Information Evaporation: The Migration Of Information To Cloud Computing Platforms

David Reavis, Texas A&M University-Texarkana, USA

ABSTRACT

The physical location for data used in every organization ebbs and flows as technology improves. In the early years of computing, data were stored on the central system because that was the only choice. As communication technology advanced, a decentralized model became popular and data were stored nearer to the place it would be used. Another leap in telecommunications prompted a move back to centralized data storage, mostly because access speeds allowed the data to be used remotely with minimal time lapse due to transmission distance. The most recent transition for housing data is to move data from various databases, some centralized and some localized, into the cloud. The benefits of moving information to a cloud computing environment have made it attractive to organizations recently. Converting data from one platform to another is done regularly by IT professionals. In each of the transitions described above, data had to be converted in some way and transitions to updated computing platforms are not uncommon. In this paper, the term information evaporation will be used to distinguish the move of information to the cloud from other conversion activities, such as system upgrades or platform transitions. Converting data from a traditional database environment to an Internet-based cloud computing environment requires a different approach to security, attention to avoiding creating information silos, and development of data tags, such as eXtensible Markup Language (XML), to facilitate cross platform data access.

Keywords: Cloud Computing; Computing Platform; Database; Data Conversion; Information Silo

INTRODUCTION

In nature, a system known as the hydrologic cycle, or water cycle, describes the continuous movement of water around the earth. In part of the hydrologic cycle, water which is stored on the earth's surface is acted upon by the sun and is transformed into a vapor where it rises into the air to form clouds (Perlman, 2010). Using the hydrologic cycle as a metaphor allows the term *information evaporation* to refer to the process of transforming information that resides in traditional data stores to information in the computing cloud.

Cloud computing was popularized beginning in 2007 as a result of IBM and Google's announcement of a research initiative to develop large data centers that could be accessed via the Internet (Lohr, 2007). The principle behind cloud computing is that the work (such as hardware provisioning, maintenance, upgrades, and other infrastructure-based activities) could be managed externally, separating these activities from the uses of the associated programs and information, but connected by the Internet. One key to the success of such a configuration is that resources can be virtualized, giving users flexibility to consume only the necessary resources and avoid the costs of unused capacity (Vouk, 2008).

The benefits of moving information to a cloud computing environment have made it attractive to numerous organizations in the past few years. Converting data from one platform to another is done regularly by IT professionals. In each of the transitions described above, data had to be converted in some way and transitions to updated computing platforms are not uncommon. In this paper, the term information evaporation will be used to distinguish the move of information to the cloud from other conversion activities such as system upgrades or platform transitions. The hydrologic metaphor is appropriate because the information does not simply disappear, but does undergo some important changes in the move to the cloud, just as evaporated water does not disappear but moves into a vapor state that is invisible to the human eye until it forms a cloud in the sky.

LITERATURE REVIEW

While there is not a single, widely accepted definition for cloud computing (Boeri, 2011), a good definition is explained by Armbrust (2010) as a data center (hardware, software, and connectivity) that is made available to the general public in a pay-as-you-go manner. This description describes a public cloud, where an organization may allow access from within the organization or from external points. The overriding attraction to this type of platform is that it allows for flexible provisioning of resources. When surges in processing occur, additional resources can be quickly and easily provisioned and used, then when the load drops, the resources are decommissioned until they are needed again. This reduces the up-front commitment of the organization because they do not need to invest in the hardware based on a forecast of resources needed. An organization using the cloud can simply purchase the computing power they need, then release the resources they do not need (Armbrust, 2010).

The main technology component that allows flexible provisioning of computing resources is virtualization. Virtualization allows a single machine or cluster of machines to be used by multiple users without duplicating certain aspects of the hardware. Products such as VMware™ provide the ability to run multiple instances of a server on a single machine such that memory, disk space, and other resources can be supplied as needed by the various systems (Vouk, 2008).

Buyya (2009) argues that cloud computing will eventually become a utility in the same way that water, electricity, natural gas, and telephony are utilities today. For this to occur, computing resources must be provisioned and marketed similarly to other utilities. The market for these resources is currently in its infancy and as the technology for cloud computing matures, the markets where these resources are made available will progress.

In June 2010, the Pew Research Center reported that a poll of technology experts showed that 71% of these experts expect most people to access software applications in a cloud environment (Boeri, 2011). Despite this expectation, many organizations are skeptical of cloud computing because of the risks related to security and confidentiality. The federal government is particularly concerned with security, and because of regulations in the Federal Information Security Management Act (FISMA), cloud vendors face significant challenges in meeting these government requirements. FISMA specifies over 120 different controls on information systems that address issues such as the trustworthiness of administrators, contingency planning, and backup plans for a vendor who goes out of business (Hoover, 2010).

In addition to the security concerns, other issues that may cause organizations to hesitate to use the cloud computing paradigm are those about business continuity and service availability, data lock-in, and data transfer bottlenecks. The availability discussion may result in a service level agreement between an organization and a cloud computing vendor, but many companies are reluctant to allow a single point of failure (the cloud) as a risk to system availability. Data lock-in occurs when the storage application program interface (API) limits the user to certain methods of data transfer that are proprietary to the cloud vendor. This is a danger because of the lack of standards in data storage for cloud environments. This lack of standardization is a key issue in information evaporation because once data has been moved to the cloud environment, a significant part of the value of the cloud is derived from accessing the data from diverse client platforms. The remaining objection, data transfer bottlenecks, may be overcome in some environments through data compression techniques, but in some environments there is simply still too much data moving from system to system for cloud computing to be a viable option (Armbrust, 2010).

OBJECTIVE OF THE STUDY

The objective of this study is to identify and describe the key differences between converting data from one traditional platform to another and converting data and the information it represents from a traditional platform to the cloud computing environment. These differences include an emphasis on security, the need for avoiding creating information silos, and using data representation standards such as XML. The primary exemplar for discussion will be the case of data for the Susan G. Komen Race for the Cure ® being converted for use in an application developed by Salesforce.com™, a cloud computing provider. This conversion took place during 2010 and involved data conversion from over 120 affiliates. The information moved to the cloud environment, in this case, is indicative of

many of the issues faced by organizations seeking to take advantage of information evaporation (i.e., moving their information to the cloud computing platform). (salesforce.com, 2010)

There are some practices in converting data to the cloud environment that are no different than any other conversion from one platform to another. These practices include making sure that data is appropriately normalized, table relationships in the target system exist to support the data structures, any differences in data types are resolved, and data scrubbing has been done to ensure a clean conversion. These types of conversion issues are common in any conversion and must be addressed in information evaporation as well.

Security

The most frequently cited concern with cloud computing is how enterprise data will be safeguarded once it is moved to a third party, cloud-based environment (Harris, 2009). In the above cited case, one affiliate's major concern was the protection of email addresses of minors. In other situations, sensitive information, such as financial records, sales records, human resource information, and many other data points, causes organizations to be very careful about who has access to these records. The ability of the organization to control access may be diminished once the data are moved out of the local data center and into the cloud.

Security also becomes an issue when chains of custody are important - an issue related to e-discovery. If an organization's cloud-based data becomes part of a legal proceeding, the courts frequently require the organization to lock down information so that it cannot be altered. The cloud computing environment adds yet another level of complexity to this situation because the data, and likely the whole computing environment, are part of a virtual system (Boeri, 2011).

Cloud vendors are keenly aware of the security issue and continue to take steps toward a more secure environment. One of the essential elements in providing the desired security is data encryption. By encrypting data on the disk, companies can reduce their exposure to the risk that someone could steal their data simply by copying it from the database. Recent advances in cloud data encryption provide users with the ability to exercise complete control over the encryption keys and encryption and decryption processes by having the encryption keys reside with the enterprise instead of with the cloud provider (Vijayan, 2011). While data encryption is common in some industries, many traditional data center-based systems do not encrypt data that are stored within the confines of their enterprise.

The globalization of computing in the cloud also brings the issue of handling personally identifiable information (PII) into question. Different countries have different requirements on how PII is stored, accessed, and protected. Having to confirm the physical location of servers and verify that the laws in that country do not conflict with organizational practices adds time and cost to the conversion from a traditional system to a cloud-based architecture (Boeri, 2011).

Additional security for data in a public cloud may be achieved through a service level agreement (SLA) between the cloud vendor and the organization using the cloud services. Many organizations use SLAs to ensure uptime, disaster recovery and business continuity, and traffic analysis; but including physical security, logical access, proper authentication, and protection of off-site data storage are factors that must be included in the cloud computing scenario. Access should not only address the data itself but may include access to servers, programs, and the Internet (Kandukuri, 2009).

Because security is one of the major impediments for companies moving to cloud computing, a variety of organizations are responding to the challenge of ensuring information security in the cloud. One of these organizations is the Cloud Security Alliance (CSA) that issued a guidance paper in 2009 touting access management policies as one of the keys to minimizing administrator risks. Access management policies are concerned with the various roles in the computing environment and what specific duties are assigned to each role. To provide the quick provisioning of hardware that makes the cloud environment so attractive, the vendor may need "root" or "Super User" access to the hyper visor or server level. This all-powerful authority over the system must then be accountable for areas where privileges are allowed (Craig, 2010).

When operating information systems in a cloud environment, the emphasis for security is normally the data, servers, and any element of the network that resides in the cloud. Because of this emphasis, the client portion of the system is sometimes overlooked. One of the most frequently used client-side tools is the Web browser, which may be enhanced by plug-ins and extensions which are noted for security risks. The next wave of cloud access will be mobile devices, which also provide browsing capability through lightweight applications (apps) that may introduce further vulnerabilities to data security in the cloud (Jansen, W. & Grance, T., 2011).

Security tops the list of concerns for many organizations because once data has been moved to a cloud-based system, there are numerous risks that must be addressed, some of which are much easier to address in a traditional environment. The root cause of the security issues arises from the fact that the data are physically stored in a remote system and unless there are specific SLAs which guarantee all access privileges, the data may be at risk. The methods described above represent some of the potential ways to mitigate this risk, but security remains significantly more complex in the cloud than in the traditional data center.

Information Silos

In the Susan G. Komen Race for the Cure ® case, prior to information evaporation, data existed in various forms on computers and in filing cabinets at more than 120 affiliates. Some affiliates used Microsoft Access on a personal computer to store their race participants and sponsor data and others used spreadsheets or custom software. One of the biggest advantages of this move was that the data from these diverse sources were merged into a single platform. In terms of avoiding information silos, that is the ideal. An information silo develops when an information system is incapable of reciprocal operation with other related information systems (Short, 2010).

Organizations that focus on data at a department or location level, rather than at the enterprise level, risk building information silos. As an example, if an organization's Human Resource department uses an internal team to acquire systems that meet its needs using a cloud-based platform without considering how the data might be needed in other areas of the company, then the HR department has likely created an information silo. The remaining legacy systems could potentially benefit from the cost, overtime, benefit, or other information that might be stored in the new system, but the integration of an older legacy system and a cloud-based system can be problematic (Krigsman, 2010). Information silos may also make it difficult to optimize supply chains, integrate with external partner systems, and take advantage of market opportunities (Vayghan, 2007).

Another example of information silos is evident in various university and public libraries. Current systems in these organizations are frequently based on pre-internet technology and contain duplications for the same information on an order of magnitude that may not be found in any other type of business. Considering the number of copies of cataloging data there are for popular periodicals in libraries across the spectrum of these public organizations, there are likely thousands upon thousands of entries for a single periodical. In a cloud-based environment, the possibility of reducing some of this duplication could greatly decrease the storage space required (Goldner, 2011).

As an organization begins to plan for data evaporation, it must consider how cloud-based systems will interoperate with other systems. If moving a system to the cloud creates an information silo, the implications for other systems should be considered. Options may include an enterprise application integration (EAI) tool, reliance on cloud vendor APIs, or internally-developed solutions. In some cases, the organization may choose to simply live with the silo. Regardless of the remedy, the effect of moving systems to the cloud, while continuing to operate other internal systems, carries some risk and the potential for added operational cost combined with the potential for inefficient system interoperations.

Standards

Many legacy systems have been built on relational database (RDB) technology. As RDB technology has improved, systems have been updated to take advantage of more powerful database functionality. McNurlin (2009, p.377) defines the process of moving a legacy system from one platform to another as reengineering. Reengineering systems to function in the cloud require the developer to use an approach for storing data that will allow efficient

access in an environment where no widely accepted standards have been developed. An emerging standard for data storage on Internet-based platforms is the eXtensible Markup Language (XML) (Knight, 2009). Other Web formats, such as JSON, RDF, and Wiki text, are also widely used in cloud computing (Khatchadourian, 2010). These standards provide the important ability to store and retrieve structured, semi-structured, and unstructured data (Benedikt, 2010).

As an organization moves software functionality to the cloud, there are some legacy systems that may not be well suited for the new environment. In cases where traditional systems and cloud-based systems co-exist in an organization, the need for interoperability among the various systems may occur. An XML application program interface (API) is one of the ways that developers pass information to and from the various platforms. Using vendor APIs allows legacy systems and cloud-based systems to work together to some degree. One complication is that some legacy systems do not support XML and the communication for such a system may need additional coding to supply the cloud system with appropriate metadata for a given transaction (Knight, 2009).

The potential problem of a lack of interoperability is not an issue when an organization, such as the Susan G. Komen Race for the Cure ®, moves all functionality to the cloud. In this case, the metadata needed to create XML tags were created as part of the conversion process and included in the initial table population of the cloud database. This is one of the differences between moving from one traditional platform to another and to a cloud-based system. The traditional target may or may not require additional metadata to support data storage, whereas most cloud-based systems use XML (or similar) tags to support data interoperability among Internet-based platforms.

IMPLICATIONS

The information and case elements from the Susan G. Komen Race for the Cure ® case indicate that there are differences between moving from one traditional information system to a new system on a similar platform as well as moving from a traditional information system to a cloud computing environment. These differences include a much stronger emphasis on security in the cloud environment, the risk of creating information silos, and the lack of well-defined standards in the cloud. There are likely other differences in the approach to moving to the cloud-based computing platform. Organizations that are considering using information evaporation should address these issues in any system migration project and be aware that the cloud technology is wholly different than most of the existing legacy environments.

Because of the attractive nature of quick provisioning of virtualized resources, many enterprises are evaluating decisions on various projects that include cloud-based options. IT experts agree that this platform will increase in popularity because of the benefits in flexible resource allocation and despite the risks in security and availability. As more companies use information evaporation, the cloud industry will mature and provide better interoperability options, better security, and better control over the available resources.

CONCLUSIONS

The differences between moving from one traditional information system to a new system on a similar platform and moving from a traditional information system to a cloud computing environment are significant. Since moving information from a traditional environment to an Internet-based system is relatively new, the term information evaporation has been coined to help conceptualize the process. Using the hydrologic cycle as a metaphor for moving information to the cloud may help IT practitioners and users gain a needed perspective on this new and unique process.

RECOMMENDATIONS FOR FUTURE RESEARCH

As more enterprises move computing functions to the cloud environment, other areas of research may include investigating the global implications of data storage across international borders, investigating more efficient ways of moving large data sets to and from the cloud, and the evolving standards for data storage in cloud

computing environments. Research also needs to be undertaken to isolate the actual risks of data loss as opposed to the perceived risks for which cloud vendors are currently expending significant resources to address.

ACKNOWLEDGEMENT

This research project was made possible in part by a grant from the Women for A&M-Texarkana.

AUTHOR INFORMATION

David Reavis teaches both Web-based and traditionally formatted classes in management information systems at Texas A&M University-Texarkana. His work experience includes software development roles for manufacturing companies such as Cooper Tire and Rubber Co. and Alcoa. Reavis received his undergraduate BBA in computer information systems from Southern Arkansas University, his MBA from Texas A&M University-Texarkana and his Ph.D. in information systems from Nova Southeastern University. E-mail: david.reavis@tamut.edu

REFERENCES

1. Armbrust, M., Fox, A., Griffith, R., Joseph, A., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., and Zaharia, M. (2010). A View of Cloud Computing. *Communications of the ACM. 53*(4).
2. Benedikt, M., Florescu, D., Gardner, P., Guerrini, G., Mesiti, M., and Waller, E. (2010). Report on the EDBT/ICDT 2010 workshop on updates in XML. *SIGMOD Rec.* 39, 1 (September 2010), 54-57. http://doi.acm.org/10.1145/1860702.1860713
3. Boeri, R. (2011). Cloudy Content. *EContent.34*(6). 27.
4. Buyya, R., Yeo, C., Venugopal, S., Broberg, J., and Branic, I. (2009). Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility. *Future Generation Computer Systems*. In Press.
5. Craig, D. (2010, September 13, 2010). Access Management Policies Hold Keys to Minimizing Administrator Risks. *Construction Cloud Computing*. Retrieved on February 10, 2011 from http://www.constructioncloudcomputing.com/2010/09/13/access-management-policies-hold-keys-to-minimizing-administrator-risks/
6. Goldner, M. (2011). Winds of Change: Libraries and Cloud Computing. *Multimedia Information Technology*. *37*(3). 24-28.
7. Harris, J., Daugherty, P., and Tobolski, J. (2009). What the Enterprise Needs to Know About Cloud Computing. *Accenture*. Retrieved January 10, 2011 from http://www.accenture.com/us-en/Pages/insight-enterprise-cloud-computing-summary.aspx
8. Hoover, J. (2010, July). Cloud Compliance in Government. *Information Week Analytics*.
9. Jansen, W. & Grance, T. (2011). Guidelines on Security and Privacy in Public Cloud Computing. Retrieved from the National Institute of Standards and Technology website: http://www.nist.gov/customcf/get_pdf.cfm?pub_id=909494
10. Kandukuri, B. R., Paturi, R., and Rakshit, A. (2009). Cloud Security Issues. 2009 IEEE International Conference on Services Computing. DOI 10.1109/SCC.2009.84
11. Khatchadourian, S., Consens, M., and Simeon, J. (2010) Web data processing on the cloud. In Proceedings of the 2010 Conference of the Center for Advanced Studies on Collaborative Research (CASCON '10). ACM, New York, NY, USA, 356-358. DOI=10.1145/1923947.1923990 http://doi.acm.org/10.1145/1923947.1923990
12. Knight, R. (2009, June 30). The New Role of XML in Cloud Data Integration. IBM. Retrieved February 14, 2011 from http://www.ibm.com/developerworks/webservices/library/x-sftoeap/index.html
13. Krigsman, M. (2010, January 14). Information Silos and IT Governance Failure. IT Project Failures. Retrieved January 22, 2011 from http://www.zdnet.com/blog/projectfailures/information-silos-and-it-governance-failure/7975
14. Lohr, S. (2007, October 8). Google and I.B.M. Join in 'Cloud Computing' Research, *New York Times*. Retrieved from: http://www.nytimes.com/2007/10/08/technology/08cloud.html?r=1&ei=5088&en=92a8c77c354521ba&ex=1349582400&oref=slogin&partner=rssnyt&emc=rss&pagewanted=print

15. Mcnurlin, B.C., Sprague, R. H., Bui, T. (2009). *Information Systems Management in Practice*. Upper Saddle River, NJ: Prentice Hall
16. Perlman, H. (2010) Evapotranspiration. U. S. Geological Survey. Retrieve January 15, 2011 from http://ga.water.usgs.gov/edu/watercyclesummary.html
17. Salesforce.com, (2010). Susan G. Komen for the Cure ® completes tasks that used to take months in seconds with Force.com cloud platform. Retrieved February 15, 2011from: http://www.salesforce.com/showcase/stories/susang-komen.jsp
18. Short, K. (2010). Information Data Silo. Technology, Marketing and Business Process Management. Retrieved on February 15, 2011 from http://kellyrshort.com/lpinformationsilo.html
19. Vauk, M.A. (2008). Cloud Computing – Issues, Research, and Implementations. *Journal of Computing and Information Technology - CIT* 16(4) 235-246.
20. Vayghan, J. A.; Garfinkle, S. M.; Walenta, C.; Healy, D. C.; Valentin, Z.; (2007). The internal information transformation of IBM, *IBM Systems Journal 46*(4), pp.669-683. doi: 10.1147/sj.464.0669
21. Vijayan, J. (2011, February 10). Vendors Tap into Cloud Security Concerns with New Encryption Tools. *Computerworld*. Retrieved February 12, 2011 from http://www.computerworld.com/s/article/9208882/Vendors_tap_into_cloud_security_concerns_with_new_encryption_tools

NOTES