IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-7, Issue-10, Oct 2019 E-ISSN: 2347-2693

Comparative Analysis Study of Various Techniques of Energy Efficiency in Cloud Computing

Neeshu Sharma[1*], Suresh Kumar Kaswan[2]

[1,2]Dept. of Computer Science & Engg., School of Engg. & Technology, RIMT University, Punjab, India

[*]*Corresponding Author: neeshusharma2908@gmail.com*

DOI: https://doi.org/10.26438/ijcse/v7i10.229234 | Available online at: www.ijcseonline.org

Accepted: 07/Oct/2019, Published: 31/Oct/2019

Abstract— Cloud Computing is one of the most emerging field for research now a days. The cloud is responsible to provide various set of services to users which requires a lot of energy. As they are growing up with a rapid rate, the burden on cloud is increasing daily. Various researchers are working on cloud efficiency with major factor as energy efficiency. As energy efficiency will not only increase user handling rate but also decrease overall global cost and pollution. In this paper, various previous techniques used for energy efficiency are discussed on the basis various performance parameters to analyse the best available techniques.

Keywords: Cloud Computing, Machine Learning, Deep Learning, Convolutional neural networks (CNN)

I. INTRODUCTION

Distributed computing frameworks are intended to help the openness and arrangement of different administrations situated applications by the clients. Cloud registering administrations are made accessible through the server firms or server farms. To take care of the developing demand for calculations and vast volume of information, the cloud figuring conditions gives superior servers and rapid mass capacity gadgets. These assets are the real wellspring of the power utilization in server farms alongside cooling and cooling gear. Also the vitality utilization in the cloud is relative to the asset use and server farms are nearly the world's most astounding shoppers of power. Then again, Cloud server farm can lessen the add up to vitality devoured through errand combination and server solidification utilizing the virtualization by outstanding tasks at hand can have a similar server and unused servers can be turned off. The aggregate processing intensity of the Cloud server farm is the whole of the processing intensity of the individual physical machine. Mists utilizes virtualization innovation in server farms to designate assets for the benefits according to require. Mists gives three dimensions of access to the clients: SaaS, PaaS , and IaaS. The undertaking started by the client can vary enormously from client to the client. Substances in the Cloud are self-governing and self-intrigued; in any case, they will share their assets and administrations to accomplish their individual and aggregate objectives. In such an open domain, the booking choice is a test given the decentralized idea of the earth. Every element has explicit necessities and destinations that need to accomplish. Server union are permitting the various servers running on a solitary physical server at the same time to limit the vitality devoured in a server farm. Running the different servers on a solitary physical server are acknowledged through virtual machine idea. The assignment solidification likewise know as server/remaining task at hand union issue. This vitality effective asset assignment keeps up the use of all registering assets and circulates virtual machines such that the vitality utilization can limit. The objective of these calculations is to keep up accessibility to process hubs while diminishing the aggregate energy devoured by the cloud foundation.

Machine learning calculations are ending up increasingly more well-known because of the accessibility of vast volumes of information and the headways in equipment that makes it conceivable to examine this information. These calculations are available in extensive scale server farms, which represent 3% of the worldwide vitality consumption1. AI planning. Furthermore, we can find AI techniques not only predicting behaviors or situations but also managing workflows among machines of a distributed system. These methods are basically AI planning, methods for planning, and scheduling events using as guide a set of operators and a set of observations. Some works done by [4,5] show methods for scheduling jobs on a grid environment generating, for each job, resource requirements and available resource workflows. These workflows are searched and used by AI planning methods to schedule the application execution matching resource requirements with resource availability.

Convolutional neural systems chiefly comprise of three sorts of layers: convolutional layers (Conv), pooling layers (Pooling) and completely associated layers (FC). Each layer may contain thousands to a great many neurons. A solitary neuron takes a few sources of info, figures their weighted aggregate, and sends the yield to the neurons in the following layer. Thusly, unmistakable layers apply distinctive tasks to their data sources and create yields for the layers that pursue. Convolutional layers: These layers apply convolutions to the contribution with a few channels and add a predisposition term to the outcomes. Frequently, a nonlinear capacity (called initiation work) is additionally connected to the outcomes. Convolutional layers abuse spatial network and shared loads. The parameters of a convolutional layer are diminished significantly contrasted with a normal shrouded layer of a MLP. Convolutional layers are the most computational escalated layers in CNNs. Pooling layers: These layers play out a nonlinear down sampling activity on the information. They parcel the contribution to a lot of sub-locales and yield tested outcomes from these sub-areas. In light of their examining strategy, pooling layers can be classified into: greatest pooling, normal pooling and stochastic pooling. Pooling layers dynamically lessen the measure of parameters and in addition control display overfitting. Pooling layers are normally put between two convolutional layers. Completely associated layers: Unlike in convolutional layers, neurons in FC layers have full associations with all yield from the previous layers. As a result, a FC layer has a lot a larger number of parameters than a convolutional layer. In any case, since convolution activities are supplanted by duplications, completely associated layers require less computational power.

Deep learning utilizes profound neural systems to illuminate machine learning errands. Neural systems comprise of an accumulation of neurons and associations. A neuron gets numerous contributions from ancestor neurons furthermore, produces one yield. The yield is a weighted aggregate of the sources of info pursued by the neuron's initiation work, which is normally nonlinear. Neurons are sorted out as layers. Neurons in the same layer are not associated. Neurons with no forerunner are called input neurons, neurons with no successor are called yield neurons. On the off chance that the quantity of layers between the info neuron and the yield neuron is extensive, at that point it is called profound neural system. There is no strict definition, however in general, with in excess of eight layers it is considered "profound" [2]. Current profound neural systems can have many layers [5]. Neurons are wired through associations. Every association exchanges the yield of a neuron I to the contribution of another neuron j. Every association has a load that will be duplicated with the actuation, which will increment or decline the flag. This weight will be balanced amid the learning procedure, and this procedure is called preparing

II. LITERATURE REVIEW

Advancements, for example, parallel programming and virtualization empowered various clients and applications sharing registering assets, i.e., multi-tenure. Server bunches, in spite of the fact that not at the size of present day server farms, are situations where multi-tenure ought to be bolstered. Heath et al. [2] centered on the issue of dispersing customer solicitations to the servers in a heterogeneous bunch with the goal that a decent tradeoff between throughput and vitality investment funds will be found. They structured a group with the capacity to arrange itself to improve for a assortment of measurements (or a mix thereof), among which vitality utilization was highlighted and decreasing it was picked as a primary goal. They created investigative models for demand appropriations and asset usage, and after that utilized recreated toughening to discover a demand dissemination that limits power to-throughput proportion. Their technique offered over 40% reserve funds in vitality utilization.

Chen et al. [3] structured calculations to limit the number of running servers by means of dynamic provisioning and circulate the load to these servers so that sets aside to 30% vitality. Urgaonkar et al. [4] conceived an online control calculation to perform affirmation control and asset the board in a server farm utilizing Lyapunov Optimization. They defined stochastic enhancement issues whose arrangements amplify a joint utility consolidating application throughput and the sum of vitality spared.

Beloglazov et al. [5], [6] proposed strategies for energy efficient cloud asset assignment and planning calculations to meet nature of administration requests while looking after low control utilization. They made utilization of dynamic virtual machine (VM) assignment and live relocation (development from one physical hub to another) to accomplish diminished vitality utilization. The calculation for VM position is a Modified Best Fit Decreasing calculation, and the calculation for choosing which VMs to move hopes to limit the quantity of VM relocations expected to bring down CPU usage underneath a limit at all hosts.

Gandhi et al. [7] utilized a blend of prescient and responsive asset provisioning to fulfill benefit level assertions (SLA) while setting aside to 35% vitality. The proposed framework breaks down long haul request accounts to set up a base outstanding burden, bolsters this base remaining task at hand to a prescient controller that allots simply enough assets to meet SLA necessities, also, increases it with a basic receptive controller to deal with situations where real interest is higher than the anticipated interest.

Van et al. [8] created dynamic VM provisioning and situation administrators that endeavor to guarantee SLA consistence while decreasing vitality utilization, and place VMs on the least number of physical machines by means of live relocation. Uniting undertakings in a server farm can diminish the number of servers required, yet discharged assets may at present expend vitality in the inactive state.

Lee and Zumaya [9] thought about inactive control attract expansion to dynamic vitality utilization and created heuristics to limit add up to vitality utilization. Xu and Fortes [10] utilized a hereditary calculation bolstered with multi-target streamlining to all the while limit the measure of squandered assets, vitality utilization and warm dispersal costs.

Shen et al. [11] displayed CloudScale, a forecast based online framework for versatile cloud asset assignment. Their asset request indicator utilizes a quick Fourier change to distinguish a mark that can be utilized to evaluate future requests. On the off chance that a mark isn't discovered, the indicator, utilizes a discrete-time Markov chain. Moreover, Cloud Scale employments preemptive VM movement to stay away from clashes, alongside powerful voltage and recurrence scaling to spare vitality. The outcome is that CloudScale can spare 8-10% aggregate vitality utilization.
Heller et al. [12] presented ElasticTree, a power administrator with an attention on the server farm organize components (joins and switches). ElasticTree screens traffic conditions in the information focus, and essentially kills the switches and connections in the event that they are not required. The creators utilized a mix of straight programming, covetous canister pressing, and topology-mindful heuristic ways to deal with accomplish up to half decrease in system vitality use.

Berl et al. [13] gave a progressively exhaustive study of vitality effective distributed computing arrangements. Liao et al. [14] utilized machine figuring out how to locate the best setup for memory prefacers. They utilized an assortment of calculations including closest neighbor, innocent Bayes, C4.5 choice tree, Ripper, bolster vector machines, calculated relapse, multi-layer perceptron, and spiral premise work. Bodik et al. [15] connected factual machine learning models, for example, straight and LOESS relapse, to ideal control for server farms. Foreseeing how much time and assets will be spent by applications is important to have the capacity to plan occupations proficiently.

III. CONCLUSION

In this work an extensive comprehensive study was conducted to analyze power usage and behavior of various energy efficient algorithms with various well known DNN techniques like CNN models and other GPU based training frameworks. Various other power models are also reviewed from literature based on ML and AI. The major favorable findings indicating the usefulness and accuracy DNN based models over ML and AI. So, a detailed workload characterization will be proposed to facilitate the best framework for energy efficiency in cloud computing with deep neural network solution.

IV. FUTURE WORK AND METHODOLOGY

This work will explore the use of machine learning methods to predict the occurrence of correlated failures. With the prediction results, VM consolidation mechanism will can also be adopted to further optimize the fault-tolerance and energy consumption of the cloud computing systems.

Table:1 Conclusive Literature Review

Sr. No.	Author's	Title of the Paper	Study	Conclusions
1.	Albert Greenberg, James Hamilton, David A. Maltz, Parveen Patel	The Cost of a Cloud:Research Problems in Data Center Networks	The resources inside the data centers often operate at low utilization due to resource stranding and fragmentation. To attack this first problem, we propose (1) increasing network agility, and (2) providing appropriate incentives to shape resource consumption. Second, cloud service providers are building out geo-distributed networks of data centers.	To improve data center efficiency. First, we need to increase internal data center network agility, to fight resource fragmentation and to get more work out of fewer servers – reducing costs across the board. Second, we need to pursue the design of algorithms and market mechanisms for resource consumption shaping that improve data center efficiency.
2	T. Heath, B. Diniz,E.V.Carrera,W. Meira Jr, and R. Bianchini	Energy Conservation in Heterogeneous Server Clusters	A Cooperative Web server for a heterogeneous cluster that uses modeling and optimization to minimize the energy consumed per request.	We conclude that Web servers need to self-configure intelligently on heterogeneous clusters for higher energy savings. We also conclude that the style of modeling that allows our system to self-configure should be more widely applied in the systems community.

3.	Gong Chen, Wenbo He, Jie Liu , Suman Nath , Leonidas Rigas , Lin Xiao† , Feng Zhao	Energy-Aware Server Provisioning and Load Dispatching for Connection-Intensive Internet Services	Study on unique properties, performance, and power models of connection servers, based on a real data trace collected from the deployed Windows Live Messenger.	An algorithm (hybrid algorithm for load balancing) can save a significant amount of energy .
4.	Rahul Urgaonkar, Ulas C. Kozat, Ken Igarashi, Michael J. Neely	Dynamic Resource Allocation and Power Management in Virtualized Data Centers	An alternate approach that makes use of the queueing information available in the system to make online control decisions	We have used an alternate approach that makes use of the queueing information available in the system to make online control decisions.
5.	Anton Beloglazov* and Rajkumar Buyya	Energy Efficient Resource Management in Virtualized Cloud Data Centers	We present first results of simulation-driven evaluation of heuristics for dynamic reallocation of VMs using live migration according to current requirements for CPU performance. The results show that the proposed technique brings substantial energy savings, while ensuring reliable QoS.	Defined the problem of minimizing the energy consumption while meeting QoS requirements and stated the requirements for VM allocation policies.
6	Anton Beloglazov a, Jemal Abawajy b , Rajkumar Buyya a	Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing	A survey of research in energy-efficient computing and propose: (a) architectural principles for energy-efficient management of Clouds; (b) energy-efficient resource allocation policies and scheduling algorithms considering QoS expectations and power usage characteristics of the devices;	The development of a software platform that supports the energy-efficient management and allocation of Cloud data center resources. In order to reduce the cost of software engineering, we will extensively reuse existing Cloud middleware and associated technologies.
7.	Yuan Chen, Daniel Gmach, Martin Arlitt, Manish Marwah, Anshul Gandhi	Minimizing Data Center SLA Violations and Power Consumption via Hybrid Resource Provisioning	Two techniques to dynamically allocate capacity: predictive provisioning handles the estimated base workload at coarse time scales (e.g., hours or days) and reactive provisioning handles any excess workload at finer time scales (e.g., minutes).	Proposed a hybrid approach that proactively allocates resources for the predictable demand pattern and leverages a reactive controller to deal with excess demand.
8	Hien Nguyen Van; Tran, F.D.; Menaud, J.-M. Cloud Computing (CLOUD	Performance and power management for cloud infrastructures," in Cloud Computing (CLOUD)	A dynamic resource provisioning system is needed capable of addressing two main issues like How much resource (CPU, memory...) to allocate to hosted applications? –Where to place the application workloads within the datacenter to maximize energy savings?	Addressed the problem: –Resource allocation in Cloud infrastructures – Application performance and energy cost while providing –Cloud administrator high-level knobs to control the resource –Management system with regard to application-level SLAs –Datacenter exploitation costs
9.	Young Choon Lee · Albert Y. Zomaya	Energy efficient utilization of resources in cloud computing systems	Task consolidation is an effective method to increase resource utilization and in turn reduces energy consumption. Recent studies identified that server energy consumption scales linearly with (processor) resource utilization.	Resource utilization directly relates to energy consumption, we have successfully modeled their relationship and developed two energy-conscious task consolidation heuristics.
10.	Ahmed Sallam1,2 and Kenli Li1,∗	A Multi-objective Virtual Machine Migration Policy in Cloud Systems	Study on the migration process as a multiobjective problem where the objectives are typically non-commensurable. Therefore, we propose a novel migration policy consolidated by a new elastic multi-objective optimization strategy to evaluate different objectives (including migration cost) simultaneously, and to provide the flexibility for manipulating different cases.	Virtual machine migration is a potential solution for many critical situations in a Cloud environment such as load imbalance, lower utilization and workload hotspots resulting from the dynamic fluctuation of the system workload.
11	Zhiming Shen, Sethuraman Subbiah, Xiaohui Gu,	CloudScale: Elastic Resource Scaling for Multi-Tenant Cloud Systems	System that automates finegrained elastic resource scaling for multi-tenant cloud computing infrastructures. CloudScale employs online resource demand predictionand prediction	An interactive application typically needs to avoid high resource pressure for maintaining sufficient resources to serve any requests as soon as they arrive.

			error handling to achieve adaptive resource allocation without assuming any prior knowledge	
12	Brandon Heller, Srini Seetharaman, Priya Mahadevan,	ElasticTree: Saving Energy in Data Center Networks	Elastic Tree, a network-wide power1 manager, which dynamically adjusts the set of active network elements — links and switches—to satisfy changing data center traffic loads.	Built Elastic Tree, which through data-center-wide traffic management and control, introduces energy proportionality in today's non-energy proportional networks.
13	Mehmet Demirci	A Survey of Machine Learning Applications for Energy-Efficient Resource Management in Cloud Computing Environments	Studying non-ML proposals to energy conservation in data centers, and also how ML has been applied towards other objectives in the cloud	The most common use of machine learning in this context is for predicting future resource demands. Accurate estimation of these demands can allow cloud providers to develop intelligent resource management policies which rely on task scheduling and consolidation to turn on the minimum number of machines in the data center,thus conserving energy.
14	Shih-wei Liao1,2, Tzu-Han Hung3, Donald Nguyen4,	Machine Learning-Based Prefetch Optimization for Data Center Applications	Study on the effectiveness of different processor prefetch con_gurations, which can greatly inuence the performance of memory system and the overall data center.	Design and Implement a framework for optimizing parameter values based on machine learning.
15	Peter Bod´ık, Rean Griffith, Charles Sutton, Armando Fox, Michael Jordan, David Patterson	Statistical Machine Learning Makes Automatic Control Practical for Internet Datacenters	Rich statistical models of the application performance, simulation-based methods for finding an optimal control policy, and change-point methods to find abrupt changes in performance.	Shortcomings of automating datacenters using closed-loop control can be addressed by replacing simple techniques of modeling and model management with more sophisticated techniques imported from statistical machine learning.

REFERENCES

[1] A. Greenberg, J. Hamilton, D. A. Maltz, and P. Patel, "The cost of a cloud: research problems in data center networks," ACM SIGCOMM computer communication review, vol. 39, no. 1, pp. 68–73, 2008.

[2] T. Heath, B. Diniz, E. V. Carrera, W. Meira Jr, and R. Bianchini, "Energy conservation in heterogeneous server clusters," in Proceedings of the tenth ACM SIGPLAN symposium on Principles and practice of parallel programming. ACM, 2005, pp. 186–195.

[3] G. Chen, W. He, J. Liu, S. Nath, L. Rigas, L. Xiao, and F. Zhao, "Energyaware server provisioning and load dispatching for connection-intensive internet services." in NSDI, vol. 8, 2008, pp. 337–350.

[4] R. Urgaonkar, U. C. Kozat, K. Igarashi, and M. J. Neely, "Dynamic resource allocation and power management in virtualized data centers," in Network Operations and Management Symposium (NOMS), 2010 IEEE. IEEE, 2010, pp. 479–486.

[5] A. Beloglazov and R. Buyya, "Energy efficient resource management in virtualized cloud data centers," in Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing. IEEE Computer Society, 2010, pp. 826–831.

[6] A. Beloglazov, J. Abawajy, and R. Buyya, "Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing," Future generation computer systems, vol. 28, no. 5, pp. 755–768, 2012.

[7] A. Gandhi, Y. Chen, D. Gmach, M. Arlitt, and M. Marwah, "Minimizing data center sla violations and power consumption via hybrid resource provisioning," in Green Computing Conference and Workshops (IGCC), 2011 International. IEEE, 2011, pp. 1–8.

[8] H. N. Van, F. D. Tran, and J.-M. Menaud, "Performance and power management for cloud infrastructures," in Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on. IEEE, 2010, pp. 329–336.

[9] Y. C. Lee and A. Y. Zomaya, "Energy efficient utilization of resources in cloud computing systems," The Journal of Supercomputing, vol. 60, no. 2, pp. 268–280, 2012.

[10] J. Xu and J. A. Fortes, "Multi-objective virtual machine placement in virtualized data center environments," in Green Computing and Communications (GreenCom), 2010 IEEE/ACM Int'l Conference on & Int'l Conference on Cyber, Physical and Social Computing (CPSCom). IEEE, 2010, pp. 179–188.

[11] Z. Shen, S. Subbiah, X. Gu, and J. Wilkes, "Cloudscale: elastic resource scaling for multi-tenant cloud systems," in Proceedings of the 2nd ACM Symposium on Cloud Computing. ACM, 2011, p. 5.

[12] B. Heller, S. Seetharaman, P. Mahadevan, Y. Yiakoumis, P. Sharma, S. Banerjee, and N. McKeown, "Elastictree: Saving energy in data center networks." in NSDI, vol. 10, 2010, pp. 249–264.

[13] A. Berl, E. Gelenbe, M. Di Girolamo, G. Giuliani, H. De Meer, M. Q. Dang, and K. Pentikousis, "Energy-efficient cloud computing," The computer journal, vol. 53, no. 7, pp. 1045–1051, 2010.
[14] S.-w. Liao, T.-H. Hung, D. Nguyen, C. Chou, C. Tu, and H. Zhou, "Machine learning-based prefetch optimization for data center applications," in Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis. ACM, 2009, p. 56.
[15] P. Bodık, R. Griffith, C. Sutton, A. Fox, M. Jordan, and D. Patterson, "Statistical machine learning makes automatic control practical for internet datacenters," in Proceedings of the 2009 conference on Hot topics in cloud computing, 2009, pp. 12–12.

Authors Profile

Neeshu Sharma, Research Scholar, RIMT University, is working for Chandigarh University as an Assistant Professor in Department of Computer Science & Engineering. Her Expert area is deepa learning, nueral networks and cloud computing. She having 10 years of experience and research orieted person. She published 10 research papers and attended various conferences and FDPs at intenational and national level.

Suresh Kumar Kaswan, pursued Bachelor of Technology in Computer Science & Engineering from Kurukshetra University, India and Master of Technology in Computer Science & Engineering from C.D.L. University, India in 2005 and 2008 respectively. He is Ph.D. in CSE and currently working as Professor2, Department of Computer Science & Engg., School of Engg. & Technology, RIMT University, Punjab. He is a member of ISROSET since 2018, a life member. He has published 20 research papers in reputed international and national journal & conferences. His main research work focuses on ICT infrastructure and tools for grassroots & society developments. He has 14 years of technical teaching, administration experience.

