Cooperatively Building Web Scale for Libraries

Robin Murray
Robin Murray, VP Global Product Management

OCLC Global Council
April 11, 2011
Collaboratively Building Web-Scale with 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 : The Web-Scale Platform
- Collaboratively building Web-Scale with 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, ...).”

Lorcan Dempsey
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.
And Scale Matters...

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

=>

Web Scale is critical for libraries
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
Libraries and Web-Scale?
OCLC: Collaboratively Building Web-Scale with Libraries
Data: WorldCat Growth since 1998

 Millions of records

<table>
<thead>
<tr>
<th>Year</th>
<th>1998</th>
<th>1999</th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>39</td>
<td>41</td>
<td>44</td>
<td>47</td>
<td>50</td>
<td>52</td>
<td>55</td>
<td>61</td>
<td>67</td>
<td>86</td>
<td>108</td>
<td>139</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, OAIster)</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:</strong> 4.5 million records</td>
<td></td>
</tr>
</tbody>
</table>
Community: The OCLC Cooperative

72,035 libraries in 171 countries
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
Infrastructure: OCLC Web-Scale Product Strategy

- 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

Libraries worldwide
1,212,383

Books: physical processing
15,517,196,010

Back-office transactions
61,879,349

OPAC searches
105,607,800,600

Database searches
36,555,852,000

Circulation / ILL
4,983,393,968
+ Adds/deletes; patron record maintenance, etc.

__________________________

Annual transactions
166,041,975,140

=> 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>Asynch. Transactions</td>
</tr>
</tbody>
</table>

- Judicious Caching
- Embrace Open Standards
- Optimistic Locking
- Stateless Services
- Highly Layered
- Data Redundancy
- Avoid Distributed Transactions
- Temporary data inconsistency
- Discoverable Services
- Versioned APIs
- Replication & Failover
## Design for Web-Scale

<table>
<thead>
<tr>
<th>Goals</th>
<th>Architecture Features</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Responsive</strong></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>Discoverable Services</td>
</tr>
<tr>
<td></td>
<td>Judith Caching</td>
</tr>
<tr>
<td></td>
<td>“Shared Nothing” Architecture</td>
</tr>
<tr>
<td></td>
<td>Network savvy APIs</td>
</tr>
<tr>
<td></td>
<td>Optimistic Locking</td>
</tr>
<tr>
<td></td>
<td>Stateless Services</td>
</tr>
<tr>
<td></td>
<td>Highly Layered</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>
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 cooperative 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.
The Platform

- What is it?
  - The core underlying all OCLC applications. Opened to enable...
    - Innovation, Publishing, Sharing...

- Some early examples
  - Plugging additional features into an OCLC application...
  - Surfacing OCLC services in a 3rd party environment...
  - A 3rd party surfacing library services in their app...
The Platform - The core infrastructure underlying all OCLC applications...
The Platform - The core infrastructure underlying all OCLC applications...

I want to Innovate and Integrate

I want to Expose and Share Innovations

I want to Benefit from others’ Innovations

The Platform - The core infrastructure underlying all OCLC applications...

I want to Innovate and Integrate

I want to Expose and Share Innovations

I want to Benefit from others’ Innovations

The Platform - The core infrastructure underlying all OCLC applications...
The Platform - The core infrastructure underlying all OCLC applications ...

I want to Innovate and Integrate

I want to Expose and Share Innovations

I want to Benefit from others’ Innovations

Ability to create apps.
(service catalog, service directory)

Ability to publish apps.

Ability to find & install apps.
(App Store)

Ability to share
(Community Site)

Data Layer

WorldCat

Business Logic Services

Core Data Services

Registry
KBWC
WorldCat
Identifiers
X-ID
...

The Platform - The core infrastructure underlying all OCLC applications ...
Building Apps

Service Catalog - provides full documentation to the platform

Community Site - provides help and collaboration facilities
Developer Network
‘Hackathons’
### Manage Apps

**Add an App to the repository (max size: 5120K)**

<table>
<thead>
<tr>
<th>ID</th>
<th>App Name</th>
<th>Status</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>237</td>
<td>Search the Gail Borden Public Library catalog</td>
<td>Approved</td>
<td></td>
</tr>
<tr>
<td>238</td>
<td>The Librarian's Book Reoogle</td>
<td>Approved</td>
<td></td>
</tr>
<tr>
<td>239</td>
<td>Montréal Library Network</td>
<td>Approved</td>
<td></td>
</tr>
<tr>
<td>240</td>
<td>Riverside Public Library</td>
<td>Approved</td>
<td></td>
</tr>
<tr>
<td>241</td>
<td>Frogger</td>
<td>Approved</td>
<td></td>
</tr>
<tr>
<td>242</td>
<td>Hamster</td>
<td>Approved</td>
<td></td>
</tr>
<tr>
<td>243</td>
<td>Tetrix</td>
<td>Approved</td>
<td></td>
</tr>
</tbody>
</table>
**Finding & Installing Apps : The App Gallery**

**Search the Gail Borden Public Library catalog**
Web Service: General
Developer: Gail Borden Public Library
Search the Gail Borden Public Library catalog for books, movies, and more.

**The Librarian's Book Revoogle**
Web Service: General
Developer: Paul K.
A search engine for finding online book reviews written by librarians, staff, or library users. Library and librarian websites are included.

**Montréal Library Network**
Web Service: General
Developer: Jonathan Latreille
Search for book(s) in NELLIGAN catalogue from Montréal library network

**Riverside Public Library**
Web Service: General
Bestseller 1.5

Description: The Bestseller widget aggregates data from online sources such as the New York Times and Oprah Bookclub lists.

Organization: 6559

Create Date: Wed Feb 23 13:28:16 EST 2011

Author: Harnish

Author Email: harnishk@oclc.org

Locales: 

Categories: [General]

Features: [analytics]

Screenshot:
### Early App Examples

**Current Budget:** $4,573.21  
**Remaining Budget:** $4,348.43  

[http://redlaser.com](http://redlaser.com)
Current Status...

- 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
- Open Platform
  - DevNet now; Pilot March; Full Platform mid-2011
Collaboratively Building Web-Scale with Libraries: Summary

• Web-Scale is Critical for Libraries
• Web-Scale: Data, Community, Infrastructure
• OCLC has unique position in helping libraries achieve Web-Scale
• Web-Scale Platform - designed for collective innovation