Rethink Cloud Strategies for Cost Effective Enterprise Storage Management

Laz Vekiarides
ClearSky Data
Enterprise Storage Today

- **Flash**
- **Mid-Range**
- **Scale Out**

Complex, costly silos

<table>
<thead>
<tr>
<th>Capacity</th>
<th>$/TB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flash</td>
<td>$$$</td>
</tr>
<tr>
<td>Mid-Range</td>
<td>$$</td>
</tr>
<tr>
<td>Scale Out</td>
<td>$</td>
</tr>
</tbody>
</table>
Then There is Disaster Recovery

Capacity

$/TB

Flash

Mid-Range

Scale Out

More copies

$/TB

... And Then the Upgrades
What Enterprises Really Want

- High Performance Where It's Needed
- Enterprise Availability & Security
- Cloud Economics & Scalability

Capacity vs. $/TB
Today, Cloud Storage is “Where Data Goes to Die”

- Highest Performance
- Availability
- Tightest Recovery Objectives

- Application Integration
- Manageability
- Agility
- Flexibility

- Retention
- Recovery times
- Security
- Cost
Closer Look At Storage Clouds

• Elastic, reliable, durable
• Pay as-you-go
• No hardware – mostly
• Really “feels” like software

• Seems well priced

BUT…

Look at the transaction costs!
Closer Look At Storage Clouds

- Elastic, reliable, durable
- Pay as-you-go
- No hardware – mostly
- Really “feels” like software

- Seems well priced

BUT...

Look at the transaction costs!

Don’t forget transfer costs!
A Simple “Enterprise” Cost Example

- **Assumptions:**
  - 100 TB of storage
  - 10K 8K write IOPS average
  - 30K 8K read IOPS
  - Trivial mapping between objects and I/Os
  - No caching
  - Use S3 rates

- **Results get VERY expensive with heavy use**

- **Caching and packing provide only minimal cost benefit**

<table>
<thead>
<tr>
<th>Item</th>
<th>Cost (USD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage (100TB @ $29/mo)</td>
<td>$2900</td>
</tr>
<tr>
<td>PUT requests (10K/s-30 days)</td>
<td>$129,600</td>
</tr>
<tr>
<td>GET requests (30K/s-30 days)</td>
<td>$31,104</td>
</tr>
<tr>
<td>Bytes read</td>
<td>$31,104</td>
</tr>
<tr>
<td>Bytes written</td>
<td>0</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>$194,708</strong></td>
</tr>
</tbody>
</table>

That’s per month!
It’s the Latency, Stupid

(Apologies to Stuart Cheshire)

- Data travels at the speed of light
- Fast - but finite
  - $3 \times 10^8$ meters per second
  - 186000 miles per second
- Example: Boston to San Francisco
  - 2740 miles
  - 29.4 milliseconds RT
- There are more delays
  - Light travels more slowly in fiber
  - Fiber-optic repeaters every few hundred miles
  - Switches, routers
  - Protocols, virtualization, etc.
- End result: ~60ms
So, Where Exactly Is “The Cloud”?

- Amazon East is near Ashburn, VA
- West is in San Francisco
- Boston is closest to East
- Best case numbers:
  - ~12ms round trip
  - From Markley via Direct Connect Ethernet
  - Does not include time to actually access the storage
What Latency Do Applications Expect?

- Applications see end-to-end latency
  - OS/Hypervisor
  - LAN/Network
  - Storage
  - Compute
- Storage is largest part
- Best practice <10ms

**Obvious conclusion:** Compute must live near data
Some Applications Can Work Today

- Low performance, low intensity
  - Archive storage
  - Backup
  - Cold data
- Predominantly files
- Could be objects
  - Images
  - Movies
- Tape/removable disk use cases
So the conclusion?

- Cloud data is reasonably cheap as long as you don’t access it much.
- Data, apps, and users are happiest when they are nearest to each other.
  - Network latency matters. Pick a good carrier.
  - Location matters. Pick a good datacenter 😊
- With a gateway, using cloud storage remotely is only suitable for cold data.

ClearSky Data: Enterprise Storage Is About to Change Forever.
Backups
What is “Cloud Storage”?

- Scale-out object storage:
  - Blobs of data and tags (Name, size, MD5 hash, etc)
  - Simple PUT, GET, DELETE operations
  - Accessed remotely or within a public cloud

- Shamelessly lifted from S3 literature:
  - Reliable
  - Scalable
  - Fast
  - Inexpensive
  - Simple

- Caveats:
  - NOT a file system
  - NOT a block (i.e. disk) device
Requirements For Data Storage Evolve Over The Course Of Its Life

- Highest Performance
- Availability
- Tightest RPO/RTO

Primary

- Application Integration
- Manageability
- Agility
- Flexibility

Nearline

- Retention
- Recovery times
- Security
- Cost

Offline
Today, We Use Cloud For Data Only At The End of Life

- Highest Performance
- Availability
- Tightest RPO/RTO

Primary

Application Integration
Manageability
Agility
Flexibility

Nearline

Retention
Recovery times
Security
Cost

Offline