ISSN:0975-9646
Manjunatha Swamy C et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (5) , 2015, 4357-4360

Dynamic Resource Scheduling In Cloud Computing Using Genetic Algorithm: A Survey

Manjunatha Swamy C[#1], Anil Kumar B H[*2], Kiran B[#3]
[#]*Dept. of Computer Science and Engineering, ATME College of Engineering, Mysore, India*
[*]*Dept. of Computer Science and Engineering, ATME College of Engineering, Mysore, India*

Abstract **- Resource scheduling and allocation is very critical issue in cloud computing. Any computation is carried out when there is sufficient or proportionate resources available. Services are provided to the customers or end users with the proper analysis of resources. Cloud computing plays a vital role is a model for enabling ubiquitous network access to a shared pool of configurable computing resources. Any cloud provides services mainly three ways software as a service (SaaS), platform as a service (PaaS) and infrastructure as a service (IaaS). Infrastructure as a service in cloud grabs much attention in Cloud computing. To utilize resources more efficiently an optimized scheduling algorithm is used to address cloud scheduling problems. By deploying virtual machines in appropriate locations to improve the speed of locating best allocation method which intern permit maximum utilization of resources available. The scheduling resources to overcome unbalance in assignment problem, in this parallel genetic algorithm concept is used which is much faster than traditional genetic algorithm [1].**

Keywords **- Cloud computing; Resource Scheduling; VRaaS (virtual resource as a service); VMware; Parallel genetic algorithm; Assignment problems.**

I. INTRODUCTION

The cloud computing has great potential of providing robust computational power to the society at reduced cost. The dynamic resource scheduling model for a public cloud which has numerous nodes with distributed computing environments with many different geographic locations.

To be competitive, corporations must minimize inefficiencies and maximize productivity. In manufacturing, productivity is inherently linked to how well you can optimize the resources you have, reduce waste and increase efficiency. Finding the best way to maximize efficiency in a manufacturing process can be extremely complex. Even on simple projects, there are multiple inputs, multiple steps, many constraints and limited resources. In general a resource constrained scheduling problem consists of: A set of jobs that must be executed, finite set of resources that can be used to complete each job, associated with set of constraints that must be satisfied. Constraints basically of two types in Temporal Constraints the time window to complete the task, Procedural Constraints is the order each task must be completed and Resource Constraints is the resource available with set of objectives to evaluate the scheduling performance.

Clouds can be used to provide on-demand capacity as a utility, although the realization of this idea can differ among various cloud providers, the most flexible approach is the provisioning of virtualized resources as a service (VRaaS). Cloud typically includes a large amount of virtual and physical servers and efficiently managing this virtual infrastructure becomes a major concern. VMware is a tool running as a application in the user area exist above the kernel) Virtual Center, Platform Orchestrator, have emerged to manage virtual infrastructures, providing a centralized control platform for the automatic deployment and monitoring of virtual machines (VMs) in resource pools. By this can decide where to place exactly the virtual machines. Open-source cloud systems like Eucalyptus, Open Nebula and Nimbus these to decide the allocation of resources, Eucalyptus uses Greedy (first fit) and Round robin algorithm [4] [5], Open Nebula uses a ranking algorithm [6], and Nimbus uses some customizable tools like PBS and SGE [7].But none of the above strategies is both automated scheduling and at the same time considering the maximum usage of resources.

Genetic Algorithm is a good choice to deal with scheduling problem, the scheduling policies which used by GA improved the utilization rate of resources when VMs are allocated. Genetic Algorithms (GA) is powerful search techniques that are used to solve difficult problems in many disciplines. Parallel Genetic Algorithm (PGA) is parallel implementations of GAs which can provide considerable access in terms of performance and scalability. PGAs can easily be implemented on networks of heterogeneous computers or on parallel mainframes. PGA is schedule the resources in the cloud more effectively and efficiently along with the speed of finding the best allocation sequence has been obtained with fast convergence.

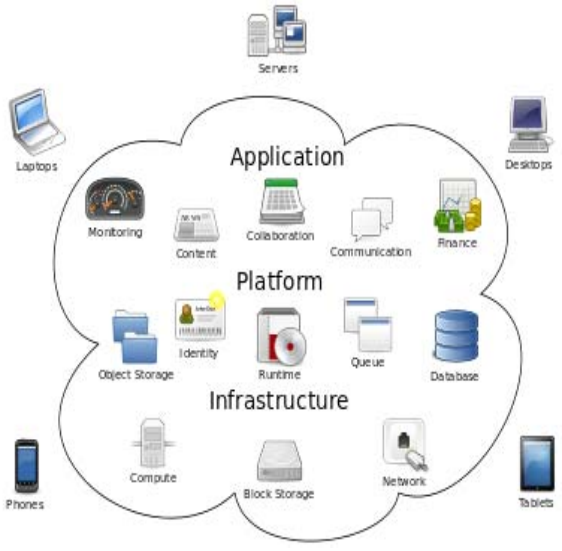

Fig. 1 Cloud Computing Environment

Resource as a service (RaaS) cloud in this customers will purchase "seed virtual machines" containing a baseline of resources as well as an "economic agent" for adding additional capacity. The agent will make decisions based on the current prices of those resources, the changing load the machine should handle and the client's subjective valuation of those different resources at different points in time."

Cloud service providers software would also incorporate economic agents to represent their own interests. Customer agents could also negotiate with those controlled by other customers, who may have extra resource to sell.

Pricing for various resource types would be market-driven according to supply and demand. This also be applied to SLAs (service level agreements), when supply is insufficient for serving all clients the provider can starve clients with lower-priority SLAs when only 90% availability) by raising the price of resources."

Making the RaaS cloud concept a reality will require new types of applications and system software, the researcher's state. In RaaS cloud, virtual machines never know the precise amount of resources that will be available to them at any given time. This requires software running in those virtual machines to adapt to changing resource availability and exploit whatever resources the software is consist of and provide if available.

Virtual machine ware (VMware) works based on the virtualization concept which act as one cloud, any application and any device. As a application runs in user space installed above the operating systems. Virtualization is process of creating exact replica of physical bare hardware which is not real but those who accessing the VM are in illusion that he has an access to original hardware. Method of dividing available system resources logically to service different applications which is some time hardware virtualization or software virtualization.

VMware software provides a completely virtualized set of hardware to the guest operating system. VMware software virtualizes the hardware for a video adapter, a network adapter, and hard disk adapters. The host provides pass-through drivers for guest USB, serial, and parallel devices. In this way, VMware virtual machines become highly portable between computers, because every host looks nearly identical to the guest. In practice, a system administrator can pause operations on a virtual machine guest, move or copy that guest to another physical computer, and there resume execution exactly at the point of suspension. Alternatively, for enterprise servers, a feature called vMotion allows the migration of operational guest virtual machines between similar but separate hardware hosts sharing the same storage (or, with vMotionHYPERLINK "https://en.wikipedia.org/w/index.php?title=VMotion_Storage&action=edit&redlink=1" Storage, separate storage can be used, too). Each of these transitions is completely transparent to any users on the virtual machine at the time it is being migrated.

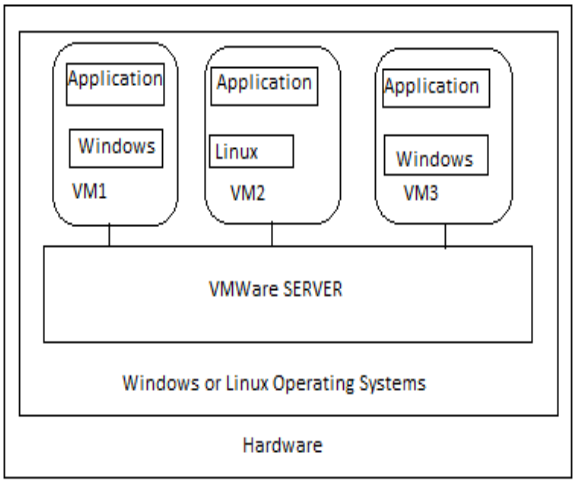

Fig. 2 Architecture of VMWare

II. PROBLEM DEFINITION

Schedulers used in PGA are unique and separate this can access information's on all requests received by the core, based on the particular request which keep tracks current allocations, creating and updating a resource schedule and sending the appropriate deployment commands to the cloud Core. There is request is referred as instance requests (IRs) and m idle or partially idle computing nodes available in the cloud. At this stage the main problem is to find the allocation sequence which makes the utilization rate of the resource to be achieved as maximum. VMs are allocated on physical resources according to a ranking algorithm even though which fails to achieve higher VMs utilization rate. To achieve an optimal VMs allocation sequence, GA seems to be a good choice, but it can be very demanding in terms of computation load and memory. With this Parallel Genetic Algorithm (PGA) has been promoted, the parallel implementations of GA, which can provide considerable gains in terms of performance and scalability. PGA can easily be implemented to address resource scheduling problem.

A. Model

In this a scheduler is available at each node in the cloud. Mainly three activities are carried out in the scheduling. At first the system maintains the list of idle resources, list of virtual machines availability will be updated all at regular intervals of time and each time new VM requests may come or VMs are in shutdown state or any change in the physical resources are being detected. Then use the PGA to find the optimal allocation sequence pattern. Later allocate the specified physical machines to the VM requests. Several important factors we must consider in the process of solving unbalanced assignment problem with our PGA. In assigning the requested resources need to consider different genetic algorithms parameters such as chromosome representation, suitable fitness function design and use of proper migration method.

Chromosome representation is most important feature of genetic algorithm. In this system approach encoding scheme is used. First, proper label allocated sequentially to the IRs and mark each of the CPU cores, memory size and hard disk size be related, e.g. 1.(1 core,512M, 10G); 2.(2 cores, 1000M, 20G). There is computing node each one is numbered in order and the size of idle resources of each node is should be known and monitored. The allocation is done to each computing node in parallel by representing an integer and arrange from 1 to m. If the chromosome pattern is (8, 2, 4) it means the first IR is to place at the node in which the number is 8, the second IR is place to the node which the number is 2 and similar way the third IR is place at the node which the number is 4.

B. Operators Used in PGA

In parallel genetic algorithm different parameters need to be set. The migration of individuals from one node to another is controlled by several parameters [8].

- The topology or structure that defines the connections between the subpopulations, commonly used topologies include hypercube, 2D or 3D mesh topology. The rate of migration that controls how many individuals going to migrate. Migration scheme, that controls which individuals from the source node, which may be best or worst or random is migrate to another node then which individuals are going to be replaced (worst, random, etc.)
- Migration interval that determines the frequency of migrations, the parallel model we use here is coarse-grained model and utilize one-way ring topology to connect subpopulations. In figure.2 migration of n number of information request form first node to sub nodes in the tree. Based on the availability of resources interns of idle nodes will be selected for allocation and to service the requests, every ten generations migration once with a migration rate of 10% and the optimal individuals emigrate, worst individuals are replaced.

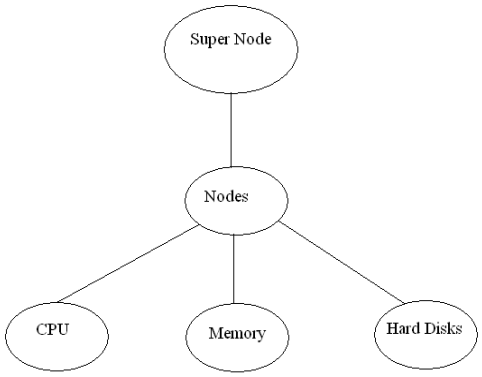

Fig. 3 Schematic view of modules in the system

In this system considering one computing node which launch one instance at a scheduling time. Here one job will be allocated for one time; one physical machine can run many virtual machines and gathering all the idle resources of each computing node. Each idle resource is appearing at the scheduling time except the nodes that don't have any free resources to provide. For an instance system which is four cores computing node that half of its CPUs is in idle. We will take the rest part of its CPUs is that will be considered at the scheduling time.

III. FITNESS FUNCTION

One of the flexibility in genetic algorithm is fitness function. In this system CPU the number of cores, memory capacity and disk capacity is considered to achieve maximum utilization of resources. Java genetic algorithm package can be used which is a combination of genetic algorithm and genetic programming component as a framework that can be easily used in evolutionary principles. GA is so powerful technology available to solve problems with a huge solution space having limited time and CPU power.

Java genetic algorithm package (JGAP) framework which also supports multiple threads concept to implement and tested on windows XP. Virtual machines in the system offer different services to the users by accepting request as Information request (IR). When the request arrives at the system, scheduler will updates the list of idle resources and assign the request to particular node, nodes associated with active server will provide services. The speed of the parallel genetic algorithm is high with genetic algorithm in finding best allocation scheme and utilization rate factor for PGA is better than GA. Due to multiple thread concepts in PGA performance will be more acceptable with GA.

IV. CONCLUSION

By using genetic algorithm to address scheduling of resources problem in cloud this system associated with parallel genetic algorithm (PGA) to achieve the optimization and sometime sub-optimization for cloud scheduling problems. Genetic operators like mutation, cross over etc are deployed in the system to obtain desired performance with efficiency in resources allocation. Representing scheduling problem as unbalanced assignment problem with server node concept and proper scheduler to list the idle nodes to assign incoming request for the specific node to perform useful task. Compared to traditional GA, this parallel genetic algorithm improved the speed of locating the best allocation. Utilization rate of the resources is also increases with precision.

V. FUTURE WORK

System can be designed and deployed with dynamic load balancing concept in cloud environment; Bigdata concept on cloud is another approach to consider, cloud environment consists of numerous nodes to be monitored properly to move the incoming request dynamically to the idle or normal nodes. Number of requests by the users will be satisfied by reassign the incoming request to other nodes to process when previous node becomes overloaded. Any nodes in the cloud are in three states as Idle, busy, normal. Using proper well defined algorithm can obtain scalability and efficiency in providing unbreakable services.

ACKNOWLEDGEMENT

I would like to thank Prof. M. S. Veerendra Kumar, HOD, Dept. of CS&E, ATME, Mysore and faculty members for

their valuable suggestions and comments throughout of this work.

REFERENCES

[1] I. W. Flockhart and N. J. Radcliffe, "A genetic algorithm-based approach to data mining," in Proc. 2nd Int. Conf. Knowledge Discovery Data Mining (KDD-96). Portland, OR, Aug. 2–4, 1996, page 299.

[2] B. Adler, Load balancing in the cloud: Tools, tips and techniques, http://www.rightscale. Com / info center/ whitepapers/ Load-Balancing-in-the-Cloud.pdf, 2012.

[3] R. Hunter, The why of cloud, http://www.gartner.com/ DisplayDocument?doc cd=226469&ref= g noreg, 2012.

[4] Eucalyptus,http://www.eucalyptus.com.

[5] Daniel Nurmi, Rich Wolski, Chris Grzegorczyk, Graziano Obertelli, Sunil Soman, Lamia Youseff, and Dmitrii Zagorodnov. 2009. "The Eucalyptus Open-Source Cloud-Computing System," In Proceedings of the 2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID 09). IEEE Computer Society, Washington, DC, USA,pp.124-131. doi:10.1109/CCGRID.2009.93.

[6] OpenNebular,http://www.opennebula.org.

[7] Nimbus, http://nimbusproject.org.

[8] Nowostawski, M. Poli,R,, "Parallel Genetic Algorithm Taxonomy," Knowledge-Based Intelligent Information Engineering Systems, 1999. Third International Conference, Dec 1999, pp.88 – 92,doi:10.1109/KES.1999.820127.