Archiving On-Premise and in the Cloud

March 2015
Cloud Storage

Storage accessed over a network via web services APIs.

http://swift.example.com/v1/account/container/object

Source: http://docs.openstack.org/admin-guide-cloud/content/objectstorage_characteristics.html
OpenStack Swift Object Storage Overview

Source: http://docs.openstack.org/admin-guide-cloud/content/section_storage-nodes.html
Benefits of Cloud Storage

- Limitless scalability
- Custom metadata
- Single namespace
- Simplified management
Why Are Companies Moving to the Cloud?

- Pay-as-you-go model
- Improved agility (time-to-market)
- Competitive advantage
- Stronger security
## Trade-offs between On-premise and Cloud

Business Needs Should Drive the Solution

<table>
<thead>
<tr>
<th></th>
<th>On-Premise</th>
<th>Cloud</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latency</td>
<td>Low due to locality</td>
<td>Higher due to network</td>
</tr>
<tr>
<td>Throughput</td>
<td>Fast Performance</td>
<td>Limited by bandwidth</td>
</tr>
<tr>
<td>Security</td>
<td>Data on-site</td>
<td>Third party</td>
</tr>
<tr>
<td>Investment</td>
<td>Up-front</td>
<td>Pay as you go</td>
</tr>
<tr>
<td>Infrastructure</td>
<td>IT staff to manage</td>
<td>Provider managed</td>
</tr>
<tr>
<td>Customization</td>
<td>More</td>
<td>Less</td>
</tr>
<tr>
<td>Compliance</td>
<td>Tailored</td>
<td>Standards supported?</td>
</tr>
<tr>
<td>Stability</td>
<td>Owned</td>
<td>Risk of going away</td>
</tr>
</tbody>
</table>
# Data Transfer

<table>
<thead>
<tr>
<th>Available Internet Connection</th>
<th>Theoretical Min. Number of Days to Transfer 1TB at 80% Network Utilization</th>
<th>When to Consider AWS Import/Export?</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1 (1.544Mbps)</td>
<td>82 days</td>
<td>100GB or more</td>
</tr>
<tr>
<td>10Mbps</td>
<td>13 days</td>
<td>600GB or more</td>
</tr>
<tr>
<td>T3 (44.736Mbps)</td>
<td>3 days</td>
<td>2TB or more</td>
</tr>
<tr>
<td>100Mbps</td>
<td>1 to 2 days</td>
<td>5TB or more</td>
</tr>
<tr>
<td>1000Mbps</td>
<td>Less than 1 day</td>
<td>60TB or more</td>
</tr>
</tbody>
</table>

http://aws.amazon.com/importexport/details/
Public Cloud Storage Use Cases

- Backup
- Archive
- Content Storage and Distribution
  - Office productivity
  - Media sharing
- Cloud Native Application Data
Data Protection 3-2-1 Rule
3 Copies - 2 Different Medias - 1 Offsite

The 3-2-1 rule ensures **logical** and **physical** data protection.

3 Total Data Copies

- **Copy 1**: Production
- **Copy 2**: Backup to disk or Tape
- **Copy 3**: Offsite Copy

Backstop = Last Line of Defense
## Preservation Considerations with Cloud Storage

<table>
<thead>
<tr>
<th>Preservation Requirement</th>
<th>Cloud Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive metadata</td>
<td>Cloud storage provides custom metadata</td>
</tr>
<tr>
<td>Fixity</td>
<td>Is check done outside of cloud storage?</td>
</tr>
<tr>
<td>System security and access control</td>
<td>Are they sufficient?</td>
</tr>
<tr>
<td>Auditable Event Tracking</td>
<td>Are there records of actions associated with an asset?</td>
</tr>
<tr>
<td>Immutability</td>
<td>Is there versioning so that originals are unchanged?</td>
</tr>
<tr>
<td>Transcoding</td>
<td>Does it require a recall and new upload?</td>
</tr>
</tbody>
</table>
IDC is observing a shifting trend in cloud and Web-scale architectures to a multitier storage strategy in order to sustain growth while maintaining existing data.

IDC
Technology Assessment: Cold Storage is Hot Again
90% of an Organization's Data is Passive

Facebook Photo Access Patterns

Note: Data is from the Open Compute Summit IV, January 2013, Santa Clara, California.
Source: Facebook, 2013
10PB, 5 Year TCO Comparison

On Premise Tape vs Public Cloud Storage

- Amazon Glacier Pricing
  - $.01 per GB/month
  - Upload and retrieval Requests = $0.05 per 1,000
  - Data Transfer IN To Amazon Glacier = $0.00
  - Data Transfer OUT From Amazon Glacier To Internet = $.05 to $.09 per GB

- The cost to store 10PB for 5 years is $6 million
  - Does not include costs for upload, transferring data out, or bandwidth
  - 10gigE link would take 92+ days to transfer 10PB
  - Reading all 10PB would cost $500,000.
10PB, 5 Year TCO Comparison
On Premise Tape vs Public Cloud Storage

- **Tape Pricing**
  - $296,279.80 4,000 slot library including installation
  - $160,000.00 4,000 LTO6 tapes
  - $96,969.60 8 LTO6 tape drives
  - $232,547.65 5 years of Premier Support
  - $3,997.19 Power: 702 watts at $.13/kilowatt-hour
  - $63,750.00 Floor space: 51 sqft at $250 per sqft/year

- **Total Cost $853,544.24**
  - Glacier is 7x the cost over 5 years
“Glacier is almost 10 times as expensive as an on-premise tape system with support.”

Jack Clark, ZDNet

AWS Glacier’s dazzling price benefits melt next to the cost of tape
Oracle Sponsors OpenStack Foundation

December 10, 2013

“Oracle also plans to integrate OpenStack Object Storage into its storage portfolio, providing customers with access via OpenStack APIs to Oracle ZFS Storage Appliance and Pillar Axiom storage systems for object storage and StorageTek tape solutions for deep archiving and data protection.”

Tape Web Interface for Cold Storage / Archive

Simplify the use of Tape

Data Accessed via a web services API

Data migrated to low-cost tape based on an automated policy
Swift Tape Deployment Diagram (Logical)

- External SWIFT Client
- Swift Proxy
- Oracle HSM
- Storage Network
- Up to 512 Nodes

REST connections:
- External SWIFT Client to Swift Proxy
- Swift Proxy to Oracle HSM
- Oracle HSM to Storage Nodes
- Storage Nodes to QFS Clients

Firewall separation between External SWIFT Client and Storage Network.
Summary

- Understand the trade-offs between on-premise vs cloud
  - Business needs drive the technology decision
- Remember the 3-2-1 Rule (3 copies, 2 mediums, 1 offsite)
- There is a trend towards multiple tiers of cloud storage
  - Tape behind a web interface can provide a low-cost, lower SLA tier