A Novel Approach to Enhance the Performance of Cloud Computing Using Load Balancing in File Systems

Pradheep M [1],Anjandeep Kaur Rai [2] ,Anup Parkash Singh [3]

[1, 2] *Post-Graduate Students, Department of Information Technology,Lovely Professional University, Punjab, India*
[3] *Assistant professor, Department of Computer Science and Engineering, Lovely Professional University, India˙*

Abstract: Cloud computing is an emerging technology. It uses the different computing and storage services over the Internet. In cloud computing, the internet is viewed as a cloud. File system backup is a particularly attractive application for outsourcing to the cloud because it is relatively simple, the growth of disk capacity relative to tape capacity has created efficiency and cost inflection point, and the cloud offers easy off-site storage, always a key concern for load balancing. In this paper, we are going to enhance the performance of file system. For this purpose, we can integrate the load balancing and round robin algorithm.

Keywords: Cloud Computing, Security, File System, Schedulers.

I. INTRODUCTION

Cloud Computing is the biggest technology advancement now a days. It has taken computing in initial to the next level. Cloud computing provides the information technology as a service. Cloud computing uses the internet and the central remote servers to support different data and applications. It is an internet based technology. It allows the users to find their personal files at any computer with internet access. Cloud computing is flexible in nature. It allocates the resources on the authority request. [2] Cloud computing provides the act of uniting. It is an emerging technology, that is used to provide various computing and storage services over the Internet. In cloud computing, the internet is viewed as a cloud. By the use of cloud computing, the capital and operational costs can be cut.
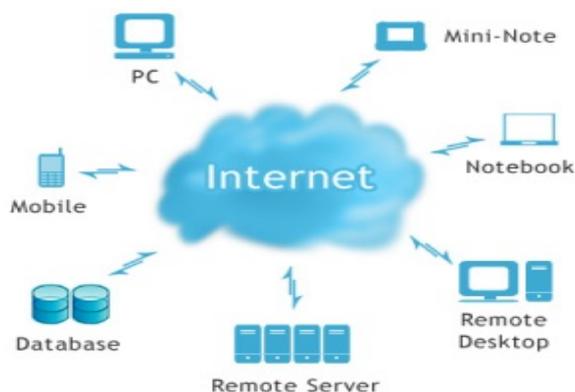
Figure 1: A cloud is used in network diagrams to depict the Internet [1]

Uses of Cloud Computing:

In the older days every company was to license their software through CDs DVDs and when it was to come on upgrading, they were to face lots of problems. When cloud computing comes as a service part like rental the cost of supplying and vendor system could be reduced, where the software comes to any organization directly. Cloud computing incorporates the infrastructure, platform, and software as services. These service providers rent data center hardware and software to deliver storage and computing services through the Internet. Internet users can receive services from a cloud as if they were employing a super computer which be using cloud computing. To storing data in the cloud instead of on their own devices and it making ubiquitous data access possible. They can run their applications on much more powerful cloud computing platforms with software deployed in the cloud which mitigating the users burden of full software installation and continual upgrade on their local devices.

II. VIRTUALIZATION:

In the cloud computing, clouds are divided into three parts, as, Public Clouds, Private Clouds and the Hybrid Clouds. Virtualization is very useful concept in context of cloud systems. Virtualization means something that is not real. [3] Virtualization is a software implementation of computer. It helps to execute different programs like a real machine. Virtualization is related to cloud computing, because it can be used by end users. The end users can use various services of the cloud. Virtualization can be classified in two different parts:

- Full virtualization
- Partial virtualization

Full Virtualization

In case of full virtualization, the complete installation of one machine is done on another machine. It will result in a virtual machine which will have all the software that are presents in the actual server. The full virtualization is used for many purposes.

- Sharing a computer system among different users
- It helps to Isolates users from each other and from the control program
- Emulating hardware on another machine
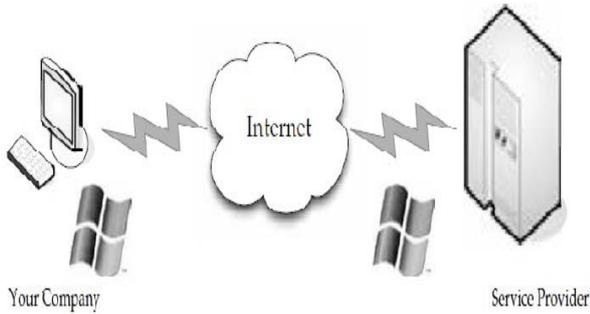
Figure 2: Full virtualization

Partial virtualization

In partial virtualization, the hardware allows multiple operating systems to run on single machine by efficient use of system resources such as memory and processor. In this case, the services are not fully available. These services are provided partially.

It has the following advantages:

- **Disaster recovery:** It helps in the disaster recovery.
- **Migration:** in the virtualization, hardware can be replaced easily, hence migration of different parts of a new machine is faster.
- **Capacity management:** In a virtualized environment, it is easier and faster to add more hard drive capacity and processing power.
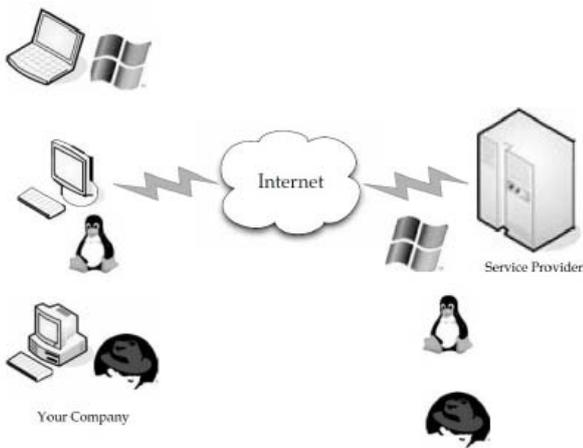

Figure 3: Partial virtualization

File System in cloud computing:

Now days, the growth of communication and the technologies has transformed the way we live and work. The cloud computing is an emerging technology. In the cloud computing storage area is used. For storing the information in storage area various files are required. To maintain these files file systems are used. These files systems include:

- **Hadoop Distributed File System:** these file systems deploys in large scale distributed system. These file system includes facebook, google, yahoo and so on.

These file systems use a name node to keep a list of all files in the cloud and their respective metadata. In this the name node has to manage all the files related operations, these files related operation includes open, copy, move, delete, update, etc.

- **Network File System:** It is a way to share files between different machines on a network. these files are located on the client local hard drive. One of the disadvantages of NFS is that it tries to make a remote file system that appears as a local file system.
- **Andrew File System:** It is a distributed networked system, which uses a set of trusted servers to present a homogeneous, location transparent file name space to all the client workstations. AFS uses Kerberos for authentication and implements access control lists on directories for users and groups.
- **Blue Sky File System:** The blue sky film system uses the object data structures maintained in the file system. and their organization in a log-structured format. The Blue Sky file systems use to cleans the logs comprising file system.
- **Lustre File System:** It is a scalable parallel distributed file system
- **Mirror File System:** MFS has two unique features that enable it to provide benefits for Cloud Computing users and service providers. Replication at the File System Software Module Level allows file replication between two separate systems in real time. A user can have two copies of the same file on two systems, one on the local desktop, the other one on the Cloud server. This helps to make transition or migration from local computing to Cloud Computing easy and seamless. For the service provider, copies of the same file can be replicated in multiple data centers in real time, which helps to make the operation of the Cloud service very efficient and always available. MFS does not restrict the storage systems to be shared and confined physically to one data center, the service providers can scale the operation up from one data center to multiple data centers easily as needed and vice versa. Each data center has its own system and its own storage, but contains information identical to that of the other data centers. The distributed capability of MFS makes the Cloud Computing service much more reliable and efficient.

III.   LITERATURE SURVEY

Michael Vrable, et.al,(2012): discuss about the blue sky file system that is used in the cloud storage. In this paper, we explore an approach for bridging these domains for one particular application: network file service.[4] In particular, we are concerned with the extent to which traditional network file service can be replaced with commodity cloud services.

Jasmin James, et.al, (2012): discuss about the security in cloud computing. Cloud computing is fast growing area in computing research. With the advancement of the Cloud, many new possibilities are coming into picture, like how applications can be built and how different services can be offered to the end user through Virtualization. There are the

cloud service providers who provide large scaled computing infrastructure defined on usage, and provide the infrastructure services in a very flexible manner. The virtualization forms the foundation of cloud technology where[6] Virtualization is an emerging IT paradigm that separates computing functions and technology implementations from physical hardware. By using virtualization, users can access servers without knowing specific server details. The virtualization layer will execute user request for computing resources by accessing appropriate resources. In this paper, author firstly analyses the different Virtual Machine (VM) load balancing algorithms. Secondly, a new VM load balancing algorithm has been proposed and implemented for an IaaS framework in Simulated cloud computing environment.

Tejinder Sharma, et.al,(2013): in this paper author discuss about the cloud computing. As, the computer networks are still in their infancy, but they grow up and become sophisticated. Cloud computing is emerging as a new paradigm of large scale distributed computing. It has moved computing and data away from desktop and portable PCs, into large data centers. It has the capability to harness the power of Internet and wide area network to use resources that are available remotely.[5] There are many security issues in the cloud computing. In this paper, author discuss about the various scheduling problems. One of the challenging scheduling problems in Cloud datacenters is to take the allocation and migration of reconfigurable virtual machines into consideration as well as the integrated features of hosting physical machines. In order to select the virtual nodes for executing the task, Load balancing is a methodology to distribute workload across multiple computers. The main objective of this paper to propose efficient and enhanced scheduling algorithm that can maintain the load balancing and provides better improved strategies through efficient job scheduling and modified resource allocation techniques.

IV. PROPOSED WORK

In cloud computing, the platform, computing and software can be used as services. It is the form of utility computing, in which customer need not own the necessary infrastructure and pay for only what they use. The computing resources are delivered as virtual machines. In such a scenario, task scheduling algorithms play an important role where the aim is to schedule the tasks effectively. It helps to reduce the turnaround time and improve resource utilization.

V. METHODOLOGY:

Load balancing in the cloud computing mainly impact on the performance of file system. With the load balancing technique improve the efficiency of the file system. In this thesis mainly work with the better load balance and cloud partition under the different situations. In this proposed system to develop such a file system which can execute N number of jobs on 72 processors which can take less time and work more efficiently as compared to existing file systems. To achieve this goal the proposed file system uses to approaches which are time sharing and space sharing.

Time sharing approach helps to balance the load of number of jobs on processors and also helps to allocate that job the processor can execute according to its capacity which results in getting less weight time for the jobs. After this the time sharing technique execute jobs which are allocated jobs according to job sharing techniques and result in producing less response time then the existing file systems. The space sharing technique also allows splitting the job on different processors if one processor is not able to fulfill the requirements of the job then the job will be split on the different processors which makes job to be executed in less time. The proposed system also uses the demand driven approach to make the results of the file system more efficient. In the demand driven approach if the processor is in idle state then it will demand its parent for the jobs and if the parent does not have the job then it will demand the job from his parent will makes the file system to wait less for the jobs. In the work load model all tasks of jobs have equal service demand. Job cumulative service demand is dividing into maximum jobs and each job will have a demand of minimum time. This work load shows the advantage of space sharing policy.

I) *Job Selection*: Job selection policy is used to select the jobs in the queue. The global scheduler consist the jobs in the queue. The aim of scheduling policy is to carry the job from the queue in some manner. So we use First Come First Serve policy. It is one of the simple policies and it has less overhead as compare to other policies. It implements just one queue which holds the tasks in order they come in. The job is served in arrival order.

II) *Selecting Site*: The Site/Cluster is selected on which the job runs. The Most-fit policy is used to select the cluster. The Most-fit policy is used to reduce resource fragmentation by choosing the appropriate cluster which waste less number of processor and by taking care of the other jobs in the queue.

Each cluster which has enough processors for the waiting job, the file system performs a series of simulated activities, to measure how many immediate subsequent allocations can follow the allocation decision. After each cluster has been checked, the file system selects the cluster with the largest number of immediate subsequent allocations to perform current job allocation.

If there is no single site having enough free processors Multi-site execution co-allocation will be used, this policy tries to run a parallel job across several sites. However, a parallel job might take much longer execution time when running across site boundaries. This is because the speed and bandwidth of inter-site network is usually much slower and less than those of intra-site network. The Largest-First selection is done in case of Multi-site execution. The file system first sorts the sites in the computing grid into decreasing order according to the number of available processors BPUs. Then, the file system repeatedly picks up a site according to the sorted order until the total number of available processors on all selected sites is larger than or equal to the requirement of the waiting parallel job.

LEVEL 0

LEVEL 1

LEVEL 2

File System

Sub Systems

further Sub systems
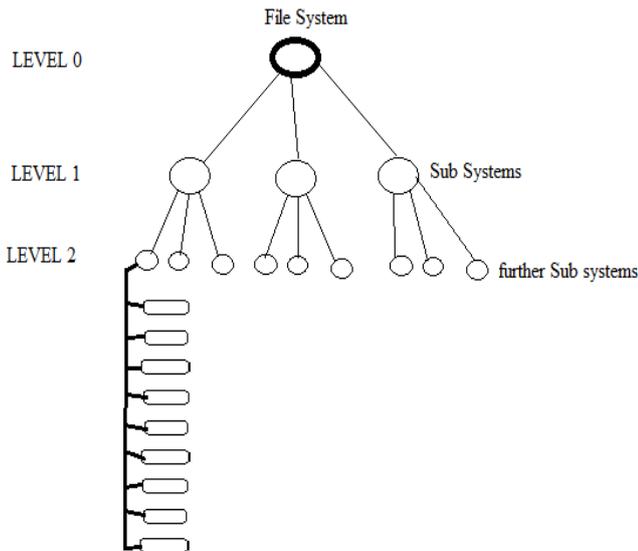
Figure 4: hierarchy of scheduling in file system

- In the first step define the tree structure in which main file system divide into three sub file systems. After these three sub file system divide into further three nodes, it contain total 9 nodes in the file system. Each node connected with 8 processors so the total 72 number of processors in the file system.

- All the processors find their BPU request according to the total sum of their BPUs range. In the file system all the processors perform the task divide equally on the processors.

- At time when all the jobs equally divide on the processors apply the round robin for execution of the processor this method to find the waiting time and response time of the job.

- Round robin: Each process is assigned a time interval, called its quantum, which it is allowed to run. If the process is still running at the end of the quantum, the CPU is pre-empted and given to another process. If the process has blocked or finished before the quantum has elapsed, the CPU switching is done when the process blocks, of course. All the scheduler needs to do is maintain a list of run able processes. When the process uses up its quantum, it is put on the end of the list. The only interesting issue with round robin is the length of the quantum. Switching from one process to another requires a certain amount of time for doing the administration saving and loading registers and memory maps, updating various tables and lists, flushing and reloading the memory cache in the file system.

REFERENCES:

1. http://cloudhostingprovider.co.uk/cloud-computing/basic-knowledge-about-cloud-computing
2. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia. (2009, Feb. 10). Above the clouds: A Berkeley view of cloud computing. EECS Dept., Univ. California, Berkeley, No. UCB/EECS-2009-28
3. Anthony T.Velte, Toby J.Velte, Robert Elsenpeter, Cloud Computing A Practical Approach, TATA McGRAW-HILL Edition 2010.
4. Michael Vrable_, Stefan Savage, and Geoffrey M. Voelker, BlueSky: A Cloud-Backed File System for the Enterprise
5. Tejinder Sharma, Vijay Kumar Banga, Efficient and Enhanced Algorithm in Cloud Computing, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3, Issue-1, March 2013
6. Jasmin James, Dr. Bhupendra Verma, EFFICIENT VM LOAD BALANCING ALGORITHM FOR A CLOUD COMPUTING ENVIRONMENT, Jasmin James et al. / International Journal on Computer Science and Engineering (IJCSE)