Cloud Storage Use Cases

Version 0.5 rev 0

Publication of this Trial-Use Draft Specification for trial use and comment has been approved by the SNIA Technical Council and the Cloud Storage TWG. Distribution of this draft specification for comment shall not continue beyond 6 months from the date of publication. It is expected, but not certain that following this 6 month period, this draft specification, revised as necessary will be submitted to SNIA Technical Council for final approval. Discussions concerning this document should be held on snia-cloud@googlegroups.com, suggestions for revision and concrete content addition should be directed to http://snia.org/feedback.

Trial-Use Draft

June 8, 2009
## Revision History

<table>
<thead>
<tr>
<th>Revision</th>
<th>Date</th>
<th>Sections</th>
<th>Originator:</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The SNIA hereby grants permission for individuals to use this document for personal use only, and for corporations and other business entities to use this document for internal use only (including internal copying, distribution, and display) provided that:

1. Any text, diagram, chart, table or definition reproduced must be reproduced in its entirety with no alteration, and,

2. Any document, printed or electronic, in which material from this document (or any portion hereof) is reproduced must acknowledge the SNIA copyright on that material, and must credit the SNIA for granting permission for its reuse.

Other than as explicitly provided above, you may not make any commercial use of this document, sell any or this entire document, or distribute this document to third parties. All rights not explicitly granted are expressly reserved to SNIA. Permission to use this document for purposes other than those enumerated above may be requested by e-mailing tcmd@snia.org please include the identity of the requesting individual and/or company and a brief description of the purpose, nature, and scope of the requested use.

Copyright © 2009 Storage Networking Industry Association.
Table of Contents

1 OVERVIEW ......................................................................................................................... 6

2 USE CASES ......................................................................................................................... 7
    2.1 WEB FACING APPLICATIONS......................................................................................... 7
        2.1.1 Media Streaming ................................................................................................ 7
        2.1.2 Social Media Sites ............................................................................................ 7
        2.1.3 Specialized Content Storage & Sharing (ex: Pictures) .................................... 7
    2.2 UNSTRUCTURED DATA STORAGE ................................................................ .......... 7
        2.2.1 General Content Storage ("Cloud Drive") ......................................................... 7
        2.2.2 General Content Storage with Synchronization to/from the cloud .................. 8
    2.3 BACKUP TO THE CLOUD ......................................................................................... 8
        2.3.1 Backup Software running on, some, local machines – destination Cloud Storage ................................................................. 8
        2.3.2 Backup of Cloud Computing Data .................................................................... 9
        2.3.3 Backup from one cloud provider to the other .................................................. 9
        2.3.4 Restore (i.e. Give me back all my laptop data) ................................................ 9
    2.4 ARCHIVE/RETENTION TO THE CLOUD ................................................................. 9
        2.4.1 Retention Period: “Keep my files for X amount of time” .................................. 10
        2.4.2 Secure Deletion: “When it’s gone, it’s REALLY gone” ..................................... 10
        2.4.3 eDiscovery : “Satisfy my subpoenas“ .............................................................. 10
    2.5 PRESERVATION IN THE CLOUD ............................................................................. 10
        2.5.1 Libraries & University archives/repositories .................................................... 10
    2.6 DATABASES IN THE CLOUD ..................................................................................... 11
        2.6.1 Cloud Table Storage ....................................................................................... 11
    2.7 STORAGE FOR CLOUD COMPUTING ................................................................. 11
        2.7.1 Storage for IaaS ............................................................................................. 12
        2.7.2 Storage for PaaS ........................................................................................... 12
        2.7.3 Storage for SaaS (Software as a Service) ...................................................... 12
    2.8 CONTENT DISTRIBUTION (PROPAGATING DATA GEOGRAPHICALLY) ............... 12
    2.9 CLOUD STORAGE PEERING (I.E. “INTERCLOUD STORAGE”) ......................... 12

3 REQUIREMENTS ................................................................................................................. 13
    3.1 DATA ACCESS ........................................................................................................... 13
        3.1.1 Performance ...................................................................................................... 13
        3.1.2 Availability ...................................................................................................... 13
        3.1.3 Security ............................................................................................................ 14
        3.1.4 Migration (in/out) ......................................................................................... 14
    3.2 STORAGE MANAGEMENT ......................................................................................... 14
        3.2.1 Provisioning ...................................................................................................... 14
        3.2.2 Snapshot/Clone .............................................................................................. 14
3.2.3 Metering ................................................................. 14
3.2.4 Security ........................................................................ 14
3.2.5 Monitoring ................................................................. 14

3.3 DATA MANAGEMENT ......................................................... 14
3.3.1 Metadata support .......................................................... 14
3.3.2 Deletion .......................................................................... 14
3.3.3 Rename .......................................................................... 15
3.3.4 Data Movement .............................................................. 15
3.3.5 Synchronization Support .................................................. 15

3.4 POLICY ............................................................................. 15
3.4.1 QoS for Storage .............................................................. 15
3.4.2 Retention for Data .......................................................... 15

3.5 DATA SEARCH ................................................................. 15
3.5.1 Indexed data ................................................................. 15
3.5.2 Metadata ...................................................................... 15
1 Overview

This document summarizes the use cases in various Cloud Storage offerings.
2 Use Cases

2.1 Web Facing Applications

Web facing applications will typically use a Cloud Storage offering that provides the data directly to the user’s browser using a URL. The data is typed (MIME) and the browser invokes the appropriate application to view the data.

2.1.1 Media Streaming

Media (audio, video) files are served as a stream of data, allowing use of parts of the data within the file without requiring all data in the file to have been received by the client.
Example(s): YouTube

2.1.2 Social Media Sites

Social media sites include Myspace, Facebook, Twitter, Blogs, etc. Cloud Storage is used as an auxiliary storage space augmenting the web facing social application.
Example(s): Joyent

2.1.3 Specialized Content Storage & Sharing (ex: Pictures)

Pictures and content are stored in Cloud Storage (URL based typically). A content management system is used to keep track of additional metadata associated with the data.
Example(s): Smugmug is an example of this.

2.2 Unstructured Data Storage

2.2.1 General Content Storage (“Cloud Drive”)

This is a pre-allocated storage space (LUN, Filesystem) that is exported via standard client protocols (ex: WebDAV, NFS, CIFS), and “mounted” on a local machine. Normal POSIX semantics are available at that point for creating/reading/writing/deleting the files.
Example(s): A number of vendors have offerings in this space. A sub case is cloud desktop. Examples: iCloud, ThinkGrid

2.2.2 **General Content Storage with Synchronization to/from the cloud**

This is the ability to synchronize local client data, from multiple clients, with a Cloud Storage version. Changes are detected and then synchronization is done asynchronously and opportunistically. Access may or may not be through standard file protocols and URLs. Clients and servers have a way of sharing state describing what has/needs to be synced.

File sync management may be done through:

1) List of exclude/include files, or

2) A dedicated “folder” for synchronize between machines

*Note: Synchronization may be layered on top of the standards we define in this TWG, however, there will be no attempt to standardize such sync’ing protocols in this TWG.*

There are generally three elements to such syncing:

1. Client side interface for file input/output
2. Server side method for data presentation to clients
3. Policy for the synchronization and visibility control

Example(s):
Implementations: MobileMe, Windows Live Mesh, Combination of Google Docs and Google Gears.
Frameworks: Adobe Air

---

2.3 **Backup to the Cloud**

![Backup diagram](image)

2.3.1 **Backup Software running on, some, local machines – destination Cloud Storage**

This is local backup software or backup server using Cloud Storage as the destination of backup data.

2.3.1.1 **Backup Software on each machine (i.e. TimeFinder)**

This is a backup application that only backs up a single machine.
Example(s): Iron Mountain’s Connected Backup for PC
2.3.1.2 **Backup Server based for multiple local machines**
This is a local, central backup server that aggregates the use of the Cloud Storage for one location. It generally takes the form of an appliance giving the user an interface to manage the appliance. Also the appliance would back up its own metadata.
Examples: Cirrusstore
Note: Zmanda provides software that can be deployed this way.

2.3.1.3 **File Server Appliance locally with embedded backup to Cloud**
This is a NAS server with integrated backup to the Cloud.
Examples: Datto, Seagate Free Agent, Iomega’s StoreCenter IX2, HP’s MediaSmart Server, Cachengo

Note on Server implementations: A common technique used by some local servers is to have the client computers turn on data sharing (i.e. becoming a CIFS server). Then having the local backup server become a CIFS client of the backup client and then backing up the data in that manner. This is an elegant way to circumvent having to install 3rd party backup software on the backup clients.

2.3.2 **Backup of Cloud Computing Data**
Backup of the data used in Cloud Computing (IaaS).
Example(s): vSphere (includes de-dup as well)

2.3.3 **Backup from one cloud provider to the other**
This is the case of using a second cloud provider as the target of backup data from the first cloud provider.

2.3.4 **Restore (i.e. Give me back all my laptop data)**
This is the obvious reason of why you are doing the backups in the first place, needing to restore. Most solutions allow for both online restores as well as physical shipment of media to the customer.
Examples: Mozy ships DVDs of data. R1Soft is doing bare metal restore.

2.4 **Archive/Retention to the Cloud**
This is the use case of using the Cloud Storage for archiving of data. Theoretically XAM should be an ideal interface for this.
Example(s): Iron Mountain - VFS (not XAM based)
Considerations for the use case:

1) Does the user maintain a local copy?

2) Does the Cloud provider do virus scans or other operations on your behalf? It’s not useful to have to pull the files back over the wire.

2.4.1 Retention Period: “Keep my files for X amount of time”
This is the case where you define the period of time that you guarantee files will be retained.

2.4.2 Secure Deletion: “When it’s gone, it’s REALLY gone”
This is the case where the service provider provides a means of deleting data in such a way that it’s truly gone, i.e. not recoverable by any means. A common method for this is encrypting the data at rest and then shredding the encryption keys.

2.4.3 eDiscovery: “Satisfy my subpoenas“
This is providing a service such that when certain documents are required to be produced for a court case, the appropriate documents are produced without undue time or costs.
Note: Email archiving may be a specialization of this case.

2.5 Preservation in the Cloud
Preservation is distinguished from Archive/Retention in that the goal of preservation is to actively maintain the upkeep of information, most likely for long periods of time.

2.5.1 Libraries & University archives/repositories
Many of these repositories leverage RDF as a way to describe the data and it’s relation to other data. Fedora Commons and other content managers are used to keep track of the metadata. Preserving of machine images along with the data so that a user can at a minimum display the information using the application that generated the data.
Example(s): Fedorazon
2.6 Databases in the Cloud

2.6.1 Cloud Table Storage

Cloud Table Storage falls into the following categories:
1) Horizontally Scalable, Object-Relational:
   Examples: Microsoft Azure Tables, Google BigTable, (hyperTable), SimpleDB
2) Vertically Scalable, Traditional Relational:
   Example(s): SQL services
3) Document Model:
   Example(s): CouchDB

Note: We will not be standardizing a Cloud Data Storage Interface for Cloud Table Storage at this time, but the Cloud Data Management Interface is applicable regardless.

2.7 Storage for Cloud Computing
2.7.1 Storage for IaaS
Traditional data storage which is accessible as part of the computing infrastructure. Example: EC2 leverages S3 as if it were a private cloud.

2.7.1.1 Image Storage
The image of the Guest OS which is made available to hypervisors for starting a VM.

2.7.1.2 Guest Auxiliary Storage
Provision the storage space, at a given QoS, which the guest needs beyond the boot storage.

2.7.1.3 Application Image Storage
The application is maintained in the cloud and invoked locally. Maintenance of the application is done centrally. Streamed in using a statistically probable execution order so that not all bits need be present to start executing. 
This is a function of the distribution network and can be layered on top of the interfaces we are defining.

2.7.2 Storage for PaaS
Note: listing here for completeness, however this type of storage is not usually surfaced or manageable from this TWGs point of view.

2.7.3 Storage for SaaS (Software as a Service)
Note: listing here for completeness, however this type of storage is not usually surfaced or manageable from this TWGs point of view.
Examples: Salesforce.com uses S3

2.8 Content Distribution (Propagating Data Geographically)
Distributing data globally for the purposes of decreasing latency and increasing scalability.
Examples:
1) “Hot” media serving – move to point of presence, replicate out to caches, then recover the resources when unused.
2) Data transformation in route (i.e. localization, NTSC->PAL)
Note: This TWG will support this use case as a layering on top of the standard interfaces.

2.9 Cloud Storage Peering (i.e. “Intercloud” Storage)
This is the concept of having the Storage Clouds of different Cloud Storage Providers being able to interoperate between each other (in other words doing for Storage Clouds what the Internet did for separate, proprietary networks).
Possible Characteristics:
• Shared storage and replication between cloud storage offerings.
• Distribute the data across cloud storage providers (possibly via a storage broker that provides a blended rate).
• Data may be erasure encoded as well as replicated.
• Caching and distribution between the client and “dumb storage” provider, geographic staging and replication.
• Activation and de-activation relative to some trust model – activation requires assembly from the erasure coded blocks, decoding and decryption, de-activation involve encryption, encoding and distribution
• Topology of the network is more nuanced than the typical two tier processing model and more dynamic as well.

Examples:
1. “Federated” Cloud Storage
2. “Cloud Exchange”
3. Cloud “Bursting”, offloading, Hybrid Internal/External Clouds

3 Requirements

For the “SNIA Cloud Storage Data Storage Interface” and the “SNIA Cloud Storage Data Management Interface”.

3.1 Data Access

• Object & Containers shall be accessible through a (MIME-Typed) URL that can be embedded in a web page and accessible through a browser. IETF RFC 2616 based retrieval of objects. It should be possible to “fix” that URL as part of the service (doesn’t change while the object is with that vendor).
• Objects & Containers shall also be referable through a location independent URN.

3.1.1 Performance

We shall standardize the expression of performance requirements for data access with Data System Metadata. Vendors should be able to extend this for specific performance capabilities.

3.1.2 Availability

We shall standardize the expression of availability requirements for data access with Data System Metadata. Vendors should be able to extend this for specific availability capabilities.
3.1.3 **Security**
We shall standardize the expression of security requirements for data access with Storage System Metadata. Vendors should be able to extend this for specific security capabilities.

3.1.4 **Migration In/Out (Import/Export by a provider)**
We shall define a standard format for interoperable movement of data and its associated metadata between Cloud Vendors.

### 3.2 Storage Management

3.2.1 **Provisioning**
We shall define an interface to create a container and associate default data requirements with it. We shall support both hard and soft containers as well as nested soft containers.

3.2.2 **Snapshot/Clone**
We shall provide operations for snapshot and clone of a container.

3.2.3 **Metering**
We shall define an interface for metering storage capacity utilization and bandwidth usage on a container and object basis. The storage equivalent of a “Call Detail Record” shall be available through the interface. The interface will report results of performance against the requirements.

3.2.4 **Security**
Both the data storage and data management interfaces shall support: authentication, authorization, auditing, privacy and integrity.

3.2.5 **Monitoring**
The aggregate and individual statistics on containers and objects shall be available through the data management interface.

### 3.3 Data Management

3.3.1 **Metadata support**
We shall define Storage System Metadata, Data System Metadata and User Metadata as part of the interface.

3.3.2 **Deletion**
We shall define a standard means of deleting containers and objects. Secure deletion (shredding) shall be defined, as well as other dispositions.
3.3.3 **Rename**
We shall define a standard means of changing URN of an object or a container.

3.3.4 **Data Movement**
We shall define a standard means of moving an object or container from one container to another. (i.e. change the URL for an object or container)

3.3.5 **Synchronization Support**
We shall define a standard means of notification of changes to a container and it’s data objects sufficient to enable synchronization.

3.4 **Policy**

3.4.1 **QoS for Storage**
We shall standardize the Data System Metadata that allows Policies to drive data services within the cloud.

3.4.2 **Retention for Data**
We shall standardize retention of Data System Metadata compatible with XAM.

3.5 **Data Search**

3.5.1 **Indexed data**
We shall define a standard compatible with XAM level 2 query.

3.5.2 **Metadata**
We shall define a standard compatible with XAM level 1 query.