A comparative study of energy and task efficient load balancing algorithms in cloud computing

A D Gaikwad¹, K R Singh², S D Kamble³ and M M Raghuwanshi⁴

¹ Assistant Professor, YCCE, Nagpur, India
²,³ Associate Professor, YCCE, Nagpur, India
⁴ Professor, GHRCEM, Pune, India

E-mail: amolgaikwad.ag@gmail.com, singhkavita19@yahoo.co.in, shailesh_2kin@rediffmail.com, m_raghuwanshi@rediffmail.com

Abstract. The future growth of the Internet of Services has fundamentally changed the emergence of cloud computing. Cloud data centres serve multiple tenant demands for cloud applications that discharge vast amounts of electricity, leading to high operating costs and environmental diffusion of carbon dioxide (CO2). To fix this is the need for preservation to enable potential use by building a new structure and measuring the effect in a cloud data centre. Consequently, the use of pruned electricity reduces the cost of processing power. In order to meet energy-efficient data centres in the cloud, adjusting to optimal load balancing processing a good way for energy savings. To minimise the large energy use of cloud data centres herds, this focuses on increasing efficacy by breaking the workload evenly. In this paper, we plan to provide a comprehensive comparative analysis in cloud computing of current load balancing algorithms.

Index Terms— Cloud computing, Load balancing, Energy efficiency, Green computing

1. Introduction

Cloud Computing is an emerging modern platform for broad based technologies, for example: Computing and data from computers and PCs have been transferred into larger desktops have been reinventing their customers' studios on top of PCs Large businesses are interested in cloud computing in order to expand their technologies, its reachability and the economy associated with applications. Cloud computing distinguishes ISO as a business model (IaaS), Application as a Service (PaaS), Software as a Service (SaaS), and Protection as a Service (SEaaS) into the device models. These four shared in a single online portal which is powered by a virtual computer (VMs). VM's build the imagination of a dedicated computer for laptops. As a result of the improvement also improves according to the cost specifications, the host produces Vms. load exceeds the threshold values given to each system output is altered. Thus, reaching the power of datacentres is a difficulty. So, in such an environment, data centres were found to absorb 0.5 percent of the global demand for electricity in 2018. A annual improvement trend in data centres Last year, data centres' general capacity utilization has exponentially. Cloud storage eliminates consumption of energy due to the usage of the VM and its proper relocation during the load balance. Via high power demand the data centres deliver large amounts of carbon dioxide (CO2). Think it again is another key concept. It includes power equation counselling methods to minimize turbines for cloud computing [3]. Other innovations such as the improvement of software applications, the virtualization of computing resources
and dynamic voltage frequency scaling (DVFS) and the use of energy efficient hardware will provide high cloud power. Cloud storage eliminates consumption of energy due to the usage of the VM and its proper relocation during the load balance. Via high power demand the data centres deliver large amounts of carbon dioxide (C02). Think it again is another key concept. It includes power equation counselling methods to minimize turbines for cloud computing [3]. Other innovations such as the improvement of software applications, the virtualization of computing resources and dynamic voltage frequency scaling (DVFS) and the use of energy efficient hardware will provide high cloud power.

1.1. Cloud
The Green Cloud is a talking point that relates to a type of distribution of internet services that has many implications for the environment. Cloud data centres are more powerful than conventional data centres and use green data. Cloud Computing will assist with energy savings by deploying the great scale of virtualization, according to the latest Microsoft research.

1.2. Wide Scalability
For numerous clients, such as Google, Amazon, IBM Cloud and more than one million suppliers, Green Cloud runs on a very large scale. Green cloud computing makes it possible for consumers to get services anywhere, anywhere.

1.3. fault tolerance
With its efficient error tolerance, Cloud has the best service efficiency. It is more efficient to use cloud computing than any local device on a laptop.

1.4. Availability
Applications run in an independent platform environment with good accuracy and performance.

2. Green computing in clouds
The practise of introducing policies and procedures to increase the performance of computing resources is Green Computing [6], or Green IT In order to cut back on energy, the use and impact of their use on the ecosystem [10][12]. It requires the ability to gain fast and scalable access to high-end computing resources as High-Performance Computing (HPC) is prevalent in business and financial IT implementations. Through allowing use of data centres, cloud computing offers this computing infrastructure. This enables HPC users to access their applications and results, anywhere from the cloud, on demand and payable [7]. On distant wide-band data networks, cloud computing centres have allowed high speed computer networks which have led them to function and perform applications more efficiently when opposed to local private computers. Those centres do cost under single application software licences that run on applications clusters for the service of operation and hosting on-site data centres[11]. Nevertheless, the explosion of cloud computing networks and rising demand have increased considerably the stable energy use of data centres which is a key problem for industry and society as a major obstacle. [8]This rise in demand for electricity can not only increase energy prices but it raise carbon emissions as well. The major investment of energy results in a smaller benefit range from the cloud companies and further emissions of carbon are not best appropriate for the environment. Therefore, both from the point of view of the cloud provider and the climate, energy-efficient technologies that can address high energy usage are needed.

3. Essential characteristics of cloud computing
3.1 Self-Services on Demand
Via cloud service providers, a business application should be safe. The customer will have access to cloud services and online changes will be made to the services. The cloud application services are provided as needed, such as email, software apps, with no human contact with and service provider required. Amazon Web Services (AWS), Microsoft, Google, IBM and Salesforce.com are cloud service providers that provide self-service on demand.

3.2 Wide Access to Networks
Services for organization executives using their smartphones, tablets and laptops. They can use the access point to access cloud services wherever they are located. With the support of mobility features, the employee can work full time on the cloud project so that the organization can generate successful sales and services. Cloud resources can be collected over the network and accessed for deployment through standard mechanisms of different cloud models such as private, public, and hybrid.

3.3 Pooling of Resources
As computing is extended by customer demand, resources are shared together, and virtual resources are dynamically distributed to the consumer for higher service continuity. Resources include encoding, memory, bandwidth of the network, application services, etc.

3.4 Swift Elasticity
Cloud services are versatile and can be delivered to consumers easily and elastically. Customers can use the facilities anytime at any time. Based on the users, the facilities are easily scaled.

3.5 The Service Assessed
Through utilizing resource utilization, cloud computing resources are tracked, calculated, managed and transparently reportable to the user. The computational capabilities for the economical use of resources are also optimized.

3.6 Multi Tenaciousness
The Cloud Protection Alliance supports cloud computing. It refers to the need for multiple client constituencies for policy-driven compliance, segmentation, separation, governance, service levels, and chargeback/billing models. Consumers may use the service offerings of a public cloud provider or simply be from the same entity, such as entirely different business divisions rather than separate structures.

4. Load balancing algorithms review
A. Static Load Balancing: These algorithms rely on the completion time of the mission. [1]. In static algorithms, at compile time, decisions are made about load balancing. These are restricted to the environment where there are few variations in load. The current state of the system does not depend on these algorithms. The traffic is distributed equally between the servers by a static load balancer algorithm. When transmitting the load, it does not use the device data and is less.

B. Dynamic Load Balancing
The dynamic algorithm is determined by different node characteristics, such as network strengths and bandwidth. The need for this the node is continuously tested and is usually
difficult to implement. In the cloud computing world, dynamic algorithms are well adapted since they delegate function and assign the servers appropriate weights at runtime [1]

5. The need for energy aware load balancer

Computer machinery output has advanced, much faster than its power efficiency, over the past two decades. [3]. The energy used by the idle system should be like 0 in an ideal atmosphere and should be increased linearly with the system. Machines which need to extract energy at idle linearly also have to be implemented by over half of the energy they consume at full charge in today’s world. During an extended period, the data obtained suggests that the standard operating system for data centre servers will be much less than the optimal regime for use of resources. As the load is rising, in an energy proportionate model that is not powered low will consume low, very little energy as low and, last of all, more energy with a low load. The easiest, energy appropriate platform is to constantly run at 100 per cent efficiency.

Understanding the vital use of electricity at the cloud storage centres which is predicted to increase spectacularly in the potential power of the research community's participation in energy utilisation of resources and implementation plans, as well the mechanisms to enact those policies. Green Computing for CLOUDs (or Load balancing is indispensable) should also be accomplished, that can be realised with the help of the following two factors:

- Low energy consumption - It aims to avoid overheating by balancing the workload across all cloud nodes, reducing the volume of energy consumed.
- Reducing Carbon Pollution - Economic energy use and productive greenhouse emissions, fading close together. The bigger the energy used, the better the carbon footprint. With the aid of load balancing the decrease of energy consumption, so does carbon contamination that leads to sustainable green calculation.

Resource usage and energy management are not always the key subjects of conversation. As we have studied under cloud computing discourse, hardware usage and energy conservation are not always. But the usage of energy can be held to a minimum with careful balance of loads, which not only reduces costs, It also greens the business. Load balancing has also made components of cloud infrastructure scalability. However, boost the efficiency of dispersed activities will help minimise power demand and consumption of green cards by can compliance. A big necessity will also be the inclusion of alternative technologies covering a tremendous amount of energy efficiency for a large variety of cloud providers and from the environmental perspective, too. Knowledge that cloud web energy usage is key and will dramatically further increase in the future gives research team engagement in the control of energy resources as well as instrument placement of policies.

6. Metrics for load balancing

Load balancing is important to expand the vibrant local workload in cloud processing and distribute equal across all nodes. By allowing an important and equal distribution of resources to computer outlets, it allows to achieve high costume management. The Capital, resources. Requirement of proper load balancing aids to reduce resource usage, fight fail over, allow adaptability, stopped bottlenecks and excess supply, etc.30][31]. The numerous qualitative metrics or criteria that are deemed in cloud computing, load balancing is addressed as follows:

a. Throughput: the overall number of assets completed. If this instrument should work harder, a substantial output is required.
b. Associated Overhead: The amount of overhead created by the execution of the load balancing algorithm. Just a minimum general overhead is possible to execute the algorithm effectively.

c. Fault tolerant: it is the ability of the algorithm under conditions of error, the algorithm's ability to run correctly and consistently at any arbitrary machine node.

d. Migration time: The time taken to transfer a job from a machine to another computer in the system or relay it. To boost the system's performance, this will be the least time needed.

e. Response time: It's the minimal time needed to respond to a deck performing a complex algorithm relating load balancing.

f. Resource Usage: This is the extent to which the resourcing from your network. A usage algorithm providing optimum effective use of services is generated.

g. Scalability: Describes the system's ability to achieve a load balancing algorithm with a small number of processors or computers.

h. Performance: After performing load balancing, it reflects the efficiency of the system. If all the above criteria are optimally met, the system's efficiency will be greatly improved.

**Table no 1.** Comparison of existing scheduling algorithms

| Scheduling Method | Parameters Considered | Advantages | Disadvantages |
|-------------------|----------------------|------------|---------------|
| First Come First Serve | Arrival time | Simple Execution | Any other scheduled criteria are not considered |
| Round Robin | Time Arrival, Time quantum Load | Fairer balance of loads | Pre-emption is required |
| Opportunistic Load Balancing | Time Bound | Utilizing Best Resources | Weak make span |
| Execution Time Algorithm Minimum | Time Bound Completion Time | Quickest machine is selected for Scheduling | Poor Load Balanced |
| Completion Time Algorithm | Makespan, Expected completion time Makespan, Efficiency | Fairer balance of loads | No optimization | Selection of Resourced |
| Min-Min, Max-Min Genetic Algorithm | Makespan, Great Output and Performance | Fairer makespan, Completeness | Poor Load Balanced |
| Switching       | Performance, Optimization | Fairer balance of loads | Poor Switching of Loads in terms of cost and time |
|-----------------|---------------------------|-------------------------|--------------------------------------------------|
| Algorithm       | Makespan, Load balancing, Performance | Fairer Selection of Machine | Poor Criteria for Resource Selection |
| K-percent Best  | Makespan, Performance     | Better Energy Saving and Makespan | Complex Execution |
| Energy efficient| Energy                    | Fairer load balancing and Communication | Work for Light Weight Application |
| method using    | Consumption, Makespan, Execution time | Fairer load balancing and Energy Saving | Complex Execution and QOS Factor. |
| DVFS            | DENS Parameter Consider- Load, Energy and Congestion | Fairer load balancing and Energy Saving | Low Makespan |
| e-STAB Task     | Energy                    | Cost and Performance is considered for load Balancing | QOS Factor |
| Scheduling & Server Provisioning | Processing cost, Makespan | Fairer Load Balancing and Priority Based Balancing | Makespan factor to be improved |
| Improved Cost Based Algorithm | Completion Time | Total Completion Time not Mention | |
| Priority based Job Scheduling Algorithm | Fault and Cost factor | Highest resource has the advantages | |
7. Load balancing challenges
Cloud computing research is also in its early stages and some basic research is still ongoing. Early experiments in the cloud are still underway as some fundamental science is also under way. The research community is unresolved, especially problems when cargo juggling is a vital tool of the computing industry. In automated service: elasticity, capital to be represented or hands would be a key function. How will cloud platforms then be used or launched by retaining the same effectiveness of traditional systems and optimal capital. The primary two factors are Virtual Machine and Energy Aware management which is the main research is underway.

Virtual Machines: With virtualization, it is possible to view an entire machine as a file or a collection of files, to unload a heavily loaded physical machine, to transfer a virtual machine between physical machines. The primary aim is to spread the load within a datacenter or datacenter package. How can the load be dynamically spread as the virtual machine is transferred to prevent bottlenecks in cloud computer systems.

7.1. Energy Aware Management
If the optimum use of storage is implemented, a complete replication algorithm would not have been contemplated. This is because the same data can be saved in all replication nodes. Since there is a need for further room, it induces greater expense through robust replication algorithms. However, partial algorithms of replication, based on node capabilities, can allow some sections of data sets (totaling on an overlap) in each node to be saved, including, for example, processing speed and capacity[3] But this may lead to greater availability, but it enhances the sophistication of load balance algorithms, which aim to understand the availability of the components of a collection of data across different regions.

7.2. Replication (Storage)
If the optimum use of storage is implemented, a complete replication algorithm would not have been contemplated. This is because the same data can be saved in all replication nodes. Since there is a need for further room, it induces greater expense through robust replication algorithms. However, partial algorithms of replication, based on node capabilities, can allow some sections of data sets (totaling on an overlap) in each node to be saved, including, for example, processing speed and capacity[3] But this may lead to greater availability, but it enhances the sophistication of load balance algorithms, which aim to understand the availability of the components of a collection of data across different regions.

8. Open issues
Much analysis concerning energy consciousness preparation was undertaken using metaheuristic techniques for reducing energy consumption. Computing products yield a great deal of energy which makes them more suitable for susceptible errors and would typically increase the performance of systems and reduce device life. Cooling is very important in order to keep the temperature within appropriate boundaries. In the Google survey, 50% of the energy from the computer infrastructure is expected to be used as means for cooling equipment running electricity. Projects and findings to cut heat created by electricity and to de-heat the cooling systems of heated machines can be launched on the software side. It is possible to monitor the thermal state of physical devices; thus, virtual virtual equipment may be moved from the overheated physical device to mitigate heat induced. The aims of designing methods and ways of transacting virtual engine optimization, which retain it as a temperature, and there is no increase as well as reduction
of migratory overhead and declining efficiency. Scheduling protection and privacy-aware are another area that must be resolved using metaheuristic methodology as well.

9. Conclusion
The sector has commonly welcomed cloud storage, but certain recent subjects, such as load balancing, virtual equipment relocation, server consolidation, energy efficiency, etc, have not yet been discussed. Inter alia, the problem of load balance is fundamental to those challenges, namely, the consistent allocation of the workload to nodes in all the cloud for a major customer to accomplish meeting and resource consumption ratios. This ensures each device is organised effectively, and equally, The new Load Balancer policies were checked focus mostly on reduced overhead, operation turnaround times and efficiency increases etc, but all processes acknowledged the concerns of energy use and emissions from carbon. There is, however, a need to develop an effective load balancing solution that can increase performance for cloud computing.

10. References
[1] Ammar Rayes, Bechir Hamdaoui, Mehdi Dabbagh and Mohsen Guizani, “Energy-Efficient Resource Allocation and Provisioning Framework for Cloud Data Centers”, IEEE 2015.
[2] R. W. Lucky, “Cloud computing”, IEEE Journal of Spectrum, Vol. 46, No. 5, May 2009, pages 27-45.
[3] M. D. Dikaiakos, G. Pallis, D. Katso, P. Mehta, and A. Vakali, “Cloud Computing: Distributed Internet Computing for IT and Scientific Research”, IEEE Journal of Network Computing, Vol. 13, No. 5, September/October 2009, pages 10-13.
[4] G. Pallis, “Cloud Computing: The New Frontier of Internet Computing”, IEEE Journal of Network Computing, Vol. 14, No. 5, September/October 2010, pages 70-73.
[5] 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.
[6] 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.
[7] S. K. Garg, C. S. Yeob, A. Anandasisvame, and R. Buyya, “Environment-conscious scheduling of HPC applications on distributed Cloud-oriented data centers”, Journal of Parallel and Distributed Computing, Elsevier, Vol. 70, No. 6, May 2010, pages 1-18.
[8] K. M. Nagothu, B. Kelley, J. Prevost, and M. Jamshidi, “Ultra low energy cloud computing using adaptive load prediction”, Proceedings of IEEE World Automation Congress(WAC), Kobe, September 2010, pages 1-7.
[9] B. P. Rimal, E. Choi, and I. Lumb, “A Taxonomy, Survey, and Issues of Cloud Computing Ecosystems, Cloud Computing: Principles, Systems and Applications”, Computer Communications and Networks, Chapter 2, pages 21-46, DOI 10.1007/978-1-84996-241-2, Springer – V erlagLondonLimited, 2010.
[10] S. Kabiraj, V. Topka, and R. C. Walke, “Going Green: A Holistic Approach to Transform Businesses”, International Journal of Managing Information Technology (IJMIT), Vol. 2, No. 3, August 2010, pages 22-31.
[11] K. T. Rao, P. S. Kiran, and L. S. S. Reddy, “Energy Efficiency in Datacenters through Virtualization: A Case Study”, Global Journal of Computer Science and Technology, Vol. 10, No. 3, April 2010, pages 2-6.
[12] J. Baliga, R. W. A. Ayre, K. Hinton, and R. S. Tucker, “Green Cloud Computing: Balancing Energy in Processing, Storage, and Transport”, Proceedings of the IEEE, Vol. 99, No. 1, January 2011, pages 149-167.
[13] A. M. Alakel, “A Guide to dynamic Load balancing in Distributed Computer Systems”, International Journal of Computer Science and Network Security (IJCNS), Vol. 10, No. 6, June 2010, pages 153-160.
[14] 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.
[15] 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 Advanced Mechatronics and Automation (ICIMA), Wuhan, China, May 2010, pages 240-243.
[16] Y. Zhao, and W. Huang, “Adaptive Distributed Load Balancing Algorithm based on Live Migration of Virtual Machines in Cloud”, Proceedings of 5th IEEE International Joint Conference on INC, IMS and IDC, Seoul, Republic of Korea, August 2009, pages 170-175.
[17] R. Stanojevic, and R. Shorten, “Load balancing vs.distributed rate limiting: a unifying framework for cloud control”, Proceedings of IEEE ICC, Dresden, Germany, August 2009, pages 1-6.
[18] 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.
[19] Q. Zhang, L. Cheng, and R. Boutaba, “Cloud computing: state-of-the-art and research challenges”, Journal of Internet Services and Applications, Vol. 1, No. 1, April 2010, pages 7-18.
[20] 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.

[21] 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.

[22] Y. Lua, Q. Xiea, G. Klioth, A. Gellerb, J. R. Larushb, 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.

[23] Sung Ho Jang, Tae Young Kim, Jae Kwon Kim, Jong Sik Lee School, “The Study of Genetic Algorithm-based Task Scheduling for Cloud Computing”, International Journal of Control and Automation Vol. 5, No. 4, December, 2012.

[24] 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.

[25] 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.

[26] A. 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

[27] 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.

[28] 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.

[29] A. Singh, M. Korupolu, and D. Mohapatra, “Server-storage virtualization: integration and load balancing in data centers”, Proceedings of the ACM/IEEE conference on

[30] Monir Abdullah, Mohamed Othman, Cost Based Multi QoS Job Scheduling using Divisible Load Theory in Cloud Computing, International Conference on Computational Science, ICCS 2013.

[31] Peter Mell, Timothy Grance, “The NIST definition of Cloud Computing (September, 2011)”, Accessed on May, 2014.

[32] Tracy D. Braun, Howard Jay Siegel, Noah Beck, “A Comparison of Eleven Static Heuristics or Mapping a Class of Independent Tasks onto Heterogeneous Distributed Computing Systems”, Journal of Parallel and Distributed Computing 61, pp. 810-837, 2001.

[33] Weicheng Huai, Zhuzhong Qian, Xin Li, Gangyi Luo, and Sanglu Lu, “Energy Aware Task Scheduling in Data Centers, Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable.

[34] Teena Mathew, K. Chandra Sekaran, John Jose, “Study and Analysis of Various Task Scheduling Algorithms in the Cloud Computing Environment”, International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE 2014.

[35] S.Nagadevi1, K.Satyaapriya2, Dr.D.Malathy3, “A Survey On Economic Cloud Schedulers For Optimized Task Scheduling”, International Journal of Advanced Engineering Technology, 2013.