



DuraCloud

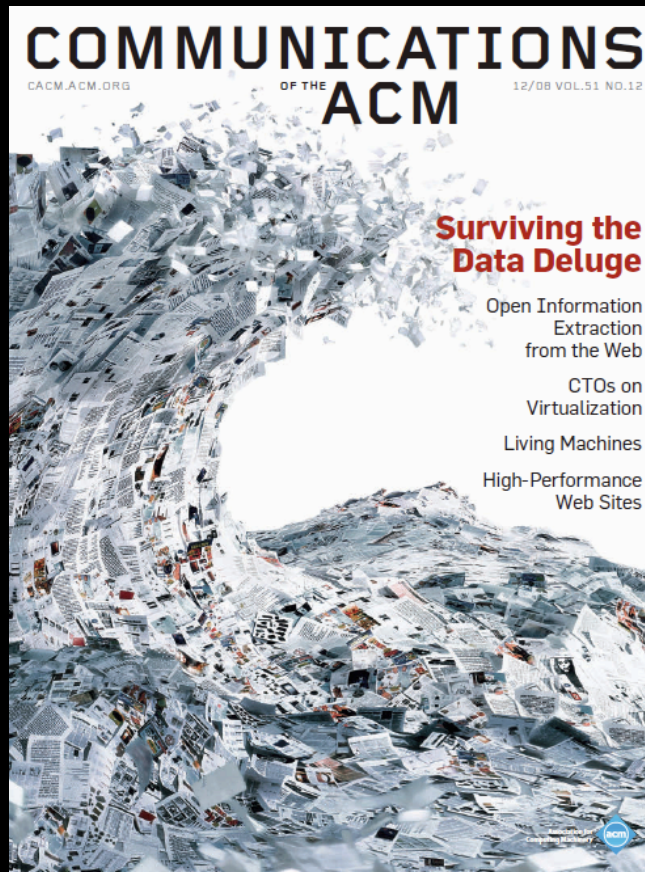
A service provided by



Sandy Payette and Michele Kimpton



December 2008



Flickr photos.). Emerging data CI and management environments and systems, including IRODS (www.irods.org/), LOCKSS (www.lockss.org/lockss/Home), the Fedora Commons (www.fedora-commons.org/), and D-Space (www.dspace.org/), are beginning to develop and incorporate mechanisms that implement relevant policies and procedures. Over the next decade, the

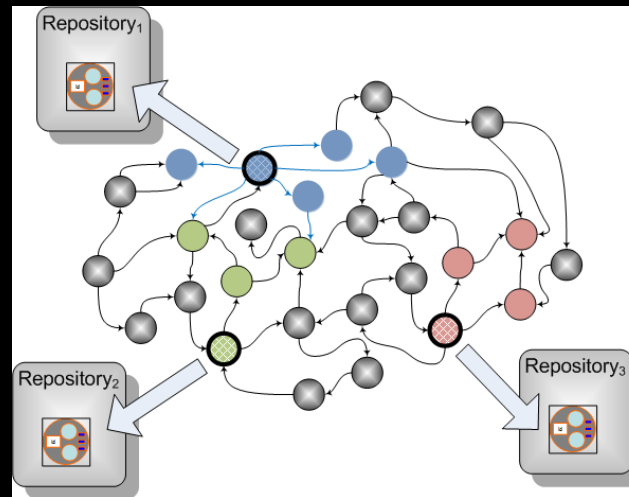
page 53

The commercialization of data storage and services contributes an important component of the data CI environment needed to harness the potential of our information-rich world. However, private-sector storage and services are not the solution to all digital data needs. For some digital data considered to be "in the public interest" (such as census data, official records, critical scientific data collections, and a variety of irreplaceable data), a greater level of trust, monitoring, replication, and accountability is required to minimize the likelihood of loss or damage and ensure the data will be there for a very long time. For such community data sets, stewardship by a trusted entity (such as libraries, archives, museums, universities, and institutional repositories), whose mission is the public good rather than profit, is generally required.

page 55

Vision: Federated Repositories and Cyberinfrastructure

Heaven



DuraCloud

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).



Berkeley Definition of Cloud

“The services themselves have long been referred to as **Software as a Service (SaaS)**. The datacenter hardware and software is what we will call a **Cloud**. When a Cloud is made available in a **pay-as-you-go manner** to the general public, we call it a **Public Cloud**; the service being sold is **Utility Computing**.

We use the term **Private Cloud** to refer to internal datacenters of a business or other organization, not made available to the general public.”

Examples of Cloud Services

- **Software as a Service (SAAS)**
 - e.g. , Google Apps
- **Cloud Computing**
 - e.g., Amazon Elastic Compute Cloud (EC2)
- **Cloud Storage**
 - e.g., Amazon Simple Storage Service (S3)

Cloud Services

Elastic web-based infrastructure for storage and compute

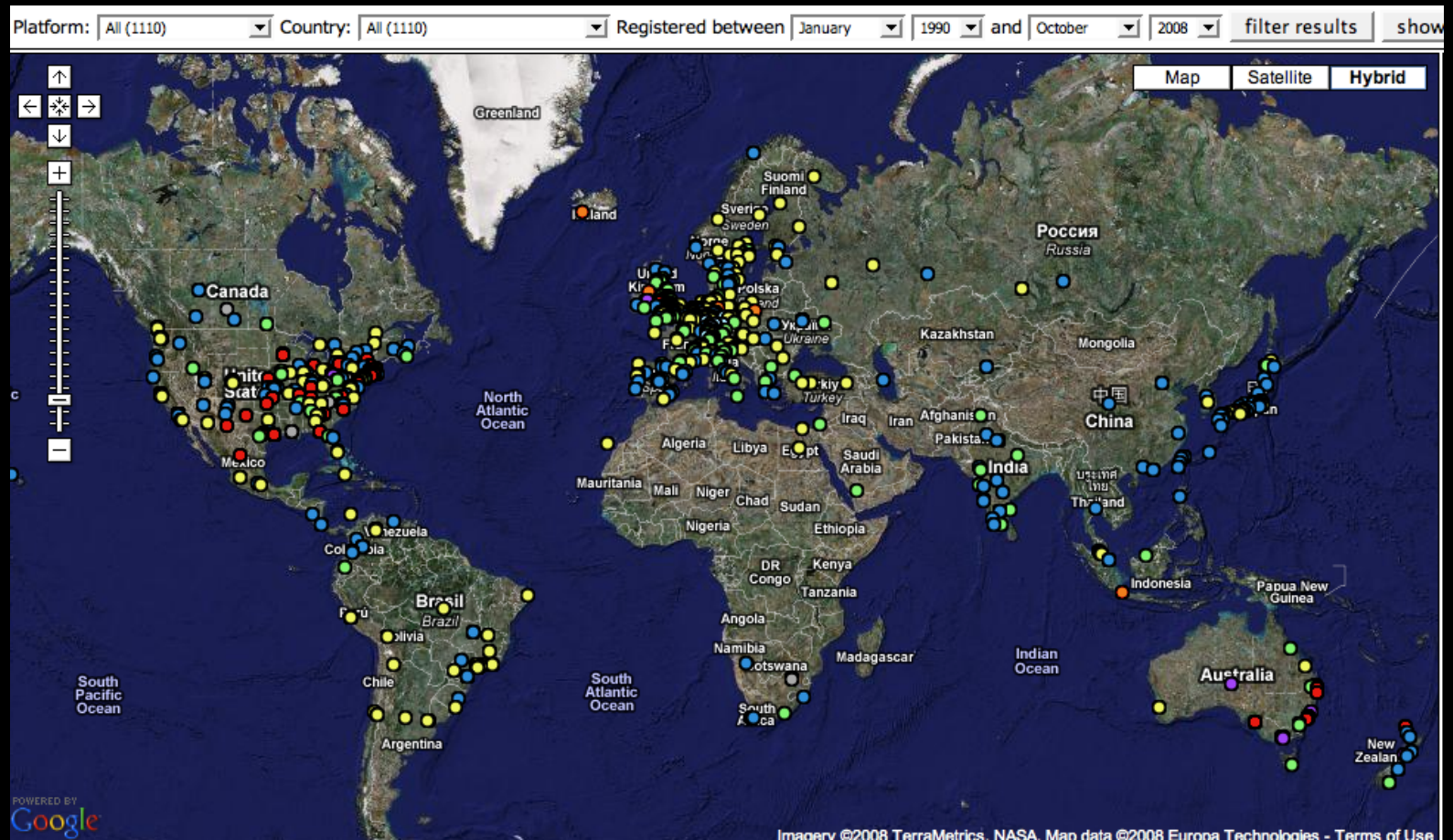


DuraCloud Proposition

Trust and durability in the cloud

What have we learned from our users?

Focus Groups
Site Visits
Forums



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*



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

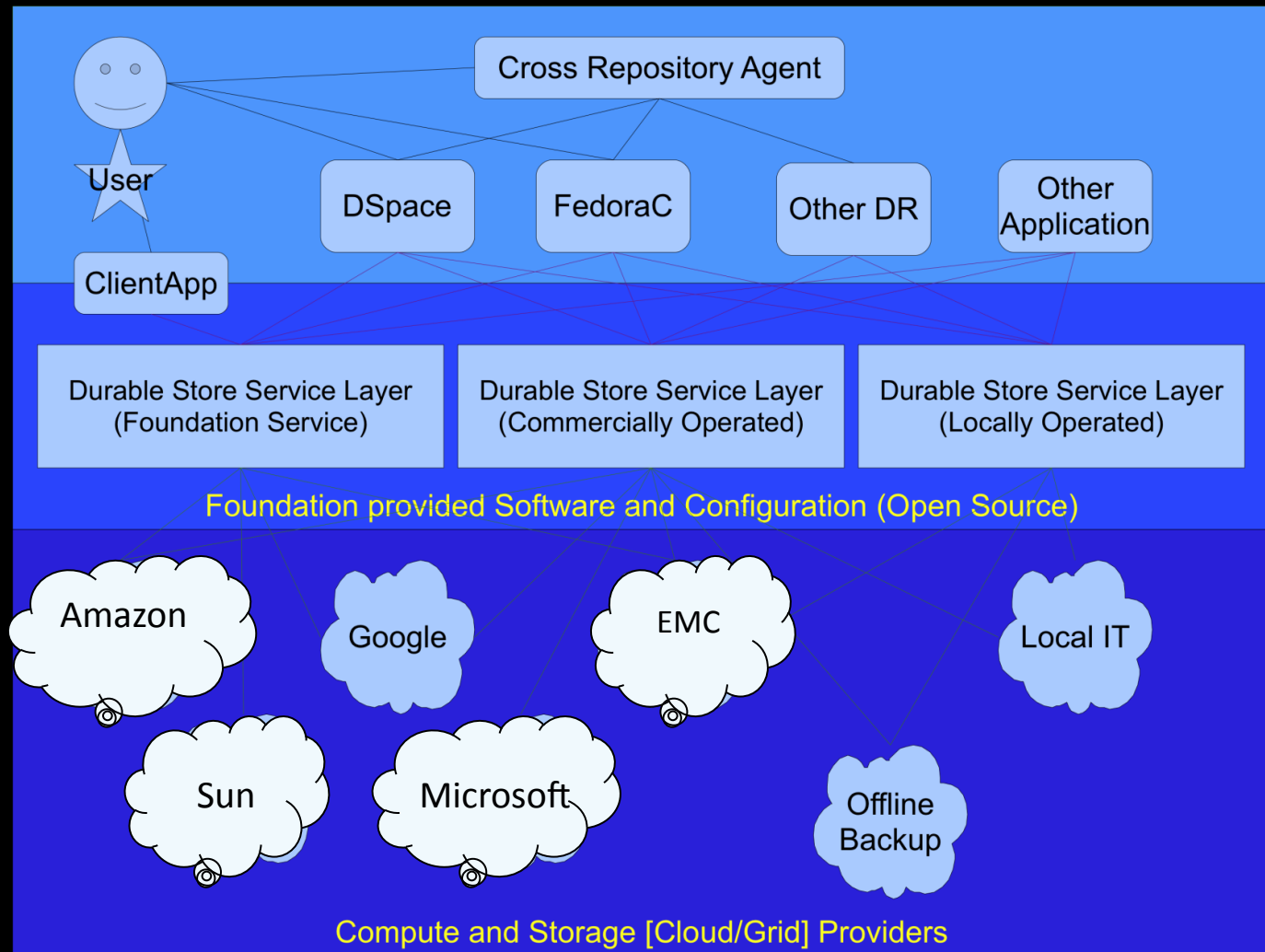


Chinese Menu
of Service Options



DuraCloud

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

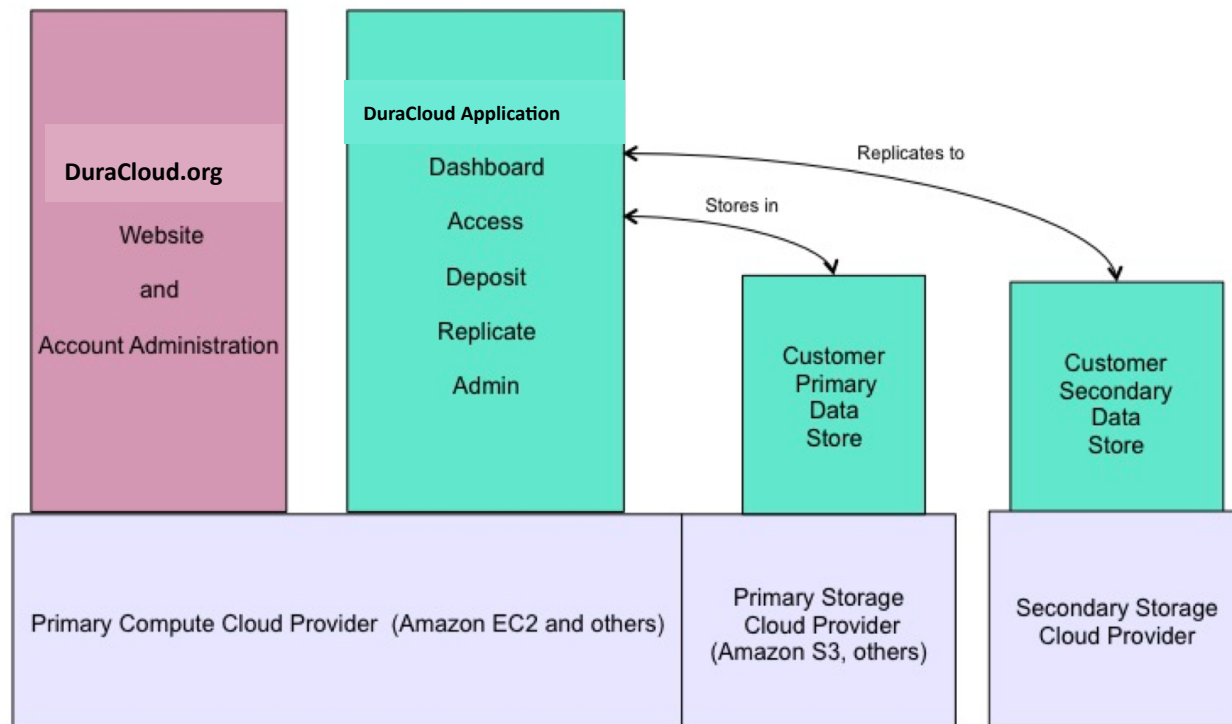


DuraSpace
Mediating
Service

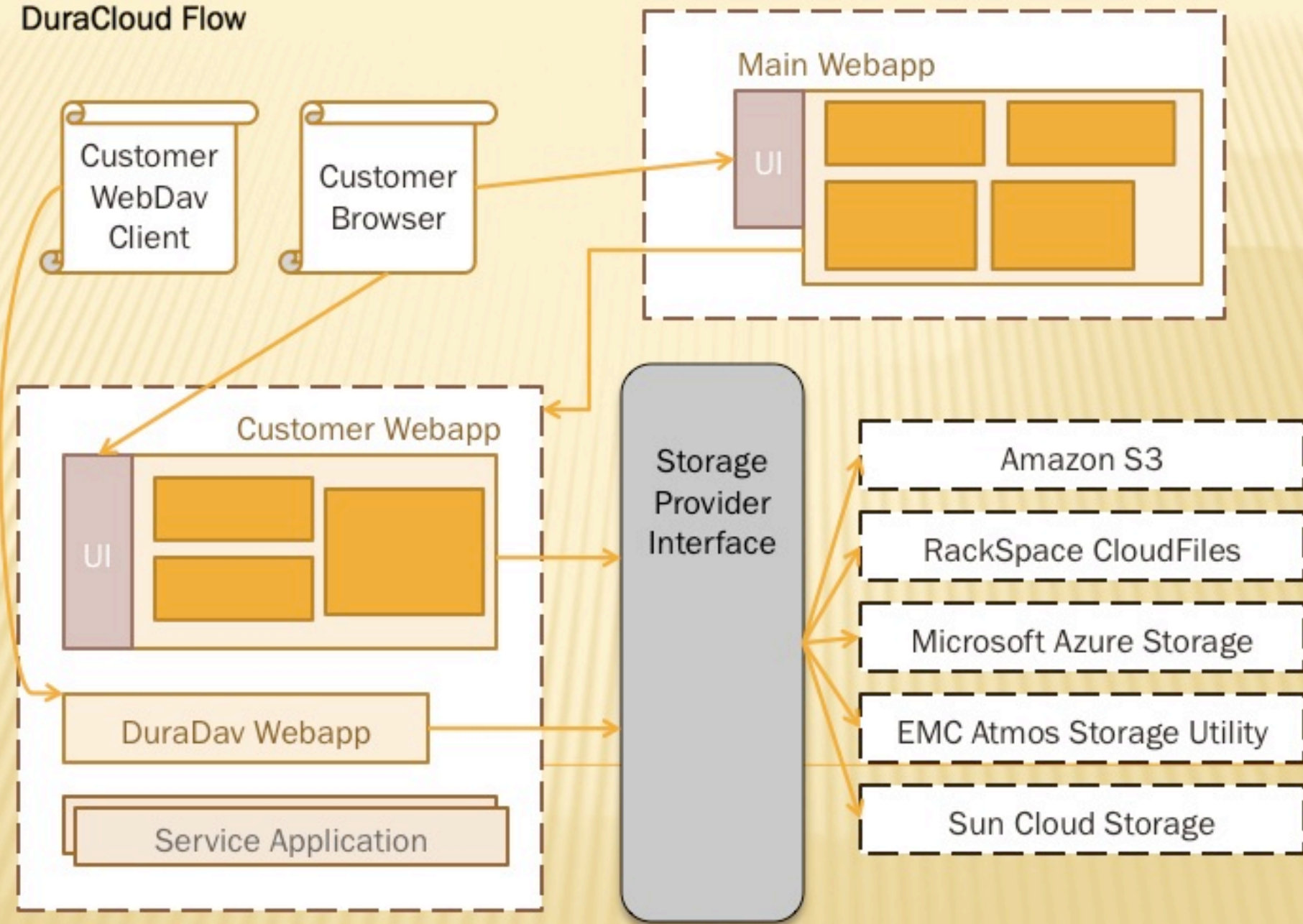
DuraCloud

Basic Architecture

DuraCloud – Notional View



DuraCloud Flow



DuraCloud

Making the Cloud Durable

Use Cases

Partnerships and Pilots

Challenge

Digital preservation is essential but difficult to implement

- Tools and processes unproven
- Limited IT support
- Capital expenditures limited
- 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

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

Initial capabilities













- Replication, up to three providers (including local store)
- Web based “Dashboard”
- Data integrity checking and monitoring
- Can push content from DSpace and Fedora repository platforms via plug-ins
- Pay per use
- Initial compute services on content

Additional services

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

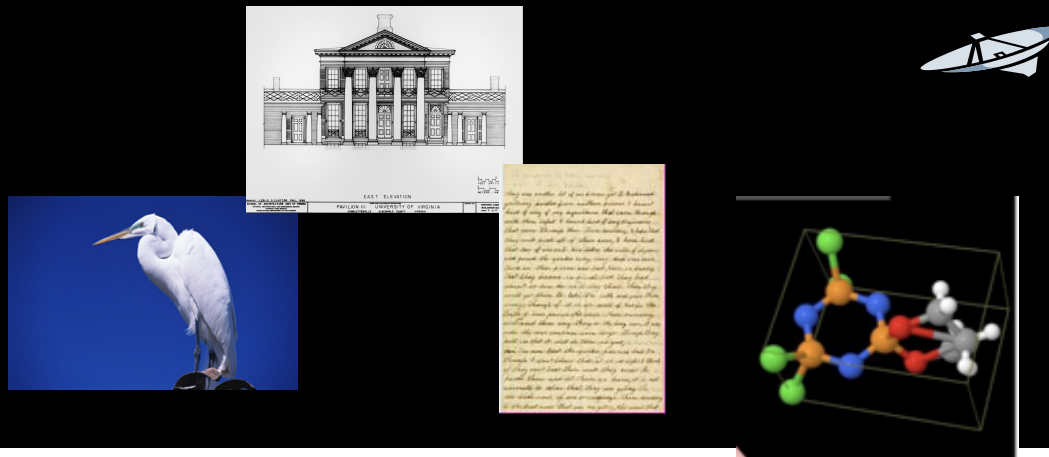
<h1>老四川</h1>		<h2>Supps</h2>		<h2>Hot Rice & Noodles</h2>	
Original 1 Street Sichuan Restaurant		361 Baby Bone Soup	\$1.75	401 Steamed Beef with Noodles	\$2.95
Sichuan Gourmet		362 Chicken Bone Soup	\$1.75	402 Fried Pork Chops w/ Bean Paste	\$2.45
Delivery: 767-676-7228 or 767-676-7248		363 Noodle Soup (200g)	\$1.75	403 Beef & Pork Stew (Szechuan Style)	\$2.45
200g (7oz) Minimum Delivery		364 Noodle Soup (300g)	\$2.00	404 Beef Stew (Szechuan Style)	\$2.45
Min. 100g (3.5oz) Delivery		365 Noodle Soup (400g)	\$2.25	405 Beef Stew (Szechuan Style)	\$2.45
Min. 200g (7oz) Delivery		366 Noodle Soup (500g)	\$2.50	406 Beef Stew (Szechuan Style)	\$2.45
Min. 300g (10.5oz) Delivery		367 Noodle Soup (600g)	\$2.75	407 Beef Stew (Szechuan Style)	\$2.45
Min. 400g (14oz) Delivery		368 Noodle Soup (700g)	\$3.00	408 Beef Stew (Szechuan Style)	\$2.45
Min. 500g (17.5oz) Delivery		369 Noodle Soup (800g)	\$3.25	409 Beef Stew (Szechuan Style)	\$2.45
Min. 600g (21oz) Delivery		370 Noodle Soup (900g)	\$3.50	410 Beef Stew (Szechuan Style)	\$2.45
Min. 700g (24.5oz) Delivery		371 Noodle Soup (1000g)	\$3.75	411 Beef Stew (Szechuan Style)	\$2.45
Min. 800g (28oz) Delivery		372 Noodle Soup (1100g)	\$4.00	412 Beef Stew (Szechuan Style)	\$2.45
Min. 900g (31.5oz) Delivery		373 Noodle Soup (1200g)	\$4.25	413 Beef Stew (Szechuan Style)	\$2.45
Min. 1000g (35oz) Delivery		374 Noodle Soup (1300g)	\$4.50	414 Beef Stew (Szechuan Style)	\$2.45
Min. 1100g (38.5oz) Delivery		375 Noodle Soup (1400g)	\$4.75	415 Beef Stew (Szechuan Style)	\$2.45
Min. 1200g (42oz) Delivery		376 Noodle Soup (1500g)	\$5.00	416 Beef Stew (Szechuan Style)	\$2.45
Min. 1300g (45.5oz) Delivery		377 Noodle Soup (1600g)	\$5.25	417 Beef Stew (Szechuan Style)	\$2.45
Min. 1400g (49oz) Delivery		378 Noodle Soup (1700g)	\$5.50	418 Beef Stew (Szechuan Style)	\$2.45
Min. 1500g (52.5oz) Delivery		379 Noodle Soup (1800g)	\$5.75	419 Beef Stew (Szechuan Style)	\$2.45
Min. 1600g (56oz) Delivery		380 Noodle Soup (1900g)	\$6.00	420 Beef Stew (Szechuan Style)	\$2.45
Min. 1700g (59.5oz) Delivery		381 Noodle Soup (2000g)	\$6.25	421 Beef Stew (Szechuan Style)	\$2.45
Min. 1800g (63oz) Delivery		382 Noodle Soup (2100g)	\$6.50	422 Beef Stew (Szechuan Style)	\$2.45
Min. 1900g (66.5oz) Delivery		383 Noodle Soup (2200g)	\$6.75	423 Beef Stew (Szechuan Style)	\$2.45
Min. 2000g (70oz) Delivery		384 Noodle Soup (2300g)	\$7.00	424 Beef Stew (Szechuan Style)	\$2.45
Min. 2100g (73.5oz) Delivery		385 Noodle Soup (2400g)	\$7.25	425 Beef Stew (Szechuan Style)	\$2.45
Min. 2200g (77oz) Delivery		386 Noodle Soup (2500g)	\$7.50	426 Beef Stew (Szechuan Style)	\$2.45
Min. 2300g (80.5oz) Delivery		387 Noodle Soup (2600g)	\$7.75	427 Beef Stew (Szechuan Style)	\$2.45
Min. 2400g (84oz) Delivery		388 Noodle Soup (2700g)	\$8.00	428 Beef Stew (Szechuan Style)	\$2.45
Min. 2500g (87.5oz) Delivery		389 Noodle Soup (2800g)	\$8.25	429 Beef Stew (Szechuan Style)	\$2.45
Min. 2600g (91oz) Delivery		390 Noodle Soup (2900g)	\$8.50	430 Beef Stew (Szechuan Style)	\$2.45
Min. 2700g (94.5oz) Delivery		391 Noodle Soup (3000g)	\$8.75	431 Beef Stew (Szechuan Style)	\$2.45
Min. 2800g (98oz) Delivery		392 Noodle Soup (3100g)	\$9.00	432 Beef Stew (Szechuan Style)	\$2.45
Min. 2900g (101.5oz) Delivery		393 Noodle Soup (3200g)	\$9.25	433 Beef Stew (Szechuan Style)	\$2.45
Min. 3000g (105oz) Delivery		394 Noodle Soup (3300g)	\$9.50	434 Beef Stew (Szechuan Style)	\$2.45
Min. 3100g (108.5oz) Delivery		395 Noodle Soup (3400g)	\$9.75	435 Beef Stew (Szechuan Style)	\$2.45
Min. 3200g (112oz) Delivery		396 Noodle Soup (3500g)	\$10.00	436 Beef Stew (Szechuan Style)	\$2.45
Min. 3300g (115.5oz) Delivery		397 Noodle Soup (3600g)	\$10.25	437 Beef Stew (Szechuan Style)	\$2.45
Min. 3400g (119oz) Delivery		398 Noodle Soup (3700g)	\$10.50	438 Beef Stew (Szechuan Style)	\$2.45
Min. 3500g (122.5oz) Delivery		399 Noodle Soup (3800g)	\$10.75	439 Beef Stew (Szechuan Style)	\$2.45
Min. 3600g (126oz) Delivery		400 Noodle Soup (3900g)	\$11.00	440 Beef Stew (Szechuan Style)	\$2.45
Min. 3700g (129.5oz) Delivery		401 Noodle Soup (4000g)	\$11.25	441 Beef Stew (Szechuan Style)	\$2.45
Min. 3800g (133oz) Delivery		402 Noodle Soup (4100g)	\$11.50	442 Beef Stew (Szechuan Style)	\$2

Enable others to build and deploy services and apps in DuraSpace environment

 <p>Bloons Digital Goldfish Ltd Released Apr 05, 2009 \$0.99 BUY APP</p>	 <p>Mover Infinite Labs Released May 11, 2009 Free GET APP</p>	 <p>Ewe Stash - Knitting Treeness, LLC Released May 11, 2009 \$2.99 BUY APP</p>
 <p>Trixel ADEPT Games, LLC Released Apr 27, 2009 \$0.99 BUY APP</p>	 <p>Spawn Illuminati Elements of Design Released Feb 23, 2009 \$1.99 BUY APP</p>	 <p>World of Tunes Com2uS Corp. Released May 13, 2009 \$1.99 BUY APP</p>
 <p>Angels&Demons Sony Pictures Television Released Apr 20, 2009 \$4.99 BUY APP</p>	 <p>Gravulous SilentLogic Studios Released May 01, 2009 \$0.99 BUY APP</p>	 <p>STAR TREK® Electronic Arts Released May 05, 2009 \$2.99 BUY APP</p>
 <p>Myst Cyan Worlds Released May 02, 2009 \$5.99 BUY APP</p>	 <p>Vir2L® Drums Vir2L Studios Released Apr 17, 2009 \$0.99 BUY APP</p>	 <p>Dane Cook Global Zannel, Inc. Released Apr 16, 2009 \$2.99 BUY APP</p>

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
- 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 for end user
- 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
 - Amazon
 - Sun
 - Microsoft
 - EMC
- Selected initial 3 pilot partners
 - New York Public Library
 - Biodiversity Heritage Library
 - TBD (in selection process now)

Timeline

- Alpha DuraCloud service – June 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: <http://duraspace.org>

DuraCloud Service: <http://duracloud.org> (soon)

