OCLC Systems & Services: International digital library perspectives

An overview of virtual and cloud computing
H. Frank Cervone

Accepted May 2010

MANAGING DIGITAL LIBRARIES: THE VIEW FROM 30,000 FEET

An overview of virtual and cloud computing

H. Frank Cervone
Information Services, Purdue University Calumet, Hammond, Indiana, USA

Abstract

Purpose – The purpose of this paper is to define and describe virtualization of servers and cloud computing. Understanding the differences and similarities between the two technology models will help digital library managers make better decisions related to hosting of application services.

Design/methodology/approach – The paper is a general overview of the principles and techniques of virtualization and cloud computing.

Findings – Both virtualization and cloud computing can be effective methods of optimizing hardware resources used to run digital library applications; however, system managers should not overlook the potential for security problems and management problems given the outsourced nature of the computing resources.

Originality/value – The paper fills a gap in the digital library project management literature by providing an overview of the server virtualization and cloud computing models which could be applied to digital library projects.

Keywords Computer applications, Digital libraries, Project management, File servers

Paper type General review

In the last year or so, the buzz around cloud and virtual computing has grown almost exponentially. Not surprisingly, confusion and cynicism have also grown in direct proportion to the seemingly unending hype generated by these two concepts. The extent of this schizophrenic state of opinion related to cloud computing was demonstrated by Larry Ellison, president of Oracle, during a talk at the Oracle OpenWorld conference:

> The interesting thing about cloud computing is that we've redefined cloud computing to include everything that we already do. I can't think of anything that isn't cloud computing with all of these announcements. The computer industry is the only industry that is more fashion-driven than women's fashion. Maybe I'm an idiot, but I have no idea what anyone is talking about. What is it? It's complete gibberish. It's insane. When is this idiocy going to stop? (Farber, 2008).

What is obvious is that virtual and cloud computing are changing the way information technology is implemented in organizations today. The question is no longer if virtual and cloud computing will be a reality, but instead when. Adoption of cloud computing is definitely accelerating (Avery, 2009) and much of this is being driven by requirements for organizations to be more adaptable and flexible in responding to IT needs.

OCLC Systems & Services: International digital library perspectives
Vol. 26 No. 3, 2010
pp. 162-165

1065-075X
DOI 10.1108/10650751011073607

Contrasting traditional computing with virtualized computing

Traditionally, organizations have implemented applications by installing software for the application on one or more physical servers. In many cases, multiple servers may be used when an application services a high transaction volume or when there needs to be a high level of assurance that the failure of a single server will not cause an application to completely fail. A classic example of this type of application is a university's student registration system. During peak registration periods, transaction levels will be very high and the university needs to ensure that the application will not go down due to a hardware failure. In the middle of the semester, however, the application will be barely used and an outage would likely not affect many people.

In this traditional model, servers are considered as a single, functional unit that encompasses the hardware, operating system, application, utilities, and storage for the application. When the capacity of the server is exceeded, additional hardware or storage must be added to the physical machine. It is possible to set up clusters of servers, all of which are configured exactly the same, to ensure greater fault tolerance. This helps protect the organization from the undesirable consequences of a server hardware failure. However, there are limitations on the scalability of clusters due to constraints in either the operating system or database management software. An additional factor is that not all applications will function correctly in a clustered environment.

The traditional server model has been in use for several decades because it is relatively easy to implement and deploy. Moreover, conceptually the model is not unduly complicated and almost all applications have been designed using this model. The downside is that it has inherent limitations related to scalability both up and down. Increasing scalability requires additional hardware that exactly duplicates the current physical hardware. On the other hand, when a server is not busy it is impossible to use that extra capacity for other workloads. Given the insular nature of the model, disaster recovery is complicated because a replica of the entire environment needs to be recreated to reestablish functionality.

In a virtual server model, a different type of operating system runs on the physical hardware. Known as a hypervisor or a virtual machine manager, this operating system is specialized and designed to run other operating systems as applications. By doing this, the traditional server environment can be encapsulated in an environment which provides a layer of isolation from the underlying hardware. A hypervisor can run multiple operating system environments and their associated applications in individual virtual machines, each of which runs on a single physical machine. Each virtual machine is presented with an environmental context that makes it appear as if they were running natively as the only machine on the physical hardware. It is the job of the hypervisor to simulate, emulate, or enable all of the functionality necessary for the virtual machine to operate normally and with all of its functionality.

In many virtual environments, however, the physical machines run as a cluster through cooperating hypervisors. This is to provide fallover capability if a physical server should fail, just as would occur with a regular physical hardware cluster. What is different, however, is that the fallover virtual machine does not have to run in parallel with the "real" virtual machine. When the hardware of a virtual machine running on a cluster fails, the operating environment of the virtual machines on that server can be transparently swapped onto another physical server on the fly avoiding an interruption of service.

Because of this flexibility in being able to move the virtual servers around the cluster, scaling resources to meet the needs of virtual machines is also simplified. In most virtual server implementations, several virtual machines are assigned to a single physical server based on the assumption that none of the virtual machines will need the entire computing power of the physical server. When a particular virtual machine does consume all the processing cycles of the physical server, the hypervisor could migrate the other less busy virtual machines to other physical servers to ensure no virtual machine's performance is significantly degraded. Additionally, if a particular virtual machine exceeds the processing capacity of its current physical server, the hypervisor is able to enable additional instances of the virtual machine on other physical servers within the cluster to keep the performance of the virtual machine's applications at an acceptable level. How these virtual machine transfer operations are implemented varies significantly depending on the capabilities of the particular hypervisor but conceptually the implementation of the functionality is the same across varieties of hypervisor.

From virtualized computing to cloud computing

Cloud (common location independent, on-line utility on demand) computing takes this concept of virtualization even further and adds a couple of additional twists as well. In a cloud computing environment, the organization running an application does not typically own the physical hardware used for the applications. In fact, when running applications in the cloud, an organization does not usually know exactly where the computation work of the applications is being processed.

Cloud computing provides an organization with appreciably more flexibility and scalability to satisfy computing needs. An organization can start out running an application on a single virtual server and ramp up as necessary because the cloud provider hosts the applications on virtual servers. New virtual servers can be created in a matter of minutes and, depending on demand, a server can be deactivated in a similarly short amount of time. The computing power to run the applications is purchased on an "as needed" basis and paid for based on actual usage.

Public clouds, such as Amazon's EC2 service, are available to any person or organization that wants to use the services provided. Private clouds, on the other hand, are typically contracted for by an organization with a provider on a one-on-one basis. That is, the private cloud is only available to and is only used by the contracting organization. Private clouds provide an organization with a greater degree of control over the operation of the cloud as well as providing a higher level of security since the cloud is not shared with others.

One of the significant advantages of cloud computing is the potential cost savings that can be gained. Usually cloud computing has little or no upfront capital costs. For the most part, operational responsibilities are shifted to the cloud provider, who is then responsible for the on-going maintenance of the hardware used by the cloud. A major portion of an organization's cost savings are gained by taking advantage of the variable transaction demands most applications have and redistributing unused computing cycles during the slower periods of one application to more demanding applications running at the same time. Because of this ability to shift resources, additional hardware purchases can be deferred until the entire cloud cluster's overall application load requires more computing power, rather than just a single application.

Concerns

While one of the advantages of a cloud is that an organization does not need to know the details of the physical hardware, an organization will often want to know where the application's cloud "lives." One of the ways cloud providers keep expenses down is to place data center in locations where the cost of real estate, utilities, and labor are low. Given this scenario, many clouds may be hosted in foreign countries.

The major problem with clouds that are hosted internationally is that the application and data are subject to the laws and policies of the host nation. For example, many Canadian provinces have made it illegal for applications in their province to be hosted in the USA because the data would then be subject to provisions of the Patriot Act. Similarly, some applications in the USA cannot be hosted overseas because of restrictions in the export of computer system technology.

Because of this, where a cloud is hosted is important in the context of concerns related to compliance, auditability, and eDiscovery. Since cloud computing is a variation on computer outsourcing, an organization will need to have a high degree of confidence in the security procedures and protocols of their cloud computing provider. Depending on the nature of the application and data hosted in the cloud, an organization may have to deal with issues related to HIPPA, FERPA, PCI, GLBA, and other mandates or regulatory agencies. An organization needs to ensure that their provider can address any requests for information related to regulatory or statutory issues. The Cloud Security Alliance (2009) has developed a comprehensive guide to help organizations sort out the issues related to many of these concerns.

Perhaps most importantly, an organization can best protect itself by only working with cloud providers that are committed to transparency. There is no valid reason that prevents cloud providers from being able to disclose their security practices and hosting procedures to the organization purchasing cloud computing services. Microsoft has recently promoted the Cloud Computing Advancement Act (Microsoft Corporation, 2010) as a way to provide a greater degree of transparency and openness in the cloud computing arena. This, along with the work of the Cloud Security Alliance, provides a solid framework for organizations to address the complex issues related to cloud computing.

References

Avery, P. (2009), "Research indicates increase in cloud computing", *IT Business Edge*, August 26, available at: www.itbusinessedge.com/cm/community/kn/blog/research-indicates-increase-in-cloud-computing/?cs=35256

Cloud Security Alliance (2009), *Security Guidance for Critical Areas of Focus in Cloud Computing, V2.1*, available at: www.cloudsecurityalliance.org/csaguide.pdf

Farber, D. (2008), "Oracle's Ellison nails cloud computing", *CNET News*, September 28, available at: http://news.cnet.com/8301-13953_3-10052188-80.html

Microsoft Corporation (2010), *Cloud Computing for Business and Society*, available at: www.microsoft.com/presspass/presskits/cloudpolicy/

Corresponding author

H. Frank Cervone can be contacted at: fcervone@calumet.purdue.edu

This article has been cited by:

1. Tiago Oliveira, Manoj Thomas, Mariana Espadanal. 2014. Assessing the determinants of cloud computing adoption: An analysis of the manufacturing and services sectors. *Information & Management* **51**:5, 497-510. [CrossRef]
2. References 229-257. [CrossRef]
3. Louay Karadsheh, Samer Alhawari. 2013. Applying Security Policies in Small Business Utilizing Cloud Computing Technologies. *International Journal of Cloud Applications and Computing* **1**:2, 29-40. [CrossRef]
4. Judith Mavodza. 2013. The impact of cloud computing on the future of academic library practices and services. *New Library World* **114**:3/4, 132-141. [Abstract] [Full Text] [PDF]
5. Weiling Liu, Huibin (Heather) Cai. 2013. Embracing the shift to cloud computing: knowledge and skills for systems librarians. *OCLC Systems & Services: International digital library perspectives* **29**:1, 22-29. [Abstract] [Full Text] [PDF]
6. Gobinda Chowdhury. 2012. An agenda for green information retrieval research. *Information Processing & Management* **48**:6, 1067-1077. [CrossRef]
7. Bay Arinze, Murugan Anandarajan. 2012. Factors that Determine the Adoption of Cloud Computing. *International Journal of Enterprise Information Systems* **6**:4, 55-68. [CrossRef]
8. Gobinda Chowdhury. 2012. How digital information services can reduce greenhouse gas emissions. *Online Information Review* **36**:4, 489-506. [Abstract] [Full Text] [PDF]
9. Louay Karadsheh. 2012. Applying security policies and service level agreement to IaaS service model to enhance security and transition. *Computers & Security* **31**:3, 315-326. [CrossRef]
10. Gobinda Chowdhury. 2012. Building environmentally sustainable information services: A green is research agenda. *Journal of the American Society for Information Science and Technology* **63**:4, 633-647. [CrossRef]
11. Alex Mu-Hsing Kuo. 2011. Opportunities and Challenges of Cloud Computing to Improve Health Care Services. *Journal of Medical Internet Research* **13**:3, e67. [CrossRef]
12. Bay Arinze, Murugan AnandarajanFactors that Determine the Adoption of Cloud Computing 210-223. [CrossRef]
13. Satish C. Sharma, Harshila BagoriaLibraries and Cloud Computing Models: 124-149. [CrossRef]
14. Gobinda G. ChowdhuryContent in the Cloud: 227-238. [CrossRef]
15. Louay Karadsheh, Samer AlhawariApplying Security Policies in Small Business Utilizing Cloud Computing Technologies 112-124. [CrossRef]
16. Satish C. Sharma, Harshila BagoriaLibraries and Cloud Computing Models: 72-97. [CrossRef]