

Addressing the Uncontrolled Proliferation of Unstructured Data with InfiniBox® and Hammerspace

Over the last year, Gartner end users report 30% to 60% growth for their file data, and by 2026 large enterprises will triple their unstructured data capacity across their on-premises, edge, and public cloud locations, compared to 2023.¹ In addition, by 2028 Gartner estimates that 70% of file and object data will be deployed on a consolidated, unstructured data storage platform, up from 35% in early 2023. Most importantly, Gartner predicts that all storage products will include cyber storage capabilities focused on active defense from cyber events beyond recovery, up from 10% in early 2023.²

As unstructured data continues to proliferate at unprecedented rates through organizations, the need to manage that data, ensure its security, and provide seamless access to it regardless of its location has become even more important. InfiniBox provides a multi-petabyte scale platform for file and block workloads with guaranteed 100% availability, guaranteed performance, and InfiniSafe® cyber recoverability guarantees. Native InfiniBox NAS capabilities include simultaneous dual protocol access via NFS and SMB protocols, filesystem snapshots, compression, replication, WORM, and billion-file scalability for large unstructured data sets. However, customers invariably have unstructured data across other storage types, multiple data centers, and in the public cloud. To help support our customers' various file access requirements, we have partnered with Hammerspace to provide sophisticated file indexing and automated data orchestration capabilities. This provides customers with a global file system (GFS) and data management system that bridges heterogeneous on-premises and cloud storage platforms to better manage their distributed data.

A global file system (GFS) presents unified file access with standard protocols via a single global namespace that spans multiple storage silos and locations (on-prem, cloud, or edge). This means that users or applications mounting the file system at any of the locations will see the same file and folder structure as clients in any other location. The benefit of using a GFS is that all users everywhere are sharing the same file metadata, rather than wrangling forked file copies across multiple locations. Not only does this reduce copy sprawl and the resulting complexity for users and IT administrators, a GFS also helps increase storage utilization and enables far greater efficiency with the amount of data that is transferred between silos and sites.

A GFS works by keeping a file's metadata, such as its creation date, size, name, and location, separate from the main payload of the file content (sometimes referred to as the essence data). The metadata is continuously synced between locations providing a universal access layer, while the payload is only transferred when required as a background operation transparent to users and applications. Since the metadata for even large files takes up only a couple of kilobytes, using this system in production ends up requiring less data transfer than a full syncing system based upon file copies.

The GFS abstracts these concepts away from the user. From an application's point of view, the file system behaves the same as a local file system. The result is that the workflows built on top of local file systems will not need any modifications to work with a GFS.

This cross-platform, multi-site global namespace enables customers to manage distributed file storage and seamlessly presents access to file data regardless of its physical location. The factors driving the need for this global file system are:

- ▶ Data center consolidation and inevitable IT tech refresh
- ▶ The shift to more agile application environments
- ▶ A shifting landscape including hybrid environments of one or more on-premises data centers, plus multiple cloud service providers and regions

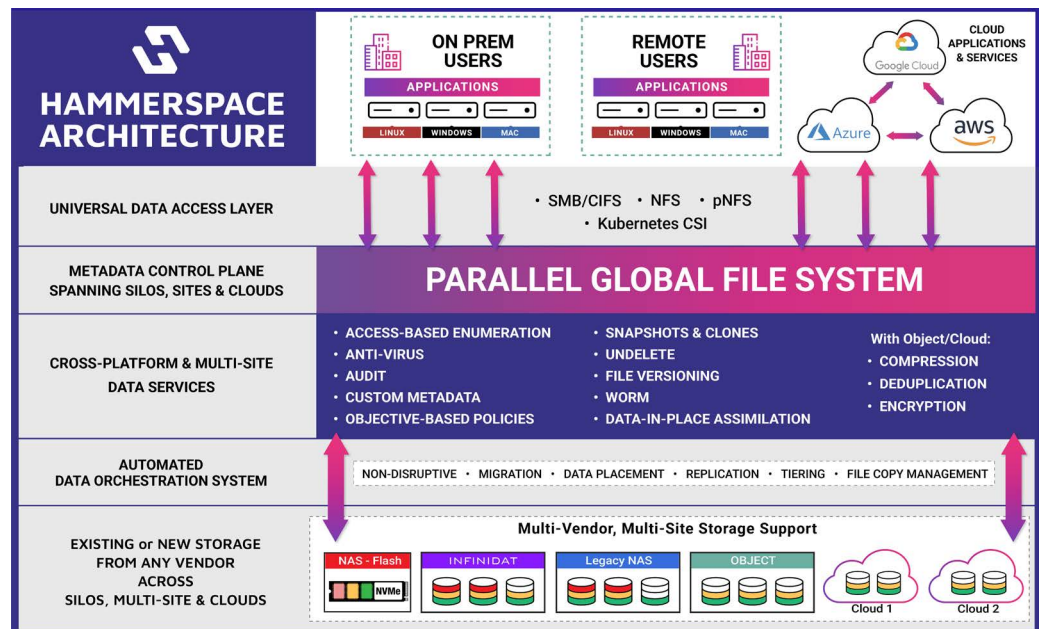
“Infinidat Federal has partnered with Hammerspace to provide sophisticated file indexing and automated data orchestration capabilities. This provides customers with a global file system (GFS) and data management system that bridges heterogeneous on-premises and cloud storage platforms to better manage their distributed data.”

- ▶ Operational silos must be reduced to simplify IT management tasks, which enables better collaboration across interdepartmental teams while also enforcing cross-platform data services
- ▶ New use cases must be rapidly supported for AI/ML and other innovative data-intensive workflows

Hammerspace

Hammerspace is a software-defined data orchestration system that provides unified file access via a high-performance Parallel Global File System which can span different storage types from any vendor, as well as across geographic locations, public and private clouds, and cloud regions. Designed to make data a global resource across distributed or otherwise incompatible storage platforms from any vendor, Hammerspace presents a cross-platform global namespace where users and applications can have direct multi-protocol access to all files on any storage anywhere.

Hammerspace has a full suite of data services and automated orchestration services designed to protect data, and manage its placement and replication based on policies.



The Solution

Unlike traditional “unified” storage, which is an open source file system bolted onto a block array, InfiniBox was built from the ground up to deliver petabyte-scale NAS and SAN in parallel. The InfiniBox File System (IFS) is a first-level peer to the block service and is a clean-sheet design that employs a stateless, functional programming paradigm to deliver petabyte-scale NAS services that couple directly with Neural Cache. Along with conventional NAS workloads, InfiniBox excels at extreme use cases, such as trillions of objects in a file system and millions of files per directory, which are simply not possible with traditional NAS filers supporting SMB and NFS.

With Hammerspace, users and applications throughout the enterprise interface with InfiniBox, third party storage, or cloud repositories via a common unified file directory, regardless of where they reside. As a standards-based platform, this greatly simplifies the administration of data and enables customers seamless access from anywhere in the world. No proprietary client software is needed, since users and applications are accessing files via the same standard SMB and/or NFS file protocols they are used to. Further, with the suite of data services (the blue section in above diagram), Hammerspace enhances the native data protection offered with InfiniBox and extends that protection across the rest of the unstructured data not sitting on the InfiniBox.

The platform’s automated data orchestration capabilities enable customers to create policies to manage the lifecycle of their unstructured data and place it where it is best suited. This data orchestration capability can be automated with objective-based policies, triggered by any combination of metadata including file system attributes and custom metadata tags. For example, policies can be used to transparently migrate data from a retiring platform to tier 1 storage on InfiniBox, or to move data to a deep archive in the cloud, as needed. These policies can be set at a file-granular level, and are completely non-disruptive to user access, since with Hammerspace all such data placement actions now become a background operation, even on live files. Users simply see their files in the same expected shares and folder structure via the unified metadata layer, regardless of where the files are today or may move to tomorrow. Hammerspace also provides antivirus capabilities, using standard Internet Content Adaptation Protocol (ICAP) to implement virus scanning in transparent HTTP proxy caches.

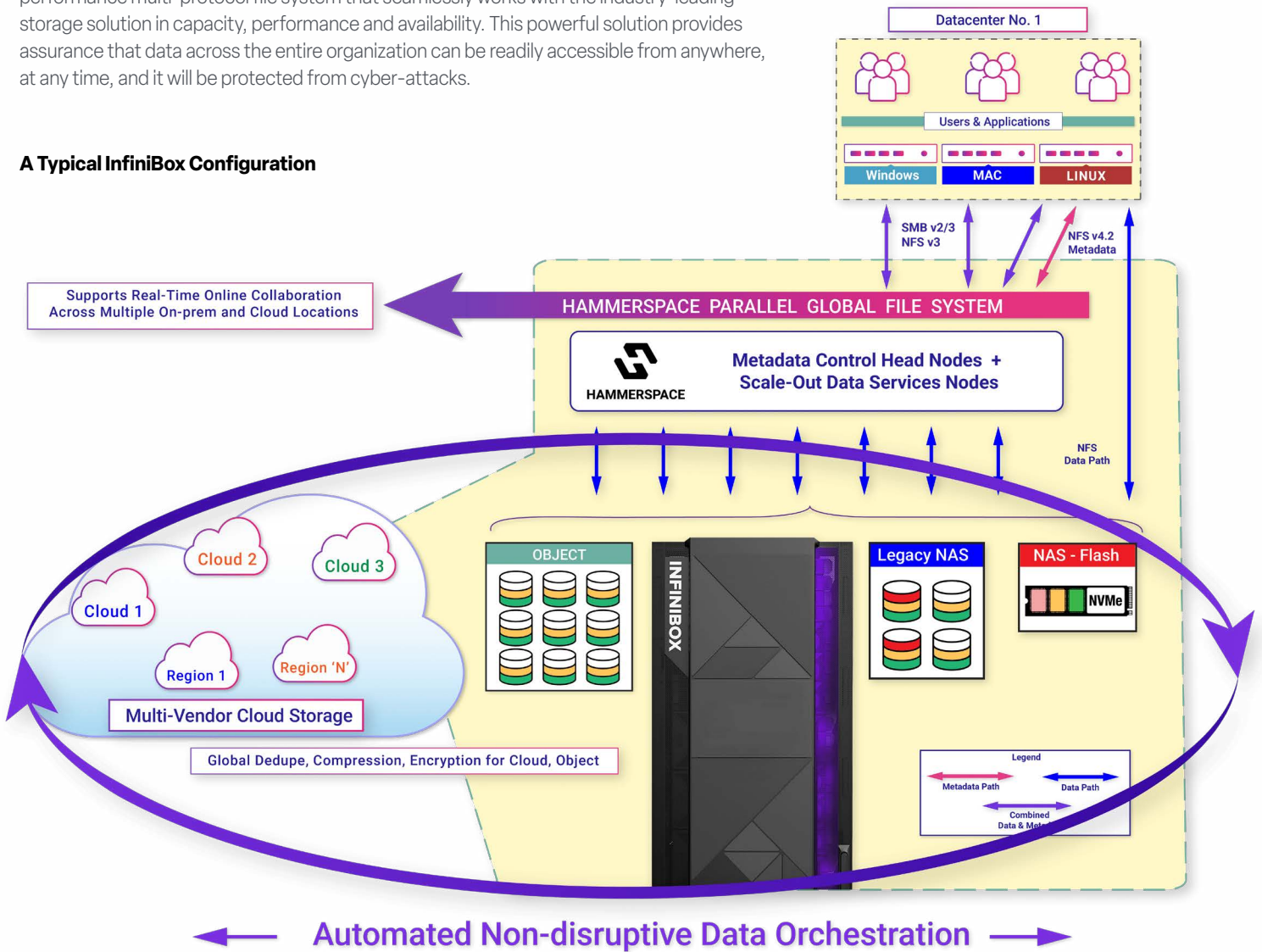
With the Hammerspace solution, InfiniBox can be installed in existing infrastructures and immediately participate in the global file system. Customers can choose to migrate sensitive data to InfiniBox to take advantage of the cyber resilience features of InfiniSafe, including immutable snapshots with recovery within one minute, air-gapping, and malware scanning with InfiniSafe Cyber Detection. As data is aged and access is not as critical, policies can move it to other archive tiers for long term storage.

Hammerspace scales to meet the volume needs of the customer environment. Deployed as a fully-integrated software solution on bare-metal servers, VMs, or cloud machine instances, Hammerspace software includes Anvil servers for metadata management, and DSX data services nodes,

which handle all I/O and data actions. There is no one-size-fits-all specification for the server requirements for Anvil or DSX nodes, which means the system can be tuned to the specific load requirements of the customer's use cases. This enables the system to be dialed-in to minimize unnecessary infrastructure expenses, and also to dynamically scale-out when needed to saturate a high-performing infrastructure without disrupting user or application access.

The partnership between Infinidat Federal and Hammerspace offers customers a high-performance multi-protocol file system that seamlessly works with the industry-leading storage solution in capacity, performance and availability. This powerful solution provides assurance that data across the entire organization can be readily accessible from anywhere, at any time, and it will be protected from cyber-attacks.

A Typical InfiniBox Configuration



If you are struggling with uncontrolled sprawl of unstructured data or need storage capabilities with new data access for data analytics, we've got your solution! Contact your Infinidat Federal sales team for this comprehensive solution to securely manage your data, provide cyber resilience, guarantee its availability and performance, and place it where it makes the most sense. Not a current customer? We are easy to reach at sales@infinidatfederal.com.

¹ Modernize Your File Storage and Data Services for the Hybrid Cloud Future, 28 February 2023

² Magic Quadrant for Distributed File Systems and Object Storage, 1 November 2023