Interoperable Cloud Storage with the CDMI Standard

Mark Carlson, SNIA TC and Oracle
Co-Chair, SNIA Cloud Storage TWG
SNIA Legal Notice

- The material contained in this tutorial is copyrighted by the SNIA.
- Member companies and individual members may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced in their entirety without modification
  - The SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA Education Committee.
- Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.
- The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.
Abstract

The SNIA has published the CDMI Cloud Storage standard for implementation by cloud storage vendors as well as Public and Private clouds. This tutorial will provide an overview of the features of the new standard and explain how interoperability between clouds is achieved.

- Now that the standard is available, what should you be requiring from your cloud vendors?
- How can you expect this standard to roll out in implementations?

Learning Objectives

- How this cloud storage standard can achieve interoperability and what this interoperability means to you.
- How you can use the cloud storage standard both internally for private clouds as well as for public clouds.
- Who has plans to implement CDMI and how will the implementations roll out.
Agenda

- Some background on cloud storage
  - CDMI Overview
  - What Is Cloud Storage Used For?

- SNIA Cloud Efforts
  - Cloud Storage TWG
  - Cloud Storage Initiative

- The Cloud Storage Reference Model
  - Existing Cloud APIs
  - Leveraging the Storage Industry Resource Domain Model
  - The Big Picture
  - Using a RESTful protocol
  - Why an Industry Standard?

- CDMI Status

- Object Model, Deployment possibilities
CDMI Overview

- Data Portability Standard
  - Move Data (and most importantly – Metadata) from cloud to cloud

- Advanced Cloud Services
  - Data System Metadata allows cloud vendors to up-sell!
  - Specialized storage clouds for specific use cases

- Logging, Security, Audit Trails

- Extensible to accommodate rapid innovation in cloud market

- Moving on to ISO standardization
What Is Cloud Storage Used For?

- Elastic demand for web based media (video, eBooks, audio)
- Backup to the cloud
  - Restore, Recovery, “Seed” the backup with hard drive
- Sync of files to the cloud and multiple devices
  - Internet “Drive” secondary storage
- Archive to the cloud
  - Including Compliance, Retention and eDiscovery
- Storage for Cloud Computing
  - Support for legacy storage interfaces key
Storage Vendors, Cloud Providers, Developers
- >200 Technical Work Group members
- Google group for broader community (> 450 members): http://groups.google.com/group/snia-cloud

Cloud Data Management Interface (CDMI) v 1.0
- SNIA Architecture Standard
- Next step ISO standardization

CDMI Reference Implementation
- Java based, uses any POSIX filesystem
- Open Source License (BSD)
SNIA Cloud Storage Initiative

- Gaining Momentum for Cloud Storage
  - Supporting the development and adoption of CDMI, Cloud Storage
  - Marketing, Outreach, Education on Cloud Storage
  - Requirements gathering
  - Premier Organization promoting Cloud Storage and associated Standards
- 28 Member companies and growing
  - Multiple events including Cloud Burst event focused exclusively on cloud storage
Cloud Data Management Interface

- Applicable to three types of Cloud Storage:
  - Cloud Storage for Cloud Computing
    - Whitepaper at snia.org/cloud – the management interface for the lifecycle of storage in a compute cloud
  - Public Storage Cloud
    - Whitepaper at snia.org/cloud – both a Data Path for the Cloud and a Management Path for the Cloud Data
  - Private Cloud Storage
    - As well as hybrid clouds
    - An API for Storage Vendors selling into Cloud based solutions

- Semantics
  - Simple Containers and Data Objects with tagged Metadata
  - Data System Metadata expresses the data requirements

- Protocol
  - RESTful HTTP as “core” interface style
  - JSON (JavaScript Object Notation) – format of the representations are extensible
**Represention State Transfer**
- Started with [Dissertation by Roy Fielding](#) outlining the principles
  - A form of web services (but not based on WS-*)

**Addressability**
- Every object (resource) is addressable through a unique identifier

**Uniform, Constrained Interface**
- Use only HTTP verbs and model other semantics in the data model
- Allows for Familiarity (low learning curve), Interoperability and Scalability

**Representation Oriented**
- Complexity is in the representations

**Communicate Statelessly**
- No persistent client-server connections
Why a RESTful approach for a Cloud Storage Standard

- Simplicity Rules!
- Common Infrastructures in many Languages on many Platforms
- Low learning curve leads to developer adoption
- Developer adoption creates eco-system around API
- Eco-system eases adoption by vendors and customers
- Scale-out implementation feasibility
Why not just a common library?

- There are several common Cloud Libraries available
  - **Libcloud** (python), **Jclouds** (Java), **Simple Cloud** (PHP), etc.
  - They all write adapters from the common library to each of the proprietary interfaces
  - Adapters must be maintained as interfaces evolve
  - Library is under control of 3rd party, so vendors not likely to support directly
  - Each language ends up propagating it’s own common library with no common semantics between them
Why not just adopt one of the existing interfaces?

- Despite the “open” licensing of several existing cloud storage interfaces, they all remain under the change control of a single vendor.
- No cloud vendor wants to have a competitor have change control over their interface.
  - Thus they release their own interface which they do have change control over.
- This leads to the propagation of multiple interfaces, each essentially locking developers/customers into that service.
- CDMI is under change control of a standards body, accommodates requirements from multiple vendors and can be extended for proprietary functions.
Clients can be in the cloud or enterprise and provide additional services (computing, data, etc.).

Clients acting in the role of using a Data Storage Interface

Management of the Cloud Storage can be standalone or part of the overall management of your cloud computing.

Clients acting in the role of Managing Data/Storage
Models for Cloud Ecology

- **Cloud Federation**
- **Computing Cloud**
- **Object Storage Cloud**
- **Distribution Cloud**

- Multiple Distribution Points
- Cloud Peering
- Data Usage
CDMI Overview

CDMI Basic flow:

CDMI Client issues requests

HTTPS: PUT, GET, POST, DELETE
Mime Type: application/cdmi-______
cdmi-object, cdmi-container, cdmi-queue, cdmi-domain, cdmi-capability*
Data, Metadata

CDMI Implementation issues response

HTTP Status (200 OK, 201 Created, etc.)
Mime Type: application/cdmi-______
cdmi-object, cdmi-container, cdmi-queue, cdmi-domain, cdmi-capability
Data, Metadata

*CDMI MIME Types are Registered with IANA
Model for the Interface

The resources which are accessed through the RESTful interface
# CDMI is maturing as a standard

<table>
<thead>
<tr>
<th>Maturity Level*</th>
<th>Description</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. No Standards</td>
<td>Standardization needed</td>
<td>Encourage standards development</td>
</tr>
<tr>
<td>2. Under Development</td>
<td>Discussions within standards groups. Open source project launched.</td>
<td>Monitor and provide feedback to standards development</td>
</tr>
<tr>
<td>3. Specification Document Published</td>
<td>Initial specification posted for public review</td>
<td>Review specification and plan testing</td>
</tr>
<tr>
<td>4. Initial Reference Implementation</td>
<td>Reference implementation available</td>
<td>Evaluate reference implementation</td>
</tr>
<tr>
<td>5. Early Third Party Testing</td>
<td>Evaluation in test environments</td>
<td>Pilot Projects should consider use</td>
</tr>
<tr>
<td>6. Initial Production Implementations</td>
<td>Successful use in production</td>
<td>Mainstream projects should consider use</td>
</tr>
<tr>
<td>7. Many Deployments</td>
<td>Widespread use by many groups</td>
<td>Projects should use the standard as a default</td>
</tr>
<tr>
<td>8. Accepted Standard</td>
<td>De facto or de jure acceptance as a standards</td>
<td>Projects should use unless special circumstances require exemption</td>
</tr>
<tr>
<td>9. Aging Standards</td>
<td>Newer standards are under development</td>
<td>Projects should explore alternatives</td>
</tr>
</tbody>
</table>

*Source: Draft NIST Cloud Standards Roadmap
CDMI addresses SAJACC Use Cases

- CDMI is an HTTP/RESTful protocol with TLS support for securing the data, metadata, and communications.
  - CDMI Content Types (MIME) are standardized by IANA (IETF RFC).

SAJACC Use Case 3.4: Copy Data Objects into a Cloud

CDMI Client issues requests:
- HTTPS: PUT
  - ContentType: application/cdmi-object
  - Data, Metadata

CDMI Implementation issues response:
- HTTP Status (200 OK)
  - ContentType: application/cdmi-object
  - Metadata
CDMI uses existing standards

CDMI data objects can be accessed by standard browsers and internet tools (subject to owner’s access)

SAJACC Use Case 3.5: *Copy Data Objects out of a Cloud*

CDMI Client issues requests

- HTTPS: GET
  - ContentType: application/cdmi-object

CDMI Implementation issues response

- HTTP Status (200 OK)
  - ContentType: application/cdmi-object
  - Data, Metadata

Browser issues requests

- HTTP: GET
  - Accept: *

Implementation issues response

- HTTP Status (200 OK)
  - ContentType: (based on data type)
  - Data
CDMI defines interoperable services

- CDMI data objects may “order” data services from the cloud
  - Secure Erasure, Encryption, Replication, Retention, Backup/Restore, Tiering, Hashing, Preservation, etc. (extensible)
  - Done through Data System Metadata (key/value) on the Containers or Objects

SAJACC Use Case 3.6: *Erase Data Objects in a Cloud*

CDMI Client issues requests

HTTPS: DELETE
ContentType: application/cdmi-object

CDMI Implementation issues response

HTTP Status (200 OK)
CDMI enables Data Portability

- CDMI standard defines an interoperable format for moving data and associated metadata between cloud providers interoperably
- And ensuring that the new cloud provides the same services

SAJACC Use Case 4.1:
Copy Data Objects between Cloud Providers

CDMI Client issues requests
HTTPS: PUT
ContentType: application/cdmi-object
Metadata: serialize
(source container/object)

CDMI Implementation issues response
HTTP Status (200 OK)
ContentType: application/cdmi-object
Metadata

Data Storage Cloud
FeedEx a Disk

CDMI Client issues requests
HTTPS: PUT
ContentType: application/cdmi-object
Metadata: deserialize
(destination container/object)

CDMI Implementation issues response
HTTP Status (200 OK)
Domains of Resource Management

**Information Resource Domain**
Services understand the semantics of the content in context

**Data Resource Domain**
The content is opaque to the Services and without context

**Storage Resource Domain**
The bits are contained by these Services

**Information Policies**
Data classified according to importance to organization

**Data Policies**
Data treated according to requirements, lifecycle

**Storage Policies**
Ensure correct and reliable operation
All of these interfaces support some or all of this model. The key to retaining the simplicity of the cloud, however, is in the use of metadata to drive the underlying services so that users need not manage the services themselves.
How does CDMI fit into a storage cloud?

- **Small Private Cloud**
  - Deployed as a layer above NAS box, or may also be embedded

- **CDMI Containers and Objects** are mapped to a mounted filesystem’s directories and files

- **CDMI can also be used** to configure NAS storage not available through CDMI data path
How does CDMI fit into a storage cloud?

- **Large Scale out Cloud**
  - Deployed as a horizontal set of parallel filesystem clients with requests balanced across them
  - Storage is implemented by a set of data servers with a common metadata server
For More information

❖ One Web Site to Remember: http://snia.org/cloud

❖ Large Cloud Storage Community
  • http://groups.google.com/group/snia-cloud
  • http://twitter.com/SNIAcloud (@SNIAcloud)
  • http://www.google.com/profiles/SNIAcloud
Other Cloud Tutorials

- Cloud Backup and Recovery Requirements
- The Role of WAN Optimization in Cloud Infrastructures
- Securing the Cloud - Using Encryption and Key Management to solve today's Cloud Security challenges

Also visit the Cloud Storage Hands-On Lab

Pavilion at SNW®
Q&A / Feedback

Please send any questions or comments on this presentation to SNIA: trackcloudtechnologies@snia.org

Many thanks to the following individuals for their contributions to this tutorial.
- SNIA Education Committee

Scott Baker
David Slik
Rich Ramos
Members of the SNIA Cloud Storage TWG