Thank you

for more information

DuraSpace organization
  • http://duraspace.org

DuraCloud project
  • http://duracloud.org

awoods@duraspace.org
Preservation Infrastructure in the Cloud

CNI, April 2010

Andrew Woods
awoods@duraspace.org
Outline

- what is DuraCloud?
- pilot partners
- architecture
- quick look
- roadmap
Organizational Mission

- promoting durable, persistent access to digital data
- creating practical solutions in collaboration with academic, scientific, cultural, and technological communities
- development/deployment of open technologies
What is it?

- cloud-based service offered by the not for profit organization, DuraSpace
- cloud storage across multiple commercial & non-commercial providers
- an open source, cloud storage/compute application...focused on preservation support and...data access for reuse and sharing
- an open canvas for cloud-based services
Key notions

- mediated cloud storage
- ready to use, value-added services
- dead simple
cloud-based services
Initial Pilot Partners

BHL

NYPL

WGBH
Use Cases

- geographically distributed online backup
- bulk image conversion
- jpeg2000 serving/viewing
- international replication
- streaming video
- access to corpus for processing
- ingest staging area
Lessons learned

- bulk ingest: chunking / stitching
- bulk ingest: parallel upload
- bulk ingest: hard disk process

- bulk image conversion

- de-mystifying third party hosted content

- asynchronous nature of cloud storage
sign up → launch → primary instance
http://demo.duracloud.org:8080/duradmin
REST-API examples

- **get contents of space**
  
  http://your-domain.edu/durastore/images

- **get content**
  
  http://your-domain.edu/durastore/images/rome.jpg

- **deploy service**
  
  http://your-domain.edu/durasync/services/image-conversion-1.0.0
primary instance
Security
underlying storage provider (e.g. S3) access control
DuraCloud
application security

underlying storage provider (e.g. S3) access control
channel security
(apache ssl/tls)
instance firewall
Roadmap

- series of internal releases through Fall '10
  - repository integration (DSpace & Fedora)
  - expanded storage/compute provider support
  - expanded service selection

- open sourcing Summer '10

- second small set of pilot partners

- extended pilots Fall '10

- public beta release early '11
Roadmap

- series of internal releases through Fall '10
  - repository integration (DSpace & Fedora)
  - expanded storage/compute provider support
  - expanded service selection

- open sourcing Summer '10

- second small set of pilot partners

- extended pilots Fall '10

- public beta release early '11

Upcoming service developments focused on preservation support and access

- indexing / search
- bit integrity assurance
- expanded replication (on demand, filtered, etc)
- video streaming
- auditing
Roadmap

- series of internal releases through Fall '10
  - repository integration (DSpace & Fedora)
  - expanded storage/compute provider support
  - expanded service selection

- open sourcing Summer '10

- second small set of pilot partners

- extended pilots Fall '10

- public beta release early '11
Thank you

for more information

DuraSpace organization
  • http://duraspace.org

DuraCloud project
  • http://duracloud.org

awoods@duraspace.org
Roadmap

- series of internal releases through Fall '10
  - repository integration (DSpace & Fedora)
  - expanded storage/compute provider support
  - expanded service selection

- open sourcing Summer '10

- second small set of pilot partners

- extended pilots Fall '10

- public beta release early '11