

# DuraCloud

Managing durable data in the cloud

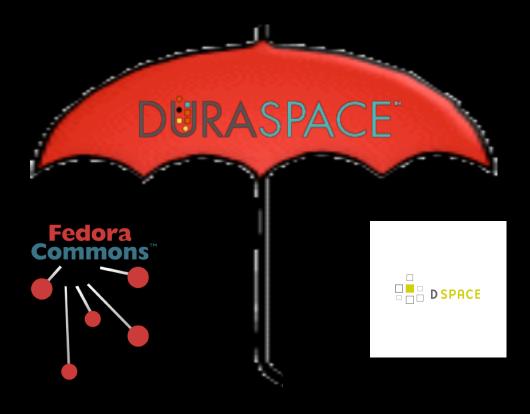


Michele Kimpton, Director DuraSpace





## Open Source Portfolio











#### Goals of DuraSpace

#### Stewardship:

Support and align open source development communities for DSpace and Fedora

#### Innovation:

Think beyond existing platforms

New strategies for enabling access and preservation of digital content

#### Sustainability:

Develop business model to sustain the nonprofit and open technologies we support





#### **Emergence of Infrastructure**

#### **Systems**





#### **Networks**



Integrate components
Central control
Dedicated/specialized gateways
More closed
More preconceived

Integrate systems
Distributed control
Generic gateways
More open
More reconfigurable

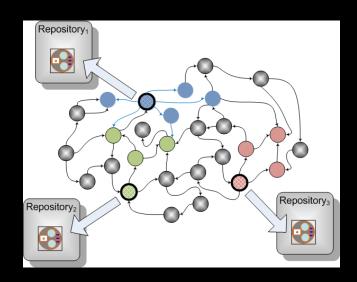


Source: Understanding Infrastructure: Lessons for New Scientific Infrastructure, http://deepblue.lib.umich.edu/handle/2027.42/49353



# Vision: Federated Repositories and Cyberinfrastructure

Heaven









### What About the Cloud?

A style of computing where massively scalable IT-related capabilities are provided "as a service" using Internet technologies to multiple external customers.

(Gartner, 6/08).







## **Cloud Services**

Elastic web-based infrastructure for storage and compute











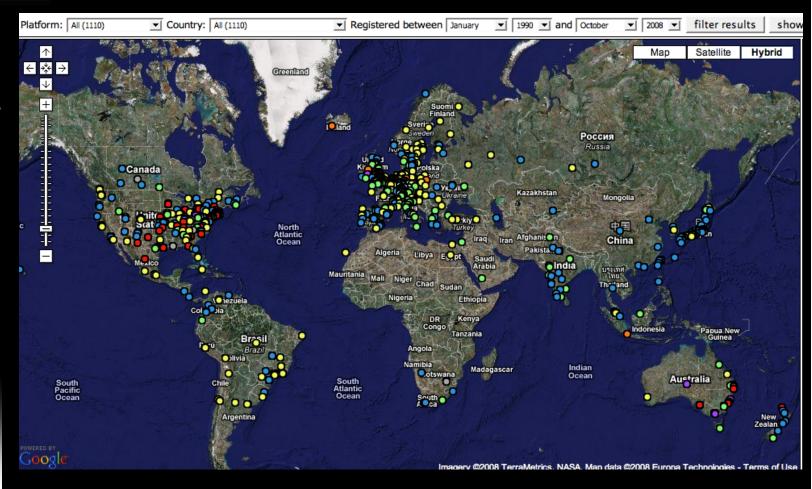






### What have we learned from our users?

Focus Groups
Site Visits





Over 750 organizations using DSpace or Fedora worldwide



## Challenge

Digital preservation is essential but difficult to implement

- Tools and processes unproven
- Limited IT support
- Resources unavailable
- Task can be overwhelming (replication, migration, emulation, etc.)





## Challenge

Barriers to making digital content more accessible and useful to researchers

- Systems not interoperable
- Heterogeneous applications/platforms
- Lack of commons standards
- Non-elastic compute capability





## Advantages – Cloud Services

- Flexibility
- Scalability
- Pay for use
- Easy to implement
- Cost





## **Economies of Scale and Cost**

# Public cloud providers drive cost down through scale, location and virtualization technology

Technology*	Cost Medium Datacenter	Cost Large Datacenter
Network	\$95 per Mbit/sec/mo	\$13 per Mbit/sec/mo
Storage	\$2.20 per Gbyte/mo	\$.40 per Gbyte/mo
Admin	140 servers/admin	>1000 servers/admin

Large Datacenters (tens of thousands of computers)
Medium Datacenters (thousands)



Source: Hamilton, Internet-Scale Service Efficiency,, LADIS Workshop (Sept 08)



### Issues

- Stability
- Transparency
- Data lock in
- SLA's
- Trust

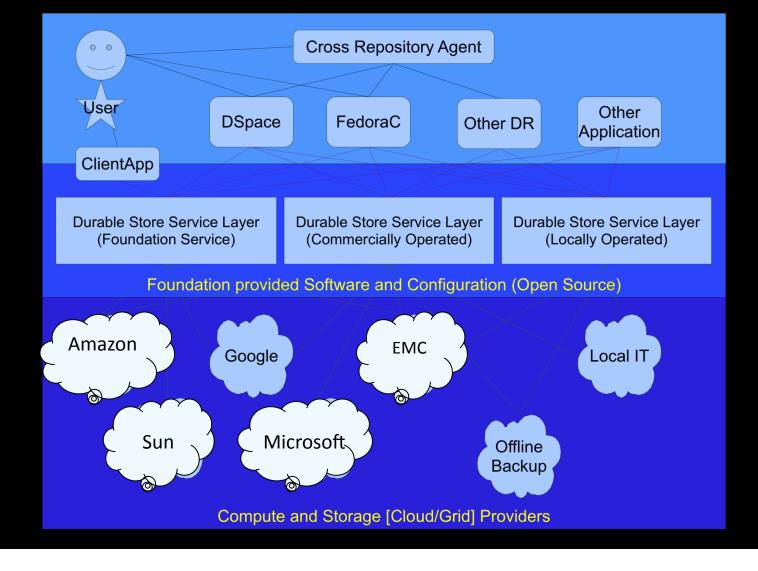




### DuraCloud

# Trusted management of and access to durable digital assets in the cloud

DuraSpace Mediating Service







## DuraCloud - basics

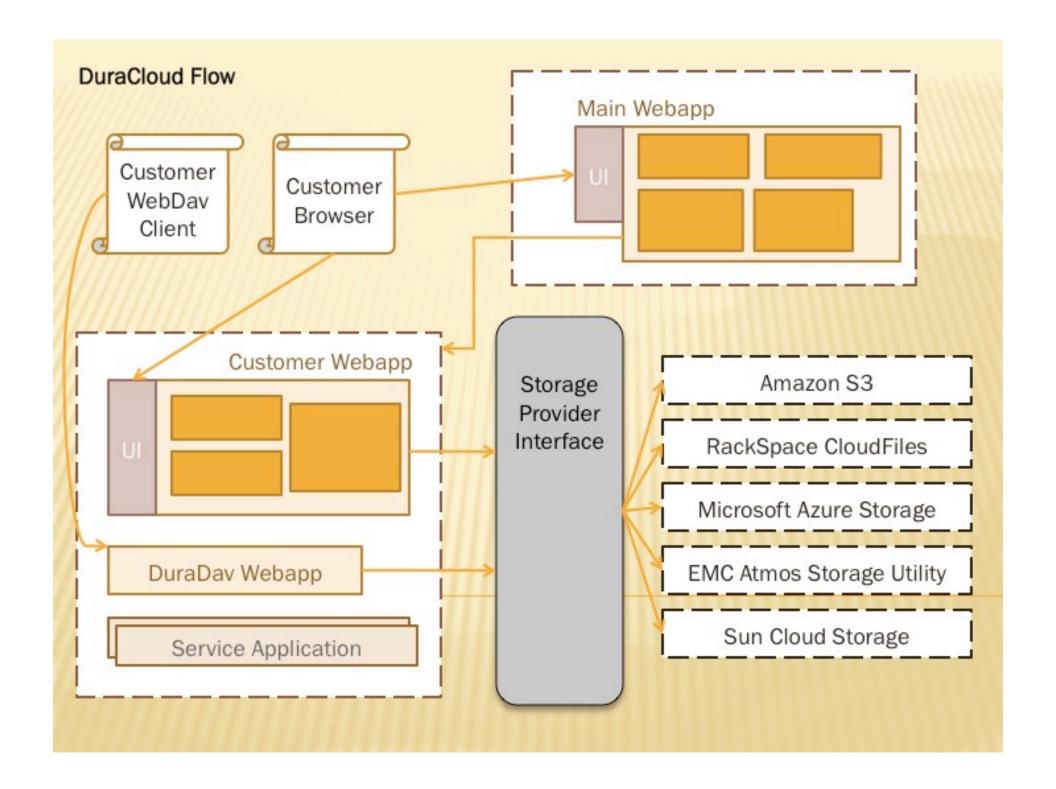


Chinese Menu of Service Options

Replicate to multiple storage providers
Replicate to multiple geographic areas
Monitor and audit digital assets
Compute services in cloud next to content



Hosted by DuraSpace not-for-profit org Partnerships with cloud providers "Pay for use" for services and storage Available to run internally- open source





## Additional services

- Other DuraSpace-provided services on top of content stored in the cloud
  - Search
  - Aggregation
  - Streaming
  - Migration
  - Hosting repositories







# Enable others to build and deploy services and apps in DuraCloud environment



Bloons
Digital Goldfish Ltd
Released Apr 05, 2009
\$0.99 BUY APP



Mover Infinite Labs Released May 11, 2009 Free GET APP



Ewe Stash - Knitting a Treeness, LLC Released May 11, 2009 \$2,99 (BUY APP)



Trixel
ADEPT Games, LLC
Released Apr 27, 2009
\$0.99 BUY APP



Spawn Illuminati Elements of Design Released Feb 23, 2009 \$1.99 BUY APP



World of Tunes Com2uS Corp. Released May 13, 2009 \$1.99 BUY APP



Angels&Demons
Sony Pictures Television
Released Apr 20, 2009
\$4,99 BUY APP



Gravulous SilentLogic Studios Released May 01, 2009 \$0.99 BUY APP



STAR TREK ®
Electronic Arts
Released May 05, 2009
\$2.99 BUY APP



Myst Cyan Worlds Released May 02, 2009 \$5.99 BUY APP



Vir2L® Drums
Vir2L Studios
Released Apr 17, 2009
\$0.99 BUY APP



Dane Cook Global Zannel, Inc. Released Apr 16, 2009 \$2.99 BUY APP

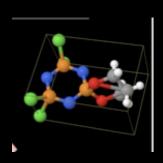




# Use Cases: DuraCloud with Cloud Storage

- Online backup for text, images, datasets, video, audio
- Enable preservation via multiple copies, geographies, administrations
- Elastic provisioning of temporary or permanent storage for projects or jobs











# Use Cases: DuraCloud with Cloud Compute

- Streaming service for video
- Hosting JPEG2000 image engine
- Indexing and other processing heavy jobs
- Repositories in cloud
- Data and text mining over open data
- Aggregation and web 2.0 tools on open content and collections





# DuraCloud Underlying software

- Open core
  - ✓ Core components available for others to build on and run
  - ✓ Open source apache license
- Architecture to create cloud networks
  - ✓ Public clouds
  - ✓ Private clouds
  - ✓ University consortia
- Also useful in research partnerships





### Critical success factors

- Ease of use simplicity
- Trusted partner within community
- Cost effective
- Elastic, scalable, flexible
- Establish key partnerships with cloud preferred cloud service providers
- Build community of developers and users





### Partners and Pilots

Selected initial cloud providers









Selected 2 initial pilot partners









### Pilot use cases

Ingest large quantity of material

Replicate to multiple cloud platforms

Manage replication and monitoring

Run services





#### Timeline

- Initial open source release
   – summer 2009
- Begin pilots September 2009
- Pilot data loading and testing Fall 2009
- Plug-ins for repository platforms Q4 2009
- Beta for repository community Q1 2010
- Pilot testing with compute services Q1 2010
- Report pilot results Q1 2010
- Launch production service Q2 2010





#### For more information:

#### DuraSpace Organization: <a href="http://duraspace.org">http://duraspace.org</a>

