Implementing, Serving, and Using Cloud Storage

November 2013

Table of Contents

Introduction 1
The Business Case for Cloud Storage 1
Applications for Cloud Storage 2
Backup 3
Archive 3
Application Data Storage 4
Other Applications for Cloud Storage 5
The Requirement for a Cloud Storage Standard 6
Cloud Storage—an Abstract Model 6
Cloud Storage—the Reality 6
The Complete Picture 8
The Cloud Data Management Interface 9
How CDMI Works 9
Deploying Cloud—Public, Private, and Hybrid 10
CDMI Advantages for Private and Hybrid Storage Clouds 11
CDMI Advantages for the Public Cloud 11
Example of Deploying a Hybrid Backup Cloud 12
How CDMI Will Roll Out 13
Conclusion 14
About the CSI 15
About the SNIA 15

List of Figures

Figure 1 – Storage Cloud and Other Cloud Services 2
Figure 2 – Data Storage as a Service 6
Figure 3 – Existing Interface Standards for Data Storage 7
Figure 4 – Cloud Storage Management 8
Figure 5 – CDMI Basic Data Flow 10
Figure 6 – Hybrid Backup Cloud 13
Figure 7 - Implementation of Cloud Architecture 13

Introduction

Organizations of all types are trying to control costs and satisfy increasing demands at the same time—demands created by explosive data growth and ever-changing regulations. To address these challenges, storage industry professionals are turning to cloud computing and cloud storage solutions.

Cloud computing is not a new technology, but rather, a new business model that encompasses a set of technologies—such as server virtualization—that reduce the cost of using information technology (IT) resources. Cloud computing takes advantage of Web-based mechanisms that allow scalable, virtualized IT resources to be provided as a service over a network. The advantages of cloud storage and other cloud services include "pay as you go" (i.e., billing only for services consumed—no fixed costs), the perception of infinite capacity (elasticity), and the simplicity of use/management.

When virtualized storage is available on demand over a network, organizations are not required to buy or provision storage capacity before storing data. As a result, organizations can save a significant amount of money on storage costs because they typically only pay for the storage that they actually use.

When SNIA recognized the significant changes in the way that organizations use storage, it developed the Cloud Data Management Interface (CDMI) standard for cloud storage vendors and others to use when implementing their own public and private clouds. As the interest in cloud storage grows, SNIA is also developing a series of education programs. These programs provide advice and recommendations for service providers when deploying clouds, and include use cases for organizations when adopting the technology.

The Business Case for Cloud Storage

Forrester research defines cloud computing as "A pool of abstracted, highly scalable, and managed compute infrastructure capable of hosting end-customer applications and billed by consumption ("Is Cloud Computing Ready for the Enterprise?" Forrester Research, Inc.).

From a computing perspective, cloud computing enables organizations with flexible IT environments to reduce capital costs by providing a 'pay-as-you-go' service. Pay-as-you-go means that the organizations incur costs for services that they use, not for what they require to sustain peak use.

From a storage perspective, cloud storage meets a variety of demands. These demands have traditionally been met in one of two ways: by investing in private disk storage—directly attached disks or network connected SAN and NAS systems—or by longer-term retention media, like tape for backup and archiving. Cloud storage promises a more efficient and flexible alternative for a variety of storage use cases—from Web-based media, such as video, audio, and electronic books—to archiving, where compliance, retention, and e-discovery are simplified. New requirements are also evolving, such as storage for cloud computing and cloud-based applications.

To meet these diverse requirements, cloud storage has developed into three major implementations: public, private, and hybrid.

- A public implementation is a secure, multi-tenant environment that is externally available to all users.
- A private implementation is a secure, single-entity environment, either inside or outside an organization's firewall.
- A hybrid implementation is a combination of public and private clouds.

As shown in Figure 1 – Storage Cloud and Other Cloud Services, the service delivery models can offer Storage-as-a-Service (SaaS), Platform-as-a-Service (PaaS), and Infrastructure-as-a-Service (IaaS) for all three implementation models.

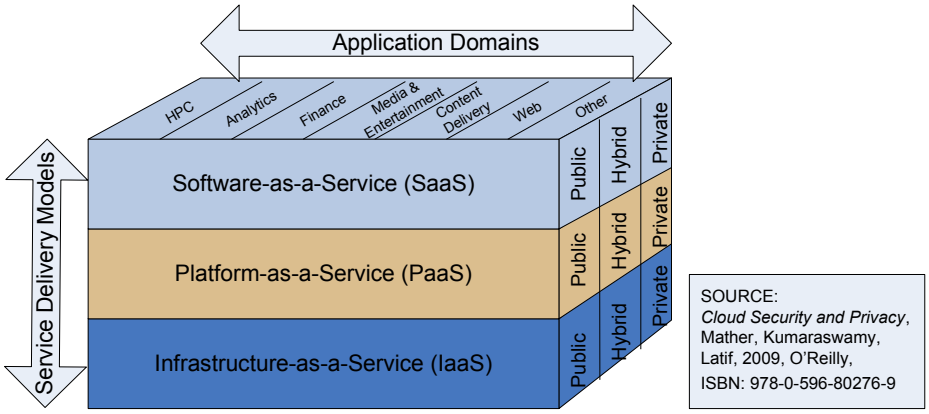

Figure 1 – Storage Cloud and Other Cloud Services

The terms *multi-tenant* or *multi-tenancy* are not new; both have been used to describe application architectures designed to support multiple users—the "tenants"—for many years. With the advent of cloud computing, this terminology has simply been extended to include any cloud architecture. Security in a multi-tenancy environment is essential, covers all aspects of the internal and external environment, and extends from the application through the server, network, and storage layers.

Equally important to the cloud storage provider is the allocation of costs. All types of clouds will require metering and billing—externally provided clouds that are dedicated to a single organization, multi-tenancy clouds, and even internally provided cloud storage (or at least, cross charging to internal business units).

The challenge for cloud providers is to show that cloud computing can meet a potential user's peak demands without expanding existing facilities and at a price that is less than or equal to the non-cloud alternative.

Applications for Cloud Storage

Before investing in cloud storage, providers and users must understand the options for cloud storage and how to create a strong business case. Some obvious and specific use cases that lend themselves to cloud storage include backup, archive, and application data storage.

Backup

In all organizations, business-critical data must be secure, available at short notice, and restorable to a specific time in the past. Traditionally, organizations use backup software and agents on file servers or desktops to back up their data to a mixture of specific disk systems and tape. Using cloud storage, organizations would use the same or similar backup software and agents, but instead, would back up their data to the cloud, which should have sufficient capacity and moderate latency to meet backup and recovery objectives.

When moving backup applications to the cloud, considerations include:

- Cost – Is cloud storage less expensive over time than existing alternatives?
- Capacity – Can the cloud handle the required daily, weekly, and monthly capacities and provide enough capacity for the extended periods that backup data is often held?
- Latency – Is latency low enough to meet backup and recovery objectives, but not so low as to make cloud storage too expensive?
- Security and privacy – Is the data secured from tampering or third-party access? Does the legal jurisdiction where the service is provided meet privacy requirements?

Cloud storage can offer many advantages, not only lower costs, but also reduced chargebacks to business units.

- With cloud storage, organizations shift the burden of meeting Service Level Agreements (SLAs) to a service provider.
- Traditional backup environments require capital investment; the largest users are often the only ones who experience the economies of scale and efficiencies that traditional solutions offer.
- With cloud storage, users avoid overbuying capacity to lessen their risk of running out of capacity, which provides a "quick fix" but at a higher price.
- A public backup cloud can help turn capital expenses into operating expenses.
- A private backup cloud has the potential to leverage a dedicated backup cloud in an off-site location, improving the security of disaster recovery.
- A hybrid backup cloud allows managing backups to local, public, and/or private clouds to meet the varying requirements of cost, availability, latency, and security.

Archive

Organizations are increasingly being forced to retain larger volumes of data for longer periods, from decades to a century or more. The traditional approach is to back up long-term retention data to an external storage media, often tape, and keep it stored off site.

Using cloud storage, backup data may be sent to a cloud archive that provides low-cost, high-capacity archive storage.

When moving long-term retention data to the cloud, considerations include:

- Cost – Is cloud storage less expensive over time than long-term media such as tape?
- Capacity and duration – Can the cloud handle the volume and proposed data retention requirements?
- Refresh – Since technology rarely lasts a century, can the data be easily retrieved and moved to another storage medium in the future?
- Security and privacy – Is the data secured from tampering or third-party access? Does the legal jurisdiction where the service is provided meet privacy requirements?

Cloud storage can offer many advantages, not only lower costs, but also reduced chargebacks to business units.

- Latency is less important than cost, given that archive data is not normally required immediately.
- Cloud storage allows users to shift the burden of meeting compliance to a service provider, whereas a traditional environment requires users to have long-term archive expertise and to comply with regulatory, compliance, and legal requirements.
- A private backup cloud has the potential to leverage a dedicated archive cloud that is off site, thereby improving security and making it unnecessary to move tapes.
- A hybrid backup cloud allows different models for different regions of the world or business units to comply with regulatory, compliance, or legal requirements.

Application Data Storage

Business-critical applications and supporting applications require temporary and permanent data storage, which is normally supplied by internal local disk, external local disk, NAS, or SAN. Apart from the challenges posed by backing up user data from a wide variety of disparate sources, application data has stricter access and latency requirements.

When moving application data to the cloud, some considerations include:

- Cost – Is cloud storage less expensive than internal local disk, external local disk, SAN or NAS?
- Latency – Does cloud storage provide the kind of latency that applications and users expect?
- Accessibility – Is the data available from multiple locations?

- Security and privacy – Is the data secured from tampering or third-party access? Does the legal jurisdiction where the service is provided meet privacy requirements?

Cloud storage can offer many advantages, including lower costs that result in reduced chargebacks to business units. Additional advantages include:

- Backup is more easily resolved when the primary application data is already in the cloud.
- Users shift the burden of meeting SLAs to a service provider.
- A private cloud has the potential to significantly reduce data and backup management costs.
- A hybrid backup cloud allows data to be close to the application where latency is critical.
- Cloud solutions can provide superior geographic latency, data protection, and recovery levels for certain distributed applications.

Other Applications for Cloud Storage

The examples above show some common themes when evaluating a cloud solution, but there are also many other applications for cloud storage. When assessing the viability of a cloud solution, consider the following business uses that also apply:

- Utility model – Does the cloud provider recognize the key role of storage in cloud provisioning?
- Non-intrusive/non-disruptive change – Is it easy and cost effective to change existing infrastructures to cloud storage, and does the cloud provider offer vendor-neutral support for various platforms?
- Rapid, flexible provisioning – Can the cloud solution quickly and non-disruptively expand, shrink, provision, and de-provision storage on demand?
- Universal access – Can the cloud solution store any type of application data and provide access through any standard network access protocols?
- Autonomic – Does the cloud solution provide always-on operation, zero- to low-touch administration, fully automated management, and data mobility?
- Secure and protected data – Does the cloud solution provide controlled access to data with adequate privacy and security? Is the data protection policy widely enforced?
- Self-service – Does the cloud solution provide on-demand provisioning of capacity with pre-assigned service levels and application- and OS-consistent restores?
- Measured service – Can the cloud solution meter SLAs and provide capacity reports? Is chargeback and billing available and in understandable units for managing user costs?

In summary, cloud storage will prove its worth if it supports business as usual, but better.

The Requirement for a Cloud Storage Standard

Because it is relatively easy for cloud storage solutions to improve business processes, interest in and adoption of cloud storage solutions are growing. Until now, there have been few standards that simplify and allow interoperability across disparate cloud solutions—standards that organizations need if they want to enjoy the benefits of an open, competitive marketplace in cloud storage.

Because the variety of use cases rarely share a common interface to cloud storage, SNIA formed a technical working group (TWG) with over 75 members to develop a standard for cloud storage. In June 2009, it published a use cases and reference model, and in July 2010, the TWG published the first draft standard, the Cloud Data Management Interface (CDMI), with the intention of gaining future ISO and ANSI certification. Currently, the TWG is working on a cross-platform reference architecture to encourage adoption of this new standard.

Cloud Storage—an Abstract Model

The simplest way to describe cloud storage is this: A cloud represents a "fuzzy" container for data, and the user doesn't really care how the cloud provider implements, operates, or manages the cloud. A client, through the medium of a network, makes requests to the cloud storage to securely store and subsequently retrieve data at an agreed level of service.

Although seemingly abstract and complex, cloud storage is actually rather simple. Regardless of data type, cloud storage is a pool of resources that are provided in small increments with the appearance of infinite capacity. In other words, cloud storage is virtualized storage on demand and is more formally called 'Data Storage as a Service' (DaaS). DaaS is defined as "Delivery over a network of appropriately configured virtual storage and related data services, based on a request for a given service level." (See Figure 2.)

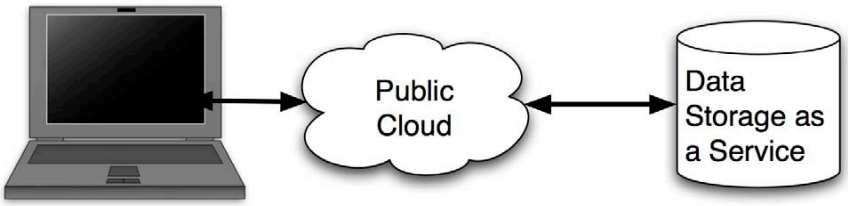

Figure 2 – Data Storage as a Service

Cloud Storage—the Reality

Today's IT environment is composed of various products that are intended to store, protect, secure, and make available the information used by businesses and business processes. These products encompass elements used in both the data path and control path between the user and the eventual location of that information.

As vendors and suppliers of cloud services have delivered early implementations to users, they have tended to supply a multitude of interfaces that have been re-purposed for DaaS, such as block-based access via iSCSI; POSIX interfaces (NFS, CIFS, and WebDAV); object-based CRUD (Create, Read, Update, Delete) interfaces over HTTP; and a plethora of other proprietary interfaces for database or table access (see Figure 3 – Existing Interface Standards for Data Storage).

Compared to the simplicity of the abstract cloud model, the existing cloud storage model is rather complex because there are so many interfaces that are required to meet the different demands of end users for accessing storage.

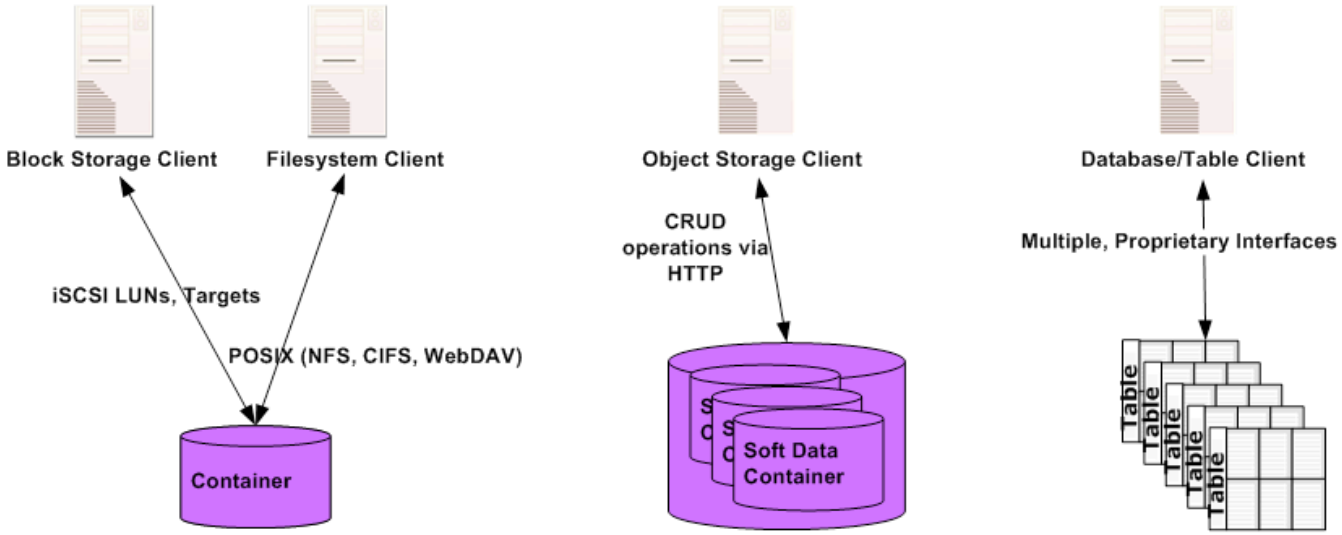

Figure 3 – Existing Interface Standards for Data Storage

Standards exist and are emerging for interoperability between these elements; however, what is missing is a comprehensive description of where interoperability is needed and where standards can be best applied. All of these interfaces use parts of SNIA's Resource Domain model, which sets out these elements and describes a logical view of their functions and capabilities using a descriptive taxonomy. The purpose of this model is to retain the simplicity of the abstract cloud model by using metadata to drive the underlying services, so that users do not have to manage the service themselves.

With a model in place, it's possible to identify existing standards, where appropriate, and identify areas where new standards might be needed. Although a variety of vendors offer "open" licenses for cloud storage interfaces and sets of preexisting libraries that provide similar functionality, no vendor wants a competitor to control the specification of the interface. In addition, multiple "standards" will proliferate, locking users into proprietary architectures. The SNIA's response has been to develop, using this Resource Domain model, the CDMI, an extensible standard that accommodates vendors' requirements and ensures consistency and interoperability for users.

The Complete Picture

Clients can adapt their existing, proprietary interface, while still allowing access via CDMI or standards-based interfaces, such as POSIX, XAM and so on. They can adapt their interface by:

- Abstracting the data and storage services into data storage cloud containers
- Providing a consistent and uniform interface to gain access through CDMI

The CDMI defines the functional interface that applications can use to create, retrieve, update, and delete data elements from the cloud. As part of this interface, the client can 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 (see Figure 4 – Cloud Storage Management).

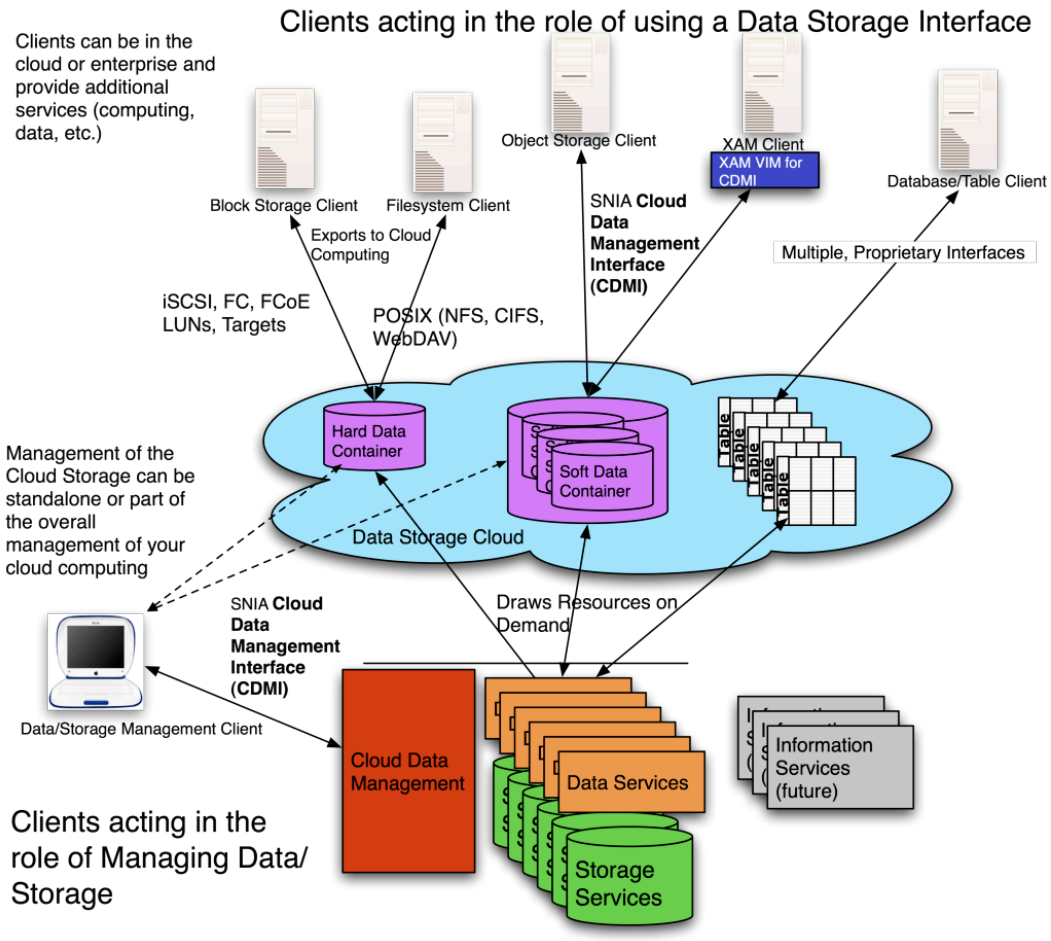

Figure 4 – Cloud Storage Management

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 full capabilities of the storage being used.

The CDMI applies to three types of storage:

- Cloud storage for cloud computing as a management interface for the storage lifecycle
- Public cloud storage, both for management and data access
- Private and hybrid cloud storage, where it can act as an API for vendor solutions

The Cloud Data Management Interface

Designed to enable interoperable cloud storage and data management, the CDMI specification is aggressively addressing a total cloud storage solution. This solution helps users avoid the chaos of proprietary advances and partial-solution APIs that would erode the integrity of the cloud model.

Easy to implement, CDMI integrates and is interoperable with various types of client applications and is designed to be compatible with current public cloud storage offerings. CDMI offers standard approaches to data portability, compliance, and security. It also offers the ability to connect one cloud provider to another, enabling compatibility among cloud vendors.

How CDMI Works

Providing both a data path to the cloud service and a management path for the cloud data, CDMI is the functional interface that applications use to create, retrieve, update, and delete data elements in 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.

The semantics of CDMI are straightforward; simple containers and data objects are tagged with metadata—some of which are metadata that describe the data requirements of the object or container. The protocol for accessing the data and metadata is RESTful HTTP, first outlined by Roy Fielding[1] (see Figure 5 – CDMI Basic Data Flow).

[1] http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm Architectural Styles and the Design of Network-based Software Architectures, Roy Thomas Fielding, 2000

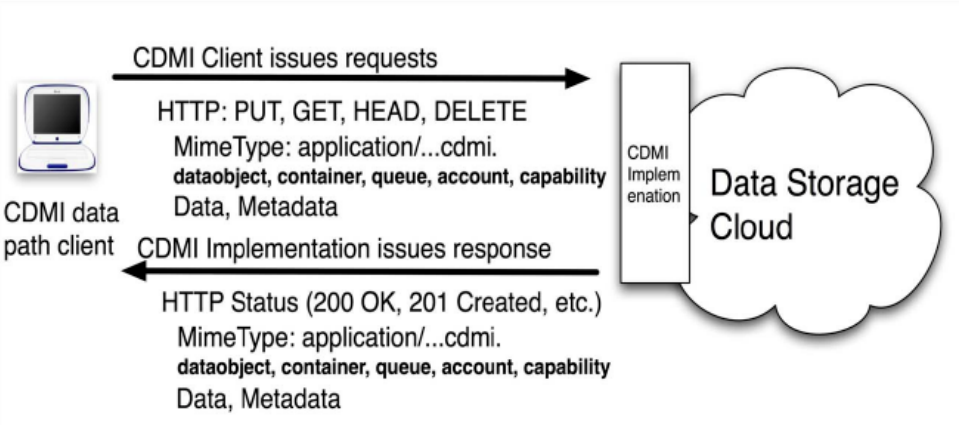

Figure 5 – CDMI Basic Data Flow

Using RESTful techniques allows CDMI to provide:

- Unique names – Every object is addressable by a unique identifier.
- Uniformity – The interface uses only HTTP verbs with other semantics carried in the data model.
- Simplicity – The complexity is encapsulated in the representations.
- Statelessness – A lack of persistent client-side connections simplifies implementation.

The format of the representations (the data carried by the HTTP requests) is in JavaScript Object Notation (JSON) that allows great flexibility, readability by both humans and machines, and extensibility.

This flexibility gives CDMI the following advantages as an interface specification:

- First, it is well supported by many infrastructures and programming languages.
- Second, a small learning curve should encourage adoption and an "ecosystem" of support and code around the API.

Vendors and users will find these reasons compelling when developing a cloud using this new, open interface.

Deploying Cloud—Public, Private, and Hybrid

Public storage clouds increase the efficiency of offering storage capacity through the use of multi-tenancy solutions, meaning that the same, shared storage infrastructure can service multiple users.

Private storage clouds typically exist behind an organization's firewall and are deployed for internal users. They can be located in an enterprise data center or hosted at a collocation facility that is

possibly owned by a third party, like a service provider. Private clouds are designed to take advantage of the flexibility and management simplicity of the cloud model, which means that when cloud storage is implemented, users can set up and manage their own environment.

Hybrid storage clouds are a combination of both public and private storage clouds.

CDMI Advantages for Private and Hybrid Storage Clouds

CDMI considerably benefits the long-term needs of ongoing data management, since many users may be concerned about the potential of being locked into a particular cloud interface. By offering the promise of portability by standardizing a cloud data import and export format, CDMI allows users to take advantage of private and hybrid cloud offerings without depending on a particular platform in the future.

CDMI offers a standard format that can be used to move data and its associated metadata between clouds. Using this standard means that data can be moved (even out of band via a disk drive) from one cloud to another while maintaining the data requirements and ensuring that the new cloud meets these requirements.

Those interested in private and hybrid cloud models often need flexibility, since some users may initially test cloud technologies in a private implementation and then want to incorporate a hybrid storage cloud at a later date. CDMI also provides the ability to move data between different cloud implementations as they become better able to meet the data requirements.

In a hybrid cloud, CDMI helps further mitigate the security risks of using a public cloud. Using metadata to manage the security of data within containers, CDMI allows client applications to use data services such as encryption. Using data services enables administrative and management applications to control data access and management within the systems, including managing accounts and ensuring security and programmatic monitoring/billing information on usage, even for storage that is accessible by other data path protocols.

CDMI allows access control of legacy interfaces that are used when containers are exported via those other protocols. It also provides for a secure delete mechanism (with various algorithms) for data that has expired.

CDMI Advantages for the Public Cloud

Security is one of the key objections to implementing a public cloud. CDMI helps to mitigate the security risks by using metadata to manage the data within containers. Using metadata exposes storage and data services to client applications, which allows the client applications to use their administrative and management applications to control data movement within and between systems—data such as information to manage accounts, security access, and monitoring/billing, even for storage that other protocols can access.

For example, in public cloud storage environments, accounts are often created programmatically and account user membership and credentials are often populated using similar interfaces. By providing

access to user membership, CDMI enables self-enrollment, automatic provisioning, and other advanced self-service capabilities, either directly through the CDMI standard or through software systems that interface using CDMI. The account membership capability provides information and allows the specification of users and groups of users that are allowed to access the account via CDMI and other access protocols.

Authorization and authentication in the data path of CDMI is done using the same mechanism as in NFS—a secure standard in wide use where the threat models are well known. CDMI also allows control over the access control of legacy interfaces that are used when containers are exported via those protocols. CDMI has data system metadata that allows the data to be encrypted when stored on the underlying infrastructure—an important capability in multi-tenant situations, such as public clouds.

CDMI also benefits the long-term needs of ongoing data management, since many users may be concerned about the potential of being locked into a particular public cloud implementation. By offering portability by standardizing cloud management and interfaces, CDMI allows users to take advantage of public cloud offerings, without depending on a particular platform. CDMI also offers a standard format that can be used to move data and its associated metadata within a cloud or between clouds.

Example of Deploying a Hybrid Backup Cloud

The steps to deploying a cloud are as follows:

1. Determine your storage cloud use case.
2. Diagnose the potential bottlenecks for your deployment scenario.
3. Define public, private, or hybrid deployment.
4. Document concerns with cloud deployment style.
5. Design the deployment.

In the example shown in Figure 6 – Hybrid Backup Cloud, the storage cloud is being used as a complementary backup environment. The backup is possibly for less important data or data that isn't required immediately. The backup may also be an overflow buffer for projects that require backup services that can't be accommodated by the in-house backup environment without considerable incremental cost.

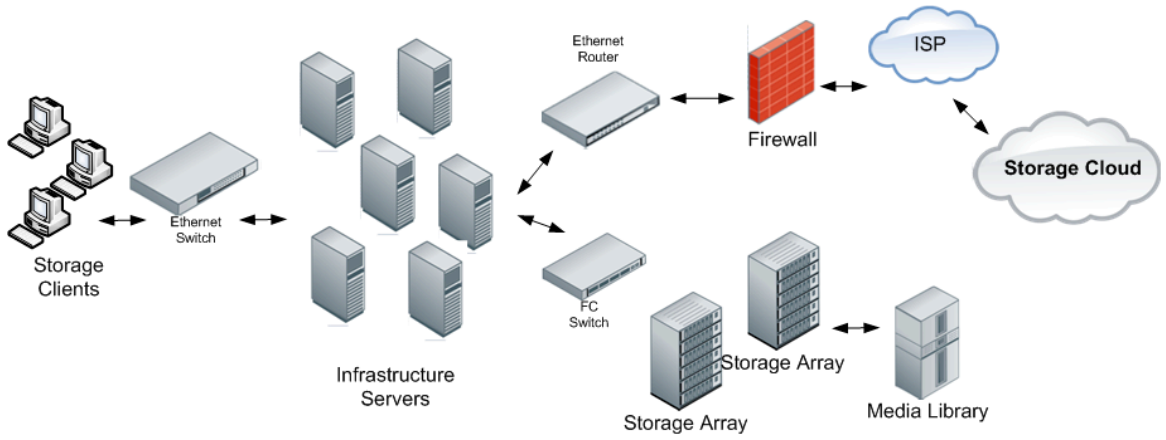

Figure 6 – Hybrid Backup Cloud

As is the case when re-architecting an IT environment, bottlenecks become important considerations. In the example, we might ask, do we have enough bandwidth to support the cloud backup streams? What kind of Recovery Time Objective (RTO), and hence, recovery bandwidth, will be required should a restore become imperative?

From a design perspective, SNIA provides a CDMI Reference Implementation Developer's Guide that can serve as a template for implementing cloud architectures (see Figure 7 - Implementation of Cloud Architecture). Because much of the functionality of CDMI is exposed through simple RESTful interfaces with easily parsed JSON representation for the data payload, most developments will use readily and widely available components to construct rapidly developed, deployed, and flexible solutions.

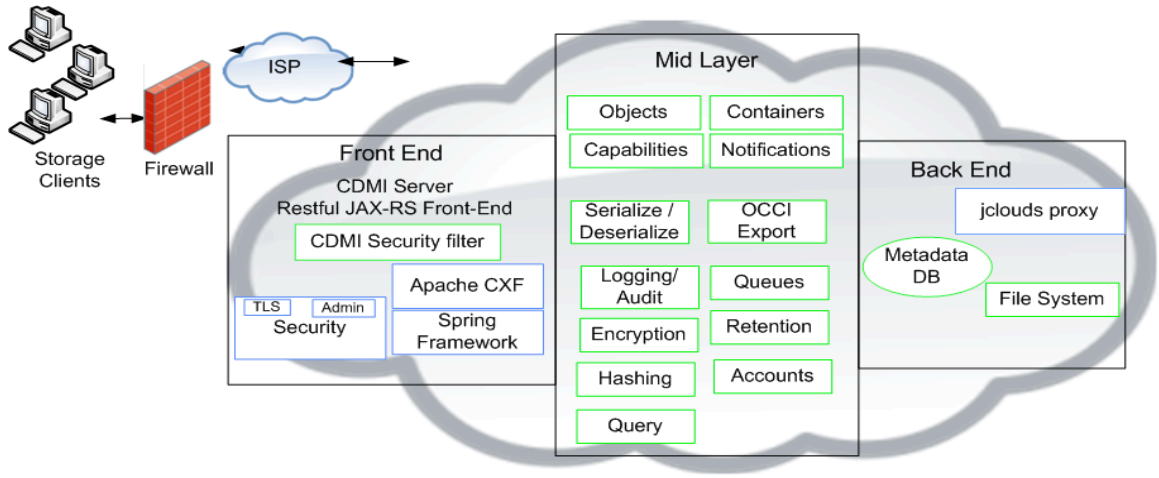

Figure 7 - Implementation of Cloud Architecture

How CDMI Will Roll Out

Since CDMI can be used as both a data path and a management path, there are several ways to deploy an implementation of CDMI for a private, public, or hybrid cloud. First of all, CDMI can be deployed

side by side with existing proprietary interfaces. This deployment allows existing applications to move over to the standard interface when they want to take advantage of the features. The side-by-side deployment allows the same data to be accessed via either interface, and no movement of data is required. As the cloud provider adds additional data services and capabilities to their service, the application can use the CDMI interface to ensure that the existing data requirements are being met using those new services. If the cloud provider implements the CDMI accounting, users can use that function to administer the security and to programmatically access their bill.

CDMI has many capabilities and not all will be implemented by every cloud offering. CDMI has capability resources that let a client application programmatically find out which capabilities are actually implemented before trying to use them. This feature also allows new cloud offerings to use CDMI as the initial interface for their service, expanding the implementation of CDMI as the offering increases capabilities. CDMI is also extensible to accommodate services and features that are not yet standardized, which means that those functions will not require separate, proprietary interfaces.

Conclusion

There are many advantages to using public, private, and hybrid clouds as part of an organization's long-term storage strategy. In private/hybrid models, internal IT departments have more control of their data (vs. public clouds) without needing to actively manage it, resulting in significantly lower costs than traditional storage. Public clouds offer the availability of storage capacity on a pay-as-you-go basis, combined with scalability and ease of use and can offer much needed cost savings.

CDMI, by enabling interoperable cloud storage and data management, provides a new paradigm of managing all cloud environments, while maintaining the simplicity that makes the cloud approach attractive. With its total cloud storage solution, CDMI is helping users avoid concerns around proprietary cloud advances and partial-solution APIs, while helping to maintain the integrity of the cloud model.

About the CSI

The SNIA Cloud Storage Initiative (CSI) was created to foster the growth and success of the market for cloud storage. Members of the SNIA CSI work together to educate the vendor and user communities about cloud storage, perform market outreach that highlights the virtues of cloud storage, collaborate with other industry associations on cloud storage technical work, and coordinate with SNIA Regional Affiliates to ensure that the results of CSI activities are felt worldwide. The CSI, along with 140 individuals from more than 30 organizations, promotes the adoption of standardization through the Cloud Data Management Interface (CDMI) standard specification. For more information or to get involved, visit the SNIA CSI website at http://www.snia.org/cloud.

About the SNIA

The Storage Networking Industry Association (SNIA) is a not-for-profit global organization made up of some 400 member companies spanning virtually the entire storage industry. SNIA's mission is to lead the storage industry worldwide in developing and promoting standards, technologies, and educational services to empower organizations in the management of information. To this end, the SNIA is uniquely committed to delivering standards, education, and services that will propel open storage networking solutions into the broader market. For additional information, visit the SNIA web site at www.snia.org.