Advances in Science and Technology Research Journal
Volume 10, No. 30, June 2016, pages 38–50
DOI: 10.12913/22998624/62746

Review Article

RESOURCE SCHEDULING IN CLOUD ENVIRONMET: A SURVEY

Neeraj Mangla[1], Manpreet Singh[1], Sanjeev Kumar Rana[1]

[1] Computer Engineering Department, Engineering College, Maharishi Markandeshwar University Mullana, 133207 Ambala, India, e-mail: erneerajynr@gmail.com

Received: 2016.04.15
Accepted: 2016.05.10
Published: 2016.06.01

ABSTRACT

Cloud Computing offers the avant-garde services at a stretch that are too attractive for any cloud user to ignore. With its growing application and popularization, IT companies are rapidly deploying distributed data centers globally, posing numerous challenges in terms of scheduling of resources under different administrative domains. This perspective brings out certain vital factors for efficient scheduling of resources providing a wide genre of characteristics, diversity in context of level of service agreements and that too with user-contingent elasticity. In this paper, a comprehensive survey of research related to various aspects of cloud resource scheduling is provided. A comparative analysis of various resource scheduling techniques focusing on key performance parameters like Energy efficiency, Virtual Machine allocation and migration, Cost-effectiveness and Service-Level Agreement is also presented.

Keywords: resource scheduling, energy conservation, cloud computing, virtualization, service-level agreement.

INTRODUCTION

Cloud computing is the new cost-efficient computing standard that delivers on-demand access to services on pay-per-usage basis [1, 2]. The unwavering services offered by cloud computing are realized through its innovative global data centers that are firmed on virtualized compute and storage technologies [3]. It is intended for the cloud users to multifold the prospects by accessing leased infrastructure and software applications ubiquitously and unrestrictive in time [4]. The ideology is firmed on the grounds of 're-usability of IT capabilities'. The traditional computing archetypes become outmoded by cloud computing due to its expansive horizons across organizational boundaries [5].

The cloud computing paradigm offers numerous benefits to both cloud customers and service providers. The aim of service provider is to maximize the profit by efficient usage of its datacenter resources through virtualization technology [6] and effective scheduling within the constraint of Service-Level Agreement [7] with cloud users and limited power budget [8, 9]. From the perspective of cloud's user, the focus is on application performance, availability of services, cost-effectiveness [10, 11] and adaptability to the changing requirements.

The resource scheduling in cloud environment is always a complicated task due to geographical distribution of resources having varying load conditions, different user's requirements and price models [12]. A lot of research work dealing with the cloud resource scheduling problem has been carved by many researchers [13, 14, 35, 36, 58, 59, 71, 72]. This paper provides a detailed survey of the prevailing resource scheduling techniques focusing on the promising features and challenges of cloud computing.

LITERATURE SELECTION PROCESS

In order to select the relevant papers, the following keywords were used on a number of search databases (IEEE, ACM, Elsevier, Springer, and Google Scholar):

- Virtual machine allocation,
- Virtual machine migration,
- Virtualization,
- Cloud resource scheduling,
- SLA-aware resource scheduling,
- Profit maximization,
- Cost-effective scheduling,
- Energy-aware scheduling.

In total, 278 potential papers along with abstracts published within the period from 2007 to 2016 were collected. Finally, 93 full length most relevant papers for this literature survey was selected and analyzed. The year-wise and publisher-wise bifurcation of selected papers is presented in Figure 1 and Figure 2.

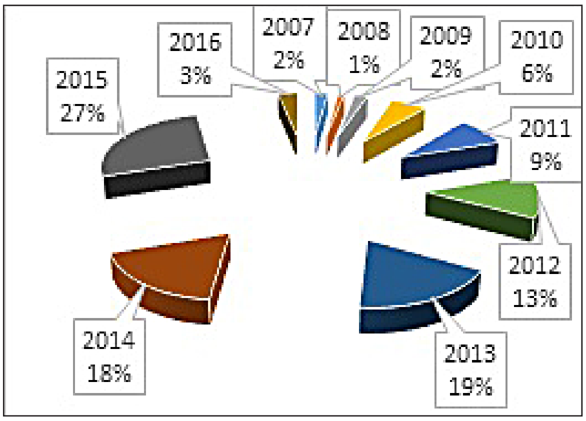

Fig. 1. The year-wise bifurcation of selected papers

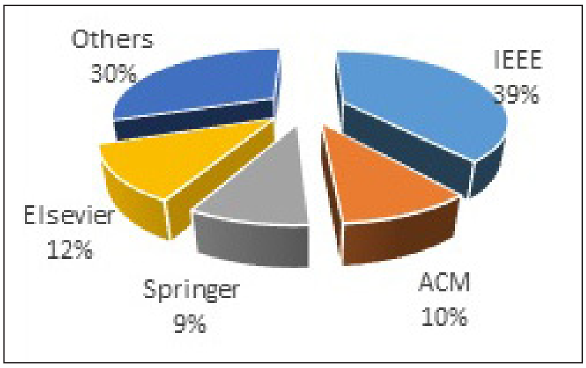

Fig. 2. The publisher-wise bifurcation of selected papers

LITERATURE REVIEW

Efficient resource provisioning with focus on satisfying user service requirements by taking into account both the economic and environmental viability, lays the foundation for the success of commercial competition [22, 47, 65, 78]. In the subsequent sections, a classification of the resource scheduling problem in cloud environment as shown in Figure 3 and its existing on hand solutions have been reviewed in brief.

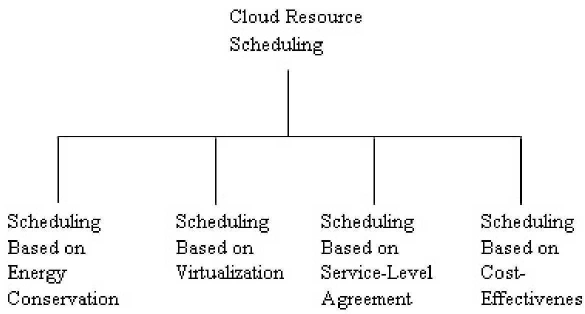

Fig. 3. Classification of the resource scheduling problem in cloud environment

Scheduling based on energy conservation

With the ever-increasing needs of Cloud infrastructure, the energy consumption of datacenters has increased dramatically raising the apprehension for both government and service providers to consume energy effectively [47, 51]. Assessment of literature is indicative of the fact that servers in many datacenters usually operate at 30% of their full capacity [57]. Incorporating energy reduction activities during the phases of low utilization is a centre of attraction and thus is the target of most of the research related to energy saving in cloud.

An enabling VM scheduling policy is presented in [37] to reduce the energy-losses caused by I/O virtualization mechanism while handling intensive mixed-workloads in particular. In [39], authors proposed a philosophical architecture for energy management of Clouds. The energy-oriented resource allocation policies and scheduling approaches are also discussed which considers the two important characteristics: QoS expectations and power usage of resources. An enterprise data center-applicable framework has been formulated in [54] that incorporate a computation proficient heuristic design which offers speedy placement solutions sensitive to the workload. In [53], the authors intends to highlight the ambiguity of the computing environment and a scheduling design to optimize the influence of uncertainty on the task scheduling capability for a cloud data center. This framework has been utilized by the authors to implement a novel scheduling algorithm that vigorously gain benefit of the proactive and reactive scheduling methods, for scheduling real-time, aperiodic, independent tasks.

Table 1. Comparison of various energy conservation based resource scheduling techniques

Technique	Performance metrics	Environment	Results	Refer-red work
Energy based Efficient Resource Scheduling Algorithm	Energy efficiency	CloudSim	Proposed algorithm (23.39%) reduces more energy than EEUR (18.71%) [55] during migrations.	[36]
VM consolidation based on utilization, network topologies and thermal state of servers	Energy efficiency	CloudSim	Proposed heuristic reduces energy by 83% and 66% as compared to a non-power aware system and a system based on DVFS technique.	[38]
Energy aware scheduling algorithm based on processor and disk workloads	Energy efficiency, Execution time	Simulation (MATLAB)	Proposed algorithm shows 24.9% power savings and 1.2% performance degradation in comparison with other scheduling approaches.	[40]
Thermal aware resource allocation technique	Energy efficiency, Cost	Simulation (Environment not mentioned)	17% improvement in terms of maximizing reward and 9% reduction in power consumption is reported in proposed technique.	[41]
Heuristic based on performance-per-watt for VM placement	Energy efficiency	CloudSim	35 % of total energy saving in proposed heuristic in comparison with the traditional allocation heuristics.	[42]
Power-aware VM allocation heuristic	Energy efficiency	CloudSim	Proposed heuristics results in energy saving of 22.4% and 16.0% when simulated with power--aware best-fit decreasing and vector bin-packing norm-based greedy algorithms.	[43]
Reinforcement Learning--based VM Consolidation method		CloudSim	Energy saving of 12.5%,19.4%, 22.6% and 28.5% can be attained in comparison with LR, MAD, THR and IQR [56] in the real workload. SLA violation rate is less in proposed solution than other techniques.	[44]
Scheduling strategy capable of providing virtual clusters that reactively focus on energy optimization to perform consolidation	Energy efficiency, SLA violation rate	Simulation (Environment not mentioned)	Energy saving in proposed algorithm over OBFIT is 23.6%, 16.9% and 72.4% for the average task length ratios of 0.01, 0.1 and 1, respectively. The working efficiency of proposed algorithm is improved over OBFIT as 26.2%, 20.3% and 219.7% for the average task length ratios of 0.01, 0.1 and 1, respectively.	[45]
A PSO based energy-aware VM allocation algorithm	Energy efficiency, Resource utilization	CloudSim	Energy savings and utilization of datacenter resources in proposed algorithm are significant at the same time.	[46]
A novel heuristic based on multi-criteria decision making method for determining the lightly loaded nodes and nodes for VM consolidation	Energy efficiency, SLA violation rate, Number of VM migration.	CloudSim	The proposed approach shows upto 46%, 99%, and 95% reduction in energy used, deadline violations, and VM migrations respectively as compared with similar heuristics.	[48]
VM management architecture for Snooze (a private cloud)	Energy efficiency	Snooze	Result confirmed the effectiveness of proposed solution as energy conservation of upto 67% is achieved on a realistic workload.	[49]
An improved PSO for optimal VM placement	Energy efficiency	Java based simulator	Proposed approach significantly reduces energy by 13–23% as compared to other heuristic approaches	[50]

In [52], a predictive design which aims to combine the machine learning clustering and stochastic theory to estimate both the number of VM requests and the amount of cloud resources associated with each request is formulated. An amalgamated resource conditioning framework depending upon this method has been used by the authors to make suitable energy-aware resource supervised decisions which is further evaluated using real Google traces collected over a 29-day period from a Google cluster containing over 12,500 PMs. Comparison of various energy conservation based resource scheduling techniques is presented in Table 1.

Scheduling based on virtualization

Virtualization technology allows different applications to be allocated on the single Physical Machine (PM) in logically secluded VMs. The use of virtualization technique permits the migration of live Virtual Machines and their consolidation on lesser number of PMs resulting in high utilization of the available physical resources, reducing the energy consumption and capital cost associated with the cloud datacenter [79,85]. Virtual Machine allocation and migration in cloud environment is a challenging task. The concept behind VM allocation is the mapping between VM and PM with an objective to maximize application performance, energy saving, or augment the provider's revenue [80].

In [86], authors presented a Virtual Computing Laboratory framework model using the concept of private cloud by extending the open source IaaS solution Eucalyptus. A mapping algorithm for VMs based on rules and the principles of set theoretic is also presented. The algorithmic design is projected towards being able to autonomic plotting between VMs and datacenter resources. A system based on virtualization for the allocation of data center resources dynamically on the basis of demands of the application is presented [87]. In parallel, the optimized number of servers henceforth supports the green computing. The concept of "skewness" is put forward to determine the non-proportionality in the multi-dimensional resource utilization of a server. It is also shown that different types of workloads can be combined efficiently and overall utilization of server resources is improved upon minimization of skewness. A group of heuristics is also developed that is able to effectively save the energy while avoiding the system overloading. Efficiency of this algorithm was adjudged through the trace driven simulation and henceforth the results of experiments. In [88], a combination of ant colony optimization (ACO) and VM dynamic forecast scheduling (VM_DFS) to perform VM scheduling is presented. In this algorithm through analysis of historical memory consumption in each PM, future memory consumption forecast of VMs and their allocation on the cloud resources is performed. This methodology is experimented in MATLAB for both homogeneous and heterogeneous mode and results indicate that the proposed algorithm produces lower resource wastage than other traditional approaches and better load balancing among PMs.

Virtual Machine migration is of utmost importance in implementing resource management strategies for the optimization of performance metrics such as consumption of energy, utilization of resource and QoS. The primer challenge for VM migration in terms of service downtime and high network utilization is discussed [84]. In [81], the authors put forward a novel model for optimization which laid its basis on linear programming along with an automatic approach for VM migration in self-managing virtualized environments. The experimentation yielded the applicability of this approach to effectively determine which virtual machines should be migrated and on which physical machines to host them while minimizing operational and migration costs. In [82], authors proposed a Linear Programming formulation and heuristics to control VM migration that gives priority to the virtual machines with steady capacity. In order to draw comparison between this migration-control approaches with the well-established eager-migration-based solutions, the simulations are implemented using TU-Berlin and Google data center workloads. The results confirmed the reduction in the number of migrations with minimal penalty in the number of physical servers, if the migration of VMs with steady capacity is avoided. Basic outline for cloud brokering and multi-cloud VM administration is proposed [83]. Descriptive algorithms were also proposed by the authors that paved way for the efficient placement of applications in multi-cloud environments. Incorporation of price and performance along with constraints such as hardware configuration, load balancing is also incorporated in the placement model. In contrast to single-cloud deployment, the proposed multi-cloud placement algorithms yielded improvement in performance with lower costs during an evaluation against commercial clouds. An attempt to decide the appropriate time for VM migration has been made by the authors [84] in which they have articulated an application-oriented live migration model exploring the application level information, in conjunction to the state-of-art system level metric. The experiment was conducted on three real applications with due consideration to the application specific to catalyze the VM live migration. It depicted a significant drop in the network overhead up to 42% and decrease in live migration time up to 63%. Table 2 elucidates a comparative study of various virtualization based resource scheduling techniques.

Scheduling based on service-level agreement

The cloud service providers and their customers have to negotiate a Service Level Agreement (SLA), which basically outlines the service requirements and the assurance in the delivery of a service. Violation of the SLA is a key issue as it

Table 2. Comparative study of various virtualization based resource scheduling techniques

Technique	Performance metrics	Environment	Results	Referred work
Virtual machine allocation is based on random allocation strategy, sequence allocation strategy and greedy allocation strategy	Makespan	CloudSim	Makespan of the greedy strategy is least, largest for random strategy and for sequence strategies it is moderate	[89]
Autonomous synchroni-zation--aware VM Sched-uling (SVS) algorithm. Integration of proposed algorithm into Xen VMM scheduler	Normalized exe-cution time, CPU fairness	Real cluster envi-ronment with NPB benchmark and real-world trace	Proposed solution yield better results for tightly-coupled parallel applications when compared with their execution on Xen's Credit scheduler, balance scheduler, and hybrid scheduler	[72]
Architecture aware Dynamic Resource Man-age-ment Scheme (A-DRM) with emphasis on the effect of micro architecture-level interference on VM migration	Number of VM migration, Me-mory bandwidth utilization	A cluster of four homogeneous NUMA servers and a Network-Attached Storage(NAS) using KVM and QEMU platform	A-DRM can enhance the performan-ce of virtual machines and average cluster-wide memory bandwidth utilization by up to 26.55% and 17% respectively as compared to a traditio-nal DRM scheme	[73]
Genetic Algorithm	Resource utili-zation, Energy efficiency	CloudSim	Proposed algorithm save energy by approximately13% in comparison to the baseline scheduling algorithm	[90]
Priority based allocation of jobs to VMs	Application exe-cution time	CloudSim	Less overhead in executing all submitted jobs, when compared with creation of new VM	[91]
VM management based on workload requirements, existing resource capacity, and defined provisioning policies	Throughput, Execution time	Harmony grid envi-ronment	Proposed solution is providing up to 20% more services than the heuristics that are able to satisfy the same requirements	[74]
Linear programming models are proposed for VM allocation	Total execution time	COIN-OR CBC solver	Empirical results exemplify the perfor-mance of the three proposed models	[75]
Scheduling based on VM confi-guration	Makespan, Cost	CloudSim	The results highlights the factors affecting VM configuration in order to handle user and provider preferences	[76]
Design and development of the autonomic and decentralized mechanisms for Dynamic Virtual Machine (DVM) management	SLA violation rate, Energy efficiency	C#.NET to develop an event--driven simulation environment	DVM strategy achieves less power consumption, leading to more than 20% energy savings	[77]
A Mixed Integer Programming (MIP) formulation of the application placement problem	Resource utili-zation	Java based simu-lator	Utilization rate of round-robin is between 73.8% and 90.2% while utili-zation rate of the proposed heuristics is between 80.5% and 96.4%	[78]
An approach named GraspCC--festo produce the optimal esti-mation of the amount of virtual machines to allocate for each workflow.	Execution time	ANSI C based simulator	The execution of the workflow fragment using the amount of virtual machines following the allocation given b GraspCC was about 64% smaller than execution that followed the configuration provided by SciDim	[93]

tends to make customers malcontent and eventually their level of satisfaction declines.

In [2], authors proposed the following to reduce the deadline miss rate: (1) design and modeling of the Analytics as a Service (AaaS) platform; (2) formulation of the resource scheduling problem on the basis of mixed Integer Linear Programming (ILP) model; (3) an algorithm to ensure admission control and resource scheduling. In [68], author's contribution is: (1) a novel domain specific language that is able to illustrate the QoS-oriented SLA associated with cloud services; (2) a general control-theoretic approach for managing cloud service SLA and; (3) apply the proposed language and control approach to guarantee SLA in various cases studies, ranging from cloud-based MapReduce service, to locking service, and higher-level e-commerce service. With the objective of reducing the effects of SLA violations to offer a phenomenal proactive resource allocation approach, authors in [69], aimed on two user's hidden characteristics viz.-a-viz. willingness to pay for service and willingness to pay for certainty. In addition, holistic approaches based on learning automaton for the approximation of these characteristics are also instilled. An OpenStack version of the Generic SLA Manager, besides its methodical approaches for VM selection and allocation during live migration of VMs is presented [70].A use case is simulated, where IaaS (OpenStack-SLAM) and PaaS (OpenShift) are coalesced to evaluate the performance and effectiveness of the proposed VM placement strategies, in case of multi-domain SLA pricing & penalty model. Comparison of various SLA based resource scheduling techniques is presented in Table 3.

Scheduling based on cost-effectiveness

In cloud, service providers want to minimize resource rental costs while still meeting workload demands and cloud users look forward to the lowest possible prices for the resources they lease. It is worthwhile to mention that optimized VM placement [76], reducing application makespan [16], dynamic resource renting schemes [26] can satisfy the monetary requirements of both providers and users.

Table 3. Comparison of various SLA based resource scheduling techniques

Technique	Performance metrics	Environment	Results	Referred work
Automated SLA management comprising negotiation and provisioning	SLA violation rate, Energy efficiency, Cost	CloudSim	In simulation scenario, an additional CPU power in terms of 100 MIPS lead to the optimal result: the SLA average violation rate decreases to 10.5%, de-crease of ROI (0.68%), increase of energy con-sumption (4.09%).	[58]
SLA-aware and profitability oriented scheduling of cloud services	SLA violation rate, Cost	Simulation (Environment not mentioned)	Heuristic cost-based scheduling show significant improvement over FCFS and SJF in terms of cost and deadline enforcement	[59]
Amalgamation of market oriented provisioning policies and virtualization techniques to provide elasticity of resources	SLA violation rate	Aneka platform, Amazon EC2	Aneka due to its capability of allocating resources dynamically seems to be a good option for satisfying application QoS requirements	[60]
A negotiation based Adaptive Scheduler	SLA violation rate	Hadoop	The investigative study demonstrates the benefits of the proposed scheduler over existing schedulers in Hadoop in terms of Resources Availability, Priority Basis Allocation, Uniform Resource Distribution, Guaranteed Service and Negotiation	[61]
Aa novel SLA-driven architecture based in the WS-Agreement specification for the automatic provision, scheduling, allocation and dynamic management of Cloud resources	SLA violation rate, Cost	OpenNebula, jLinpack service as a Virtual Service	Proposed solution provides minimum cost and maximum efficiency under heterogeneous environment	[62]
VM scheduling focusing on price, service initiation time and data transfer time	SLA violation rate, Cost	CloudSim	Proposed algorithm re-duces cost around 50% by making use of only 60% of VMs in number SLA violation rate in pro-posed algorithm when the service initiation time is 'very short' and 'very long' is less than 13%	[63]
SLA aware and capacity planning solution for Service Consolidation in OpenShift PaaS.	SLA violation rate, Energy effi--ciency	Simulation (Environment not mentioned)	Simulated Annealing comes out to be the maxi-mum yielding algorithm for implementing most of the policies	[64]
Online energy-aware resource scheduling framework	SLA violation rate, Energy effi--ciency	Simulation (Environment not mentioned)	Energy consumption and SLA violation rate of proposed solution is reduced by 21% and 16% than the genetic algorithm in average	[65]
An autonomic computing based multi-tier architecture to handle fluctuating workloads	SLA violation rate, Cost	Simulation MATLAB	Proposed dynamic alloca-tion method attains 15–20% higher total profit than other Static and Dynamic resources allo-cation strategy	[66]
Automatic service selection for multi-cloud environment taking into account the SLA claims of SaaS providers	Service ranking	Java based simulator	Automatic service selection for multi-cloud environ-ment taking into account the SLA claims of SaaS providers Service ranking Java based simulator Simulation-based eva-lua-tion and a comparison with a utility-based match-ing algorithm shows the effectiveness of pro-posed approach in select-ing a set of services satis-fying SLA parameters	[67]

A cost-aware runtime strategy to store the datasets of the generated application in the cloud as a decision support system by making an estimate of the trade-off between computation, storage and bandwidth is proposed [30]. In [31], authors formalized an optimization framework for MapReduce over multi-cloud including virtual machine and data transfer costs. In addition, a decentralized resource management middleware that considers multi-optimization is designed. In [32], authors proposed an innovative methodology to manage virtualized data centers according to multiple facets (energy efficiency, virtualization overheads, and SLA violation penalties) when placing VMs in data center nodes and maximizing the provider's profit. The experiments demonstrated that the proposed model is capable of increasing the provider's revenue by 30% and can handle certain issues including resource heterogeneity. The outline of scheduling prioritized workflow ensembles under budget and deadline constraints is discussed [33]. The authors developed a genre of dynamic and static algorithms to schedule tasks and resource provisioning that depends upon the workflow structure information and estimates of task runtimes. The proposed algorithms are then simulated using a simulator built on top of CloudSim, which designs the infrastructure and the application, keeping under consideration the uncertainties in task runtime estimates, provisioning delays and failures. In [34], authors devised the following approaches in order to minimize the rental costs while meeting users' computing needs: (1) an online algorithm called Online Cost-efficient Scheduling (OCS) that makes use of a priority function to determine the urgency of requests, where the value of the priority function is based on the expected VM speed to complete a request, and assigns faster VM instances to requests with higher priorities; (2) an algorithm called Dynamic Resource Planning (DRP) to terminate unneeded VM instances before the next billing cycle starts based on prediction of daily patterns of resource usage from interactive services using queuing analysis; (3) an algorithm called Cost-conscious Scheduling algorithm (CCS) to dispatch batch jobs. The core of CCS is workload partitioning, which splits batch jobs across a large number of time slots according to the remaining resource capacity and spot instance pricing. Comparison of various Cost-effective cloud resource scheduling techniques is presented in Table 4.

Table 4. Comparison of various Cost-effective cloud resource scheduling techniques

Technique	Performance metrics	Environment	Results	Referred work
A technique to identify the potential data sets that need to be encrypted for handling the privacy requirements of customers	Cost	U-Cloud with KVM virtualization software, OpenStack and Hadoop	The privacy-keeping cost of potential data sets in proposed approach is at least 40% less than the cost incurred in encrypting all data sets for privacy preserving	[13]
An architecture to enable coordinated dynamic provisioning of public Cloud resources and scheduling of deadline-constrained applications In addition, a novel approach for billing users for the utilization of public Cloud resources	Cost, Resource Utilization	CloudSim	the proposed strategy can reduce the total utilization of public Cloud services by up to 20% without any impact in the capacity of meeting ap-plication deadlines	[15]
A PSO based scheduling of workflows on IaaS clouds	Cost, Makespan, Deadline con-straint evalua-tion	CloudSim and four different workflows: Montage, LIGO, SIPHT and Cyber-Shake	The proposed solution exhibit better perfor-mance than the other similar approaches in terms of meeting application's deadline and gen-erating schedules with lower execution cost	[16]
A holistic double renting scheme for service providers	Cost	Analytical solution	The results show that proposed scheme outperforms the Single-Quality-Unguaranteed (SQU) renting scheme in terms of both the quality of service offered and profit	[17]

Table 4. Continued

PSO based model for reso-urce scheduling	Cost	JSwarm package	Proposed solution is capable of saving cost at least three times as compared to "Best Resource Selection" (BRS) heuristic	[27]
A QoS-based selling me-chanism for batch jobs in a multi-tenant OpenStack cluster	Cost	OpenStack Sahara environment	Revenue of cloud providers in proposed mechanism can be increased by 40% as compared to a fixed, per-node-period pricing method	[18]
PSO based model for task--resource mapping	Cost, Execution time		Results show that at high load, propo-sed meta-heuristic converges faster and optimize better than crossover and muta--tion PSO (CM-PSO) and PSO algorithm embed in local search (L-PSO) The running time of the L-PSO is almost three times and the CM-PSO is nearly two time as that of proposed meta-heuristic at high load	[19]
Implementing VM placement as a Multi-level Generalized Assignment Problem using first-fit heuristic	Cost, SLA violation rate, Energy effi-ciency		Profit loss within 15% in worst scenario, and the within 10% for power aware system SLA violation rate is under 15% in the worst scenario under all workload conditions	[20]
Improved GA to handle data-intensive services	Cost, Execution time	MATLAB	Proposed GA shows gradual increases in execution time as compared to Mixed Integer Programming approach with the increase in number of concrete services	[21]
Optimizing the objec-tives of cloud users and service providers (IaaS, SaaS) with efficient resource provisio-ning	Cost, Resource utilization	Simulation (Environment not men-tioned)	At job arrival rate=0.4, the proposed algorithm exhibits 16% and 12% more utilization than DRP_exponential Pricing and DRP_linear Pricing [28] respectively At job arrival rate=0.3, resource usage cost is 9% less than DRP_linear pricing	[22]
A framework to formu-late bidding price for user along with an algorithm to compu-te Nash equilibrium solution	Cost	Simulation (Environment not men-tioned)	The obtained near-equilibrium solution is close to the equilibrium one	[23]
A cloud service request model with SLA constraints is established and based on the request model a cost-aware service request scheduling using genetic approach is presented	Cost, Resource Utilization	Java based simulator	The proposed algorithm achieves higher resource utilization (80% on simulated request data set) in comparison with the other three revenue-aware algorithms i.e. BL-small (30%), BL-large (35%) and BL--xlarge (45%) [29] The operational profits of three revenue--aware algorithms is lower than proposed algorithm	[24]
A cost-effective re-source management framework called Cura is designed: • to provide a cost-effective solution to efficiently han-dle MapReduce produc-tion workloads • leverages MapReduce profiling to automatically create the best cluster configuration for the jobs • implements a globally ef-ficient resource allocation scheme	Cost, Execution time	Java based simulator	The experimental results using Facebook-like workload traces show that proposed techniques lead to more than 80% reduction in the cloud compute infrastructure cost with up to 65% reduction in job response times	[25]
A dynamic virtual resource renting method that attempts to dynamically adjust the vir-tual resource rental strategy according to price distribu-tion and task urgency	Cost	Java based simulator	The simulation results show that average rental cost of proposed method is much lower and average profit is the highest among other traditional revenue-aware algorithms	[26]

CONCLUSION

Resource scheduling always remained an active area of research due to world-wide uncontrolled growth of datacenters to cope up with the growing demands of cloud infrastructure. Based on the literature survey, the challenges posed by the cloud environment itself in terms of high power energy consumption of data centers, customer satisfaction, provider profitability etc. are identified. Furthermore, a critical evaluation of the existing on-hand cloud resource scheduling techniques on the basis of selected parameters from the literature is carried out. With a belief, that an up-to-date review of the resource scheduling literature presented in this paper will surely help the researchers and developers in selecting the most appropriate techniques to manage resources in cloud environment under the given constraints.

REFERENCES

1. Demchenko, Y., Van der Ham, J., Yakovenko, V., De Laat, C., Ghijsen, M., & Cristea, M. On-demand provisioning of cloud and grid based infrastructure services for collaborative projects and groups. In: Collaboration Technologies and Systems (CTS), International Conference, IEEE, May 2011, 134–142.
2. Buyya, R., Yeo, C. S., Venugopal, S., Broberg, J., & Brandic, I. Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. Future Generation computer systems, 25 (6), 2009, 599–616.
3. Shawish, A., & Salama, M. Cloud computing: paradigms and technologies. In Inter-cooperative Collective Intelligence: Techniques and Applications (pp. 39-67). Springer Berlin Heidelberg 2014.
4. Whaiduzzaman, M., Gani, A., Anuar, N.B., Shiraz, M., Haque, M.N., & Haque, I.T. Cloud service selection using multicriteria decision analysis. The Scientific World Journal, 2014.
5. Foster, I., Zhao, Y., Raicu, I., & Lu, S. Cloud computing and grid computing 360-degree compared. In: Grid Computing Environments Workshop, IEEE. 2008, November, 1–10.
6. Huang, J., Liu, Y., & Duan, Q. Service provisioning in virtualization-based Cloud computing: Modeling and optimization. In Global Communications Conference (GLOBECOM), IEEE. 2012, December, 1710–1715.
7. Mohamadi, A., & Barani, S. A review on approaches in service level agreement in cloud computing environment. In Fuzzy and Intelligent Systems (CFIS), 4th Iranian Joint Congress, IEEE. 2015, September, 1–4.
8. Okada, T. K., De La Fuente Vigliotti, A., Macedo Batista, D., & Goldman vel Lejbman, A. Consolidation of VMs to improve energy efficiency in cloud computing environments. In Computer Networks and Distributed Systems (SBRC), XXXIII Brazilian Symposium, IEEE. 2015, May, 150–158.
9. Hou, R., Jiang, T., Zhang, L., Qi, P., Dong, J., Wang, H., & Zhang, S. Cost effective data center servers. In High Performance Computer Architecture (HPCA2013), 19th International Symposium, IEEE. 2013, February, 179–187.
10. Gupta, A., & Kalé, L. V. Towards efficient mapping, scheduling, and execution of HPC applications on platforms in cloud. In Parallel and Distributed Processing Symposium Workshops & PhD Forum (IPDPSW), 27th International, IEEE. 2013, May, 2294–2297.
11. Han, R., Ghanem, M. M., Guo, L., Guo, Y., & Osmond, M. Enabling cost-aware and adaptive elasticity of multi-tier cloud applications. Future Generation Computer Systems, 32, 2014, 82–98.
12. Weingärtner, R., Bräscher, G. B., & Westphall, C. B. Cloud resource management: A survey on forecasting and profiling models. Journal of Network and Computer Applications, 47, 2015, 99–106.
13. Zhang, X., Liu, C., Nepal, S., Pandey, S., & Chen, J. A privacy leakage upper bound constraint-based approach for cost-effective privacy preserving of intermediate data sets in cloud. Parallel and Distributed Systems, IEEE Transactions, 24 (6), 2013, 1192–1202.
14. Verma, A., & Kaushal, S. Deadline and budget distribution based cost-time optimization workflow scheduling algorithm for cloud. In IJCA Proceedings on international conference on recent advances and future trends in information technology (iRAFIT 2012, April).
15. Calheiros, R. N., & Buyya, R. Cost-effective provisioning and scheduling of deadline-constrained applications in hybrid clouds. In Web Information Systems Engineering-WISE. Springer Berlin Heidelberg 2012.
16. Rodriguez, M. A., & Buyya, R. Deadline based resource provisioning and scheduling algorithm for scientific workflows on clouds. Cloud Computing, IEEE Transactions on, 2 (2), 2014, 222–235.
17. Mei, J., Li, K., Ouyang, A., & Li, K. A profit maximization scheme with guaranteed quality of service in cloud computing. Computers, IEEE Transactions, 64 (11), 2015, 3064–3078.
18. Sandholm, T., Ward, J., Balestrieri, F., & Huberman, B. A. QoS-Based Pricing and Scheduling of Batch Jobs in OpenStack Clouds. 2015, arXiv pre-

print arXiv:1504.07283.

19. Guo, L., Zhao, S., Shen, S., & Jiang, C. Task scheduling optimization in cloud computing based on heuristic algorithm. Journal of Networks, 7 (3), 2012, 547–553.
20. Shi, W., & Hong, B. Towards profitable virtual machine placement in the data center. In Utility and Cloud Computing (UCC), Fourth IEEE International Conference, 2011, December, 138–145.
21. Wang, L., Shen, J., Luo, J., & Dong, F. An improved genetic algorithm for cost-effective data-intensive service composition. In Semantics, Knowledge and Grids (SKG), Ninth International Conference, IEEE. 2013, October, 105–112.
22. Li, C., & Li, L. Efficient resource allocation for optimizing objectives of cloud users, IaaS provider and SaaS provider in cloud environment. The Journal of Supercomputing, 65 (2), 2013, 866–885.
23. Li, K., Liu, C., Li, K., & Zomaya, A. A Framework of Price Bidding Configurations for Resource Usage in Cloud Computing, 2015.
24. Liu, Z., Wang, S., Sun, Q., Zou, H., & Yang, F. Cost-aware cloud service request scheduling for SaaS providers. The Computer Journal 2013, bxt009.
25. Palanisamy, B., Singh, A., & Liu, L. Cost-effective resource provisioning for mapreduce in a cloud. Parallel and Distributed Systems, IEEE Transactions, 26 (5), 2015, 1265–1279.
26. Zhou, A., Wang, S., Sun, Q., Zou, H., & Yang, F. Dynamic Virtual Resource Renting Method for Maximizing the Profits of a Cloud Service Provider in a Dynamic Pricing Model. In Parallel and Distributed Systems (ICPADS), International Conference, IEEE. 2013, December, 118–125.
27. Pandey, S., Wu, L., Guru, S.M., & Buyya, R.A particle swarm optimization-based heuristic for scheduling workflow applications in cloud computing environments. In Advanced information networking and applications (AINA), 24th IEEE international conference, 2010, April, 400–407.
28. Zhu, Q., & Agrawal, G. Resource provisioning with budget constraints for adaptive applications in cloud environments. In Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing, ACM. 2010, June, 304–307.
29. Chen, J., Wang, C., Zhou, B. B., Sun, L., Lee, Y. C., & Zomaya, A. Y. Tradeoffs between profit and customer satisfaction for service provisioning in the cloud. In Proceedings of the 20th international symposium on High performance distributed computing, ACM. 2011, June, 229–238.
30. Yuan, D., Cui, L., Liu, X., Fu, E., & Yang, Y. A Cost-Effective Strategy for Storing Scientific Datasets with Multiple Service Providers in the Cloud, 2016. arXiv preprint arXiv:1601.07028.
31. Imai, S., Patterson, S., & Varela, C. A. Cost-Efficient High-Performance Internet-Scale Data Analytics over Multi-Cloud Environments. In Cluster, Cloud and Grid Computing (CCGrid), 15th IEEE/ACM International Symposium, 2015, May, 793–796.
32. Goiri, Í., Berral, J. L., Fitó, J. O., Julià, F., Nou, R., Guitart, J., Torres, J. Energy-efficient and multifaceted resource management for profit-driven virtualized data centers. Future Generation Computer Systems, 28 (5), 2012, 718–731.
33. Li, J., Su, S., Cheng, X., Song, M., Ma, L., & Wang, J. Cost-efficient coordinated scheduling for leasing cloud resources on hybrid workloads. Parallel Computing, 44, 2015, 1–17.
34. Malawski, M., Juve, G., Deelman, E., & Nabrzyski, J. Algorithms for cost-and deadline-constrained provisioning for scientific workflow ensembles in IaaS clouds. Future Generation Computer Systems, 48, 2015, 1–18.
35. Videv, S., & Haas, H. Energy-efficient scheduling and bandwidth-energy efficiency trade-off with low load. In Communications (ICC), IEEE International Conference, 2011, June, 1–5.
36. Singh, S., & Chana, I. Energy based efficient resource scheduling: a step towards green computing. Int J Energy Inf Commun, 5 (2), 2014, 35–52.
37. Liu, D., & Han, N. An Energy-efficient Task Scheduler in Virtualized Cloud Platforms. International Journal of Grid and Distributed Computing, 7 (3), 2014, 123–134.
38. Beloglazov, A., & Buyya, R. Energy efficient resource management in virtualized cloud data centers. In Proceedings of the 10th IEEE/ACM international conference on cluster, cloud and grid computing, 2010, May, 826–831.
39. Buyya, R., Beloglazov, A., & Abawajy, J. Energy-efficient management of data center resources for cloud computing: a vision, architectural elements, and open challenges, 2010. arXiv preprint arXiv:1006.0308.
40. Sharifi M, Salimi H, Najafzadeh M. Power-efficient distributed scheduling of virtual machines using workload-aware consolidation techniques. The Journal of Supercomputing, Springer 2012.
41. Al-Qawasmeh, A.M., Pasricha, S., Maciejewski, A.A., & Siegel, H.J. Power and thermal-aware workload allocation in heterogeneous data centers. Computers, IEEE Transactions, 64 (2), 2015, 477–491.
42. Quang-Hung, N., Thoai, N., & Son, N. T. EPOBF: energy efficient allocation of virtual machines in high performance computing cloud. In: Transactions on Large-Scale Data-and Knowledge-Cen-

tered Systems XVI, 2014, 71–86.

43. Quang-Hung, N., Le, D. K., Thoai, N., & Son, N. T. Heuristics for energy-aware vm allocation in hpc clouds. In: Future Data and Security Engineering, Springer International Publishing 2014, 248–261.
44. Farahnakian, F., Liljeberg, P., & Plosila, J. Energy-efficient virtual machines consolidation in cloud data centers using reinforcement learning. In: Parallel, Distributed and Network-Based Processing (PDP), 2014 22nd Euromicro International Conference, IEEE. 2014, February, 500–507.
45. Sampaio, A. M., & Barbosa, J. G. Towards high-available and energy-efficient virtual computing environments in the cloud. Future Generation Computer Systems, 40, 2014, 30–43.
46. Xiong, A. P., & Xu, C. X. Energy efficient multiresource allocation of virtual machine based on PSO in cloud data center. Mathematical Problems in Engineering, 2014.
47. Dong, D., & Herbert, J. Energy efficient vm placement supported by data analytic service. In Cluster, Cloud and Grid Computing (CCGrid), 13th IEEE/ACM International Symposium, 2013, May, 648–655.
48. Arianyan, E., Taheri, H., & Sharifian, S. Novel energy and SLA efficient resource management heuristics for consolidation of virtual machines in cloud data centers. Computers & Electrical Engineering, 47, 2015, 222–240.
49. Feller, E., Rohr, C., Margery, D., & Morin, C. Energy management in IaaS clouds: a holistic approach. In Cloud Computing (CLOUD), IEEE 5th International Conference, 2012, June, 204–212.
50. Wang, S., Liu, Z., Zheng, Z., Sun, Q., & Yang, F. Particle swarm optimization for energy-aware virtual machine placement optimization in virtualized data centers. In Parallel and Distributed Systems (ICPADS), International Conference, IEEE. 2013, December, 102–109.
51. Uddin, M., Darabidarabkhani, Y., Shah, A., & Memon, J. Evaluating power efficient algorithms for efficiency and carbon emissions in cloud data centers: A review. Renewable and Sustainable Energy Reviews, 51, 2015, 1553–1563.
52. Dabbagh, M., Hamdaoui, B., Guizani, M., & Rayes, A. Energy-efficient resource allocation and provisioning framework for cloud data centers. Network and Service Management, IEEE Transactions, 12 (3), 2015, 377–391.
53. Chen, H., Zhu, X., Guo, H., Zhu, J., Qin, X., & Wu, J. Towards energy-efficient scheduling for real-time tasks under uncertain cloud computing environment. Journal of Systems and Software, 99, 2015, 20–35.
54. Zhang, Z., Hsu, C. C., & Chang, M. Cool Cloud: A Practical Dynamic Virtual Machine Placement Framework for Energy Aware Data Centers. In Cloud Computing (CLOUD), IEEE 8th International Conference, 2015, June, 758–765.
55. Lee, Y.C., & Zomaya, A.Y. Energy efficient utilization of resources in cloud computing systems. The Journal of Supercomputing, 60 (2), 2012, 268–280.
56. Beloglazov, A., & Buyya, R. Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers. Concurrency and Computation: Practice and Experience, 24 (13), 2012, 1397–1420.
57. Barroso, L.A., & Hölzle, U. The case for energy-proportional computing. Computer 12, 2007, 33–37.
58. Lu, K., Yahyapour, R., Wieder, P., Yaqub, E., Abdullah, M., Schloer, B., & Kotsokalis, C. Fault-tolerant Service Level Agreement lifecycle management in clouds using actor system. Future Generation Computer Systems, 54, 2016, 247–259.
59. Moon, H. J., Chi, Y., & Hacigumus, H. SLA-aware profit optimization in cloud services via resource scheduling. In Services (SERVICES-1), 6th World Congress, IEEE. 2010, July, 152–153.
60. Buyya, R., Garg, S. K., & Calheiros, R. N. SLA-oriented resource provisioning for cloud computing: Challenges, architecture, and solutions. In Cloud and Service Computing (CSC), International Conference, IEEE. 2011, December, 1–10.
61. Nayak, D., Martha, V. S., Threm, D., Ramaswamy, S., Prince, S., & Fatimberger, G. Adaptive scheduling in the cloud—SLA for Hadoop job scheduling. In Science and Information Conference (SAI), IEEE. 2015, July, 832–837.
62. García, A. G., Espert, I. B., & García, V. H. SLA-driven dynamic cloud resource management. Future Generation Computer Systems, 31, 2014, 1–11.
63. Wu, L., Garg, S. K., & Buyya, R. SLA-based resource allocation for software as a service provider (SaaS) in cloud computing environments. In Cluster, Cloud and Grid Computing (CCGrid), 11th IEEE/ACM International Symposium, 2011, May, 195–204.
64. Yaqub, E., Yahyapour, R., Wieder, P., Jehangiri, A. I., Lu, K., & Kotsokalis, C. Metaheuristics-based planning and optimization for sla-aware resource management in paas clouds. In Proceedings of the IEEE/ACM 7th International Conference on Utility and Cloud Computing, 2014, December, 288–297.
65. Rajabi, A., Faragardi, H. R., & Yazdani, N. Communication-aware and energy-efficient resource provisioning for real-time cloud services. In Computer Architecture and Digital Systems (CADS), 2013 17th CSI International Symposium, IEEE. 2013, October, 125–129.

66. Bi, J., Yuan, H., Tie, M., & Tan, W. SLA-based optimisation of virtualised resource for multi-tier web applications in cloud data centres. Enterprise Information Systems, 9 (7), 2015, 743–767.
67. Farokhi, S., Jrad, F., Brandic, I., & Streit, A. HS4MC–Hierarchical SLA-based Service Selection for Multi-Cloud Environments. In: CLOSER 2014, 722–734.
68. Serrano D, Bouchenak S, Kouki Y, de Oliveira Jr FA, Ledoux T, Lejeune J, Sopena J, Arantes L, Sens P. SLA guarantees for cloud services, In Future Generation Computer Systems. In: Future Generation Computer Systems, 2016, 233–246.
69. Morshedlou, H., & Meybodi, M. R. Decreasing impact of sla violations: a proactive resource allocation approach for cloud computing environments. Cloud Computing, IEEE Transactions, 2 (2), 2014, 156–167.
70. Lu, K., Yahyapour, R., Wieder, P., Kotsokalis, C., Yaqub, E., & Jehangiri, A. I. Qos-aware vm placement in multi-domain service level agreements scenarios. In IEEE Sixth International Conference on Cloud Computing, 2013, June, 661–668.
71. Selvi, S. T., SathiaBham, P. R., Architha, S., Kaarunya, T., & Vinothini, K. Scheduling In Virtualized Grid Environment Using Hybrid Approach. International Journal of Grid Computing & Applications (IJGCA), 2010, Vol, 1.
72. Wu, S., Chen, H., Di, S., Zhou, B., Xie, Z., Jin, H., & Shi, X. Synchronization-Aware Scheduling for Virtual Clusters in Cloud. Parallel and Distributed Systems, IEEE Transactions on, 26 (10), 2015, 2890–2902.
73. Wang, H., Isci, C., Subramanian, L., Choi, J., Qian, D., & Mutlu, O. A-DRM: Architecture-aware distributed resource management of virtualized clusters. In Proceedings of the 11th ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments, 2015, March, 93–106. ACM.
74. Garbacki, P., & Naik, V. K. Efficient resource virtualization and sharing strategies for heterogeneous grid environments. In Integrated Network Management, IM'07. 10th IFIP/IEEE International Symposium, 2007, May, 40–49.
75. Nguyen, Q.T., Quang–Hung, N., Tuong, N.H., Tran, V.H., & Thoai, N. Virtual machine allocation in cloud computing for minimizing total execution time on each machine. In Computing, Management and Telecommunications (ComManTel), International Conference, IEEE. 2013, January, 241–245.
76. Sahal, R., & Omara, F. A. Effective virtual machine configuration for cloud environment. In Informatics and Systems (INFOS), 9th International Conference, IEEE. 2014, December, PDC-15.
77. Wang, X., Liu, X., Fan, L., & Jia, X. A decentralized virtual machine migration approach of data centers for cloud computing. Mathematical Problems in Engineering, 2013.
78. Aral, A., & Ovatman, T. Improving Resource Utilization in Cloud Environments using Application Placement Heuristics. In: CLOSER, 2014, 527–534.
79. Li, K., Wu, J., & Blaisse, A. Elasticity-aware virtual machine placement for cloud datacenters. In Cloud Networking (CloudNet), IEEE 2nd International Conference, 2013, November, 99–107.
80. Li, W., Tordsson, J., & Elmroth, E. Virtual machine placement for predictable and time-constrained peak loads. In Economics of Grids, Clouds, Systems, and Services, 2011, 120–134.
81. Park, J. G., Kim, J. M., Choi, H., & Woo, Y. C. Virtual machine migration in self-managing virtualized server environments. In Advanced Communication Technology, ICACT 2009. 11th International Conference, IEEE. 2009, February, Vol. 3, 2077–2083.
82. Ferreto, T. C., Netto, M. A., Calheiros, R. N., & De Rose, C. A. Server consolidation with migration control for virtualized data centers. Future Generation Computer Systems, 27 (8), 2011, 1027–1034.
83. Tordsson, J., Montero, R. S., Moreno-Vozmediano, R., & Llorente, I. M. Cloud brokering mechanisms for optimized placement of virtual machines across multiple providers. Future Generation Computer Systems, 28 (2), 2012, 358–367.
84. Baruchi, A., Toshimi Midorikawa, E., & Netto, M. A. Improving Virtual Machine live migration via application-level workload analysis. In Network and Service Management (CNSM), 10th International Conference, IEEE. 2014, Nov. 163–168.
85. Hieu, N. T., Di Francesco, M., & Yla-Jaaski, A. Virtual Machine Consolidation with Usage Prediction for Energy-Efficient Cloud Data Centers. In Cloud Computing (CLOUD), IEEE 8th International Conference, 2015, June, 750–757.
86. Ezugwu, A. E., Buhari, S. M., & Junaidu, S. B. Virtual machine allocation in cloud computing environment. International Journal of Cloud Applications and Computing (IJCAC), 3 (2), 2013, 47–60.
87. Z. Xiao, W. Song and Q. Chen. Dynamic Resource Allocation Using Virtual Machines for Cloud Computing Environment. In IEEE Transactions on Parallel and Distributed Systems, 24, 6, June 2013, 1107–1117.
88. M. Seddigh, H. Taheri and S. Sharifian. Dynamic prediction scheduling for virtual machine placement via ant colony optimization. Signal Processing and Intelligent Systems Conference (SPIS), Tehran, Iran, 2015 , 104–108.
89. Xu, X., Hu, H., Hu, N., & Ying, W. Cloud task and virtual machine allocation strategy in cloud computing environment. In Network Computing and

Information Security, 2012, 113–120.

90. Quang-Hung, N., Nien, P. D., Nam, N. H., Tuong, N. H., & Thoai, N. A genetic algorithm for power-aware virtual machine allocation in private cloud. In Information and Communication Technology 2013, 183–191.
91. Saraswathi, A. T., Kalaashri, Y. R. A., & Padmavathi, S. Dynamic resource allocation scheme in cloud computing. Procedia Computer Science, 47, 2015, 30–36.
92. Coutinho, R. D. C., Drummond, L. M., Frota, Y., & de Oliveira, D. Optimizing virtual machine allocation for parallel scientific workflows in federated clouds. Future Generation Computer Systems, 46, 2015, 51–68.
93. de Oliveira, D., Viana, V., Ogasawara, E., Ocaña, K., & Mattoso, M. Dimensioning the virtual cluster for parallel scientific workflows in clouds. In Proceedings of the 4th ACM workshop on Scientific cloud computing, 2013, June, 5–12.