# GIGAOM

Image credit: Ben Slater

**Enrico Signoretti, Max Mortillaro**
Sep 22, 2021

# GigaOm Radar for Cloud File Systems v2.0

Cloud File Storage, Data Center, Data Storage

# GigaOm Radar for Cloud File Systems

## Table of Contents

# 1. Summary

File storage is a critical component of every hybrid cloud strategy, and enterprises often prefer this type of storage over block and object storage, in particular for big data, artificial intelligence, and collaboration. So we decided to focus our assessment of the cloud file storage sector on two parts: this one, which centers on big data and AI, and another on distributed cloud file storage that concentrates more on collaboration.

Cloud providers didn't initially offer file storage services, which allowed multiple storage vendors to jump in with products and services to fill that gap. With the COVID pandemic still ongoing, the increasing need for data mobility, and the large number of workloads moving across on-premises and cloud infrastructures, file storage is simply better—easier to use and more accessible than other forms of storage.

Lift-and-shift migrations to the cloud are increasingly common scenarios and enterprises often want to keep the environment as identical as possible to the original one. File storage is a key factor in accomplishing this, but simplicity and performance are important as well.

File systems still provide the best combination of performance, usability, and scalability for many workloads. It is still the primary interface for the majority of big data, artificial intelligence/machine learning (AI/ML), and high-performing computing (HPC) applications, and today it usually offers data services such as snapshots to improve data management operations.

In recent years, file systems also have become more cloud-friendly, showing better integrations with object storage, which enables better scalability, a better balance of speed and cost, and advanced features for data migration and disaster recovery.

Both traditional storage vendors and cloud providers now offer file services or solutions that can run both on-premises and in the cloud. Their approaches are different, though, and it can be very difficult to find a solution that both meets today's needs and can evolve to face future challenges. Cloud providers generally offer the best integration across the entire stack, but also raise the risk of lock-in, and services are not always the best in class. On the other hand, solutions from storage vendors typically provide more flexibility, performance, and scalability, but can be less efficient or lack the level of integration offered by an end-to-end solution.

## HOW TO READ THIS REPORT

This GigaOm report is one of a series of documents that helps IT organizations assess competing solutions in the context of well-defined features and criteria. For a fuller understanding consider reviewing the following reports:

**Key Criteria report**: A detailed market sector analysis that assesses the impact that key product features and criteria have on top-line solution characteristics—such as scalability, performance, and TCO—that drive purchase decisions.

**GigaOm Radar report**: A forward-looking analysis that plots the relative value and progression of vendor solutions along multiple axes based on strategy and execution. The Radar report includes a breakdown of each vendor's offering in the sector.

**Solution Profile**: An in-depth vendor analysis that builds on the framework developed in the Key Criteria and Radar reports to assess a company's engagement within a technology sector. This analysis includes forward-looking guidance around both strategy and product.

# 2. Market Categories and Deployment Types

For a better understanding of the market and vendor positioning (**Table 1**), we assess how well solutions for cloud file systems are positioned to serve specific market segments:

- **Small-to-medium enterprise:** In this category, we assess solutions on their ability to meet the needs of organizations ranging from small businesses to medium-sized companies. Also assessed are departmental use cases in large enterprises, where ease of use and deployment are more important than extensive management functionality, data mobility, and feature set.

- **Large enterprise:** Here, offerings are assessed on their ability to support large and business-critical projects. Optimal solutions in this category will have a strong focus on flexibility, performance, data services, and features to improve security and data protection. Scalability is another big differentiator, as is the ability to deploy the same service in different environments.

- **Specialized:** Optimal solutions will be designed for specific workloads and use cases, such as big data analytics and high-performance computing.

In addition, we recognize two deployment models for solutions in this report: cloud-only (SaaS), or hybrid and multi-cloud.

- **SaaS:** The solution is available in the cloud as a managed service. Often designed, deployed, and managed by the service provider or the storage vendor, it is available only from that specific provider. The big advantages of this type of solution are its simplicity and the integration with other services offered by the cloud service provider.

- **Hybrid and multi-cloud solutions:** These solutions are meant to be installed both on-premises and in the cloud, allowing customers to build hybrid or multi-cloud storage infrastructures. Integrating with a single cloud provider could be limited compared to the other option and more complex to deploy and manage. On the other hand, these solutions are more flexible, and the user usually has more control over the entire stack with regard to resource allocation and tuning. They can be deployed in the form of a virtual appliance, like a traditional NAS filer but in the cloud, or as a software component that can be installed on a Linux VM (that is, a file system).

*Table 1. Vendor Positioning*

| | MARKET SEGMENT | | | DEPLOYMENT MODELS | |
|---|---|---|---|---|---|
| | Small to Medium Enterprise | Large Enterprise | Specialized | SaaS | Hybrid and Multi-Cloud |
| Amazon | ++ | ++ | +++ | +++ | − |
| DDN | ++ | +++ | +++ | ++ | +++ |
| Google | ++ | ++ | ++ | +++ | − |
| Hammerspace | ++ | +++ | +++ | − | +++ |
| IBM | + | ++ | +++ | − | ++ |
| Microsoft | ++ | ++ | +++ | +++ | + |
| Nasuni | +++ | +++ | ++ | +++ | − |
| NetApp | +++ | +++ | ++ | +++ | +++ |
| ObjectiveFS | +++ | ++ | ++ | − | +++ |
| Oracle | ++ | ++ | +++ | +++ | − |
| Panzura | +++ | +++ | ++ | +++ | − |
| Qumulo | ++ | +++ | +++ | +++ | +++ |
| Scality | + | + | +++ | − | ++ |
| Weka | ++ | +++ | +++ | − | +++ |
| Zadara | +++ | +++ | ++ | +++ | ++ |

+++  Exceptional: Outstanding focus and execution
++  Capable: Good but with room for improvement
+  Limited: Lacking in execution and use cases
−  Not applicable or absent

Source: GigaOm 2021

# 3. Key Criteria Comparison

Building on the findings from the GigaOm report, *Key Criteria for Evaluating Cloud File Systems*, **Table 2** summarizes how each vendor included in this research performs in the areas that we consider differentiating and critical in this sector. The objective is to give the reader a snapshot of the technical capabilities of different solutions and define the perimeter of the market landscape.

*Table 2. Key Criteria Comparison*

| | KEY CRITERIA | | | | | | |
|---|---|---|---|---|---|---|---|
| | Global Namespace | Hybrid and Multi-Cloud | Integration with Object Storage | Data Management | Analytics | Advanced Security | Edge Deployments |
| Amazon | ++ | + | ++ | + | + | ++ | + |
| DDN | +++ | +++ | +++ | ++ | ++ | ++ | + |
| Google | + | − | + | ++ | + | + | − |
| Hammerspace | +++ | +++ | +++ | +++ | ++ | ++ | +++ |
| IBM | +++ | +++ | ++ | +++ | ++ | ++ | + |
| Microsoft | + | + | ++ | ++ | + | ++ | ++ |
| Nasuni | +++ | +++ | ++ | ++ | +++ | +++ | +++ |
| NetApp | +++ | +++ | +++ | ++ | +++ | +++ | +++ |
| ObjectiveFS | ++ | +++ | +++ | + | ++ | ++ | ++ |
| Oracle | ++ | + | + | ++ | ++ | + | − |
| Panzura | +++ | +++ | +++ | + | +++ | +++ | +++ |
| Qumulo | +++ | +++ | +++ | ++ | +++ | +++ | ++ |
| Scality | +++ | ++ | ++ | + | + | ++ | ++ |
| Weka | +++ | +++ | +++ | + | ++ | +++ | +++ |
| Zadara | ++ | +++ | ++ | + | ++ | ++ | ++ |

+++ Exceptional: Outstanding focus and execution
++ Capable: Good but with room for improvement
+ Limited: Lacking in execution and use cases
− Not applicable or absent

Source: GigaOm 2021

**Table 3** compares the vendors in this sector based on evaluation metrics, which are top-line characteristics of each solution that help define their impact on the organization.

*Table 3. Evaluation Metrics Comparison*

| | Architecture and Scalability | Flexibility | Efficiency | Performance | Manageability and Ease of Use | Security |
|---|---|---|---|---|---|---|
| Amazon | ++ | ++ | ++ | ++ | ++ | ++ |
| DDN | ++ | +++ | +++ | +++ | ++ | ++ |
| Google | ++ | + | ++ | +++ | ++ | + |
| Hammerspace | ++ | +++ | +++ | ++ | +++ | ++ |
| IBM | ++ | ++ | ++ | +++ | ++ | ++ |
| Microsoft | ++ | ++ | +++ | +++ | ++ | ++ |
| Nasuni | +++ | ++ | +++ | ++ | +++ | ++ |
| NetApp | +++ | +++ | +++ | +++ | +++ | +++ |
| ObjectiveFS | +++ | +++ | ++ | ++ | ++ | ++ |
| Oracle | +++ | +++ | ++ | +++ | ++ | ++ |
| Panzura | +++ | ++ | +++ | ++ | +++ | ++ |
| Qumulo | +++ | +++ | ++ | +++ | +++ | +++ |
| Scality | ++ | + | +++ | + | ++ | ++ |
| Weka | +++ | +++ | ++ | +++ | ++ | +++ |
| Zadara | +++ | +++ | +++ | +++ | +++ | +++ |

+++ Exceptional: Outstanding focus and execution
++ Capable: Good but with room for improvement
+ Limited: Lacking in execution and use cases
– Not applicable or absent

Source: GigaOm 2021

By combining the information provided in the tables above, the reader can develop a clear understanding of the technical solutions available in the market.

# 4. GigaOm Radar

This report synthesizes the analysis of key criteria and their impact on evaluation metrics to inform the GigaOm Radar graphic in **Figure 1**. The resulting chart is a forward-looking perspective on all the vendors in this report, based on their products' technical capabilities and feature sets.
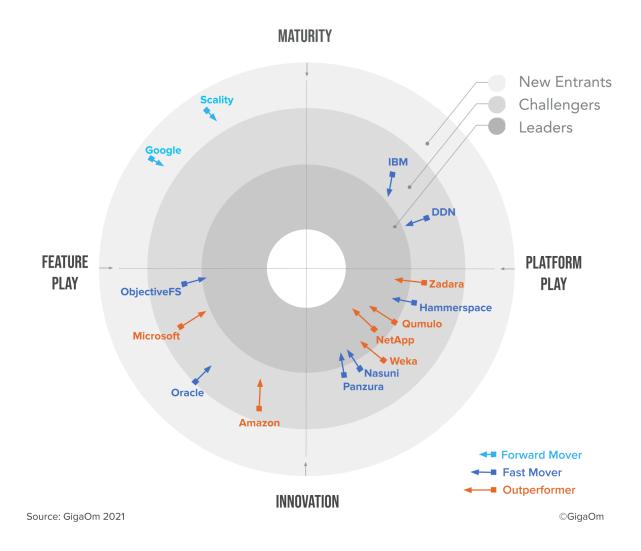


*Figure 1. GigaOm Radar for Cloud File Systems*

The GigaOm Radar plots vendor solutions across a series of concentric rings, with those set closer to center judged to be of higher overall value. The chart characterizes each vendor on two axes—Maturity versus Innovation, and Feature Play versus Platform Play—while providing an arrow that projects each solution's evolution over the coming 12 to 18 months.

As you can see in the Radar chart in **Figure 1**, the lower-right area shows seven vendors closely competing in a zone that focuses on innovation and a platform approach to cloud file systems. Leading the competition is NetApp, whose coherent multi-cloud strategy and outstanding execution

translates into a continuous data management plane (across clouds and locations), a comprehensive enterprise-grade feature set, and varied, flexible deployment options. These features are complemented by outstanding analytics and reporting capabilities. The completeness of NetApp's offering, combined with the ability to address a very broad set of storage use cases beyond cloud file storage, and the availability as a Tier 1 offering both on AWS and Azure, makes it a compelling choice.

Qumulo offers an excellent software-based, vendor-agnostic file system platform that spans both private and public clouds. The solution is built around a well-engineered core data services set and delivers consistent performance, deep object storage integration, and outstanding data mobility features.

Weka's solution doesn't compromise on performance, robustness, or scalability. It combines excellent architectural foundations, outstanding object storage integration, and a specific focus on the most demanding use cases, such as AI/ML, HPC, and HFT.

Zadara delivers its solution as-a-service primarily via managed service providers, with some deployments directly on customer premises. It offers a comprehensive platform that integrates storage, compute, and networking with great analytics and excellent data mobility and data protection options.

Hammerspace takes a slightly different approach to storage. Its global file system seeks to abstract the underlying storage, providing a single namespace where metadata reigns supreme and policy-based data management features automated data mobility, data protection, and tiering features. From a core data services perspective, the solution is complete and takes a novel approach to managing cloud file systems.

Nasuni's solution primarily caters to the needs of distributed cloud file storage, but it can address some cloud file system use cases as well. While it's innovative, offering many enterprise-grade features, Nasuni has a back end that is powered by object storage, an architectural choice that may impact the platform's ability to meet the performance, throughput, and latency requirements of demanding workloads typically running on cloud file systems.

Panzura, although architected differently, addresses the same use cases and verticals. Its global cloud file system is perfectly suited for distributed file access, while also able to support cloud file system requirements. Panzura will, however, have challenges meeting the high throughput and low latency requirements of the most demanding applications, which typically require performance-tuned cloud file systems.

The lower-left area includes three major cloud providers—Amazon, Microsoft, and Oracle—as well as ObjectiveFS, and includes offerings that enable either a specific focus or multiple choices, but without an end-to-end platform solution. ObjectiveFS is the only non-public cloud vendor in this area; its solution provides a cloud file system that uses an object store in the backend. The solution offers consistent performance, great scalability and deployment choices, and it excels in terms of security and multi-tenancy.

Microsoft follows with a broad portfolio of cloud file system offerings, each with multiple performance/cost tiers, that deliver great flexibility in terms of consumption and deployment options. Among these, the most mature enterprise-grade offering is Azure NetApp Files, as the other offerings have more limited feature sets. Microsoft's challenge is to provide a consistent experience in terms of unified management and feature set to its customers across tiers.

Amazon has a number of cloud file system offerings. Among these, Elastic File Storage (EFS), FSx for Windows, and FSx for Lustre are the most popular, along with the recently introduced Amazon FSx for NetApp ONTAP from AWS. EFS delivers a rich feature set for NFS workloads with excellent performance, resiliency, and data durability, while FSx for both Windows and Lustre provide targeted features to Windows and Linux users. Unfortunately, management, monitoring and observability features are scattered across many dashboards, and advanced data protection remains a tedious, manual endeavor, diluting value for customers. Amazon FSx for NetApp ONTAP, as the newest offering, brings AWS a multiprotocol cloud file system offering with the enterprise-grade features for performance, resiliency, and management that have long been requirements for data center environments.

Somewhat behind is Oracle Cloud with Oracle Cloud Infrastructure File Storage, an offering built around strong data durability, high availability, and massive scalability. The interesting aspects of this solution currently are its robustness and focus on data management, while other areas need further development to catch up with the competition. Oracle also offers compelling options for HPC-oriented organizations with its OCI HPC File System stacks, and provides an OCI ZFS image as well.

The upper-right corner shows IBM and DDN, two companies with mature and proven cloud file storage solutions. IBM's Spectrum Scale provides a robust and scalable architecture capable of delivering great performance, support for modern workloads, and, with Watson Data Discovery, outstanding data management capabilities. DDN focuses strongly on AI and HPC workloads with its Lustre-based Exascaler EXA5 appliance, which delivers scalability, performance, and multi-tenancy—key capabilities for these types of workloads.

Finally, Scality and Google are located in the upper-left corner. Scality's SOFS delivers outstanding performance to sequential and throughput-intensive workloads, with a focus on scalability, parallelism, and aggregated throughput. Google refocused its cloud file storage solutions into Google Filestore, offering a number of performance tiers with great raw performance and throughput potential. However, capabilities remain very limited compared to the other competitors.

## INSIDE THE GIGAOM RADAR

The GigaOm Radar weighs each vendor's execution, roadmap, and ability to innovate to plot solutions along two axes, each set as opposing pairs. On the Y axis, **Maturity** recognizes solution stability, strength of ecosystem, and a conservative stance, while **Innovation** highlights technical innovation and a more aggressive approach. On the X axis, **Feature Play** connotes a narrow focus on niche or cutting-edge functionality, while **Platform Play** displays a broader platform focus and commitment to a comprehensive feature set.

The closer to center a solution sits, the better its execution and value, with top performers occupying the inner Leaders circle. The centermost circle is almost always empty, reserved for highly mature and consolidated markets that lack space for further innovation.

The GigaOm Radar offers a forward-looking assessment, plotting the current and projected position of each solution over a 12- to 18-month window. Arrows indicate travel based on strategy and pace of innovation, with vendors designated as Forward Movers, Fast Movers, or Outperformers based on their rate of progression.

Note that the Radar excludes vendor market share as a metric. The focus is on forward-looking analysis that emphasizes the value of innovation and differentiation over incumbent market position.

# 5. Vendor Insights

## Amazon

Amazon Web Services provides a myriad of offerings that either overlap each other or cover narrow use cases. For cloud file systems, it offers four services: EFS, FSx for Windows File Server, FSx for Lustre, and the newly introduced FSx for NetApp ONTAP.

As its name implies, Amazon FSx for Windows File Server provides fully managed, native Windows file-sharing services using the SMB protocol. It supports user quotas, user-initiated file restores, and Windows access control lists. The service integrates with Windows-based Active Directory (AD) or AWS Microsoft-managed AD, and leverages DFS to provide single namespace capabilities.

Amazon FSx for Lustre implements a POSIX-compliant file system that natively integrates with Linux workloads and is accessible by Amazon EC2 instances or on-premises workloads. The solution is linked with AWS S3 buckets, whose objects are then transparently presented as files. Applications can manipulate those objects as files, while FSx automatically ensures changes are committed in the object back end.

FSx for NetApp ONTAP was introduced very recently and provides a proven, enterprise-grade cloud file storage experience to the AWS platform that is natively integrated into the AWS consumption model, while offering superior technical capabilities and a true multi-cloud implementation, thanks to NetApp's presence on all major cloud platforms. FSx for ONTAP supports NFS, SMB, and iSCSI, and can be deployed across multiple availability zones through an active-standby model that supports synchronous replication. In the case of an Availability Zone (AZ, a One Zone storage class) unavailability, failover/failback operations are automated and transparent. The solution follows NetApp ONTAP principles and scales to multiple petabytes in a single namespace.

Amazon EFS is a massively parallel file storage solution that is consumable as-a-service. It provides shared access to thousands of workloads (EC2 instances, ECS, EKS, Fargate, and Lamba), making it particularly suited to handle latency-sensitive workloads with high throughput requirements.

EFS' back-end architecture is completely transparent to customers. The solution can scale up to petabytes and provides a pay-as-you-go model by which organizations are charged for the capacity used. File systems can be hosted within a single AZ or across multiple AZs when applications require multi-zone resiliency. Regardless of the hosting model, data is stored redundantly with 11 9s of durability.

Besides the zones, two storage classes exist: EFS Standard and EFS Standard-IA (infrequent access). Data is available through a single storage namespace and is moved transparently between standard and infrequent access storage classes based on file usage patterns, an automatic tiering feature that reduces storage costs for infrequently accessed files. Centralized data protection capabilities are available with AWS Backup, which allows scheduling of backup jobs as well as defining retention

policies. Data mobility is supported with AWS DataSync, which allows the transfer of files among separate EFS filesystems, AWS S3 buckets, or other locations or protocols, whether in the cloud or on-premises.

Organizations seeking data management capabilities will find that Amazon file services offer a rich set of connections through comprehensive REST APIs and integrations with other AWS services as mentioned above, allowing those file services to serve traditional workloads, containers, and serverless functions.

Analytics and monitoring are handled through several endpoints: The AWS console provides a general overview, with specialized consoles such as the Amazon EFS console covering several metrics about current usage, the number of mount targets, and lifecycle state. The CloudWatch home page provides a more classical overview of metrics, health, and alarms. Automated monitoring can be achieved using CloudWatch Alarms, CloudWatch Logs, and CloudWatch Events, while AWS CloudTrail Log Monitoring provides extra log monitoring options.

Although Amazon file services incorporate standard security features such as in-flight and at-rest encryption, this doesn't include any advanced security features as per the definition in the GigaOm Key Criteria for Cloud File Storage, such as ransomware protection or suspicious user activity monitoring. The AWS platform provides tools to implement such protections, but they must be implemented manually.

**Strengths:** Amazon offers a broad set of cloud file storage solutions. Among these, EFS combines high performance and the simplicity of cloud consumption with solid data availability and an appealing data tiering feature. The solution delivers value thanks to a well-conceived, seamlessly scalable architecture that supports a broad variety of workloads. The availability of FSx for NetApp ONTAP is also a major plus and should not be underestimated because it gives Amazon a serious boost in terms of capabilities and credibility among large enterprises accustomed to NetApp solutions.

**Challenges:** As with most public cloud offerings, analytics and management capabilities are scattered across various standalone products, making it difficult for administrators to configure and monitor the solution seamlessly. Advanced security capabilities require manual tuning, and log monitoring can be complex to configure. Finally, the variety of solutions dilutes the value proposition, unnecessarily increases complexity, and forces organizations to scatter their workloads across various platforms.

## DDN Exascaler

DDN's Exascaler delivers a parallel file system that provides performance, scalability, reliability, and simplicity. With it, DDN offers a data platform that is capable of enabling and accelerating a wide range of data-intensive workflows at scale.

DDN Exascaler was created with the fast and scalable open parallel file system, Lustre, which is the most popular file system for scale-out computing and has been proven and hardened in the most

demanding HPC environments. Lustre and EXAScaler continue to be developed by a very active, dedicated, and talented team, most of whom now work at DDN.

The DDN EXAScaler appliances combine the parallel file system software with a fast hyperconverged data storage platform in a package that's easy to deploy, and is managed and backed by the leaders in data at scale. Built with AI and HPC workloads in mind, DDN excels in GPU integration, with the first GPU-direct integration. The EXAScaler client is deployed into the GPU node, enabling RDMA access as well as monitoring of application access patterns from the GPU client all the way to the disk, providing outstanding levels of visibility of workloads. DDN is also the only certified/supported storage for NVIDIA DGX SuperPOD, a feature that allows DDN customers to run the solution as a hosted AI cloud.

DDN Exascaler's fast parallel architecture enables scalability and performance, supporting low-latency workloads and high-bandwidth applications such as GPU-based workloads, AI frameworks, and Kubernetes-based applications. Moreover, the DDN EXAScaler solution can grow with your data at scale and its intelligent management tools manage data across tiers.

Data security is built in, with secure multi-tenancy, encryption, end-to-end data protection, and replication services baked into the product and providing a well-balanced solution to the customer. In addition, Lustre's capabilities around changelog data and audit logs are built into the Exascaler product, providing better insights for customers into their data. Unfortunately, ransomware protection is not yet completely incorporated into the solution.

Besides the physical EXA5 appliance, a cloud-based solution branded EXAScaler Cloud runs natively on AWS, Azure, and GCP, and can be obtained easily from each cloud provider's marketplace. Features such as cloud sync enable multi-cloud and hybrid data management capabilities within EXAScaler for archive, data protection, and bursting of cloud workloads.

Also worth a mention is DDN DataFlow, a data management platform that is tightly integrated with EXAScaler. Although it's a separate product, a broad majority of DDN users rely on DataFlow for platform migration, archive and data protection use cases, and data movement across cloud, repatriation, etc.

**Strengths:** DDN Exascaler is built on top of the Lustre parallel file system and offers a scalable and performant solution that gives its customers a secure and flexible system with multi-tenancy, encryption, replication, and more. The solution particularly shines thanks to its outstanding GPU integration capabilities, an area where DDN is recognized as a leader.

**Challenges:** Ransomware protection capabilities are missing currently.

## Google Filestore

Available on the GCP public cloud, Filestore is a fully managed NAS solution for the Google Compute

Engine and GKE-powered Kubernetes instances. The solution focuses on high-performance workloads, scales up to hundreds of TBs, and is available on three performance tiers (Basic HDD, Basic SSD, and High Scale SSD), each with different capacity, throughput, and IOPS characteristics.

The solution is native to the Google Cloud environment and is therefore not available on-premises or on other cloud platforms. It doesn't provide a global namespace; instead, customers get one namespace of up to 100TB per share.

Filestore has an incremental backup capability that was launched in beta in September 2020, and currently provides the ability to create backups within or across regions. Backups are globally addressable, allowing restores in any GCP region. As of this report's publication date, the capability is still in beta, so organizations willing to rely on this feature should expect very limited options in terms of both feature set and manageability. Google currently recommends organizations leverage ecosystem partners for enterprise-grade data protection capabilities. Data mobility capabilities are limited to the use of command-line tools such as rsync or scp, and these tools can be used also to copy data to cloud storage buckets, Google's object storage solution.

Filestore includes a set of REST APIs that can be used for data management activities. Data analytics provide basic metrics and the ability to configure alerts.

The solution implements industry-standard security features, but there are no capabilities for auditing user activities (except manually parsing logs) or for protecting against ransomware. Organizations can, however, create Google Cloud storage buckets with the Bucket Lock functionality, and use data mobility tools to copy data to the object store.

**Strengths:** Google Filestore is an interesting solution for organizations that rely heavily on the Google Cloud Platform. It provides a native experience with high throughput and sustained performance for latency-sensitive workloads.

**Challenges:** The solution lacks maturity in many areas, provides limited scalability, and requires further development to be seriously considered for production use.

## Hammerspace

Hammerspace's global file system helps overcome the siloed nature of hybrid cloud file storage by providing a single file system regardless of a site's geographic location or whether storage is on-premises or cloud-based, and by separating the control plane (metadata) from the data plane (where data actually resides).

The solution lets customers utilize, access, store, protect, and move data around the world through a single global namespace, and the user has no need to know where the resources are physically located. The product revolves around the astute use of metadata across file system standards and includes telemetry data (such as IOPS, throughput, and latency), as well as user-defined and analytics-

harvested metadata that allow users to rapidly view, filter, and search the metadata in place instead of relying on file names.

Hammerspace can be deployed on-premises or to the cloud, with support for AWS, Azure, and GCP. It implements share-level snapshots as well as comprehensive replication capabilities, allowing files to be replicated automatically across different sites through the Hammerspace Policy Engine. Manual replication activities are available on-demand as well. These capabilities allow organizations to implement multi-site, active-active disaster recovery with automated failover and failback. Integration with object storage is also a core capability of Hammerspace because data not only can be replicated or saved to the cloud, but also be automatically tiered on object storage, thus reducing the on-premises data footprint and leveraging cloud economics to keep storage spend under control.

One of Hammerspace's key features is its Autonomic Data Management component. This is a machine learning engine that runs a continuous market economy simulation that, when combined with telemetry data from a customer's environment, helps make real-time, cross-cloud data placement decisions based on performance and cost. Although Hammerspace categorizes this feature as data management, in the context of the GigaOm Key Criteria for Cloud File Storage, this capability is more related to the key criteria for Hybrid and Multi-Cloud and Integration with Object Storage.

Ransomware protection is offered by Hammerspace through immutable file shares with global snapshot capabilities as well as an undelete function and file versioning, allowing users to revert back to a file version not affected by ransomware-related data corruption.

Organizations evaluating HammerSpace should take into account the impact of the global snapshot feature on replication intervals when scaling up the number of Hammerspace sites. With eight sites, Hammerspace performs at its best with replication intervals of a few seconds, but with 24 or 32 sites, this interval will go to 20+ seconds with even bigger nodes.

**Strengths:** Hammerspace's Distributed Global File System offers a very balanced set of capabilities with replication and multi- and hybrid cloud capabilities through the power of metadata.

**Challenges:** For widely distributed organizations, the global snapshot feature can have a detrimental impact on replication intervals. This can pose a challenge to organizations with an elevated data change rate.

## IBM Spectrum Scale

IBM provides cloud file storage via a robust and proven software-defined storage solution: IBM Spectrum Scale, a high-performance offering based on IBM's global parallel file system (GPFS).

Two key features of IBM Spectrum Scale are its scalability and flexible architecture. The product can handle several building blocks on the back end: IBM NVMe flash nodes, Red Hat OpenShift nodes, capacity, object storage, and multi-vendor NFS nodes. The solution presents several file interfaces,

such as SMB, NFS, POSIX-compliant, and HDFS (Hadoop), as well as an S3-compatible object interface, making it a versatile choice for environments with multiple types of workloads. Data placement is taken care of by the IBM Spectrum Scale clients, which spread the load across storage nodes in a cluster.

The solution offers a single namespace that is manageable, and migration policies that enable transparent data movement across storage pools without impacting the user experience.

IBM Spectrum Scale supports remote sites and offers various data caching options as well as snapshot support and multisite replication capabilities. The solution includes policy-driven storage management features that allow organizations to automate data placement on the various building blocks based on the characteristics of the data and the cost of the underlying storage, including a feature called Transparent Cloud Tiering that allows users to tier files to cloud object storage with an efficient replication mechanism.

The solution includes a management interface that provides monitoring capabilities for tracking data usage profiles and patterns. Comprehensive data management capabilities are provided through an additional service, IBM Watson Data Discovery.

The latest release of IBM Spectrum Scale includes file audit logging capabilities to track user access across all protocols and platforms, a key security requirement for modern cloud file systems. The solution currently offers no built-in ransomware protection capabilities, although other IBM or third-party products can integrate with Spectrum Scale.

The solution continues to be popular within the HPC community, and IBM also positions Spectrum Scale as an optimized solution for AI use cases. Operating IBM Spectrum Scale for edge use cases should be possible before the end of 2021 when IBM Spectrum Fusion, a containerized version of Spectrum Scale (which should also be consumable in an HCI deployment model) should be available.

**Strengths:** While Spectrum Scale has been around for a long time, the solution's solid architecture and mature enterprise capabilities mean it is still a relevant choice when it comes to cloud file systems. The solution offers multiple enterprise-grade capabilities and will cater to organizations looking to support diverse storage needs in a unified high-performance platform. The product has excellent multi-platform capabilities that extend beyond x86 architectures.

**Challenges:** IBM Spectrum Scale may appear complex and intimidating to small organizations looking for a simple cloud file system solution. Though IBM Watson Data Discovery provides excellent data management capabilities, it is an add-on solution that incurs an extra charge. Advanced analytics capabilities are lacking and must be developed.

# Microsoft Azure

Microsoft offers a number of SaaS-based cloud file storage solutions through its Azure Storage

portfolio. Some of these solutions, such as Azure File Sync, can also be installed on-premises or be cloud hosted (deployed in the cloud and managed by the customer).

The solution portfolio is split into multiple tiers based on performance, cost, and use case. The first solution, Azure Blob, provides file-based access (REST, NFS 3.0, and HDFS via the ABFS driver for Big Data analytics) to an object storage back end with a focus on large, read-heavy sequential access workloads such as large-scale analytics data, backing up and archiving, media rendering, genomic sequencing, and so forth. This solution offers the lowest storage cost among Microsoft's cloud file storage solutions, with performance of up to 20K IOPS per volume and several performance tiers. Blob also offers high-throughput block blobs, which provide better performance for workloads that require very high throughput with single blobs. On request, IOPS can be increased to up to 100K.

The second solution, Azure Files, uses the same hardware as Azure Blob but implements full POSIX filesystem support with the NFS 4.1 protocol (as well as REST, API and SMB). The solution is oriented more toward random access workloads, ideally with in-place data updates. Slightly pricier, this solution achieves up to 100K IOPS per volume and offers several performance tiers (cool, hot, transaction optimized, premium).

Third in the portfolio, Azure NetApp Files consists of a fully managed service that uses NetApp's hardware-based on ONTAP and is fully integrated in the Microsoft Azure cloud. This solution offers the highest performance of all, with up to 460K IOPS per volume, and is the most expensive of the three offerings. Like Azure Blob, it also comes with several performance tiers. Azure Blob and Azure Files are broadly available, while Azure NetApp Files is available in selected regions.

Global namespaces are supported with Azure File Sync through the use of DFS Namespaces, but there is no global namespace capability available to federate the various solutions and tiers offered across the Azure cloud file storage portfolio.

Besides Azure File Sync, Azure offers a variety of data replication and redundancy options. Redundancy can be set up locally or at the availability zone level. Geo-redundancy is possible either within one zone or across multiple zones, although currently only shares under 5TB are supported with Azure Files. Backup and restore across multiple regions is also available with Azure Backup, but users need to take into account possible additional costs when restoring across regions. Organizations looking at Azure Backup should consider the available schedules, as the lowest granularity level is daily backups.

Putting aside Azure NetApp Files, which relies on ONTAP, both Azure Blob and Azure Files are based on an object storage back end. Multi-cloud capabilities are non-existent as Microsoft owns the entire ecosystem, but this fact is counterbalanced by the many zones and regions available. Several storage tiers are offered for Azure Blob and Azure Files: Premium (SSD-based, I/O-intensive workloads, low latency), Hot (general purpose scenarios, HDD-based), Cool (cost efficient, online archive, based on HDD as well), and Archive storage (ultra-low cost store). Automated tiering capabilities are partially present across the Azure Blob and Azure Files offerings: Azure Blob offers lifecycle management and a policy-based automated tiering solution, while the on-premises Azure File Sync solution enables

offloading of on-premises files to the cloud.

Because it is based on Microsoft Azure, the storage portfolio offers rich integration capabilities through APIs for data management purposes. Observability and analytics are possible via the Azure Monitor single pane of glass management interface, which also incorporates Azure Monitor Storage Insights. Storage Insights allows the organization to view macro-level information around storage usage at scale, but it also lets users drill down into particular storage accounts for in-depth metrics (such as latency or transactions per second) or to diagnose issues. A global view of capacity usage and detailed logs are also offered.

Azure Files services provides incremental read-only backups as a way to protect against ransomware. Up to 200 snapshots per share are supported, with up to 10 years retention. Those read-only snapshots can be taken either manually or through the use of Azure Backup. The capability comes with a soft delete feature that acts as a recycle bin and allows the recovery of deleted files within a certain timeframe.

The on-premises Azure File Sync solution can be used for edge deployments, although in this case, edge would refer primarily to remote office/branch office use cases. It is also common to deploy it in various clouds to act as a local cache for distributed cloud environments. For example, primary data could reside in the U.S. while Azure File Sync would be used in Europe.

**Strengths:** Microsoft offers a broad portfolio with multiple options, protocols, use cases and performance tiers that allows organizations to consume cloud file storage in a cost-efficient manner.

**Challenges:** There are no global namespace management capabilities to abstract the underlying file share complexity for the end user. Some features, such as automatic tiering, are partially implemented and would benefit from being present across all offerings. There are also limitations based on the different share types. The various offerings can appear to be too complex and therefore intimidating for smaller organizations.

## Nasuni

Nasuni offers a SaaS solution for enterprise file services, with an object-based global file system as its main engine and with many familiar file interfaces, including SMB and NFS. It is integrated with all major cloud providers and works with on-premises S3-compatible object stores. Many Nasuni customers implement the solution to replace traditional NAS systems, and its characteristics enable users to replace several additional infrastructure components as well, such as backup, archiving platforms, and more. Recently the company added features for improved ransomware recovery and advanced data management, making the solution even more compelling to enterprise users.

Nasuni offers a global file system called UniFS, which provides a layer that separates files from storage resources, managing one master copy of data in public or private cloud object storage while distributing data access. The global file system manages all metadata, such as versioning, access control, audit records, and locking, and provides access to files via standard protocols such as SMB

and NFS. Files in active use are cached using Nasuni's Edge Appliances, so users benefit from high performance access through existing drive mappings and share points. All files, including files in use across multiple local caches, have their master copies stored in cloud object storage so they are globally accessible from any access point.

The Nasuni Management Console delivers centralized management of the global edge appliances, volumes, snapshots, recoveries, protocols, shares, and more. The web-based interface can be used for point-and-click configuration, but Nasuni also offers a REST API method for automated monitoring, provisioning, and reporting across any number of sites. In addition, the Nasuni Health Monitor reports to the Nasuni Management Console on the health of the CPU, directory services, disk, file system, memory, NFS, network, SMB, services, and so on. Nasuni also integrates with tools like Grafana and Splunk for further analytics.

Nasuni's security capabilities allow customers to configure file system auditing and logging for operations on volumes. Syslog Export enables notifications and auditing messages to be sent to syslog servers. Auditing volume events such as Create, Delete, Rename, and Security aid customers in identifying and recovering from ransomware attacks. As an alternative to its own auditing options, Nasuni also supports auditing by Varonis. Nasuni protects files efficiently and quickly, and has the ability to dial back to a pre-ransomware state in seconds or minutes. Nasuni Continuous File Versioning allows a nearly infinite number of snapshots as immutable copies, so customers can easily recover in minutes to any point in time to quickly mitigate a ransomware attack.

Nasuni Edge Appliances are lightweight VMs or hardware appliances that cache frequently accessed files wherever SMB or NFS access is needed with performance as well. They can be deployed on-premises or in the cloud to replace legacy file servers and NAS devices. The Nasuni Edge Appliances encrypt and dedupe files, then snapshot them at frequent intervals to the cloud where they are written to object storage in read-only format.

**Strengths:** Nasuni offers a great and efficient distributed file system solution to its customers that is secure and scalable. The solution offers protection against ransomware at a very fine level, and with the edge appliances, customers can access their frequently used data in a fast and secure way.

**Challenges:** The solution focuses primarily on distributed data and data availability, whereas cloud file systems are tuned primarily to deliver high performance, high throughput, and low latency to performance-oriented workloads.

## NetApp

NetApp has an interesting approach to cloud file systems based on its Cloud Volumes platform. This solution offers many deployment choices to organizations. Besides on-premises deployments, NetApp has forged unique partnerships with the three major public cloud providers to offer a native NetApp experience that is tightly integrated with the public cloud platform, not only from a performance and technical integration perspective, but also with regard to charging and management. This is particularly the case with Azure NetApp Files and now Amazon FSx for NetApp

ONTAP. These are Tier 1 cloud file services directly offered and managed by Azure and AWS that provide a seamless experience to the user, regardless of the cloud platform they use.

Cloud Volumes implements a global namespace that abstracts multiple deployments and locations regardless of distance. Several intelligent caching mechanisms combined with global file locking capabilities enable a seamless, latency-free experience that makes data accessible at local access speeds from local cache instances.

Based on ONTAP, Cloud Volumes has been architected to support hybrid deployments natively, whether on-premises or in the cloud. Cloud Volumes abstracts the underlying cloud infrastructure and deployment model to present users with a single control plane, and the management of all systems is unified under Cloud Manager, NetApp's management console. Because ONTAP is at the heart of Cloud Volumes, all of the data services provided on-premises or in the cloud are the same: discovery, deployment, protection, governance, migration and tiering. Tiering, replication, and data mobility capabilities are outstanding and enable a seamless, fully hybrid experience that lets organizations decide where primary data resides, where infrequently accessed data gets tiered to, and where data copies and backups used for disaster recovery should be replicated to, notably thanks to NetApp's Cloud Backup service.

Integration with object storage is a key part of the solution, and policy-based data placement allows automated, transparent data tiering on-premises with NetApp StorageGRID, or in the cloud with AWS S3, Azure Blob Storage, or Google Cloud Storage, with the ability to recall requested files from the object tier. Object storage integration also extends to backup and DR use cases: With Cloud Backup, backup data can be written to object stores using block-level, incremental-forever technology.

Data management capabilities are enabled by consistent APIs that allow data copies to be created as needed; the platform also offers strong data analytics capabilities through Cloud Manager (which has integrated dashboards fed by NetApp's Cloud Insights service), and particularly through Cloud Data Sense, one of Cloud Manager's accessible services. This service provides insights around data owners, location, access frequency, and data privileges, as well as potential access vulnerabilities, with manual or automated policy-based actions. Organizations have the ability to generate compliance and audit reports such as DSARs; HIPAA and GDPR regulatory reports also can be run in real time on all Cloud Volumes data stores.

NetApp provides advanced security measures against ransomware and suspicious user or file activities through NetApp FPolicy and snapshot capabilities. The FPolicy solution is integrated into Cloud Volumes ONTAP and provides prevention abilities. It allows file operations to be monitored and blocked as a preventive measure. It includes detection of common ransomware file extensions as well as integration capabilities with third-party technology partners such as Varonis, Veritas, and others. Visibility of FPolicy activities is possible through NetApp Cloud Insights, which analyzes user behavior to identify file access anomalies to preempt possible risks from outsiders, ransomware attacks, or rogue users.

From a remediation perspective, immutable point-in-time NetApp snapshot copies provide the ability

to revert to a healthy state. Organizations can also enable Cloud WORM, an additional write-once, read-many capability, when they create new Cloud Volumes ONTAP instances. This feature is powered by NetApp SnapLock and provides long-term snapshot retention that can be used not only for ransomware protection, but also for regulatory and compliance purposes.

The solution supports flexible deployment models that also take into consideration edge use cases. From Cloud Manager, customers can enable the Global File Cache service for branch locations, remote sites, or regional hyperscalers' points of presence, to enable local-speed, low-latency access to centralized shares through a single global namespace with full global file locking capabilities.

Cloud Volumes is available as-a-service on AWS, Azure, and GCP.

**Strengths:** NetApp Cloud Volumes offers a complete enterprise-grade feature set and a continuous data management plane, both of which are complemented by comprehensive and flexible deployment models. Monitoring and management are intuitive and simple to use; analytics and reporting capabilities are tailored to cope with regulatory demands; and advanced security capabilities help mitigate today's ransomware challenges. The addition of ONTAP as an AWS managed service through Amazon FSx for NetApp ONTAP puts NetApp in a position to deliver a stronger multi-cloud experience to its customers.

**Challenges:** Although not necessarily a challenge, NetApp's offering and ecosystem are very rich and comprehensive. Without proper guidance, some organizations might feel intimidated.

# ObjectiveFS

ObjectiveFS is a cloud file storage platform that supports on-premises, hybrid, and cloud-based deployments. Its POSIX filesystem can be accessed as one or many directories by clients, and it uses an object store on the back end. Data is written directly to the object store without any intermediate servers. ObjectiveFS runs locally on servers through client-side software, providing local-disk-speed performance. The solution scales simply and without disruption by adding ObjectiveFS nodes to an existing environment. The solution is massively scalable to thousands of servers and petabytes of storage.

ObjectiveFS offers a global namespace where all of the updates are synchronized through the object store back end. The solution supports cloud-based or on-premises S3-compatible and Microsoft Azure object stores. It uses its own log-structured implementation to write data to the object store back end by bundling many small writes together into a single object. The same technique then can be used for read operations by accessing only the relevant portion of the object. The solution implementation also uses a method called compaction, which bundles metadata and data into a single object for faster access. Storage-class-aware support ensures that policies can be used to implement intelligent data tiering and move data across tiers based on usage. To ensure performance requirements are met, ObjectiveFS offers several levels of caching that can be used together.

Users can deploy the solution across multiple locations (multi-region and multi-cloud). The solution

delivers flexible deployment choices, allowing storage and compute to run in different locations and across different clouds.

The solution currently offers no data management capabilities and relies on third-party integrations to analyze, reprocess, or augment data. Built-in analytics provide latency heatmaps as well as performance-tuning information, and the solution also logs access to data, including access to data types, access requests, and cache hit rate.

ObjectiveFS provides comprehensive security features, such as data in-flight and at-rest encryption. The solution supports multi-tenancy, and therefore, data is encrypted using separate encryption keys, making it accessible only to the tenant that owns the data.

One of the most interesting features of ObjectiveFS is the inclusion of a "workload adaptive heuristics" mechanism that supports hundreds of millions of files and tunes the file system to ensure consistent performance is delivered regardless of the I/O activity profile (read vs. write, sequential vs. random) and the size of the files, handling many small files or large terabyte-sized files at the same performance levels.

**Strengths:** ObjectiveFS provides a highly scalable and robust solution at consistent performance levels regardless of the data type. It delivers flexible deployment options and comes with excellent security and multi-tenancy features.

**Challenges:** The solution currently lacks any kind of data management capabilities.

## Oracle Cloud

Oracle provides cloud file system options via three offerings: Oracle Cloud Infrastructure File Storage, Oracle HPC File System stacks, and Oracle ZFS.

File Storage is the cloud file system solution developed by Oracle on its Oracle Cloud Infrastructure (OCI) platform. Delivered as-a-service, the solution provides an automatically scalable, fully managed elastic file system. Up to 100 file systems can be created in each availability domain, and each of those file systems can grow up to 8 exabytes. Optimized for parallelized workloads, the solution also focuses on high availability and data durability, with 5-way replication across different fault domains. The solution implements an interesting mechanism named "eventual overwrite" for data deletion. Each file is created in the file system with its own encryption key. When a file is deleted, the key is destroyed and the file becomes inaccessible. The same mechanism is used for an entire file system deletion. Periodically, inaccessible files and file systems are purged to free space and eradicate residual data.

Oracle File Storage supports snapshots as well as clones. The clone feature allows a filesystem to be made available instantaneously for read and write access while inheriting snapshots from the original source, making copies immediately available for test and development use cases, allowing

organizations to significantly reduce the time needed to create copies of their production environment for validation purposes. No backup feature currently exists, although third-party tools can be used to copy data across OCI domains, regions, OCI Object Storage, or on-premises storage.

Data management capabilities primarily reside in the use of REST APIs. These can be combined with the clone feature to automate fast copy provisioning operations that execute workloads on copies of the primary data sets. A management console provides an overview of existing file systems and provides usage and metering information, both at the file system and mount target levels. Administrators also get a view of the system health with performance metrics, and can configure alarms and notifications in the general monitoring interface for Oracle Cloud Infrastructure.

Currently, ransomware protection and user activity monitoring capabilities are not available on Oracle File Storage. Through Oracle's Dedicated Region offering, all OCI services can be deployed on the customer premises. Connectivity to Oracle Cloud is made possible through a FastConnect private connection and the OCI-Azure Interconnect. Oracle also offers a no-cost storage gateway for basic file interactions on-premises, with back-end connectivity to OCI Object storage.

OCI HPC File System stacks is an OCI offering dedicated to high-performance computing workloads in an organization that require the use of traditional HPC parallel file systems such as BeeGFS, IBM Spectrum Scale, GlusterFS, or Lustre. An open source license is used when the file system supports it; otherwise, the customer has to cover licensing costs through a bring-your-own-license (BYOL) model. OCI allows those HPC File System stacks to be deployed via a wizard and includes Terraform automation support. Although the feature set depends on the file system used, OCI can provide up to 500GB/s of bandwidth on this offering.

Organizations also can opt for the Oracle ZFS image option, a marketplace image that can be configured as bare metal or a VM and supports ZFS. Currently, only single node deployments exist, but clustering options are coming in the near future. Each image can scale to 960TB, providing support for NFS and SMB with AD integration. The solution fully supports replication, snapshots, clones, and cloud snapshots, with several DR options. Oracle also provides sizing options to select the optimal image based on the expected number of clients. This service operates under a BYOL model to which the organization adds the cost of compute and block storage.

All three offerings include encryption and a key management system that also supports multi-tenant key management.

**Strengths:** Oracle Cloud Infrastructure offers an attractive palette of cloud file services, starting with Oracle File Storage. An interesting platform built for high availability, durability, and with excellent scalability, Oracle File Storage also includes useful cloning features that will appeal to organizations working with massive data sets. OCI HPC File System stacks as well as the Oracle ZFS image offer industry-proven alternatives to organizations seeking to shift on-premises file-based workloads to the cloud, notably in the HPC space.

**Challenges:** The offerings are currently limited in several areas where improvements would be

welcome, notably around advanced data protection features and monitoring capabilities.

## Panzura

Panzura offers a cloud file system, CloudFS. The solution works across sites (public and private clouds) and provides a single data plane with local file operation performance, automated file locking, and immediate global data consistency.

The solution implements a global namespace and tackles data integrity requirements through a global file-locking mechanism that provides real-time data consistency regardless of where a file is accessed from around the world. It also provides efficient snapshot management with version control and allows administrators to configure retention policies as needed. Besides the high-performance characteristics of the solution, backup and disaster recovery capabilities are offered as well.

Panzura relies on S3 object stores and supports a broad range of object storage solutions, whether in the public cloud (AWS S3, Azure Blob) or on-premises, with Cloudian or IBM Cloud Object Storage (COS). A feature called Cloud Mirroring allows multi-back-end capabilities by writing data to a second cloud storage provider to ensure data is always available, even if a failure occurs at one of the cloud storage providers. Tiering and archiving also are implemented in Panzura.

Analytics capabilities are offered through Panzura Data Services, a set of advanced features that provide global search, user auditing, one-click file restoration, and monitoring functions aimed at core metrics and storage consumption, showing, for example, frequency of access, active users, and the health of the environment. For data management, Panzura provides various API services that allow users to connect their data management tools to Panzura. Panzura Data Services also allows the detection of infrequently accessed data so that subsequent action can be taken.

Security capabilities include user auditing (through Panzura Data Services) as well as ransomware protection. Ransomware protection is handled with a combination of immutable data (a WORM S3 backend) and read-only snapshots taken every 60 seconds at the global filer level, regularly moving data to the immutable object store, and allowing seamless data recovery in case of a ransomware attack through the same mechanism an organization would use to recover data under normal circumstances (backup). Key management for encryption (at rest and in flight) occurs through a bring-your-own key management system model. The solution also includes a Secure Erase feature that removes all versions of a deleted file and subsequently overwrites the deleted data with zeros, a feature available even with cloud-based object storage.

**Strengths:** Panzura provides a cloud file system that offers local-access performance levels with global availability, data consistency, tiered storage, and multi-backend capabilities. Panzura Data Services delivers advanced analytics and data management capabilities that help organizations better understand and manage their data footprint.

**Challenges:** The Panzura solution has been architected primarily as a distributed cloud file storage solution and is also good for general purpose enterprise workloads. While it technically meets the

requirements for a cloud file system and provides local-grade performance, the solution cannot meet demanding high performance workload requirements for which high throughput, high IOPS, and ultra low latency are essential.

# Qumulo

Qumulo has developed a software-based, vendor-agnostic cloud file system that can be deployed on-premises, in the cloud, or even delivered through hardware vendor partnerships. The solution provides a comprehensive set of enterprise-grade data services branded Qumulo Core. These handle core storage operations (scalability, performance) as well as data replication and mobility, security, ransomware protection, data integration, and analytics.

Qumulo supports hybrid and cloud-based deployments. Cloud services are delivered through Cloud Q, a set of solutions designed specifically for the cloud that leverage Qumulo Core services. Organizations can either deploy Cloud Q through their preferred public cloud marketplace (the solution supports AWS, Azure, and GCP) or choose to deploy Qumulo as a fully managed SaaS offering on Microsoft Azure. AWS Outposts is also supported, and a comprehensive partnership with AWS is also in place (WAF certification, AWS Quick Start, and so on). Qumulo is expanding its delivery models as well through storage-as-a-service partnerships with HPE GreenLake and others.

The solution scales linearly, both from a performance and capacity perspective, providing a single namespace with limitless capacity that supports billions of large and small files and provides the ability to use nearly 100% of usable storage through efficient erasure code techniques. It also supports automatic data rebalancing when nodes or instances are added. The namespace allows for real-time queries and aggregation of metadata, greatly reducing search times.

Data protection and replication, as well as mobility use cases, are well covered and include snapshots and snapshot-based replication to the cloud, continuous replication, and disaster recovery support with failover capabilities. Qumulo SHIFT is a built-in data service that moves data to AWS S3 object stores with built-in replication, including support for immutable snapshots. Although currently in technical preview, Qumulo will support the ability to shift data back from S3, which allows bidirectional data movements to and from S3 object storage, providing organizations with more flexibility and better cost control.

Qumulo includes a comprehensive set of REST APIs that can be used not only to perform proactive management but also to automate file system operations. The solution comes with a powerful data analytics engine that provides real-time operational analytics (across all files, directories, metrics, users, and workloads), capacity awareness, and predictive capacity trends, with the ability to "time travel" through performance data.

Advanced security features include ransomware protection snapshots, and immutable snapshots replicated to the cloud, as well as audit logging to review user activity.

**Strengths:** Qumulo offers a comprehensive cloud file storage solution that is simple to manage and

implement. It has a rich and complete data services set combined with a broad choice of deployment models, making it one of the most compelling cloud file storage offerings currently available.

**Challenges:** Although the solution is very complete, some important features are currently still on the roadmap. Among these, data reduction improvements and multi-tenancy should be mentioned.

# Scality SOFS

Scality offers a cloud file system solution called Scality SOFS (Scale Out File System) that is built on top of Scality RING and leverages Microsoft Azure. SOFS consists of a POSIX layer integrated into an object-based storage architecture that uses a database (Azure CosmosDB) for a POSIX representation of objects (metadata) into the object store (Azure Blob Storage). SOFS is deployed via stateless VM images for Azure Cloud, and the solution can be scaled as necessary by adding more VMs. This solution offers high aggregate throughput via the SMB, NFS, and FUSE protocols, with hundreds to thousands of Mbps of throughput per interface, and is positioned for throughput and sequential access workloads.

Scality SOFS is already available as an on-premises offering through Scality RING, but the cloud-based solution is built to run specifically on Azure. Each Scality RING on which Scality SOFS runs offers a global namespace, making global metadata search possible within that RING.

Object storage is at the heart of Scality SOFS, as the solution relies on Azure Blob Storage as its storage backend. Although only Azure Blob Storage is used, this cloud object storage solution has multiple tiers, and Scality SOFS is able to take advantage of the various tiers to optimize the balance between performance and cost.

Management is done through the RING Supervisor GUI, which provides standard analytics capabilities. This GUI also manages SOFS volumes and connectors, and provides insights into the infrastructure as well as utilization metrics that can be used for capacity planning (although an integrated capacity planning function doesn't exist).

Protection against ransomware is implemented through immutable (WORM) volumes, a feature that is configurable at the volume level.

Organizations considering Scality SOFS should take into account the specificity of the solution and its focus on high throughput and sequential workloads. Use cases that are ideal for Scality SOFS include data lakes for massive log retention, long-term data retention of medical imaging data, and as HPC storage systems, all solutions for which scale and aggregate throughput or sequential access are essential.

**Strengths:** A robust cloud file system solution specifically designed for sequential and throughput-intensive workloads, Scality SOFS takes advantage of inexpensive object storage to deliver massive scalability, parallelism, and aggregated throughput.

**Challenges:** Data management and data analytics capabilities are nonexistent. The cloud edition of Scality SOFS currently supports only Microsoft Azure.

# Weka.io

Weka.io's WekaFS cloud file system solution offers a single data platform with mixed workload capabilities and multi-protocol support (SMB, NFS, S3, POSIX, GPU Direct, and the Kubernetes CSI).

WekaFS is deployed as a set of containers and offers multiple deployment options on-premises (bare metal, containerized, virtual) as well in the cloud. It can run either fully on-prem, in hybrid mode, or solely in the cloud, and all deployments can be managed through a single management console.

The solution implements a global namespace that spans across locations (on-prem, cloud) and performance tiers, with automatic scalability in the cloud, presenting users with a unified namespace that abstracts the underlying complexity and enables transparent background data movements between tiers.

The global namespace natively supports and expands into S3 object storage bidirectionally thanks to dynamic data tiering, which automatically pushes data to the object tier when capacity runs low on the NVMe flash tier. Both tiers (flash-based and object) can scale independently. A feature called Snap-To-Object allows data and metadata to be committed to snapshots for backup, archive, and asynchronous mirroring. This feature can be used also for cloud-only use cases in AWS, such as pausing or restarting a cluster, protecting against single availability zone failure, or migrating file systems across regions.

Data management capabilities primarily include the ability to create copies of the data through snapshots, for example with DevOps use cases where jobs or functions are executed against data copies instead of the primary data set. API integrations (including serverless) are possible through Weka.io APIs. On the analytics side, Weka's monitoring platform captures and provides telemetry data about the environment, with the ability to deep dive into certain metrics all the way down to filesystem calls. Weka also supplies a proactive cloud-based monitoring service called Weka Home that collects telemetry data (events and statistics) and provides proactive support in case of detected irregularities.

The solution supports advanced security capabilities with Weka Snap-To-Object, which also allows data to be protected with immutable object-based snapshots and thus safeguards data against ransomware attacks. Weka Home monitors for anomalies in access patterns and enables the identification of suspicious operations. The solution also supports log forwarding to inhibit tampering by a malicious actor or a rogue administrator.

WekaFS supports edge aggregation deployments, which enable smaller footprint clusters to be deployed alongside embedded IoT devices, for example with autonomous vehicles. Finally, organizations can deploy WekaFS directly on AWS from the marketplace, and also run a certified WekaFS deployment on AWS Outposts, thanks to Weka.io's ISV Accelerate partnership with AWS.

Although versatile, the solution is particularly useful in demanding environments that require low latency, high performance, and cloud scalability, such as AI/ML, life sciences, financial trading, HPC, media rendering and visual effects, as well as electronic design and automation, and engineering DevOps.

**Strengths:** Weka.io has architected a robust and seamlessly scalable high-performance storage solution with comprehensive deployment options, automated tiering, and a rich set of services via a single platform that eliminates the need to copy data through various dedicated storage tiers. Its single namespace encompassing file and object storage reduces infrastructure sprawl and complexity to the benefit of users and organizations alike.

**Challenges:** Weka.io's strong focus on performance and scalability eclipses the growing need for data analysis within organizations. From a key criteria perspective, data management is an area needing improvement, with no current capabilities available.

## Zadara

Zadara Edge Cloud Services is an interesting solution that is aimed primarily at managed service providers and some larger enterprises. The solution is available globally through 300 cloud partners on six continents and consists of an elastic infrastructure layer comprising compute, networking, and storage capabilities for which cost is based on usage. The storage offering, named zStorage, consists of one or more Virtual Private Storage Arrays (VSPAs) that can be deployed on SSD, hybrid, and HDD media types. A VSPA is able to serve block, file, and object storage simultaneously. Various VPSAs can be created, each with its own engine type (which dictates performance) and its own set of drives, including spares.

Global namespaces are supported for file-based storage, with a capacity limit of up to 0.5PB, after which a new namespace must be created.

The solution offers thinly provisioned snapshots as well as cloning capabilities, which can be local or remote. The snapshot-based asynchronous remote mirroring feature makes possible replication to a different pool within the same VPSA, to a different local or remote VSPA, or even to a different cloud provider. The replicated data is encrypted and compressed before being transferred to the destination. The solution also allows for many-to-many relationships, which enables cross-VPSA replication in active-active replication scenarios. Cloning capabilities are also available remotely and can be used for rapid migration of volumes between VPSAs because the data can be made available instantly (although a dependency with the source data remains until all of the data has been copied in the background).

Native backup and restore capabilities leverage object storage integration with AWS S3, Google Cloud Storage, Zadara VPSA Object Storage, and other S3-compatible object stores. Object storage can be used by Zadara for audit and data retention purposes. Zadara supports AWS Direct Connect as well as Azure ExpressRoute, both of which allow a single volume to be made available to workloads residing in multiple public clouds, enabling the use of a single dataset across multiple

locations or clouds. When deployed on flash, zStorage supports an auto-tiering capability that recognizes hot data and places it on the flash/high-performance tier, while less frequently accessed data is tiered either on lower-cost hard disks or S3-compatible object storage.

Zadara File Lifecycle Management services provide data management and analytics capabilities to the solution, including growth trends (overall and by file type), capacity utilization across several metrics, and usage statistics by owners and groups.

Zadara natively supports access auditing for files accessed through NFS and SMB. Auditing data is segregated or made accessible only by administrators and can be uploaded to a remote S3 repository for long-term retention. Although no native ransomware protection capabilities exist, Zadara partners with Veeam to provide such protection through Veeam's Scale-Out Backup Repository immutability features.

Zadara's Federated Edge Program allows MSPs to rapidly deploy Zadara at the edge, enabling MSPs to provision a turnkey infrastructure closer to their customers while adhering to the Zadara Cloud operating model. Zadara provides the necessary hardware and software, and revenues are shared between Zadara and the Federated Edge partners.

Emerging technologies such as data classification, compliance, and data sovereignty are currently absent.

**Strengths:** Zadara Edge Cloud Services delivers comprehensive file storage service capabilities via a platform with rich compute, storage, and networking support. Remote cloning and mirroring capabilities offer a seamless experience complemented by object storage tiering and multi-cloud support. Analytics provide multi-dimensional information about trends, capacity, and user statistics. File auditing functionalities with long-term retention can be useful for legal purposes.

**Challenges:** The 0.5PB capacity limit on namespaces can become an operational hurdle for organizations with many teams working on very large datasets such as cloud-based AI, HPC, and big data workloads.

# 6. Analyst's Take

The cloud file system market is interesting: Many would believe that, thanks to their dominance and massive market share, public cloud providers would provide the most comprehensive solutions. Nothing is further from the truth.

Public cloud providers range from simpler offerings (Oracle, Google) to a broad swath of more complex cloud file system solutions (Azure, Amazon), with some overlaps and niche use cases. The primary concern with these is that, with a few notable exceptions (such as Tier 1 partnerships with vendors such as NetApp), some of these public cloud solutions need additional adjustments and improvements to fit the needs of the enterprise. In the context of the *GigaOm Key Criteria Report for Evaluating File-Based Cloud Storage*, these solutions show gaps in key criteria coverage around management, monitoring, and advanced security implementation. To their credit, public cloud solutions generally offer seamless scalability and straightforward pay-as-you-go consumption options.

In contrast, storage vendors specializing in cloud file systems may have more narrowly focused solutions but with better enterprise-grade capabilities and more complete feature sets. Among these, many solutions can run on public clouds and offer cloud-like consumption models while delivering compelling value and the ability to operate seamlessly using a hybrid cloud model.

As organizations put more emphasis on shifting and redesigning file-based, performance-sensitive workloads to the cloud, demand for cloud file systems will continue to grow. Specialized storage vendors are currently in a better position to meet that demand, thanks to complete feature sets oriented toward enterprise requirements, while public cloud providers lag behind, except those who partner with specialized storage vendors and have a dedicated offering. Public cloud vendors will need to work on both portfolio rationalization and richer feature sets, otherwise the sprawl of offerings will continue to grow, to the detriment of end-user needs.

# 7. About Enrico Signoretti

Enrico Signoretti has more than 25 years in technical product strategy and management roles. He has advised mid-market and large enterprises across numerous industries, and worked with a range of software companies from small ISVs to global providers.

Enrico is an internationally renowned expert on data storage—and a visionary, author, blogger, and speaker on the topic. He has tracked the evolution of the storage industry for years, as a Gigaom Research Analyst, an independent analyst, and as a contributor to the Register.

# 8. About Max Mortillaro

Max Mortillaro is an independent industry analyst with a focus on storage, multi-cloud & hybrid cloud, data management, and data protection.

Max carries over 20 years of experience in the IT industry, having worked for organizations across various verticals such as the French Ministry of Foreign Affairs, HSBC, Dimension Data, and Novartis to cite the most prominent ones. Max remains a technology practitioner at heart and currently provides technological advice and management support, driving the qualification and release to production of new IT infrastructure initiatives in the heavily regulated pharmaceutical sector.

Besides publishing content/research on the TECHunplugged.io blog, Gestalt IT, Amazic World, and other outlets, Max is also regularly participating in podcasts or discussion panels. He has been a long-time Tech Field Day Alumni, former VMUG leader, and active member of the IT infrastructure community. He has also continuously been running his own technology blog kamshin.com since 2008, where his passion for content creation started.

Max is an advocate for online security, privacy, encryption, and digital rights. When not working on projects or creating content, Max loves to spend time with his wife and two sons, either busy cooking delicious meals or trekking/mountain biking.

# 9. About GigaOm

GigaOm provides technical, operational, and business advice for IT's strategic digital enterprise and business initiatives. Enterprise business leaders, CIOs, and technology organizations partner with GigaOm for practical, actionable, strategic, and visionary advice for modernizing and transforming their business. GigaOm's advice empowers enterprises to successfully compete in an increasingly complicated business atmosphere that requires a solid understanding of constantly changing customer demands.

GigaOm works directly with enterprises both inside and outside of the IT organization to apply proven research and methodologies designed to avoid pitfalls and roadblocks while balancing risk and innovation. Research methodologies include but are not limited to adoption and benchmarking surveys, use cases, interviews, ROI/TCO, market landscapes, strategic trends, and technical benchmarks. Our analysts possess 20+ years of experience advising a spectrum of clients from early adopters to mainstream enterprises.

GigaOm's perspective is that of the unbiased enterprise practitioner. Through this perspective, GigaOm connects with engaged and loyal subscribers on a deep and meaningful level.

# 10. Copyright