大學圖書館
19 卷 1 期（民國 104 年 3 月） 頁 21-40

The Development of a SOA and APIs Based Chinese Cloud Cataloguing Platform

張迺貞 Naicheng Chang
大同大學通識教育中心副教授
Associate Professor, General Education Center, Tatung University
E-mail：ncchang@ttu.edu.tw

蔡育欽 Yuchin Tsai
國家實驗研究院國家網路與計算中心副研究員
Associate Researcher, National Center for High-performance Computing, NARL, NSC
E-mail：thomas@nchc.narl.org.tw

Alan Hopkinson
Chairman of the Permanent UNIMARC Committee of IFLA
E-mail：a.hopkinson@mdx.ac.uk

【Abstract】

The purpose of this project is to build the first Chinese cloud resources pooling platform with an on-demand self-service infrastructure for libraries and Library Management System (LMS) vendors by developing a Service-Oriented Architecture (SOA) and Application Programming Interfaces (APIs) procedure to access heterogeneous library bibliographic records. It aims to provide a cloud computing cataloguing service to libraries, academic and LMS vendors in Taiwan for future use.

The system platform design was developed under two stages: 1) constructing the cloud computing infrastructure includes setting up hardware facilities, integrating Internet resources and using a set of widely accepted free or open source virtual technologies, 2) deploying the cataloguing technology includes the development of software, using developers and a team of library professionals to reduce the technology gap and to ensure that library needs are taken into account.

The project demonstrates a potentially workable process in terms of availability and feasibility of a Chinese library cataloguing platform that is based on cloud computing

臺灣大學學術期刊資料庫

收稿日期：民國 103 年 10 月 8 日；接受日期：民國 103 年 12 月 18 日 doi:10.6146/univj.19-1.02

infrastructure. It has the potential to reduce dependence on library management systems. The library cataloguing platform initially for research purposes which will eventually serve as a model for 21st century software on demand and cloud computing platforms.

Keywords：cloud computing; Software as a Service (SaaS); Service-Oriented Architecture (SOA); library cataloguing platform

I. A new business model in the computing world

Because of advances in information technology network infrastructure, large amounts of data have moved to the Internet, to become resources that are available on demand, and this forms the basis of cloud computing. In a cloud environment, users do not maintain their own network resources, such as hardware, software or services. Network resources are provided as services over the web, through remote data centers on a subscription basis. The five essential characteristics of cloud computing are on-demand self-service, broad network access, resource pooling, rapid elasticity or expansion and a measured service (Mell and Grance, 2010). Cloud computing has reshaped the appearance of the information industry value chain, which has initiated a software and service-oriented era of competition. This can be seen in the three service models defined by the National Institute of Standards and Technology (NIST). In Infrastructure as a Service (IaaS), institutions outsource IT infrastructure to a cloud provider, to support their daily operation. The Amazon EC2 in the Amazon Web Service is an example of this. Platform as a Service (PaaS) allows users to deploy in-house applications on the cloud infrastructure, under the providers' cloud environment. The Google App Engine is an example of this. Software as a Service (SaaS) is a software delivery model in which software and its associated data are hosted centrally in the cloud and are typically accessed by users using a web browser over the Internet. SaaS has become a common delivery model for many business applications. Google features several web applications, such as Google Docs, which have a similar functionality to traditional office suites. In terms of deployment strategies, there are private clouds, community clouds, public clouds and hybrid clouds, which together provide ways to deliver cloud services.

Libraries often allocate a large amount of their budget to building infrastructure by purchasing software, purchasing and maintaining hardware, purchasing networking equipment and increasing storage space. Cloud computing is crucial to transforming the Library Management System (LMS) that provides services for a library, such as management of access to electronic

臺灣大學學術期刊資料庫

journals, as an additional module for an existing LMS (Dempsey, 2010; Mitchell, 2010; Peters, 2010). Unlike the traditional method of library management, SaaS allows libraries to access resources by paying a subscription fee. Because the software is hosted remotely, libraries can also invest less than usual in hardware. SaaS is easier to install, set-up and requires less daily upkeep and maintenance. With SaaS, the LMS vendors can concentrate on managing the SaaS platform, instead of providing separate services to different customers. Libraries can make strategic plans for the allocation of resources and offer better services than they could with in-house solutions. Further cost benefits for libraries are that libraries could request cloud-based services on demand, depending on their size and needs, so they can concentrate more on library services. Other benefits for LMS vendors would be a reduction in the maintenance cost and the cost of IT human resources, which would also increases the quality of commercial services. The Symphony and Horizon platforms from SirsiDynix, which provide integrated library systems that include library management tools in the SaaS environment, are examples of this kind (Rapp and Fialkoff, 2010).

The Service-Oriented Architecture (SOA) is an architecture model with a component of standardized web service technologies. It is a key concept and plays a crucial role in cloud computing-based services. The libraries or LMS vendors aggregate the cloud resources using an Internet Application Programming Interface (API) library and create cloud application services on the SOA platform. Figure 1 shows a virtual cloud platform that was built by the research team, using Kernel-based Virtual Machine (KVM) technology and the Cassandra database management system. The platform supports MARC (Machine-Readable Cataloguing), FRBR (Functional Requirements for Bibliographic Records), RDA (Resource Description and Access) and Dublin Core formats. It also supports data exchange using ISO 2709, supplemented by Z39.50. Using SOA API architecture, institutions and library developers can customize by developing Internet-based applications. The purpose of this project, therefore, is to build the first Chinese cloud resources pooling platform with an on-demand self-service infrastructure for libraries and LMS vendors by developing a SOA and APIs procedure to access heterogeneous library bibliographic records. It aims to provide a cloud computing cataloguing service to libraries, academic and LMS vendors in Taiwan for future use.

A. Library services platform

The technology allows library professionals and LMS vendors to use cloud computing to provide better quality library services more economically. Prestigious LMS's, such as Serials Solutions (2010),

臺灣大學圖書館 期刊資料庫

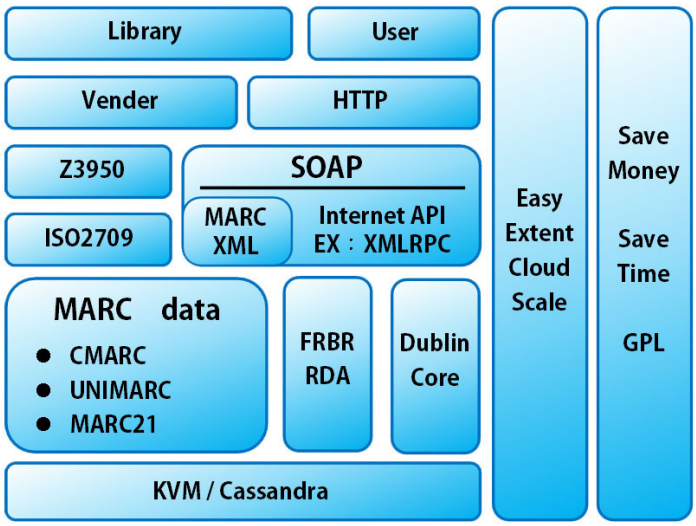


Figure 1: The performance benefits of a cloud cataloguing module

SirsiDynix, Innnovative Interfaces and Ex Libris, have also incorporated the technology into their products (Rapp and Fialkoff, 2010; Shimshock, 2010). It is clear that integrating cloud technology into library services will be the future trend for library services (Breeding, 2012; Collins and Rathemacher, 2010; Han, 2010; Lloret, 2012; Mavodza, 2013; Pace, 2009; Trehub and Wilson, 2010; Wilson, 2012; Yang 2012). Akerman (2007) at the National Research Council of Canada and the Kuali Foundation (2013) had made experiments on SOA and proved to be useful. Breeding (2010) identified these new opportunities for web services and SOA and predicted that the future library environment would increasingly embrace openness, SOA- and APIs-based library services. A number of studies have also noted that sharing is another technology that will come to libraries (Breeding, 2012; Goldner, 2011; Goldner and Birch, 2012; Wilson, 2012).

It is the authors' view that for an efficient SOA and APIs based cloud platform, the architecture must contain the following features: 1) a browser-based management interface; 2) a means of reducing the total cost of the service through reuse, a shared federation of resources, standards compliance; 3) resources and applications that can be shared on the Internet; 4) interoperability and aggregation of heterogeneous resources and 5) a flexible, reusable and integrated SOA-based interface.

Both Han (2010) and Yang (2012) noted that the cloud approach is not always a cost saving solution. At present, more understanding of cloud issues is required to

臺灣大學圖書館 期刊資料庫

help communities in Taiwan in to address the reality of using a cloud. A library management system consists of a number of modules. Of these, the cataloguing module is the most complicated and there is a wide gap between systems development (i.e. LMS vendors) and real world cataloguing practices (i.e. libraries), which results in a longer period of time for system implementation. The purposes of this project are: 1) to determine the availability and feasibility of a library cataloguing platform that is based on cloud computing infrastructure; 2) to reduce dependence on library management systems; 3) to build a cloud resources pooling platform with an on-demand self-service infrastructure for libraries and LMS vendors; 4) to develop a SOA and APIs procedure to access heterogeneous library bibliographic records and 5) to provide a cloud computing cataloguing service to libraries, academia and LMS vendors for future use.

B. A library cataloguing platform using cloud computing

In future, more worldwide LMS vendors will host cloud applications at their end and customers will access those applications using the Internet as a communication channel. However, academic libraries in Taiwan have found themselves in the embarrassing situation of not being able to embrace a LMS cloud solution. There follows a discussion of the two main reasons why a local or western LMS cannot properly deploy cloud services in higher education libraries in Taiwan.

1. Complex bibliographic formats

Most library bibliographic data are stored in one of the implementations of ISO 2709, known as MARC, which allows the worldwide sharing of records that are created in one library. However, there are post-processing difficulties in displaying the bibliographic data. For example, there are multiple internal codes, such as Chinese Character Code for Information Interchange (CCCII) and Big5, which are the most widely used Chinese internal codes in libraries in Taiwan (Mao and Hsu, 2006). Although CMARC is closely linked with UNIMARC, issues with multiple internal codes, that is, conversion between CCCII, Big5 and utf8 (Unicode), and complicated MARC formats, that is, conversion between MARC formats and other schema such as Dublin Core, could be potentially complex for both libraries and LMS vendors. As an example, the first author's affiliation, the Tatung University Library was established in 1956 with two hundred thousand holdings. The Library has been using an out-of-date Dynix systems product, since 1992. The systems use CCCII as a core-code for library operation, such as cataloguing and circulation, and auto-convert the CCCII code to Unicode when displaying in OPAC.

The MARC formats used in higher education libraries in Taiwan are complicated.

臺灣大學學術期刊資料庫

CMARC3 is the most widely used MARC format in the library community, even though the fourth edition was published in 1997 and the later CMARC4 was updated in 2001. In addition, in order to avoid any loss of information during the conversion between MARC formats, most libraries process Chinese materials using CMARC and use USMARC/MARC21 for Western material (Hsu et al., 2011). In December 2011, the National Central Library in Taiwan decided to move from CMARC to MARC21 and began to produce MARC21-only bibliographic records. It has implemented RDA for new Western materials, starting in March 2013 (NCL, 2013). However, moving to MARC21 involves a substantial investment in time, money, labour and technology. More than half of the libraries have taken a wait-and-see approach to a MARC21 transfer (Chang, 2012). This has increased the difficulties in promoting RDA in Taiwan. The structure and future development of RDA are also very different from AACR and Chinese Cataloging Rules, so the majority of libraries are reluctant to use RDA for their cataloguing operations (NCL, 2014).

Because of budgetary limitations, it is difficult for libraries to make plans for the transfer of Chinese or Western bibliographic formats from CMARC3 to MARC21, to begin a trial project to map records to FRBR, or to begin a RDA test bed to evaluate the workflow before formally adopting the scheme for their cataloguing operations. As most libraries do not have adequate in-house technical support, if the LMS vendors provide satisfactorily customized services, including hosting, migration assistance, staff training, support, software maintenance and development, the cost to the vendors increases. This is true in the case of the Tatung University Library. The Library recently bought the Koha library system (in Unicode environment) for several reasons. Firstly, Koha is a mature, integrated library system with many advantages. The Library also uses MARC21 to deal with all types of resources and Koha provides the default MARC21 template. Lastly, an open-source LMS, like Koha, is the only option, with very limited budget and insufficient in-house technical support.

2. Ownership vs. leasing

In a traditional computing environment, libraries have total control of their software and hardware. The data stays within the institutions. In a cloud computing environment, some of the control is surrendered to the cloud service providers. This is not acceptable, psychologically and practically, for libraries. Psychologically, Chinese people require stronger control over their assets than Western citizens. Practically, in the inherent accounting structure, the physical parts of library holdings, as well as software and IT infrastructure, are items listed in the asset category, according to rules made by the government. However, the

臺灣人學學術期刊資料庫

cost of a subscription-based cloud service is a recurring cost in library expenditure. This would be an extra burden, when library budgets are reducing annually, so a cloud solution might not save libraries money (Han, 2010).

In addition to these two obstacles, libraries routinely implement a number of applications that require separate management processes and IT platforms, which results in a relatively complex environment that the libraries have to manage. From the economic point of view, an SOA- and APIs-based cataloguing cloud platform with readily accessible and sharable resources, such as the MARC format transfer mechanism, and a workable cataloguing interface could be beneficial. The platform supports for task workflows could be used to ensure task-based tactical deployment of solutions with adequate functions, would be faster to deploy, easy to upgrade and would be cheap.

This project seeks to develop the first Chinese cloud computing bibliographic data processing center, for use by libraries, academic and commercial communities in Taiwan, for cooperation and community building. The platform is based on library standards compliance, which allows libraries to do cataloguing online and to access resources or to download bibliographic data. It covers functions including: 1) creating, editing and deleting bibliographic data; 2) classification and cataloguing; 3) supporting conversion, such as MARC, Dublin Core, FRBR and RDA; 4) cataloguing data publicity and access and 5) the use of remote call handling data.

II. System platform design

This project uses the cloud computing infrastructure in the National Center for High-Performance Computing (NCHC) (2014) at the National Applied Research Laboratory, to which one of the co-authors is affiliated. The project integrates authors' previous research results using multiple MARC formats and a double byte character set (Chang et al., 2010) and the FRBR research (Chang et al., 2013). The project has two stages. Firstly, constructing the cloud computing infrastructure includes setting up hardware facilities, integrating Internet resources and using a set of widely accepted free or open source virtual technologies. Secondly, deploying the cataloguing technology includes the development of software, using developers and a team of library professionals to reduce the technology gap and to ensure that library needs are taken into account.

A. Cloud computing infrastructure

A number of underlying technologies make cloud computing possible. The main enabling technology is virtualization (wiki, 2014). In this project, virtualization means platform virtualization, which refers to the creation of virtual machines that act like real

臺灣人學學術期刊資料庫

computers with an operating system. This is performed within a cloud environment across a set of services, using a hypervisor or virtual machine monitor (VMM), which sits between the hardware and the operating system. The advantages of the infrastructure are that it allows each one or more virtualized operating systems to perform simultaneous work, it simplifies management by consolidating servers and security and reliability and device independence are improved by the use of the hypervisor architecture. This load balancing architecture gives crucial advantages, in terms of wide availability and high scalability in a cloud computing environment. The cloud cataloguing system that uses this architecture, which is outlined in Figure 2, allows the virtual machine to continue operations if the host machine fails. A number of hypervisor technologies are currently available for open platforms. Younge et al. (2011) studied the four most popular and feature-rich of these and concluded that the KVM hypervisor is the best overall choice of a fully featured virtualization technology for use within high performance computing cloud environments.

After an analysis of the functional needs for load balancing and the use of more effective risk management, the distributed open platform, Linux Ubuntu, version 11.20X86_64bit, was chosen as the operation system, because it provides full KVM in a virtualization environment. This project also uses Quick EMUlator (QEMU) as a CPU emulation manager, to execute multiple virtual CPUs in parallel for user-level processes and to act as the communication channel with the host machine. Linux Ubuntu uses KVM as

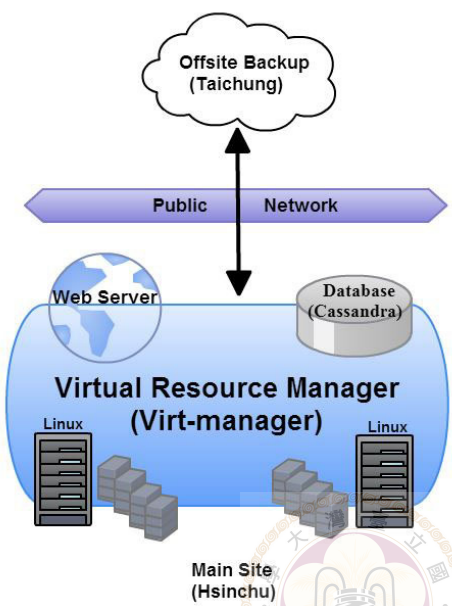


Figure 2: The load balancing architecture for the project

臺灣大學學術期刊資料庫

the back-end virtualization technology and the virtual machine manager, virt-manager, as the front-end, to manage virtual machines. Figure 3 shows the graphical interface of virt-manager. At the front-end, there is Libvirt, which is a virtualization API that interacts with the virtualization processes in the Linux platform.

The architecture deployed has two sets of central processing units, Intel Xeon 48 processors, as host machines, each with a 512GB system memory and 10TB of storage space. The host machines have four virtual operating systems in each processor. Each virtual machine has two central processing units, each with a 8GB system memory and 1T of storage space. The data is stored is separate geographical locations, in NCHC's Hsinchu Headquarters and in the Taichung Branch, which serve as the offsite backup and the Virtual Private Network (VPN). In order to manage the virtual machines effectively at different sites, the project uses two open source solutions, developed by the Free Software Lab at NCHC (2010). Clonezilla (2012) is used to provide simultaneous large bare metal backup and recovery, either locally or remotely. The DRBL (Diskless Remote Boot in Linux) (2012) is used to manage the deployment of the Linux operating system across client-side users. The DRBL uses distributed hardware resources that allow client-side users to fully access local hardware. The software were chosen because one of the co-authors is a member of the software development team.

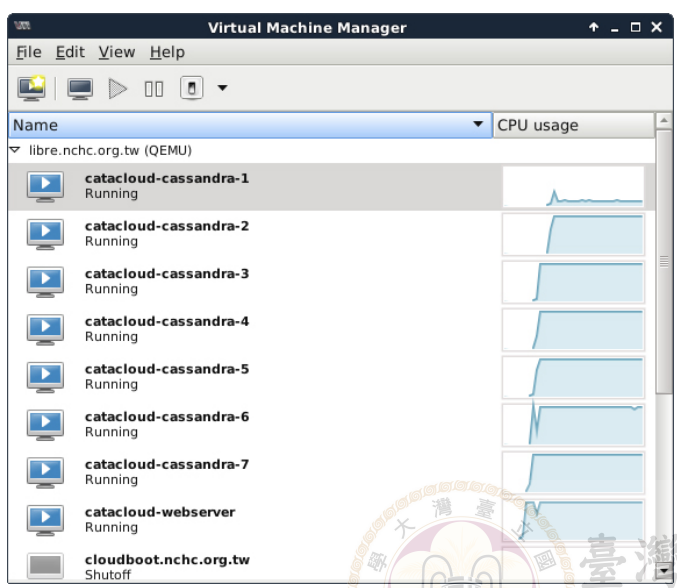


Figure 3: The graphical interface of virt-manager

臺灣大學學術期刊資料庫

1. NoSQL solution

Cassandra is an open source distributed NoSQL (Not just a SQL) that uses a key–value to store millions of data records. This is particularly useful for working with a huge quantity of library data, where the data's nature does not require a strong relational model. As shown in Table 1, the publisher and the year of publishing are grouped as a column with a sequential number (SN) 5. The sequential numbers, 2-5, give information about the book title, the author, the ISBN and the publisher, which are grouped as a column family (CF).

A column can be added to one or multiple rows at any time. This is also very useful for FRBR. When processing the conversion of MARC data elements to the FRBR model, using the mapping algorithm that is provided by the system, the system clusters data by creating structural FRBR work-sets with appropriate attributes and values. This is similar to Cassandra's data structure (columns), up to the root of the tree (the cluster). In Cassandra, the keyspace is a namespace that contains column families. This project defines a keyspace, "catalogue", with a column family named "record". The Cassandra::Simple module is used to manage records, using the Cassandra API.

2. Data structure

This project uses MySQL to manage the cloud website. The reference tables are shown in Figure 4.

The function of each table in the database is detailed in Table 2.

B. Deploying the cataloguing technology

The platform enables users to perform cataloguing, using an intuitive, web based interface, as shown in Figure 5. The domain name (http://catacloud.nchc.org.tw) is still under application. The system uses a standard web browser. The project incorporates MARC/Perl (n.d.) modules into the

Table 1: Example of Cassandra bibliographic record

SN	Value1	Value2	Value3	Value4	…
1	a	SN	1		
2	a	Title	Complete guide to GCC		
3	a	Author	Kurt Wall		
4	a	ISBN	95752758089		
5	a	Publisher	Bosho	2008	
6	b	...	...		

臺灣大學學術期刊資料庫

Table 2: Functions of tables in the database

Table	Function
Metadata	The Metametadata records all of the processes for bibliographic data, including the bibliographic format, the title, the author, the ISBN/ISSN, the processing time and the processing behavior.
Type	The Type is the ID for the bibliographic format and the name of the bibliographic format.
Template	The Template is used as the house bibliographic format for libraries or commercial users.
Library	The Library holds the user IDs of libraries or commercial organizations and their names.
User	The User holds the user IDs and their names.
News	News is used for announcements.
Alternative_cas_data	The Alternative_cas_data synchronizes data in Cassandra for backup.
[cmarc3|cmarc4|marc21|frrb|dc] structure	These are the bibliographic formats and their fields, in detail.

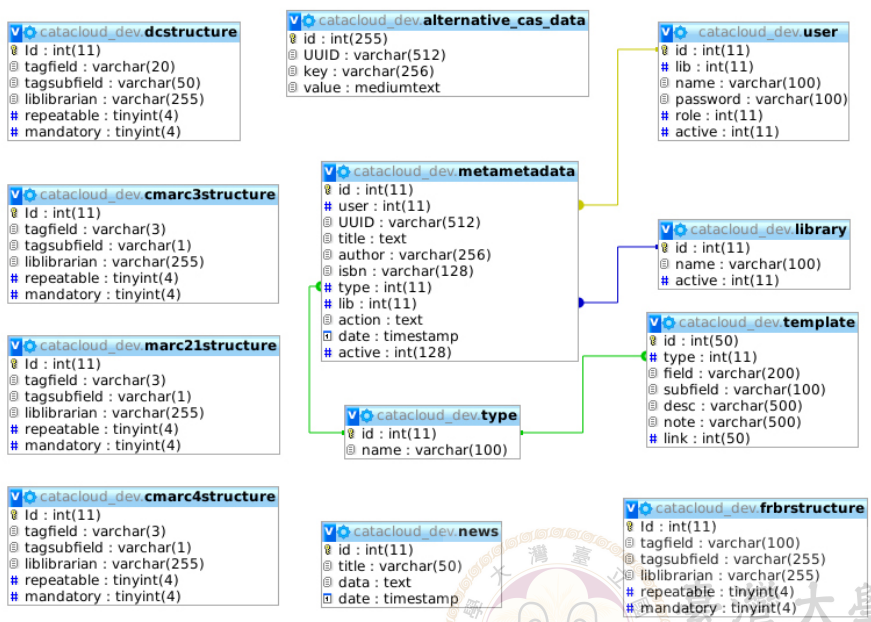


Figure 4: Reference tables in the database

臺灣大學學術期刊資料庫

cataloguing environment, which is entirely under Perl, to reduce the effort expended on system development.

The MARC modules are used to manage library records, as described below.

add/edit/delete records

The Perl based MARC::Record framework provides a scalable approach for the processing of MARC data in the Perl system environment. In the framework, MARC::Batch handles the files of MARC::Record objects, MARC::Field handles MARC fields and the MARC::File handles MARC files. Below is an example of the addition of a new field using the MARC::Field module.

```
my $author = MARC::Field->new(
                100, "1", " ", a => "Arnosky, Jim."
                );
$marc->add_fields( $author );
```

The project uses NET::Z3950, to allow efficient search and retrieval of MARC records from the two most comprehensive and most-used Taiwanese record databases: the National Bibliographic Information Network (NBINET) and the OCLC WorldCat. MARC::Crosswalk::DublinCore provides the XML file for Dublin Core. A CMARC3 to MARC21 conversion system has been developed by the National Central



Figure 5: The login view of the cloud cataloguing platform

臺灣大學學術期刊資料庫

Library in Taiwan (NCL, 2009). The results of authors' previous studies with FRBR tools show that the LibFRBR platform supports FRBR records, as shown in Figure 6. The FRBR tools include LibFRBR, an application function library used to convert bibliographic records into FRBRized structures in Koha, and the FRBR Cataloguing Interface (Chang et al., 2013). The Chinese version of RDA will be available at the end of 2014 and will be published in 2015 by the National Central Library, so the RDA will be a continuation of this project and the results will also be shared through the platform.

1. Remote call handling data

An open platform for data access and message transmission is developed using HTTP and Simple Object Access Protocol (SOAP). Users set up application programming for online cataloguing and make use of the platform resources. The library and the LMS vendor communities can also share their custom applications across the platform. A summary of the method of SOAP API is shown in Table 3.



Figure 6: Processing a FRBR record

臺灣大學學術期刊資料庫

Table 3: Method summary of SOAP API in the study

doAuthenticate()	◆an authentication interface to process the http authentication ◆Users require a username and password to carry out the subsequent activity
ListBiblio($type = '', $limit = '')	◆for listing bibliographic data and individual UUID's ◆Users provide a type parameter with a bibliographic format ID ◆Users provide a limit parameter and limit the result to a query using a LIMIT clause
AddBiblio($type, $data)	◆for adding new bibliographic data into the system ◆Users provide a type parameter with a bibliographic format ID ◆Users provide a data parameter with bibliographic data
EditBiblio($ouuid, $type, $data)	◆for editing bibliographic data ◆Users provide a UUID with the bibliographic data to be edited ◆Users provide a type parameter with a bibliographic format ID ◆Users provide a data parameter with edited bibliographic data
DeleteBiblio($uuid, $type)	◆for deleting bibliographic data ◆Users provide a UUID with bibliographic data to be deleted ◆Users provide a type parameter with a bibliographic format ID
DetailBiblio($uuid)	◆for accessing bibliographic data ◆Users provide a UUID with bibliographic data to be accessed
GetTemplate($library)	◆for choosing a template to edit house bibliographic records ◆Users provide a library ID

臺灣大學學術期刊資料庫

III. Discussion and evaluation

Wide availability and data safety are critical in a cloud computing environment. This project minimizes the security concerns about host stability and data recovery capability by using virtualization technology and by locating the Cassandra databases in widely geographically distributed sites. However, the administration of a cloud system administration presents the same technical and business issues as locally deployed systems. The following paragraph discusses the technical and business challenges that the technical development poses.

A. Server downtime

Server failure is often automatic and usually occurs without warning, which can lead to loss of service and loss of data. This project uses virtualization technology to ensure wide availability and high reliability and to provide failure capability in servers when continuous availability is required. If there is to be no single point of failure, the libraries or LMS vendors ensure high availability by mirroring their data in the traditional servers.

An experiment involving multiple users from authors' institutions was conducted during the period of summer 2013 to February 2014, using the cloud platform shown in Figure 7.

The graph shows that a significant portion (60%-75%) of the user traffic on the cloud platform is as a result of uploads or other cataloguing operations. The user traffic reaches its peak in the summer and winter breaks, as libraries have more time to practice MARC format conversions and to experiment with FRBR/RDA mappings and other new

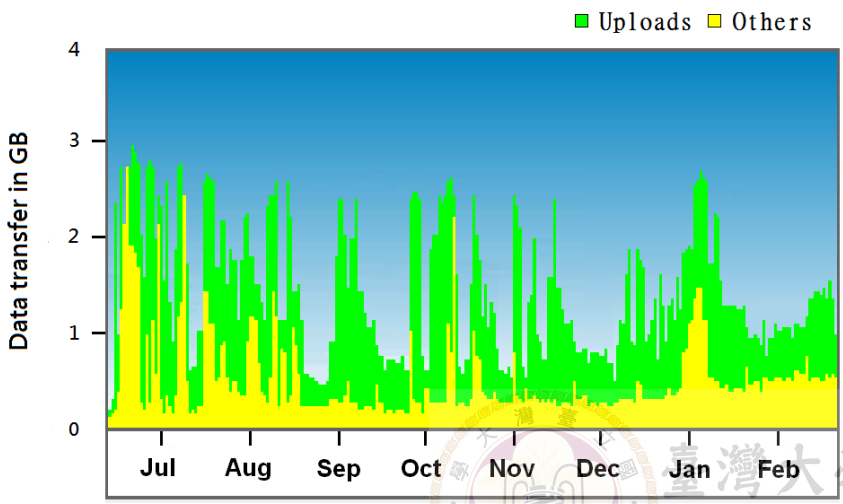


Figure 7: User traffic

臺灣大學學術期刊資料庫

features. The average traffic load per month is around 1 GB. A technology assessment was also performed, using 200 bibliographic records, which was supported by SRIS, which is the only Koha commercial service provider in Taiwan, because of the very small market. They have created a concise trial version of RDA framework for Chinese environments that are ready for libraries to experiment with. It consists of the basic RDA tags, 336, 337, 338 and a few others. An official Chinese RDA User Guide is scheduled for release in early 2015. No negative feedback on hardware/software capacity was noted during the experiment.

B. Performance issues

The NCHC has a national-level remote backup strategy in three business sites in the north, center and the south of Taiwan, to avoid a disaster resulting from the loss of information. This project uses the NCHC's two interconnected business sites (Hsinchu and Taichung) for offsite storage, replication and backup, so the libraries or LMS vendors that use the cloud have guaranteed reliable storage.

C. Internet service outage

The NCHC is Taiwan's only national-level supercomputing center. It provides High Performance Computing (HPC) services to science, research and other interdisciplinary fields. Logically, the national-level service is not easily suspended or interrupted.

D. Internet bandwidth

The NCHC developed the TaiWan Advanced Research and Education Network (TWAREN) (2009), which uses advanced all-optical network technology to provide timely and rapid data recovery. Its domestic backbone is connected to the Taiwanese Academic Network (TANet), which serves higher education establishments in Taiwan. The individual institutional network bandwidth is a 1Gpbs and IPv4/IPv6 dual stack network service environment, so it is sufficient to support a cloud computing service.

E. Network problems caused by instability

The cloud cataloguing system works using a web browser, so if there is a network failure, cataloguing work cannot proceed via a HTTP protocol. A possible solution is that, currently, browsers are HTML5 compliant, so offline processing is possible. A small portion of the cataloguing work could be completed offline, but this is not possible for all of the cataloguing work. Another alternative could be that libraries or LMS vendors manage offline data using the SOAP provided on the cloud platform, to provide an ongoing cataloguing service.

臺灣人學學術期刊資料庫

F. Cost analysis and sustainability

Cost analysis indicators are present for the areas of hardware, software and human labor. As mentioned in Section 2.2, the NCHC supports this research with their virtual host, which reduces the hardware expense enormously. The project also uses widely accepted free software or open source in all areas and authorizes source code sharing that is developed by this project in Github (2014), a web-based hosting service for open source projects, distributed as a GPL for further application. However, the cost is large, in terms of human labor hours and software integration and specialization. It is seen that using Cassandra significantly reduces the human labor burden by using multiple data centers and real-time analysis of the growing amounts of data. The government will continue to invest in the NCHC as an ongoing sustainable policy, because TWAREN at NCHC plays a critical role in providing network connections to academic research institutions that foster the creativity and talent that Taiwan needs to maintain competitiveness in the new world economy.

We undertook a trial cost assessment for the case of the Tatung University Library, based on the pricing of hicloud CaaS (Computing as a Service) by Chunghwa Telecom (CHT), Taiwan's largest provider of fixed line services, mobile services, broadband access service and Internet service. The hardware cost involves the leasing of hicloud CaaS Linux CentOS6.3, 64bit, standard model M, equipped with 2HCU(GHz), 4GB RAM, HD 100GB, 1 individual exclusive IP at USD 0.157 per hour. (note: each Hicloud Computing Unit (HCU) adapted in the standard model M has a computing capability that is equivalent to that of 1.0 GHz 2007 Xeon processor.) Staffing cost is USD1,000/year and 5 percent of a system administrator's time to manage the server, at a USD20,000 annual salary. The cost assessment demonstrates that the budget is reduced, which should encourage library professionals and LMS vendors to proceed with innovative initiatives.

IV. Conclusion

Cloud computing technology has seen significant development in the government, commercial and academic sectors in Taiwan, but no research or practically workable cloud solution has been provided for the library sector that is suitable for use in a Chinese cataloguing environment. The project demonstrates the originality of a SOA and APIs based Chinese cloud cataloguing platform with an architecture containing features of interoperability and aggregation of heterogeneous resources and applications that can be shared on the Internet, a means of possible reducing the cost of the service through reuse, a shared federation of resources and standards compliance. The SOA and APIs based platform gives libraries

臺灣人學學術期刊資料庫

and LMS vendors the flexibility to start a cataloguing task quickly, without the need for large investment in complex computing infrastructure. This platform also provides access to resources and a functionality to create new services, beyond those already delivered. Small institutions can also benefit from the savings that arise from its use. The cloud cataloguing platform is still under construction and not yet opens to the public. This project gives a working library cataloguing platform, initially for research purposes, which will eventually serve as a model for 21st century software on demand and cloud computing platforms. The results of this research will define the evolving role of the library, in terms of cloud computing. When the platform matures, the library community and the LMS vendors could establish a consortium to share cloud experiences, recommend solutions and give cooperative timely help and technical support.

Acknowledgement

Financial support for this research from the National Science Council, Taiwan, under the grant NSC 100-2410-H-036-002 is gratefully acknowledged.

References

Akerman, R. (2007). Library service-oriented architecture to enhance access to science. *The 28th annual conference of the International Association of Technological University Libraries (IATUL)*, June 11 - 14, 2007 at the Royal Institute of Technology, Stockholm, Sweden, Retrieved June 15, 2014, from eprints. rclis.org/10079/1/IATUL2007-RichardAkerman.pdf

Breeding, M. (2010). Opening up library systems through web services and SOA: hype or reality? *Library Technology Reports*. Retrived from ALA TechSource.

Breeding, M. (2012). Current and future trends in information technologies for information units. *El Profesional de la Informacion*, enero-febrero, *21*(1), 9-16. doi:10.3145/epi.2012.ene.02

Chang, N., Tsai, Y. and Hopkinson, A. (2010). An evaluation of implementing Koha in a Chinese language environment. *Program: electronic library and information systems*, *44*(4), 342-356. doi:10.1108/00330331011083239

Chang, N., Tsai, Y., Dunsire, G. and Hopkinson, A. (2013). Experimenting with implementing FRBR in a Chinese Koha system. *Library Hi Tech News, 30*(10), 10-20. doi:10.1108/CHTN-09-2013-0054

臺灣大學學術期刊資料庫

Chang, H. (2012). The Survey of usage of MARC21 and RDA of libraries in Taiwan. *MARC21 and RDA Forum*, National Central Library, 23 November, 2012.

Clonezilla (2012). *The Free and Open Source Software for Disk Imaging and Cloning.* Retrieved June 15, 2014, from http://clonezilla.org/

Collins, M. and Rathemacher, A. (2010), Open forum: the future of library systems. *the Serials Librarian, 58*, 167-173. doi:10.1080/03615261003625703

Dempsey, L. (2010). *Discovery layers: top tech trends 2*. Retrieved June 15, 2014, from http://orweblog.oclc.org/ archives/002116.html

DRBL (2012). *Diskless Remote Boot in Linux.* Retrieved June 15, 2014, from http://drbl.sourceforge.net/about/

Github (2014). *Catacloud.* Retrieved June 15, 2014, from https://github.com/Thomas-Tsai/catacloud

Goldner, M. (2011). Winds of change: libraries and cloud computing. *Multimedia Information & Technology, 37*(3), 24-28.

Goldner, M. and Birch, K. (2012). Resource sharing in a cloud computing age. *Interblending & Document supply, 40*(1), 4-11. doi:10.1108/02641611211214224

Han, Y. (2010). On the Clouds: a new way of computing. *Information Technology and Libraries*, June 2010, 87-92.

Hsu L., Cheng, Y. and Niu. H. (2011). The research and analysis of CMARC3 and MARC 21 bibliographic format mapping: a case study of indicators and codes. *National Central Library Bulletin, 100*(1), 99-132.

Kuali Foundation. (2013). *Open Library Environment.* Retrieved June 15, 2014, from http://www.kuali.org/ole/modules/systems-integration-entity

Lloret, N. (2012). "Cloud computing" in library automation: benefits and drawbacks. *The Bottom Line: Managing Library Finances, 25*(3), 110-114. MARC/Perl (n.d.), Retrieved June 15, 2014, from http://marcpm.sourceforge.net/

Mavodza, J. (2013). The Impact of Cloud Computing on the Future of Academic Library Practices and Services. *New Library World, 114*(3/4), 132-141. doi:10.1108/03074801311304041

Mell, P and Grance, T. (2010). *NIST Definition of Cloud Computing v15*. Retrieved June 15, 2014, from http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc

Mitchell, E. (2010). Using cloud services for library IT infrastructure. *The Code4library Journal*, Issue 9, Retrieved June 15, 2014, from http://journal.code4lib.org/articles/2510

NCHC (2010). *The Free Software Lab at the NCHC.* Retrieved June 15, 2014, from http://free.nchc.org.tw/pmwiki/

臺灣大學學術期刊資料庫

NCHC (2014). *HPC services.* Retrieved June 15, 2014, from http://www.nchc.org.tw/en/services/supercomputing/

NCL. (2009). *The National Central Library Machine readable cataloging format converting system.* Retrieved June 15, 2014, from http://catbase2.ncl.edu.tw/c2m21.2/index_login.php

NCL. (2013). *The National Central Library RDA Working Group Report on the RDA Implementation, the 12th meeting on 11 April 2013*. Retrieved June 15, 2014, from http://catweb.ncl.edu.tw/portal_b1_page.php?button_num=b1&cnt_id=409

NCL. (2014). *The survey results of adopting RDA of libraries in Taiwan.* Retrieved October 5, 2014, from http://catweb.ncl.edu.tw/portal_b1_page.php?button_num=b1&cnt_id=448

Pace, A. (2009). 21st century library systems. *Journal of Library Administration, 49*(6), 641-650. doi:10.1080/01930820903238834

Peters, C. (2010, March 6). *What is cloud computing and how will it affect libraries?* Retrieved June 15, 2014, from http://www.techsoupforlibraries.org/blog/??what-is-cloud-computing-and-how-will-it-affect-libraries

Rapp, D. and Fialkoff, F. (2010). Corner office interview: SirsiDynix's Rautenstrauch and Nelson. *Library Journal, 135*(16), 32-33.

Serials Solutions launches Aqua Browser SaaS Version. (2010, March 24). *Library Technology Guides.* Retrieved June 15, 2014, from http://www.librarytechnology.org/ ltg-displaytext.pl?RC=14657

Shimshock, G. (2010, January 17). *Encore offers new article integration capabilities.* Retrieved June 15, 2014, from http://www.iii.com/news/pr_display.php?id=451

Trehub, A. and Wilson, T.C. (2010). Keeping it simple: the Alabama Digital Preservation Network (ADPNet). *Library Hi Tech, 28*(2), 245-258. doi:10.1108/07378831011047659

TWAREN. (2009). *TaiWan Advanced Research and Education Network.* Retrieved June 15, 2014, from http://www.twaren.net/english/AboutTwaren/Initiative/

Wikipedia. (2014). *Virtualization.* Retrieved June 15, 2014, from http://en.wikipedia.org/wiki/Virtualization

Wilson, K. (2012). Introducing the next generation of library management systems. *Serials Review, 38*(2), 110-123. doi:10.1016/j.serrev.2012.04.003

Yang, S. Q. (2012). Move into the cloud, shall we? *Library Hi Tech News, 4*(1), 4-7. doi:10.1108/07419051211223417

Younge, A. J. et al. (2011). Analysis of virtualization technologies for high performance computing environments. *2011 IEEE 4th International Conference on Cloud Computing* (pp.9-16). doi:10.1109/CLOUD.2011.29

臺灣大學學術期刊資料庫