panzura

Panzura White Paper

Panzura Freedom NAS Filer: Technology in Detail

The Panzura Freedom Family, powered by the industry's first purpose built global cloud filesystem, called the Panzura CloudFS is a next generation multi-cloud NAS filer. The Freedom Filer provides today's enterprise with a cluster based solution that spans data centers, office sites and compute clouds enabling local, hybrid and in-cloud data workflows for NFS, SMB and mobile clients. Panzura's software defined storage solution provides local performance with the economics, scalability, and durability of the cloud.

Organizations can consolidate their unstructured data in the cloud, eliminating islands of storage across sites and legacy filers. Panzura's Freedom NAS filers provide access to data locally at high speeds without any modifications to existing applications or clients. The Panzura Freedom filers make deploying cloud storage and a global file system easy and transparent to users.

Executive Summary

Today, enterprise IT executives struggle with storage growth, storage capacity balancing, and data mobility. Traditional technologies such as tape, SAN, and NAS were more than sufficient to meet the needs of the past but an increasingly distributed workforce generating massive amounts of data from multiple platforms forced enterprises to consider new storage paradigms, in particular cloud storage. But adopting the cloud as a storage tier can be terribly problematic. Integrating with existing IT environments, ensuring data security, and managing data across sites plus the cloud is not a trivial exercise.

Panzura created a solution based on a global filesystem and unified namespace that makes adding and using cloud storage seamless and secure while enabling global file sharing, cloud-integrated NAS, and data protection, including archiving, disaster recovery, and backup. Panzura Freedom Filers™ deploy quickly and easily without changes to existing infrastructures or applications, all while securing data with snapshots and military-grade encryption. Panzura Freedom Filers make cloud storage a seamless, viable storage tier for enterprises of all sizes.

Unstructured file storage is the fastest growing category of data. (Figure 1). This is due to the explosive growth of machine generated data such as application log files, machine learning output, IoT telemetry or sensor data, 4K video and 3D imaging. File-based NAS storage has gained prominence, both due to this explosive growth in unstructured data creation and to its simplicity, ease-of-use, and highly integrated suite of capabilities (e.g. application and user-accessible data recovery via snapshots). Network attached file storage (NAS) systems facilitate shared access to content among multiple client computers and application servers running Network File System (NFS) and Server Message Block (SMB) protocols. Today's increasingly mobile workforce demand access to that same unstructured data on their personal devices like IOS and Android phones and tablets, web browser as well as Windows and Mac applications disconnected from the corporate LAN.

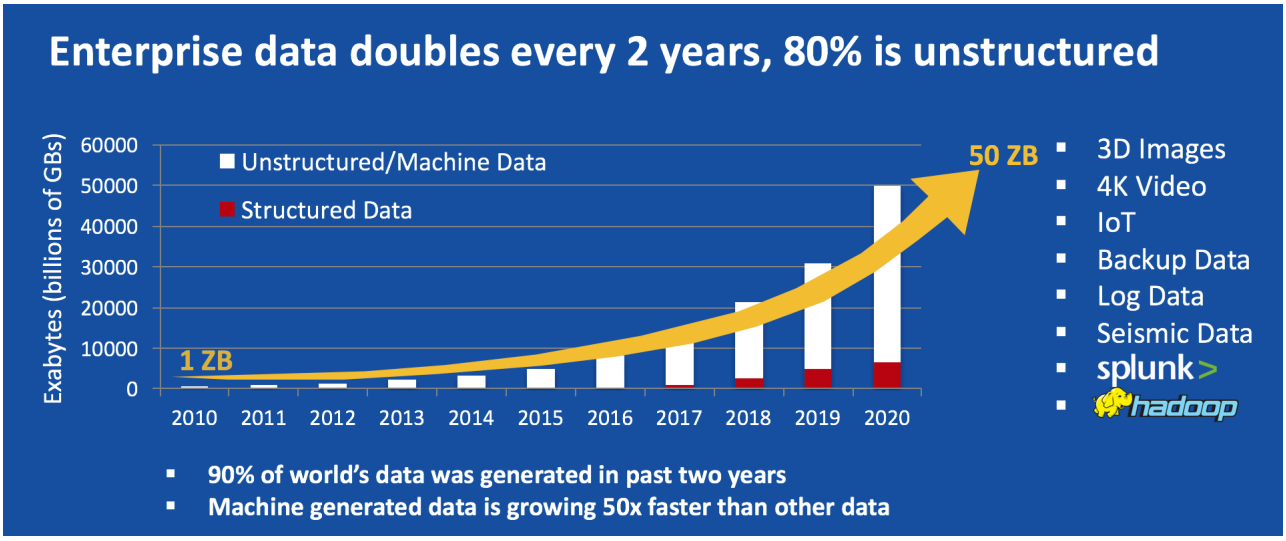


Figure 1: Unstructured data growth projections

However, even with these advantages, the file storage needs of the modern, globally-distributed enterprise are not being adequately served. Despite the plethora of Internet-based communications and productivity tools that have arisen to help remote teams remain effective, sharing files on an enterprise-wide basis remains very challenging.

This white paper examines the Panzura Freedom Filer, and in particular the Panzura CloudFS filesystem, in detail and explores key areas critical to system design including performance, scalability, data integrity, and security. But perhaps most

importantly, it explains in detail what distinguishes the Panzura Freedom Filer from other file storage solutions; the ability to leverage the Internet to provide ***global multi-cloud*** integrated NAS to a decentralized enterprise and the capability to leverage a centralized storage pool in the form of private or public cloud storage to simplify capacity growth, management, and data protection. The intended audience for this paper is technically knowledgeable about data storage and filesystems

> *"As unstructured data and metadata continue to grow upwards of 60% to 80% on average per year, enterprise IT managers face an unwinnable struggle to maintain or reduce costs while ensuring user needs (i.e., SLAs) are met."*

Ongoing Storage Challenges

As unstructured data and metadata continue to grow upwards of 60% to 80% on average per year, enterprise IT managers face an unwinnable struggle to maintain or reduce costs while ensuring user and application needs (i.e., SLAs) are met. Traditional means of storing data, primarily local NAS, and protecting data, primarily tape or disk-to-disk solutions, do not scale quickly or economically enough to accommodate the growth in demand from massive expansion in data generated by individuals, applications and the digital transformation of most organizational workflows. In addition, constantly connected individuals demand that data be available anytime, from anywhere, without reduced performance or information lag. Limited budgets, limited technology alternatives, strict user and application demands have put IT on an unsustainable trajectory that must be addressed without negatively impacting performance.

Data storage challenge The primary technology for storing unstructured enterprise data today is high-performance, high-cost network-attached storage (NAS). With this technology, the standard method for addressing data growth is to add additional expensive filers containing spindles or SSD based flash arrays. In addition to the high hardware costs, more disks/SSDs mean more datacenter space, power and cooling requirements, and added off site capacity for replication. Because capacity expansion takes time, IT managers must try to forecast storage needs and forward provision to accommodate these potentially incorrect forecasts. Unanticipated spikes in storage demand send IT managers scrambling for added capacity and overly aggressive forecasts result in investment in idle capacity. As enterprises adopt newer cloud based technologies like Dropbox and Box to meet the increasingly mobile workforce, these services lead to even more disparate islands of storage to be managed. Additionally as different departments independently launch projects into the cloud for development, test and simulation workflows the enterprise is evolving into an ever increasingly complex multi-cloud vendor ecosystem.

When multiple sites are taken into consideration, using standard NAS can often result in over-provisioning at some sites and under-provisioning at others, as well as in storage "islands" where data at one site is not visible to users and applications at other sites. The only way for users at one site to view files created or edited at another site is to save copies of files stored off-site to their own location, resulting data sprawl or duplication of files and commensurate overspending on expensive storage, not to mention significant version control challenges. This problem is compounded when backup and archiving also occurs locally, since duplicate copies on islands of storage means greater storage needed for data protection. As an alternative to customer-owned NAS, some vendors offer what they call Cloud NAS. Almost all of these solutions suffer from limitations in performance, scale, or both, and none are yet enterprise-class.

What about the nature of the data itself? On-site NAS today supports both structured and unstructured data storage. High-performance applications like databases that utilize structured data require block storage in order to provide the response times and synchronous replication speeds needed to avoid applications timing out. This storage often uses high-performance drives or, with growing frequency, SSD storage. iSCSI is a common interface used for block storage to provide direct disk access to these applications. But applications using unstructured data (which represents 80% of data under management,

on average) store it in filesystems (which provide the structure), not as blocks, and usually use interfaces like SMB or NFS unless the applications are rewritten for iSCSI. For the most part, block storage interfaces like iSCSI do not lend themselves to applications using unstructured data.

In addition to compatibility, block storage interfaces suffer other shortcomings relative to file-based protocols. Because block-based applications are primarily limited to single-node storage targets, their ability to scale can be quite limited compared to file-based storage systems spanning multiple servers. And unlike with unstructured data, replication of structured data requires that the disks at the replication site be identical to those at the source site. Since applications using unstructured data are disk agnostic and can address multiple servers, they are particularly suited to solutions with SMB or NFS interfaces and that target scalable storage, particularly object storage. Thus for optimal storage performance and cost control, administrators devote special attention to tier storage according to the specific requirements of each class of user or application. (Figure 2)

	Block	File
Example Application Interface	iSCSI	Standard SMB/NFS
Application Types	Databases, Exchange	Departmental Shares, Home Directories, Application Logs, Video & Images, IoT telemetry and sensor data, Productivity Applications, etc.
Share of Data	<20%	>80%
Scalability	Limited	Massive
Replication Storage Type	Identical	Mixed OK

Figure 2: Comparing aspects of block- and file-based storage

DATA PROTECTION CHALLENGE Data protection is comprised of archival, backup, and disaster recovery (DR). The primary purpose is to maintain access to data that is no longer regularly used or to be able to recover current data if it is lost. For archiving, the authoritative copy of the data may be stored off site and recalled when needed. With backup, the data remains in use, or at least kept in primary storage, and a copy is made and stored for retrieval if the original data is lost. The most common method for protecting data is using a software application to direct data to disk or tape. Magnetic tape has been used for data storage for over 60 years and the technology has not changed all that much during that time. It is still in use due primarily to inertia (the devil you know…) and its perception as being cheap on a $/GB basis. Using tape, however, is very cumbersome, time consuming, and prone to error, making it a much derided medium for data backups and archiving.

With steep reductions in the cost of disk and the development of deduplication over the last decade, disk-to-disk archive and backup has gained more and more share from tape. Disk targets range from removable (very slow) optical disk and commodity magnetic disk to specialized backup and archiving appliances. But all disk-to-disk backup still suffers from one or more of the following major drawbacks: high cost, limited functionality, vendor lock-in, limited scalability, and cumbersome deployment and management. Sometimes disk-to-disk-to-tape methodologies are also deployed.

More recently, disk-to-disk-to-cloud data protection solutions have appeared, offering the scalability, availability, and economy

of the cloud as a storage target. While theoretically, using the cloud is quite appealing, in practice, integrating cloud storage into an established IT infrastructure can be incredibly problematic due to issues like latency, communication protocols, and data security. The primary concern being the time to restore due to limited bandwidth and highly latent links.

DR can leverage both backup and archival while centering around bringing operations back up when a site either partially or fully fails. DR involves rebuilding site functionality as quickly as possible so as to minimize the impact on overall IT operations. This rebuilding can occur either in the same location or off-site at another location. Traditional reliance on tape, with its complex off-site logistics and slow data search/access, has made rapid tape-based DR unachievable. Replication (mirroring or storing backup data in one location to another) is a common but potentially expensive way to implement a DR strategy if it involves full hardware duplication, which doubles datacenter capital costs. DR planning is challenging, time consuming, and difficult to get right. For these reasons, DR is often put off or avoided as long as possible, with organizations adopting a "hope for the best" strategy.

Introduction

The data networking and data storage industries have both been around for over thirty years yet have largely grown independently of each other. Vendors either take responsibility for transporting bits from one location to another (networking), or for making sure bits are persistently stored for later retrieval (storage). Meanwhile, an entirely different crop of companies have emerged to create software and hardware that leverages the power of the network to improve productivity. The scale and speed of today's business would be impossible without tools such as VoIP, web conferencing, telepresence, e-mail, and chat that we now take for granted.

With a management pedigree spanning both networking and storage, Panzura examined the parallel but separate development of networking and storage technologies and asked, "Why?" Why aren't networking technologies being exploited to solve business and technical challenges with storage systems? Why in today's connected world are there still geographic islands of storage? Why, despite pervasive Internet availability, are people still resorting to burning DVDs and mailing flash drives as a method of distributing files? Why is file management localized in a connected enterprise?

File-based NAS storage has historically communicated on the LAN only, using it as the connection between the filer and client computers in a traditional client-server model. Panzura extends the filesystem beyond the LAN by leveraging Internet connectivity, allowing the Freedom NAS filers to communicate not just with LAN-based clients, but also with remote Freedom filers to leverage a unified, global filesystem and unified namespace that breaks the location barrier of traditional NAS.

Panzura incorporates several networking and storage innovations to globalize the filesystem across sites by leveraging the Internet. It has been engineered as a software defined, multi-cloud global system from the ground up, and while it can be deployed in a single site like other file storage products, its true power is realized when it spans multiple clouds or sites unifying both on-premis, hybrid and in-cloud workflows either in a public or private deployment.

Freedom CloudFS Overview

To address the storage tiering and data protection challenges outlined above, Panzura examined how data is created, stored, and consumed within an enterprise-class organization. By developing a distributed filesystem specifically designed to accommodate highly latent remote object stores by incorporating network acceleration technology, Panzura was able to overcome the main limitations preventing enterprises from successfully integrating cloud storage into their infrastructure.

The Panzura Freedom filer architecture is comprised of four major component blocks: the Freedom Interfaces, the Freedom Data Path, the Freedom API and the Freedom CloudFS. (Figure 3) Together, they provide a multi-cloud file services platform

that enables high performance tiered NAS, global file collaboration, active archiving, backup, and DR across all enterprise locations. The Freedom filer architecture enables the enterprise to consolidate their unstructured data and eliminating islands of storage. Key features of the Panzura Freedom filer solution are described in this section.

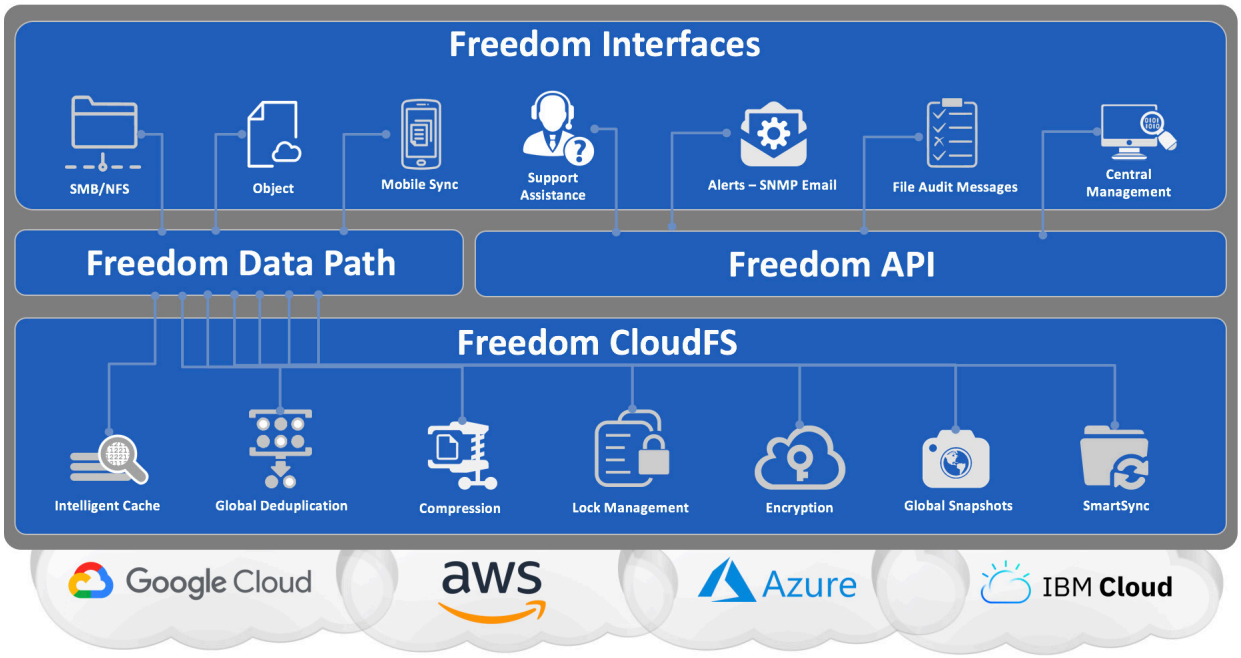


Figure 3: A block diagram of the major functional components within a Panzura File Services Controller

File Based Storage

As discussed above, storage optimized for block data will be very different from that optimized for file data, due to different requirements. Panzura developed a high-performance file-based global storage platform for the cloud to address the 80% of current data that is unstructured. By supporting NFS and SMB transfer protocols commonly used by most applications, Freedom Filers can plug into existing IT infrastructures without any changes while connecting to all major cloud storage platforms, simplifying deployment and minimizing impact on operations. All data is managed under a global filesystem, simplifying user interaction and system administration while tying into enterprise applications and targeting both local disk and the cloud.

Cloud Object Storage

Object storage, the typical storage system used in the cloud, breaks data up and stores it as flexibly-sized containers or chunks that can be individually addressed, manipulated and stored in many locations, not tied to any particular disk. Each object usually has some associated metadata. Object storage can scale to billions of objects and exabytes of capacity while protecting data with greater effectiveness than RAID. In addition, due to the discrete scale-out architecture of object storage, drive failures have little impact and self-healing replication functions recover very rapidly vs. weeks for large capacity RAID systems. This combination of scale and robustness make object storage an ideal target for warehousing enterprise data. The Panzura Freedom family of filers interface directly with all major cloud object storage APIs and related storage tiers, avoiding vendor lock-in, and leverage object-based cloud storage as a data warehouse to provide massive scale and availability at a very compelling cost structure.

Panzura Snapshot Technology

Snapshots for Consistency

Snapshots capture the state of a filesystem at a given point in time. For example, if blocks A, B, and C of a file are written and snapshot 1 is taken, that snapshot captures blocks A, B, and C to represent the file (Figure 4). If someone then edits the file so that block C1 replaces C and snapshot 2 is taken, the data pointers in the snapshot file blocks A, B, and C now point to A, B, and C1. Block C is still retained but not referenced in snapshot 2. If someone wanted to recover to the original state, they can restore snapshot 1, then the system will point back to A, B, and C, ignoring C1.

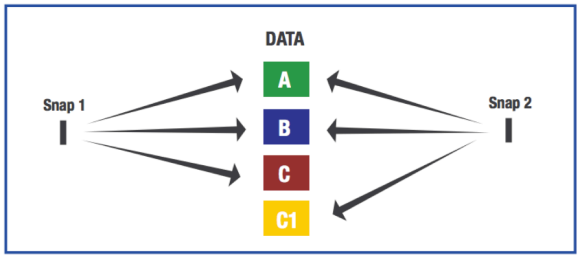


Figure 4: Snapshots maintain filesystem consistency

By using snapshots for creating and saving an ongoing series of recovery points for different stages in data's lifetime, a consistent state of the filesystem can always be restored in the event of a data loss.

Snapshots for Currency

Panzura uses **differences** between consecutive snapshots both to maintain filesystem consistency as well as to protect data in the filesystem. In a process called syncing, the Panzura filesystem takes the net changes to metadata and data between consecutive snapshots and sends them to the cloud. The metadata portion of these changes is retrieved from the cloud by all other Panzura Freedom filers in the configured CloudFS, where they are used to update the state of the filesystem and maintain currency (Figure 5). This system updating occurs continuously across all filers, with each filer sending and receiving extremely small metadata snapshot deltas and using them to update the filesystem seamlessly and transparently.

For example a controller in Brazil (red in Figure 5) takes Snap 1 and then later takes Snap 2. The difference in metadata between Snap 1 and Snap 2 for Brazil is shown in red as 1-2Meta. The difference in data between Snap1 and Snap2 for Brazil is shown in red as 1-2Data. Brazil sends its 1-2Meta and 1-2Data to update the cloud, as do all other controllers in the infrastructure. Brazil also receives back metadata updates for all other controllers (shown as 1-2Meta in green for India and in blue for South Africa in Figure 5).

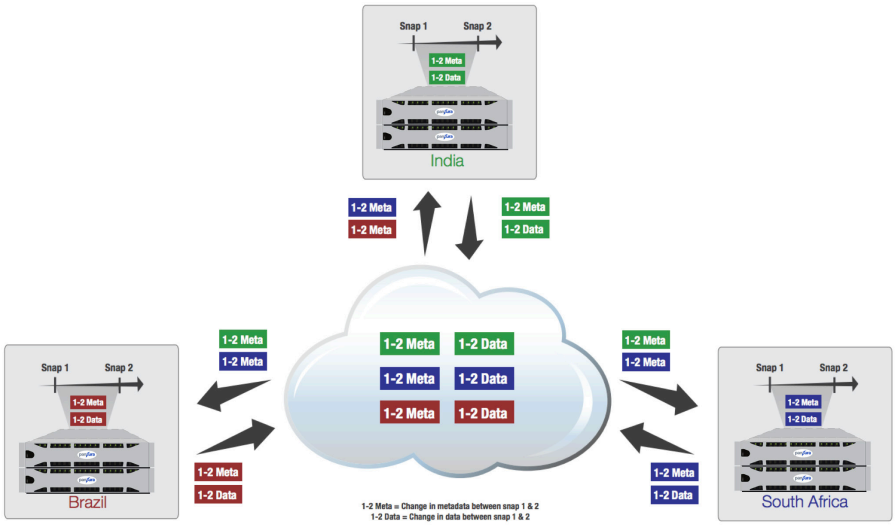


Figure 5: Snapshots' role in global replication

All of the changes in data and metadata are stored and tracked sequentially in time such that should a data loss occur at a Freedom filer or in the cloud, data can be restored to any previous state at which a snapshot was taken, without the need to follow a separate backup process.

It is important to reiterate that the size of these snapshot deltas are exceptionally small relative to the data in the filesystem; thus they can be captured continuously and use bandwidth and capacity very efficiently. The result is the Holy Grail of a global filesystem: a solution that requires almost no overhead but provides near real-time, continuous rapid updates across all sites for global filesystem currency.

The key to a current global filesystem is accurate and efficient transfer of only that data that is needed to make sure the filesystem views of each Freedom filer remain current. Panzura snapshot technology enables currency across a globally-dispersed filesystem with minimal overhead, providing local NAS responsiveness to a worldwide infrastructure.

Snapshots for Efficiency

In addition to system snapshots used to maintain consistency and currency, Panzura Freedom filers also have no practical limit for user-managed snapshots. This category of snapshots allows users to recover data on their own by simply finding the desired snapshot in their inventory and restoring it. This self-service recovery greatly reduces demands on IT by allowing users to recover data on their own, without IT intervention. Policies around user-managed snapshots (frequency, age, etc.) are defined by IT administration.

For example, a Microsoft Windows user in India travels to Brazil and realizes she needs a file that she deleted 3 months ago. She directs her Windows Explorer to the local Brazil Quicksilver controller and navigates to her snapshot folder, finds the date/time that corresponds to the filesystem view that contains the file she wants to recover, opens that snapshot, and navigates to the file or files she needs to recover, then just drags and drops the needed file(s) into her current filesystem location where she wants them restored. Within minutes, she has recovered whatever files she needs and can continue with her work, all without involving anyone from IT.

IT administration can dynamically change snapshot policies as needed to balance frequency and duration for optimal system performance and user satisfaction.

Snapshots Benefits

Panzura snapshot technology provides three major benefits: Global filesystem consistency, currency, and efficiency. Continuous snapshots provide very granular recovery points so that in the event of a data loss, a consistent filesystem state can be restored with minimal disruption or delay.

By syncing all filesystem views globally in real-time, Panzura snapshot technology provides all users in all locations with a current view of the entire filesystem, allowing them to experience cloud storage as if it were local, finally solving the key inhibitor to a true global filesystem.

By empowering users to recover their own data as needed, Panzura snapshot technology offloads a key aspect of user support, freeing up time for strategic IT projects. The Panzura Freedom Filer brings the power of the cloud to enterprises without sacrificing the user experience.

Intelligent Caching

Intelligent Read Cache

Panzura CloudFS utilizes a user-definable percentage of the local storage as the IRC to intelligently track hot, warm, and cold file block structures as they are accessed. This caching dramatically increases the I/O performance of reads by servicing them from local cached storage rather than from external cloud storage. The filesystem also buffers against variations in cloud availability to help maintain consistent read/write response times.

SmartCache Policies

SmartCache is a dynamically managed caching technology that allows administrators to create intelligent caching policies based on defined rules. SmartCache policies provide a flexible method for the storage administrator to directly manage and influence the performance and availability of reads for explicit types of data via specific policies. Caching policies provide two basic functions. The first function is pinned data. Pinned data keeps data available on local storage using flexible wildcard policy rules. Pinning is a forced action and executed against full files whereas IRC caching is a read-stimulated action executed against frequently accessed blocks within a file. Pinned data results in a 100% local read guarantee whereas IRC is deterministic based on previous I/O read patterns within the local disk. The second is Auto-Caching which has the system automatically cache data locally based on defined rules. However, auto-cached data can be evicted for requested hot data, if needed.

The pinned or auto-cache data is a subset of the total IRC storage tier. Pinned data is considered high-priority cached data that is never evicted unless authorized by the administrator, whereas auto-cached (cached based on wildcard rules) or IRCcached (data blocks automatically cached based on observed usage patterns) can be evicted by the system if needed to make space for more frequently accessed data. The balancing of pinning and IRC is delicate as a pinning rule will force data blocks to be logically placed inside the IRC, consuming IRC space, which may affect the IRC utilization and efficiency in ways that the administrator may not have considered. Because pinned policies are of the highest priority and override caching rules based on observed behavior, careful attention should be given to those policies so as to not consume all of the local storage leaving little for actual hot data. New functionality added in the 7.1 release provides an even higher degree of automated caching capabilities called Auto Pre-populate. If enabled, the filer will automatically pre-populate or pre-cache files based on ownership changes between filers in a CloudFS. This is particularly helpful in collaborative workflows where users at different sites are working on the same datasets. As the filer detects ownership changes between locations it will automatically cache data in the same directory in anticipation of user read requests on those files between sites.

The SmartCache policy allows administrators to define wildcard based rule sets. In this way administrators can define matching rules for specific directories within the filesystem or even specific file types across the entire filesystem. For each defined rule the administrator can define caching actions for that given rule. Rule actions are as follows;

- **Auto Cache:** This is the standard behavior of the filer. The filer will evict blocks of a file as needed to accommodate new data.
- **Deny:** When this action is selected, the creation of files with names matching the glob expression is not allowed for that filesystem.
- **Pinned:** Applies a 'last out' policy to data. It avoids evicting data unless the cache is full and new priority data is being ingested. Pinned data will be evicted only as a last resort after other data to maintain normal operations.
- **Not Replicated:** Causes the data not to be copied to the cloud. This creates an unprotected scratch or temp space and should be used cautiously since the data is persisted locally but considered temporary as it is not available in the case of a DR event.

Local Storage Usage

A portion of the local disk storage is allocated for IRC. This portion is configurable and is set to 40% by default. Over time and through general usage, the system dynamically populates the IRC with hot data blocks from all of the files being read by users. The most optimal and efficient IRC configuration is to have the IRC 100% full of hot blocks, with cold blocks being evicted to the cloud. In this case, a high percentage of reads are serviced directly from the IRC rather than from the cloud. This is the optimal caching state, but is harder to achieve the more pinning rules are added.

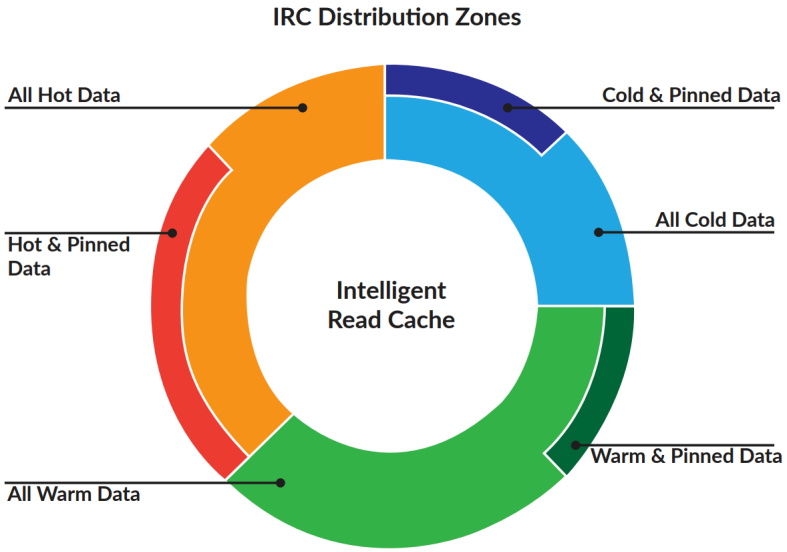
Blocks residing in IRC are characterized by a combination of 3 different temperature states, 2 modification states, and 2 protections states. These are:

- Pinned – Blocks that have been pinned receive the highest priority in the IRC and are the last to be evicted, but only if critical write space is needed.
- Hot – Blocks frequently being accessed for reads. The goal is to have all hot blocks in the IRC.
- Warm – Blocks that were recently hot but have not been read as recently as any of the hot blocks. They will be evicted after cold but before any hot blocks if extra IRC space is needed.
- Cold – Blocks that have not been accessed for 30 days. These are the first blocks to be evicted when the IRC needs space for pinned, hot, or warm blocks. There should always be some cold blocks as this indicates that the IRC completely holds all pinned, hot, and warm blocks.
- Recently modified – Blocks that have been written to as part of updates to a file.
- Not modified – Blocks that have not been written.
- Protected in the cloud – Blocks that have been successfully uploaded to the cloud storage.
- Not yet cloud protected – Blocks that are pending upload the cloud.

Pinning consumes IRC space by forcing complete files into the IRC. The amount of space consumed by pinned data depends on the aggressiveness of the pinning policy and the number of rules. There are many objectives and use cases for pinning, not all of which can be documented in this paper. Pinning is a technology designed for the administrator to satisfy user or site needs. From a general perspective, pinning overrides the IRC's auto-caching logic to disable eviction indefinitely for specific blocks. Because of this, careful attention to specific pinned rules should be given to prevent a rule that could cause thrashing

of the local cache space (rotating eviction of data with new data due to reduced cache capacity). It is recommended that administrators utilize the Auto-Caching action or enable the Auto Pre-populate feature where possible.

The following picture graphically represents how the IRC is logically divided into multiple zones. The main zones are the cold, warm and hot segments as classified by recent read activity. The small outsize curved bar shows how much data (as a %) within each zone is pinned while the main segment shows the total amount of IRC auto-cached data.

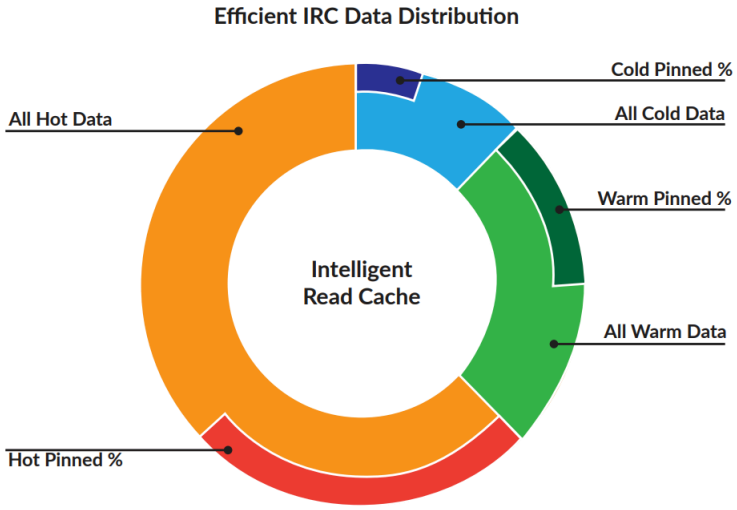


As more data is written to the cache, it is highly unlikely that the IRC will contain "recently modified cold data," as a recent read operation would have automatically changed that data to a hot or warm state. Panzura CloudFS is designed to transition all data into the cloud as quickly as possible. Data is always uploaded to the cloud before becoming hot, warm, or cold based on any recent read activity.

An important aspect of pinning is that when data is pinned, that data is only evicted from IRC if the administrator changes the pinning policy or space is needed for writes and all other hot, warm, and cold data has been evicted. Pinned data is considered high-priority IRC data. Inversely, auto-cached IRC data is treated as low-priority cache data that can be evicted automatically by the system as IRC space is needed for new hot data. As more pinned data consumes the IRC, the usable auto-cache zone is reduced, which eventually negatively affects the most frequently read data causing it to be evicted and then re-read continuously. Therefore aggressive policies that pin large amounts of data should be used sparingly as this could cause excessive local disk I/O and reduce performance.

Ideally, most of the data that applications need should be resident in the IRC storage cache. The above diagram depicts a case where all of the hot and warm data is auto-cached with some cold data and some is pinned. Overall, most of the IRC local-disk space is being used by active data (hot+warm). The amount of cold pinned files should always be monitored as it indicates a pinning rule that is no longer accurate or potentially no longer needed. Those rules should be removed from the system.

In an inefficient configuration, the IRC could contain cold data which is 100% pinned. This is undesirable as that cold data is permanently locked in place and has forced out some warm auto-cache data. This could represent a system where pinned data is not useful to other users or applications so they are not being followed by reads.



If the majority of the data sitting in the IRC is cold, then the IRC not really being leveraged. Worse yet, improperly configured pinning rules can result in 100% of the cold IRC zone being consumed by permanently pinned data, making the hot and warm zones of the IRC very small. This state puts the performance of the system at risk if a read-storm occurs for files outside of the pinned zones. The system has difficulty managing such a read-storm because it tries quickly to populate the tiny hot and warm zones of the IRC with many blocks but the IRC evictor has to perform an excessive amount of initial pre-work to flush the warm and then hot IRC zone (avoiding pinned objects) and then keep up with the remaining IRC read populations because of the limited IRC space. This results in a very high eviction rate being associated to reads and cache thrashing, ultimately resulting in bad performance. In this use case, the pinning rules should be evaluated and unnecessary rules should be removed.

Global Filesystem

The heart of any storage system for unstructured data is the file system. Key filesystems that have shaped the market include VxFS (Veritas), NTFS (Microsoft), WAFL (NetApp), and ZFS (Sun). A successful filesystem must be highly scalable, high performing, flexible, and manageable. NetApp built much of its success around WAFL and its ONTAP OS. WAFL combined RAID and the disk device manager with the file system plus replication and snapshots (limited per volume). Its primary target is HDD. ZFS put all these elements plus encryption and deduplication in one stack, is massively scalable, and targets HDD and SSD natively but has no native cloud integration.

The Panzura CloudFS file system was engineered to closely manage how files are managed and stored to provide seamless, high-performance, and robust multi-cloud data management. It improves on WAFL and ZFS while integrating cloud storage as a native capability.

Any user at any location can view and access files created by anyone, anywhere, at any time. The file system dynamically coordinates where files get stored, what gets sent to the cloud, who has edit and access rights, what files get locally cached for improved performance, and how data, metadata, and snapshots are managed. The structure of the file system has no practical limit for the number of user-managed snapshots per CloudFS. Panzura's innovative use of metadata and snapshots for file system updates, combined with unique caching and pinning capabilities in the Freedom Filers, allows customers to

view data and interact through an enterprise-wide file system that is continually updated in real time. Support for extended file system access control lists (ACLs) empowers administrators to set policies that determine what access and management functions per file will be available on a per user basis. Because the file system is global and shared across all filers and all data is also stored in the cloud, all data is always available on any filer in the CloudFS, even if network connections are temporarily lost to one site for any reason.

Global Namespace

At the highest level, the Panzura global namespace is an in-band file system fabric that integrates multiple physical file systems into a single space and is mounted locally on each node. The entire global namespace has the root label of the distributed cloud file system.

As an example, the following 2 global namespace paths point to the same directory (\projects\team20) and are visible from both nodes as well as locally on nodes cc1-ca (California) and cc1-hi (Hawaii).

H:\ -> \\cc1-ca\cloudfs\cc1-ca\projects\team20

J:\ -> \\cc1-hi\cloudfs\cc1-ca\projects\team20

It is important to note some fundamental differences between the Panzura global namespace fabric and conventional global namespace (GNS) concepts. Unlike Panzura's global namespace, conventional GNS architectures require a database process on each storage system (either in-band or side-band on a separate dedicated GNS system). These distributed databases own and manage all file system metadata transactions. File operations are intercepted in-band, processed and acted on by the distributed database instances before each file operation is allowed to complete. Changes to file metadata anywhere within the GNS require complex out-of-band distributed database replication, synchronization, arbitration and resolution while simultaneously attempting to provide real-time access to the file with guaranteed consistency. The reliance on multiple distributed database processes could introduce complexities, in-band latencies and operation challenges that fail to scale at global multi-site levels. Examples of conventional GNS solutions are Microsoft DFS, F5 ARX and EMC Rainfinity. Additionally, these GNS architectures are somewhat limited in their ability to offer capabilities like global snapshots. The Panzura global namespace has no reliance on underlying distributed databases and avoids common GNS limitations (e.g. speed, transactional data coherence, write order fidelity, open files, precision, in-band operation, global snapshots). At a fundamental level, the file system fabric is the namespace engine.

Global File Locking

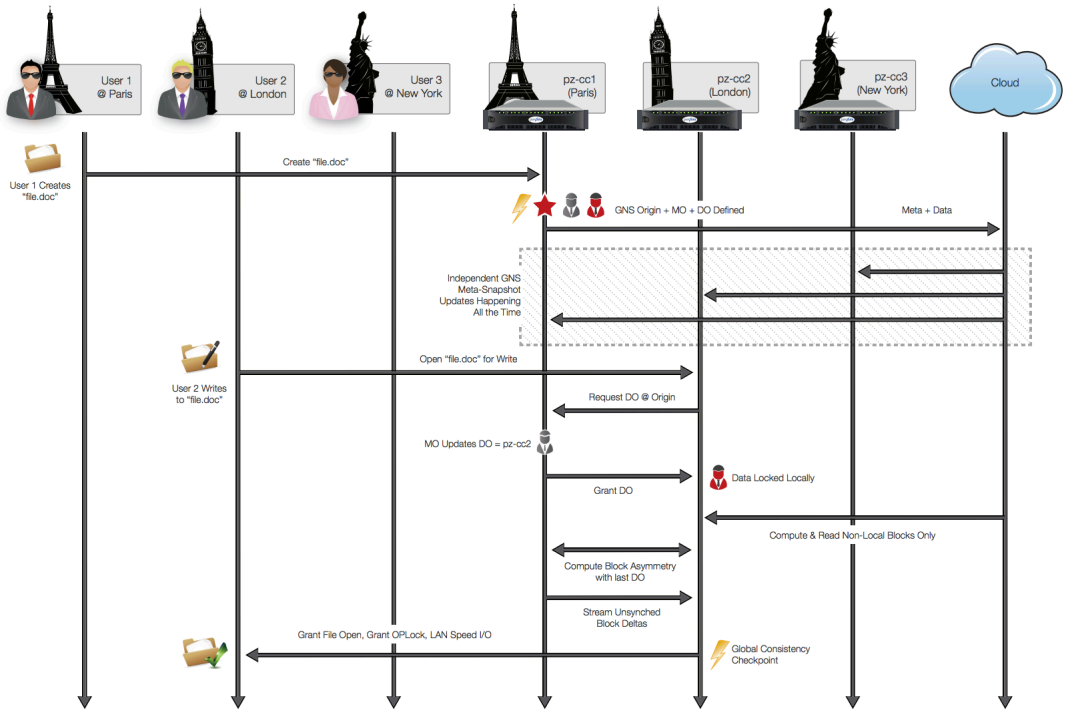
The below diagram is used to discuss the lower level details and concepts of data ownership, data mobility and global locking.

The following core concepts and symbols are used to describe the locking flow:

- **The Origin** – the node where the file was originally created.
- **The Data Owner** – Freedom Flash Cache grants ownership of the data by the Origin. The payload is built on this system and the authoritative data instance is managed from here.
- **Ownership Metadata (OM)** – the metadata showing the Data Owner (DO) state.
- **The Authoritative Write Node** – the Freedom system where locks for a data write operation are executed. This will normally be the DO. Writes here always happen at LAN speed.

- **Traditional File Lock** – a lock issued against a file by a file system, a server or an application. This lock may consist of extensive application specific meta-information and be written into part(s) of the file payload and/or its file system metadata.
- **File Coherency Locks** – these are file system locks that are issued by applications in order to arbitrate guaranteed consistency between applications writing/reading to a single file. An example is: Microsoft Office Application locks.
- **Opportunistic Caching Locks** – a delegated right issued by a file server protocol engine for a remote client to cache a file locally to increase client-side performance. This is not necessarily a guaranteed write lock. This delegation may be revoked by a file server at anytime. An example is: Microsoft SMB OPLOCKS.
- **Data Asymmetry Resolution (DAR)** – Panzura technology that efficiently resolves differences between remote sets of files and transports the required blocks to the DO.

Data Ownership, Data Locking and Data Mobility



The Panzura Distributed File System was designed so that data and metadata are physically decoupled. This decoupling enables the file system to be highly flexible in referencing which physical blocks are used to construct a file. Global distributed file locking leverages this flexibility by assigning a DO to all files. This DO state is held within the OM of each file and is easily transported via snapshots. A Freedom system that wants to be the new DO communicates with the Origin, whose location is defined by a unique unified namespace path for each file when the file was originally created.

Within distributed file locking, DO states naturally flow from node-to-node. DO transitions are frequent events and are negotiated via small real-time peer-to-peer communications among Freedom systems. As the DO flows to a new node, that new node instantly becomes the authoritative write node. All new writes to the file will now happen at the new DO node at full LAN Speed. Note: The Origin is never involved in the I/O data-path during a write operation once the DO successfully migrates.

The final step in assimilating blocks after a DO transition is to resolve any data asymmetry. This involves a direct peer-to-peer communication between the Origin and the new DO, and possibly the current DO (which might not be the Origin). Within this peer-to-peer stream the OM computes a final delta list of real-time changes that may have occurred since the DO changed. This list, which can be as small as a single file system block, is streamed directly to the new DO via a secure optimized data channel. The new DO processes all remaining blocks deltas, making the file current and consistent.

All file reads and writes from that Freedom system now happen as local I/O operations on the new DO. The DO retains full read/write ownership until a new DO transition occurs.

Below are two diagrams that describe two scenarios showing the transactional details of the distributed file locking architecture detailed above.

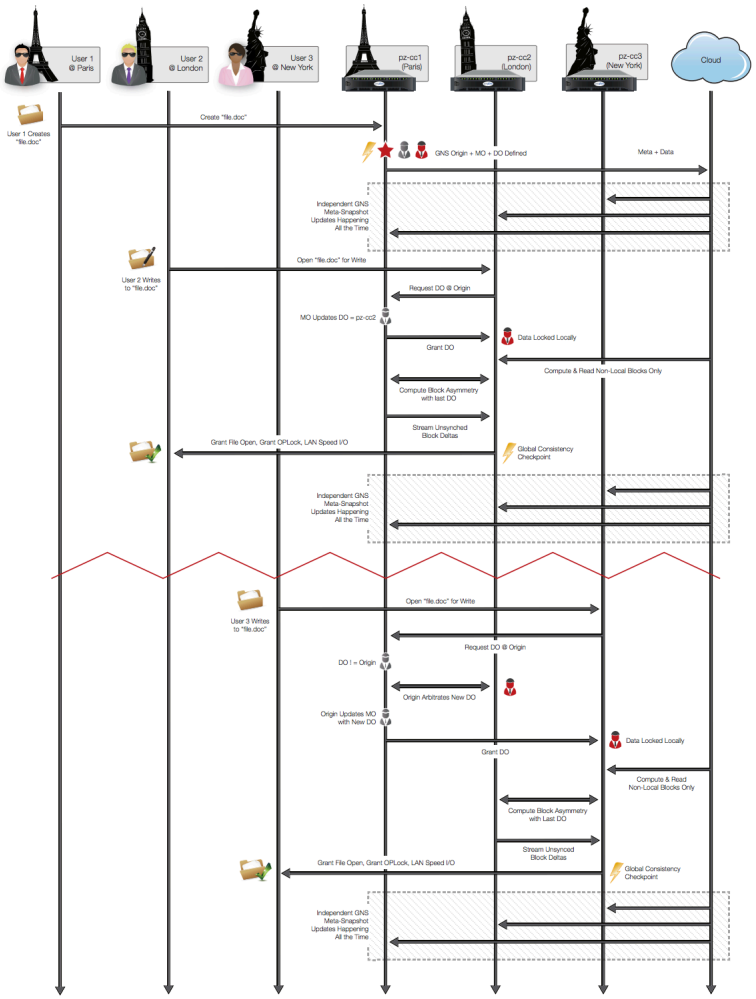
A 2-User Transaction

In this transaction two users will open the same file for write, from one unique global namespace path. Each user will experience LAN speed I/O for the read and write operations as well as 100% data consistency via the data-ownership locking and datablock mobility.

User 1 is in Paris and creates a new file called “File. doc” on the Paris node. This transaction defines the file Origin because it assigns the unique global namespace path to a new file. The metadata is updated and DO is assigned as pz-cc1 in Paris.

The Paris node will eventually write the metadata and payload for File.doc to the cloud. This will happen independently of the original write. All nodes in the Panzura Distributed Cloud File System global namespace will independently read from the cloud the metadata of all other nodes from their private regions and update their file systems and namespace view.

Later in the sequence, User 2 (in London) tried to write to the same file by accessing the unique global namespace path on his local system in London (pzcc2). This is the start of the global read-write locking transaction. Controller pz-cc2 will request DO from the Origin. The Origin evaluates the OM state and discovers that it itself is the current DO. Controller pz-cc2 is granted DO and the OM is updated by the Origin.



From here, pz-cc2 will resolve the location of all blocks. Some may come from the local file system instance and some may be transported in from the cloud. As a final consistency state check, pz-cc2 asks the previous DO if it holds any in-flight data blocks that have not yet been synced to the cloud. This resolves any data asymmetry. The processes occur in parallel to all other block reads and results in the entire state of the file being fabricated on pz-cc2 with guaranteed data consistency. At this point pz-cc2 will serve the file to the client at local LAN speeds. All other nodes are updated via the normal snapshot process.

A 3-User Transaction

In this transaction three users will open the same file for write, from one unique unified namespace path. Each user will experience LAN speed I/O for the read and write operations as well as 100% data consistency via the data ownership locking and data block mobility.

User 3 is in New York and wants to write to “File.doc”. His application opens the path in the global namespace via his local view, which is consistent. During the initial local file open operation, the New York Panzura Freedom Flash Cache (pz-cc3) evaluates the global namespace path and identifies the Paris Freedom Flash Cache (pz-cc1) as the Origin. The New York Freedom Flash Cache will contact the Paris appliance and request DO. Paris evaluates the OM state and discovers that New York is not the current DO; the London appliance is (pz-cc2). Paris will now arbitrate a DO transition. As London relinquishes the DO, Paris updates the OM with a new DO of New York and advises New York that it is now the new DO.

At all times, New York knew that Paris was not the DO and that London was the last DO but in order to enforce consistency, the Origin was leveraged as the authoritative arbiter of the transition since it must also update the OM. New York is now the new authoritative DO and will compute any missing blocks that are not localized in the New York file system. These blocks can come from multiple locations in parallel.

As part of this process, pz-cc3 (New York) will communicate directly with London (pz-cc2) to compute any in-flight blocks not yet in the cloud from the last Authoritative Write Node (i.e. the previous DO which was pz-cc2, London). Any in-flight blocks are peer-to-peer streamed directly between London and New York to resolve data asymmetry. The result is that the full block structure in New York (pz-cc3) has guaranteed data consistency. At this point pz-cc2 will serve the file to the client at local LAN speeds. All other Freedom systems are updated via the normal snapshot process.

GLOBAL DEDUPLICATION: Unlike other deduplication solutions, which were designed to offset inherent data duplication in localized, inefficient file systems, Panzura designed an interconnected, global file system that stops file-level duplication before data gets stored. Since only unique copies of files across all sites are preserved by the filesystem, data is deduplicated before it is ever stored. Capacity is optimized further by running advanced, inline block-level deduplication on any data that gets stored on the network to remove blocks common across different files. Unlike any other deduplication provider, Panzura embeds the deduplication reference table in metadata, which is instantly shared among all Freedom Filers. This inline deduplication method removes data redundancy across filers, rather than just based on data seen by a single controller. Thus each controller in the network benefits from data seen by all other controllers, ensuring even greater capacity reduction, guaranteeing all data in the cloud is unique, and driving down cloud storage and network capacity (and cost) consumed by the enterprise.

MILITARY-GRADE ENCRYPTION: One of the top concerns most frequently expressed by IT professionals about cloud storage is data security. Because data is being transmitted to and stored by a 3rd-party cloud storage provider outside the corporate firewall, some worry that their data will be exposed and at risk for theft. The perception is that keeping data inside the firewall is inherently safer. This concern must be overcome by any cloud storage solution before it can become mainstream within an enterprise.

Panzura addresses data security concerns directly by applying military-grade encryption to all data stored in the cloud. Each Freedom Filer applies AES-256-CBC encryption for all data at rest in the cloud. In addition, all data transmitted to or from the cloud is encrypted with TLS v1.2 to prevent access via interception. Encryption keys are managed by the enterprise, never stored in the cloud. This complete, robust two-tier encryption solution is in addition to the typical multi-layer security provided by mainstream cloud storage providers. In some cases, customers find that the combined security of a Panzura+cloud solution is greater than they can reasonably achieve within their own infrastructure, making cloud storage safer than some private cloud deployments

Summary

The cloud offers tremendous potential for enterprises to reduce storage costs, improve productivity, and reduce data availability risk. Tapping that potential fully and effectively can provide significant competitive advantage while reducing both business and technological risk. To date, enterprises attempting to fully integrate the cloud as a storage tier have been faced with building their own limited-capability solution by kludging together different technologies from various vendors, many of which were never designed to be used with cloud storage. This "Frankenstein" cloud storage implementation fails to realize the full benefits of cloud storage while consuming precious IT resources in implementation and management.

Panzura's Freedom Filer breaks this cycle with a fully cloud-integrated enterprise storage solution to handle NAS, active archiving, DR, and backup. By designing a global filesystem and namespace with cloud integration at a fundamental level, the Panzura solution brings the cloud as a seamless storage tier for the first time while enabling global file sharing and full access to all files in the system from any location at any time.

This game-changing technology finally brings the full power and benefits of cloud storage to enterprise customers, helping to break the unending on-site storage expansion cycle while eliminating islands of storage that inhibit cross-site application or user interaction and productivity and real-time data protection. Panzura makes deploying cloud storage and a global filesystem easy and transparent to users.

Panzura, Inc. | 695 Campbell Technology Pkwy #225, Campbell, CA, USA | 855-PANZURA | www.panzura.com