Cloud Computing in KAUST Library: Beyond the Remote Hosting

Author:
Yi Yu, systems librarian
Library of King Abdullah University of Science and Technology, Saudi Arabia
yi.yu@kaust.edu.sa

Abstract:
Enterprise computing is the key strategic approach for KAUST to build its modern IT landscape. In such a strategic direction and technical environment, the library tries to establish library technology by catching new trends which also make the library to function more efficient and sufficient. This paper tells the story about cloud computing development in KAUST library. It uses real world scenarios and first-hand experiences to describe what the cloud computing means for KAUST library, what the difficulties we met during the implementation, how it affects library functional performance and work procedure, how it impacts the style and modal of library technical service and systems administrations, how it changes the relationships and cooperation among library, campus IT and vendors, what the benefit and disadvantage we get. The story will share the knowledge and lessons that KAUST library learnt during the cloud computing development, and will also point out the future direction of the cloud computing which KAUST library supposes toward.
**Introduction**

King Abdullah University of Science and Technology (KAUST) is a new science university opened in Saudi Arabia in September of 2009. Named after the nation’s leader whose vision for a new “House of Wisdom” as one of the world's great institutions of research, “KAUST”, as the university is called, removed its scaffolding and cranes to reveal an ultra-modern campus on the shore of the Red Sea -- an ancient fishing village called Thuwal beside the thriving metropolitan city Jeddah. Just like its parent organization, KAUST Library is a brand new library which was built from scratch. During 4 years development, KAUST library successfully grew into a medium sized academic and research library which includes 25 staff members, over 10,000 print collections and about 50 subscription of electronic resources covering the major science databases and e-journals. Besides presenting the new generation library by its spacious and artistic modern architecture, the library was also equipped the most advanced facilities and library systems.

As enterprise campus is the key strategic approach for KAUST, enterprise computing naturally becomes a direction to build its modern IT landscape. In such strategic and technical environment, KAUST library tries to establish library systems by catching new trends and maximally leverage the benefits of utilization of the new technology. Go to cloud is one of these efforts and achievements.

**What is cloud computing and why go to cloud?**

The term “cloud computing” is more like a metaphor instead of an accurate definition. As the word “cloud" means "the Internet" conventionally, the phrase cloud computing presents "a type of Internet-based computing," where different services -- such as servers, storage and applications -- are delivered to an organization's computers and devices through the Internet.\(^i\)

For the popularity of the term, it can be attributed to its hosted services in the sense of application service provisioning that run client server software on a remote location.\(^ii\) That is what KAUST Library cloud computing practice focuses on – however, the experiences and lessons we got not just remote hosting service.
From the very beginning, as one of the major server applications in the campus, KAUST library system was hosted locally in KAUST IT Data Center. When cloudy computing became a new approach for campus IT infrastructure, we were thinking of the possibility and feasibility of using cloudy computing for library systems. After a discussion, we thought that since Internet connections are no longer a barrier, there are many benefits for the library to apply cloudy computing in our surroundings. First of all, by switching to the remote servers hosted by our vendor, we are able to have more dedicated and professional technical support for the library. Secondly, full package services such as upgrading, database maintenance and backup will allow the library’s technological staff to do more strategic activities rather than routine operation tasks. The benefit will also include reducing the mistakes due to misunderstanding between vendors and campus IT, quicker trouble-shooting and problem solving, and many other advantages. Plus, Due to special security concerns in this special university, putting all of the ILS application in the cloudy will be beneficial for the security of the entire campus network.

**How to implement?**

The first step was to migrate server to vendor hosted location. As many tasks needed to do such as set migration schedule, change DNS entry, open port for remote host, update IP address for electronic resources, etc., many units among the library, campus IT and hosting vendor were involved, many people needed to work together to make the migration success. Even we did tremendous work in communication and coordination for preparation, as well as tired to predict the potential problems as much as possible, there were still unexpected situation occurred during the period of the migration. For example, the window of the server shutdown was chosen based on KAUST IT CIO’s suggestion and agreed by the three parts of vendor, campus IT and library, however, it incurred a great complain from reference team since they scheduled a tour for new employee and arrange a demo of library online catalog in that day. For the special security reason, many port in campus network are closed unless you request it open through certain procedure, however, this situation had never been experienced by the hosting vendor before. Thus after the migration, we could not access library online catalog although it was no problem to connect server through client application. It caused a big panic in the library. Because updating IP address of electronic resources was not list in the migration “to do list” from the
vendor, we missed to inform the e-publishers to change the IP address accordingly. It made the migration totally failed, thus we need rollback to old server. When everything seemed ok for the second time of migration, we fond the self-check machines did not work as the new server IP was not be able to be updated due to the admin interface could not display. After couple days wrestling, we found that it was the installation of new anti-virus software which blocked the display. Even all the connection from the public interface looked normal, there were still emergent reports poured to the systems librarian: many staff members in the library were complaining that their client-server connection was broken. Actually that was due to they left a space when they type the new server IP in the configuration file! Such panic and chaos came one after another during that days. “Fail miserably in a very easy task” is not a cliché. In this situation, preparation in great care, coordination in sincerity, and communication in detail are the keys for making process easy. Certainly, solid technical knowledge would give a great help to get things done.

What problem in the cloud?

After our integrated library systems moved to remote datacenter hosted by the system vendor, the library staff members, who use different modules of the application to process library routine work like catalog, circulation, acquisition, etc., found that the session would automatically disconnect if they did not act the process for several minutes. This situation never occurred when KAUST IT locally hosted the server. The user was really annoyed since he/she always needed to restart the module even it only was inactive for a couple of minutes.
Figure 1 - disconnection message from client application of library system

For debugging, KAUST IT network team installed a network packet analyzer called Wireshark on several workstations and had it closely monitor the record of disconnection pattern and understand the transaction behavior, as well as to capture the packet data for identification of the cause on the packet and protocol level. After investigations based on the captured data, the inter network situation seemed getting clear:

- The disconnection occurs intermittently and in a random pattern;
- The Client-Server communication uses a keep-alive mechanism, on ports other than the ports used by the SCCP protocol;
- Protocol inspection for the SCCP protocol was required by the application vendor to be turned off on the firewall by KAUST IT;
- Packet captures on both the server side (Internet) and the client side (Inside) had shown TCP RST received on both sides.
This tells that a device in between is sending RST frames in both directions to terminate the session after a certain timeout period. IT Security increased the period after which idle open TCP sessions are reset to 8 hours (60 minutes is the default). However, the disconnection frequency is much less than 8 hours, therefore, this could not be the cause.

**How to resolve?**

For figuring out which part of the whole chain caused the disconnection, KAUST IT examined potential nodes by identifying internal and external traffic paths.

![Figure 2 - traffic path, inside is marked green while outside is marked red.](image)

The nodes to be tested in the outside path are:
- International Gateway (IGW)
- KAUST MPLS Core

The nodes to be tested in the inside path are:
- Firewall
- Anagran
- F5 Load Balancer
An exclusive method was used to investigate. The results show all the test machines which were placed right before IGW, the Virtual routing and forwarding (VRF) and firewall were not experienced the disconnection, while reference machine on the Data VLAN was experiencing the disconnections. Thus it is reasonable to conclude that it is F5 load balancer caused the disconnection was caused.

**The disconnection problem comes again**

However, after about a year, in the August of 2014, the disconnection problem came again. When we reported this issue to the vendor Innovative Interface, they told that the disconnection could be related to the current 2011 1.4 release. They said that they discovered in this release that Linux did not handle small network glitches or disconnects resulting in the error listed. They had later improved this functionality in 2011 1.5 and above as first step which in similar cases has proven to be successful. They suggested us to upgrade Millennium to the latest release 2011 1.6. After the upgrade, they advised us to clear the jar cache for every affected workstation. However the disconnection problem still was not resolved by this procedure. In this situation,
we experienced another, much more painful investigation and debugging process.

First of all, we coordinated vendor Innovative Interface and campus IT and let them work together for the investigation. As the vendor could not replicate the error from remote access, we narrowed cause of the disconnection wherefore from campus network. However, the people of KAUST IT network insisted there was no change for the campus network. Thus we went back the vendor for help. As the same as before, the vendor suggested us to use WireShark checking TCP communications, especially checking for the use of packet shaping devices on our network \(^1\), such as F5 Load Balancer. The expert of the vendor also advised us to turn the ALG feature off as Juniper firewalls create Application Layer Gateways(ALG) for some applications and it can mistake Innovative traffic for SCCP traffic.

During the period of approaching campus IT security team for checking the firewall settings, we got the information from IT network team said that the F5 load balancer was no longer in the path to the Internet. That means there was no packet shaping issue in the path. Thus one possible cause of the disconnection was excluded. The followed question was: does Juniper firewall cause the disconnection? Quickly, the security team confirmed us that ALG was disabled on our firewalls for SCCP. However, we found that the disconnection still happened randomly.

The investigation followed by was tedious and distressful. Simply saying, KAUST IT with Innovative work together with a join session. The idea was KAUST IT collected trace logs on firewall, along with using WireShark traced logs collected on client PC, Innovative collected the logs on server side. Then the two sides correlated the logs to identify which segment is causing the issue. The captured records from the WireShark which installed in the client workstation caught the disconnection information. However, The WireShark that Innovative ran on the server did not reveal any errors or issue at the same time. We sent the log file on client side to Innovative for investigation. The expert of the vendor pointed two signs of the error evidence: 1) bad checksums (IPv4) - a sign that there was TCP-offload happening on the NICs below the

\(^1\) These devices are typically used to optimize network performance by restricting or delaying packets that meet certain criteria that are unrelated to the mission of the institution, such as Gnutella traffic.
layer WireShark captures packets on the client side (see the next image-1); 2) Cisco firewall devices altered some packets at times which might create disconnection.

Image -1: Error information shown on Wireshark analasys tool

The expert of Innovative thought the issue appeared to be local from the PC that we tested on, and it looked like a SCCP “fixup” was altering the traffic\(^2\). The vendor recommended that KAUST IT firewall administrator configuring the Cisco products to disable SCCP “fixup”, and also check if there was any limitation of the number for the connected session in the firewall. The response from KAUST IT was quite disappointed: 1) there was no “fixup” configured for the related ports on firewall; 2) there is no limitation of sessions on the firewall for the related ports. During another round of TCP dump checking, the expert of vendor found out there was a few of TCP Retransmission’s from the client to the server\(^3\). However, the KAUST IT people thought that it was an expected behavior as the security measures the firewall blocking any icmp packets exceeding size 1000 bytes. As the investigation seemed like searching a pin in a haystack the expert, the expert of the vendor suggested another way to check: added port 2002 and configured Millennium to use port 2002 as well, because they guessed there was perhaps something conflicting on this the port 2000 which commonly used to VOIP\(^4\). However this striving was also failed. The disconnection was still appeared.

\(^2\) Cisco PIX firewalls and ASA devices have the ability to alter certain connections as they traverse the firewall. Cisco PIX refers to this as a “fixup,” which can be enabled or disabled for several network services including the Skinny Client Control Protocol (SCCP/skinny). Unfortunately both SCCP and Millennium use port 2000/TCP. When a Cisco firewall sees traffic on port 2000/TCP it assumes it is a SCCP connection and attempts to alter the traffic. This could create problems with Millennium connections.

\(^3\) Retransmissions (RST) are a result network connectivity drops, or firewall filtering/inspections.

\(^4\) Voice over Internet Protocol is a category of hardware and software that enables people to use the Internet as the transmission medium for telephone calls by sending voice data in packets using IP rather than by traditional circuit transmissions of the PSTN.
As the investigation resulted in a stalemate, an accident finding gave a hope. One day, the systems librarian of the library found her could not access her campus portal. She asked IT help desk for help. The desktop specialist checked the Internet option for her workstation and moved out gcproxy, which was provided by KAUST IT to the campus as another international link\(^5\), on the LAN Settings. After removed the gcproxy, the systems librarian could immediately access her portal after that. She immediately associated this with the disconnection issue. She asked the IT desktop specialist if gcproxy affects server-client connection, the specialist said it does as the internet option in computer not only control the browser’s internet connection, it controls whole machine’s internet connection as well. The systems librarian immediately reported this finding to IT network team, they thought we should test in another direction: run Millennium in the Internet environment without gcproxy. For including more people to join the test, we sent out notice and instruction about how to checked out gcproxy from Internet option on the workstations, and also requested the library staffs reporting the observation of how the Millennium performance after he/she check out the gcproxy. The result was exciting: we have not received the disconnection error report since then.

**Lessons and next step**

There are several points worth to be written down about our practice of cloud computing. First, the crucial step is to make correct decision. As there are many discussions about pro and con of cloud computing, it is necessary to investigate and analyze the real need and reality before making the decision. Considering there many units of the different organizations will involve into this process, it is also very important to get understanding and support from skateboard and leadership. During the process of implementation, each step and detail should be considered

\(^5\) gcproxy link terminates at Netherlight in Amsterdam and is subdivided for commodity Internet access, for connections to research and education networks in Europe, as well as for private links supporting KAUST programs. KAUST commodity Internet link is via dual 120 Mb (Megabits per second) links to Mobily, one of Saudi Arabia’s largest telecom companies. The purpose for supplying gcproxy link by KAUST IT is to fast the internet speed for KAUST academic community. However, this proxy caused the interruption of many other applications. It is stopped by most of the end user’s usage now.
thoroughly. Any neglect or careless may cause big disaster just like we experienced rollback to old server. Some problems might not be escapable as it was a new experience for us. As everything is in the “cloud” now, the network and security issues become prominent. Things might become more complicated and invisible. It does not only bring new challenge to KAUST IT, it also raises new questions to the server application, such as how the system adapt to the new network environment. For example, during the period of resolving disconnection problem, the server/host vendor and KAUST IT used to argue long time about who should take the responsibility for the disconnection. Both of them thought the problem was not theirs. Even the final result shows that the problem caused by the configuration of F5 Load Balancer, KAUST IT still believes that the reason is due to the application not developed according to the requirement of remote hosting. They said that this problem did not happen as they use the similar F5 configuration for other remote hosted applications.

As the cloud computing is a new practice both for the hosting vendor who is also the application vendor and the KAUST library, many things will change according to the new situation. There are still a lot of unclear issues and gray areas. We need review all of the documents including contracts, agreements, emergency procedures, etc. We will discuss with the suppliers, vendors and KAUST IT to revise and complete the documents based on the new technic environment. For the systems librarian, cloud computing places this specialist to a new space. Instead of only focus on the functionalities and performance of the library system, the systems librarian needs to be wearing about whole chain of the system that crosses the border of internal and external network. Thus, the systems librarian needs to have broader vision regarding the organization and technology, better understand about the breadth of different technologies and platforms, to own the capability to think outside the box, and even need to learn the language of business. The knowledge and skills of network and security should be most important and prominent for the systems librarian in the cloud-computing era.

**Conclusion**

KAUST Library moved its integrated library systems to remote host for coping with the cloud computing approach of KAUST IT. Server migration was a major task for the realization of remote hosting. Careful preparation, full cooperation, and smooth communication are the keys
for the success of migration. During the working on the cloud computing, we encountered various difficulties and problems including client-server disconnection. We used exclusive method to investigate and found the setting of F5 Load Balancer is the cause of the disconnection. We will continue to work with KAUST IT and application vendor and host supplier to improve the service and management. The library technic staff will continue to learn new knowledge both in technology and business for catching the new trends of computing service.
Reference


