Cloud Storage: Adoption, Practice and Deployment

An Outlook Report from Storage Strategies NOW

April 4, 2011

By Deni Connor, Patrick H. Corrigan and James E. Bagley
Client Relations: Phylis Bockelman

Storage Strategies NOW
8815 Mountain Path Circle
Austin, Texas 78759
(512) 345-3850

SSG-NOW.COM

Sponsors 4
What is cloud storage? 5
How is cloud storage deployed? 7
Public Cloud 7
Private Cloud 7
Hybrid Cloud 7
Community Cloud 7
Cooperative Cloud 7
Cloud Requirements 8
Multi-tenancy 8
Security 8
Secure Transmission Channel 8
Data Verification 9
Audit Trails 9
Performance 9
Quality of Service (QoS) 9
Data Protection and Availability 10
Customer Data Protection 11
Retention, Compliance, eDiscovery 11
Manageability 11
Data & Metadata Portability 11
Metering and Billing 12
The Cloud Storage Market – An IT Professionals Survey 13
Cloud Storage Trends 26
Personal Cloud Servers 26
Customer Name: Sony Pictures Imageworks 27
Cloud Storage APIs 28
Block Systems, File Systems and Object Storage 28
Block Interfaces 28
File Interfaces 29
Common Internet File System (CIFS) 29
Network File System (NFS) 29
Other File Systems 30
Object Storage 30
REST (or ReST) 30
SOAP 30

Cloud APIs by Vendor 31
Amazon S3 APIs 31
Eucalyptus APIs 31
Mezeo APIs 31
Nirvanix APIs 32
Open Grid Forum Open Cloud Computing Interface (*OCCI*) 32
OpenStack Cloud Computing and Storage 32
Rackspace Cloud Files APIs 32
SMEStorage 33
SNIA Cloud Data Management Interface 33
SNIA Content-aware eXtensible Access Method Storage API (XAM) 34
VMware vCloud API 34
Windows Azure Storage Services REST API 35
Cloud Computing Initiatives 36
Cloud Computing Interoperability Forum 36
Distributed Management Platform Task Force, Inc. (DMTF) 36
Open Data Center Alliance 36
Development Platforms: EMC Atmos 37
Best practices 38
Cloud Gateway and Enablement Providers 39
Vendor Name: Avere Systems 39
Cloud Gateway and Enablement Vendors Tables 44

Sponsors

AVERE

SNIA
Cloud Storage Initiative

What is cloud storage?

There are nearly as many definitions of cloud storage as there are providers of cloud services. In simplest terms, cloud storage is data storage or services hosted remotely on servers and storage devices on the Internet or a similar private network, usually hosted by a third party.

Cloud storage is a subset of cloud computing, in which the term cloud refers to the wide area network infrastructure, including switches and routers, for a packet-switched network. When capitalized, cloud usually refers to the public data network, including the Internet.

Cloud computing has probably been best defined by the National Institute of Standards and Technology (NIST) as:

> "Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."

In general, cloud storage has the following attributes:

- Resource pooling and multi-tenancy. Multiple customers can use the same storage infrastructure. Resources are pooled and multiple cloud resources (servers, storage and applications) are assigned and unassigned to customers as needed. Resources are location-independent and transparent to the user -- in most cases the customer does not control or need know about the actual location of a particular resource. A customer's data is protected from access by other customers. In the case of private clouds, multiple departments or business units within an organization share infrastructure and applications.

- Scalable and elastic. Virtualized storage that is for all practical purposes infinitely and immediately expandable. Storage can be expanded or contracted as needed easily and efficiently.

- Accessible via standard Internet APIs and communications protocols including HTTP, FTP, XML, SOAP and REST.

- Service-based. Customers typically have no capital costs (CAPEX) and pay for storage as a service (OPEX).

- Pricing is normally based on usage. Customers typically pay a per-gigabyte rate for upload and download and a per-gigabyte fee for monthly storage. In addition, some providers charge for each data access request based on reads, writes, etc.

- Shared and collaborative. Because cloud storage is usually accessed via the Internet it allows data access from multiple locations and multiple users.

- On-demand self-service. Customers can typically manage their storage service using some sort of management console. Customer service agents are usually only needed for problem resolution and similar tasks.

From the customer point of view, the storage appears as one or more volumes or servers. As more storage is needed it is available, eliminating much of the need for buying new local servers and storage. Capacity planning shifts from capital expense (CAPEX) and staffing expenditures to operation expense (OPEX) for upload, download and storage costs. Staffing time can also be reduced with cloud storage, allowing the reallocation of IT staff to more strategic business initiatives. Data is deployed to cloud storage either through Web-based applications or through Web services application programming interfaces (APIs). Web-based applications are often used for manual access to data or management functions, while APIs are used for more automated or transparent approaches. Since standard APIs and communications protocols are used, the physical location of the data becomes irrelevant, since it can be made available virtually anywhere via the Internet or private network. This also means that cloud data can be easily replicated to multiple locations for fault tolerance, high availability and other purposes, often without involvement of the customer.

APIs are the particular rules and specifications that allow software programs to access particular services. Cloud storage APIs are the rules and specifications for accessing cloud storage services. Cloud storage APIs are usually based on *simple object access protocol* (SOAP) or *representational state transfer* (REST). Cloud storage providers such as Amazon, Nirvanix and Mezeo Software have created their own proprietary APIs for accessing their storage resources. Although a number of providers allow the use of Amazon's API to access their services, not all do. Also, other services and service-specific APIs, such as the Nirvanix offering, provide features that are different from Amazon's.

A number of organizations have developed or are developing methods to generalize cloud storage access. The most common approach is to create an abstraction layer or a set of programming libraries that "front-end" proprietary APIs, effectively creating a common API for supported services.

Cloud Data Management Interface (CDMI)
Probably the most ambitious project is being spearheaded by the Storage Networking Industry Association (SNIA). SNIA's Cloud Data Management Interface (CDMI) is an attempt to standardize the methods for accessing and managing cloud data. SNIA's website (http://www.snia.org/tech_activities/standards/curr_standards/cdmi) gives the following description of CDMI and its purpose:

> "The Cloud Data Management Interface defines the functional interface that applications will use to create, retrieve, update and delete data elements from the Cloud. As part of this interface the client will be able to discover the capabilities of the cloud storage offering and use this interface to manage containers and the data that is placed in them. In addition, metadata can be set on containers and their contained data elements through this interface.
>
> "This interface is also used by administrative and management applications to manage containers, accounts, security access and monitoring/billing information, even for storage that is accessible by other protocols. The capabilities of the underlying storage and data services are exposed so that clients can understand the offering."

The Cloud Storage Library
The Cloud Storage Library project is made up of a PHP client library that lets customers integrate with common APIs that both Amazon S3 and Nirvanix support. http://code.google.com/p/cloudstorageapi

CloudStorage.API
The CloudStorage.API can be used to connect application to cloud storage and enable persistence ignorance and testability for cloud applications. http://cloudstorageapi.codeplex.com/

Cloud loop
Cloud loop is a universal adapter for cloud storage that gives users a single common interface for storing and managing data in the cloud. http://www.cloudloop.com

Simple Cloud API
The Simple Cloud API brings cloud technologies to PHP and the PHPilosophy to the cloud, starting with common interfaces for three cloud application services – file, document and simple queue services. http://simplecloudapi.org

How is cloud storage deployed?

Public Cloud

In the largest sense, the public cloud is all of the commercially available, shared-tenancy computing resources, based on the cloud computing model, available via the Internet.

Private Cloud

There are those who say the term private cloud is just a marketing phrase and others who say it is an oxymoron, since the "cloud" traditionally exists outside of corporate walls. For our purposes, however, a private cloud entails the use of cloud technology in a manner such that the infrastructure is only used or controlled by a single (corporate) entity and is typically not shared with other organizations. Private cloud resources can be located anywhere, including a corporate data center or a service provider's facility, but are accessed via private networks or virtual private networks (VPNs). We will not specifically address the private cloud in this report.

Hybrid Cloud

The term *hybrid cloud* has two definitions–either a combination of cloud services and local physical hardware to provide a single service, or two or more clouds (public, private or community) that are used together but appear to the user as a single service. The second definition is also called a *combined cloud*. We will use the first definition here.

For storage, a hybrid cloud typically consists of a local appliance that provides storage, caching, deduplication, backup or other services as a front end for cloud storage services. Because the appliance stores or caches cloud data locally it can reduce latency and effectively improve performance It can also reduce access costs since local requests for already cached data do not have to be serviced by the cloud storage provider. Companies that provide hybrid appliances include Axcient, Barracuda Networks, CTERA and StorSimple.

Community Cloud

A *community cloud* has the shared tenancy aspects of a public cloud service but limits access to a few organizations. Community clouds include various clouds being implemented by U.S. Government agencies such as the Defense Information Systems Agency (DISA), the General Services Administration (GSA) and the National Space and Aeronautics Agency (NASA). It also includes services like Google's GovCloud, which provides secure cloud services to local government agencies.

Cooperative Cloud

In a cooperative cloud participants contribute unused local storage for use by other cloud members. Each member stores data on other members' available and contributed disk space. A member of the co-op typically contributes disk space equal to the amount they consume in the cloud. Currently the only commercial cooperative cloud that we are aware of is managed by Symform (http://www.symform.com).

Cloud Requirements

Multi-tenancy

In a multi-tenant model, the provider's resources are pooled to serve many customers, with different physical and virtual resources dynamically assigned and reassigned as needed. This elastic, shared-storage model allows the provider to cost-effectively provision resources and bill by usage rather than available storage.

Security

Secure cloud storage requires a secure transmission channel, methods to make sure the data itself is secure and physical security at the storage site. Securing data usually entails using a combination of encryption, authentication and authorization.

Encryption is the process of scrambling data in such a manner as to make it unreadable without special information, called a key, to make it readable again. Encryption can be used to safeguard data while it is being transmitted (in transit) as well as while it is being stored (at rest) in the cloud.

Authentication is the process of determining someone or something is who or what they say they are. Authentication can employ passwords, biometrics, identifying tokens and other means. Generally speaking, the more complex the authentication process the more secure it is. Long passwords of numbers and uppercase and lowercase are more secure than simple passwords. Two-factor authentication, employing, for example, a physical token and a complex password, is more secure yet.

Authorization determines who or what has access to data and what level of each authorized person or device has. For secure cloud storage, access must be controlled for the transmission channel, the data source site and the cloud storage sites.

Physical security must be provided for the cloud storage site. Even if data is encrypted, physical hardware and communications channels can be sabotaged.

Secure Transmission Channel

The four primary methods used to secure network communications are:

- Transport Layer Security (TLS) and Secure Sockets Layer (SSL). TLS, like its predecessor SSL, is a cryptographic protocol that provides data security across a network. TLS is a client/server protocol that uses a handshaking procedure to establish a secure connection.

- Hypertext Transfer Protocol Secure (HTTPS). HTTPS combines the standard Internet HTTP protocol with TLS to secure web-based transactions.

- Private Networks. A private network uses dedicated connections to link sites. An example of this is the U.S. Courts' *data communications network* (DCN), which connects the various facilities of the courts. A private network provides security because, for the most part, access is limited to those within that particular organization. Security can potentially be compromised, however, because most private networks have gateways connected to the public Internet. Connecting to a public cloud service from a private network would require the use of a gateway.

- Virtual Private Networks (VPNs). A VPN creates a virtual, encrypted "tunnel" connecting two or more locations across the Internet. A VPN connection requires some method of authentication for a user or device to join the network. VPN connections can be made by individual user computers, servers, or network routers and gateways. VPNs provide most of the benefits of private networks without the cost.

Data Verification

How do you know that the data you are storing in the cloud are actually there? Due diligence requires tools and procedures to verify cloud data. Unfortunately, the standard methods of checksums, file and object comparisons, attribute verification and database transaction verification all create additional network traffic and data accesses, which can degrade performance and, depending on the service provider, engender additional costs.

At this time there appear to be no standard methods for cloud data verification, although several groups of researchers are working on the issue:

http://eprint.iacr.org/2009/281.pdf
http://www.iaeng.org/publication/WCE2010/WCE2010_pp485-486.pdf
http://www.cs.odu.edu/~mukka/cs775s10/Presentations/Security.6.pdf

Audit Trails

Most organizations, especially medium and large organizations, need the ability to verify not only who has access to cloud data, but, in many cases, who specifically has accessed that data. When choosing cloud storage services it is important to make sure that the audit information you require is easily and quickly available.

Performance

Cloud storage performance can be categorized in two ways–actual transmission speed and latency. Although slow transmission speed will cause latency, there are other factors involved as well. Some of the factors that affect cloud storage performance include:

- Available network bandwidth is a major factor. Unless high-speed dedicated links over relatively short distances are employed, the network will be a significant bottleneck.
- Efficiency of the systems employed to provide multi-tenant concurrent access at the storage provider's data center.
- Use of compression and deduplication. These functions create latency. Compressing or deduplicating data prior to transmission, however, can decrease the time and bandwidth needed for transmission. , in many cases, can improve overall performance.
- Caching. Caching data locally provides immediate access to the cached data and also typically reduces the amount of data that must be uploaded and downloaded from the remote storage site.

Quality of Service (QoS)

Quality of service (QoS) refers to the ability to guarantee performance, apply different levels of performance for different classes of data, or both. QoS applies to cloud storage as well as the network itself. Network, QoS, however, is beyond the scope of this report.

Since cloud computing and cloud storage are based on a multi-tenant model, some mechanism should be in place to guarantee that each tenant can effectively access its data without undue performance degradation. In a multi-tenancy model, costs can go down as the number of tenants increase, but so can performance for each tenant. In addition, certain types of data may be more time-sensitive than others. For example, looking up credit card information for eCommerce is usually a higher priority than storing data for archive purposes.

A number of providers offer tools to monitor QoS but few offer their customers QoS management tools. Most providers have service level agreements (SLAs), but in many cases they are quite vague and open-ended. For example, one provider's SLA states:

> "{provider} will use commercially reasonable efforts to make {service} available with a Monthly Uptime Percentage…of at least 99.9%…"

Data Protection and Availability

To assure that data is protected from loss and available when needed, cloud storage providers generally take some or all of the following measures:

- Physical site security. Cloud storage providers typically provide some level of physical security for their servers. The degree of security, however, can vary greatly from provider to provider. Smaller cloud storage sites might be housed in a co-location facility with minimal access control.
- Protection against power loss. A cloud storage site should be equipped with backup power to prevent failure in case of a power outage.
- Protection against loss of network access. Some data centers use multiple network connections, supplied by separate carriers. with entry points at different locations in the building, to guard against loss of communications due to cable cuts and similar events.
- Data redundancy. Storage servers may use RAID-type methods to provide for data availability in case of drive failure.
- Server redundancy and server fail-over. Storage sites typically employ some type of server fail-over mechanism to provide continued data and service availability in case of a server failure.
- Redundant data sites. Many vendors mirror data between multiple sites to provide data availability in case of a site disaster. Some providers, such as Amazon S3, allow the customer to specify the level of redundancy required for different classes of data. Reduced redundancy generally means lower costs but greater risks of data loss.
- Levels of redundancy. Some providers offer different levels of redundancy at different price points. The idea is to reduce storage costs for less-critical data. For example, Amazon offers Reduced Redundancy Storage (RRS), which "stores objects on multiple devices across multiple facilities, providing 400 times the durability of a typical disk drive, but does not replicate objects as many times as standard Amazon S3 storage. thus is even more cost effective." (http://aws.amazon.com/s3/#pricing)
- Versioning and data retention. Versioning provides the ability to maintain multiple versions of data. Some vendors provide versioning for data objects stored on their servers. With many providers, data retention policies can be applied to determine the number of copies of files or data object to be retained and the period of retention. Versioning can also be provided by various storage and backup systems and applications.
- Accessibility of cloud storage as live data. In many cases, customers use local servers to provide live data and use cloud storage for backup or replication only. In such cases, being able to access data stored in the cloud as if it is on a local server or volume can help prevent downtime if local data becomes unavailable.
- Backup to tape or other media. Even redundant data can be lost or corrupted, so backing up data is still important. For example, an attempted software upgrade caused Google to lose data for 150,000 Gmail accounts. Although it took several days, they were able to restore most of the lost data from backup tapes.

- Data availability and retention in case of billing or contract disputes. Different providers have different policies governing access to and retention of data if billing or other disputes arise.

Customer Data Protection

Even if cloud storage vendors provide reasonable data protection, in most cases it is a good practice for customers to protect data locally as well. This protects the customer in case the cloud storage service, for whatever reason, is unavailable. Also, recovering data from local storage is typically faster than recovering from cloud storage. This is especially important if a significant amount of data needs to be recovered in a short period of time.

Retention, Compliance, eDiscovery

Long-term data retention, regulatory compliance. Federal eDiscovery requirements all present challenges when data is stored in the cloud. Security and access control, searchability of data and long-term data integrity are major issues with cloud storage. These concerns must be addressed by IT and legal staff and should be part of any service level agreement (SLA) with the cloud service provider.

Manageability

Currently each provider has their own mechanism for managing cloud storage and services. Different providers offer different management capabilities. IT staff must often use different management tools and management consoles to manage storage services from different providers.

The Storage Networking Industry Association (SNIA) has created two standardization initiatives to address these issues.

SNIA's Cloud Data Management Interface (CDMI) addresses many of these issues. CDMI focuses on the storage side of cloud data. You can find information about the initiative here:

http://www.snia.org/forums/csi/programs/CDMIportal

SNIA's eXtensible Access Method (XAM) is a standardized access method for accessing cloud data. It provides a standardized approach for storing, archiving and accessing data and its related metadata. One major goal of the XAM initiative is to assure data portability between cloud providers and cloud storage applications. XAM APIs can be used to build applications that conform to the standard. For more information:

http://www.snia.org/forums/xam

Both CDMI and XAM are described in detail later in this report.

Data & Metadata Portability

Because of the proprietary nature of cloud, the portability of data and metadata between cloud providers affect cloud-based applications and virtual machines as well as specialized storage sites, such as those dedicated to graphics, photos, etc.

Metadata can include security policies, access control lists, rankings and display order for graphics files on sites like Flickr, comments and much more. Unfortunately this metadata is not easily transportable between providers. First, it is often unclear who owns the metadata—the customer or the provider. Second, there are no standards for metadata, so each provider is free to maintain and store it in its own way. Third, many providers consider their metadata mechanisms part of their added value and do not want to make its transfer easy.

The portability of the data itself is a less complex issue, but still a major problem. For example, many graphics sites have no mechanism for bulk transfer of data—the client can only transfer one image at a time.

SNIA's CDMI attempts to address many of these issues. Until portability standards are established, it is a good idea to maintain local copies of data (always a good idea anyway) and detailed records of security policies, access control lists and other critical cloud metadata.

Metering and Billing

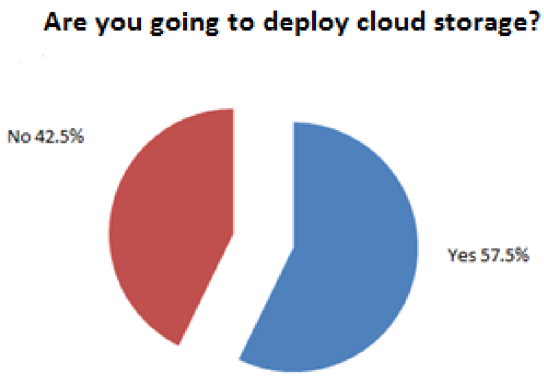
Cloud storage billing is typically based on the following:

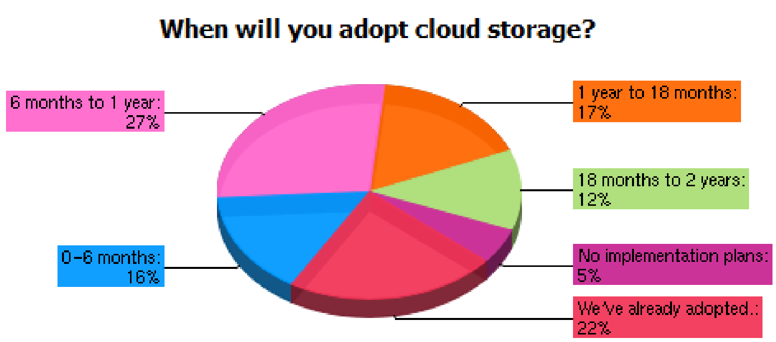
- Amount of data uploaded
- Amount of data downloaded
- Amount of data stored
- Number of requests and type of request

Most providers charge for data upload, download and actual storage capacity. Some charge for requests. For example, at this writing, Amazon S3's posted charges for requests are either $0.10 per 1,000 requests or $0.01 per 10,000 requests, depending on request type. Others such as Nasuni offer a total package price that includes the cost of their cloud gateway and the cost of storing data at one of its many partners.

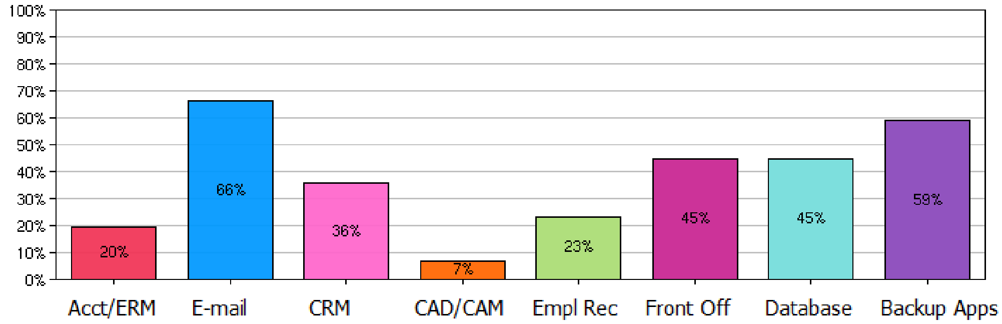
The Cloud Storage Market – An IT Professionals Survey

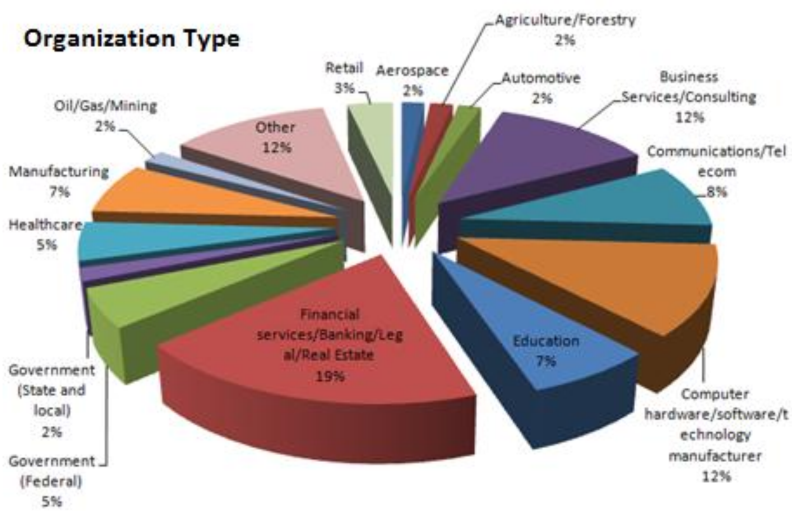
The Storage Networking Industry Association (SNIA) and Storage Strategies NOW (SSG-NOW) conducted a survey of IT professionals, both at the CIO-level and of operations staff, on their plans for deploying cloud storage. The survey, sent to 5,000 recipients during the first quarter of 2011, netted 133 qualified respondents in North America. The results of the survey and the significance of the results are related here.



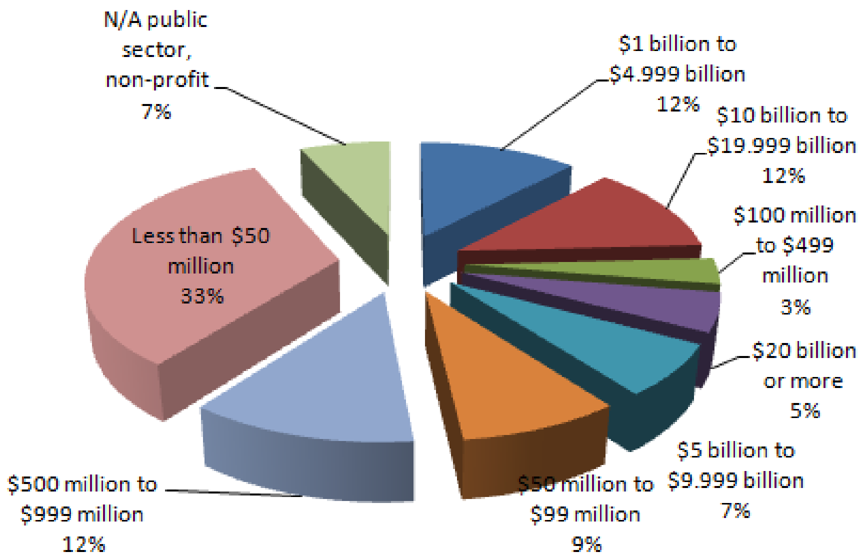


What applications are you most likely to store in the storage cloud?

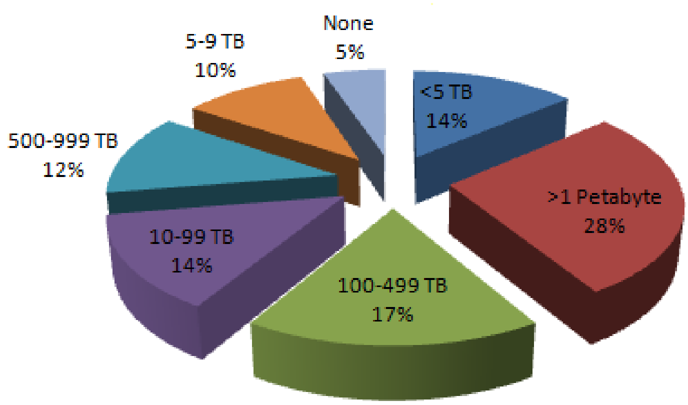


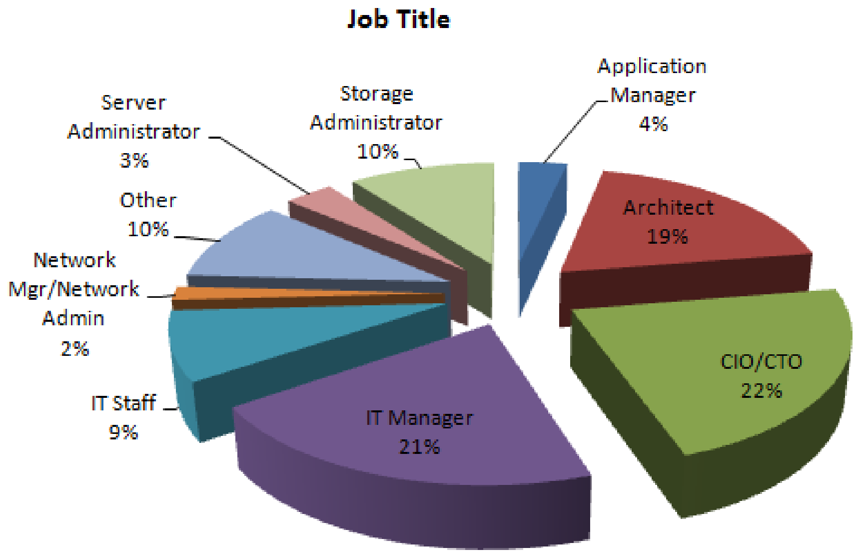


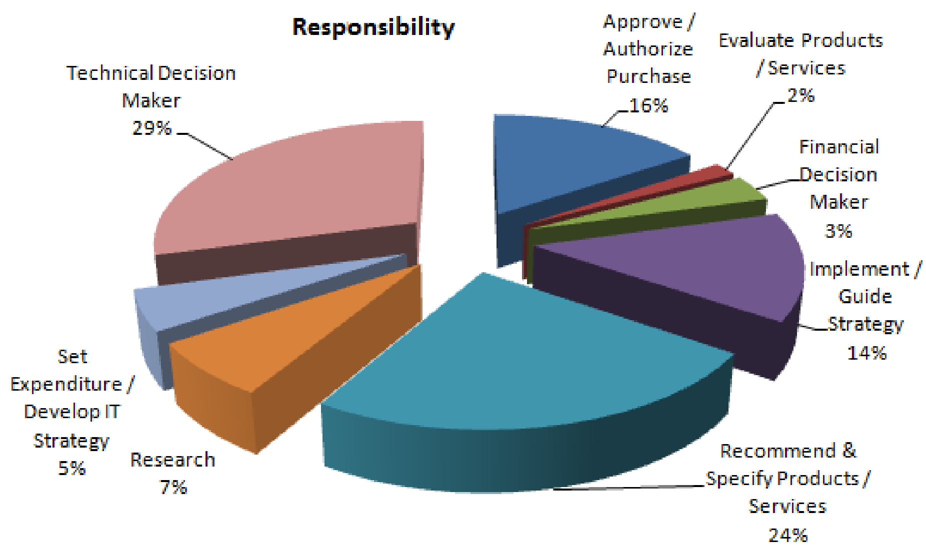
Annual revenue

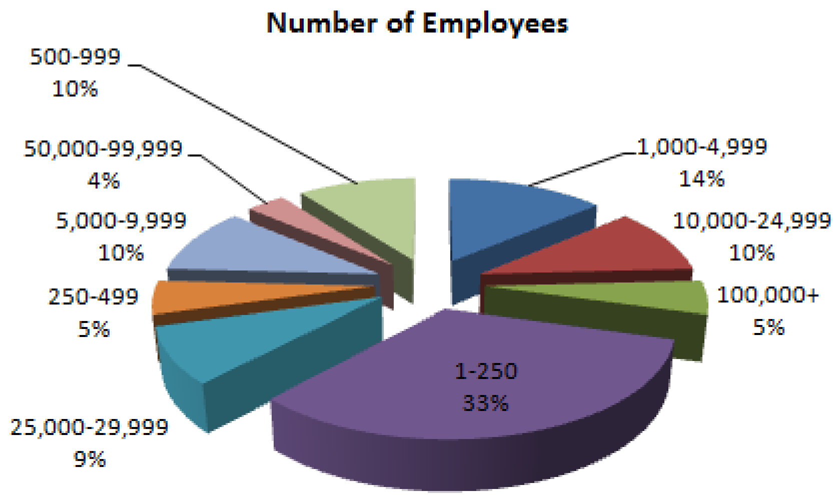


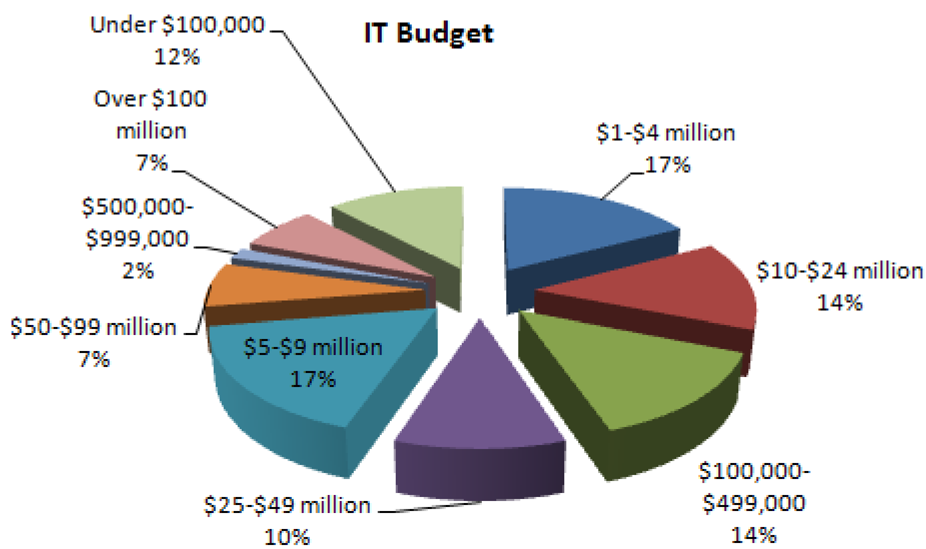
Amount of storage managed




Job Title
Application Manager 4%
Architect 19%
CIO/CTO 22%
IT Manager 21%
IT Staff 9%
Network Mgr/Network Admin 2%
Other 10%
Server Administrator 3%
Storage Administrator 10%


Responsibility
Approve / Authorize Purchase 16%
Evaluate Products / Services 2%
Financial Decision Maker 3%
Implement / Guide Strategy 14%
Recommend & Specify Products / Services 24%
Research 7%
Set Expenditure / Develop IT Strategy 5%
Technical Decision Maker 29%

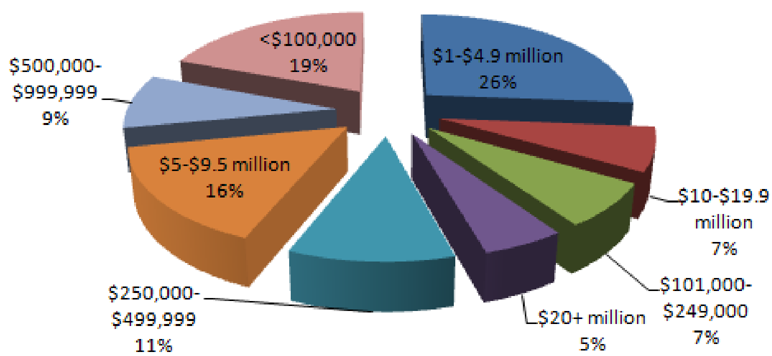

Number of Employees
500-999
10%
50,000-99,999
4%
5,000-9,999
10%
250-499
5%
25,000-29,999
9%
1,000-4,999
14%
10,000-24,999
10%
100,000+
5%
1-250
33%


IT Budget
Under $100,000
12%
Over $100
million
7%
$500,000-
$999,000
2%
$50-$99 million
7%
$5-$9 million
17%
$25-$49 million
10%
$1-$4 million
17%
$10-$24 million
14%
$100,000-
$499,000
14%

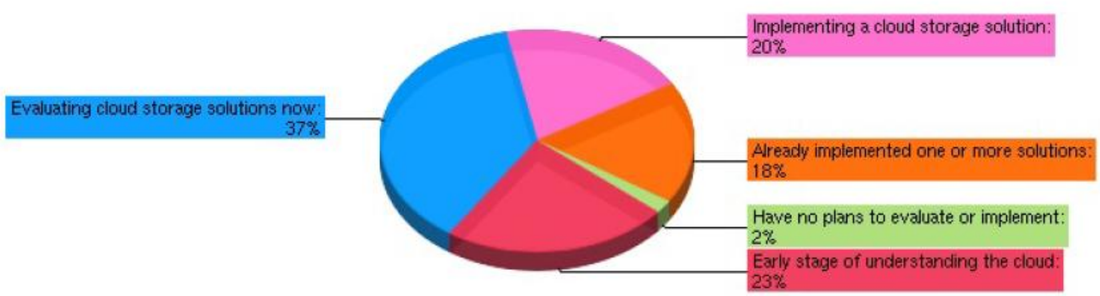
What are your organization's most important IT priorities over the next 12-18 months? (Please rank by importance, 1=Most important to 10=Least important)

Improve data backup and recovery	1
Increased use of server virtualization	2
Information security initiatives	3
Business continuity/disaster recovery	4
Manage data growth	5
Upgrade network infrastructure	6
Major application deployments	7
Data center consolidation	8
Large-scale desktop/laptop PC deployments	9
Regulatory compliance initiatives	10

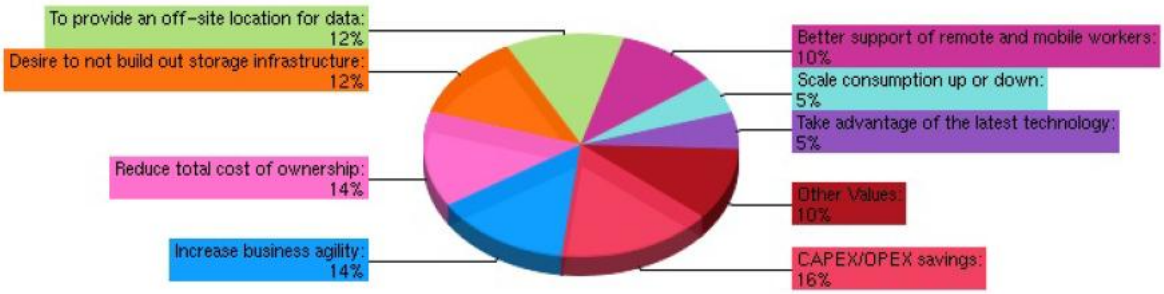
Annual Storage Expenditures



What is your implementation stage for cloud storage?



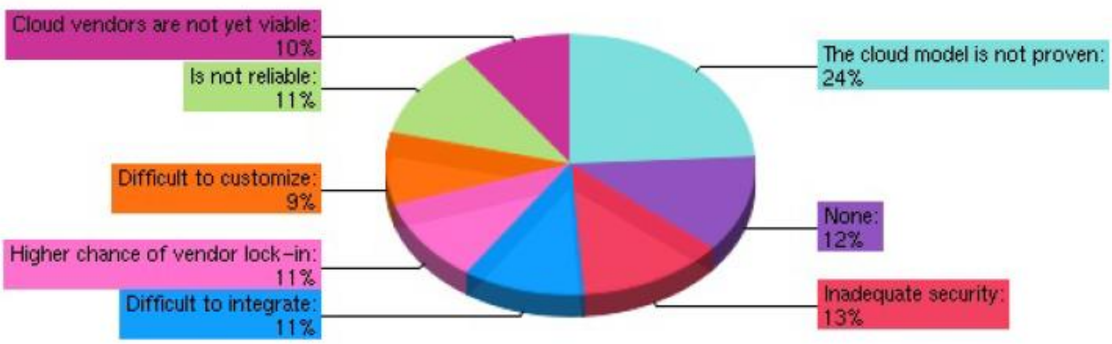
What is your primary reason for storing data in the storage cloud?



What storage applications are you planning to or have deployed in the storage cloud (Select all that apply)

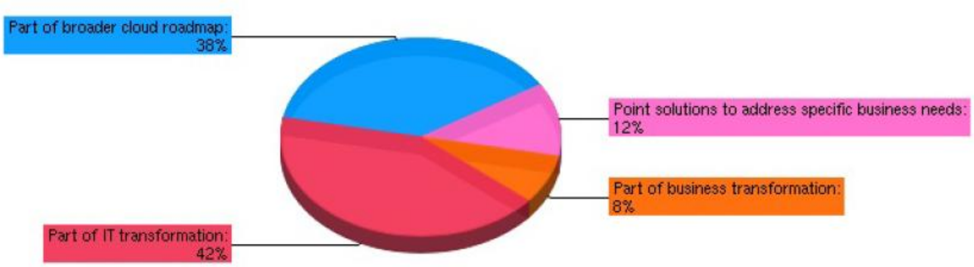
	Cloud 2011	Cloud 2012-2014	Private cloud 2011	Private cloud 2012-2014	Public/hybrid Cloud 2011	Public/hybrid Cloud 2012-2014	Do not expect to deploy in cloud	Totals
Primary backup and restore	15.2%	20.3%	16.5%	16.5%	7.6%	6.3%	17.7%	100%
Second copy of backup data	12.7%	29.6%	15.5%	11.3%	5.6%	7.0%	18.3%	100%
Archiving and retention	11.4%	19.0%	21.5%	20.3%	7.6%	10.1%	10.1%	100%
Data warehousing	4.6%	13.8%	13.8%	20.0%	6.2%	3.1%	38.5%	100%
Primary data, Tier 1 Storage	1.6%	14.8%	9.8%	18.0%	3.3%	1.6%	50.8%	100%
Secondary data, Tier 2 Storage	6.5%	19.4%	3.2%	16.1%	4.8%	12.9%	37.1%	100%
Business continuity and disaster recovery	9.1%	28.8%	10.6%	21.2%	3.0%	9.1%	18.2%	10I0%
Storage bursting	3.3%	13.3%	5.0%	11.7%	6.7%	13.3%	46.7%	100%
Data tiering	4.7%	15.6%	7.8%	15.6%	10.9%	10.9%	34.4%	100%
eDiscovery/Compliance	1.6%	11.5%	9.8%	19.7%	6.6%	3.3%	47.55	100%
Replace centralized computing/global access	1.6%	13.1%	11.5%	11.5%	4.9%	8.2%	49.2%	100%
Host files, unstructured data	10.4%	14.9%	14.9%	19.4%	4.5%	10.4%	25.4%	100%
Database, transactional structured data	9.1%	12.1%	6.1%	16.7%	6.1%	6.1%	43.9%	100%
E-mail	19.8%	17.3%	12.3%	14.8%	4.9%	7.4%	23.5%	100%
None of the above	6.3%	1.6%	7.9%	3.2%	1.6%	3.2%	76.2%	100%
Other	0.09%	3.6%	5.4%	8.9%	0.0%	0.0%	82.1%	100%

What is the single biggest misperception about cloud storage?



How long will you retain data in the storage cloud?

What are your future cloud storage plans?



What were the biggest concerns that you or your organization has or had to overcome before signing up for cloud storage? (Please type in 1, 2, or 3 to select your 3 largest concerns. 1=Largest concern.)

Security	1
Performance (i.e. latency)	2
Price/Pricing model	3
Availability	4
Regulatory and/or compliance requirements	5
Internal company culture	6
Vendor lock-in	7
Difficult to get started (i.e. need to learn new tools	8
Unable to license my desired software on the cloud storage service	9
Unable to customize to my needs	10

What cloud enablement and gateway vendors are you planning to use for your public cloud storage plans? (Please select your top 3.)

Other	1
CommVault/Dell	2
Symantec	3
Hitachi Data Systems	4
F5	5
Nirvanix	6
Avere Systems	7
Nasuni	8
StorSimple	9
CTERA	10
Gladinet	11
Cleversafe	12
Scality	13
Cirtas	14
Panzura	15
Mezeo Software	16
Seven10 Storage Software	17
DataDirect Networks	18
Twinstrata	19

What cloud infrastructure vendors are you planning to use for your public cloud storage plans? (Please select your top three)

EMC	1
Other	2
HP	3
Hitachi Data Systems	4
IBM	5
Oracle	6
3PAR	7
NetApp	8
Avere Systems	9
DataDirect	10
Cleversafe	11
Caringo	12
Nexsan	13
F5	14

Zetta	15
Gluster	16
Permabit	17
Nexenta	18
Mezeo Software	19
RELDATA	20
IceWeb	21

Please select the top three (3) vendors of products or services for your private cloud storage plans.

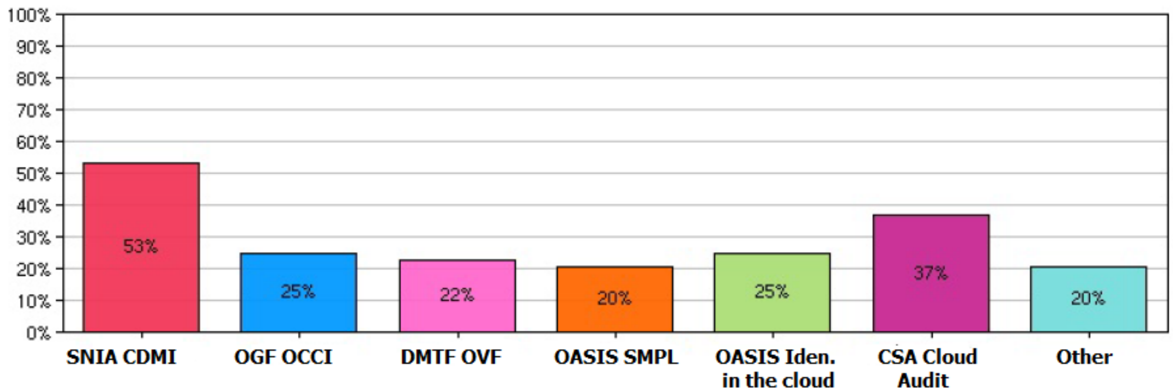
EMC	1
Other	2
CommVault/Dell	3
HP	4
IBM	5
Hitachi Data Systems	6
NetApp	7
DataDirect Networks	8
Cirtas	9
Oracle	10
3PAR	11
Gladinet	12
Gluster	13
Nasuni	14
CTERA	15
Caringo	16
Avere Systems	17
Symantec	18
StorSimple	19
Cleversafe	20
F5	21
Nexsan	22
RELDATA	23
IceWeb	24
Nexenta	25
Nirvanix	26
Zetta	27
Mezeo Software	28
Panzura	29
Permabit	30
Twinstrata	31
Scality	32
Seven10 Storage Software	33

Why did you originally sign up with your cloud storage service provider? (Please select the three most important reasons.)

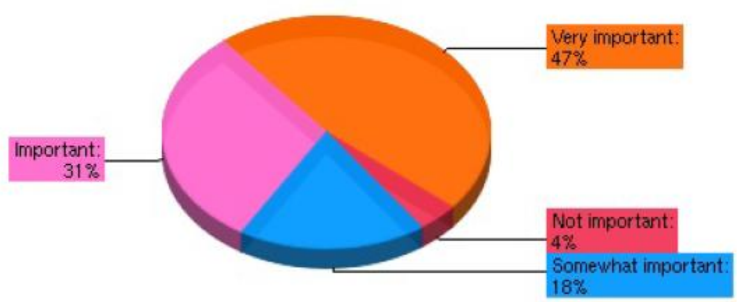
On-demand pricing (pay only for what I use, no CAPEX)	1
Speed-to-market (ability to implement a solution in a short period of time	2
Trust the vendor's ability to operate/leverage/use world class datacenters	3
Outsource IT infrastructure I) (focus on other tasks)	4
Flexibility (choice of operating system, software, programming languages)	5
Scalability (add resources on demand)	6

Service features and functionality	7
Availability (SLAs)	8
Other	9
Global footprint	10
My boss told me to	11

What cloud storage standards will you require being included in RFPs?



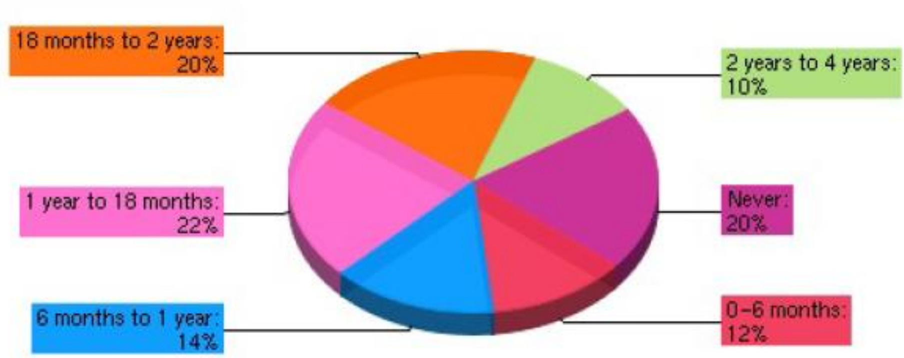
How important are standards for public storage clouds?



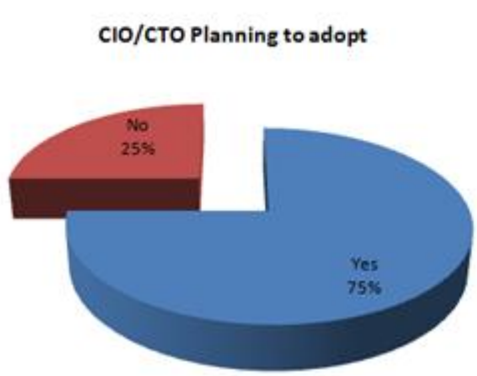
Which of these standards are important for public/hybrid storage cloud?

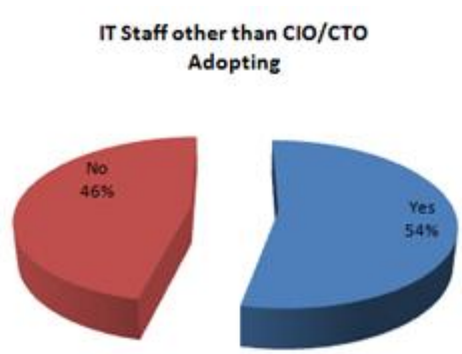
	1=Not important	2=Somewhat important	3=Important	4=Very Important	Totals
SNIA	20.0%	22.0%	28.0%	30.0%	100%
OGF OCCI	20.0%	42.0%	28.0%	10.0%	100%
DMTF OVF	18.0%	32.0%	42.0%	8.0%	100%
OASIS SMPL	20.0%	36.0%	30.0%	14.0%	100%
OASIS Identity in the Cloud	22.0%	26.0%	38.0%	14.0%	100%
CSA Cloud Audit	20.0%	32/0%	26/0%	22.0%	100%

When are you going to require vendors to comply with cloud storage standards?

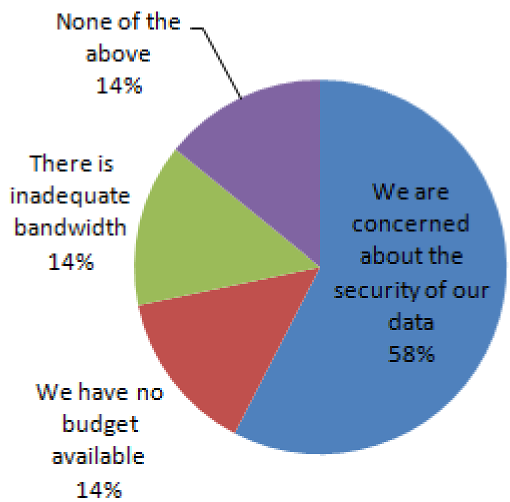


By CIO/CTO

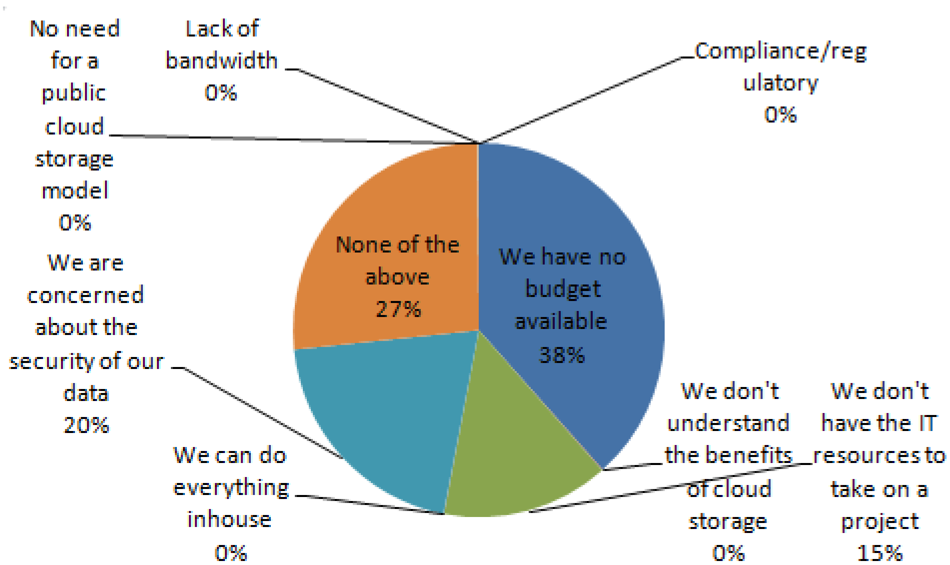




CIO/CTO Not Adopting



IT Operations Not Adopting



Cloud Storage Trends

Public cloud storage has primarily been adopted by small and medium-sized businesses and mainly used for archiving of email, front office applications and secondary storage and backup. Hybrid backup appliances play a major role here since they provide both local and cloud storage through a single interface.

Some larger organizations are using cloud storage as part of their IT infrastructure, but a greater number of them are mostly testing the waters, usually by hosting secondary data in the cloud. Concerns about security, regulatory compliance, data integrity and availability, as well as the lack of tools and standards for auditing and data verification, have been impediments to more wide-scale adoption of cloud storage. The lack of effective standards in APIs, at least until recently. differences in service offerings and billing models from different providers has only added to the confusion.

Because the storage as a service (STaaS) model of cloud storage requires different accounting methods for storage cost, internal corporate practices and policies can also be an impediment to quick adoption. It is not always easy to directly compare the costs of local storage with the costs of cloud storage. often internal charge-back policies need to be modified as well.

As the issues of standards, security, accessibility, billing, etc. get resolved we can expect more organizations to adopt cloud storage as part of their IT infrastructure.

Personal Cloud Servers

One relatively new development is the personal cloud server. Although they don't fulfill all the requirements of the definition of cloud storage (multi-tenancy, usage-based pricing, etc.) they do provide cost-effective storage that can be accessed via the Internet. Some examples personal cloud storage servers are Pogoplug (and devices that use the Pogoplug service) and the Iomega cloud products.

Pogoplug (http://www.pogoplug.com) provides network-attached devices that support external USB 2.0 hard disks. When the devices are set up and registered with the Pogoplug service they are accessible from the Internet. Seagate Technology (http://www.seagate.com/www/en-us/products/network_storage/file-sharing) provides a device called GoFlexNet that docks up to two of Seagate's GoFlex portable hard disks. GoFlexNet, which uses the Pogoplug service, also has a USB 2.0 port for an external disk as well. These devices offer different levels of security, sharing and access control.

Iomega (http://iomega.com) has added similar cloud capabilities to its ix2-200 and ix4-200d network storage devices. In addition, they have introduced a Home Media Network Hard Drive, Cloud Edition targeted to the home market.

Customer Name: Sony Pictures Imageworks

Type of Organization: Media and Entertainment

Location and Environment: Culver City, Calif.

Contact name: Nick Bali, Sr. Systems Engineer, Systems R&D

Amount and type of data managed: 1 petabyte

Infrastructure: Avere FXT cluster with replication and backup to tape

Challenges:

- Multiple employees working on data concurrently and data being exchanged between worksites
- Inefficient access to data using WAN accelerators, Wide Area File Services and caching products
- Each site keeps 2TB of data; 20-30% of which constantly changes and need to be replicated to headquarters.
- “Looking at the read/write workloads at remote sites, we found that the majority of traffic is reads,” says Bali. “Our goal was to solve this issue of edge data access, something that we struggled with for some time.”
- Capacity requirements too great for memory-based approaches
- Proprietary formats required a change to single-vendor solution
- Public cloud storage cost-prohibitive

Solution:

- Deployed Avere’s tiered NAS, which separate data delivery from data retention and deliver both more efficiently
- Data moved to optimal storage tier (SSD or HDD) on Avere FXT based on demand. Inactive data moved to traditional NAS
- Clustered NAS system provides high-availability
- Algorithms monitor data access and adjust performance, distribute workload and minimize disk access
- Scalable system which can be increased by adding storage or appliances
- Three node cluster points to data center in Los Angeles, ensuring that data is always up to date

Benefit:

- Read/write speeds if remote users are staggering
- Meeting recovery point objectives
- Heterogeneous deployment of Avere FXT appliances and traditional NAS protect existing investments
- “Imageworks was able to move workers to the Avere tiered NAS system utterly behind the scenes, without any interruption in service,” says Bali. “It’s completely transparent to them. I can add new volumes and new users and they don’t see it.”

Cloud Storage APIs

Cloud Application Programming Interfaces (APIs), as the name implies, are sets of program function calls that must be applied to any application programs that access cloud storage. Since most applications access storage, applying modifications to all programs is a daunting task. one that most enterprises will want to avoid.

As of now, there is no standard API available for cloud storage, although standards development initiatives are underway from the Storage Network Industry Association (SNIA) and the Distributed Management Task Force (DMTF). These standards are in their early formation and the specification and approval of these standards through the national (ANSI) and international (ISO) standards organizations will take many years.

Without a standard, cloud storage users must either choose a cloud provider and develop the applications modifications necessary, or use a gateway appliance (hardware or software). This is a primary roadblock to cloud storage adoption. Aside from the concerns regarding data access and security, the fear of vendor lock-in and data getting erased over a contract dispute are serious issues for organizations of all sizes. This has relegated cloud storage to secondary copies of backup data, archive, eDiscovery data and disaster recovery compliance for many organizations.

Within this application subset, many successful vendors of data protection software and appliances exist that provide cloud interfaces embedded in the protection application and related appliances. For many organizations, this is a powerful deployment of the cloud. Rather than managing or renting an offsite facility for disaster recovery, cloud storage, often coupled with an appliance to provide local file access and reduce bandwidth requirements for data movement to the cloud, is a perfect solution for these critical, but generic, applications.

In the very long term, the importance of standards adoption will be reflected in a heterogeneous access to a wide variety of cloud storage providers, much as we now take the use of WiFi for granted. At one time, choice of wireless network equipment vendors was necessary because each supplier had their own protocols and equipment was not interoperable. Client radio interface cards and access point routers were more than 100 times more expensive than today. Today we also take for granted the three prong connector used to access the electric power grid.

In the future, cloud storage will be available in a similar manner. But for the time being, the APIs necessary to provision, secure, read and write cloud storage will be a necessary consideration. As one cloud gateway appliance vendor put it "proprietary cloud storage APIs make every service provider a Hotel California – you can check out but you can't leave."

Block Systems, File Systems and Object Storage

There are three basic methods of accessing stored data. Block storage is the lowest level of access and is how individual drives and Storage Area Networks are accessed. File systems provide a higher level of access but have operating system dependencies because each operating system provides the low level drivers and file management structure. Object storage is a newer and higher level of storage which defines storage in terms of its use, be it a structured database, unstructured files like email and text documents or compressed binary format files such as music and pictures. The advantage of object storage is that the system providing access has a higher level of definition of the intended use of the storage in advance. This makes typical storage operations such as deduplication, replication, compression and updates more efficient. The disadvantage is that an object store needs, guess what - an API.

Block Interfaces

The simplest way to provide generic access to storage is by faking the using computer into thinking it is talking to a locally attached drive. As always, there are a number of ways of pulling this off. We will focus on the most commonly deployed of these interfaces. A grandfather of all drive media interfaces is the Small Computer System Interface (SCSI – pronounced 'scuzzy'). This interface includes a data packet format that can be emulated with a software driver that takes standard drive command packets and adds an Internet Protocol envelope that can be routed over a local or wide area network. This interface is known as iSCSI – pronounced 'i-scuzzy') and operates over standard Ethernet equipment. The deployment of iSCSI to provide file access to remote files over the internet has been in use for many years, using an approach called a Virtual Private Network (VPN). With appropriate network configuration establishment. after a secure log-in procedure, a remote user can access files using block level SCSI commands, thus 'faking' the using computer into thinking it is talking to a local drive. While this interface is inefficient over a wide area network connection, it is reliable and drivers and support for iSCSI is available in all modern operating systems. It is a simple implementation and is very often incorporated with a cloud gateway appliance to provide a plug-and-play capability for access to the cloud. The gateway device, which has implemented a cloud API compatible with one or more cloud storage providers, simplifies and aggregates communications from the using computer network to improve bandwidth utilization and increase throughput.

File Interfaces

File interfaces have been in use for many years but require a higher level of operating system integration. The primary advantage of a file system is that the applications are separated from the low-level disk emulation operations. higher efficiencies can be gained. In addition, systems managers can implement file naming conventions that can span across many hundreds, even thousands, of client computers. The common buzzwords for this capability are 'Global Name Space. While a number of file systems are in use, the predominant network file systems are the Common Internet File System (CIFS) and the Network File System (NFS). While cloud storage service providers do not directly support these file systems due to the amount of bandwidth that they consume, most cloud access appliances support either CIFS, NFS, or both.

Common Internet File System (CIFS)

A direct descendant of the Microsoft Disk Operating System, CIFS is widely deployed and supported by many generations of Microsoft client and server operating systems. is therefore often the default file system in many Network Area Storage (NAS) systems and cloud appliances. Documented and maintained by Microsoft, CIFS is considered a baseline for most file server, or 'Filer' uses. Based on the widely used SMB (Server Message Block) network protocol, CIFS has become a key file sharing protocol due to its ubiquitous distribution and its inclusion of enhancements that improve suitability for internet file sharing. It is an integral part of workstation and server operating systems as well as embedded and appliance systems, including cloud appliances, file acceleration devices and wide area network optimizers.

Network File System (NFS)

NFS was developed by Sun Microsystems and first specified way back in 1984. NFS uses the Open Network Computing Remote Procedure Call (ONC RPC) system, which provides an open standard defined and published by the Internet Engineering Task Force (IETF) describing 'methods, behaviors, research or innovations' applicable to the systems which support NFS via Internet Protocol. Early versions of NFS include v1 and v2. NFS v2 was the first version released outside of Sun Microsystems. Open platform releases controlled by the IETF are set by Request for Comment (RFC) documents, which are incrementally issued until a consensus standard is defined. NFSv2 was defined by RFC 1094 in 1989. It had the then-reasonable limitation that only the first 2GB of a file could be accessed. NFSv2 used User Datagram Protocol, which could efficiently execute storage commands without maintaining a prior state. While fast and efficient, UDP has limitations that would haunt the early implementations of NFS.

NFSv3 was defined by RFC 1813 in 1995 and had the features that have made it the gold standard for open file protocols for over fifteen years. Features included 64bit file addressing and the inclusion of TCP protocol which allows routing of commands over a Wide Area Network (WAN). NFSv3 remains the predominant NFS implementation.

Version 4 (RFC 3010, December 2000; revised in RFC 3530, April 2003) was influenced by Andrew File System (from Carnegie-Melon University 'AFS') and CIFS. But it took until January of 2010 for NFSv4.1 to be finalized, partly because Sun Microsystems turned support over to the IETF in 2000, which retarded development and adoption. Most vendors that support NFSv3 are currently developing NFSv4 somewhere on the roadmap. CIFS compatibility is usually the stated reason. Parallel NFS is also a new variant the provides multiple paths to the file system that is prevalent on appliance and NAS system roadmaps.

NFS is common to open Unix and Linux operating systems including Solaris, AIX and HP-UX.

Other File Systems

Modern open-source file systems include Gluster File System (GFS) and Zettabyte File System (ZFS). Both have tremendous features for internal use on storage systems that support both block and file interfaces, but are usually too complex to be considered for a cloud API. Cloud storage infrastructure and access appliances have been developed using these file systems for the underlying storage.

Object Storage

Because it helps cloud storage providers to implement more efficient systems, proprietary APIs and the cloud storage standards including the SNIA CDMI and the DTMF are heavily focused on object storage features. In the near term, this will probably enhance the gateway appliance market as vendors in that area can claim support for current generic interfaces (iSCSI, CIFS, NFS) and roadmap support for coming industry standards. Representational State Transfer (REST) is a web-based architecture that uses the Hyper-Text Transfer Protocol (HTTP) as a means to communicate to Internet-based object storage systems. Simple Object Access Protocol (SOAP) provides programming within the Extensible Markup Language (XML) message format and can support a wide variety of user developed storage functions.

REST (or ReST)

REST is particularly applicable to cloud storage applications as it can be implemented completely within HTTP (web page programming) and has the communications features that we take for granted when using the worldwide web. Clients can issue requests to servers without concern as to the location or intermediary devices routing the request. Servers in turn can send programming information back to the client and have it execute the code. REST is a stateless architecture, which means that each message contains everything that the client or server needs to execute the request. By using Universal Resource Locator (URL) and Universal Resource Identifier (URI) from a browser or direct web interface, information can be exchanged between clients and cloud servers by using simple 'Get', 'Put', 'Post' and 'Delete' functions, which address the storage resource using a standard web address. Since REST is an architecture rather than a specific protocol, reference to the cloud storage provider's specific command functions is necessary.

SOAP

As its name (Simple Object Transfer Protocol) implies, SOAP is a protocol with a specific protocol standard, although with additional functionality by using an XML messaging format. This provides an application development environment which is very flexible and can provide considerable functionality for development of many applications, including cloud storage operations. SOAP is an industry standard managed by the W3C committee which maintains a wide range of standards within www protocols (hence W3C).

As the SOAP abstract states in the latest W3C version of the standard (April 2007):

> 'SOAP Version 1.2 is a lightweight protocol intended for exchanging structured information in a decentralized, distributed environment. "Part 1: Messaging Framework" defines, using XML technologies, an extensible messaging framework containing a message construct that can be exchanged over a variety of underlying protocols.'

Cloud APIs by Vendor

Amazon S3 APIs

Amazon Simple Storage Service (S3) was launched in 2006 as the first web service from the venerable merchandise purveyor. S3 stores objects up to 5TB in size, accompanied by up to 2KB of metadata. Objects are organized into buckets that must be owned by an Amazon Web Service account identified by a unique, user-assigned key. Virtual machine images used by the Amazon Elastic Compute Cloud (EC2) can be exported to S3 for permanent storage.
Buckets and objects can be created, listed and retrieved using either a REST or a SOAP interface. Additionally, objects can be downloaded using the HTTP GET interface and the BitTorrent protocol.

Requests are authorized using an access control list associated with each bucket and object. Bucket names and keys are chosen so that objects are addressable using HTTP URLs:

http://s3.amazonaws.com/bucket/key
http://bucket.s3.amazonaws.com/key

Objects are accessible by unmodified HTTP clients so S3 can be used to replace existing web hosting infrastructure. The authentication mechanism allows the bucket owner to create an authenticated URL that is valid for a specified period. The URL can be handed off to a third-party for access. Items can also be served as a BitTorrent feed. S3 can provide an interface to any BitTorrent client.

Eucalyptus APIs

Eucalyptus Systems is a relatively young venture-backed company that provides software infrastructure for developing private clouds. The APIs used are identical to Amazon S3 and EC2.

Mezeo APIs

Mezeo is a relatively young company with a very rich client application functionality implemented within REST constraints. Mezeo is currently supported by Gladinet, TwinStrata and CTERA cloud infrastructure and gateway offerings.

Methods supported include: GET, PUT, POST, DELETE, LOCK, UNLOCK and RESTORE Resources (note plurals indicate groups) include:

v bCloud, Account, Contacts, Contact, Container, ContainerContents, File, FileContent, Projects, Project, ProjectAssignments, ProjectAssignment, Shares, MetaContainers, MetaContainer, Permissions, Permission, RecycleBin, DeletedObject

See the Mezeo developer documentation at:
http://developer.mezeo.com/mezeo-developer-center/documentation/api-documentation/

Nirvanix APIs

Nirvanix came out of stealth mode in 2007 and operates seven data centers in the United States, Europe and Asia. Nirvanix provides both a software appliance that can be installed on customer equipment, as well as a REST or SOAP web APIs.

The CloudNAS software appliance, once installed, can be mounted as any storage device and provides CIFS and NFS support. The simplicity of this approach means that the global namespace feature of NAS can extend into a boundless, worldwide storage array that is transparent to end users. The system features encryption both in transit and once stored in the cloud. CloudNAS is a key to Nirvanix' rapid expansion into significant end user accounts.

In addition, the Nirvanix Web Services API feature-rich REST and SOAP interface exists for clients who wish to use this deeper control structure, which can be used over both HTTP and HTTPS. Methods supported provide functions for Authentication, Accounting, File System Management including base file and folder commands such as create, rename, copy, move. delete. Add, modify, delete. search custom metadata for files and folders. Transfer to upload or download files or parts of files, sharing, image operations, video format commands and audio file control. The CloudNAS appliance is written using this API.

Open Grid Forum Open Cloud Computing Interface (*OCCI*)

The Open Grid Forum (OGF) is a European-based standards body that is developing a standard for an Open Cloud Computing Interface (OCCI). This standard is at a very early stage but stakeholders in the cloud computing services industry should monitor the developing standards closely. OCCI is included by reference in the SNIA CDMI specification (see a following heading). There are two documents currently out for review regarding the OCCI, both released at the end of January, 2011. The first is a specification rendering of a cloud computing API using REST architecture. The second is a document describing automated negotiation and reservation of cloud computing services. Both documents can be reviewed by using this link: http://www.ogf.org/gf/docs/?public_comment

OpenStack Cloud Computing and Storage

OpenStack is an interest group developing a massively scalable cloud operating system platform in an open source environment. Open source code can be freely obtained under the Apache 2.0 license. Contributors include Dell, Rackspace, NASA, Citrix, Cisco, Canonical and over 50 other organizations. There are two interrelated projects, cloud computing and object storage, which are released independently but with the same code name. Early releases, code named version Bexar (no doubt after the Texas county where the Apaches once roamed and home of the Alamo) were completed in February of 2010, with the next scheduled releases, named Cactus, scheduled for Q2 2011. Information about OpenStack can be accessed using this link: http://www.openstack.org/index.php

Rackspace Cloud Files APIs

Rackspace, using OpenStack, has recently made its Cloud Files service generally available. A control panel GUI can be used to first establish an account and then move and manage files from a client/server into the Rackspace cloud. The GUI resembles a typical file management screen. Rackspace supports a REST API as well as language-specific API with documents and assets for C#/.NET, Python, PHP, Java and Ruby. Rackspace also publishes an interesting set of caveats and limitations regarding its Cloud Files service. As these limitations apply to any cloud storage service without a NAS or block-level storage gateway, they should be viewed as universal, or at least become discussion points with any cloud storage provider:

> Native support within your operating system. It is not possible to "mount" or "map" the Cloud Files storage system as a virtual hard-disk on your computer or server.

Disk mirroring or backup solutions that require byte/block level differences. There are no concepts of "appending data" or "file locking" operations within Cloud Files.

Data can be organized into storage compartments called "Containers", but Containers cannot be nested. Since there is only one top level of organization, you will not be able to upload nested directory/folder structures into Cloud Files unless a transformation is performed to flatten the structure.

There is no built-in concept of permissions or access controls in the Cloud Files system such as you may be accustomed to in a traditional file system.

There are no transcoding capabilities to read a file in one format and upload it in one or more other formats into Cloud Files.

Rackspace, as a long-term web hosting and services firm, has a well established support organization and has recently announced extended support by this group for cloud storage implementation services. Information on Cloud Files can be accessed by this link: http://docs.rackspacecloud.com/files/api/v1/cf-intro-20110124.pdf

SMEStorage

SMEStorage is a London, UK company that provides a server gateway appliance and client driver support for workstations, laptops, iPads, iPhones and Motorola Droid smart phones. It has built-in support for a large number of cloud storage services, currently including:

Amazon S3
RackSpace Cloud Files
Box.net
MobileMe
Microsoft SkyDrive
Microsoft Live Mesh
Evernote (import only)
Google Docs (currently import only)
Email-as-a-Cloud (the ability to use any email account as a storage cloud)
GMail-as-a-Cloud (the ability to use Google mail as a storage cloud)
FTP-as-a-Cloud (the ability to use FTP as a storage cloud)
WebDav (any WebDav enabled cloud)

Access is also provided via a GUI and includes file synchronization and recovery functions. It also supports a REST-based API that works above any supported cloud service and that has bindings for .NET, PHP, PERL, Ruby and Java. Operating systems support includes Windows, Macintosh and Linux.

SNIA Cloud Data Management Interface

The Storage Network Industry Association (SNIA) is the leading trade association in the computer storage field and a Platinum sponsor of this report. The SNIA Cloud Initiative has been focused on developing an all-encompassing standard that addresses legacy interfaces such as iSCSI, CIFS and NFS along with a substantial REST architected object store standard that covers a broad set of methods from the standard Get, Put, Post, Erase to methods that retrieve information regarding the cloud providers' CDMI implementation status. CDMI uses REST URI/URL namespace based on 'containers' (similar to the Amazon S3 'buckets') and a substantial lexicon of 'Methods.' The team and companies that contributed to the proposed architecture deserve tremendous recognition. CDMI Version 1.0 was released in April 2010 as a SNIA Technical Position, a beginning point on the long road to the development of an international standard. As stated at the beginning of this section, many years will pass before an ISO level standard is final, but this is a very complete document and provides a rationalized content for a cloud storage standard. CDMI can be considered a superset of APIs offered by the participants in the cloud storage and services industry.

In SNIA's words:

"SNIA standards are primarily related to data, storage. information management and address such challenges as interoperability, usability and complexity. They start out in SNIA-sponsored Technical Working Groups (member collaboration efforts centering on best practices) or as contributions from independent vendor collaborations. are subject to a rigorous processes involving checks and technology reviews by SNIA's Members, Technical Council. Board of Directors, conducted under provisions that respect and protect intellectual property and address licensing of essential claims."

A SNIA standard that has completed these processes is labeled a SNIA Technical Position, which signifies that the SNIA endorses and recommends the ideas, methodologies. technologies described. SNIA may take SNIA Technical Positions to ANSI, ISO. other national and international standards bodies for additional endorsement.'
Information regarding the SNIA Cloud Storage Initiative can be found at:

http://www.snia.org/cloud

The SNIA CDMI Version 1.0 Technical Position document can be found at:
http://www.snia.org/tech_activities/standards/curr_standards/cdmi/CDMI_SNIA_Architecture_v1.0.pdf

SNIA Content-aware eXtensible Access Method Storage API (XAM)

The XAM Interface specification defines a standard access API between 'Consumers' (application and management software) and 'Providers' (storage systems) to manage information storage services. XAM includes metadata definitions to accompany data to achieve application interoperability, storage transparency. automation for Information Life Cycle Management (ILM)-based practices, long term records retention. information security. XAM will be expanded over time to include other data types as well as support additional implementations based on the XAM API to XAM conformant storage systems. XAM is included by reference in the CDMI standard.
XAM is defined in three SNIA Technical Position documents, the latest versions 1.01 published in June 2009.
Part 1 addresses architecture:
http://www.snia.org/tech_activities/standards/curr_standards/xam/XAM_Arch_v1.01.pdf

Part 2 addresses a C programming language API:
http://www.snia.org/tech_activities/standards/curr_standards/xam/XAM_C_API_v1.01.pdf

Part 3 addresses a Java programming language API:
http://www.snia.org/tech_activities/standards/curr_standards/xam/XAM_Java_API_v1.01.pdf

VMware vCloud API

VMware released its vCloud API version 1.0 in April of 2010 with an update in August of 2010. The REST architected API includes HTTP URL/URI methods for browsing containers including log-in and authentication, provisioning, data center operations, re-configuring a vApp or re-configuring a Virtual Machine. Within a vCloud environment, a complete set of cloud storage and elastic computing is provided. The specification is concise and the API is consistent with VMware's operational command structure.

While the API is simple and concise, an extensive VMware infrastructure is necessary to operate the vCloud functionality. A white paper discussing the architecture can be found at:
http://www.vmware.com/files/pdf/VMware-Architecting-vCloud-WP.pdf

The vCloud API specification can be found at:
http://communities.vmware.com/docs/DOC-12464

Windows Azure Storage Services REST API

Windows Azure supports a REST URL/URI API with methods to access Account services, 'Blob' services, Queue services and Table services. The system provides persistent, redundant storage. A very well-developed and supported service, Azure has operated with the same version since September 2009. Account services are accessed first to establish ownership of containers that support two types of blobs, one optimized for streaming operations and another for random access. Queue support is for messaging between Azure services. is particularly useful in email-like operations. Table services allow the creation of tabled data sets which can be used for load balancing across storage nodes. No specific database schema is enforced, so users are free to develop data structures within the table service. With Microsoft support and a large and loyal ecosystem, Azure is a major power in the cloud storage provider industry and stakeholders should ensure continuing compatibility with this architecture. The SNIA CDMI is fully compliant with the Queue and Table methods.

The Azure API is documented at:
http://msdn.microsoft.com/en-us/library/dd179355.aspx

Cloud Computing Initiatives

Cloud computing initiatives are related to cloud storage initiatives in that cloud storage is often used for elastic computing resources. Development and testing of large VM deployments is a typical application for elastic computing because a large configuration can be quickly deployed, along with test data sets, without investment in owned test infrastructure. In order to achieve the longer-term vision of an unlimited computing scalability, with rapid expansion and contraction based on the using organization needs will require further advances in standards. As with all interoperability standards, these usually begin with academic exercises which are followed by interest group formations of manufacturers, users and industry associations. With stakeholders spanning high performance computing users in diverse fields ranging from aerospace to government. Following is a discussion of some of the major initiatives in this area.

Cloud Computing Interoperability Forum

More of a tradeshow platform than a serious standards group, this loose organization publishes a blog and provides a medium for advertising cloud-centric trade shows. Information can be found at:
http://groups.google.com/group/cloudforum

Distributed Management Platform Task Force, Inc. (DMTF)

The DMTF is a major standards body operating within the WT3 standards work group. DMTF is collaborating with SNIA on the CDMI standard and will likely be the roadmap towards both ANSI and ISO standard adoption. A recent collaboration between, SNIA, DMTF and NIST, the United States National Institute of Testing indicates the level of interest and support for cloud storage and cloud computing standards development. As we have stated many times in this report, international standards take many years and the donation of many technical resources. Our assessment is that this organization and collaborators will be the pathway for the ultimate 'three-prong' interface to utility-like resources for cloud computing and cloud storage resources.

A wealth of information regarding the task force can be accessed at: http://www.w3.org/

Open Data Center Alliance

The Open Data Center Alliance is a refreshingly end-user directed organization dedicated to encouraging vendors to provide interoperable products. A reading of this section indicates that there are many vendors with cloud offerings that require an end-user to check into the proverbial '*Hotel California*'. It is obvious to the casual observer that this puts the cloud vendor in a powerful position. the ODCA is a counter punch to establish an end user position in the standards and negotiation process. With such powerful organizations as BMW, Capgemini, China Life Insurance Company Ltd., China Unicom Group, Deutsche Bank, JPMorgan Chase, Lockheed Martin, Marriott International, National Australia Bank, Shell Global Solutions, Terremark and UBS on its steering committee, the Open Data Center Alliance is a serious developer of best practices for the large data center. For more information see: http://www.opendatacenteralliance.org/

Development Platforms: EMC Atmos

Due to the scale of integration to develop support for vCloud, EMC created a cloud platform as both cloud service and demonstration capability. EMC subsequently decided that operating as a cloud service provider and potentially competing with its customers, the best use of Atmos was as a turn-key platform that could be deployed by EMC enterprise customers either as a private cloud solution or a public cloud service platform. If your organization needs a turn-key, multi-petabyte data storage and distribution solution. if you have a big enough checkbook, Atmos could be what you need. The system combines massive scalability with automated data placement to help deliver content and information services anywhere in the world. For more information see: http://www.emc.com/products/detail/software/atmos.htm

Best practices

Moving to cloud storage provides many benefits, but also bears some risk as well. As with any other major IT endeavor, proceed with caution. Doing a little homework and study in advance can help prevent problems later on. The following, while not complete, is a list of some things you should consider when implementing cloud storage:

- Not all cloud services are the same. Make sure your service provider meets you needs. Ask yourself the following questions:
    - What degree of redundancy is provided? Some service providers offer multi-site redundancy, some only provide single-site redundancy (mirroring, etc.) while some provide no redundancy at all.
    - Is there automatic fail-over in case of a disk/server/site failure?
    - Does the provider offer versioning, or do they only store the most current version of a file or data object? Is there a retention period for deleted files? If they do offer versioning, how flexible is the retention policy?
    - Does the provider back up data? If so, what is the backup cycle and retention policy? In an emergency how long will it take to restore data? Even cloud service providers can have hardware and software failures that can cause data loss.
    - Does the provider offer an easy-to-use management console? Is it web-based and can it be accessed from any location in case of emergency?
    - Does the pricing structure fit your business model? For example, some vendors charge for every file access (read, write, etc.) in addition to per-gigabyte upload and download charges. If you are moving large blocks of data those access charges will be minimal. If you are doing primarily database lookups and updates, however, they can add up quickly.
- Determine the cloud deployment model that is best for you. Public, private, hybrid, community or cooperative models all have their pros and cons, but each does not fit every business model.
- Don't put your eggs in one basket. As with local storage make sure you back up your data.
- Do you need local storage as well as cloud storage? Local storage provides immediate access to your data, even if there are network issues. If this is important to you than consider replicating your cloud data locally.
- Read the contract. What liability does the provider have in case of data loss, etc.? What are the service commitments, if any? What happens if there is a billing dispute? Can your data be held hostage, or, even worse, deleted?
- Take things a step at a time. While cloud storage can solve many problems it can create new ones as well. There will be a learning curve for you and your staff. Taking small steps, at least initially, limits your exposure if things don't go as anticipated.

Cloud Gateway and Enablement Providers

A cloud storage gateway translates cloud storage APIs to file-based interfaces such as Windows Common Internet File System (CIFS) or Unix/Linux Network File System (NFS) or storage transport protocols such as iSCSI or Fibre Channel. The gateway, which resides at the customer premises, can be a dedicated appliance or server, a virtual appliance or server, or a software application running on a general purpose system. Dedicated gateways often provide additional features such compression, deduplication, encryption and caching as well as backup and recovery capabilities. In effect, a cloud storage gateway attempts to make cloud storage appear as local storage.

Vendor Name: Avere Systems

Product Name: FXT Series
Link to website: www.averesystems.com
Link to data sheet: www.averesystems.com/Avere%20FXT%20Data%20Sheet.pdf

Product Description: Avere Systems makes appliances that enable solid-state drive tiered network attached storage (NAS). While not typically a cloud gateway, the FXT appliances allow access to public and private clouds. The appliances keep copies of the most frequently accessed data close to the users on a local FXT SSD-based node or cluster of nodes and communicates with the mass storage in the cloud over the WAN. By staging the hot data on the FXT SSD nodes, customers maintain the fast data access of local storage with the unlimited scalability of cloud storage. Instead of limiting cloud storage to backup and other secondary storage applications, Avere FXT appliances allow the cloud to be used for primary storage by mitigating the performance and latency concerns and delivering application performance that is beyond traditional local NAS systems.

The FXT hardware platform consists of three models – the FXT 2300, 2500 and 2700. Each appliance contains DRAM and NVRAM to accelerate the performance of data. Data is stored in the 64KB bank of DRAM. Write performance is accelerated by 1GB of NVRAM. Each appliance contains eight 15K SAS disk drives or eight SLC flash-based solid state drives.

The FXT 2300 contains eight 146GB 15K SAS drives and supports an application working set of 1.2TB raw per appliance and up to 29TB raw per cluster. The FXT 2500 contains eight 450GB 15K SAS drives and supports an application working set of 3.6TB raw per appliance and up to 90TB raw per cluster. The FXT 2700 contains eight 64GB SLC Flash SSDs and supports an application working set of 512GB raw per appliance and up to 13TB raw per cluster.
The company has recently introduced Avere 2.0, a global namespace (GNS), which gives visibility into a distributed NAS infrastructure of Avere FXT Series Appliances and other vendors' filers. Avere 2.0 supports NFS and CIFS and unifies NAS systems into a single namespace that can be viewed from one centralized management interface.

With Avere's GNS, storage administrators can easily create and manage logical groupings of file-based resources regardless of physical location, presenting clients running NFS or CIFS with simplified and transparent access to data with a single mount point. Moreover, when data is moved from one storage server to another within the network, client access is automatically rerouted with no disruption or downtime. GNS is available immediately as part of the 2.0 release of the Avere operating system.

Avere's GNS is part of the company's A-3 Architecture, the technological foundation on which its solutions are built upon. This new storage virtualization capability joins Avere's Tiered File System (TFS) – a file system designed from the ground up to be deployed on multiple storage tiers for maximum efficiency – its linear performance scaling via scale-out clustering and visibility into data storage operations as a technology for building out the next generation of NAS systems.

An example of the Avere FXT being used in a private cloud environment is Sony Pictures Imageworks. The FXT gives Sony the ability to deploy high-performance NAS to employees located in its digital animation studio in New Mexico at a reasonable cost and with minimal administration. The FXT architectures gives employees the ability to access network files at Sony headquarters for a reduction in latency, making user access as fast as local access.
The Avere Appliances also support VMware.

Cost: The FXT Series Appliances start at $52,500.

Cloud Gateway and Enablement Vendors Tables

	#U in height	Rack-mount	Raw capacity	Logical capacity	# Ports GbE	# Ports FC	Public cloud	Private Cloud	Virtual Appliance	Cost	Cost/GB/local cache	Cost/GB/cloud storage per month	Other
Avere Systems FXT Series		X	3.6TB/ node*		2 10GbE +1Gbe or 10 1GbE		X	X		$52,500			*512GB SSD, 1GB NVRAM, 63GB DRAM, 3.6TB SAS; 25 nodes per cluster

Cloud Gateway and Enablement Vendors: NAS/SAN Capability, Communications Features

	NAS	SAN	iSCSI	CIFS	NFS v3	NFS v4	HTTP/HTTPS	Structured Data	Unstructured Data	REST	HTTP	FTP	WebDAV	Rsync	POSIX-compatible file systems
Avere Systems FXT Series	X			X	X				X						

Cloud Gateway and Enablement Vendors: Hypervisor Support, APIs deployed

	VMware ESX	VMware ESXi	VMware vSphere	Hyper-V	Xen	Citrix XenServer	Other	Amazon S3 APIs	Eucalyptus APIs	Mezeo	Nirvanix	Rackspace Cloud Files	Simple Cloud	SME Storage	SNIA CDMI	SNIA XAM	VMware vCloud API	Windows Azure REST API	Other
Avere Systems FXT Series	X	X	X				Red Hat Enterprise												

	VMware ESX	VMware ESXi	VMware vSphere	Hyper-V	Xen	Citrix XenServer	Other	Amazon S3 APIs	Eucalyptus APIs	Mezeo	Nirvanix	Rackspace Cloud Files	Simple Cloud	SME Storage	SNIA CDMI	SNIA XAM	VMware vCloud API	Windows Azure REST API	Other
							Virtualization												