SNIA

Advancing storage & information technology

Managing Data Storage in the Public Cloud

October 2009

Table of Contents

List of Figures

© 2009 STORAGE NETWORKING INDUSTRY ASSOCIATION

Introduction

With organizations struggling to control costs, they face increasing demands to keep pace with explosive data growth and ever-changing regulations. To address these challenges, storage industry professionals are increasingly turning to cloud computing and cloud storage solutions.

Cloud computing is not a new technology in itself – it is a new business model wrapped around technologies, such as server virtualization, to reduce the cost of using information technology resources. Cloud computing takes advantage of Web-based technologies to allow scalable, virtualized IT resources to be provided as a service over the network.

This approach offers great potential for the storage and management of data. The advantages of cloud storage are the same as those that define other cloud services: pay as you go, the perception of infinite capacity (elasticity), and the simplicity of use/management.

When virtualized storage is available on demand over a network, an organization is freed from the need to purchase – or often even provision – its storage capacity before proceeding with data storage. The user only pays for the amount of storage actually consumed, offering significant cost-savings.

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 evaluating cloud implementations, there are three types 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 types, this paper discusses public cloud storage solutions and the emerging methods available to best manage these environments.

What is a Public Cloud?

Public cloud solutions are the most well-known examples of cloud storage. In a public cloud implementation, 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.

In this pay-per-use model, public cloud vendors provide applications, compute platforms and storage to the general public, delivering significant economies of scale. For storage, the difference between the purchase of a dedicated local appliance and the use of a public cloud is not the functional interface, but merely the fact that the storage is delivered on demand. The customer pays for either what they

© 2009 STORAGE NETWORKING INDUSTRY ASSOCIATION

actually use or in other cases, what they have allocated for use. As an extension of the financial benefits, public clouds offer a scalability that is often beyond what a user would be able to afford otherwise.

Publicly accessible clouds offer storage capacity through the use of multi-tenancy solutions, meaning multiple customers are serviced at once from the same infrastructure. This results some common concerns when evaluating public cloud solutions, including security and privacy, as well as the possibilities of latency and compliance issues.

When considering the use of public cloud options for data storage, attention should be given to the management – both now and in the future – of both the clouds and the data, as well as the integration of the Cloud service usage with internal IT.

Managing Data Storage as a Service

As an emerging technology, many early providers of public cloud storage offerings have focused on a "best effort" quality of storage service, with minimal additional data services offered. In addition to the concerns around public clouds noted above, there has also been trepidation about portability and vendor lock-in when working with cloud offerings.

These concerns are addressed by managing 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.

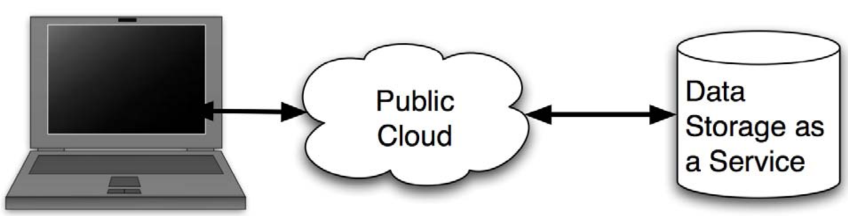

Figure 1: DaaS

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

© 2009 STORAGE NETWORKING INDUSTRY ASSOCIATION

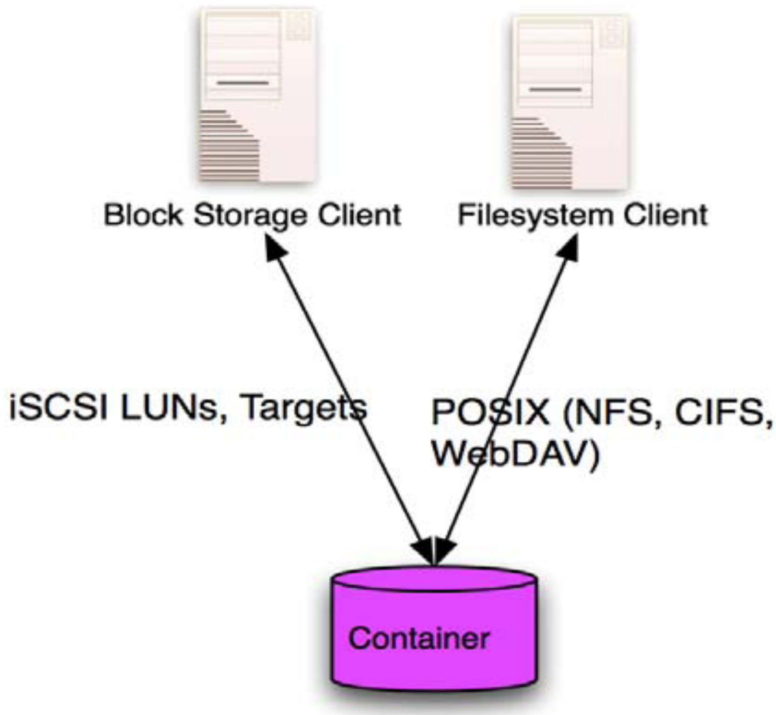

Figure 2: Existing Data Storage Interface Standards

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.

The SNIA has developed an interface that offers DaaS while leveraging the SNIA 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 public cloud implementations, is shown in Figure 3.

© 2009 STORAGE NETWORKING INDUSTRY ASSOCIATION

SNIA

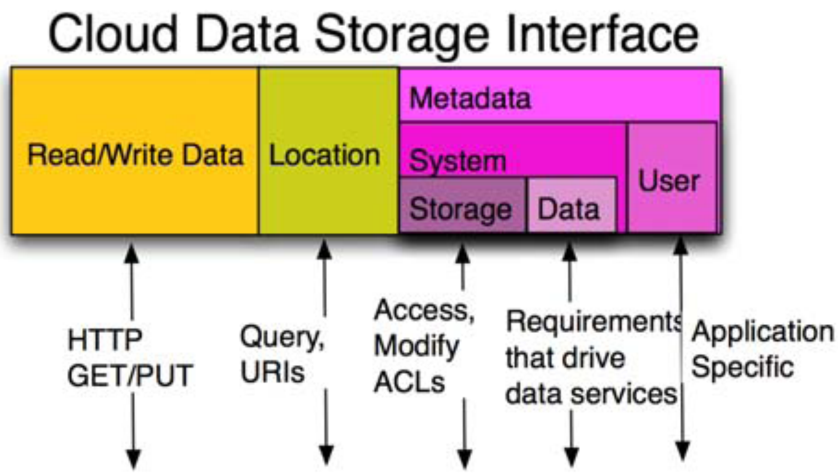

Figure 3: Using the SNIA Resource Domain Model

Improving Public Cloud Storage – CDMI

Building on this use of the SNIA Resource Domain Model, the management of public clouds (as well as private and hybrid clouds) is addressed by the SNIA in its new Cloud Data Management Interface (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 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 (like Amazon™, Iron Mountain®, Nirvanix®, etc.). 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.

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 4 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 are 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).

© 2009 STORAGE NETWORKING INDUSTRY ASSOCIATION
SNIA

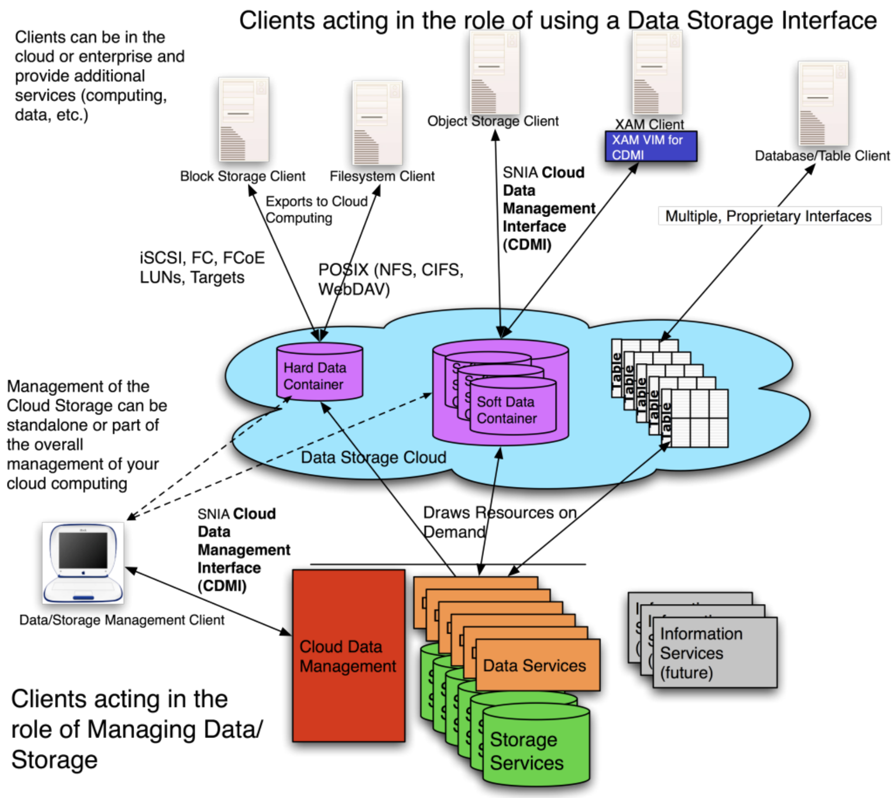

Figure 4: Cloud Storage Management

For 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 4. 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.

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.

© 2009 STORAGE NETWORKING INDUSTRY ASSOCIATION
SNIA

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 proscribing 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 5: CDMI Interface Model:

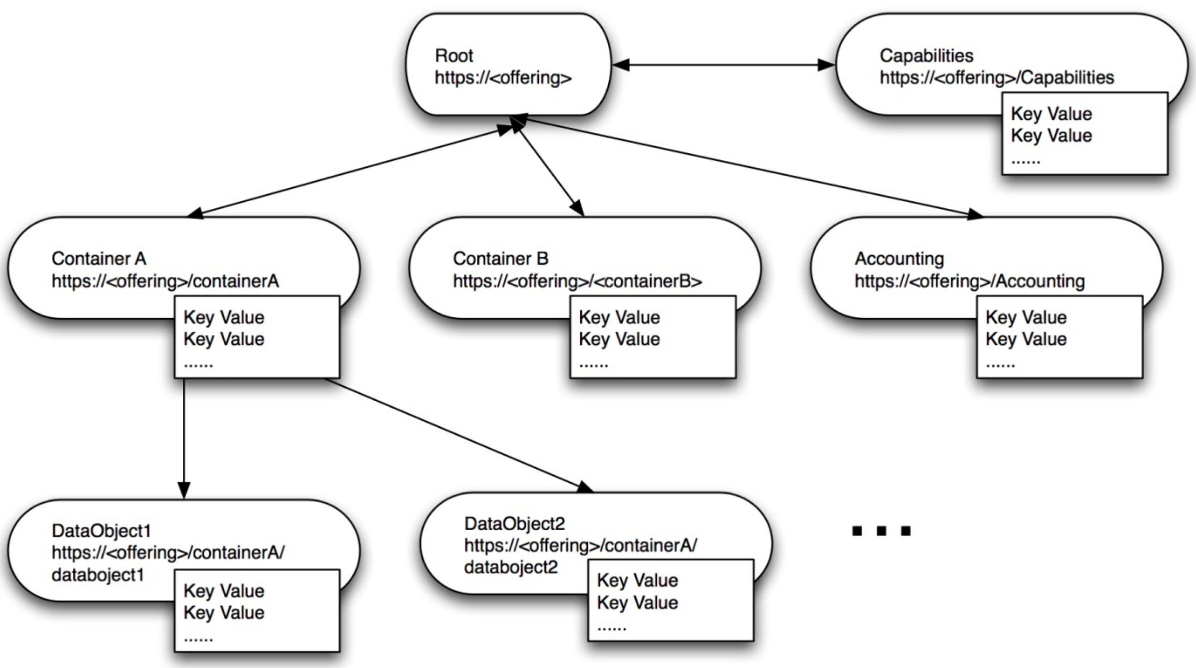

Figure 5: CDMI Interface Model

© 2009 STORAGE NETWORKING INDUSTRY ASSOCIATION

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 the Public Cloud

With one of the key objections to public cloud implementations being security, CDMI assists in easing these concerns by helping to mitigate the security risks. Using metadata to manage the data within containers, CDMI allows storage and data services to be exposed to client applications. This enables administrative and management applications to control data movement within the systems, including accounts, security access and monitoring/billing information, even for storage that is accessible by other protocols.

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 end 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 interface used when containers are exported via those protocols. CDMI also 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. Lastly, CDMI provides for a secure delete mechanism (with various algorithms) for data that has expired

CDMI is also beneficial when one considers ongoing data management needs long-term, since the potential of getting locked into a particular public cloud implementation is a critical concern for many. By offering the promise of portability through the standardization of cloud management and interfaces, CDMI can allow users to take advantage of public cloud offerings without future dependence on a particular platform. CDMI also offers a standard format which can be used to move data and its associated metadata within a cloud or between clouds. This means that data can be moved (even out

© 2009 STORAGE NETWORKING INDUSTRY ASSOCIATION

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.

How CDMI Will Roll Out

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 a public cloud. First of all, CDMI can be deployed side by side with existing proprietary interfaces. This allows existing applications to move over to the standard interface as they desire to take advantage of the features. The side-by-side deployment allows the dame 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. If the cloud provider implements the CDMI accounting, that can be used by the customer to administer the security and 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 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 also 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 public clouds as part of an organization's long-term storage strategy. The availability of storage capacity on a pay as you go basis, combined with scalability and ease of use can offer much-needed cost-savings. However, concerns about security, privacy and vendor lock-in have prevented many users from taking advantage of the benefits of public cloud storage.

CDMI, by enabling interoperable cloud storage and data management, provides a new level of management to public 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 public 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

© 2009 STORAGE NETWORKING INDUSTRY ASSOCIATION

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.

© 2009 STORAGE NETWORKING INDUSTRY ASSOCIATION
SNIA