

# Cloud Resource Management: Comparative Analysis and Research Issues

Harvinder Singh<sup>1</sup>, Anshu Bhasin<sup>1</sup>, Parag Ravikant Kaveri<sup>2</sup>, Vinay Chavan<sup>3</sup>

**Abstract**— Cloud resource management is momentous for efficient resource allocation and scheduling that requires for fulfilling customers' expectations. But, it is difficult to predict an appropriate matching in a heterogeneous and dynamic cloud environment that leads to performance degradation and SLA violation. Thus, resource management is a challenging task that may be compromised because of the inappropriate allocation of the required resource. This paper presents a systematic review and analytical comparisons of existing surveys, research work exists on SLA, resource allocation and resource scheduling in cloud computing. Further, discussion on open research issues, current status and future research directions in the field of cloud resource management.

**Index Terms**— Resource allocation, Resource scheduling, QoS, SLA, Heterogeneity, Scalability, VM management, Resource utilization, Energy consumption, Security, Monitoring.

## 1 INTRODUCTION

Cloud computing providing virtual resources to customers under the pay-per-usage model managed by service providers. It includes infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS). Service providers' priorities are to fulfill customers' expectations for competitive costs. Resource management needs the optimal usage of virtual resources to respond on a large scale. It is challenging to allocate a suitable resource on heterogeneous and dynamic task requirements that impacts performance and SLA. The quality of service (QoS) can be fulfilled by considering availability, scalability, utilization, cost, time, energy consumption and so forth. Cloud has the capacity to provide service according to the behavioral of applications. Organizations such as banking, health care system, educational platform, and e-commerce are using cloud services for storing and retrieving data [Pietri et al. 2016]. It has eliminated the need for purchasing physical resources [Jamshidi et al. 2013]. Cloud service has become an integral part of our daily lives that fulfill information technology (IT) needs by cost-effectiveness and usability [Buyya et al. 2009]. To realize this, there is a challenge to ensure that guarantees QoS requirements and SLA by managing resource heterogeneity, dynamism, complexity, and uncertainty. It can distrust consumer and provider relations where pricing policies are according to QoS parameters [Mustafa et al. 2015].

Cloud computing features can efficiently manage varied application requirements need to be explored. Existing resource management techniques are unable to such a customizing environment to achieve important QoS parameters and avoid SLA violations. This survey has been conducted to provide a hands-on-information in the field of cloud resource management. In a cloud environment unpredictability and uncertainty, the problem causes inappropriate matching. It needs to consider the following:

- SLA rules: Consumers' are paying and expecting cloud service at a reasonable cost and time. But, it is challenging to provide expected performance and SLA violation.
- Availability: Resource availability plays an important role to process users' instant demand. But, dynamic reallocation leads to network congestion and energy consumption.
- Dynamic environment: The cloud environment is dynamic and unpredictable. It is difficult to manage dynamism and uncertainty.
- Heterogeneity: Cloud computing comprising heterogeneous resources. It can fulfill varied application demands. But, traditional techniques are unable to manage heterogeneity.
- Geographical distance: Cloud data centers are located in different geographical regions that require the proper distribution of isolated resources for higher utilization. But, existing resource management techniques are unable to manage diversified network resources.

Moreover, existing techniques need an extension to customize emerging platforms, such as edge computing, containers and hybrid cloud component [Gupta et al. 2017]. The main issue in this context to enhance availability, utilization, and elasticity that helps to avoid SLA violation [Zhan et al. 2015; Singh et al. 2015]. Therefore, a standardized technique should have existed for efficient cloud resource management.

### 1.1 Significance of Cloud Resource Management (CRM)

The motive of cloud resource management (CRM) is to provide applications based services by efficiently managing runtime resources. The traditional cloud would be replaced by hybrid components. There is a need for managing

- Harvinder Singh is currently pursuing Ph.D in computer application in IKGujral Punjab Technical University Kapurthala, India E-mail: rs.hschaahal@ptu.ac.in
- Dr. Anshu Bhasin is currently Faculty in computer science and Engg at IKGujral Punjab Technical University Kapurthala, India E-mail: dr.anshubhasin@ptu.ac.in
- Dr. Parag Ravikant Kaveri is currently Faculty in computer studies at Symbiosis Institute of Computer Studies and Research (SICSR), Symbiosis International (Deemed University), Pune, India E-mail: parag.kaveri@sicsr.ac.in
- Dr. Vinay Chavan is currently Faculty in computer science at Seth Kesarimal Porwal College, Kamptee, Nagpur, India E-mail: drovinaychavan@yahoo.com

heterogeneity, dynamism, and uncertainty while scheduling cloud resources. Cloud resource scheduling is an NP-hard problem impacts performance. To overcome such a problem, the cloud needs a novel technique that can manage according to its characteristics [Singh H. et al., 2017; Singh H. et al., 2019; Singh H. et al., 2019]. This comprehensive survey will help researchers for developing a standard framework to enhance the quality of cloud services, cost-effectiveness and usability.

## 1.2 Motivation for CRM

The objective of this study is a discussion on resource management techniques: resource provisioning, allocation, scheduling, and monitoring from the perspective of QoS parameters and SLA. It also discusses the challenges on resource provisioning, resource allocation, resource mapping, resource scheduling, application scheduling, VM management, workload management, QoS parameters, SLA, Big data, latency, energy consumption, utilization, scalability, security, and monitoring in cloud resource management.

Cloud services are based on the pay-per-usage model. To retain quality service while responding on large scale optimization of allocated resources required. Growing application domains, heterogeneity, complexity, and unpredictability are the obstacle in cloud resource management.

## 1.3 Comparison between Existing Surveys and Ours

Cloud services demand is growing tremendously. To efficiently handle deliveries by protecting the unnatural condition, it needs a better resource management solution. This section has been conducted a discussion and comparative analysis (table 1) on existing surveys in cloud resource management.

Vinothina et al. [2012] survey on resource allocation strategies focuses on Service level agreement (SLA) to identify utilization of resources. It is an in-depth discussion on optimal allocation to strengthen cloud services. Tinghuai et al. [2014] conducted a review of existing resource allocation and scheduling techniques to attain SLA by providing cost-effective services. The scheduling strategies are considered locality-aware, reliability-aware and energy-aware for resource management. Manvi et al. [2014] review resource provisioning, allocation and scheduling techniques for efficient management of infrastructure resources. The QoS requirements and SLA violation concerns were also discussed to provide cost-effective services. Singh et al. [2015] presented a survey on QoS aware resource management techniques in cloud computing. The importance of optimization-aware self-management techniques also discussed. It shows the impact of QoS parameters on SLA, it also gives suggestions further development of standardizing resource management technique. Mustafa et al. [2015] present a review and taxonomy of resource management techniques. It discusses challenges in energy-aware, workload-aware and network resources for managing SLA and profit in hybrid and mobile cloud. Alkhank et al. [2015] detailed taxonomy on workflow scheduling for efficient resource management. It discusses challenges in resource allocation and scheduling to provide cost-aware services while managing data-intensive applications. Zhan et al. [2015] presented, taxonomy on resource scheduling schemes to provide application-aware

cost-effective cloud services. It discusses an evolution aware method to proper utilize IaaS, PaaS and SaaS resources while managing energy and profit. Madni et al. [2016] review resource scheduling techniques for efficient resource allocation and scheduling. It discusses different resource scheduling challenges and impacts on QoS parameters and SLA. The paper directs suggestions for cost-aware VM allocation and management. Pietri et al. [2016] review on VM configuration and placement strategies for efficient VM mapping. It highlights the issues in VM reallocation for varied application requirements, reduction of energy SLA violation. It also discusses various application metrics and platforms for appropriate VM placement to manage resource utilization and profit. Zhang et al. [2016] present a review on different categories of resource provisioning algorithms in cloud computing. It identified some issues of VMs migration, availability, and scalability that can lead to SLA violation. Singh et al. [2016] presented a survey on cloud resource management (resource provisioning and scheduling), the importance of algorithm accuracy for the selection of specific workload execution. It highlights issues for efficient resource distribution, impact on QoS and SLA. Further, Singh et al. [2016] present a broad survey on resource scheduling algorithms (RSA) in cloud computing. It discusses various categories of RSA for efficient resource management. The paper also highlights the challenging issues in resource provisioning and scheduling. Mezni et al. [2018] review on various uncertainty aware approaches in cloud computing. It discusses performance affecting uncertainty aware factor, and future research directions for cloud service life-cycle management. Kumar M. et al. [2019] presented a review of different resource scheduling techniques for better resource management. It discusses the benefits and limitations of heuristic, metaheuristic, and hybrid scheduling techniques; it also discusses various simulation tools.

- This research survey is an initiative that reviews on existing surveys. It incorporated all the characteristics of sustainable cloud computing. This research work presents an analytical discussion on existing surveys that result in open research issues.
- A systematic review of resource management techniques in cloud computing.
- Discussion on challenging and current issues for identifying future scope in the related domain.
- It organizes resource management techniques based on strategies followed by a method.
- Open challenging issues still impacting on performance and service quality has been discussed.
- To overcome the issues, it presented suggestions for researchers and academia.

Table 1 shows a comparative analysis of existing surveys where SLA, QoS, resource allocation, and resource scheduling have been considered in the majority of surveys. The research work needs other important factors: virtualization, energy consumption, utilization, workload management, monitoring, and cost. Despite limited surveys on open issues and further

Table 1. Comparison between Existing Surveys and Present Survey

Criteria	Our survey	Vinothina et al. [2012]	Tinghuai et al. [2014]	Manvi et al. [2014]	Singh et al. [2015]	Mustafa et al. [2015]	Alkhank et al. [2015]	Zhan et al. [2015]	Madni et al. [2016]	Pietri et al. [2016]	Zhang et al. [2016]	Singh et al. [2016]	Singh et al. 2016	Mezni et al. [2018]	Kumar M. et al. [2019]
Survey	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Virtualization	√				√	√	√	√	√	√	√	√			
Energy	√		√		√	√		√	√	√	√				
Utilization	√	√					√		√	√		√			√
Workload management					√	√		√		√			√	√	
Resource allocation	√	√	√	√	√					√	√	√	√	√	√
Resource scheduling	√		√	√	√	√	√	√	√	√		√	√	√	√
Monitoring	√				√							√			
SLA	√	√	√	√	√	√	√	√	√	√	√	√	√		√
Cost	√		√	√			√	√	√	√	√	√			
Qi's	√		√	√	√	√	√		√	√		√		√	√
Review+ further research guidelines	√				√			√	√			√		√	

research directions, other surveys presented the only review of existing research work.

## 1.4 Paper Structure

The paper is organized as follows: section 2 presents cloud resource management mechanism by following subsections: SLA- QoS: parameters-validation check, services: demand-monitoring, resources: VMs-availability; resource allocation policies- heuristic and metaheuristic; resource scheduling policies- heuristic and metaheuristic; section 3 discussion on open research challenges and future research directions; and section 4 conclusions and summary of the work.

## 2 CLOUD RESOURCE MANAGEMENT MECHANISM

This research work has adopted a systematic way to present existing works on cloud resource management. Cloud resources are providing services to millions of users under different models based on customers' budgets and requirements. Therefore, the resource management (RM) technique should monitor dynamism and scalability while scheduling [Singh et al., 2017; Manvi et al., 2014]. It is challenging to meet customer reliability and provider expectations without compromising on performance and SLA violation. It has been identified that service quality, economic approach and resource usability are important factors that need to be considered in the standard resource management technique. Figure 1: presents cloud resource management mechanism: (i) SLA (ii) resource allocation (iii) resource scheduling.

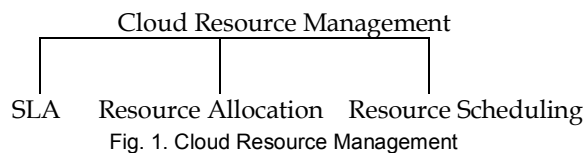


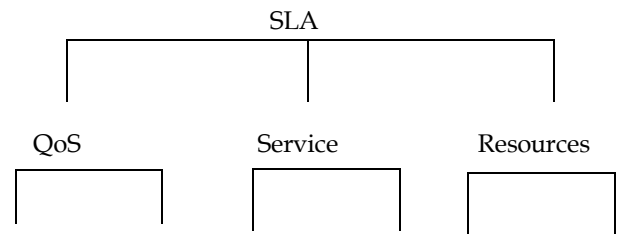
Fig. 1. Cloud Resource Management

### 2.1 SLA

Cloud computing paradigm based on the concept of Service-Oriented Architecture (SOA) where a legal agreement exists between the provider and consumer [Jula et al. 2014]. The provider needs to ensure guarantees of SLA by meeting users' requirements.

In this context, Buyya et al. [2011] presented an SLA-aware resource allocation model for providing business-oriented services. It was based on the allocation of services to meet application demand by managing SLA. It also highlights the complexity of application workload, and execution of specific service in terms of QoS by managing the budget. Goudarzi et al. [2012] have considered SLA constraints using a heuristic method for performance optimization. It manages VM allocation and placement to meet users' requests while managing operational and migration costs. Mustafa et al. [2015] discussed SLA, violation and penalty measures. It highlights that such techniques are unable to provide service according to expected QoS parameters. It also leads to performance degradation and SLA violation. Suprakash et al. [2019] presented SLA aware model to provide quality of service. This work focuses on 100% resource utilization by monitoring that ensures effective management of SLA. A

catalog based model adopted for getting the current state of resource which help to rearrange underutilized resources. Figure 1 shows different aspects of SLA which results in guaranteeing SLA and QoS parameters.



Parameters Validation check Demand Monitoring VMs Availability

Fig. 2. Classification of SLA

#### 2.1.1 QoS: Parameters- Validation Check

Plenty of research work has done on SLA and its violation with respect to QoS parameters. Beloglazov et al. [2012] presented "energy-aware resource allocation" (EA-RA), a negotiation based process. It considers QoS parameters and greenhouse gas emissions through optimization to reduce cost and energy consumption. Feng et al. [2013] presented a QoS based load aware method to improve utilization through Cloud Virtual technology virtualizes cloud data center resources that run instances of resources or applications. A single physical machine runs multiple application services. But, particular VM instance allocation to a specific task is complicated due to heterogeneity and dynamic environment [Chavan et al. 2015].

In this context, Liu et al. [2014] presented a high availability (HA) method for managing cloud infrastructure. The research work discussed techniques developed for increasing the reliability and availability in the cloud. It highlights challenges and developed benchmarking based method that estimates available for the deployment of resources to a customer. Chavan et al. [2014] presented a higher resource availability method using clustering virtual machines in data centers. It optimizes shared resources to get better results for scalability and availability. The work performed auto-VM migration to manage availability and balancing that improves utilization. Nabi et al. [2016] presented a survey on current availability solutions in the cloud designed. This taxonomy discussed service Availability Forum, concepts, and mechanisms for baseline evaluation. It categorized proposed solutions, differences and similarities between various solutions in terms of availability and their impact on performance and SLA. Pietri et al. [2016] survey on mapping VMs onto PMs and categorize techniques based on mapping properties impact availability. This research work discussed scheduling actions, triggers, optimization goals, metrics, applications, and platforms used for evaluation in different techniques. Mishra et al. [2018] presented an "Energy-efficient VM-placement" (EE-VMsP) algorithm for optimal mapping. It performed VMs configuration and placement to reduce energy consumption and makespan by appropriate handling of heterogeneous and dynamic tasks.

It was concluded that resource availability can help while distributing VMs. Further, matching techniques manage



adaptability and throughput utilization of resources. Therefore, the RM technique should be able to release unnecessary active resources at runtime to organize availability, auto-scaling and clustering. To control elastic scaling and energy consumption, it efficiently distributes physical resources. Ding et al. [2014] presented a QoS-aware resource matching framework by identifying resource requirements using the negotiation process. The method adopts multi-attribute matching for resource provisioning.

Singh et al. [2015] presented a QoS-aware resource provisioning framework, K-means based clustering mechanism was used to group similar workloads. It identifies and classifies workload before resource provisioning to reduce cost, time and energy. Serrano et al. [2016] presented SLAaaS (SLA-aware Service) model to integrate QoS and SLA. It addresses SLA violation problems and formulates the CSLA language as a proposed solution. Singh et al. [2017] presented STAR, an optimization scheme that improves the probability to meet QoS parameters, performance and minimize SLA violation.

### 2.1.2 Services: Demand- Monitoring

The cloud environment provides cost-effective services under the pay-per-usage model. Thus providers need to ensure service should be according to demand and monitoring of allocated resources. Monitoring is an important aspect of cloud resource management. Cloud is facing challenges to manage services on-demand, scalability, elasticity, and uncertainty [Zhao et al. 2014; Syed et al. 2017].

Ward et al. [2014] survey of contemporary monitoring tools and examine the designs and challenges in cloud monitoring. It concluded that the cloud needs distributed monitoring support tools. These tools should have extensive range of resource monitoring to manage scalability, elasticity, performance on demand of consumers. Rodrigues et al. [2016] highlighted the importance of monitoring and performance affecting factors: scalability, elasticity, and migration. It shows that monitoring should be performed at different levels to exploit cloud resources. Prasad et al. [2018] conducted a survey on monitoring that considers the importance of resource availability at the IaaS level. It shows that availability can improve service quality and utilization. The work highlights availability and monitoring to help in service quality and performance by improving utilization. Wang et al. [2018] proposed a "self-adaptive monitoring approach" (SA-MA) to estimate the running state of the system, anomaly degree, and predict faults detection using principal component analysis. It efficiently detects abnormal state and lowering the monitoring overheads.

The monitoring of allocated resources can improve availability and usability. Traditional RM techniques are unable to provide such an environment for runtime workload identification that can monitor resource state. The survey shows that limited work has been done on monitoring for managing elastic scaling and expected service quality. Further research work can be done on service-aware, application-aware and resource-aware monitoring for efficient cloud services.

### 2.1.3 RESOURCES: VMS- AVAILABILITY

Cloud Virtual technology virtualizes cloud data center resources that run instances of resources or applications. A single physical machine runs multiple application services. But, particular VM instance allocation to a specific task is complicated due to heterogeneity and dynamic environment [Chavan et al. 2015].

In this context, Liu et al. [2014] presented a high availability (HA) method for managing cloud infrastructure. The research work discussed techniques developed for increasing the reliability and availability in the cloud. It highlights challenges and developed benchmarking based method that estimates available for the deployment of resources to a customer. Chavan et al. [2014] presented a higher resource availability method using clustering virtual machines in data centers. It optimizes shared resources to get better results for scalability and availability. The work performed auto-VM migration to manage availability and balancing that improves utilization. Nabi et al. [2016] presented a survey on current availability solutions in the cloud designed. This taxonomy discussed service Availability Forum, concepts, and mechanisms for baseline evaluation. It categorized proposed solutions, differences and similarities between various solutions in terms of availability and their impact on performance and SLA. Pietri et al. [2016] survey on mapping VMs onto PMs and categorize techniques based on mapping properties impact availability. This research work discussed scheduling actions, triggers, optimization goals, metrics, applications, and platforms used for evaluation in different techniques. Mishra et al. [2018] presented an "Energy-efficient VM-placement" (EE-VMsP) algorithm for optimal mapping. It performed VMs configuration and placement to reduce energy consumption and makespan by appropriate handling of heterogeneous and dynamic tasks.

It was concluded that resource availability can help while distributing VMs. Further, matching techniques manage adaptability and throughput utilization of resources. Therefore, the RM technique should be able to release unnecessary active resources at runtime to organize availability.

### 2.2 Resource Allocation Policies

The resource allocation performs by investigating task requirements for fulfilling customers' expectations. It is difficult to perform suitable resource allocation according to the application workload. Due to the growing complexity of the data center and application requirements problem becomes challenging. It needs to ensure user demand by allocating suitable resources without impacting energy consumption. While inappropriate allocation leads to performance degradation, and violation of SLA.

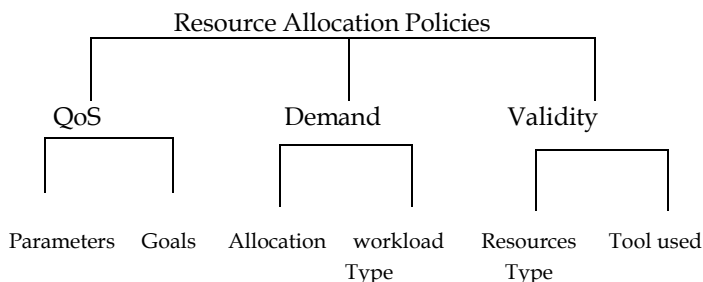


Fig. 3. Classification of Resource Allocation Policies

To address this problem, several researchers have presented different solutions. In the following sections, resource allocation policies are categorized based on heuristic and metaheuristic techniques in the perspective of QoS, demand, and validity, as depicted in figure 3.

Heuristic methods are used in various calculation patterns for finding an exact solution. It can provide a near solution to save energy but unable to predict suitable solutions [Tinghuai et al. 2014]. Metaheuristic approach follows the optimization process for appropriate matching using the convergence process [Vinothina et al. 2012; Zhang et al. 2016]. It provides a more accurate solution by managing resource usability and service quality.

### 2.2.1 Heuristic Resource Allocation Policies

Lee et al. [2010] presented “energy-conscious task consolidation” (ECTC) and “maximize resource utilization” (MaxUtil) for task consolidation using heuristics. The heuristic approach followed by the algorithm has efficiently reduced cost, energy consumption and maximize resource utilization. Nosrati et al. [2016] presented, “energy-efficient and latency optimized” (EELo) resource allocation method. It accomplished the identification of communication latencies and the geographical distance of the system using an optimization technique to minimize energy consumption. Verma et al. (2016) presented, “dynamic resource demand prediction and allocation framework” (DRDP-RF). It avoids unnecessary computational cost and time by predicting the demands, improves resource utilization and performance.

Beloglazov et al. [2012] presented a “modified best fit decreasing” (MBFD) method for resource provisioning and allocation to manage energy consumption and SLA violation. The VMs are sorted in utilization state; it improves the benefits of heterogeneity. Xiao et al [2013] proposed a fast up and slow down (FUSD) resource allocation method to overcome the workload of the system and minimize the number of active servers to reduce energy consumption. The skewness concept was used to measure resource utilization (scale up and down) for combining Virtual Machines.

Wu et al. [2011] presented the ProfminVio algorithm for mapping and scheduling based on customer dynamic demand. It focuses on maximizing profit and maintaining SLA policies. Ding et al. [2014] presented a QoS-aware resource recommendation (QoS-RR) method that performs on attribute-based matching for “price utility and group customer evaluation” which guarantees QoS requirements. The resource matching has considered with an objective to manage performance and QoS. Singh et al. [2015] presented QoS based resource provisioning (QoS-RP) technique, the k-means

clustering technique is used to identify and categorize requirements. It reduced execution cost and time without SLA violation.

Koch et al. [2016] presented a “probabilistic workload-aware dynamic resource allocation” (PWA-DRA) technique that improves the utilization of resources. The maximum likelihood estimation (MLE) method for optimization which considers utilizing heterogeneity for appropriate service; its’ execution performed to promote digital education for institutions. Peng et al. [2016] presented an “application type based resource allocation” (AT-RA) strategy. It adopted the resource-aware workload management scheme to map tasks on appropriate VMs.

Thein et al. [2018] proposed an “Energy-efficient Cloud Infrastructure Resource Allocation” (EE-RA) framework to efficient use of cloud infrastructure. It used reinforcement learning mechanism and fuzzy logic for effective resource allocation to manage revenue, cost and SLA violation and data center resources. Liu et al. [2018] presented Cloud\_RRSF, a multi-objective optimization algorithm for reliable resource allocation. algorithm to reinstate the failure of cloud service during resource allocation. It concentrates on performance management, resource utilization, and energy consumption to meet QoS requirements for improving consumer’s and provider’s relations.

### 2.2.2 Metaheuristic Resource Allocation Policies

Nuttapong et al. [2012] proposed a resource provisioning method based on cost optimization (RP-CO) using PSO. It has taken task characteristics and resource configuration to decide purchasing decision instances or reservations. Rodriguez et al. [2014] presented a “resource provisioning and scheduling algorithm” (RPSA) for scientific workflow application using PSO. To manage heterogeneity, it concentrates on elasticity and dynamicity. Fereshteh et al. [2017] presented a multi-objective PSO crowding distance (MOPSO-CD) algorithm to perform service allocation for maximizing the utilities of customers and providers. This method performed a step-wise selection of appropriate service concentrates on revenue and utilization.

Sivadon et al. [2012] presented optimal cloud resource provisioning (OCRP) algorithm that provision resources by following the demand under different stages and price uncertainties. The sample-average approximation (SAA) approach is used to estimate the cost of provisioning. Xue et al. [2014] proposed an improved differential algorithm for cloud task scheduling (IDA-CTS) using DE. The zooming factor, mutation and crossover strategy is adopted for selection to reduce makespan and energy consumption.

Aarti et al. [2015] presented, “agent-based automated service composition” (A2SC) resource provisioning scheme. The mobile agents potentially manage resource distribution to execute an application on minimum cost. Vakilinia et al. [2015] present a resource allocation model based on the Poisson process (RAM-PP) where jobs (constant and dynamic) are

Table. 2. Comparative Analysis of resource allocation techniques based on QoS parameters

Reference	Algorithm	QoS Parameters	Goals	Allocation environment	Workload type	Resources	Tool used
Lee et al. [2010]	ECTC MaxUtil	Utilization, energy, performance	Task consolidation heuristics in a way to reduce energy consumption and maximize resource utilization	Dynamic	Homogeneous	IaaS, VM	Gaussian method
Wu et al. [2011]	ProfminVio	Cost, profit, SLA	Identify customer varied request to assign heterogeneity based VMs	Dynamic	Heterogeneous	IaaS, VM, SaaS	Cloudsim
Beloglazov et al. [2012]	MBFD	Energy, SLA, utilization	Thresholds based energy-aware VMs migration while considering Qi's	Dynamic	Heterogeneous	IaaS, VMs	Cloudsim
Xiao et al. 2013	FUSD	Energy, performance	Hot spot mitigation and green computing	Dynamic	Synthetic workload	IaaS, CPU, VMs	Trace-driven simulation
Ding et al. [2014]	QoS-RR	Cost, profit	Resource matching performed in a way to meet Qi's	Dynamic	Heterogeneous	VMs	.NET and SQL Server
Singh et al. [2015]	QoS-RP	Cost, time	Workload assigned to suitable VM to fulfill QoS using K-means clustering	Dynamic	Heterogeneous	IaaS, VMs	Cloudsim
Koch et al. [2016]	PWA-DRA)	Cost	Probabilistic resource allocation scheme for education institutions	Dynamic	Synthetic workloads	IaaS	Discrete-event simulation
Peng et al. [2016]	AT-RA	Utilization, load balancing	Workload and resource-aware strategy to allocate only the required resource	Dynamic	Heterogeneous	VMs, CPU	Cloudsim, cloudstack
Nosrati et al. [2016]	EELO	Energy, response time	Energy-efficient latency optimization especially for media resource allocation	Static	Homogeneous	VMs, network resources	Cloudsim
Verma et al. 2016	DRDP-RF	Utilization	Resource allocation performed by predicting tenants demand	Dynamic	Heterogeneous	SaaS, VMs	Cloudsim
Nuttapong et al. [2012]	RP-CO	Cost	Optimal resource provisioning to perform best allocation decision	Dynamic	Heterogeneous	VMs	Workflow
Sivadon et al. 2012	OCRP	Cost	Different provisioning stages considered to reduce the cost of resource utilization	Dynamic	Heterogeneous	VMs	GLPK
Xue et al [2014]	IDA-CTS	Makespan, energy	Optimal scheduling using DE to minimize makespan and energy consumption	Dynamic	Homogeneous	VMs, MIPS	Cloudsim
Rodriguez et al. 2014	RPSA	Cost, deadline	Heterogeneity, elasticity, and dynamicity considered using PSO	Dynamic	Heterogeneous	IaaS, VMs	Cloudsim
Singh et al. 2015	A2SC	Cost	Identified dynamic requirements and runtime resource management	Dynamic	Heterogeneous	Datacenter	JAVA
Vakilinia et al. [2015]	RAM-PP	Job arrival rate, utilization	Homogenous and heterogeneous VMs considered according to jobs	Dynamic	Homogeneous	VMs, CPU	Discrete event simulation
Alexander et al. [2016]	LA-CBA	Cost, makespan, utilization	Allocation and workload managing strategy using Cuckoo-search	Dynamic	Heterogeneous	VMs, CPU	Cloudsim
Pillai et al. [2016]	RA-UPGT	Cost, response time	Allocation based on coalition formation with varied capabilities of agents	Dynamic	Heterogeneous	VMs	JAVA
Samimi et al. 2016	CDARA	Pricing	Auction-based market-oriented pricing model for resource allocation	Dynamic	Heterogeneous	VMs, MIPS	Cloudsim
Fereshteh et al. 2017	MOPSO-CD	Time, revenue, utilization	Revenue, time and utilization considered for users and providers	Dynamic	Heterogeneous	MIPS	Matlab
Thein et al. [2018]	EE-RA	Energy, profit, scalability, SLA	Energy consumption, utilization, and SLA based resource allocation performed	Dynamic	Heterogeneous	VMs, CPU	Cloudsim
Tafsiri et al. [2018]	CDA-RA	Profit, efficiency, utilization	Fairness usage of resources performed to improve efficiency, profit, and utilization	Dynamic	Heterogeneous	VMs	Cloudsim
Liu et al. [2018]	Cloud_RRSSF	Reliability, utilization, energy consumption	Reliable resource allocation performed to reduce energy consumption	Dynamic	Heterogeneous	Host, CPU	Cloudsim

Submitted by users. The method performed job blocking probability and distribution to manage utilization and cost. Samimi et al. [2016] presented, "combinatorial double auction resource allocation" (CDARA) model, it considers economic benefit for service providers and customers. The broker analysis task types and quantities the appropriate resource and budget. Tafsiri et al. [2018] presented a "Combinatorial double auction-based resource allocation" (CDA-RA) algorithm to strength provider and consumer relations. The technique followed a linear optimization mechanism to benefits the pricing of resource usage.

Alexander et al. [2016] presented a load-aware Cuckoo based allocation (LA-CBA) method for efficient resource management. A load-aware scheme is performed by virtual clusters and virtual machines in the datacenter. It manages deadline constraints without affecting computational cost, makespan, and utilization. Pillai et al. [2016] presented "resource allocation using the uncertainty principle of game theory" (RA-UPGT) for managing VMs according to demand. It performed demand-aware and topology-aware allocation based on coalition formation to manage task allocation time and communication cost.

### 2.2.3 Data Analysis

This section presents a comparative analysis of publications, sections 2.2.1 and section 2.2.2 discussed allocation policies which are tabulated in table 2. The analysis was done based on QoS parameters, prime goals of the work, scheduling environment and workload type, resources considered in experimentation and implementation tool. The graphical data shown in figure 4 and figure 5, represents QoS parameters, and resources considered in experimentation.

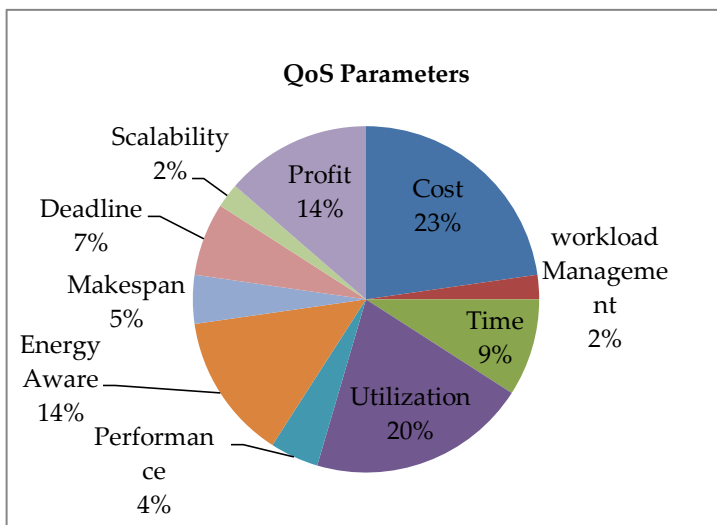


Fig. 4. Comparative analysis of resource allocation publications based on QoS parameters

Figure 4 shows the percentage of QoS parameters considered (cost, time, utilization, energy-aware, makespan, scalability, workload management, and profit) in research publications discussed in section 2.2.1 and 2.2.2. The graphical data demonstrates that cost examined more than other factors with 23% of the research papers. For the rest of the factors 21%

of research papers concentrate on utilization; time 9% and makespan 5%; energy-aware 14%; workload management and scalability 2%; deadline 7%; profit 14%; performance 5%. It shows publications have identified cost as a prime factor by researchers. Other factors such as scalability, elasticity, latency are still needed to be considered for better cloud resource management. Figure 5 shows the percentage of different resources used (VMs, host, data center, CPU, MIPS, Memory, IaaS, PaaS, SaaS and network resources) in experimentation. The graphical data demonstrates that VMs consideration 49% in the research publications. For the rest of the factors, 17% of research papers concentrate on IaaS; 15% CPU; 7% MIPS; data center 3%, Host and network resources 2%; SaaS 5%. This shows that VM is more attentive and utilized among resources by researchers; rest factors such as host, datacenter, PaaS, SaaS, network resources and memory can be explored in further research works. The comparative analysis shows the dynamic allocation for appropriate management of heterogeneity and QoS requirements. The existing techniques considered VM as an important factor for resources; hence work can be done on other parameters. The majority of resource allocation policies are implemented on cloudsim tool.

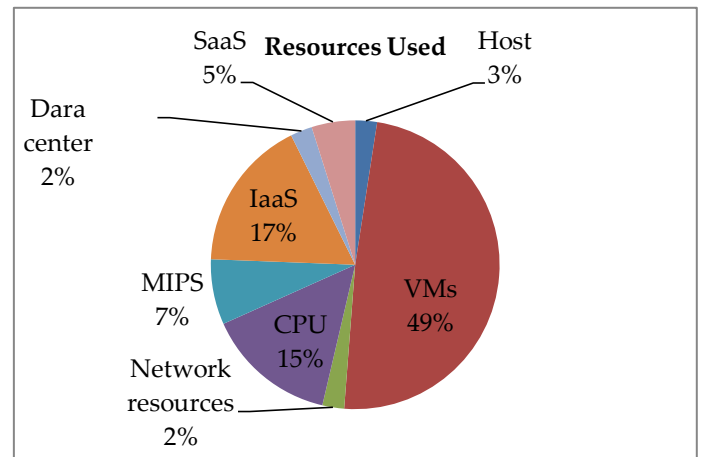


Fig. 5. Comparative analysis of resource allocation publications based on resources considered for the experiments

### 2.3 Resource Scheduling

In this section, resources scheduling has considered as a part of the study- challenges in resource scheduling, previous works on scheduling problems, benefits of various scheduling algorithms. Resource scheduling is the procedure of trading between tasks and resources [Alkhank et al. 2016]. The appropriate scheduling can fulfill task requirements.

In this regard, the research community has proposed different resource scheduling techniques discussed in the following. Heuristic and metaheuristic based resource scheduling techniques have been discussed in terms of QoS, demand, and validity depicted in figure 6. It is difficult to schedule resources at runtime while the workload is heterogeneous and dynamic.



### 2.3.1 Heuristic Resource Scheduling Policies

Sindhu et al. [2011] proposed the “longest cloudlet fastest processing element” (LCFP) and “shortest cloudlet fastest processing element” (SCFP) algorithms for efficient task scheduling. The sorting method designed to sort length-wise a list of VMs (PEs) to map tasks on sufficient resources.

Warneke et al. [2011] presented a scheme of dynamic resource scheduling (Nephele) to process data simultaneously. The Nephele performed allocation/ reallocation of resources for different types of jobs to manage cost, time and utilization. Lakhani et al [2013] proposed an optimization-based resource scheduling task grouping (ORS-TG) algorithm to group similar resource requirements. The task assignment is performed to assign a suitable VM based on MIPS requirements to manage utilization, cost and time.

Calheiros et al. [2014] presented EIPR, an enhanced IaaS cloud partial critical path (IC-PCP) for provisioning and scheduling. The priorities of tasks are considered based on different time slots that managed performance and makespan. Zhang et al. [2015] presented PRISM as a “fine-grain resource-aware” MapReduce scheduler. The phase-level scheduling scheme has been used to allocate certain resources according to each task execution that reduced idle time and makespan.

Chen et al. [2015] presented uncertainty-aware scheduling algorithm PRS to schedule dynamic workload according to resource availability. The interval number theory has been used to describe the uncertainty that impacts on scaling, energy, and utilization. Qizhi et al. [2016] present dynamic virtual resource management (DV-RM) framework to cope with traffic burst using a Gompertz curve and average moving method. The auto regression model and analyzing time series are used for predicting workload to guarantee SLA and cost.

Singh et al. [2016] proposed resource provisioning and scheduling (RPS) techniques to fulfill QoS parameters which run a scheduling policy based on suitability. It efficiently executes workload on resources with minimum execution cost, time and energy. Kong et al. [2016] present an auction mechanism based resource scheduling algorithm (AM-BRS) for pricing adaptive VM scheduling. The algorithm major factors are network bandwidth, auction deadline, pricing, and utilization. Ali et al. [2016] presented group tasks scheduling (GTS) algorithm, it schedule tasks by classifying into five different categories and four types of attributes. The scheduling performed to select a category and individual task for low latency.

Panda et al. [2018] proposed a pair-based task scheduling (PTS) algorithm that pair unequal tasks for efficient resource utilization. It uses an optimization mechanism to consider start time, duration and end time of task execution for managing resource availability. Panda et al. proposed a cloud task partitioning scheduling (CTPS) algorithm for online and offline scheduling. This paper also extends min-min and max\_min task partitioning scheduling algorithms that execute utilize heterogeneous resources for reducing execution time, makespan and cost.

### 2.3.2 Metaheuristic Resource Scheduling Policies

In this regard, Pandey et al [2010] presented PSO based heuristic technique (PSO-H) for optimal scheduling. It performed an optimal global solution within a reasonable

amount of time and cost. Babu et al. [2013] proposed an algorithm, “Honey bee behavior inspired load balancing” (HBB-LB) for an optimal load. The VM grouping scheme efficiently manages the load balancing, network problems, and scalability without affecting execution time and makespan. Pacini et al. [2013] presented dynamic scheduling based on ant colony optimization (DS-ACO). The custom VMs creation is performed based on job requests for faster execution, utilization, and balances the load on hosts.

Tsai et al. [2013] proposed an “improved differential evolution algorithm” (IDEA) method for task scheduling. This technique is based on Taguchi and differential evolution to enhance exploration and exploitation. Zuo et al. [2014] presented a “self-adaptive learning particle swarm optimization” (SLPSO) scheduling technique that performed based on particle decoding. The rank-order system has used for a permutation of tasks to evaluate particle, deadline, and QoS.

Tsai et al. [2014] presented a novel technique “hyper-heuristic scheduling algorithm” (HHSA) for optimal scheduling. It used detection and perturbation operator to find the best solution and minimize makespan by a low-level algorithm. Pacini et al. [2015] proposed a cloud scheduler based on ACO and genetic algorithm (CS-ACO-GA) for dynamic load information. The VM suitability is controlled content delivery, network message, throughput and response time. It manages performance and makespan. Duan et al. [2016] presented an energy-aware prediction model (PreAntPolicy) for the scheduling of heterogeneous VMs using an improved ant colony algorithm. It performed resource-intensive for application demand by dynamic provisioning of heterogeneous resources.

Singh et al. [2017] presented an efficient strategy (BULLET) for resource distribution using PSO with an objective to reduce cost, time and energy while meeting users' satisfaction and workload deadline. Zhou et al. [2017] presented hybrid glowworm swarm optimization (HGSO) for efficient scheduling. It also used quantum behavior that accelerates the convergence to easily escape from local optima to global best. It efficiently reduces execution cost, time and makespan. Mansouri et al. [2019] proposed a hybrid task scheduling algorithm named (FMPSO) using the fuzzy system and modified PSO. Crossover, mutation and velocity update techniques were used to improve performance and reduction of makespan.

### 2.3.3 Data Analysis

This section presents a comparative analysis of publications discussed in section 2.3.1 and 2.3.2. The comparative analysis of the above-discussed research work has done based on QoS parameters, the goal of the work, scheduling environment, workload type, resource considered in experimentation, and

Table. 3. Comparative Analysis of resource scheduling techniques based on QoS parameters

Reference	Algorithm	Parameters	Goals	Scheduling environment	Workload type	Resources	Tool used
Sindhu et al. [2011]	LCFP, SCFP	Makespan, utilization	Sorting order performed to schedule a task on appropriate VM	Static	Homogeneous	VMs, MIPS	Cloudsim
Warneke et al. 2011	Nephele	Cost, time, utilization	Dynamic task assigning to reduce processing cost	Dynamic	Heterogeneous	IaaS	Hadoop
Lakhani et al [2013]	ORS-TG	Cost, time	Grouping based on similar characteristics resources demand	Dynamic	Heterogeneous	VMs, MIPS	Cloudsim
Calheiros et al. 2014	EIPR	Makespan, performance	Task replication performed to avoid deadline that reduces budget	Dynamic	Heterogeneous	IaaS, VMs, memory	Cloudsim
Zhang et al. 2015	PRISM	Utilization, makespan	phase-level scheduler to allocate required resources for jobs	Static	Homogeneous	IaaS, CPU, memory	Hadoop 0.20.2.
Chen et al. 2015	PRS	Energy-aware, utilization	Uncertainty and resource-aware scheduling to manage scaling and heterogeneity	Dynamic	Heterogeneous	VMs, CPU, network resources	CloudStack/ CloudSim
Qizhi et al. 2016	DV-RM	Availability, cost	Scheduling performed using workload load forecasting method	Dynamic	Heterogeneous	VMs, CPU	JAVA
Kong et al. 2016	AM-BRS	Profit, utilization	Pricing adaptive scheduling using an auction mechanism	Dynamic	Homogeneous	VMs	Cloudsim
Ali et al. [2016]	GTS	Latency, execution time, load balancing	Task grouping based on attributes to manage latency and workload	Dynamic	Heterogeneous	Network resources	JAVA
Singh et al.	RP-RS	Cost, time and energy	Load-aware resource provisioning and scheduling using clustering	Dynamic	Heterogeneous	IaaS, VMs	Cloudsim
Pandey et al 2010	PSO-H	Cost	Mapping performed based on optimal dependency using PSO	Dynamic	Homogeneous	CPU, network resources	JSwarm
Babu et al. [2013]	HBB-LB	Makespan, load balancing	Load balancing performed based on VM capacity and current state	Dynamic	Heterogeneous	VMs, network resources	Cloudsim
Pacini et al. 2013	DS-ACO	Load balancing	Custom VM scheduling based on host load	Dynamic	Heterogeneous	Host, VMs, CPU	Cloudsim
Tsai et al. 2013	IDEA	Cost, makespan	Optimal solution based resource task matching performed	Dynamic	Homogeneous	IaaS	
Zuo et al. [2014]	SL-PSO	Profit, cost	Cost-effective scheduling to manage deadline and profit	Dynamic	Homogeneous	VMs, CPU	Matlab 7.0
Tsai et al. 2014	HHSA	Makespan, time	To find a better solution for leveraging the strengths of low-level algorithms	Dynamic	Heterogeneous	VMs, CPU	Cloudsim/ Hadoop
Pacini et al [2015]	CS-ACO-GA	Utilization, response time	Load aware scheduling to get throughput and response time	Dynamic	Homogeneous	Host, VMs	Cloudsim
Duan et al. [2016]	PreAntPolicy	Energy, utilization	Energy-aware heterogeneous VM scheduling scheme	Dynamic	Heterogeneous	VMs, CPU	Cloudsim
Singh et al. [2017]	BULLET	Cost, time and energy	Workload execution performed based on QoS using PSO	Dynamic	Heterogeneous	IaaS, PaaS, SaaS	Cloudsim
Zhou et al. [2017]	HGSO	Cost, time, makespan	Hybridization of evolutionary, quantum transition, and random walk to find global best.	Dynamic	Heterogeneous	VMs MIPS	Cloudsim
Panda et al. [2018]	PTS	Execution time, layover time	Task scheduling technique evaluates execution time to reduce layover time	Dynamic	Heterogeneous	VMs MIPS	MATLAB R2014a
Panda et al. [2018]	CTPS	Makespan, interval time	Extended scheduling algorithms proposed to reduce execution time and makespan	Dynamic	Heterogeneous	VMs	MATLAB R2014a
Mansouri et al. [2019]	FMPPO	Makespan, efficiency	Optimal solution performed for appropriate task scheduling using dynamic velocity	Dynamic	Heterogeneous	VMs	Cloudsim

tool used for implementing research proposal, represented by table 3. Further, the graphical results of the comparative analysis of table 3 presented in figure 7 and figure 8.

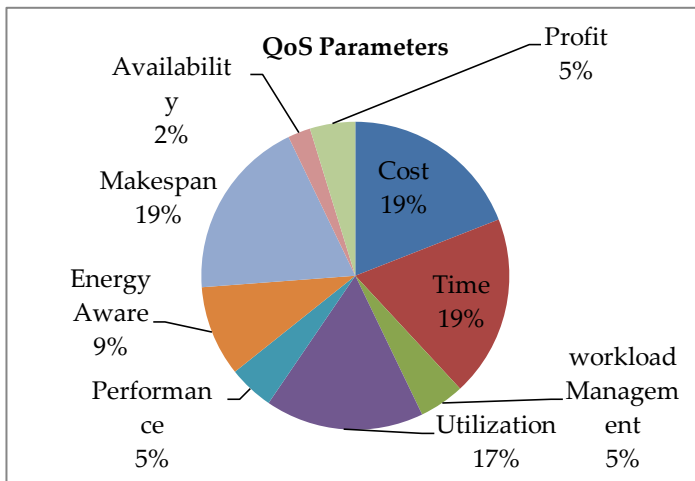


Fig. 7. Comparative analysis of resource scheduling publications based on QoS parameters

Figure 7 shows the percentage of QoS parameters considered (cost, time, utilization, energy awareness, makespan, scalability, workload management, and profit) in publications. The graphical data demonstrates that cost, time and makespan have considered more than other factors with 19% of the research papers. For the rest of the factors 17% research papers concentrate on utilization; 9% energy-aware; workload management, performance and profit 5%; availability 2%. It shows cost has examined more than all other parameters by the researchers; scalability and elasticity still need more attention

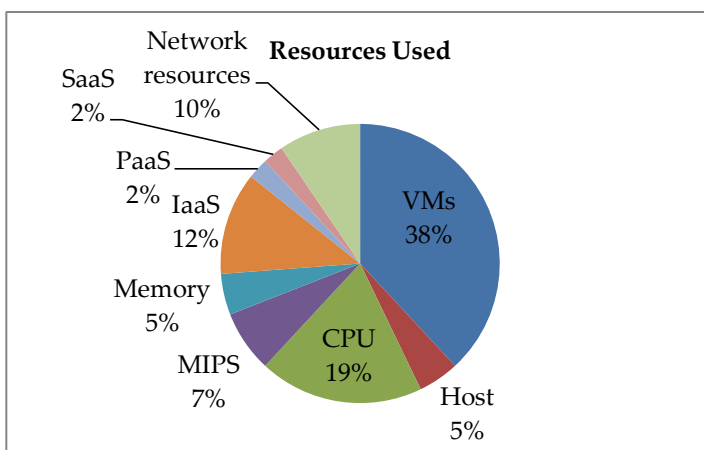


Fig. 8. Comparative analysis of resource scheduling publications based on resources considered for the experiments

Figure 8 shows the percentage of different resource uses (VMs, host, data center, CPU, MIPS, Memory, IaaS, PaaS, SaaS and network resources) in experimentation. The graphical data clearly demonstrates that VMs considered and analyzed more than other resources with 38% of the research papers. For the rest of the factors 19% of research papers concentrate

on CPU; IaaS 12%; 10% network resources; 5% memory and MIPS 7%; SaaS and PaaS 2%. This shows that VMs utilized as the primer of all resources by the researchers; the rest of the resources need more work for experimentation.

It shows that dynamic, heterogeneity, QoS requirements and VMs have been widely considered. Cloudsim is most using tool by researchers for algorithm implementation as an experimental tool.

### 3 OPEN CHALLENGING ISSUES AND FUTURE RESEARCH DIRECTIONS

This survey was conducted with a wide range of existing research work in cloud resource management. The literature considered for the study includes conferences, journals, articles, books and other published reports on resource management in cloud computing. The literature work has been organized from more than 100 prominent and relevant research publications. The survey on QoS, resource allocation and resource scheduling in cloud resource management was considered during the study. [Mezni et al. 2018; Singh et al. 2016; Kumar M. et al. 2019]

This literature survey shows that challenges of resource management in cloud computing exist at various levels of hardware, software, and networking. Due to several approaches and affordable services, cloud service demand is growing tremendously. The provider is responsible to identify consumer requests, availability of resources, current workload demand, resource monitoring, and suitable matchmaking. It has become a challenge to consider resource distribution under an unpredictable, heterogeneous and dynamic environment to ensure accurate matching based on unexpected demand from the industries. Despite, number of research work that exists in cloud resource management there are some open research challenges that are still influencing the performance and quality of cloud services.

#### 3.1 Open Challenging Issues

The survey contributes to existing literature with an extension of open research challenges and suggestions for academic professionals and tentative researchers. The survey identifies and highlights open research issues to manage resource specifications in accordance with workload based QoS requirements. The research community continues doing research efforts to provide adequate service based on potential growth in cloud services. Despite, a huge development in the field, it was very difficult to manage a large amount of data according to industry expectations. The emerging growth of cloud computing is still facing open research issues that impact performance. Some of the identified open research issues that need to be resolved are discussed below in table 4.

#### 3.2 FUTURE RESEARCH DIRECTIONS

This section elucidates the future research directions for resource management in cloud computing. Resource management is challenging in a cloud that shows unexpected results due to an inappropriate match. The suitable resource to

Table 4. Open Challenging Issues

Resource categorization	Description of Open Issues
Resource provisioning	<ol style="list-style-type: none"> <li>1. Efficient resource provisioning needs resource availability of resources.</li> <li>2. It is difficult to reserve resources for provisioning to provide instant response and avoid deadline of tasks.</li> <li>3. The workload analyzer cannot evaluate the dynamic and heterogeneous workload accurately to perform efficient provisioning.</li> </ol>
Resource allocation	<ol style="list-style-type: none"> <li>1. The optimal workload analysis for resource allocation is challenging.</li> <li>2. QoS parameters and SLA management are difficult without compromising performance.</li> <li>3. Resource allocation needs dynamic provisioning to improve utilization and reliability is also challenging.</li> </ol>
Resource mapping	<ol style="list-style-type: none"> <li>1. Mapping is difficult in the cloud; it requires a hybrid nature for mapping assignment.</li> <li>2. It is difficult to predict workload requirements based on historical data. It is an optimization problem to pair suitable tasks to a resource.</li> <li>3. Accurate evaluation of mapping is difficult due to the dynamic and unpredictable nature of consumer demand.</li> </ol>
Resource scheduling	<ol style="list-style-type: none"> <li>1. It is difficult to identify different workload patterns with general scheduling techniques.</li> <li>2. Runtime scheduling and location-aware allocation is a challenging scheduling problem.</li> <li>3. It is a challenge to efficient scheduling while managing cost, energy, and profit.</li> </ol>
Application scheduling	<ol style="list-style-type: none"> <li>1. Cloud consumer expectations are to access cost and time effective services. So, providers ensure to schedule workload based on application requirements.</li> <li>2. It is important to meet application QoS requirements to avoid deadlines of tasks.</li> <li>3. Application-aware scheduling is challenging to meet utilization and performance without SLA violation.</li> </ol>
VM management	<ol style="list-style-type: none"> <li>1. What is the impact of VM instance creation on utilization and energy consumption?</li> <li>2. How VM geographical location and migration impact on latency and performance.</li> <li>3. What negotiation criteria suitable for VM configuration and placement.</li> </ol>
Workload management	<ol style="list-style-type: none"> <li>1. Workload management is challenging in resource management.</li> <li>2. It is difficult to manage resources according to their specifications due to unspecified and heterogeneous workloads.</li> <li>3. It is difficult to organize runtime workload and migration due to task dependency and unpredictable workload.</li> </ol>
QoS parameters	<ol style="list-style-type: none"> <li>1. Providers' aim is to provide cloud services based on expected performance that compromises on other QoS parameters.</li> <li>2. Customers are paying for cloud services expect to fulfill QoS requirements.</li> <li>3. How VM provisioning, scheduling, consolidation, migration and monitoring impact efficiency, and QoS parameters.</li> </ol>
SLA	<ol style="list-style-type: none"> <li>1. Cloud services are based on pay-per-usage where SLA management is challenging.</li> <li>2. Inappropriate matching of resources to workload can't fulfill QoS and SLA requirements.</li> <li>3. Resource management is challenging due to dynamic, heterogeneous and uncertainty that can violate SLA.</li> </ol>
Big data	<ol style="list-style-type: none"> <li>1. It is a challenge to retrieve and deliver huge data according to customers' expectations.</li> <li>2. It is also difficult to properly distribute the workload on resources based on varied application requirements.</li> <li>3. Resource distribution to simultaneously millions of requests is also challenging due to a large amount of data (homogeneous/heterogeneous).</li> </ol>
Latency	<ol style="list-style-type: none"> <li>1. Millions of users' concurrent requests of resources are difficult to manage.</li> <li>2. Dynamic assignment of resources and workload migration leads to high latency.</li> <li>3. Location-aware resource allocation is difficult to find a suitable route.</li> </ol>
Energy consumption	<ol style="list-style-type: none"> <li>1. How the allocation decision impacts energy consumption.</li> <li>2. How different application requirements impact energy consumption and utilization?</li> <li>3. Accurate availability of the resource is difficult to reduce energy consumption for resource discovery.</li> </ol>
Utilization	<ol style="list-style-type: none"> <li>1. How scheduling criteria impacts on utilization?</li> <li>2. How dynamic resource allocation effects VM and host utilization?</li> <li>3. How VM instance impact VM utilization?</li> </ol>
Scalability	<ol style="list-style-type: none"> <li>1. How scheduling technique impact scalability based on application requirements?</li> <li>2. How available resource can be scaled to manage demand extensively?</li> <li>3. It is challenging to scale up and down based on a dynamic and heterogeneous workload.</li> </ol>
Security	<ol style="list-style-type: none"> <li>1. Security is an important QoS parameter to manage reliability and SLA.</li> <li>2. To prevent from unauthorized access using firewalls and optimization tools may harm to performance.</li> <li>3. To secure consumer data from theft and harmful attacks is a challenging issue.</li> </ol>
Monitoring	<ol style="list-style-type: none"> <li>1. Allocated resource monitoring in a dynamic cloud environment is difficult.</li> <li>2. It is also difficult to optimize monitoring while a pattern of demand is unpredictable.</li> <li>3. Monitoring of current usage resources is challenging while migrating data to other VM.</li> </ol>



workload provides expected performance and fulfills QoS requirements. This survey has evaluated existing work in the field of resource management. Further, it has presented open research challenges, conclusions of the survey. The potential growth in cloud computing needs consistent research work in the field of resource management. A standardized technique requires for efficient resource allocation and scheduling that can properly utilize cloud resources. It would perform appropriate usage of resources; manage reliability and consistency between providers and consumers. This research survey discusses research challenges, suggestions for further research work that should be considered in the context of resource management. The key findings of the cloud resource management mechanism from the literature would various benefits. Some of the key findings are:

- Resource matching should be done based on workload patterns and resource configuration. An autonomic technique would improve resource utilization and performance.
- Workload management should be done based on availability and scalability. It would help to SLA management, overload and under load resource provisioning.
- Resource scheduling should be done after identify resource provisioning to perform cost-effective workload execution.
- Resource scheduling is an optimization problem, and runtime environment can identify workload requirements, availability and SLA to place appropriate migration and scheduling.
- Effective resources management helps to improve utilization, energy consumption and cost reduction that can manage availability, scalability, and profit.
- Resource management is challenging to manage QoS requirements and SLA.
- An authentication detection requires to secure user data under certain situations should meet to fulfill customers' expectations.
- An autonomic technique requires determining resource configuration, availability, and scalability for workload requirements (homogeneous and heterogeneous) to process millions of concurrent requests.
- The dynamic environment should adapt to identify dynamic workload fluctuation to perform appropriate VM configuration and placement.
- To fulfill customers' QoS requirements and SLA resource provisioning criteria should be able to identify demand criteria.

These future research directions if implicating in further research development for fulfilling resource management in cloud computing would enhance the quality of cloud service and help in QoS parameters and SLA.

#### 4 CONCLUSIONS AND SUMMARY

In this survey paper, a systematic literature review was conducted on cloud resource management. The research presents a cloud resource management mechanism with respect to several parameters including SLA, resource

allocation, and scheduling with their testbed QoS (parameters and validation check), service (demand and monitoring), and resources (VMs and availability). This literature has highlighted some open research issues that are impacting on the cloud. It has noticed that the best resource discovery for the best pairing of workload to the resource is challenging. Providers' efforts are to manage SLA violations and QoS parameters. We have identified the effect of workload matching on resource management and utilization.

The survey also identifies future research work in the direction of autonomic scaling and elasticity based on resource allocation and scheduling to select a suitable resource for managing runtime workload on resources to be considered. The research work contributed to other than QoS parameters and resource usage (hardware/software) based on application dynamic constraints that can be done. It is also challenging for providers to manage the growing demand for cloud service to ensure reliability and profits. The scalability of cloud resources according to scaling in demand in the perspective of QoS parameters and SLA is a potential research area in cloud resource management. The resource reservation to avoid deadlock is also effects on utilization that need to develop runtime optimal usage of resources. Following are the facts can be concluded:

- ✓ Resource provisioning and scheduling in the cloud environment is an optimization problem.
- ✓ It is challenging to reserve a desirable resource for future predictions.
- ✓ The static environment can reduce energy consumption and cost. But, it leads to under-provisioning and overprovisioning, unable to perform proper workload distribution.
- ✓ Dynamic resource allocation can exploit cloud resources while inappropriate workload allocation leads to energy consumption and communication overheads.
- ✓ Efficient resource provisioning from the perspective of QoS parameters and SLA is challenging.
- ✓ There is a need for an efficient technique that would manage resources according to different application design.
- ✓ What is a suitable schedule plan to efficiently meet the requires of different services?
- ✓ How synchronization improves availability and solution construction.
- ✓ It is difficult to manage under-provisioning and over-provisioning without optimal scalability.
- ✓ How delivered service can be profitable and competitive in the perspective of providers and consumer expectations.
- ✓ Security is important while considering SLA and performance.

#### REFERENCES

- [1] Abdullah Monir, Lu Kuan, Wieder Philipp, Yahyapour Ramin1, "A Heuristic-Based Approach for Dynamic VMs Consolidation in Cloud Data Centers, Springer," Arab J Sci Eng, vol. 42, pp. 3535- 3349, 2017. DOI 10.1007/s13369-017-2580-5
- [2] Alexander Aneena Ann, Joseph Divya Lissia, "An Efficient Resource Management For Prioritized Users In

- Cloud Environment Using Cuckoo Search Algorithm," ELSEVIER Procedia Technology, vol. 25, pp. 341-348, 2016. DOI:10.1016/j.protcy.2016.08.116
- [3] Ali Hend Gamal El-Din, Saroit Imane Aly, Kotb Amira Mohamed, "Grouped tasks scheduling algorithm based on QoS in cloud computing network," Egyptian Informatics Journal, 2016. <http://dx.doi.org/10.1016/j.eij.2016.07.002>
- [4] Alkhank Ehab Nabel, Lee Sai Peck, Khan Sai Ur Rehman, "Cost-aware challenges for workflow scheduling approaches in cloud computing environments: Taxonomy and opportunities," ELSEVIER, Future Generation Computer System, vol. 50, pp.3-21, 2015. <http://dx.doi.org/10.1016/j.future.2015.01.007>
- [5] Abdelsamea Amany, Ali A. El-Moursy, Elsayed E. Hemayed, Hesham Eldeeb, "Virtual machine consolidation enhancement using hybrid regression algorithms," Egyptian Informatics Journal, vol. 18, no. 3, pp. 161-170, 2017. <https://doi.org/10.1016/j.eij.2016.12.002>.
- [6] Beloglazov Anton, Abawajy Jemal, Buyya Rajkumar, "Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing," ELSEVIER Future Generation Computer System, vol. 28, pp. 755-768, 2012. DOI:10.1016/j.future.2011.04.017
- [7] Babu L.D. Dhinesh and Venkata Krishna P., "Honey bee behaviour inspired load balancing of tasks in cloud computing environments," ELSEVIER Applied Soft Computing, vol. 13, pp. 2292-2303, 2013. <http://dx.doi.org/10.1016/j.asoc.2013.01.025>
- [8] Buyya Rajkumar, Yeo Chin Shin, Venugopal Srikumar, Broberg James, Brandic Ivona, "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility," ELSEVIER Future Generation Computer Systems, vol. 25, no.6, pp. 59-616, 2009. DOI: 10.1016/j.future.2008.12.001
- [9] Buyya Rajkumar, Garg Saurabh Kumar, Calheiros Rodrigo N., "SLA-oriented resource provisioning for cloud computing: Challenges, architecture, and solutions," International Conference on Cloud and Service Computing, Hong Kong, pp. 1-10, 2011. doi: 10.1109/CSC.2011.6138522
- [10] Calheiros Rodrigo N. and Buyya Rajkumar, "Meeting Deadline of Scientific Workflows in Public Clouds with Tasks Replication," IEEE Trans. Parallel and Distributed, vol. 25, no. 7, pp. 1787-1796, 2014. DOI 10.1109/TPDS.2013.238
- [11] Chavan Vinay, Kaveri Parag Ravikant, "Clustered virtual machines for higher availability of resources with improved scalability in cloud computing," 1st International Conference on Networks & Soft Computing (ICNSC2014), Guntur, pp. 221-225, 2014. doi: 10.1109/CNSC.2014.6906707
- [12] Chavan Vinay, Kaveri Parag Ravikant, "Shared resource clustering for load balancing and availability in cloud," 2nd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, pp. 1004-1007, 2015.
- [13] Chen Yi-Hsuan, Chen Chi-Yuan, "Service Oriented Cloud VM Placement Strategy for Internet of Things," IEEE Access, vol. 5, pp. 25396-25407, 2017. doi: 10.51109/ACCESS.2017.2769667
- [14] Chen Huangke, Zhu Xiaomin, Guo Hui, Zhu Jiangnan, Qin Xiao, Wu Jianhong, "Towards energy-efficient scheduling for real-time tasks under uncertain cloud computing environment," In Journal of Systems and Software, vol. 99, pp. 20-35, 2015. ISSN 0164-1212. <https://doi.org/10.1016/j.jss.2014.08.065>
- [15] Choudhary Vidyanand, Vithayathil Joseph, "The Impact of Cloud Computing: Should the IT Department be Organized as a Cost Center or a Profit Center?," Emerald Journal of Management Information System, vol. 30, no. 2, pp-67-100, 2015. DOI: 10.2753/MIS0742-1222300203
- [16] Chowdhury Mohammed Rashid, Mahmud Mohammed Raihan, Rehman Rashedur M., "Implementation and performance analysis of various VM placement strategies in Cloudsim," Springer Open Journal of Cloud Computing: Advances, Systems and Application, pp. 4-20, 2015. DOI 10.111186/s13677-015-0045-5
- [17] Chowdhury Mohammed Rashid, Mahmud Mohammad Raihan, Rahman Rashedur M, "Clustered based VM placement strategies," IEEE/ACIS 14th International Conference on Computer and Information Science (ICIS), Las Vegas, NV, pp. 247-252, 2015. doi: 10.1109/ICIS.2015.7166601
- [18] Ding Shuai, Chengyi Xia, Qiong Cai, Kaile Zhou, Shanlin Yang, "QoS-aware resource matching and recommendation for cloud computing system," ELSEVIER Applied Mathematics and Computation, vol. 247, pp. 941-950, 2014. <http://dx.doi.org/10.1016/j.amc.2014.09.058>
- [19] Duan Hancong, Chen Chao, Min Geyong, Wu Yu, "Energy-aware scheduling of virtual machines in heterogeneous cloud computing systems," ELSEVIER, Future Gen. Comp. Sys, 2016. <http://dx.doi.org/10.1016/j.future.2016.02.016>.
- [20] Feng Ye, Zhijian Wang, Feng Xu, Yuanchao Zhou, Fachao Zhou, Shaosong Yang, "A Novel Cloud Load Balancing Mechanism in Premise of Ensuring Qi's," Taylor & Francis Intelligent Automation and Soft Computing, vol. 19, no. 2, pp. 151-163, 2013. <http://dx.doi.org/10.1080/10798587.2013.786968>
- [21] Ferdous Md Hasanul , MurshedManzur , Calheiros Rodrigo N. , Buyya Rajkumar, "Multi-objective, Decentralized Dynamic Virtual Machine Consolidation using ACO Metaheuristic in Computing Clouds," Wiley Concurrency and Computation: Practice and experience, pp. 1-40, 2016. DOI: 10.1002/cpe.
- [22] Fereshteh Sheikholeslami, Nima Jafari Navimipour, "Service allocation in the cloud environments using multi-objective particle swarm optimization algorithm based on crowding distance," ELSEVIER Swarm and Evolutionary Computation, vol. 35, pp. 53-64, ISSN 2210-6502, 2017. <https://doi.org/10.1016/j.swevo.2017.02.007>.
- [23] Ghaderi Abdolsalam , Jabalameli Mohammad Saeed, Barzinpour Farnaz, Rahmaniani Ragheb, "An Efficient Hybrid Particle Swarm Optimization Algorithm for Solving the Uncapacitated Continuous Location-Allocation Problem," Springer Networks and Spatial Economics, 2012. DOI 10.1007/s11067-011-9162-y

- [24] Goudarzi Hadi, Ghasemazar Mohammad, Pedram Massoud, "SLA-based Optimization of Power and Migration Cost in Cloud Computing," 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (ccgrid 2012), Ottawa, ON, pp. 172-179, 2012. doi: 10.1109/CCGrid.2012.112
- [25] Gupta Harshit, Dastjerdi Amir Vahid, Ghosh Soumya K., Buyya Rajkumar, "iFogSim: A toolkit for modeling and simulation of resource management techniques in the Internet of Things, Edge and Fog computing environments," WILEY, software: practice and experience, vol. 47, no. 9, pp. 2175-1296, 2017. https://doi.org/10.1002/spe.2509
- [26] <http://www.rightscale.com>
- [27] Julia Amin, Sundararajan Elankovan and Othman Zalinda, "Cloud Computing Service Composition: A Systematic Literature Review," ELSEVIER, Expert Systems with Applications, vol. 41, pp. 3809-3824, 2014. <http://dx.doi.org/10.1016/j.eswa.2013.12.017>
- [28] Jamshidi Pooyan, Ahmad Aakash, and Pahl Claus, "Cloud Migration Research: A Systematic Review," IEEE Transactions on Cloud Computing, vol. 1, No. 2, pp. 142-157, 2013.
- [29] Kalra Mala, Singh Sarbjeet, "A review of metaheuristic scheduling techniques in cloud computing," Egyptian Informatics Journal, 2015. <http://dx.doi.org/10.1016/j.eij.2015.07.001>
- [30] Khan Md Anit, Paplinski Andrew, Khan Abdul Malik, Murshed Manzur, and Buyya Rajkumar, "Dynamic Virtual Machine Consolidation Algorithms for Energy-Efficient Cloud Resource Management: A Review," Springer International Publishing AG 2018 W. Rivera (ed.), Sustainable Cloud and Energy Services, 2018. DOI 10.1007/978-3-319-62238-5\_6
- [31] Koch Fernando, Assuncao Marcos D., Cardonha Carlos, Netto Marco A.S., "Optimising resource costs of cloud computing for education," ELSEVIER Future Generation Computer Systems, vol. 55, pp. 473-479, 2016. <http://dx.doi.org/10.1016/j.future.2015.03.013>
- [32] Kong Weiwei, Lei Yang, Ma Jing, "Virtual machine resource scheduling algorithm for cloud computing based on auction mechanism," ELSEVIER Optik, vol. 127, pp.5099-5104, 2016. <http://dx.doi.org/10.1016/j.ijleo.2016.02.061>
- [33] Kumar Mohit, Sharma S.C., Goel Anubhav, Singh S.P., "A comprehensive survey for scheduling techniques in cloud computing," Journal of Network and Computer Applications, vol. 143, pp. 1-33, 2019. doi.org/10.1016/j.jnca.2019.06.006
- [34] Kumar Ajay, Bawa Seema, "A comparative review of meta-heuristic approaches to optimize the SLA violation costs for dynamic execution of cloud services," pp.1-14, 2019. <https://doi.org/10.1007/s00500-019-04155-4>
- [35] Lakhani Jignesh, Bheda Hitesh A., "An Approach to Optimized Resource Scheduling using Task Grouping in Cloud," IJARCSSE, vol. 3, no. 9, pp.594-598, 2013.
- [36] Lee Young Choon, Zomaya Albert Y., "Energy efficient utilization of resources in cloud computing systems," Springer J. Supercomputing, 2010. DOI 10.1007/s11227-010-0421-3
- [37] Liu Xiao Xi, Qiu Jian, Zhang Jian Ming, "High Availability Benchmarking for Cloud Management Infrastructure," In International Conference on Service Sciences, Wuxi CHINA, pp. 163-168, 2014. doi: 10.1109/ICSS.2014.28
- [38] Liu Dan, Sui Xin, Li Li, Jiang Zhengang, Wang Huan, Zhang Zetian, Zeng Yan, "A cloud service adaptive framework based on reliable resource allocation," Future Generation Computer Systems, vol. 89, pp. 455-463, 2018. doi.org/10.1016/j.future.2018.05.059
- [39] Madni Syed Hamid Hussain, Latiff Muhammad Shafie Abd, Coulibaly Yahaya, Abdulhamid Shafi, "Resource scheduling for infrastructure as a service (IaaS) in cloud computing: Challenges and opportunities," ELSEVIER, Journal of Network and Computer Applications, vol. 6, pp.173-200, 2016. <http://dx.doi.org/10.1016/j.jnca.2016.04.016>
- [40] Manvi S. Sunilkumar, Shyam G. Krishan, "Resource management for Infrastructure as a Service (IaaS) cloud computing: A survey," ELSEVIER Journal of Network and Computer Applications, vol. 41, pp. 424-440, 2014. <http://dx.doi.org/10.1016/j.jnca.2013.10.004>
- [41] Mansouri Najme, Zade Behnam Mohammad Hasani, Javidi Mohammad Masoud, "Hybrid task scheduling strategy for cloud computing by modified particle swarm optimization and fuzzy theory," Computers & Industrial Engineering, vol. 130, pp. 597-633, 2019. doi.org/10.1016/j.cie.2019.03.006.
- [42] Masdari Mohammad, Nabavi Sayyid Shahab, Ahmadi Vafa, "An overview of virtual machine placement schemes in cloud computing," Journal of Network and Computer Applications, vol. 66, pp. 106-127, 2016. <https://doi.org/10.1016/j.jnca.2016.01.011>
- [43] Mezni Haithem, Aridhi Sabeur, Hadjali Allel, "The uncertain cloud: State of the art and research challenges," International Journal of Approximate Reasoning, vol. 103, pp.139-151, 2018. doi.org/10.1016/j.ijar.2018.09.009
- [44] Mishra Sambit Kumar, Puthal Deepak, Sahoo Bibhudatta, Jayaraman Prem Prakash, Jun Song, Zomaya Albert Y., Ranjan Rajiv, "Energy-efficient VM-placement in cloud data center," Sustainable Computing: Informatics and Systems, vol. 20, pp. 48-55, 2018. doi.org/10.1016/j.suscom.2018.01.002
- [45] Mustafa Saad, Nazir Babar, Hayat Amir, Khan Atta ur Rehman, Madni Sajjad A., "Resource management in cloud computing: Taxonomy, prospects, and challenges," ELSEVIER Computers and Electrical Engineering, vol. 47, pp.186-203, 2015. <http://dx.doi.org/10.1016/j.compeleceng.2015.07.021>
- [46] Nabi Mina, Toeroe Maria, Khendek Ferhat, "Availability in the cloud: State of the art," Journal of Network and Computer Applications, vol. 60, pp. 54-67, 2016. <https://doi.org/10.1016/j.jnca.2015.11.014>
- [47] Nosrati Masoud, Karimi Ronak, "Energy efficient and latency optimized media resource allocation," Emerald International Journal of Web Information Systems, vol. 12, no. 1, pp. 2-17, 2016. DOI 10.1108/IJWIS-10-2015-0031
- [48] N. Nuttapong, S. Booncharoen, A. Tiranee, "Cost Optimization in Cloud Provisioning using Particle Swarm Optimization," IEEE 9th Int'l Electrical Engg./Electronics, Comp., Telecommunications and Information



- Technology, 2012.
- [49] Pacini Elina, Mateos Cristian, Garcia Garino Carlos, "Dynamic Scheduling of Scientific Experiments on Clouds using Ant Colony Optimization," Proceeding of 3rd Intl' Conf. on Parallel, Distributed, Grid and Cloud Computing for Engg., Civil-Comp Press, Scotland.
- [50] Pacini Elina, Mateos Cristian, Garcia Garino Carlos, "Balancing throughput and response time in online scientific Clouds via Ant Colony Optimization," ELSEVIER Advances in Engineering Software, vol. 84, pp. 31-47, 2015. <http://dx.doi.org/10.1016/j.advengsoft.2015.01.005>
- [51] Panda Sanjaya Kumar, Das Satyabrata, "Task Partitioning Scheduling Algorithms for Heterogeneous Multi-Cloud Environment," Arabian Journal for Science and Engineering, vol. 43, no. 2, pp. 913-933, 2018. DOI:10.1007/s13369-017-2798-2
- [52] Panda Sanjaya Kumar, Nanda Shradha Surachita, Bhoi Sourav Kumar, "A pair-based task scheduling algorithm for cloud computing environment," Journal of King Saud University - Computer and Information Sciences, 2018. <https://doi.org/10.1016/j.jksuci.2018.10.001>
- [53] Pandey Suraj, Wu Linlin, Mayura Guru Siddeswara, Buyya Rajkumar, "A Particle Swarm Optimization-based Heuristic for Scheduling Workflow Applications in Cloud Computing Environments," 24th IEEE Int'l Conf. on advanced info. Networking & App., pp. 400-407, 2010. DOI: 10.1109/AINA.2010.31
- [54] Peng Jun-Jie, Zhi Xiao-Fei, Xie Xiao-Lan, "Application type based resource allocation strategy in cloud environment," ELSEVIER Microprocessors and Microsystems, pp.1-7, 2016. <http://dx.doi.org/10.1016/j.micpro.2016.09.014>
- [55] Pietri Iliia, Sakellariou Rizos, "Mapping Virtual Machines onto Physical Machines in Cloud Computing: A Survey," ACM Computing Survey, vol. 49, no. 3, pp. 1-30, 2016. DOI: <http://dx.doi.org/10.1145/2983575>
- [56] Pillai Parvathy S., Rao Shrisha, "Resource Allocation in Cloud Computing Using the Uncertainty Principle of Game Theory," IEEE Systems Journal, vol. 10, no. 2, pp. 667-648, 2016. DOI 10.1109/JSYST.2014.2314861
- [57] Prasad Vivek Kumar, Bhavsar Madhuri, "Efficient Resource Monitoring and Prediction Techniques in an IaaS Level of Cloud Computing: Survey," In: Patel Z., Gupta S. (eds) Future Internet Technologies and Trends. ICFITT 2017. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 220. Springer, Cham, 2018. [https://doi.org/10.1007/978-3-319-73712-6\\_5](https://doi.org/10.1007/978-3-319-73712-6_5)
- [58] Qizhi Zhang, Haopeng Chen, Yuxi Shen, Sixiang Ma, Heng Lu, "Optimization of virtual resource management for cloud applications to cope with traffic burst," In Future Generation Computer Systems, vol. 58, pp. 42-55, 2016. ISSN 0167-739X, <https://doi.org/10.1016/j.future.2015.12.011>
- [59] Rodriguez Maria Alejandra and Buyya Rajkumar, "Deadline Based Resource Provisioning and Scheduling Algorithm for Scientific Workflows on Clouds," IEEE Transactions on Cloud Computing, vol. 2, no. 2, pp. 222-235, 2014. doi: 10.1109/TCC.2014.2314655
- [60] Rodrigues Da Cunha, G., Calheiros Rodrigo N., Guimaraes Vinicius T., Santos, Glederson L. D., Carvalho Marcio B. D., Granville Lisandro Z., Tarouco, Liane M. R., Buyya Rajkumar, "Monitoring of cloud computing environments: concepts, solutions, trends, and future directions," In: Proceedings of the 31st Annual ACM Symposium on Applied Computing, pp. 378-383, 2016. DOI: <http://dx.doi.org/10.1145/2851613.2851619>
- [61] Samimi Parnia, Teimouri Youness, Mukhtar Muriati, "A combinatorial double auction resource allocation model in cloud computing," ELSEVIER Information Sciences, vol. 357, pp. 201-216, 2016. <http://dx.doi.org/10.1016/j.ins.2014.02.008>
- [62] Serrano Damián, Bouchenak Sara, Kouki Yousri, Oliveira Jr. Frederico Alvares de, Ledoux Thomas, Lejeune Jonathan, Sopena Julien, Arantes Luciana, Sens Pierre, SLA guarantees for cloud services, Future Generation Computer Systems, vol. 54, pp. 233-246, 2016. <https://doi.org/10.1016/j.future.2015.03.018>
- [63] Shrimali Bela, Patel Hireen, "Multi-objective optimization oriented policy for performance and energy efficient resource allocation in Cloud environment," Journal of King Saud University - Computer and Information Sciences, 2017. <https://doi.org/10.1016/j.jksuci.2017.12.001>
- [64] Sindhu S. and Saswati Mukherjee, "Efficient Task Scheduling Algorithm for Cloud Computing Environment," Springer High Performance Architecture and Grid Computing, Comm. In Computer & Information Sci., vol. 169, pp. 79-83, 2011. [https://doi.org/10.1007/978-3-642-22577-2\\_11](https://doi.org/10.1007/978-3-642-22577-2_11)
- [65] Singh Sukhpal, Inderveer Chana, "Q-aware: Quality of service based cloud resource provisioning. ELSEVIER," Computers and Electrical Engg., vol. 42, pp. 138-160, 2015. <http://dx.doi.org/10.1016/j.compeleceng.2015.02.003>
- [66] Singh Sukhpal, Inderveer Chana, "Resource provisioning and scheduling in clouds:QoS perspective," Springer Journal of Supercomputing, 2016. DOI 10.1007/s11227-016-1626-x
- [67] Singh Sukhpal, Inderveer Chana, "QoS-Aware Autonomic Resource Management in Cloud Computing: A Systematic Review," ACM Computing Survey, vol. 48, no. 3, pp. 1-46, 2015. DOI: <http://dx.doi.org/10.1145/2843889>
- [68] Singh Sukhpal, Inderveer Chana, "Cloud resource provisioning: survey, status and future research directions," Knowledge and Information Systems, vol. 49, Issue 3, pp 1005-1069, 2016. doi.org/10.1007/s10115-016-0922-3
- [69] Singh Sukhpal, Inderveer Chana, "A Survey on Resource Scheduling in Cloud Computing: Issues and Challenges," Journal of Grid Computing, vol. 14, Issue 2, pp 217-264, 2016. doi.org/10.1007/s10723-015-9359-2
- [70] Singh Sukhpal, Chana Inderveer, Buyya Rajkumar, "STAR: SLA-aware Autonomic Management of Cloud Resources," in IEEE Transactions on Cloud Computing, vol. pp, no. 99, 2017. doi: 10.1109/TCC.2017.2648788
- [71] Sukhpal Singh Gill, Rajkumar Buyya, Inderveer Chana, Maninder Singh, "Ajith Abraham BULLET: Particle Swarm Optimization Based Scheduling Technique for Provisioned Cloud Resources," Springer J Netw Syst



- Manage, 2017. DOI 10.1007/s10922-017-9419-y
- [72] Singh Harvinder, Bhasin Anshu, "Efficient Resource Management Technique for Performance Improvement in Cloud Computing," *Indian Journal of Comp. Sci. & Engg.* vol. 8, no. 1, pp.33-39.
- [73] [73] Singh Harvinder, Bhasin Anshu, Kaveri Parag R., "QoS based Efficient Resource Allocation and Scheduling in Cloud Computing," *International Journal of Technology and Human Interaction*, IGI Global, vol. 15 (4), pp.13- 29, 2019. DOI: 10.4018/IJTHI.2019100102
- [74] Singh Harvinder, Singh Gurdev, "A Survey Paper on Task Scheduling Methods in Cluster Computing Environment for High Performance," 5th, ACCT, IEEE, pp. 241-246, 2015. DOI: 978-1-4799-8487-9
- [75] Singh Harvinder, Bhasin Anshu, Kaveri Parag R., "SECURE : Efficient resource scheduling by swarm in cloud computing," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 22, no. 2, pp.127-137, 2019. DOI: 10.1080/09720529.2019.1576334
- [76] Singh Aarti, Juneja Dimple, Malhotra Manisha, "A novel agent based autonomous and service composition framework for cost optimization of resource provisioning in cloud computing," *ELSEVIER Journal of King Saud University Comp. and Information Sci.*, vol. 29, no. 1, pp. 19-28, 2017. <http://dx.doi.org/10.1016/j.jksuci.2015.09.001>
- [77] Sivadon Chaisiri, Bu-Sung Lee, Dusit Niyato, "Optimization of Resource Provisioning Cost in Cloud Computing," *IEEE Transactions of Service Computing*, vol. 5, no. 2, pp. 164-177, 2012. DOI 10.1109/TSC.2011.7
- [78] Suprakash S & Balakannan S P., "Service Level Agreement Based Catalogue Management and Resource Provisioning in Cloud for Optimal Resource Utilization," *Springer Journal Mobile Netw Appl*, vol. 24, pp. 1853-1861, 2019. doi:10.1007/s11036-019-01382-9
- [79] Tafsiri Seyedeh Aso, Yousefi Saleh, "Combinatorial double auction-based resource allocation mechanism in cloud computing market," *Journal of Systems and Software*, vol. 137, pp. 322-334, 2018. doi.org/10.1016/j.jss.2017.11.044
- [80] Thein Thandar, Myo Myint Myat, Parvin Sazia, Gawanmeh Amjad, "Reinforcement learning based methodology for energy-efficient resource allocation in cloud data centers," *Journal of King Saud University - Computer and Information Sciences*, 2018. doi.org/10.1016/j.jksuci.2018.11.005.
- [81] Tinghuai Ma, Ya-Chu, Licheng Zhao, Otgonbayar Ankhbayar, "Resource Allocation and Scheduling in Cloud Computing: Policy and Algorithm," *IETE Technical Review*, Taylor & Francis, vol. 31, no. 1, pp. 4-16, 2014. <http://dx.doi.org/10.1080/02564602.2014.890837>
- [82] Tsai Jinn-Tsong, Fang Jia-Cen, Chou Jyh-Horng, "Optimized task scheduling and resource allocation on cloud computing environment using improved differential evolution algorithm," *ELSEVIER Computers & Operations Research*, vol. 40, no. 12, pp. 3045-3055, 2013. <http://dx.doi.org/10.1016/j.cor.2013.06.012>
- [83] Tsai Chun-Wei, Wei-Cheng Huang, Meng-Hsiu Chiang, Ming-Chao Chiang, and Chu-Sing Yang, "A Hyper-Heuristic Scheduling Algorithm for Cloud," in *IEEE Transactions on Cloud Computing*, vol. 2, no. 2, pp. 236-250, 2014. doi: 10.1109/TCC.2014.2315797
- [84] Vakilinia Shahin, Ali Mustafa Mehmet, Qiu Dongyu, "Modeling of the resource allocation in cloud computing centers," *ELSEVIER Computer Networks*, vol. 91, pp. 453-470, 2015. <http://dx.doi.org/10.1016/j.comnet.2015.08.030>
- [85] Vinothina V., Sridharan R., Padmavathi Ganpathi G., "A Survey on Resource Allocation Strategies in Cloud Computing," *International Journal of Advanced Comp. Sci. and App.*, vol. 3, no. 6, pp. 9-102, 2012.
- [86] Verma Manish, Gangadharan G.R., Narendra Nanjangud C., Vadlamani Ravi, Inamdar Vidyadhar, Ramachandran Lakshmi, Calheiros Rodrigo N., Buyya Rajkumar, "Dynamic resource demand prediction and allocation in multi-tenant service clouds," *Wiley Concurrency and Computation: Practice and Experience*, 2016. DOI: 10.1002/cpe.3676
- [87] Wang Tao, Xu Jiwei, Zhang Wenbo, Gu Zeyu, Zhong Hua, "Self-adaptive cloud monitoring with online anomaly detection," *Future Generation Computer Systems*, vol. 80, pp. 89-101. doi.org/10.1016/j.future.2017.09.067A
- [88] Ward Jonathan Stuart, Barker Adam, "Observing the clouds: a survey and taxonomy of cloud monitoring," *Springer Journal of Cloud Computing*, vol. 3, no. 24, 2014. <https://doi.org/10.1186/s13677-014-0024-2>
- [89] Warneke Daniel and Kao Odej, "Exploiting Dynamic Resource Allocation for Efficient Parallel Data Processing in the Cloud," *IEEE Transaction on Parallel and Distributed Systems*, vol. 22 (6), pp. 985-997, 2011. DOI 10.1109/TPDS.2011.65
- [90] Wu Linlin, Garg Saurabh Kumar, Buyya Rajkumar, "SLA-based Resource Allocation for Software as a Service Provider (SaaS) in Cloud Computing Environment," 11th IEEE/ACM Int'l Symposium on Cluster, Cloud and Grid Computing, pp. 195-204, 2011. DOI 10.1109/CCGrid.2011.51
- [91] Xiao Zhen, Song Weijia and Chen Qi, "Dynamic Resource Allocation Using Virtual Machines for Cloud Computing Environment," *IEEE Tran. Parallel and Distributed*, vol. 24, no. 6, pp. 1107-1116, 2013. DOI 10.1109/TPDS.2012.283.
- [92] Xue Jing, Li Liutao, Zhao SaiSai and Jiao Litao, "A Study of Task Scheduling Based on Differential Evolution Algorithm in Cloud Computing," 6th IEEE Int'l Conf. Computational Intelligence and Communication Network, pp. 637-640, 2014. DOI 10.1109/CICN.2014.142
- [93] Xiao-Fang Liu, Zhi-Hui Zhan, Jeremiah D. Deng, Yun Li, Tianlong Gu, Jun Zhang, "An Energy Efficient Ant Colony System for Virtual Machine Placement in Cloud Computing," *IEEE Transactions on Evolutionary Computation*, vol. 22, no. 1, pp. 113-128, 2016. doi: 10.1109/TEVC.2016.2623803
- [94] Zhan Zhi-Hui, Liu Xiao-Fang, Gong Yue-Jiao, Zhang Jun, Chung Henry Shu-hung, Li Yun, "Cloud Computing Resource Scheduling and a Survey of Its Evolutionary Approaches," *ACM Computing Survey*, vol. 47, no. 4, pp. 1-33, 2015. DOI: <http://dx.doi.org/10.1145/2788397>
- [95] Zhang Jiangtao, Huang Hejiao, Wang Xuan, "Resource provision algorithms in cloud computing: A survey,"

ELSEVIER Journal of Network and Computer Applications, vol. 64, pp. 23-42, 2016.  
<http://dx.doi.org/10.1016/j.jnca.2015.12.018>

- [96] Zuo Xingquan, Zhang Guoxiang and Tan Wei, "Self-Adaptive Learning PSO-Based Deadline Constrained Task Scheduling for Hybrid IaaS Cloud," IEEE Transaction Automation Science and Engineering, vol. 11, no.2, pp. 564-573, 2014. 10.1109/TASE.2013.2272758
- [97] Zhang Qi, Zhani Mohamed Faeth, Yang Yuke, Boutaba Raouf and Wong Bernard, "PRISM- Fine-Grained Resource-Aware Scheduling for MapReduce," IEEE Trans. Cloud Computing, vol. 3, no. 2, pp. 182-194, 2015. DOI 10.1109/TCC.2014.2379096
- [98] Zhou Jing & Dong Shoubin, "Hybrid glowworm swarm optimization for task scheduling in the cloud environment. Engineering Optimization," vol. 50, no. 6, pp. 949-964, 2017. DOI: 10.1080/0305215X.2017.1361418