Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 5 (2023)

Innovative Approach to Dynamic Cloud Workflow Management and Scheduling for Big Data Applications

[1] Sreedevi Gangolu, [2] Dr. Bajrang Lal

[1]Research Scholar, Department of Computer Science and Engineering, Singhania University, Pacheri Bari, Jhunjhunu, Rajasthan.

[2]Professor, Department of Computer Science and Engineering, Singhania University, Pacheri Bari, Jhunjhunu, Rajasthan

Abstract

Most recent innovations in technology, such as cloud computing and distributed file systems, the problems that huge newline data has introduced are solved by in-memory computing and parallel computing. Computing resources for current pools may be distributed according to demand and scientific computing can benefit from the latest advancements in virtualization technology and cloud computing. Computing power, storage, platforms, and software applications are all part of the abstracted and virtualized resources that customers have access to over the internet through cloud computing. Efficient storage and processing of big data is necessary for knowledge and information extraction from it. Resources for processing data are not keeping pace with its exponential growth. Therefore, it is a tough newline task to manage and interpret information from enormous datasets. Computational time is proportional to the size of the dataset. In addition, the newline workflow has become more complex, with several subtasks that must be newline executed simultaneously or sequentially.

Keywords: *customers, internet, technologies. Improving, optimising, efficiency, cloud computing.*

Introduction

Integrating processes in a cloud computing setting is made easy using cloud workflow technologies. Improving and optimising business processes, increasing efficiency, achieving greater control over business processes, increasing flexibility in business processes, and enhancing customer service quality are all possible with cloud workflow management. Automating and streamlining cloud computing applications has never been easier than with cloud workflows, which facilitate application development, execution, management, and monitoring.

Traditional workflow methods and techniques may not be suitable for tackling related issues in cloud computing environments because of the cloud's dynamic, distributed, heterogeneous, and scalable nature. Instead, it is necessary to design novel approaches that align with the characteristics of cloud computing resources and applications. This entails investigating cloud computing workflow architecture, dynamic process models, resource management models, and dynamic scheduling algorithms. A growing number of research groups across the world are looking at potential uses for molecular data in cloud computing settings. The author tested two different cloudbased setups, comparing their respective cost and performance characteristics. According to their research, a cloud architecture opens up a plethora of administration duties for controlling cloud-based storage, compute, and networking resources. The authors opted for a subtle, unobtrusive method that works just with interlaced logs—a kind of data that is abundant in cloud environments. There may or may not be error messages in the logs, but the discrepancies discovered during analysis point to possible execution problems. To better schedule applications to resources in the cloud, the authors presented an improved particle swarm optimisation (IPSO) technique in. By distributing jobs among available resources, the IPSO algorithm reduces overall costs. The findings demonstrated that the enhanced algorithm outperformed the conventional particle swarm approach.

The practicality and efficacy of the suggested approach were shown by numerical instances and simulation trials. By using a software-only approach to deploying a dynamic software infrastructure, the authors of suggested an alternate architecture that works well with cloud computing. In order to address security and scalability concerns, the authors implemented this design to circumvent certain constraints. A conceptual model for the comprehensive management of all resources was created by the author in order to improve energy efficiency and reduce the carbon footprint of a cloud data centre (CDC). In order to provide the groundwork for future practical advancements, the author addressed the interplay between energy and dependability in regard to sustainable cloud computing. Model as a Service (Maas) is the name of the authors' innovative cloud computing architecture. An adaptable and efficient approach to assessing parameters' uncertainty and time-varying properties was laid forth in this work. Recently, there has been a lot of buzz around group data sharing in cloud settings. As cloud computing continues to gain traction, the issue of how to ensure safe and effective data exchange on the cloud is taking on more urgency.

An ant colony system (ACS)-based approach to virtual machine placement (VMP) was devised by the author in. The results demonstrated that, in comparison to conventional algorithms, OEMACS performs better. The cloud workflow scheduling approach is the backbone of any cloud workflow engine. Workflow management systems must first find a trustworthy service provider that satisfies the user's trusted domain quality of service requirements in order to operate workflow tasks in a cloud environment. The provider must then equitably distribute the virtual computing resources in their data centre to complete the workflow. The first step in running a cloud workflow is selecting an acceptable service provider and mapping the sub-workflows that can operate concurrently with the pertinent web services; the second step is optimising the sub-workflows by mapping them to the applicable virtual computing resources. The ability to access and pay for computer services whenever needed is the main differentiator between the cloud and conventional computing models. However, conventional workflow approaches and technology are illequipped to handle issues that develop in cloud workflow management because of the cloud's inherent autonomy, diversity, and constant change. Research into massive-scale data processing technologies is booming right now, with a plethora of important studies conducted both domestically and internationally. The authors discovered that research in the fields of biomedicine and health informatics is making more use of big data technologies in.

Massive amounts of patient data are being collected as a result of the adoption of EHRs, and next-generation sequencing technology has the capacity to process billions of DNA sequences every day. There have been several new findings and approaches announced in the last five years in the rapidly expanding area of big data applications in healthcare. The authors discovered in that by using centralized cloud administration and flexible resource allocation, streaming media application providers may significantly lower their operating costs. Based on each viewer's local memory, the writers of the study evaluated the optimum deployment problem (ODP).

Literature Review

Hafidha Al-Barashdi et al (2019) Research on big data has been getting a lot of attention as of late. Academic library Big Data analytics, on the other hand, face two main obstacles: the sheer complexity of the methods and algorithms used and the sheer amount, velocity, and diversity of data. The major objective of this research is to identify the methods and resources that academic libraries may use to analyse Big Data and to calculate the benefits that libraries can reap from this trend. Also, how can we get librarians involved with Big Data? That's one of the research issues our project aims to address. In Big Data, what new avenues of inquiry are on the horizon? When it comes to academic libraries and big data, what have researchers failed to cover? A thorough literature assessment of academic libraries' Big Data analysis over the last seven years was carried out to provide a much clearer picture of the benefits of Big Data in academic libraries and their future research directions. Academic libraries contained 37 articles pertaining to Big Data as a result of the search. The findings showed that even though there was a lot of study on the subject, very few studies addressed the implications of Big Data for academic libraries, namely the methods and instruments for analysis. We talk about how academic libraries can use Big Data and how it will change the way we conduct research in the future. Additionally, this report emphasizes how university libraries are leading the way in Big Data research.

Yannian Hu et al (2019) Cloud workflow management and scheduling research is critically needed due to the increasing prevalence and use of large data. The good news is that appropriate methodologies for efficient analysis do exist. The essay examines big data from several angles and comes to the following findings on smart cloud workflow scheduling and management: There is an average and a maximum reduction in reaction time of 58.26% when compared to the original Storm system. 23.18%; memory use jumped 88.7%, or 71.16% on average; and CPU resource utilisation jumped 17.96%, or 11.39% on average. Algorithms MOACO and CCA outperform GA algorithm when it comes to optimising the dynamic combination of web services, with MOACO outperforming CCA algorithm on average. Additionally, the article suggests a method for scheduling cloud workflows that makes use of smart algorithms and tweaks the apparent approach for combining resources provided by cloud services in order to accomplish two-tier scheduling of cloud workflow jobs. We have developed and enhanced three exemplary intelligent algorithms for scheduling optimisation after studying them: the ACO, PSO, and GA algorithms. It is quite evident that various algorithms provide substantially different optimum values for the same situation when tested with diverse scenarios. On the other hand, there is a strong agreement between the mean curve and the ideal solution curve.

Mainak Adhikari et al (2019) A new trend in the distributed environment is workflow scheduling, which is difficult since it must adhere to many different quality of service (QoS) requirements. Applications that tackle enterprise- or science-scale issues are received by the cloud in the form of workflows, which are sets of interdependent tasks. The methods for scheduling workflows in the cloud have been the subject of much research, and this article summarizes the main points. Based on their goals and execution model, this article categorizes and examines the features of different workflow scheduling approaches. Furthermore, new possibilities and needs for distributed process scheduling are emerging as a result of current technology breakthroughs and paradigm shifts like serverless and fog computing. Web applications, event-driven apps, and the Internet of Things (IoT) are some examples of background chores that serverless infrastructures are primarily built to handle. The Fog computing paradigm was created to tackle the growing resource needs and overcome the limitations of the cloud-centric Internet of Things. In light of these new developments in cloud computing, the essay goes on to talk about workflow scheduling.

Bin Zhang et al (2018) A software-as-a-service (SaaS) platform for intelligent big data is built using a lightweight cloud workflow system that is based on representational state transfer, according to this research. In order to increase the process's interaction and reaction time, the system enables quick creation and flexible deployment of the business analysis process, and it also allows the dynamic building and operation of an intelligent data analysis application. In order to enable users to do batch and streaming computing concurrently, the suggested system incorporates online-streaming analysis and offline-batch models. Users have the ability to customize a set of workflow procedures and use cloud capabilities for big data analysis. The research sheds light on the cloud workflow system's architecture, application modelling, customization, dynamic building, and scheduling. To improve efficiency, it is suggested to use a chain workflow foundation mechanism that combines many analytic components into one. The system's analytical capabilities are validated via the provision of four real-world application examples. The experimental findings demonstrate that the suggested system is capable of handling several users logging in at the same time and makes good use of data analysis methods. Network operators have adopted the suggested SaaS workflow solution, and it has proven successful for them.

Yong Zhao et al (2014) Both the academic and business communities are showing enormous interest in cloud computing. The phrase "Cloud Workflow" is used in this context to refer to the process of planning, carrying out and tracking the creation of extensive cloud-based scientific procedures. It also includes the data management and computer resource management required to make these workflows possible to execute Before delving into the implications of merging Clouds and processes, this study examines the disparity between these two supplementary technologies. Our reference system for managing scientific workflows in the cloud is then introduced, followed by a presentation of the main difficulties in supporting these workflows. We wrap off with sharing what we learned about cloud integration with Swift, a scientific workflow management system.

Workflow Scheduling Based On Novel Genetic Algorithm

A novel method for addressing problems, swarm intelligence draws on the social activities of insects and other animals. Specifically, ant colony optimization—a general purpose optimisation technique—has been the most researched and effective of the many methodologies and techniques inspired by ants. Cloud computing made use of the method for determining the shortest route. For optimal performance on the cloud, try using a genetic algorithm or an ant colony technique. Identifying the structure's interfaces that allow goods to interoperate at different levels is what the WfMC (Workflow Management Coalition) means when they say workflow. Scheduling workflows involves taking into account pre-defined criteria and then allocating workflow tasks to the resource pool that is most suited to complete them. The NP-complete issue is the main obstacle in workflow management systems.

Experimental Results

This section displays the intended measurements, input data, result outcomes, and analysis of the research outcomes. Performance parameters that are considered are makespan (lapse time), resource utilisation, and deadline hit. This study compares two heuristic algorithms: the Proposed Niche Genetic Algorithm (also known as Proposed GA) and the Genetic Algorithm. The results indicate that the Proposed Genetic algorithm generates a better scheduling solution when it is run via simulations. The metrics, makespan (lapse time), resource utilisation, and deadline hit were used to analyse the trial and test operations. We compare the workflow's individual tasks to input and output files of varying sizes.

Table 1: Evaluation of workflow scheduling algorithm using parameters of diversified nature

Algorithm	Resource Matrix	Makespan (Lapse Time in seconds)	Deadline Hit (%)	Resource Utilization (%)
GA	128 x 8	30	56	37
Proposed GA		24	69	51
GA	256 x 16	37	64	41
Proposed GA		27	78	62
GA	512 x 32	36	69	38
Proposed GA		28	83	61
GA	1024 x 64	45	73	46
Proposed GA		32	91	72

The results and data from the examination of many diverse factors, such as makespan (lapse time), deadline hit, and resource utilisation, are shown in Table 1. Maximum deadline hit, maximum resource utilisation, and minimum makespan are all features of the proposed GA.

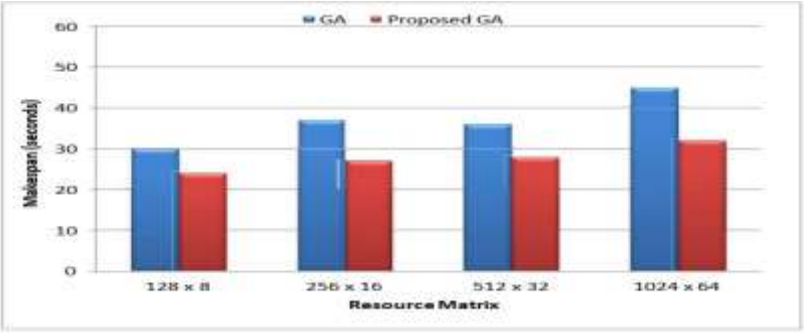

Figure 1 Comparison of Makespan (turnaround time)

Both the makespan for GA and the proposed GA methods were evaluated, as shown in Figure 1. When compared to the conventional genetic algorithm-based scheduler, the suggested technique has a shorter makespan.

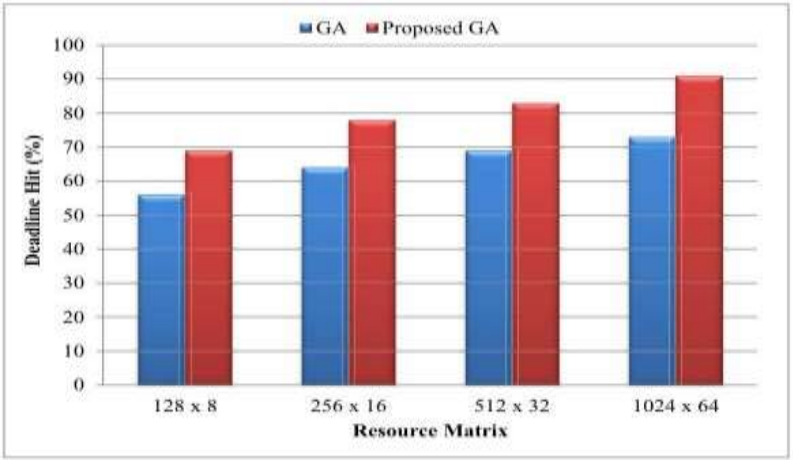

Figure 2: Comparison of Deadline hit

Figure 2 shows the contrast in relation to the deadlines met. In contrast to conventional genetic algorithms, the suggested improved GA has achieved a higher rate of deadline hit.

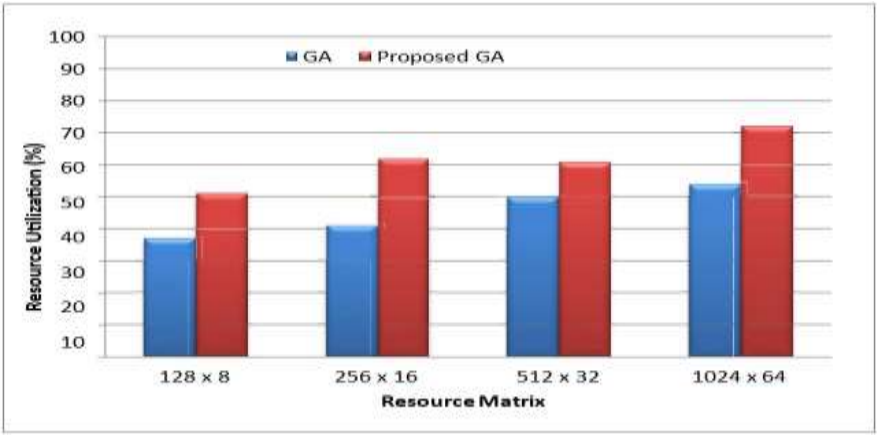

Figure 3: Comparison of Resource Utilization

Both the GA and the planned GA's resource utilisation parameters are shown in figure 3. The suggested approach perfectly matches the workload to the resources that are at hand. The suggested workflow scheduling method achieves more equitable resource utilisation than the genetic algorithm-based scheduling method.

Workflow Scheduling Based On Chaotic Whale Optimizer

Through cloud computing, users and businesses have access to a common pool of resources such as data storage, networks, computing power, and specialised applications. Generating an efficient workflow scheduling algorithm based on the client's needs might help lessen the unpredictability that comes with workload allocation and scheduling in a cloud environment. on the literature, you may find a number of optimisation algorithms that centre on Swarm based optimisation on the cloud.

For huge databases, Whale Optimisation is a new bio-inspired optimisation algorithm that takes its cues from the unique bubble net hunting technique used by humpback whales. Its use of adaptive approaches allows for a reduction in processing time even for very complicated situations. Because of its cheap computing cost, broad variety of applications, and relative simplicity, whale optimisation has recently gained popularity. Reactive voltage control, data mining, chemical engineering, pattern recognition, and environmental engineering are just a few of

the fields that have made use of Whale optimisation. Problems involving scheduling and job allocation that are NP-Hard have also been tackled using the Whale optimisation.

When it comes to scheduling, genetic algorithms are a great way to go. They work by looking at potential solutions, or schedules, for each job and then allocating resources accordingly. The fundamental idea behind genetic algorithms is the process of population generation in biology. This project aims to improve the whale optimizer for workflow scheduling problems by combining Chaotic Maps Mechanism with whale optimisation. Getting stuck in local optima is the biggest issue with Whale optimisation. Using chaotic maps to calculate and automatically adjust the optimisation algorithm's internal parameters might be an effective solution to this challenge and a way to break out of the local optima. This works well for difficult situations since the suggested method becomes better at finding the optimal answer as iterations go. Complex and multi-model goal functions may be optimized using the updated technique.

Experimental Results

In this part, we provide the experimental setup and comparative metrics that will be used to evaluate the proposed method on the cloud environment. Multiple metrics, including makespan (also known as lapse time or turnaround time), deadline hit, and resource utilisation, were used to assess the test runs. Each step of the process is reviewed and assessed in relation to different input and output file sizes.

Table 2: Evaluation of Workflow Scheduling Algorithm Using Parameters of Diversified Nature

Algorithm	Resource Matrix	Makespan (turnaround time in seconds)	Deadline Hit (%)	Resource Utilization (%)
WOS	128 x 8	43	58	45
CWOAS		35	70	56
WOS	256 x 16	47	66	51
CWOAS		38	79	62
WOS	512 x 32	48	69	49
CWOAS		38	74	61
WOS	1024 x 64	54	78	53
CWOAS		43	92	72

Using metrics like makespan (turnaround time), deadline hit, and resource utilisation, Table 2 compares and contrasts different scheduling algorithms. The values are computed using the resource matrix, which is used for various measures. Compared to schedulers based on whale optimisation, the proposed CWOAS is much more efficient.

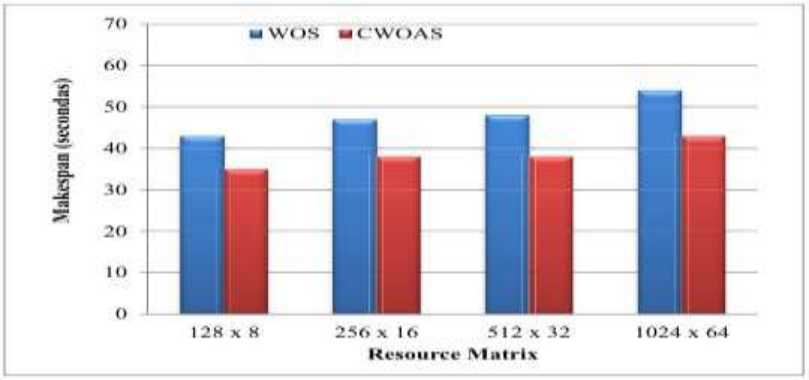

Figure 4: Comparison of Makespan

Both the WOS and the CWOAS, which is based on the proposed Chaotic Whale Optimizer Algorithm, have makespan diagrams in Figure 4. When comparing the proposed CWOAS scheduler technique to the WOS scheduler method, the makespan is lower in the former.

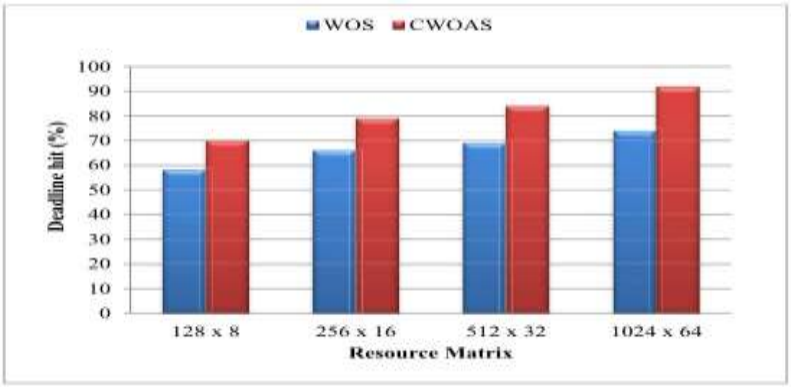

Figure 5: Comparison of Deadline hit

The comparison of the deadlines hit for WOS and CWOAS is shown in Figure 5. When compared to the WOS Scheduler, the proposed CWOAS approach has a better rate of completion.

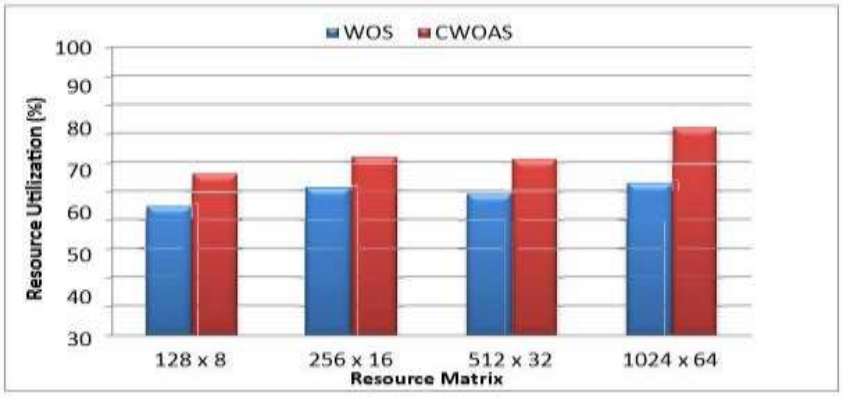

Figure 6: Comparison of Resource Utilization

For the purpose of comparing resource use, figure 6 is provided. The suggested schedulers assess the work assigned by the client computer. In comparison to the whale optimisation-based scheduler, the suggested CWOAS scheduler approach has balanced the jobs.

Conclusion

It is important to proceed with caution while scheduling workflows on the cloud. The effect of such breaches is growing as more and more organisations go to the cloud. Using the suggested evolutionary algorithm, chaotic whale optimisation methods, and antlion dynamic workflow scheduling algorithm, this study gives a summary of cloud computing. Learning how cloud-based workflow scheduling algorithms function is the main goal of this project. Cloud computing offers a wide range of services and may be implemented in many ways. Famous methods for scheduling processes in distributed systems include Antlion, the Whale Optimisation algorithm, and genetic algorithms.

References

[1] Barrett, E., Howley, E., Duggan, J.: Applying reinforcement learning towards automating resource allocation and application scalability in the cloud. Concur. Comput. Pract. Exp. 25(12), 1656–1674 (2013)

[2] Fakhfakh, F., Kacem, H.H., Kacem, A.H.: A provisioning approach of cloud resources for dynamic workflows. In: 8th IEEE International Conference on Cloud Computing, pp. 469– 476. IEEE (2015)

[3] Hu, Yannian & Wang, Hui & Ma, Wenge. (2019). Intelligent Cloud Workflow Management and Scheduling Method for Big Data Applications. 10.21203/rs.2.19246/v1.

[4] Jungmann, A., Kleinjohann, B.: Learning recommendation system for automated service composition. In: 2013 IEEE International Conference on Services Computing, pp. 97–104 (2013)

[5] Kaelbling, L.P., Littman, M.L., Moore, A.W.: Reinforcement learning: a survey. J. Artif. Intell. Res. 4, 237–285 (1996)

[6] Masdari, M., ValiKardan, S., Shahi, Z., Azar, S.I.: Towards workflow scheduling in cloud computing: a comprehensive analysis. J. Netw. Comput. Appl. 66, 64–82 (2016)

[7] Salman, M. S., Abdullah, M. K. J., & Sahid, S. N. Z. (2020). Big Data Analytic Concepts in Libraries: A Systematic Literature Review. International Journal of Academic Research in Progressive Education and Development, 9(2), 345–362.

[8] Smanchat, S., Viriyapant, K.: Taxonomies of workflow scheduling problem and techniques in the cloud. Future Gener. Comput. Syst. 52, 1–12 (2015)

[9] Wang, H., Zhou, X., Zhou, X., Liu, W., Li, W., Bouguettaya, A.: Adaptive service composition based on reinforcement learning. In: Maglio, P.P., Weske, M., Yang, J., Fantinato, M. (eds.) ICSOC 2010. LNCS, vol. 6470, pp. 92–107. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-17358-5_7

[10] Wei, Y., Pan, L., Yuan, D., Liu, S., Wu, L., Meng, X.: A cost-optimal service selection approach for collaborative workflow execution in clouds. In: 20th IEEE International Conference on Computer Supported Cooperative Work in Design, pp. 351–356 (2016)

[11] Wu, D., Rosen, D.W., Wang, L., Schaefer, D.: Cloud-based design and manufacturing: a new paradigm in digital manufacturing and design innovation. Comput. Aided Des. 59, 1–14 (2015)

[12] Yan, Y., Zhang, B., Guo, J.: An adaptive decision making approach based on reinforcement learning for selfmanaged cloud applications. In: 2016 IEEE International Conference on Web Services, pp. 720–723 (2016)

[13] Yong Zhao, Youfu Li, Ioan Raicu, Shiyong Lu, Wenhong Tian, Heng Liu (2014) Enabling scalable scientific workflow management in the Cloud.

[14] Zhang, B.; Yu, L.; Feng, Y.; Liu, L.; Zhao, S. Application of Workflow Technology for Big Data Analysis Service. Appl. Sci. 2018, 8, 591. https://doi.org/10.3390/app8040591