IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-11, Nov 2018 E-ISSN: 2347-2693

Providing the Resources by using scheduling algorithm in Cloud Computing

B. Muni Lavanya

CSE Department, JNTUACEP, Pulivendala, India

*Corresponding Author: munilavanya45@gmail.com, Ph.No: 9490080261

Available online at: www.ijcseonline.org

Accepted: 25/Nov/2018, Published: 30/Nov/2018

Abstract— Cloud computing is an environment for providing different services on the premise of demand and the need. It can able to build the virtualization, distributed computing to assist the cost-efficient usage method to adequate the resources with the usage of resource scalability, computing resources and on-demand resources. It can manage the user requests and also make available the resources by implementing the new virtual machines. The virtual machine can qualify for assigning the resources premises on the availability of resources. Cloud computing assists to load balancing and avoid the delay in running of the job. Based on the characteristics of the job new virtual machines are implemented. This paper can deal with the assigning of resources to the job based on its requirement and also consider the priorities of the job. Here we distinguish the various types of leases with their corresponding priorities. According to that priority, the execution can be processed. The main aim of this work is to delay the low priority job (consists high deadline of the job) and schedule the high priority job (having a low deadline of the job) and dynamically allocates VM resources to the user job within the deadline.

Keywords— CloudSim, Virtual machines, SLA

I. INTRODUCTION

Cloud is a computing technology based on parallel, distributed and grid computing. Cloud can have properties such as pliancy and scalability on providing the services based on low cost and high reliability. As compared to the typical distributed system it is complex to maintain worthy of trust in providing the services. Getting reliability in cloud computing is hard and complex. In cloudcomputing, the term reliability concludes that providing a failure-free operation (or) providing the quality of service based on user request to achieving the user satisfaction. We need to make available quality in the service. The term reliability states that it is a mixture of quality attributes such as fault tolerance, availability and fault recovery. Based on uptime and downtime taken by the service can be stated as the availability. Availability is one the major factor to provide the reliability of service. According to user's view downtime is to be downtime and according to service provider's view downtime can be distinguished into two forms such as planned downtime and unplanned downtime. The business organizations are always to decrease the own unplanned time to maintain reliability issues.

Scheduling is a major part ofthe operating system. Scheduling in the operating system can decide such that schedule the problem to the CPU. When the different problem can be given to the resources, then all the problems can wait in the running queue. Based on the schedule the job can be allocated to the CPU. Turnaround time is represented as the time taken between arriving the job and as well as completion of job which includes waiting time and execution time. The response time denotes that the how faster response can be received from the system to execute a job. Based on a number of jobs can be completed in unit time can be stated throughput of the system. Minimum response time is one of the performance metrics which decides user satisfaction who can expect less response time.

In the typical production management system,the cloud environment can allocatemany jobs. With the help of using the scheduler, the corresponding job can be formed by the cloud. The scheduler can have the software to interface the defined workflowsanddependencies and to execute the submitted jobs dynamically and automatically. To run the user's job, all the required VM images are stored and pre-configured by the Cloud Broker. All the en-queued jobs are to be queued to schedule the jobs in a sequential execution. At every moment scheduler performs five tasks:
(I) Analyze the incoming future workloads
(II) In advance arrange the necessary VMs
(III)Virtual machines are scheduled to corresponding jobs.
(IV) When the billing time unit (BTU) is to be increased then release the idle VMs
(V) When the waiting time of the job increases, then create the new VM.

The significance of Resource Allocation:

In distributed cloud computing, Resource Allocation (RA) is the technology toward the approachable resources to the corresponding cloud applications over the web. Resource assignment starves services if the portion isn't managed precisely. Resource provisioning takes care of that issue by enabling the service provider to deal with the resources for every individual module. Resource Allocation Strategy (RAS) is tied in with joining cloud provider activities for utilizing and distributing rare resources inside the limit of cloud conditionto address the issues of the cloud application. It requires the sort and measure of resources required by every application with a specific end goal to finish a user work. The request and time of allotment of resources are likewise a contribution for an optimal RAS. An optimal RAS ought to keep away from the accompanying criteria as follows:

a) Resource contention situation emerges when two applications try to access the same resource in the meantime.

b) The scarcity of resources emerges when there are limited resources.

c) Resource fragmentation situation emerges when the resources are separated. [There will be sufficient resources but not able to provide the required application.]

d) Over-provisioning of resources emerges when the application gets excess resources than the requested one.

e) Under-provisioning of resources happens when the application is allotted with fewer numbers of resources than the demand.

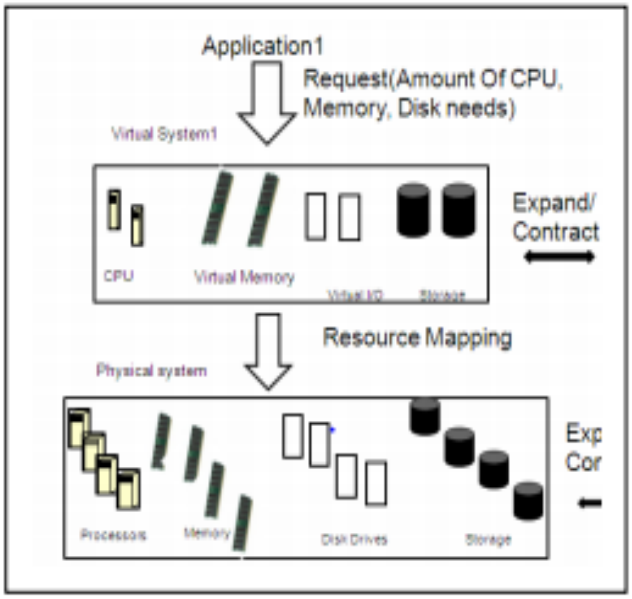

Figure 1: Mapping of virtual to physical resources

The objective of cloud computing is to enable clients to take profit by these technologies, without the requirement for deep learning about or expertise with every last one of them. The cloud intends to cut expenses and enables the clients to focus on their center business as opposed to being hindered by IT obstacles. The guideline to empowering development for cloud computing is virtualization.

Virtualization software isolates a physical computing gadget into at least one "virtual" gadgets, every one of which can be successfully utilized and managed to perform processing computing tasks. With a workingsystem-level virtualization making a versatile arrangement of multi independent computing gadgets, idle computing resources can be scheduled and utilized more effectively.

Virtualization gives the agility required to accelerate IT activities and decreases cost by expanding infrastructure utilization. Autonomic computing computerizes the strategy through which the user can give the resources on-demand. By limiting user inclusion, mechanization accelerates the procedure, decreases work costs and lessens the possibility of human mistakes.

Clients repeatedly face various difficult business issues. Cloud computing gets ideas from Service-oriented Architecture (SOA) that can empower the client to break these issues into services that can be fused to answer. Cloud computing gives the most of its advantages as assets as services and makes use of the well-established standards and best practices got in the space of SOA to empower worldwide and simple access to cloud benefits standardized.

Cloud computing additionally uses ideas from utility processing to give metrics to the servers which can be utilized. Such measurements are at the focal point of the general public cloud pay-per-utilize techniques. Likewise, evaluated services are a fundamental bit of the contribution to the autonomic computing, empowering services to scale on-demand and to perform modified disappointment recovery. Cloud computing is a sort of matrix processing; it has developed by tending to the QoS (Quality of Service) and reliability issues. Cloud computing gives the tools and advancements to implement data/compute concentrated parallel applications with substantially more reasonable costs contrasted with parallel computing methods.

II. RELATED WORKS

Cloud computing is efficient for virtualization to maintain and to allocate virtual machines efficiently in the cloud environment. In our work, we used to analyze and to enhance the characteristics of the cloud environment. We mainly consider cost and reliability are the two major components to allocate resources dynamically. Our main intention is to implement the non-profit based scheduling algorithm. To improves the performance and reliability of the cloud by decreasing the waiting and execution time for allocating resources. Then finally we can increase the user satisfaction by creating the new VM.

Weiwei Lin [1] stated a method to consider the threshold for dynamic allocation can be done Based on the dynamic allocation of resources according to the load changes. Threshold strategy can optimize the choice of resource reallocation CloudSim can be utilized to enhance the performance of allocating the resources contrast to previous techniques. Saraswathi et al. improve a method that based on characteristics of the job resource allocation is to be done [2]. In this method, they use preemptive scheduling algorithm that means low priority job can be preempted if a high priority job can enter into the queue. Scheduling the new job by suspending the running job is not an efficient algorithm to allocate the resources.

Kumar and Saxena [3] proposed a technique to compare the performance with the well-known VCG mechanism. In real time cloud resource allocation system most of the resources providers can have a fixed cost and also higher overhead to allocating the resources. A hybrid cloud environment can userule-based provisioning algorithm to allocating the resources [4]. In this proposed method they didn't consider cost, reliability to allocating the resources.

Based on user requirements we need to allocate resources according to cost and reliability which is the most complex to manage the cost and reliability. To determine the smallest number of servers a performance model is to be used [5], [19]. In this approach, they won't consider cost and also reliability to allocating the resources.

Priya Gupta[6] implements the allocation of VM to the user according to analyzing the characteristics of the job. While executing the high priority job the low priority job doesn't affect the execution of high priority job. They dynamically allocate VM to the user. To planning the projects a cellular automation entropy method can be used to making the decision [7],[8] this model is not designed for the environment that can be more challenging. While executing the high priority job the low priority job doesn't affect the execution of high priority job. They dynamically allocate VM to the user.

Resource provisioning is one of the genuine tasks in large-scale distributed systems, for instance, federated Grids. Recently, the resource management systems in these situations have been established to utilize the lease distraction and virtual machines (VMs) for resource provisioning. In the large-scale distributed systems, resource providers are capable of serving requests from external users along with their local users. The problem identified when there are no sufficient resources for local users, who have a higher priority than external ones, and need resources desperately. This issue could be solved by pre-empting VM-based leases from external users and allocating them to the local ones [9], [17].

Grid'5000, a 5000 CPUs nation-wide infrastructure for research in Grid computing. Grid'5000 is designed to provide a scientific tool for computer scientists similar to the large-scale instruments used by physicists, astronomers and biologists. The motivations, design, architecture, configuration examples of Grid'5000 and performance results for the reconfiguration subsystem [11], [18].

The fundamental resource provisioning abstraction in Haizea is the lease. Naturally, a lease is some agreement where one party agrees to deliver a set of resources (an apartment, a car, etc.) to another party. At the point When a user needs to request computational resources from Haizea, it does as such as a lease. When applied to computational resources, the lease abstraction is a great and general develop with the help ofa lot of nuances. The Haizea Manual contains a detailed definition of leases and these various types of leases upheld by Haizea (see below for a quick list of supported lease types) [10].

A newly designed process for dynamic resource administration in a cluster designer called Cluster on-Demand (COD). COD can access servers from a systematic pool to different virtual clusters (multiple v clusters), without considering the configured software flat form, namespaces, usable user controls, and also network storage volumes. The experiments are utilizing the most popular and suitable Sun Grid Engine batch assigned to demonstrate that dynamic virtual clusters are an authorizing abstraction for most forward resource management in computing usage and grids [12], [20].

An architecture of grid is expanded with peer-to-peer links among the sites based on the equal administrative control. To manage this architecture, we engage with the key conception of allocating matchmaking, which is no permanent (i.e., temporarily) with unsighted resources from remote sites to local flatform.With trace-related simulations is to compute an output based on the multiple infrastructural and load states, it outperforms other methodologies to inter-operating grids. Especially, we derived that delegated matchmaking improves up to 60% more effective throughput and finishes 26% more jobs than its top alternative [13], [14],[15]

The Cloud Computing delivers three important types of Infrastructure as a Service (IaaS) resources on demand which are: computing, networking, and storage. Computing resources are a collection of Physical Machines (PMs), each comprised of one or more processors, memory, network interface and local I/O, which together provide the computational capacity of a Cloud environment (e.g.,a virtual machine). These PMs may be required to interconnect with a high-bandwidth network by utilizing the networking resources (e.g.,a virtual switch). The Cloud storage resources are entitled as storage services [16], [23].

Infrastructure as a service (IaaS) introduces the online services that give high-level APIs utilized to dereference of different low-level details of underlying network architecture like physical computing resources, based on userlocation, amount of data partitioning, scaling, providingthe security, an also backup,etc. PaaS vendors givea chance to the improvement of the related platform to application developers.The provider can mainly developa toolkit and also standards for the establishment, channels for distribution cause and payment. In the PaaS designs, cloud providers provide a computing environment, mainly consists operating system, programming-language environment, major database, and implemented a web server. [21], [22]. Application developers can develop and run their related software solutions to issues on a cloud platform without examining the cost and as well as the complexity of buying and administrating the corresponding hardware and software proto layers. In the software as a service (SaaS) design, users can able to utilize their application software and also their databases. Cloud providers can have to access the infrastructure and platforms that can be used to run the applications. SaaS is periodically suggested as "on-demand software" and is usually considered on pay-per-use premises or utilizing with a subscription fee.

III. PROPOSED WORK

```
Procedure 1: Selecting a high priority job for the execution.
1: Input: Jobs in queue for execution; Adding a New high priority job into queue
       : Threshold;
2: Begin
3: For each job is executing in a queue
4:      if (lease == suspendable || cancellable)// check lease type//
5:         candidateSet.add(job)
6:          end if
7: end for
8: For each job present in candidateSet
9:     if (deadline.job < deadline.job of
                                  new high priority)
11:                candidateSet.remove(job)
12:     end if
13: end for
14: For each job present in candidateSet
15:     if (execution.job > Threshold)
16:        candidateSet.remove(job)
17:     end if
18: end for
19: if (candidateSet.count > 1)
20:     candidateSet.select(1)
21: end if
22: End
23: Output: Selected Job for execution of new high priority job
```

From the virtualization, cloud computing has been understoodas the factor to enhance efficiency and agility. The virtualization has been helped for the efficient use of hardware resources. In order to reduce the physical servers in the cloud computing environment virtual machines are to be allocated to the users based on their job performance. To meet out the service level agreement (SLA) VM resources are to be allocated based on their characteristics of the job. SLA is an agreement between a client and the service provider. These are output based and depends on various parameters such as availability, and qualityof the resources.In this paper, we proposed a method such that VM resources are allocated dynamically by configuring the virtual resource periodically.

In procedure 1, we set various priorities for the data, and the procedure shows the process of selecting a high priority for the execution based on the threshold values. The procedure is performed when the jobs are in the queue for execution.
While executing a job all the resources are to be allocated to the corresponding job when a new high priority job enter into the queue then suspend the execution of low priority job and allocate them all the resources to the high priority job. And provide the opportunity to use the resources to the high priority job. When a job is arriving, then it checks the availability of VMs if the required VMs are available then it allows to run the job on the VM. If there is no availability of VMs, then there is no possibility of running the job. The VM can check the low priority job which can have the corresponding job. In between a high priority can enters then it to be pre-empted and scheduled to high priority job. Then continue the execution of low priority job.
We assumed three various types of leases to associate with the jobs are:

(I) Cancellable
(II) Suspendable
(III) Non-preemptable

Consider the lease type from the list if the algorithm found two or more same priority jobs. Based on lease type job execution can be processed. If the lease of the job belongs to non-preemptable, then there is no need to consider candidate set. If the priority of cancellable lease can be higher than the suspendable lease type, then such type of leases can be killed. The jobs those can belong to the suspendable lease those can be suspended and then resumed. The level of completion of the job can be considered if two or more low priority jobs can be entered into the queue. The job with the completion of minimum portion then that job can consider for preemption.

Algorithm 1, gives the steps in detail for high priority job execution when all the jobs are available,and resources are allocated in contrast to procedure 1.


```
1: Algorithm 1: high priority job execution when all the available existing
resources are allocated
2: Input: New job, all jobs running in host
3: Begin
4: New job entered into queue
5: if (New deadline,job< all jobs running in host)
6:      job consists high priority = New job
            If (check VM is available)
7:              assign High priority job to that VM
8:                  else
9:                   high priority Job execution
10:                   priority job ();
11:                     Suspend (Suspend job)
12:                         for the suspended VM allocate the high priority job
13:          end if
14:          all jobs are to be executed on VM
15:          if (completion of a job which is running in VM)
16:            resume (Suspend job)
17:                 To that VM allocate the suspended job
18:          end if
19:   resumed job is executed
20: End
21: Output: After execution of all jobs submitted to the host
```

IV. RESULTS AND DISCUSSIONS

In our paper, we use the cloudSim, and simulated the data center with the two new hosts each with two PES and then we are created the two VMs which all in a nee of one PE. Based on the requirement of a number ofPEs in the host and number of PE's required to the VMs we can allocate VMs as the hosts. The jobis given to VM for execution. The execution can be doneby first come-first serve scheduling algorithm. The deadlineforthe next job is checked. If the deadline of the job less than the next two jobs then it has high priority otherwise the jobbelongs to low priority.

The low priority job can be executed after the execution of any of another job. For the execution of high priority job,the low priority jobs are to be suspended and then high priority job can get executed. After the execution of the higher priority job, it can compare the suspended job with the other jobs in the queue and then selects the high length of the job. Here high priority means that maximum length. Corresponding to the priorities of the all lease types the execution of the jobs can be processed. After the completion of selecting the high priority job that to be executed on the VM and remaining jobs are to be suspended. Then allocate the suspended job if any of the jobs can execute. The same process can be followed to all the other incoming jobs.

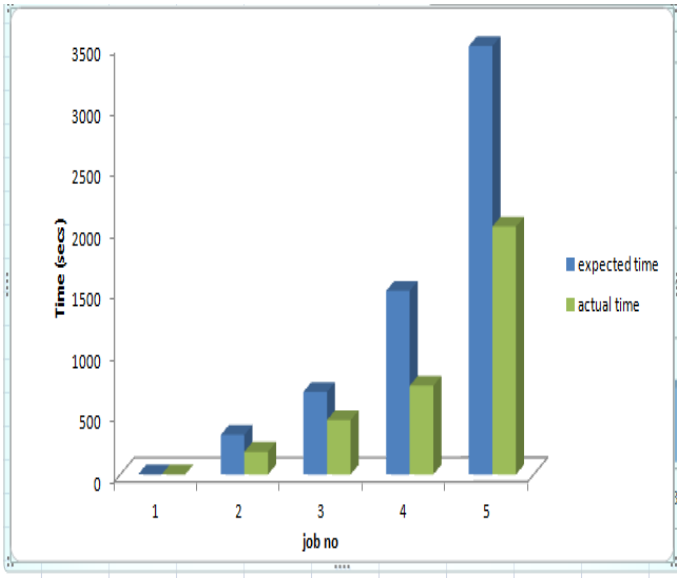

Figure 2: Jobs time comparison

The graph can show that four different jobs of variation between expected time and actual time. Initially, the expected time is very high. After applying our approach, we can gradually become smaller the execution time which can be stated as actual time. The actual time increases when the expected time is increased. Like that we can state our approach can work efficiently and provides the reliable service to the user. From the results, we can state that the new approach is to work more effectively as compared to previous approaches in terms of decreasing the execution time for a job. While the execution time of the jobs decreases then the throughput of the allocating resources to the job will be increased.

In figure 3 it shows the allocation of the number of PEs based on the requirement of VMs. New hosts are to be createdby needy of number of PEs to the virtual machine and accessibility of a number of PEs in the host. Here we can see that two hosts with the corresponding two PEs are to be created in the datacenter.

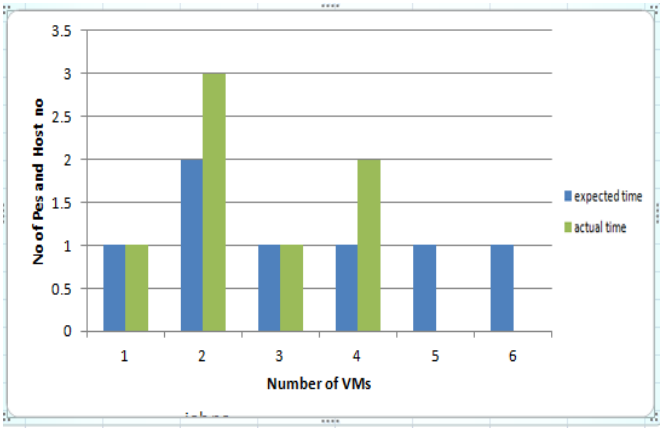

Figure 3: Allocation of VMs on the hosts based on PES

It can simulate the creation of new VMs 1,2,3,4 on the hosts,and we also observe that failure creation of VMs 4 and 5. This failure can occur because of a lack of PEs availability of hosts in the data center. This issue is resolved by creating the new hosts on the data center. Based on accessibility number of PEs in the host, a new data center is to be created. From the figures, we can see how the proposed technique lessens the time when compared to the existing techniques.

V. CONCLUSION

In this paper, a new method is implemented to execute a high priority job. This method reduces the unwanted creation of new virtual machines for the new incoming jobs and also help the re-usage of available PEs to execute the job. The new algorithm can suspend the low priority job when the high priority job can enter into the queue. This method has less overhead to creating the new VM. In the future, we are planning to implement the new approach in real time with the help of economy-related preemption methods in the cloud environment.

REFERENCES

[1] Zhao M, Figueiredo RJ. Experimental study of virtual migration in support of reservation of cluster resource of the 3rd Inter. Workshop on virtualization Tech Distributed Computing. ACM: Barcelona, 2007. pp. 5

[2] Saraswathi AT, Kalaashri. Y.RA, Dr. S. Padmavathi. "Dynamic resource allocation scheme in cloud computing". Procedia computer science 47b (2015). Pages: 30-36.

[3] Kumar, N, Saxena,S. "A preferred based resource allocation in cloud computing systems". In Proceedings of computer sci. 57, 104-111, 2015.

[4] R. Grewal, P. Patreiya, "A rule-based approach for effective resource provisioning in hybrid cloud environment", in advances in intelligent Systems and computing, vol. 203, Springer ,berlin Heidelberg, pp. 41-57, 2013.

[5] Ye Hu, JohnyWong , Gabriel Iszalai , Marlin Litoiu, "resource provisioning for cloud computing" Proceedings of the 2009 conference of the center for advanced studies on collaborative Research, November 02-05, 2009, Ontario, Canada.

[6] Priya Gupta, MakrandSamavatsar, Upendra Singh" Cloud computing through dynamic resource allocation scheme".International conference on electronics, communication and aerospace technology ICECA 2017.

[7] Sotomayor B, Montero RS, Llorente IM, " Foster I.Resource leasing and the art of suspending virtual machines ". In Proc. of the 11th IEEE Inter Conference on High Performance Computing and Communications, USA, 2009. Pages: 59-68.

[8] Chunlin Li, La Yuan Li. "Optimal resource provisioning for cloud computing". The Journal of Super computing, 2012. Vol. 62, Issue pp. 989-1022,

[9] Amini Salehi M, Javadi B and Buyya R. "Resource Provisioning based on Preempting Virtual Machines in Distributed Systems". The Journal of Concurrency and Computation: Practice and Experience, 2013. Vol. 26, No. 2. Pages: 412-433.

[10] http://Haizea.cs.uchicago.edu/

[11] Bolze R, Cappello F, Caron E, Daydé M, Desprez F, Jeannot E, Jégou Y, Lanteri S, Leduc J, Melab N, MornetG,Namyst R, Primet P, Quetier B, Richard O, El-Ghazali T, Touche I. Grid'5000: a large scale and highly reconfigurable experimental Grid testbed. International Journal of High Performance Computing Applications 2006;20(4):481–497.

[12] Chase JS, Irwin DE, Grit LE, Moore JD, Sprenkle SE. Dynamic virtual clusters in a Grid site manager. Proceedingsof the 12th IEEE International Symposium on High Performance Distributed Computing, USA, 2003; 90–98.

[13] Amini Salehi M, Javadi B, Buyya R. Resource provisioning based on leases preemption in InterGrid. In Proceedingof the 34th Australasian Computer Science Conference (ACSC'11), Vol. 113, CRPIT. ACS: Perth, Australia, 2011;25–34.

[14] Tsafrir D, Etsion Y, Feitelson DG. Backfilling using system-generated predictions rather than user runtime estimates.IEEE Transactions on Parallel and Distributed Systems 2007; 18(6):789–803.

[15] Iosup A, Epema DHJ, Tannenbaum T, Farrellee M, Livny M. Inter-operating grids through delegated matchmaking.In Proceedings of the ACM/IEEE Conference on Supercomputing (SC '07). ACM: USA, 2007; 1–12.

[16] I. Foster and C. Kesselman, editors. The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann Pub., 1999.

[17] J. P. Jones and B. Nitzberg. Scheduling for parallel supercomputing: A historical perspective of achievable utilization. In JSSPP, volume 1659 of LNCS, pages 1–16, 1999.

[18] U. Lublin and D. G. Feitelson. The workload on parallel supercomputers: modeling the characteristics of rigid jobs. J. PDC, 63(11):1105–1122, 2003.

[19] A. Iosup, M. Jan, O. Sonmez, and D. Epema. The characteristics and performance of groups of jobs in grids. In Euro-Par, LNCS, 2007.

[20] H. H. Mohamed and D. H. J. Epema. Experiences with the koala co-allocating scheduler in multiclusters. In CCGrid, pages 784–791. IEEE CS, 2005.

[21] D. G. Feitelson and L. Rudolph. Metrics and benchmarking for parallel job scheduling. In IPPS/SPDP, volume 1459 of LNCS, pages 1–24, 1998.

[22] A. Iosup, C. Dumitrescu, D. H. Epema, H. Li, and L. Wolters. How are real grids used? The analysis of four grid traces and its implications. In GRID, pages 262–270. IEEE CS, 2006.

[23] Mohamed Graiet; AmelMammar; SouhaBoubaker; Walid Gaaloul. "Towards Correct Cloud Resource Allocation in Business Processes" IEEE Transactions on Services Computing, Year: 2017, Volume: 10, Issue: 1 Pages: 23 - 36.

Authors Profile

Mrs. B Muni Lavanya currently working as Lecturer in JNTUA College of Engineering, Pulivendula. She has been working on Wireless Mesh Networks since 2015 and is interested in the areas of research Parallel Processing, Big Data and Cloud Computing.

