Public Health Applications in the Cloud

Wayne A. Loschen, Miles A. Stewart, and Joseph S. Lombardo

Public health departments use information technology every day to help protect their communities from disease. With the advent of cloud technology, public health professionals are now discussing whether and how cloud technology should be used. This article describes the advantages and disadvantages of putting applications in the cloud, along with lessons learned from moving an existing public health disease surveillance system, Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE), from a locally installed system to a cloud environment. The article includes specific details on the use of the Amazon GovCloud environment, performance issues, and potential issues that could arise from use of the cloud in international settings. Although cloud technologies will not solve every problem, they provide a powerful and flexible foundation for applications—something that public health professionals should consider using for their future applications.

INTRODUCTION

For the past 15 years, the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE) has been used by public health officials to monitor for outbreaks of disease. A web-based system, ESSENCE has primarily been installed in individual public health jurisdictions' facilities to support the local users. These jurisdictions may be at the city, county, state, or even multistate level. Each system is configured and deployed so that it scales depending on the needs of the individual jurisdiction.

With constantly reduced budgets, it is important for public health jurisdictions to strive for efficiency with all their expenditures, especially information technology (IT) costs. Some IT costs can be extremely burdensome; jurisdictions in Virginia have been charged more than $200,000 per server for a year of maintenance and San Diego has had costs of $125,000 per server per year. These types of expenses are forcing public health departments to develop strategies to find more efficient ways to use IT systems. They need IT systems that are

reliable, scalable, and cost effective. One strategy is cloud computing.

WHAT IS THE CLOUD?

Cloud computing is a buzzword that is often used and can have different meanings depending on context. The term can often be used to define computing resources that run algorithms across thousands of computers in parallel. It can also be used to define massive storage capabilities that can manage petabytes or exabytes of data. Some people will use the term to describe web-based applications that serve millions of users, such as Gmail or Amazon. Each of these examples illustrates an aspect of cloud computing. Each example entails a number of virtual computing resources that can be created with the click of a mouse or automatically by computer applications.

At its simplest a cloud is just that, a configurable set of virtual computers that can be allocated dynamically. When dealing with clouds, the types of virtual computing resources that can be configured vary from basic computing resources, instances with Windows or Linux as the only installed software, or entire application systems that come completely configured and ready to operate with the click of a button. To describe these different situations, terms such as *Infrastructure as a Service* (IaaS), *Platform as a Service* (PaaS), and *Software as a Service* (SaaS) are used. Each cloud type describes the level of bundled software and capabilities being provided.

BENEFITS OF A CLOUD

Independent of the type of cloud used, the resources can be arranged together in a multitude of ways to meet the needs of the user. A system can be designed to take advantage of tens or hundreds of individual cloud computers for a short time to perform a calculation and then release those resources back into the pool to become available to other systems in the cloud. This ability to configure a resource, use it, and then release it allows for increased system flexibility. Instead of purchasing equipment that may be used for only minutes a day and would then sit idle otherwise, organizations can use resources more efficiently, leading to lower costs.

Beyond efficiency, the cloud can provide a flexibility that can change the way a system works. Simple examples include dealing with full hard drives. Almost every system will experience complications as its hard drive becomes full. Frantically, administrators will try to remedy the problem and frequently they will eventually have to purchase bigger hard drives or additional systems. With cloud technology, adding an additional hard drive takes about 30 seconds and costs approximately $0.10 per gigabyte per month,[1] much cheaper and quicker than ordering a new or replacement hard drive for a local server and installing it. Additional resources can be added on demand in the form of storage or computational power. Just like additional storage capacity, adding additional servers or migrating to more powerful platforms can be accomplished in the order minutes instead of weeks. This flexibility allows for systems to start small and scale with need rather than requiring large amounts of money initially for storage and computational power that might seldom be needed.

After efficiency and flexibility, the next best aspect of cloud technology is its reliability. Although there have been noticeable exceptions, such as the 36+-hour downtime of Amazon's Elastic Compute Cloud (EC2),[2] most clouds are built across multiple physical locations, allowing a system to remain active even if one of its data centers stops working. Losing an entire data center could impact performance, but if the system takes advantage of the redundancy available in the cloud environment, the system can typically remain active. For the majority of administrators who run a single server room in a single building, a cloud is a substantial improvement to reliability.

Unfortunately, there are two factors working against the reliability of clouds. The first is the lack of control if something bad does occur. In the case of a server failure in a non-cloud environment, the administrator can immediately begin putting solutions in place to fix the issue; however, with a cloud, the administrator is at the mercy of the cloud host to fix any issues. (Depending on the local administrator's level of skill with the particular issue, this actually could be an advantage instead of a detriment.) The second factor is network reliability. Most organizations use two types of networks: internal and Internet. It is possible that an organization could keep its internal network functioning during an outage on the network connecting it to the Internet. When the system is being served to a local user base via the internal network, the loss of Internet connectivity would affect a cloud-based solution more than a locally hosted one. When users require Internet connectivity to reach the server, the reliability of an organization's network compared with that of the redundant networks of many cloud operators actually favors the cloud-based solutions. Knowledge of the individual situation will help determine the reliability of cloud solutions relative to a locally hosted application for that particular situation.

ISSUES WITH THE CLOUD

Although there are many benefits to using a cloud, cloud technology is by no means a perfect solution and there are concerns that must be addressed. The first concern is security. Hosting sensitive information in servers that sit in someone else's data center can be a deterrent. Each cloud must help secure an organization's data from attacks through the network, the underlying hardware, in the physical space in which the data centers live, and from insider attacks from legitimate personnel. Each aspect is different.

For network security, the strategies used for a normal data center are similar to those used in a cloud data center. A cloud-created server connected to the Internet is not much different than a real server connected to the Internet. A cloud provider can add some extra protections, such as firewall tools and anti-spoofing applications, to protect itself and others from network attacks, but these protections can also be incorporated in non-cloud environments. A majority of the network security concerns are at the application and operating system levels. These areas are the responsibility of the system administrators who are using the cloud resources, not the responsibility of the cloud provider. If the application has a vulnerability that allows its data to be accessed, it does not matter whether that system is running on a cloud or in a local data center—the vulnerability is the same.

From the hardware perspective, many cloud providers rely on encryption and software systems to protect and control access to the data they are storing for various organizations. When using a cloud, an organization assumes additional risk compared with having its own physical server. In a cloud environment, a single physical machine may be running virtual servers for multiple organizations. Theoretically, this is less secure than having physical separation between the different systems because software is the only thing separating the shared use of memory and computational resources. There are three factors mitigating this risk. In addition to the expected improvements in the underlying cloud software used to manage security, some cloud providers have begun to allow more dedicated options that create cloud resources that use physical resources completely and do not share those physical resources across organizations. This setup eliminates one potential avenue of attack by preventing individuals outside of one organization from obtaining the data of another organization because access to the physical device is not shared; however, this sort of setup adds costs. The second factor is the creation of specialized clouds, such as Amazon's GovCloud. These specialized clouds have higher levels of security and require that operators have particular backgrounds, such as being U.S. citizens, before they are given rights to work on the physical hardware the cloud is operating on. The last factor is the threat of revenue loss for the companies selling cloud solutions. Clouds are becoming big business and if they are proven insecure, these companies stand to lose a great deal of money. This threat will help keep the pressure on vendors to keep clouds as secure as possible.

The last aspect is the physical security of the system itself. With all the complexity of network security and the software systems to manage virtual access control, it is still possible for someone to just walk into a server room, remove a hard drive, and walk away with it. This risk may or may not be greater for those organizations with systems in a cloud environment. Certainly in some situations, vendors providing cloud environments have greater security in place to prevent physical theft. However, even if there is increased security, an organization assumes risk when allowing employees of another organization to have physical access to its servers. Unless an organization is already using outside vendors to support a local data center, physical access by non-employees will always be a risk of working with cloud vendors. Vendors can mitigate this risk by requiring certain qualities in their employees, such as Amazon GovCloud's requirement that employees be U.S. citizens. Similar to the hardware aspect, the threat of lost business because of data theft keeps many vendors vigilant against physical security threats.

In addition to security, the true cost benefits of using a cloud can be difficult to understand. In some situations a cloud solution can be more cost effective than purchasing hardware. This is especially true in dynamic and unknown circumstances that may require different hardware configurations frequently. Examples include web applications that must be able to scale to handle a large event for a short period of time, systems that must be able to frequently increase storage capacities, and architectures that might require researching many different hardware and software configurations. Each of these situations can benefit from a cloud vendor's pricing system, which allows administrators to pay only for what they use when they use it. Many cloud vendors charge by the hour per computational resource, gigabyte of storage, or gigabyte of network traffic.

If the needs of the system are more known and static, however, using cloud resources may not provide any cost savings. Cloud providers can provide discounted pricing for customers that reserve resources for longer periods of time, such as 1- or 3-year reservations. These discounts, however, still may not result in a cheaper cost for a server when compared to purchasing one directly. However, it can be difficult to determine the true cost of ownership of self-owned servers. It may be hard to understand and compute power, cooling, and manpower costs associated with maintenance and security for a single server, and these costs are specific to a particular organization. In general, cloud resources can be at least comparable and in some cases more cost effective than hosting an in-house system.

MIGRATING TO A CLOUD ENVIRONMENT

After considering all of these factors, the Centers for Disease Control and Prevention (CDC) made the choice to use Amazon GovCloud as the cloud platform for its BioSense 2.0 efforts.[3] To provide public health entities with additional analytical tools to use with their data in the cloud, the ESSENCE system was also adapted to work in Amazon GovCloud. The process of migrating an existing system that worked in a non-cloud environment to a cloud environment is ongoing but so far has been successful. Even so, during the process many

lessons have been learned about the Amazon environment, cloud architectures, and possible adaptations to be considered for the future.

Amazon GovCloud is a specialized environment that is based on the Amazon Elastic Compute Cloud, or EC2. Using Amazon Web Services (AWS), GovCloud caters to systems that must meet certain U.S. government requirements in order to be visible on the Internet. From their website: "AWS GovCloud is an isolated AWS Region designed to allow US government agencies and customers to move sensitive workloads into the cloud by addressing their specific regulatory and compliance requirements."[4]

The first hurdle in using GovCloud was determining what resources were needed. GovCloud offers many different options at many different price points.[5] The ESSENCE system is a Microsoft Windows-based system that typically runs on three to five servers. In the architecture, the servers are constantly on and constantly busy. Therefore, there was no need for on-demand-style servers that could turn on and off as needed. Instead, options that gave discounts for longer-term reservations were more useful. Amazon provides two types of discounts in this regard: an option to pay a smaller up-front cost with a discounted hourly rate or an option to pay a larger up-front cost with no hourly rate. This second option was available only in GovCloud and was not available in the normal EC2 environment. For the needs of our pilot project, the second option for a 1-year reservation was chosen (3-year reservations are also available). It was easy to compute the hourly rate in addition to the up-front cost to determine how many months of constant usage would be required to make the second option more cost effective. This worked out to be close to 8 months depending on the specific configuration chosen.

The next decision was to determine the size required for the computation resource. Amazon GovCloud offers general-purpose instances that have "small," "medium," "large," "extra-large," and "double-extra-large" options that range from 1.7 to 30 GB of memory and from 1 to 26 EC2 compute units (synonymous to CPUs of a standard size). In addition, other nonstandard instances are available, including micro, memory-optimized, CPU-optimized, storage-optimized, and GPU instances. Inside many of these types are levels of servers including the small-, medium-, and large-style choices. After studying all the options, two standard large Windows servers and two high-memory extra-large Windows servers with Microsoft SQL Server were chosen for the ESSENCE project.

By default, each of these particular servers has a smaller (20 or 30 GB) hard drive that is attached. With two clicks of the mouse, additional storage space can be quickly assigned to each server and attached as a new drive. There are also processes in place to expand the size of existing drives, but adding entirely new drives worked for our purposes. It is also possible to assign Internet-accessible IP addresses, set up firewall rules between the servers and the Internet and between each of the servers, and set up keys and security permissions for management. Before connecting to the new servers, the firewall was configured to open up only certain ports needed; some administrative ports were limited to only specific IP addresses. The next step involved opening up similar ports on the local firewall so that administration could occur. Once all the networking had been configured, it was possible to use a remote desktop connection to begin installing the ESSENCE system in the cloud.

Our belief that the network was fully configured proved false early in the process as we encountered problem after problem with the servers transmitting between themselves and with the Internet. In the end, two culprits were found to cause most of the problems: Windows firewalls on the individual machines and the Amazon mechanisms for configuring firewalls. As with any Windows server, on a cloud or not, certain ports may be blocked by default and new firewall rules must be created to allow that traffic through. What made the problem slightly more complex was the nature of IP addresses inside the cloud. Amazon gives each server an internal IP address to the local network that is created just for individual accounts. In addition, for an additional cost, Internet-accessible IP addresses (called "Elastic" IP addresses) can be enabled as well. Finally, when using the firewall tools for the Amazon network, an administrator can assign rules by security groups. Servers can then be assigned to security groups, and the rules can follow a group instead of individual servers. Because of each of these options, the administrator must take special care to use the correct IP addresses in the Windows firewall relative to what is allowed in the Amazon firewall. If one firewall allows IP connections using the Internet-accessible addresses and the other firewall allows only the internal IP address ranges, traffic will remain blocked.

Eventually all the networking issues were resolved, and the next step in the process involved adding the necessary security and management software to the servers. Just like a server in a local data center, each server must have patches installed, antivirus software loaded, and monitoring tools enabled. All of the best practices used to maintain a secure platform must be performed. In addition to the security maintenance activities, backup steps must also be developed. These steps can take advantage of cloud capabilities because Amazon offers administrators the ability to take snapshots of currently running systems. These snapshots can be stored and re-created whenever needed. One additional discovery was that all GovCloud servers use GMT time by default, so we had to make sure all backup schedules took this into account.

Finally, with the servers created, the network configured, and the security software installed, the ESSENCE system was loaded onto the cloud servers. This process almost went smoothly, with the only problem being that Microsoft SQL Server was not installed on two of the

images as expected. After correspondence with Amazon GovCloud's representatives, a new image including SQL Server was available. At this point, ESSENCE was installed as if the servers were sitting in a local environment. Finally, we used the local Johns Hopkins University Applied Physics Laboratory (APL) domain name system (DNS) server to create an entry for our cloud web server so that we could give users a URL that was easier to remember.

PERFORMANCE TESTING

To evaluate the practical performance of this system in the cloud and to assess the validity of the suggested advantages and disadvantages highlighted in this article, a test was designed to evaluate the number of concurrent users the system could support in the cloud. Using a tool called JMeter and a set of common usage patterns from historic ESSENCE users, a series of workflows were designed that simulated a normal user's behavior in ESSENCE. These behaviors included performing queries, viewing time-series graphs, and then viewing data details pages, for example, among other common usage patterns. The JMeter tool could then be set up to simulate any number of concurrent users attempting to use a website.

A test was designed to evaluate four combinations of servers:

- Web server: standard large; and database server: high memory extra-large (XL)
- Web server: standard large; and database server: high memory double extra-large (2XL)
- Web server: high CPU extra-large; and database server: high memory extra-large (XL)
- Web server: high CPU extra-large; and database server: high memory double extra-large (2XL)

The tests showed that the web server machine did not affect performance significantly. The size of the database server, however, did impact performance. Tests were performed with 10, 100, 150, 200, 250, and 300 concurrent users. The results are shown in Table 1 and Figure 1.

Based on these results, the XL servers are comparable to the 2XL servers if the number of concurrent users is 100 or less. This can be seen in similar average times. If the system is required to support more than 100 concurrent users actively using the system, then the more powerful server should be used. In our experience with ESSENCE usage in the past, we estimate that systems that have more than 200 registered users rarely have more than 20 concurrent users. This would indicate that, depending on their activity level, at least 1000 users could use the XL database server without a performance hit.

This performance testing proved that multiple jurisdictions could share cloud resources and still expect the level of performance that they are accustomed to in a local version of ESSENCE. This sharing of resources could decrease the costs for individual jurisdictions and allow for an easier path for scalability and flexibility for

Table 1. Performance results of ESSENCE usage based on XL- and 2XL-sized database servers

No. of Users	Min. Time (ms)	Max. Time (ms)	Avg. Time (ms)	Success Rate	No. of Samples
XL Database Server					
10	336	4,112	680	100	398
100	47	47,747	2,891	100	2,978
150	1,283	69,672	19,613	100	1,739
200	896	104,924	31,119	100	1,652
250	528	99,719	35,893	99.68	1,864
300	992	93,779	29,464	74.92	2,624
2XL Database Server					
10	42	6,175	467	100	367
100	42	19,513	2,387	100	3,150
150	43	22,221	1,909	100	4,930
200	50	72,870	7,128	100	4,248
250	44	147,909	14,270	100	3,596
300	292	171,364	26,562	76.09	2,769

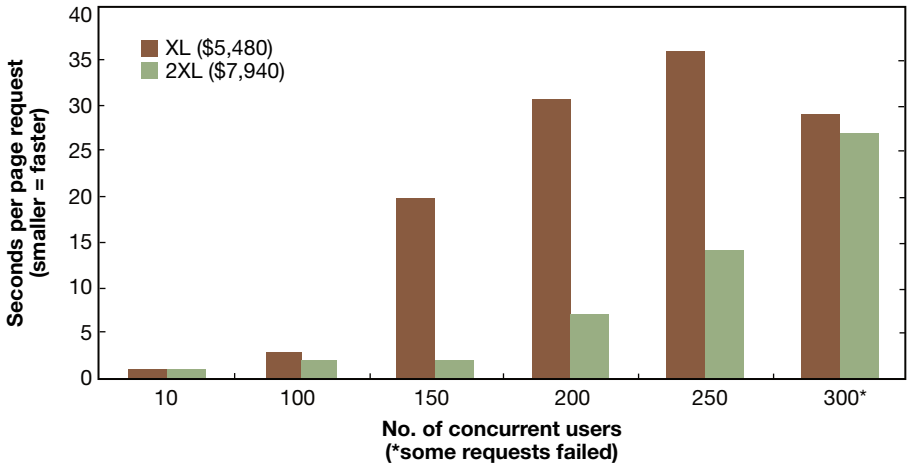

Figure 1. Performance results of ESSENCE usage based on XL- and 2XL-sized database servers.

the participating jurisdictions. Anecdotally, users of the cloud version of ESSENCE also found the day-to-day performance to be faster than it was in their local instances. So in addition to the efficiency, perceptions of performance may even be improved in a cloud instance over a local instance.

INTERNATIONAL USE CONSIDERATIONS

Although the CDC and other U.S. public health entities have begun using cloud technology, there may be additional barriers to its use in some international settings. Although there are many cloud services available throughout the world, some countries do not have cloud providers located inside their borders. This is not a technical barrier as long as the public health entity has an Internet connection, but it can be a political barrier. The issues of data privacy are serious and some jurisdictions have policies that forbid data from leaving their countries. In addition to the political issues, there are technical concerns if the public health entity does not have reliable or robust Internet connectivity. Although the cloud services may be available, if the public health entity cannot connect to the service to provide or analyze its data, the value of using a cloud is greatly diminished. Even if the entity has reliable Internet connectivity, if there is not enough bandwidth to allow the data to be transferred in a timely manner from the data providers to the cloud application, this too can render a cloud-based solution unusable. In these situations, preference may be given to local solutions that can incorporate data collection from a variety of methods, including hand-entering or loading collected data from media or text messaging-based collection systems. These collection methods can bypass use of a network and allow analysis to be performed despite the IT limitations. Although many public health entities in an international setting may be perfectly suited to use cloud technology, it is not something that should be assumed without first considering these possible barriers.

CONCLUSIONS

Although cloud technology is not a perfect solution to every problem, the ability to have a scalable and flexible foundation for applications is useful for tools such as ESSENCE. The cloud platform provides the ability to host a large number of users on a single hardware platform and therefore share costs across multiple jurisdictions, providing a more cost-efficient model for IT system hosting. It also has the added benefit of being able to adapt in the future by assigning additional resources to the application if needed. Both of these aspects allow for a cloud version of ESSENCE to be a much more efficient solution compared with numerous individual instances. Although technology and pricing are not the only issues to consider when deciding on the utility of a cloud-based application, we have shown through this pilot and our performance testing that a cloud-based system could be a valuable alternative that should be considered.

ACKNOWLEDGMENTS: The ESSENCE in the cloud initiative was supported by the CDC's Division of Notifiable Diseases and Healthcare Information (DNDHI) BioSense Program. Also, the project could not have succeeded without the hard work of many individuals from all three partner organizations for this project: CDC's BioSense team, Tarrant County Public Health Department, and APL's ESSENCE team.

REFERENCES

[1]"Amazon EC2 Pricing," Amazon.com, http://aws.amazon.com/ec2/pricing/ (accessed 26 Jul 2013).
[2]Goldman, D., "Why Amazon's Cloud Titanic Went Down," CNN.com, http://money.cnn.com/2011/04/22/technology/amazon_ec2_cloud_outage/index.htm (22 Apr 2014).
[3]Kass-Hout, T. A., Spears, K. L., Brownstein, J. S., Alletto, M., Freifeld, C. C., et al., "CDC's BioSense 2.0: Bringing Together the Science and Practice of Public Health Surveillance," *AJPM Prevention in Practice*, http://ajpmonline.wordpress.com/2011/11/15/cdc%E2%80%99s-biosense-2-0-bringing-together-the-science-and-practice-of-public-health-surveillance-4 (15 Nov 2011).
[4]"Amazon GovCloud (US) Region–Government Cloud Computing, Amazon.com, http://aws.amazon.com/govcloud-us/ (accessed 26 Jul 2013).
[5]"AWS GovCloud (US) Region Pricing," Amazon.com, http://aws.amazon.com/govcloud-us/pricing/ (accessed 26 Jul 2013).

The Authors

Wayne A. Loschen is the Technical Lead for ESSENCE-related projects. He is an APL Senior Professional Staff member in the Bio-Threat Awareness Systems (QAI) Group. For ESSENCE projects, he provides technical guidance and software development and works with local public health partners to improve the utility of ESSENCE for its users. **Miles A. Stewart** is a software engineer in the QAI Group. He is responsible for back-end development, performance testing, and general feature development for the ESSENCE project. **Joseph S. Lombardo** is the Principal Investigator (PI) for ESSENCE and has led the program since its inception in 1997. For further information on the work reported here, contact Joseph Lombardo. His e-mail address is joe.lombardo@jhuapl.edu.

The *Johns Hopkins APL Technical Digest* can be accessed electronically at **www.jhuapl.edu/techdigest**.