International Journal of Computing and Digital Systems
2024, VOL. 17, NO. 1, 1–10
http://dx.doi.org/10.12785/ijcds/XXXXXX

Resource Utilization Performance of Complex Workflows on the Public Cloud: A Simulation-Based Approach

Faris Llwaah[1]

[1]*Department of Cyber Security,College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq*

Abstract: Public cloud computing has emerged as a popular computing platform. To benefit from its pay-as-you go billing model, it is important to allocate enough public cloud resources to execute complex scientific workflows promptly in an environment consisting of various virtual machines. By provisioning more resources than required, applications will finish in a short amount of time, resource provisioning costs may significantly increase. In practice, cloud users employ cost-benefit analysis to determine the optimal resource provisioning strategy or general guidelines, or they use optimization-based methods to predict application execution time and proactively allocate public cloud resources to run complex workflows. The effectiveness of these methods can be greatly influenced by the burden associated with accessing cloud resources due to high costs, which can impact the accuracy of prediction models, application complexity, as well as the available budget. Therefore, previous studies have indicated the performance of these strategies.
This research contributes by presenting a simulation tool that employs empirical rules for self-allocating public cloud resources to execute complex workflows. The suggested simulation platform (WorkFlowSim) includes specific features that assist in estimating the time needed for workflow execution. We have approved three workflows, each with different structures and sizes. One of these types contains real data from previous operations on the Microsoft Azure cloud, which helps ensure results and gain more credibility. The main goal of the study is to normalize a platform to accurately and efficiently determine cloud resource usage predictions. Therefore, as an achievement verified by experiments the minimum squared error (MSE) of the estimated runtime has calculated to be 454.1161, the minimal predict value when the resource count is 12, and is 100% consistent with the actual execution.

Keywords: Cloud Computing, Public Cloud Computing, Workflow, WorkflowSim, Azure Cloud, Resource Utilization Prediction, Resource Provisioning

1. INTRODUCTION

The ecosystem of big data is rapidly growing due to its efficiency in managing, processing, and analyzing large data sets. The interconnected processes needed to manage a task or multiple tasks are generally known as workflows, which play a vital role in streamlining operations within organizations by automating and monitoring processes, ensuring that tasks are completed efficiently and on time [1], [2]. Scheduling workflows in a dynamic computing environment is a difficult task that requires the optimization of numerous variables in order to achieve a good results. One critical factor is the efficient utilization of available resources to guarantee that workflows are executed within the specified time frame, as well as to reduce costs and prevent congestion.

Moreover, workflows have the ability to significantly improve system performance when they are deployed in the cloud. Workflow applications may effectively decrease execution time, improve resource allocation, and enable efficient interaction amongst distributed processes by utilizing the scalability and flexibility available via cloud platforms [3].

Additionally, an important factor that impacts the overall performance of the cloud computing is the use of resources in complex workflow. By guaranteeing that computing tasks are completed in parallel, efficient resource allocation minimizes time and maximizes the utilization of available resources. Consequently, this results in enhanced system performance, reduced costs, and raised output, therefore, guaranteeing the provision of high-quality services (QoS) [4].

As a scope of big data continues to emerge, we cannot overlook the importance of optimizing big data workflows. Also the ongoing growth of data and the increasing complexity of data driven applications, the resources management of the workflow is becoming increasingly difficult for gaining meaningful insights and ensuring the success of data-driven applications. Therefore, these reasons lead to an

E-mail address: f.llwaah@uomosul.edu.iq

expansion in research for big data tools that can handle large datasets with speed and efficiency. Workflows play a vital role in managing and processing this data [5]. The main impetus of this paper is to propose an alternative approach to evaluate and predict the necessary resources to complete the work of such applications on cloud computing.

Yet, emerging budget-constrained application service level agreements (SLAs) often contract for providing cloud resource utilization performance in terms of multi-criteria mapping that is dynamically determined by different cloud resource costs [6]. As a result, the primary challenge is that work demands often change over time, making it difficult to predict and optimize during operational periods in the real cloud [7]. Therefore, service providers are encouraged to reuse published modeling studies of resource utilization costs to ensure satisfactory savings in time and expenses within agreed response times. However, employing accurate measurements for modeling public cloud resources presents several issues: (1) results depend on how they are calibrated; (2) without a comprehensive list of criteria, research studies are often repeated due to the vast diversity, making cost comparisons nearly impossible; and (3) inconsistent measurements put service providers at risk of losing contracts.

As noted in the era of big data, workflow management systems are utilized to configure application workflows for complex data processing operations across various fields, including scientific research, industry, and sensor technology. Key features of these workflow management systems include transparent execution, rapid development, and process simulation. Generally, a simulation-based approach provides a cost-effective and low-risk prediction of the proposed workflow model before committing actual investment in resources [8].

This work leverages workflow modeling to execute independent big data processing tasks on large-scale resource infrastructures with simulations. The proposed system allows for experimentation in resource allocation for scaling up or down, rapid data replication for task sub-entries without incurring financial costs, and the option to use either private or public cloud environments as primary tools for fair and economical resource allocation and specification.

The primary contributions of this work are: This is a novel approach that is different from previous approaches because it is free and can be repeated infinitely at no charge. This approach also mitigates the consequences of over-provisioning, i.e., eliminating increased costs and the waste of resources. It also helps mitigate the consequences of under-provisioning, i.e., resolving service level agreement (SLA) violations.

This paper is organized as follows:

Section 2 presents our motivation to improve the performance of cloud computing through simulation platform. In Section 3, we discuss a key of previous works. Section 4 displays the methodology of research. In Section 5, we describe the structure and concepts of the simulation platform that is being chosen. Section 6 outlines the chosen workflow for our experiments. Section 7 is dedicated to illustrating our experiments and presenting the results. Section 8 includes the future work and conclusions.

2. Motivations to improve cloud computing performance using simulation

The simulation can model the behavior of organizations in detail under different conditions in a way that is often not feasible or possible in the real world. So, the usage of simulation in cloud computing research has several advantages, e.g., it may be used as a new alternative to an online testing account, an opportunity to evaluate different strategies and options, and modeling different (complex) situations before applying optimization algorithms. Despite the ability of simulation to replicate real-world environments, some people argue that the results of a simulation study are always debatable and that the reader never conceives that the result is soundly credible. Xiaolong Xu et al. (2020) [9], illustrated that the application of simulations should be a balancing act among the following trade-offs: time and cost vs. accuracy, quantity, and richness of available data.

As a result, one of the significant limitations that are often opined by researchers in cloud computing is the potential risk of a mismatch between the outcomes of a simulated scenario and those that would have emerged if the study was conducted over real cloud systems. The work [10] mentioned that simulation is considered to be a sound scientific method in the context of cloud computing, even though some researchers continue to question its reliability. Also, different problems ranging from reliability to verifiability, validity of results, and investment regarding time, cost and quality remain to be issues that need to be addressed in simulation studies in cloud computing research. The used method must be chosen depending on the goal of the study. Moreover, it is emphasized that simulation studies should always be carried out after a specific period of development and be preceded by a few case studies and/or some theoretical studies.

A. Advantages of Using Simulation

The main advantages of using simulation in improving the performance of cloud computing include reducing costs and the amount of time necessary to test and develop systems. This also benefits in lowering the risk and expense associated with the experimental changes to the cloud computing environment. Moreover, simulation enables the allocation of tasks and improves the utilization of computing resources in the cloud, therefore ensuring both their efficacy and productivity.

As we mentioned above, the benefits of simulation create outlines for potential future research in cloud computing. Because our research addresses the issue of resource utilization in the cloud computing environment, this is one of the reasons to select simulation for related research.

3. Related Work

Complex workflows require an immense amount of computational resources, and the public cloud environment can provide massive on-demand resources to handle such computational needs at a low cost. Numerous studies have been conducted to enhance the resource utilization and performance of complex workflows in the cloud. Although much progress has been made, the development of new methods to improve the scalability of workflows lags in these works, and complicated workflow research is relatively rare. To efficiently address complex workflows in large-scale cloud environments, a clear understanding of current related research on complex workflows is greatly needed. A systematic literature review, or meta-analysis, is a method of revitalizing and allowing the current understanding or comprehension of a complicated operation and administration by rigorously detecting, selecting, and summarizing relevant earlier investigations [11], [12].

A. Resource Management and Optimization in Cloud Computing

Several researches have concentrated on using mongrel algorithms to improve resource management in cloud computing systems. For instance, the study [13] under discussion is a novel approach that integrates machine learning with task scheduling in order to optimize the efficiency of cloud resources. This approach highlights the continuous adaptation to increasing tasks, which leads to improved resource allocation and cost efficiency. The authors in the research [14], made an effort to simulate the flow of big data generated using a Business Process Modeling Notation (BPMN) framework to optimize cloud resources.

The framework's capabilities were demonstrated using real data on workflows. To achieve this, a data-specific framework was developed to evaluate performance and was later tested on workflows consisting of 13, 52, and 104 tasks. The results indicated that the framework is suitable for evaluation and can be utilized for big data processing operations. Consequently, it distributed the total execution time across all tasks and improved the deployment of cloud resources efficiently and economically, providing crucial insights for decision-makers regarding their core business processes.

The study [15] was done to consider the challenges related to predicting how much resource is used in order to address the issue of an over-provisioning and under-provisioning. Over-provisioning causes excessive energy consumption and increased costs, while under-provisioning can lead to violations of Service-level agreements (SLAs) and a decrease in QoS. existing methods primarily address the prediction of the utilization of individual resources, for instance CPU, Main memory, Bandwidth Network, or Servers dedicated to cloud applications, these resources are often disregarded in the prediction of total utilization. This research uses FLNN and hybrid GA-PSO to leverage many resources. Data from Google's cluster tracking can assess the model's efficacy.

Similarly, the machine learning approaches can be used to create models that produce accurate predictions and outcomes, thereby preventing both resource overprovisioning and under provisioning. The use of an ensemble approach in machine learning is particularly beneficial as it combines multiple predictors instead of relying on just one. The article [16] presented a new method for workload prediction based on stack propagation. When the proposed model was compared to individual models and baseline models in a validation experiment, the ensemble approach improved prediction accuracy by 2%, as indicated by the results.

Based on our knowledge of previous works, it appears that this is the novel work to test the performance of cloud resource management strategies through simulation platform with regard to the load of accessing cloud resources.

B. Resource Management By Simulation Tools

The tools to represent and simulate entities and events support all discrete event simulation framework [17]. In addition, it is necessary to provide an interface that facilitates both the transfer and reception of events. The simulation framework manages the life cycle of each entity and the creation and distribution of events [18]. The availability of essential infrastructure, it becomes possible to develop more advanced tools on top of it. Furthermore, there are abstract concepts present within the simulation environment for resource management and scheduling, for example optimization and scheduling algorithms, which have no physical existence. Using a reliable and proven discrete event simulation framework and building our desired simulator on top of it is a very efficient method. It can provide us with the basic equipment we need, allowing us to concentrate on adding more advanced features.

CloudSim, [8] is a widely used platform simulator designed specifically for simulating the behavior of scheduling and resource management methods for workloads represented as separate tasks. This simulator offers a wide range of vital components for simulating a cloud environment. This is a toolkit that is open-source, allowing for the reprogramming of many features. Although CloudSim offers tools and may be partially reprogrammed, it is not suitable for simulating workflow execution in a cloud computing environment. This simulator allows the realistic representation of independent tasks and their features, including the number of instructions to be performed correctly. However, it lacks support for functionality like the interdependence among a group of tasks, which is necessary for physically presenting the process flow. Another constraint is the strictness of the arrival time for each task, requiring all tasks to be introduced into the simulation environment at time zero.

WorkflowSim, [19] is built on the foundation of CloudSim, but with a different mission: to model the execution of scientific procedures in distributed systems. This simulator allows the user to simulate scientific workflows by using a directed acyclic graph as its computing model. WorkflowSim considers many types of delays, including queue delay, postscript delay, and workflow engine delay, when conducting the simulation. It has the ability to simulate unpredictable faults in the system and react to them based on the users' settings, making it a very suitable option for simulating fault-tolerant systems and algorithms. An issue of this simulator is its absence of capability for the simultaneous execution of multiple tasks. Upon application to the workflow engine component, all jobs inside each workflow will get combined, making it impossible to discern which task relates to which workflow. Additionally, WorkflowSim offers support for a star-like topology in networking, where the central node functions as the broker and the other nodes linked to it serve as data centers. These limitations make WorkflowSim inappropriate for decentralized systems.

C. Complex Workflows

Complex workflows can be defined as a sets of connected tasks or jobs that are executed throughout several cloud resources to accomplish a certain goal, typically on time and within budget. Scientific processing, business processes, and massive data processing are examples of these workflows.

1) Components And Definition

A complex workflow is made up of multiple independent tasks or sub-workflows that operate cooperatively for completing specific functions, such as finishing assigned work within specified time constraints. The available cloud environments are capable of supporting such requirements and provide vast computational, storage, and other resources on demand. Based on the cost factor of these resources, the organization and independent users can also have autoscalers in dynamic and on-demand scenarios to match the requirements of the tasks getting executed in the form of workflows [20]. In order to specify and analyze the goals for complex workflows, one must use a workflow composition language to define the required complex goals, like throughput, fast speed execution, etc. A complex workflow is composed of multiple integrated goals. A goal can be the composite of at least one visited node or service [21].

2) Complex Workflows Challenges

The adaptability and expandability of public cloud resources are recognized for simplifying intricate workflows and enhancing efficiency and productivity. However, effectively managing large amounts of data and completion times presents a significant challenge, potentially impeding real-time data processing and prolonging processing times. In order to tackle this problem, it is necessary to improve efficiency in task design and storage use. Describing difficult tasks into smaller subtasks has continuously shown advances in query performance and efficiency of computation. Optimization of current data storage systems, especially for complex variable inquiries, is crucial. In light of their adaptability and scalability, public cloud services offer a cost-effective and efficient solution to these issues, making them highly suitable for data management and workflow enhancement. Utilizing public cloud resources can significantly boost productivity and streamline processes [22]. The work [21] was explored cloud resource management, particularly focusing on how workflows in scientific computing environments can be optimized through efficient resource utilization.

4. METHODOLOGY of RESEARCH

The following steps make up the research methodology used in this research:

- **Research Comprehension:** The research focuses on optimizing the efficient use of cloud computing resources during the execution of big data workflows. The aim of the research is to accurately predict the utilization of virtual engines, which includes the central processing unit, memory, and network bandwidth, to execute those workflows, using WorkflowSim.

- **Complex Workflows Collection:** We will conduct the experiments using big data workflows from Pegasus and the Next Generation Sequences (NGS) pipeline, based on real data in cloud computing. This data includes the number of resources and execution durations associated with the volume of input data during real execution.

- **Workflows Pre-processing:** We map out the architecture for the selected types in the application to understand the structure and complexity of big data workflows. This includes examining the collection of input data and the dependencies between the tasks that make up the workflow, as well as preparing the simulation platform with XML files for training, operation, and prediction using the simulation platform.

- **Predicting Resource Efficiency:** We employ a simulation platform and selected big data workflows for training applications to predict optimal resource utilization in the minimum possible time, thereby reducing costs. The simulation platform stands out by its ability to provide limitless experimentation at no cost, allowing cloud users to acquire the most optimal information to support the implementation of these selected workflows or others in a real cloud environment without the possibility of over- or under-provisioning.

- **Evaluation:** We calculate the metrics, that are often used to evaluate the prediction accuracy. For example, the mean squared error is an indicator that computes the average of the squares of the errors, which discourages bigger mistakes significantly more than the mean absolute error. The R-squared metric indicates the percentage of difference in the dependent variable that can be accounted for by the different times, while the mean absolute error measures the average absolute difference between the predicted and actual times.

- **Performance criteria:** The main aim is workflow execution to lower costs, improve cloud computing performance, and maximize resource usage.
- **Experiencing and Feedback:** The final step we, involve comprehensively evaluating the predictive results for all selected workflows by systematically describing the graphical charts according to the data and variables controlled within the simulation platform.

5. WorkflowSim Concepts

In this section, we introduce a simulation tool that is the basis of our work called WorkflowSim. It is developed to manage and monitor scientific workflow completion and quality. This is achieved through robust, discrete event simulation mechanisms, as mentioned before. In the following subsequent section, we provide an introduction to WorkflowSim.

A. *Introduction to WorkflowSim*

WorkflowSim serves workflow management researchers and developers. It simulates cloud computing-specific workflow methods for scheduling that are simple and efficient. As an extension of cloud simulation tools, WorkflowSim presents the major architectural idea of CloudSim so that users can comfortably use it and extend it based on the needs of their research.

The major idea presented by the architecture of CloudSim is the division of the simulation components of a cloud-based simulation into several layers. This kind of layering makes the user's task easy to do and easy to integrate. On the same principle, the core idea presented by WorkflowSim is an efficient framework for simulating an algorithm for scheduling Directed Acyclic Graphs (DAGs)-based workflows in cloud as well as cluster environments. WorkflowSim is adaptable and extensible so that users can also simulate the scheduling of workflows on utility grids and other grid computing-based platforms.

B. *Conceptual Framework*

A conceptual framework organizes coherent ideas into meanings and places them within a structural framework to support these meanings. The importance of a conceptual framework traces its roots to many different areas of research, such as utilizing the available resources in the cloud computing environment to perform big data analytics, which consists of computational resources as well as data storage management resources [23]. As cloud environments are generally seen to be composed of multiple data centers, a high-performance generalized service-oriented architecture (SOA) for virtualized cloud environments with the capacity of elastic scalability at web scale is useful and forms the basis of WorkflowSim [24]. Furthermore, the WorkflowSim platform is utilized to estimate the time needed to execute a workflow on big data and the amount of time throughout which the security of that data will be ensured. For further information, see these articles [25],[26].

C. *WorkflowSim Architecture*

The major architectural idea presented by CloudSim is the division of simulation components in a cloud-based simulation. Similarly, the architecture of WorkflowSim presents the complete idea of simulating the scheduling and execution of workflows in the field of cloud computing. The Fig. 1 illustrates the architecture of WorkflowSim.

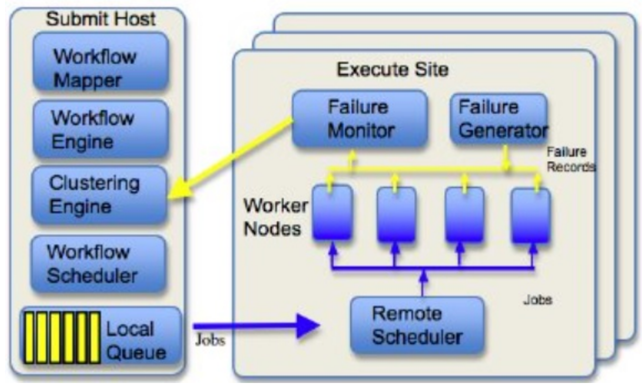

Figure 1. The WorkflowSim Architecture [19]

6. Studies of cases

In this section, we investigate the efficiency of resource utilization for three distinct complex workflows, two of which are detailed in [27], Montage and CyberShake. The other one is illustrated in [28], NGS pipeline.

- A- Montage is a procedure that has been employed to create astronomical image mosaics that meet a particular goal. As is illustrated in Fig. 2, which shows the small portion of a typical Montage workflow that is relevant to flow. The graphic represents the conceptual flow graph of the abstract form. Additionally, the figure demonstrated the nodes, the actual procedure would also involve nodes for data transmission and registration. The process can be categorized into multiple levels, as proposed in the referenced article [27].
  In the flow process, the levels of responsibility are illustrated by the numerical values associated with the vertices of the graph. The process requires two components: the input photos in the commonly-used FITS format and a ”template header” that describes the composition of the mosaic. Each input picture is first re-projected into the output mosaic's space of ordinates; next, the background of the re-projected images is removed; and finally, the final output picture is composed of the three parts.
- B- The CyberShake workflow is a scientific workflow used in the field of seismology to simulate and

analyze earthquake hazards.The CyberShake is available at the Southern California Earthquake Center (SCEC)[29]. Its project simulates earth motions for a specific area in order to produce Probabilistic Seismic Hazard Analysis (PSHA) data. The workflow makes use of high-performance computing resources and is designed to handle complex computations, including several steps of analyzing data and processing. to perform these tasks efficiently. The workflow is modular, consisting of distinct stages such as data preparation, wave propagation simulation, intensity measure calculation, and hazard curve computation. As displayed in Fig. 3, which shows the small portion of a typical CyberShake workflow that is relevant to flow.

- C- For the purpose of examining the modeling of resource utilization at a cloud simulator, we have selected a more complex instance study centered around the Next Generation Sequences (NGS) workflow. This method is used to carry out Whole Exome Sequencing (WES) data analysis workflows at the Laboratory for Genetic Medicine.
The workflow frequently handles the NGS pipeline processing phases, which are annotation (Annovar), saturation analysis (bedTools), alignment (BWA), sequencing re-calibration, content filtering, cleaning (Picard), and differences contacting and re-calibration (GATK). As seen in Fig. 4, the structure graph. Including the upper level structure that controls the flow of data and eight lower level processes, each has responsible for executing a particular pipeline step. Each phase involves the simultaneous execution of sub-workflows over several samples or areas corresponding to a certain sub-chromosome. The greatest level of complexity is characterized by the initiation of (N) sub-workflows for a specific step. All of these sub-workflows must be completed before proceeding to the next phase.
In summary, the workflow comprises three essential phases: (1) Generate the original sequencing for variant identification plus compute the amount of coverage to ensure that each sample size is around 36 gigabytes. (3) Variant identification and re-calibration; (4) Variant discrimination and annotation.
The implementation of Stages 1 and 3 comprises a loop, whereby each tool is invoked on individual samples. Given the independence of the samples in these two phases, the process of parallelization at this level is uncomplicated. Stage 2, in contrast, executes just once for each input sample, rendering processing across samples unfeasible. Nevertheless, the tools used in Stage 2 possess the capability to function separately on each chromosome (or even on the smaller in size sub-chromosomal zones). However, there is still a significant level of parallelism at this stage. It is worth noting that each exome in each sample remains divided according to chromosomal boundaries. This phenomenon is commonly known as chromosomal split [25].

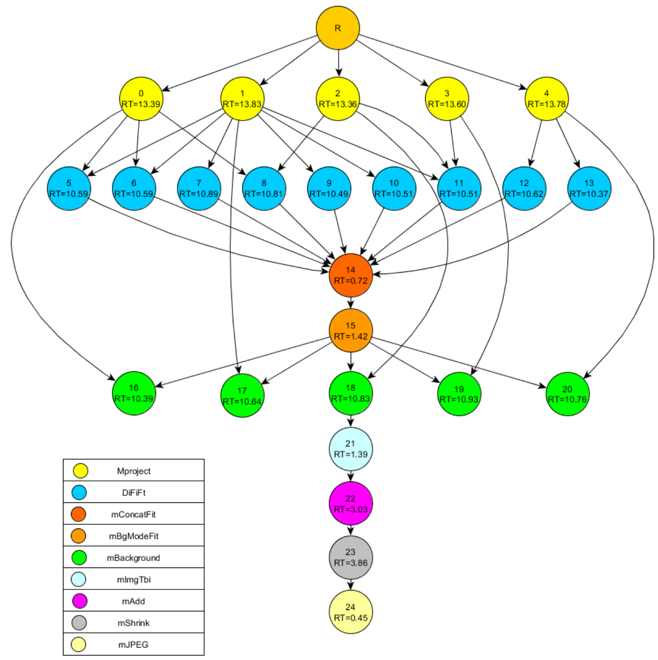

Figure 2. The structure of Montage workflow

7. Resource Utilization experiment

In our experiment, the WorkflowSim 1.1.0 framework that we utilized has an openly accessible number of VMs in a datacenter; each one of these VMs represents an engine in the real cloud. Each virtual machine has a total of 512MB of RAM and one processor. The speed of VMs is spread out uniformly between 1000 and 2500. Additionally, the bandwidth is 1000 Mbps and the maximum transfer rate is 15. The environmental factors relevant to the configuration and associated expenses are documented in the Table I.

TABLE I. Resources utilized in the assessment experiment and their configuration parameters.

Configuration Parameters	
VM No.	From 1-90
CPU MIPS	1000
CPU	3.0
Bandwidth	1000
RAM Size	512 MB
VMM Type	Xen
Image Size	10000

The initial experiment was carried out with the guidance of the WorkflowSim platform and used the complex workflow outlined in the previous section 6, Montage, which handles huge amounts of data, including 1000 different tasks. The results, which are shown in the Fig. 5, predict the

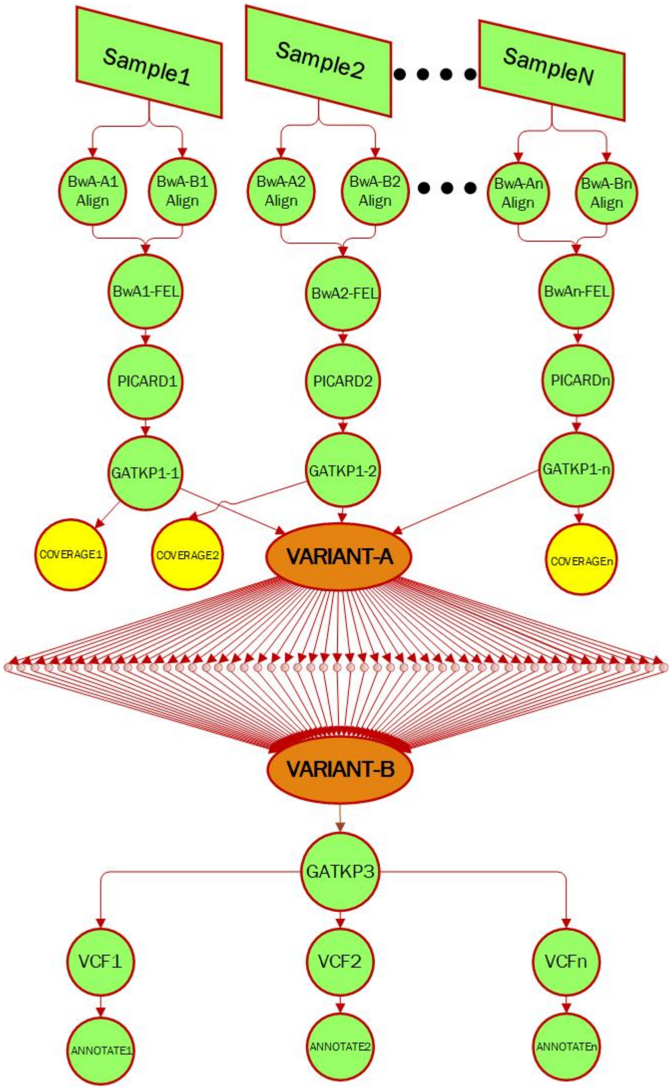

Figure 3. The structure of Cybershake workflow

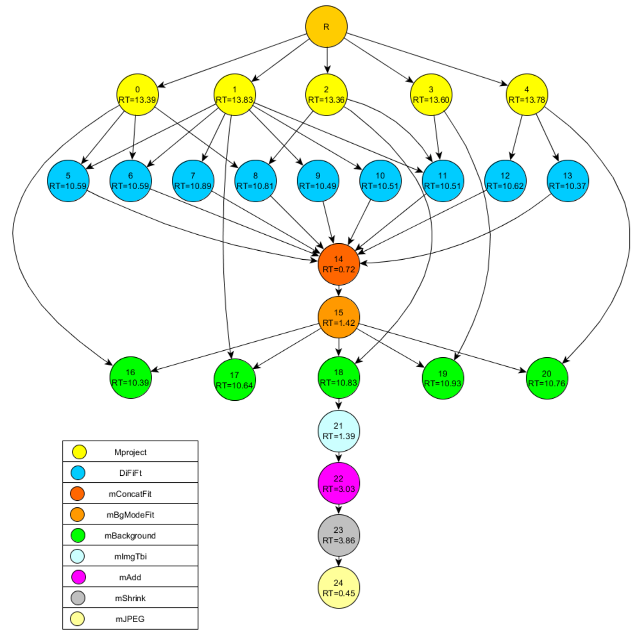

Figure 4. The structure of NGS pipeline workflow

resource consumption required for implementing the exact same workflow on a real cloud platform, ensuring optimal performance and a reasonably equitable resource use.

We conducted a number of operational attempts for the Montage workflow, ranging from 1 to 90. We estimated this number of attempts by taking the first number and increasing it by one, then continuing until we reached a point where increasing the number of resources had no effect on the execution time; in other words, having an excess of resources is not beneficial.

The graph chart in the Fig. 5 that illustrates the efficient execution of a Montage workflow with resource management demonstrates the following: The purple line graph that represents the executing time shows a significant decrease in time as the number of resources utilized increases from 1 to 10. The time graph begins to separate out slowly and intersects with the red resource graph at the point of utilizing 11 resources, this is considered the optimal number of resources that concords with time and cost. The second portion of the time graph exhibits a flat path with a slight decrease in slope as resources increase from 12 to 80, this indicates that resources are being consumed without a significant increase in operational yield. The final stretch of the time graph, which is shown in yellow, follows a straight line at 100% when resources increase from 80 to 90, because of the inability to reduce the executing time while still increasing the resources. See the specifics of the execution outcomes by examining the table II.

Furthermore, the second experiment was performed using the WorkflowSim platform and used the Cyber-Shake workflow outlined in section 6. This workflow has been created to manage massive amounts of data, which includes 50 individual tasks. The results are presented in the Fig. 6. In addition, the objective of this experiment is to predict accurately the resource consumption needed to execute the chosen workflow on a real cloud platform.

To evaluate the efficiency of running various workflows on the proposed simulation platform, we selected this workflow with an unusual structure and size that is smaller. A number of operational experiments have been conducted for the CyberShake workflow, including a range of 1 to 20 attempts. Furthermore, the number of attempts was determined by gradually raising the initial number by one, and so forth, until a point was reached where increasing the resources did not affect the execution time. In simple terms, a lot of resources is not beneficial.

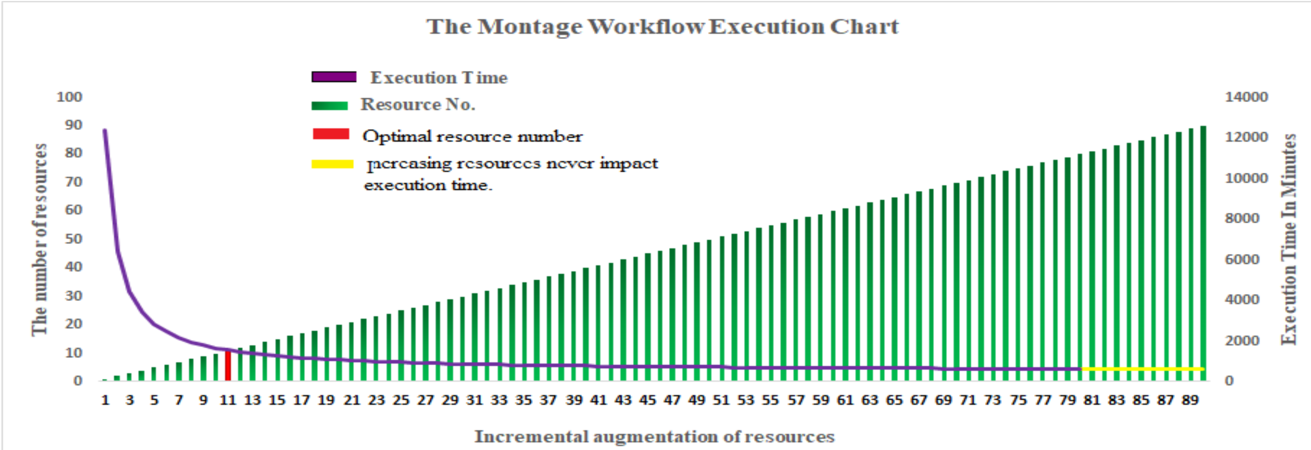

Figure 5. The Montage workflow execution chart

The graph in Fig. 6 illustrates the allocation of resources used during the execution of the CyberShake workflow. Furthermore, the graph depicts the behavior of execution time. Whereas the orange line shows the decrease in execution time as it improves from left to right with the increase in resource number from 1 to 5.

The time graph begins to part slowly and meets the red resource chart at the 6 resource consumption point that is termed optimal since it corresponds with time. Its further increase to 7, 8, and 9 depicts the second part of the time graph moving flatly with a slight decrease in slope, showing that it is just consuming resources without any decreasing operational time strongly.

The green portion of the chart is flat along a line at 100% as resources increase from 10 to 20 in the final segment because increasing resources is not being able to decrease the running time. Table III discloses the details of the implementation results.

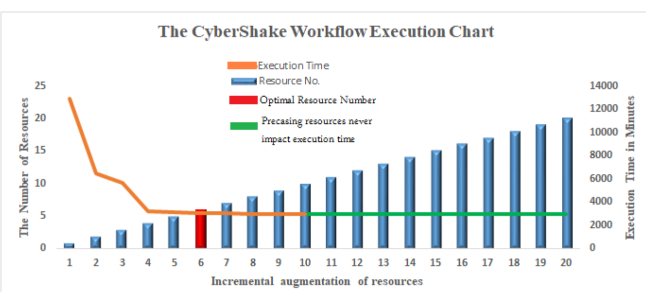

Figure 6. The CyberShake workflow execution chart

The final experiment was also conducted based on the WorkflowSim platform, using the NGS pipeline workflow that we introduced in Section 6. This workflow, an approach to big data, had a minimum reproducibility of six samples, with a total of 101 tasks.

The Fig. 7 presents the results. Moreover, the key reason for conducting this experiment is to predict the execution time by using the simulation platform and comparing it with the real-time previously obtained from running it on Microsoft's public Azure cloud. In order to enhance the analysis of previous experiment results, we compared the resource consumption required for execution on the real cloud with that from the simulation platform.

The chart in Fig. 7 demonstrates how the NGS workflow is running in actual time with resource management along with the forecasted number of resources on our platform. The blue line stands for real-time workflow execution on the Microsoft Azure Cloud, and this is acted upon when comparing the number of resources predicted by the simulation platform.

The orange graph represents estimated time while decreasing from highest to lowest until it reaches an optimal value of 12 resources. A red straight line gradually moves from left to right as time decreases, while resource consumption increases from 1 to 11. The intersection point between the red straight line and the resource consumption line is at 12, which reflects the ideal prediction state because it coincides with real-time data.

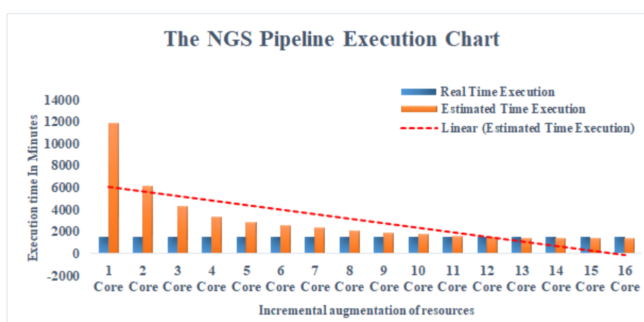

Figure 7. The NGS pipeline execution chart

TABLE II. Summary table of the first experiment

No. of Attempts	Resource No.	The change in the execution time graph
10	From 1 To 10	A significant decline in execution time from 12384.23 to 1665.95
1	11 resources Only	The best amount of resources can be used, the time is 1567.5
68	From 12 To 80	Small execution time change from 1463.99 to 632.5 despite abundant resources.
10	From 80 To 90	Without changing the execution time, the time is 632.57

TABLE III. Summary table of the second experiment

No. of Attempts	Resource No.	The change in the execution time graph
5	From 1 To 5	A significant decline in execution time from 12889.29 to 3155.61
1	6 resources Only	The best amount of resources can be used, the time is 3071.59
3	7, 8, and 9	Small execution time change from 3026.26 to 2974.12 despite abundant resources.
10	From 10 To 20	Without changing the execution time, the time is 2953.94

TABLE IV. Summary table of the third experiment

No. of Attempts	Resource No.	The change in the execution time graph
11	From 1 To 11	A significant decline in execution time from 11872.87 to 1568.1
1	12	Optimal resource No. when Estimated time 1509.95 approximately equal to Actual execution time 1531.26.
N	13 and above	Increasing resources becomes useless and leads to very high costs.

8. Conclusions and Future Work

This work proposed a comprehensive approach using a simulation platform to predict resource utilization when operating complex workflows in public cloud computing environments. Efficient resource management in cloud environments is one of the biggest challenges facing organizations, especially when dealing with complex workflows that require intensive and dynamic allocation of computing resources. Three different types of workflows were utilized to show different scenarios involved with applications with different computing requirements. Each workflow type has a unique property, such as size, structure, and others with real data. When compared to real data, this produced results that were highly accurate and acceptable, with an almost optimal success rate of 100%.

The focus of the study was to test the accuracy of the resource utilization prediction of the simulation platform, as shown in the third experiment. The minimum square error (MSE) of the runtime estimate was calculated to be 454.1161, the lowest value among all predictions when the resource count was 12, and 100% consistent with the actual execution.

Therefore, this study focused on the accuracy of simulation in identifying the right point of resource allocation to enhance efficiency and cut down operational costs. The findings of this study will give recommendations that would help enhance strategies on resource allocations in public cloud environments and prevent over-provisioning or under-provisioning resources that could result in poor performance of systems and applications. This research deems itself an essential stepping stone for simulative technology to be used as an effective tool in proactive and premeditated cloud resource management.

A. Abbreviations and Acronyms

The following is an explanation of some abbreviations that appeared in the research paper.

QoS- Quality of service.

SLAs- Service Level Agreements.

CPU- Central Processing Unit

MIPS- Million Instructions Per Second.

NGS- Next Generation Sequences.

XML- Extensible Markup Language.

SOA- Service Oriented Architecture.

References

[1] D. Fernández-Cerero, A. Fernández-Montes, A. Jakóbik, J. Kołodziej, and M. Toro, "Score: Simulator for cloud optimization of resources and energy consumption," *Simulation Modelling Practice and Theory*, vol. 82, pp. 160–173, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1569190X18300030

[2] F. Llwaah, J. Cala, and N. Thomas, "Simulation of runtime performance of big data workflows on the cloud," 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:40273210

[3] F. Ghedass and F. Ben Charrada, "A multi-view learning approach for the autonomic management of big services," in *Web Information Systems Engineering – WISE 2021*, W. Zhang, L. Zou, Z. Maamar, and L. Chen, Eds. Cham: Springer International Publishing, 2021, pp. 463–479.

[4] A. Mukherjee, D. De, and R. Buyya, *Cloud Computing Resource Management*. Singapore: Springer Nature Singapore, 2024, pp. 17–37. [Online]. Available: https://doi.org/10.1007/978-981-97-2644-8_2

[5] L. Wang, R. Ranjan, J. Chen, and B. Benatallah, *Cloud Computing: Methodology, Systems, and Applications*, 01 2011.

[6] S. K. Garg, S. Versteeg, and R. Buyya, "A framework for ranking of cloud computing services," *Future Generation Computer Systems*, vol. 29, no. 4, pp. 1012–1023, 2013, special Section: Utility and Cloud Computing. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X12001422

[7] R. Buyya, C. S. Yeo, and S. Venugopal, "Market-oriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities," in *2008 10th IEEE International Conference on High Performance Computing and Communications*, 2008, pp. 5–13.

[8] R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. F. De Rose, and R. Buyya, "Cloudsim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms," *Softw. Pract. Exper.*, vol. 41, no. 1, p. 23–50, jan 2011. [Online]. Available: https://doi.org/10.1002/spe.995

[9] X. Xu, X. Zhang, M. Khan, W. Dou, S. Xue, and S. Yu, "A balanced virtual machine scheduling method for energy-performance trade-offs in cyber-physical cloud systems," *Future Generation Computer Systems*, vol. 105, pp. 789–799, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X17318927

[10] I. Bambrik, "A survey on cloud computing simulation and modeling," *SN Computer Science*, vol. 1, 08 2020.

[11] M. Adel Serhani, H. T. El-Kassabi, K. Shuaib, A. N. Navaz, B. Benatallah, and A. Beheshti, "Self-adapting cloud services orchestration for fulfilling intensive sensory data-driven iot workflows," *Future Generation Computer Systems*, vol. 108, pp. 583–597, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X19316231

[12] B. M. Miloud Khaldi, Mohammed Rebbah and O. Smail, "Fault tolerance for a scientific workflow system in a cloud computing environment," *International Journal of Computers and Applications*, vol. 42, no. 7, pp. 705–714, 2020.

[13] Y. Gong, J. Huang, B. Liu, J. Xu, B. Wu, and Y. Zhang, "Dynamic resource allocation for virtual machine migration optimization using machine learning," *arXiv preprint arXiv:2403.13619*, 2024.

[14] S. D. Simić, N. Tanković, and D. Etinger, "Big data bpmn workflow resource optimization in the cloud," *Parallel Computing*, vol. 117, p. 103025, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167819123000315

[15] S. Malik, M. Tahir, M. Sardaraz, and A. Alourani, "A resource

utilization prediction model for cloud data centers using evolutionary algorithms and machine learning techniques," *Applied Sciences*, vol. 12, no. 4, p. 2160, 2022.

[16] R. Shaikh, C. H. Muntean, and S. Gupta, "Prediction of resource utilisation in cloud computing using machine learning." in *CLOSER*, 2024, pp. 103–114.

[17] J. Banks, J. II, B. Nelson, and D. Nicol, *Discret-Event System Simulation*, 01 2010.

[18] A. M. Law, W. D. Kelton, and W. D. Kelton, *Simulation modeling and analysis*. Mcgraw-hill New York, 2007, vol. 3.

[19] W. Chen and E. Deelman, "Workflowsim: A toolkit for simulating scientific workflows in distributed environments," in *2012 IEEE 8th international conference on E-science*. IEEE, 2012, pp. 1–8.

[20] T. Ben-Nun, T. Gamblin, D. S. Hollman, H. Krishnan, and C. J. Newburn, "Workflows are the new applications: Challenges in performance, portability, and productivity," in *2020 IEEE/ACM International Workshop on Performance, Portability and Productivity in HPC (P3HPC)*, 2020, pp. 57–69.

[21] N. M. Gonzalez, T. C. M. d. B. Carvalho, and C. C. Miers, "Cloud resource management: towards efficient execution of large-scale scientific applications and workflows on complex infrastructures," *Journal of Cloud Computing*, vol. 6, pp. 1–20, 2017.

[22] X. Ma, H. Xu, H. Gao, and M. Bian, "Real-time multiple-workflow scheduling in cloud environments," *IEEE Transactions on Network and Service Management*, vol. 18, no. 4, pp. 4002–4018, 2021.

[23] S. Mustafa, B. Nazir, A. Hayat, S. A. Madani *et al.*, "Resource management in cloud computing: Taxonomy, prospects, and challenges," *Computers & Electrical Engineering*, vol. 47, pp. 186–203, 2015.

[24] R. Buyya, R. Ranjan, and R. N. Calheiros, "Modeling and simulation of scalable cloud computing environments and the cloudsim toolkit: Challenges and opportunities," in *2009 international conference on high performance computing & simulation*, 2009, pp. 1–11.

[25] F. Llwaah, J. Cała, and N. Thomas, "Runtime performance prediction of big data workflows with i/o-aware simulation," in *Proceedings of the 11th EAI International Conference on Performance Evaluation Methodologies and Tools*, 2017, pp. 74–81.

[26] F. Llwaah, "Security time prediction of big data workflows with aes algorithm-aware simulation," *International Journal of Computing and Digital Systems*, vol. 16, no. 1, pp. 1–11, 2024.

[27] E. Deelman, G. Singh, M.-H. Su, J. Blythe, Y. Gil, C. Kesselman, G. Mehta, K. Vahi, G. B. Berriman, J. Good, A. Laity, J. C. Jacob, and D. S. Katz, "Pegasus: a framework for mapping complex scientific workflows onto distributed systems," *Scientific Programming Journal*, vol. 13, no. 3, pp. 219–237, 2005. [Online]. Available: http://pegasus.isi.edu/publications/Sci.pdf

[28] J. Cała, Y. Xu, E. A. Wijaya, and P. Missier, "From scripted HPC-based NGS pipelines to workflows on the cloud," in *First Int. Work. Cloud Bio (C4Bio 2014)*, 2014.

[29] R. Graves, T. H. Jordan, S. Callaghan, E. Deelman, E. Field, G. Juve, C. Kesselman, P. Maechling, G. Mehta, K. Milner *et al.*, "Cybershake: A physics-based seismic hazard model for southern california," *Pure and Applied Geophysics*, vol. 168, pp. 367–381, 2011.

Dr. Faris Llwaah achieved a Ph.D. in computer science in 2018 from Newcastle University, UK, in the High Performance of Computing on the Cloud. He received a master's in computer science from Al-Nahrain University in Baghdad, Iraq in 1993. His degree is in computer science from Iraq's University of Mosul. His job is Cyber Security coordinator. He is a Cyber Security Department scientific committee member and rapporteur. He is on the department's examination committee. He has higher education teaching expertise. He is very interested in the performance of all cloud computing applications and computer operating systems.