A Review of Load Balancing Scenario in Cloud Computing Using Different Heuristic Based Approach

Shatrunjay Panday
M Tech Software Engineering
RKDF Institute Of Science & Technology Bhopal
E-mail:- shatrunjay.panday@gmail.com

Mr. Daya Shankar Pandey
Assistant Professor ,
RKDF Institute Of Science & Technology Bhopal
E-mail:- er_daya@rediffmail.com

ABSTRACT
The limited resource and unlimited number of user always wanted the process of balancing condition. The process of load balancing improves the efficiency of cloud system environment. The process of cloud load balancing used various methods and algorithm for process of balancing. The dynamic load balancing technique used heuristic and queuing based searching and allocation technique for the balancing of job in cloud environment. The various researcher and scientists used dynamic and static load balancing technique. in dynamic load balancing technique used heuristic function such as ACO, PSO, Genetic algorithm and many more guided searching algorithms. In this paper proposed graph based technique for the allocation of job in load balancing process. The graph based technique basically used the concept of sharing of virtual machine. the shared virtual machine allocated the job in dedicated time period for the execution of process.

Keyword: Cloud Computing, Task Scheduling, Resource Allocation, ACO, PSO, GSO.

INTRODUCTION

Cloud Computing is an attracting technology in the field of computer science. Cloud Computing involves sharing of resources. Cloud Computing is made up by aggregating two terms in the field of technology, First term is Cloud and the second term is computing [9]. Cloud is a pool of heterogeneous resources. It is a mesh of huge infrastructure and has no relevance with its name "Cloud". 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. Computation in cloud is done with the aim to achieve maximum resource utilization with higher availability at minimized cost. Cloud is a collection of heterogeneous resources and requirements of these resources can change dynamically [8, 13].

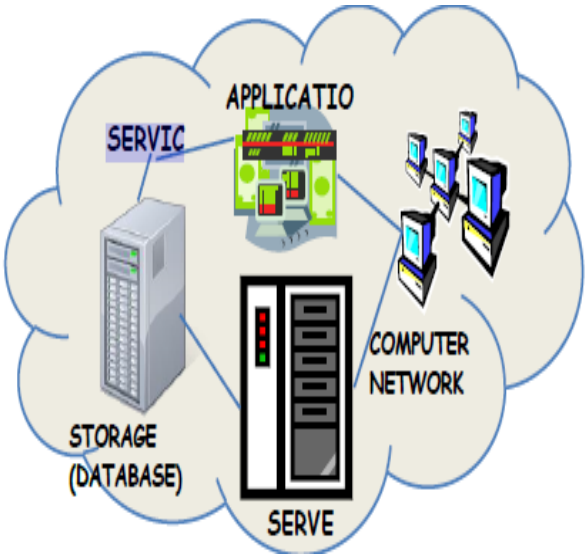

Figure 1: Basic structure of a cloud.

Cloud providers are always interested in maximizing the resources utilization and the associated revenues, by trimming down energy consumption and operational expenses, while on the other hand cloud users are interested in minimizing response time and optimizing overall application throughput [2]. The cloud offers computing facilities on-demand, served over a network from a shared pool of resources, typically residing in a large data center. Cloud computing offers a new approach to the way applications are written and deployed. One of the primary benefits of cloud computing is elasticity, whereby resources can be automatically expanded during demand spikes, or contracted during demand lulls; thereby improving user experience and system utilization [6].

Currently Cloud computing is an emerging concept which efficiently combining many fields of computing. Cloud computing provides us various services and software and processing capacity over the internet [8]. The biggest advantage of cloud computing is that it is capable of handling a huge amount of growing work in a predestined manner for the usage of the business customers. Here in cloud computing the main concept for cloud computing is virtualization which

generalizes the physical infrastructure and makes it easy to use and managed [10].

II SERVICE MODELS

A service model determines the types of computer resources offered to consumers. Three main types of cloud services are software (SaaS), platform (PaaS) and infrastructure (IaaS) covers a huge range of services. However, new service models are continuously emerging [7].

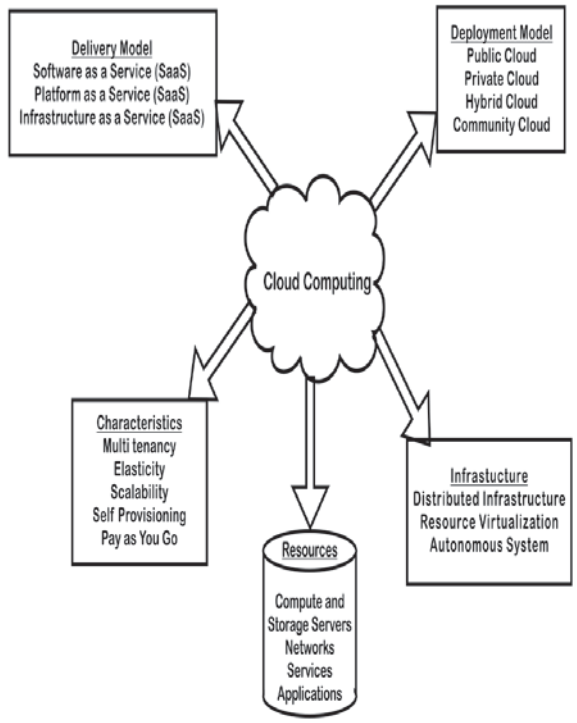

Figure 2: Cloud Computing Framework.

SOFTWARE AS A SERVICE: SAAS is a new model of how software is delivered. SaaS refers to software that is accessed via a web browser and is paid on a subscription basis (monthly or yearly). Standard application software functionality is offered within a cloud. Example: Google Apps (Google Docs), SalesForce.com are among the providers of this kind of cloud computing [4].

PLATFORM AS A SERVICE: PaaS offerings facilitate deployment of applications without the cost and complexity of buying and managing the underlying hardware and software and provisioning hosting capabilities, providing all of the facilities required to support the complete life cycle of building and delivering web applications and services entirely available from the Internet. PaaS offerings may include facilities for application design, application development, testing, deployment and hosting as well as application services such as team collaboration, web service integration and marshalling, database integration, security, scalability, storage, persistence, state management, application versioning, application instrumentation and developer community facilitation. These services may be provisioned as an integrated solution over the web. Some PaaS providers are Microsoft Azure, Bungee Connect, WorkXpress, and Force.com for Google App Engine, OrangeScape, among many others [3].

INFRASTRUCTURE AS A SERVICE: Cloud computing provides different services to different users such as SaaS, IaaS, and PaaS. Users can avail these services in a Pay per-Use-On-Demand model, which can access shared IT resources like server, data storage, application, network and so on through internet. IaaS offers storage and compute resources that developers and IT organizations use to deliver custom business solutions. The customer accesses those services with defined interfaces. In cloud computing the underlying infrastructure that provides the service may be very sophisticated indeed. However, the user doesn't necessarily need to understand this infrastructure to use it. The IaaS customer rents computing resources instead of buying and installing them in their own data center. Infrastructure as a Service is one of the example of cloud computing.

III TASK SCHEDULING

As the number of users of Cloud computing Systems increased, the tasks to be scheduled in Cloud increased proportionally. Therefore, there is a need for better algorithms to schedule tasks on these systems. Algorithms required to schedule tasks are service oriented and differ in different environments. Task Scheduling algorithms in cloud computing aim at minimizing the make-span of tasks with minimum resources efficiently. Cloud computing, uses low-power hosts to achieve high usability. The cloud computing refers to a class of systems and applications that employ distributed resources to perform a function in a decentralized manner. Cloud computing is to utilize the computing resources (service nodes) on the network to facilitate the execution of complicated tasks that require large-scale computation. Thus, the selecting nodes for executing a task in the cloud computing must be considered. A task is an activity that uses set of inputs to produce a set of outputs. In Cloud computing, user applications will run on virtual systems where distributed resources are allocated dynamically. Dynamic load-balancing mechanism has to allocate tasks to the processors dynamically as they arrive. Redistribution of tasks has to take place when some processors are overloaded. Every application is completely different in nature and independent where some require more CPU time to compute complex task, and some others may need more memory to store data. Different scheduling algorithms can be used depending on the type of the task to be scheduled. The scheduling algorithms can utilize better executing efficiency and maintain the load balancing of system. The efficiency of the cloud depends on the algorithms used for task scheduling.

The performance of cloud depends on the task scheduling algorithms used in the scheduling components or brokering components. Scheduling parallel applications modeled by Directed Acyclic Graphs onto a network of heterogeneous computers is a NP-Complete problem. A number of algorithms have been proposed in the past to solve the task-scheduling problem for heterogeneous network of computers. However, none of these algorithms can be extended to cloud

computing systems which are also heterogeneous computing systems. Since cloud computing systems have a high degree of unpredictability with respect to resource availability and network bandwidth, task scheduling algorithms for cloud computing systems should incorporate the latency caused by unpredictable resource availability. The present study involves surveying the different task scheduling algorithms developed for cloud environment.

IV RELATED WORK

[1] In CC environment Several parameters are available in literature for measuring efficiency of a load balancing algorithm. 1. Throughput: it must ensure increased throughput at minimal expense. If a load balancing algorithm doesn't increase system throughput, it defeats its own purpose. 2. Reliability: it must be reliable, since process failure while transferring job from one location to another may lead to increased waiting time and customer have no satisfaction. 3. Adaptability: it must be capable of adapting the dynamically changing user requests and gives facilities of task allocation in minimal amount of time. 4. Fault Tolerance: it must ensure fault tolerance, here in case of a problem complete load balancing mechanism does not stop working in the system. 5. Waiting Time: it should reduce wait time of a task for allocation of resources to it. Cloud Computing revolves around internet based acquisition and release of resources from a data center.

[2] In this research author discuss on Cloud Computing they deploy groups of remote servers and software networks that allow centralized data storage and online access to computer services or resources. Load balancing has become one of the key issues with the rapid growth of web traffic and the services available under cloud environments. It is the job of Load Balancer to distribute the load among different virtual machines so that all the nodes get equally loaded. To attain the maximum throughput and minimum time various researchers throughout the world proposed many algorithms and approaches.

[3] Author here discuss on a new Bee Swarm optimization algorithm called Bees Life Algorithm (BLA) applied to efficiently schedule computation jobs among processing resources onto the cloud datacenters. It is considered as NP-Complete problem and it aims at spreading the workloads among the processing resources in an optimal fashion to reduce the total execution time of jobs and then, to improve the effectiveness of the whole cloud computing services. BLA has been inspired by bees' life in nature represented in their most important behaviors which are reproduction and food source searching. In order to evaluate the effectiveness and the performance of this proposal, a set of experimental tests has been conducted.

[4] In this paper author present Cloud computing has been considered as the new computing paradigm that would offer computer resources over the Internet as service. With the widespread use of cloud, computing would become another utility similar to electricity, water, gas and telephony where the customer would be paying only for the services consumed contrary to the current practice of paying a monthly or annual fixed charge irrespective of use. For cloud computing to become accepted by everybody, several issues need to be resolved. One of the most important issues to be addressed is cloud security. Trust management is one of the important components of cloud security that requires special attention.

[5] Author in this paper define the Although smartphones are increasingly becoming more and more powerful, enabling pervasiveness is severely hindered by the resource limitations of mobile devices. The combination of social interactions and mobile devices in the form of 'crowd computing' has the potential to surpass these limitations. In this paper, they introduce Honeybee a crowd computing framework for mobile devices. Honeybee enables mobile devices to share work, utilize local resources and human collaboration in the mobile context. It employs 'work stealing' to effectively load balance tasks across nodes that are a priori unknown. They describe the design of Honeybee, and report initial experimental data from applications implemented using Honeybee.

[6] Author discuss here the cloud computing load balancing is required to achieve evenly distribute load among the nodes and to efficiently make use of the resources Load balancing ensures that all the processor in the system or every node in the network does approximately the equal amount of work at any instant of time. This technique can be sender initiated, receiver initiated or symmetric type (combination of sender initiated and receiver initiated types). This paper presents the features and discussed about the pros and cons of various load balancing algorithm in the authors style. Various consideration of the algorithm like throughput, performance, fault tolerance, migration time, response time etc has been discussed.

[7] In this research, they define the honey bee foraging mechanism for load balancing is improved by random stealing technique. For finding the state of a virtual machine, the deviation of virtual machine load is calculated and checked for confinement within a threshold condition set. With random stealing method, tasks are stolen from a random Virtual machine when a VM is idle. It thus saves the idle time of the processing elements in the Virtual machine. The performance evaluation is done by using cloud-Sim. Simulation results show that the modified honey bee foraging technique with random stealing reduces the make span of algorithm execution together with balancing system load and reduces the idle time of Virtual machine.

[8] In order to provide valuable information and influence the decision-making process of a load balancer, thus maintaining optimal load balancing in hosted (or cloud) environments, it is not enough just to provide information from networking part of the computer system or from external load balancer. Load balancing models and algorithms proposed in the literature or applied in open-source or commercial load balancers rely either on session-switching at the application layer, packet-switching mode at the network layer or processor load balancing mode. The analysis of detected issues for those load balancing algorithms is

presented in this paper, as a preparation phase for a new load balancing model (algorithm) proposition.

V PROBLEM STATEMENT

The process of load balancing and task scheduling impart a major role in success of cloud computing. In review process we found that various factor effects the performance of cloud computational in concerns of storage of network data and sharing of resource. The sharing of cloud resource generates a network overload, the network overload arises the problem of bandwidth and stack overflow. Some problem related to cloud computing is given below.

1. Increasing the time span for process [11]
2. Failure of resource allocation [12]
3. Traffic overhead of network [13]
4. Waste of resource [4]
5. Cost of monitoring of resource [5]
6. Process feedback system [6]

For the minimization of this entire problem, used heuristic function by various author in load balancing in cloud computing

VI PROPOSED APPROACH

For load balancing and task scheduling of cloud computing used meta-heuristic function concerned with the design of intelligent multi-agent systems by taking inspiration from the collective behaviors of social insects and other animal societies. They are characterized by a decentralized way of working that mimics the behavior of the swarm. The optimization needs a level of solution approach of multi-criteria of problem formulation. Some optimization of resource method discuss here. Genetic Algorithms or Genetic algorithms are search algorithms based on the mechanism of natural selection, better known by the evolutionary process [8]. In the process of evolution, the individual is continuously changing genes to adapt to the environment of his life. Only individuals who are strong can survive.

The main idea of GSO is to model a problem as the search for a minimum cost path in a graph. Artificial glowworm as if walks on this graph, looking for cheaper paths. Each ant has a rather simple behavior capable of finding relatively costlier paths. Cheaper paths are found as the emergent result of the global cooperation among ants in the colony. The behavior of artificial ants is inspired from real ants: they lay pheromone trails (obviously in a mathematical form) on the graph edges and choose their path with respect to probabilities that depend on neighbor's swarm [5]. For the improvement of the system performance in cloud computing environment used queue based routing protocol for time quantum theory for the allocation of job.

VII CONCLUSION AND FUTURE WORK

The efficiency of cloud based infrastructure based on load balancing factor of different component of cloud computing. The load balancer plays a major role in cloud based services. The cloud based services interact with user and dedicated cloud infrastructure. The interaction of user and cloud operation request to transfer the load to virtual machine and other resources. For the improvement of load efficiency in cloud computing various researcher and cloud designer used swarm based job and task scheduling technique. The swarm based task scheduling technique is very efficient in compassion of old and traditional technique such and FCFS and round robin technique. In this paper, we have study a various research paper based on task scheduling in cloud computing environment.

REFERENCES:-

[1] Gaochao Xu, Junjie Pang and Xiaodong Fu "A Load Balancing Model Based on Cloud Partitioning for the Public Cloud", TSINGHUA SCIENCE AND TECHNOLOGY, 2013, Pp 34-39.

[2] Siva Theja Maguluri and R.Srikan "Scheduling Jobs With Unknown Duration in Clouds", IEEE, 2014, Pp 1938-1951.

[3] Yuan Zhang, Xiaoming Fu and K. K. Ramakrishnan "Fine-Grained Multi-Resource Scheduling in Cloud Datacenters", IEEE, 2014, Pp 1-6.

[4] V. Suresh Kumar and M. Aramudhan "TRUST BASED RESOURCE SELECTION AND LIST SCHEDULING IN CLOUD COMPUTING", IJAET, 2014, Pp 2455-2463.

[5] Hassan Takabi, James B. D. Joshi and Gail-Joon aHn "Security and Privacy Challenges in Cloud Computing Environments", IEEE, 2010, Pp 24-31.

[6] Mayanka Katyal and Atul Mishra "A Comparative Study of Load Balancing Algorithms in Cloud Computing Environment", International Journal of Distributed and Cloud Computing, 2013, Pp 5-14.

[7] Mate J. Csorba, Hein Meling and Poul E. Heegaard "Ant System for Service Deployment in Private and Public Clouds", BADS'10, 2010, Pp 1-10.

[8] Ye Huang, Nik Bessis, Peter Norrington, Pierre Kuonen and Beat Hirsbrunner "Exploring decentralized dynamic scheduling for grids and clouds using the community-aware scheduling algorithm", Elsevier, 2011, Pp 1-14.

[9] Parveen Patel, Deepak Bansal, Lihua Yuan, Ashwin Murthy, Albert Greenberg, David A. Maltz, Randy Kern, Hemant Kumar, Marios Zikos, Hongyu Wu, Changhoon Kim and Naveen Karri "Ananta: Cloud Scale Load Balancing", SIGCOMM, 2013, Pp 207-218.

[10] Klaithem Al Nuaimi, Nader Mohamed, Mariam Al Nuaimi and Jameela Al-Jaroodi "A Survey of Load Balancing in Cloud Computing: Challenges and Algorithms", IEEE, 2012, Pp 137-142.

[11] Linh T.X. Phan, Zhuoyao Zhang, Qi Zheng, Boon Thau Loo and Insup Lee "An Empirical Analysis of Scheduling Techniques for Real-Time Cloud-Based Data Processing", IEEE, 2011, Pp 3-10.

[12] Vijayalakshmi A. Lepakshi and Dr. Prashanth C S R "A Study on Task Scheduling Algorithms in Cloud Computing",

International Journal of Engineering and Innovative Technology, 2013, Pp 119-125.

[13] Azizkhan F Pathan and S. B. Mallikarjuna "A Load Balancing Model Based on Cloud Partitioning for the Public Cloud", International Journal of Information & Computation Technology, 2014, Pp 1605-1610.

[14] Joseph Doyle, Robert Shorten and Donal O'Mahony "Stratus: Load Balancing the Cloud for Carbon Emissions Control", IEEE, 2014, Pp 1-14.