

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 1, Issue 8, October 2013

## A Survey on Dynamic Optimization of Multi attributes Resource Allocation in Self-Organizing Clouds

Richu Jimmy<sup>1</sup>, S. Vanitha<sup>2</sup>, Daniel .D<sup>3</sup>

PG Scholar, Dept. of CSE, Dr. N.G.P Institute of Technology. Coimbatore, India<sup>1</sup>

Assistant Professor, Dept. of CSE Dr. N.G.P Institute of Technology. Coimbatore, India<sup>2</sup>

Assistant Professor, Dept. of CSE Dr. N.G.P Institute of Technology. Coimbatore, India<sup>3</sup>

**ABSTRACT:** Virtual Machine is a software implementation of a computing environment in which operating system can be installed & run. By combining the concept of volunteer computing with cloud architectures, we can build self organizing clouds (SOC). It gathers unused resources over internet. Here each participant can act as both resource provider and requester. Fully distributed, Virtual Machine multiplexing resource allocation scheme is used to manage decentralized resources. Proportional share model aims at providing maximized resource utilization and optimal execution efficiency. Pointer-Gossiping CAN (PG-CAN) is used to locate qualified nodes with minimum contention. Existing solutions produces heavy weight query messages per tasks. These messages will utilize large amount of resources for computation. The above algorithms along with PG-CAN set up are used in order to produce light weight queries per task in our SOC. This set up can produce an improvement in system output by 15-60 percent than peer to peer model which was used earlier.

**Keywords :** Cloud computing, Virtual Machine multiplexing resource allocation, SOC, CAN, PG-CAN .

### I. INTRODUCTION

Cloud computing is an internet based computing where infinite numbers of computers are coupled to a cloud server. It has the potential to transform large part of IT industry making software even more attractive as a service and has the capability to change, the way hardware is designed. Its main objective is to gain the efficiency of the shared resources. Cloud resources are shared by multiple users. It can be reallocated as per demand.

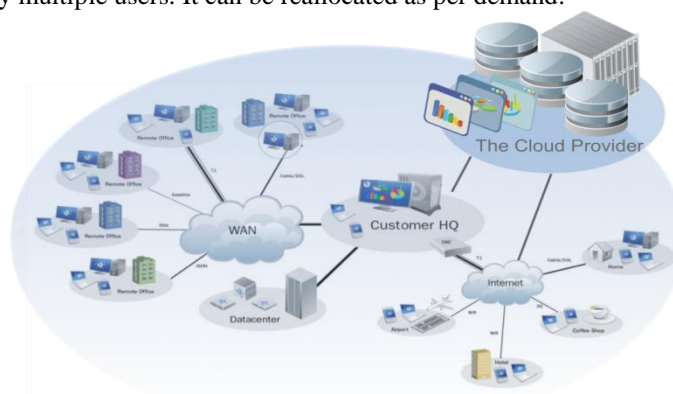


Figure 1: Architecture of Cloud Computing

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 1, Issue 8, October 2013

Cloud computing consists of three elementary service models: 1) platform as a service, 2) infrastructure as a service 3) software as a service.

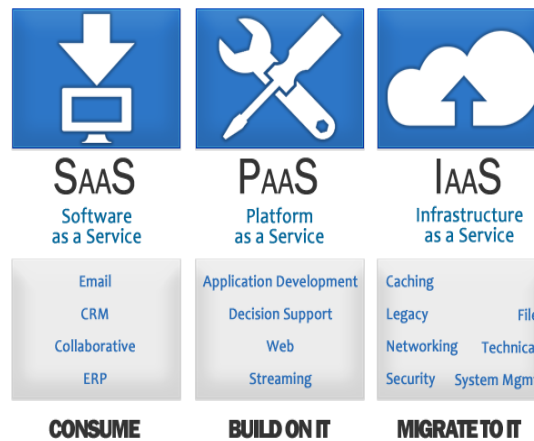


Figure 2: Service Models

Software as a Service helps the application to be provided as a service for the customers across the internet. Here the cloud provider builds in and operates the application software in cloud and this could be used by the customers according to their needs. It reduces the burden of software maintenance and support. Some of its examples are human resources, financial planning etc. Platform as a Service offers services like storage and database. Some of its advantages are security, reliability, scalability etc. Google Application Engine is an example of Platform as a Service .Infrastructure as a Service offers services like VM disk image library, firewalls, IP Address, VLANs etc.

There are four main types of deployment models in cloud computing: 1) private cloud 2) public cloud 3) community cloud and 4) hybrid cloud [1]. In public cloud, resources can be shared between organizations for maximum cost efficiency. Data confidentiality and security will be very high in private clouds. In community clouds, infrastructure is shared between several organizations with common interest. Hybrid cloud is a combination of at least one private cloud and one public cloud. Government Organizations is an example of community cloud. Some of the characteristics of cloud computing are security, virtualization, reliability, performance, device and location independence etc.

Most of these Cloud platform architectures support simple keyword- based requests and can't response complex queries efficiently .It occurs mainly due to the lack of efficiency in multi-attribute index techniques. Existing Cloud platform providers therefore improved the performance of indexing techniques for complex requests. Multi-attribute indexing schemes had been proposed for Cloud platform to manage the huge and variety data to address the complex queries efficiently. Main issues in cloud computing are Security, performance, Data and reliability. Here I am focussing on mainly performance issues in cloud computing.

The rest of the paper is organized as follows: Chapter 2 explains the related works, Chapter 3 gives a detailed presentation about the optimization of multi attribute resources in SOC. Chapter 4 gives the performance evaluation and chapter 5 tells conclusion.

## II. RELATED WORKS

Cloud Computing, is one of the most trending technologies in today's world. It has the capability to transform a large part of the IT industry which makes the software more attractive a service. Today Developers do not require the



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 1, Issue 8, October 2013

large capital to operate their tasks. Large numbers of computers are connected by a high speed network. Users access the sites provided by the cloud environment through public Internet. Cloud server has the capability to provide both applications and hardware systems. Here applications are provided as a service. Utility Computing is a type of computing which packages different computer resources and sells them according to users demand. It will be preferable for them to run a Private Cloud. Cloud computing use the concept of pay-for-use licensing model to match the users demand. Infrastructures Software is running on VMs. Examples of Cloud Computing vendors are Amazon Web Services, Microsoft Azure, Google Application Engine etc. They provide Computation model (VM), Storage model and Networking model.

Some of the obstacles that had been faced during the growth cloud computing are availability of service, data confidentiality, performance unpredictability etc. Lack of availability of services can be solved by using multiple cloud providers. Data confidentiality can be tackled by the method of encryption and firewalls. Performance Unpredictability can be reduced by VM Support and Flash Memory. Software licensing problem can be avoided by using pay- for- use licensing model. There are various factors which really influences these companies to become cloud providers. Some of the reasons are that it enables them to earn lot of money easily. Another reason is that it provides maximum advantage for their investment.

Some of the latest Technology trends that had been described in this paper are that Amazon Web Services began to provide pay-as-you-go computing with no contract i.e. all customers need is a credit card facility only. A second development was selling hardware-level virtual machines, granting consumers to prefer their own software requirement without disturbing each other while sharing the same hardware and thereby reducing costs to much extend [1].

The Content Addressable Network (CAN) is a distributed peer to peer infrastructure which provides Hash Table Functionality on Internet. CAN node maintains routing Table which holds IP Address and virtual coordinate zones of its neighbouring nodes. When a node routes a message towards its destination point it first determine which neighbouring node is closest to its destination point. Then it will search for that particular node's IP Address through routing table [13,15]. Hash Table plays a major role in today's software systems. It maps key onto values. The CAN is scalable, fault-tolerant and completely self-organized. The best examples of current Internet systems that could be improved by CAN are peer to-peer file sharing systems such as Napster and Gnutella. In these systems, files are stored at the peers and are transferred directly between them. This paper mainly focuses on the design of CAN. The basic operations performed on a CAN are the insertion, lookup and deletion of (key, value) pairs [6, 9 and 11].

CAN consist of numerous individual nodes. Each node stores a chunk (called a *zone*) of the entire hash table and also holds information about neighboring zones in the table. Requests for a particular key are routed by intermediate CAN towards the CAN node whose zone contains it. Our CAN design is completely distributed and are scalable and fault tolerant too. This design is implemented completely at the application level [9, 12].

The three main features of this design are: CAN routing, CAN construction, CAN maintenance etc. CAN node maintain coordinate routing table which contains IP address and virtual coordinate zone of its neighbors in the coordinate space. Here routing takes place in a straight line path from source to destination in Cartesian space. The whole CAN is divided into nodes. CAN grows simultaneously when a new node is joined to the system. This can be implemented by the process of node splitting. It consists of three processes:

1. The new node must find a node already in the CAN.
2. With the help of CAN routing mechanisms, it must find a node whose zone can be split.
3. The neighbours of this zone must be notified so that routing process can include the new node.

When a node leaves the CAN, its zone must be occupied by its neighboring nodes. If it forms a valid single zone, then it is fine. Otherwise we have to reconsider other neighboring nodes with less zone space. The CAN also needs to be strong from node or network failures. During this situation one or more nodes become unreachable. This can be solved by using takeover algorithm. The delayed absence of an update message from that particular neighboring node indicates its failure. Path length, latency, hash table availability are some of the parameters used to evaluate system performance.



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 1, Issue 8, October 2013

Some of the design improvements described in this paper are multidimensional co-ordinate space, multiple hash functions [2].

The difficulties in allocating shared resources among multiple users can be realized by using proportional-share market mechanism. Here each user divides his budget and receives a small part of required resources relative to his bid. We use resource allocation game under this mechanism to revise the efficiency and fairness of the equilibrium in this game. Usually, the sharing is completed through first come first serve and round-robin time sharing policies. However it does not have capability to express the priority of their tasks [18, 20].

Another method for allocating resources to systems is to use market-based approach. In this approach, users put forward bids for resources, and are owed according to some pre-defined resource allocation mechanisms. We explore the efficiency and fairness of the proportional-share allocation mechanism, by considering the linear utility function model. Utility function is weighted sum of tiny proportions, users receive from resources. Each user tries to achieve most of utility functions by dividing the budget cleverly. Main results that had been provided in this paper are survival of Nash Equilibria, design of algorithms to amplify utility, simulation results for efficiency and fairness. Best response algorithm and local greedy adjustment algorithm helps to increase the utility of corresponding users [3].

Virtualization plays a major role in cloud computing. At the beginning itself, each VM is provided with specific amount of computing resources. Resource provisioning relates to allocating resources to VMs to perform their corresponding tasks. Efficient resource provisioning can be attained by two processes: (1) Static resource provisioning. (2) Dynamic resource provisioning.

In static resource provisioning, VMs are created with precise size and are build up onto a set of physical servers. Here VM capacity cannot be modified .But comparing with the case of dynamic resource provisioning, VM capacity can be tuned in accordance to workload variations. Static provisioning is implemented at basic stage of capacity planning [17, 18]. VM sizing is the most fundamental step in both of these resource provisioning. It specifies evaluation of the amount of resources that have to be allocated to VM. Some of the popular methods were VM-by-VM basis and Joint-VM provisioning method. In VM-by-VM basis, each VM has expected size based on workload pattern. Multiple VMs are joined together to form joint VM. It is based on assessment of combined capacity needs. Main results of this paper are: SLA (service level agreement) model that maps application routine requirements to resource claim requirements, VM selection algorithm that explores to find those VMs with the majority consistent demand patterns. It will provide high capacity investments etc. [4]

### III. OPTIMIZATION IN SELF ORGANIZING CLOUDS

Cloud computing focus on allocation of resources to match the needs of users demand over the network. Most of the cloud architectures, which are built on top of a centralized architecture, may suffer from several problems like denial-of-service (DoS) attacks, limited availability of computational resources etc. So we use the concept of self organizing clouds. They are mainly used to gather the unused resources over the internet. Here each participants act as both resource requestor and resource provider. They can locate nodes with required resources.

Here two issues are considered: 1) Multi attribute range query problem. 2) Optimization of task's execution problem. We propose dynamic optimal proportional-share (DOPS) by combining it with proportional share model. To locate the qualified nodes in SOC, we use pointer gossiping CAN. They are mainly used to locate the qualified nodes with minimum contention. Therefore in order to establish this setup, we have to first create cloud environment [21]. Here large number of computers is connected by a high speed network. Users can access the sites hosted by the cloud environment. Multiple cloud users can request number of cloud services simultaneously [4, 17].

When a task is submitted with expected resource vector and weight vector of that submitted task we have to find out two major issues: 1) How to locate a qualified nodes for executing their task 2) How to optimize task's execution time. Optimal resource allocation takes place if it satisfies two conditions: 1) the product of price vector and expected



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 1, Issue 8, October 2013

resource vector should be less than the budget. 2) Expected resource vector should be less than or equal to the available resource vector. Optimal solutions must satisfy Karush-Kuhn-Tucker (KKT) conditions [5, 3]. In order to solve the simultaneous equations and inequalities two traditional methods were used: 1) brute-force method 2) Interior point method. Brute Force method is randomly selecting an input to get an output. No specific criteria will be needed for selecting them. Interior point method is the process of collecting information and trying to get required output on its basis.

Main aim of Optimal Proportional-share Allocation is to redistribute the available resources which are already used by the running tasks. It consists of two main procedures: 1) Slice Handler 2) Event Handler. In slice handler, resources are reallocated periodically, so that running task can attain additional resources according to their demand. In Event Handler, resources are redistributed on the basis of certain criteria's such as task arrival and task completion. Here even if additional resources are there, they cannot be re allocated till the end of task completion.

Pointer-Gossiping CAN helps to allocate the task to free c.p.u with minimum contention. Here we use content addressable network (CAN) as the distributed hash table (DHT) overlay to provide multidimensional feature. It consists of two algorithms: 1) Pointer-sender: When the extra resources are free in a particular node, it will be propagated in backward direction to notify the presence of resources to other nodes. 2) Pointer-Relay: It can be explained as reply from the nodes explained above i.e. the reply from these nodes in downward direction to other nodes.

## IV. PERFORMANCE EVALUATION

Here all the proposed approaches are evaluated to compare the performance of all approaches. We analyse our proposed scheme in terms of simulation time and total allocation. Based on the comparison and the results from the experiment we can state that proposed approach works better than the other existing system. Proportional-share scheduler is used to multiplex c.p.u resources among virtual machines. It will be implemented according to Xen's credit-scheduler. We use PeerSim to implement PG-CAN Setup. Hardware Configuration of each node is chosen without conscious decision, according to system parameters like number of nodes, number of processors per node, computation rate per processor, I/O speed per node, memory size per node etc. we can derive min\_ capacity and max \_capacity at each resource dimension.

We will compare execution efficiency of SOC to P2P desktop Grid to compare the efficiency of DOPS algorithm. We also consider different types of resource demand such as CPU, disk speed, network, memory size and disk space. Later we will compare PG-CAN with hopping manner and spreading manner to our existing solutions. The final output is that we can produce efficient resource allocation and optimal execution time.

## V. CONCLUSION

Main achievements that can be developed are: Task's resources are designated according to limits of user's budget, Maximum resource application, Usage of light weight queries, less chance of contention etc. My future work is to adaptively schedule multi attribute resources in self organizing clouds with a suitable algorithm so that any number of tasks can be resolved.

## REFERENCES

- [1] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwinski, G. Lee, D.A. Patterson, A. Rabkin, I.Stoica, and M. Zaharia, "Above the Clouds: A Berkeley View of Cloud Computing," Technical Report UCB/EECS-2009-28, pp. 6-25, Feb. 2009.
- [2] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker, "A Scalable Content-Addressable Network," Proc. ACM SIGCOMM'01, pp. 161-172, 2001.
- [3] M. Feldman, K. Lai, and L. Zhang, "The Proportional-Share Allocation Market for Computational Resources," IEEE Trans.Parallel and Distributed Systems, vol. 20, no. 8, pp. 1075-1088, Aug.2009.



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 1, Issue 8, October 2013

- [4] X. Meng et al., "Efficient Resource Provisioning in Compute Clouds via vm Multiplexing," Proc. IEEE Seventh Int'l Conf. Autonomic Computing (ICAC '10), pp. 11-20, 2010.
- [5] D.P. Anderson, "Boinc: A System for Public-Resource Computing and Storage," Proc. IEEE/ACM Fifth Int'l Workshop Grid Computing, pp. 4-10, 2004.
- [6] S. Govindan, J. Choi, B. Urgaonkar, A. Sivasubramaniam, and A. Baldini, "Statistical Profiling- Based Techniques for Effective Power Provisioning in Data Centers," Proc. Fourth ACM Conf. European Conf. Computer Systems (EuroSys '09), pp. 317-330, 2009.
- [7] S. Soltesz, H. Poetzl, M.E. Fluczynski, A. Bavier, and L. Peterson, "Container-Based Operating System Virtualization: A Scalable, High-Performance Alternative to Hypervisors," Proc. Second ACM Int'l European Conf. Computer Systems (Euro '07), pp.275-287, 2007.
- [8] J.S. Kim et al., "Using Content-Addressable Network for Load Balancing in Desktop Grids," Proc. 16th ACM Int'l Symp. High Performance Distributed Computing (HPDC '07), pp. 189-198, 2007.
- [9] Y. Drougas and V. Kalogeraki, "A Fair Resource Allocation Algorithm for Peer-to-Peer Overlays," Proc. IEEE INFOCOM '05, pp. 2853-2858, 2005.
- [10] J. Cao, F.B. Liu, and C.Z. Xu, "P2pgrid: Integrating P2P Networks Into the Grid Environment: Research Articles," vol. 19, no. 7, pp. 1023-1046, 2007.
- [11] P. Switalski and F. Seredynski, "Generalized Extremal Optimization for Solving Multiprocessor Task Scheduling Problem," Proc. Seventh Int'l Conf. Simulated Evolution and Learning, pp. 161-169, 2008.
- [12] G. Singh, C. Kesselman, and E. Deelman, "A Provisioning Model and its Comparison with Best-Effort for Performance-Cost Optimization in Grids," Proc. 16th ACM Symp. High Performance Distributed Computing (HPDC '07), pp.117-126, 2007.
- [13] C.B. Lee and A.E. Snavey, "Precise and Realistic Utility Functions for User-Centric Performance Analysis of Schedulers," Proc. 16th ACM Int'l Symp. High Performance Distributed Computing (HPDC '07), pp. 107-116, 2007.
- [14] D. Li, J. Cao, X. Lu, and K.C.C. Chen, "Efficient Range Query Processing in Peer-to-Peer Systems," IEEE Trans. Knowledge and Data Eng., vol. 21, no. 1, pp. 78-91, Jan.2009.
- [15] J. Wang, S. Wu, H. Gao, J. Li, and B.C. Ooi, "Indexing Multi- Dimensional Data in a Cloud System," Proc. ACM Int'l Conf. Management of Data (SIGMOD '10), pp. 591-602, 2010.
- [16] Q. Zheng, H. Yang, and Y. Sun, "How to Avoid Herd: A Novel Stochastic Algorithm in Grid Scheduling," Proc. 15th ACM Int'l Symp. High Performance Distributed Computing (HPDC '06), pp. 267-278, 2006.

## BIOGRAPHY



### Richu Jimmy

She received the B.Tech degree in Computer Science & Engineering from College Of Engineering, Adoor, Kerala in 2012. She is currently with the Post graduate in Dr. N.G.P Institute of Technology now works on the research project in parallel and distributed Systems and cloud computing.



### S. Vanitha

She had received her B.E degree in Computer Science and Engineering from Vivekanandha college of Engineering for women in 2005. She has completed her M.E degree in Software Engineering from Anna University, Regional centre-Coimbatore. She is currently working as Assistant Professor in the Department of Computer Science and Engineering, Dr. N.G.P Institute of Technology, Coimbatore, India. She is the lifelong member of ISTE.



### D. Daniel

He received the B.E degree in Information Technology from Karunya University in 2009. He Completed his Post graduate in Karunya University and now works as assistant professor at Dr.NGP Institute of Technology.