Road-show
8-10 Aug 2010

Achieving Web-Scale for Libraries

Robin Murray
Vice President, Global Product Management
OCLC
Achieving Web-Scale for Libraries

- What is Web-Scale?
  - Is it the same as “The Cloud”?
  - Examples of Web-Scale
- Data, Community, Infrastructure
- OCLC and Web-Scale
  - Data, Community, Infrastructure
  - OCLC Product Strategy
- Libraries building Web-Scale for Libraries
  - Where we are today…
“'Web-scale' refers to how major web presences architect systems and services to scale as use grows. But it also seems evocative in a broader way of the general attributes of the large gravitational hubs which are such a feature of the current web (eBay, Amazon, Google, WikiPedia, ...).”
The Web is all about **scale**, finding ways to attract the most users for **centralized resources**, spreading those costs over larger and larger **audiences** as the technology gets more and more capable.

Chris Anderson
And Scale Matters...

In a web-economy the rich get richer and...

=>

Web Scale is critical for libraries
Web-Scale and Cloud Computing

A style of computing in which scalable and elastic IT-enabled capabilities are delivered as a service to external customers using Internet technologies.

-Gartner Group

Simple: Web-based applications delivered remotely.

Cloud = Infrastructure
Web-Scale is more than just Infrastructure
Web-Scale: examples

Infrastructure
Data
Community

eBay
Amazon
Facebook
Google
Libraries and Web-Scale?
OCLC: Helping Libraries Build Web-Scale for Libraries

Infrastructure

Data

Community
Data: WorldCat Growth since 1998

Millions of records

<table>
<thead>
<tr>
<th>Year</th>
<th>Records (Millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998</td>
<td>39</td>
</tr>
<tr>
<td>1999</td>
<td>41</td>
</tr>
<tr>
<td>2000</td>
<td>44</td>
</tr>
<tr>
<td>2001</td>
<td>47</td>
</tr>
<tr>
<td>2002</td>
<td>50</td>
</tr>
<tr>
<td>2003</td>
<td>52</td>
</tr>
<tr>
<td>2004</td>
<td>55</td>
</tr>
<tr>
<td>2005</td>
<td>61</td>
</tr>
<tr>
<td>2006</td>
<td>67</td>
</tr>
<tr>
<td>2007</td>
<td>86</td>
</tr>
<tr>
<td>2008</td>
<td>108</td>
</tr>
<tr>
<td>2009</td>
<td>139</td>
</tr>
<tr>
<td>2010</td>
<td>170</td>
</tr>
</tbody>
</table>
**Data: WorldCat across Print, License and Digital Data**

1.9 billion items and growing!

<table>
<thead>
<tr>
<th>Physical holdings in WorldCat</th>
<th>Licensed digital content in library collections</th>
<th>Local library content being digitized</th>
</tr>
</thead>
<tbody>
<tr>
<td>170 million bib records</td>
<td>325 million electronic database records</td>
<td>30 million items (Google, HathiTrust, OAlster)</td>
</tr>
<tr>
<td>3.6 million digital items</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5 billion holdings</td>
<td><strong>NEW! JSTOR Metadata: 4.5 million records</strong></td>
<td></td>
</tr>
</tbody>
</table>
Community: The OCLC Cooperative

72,035 libraries in 171 countries

- 1,418
- 55,820
- 1,091
- 1,752
- 5,715
- 1,800
- 381
- 4,058
Community: New OCLC governance structure

- Members
- Regional Council
- Global Council
- Board of Trustees
Community: New OCLC governance structure

Members

Americas Regional Council

Asia Pacific Regional Council

EMEA Regional Council

Global Council

Board of Trustees
1. Web-Scale is critical for libraries
   - In a web-economy the rich get richer and...

2. OCLC is the only organization that could deliver web scale for libraries
   - Data, Community, Infrastructure
   - Opportunity and Obligation
Design for Library Web-Scale

Design for Scale

Design for Community

- An Open Platform for “Collective Innovation”

Design for Capability

- D2D; License Management; Circulation & Acquisitions; Analytics; 3rd Party Apps...

Design for Economy

- Reduce costs
# “Library Web scale”

Worldwide libraries and worldwide library transactions

<table>
<thead>
<tr>
<th>Libraries worldwide</th>
<th>1,212,383</th>
</tr>
</thead>
<tbody>
<tr>
<td>Books: physical processing</td>
<td>15,517,196,010</td>
</tr>
<tr>
<td>Back-office transactions</td>
<td>61,879,349</td>
</tr>
<tr>
<td>OPAC searches</td>
<td>105,607,800,600</td>
</tr>
<tr>
<td>Database searches</td>
<td>36,555,852,000</td>
</tr>
<tr>
<td>Circulation / ILL</td>
<td>4,983,393,968</td>
</tr>
<tr>
<td>+ Adds/deletes; patron record maintenance, etc.</td>
<td></td>
</tr>
</tbody>
</table>

---

Annual transactions | 166,041,975,140 |

Possible with a small farm of commodity servers in the cloud with appropriately architected software.

=> Massive infrastructure cost reductions possible for libraries.

- 18,954,563 transactions / day
- 5,265 transactions / second
## Design for Web-Scale

<table>
<thead>
<tr>
<th>Goals</th>
<th>Architecture Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responsive</td>
<td>Service Oriented Architecture</td>
</tr>
<tr>
<td>Massively Scalable</td>
<td>Partition by data and domain</td>
</tr>
<tr>
<td>Highly Fault Tolerant</td>
<td>“Shared Nothing” Architecture</td>
</tr>
<tr>
<td>Suitable for Public Consumption</td>
<td>Embrace Open Standards</td>
</tr>
<tr>
<td></td>
<td>Judicious Caching</td>
</tr>
<tr>
<td></td>
<td>Optimistic Locking</td>
</tr>
<tr>
<td></td>
<td>Asynch. Transactions</td>
</tr>
<tr>
<td></td>
<td>Network savvy APIs</td>
</tr>
<tr>
<td></td>
<td>Stateless Services</td>
</tr>
<tr>
<td></td>
<td>Highly Layered</td>
</tr>
<tr>
<td></td>
<td>Versioned APIs</td>
</tr>
<tr>
<td></td>
<td>Discoverable Services</td>
</tr>
<tr>
<td></td>
<td>Data Redundancy</td>
</tr>
<tr>
<td></td>
<td>Replication &amp; Failover</td>
</tr>
<tr>
<td></td>
<td>Temporary data inconsistency</td>
</tr>
<tr>
<td></td>
<td>Avoid Distributed Transactions</td>
</tr>
</tbody>
</table>
### Design for Web-Scale

<table>
<thead>
<tr>
<th>Goals</th>
<th>Architecture Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responsive</td>
<td>Service Oriented Architecture</td>
</tr>
<tr>
<td>Massively Scalable</td>
<td>Partition by data and domain</td>
</tr>
<tr>
<td>Highly Fault Tolerant</td>
<td>Embrace Open Standards</td>
</tr>
<tr>
<td>Suitable for Public Consumption</td>
<td>Judicious Caching</td>
</tr>
</tbody>
</table>

- **Judicious Caching**
- **Optimistic Locking**
- **Highly Layered**
- **Data Redundancy**
- **Discoverable Services**
- **Asynch. Transactions**
- **Network savvy APIs**
- **Stateless Services**
- **Versioned APIs**
- **Replication & Failover**
- **Temporary data inconsistency**
Design for Community: Collective Innovation
Infrastructure: OCLC Product Strategy

**Open and Extensible Platform**
built on an *extended view of WorldCat*.

- **“Open”** – 3rd-party systems can make use of core services in a supplier-neutral manner – supporting the widest possible reach of the co-operative and use of the platform.
- **Extensible**” – users, third-party suppliers and the library development community can add services and applications – fostering collective innovation.
- **“Extended View of WorldCat”** – the collection of databases that represent data for purchased, licensed and digital content, exposed through a rich range of network-level data services.
Design for Functionality: Applications

- Global Library Exposure: WorldCat.org
  - Search-engine syndication; Google books; affiliate sites...
  - >1M referrals to libraries/month

- Discovery to Delivery: WorldCat Local
  - ~1200 libraries live with WorldCat Local

- Circulation & Acquisitions
  - First sites live now.

- Knowledge Base Management
  - Available Now. Free as part of cataloguing

- License Management
  - Pilot Jan 2012

- Open Platform
  - DevNet now; Full Platform mid-2012
3rd Party Apps

- ~50 Apps built on WorldCat API service
- Full platform release will see 100’s of services available ...

http://redlaser.com
Design for Economy

- Reduced Systems Costs
  - Design for scale => Low cost subscription services

- Reduce Running Costs
  - IT Infrastructure
  - IT Management
  - Upgrade cycles...

- Reduced Workflow Costs
  - Cataloguing
  - Acquisitions
  - Circ/ILL
  - Knowledge Base Management
  - Vendor Management
  - ...

[Image of OCLC logo]
Libraries Building Web-Scale for Libraries: Summary

- Web-Scale is Critical for Libraries
- Web-Scale: Data, Community, Infrastructure
- OCLC has unique position in helping libraries achieve Web-Scale
- Benefits to Libraries:
  - Improved Service - increased exposure, syndication
  - Gain from collective innovation
  - Streamlined Workflows
  - Reduced Costs
Achieving Web-Scale for Libraries

Robin Murray
Vice President, Global Product Management
OCLC