Design and Implementation of Resource Aware and Load Balancing in Cloud Environment

Sanjoli Gupta[1] Meenakshi Gupta[2]
[1]M.Tech Student [2]Assistant Professor
[1,2]Department of Computer Science & Engineering
[1,2]RIMT-IET

Abstract— Cloud computing is a new class of network based computing that provides the computing resources as a service over a network to the customers on their demand. The unique concept of cloud computing creates new opportunities for Business and IT enterprises to achieve their goals. In cloud computing, usually there are many jobs have to be executed with the available resources to achieve optimal performance, minimal total time for completion, shortest response time, and efficient utilization of resources etc. To achieve these objectives and high performance, it is important to design and develop a multi objective scheduling algorithm. Hence it is most challenging to schedule the tasks along with satisfying the user's Quality of Service requirements. The paper aims to improve the performance of CPU, memory and network operations by reducing the load of a virtual machine (VM) by using Load Balancing Method. Finally, it optimizes the resource utilization by using Resource Aware Scheduling Algorithm.

Key words: Cloud Computing, Hyper-Heuristic Scheduling Algorithm; Load Balancing; Resource-Aware Scheduling Algorithm; Makespan Database

I. INTRODUCTION

Cloud computing could be a promising and up-coming innovation which allows the purchasers to pay as they need, it empowers facilitating of pervasive applications from client, explorative, and business areas. Cloud computing is golf stroking forth utility-situated IT administrations to purchasers round the world. The developing expense of calibration and overseeing laptop frameworks is prompting out-sourcing of business administration's to facilitating focuses. It provides with a capable and an upgraded approach to perform the occupations that were place along by the purchasers in numerous views, maybe, openness, flexibility and flexibility. Cloud computing makes the long-held long for utility registering conceivable with a pay- as-you-go, endlessly versatile, all around accessible framework. With cloud computing, you'll begin very little and switch resolute be immense fast. That's the explanation cloud computing is progressive, in spite of the likelihood that the innovation it's supported is organic process. Cloud computing takes the innovation, administrations, and applications that are like those on the web and remodel them into a self-administration utility. Cloud computing could be a reflection seeable of the thought of pooling physical qualities and exhibiting them as a virtual asset. It's another model for provisioning assets, for composition applications, and for stage autonomous consumer access to administrations. Clouds will are available varied types, and also the administrations and applications that keep running on clouds might presumably be sent by a cloud administration provider.

A. Cloud Types

Cloud computing is that the methodology for utilizing out of reach servers on the online to handle, store and method info as hostile utilizing a digital computer. Cloud computing is isolated into 2 specific arrangements of models:

B. Deployment Models

This refers to the placement and management of the cloud's infrastructure.

C. Service Models

This consists of the actual kinds of services that you just will access on the cloud computing platform.

II. RELATED WORK

The problem of developing a geographical load balancing (GLB) scheme for distributed Internet data centers (IDCs) when they are price-makers in the deregulated electricity markets, i.e., GLB may impose impact on electricity prices in IDC locations due to the large power consumption of IDCs. Taking into account the dynamic characteristics (e.g., time-varying power demands and generation outputs) and actual physical constraints (e.g., active/reactive power flow balancing, transmission congestion, and network loss) of smart grids, it is difficult to obtain price impact models with analytical expressions accurately. As a result, it is challenging to design an efficient GLB scheme without requiring price impact models. To overcome the above challenge, we propose a price-sensitivity aware GLB scheme. The key idea of the proposed scheme is to impose proper limits on the workloads allocated to the IDC locations with high price-sensitivity coefficients by exploiting a number of interactive information between IDCs and main grids, so that the sudden increase in the total cost could be avoided. Here, the price-sensitivity coefficient is defined as the ratio of the percentage change in price to the percentage change in IDC power demand.[1]

Cloud Computing is merely defined as delivering IT services over the Internet. It is ubiquitous, on-demand provisioning of resources. The fundamental principle of reusability of IT capabilities is the basis of cloud computing. To meet the growing demand for computations and large volumes of data, the model of cloud computing environment provides high performance servers and persistent storage devices like hard disk drives. These resources are the major source of the enormous power consumption in cloud data center along with air conditioning and cooling equipment. Energy consumption and diffusion of carbon that contaminate the atmosphere go hand in hand. The more the energy consumed, higher is the carbon footprint, which in turn maximizes the operating cost. Energy aware load balancer like Weighted First-cum-First-Served, helps in avoiding overheating and overloading of servers, by evenly

balancing the workload across all the servers in the data center, hence reducing the amount of energy consumed. Dynamic placement of virtual machines in the data center enables to ensure availability of computing resources, tries to minimize overloaded servers and save the power consumption by shutting down some under-utilized, idle servers and thereby in maximizing the usage of resources. The main aim of the dynamic VM placement ensures the mapping of virtual machines to physical machines at run time, in such a way that the physical machine can be utilized to their maximum efficiency, and is non-preemptive in nature, where the already running VMs are neither shut down nor interrupted but in turn provides a list of live VM migrations to be executed to reach the optimal solution, thereby saves energy consumption to a greater extent. The central objective of this work is dedicated to develop an energy efficient resource provisioning framework with dynamic virtual machine placement using energy aware load balancer in cloud.[2]

Many of today's cloud datacenters are powered by electricity generated from brown energy (e.g., fuel fossil and oil), which directly translates into severe harm to the environment. To reduce the carbon footprint and the operating cost of datacenters, there is a clear trend to migrate to green datacenters, which are entirely (or partially) powered by renewable energy. However, given a portfolio with multiple off- and on-site power supplies (such as power grid, renewables, and batteries), it is challenging for datacenter operators to conduct efficient power management and thus meet the highly dynamic user demand. In this paper, we proposed an online algorithm, which is called EcoPower, to perform eco-aware power management and load scheduling jointly for geographically distributed cloud datacenters. Our objective is to minimize the time-average eco-aware power cost of cloud datacenters while still ensuring the qualityof-experience (QoE) constraint of user requests. To this end, we formulated the problem into a constrained stochastic optimization problem and apply the Lyapunov optimization theory to design an online control algorithm, which approaches the optimality with explicitly provable upper bounds. We also conducted extensive trace-driven simulations, and our results show that our proposed EcoPower algorithm can achieve a good balance between power cost savings, environment protection, and the user QoE, with the eco-aware power cost being cut down by over 20%. We found that wind dominant, solar complementary is a better strategy for cloud datacenters to integrate renewables into their power supply.[3]

Aiming at the current problems that most physical hosts in the cloud data center are so overloaded that it makes the whole cloud data center' load imbalanced and that existing load balancing approaches have relatively high complexity, this paper has focused on the selection problem of physical hosts for deploying requested tasks and proposed a novel heuristic approach called LB-BC (Load Balancing based on Bayes and Clustering). Most previous works, generally, utilize a series of algorithms through optimizing the candidate target hosts within an algorithm cycle and then picking out the optimal target hosts to achieve the immediate load balancing effect. However, the immediate effect doesn't guarantee high execution efficiency for the next task although it has abilities in achieving high resource utilization. Based on this argument, LB-BC introduces the concept of achieving the overall load balancing in a long-term process in contrast to the immediate load balancing approaches in the current literature. LB-BC makes a limited constraint about all physical hosts aiming to achieve a task deployment approach with global search capability in terms of the performance function of computing resource. The Bayes theorem is combined with the clustering process to obtain the optimal clustering set of physical hosts finally. Simulation results show that compared with the existing works, the proposed approach has reduced the failure number of task deployment events obviously, improved the throughput, and optimized the external services performance of cloud data centers.[4]

Distributing the system workload and balancing all incoming requests among all processing nodes in cloud computing environments is one of the important challenges in today cloud computing world. Many load balancing algorithms and approaches have been proposed for distributed and cloud computing systems. In addition the broker policy for distributing the workload among different datacenters in a cloud environment is one of the important factors for improving the system performance. In this, authors present an analytical comparison for the combinations of VM load balancing algorithms and different broker policies. They evaluate these approaches by simulating on CloudAnalyst simulator and the final results are presented based on different parameters. The results of this research specify the best possible combinations.[5]

Energy consumption has become a major challenge in cloud computing infrastructures. Cloud computing data centers consume enormous amount of electrical power resulting in high amount of carbon dioxide that affects the green environment as well as high operational costs for cloud providers. On the other hand, reducing the energy consumption would negatively impact the SLA (Service Level Agreement) that is a crucial concern in any resource allocation policy. In this paper, authors propose a novel power aware load balancing method, named ICAMMT to manage power consumption in cloud computing data centers. Authors have exploited the Imperialism Competitive Algorithm (ICA) for detecting over utilized hosts and then we migrate one or several virtual machines of these hosts to the other hosts to decrease their utilization. Finally, they consider other hosts as underutilized host and if it is possible, we migrate all of their VMs to the other hosts and switch them to the sleep mode. The results indicate that our method as compared to the previously proposed resource allocation policies such as LR-MMT (local Regression-Minimum Migration Time), MAD-MMT (Median Absolute Deviation- Minimum Migration Time), Bee-MMT (Bee colony algorithm- Minimum Migration Time) and non-Power aware policy offers least power consumption and SLA violation.[6]

Load balancing plays a vital role in Cloud computing to enhance throughput, optimize resource use and reduce response time. The main features to be considered while selecting a load balancing algorithm for cloud is the ability of the algorithm to address distributed network, dynamic environment and self regulation. Biased random sampling is one such algorithm; it allocates jobs by performing a random walk in the network. The selection of

neighbour is uniformly distributed among the neighbour nodes in case of biased random sampling algorithm. Improved version of random sampling algorithm uses cost based load computation to select node for random walk. This paper introduces neighbour awareness and prediction mechanisms to further improve the selection process of nodes for random walk. The proposed algorithm selects the least loaded node from the neighbour list for the random walk. This can be achieved by computing probability of each neighbour based on perceived load of the respective neighbour. Thus the probability of choosing lightly loaded node can be increased and hence the job waiting time can be decreased further.[7]

With the continuing growth of cloud computing services, power consumption has become one of the most challenging issues in data center environments. With the support of today's virtualization technology, the efficiency and flexibility of data center management can be greatly enhanced, creating great energy saving opportunities. However, effective energy aware design is a non-trivial task, considering the size of the data center, the dynamic fluctuation of workloads and the variation of computing resource requests. In this, authors propose CoolCloud: a practical solution for managing the mappings of VMs to physical servers. This framework solves the problem of finding the most energy efficient way (least resource wastage and least power consumption) of placing the VMs considering their resource requirements. Experiment result demonstrates our design can effectively improve data center energy efficiency and scales well to large size data centers. Comparing with industry leading product VMware's Distributed Resource Scheduler (DRS), our design offers better performance in both load balancing and power consumption.[8]

III. PROPOSED SCHEME

A. *Objectives*

1) To analyze and reduce the makespan time of jobs by using low-level heuristic and Hyper-Heuristic scheduling algorithms.
2) To balance the load of a virtual machine (VM) by using Task-based System Load Balancing Method.
3) To optimize the resource utilization by using Resource Aware Scheduling Algorithm (RASA).
4) To compare the results of Hyper-Heuristic Scheduling Algorithm (HHSA) with enhanced proposed scheduling algorithm for makespan time.

B. *Methodology*

1) Hyper-Heuristic

A brand new superior hyper-heuristic formula is planned for programming on cloud computing systems to decrease the create span. The 2 detection operators one for diversity detection and one for improvement detection square measure planned for the planned formula to manage the effectiveness to use the low-level heuristic formula. During this we tend to compared with the low level heuristic algorithms like SLPSO, Max-Min, and FIFO.

2) Task-Based System Load leveling Method

Load balance of a virtual machines is achieved by 1st mapping tasks to VM's so all the VM to host resources, exploitation the Task-Based System Load leveling methodology. This formula ensures the system load leveling through solely transferring additional tasks from associate degree full VM rather than migrating the whole full VM.

3) RASA (Resource-Aware-Scheduling-Algorithm)

Resource utilization is achieved by exploitation RASA a combined approach of Max-Min and Min-Min strategy) to any optimize the resources in terms of accuracy and potency. to realize this, it 1st estimates the completion time and length of the tasks on every of the accessible cloud resources so applies the Max-Min and Min-Min algorithms, as an alternative. Tiny tasks square measure dead by exploitation Min-Min strategy before the big ones. To avoid delays within the execution of enormous tasks to support concurrency within the execution of enormous and little tasks Max-Min strategy is employed.

The planned formula leverages the strength of different 2 low-level programming algorithms i.e. first in first out and Max-Min, whereas not increasing them makespan time, by running one and only 1 low-level formula at every iteration in every cycle. Load of virtual machines square measure balanced through directly transferring the additional tasks from the full VM's rather than migrating the whole overload VM by exploitation the Task based mostly System Load leveling methodology, and to extend the potency and accuracy of resource utilization, Resource-Aware-scheduling formula is employed.

IV. ALGORITHM

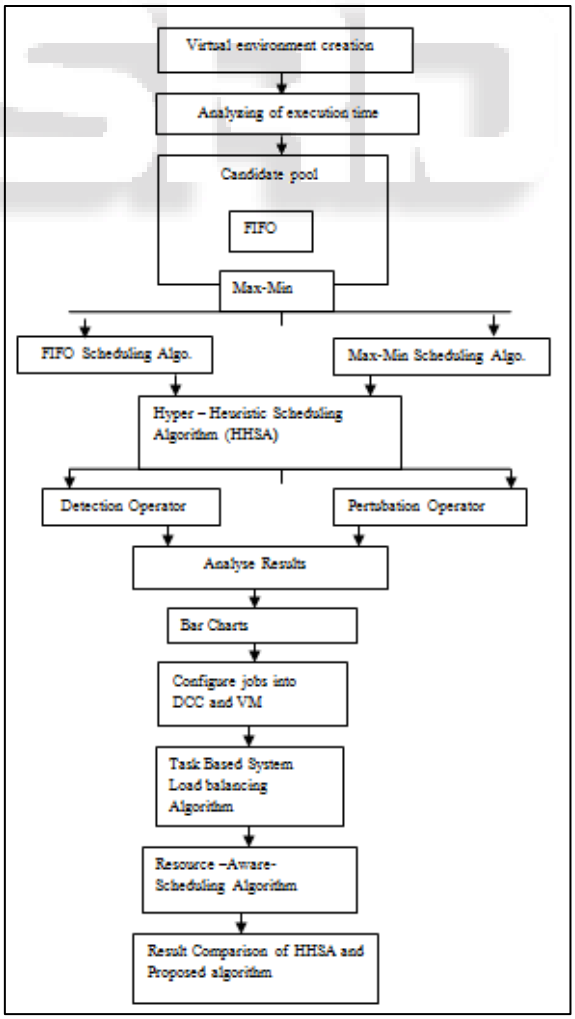

Fig. 1:

V. EXPERIMENTAL RESULT

The proposed method has been implemented in Netbeans IDE and the experimental results have been presented.

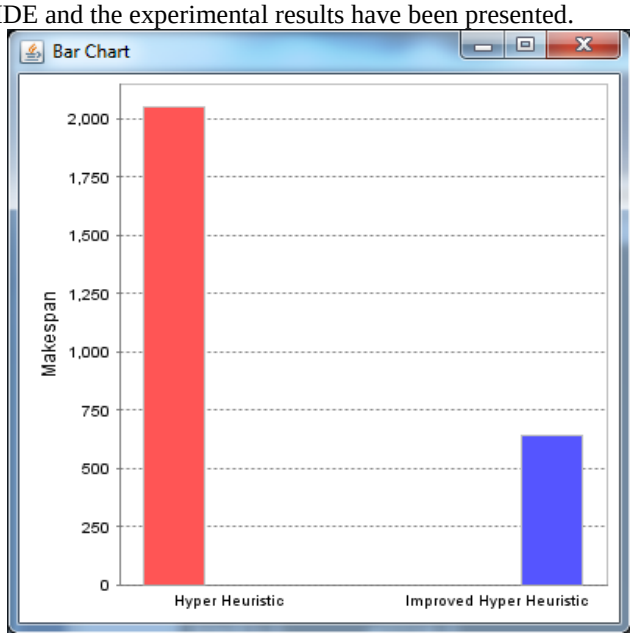

Fig. 2: Makespan time for HHSA(1900ms) and Proposed algorithm (587ms)

Improved hyper heuristic algorithm combines two low-level scheduling algorithms i.e. FIFO and Max-Min to initiate ideal scheduling solutions with minimized calculation time. From the pool of candidate one algorithm is choose as heuristic algorithm. Two operators are utilized i.e. diversity detection operator that automatically figure out which algorithm is picked and pertubation operator to optimize the solutions created by each of these algorithms to further enhance makespan time. It has been observed that makespan time for hyper heuristic algorithm is very high as compared to proposed approach i.e.improved hyper heuristic algorithm. Hence the proposed approach shows better results.

VI. CONCLUSION

Job scheduling is one of the most famous problems in cloud computing so; there is always a chance of modification of previously completed work in this particular field. In this research an effective adjustment of the high –level heuristic Algorithm for a job scheduling problem to minimize the aggregate makespan time of given arrangement is displayed. As conclusion land at concerning the utilization of Hyper-Heuristics, a combined Scheduling technique will prove to generate better results than individual heuristic techniques used for solving problem. The proposed algorithm leverages the strength of other two low-level scheduling algorithms i.e. FIFO and Max-Min, while not increasing the makespan time, by running one and only one low-level algorithm at each iteration in each cycle. Load of virtual machines are balanced through Task Based System Load Balancing Method, and to increase the utilization and efficiency of resources, Resource-Aware-scheduling algorithm is used. The simulation results shows that the proposed Scheduling algorithm reaches its objective: compared with the Hyper-Heuristic Scheduling algorithm, it deeply reduces the total makespan time of jobs, increase the overall performance of the whole system of Cloud computing environment.

REFERENCES

[1] Liang Yu, Tao Jiang, Senior and Yulong Zou ,"Price-Sensitivity Aware Load Balancing for Geographically Distributed Internet Data Centers in Smart Grid Environment," In: IEEE Transactions on cloud computing , Volume: PP, Issue:99, 2016.

[2] Pavithra.B. , Ranjana.R.," Energy Efficient Resource Provisioning with Dynamic VM Placement Using Energy Aware Load Balancer in Cloud ,"In: 2016 International conference on information communication and Embedded systems (ICICES), IEEE,2016.

[3] Xiang Deng, Di Wu, Junfeng Shen and Jian He," Eco-Aware Online Power Management and Load Scheduling for Green Cloud Datacenters, "In: IEEE system journal, Volume:10 , Issue:1, IEEE,2016.

[4] Jia Zhao, Kun Yang, Xiaohui Wei, Yan Ding, Liang Hu, Gaochao Xu," A Heuristic Clustering-based Task Deployment Approach for Load Balancing Using Bayes Theorem in Cloud Environment, "In: IEEE Transactions on Parallel and distributed systems, Volume:27 , Issue:2,IEEE,2016.

[5] Mohammad Reza Mesbahi, Mahnaz Hashemi , Amir Masoud Rahmani," Performance Evaluation and Analysis of Load Balancing Algorithms in Cloud Computing Environments, "In: 2016 Second International Conference on Web Research(ICWR),IEEE,2016.

[6] S.Yakhchi, S.M.Ghafari, M.Yakhchi M.Fazeli, A.Patooghy," ICA-MMT: A Load Balancing Method in Cloud Computing Environment,"In:Web applications and Networking(WSWAN),2nd world symposium,IEEE,2015.

[7] Ariharan v, Sheeja S Manakattu," Neighbour Aware Random Sampling (NARS) algorithm for load balancing in Cloud computing, "In: Electrical Computer and Communication Technologies (ICECCT),IEEE International Conference,IEEE,2015.

[8] Zhiming Zhang, Chan-Ching Hsu, Morris Chang," CoolCloud: A Practical Dynamic Virtual Machine Placement Framework for Energy Aware Data Centers,"In:2015 IEEE 8th International Conference on Cloud Computing ,IEEE,2015.

