



Give Your Data Management a Boost:

Overcome the Limits of Your Storage
Infrastructure for Unstructured Data





Table of Contents

Building Scale for Unstructured Data

Large Telecommunications Provider Supercharged
its Storage

Navigating the Exponential Growth of Unstructured Data

The Real Struggle of Managing Unstructured Data in Certain
Industries

The Rise of File Data in the Public Cloud

Top 10 Benefits of Hybrid Cloud File Storage [Qumulo 2020]



Building Scale for Unstructured Data

Organizations are facing exponential data growth and performance demands. Their traditional storage infrastructure is just not able to support their business requirements.

To succeed, organizations must address their challenges by adopting the most cost-effective, scalable, flexible and high-performing platforms suited to their specific data and application needs. Determining that platform impacts and is shaped by their data management, private cloud and custom hybrid cloud strategies. ePlus helps organizations across all industries address these challenges every day.

Many organizations require massively scalable file and object storage platforms in the hybrid cloud, for data including:

- application data
- media
- log data
- archive
- long-term backup retention
- big data analytics

In these cases, file storage platforms are the ideal solution—notably for organizations in industries including:

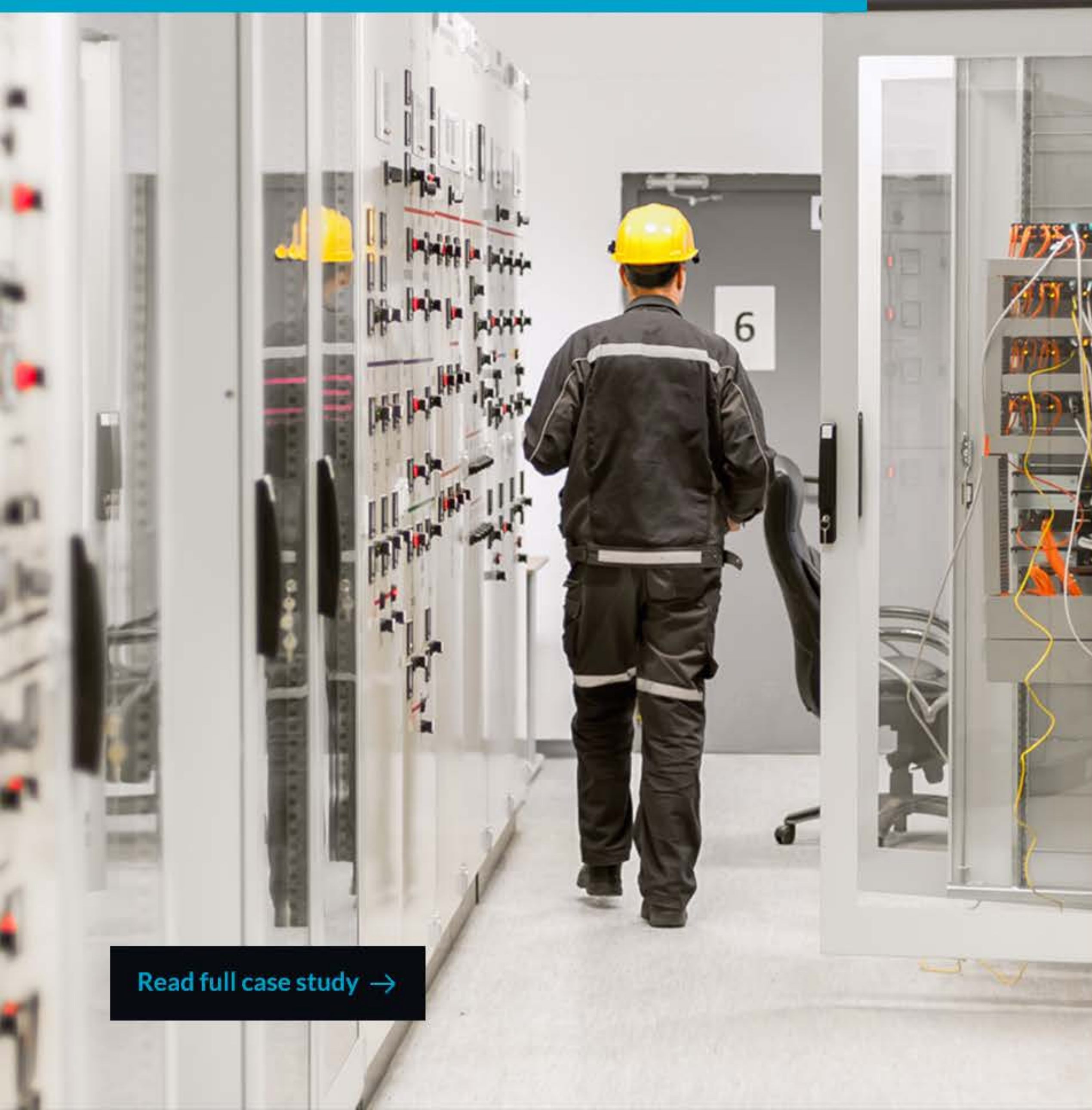
- media and entertainment
- gaming
- telecommunications
- utilities
- healthcare
- video surveillance
- manufacturing
- libraries

Large Telecommunications Provider Supercharged its Storage.

ePlus and Qumulo helped a major telecommunications company handle its massive amounts of unstructured data with hybrid cloud file storage that delivered the necessary capacity, reliability and performance the company needed.



Large Telecommunications Provider Supercharged its Storage



One of the largest telecommunications providers in the world needed to replace its aging scale-out system with a modern storage infrastructure to handle the immense volume and velocity of machine log data. ePlus, with the help of Qumulo and its hybrid cloud file storage, delivered the necessary capacity, reliability and performance the company needed.

Large telecommunications providers make billions of endpoint connections each day, and every connection generates log data that must be ingested, stored and processed to identify events or anomalies. These logs represent daily terabytes of machine data, amounting to multiple petabytes of data over time. Storing this immense volume of data requires incredible capacity and performance. The data also needs to have uninterrupted availability, so it can be analyzed in real-time. Constant uptime takes both reliable equipment and impeccable customer service. That was the challenge facing this top U.S.-based carrier.

When you're ingesting terabytes of data each day from more than 60 billion incoming events, you need storage capacity and performance that's very, very scalable," said a high-level executive. "You can't afford a misfire with the production system."

After the telecommunications provider determined its current solution set was too risky for an upgrade, and decided that object storage would not be able to provide much needed protocols, the company approached ePlus and Qumulo for a scalable file storage solution.

Qumulo software defined technology offered the telecommunications carrier the storage performance and scale the company required, along with the flexible API based programmability that the object system had promised.

[Read full case study →](#)

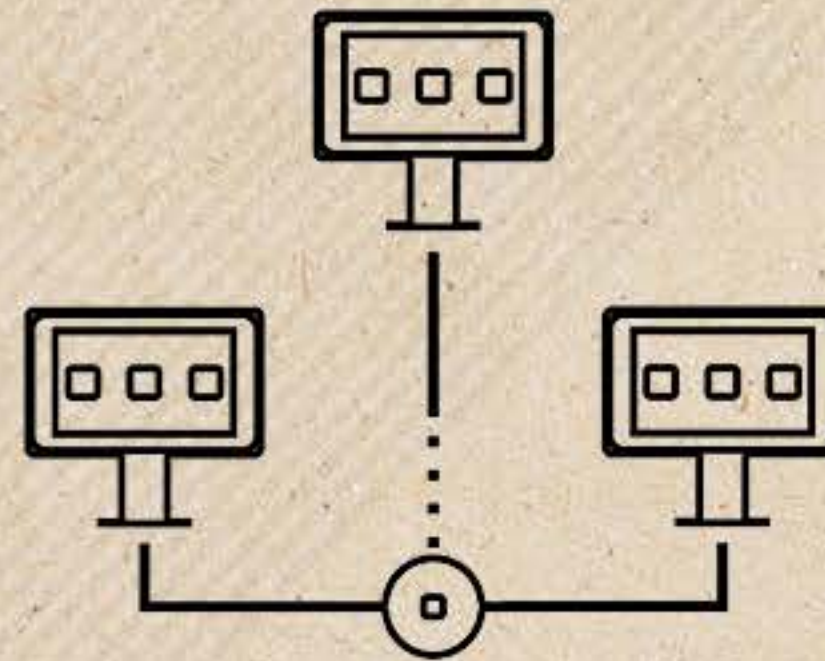


Distributed File vs. Object Storage

Distributed file systems and object storage differ in how they work.

Object storage houses data in structures called "objects," and serves clients data via RESTful HTTP APIs, such as Amazon Simple Storage Service (S3) and OpenStack.

Distributed file storage clusters multiple storage nodes together in a single parallel file system, with a storage pool to provide multiple hosts with high-bandwidth data access in parallel. Data is distributed over multiple nodes to deliver data availability and resilience, and to provide high throughput and scale capacity in a linear way.



Distributed file systems and object storage are data storage architectures for unstructured data. They are defined by Gartner as software and hardware solutions that support object/scale-out file technology to address requirements for unstructured data growth.¹

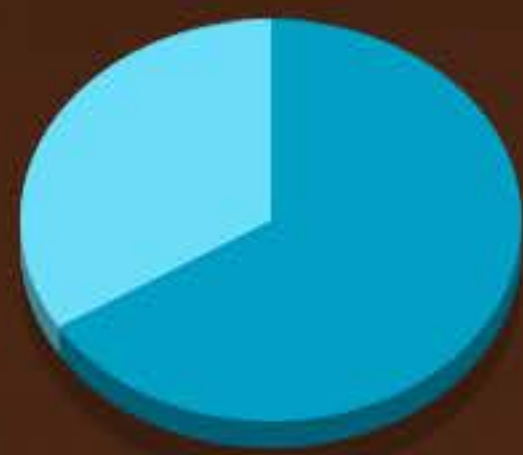
Navigating the exponential growth of unstructured data driven by the digitization of everything

By Ben Gitenstein, VP of products and solutions, Qumulo

We have far surpassed the days of “big” data, and are deep into the realm of massive, mind-numbing, astronomic amounts of data—with no end of this growth trajectory in sight. In fact, during the last 10 years, the amount of data that emerged in the world has increased from 1.2 trillion gigabytes to 59 trillion gigabytes—that’s nearly 5,000% growth.²

These days, everything that can be turned into digital content or a completely digitized workflow is being digitized. Digitization is happening in movie making, manufacturing, game development, banking, healthcare and more. It’s happening in our libraries. Today’s libraries are taking all of those primary sources, books and magazines and digitizing them so people can access them at home, on their personal devices. This is not only convenient but also necessary, as discovered during pandemic-driven social distancing.

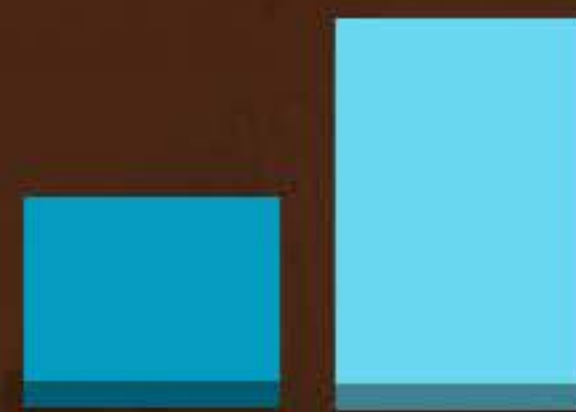
What is creating the massive expansion? Consider these scenarios:²



65% of the world's Gross Domestic Product (GDP) will be digitized by 2022.



Video streaming is predicted to add 52 million users by 2024.



1.4 billion smart home devices will be shipped worldwide in 2024, up from 854 million in 2018.



45% of repetitive work will be automated or augmented by “digital co-workers.”

Only 18% of organizations in a recent survey reported they know how to leverage such data.^{3,4}

As digital experiences are created all around us on such a massive scale, there is the need for IT professionals like you to figure out how to work with this data.

You need to understand:

- what your data is
- who is using it
- how they're using it
- who is creating it

You need insight into:

- how many people are sharing the same data
- what is their interaction with it?
- how is it evolving?

You need to ensure data is accessible to your end users, yet your existing systems may not work in that environment. You can't put 50 petabytes inside of a legacy server. It just doesn't work. You need to determine how to secure your data as it moves back and forth across your entire environment, and that includes the core, the edge and everything in between.

If your organization faces challenges around understanding, moving and organizing unstructured data, you're not alone. Only 18% of organizations in a recent survey reported they know how to leverage such data.^{3,4}

Since a majority of data (80% to 90%, according to multiple analyst estimates) is unstructured information such as text, video, audio, web logs, social media, etc.,⁴ companies that do learn to master this untapped resource will have real potential to create a lasting competitive advantage.

The real struggle of managing unstructured data, especially in certain high-performance industries

By Tim Kresler, Director of Systems Engineering, Qumulo

IDC predicts that 80% of worldwide data will be unstructured by 2025,⁵ so it makes sense that companies continue to look for new ways to store their extreme volumes of data securely. As they do this, they are also seeking ways to scale endlessly, bring applications to market quickly, help customers access and move their data freely and extend beyond existing business limits by leveraging public cloud.

IT professionals in charge of all this find it challenging enough. However, if you happen to work in an industry that creates massive amounts of unstructured data, you may have an extra set of unique challenges.

For instance

- How would you capture and store critical data coming in from a genomic sequencer?
- What would you do with the boundless files you need to store from a high-resolution video camera? Certain industries are leveraging video more than ever—for security purposes, surveillance, tracking, training and more.
- Think of hospitals that need video coverage for every surgery. What do they do with all that data? Where do they store it? How can they secure it or access it later to leverage for research and innovation? These are the questions companies are asking.



A group of people, likely in a gaming or esports arena, are seated at long tables with multiple computer monitors. They are all wearing large, professional-grade headsets with microphones. The room is dimly lit with a strong blue ambient light. In the foreground, a man is smiling and looking towards his monitor. Behind him, a woman and another man are also focused on their screens. The background shows more people and equipment, creating a sense of a large-scale event or competition.

Impacted Industries

Multitudes of companies in most industries are successfully leveraging artificial intelligence (AI) to help them distill meaning from the volumes of unstructured data they generate each day. Companies that are able to capture, track, access, store, manage and secure this unstructured data efficiently are positioned for greater success. Here are a few industries most impacted by unstructured data.

Healthcare

Hospitals often have large-scale data that is difficult to handle within existing database management systems. Massive amounts of unstructured data come from medical video and imaging devices (e.g., endoscopes, surgery robots, laparoscopes, etc.). Biosignal data is another big data generator, with images displayed on patient monitors, in operating rooms, or from patients' wearable monitoring devices. Since it is hard to manage this type of data within the structure of Electronic Medical Records, the data is often ignored, unsaved or abandoned, which results in many missed opportunities.

Media & Entertainment

Almost every aspect of media and entertainment has gone digital, from content creation and production to broadcast and content delivery. Not only do media companies need to support complex, demanding workflows for visual effects and computer-generated film animation, but they also have multitudes of animation and visual effects teams across the world that need instant, secure access. The amount of content being created is increasing as well as the demand for higher resolution content, thus creating endless amounts of unstructured data that needs to be stored, secured, accessed and managed.

Gaming

As one of the most competitive media industries in the world today, gaming companies continually need to accelerate gaming design and production with fast, reliable accessibility to all their data. Rapid scalability is also essential to meeting aggressive production demands. Gaming companies have a lot to gain by collecting and evaluating unstructured consumer data. By studying it, these companies can learn from consumer behaviors and preferences, and use that information to develop games that will be played over and over again. Efficient unstructured data management also allows in-house developers and remote artists to collaborate with multiple studios, working on projects simultaneously to achieve faster turnaround times.

Manufacturing

The amount of unstructured data manufacturers need to manage is growing exponentially. There is a deluge of different types of data, including standard office documents as well as social media, email, engineering CAD files, videos etc. These unstructured files contain a large volume of data that is essential to running a manufacturing business. Manufacturers that are able to successfully manage this unstructured data are often able to improve the quality and consistency of their processes. There is a big focus on compliant-driven data that follows specific quality standards, either for compliance reasons or because a customer has requested it. Effective management of this data helps manufacturers prove that they have followed all necessary quality processes correctly.



The Rise of File Storage and the Public Cloud

By Ben Gitenstein, VP of products and solutions, Qumulo

In a short time, cloud and file data have each gone through significant shifts.

Cloud

A few years ago, the cloud was a big idea that had become very popular, but was still mainly a playground for developers building new applications and a small subset of enterprise workflows. It wasn't the default standard. Since then, the public cloud has rapidly become the standard way in which enterprises run their business; instead of having to prove why you should move a workload to the cloud, today you have to prove why you shouldn't.

File Storage

In the early days of cloud, unstructured data meant object storage and S3 to public cloud providers led by Amazon. The founders of Qumulo did not agree. In fact, they had a fundamental passion for and belief in the importance of file systems as a way to help organizations use and manage data to achieve their objectives and mission. They believed that while object and S3 is fine for a large swath of unstructured data, an important subset of unstructured data should be handled by file services in their native format.

When is File Better than Object?

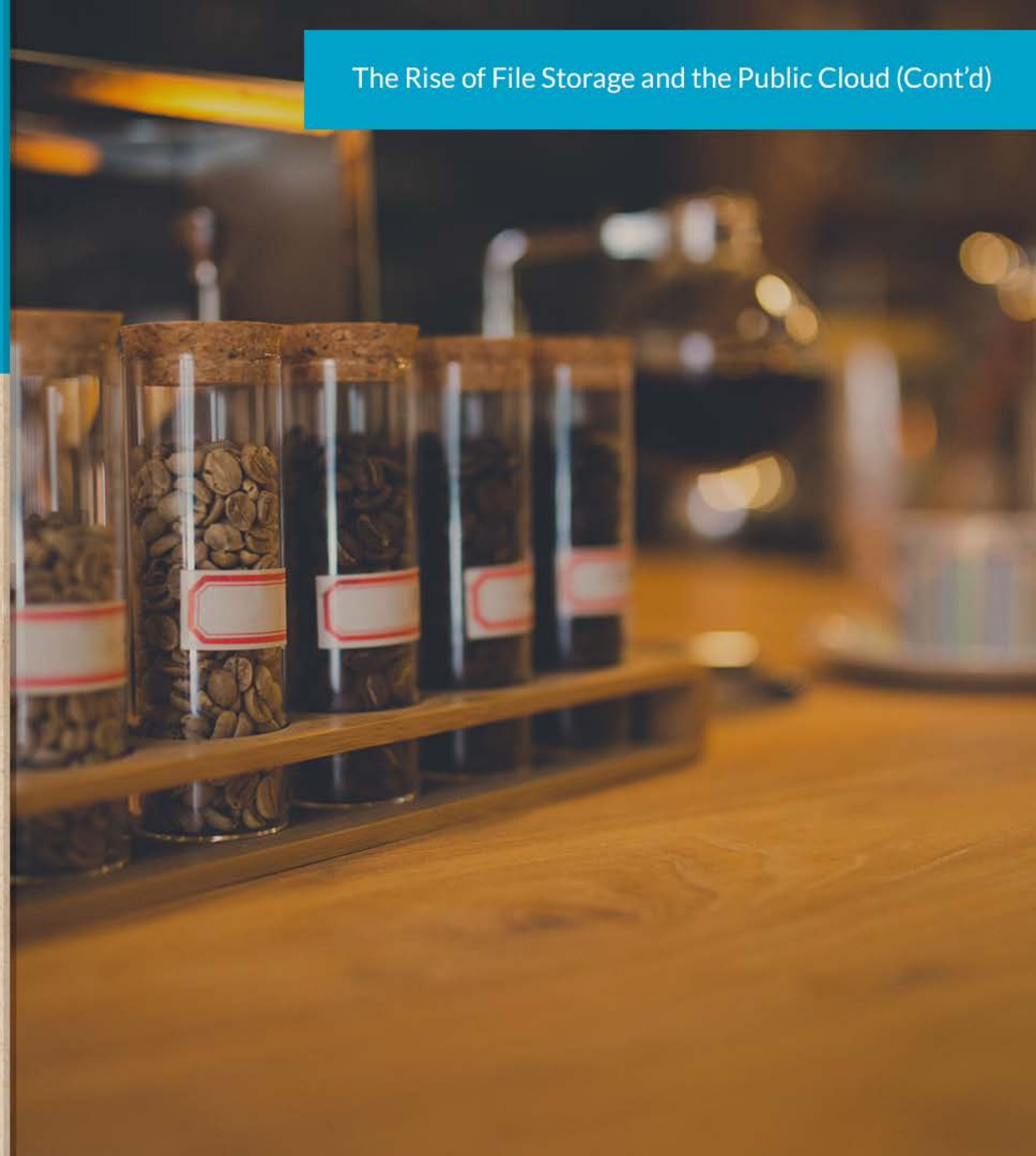
- When you have unstructured data that is actively used, interactive data
- When you need to organize data so people can use it
- When it comes to security, because there's a rich directory ecosystem of security tools
- When you have interactive workflows where you need to make lots of small changes to the data, such as updating logs on a log file from a manufacturing device
- When you need to interoperate with a bunch of applications and end-user clients like windows and Mac.

With the explosion of unstructured data, scale increased. Public cloud providers eventually came to agree that file was better for some unstructured data, yet as object companies at heart, they are still wired to drive all unstructured data into object and S3. By contrast, Qumulo is still full of passionate file believers. The company has never tried to shift to file plus object or file plus block. Its engineering spend and focus remain firmly on file.

For its first few years, the company focused solely on building a file product companies could put in their data centers. More recently, Qumulo's mission shifted to providing first-class file services at scale in the public cloud for workloads that need interactive performance, where end-users need to interact with the data, where performance mattered, and where complicity and experience mattered.

By moving to the cloud, Qumulo has been able to offer many new capabilities to its data center customers, such as speeding time to value, enabling elastic infrastructure and saving money when public cloud providers institute volume price drops. It also allows organizations to openly and freely move their data without vendor lock-in, enabling organizations to take any data from any Qumulo file system, replicate that data into Amazon's S3 service as a native object, and do whatever they want with it, such as saving a second copy as a backup, archiving cold data no longer in use, or collaborating globally.

The Rise of File Storage and the Public Cloud (Cont'd)





Top 10 Benefits of Hybrid Cloud File Storage

Why You Should Consider Hybrid Cloud Architecture



1

IT administrators can optimize their time and resources with storage management that scales across on-prem, hybrid cloud and cloud environments.

Minimize IT resources by managing all storage platforms, on-prem and/or multi-cloud, via the same web interface and using the exact same software. Have standard enterprise file management tools such as quotas and snapshots so that administrators can protect data and avoid cost overruns.

2

Need additional IT resources ASAP? Seamlessly spin up instances in the cloud, and shut them down just as easily.

Quickly and easily expand your compute and storage IT resources in the cloud to handle unexpected or peak storage requirements. Spin-up cloud instances quickly and seamlessly when required for additional resources and shut them down just as easily when no longer needed - all managed through the same web interface.

3

Gain insights and prevent expensive outages with real-time analytics of your data where it resides—on-prem or in the cloud.

Obtain real-time visibility into the state of your data and storage with powerful analytics that provides actionable information. For example, being able to immediately identify, in real-time, an application that is consuming too much storage to immediately set a quota.



4

Cloud file lakes make it easy to store and move file data.

Data stored as file shouldn't have to be converted to object to be able to leverage the cloud. Cloud file lakes, in contrast to cloud data lakes, offer customers the ability to store file data via S3 as it was intended to be stored: as file. The file lake creates a single scalable file namespace in a public cloud, with the features and capabilities of a modern file system. These cloud file lakes enable the file data to easily be moved on-prem for local manipulation to import and export the data to object data lakes.

5

Cost efficiently protect your data.

Easily and automatically move data from on-prem to the cloud, or replicate between multi-cloud environments for efficient data protection.

6

Workflows and security can be streamlined for multi-protocol environments.

Windows, Linux, and Mac clients (and applications) can access the same data without any customization and from the same namespace. Ensure access to data from any server, from anywhere so your people are always able to work.

7

You don't have to sacrifice permissions control in the cloud.

Integrate with Active Directory and LDAP and offer granular permission control (across Windows/Mac/Linux). Also, a detailed audit trail is maintained.

8

With the cloud, you can easily and cost effectively manage your relentless petabyte data growth.

Scale to petabytes, hundreds of GB/s and hundreds of thousands of IOPS in a single namespace. Seamlessly grow the storage environment just by adding an additional node. No down time and no need for a storage specialist to tune the storage.

9

Protect investments while meeting your applications needs.

For on-prem performance of applications with low latency requirements, protect your investments by using the latest, fast storage that will take you into the future. Capitalize on the speed provided by SSDs with all-NVMe file storage.

10

Storage can be an integral part of your overall IT reporting and analytic tools.

Programmable control (APIs) enables the integration of data and storage metrics to ensure it is an integral part of the overall IT management systems and analytic tools. These APIs also enable automation of mundane tasks to optimize your IT administrative resources.





See how ePlus and
Qumulo can help
you harness the
power of your
unstructured data.

[Learn more →](#)

References

1) Gartner Peer Insights, Swift, 2021. 2) "54 predictions about the state of data in 2021," Forbes, December 30, 2020. 3) "Analytics and AI-driven enterprises thrive in the age of with: the culture catalyst," Deloitte Insights, July 2019. 4) "Tapping the power of unstructured data," MIT, February 1, 2021. 5) "The Future of Data Revolution will be Unstructured Data," Analytics Insight, IDC, October 2021.