INTERNATIONAL JOURNAL OF TECHNOLOGY AND COMPUTING (IJTC)
ISSN-2455-099X,
Volume 2, Issue 4 April 2016.

A NEW ERA TO BALANCE THE LOAD ON CLOUD USING VECTOR DOT LOAD BALANCING METHOD

Simranjeet Sidhu*, Er. Manish Mittal

ASRA College of Engineering and Technology, Bhawanigarh
simran.sidhu92@yahoo.com*,

A B S T R A C T

Cloud computing has recently become popular due to the maturity of related technologies such as network devices, software applications and hardware capacities. Resources in these systems can be widely distributed and the scale of resources involved can range from several servers to an entire data center. In Cloud Computing Load balancing has a great role. To allocate and balance the load of the resources among the various components and nodes load balancing is required. Load balancing aims to optimize resource use, maximize throughput, minimize response time, and avoid overload of any single resource. In the paper, we have studied and implemented three algorithm using Java Programming and simulate the algorithms on CloudSim. The Main contribution of CloudSim is to provide a holistic software framework for modeling Cloud computing environments and performance testing application services. And, we have proved that our proposed algorithm Vector Dot gives the better results as compared to Active clustering and Join Idle queue. We analyzed the results on the basis of different performance parameters such as Response time, Execution time and Energy consumption.

Keywords: Load balancing, Vector Dot, Active Clustering, Join idle queue, Paas, Iaas, Saas.

I. INTRODUCTION

Cloud computing presents a delivery model for IT services based on internet and provides a scalable service to easily consume over the internet on an as-needed basis. A major feature is cloud service is that data processed on clouds are often outsourced, leading to number of issues related to accountability including the handling of personally identifiable information. Cloud computing, a framework for enabling convenient, and on-demand network access to a shared pool of computing resources, is emerging as a new paradigm of large-scale distributed computing. It has widely been adopted by the industry, though there are many existing issues like Load Balancing, Virtual Machine Migration, Server Consolidation, Energy Management, etc. that are not fully addressed. In case of Cloud computing services can be used from diverse and wide spread resources, rather than remote servers or local machines. There is no standard definition of Cloud computing. Generally it consists of a bunc*h of distributed servers known as masters, providing demanded services and resources to different clients known as clients in a network with scalability and reliability of datacenter. The distributed computers

provide on-demand services. Services may be of software resources (e.g. Software as a Service, SaaS) or physical resources (e.g. Platform as a Service, PaaS) or hardware/infrastructure (e.g. Hardware as a Service, HaaS or Infrastructure as a Service, IaaS). AmazonEC2 (Amazon Elastic Compute Cloud) is an example of cloud computing services [2]. Some research aims to improve resource utilization while others aim to reduce energy consumption. The goals of both are to reduce costs for data centers. Due to the large size of many data centers, the financial savings are substantial.

A. CLOUD COMPUTING APPLICATION ARCHITECTURE

As we know that cloud computing is the shift of computing to a host of hardware infrastructure that is distributed in the cloud. The commodity hardware infrastructure consists of the various low cost data servers that are connected to the system and provide their storage and processing and other computing resources to the application. Cloud computing involves running applications on virtual servers that are allocated on this distributed hardware infrastructure available in the cloud. These virtual servers are made in such a way that the different service level agreements and reliability issues are met. There may be multiple instances of the same virtual server accessing the different parts of the hardware infrastructure available. This is to make sure that there are multiple copies of the applications which are ready to take over on another one's failure. The virtual server distributes the processing between the infrastructure and the computing is done and the result returned.

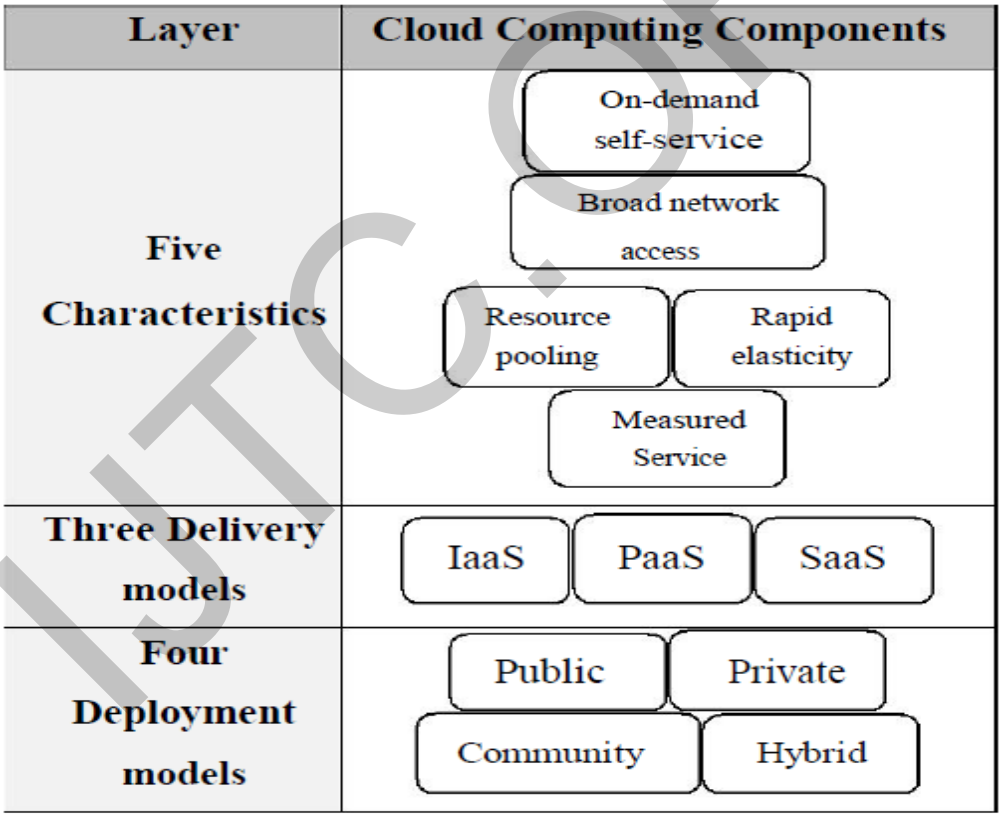

Fig 1. Cloud Computing Application Architecture

There will be a workload distribution management system, also known as the grid engine, for managing the different requests coming to the virtual servers. This engine will take care of the creation of multiple copies and also the preservation of integrity of the data that is stored in the infrastructure. This will also adjust itself such that even on heavier load, the processing is completed as per the requirements. The different workload management systems are hidden from the users. For the user, the processing is done

and the result is obtained. There is no question of where it was done and how it was done. The users are billed based on the usage of the system - as said before - the commodity is now cycles and bytes. The billing is usually on the basis of usage per CPU per hour or GB data transfer per hour.

B. LOAD BALANCING ON CLOUD COMPUTING

Load balancing in large distributed server systems is a complex optimization problem of critical importance in cloud systems and data centers. Load balancing algorithms are classified as static and dynamic algorithms. Static algorithms are mostly suitable for homogeneous and stable environments and can produce very good results in these environments. However, they are usually not flexible and cannot match the dynamic changes to the attributes during the execution time. Dynamic algorithms are more flexible and take into consideration different types of attributes in the system both prior to and during run-time [2]. These algorithms can adapt to changes and provide better results in heterogeneous and dynamic environments. However, as the distribution attributes become more complex and dynamic. As a result some of these algorithms could become inefficient and cause more overhead than necessary resulting in an overall degradation of the services performance.

C. TYPES OF LOAD BALANCING ALGORITHMS

Depending on who initiated the process, load balancing algorithms can be of three catagories as given in [4]:

- Sender Initiated: If the load balancing algorithm is initialized by the sender.
- Receiver Initiated: If the load balancing algorithm is initiated by the receive
- 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 as given in [4]:
- Static: It doesn't depend on the current state of the system. Prior knowledge of the system is needed.
- 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.

D. Goals of Load balancing

As given in [4], the goals of load balancing are :

- To improve the performance substantially
- To have a backup plan in case the system fails even partially
- To maintain the system stability
- To accommodate future modification in the system

II. RELATED WORK

Dr. Vijay Laxmi et al. [1] in 2014 presented a load balancing is essential for efficient operations in distributed environments. As Cloud Computing is growing rapidly and clients are demanding more services and better results, load balancing for the Cloud has become a very interesting and important research area. Load balancing ensures that all the processor in the system or every node in the network does approximately the equal amount of work at any instant of time. Our objective is to implement an

effective load balancing algorithm for balancing the load on cloud. In this thesis, we have studied and implemented three algorithm using Java Programming and simulate the algorithms on CloudSim.

Saeed Javanmardi et al.[2] in 2014 presented a hybrid job scheduling approach, which considers the load balancing of the system and reduces total execution time and execution cost. We try to modify the standard Genetic algorithm and to reduce the iteration of creating population with the aid of fuzzy theory. The main goal of this research is to assign the jobs to the resources with considering the VM MIPS and length of jobs. The new algorithm assigns the jobs to the resources with considering the job length and resources capacities.

Tom Guérout *et al.* [3] in 2014 discussed analysis of studies on Clouds modeling, Clouds scheduling, and actual SLAs of SaaS providers. Based on these analyses, they proposed a Cloud architecture modeling that includes the DVFS, but especially a modeling of Clouds Quality of Service parameters. This list contains definitions, measurable and reusable metrics for non-functional parameters. The defined QoS metrics are measurable and reusable in any scheduling approach for Clouds. The use of these QoS models is done through the performance analysis of three scheduling approaches considering four QoS parameters.

Bernardetta Addis *et al.* [4] in 2014 proposed a new optimization framework for the management of the energy usage in an integrated system for Cloud services that includes both service centers and communication networks for accessing and interconnecting them. The optimization framework considers a PaaS scenario where VMs serving an application can be allocated to a set of SCs geographically distributed and traffic load coming from different world regions can be assigned to VMs in order to optimize the energy cost and minimize CO2 emissions. Numerical results, on a set of randomly generated instances and a case study representative of a large Cloud provider, shows that the availability of green energy have a big impact on optimal energy management policies and that the contribution of the network is far from being negligible.

B.Sujana et al. [7] in 2013 presented a cloud computing presents a delivery model for IT services based on internet and provide a scalable service to easily consume over the internet on an as-needed basis. A major feature is cloud service is that data processed on clouds are often outsourced, leading to number of issues related to accountability including the handling of personally identifiable information. Dealing with "single cloud" providers is predicted to become less popular with customers due to risks of service availability failure and the possibility of malicious insiders in the single cloud.

III. RESEARCH PROBLEM

In a cloud environment, there may be any number of host machines and each host machine has different-different load due to virtual machines as per the client's demand. The load of a host machine may be of various types such as CPU load, Memory load, Storage load and Network related load etc. If the load of any host machine exceeds its capacity then it affects its efficiency. In runtime, any client application service may change their resource (CPU, RAM, Storage and Bandwidth etc.) demand and this causes the host system to be imbalanced. If this imbalanced situation occurs due to overloading then system is balanced using load balancing techniques by distributing the extra workload to the whole clouds host heaving light loads. This helps to improve the overall performance of the cloud system. In the paper, we are proposing the Vector Dot method to balance the load on the cloud and compared it with the existing load balancing methods such as active clustering and join idle queue. The main objective of the research is balance the load on cloud and consumes less energy as compared to previous, on the cloud by using proposed method. Also we have to prove that our proposed

technique is more efficient for load balancing and energy consumption on cloud as compared to previous.

A. Metrics for Load Balancing In Clouds

Various metrics will be considered in load balancing techniques in cloud computing are discussed below

- **Throughput** 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 Associated** 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** is the ability of an algorithm to perform uniform load balancing in spite of arbitrary node or link failure. The load balancing should be a good fault-tolerant technique.
- **Response Time** 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** is used to check the utilization of re-sources. It should be optimized for an efficient load balancing.
- **Scalability** is the ability of an algorithm to perform load balancing for a system with any finite number of nodes. This metric should be improved.
- **Performance** is used to check the efficiency of the system. This has to be improved at a reasonable cost, e.g., reduce task response time while keeping acceptable delays.

IV. RESEARCH OBEJECTIVES

The Objective of the proposed work is to balance the load and energy efficient using Dynamic load balancing model for cloud computing architecture.

The Objectives of the research work are:

- To optimize the performance of cloud computing architecture.
- To implement proposed method using java programming and simulate on cloud computing environment using CloudSim toolkit.
- To compare the Vector Dot with exiting load balancing techniques such as Active Clustering and Join- Idle queue for energy efficiency.
- To Analyze the behavior of research work using following parameters-
  - Energy Efficient Metric
  - Execution time
  - Response time
  - Resource utilization
  - Fault Tolerance
  - Scalability
  - Throughput
- To Evaluate the performance and behavior of proposed Vector Dot dynamic load balancing technique by comparing it with existing load balancing methods such as Join-Idle queue and Active Clustering.

V. STEPS TO IMPLEMENT THE RESEARCH WORK

A. Setup server_config.xml

In this research work we are using five servers having their different IDs, names, IP address, speed and RAM. The number of jobs can be increased or wane as per the requirement. As by increasing the number of jobs the speed of server has to be increased so that it cannot affect the overall performance of the system. The parameters of job are id, requestType, arrival Time and length. The arrival time should be in increasing order.

B. Initialize the Tomcat Server for project execution

To run the project on server Tomcat Server should be initialized. The Tomcat server is available in different versions. In this Research work Tomcat v6.0 server is configured. The steps for configure Tomcat v6.0 Server on Eclipse interface is following as:

a) Right click on project and then click on run on server.

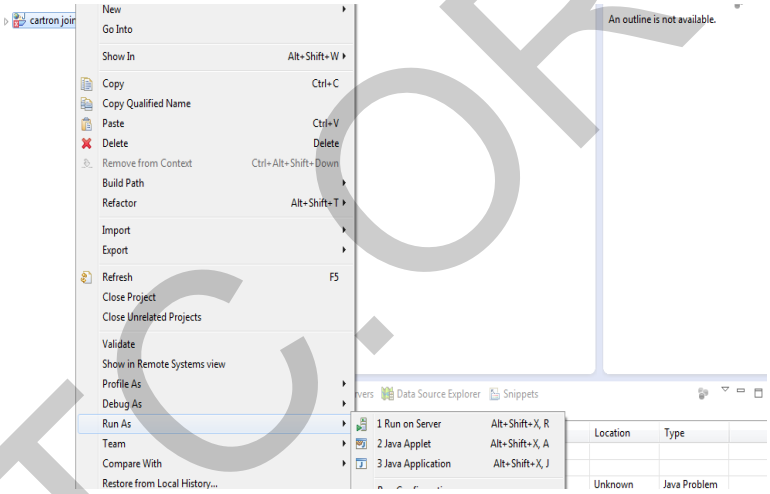

Fig. 2. project execution screen.

C. Simulation Results using Tomcat Server

D. When all algorithm are selected for load balancing, simulation results are as depicted in fig.3.

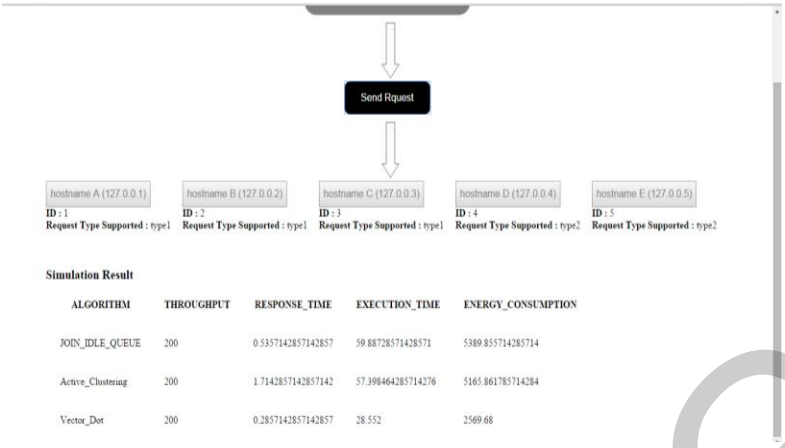

Fig. 3 Simulation results of Vector Dot, Join idle and Active Clustering

VI. RESULT ANALYSIS

In the result analysis, we have different output tables and graphical charts to compare the proposed method with other techniques.

A. Output Tables

The output tables shows the values for different parameters like throughput, response time,execution time and energy consumption when a particular technique is used for dynamic load balancing with the distinct number of jobs. The different tables are drawn for three dynamic load balancing techniques and then for a particular number of jobs these are compared and the experimental results show that our proposed model gives the best results in terms of energy consumption, execution time, response time and throughput. All three algorithms are implemented and compared on CloudSim tool for energy efficiency and load balancing. Table 1 depicts the result of different parameters for five jobs. During the comparison, Vector dot technique is counted as best model for producing the good results according to the user requirements.

Table 1. Different algorithms are compared for 5 Jobs.

Algorithms	Number of jobs/Tasks	Response Time	Execution Time	Energy Consumption
Active Clustering	For #5	1.714	57.39	5165
Join idle queue	For #5	0.535	59.88	5389
Vector Dot	For #5	0.285	28.55	2569

Table 2 depicts the result of different parameters for six jobs.

Algorithms	Number of jobs/Tasks	Response Time	Execution Time	Energy Consumption
Active Clustering	For #6	2.9	79.82	6043.8
Join idle queue	For #6	1.66	83.05	7474.5
Vector Dot	For #6	2.1	73.82	6643.8

Table 3. Depicts the result of different parameters for eleven jobs.

Algorithms	Number of jobs/Tasks	Response Time	Execution Time	Energy Consumption
Active Clustering	For #11	4.3	96.3	5333.2
Join idle queue	For #11	0.2	144.3	128288
Vector Dot	For #11	0.22	54.75	4927.5

All three algorithms are compared for energy efficiency and load balancing. The results show that Vector dot technique is the better because which consumes less energy and all the tasks are executed in less time with no delay. It concluded that Vector dot is best in the energy efficient technique in cloud computing.

B. GRAPHICAL CHARTS:

Graphical charts shows the comparison between different techniques based on distinct number of jobs considering discrete parameters like energy consumption, response time and execution time.

Fig.4 reveals the information about the energy consumption of datacenter and the results show that Vector Dot technique consumed less energy as compared to Active Clustering and Join idle queue. Therfore our proposed model is best in case of energy efficient.

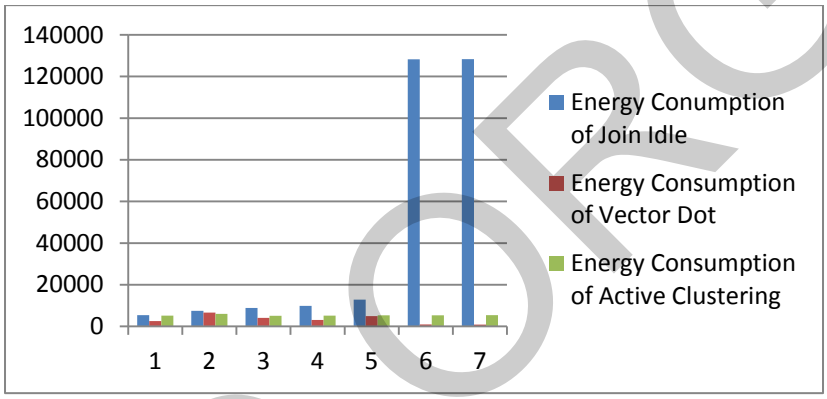

Fig. 4 Energy consumption comparison based on distinct number of jobs.

Response time is the time taken to respond by a particular load balancing algorithm in a distributed system. The response time is calculated by varying the number of jobs. Fig.5 depicts that the response time is minimum in case of Vector Dot technique.

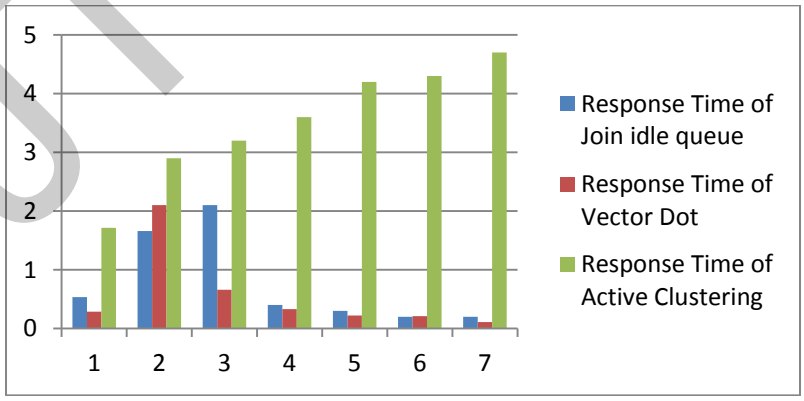

Fig. 5 Response Time comparison based on distinct number of jobs.

Execution time is estimated by calculating the total number of jobs executed within a fixed span of time. Fig.6 shows the execution time taken by three techniques for distinct number of jobs.

After implementation results show that the execution time of Vector Dot is less as compared to Active clustering and Join Idle queue.

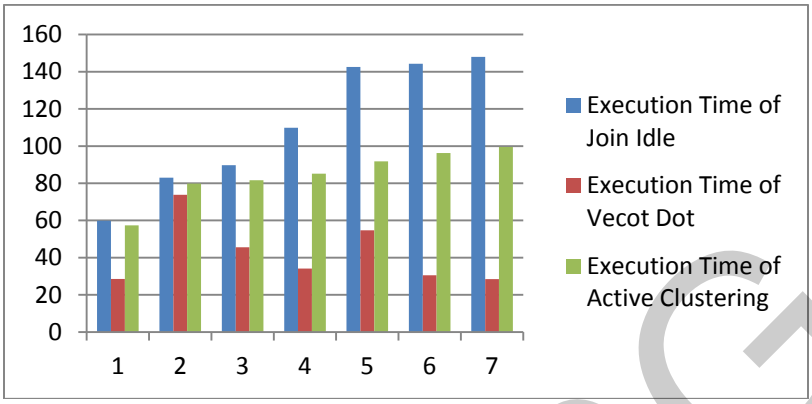

Fig. 6 Execution time comparison based on distinct number of jobs

VII. CONLUSIONS AND FUTURE SCOPE

In recent years, energy efficiency has emerged as one of the most important design requirements for modern computing systems, ranging from single servers to data centers and Clouds, as they continue to consume enormous amounts of electrical power. Apart from high operating costs incurred by computing resources, this leads to significant emissions of CO2 into the environment. For example, currently, IT infrastructures contribute about 2% of the total CO2 footprints. Unless energy-efficient techniques and algorithms to manage computing resources are developed, its contribution in the world's energy consumption and CO2 emissions is expected to rapidly grow. It has been shown that proper load balancing of computing resources can lead to a significant reduction of the energy consumption by a system, while still meeting the performance requirements. A relaxation of the performance constraints usually results in a further decrease of the energy consumption. Load balancing that is required to distribute the excess dynamic local workload evenly to all the nodes in the whole Cloud to achieve a high user satisfaction and resource utilization ratio.

In the research work we have proposed and implemented a Vector Dot on cloud environment using CloudSim Toolkit. And compared it with the Active Clustering and Join Idle queue. The results show that proposed technique is much better than the existing load balancing methods in terms of Response time, Execution Time, and Throughput. We also concluded that Vector Dot technique consumes less energy than Central Load Balancer.

FUTURE WORK

Cloud Computing is a vast concept and energy efficiency plays a very important role in case of Clouds. There is a huge scope of improvement in this area. We have implemented only two dynamic load balancing algorithms. But there are still other approaches that can be applied to balance the load and energy consumption in clouds. The performance of the given algorithms can also be increased by varying different parameters. We can also move our research work on any Private Cloud for the Security and further enhancements.

REFERENCES

[1] Ashima Narang, Dr. Vijay Laxmi "Comparision of a New Approach of Balancing the Load in Cloud Environment with the Existing Techniques" published in the INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY, in the October, 2014.
[2] Saeed Javanmardi , Mohammad Shojafar, Danilo Amendola, Nicola Cordeschi "Hybrid Job scheduling Algorithm for Cloud computing Environment" published in SpringerVerlag Berlin Heidelberg 2014.
[3] Tom Guérout, Samir Medjiah, Georges Da Costa, Thierry Monteil (2014)," *Quality of service modeling for green scheduling in Clouds*", In Elsevier.
[4] Bernardetta Addis, Danilo Ardagna, Antonio Capone , Giuliana Carello (2014)," Energy-aware joint management of networks and Cloud infrastructures", In Elsevier Computer Networks, pp.75–95.
[5] Rajesh Gorge Rajan and V.Jeyakrishnan *"A Survey on Load Balancing in Cloud Computing Environments"*Vol-2.Issue-12,December 2013.
[6] Dongliang Zhang,Changjun Jiang,Shu Li, " *A fast Adaptive Load balancing method for parallel practical based simulations*", Simulation Modelling Practice and theory 17(2009) 1032-1042.
[7] M.Randles,D.Lamb and A.Bendiab, "*A Comparative Study into distributed load balancing algorithm for cloud computing*" IEEE 24th International Conference on Advanced Information and Networking and Applications Workshops,pp.551-556,April 2010.
[8] Suriya Begum, Dr. Prashanth C.S.R, "*Review of load balancing in cloud Computing".* 10, Issue 1, No 2, January 2013.
[9] Rahul Bhoyar,"*Cloud Computing: Service Models ,Types,database and Issues*",Volume 3, Issue 3, March 2013.
[10] Argha Roy, Diptam Dutta, "*Dynamic Load balancing: Improve efficiency in cloud Computing*", Vol 15, issue 2, ISSN: 2278-9359,December 2013.
[11] Meysam Orouskhani *"Novel Cat Swarn Optimization algorithm for Unconstraint Optimization Problem",* IJITCS, vol.5, no.11, pp.32- 41, 2013.
[12] Mohiuddin Ahmed, Abu Sina Md. Raju Chowdhury, Mustaq Ahmed, Md. Mahmudul, *"An Advanced Survey on Cloud Computing and State-of-the-art",* Issue 1, No 1, January 2012.
[13] Jaspreet Kaur, "*The comparison of Load balancing Algorithm in cloud*",Vol. 2, Issue 3, May-Jun 2012.
[14] Ram Prasad Padhay,P Gautam Prasad Rao, "Load Balancing in Cloud Computing System",107CSO46.
[15] Jitendra Bhatia,Trith Patel,Harshal Trivedi ,Vishrut Majmudar, "*HTV Dynamic Load balancing For Virtual Machine Instances In Cloud*" , 18 Dec 2012, Pages 1-20 IEEE.
[16] Parin.V.Patel ,Hitesh.D.Patel, Pinal.J.Patel, "*A Survey Of Load Balancing In Cloud Computing*" IJERT, Vol.1,Issue 9,November 2012.
[17] P.Salot, "*A Survey of various Scheduling algorithm in cloud computing Environment*" ,IJRET ,Volume:2,Issue:2,Feb 2012.
[18] Y.Fang,F.Wang ,And J.Ge , "*A Task Scheduling algorithm Based on Load Balancing in Cloud Computing*" , Web Information Systems and mining,Lectures Notes in Computer Science,Vol 6318, Pages-271-277,2010.
[19] J.Kaur,"*Comparison of Load Balancing in cloud*", IJERT, Vol.2,No:3,pp. 1169-73,2012.
[20] Sujit Tilak, "*A Survey of Various Scheduling Algorithms in Cloud Environment",* IJEI, Volume 1, Issue 2 (September 2012) PP: 36-39.
[21] Saeed Parsa and Reza Entezari-Maleki," RASA: A New Task Scheduling Algorithm in Grid Environment" in World Applied Sciences Journal 7 (Special Issue of Computer & IT): 152-160, 2009.Berry M. W., Dumais S. T., O'Brien G. W. Using linear algebra for intelligent information retrieval, SIAM Review, 1995, 37, pp. 573-595.

[22] S. K. Garg, C. S. Yeob, A. Anandasivamc, and R. Buyya,"Environment-conscious scheduling of HPC applications ondistributed Cloud-oriented data centers", Journal of Parallel and Distributed Computing, Elsevier, Vol. 70, No. 6, May 2010, pages 1-18.