TNC Transactions on Networks and Communications
SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
VOLUME 2, ISSUE 4
ISSN: 2054 - 7420

Impact of Data-Centre and User-Base Location On Overall Response-Time In A Cloud-Computing Environment

Amina Rashid, Javed Parvez
Department of Computer Science University of Kashmir, Srinagar, India
amina.rashid44@gmail.com; javed_parvez@kashmiruniversity.ac.in

ABSTRACT

Cloud computing is a key component as well as a measure of success for various organizations today. Apart from benefits obtained, it is important to take into account the location of user-base and data-centre, which is essential for performance and security reasons. This information is required since the location of data-centre and user-base can impact the overall response time. In this paper we evaluate the effect on overall response time, of relevant factors such as the location of data-centre and the serviced user-base.

Keywords: Hypervisor, Server Consolidation, Cloud Service Provider (CSP)

1 Introduction

Cloud is essentially a metaphor for the Internet [10].Cloud computing is considered nowadays as a fast growing area in computing research and industry as well. Cloud Computing uses an approach wherein platform, hardware, service are treated as a utility. This utility is metered in cloud computing environment.

Cloud computing is a model, wherein pooling of available shared resources is done. It may mean data centre hosting and understood as utility computing or grid computing [1][2].Cloud computing aims at offering distributed, virtualized and elastic resources as utilities to end users, and has the potential to support full realization of "computing as utility" in near future[14].In cloud computing, there are two important terms data-centre and user-base. Data-centre is used for providing services to the users whose requests are directed to it. User-base can be any organization or a small company, comprising a cluster of users, which gets its requests catered by the data-centre.

In cloud computing, the concept of virtualization technology is used, which enables service providers to run virtual machines and also to share the underlying resources. The software layer which provides virtualization is called hypervisor. Hypervisor emulates the underlying hardware resources with respect to different operating systems. Operating system has the direct access to the underlying hardware. But in case of virtualization, operating systems access

DOI: 10.14738/tnc.24.358
Publication Date: 24th August 2014
URL: http://dx.doi.org/10.14738/tnc.24.358

hardware through hypervisor. The hypervisor executes the privileged instruction on behalf of the virtual machine.

In cloud computing, virtualization technology is used to dynamically allocate or deallocate the resources for an application. Virtualization also helps to co-locate virtual machines to a small number of physical machines, so that the number of active physical machines can be reduced. This approach is called as server consolidation.

1.1 Cloud Computing: Security Concerns

Most of the users of cloud have expectations for the security of their data .The cloud owner and operators have responsibility of ensuring various security measures, and standards and protocols followed appropriately. There are two main aspects of security controls in cloud implementation. First, the presence of control and the second is the effectiveness or robustness of control. Associated with cloud computing are various security concerns. Some of these include:

1. Network Availability: Cloud computing value can be realized only if the network connectivity and bandwidth meet the minimum needs.
2. Cloud Provider Viability: Cloud Providers are relatively new to business and hence there are questions regarding the viability and commitment of the provider.
3. Disaster Recovery and Business Continuity: The users of cloud service require to have confidence that their operations and services will continue even if the cloud providers production environment is subject to disaster.
4. Security Incidents: The users need to be appropriately informed by the provider when an incident occurs. Users may also require provider support to respond to audit or assessment findings
5. Transparency: When a cloud provider does not expose details of their internal policy or technology implementation, the users must trust the cloud provider's security claims.

1.2 Business and IR Perspective

Business organizations are now compelled to deliver IT-enabled services via Internet that are built for end-user to be in control is what is now known as cloud computing. Cloud computing is emerging consumption and delivery model. It enables provisioning of standardized business and computing services through shared infrastructure, wherein end-user controls interaction to accomplish business task. Enterprise resources like processing power, storage, databases are no longer confined to enterprise only. Now, abstract or virtual resources are tapped whenever needed. From computing resources point of view, everything is provisioned by cloud.

2 Cloud Service Provider (CSP)

Business establishments put up a constant pressure on their IT managers for reduced budgets. In current scenario, the need of flexibility and competitive edge are essential requirements for business [8]. Cloud vendors are experiencing growth rates of 50% per annum[11].Such

scenarios lead to the requirement of cost effective and efficient solutions, which are very well provided by cloud computing environment, wherein storage and computing are provided by the infrastructure not owned by the organization. When adopting a public cloud, consumer does not need to be operationally concerned with the details of the underlying cloud infrastructure. However, there are several questions for customers that have to do with security and governance of the cloud service.

Customers of a public cloud service have expectations that the data they put into the service will have integrity and be protected. Customers trust that the CSP will offer the appropriate level of security and governance.

The Cloud Service Provider (CSP) should definitely have the capability to continue the services despite any disaster conditions, if they occur, which may include earthquakes, flood, fire etc. This capability is expected from CSP because the disaster can affect the cloud, and hence recovery measures or plans should be followed periodically and tested. The CSP should also provide a recovery process which in itself should not compromise the security of data.

Cloud Service Providers provide business continuity, recovery, backup through self-healing, but this is not possible to determine with any specificity where data processing takes place within cloud infrastructure [13].

Cloud providers have to safeguard the privacy and security of personal and confidential data that they hold of any organization and users. It is essential for cloud providers to ensure to their users that the security of their data is not compromised.

Hence, various questions and security risks are involved in itself while selecting a cloud service provider, especially while considering the factor of security of data. There are various security-related issues which need to be considered for a cloud service provider, some of the associated concerns are as follows:

1. Policies for Security: Organization which have strict enforcement of security policies, surely give an indication of how seriously the organization is taking the responsibility for security.
2. Independence of Security Staff: Organization maintaining separate staff for security and operations within the organization. Security staff can report independently but need to work in close cooperation to the operations staff.
3. Changes Documented: Any change should be documented well, reviewed and also approved. Change made can be in any aspect but the
4. Authority to make changes and how should be well delineated.
5. Safe Upgrades and also Patch Management: Safe and timely, upgrades and patch management should be done to limit exposure and hence providing security on an on-going basis.

6. Timely Scans: There should be timely scans made to infrastructure and vulnerability assessment should be made. Any issues detected should be evaluated in respect to their potential impact and immediately required corrective steps should be taken.
7. Forensics and Legal Management: To meet forensics and legal requirements any security logs should be maintained long enough. Security logs contribute to knowledge which may provide proper information regarding any breach of security if occurred and hence enabling to understand the scope and its potential impact.
8. Management of Incidents: The customer or a user of cloud should be well aware of any security related incident and the process related to it, hence being transparent to it and the same should also be kept well documented.

Cloud users depend on the providers for the services. These include three types of categories[15]:

I. Cloud Service Provider: Cloud Service Provider, having direct but contractual relationship with the consumer of the service .
II. Cloud infrastructure provider: Provides Cloud Service Provider with some form of infrastructure.
III. Communication Service Provider: Provides transmission service enabling consumer to communicate with Cloud Service Provider.

3 Data Centre Location

Considering the various benefits provided by using cloud computing or cloud service, one tends to forget the importance of the location where data will be stored or in other words the location where cloud is installed. Pooling resources in cloud model allows for greater flexibility and innovation for dynamic business demands [9]:

Traditionally, in a data centre, each application runs in silos, silo is sized for peak load. Here, there is no way to share the capacity between silos, we

need to carry enough capacity for peak workload of every application. Moving to shared or pooled resources will increase utilization rates and carry enough capacity spread across all workloads.

Pooling resources in cloud model allows for greater flexibility and faster innovation for dynamic business demands. If your business is growing fast, has high frequency of new projects, or experiences a sudden spike in demand, we can build new solutions for each of those initiatives much faster without affecting overall performance.

Rise in public computing utilities has led to increased need for better security of the data. Not known to many is the fact that the location of data or the data centre is governed by certain

laws, under whose jurisdiction they fall. The location of the data centre that implements the service utility has both direct and indirect implications. The customer must be aware of the jurisdiction of the nodes that form the cloud fall in.There are certain laws governing the transfer of data, and also what kind of data can be transferred, like what personal data can be collected and where it can be transferred to, even if this transfer is required for backup process.

The main compliance concerns with transborder data flows include whether the laws in the jurisdiction where the data was collected permit the flow, whether those laws continue to apply to the data post-transfer, and whether the laws at the destination present additional risks or benefits [4]. Technical, physical and administrative safeguards, such as access controls, often apply. For example, European data protection laws [5] may impose additional obligations on the handling and processing of European data transferred to the U.S.

The use of cloud computing has increased, this could lead to reduction in demand for high storage capacity consumer end devices, due to cheaper low storage devices that stream all content via cloud is becoming more popular. Jake Gardner explains that unregulated usage is beneficial for IT and tech moguls like Amazon, anonymous nature of cost consumption of cloud usage makes it difficult for business to evaluate and incorporate it into their business plans [12].

4 Simulation Framework, Setup & Result Analysis

Various simulation toolkits have been developed in market that can be very well used for simulating the cloud environment and thereby providing an understanding of large scaled applications floating on internet. Out of the various simulation toolkits available in the market, we are making use of cloud analyst simulation toolkit. This toolkit separates simulation experiments from programming exercise, which in return enables the analyst to dig out on the simulators parameters rather than concentrating on the programming part of it.The graphical output provided by the simulator is one of the key features, highlighting the response time and data processing time for the analyst.

Here we are trying to show that the location of data and data centre does not only affect the security issues and legal issues but also affect the overall response time. In the cloud analyst simulator the world is divided into 6 regions that coincide with the 6 main continents in world. Depending upon the location of data centre the overall response time between the data centre and user base is highly affected, this was established by dividing the entire coverage area into six regions, labelled as R0, R1, R2, R3, R4, R5, respectively.

Six scenarios were created wherein all other aspects of simulation were kept constant which included:

I. Simulation Duration:60 min
II. Service Broker Policy: Closest Data Centre
III. Load balancing policy across VMs in single data centre: Round Robin

IV. Request Grouping factor in UserBase:10
V. Request Grouping factor in DataCentre:10
VI. Executable instruction length per request:100

Using above parameters and changing the location of the user base and keeping the location of data centre as constant 6 scenarios were evaluated as under:

Scenario 1 : Data Centre: Region 0 User Base: Region 0

Overall Response Time:

Average(ms)	Minimum(ms)	Maximum(ms)
50.09	39.55	61.61

Data Centre Processing Time:

Average(ms)	Minimum(ms)	Maximum(ms)
0.48	0.01	0.86

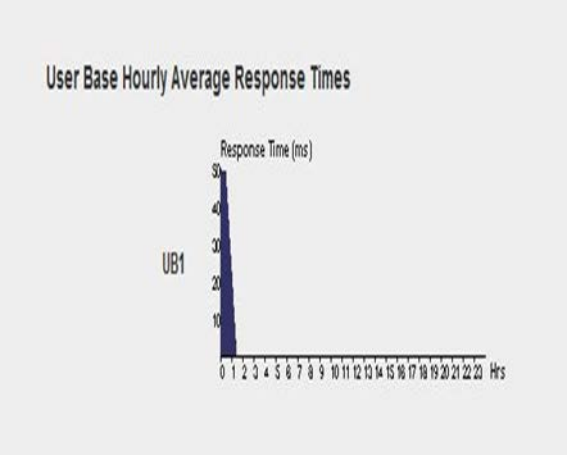

Figure 1: User Base Hourly Average Response Time

Scenario 2 : Data Centre: Region 0 User Base: Region 1

Overall Response Time:

Average(ms)	Minimum(ms)	Maximum(ms)
200.07	158.06	246.12

Data Centre Processing Time

Average(ms)	Minimum(ms)	Maximum(ms)
0.10	0.02	0.11

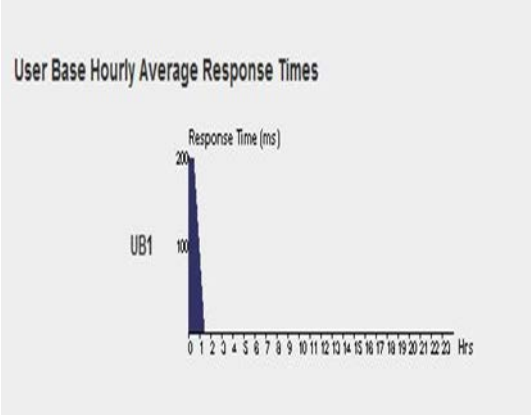

Figure 2: User Base Hourly Average Response Time

Scenario 3 : Data Centre: Region 0 User Base: Region 2

Overall Response Time:

Average(ms)	Minimum(ms)	Maximum(ms)
300.06	237.06	369.12

Data Centre Processing Time

Average(ms)	Minimum(ms)	Maximum(ms)
0.34	0.02	0.61

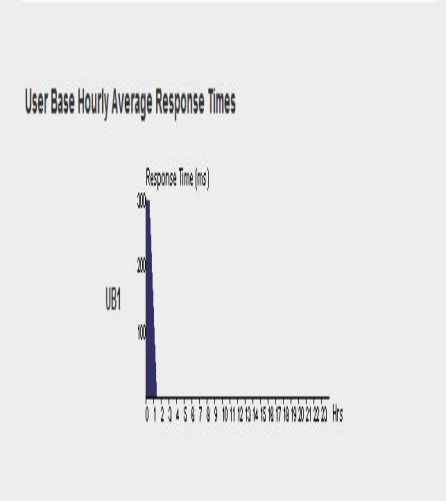

Figure 3: User Base Hourly Average Response Time

Scenario 4 : Data Centre: Region 0 User Base: Region 3

Overall Response Time:

Average(ms)	Minimum(ms)	Maximum(ms)
500.02	395.06	615.12

Data Centre Processing Time

Average(ms)	Minimum(ms)	Maximum(ms)
0.34	0.02	0.61

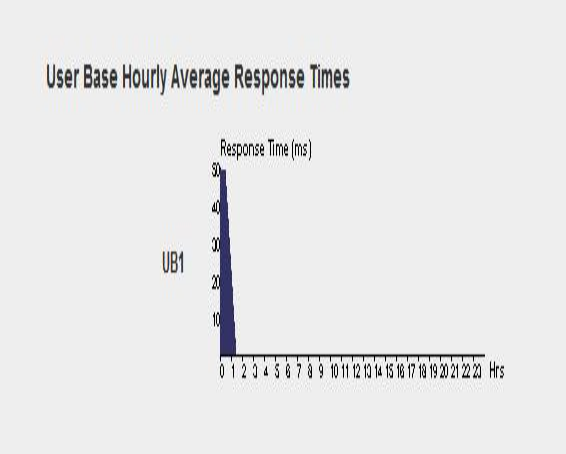

Figure 4: User Base Hourly Average Response Time

Scenario 5 : Data Centre: Region 0 User Base: Region 4

Overall Response Time:

Average(ms)	Minimum(ms)	Maximum(ms)
500.02	395.06	615.12

Data Centre Processing Time

Average(ms)	Minimum(ms)	Maximum(ms)
0.34	0.02	0.61

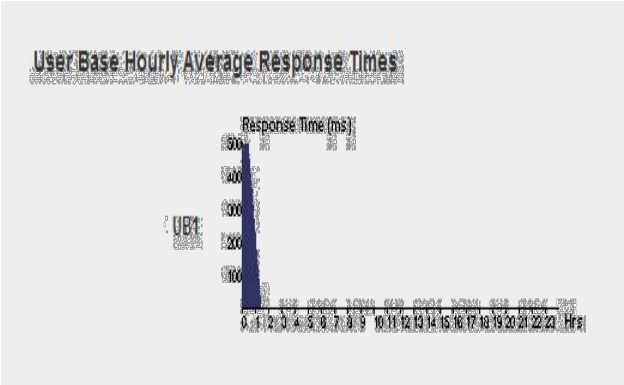

Figure 5: User Base Hourly Average Response Time

Scenario 6 : Data Centre: Region 0 User Base: Region 6

Overall Response Time:

Average(ms)	Minimum(ms)	Maximum(ms)
200.07	158.06	246.12

Data Centre Processing Time

Average(ms)	Minimum(ms)	Maximum(ms)
0.10	0.02	0.11

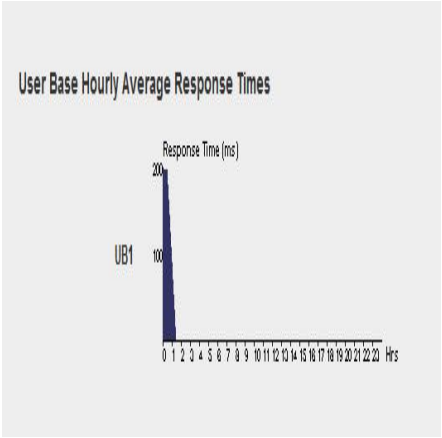

Figure 6: User Base Hourly Average Response Time

From the above scenarios we can easily confer that if the data centre and the user base are in the same region the overall response time calculated at average is very low and goes on increasing as the location of user base is altered, whileas the location of data centre is kept same, which was previously set to region 0 i.e. R0.

Also from the above scenarios it can be easily viewed that the overall response calculated for the regions R1 and R5 is same and so are their data processing time. Similarly, the overall response time calculated for regions R3 and R4 are same and so are their processing times. The overall response time calculated for region R0 is far less compared to any other region in which

user base calculated, hence confirming that if the region of the data centre and user base is same then the response time is far less. Similarly, the overall response times of regions R0 and R2 are entirely different from other regions and do not match to other regions. The data centre processing time calculated for region R0 is entirely different, whileas the data centre processing times calculated for regions R2, R3 and R4 are same.

5 Conclusion

At various instances cloud computing is advocated as providing infinite capacity on demand, but the real part of it is that the capacity of data centres in real world is finite[6].Installing and setting up data centre is not an easy task, it involves various sensitive issues like political and legal issues. Cloud computing is a popular paradigm now-a-days, wherein cloud providers offer scalable resources over Internet. Various providers like Amazons EC2, Google's AppEngine, IBM's Blue Cloud, and Microsoft's Azure provide services to the customers which include primarily storage and computing [7]. From the above scenarios we can well comprehend that there are some regions which have same overall response time, whileas some other regions have less overall response time compared to other. Now depending upon the allowance in a particular region, the cloud service provider can be allowed or rejected to set up a data centre in that region. This in return can affect the performance provided by the cloud service provider in respect of overall response as seen in above scenarios. Also it can lead to an additional charge to the customer if the data centre it is accessing does not fall in the region in its specified jurisdiction.Cloud computing is considered to be in its initial stages,lot more needs to be explored in this area.Various issues are yet to be solved,some of which include the audit.

REFERENCES

[1]. Lijun Mei,W.K.Chan,T.H.Tse,"A Tale of Clouds: Parad igm Comparisons and Some Thoughts on Research Issues",I EEE Asia-Pacific Services Computing Conference,2008

[2]. Thomas B Winans,JohnSeely Brown, "Cloud Computing, A Collection of working papers",2009.

[3]. http://technet.microsoft.com/en-us/magazine/hh536219.aspx

[4]. Amina Rashid,JavedParvez,"Security Issues in Cloud Computing:An Overview"

[5]. http://ec.europa.eu/j ustice/ data -protection/index_en .html.

[6]. Zhen Xiao,Qi Chen, Haipeng Luo," Automatic Scaling of Internet Applications for Cloud Computing Services".

[7]. Haiying Shen*, Guoxin Liu," An Efficient and Trust worthy Resource Sharing Platform for Collaborative Cloud Computing" .

[8]. http://viewer.media.bitpipe.com/1078177630_947/1267474882_422/WP_DC_DataCenterCloudComputing1.pdf.

[9]. http://www.oracle.com/us/products/database/cloud-computing-guide-1561727.pdf.

[10]. http://www.netlingo.com/word/cloud-computing.php.

[11]. http://www.fsn.co.uk/channel_outsourcing/the_economy_is_flat_so_why_are_financials_cloud_vendors_growing_at_more_than_90_perce nt_per_annum#.UbmtsPlJPGA/

[12]. http://www.wired.com/2013/01/beware-7-sins-of-cloud-computing.

[13]. McKinley,P.K.,Samimi,F.A.,Shapiro,J.K.,Chipping,T.:Service clouds:a distributed infrastructure for constructing autonomic communication services,In : Dependable,Autonomic and Secure Computing,IEEE,12-14 Dec 2011,Sydney,Australia,341-348(2006)

