White Paper

# Multi-cloud File Storage Design Considerations

## How Efficient Metadata Handling Is Essential to Maximizing the Value of File Data across Multi-cloud Infrastructure

By Scott Sinclair, ESG Senior Analyst

February 2019

This ESG White Paper was commissioned by InfiniteIO and is distributed under license from ESG.

## Contents

## Introduction

The modern business climate has evolved dramatically with the rise of the digital transformation age. IT can no longer remain a subservient cost center. Instead, IT must leverage its expertise to transform into a critical business enabler of net new revenue and cost saving opportunities. In the digital economy, data is a strategic weapon, and businesses that cannot wield data effectively are at a massive competitive disadvantage.

According to ESG's research on IT spending intentions, 86% of IT decision makers agree with the statement, "If we do not embrace digital transformation, we will be a less competitive and/or effective organization."[1]

For the business, the rewards are massive. For IT, however, the transformation is often daunting. Its day job of keeping the business running is not going away, so IT must simultaneously find the extra cycles, budget, and resources to architect and execute this end-to-end business transformation rooted in leveraging data. And rarely do businesses, small or large, have the resources to scale personnel or processes in line with demands.

While multiple data types fuel these transformations, it is the file data environment that is often the most unwieldy and the most complex. The complexity of file data stems from massive surges in both content generation and retention that have been constant for nearly a decade. The top three identified workloads driving data growth are digital media (31%), collaboration (30%), and business intelligence (29%), and all three comprise file content.[2]

File growth, however, only scratches the surface of the real problem. For decades, the infrastructure designed to serve file content was architected for both preservation and cost control. The emergence of analytics, artificial intelligence (AI), and the Internet of Things (IoT), however, changes that dynamic. Now, speed is of the essence. Enabling business opportunity necessitates that the right data is identified, located, and isolated on high-performance infrastructure, quickly—or the business initiatives that leverage that data will suffer.

This is a small problem when you have a dozen or so terabytes (TBs) of data. Modern file environments, though, often measure 1,000 times larger, in the dozens of PBs or billions of files under management. Adding complexity, the rise of public cloud services and the edge has dispersed these massive data volumes across a disaggregated environment.

A glimmer of hope does exist, however, with recent innovations in metadata handling, and one innovator that you have likely not heard of, InfiniteIO, is taking a novel approach to addressing the challenge of multi-cloud file storage, one they define as the true hybrid cloud. The idea is simply that if metadata could be consolidated and then harnessed quickly and easily, then the file data universe would seemingly shrink and become far more manageable. To that end, InfiniteIO's architecture abstracts and accelerates the metadata for all file content in a multi-cloud ecosystem with zero impact to applications, simultaneously accelerating multi-cloud file storage while minimizing the cost.

## The Rise of Multi-cloud File Storage and Disaggregated IT

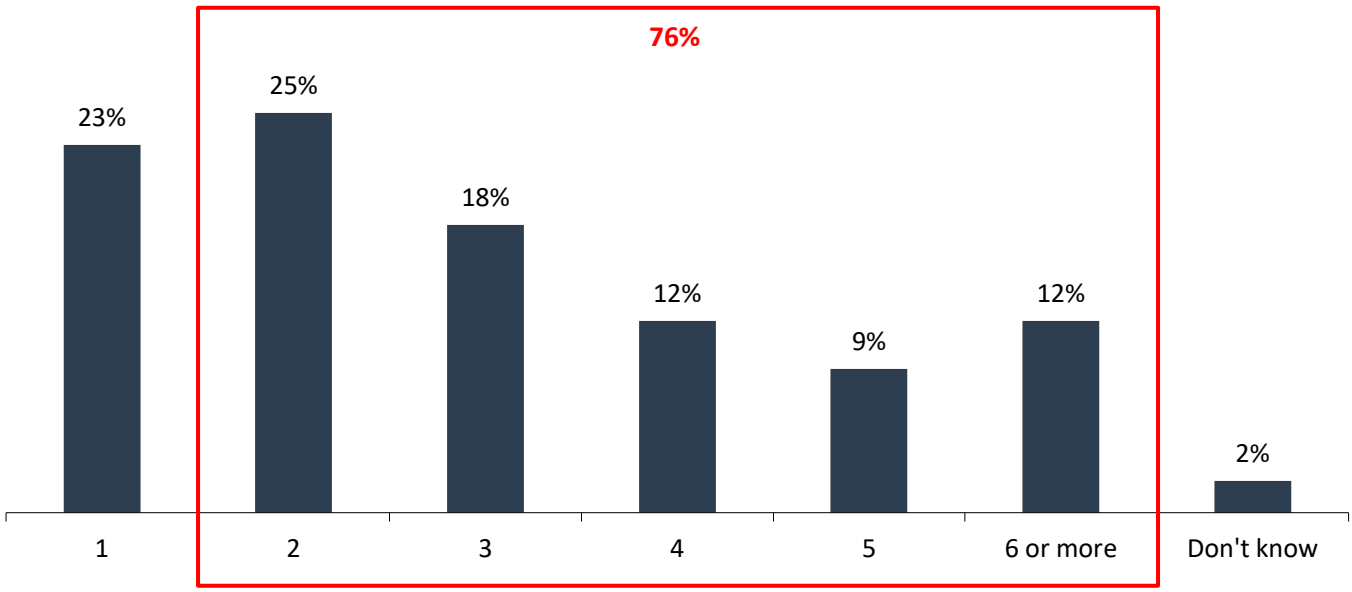
Simultaneously scaling capacity and performance quickly becomes unsustainable with traditional NAS systems. In an ideal world, all data would live on the fastest storage tier, such as all-flash, but for massive scale, often multi-petabyte file environments, that is impractical. The mounting silos of file storage capacity burden IT dearly. Increased data capacities are the most common driver of IT complexity, with 41% of IT decision makers who have experienced an increase in IT complexity citing higher data volumes as a cause.[3]

In response, IT organizations have turned to the public cloud, as over one-third (35%) of storage decision makers identified using "cloud storage" as a way to source capacity instead of purchasing new on-premises infrastructure as an initiative

[1] Source: ESG Master Survey Results, *2018 IT Spending Intentions Survey*, December 2017.
[2] Source: ESG Master Survey Results, *2017 General Storage Trends*, November 2017.
[3] Source: ESG Master Survey Results, *2018 IT Spending Intentions Survey*, December 2017.

expected to impact their organization's storage spending.[4] As a result, businesses are opting to integrate multiple public cloud infrastructure providers into their data storage ecosystem. According to ESG research, 76% of IaaS users leverage more than one public cloud infrastructure provider.[5]

**Figure 1. Number of Unique Public Cloud Infrastructure Providers**




*Source: Enterprise Strategy Group*

Despite the benefits, the addition of cloud services introduces performance impacts and new cost considerations, often in the form of egress fees, beyond those involved with on-premises file storage. Several of these new cost considerations are often underestimated by IT and the business. Additionally, regulatory compliance and data sensitivity often dictate where files can reside, further increasing both the need for a distributed infrastructure and the complexity of managing it. As a result, modern file storage environments are becoming increasingly disaggregated, complex, and costly. This increased complexity slows down IT service delivery, and, in a digitally driven economy, slower IT services translate into a less competitive business.

IT needs a new solution for enterprise multi-cloud file data storage and management, something that can bring consolidation, order, and high performance to a hybrid or multi-cloud environment while offering freedom in infrastructure choice and flexibility without impacting workloads.

## The Cloud: Highly Valuable, but not a Panacea

With the excitement and rampant adoption of public cloud services, many businesses have seen successes, but not every migration has gone according to plan. According to ESG research, 41% of IT organizations have pulled at least one workload back from public cloud infrastructure services to be run on-premises. While there are multiple factors behind these decisions, the most commonly identified rationales behind the moves back include data security (38%), cost (25%), and regulatory compliance concerns (23%), all of which involve file content (see Figure 2).

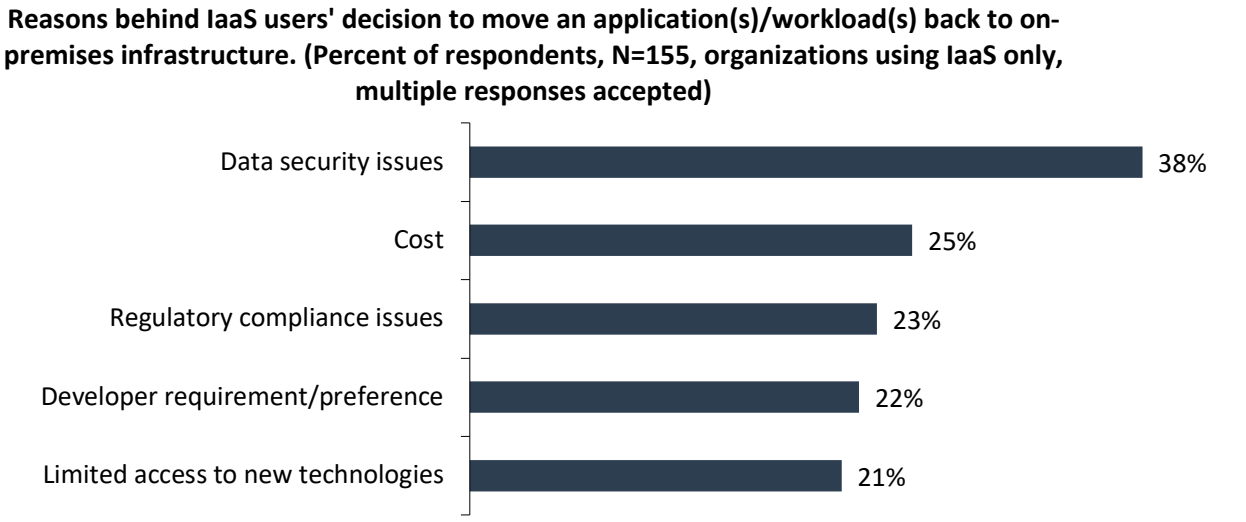
[4] Source: ESG Master Survey Results, *2017 General Storage Trends*, November 2017.
[5] Source: ESG Master Survey Results, *2018 IT Spending Intentions Survey*, December 2017.

When IT decision makers who had moved a workload back from the public cloud were asked to identify the on-premises storage solutions that replaced the public cloud:

- 27% identified that the storage for at least one workload was deployed on-premises with disk-based or hybrid (HDDs and SDDs) file storage.
- 24% identified that the storage for at least one workload was deployed on-premises with all-flash file storage, reinforcing the need for greater file storage performance.[6]

The bottom line is that simply moving everything to the cloud is not a practical option. IT needs to effectively leverage both on- and off-premises infrastructure.

**Figure 2. Top Five Rationales for Moving a Workload Back from Public Cloud Infrastructure Services (IaaS)**




*Source: Enterprise Strategy Group*

## The Role of Metadata in Addressing Multi-cloud File Storage Complexity

As hybrid and multi-cloud become the new normal, IT becomes increasingly disaggregated, and businesses at the forefront of the digital transformation age face a dilemma. The quantity, quality, and pace of data analysis and machine learning translate into great business opportunity but increasing the volume of data greatly diminishes the pace at which insights can be captured. The time lost is too expensive given the value of these initiatives. It is here that metadata handling can play an integral role.

For those not familiar with metadata, a brief primer may be helpful: It is the data about the file data. In a typical file storage environment, most of the requests for files are actually for metadata, as file metadata includes attributes such as when was the file created, who created it, who can read it, and the file size. Directory views and file content search operations all access metadata. In addition, multiple applications, such as for backup and storage functions, actively scan metadata for changes.

When the metadata is stored on high-performing infrastructure, these activities take less time. Conversely, slow metadata responses drag down file storage usage as scans, directory views, and searches all take longer. These delays hinder

[6] Source: ESG Master Survey Results, *Tipping Point: Striking the Hybrid Cloud Balance*, October 2018.

operations on file data and slow down applications that require metadata scans—for example, increasing the time required to back up or replicate a file environment.

For local storage that measures in the terabytes, the specific metadata storage design typically does not matter. As file counts and capacities scale, systems that once offered high-performance access can slow significantly. For file environments that measure in the tens of petabytes and are distributed across multiple data centers, multiple clouds, and the edge, the right metadata storage architecture is essential.

To optimize metadata handling, some storage systems store the metadata with the data, offering a scale-out architecture to accelerate performance. Some architectures offload the metadata to a separate storage pool, allowing the metadata to be stored on a high-performing storage tier. While both architectures have their advantages, the modern file storage ecosystem is dispersed across a variety of platforms and providers. Leveraging just one architecture for everything is inefficient and typically impractical.

Integrating public cloud infrastructure services also adds a new wrinkle. With some public cloud providers, metadata access requires a recall and reconstitution of the entire file, which not only slows down performance but also adds an egress charge to the operation. These egress charges involved with metadata operations add unpredictable, excessive, and unnecessary costs to cloud-based file stores.

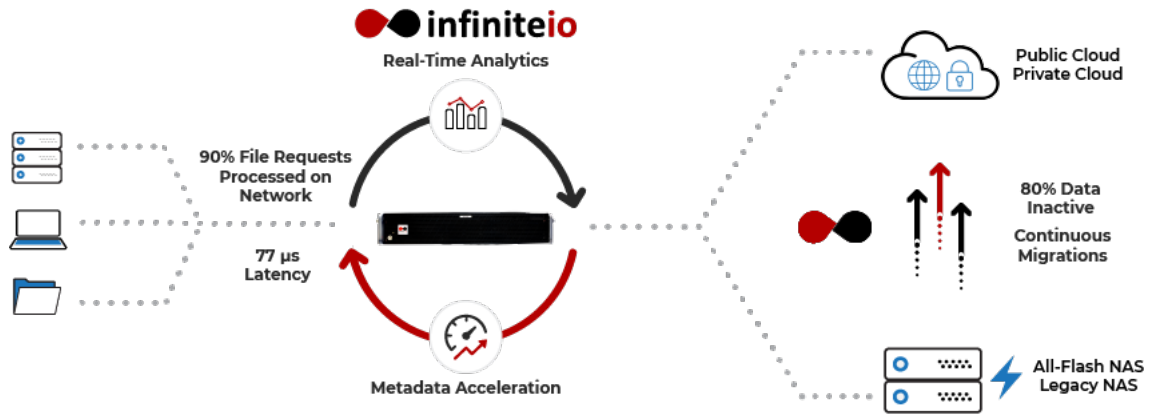
Efficient metadata handling is required for not just one platform, but the entire multi-cloud file environment. As demands grow, efficient metadata handing will be essential for businesses to maximize the value of their file data, control storage infrastructure costs, and ensure the successful integration of public cloud infrastructure.

## Transform File Storage with InfiniteIO's True Hybrid Cloud

Vendors are often content to focus innovation solely on their own storage platform. As a result, few solutions address metadata handling at a multi-platform, multi-cloud, and multi-workload level. One emerging innovator, InfiniteIO, believes metadata is the key to unlocking the true hybrid cloud and has developed a novel approach to addressing this multi-cloud metadata issue.

InfiniteIO's system, the NSC-110, delivers high-performance local metadata abstraction via deep packet inspection of both on- and off-premises file stores and puts the metadata in a memory-based metadata map. Leveraging memory to store metadata accelerates the metadata operations, and even massive, highly distributed file environments appear as if they reside fully on high-performance local storage.

**Figure 3. InfiniteIO Metadata Acceleration**




*Source: InfiniteIO*

By leveraging deep packet inspection to learn metadata in real time, the NSC-110 sits invisibly between the clients and the existing file systems, delivering the benefits of metadata acceleration without massive investments in new infrastructure, time-consuming and costly file migrations, or even disruptive changes to existing mount points. InfiniteIO then offers a variety of policy-based tools to automatically identify and then migrate data to the optimal tier in the multi-cloud file ecosystem, and all data movement is encrypted for better security.

This design stems from InfiniteIO's true hybrid cloud ideal, which is built upon five fundamental principles:

1. **Real-time insights on data center and cloud traffic:** As stated previously, the speed of data analysis is of the essence. Automating data management policies using machine learning is a good first step. Given the rise of multi-cloud environments, however, the business initiatives that leverage large amounts of data will continue to suffer unless the analytics can keep up with changes to file metadata and provide a real-time view of traffic between the data center and the cloud. Identifying changes to files by scanning and rescanning petabytes of data, which could take weeks to complete over and over, defeats the purpose. The optimal hybrid cloud will have a live, granular view of hot and cold files over the entire environment, so that any policy automation will always know the true state of each file no matter where it is located.

2. **Integration into existing file infrastructure with zero workload impact:** Accessing innovation often necessitates disruption. Deploying a new file storage solution requires the ripping and replacement of existing infrastructure. Even a software-based file abstraction layer introduces new mount points, disrupting workloads, adding risk, and often introducing downtime. InfiniteIO believes that true hybrid cloud innovation should be deployed without disruption and without new mount points, while enabling businesses to maintain their existing infrastructure investment. Without these characteristics, any solution would likely be a nonstarter for an IT organization undergoing massive transformation due to the upfront disruption and the considerable risk involved.

3. **High-performance metadata access for cloud data without a file recall:** Conceptually, a true hybrid cloud should present a consistent infrastructure experience regardless of whether the data resides on-premises or off-premises. While there are some limits due to laws of physics that cannot ever be resolved fully (such as transmitting data over distance), a true hybrid cloud solution should do everything possible to minimize egress fees, such as those incurred during metadata operations. In other words, businesses should be able to make the data storage location decision based on what is right for the business, without significantly impacting the applications. Achieving this goal requires high-performance metadata operations regardless of the data's location.

4. **Transparent private and public cloud storage utilization:** Capacity limitations and cloud egress fees often play a major role in determining the overall cost of file storage, and therefore often impact the design of the file storage environment. According to InfiniteIO, true hybrid cloud solutions can address both these factors by presenting consistent metadata but allow the files to move independently. This allows enterprises to migrate inactive/cold files to private or public cloud storage and fully utilize their practically limitless capacity. Additionally, whether data is stored locally or migrated to cloud storage, the applications should neither know or care. Automating migration based on real-time analytics will ensure that any files that have to be recalled will move to the cloud once they become inactive again.

5. **Scale with simplicity:** Organizations must be able to manage future data growth without adding complexity. Some file storage systems leave a stub or a symbolic link of the file on the local file store when the data is migrated to the cloud. This model requires access to file data on the cloud to always go through the local storage device, slowing down file accesses and introducing complexity if the organization chooses to remove the local storage system in the future but retain the data on the cloud. Others require new mount points or other network reconfigurations. As capacity grows to petabytes or billions of files, InfiniteIO believes that hybrid clouds should

scale data management without adding complexity. That means no stubs, symbolic links, VM sprawl, third-party file systems, or modifications to the clients or network required. All of these represent complexity, or worse, points of failure.

## The Transformation Benefits of Multi-cloud Metadata Abstraction and Acceleration

Leveraging its true hybrid cloud blueprint, InfiniteIO's technology offers a transformational blend of business benefits for application acceleration and risk reduction. Some specific benefits include:

- The "no lock-in" flexibility of InfiniteIO's approach: InfiniteIO describes the NSC-110 as a network storage controller. The "network" term in the description highlights InfiniteIO's ability to install as a clustered/redundant solution invisibly—as if it were a network switch in front of existing storage systems, rather than a new storage target.

  As a result, InfiniteIO metadata abstraction requires no new mount points. There is no disruption to existing applications and no need to alter existing applications. InfiniteIO also supports the existing in-place file storage infrastructure; file data can stay where it lives. This approach offers tremendous benefits relative to more common file virtualization solutions by removing the cost, complexity, and risk associated with the transition to new mount points. As nothing in the existing environment needs to change, there's no risk of making a configuration mistake.

- High-performance file storage while reducing cost of multi-cloud infrastructure: InfiniteIO's metadata abstraction extends across on- and off-premises file stores, delivering local flash-like performance for directory views, scans, and searches of file data on the public cloud as well as on-premises storage.

  With the metadata abstracted and located in high-performance memory, file directory views, scans, and searches for multi-cloud environments operate as if the data were on high-performance local storage. These operations are vital for businesses that want to leverage files for analytics, businesses intelligence, or AI workloads. Leveraging the cloud is easier, faster, and more effective. This acceleration also introduces the opportunity to incorporate lower cost public cloud storage tiers. With the metadata stored locally, organizations can leverage the cloud more freely without concern over metadata scans or operations generating unexpected and costly egress fees.

- Automatic location of inactive data across a heterogeneous multi-cloud environment: InfiniteIO's Infiniview portal allows admins to create policies for data migration based on a wide variety of variables and more than just file activity. Examples of options include last access date, file type, and user ID. Infiniview also allows for the use of wildcards and variable combinations to create policies that continuously migrate files. When combined with its invisible deployment approach and its ability to leverage a wide variety of heterogeneous file storage ecosystems both on- and off-premises, the power of this level of automation is greatly increased, offering the potential for substantial operational savings. With the administrative cost associated with maintaining an operationally efficient file ecosystem greatly reduced, personnel can be freed to focus their attention on the real goal: maximizing the value of the file data.

- Automated tiering to archive or to the cloud without slowing down the metadata: Building on the benefits of high-performance abstraction with automated file search and identification, InfiniteIO can also automatically migrate inactive file data across multiple file stores to low-cost cloud/object storage. Additionally, administrator intervention is not necessary to recall a file. Combined with high-performance metadata abstraction, all files are presented as local across a hybrid or multi-cloud environment. Native object workflows transparently flow through and are serviced directly by the attached public cloud or object storage. Data is not locked into a specific storage vendor once moved because the data is stored natively at the destination. As a result, data migrated to the cloud can fully leverage the services offered by the cloud provider.

## The Bigger Truth

File data is a strategic and incredibly valuable asset. Businesses that can quickly and effectively harness these assets find success, while those that cannot often struggle to stay relevant. This is the reality of the digital economy and the driving force behind digital transformation. Analytics, AI, and IoT are all predicated on the creation of content and the ability to quickly analyze the data, identify insights, and adjust decisions, processes, and strategies accordingly.

With continued growth in hybrid and multi-cloud adoption, the future of file storage will likely continue to be just as diverse as it is today, if not more so with capacity growth showing no signs of slowing down. InfiniteIO, however, makes diverse file ecosystems seem smaller, more manageable, and more accessible, without changing the existing environment. Business have greater freedom to choose the right platform for their data and more freedom to adjust. InfiniteIO calls this the true hybrid cloud. And with it, businesses are freed to focus attentions on leveraging their digital assets to their fullest extent, rather than spending their time just maintaining the status quo. While still a young player in the enterprise file storage space, InfiniteIO's technology has tremendous potential. And its invisible, no-risk deployment approach makes it easier to experience what the technology can do.