Global Journal of Computer Science and Technology
Cloud and Distributed
Volume 13 Issue 1 Version 1.0 Year 2013
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Cloud Computing: Performance Implications and Challenges

By Abhishek Aggarwal & Dr. Sona Malhotra
Kurukshetra University, India

Abstract - Cloud computing is a new paradigm in the field of distributed computing. The objective of cloud computing is to provide various computing resources over the internet in the form of service to number of cloud consumers. Cloud provides the computing environment to organization in a cost effective manner and give flexibility to increase the number of resources as required during peak load time. In this paper we have tried to highlight some of the major challenges like security, availability of cloud services, reliability and auto-provisioning of cloud resources etc. which need to be addressed by researchers. Certainly there are some performance implications which also need to be resolved in order to get maximum output from the cloud so we need to manage the cloud resources in optimized way to increase the performance of cloud and its adaptability among different organization.

Keywords : *cloud, services, performance implications, cloud bursting, load balancing.*

GJCST-B Classification : *C.2.4*

Strictly as per the compliance and regulations of:

Cloud Computing: Performance Implications and Challenges

Abhishek Aggarwal [α] & Dr. Sona Malhotra [σ]

Abstract - Cloud computing is a new paradigm in the field of distributed computing. The objective of cloud computing is to provide various computing resources over the internet in the form of service to number of cloud consumers. Cloud provides the computing environment to organization in a cost effective manner and give flexibility to increase the number of resources as required during peak load time. In this paper we have tried to highlight some of the major challenges like security, availability of cloud services, reliability and auto-provisioning of cloud resources etc. which need to be addressed by researchers. Certainly there are some performance implications which also need to be resolved in order to get maximum output from the cloud so we need to manage the cloud resources in optimized way to increase the performance of cloud and its adaptability among different organization.

Keywords : cloud, services, performance implications, cloud bursting, load balancing.

I. Introduction

Cloud computing is treated as a new model for computing which aims to provide reliable, low cost and customizable as per the requirement of user and guaranteed computing dynamic environments for end-users over the internet. It aims for enabling convenient, on demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction [1].The advantages of cloud computing over traditional computing include: agility, lower entry cost, device independency, location independency and scalability. There are lots of characteristics and services offered by cloud computing which makes it different than traditional web services or service oriented architecture. We can understand cloud computing on the basis of its type i.e. how we manage the cloud in order to fulfill the requirement of cloud consumers and type of services provided by the cloud. On the basis of services model there are four types of services offered by a cloud i.e. SaaS (Software as a Service), PaaS (Platform as a Service), IaaS (Infrastructure as a Service), and DaaS (Data as a Service) a kind of IaaS.

Author : Ph.D Research Scholar at Deptt. of CSE in UIET, Kurukshetra University, Kurukshetra, Haryana, India.
E-mail : abhisoft@rediffmail.com
Author : Working as Assistant Professor at Deptt. of CSE in UIET, Kurukshetra University, Kurukshetra, Haryana, India.
E-mail : sona_malhotra@yahoo.com

In contrast with the service model, on the basis of deployment model we can categorize cloud into four different type i.e. private cloud, community cloud, public cloud and hybrid cloud. There are a number of challenges which are addressed by researchers and practitioners in the field of cloud computing as briefly presented as Performance, Security & Privacy, Platform Control, Bandwidth Cost, Interoperability, Service availability and Reliability etc. In the cloud computing environment, when we shift ourselves to public cloud there may be several factors affecting the performance of service as listed as 1) Delay in services 2) Availability of services 3) Different standards of cloud vendors 4) Data location and relocation 5) Degree of coupling among computing components etc.

This Paper consists of various sections, In the overview section we explained the cloud computing, its different models and architectures of cloud services. Thereafter, we explained the different performance issues and challenges while migrating to the public cloud and then we find the future research scope in the field of cloud computing model. Finally, conclusion is drawn in last section.

II. Overview of Cloud Computing

Cloud Computing is a branch of computing to provide shared pool of customizable Resources like Application, Platform and infrastructure as a service to different cloud consumers, SME and other cloud vendors. There are two models in the cloud computing, one on the basis of services and other on the basis of deployment. Cloud computing can be viewed as a collection of services, which can be presented as a layered cloud computing architecture, as shown in fig.1

a) Types of Service

SaaS appears on the top of stack in fig. 1 and allow cloud consumers to access the services remotely and on rental basis i.e. “pay-as-you-go”. Consumers may access the services depends upon SLA “Service Level Agreement”. It save the users from the troubles of software deployment and maintenance, and, software is often shared by the multiple users, automatically updated in cloud and no additional licenses need to be purchased.

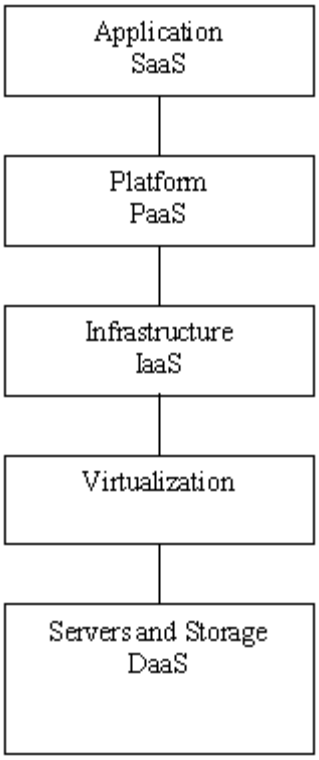

Figure 1 : Layered architecture of cloud computing

Web based mail and Google Docs are best examples of SaaS.

PaaS is a service which enables the users to develop their own applications using the platform of different cloud vendors. It provides complete development environment with a set of services to design, develop, test, deploy and monitor the application on the cloud [4]. End user may not know that on the cloud, which server hosts the application. Storage space of the application may be increased or decreased as per the need of application. Google App Engine and Microsoft Azure are examples of PaaS.

IaaS is a service in the form of infrastructure. So, instead of having high cost data centre maintaining at their own end consumers may use the storage and computing resources like CPU, Hard Disk or other I/O devices from different vendors. Amazon EC2 and Rack space are examples of IaaS. IaaS is virtualized over the set of different servers which may be physically located to different locations so the cloud vendors may setup VM in order to process the user's request uninterruptedly. DaaS is a form of IaaS where logical vs. physical mapping need to takes place using virtualization.

There are lot of risks and advantages associated with different service models e.g. in SaaS user has very limited scope of customization and difficulty may arises in data integration. In contrast to this user does not need to worry about the updates of service.

b) Characteristics of Cloud Computing

There are some essential characteristics of cloud computing services as follow:

i. *Sharing of resources using virtualization*

In a cloud environment, multiple computing resources of different kind may be pooled together and virtualized in order to provide services to different category of users to support multi-tenant model. Resources are dynamically assigned as per the demand of users. Sharing of resources enable the economy of scale and specialization. Specialized resources are pooled to cater the users of one category. Resources are hidden from users and consumers who have no idea about the physical location of resources like CPU, Storage and DBMS etc using virtualization.

ii. *Demand may change very rapidly*

All the resources must be readily available as per the demand of users. Cloud vendors must immediately fulfill the requirement of consumers and release the resources when the task is completed. Vendors must calculate the peak load of all the consumers in order to provide uninterrupted services to them.

iii. *Measuring the services used*

Cloud infrastructure is able to provide some mechanism to measure the services being used by consumer and generate appropriate billing such that no conflict arises. Some monitoring services may be used for the accurate measurement.

iv. *On-demand self service*

Cloud is able to provide set of services automatically without human interaction. User interface may be provided to avail the services and check the usage and billing information with complete transparency.

v. *Support for Different heterogeneous devices*

Computing resources may be used over the internet by broad range of devices and platform using client application (e.g. Web Browser).

c) Deployment Models

On the basis of where cloud services have been deployed clouds are categorized in to different categories like:

i. *Private Cloud*

These are proprietary networks normally resides and most often used by the organizations. All the services are deployed and organized within the organization managed by third party or organization itself. Private cloud doesn't make any sense because a lot of infrastructural and management cost is involved. Only mission critical application should be deployed as a part of private cloud in order to secure them from outside attack. In this type of cloud data and applications are more secure but special management skills are required to maintain it at their own end.

ii. *Public Cloud*

It is the main stream of cloud computing where services are publically deployed. Data and applications are hosted by third party and managed by service providers. Resources are provided free of cost or by charging an amount on pay-as-you-go basis. Cloud vendors shall be fully responsible for availability of service or computing resources but careful supervision is required by the enterprises to check the services of cloud providers.

iii. *Community Cloud*

Where set of computing resources are shared by a particular group of community instead publically. People who have similar and shared backgrounds and requirements may form the community cloud, this way we can reduce the computing cost and increase the security by limiting the access of resources to a particular set of users.

iv. *Hybrid Cloud*

Sometimes it is required to deploy data and applications within the organization (for mission critical processes) and sometimes for an external or outside organization. Hybrid cloud may target very effectively such organizations because of enhanced control and management by the enterprise itself. A clear cut distinction should be made between management responsibilities of the organization and cloud vendors.

v. *Virtual Private Cloud*

It is a secure and seamless bridge between an organization's existing IT infrastructure and the public cloud. It is public because it uses the computing resources of public cloud for users; however it is virtually private because the connection between IT legacy and cloud is secured through a virtual private network. Thereby having a security advantage of private cloud user can still enjoy pay-as-per-use on these public isolated resources.

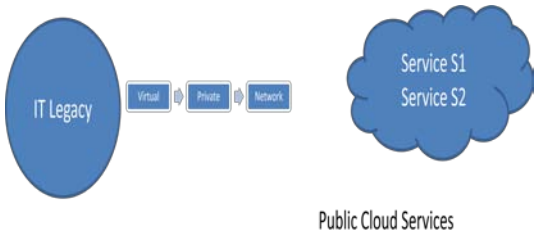

Figure 2 : Virtual Private Cloud

VPC is a perfect balance between the private cloud for control purpose and public cloud as far as the flexibility is concerned.

III. Performance Issues and Challenges

a) Performance Degradation

As we know internet is the back bone of cloud computing so performance is limited by the speed of internet. Data intensive or transaction intensive applications are highly effective when migrating to the public cloud. So, for such applications, performance is a major concern. We should provide some mechanism to categorize the applications and decide to migrate it to the public cloud or not.

b) Security & Privacy

It is the most important issue in the area of cloud computing. According to M. Kretzschmar [5] for collaborative cloud some cloud security management issues are there. Cloud security management infrastructure has to be managed and integrated with in cloud security management system.

c) Availability

Sometimes demand increases very rapidly and resources are not available, resulting delay in the services, so, non availability of services[8] during the peak load is a major concern. It may be overcome by distributing the load to some other resources and balancing the load to various resources within cloud or to other cloud also. some of the issues like non availability of service which lead to lack of reliability, outage and vendor lock-in etc. are highlighted and targeted by proposed 3-tier cloud deployment architecture[11] over two tier deployment architecture. An addition service provide layer which consists of four components outage handling data centre, interface to cloud, value added services and interface to client.

d) Lack of Support for multi-tenancy

In the cloud computing, different types of users are using services from the cloud. Sometime a company of repute share resources with notorious user with a criminal mind, so in multi-tenant environment security of data is always a major concern. Security issues has played the most important role in hindering cloud computing as the resources are being shared by multiple users using multi-tenant model.

e) Interoperability

Sometime cloud vendors are not able to cater the services to cloud consumers and may seek specialized services to other cloud vendors. Integration of services within a different cloud or with existing legacy system is always a headache for the cloud vendors. Lack of standards being adopted in the industry creates problem of interoperability. Existing legacy IT needs integration with different clouds in order to user services as described in Fig 3.

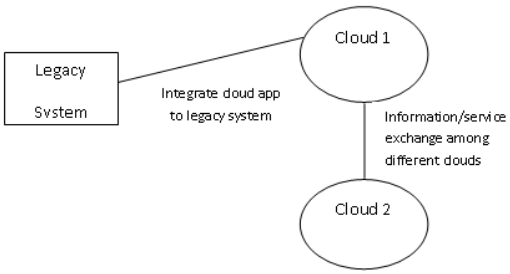

Figure 3 : Integration of services in different clouds

As cloud computing matures, the ability to support interoperability becomes more important [6].

f) Service Level Agreement

SLA is one of the major issues in cloud computing. Lack of well defined SLA by cloud providers lead to a problem for cloud consumers. What is guaranteed uptime? What are the repercussions if the provider fails to meet the standards? What happens to customer data if the company moves to different provider? There are some common questions which arise in the mind of users and there is a need to give sufficient attention to answer these questions. Some standards need to design for setting up SLA in a proper manner. SLA specifications need to be provided in such a way that they can cover most of the consumer expectation and resource allocation mechanism on the cloud.

g) Performance instability and Load balancing

It has been found in various surveys that Amazon, Microsoft and Google suffered from variations in performance and availability due to variations in the load. Specifically, the researchers measured how the cloud vendors scaled up and responded to immediate requirements of 3000 concurrent cloud users resulting sometimes; we can't predict the changes in performance because of the variations in demand. In order to manage the resources efficiently S. Wang et.al. [9] Proposed a two phase load balancing algorithm that combines OLB and LBMM scheduling algorithm in three level cloud computing environment. R.Lee et.al. [10] described two new load balancing policies in dynamic manner. These policies dispatches workload based on the dynamic comparison of the latest resource capacity available in each server. Unfortunately, Server capability varies in practice and is not easy to record in ordered position, which will cause non resource aware load balancing algorithms to distribute workloads evenly.

h) Data Storage and Data Processing

Every Enterprise has some set of sensitive and confidential data which needs to be carefully stored and processed. Each cloud consumer worried about the location of their organizational data so special care shall be given to the location [14] and processing techniques for such private data. Moreover data needs to transfer from one cloud to another, so special attention and preventive measures should be given to the migration and security of data respectively. For Data intensive application performance is purely depends up on high speed internet connection [15].

i) Resource Provisioning Policies

Cloud has a capability to provide resources on demand so an auto provisioning of resources must be supported by cloud in the peak load hours. In order to improve the performance of cloud during peak load a number of resource provisioning policies have been evolved. A.Iosup et.al. [7] analyses provisioning and allocation policies over three IaaS clouds , including Amazon EC2. He has compared various static and dynamic provisioning policies and their performance in different workload pattern over different IaaS clouds. Scheduling of jobs is also a core and challenging issue in cloud computing. L.Li et.al. [12] analyses the different QoS requirements of cloud computing resources. He builds the non-preemptive priority queuing model for the jobs and then build the system cost function. P. Gupta et.al. [13] explained different job scheduling methodologies for web application and web server in cloud computing environment. He has targeted various issues like virtual resources and queuing strategies.

IV. Related Research Scope

In Cloud computing research, both industry and academia have been active and several research activities have been carried out in past few years. Several architectures have been proposed to target the issues and performance implications discussed so far. We can throw some light on the core issues in the field of cloud computing. In order to minimize the cost, multi-tenant architecture has been proposed with minimum interference among cloud consumers. As we know computing components are tightly coupled so we need to minimize the coupling among components within the same cloud so that they may be used in intra-cloud environment and cater the needs of other cloud users also. Secondly we can carry out research on the scheduling of computing resources in order to give maximum resource utilization without making delay in the services to cloud consumers. There is a need to distribute the load of computing resources among different clouds for balancing the load where cloud interoperability remains one of the major concerns. We need to develop some standard techniques or protocols so that services may be frequently used among clouds without any interruption. When we access the services and data from other clouds the data shall migrate from one location to other, so special technique may be find out to secure the data during transmission. Because of widely relocation of data, cloud computing is always be the favorite place for the hackers, so special encryption technique needs to find out so that data shall be fully controlled and monitored by the cloud vendors and

consumers. A function 'Cloud Manager' [3] must be available, to at least assign the request to the server. Cloud Manager is like an intermediate between the client side and server side infrastructure. It performs the various functions at the gateway of cloud including monitoring of available capacity of various hosts, load balancing and usage accounting. We need to optimize the functionality of cloud manager in order to improve the efficiency and performance of cloud itself.

V. Conclusion

In this paper, the author discussed the performance implications and challenges in the field of cloud computing. We have discussed the different models of cloud computing as far as services and deployment of clouds are concerned. There is a wide scope of research in the field of cloud computing to target the issues and challenges discussed so far. Several architectures need to be evolved in order to give high quality cloud computing services.

References Références Referencias

1. Huhns M. N. and Singh M. P., "Service-Oriented Computing: Key Concepts and Principles," IEEE Internet Computing, vol. 09, pp. 75 - 81, 2005. M. N. Huhns, M. P. (2005). *Service-Oriented Computing: Key Concepts and Principles* (Vol. 09). IEEE Internet Computing.
2. Dabrowski C., "Reliability in grid computing system, concurrency and computation: practice and experience" pp. 927-959, 2009.
3. Clarke R. "User Requirements for Cloud Computing Architecture" in 10th IEEE/ACM international conference on cluster, cloud and grid computing, pp. 625-630, 2010.
4. Tsai Wei-Tek, Sun Xin, Balasooriya J. "Service Oriented Cloud Computing Architecture" in 7th International Conference in Information Technology, pp. 684-689, 2010.
5. M. Kretzschmar and S. Hanigk, "Security Management interoperability challenges for Collaborative Clouds", Proceedings of 4th International DMTF Academic Alliance Workshop on Systems and Virtualization Management, 2010, pp. 43-49.
6. Scott Dowell, "Cloud to Cloud Interoperability", Proc. of the 2011 6th International Conference on System of Systems Engineering, Albuquerque, New Mexico, USA, pp. 258-263.
7. D. Villegas, A. Antoniou, S. Sadjadi, A. Losup, "An analysis of provisioning and allocation policies for infrastructure-as-a-service clouds", Proceedings of 12th IEEE/ACM International Symposium on cluster, cloud and grid computing", 2012, pp. 612-619.
8. M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, M. Zaharia, "Above the Clouds: A Berkeley View of Cloud Computing", EECS, University of California at Berkeley, Tech.Rep.UCB/EECS-2009-28.
9. S. Wang, W. Liao, "Towards a Load Balancing in a Three-level Cloud Computing Network", Proceedings of 3rd International conference on Computer Science and Information technology, July 2010, pp. 108-113.
10. R. Lee, B. Jeng, "Load-Balancing Tactics in Cloud", Proceedings of 2011 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, pp. 447-454.
11. M. Phatak and Kamlesh. V.N, "On Cloud Computing Deployment Architecture", Proceedings of International Conference on Advances in ICT for Emerging Regions (ICTer), 2010, pp. 11-14.
12. L. Li, "An Optimistic Differentiated Service Job Scheduling System for Cloud Computing Service Users and Providers", Proceedings of 3rd International Conference on Multimedia and Ubiquitous Engineering, 2009, pp. 295-299.
13. P.K Gupta, N. Rakesh, "Different Job Scheduling Methodologies for Web Application and Web Server in a Cloud Computing Environment", Proceedings of 3rd International Conference on Emerging Trends in Engineering and Technology, pp. 569-572.
14. Z. Mahmood, "Data Location and Security Issues in Cloud Computing", Proceedings of International Conference on Emerging Intelligent Data and Web Technologies, 2011, pp.49-53.
15. Khalid, H. Mujtaba, "Data Processing Issues in Cloud Computing", Proceedings of 2nd International Conference on Machine Vision, 2009, pp. 301-304.

This page is intentionally left blank