SNIA
Advancing storage & information technology

Managing Private and Hybrid Clouds for Data Storage

January 2010

Table of Contents

List of Figures

Introduction

Many organizations, driven by the opportunities for significant cost-savings, are considering cloud computing and cloud storage solutions, which take advantage of Web-based technologies to allow scalable, virtualized IT resources to be provided as a service over the network. Not a new technology in itself, cloud computing is a new business model wrapped around existing technologies, such as server virtualization, to make the use of information technology resources more efficient.

The advantages of cloud storage are the same as those that define other cloud services: a high degree of scalability and elasticity, along with simplified management. When virtualized storage is available on demand over a network, an organization can be freed from the need to purchase – or often even provision – its storage capacity before proceeding with data storage.

Cloud storage can be implemented in many different ways. For example: local data (such as on a laptop) can be backed up to cloud storage; a virtual disk can be "synched" to the cloud and distributed to other computers; and the cloud can be used as an archive to retain (under policy) data for regulatory or other purposes. When considering cloud options for data storage, planning for the current and future management of both the clouds and the data is critical, as well as the integration of the cloud service usage with internal IT.

There are three major deployment models of cloud storage to be considered: public, private, and hybrid models (many users implement a combination of these). While the Storage Networking Industry Association™ (SNIA) is addressing all three, this paper discusses private and hybrid cloud storage solutions and the emerging methods available to best manage these environments.

What is a Private Storage Cloud?

For many, the term "cloud" brings to mind public cloud solutions in which an organization accesses third-party resources (like Amazon S3™, Iron Mountain®, Google™, etc.) on an as-needed basis, without the requirement to invest in additional internal infrastructure. However, the cloud model can also be implemented for existing legacy, already "owned" storage in what is termed a "private storage cloud."

Storage clouds increase the efficiency of offering storage capacity through the use of multi-tenancy solutions – meaning multiple customers are serviced at once from the same shared storage infrastructure.

Private storage clouds will typically exist behind an organization's firewall and are deployed for internal customers. Private storage clouds can be located in an enterprise data center, but can also be hosted at a collocation facility possibly owned by a third party like a service provider. Designed to take advantage of the elasticity and management simplicity of the cloud model, a private storage cloud implementation allows the customer to set up and manage their own environment.

A key component of public clouds, which will also see use in private enterprise clouds, is the ability to meter usage and bill appropriately for usage of the storage cloud. In an enterprise environment, the IT department, outsourced or not, becomes the cloud provider for the other departments of the organization. Part of the IT department's budget is then derived from this billing of the other departments. What this does is highlight the total cost of providing that storage and bundle it into the bill they give their internal customers. The benefit to the organization overall is that there is now a single set of metrics (the billable rates) that the IT organization can utilize to reduce costs and achieve efficiencies.

Of course, some vendors may try to sell their storage as "cloud capable" or "cloud ready," but without the metering and billing capability the IT organization is forced to put those pieces together itself. What is needed is an interface into those capabilities for each storage component, allowing the IT organization to operate its infrastructure as a cloud without having to roll its own instrumentation of that infrastructure.

The SNIA's Cloud Data Management Interface (CDMI) standard provides just such an interface for storage vendors to implement the required metering of storage and data service usage, as well as the interface to feed the billing applications that IT organizations will be putting into place. By using a standard interface, the IT development group does not have to write adapters to each different vendor's proprietary interface.

What is a Hybrid Storage Cloud?

As the name implies, a hybrid storage cloud uses a combination of both public and private storage clouds. For example, an organization may elect to store sensitive data in the private storage cloud, while utilizing the elastic capacity and cost-effective public storage cloud for the storage of less sensitive data. Another example is in conjunction with hybrid computing clouds, the data from an internal storage cloud can be made available (replicated) to a public storage cloud for cloud computing tasks that need to access the same data that is available internally (when elastic demand for computing exceeds the internal resources). Lastly, a combination of public and private clouds can be used to implement storage tiering, with lower cost tiers being implemented by a public cloud provider for data that no longer needs the latency of local access.

Hybrid implementations are often useful for archiving and backup functions, allowing local data to be replicated to the public cloud, thus lowering storage costs. Though hybrid clouds require more IT management than pure-play public clouds, many vendors offer rules-based solutions that enable simplified management across these resources, allowing organizations to develop a customized solution that balances cost benefits with security and other key concerns.

A hybrid cloud points out the need for a storage standard that provides common interface for both the private as well as the public components in this "federation" of clouds. The standard must also provide for the portable movement and migration of data and its associated requirements back and forth between the clouds. CDMI provides these capabilities, as well as optional security measures that include transport security, user and entity authentication, authorization and access controls, data integrity, data and media sanitization, data retention, protections against malware, data at-rest encryption, and security capability queries.

Managing Data Storage as a Service

As with public clouds, concerns around portability and vendor lock-in continue to be a key consideration when working with any cloud offerings.

These concerns are addressed by managing all cloud storage using a Data Storage as a Service (DaaS) approach. DaaS is defined by SNIA (http://www.snia.org/education/dictionary/) as the delivery of virtualized storage on demand over a network appropriately configured (with virtual storage and related data services), based on a request for a given service level.

By abstracting data storage behind a set of service interfaces and delivering it on demand, a wide range of actual offerings and implementations are possible. The SNIA has developed an interface that offers DaaS while leveraging the SNIA Storage Industry Resource Domain Model to minimize complexity and ensure that cloud storage remains user-friendly. This model, which allows both legacy and new applications to be used with private and hybrid cloud implementations, is shown in Figure 1.

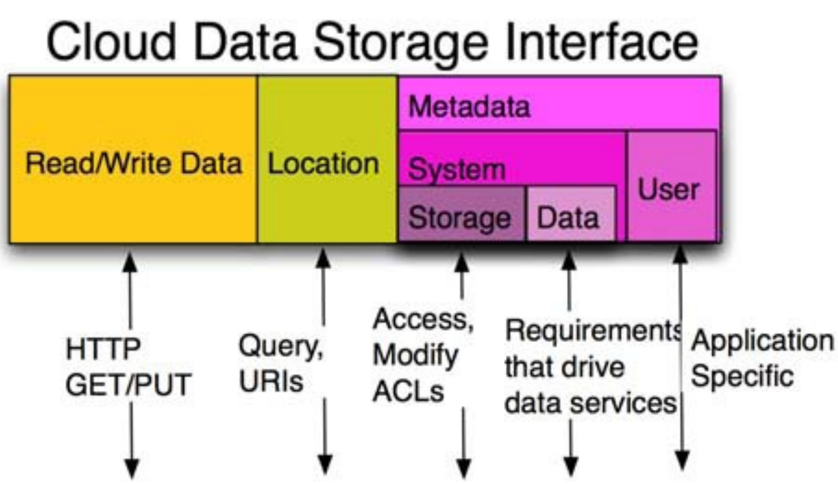

Figure 1: Using the SNIA Resource Domain Model

SNIA

Improving Private and Hybrid Cloud Storage – CDMI

Building on this use of the SNIA Storage Industry Resource Domain Model, the management of private and hybrid clouds (as well as public clouds) is addressed by the SNIA in CDMI. Designed to enable interoperable cloud storage and data management, the CDMI specification is aggressively addressing a total cloud storage solution – helping users avoid the chaos of proprietary interfaces and partial solution APIs that would erode the benefits of the cloud model.

CDMI offers standard approaches to data portability, compliance and security, as well as the ability to connect one cloud provider to another, enabling compatibility among cloud vendors – particularly critical in a hybrid cloud implementation. Easy to implement, CDMI integrates and is interoperable with various types of client applications and is designed to be compatible with current cloud storage offerings.

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 will use to Create, Retrieve, Update and Delete (CRUD semantics) 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.

Figure 2 shows multiple types of cloud data storage interfaces that are able to support current, legacy and new applications. All of the interfaces allow storage to be provided on demand, drawn from a pool of capacity provided by storage services. The data services can be applied to individual data elements as determined by the data system metadata. Metadata specifies the data requirements on the basis of individual data elements or on groups of data elements (containers).

SNIA

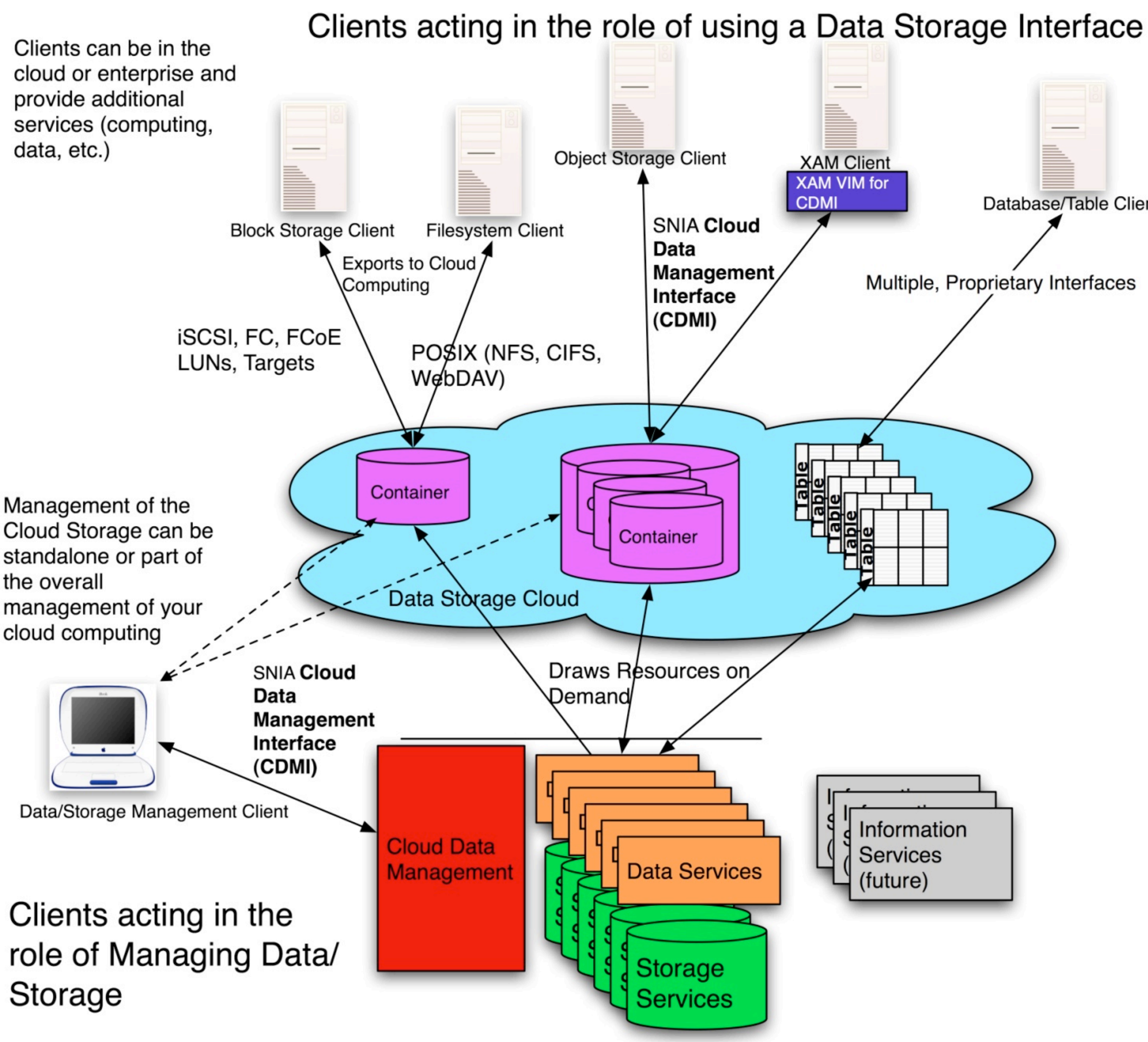

Figure 2: Cloud Storage Management

When used as the functional data path interface for data storage, CDMI accesses every data object through a separate URI (Uniform Resource Identifier). Since objects can be fetched using the standard HTTP protocol employing RESTful (Representational State Transfer) operations, each data element can be managed as a separate resource.

CDMI provides not only a data object interface, it also can be used to manage containers exported for use by cloud computing infrastructures (as shown in Figure 2). The notion of a container is used in CDMI as an abstraction of the underlying storage space in a cloud. This is not only a useful abstraction to represent storage space, but a container also serves to represent a grouping of the data stored in it, and a point of control for applying data services in the aggregate.

An important part of any DaaS offering is the support of legacy clients, often used in private and hybrid cloud models. This is accommodated with existing standard protocols such as iSCSI for block and CIFS/NFS or WebDAV for file network storage as shown below:

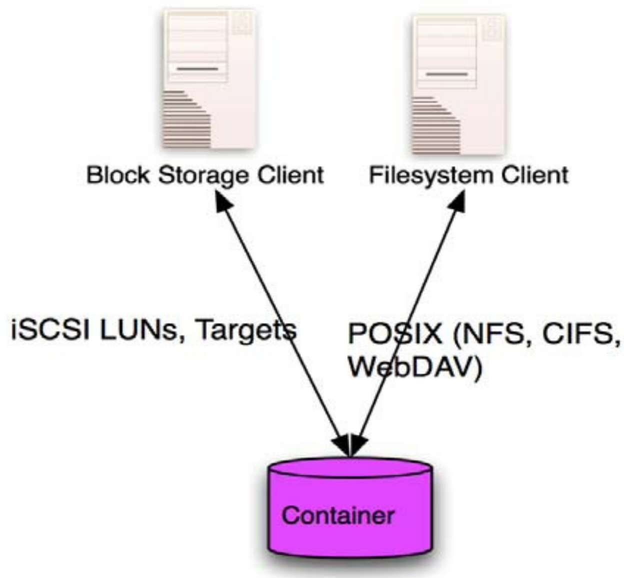

Figure 3: Existing Data Storage Interface Standards

In the case of block storage, a LUN or virtual volume is the granularity of allocation. For file protocols, a filesystem is the unit of granularity. In either case, the actual storage space can be thin provisioned and billed for based on actual usage. Data services such as compression and de-duplication can be used to further reduce the actual space consumed.

Management of legacy interfaces on containers in CDMI is handled through an "export" function that enable various protocols to access a CDMI container. The access controls for these protocols are part of the CDMI interface to ensure secure, protected access. Typically, the container is thin provisioned at an "advertised" size the applications see through these protocols (i.e., LUN size). This size can be arbitrarily large in order to not run out of space during normal operations. CDMI accounting shows the actual usage within this size that is part of the actual bill.

Metadata in CDMI

CDMI uses many different types of metadata, including HTTP metadata, data system metadata, user metadata, and storage system metadata. To address the requirements of enterprise applications and the data managed by them, this use of metadata allows CDMI to deliver simplicity through a standard interface. CDMI leverages previous SNIA standards such as the eXtensible Access Method (XAM) for metadata on each data element. In particular, XAM has metadata that drives retention data services useful in compliance and eDiscovery.

CDMI's use of metadata extends from individual data elements and can apply to containers of data, as well. Thus, any data placed into a container essentially inherits the data system metadata of the container into which it was placed. When creating a new container within an existing container, the new container would similarly inherit the metadata settings of its parent container. Of course, the data system metadata can be overridden at the container or individual data element level, as desired.

The extension of metadata to managing containers, not just data, enables a reduction in the number of paradigms for managing the components of storage – a significant cost savings. By supporting metadata in a cloud storage interface standard and prescribing how the storage and data system metadata is interpreted to meet the requirements of the data, the simplicity required by the cloud storage paradigm is maintained, while still addressing the requirements of enterprise applications and their data.

The model behind the Cloud Data Management Interface is shown in Figure 4: CDMI Interface Model:

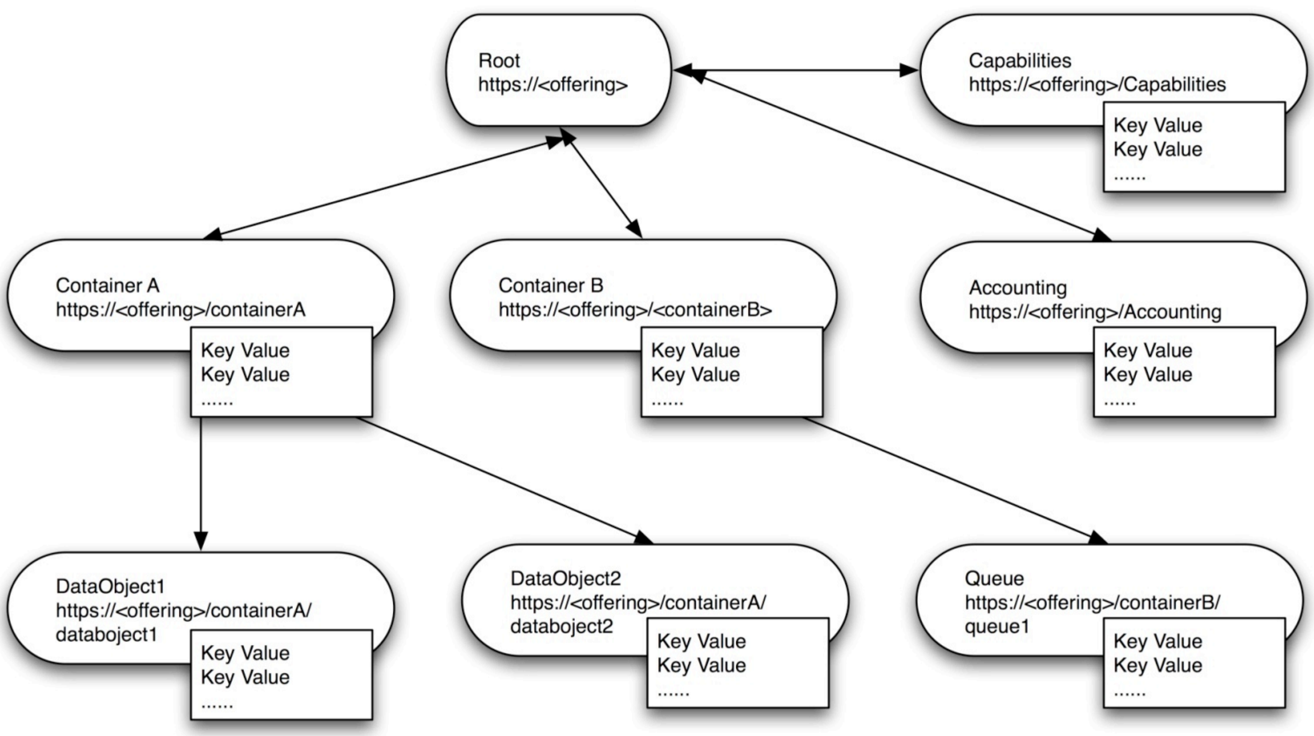

Figure 4: CDMI Interface Model

For data storage operations, the client of the interface only needs to know about containers and data objects. All implementations are required to support at least one level of containers—a sort of grouping of data objects. As shown in Figure 5, the client may do a PUT to the container URI and create a new container with the specified name. The KEY/VALUE metadata is optional. Once a container is created, a client may do a PUT to create a data object URI. A subsequent GET will fetch the actual data object. The only metadata KEY/VALUE required on the data object PUT is content type (MIME). Other KEY/VALUE pairs can be used to specify the data requirements at the data object or container level. This metadata is defined in the CDMI specification.

CDMI Advantages for Private and Hybrid Storage Clouds

CDMI is highly beneficial when considering ongoing data management needs long-term, since the potential of getting locked into a particular cloud interface is a critical concern for many. By offering the promise of portability through the standardization of a cloud data import and export format, CDMI can allow users to take advantage of private and hybrid cloud offerings without future dependence on a particular platform.

CDMI offers a standard format which can be used to move data and its associated metadata between clouds. This 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 they are met by the new cloud.

Those interested in private and hybrid cloud models are often particularly in need of flexibility, since some users may initially test cloud technologies in a private implementation, and then look to incorporate hybrid approaches at a later date. CDMI also provides the ability to move data between the 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 the public cloud usage. Using metadata to manage the security of data within containers, CDMI allows data services such as encryption to be utilized by client applications. This enables administrative and management applications to control data access and management within the systems, including managing accounts, ensuring security and programmatic monitoring/billing information on usage, even for storage that is accessible by other data path protocols.

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

How to Roll Out CDMI

Since CDMI can be used as both a Data Path and a Management Path, there are several ways to roll out an implementation of CDMI for private and hybrid clouds. First, CDMI can be deployed side by side with existing proprietary interfaces. This allows existing applications to move over to the standard interface as they take advantage of the new features. The side-by-side deployment allows the 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 CDMI interface can be used by the application to ensure that the existing data's requirements are being met using those new services.

Because the standard has many features and not all will be implemented by every cloud offering, CDMI has capability resources that allow a client application to programmatically discover which capabilities are actually implemented before trying to use them. This also allows new cloud offerings to use CDMI as the initial interface for their service, expanding the implementation of CDMI as their offering increases in capabilities.

CDMI is extensible to accommodate services and features that are not yet standardized, obviating the need for separate proprietary interfaces for those functions.

Conclusion

There are many advantages to using private and hybrid clouds as part of an organization's long-term storage strategy. In these 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.

CDMI, by enabling interoperable cloud storage and data management, provides a new paradigm of management for private and hybrid 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, helping 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 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.