Semiconductor Optoelectronics, *Vol. 41 No. 12 (2022), 1686-1695*
https://bdtgd.cn/

Semiconductor Optoelectronics
ISSN: 1001-5868

DEADLINE-BASED VIRTUAL MACHINE OPTIMIZATION FOR IMPROVED RESOURCE UTILIZATION

Hemanta Dey [1], Dr. Akash Saxena [2]

[1]Research Scholar, Department of Computer Science , Mansarovar Global University, Sehore, M.P., India.

[2] Research Guide, Department of Computer Science , Mansarovar Global University, Sehore, M.P., India.

ABSTRACT

A Deadline-Based Virtual Machine Optimization Technique," focusing on the key aspects of the resource optimization method. The technique, called Deadline-Based Virtual Machine Optimization (DBVO), aims to improve resource efficiency by dynamically adjusting virtual machine configurations to meet application-specific deadlines. This study introduces a unique technique called Deadline Based VM Optimization (DBVO) for the optimization of Virtual Machines (VMs). In this suggested methodology, the major criteria considered are the time factor, namely the execution time and deadline. As time progresses, deadlines get shorter. Hence, in the suggested methodology, the imminent deadline is taken into account, and only the necessary virtual machines (VMs) are started. This results in minimal virtual machine startup and optimal resource use. The findings gained in this study are compared with many current methodologies.

Keywords: - Virtual Machine, Deadline, Strategy, Technique & Approach.

INTRODUCTION

Optimizing Virtual Machines (VMs) has become an indispensable component of modern computing ecosystems, particularly in cloud computing environments. As the demand for scalable, flexible, and efficient computing resources continues to grow, virtualization technologies and their optimization have emerged as critical tools in addressing the complex challenges of resource allocation, performance management, and cost-effectiveness. This essay delves into the multifaceted domain of virtual machine optimization, aiming to explore the fundamental concepts, techniques, and implications associated with this ever-evolving field. Through an in-depth analysis, we will navigate the various facets of VM optimization, from its historical origins to its current state, shedding light on its significance in contemporary computing landscapes and providing insights into the promising future it holds.

The advent of virtualization technology marked a paradigm shift in the way computing resources are provisioned and managed. Virtual Machines, often abbreviated as VMs, are selfcontained software instances that emulate physical hardware. By decoupling the operating system and applications from the underlying hardware, VMs enable more efficient utilization of computing resources, making it possible to run multiple operating systems and applications on a single physical server. This capability revolutionized data centers, allowing for greater flexibility and scalability while significantly reducing hardware costs.

However, the flexibility offered by VMs also introduced new challenges. Efficiently allocating resources, ensuring consistent performance, and optimizing the use of physical infrastructure became paramount. Virtual machine optimization emerged as a solution to these challenges, focusing on strategies to enhance resource utilization, improve performance, and reduce operational costs. Throughout this essay, we will dissect the evolution of VM optimization, explore the methods and tools used in this field, and assess its implications in the ever-evolving landscape of cloud computing, data centers, and enterprise IT environments.

To comprehend the significance of virtual machine optimization, it is crucial to trace its historical development. Virtualization itself dates back to the 1960s when IBM introduced the concept of virtual machines on their mainframe systems. However, the modern era of virtualization, often associated with the rise of VMware in the early 2000s, marked a turning point. VMware's introduction of x86 virtualization technology, which allowed multiple operating systems to run concurrently on a single physical server, laid the foundation for the virtualization revolution. This innovation sparked a transformation in data centers and cloud computing, fostering the need for advanced optimization techniques.

As the adoption of virtualization technology expanded, the challenges of managing VMs also grew. Virtual machine sprawl, a phenomenon where organizations created more VMs than they could effectively manage, became a common issue. Without proper optimization, VMs would often remain over-provisioned, resulting in underutilized hardware resources and increased operational costs. This dilemma prompted the development of various optimization techniques and tools aimed at streamlining resource allocation, enhancing performance, and reducing waste.

One of the primary areas of focus in virtual machine optimization is resource allocation. Efficiently distributing CPU, memory, storage, and network resources among VMs is essential to ensure optimal performance. Over allocation or under allocation of resources can lead to issues such as resource contention or suboptimal application performance. Techniques like Dynamic Resource Allocation (DRA) and Quality of Service (QoS) policies have been developed to address these challenges, enabling VMs to dynamically adjust their resource allocations based on workload demands.

In addition to resource allocation, performance optimization is another key aspect of VM management. Organizations often rely on virtualization to run mission-critical applications, and any performance degradation can lead to significant business disruptions. Performance optimization strategies encompass techniques such as load balancing, workload migration, and performance monitoring. These methods ensure that VMs operate at their peak efficiency, minimizing bottlenecks and latency, and providing consistent and reliable performance for users.

Cost optimization is yet another critical facet of virtual machine optimization, especially in cloud computing environments. By efficiently managing resources and ensuring that VMs only consume what they need, organizations can reduce operational costs significantly. Techniques such as auto-scaling and cost monitoring help control expenses while maintaining high levels of service. This cost-effective approach aligns with the principles of cloud computing, where the pay-as-you-go model encourages resource optimization to minimize expenses.

The tools and technologies developed for virtual machine optimization have evolved in tandem with the growing demands of the industry. Virtualization platforms, like VMware vSphere and Microsoft Hyper-V, have integrated optimization features to simplify resource allocation and management. Additionally, specialized management tools and software, such as VMware vRealize Suite and Citrix XenCenter, have been developed to provide more advanced optimization capabilities.

As cloud computing has become a cornerstone of modern IT infrastructure, virtual machine optimization has gained even greater prominence. Cloud providers, such as Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform, have integrated sophisticated optimization features into their services. These features include auto-scaling, predictive scaling, and intelligent resource allocation to ensure that VMs running in the cloud are cost-efficient and performant.

While virtual machine optimization has made significant strides in addressing resource allocation, performance, and cost, the field continues to evolve. The rise of containerization technologies, such as Docker and Kubernetes, has introduced new challenges and opportunities for optimization. Containers offer lightweight, fast-starting environments that are particularly well-suited for micro services architectures, but optimizing resource allocation and performance in containerized environments requires a different approach.

Furthermore, the emergence of server less computing, represented by platforms like AWS Lambda and Azure Functions, has created a paradigm shift where developers are abstracted from the underlying infrastructure entirely. In this context, optimization focuses on the efficient execution of code functions rather than managing virtual machines. Serverless optimization strategies are still evolving, and they emphasize granular resource allocation and cost-effective execution.

Looking to the future, virtual machine optimization will continue to adapt and expand to meet the evolving needs of IT environments. AI and machine learning technologies are being integrated into optimization processes to provide predictive analytics and autonomous management. As virtualization technologies, containers, and server less computing continue to coexist and complement each other, the role of optimization will become increasingly complex, requiring holistic approaches that consider multiple layers of the stack. Virtual machine optimization stands as a critical discipline in the realm of computing, with a rich history and a promising future. It has evolved alongside virtualization technology, providing solutions to the challenges of resource allocation, performance management, and cost-effectiveness. As cloud computing, containers, and server less paradigms reshape the IT landscape, the field of virtual machine optimization will remain essential in ensuring the efficient use of computing resources and the delivery of high-performance, cost-effective services. Its enduring significance in the ever-changing world of technology underscores the need for continuous research, innovation, and adaptation in this dynamic domain.

REVIEW OF LITERATURE

Gupta, Praveen & Singh, Chour. (2021). Cloud computing (CC) is a burgeoning field that is now seeing growth and progress in both the commercial sector and the fields of information technology (IT) and mobile computing. Instead of opting for direct acquisition, the use of resources such as software, central processing units (CPUs), memory, and input/output (I/O) equipment is utilized, and the billing is calculated based on their relative usage. The significant expansion of cloud computing (CC) has resulted in a substantial rise in energy consumption, since data centers include a diverse array of computer systems. Cloud service providers are now actively seeking ecologically friendly strategies to reduce energy usage and minimize carbon emissions. The main focal point of interest revolved on task planning, with a specific focus on enhancing resource allocation and reducing energy use. This study presents an innovative methodology that use deep learning techniques to address the inherent challenges associated with task planning, with a specific focus on enhancing energy efficiency. The aim of this study is to create algorithms that are optimized to effectively overcome the limits found in existing approaches for task planning. The approach being proposed seeks to address two fundamental elements: the temporal requirements for completion and the efficient use of resources. Furthermore, it aims to conform to a defined approach in order to set a certain time period. Within a heterogeneous setting, we undertake an examination of the methodology used to assess the effectiveness and efficiency.

Saxena, Deepika & Singh, Ashutosh et al., (2021) The inherent flexibility of cloud resources allows cloud users to effectively manage their resource needs by scaling up or down as required. The objective of this framework is to use proactive prediction techniques to anticipate resource utilization on servers, followed by the equitable distribution of workload. The use of technology facilitates the saving of energy via the reduction of operating servers, the optimization of virtual machine migrations, and the maximization of resource utilization. A novel approach has been developed and deployed to forecast the allocation of online resources on virtual machines (VMs). In addition, this research introduces innovative approaches for the placement and migration of virtual machines (VMs) with several objectives. The primary goals are to reduce network traffic and optimize power use inside datacenters. After conducting a comparative analysis between the recommended framework and current state-of-the-art methodologies, it becomes apparent that the former demonstrates superiority in several performance parameters. The OP-MLB framework exhibits a substantial improvement in power conservation, with an increase of up to 85.3% in comparison to the Best-Fit approach.

Song, Fei & Huang et al., (2014) This paper primarily examines the energy efficiency and scalability challenges faced by modern data centers, using the prominent methodology of convex optimization theory. A novel approach to deploying virtual machines (VMs) is proposed as a potential solution for mitigating the issues discussed before in the context of large-scale systems. The primary subject of this discourse is on the approach used for virtual machine (VM) deployment, which conforms to the server limitations of physical machines. The process begins by first scrutinizing the definitions of VM placement fairness and utility function. Subsequently, the virtual machine (VM) placement is approached as an optimization problem, considering the inherent relationships and traffic across VMs. In light of the existing structural disparities observed in recently proposed data center designs, our objective is to undertake a thorough comparative analysis. This study aims to evaluate the impact of network architectures, server limitations, and application dependencies on the potential performance improvements that can be attained through optimization-based virtual machine (VM) placement. When juxtaposed with preceding systems, this research reveals improvements in performance from several perspectives. The aforementioned benefits include a decrease in the use of tangible hardware, a reduction in the expenses associated with intercommunication among virtual machines, and enhancements in the energy efficiency and scalability of data centers.

Patel, Suhradam & Bhujade et al., (2013) Cloud computing is a paradigm that enables the delivery of IT resources in the form of services via the Internet. The future of application deployment is increasingly reliant on the Quality of Service offered by service providers. Virtualization serves as the foundation for several functionalities, offering several advantages such as mobility, optimal usage of hardware resources, simplified maintenance, cost-effectiveness, and disadvantages such as potential performance decrease when virtualized resources are not evenly assigned. This article categorizes user requests based on the kind of deadline, distinguishing between unrestricted and tight deadlines.

Khalid, Omer & Maljevic et al., (2010) The advent of virtualization technology has facilitated the separation of programs from the underlying hardware, resulting in many advantages such as enhanced mobility, improved control over the execution environment, and increased isolation. The use of this technology has gained significant acceptance in both scientific grid computing and commercial cloud computing environments. Virtualization, despite its advantageous aspects, imposes a performance drawback that may be substantial for systems operating in unpredictable environments, such as High Performance Computing (HPC) applications. These applications are characterized by stringent task deadlines and dependencies on other jobs prior to execution. The primary challenge is the need to reconcile the performance demands of a given operation with the performance capabilities provided by virtualization technology, particularly when executing these processes inside virtual machines. This study introduces a unique methodology for optimizing task deadlines in virtual machines. The proposed strategy involves the development of a deadline-aware algorithm that is capable of responding to real-time job execution delays. Furthermore, the algorithm dynamically optimizes jobs to ensure their compliance with deadline commitments. The methodologies used in our study draw upon principles from both signal processing and statistical methods. We provide a comprehensive analysis of their respective performances, including an evaluation of their effects on the utilization rate of hardware resources. These findings are reported in the subsequent sections of this work.

Naha, Ranesh & Garg et al., (2019) The existing literature on Fog computing mostly focuses on resource supply, neglecting the dynamic nature of users' demand. In order to tackle the challenge of meeting time-sensitive dynamic user needs, we suggest the implementation of resource allocation and provisioning algorithms that use a hybrid and hierarchical approach, including resource ranking and the supply of resources. The algorithms under consideration are assessed inside a simulated environment, whereby the CloudSim toolset is expanded to replicate a practical Fog setting. The experimental findings suggest that the performance of the suggested algorithms surpasses that of current algorithms in terms of total data processing time, instance cost, and network latency, particularly when the number of application submissions increases. The average processing time and cost exhibit a reduction of 12% and 15% respectively, in comparison to the currently available alternatives.

Hwang, Eunji & Kim et al., (2012) Cloud computing provides a diverse array of services referred to as Anything as a Service (XaaS), which allows for the creation and implementation of different applications on the Cloud without the need to consider specific platforms. Cloud computing systems provide virtual computers that may be easily used to develop data-driven applications. This optimization is aimed at minimizing the associated costs while ensuring that the deadlines for completion are not compromised. In this study, we provide two distinct resource provisioning methodologies: one that relies on specified pricing principles and another that is centered upon the concept of deadline-aware tasks packing. During the process of conducting simulations, it is important to thoroughly assess and scrutinize them using a range of different methodologies.

Wang, Xiaohui & Haoran et al., (2018) Allocating virtual resources in the cloud is becoming more important. Using virtualization technology in the data center, virtual computers are deployed on real machines to perform a variety of users' tasks. However, in this procedure, it is necessary to take into account the overall load distribution, energy consumption, and resource use of physical equipment. As a result, two models are set up, one for each optimization goal. The primary model is constructed so as to have as little unbalanced load as possible. The second model is constructed to make the most efficient use of available resources while using as little energy as possible. We present a novel approach termed resampled binary particle swarm optimization (RBPSO) to improve the placement of virtual machines. RBPSO is an extension of BPSO that incorporates resampling, mutation, and a minor vibration mechanism to increase the algorithm's global search capability by preserving population variety and decreasing unnecessary computation. The RBPSO is then applied to the cloud computing deployment issue of virtual machines. The results of the trials validate the reasonableness of the proposed model, demonstrating that RBPSO outperforms both BPSO and GA.

Kong, Weiwei & Lei et al., (2016) A novel adaptive VM resource scheduling method based on auction mechanism is proposed to address the issue of scheduling virtual machines (VMs) in the cloud, taking into account a variety of aspects such as network bandwidth and auction deadline. Before the competition deadline, the bids of individual customers are ordered.

Second, virtual machine (VM) resources are set in accordance with the cloud service providers' minimal charges once the client group has been checked. Last but not least, the final payment price may be calculated by factoring in typical payment levels and competitive payment levels, ensuring that the requested client tasks can be accomplished using the available VM resource. The simulation experiments validate that the suggested method may improve cloud service quality, cloud service provider revenues, and VM resource consumption.

RESEARCH METHODOLOGY

Tasks with a deadline are managed using a novel strategy. Following submission of tasks in deadline ascending sequence, the deadline will reduce as each job is completed. To track the jobs whose due dates are approaching faster than the next deadline, we use the nxt deadline variable. Once the necessary tasks have been identified, additional resources may be established and assigned. The deadline-based tasks are handled in the current article using a dynamic method.

DATA ANALYSIS AND INTERPRETATION

The section compares the suggested methods to those already in use. The same values for the parameters are used as in previous methods.

- **Comparison of Results**

This study presents a comparative analysis of the suggested DBVO (Deadline-Based Virtualization Optimization) technique and the NDF (Nearest Deadline First) strategy. The NDF technique included the execution of tasks only based on deadlines that were in close proximity to the finish, without any optimization of resources. In the suggested technique, the fundamental purpose of the DBVO approach is to effectively manage jobs with time constraints while optimizing resource use. As the passage of time occurs, the allotted time for completing activities diminishes.

In the first scenario, the comparison is made between the numbers of resources begun during execution, with the objective of minimizing resource waste by initiating the smallest amount of resources. Figure 1 illustrates a comparative analysis of the quantities of resources that were begun throughout the execution process.

In this graphical representation, horizontal lines are used to indicate the quantity of tasks, while the vertical line represents the allocation of resources required for the timely execution of these activities. The findings are shown in Figure 1, which demonstrates that the suggested technique effectively accomplishes tasks within the designated timeframe while also minimizing resource use from the pool.

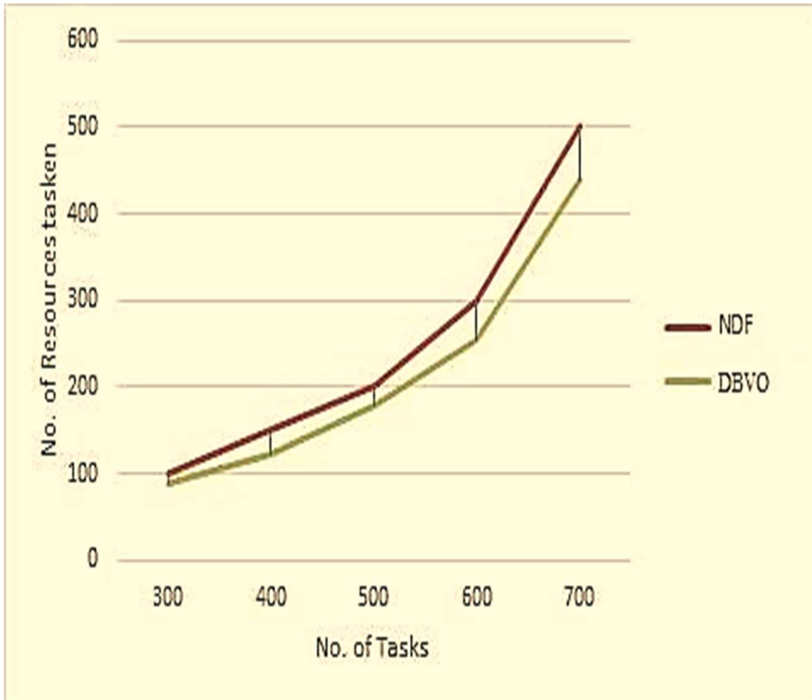

Figure 1: No. of Resources Compared

In addition, the DBVO technique also took into consideration another parameter, namely make-span. The objective of any scheduling strategy is to minimize the makespan. The minimization of makespan is achieved by balancing the load, since this results in an increase in the makespan when compared to the suggested strategy. In order to enhance the credibility of the suggested methodology, an analysis was conducted. The findings are shown in Figure 2.

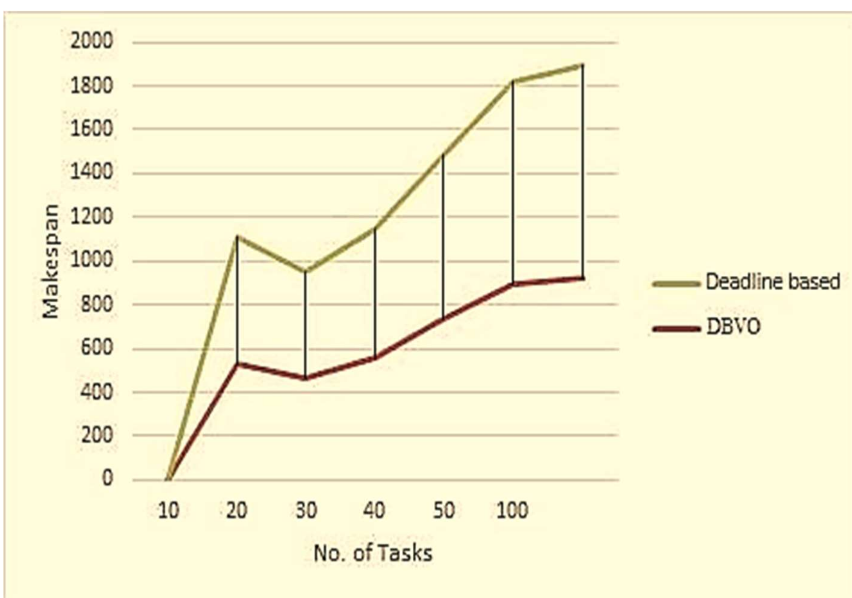

Figure 2: Comparison of Makespan

According to Figure 2, the suggested technique exhibits a lower make-span in comparison to the present approach.

CONCLUSION

A potential strategy to improve resource consumption and overall performance in cloud computing settings is one that is based on the optimization of virtual machines according to a deadline. This technique assures effective resource usage, increased performance, and a greater quality of service since it allots resources based on the deadlines that are particular to each applications.

Not only does this avoid resource contention and underutilization, but it also makes it possible for cloud providers to guarantee constant and predictable performance, which is very helpful for applications that have rigorous needs. In today's dynamic and demanding computing market, the adoption of deadline-based optimization may lead to more cost-effective and sustainable resource management, which is beneficial to both providers and end-users. As cloud computing continues to expand, adopting deadline-based optimization can lead to these benefits.

REFERENCES

1. Gupta, Praveen & Singh, Chour. (2021). An Optimized Energy-Efficient Task Scheduling Algorithm based on Deep Learning Technique for Energy Consumption.

2. Meng, Xiaoqiao & Isci, Canturk & Kephart, Jeffrey & Zhang, Li & Bouillet, Eric & Pendarakis, Dimitrios. (2010). Efficient resource provisioning in compute clouds via VM multiplexing. Proceeding of the 7th International Conference on Autonomic Computing, ICAC '10 and Co-located Workshops. 11-20. 10.1145/1809049.1809052.

3. Nasr, Aida & Chronopoulos, A. & El-Bahnasawy, Nirmeen & Attiya, Gamal & El-Sayed, Ayman. (2019). A novel water pressure change optimization technique for solving scheduling problem in cloud computing. Cluster Computing. 22. 10.1007/s10586-018-2867-7.

4. Saxena, Deepika & Singh, Ashutosh & Buyya, Rajkumar. (2021). OP-MLB: An Online VM Prediction based Multi-objective Load Balancing Framework for Resource Management at Cloud Datacenter. IEEE Transactions on Cloud Computing. PP. 1-1. 10.1109/TCC.2021.3059096.

5. Patel, Suhradam & Bhujade, Rakesh & Ullah, Asif & Kathrotia, Sugney. (2013). Deadline based Virtual Machine Provisioning to improve QoS in Cloud Computing. 219-224. 10.1109/ICACCI.2013.6637174.

6. Shin, SaeMi & Kim, Yena & Lee, SuKyoung. (2015). Deadline-guaranteed scheduling algorithm with improved resource utilization for cloud computing. 2015 12th Annual IEEE

Consumer Communications and Networking Conference, CCNC 2015. 814-819. 10.1109/CCNC.2015.7158082.

7. Khalid, Omer & Maljevic, Ivo & Anthony, Richard & Petridis, Miltos & Parrott, Kevin & Schulz, Markus. (2010). Deadline Aware Virtual Machine Scheduler for Grid and Cloud Computing. Advanced Information Networking and Applications Workshops, International Conference on. 85-90. 10.1109/WAINA.2010.107.

8. Naha, Ranesh & Garg, Saurabh & Chan, Andrew & Battula, Sudheer. (2019). Deadline-based dynamic resource allocation and provisioning algorithms in Fog-Cloud environment. Future Generation Computer Systems. 104. 10.1016/j.future.2019.10.018.

9. Hwang, Eunji & Kim, Kyong Hoon. (2012). Minimizing Cost of Virtual Machines for Deadline-Constrained MapReduce Applications in the Cloud. 130-138. 10.1109/Grid.2012.19.

10. Wang, Xiaohui & Haoran, Gu & Yue, Yuxian. (2018). The optimization of virtual resource allocation in cloud computing based on RBPSO. Concurrency and Computation: Practice and Experience. 32. 10.1002/cpe.5113.

11. Kong, Weiwei & Lei, Yang & Ma, Jing. (2016). Virtual Machine Resource Scheduling Algorithm for Cloud Computing Based on Auction Mechanism. Optik - International Journal for Light and Electron Optics. 127. 10.1016/j.ijleo.2016.02.061.

12. Rawat, Pradeep & Dimri, Priti & Saroha, Gyanendra. (2018). Virtual machine allocation to the task using an optimization method in cloud computing environment. International Journal of Information Technology. 12. 1-9. 10.1007/s41870-018-0242-9.

13. Ding, Youwei & Qin, Xiaolin & Liu, Liang & Wang, Taochun. (2015). Energy efficient scheduling of virtual machines in cloud with deadline constraint. Future Generation Computer Systems. 50. 10.1016/j.future.2015.02.001.

14. Rodríguez, Maria & Buyya, Rajkumar. (2014). Deadline Based Resource Provisioningand Scheduling Algorithm for Scientific Workflows on Clouds. Cloud Computing, IEEE Transactions on. 2. 222-235. 10.1109/TCC.2014.2314655.