

IDC MarketScape: Worldwide Distributed Scale-Out File System 2022 Vendor Assessment

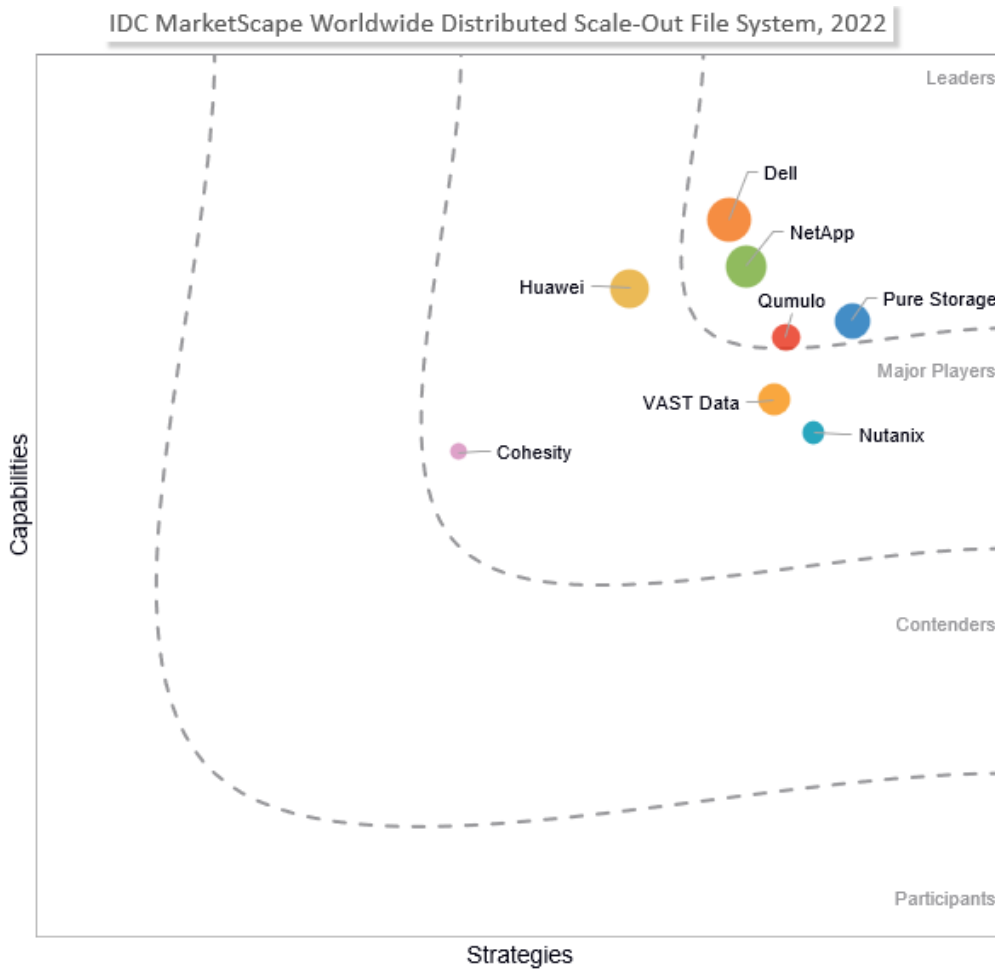
Eric Burgener

THIS IDC MARKETSCAPE EXCERPT FEATURES QUMULO

IDC MARKETSCAPE FIGURE

FIGURE 1

IDC MarketScape Worldwide Distributed Scale-Out File System Vendor Assessment



Source: IDC, 2022

Please see the Appendix for detailed methodology, market definition, and scoring criteria.

IN THIS EXCERPT

The content for this excerpt was taken directly from IDC MarketScape: Worldwide Distributed Scale-Out File System 2022 Vendor Assessment (Doc # US49015322). All or parts of the following sections are included in this excerpt: IDC Opinion, IDC MarketScape Vendor Inclusion Criteria, Essential Guidance, Vendor Summary Profile, Appendix and Learn More. Also included is Figure 1.

IDC OPINION

Over the next five years, scale-out file systems will be widely deployed by enterprises looking to consolidate file-based workloads, improve file-based infrastructure efficiencies, and handle many of the performance and scalability requirements of modernized applications that are very data intensive. All of the products evaluated here will be able to do that very well for most enterprises, although there are some differences in top-end performance and scalability and ease of use between offerings – that is why Figure 1 has many of the vendors clustered closely together. What the reader should note, however, is that there can be significant differences between vendors in their architectures, product strategies, areas of focus, and software-defined flexibility that should be evaluated as purchase decisions are made.

The "Advice for Technology Buyers" section is probably the most important section to read for those who will be involved in making a purchase decision. This section introduces a number of strategic questions enterprises should ask themselves when determining what is most important in selecting a scale-out file system offering. As an example, all evaluated products can support a 1PB file system, but what each system looks like, how easy it is to manage and upgrade, how much it costs and, in general, how it gets there can be very different. There is no "best" offering in this market, but there are certain products that are better suited for certain workloads and will cater better to certain objectives like top-end performance and scalability, ease of use and management, lower energy and floorspace consumption, hybrid cloud capabilities, and how different access methods are supported.

Enterprises can expect a lot more innovation to occur in the scale-out file market going forward, driven primarily by the fact that 80% of the data that will be created over the next five years will be file and/or object based. If enterprises just need to simplify basic file sharing (home directories, etc.), there are a lot of very viable options (some of which are mentioned in the "Vendors to Watch" section). Modernized applications, particularly those using artificial intelligence (AI) or those which are very data intensive, will have additional demands that may not be well met by the simpler products, and that's where enterprises will need to turn to true distributed scale-out file system platforms.

IDC MARKETSCAPE VENDOR INCLUSION CRITERIA

This IDC study assesses the capabilities and business strategies of popular suppliers in the distributed scale-out file-based storage market segment. For a complete definition of distributed scale-out file systems (and a discussion of the new file-based storage taxonomy that IDC introduced in July 2021), see *Reclassifying File Storage – A New Approach for the Future of Digital Infrastructure* (IDC #US48051221, July 2021). This evaluation is based on a comprehensive framework and a set of

parameters that gauge the success of a supplier in delivering a scale-out file-based storage solution to the enterprise market.

To be evaluated in this study, a vendor needs to have a scale-out file-based storage platform:

- **That conforms to IDC's taxonomy.** According to *Reclassifying File Storage – A New Approach for the Future of Digital Infrastructure* (IDC #US48051221, July 2021), assessed products need to meet the definition of a distributed scale-out file system platform or a clustered scale-up file system that is sold primarily against distributed scale-out file systems.
- **Whose intellectual property (IP) is fully owned by the vendor.** The vendor being assessed has developed the distributed scale-out file-based storage solution in-house or obtained the technology through acquisition.
- **That was generally available by September 2021 and generates at least \$30 million in annual revenue.** This is to ensure that the vendor product has at least some level of maturity and market traction.

ADVICE FOR TECHNOLOGY BUYERS

Given that the vendors in this assessment are using widely varying product strategies, an important place to start the evaluation process for an enterprise is to understand which of the different approaches appeal to the enterprise and/or are a better fit for its needs. Do you like the idea of being able to manage block-, file-, and object-based workloads on the same storage system through a unified management interface? Do you prefer unified storage (which can avoid semantic loss issues but will use more storage capacity to provide multiprotocol access to the same data object) or multiprotocol access (which uses less storage capacity but where semantic loss may be an issue)? Are you a federal agency that requires FIPS 140-2 compliant encryption? Do you prefer a storage architecture built around server-based storage nodes or are you open to different architectures that may offer differentiators in certain environments? Six of the vendors assessed use server-based storage nodes (although some of them have some proprietary content), while two – NetApp and Pure – use different architectures.

Would you prefer to use traditional access methods like NFS and SMB but also have access to an intelligent client that offers significantly more parallelization if/when you might need it? Other vendors will tell you how they've extended the performance capabilities of NFS over TCP beyond the 2GBps limit per mount point with nconnect or features specific to their platform that still use the standard NFS client (for example) so you don't have to deploy an intelligent client. Do you require NDMP support? Are you interested in the idea of a cacheless architecture that can offer very high degrees of data concurrency or do more traditional cache-based architectures meet your needs just fine? Do you need POSIX compliance? POSIX really isn't the future, but there are hundreds of thousands of already deployed applications that use it.

Do you have a preference for an HCI-based architecture (like Cohesity or Nutanix) or a disaggregated storage approach? Do you want to buy your solution from a major OEM (Cisco sells Cohesity, Dell sells Nutanix, and HPE sells Qumulo) or would you prefer to buy it from the developing vendor directly (or a channel partner of theirs)? Do you like the idea of combining data protection and enterprise file sharing under a single system or not? While this is not an exhaustive list of questions, these are the kinds of questions an IT manager should ponder when evaluating scale-out file systems for enterprise workloads.

As with most enterprise workloads, high availability (HA) is important and enterprise file sharing is no exception. Solutions that have been around for a long time tend to have an extensive, proven feature set in this area. Understand your recovery point objectives (RPOs) and recovery time objectives (RTOs) for both local and disaster recovery, and match that with capabilities in the scale-out file system offerings. Tunable erasure coding (EC) (so data durability and capacity utilization can be set differently for different workloads), snapshots, replication, a simple "snap to object" feature that makes it very easy to back up the entire namespace to an external object store, air-gap protection to defend against ransomware, and integration with third-party backup products like Commvault and Veritas, all these are features that can impact data protection workflows, availability, and recovery times.

Ease of management at scale is another differentiating area. There are many challenges in managing scale-out file system environments, and there has been a lot of employee interchange between the various scale-out file system players in the past 20 years. The challenges are well known at all vendors, but how they address them varies. If you have managed a scale-out file system before, what are your hot-button issues?

- Do you need absolutely the lowest latencies for random small file accesses or are sub-millisecond average response times good enough?
- Are you trying to consolidate workloads across your data stage pipelines that need both native and intelligent client-based access methods?
- Do you want to be able to rapidly create delta differentials for backup purposes without having to walk all the file trees?
- Do you want particularly low-capacity utilization of on-disk data protection options at your target level of durability because you have multiple petabytes of data under management?
- Do you need support for compression and/or deduplication because your data sets can benefit significantly from these technologies (or not, since much unstructured data does not compress and/or deduplicate very well)?
- Are disruptive upgrades and slow disruptive recovery in SMB environments a particular pain point?
- Are you particularly concerned about large capacity drive rebuild times or how easy and nondisruptive it is to expand the cluster by adding a new node?
- Are you concerned about how easy and efficient is it to use file quota management systems?

These (and many more) are all issues many scale-out file system administrators have struggled with.

The key to selecting a platform best suited for your requirements is to thoroughly understand your needs and preferences up front. The vendors assessed here all provide a range of performance, scalability, availability, and core functionality that meet the requirements for most enterprise file-based workloads, but among the eight vendors, there are very different ways to get there and very different emphases in their product designs. List what is most important to you, and map that to the vendor offerings. Doing that will require going beyond this document since we do not provide direct head-to-head comparisons between vendors. IDC has, however, published a number of technical reviews of different vendor offerings in separate research, discussing the benefits of the approaches they have taken.

VENDOR SUMMARY PROFILES

This section briefly explains IDC's key observations resulting in a vendor's position in the IDC MarketScape. While every vendor is evaluated against each of the criteria outlined in the Appendix, the description here provides a summary of each vendor's strengths and challenges.

Qumulo

Qumulo is positioned in the Leaders category in the 2022 IDC MarketScape for worldwide distributed scale-out file system.

Founded in 2012, Seattle-based Qumulo is a distributed scale-out file system vendor that has been shipping its Qumulo Core file system platform since 2015. Although privately held, Qumulo reached unicorn status with a \$1.2 billion valuation in July 2020 and its recent FY 4Q22 results indicated 75% sequential growth over its third quarter results. The vendor completed its sixth round of funding in July 2020, bringing the total raised to \$351 million. Qumulo sells its file system software on a subscription basis. Nearing 1,000 customers, the vendor is significantly benefiting from digital transformation in enterprises struggling to implement a more efficient strategy for unstructured data management at scale.

Strengths

Differentiators that define Qumulo include flexibility of deployment (because of its software-defined design), simplicity and efficiency of management at petabyte scale (because of features specifically designed to address well-known scale-out file system issues with first-generation distributed scale-out file system platforms), and real-time analytics (that provide comprehensive visibility into file system metrics that enable more effective management of large-scale environments).

Qumulo Core is a software product that can be deployed on commodity server-based storage from a variety of different suppliers and is available as an appliance from channel providers like Arrow, Fujitsu, and HPE and as a cloud-based file storage service from AWS, Microsoft Azure, and Google Cloud. Its software-defined nature supports mixed storage cluster configurations with NVMe and hybrid nodes, delivers exactly the same functionality in on-premises or public cloud-based deployments, and easily accommodates new storage device types and multigenerational technology refresh. Qumulo cloud instances (referred to as Cloud Q, while its on-premises deployments are referred to as Server Q) differentiate themselves from native cloud-based file services on scalability, multiprotocol support, and data visibility (through their onboard real-time analytics capabilities). The vendor offers a unified management console that spans on-premises and off-premises deployment locations for easier management in hybrid multicloud environments.

Architecturally, the system's scalability in terms of file sizes and file system sizes is impressive. Although the company has not validated it for use in production, it claims that its architecture supports single file sizes up to 9EB and unlimited file system sizes. Qumulo does certification work for production use based on customer requirements though, and to date, the vendor has certified a 40PB cluster size (all in a single namespace). These numbers stand in stark contrast to file and file system size limitations of some of the vendors assessed in this document.

While all scale-out file system vendors claim ease of management, Qumulo can point to a variety of specific features that are very attractive to administrators already well versed in managing these types of environments – how quota management is handled with an in-band method that is much more

efficient at keeping it in sync with the actual file system, how namespacewide delta differentials are generated within 15-30 seconds for backup purposes, the use of heat metric-based intelligent data placement to optimize cache hit rates, and a number of optimizations that improve the system's ability to efficiently handle billions of small files. (Note that it is too detailed to list here but definitely an area that should be explored by technical decision makers for whom that is important.)

On-disk data protection is purely software defined and operates at the block (rather than the file) level using a flexible EC-like approach (the vendor's platform includes an integrated volume manager/file system layer). This not only enables much faster drive rebuilds but lets customers mix and match different device sizes in the system as well. In addition to these capabilities, Qumulo supports a variety of other features like snapshots, replication, changed block tracking that does not require B-tree walks, air-gap protection to defend against ransomware, and integration with third-party backup products to support high availability and fast recovery. Qumulo is widely used in its installed base for mission-critical file-based workloads.

Qumulo is fully committed to using native access methods like NFS, SMB, and HTTP instead of intelligent clients because of the ease-of-use differences. It uses a multiprotocol access design that maximizes capacity utilization, uses cross-protocol permissions and identity management, and uses 256-bit encryption for all data at rest. Replicated data in flight is encrypted, as is all access using the SMB protocol. The vendor is currently in the process of obtaining FIPS 140-2 certification from the U.S. government. Rich REST APIs enable easy workflow automation integration.

Challenges

While no longer a start-up, Qumulo is still a relatively small player. All Qumulo fulfillment flows through channels, but the vendor does provide a single point of support contact for all its different appliance-based models. Alternatively, customers can also buy its product through established storage OEMs like HPE if there is concern about the vendor's size.

Some features that other vendors offer are missing. Qumulo does not yet support compression and data deduplication, and while it is likely to introduce compression at some point, Qumulo does not intend to introduce dedupe – primarily because the data sets its customers manage rarely benefit from it. S3 is not a supported access method, although the vendor plans to remedy that by the end of 2022. Some multitenant management features like access zones (which allow administrators to partition a cluster into multiple virtual containers to isolate data) are not supported, so customers may want to closely review the feature set to ensure that Qumulo's capabilities meet their requirements.

Qumulo is a firm believer in software-defined systems and tends not to support hardware assist for features like encryption (although it does support that in its HPE Apollo-based appliance). And while it is not necessarily a challenge, Qumulo does not aim to offer the lowest latencies, the highest throughput to a single file or a platform well suited for cheap and deep cold archives – it targets the large middle of the file storage market with a system that offers very flexible deployment options, is easy to manage at scale, and provides comprehensive data visibility to inform better data management.

Consider Qumulo When

Qumulo Core is a good fit for not only more traditional file sharing but also the newer big data analytics workloads being deployed as part of digital transformation that use NFS and/or SMB. Enterprises with hybrid multicloud environments will appreciate the deployment flexibility and cloud support capabilities,

and administrators with prior experience with distributed scale-out file systems will appreciate many of the ease-of-use features (including in particular the visibility enabled by the Qumulo Aware component) that really prove their worth as data under management scales to a petabyte and beyond. Key verticals where the vendor has enjoyed significant success include advertising, media and entertainment, manufacturing, technology, software and telecom, health and biotechnology, and professional, technical, and business services. Qumulo already also has a good federal business but expects that to ramp further once it achieves FIPS 140-2 certification by the end of 2022.

APPENDIX

Reading an IDC MarketScape Graph

For the purposes of this analysis, IDC divided potential key measures for success into two primary categories: capabilities and strategies.

Positioning on the y-axis reflects the vendor's current capabilities and menu of services and how well aligned the vendor is to customer needs. The capabilities category focuses on the capabilities of the company and product today, here and now. Under this category, IDC analysts will look at how well a vendor is building/delivering capabilities that enable it to execute its chosen strategy in the market.

Positioning on the x-axis, or strategies axis, indicates how well the vendor's future strategy aligns with what customers will require in three to five years. The strategies category focuses on high-level decisions and underlying assumptions about offerings, customer segments, and business and go-to-market plans for the next three to five years.

The size of the individual vendor markers in the IDC MarketScape represents the market share of each individual vendor within the specific market segment being assessed, not the overall storage-related revenue of the vendor.

Several suppliers offer different file system offerings, although they do not all necessarily compete in the distributed scale-out file system segment. In cases where the vendor offers two scale-out file system types, IDC has worked with the vendor to select the product that most closely fits within the inclusion criteria of this study.

IDC MarketScape Methodology

IDC MarketScape criteria selection, weightings, and vendor scores represent well-researched IDC judgment about the market and specific vendors. IDC analysts tailor the range of standard characteristics by which vendors are measured through structured discussions, surveys, and interviews with market leaders, participants, and end users. Market weightings are based on user interviews, buyer surveys, and the input of IDC experts in each market. IDC analysts base individual vendor scores, and ultimately vendor positions on the IDC MarketScape, on detailed surveys and interviews with the vendors, publicly available information, and end-user experiences in an effort to provide an accurate and consistent assessment of each vendor's characteristics, behavior, and capability.

Market Definition

In July 2021, IDC introduced a new taxonomy for the file system market. There are four segments to the file system market: scale-up file storage, scale-up clusters, distributed scale-out file storage, and parallel scale-out file storage. The scale-up segment is small and shrinking in size, while all the growth

is being driven by scale-out products. Briefly, scale-out file systems distribute data across nodes while presenting a single data access namespace. There are some differences, however, in how data is distributed between scale-up clusters and scale-out file storage. In scale-up clusters, data is rarely ever distributed across nodes, and the throughput to a given file is limited to the bandwidth of the single node from which it is served. In scale-out clusters, data in a single file can be distributed across nodes, a design which can improve access performance, data concurrency, and recovery time.

Scale-up clusters and distributed scale-out file storage routinely compete for the same business in enterprises, and this vendor assessment includes vendors from both segments. For more detail on how each of these segments is defined, see *Reclassifying File Storage – A New Approach for the Future of Digital Infrastructure* (IDC #US48051221, July 2021).

LEARN MORE

Related Research

- *Qumulo Exhibiting Strong Momentum as It Serves the Evolving Unstructured Data Storage Needs of Enterprise Customers* (IDC #US48896622, March 2022)
- *VAST Data: A Technical Deep-Dive Look at a Compelling New Scale-Out Storage Architecture* (IDC #US48805222, February 2022)
- *Worldwide File- and Object-Based Storage Forecast, 2021-2025: New Enterprise Workloads Driving Strong Growth* (IDC #US48403021, December 2021)
- *Hyperconverged Infrastructure Adoption Trends – 3Q21: Building Block for Hybrid Cloud Infrastructure* (IDC #US48308121, October 2021)
- *Enterprise Workloads Resulting in Broader Adoption of Scale-Out File Storage Architectures* (IDC #US48305121, October 2021)
- *How to Compare Distributed Scale-Out File Storage Platforms for Use with Enterprise Workloads* (IDC #US48191621, September 2021)
- *Reclassifying File Storage – A New Approach for the Future of Digital Infrastructure* (IDC #US48051221, July 2021)

Synopsis

This IDC study represents a vendor assessment model called the IDC MarketScope. This study is a quantitative and qualitative assessment of the characteristics that assess a vendor's current and future success in the relevant market or market segment and provide a measure of the vendor's ability to become a leader or maintain leadership.

The distributed scale-out file system market segment, which is part of the file-based storage market, is an example of a large, maturing market that is still exhibiting low double-digit growth. This document assesses the capabilities and strategies of key vendors of scale-out file-based platforms. While seven of the assessed vendors have distributed scale-out file system designs, one of the vendors (NetApp) actually uses a scale-up cluster design but still meets the inclusion criteria of this vendor assessment study.

"While all evaluated vendors tout the performance, scalability, and ease of use of their file-based storage offerings, a closer look reveals important distinctions in how vendors define these metrics and build their products to achieve them," said Eric Burgener, research vice president, Infrastructure Systems, Platforms and Technologies Group, IDC. "To select the right product, enterprises need to

understand the architectural differences between the different vendor approaches, understand the implications of those choices for their workloads, and then choose the product which best fits their requirements."

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

140 Kendrick Street
Building B
Needham, MA 02494
USA
508.872.8200
Twitter: @IDC
blogs.idc.com
www.idc.com

Copyright and Trademark Notice

This IDC research document was published as part of an IDC continuous intelligence service, providing written research, analyst interactions, telebriefings, and conferences. Visit www.idc.com to learn more about IDC subscription and consulting services. To view a list of IDC offices worldwide, visit www.idc.com/offices. Please contact the IDC Hotline at 800.343.4952, ext. 7988 (or +1.508.988.7988) or sales@idc.com for information on applying the price of this document toward the purchase of an IDC service or for information on additional copies or web rights. IDC and IDC MarketScape are trademarks of International Data Group, Inc.

Copyright 2022 IDC. Reproduction is forbidden unless authorized. All rights reserved.

