Three Proven New Ideas That Will Completely Change Your Cost Model for Big Data Storage

Mark Seamans
FileTek
www.FileTek.com
Three Proven New Ideas That Will Completely Change Your Cost Model for Big Data Storage

Mark Seamans
FileTek
www.FileTek.com
The Reality

STORAGE
The Reality

- Email
- New Documents
- Active Project Info
- Database Contents
- Active Research
- Movies Under Development

- Finalized Documents, PPTs, XLS
- Research Results and Data
- Genomic Sequence Info
- Transcoded Media Assets
- Seismic Information
- Reference Materials
- Corporate Records
- Manufacturing Data
- Digital Forensic Evidence
The Problem

TOP 10 PREDICTIONS

IDC Predictions 2012: Competing for 2020

Frank Gens

PREDCITIONS

- Big Data will join mobile and cloud as the next "must have" competency as the volume of digital content grows to 2.7ZB (1ZB = 1 billion terabytes) in 2012, up 48% from 2011, rocketing toward 8ZB by 2015.
Storage Options and Cost Profiles

- Wide Variety of Disk Storage Alternatives
  - RAID, JBOD, SAN, NAS, SATA, SAS, SSD, Hybrid, Erasure-based Encoding
- Tape-based Storage Continues to March On
  - LTO Standard Promises Large Capacity Increases, Higher Reliability, Lower Power Consumption and Very Low Cost Per TB
- Cloud-based Storage Is an Emerging Option
  - S3 and Similar
  - Amazon Glacier
Tape-based Storage Considerations

Technology Price/GB Projections
…and Tape’s Advantage Is Accelerating

Source: Oracle: 2012 Tape Summit
## Tape-based Storage Considerations

### It’s Not Only About Cost/TB

<table>
<thead>
<tr>
<th></th>
<th>Disk</th>
<th>Tape</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max shelf life (bit rot)</td>
<td>10 years</td>
<td>30 years</td>
</tr>
<tr>
<td>Best practices for data migration to new technology</td>
<td>3-5 years</td>
<td>8-12 years</td>
</tr>
<tr>
<td>Uncorrected Bit Error Rate, Probability (avg 1 error in x TB)</td>
<td>$10^{-14}$ (~10’s of TB)</td>
<td>$10^{-19}$ (~1 million TB)</td>
</tr>
<tr>
<td>Power and cooling</td>
<td>290X</td>
<td>X</td>
</tr>
</tbody>
</table>

“The cost of energy alone for the average disk-based (archive) solution exceeds the entire TCO of the average tape-based solution.”

The Clipper Group, “In Search of the Long-term Archiving Solution, December 2010

If a vendor offers the disk for “free,” it is a BAD DEAL!!!!

Source: Oracle: 2012 Tape Summit
Amazon Glacier

- Low cost cloud storage targeted at long term retention, infrequent data access, and 4-hour data access time from request to availability

<table>
<thead>
<tr>
<th>Storage</th>
<th>Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amount of data stored</td>
<td>Storage Cost</td>
</tr>
<tr>
<td>500 TB</td>
<td>$307,200.00</td>
</tr>
<tr>
<td>Duration data was kept</td>
<td>Retrieval Cost</td>
</tr>
<tr>
<td>1800 days</td>
<td>$44,544.00</td>
</tr>
<tr>
<td>Early deletion</td>
<td>Deletion Cost</td>
</tr>
<tr>
<td>Amount of data to be deleted</td>
<td>$0.00</td>
</tr>
<tr>
<td>0 GB</td>
<td>Transfer Cost</td>
</tr>
<tr>
<td>Duration data was kept</td>
<td>$2,603.91</td>
</tr>
<tr>
<td>Retrieval</td>
<td>Total Cost</td>
</tr>
<tr>
<td>Amount of data to be retrieved</td>
<td>$354,347.91</td>
</tr>
<tr>
<td>25 TB</td>
<td></td>
</tr>
</tbody>
</table>
- What is the new way?
- How do I get there?
- How much could I save?
- Simple User Model
- Rich Analysis & Migration
- CFO-ready ROI Model
Key Concepts

- Not about picking storage technology but rather about implementing a strategy for managing storage
- Anticipate the fact that the ‘next big thing’ in storage technology will always be on the horizon
- Look for ways to blend technologies to match capabilities to requirements while optimizing cost
- Provide access to data that is transparent to users
Intelligent Storage Management
Intelligent Storage Management

Application Information Access

- Policy Management
- Active Validation & Repair
- CIFS Interface
- NAS Interface
- FTP Interface
- System Mirroring
- Integrated Backup
- File Version Management
- Point-in-Time File Access
- WORM Tape Support

Direct User Information Access

- Global Name Space Mgmt.
- System Health Monitor
- Soft/Hard Deletes
- Automated Data Migration
- File Stubbing & Links
- Silent Corruption Detection
- Media Migration Mgmt.
- MD5 Checksum
- Collector Retention
- Onsite/Offsite Data Mgmt.
- Data Encryption
- Integrated Backup/Archive
- Optimized Retrieval based on File Size and Media
- Audit Trail
- LTFS Support
- Capacity Reporting
- System Alerts
- Intelligent Buffering
- Media Consolidation

Storage
Disk – Tape – SSD – Cloud - Other
Dynamic Storage Mix

- Primary = Disk
- Backup = Tape

- Primary = Tape
- Backup = Cloud
- Archive = Tape (Offsite)
Analysis to Understand Your Data

- Utilization by:
  - Creation Date
  - Last Access Date
  - File Size
  - File Type
  - Number of Files
  - More
  - In-Place Analysis
  - Export to Excel or other external apps
Policy Driven Data Migration

- Migration by:
  - Copy
  - Migrate
  - Synchronize

- Inclusion and Exclusion Management

- Intelligent Linking

- Hash Verification

- Schedule Management
Value Realization Modeler (VRM)

- Storage Migration ("Active Archive") Savings
- Data Mirroring ("Active/Active") Savings
- Backup Replacement Savings
- Environmental/Facility ("Green") Savings
Results of a published case study conducted by the Case Study Forum with the participation of Carnegie Institute Department of Embryology established the following:

**ROI in Action: Carnegie Institute**

- **ROI** – 781%
- **Payback** – 3 months
- **Cumulative Net Value** - $773,900
- **Net Present Value** - $626,372
Results of a published case study conducted by the Case Study Forum with the participation of Wistar Institute:

ROI in Action: Wistar Institute

ROI – 239%
Payback – 16 months
Cumulative Net Value - $670,000
Net Present Value - $490,000
A Complete Approach

Analyze → Migrate → Manage

Compelling Business ROI Model
Summary

- Big Data breaks many of the old school approaches to storing, backing up and restoring data
- The exploding requirements for storage are accelerating the pace of change in storage technology
- Anticipating that change and putting systems & processes in place for understanding and managing your data and its lifecycle are critical
Leading organizations are treating storage evolution and optimization as a core competence and *part of the process* rather than having it be long series of disjointed IT projects.

The results:
- A smoother process for IT,
- A seamless experience for users, and
- The lowest possible cost for the organization.