

DuraCloud

Managing durable data in the cloud

The logo for DuraSpace, featuring the word "DURA" in a grey, sans-serif font above the word "SPACE" in a teal, sans-serif font. A small graphic of colored squares (orange, yellow, black, teal) is positioned to the left of the "A" in "DURA".

Michele Kimpton, Director DuraSpace

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 non-profit and open technologies we support

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



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 - basics

- 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



Chinese Menu
of Service Options



Additional services

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



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*



For more information:

DuraSpace Organization: <http://duraspace.org>

The screenshot shows the DuraSpace website homepage. At the top left is the DuraSpace logo. To its right is a navigation menu with links for HOME, PRESS RELEASE, FAQ, PEOPLE, and LOGOS. The main content area features a central announcement titled "Introducing DuraSpace™" with the subtitle "an organization creating open technologies for durable digital content." Below this, two paragraphs describe the merger of Fedora Commons and the DSpace Foundation. A right-hand sidebar contains sections for "Announcements" and "Projects". The "Announcements" section lists several recent events with dates. The "Projects" section lists DSpace, Fedora Commons, and Mulgara. At the bottom of the page, there is a large DuraSpace logo and a smaller logo for the DSpace and Fedora Commons partnership.

DURASPACE™ HOME PRESS RELEASE FAQ PEOPLE LOGOS

Introducing DuraSpace™

an organization creating open technologies for durable digital content.

Fedora Commons and the DSpace Foundation, two of the largest providers of open source software for managing and providing access to digital content, are joining their organizations to pursue a common mission. Jointly, they will provide *leadership and innovation* in open source technologies for global communities who manage, preserve, and provide access to digital content.

The joined organization, named "DuraSpace," will *sustain and grow* its flagship repository platforms - Fedora and DSpace. DuraSpace will also expand its portfolio by offering new technologies and services that respond to the dynamic environment of the Web and to new requirements from existing and future users. DuraSpace will focus on supporting existing communities and will also engage a larger and more diverse group of stakeholders in support of its not-for-profit mission.

Announcements

Available Now
DuraSpace Twitter Feed...

May 12th
DuraSpace Press Release Available...

May 19th
OR09 Developer Challenge Winner...

May 24-29th
DuraSpace Presence at Sun PASIG...

June 18th
JumpBox for DSpace...

Projects

DSpace

Fedora Commons

Mulgara

DURASPACE™

DSpace | Fedora Commons

