IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 15, Issue 2 (Nov. - Dec. 2013), PP 22-29
www.iosrjournals.org

Load Balancing In Cloud Computing:A Review

Shiny
[1](*M.Tech, Central University of Punjab,Bathinda, Punjab, India)*

Abstract: *As the IT industry is growing day by day, the need of computing and storage is increasing rapidly. The amount of data exchanged over the network is constantly increasing. Thus the process of this increasing mass of data requires more computer equipment to meet the various needs of the organizations. To better capitalize their investment, the over-equipped organizations open their infrastructures to others by exploiting the Internet and other important technologies such as virtualization by creating a new computing model: the cloud computing. Cloud computing is one of the significant milestones in recent times in the history of computers. The basic concept of cloud computing is to provide a platform for sharing of resources which includes software and infrastructure with the help of virtualization. This paper presents a brief review of cloud computing. The main emphasize of this paper is on the load balancing technique in cloud computing.*

Keywords: *Cloud Computing, Load Balancing, Dynamic Load Balancing, Virtualization, Data Center.*

I. Introduction

Cloud computing get its name as a metaphor for the Internet. Typically, the Internet is represented in network diagrams as a cloud [1]. The term "cloud"[2]originates from the world of telecommunications when providers began using virtual private network (VPN) services for data communications. Cloud computing simply means Internet Computing, generally the internet is seen as collection of clouds; thus the world cloud computing is defined as utilizing the internet to provide technology enabled services to the people and organizations[3].According to NIST(National Institute of Standards and Technology), cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources i.e. network servers, storage applications and services)[2].

The evolution of cloud computing over the past few years is one of the major advances in the history of computing. Cloud computing is a recent trend in IT that moves computing and data away from desktop and portable PC's into large data centers.The main advantage of cloud computing is that customer do not have to pay for infrastructures, its installation and required man power to handle such infrastructure and maintenance [2]. Cloud computing is independent computing it is totally different from grid and utility computing. Cloud computing is cheaper than other computing models;zero maintenance cost is involved since the service provider is responsible for the availability of services and clients are free from maintenance and management problems of resource machines. Due to this feature, cloud computing is also known as utility computing or "IT on demand"[3].In contrast to previous paradigms (clusters and grid computing), cloud computing is not application-oriented but service-oriented; it offers on-demand virtualized resources as measurable and billable utilities[4].

Cloud computing is a distributed computing paradigm that focuses on providing a wide range of users with distributed access to scalable, virtualized hardware and/or software infrastructure over the internet. It involves virtualization, distributed computing, networking, software and web services.The concept of cloud computing has significantly changed the field of parallel and distributed computing systems today[5].It has emerged as apopular solution to provide cheap and easy access to externalized IT resources [4]. Cloud computing deals with virtualization, scalability, interoperability, quality of service and the delivery models of the cloud, namely public, private and hybrid[2].Through virtualization, cloud computing is able to address with the same physical infrastructure a large client base with different computational needs[4].The rapid growth in the field of cloud computing also increases severe security concerns. Lack of security is the only hurdle in wide adoption of cloud computing[3].

The main objective of this paper is to give an outline of cloud computing. Section 2 demonstrates the architecture of cloud computing. The advantages and disadvantages of the cloud are discussed in Section 3. Section 4 and 5 explains the basic concept of load balancing and dynamic load balancing respectively and Section 6 concludes the paper.

II. Architecture Of Cloud Computing

Cloud computing architecture refers to the components required for cloud computing. Cloud computing typically involves multiple cloud components communicating with each other over a loose

coupling mechanism such as messaging queue [6]. Cloud computing architecture can be divided into two sections: front end and back end. They both are connected with each other through a network, usually the internet. The front end is what the user (client) sees whereas the back end is the cloud of the system. Front end has client's computer and the application required to access the cloud and the back end has the cloud computing services like various computers, servers and data storage [2].

2.1. Cloud Computing Services

The term services in cloud computing is the concept of being able to use reusable, fine-grained components across a vendor's network[7].Cloud computing services can be broadly classified into three categories:SaaS (Software-as-a-Service), in which provides software applications and programs like documents editing services, messaging services, etc. to users; PaaS (Platform-as-a-Service), in which it provides platforms like operating systems or software development environment to programmers and users for building and running your programs; and IaaS (Infrastructure as a Service), in which infrastructure such as hardware and storage, are provided to users as a service [14].

2.1.1. Software as a Service (SaaS)

SaaS, sometimes referred to as "software on demand". SaaS constitutes a mojor role in all th cloud computing offerings [14].Software as a Service (SaaS) is the model in which an application is hosted as a service to customers who access it via the Internet[1]. SaaS is software that is owned, delivered and managed remotely by one or more providers and is offered in a pay-as-per-use manner[4]. SaaS focuses on providing users with business specific capabilities such as e-mail or customer management[5]. The typical user of SaaS offering usually has neither knowledge nor control about the underlying infrastructure[4]. One of the example of SaaS provider is Google Apps that provides large suite of web based applications for many business applications including accounting, enterprise resource management (ERP), human resource management (HRM), customer relationship management (CRM) and security device manager (SDM).

2.1.2. Platform as a Service (PaaS)

PaaS is a service model cloud computing. In this model, client creates the software using tools and libraries from the provider[5]. The client controls the applications that run in the environment, but does not control the operating system, hardware and network infrastructure on which they are running[4]. The provider provides the network, servers and storage. One of the example of PaaS is Google App Engine that provides clients to run their applications on Google's infrastructure [5].PaaS services include application design, development, testing, deployment and hosting. Other services include team collaboration, web service integration, database integration, security, scalability, storage, state management and versioning.PaaS also supports web development interfaces such as Simple Object Access Protocol (SOAP) and Representational State Transfer (REST), which allows the construction of multiple web services, sometimes called mashups. A downfall to PaaS is a lack of interoperability and portability among providers [1].

2.1.3. Infrastructure as a Service (IaaS)

IaaS, also known as cloud infrastructure services, delivers computer infrastructure-typically a platform virtualization outsourced service. IaaS model provides a virtual data center within the cloud. IaaS provides servers (physical and virtualized), cloud-based data storage, etc[8].The client need not purchase the required servers, data center or the network resources. The key advantage is that customers need to pay only for the time duration they use the service[2]. 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 EC2 infrastructure[5].

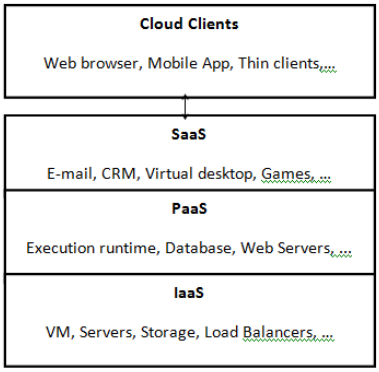

Fig. 1. Architecture of Cloud Computing

2.2. Components of Cloud

Cloud computing is the dynamic provisioning of information technology capabilities (hardware, software or services) from third parties over a network [5]. One of the main reasons cloud computing becomes popular is due to the adoption of businesses as the easier way to implement business processes [9].A cloud computing model generally consists of three major components: clients, datacenter and distributed servers.

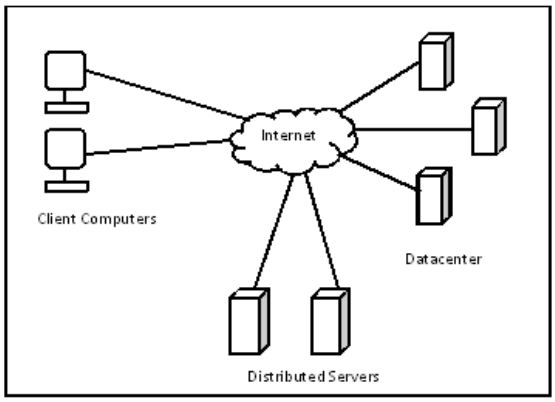

Fig. 2. Cloud Components

2.2.1. Clients

Clients are the devices that the end users interact with to manage their information on the cloud. Clients generally fall into three categories[1, 5]:

- **Mobile:** Mobile devices include PDAs or smartphones, like a Blackberry, Windows Mobile Smartphone or an iPhone.
- **Thin:** Clients are computers that do not have internal hard drives, but rather let the servers do all the work, but then display the information.
- **Thick:** This type of client is a regular computer, using a web browser like Firefox, Google Chrome or Internet Explorer to connect the different cloud. Following are the some benefits of thin clients:
    - › Lower hardware costs
    - › Lower IT costs
    - › Data Security
    - › Less power consumption
    - › Less noise
    - › Ease of repair or replacement

2.2.2. Datacenter

The datacenter is nothing but the collection of servers hosting different applications. It could be a large room in the basement of your building or a room full of servers on the other side of the world that you access via Internet[1].An end user connects to the datacenter to subscribe different applications [5].

2.2.3. Distributed Servers

A server, which actively checks the services of their hosts, is known as distributed server. Distributed servers are the part of a cloud which is available throughout the internet hosting different

applications. But while using the application from the cloud, the user would feel that he/she is using this application from its own machine [5].

2.3. Types of Cloud

On the basis of accessibility, clouds can be deployed using any of the following strategies:

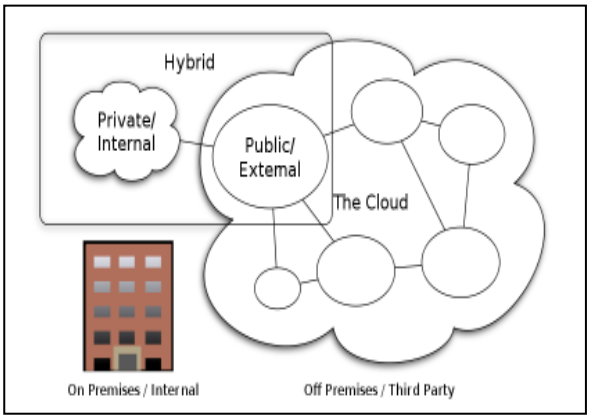

Fig. 3. Types of Cloud [16]

2.3.1. Public Cloud

Public clouds are made available to the general public by a service provider who hosts the cloud infrastructure. Generally, public cloud providers like Amazon AWS, Microsoft and Google own and operate the infrastructure and offer access over the Internet [10].Users need to pay only for the time duration they use the service i.e. pay-per-use [2]. A public cloud does not mean that a user's data is publically visible; public cloud vendors typically provide an access control mechanism for their users [5]. Clients do not need to purchase hardware to get service and can also scale their user on demand[6]. Public clouds provide an elastic, cost effective means to deploy solutions.

2.3.2. Private Cloud

Private cloud means using a cloud infrastructure solely by one customer or organization [11]. In a private cloud-based service, resources are deployed inside a firewall and managed by client's organization without the restriction of network bandwidth and security exposures [5]. Organization owns the hardware and software infrastructure, manages cloud and controls access to its resources [6]. The main advantage of private cloud is, it is easier to manage security, maintenance and upgrades and also provides more control over the deployment and use. As compared to public clouds, where all resources and applications were managed by the service provider, in private cloud these services are pooled together and made available for users at the organizational level [2].

2.3.3. Community Cloud

A community cloud is a cloud service model that provides a cloud computing solution to a limited number of individuals or organizations that is governed, managed and secured commonly by all the participating organizations or a third party managed service provider. It is multi-tenant infrastructure that is controlled and used by a group of organizations that have shared interests such as specific security requirements. Community clouds are designed for business and organizations working on joint projects, applications or research which requires a central cloud computing facility for building, managing and executing such projects[12].

2.3.4. Hybrid Cloud

A hybrid cloud is a cloud computing environment in which an organization provides and manages some in-house resources and has others provided externally [15]. Hybrid cloud is a combination of private and public cloud. In this, a private cloud is linked to one or more external cloud services. The goal is to combine services and data from a variety of cloud models to create a unified automated and well-managed computing environment [13].

III. Advantages And Disadvantages

Cloud computing is a disruptive technology that is changing the way the enterprises look to meet their IT hardware and software requirements. Cloud computing is a mix of the latest ideas, technology and delivery models including IaaS, PaaS and SaaS[17].

3.1. Advantages

Following are the advantages offered by cloud computing:

3.1.1. Flexibility

The major benefit of cloud computing is that there is no limitation of place and medium. One can access his/her applications and data anywhere in the world, on any system [18].But the thing is you need a computer or laptop or smartphone or Android or Blackberry and other applicable devices with internet connections [19].

3.1.2. Low Cost

Cloud computing is Cap-Ex free means capital expenditure free. There is no need to spend big money on hardware, software and licensing fees[20]. The cost of using cloud resources is very economical for resources such as centralized, heal estate, bandwidth and power. Users will also save money on software updates, management costs and data storage costs [18]. Everything is stetted up in host, which automatically saves the time and money for any organizations [19].

3.1.3. Cloud is environment friendly

In general, the cloud is more efficient than the typical IT infrastructure normally scales down, freeing up resources and consuming less power. At any moment, only the resources that are truly needed are consumed by the system [21].

3.1.4. Scalability

Scalability is a built-in feature of cloud deployments[21]. Enterprises no longer have to invest time in buying and setting up the hardware, software and other resources necessary for a new application. They can quickly scale up or scale down their usage of services on the cloud as per market demands, during hours of maximum activity, while launching sales, campaigns, etc.[17].

3.1.5. More storage capacity

Storage capacity is also one of the main benefits of cloud computing, as it can store more data as compared to a personal computer. It eliminates worries about running out of storage space [21]. Everything is online, store your entire data in cloud and can access at any time in browser [19].

3.2. Disadvantages

3.2.1. Security and Privacy

Cloud computing means that all of your stuff is on the web. Security and privacy are the biggest concerns about cloud computing which hamper the growth of cloud. Security issues such as data loss, phishing, andbotnet pose serious threats to organizations data and software.Moreover, the multi-tenancymodel and the pooled computing resources in cloud computing has introduced new security challenges that require novel techniques to tackle with [22].

Privacy is another big issue with cloud computing server. To make cloud servers more secure to ensure that a client's data is not accessed by any unauthorized users, cloud service providers have developed password protected accounts, security servers through which all data being transferred must pass and data encryption technique[18].

3.2.2. Migration

Migration problem is also a big issue about cloud computing. If you want to move from one cloud to another i.e. from one hosting provider to another, have to face more problems. It's not easy to move to another hosting provider because of migration process will take time to transfer files, which indirectly your business in offline for some time/days [19].

3.2.3. Requires a constant internet connection

This is the panic situation for business owner, when site goes offline for some time [19]. It makes your business dependent on the reliability of your Internet connection. If you do not have an Internet connection, you cannot access anything even your on data. A dead Internet connection means no work.

Web based apps often require a lot of bandwidth to download [18]. Even Amazon, Google and Apple websites faced this problem. There are two ways to mitigate this risk[23]:

- Make sure that you are using an enterprise class Internet connection: Enterprise resource connections are more expensive but provide much better fault tolerance and repair service that consumer-class connections do.
- Provide redundant connections if you can: If one connection fails, traffic can be rerouted through alternative connections.

IV. Load Balancing

As cloud computing is growing rapidly and clients are demanding more results and better services, load balancing for cloud has become a very important and interesting concept. Load balancing is a new technique that facilitates networks and resources by providing a maximum throughput with minimum response time [25]. It is a generic term used for distributing the workload across one or more servers, network interfaces, hard drives or other computing resources to enhance both resource utilization and job response time. The load can be CPU load, memory capacity, delay or network load. Load balancing ensures that all the processors in the system or every node in the network does approximately the equal amount of work at any instant of time. While balancing the load, certain types of information such as number of jobs waiting in queue, job arrival rate, CPU processing rate and so forth at each processors, may be exchanged among the processors for improving the overall performance [24].

Load balancers are used for assigning load to different virtual machines in such a way that none of the nodes gets loaded heavily or lightly. The load balancing needs to be done properly because failure in any one node can lead to unavailability of data [27]. The load balancer accepts multiple requests from the client and distributing each of them across multiple computers or network devices based on how busy the or network device is. Load balancing helps to prevent a server or network device from getting vanquish with requests and also helps in distribution of work. If the load balancer is not available the client can wait for long time and process their request in the particular server only where the client give request [26].

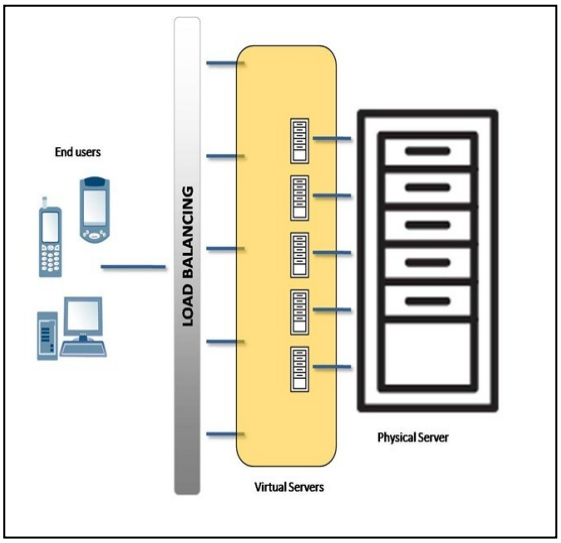

Fig.4. Load Balancer [28]

4.1. Goals of load balancing

The main aim of load balancing is as follows:

- To maintain the firmness of the system
- To significantly improve the performance of the system
- To have a backup plan in case of any failure
- To encompass the future modification of the system

4.2. Types of Load Balancing Algorithms

The load balancing algorithms are classified into two categories: static and dynamic load balancing.

4.2.1. Static Load Balancing Algorithm

In static load balancing algorithms, the performance of the processors is determined at the beginning of the execution, it does not depend on current state of the system. The goal of static load balancing is to reduce the overall execution time of a synchronous program while minimizing the communication delays. These algorithms are mostly suitable for homogeneous and stable environments and can produce very good results [30].Some of the examples of static load balancing algorithms are: Round Robin algorithm, Randomized algorithm and Threshold algorithm.

4.2.2. Dynamic Load Balancing Algorithm

In dynamic load balancing algorithm, the decisions on load balancing are based on the current state of the system; no prior knowledge is needed [30]. The main advantage of dynamic load balancing is that if any node fails, it will not stop the system; it will only affect the performance of the system. These algorithms are more flexible than static algorithms, can easily adapt to changes and provide better results in heterogeneous and dynamic environments[30].Dynamic load balancer uses policies for keeping the track of updated information. There are four policies for dynamic load balancers: transfer policy, selection policy, location policy and information policy[27].

V. Dynamic Load Balancing

A load balancing algorithm which is dynamic in nature does not consider the previous state or behavior of the system; it depends upon the current behavior of the system. The important things to examine while developing such algorithms are: load estimation, comparison of load, stability of different system, performance of the system, interaction between the nodes, nature of work to be transferred, selection of nodes, etc. [26]. The task of load balancing is shared among distributed nodes. In a distributed system, dynamic load balancing can be done in two different ways: distributed and non-distributed.

5.1. Distributed Dynamic Load Balancing Algorithm

In the distributed one, the dynamic load balancing algorithm is executed by all nodes present in the system and the task of scheduling is shared among them[27]. The interaction among the nodes to achieve load balancing can take two forms: cooperative and non-cooperative. In the cooperative, the nodes work together to achieve the common objective, for example to improve the response time, etc. whereas in non-cooperative, each node works independently towards a goal local to it, for example to improve the response time of a local task, etc.[28], [29]. In dynamic load balancing systems, the nodes can interact with each other generating more messages as compared to non-distributed ones because each node in the system needs to interact with every other node[27].

5.2. Non-Distributed Load Balancing Algorithm

In the non-distributed or undistributed, the nodes work individually in order to achieve a common goal. Non-distributed dynamic load balancing algorithms are further classified into two: centralized and semi-centralized.

5.2.1. Semi-distributed Dynamic Load Balancing

In semi-distributed dynamic load balancing, the nodes of the system are partitioned into clusters, where the load balancing in each cluster is of centralized form. A central node is elected in each cluster by appropriate election technique which takes care of load balancing within that cluster. Therefore, the load balancing of whole system is done via the central nodes of each cluster[29].

5.2.2. Centralized Dynamic Load Balancing

In centralized dynamic load balancing, the algorithm is executed only by a single node in the whole system i.e. central node. This node is completely responsible for load balancing of the whole system and rest of the nodes interacts only with the central node[29].

VI. Conclusion

Cloud computing is an emerging field of information technology (IT). It enables a wide range of users to access distributed, scalable, virtualized, hardware and/or software infrastructure over the Internet.Load balancing is one of the leading issue of cloud computing. So there is need for a well ordered load balancing algorithm for efficient utilization of resources.

This paper is an earnest effort to unveil the concept of load balancing and its type's, especially dynamic load balancing. But, still there are miles to go. As cloud computing is yet in its infancy and there are many more open issues like security, resource utilization, etc. that need to be explored.

References

[1] A.T. Velte, T.J. Veltey and R. Elsenpeter, *Cloud Computing: A Practical Approach* (Tata McGraw-Hill Education Private Limited, New Delhi, Edition 2010).

[2] Y. Jadeja and K. Modi, "Cloud Computing-Concepts, Architecture and Challenges", *International Conference on Computing, Electronics and Electrical Technologies*, 2012, 877-880.

[3] F.B. Shaikh and S. Haider, "Security Threats in Cloud Computing", *Internet Technology and Secured Transactions*,2011, 214-219.

[4] J. Srinivas, K.V.S.Reddy and A.M. Qyser, "Cloud Computing Basics", *International Journal of Advanced Research in Computer and Communication Engineering, 1(5)*, July 2012.

[5] S. Ray and A.D. Sarkar, "Execution Analysis of Load Balancing Algorithms in Cloud Computing Environment",*International Journal on Cloud Computing Services and Architecture*, *2(5)*, October 2012.

[6] e2networks.com/cloud-servers-india/cloud-computing-architecture/

[7] D.K. Kumar, G.V. Rao and G.S. Rao, "Cloud Computing: An Analysis of Its Challenges and Security Issues", *International Journal of Computer Science and Network, 1(5)*, October 2012.

[8] K. Jamsa, *Cloud Computing* (Jones & Bartlett Learning, 5 Wall Street, Burlington, USA, 2013).

[9] http://cloud.siliconindia.com/forum/Cloud-Computing-Basic-Components-qid-7970-catid-39.html

[10] http://blog.appcore.com/blog/bid/167543/Types-of-Cloud-Computing-Private-Public-and-Hybrid-Clouds

[11] http://www.globaldots.com/cloud-computing-types-of-cloud/

[12] http://www.techopedia.com/definition/26559/community-cloud

[13] KumarVikas and PradhanPrasann, "Role of Service Level Agreements in SaaS Business Scenario", *The IUP Journal of Information Technology, 9(1)*, March 2013.

[14] www.dummies.com/how-to/content/what-is-hybrid-cloud-computing.html

[15] http://searchcloudcomputing.techtarget.com/definition/hybrid-cloud

[16] http://2.bp.blogspot.com/_OhIk3F9lWZE/TNF1-eK65dI/AAAAAAAAAAo/UDHCtGyC4rc/s400/800px-Cloud_computing_types.svg.png

[17] http://www.financialforce.com/resources/research/salesforce-platform/cloud-computing-advantages/

[18] www.techinmind.com/waht-is-cloud-computing-what-are-its-advantages-and-disadvantages/#

[19] cloudcomputingadvices.com/cloud-computing-advantages-disadvantages/

[20] www.verio.com/resource-center/articles/cloud-computing-benefits/

[21] www.javacodegeeks.com/2013/04/advantages-and-disadvantages-of-cloud-computing-cloud-computing-pros-and-cons.html

[22] Kuyoro S.O., Ibikunle F. and Awodele O., "Cloud Computing Security Issues and Challenges", *International Journal of Computer Networks (IJCN), 3(5)*, 2011.

[23] www.dummies.com/how-to/content/disadvantages-of-cloud-computing-for-networks.html

[24] N. Sran and N. Kaur, "Comparative Analysis of Existing Load Balancing Techniques in Cloud Computing", *International Journal of Engineering Science Invention, 2(1)*, January 2013.

[25] Z. Chaczko, V. Mahadevan, S. Aslanzadeh and C. Mcdermid, "Availability and Load Balancing in Cloud Computing", *International Conference on Computer and Software Modeling*, *14*, 2011.

[26] A. Roy and D. Dutta, "Dynamic Load Balancing: Improve Efficiency in Cloud Computing", *International Journal of Emerging Research in Management & Technology*, 2(4), April 2013.

[27] S.S. Moharana, R.D. Ramesh and D. Powar, "Analysis of Load Balancers in Cloud Computing", *International Journal of Computer Science and Engineering*, *2(2)*, May 2013.

[28] https:// devcentral.f5. com/ weblogs /images/ devcentral_f5_com/weblogs/macvittie/WindowsLiveWriter/loadbalancingisakeycomponenttobuildingcl_ 3A92/load- balancing_2.jpg

[29] Y.R. Kumar, M.M. Priya and K.S. Chatrapati, "Effective Distributed Dynamic Load Balancing For The Clouds", *International Journal of Engineering Research & Technology, 2(2)*, February 2013.

[30] K. Al Nuaimi, N. Mohamed, M. Al Nuaimi and J. Al-Jarrodi, "A Survey of Load Balancing in Cloud Computing: Challenges and Algorithms", *Network Cloud Computing and Applications*, 2012, 137-142.