It’s dangerous to go alone!

How about *we* do this!?
Questions!
Our needs
Sustainability
Recognition of the benefits of preservation by decision makers
Process for Selecting digital materials with long-term value
Incentive for decision makers to preserve in the public interest
appropriate *Organization* and *Governance* of digital preservation activities
mechanisms to secure an ongoing, efficient *Allocation of Resources* for digital preservation activities
timely actions to ensure access
Locally or Consortially?
Technology ??
Server Virtualization

Source: http://www.techcrates.com/use-of-virtual-servers/
Scale x Complexity > Skills

Source: http://www.theregister.co.uk/2012/03/26/opscode_funding_chef_momentum/
Configuration Management / DevOps
AUTOMATE

ALL THE THINGS!

Hardware commoditization / open source hardware
Intel-based x86 servers dominant in data centres for 10 years
Nothing particular to distinguish the products of different hardware vendors
Software processes, virtualization, configuration management all well defined
Hardware: Does it really matter anymore?

Moderated by Lawrence Dignan | January 9, 2012 -- 07:00 GMT (23:00 PST)

Summary: At the 2012 Consumer Electronics Show, hardware will take a backseat to software -- perhaps for good. Is hardware innovation alive and well?

---

It's alive and well

or

Hardware? Ho-hum

Best Argument: It's alive and well

---

Opening Statements

Hardware is what makes it possible

David Chernicoff: No matter how interesting software is, or how exciting the user experience becomes, without continued innovation in hardware technologies, software will begin to

Merely a matter of preference

Chris Dawson: As the world moves online, hardware has become nothing more than a religion. I tend to use Macs because I like the OS and the hardware takes the abuse I give it. On

---

Source: http://www.zdnet.com/debate/hardware-does-it-really-matter-anymore/10086653/
As Companies Focus on Services, Will Hardware Become Irrelevant?

Published: November 07, 2012 in Knowledge@Wharton

In recent weeks, Amazon.com, Microsoft and Apple have all held splashy events to tout their newest tablets -- the Kindle Fire HD, the Surface and the iPad Mini. But even as Amazon CEO Jeff Bezos talked up the specs of the new Kindle Fire, he noted that the market is becoming less about the gadgets themselves, and more about how companies can make money off additional services and purchases that appeal to device users.

Technology companies now see the buying of a tablet as merely the beginning of a long relationship with a consumer -- a relationship that will extend across multiple platforms and content types, with users periodically paying subscription fees or purchasing individual applications, music, movies and television shows. And if the firm is assured of a future payoff, it can offset the losses from charging less for the device.

But making such a transition is easier said than done, according to Wharton legal studies and business ethics professor Andrea Matwyshyn. "At present, this model is limited to only a few players," she says. "The services model definitely works for Amazon, given its content library with books and its expansion into [streaming] movies..."
Rethinking the hardware entirely: open source hardware
5 System View

Figure 1 System Overview

Software on commodity hardware can replicate costly vendor systems: Software defined storage
Vendor storage: cost vs. budget vs. growth
Software defined storage
## Storage Options & Cost

<table>
<thead>
<tr>
<th>Storage Type</th>
<th>Price/TB/Yr</th>
<th>Price/1.2PB/Yr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon S3 (1)(5)</td>
<td>$829.59</td>
<td>$995,512</td>
</tr>
<tr>
<td>Rackspace (2)(5)</td>
<td>$960.10</td>
<td>$1,152,123</td>
</tr>
<tr>
<td>Duracloud Preservation (3)(5)</td>
<td>$1,000.63</td>
<td>$1,200,750</td>
</tr>
<tr>
<td>Microsoft Azure (4)(5)</td>
<td>$829.59</td>
<td>$995,512</td>
</tr>
</tbody>
</table>

(2) [http://www.rackspace.com/cloud/files/pricing/](http://www.rackspace.com/cloud/files/pricing/)
(3) [http://www.duracloud.org/pricing](http://www.duracloud.org/pricing)
(5) USD on a Tiered “Pay As You Go” Plan, data is geographically redundant
Storage Node

Commodity Hardware

- Off the shelf consumer grade disks
- Software Replication (3X)
- Fixity & Self healing
- Geographically distributed
- Scalable… HIGHLY Scalable!
- Object Storage
Block vs. Object Storage

Block Storage

- Traditional storage
- SCSI or ATA or iSCSI or FC
- Access data via OS
- “C:” or “/dev/sda” or “Macintosh HD”
- Attached to one server at a time
Object Storage

- Access data via HTTP
- GET, POST, DELETE
- Build applications using REST API

```
swift -A https://oculcloud/auth/v1.0 -U system:root -K testpass upload thumbnails dataverse/logo.png
swift -A https://oculcloud/auth/v1.0 -U system:root -K testpass upload myfiles dataverse/head_right.png
```

- Static content (PDFs, jpgs)
- All files have a URL
Cyberduck Demo
## Storage Options & Cost

<table>
<thead>
<tr>
<th>Storage Type</th>
<th>Price/TB/Yr</th>
<th>Price/1.2PB/Yr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon S3 (1)(5)</td>
<td>$829.59</td>
<td>$995,512</td>
</tr>
<tr>
<td>Rackspace (2)(5)</td>
<td>$960.10</td>
<td>$1,152,123</td>
</tr>
<tr>
<td>Duracloud Preservation (3)(5)</td>
<td>$1,000.63</td>
<td>$1,200,750</td>
</tr>
<tr>
<td>Microsoft Azure (4)(5)</td>
<td>$829.59</td>
<td>$995,512</td>
</tr>
<tr>
<td>OCUL (6)</td>
<td>$350</td>
<td>$420,000</td>
</tr>
</tbody>
</table>

(2) [http://www.rackspace.com/cloud/files/pricing/](http://www.rackspace.com/cloud/files/pricing/)
(3) [http://www.duracloud.org/pricing](http://www.duracloud.org/pricing)
(5) USD on a Tiered "Pay As You Go" Plan, data is geographically redundant
(6) Includes cost of additional staff
So what?
We’re envisioning a solution that can help us out in a lot of different areas:

- Scholars Portal platforms
- Dataverse/other RDM initiatives
- Institutional Repositories
- New applications
- Just space in general!

with lots of different functional requirements:

- Autonomic digital preservation activities
- Lots of different interfaces
- Integration with HPC centers
Important because:

- Cost
- Jurisdiction
- Self-sufficiency
Which brings us to the point....
IT'S DANGEROUS TO GO ALONE! TAKE THIS.
IT'S DANGEROUS TO GO ALONE! TAKE THIS.
IT'S DANGEROUS TO GO ALONE! TAKE THIS.
IT'S DANGEROUS TO GO ALONE! TAKE THIS.
IT'S DANGEROUS TO GO ALONE! TAKE THIS.
If it’s worth doing here....
....it’s worth doing here!
Contact

Nick Ruest
ruestn@yorku.ca/@ruebot

Steve Marks
steve@scholarsportal.info/@sbmarks

Amaz Taufique
amaz@scholarsportal.info/@amazofficial

Graham Stewart graham.stewart@utoronto.ca