Efficient System for Load Balancing in Cloud using Artificial Neural Network

Ashwini Y. Gudadhe[1], Mukul Pande[2]

PG Student, Department of Information Technology, Tulsiramji Gaikwad-Patil College of Engg. & Tech., Nagpur,

Maharashtra, India[1]

Assistant Professor, Department of Information Technology, Tulsiramji Gaikwad-Patil College of Engg. & Tech.,

Nagpur, Maharashtra, India[2]

Abstract: Cloud computing is the long dreamed vision of computing as a utility, where users can remotely store their data into the cloud so as to enjoy the on-demand high quality applications and services from a shared pool of configurable computing resources. 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, experimental results showed that scheduling using Artificial Neural Network (ANN) can perform better scheduling than existing approach.

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

I. INTRODUCTION

Cloud computing is an emerging paradigm that accesses the network and share computing resources with expedient and minimal management efforts. 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].

Aim of Cloud computing is to maximum the benefit of distributed resources and aggregate them to be able solve large scale computation problems. It provides the computing services for users as public utility, which is available to organizations and individual [9]. Service providers provide the services to the subscribers on contractual basis. They charge the subscribers according to the provided services. Users need to pay for the provided service with the payment system. Thus, cost reduction is considered one of the main advantages of using cloud computing. Again service providers guarantee the quality of the provided services such as data processing, data storage and data access [10].

Infrastructure refers to both the applications delivered to end users as services over the Internet and the hardware and system software in datacenters that is responsible for providing those services. 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 [5]

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].

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].

Figure 1. Cloud computing paradigm

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. Fallowing Figure shows Basic Structure of Artificial Neural Network (ANN).
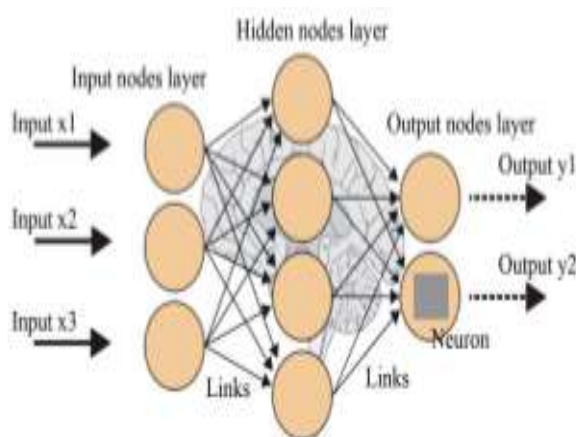

Figure 2. Basic structure of an Artificial Neural Network (ANN) [22]

 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. Kliotb, 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].

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.
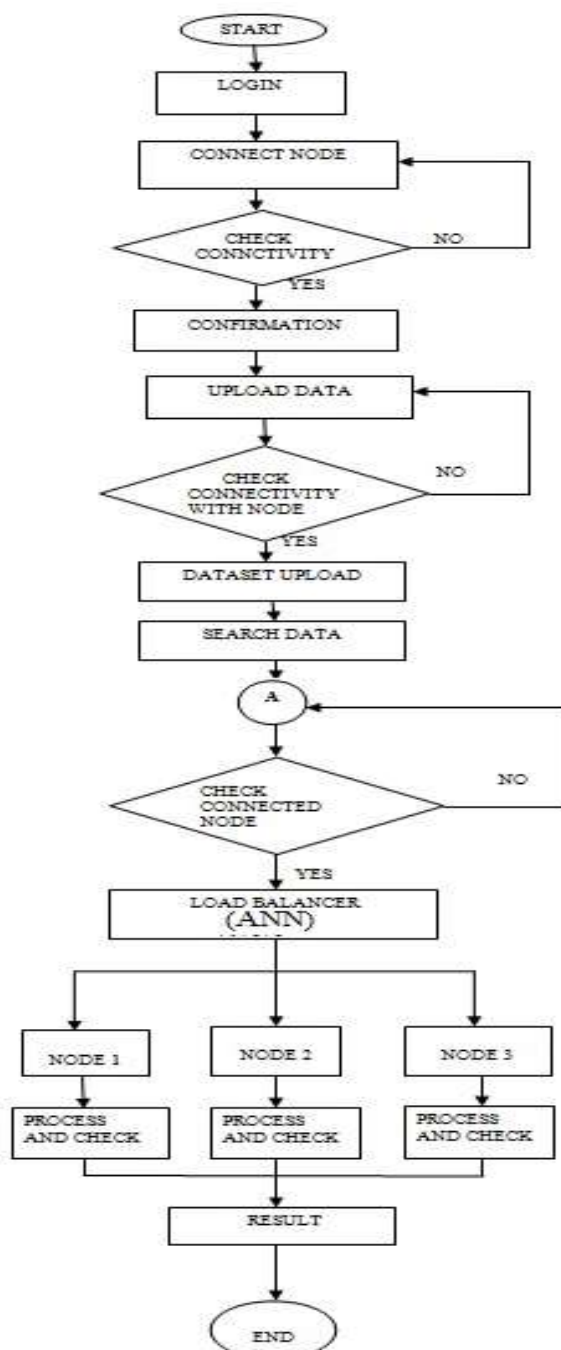

Figure 3: Flow of proposed work

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. RESULTS & DISCUSSION

In this section, we have discussed about steps which are involved in complete execution. This is homepage for proposed system. It consist login of client, Business service provider (BSP) login, Infrastructure Service provider (ISP) login and additional login admin. Admin is main user of system.
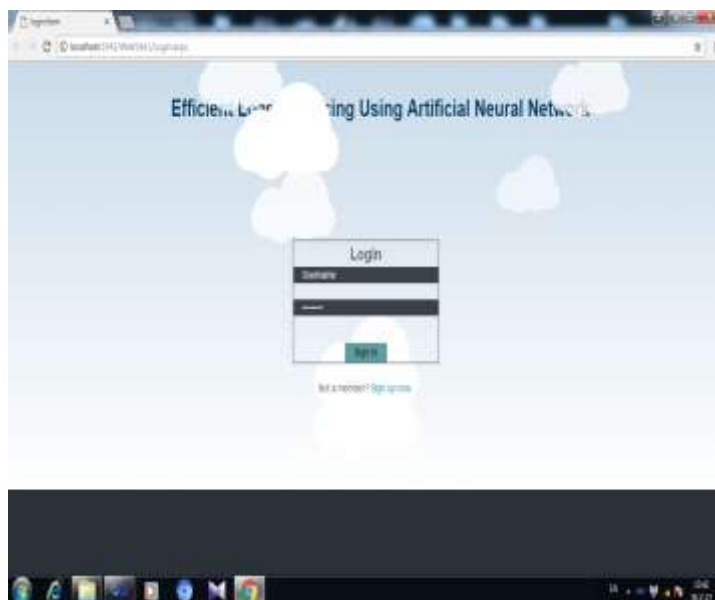

Figure 4. Homepage

Second step is to perform registration process. This is registration page. This page provides registration for servers, suppose want to add new server to improve performance.
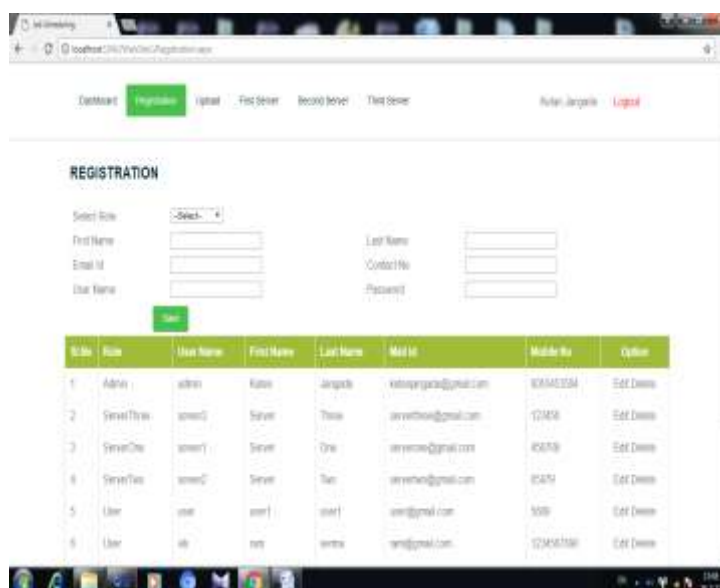

Figure 5. Registration Process

Client can also go through registration process which consists of menus full name, User Name, Password, Email address and contact number as input fields. These details needed for registration process. All registered rolls are visible with this dashboard.
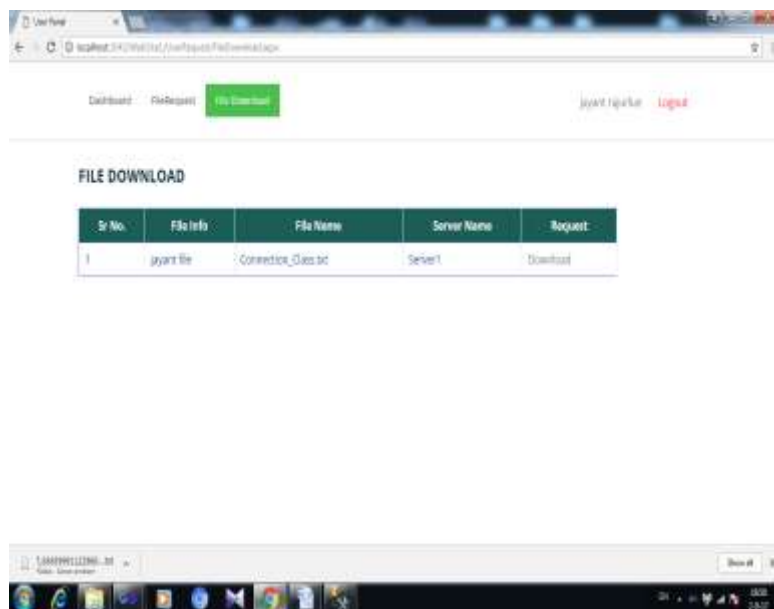
Figure 6. Resource available for download

Once scheduling process is over, resources will be available to respective client. It also shows that from which server resource is allocated as multiple server are maintained.

Figure 7. Result analysis

The figure 7 shows the comparison between existing and proposed system in terms time needed to execute the request. The Y- axis shows time of systems to execute process and X-axis shows the existing and proposed algorithm. The proposed system execute process in less time than existing work because of appropriate selection of cloud and well managed scheduling process and availability of multiple servers.

V. CONCLUSION

In this paper we have presented load balancing approach with the help of ANN. Load Balancing is an essential task in Cloud Computing environment to achieve maximum utilization of resources. One of the loads balancing scheme that provide easiest managing of request to different servers because dynamic load balancing algorithm are difficult to simulate but are best suited in heterogeneous environment of cloud computing. One of the best way to perform job scheduling is to 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. Experimental results showed that proposed algorithm works better than existing one in terms of time and load balancing.

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.

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

[25] Siddhi Raut, Lalit Dole, Jayant Rajurkar," Analysis of Wireless Sensor Network Protocols using Data Mining Techniques", International Journal of Advanced Engineering Research and Science (IJAERS) [Vol-3, Issue-9, Sept- 2016], ISSN: 2349-6495(P) | 2456-1908(O),pp no 90-94.

[26] 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.

[27] 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.

[28] 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.