CHAPTER 2
LITERATURE REVIEW

2.1 INTRODUCTION

Cloud Computing tends to be an interesting technology, with huge potentials in enterprises and marketing related applications. Clouds design the infrastructure in a better way so as to access the applications and the associated data from anywhere. Enterprises pay to use the resources in cloud service providers for use of storage and further computational purposes. As a result, the infrastructure cost is much reduced in cloud environment. One of the major drawbacks in Cloud Computing is associated in optimizing the resources being allocated to the cloud infrastructure.

Because of the individuality of the model, resource allocation is carried out with the objective of minimizing the costs related to it. The other challenges of resource allocations are gathering customer demands and application requirements. Resource scheduling methods of Cloud Computing are commonly based on setting up and optimization of most of the important physical resources.

Interference Aware Resource Allocation (IARA) is used in the first work for proper resource allocation by discarding the interferences happening in between the communication. IARA employs a resource scheduler in order to

allocate the resources with optimal energy as well as bandwidth consumption. IARA schedules to achieve sub-optimization for cloud computing problems. IARA Scheduling Technique achieves both the allocation of resources and utilization of system resource.

Moreover, Adaptive Load Balancing (ALB) Approach in the second work exploits to balance the load in cloud environment. Balances load from every cluster group by using ALB to minimize the bandwidth and energy consumption with the adoption of repetitive query messaging. In general, ALB collects information about the current load of the other group and calculates the average energy and the bandwidth consumption of each group. ALB Method not only balances the energy consumption but also enhances the utilization of the resources with it minimal bandwidth usage.

The final work develops Genetic Clustering with Workload Multi-task (GCWM) Scheduler, based on clustering of a similar workload using genetic methods to perform many tasks simultaneously. GCWM Scheduler Method is used to minimize the computational cost and complexities involved during computation. In GCWM Scheduler methods are applied to cluster high range of nodes with the initial population, such as task selection, crossover and mutation operators for workload management. Fitness function is adopted with GCWM that performs cluster related task in cloud zone and communicates with each other efficiently. GCWM is more suitable in distributing the computing

resources and it ensures the multi-tasking operation with efficient users' communication.

2.2. DIFFERENT RESOURCE ALLOCATION TECHNIQUE IN CLOUD COMPUTING ENVIRONMENT

Most of the researchers have focused on allocating the resources properly with the design of various Resource Allocation techniques. Resource Allocation techniques arrange the resources, based on the priority of the processing ability, cloud storage and speed to resolve the cloud computing problem. The existing methods have been developed with certain localization based algorithms for preserving the high energy utilization and bandwidth consumption in the cloud environment. Studies also have proved better results in delivering the superior resource allocation methods. A few existing methods, that intend to establish a better resource scheduling policy as well as high utilization of resources are elaborated in the sub-sections.

2.2.1. Energy-efficient Resource Allocation in Cloud Computing

The most initial method developed in data processing of the cloud computing environment is Nephele. Much attention on efficient parallel data processing in cloud computing environment is done with Nephel. Nephele is a structure of data processing which openly exploits a dynamic resource allocation, supported by IaaS clouds for both, job execution and allocation. A definite task of processing a job can be prearranged to the different types of the virtual machines. This is routinely instantiated and

terminated during job execution. The authors enhance the evaluations of techniques with the analysis, in terms of processing jobs on an IaaS cloud computing environment.

Next, in parallel data processing, some of the structural schemes have also been established. One such method is the designed architectural frame and attitude for an energy-efficient Cloud computing proposed by Anton Beloglazov, et al., (2012) [1]. This technique is used to resource conditioning and scheduling algorithms for energy-efficient management of cloud infrastructure. The client application side is used with energy-aware allocation heuristics and the condition data center resources enhance the energy efficiency of the data center. Moreover, the technique leads to dispensed negotiated Quality of Service (QoS). Energy-efficient resource allocation strategy and scheduling algorithms are facilitating QoS prospect and energy utilization features of the devices. The result, provide better energy efficiency level with a cost saving under dynamic workload scenario

Cloud computing has more benefits for the deployment of data-intensive applications. Donald Kossmann, et al., (2010) [12] reveal that the system, mainly is used to reduce cost with a pay-as-you-go business model. The workload level is increased, based on the throughput level of the server. The authors focus on the transaction processing, such as read and updates workload level with OLAP. Moreover, the wide objective is on effective cloud

computing emergence for database applications and the report includes results of an inclusive evaluation. As a result cost level and performance of the service differ significantly, based on the workload.

Hatakeyama. K., et al., (2008) [19] have worked on cooperation computing and network resources much like are essential to be allocated all together to each service request. The researchers have developed the optimal joint resource allocation scheme for the multiple resource types. This method is used to achieve an effective use of the resource utilization. The authors have also designed fair joint multiple resource allocation method for an effort to equalize the entire amount of the key resources assigned for each service in each time block. Key resource is a type of resource with the full size of a requested resource, which is considered as the biggest proportion of the highest resource type. However, this method is only applicable to enable fair allocation between there two distinct services.

In addition to the energy efficient resource allocation factor, load balancing is also considered as one of the significant aspects in cloud environment. Studies on load balancing face trouble in VM resources scheduling. Jianhua Gu, et al., (2012) [25] introduce a scheduling strategy on load balancing of VM resources, based on genetic algorithm. Depending on the chronological data and the present state of the system and via genetic algorithm, this policy computes the influence measures on the system, after

exploitation of the required VM resources. After selecting the least-affective solution, the system aims to achieve the best load balancing as well as to reduce or negotiate dynamic migration. Simultaneously this method provides a variation rate to illustrate the load variations of the virtual machine. Further, it establishes the average load distance to calculate the whole load balancing consequence of the algorithm. As a result, this system provides better load balancing in the virtual machine with scheduling and improved good resource utility.

In cloud computing, computational resources are offered to remote users in the structure of leases. Jiyin Li, et al., (2010) [26] have developed an adaptive resource allocation algorithm with preemptable tasks in cloud system. For a cloud user, the work sends request to the multiple cloud services, simultaneously in favour of the cloud infrastructure. In this situation, the system needs parallel processing in the cloud system to improve its performance. A applying parallel processing in cloud computing, is very important to run the allocation of the resources as well as to schedule the tasks execution order. Resource Allocation Mechanism is used with a pre-defined job execution cash sow to enhance the bandwidth utilization of the cloud environment. The adaptive resource allocation algorithm is based on the updation of the definite task executions.

Kumar, K., et al., (2011) [34] offer a proficient, infrastructure as a service model to allocate the resources for real-time tasks. Real time tasks must be completed earlier than the given deadline, Cloud computing provides a selection of the resources with various speed and cost. In addition, in cloud computing, allocations of the resources can be scaled up depending on its requirements. This system is called elasticity and it has a key variance of obtainable multi-processor task allocations. The authors concentrate on the problem of resource allocation for a set of real time tasks and as such the economic cost is reduced. The authors also introduce polynomial-time solution for the efficient allocation of the resources.

The workflow is used to illustrate the variety of applications, concerning high processing and storage demands. As public cloud charge is used in a per-use basis, private clouds owned by users can be exploited with no charge. Luiz Fernando Bittencourt, and Edmundo Roberto Mauro Madeira (2011) [39] establish the Hybrid Cloud Optimized Cost Scheduling Algorithm to decide the resources in the public cloud and fit it to the private cloud. In general, the system provides sufficient processing energy to perform the workflow, within the given execution time. This algorithm is used to reduce the cost as well as to attain the desired execution time. The cooperation of the public and private cloud is called hybrid cloud. The hybrid cloud user has elasticity provided by the public resources cloud and it can be collected in the private resource pool as

the needed resources. The resource management and scheduling are further discussed in the sub-sections with some existing methods.

2.2.2 Efficient Resource Management and Scheduling in Cloud Environment

Marco A.S. Netto., et al., (2011) [46] have developed Meta schedulers co-allocation resources is developed by requesting a static number of processors and handling time for each cluster. Static requests, described by the users limit the primary scheduling and avoid rescheduling of the applications to the other resource sets. The greatest trouble with the heterogeneous environment is the measure of application execution times used by the users. Meta scheduler is widely used to perform predictions in automatic resource selection. Resource co-allocation, through rescheduling will maintain dependence on performance predictions for the multi-cluster iterative parallel applications. Iterative applications are used to recover many problems, such as knowledge and engineering and comprises large-scale computations, based on asynchronous model.

The main problem of cloud computing is to increase the energy consumption in the data center. As the main data center stays underutilized, an essential amount of energy is preserved by shifting the VM functions of the unused machines to the other machines and hiding such unused machines. Mehta, A., et al., (2011) [47] concentrate on to design the level for energy-efficient cloud data centers. This system creates with use of chronological

traffic data from the data centers through the service request prediction model. The authors to enable the classification of the number of the active servers, needed at a given instant, thus making it probable hibernation for the underutilized servers. As a result, the work is able to reduce the energy consumption level in the cloud environment.

Shakti Mishra, et al., (2010) [53] introduce a priority-based scheduling for running the queues and also employ the multilevel feedback queue scheduling, used for relocating tasks without degrading the local job performance level. These scheduling methods are used to increase the overall throughput level. The computation intensive application tends to transfer the process, which reduces the response time in the cloud environment. The clustering method offeres to enhance the availability of the service and thereby distributing the computational work load within the computational intensive application. The authors achieve the computing power with clustering usage effectively and assigns then to all the available nodes.

Considering the processor, more attention is paid to the multiple execution cores in the cloud computing era. The execution core is perceived as a distinct logical processor by the software. The operating system is optimized for Symmetric Multi processing and to improve the processor count up immediately with the advantage for the multiple execution cores. Siddha S., et al., (2007) [54] introduce the various multi core topologies and related

processor power management system provides new optimization opportunities to the process scheduler.

In cloud computing, an essential method to assign both the processing ability and network bandwidth is required to access it. Takuro Tomita and Shin-ichi Kuribayashi, (2011) [19] a congestion control method with the intention of minimizing the size of the needed resource for the congested resource type. This method is used to attain effectiveness by exploiting of the resources in a congested situation. Besides, the authors enhance the congestion control method to facilitate a fair resource allocation level in a congested situation. The fairness in a congested situation, author supposes the multiple resources are assigned at the same time to each request of service.

The next authors concentrate on the fair allocation system in the cloud environment. Xindong You, et al., (2011) [64] discuss the adoption of Automatic Resource Allocation Strategy, based on Market Mechanism (ARAS-M) for improving the resource utilization of the huge data centers as delivering services with elevated QoS to the Cloud Clients. The first system of the architecture of market model of ARAS-M aims to build the QoS-refectitive utility function and it is considered to have dissimilar resource requirements of the Cloud Clients. The authors justify the optimality of the equilibrium state of ARAS-M. The genetic algorithm is dependent on the automatic price adjusting algorithm, which is used for achieving the equilibrium state of ARAS-M. As a

result, ARAS-M is used to get a better resource allocation and also to achieve load balancing and the equilibrium state in cloud environment.

Resource scheduling is a main process for the cloud environment, such as IaaS. Zhongni Zheng, et al., (2011) [73] introduce the optimized scheduling algorithm, used to achieve better effective use of resource and also to attain optimization and sub optimization in cloud computing environment. The VM is reliable to maximize the fact of discovering the best allocation of the resources. The authors consider the mathematically, scheduling system under unbalance assignment problem. The scheduling system is used to parallel the genetic algorithm and the traditional genetic algorithm. The traditional genetic algorithm is faster than the level of the scheduling policy. As a result, this method is used to enhance the speed of the resource allocation and resource utilization.

Better development of cloud computing requires the exact quantitative analysis of the cloud data centers. As a perfect modeling of the cloud centers, it is not possible due to the nature of the cloud centers and diversity of user requests. Khazaei and Misic, (2011) [35] the author introduce an innovative approximate analytical model for the quantitative analysis of the cloud data server farms. They have made progress to attain an accurate evaluation of the complete probability distribution of the request, response time and the other significant performance indicators. The estimated analytical model supports the

cloud operator to evaluate the link between the number of servers and the input buffer size with a single side.

Cloud computing with its assured unlimited computation, storage and bandwidth, becomes a more interesting option for many organizations. A cloud provides mature, service dependent application required to dynamically recompose and to self adapt to the changing QoS requirement. Nallur Bahsoon, (2012) [49] has established a decentralized method used in market based heuristics for such self-adaption. Moreover, the system is used to double-auction and to agree application with the select service to choose in cloud environment. As a result, self-adaption mechanism is more effective for individual application, in addition to the collective perception of all the applications adapting simultaneously.

2.2.3. Resource Management with Improving Scheduling Effects

A famous protocol is popular for dynamic resource Management in huge Cloud Environments. Wuhib, F. et al., (2012) [61] have mainly used the gossip protocol for three contributions, such as to make certain fair resource allocation surrounded by sites/applications and dynamically to adapt the assignment of load changes. The work also combines both the number of physical machines and sites/applications in scales. The resource allocation issues dynamically increase the cloud utility based on CPU memory constraints. A protocol used to calculate the optimal solution, without regard to memory constraints, verify

accuracy and convergence properties. The protocol provides a better efficient heuristic solution for the whole problem, which includes reduction of cost for adapting an allocation. The protocol incessantly carries out a dynamic, local input and does not need global synchronization.

Beaumont, Oliver; et al., (2013) [8] consider the problem of conveying a set of clients, with demands to a set of servers with ability and degree constraints. Investigation is made to find an allocation, such as the number of clients assigned to a server is lesser than, the server degree and the whole demand is lesser than the server capacity, as well as increase throughput level. This problem has numerous natural applications in the context self-governing task scheduling or virtual machine allocation. The authors consider both the offline and online versions of the problem. The authors demonstrate degree constraint on the maximal number of clients that the server can hold as sensible in many contexts.

Electricity expenditure encompasses an important fraction of the full amount of the operating cost in the data centers. Guo, Y. and Fang, Y. (2013) [17] preserve the energy storage capabilities in data centers to minimize the electricity cost below the wholesale electricity markets. The authors employ a stochastic program used for the center-level load balancing, server based configuration and battery management as simultaneously providing assurance in the quality of the service in the end user. The authors use Lyapunov

Optimization Technique for cost saving and for energy storage capacity within an online algorithm. The authors consider the power cost minimization, using energy storage in the multiple data center with regard to both the spatial and temporal variations in the extensive electricity prices and workload arrival processes.

Wang, Z. et al., (2013) [62] have designed the Combinatorial Statistical Analysis on Cloud (COSAC), for a huge scale data set over Map reduces based on cloud environment. COSAC controls the two important phases, such as the distribution and the statistical analysis phase. The distribution phase analyses the new load balancing system dispensing the combination detail tasks across the processing units. The quantitative analysis stage at every unit optimizes the processing of the assigned fusions by extracting the computations that can be reused. COSAC also maintains a more practical scenario, where only an elected subset of objects requirements are to be evaluated against all objects. As a result the authors achieve a better efficiency, which is scalable and flexible.

The Cloud computing is to allow the business customers to scale their changeable resource usage based on requirements. Zhen Xiao et al., (2013) [74] explore the virtualization technology, used to assign the data center resource vigorously, based on the application demands and hold up the green computing by optimizing the wage in a number of servers. Besides it adopts

Skewness Method in order to compute inequality in the multidimensional resource utilization of the server. The authors combine together the different types of the workloads adequately and get better overall utilization of the server resources. In addition, the authors expand a set of heuristics that avoids overload in the system efficiently as saving energy is used.

Even mobile computing is a promising technology to enhance the quality level of the mobile service. Kaewpuang, R. et al., (2013) [36] consider the resource sharing issues to support the mobile applications in mobile cloud computing environment. In this environment, the service provider can assist to construct the resource pool so as to distribute its own resource with each other. The authors introduce a framework for the resource allocation with the resource management and to increase the resource utilization. The authors increase the benefit of the mobile cloud service providers. The authors establish and recover the optimization models to acquire the optimal number of application illustration that can be sustained to maximize the revenue of the service providers as meeting the resource necessities of the mobile applications.

Sheng Di and Cho-Li Wang (2013) [55] work discuss the distributed VM-multiplexing resource allocation method for controlling the decentralized resource. This method not only attains the increased resource utilization using

Proportional Share Model (PSM), but also provides provably the best execution efficiency. The authors also introduce the multi-attribute range query protocol which is used capable nodes. This protocol is used to find for each one task of its efficient resources under a randomized strategy that moderates the disputation among the requesters. As a result the authors use this protocol to get a better throughput level compared to the existing system.

Addis et al. (2003) [2] discuss in their work the worldwide attention in the delivery of computing and storage capability as a service carries on growing at a rapid pace. Moreover, the complication of such cloud computing centers need advanced resource management resolution. So that it is able to actively incorporate the cloud platform as a continuous service and gives guarantee. The authors introduce the resource allocation approach for virtualized cloud environments. The authors use assure performance and accessibility guarantees and reduce the energy costs at the very outsized cloud service centers. Moreover, shared hierarchical structure is adopted, based on the combined numerical non-linear optimization of resource allocation phase at multiple timescales.

2.3. LOAD BALANCING APPROACH IN CLOUD ENVIRONMENT

Most of the investigations in cloud environment also include issues related to load imbalance. Researchers attempt to deliver prominent load balancing schemes with energy efficient and bandwidth utilization in the cloud

environment. The methods compose algorithms to balance the load from each cloud nodes by reducing the bandwidth and energy consumption. With efficient methods, information can be gathered about the progress of the load balance so as to calculate the average energy and bandwidth consumption of each cloud node. Further, the sub-section elaborates the existing load balancing schemes.

2.3.1. Different Load Balancing Techniques in Cloud Infrastructure

Cloud computing based IT industry requires a model to attain reliability and misapprehension of its infinite capacity. It is necessary for each of the resources to be virtualized to conceal the execution of the way they are multiplexed and distributed. The parallel processing allocates the number of servers, central processing unit and resource exploitation in PaaS. Jianfeng Yan and Wen-Syan Li, (2009) [33] concentrate on automatic calibration resource allocation used to the logical task in parallel processing. The authors achieve possible exchange between the parallelism profit and overheads. The Calibrating algorithm is used to allocate the resources with illustration runtime statistic information.

Hybrid cloud is a cloud computing environment in which an association provides and controls some of the interior resources and others provide the exterior resources. Yan Zhu., and Shanbiao Wang., (2012) [70] emphasize that this environment could can carry the irreversible losses to the clients due to the need of integrity verification method for distributed data outsourcing.

Moreover, the focus is on collaborative integrity verification method in hybrid clouds so as to carry the scalable service and data migration. The existing system upholds the client data with multiple cloud service providers. The authors establish collaborative provable data possession method to get used to the homomorphic authenticates responses and hash index hierarchy. The authors use optimization to establish a security, based on multi-prover zero-knowledge proof system.

Currently, many organizations outsource data storage in the cloud so that a member or the owner of an association can simply share the data with the other members. Boyang Wang., et al., (2011) [7] introduce trouble-free and effective publicly verifiable system to allow the data integrity in the cloud without given up the anonymity of neither the data owners nor the requiring important verification Meta data. The authors establish that a security mediator (SEM) is capable to construct the signature on the extracted data for the data clients. This method decouples the anonymity protection method from provable data possession. The authors also enhance the multi security mediator used to avoid the potential single point of failure.

Cloud computing is defined as a deliverance of product quite than service. Cloud computing is to share the service based on internet. Lots of users position their data in the cloud. Nevertheless, the reality of the users still have physical possession of probably huge size of sensitive data creates the data

integrity. Jachak K.B., et al., (2012) [32] concentrate on data integrity and security in cloud computing environment. Security is achieved by signing the data block by to sending to cloud. Boneh–Lynn–Shacham (BLS) Algorithm, used to perform the signing system is more protected when compared to the other algorithms. The third party auditor is used to verify the data and the accuracy of the data in cloud data storage. Even the authors use a public key based homomorphic authenticator to reach the protective level with arbitrary masking privacy and maintaining public auditing. Bilinear aggregate certificate system is used to attain batch auditing which minimizes the computational overhead.

Managing the assignment of the virtual machine in a cloud at the physical resource is a solution requirement for the success of clouds. Imad M. Abbadi., and Anbang Ruan., (2013) [24] introduce the innovative cloud scheduler considering equally the requirement of the user and infrastructure properties. In addition, the system focuses on guaranteed users and their virtual resources are hosted using the physical resource contest and their necessities without receiving the users concerned with considerate details of cloud infrastructure. Work exposed prototype is based on open stack and cloud scheduler providing trust management of cloud and trust status of the cloud infrastructure.

Cloud computing is to facilitate extremely scalable services to be effortlessly consumed over Internet on as-needed center. A main characteristic of cloud services is that the users' data are frequently processed remotely in indefinite machines that users do not own or control as losing the data is an important issue in cloud environment. Smitha Sundareswaran., et al., (2012) [56] introduce a new highly decentralized information accountability structure to carry facilitates real usage of user data in cloud environment. The work is established on object-centered method to provide a logging method in cooperation with the user's data and policies. And Besides, it also provides distributed auditing mechanisms.

Cloud computing inexpensively allows the customers with restricted computational resources to outsource the large-scale computations in Cloud. Nevertheless, how to defend the customers' secret data concerned in computations becomes a main security concern. A secure outsourcing method is used to recover large scale of liner equation in cloud environment. The authors have applied conventional approaches like Gaussian elimination or LU decomposition. Cong Wang, et al., (2012) [10] build a secure LE outsourcing method using fully different iterative method, which is easier to execute and demands comparatively simpler matrix-vector operations. Particularly, this method allows the customer, to securely connect to the cloud for iteratively discovery consecutive approximations to LE solution. The author introduce numerical features of matrix-vector operations for robust

cheating detection in cloud environment. An efficient result verification method used to tolerate customer to confirm all answers is obtained from the preceding repetitive estimates in one group with elevated probability.

2.3.2. Certain Load balancing Technique with Minimal Bandwidth Usage

Some recently distributed applications need cooperation between the multiple geographically separated computing services. Moreover, it executes intensive computing at the end sites and huge scale data are transmitted in the wide area network. It has been extensively recognized that WDM networks are cost effectual so as to hold data transfers in these types of data-intensive applications. Nevertheless, neither conventional approaches to instituting light paths between specified source destination pairs. Xin Liu., et al., (2010) [65] develop a joint resource allocation in cooperation with the computing resource of network. This method provides a better resource allocation with up to date distributed computing application.

The recent distributed system is used to allow the users distribution and trade of the resource in the cloud computing environment. Users buy the various resources, such as bandwidth, computing power and storage system from one or additional cloud providers for a restricted period of time with changeable or fixed price. Haresh M V., et al., (2011) [22] use federated cloud system for resource sharing, with an improving scalability level. The resource allocation in the cloud is a challenging issue. The resource allocation agent

based method is used to enhance the resource allocation approach. In this method, the user need not know who the cloud service provider is and where the resources reside in. The consumer gets resource with a, fewer prices. This resource allocation method holds three agents, such as the consumer agent, the resource brokering and the provider agent.

Cloud computing supports a wide range of computation power and storage capability to facilitate the clients to deploy the computation and data-intensive applications without the infrastructure deal. The privacy preserving in the intermediate data sets suit demanding issues since adversaries can improve the privacy-sensitive information by evaluating multiple the in-between data sets. Xuyun Zhang., et al., (2012) [66] provide are upper bound privacy leakage constraint-based method which is used to discover an intermediate data set which is to be encrypted. The privacy preserving cost can be reduced as the privacy requirement in data owner be able to satisfy.

The cloud computing system is a raising computing system to enable users to distantly load their data into a cloud so as to benefit from the scalable services on-demand. Guojun Wang, et al., (2010) [18] analyse the enterprises in using with a restricted budget, to attain cost saving and improvements of productivity, through cloud-based services. Moreover, the work introduced Hierarchical Identity-Based Encryption (HIBE) System, used for enterprises to competently distribute the confidential data on the cloud servers. The authors

also attain to combine Hierarchical Identity-Based Encryption and cipher text-policy attribute-based encryption (CP-ABE) for improving the performance level. They have the applied proxy re-encryption and lazy re-encryption for cloud data integrity and confidentiality.

For example, considering the personal health record in the performance analysis of cloud based protection can prove better performance in outsourcing sensitive data. The personal health record is a promising patient-centric model of health information trade, which is frequently outsourced to be stored at the third party. Nevertheless, there have been broad privacy concerns as personal health information can be descriptive to those third party servers and to illegal parties. Ming Li., et al., (2012) [48] Verify the patient-centric framework for data access control method in PHR. Moreover, attribute based encryption is used to encrypts the patient's data in PHR. This method is used to provide scalability in key management, flexible access, and competent user revocation. The authors concentrate on the multiple data owner state, and separate the users in PHR system into multiple security area and really decreasing the key management complexity for the owners and users. Multi authority attribute based encryption provides a high level of privacy in the patient data assurance in PHR. As a result these methods improve the security and efficiency in the cloud environment.

Cloud computing has been constructed upon and improvement of distributed computing, grid computing and virtualization. As the cost of all the tasks of cloud resources is dissimilar with one another, scheduling tasks to users in the cloud is not same as in the conventional scheduling methods. Mrs.S.Selvarani., et al., (2010) [44] mainly focuses on schedule task groups in the cloud computing environment. The resources encompass the different resource costs and computation performance. The authors introduce a cost-based scheduling algorithm and it is used for constructing efficient mapping of the tasks to the accessible resources in the cloud. This scheduling is used to enhance the communication and computation ratio by combination of user tasks, according to exacting cloud resource's processing ability and it sends the collection of jobs to the resource.

Muhammad Abdullah Adnan., et al., (2012) [45] have recommended a newly geographical load balancing methods for the data centers, hosting cloud computation to reduce the energy cost by the use of the present price variation across regions. The authors used flexibility from Service Level Agreements (SLAs) for distinguishing the loads, under covered latency needs. The authors introduce a geographical load balancing method to balance the load with cost saving. The authors examine how much workload is to be performed in each data center and how much workload is to be postponed and shifted to the other data centers, for power saving and conventional deadlines. The authors use offline arrangements for remote load balancing trouble with dynamic deferral.

This method provides online algorithms to decide the task of workload to data centers and relocation of workload between than data centers to adapt with the dynamic electricity price changes.

Cloud Computing has been predicted as the next-generation architecture of IT Enterprise. Besides, the cloud environment moves the application software and databases to centralized huge data centers, where managing of data and services may not be completely dependable. Qian Wang., et al., (2012) [10] attain high data integrity in cloud data storage environment. The system evaluates the job of allowing the third party auditor (TPA), on behalf of the cloud client, to confirm the integrity of the dynamic data stored in cloud environment. The authors use the data operation in data dynamic via more general representation, such as insertion deletion and block modification. The Merkle Hash Tree construction is used for block tag authentication. An effective handling of auditing task will improve the data integrity level of the cloud system.

The cloud computing is one of the main parts of the real world life. This service, supported by infrastructure is known as Internet Data Center (IDC). As demand for cloud computing services rise, energy is preserved by the Internet Data Center. Both academia and industry have compensated great concentration to energy is management of IDCs. Jianying Luo, et al., (2014) [31] point out that an Energy Cost Optimization Internet data center algorithm

is used to sequential the diversity of the electricity price. Dynamically schedule workload is to perform on IDC servers throughout an input queue. This algorithm helps to reduce the cost and assures service delay limits and improved workload drop if the service delay limit is adequately large.

2.3.3. Organizing Resource with Load balancing Factor

Zhiguo Wan, et al., (2012) [75] stress that cloud computing related variety of methods intends to organize resource with Attribute-Based Encryption (ABE) as that is used for outsourcing information. But attribute based encryption is not flexible with fine-grained access control for outsource information in cloud. In order to overcome the encryption problem, the work develops Hierarchical Attribute-Set-Based Encryption (HASBE) to increase the cipher text Approach along with Attribute Set Based Encryption (ASBE) for hierarchical structure to clients. This method not only attains scalability also acquires inherit flexible and well-designed access control, supporting the compound attributes of ASBE. Further, HASBE use much valuable information right to use termination occurrence with client elimination efficiently.

For the various multicore server processors across the clouds and data centers, collective performance of cloud can be optimized by load sharing and balancing. Energy efficiency is a significant issue for large scale server in present and future data centers. Multicore processor technology offers

innovative levels of performance and energy efficiency. Junwei Cao, et al., (2014) [30] offer better power consumption which is proved with performance constrained load distribution methods for the present and future large-scale data centers the in cloud system. The authors mainly focus on optimal power allocation and load sharing for the numerous varied multicore server processors across the clouds and data centers. The optimal power allocation and load distribution are used for optimization in cloud environment. As a result, this method is used to achieve power and performance optimization in cloud data center.

Cloud computing is an effectiven computing model that has developed into a solid base for a broad array of enterprises and end-user applications. Mohamed Abu Sharkh, et al., (2013) [43] in their work allows the providers to offer changeable service portfolios that vary in resource configurations and provider services. A widespread solution used in resource allocation is essential in cloud computing service provider. Any resource allocation form has to be regarded as a computational resource in addition to the network resources to exactly imitate the practical demands. Another feature, that should be reflecting on as provisioning resource is energy consumption. This feature is receiving more concentration from the industrial and government parties. The authors have used resource allocation algorithms to achieve the task scheduling of virtual machines on servers inhabiting in data centers. The author have

provided a reference when designing a complete energy-aware resource allocation form for cloud computing data centers.

Yang Tang, et al., (2012) [71] have designed File Assured Deletion (FADE) to secure the overlay cloud storage mechanism to attain fine-grained, policy-based access control and file assured deletion. This mechanism related outsources files with the access system, and confidently deletes the files to create them unrecoverable to each one upon revocations of the file access strategy. The authors attain the security goal, FADE is based on a set of cryptographic key operations which are self-sustained by a quorum of key managers that are self-governing of the third-party clouds. Particularly, FADE System performs as an overlay method in the cloud data storage service. The FADE Mechanism provides security defense for the outsourced data, and initiates only the smallest performance and financial cost overhead. This mechanism provides to integrate the value added security features in the cloud storage service.

The Load balancing in the cloud computing environment has a significant substance on performance. The authors achieve better load balancing in the cloud environment to enhance the user satisfaction level. Xu, and Gaochao, (2013) [67] are able to expose a better load balancing technique model for the public cloud system dependent on cloud partitioning with switch mechanism to select different levels for the different conditions. This load balancing algorithm is applied in game theory to balance the load and

enhance the efficiency level in the public cloud system. The load balancing is done with the use of the main controller and balancer.

The controller first allocates the jobs to the appropriate cloud partition and then communicates with the balancers in all the partitions to restore the status information. The balancers in each partition collects the current status information from all nodes and then decide the right level to allocate the jobs. The Cloud computing environment has various problems in resource management and load balancing technique. Load balancing is a procedure of balancing the workload between the multiple nodes. Jamuna and Anadha Kumar (2013) [29] have an formulated an optimized load balancing technique with Cloud Clustering Technique and it is used in a much better way in cloud environment. The public cloud environment has massive nodes over huge and various geographical locations. This techniques partitions the cloud environment into several partitions by creating use of cloud clustering method and assists the providers to reduce the process of load balancing. As a result load balancing method provides higher performance and stability in the cloud.

In a company of many inexpensive advantages of cloud computing, many organizations have been allowing to move their information systems to the cloud. Nevertheless, a significant problem in public clouds is how to selectively distribute the data, based on fine-grained attribute based access control strategy. Mohamed Nabeel and Elisa Bertino (2012) [42] in their work assure confidentiality of the data and discuss how to conserve the privacy of

users from the cloud. The authors concentrate on two access control systems, such as Attribute Based Access Control (ABAC) Scheme and fine-grained access control, based on the strategy, using the individuality attributes, over the encrypted data. While dealing with the third-party cloud services, an essential issue is individual attributes in the access control strategy may disclose the privacy-sensitive information about the users, organizations and leak the secret information about the content.

The propagation of cloud computing has the support of the broad deployment of large-scale datacenters with remarkable energy consumption and high carbon emission. The authors concentrate to reduce the energy cost and carbon footprint, and the rising number of cloud service providers have measured the green data centers with renewable power sources, such as solar or wind. Nevertheless in the different stable provider of grid energy, it is demanding to use and recognize the renewable energy due to its uncertain, irregular and variable nature. Wei Deng, et al., (2014) [63] provide taxonomy of the state-of-the-art research in applying renewable power in cloud computing datacenters from five types of features, such as generation models and prediction technique of renewable power, ability planning of green datacenters, intra-datacenter workload scheduling and load balancing across geologically dispersed datacenters.

The power management is a becoming an increasing significant problem for the internet service supported by the multiple geo-dispersed data center. Rahman, et al., (2014) [52] discuss that the data centers energy saving and cost are becoming inappropriately high and important it increases the load on both the energy resource and cloud environment. The smart grid offers a possible way for dynamic and competent power management of data centers. The power management based geographic load balancing is used effectively operate many features of the smart grid.

As a known fact, cloud computing is the latest and imminent paradigm that suggests huge benefits, such as abridged time for marketing limitless computing power and flexible computing abilities. It is a representation to offer an on-demand network access to a mutual pool of computing resources which include a large number of Load Balancing, Scheduling, etc. Divya Chaudhary and Rajender Singh Chhillar, (2013) [13] introduce load balancing mechanism to share the workload consistently to all the nodes in a system in order to attain a high resource of utilization and user satisfaction level. It assists in the allocation and de-allocation of the occurrence in the applications without failure. The main task is to competently and efficiently use the infrastructure obtainable for the cloud for processing the users' requests. It is a process of reconvening the entire load to the individual nodes of the shared system to construct the resource utilization effectual and to get a better response time for

the job. At the same time, it eliminates the imbalances in the modes as some are overloaded and some others are underloaded.

Cloud computing is to be organized in data centres where substantial machines are virtualized. More number of virtual machines have exceeding virtualization. Ajith Singh and hemalatha, (2012) [3] concentrate on load balancing approach. The data centres are constructed with lots of systems, where balancing is a complicated task, especially for cloud computing. The authors introduce the semi-distributed load balancing in cloud computing. This load balancing algorithm is used to get better load balancing during cloud computing when it is applied to each central part of the clusters.

In this modern, highly technologized the appearance of cloud computing and its applications and uses in load balancing have been elevated to the utmost level. The number of the users' right to use this service is rising drastically day by day. As the cloud is prepared to datacenter is extremely powerful to hold large numbers of users at rest with load balancing is essential. Soumya Ranjan Jena and Zulfikhar Ahmad, (2013) [57] have enhance the process of load balancing algorithm and distributed the load by reducing the response time cost, energy utilization and overhead.

The authors concentrate on load balancing in cloud environment. The major load balancing problems in cloud computing are load calculation and load distribution. The authors have handled many load balancing techniques to

disperse the tasks accurately. Pranesh Das et al., (2013) [50] have introduced in their work better Load Balancing Technique for Virtualization and Fault Tolerance in Cloud Computing (LBVFT) so as to allocate the task to the virtual nodes. The system facilitates the cloud manager and the decision maker is in a better position to handle virtualization, load balancing and the fault, in the cloud. Besides Load Balancing Technique for Virtualization and Fault Tolerance is to allocate the tasks to the virtual nodes, based on their success rate and prior load status. The load allocation is done by the load balancer of the cloud manager, based on its higher success rate and lower load obtainable nodes.

2.4. DIFFERENT TECHNIQUES IN MULTI-TASK HANDLING

Most of the scheduling algorithms intend in balancing many tasks with minimum bandwidth and energy utilization. Even the performance of a single task handling efficiency boosts the multi-task scheduling. The foremost thing in managing the multi-task is the appropriate equalization of workload. Each cloud node accepts a suitable load in order to perform the multi-task. The Various existing techniques propose cluster based scheduling for multi-tasking, with the objective of energy consumption and lower bandwidth utilization criteria. All the current schemes use distributed computing resources to justify a better tasking operation with an efficient user communication.

2.4.1. Clustering Related Multi-tasking in Cloud Environment

The scheduling is a many task flow in the distributed computing environment with distinguished NP-hard problems. The problems lie in fulfilling the multiple objectives that may be contradictory in nature. For illustration, it is complicated to reduce the Makespan of many tasks, as reducing the resource cost and conserving the fault tolerance and quality of service (QoS) are hard. Fan Zhanga., et al., (2013) [16] reveal a better suboptimal or sufficiently good schedule algorithm for the smooth multitasking workflows on of the cloud environment. Moreover, Multi-Objective Scheduling (MOS) Method is used in the cloud, depending on ordinal Optimization (OO) Scheme. The authors expand OO scheme to meet the exceptional demands from the cloud environments that apply to the virtual clusters of the servers of the multiple data centers. As a result, this method is used to reduce the scheduled overhead time and increase optimal the performance.

The most famous and common technique is the provable data possession that is for guaranteeing the data integrity in storage outsourcing. Yan Zhu., et al., prove (2012) [70] that the data possession is sufficient for the shared cloud storage to support the scalability of the service and data migrations. As an enhancement of the provable data possession, a novel cooperative provable data possession is designed with homomorphic verifiable response and hash index hierarchy. The cooperative provable data possession proves security, the

based on the multiprover zero-knowledge proof system so the to suit its comprehensiveness, knowledge soundness, and zero-knowledge properties. Moreover, the system performance optimization mechanisms of the efficient methods select the optimal parameter values with the minimum computation costs of clients and storage space service providers. As a result, this method reduces computation and communication overhead level.

In IaaS cloud computing, the computational resources favour the remote users in the form of a rent. For a cloud user, the service provider requests the multiple cloud services at the same time. Jiayin Li., et al., (2012) [28] exploit parallel processing in cloud computing, which is essential to execute the mechanism to assign the resource and schedule execution order of the tasks. Besides, a resource optimization system used with the preemptable task execution is exposed with improvement in the utilization of the resources in the cloud infrastructure. The authors introduce two online dynamic resource allocation algorithms for IaaS cloud environment with preemptable tasks. This algorithm is used to modify the resource assignment, dynamically depending to modernize the information of the definite task execution.

Konstantinos Tsakalozos., et al., (2013) [37] represent that IaaS clouds offer complete virtual infrastructures for the dispersed processing concealing all the physical underlying machinery. Nefeli is appropriate for virtual infrastructure gateway that lifts this constraint. Throughout Nefeli, the

cloud customers provide deployment hints on probable mapping of the VMs to their physical nodes. Such hints comprise collocation and anticollocation of VMs. The consumers allocate only the features of the virtual infrastructure and stay at all times disbelieving the to cloud center physical characteristics. The number of consumer offers hint is increased with a high level placement strategy particularly by cloud administration. Finally, Nefeli incorporates the events and alters the virtual machine deployment, rendering them with a fewer bottlenecks so as to enhance the quality of the rendered service.

Cloud computing involves solution requirement for resource configuration in a real-time manner. Xian ping Bu., et al., (2013) [68] have investigated the virtualized background and both the virtual machines and hosted applications are required to be configured to adjust the system dynamics. With much investigation, the Co Tuner is developed from the coordinated configuration of the virtual machines and inhabitant the applications. This system is the main structure of the standard with model free reinforcement learning system. The authors add simplex method and reinforcement knowledge with improving system knowledge.

Cloud computing provides pretty option for businesses to charge an appropriate size Map Reduce cluster, use resources as a service, and balance only the resources that are devoted. A main confront in such environments is to enlarge the utilization of Map Reduce clusters to reduce the cost. Abhishek

Verma., et al., (2013) [4] have attained the goal to enhance the execution of Map reduces jobs on cluster. Moreover, the system executes offline analysis with a set of jobs on the new data to assess the performance profit of the various optimization techniques. Besides it employs a simple abstraction in each Map reduce job that symbolizes a pair of map minimize stage duration. However, the focus is made only on the optimal two stage job scheduling using Johnson Algorithm.

2.4.2. Resource Allocation and Job Execution in Cloud Environment

Yong Cui., et al., (2013) [69] Stress that the unbalanced traffic demands of dissimilar data center applications are significant problems in designing Data Center Networks (DCN). A hybrid data center network is adopted for a better resolution to exploit the wireless transmissions in data center networks. Researchers concentrate on the obstruction problem, caused by only some hot nodes to get a better global performance. DCN model suits the wireless transmission in data center network by allowing, both the wireless interference and adaptive transmission rate.

Alexandru Iosup., et al., (2010) [5] say that cloud computing is rising a profitable infrastructure paradigm that assures to remove the requirement for the sustaining expensive computing facilities by companies and institutes alike. Throughout the use of virtualization and resource time distribution, clouds serve with a single set of physical resources and huge user base with dissimilar

needs as discussed. Nevertheless, the present commercied clouds have been constructing to support the web and small database workloads, which are very dissimilar from the characteristic scientific computing workloads. Furthermore, use of virtualization and resource time distribution may begin significant performance penalties for challenging the scientific computing workloads. The work enumerates presence in real scientific computing workloads of the multi-task computing users. The users who employ insecure coupled applications encompass many tasks to attain the scientific goals.

Haiyan Guan., et al., (2013) [23] have proved that the light detection and ranging methods to be more special tools to gather, within a small time, three dimensional points. The clouds used with high density, high accuracy and significantly specify outside information pertaining to the environmental objects like the feature extraction and 3-D reconstruction in CAD system. The authors introduce and innovative system of network based computational resource to proceed, control and share the enormous quantities of the idler data in the cloud. The process of virtualization scheme is rising to a high level and has been efficient in the organization of the huge scale lidar point clouds.

Cloud Computing is referred to as computing as a service rather than a product, such as shared resources, software information, etc. Cloud computing can be used to send outing user tasks or jobs to obtain the system resources, like storage and software. V. Venkatesa Kumar and K. Dinesh., (2012) [60]

point out that the Scheduling algorithm is used to task the dispatching of the users. Moreover, they explore job scheduling using fuzzy neural network algorithm. The user task can be split based on QoS parameter, such as bandwidth memory, CPU utilization and size. The split tasks are specified to the fuzzier input values, which are converted in to a range of 0 and 1. The NN consists input layers, hidden layers and output layers to regulate the weight of the task user so as to be equivalent with the system resources. The de-fuzzier is an overturn operation of the fuzzier system.

Hamzeh Khazaei., et al., (2012) [21] elaborate the concert of cloud centers with an elevated degree of virtualization and poisson batch task arrivals. They incorporate the logical model and authenticate it with a self-governing simulation model. Task service times are representation of a common probability distribution. But representation also accounts for a descent in performance, which is appropriate to the workload at each node. Besides, the author shows that the behavior a of they cloud center may be enhanced, if received requests are divided based on their coefficience in service time and batch size.

A single cloud service cannot satisfy all the user needs, a composition of cloud services is essential. Cloud service composition involves various tasks, such as recognition, compatibility checking, assortment, and deployment. Cloud is a compound process and users discover it a complicated process and

they struggle to choose the best one among the hundreds, if not thousands, of probable compositions accessible. Amir Vahid Dastjerdi., and Rajkumar Buyya., (2012) [6] represent the structure based algorithm to shorten the cloud service composition for the unskilled consumer. Employed ontology has been based on to examine the cloud service compatibility by applying motivation on expert knowledge as well as to reduce the efforts of the users in expressing preference, author applied combination of evolutionary algorithm and fuzzy logic composition optimization.

Cloud Computing is an interesting and helpful area in up to date distributed computing. Even cloud gives an improved way of scheduling jobs in limited conditions. Jianhua Tang., et al., (2013) [27] the scheduler sustains balances between QOS and equality among jobs so efficiency may be enlarged as elaborated in M. Vijayalakshmi., and Venkatesa Kumar., (2014) [40]. As the number of cloud users rise exponentially the demand for resources, like virtual machines have also amplified. So, scheduling is the most important problem in cloud computing. The authors have achieved practical comparison of the different job scheduling algorithms by scheduling the virtual machines. The optimization technique is to decide issues in scheduling based algorithm.

2.4.3. Different Job Scheduling Algorithm in Cloud Computing Platforms

Cloud computing platform offers an illusion of endless computing resources to the cloud users because they can raise and reduce their

consumption rate according to their demands. Simultaneously, cloud environment creates a number of challenges. Two groups of actors in cloud computing environments, the cloud providers and the cloud users follow different goals; the providers desire to maximize the revenue by attaining a high resource utilization. Cloud users desire to reduce then fixed cost as meeting their performance requirements. Ms. Shubhangi and D. Patil1. as well as Dr. S. C. Mehrotra., (2012) [41] opine that it is complicated to assign resource in equally optimal way to lack of information distributed between them.

Jianhua Tang., et al., (2013) [27] elaborate the issues related to the optimally redirecting user requests in Cloud-Centric Media Network (CCMN) to the numerous targeted VM. The often scale their service abilities in order to reduce the cost function that incorporates the service response time, computing costs, and routing costs. Moreover, the work facilitates the request arrival process to switch between the usual and flash crowd modes to model user requests of CCMN. The review enumerates tradeoff in flash crowd identification delay and false alarm frequency, request assignment rates and service ability as virtual machine.

In cloud environment, the principle logic of implementation needs a resource allocation because of the high task of the resource share. Resource Scheduling is a difficult task in cloud computing environment since it has many

optional computers with unreliable capacities. Vignesh V., et al., (2013) [59] the author introduce a resource model, used for job-oriented resource scheduling in the cloud computing platforms. Resource assignment task is scheduled for processing and it gives obtainable resources and user favourite. The computing resources can be assigned according to the rank of the job. Moreover, the work creates analysis of the resource scheduling algorithm. Three algorithms, such as round robin, preemptive priority and the shortest remaining time achieve the time parameter.

Network I/O virtualization plays a significant function in cloud computing. En-Hao Chang., et al., (2013) [14] concentrate on the system broad virtualization problem of TCP/IP Offload Engine (TOE). The work involves three important factors that influence the performance of TOE: I/O, virtualization architectures, quality of service and virtual machine monitor scheduler. The device emulation is dependent on TOE and the VMM control socket connections is connected to TOE straightforwardly**.** It is to eliminate the packet copy and demultiplexing overheads as materialize in virtualization of layer 2 network card. The direct I/O access architecture offers per virtual machine, depending upon the physical interface control that assists in removing most of the VMM interventions. The authors establishe VMM scheduler with preemptive I/O scheduling and programmable I/O command dispatcher with deficit weighed round robin (DWRR) strategy**.** This method is efficient to make certain service fairness and simultaneously maximize TOE utilization.

Karthik, Ramaraj and Kannan, (2013) [38] point out that job scheduling is one of the complex problems in Cloud Computing platforms. Their work is intended to group the method so as to expand the combinational backfill algorithm dependenting on smadium and long queue method, using random fashion. This algorithm is used to improve the resource gap, minimize the system inactive time and support to attain high resource usage and offer the quality of system in cloud computing environment. In order to create most competent use of the resources, achieve optimization for cloud scheduling problems is used. It is not probable to forecast the job execution time in the cloud environment. Moreover, the work presents Optimized Resource Filling (ORF) correctly consuming resources and raising the unused obtainable working space and decrease starvation. When compared to the traditional and stability spiral method, the authors indent smadium for cloud resource management and ORF so as to make an effort to fill up the unused space produced by the scheduler.

The cloud environment is one of the most split able environments where multiple clients are associated to the common environment so as to right the use services and products. A cloud environment can be public or private. According to Harpreet Kaur and Maninder Singh (2013) [20] in such case, the method is required to execute efficient scheduling and resource allocation. Resource relocation is to switch to the assigned process in some added virtual

machine or cloud to liberate the load and to carry out effective execution of the cloud requests.

Cloud computing is a resolution for the dispensation of large amounts of data. So, Google has initiated map reduce as a programming representation for huge scale data applications in cloud platforms. Data processing and parallel computing are done by using Map Reduce. Apache Hadoop is an open source accomplishment of Map Reduce. Forough zare, et al., (2014) [15] opines that job Shop Scheduling Problem (JSSP) is a significant problem and that is one of the most popular NP hards. It is essential to find a quicker solution for the large scale problems. JSSP integrates fuzzy neural network with Map reduce representation to resolve the job shop scheduling trouble. As a result, the authors achieve better results in terms of high convergence speed with less execution time.

Pradeep kumar and Amandeep Verma (2012) [51] View that cloud computing is a new promising technology and is being accepted day by day because of its great features. The main significant problem in cloud computing is the scheduling of the users' requests so that the requested tasks can be accomplished with minimum time, according to user definite time. A good scheduling technique also assists in efficient utilization of the resources. Various scheduling algorithms, such as Min-Min, Max-Min, X-Sufferage, Genetic Algorithm, Particle Swarm Optimization etc are available, the author

use three scheduling algorithm such as Min-Min, Max-Min and Genetic Algorithm. The genetic algorithm is used to enhance the sample data set and to combine with Min-Min and Max-Min in genetic algorithm. As a final point, all these existing methods include their own advantages and disadvantages. The limitations in the existing works are elaborated the in next section.

2.5. RESEARCH GAP

The optimal joint resource allocation scheme is appropriate even for multiple resource types as the allocation task is simple and it is done at the same time to each service request. Similarly, fair joint resource allocation method is able to equalize the complete range of key resources, assigned for each service at each time block. However, the main issue of these schemes is that it does not support on prioritization of the tasks and the resources and assigned as per request.

The scheduling strategy that intends in achieving better load balancing of the VM resources is considerable in organizing the resources. In the scheduling strategy, much time complexity is taken for selecting the least-affective solution to achieve the optimal load balancing as well as dynamic migration. Though, the system delivers efficient load balancing in the virtual machine with scheduling and improves its resource utility, the time complexity is high.

The infrastructure as a Service model offeres better allocation of resources for the real-time tasks. The polynomial-time solution in a IaaS model is used for efficient resource allocation in cloud computing environment. But, this model fails in considering the delay of the task's time, as well as in noticing the huge rent.

The priority-based scheduling running in queues and multilevel feedback queue scheduling are useful for relocating the tasks without degrading the local job performance level. Priority based scheduling scheme achieves the computing power with clustering used effectively and assigned rather than between all available nodes. However, Priority based scheduling system is not convincing with respect to user requirements in the cloud environment.

The virtualization technology for the purpose of efficient data center resource allocate often fulfills the application on-demands. Moreover, virtualization technology holds up green computing by optimizing the load to the number of servers. Virtualization technology is used to expand a set of heuristics to avoid overloading the in system efficiently as energy consumption. But, virtualization technology does not support the task scheduling process in cloud environment.

The cloud scheduler manages the equal requirements of users and infrastructure properties and it ensures the users about of their virtual resources host using physical resources contest their necessities without receiving users concerned with considerate details of cloud infrastructure. The

resource scheduling is unsuccessful in establishing a trustworthy collection and calculation of other properties.

The cost-based scheduling algorithm elaborated in literature for constructing efficient mapping of tasks supports better access resources in cloud. Cost-based scheduling is used to enhance the communication and computation ratio by a combination of the user tasks. However, the improvement of algorithm does not concentrate on independent task scheduling in the cloud environment.

Nefeli is used for virtual infrastructure gateway that lifts this constraint. Throughout Nefeli, the cloud customers provide deployment hints on the probable mapping of VMs to physical nodes. Such hints comprise collocation and anticollocation of the virtual machines**.** Nefeli fails to offer deployment hints with effective handling of the deployment of virtual infrastructures in the context of real large cloud installations.

The resources needed for the computation in network interrelated model for cloud computing platform is more done with Condor-based process virtualization platform. Moreover, in Condor-based process virtualization, the resources handle vast quantities of sensitive data. But, Condor-based process virtualization model fails to provide a parallel distributed and network based processing workflows.

The arrival process based on request of the user between normal and flash crowd model are redirected to VMs in order of non-decreasing prices.

Elastic Service Scaling is employed in flash crowd model at Cloud-Centric Media Networks struggles to achieve multiple dispatchers. In addition, there are needs for additional focus on the design of distributed redirection and service scaling strategies. Therefore, the proposed works consider the above limits, depending on which proposals emerge as appropriate techniques to fulfill the drawbacks with contribution as detailed in next section.

2.6. CONTRIBUTION OF THE THESIS

The main contributions of the thesis are as follows

(i) To enhance the resource scheduling process with the proposed Inference Aware Resource Allocation (IARA) Based Resource Scheduler

(ii) To allocate the resources with optimal energy and bandwidth consumption by using IARA Technique, which resolves the sub-optimization problems

(iii) To improve the energy consumption in cloud infrastructures, using Adaptive Load Balancing Approach (ALB)

(iv) To balance the load in the resource allocation process, through cluster formation by employing ALB Algorithm

(v) To deal the workload management on multi-tasking with the development of Genetic Clustering with Workload Multi-task (GCWM) Scheduler Scheme

(vi) To reduce the computational cost and complexity involved during computation using genetic concepts of GCWM Scheduler

(vii) Finally, to ensure multi-tasking operation with efficient user communication through GCWM in the distributed computing resources.