）。

热数据集（持久层中频繁读取或最近读取的数据）缓存在内存中。在混合配置中，热数据集也缓存在 SSD 中（图 5）。在将 HDD 用于持久存储的配置中，把频繁使用的数据存储在缓存层中有助于提高工作负载的性能。例如，当应用和虚拟机修改数据时，数据可能会从缓存中读取，因此，通常不需要读取和展开旋转磁盘上的数据。因为 HX 数据平台将缓存层从持久层分离，所以您可以单独扩展 I/O 性能和存储容量。但是，全闪存配置不使用读缓存，因为数据缓存不提供任何性能优势；持久数据副本已位于高性能 SSD 中。在这些配置中，使用 SSD 实施的读缓存可能会成为瓶颈并妨碍系统为整组 SSD 使用汇聚带宽。

数据优化

HX 数据平台提供细致详尽的线内重复数据删除和可变块线内压缩，这些功能始终为缓存（SSD 和内存）层和容量（SSD 或 HDD）层中的对象开启。与其他解决方案不同，其他解决方案需要您关闭这些功能来维持性能，思科数据平台的重复数据删除和压缩功能旨在维持和增强性能并显著降低物理存储容量要求。

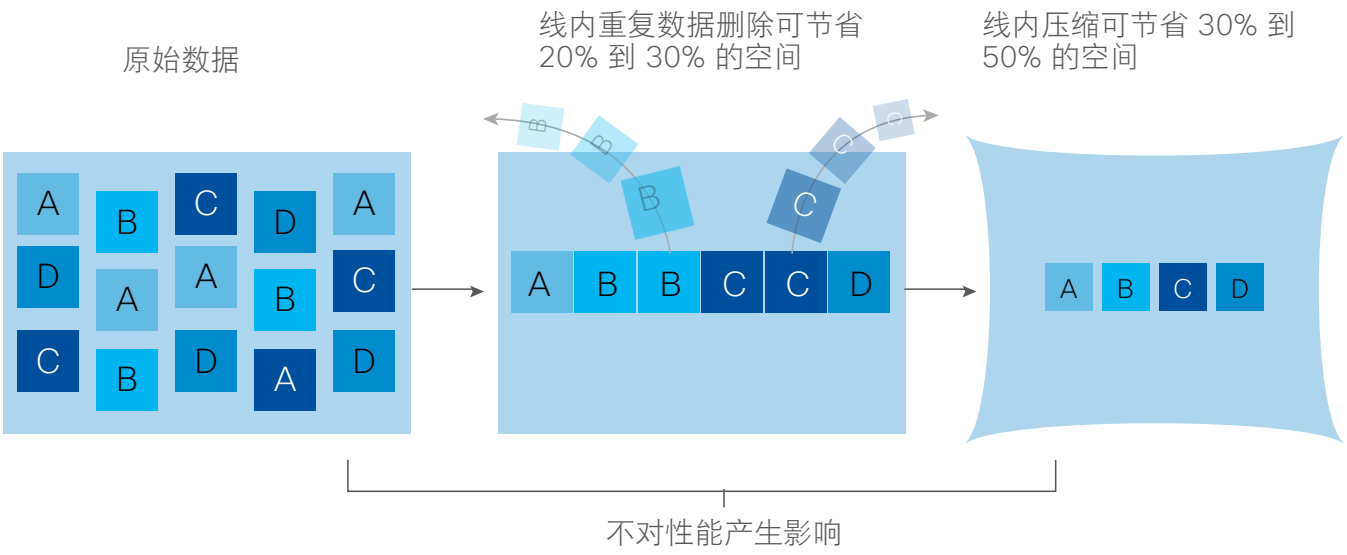


图 6：思科 HyperFlex HX 数据平台优化数据存储，并且不对性能产生影响

重复数据删除

重复数据删除用于集群中的所有存储，包括内存、SSD 和 HDD。通过使用正在申请专利的 Top-K 多数算法，该平台应用了实证研究得出的结论，实证研究显示，大多数数据在分切成小数据块时，有基于少数数据块的显著重复数据删除潜力。通过指纹识别和索引这些频繁使用的数据块，仅用少量内存即可实现高重复数据删除率，这是集群节点中的高价值资源（图 6）。不仅会删除持久层中的重复数据来节约空间；当数据读取至混合配置中的缓存层时，仍可删除重复数据。该方法允许在缓存层中存储较大的工作集，从而为使用速度较慢的 HDD 的配置加速读取性能。

线内压缩

HX 数据平台对数据集使用高性能线内压缩来节省存储容量。虽然其他产品可以提供压缩功能，但许多产品的压缩功能会对性能造成不利影响。在与之相比，思科数据平台使用 CPU-卸载指令来减少压缩操作对性能的影响。此外，结构化日志分布式对象层不会对之前压缩数据的修改（写入操作）产生任何影响。相反，传入

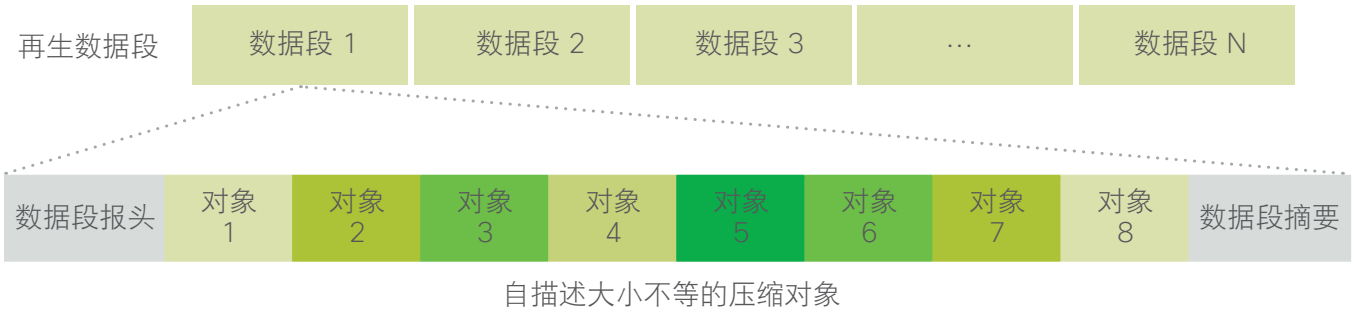


图 7：思科 HyperFlex HX 数据平台的结构化日志文件系统数据布局

修改被压缩并写入至新位置，现有（旧）数据标记为删除，除非需要在快照中保留该数据。注意，不需要在写入操作前读取正在修改的数据。该功能可避免典型读取-修改-写入的不利后果，显著提高写入性能。

结构化日志分布式对象

在 HX 数据平台中，结构化日志分布式对象存储层对通过重复数据删除引擎筛选的数据进行分组并压缩至可自寻址的对象。系统将这些对象以结构化日志、顺序方式写入磁盘。所有传入 I/O（包括随机 I/O）按顺序写入缓存（SSD 和内存）层和持久 (SSD 或 HDD) 层。该对象分布到集群中的所有节点以统一使用存储容量。

通过使用顺序布局，该平台帮助提高闪存耐用性并充分利用 HDD 的读写性能特性，HDD 的读写性能特性非常适合顺序 I/O 操作。由于没有使用读取-修改-写入操作，所以压缩、快照和克隆操作对整体性能的影响很小或根本没有影响。

数据块压缩至对象，并以固定大小的数据段按顺序安置，然后数据段再以结构化日志的方式按顺序安置（图 7）。结构化日志数据段中的每个压缩对象都可使用密钥进行唯一寻址，每个密钥用一个校验和进行指纹识别和存储以提供高水平的数据完整性。此外，按时间顺序写入对象有助于平台从媒介或节点故障中快速恢复，只需要重新写入系统发生故障后而没有写入的数据。

加密

通过使用可选自加密驱动器 (SED)，HX 数据平台可以加密数据平台上的缓存层和持久层。通过与企业密钥管理软件或口令保护的密钥集成，静态数据加密可帮助您符合健康保险转移与责任法案 (HIPAA)、支付卡行业数据安全标准 (PCI-DSS)、联邦信息安全管理法案 (FISMA) 和萨班斯法案的要求。该平台本身已针对联邦信息处理标准 (FIPS) 140-1 进行强化，采用密钥管理的加密驱动器符合 FIPS 140-2 标准。

数据服务

HX 数据平台提供可扩展的节省空间型数据实施，包括精简配置、空间回收、基于指针的快照以及克隆，且不影响性能。

精简配置

该平台通过消除预测、购买和安装可能会长时间闲置的磁盘容量的需求来有效利用存储。虚拟数据容器可为应用表示任何数量的逻辑空间，而所需的物理存储空间数量由写入的数据决定。因此，您可以在现有节点上扩展存储，也可以根据业务需求通过增添更多存储密集型节点来扩展集群，从而消除在您需要大量存储之前购买存储的需求。

快照

HX 数据平台使用基于元数据的零复制快照改善备份操作和远程复制，这是要求不间断数据可用性企业所需的关键功能。空间有效的快照使您能够进行频繁的数据在线备份，而无需担心消耗物理存储容量。数据可以离线移动或从这些快照中立即恢复。

- 快速快照更新：当快照中包含修改数据时，修改数据会写入到新位置，而元数据会得到更新，无需读取-修改-写入操作。
- 快速快照删除：您可以快速删除快照。该平台仅删除位于 SSD 的少量元数据，而不需要使用增量磁盘技术解决方案所必须的长时间整合过程。
- 高度有针对性快照：使用 HX 数据平台，您可以以单个文件为基础产生快照。在虚拟环境中，这些文件映射到虚拟机中的驱动器上。这些灵活的特异性使您能够在不同的虚拟机上应用不同的快照策略。

节省空间的快速克隆

在 HX 数据平台中，克隆是可写入快照，用于快速调配项目（如用于测试和开发环境的虚拟桌面和应用）。这些节省空间的快速克隆可迅速复制存储卷，因此只需进行元数据操作就可以复制虚拟机，实际数据复制只在进行写入操作时进行。使用这种方法，可在几分钟内创建和删除数以百计的克隆。与全复制方法相比，该方法可节约大量时间，增加 IT 灵活性，并提高 IT 人员的工作效率。

系统在创建克隆时删除重复数据。当不同克隆版本开始出现不同时，它们中的共同数据会被共享，只有独特的数据会占用新存储空间。重复数据删除引擎消除已分离克隆中的数据副本，以进一步减少克隆的存储涉及面。因此，您可以配置大量应用环境，而无需担心存储容量的使用。

数据复制和可用性

在 HX 数据平台中，结构化日志分布式对象层复制传入数据，进而提高数据可用性。根据您设置的策略，写入至写入缓存的数据可在应用确认写入操作前同步复制到位于不同节点的一个或多个 SSD 中。此方法在保护数据免受 SSD 或节点故障影响的同时，快速确认传入写入操作。如果出现 SSD 或节点故障，系统将使用该数据的可用副本在其他 SSD 或节点上快速重新创建副本

结构化日志分布式对象层同时复制从写入层移至容量层的数据。复制的数据同样免受 SSD、HDD 或节点故障影响。在存在两个副本或总共三个数据副本的情况下，集群可经受两个 SSD、两个 HDD 或两个节点发生无关的故障，而无数据损失的风险。无关的故障是指不同物理节点上发生的故障。同一个节点上发生的故障会影响同一数据副本，

视为一个故障。例如，如果节点上的一个磁盘故障导致同一节点上的另一磁盘也发生故障，那么这种关联故障在系统中算作一个故障。这种情况下，该集群可以承受其他节点上发生另外一个不关联的故障。请参阅 Cisco HyperFlex HX 数据平台系统管理员指南了解容错配置和设置的完整列表。如果思科 HyperFlex 控制器软件出现问题，位于该节点的应用的数据请求将自动路由到该集群的其他控制器。该功能还可用于在滚动的基础上升级或维护控制器软件，而不影响集群或数据的可用性。该自愈功能是 HX 数据平台非常适合生产应用的原因之一。

此外，本地复制可将一致的集群数据传输到本地或远程集群。借助本地复制，您可以在本地或远程环境中创建快照和存储您环境中的时间点副本以用于备份和灾难恢复目的。

数据再平衡

分布式文件系统需要稳健的数据再平衡功能。在 HX 数据平台中，元数据访问不会产生任何开销，且再平衡极其有效。再平衡是一种出现于缓存层和持久层的非破坏性在线过程，数据移动时保持良好水平的特异性，以提高存储容量的利用率。增添或移除节点和驱动器或当节点和驱动器出现故障时，该平台会自动再平衡现有数据。在集群中增添新节点时，其容量和性能可用于新数据和现有数据。再平衡引擎将现有数据分布至新节点，并帮助确保从容量和性能角度来看，集群中的所有节点均统一使用。如果节点出现故障或被从集群中移除，再平衡引擎重建并将故障或已移除节点中的数据副本分布至集群中的可用节点。

在线升级

思科 HyperFlex HX 系列节点和 HX 数据平台支持在线升级，便于您在不中断业务的情况下扩展和更新环境。您可以轻松扩展物理资源；添加处理功能；下载和安装 BIOS、驱动器、虚拟机监控程序、固件和思科 UCS 管理器更新、增强和漏洞修复。

结论

思科 HyperFlex HX 数据平台为支持新 IT 消费模式的超融合基础设施部署革新了数据存储。该平台的架构和软件定义的存储方法为您提供针对特定用途的高性能分布式文件系统，提供一系列企业级数据管理服务。通过重新定义分布式存储技术这一创新举措，数据平台为您提供实现自适应 IT 基础设施的超融合基础设施。

如需思科 HyperFlex 系统的更多信息，请访问
http://www.cisco.com/go/hyperflex

采用 Intel® Xeon® 处理器的思科 HyperFlexTM 系统