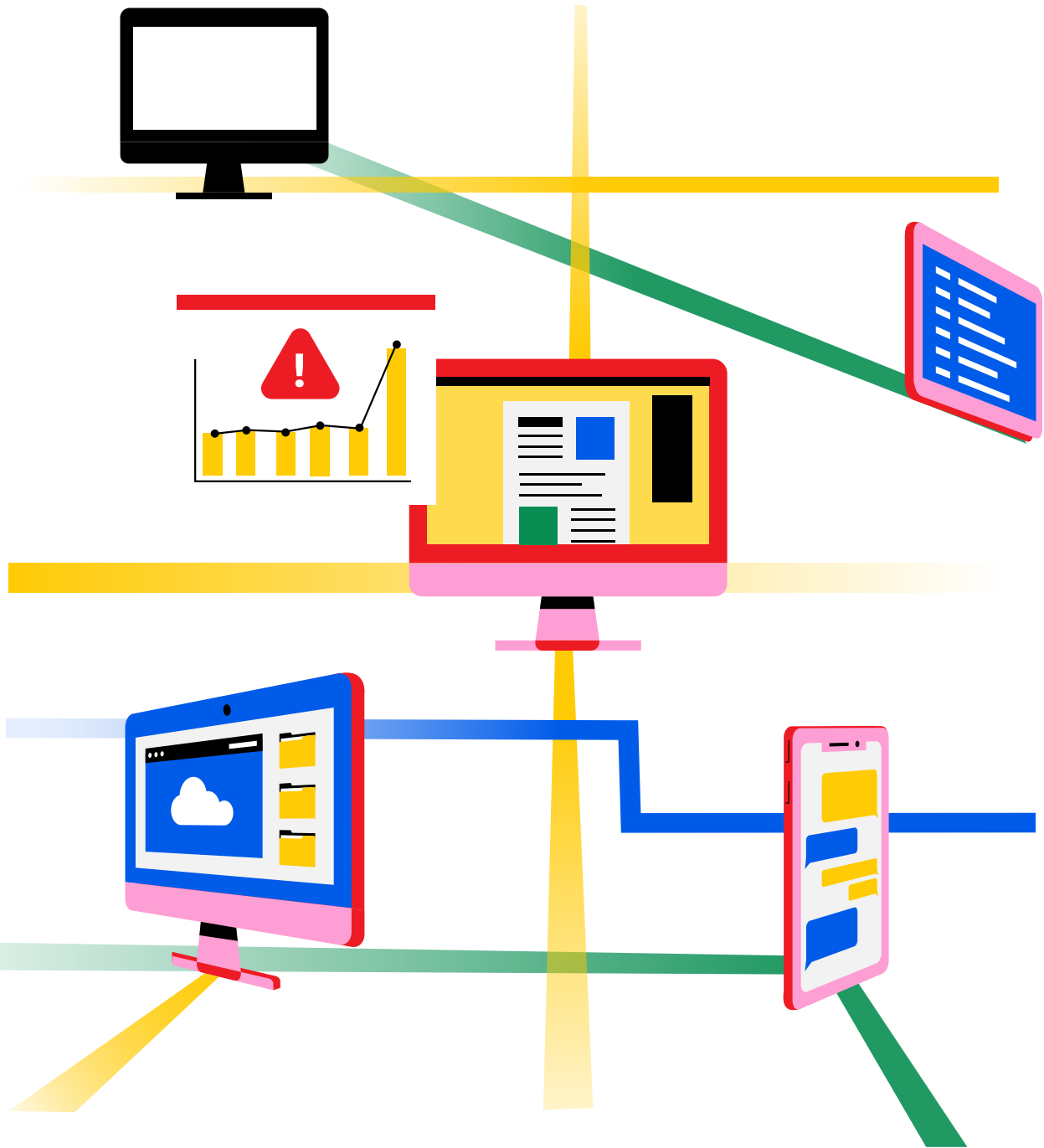


# Panzura Data Services

WHITEPAPER



# CONTENTS



- 3. Introduction
- 4. Panzura Data Services Overview
- 5. The Basic Tier
- 5. Pulse
- 5. Alerts
- 6. Inventory
- 7. The Search Tier
- 7. Search
- 8. Recovery
- 8. Quotas
- 9. Analytics
- 9. The Audit Tier
- 10. Example One – Mass File Deletion
- 10. Example Two – Avoiding Fines for Breaches of Regulations Through Self-Reporting
- 11. Example Three – Legal Hold
- 12. Analytics
- 12. Audit Log Retention
- 13. Summary



# Complete Data Visibility and Management

A modern organization's greatest asset is its data. Forward-thinking leadership understands the impact it can have on strategy and their ability to make better decisions. Gaining a competitive edge, or achieving ambitious goals depends on the ability to intelligently manage, find, audit, analyze and work with trustworthy data quickly, easily and efficiently.

One of the greatest challenges in realizing the value of this data asset is that unstructured data remains widely distributed across storage silos, locations and applications. This makes even finding files extremely challenging, let alone making them available for analysis and insights.


For users, every minute spent searching for files is a drain on productivity and innovation. In a 2021 Asana survey of US professionals, 54% reported wasting time looking for files, and 58% ranked findability as a top-3 problem to solve.

The tendency of users to recreate files they cannot find quickly enough creates yet more data to sort through and puts more pressure on storage.

For IT teams, finding and restoring deleted files from backups, trying to secure sensitive data, proving it's secure and compliant with regulations, and monitoring the health of file networks occupies a punishing amount of time.

This all comes at a very real cost. Just a few minutes a day, multiplied over a workforce, amounts to thousands of hours wasted every year.

It's not just about dead time. The question is, what else could the same people be doing to move projects ahead, using the same time?



It doesn't just mean that work gets interrupted, and projects take longer to complete. It means lost opportunities for innovation – the kind of innovation that can put organizations on a completely different trajectory.

## **The Need For a Complete Data Management Solution**

As data volumes grow, regulatory requirements and cyber threats increase, the need for data management tools – tools to find files, check on file access, recover deleted or damaged files, spot behavioral anomalies and provide file system observability – becomes increasingly pressing.

Individual solutions for findability, auditing, file recovery and visibility add more management complexity, while still not delivering a complete view over all the relevant data locations.

For the organization, the impact of being unable to manage the entire data set ranges from lost productivity and lost opportunity – which can be difficult to accurately assess – to heavy fines through inability to recognize when data regulations have been breached and take corrective action.

## **Panzura Data Services Overview**

Panzura Data Services is a SaaS data management solution providing a single, unified view and management of unstructured data, regardless of where it's stored.

Primarily designed to work hand-in-glove with Panzura's global file system CloudFS, Data Services ingests both metadata and audit log data from CloudFS when connected, to enable visibility and observability over the global file system and related infrastructure, and offer lightning fast file search, audit, recovery and analysis across files in CloudFS.

Additionally, other SMB or NFS storage devices such as NetApp and PowerScale/Isilon can be connected to Data Services to offer findability across an organization's entire file network. Data Services ingests metadata from these devices to enable search, but does not offer audit or file restoration for data stored on devices unrelated to CloudFS.

Data Services offers three tiers of account service, which provide visibility, file search and file audit.

## The Basic Tier: File System Visibility

Data Services offers file system visibility and observability to all Panzura customers, as part of CloudFS.

Available without any additional subscription, this basic account tier provides monitoring and visibility of four file network and storage elements:

- ✓ Data Services Pulse, which monitors and reports on core storage data, file network activity, network health, and cloud connectivity metrics
- ✓ Configurable alerts triggered when storage, system and cloud thresholds are exceeded and may require attention
- ✓ CloudFS node inventory
- ✓ CloudFS dataset inventory

### Pulse

Data Services Pulse monitors the operational metrics that let you know whether file system infrastructure and associated cloud connectivity are operating as expected, or if something needs attention.

Pulse offers overviews of five separate areas: system, storage, cloud, events and high availability and includes graphed analytics to show you what normal looks like, while making anomalous activity obvious.

### System

This section contains health metrics for CloudFS system nodes; the virtual machines that hold the file system metadata and cache frequently-used files for a local-feeling file operation experience.

The system metrics captured are:

- ✓ CPU utilization and load peaks
- ✓ Memory utilization
- ✓ Bandwidth limits

- ✓ Network traffic
- ✓ SMB connections
- ✓ Disk input and output stats

## Storage

This section contains health metrics for CloudFS storage;

The metrics captured are:

- ✓ Local and cloud disk usage
- ✓ Metadata storage utilization
- ✓ Cache statistics, and hit and miss ratios
- ✓ Managed capacity usage
- ✓ LAN upload and download
- ✓ WAN upload and download
- ✓ Snapshot status

## Cloud

This section contains health metrics for cloud connectivity, as well as site-to-site communication.

The metrics captured are:

- ✓ Cloud upload and download failure rate
- ✓ Site-to-site latency
- ✓ Snapshot sync per minute

## Events

This section records significant activity such as the expiry of any licenses applied to Panzura nodes, or performance issues such as excessive latency.

## High Availability (HA)

The high availability section replicates the metrics available in the system section above, for nodes designated as high availability nodes.

## Alerts

Configurable alerts are available for every metric captured by Pulse. These alerts can be set to provide proactive warnings triggered when a threshold is exceeded and may need attention.

Each alert's threshold, method of notification and recipient can be independently configured, and alerts can be delivered by email, Slack or text message.

Additionally, the alerts section supports

## Inventory

Data Services includes inventories of CloudFS nodes, and associated datasets, providing visibility over CloudFS configurations.

The CloudFS node inventory includes every local instance within a CloudFS deployment, along with its current connectivity. The CloudFS dataset inventory reflects where authoritative data is stored, and which CloudFS ring it belongs to.

## The Search Tier: Solving Findability Problems and Recovering Deleted or Damaged Files

Data Services search is designed to find files in near real time, even when querying hundreds of millions of files spread over multiple locations, including CloudFS deployments, as well as other Windows file shares such as NetApp and PowerScale/Isilon devices.

The Data Services search tier includes search, file recovery, soft user quotas and data analytics.

## Search

Search offers a free-text search field into which any known parameters such as file name, or file extension can be entered. Several additional filters are available to refine search results by date, file age and size, and to specify whether the target is a file or folder.

Search results are based on file metadata and include

- ✓ file name and extension
- ✓ location
- ✓ node it was created on
- ✓ file size
- ✓ date created on CloudFS

✓ date last modified

Search includes a number of additional time-saving features for IT personnel, such as the ability to save frequently executed searches.

Search references files available in live file systems, and does not reference snapshots. As a result, deleted files will not be detected by searches using the Data Services search tier.

## Recovery

Recovery searches both CloudFS and CloudFS snapshots to return search results that can be restored to their previous state and location from within Data Services.

Recovery offers the same search functionality and filters as the search operation itself, but includes one more result – snapshots. Search results include every available snapshot the file has been captured by, enabling point-in-time restoration to either the file's original storage location, or a new location as specified.

Unlike restoring files from a backup, using Data Services to restore from a snapshot involves updating metadata, rather than file data itself. This allows rapid file recovery.

This is made possible through immutable data; CloudFS stores new and changed data in the cloud or object store using non-destructive writes, while existing data blocks remain unchanged. Metadata pointers are then updated in real time to reflect which data blocks comprise a file at any given time.

Once captured by snapshots according to an organization's snapshot schedule, files can be restored to their previous state by simply restoring the metadata pointers to that point in time.

Metadata is a fraction of the size of the file itself, so restoration is fast, and consumes only a tiny amount of system resources.

## Quotas

Soft user quotas allow administrators to establish quotas for individual users and user groups based on their home directories, as well as set alerts that trigger when configurable thresholds have been exceeded.



These can be used to warn administrators and users when storage thresholds are being approached, to allow proactive action to be taken.

## **Analytics**

Data Services analytics offered in the search tier offers an overview and allows analysis of data storage, allowing for understanding of what's consuming space in both CloudFS and other connected file shares.

Data analytics shows hot, warm and cold data stored by age, and size, as well as storage distribution by file size and file type. Additionally, a running total of data added by day is shown, to allow unexpected spikes to be easily identified. These metrics are available for both CloudFS and other connected SMB/NFS file shares.

## **The Audit Tier: Solving Audit and Compliance Problems**

Data Services audit turns CloudFS audit logs into meaningful information by searching through audit records – processed syslog events – and returning clear and comprehensive audit trails in near real time.

Audit detects and reports on the following file actions to provide a complete view of user actions taken on within CloudFS:

- ✓ copy
- ✓ create file
- ✓ create folder
- ✓ lock file
- ✓ write file
- ✓ move file
- ✓ read file
- ✓ remove file
- ✓ remove folder
- ✓ remove permissions
- ✓ rename
- ✓ set attribute
- ✓ set permissions

Audit uses the same powerful and burstable search functionality available in the search tier and is capable of returning millions of results in under a second.

The free-text search field is used for any known file identifiers, such as file name, extension, or last-known location, and filters further refine the search criteria by audit

action, age of file or user.

This provides IT personnel with the valuable ability to rapidly query millions of files using known parameters, and to investigate individual file audit logs before further refining searches.

To understand the power of audit in combination with CloudFS, let's take a look at some real examples.

### **Example One – Mass File Deletion**

An agency had 6 million files deleted between one day and the next and wanted to understand when it had happened in order to be able to restore the right files from the relevant CloudFS snapshots, as well as to identify the user who had performed the deletion.

If the name of an individual file known to have been deleted in the mass deletion is available, search can first be used to locate that file. Viewing its audit trail will reveal the deletion action, when it took place and the user responsible. The user and time parameters can then be used to filter for all files deleted by that user, within the specified timeframe.

Alternatively, if no individual details are known, filters can be used at the outset. In this scenario the volume of files removed suggests that directories were deleted, so the first filter applied would be to identify directory deletions between yesterday and today.

Using those results, the search can be further refined to show files within those directories. Viewing the audit trail of affected files then reveals the time of deletion and responsible user.

Now, filters can be set to show all files deleted by that user, within the timeframe, for a definitive list of files affected.

### **Example Two – Avoiding Fines for Breaches of Regulations Through Self-Reporting**

An international firm headquartered in France, with offices and customers throughout Europe and the USA is required to ensure that European-based data remains within the geographic confines of the European Union.

The firm creates two CloudFS rings. One global file system ring is based in Europe and the other in the USA. Each allows offices from within their respective regions to work from an authoritative dataset, as each works off a mapped drive available in their region.

In theory, no data should spill out of Europe into the USA, or vice versa. However, over \$1B in GDPR fines levied between 2019 and mid-2021 shows that it's all too easy for data to move beyond its regulated boundaries. Self-reporting, and being able to prove that spilled data has not been incorrectly accessed or used can avoid the otherwise inevitable fines.

In this example, European client files are accidentally saved onto a mapped drive accessible from within the USA, and as a result are immediately out of regulatory compliance.

However, IT personnel deliberately use Data Services audit to track file creation on nodes belonging to both CloudFS rings, checking for this type of file movement.

By first removing the European files from the US-accessible drive, and then using Data Services audit trail to show that the files had not been accessed and had subsequently been deleted before any effective breach occurred, the firm can report their continued compliance, and avoid paying a fine.

### **Example Three – Legal Hold**

A firm receives a legal hold notice instructing them to collate and preserve specific data within a date range, pending litigation.

For an enterprise using traditional storage, along with regular backups and offsite archival processes, this would require identifying and recovering data from multiple backups, and then sorting through it to determine relevancy by date. The older the data is, the more time-consuming and inefficient it is to retrieve and accurately assess.

Backups tend to be weekly or monthly, or even yearly, based on the age of the data requested, so this type of recovery increases the risk of presenting more data than has been requested because backup date ranges don't align with the necessary date ranges.

In this scenario, the effort required to identify and collate the relevant data – and nothing but the relevant data – is substantial and consumes an enormous amount of IT and

subject-matter-expert time.

Submitting excess data exponentially increases the firm's exposure in both this and future litigation, and it's difficult to estimate the potential impact to the firm of getting this wrong.

By contrast, the firm using CloudFS and Panzura Data Services can identify the relevant data, within precisely the date range requested, and held within only the relevant file paths, within minutes.

They use audit filters to set the date range, and the free-text search field to specify the file path data is or was held on, and/or users who worked on the relevant data to produce a definitive list of files that meet the requested criteria. If required, files can be restored to the state they were in within the specified timeframe, using Data Services' recovery feature.

Regardless of its current location, all of this data can now be recovered to a new directory that they can then make available to their legal team, and eventually to an external legal team as required.

Again, this takes just minutes. Overall savings delivered by this modern approach to data management include thousands of hours spent finding and verifying data as well as mitigating potentially ruinous legal jeopardy.

## **Analytics**

Data Services analytics offered in the audit tier offers an overview and allows analysis of data activity, allowing for understanding of how data in CloudFS is being used. This dashboard shows the most active users, most frequently accessed files and folders, and can be filtered by date range for the most relevant view.

Every metric allows a one-click deep dive into the files or usage that comprise it. For example, clicking on a user displays all user activity for the given timeframe.

## **Audit log retention**

By default, audit logs are retained by Data Services for 90 days and as logs reach the 90-day mark, they are auto-deleted. Additional retention licenses allow log retention for up to 5 years where required.



# SUMMARY

Replacing individual search, audit and monitoring tools with a single pane of glass that also incorporates CloudFS management capabilities offers tremendous potential for organizations to substantially reduce the time IT teams spend on operational activity.

## The business value delivered by Panzura includes:

### **Tightly integrated visibility**

Panzura Data Services vastly reduces the amount of time spent finding files and file activity by providing visibility, findability and activity monitoring over disparate file systems, from one elegant dashboard.

### **Easy data restoration**

Restoring deleted or damaged files from backups is time-consuming and incredibly inefficient. With CloudFS, Data Services finds and restores files in seconds by searching through snapshots and letting you retrieve the version you want.

### **Improved data compliance**

Data Services' ability to track file movement and access allows organizations to monitor and prove data compliance as well as track data spillage, to avoid fines for potential breaches.

### **Observability brings predictability**

Complete visibility over file systems and associated infrastructure allows detection of issues requiring early intervention, as well as providing a clear view of what normal operation looks like.