Critical Capabilities for Object Storage

Published 21 October 2020 - ID G00465540 - 32 min read

By Analysts Chandra Mukhyala, Julia Palmer, Jerry Rozeman, Robert Preston

Initiatives: Data Center Infrastructure

Growing adoption of the Amazon S3 protocol by application vendors is expanding interest in on-premises object storage to store static data. In this research, we evaluate 11 object storage products against eight critical capabilities in use cases relevant to I&O leaders.

This Critical Capabilities is related to other research:
Magic Quadrant for Distributed File Systems and Object Storage

Overview

Key Findings

- Traditional enterprise applications that were dependent on NFS and SMB protocols are increasing their adoption of the Amazon S3 protocol to benefit from the scalability and geographical distribution available in object storage.

- Given the rise of the Amazon S3 protocol as a de facto access protocol for object storage, distributed file storage vendors are also offering S3 protocol access for objects stored in file systems.

- With increasing options for low-cost storage tiers and granular security, public cloud continues to increase its attractiveness as a viable alternative to on-premises object storage.

- Object storage vendors continue to add file protocol access, but the file access mostly remains a gateway and is not scalable and resilient like the underlying object storage.

Recommendations

I&O leaders responsible for unstructured data infrastructure should:

- Evaluate object storage systems for data that is mostly read-only once the data is written to eliminate the performance issues involved in modifying objects.

- Deploy object storage for applications that support the Amazon S3 API, instead of file interfaces, unless the object storage system has a scalable native file support to fully benefit from the scalability of the object store.
Strategic Planning Assumptions

By 2024, large enterprises will triple their unstructured data stored as file or object storage on-premises, at the edge or in the public cloud, compared with 2020.

By 2024, 40% of I&O leaders will implement at least one hybrid cloud storage capability, up from 15% in 2020.

By 2024, 50% of the global unstructured data storage capacity will be deployed as software-defined storage (SDS) on-premises or in the public cloud, up from less than 20% in 2020.

What You Need to Know

Many data storage requirements that object storage systems were created for originally are now also being met by distributed file systems on-premises, except at a smaller scale. In addition, public cloud continues to be an attractive alternative for on-premises object storage due to new lower-cost storage tiers and the overall agility of public cloud.

Key requirements that object storage systems were created for include:

- **An HTTP-based protocol for accessing the underlying objects, not just locally but also from geographically distributed locations.** Amazon Simple Storage Service (S3) has become the de facto standard for web-based access. Given S3’s popularity among application developers, many distributed file systems are also offering the same access protocol. Developers dealing with massive amounts of unstructured data like the simple means by which they can put and get objects using the S3 protocol. In 2019, many independent software vendor (ISV) applications started supporting S3 as an access protocol. Initially, backup and archive ISV applications started supporting S3 interfaces. But the same trend is now also seen in industry-specific applications, such as picture archiving and communication system and media asset management applications, where Network File System (NFS) and/or Server Message Block (SMB) used to be the only access protocols supported. High-performance analytics applications are also now supporting the S3 protocol to write persistent data to object storage after processing the data in-memory.

- **Low-cost storage from leveraging industry-standard x86 servers.** Object storage systems typically adopt some form of software-based erasure coding to protect from drive, node or other hardware failures. This method of data protection provides a more resilient system against
What remains unique to object storage and difficult for file storage systems to emulate is the elimination of the overhead involved in the number of operations required to write or read an object. Objects are written in a flat namespace, making them extremely scalable without the overhead of managing file system trees or the overhead of file locking present in file systems. This scalability can easily expand across geographical locations, making them the ideal storage system for collaboration use cases. In addition, object storage is ideally suited for applications that can leverage metadata associated with an object. Object storage allows for a variable amount of metadata that can be attached to the object to uniquely identify an object and to fetch it very quickly from massive unstructured data volumes.

The number of applications that require the massive global scalability or leverage custom metadata remains small in the on-premises world. This in turn makes the need for object storage systems relatively small compared with distributed file systems. Distributed file systems now support the S3 protocols and software-based protection from disk and hardware failures using erasure coding. Consequently, distributed file systems address two of the four reasons that object storage products were originally developed for.

Infrastructure and operations (I&O) leaders should pay careful attention to what capabilities they need from a distributed storage system and make careful assessment of what storage is best-suited for a given application — distributed file or object. They should evaluate:

- Distributed file systems, rather than object storage products, for workloads that frequently modify the underlying objects
- Object storage products for use cases that require massive scaling across geographical locations, and for applications leveraging custom metadata describing the objects

I&O leaders who need highly scalable, self-healing and cost-effective storage platforms for large amounts of unstructured data should evaluate the suitability of object storage platforms. They should use this research to identify appropriate products for their use cases.

**Analysis**

**Critical Capabilities Use-Case Graphics**

*Figure 1: Vendors’ Product Scores for Analytics Use Case*
### Product or Service Scores for Analytics

<table>
<thead>
<tr>
<th>Vendor</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloudian</td>
<td>4.23</td>
</tr>
<tr>
<td>NetApp</td>
<td>4.18</td>
</tr>
<tr>
<td>IBM</td>
<td>4.03</td>
</tr>
<tr>
<td>Huawei</td>
<td>4.02</td>
</tr>
<tr>
<td>Hitachi Vantara</td>
<td>3.94</td>
</tr>
<tr>
<td>Scality</td>
<td>3.79</td>
</tr>
<tr>
<td>Dell Technologies</td>
<td>3.75</td>
</tr>
<tr>
<td>Red Hat</td>
<td>3.71</td>
</tr>
<tr>
<td>Quantum</td>
<td>3.53</td>
</tr>
<tr>
<td>Caringo</td>
<td>3.17</td>
</tr>
<tr>
<td>DDN</td>
<td>3.02</td>
</tr>
</tbody>
</table>

As of 20 August 2020

Source: Gartner (October 2020)

**Figure 2: Vendors’ Product Scores for Archiving Use Case**
<table>
<thead>
<tr>
<th>Product or Service</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloudian</td>
<td>4.21</td>
</tr>
<tr>
<td>NetApp</td>
<td>4.19</td>
</tr>
<tr>
<td>Huawei</td>
<td>4.02</td>
</tr>
<tr>
<td>IBM</td>
<td>4.01</td>
</tr>
<tr>
<td>Hitachi Vantara</td>
<td>4.00</td>
</tr>
<tr>
<td>Dell Technologies</td>
<td>3.78</td>
</tr>
<tr>
<td>Scality</td>
<td>3.78</td>
</tr>
<tr>
<td>Red Hat</td>
<td>3.73</td>
</tr>
<tr>
<td>Quantum</td>
<td>3.52</td>
</tr>
<tr>
<td>Caringo</td>
<td>3.16</td>
</tr>
<tr>
<td>DDN</td>
<td>3.00</td>
</tr>
</tbody>
</table>

As of 20 August 2020

Source: Gartner (October 2020)

Figure 3: Vendors’ Product Scores for Backup Use Case
## Product or Service Scores for Backup

<table>
<thead>
<tr>
<th>Vendor</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloudian</td>
<td>4.20</td>
</tr>
<tr>
<td>NetApp</td>
<td>4.17</td>
</tr>
<tr>
<td>Huawei</td>
<td>4.02</td>
</tr>
<tr>
<td>IBM</td>
<td>4.01</td>
</tr>
<tr>
<td>Hitachi Vantara</td>
<td>3.94</td>
</tr>
<tr>
<td>Scality</td>
<td>3.82</td>
</tr>
<tr>
<td>Dell Technologies</td>
<td>3.73</td>
</tr>
<tr>
<td>Red Hat</td>
<td>3.72</td>
</tr>
<tr>
<td>Quantum</td>
<td>3.51</td>
</tr>
<tr>
<td>Caringo</td>
<td>3.17</td>
</tr>
<tr>
<td>DDN</td>
<td>3.05</td>
</tr>
</tbody>
</table>

As of 20 August 2020

Source: Gartner (October 2020)

**Figure 4: Vendors’ Product Scores for Cloud Storage Use Case**
## Product or Service Scores for Cloud Storage

<table>
<thead>
<tr>
<th>Vendor</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloudian</td>
<td>4.23</td>
</tr>
<tr>
<td>NetApp</td>
<td>4.17</td>
</tr>
<tr>
<td>IBM</td>
<td>4.03</td>
</tr>
<tr>
<td>Huawei</td>
<td>4.02</td>
</tr>
<tr>
<td>Hitachi Vantara</td>
<td>3.94</td>
</tr>
<tr>
<td>Scality</td>
<td>3.82</td>
</tr>
<tr>
<td>Dell Technologies</td>
<td>3.75</td>
</tr>
<tr>
<td>Red Hat</td>
<td>3.73</td>
</tr>
<tr>
<td>Quantum</td>
<td>3.51</td>
</tr>
<tr>
<td>Caringo</td>
<td>3.16</td>
</tr>
<tr>
<td>DDN</td>
<td>3.01</td>
</tr>
</tbody>
</table>

As of 20 August 2020

Source: Gartner (October 2020)

**Figure 5: Vendors’ Product Scores for Hybrid Cloud Storage Use Case**
**Vendors**

**Caringo**

Caringo's object storage platform, Swarm, is a key-value object store that does not depend on any underlying file system, giving it a highly scalable architecture. Swarm is often paired with FileFly, a Windows application for migrating file-based data to and from Windows and NetApp filers. SwarmNFS allows ingesting of data over NFS into Swarm as objects that can be accessed over S3, Hadoop Distributed File System (HDFS), SMB and NFS. Caringo Drive is a virtual drive that allows dropping files from Windows or macOS into Swarm. Swarm's access control settings are widespread, with support for cluster, tenant, domain, bucket and individual objects.

Swarm is a good fit for governance-oriented archiving because the product has support for write once, read many (WORM), content authenticity and legal hold capabilities. Caringo Swarm can start small at 5TB and has production deployments going above 42PB.

New in the past 12 months is Caringo Swarm 11, which introduced Partial File Restore (PFR). PFR takes a clip of a large video file for rapid turnaround of a specific portion of a video directly from the archive layer for editing, internal sharing or streaming.
Caringo positions the object storage capabilities of Swarm for cloud storage, archiving and backup use cases, as well as HDFS analytics. Caringo Swarm scored below average across all the use cases relative to other products evaluated in this report.

Cloudian

Cloudian HyperStore is a scale-out native object storage platform that is available as a software-only offering. It runs on commodity hardware from Cisco, Hewlett Packard Enterprise (HPE), Lenovo, Seagate, Quanta Cloud Technology (QCT) and Supermicro and can be deployed as a virtual machine (VM) or as a fully integrated HyperStore appliance. HyperStore also supports file services through an external HyperFile network-attached storage (NAS) gateway controller. The solution is implemented by deploying individual nodes into a logical solution that supports both a local or geodistributed implementation.

HyperStore scores well across all capabilities. It supports a broad ecosystem of server vendors, maintains high compatibility with the S3 API and has a well-designed dashboard. HyperStore also supports dissimilar hardware configurations while supporting very-high-density nodes. It supports geodistributed erasure coding, as well as data efficiency features, such as deduplication and compression, while supporting all-flash configurations together with performance management capabilities like quality of service (QoS). In addition, HyperStore is highly flexible as it can be deployed as a native appliance or software-only solution.

However, Cloudian lacks a native, integrated file system capability that doesn't require the deployment of the HyperFile NAS controllers on top of the HyperStore object storage platform. In addition, the HyperFile NAS controllers do not offer the same scale-out capabilities as the HyperStore nodes.

Cloudian HyperStore has the richest feature set among the products compared in this report, leading to highest scores for critical capabilities and the top rank in all use cases. HyperStore is mostly deployed in backup and archiving and private cloud storage use cases.

DDN

DDN’s WOS platform is an object storage offering that runs on-premises and was released in 2016. DDN positions WOS for long-term repository and typically as a tier for its distributed file storage offering, EXAScaler. WOS is a true object storage system in that it does not rely on any file system or file-system-like operations to the underlying disk, which minimizes disk operations with as little as a single-disk operation for reads, and two for writes.

DDN’s WOS scales out in clusters of up to 256 nodes, and it supports clustering of up to 8,192 nodes across multiple geographies. Users can combine up to 32 clusters to build out an exabyte namespace, distributed geographically. WOS supports highly configurable erasure coding to meet varying availability and performance needs. It is accessible over its own RESTful API and also supports Amazon S3, but S3 implementation uses Apache HBase as an external metadata store that is separate from WOS’s native metadata store. Effectively, DDN has written a new object store
around its existing object store, which makes scalability beyond single-site deployments challenging. In addition, WOS’s S3 implementation does not support server-side encryption.

A key use case for DDN WOS is archiving from EXAScaler. While DDN continues to sell WOS, mainly to existing WOS customers, there has been no feature development on the platform since 2019. DDN is focusing all object storage development on EXAScaler front- and back-end connectivity.

**Dell Technologies**

Dell EMC ECS is a scale-out object storage offering that can be deployed on-premises on purpose-built, Dell EMC PowerEdge-based or third-party qualified servers. With VMware Cloud Director support, third-party service providers can offer ECS as a service for VMware customers. ECS supports both object and file access through NFS, but it is typically accessed through object interfaces such as S3. It is sold across all verticals, with customers ranging from very large enterprises to small and midsize enterprises. Dell Technologies is one of the few vendors with separate products for file storage and object storage, and ECS is frequently positioned as a long-term storage repository to complement PowerScale high-performance file storage.

ECS is a software-based scale-out object storage platform, and while it does not have architectural limits on scaling the number of nodes or objects, the current maximum raw capacity supported per node is 1,080TB. ECS uses erasure coding for protection from disk and other hardware failures, but it supports only two configurations of 12+4 and 10+2 data and parity chunks. ECS supports in-line compression but has no native support for deduplication. It also supports in-line encryption but not self-encrypting drives. ECS lacks the native ability to migrate data between different classes of storage within ECS or to tier data from ECS to public clouds. However, DataIQ provides this functionality at no incremental cost to the customer. In addition, it cannot be deployed in any of the major public clouds. ECS supports the Amazon S3 protocol for object access, but it does not currently support bucket notifications. In terms of performance, ECS currently does not support all-flash models and has no support for QoS.

While ECS ranked at the same position across almost all use cases, it is positioned more frequently for archive, backup and global content sharing use cases. Cloud-native applications leveraging the S3 interface continues to be a growing use case for ECS.

**Hitachi Vantara**

Hitachi Vantara’s Hitachi Content Platform (HCP) portfolio is a mature solution that has been on the market for 14 years and is used in production environments as large as 75PB in a single namespace. Hitachi has four offerings in its portfolio, with HCP at the center as the object storage solution. In addition to the core product, Hitachi offers HCP Anywhere for file synchronization and sharing and for end-user data protection. HCP Gateway and HCP Anywhere Edge are cloud file gateways, and Hitachi Content Intelligence provides search and analytics insights.
HCP can be delivered in a variety of deployment models, ranging from a fully integrated appliance and nodes connected to arrays, to commodity-based, software-only and delivered as a managed service. In February 2019, Hitachi Vantara launched its first-generation all-flash HCP configuration to improve performance for latency-sensitive workloads. Hitachi Vantara’s Everflex utility-style license model for customers deploying in the cloud or in bare-metal environments is charged per terabyte of data stored for one month.

HCP scores higher for archive and hybrid cloud use cases than other use cases because of high scores for security and multitenancy. The remaining use cases are also a good fit for HCP, and have above-average scores.

**Huawei**

Huawei OceanStor 100D is a recent rebrand of Huawei FusionStorage. While here we evaluate only OceanStor 100D object capability, OceanStor 100D also supports block and HDFS interfaces, with native file protocols becoming available in the second half of 2020. OceanStor 100D is deployed on-premises in enterprise data centers, but it also serves as the underlying storage layer for Huawei’s hyperconverged infrastructure (HCI) platform and Huawei’s public cloud infrastructure as a service (IaaS) offerings. The product is most frequently deployed as a preintegrated appliance and is popular in the service provider, government and financial sectors. Each cluster can start with three nodes and scale up to 4,096 nodes in the cluster. The OceanStor 100D administrator interface is a single pane of glass to manage all storage interfaces, including object storage APIs. OceanStor 100D natively supports the ability of tiering to public cloud storage, but it is limited to Amazon Web Services (AWS), Alibaba and Huawei Cloud. OceanStor 100D does not have multisite replication (currently in development) or data tiering between different classes of storage in the cluster.

Over the last 12 months, the product has benefited from major R&D investment and added features. These include Active Directory and LDAP authentication, artificial intelligence for IT operations (AIOps) capabilities, replication, tiering to Blu-ray storage, and data compression. In addition, OceanStor 100D now has high-density 5U 120 disks nodes and the ability to have all-flash configurations. OceanStor 100D currently is sold mainly as block and object storage, including providing HDFS storage in some projects. In the future, after the file protocol and protocol interoperability are supported and mature, Huawei will position OceanStor 100D as a unified block/file/object/HDFS flagship platform incorporating capabilities of the OceanStor 9000 product line.

Huawei positions the object storage capabilities of OceanStor 100D for cloud storage, archiving and backup use cases, as well as HDFS analytics. Huawei OceanStor 100D performed solidly across all the use cases of object storage.

**IBM**

IBM Cloud Object Storage (COS) is highly scalable cloud-native object storage that is also the underlying object storage for IBM Public Cloud. IBM COS has one of the largest deployments of
A key strength of IBM COS is its geodispersed erasure coding, which can be configured for an entire site failure without any data loss. IBM COS is available as IBM branded appliances and on certified hardware from all leading x86 server vendors that is sold on-premises, but it is also available as a service through IBM Cloud.

IBM COS, like most other object storage products in this market, benefits from the growth of Amazon S3 protocol adoption among the ISVs that manage unstructured data. IBM COS's primary advantage is that it gets visibility into public cloud storage requirements, and the corresponding features also become available for on-premises users to support cloud-native workloads in addition to long-term repository use cases. IBM COS excels on scalability, large-scale geographically dispersed erasure coding and support for cloud-native workloads. Where it lacks is a strong storage efficiency story because of the absence of common enterprise features such as compression and deduplication. IBM COS does not have performance optimization features for flash media as a caching tier, nor does it support mixed-flash or hybrid flash configurations. In addition, IBM COS does not have any native support for file access through common file access protocols such as NFS or SMB.

IBM COS ranks higher for hybrid cloud storage, cloud storage and analytics use cases than other use cases because of high scores for interoperability. It is certified with a long list of analytics applications and other common ISV applications for unstructured data. The remaining use cases are also a good fit for IBM and have scores that are close to the top use-case scores.

**NetApp**

NetApp StorageGRID is a mature object storage platform that has been in production for more than 19 years. NetApp StorageGRID is available as a virtual or physical appliance. StorageGRID can be deployed on engineered appliances, VMware environments or in bare-metal (Docker) deployments on Red Hat Enterprise Linux (RHEL), CentOS, Debian or Ubuntu Linux distributions. Additionally, StorageGRID can be deployed on NetApp’s HCI offering. StorageGRID can start small with three nodes at 36TB and can support up to 560PB in a single namespace. Multiple namespaces can be federated for scaling to additional capacity. StorageGRID erasure-coding profiles allow configuring an erasure-coding scheme, as well as a storage pool (the storage pool where erasure-coded object data will be stored). StorageGRID supports any combination of multiple erasure-coding schemes (both single site and multisite), as well as replicas for flexible data protection.

NetApp FabricPool enables tiering of cold data from a performance tier on an ONTAP all-flash array that supports distributed file access (NFS v.3, NFS v.4, SMB) to a cost-effective capacity tier on StorageGRID. StorageGRID provides compression and encryption features, and it depends on NetApp’s core product ONTAP to provide file interfaces and storage efficiency from deduplication.

StorageGRID supports up to 16 sites in a single cluster. By using cloud tiering (Cloud Storage Pools) and bucket replication (CloudMirror), StorageGRID can replicate data to additional StorageGRID or other S3-compatible clouds. Information life cycle management (ILM) rules can be configured to manage individual objects, buckets and containers, and tenants to achieve whatever
data protection or performance objective is required. While StorageGRID supports any enterprise-grade load balancer, it provides a high-availability load balancer that supports advanced QoS and monitoring features.

StorageGRID depends on NetApp’s core product ONTAP to provide file interfaces and storage efficiency features, such as compression and deduplication. NetApp positions StorageGRID for cloud storage, archiving and backup use cases, as well as HDFS analytics. NetApp StorageGRID scored second place in all the use cases of object storage.

**Quantum**

ActiveScale is the most recent acquisition in Quantum's storage portfolio. ActiveScale is a highly scalable scale-out native object storage platform that is sold as an appliance. It comes in two models: the low-capacity P100 that scales to 27PB; and the high-capacity X100 that scales to 74PB. The solution is implemented by deploying individual nodes into a logical solution that supports both a local or geodistributed implementation.

ActiveScale scores well for the quality of management as it has a well-designed dashboard with useful at-a-glance metrics, such as the percentage of capacity used, data durability state and system performance characteristics. ActiveScale also scores well for interoperability and capacity of the solution. ActiveScale scores lower for efficiency, security and performance capabilities as it is missing and still catching up on many features. Some examples are missing features for deduplication and compression, tiering, WORM, QoS, and flash support. ActiveScale is also missing a software-native implementation.

Quantum ActiveScale is consistently positioned across all use cases. ActiveScale's most popular use cases are backup, archiving and private cloud unstructured datasets.

**Red Hat**

Red Hat is the primary developer behind Ceph Storage, an open-source storage solution that delivers block, file and object storage capabilities. However, in this research, it is evaluated only on its object storage capabilities. Ceph's storage software development continuously benefits from the contribution of a vibrant open-source community, with Red Hat as leading contributor. Ceph supports unlimited nodes and capacity, and it automatically rebalances and scales when adding additional servers to the pool. While Ceph’s file system NFS support is available, it is not widely used in production deployments.

Ceph is sold as a software-only storage product and has reference architectures with OEM and ODM server hardware vendors. Ceph receives significant R&D focus as Red Hat’s container storage (OpenShift Container Storage) is now based on Ceph technology. Ceph is a highly tunable product, but performance depends on hardware selection, and thus the product does not provide guaranteed performance. Ceph is the best fit for large organizations where storage administrators are comfortable with scripting and are proficient in Linux skills.
Ceph has fewer ISV qualifications for backup, archiving, surveillance and gateways products compared with its competition. It is also missing some enterprise capabilities, such as deduplication, WORM and tiering. Some Ceph customers still cite the need for better Day 2 operations tools and capabilities, which should improve with the new dashboard release in Red Hat Ceph Storage 4.0. Ceph does not have much public cloud integration as its multicloud gateway capabilities are available only through Red Hat OpenShift Container Storage. Over the last 12 months, Ceph was replatformed to become the primary platform for OpenShift Container Storage. It added a web front end for object storage gateway and object bucket notification, as well as improved back-end storage performance and a simplified installation utility based on Ansible. It now has improved security capabilities and improved QoS features.

Red Hat is not a traditional storage company and often positions its product as part of the Red Hat infrastructure portfolio, which includes OpenShift and OpenStack. The primary use cases for Ceph are cloud storage, backup and analytics.

**Scality**

Scality RING is a highly scalable software-defined distributed storage platform with native integrated support for file protocols and object storage that can be deployed on x86 commodity hardware in a local or geodistributed implementation. Scality goes to market with several hardware partners, including through an OEM arrangement with HPE. Scality is also on the price list for resell with Cisco and Supermicro, and it meets in the channel with Western Digital, Lenovo and Gigabyte.

RING scores well across all capabilities, but it scores especially high on resilience capabilities.

Examples are its 100% guaranteed availability program with compensation, WORM support, and broad data replication and availability capabilities, such as stretch cluster implementations and/or synchronous replication. Customers can also perform their own maintenance and upgrades without vendor support, if they like. RING’s main differentiator remains its capability to provide multiprotocol access to the same objects and its capability to distribute and access data across multiple clouds with its Zenko technology. While Scality already offers comprehensive API coverage, it could still improve its overall compatibility. In addition, Scality could improve its overall ease of deployment and operations.

RING scores consistently across all use cases, but it is often deployed for large datasets in backup and archiving, video and content distribution, hosted email, digital and medical imaging, and HPC archiving solutions.

**Context**

The first generation of object storage, introduced in early 2000, manifested as content-addressed storage (CAS). During the late 2000s, the second phase of object storage shifted the product focus to cloud uses, with a development emphasis on a cost-effective cloud storage infrastructure with erasure codes for storage-efficient protection and better WAN support.
Cloud computing, as a category, was started in 2006 with the introduction of Amazon S3, an object storage service. The current market for object storage is largely being driven by S3’s originator, AWS. There’s a symbiotic relationship between Amazon S3 and other object storage services and products. AWS innovations and use cases on S3 guide the market for vendors offering public-cloud-based and on-premises object storage.

However, there’s a delay between when other object storage vendors implement new functionality and when the use cases materialize. In many instances, the public cloud use cases of object storage never materialize in the enterprise data center because of differences between the surrounding set of solutions and the buyers in each segment. This often results in Mode 1 workloads using on-premises object storage and Mode 2 workloads in the public cloud.

Product/Service Class Definition

Object storage refers to storage hardware and software infrastructure that house data in structures called “objects” and serve hosts via protocols (such as HTTP) and APIs (such as Amazon S3). Conceptually, objects are similar to files in that they are composed of content and metadata. In general, objects support richer metadata than file storage by enabling users or applications to assign attributes to objects that can be used for administrative purposes, data mining and information management.

Critical Capabilities Definition

Object storage products often outscore traditional block and file storage products in capacity scalability, security/multitenancy, total cost of ownership (TCO) and manageability, although they tend to lag in performance, interoperability and efficiency. Given the nascent state of the market, several features that clients expect in a traditional NAS system may be absent or less developed in object storage products due to design considerations or product immaturity. Clients need to understand these trade-offs during the procurement process.

Enterprises should consider the following eight critical capabilities when deploying object storage products. Enterprises can work toward these goals by evaluating object storage products in all capability areas.

Storage Efficiency

A product’s ability to support storage efficiency technologies, such as compression, single-instance storage/deduplication, tiering and configurable erasure coding to reduce TCO.

Security and Multitenancy

The native security features embedded in the platform that provide granular access control, enable enterprises to encrypt information, provide robust multitenancy and QoS, offer data immutability, and ensure compliance with regulatory requirements.

Capacity
The ability of the product to support growth in capacity in a nearly linear manner. It examines object storage capacity scalability limitations in theoretical and real-world configurations, such as maximum theoretical capacity, object size and production deployment.

**Interoperability**

The ability of the product to support third-party ISV applications across horizontal and vertical industry-specific use cases, public cloud APIs, and various deployment models.

**Manageability**

The automation, management, monitoring, and reporting tools and programs supported by the product. In addition, ease of setup and configuration, as well as metadata management capabilities, are considered.

These tools and programs can include single-pane management consoles, monitoring systems and reporting tools. They are designed to help personnel seamlessly manage systems, monitor system usage and efficiencies, and anticipate and correct system alarms and fault conditions before or soon after they occur.

**Performance**

The per-node and aggregated throughput for reads and writes that can be delivered by the cluster in real-world configurations. It also includes support for flash for metadata and data for latency-sensitive applications.

**Resilience**

The platform capabilities for providing high system availability and uptime. Options include high tolerance for simultaneous disk and/or node failures, fault isolation techniques, built-in protection against data corruption, and data protection techniques such as erasure coding and replication.

Features are designed to meet users’ recovery point objectives (RPOs) and recovery time objectives (RTOs). There are several methods for data protection in today’s object storage products. Redundant array of independent disks (RAID) is becoming less popular due to huge capacity overheads and long rebuild times.

The simplest way to protect data is replication, which stores multiple copies of the data locally or in a distributed manner. A more innovative data protection scheme is erasure coding, which breaks up data into “n” fragments and “m” additional fragments across n+m nodes, offering clients configurable choices depending on their cost and data protection requirements. Enterprises often combine erasure coding and replication because the former performs well with large files, whereas the latter works well with large numbers of small files. WAN costs and performance considerations in distributed environments are also factors.

**Value**

This is the price of the product relative to the capabilities an enterprise stands to experience.
Use Cases

Analytics
This applies to storage consumed by big data analytics applications and packaged business intelligence (BI) applications for domain or business problems.

Archiving
This applies to a product’s ability to provide a cost-effective, scalable and long-term data storage repository.

This earliest enterprise use case for object storage products has been used for more than a decade.

Backup
I&O leaders have used object storage products as backup targets for years because they provide added scalability for large backup datasets.

Cloud Storage
This is the most prominent use case for object storage products. Most popular consumer and enterprise public clouds are built on an object storage foundation.

Hybrid Cloud Storage
In this use case, enterprises use object storage solutions to enable a hybrid cloud workflow between on-premises and public cloud IaaS platforms.

Vendors Added and Dropped

Added
- Quantum (ActiveScale product acquired from Western Digital)

Dropped
- Western Digital (ActiveScale product acquired by Quantum)
- SwiftStack (acquired by NVIDIA and no longer sold as stand-alone object storage)
- SUSE (did not meet inclusion criteria)

Inclusion Criteria
To qualify for inclusion, vendors must meet all the following requirements:

- The vendor must have more than $10 million of recognized product revenue over the last four quarters (as of May 2020) for the distributed file systems and/or object storage solutions
between 1 May 2019 and 30 April 2020. Also, it should have at least 75 production customers each consuming more than 500TB raw capacity through either distributed file or object storage protocols only. Vendors must provide reference materials to support this criterion. This may require proof in the form of a confidential list of representative customers outlining deployed capacity per product (75 customers with more than 500TB each). If the vendor cannot share customer names, they can be anonymized as “large manufacturing company” or “small service provider.”

- The product must be installed in at least three major geographies. The vendor will provide evidence of a minimum of 25 production customers brought to revenue in each of the three geographies (North America, EMEA, Asia/Pacific and South America). This may require proof in the form of a confidential list of representative customers from diverse geographies (i.e., 25 customers of at least 500TB each in each of three geographies).

- The product should be deployed across at least three use cases that are outlined in this report. Vendors must provide reference materials to support this criterion.

- The product must be designed for primarily on-premises workloads and not as a pass-through solution where data will be permanently stored elsewhere.

- Products should not be offered exclusively as a service.

- The vendor should own the storage software intellectual property and be a product developer. If a product is built on top of open-source software, the vendor must be one of the top 10 active contributors to the community (in terms of code contribution over the past 12 months).

- The vendor must have a product including features and capabilities generally available before 4 April 2020 that meet the following criteria.

Packaging:

- The product must be sold as either an appliance or software-based storage solution.

- The product must be available for purchase and consumed as a stand-alone file- and/or object-storage-only product and not as part of an integrated, converged or hyperconverged system with a compute and hypervisor bundle.

Product capabilities:

- The product must have file and/or object access to the common namespace/file system.

- The product must have a fully distributed architecture where data and metadata are distributed, replicated or erasure-coded over the network across multiple nodes in the cluster. The product
must have the ability to handle disk, enclosure or node failures in a graceful manner without impacting availability.

- The product must be a single file system capable of expanding beyond 500TB.
- The product must have a global namespace capable of 2PB expansion.
- The cluster must span more than four nodes.
- The product must offer support for horizontal scaling of capacity and throughput in a cluster mode or in independent node additions with a global namespace/file system.

Note: A fully distributed architecture is a distributed computing architecture in which each node is independent and self-sufficient, and there is no single point of contention across the system. More specifically, none of the nodes share memory or disk storage. People typically contrast distributed design systems with systems that keep a large amount of centrally stored state information, whether in a database, an application or metadata server, or any other similar single point of contention.

<table>
<thead>
<tr>
<th>Critical Capabilities</th>
<th>Analytics</th>
<th>Archiving</th>
<th>Backup</th>
<th>Cloud Storage</th>
<th>Hybrid Cloud Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage Efficiency</td>
<td>12%</td>
<td>13%</td>
<td>14%</td>
<td>12%</td>
<td>10%</td>
</tr>
<tr>
<td>Security and Multitenancy</td>
<td>12%</td>
<td>14%</td>
<td>10%</td>
<td>12%</td>
<td>10%</td>
</tr>
<tr>
<td>Capacity</td>
<td>13%</td>
<td>13%</td>
<td>14%</td>
<td>12%</td>
<td>11%</td>
</tr>
<tr>
<td>Interoperability</td>
<td>14%</td>
<td>13%</td>
<td>11%</td>
<td>13%</td>
<td>20%</td>
</tr>
<tr>
<td>Manageability</td>
<td>13%</td>
<td>11%</td>
<td>11%</td>
<td>13%</td>
<td>12%</td>
</tr>
<tr>
<td>Performance</td>
<td>13%</td>
<td>8%</td>
<td>11%</td>
<td>12%</td>
<td>13%</td>
</tr>
<tr>
<td>Resilience</td>
<td>10%</td>
<td>13%</td>
<td>14%</td>
<td>13%</td>
<td>10%</td>
</tr>
<tr>
<td>Value</td>
<td>13%</td>
<td>15%</td>
<td>15%</td>
<td>13%</td>
<td>14%</td>
</tr>
</tbody>
</table>
This methodology requires analysts to identify the critical capabilities for a class of products/services. Each capability is then weighed in terms of its relative importance for specific product/service use cases.

**Critical Capabilities Rating**

Each of the products/services that meet our inclusion criteria has been evaluated on the critical capabilities on a scale from 1.0 to 5.0.

<table>
<thead>
<tr>
<th>Critical Capabilities</th>
<th>Caringo</th>
<th>Cloudian</th>
<th>DDN</th>
<th>Dell Technologies</th>
<th>HVI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage Efficiency</td>
<td>3</td>
<td>3.5</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Security and Multitenancy</td>
<td>3</td>
<td>4.5</td>
<td>2</td>
<td>3.8</td>
<td>4.5</td>
</tr>
<tr>
<td>Capacity</td>
<td>3.8</td>
<td>4.3</td>
<td>4.1</td>
<td>4.1</td>
<td>4.3</td>
</tr>
<tr>
<td>Interoperability</td>
<td>3</td>
<td>4.4</td>
<td>3</td>
<td>4.3</td>
<td>4.4</td>
</tr>
<tr>
<td>Manageability</td>
<td>3.5</td>
<td>4.2</td>
<td>3</td>
<td>4.3</td>
<td>3.5</td>
</tr>
<tr>
<td>Performance</td>
<td>3</td>
<td>4.5</td>
<td>3</td>
<td>2.8</td>
<td>3</td>
</tr>
<tr>
<td>Resilience</td>
<td>3</td>
<td>4.3</td>
<td>3</td>
<td>3.7</td>
<td>3.7</td>
</tr>
<tr>
<td>Value</td>
<td>3</td>
<td>4.1</td>
<td>3</td>
<td>3.9</td>
<td>4.1</td>
</tr>
</tbody>
</table>

As of 20 August 2020
Table 3 shows the product/service scores for each use case. The scores, which are generated by multiplying the use-case weightings by the product/service ratings, summarize how well the critical capabilities are met for each use case.

### Table 3: Product Score in Use Cases

<table>
<thead>
<tr>
<th>Use Cases</th>
<th>Caringo</th>
<th>Cloudian</th>
<th>DDN</th>
<th>Dell Technologies</th>
<th>Hitachi Vantara</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytics</td>
<td>3.17</td>
<td>4.23</td>
<td>3.02</td>
<td>3.75</td>
<td>3.94</td>
</tr>
<tr>
<td>Archiving</td>
<td>3.16</td>
<td>4.21</td>
<td>3</td>
<td>3.78</td>
<td>4</td>
</tr>
<tr>
<td>Backup</td>
<td>3.17</td>
<td>4.2</td>
<td>3.05</td>
<td>3.73</td>
<td>3.94</td>
</tr>
<tr>
<td>Cloud Storage</td>
<td>3.16</td>
<td>4.23</td>
<td>3.01</td>
<td>3.75</td>
<td>3.94</td>
</tr>
<tr>
<td>Hybrid Cloud Storage</td>
<td>3.15</td>
<td>4.25</td>
<td>3.02</td>
<td>3.79</td>
<td>3.96</td>
</tr>
</tbody>
</table>

As of 20 August 2020

Source: Gartner (October 2020)

To determine an overall score for each product/service in the use cases, multiply the ratings in Table 2 by the weightings shown in Table 1.

### Critical Capabilities Methodology

This methodology requires analysts to identify the critical capabilities for a class of products or services. Each capability is then weighted in terms of its relative importance for specific product or service use cases. Next, products/services are rated in terms of how well they achieve each of the critical capabilities. A score that summarizes how well they meet the critical capabilities for each use case is then calculated for each product/service.

"Critical capabilities" are attributes that differentiate products/services in a class in terms of their quality and performance. Gartner recommends that users consider the set of critical capabilities as some of the most important criteria for acquisition decisions.
In defining the product/service category for evaluation, the analyst first identifies the leading uses for the products/services in this market. What needs are end-users looking to fulfill, when considering products/services in this market? Use cases should match common client deployment scenarios. These distinct client scenarios define the Use Cases.

The analyst then identifies the critical capabilities. These capabilities are generalized groups of features commonly required by this class of products/services. Each capability is assigned a level of importance in fulfilling that particular need; some sets of features are more important than others, depending on the use case being evaluated.

Each vendor’s product or service is evaluated in terms of how well it delivers each capability, on a five-point scale. These ratings are displayed side-by-side for all vendors, allowing easy comparisons between the different sets of features.

Ratings and summary scores range from 1.0 to 5.0:

1 = Poor or Absent: most or all defined requirements for a capability are not achieved

2 = Fair: some requirements are not achieved

3 = Good: meets requirements

4 = Excellent: meets or exceeds some requirements

5 = Outstanding: significantly exceeds requirements

To determine an overall score for each product in the use cases, the product ratings are multiplied by the weightings to come up with the product score in use cases.

The critical capabilities Gartner has selected do not represent all capabilities for any product; therefore, may not represent those most important for a specific use situation or business objective. Clients should use a critical capabilities analysis as one of several sources of input about a product before making a product/service decision.

**Document Revision History**

Critical Capabilities for Object Storage - 25 November 2019

Critical Capabilities for Object Storage - 30 January 2019

Critical Capabilities for Object Storage - 25 January 2018

Critical Capabilities for Object Storage - 31 March 2016

Critical Capabilities for Object Storage - 11 February 2014

**Recommended by the Authors**
Magic Quadrant for Distributed File Systems and Object Storage
Critical Capabilities for Distributed File Systems
2020 Strategic Roadmap for Storage
Competitive Landscape: Infrastructure Software-Defined Storage
How Products and Services Are Evaluated in Gartner Critical Capabilities

Recommended For You
Gartner Peer Insights 'Voice of the Customer': Primary Storage
2020 Strategic Roadmap for Storage
Magic Quadrant for Primary Storage
Gartner Peer Insights 'Voice of the Customer': Distributed File Systems and Object Storage
Market Guide for Servers

Supporting Initiatives

© 2020 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates. This publication may not be reproduced or distributed in any form without Gartner's prior written permission. It consists of the opinions of Gartner's research organization, which should not be construed as statements of fact. While the information contained in this publication has been obtained from sources believed to be reliable, Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although Gartner research may address legal and financial issues, Gartner does not provide legal or investment advice and its research should not be construed or used as such. Your access and use of this publication are governed by Gartner's Usage Policy. Gartner prides itself on its reputation for independence and objectivity. Its research is produced independently by its research organization without input or influence from any third party. For further information, see "Guiding Principles on Independence and Objectivity."