

Dynamic Allocation of Resources in Self - Organizing Cloud with Fault-Tolerance

Mohini Gupta
PG Scholar,

Dept. of Computer Science and Engg,
SRM University, Chennai, India

Mrs. G Abirami
Assistant Professor,

Dept. of Computer Science and Engg,
SRM University, Chennai, India

ABSTRACT - Cloud computing, with its great potentials in low cost and on-demand services, is a promising computing platform. Cloud computing offers dynamic flexible resource allocation for reliable and guaranteed services in pay-as-you-use manner. Self-organizing cloud (SOC) is decentralized approach, uses Gossiping algorithm to know the status of neighboring nodes. For distributing load, it uses light-weight query on content addressable network with low contention. Proposed system uses the fault tolerance support for a SOC system. In order to enhance the reliability of a system fault tolerance support is needed. The main challenge is to detect and tolerate the fault in a virtual environment where all the participating resources are also virtual. Proposed method uses a deadline-driven algorithm for calculating the upper bound of task i.e. maximum time needed to execute a task. If a task is completed within the approximate upper bound then it is stated that no fault has been occurred. In occurrence of a fault the task will not complete under the dead line.

Keywords: Cloud computing, fault tolerance, reliability.

I. INTRODUCTION

Cloud Computing is a type of computing that relies on sharing computing resources rather than having local servers or personal devices to handle applications. In cloud computing, the word cloud is used as a metaphor for “the internet” so the phrase cloud computing means “A type of internet based computing” where different services – such as servers, storage and applications are delivered to an organization’s computers and devices through internet [10]. Generally it consists of a bunch of distributed servers known as masters, providing demanded services and resources to different clients known as clients in a network with scalability and reliability of datacenter (Fig. 1). The distributed computers provide on-demand services.

A. Types of Cloud

Public Cloud - Public clouds are owned and operated by companies that use them to offer rapid access to affordable computing resources to other organizations or individuals [7]. With public cloud services, users don’t need to purchase hardware, software or supporting infrastructure, which is owned and managed by providers. The public cloud is defined as a multi-tenant environment, where

client buy computing environment that is shared with a number of other clients or tenants.

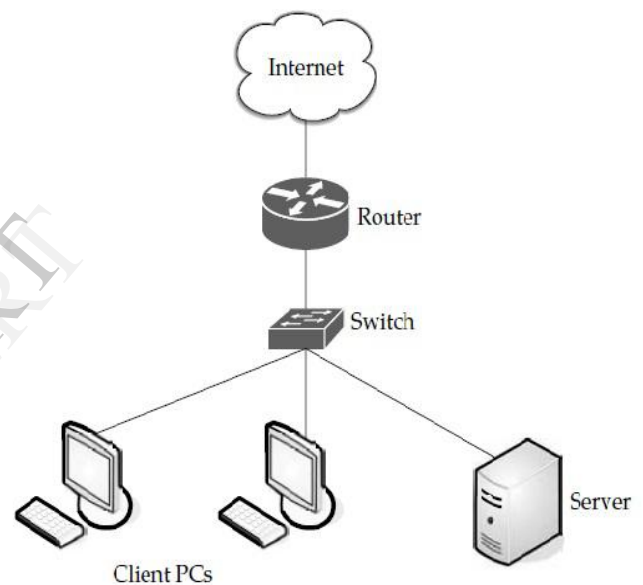


Fig 1: Cloud Computing

Private Cloud - A private cloud is owned and operated by a single company that controls the way virtualized resources and automated services are customized and used by various lines of business and constituent groups [7]. A private cloud is what you need when you have projects high-security requirements or performance sensitive applications.

The use of a private cloud can change how organizational and trust boundaries are defined and applied. The actual administration of a private cloud environment may be carried out by internal or outsourced staff.

Hybrid Cloud - A hybrid cloud uses a private cloud foundation combined with the strategic use of public cloud services [7]. The reality is a private cloud can’t exist in isolation from the rest of a company’s IT resources and the public cloud. Most companies with private clouds will

evolve to manage workloads across data centers, private clouds and public clouds—thereby creating hybrid clouds.

Community Cloud - A community cloud is shared among two or more organizations that have similar cloud requirements. In other word, a community cloud is similar to a public cloud except that its access is limited to a specific community of cloud consumers. The community cloud may be jointly owned by the community members or by a third-party cloud provider that provisions a public cloud with limited access. The member cloud consumers of the community typically share the responsibility for defining and evolving the community cloud

The cloud isn't a technology. It's more of an approach to building IT services - an approach that harnesses the power of servers, as well as virtualization technologies that combine servers into large computing pools and divide single servers into multiple virtual machines. And there are several different deployment models for implementing cloud technology.

B. Types of Services

Everyone is talking about cloud computing today, but not everyone means the same thing when they do. While there is this general idea behind the cloud - that applications or other business functions exist somewhere away from the business itself - there are much iteration that companies look to in order to actually use the technology. Services may be of software resources (e.g. Software as a Service, SaaS) or physical resources (e.g. Platform as a Service, PaaS) or hardware/ infrastructure (e.g. Hardware as a Service, HaaS or Infrastructure as a Service, IaaS)(Fig.2).

SaaS - Cloud-based applications—or software as a service (SaaS)—run on distant computers “in the cloud” that are owned and operated by others and that connect to users’ computers via the Internet and, usually, a web browser.

SaaS is a process of software delivery that permits data to be accessed from any device with an Internet connection and web browser. In this web-based model, software vendors host and maintain the servers, databases and code that constitute an application.

PaaS - Platform as a service provides a cloud-based environment with everything required to support the complete life cycle of building and delivering web-based (cloud) applications—without the cost and complexity of buying and managing the underlying hardware, software, provisioning and hosting. The PaaS delivery model represents a pre-defined ready-to-use environment typically comprised of already deployed and configured resources.

IaaS - Infrastructure as a service provides companies with computing resources including servers, networking, storage, and data centre space on a pay-per-use basis. In an IaaS agreement, the subscriber completely outsources the

storage and resources, such as hardware and software that they need.

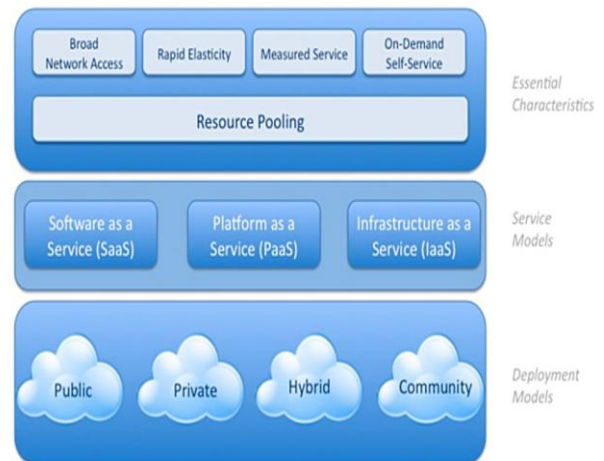


Fig 2: Cloud Computing Services

Cloud computing is changing the mode, how and where the computing is going to be demonstrate. It has earned lot of attention to be used as a computing model for various applications. But the people are still not eager to use it for real time applications [3]. But now some researchers and cloud vendors are working to provide the power of cloud computing and related benefits to the real time applications. A few cloud operators have started real time cloud support. Self-organizing cloud (SOC), which can connect a large number of computers on the Internet by a P2P connection. In SOC environment, each participating node acts as both a resource provider and a resource consumer [1]. Self-organization is a process where some form of global order or coordination arises out of the local interactions between the components of an initially disordered system [9]. This process is spontaneous: it is not directed or controlled by any agent or subsystem inside or outside of the system

In cloud computing, Resource Allocation is the process of assigning available resources to the needed cloud applications over the internet. Resource allocation starves services if the allocation is not managed precisely [4, 7]. Resource provisioning solves that problem by allowing the service providers to manage the resources for each individual module. Computer systems may fail due to hardware and software faults Use of cloud infrastructure for real time applications increases the chances of errors. To build better fault-tolerant distributed applications that can adapt to constant changes in environments and user requirements, it is necessary to separate fault-tolerant computing policies and mechanisms from application programs. Many real time systems are also safety Critical systems, so they require a higher level of fault tolerance.

II. RELATED WORK

The Cloud computing is an on demand technology because it offers dynamic and multifaceted resource allocation for reliable and guaranteed services in pay as-you-use manner to users. It is a technology that permits users to use applications without installation and access their personal files at any computer with the assistance of internet access. Because of the flexible nature of cloud computing, we can quickly access resources from cloud providers when we needed. To gain the maximum benefits, it is required to allocate the resources optimally to the task. However, it is important and complicated to allocate the resource to all the tasks in cloud computing (Fig. 3). For resource allocation many policies have been introduced [12]. Resources can be allocated on the basis of task execution time, task behavior, on the basis of auction, on the basis of SLA [12]. In this paper, FCFS algorithm is used.

The use of cloud infrastructure for real time application is quite new. Most of the real time applications need to provide the fault tolerance support [2, 3]. A lot of work has been done in the area of fault tolerance for real time systems. But there is lot of research work available in fault tolerance of real time application running on cloud environment. With cloud computing services, it is more efficient to realize resource discovery, resource matching, task scheduling and execution.

III. PROPOSED MODEL

A. RESOURCE ALLOCATION MODEL

In this paper, Self-organizing clouds use a cloud server that will be resource provider and consumer both [1]. The SOC server will initialize the resources for use and it is also the consumer of that resources. After the initializing of resources it is available for the other user. In this paper, resources are allocated dynamically on the basis of arrival of tasks; as soon as the tasks are arrived the resources are allocated to them, if it is available on "First come first serve basis" (Fig. 4). If the resources are not available then task or process has to wait for resources. Cloud server first initializes the resources in order to use by the tasks. If the resources are not initialized then tasks are not able to use them, because everything is under the control of the SOC cloud server. In cloud environment resources are nothing but memory storage, CPU, processor speed, network bandwidth etc [6]. Users request for the memory to store the data related to task likewise CPU is used for execution of the task.

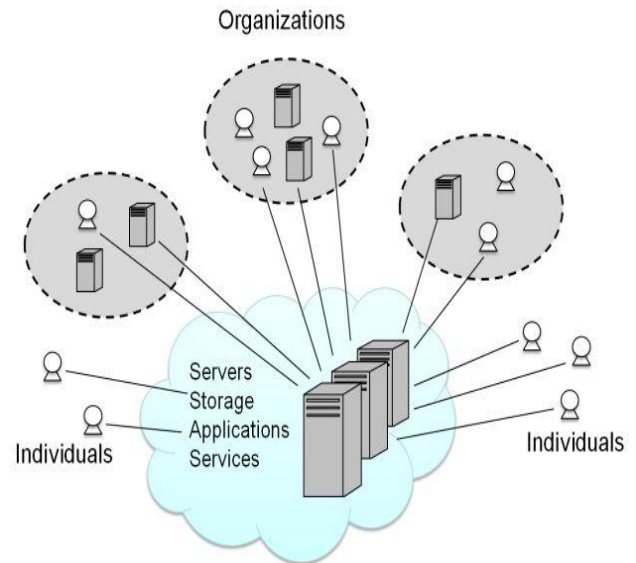


Fig 3: Resource allocation in cloud

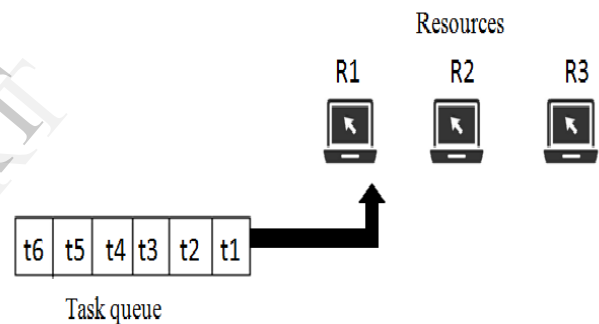


Fig 4: Resource allocation using FCFS

B. TYPES OF FAULTS IN COMPUTING ENVIRONMENT

Fault tolerance is a major concern to guarantee availability and reliability of critical services as well as application execution. Fault tolerance is one of the key issues amongst all. It is concerned with all the techniques necessary to enable a system to tolerate software faults remaining in the system after its development. Based on fault tolerance policies various fault tolerance techniques can be used [13].

1. Reactive fault tolerance

Reactive fault tolerance policies reduce the effect of failures on application execution when the failure effectively occurs [13]. It means after the occurrence of fault, it reduces the impact of fault on the system.

2. Proactive Fault Tolerance

The principle of proactive fault tolerance policies

is to avoid recovery from faults, errors and failures by predicting them and proactively replace the suspected factors by other working components [13].

In this paper, the reactive fault tolerance policy is used. The resource completes the execution of task and tolerates the faults.

C. LOAD BALANCING

Distributing processing and communications activity evenly across a computer network so that no single device is overwhelmed. Load balancing is especially important for networks where it's difficult to predict the number of requests that will be issued to a server. In SOC environment, Load balancing aims to optimize resource use, maximize throughput, minimize response time, and avoid overload of any one of the resources [10]. Using multiple components with load balancing instead of a single component may increase reliability through redundancy. Cloud computing is a vast concept. Many of the algorithms for load balancing in cloud computing have been proposed [11]. The common approach for balancing load between the resources is to use a centralized *Load balancer* [5]. But SOC is a decentralized approach, uses *Pointer – gossiping Content addressable network* (PG-CAN) for getting the status of neighboring resources whether they are idle or in use. In load balancing technique, the tasks are allocated to the resources according to their CPU speed, memory etc. If resources suppose R1 and R2 is available for executing a task and R2 has more processor speed than R1 then task will be allocated to R2. In overloaded condition SOC distribute their load to other resources. If one resource is having more loads then it is given to other under loaded resource. The resources distribute load by gossiping to each other and qualified node will be used for the task execution.

D. FAULT TOLERANCE MECHANISM

Fault tolerance is the ability of a system to respond gracefully to an unexpected hardware or software failure. In general, a failure represents the condition in which the system deviates from fulfilling its intended functionality or the expected behavior. A failure happens due to an error; that is, due to reaching an invalid system state. The proposed method uses Deadline-Driven Approach for tolerating the faults. In cloud environment the task must be completed within time. In the competitive situation, majority of tasks can be guaranteed to be completed within deadlines. In this approach the upper bound [8], the maximum time for execution of a task is calculated, e.g. A user wants to upload a file and the size of the file is 5 MB and network bandwidth at time of uploading the file is 1Mbps then the time taken by the task to complete can be calculated, the task will take 5 second to upload on network. First, the start time and end time of execution of a task is calculated based on some attributes like file size etc. This upper bound is the estimated time for the task to

complete and it may be less than the actual time. If the task takes more time than the estimated time i.e. upper bound then the system must have some faults. In the condition, when fault occur the task is split in subtask and then it will be executed.

III. PERFORMANCE ANALYSIS

Cloud computing has come out as an important paradigm for accessing distributed computing resources. In addition, the performance of the resources for each task is analyzed and come up with a performance measurement chart (Fig. 5). This analysis is done on the basis of usage of the resources. Suppose R1 is used by more tasks i.e. R1 executes more task than R2 then its performance will be more. This will show how many tasks are executed by the resource. This chart will vary according to usage of the resources. In cloud computing, performance of resources like memory, CPU, Disk, network can be analyzed, for this some specific tools can be used.

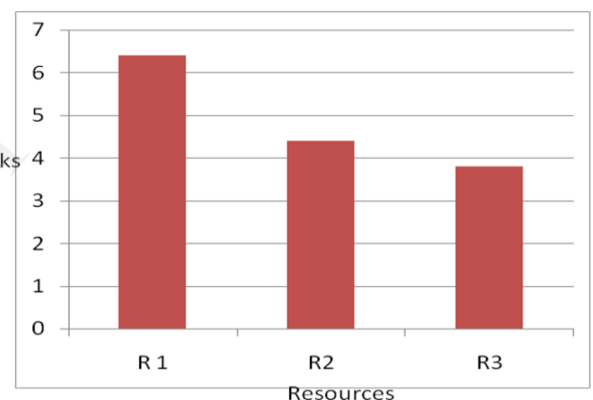


Fig 5: Performance analysis

V. CONCLUSION

In this paper, we propose a resource allocation in SOC environment that will balance the load on resources and detect the fault in the system. In addition we will calculate the upper bound of the task i.e. the maximum execution time for the task. This will be the expected time. When the resources allocated to task are sufficient, we can guarantee task's execution time always within its upper bound. If the task is taking more time than the expected deadline then there must be some fault. The performance chart of the resources will show how much a resource is utilized.

The fault tolerance mechanism in a system provide reliability and reliable system minimize overall payment of the system. So, this paper mainly focused on the reliability of the SOC system.

REFERENCES

1. Sheng Di, Member, IEEE, and Cho-Li Wang, "Dynamic Optimization of Multiattribute Resource Allocation in Self-Organizing Clouds", IEEE Transaction on parallel and distributed systems , Vol. 24, No. 3, March 2013.
2. Pardeep Kumar, Shiv Kumar Gupta, "Abstract Model of Fault Tolerance Algorithm in Cloud Computing Communication Networks", International Journal on Computer Science and Engineering (IJCSE) .
3. Sheheryar Malik, Fabrice Huet, "Adaptive Fault Tolerance in Real Time Cloud Computing", IEEE World Congress on Services, 2011.
4. Anshul Rai, Ranjita Bhagwan, Saikat Guha , "Generalized Resource Allocation for the Cloud".
5. N.Chandrakala, Dr. P.Sivaprakasam, "Analysis of Fault Tolerance Approaches in Dynamic Cloud Computing", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 2, February 2013.
6. K. Rasmi and V. Vivek, "Resource Management Techniques in Cloud Environment - A Brief Survey", International Journal of Innovation and Applied Studies ISSN 2028-9324 Vol. 2 No. 4 Apr. 2013.
7. Chandrashekhar S. Pawar and R.B.Wagh, "A review of resource allocation policies in cloud computing" , World Journal of Science and Technology 2012 .
8. Sheng Di, Member, Cho-Li Wang "Error-Tolerant Resource Allocation and Payment Minimization for Cloud System", IEEE Transaction on parallel and distributed systems , Vol. 24, No. 6, June 2013.
9. <http://en.wikipedia.org/wiki/Self-organization>
10. http://www.webopedia.com/TERM/C/cloud_computing.html
11. Hsiao, Hung-Chang, Chung, Hsueh-Yi, Shen, Haiying , Chao, Yu-Chang, "Load Rebalancing for Distributed File Systems in Clouds", Volume:24, May 2013
12. V.Vinothina, Dr.R.Sridaran, Dr.Padmavathi Ganapathi, "A Survey on Resource Allocation Strategies in Cloud Computing", International Journal of Advanced Computer Science and Applications, Vol. 3, No.6, 2012
13. Anju Bala, Inderveer Chana, "Fault Tolerance- Challenges, Techniques and Implementation in Cloud Computing", IJCSI, Vol. 9, Issue 1, No 1, January 2012