



Cloud object storage

A comparison of Dell EMC Elastic Cloud Storage and public cloud object storage solutions

As the need to ingest, store, and process data means more to businesses than ever before, choosing a cloud object storage solution that meets these needs becomes ever more critical. Cloud-based object storage allows applications, developers, and administrators to access and control data via APIs that rely on HTTP or HTTPS, ensuring integration with current and future applications. Cloud object storage can also extend beyond a single datacenter or a single cloud object storage provider, allowing businesses to scale more effectively.

Every business has different requirements for a cloud object storage solution. For instance, industry regulations and data sovereignty laws may influence some deployments, while the cost associated with deploying a solution is a ubiquitous concern. Certain applications require storage to have low latency and high performance, all while administrators prevent security breaches. In this paper, we take a look at many of the factors and features that businesses might consider when deploying cloud object storage.

We reviewed cloud-based object storage solutions from four competitors: Amazon Simple Storage Service™ (S3), Dell EMC™ Elastic Cloud Storage™ (ECS), Google Cloud Storage™, and Microsoft® Azure® Blob storage. This paper draws on publicly available information such as marketing material, user guides, manuals, and other documentation.

Executive summary

The four vendors we investigated—Amazon, Dell EMC, Google, and Microsoft—all provide usable cloud object storage. However, many of the features and implementations differ on each platform.

The public cloud competitors we analyzed all charge their customers on a monthly basis.^{1,2,3} This may include storage costs, bandwidth fees, costs associated with each API call, and more expensive redundancy settings, as well as costs associated with higher-performance storage. This means as storage needs grow or API requests increase in frequency, the cost per month increases as well. The three public cloud vendors in this paper offer free data ingress to their platform—but charge when it's time to move data out.^{1,2,3} When choosing an on-site solution like Dell EMC ECS, businesses can finance or purchase outright the capacity they need, free of demand-based costs. This can make the cost of object storage easier to manage and budget.

The public cloud providers in this paper have a wide array of products available for use under their brand. For instance, each of the three public cloud providers provide infrastructure-as-a-service and database-as-a-service technologies.^{4,5,6} Amazon currently has the largest market share in the public cloud space by a considerable margin, but Microsoft and Google are growing their cloud businesses at a rapid pace.⁷

When it comes to performance, accessing latency-critical data might favor an onsite solution. Because the solution and associated hardware are connected to the company network, data doesn't need to travel across less-than-consistent Internet connections. With Dell EMC ECS, administrators also gain visibility into many performance metrics such as individual disk performance and utilization, whereas most public cloud providers don't provide visibility.

Legacy applications and big data are another area where flexible storage is important. While many object storage solutions support a folder-like structure in their buckets using character delimiters in object keys, Dell EMC ECS provides additional protocols on a single appliance—such as HDFS and NFS. The multiple protocols available through ECS also provide interoperability, allowing data ingested with one protocol to be accessed with a different protocol. Amazon and Microsoft include other types of storage beyond object storage in their portfolios, but none provide interoperability or include the wide array of APIs that ECS is capable of supporting.

For developers, the ability to use a wide range of popular APIs means more flexibility when developing or implementing applications. With the large AWS market share lead, the Amazon S3 API might be one of the more popular object storage APIs—meaning many applications are built to use this API.⁷ OpenStack Swift is also growing in popularity, with many developers currently advancing the platform.⁸ By supporting multiple popular APIs, Dell EMC ECS can allow developers and businesses the ability to leverage greater community support and compatible applications that come with these APIs. The ability to search objects based on metadata can also aid developers. Only Dell EMC and Microsoft provide such an ability, but the Microsoft Azure indexing feature is in preview and is not recommended for production environments.

1 <https://cloud.google.com/storage/pricing>

2 <https://aws.amazon.com/s3/pricing/>

3 <https://azure.microsoft.com/en-us/pricing/details/storage/>

4 <https://aws.amazon.com/>

5 <https://azure.microsoft.com/en-us/>

6 <https://cloud.google.com/products/>

7 <https://www.srgresearch.com/articles/aws-remains-dominant-despite-microsoft-and-google-growth-surges>

8 <http://stackalytics.com/?release=all&module=swift-group&metric=commits>

Providing cloud object storage

With the rise of emerging technologies such as the Internet of Things, mobile devices, and streaming media, cloud storage providers have had to keep up with increased demand. Over time, object storage solutions have matured, gaining more customers, better direction, and more features. Many of these features are designed to help businesses deal with common issues like scalability, cost, and ease of use.

Since object storage serves multiple purposes, some public cloud providers have created various service level tiers. These tiers differ in terms of pricing, performance, and recommended usage. For instance, backups may be deployed on cheaper, slower storage, while streaming media uses higher-performance storage.

Public cloud providers charge for their service on a per-month rolling basis, shifting costs to an operational expenditure. Many public cloud vendors provide free data ingress to their platform, but charge based on the number of API calls, the amount of data stored on the platform, and for bandwidth.⁹ On-site object storage deployments aren't subject to the same variable monthly fees as public cloud providers.

Amazon

Amazon Web Services (AWS) provides public cloud solutions including storage, networking, and more.¹⁰ Amazon launched AWS in 2006, introducing services such as Amazon Simple Storage Service and Elastic Compute Cloud. Amazon offers free tiers for many of these services, allowing users to try them on a limited scale before they decide to purchase.¹¹

While AWS doesn't fully disclose all hardware details, Intel processors provide the compute.⁹ Amazon doesn't offer on-premise hardware storage solutions, but does provide software appliances like the AWS Storage Gateway, which connects on-site storage infrastructure with the AWS cloud.¹²

Amazon Web Services markets their object storage solution under the Amazon Simple Storage Service (S3) brand.¹³ The service has three tiers:¹⁴

- Standard
- Standard—Infrequent Access (IA)
- Amazon Glacier

Amazon describes Glacier as a backup solution for extremely infrequently accessed data—the retrieval latency is measured in hours instead of milliseconds. Guaranteed availability differs for each tier. For instance, Amazon Glacier has no guaranteed percentage of availability.

Amazon S3 charges customers based on the amount of data they store on the platform per month, as well as per HTTP request administrators make on the objects. Pricing varies by region, storage tier, and other settings. Ingest of data into Amazon S3 is free. However, data transfer to other AWS Regions and the Internet is charged on a GB-per-month basis.² With Amazon S3, as the amount of data stored increases, API calls increase in frequency, or data is moved off the platform—the cost per month increases.

⁹ <https://aws.amazon.com/ec2/instance-types/>

¹⁰ <https://aws.amazon.com/>

¹¹ <https://aws.amazon.com/free/>

¹² <https://aws.amazon.com/storagegateway/>

¹³ <https://aws.amazon.com/s3/>

¹⁴ <https://aws.amazon.com/s3/storage-classes/>

Amazon S3 is currently available in 13 regions worldwide. There are four regions in the United States, each with two to five availability zones, which are discrete AWS facilities.¹⁵ One of the regions in the US is reserved for government use. There are a total of 11 availability zones for non-government use in the US.

Dell EMC

Dell EMC offers the Elastic Cloud Storage (ECS) appliance for cloud object storage.¹⁶ This is the third generation of Dell EMC object storage appliances, launched in 2014.¹⁷ The appliance is available as either a turnkey, fully integrated hardware appliance, or as a software appliance that can run on commodity or third-party hardware.¹⁸ Administrators can customize and configure the hardware appliance depending on business needs and expected growth.¹⁹ The software appliance is available free of charge for non-production environments.²⁰

To access object storage on the ECS system, Dell EMC has built in support for various APIs, such as:

- Amazon S3
- OpenStack® Swift
- Dell EMC Atmos®
- Dell EMC Centera® CAS

Along with the array of supported APIs used to interact with object storage, ECS also provides support for various storage protocols such as HDFS and NFS.²¹ Providing multiple protocols can help consolidate hardware resources while keeping support for legacy applications that might need NFS.

As Dell EMC ECS is available in multiple configurations depending on the business need and necessary capacity, the cost can also be structured based on the available budget. Dell EMC Financial Services can provide flexible options for the procurement of Dell EMC hardware, including financing.²² Unlike many public cloud solutions, Dell EMC ECS doesn't charge monthly per GB, HTTP GET request, or data egress, helping to limit 'hidden' or unexpected costs.²¹

Administrators can deploy the ECS solution anywhere in the world, as the only requirement is space in a datacenter. Businesses can deploy the ECS appliance directly or to service providers to provide object storage to clients. When businesses deploy ECS on their premise, they have absolute control over where data resides.

15 <https://aws.amazon.com/about-aws/global-infrastructure/>

16 <http://www.emc.com/storage/ecs/index.htm>

17 <http://www.emc.com/about/news/press/2014/20140505-01.htm>

18 <https://www.emc.com/collateral/white-papers/h14071-ecs-architectural-guide-wp.pdf>

19 <https://www.emc.com/collateral/specification-sheet/h13117-emc-ecs-appliance-ss.pdf>

20 <http://www.emc.com/products-solutions/trial-software-download/ecs.htm>

21 <http://www.emc.com/infographics/ecs-infographic.htm>

22 <https://store.emc.com/us/Product-Family/EMC-Elastic-Cloud-Storage-Products/EMC-Elastic-Cloud-Storage/p/EMC-Elastic-Cloud-Storage-ECS-Appliance#Finance>

Google

Google Cloud Platform and Google Cloud Storage both launched in 2010.²³ Google Cloud Platform includes services such as hosted virtual machines, containers, networking, and storage options.²⁴ Google Cloud Platform focuses on public cloud offerings and doesn't offer hardware or software designed to run in an on-premise datacenter. Google does not disclose the hardware on which they provide services beyond the fact that they use Intel processors.²⁵

To help customers more easily integrate the Google public cloud with existing infrastructure, Google provides connectors and plugins, such as Cloud Interconnect and the Google Cloud Storage Connector for Hadoop.^{26,27}

Object storage is the only type of storage offered under the Google Cloud Storage brand.²⁸ There are three storage classes:

- Standard Storage
- Durable Reduced Availability (DRA)
- Cloud Storage Nearline

Both DRA and Standard Storage have access latencies measured in milliseconds, while Cloud Storage Nearline is aimed at less frequently accessed data—such as backups and records. Each tier has different guarantees of availability and cost.

The three tiers of Google Cloud Storage all have different prices per GB per month and the cost remains constant throughout the different available regions in the United States. Data ingress to the Google Cloud Storage platform is free, but data transfer between regions incurs a cost per GB. Operations performed using the Google Storage APIs incur a cost per action, and the cost of data egress depends on the number of GB transferred per month.¹ Using the Google Cloud Storage platform means cost increases as API calls increase in frequency, as storage needs grow, or as data is transferred off the platform or to a different region.

At the time of this writing, there are three regions in the US—US Central, US East, and US West.²⁹ Storage can also be deployed to a “multi-region location,” which means data is replicated to multiple regions. There is one multi-region location in the US.

23 <https://techcrunch.com/2010/05/18/google-to-launch-amazon-s3-competitor-google-storage-at-io/>

24 <https://cloud.google.com/products/>

25 <https://cloud.google.com/compute/docs/machine-types>

26 <https://cloud.google.com/interconnect/>

27 <https://cloud.google.com/storage/docs/google-integration>

28 <https://cloud.google.com/storage/>

29 <https://cloud.google.com/about/locations/>

Microsoft

Microsoft launched Windows® Azure in 2010. In 2014, Microsoft rebranded the service as Microsoft Azure and continued to expand their offerings. The Azure brand now encompasses a variety of services spanning compute power, various storage options, and more. Microsoft also partners with hardware vendors to provide on-premise private cloud solutions that are Azure-consistent.³⁰ The Microsoft Azure storage umbrella encompasses options such as Blob storage, file storage, premium storage, and more.³¹

Azure Blob storage is the Microsoft object storage offering and offers two tiers:³²

- Hot storage tier
- Cool storage tier

The tiers have similar performance but different cost structures and guarantees of availability.³³ Azure Blob storage pricing is metered in GB and is charged monthly. Cost differs based on redundancy level, region, storage tier, and type of blob. Operations, data ingress, and data egress all incur costs on a per-operation and per GB-per-month basis, respectively. Only certain types, tiers, and redundancy levels have free data ingress.³ Azure Blob storage increases in cost with an increase in API calls, data redundancy levels, amount of storage, and data egress.

Azure Blob storage also offers three types of blobs: block blobs, append blobs, and page blobs.³⁴ Block blobs are used for streaming operations, media files, documents, etc. Append blobs allow append operations for things such as log files, where data is written to the end of a blob. Page blobs support random read and write operations.

Azure Blob storage is currently available in six regions in the US and sixteen regions worldwide.³⁵ The four regions in the US include Central US, East US, East US 2, West US 2, West Central US, and South Central US.

Competitive cloud object storage features

When it comes to cloud object storage, many features can make or break a deployment. In the following sections, we describe and analyze a subset of the features available with each cloud object storage service. Note that each of the four competitors we discuss markets and implements cloud object storage differently.

We chose features that matter to businesses: Having robust access control can help increase security, which is crucial to business operations as security threats and data breaches become more common. Robust protocol options can aid developers and allow compatibility with existing applications, while deep visibility into storage operations can help administrators troubleshoot bottlenecks and optimize performance.

Access control and access logging

Security on an access level is paramount. While some data is public, other content must remain secure. Providers have different methods to control who has access to what data. To track who has accessed what data, some providers track and log all data operations. Some cloud deployments may require multi-tenancy, sub-tenants, or single tenant options. Providing administrators these options is important for many cloud use cases.

30 <https://www.microsoft.com/en-us/cloud-platform/cloud-platform-system>

31 <https://azure.microsoft.com/en-us/services/storage/>

32 <https://azure.microsoft.com/en-us/documentation/articles/storage-blob-storage-tiers/#comparison-between-the-storage-tiers>

33 <https://azure.microsoft.com/en-us/documentation/articles/storage-blob-storage-tiers/>

34 <https://azure.microsoft.com/en-us/documentation/articles/storage-introduction/#blob-storage>

35 <https://azure.microsoft.com/en-us/regions/#services>

Amazon

Amazon S3 provides ACLs, as well as per-bucket policies. The AWS Identity and Access Management (IAM) tool integrates with Amazon S3 to control and manage users, groups, and permissions.^{36,37} By using ACLs, administrators can grant read/write permissions to other AWS accounts.³⁸ ACLs can be managed through the SDKs and APIs that Amazon publishes, as well as the traditional AWS Management Console. Amazon S3 also supports time-limited access to certain objects.³⁹

When integrated with other AWS products, S3 can also support Microsoft Active Directory®, allowing administrators to manage access using existing infrastructure. Administrators can delegate access to many AWS products using this service.⁴⁰

Access logging is disabled by default, but Amazon S3 provides options for log delivery.⁴¹ Administrators can deliver logs to a target bucket, allowing a single bucket to hold separate logs for multiple source buckets, which are the buckets under logged access. S3 provides log files in space-delimited fields.⁴²

Dell EMC

Dell EMC provides integration with Active Directory and Lightweight Directory Access Protocol (LDAP), along with OpenStack Keystone to ensure support for many existing types of infrastructure. Active Directory groups can be leveraged to manage access, and ECS also provides support for multi-domain scenarios. The Dell EMC implementation of NFS can be configured to use access control lists (ACLs), and Dell EMC ECS also extends the Amazon S3 API to include support for rich Dell EMC ACLs. With Keystone, administrators can use the available OpenStack APIs and client libraries to manage access.

Dell EMC ECS also provides robust logging—administrators can view and download logs for the many protocols that ECS supports, as well as various authentication methods available.⁴³ The ECS Portal Monitoring page also provides event logging for the ECS API and much more.⁴⁴

ECS provides support for multiple authentication and access methods, allowing it to integrate with various types of infrastructure. ECS includes support for enterprise single-tenant, enterprise multi-tenant, cloud service provider single-tenant, and cloud service provider multi-tenant scenarios. Sub-tenants can be set up by using permissions available on buckets, and buckets and objects are configured with namespaces that allow access by certain tenants.⁴⁵

36 <https://aws.amazon.com/iam/>

37 <http://docs.aws.amazon.com/AmazonS3/latest/dev/s3-access-control.html>

38 http://docs.aws.amazon.com/AmazonS3/latest/dev/S3_ACLs_UsingACLs.html

39 <http://docs.aws.amazon.com/AmazonS3/latest/dev/ShareObjectPreSignedURL.html>

40 <https://aws.amazon.com/directoryservice/>

41 <http://docs.aws.amazon.com/AmazonS3/latest/dev/ServerLogs.html>

42 <http://docs.aws.amazon.com/AmazonS3/latest/dev/LogFormat.html>

43 http://www.emc.com/techpubs/ecs/ecs_logs-1.htm

44 https://www.emc.com/techpubs/ecs/ecs_monitor_events-1.htm

45 <http://www.emc.com/collateral/TechnicalDocument/docu59635.pdf>

Google

Google provides a few options for access control, including Google Cloud Identity and Access Management (IAM) permissions, ACLs, signed URLs, and signed policy documents.⁴⁶ Access control lists can apply to buckets or individual objects and are managed through the Google Cloud Console, gsutil, and both JSON and XML APIs.⁴⁷ Google IAM provides granular permissions that govern the roles and permissions of buckets and objects.⁴⁸ Google also provides time-limited access to resources.⁴⁹

Access logs are available in CSV format and are delivered to a specified bucket.⁵⁰ Administrators can configure and manage access logging through gsutil and the JSON and XML APIs. Logs are generated once daily and contain the previous day's usage.

Microsoft

Azure Active Directory role-based access control (RBAC) can govern access to containers and data.⁵¹ With Active Directory, administrators can integrate existing on-premise Active Directory servers, allowing for a single method of access control with on-premise and public cloud. Microsoft Azure Blob storage provides for RBAC and includes various roles that each have specific Actions and Not Actions they can or cannot perform.⁵² Through shared access signatures (SAS), administrators can provide time-limited, delegated access to resources without sharing an account key or credentials.⁵³

Access logs are saved in a dedicated container named \$logs and provide information on failed accesses as well as successful ones.⁵⁴ Logs are provided in a specific format outlined by Microsoft.⁵⁵

Change notifications

When data is changed or updated, some applications may use information on the changes to perform useful tasks. For instance, administrators can use change notification to move objects to faster storage if they've been recently changed. There are many ways to provide change notification to outside applications, and each provider provides the notifications slightly differently.

Amazon

Amazon S3 relies on AWS tools such as Amazon Simple Notification Service (SNS), Amazon Simple Queue Service (SQS), and AWS Lambda to provide event notifications.⁵⁶ The S3 Event Notifications provide monitoring of deletions, creations, and the loss of an object of the RRS storage class.⁵⁷ Amazon SNS can push messages to mobile devices and distributed services. Amazon SQS can transmit any amount of data and allows the addition of event notifications to the sending queue. AWS Lambda can run custom code in response to events and extend automation into other Amazon cloud solutions and tools.

46 <https://cloud.google.com/storage/docs/access-control/>

47 <https://cloud.google.com/storage/docs/access-control/create-manage-lists>

48 <https://cloud.google.com/storage/docs/access-control/iam>

49 <https://cloud.google.com/storage/docs/access-control/signed-urls>

50 <https://cloud.google.com/storage/docs/access-logs>

51 <https://azure.microsoft.com/en-us/documentation/articles/role-based-access-control-configure/>

52 <https://azure.microsoft.com/en-us/documentation/articles/storage-security-guide/>

53 <https://azure.microsoft.com/en-us/documentation/articles/storage-dotnet-shared-access-signature-part-1/>

54 <https://msdn.microsoft.com/en-us/library/azure/dn782840.aspx>

55 <https://msdn.microsoft.com/library/hh343259.aspx>

56 <http://docs.aws.amazon.com/AmazonS3/latest/dev/NotificationHowTo.html>

57 <http://docs.aws.amazon.com/AmazonS3/latest/dev/NotificationHowTo.html#supported-notification-event-types>

Dell EMC

Dell EMC ECS currently doesn't support change notifications.⁵⁸

Google

Google allows client applications or services to subscribe to a notification channel by first sending a watch request on a bucket.⁵⁹ However, only the Google JSON API supports notifications. Any application or server that can process an HTTPS POST request can receive these notifications. Google provides notifications when an object is created, changed, or removed from the bucket.

Microsoft

While native change notification isn't baked into Azure Blob storage, Microsoft has developed guides to help create services that can monitor containers for changes. By using the Azure WebJobs SDK, administrators can create jobs to perform change notification.⁶⁰ However, Microsoft reports these can have delays that may be longer than several minutes.

Multi-protocol storage access

Organizations can use many types of storage, even legacy solutions such as CIFS or NFS. It may be cost-prohibitive to convert these applications to use object storage, and even then many can't support the necessary APIs. Being able to use the same solution to provide different types of storage can streamline the administrator and end-user experience, as well as build in more features for the same cost.

Amazon

Amazon S3 supports only object storage, but AWS includes other types of storage in their portfolio. For example, Amazon Elastic Block Store (EBS) provides block storage to use with Amazon EC2 virtual machines, as well as Amazon Elastic File System (EFS) to provide file storage for use with EC2 virtual machines.^{61,62} Objects stored in Amazon S3 can't be accessed through other Amazon products such as EBS or EFS.⁶³

Dell EMC

Dell EMC ECS provides multi-protocol support on the same appliance. For instance, ECS supports object storage, NFS, and HDFS.⁶⁴ This allows ECS to support multiple types of applications that include legacy systems and big data applications. In addition to multi-protocol support, Dell EMC ECS also provides multi-protocol access, meaning data ingested through one protocol can be accessed using a different protocol. For instance, ECS provides interoperability between HDFS, S3, and NFS. Dell EMC also provides Dell EMC-CIFS, a tool allowing users to view storage as a disk on a Microsoft Windows-based system.⁶⁵ It uses the Amazon S3 API.

58 http://www.emc.com/techpubs/ecs/ecs_s3_supported_features-1.htm#SECTION_34FEDE20BDD1400495B-904F0C4E48209

59 <https://cloud.google.com/storage/docs/object-change-notification>

60 <https://azure.microsoft.com/en-us/documentation/articles/websites-dotnet-webjobs-sdk-storage-blobs-how-to/>

61 <https://aws.amazon.com/ebs/>

62 <https://aws.amazon.com/efs/>

63 <https://aws.amazon.com/>

64 <https://www.emc.com/collateral/white-papers/h14071-ecs-architectural-guide-wp.pdf>

65 <https://www.emc.com/collateral/white-papers/h15277-emc-cifs-ecs-architecture-overview.pdf>

Google

Google Cloud Storage only supports object storage, but there are other options under the Google Cloud Platform umbrella—such as Persistent Disk, which provides network-attached block storage.⁶⁶

Microsoft

Microsoft Azure Blob storage only supports object storage, but the Azure storage brand encompasses other solutions such as file storage, which provides file shares that use SMB 3.0, as well as premium storage, which provides high-performance solid-state drive-based storage for Azure virtual machines.⁶⁷ Data stored in Azure Blob storage isn't accessible using SMB 3.0 or any method other than the Azure Blob storage APIs.⁶⁸ Users may also access Azure Storage using the Azure Storage Explorer, which has apps for Windows, macOS, and Linux.⁶⁹

Available APIs and client libraries/SDKs

It's important to have Application Program Interfaces (APIs) that are well documented and easy to use. Since most if not all interactions with the object storage solution use these APIs, it's also important that current and future applications can or already use the available APIs. Some APIs have achieved high market share—meaning many developers use that platform and have developed applications that depend on it.

Amazon

Amazon S3 provides access to the S3 REST API. S3 still maintains support for the older SOAP API, but it is now available only over HTTPS.⁷⁰ Amazon provides client libraries to ease development in Java, .NET, PHP, Ruby, and Python.⁷¹ Amazon S3 also provides many code examples for developers to use.⁷²

Dell EMC

Dell EMC ECS provides support for multiple APIs that provide compatibility with those from third-party vendors such as Amazon S3 and OpenStack Swift.⁷³ In addition, Dell EMC has developed other APIs that the ECS appliance supports, such as Centera Content Addressable Storage (CAS) and the Dell EMC Atmos API. Dell EMC provides resources for developers to develop with Atmos or CAS.⁷⁴ With four ways to access object storage, Dell EMC ensures developers have robust capabilities to integrate object storage into applications. Because Dell EMC ECS supports the Amazon S3 API, many of the code examples and other resources available online for S3 can be of use. Applications that depend on the S3 API can also use Dell EMC ECS.

66 <https://cloud.google.com/docs/storing-your-data>

67 <https://azure.microsoft.com/en-us/services/storage/>

68 <https://azure.microsoft.com/en-us/services/storage/blobs/> documentation

69 <http://storageexplorer.com/>

70 <http://docs.aws.amazon.com/AmazonS3/latest/API/Welcome.html>

71 <https://aws.amazon.com/code>

72 <https://aws.amazon.com/code/Amazon-S3?browse=1>

73 <https://www.emc.com/collateral/white-papers/h14071-ecs-architectural-guide-wp.pdf>

74 <http://www.emc.com/collateral/TechnicalDocument/docu59165.pdf>

Google

Google Cloud Storage provides access to two APIs, including the Google Cloud Storage JSON API and the Google Cloud Storage XML API.⁷⁵ Both APIs are RESTful, but some features are available only with the JSON API or the XML API. Google also provides many examples and client libraries for languages such as Python, Objective C, .NET, and Java.⁷⁶ To transfer data to Google Cloud Storage, the Google Cloud Storage Transfer API helps to get external storage into a Google Cloud Storage bucket.⁷⁷

Microsoft

Administrators can access Microsoft Azure Blob storage via the Azure Blob Service REST API.⁷⁸ Microsoft also provides examples and libraries in languages such as .NET, Node.js, Java, C++, PHP, Ruby, and Python.⁷⁹ Microsoft also provides an emulator where developers can easily test and develop code without touching production environments.⁸⁰

Visibility, management, and control

To manage a storage system, it's important to have an easy-to-use portal or other interfaces to make changes as needed. Insights into the performance of a system can be valuable for administrators, helping them eliminate bottlenecks and inefficiencies.

Amazon

Administrators can manage Amazon S3 through the AWS Management Console or the AWS CLI. They can also control S3 via the AWS SDKs and APIs. The AWS Management Console is also available as a mobile app.⁸¹ All storage functions can be controlled through the AWS Management Console, a web portal. The AWS Management Portal for S3 provides the ability to upload, download, and view objects, as well as access control and other functions.⁸²

Dell EMC

The Dell EMC ECS Management API is a RESTful API that provides all management functions of the ECS appliance.⁸³ With the API, administrators and businesses can develop applications and custom portals for managing their deployments of single and multiple ECS appliances.⁸⁴

The ECS Management portal is built using the ECS Management API and provides a GUI for the management of the entire ECS stack, including the hardware. From viewing hardware and disk health to metering object storage buckets, the ECS portal provides a single pane to manage it.⁸⁵

Dell EMC ECS also provides a CLI that can be used for management, and it is also built using the ECS Management API.⁸⁶ It supports installation on systems running Red Hat® Enterprise Linux® (RHEL), SUSE Linux,

75 <https://cloud.google.com/storage/docs/apis>

76 https://cloud.google.com/storage/docs/json_api/v1/libraries

77 <https://cloud.google.com/storage/transfer/reference/rest/>

78 <https://msdn.microsoft.com/en-us/library/azure/dd135733.aspx>

79 <https://azure.microsoft.com/en-us/documentation/articles/storage-nodejs-how-to-use-blob-storage/>

80 <https://azure.microsoft.com/en-us/documentation/articles/storage-use-emulator/>

81 <https://aws.amazon.com/console/mobile/>

82 <https://aws.amazon.com/s3/details/>

83 http://www.emc.com/techpubs/ecs/ecs_api_object_control_service-1.htm

84 <https://www.emc.com/collateral/white-papers/h14071-ecs-architectural-guide-wp.pdf>

85 http://www.emc.com/techpubs/ecs/ecs_portal_use-1.htm

86 http://www.emc.com/techpubs/ecs/ecs_cli_install-1.htm

and Microsoft Windows.

Since ECS is compatible with the Amazon S3 API, OpenStack Swift API, and other APIs, some management functions are also available through these APIs. For instance, key management is available through the Amazon S3 API.

Google

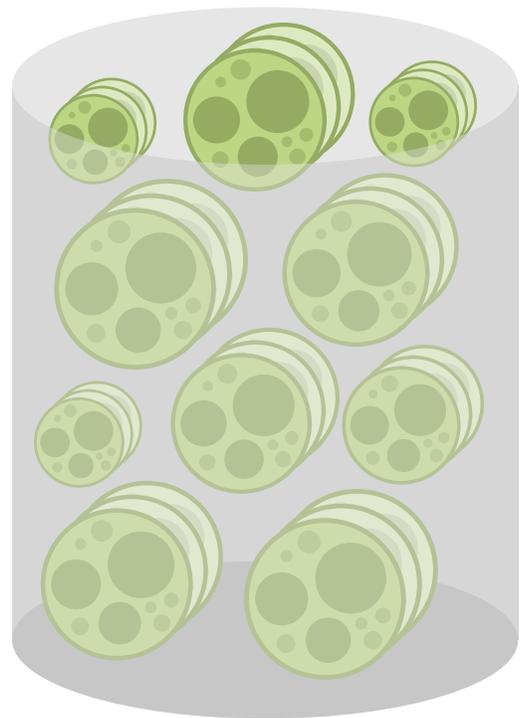
Google Cloud Storage provides the Google Cloud Platform Console, a web-based portal for management.⁸⁷ The Google Cloud Platform Console provides full access and management to storage functions such as uploading, downloading, previewing, and management tasks. The gsutil Tool is a command line interface developed in Python that provides access to many management functions.⁸⁸ Google also allows management through various client libraries and SDKs.⁸⁹

Microsoft

Microsoft Azure Blob storage provides the Azure Portal, from which storage administrators can perform many tasks relating to blob storage, such as uploading, downloading, and creating storage accounts.^{90,91} The Azure PowerShell module allows administrators to use PowerShell to automate tasks, storage accounts, containers and blobs.⁹² The Azure CLI provides a cross-platform method of interacting with Azure and Azure Blob storage, as the CLI can run on OS X, Windows, and Linux.⁹³ Through the available storage client libraries like the one available for .NET, administrators can also manage storage functions.⁹⁴

Versioning and snapshots

Object versioning allows users to undelete data or recover previous versions of objects. For instance, when a newer copy of an object is uploaded, the previous copy is stored as a version with metadata describing its version and relationship to the current object.



87 <https://console.cloud.google.com/start>

88 <https://cloud.google.com/storage/docs/gsutil>

89 <https://cloud.google.com/storage/docs/xml-api/overview>

90 <https://azure.microsoft.com/en-us/features/azure-portal/>

91 <https://azure.microsoft.com/en-us/documentation/articles/storage-create-storage-account/>

92 <https://azure.microsoft.com/en-us/documentation/articles/storage-powershell-guide-full/>

93 <https://azure.microsoft.com/en-us/documentation/articles/storage-azure-cli/>

94 <https://azure.microsoft.com/en-us/documentation/articles/storage-samples/>

Amazon

Amazon S3 supports storing variants of an object in the same bucket.⁹⁵ If an object is deleted, Amazon inserts a delete marker, allowing users to restore the previous version. Administrators can delete versions without affecting the current version, and Amazon S3 also supports lifecycle policies, which automate the deletion of versions based on administrator-defined rules.^{96,97} Administrators can temporarily disable versioning if necessary, but once versioning is enabled on a particular bucket, it cannot be completely disabled. Versioning is enabled on a per-bucket basis.

Metadata for versioning uses two objects that have the same key value, but different version IDs.

Dell EMC

The Dell EMC ECS solution supports Amazon S3-style versioning. Administrators can restore or retrieve objects using the S3 API.⁹⁸

Google

Google provides object versioning that allows users to undelete data or roll back to previous versions of data.⁹⁹ Object versioning can be enabled or disabled on a per-bucket basis. Google Cloud Storage also supports object lifecycle management, where administrators can configure rules for automatic deletion.¹⁰⁰

With respect to versioning metadata, every object has a “generation” and a “metageneration.” The generation is updated every time an object is overwritten. The metageneration is updated every time the metadata of an object is changed.

Microsoft

Microsoft allows users to take snapshots of blobs, which are read-only versions of a blob taken at a single point in time.¹⁰¹

Snapshots have the same name as the original blob, but have a different DateTime value, which describes the time at which the snapshot was taken.

Multi-site, replication, consistency, and failover capabilities

With public cloud providers, customers acknowledge a service level agreement (SLA) that defines below what percentage of availability customers receive a service credit or refund. Durability of data is also guaranteed in many SLAs. Ensuring that data is replicated to multiple sites is the first step to ensuring data is resilient to faults.

95 <http://docs.aws.amazon.com/AmazonS3/latest/dev/Versioning.html>

96 <http://docs.aws.amazon.com/AmazonS3/latest/dev/DeletingObjectVersions.html>

97 <http://docs.aws.amazon.com/AmazonS3/latest/dev/object-lifecycle-mgmt.html>

98 <https://www.emc.com/collateral/white-papers/h14071-ecs-architectural-guide-wp.pdf>

99 <https://cloud.google.com/storage/docs/object-versioning>

100 <https://cloud.google.com/storage/docs/lifecycle>

101 <https://msdn.microsoft.com/en-us/library/azure/hh488361.aspx>

Amazon

By default, Amazon S3 stores data in two separate availability zones, which are discrete AWS facilities. Data stored with Amazon S3 is designed to service the concurrent loss of data in two facilities.¹⁰² Amazon S3 provides features such as cross-region replication, which allows administrators to copy S3 objects to a second AWS region.¹⁰³

Amazon S3 provides read-after-write consistency for PUTS of new objects in a bucket, and provides eventual consistency for PUTS that overwrite current objects.¹⁰⁴ If a GET or HEAD operation is performed on an object that hasn't been created yet, Amazon S3 offers eventual consistency for read-after-write.

Dell EMC

The Dell EMC ECS provides local replication as well as the flexibility to deploy ECS in multiple locations. Upon data ingestion, ECS ensures all data is triple-mirrored, with each copy of data managed by a separate storage node. Later, in the background, ECS begins to perform erasure coding. Erasure coding (12+4 scheme) splits data into 16 fragments and distributes them across the available nodes in the ECS. After data is erasure coded, the triple-mirrored copies are removed. ECS can read the data without any decoding or reconstruction. For more information on the erasure coding methods possible with ECS, see the ECS Architectural Guide.¹⁰⁵

When deployed to multiple sites, ECS can be geographically resistant to disaster. Multiple ECS deployments can be managed through a single pane and help ensure data is geo-replicated. With four or more ECS deployments, ECS actually saves storage space while data remains highly available and durable.

A multi-site deployment of Dell EMC ECS provides strong consistency when functioning normally. When a location in a ECS multi-site deployment fails in some fashion, Dell EMC ECS reverts to an eventual consistency model. Depending on what location is designated as the primary location, ECS designates which site acquires object ownership. After the failover situation is resolved, ECS automatically returns to a strong consistency model.

Google

Google provides information on data replication in their FAQ: "Your data is replicated in multiple data centers that are geographically distributed for greater availability."¹⁰⁶ Administrators can configure Google Cloud Storage buckets for multi-region locations, meaning data is distributed in more than one region.¹⁰⁷

Google Storage offers strong consistency for read-after-write, read-after-metadata-update, and read-after-delete. Certain actions, like listing objects in buckets, are eventually consistent operations.¹⁰⁸ Google allows the user to control the period of time an object is cached, which may cause an object to not appear strongly consistent.

102 <https://aws.amazon.com/s3/faqs/>

103 <http://docs.aws.amazon.com/AmazonS3/latest/dev/crr.html>

104 <http://docs.aws.amazon.com/AmazonS3/latest/dev/Introduction.html#ConsistencyModel>

105 <https://www.emc.com/collateral/white-papers/h14071-ecs-architectural-guide-wp.pdf>

106 <https://cloud.google.com/storage/docs/faq>

107 <https://cloud.google.com/storage/docs/bucket-locations>

108 <https://cloud.google.com/storage/docs/consistency>

Microsoft

During Azure storage account creation, users can choose a replication setting. Currently, there are four options:¹⁰⁹

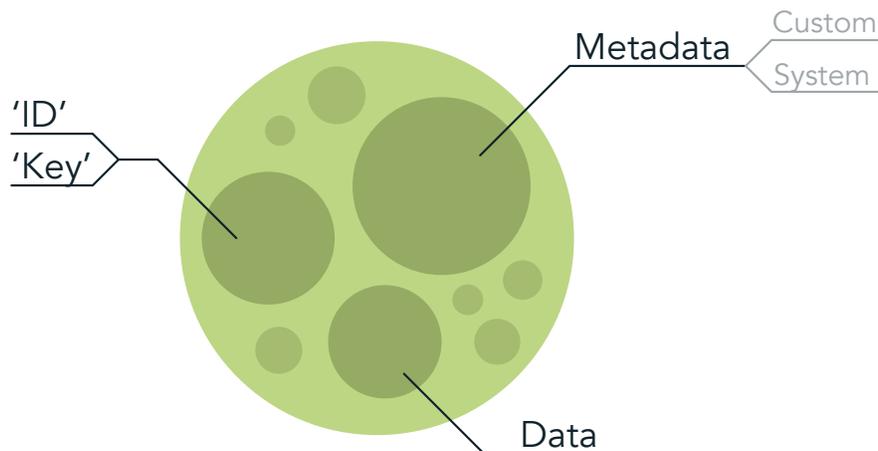
- Locally redundant storage
- Zone-redundant storage
- Geo-redundant storage
- Read-access geo-redundant storage

Locally redundant storage ensures data is replicated three times to ensure resistance to upgrades and faults in the same facility. Zone-redundant ensures data is replicated three times among separate facilities. Geo-redundant storage stores data in facilities hundreds of miles apart and keeps six copies of the data. Read-access geo-redundant storage simply provides read access to the non-primary copies of the data. Microsoft also provides guidance on what to do if an Azure outage occurs, such as while transferring data stored at a secondary location or pointing your applications at a secondary resource.¹¹⁰

Microsoft provides both optimistic concurrency and pessimistic concurrency when dealing with blobs.¹¹¹ If users don't designate a method, Azure Blob storage defaults to a last-write-wins strategy. Optimistic concurrency means update operations verify whether the data has changed since the last read. Pessimistic concurrency allows users to acquire exclusive leases on certain blobs.

Metadata and metadata search

With object storage, data is stored in a non-hierarchical manner, and each bit of data, or "object," has associated metadata that contains information on the data contained in the object or other custom data. Metadata may be configured by the user or set by the system. For instance, an object might have a custom metadata field describing the business organization that created it, and other system metadata describing the object's time of creation.



109 <https://azure.microsoft.com/en-us/documentation/articles/storage-redundancy/>

110 <https://azure.microsoft.com/en-us/documentation/articles/storage-disaster-recovery-guidance/>

111 <https://azure.microsoft.com/en-us/documentation/articles/storage-concurrency/>

Amazon

Each Amazon S3 object includes three components—the data, a key, and metadata.¹¹² The key value ensures all objects are uniquely identified. Object metadata is set at the time of upload—after upload, the metadata cannot be changed. The key value can have certain delimiters that can help the AWS portal infer folder structure, much like traditional file systems.

System metadata is metadata that Amazon S3 has predefined. Users can edit some types of system metadata, but other metadata such as encryption type can be changed depending on user and bucket settings. Custom metadata can be any set of key-value pairs as defined by Amazon S3.

Amazon allows users to search objects by the object key.¹¹³

Dell EMC

The ECS solution provides object storage that includes the actual data, system metadata, and custom user metadata. System metadata are fields defined by Dell EMC that may or may not be user-configurable. Custom user metadata is any other information defined by the user that is formatted as key-value pair.¹¹⁴ Metadata is available using the various APIs supported by Dell EMC ECS, such as the Amazon S3 API.

ECS maintains an index of up to five metadata fields. These are then searchable based on this metadata. Thus, metadata searches can return objects and their associated data. This searching and indexing feature is part of the Dell EMC Metadata Search Extension.¹¹⁵

Google

Google Cloud Storage objects have an object name, metadata, and the actual data. The object name is a unique value that identifies the object. Certain metadata is set by Google and can't be changed.¹¹⁶ There is a range of metadata fields that every object must have, and custom metadata can be changed or added after the object is uploaded.¹¹⁷

Google provides the ability to get metadata by the object name.

Microsoft

Microsoft Azure blobs contain system properties and user-defined metadata, as well as the actual data and a unique name for the blob. User-defined metadata can contain any information the user needs for their application. System properties are defined by Microsoft Azure and cannot be changed; however, some of their values are user-configurable.¹¹⁸ The user can change metadata after the object is uploaded.

Microsoft Azure supports a preview version of blob indexing, which allows users to index and search a pre-defined list of file types like PDF, text files, Microsoft Office documents, and more. User-defined metadata is also indexed for searching.¹¹⁹ Microsoft emphasizes that this feature is in preview and shouldn't be used for production environments.

112 <http://docs.aws.amazon.com/AmazonS3/latest/dev/UsingMetadata.html>

113 <http://docs.aws.amazon.com/AmazonS3/latest/UG/searching-for-objects-by-prefix.html>

114 <https://www.emc.com/collateral/white-papers/h14071-ecs-architectural-guide-wp.pdf>

115 <http://www.emc.com/collateral/TechnicalDocument/docu62943.pdf>

116 <https://cloud.google.com/storage/docs/gsutil/addlhelp/WorkingWithObjectMetadata>

117 <https://cloud.google.com/storage/docs/gsutil/commands/setmeta>

118 <https://azure.microsoft.com/en-us/documentation/articles/storage-properties-metadata/>

119 <https://azure.microsoft.com/en-us/documentation/articles/search-howto-indexing-azure-blob-storage/>

Cloud object storage features

In the table below, we provide an at-a-glance overview of the features we analyzed. While this is not a comprehensive list of all features available with every platform and competitor, we selected features that may be important to a business looking for an object storage solution.

Access control			
Amazon S3	Dell EMC ECS	Google Cloud Storage	Microsoft Azure Blob storage
Uses a combination of ACLs and AWS Identity and Access Management. Other AWS products may also integrate with S3, such as the AWS Directory Service	Provides integration with Active Directory, Lightweight Directory Access Protocol, and Keystone	Provides ACLs, Google Cloud Identity and Access Management, signed URLs, and signed policy documents	Uses Azure Active Directory role-based access control
Ability to provide time-limited access to certain resources	ECS has various protocols that can be configured independently	Provides access logs in CSV format	Able to provide time-limited access to resources
Access logs provided in space delimited files	All protocols are monitored and logged separately. Logs are also available for ECS disk operations and management functions Extends the S3 API to include rich ACLs		Access logs are provided in a specific format outlined on their website
Object change notifications			
Amazon S3	Dell EMC ECS	Google Cloud Storage	Microsoft Azure Blob storage
Depends on AWS products like Amazon Simple Notification Service, Amazon Simple Queue Service, and AWS Lambda	Does not currently support change notifications	Applications can subscribe to change notifications by sending a watch request on a bucket Provides notifications via HTTPS POST requests	Does not support native change notifications The Azure WebJobs SDK allows the creation of custom tasks that can perform change notification
Storage protocols			
Amazon S3	Dell EMC ECS	Google Cloud Storage	Microsoft Azure Blob storage
Only supports object storage; however, AWS also provides solutions like AWS Elastic File System (EFS) which provides files storage for use with AWS Elastic Compute Cloud (EC2) instances	Supports multiple protocols including object storage, NFS, and Apache® Hadoop® Distributed File System (HDFS)	Only supports object storage	Only supports object storage, but Azure File storage provides SMB-based file sharing
Supported APIs			
Amazon S3	Dell EMC ECS	Google Cloud Storage	Microsoft Azure Blob storage
Supports the Amazon S3 REST API	Supports multiple APIs such as Amazon S3, OpenStack Swift, Centra Content Addressable Storage (CAS), and the Dell EMC Atmos API	Supports the Google Cloud Storage JSON API and the Google Cloud Storage XML API. Both APIs are RESTful	Blob storage is accessed through the Azure Blob Service REST API

Management

Amazon S3	Dell EMC ECS	Google Cloud Storage	Microsoft Azure Blob storage
Provides the AWS Management Console (a web-based portal), the AWS Command Line Interface, and control through the S3 API	Supports management through the RESTful ECS Management API	The Google Cloud Platform Console provides a web-based portal for management, while the gsutil Tool provides a CLI for management	The Azure Portal provides storage administrators with a web-based management portal
	Includes the ECS Management Portal, allowing management of the entire stack from object storage to disk performance	Also provides management through the Google APIs	A PowerShell™ module integrates Azure management into PowerShell, while the Azure CLI provides support on other platforms
	Also supports management through various APIs such as S3, Swift, and others		

Versioning

Amazon S3	Dell EMC ECS	Google Cloud Storage	Microsoft Azure Blob storage
Versioning is enabled per bucket. Lifecycle Management allows automatic deletion of older versions	Supports Amazon S3-style versioning for easy retrieval and restoration through the S3 API	Object versioning can be enabled or disabled on a per-bucket basis	Microsoft Azure Blob storage provides snapshots of blobs, taken at a single point in time. These are manual
Versioning cannot be turned off once turned on		Lifecycle management allows older versions of an object to be automatically deleted	

Multi-site, failover, and replication capabilities

Amazon S3	Dell EMC ECS	Google Cloud Storage	Microsoft Azure Blob storage
Amazon reports data is stored in at least two availability zones (discrete facilities)	Provides local disk failure tolerance, and ECS deployments can be linked and managed as one	Google reports “Your data is replicated in multiple data centers that are geographically distributed for greater availability” ¹²⁰	Azure users configure replication settings upon account creation—these range from locally redundant to geo-redundant storage
Cross-region replication allows buckets to be replicated to a second region	Active-Active architecture means storage can be accessed from the closest deployment, reducing latency		Geo-redundant storage replicates data at least six times in facilities hundreds of miles apart
	The level of failure tolerance is configured by the business, not another provider		

Metadata search capabilities

Amazon S3	Dell EMC ECS	Google Cloud Storage	Microsoft Azure Blob storage
Administrators may retrieve objects by ‘key’ or name	ECS provides metadata search and indexing capabilities as part of the Dell EMC Metadata Search Extension	Google provides the ability to get an object by the object name	In Preview, not recommended for production environments: blob indexing allows metadata search, as well as indexing of certain file types such as PDFs and Microsoft Office documents
	Metadata is available through the various APIs supported by ECS		

120 <https://cloud.google.com/storage/docs/faq>

Private, hybrid, or public?

The three public cloud vendors in this paper all deliver the many benefits of public cloud options. For instance, the public cloud providers have made the process of provisioning IT infrastructure increasingly easy—shortening the time it takes to get to production. With no datacenter, hardware acquisition, or provisioning required, businesses may simply pay per month for services from Amazon, Microsoft, or Google. Low up-front costs can make this model attractive.

However, with the public cloud come other concerns. The public cloud presents problems that arise from data sovereignty, industry regulations, and the potential for vendor lock-in. If the company uses applications or only develops applications that use the Google Storage API, for instance, those applications won't work if the enterprise needs to move data to an on-site datacenter or another public cloud vendor. Because the cost of public cloud solutions may change based on the number of API calls or the number of GB stored, monthly bills may vary, potentially disrupting budgets and company finances.

Because Dell EMC ECS allows businesses to deploy object storage wherever it is needed, it enables greater control over the location of data, helping address data residency and sovereignty concerns. For instance, ECS provides deep insights into disk performance, network performance, CPU utilization, and many other metrics to aid administrators and reduce bottlenecks. In addition, security is controlled directly by company administrators, allowing them to implement practices that meet industry regulations that come with storing sensitive data.

With hybrid cloud, businesses can leverage both public cloud providers and their own on-premise resources. The Dell EMC ECS solution provides an array of APIs that are compatible with public cloud providers, allowing developers to use both public cloud resources and on-premise resources. This can help reduce vendor lock-in and limit dependence on a single cloud provider or API. While some public cloud providers like Amazon provide solutions such as the AWS Storage Gateway, which helps integrate on-premises storage in the cloud, no public cloud provider offers a turnkey solution to enable hybrid cloud capabilities like Dell EMC ECS.

Conclusion

We examined cloud object storage solutions from four providers—Amazon Simple Storage Service (S3), Dell EMC Elastic Cloud Storage (ECS), Google Cloud Storage, and Microsoft Azure Blob storage.

We found identifiable differences in their feature sets and overall implementations of cloud object storage. Amazon, Dell EMC, Google, and Microsoft each have their own unique APIs and management platforms—meaning developers must use that API exclusively. The Dell EMC ECS solution provides various protocols, APIs, and physical storage options to create flexible storage for many different applications. While some public cloud providers have years in the object storage market and large market share, it's important to keep storage and APIs vendor-agnostic and open to avoid lock-in.

The cloud storage space is always changing and evolving to meet customer demands. We have created a snapshot guide of these four cloud object storage providers as the providers presented them in public materials available on this report's date. We continue to expect these offerings to grow in size and features in this competitive field.

This project was commissioned by Dell EMC.



Facts matter.®

Principled Technologies is a registered trademark of Principled Technologies, Inc.
All other product names are the trademarks of their respective owners.

DISCLAIMER OF WARRANTIES; LIMITATION OF LIABILITY:

Principled Technologies, Inc. has made reasonable efforts to ensure the accuracy and validity of its testing, however, Principled Technologies, Inc. specifically disclaims any warranty, expressed or implied, relating to the test results and analysis, their accuracy, completeness or quality, including any implied warranty of fitness for any particular purpose. All persons or entities relying on the results of any testing do so at their own risk, and agree that Principled Technologies, Inc., its employees and its subcontractors shall have no liability whatsoever from any claim of loss or damage on account of any alleged error or defect in any testing procedure or result.

In no event shall Principled Technologies, Inc. be liable for indirect, special, incidental, or consequential damages in connection with its testing, even if advised of the possibility of such damages. In no event shall Principled Technologies, Inc.'s liability, including for direct damages, exceed the amounts paid in connection with Principled Technologies, Inc.'s testing. Customer's sole and exclusive remedies are as set forth herein.