

Review on Smart Cloud For Effective Load Balancing Using Artificial Neural Network

Ashwini Y. Gudadhe
PG student

Department of Information Technology,
Tulsiramji Gaikwad-Patil College of Engg. & Tech.,
Nagpur, Maharashtra, India
ashwini0129@gmail.com

Mukul Pande
Assistant Professor

Department of Information Technology,
Tulsiramji Gaikwad-Patil College of Engg. & Tech.,
Nagpur, Maharashtra, India
Mukulpande2002@gmail.com

Abstract— Cloud computing is a powerful paradigm in modern computing, offering a better utilization of IT resources and at the same time reducing the operational costs and increasing the elasticity of the IT infrastructure. Clouds are available in various form, started from general purpose public clouds having virtually unlimited resources, to small private clouds which is used inside enterprises for specific business purpose. Job scheduling is one of the biggest issues in cloud computing. Main motivation is to schedule users' requests to allocate resources to these requests to finish the tasks in minimum time. In this paper, we presented comprehensive reviews on different techniques use for scheduling. We have proposed implementation of artificial neural networks to optimize the job scheduling results in cloud.

Keywords— Cloud computing, service-level agreement (SLA), Job Scheduling, Artificial Intelligence, Artificial Neural Networks (ANN), Load balancing.

I. INTRODUCTION

Cloud computing helps business shift their focus to developing better business applications that will bring true business value [1]. Cloud computing mainly provide four types of service such as Infrastructure as a service (IaaS), examples includes Amazon Web Services, Secondly Software as a Service (SaaS), Third is Platform as a Service (PaaS) such as Google Apps and last is Communication as a service or CaaS [2][3][4]. Clouds are deployed on physical infrastructure where Cloud middleware is implemented for delivering services to customers. Such an infrastructure and middleware differ in their services, administrative domain and access to users. There are three types of Cloud deployments models namely Public Cloud, Private Cloud and Hybrid Cloud. Due to the exponential growth of cloud computing, it has been widely adopted by the industry and there is a rapid expansion in data-centers[1].

Public clouds like Amazon, Microsoft Azure, Google are general purpose clouds and the main characteristic of public

cloud is that they should be able to run any kind of workload at the specific time requested by the customer. To make this possible, the resources of the cloud must be always able to accommodate the next customer request. In large clouds where the customer request represents only a small fraction of the total amount of physical resources available, it is possible for the cloud provider to forecast the required amount of resources and plan accordingly [5].

Secondly, private clouds are restricted in resources and the request based First In First Served (FIFS) approach raises several problems, because the request can represent a significant fraction of the overall cloud resources. In private clouds approval process need to perform, in which an administrator will have to decide which requests are being approved or rejected based on the cloud load at that given time interval. When end user will a request for a set of workloads, the system has to logically lock the corresponding set of physical resources in the cloud until the request is approved or rejected by a cloud administrator. Such locking approach might prevent other end users to submit requests and so the overall service provided to end users will not be ideal.

On the other hand, private clouds usually have a very specific designated purpose and can benefit of additional upfront information. The additional information refers to the typical workload types or workload scheduling patterns while decision power and flexibility refers to the fact that in some scenarios, the customer does not require a specific time frame for the workload execution, and allows the cloud provider to schedule the workload based on a set of given rules. Applications of such clouds are in e-Learning [8], software testing, data mining and reporting, payroll are the Valid Applications [9].

Cloud computing system has many characteristics. Cloud computing users can access the resources via the Internet regardless of the users location or the machine type at the minimum cost. Its implementation and configuration are required the minimum skills [10]. Cloud computing resources utilization is proficient due to its sharing and scheduling between multiple users. Cloud computing users do not need to concern about resources and system maintenance, which is

being performed at the provided side. Finally, the security issues in cloud computing can be solved easier than the issues in the traditional systems that is being solved by specialized people and resources at the provider side using several traditional security methods such as encryption methods and Hash functions [11,13,17].



Figure. Cloud computing paradigm.

Artificial Neural Network (ANN) is an information processing paradigm that simulates the human brain neural which is shown in Figure. It designed to perform the same task as the way that the human brain executes a specific task or function. The adaptive nature of this network is consider one of the most important feature, where “learning by example” is used to solve the problems. This model is used to solve complex or ambiguous systems problems, pattern classification and recognition [18][19].

ANN can give very great result when it used with complex systems that has not fully understandable relationships[20]. ANN model has three main issues: network topology, transfer function, and training algorithm. ANN consists of processing units, weighted connections, activation rule, and learning rules. Following Figure shows Basic Structure of Artificial Neural Network (ANN).

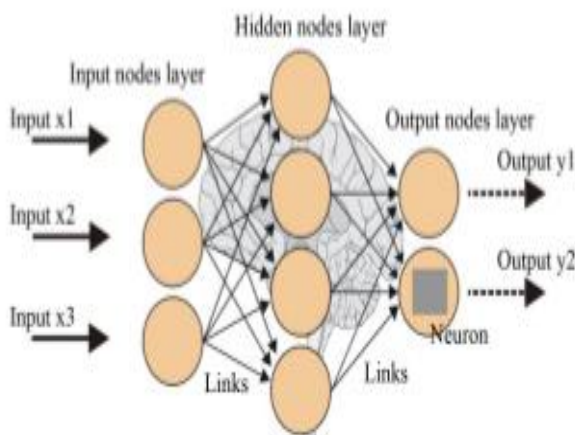


Figure. Basic structure of an Artificial Neural Network (ANN) [22].

Neural network consist of three or more layers and each layer has number of processing unit that called neurons [21]. It has input layer, output layer and hidden layers. ANN link the input layers with the output layers using hidden layers with nonlinear transformation function and weighted connections. Artificial neural networks can have different number of layers and different number of nodes. The nature of the problem and the degree of complexity are controlled the number of hidden layers and their neurons. The nonlinear transformation functions give an advantage over the predictable functions.

II. LITERATURE REVIEW

In this section, we have presented reviews on recent works which is relevant to the job scheduling in cloud.

H. Mehta, P. Kanungo, and M. Chandwani (2011) proposed approach for job scheduling using Decentralized content aware. This makes the uses of unique and special property(USP) of requests and computing nodes to help scheduler to decide the best node for processing the requests. Furthermore it uses the content information to narrow down the search. Its improved the searching performance hence increasing overall performance and also reduces idle time of the nodes [23].

Y. Lua, Q. Xiea, G. Klioib, A. Gellerb, (2011) presented mechanism called Join-Idle-Queue. They used a protocol to limit redirection rates to avoid remote servers overloading. They also uses a middleware to support this protocol and also make uses a heuristic to tolerate abrupt load changes [24].

M. Randles, D. Lamb, and A. Taleb-Bendiab(2010) used Active Clustering [25]. Description includes Optimizes job assignment by connecting similar services by local re-wiring .This approach performs better with high resources . Secondly utilizes the increased system resources to increase throughput. Also, Degrades as system diversity increases .

V. Nae, R. Prodan, and T. Fahringer, (2010) used approach, Event-driven [26].It make the uses complete capacity event as input, analyzes its components and generates the game session load balancing actions . Benefits includes the Capability of scaling up and down a game session on multiple resources according to the variable user load 2. Occasional QoS breaches as low as 0.66%[1].

A. Singh, M. Korupolu, and D. Mohapatra(2008) make the use of Vector Dot [27] ItUses dot product to distinguish node based on the item requirement. It has benefits to Handles hierarchical and multidimensional resource constraints. Secondly it removes overloads on server, switch and storage.

III. PROPOSE WORK

The first step includes in is to create user database(DB) which consist of number of registered customers to access the cloud services.

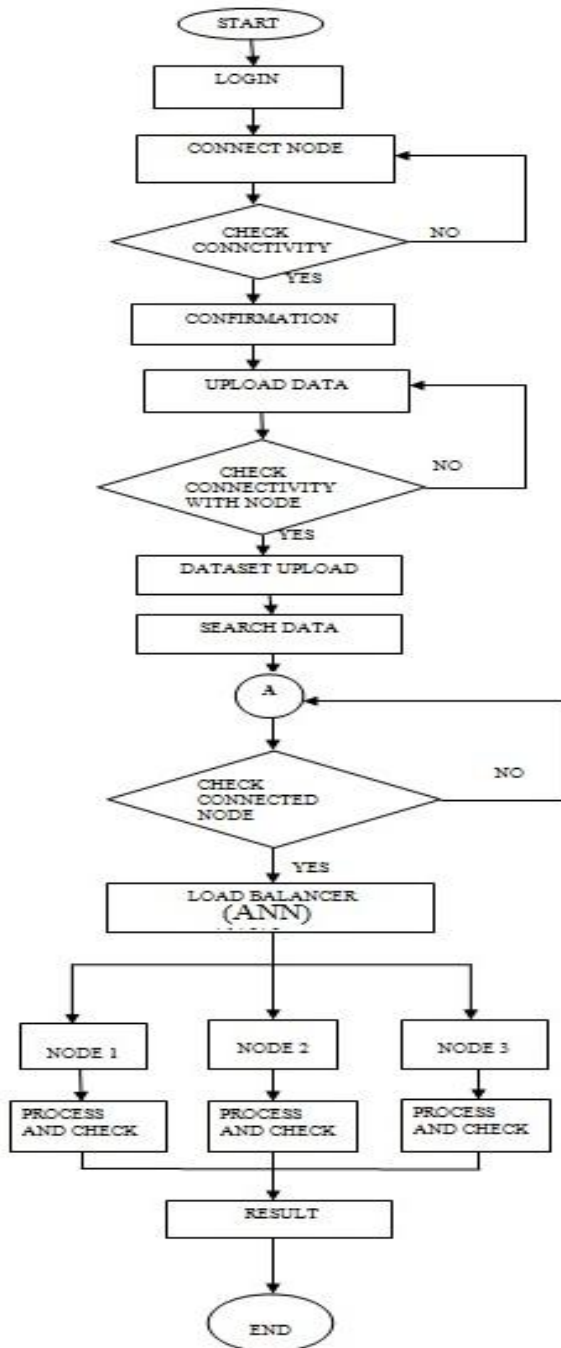


Figure: Flow of proposed work

Registered user is verified by business service provider (BSP) then processes the request. Next step is to make request to required file by uploading on cloud. If Requested file is

easily available then it respond by making file available else request is handle by ANN job scheduler for better job scheduling.

IV. Conclusion

In this review, many researches tried to solve the problem of job scheduling by using in many approaches. But they faced many problems which is mentioned. One of the best way to perform job scheduling is ti use ANN. Neural Networks designed to mimic the way that the human brain executes a specific task or function. Its most important feature is the adaptive nature, where “learning by example” is used to solve complex or ambiguous systems problems, pattern classification and recognition. ANNs are trained using different learning rates, parameters and propagation methods. It learn by changing the connections between the input and output layers. It gives better results than existing one

REFERENCES

- [1] Nada M. Al Sallami, Ali Al, daoud, Sarmad A Al Alousi, " Load Balancing with Neural Network", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 4, No. 10, 2013, pp No.138-145.
- [2] Abdulaziz Aljabre, " Cloud Computing for Increased Business Value", International Journal of Business and Social Science, Vol. 3 No. 1; January 2012.
- [3] K. Dinesh, G. Poornima, K. Kiruthika, " Efficient Resources Allocation for Different Jobs in Cloud", International Journal of Computer Applications (0975 – 8887) Volume 56– No.10, October 2012.
- [4] Gary Garrison, Sanghyun Kim, and Robin L. Wakefield, " Success Factors for Deploying Cloud Computing", communications of the ACM | September 2012 | vol. 55 | no. 9, doi:10.1145/2330667.2330685.
- [5] Amazon Elastic Compute Cloud (EC2): <http://aws.amazon.com/ec2/> accessed Dec 2011.
- [6] Microsoft Azure: <http://www.windowsazure.com/en-us/> accessed Dec 2011.
- [7] GoogleCloud: <http://www.google.com/apps/intl/en/business/cloud.html>, accessed Dec 2011.
- [8] Sam Averitt, Michael Bugaev, Aaron Peeler, Henry Shaffer, Eric Sills, Sarah Stein, Josh Thompson, Mladen Vouk, Virtual Computing Laboratory (VCL), International Conference on Virtual Computing Initiative, May 7-8, 2007, IBM Corp., Research Triangle Park, NC, pp. 1-16.
- [9] Cloud Computing Use Cases, A white paper produced by the Cloud Computing Use Case Discussion Group, Version 4.0, 2 July 2010.
- [10] Malathi, M. (2011) Cloud Computing Concepts. 3rd International Conference on Electronics Computer Technology (ICECT).
- [11] Karajeh, H., Maqableh, M. and Masa'deh, R. (2014) Security of Cloud Computing Environment. 23rd IBIMA Conference on Vision 2020: Sustainable Growth, Economic Development, and Global Competitiveness.
- [12] Mathur, P. and Nishchal, N. (2010) Cloud Computing: New Challenge to the Entire Computer Industry. 2010 1st International Conference on Parallel Distributed and Grid Computing (PDGC), Solan, 28-30 October 2010, 223-228.
- [13] Maqableh, M., Samsudin, A. and Alia, M. (2008) New Hash Function Based on Chaos Theory (CHA-1). 20-27.

- [14] Maqableh, M.M. (2010) Secure Hash Functions Based on Chaotic Maps for E-Commerce Applications. *International Journal of Information Technology and Management Information System (IJITMIS)*, 1, 12-22.
- [15] Maqableh, M.M. (2010) Fast Hash Function Based on BCCM Encryption Algorithm for E-Commerce (HFBCCM). 5th International Conference on E-Commerce in Developing Countries: With Focus on Export, Le Havre, 15-16 September 2010, 55-64.
- [16] Maqableh, M.M. (2011) Fast Parallel Keyed Hash Functions Based on Chaotic Maps (PKHC). *Western European Workshop on Research in Cryptology, Lecture Notes in Computer Science, Weimar, 20-22 July 2011*, 33-40.
- [17] Maqableh, M.M. (2012) Analysis and Design Security Primitives Based on Chaotic Systems for E-Commerce. Durham University, Durham.
- [18] Benny Karunakar, D. and Datta, G.L. (2007) Controlling Green Sand Mould Properties Using Artificial Neural Networks and Genetic Algorithms—A Comparison. *Applied Clay Science*, 37, 58-66. <http://dx.doi.org/10.1016/j.clay.2006.11.005>.
- [19] Abdella, M. and Marwala, T. (2005) The Use of Genetic Algorithms and Neural Networks to Approximate Missing Data in Database. *Computing and Informatics*, 24, 577-589.
- [20] Varahrami, V. (2010) Application of Genetic Algorithm to Neural Network Forecasting of Short-Term Water Demand. *International Conference on Applied Economics—ICOAE, Athens, 26-28 August 2010*, 783-787
- [21] Chen, C.R. and Ramaswamy, H.S. (2002) Modeling and Optimization of Variable Retort Temperature (VRT) Thermal Processing Using Coupled Neural Networks and Genetic Algorithms. *Journal of Food Engineering*, 53, 209-220. [http://dx.doi.org/10.1016/S0260-8774\(01\)00159-5](http://dx.doi.org/10.1016/S0260-8774(01)00159-5).
- [22] Tadiou, K.M. (2014) The Future of Human Evolution. <http://futurehumanevolution.com/artificial-intelligence-future-human-evolution/artificial-neural-networks>.
- [23] 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.
- [24] Y. Lua, Q. Xiea, G. Klioth, 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.
- [25] 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.
- [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] A. 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.