Journal of Network and Computer Applications 68 (2016) 173–200

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Review

Resource scheduling for infrastructure as a service (IaaS) in cloud computing: Challenges and opportunities

Syed Hamid Hussain Madni*, Muhammad Shafie Abd Latiff, Yahaya Coulibaly, Shafi'i Muhammad Abdulhamid

Faculty of Computing, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia

ARTICLE INFO

Article history:
Received 5 August 2015
Received in revised form
19 April 2016
Accepted 19 April 2016
Available online 22 April 2016

Keywords:
Resource management
IaaS cloud
Resource scheduling
Resource discovery
Resource allocation

ABSTRACT

Resource scheduling assigns the precise and accurate task to CPU, network, and storage. The aim behind this is the extreme usage of resources. However, well organized scheduling is needed for both cloud providers and cloud users. This paper is a chronological study of recent issues of resource scheduling in IaaS cloud computing environment. In our study, we investigate resource scheduling schemes and algorithms used by different researchers and categorize these approaches on the basis of problems addressed, schemes and the parameters used in evaluating different approaches. Based on various studies considered in this survey, we perceive that different schemes and algorithms did not consider some essential parameters and enhancement is requisite to improve the performance of the existing schemes. Furthermore, this study will trigger new and innovative methods of handling the problems of resource scheduling in the cloud and will help researchers in understanding the existing methodologies for the future adaptation and enhancements.

Contents

* Corresponding author.
E-mail addresses: madni4all@yahoo.com (S.H.H. Madni), shafie@utm.my (M.S.A. Latiff), coulibaly@utm.my (Y. Coulibaly), shafii.abdulhamid@futminna.edu.ng (S.M. Abdulhamid).

http://dx.doi.org/10.1016/j.jnca.2016.04.016
1084-8045/

1. Introduction

Cloud computing technology is virtualized, became a milestone and deals with many services across the Internet. It's primarily intentions are at high performance, scalability, fault-tolerant, high availability, utilization, reliable and easy to use, monitor, manage, and provision efficiency and economical. The hardware is compact to a service, fully incorporated and hardware agnostic software stack is provided as a substitute of Quality of Services (QoS) (Sandholm and Lee, 2014). From the cloud providers, large scale of resources need to be allocated and assigned to thousands of distributed cloud users, dynamically, fairly and most importantly in a profitable way. From the cloud users point of view, user economy is a driven entity when they make the decision to use the cloud services. The major issues that are commonly interrelated with IaaS in cloud computing are resource management, data management, network infrastructure management, virtualization and multi-tenancy, application programming interfaces (APIs), interoperability, security, etc. (Manvi and Krishna Shyam, 2014, Mustafa et al., 2015, Durao et al., 2014, Yang and Tate, 2012).

In the perspective of cloud computing, resource management is the procedure of distributing the computing, networking, storage, nodes and virtual machines (VMs) to a set of applications. In a way that it tries to achieve the performance objectives of the applications, the cloud resources for the providers and users jointly. The cloud providers' main objective is efficient and effective resource utilization within the limits of Service Level Agreements (SLAs) with the cloud users. Efficient resource usage is classically managed through virtualization technologies, which enables statistical multiplexing of resources through clients and requests (Jennings and Stadler, 2014). Resource scheduling assigns the precise and accurate task to CPU, network, and storage. The main aim behind resource scheduling is the extreme usage of resources. However, well organized scheduling is needed for both cloud providers and cloud users (Achar et al., 2012). The most challenging problem is the resource scheduling for IaaS in cloud computing, is handling and providing efficient utilization of resources (Manvi and Krishna Shyam, 2014). The problem has attained a lot of concentration from the research groups in the last few years. Hence, the source of limited resources, environmental requirements, resources heterogeneity, locality limitations and dynamic types of resource demand. Therefore, it is essential to have an efficient and effective resource scheduling procedure that is best applicable for the cloud computing environments.

Previous review and survey papers focused on a particular domain like mobile cloud computing (Dinh et al., 2013; Qi and Gani, 2012), energy efficiency in data centers (Kaur and Chana, 2015; Mastelic et al., 2014), load balancing (Gabi et al., 2015; Kansal and Chana, 2012), resource allocation (Madni et al., 2016; Ma et al., 2014), resource provisioning (Zhang et al., 2016) resource scheduling (Kalra and Singh, 2015; Tsai and Rodrigues, 2014; Zhang and Su, 2014) in the cloud computing (detailed in Section 2).

Our contributions in this paper are as follows:

- We present the classification schemes and an updated descriptive literature review of resource scheduling for cloud computing research.
- We review and analyze the different issues and approaches used in the resource scheduling techniques, while highlighting their benefits and weaknesses.
- We highlight the performance metrics used to evaluate existing literatures.
- We summarize the suggested future works as stated in previous researches, which helps to shape the direction of current and future works.

However, analyzing existing techniques and understanding their focus work is necessary for developing some additional applicable technique that can be an improvement of the existing techniques to take advantages from earlier studies. This paper helps future researchers to understand clearly the recent status, needs, future requirements and to locate the loopholes responsible for inefficiency in resource scheduling in cloud computing.

The remaining sections of the paper are structured as follows: Section 2 is based on some related survey and review of researches in resource scheduling in the cloud computing. Section 3 describe the scenario of resource scheduling problem in cloud computing and classification of resource scheduling schemes. Section 3 focuses on the analysis of the study. The parameters used to evaluate existing literatures are presented and analyzed in Section 4. In Section 5, we suggest some future research areas in cloud computing environment and present the conclusion and recommendations in Section 6.

2. Related works

The cloud providers and users requirements in cloud computing is surveyed and explained in detail by (Venters and Whitley, 2012). The review of the technology has been done by interviews with the cloud providers and users having distinct requirements. It also states about the features of the cloud computing and the users' demand. It provides two main dimensions, the technological and the service dimension. The main challenges in research of cloud computing are the efficiency, creativity and simplicity, with considering the knowledge and the trust among the cloud providers and users. Furthermore, the main intention of a review is to appraise the different concepts and scheduling algorithms operated for the on demand grid as a service (GaaS) cloud related to the scheduling parameters used by current researches by Abdulhamid et al. (2014a). Researchers also study the cloud infrastructures, grid middlewares and the problems identified by numerous researchers in previous studies.

Tsai and Rodrigues (2014) review the literatures about metaheuristic scheduling techniques for cloud computing and present the applications using meta-heuristics, some main issues and challenges. Moreover, provides the instructions for researchers to move on meta-heuristics techniques instead of traditional scheduling technique in cloud computing. Further, Kalra and Singh (2015) deliver an inclusive survey and comparative analysis of various meta-heuristic scheduling techniques for cloud and grid environments including the Ant Colony Optimization (ACO), BAT algorithm, Genetic Algorithm (GA), League Championship Algorithm (LCA) and Particle Swarm Optimization (PSO). Optimization metrics, observations, open issues and challenges for further researches in cloud computing are also presented.

However, Madni et al. (2016) present an appraisal of various types of resource allocation meta-heuristic algorithms that have been utilized in IaaS cloud computing environment. Furthermore, elaborates the various issues addressed through the resource allocation meta-heuristics algorithms, the comparative parameters and also the experimental tools used for validation of the various techniques. The review and classification can serve as a foundation for further researches in resource allocation in IaaS cloud computing. Similarly, Zhang et al. (2016) delivers a comprehensive summary of the modern resource provision algorithms, intending on the state-of-art, existing techniques, and addressed objectives in different deployment stages. Some newly issued features of cloud computing and insufficiency of the existing techniques are highlighted to inspire future research guidelines.

The classifications and terminologies for energy efficiency by a systematic review of cloud computing is presented by Mastelic et al. (2014). It provides a summary of previous technologies, research work, and projects for all domain of information and communications technology hardware supporting in data centers and their equipment in the cloud computing. Finally, point out the existing challenges and future research guidelines. Kaur and Chana (2015) also systematically evaluates and reviews the techniques relating to energy efficiency become conscious through cloud computing. Based on the analysis, it can be concluded that the best software based energy efficient solution to minimize energy consumption could be obtained through energy-aware job scheduling to appropriate resources. Thus, the survey is a study of existing energy aware resource scheduling techniques in cloud computing. The techniques surveyed attain a desired level of performance based on different metrics, with the primary attention on energy savings.

Kansal and Chana (2012) present a systematic review of existing load balancing techniques for cloud computing. The review determines that all the existing previous techniques are generally emphasized on decreasing interrelated workload, response time and enhancing the performance. Several parameters are defined and used to contrast the current techniques. Conversely, Gabi et al. (2015) discover that the existing load balancing task scheduling techniques for cloud computing have failed to address scalability as an important issue. Some of the recent techniques are only good for the homogenous cloud computing. However, none of the existing techniques function effectively in federal cloud computing environment.

An overview of Mobile Cloud Computing (MCC), which includes the architecture, applications, challenges and their existing solutions is provided by Dinh et al. (2013). MCC is one of the emerging trends that provide the advantages of both mobility and remote computing. The paper discusses the architecture and applications supported by MCC that includes commerce, education, healthcare, mobile gaming, etc. The open issues related from communication and computing sides of the MCC have been highlighted like network access management, quality of service and service convergence. Furthermore, in another similar work by Qi and Gani (2012) discuss more about the challenges and future work in MCC. The major challenges are emphasized on limitations of mobile devices, quality of communication, division of application services, data delivery and task division.

Resource scheduling and allocation play a vital role in cloud computing mostly to develop execution efficiency and utilization of resources, energy saving, users QoS requirements satisfaction and increasing the profit of the cloud providers. Furthermore, its algorithm and policy directly influence the cloud cost and performance. Ma et al. (2014) discuss five important issues of resource scheduling and allocation in cloud computing including the locality aware task scheduling, reliability aware scheduling, energy aware resource allocation and scheduling, SaaS layer resource allocation and scheduling, and workflow scheduling. Moreover, they perform a discussion and comprehensive analysis of numerous current resource allocation and scheduling policies and algorithms of the existing problems in terms of diverse parameters. In addition to this, Ma et al. (2014) also mark a thorough analysis of five problems with their algorithm. However, Zhang and Su (2014) review the research performance and general cloud data center resource scheduling procedures with their issues. Initially, define the concept of cloud data center architecture and resource scheduling. Then, describes the procedure for the resources scheduling for the cloud data center, with aspect to the efficient scheduling of cloud resources and low power scheduling of tasks. The paper indicates the issues in the existing area of cloud data center related to the multiple resource scheduling. The profit and utilization of resources are very less for the cloud providers and the usage of energy in the data center is much more. Therefore, there is still a need to enhance the resource scheduling for the data center for future research.

Huang and Ou (2014) explain the PSO is a kind of replicated evolutionary computation algorithm of the congregate foraging behavior in cloud environment by using the PSO for task scheduling and it can get favorable results. Moreover, the ACO has excellent and well manufactured computing process along with the advantages of extension as it consists of a significant part of cloud environment task completion and scheduling issues. Furthermore, the study analyzes the standard and appliance with layered structural design of CloudSim. The architecture includes the resource layer, code layer, service layer, VM service layer, network layer, and structure interface layer. Depending on the work, considering the task scheduling as a research object in the cloud environment. CloudSim is anticipated and stretched by the five task scheduling algorithms. After that, a specific cloud simulation prospect is prepared, and simulations are conducted five times for each algorithm. Simulation and the enhancement of task scheduling approach for cloud computing in CloudSim come to be a valuable way for the researchers (He et al., 2013).

Resource scheduling indicates the process of organizing the resources among the different cloud users according to certain rules and regulation of resources usage under a specified cloud environment. Resource scheduling in resource management is the basic technology of cloud computing. It reviews the algorithms to increase the performance, involving with dynamic scheduling depend upon the threshold, optimized genetic algorithm with double suitability and enhanced ant colony algorithm for scheduling proposed by Huang et al. (2013). While the following areas are involved in Mapreduce scheduling research like, graph methods, utility based optimization, dynamic priority, customization, temporary weight modifications, prediction, adaptive scheduling, equality between several users, study of mapreduce inter dependence and improving the reducing phases. The main task is to improve the overall performance, enhance throughput and response time, provides improved locality and fairness. The open area of research for new application, improvement of the makespan and enhancing the fairness among multiple users (Elghoneimy et al., 2012).

3. Resource scheduling

This section explains the resource scheduling problem in cloud computing environment and also presents a classification of the resource scheduling schemes.

3.1. Resource scheduling problem

In simple words, scheduling is a way of determining which activity should be performed. It is a demand of resources to the applicable consignment of tasks to the resources accessible for processing, network and storage, such that there is an extreme usage of resources. Well-organized resource scheduling is compulsory for cloud providers besides cloud users, which is also the key technology of cloud computing (Achar et al., 2012). Scheduling algorithms are generally used to decrease the execution cost and time. Scheduling tackles the problem of which resources needed to be allocated to the received task or cloudlet. Efficient scheduling algorithms should consider the balancing the load of the system, the total execution time of the available resources, quick recovery, fault tolerance and migration with no interruption of service (Javanmardi et al., 2014).

Similar to the general scheduling problem, resource scheduling problem in cloud computing is also to find the optimal planning, it can explain with the help of an expression

$$\sum_{x=1}^{m,n} (R_x + S_x) \times T_x \rightarrow U_x^{\mathbf{Z}}$$

that assigns m required numbers of cloudlets/task (or virtual resources) $T = (T_1, T_2, T_3, \ldots\ldots\ldots, T_m)$ onto n available physical resources in cloud data centers $R = (R_1, R_2, R_3, \ldots\ldots\ldots, R_n)$ and $S = (S_1, S_2, S_3, \ldots\ldots\ldots, S_n)$ to the cloud user $U = (U_1, U_2, U_3, \ldots\ldots\ldots, U_n)$ such that the fitness of z particular objectives $F = (F_1, F_2, F_3, \ldots\ldots\ldots, F_z)$ are maximized.

The problem of resource scheduling in cloud computing is illustrated in Fig. 1. Furthermore, the scheduling of resource distribution is based on various aspects for example, how many quantity of resources is required in overall scenario, the number of cloud providers providing the services, how much availability of resources are available in their data center and many other conditions. The following objectives are often considered for the optimal resource scheduling in cloud computing those are cost, time, makespan, QoS, energy, load balancing, availability, reliability, failure rate, etc.

3.2. Classification of resource scheduling

The resource scheduling schemes are classify into six hybrid categories including cost aware resource scheduling, efficiency aware resource scheduling, energy aware resource scheduling, load balancing aware resource scheduling, QoS aware resource scheduling and utilization aware resource scheduling as shown in Fig. 2. The purpose of these classifications is to build the basis for future researchers in cloud computing environment. This classification is based on the parameter used in the evaluation of the performance in the various studies for the resource scheduling. The ovals shapes in Fig. 2 represent the subcategories of six hybrid approaches and explain and evaluate as a performance metrics in Section 5 with help of Tables 6–12 and Figs. 3–5.

The first classification focuses on cost aware resource scheduling, which includes cloud providers' revenue and profit, users' expenditure, cost of resources, and total cost. While the second category focuses on efficiency aware resource scheduling, that enhances the performance by including priority, reducing the execution time, execution cost and makespan, also, increasing the bandwidth and speed. The third type presents the energy aware resource scheduling that elaborates on minimizing the power and energy consumptions in the data centers. Fourth group represents

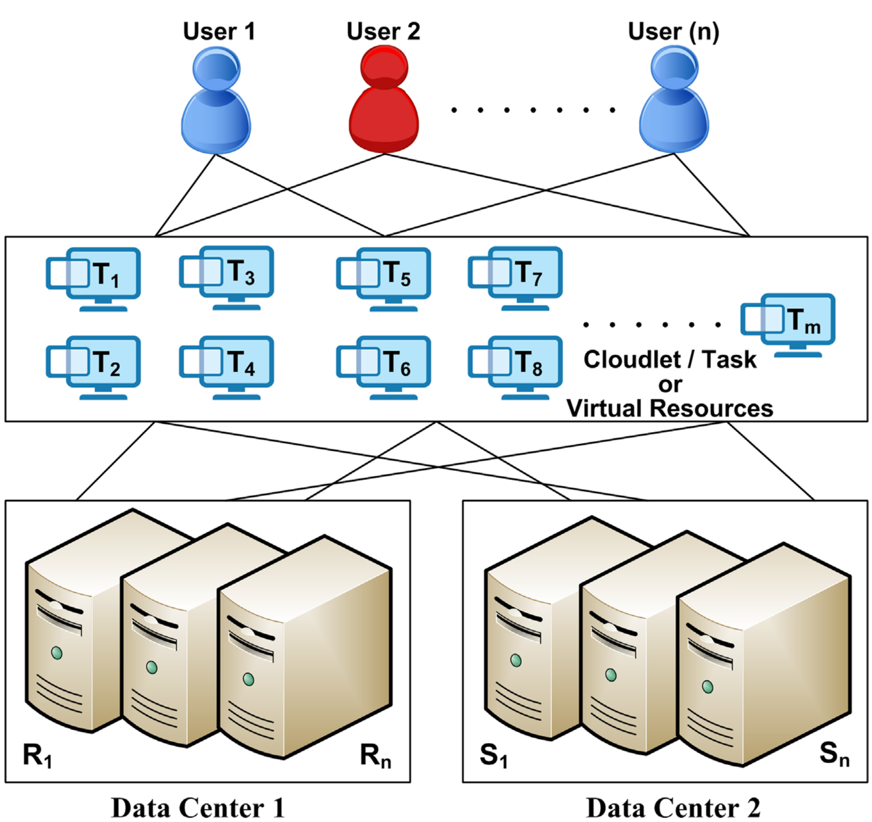

Fig. 1. Resources scheduling problem in cloud computing.

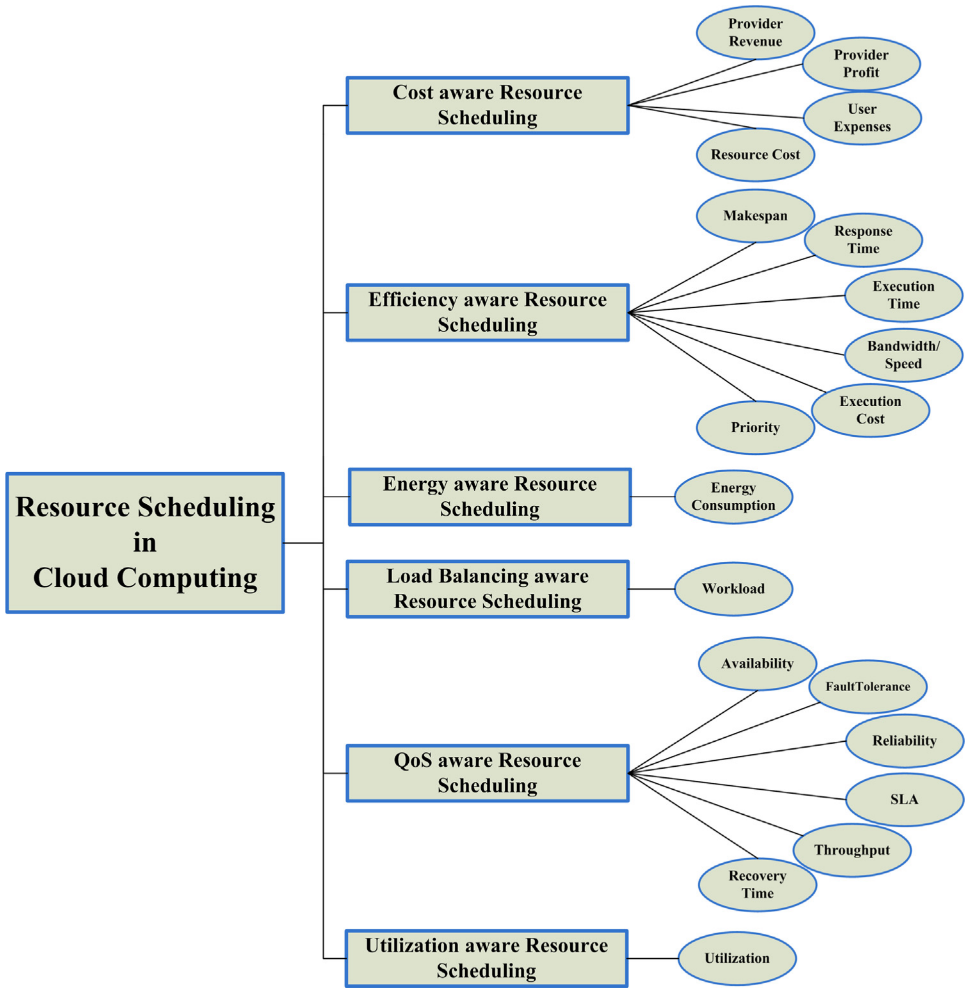

Fig. 2. Classification of resource scheduling in cloud computing.

the load balancing aware resource scheduling by efficiently managing the workload of multiple users among different data centers. Fifth class presents QoS aware resource scheduling that deals with the deals with reliability, availability, SLA, fault tolerance, throughput and recovery time. Finally, the sixth category deals with utilization aware resource scheduling. It focuses on maximizing the usage in an efficient way. The details of the classification stated above are explained in Section 4.

4. Analysis of resource scheduling

This section reviews the resource scheduling schemes and algorithms which have been used in the existing research work. However, analyzing existing techniques and understanding their focus is necessary for developing some additional applicable techniques and schemes. It can be an enhancement to the existing techniques or take the advantages from the previous studies.

4.1. Cost aware resource scheduling

Cost aware resource scheduling plays an important role in cloud computing, because as the definition of cloud declares that it delivers the services in cheapest amounts (Qian et al., 2009). A cloud provider is responsible for delivering the users' demands as a service, which results in provider revenue, profit and user expenditures. The growth of revenue and profit with maximum utilization of resources is a desire of every cloud provider. On the other side, the cloud users' desire is to get services with high performance and minimum expenses (Zhang et al., 2010). An efficient resource scheduling improves the overall system performance and helps cloud provider to deliver resources as the desire or requirement of users, with maximum utilization and economically (Li and Guo, 2010).

The main objective is to enhance the internal resource utilization and reduce the cost of outsource tasks for the cloud users. The Cuckoo Search driven Particle Swarm Optimization (CS_PSO) approach is formulated as an integer programming model to resolve this problematic issue. It execute the local search extra professionally and avoids the local optima problem of PSO. The simulation outcomes show that the proposed approach obtained high average profit as compared to the Standard PSO and Self Adaptive Learning PSO for the problem of non-trivial size (Raju et al., 2016).

Bansal et al. (2015) examine that the QoS scheduling algorithm works efficiently with makespan, latency, load balancing and cost factor. This algorithm efficient with first three parameters except the cost is introduced by Wu et al. (2013). This paper adds the additional parameter of cost factor to complete this algorithm with the optimal results.

In the paper, a new modified cost effective algorithm is proposed which minimizes the overall resource cost along with workload balancing. Simulations results demonstrate that the proposed algorithm performs better than the greedy algorithm, to reduce the overall resource cost. Furthermore, the technique deals with the energy efficiency, and also contribute to the green computing (Kapur, 2015).

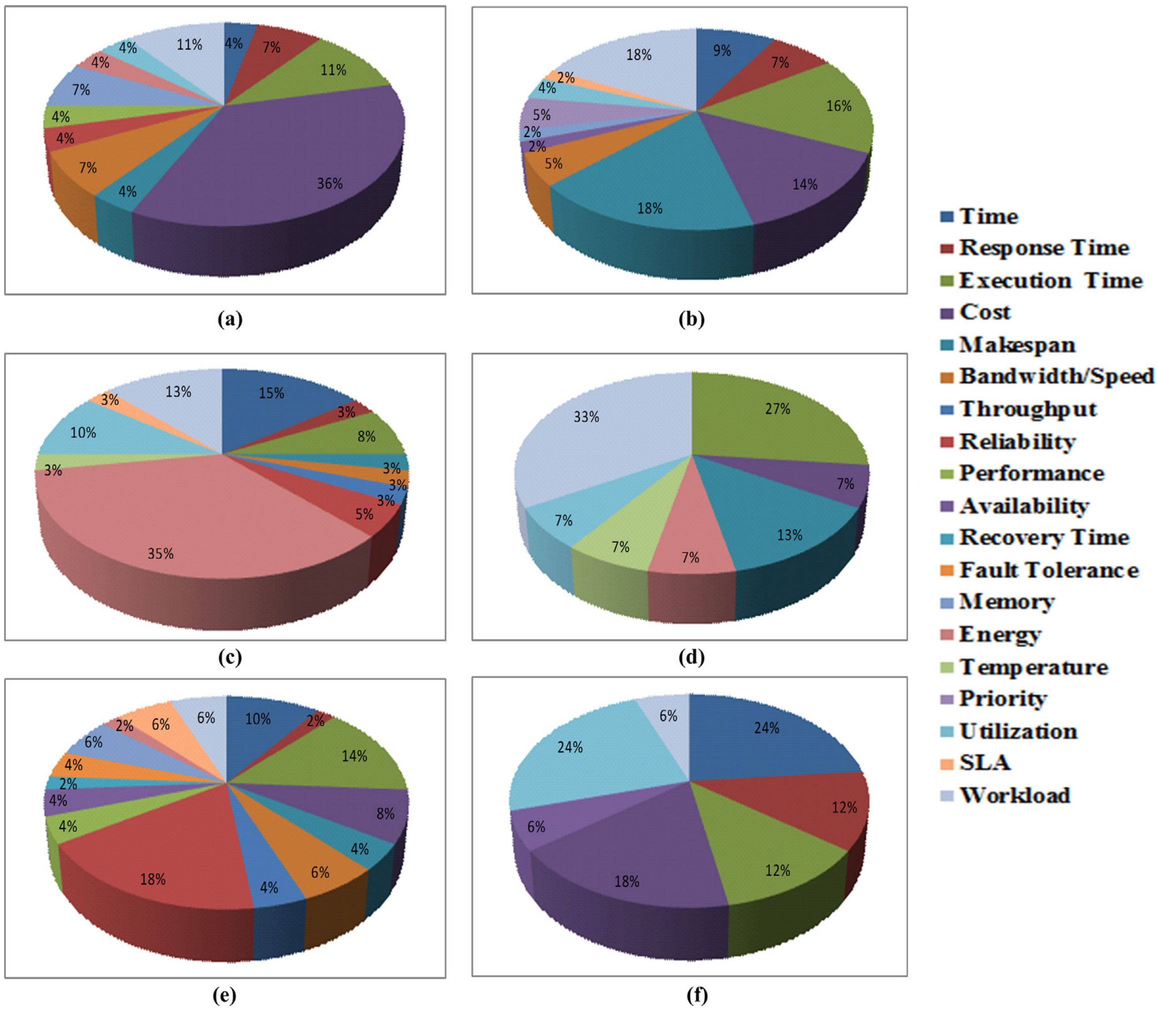

Fig. 3. Percentage of (a) Cost aware (b) Efficiency aware (c) Energy aware (d) Load balancing aware (e) QoS aware and (f) Utilization aware resource scheduling parameters.

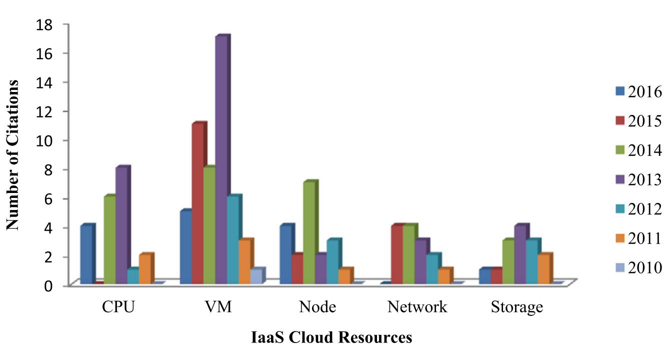

Fig. 4. Analysis of IaaS resources for scheduling from 2010 to 2016.

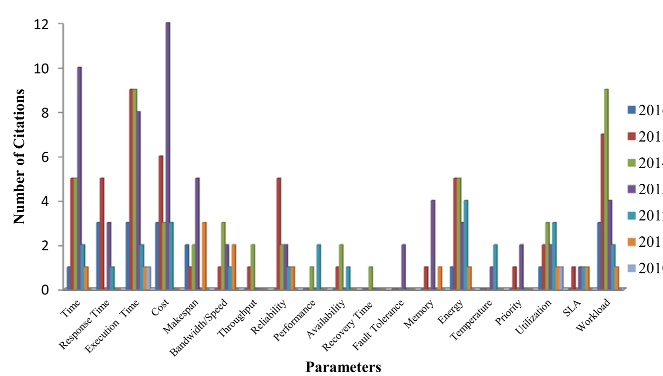

Fig. 5. Analysis of resource scheduling parameters from 2010 to 2016.

Zuo et al. (2015) propose a resource cost architecture that describes the requirements of the task for resources and shows the association between resource costs and user expenditure. Multi objective optimization scheduling architecture controls the performance and cost as the expenditure constraints of the optimization problem. Moreover, this model deals with the multi objective optimization of the user expense, time deadline, resource utility and optimal span. Finally, the paper improves the ant colony optimization algorithm to provide a solution for the optimization problem. Furthermore, the proposed algorithm always falls into the local optimum.

In the study, Netjinda et al. (2014) intend to provide an optimal scheduling framework for workflows that considers all the available purchasing options delivered by the cloud providers. The goal is to find the minimum cost of purchasing the IaaS cloud for the workload with dependency by optimizing the number of purchased instances, instance type, purchasing model and dependent task scheduling. Netjinda et al. (2014) focus on scenarios that require static task scheduling and assume that the given workflows are periodically executed over a period of time. The communication overhead is ignored. The maximum number of cloud instance, which can be purchased, is specified beforehand. To effectively discover optimal solutions, an algorithm requires two mechanisms, exploration, and exploitation. In this work, the technique based on a meta-heuristic PSO algorithm is considered because of its exploitation and exploration capabilities. After the adjustment of the few parameters, PSO can be easily implemented. Compared with other optimization algorithms in the meta-heuristic family and the quality of the results are sufficiently good.

Liu et al. (2013) design a structure for job resource mapping in order to reduce the entire implementation expenditure of the required service and the revenue production of cloud providers in cloud computing environment along with the setup of heuristic scheduling algorithm, which utilizes PSO algorithm in order to work out with job resource mappings that constitute on the planned model. In this regard, PSO algorithm starts with the informal start of particles positions and the speed. However, each particle vector is an expectant solution of the core problematic issue. However, the particles get a task in accordance with the volume of particle and genetic algorithms end up efficiently in task scheduling problem in cloud computing.

Cloud based online Hybrid scheduling policy (CoH) is presented by Shen et al. (2013). It finds the best balance of billing and PMs formations that keeps the cost of resources at lowest. The resource allocation and scheduling problem causes an Integer Programming Problems (IPP). The time to take a decision is to be kept low as CoH requires online execution. The policy takes the results of IPP, and adopts the best for scheduling decision. The significant findings of the work includes the novel online scheduling policy based on CoH, that take the scheduling decision based on IPP using heuristics approaches. The policy is enhanced that also uses reserved instances to minimize the cost.

Tiwari et al. (2013) in their research work present an approach, which is based on the inner performance of the virtual machines and data centers, with providing the cloud service providers rating. In the current situation, the number of cloud providers are growing day in and day out. Using this information scheduling of cloud providers require a method to search an optimal providers data to assist the request of scheduling. Services can be allocated to the optimal service providers by SRS. In the case of context aware application organized in geographically distributed data centers which form a cloud, the problem of scheduling and request dynamic allocation is discussed.

In the research carried by Yi et al. (2013) propose the model of MILP and recommend a Best Fit algorithm based on various different jobs scheduling approaches, to solve the resource scheduling problem in the grid and cloud computing. For the input data, it takes multiple jobs' structures that contains parallel or sequential sub cloudlets. The MILP model solves the problems with the optimal solution. However, it consumes more time when faces heavy load on input traffic. The optimal solution is obtained by the proposed heuristic algorithms in a short period of time.

In the paper, the proposed model comprises of one distributed and dynamically scaling cloud computing cluster of virtual machines. The workload contains the parallel cloudlets that are either large or small based quantified cloudlet size parameters before the simulation. Compared with AFCFS and LJFS, the paper evaluates cost efficiency and their performance in cloud computing. Moreover, it adds and removes the VMs from the system based on the load of the system at the specific time (Moschakis and Karatza, 2012). All cost aware resource scheduling techniques and comparison are shown in Table 1. The techniques are compared according to problem addressed, scheduling algorithms, policies and strategies with the primary advantages and disadvantages. Furthermore, the parameters used for cost aware resource scheduling is shown in Table 7.

4.2. Efficiency aware resource scheduling

Efficiency aware resource scheduling expresses the amount of resources consumed for processing, depending upon the targeted resources to enhance the efficiency. The effective resource scheduling helps to improve and enhance the response time, execution time, makespan, bandwidth/speed and priority (Younge et al., 2010). It is a collection of service performance that indicates the degree of satisfaction of cloud user for the IaaS resources or services, which is a desire of cloud users to acquire a service that should be more proficient economically and efficiently (Puthal et al., 2015).

A discrete version of Symbiotic Organism Search is proposed and applied to solve task scheduling problem in cloud computing environment. It is inspired by the mutual interaction between organisms in the ecosystem. Task scheduling problem is a form of discrete optimization which is known to be N-Complete. The performance of the proposed method is evaluated using four different distribution statistically based generated data sets. The proposed algorithm is outperformed than PSO algorithm, which is one of the states of the art metaheuristic methods for solving optimization problems (Abdullahi et al., 2016).

Bee Colony algorithm is enhanced and it is used to efficiently schedule the balance of load between nodes in dynamic cloud environment. The improved technique tries to attain the minimum makespan and reduce the no of migration, with providing the efficiency to the cloud users. Here, its power is also used to eliminates the tasks from over loaded VMs and submitted to the suitable under loaded VMs. (Babu and Samuel, 2016)

Bansal et al. (2016) demonstrate the implementation vision of scheduling techniques for task scheduling with simulation results within CloudSim by considering the influence of execution cost and load balancing. In the comparison of optimized and traditional algorithms, the results show that optimized algorithms are always performed better than the traditional ones.

In order to resolve the issue of task scheduling, Ma et al. (2016) suggest the innovative dynamic algorithm based on Improved Genetic Algorithm for cloud computing. The proposed technique provides the full attention to the dynamic appearance of the cloud computing. Simulation results show that IGA effectively reduce the response and execution time for task scheduling, so it can enhance the performance with throughput in cloud computing.

Considering the requirements and preferences of users, allocating jobs to the most appropriate resources is one of the most vital issues for job scheduling. To address this issue, FUGE which is a hybrid approach base on genetic algorithm and fuzzy theory is presented by Shojafar et al. (2015). The aim of this approach is optimal load balancing in consideration with cost and execution time. Modification is completed by the Standard Genetic Algorithm (SGA) and devising a fuzzy based steady state GA in order to improve SGA performance in terms of makespan. By considering VMs, processing speed, bandwidth, memory, and the job lengths jobs are assigned to resources by the FUGE algorithm.

Min-Min algorithm is used to decrease the makespan of tasks by seeing the task length. Keeping this in attention, cloud providers should accomplish cloud user demands. Thus, this scheduling algorithm that considers both cloud user requirement and resources obtainability. The improved scheduling algorithm is familiarized by Thomas et al. (2015). High priority tasks are not given any special status when they work out. The proposed method considers all of these aspects. The outcomes illustrate a substantial improvement in the utilization of resources.

To resolve workflow scheduling is identified as a NP-hard problem, for these types of problems exhaustive procedures cannot be implemented. Consequently, a non-exhaustive optimization technique is a well choice for solving such problems. In the research, Raghavan et al. (2015) use a meta-heuristic algorithm known as Bat algorithm and some features Binary Bat algorithm for workflow scheduling applications in cloud computing. Bat algorithm is superior executed than other meta-heuristic algorithms. The Bat algorithm is

Table 1
Cost aware resource scheduling.

Reference	Algorithm, policy or strategy	Problems addressed	Improvement /-achievement	Weakness/limitations
Raju et al. (2016)	• Cockoo search driven particle swarm optimization	Deadline constrained task scheduling	Improve the performance	Focus only on the cloud provider profit
Bansal et al. (2015)	• QoS driven scheduling algorithm	Allocation cost for task scheduling	Reducing the cost	Compare with only basic technique
Kapur (2015)	• Cost Effective Resource Scheduling (CERS) algorithm.	Workload balancing	Improve the performance	Cost is fully dependant on the time and compare with only one traditional technique.
Zuo et al. (2015)	• Performance and Budget Cost based Ant Colony Optimization Algorithm (PBACO)	Multi objective Task Scheduling	Improve the performance	Major comparison is done with traditional algorithms
Netjinda et al. (2014)	• **Algorithm 1** Partial Swarm Optimization • **Algorithm 2** Task Scheduling • **Algorithm 3** Variable Neighbourhood Search • **Algorithm 4** Workflow Scheduling Cost Optimization	To find the minimum cost of purchasing the IaaS cloud for the workload	Show improvement	Results are close to ideal cost
Liu et al. (2013)	• Directed Acyclic Graph (DAG) • Particle Swarm Optimization	Optimal resource scheduling	Reduce the energy and cost with increasing the high profit.	Results of random algorithm are not good due to algorithm does not optimize the QoS and rejects a number of realistic requests.
Shen et al. (2013)	• Cloud-based, online Hybrid scheduling policy (CoH)	Integer programming problems	Reduce cost and better performance	
Tiwari et al. (2013)	Rough set theory • The ROSP algorithm	Optimal resource scheduling	Improve performance	Do not compare with existing algorithms
Yi et al. (2013)	• **Algorithm 1** Best-Fit Algorithm • **Algorithm 2** Bandwidth resource allocation	Job scheduling for reducing the cost of user	Reduce cost and better performance	MILP is time consuming when the traffic is heavy.
Moschakis and Karatza (2012)	• **Algorithm 1** Adaptive first come first serve • **Algorithm 2** Largest job first served	Virtual machine management	Improve the performance and minimize the cost	Compare with only basic technique

associated with Best Resource Selection (BRS) algorithm and the conclusions show that the bat algorithm is fifty percent better than BRS algorithm considered there.

Diminishing the makespan of tasks scheduling in IaaS cloud is also a NP-hard problem using the league championship algorithm (LCA), which is the foremost intention of the study. The LCA is a sports-inspired population based algorithmic structure for universal optimization over an uninterrupted examines space. This implies great prospects of acting out well in this field as it had succeeded in solving other NP complete problems in other areas of research. From the simulation results, it shows that, the LCA helps cloud users to keep cost for the time used than the LJF, BEF or the FCFS scheduling algorithms, as it takes lesser time for the makespan to process the tasks (Abdulhamid et al., 2014b).

Javanmardi et al. (2014) suggest a hybrid job scheduling method that is used reducing the total execution time, execution cost and load balancing. The standard GA is modified and to decrease the iteration of generating population with the support of fuzzy theory. The jobs are allocated to the resources with considering the VM capacity and jobs size is a primary objective. The experiments results show the efficiency of the suggested method with reducing the execution time, execution cost and load balancing.

Regardless of this, it is considered as a subject of great dispute for effective scheduling algorithm designing and execution. Lin et al., (2014) designed a non-linear programming method in order to resolve the restricted multiport model and have designed a novel task scheduling algorithm, the bandwidth aware task scheduling (BATS) algorithm for the allocation of appropriate quantity of tasks to VM while considering the space, CPU, power, and network bandwidth. However, commonly used cloud computing task scheduling algorithms prefer the processing and memory resources in task rescheduling. Moreover, on the other hand constraints of network bandwidth in task scheduling is a topic under consideration. In this regard, none of them prove the usage of all three resources. While, keeping in view the shortcomings of the ACO algorithm and the inequality of load of fundamental machines in task scheduling process.

In a cloud computing environment the characteristics of task scheduling are discussed by Liu et al. (2014a), which discusses an algorithm for task scheduling that depends upon genetic-ant colony algorithm. The improvement is having a strong optimistic feedback of ACO and taking into account the convergence rate of the algorithm. However, the convergence rate is strongly affected by the choice of the initial pheromone. The global search capability of the GA is used by it to solve the optimal solution quickly and then converts it into the initial pheromone of ACO. Under same conditions, the results shown by simulation experiments suggest that this algorithm exceed the weights of GA and ACO, even it has an efficiency advantage in large scale environment.

Effective resource scheduling can minimize the task finishing time, increase the utilization of computing resources in cloud computing. Therefore, improvement of the performance and the quality of services are the most core parameters considered in the environment of cloud computing. In this regard, it depends on the QoS aware scheduling algorithm and parameters created for the resource and task scheduling in the cloud. It can be capable of differentiating among user's service quality necessities, to deliver users within the agreement with the needs of resources (Pan et al., 2014).

The improved ant colony algorithm depend on partial swarm optimization is known as ACA-PSO is proposed by Yang (2014). Initially, the ants are in the lineup with ant colony algorithm for the completion of the traverse, and re-arrangement of the solutions, while keeping in view the confined and universal solutions. While ACA-PSO controls the shortcomings of and algorithm as it easily gets into confined solutions in cloud computing resource scheduling.

Ergu et al. (2013) propose a model for task oriented resource allocation in cloud computing. The cloudlets are associated considering bandwidth, costs, network complete time, and reliability. The weights of cloudlets are calculated using the Analytic Hierarchy Process (AHP) and the computing resources are allocated. In the proposed model the two examples state that the weights of the cloudlets are changed with multiple values of cloud resources which are less than the specified threshold. The outcomes show that the inconsistent elements and enhances the consistency ratio, when the cloudlets weights are allocated dynamically.

In the research, Le et al. (2013) suggest a deadline based resource management scheme that dynamically allocates the resources. The proposed scheme can be separated in two portions one is resource provisioning and task scheduling. Resource provisioning provides a way to acquire or release VMs based on user demands and introducing additional VM startups to reduce costs. Furthermore, the other portion deals with the scheduling, to control the order of execution. The jobs are executed in VM sequentially based on to the predefined policy. The final objective of the research is to make sure that the submitted tasks should to be completed before the deadline.

For reducing the transferring time, execution time and execution cost, Ramezani et al. (2013) improve the complete multi objective model for task scheduling optimization. Conflict the consideration between the tasks, the authority of PSO algorithm in accuracy and speed. In order to deliver an optimal solution for the presented model, a multi objective algorithm that is based on Multi Objective PSO (MOPSO) method is proposed. Jswarm package to multi objective Jswarm (MOJ) package has been used to calculate and implement the proposed model, with extending Cloudsim toolkit put on MOJ as its task scheduling algorithm. Optimal task organization among VMs is defined by MOJ in Cloudsim according to MOPSO algorithm.

Herein, Sindhu and Mukherjee (2013) propose an innovative scheduling algorithm depending upon GA that is applications and resources centric. Furthermore, a multi objective GA optimizes the makespan and processor utilization in an efficient way. The change in the numbers of VMs in accordance with the number of hosts cannot be competently scheduled by a single scheduling technique. Therefore, they try to map in the VMs.

The Taguchi method and a Differential Evolution Algorithm (DEA) are combined in the Improved Differential Evolution Algorithm (IDEA). DEA uses few control parameters and has a dominant universal exploration capability on macro-space. To exploit the better individuals on micro-space to be potential offspring the systematic reasoning ability of the Taguchi method is used. Therefore, the presented IDEA is a balanced and more enhanced on utilization and consideration. Delivery cost, processing is contained by the cost model while time model contains waiting and receiving time. The non-dominated categorization procedure, multi-objective optimization method, not with the normalized single objective method, is realistic to search the Pareto front of makespan and cost (Tsai et al., 2013).

Task schedule algorithms openly focus on the speed and quality of the scheduling. Min-Min algorithm has the characteristic of simple and shortest completion time for the task. The least possible completion time for every single task is calculated according to completely physical resources with the selection and assigning. The recently mapped task is unconcerned to express for the least tasks for distribution, and the procedure repeats until all scheduling task set is unfilled. In the enhanced Min-Min algorithm provides QoS to the resources, includes scheduling completion time, reliability and cost. The outcomes show that the suggested algorithm is effective for the scheduling in the cloud computing environment (Wang and Yu, 2013).

In the hybrid cloud environment, Wang et al. (2013) suggest the Adaptive Scheduling with QoS satisfaction algorithm, known as AsQ. The proposed algorithm utilizes the execution time and multiple optimization techniques to find out an optimal resource allocation. Such that the utilization rate of the cloud, the renting expense, and the makespan are enhanced.Based on the algorithm, a fast scheduling technique is used to accomplish a total optimization of the cost and the deadline. The modified max min algorithm provides better results without consuming additional time. Whereas, the AsQ do not consume much time to take a decision. When a task with a deadline cannot be executed on time, the AsQ sends some more tasks to the public cloud. The remaining tasks in a private cloud, if any is implemented before the deadline and to satisfy the budget constraint. With the proposed techniques, the tasks with user constraints can be accepted or rejected.

Wu et al. (2013) suggest a task scheduling algorithm depend upon QoS driven for cloud computing. The Task Scheduling QoS (TS_QoS) algorithm calculates the priority of task according to the characteristics of task and then arranges the tasks with aspect to the priority. It also estimates the completion time of each task on various services, and schedules them quickly according to the sorted task queue. In this way, the algorithm accomplishes improved performance and balancing the load by QoS driven from both completion time and priority of task.

The paper presents a scheduling technique for multiple VMs migrations. The objectives of the paper are the analysis and formulation of VMs migration. The proposed algorithm consists of two portions. The first is used for finding the minimum cost VM migration path from all distributed VMs to PMs. The second part of the algorithm is creating the optimization migration sequence and accelerating multiple VM migration process based on parallel processing techniques. The simulations are used to validate the efficiency of the algorithm (Zhang et al., 2013c).

To reduce the problems of cloud computing data centers in resource management. Furthermore, it provides the assurance of better quality QoS that the cloud computing can supply. The ACO is applied for efficient resource scheduling according to the real QoS parameters requirement of the environment for the cloud computing (Zhu and Liang, 2013). First it gets the cloudlet, then sorts and classifies the cloudlet based on its priority. Classification depends upon the different QoS requirements. Further, it allocates and schedules the resource with ACO according to QoS demands. Finally, the requirements of QoS are analyzed and the shortest path is bind with the cloudlet and resource to run the task. The simulation results show a better performance by achieving optimum completion time.

Innovative scheduling algorithm, which competently schedules the computational tasks in a cloud environment and produced tree based data structure known as a Virtual Machine Tree. They reformed DFS, which is used the appropriate VM for implementation. The Simulation is executed for fluctuating a number of VMs and workload traces (Achar et al., 2012).

Zhong-wen and Kai (2012) proposes a scheduling technique based on the parameters of the cost, time and trust for cloud computing resources. The algorithm is appropriate for resource scheduling and provides high efficiency in large scale cloud computing. The experimental results show the competence of the design and users are capable to achieve the optimal solution by the making subsets of tree pruning and decision.

Sindhu and Mukherjee (2011) determine the use of two scheduling algorithms Shortest Cloudlet Fastest Processing Element (SCFP) and Longest Cloudlet Fastest Processing Element (LCFP) for the task scheduling. Algorithms reflect the processing constraint of a task and the computational capability of a resource during the scheduling assessments process. The whole makespan to accomplish the tasks are used as the metric to calculate the outcomes of the suggested algorithms. Table 2 compares the techniques according to the parameters used to evaluate the efficiency of resource scheduling, and further details of the parameters used in efficiency aware resource scheduling are shown in Table 8.

4.3. Energy aware resource scheduling

Energy aware resource scheduling techniques are required to overcome the problems emerging due to high energy consumption in the data centers. Under the cloud computing, diminishing the energy consumption and saving the expenses due to energy are substantial for the data centers and cloud providers (Buyya et al., 2010). Data is increasing so rapidly that progressively larger servers and disks are needed to process them quickly within the mandatory time period. The lost or wastage of idle power is a major cause of energy ineffectiveness (Beloglazov and Buyya, 2010). Green computing is proposed to achieve the efficient processing and utilization of resources, by minimizing the energy consumption (Pandi and Somasundaram, 2016; Singh, 2015).

An efficient prediction model based on fractal mathematics is developed to assist the algorithm that decides to turn on/off hosts. Through this way, it helps to avoid the performance and energy loss, which is triggered by instantaneous peak loads on account of scheduling. An improved ant colony algorithm is proposed that is applied for virtual machine execution for achieving optimize energy consumption and make it cost effectively to meet resource intensive application demands in cloud computing environment (Duan et al., 2016).

Ding et al. (2015) concentrate on the dynamic VMs scheduling to attain energy efficiency and fulfill deadline constraints with various PMs in cloud. The sufficient numbers of PMs are present to process specific number of VMs with optimal frequency and each PM operates on at least the optimal frequency. The required best power ratio performance of PMs obtain from the idea of using diversified physical machines, Where VMs are allocated the power ratio before the PMs to increase the performance of power ratio. In the cloud computing the scheduling is divided into some small time event to reconfigured and consolidate the computation resource after each time event to reduce the energy consumption. The deadline time event is necessary to operate for managing the VMs resources from the PMs.

The information about the average time response of integer programming problem in data center for task assignments, which are limited in dynamic servers is provided by Dong et al. (2015). The dynamic servers in data center consume more energy and less efficient in scheduling to enhance these parameter with a greedy task scheduling. Whereas the maximum energy saving schemes for data center servers are used in Most Efficient Server First (MESF) task scheduling scheme. With this scheme the average task response time decreases, with the cost for the server up-gradation and reduce energy. Also it protects the energy at the cost of extended task response times, although within the maximum limitation.

Hosseinimotlagh et al. (2015) suggest a Smart Energy-aware Task Scheduling (SEAT) VMs scheduling algorithm, which intentions to extend the optimal level of utilization by proposing extra calculating energy to VMs of cloud providers. The proposed SEATS algorithm creates servers to accomplish their VMs more rapidly to achieve their optimal utilization levels without requiring to transfer VMs, which ultimately information to decreasing the energy consumption. In this way, the indolent servers are turned off to save energy. Experiments show that the suggested policy not only reduces energy consumptions, but on the other hand also reduces the turnaround times of real-time tasks in the cloud.

Furthermore, Jena (2015) propose multi objective optimization

Table 2
Efficiency aware resource scheduling.

Reference	Algorithm, policy or strategy	Problems addressed	Improvement /-achievement	Weakness/limitations
Abdullahi et al. (2016)	• Symbiotic organism search	Task scheduling	Improved the performance by reducing Makespan, response time and degree of imbalance	Focus on load balancing
Babu and Samuel (2016)	• Enhance bee colony	Task scheduling and load balancing	Improved the makespan and migration time	Compare the results only with basic Bee Colony algorithm
Bansal et al. (2016)	• First come first serve • Virtual machine tree • Particle swarm optimization • QoS-driven • Activity based costing	Task scheduling	Point out the issue of performance with traditional algorithms	Do not make the comparison of meta-heuristic algorithm and consider only the execution cost and load balancing
Ma et al. (2016)	• Improved Genetic algorithm	Task scheduling	Improve the execution and response time	Do not show the evaluation of throughput in the results
Raghavan et al. (2015)	• Bat algorithm	Workflow scheduling	Minimize the execution cost	Do not consider the overall efficiency
Shojafar et al. (2015)	• FUGE algorithm	Job scheduling	Improve the execution cost and time.	Do not care about the energy consumption
Thomas et al. (2015)	• Min–Min algorithm • The credit system based on task length • The credit system based on task priority	Task length and task priority aware scheduling	Makespan of the task is lesser and enhance the utilization of resources	Do not compare the performance with existing algorithms
Abdulhamid et al. (2014b)	• League Championship Algorithm (LCA)	Minimizing the makespan	Improve the execution time	Compare with basic three algorithms
Javanmardi et al. (2014)	• Genetic algorithm modified with fuzzy theory	Job scheduling	Improve the performance regarding execution cost	Focus on only execution cost
Lin et al. (2014)	• Nonlinear programming model • Bandwidth aware Task Scheduling (BATS) algorithm	NP complete problem	Enhance the efficiency of task scheduling and resource utilization. Increase revenue and save cost of cloud provider	Wastage of resources
Liu et al. (2014a)	• Genetic-Ant Colony scheduling algorithm (GA-ACO)	Slow convergence problem	Improve the searching efficiency of algorithm (Searching and execution time)	The number of tasks is less significant and the resources are more enough.
Pan et al. (2014)	• Management system of task scheduling and resource allocation of cloud computing	QoS constraints resources task allocation	Decrease the task execution time and increase the efficiency of computing resources custom	Not implemented and simulated
Yang (2014)	• Ant Colony Algorithm based on PSO (ACA-PSO)	To enhance the efficiency of resource scheduling	Improve the convergence speed, escape deceiving into local optimum solution	Focus on local optimum
Ergu et al. (2013)	• Analytic hierarchy process	Task oriented resource allocation	Enhance the consistency ratio	Illustrate only examples mathematically
Le et al. (2013)	• First-Come-First-Service (FCFS), • Shortest Job First (SJF) • Nearest Deadline First (NDF).	Providing the cloud resources dynamically	Flexible resource allocation and scheduling with better performance	Focus on only basic algorithms
Ramezani et al. (2013)	• Multi-objective algorithm based on multi-objective PSO (MOPSO) • Multi-objective Jswarm (MOJ) package	To reduce the transferring time, execution cost, and time for the task.	Better Performance	Do not compare the performance with existing algorithms
Sindhu and Mukherjee (2013)	• Multi objective genetic algorithm based scheduler	VM scheduling	Fast processing, minimize makespan and processing utilization	
Tsai et al. (2013)	• Improved Differential Evolution Algorithm (IDEA)	To decrease task execution time and cost	Better performance	Focus only on time and cost
Wang and Yu (2013)	• Min–Min algorithm	To maximize the efficiency	Improve the performance	Min-Min algorithm has two sides. It is a possible Load imbalance and resources utilization rate is less. And traditional Min-Min algorithm does not come across various tasks of diverse service quality necessities.
Wang et al. (2013)	• Adaptive Scheduling with QoS Satisfaction (AsQ)	A variation of the multi-dimension multichoice knapsack problem	Optimization of cost and achievement of deadline constraints	Do not estimate the performance of the AsQ while AsQ spends about 77% more cost than the COSHIC
Wu et al. (2013)	• Task Scheduling Quality of Service (TS_QoS) Algorithm	QoS aware Task Scheduling	Minimize the completion time (Makespan, Load Balancing)	Algorithm is based on the priority
Zhang et al. (2013c)	• Host Selection Algorithm for Virtual Machine	To decrease the migration time and increase speed of the VM in the migration process	Reduce the VM migration time and improve the performance	Highly complex, and its performance is not up to the mark in case of large scale multi-VM migration.
Zhu and Liang (2013)	• The resource scheduling strategy based on Ant Colony Algorithm	Deals with load balancing, QoS and efficiency parameters	Minimize execution time of cloudlet	Do not compare the performance with existing algorithms
Achar et al. (2012)	• Prioritizing • Virtual Machine Tree Grouping	To maximize the efficiency	Improve the performance	Compare with the simple algorithm like FCFS and Priority Based Algorithm.

Table 2 (*continued*)

Reference	Algorithm, policy or strategy	Problems addressed	Improvement /-achievement	Weakness/limitations
	• Virtual Machine Selection • **Algorithm 1**: Modified DFS			
Zhong-wen and Kai (2012)	• The Time Cost Trust Model (based on the subset tree algorithm)	To maximize the user's individual utility and satisfactions.	Improve Performance and decision making.	
Sindhu and Mukherjee (2011)	• Longest Cloudlet Fastest Processing Element (LCFP) • Shortest Cloudlet Fastest Processing Element (SCFP)	Schedule the tasks effectively, decrease the turnaround time and enhance resource utilization	With increasing the number of tasks, LCFP exhibits better performance as compared to SCFP and FCFS.	Do not focus and use on heuristic algorithms

algorithm based on PSO, which can solve the task scheduling problem in cloud computing, where of the number of data center and user job changing. In dynamically changing cloud computing, resources are allocated optimally. Therefore, multi-objective nested PSO based algorithm is recommended. The proposed algorithm effectively utilize the resources, reduces energy and makespan. The experimental results illustrated that the proposed methods (MOPSO) perform better than the BRS and RSA.

Jiankang et al. (2014) recommend VM scheduling techniques multiple resource constraints, such as the physical server size and network link capacity, to reduce the numbers of active PMs and network elements considering energy consumption. The design of a two level heuristic algorithm is given for VM scheduling in placement and migration. Compared with previous the works proposed algorithm achieves better performance in the simulation. Resource overloading is still a problem and live migration does not support the alteration of virtual machines.

VMs are installed on PMs and countless power is expended to service the servers in cloud data center. More physical servers require extra budget and energy consumption. Consequently, the Virtual Machine Placement (VMP) problematic issue is substantial in cloud computing. Liu et al. (2014b) propose an approach based on ACO to resolve the VMP problematic issue, known as ACO-VMP, to efficiently use the physical resources and minimize the quantity of running physical servers.

Due to the increment of size and number of data centers, energy consumptions are raised up in data center since last few years. Shuja et al. (2014) anticipate a data center wide energy efficient resource scheduling framework (DCEERS) for resources scheduling with respect to the need of the throughput. The data center is structured as a multi service stream network with several source, sink nodes and different limits on movement from one stage to other. The optimization of lowest cost multi service stream provides the least number of resources required for the request of existing workload in the data center. For complex lowest cost multi service stream problem, one of the fast heuristic algorithm, Benders decomposition is worked to discover the optimal solution for the accessible data center. Simulations illustrate that proposed framework use up for extra energy as equated to other heuristics algorithms.

The comprehensive review of partial swarm optimization algorithm and application in cloud computing using the CloudSim system architecture is presented in Xiong and Wu (2014). Resources scheduling problematic issue are considered comprehensively and achieve the performance, scalability and reliability. As a final point, the requests of the users associated the area of cloud computing is found. Three parameters are examined to determine the fitness values dynamically with using the PSO as resource scheduling strategy on cloud computing. The experiment results show that the suggested resource scheduling approach decreases the usage of energy and increases the performance of data center, this strategy has a very virtuous influence on the cloud computing.

Ghribi et al. (2013) explore the NP-hard problematic issue of VM placement in cloud data centers. The main contribution is using algorithms for placement and consolidation that reduce both the migration costs and energy consumption. They anticipated a linear integer program consistent to an accurate distribution of an efficient specific VM migration algorithm to minimize the energy consumption through the alliance. Sufficient energy savings is achieved based on the workloads of the systems with fewer loads and the results are significantly improved.

Beloglazov et al. (2012) explore energy aware resource allocation and scheduling algorithms that allocates the cloud resources to cloud users in a way, which enhances the energy efficiency of a data center without compromising the SLA. Autonomic and energy aware mechanisms for self-management are used to attain the energy efficiency. Proposed algorithms for energy efficient mapping of VMs is appropriate for cloud resources. The simulation outcomes show that proposed approach decreases the energy consumption in cloud data centers.

Chen et al. (2012) suggest a cloud computing resource scheduling policy depend on the genetic algorithm with numerous fitness, for the purpose of improving the utilization of resources and saving the cost of energy in cloud computing. Chen et al. (2012) use a pre-migration approach based on three load degrees such as CPU utilization, throughput of network, rate of disk I/O, which are measured in the algorithm corresponding. In order to acquire a roughly optimal solution, and assume the hybrid genetic algorithm mutual with knapsack problem through various fitness and results show the efficiency of the algorithm. The algorithm succeeds the aim of enhancing the utilization of resources and saving the cost of energy by runtime resource scheduling.

In the paper, Luo et al. (2012) describe the energy efficiency in cloud computing. The energy consumption model is proposed, and sorts the computing resource into four different categories that includes CPU, memory, storage, and networks. Furthermore, design the various regulation and strategies for multiple components. Moreover, the paper proposes a dynamic resource scheduling algorithm based on energy optimization of cloud resources with evaluation methodology. The simulation results show a better performance of algorithm in terms of energy consumption.

The systematic performance model of the data centers is presented by Van Do and Rotter (2012). It takes into account the quality of service (QoS) assured to users and the energy functional usage in the data centers. Various scheduling algorithms ensure the regular heat emission and energy consumption of servers as well as the blocking probabilities of on demand requests. With the exception of energy consumption is potential in the functional range with the distribution of VMs to PMs on the basis of priority depends on the mathematical results of association of different allocation strategies.

Mezmaz et al. (2011) present a parallel bi-objective hybrid genetic algorithm to solve task scheduling problem in cloud computing. The main objectives are enhancing the makespan and energy consumption. Proposed method approach is appraised with the fast fourier transformation task graph which is a real world application. Results show that proposed algorithm increases on average the results, particularly in energy consumption. Indeed, the energy usage is reduced by 47.5% and the makespan by 12%. Table 3 comprehensively compares different energy aware resource scheduling techniques, while the parameters used for these techniques are shown in Table 9.

4.4. Load balancing aware resource scheduling

Load balancing is a feasible process that improves VMs and data center overloaded with computing cloudlets, tasks or jobs, through sharing loads across data center infrastructures to achieve a proficient performance of the systems (Nuaimi et al., 2012). Efficient allocation and scheduling must ensure that resources are easily available on demand and proficiently utilize under condition of high/low load by saving energy and cost (Katyal and Mishra, 2014).

An ideal scheduling algorithm is vital to resolve the load balance difficulties which can not only stable the load, but also can meet the user's requirements. An optimal load balance algorithm is suggested by Pan et al. (2015), which can improve assembly of the systems and schedule the tasks to virtual machines further professionally. Execution time of all tasks in the similar system is less than others algorithms.

Cho et al. (2014) propose a hybrid meta-heuristic VM scheduling technique with a load balancing algorithm. The proposed technique is used to customize VM requirements and considers a number of resources to achieve load balancing. Pre-reject module is used to decrease the scheduling time. Furthermore, to enhance the performance of resource scheduling PSO operator is added to the ACO technique. The proposed algorithm is implemented with a limited amount of information and uses the workload of historical requests to predict the workload of new input requests. Experimental outcomes express that the performance ACOPS is superior to the basic ACO and also it is better in balancing the load of system as compared to existing techniques.

Wang et al. (2014) propose the least job time consuming and load balancing genetic algorithm (JLGA) to optimize the task allocation sequence in a dynamic cloud computing. It reduces the makespan of tasks with balancing the load of the entire systems. The algorithm assumes greedy algorithm for initializing the population, brings in variance to describe the load intensive between nodes and weights of multi fitness function. The simulations results show that the proposed algorithm is suitable for balancing the whole system's load efficiently.

The paper suggests the Ant Colony Optimization Load Balancing algorithm (ACO-LB). However, simulation outcomes express that the enhanced ant colony optimization scheduling algorithm, not only minimize the task completion time. It also helps in the arrangement of tasks given to virtual machines while keeping in view computing potential. Due to the virtual machines stack be in a reasonable situation, and it keeps away from the waste of sources and other issues. Moreover, ACO-LB algorithm can effectively arrange the suitable resources for job completion and helps in resources allocation and generation (Xue et al., 2014).

An improved version of task scheduling in cloud computing is presented by Zhao et al. (2014) that embrace the intelligence firefly algorithm. With the intelligence firefly algorithm in the cloud computing the study shows the greatest solution for task scheduling. However, through simulation experiments the enhanced firefly algorithm searches the universal optimal solution. The development of this algorithm sets a theoretical base to additional develop the system's resources allocation consequence below cloud computing. The extending network and balancing problem of network load are solved by this method. Also, the global convergence of the algorithm is determined by it.

Mhedheb et al. (2013) suggest and implement a load and thermal aware scheduling mechanism, which is competent for avoiding the occurrence of over loading and over heating of the PMs.This proposed algorithm first manages the initial scheduling and after that looks at the change of workload and temperature on the host. In the situation of overloading or overheating, the VM is shifted to another PM, hence to avoid hot spots high temperature. The experimental results show the benefit of the proposed mechanism regarding energy consumption.

In the paper Li et al. (2011) proposes the Load Balancing Ant Colony Optimization (LBACO) algorithm, to achieve optimal load balancing in tasks scheduling. The simulations show the LBACO algorithm in applications with the number of tasks varying from 100 to 500. The simulation results show that the LBACO optimizes the entire system's load effectively. Dynamic workload of task is handled by LBACO in all conditions, and performs better than the FCFS and ACO algorithms. Table 4 comprehensively compares the previous mentioned techniques that are used for load balancing aware resource scheduling, while the parameters used for load balancing techniques are shown in Table 10.

4.5. QoS aware resource scheduling

QoS aware resource scheduling is a key issue in a cloud computing. It implies to schedule efficiently and demanded task of users to different resources according to the QoS, which focuses on

Table 3
Energy aware resource scheduling.

Reference	Algorithm, policy or strategy	Problems addressed	Improvement/achievement	Weakness/limitations
Duan et al. (2016)	• **Fractal Prediction** • **Improved Ant Colony Algorithm**	Energy aware scheduling for virtual machines	Improve the CPU load prediction and utilization with reducing energy	Compare with only traditional algorithm
Ding et al. (2015)	• **Algorithm 1** EEVS • **Algorithm 2** VM Allocation • **Algorithm 3** VM Process • **Algorithm 4** Reconfiguration	Reduce the energy consumption	Consumes less energy and processes more VMs successfully	Ignore the processing power and VM Migration
Dong et al. (2015)	• Most Efficient Server First (MESF) task scheduling scheme	Reduction of energy while keeping the response within a constrained time	Reduce the data center energy consumption	Focus only on energy consumption
Hosseinimotlagh et al. (2015)	• Smart Energy-aware Task Scheduling (SEAT)	VM scheduling and allocation of task	Decrease energy consumptions and reduces the turnaround times of real-time tasks.	Focus only on energy consumption
Jena (2015)	• Multi-Objective Particle Swarm Optimization (MOPSO)	Task scheduling	Reduce energy consumption, makespan and increase the profit	Focus only on only energy and makespan
Jiankang et al. (2014)	• VM-P algorithm • VM-Mig algorithm	VM scheduling	Better performance	Resource overloading and live migration does not control alteration of VM.
Liu et al. (2014b)	• Ant Colony Optimization Virtual Machine Placement (ACO-VMP)	VM placement	Reduce the energy consumption	Do not compare with other algorithms
Shuja et al. (2014)	• Data Center wide Energy Efficient Resource Scheduling framework (DCEERS)	Minimum cost multi commodity flow inside the data center	Reduce energy consumption	Throughput differs for resource scheduling reliant on number of resources
Xiong and Wu (2014)	• PSO algorithm	Resource scheduling in data centers	Reduce the power consumption and enhance the performance of data center,	Focus only on energy consumption
Ghribi et al. (2013)	• Exact VM Allocation Algorithm	VM migration	Consolidation and save energy consumption	Compare with only basic algorithm
Beloglazov et al. (2012)	• **Algorithm 1:** Modified Best Fit Decreasing (MBFD) • **Algorithm 2:** Minimization of Migrations (MM)	Reducing power consumption of a data center	Reduce the energy consumption in data centers.	
Chen et al. (2012)	• Hybrid genetic algorithm	Reduce the energy consumption and cost for the data centers	Improve the resource utilization and saving energy cost	
Luo et al. (2012)	• Resource scheduling algorithm for isomorphism nodes	Energy aware resource allocation	Effectively save energy.	If algorithm adjusts the energy consumption, it may reduce system performance, which increase the execution time,
Van Do and Rotter (2012)	• Analytic performance model	Energy and heat aware allocation	Energy consumption can be achieved	Focus only on energy consumption
Mezmaz et al. (2011)	• Hybrid genetic algorithm	Task Scheduling	Minimize the makespan and energy consumption	Do not compare with existing techniques

Table 4
Load balancing aware resource scheduling.

Reference	Algorithm, policy or strategy	Problems addressed	Improvement/achievement	Weakness/limitations
Pan et al. (2015)	• Interaction Artificial Bee Colony (IABC) Algorithm	Load balancing	Algorithm is more efficient in all three cases and less execution time.	Compare only with Artificial Bee Colony algorithm
Cho et al. (2014)	• Ant Colony Optimization with Particle Swarm (ACOPS)	VM scheduling	Improve the performance and reduce the workload	User requests are categorized into only three types
Wang et al. (2014)	• **Algorithm 1** Job spanning time and Load balancing Genetic Algorithm (JLGA) • **Algorithm 2** Greedy Initialization (GI)	Load balancing	Better performance	Avoid the priority
Xue et al. (2014)	• ACO-LB (Load balancing optimization algorithm based on ant colony algorithm)	Load balancing	Better performance and improve utilization rate	Ignore the cost
Zhao et al. (2014)	• Firefly algorithm	To make full use of the resources	Improve the performance	Practical implementation
Mhedheb et al. (2013)	• Thermal-aware vm scheduling scheme	VM scheduling	Reduce the energy consumption	Do not compare with other algorithms
Li et al. (2011)	• Load Balancing Ant Colony Optimization (LBACO)	Task scheduling	Improve the performance	Fix the range of task 100 to 500 in simulation

the availability, reliability, throughput, recovery time, fault tolerance and SLA of both cloud provider and users (Abdelmaboud et al., 2015). At the time of scheduling the resources, the QoS for the requested demand should not decrease. The poor utilization of resources, increasing the failure rates, non-availability of resources, violates the SLA are major issues to uncertainty of QoS (Ardagna et al., 2014).

The Proposed technique uses heuristic algorithm using Load Balancing Mutation a particle swarm optimization (LBMPSO), to achieve reliability in task scheduling. LBMPSO considers execution time, cost, makespan, and load balancing among the tasks and VMs. The results show that LBMPSO improves the performance of the mentioned parameters. Especially, it improves the availability and reliability of cloud resources (Awad et al., 2015).

Many algorithms are suggested to resolve the problems like Ant colony, priority and cost based algorithms, but these algorithms consider cloud environment as non-fault, which indicates to cut down in performance of current algorithms. Therefore, fault and load aware Honey Bee scheduling algorithm is recommended for IaaS cloud. The proposed algorithm takes into consideration failure rate and workload on a data center to enhance the performance and QoS in IaaS cloud environment (Gupta and Ghrera, 2015).

VMs scheduling problem is distinct from an allocation of a set of VMs to a set of PMs. Kumar and Raza (2015) focus on PSO based VM scheduling approach for VM placement in IaaS cloud. The approach emphases on efficient VM allocation to decrease the wastage of resource and quantity of used servers. Simulation tests show the perceive allocation of VMs to the PMs and to estimate the suggested algorithm with respect to performance and scalability.

Lakra and Yadav (2015) propose a multi objective task scheduling technique for mapping tasks to VMs in order to enhance the throughput of the data center, and to further decrease the cost according to SLA. This algorithm is simulated using CloudSim and the conclusion shows that the algorithm performs better by minimizing the execution time and increasing the throughput.

The objective of research work comes out from the challenges to search the best resource and workload according to user demand. In real scenario, the three main QoS parameters are considered for efficient utilization of resources. These include minimization of the execution cost of resources, reducing energy consumption and execution time of workloads. The main emphasis of the research work is to recommend the cloud workload management framework, clustering of workloads through machine learning techniques. Proposed resource scheduling policies deal with the cost time based policy, time based policy, cost based policy and bargaining based scheduling policy. The proposed optimization technique is used for the minimization of the execution cost and time for resource scheduling. Further, reducing the energy usage and enhancing the perform of existing scheduling algorithms if used in Singh and Chana (2015).

Hung et al. (2014) explain a procedure for task scheduling while keeping in view the conflicts related to network and expenditure of cloud in order to minimize the revival time for the improvement and betterment of consistency and accessibility of cloud services. However, network contention and cost aware scheduling algorithm constitute on the exchange between network contention and the economic expenditure. Moreover, in comparison with other available methods, this procedure performs well as compared to others instead of preferable processing time in accordance to amount given by the consumers. However, the processing time consists of the time required for recovery as in the case of failure. Moreover, task scheduling is considered as primary issue for attaining high competence in cloud computing.

Regardless of this, Kumar et al. (2014) suggest a three step method that constitutes on resource selection, setting up of user's requests with the mutual resources and the novel resource allocation along with Adaptive Job Scheduling algorithm which helps in improving the QoS distributed by clouds. However, regarding job scheduling a new weight matrix is utilized effectively to plan the jobs in accordance with accessible resources. Moreover, planned and projected approach enhances the consistency of available resources for job and helps in the reduction of time for completion of job as it enhances the QoS distributed to end users and in the end it estimates and suggests a methodology while utilizing the recognized heuristics.

Storage virtualization is considered as main characteristics of cloud storage, and it varies from conventional storage, as it suggests the idea of storage device from substantial to logical. In this regard, for enterprise storage resource consumption it makes a straightforward combined structure. However presently, all highly performing disseminated files systems utilize in cloud storage platform employ the concept of GFS of Google, Blue Cloud of IBM and S3 of Amazon. As this is the base for an enhanced and enhanced version of cloud storage model anticipated by Wang (2014).

In the IaaS cloud computing, Gupta et al. (2013) use scheduling algorithm to allocate a VM and schedule the request of the user. However, there are dynamic natures of user who have a smaller amount of resources and has funded less, user with the huge amount of resources and has funded more, and an open user or a public user. Therefore, charming into concern basic scheduling algorithm, where a user demand is assigned a VM depend on the load and the cost of data center. On the other hand, it is not charming into concern the properties of the data center. Due to this purpose, a data center with high QoS is assigned to a public user and the demand from the other user who has funded more will be distributed to a data center with low QoS. In this regard, a trust management model is recommended to reduce this problem, by attractive into concern VM monitor features which differ from data center to data center. Then the scheduling algorithm uses these trust value to increase the efficiency of the resource scheduling.

Using the fuzzy clustering, Li et al. (2013) propose the model and algorithm for dissolving the appropriate resources to tasks, which specifically attain the requests of tasks and standby the powerful resources for upcoming demand. Further, the algorithm can efficiently escape the allocating of powerful resources to simple or medium scale tasks or assigning poor resources to complex large scale tasks. It can because of failure scheduling of tasks or misuse of resources. With the concern of the reliability of resources, proposed algorithm can intensely recover the task scheduling and stability of resource management in the cloud.

The lease elimination is not possible in the current Haizea systems. In the proposed system, the lease cancelation is implemented. In the proposed technique the user can cancel the scheduled lease, before the start of its execution. Furthermore, the cloud providers can cancel the lease when the data or demand details are not available at the execution time. Whenever the cloud provider or any user vacates any slot, the vacant slot will be used for another unscheduled lease. It greatly improves the resource utilization (Nivodhini et al., 2013).

In the study, Sun et al. (2013) review the IaaS model for public cloud and analyze the performance stream according to the waiting line concept. The main aim is to increase the performance of distinct VM and platform usage, with give out one straining algorithm depend on the user request to discover an optimal resource for user's VM request. Based on the model, researchers give out the backup and disaster recovery algorithm for the cloud computing platform. This algorithm is confirmed on the cloud network platform, which enhances the QoS of the entire platform.

Reliability based model assists the cloud scheduler in the scheduling of tasks with the cloud infrastructure and help in performing fault tolerance. There are separate reliability assessment algorithms for general applications and real time applications. The algorithm for general application is adaptive and more convergent towards failures. The algorithms for real time computing do the reliability assessment on the basis of timeliness of result and are also more convergent towards failures (Malik et al., 2012).

Resource scheduling is the basic portion of resource management in cloud computing. GA has universal optimization capability, elasticity, and implicit parallelism, which are not found in other algorithms. The effective resource scheduling to achieve an appropriate task, upgraded genetic algorithm is implemented in research of resource scheduling for cloud computing. Finally, simulation based on cloudsim results show the accuracy with the strength of the scheduling algorithm (Cui et al., 2011).

Moreover, Li and Li (2011) shows the cloud resource placement with infrastructure SLA constitute on the Pareto optimality theory that distinguishes the need of user via utility function as it makes the use of Pareto for the up gradation of resource scheduling strategy in order to accomplish the favorable resource allocation. The study explains the cloud bank model of cloud computing along with the viability of mathematical proof for using the Pareto optimality theory in resource allocation. Table 5 comprehensively compares the previous various techniques that are used for load QoS aware resource scheduling, while the parameters used for QoS techniques are shown in Table 11.

4.6. *Utilization aware resource scheduling*

Fundamentally, the success of cloud computing services depends on the proficient utilization of cloud resources. However, the cloud providers have finite number of resources and attempts to compelled them to maximum utilization (Wang et al., 2015). It is challenging to apply scheduling scheme or technique to achieve the various needs of users with utilization of all resources efficiently, when many tasks or users require a lot of resources at the same time (Brummett and Galloway, 2016).

Furthermore, the critical issue is considered as a quality of service in cloud computing, a task scheduling approach with multi-dimensional QoS constraints (including the response time, execution time and resource utilization) is recommended by Jiao et al. (2015). Under the multi-dimensional QoS constraints, on the other hand, request may conflict between the users, which are possible to boost the concern of the scheduling problem. QoS performance, several advantages and the amount of load balancing for nodes are consider in the task scheduling scheme. The immune clone algorithm is suggested for numerous QoS constraint for the task scheduling problematic issue. The review and experimental results express that the immune algorithm is performed superior to the other heuristic algorithms in the response time, execution time and utilization rate.

Resource provisioning for cloud services in an efficient way is significant for any resource allocation model. All the models consider the computing and the network resources to represent accurately and serve practical needs. The main aim of the algorithm is to execute with the intention of decreasing the delay of the connection requests. Four shared scheduling algorithms are proposed in the model that are usable for scheduling the VM in data center. It schedules the connection requests based on the available network paths. The method is based on distribution and duration priority technique that provides the least delay while keeping in view the problem constraints (Abu Sharkh et al., 2013).

Cao et al. (2013) suggest an efficient VMs provision method as well as job setting up a strategy that can make transaction between cloud providers and users. In this regard, python based

Table 5
QoS aware resource scheduling.

Reference	Algorithm, policy or strategy	Problems addressed	Improvement/achievement	Weakness/limitations
Awad et al. (2015)	• Load Balancing Mutation a particle swarm optimization (LBMPSO)	Task scheduling	Improved overall performance	Fix the no of task, VMs and data centers
Gupta and Ghrera (2015)	• Load and Fault Aware Honey Bee Scheduling Algorithm	Fault aware resource scheduling	Increasing request rate	Compare with only Basic load aware Honey Bee Algorithm
Kumar and Raza (2015)	• Particle Swarm Optimization	VM allocation	Better performance	Compare with traditional algorithms
Lakra and Yadav (2015)	• Multi-Objective Task Scheduling Algorithm	Task scheduling	Enhance throughput with minimizing throughput	Compare with traditional algorithms
Singh and Chana (2015)	• Compromised Cost Time Based (CCTB) scheduling policy • Time Based (TB) Scheduling Policy • Cost Based (CB) scheduling policy • Bargaining Based (BB) scheduling policy	Enhance the proficiency of computing resources	Provide better and optimum solution for resource scheduling issues.	Cost depends upon the no of resources
Hung et al. (2014)	• Directed Acyclic Graph • **Algorithm 1** Network contention and Cost aware scheduling	Reduce recovery time	Improve performance and efficiency	Increase the cost
Kumar et al. (2014)	• Resource Allocation and Adaptive Job Scheduling algorithm	QoS aware resource scheduling	Reduce job execution time and increase reliability and %age of resource allocation	Consider distributed resources both at local and global sites
Wang (2014)	• A scheduling algorithm based on Priority (SAP).	Cloud storage scheduling	Reduce loss rate	Use the NS2 Simulator and not compare with other optimized algorithms
Gupta et al. (2013)	• Trust Management Model • Trust and Reliability Based Scheduling Algorithm for Cloud IaaS	To enhance the resource scheduling	A request with a higher trust value is being allotted to the data center with higher trust value	Results are not compared with other models
Li et al. (2013)	• Multi-QoS and Trusted Task • Scheduling Model (MQT_TSM) • Multi-QoS and Trusted Task Scheduling Algorithm (MQT_TSA)	Multi QoS and trusted task scheduling	Improve the performance	Practical implementation
Nivodhini et al. (2013)	• Wait queue in immediate leases • Reservation queue in AR leases • Deadline queue in DLS leases	To minimize the rejection rate in immediate leases AR and DLS leases.	Reduce the rejection rate and improve the utilization.	
Sun et al. (2013)	• Disaster Recovery Model Of IaaS Public Cloud	Resource re-allocation, disaster recovery, load balance	Provide existing VMs to the cloud users.	Do not compare with existing algorithms
Malik et al. (2012)	• Reliability assessment for general & real time cloud computing • Algorithm 1: Reliability Assessment Algorithm for General Applications - Failure Convergent • Algorithm 2: Reliability Assessment Algorithm for Soft Real Time Applications • Algorithm 3: Reliability Assessment Algorithm for Hard Real Time Applications	Reliability assessment of computes instance.	The reliability model assisted the cloud scheduler in the scheduling of tasks with the cloud infrastructure and help in performing fault tolerance	
Cui et al. (2011)	• Improved Genetic Algorithm	Efficient resource scheduling	Improve efficiency	Compare with the basic algorithm
Li and Li (2011)	• Pareto optimality theory	Efficient resource scheduling	Improve performance	It does not quantize resources and consumer's requirements

simulation suite simply is taken in to assemble the model. It identifies the most pragmatic condition of VMs management in cloud computing that is randomly arrival of jobs as each job is associated with various numbers of tasks along with variation in completing time. However, VMs allocation procedures and vibrant jobs scheduling are revealed along with constraint of threshold and various scheduling policies. In this regards, experiments revealed that SJF policy is considered to be fit towards these types of circumstances and it can attain improved working in order to get higher QoS.

Resource scheduling algorithms express a vital part, where the purpose is to schedule resources effectively to decrease the turnaround time and enhance the resource utilization. Zhang et al. (2013a) deal a PSO based strategy schedules applications to cloud resource taking into account both communication cost and current workload. In addition, an innovative inertia weight is familiarized in order to become the global search and local search effectively and avoid plunging into the local optimum. To conclude, experiment with application workflows by varying its performance and convergence analysis.

Zhang et al. (2013b) determine the dynamic resource management problematic issue for multiple virtual machines in cloud computing. Performance of the CPU is a form of high sensitive resource for the QoS in the IaaS Cloud. The global CPU regulation algorithm depends upon the utility optimization theory is suggested for the multiple virtual machines CPU regulation system that contains both the local and global utility device. The experiment results show that the global CPU regulation can increase the response time, which is a full assurance to the quality of user services.

In recent studies, researchers prefer resource scheduling as it constitute on ant colony algorithm for cloud computing and it helps in controlling the disadvantages of big size of node as it results in lower distribution of resources for a single node and it makes sure that user job can be fulfilled in time. However, still there is a chance that algorithms easily can get into confined solutions and ends up in ineffective resource scheduling. Wen et al. (2012) explain that merging of ant colony optimization algorithm with particle swarm optimization algorithm in order to enhance the utilization of resource scheduling in cloud computing environment.

Kim et al. (2010) prove the performance enhancement of adaptable scheduling schemes are scheduled and executed the scientific applications, in the cloud computing environment. Organizing the computing environment, Haizea as a scheduler and OpenNebula as a middleware are used to deliver infrastructures for the cloud computing. The result shows that better performance of the VM for parallel executions in certain terms of time. Another, experiment confirms the comparison of the VM priority intensive scheduling and job scheduling, with using the VMs scheduling granularity increases usage of resources and decreases the finishing time of systematic applications on cloud computing. Various utilization aware resource scheduling techniques are compared according to the different matrix and primary differences as listed in Table 6. Moreover, the parameters used for utilization of resources are shown in Table 12.

5. Comparative analysis of the parameters used in recent studies

In this section, the resources and parameters used in evaluating the current research is presented in Tables below. The tables show that the IaaS cloud resources used by the existing researchers are CPU, VM, node, network, and storage. Resource scheduling in cloud computing means to allocate and schedule the best possible resources or tasks/cloudlets to the users according to their dynamic demands with the consideration of different parameters like time, response time, execution time, cost, makespan, bandwidth/ speed, throughput, reliability, performance, availability,

Table 6
Utilization aware resource scheduling.

Reference	Algorithm, policy or strategy	Problems addressed	Improvement/achievement	Weakness/limitations
Jiao et al. (2015)	• Immune Clone Algorithm	Multiple QoS constrained	Better performance and algorithm has less execution time than other algorithms	
Abu Sharkh et al. (2013)	• Software defined networking (SDN) • SDN controller • Equal Time Distribution Technique • Node Distance Technique • Resource Based Distribution Technique • Duration Priority Technique • Greedy Algorithm • ED-GA: Equal Time Distribution technique and Greedy algorithm. • RB-DP: Resource Based Distribution technique and Duration Priority technique.	Resource utilization with minimizing the cloud provider cost, at a same time fulfilling all the client's demands	Maximize the resource utilization and minimize the cloud provider cost	
Cao et al. (2013)	• FCFS • SJNF • SJEF • LJNF • LJEF	Effective resource utilization with reducing the user's cost	SJF policy performs better to achieve higher QoS.	Focus on only basic algorithms
Zhang et al. (2013a)	• PSO based Strategy	Optimal resource scheduling	Better performance	Weights are consider only global optimum
Zhang et al. (2013b)	• CPU Global Regulation Algorithm	Dynamic resource management for multiple virtual machine	Better performance	Do not compare with existing algorithms
Wen et al. (2012)	• ACO algorithm and PSO algorithm	Optimal resource scheduling	Improve resource utilization	Focus on local optimum
Kim et al. (2010)	• Aerodynamic applications (CFD) • Two different leases • Best-effort(BE) • Advanced-reservation (AR)	Utilization of resources	Reduced execution time and improve resource usage	Scheduling is based on priority

memory, energy, temperature, priority, utilization, SLA and workload. Heterogeneous parameters are to be considered for resource scheduling. In our survey, we show the number of parameters that have been presented in existing studies, for the purpose of comparison.

Basic intention of cloud providers is to get the most out of their revenue and profit. For this purpose, various techniques are used that increase users' satisfaction, avoid SLA violations decrease the energy consumption and enhance the resource utilization by providing on demand services. On the other hand, cloud users want to minimize the expenses and maximize the overall performance. The parameters used for cost aware resource scheduling in previous techniques are displayed in Table 7.

5.1. Cost

It refers to the total amount that is paid by the cloud users to the cloud providers against the usage or utilization of resources. The main purpose of cloud provider is to maximize the revenue and profit while cloud users want to minimize the expenses. So cost focuses on all four parameters including the provider revenue, provider profit, user expenses and resource cost (Li et al., 2009).

In cloud computing, every cloud user wants the higher performance of the service which is provided by the cloud provider. This efficiency is achieved by the reducing the makespan, execution time, responses time and increasing the bandwidth or speed. The parameters used for efficiency aware resource scheduling in existing techniques are shown in Table 8.

5.2. Makespan

It determines the maximum completion time of cloudlet or task, when the resources are allocated to the users. So it is necessary to reduce the makespan of specific cloudlet otherwise the request will not be fulfilled on time (Abdulhamid et al., 2015).

5.3. Execution cost

The execution cost is the difference between the ideal cost with Service level agreement and what is actually done in the execution of task or cloudlets (Deelman et al., 2008).

5.4. Execution time

It determines the time that is consumed by the implementation of cloudlet or task. Minimum execution time is required for cloud provider and user, to enhance the efficiency. It also effects on the energy consumption, utilization, load balancing and overall performance (Puschner and Koza, 1989; Xiong and Perros, 2009).

5.5. Response time

It refers to the time, when a cloudlet or task responses to a specific input and start processing. It can be calculated by the sum of waiting time and submission time. It also directly effects on the waiting time of the cloudlet or tasks (Bashir et al., 2013).

5.6. Bandwidth/speed

Bandwidth is also defined as the amount of data that can be transferred or executed in a fixed amount of time. It is usually preceded in bits or bytes per second (bps) (Buyya et al., 2009).

5.7. Priority

It refers to a cloudlet or task, that is regarded or treated more important than the others. It is right to take precedence or to proceed before others (Ghanbari and Othman, 2012).

Due to the rising demand for cloud computing, the numbers of cloud data centers are enlarged more and more, so that the energy problems of cloud computing environment are converted progressively prominent. The energy consumption of cloud data center is protected efficiently using the rational allocation of resources scheduling, with effective use of resources. The parameters used for energy aware resource scheduling in current techniques are stated in Table 9.

5.8. Energy consumption

It refers to the usage of energy or power, which is distinct in the use of energy by resources as a process of performing the operation of cloudlet or tasks. Less energy consumption can increase the profit for the cloud user and move towards the green computing (Beloglazov et al., 2012; Tziritas et al., 2013).

With the help of load balancing, it is easy to reduce the amount of workload that is placed on a server in the data centers. Overloaded and unbalanced resources are cause the SLA violation or failure of system. Load balancing is also useful for maintaining the SLA and reducing the chances of SLA violations. The parameters used for load balancing aware resource scheduling in existing techniques are shown in Table 10.

5.9. Workload

Workload is used for load balancing, it can minimize the amount of load that has been placed on VMs, servers or nodes. Unbalanced and overloaded resources can be a cause of system failure or SLA violation. Degree of Imbalance is a metric for load balancing in the cloud computing (Nuaimi et al., 2012).

Main feature of cloud computing is to maintain the quality of service (QoS) delivered to the cloud user in the form of availability, reliability, failure rate and much more. Most researchers do not consider reliability and availability for resource scheduling in cloud computing environment because of the complexity to achieve these parameters. The parameters used for QoS aware resource scheduling in current techniques are presented in Table 11.

5.10. Availability

In cloud computing, it determines the resources or tasks that are accessible, suitable and ready for the use or service in specified location and incorrect format (Nabi et al., 2016; Hassan et al., 2014).

5.11. Throughput

It refers to the rate of production or rate in which cloudlet or task can be processed in a certain period of time. It always should be minimized for a high performance (Mustafa et al., 2015).

5.12. Reliability

It determines the frequency rate in which a cloudlet or task can complete the execution or to perform its required functions under stated conditions for a specified period of time. It is highly used to avoid the failure rate of cloud computing (Armbrust et al., 2010; Faragardi et al., 2013).

5.13. Recovery time

The recovery time is the maximum acceptable length of time

Table 7
Parameters for cost aware resource scheduling.

Reference	Resources					Task/cloudlet	Parameters														
										Cost											
	CPU	VM	Node	Network	Storage		Time	Response time	Execution time	Provider revenue	Provider profit	User expenses	Resource cost	Makespan	Bandwidth/Speed	Reliability	Performance	Memory	Energy	Utilization	Workload
Raju et al. (2016)	✓	✓			✓	✓			✓		✓										
Bansal et al. (2015)		✓			✓	✓			✓				✓		✓			✓			✓
Kapur (2015)		✓					✓						✓								✓
Zuo et al. (2015)						✓						✓	✓	✓						✓	
Netjinda et al. (2014)						✓			✓	✓		✓	✓								✓
Liu et al. (2013)			✓			✓		✓			✓					✓		✓	✓		
Shen et al. (2013)		✓									✓	✓									
Tiwari et al. (2013)		✓				✓					✓		✓								
Yi et al. (2013)			✓	✓		✓	✓				✓	✓									
Moschakis and Karatza (2012)		✓				✓		✓					✓		✓		✓				

Table 8
Parameters for efficiency aware resource scheduling in IaaS.

Reference	Resources					Task/ cloudlet	Parameters											
								Efficiency										
	CPU	VM	Node	Network	Storage		Time	Response Time	Execution Time	Execution Cost	Makespan	Bandwidth/ Speed	Priority	Availability	Memory	Utilization	SLA	Workload
Abdullahi et al. (2016)						✓		✓			✓							✓
Babu and Samuel (2016)		✓	✓			✓					✓							
Bansal et al. (2016)		✓				✓				✓								✓
Ma et al. (2016)	✓		✓			✓		✓	✓									
Raghavan et al. (2015)						✓		✓		✓								
Shojafar et al. (2015)		✓	✓	✓							✓							✓
Thomas et al. (2015)		✓				✓					✓		✓					
Abdulhamid et al. (2014b)		✓				✓			✓									✓
Javanmardi et al. (2014)		✓				✓			✓		✓							✓
Lin et al. (2014)	✓	✓				✓	✓					✓						
Liu et al. (2014a)						✓			✓									
Pan et al. (2014)			✓			✓			✓					✓				✓
Yang (2014)			✓			✓	✓											
Ergu et al. (2013)						✓		✓		✓			✓					
Le et al. (2013)		✓				✓	✓		✓								✓	
Ramezani et al. (2013)	✓	✓				✓			✓	✓		✓						
Sindhu and Mukherjee (2013)	✓	✓				✓					✓							
Tsai et al. (2013)	✓					✓	✓			✓								
Wang and Yu (2013)		✓	✓			✓				✓	✓							✓
Wang et al. (2013)						✓				✓	✓							
Wu et al. (2013)	✓	✓			✓	✓					✓		✓					✓
Zhang et al. (2013c)		✓													✓			✓
Zhu and Liang (2013)	✓		✓	✓	✓				✓			✓						
Achar et al. (2012)		✓				✓			✓							✓		✓
Zhong-wen and Kai (2012)			✓				✓			✓								
Sindhu and Mukherjee (2011)		✓		✓	✓	✓					✓					✓		

Table 9
Parameters for energy aware resource scheduling in IaaS.

Reference	Resources					Task/ Cloudlet	Parameters											
	CPU	VM	Node	Network	Storage		Time	Response time	Execution time	Makespan	Bandwidth/ speed	Throughput	Reliability	Energy	Temperature	Utilization	SLA	Workload
Duan et al. (2016)	✓	✓					✓							✓		✓		✓
Ding et al. (2015)		✓							✓					✓				✓
Dong et al. (2015)				✓		✓		✓						✓				
Hosseinimotlagh et al. (2015)		✓				✓	✓						✓	✓				
Jena (2015)						✓	✓						✓	✓				
Jiankang et al. (2014)	✓	✓		✓	✓									✓				✓
Liu et al. (2014b)	✓			✓										✓		✓		
Shuja et al. (2014)							✓					✓		✓				✓
Xiong and Wu (2014)	✓	✓	✓				✓							✓				
Ghribi et al. (2013)		✓	✓				✓		✓					✓				
Beloglazov et al. (2012)		✓		✓										✓	✓		✓	
Chen et al. (2012)	✓	✓		✓	✓									✓		✓		✓
Luo et al. (2012)			✓		✓				✓					✓		✓		
Mezmaz et al. (2011)	✓					✓				✓	✓			✓				

that a cloudlet or task can be down after a failure or disaster occurrence. It is a function of the extent to which the interruption disrupts normal operations and the amount of revenue lost per unit time as a result of the disaster (Dillon et al., 2010).

5.14. *Fault tolerance*

It refers to the mechanism that keeps the track of ongoing operations. In the case of server failure, cloudlet or task should be directly shifted to another running system to confirm the required level of service. VM migration can support in relocating cloudlets or tasks from one server to another. Faults or system failures imperatively influences on QoS and leads to a major loss in business (Cheraghlou et al., 2015; Jhawar et al., 2013).

5.15. *SLA*

In cloud computing, a SLA is a contract between cloud provider and user to guarantee the required level of services. It encloses several specifics of service level that will be delivered to the cloud users, such as IaaS resources. Therefore, proposed technique must be appraised, always avoid the SLA violation and required activities should be taken. Otherwise, cloud provider is responsible for its breach has to pay the penalty to cloud users (Patel et al., 2009).

In cloud computing, optimum usage of resources in a proficient way specifies a high influence on the overall profit of the system. Effective resource utilization is suitable for increasing the profit and reducing the energy consumption by reducing the amount of resources in use. The parameters used for utilization aware resource scheduling in current techniques are mentioned in Table 12.

5.16. *Utilization*

Resource utilization is the usage of a resource in such a way that is maximized through outcomes. Maximum resource utilization can be achieved by reducing the amount of resource in use to increase the profit and minimize the energy consumption with satisfying the users' demands in an efficient way (Madni et al., 2016; Zhang et al., 2010).

Fig. 3 shows the total percentage of the parameters used for resource scheduling in cloud computing based on proposed classified categories. The percentage of the parameters is shown in clockwise direction, it starts from the time and ended on the workload, sequentially.

Fig. 3(a) illustrates a pie chart of the parameters used in cost aware resource scheduling category. In this category, the focus of the researches is based on the cost parameter to reduce the cost of users' expenses and increase the revenue or profit for the providers. It can be achieved with the help of minimizing the response time, execution time, energy consumption and makespan while maximizing the bandwidth/ speed, balancing the workload, optimal utilization of resources. These parameters are also considered and shown with percentage of pie chart.

In Fig. 3(b), the parameters used for efficiency aware resource scheduling are demonstrated. In this category, makespan, response time, execution time, execution cost and priority are the focus of the researches in order to achieve efficiency. These all parameters are considered to enhance the efficiency of the resource scheduling and utilization effectively, which is a requirement of every cloud user.

The pie chart in Fig. 3(c) expresses the parameters according to the energy aware resource secluding. In this category, more focus is given to reduce energy consumption, while the other major parameters like execution time, utilization and workload are also considered with it to reduce the usage of energy. It directly affects

Table 10
Parameters for load balancing aware resource scheduling in IaaS.

Reference	Resources					Task/cloudlet	Parameters						
	CPU	VM	Node	Network	Storage		Execution time	Cost	Makespan	Energy	Temperature	Utilization	Workload
Pan et al. (2015)		✓				✓	✓						✓
Cho et al. (2014)	✓	✓			✓		✓					✓	✓
Wang et al. (2014)			✓			✓			✓				✓
Xue et al. (2014)		✓				✓	✓						
Zhao et al. (2014)			✓				✓	✓					
Mhedheb et al. (2013)	✓	✓				✓				✓	✓		✓
Li et al. (2011)		✓				✓			✓				✓

the cloud providers' profit.

The load balancing aware resource scheduling is displayed in pie chart of Fig. 3(d). In this category workload, execution time and makespan are the important and most used parameters. Load balancing can be achieved by reducing the execution time and makespan of the task or cloudlet and it helps to enhance the performance.

Fig. 3(e) presents the pie chart of the parameters according to the QoS aware resource scheduling category. In this category the focus of the research is based on the availability, reliability, SLA violation and failure rate. However, the other parameters are also considered by the researchers to optimize the resource scheduling. There is a need to perform more work in this area, to increase the performance and better service. These parameters are also difficult in implantation that is why these are not more considerable parameters in the research.

Utilization aware resource scheduling parameters are presented in pie chart form in Fig. 3(f). Utilization of resource is focused parameter in this category. It can be attained with the help of reducing the execution time, response time and cost, while increasing the bandwidth/speed. Due to the above cause, these parameters are also considered by the researchers for optimal resource scheduling.

In cloud computing, cloud providers organize a huge amount of IaaS resources of Infrastructure as a Service, which are allocated or assigned to cloud users as on demand. The concept of IaaS is basically offering Hardware as a Service. IaaS deals computing, storage and network as standardized services over the cloud. CPUs, VMs, Storage systems, Nodes and Networks (switches, routers, and other systems) are shared resources of IaaS and made accessible to handle workloads (Mell and Grance, 2011). Fig. 4 shows the IaaS cloud resources are deliberated by various researchers in their research for resource scheduling in cloud computing from the time period of 2010–2016.

It can be observed from Fig. 4 that most of the researchers concentrated on the VMs and computational resources for the research area, whereas few of them have concentrated on the remaining resources for IaaS in the cloud computing. Hence, we know that the storage and network resource are the basic needs of the cloud computing. Moreover, cloud computing is totally dependent on these resources. The focus on the VM attention is a virtualization and fluctuating demand of the cloud users, whereas other resources are requirement of the specific users for particular tasks. Also VM generate some more problematic issues in research or real cloud computing environment like VM migration and placement.

The parameters for resource scheduling in cloud computing are examined by various researchers in their research from the time period of 2010–2016. After reviewing Fig. 5, we observed that execution time, cost, energy and workload are most useful parameters, considered by the researchers in the research field of resource scheduling in IaaS cloud computing. Even though the reliability, response time, utilization, bandwidth/speed and makespan are emphasized by some of the cloud researchers, but there is a need to focus more on these parameters. Whereas, the throughput, availability, priority, temperature and fault tolerance are the elementary parameters considered in cloud computing and only a few researchers have used these parameters in their research. The implementation of these parameters in the simulation or real test bed environment is too hard and highly compliable that is one of reason for less consideration. Also an algorithm or scheme is used to solve issue by concerning these parameters together.

In reality, cloud is a business model, where cloud providers want to reduce the expenditure with minimizing the energy/power consumption, heat generation, storage, etc. While enhances the revenue/profit with the maximum utilization of the resource. However, the cloud users want to have high performance of the service with the minimum cost and time. For these causes, time, cost, energy, execution time and workload are considered most significant parameters in the research of cloud computing.

6. Suggested future works

In this segment, we summarize and analysis the research gap of the different strategic challenges, techniques and methodologies that are perceived during the progression of literature survey of this research for performance enhancement. A number of researchers have discussed the new research challenges that are raised by cloud computing. Fig. 6 presents the different suggested works put forward by previous research works. It shows suggested ideas according to the demand values and categorized by: very high demand, high demand, medium demand and low demand.

A number of authors have discussed the new research challenges and issues that are elevated in the area of cloud computing. Some of the crucial challenges suggested and identified by some researchers for the further research in research scheduling for the cloud computing (shown in Fig. 6) for performance improvement for resource scheduling include:

- QoS aware Resource Scheduling: service refers to a certain level of performance and availability, fault tolerance and throughput of a service. This scheme is recommended for future work by Gupta and Ghrera (2015), Huang and Ou (2014), Gupta et al. (2013), He et al. (2013), Li et al. (2013), Sindhu and Mukherjee (2013), Yi et al. (2013), Wang and Yu (2013), Li and Li (2011).
- Reliability aware Resource Scheduling is also considered as part of the QoS, which is proposed by Bansal et al. (2016), Sharma et al. (2015), Malik et al. (2012) for further enhancement.
- Workflow aware Resource Scheduling: A specific scheduling strategy for mapping or monitoring the tasks in a workflow to suitable cloud resources in order to satisfy user demand (Abdullahi et al., 2016; Raghavan et al., 2015).

Table 11
Parameters for QoS aware resource scheduling in IaaS.

Reference	Resources					Task/ cloudlet	Parameters															
													Quality of Service									
	CPU	VM	Node	Network	Storage		Time	Response time	Execution time	Cost	Makespan	Bandwidth/ speed	Throughput	Reliability	Availability	Recovery time	Fault tolerance	SLA	Performance	Memory	Energy	Workload
Awad et al. (2015)		✓				✓	✓		✓	✓	✓			✓	✓							✓
Gupta and Ghrera (2015)				✓				✓						✓								
Kumar and Raza (2015)		✓		✓		✓								✓								
Lakra and Yadav (2015)		✓				✓			✓				✓									✓
Singh and Chana (2015)									✓	✓								✓			✓	✓
Hung et al. (2014)	✓			✓		✓	✓			✓		✓				✓						
Kumar et al. (2014)									✓				✓	✓	✓				✓			
Wang (2014)			✓	✓								✓		✓								
Gupta et al. (2013)		✓					✓		✓	✓				✓								
Li et al. (2013)						✓					✓			✓								
Nivodhini et al. (2013)			✓				✓										✓			✓		
Sun et al. (2013)	✓	✓		✓	✓		✓	✓		✓							✓			✓		
Malik et al. (2012)		✓			✓									✓					✓			
Cui et al. (2011)		✓				✓	✓		✓					✓								
Li and Li (2011)	✓		✓		✓							✓						✓		✓		

Table 12
Parameters for utilization aware resource scheduling in IaaS.

Reference	Resources					Task/ cloudlet	Parameters						
	CPU	VM	Node	Network	Storage		Time	Response time	Execution time	Cost	Availability	Utilization	Workload
Jiao et al. (2015)		✓	✓			✓		✓	✓			✓	
Abu Sharkh et al. (2013)		✓	✓				✓						
Cao et al. (2013)		✓							✓	✓		✓	
Zhang et al. (2013a)		✓					✓			✓			✓
Zhang et al. (2013b)	✓	✓					✓	✓				✓	
Wen et al. (2012)			✓				✓			✓	✓		
Kim et al. (2010)		✓							✓			✓	

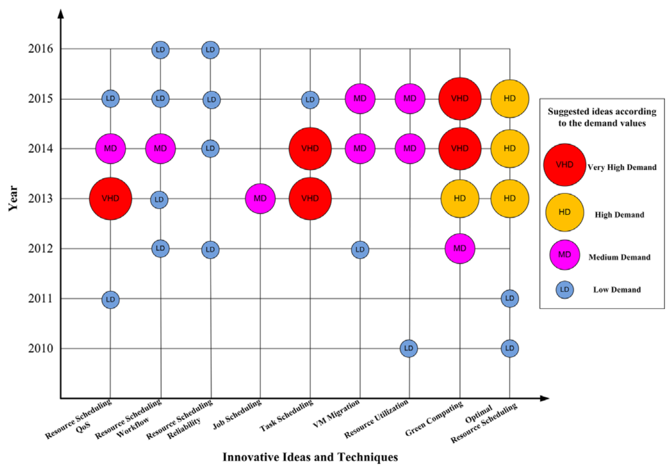

Fig. 6. Innovative ideas for cloud computing.

- Job Scheduling: In getting the best performance in cloud computing, job scheduling is considered as a primary task. Nevertheless, in geographically distributed resources many tasks will be kept in view to carry out its functions. It creates a major problem for the design of an efficient scheduling algorithm and the implementation of it in a cloud computing environment (Huang et al., 2013; Shen et al., 2013).
- Task Scheduling: The task scheduling policies and schemes have a directly effects on the efficient usage of resources and efficiency of the user's tasks in cloud computing. Therefore, to achieve optimum scheduling and allocation of users' tasks is still considered a most important issue in cloud computing (Thomas et al., 2015; Lin et al., 2014; Ma et al., 2014; Netjinda et al., 2014; He et al., 2013; Li et al., 2013; Ramezani et al., 2013; Wang and Yu, 2013).
- Optimal Resource Scheduling: In cloud computing, optimization of resource scheduling is a necessity for the utilization of resources efficiently, with considering all aspects of both cloud providers and users (Jena, 2015; Gabi et al., 2015; Singh and Chana, 2015; Pan et al., 2015; Xue et al., 2014; Zhao et al., 2014; Tiwari et al., 2013; Rimal et al., 2010).

Although cloud computing is already a conventional technology and is still developing, it is also predictable to see it settle down, converging its current diversity into more streamlined solutions. Major research challenges and issues in cloud computing are:

- VM Migration and Placement: The scheduling of VMs on PMs is a problematic issue in cloud computing. Data centers are equipped with quite a lot of PMs, each of them works for incoming VM request. Traditional approaches used for VM scheduling, including FIFO and some others. All PMs are contained in a list and the requirement for the selection of VM is checked (Kumar and Raza, 2015; Shojafar et al., 2015; Liu et al., 2014b, Chen et al., 2012; Cho et al., 2014; Jiankang et al., 2014).
- Resource Utilization: Efficient utilization or usage means that task/cloudlets use the resources allocated to them as fully as possible (Kapur, 2015; Hosseinimotlagh et al., 2015; Zhang and Su, 2014; Zhao et al., 2014; Kim et al., 2010).
- Green Computing: With the unexpected growth of data centers, particularly the appearance of cloud computing, data center is not only providing the unified servers and services. Preservation of foundation, cloud users' demands have progressed into a group of huge amounts of data processing and storing as one of the high performance computers to emphasis on. So green computing stresses on the minimum and optimum usage of energy, power, temperature and storage (Kapur, 2015; Kalra and Singh, 2015; Sharma et al., 2015; Shojafar et al., 2015; Zhang and Su, 2014; Abu Sharkh et al., 2013; Mhedheb et al., 2013; Ramezani et al., 2013; Chen et al., 2012; Luo et al., 2012; Beloglazov et al., 2012; Tsai and Rodrigues, 2014).
- Mete-heuristic methods and algorithms for resource management: Meta-heuristics methods and algorithms used to find the best solutions to optimization problems, when exact techniques proves insufficient. With meta-heuristic techniques, any type of the objective function and proceeds into consideration various objectives of resource management (Babu and Samuel, 2016; Bansal et al., 2016; Sindhu and Mukherjee, 2011).

From Fig. 6, it shows that as at 2016, more research attention is being directed at the areas of reliably and workflow aware resource scheduling. In 2015, more research attention is being directed at green computing, resource optimality, utilization and VM migration. In 2014, there are many suggestions to direct research activities towards the areas of task scheduling, green computing, resource optimality, QoS, workflow scheduling, VM migration and utilization. Before then, there was huge interest in areas of QoS, task scheduling, green computing, resource optimality, VM migration and job scheduling. These open challenges and future works will play a decisive role in describing the technological roadmap for cultivating the future IaaS cloud computing.

7. Conclusion and recommendations

The resource scheduling techniques should be as simple as possible that require less execution time, makespan and computation power so that they consume less energy and produce less heat to work as part of green computing. Furthermore, resource scheduling techniques should provide guarantee of reliability, availability and fault tolerance. There should be a procedure to keep track of ongoing operation of resource scheduling so that fault or system failure cannot affect the QoS and may not lead to a major losses in the business. Security also plays an important role in cloud user satisfaction. Whenever clients use unsecured remote or virtualized resources, there is a chance that the resources may fall into wrong hands. So, we suggest that more work should be done to provide the privacy and security with the resource scheduling in cloud computing.

Cloud computing is an interesting field of research, for the reason of its comparative innovation and explosion development. In this paper, we presented a classification scheme and a descriptive literature review of resource scheduling for IaaS in cloud computing research. A lot of scheduling algorithms, policies and strategies are available today for resource scheduling in IaaS cloud computing, but their performance evaluation remains on open discussion. Although, current resource scheduling for IaaS in cloud computing research is still perverse due to implementation and technological concerns, such as resource management, performance enhancement, cloud provider's and cloud user's satisfaction and new research theme (like green computing and big data) regarding the social and organizational inferences of cloud computing is evolving. We hope this classification and descriptive review will deliver a portrait and reference base for academics and practitioners of the modern state of resource scheduling research and stimulate additional research interest in IaaS cloud computing.

Acknowledgment

The authors are thankful to Universiti Teknologi Malaysia (UTM) for supporting this research through RUG Grant Vote no. Q. J130000.2528.05H87.

References

Abdelmaboud, A., Jawawi, D.N., Ghani, I., Elsafi, A., Kitchenham, B., 2015. Quality of service approaches in cloud computing: a systematic mapping study. J. Syst. Softw. 101, 159–179.
Abdulhamid, S.I.M., Latiff, M.S.A., Bashir, M.B., 2014a. Scheduling techniques in on-demand grid as a service cloud: a review. J. Theor. Appl. Inf. Technol. 63.
Abdulhamid, S.I.M., Latiff, M.S.A., Idris, I., 2014b. Tasks scheduling technique using league championship algorithm for makespan minimization in IaaS cloud. ARPN J. Eng. Appl. Sci. 9.
Abdulhamid, S.M., Latiff, M.S.A., Idris, I., 2015. Tasks scheduling technique using league championship algorithm for makespan minimization in IaaS cloud. ARPN J. Eng. Appl. Sci. 9, 2528–2533.
Abdullahi, M., Ngadi, M.A., Abdulhamid, S.I.M., 2016. Symbiotic organism search optimization based task scheduling in cloud computing environment. Futur. Gener. Comput. Syst. 56, 640–650.
Abu Sharkh, M., Ouda, A., Shami, A., 2013. A resource scheduling model for cloud computing data centers. In: Proceedings of the 9th IEEE International Wireless Communications and Mobile Computing Conference (IWCMC), pp. 213–218.
Achar, R., Thilagam, P., Shwetha, D., Pooja, H., 2012. Optimal scheduling of computational task in cloud using Virtual Machine Tree. In: Proceedings of the Third IEEE International Conference on Emerging Applications of Information Technology (EAIT), pp. 143–146.
Ardagna, D., Casale, G., Ciavotta, M., Pérez, J.F., Wang, W., 2014. Quality-of-service in cloud computing: modeling techniques and their applications. J. Internet Serv. Appl. 5, 1–17.
Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., 2010. A view of cloud computing. Commun. ACM 53, 50–58.
Awad, A., El-Hefnawy, N., Abdel_Kader, H., 2015. Enhanced particle swarm optimization for task scheduling in cloud computing environments. Procedia Comput. Sci. 65, 920–929.
Babu, K.R., Samuel, P., 2016. Enhanced bee colony algorithm for efficient load balancing and scheduling in cloud. Innov. Bio-Inspired Comput. Appl., Springer
Bansal, N., Awasthi, A., Bansal, S., 2016. Task scheduling algorithms with multiple factor in cloud computing environment. Inf. Syst. Des. Intell. Appl., Springer
Bansal, N., Maurya, A., Kumar, T., Singh, M., Bansal, S., 2015. Cost performance of QoS driven task scheduling in cloud computing. Procedia Comput. Sci. 57, 126–130.
Bashir, M.B., ABD Latiff, M.S., Ahmed, A.A., Yousif, A., Eltayeeb, M.E., 2013. Content-based information retrieval techniques based on grid computing: a review. IETE Tech. Rev. 30, 223–232.
Beloglazov, A., Abawajy, J., Buyya, R., 2012. Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing. Futur. Gener. Comput. Syst. 28, 755–768.
Beloglazov, A., Buyya, R., 2010. Energy efficient resource management in virtualized cloud data centers. In: Proceedings of the 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, IEEE Computer Society, pp. 826–831.
Brummett, T., Galloway, M., 2016. Towards providing resource management in a local IaaS Cloud architecture. Inf. Technol.: New Gener., Springer.
Buyya, R., Beloglazov, A., Abawajy, J., 2010. Energy-Efficient Management of Data Center Resources for Cloud Computing: A Vision, Architectural Elements, and Open Challenges. arXiv preprint arXiv:1006.0308.
Buyya, R., Yeo, C.S., Venugopal, S., Broberg, J., Brandic, I., 2009. Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility. Futur. Gener. Comput. Syst. 25, 599–616.
Cao, Y., Ro, C., Yin, J., 2013. Comparison of job scheduling policies in cloud computing. Futur. Inf. Commun. Technol. Appl., Springer.
Chen, S., Wu, J., Lu, Z., 2012. A cloud computing resource scheduling policy based on genetic algorithm with multiple fitness. In: Proceedings of the 2012 IEEE 12th International Conference on Computer and Information Technology (CIT), pp. 177–184.
Cheraghlou, M.N., Khadem-Zadeh, A., Haghparast, M., 2015. A survey of fault tolerance architecture in cloud computing. J. Netw. Comput. Appl.
Cho, K.-M., Tsai, P.-W., Tsai, C.-W., Yang, C.-S., 2014. A hybrid meta-heuristic algorithm for VM scheduling with load balancing in cloud computing. Neural Comput. Appl., 1–13.
Cui, Y.F., Li, X.M., Dong, K.W., Zhu, J.L., 2011. Cloud computing resource scheduling method research based on improved genetic algorithm. Adv. Mater. Res. 271, 552–557.
Deelman, E., Singh, G., Livny, M., Berriman, B., Good, J., 2008. The cost of doing science on the cloud: the montage example. In: Proceedings of the 2008 ACM/IEEE Conference on Supercomputing, IEEE Press, 50.
Dillon, T., Wu, C., Chang, E., 2010. Cloud computing: issues and challenges. In: Proceedings of the 24th IEEE International Conference on Advanced Information Networking and Applications (AINA), pp. 27–33.
Ding, Y., Qin, X., Liu, L., Wang, T., 2015. Energy efficient scheduling of virtual machines in cloud with deadline constraint. Futur. Gener. Comput. Syst. 50, 62–74.
Dinh, H.T., Lee, C., Niyato, D., Wang, P., 2013. A survey of mobile cloud computing: architecture, applications, and approaches. Wirel. Commun. Mob. Comput. 13, 1587–1611.
Dong, Z., Liu, N., Rojas-Cessa, R., 2015. Greedy scheduling of tasks with time constraints for energy-efficient cloud-computing data centers. J. Cloud Comput. 4, 1–14.
Duan, H., Chen, C., Min, G., Wu, Y., 2016. Energy-aware scheduling of virtual machines in heterogeneous cloud computing systems. Futur. Gener. Comput. Syst.
Durao, F., Carvalho, J.F.S., Fonseka, A., Garcia, V.C., 2014. A systematic review on cloud computing. J. Supercomput. 68, 1321–1346.
Elghoneimy, E., Bouhali, O., Alnuweiri, H. Resource allocation and scheduling in cloud computing. In: Proceedings of the IEEE International Conference on Computing, Networking and Communications (ICNC), pp. 309–314.
Ergu, D., Kou, G., Peng, Y., Shi, Y., Shi, Y., 2013. The analytic hierarchy process: task scheduling and resource allocation in cloud computing environment. J. Supercomput. 64, 835–848.
Faragardi, H.R., Shojaee, R., Tabani, H., Rajabi, A., 2013. An analytical model to evaluate reliability of cloud computing systems in the presence of QoS requirements. In: Proceedings of the IEEE/ACIS 12th International Conference on Computer and Information Science (ICIS), IEEE, pp. 315–321.
Gabi, D., Ismail, A.S., Zainal, A., 2015. Systematic review on existing load balancing techniques in cloud computing. Int. J. Comput. Appl., 125.
Ghanbari, S., Othman, M., 2012. A priority based job scheduling algorithm in cloud computing. Procedia Eng. 50, 778–785.
Ghribi, C., Hadji, M., Zeghlache, D., 2013. Energy efficient vm scheduling for cloud data centers: Exact allocation and migration algorithms. In: Proceedings of the 13th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), IEEE, pp. 671–678.
Gupta, P., Ghrera, S.P., 2015. Load and Fault Aware Honey Bee Scheduling Algorithm for Cloud Infrastructure. In: Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014. Springer, pp. 135–143.
Gupta, P., Goyal, M.K., Kumar, P., Aggarwal, A., 2013. Trust and reliability based scheduling algorithm for cloud IaaS. In: Proceedings of the Third International Conference on Trends in Information, Telecommunication and Computing. Springer, pp. 603–607.
Hassan, S., Abbas Kamboh, A., Azam, F., 2014. Analysis of Cloud Computing Performance, Scalability, Availability, & Security. In: Proceedings of the International Conference on Information Science and Applications (ICISA), IEEE, pp. 1–5.
He, Z.T., Zhang, X.Q., Zhang, H.X., Xu, Z.W., 2013. Study on new task scheduling strategy in cloud computing environment based on the simulator CloudSim. Adv. Mater. Res. 651, 829–834.
Hosseinimotlagh, S., Khunjush, F., Samadzadeh, R., 2015. SEATS: smart energy-aware task scheduling in real-time cloud computing. J. Supercomput. 71, 45–66.
Huang, L., Chen, H.-S., Hu, T.-T., 2013. Survey on resource allocation policy and job scheduling algorithms of cloud computing1. J. Softw. 8, 480–487.
Huang, M.G., Ou, Z.Q., 2014. Review of task scheduling algorithm research in cloud computing. Adv. Mater. Res., 3236–3239.
Hung, P.P., Nguyen, M.V., Aazam, M., Huh, E.N., 2014. Task scheduling for optimizing

recovery time in cloud computing. In: Proceedings of the International Conference on Computing, Management and Telecommunications (ComManTel), IEEE, pp. 188–193.
Javanmardi, S., Shojafar, M., Amendola, D., Cordeschi, N., Liu, H., Abraham, A., 2014. Hybrid Job Scheduling Algorithm for Cloud Computing Environment. In: Proceedings of the Fifth International Conference on Innovations in Bio-Inspired Computing and Applications IBICA. Springer, pp. 43–52.
Jena, R., 2015. Multi objective task scheduling in cloud environment using nested PSO framework. Procedia Comput. Sci. 57, 1219–1227.
Jennings, B., Stadler, R., 2014. Resource management in clouds: survey and research challenges. J. Netw. Syst. Manag., 1–53.
Jhawar, R., Piuri, V., Santambrogio, M., 2013. Fault tolerance management in cloud computing: a system-level perspective. Syst. J. IEEE 7, 288–297.
Jiankang, D., Hongbo, W., Yangyang, L., Shiduan, C., 2014. Virtual machine scheduling for improving energy efciency in IaaS cloud. Communications 11, 1–12.
Jiao, H., Zhang, J., Li, J., Shi, J., Li, J., 2015. Immune optimization of task scheduling on multidimensional QoS constraints. Clust. Comput. 18, 909–918.
Kalra, M., Singh, S., 2015. A review of metaheuristic scheduling techniques in cloud computing. Egypt. Inf. J. 16, 275–295.
Kansal, N.J., Chana, I., 2012. Existing load balancing techniques in cloud computing: a systematic review. J. Inf. Syst. Commun. 3, 87.
Kapur, R., 2015. A cost effective approach for resource scheduling in cloud computing. In: Proceedings of the Computer, Communication and Control (IC4), International Conference on IEEE, pp. 1–6.
Katyal, M., Mishra, A., 2014. A Comparative Study of Load Balancing Algorithms in Cloud Computing Environment arXiv preprint arXiv:1403.6918.
Kaur, T., Chana, I., 2015. Energy efficiency techniques in cloud computing: a survey and taxonomy. ACM Comput. Surv. (CSUR) 48, 22.
Kim, S., Kim, Y., Song, N., Kim, C., 2010. Adaptable scheduling schemes for scientific applications on science cloud. In: Proceedings of the Cluster Computing Workshops and Posters (CLUSTER WORKSHOPS), IEEE International Conference on IEEE, pp. 1–3.
Kumar, D., Raza, Z., 2015. A PSO Based VM Resource Scheduling Model for Cloud Computing. In: Proceedings of the Computational Intelligence & Communication Technology (CICT), IEEE International Conference on, pp. 213–219.
Kumar, N., Chilamkurti, N., Zeadally, S., Jeong, Y.-S., 2014. Achieving Quality of Service (QoS) using resource allocation and adaptive scheduling in cloud computing with grid support. Comput. J. 57, 281–290.
Lakra, A.V., Yadav, D.K., 2015. Multi-objective tasks scheduling algorithm for cloud computing throughput optimization. Procedia Comput. Sci. 48, 107–113.
Le, G., Xu, K., Song, J., 2013. Dynamic resource provisioning and scheduling with deadline constraint in elastic cloud. In: Proceedings of the Service Sciences (ICSS), International Conference on IEEE, pp. 113–117.
Li, F.F., Xie, D.Q., Gao, Y., Xie, G.W., Guo, Q.Y., 2013. Research on Multi-QoS and trusted task scheduling in cloud computing environment. Appl. Mech. Mater. 263, 1892–1896.
Li, H., Li, H., 2011. A research of resource scheduling strategy with SLA restriction for cloud computing based on Pareto optimality $M \times N$ production model. Web Inf. Syst. Min., Springer.
Li, K., Xu, G., Zhao, G., Dong, Y., Wang, D., 2011. Cloud task scheduling based on load balancing ant colony optimization. In: Proceedings of the Chinagrid Conference (ChinaGrid), Sixth Annual, IEEE, pp. 3–9.
Li, Q., Guo, Y., 2010. Optimization of resource scheduling in cloud computing. In: Proceedings of the 12th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, IEEE, pp. 315–320.
Li, X., Li, Y., Liu, T., Qiu, J., Wang, F., 2009. The method and tool of cost analysis for cloud computing. In: Proceedings of the Cloud Computing, CLOUD'09, IEEE International Conference on, IEEE, pp. 93–100.
Lin, W., Liang, C., Wang, J.Z., Buyya, R., 2014. Bandwidth-aware divisible task scheduling for cloud computing. Softw.: Pract. Exp. 44, 163–174.
Liu, C.-Y., Zou, C.-M., Wu, P., 2014a. A task scheduling algorithm based on genetic algorithm and ant colony optimization in cloud computing. In: Proceedings of the Distributed Computing and Applications to Business, Engineering and Science (DCABES), 13th International Symposium on, IEEE, pp. 68–72.
Liu, J., Luo, X.G., Zhang, X.M., Zhang, F., 2013. Job scheduling algorithm for cloud computing based on particle swarm optimization. Adv. Mater. Res. 662, 957–960.
Liu, X.-F., Zhan, Z.-H., Du, K.-J., Chen, W.-N., 2014b. Energy aware virtual machine placement scheduling in cloud computing based on ant colony optimization approach. In: Proceedings of the 2014 Conference on Genetic and Evolutionary Computation, ACM, pp. 41-48.
Luo, L., Wu, W., Di, D., Zhang, F., Yan, Y., Mao, Y., 2012. A resource scheduling algorithm of cloud computing based on energy efficient optimization methods. In: Proceedings of the Green Computing Conference (IGCC), 2012 International, IEEE, pp. 1–6.
Ma, J., Li, W., Fu, T., Yan, L., Hu, G., 2016. A novel dynamic task scheduling algorithm based on improved genetic algorithm in cloud computing. Wirel. Commun. Netw. Appl., Springer.
Ma, T., Chu, Y., Zhao, L., Ankhbayar, O., 2014. Resource allocation and scheduling in cloud computing: policy and algorithm. IETE Tech. Rev. 31, 4–16.
Madni, S.H.H., Latiff, M.S.A., Coulibaly, Y., Abdulhamid, S.I.M., 2016. An appraisal of meta-heuristic resource allocation techniques for IaaS cloud. Indian J. Sci. Technol. 9, 1–14.
Malik, S., Huet, F., Caromel, D., 2012. Reliability aware scheduling in cloud computing. In: Proceedings of the Internet Technology and Secured Transactions, International Conference for, IEEE, pp. 194–200.
Manvi, S.S., Krishna Shyam, G., 2014. Resource management for Infrastructure as a Service (IaaS) in cloud computing: a survey. J. Netw. Comput. Appl. 41, 424–440.
Mastelic, T., Oleksiak, A., Claussen, H., Brandic, I., Pierson, J.-M., Vasilakos, A.V., 2014. Cloud computing: survey on energy efficiency. ACM Comput. Surv. (CSUR) 47, 33.
Mell, P., Grance, T., 2011. The NIST definition of cloud computing [Recommendations of the National Institute of Standards and Technology-Special Publication 800-145]. NIST. Recuperado de, Washington D.C. ⟨http://csrc. nist. gov/publications/nistpubs/800-145/SP800-145.pdf⟩.
Mezmaz, M., Melab, N., Kessaci, Y., Lee, Y.C., Talbi, E.-G., Zomaya, A.Y., Tuyttens, D., 2011. A parallel bi-objective hybrid metaheuristic for energy-aware scheduling for cloud computing systems. J. Parallel Distrib. Comput. 71, 1497–1508.
Mhedheb, Y., Jrad, F., Tao, J., Zhao, J., Kołodziej, J., Streit, A., 2013. Load and thermal-aware VM scheduling on the cloud. Algorithms Archit. Parallel Process., Springer.
Moschakis, I.A., Karatza, H.D., 2012. Evaluation of gang scheduling performance and cost in a cloud computing system. J. Supercomput. 59, 975–992.
Mustafa, S., Nazir, B., Hayat, A., Madani, S.A., 2015. Resource management in cloud computing: Taxonomy, prospects, and challenges. Comput. Electr. Eng.
Nabi, M., Toeroe, M., Khendek, F., 2016. Availability in the cloud: state of the art. J. Netw. Comput. Appl. 60, 54–67.
Netjinda, N., Sirinaovakul, B., Achalakul, T., 2014. Cost optimal scheduling in IaaS for dependent workload with particle swarm optimization. J. Supercomput. 68, 1579–1603.
Nivodhini, M., Kousalya, K., Malliga, S., 2013. Algorithms to improve scheduling techniques in IaaS cloud. In: Proceedings of the Information Communication and Embedded Systems (ICICES), 2013 International Conference on, IEEE, pp. 246–250.
Nuaimi, K.A., Mohamed, N., Nuaimi, M.A., Al-Jaroodi, J., 2012. A survey of load balancing in cloud computing: challenges and algorithms. In: Proceedings of the Network Cloud Computing and Applications (NCCA), 2012 Second Symposium on, IEEE, pp. 137–142.
Pan, B.L., Wang, Y.P., Li, H.X., Qian, J., 2014. Task scheduling and resource allocation of cloud computing based on QoS. Adv. Mater. Res. 915, 1382–1385.
Pan, J.-S., Wang, H., Zhao, H., Tang, L., 2015. Interaction artificial bee colony based load balance method in cloud computing. Genet. Evolut. Comput., Springer
Pandi, K.M., Somasundaram, K., 2016. Energy efficient in virtual infrastructure and green cloud computing: a review. Indian J. Sci. Technol. 9.
Patel, P., Ranabahu, A.H., Sheth, A.P. 2009. Service Level Agreement in Cloud Computing.
Puschner, P., Koza, C., 1989. Calculating the maximum execution time of real-time programs. Real-Time Syst. 1, 159–176.
Puthal, D., Sahoo, B., Mishra, Swain, S., 2015. Cloud computing features, issues, and challenges: a big picture. In: Proceedings of the Computational Intelligence and Networks (CINE), 2015 International Conference on, IEEE, pp. 116–123.
Qi, H., Gani, A., 2012. Research on mobile cloud computing: review, trend and perspectives. In: Proceedings of the Digital Information and Communication Technology and it's Applications (DICTAP), 2012 second International Conference on, IEEE, pp. 195–202.
Qian, L., Luo, Z., Du, Y., Guo, L., 2009. Cloud Computing: An Overview. Cloud Computing.
Raghavan, S., Marimuthu, C., Sarwesh, P., Chandrasekaran, K., 2015. Bat algorithm for scheduling workflow applications in cloud. In: Proceedings of the Electronic Design, Computer Networks & Automated Verification (EDCAV), 2015 International Conference on, IEEE, pp. 139–144.
Raju, I.R.K., Varma, P.S., Sundari, M.R., Moses, G.J., 2016. Deadline aware two stage scheduling algorithm in cloud computing. Indian J. Sci. Technol. 9.
Ramezani, F., Lu, J., Hussain, F., 2013. Task scheduling optimization in cloud computing applying multi-objective particle swarm optimization. Serv.-Oriented Comput., Springer.
Rimal, B.P., Choi, E., Lumb, I., 2010. A taxonomy, survey, and issues of cloud computing ecosystems. Cloud Comput., Springer.
Sandholm, T., Lee, D., 2014. Notes on cloud computing principles. J. Cloud Comput. 3, 1–10.
Sharma, Y., Javadi, B., Si, W., 2015. On the reliability and energy efficiency in cloud computing. Parallel Distrib. Comput. 27, 111.
Shen, S., Deng, K., Iosup, A., Epema, D., 2013. Scheduling Jobs in the Cloud Using On-demand and Reserved Instances, Euro-Par 2013 Parallel Processing. Springer.
Shojafar, M., Javanmardi, S., Abolfazli, S., Cordeschi, N., 2015. FUGE: a joint meta-heuristic approach to cloud job scheduling algorithm using fuzzy theory and a genetic method. Cluster Comput., 1–16.
Shuja, J., Bilal, K., Madani, S.A., Khan, S.U., 2014. Data center energy efficient resource scheduling. Cluster Comput. 17, 1265–1277.
Sindhu, S., Mukherjee, S., 2011. Efficient task scheduling algorithms for cloud computing environment. High Perform. Archit. Grid Comput., Springer
Sindhu, S., Mukherjee, S., 2013. A genetic algorithm based scheduler for cloud environment. In: Proceedings of the Computer and Communication Technology (ICCCT), 2013 4th International Conference on, IEEE, pp. 23–27.
Singh, S., 2015. Green computing strategies & challenges. In: Proceedings of the Green Computing and Internet of Things (ICGCIoT), 2015 International Conference on, IEEE, pp. 758–760.
Singh, S., Chana, I., 2015. QRSF: QoS-aware resource scheduling framework in cloud computing. J. Supercomput. 71, 241–292.
Sun, A., Ji, T., Yue, Q., Yang, S., 2013. Virtual machine scheduling, motion and disaster recovery model for IaaS cloud computing platform. In: Proceedings of the

Conference Anthology, IEEE, IEEE, pp. 1–5.
Thomas, A., Krishnalal, G., Raj, V.J., 2015. Credit based scheduling algorithm in cloud computing environment. Procedia Comput. Sci. 46, 913–920.
Tiwari, A., Nagaraju, A., Mahrishi, M., 2013. An optimized scheduling algorithm for cloud broker using adaptive cost model. In: Proceedings of the Advance Computing Conference (IACC), 2013 IEEE 3rd International, IEEE, pp. 28–33.
Tsai, C.-W., Rodrigues, J.J., 2014. Metaheuristic scheduling for cloud: a survey. Syst. J. IEEE 8, 279–291.
Tsai, J.-T., Fang, J.-C., Chou, J.-H., 2013. Optimized task scheduling and resource allocation on cloud computing environment using improved differential evolution algorithm. Comput. Oper. Res. 40, 3045–3055.
Tziritas, N., Xu, C.-Z., Loukopoulos, T., Khan, S.U., Yu, Z., 2013. Application-aware workload consolidation to minimize both energy consumption and network load in cloud environments. In: Proceedings of the Parallel Processing (ICPP), 2013 42nd International Conference on, IEEE, pp. 449–457.
Van Do, T., Rotter, C., 2012. Comparison of scheduling schemes for on-demand IaaS requests. J. Syst. Softw. 85, 1400–1408.
Venters, W., Whitley, E.A., 2012. A critical review of cloud computing: researching desires and realities. J. Inf. Technol. 27, 179–197.
Wang, G., Yu, H.C., 2013. Task scheduling algorithm based on improved min–min algorithm in cloud computing environment. Appl. Mech. Mater. 303, 2429–2432.
Wang, H., Wang, F., Liu, J., Wang, D., Groen, J., 2015. Enabling customer-provided resources for cloud computing: potentials, challenges, and implementation. Parallel Distrib. Syst. IEEE Trans. 26, 1874–1886.
Wang, P., 2014. An improved scheduling algorithm for cloud storage. Adv. Mater. Res., 2200–2203.
Wang, T., Liu, Z., Chen, Y., Xu, Y., Dai, X., 2014. Load balancing task scheduling based on genetic algorithm in cloud computing. In: Proceedings of the Dependable, Autonomic and Secure Computing (DASC), 2014 IEEE 12th International Conference on, IEEE, pp. 146–152.
Wang, W.-J., Chang, Y.-S., Lo, W.-T., Lee, Y.-K., 2013. Adaptive scheduling for parallel tasks with QoS satisfaction for hybrid cloud environments. J. Supercomput. 66, 783–811.
Wen, X., Huang, M., Shi, J., 2012. Study on resources scheduling based on ACO algorithm and PSO algorithm in cloud computing. In: Proceedings of the Distributed Computing and Applications to Business, Engineering & Science (DCABES), 2012 11th International Symposium on, IEEE, pp. 219–222.
Wu, X., Deng, M., Zhang, R., Zeng, B., Zhou, S., 2013. A task scheduling algorithm based on QoS-driven in Cloud Computing. Procedia Comput. Sci. 17, 1162–1169.
Xiong, K., Perros, H., 2009. Service performance and analysis in cloud computing. In: Proceedings of the Services-I, 2009 World Conference on, IEEE, pp. 693–700.
Xiong, Y.Y., Wu, Y.Y., 2014. Cloud computing resource schedule strategy based on PSO algorithm. Appl. Mech. Mater. 513, 1332–1336.
Xue, S., Li, M., Xu, X., Chen, J., Xue, S., 2014. An ACO-LB algorithm for task scheduling in the cloud environment. J. Softw. 9, 466–473.
Yang, H., 2014. Improved ant colony algorithm based on PSO and its application on cloud computing resource scheduling. Adv. Mater. Res., 2192–2195.
Yang, H., Tate, M., 2012. A descriptive literature review and classification of cloud computing research. Commun. Assoc. Inf. Syst. 31, 35–60.
Yi, P., Ding, H., Ramamurthy, B., 2013. Budget-minimized resource allocation and task scheduling in distributed grid/clouds. In: Proceedings of the Computer Communications and Networks (ICCCN), 2013 22nd International Conference on, IEEE, pp. 1–8.
Younge, A.J., Von Laszewski, G., Wang, L., Lopez-Alarcon, S., Carithers, W., 2010. Efficient resource management for cloud computing environments. In: Proceedings of the Green Computing Conference, 2010 International, IEEE, pp. 357–364.
Zhang, H., Li, P., Zhou, Z., Yu, X., 2013a. A PSO-Based Hierarchical Resource Scheduling Strategy on Cloud Computing. Trust. Comput. Serv., Springer.
Zhang, J., Huang, H., Wang, X., 2016. Resource provision algorithms in cloud computing: a survey. J. Netw. Comput. Appl.
Zhang, Q., Cheng, L., Boutaba, R., 2010. Cloud computing: state-of-the-art and research challenges. J. Internet Serv. Appl. 1, 7–18.
Zhang, W., He, H., Chen, G., Sun, J., 2013b. Multiple virtual machines resource scheduling for cloud computing. Appl. Math. 7, 2089–2096.
Zhang, Y., Su, Y.M., 2014. Research on resource scheduling algorithm in cloud computing data center. Adv. Mater. Res., 2050–2053 (Trans Tech Publ).
Zhang, Z., Xiao, L., Chen, X., Peng, J., 2013c. A scheduling method for multiple virtual machines migration in cloud. Netw. Parallel Comput., Springer.
Zhao, L.F., Zhou, S.H., Chang, W.B., 2014. Task scheduling in cloud computing with improved firefly algorithm. Appl. Mech. Mater., 3189–3193 (Trans Tech Publ).
Zhong-Wen, G., Kai, Z., 2012. The research on cloud computing resource scheduling method based on time-cost-trust model. In: Proceedings of the Computer Science and Network Technology (ICCSNT), 2012 2nd International Conference on, IEEE, pp. 939–942.
Zhu, Y., Liang, H., 2013. Research for the virtual machine-oriented cloud resource scheduling algorithm. In: Proceedings of the Information Management, Innovation Management and Industrial Engineering (ICIII), 2013 6th International Conference on, IEEE, pp. 133–136.
Zuo, L., Shu, L., Dong, S., Zhu, C., Hara, T., 2015. A multi-objective optimization scheduling method based on the ant colony algorithm in cloud computing. Access IEEE 3, 2687–2699.