GIGAOM

MARKET LANDSCAPE REPORT

Key Criteria for Evaluating File-Based Cloud Storage

A critical Component for Multi-Cloud Data Storage Strategy

ENRICO SIGNORETTI

TOPICS: CLOUD CLOUD STORAGE

GIGAOM

Key Criteria for Evaluating File-Based Cloud Storage

A critical Component for Multi-Cloud Data Storage Strategy

TABLE OF CONTENTS

1. Summary

File storage is still a very important part of every enterprise data storage infrastructure, and there is no surprise that now, more than ever, users ask for file services in the cloud too.

Cloud providers neglected to add file services to their product portfolio at the beginning, concentrating on block and object storage. Albeit these two types of storage cover many use cases, and new applications can be developed to use them without needing file storage, the reality is that there are several circumstances where files are preferable:

- **Lift and Shift:** With more and more enterprises opting for the public cloud as their primary IT infrastructure, it is very common now to see "lift and shift" migrations. In this scenario, the users want to replicate the same services they had in the on-premises data center, including (POSIX)-compliant file systems, data services, and all other enterprise features they are accustomed to.
- **Simplicity:** File storage is the most user-friendly storage, and many developers still prefer this familiar interface to others because it gives them the ability to build even more portable applications while simplifying sharing machine- and human-generated data.
- **Performance:** Albeit object storage performance is improving by leaps and bounds, 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 usually offers data services such as snapshots to improve data management operations.

In the last few years, file systems have also become more cloud-friendly, showing better integrations with object storage. This brought several advantages to end users:

- **Better scalability.** Policy-driven tiering mechanisms enable us to move cold data to Amazon's Simple Storage Service (S3)-compatible storage and save precious resources in the high-performance tier.
- **Best combination of speed and $/GB**. Some file storage gateways are specifically designed to work with an object storage back-end.
- **Simplify data migrations and synchronization.** Many file storage systems can replicate data to remote file or object stores, in the cloud or on-premises. This allows us to synchronize and serve data sets across different infrastructures to optimize compute-data vicinity, hence latency.
- **Disaster Recovery (DR)**. Syncing data to a remote object store enables the user to leverage a cheaper storage repository in the cloud and populate a file system only if necessary.

These functionalities are particularly important now that vendors are optimizing their file storage for flash memory and access speed so that the user can build a multi-tier infrastructure to optimize $/GB

as well. Furthermore, users are looking at hybrid and multi-cloud for their infrastructure strategy, and this storage infrastructure design perfectly fits in this context.

With this in mind, it is easy to understand the reasons behind the success of file storage in the cloud, especially if it can be integrated with on-premises and cloud storage as well.

Report Methodology

A Key Criteria Report analyzes the most important features of a technology category to understand how they impact an enterprise and its IT organization. Features are grouped into four categories:

1. Table Stakes
2. Key Criteria
3. Critical Impact of Features on the Metrics
4. Near-term Game-changing Technology

The goal is to help organizations assess capabilities and build a mid-to-long-term infrastructure strategy. In a mature technology, the solutions are divided into three target market categories: enterprise, high-performance, and specialized solutions. In a mature market, these differ in their characteristics and how they can be integrated with existing infrastructures. That said, the assessment is more dependent on the specific user's needs and not solely on the organization's vertical.

Table Stakes

Table stakes are system characteristics and features that are important when choosing the right solution. They include architectural choices that depend on the size of the organization, its requirements, the expected growth over time, and the types of workloads. Table stakes are mature, and the implementation of these features will not add any business advantage nor significantly change the total cost of ownership (TCO) or return on investment (ROI) of the infrastructure.

Key Criteria

Key criteria features really differentiate one solution from another. Depending on real user needs, they have a positive impact on one or more of the metrics mentioned. Therefore, implementation details are essential to understanding the benefits relative to the infrastructure, processes, or business. Following table stakes and key criteria, aspects like architectural design and implementation regain importance and need to be analyzed in great detail. In some cases, the features described in the Key Criteria section are the core solution, and the rest of the system is designed around them. This could be an important benefit for organizations that see them as a real practical advantage, but it also poses some risks in the long term. In fact, over time, the differentiation introduced by a feature becomes less relevant and falls into the table-stakes group, while new system capabilities introduce new benefits or

address new needs, with a positive impact on the metrics like efficiency, manageability, flexibility and so on.

Key criteria bring several benefits to organizations of all sizes with different business needs. The section is organized to give the reader a brief description of the specific functionality or technology, its benefits in general terms, and what to expect from a good implementation. In order to give a complete picture, we also include examples of the most interesting implementation currently available in the market.

Critical Impact of Features on the Metrics

Technology, functionality, and architecture designs that have demonstrated their value are adopted by other vendors, become a standard, and lose their status as a differentiator. Initially, the implementation of these key criteria was crucial for delivering real value, perhaps with some trade-offs. The most important metrics for the evaluation of a technology solution include:

- Architecture
- Scalability
- Flexibility
- Efficiency
- Performance
- Manageability and ease-of-use

This section documents the impact that individual features have on the metrics at the moment of report publication. Each feature is scored from 1 to 5, with a score of five having the most impact on an enterprise. This is not absolute and should always be verified with the organization's requirements and use case. Strategic decisions can then be based on the impact each metric can have on the infrastructure, system management, and IT processes already in place with particular emphasis on ROI and TCO.

Near-Term Game-Changing Technology

In this report section, we analyze the most interesting technologies on the horizon over the next twelve to eighteen months. Some are already present in some form but usually as part of niche products or for addressing very specific use cases. In either case, at this stage, the implementations available are not mature enough to be grouped in key criteria. Yet when implemented correctly and efficiently, the technologies can really make a difference to the metrics.

Over time, game-changing features become key criteria, and the cycle repeats. Therefore, to get the best ROI, it is important to check what vendors are offering today and what they plan to release in the

near future.

Companion Reports

The Key Criteria Report is part of a series of documents aimed at giving the reader tools to understand technology more fully, evaluate it, and explore the market to find the best solutions for their organization.

In this context, and to get a complete view of the state of the solutions available in the market, the reader should consider the following documents:

- **Key Criteria for Evaluating Hosted Kubernetes** report is an introduction to the technology. It defines the necessary evaluation metrics, the key criteria that may be used to evaluate new solutions, and the impact of the latter on the former. It is dedicated to those end-users that are approaching a new technology for the first time or want an update on the latest evolution.
- **GigaOm Radar for Hosted Kubernetes** offers a brief 360° view of the market, including market and technical positioning of most notable vendors, a short introduction of their solutions and differentiation, including a high-level graphic comparison of the vendors.
- **Vendor Profiles for Hosted Kubernetes** are easy-to-read deep-dive documents that cover a single vendor regarding the solutions described in the other reports. They provide more details on the solution, how the vendor approached the key criteria, and the impact that its solutions have on the evaluation metrics. This document helps end-users to get a quick but complete evaluation of a single vendo

2. Cloud File Storage Primer

As mentioned earlier, file storage is a good fit in a large number of use cases, and there is also some overlap with other types of storage.

COMPARISON OF DIFFERENT TYPES OF STORAGE

STORAGE TYPE	LOCATION	PROTOCOL	DATA /ACCESS RATIO	PERFORMANCE	SCALABILITY
Block	Data center, close to servers	FC, iSCSI	N:1 - 1:1	High IOPS Low latency Consistency	Usually scale up (less than 1PB)
File (NAS)	Local network, close to clients	SMB, NFS Direct FS Mount Client SW	1:N	Higher latency Good IOPS Good throughput	Scale out (multi PB)
Object Storage	Internet, accesible from everywhere and any device	HTTP, APIs	1:Millions	Highest throughput High parallelism Higher latency	Scale out (hundreds of PBs, Exabytes)

GIGAOM

Table 1: Comparison of Different Types of Storage

Scale-up file systems are becoming less relevant in the cloud with scale-out now taking the lion's share of solutions. The complexity of scale-out solutions is often hidden by automated deployment mechanisms and easier to manage than in the past or directly managed by the service provider. At the same time, they provide better performance, reliability, and scalability. Capacity expansion and reduction are also simplified and aligned with the expectations the user has about a cloud service.

Most common deployment models are:

- **Cloud-only solutions:** available only in the cloud. Often designed, deployed, and managed by the service provider, they are available only from that specific provider. The big advantage of this type of solution is the integration with other services offered by the cloud service provider (functions, for example) and its simplicity.
- **Hybrid and multi-cloud solutions:** These solutions are meant to be installed both on-premises and in the cloud, allowing them to build hybrid or multi-cloud storage infrastructures. The integration

with the single cloud provider could be limited compared to the other option and more complex to deploy and manage. On the other hand, they are more flexible, and the user usually has more control over the entire stack for resource allocation and tuning. These solutions can be deployed in the form of virtual appliances, like a traditional network-attached storage (NAS) filer but in the cloud, or a software component that can be installed on a Linux VM (i.e., a file system).

Many types of workloads that can be served by file storage systems, and there is not a single solution for all of them. This is why most cloud providers offer several solutions in their service catalog and collaborate with third-party vendors to further expand their product line.

File storage can be accessed in two different ways:

1. **Network share:** Network protocols (network file system (NFS) and server messaging block (SMB) are available on all operating systems and can be easily configured. This is the simplest way to configure access while getting good performance. This solution is the most common and covers the majority of use cases.
2. **File system client:** Some file systems can be accessed directly, thanks to software installed in the client. The file system is seen as a local file system to the client, and all communication is optimized. This option adds some complexity from the management perspective but offers the best performance for highly demanding workloads such as big data, HPC, and ML.

Licensing and subscription models are other aspects to consider when it comes to cloud file storage. In most of the cases, this is merely a consequence of how the storage is deployed and managed:

- **As-a-Service:** Customers can subscribe to a Storage-as-a-Service (STaas) that can be accessed through a network protocol, such as NFS or SMB, or a specific client. The solution can also be a cluster of virtual machines (VMs) with the scale-out file system installed in it. The second option is becoming very common among storage vendors, with their products available on marketplaces of major cloud providers. This makes it an interesting option to build a storage layer that has the same functionalities across different clouds and on-premises installations.
- **Managed:** The service provider offers a managed infrastructure that includes the complete cluster and basic management tools. This type of solution is usually based on traditional scale-out file systems (like GlusterFS or Lustre). Usually not originally designed for the cloud, it has the advantage of being easily replicable both on-premises and with other service providers, limiting the potential lock-in and giving the user a familiar environment to operate.
- **Bring Your Own License (BYOL):** The end-user already owns the licenses, and the software is also supported for cloud deployments. In this case, it is highly likely that the end-user already wants to migrate from on-premises infrastructure to the cloud, or wants to keep full control over the infrastructure and its software components. This scenario is the most flexible and gives better control over costs, even though it is not the easiest to manage. In fact, for this type of installation, the skills of the system administrators are key to success.

The main differences between the approaches described above can be found in the total cost of ownership of the solution. The first one (Storage-as-a-Service) is the easiest to manage and probably the less tunable and efficient. The second option adds additional flexibility but is more expensive, while the latter offers the best granularity and control, but it is also the most difficult to manage.

3. Evaluation Criteria

As cloud file storage features and technology progress from near-term game-changing, to key criteria, to table stakes, Figure 1 presents the evaluation criteria and where they fit on the timeline.

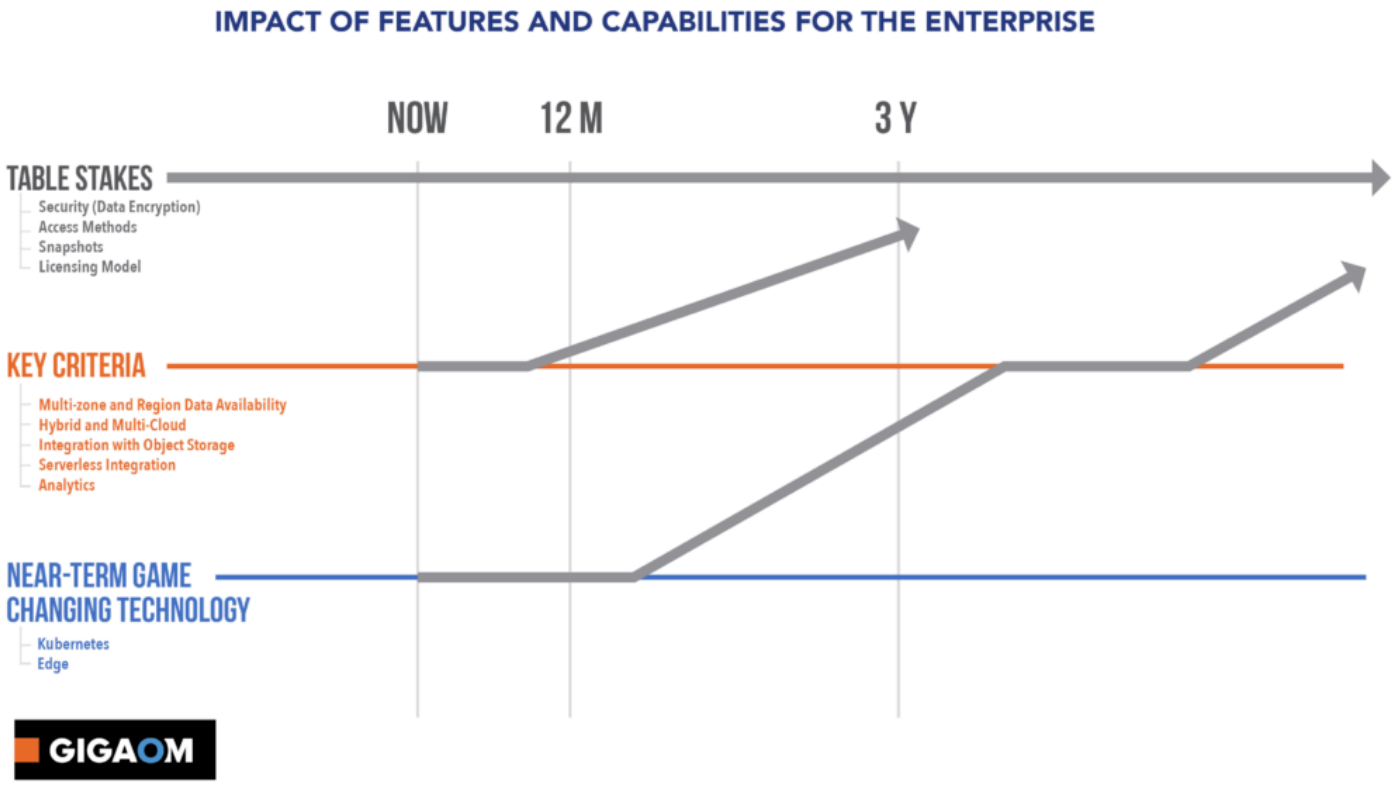

Figure 1: Timeline: Impact of Features & Capabilities for the Enterprise.

Another example of this progression is protocol support. This was a big differentiator once, with different levels of support for SMB or NFS, protocol versions, and limitations. Now, we practically take for granted that a modern file storage system supports the latest versions of SMB and NFS, with complete active directory support as well. In this context, what were considered key criteria for the evaluation of a file storage solution are now table stakes.

4. Critical Impact of Features on the Metrics

The metrics for the evaluation of cloud file storage for enterprises include:

Architecture

To get a clearer view of the potential of a file-based storage system, beyond the features available today, it is important to understand its architectural design. In fact, file systems are complex, and the architecture design has a direct impact on all the other metrics described in the following paragraphs, also giving indications on potential evolutions of key features.

Scalability

Even though it is true that the trend is toward scale-out architectures, scalability is still a challenge in many cases. Capacity is only one aspect of scalability, with other parameters such as the number of files per directory, ability to take advantage of new resources, and ability to handle metadata operations at a scale all contributing to defining how scalable a file system is.

Flexibility

A file storage system that is quickly adaptable to different workloads, data types, and supports multiple configuration options allows for a quicker response to new business needs and evolving scenarios. Even more so, the possibility to deploy it in hybrid, multi-cloud, and multi-region environments opens up additional opportunities to create a consistent storage experience across different infrastructures.

Efficiency

Often underestimated in cloud environments, an efficient file system can contribute heavily to cost savings. Data footprint reduction, such as compression and deduplication, allows enterprises to save capacity. When similar techniques are also used for remote replication, egress fees and network bandwidth can be optimized as well.

Performance

Depending on the workload and the amount of data under management, performance is another key aspect of every file system. It is important for completing a job quickly and getting results sooner, but also to save money because of the reduced amount of compute cycles wasted while waiting for data to reach the central processing unit (CPU). For this reason, performance is fundamental for practically every application that works with active data.

Manageability & Ease of Use

A simple file system does not need a management console or any other sophisticated user interface. Still, if the user wants to build a solid file-based storage infrastructure spanning across multiple clouds, things radically change. A management console for configuration tasks and monitoring dashboards, associated with analytics and application programming interfaces (APIs) to automate and simplify most repetitive tasks, becomes a crucial component of the system.

5. Table Stakes

There are some important characteristics of file-based cloud storage that are now common to most solutions available in the market and that users take for granted.

Reliability & integrity

Two of the characteristics that a modern file system should always provide, no matter if deployed on-premises or in the cloud, are reliability and data integrity. This comes before the high availability of the file service and any other advanced feature described in the following pages. Usually, several mechanisms (e.g., data scrubbing and error checking and correction – ECC) are in place to ensure that data has been saved correctly and remains safe over time. At the same time, multiple-parity data protection schemes (e.g., RAID 6 or distributed erasure coding) are designed to protect data against multiple infrastructure failures.

Security (Data Encryption)

Usually, the storage infrastructure of all major cloud providers is very secure, and many of its features are passed to users. However, for multi-cloud solutions not under the direct control of the service provider, security may be a concern. Fortunately, the majority of file-based storage systems offer the possibility to encrypt entire data volumes to limit the risks of data leaks for data at rest, and the most sophisticated ones can also encrypt data while in transit.

Attention to security does not stop here, and there is a growing number of solutions addressing security at the management level with role-based access control (RBAC) for system administrators and by providing detailed log and auditing features to improve compliance for highly regulated environments.

Access methods

NFS and Microsoft SMB are the most common protocols to access shared file system volumes from Linux and Windows operating systems. Clients are already available in the operating system, and they just need to be configured to mount the necessary resources. Compatibility is no longer an issue with practically all vendors now supporting the latest version of these protocols.

For a restricted number of use cases and products, especially when the highest performance is required, a specific client provided by the storage vendor allows the file system directly to be accessed with a proprietary protocol. This method requires the installation of software on the client node, but it is usually much more optimized that general-purpose protocols.

Some storage vendors also support additional protocols like file transfer protocol (FTP), HTTP/S-based representational state transfer (REST) APIs or even a subset of the S3 API, but they are less common

and usually are available only to ensure full compatibility with on-premises installations of the same product.

Snapshots

Snapshots are available for practically every type of file-based storage system in the market now, and the cloud doesn't make an exception. They are used as a very first backup copy or to speed up other operations and processes. Usually implemented with the "redirect on write" (RoW) methodology for efficiency reasons, the user can take as many snapshots as needed and manage their retention through user-defined policies or manually with APIs, command line interface (CLI) or user interfaces (UIs).

Licensing model

The entire IT industry now mimics the cloud, and vendors can sell both software licenses and hardware equipment in a pay-as-you-go fashion with a subscription. This helps the user to align the purchasing models and move more of the budget expense from capital expenditures (CAPEX) to operating expenses (OPEX).

Another advantage of the subscription model comes from the simplification when migrating from on-premises to the cloud. The user can always move the subscription to the cloud or, in many cases, stop it and start a new subscription from the cloud provider's market place to consolidate the billing.

6. Key Criteria

File services in the cloud are used for a large variety of use cases, and most of the users can usually opt for one of these options:

- **Simplicity:** The services available on the service provider platform are the most integrated with the ecosystem, really easy to manage, with a reasonable price, but often without the scalability and performance of other solutions. This is the best option for relatively small amounts of data, new applications, and if the user does not have plans for hybrid or multi-cloud.
- **Lift and Shift scenarios:** In case of migrations of existing workloads, the user has two options. The first is to migrate data to one of the services offered by the service providers. The second, when possible, is to deploy the same storage platform currently in use on-premises in the cloud. In either case, it is highly likely that the service provider can offer an offline migration mechanism in the form of a transportable storage appliance to move large quantities of data to their data centers. At the same time, if the customers opt for the second solution, remote data replication could be the easiest way to move data to the cloud.
- **High-performance workloads:** There is not a general rule of thumb for these kinds of workloads, and it depends on the actual performance needs, capacity, and if the data sets are needed only in one or multiple locations at the same time.

Again, it is important to note that service providers acknowledged the importance of file services and are quickly expanding the number of options available for their customers. At the same time, storage vendors are working to make sure that their storage software or virtual appliances are available and integrated with as many service providers as possible. In this context, the key criteria for evaluating file-based cloud storage should be weighed depending on the use case and overall cloud strategy.

Multi-Zone & Region Data Availability

Support for high availability across multiple zones and regions in the same cloud is a basic requirement for mission-critical applications, and avoid or minimize service disruptions in case of a disaster that impacts services in a region.

Usually, cloud providers and storage vendors offer multiple levels of service for data replication across zones or regions:

- **Synchronous replication:** Due to latency requirements, it is normally offered only for data replication across availability zones (AZ) in the same region, it is quite expensive and may have an impact on performance as well. The recovery time objective (RTO) and recovery point objective (RPO) are near zero.
- **Asynchronous replication:** The impact on performance is minimal, and usually, the user can decide

how often changes are replicated to other zones or regions. This solution is more cost-effective and applicable to longer distances, globally at times. RPO could be a limiting factor and should always be considered when designing the infrastructure.

Several cloud providers and storage vendors offer 1:N replication and cascading options with local cross-AZ synchronous and remote asynchronous replication. This is an option that is very expensive and that can be limited by the number of regions where the service is effectively available, but offers the best availability.

Hybrid & Multi-Cloud

Hybrid and multi-cloud solutions are rarely offered by cloud providers, while several storage vendors provide these options to their customers. Selecting a solution that can span across different environments enables the user to:

- **Simplify operations:** The same API, UI, and management consoles are available on all the clouds. Some vendors also offer a unified management tool that allows users to manage all physical, virtual, and cloud systems from a single UI.
- **Migrate and synchronize data across different clouds:** The solution, no matter where deployed, has an identical feature set, including remote data replication. This allows the user to bring data close to where it is consumed and keep volume synchronized.
- **Disaster recovery:** Having the possibility to keep volumes synchronized across on-premises infrastructure and the cloud simplifies DR strategy while contributing to making it more cost-effective.
- **Data repatriation:** Users with a mature hybrid cloud strategy may decide to opt for repatriation of some data and applications. Moving back an application with a stable and predictable workload to an on-premises data center can lead to substantial savings.

Hybrid and multi-cloud are the best options for a solid multi-cloud storage strategy and the one that has the best impact on TCO. Limitations of this approach could come from the limited ecosystem integration, and depending on the implementation, it could be less optimized for a specific cloud environment, leading to a worse $/GB.

Usually, storage vendors offer their solution to the CSP's marketplace. In other cases, the product is integrated into the service catalog or is available as a software or virtual appliance depending on the maturity of the solution and licensing requirements.

Integration With Object Storage

It is undeniable that object storage has been steadily growing since the very beginning of cloud

computing. Developers love it for its characteristics and its API interface, while it is also the cheapest form of storage in the cloud and notably slower than file or block storage. Another characteristic of object storage is its accessibility thanks to the S3 protocol, which uses HTTP/S as a transport layer.

Combining file and object storage has several benefits for the user:

- **Storage tiering:** With the right automation in place, less accessed files can be moved to object storage. This technique allows us to configure a smaller but high-performance file storage system and build capacity with the object storage in the back-end.
- **Backup and disaster recovery:** Some file storage systems allow organizations to take snapshots or make full copies of data volumes to object stores. This functionality is often used to make backup copies, which can also be used to populate a new instance of the file storage system in other locations. This method is slower than remote replication, but it is less expensive and does not need to instance the file store to run simultaneously.
- **File system population and de-population:** Object stores can be massive and are the perfect solution to build data lakes. Some file-based storage systems can query an object store (e.g., all metadata tags that contain a specific string or all files created after a certain date) and create a filesystem with all the objects/files that satisfy the terms of the query. This functionality is particularly useful for HPC, big data, and AI workloads. At the end of the work, new resulting files can move back to the object store.

These are only some of the possible integrations. Still, it is clear that combining scalability and low cost of object storage with the flexibility and speed of a file system can open up several opportunities for the user.

Serverless Integration

More and more cloud services now take advantage of integration with serverless functions. A function is a small piece of code that does a relatively small job, usually triggered by an event. Events can be of any type. The main advantage of serverless computing lies in the fact that the user does not need to manage anything of the infrastructure, which is totally in the hands of the service provider. In the case of file storage, functions can be associated with any file system operation like file creation, read, write, deletion, or update.

By associating functions to storage events, the user can offload part of the application logic to the infrastructure, or add new functionalities without needing to modify existing code. Possible use cases are plenty, including:

- Check file format and content integrity
- Scan the newly uploaded file for viruses or other malicious code

- Check file content for compliance
- Modify the format of newly updated files
- Create additional versions of new files. I.e., creating a low-resolution picture from a high-resolution one
- And much more

To get the proper integration with serverless functions, the storage system must be able to pass event messages to a message broker, so that the framework of a function can intercept them. This kind of architecture is pretty common now among cloud service providers, but it is less so for storage vendors.

Implementations of serverless computing functions are slightly different across cloud providers, and the same goes for APIs and events they support, making functions not very portable. This means that it is highly likely this type of feature is available only from cloud providers that own the entire solutions and infrastructure stack.

Analytics

As soon as the file storage infrastructure grows and becomes more relevant, spanning across multiple cloud environments, knowing more about data, workloads, and applications accessing it becomes crucial both for overall infrastructure efficiency and cost savings. Storage solutions designed for multi-cloud environments should always have a single management console, able to consolidate telemetry from all the environments and give a status of the entire infrastructure at first glance.

Similar to what happens for traditional storage, the user can benefit from deeper data analysis as well:

- The **system administrator** can use the information to better understand the workloads and their impact on the storage infrastructure, plan for future expansions or migrations, select the best policies for the best compromise between performance and cost.
- The **developer** can use the same information to optimize the code and see how code changes impact the infrastructure. Furthermore, depending on the type of applications and processes in place, the developer can use the same information to select the right resources when (self-)provisioning storage for applications.
- The **business manager** can use this information to analyze costs and make strategic decisions about the future of the infrastructure and its expansion.

7. Critical Features: Impact Analysis

As described earlier in the introduction, this report analyzes the impact of critical features of cloud file storage solutions available in the market. It puts them in context with the evaluation metrics that are usually at the core of strategic decisions.

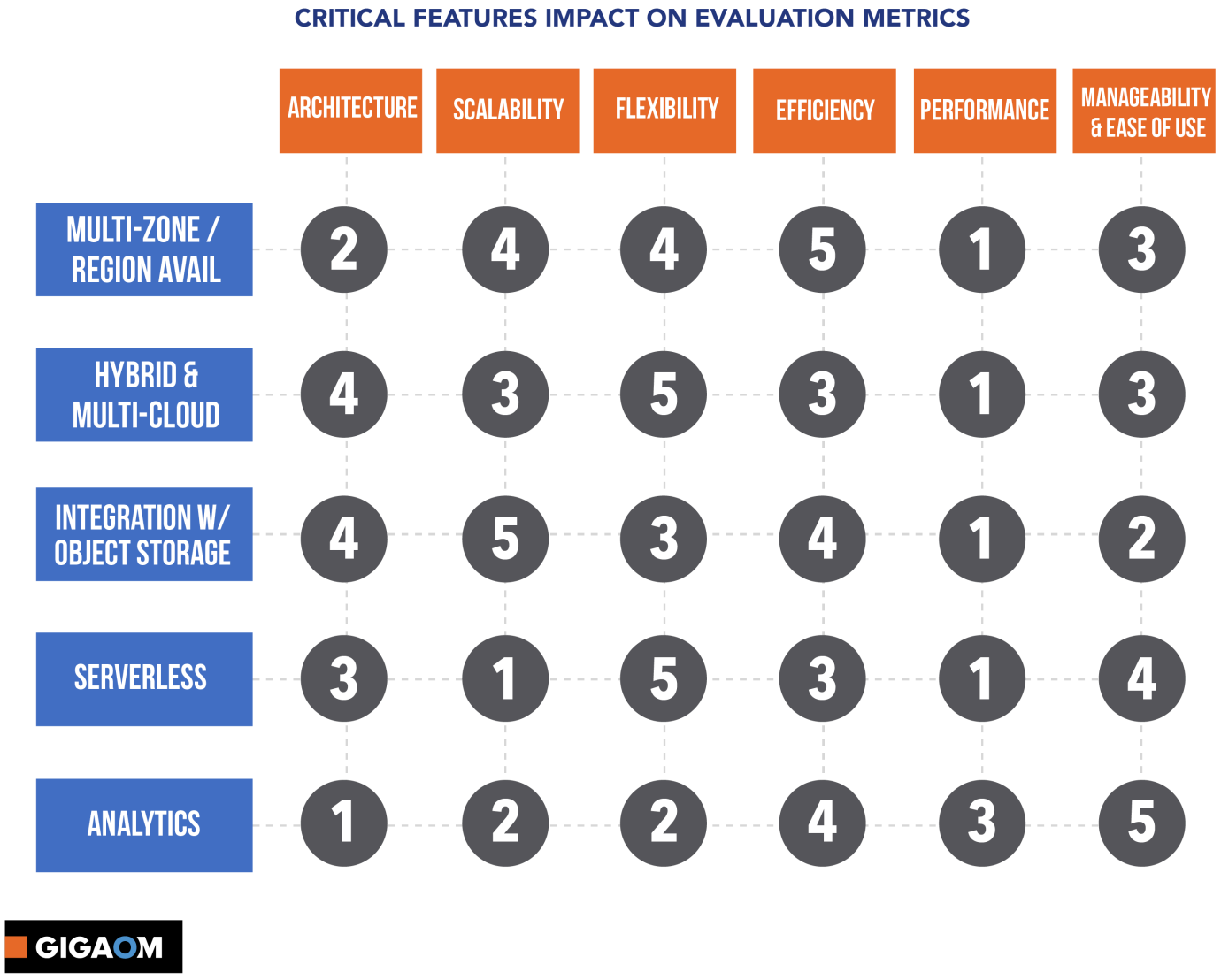

Table 2: Critical Features Impact on Evaluation Metrics

Impact on Architecture

The internal design of the storage solution is at the base of all the features exposed in the front-end and back-end. In this particular case, cloud file storage systems, the ability to deploy the file store in multiple zones for high availability, in different regions for data-application proximity, and on multiple

clouds give the user extraordinary freedom of choice, leading to a better TCO.

Impact on Scalability

Horizontal and vertical scaling are both of critical importance for file storage, especially when the user needs to scale both legacy and traditional applications. At the same time, integration with an object storage back-end can introduce a scalability factor that wouldn't be possible to get otherwise or, at least, at a reasonable cost.

Impact on Flexibility

The flexibility of a cloud storage solution can be seen from different perspectives. On the one hand, we can deploy the storage where it is needed and in the form that is more appropriate (virtual or physical appliance, service, etc.). On the other hand, the integration with the cloud ecosystem is key to exploit all the potential of the cloud infrastructure.

Impact on Efficiency

Alongside the traditional concept of storage efficiency in terms of data footprint reduction and optimized data paths, which has a direct effect on cost reduction, the efficiency of a cloud storage solution is also measured in its ability to integrate seamlessly with the rest of the cloud ecosystem and provide the necessary APIs to developers.

Impact on Performance

More and more workloads are moving to the cloud, and performance should always be one of the parameters to watch. Some storage systems have performance as a primary characteristic and can take advantage of every available resource to improve response time. In general, storage performance is important for every application that deals with active data because it helps to optimize the entire stack by reducing latency, therefore wasted CPU cycles to wait for data to compute, and reduces costs.

Impact on Manageability

A simple file repository does not need sophisticated management tools. However, when the storage infrastructure grows in size and spans across multiple clouds, getting a complete view of what is happening is fundamental to avoid risks and provide a better service level for applications and end-users. At the same time, interacting with the storage infrastructure with a consistent interface (GUI, API, and CLI) is fundamental to simplify operations.

8. Near-Term Game-Changing Technology

Kubernetes

Interest around Kubernetes has been steadily growing for quite a while now. Enterprises have started to adopt it, and they require support for persistent storage volumes. Not all storage vendors are ready with CSI plug-ins, and many of the current implementations are still immature (more on this can be found on Key Criteria for Evaluating Kubernetes Data Storage report), but it is also true that every vendor is working on it.

Kubernetes support, in combination with the ability to sync data across clouds, simplify data migration as well as application mobility across clouds. It is a key functionality for supporting hybrid and multi-cloud IT strategies.

Edge

Edge computing is now becoming relevant for every distributed organization, and how to reorganize IT infrastructures at the infrastructure edge to take advantage of the latest technology is an ongoing conversation with most IT managers. The general consensus is that it does not make sense to move to the cloud the data that is created and consumed at the Edge, meaning that the Edge infrastructure should be designed to be self-sufficient for both data and applications for a certain amount of time at least. In this context, data storage plays a key role at the Edge with every storage vendor and cloud provider looking at how to exploit this opportunity.

In most cases, a full-fledged storage infrastructure does not make sense even if data should always be available locally. To satisfy this paradigm, caching and automated tiering are the most straightforward solutions to implement, while other approaches like storage federation are beneficial in specific use cases.

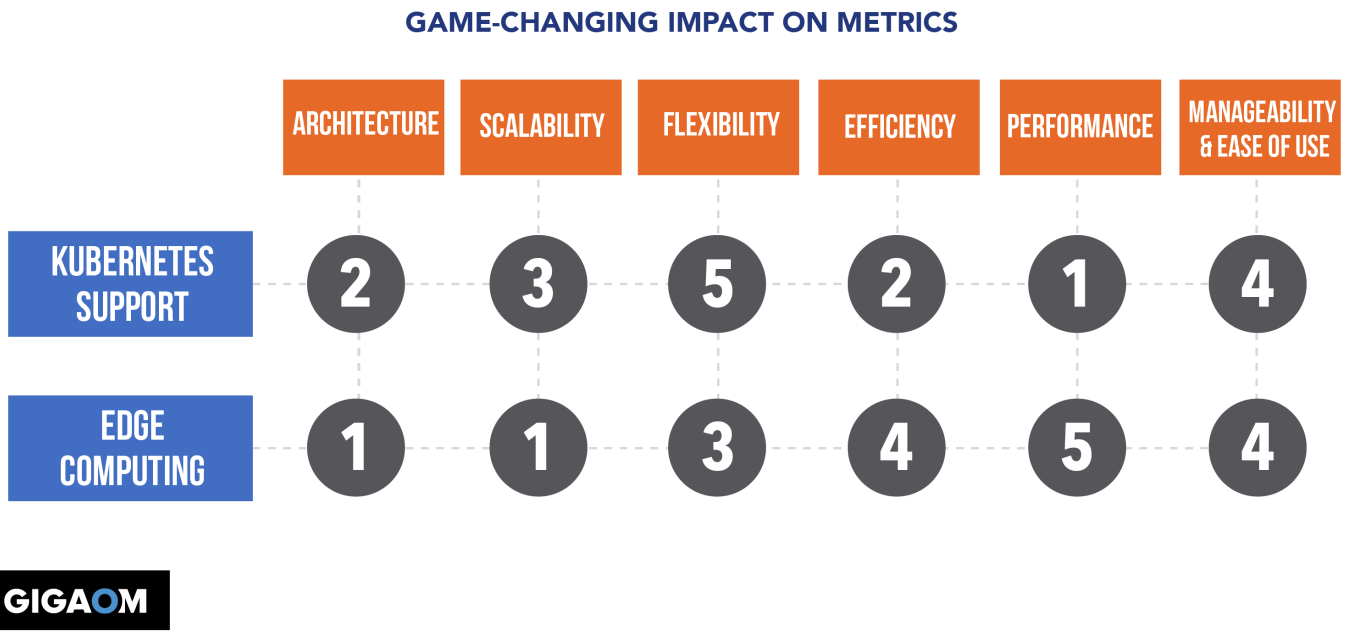

Table 3: Game-Changing Impact of features on Metrics

9. Conclusion

File storage in the cloud is gaining traction, thanks to the increasing number of enterprises adopting public cloud and moving data and applications across different environments.

For small organizations, looking at the public cloud for their primary IT infrastructure, file storage brings a familiar and user-friendly access method for both legacy applications and end-users while bridging the gap between traditional and cloud-native applications. These types of organizations are usually more concerned about the simplicity of the solution, how it integrates with the cloud ecosystem, and cost.

For larger organizations and users needing to interact with large amounts of data efficiently, file storage remains the best combination of performance and capacity. Specialized, high-performance file systems are now more cloud-friendly. At the same time, other solutions are specifically designed to work in hybrid cloud environments, simplifying data synchronization across different environments, and bringing data closer to the CPU. In this case, multi-zone, multi-region and cloud support become very relevant to execute on hybrid and multi-cloud strategies while the flexibility of the solution and its feature set are crucial to maximizing its deployment possibilities. Performance is another area where large organizations should always focus not only in terms of efficiency and speed but also for optimization and overall TCO.

10. About Enrico Signoretti

Enrico has 25+ years of industry experience in technical product strategy and management roles. He has advised mid-market and large enterprises across numerous industries and software companies ranging from small ISVs to large providers.

Enrico is an internationally renowned visionary author, blogger, and speaker on the topic of data storage. He has tracked the changes in the storage industry as a Gigaom Research Analyst, Independent Analyst and contributor to the Register.

11. 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.

12. Copyright

© Knowingly, Inc. 2020. *"Key Criteria for Evaluating File-Based Cloud Storage"* is a trademark of Knowingly, Inc.. For permission to reproduce this report, please contact sales@gigaom.com.