Winds of change:
Libraries and cloud computing

By Matt Goldner.

Executive Summary

Cloud computing is a new technology model for IT services which many businesses and organizations are adopting. It allows them to avoid locally hosting multiple servers and equipment and constantly dealing with hardware failure, software installs, upgrades and compatibility issues. For many organizations, cloud computing can simplify processes and save time and money. This article defines cloud computing and shows how it is different from other types of computing. It also discusses how cloud computing solutions could be beneficial to libraries in three basic areas: technology, data and community.

Introduction

Cloud computing can transform the way systems are built and services delivered, providing libraries with an opportunity to extend their impact.

Cloud computing has become a major topic of discussion and debate for any business or organization which relies on technology. Anyone connected to the Internet is probably using some type of cloud computing on a regular basis. Whether they are using Google's Gmail, organizing photos on Flickr or searching the Web with Bing they are engaged in cloud computing. As Geoffrey Moore points out, the interesting thing about cloud computing is it did not start as a technology for the business enterprise, but was driven by the public with services like Facebook and Flickr.[i]

Over the last few years businesses have started to see the value of cloud computing causing it to become a major technology solution for businesses and organizations around the world. Looking across the information and broader technology landscape, it is not difficult to find success stories of switching to cloud computing, disaster stories, and a great deal of debate about what cloud computing is, or isn't. The purpose of this article is to look specifically at how cloud computing can be employed by libraries and what needs to be considered before moving into a cloud computing solution.

What is cloud computing?

First there must be a definition of cloud computing for this discussion. The Gartner Group defines cloud computing as "a style of computing in which massively scalable and elastic IT-enabled capabilities are delivered as a service to external customers using Internet technologies."[ii] In various presentations KPMG breaks this into essentially four different types of cloud computing: infrastructure, platform, applications and services. To put this in more concrete terms, examples of each can be:

Matt Goldner
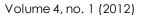
Type	What it is	Examples
Infrastructure	Buying space / time on external servers	Amazon A3 Bungee
Platform	An existing software platform to build your own applications on	Facebook
Applications	Software applications accessed with a Web browser	Google Docs Salesforce.com
Services	Ready to use services accessed with a Web browser	ADP

The above table illustrates why there are varying definitions of cloud computing. Many cloud services actually incorporate two or more of these types. For example, Google docs provide infrastructure as well as applications. It should also be noted that many cloud applications and services are actually using another providers' cloud infrastructure to run their service, which will be touched on later in this article.

How is cloud computing different?

For much of the past 25 years, software development and system engineering has centered primarily on the personal computer. The PC era was characterized by monolithic, proprietary operating systems and programs that had long development times and release cycles. In that environment, the design of software was isolated and all attention focused on a single application.

With cloud computing, hardware and functionality traditionally installed and run in a local environment is now performed on the network, in the Internet cloud. In essence, the Internet cloud becomes the development platform and the operating system to which programmers write reusable, constantly updated software components that are delivered over the network and that can be embedded or loosely coupled with other Web applications.

Libraries have been using some cloud computing services for over a decade. Online databases are accessed as cloud applications. Large union catalogs can also be defined as cloud applications. However, a look outside libraries is warranted to better understand the value proposition of cloud computing.

Why are businesses and organizations adopting cloud computing solutions?

Jeff Bezos of Amazon has repeatedly spoken of the 70/30 rule. He states that it can be demonstrated that businesses which run applications spend 70% of their time and money supporting the infrastructure required to keep their business going.[iii] This only leaves them 30% of time and money to work on innovation and ways to improve and grow their business. He goes on to show that when a business moves their core applications to a cloud-based solution, they can invert this ratio thus giving them 70% of their time and money to improve and grow their business. In a recent conversation with the director of a large academic research library, this proposition was set forth. She started laughing and said she wished they were only expending 70% of the time and money on infrastructure.

John Waters gives an example of this rule at work. He is executive director of the Minnesota Online High School (MNOHS), which supports all of its courses over the Internet for its students. In talking about why they switched to a cloud solution he states:

"We were collapsing under the weight of the need to support so many different computers ... Until recently, the school provisioned those computers by sending out CDs or its licensed software and guiding students through downloads or the open source and custom applications it uses. ... MNOHS began looking for a better way. ... the school launched a pilot program to test a system designed to move its entire operation to the cloud. ... Student work doesn't reside on computers anymore. All the applications and data are stored in the cloud. No more CDs. No

more downloads. And if a laptop dies or gets stolen, student work isn't lost."[iv]

Essentially this let the Minnesota Online High School switch their efforts from managing technology to focusing on education. This should be the purpose of cloud computing solutions, to let a business or organization focus on its core business or mission instead of technology to deliver that business or mission to its consumers.

What can cloud computing solutions do for libraries?

So turning to cloud computing and libraries, are there real problems that can be solved? The answer is yes. The library community can apply the concept of cloud computing to amplify the power of cooperation and to build a significant, unified presence on the Web. This approach to computing can help libraries save time and money while simplifying workflows.

A brief list of potential areas of improvement could include:

- Most library computer systems are built on pre-Web technology
- Systems distributed across the Net using pre-Web technology are harder and more costly to integrate
- Libraries store and maintain much of the same data hundreds and thousands of times
- With library data scatter across distributed systems the library's Web presence is weakened
- With libraries running independent systems collaboration between libraries is made difficult and expensive
- Information seekers work in common Web environments and distributed systems make it difficult to get the library into their workflow
- Many systems are only used to 10% of their capacity. Combining systems into a cloud environment reduces the carbon footprints, making libraries greener

These improvements can be grouped into three basic areas: technology, data and community. Each offers some general and some unique opportunities for libraries. Looking first at the technology that most current library systems employ, several benefits of cloud computing solutions surface.

Technology improvements

Cloud computing solutions at their essence are built on current technology and should be architected to allow for technology shifts. Looking at the explosion of mobile devices one sees how businesses and organizations operating in a cloud environment are able to adapt and deliver their services to the new devices much more quickly and less expensively.

The mainstay of libraries is the library management system (LMS, also known as the integrated library system or ILS). Library management systems were developed before the Internet and Web existed and are generally closed proprietary systems. It has been difficult and costly for these closed systems to take advantage of new technologies as they emerge. It is also challenging to integrate to external systems and libraries must rely on their vendors to do any such integration.

Over time libraries have needed to add more systems to manage their changing collections which moved from strictly physical collection management to a combination of physical, licensed and digital collections. Since each of these systems has stood alone integrating them has been difficult and at times not possible. What can change in a cloud environment for managing core library services?

First would be the possibility of open service oriented architecture. Many cloud solutions offer this type of openness with published application program interfaces (APIs) that any programmer can take advantage of. This means if a new service or technology emerges libraries will not always be dependent on a vendor or other third party to start taking advantage of these services and technologies. Existing library systems have used APIs to connect to external services but they have remained closed proprietary systems making it hard to integrate them into external services. As Andrew Pace stated it, "... demands fall short by merely asking that local systems avail themselves of other Web services rather than establishing themselves as services in their own right."[v] When library systems are deployed as open cloud solutions then the library community itself can step up to create extensions to their core services and more importantly share them throughout the community using cloud solutions. This makes it possible to integrate two services once and re-use it across the community.

Secondly libraries can get out of the business of technology and focus on collection building, patron services and innovation. Servers can be decommissioned and no longer require replacement every five years (or less). Staff no longer has to

maintain the complex software stack necessary to run local systems and worry about compatibility of the stack during upgrades. Instead technical skills can be re-deployed for extending cloud services into their environment and their environment into other cloud services.

Data efficiencies

When data is stored in the cloud it offers several advantages. Common data can now be easily shared among services and users. The need for local storage, maintenance and backups is removed. Agreements can be forged to share data that normally would be considered private to a single business or organization. And finally libraries can achieve Web scale when they massively aggregate data and users, something a cloud environment makes possible.

Like the advantages of technology deployed and accessed as cloud solutions, data storage in the cloud brings many benefits for libraries. The easy one to recognize is the same data being stored hundreds and thousands of times across libraries. Consider how many copies of the cataloging data there are for a serial publication such as the Economist. And if a change is needed to the cataloging data to keep it current each library must perform that change. When this data is maintained in the cloud, maintenance and backup of this data is now done once and if a change is needed, once one library performs the change all share it.

Another great benefit of data stored in the cloud is the opportunity for collaboration and cooperative intelligence. Libraries can agree to share pools of data for cooperative collection building, cooperative preservation or digitization, cooperative sharing of materials, etc. And with massively aggregated data new services can be created such as recommender services based on a broad base of usage data.

As stated above when library data is widely distributed across systems it makes library Web presence weak. When search engines such as Google, Yahoo and Bing can harvest from large data stores it opens the opportunity for the collective to work on search engine optimization, or the improvement of library collections appearing more relevant to search engines thus displayed higher in search results. This is a complex and ever changing task that would be prohibitive for individual libraries to accomplish. Further, aggregated data can attract a much larger aggregation of users who interact with

the data, add to it and re-use it. The result is every user adds benefit for every other user.

Community power

Libraries have a somewhat unique opportunity with cloud computing, to create an online information community network. Such a community is really two communities, the internal community of libraries collaborating within a single institution and across institutions and the external community of libraries and information seekers. The value to libraries is the "network effect" that coming together in the cloud provides. The cooperative efforts of libraries will create scale savings and efficiencies, bring wider recognition for libraries, and provide cooperative intelligence for better decision-making, and provide the platform on which libraries can innovate.

Looking externally the first community cloud computing offers is taking advantage of social media. Businesses and organizations can both build social communities around their services and participate in existing social communities such as Facebook or Twitter.

The internal community formed through the cloud offers new possibilities and efficiencies for current workflows. Starting with a single organization the simple task of collaboratively working on documents and maintaining version control either requires extensive manual processes between colleagues or a locally installed system to assist in collaboration and version control. Many librarians have discovered the power of services like Google Docs to reduce the effort of working jointly. Services like these allow them to easily share ongoing work whenever they want and wherever they are.

The potential for collaboration between libraries is truly revolutionary in a cloud environment. When data and functions are shared in the cloud libraries can make joint decisions on collection development, preservation, digitization, in real time. As demonstrated by OCLC's QuestionPoint virtual reference service and it's 24/7 cooperative a single libraries ability to assist patron's is expanded beyond the constraint of its own walls and hours of operation to become a true cloud service (last year QuestionPoint logged its five millionth answer to a reference question).

Real world examples of current library cloud solutions

To date, the main focus of libraries moving into the cloud has been discovery services, the need to disclose their vast collections on the Web. Though library OPACs attract existing patrons they are not integrated with most information seekers common workflows. So a first step for libraries has been to start massively aggregating data about their collections into common pools. OCLC's WorldCat, the first example of this, is now forty years old and pre-dates both the Web and cloud computing. Other similar union catalogs have existed throughout the world most commonly supported by national libraries and large union catalogs, such as the National Library of Australia, the Bayerische Staatsbibliothek in Germany, and Bibsys in Norway. However the advent of the Web has allowed libraries to extend this original vision in new ways.

Extending these services beyond traditional library collections is well illustrated by the National Library of Australia's (NLA) Trove. It has used the Web to accomplish two tasks. This is done by first combining the collections of Australian libraries with other important Australian and international collections and information sources such a Wikipedia and secondly to open much of this content so the public can tag it, edit it, collect it and review it.

The explosion of digitization projects in the last decade has driven this gathering of information in new directions with examples in addition to NLA's Trove to others like the Hathi Trust, OAISTER and Europeana. The Hathi trust is building a repository of digitized books and journals from major research libraries in the United States. OAISTER is a service started by the University of Michigan and now managed by OCLC which seeks to harvest all the major digital repositories around the world. Europeana is gathering the digitized collections from Europe's galleries, libraries, archives and museums. What makes these aggregations and others like them important is their intent to allow their content to be mashed up into other services and re-used.

Other benefits growing from massively aggregated data about collections is the ability to aggregate user opinion and use. LibraryThing is a good example of being able to build recommender services based on the aggregation of what thousands of people hold in their personal libraries.

However, there is no reason to extend cloud-based services only to libraries' end users. As Marshall

Breeding points out, "We can't let the current focus on front-end interfaces make us complacent about the software systems that we use to automate routine library functions." [vi]

Beyond library discovery services

It is here that libraries can look to gain new efficiencies both internally and among the entire library community. When library software suppliers create the user personas that will use their software the focus is generally on external personas but there are also many internal personas that need to take advantage of new technologies and Web capabilities. One such example has been given with reference librarians now able to both better assist their patrons online but also to build a large network of librarians globally who can answer specific questions and be available 24/7. What other personas in the library can benefit from cloud solutions?

- Acquisitions librarians managing increasingly diverse collections
- Cataloging librarians seeking to describe an ever increasing body of information and information sources the library is managing
- Serials librarians working to maintain control and access to collections spidered across the Web
- Electronic resource librarians managing burgeoning collections, and ever-changing lists of vendors

The dramatic change in library collections often blurs the lines between traditional job roles in libraries. An acquisitions librarian probably also needs to manage licenses for electronic materials as well as manage purchasing for multiple formats, often for the same item. They need to access information from suppliers, reviewers, local constituency and other staff in a unified manner. This begs for an open system deployed where it can easily be accessed by external systems and pull in data and services in from those same systems. Cloud computing solutions can create the new workflows needed by librarians because it offers the opportunity for a cooperative platform for libraries to build on. There are four key principles of a cooperative platform.

- Openness, meaning that services and data are made available to support greater interoperability, not only within and between cloud services, but also with library-developed and third-party applications;

- Extensibility, meaning that the platform can easily accommodate the addition of new services and applications, developed either by the service provider or by members of the community;
- Data richness, meaning that a library can interact with and expose a wide variety of information about purchased, licensed, and digital content through this platform; and
- Collaboration, meaning that libraries can harness the collective power of the community of libraries to innovate and share solutions.

And it is precisely this that the business world and social media have demonstrated can be done with cloud computing solutions. Through cooperative and community building libraries can have the same possibilities.

Caveat Emptor

However, if libraries are to consider moving more of their services into the cloud there are certain questions that must be addressed. Foremost is whether this service will make the library more efficient and help it offer better service to its constituency. It goes without saying that adopting technology for technologies sake is not a good management decision. Once it is determined a cloud solution does accomplish this for the library then considerations to look at are:

- Does the service have built-in scalability, reliability and security?
- Is it multi-tenancy?
- Who owns the data stored in the system and what rights does the library have to extract their data for other uses or even to leave the service entirely?
- Is it an open system so that external data and services can be economically integrated into this service and its services can be economically integrated to external services?

In discussions of cloud computing security and privacy are raised as serious concerns throughout the literature and especially by librarians. When considering a cloud application two aspects of security and especially privacy must be examined, technical and legal. Does the provider demonstrate the necessary technical expertise and explanation of their environment to insure there will be no unauthorized access to a library's data stored in the cloud? And have they considered the legal requirements of the government bodies the library is answerable to? This means it is also important to know exactly where the data is going to be stored since different countries have much different privacy

requirements and standards. And since many cloud solutions are actually running on another suppliers cloud infrastructure due diligence is required. First to be certain where data is stored and secondly what the service level agreements for access to the data and preservation of the data are for the actual infrastructure supplier. Regardless, one important take-away remains—data privacy and security are not mutually exclusive to cloud based solutions.

Multi-tenancy is critical for scalability of any cloud service. "Multi-tenancy refers to a principle in software architecture where a single instance of the software runs on a server, serving multiple client organizations (tenants). Multi-tenancy is contrasted with a multi-instance architecture where separate software instances (or hardware systems) are set up for different organizations. With a multi-tenant architecture, a software application is designed to virtually partition its data and configuration thus each client organization works with a customized virtual application instance."[vii] This is the architecture that makes cloud solutions highly scalable thus must be considered when adopting a cloud service.

Data ownership cannot be overstressed. The library must know it has complete access to all their data while using the service so they can take it and re-use it as necessary, whether in another service or simply for reporting purposes. Just as important they must be certain they can extract all of their data at any point in the future should they decide to leave the service. It is also important to know what provisions are in place for data access should the supplier go out of business.

Finally a library must know that the service is truly an open, service oriented architecture which can truly change the future of libraries. This allows libraries to shift the use of internal technical expertise from maintaining software and servers towards innovative uses of cloud services in their local environment.

Conclusion

Libraries have the opportunity to improve their services and relevance in today's information society. Cloud computing is one avenue for this move into the future. It can bring several benefits for libraries and give them a different future.

The cooperative effect of libraries using the same, shared hardware, services and data— rather than hosting hardware and software on behalf of individual libraries—can result in lowering the total costs of managing library collections and enhancing

the both library user's experience and library staff workflows.

While local library systems served an important purpose earlier in library automation they now represent a tremendous duplication of effort. Each library builds and maintains a database, buys equipment and installs and updates the software. In fact, some libraries can get stuck in perpetual upgrade mode, which involves lots of testing and retesting and time-consuming customization.

With cloud computing, all of this is taken care of transparently for the library and user. Among the benefits of a cloud computing approach:

- Take advantage of current and rapidly emerging technology to fully participate in the Web's information landscape
- Increased visibility and accessibility of collections
- Reduced duplication of effort from networked technical services and collection management
- Streamlined workflows, optimized to fully benefit from network participation
- Cooperative intelligence and improved service levels enabled by the large-scale aggregation of usage data
- Make libraries greener by sharing computing power thus reducing carbon footprints

The vision is to use cloud computing to deliver library resources, services and expertise at the point of need, within user workflows and in a manner that users want and understand.

It should free libraries from managing technology so they can focus on collection building, improved services and innovation. The cloud computing model will encourage libraries and their users to participate in a network and community of libraries by enabling them to reuse information and socialize around information. It can also create a powerful, unified presence for libraries on the Web and give users a local, group and global reach.

References

i. Moore, Geoffrey: Core Content and the Cloud.
http://www.youtube.com/watch?v=0swJCYLH2C

ii. Gartner Group:
http://www.gartner.com/it/page.jsp?id=1035013

iii. Bezos, Jeff: :
http://www.slideshare.net/goodfriday/amazon-web-services-building-a-webscale-computingarchitecture

iv. Waters, John: Up, Up and Away. In: T.H.E. Journal 37:1 (January 2010) S. 22-4, 26-7

v. Pace, Andrew: 21st Century Library Systems. In: Journal of Library Administration 49 (2009) S. 641-650

vi. Breeding, Marshall: It's time to break the mold of the original ILS. In: Computers in Libraries 27:10 (2007) S. 39-41

vii. Wikipedia: http://en.wikipedia.org/wiki/Multi-tenancy

Matt Goldner joined OCLC in October 2004 and is currently the Product and Technology Advocate for the library community.

About OCLC: Founded in 1967, OCLC is a nonprofit, membership, computer library service and research organization dedicated to the public purposes of furthering access to the world's information and reducing library costs. More than 72,000 libraries in 170 countries have used OCLC services to locate, acquire, catalog, lend, preserve and manage library materials. Researchers, students, faculty, scholars, professional librarians and other information seekers use OCLC services to obtain bibliographic, abstract and full-text information when and where they need it. OCLC and its member libraries cooperatively produce and maintain WorldCat, the world's largest online database for discovery of library resources. Search WorldCat.org on the Web. For more information, visit the OCLC website.