**LITERATURE SURVEY**

**TITLE:** ‘‘Cloud computing virtualization of resources allocation for distributed systems,’

**ABSTRACT:** Cloud computing is a new technology which managed by a third party "cloud provider" to provide the clients with services anywhere, at any time, and under various circumstances. In order to provide clients with cloud resources and satisfy their needs, cloud computing employs virtualization and resource provisioning techniques. The process of providing clients with shared virtualized resources (hardware, software, and platform) is a big challenge for the cloud provider because of over-provision and under-provision problems. Therefore, this paper highlighted some proposed approaches and scheduling algorithms applied for resource allocation within cloud computing through virtualization in the datacenter. The paper also aims to explore the role of virtualization in providing resources effectively based on clients' requirements. The results of these approaches showed that each proposed approach and scheduling algorithm has an obvious role in utilizing the shared resources of the cloud data center. The paper also explored that virtualization technique has a significant impact on enhancing the network performance, save the cost by reducing the number of Physical Machines (PM) in the datacenter, balance the load, conserve the server's energy, and allocate resources actively thus satisfying the clients' requirements. Based on our review, the availability of Virtual Machine (VM) resource and execution time of requests are the key factors to be considered in any optimal resource allocation algorithm. As a results of our analyzing for the proposed approaches is that the requests execution time and VM availability are main issues and should in consideration in any allocating resource approach.

**TITLE:** ‘‘Cloud computing: A paradigm shift in the way of computing,’’

**ABSTRACT:** Introduction Cloud computing is a new trend of computing where resources like storage, computation power, network, applications etc. are delivered as services. This services are available to the customers as subscription-based model i.e. pay-as-you go. In this model, customers can get these services on their demands regardless of where these services are hosted and customers have to pay depending on their usage of services. In cloud computing, resources are made virtual and unlimited. Also, the resources can be provisioned from anywhere i.e. always available at any location. So, cloud computing is a new paradigm where we can provision resources dynamically, deploy applications, and can access platform-independent services. Cloud computing, successor of internet computing , is a technology, where the concept of utility, scalability, on-demand services are incorporated. Figure 1 illustrates " Internet Computing " vs. " Cloud Computing " . Defining Cloud in IT According to the U.S. National Institute of Standards and Technology (NIST) , Cloud is a classical model which enable omnipresent, convenient, on-demand network access to a publicly accessible pool of configurable resources like servers, storage, network components, applications; that can be accessed, manipulated and released with minimal management effort, less cost and minimal service provider interaction. Cloud computing can be defined by the following important properties. Service on demand: Cloud users can use services on their demands, whenever they need from any place and at any time without making any direct communication with cloud service provider. Wide network access: Services can be accessed over the network using different devices (like laptops, mobile phones, PDA, tablets, office computer etc.). Services can be provisioned in any platform, which means cloud services are platform independent. Pooled Resources: In cloud computing, resources are pooled together so that cloud providers can offer multi-tenant services. Multi-tenant supports multiple users to be served at a time with physical and virtual resources. These resources can be dynamically assigned and released according to the user's choice. Increased elasticity: There is no limit for provisioning resources via cloud. So services can be easily and quickly scale in and scale out. For example, an online shopping site uses the resources from the cloud in terms of users.

**TITLE:** ‘‘Toward cloud computing: Security and performance,’’

**ABSTRACT:** Security and performance are basic requirements for any system. They are considered the criteria for the measurement of any progress in a security system. Security is an indicator that affects the level of performance through the threats that influence the performance of parts of the cloud during the rendering of services. Both security and performance demonstrate the efficiency of cloud computing which indicates that the performance and security are measurements for the extent of the development of the cloud. In this paper, the relationship between performance and security will be examined to know the extent of their impact on the progress of cloud computing.

**TITLE:** ‘‘Benefits and challenges of the adoption of cloud computing in business,’’

**ABSTRACT:** The loss of business and downturn of economics almost occur every day. Thus technology is needed in every organization. Cloud computing has played a major role in solving the inefficiencies problem in organizations and increase the growth of business thus help the organizations to stay competitive. It is required to improve and automate the traditional ways of doing business. Cloud computing has been considered as an innovative way to improve business. Overall, cloud computing enables the organizations to manage their business efficiently. Unnecessary procedural, administrative, hardware and software costs in organizations expenses are avoided using cloud computing. Although cloud computing can provide advantages but it does not mean that there are no drawbacks. Security has become the major concern in cloud and cloud attacks too. Business organizations need to be alert against the attacks to their cloud storage. Benefits and drawbacks of cloud computing in business will be explored in this paper. Some solutions also provided in this paper to overcome the drawbacks. The method has been used is secondary research, that is collecting data from published journal papers and conference papers.

**TITLE:** ‘‘Proposing a load balancing algorithm for the optimization of cloud computing applications,’’

**ABSTRACT:** Cloud Computing (CC) is a fast growing services that make use of pay per use model. The technology provides various services in terms of storage, deployment, web services etc. however the expand of these services and the tremendous increase of user demand has resulted in many challenges to keep up the performance in line with QoS measurement and SLA document provided by cloud providers to enterprises. This expand resulted in challenges such as load balancing. Besides that, user's requirements became hard to fulfil in terms of response time and deadline regarding task scheduling. To address these challenges, this research proposes an optimized algorithm with the use of Machine Learning Classification technique based on deadline constraints. The main objective of the proposed algorithm is to enhance the efficiency, optimize the server resources by considering the priority of different users' tasks and avoid server breakdown. Our proposed algorithm will address the mentioned issues and current research gap based on the recent literature.

**TITLE:** Load balancing in cloud computing: A big picture,’’

**ABSTRACT:** Scheduling or the allocation of user requests (tasks) in the cloud environment is an NP-hard optimization problem. According to the cloud infrastructure and the user requests, the cloud system is assigned with some load (that may be underloaded or overloaded or load is balanced). Situations like underloaded and overloaded cause different system failure concerning the power consumption, execution time, machine failure, etc. Therefore, load balancing is required to overcome all mentioned problems. This load balancing of tasks (those are may be dependent or independent) on virtual machines (VMs) is a significant aspect of task scheduling in clouds. There are various types of loads in the cloud network such as memory load, Computation (CPU) load, network load, etc. Load balancing is the mechanism of detecting overloaded and underloaded nodes and then balance the load among them. Researchers proposed various load balancing approaches in cloud computing to optimize different performance parameters. We have presented a taxonomy for the load balancing algorithms in the cloud. A brief explanation of considered performance parameters in the literature and their effects is presented in this paper. To analyze the performance of heuristic-based algorithms, the simulation is carried out in CloudSim simulator and the results are presented in detail.

**TITLE:** ‘‘Performance comparison of heuristic algorithms for task scheduling in IaaS cloud computing environment,’

**ABSTRACT:** Cloud computing infrastructure is suitable for meeting computational needs of large task sizes. Optimal scheduling of tasks in cloud computing environment has been proved to be an NP-complete problem, hence the need for the application of heuristic methods. Several heuristic algorithms have been developed and used in addressing this problem, but choosing the appropriate algorithm for solving task assignment problem of a particular nature is difficult since the methods are developed under different assumptions. Therefore, six rule based heuristic algorithms are implemented and used to schedule autonomous tasks in homogeneous and heterogeneous environments with the aim of comparing their performance in terms of cost, degree of imbalance, makespan and throughput. First Come First Serve (FCFS), Minimum Completion Time (MCT), Minimum Execution Time (MET), Max-min, Min-min and Sufferage are the heuristic algorithms considered for the performance comparison and analysis of task scheduling in cloud computing.

**TITLE:** ‘‘Heuristic-based load-balancing algorithm for IaaS cloud,’’

**ABSTRACT:** The tremendous growth of [virtualization technology](https://www.sciencedirect.com/topics/computer-science/virtualization-technology) in cloud environment reflects the increasing number of tasks that require the services of the virtual machines (VMs). To balance the load among the VMs and minimizing the makespan of the tasks are the challenging research issues. Many algorithms have been proposed to solve the said problem. However, they lack in finding the potential information about the resources and tasks and it may lead to the improper assignment of the tasks to the VMs. In this paper, we propose a new load balancing algorithm for Infrastructure as a Service (IaaS) cloud. We devise an efficient strategy to configure the servers based on the number of incoming tasks and their sizes to find suitable VMs for assignment and maximize the utilization of computing resource. We test the proposed algorithm through simulation runs and compare the simulation results with the existing algorithms using various performance metrics. Through comparisons, we demonstrate that the proposed algorithm performs better than the existing ones.

**TITLE:** A study on virtualization and hypervisor in cloud computing,’’

**ABSTRACT:** Virtualization is the key factor in todays’ world of cloud computer technology. Virtualization is used to create abundant ardent resources from a single physical server/entity. Hypervisor is a software tool that isolates the physical infrastructure to create and sperate the resources to work in a cloud environment. Different types of operating systems can be hosted and managed by a single physical server. Virtualization by using hypervisor is significant to progress on system security, deliver greater flexibility and reliability. Currently, there is a bigger desire to use virtualized systems in corporate enterprises which enable to reduce cost and more proficiently use resources. In this paper, I discuss different types of the hypervisors and virtualization methods and how it benefits in expansion of resources in cloud computing

**TITLE:** ‘‘A comprehensive survey for scheduling techniques in cloud computing,’’

**ABSTRACT:** Resource scheduling becomes the prominent issue in cloud computing due to rapid growth of on demand request and heterogeneous nature of cloud resources. Cloud provides dynamism, uncertainty and elasticity based services to users in pay-as-you-go fashion over the internet. In recent decade, increase in requests (diverse and complex applications) for cloud services is raising the workload in cloud environment. Inefficient scheduling techniques face the challenges of resources being over utilized and underutilized (imbalanced) which leads to either degradation in service performance (in case of over utilized) or wastage of cloud resources (in case of underutilized). The basic idea behind the scheduling is to distribute tasks (diverse and complex nature) among the cloud resources in such a manner that [scheduling algorithm](https://www.sciencedirect.com/topics/computer-science/scheduling-algorithm) avoids the problem of imbalance. Scheduling algorithm should also optimize the [key performance indicator](https://www.sciencedirect.com/topics/computer-science/key-performance-indicator) parameters like response time, makespan time, reliability, availability, energy consumption, cost, resource utilization etc. To fulfill the above-mentioned objective, many state of art scheduling algorithms have been proposed based upon heuristic, meta-heuristic and hybrid, reported in the literature. This paper provides the systematic review as well as [classification](https://www.sciencedirect.com/topics/computer-science/classification) of proposed scheduling techniques along with their advantages and limitations. We hope that our systematic and comprehensive survey work as a stepping stone for new researchers in the field of cloud computing and will be helpful for further development of scheduling technique.