Move Objects to LTFS Tape Using HTTP Web Service Interface

Matt Starr, CTO Spectra Logic
Jeff Braunstein, Developer Evangelist
Data created each year growth
40% annually

2009
1x1 = 800 Exabytes

2015
2.8x2.8 = 6.5 Zettabytes

2020
6.6x6.6 = 35 Zettabytes
Genomics

6 Cows = 1TB data

University of Washington, St. Louis
Video Data Explosion

- 8K UHD: 300 FPS
- 4K UHD: 120 FPS
- FHD: 72 FPS
- SD: 48 FPS
- 24 FPS
What Happens in an Internet Minute?

1,572,877 GB of global IP data transferred

- 10 Million ads displayed
- 347,222 Tweets
- 3.3 Million pieces of content shared
- 6.9 Million messages sent
- Netflix + Youtube = more than ½ of all traffic

- 438,801 Wiki page views
- $400 Million during Alibaba peak day sales
- 31,773 hours of music played
- 194,064 app downloads
- 34.7 Million instant messages (MIM) sent
- $133,436 in sales
- 57,870 page views
- 4,1 Million searches
- 100 hours of video uploaded
- 4.1 Million searches
- 138,889 hours of video watched
- 23,148 hours of video watched

And Future Growth is Staggering

By 2017, mobile traffic will have grown 13X in just 5 years
In 2017, there will be 3X more connected devices than people on Earth
All digital data created reached 4 zettabytes in 2013

Source: Intel
Digital System and their growth

- A $1000 Genome scan take 780 MB fully compressed
- A 2011 HiSeq-2000 scanner generate 20TB per month
- A typical security camera - 105GB of data per day.
  - City of Riverside, CA has 1000 cameras
- A digital camera is now 12MP or better
- A 4K video consumes 4.25 TB per hour
- Geospatial
  - An average ingest (rain) of 5 TB a day of new imagery.
  - 80 TB a day of refined product produced.
Storage Cost Per PB/Day over 5 yrs

OpEx per PB per day

CapEx per PB per day

US Dollars

Enterprise Production Disk

Enterprise Archive Disk

Midrange Tape

Enterprise Tape

Complexity in Deployment and use
What options, cloud?

**Traditional Storage**
- **Disk**
  - Easy to deploy
  - Expensive to run,
  - Expensive to buy
- **Tape**
  - Complex to deploy
  - Inexpensive to run
  - Inexpensive to buy

**Object and Cloud Storage**
- **Public Cloud**
  - Easy to deploy
  - Expensive to run
  - Little to buy
- **Private Cloud**
  - Complex to Deploy
  - Inexpensive to run
  - Moderate to buy
Pitfalls of Public Cloud ($/PB/Day)
Amazon

- Amazon apologizes for cloud outage and permanent data loss
  - ...After connectivity was restored, the issue of data corruption had to be addressed. By April 24th, the company had restored nearly 99% of their customers’ data at which point Amazon “began forensics on the remaining volumes.”

- Amazon has over 1 trillion objects!
- If they did lose 1% this is 10 billion objects
From Amazon

This is a notification that your volume vol-13c9647a experienced a failure due to multiple failures of the underlying hardware components and was unable to be recovered. We recommend recovering from your most recent snapshot.

We regret the loss and inconvenience.

Sincerely,
--- Dxxx Dxxxxxx
Nirvanix, VC pull the plug

- On September 16, 2013 Nirvanix notified customers that they had until September 30, 2013 to move their data off of the service at which point Nirvanix would shut down. By September 28, 2013 Nirvanix had closed its doors and shutdown all services.

- The primary lesson to learn is that you, as your organization's data custodian, can outsource the actual data to a cloud provider. But you, not the cloud provider, are ultimately responsible for your organization's data. - Howard Marks
Cloud easy to use, you don’t control your data

- Good points on Cloud Storage
  - Cloud ready available toolkits
  - Cloud storage is naturally WAN friend
  - Cloud storage sets the user expectation
    - It is put away

- Dangerous side of Cloud Storage
  - No proven SLA
  - Expensive data pipe
  - Costly get all your data back
What if...

S3/Object + Tape/Sequential Media Support + Removable Media Support w/ LTFS = Deep Simple Storage Service

Start with Object, S3 and Rest
Use Archive media as the target
Spectra S3 facilitates an Objected based Archive
What does BlackPearl and Spectra S3 do?

- DS3 Client Software
- BulkPut
- Objects Created
- Cache Drives
- FC SAS
- Tape Library

2015 Storage Developer Conference. © Spectra Logic. All Rights Reserved.
What Makes BlackPearl Revolutionary:

- **BlackPearl plus T950 with TS1150 drives**
  - 7.2 PB Usable
  - $0.09/GB

- **BlackPearl plus T950 with LTO6 drives**
  - 2.4 PB Usable
  - $0.16/GB

- **BlackPearl plus T380 with LTO6 drives**
  - 900 TB Usable
  - $0.23/GB
Advanced Bucket Management
Storage Cost Per PB/Day over 5 yrs

US Dollars

OpEx per PB per day
CapEx per PB per day

Complexity in Deployment and use

Enterprise Production Disk
Enterprise Archive Disk
Midrange Tape
Enterprise Tape
Data Movement Approaches with Spectra S3:

- Automated: Application or user built clients
- Semi-automated: NFI or scripting
- Manual: Deep Storage Browser or Command Line

```bash
#!/bin/sh
PATH=$PATH:/Users/matts/Downloads/ds3_java_cli-0.7.3-SNAPSHOT/bin
export PATH
export DS3_ACCESS_KEY=e3BiY3RyYQ==
export DS3_SECRET_KEY=Vz8m6cLy
export DS3_ENDPOINT=192.168.56.101:8080
DIR_TO_PUT="/Users/matts/archive/Assets"

d3_java_cli -http -c put_bucket -b cokemachine
d3_java_cli -http -c put_bulk -d $DIR_TO_PUT -b cokemachine
```
Questions

How to develop