RESOURCE ALLOCATION FOR JOB OPTIMZATION IN MULTI-CLOUD ENVIRONMENT

By

MOHD HAIRY BIN MOHAMADDIAH

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

August 2023

FSKTM 2023 3

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

RESOURCE ALLOCATION FOR JOB OPTIMZATION IN MULTI-CLOUD ENVIRONMENT

By

MOHD HAIRY BIN MOHAMADDIAH

August 2023

Chairman : Associate Professor Azizol bin Hj. Abdullah, PhD
Faculty : Computer Science and Information Technology

Resource management consists of three domains namely allocation, discovery, and monitoring. Resource allocation in cloud computing is a complex process that involves identifying the best pair of tasks and resources based on quality-of-service requirements. Hence, the agility of demands for job processing from the clients is a challenge for cloud service broker to efficiently allocate resources and meet the requirements of the tasks within the specified deadline.

This thesis studies the resource management problem in resource allocation of multi cloud environment. The main problem is when the allocation of resources influences the optimization of job processing and the cloud resources. It leads the resources to be underutilized or overutilized, resulting in poor resource utilization and inefficient job execution. This thesis analyses the current optimization solution, used preemption mechanism via dynamic cloud list scheduling (DCLS) and dynamic min-min scheduling (DCMMS) method. The current solution might cause higher execution time and lower the utilization rate. Therefore, it is essential to provide an efficient mechanism for resource allocation of job optimization to reduce the execution time and increase the utilization time. The resource allocation and selection mechanisms are proposed for cloud broker of job optimization in a multi-cloud environment. It also proposes high level service brokering model to support the allocation and selection mechanisms. A Multilevel Allocation mechanism (MLA) includes jobs and resources in allocation mechanism as an effort to optimize job processing and resource allocation. The allocation approach explicitly considers priority list and rank the resources for job allocation. To leverage on the feedback information and processing power of a resource, a Resource optimization Based on Reputation mechanism (REP-R) is being introduced. The proposed mechanism deals with both job and resources simultaneously. Finally, a selection mechanism method, Resource Selection Based on Job Classification, (RES-J) is being proposed to select the fit resources based on job classification. Decision tree classification is adopted for job classification; thus, it enables the discovery and optimization of resource availability due to over or under provisioning.

To simulate the proposed mechanisms, CloudSim is used to conduct an extensive simulation with a diverse set of jobs and scenarios. The findings of the mechanism show that MLA is 80% better than other DAG methods. In producing shorter schedule length, DCLS produces a better schedule length ratio (SLR) compared to the proposed mechanisms of MLA and DCMMS. However, MLA is the best possible allocation of resources or scheduling strategy to achieve the objective by minimizing the schedule length since the mechanism considers many parameters as compared to DCLS when scheduling the job. In contrast with SLR, MLA produces the best makespan among the three mechanisms. For the second mechanism, the average execution time in REP-R is 7% faster than DCMMS and it outperforms DCLS. This is due to the allocation that chooses the most reputable resources, thus minimizes the job execution time. For the third mechanism, the overall performance shows that RES-J utilizes the most resources compared to DCMMS and DCLS in booth loose and tight scenarios.

Overall, the proposed mechanism comprises multi-level, resource reputation and selection in the allocation of resources, and it shows promising results in improving resource utilization and overall performance of cloud systems. In addition, it is a strategy for the cloud broker with the aim to minimize the overall cost and optimize job scheduling.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

PERUNTUKAN SUMBER UNTUK PENGOPTIMUMAN KERJA DALAM PERSEKITARAN AWAN BERBILANG

Oleh

MOHD HAIRY BIN MOHAMADDIAH

Ogos 2023

Pengerusi : Profesor Madya Azizol bin Hj. Abdullah, PhD
Fakulti : Sains Komputer dan Teknologi Maklumat

Pengurusan sumber terdiri daripada tiga domain iaitu peruntukan, penemuan dan pemantauan. Peruntukan sumber dalam pengkomputeran awan adalah proses yang kompleks, yang melibatkan proses mengenal pasti kesesuaian di antara permintaan pelanggan dan sumber terbaik, berdasarkan keperluan kualiti perkhidmatan. Oleh itu, ketangkasan permintaan untuk pemprosesan permintaan dari pelanggan adalah satu cabaran bagi broker perkhidmatan awan untuk memperuntukkan sumber dengan cekap dan memenuhi keperluan tugas dalam tempoh masa yang ditetapkan.

Tesis ini mengkaji masalah pengurusan sumber dalam peruntukan sumber persekitaran berbilang awan. Masalah utama ialah apabila peruntukan sumber mempengaruhi pengoptimuman pemprosesan tugas dan sumber awan. Ia menyebabkan sumber-sumber yang kurang digunakan atau terlalu banyak digunakan, mengakibatkan penggunaan sumber yang lemah dan pelaksanaan kerja yang tidak cekap. p\Penyelesaian peruntukan sumber menerusi pengoptimuman semasa, menggunakan mekanisme pemintasan melalui kaedah penjadualan senarai awan dinamik (DCLS) dan kaedah penjadualan min-min dinamik (DCMMS) telah dianalisis. Kedua-dua kaedah ini boleh melambatkan tempoh masa pemprosesan kerja serta penggunaan sumber yang lemah. Oleh itu, adalah penting untuk menyediakan mekanisme yang cekap untuk peruntukan sumber pengoptimuman pekerjaan untuk mengurangkan tempoh masa pemprosesan kerja. Mekanisme peruntukan dan pemilihan dicadangkan untuk broker awan bagi pengoptimuman pekerjaan dalam persekitaran berbilang awan. Ia juga mencadangkan model broker perkhidmatan peringkat tinggi untuk menyokong peruntukan dan mekanisme pemilihan. Mekanisme Peruntukan Bertingkat (MLA) merangkumi pekerjaan dan sumber dalam mekanisme peruntukan sebagai usaha mengoptimumkan pemprosesan kerja dan peruntukan sumber. Pendekatan peruntukan secara jelas mempertimbangkan senarai keutamaan dan kedudukan sumber untuk peruntukan pekerjaan. Untuk memanfaatkan maklumat maklum balas dan kuasa pemprosesan sumber, pengoptimuman Sumber Berdasarkan mekanisme Reputasi (REP-R)

dicadangkan. Mekanisme ini, menangani pekerjaan dan sumber secara serentak. Akhir sekali, kaedah mekanisme pemilihan, Pemilihan Sumber Berdasarkan Klasifikasi Pekerjaan, (RES-J) dicadangkan untuk memilih sumber yang sesuai berdasarkan klasifikasi pekerjaan. Kaedah klasifikasi pokok keputusan diguna pakai untuk klasifikasi pekerjaan; oleh itu, ia membolehkan penemuan dan pengoptimuman ketersediaan sumber dalam situasi terlebih atau kekurangan sumber untuk memproses pekerjaan.

Untuk mengesahkan kaedah yang dicadangkan, *CloudSim* digunakan untuk melaksanakan simulasi dengan kepelbagaian jenis pekerjaan dan senario. Hasil simulasi menunjukkan MLA adalah menunjukkan keputusan 80% lebih baik daripada kaedah DAG lain. Dalam menghasilkan panjang jadual yang lebih pendek, DCLS menghasilkan nisbah panjang jadual (SLR) yang lebih baik berbanding mekanisme MLA dan DCMMS yang dicadangkan. Walau bagaimanapun, MLA adalah peruntukan sumber atau strategi penjadualan yang terbaik untuk mencapai objektif dengan meminimumkan panjang jadual kerana mekanisme ini, mempertimbangkan banyak parameter berbanding DCLS semasa menjadualkan kerja untuk diperoses. Berbeza dengan SLR, MLA menghasilkan tempoh keseluruhan kerja selesai diperoseskan terbaik, di antara tiga mekanisme. Untuk mekanisme kedua, purata masa pelaksanaan dalam REP-R adalah 7% lebih cepat daripada DCMMS dan juga mengatasi DCLS. Ini disebabkan mekanisme MLA memilih sumber yang paling bereputasi, seterusnya meminimumkan masa pelaksanaan kerja. Untuk mekanisme ketiga, prestasi keseluruhan menunjukkan bahawa RES-J menggunakan sumber yang paling optimum berbanding DCMMS dan DCLS dalam senario longgar dan ketat.

Kesimpulannya, mekanisme yang dicadangkan terdiri daripada pelbagai peringkat, reputasi sumber dan pemilihan dalam peruntukan sumber, dan ia menunjukkan hasil yang menjanjikan peningkatan penggunaan sumber dan prestasi keseluruhan sistem awan. Di samping itu, ia adalah strategi untuk broker awan dengan tujuan untuk meminimumkan kos keseluruhan dan mengoptimumkan penjadualan kerja.

ACKNOWLEDGEMENTS

First of all, i'm grateful to Allah the Almighty, for the wonderful life given to me and allowing me to complete this journey. Thank to Allah s.w.t for the air I breathe, the food i ate, the love that I am thankful for and the Iman that i always seek out.

Secondly, i wish to extend my sincere gratitude to my supervisor, Associate Prof Dr Azizol Abdullah, for his understanding, not giving up on me, his valuable comments, thought, wisdom, and his assistance to complete my study. Not to forget, my endless thanks to my supervisory committee member, Prof Dato Dr Shamala Subramaniam and Associate Prof Dr Masnida Hussin, who always supports and guiding me throughout my study.

Special thanks to Faculty Dean and Deputy Deans for their trust, time, patience, and guidance. Not to forget, thank you to all the lecturers and administration staff at the Faculty of Computer Science & Information Technology, UPM especially to Miss Noriah from the Post Graduate Section for their helpful assistance, advice, and comments.

My heartfelt thanks also go to all my family especially my late mom, my father, my siblings and their family, for all their love, support and trust that keep me strong and never give up in completing my study.

Thank you also to my best friend alias Bro Adzly, and Bro Madi, my close friend Aliyu, Arzila, Omid, Azman, Iyad, GBS Gang, Mizi, Dr Ariza, late Abang Farez and Afzey for their assistance, love, laugh, and thought that keep me going through this valuable journey.

Thank you to my bosses, Associate Prof Dr Juliana and Puan Sariani for always support and encourage me to finish my study. Special thanks to my team, BSM team for the encouragement. To my colleagues at Jabatan Infostruktur, PPII UiTM, especially Mr Mazuhan thank you for your assistance. Whenever i needed their help, they were always there to lend a helping hand. I am thankful to them for giving me the spiritual and moral support upon completion of this study.

Finally, to those who had indirectly contributed in completing my thesis, my thanks also goes to them.

Alfatihah to my late mom and abang farez.

Thank you all.

This thesis was submitted to the Senate of the Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Azizol bin Hj. Abdullah, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Shamala a/p K. Subramaniam, PhD
Professor Dato'
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Masnida binti Hussin, PhD
Associate Professor Ts.
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

ZALILAH MOHD SHARIFF, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 14 March 2024

TABLE OF CONTENTS

Page

CHAPTER

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AR	Advanced Reservation
BE	Best Effort
BW1	Basework 1
CB	Cloud Service Broker
CSA	Composable Services Architecture
CSP	Cloud service provider
CPU	Central Processing Unit
DAG	Directed Acyclic Graph
DCLS	Dynamic cloud list scheduling
DCMMS	Dynamic min-min scheduling
HEFT	Heterogenous Early Finish Time
IaaS	Infrastructure as a Service
ISMF	Infrastructure Services Modelling Framework
IT	Information Technology
IP	infrastructure provider
MLA	Multi-level Allocation
NIST	National Institute of Standards & Technology
NP	Non-deterministic polynomial time
RES-J	Resource Selection Based on Job Classification
REP-R	Resource optimization Based on Reputation mechanism
RMS	Resource Management System
SDF	Service Delivery Framework
SLA	Service Level Agreement
SP	Service Provider
VM	Virtual Machine

CHAPTER 1

INTRODUCTION

This chapter introduces the background of the research, with a brief description of cloud computing, cloud resource allocation and brokering, followed by the research motivation and problem statement, the objectives and major contributions. The thesis chapters are outlined at the end of this chapter.

1.1 Background

The innovative technology of cloud computing as part of utility computing is disruptive; thus, it requires changing how organizations strategize IT spending on infrastructure, usage of computers and the Internet (Dimitrov & Osman, 2012). The migration of applications and the usage of cloud services have revolutionized businesses globally and become a phenomenal transformation of IT services.

“Cloud computing is defined as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services)”(Mell & Grance, 2011). The computing resources in cloud, can rapidly be provisioned and released with minimal management effort or service provider interaction characteristics comprising a broad network access with its ability to access network via heterogeneous platforms, on demand self-service, when it provisions the computing power automatically (Bohn et al., 2011).

The Cloud Service Provider (quoted as service provider, SP) delivers the provisioned resources, by running the cloud software to allocate requested resources via multiple services to the subscribers. Mostly the provisioning will then be abstracted into virtual machines which make use virtualization technology. The requested resources are handled by Cloud Service Broker (quoted as CB). CB will select the best set of resources requested by the user/client subscriber and provide the resources to them (Jula et al., 2014). Subsequently, it is used by the subscribers to deploy its required applications in the provision platform. The subscribers will have full access of the provisioned resource while the provider controls the physical hardware layer and monitors the performance of the resources (Sun et al., 2010). Figure 1.1 below illustrates the visualization of the concept discussed previously regarding cloud reference model which depicts the services flow of the service model (Bohn et al., 2011). The reference model is our main reference for proposing the mechanisms of this study. It will be discussed in chapter 3.

Based on the model, cloud brokers act as middlemen between service providers and cloud providers. Numerous cloud providers rent various kinds of cloud resources to cloud brokers (Mehrotra et al., 2016). With diverse of jobs submitted to cloud provider, it will require different types of resources. These resources will be allocated accordingly. Therefore, resource allocation plays a vital role in provisioning computing resources to fulfil the demand from a client. In addition, the difficulty of allocating cloud resources has increased with the introduction of federated cloud computing systems and cloud brokerage.

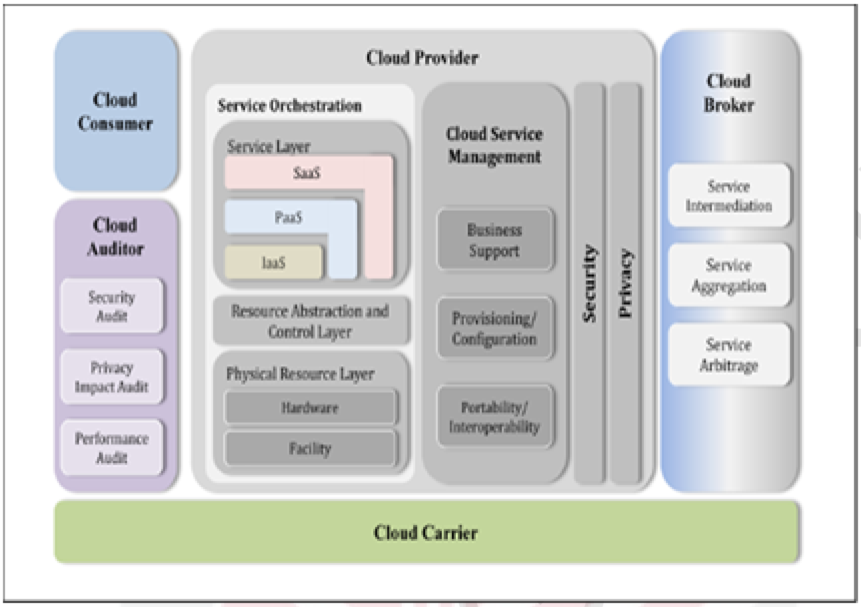

Figure 1.1 : Cloud Reference Model by NIST

1.2 Problem Statements

Resource allocation in a large-scale cloud environment is a complex task. It involves a parallel processing of scheduling heterogeneous jobs across multiple clouds. Optimization plays a crucial role in maximizing resource utilization and minimizing job execution time, particularly in scheduling the resources efficiently and effectively.

Three main problems that motivate this research are:

1. Due to the diverse and heterogenous type of jobs to be processed, there is a difficult optimization problem to allocate the resources in a way that satisfies the requirements and preferences of each user while maintaining overall system performance and efficiency (Li et al., 2012). Users demand different sets of resources based on their required jobs to be processed. This will lead to cloud brokers facing major problems in processing user's applications within a minimum time frame. Consequently, this will influence the makespan for job processing (Yousaf & Welzl, 2014) and impact the schedule length.
2. The other challenge to optimize cloud resources occurs especially when processing the jobs. Even though consolidation environments and workloads are shared across multiple cloud, current optimization mechanism does not guarantee that the jobs can be executed in timely manner (Li et al., 2012).

Furthermore, running intensive jobs in the cloud will require diversity of resources to be allocated. However, ineffective jobs allocation to the cloud resources are caused by misaligning jobs with the underlying infrastructure, which compromises system stability (Liu & Buyya, 2020). This will resort the resources to become unavailable due to resource contention (Tchernykh et al., 2015) and limited usage at the cloud provider. It will affect the job execution time. With reputable resources it will hinder the overprovisioning resources scenario, increase optimization of resources thus improve the job execution time.

3. The resources are becoming underutilized due to improper selection of resources at brokering services. It leads to the non-optimization of resources and increases a significant cost to the cloud broker, infrastructure provider and cloud user/client subscriber (Li et al., 2012). By successfully predicting peak loads, it solves over and under provisioning of cloud resources (Espadas et al., 2013). However, the absence of having an effective resource selection mechanism has caused the costly resources to be wasted during non-peak times (underutilization). Therefore, there are losses of revenues for the providers as the resource's selection is not well planned. Proper selection of the best fit resources is supported by choosing the predictive loads/incoming jobs among a variety of cloud resources that best suit the needs of the user (Qi et al., 2022). Subsequently, it selects and processes the job assigned by the cloud service provider and will fully increase the utilization of resources.

1.3 Research Objectives

The primary objective is to propose a new allocation and selection mechanisms in multi-cloud computing environment for cloud service provider, cloud broker and cloud user/client to optimize job and resource allocation and selection by utilizing the concept of job scheduling, job prioritization, resource reputation and resource classification. Specific objectives that must be fulfilled to ensure that the primary objective of the study is achieved are:

1. To propose a multi-level resource allocation mechanism called MLA that include jobs and resources in allocation mechanism as an effort for optimization by enhancing the existing task scheduling method based on jobs prioritization and resource ranking to minimize the schedule length; thus, the make-span of job processing is more efficient in a large-scale cloud environment.
2. To propose a resource optimization based on Reputation mechanism called REP-R for resource allocation to minimize the job execution time for every resource in a cloud environment.
3. To propose a resource selection vased on job classification called RES-J that will trigger the best fit resources by applying classification in machine learning technique to assist and optimize the resource availability due to over provisioning and underutilized resources in job processing. This method will improve the utilization rate of the job processing in the cloud.

1.4 Research Scope

The scope of this study is centered on resource allocation problems at service and infrastructure layers in the cloud deployment scheme. The research concentrates on multi-cloud deployment with heterogeneous resources, efficiency on allocation of job and resources and, the process selection of resources in the problem domain. In addition to the above, tasks and jobs and majority class are used interchangeably in this study, which represents the input of our proposed contribution of this thesis.

1.5 Research Significance

A large-scale heterogeneous environment of cloud computing involves dynamic and large workloads that require a certain number of resources or limited time for processing. Therefore, the demand for cloud services has been increasing tremendously especially for data intensive and scientific computing application which requires the availability and reliability of a huge number of computing resources for performing large scale experiments. The current high-performance computing solutions and installed facilities, such as clusters and super computers, can accommodate these requirements. However, these facilities are difficult to set up, maintain, and operate. Therefore, cloud computing services provide scientists with a completely new model of provisioning the computing infrastructure.

From this study, the proposed mechanism will help the cloud broker mainly, and cloud infrastructure provider to improve the service time and optimize the process of resource provisioning in the cloud. It will maximize the profit of the brokers while reducing the operation cost. The proposed mechanism should relieve the burden of the cloud users from having problems acquiring resources and processing requests. From a financial viewpoint, the provider will enjoy cost reduction for the utility, while the user will benefit from the fee reduction due to better resource utilization.

1.6 Thesis Organization

The thesis is organized in accordance with the structured thesis standard by Universiti Putra Malaysia. For this thesis, the term of ‘this study’ and ‘this thesis’ is used interchangeably.

The thesis is organized as follows:

Chapter 2: It discusses the background study of resource management problems namely in allocation and provisioning. together with different kinds of strategies that have been proposed recently in solving the problems. Some examples of the resource provisioning and allocation problem are also presented in the chapter.

Chapter 3: It explains in detail the research methodologies conducted in this thesis, such as the flow of the research, the detail of data and the configuration used in the simulations. It also discusses the implementation of our benchmark models. This chapter will also explain about brokering services in the cloud and the relationship with our mechanism. The chapter also describes the formulation of our mechanisms for brokering services. Furthermore, this chapter will detail the proposed metrics to validate our mechanisms.

Chapter 4: It explores the first type of allocation mechanism, which consists of a multi-level allocation model. This model consists of job prioritization and resource ranking model. It presents schedule length ratio and makes span comparison with other mechanism in cloud.

Chapter 5: It explores optimization mechanism by applying resource reputation method. The mechanism for resources of the resource allocation is also included. It will make full use of the information of resources to set up reputation resources; thus, improving the job execution time in resource allocation.

Chapter 6: It describes the resource selection mechanism. The mechanism will be based on job classification and selection of the best resources. It will investigate the utilization rate of the resource and when the resources are being allocated. It will discuss the result and analysis obtained from the implementation of the proposed mechanism.

Chapter 7: Finally, this chapter concludes the thesis and suggests several improvements that can be done based on this research contribution as future work.

REFERENCES

Aazam, M., & Huh, E. (2015). Cloud broker service-oriented resource management model. *Transactions on Emerging Telecommunications Technologies, 28, e2937*. https://doi.org/10.1002/ett.2937

Aceto, G., Botta, A., Donato, W. De, & Pescapè, A. (2013). Cloud monitoring: A survey. *Computer Networks*, *57*(9), 2093–2115. https://doi.org/10.1016/j.comnet.2013.04.001

Aggarwal, R. (2018). Resource provisioning and resource allocation in cloud computing environment. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology © 2018 IJSRCSEIT*, *3*(3), 31.

Ahmed El-Sayed Ibrahim. (2004). *Integrating Intelligent Methods for Scheduling in Grid Computing Systems*. [Doctoral Dissertation]. University of Connecticut.

Ali-eldin, A., Tordsson, J., Elmroth, E., & Kihl, M. (2013). *Workload classification for efficient auto-scaling of cloud resources*. https://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-87231.

Arabnejad, H., & Barbosa, J. G. (2014). A budget constrained scheduling algorithm for workflow applications. *Journal of Grid Computing*, *12*(4), 665–679. https://doi.org/10.1007/s10723-014-9294-7

Armenta-Cano, F. A., Tchernykh, A., Cortes-Mendoza, J. M., Yahyapour, R., Drozdov, A. Y., Bouvry, P., Kliazovich, D., Avetisyan, A., & Nesmachnow, S. (2017). Min_c: Heterogeneous concentration policy for energy-aware scheduling of jobs with resource contention. *Programming and Computer Software*, *43*(3), 204–215. https://doi.org/10.1134/S0361768817030021

Asghari, A., Sohrabi, M. K., & Yaghmaee, F. (2020). A cloud resource management framework for multiple online scientific workflows using cooperative reinforcement learning agents. *Computer Networks* (Vol. 179). https://doi.org/10.1016/j.comnet.2020.107340

Ayyadevara, V. K. (2018). Random Forest. In *Pro Machine Learning Algorithms*. Apress. https://doi.org/10.1007/978-1-4842-3564-5_5.

Babu, L. D. D., Gunasekaran, A., & Krishna, P. V. (2014). A decision-based pre-emptive fair scheduling strategy to process cloud computing work-flows for sustainable enterprise management. *International Journal of Business Information Systems*, *16*(4), 409. https://doi.org/10.1504/IJBIS.2014.063929

Balaji, M., Aswani, C., & Rao, G. S. V. R. K. (2018). Predictive cloud resource management framework for enterprise workloads. *Journal of King Saud University - Computer and Information Sciences*, *30*(3), 404–415. https://doi.org/10.1016/j.jksuci.2016.10.005

Bazghandi, A. (2021). A self-configurable model for cloud resource allocation. *ICCKE 2021 - 11th International Conference on Computer Engineering and Knowledge*, *Iccke*, 55–61. https://doi.org/10.1109/ICCKE54056.2021.9721523

Beloglazov, A. (2013). *Energy-efficient management of virtual machines in data centers for cloud computing*. [Doctoral Dissertation] University of Melbourne. http://hdl.handle.net/11343/38198.

Bohn, R. B., Messina, J., Liu, F., Tong, J., & Mao, J. (2011). NIST cloud computing reference architecture. *2011 IEEE World Congress on Services*, 594–596. https://doi.org/10.1109/SERVICES.2011.105

Buyya, R., & Ranjan, R. (2010). Special section: Federated resource management in grid and cloud computing systems. *Future Generation Computer Systems*, *26*(8), 1189–1191. https://doi.org/10.1016/j.future.2010.06.003

Cai, Z., Li, X., & Gupta, J. N. D. (2013). Critical path-based iterative heuristic for workflow scheduling in utility and cloud computing. In Basu, S., Pautasso, C., Zhang, L., Fu, X. (eds), *Service-Oriented Computing. ICSOC 2013. Lecture Notes in Computer Science*, vol 8274. Springer. https://doi.org/10.1007/978-3-642-45005-1_15.

Calatrava, A., Molto, G., & Hernandez, V. (2011). Combining grid and cloud resources for hybrid scientific computing executions. *2011 IEEE Third International Conference on Cloud Computing Technology and Science*, 494–501. https://doi.org/10.1109/CloudCom.2011.73

Calheiros, R., & Ranjan, R. (2011). CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. *Software: Practice …*, *August 2010*, 23–50. https://doi.org/10.1002/spe

Çavdar, D., Birke, R., Chen, L. Y., & Alagöz, F. (2015). A simulation framework for priority scheduling on heterogeneous clusters. *Future Generation Computer Systems*, *52*, 37–48. https://doi.org/http://dx.doi.org/10.1016/j.future.2015.04.008

Chao, W., & Junzheng, W. (2018). Cloud-service decision tree classification for education platform. *Cognitive Systems Research*, *52*, 234–239. https://doi.org/10.1016/j.cogsys.2018.06.021

Chauhan, S. S., Pilli, E. S., & Joshi, R. C. (2021). BGSA: Broker guided service allocation in federated cloud. *Sustainable Computing: Informatics and Systems*, *32*(September), 100609. https://doi.org/10.1016/j.suscom.2021.100609

Cheng, Y., Wu, Z., Liu, K., Wu, Q., & Wang, Y. (2019). Smart DAG tasks scheduling between trusted and untrusted entities using the MCTS method. *Sustainability (Switzerland)*, *11*(7), 1–16. https://doi.org/10.3390/su11071826

Demchenko, Y., Ham, J. V. D., & Yakovenko, V. (2011). On-demand provisioning of cloud and grid based infrastructure services for collaborative projects and groups. *2011 International Conference on Collaboration Technologies and Systems (CTS)*, Philadelphia, PA, USA, 134-142. doi: 10.1109/CTS.2011.5928675.

Dilip Kumar, S. M., Sadashiv, N., & Goudar, R. S. (2014). Priority based resource allocation and demand based pricing model in peer-to-peer clouds. *Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2014*, 1210–1216. https://doi.org/10.1109/ICACCI.2014.6968277

Dillon, T., Wu, C., & Chang, E. (2010). Cloud computing: issues and challenges. *2010 24th IEEE International Conference on Advanced Information Networking and Applications*, 27–33. https://doi.org/10.1109/AINA.2010.187

Dimitrov, M. & Osman, I. (2012). The impact of cloud computing on organizations in regard to cost and security. [Masters thesis]. Universitet UMEA.https://www.diva-portal.org/smash/get/diva2:728880/FULLTEXT02

Do, T. Van & Rotter, C. (2012). Comparison of scheduling schemes for on-demand IaaS requests. *Journal of Systems and Software, 85*(6), 1400–1408. https://doi.org/10.1016/j.jss.2012.01.019

Dubey, K., Kumar, M., & Sharma, S. C. (2018). Modified HEFT algorithm for task scheduling in cloud environment. *Procedia Computer Science*, *125*, 725–732. https://doi.org/10.1016/j.procs.2017.12.093

Emeakaroha, V. C., Brandic, I., Maurer, M., & Dustdar, S. (2010). Low level metrics to high level SLAs - LoM2HiS framework: Bridging the gap between monitored metrics and SLA parameters in cloud environments. *2010 International Conference on High Performance Computing & Simulation*, Caen, France, 48-54. doi: 10.1109/HPCS.2010.5547150

Emeras, J., Varrette, S., Guzek, M., & Bouvry, P. (2015). Evalix : classification and prediction of job resource consumption on HPC platforms. *Job Scheduling Strategies for Parallel Processing*, 102–122.

Espadas, J., Molina, A., Jiménez, G., Molina, M., Ramírez, R., & Concha, D. (2013). A tenant-based resource allocation model for scaling Software-as-a-Service applications over cloud computing infrastructures. *Future Generation Computer Systems*, *29*(1), 273–286. https://doi.org/10.1016/j.future.2011.10.013

Feitelson, D. G., Tsafrir, D., & Krakov, D. (2014). Experience with using the parallel workloads archive. *Journal of Parallel and Distributed Computing*, *74*(10), 2967–2982. https://doi.org/10.1016/j.jpdc.2014.06.013

Freitas, R. (2009). Scientific Research methods and computer science. *MAPI Seminars Workshop 2009*, 1–6. http://map.edu.pt/i/2008/map-i-research-methods-workshop-2009/RicardoFreitasFinal.pdf

Ghanbari, S. (2019). Priority-aware job scheduling algorithm in cloud computing: a multi-criteria approach. *Azerbaijan Journal of High Performance Computing*, *2*(1), 29–38. https://doi.org/10.32010/26166127.2019.2.1.29.38

Ghosh, R., Longo, F., Frattini, F., Russo, S., & Trivedi, K. (2014). Scalable analytics for IaaS cloud availability. *IEEE Transactions on Cloud Computing*, *7161*(c), 1–1. https://doi.org/10.1109/TCC.2014.2310737

Gill, S. S., & Buyya, R. (2019). Resource provisioning based scheduling framework for execution of heterogeneous and clustered workloads in clouds: from Fundamental to Autonomic Offering. *Journal of Grid Computing*, *17*(3), 385–417. https://doi.org/10.1007/s10723-017-9424-0

Guan, Q., Zhang, Z., & Fu, S. (2012). Ensemble of Bayesian predictors and decision trees for proactive failure management in cloud computing systems. *Journal of Communications*, *7*(1), 52–61. https://doi.org/10.4304/jcm.7.1.52-61

Guo, F., & Song, H. (2021). Study on optimization of course selection system based on virtual machine cluster and dynamic resource expansion. *ACM International Conference Proceeding Series*, 149–153. https://doi.org/10.1145/3459955.3460615

Hassan, H., El-Desouky, A. I., Ibrahim, A., El-Kenawy, E. S. M., & Arnous, R. (2020). Enhanced QoS-based model for trust assessment in cloud Computing environment. *IEEE Access*, *8*, 43752–43763. https://doi.org/10.1109/ACCESS.2020.2978452

Hu, Y., Wong, J., Iszlai, G., & Litoiu, M. (2009). Resource provisioning for cloud computing. *Proceedings of the 2009 Conference of the Center for Advanced Studies on Collaborative Research*, 101–111. http://dl.acm.org/citation.cfm?id=1723041

Hung, P. P., Aazam, M., Nguyen, T. D., & Huh, E. N. (2014). A solution for optimizing recovery time in cloud computing. *Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication, ICUIMC 2014*. https://doi.org/10.1145/2557977.2558029

Hung, P. P., & Huh, E. (2015). An adaptive procedure for task scheduling pptimization in mobile cloud computing. *Mathematical Problems in Engineering*, *2015*. https://doi.org/10.1155/2015/969027

Hussin, M., Abdullah, A., & Subramaniam, S. (2013). Adaptive resource allocation for reliable performance in heterogeneous distributed systems. *Algorithms and Architectures for Parallel Processing*, 51–58. http://link.springer.com/chapter/10.1007/978-3-319-03889-6_6

Hussin, M., Asilah Wati Abdul Hamid, N., & Kasmiran, K. A. (2015). Improving reliability in resource management through adaptive reinforcement learning for distributed systems. *Journal of Parallel and Distributed Computing*, *75*, 93–100. https://doi.org/10.1016/j.jpdc.2014.10.001

Hussin, M., & Latip, R. (2013). Adaptive Resource control mechanism through reputation-based scheduling in heterogeneous distributed Systems. *Journal of Computer Science*, *9*(12), 1661–1668. https://doi.org/10.3844/jcssp.2013.1661.1668

Hussin, M., Lee, Y. C., & Zomaya, A. Y. (2011). Efficient energy management using adaptive reinforcement learning-based scheduling in large-scale distributed systems. *2011 International Conference on Parallel Processing*, 385–393. https://doi.org/10.1109/ICPP.2011.18

Hussin, M., Lee, Y., & Zomaya, A. (2010). ADREA: A framework for adaptive resource allocation in distributed computing systems. *2010 11th IEEE International Conference on Parallel and Distributed Computing, Applications and Technologies*, 50–57. https://doi.org/10.1109/PDCAT.2010.19

Hwang, K. (2017). *Cloud computing for machine learning and cognitive applications.* Mit Press.

Iglesias, J. O., De Cauwer, M., Mehta, D., O'Sullivan, B., & Murphy, L. (2016). Increasing task consolidation efficiency by using more accurate resource estimations. *Future Generation Computer Systems*, *56*, 407–420. https://doi.org/10.1016/j.future.2015.08.018

Iyer, R., Illikkal, R., Tickoo, O., Zhao, L., Apparao, P., & Newell, D. (2009). VM3: Measuring, modeling and managing VM shared resources. *Computer Networks*, *53*(17), 2873–2887. https://doi.org/10.1016/j.comnet.2009.04.015

Jain, R. (1991). The art of computer systems performance analysis - techniques for experimental design, measurement, simulation, and modeling. In *Feuerlicht, G., Lamersdorf, W. (eds.) Service-Oriented Computing – ICSOC 2008 Workshops*. ICSOC 2008. Lecture Notes in Computer Science, vol 5472. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-01247-1_5

Jalaparti;, V., Nyuen;, G., Gupta;, I., & Caesar, M. (2010). Cloud Resource allocation games. *Computational Mathematics and Modeling*, *1*(4), 433–444. https://doi.org/10.1007/BF01128293

Javadi, B., Abawajy, J., & Buyya, R. (2012). Failure-aware resource provisioning for hybrid Cloud infrastructure. *Journal of Parallel and Distributed Computing*, *72*(10), 1318–1331. https://doi.org/10.1016/j.jpdc.2012.06.012

Jena, R. K. (2015). Multi objective task scheduling in cloud environment using nested PSO framework. *Procedia Computer Science*, *57*, 1219–1227. https://doi.org/http://dx.doi.org/10.1016/j.procs.2015.07.419

Jula, A., Sundararajan, E., & Othman, Z. (2014). Cloud computing service composition : A systematic literature review. *Expert Systems with Applications*, *41*, 3809–3824. https://doi.org/10.1016/j.eswa.2013.12.017

Junior, W., Oliveira, E., Santos, A., & Dias, K. (2019). A context-sensitive offloading system using machine-learning classification algorithms for mobile cloud environment. *Future Generation Computer Systems*, *90*, 503–520. https://doi.org/10.1016/j.future.2018.08.026

Kafil, M., & Ahmad, I. (1998). Optimal task assignment in heterogeneous distributed computing\nsystems. *IEEE Concurrency*, *6*(3). https://doi.org/10.1109/4434.708255

Katsaros, G., Kousiouris, G., Gogouvitis, S. V., Kyriazis, D., Menychtas, A., & Varvarigou, T. (2012). A Self-adaptive hierarchical monitoring mechanism for clouds. *Journal of Systems and Software*, *85*(5), 1029–1041. https://doi.org/10.1016/j.jss.2011.11.1043

Kecskemeti, G., Nemeth, Z., Kertesz, A., & Ranjan, R. (2018). Cloud workload prediction based on workflow execution time discrepancies. *Cluster Computing*, *9*. https://doi.org/10.1007/s10586-018-2849-9

Kolomvatsos, K., & Anagnostopoulos, C. (2019). Multi-criteria optimal task allocation at the edge. *Future Generation Computer Systems*, *93*, 358–372. https://doi.org/10.1016/j.future.2018.10.051

Krauter, K., Buyya, R., & Maheswaran, M. (2002). A taxonomy and survey of grid resource management systems for distributed computing. *Software: Practice and Experience*, *32*(2), 135–164. https://doi.org/10.1002/spe.432

Kumar, K. S. S., & Jaisankar, N. (2020). An automated resource management framework for minimizing SLA violations and negotiation in collaborative cloud. In *International Journal of Cognitive Computing in Engineering* (1, 27–35). https://doi.org/10.1016/j.ijcce.2020.09.001

Li, J., Qiu, M., Ming, Z., Quan, G., Qin, X., & Gu, Z. (2012). Online optimization for scheduling preemptable tasks on IaaS cloud systems. *Journal of Parallel and Distributed Computing*, *72*(5), 666–677. https://doi.org/10.1016/j.jpdc.2012.02.002

Liu, F. , Tong, J. , Mao, J. , Bohn, R. , Messina, J. , Badger, M., & Leaf, D. (2011). NIST cloud computing reference architecture. *Special Publication (NIST SP), National Institute of Standards and Technology*. https://doi.org/10.6028/NIST.SP.500-292, https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=909505

Liu, X., & Buyya, R. (2020). Resource management and scheduling in distributed stream processing systems: A taxonomy, review, and future directions. *ACM Computing Surveys*, *53*(3). https://doi.org/10.1145/3355399

Liu, Z., Qu, W., Liu, W., Li, Z., & Xu, Y. (2015). Resource preprocessing and optimal task scheduling in cloud computing environments. *Concurrency and Computation: Practice and Experience*, *27*(13), 3461–3482. https://doi.org/10.1002/cpe.3204

Mehrotra, R., Srivastava, S., Banicescu, I., & Abdelwahed, S. (2016). Towards an autonomic performance management approach for a cloud broker environment using a decomposition-coordination based methodology. *Future Generation Computer Systems*, *54*, 195–205. https://doi.org/10.1016/j.future.2015.03.020

Mell, P., & Grance, T. (2011). The NIST definition of cloud computing (draft). *NIST Special Publication*. http://pre-developer.att.com/home/learn/enablingtechnologies/The_NIST_Definition_of_Cloud_Computing.pdf

Moca, M., Litan, C., Silaghi, G. C., & Fedak, G. (2015). Multi-criteria and satisfaction oiented scheduling for hybrid distributed computing infrastructures. *Future Generation Computer Systems*, *55*, 428–443. https://doi.org/10.1016/j.future.2015.03.022

Moschakis, I. a., & Karatza, H. D. (2012). Evaluation of gang scheduling performance and cost in a cloud computing system. *The Journal of Supercomputing*, *59*(2), 975–992. https://doi.org/10.1007/s11227-010-0481-4

Moschakis, I. a., & Karatza, H. D. (2015). A meta-heuristic optimization approach to the scheduling of bag-of-tasks applications on heterogeneous clouds with multi-level arrivals and critical jobs. *Simulation Modelling Practice and Theory*, *57*, 1–25. https://doi.org/10.1016/j.simpat.2015.04.009

Muhuri, P. K., Rauniyar, A., & Nath, R. (2019). On arrival scheduling of real-time precedence constrained tasks on multi-processor systems using genetic algorithm. *Future Generation Computer Systems, 93,* 702-726. https://www.sciencedirect.com/science/article/abs/pii/S0167739X18312810

Muralidharan, C., & Anitha, R. (2022). Trusted cloud broker for estimating the reputation of cloud providers in federated cloud environment. *Concurrency and Computation: Practice and Experience*, *34*(1), 1–15. https://doi.org/10.1002/cpe.6537

Naha, R. K., Othman, M., & Akhter, N. (2017). Diverse approaches to cloud brokering: Innovations and issues. *International Journal of Communication Networks and Distributed Systems*, *19*(1), 99–120. https://doi.org/10.1504/IJCNDS.2017.085441

Nasr, A. A.-E., & Elbooz, S. A. (2018). Scheduling strategies in cloud computing: methods and implementations. *American Journal of Engineering and Applied Sciences*, *11*(2), 426–432. https://doi.org/10.3844/ajeassp.2018.426.432

Natingga, D. (2018). *Data science algorithms in a week: Top 7 algorithms for scientific computing, data analysis, and machine learning, 2nd Edition*. Packt Publishing.

Nayak, B., Bisoyi, B., & Pattnaik, P. K. (2023). Data center selection through service broker policy in cloud computing environment. *Materials Today: Proceedings*, *80*, 2218–2223. https://doi.org/10.1016/j.matpr.2021.06.185

Nayak, S. C., & Tripathy, C. (2018). Deadline based task scheduling using multi-criteria decision-making in cloud environment. *Ain Shams Engineering Journal*, *9*(4), 3315–3324. https://doi.org/10.1016/j.asej.2017.10.007

Newman, P., & Kotonya, G. (2015). A resource-aware framework for resource-constrained service-oriented systems. *Future Generation Computer Systems*, *47*, 161–175.

Pal, S., & Pattnaik, P. (2012). Efficient architectural framework for cloud computing. *International Journal of Cloud Computing and and Services Science*, *1*(2), 66–73. http://iaesjournal.com/online/index.php/IJ-CLOSER/article/view/513

Park, J., An, Y., & Yeom, K. (2016). Virtual cloud bank: Cloud service broker for intermediating services based on semantic analysis models. *Proceedings - 2015 IEEE 12th International Conference on Ubiquitous Intelligence and Computing, 2015 IEEE 12th International Conference on Advanced and Trusted Computing, 2015 IEEE 15th International Conference on Scalable Computing and Communications, 20*, 1022–1029. https://doi.org/10.1109/UIC-ATC-ScalCom-CBDCom-IoP.2015.191

Patel, S. J., & Bhoi, U. R. (2014). Improved priority based job scheduling algorithm in cloud computing using iterative method. *Proceedings - 2014 4th International Conference on Advances in Computing and Communications, ICACC 2014*, 199–202. https://doi.org/10.1109/ICACC.2014.55

Pavlo, A., Paulson, E., Rasin, A., Abadi, D. J., DeWitt, D. J., Madden, S., & Stonebraker, M. (2009). A comparison of approaches to large-scale data analysis. *SIGMOD-PODS'09 - Proceedings of the International Conference on Management of Data and 28th Symposium on Principles of Database Systems*, 165–178. https://doi.org/10.1145/1559845.1559865

Qi, Y., Pan, L., & Liu, S. (2022). A Lyapunov optimization-based online scheduling algorithm for service provisioning in cloud computing. *Future Generation Computer Systems*, *134*, 40–52. https://doi.org/10.1016/j.future.2022.03.037

Rezvani, M., Akbari, M. K., & Javadi, B. (2014). Resource allocation in cloud computing environments based on integer linear programming. *The Computer Journal*. https://doi.org/10.1093/comjnl/bxu024

Saleh, H., Nashaat, H., Saber, W., & Harb, H. M. (2019). IPSO task scheduling algorithm for large scale data in cloud computing environment. *IEEE Access*, *7*, 5412–5420. https://doi.org/10.1109/ACCESS.2018.2890067

San, C. C., Thwin, M. M. S., & Htun, N. L. (2019). Malicious software family classification using machine learning multi-class classifiers. *Lecture Notes in Electrical Engineering*, *481*, 423–433. https://doi.org/10.1007/978-981-13-2622-6_41

Saroha, V. K., & Rana, S. (2018). Implementing job scheduling approach in cloud environment. *IEEE International Conference on Power, Control, Signals and Instrumentation Engineering, ICPCSI 2017*, 2228–2234. https://doi.org/10.1109/ICPCSI.2017.8392113

Seifaddini, O. (2017). *Enhanced great deluge based job scheduling algorithms for grid computing*. [Doctoral Thesis].University Putra Malaysia.

Shabeera, T. P., Madhu Kumar, S. D., Salam, S. M., & Murali Krishnan, K. (2017). Optimizing VM allocation and data placement for data-intensive applications in cloud using ACO metaheuristic algorithm. *Engineering Science and Technology, an International Journal*, *20*(2), 616–628. https://doi.org/10.1016/j.jestch.2016.11.006

Shahid, M., Raza, Z., & Sajid, M. (2015). Level based batch scheduling strategy with idle slot reduction under DAG constraints for computational grid. *Journal of Systems and Software*, *108*, 110–133. https://doi.org/10.1016/j.jss.2015.06.016

Shang, Q. (2021). A dynamic resource allocation algorithm in cloud computing based on workflow and resource clustering. *Journal of Internet Technology*, *22*(2), 403–411. https://doi.org/10.3966/160792642021032202015

Shao, J., Wei, H., Wang, Q., & Mei, H. (2010). A runtime model based monitoring approach for cloud. *2010 IEEE 3rd International Conference on Cloud Computing*, 313–320. https://doi.org/10.1109/CLOUD.2010.31

Shrimali, B., & Patel, H. (2017). Multi-objective optimization oriented policy for performance and energy efficient resource allocation in Cloud environment. *Journal of King Saud University - Computer and Information Sciences*. 32(7), 860-869. https://doi.org/10.1016/j.jksuci.2017.12.001

Singh, A., Juneja, D., & Malhotra, M. (2017). A novel agent based autonomous and service composition framework for cost optimization of resource provisioning in cloud computing. *Journal of King Saud University - Computer and Information Sciences*, *29*(1), 19-28. https://doi.org/http://dx.doi.org/10.1016/j.jksuci.2015.09.001

Stillwell, M., Schanzenbach, D., Vivien, F., & Casanova, H. (2009). Resource allocation using virtual clusters. *2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid, Section III*, 260–267. https://doi.org/10.1109/CCGRID.2009.23

Sudarsan, R., & Ribbens, C. J. (2016). Combining performance and priority for scheduling resizable parallel applications. *Journal of. Parallel Distributed. Computing* , *87*, 55–66. https://doi.org/10.1016/j.jpdc.2015.09.007

Sun, Y., Xiao, Z., & Bao, D. (2010). An architecture model of management and monitoring on Cloud services resources. *2010 3rd International Conference on Advanced Computer Theory and Engineering(ICACTE)*, V3-207-V3-211. https://doi.org/10.1109/ICACTE.2010.5579654

Suthaharan, S. (2015). Machine learning models and algorithms for big data classification: Thinking with examples for effective learning. *Springer New York*.

Tang, X., Li, K., & Liao, G. (2014). An effective reliability-driven technique of allocating tasks on heterogeneous cluster systems. *Cluster Computing*. 17, 1413-1425. https://doi.org/10.1007/s10586-014-0372-1

Tapoglou, N., & Mehnen, J. (2016). Cloud-based job dispatching using multi-criteria decision making. *Procedia CIRP*, *41*, 661–666. https://doi.org/10.1016/j.procir.2015.12.081

Tchernykh, A., Schwiegelsohn, U., Alexandrov, V., & Talbi, E. (2015). Towards understanding uncertainty in cloud computing resource provisioning. *Procedia Computer Science*, *51*, 1772–1781. https://doi.org/10.1016/j.procs.2015.05.387

Tian, W., Xu, M., Chen, A., Li, G., Wang, X., & Chen, Y. (2015). Open-source simulators for cloud computing: Comparative study and challenging issues. *Simulation Modelling Practice and Theory*, *58*, 239–254. https://doi.org/10.1016/j.simpat.2015.06.002

Tong, Z., Chen, H., Deng, X., Li, K., & Li, K. (2020). A scheduling scheme in the cloud computing environment using deep Q-learning. *Information Sciences*, *512*, 1170–1191. https://doi.org/10.1016/j.ins.2019.10.035

Urgaonkar, R., & Kozat, U. (2010). Dynamic resource allocation and power management in virtualized data centers. *2010 IEEE Network Operations and Management Symposium* - NOMS 2010, Osaka, Japan, pp. 479-486, doi: 10.1109/NOMS.2010.5488484

Verma, M., Gangadharan, G. R., Narendra, N. C., Vadlamani, R., Inamdar, V., Ramachandran, L., Calheiros, R. N., & Buyya, R. (2016). Dynamic resource demand prediction and allocation in multi-tenant service clouds. *Concurrency Computation Practice and Experience*, *22*(6), 685–701. https://doi.org/10.1002/cpe

Verma, M., Gangadharan, G. R., Ravi, V., & Narendra, N. (2013). Resource demand prediction in multi-tenant service clouds. *2013 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM)*, 1–8. https://doi.org/10.1109/CCEM.2013.6684440

Vinothina, V. , Sridaran R., Ganapathi, P. (2012). A survey on resource allocation strategies in cloud computing. *International Journal Of Advanced Computer Science and Applications*, *3*(6), 97–104.

Wagle, S. S. (2016). Cloud service optimization method for multi-cloud brokering. *Proceedings - 2015 IEEE International Conference on Cloud Computing in Emerging Markets, CCEM 2015*, 132–139. https://doi.org/10.1109/CCEM.2015.25

Wang, L. (2011). *Cloud Computing: Methodology, systems, and applications*. *Taylor & Francis Group.* https://doi.org/10.1201/b11149-27

Wang, Y., Cao, S., Wang, G., Feng, Z., Zhang, C., & Guo, H. (2017). Fairness scheduling with dynamic priority for multi workflow on heterogeneous systems. *2017 2nd IEEE International Conference on Cloud Computing and Big Data Analysis, ICCCBDA 2017*, 404–409. https://doi.org/10.1109/ICCCBDA.2017.7951947

Wei, Y., Blake, M. B., & Madey, G. R. (2013). An operation-time simulation framework for UAV swarm configuration and mission planning. *Procedia Computer Science*, *18*, 1949–1958. https://doi.org/10.1016/j.procs.2013.05.364

Win, T. R., Yee, T. T., & Htoon, E. C. (2019). Optimized resource allocation model in cloud computing system. *2019 International Conference on Advanced Information Technologies, ICAIT 2019*, 49–54. https://doi.org/10.1109/AITC.2019.8920852

Witanto, J. N., Lim, H., & Atiquzzaman, M. (2018). Adaptive selection of dynamic VM consolidation algorithm using neural network for cloud resource management. *Future Generation Computer Systems*, *87*, 35–42. https://doi.org/10.1016/j.future.2018.04.075

Xhafa, F., Carretero, J., Dorronsoro, B., & Alba, E. (2009). A tabu search algorithm for scheduling independent jobs in computational grids. *Computing and Informatics*. *28*, 1001-1014.

Xu, Y., Li, K., He, L., & Truong, T. K. (2013). A DAG scheduling scheme on heterogeneous computing systems using double molecular structure-based chemical reaction optimization. *Journal of Parallel and Distributed Computing*, *73*(9), 1306–1322. https://doi.org/10.1016/j.jpdc.2013.05.005

Yakubu, I. Z., & Malathy, C. (2020). Priority based delay time scheduling for quality of service in cloud computing networks. *International Conference on Emerging Trends in Information Technology and Engineering, Ic-ETITE 2020*, 1–5. https://doi.org/10.1109/ic-ETITE47903.2020.379

Yang, X. S. (2019). Introduction to algorithms for data mining and machine learning. *Introduction to Algorithms for Data Mining and Machine Learning*, *Elsevier*. https://doi.org/10.1016/C2018-0-02034-4

Yousaf, M. M., & Welzl, M. (2014). Network-aware HEFT scheduling for grid. *The Scientific World Journal*, *2014*. https://doi.org/10.1155/2014/317284

Zhang, B., Zeng, Z., Shi, X., Yang, J., Veeravalli, B., & Li, K. (2021). A novel cooperative resource provisioning strategy for Multi-Cloud load balancing. *Journal of Parallel and Distributed Computing*, *152*, 98–107. https://doi.org/10.1016/j.jpdc.2021.02.003

Zhou, G., Tian, W., & Buyya, R. (2023). Multi-search-routes-based methods for minimizing makespan of homogeneous and heterogeneous resources in Cloud computing. *Future Generation Computer Systems*, *141*, 414–432. https://doi.org/10.1016/j.future.2022.11.031

Zhu, Z., & Fan, P. (2019). Machine learning based prediction and classification of computational jobs in cloud computing centers. *2019 15th International Wireless Communications and Mobile Computing Conference, IWCMC 2019*, 1482–1487. https://doi.org/10.1109/IWCMC.2019.8766558

Zuo, L., Dong, S., Shu, L., Zhu, C., & Han, G. (2018). A Multiqueue interlacing peak scheduling method based on tasks' classification in cloud computing. *IEEE Systems Journal*, *12*(2), 1518–1530. https://doi.org/10.1109/JSYST.2016.2542251