



DuraCloud

Enabling services for managing data in the cloud



Michele Kimpton, CBO DuraSpace

Bill Branan, Senior Developer DuraSpace



Agenda

What problem are we trying to solve

What is DuraCloud

Pilot

Timeline

Challenges (From our communities)

*Digital preservation and archiving is hard to achieve , even just **basic replication***

***Easy** and **elastic** provisioning of shared infrastructure (also across institutions!)*

*Robust compute environments for large indexing jobs, **data mining** and **analysis** of large datasets*

*Making digital **content** more **accessible** and **useable** to researchers*

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).



More definitions... (UC Berkeley RAD Lab)

Public Cloud

The service being sold is “Utility Computing”

Private Cloud

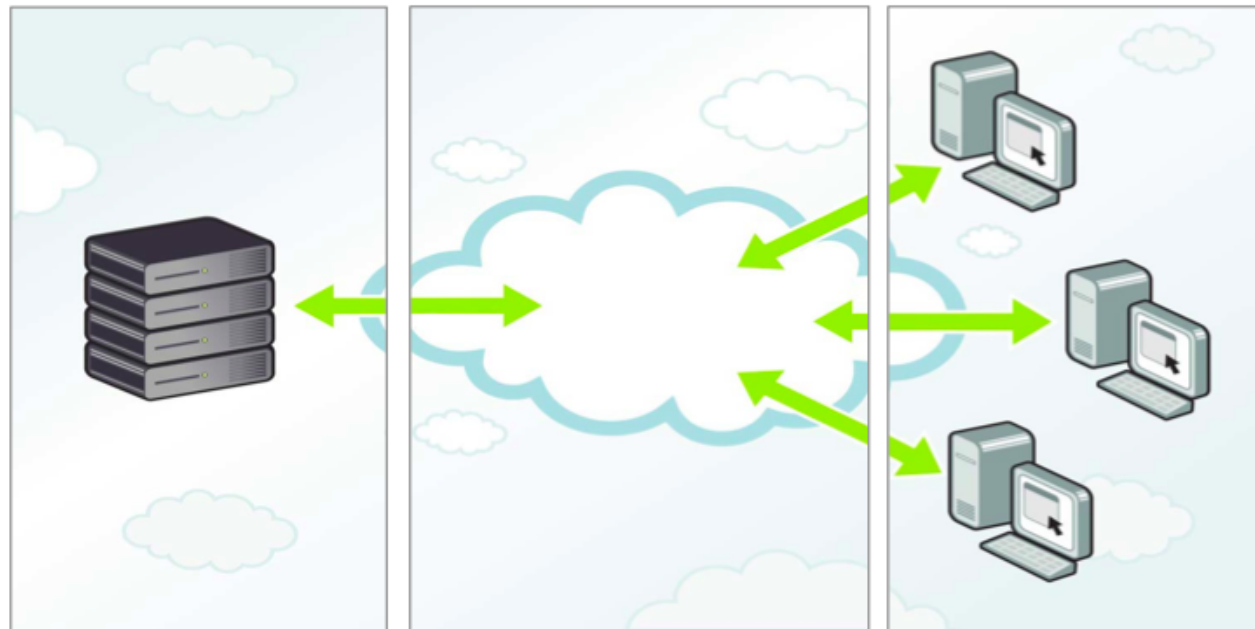
Internal datacenters of a business or other organization not made available to the general public

- ## Community Clouds

Networks of private clouds, available within a given community

Cloud services

- Software as a Service
- Infrastructure as a Service
- Platform as a Service



Public Cloud Services

Elastic web-based infrastructure for storage and compute



Focus group study

Academic CIO's and Library IT

- Institutions' perceptions of current compute and storage costs
- Current market understanding, willingness to consider cloud computing
- DuraSpace value
- Need for DuraCloud services

Cloud Computing Today

- All participants possess some understanding of cloud compute and storage
- No institution using external cloud provider to manage or store collections today
- Majority of institutions think it is likely they will use cloud computing in some capacity in the next twelve months

Cloud Computing Barriers

- Long-term trustworthiness and sustainability of solution
- Access and reliability
- Data security and privacy concerns

DuraSpace Value Proposition

“The community they have historically served brings a sense of trust and commitment to the market. Nothing is as powerful in higher education as strong references from the higher education community.”

“We would work with DuraSpace over Sun or Google any day in this market. Given the size and the strength of the education sector overall (small relative to their other markets), we can never count on the big players to not cut the services that are key to us.”

“We view DuraSpace as part of our community, not just a vendor. We have different roles, but we are all working on behalf of the same things.”

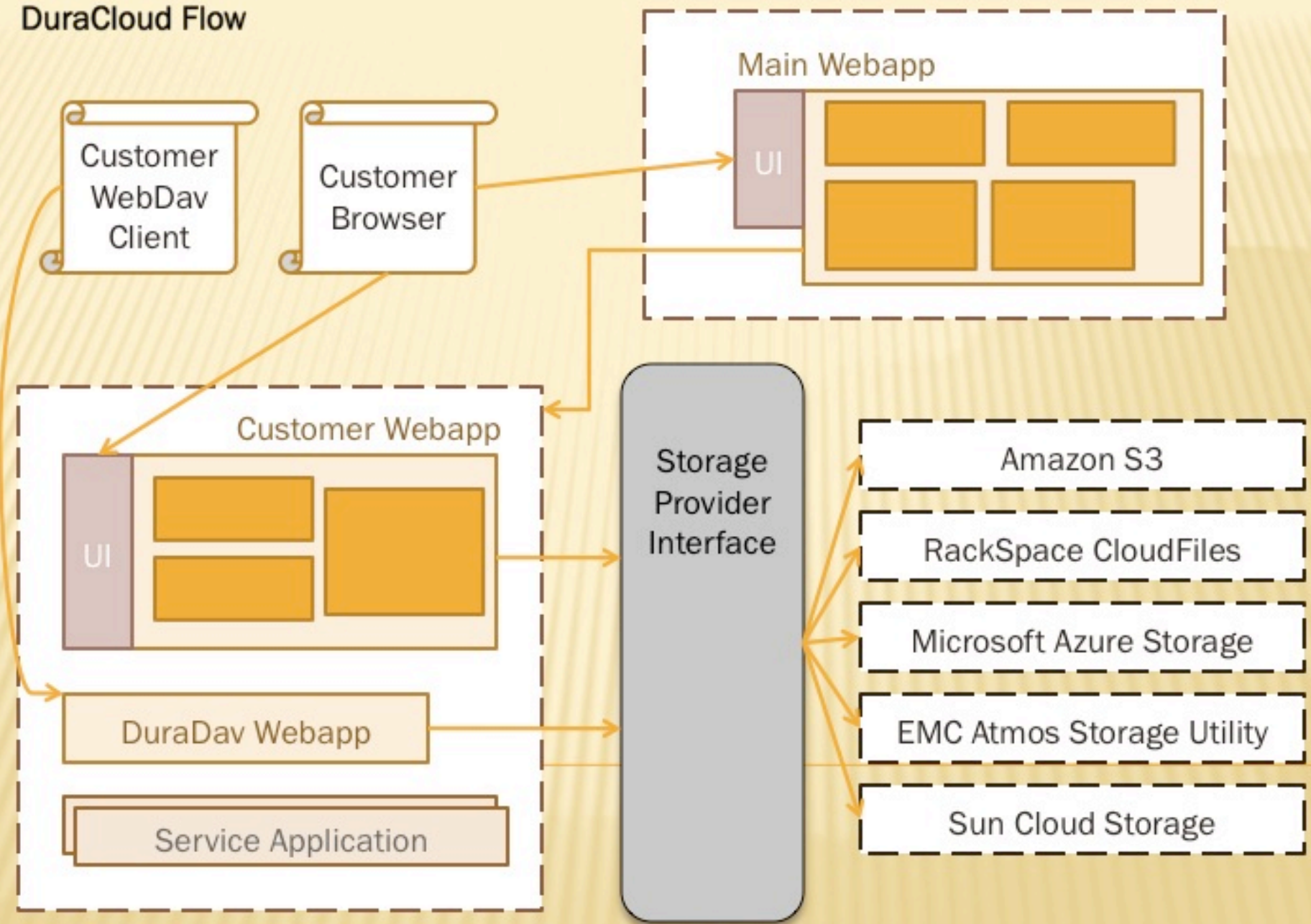
DuraCloud Proposition

Trust and durability in the cloud

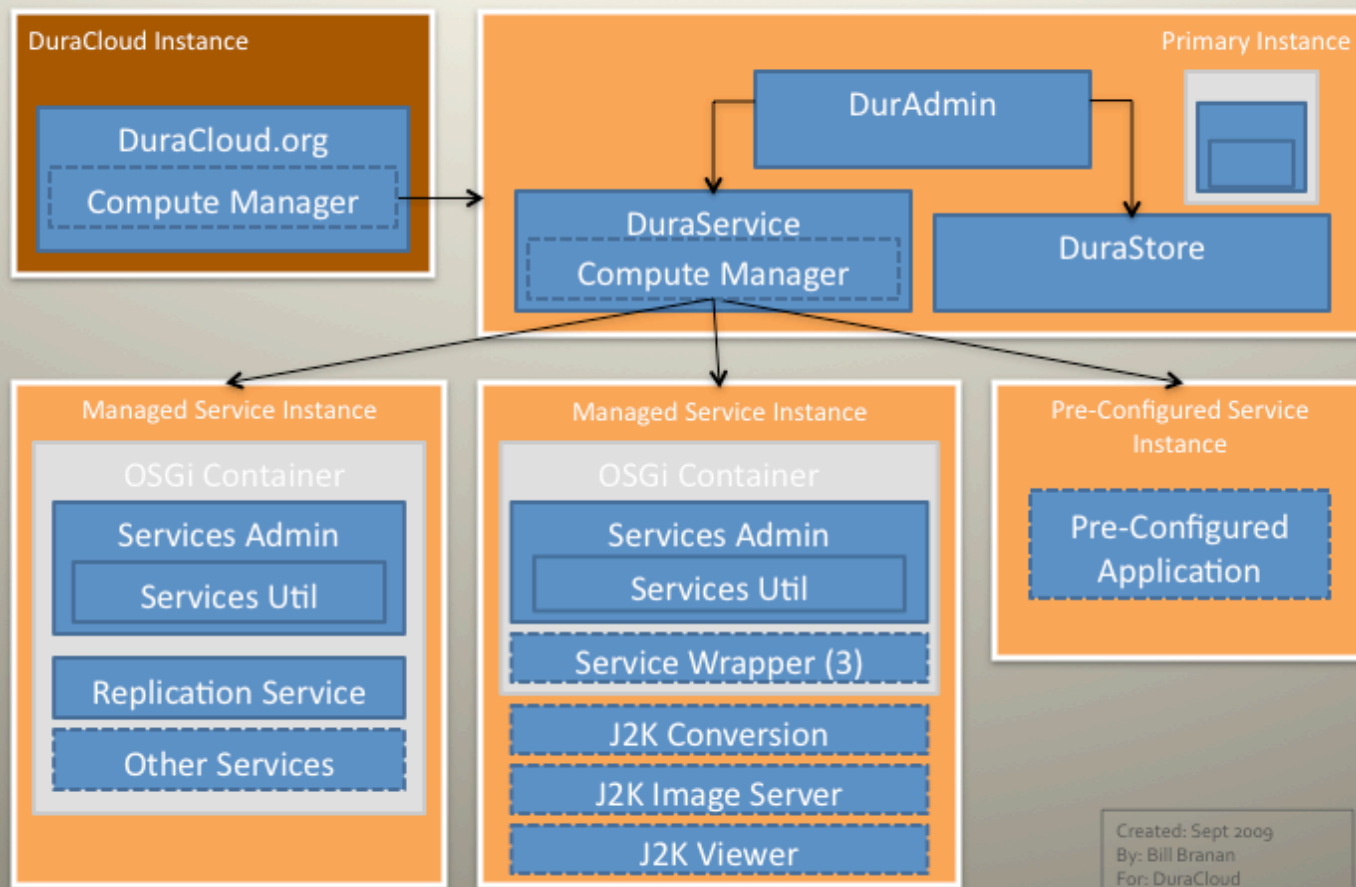
DuraCloud is a service aimed at supporting libraries, universities, and other cultural heritage organizations that wish to provide perpetual access to their digital content. The service replicates and distributes content across multiple cloud providers and enables the deployment of services to support:

- * access
- * preservation
- * re-use

DuraCloud Flow

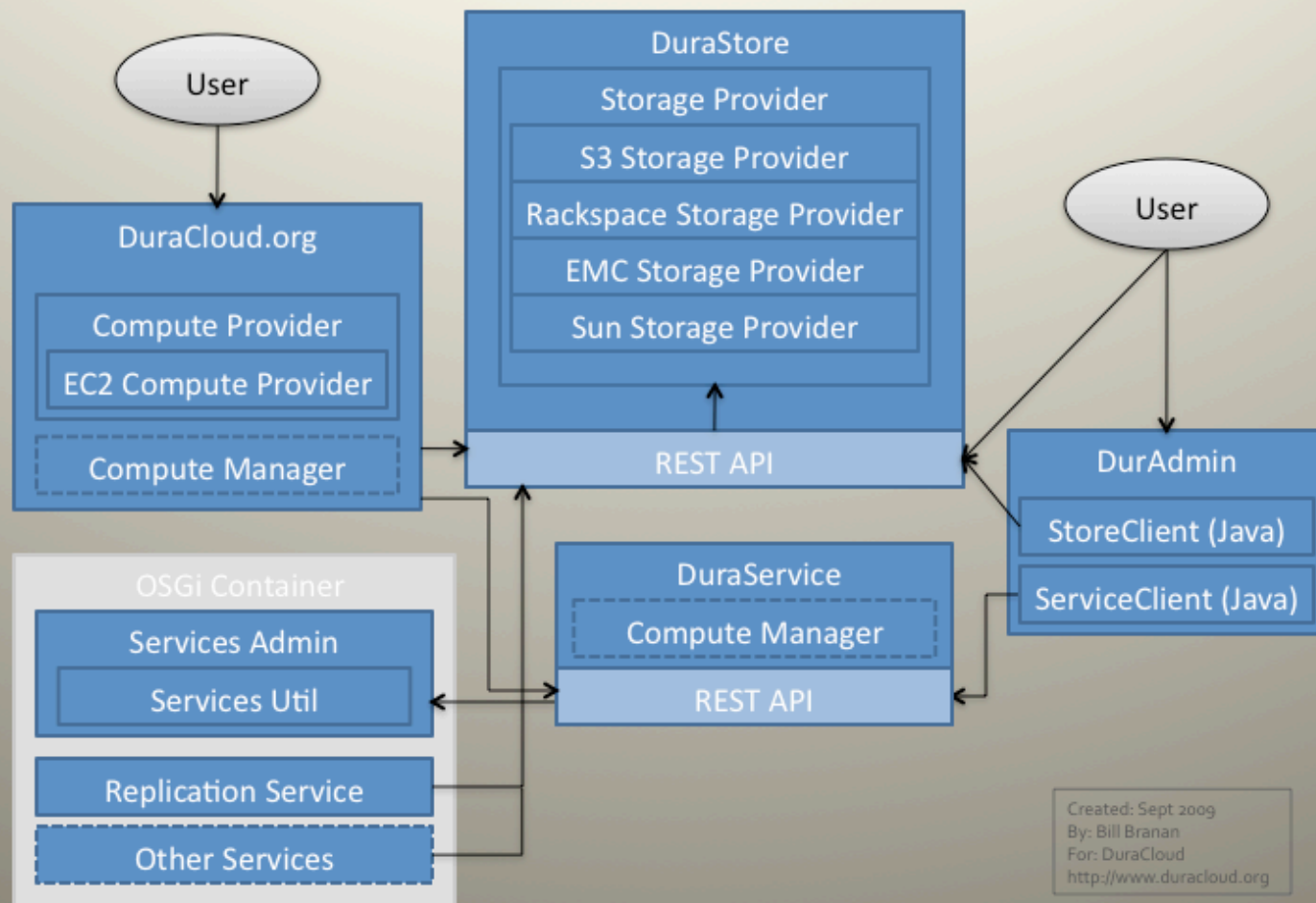


DuraCloud Preliminary Server Architecture



Created: Sept 2009
By: Bill Branan
For: DuraCloud
<http://www.duracloud.org>

DuraCloud Preliminary Project Architecture



Preservation Services

- Online backup in the cloud*
- ability to replicate content to multiple providers and locations*
- ability to synchronize backup with primary store or repository system*
- management ,monitoring, audit and repair through web based interface*

Hosted by DuraSpace not-for-profit org
Partnerships with cloud providers








Access and compute services



Application Exchange

Native Apps | Most Popular | New | Staff Picks

Secure, Reliable, Easy to use! Native apps are built and run 100% on Force.com's trusted cloud infrastructure.

	Infowelders Revenue Heat Map	Infowelders X Native
	Callidus Plan Communicator	Callidus Software Inc. X Native
	Jobsience For iPhone	Jobsience On-Demand X Native
	PaymentConnect: Native Credit Card Processing, Order Mgt, Recurring Payments	Linvio, Inc. X Native
	Email Template for iPhone and BlackBerry	Force.com Labs X Native
	Opportunity Assignment Guide V4.1.5	Vandesys X Native
	studentforce	studentforce.com LLC X Native

What DuraCloud is not...

Repository platform

Hosting service

Central archive or library

You own and manage your data, we enable technology and services utilizing the cloud

Partners and Pilots

- Selected initial cloud providers



- Selected 2 initial pilot partners





Cloud partnerships

- Preferred pricing
- Co-Development of services
- Free storage for open data pool
- Early notification of architecture/API changes
- Immediate notification of security breach
- Possible enhanced SLA's-data loss?
- Sponsorship of non-profit


NYPL pilot

Digital Gallery Collection

Explore NYPL's new online exhibition: [Treasures of the New York Public Library](#)

The New York Public Library

Digital Gallery



Lispenard Street between Church Street and Broadway ... Source: Drugstore Photographs, Or, A Trip Along the Yangtze River, 1999, by Dylan Stone. Digital ID: 518688.


NYPL Digital Gallery provides free and open access to over 700,000 images digitized from the The New York Public Library's vast collections, including illuminated manuscripts, historical maps, vintage posters, rare prints, photographs and more.

SEARCH [\[More Options\]](#)

BROWSE

- Subjects
- Names
- Library Divisions
- My Selections
- Arts & Literature
- Cities & Buildings
- Culture & Society
- History & Geography
- Industry & Technology
- Nature & Science
- Printing & Graphics

DIGITAL GALLERY PICKS



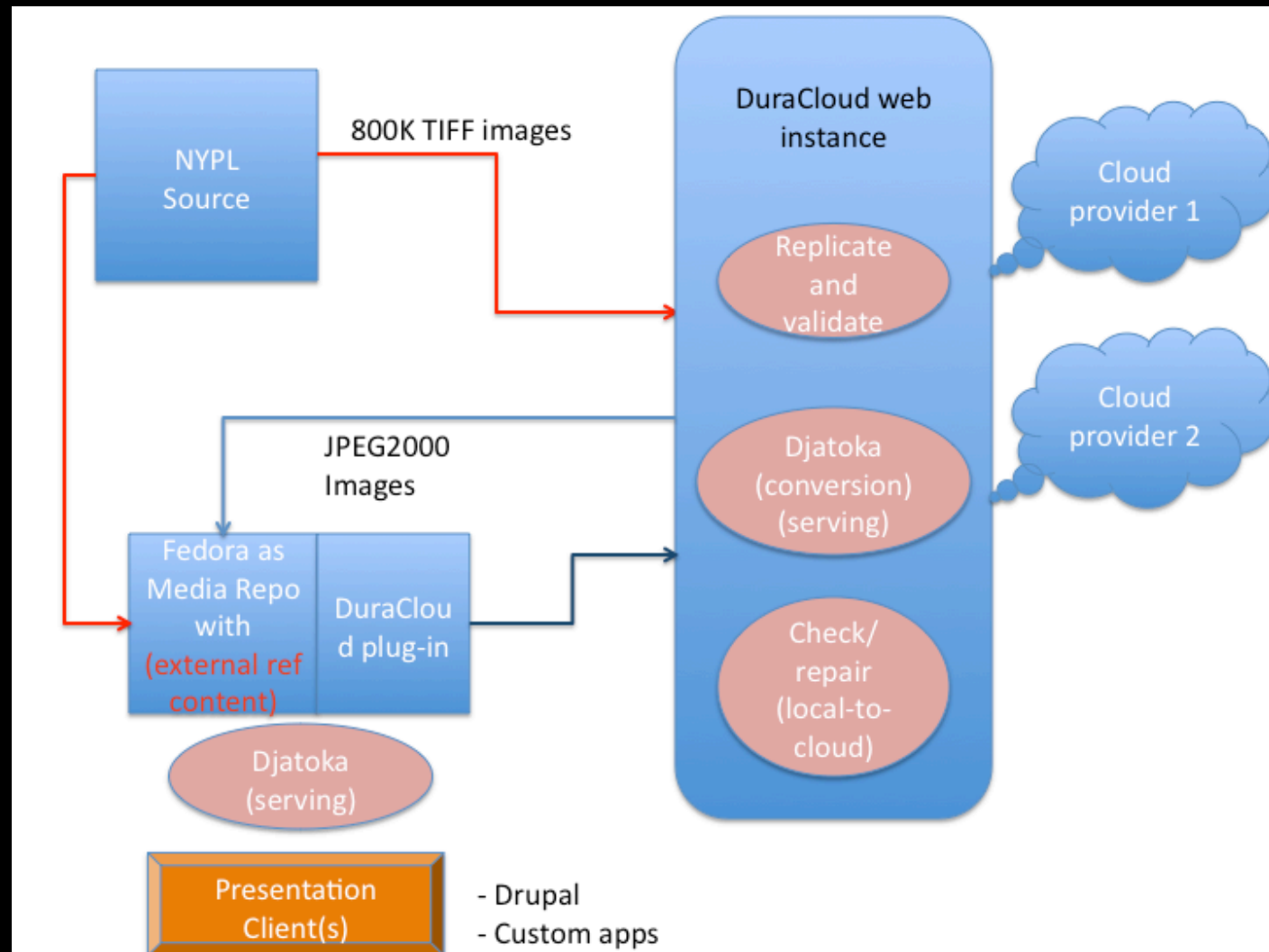
-back up copy 800k images (50 TB data)

-transformation from Tiff to JPEG 2000

-run image server in cloud

-Push JPEG 2000 back into Fedora Repository

NYPL workflow



BHL pilot

BioDiversity Heritage Library

Biodiversity Heritage Library
News and Updates from the Biodiversity Heritage Library
available online at <http://www.biodiversitylibrary.org>

Over 250 years of biodiversity literature to support the work of scientists, researchers, students and you.

BHL RECENT ADDITIONS

L'illustration horticole : v. 20 1873 (added: 09/11/2009) - 09/11/2009

L'illustration horticole : v. 19 1872 (added: 09/11/2009) - 09/11/2009

The Floricultural Cabinet and Florist's Magazine. VOL. 1 (added: 09/10/2009) - 09/10/2009

Voyage de La Pérouse autour du monde / Atlas (added: 09/10/2009) - 09/10/2009

TUESDAY, AUGUST 25, 2009

Book of the Week: BHL, EOL, and Marine Life

The dream of making biodiversity information freely available to people around the world is an ambitious goal embraced by many in the scientific community, and it represents the mission behind both the Biodiversity Heritage Library and the Encyclopedia of Life (www.eol.com), of which BHL is a cornerstone institution. While the Biodiversity Heritage Library strives to digitize the published literature

ABOUT BHL

Twelve major natural history museum libraries, botanical libraries, and research institutions have joined to form the Biodiversity Heritage Library. The participating libraries have over two million volumes of biodiversity literature collected over 200 years to support the work of scientists, researchers, and students in their home institutions and throughout the world.

LABELS

News (10)

-back up copy entire corpus (40 TB data) from multiple sources

-have multiple copies including Europe

-Do compute intensive data mining over corpus

Pilot use cases

NYPL

Replication and preservation support

Format conversion

Instant provisioning

Synchronization with repository

BHL

- Replication and preservation support
- International collaborative infrastructure
- Researcher platform for data mining

Timeline

- Begin pilots(MOU's in place) – September 2009
- DuraCloud Alpha Pilot release- Oct 2009
- Pilot data loading and testing – Fall 2009
- Beta for repository community - Q1 2010
- Pilot testing with software services Q1 2010
- Cloud partner evaluations complete-Q2 2010
- Strategic cloud partnerships in place- Q2 2010
- Pricing Model determined-Q2 2010
- Report pilot results – Q2 2010
- Launch production service Q3 2010

Pilot Success

- Can replicate content across 2 or more cloud providers through web interface
- Can manage and check and repair content through web interface
- Can perform at least one service on content
- Migrated small, medium and large datasets into cloud (up to 40 TB)
- Established pricing strategy
- Have integrated both DSpace and Fedora
- Have multiple strategic cloud partnerships in place
- Launch service mid 2010

Core Team

Project Director- myself

Senior Developer- Bill Branan

Senior Developer-Andrew Woods

Web Developer- Danny Bernstein

Technical oversight- Brad McLean

Fedora integration- Chris Wilper

DSpace integration- Tim Donohue

Marketing communications-Carol Minton
Morris

Thank You



For more information:

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

Wiki: <http://www.fedora-commons.org/>

confluence/display/duracloudpilot/

Mkimpton@duraspace.org