

Hammerspace Tier 0 Affinitization

TECH BRIEF

Introduction

Building on the improvements Hammerspace introduced into the Linux kernel in 2024 to further improve Tier 0 performance, the new Hammerspace v5.2 release contains a related enhancement called Tier 0 affinitization that adds locality-awareness to Tier 0 installations.

With Tier 0 affinitization, Hammerspace increases the flexibility of the Tier 0 architecture and reduces the amount of manual configuration required, making Tier 0 even easier to implement. This new functionality is automatic, transparent, and enabled by default.

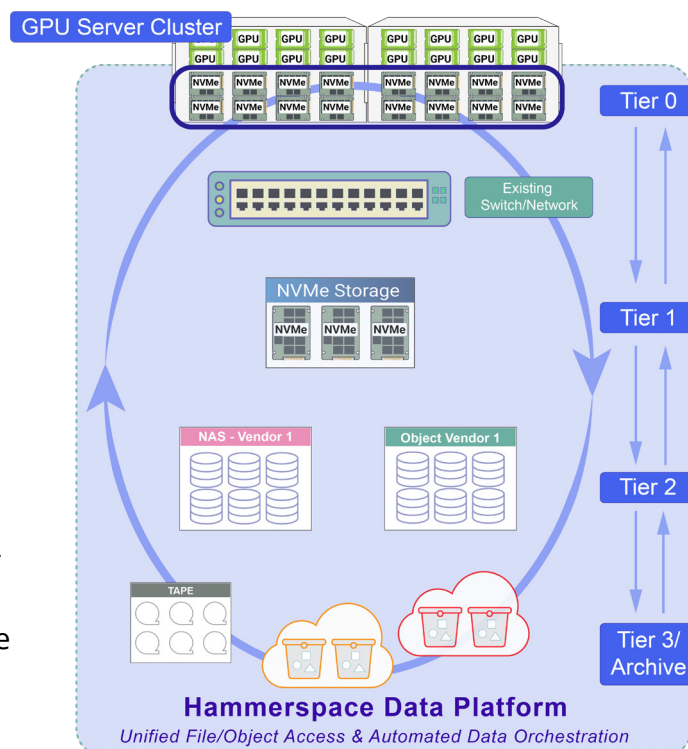
After providing background context, this Tech Brief will explain how Tier 0 affinitization works and its benefits. All of the functionality discussed below is based on NFS standards and does not require loading proprietary software or kernel patches on the compute servers.

Background

Hammerspace Tier 0

Compute servers that are clustered for AI and HPC workloads each typically contain a few NVMe SSDs. This storage is the fastest and lowest latency available, residing on the same PCIe bus with the CPUs and GPUs. But because this capacity is broken into many small silos across the servers, it is largely unused. This is a missed opportunity.

Hammerspace Tier 0 activates these formerly isolated islands of storage, bringing them together into a pNFS-mountable shared file system usable by the entire cluster. When combined with additional tiers of storage and Hammerspace data orchestration, Tier 0 becomes a standards-based high-performance foundation for data-intensive workflows.



Parallel NFS (pNFS)

Like other high-performance file systems, pNFS separates metadata communication from the data path. To perform I/O, a pNFS client first contacts the metadata server (known as an Anvil in Hammerspace) which provides the client with a layout. The layout includes a list of storage targets the client may use to perform the I/O, and the credentials needed to do so. The client then goes directly to the storage to perform I/O – the Anvil metadata server is not in the data path.

Today, per the pNFS standard, the client will use the first storage target in the layout, only moving down the list if the first target is unreachable or if (for writes) the client needs to create additional mirrors of the data. Hammerspace automatically manipulates the order storage targets appear in layouts to spread the load and direct read and write activity as needed to fulfill configured policy objectives.

Tier 0 Affinitization

Previous Behavior

Tier 0 provides the best performance when the compute node requesting I/O can use its local storage volume instead of reaching across the network to read from or write to another node. This requires the Anvil to know that the compute node requesting the layout also hosts a local Tier 0 storage volume, so the Anvil can put that volume at the top of the list in the layout.

If the Anvil can't tell when a pNFS client is also host to a Tier 0 volume, the read or write will be directed to the client-local volume in only $1/n$ cases on average, where n is the number of Tier 0 nodes in the cluster. In a five-node cluster, then, the I/O will be local only 1/5th of the time, and the proportion of local I/O decreases with increasing cluster size. This is clearly not ideal.

Previously, making the Anvil locality-aware to optimize Tier 0 performance required configuring multiple directories in the file system and setting specific Hammerspace policy objectives on those directories to direct data flow.

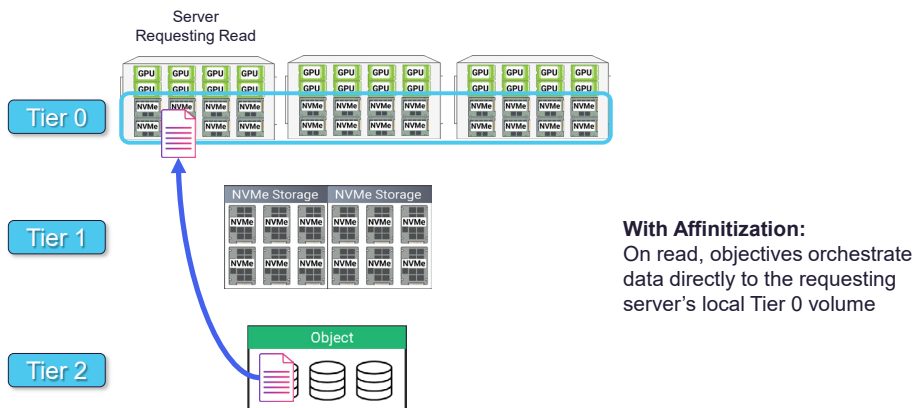
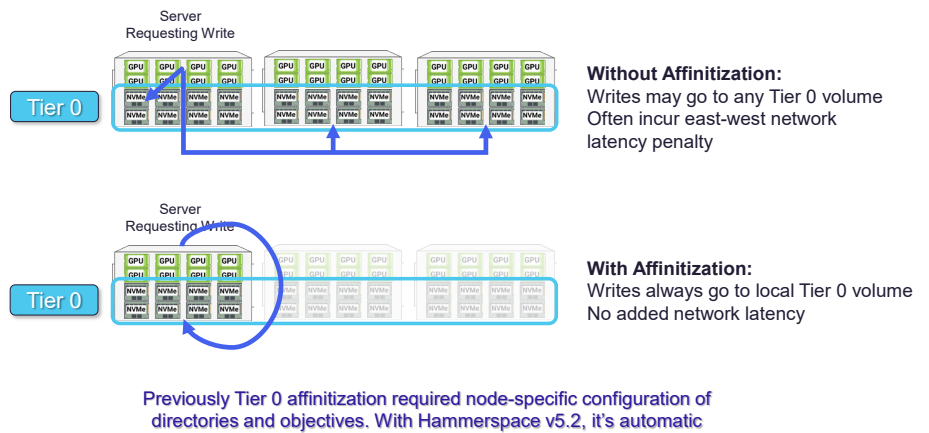
New Affinitization

Affinitization eliminates these configuration steps. Now, starting with Hammerspace v5.2, the Hammerspace Anvil automatically detects when a client requesting I/O has a local Tier 0 storage volume. The Anvil can then put the local volume at the top of the list in pNFS layouts requested by that client, ensuring that as much I/O as possible is kept local. This maximizes the performance of Tier 0 without requiring any extra configuration or management.



Remember that Tier 0 is only one tier of storage in a multi-tier environment. There is usually a Tier 1 consisting of network-attached NVMe flash, and often a colder Tier 2 or Tier 3 on object storage or cloud storage as well.

Hammerspace's highly configurable policy objectives are used to define and direct data placement and data protection behavior. This can include objectives that place instances of files on different tiers based on usage.



For example: when Tier 0 is configured and a file is accessed that resides on Tier 1, 2, or 3, objectives cause an instance of that file to be placed on Tier 0. Affinitization now adds intelligence to the I/O, ensuring that the file not only lands on one of the Tier 0 storage volumes, but is placed into the local volume on the exact compute node that needs it. This automatically ensures the highest performance access to the data.

Conclusion

With Hammerspace v5.2, Tier 0 is even better than before. The new affinitization capability is automatic and transparent, making it even easier to take advantage of the performance and efficiency benefits of Tier 0.

To learn more about Hammerspace Tier 0, visit <https://hammerspace.com/tier-0>.

To learn more about pNFS, visit the Hammerspace pNFS Resource Center: <https://hammerspace.com/parallel-nfs>.

