www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 5 Issue 12 Dec. 2016, Page No. 19288-19295

A Survey of Load Balencing Technique in Cloud Computing Environment

Manisha Verma, Somesh Kumar
Noida Institute of Engineering & Technology, Noida
Affiliated with AKTU, Lucknow, UP, India

Abstract : ***Cloud computing refers to service delivery over internet by several application which are in distributed data centers. Cloud computing has many advantages along with some issues. These issues are related with load management, reliability, data portability, various security issues and much more. In this paper our main concern is load balancing algorithms in cloud computing. The load can be network load, memory capacity, CPU load etc. The load balancing is a process of reassigning the total load to the individual nodes of the collective system to make resource utilization effective and to improve the response time of the job, simultaneously removing a condition in which some of the nodes are over loaded while some others are under loaded. This paper presents various issues of cloud computing especially related to load balancing and various load balancing algorithms or technique in cloud computing adopted in past research work have been analyzed and findings were illustrated in this paper.***

Keywords*:* Cloud computing, Load balancing

I. INTRODUCTION

Cloud computing is a model 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]. Cloud computing services are becoming the primary source of computing power for both enterprises and personal computing applications. A cloud computing platform can provide a variety of resources, including infrastructure, software, and services, to users in an on-demand fashion. To access these resources, a cloud user submits a request for resources. The cloud provider then provides the requested resources from a common resource pool (e.g., a cluster of servers), and allows the user to use these resources for a required time period. Compared to traditional "own-and-use" approaches, cloud computing services eliminate the costs of purchasing and maintaining the infrastructures for cloud users, and allow the users to dynamically scale up and down computing resources in real time based on their needs. Several cloud computing systems are now commercially available, including Amazon EC2 system , Google's AppEngine, and Microsoft's Azure.s.

Load balancing is one of the central issues in cloud computing [28]. It is a mechanism that distributes the dynamic local workload evenly across all the nodes in the whole cloud to avoid a situation where some nodes are heavily loaded while others are idle or doing little work. It helps to achieve a high user satisfaction and resource utilization ratio, hence improving the overall performance and resource utility of the system. It also ensures that every computing resource is distributed efficiently and fairly . It further prevents bottlenecks of the system which may occur due to load imbalance. When one or more components of any service fail, load balancing helps in continuation of the service by implementing fair-over, i.e. in provisioning and de-provisioning of instances of applications without fail. It also ensures that every computing resource is distributed efficiently and fairly [28] . Consumption of resources and conservation of energy are not always a prime focus of discussion in cloud computing. However, resource consumption can be kept to a minimum with proper load balancing which not only helps in reducing costs but making enterprises greener [29] . Scalability which is one of the very important features of cloud computing is also enabled by load balancing.

The rest of paper is organised as follows. Section II introduces the overview of cloud computing. Section III introduces the issues of cloud computing environment. Section IV introduces the over view of load balancing algorithms. Section V introduces the various load balancing algorithms which are used in cloud computing environment. Finally, section VI concludes this paper.

II. OVERVIEW OF CLOUD COMPUTING

Cloud computing is define as "Cloud computing is Internet-based computing, whereby shared resources, software, and information are provided to computers and other devices on demand, like the electricity grid. It is a style of computing in which dynamically scalable and often virtualized resources are provided as a service over the Internet[2].
Cloud computing is an on demand service in which shared resources, information, software and other devices are provided according to the clients requirement at specific time. Its a term which is generally used in case of Internet. The whole Internet can be viewed as a cloud. Capital and operational costs can be cut using cloud computing.

DOI: 10.18535/ijecs/v5i12.03

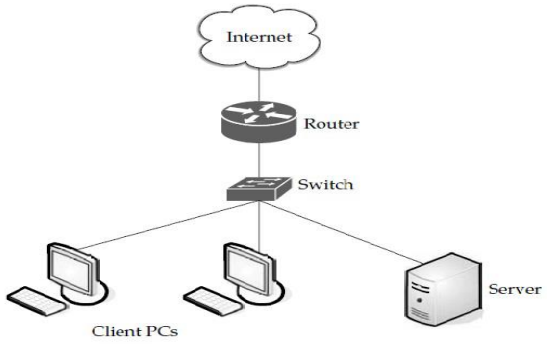

Figure 1: A cloud is used in network diagrams to depict the internet (adopted from [3]).

A. Component of Cloud Computing

A Cloud system consists of 3 major components such as clients, datacenter, and distributed servers. Each element has a definite purpose and plays a specific role

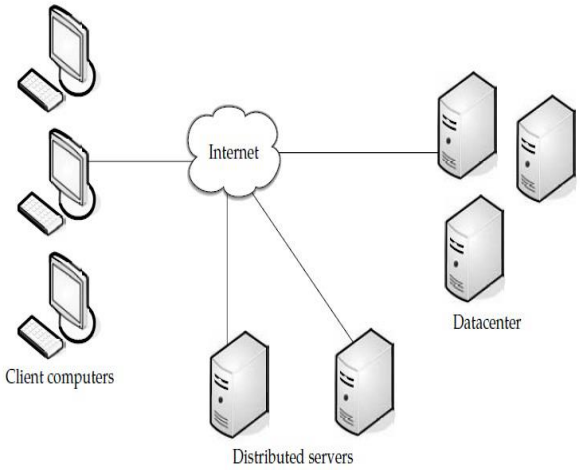

Figure 2: Components of Cloud [4]

i. Clients:

End users interact with the clients to manage information related to the cloud. Clients generally fall into three categories as given in [1]:

Mobile: Windows Mobile Smartphone, smartphones, like a Blackberry, or an iPhone.

Thin: They don't do any computation work. They only dispaly the information. Servers do all the works for them. Thin clients don't have any internal memory.

Thick: These use different browsers like IE or mozilla Firefox or Google Chrome to connect to the Internet cloud.

Now-a-days thin clients are more popular as compared to other clients because of their low price, security, low consuption of power, less noise, easily replaceble and repairable etc.

ii. Datacenter

Datacenter is nothing but a collection of servers hosting different applications. A end user connects to the datacenter to subscribe different applications. A datacenter may exist at a large distance from the clients.

Now-a-days a concept called virtualisation is used to install a software that allow multiple instances of virtual server applications.

iii. Distributed servers

Distributed servers are the parts of a cloud which are present throughout the Internet hosting different applications. But while using the application from the cloud, the user will feel that he is using this application from its own machine.

B. Architecture of Cloud Computing

Fig. 3 Shows the layered architecture of cloud computing. Cloud architecture is the design of software applications that uses internet-accessible on-demand service. Cloud architectures are underlying on infrastructure which is used only when it is needed that draw the necessary resources on demand and perform a specific job, then relinquish the unneeded resources and often dispose them after the job is done.

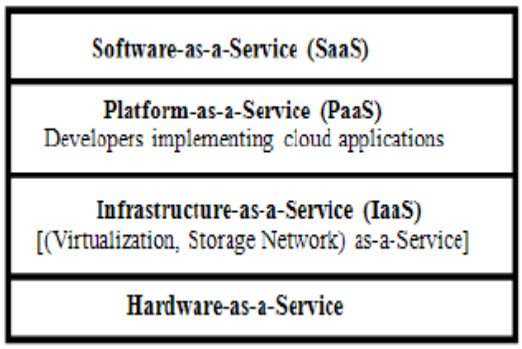

Fig. 3: Cloud Layered Architecture [5]

The services are accessible anywhere in the world, with the cloud appearing as a single point of access for all the computing needs of consumers. Cloud architectures address the key difficulties surrounding large scale data processing[5].

C. Service Model of Cloud Computing

There are three categories of cloud services such as infrastructure, platform, software. There services are delivered and consumed in real time over the internet.

i. Software-as-a-Service (SaaS):

SaaS focuses on providing users with business specific capabilities such as email or customer management. In SaaS organizations and developers can use the business specific capabilities developed by third parties in the "cloud". One of the example of SaaS provider is *Google Apps* that provides large suite of web based applications for enterprise use[6].

ii. Platform-as-a-Service (Paas):

Paas is a service model of cloud computing. In

this model clients create the software using tools and libraries from the provider. Clients also control software deployment and configuration settings. The provider provides the network, servers and storage.
One of the examples of PaaS is *Google App Engine* that provides clients to run their applications on Google's infrastructure[6].

iii. Infrastructure-as-a-Service (IaaS):

IaaS provides mainly conceptual infrastructure over the internet (e.g. compute cycles or storage). IaaS allows organizations and developers to extend their IT infrastructure on demand basis.
One of the examples of IaaS providers is *Amazon Elastic Compute Cloud (EC2)*. It provides users with a special virtual machine that can be deployed and run on the EC2 infrastructure[6].

D. Deployment Model of Cloud Computing

There are four primary cloud deployment model:
Public cloud
Private cloud
Community cloud
Hybrid cloud
Their differences lie primarily in the scope and accessed of published cloud services, as they are made available to services consumers.

CLOUD COMPUTING DEPLOYMENT MODELS	
Private Cloud	The cloud infrastructure is owned or leased by a single organization and is operated solely for that organization.
Community Cloud	Several organizations that have similar polices, objectives, aims and concerns share the cloud infrastructure.
Public Cloud	A large organization owns the cloud infrastructure and sells cloud services to industries or public.
Hybrid	It is combination of two or more clouds. It

Table 1 : Cloud Deployment models [7]

E. Advantage of Cloud Computing

Cloud computing offers various advantages such as :

Mobility: We don‟t need to carry our personal computer, because we can access our documents anytime anywhere [8].

Virtualization: In cloud computing, virtualization is a concept where users have a single view of available resources irrespective of their arrangement in physical devices. So it is advantageous for the providing the service towards users with less number of physical resources.

Scalability: Scalability is the capability of a system to increase total throughput under an increased load when resources are added. Resources can be hardware, servers, storage, and network. The user can quickly scale up or scale down the resources in cloud computing according to their need without buying the resources[8].

Maximized Storage: Users or clients in cloud computing can store more data in cloud than on private computer systems, which they use regular basis. It not only relieves them from buying extra storage space, but also improves performance of their regular system, as it is less loaded. On the other hand, data or programs are accessed anytime through internet, since they are available in cloud[8].

Low infrastructure cost - As in the clouds the user need not own the resources, it just need to pay as per the usage in terms of time, storage and services. This feature reduces the cost of owning the infrastructure.

Green Technology- The cloud computing is a green technology since it enable resource sharing among users thus not requiring large data centers that consumes a lot of power[9].

Fast Implementation- Time of Implementation of cloud for an application may be in days or sometimes in hours. You just need a valid credit card and need to fulfill some online registration formalities[9]

III. ISSUES OF CLOUD COMPUTING

There are various issues in cloud computing environment some of the technical issues in cloud computing will include load balancing, security, reliability, ownership, data back-up,data portability, multiplatform support, and intellectual property and many more. Here is a rundown on most of the current issues concerning cloud computing:

Security [10] :Usually security is the focal concern in terms of data, infrastructure, and virtualization etc. Corporate information is not only a competitive asset, but it often contains information of customers, consumers and employees that in the wrong hand, could create a civil liability and possibly criminal charges. Cloud computing can be made secure but securing cloud computing data is a contractual issues as well as a technical one.

Interoperability [5]:The issues of interoperability is needed to allow applications to be ported between clouds or to use multiple infrastructures before critical business applications are delivered from the cloud. Recently cloud computing interoperability forum(CCIF) was formed to define an organization that would enable interoperable enterprise cloud computing platform through application integration and stake holder cooperation.

Bandwidth, quality of service and data limits [11]:Cloud computing requires not just high speed, but also high quality broadband connections, that are always connected. Whilst any websites are usable on non-broadband connections or slow broadband connections; cloud-based applications are often not usable. Connection speed in Kilobyte per second (or MB/s and GB/s) is important for use of cloud computing services. Also important are Quality of Service (QoS); indicators for which include the amount of time the connections are dropped, response time (ping), and the extent of the delays in the processing of network data (latency) and loss of data (packet loss). If the benefits of cloud computing are to be reaped at a national development level then investment in access infrastructure, backbone infrastructure, the last-mile (or local loop).

Load Balancing [12]: Load balancing is a relatively new technique that facilitates networks and resources by providing a maximum throughput with minimum response time. Dividing the traffic between servers, data can be sent and received without major delay. Different kinds of algorithms are available that helps traffic loaded between available servers. A basic example of load balancing in our daily life can be related to websites. Without load balancing, users could experience delays, timeouts and possible long system responses. Load balancing solutions usually apply redundant servers which help a better distribution of the communication traffic so that the website availability is conclusively settled.

Service Level Agreement (SLA)[10]: Although cloud consumers do not have control over the underlying computing resources, they do need to ensure the quality, availability, reliability, and performance of these resources when consumers have migrated their core business functions onto their entrusted cloud. In other words, it is vital for consumers to obtain guarantees from providers on service delivery. Typically, these are provided through Service Level Agreements (SLAs) negotiated between the providers and consumers. The very first issue is the definition of SLA specifications in such a way that has an appropriate level of granularity, namely the tradeoffs between expressiveness and complicatedness, so that they can cover most of the consumer expectations and is relatively simple to be weighted, verified, evaluated, and enforced by the resource allocation mechanism on the cloud. In addition, different cloud offerings (IaaS, PaaS, and SaaS) will need to define different SLA metaspecifications. This also raises a number of implementation problems for the cloud providers. Furthermore, advanced SLA mechanisms need to constantly incorporate user feedback and customization features into the SLA evaluation framework.

Multiplatform Support [13]: More an issue for IT departments using managed services is how the cloudbased service integrates across different platforms and operating systems, e.g. OS X, Windows, Linux and thinclients. Usually, some customized adaption of the service takes care of any problem. Multiplatform support requirements will ease as more user interfaces become web-based.

Reliability [32]: Some people worry also about whether a cloud service provider is financially stable and hether their data storage system is trustworthy. Most cloud providers attempt to mollify this concern by using redundant storage techniques, but it is still possible that a service could crash or go out of business, leaving users with limited or no access to their data. A diversification of providers can help alleviate this concern, albeit at a higher cost.

IV. OVERVIEW OF LOAD BALANCING

It is a process of reassigning the total load to the individual nodes of the collective system to make resource utilization effective and to improve the response time of the job, simultaneously removing a condition in which some of the nodes are over loaded while some others are under loaded. A load balancing algorithm which is dynamic in nature does not consider the previous state or behavior of the system, that is, it depends on the present behavior of the system. The important things to consider while developing such algorithm are : estimation of load, comparison of load, stability of different system, performance of system, interaction between the nodes, nature of work to be transferred, selecting of nodes and many other ones [4] . This load considered can be in terms of CPU load, amount of memory used, delay or Network load.

A. Goals of Load balancing

The goals of load balancing are as follows :

i. To improve the performance substantially
ii. To have a backup plan in case the system fails even partially
iii. To maintain the system stability
iv. To accommodate future modification in the system

B. Types of Load balancing algorithms

Depending on who initiated the process, load balancing algorithms can be of three catagories :

i. Sender Initiated: If the load balancing algorithm is initialised by the sender
ii. Receiver Initiated: If the load balancing algorithm is initiated by the receiver
iii. Symmetric: It is the combination of both sender initiated and receiver initiated

Depending on the current state of the system, load balancing algorithms can be divided into 2 catagories :

i. Static: It doesnt depend on the current state of the system. Prior knowledge of the system is needed

ii. Dynamic: Decisions on load balancing are based on current state of the system. No prior knowledge is needed. So it is better than static approach.

C. Load Balancing Metrics

LOAD BALANCING METRICS	
Metric	Illustration
Throughput	It is used to calculate the no. of tasks whose execution has been completed. It should be high to improve the performance of the system
Overhead	It determines the amount of overhead involved while implementing a load-balancing algorithm. It is composed of overhead due to movement of tasks, inter-processor and inter-process communication. This should be minimized so that a load balancing technique can work efficiently.
Fault Tolerance	It is the time to migrate the jobs or resources from one node to other. It should be minimized in order to enhance the performance of the system.
Response Time	It is the amount of time taken to respond by a particular load balancing algorithm in a distributed system. This parameter should be minimized.
Resource Utilization	It is used to check the utilization of re-sources. It should be optimized for an efficient load balancing.
Scalability	It is the ability of an algorithm to perform load balancing for a system with any finite number of nodes. This metric should be improved.

Table 2 : Metrics in existing LB techniques in cloud computing

D. Need of Load Balancing in Cloud Computing

Load balancing in clouds is a mechanism that distributes the excess dynamic local workload evenly across all the nodes. It is used to achieve a high user satisfaction and resource utilization ratio [15], making sure that no single node is overwhelmed, hence improving the overall performance of the system. Proper load balancing can help in utilizing the available resources optimally, thereby minimizing the resource consumption. It also helps in implementing fail-over, enabling scalability, avoiding bottlenecks and over-provisioning, reducing response time etc.

Apart from the above-mentioned factors, load balancing is also required to achieve Green computing in clouds which can be done with the help of the following two factors:

• Reducing Energy Consumption - Load balancing helps in avoiding overheating by balancing the workload across all the nodes of a cloud, hence reducing the amount of energy consumed.

• Reducing Carbon Emission - Energy consumption and carbon emission go hand in hand. The more the energy consumed, higher is the carbon footprint. As the energy consumption is reduced with the help of Load balancing, so is the carbon emission helping in achieving Green computing.

v. EXISTING LOAD BALANCING TECHNIQUES

Brief reviews of few existing load balancing algorithms are presented in the following:

Scheduling strategy on load balancing of virtual machine resources: - J. Hu et al. [14] proposed a scheduling strategy on load balancing of VM resources that uses historical data and current state of the system. This strategy achieves the best load balancing and reduced dynamic migration by using a genetic algorithm. It helps in resolving the issue of load-imbalance and high cost of migration thus achieving better resource utilization.

Central load balancing policy for virtual machines - A. Bhadani et al. [15] proposed a Central Load Balancing Policy for Virtual Machines (CLBVM) that balances the load evenly in a distributed virtual machine/cloud computing environment. This policy improves the overall performance of the system but does not consider the systems that are fault-tolerant.

LBVS: Load Balancing strategy for Virtual Storage :- H. Liu et al. [16] proposed a load balancing virtual storage strategy (LBVS) that provides a large scale net data storage model and Storage as a Service model based on Cloud Storage. Storage virtualization is achieved using an architecture that is three-layered and load balancing is achieved using two load balancing modules. It helps in improving the efficiency of concurrent access by using replica balancing further reducing the response time and enhancing the capacity of disaster recovery. This strategy also helps in improving the use rate of storage resource, flexibility and robustness of the system.

Decentralized content aware load balancing: - H. Mehta et al. [17] proposed a new content aware load balancing policy named as workload and client aware policy (WCAP). It uses a unique and special property (USP) to specify the unique and special property of the requests as well as computing nodes. USP helps the scheduler to decide the best suitable node for the processing the requests. This strategy is implemented in a decentralized manner with low overhead. By using the content information to narrow down the search, this technique improves the searching performance and hence overall performance of the system. It also helps in

reducing the idle time of the computing nodes hence improving their utilization.

Server-based load balancing for Internet distributed services - A. M. Nakai et al. [18] proposed a new serverbased load balancing policy for web servers which are distributed all over the world. It helps in reducing the service response times by using a protocol that limits the redirection of requests to the closest remote servers without overloading them. A middleware is described to implement this protocol. It also uses a heuristic to help web servers to endure overloads.

Join-Idle-Queue - Y. Lua et al. [19] proposed a Join- Idle-Queue load balancing algorithm for dynamically scalable web services. This algorithm provides largescale largescale load balancing with distributed dispatchers by, first load balancing idle processors across dispatchers for the availability of idle processors at each dispatcher and then, assigning jobs to processors to reduce average queue length at each processor. By removing the load balancing work from the critical path of request processing, it effectively reduces the system load, incurs no communication overhead at job arrivals and does not increase actual response time.

A Lock-free multiprocessing solution for LB - X. Liu et al. [20] proposed a lock-free multiprocessing load balancing solution that avoids the use of shared memory in contrast to other multiprocessing load balancing solutions which use shared memory and lock to maintain a user session. It is achieved by modifying Linux kernel. This solution helps in improving the overall performance of load balancer in a multi-core environment by running multiple load-balancing processes in one load balancer.

A Task Scheduling Algorithm Based on Load Balancing - Y. Fang et al. [21] discussed a two-level task scheduling mechanism based on load balancing to meet dynamic requirements of users and obtain a high resource utilization. It achieves load balancing by first mapping tasks to virtual machines and then virtual machines to host resources thereby improving the task response time, resource utilization and overall performance of the cloud computing environment.

Honeybee Foraging Behavior - M. Randles et al. [22] investigated a decentralized honeybee-based load balancing technique that is a nature-inspired algorithm for self-organization. It achieves global load balancing through local server actions. Performance of the system is enhanced with increased system diversity but throughput is not increased with an increase in system size. It is best suited for the conditions where the diverse population of service types is required.

ACCLB (Load Balancing mechanism based on ant colony and complex network theory) - Z. Zhang et al. [23] proposed a load balancing mechanism based on ant colony and complex network theory in an open cloud computing federation. It uses small-world and scale-free characteristics of a complex network to achieve better load balancing. This technique overcomes heterogeneity, is adaptive to dynamic environments, is excellent in fault tolerance and has good scalability hence helps in improving the performance of the system.

Particle Swarm Optimization :- Wu et. al [24] has experimented with a set of workflow applications by varying their data communication costs and computation costs according to a cloud price model. First, the algorithm starts with swarm initialization using greedy randomized adaptive search procedure to guarantee each particle in the initial swarm is a feasible and efficient solution. Then, compute the potential exemplars, pbest and gbest, for particles to learn from while they are moving. The stop condition is considered as the user‟s QoS requirements, such as deadline, the budget for computation cost or data transfer cost. The particle‟s new position generation procedure has three steps: 1) select elements from the promising set of pairs with larger probability, that is, the particle learns from gbest and pbest; 2) due to the discrete property of scheduling, there are usually not enough feasible pairs in gbest to generate new position, so the particle will learn from its previous position; 3) all the unmapped tasks should choose resources from other feasible pairs. Finally, gbest will be return as optimal solution. The authors have also compared the total computation cost optimization ratio by varying the tasks number. The result shows that when the task number of the workflow becomes large, their technique optimization ratio increases relatively dramatic. It means the technique can actually achieve lower cost for executing the workflow. Experimental results show that the proposed algorithm can achieve much more cost savings and better performance on makes pan and cost optimization. Result could be better if SLA was considered. The goal of this study was to determine whether the literature on load balancing techniques in cloud computing provides a uniform and rigorous base. The papers were initially obtained in a broad search in four databases covering relevant journals, conference and workshop proceedings. Then an extensive systematic selection process was carried out to identify papers describing load balancing techniques in cloud computing. The results presented here thus give a good picture of the existing load balancing techniques in cloud computing.

Two-phase load balancing algorithm (OLB + LBMM) - S.-C. Wang et al. [25] proposed a two- phase scheduling algorithm that combines OLB (Opportunistic Load Balancing) and LBMM (Load Balance Min-Min) scheduling

algorithms to utilize better executing efficiency and maintain the load balancing of the system. OLB scheduling algorithm, keeps every node in working state to achieve the goal of load balance and LBMM scheduling algorithm is utilized to minimize the execution time of each task on the node thereby minimizing the overall completion time. This combined approach hence helps in an efficient utilization of resources and enhances the work efficiency.

Event-driven - V. Nae et al. [20] presented an eventdriven load balancing algorithm for real-time Massively Multiplayer Online Games (MMOG). This algorithm after receiving capacity events as input, analyzes its components in context of the resources and the global state of the game session, thereby generating the game session load balancing actions. It is capable of scaling up and down a game session on multiple resources according to the variable user load but has occasional QoS breaches.

VectorDot - A. Singh et al. [29] proposed a novel load balancing algorithm called VectorDot. It handles the hierarchical complexity of the data-center and multidimensionality of resource loads across servers, network switches, and storage in an agile data center that has integrated server and storage virtualization technologies. VectorDot uses dot product to distinguish nodes based on the item requirements and helps in removing overloads on servers, switches and storage nodes.

vi. CONCLUSION

Existing Load Balancing techniques that have been studied, mainly focus on reducing overhead, service response time and improving performance etc., but none of the techniques has considered the energy consumption and carbon emission factors. Therefore, there is a need to develop an Energy-efficient load balancing technique that can improve the performance of cloud computing workload across all the nodes of a Cloud, hence reducing energy consumption. along with maximum resource utilization, in turn reducing energy consumption as well as carbon emission to an extent that will help achieve Green Computing. Data security includes the specific controls and technologies used to enforce information governance. The Cloud Computing provider must assure the data owner that they provide full disclosure (aka ‘transparency’) regarding security practices and procedures as stated in their SLAs. Data security lifecycle is an essential part of cloud computing governance, yet its emphasizing different elements such as location of data, access of data and control of data and mapping of lifecycle and Function vs. Controls.

REFERENCES

[1] NIST (National Institute of Standards and Technology). http://csrc.nist.gov/groups/SNS/cloud-computing/

[2] Wikipedia, the free encyclopedia

[3] Anthony T.Velte, Toby J.Velte, Robert Elsenpeter, Cloud Computing A Practical Approach, TATA McGRAW-HILL Edition 2010.

[4] Michael Armbrust, Armando Fox, Gunho Lee, Ion Stoica(2009) “Above the Clouds :A Berkeley View of Cloud Computing” University of California at Berkeley Technical Report No. UCB/EECS- 2009-28

[5] Bhaskar Prasad Rimal ,Eunm Choi, Ian Lumb, “A Taxonomy and Survey of Cloud Computing System” 2009 Fifth International Joint Conference on INC, IMS and IDC 978-0-7695-3769-6/09 $26.00 © 2009 IEEE

[6] Soumya Ray and Ajanta De Sarkar “EXECUTION ANALYSIS OF LOAD BALANCING ALGORITHMS IN CLOUD COMPUTING ENVIRONMENT” International Journal on Cloud Computing: Services and Architecture (IJCCSA),Vol.2, No.5, October 2012

[7] Suriya Begum, Dr. Prashanth C.S.R “Review of load balancing in cloud computing” IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013 ISSN (Print): 1694-0784 |

[8]] Rajkumar Buyya, Chee Shin Yeo, and Srikumar Venugopal, Market- Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities, HPCC 2008: 5-13

[9] Amazon web service, [Online]. Available: http://aws.amazon.com/

[10] Kuyoro S. O., Ibikunle F., Awodele O., Cloud Computing SecurityIssues and Challenges, International Journal of Computer Networks(IJCN), Volume (3) : Issue (5) : 2011

[11] Southern African Internet Governance Forum Issue Papers1 No. 1 of 5 Emerging Issues: Cloud Computing

[12] Jaspreet kaur, Comparison of load balancing algorithms in a Cloud, International Journal of Engineering Research and Applications, Vol. 2,Issue 3, May-Jun 2012, pp.1169-1173

[13] Raksha Nawal, Aashish Gupta, Pragya Nagar, Rajesh Gurjar,Emergence, Performance And Issues Challenging Cloud Computing, Proceedings of the NCNTE-2012, Third Biennial National Conference on Nascent Technologies, 2012

[14] J. Hu, J. Gu, G. Sun, and T. Zhao, “A Scheduling Strategy on Load Balancing of Virtual Machine Resources in Cloud Computing Environment”, Third International Symposium on Parallel Architectures, Algorithms and Programming (PAAP), 2010, pages 89-96.

[15] Bhadani, and S. Chaudhary, “Performance evaluation of web servers using central load balancing policy over virtual machines on cloud”, Proceedings of the Third Annual ACM Bangalore Conference (COMPUTE), January 2010.

[16] H. Liu, S. Liu, X. Meng, C. Yang, and Y. Zhang, “LBVS: A Load Balancing Strategy for Virtual Storage”, International Conference on Service Sciences (ICSS), IEEE, 2010, pages 257-262.

[17] H. Mehta, P. Kanungo, and M. Chandwani, “Decentralized content aware load balancing algorithm for distributed computing environments”, Proceedings of the International Conference Workshop on Emerging Trends in Technology (ICWET), February 2011, pages 370-375.

[18] A. M. Nakai, E. Madeira, and L. E. Buzato, “Load Balancing for Internet Distributed Services Using Limited Redirection

Rates", 5th IEEE Latin-American Symposium on Dependable Computing (LADC), 2011, pages 156-165.

[19] Y. Lua, Q. Xiea, G. Kliotb, A. Gellerb, J. R. Larusb, and A. Greenber, "Join-Idle-Queue: A novel load balancing algorithm for dynamically scalable web services", An international Journal on Performance evaluation, In Press, Accepted Manuscript, Available online 3 August 2011.

[20] Xi. Liu, Lei. Pan, Chong-Jun. Wang, and Jun-Yuan. Xie, "A Lock-Free Solution for Load Balancing in Multi-Core Environment", 3rd IEEE International Workshop on Intelligent Systems and Applications (ISA), 2011, pages 1-4.

[21] Y. Fang, F. Wang, and J. Ge, "A Task Scheduling Algorithm Based on Load Balancing in Cloud Computing", Web Information Systems and Mining, Lecture Notes in Computer Science, Vol. 6318, 2010, pages 271-277.

[22] M. Randles, D. Lamb, and A. Taleb-Bendiab, "A Comparative Study into Distributed Load Balancing Algorithms for Cloud Computing", Proceedings of 24th IEEE International Conference on Advanced Information Networking and Applications Workshops, Perth, Australia,April 2010, pages 551-556.

[23] Z. Zhang, and X. Zhang, "A Load Balancing Mechanism Based on Ant Colony and Complex Network Theory in Open Cloud Computing Federation", Proceedings of 2nd International Conference on Industrial Mechatronics and Automation (ICIMA), Wuhan, China, May 2010, pages 240- 243.

[24] Zhangjun Wu, Zhiwei Ni, Lichuan Gu, Xiao Liu, A Revised Discrete Particle Swarm Optimization for Cloud Workflow Scheduling, IEEE 2010

[25] S. Wang, K. Yan, W. Liao, and S. Wang, "Towards a Load Balancing in a Three-level Cloud Computing Network", Proceedings of the 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), Chengdu, China, September 2010, pages 108-113.

[26] V. Nae, R. Prodan, and T. Fahringer, "Cost-Efficient Hosting and Load Balancing of Massively Multiplayer Online Games", Proceedings of the 11th IEEE/ACM International Conference on Grid Computing (Grid), IEEE Computer Society, October 2010, pages 9-17.

[27] Singh, M. Korupolu, and D. Mohapatra, "Server-storage virtualization: integration and load balancing in data centers", Proceedings of the ACM/IEEE conference on Supercomputing (SC), November 2008.

[28] B. P. Rima, E. Choi, and I. Lumb, "A Taxonomy and Survey of Cloud Computing Systems", Proceedings of 5th IEEE International Joint Conference on INC, IMS and IDC, Seoul, Korea, August 2009, pages 44-51.

[29] R. Mata-Toledo, and P. Gupta, "Green data center: how green can we perform", Journal of Technology Research, Academic and Business Research Institute, Vol. 2, No. 1, May 2010, pages 1-8.

[30] Jing Deng Scott C.-H. Huang Yunghsiang S. Han and Julia H. Deng, Fault-Tolerant and Reliable Computation in Cloud Computing, GLOBECOM Workshops (GC Wkshps), 2010 IEEE