Comparison Among Cloud Technologies and Cloud Performance

Omar M. Ahmed1*, Lailan M. Haji2, Hanan M. Shukur3, Rizgar R. Zebari4, Shakir M. Abas5, Mohammad A. M. Sadeeq6

1Zakho Technical Institute, Duhok polytechnic University, Kurdistan Region –Iraq, omar.alzakholi@dpu.edu.krd
2Department of Computer Science, University of Zakho, Kurdistan Region –Iraq, lailan.Haji@uoz.edu.krd
3Department of Computer Techniques Engineering, Al Kitab University, Iraq, hanam89md@gmail.com
4Duhok polytechnic University, Kurdistan Region –Iraq, rizgar.ramadhan@dpu.edu.krd
5Duhok polytechnic University, Kurdistan Region –Iraq, Shakirdu@yahoo.com
6Duhok polytechnic University, Kurdistan Region –Iraq, mohammad.abdulrazaq@dpu.edu.krd
* Correspondence: omar.alzakholi@dpu.edu.krd

Abstract
The cloud is the best method used for the utilization and organization of data. The cloud provides many resources for us via the internet. There are many technologies used in cloud computing systems; each one uses a different kind of protocols and methods. Many tasks can execute on different servers per second, which cannot execute on their computer. The most popular technologies used in the cloud system are Hadoop, Dryad, and another map reducing framework. Also, there are many tools used to optimize the performance of the cloud system, such as Cap3, HEP, and Cloudburst. This paper reviews in detail the cloud computing system, its used technologies, and the best technologies used with it according to multiple factors and criteria such as the procedure cost, speed cons and pros. Moreover, A comprehensive comparison of the tools used for the utilization of cloud computing systems is presented.

Keywords: Cloud, Hadoop, Dryad, Cap3, HEP, Cloudburst, Cloud Technologies

Received: Marc 21, 2020 / Accepted: April 22, 2020 / Online: April 23, 2020

I. INTRODUCTION
The computing of scientific solutions requires creating mathematical models and techniques of the numerical solution for the majority of problems. These types of models unusually need a massive time and effort of processing and many resources to compute a large number of complex procedures in determining time [1],[2]. These can be solved initially addressed by high-performance computing infrastructures such as a pool of networking computers, clustering, or by high CPU cycle with some software such as Condor [3],[4],[5]. However, with networking, a new approach is available to make complex procedures easier by distributing the tasks on cloud machines over the world [6],[7].

Cloud computing is one of the best technologies that provide information technology services and solve many problems [8],[9]. Cloud computing using the virtualization technologies, which provide the end-users many of services such as hardware resources and applications levels [10],[11]. Besides, cloud computing has other features that can benefit the scientists by the scaling of downing the infrastructure of computing depending on the user’s budget and the requirement of applications. However, by using cloud computing technology, we can provide a large number of distributed infrastructures with a suitable environment [12],[13].

Cloud computing and the technologies of cloud considered two class categories of technologies used in the cloud computing system. The cloud term means there are many services such as Platform as a service (PaaS), Infrastructure as a Service (IaaS), etc. Virtualization is a base role applied in the cloud computing system [14]. Cloud technology refers to the difference of runtimes software, which used on cloud computing systems such as Hadoop, Dryad, and communication framework such as HDFS (Hadoop Distributed File System), Amazon S3, etc [15]. In the present, there are many availabilities of cloud service applications which used primarily on cloud computing systems such as Nimbus and Eucalyptus, which allow different companies to develop the clouds to raise the efficiency of
The Hadoop architecture is simple - use Figure 1 shows HDFS architecture. The Hadoop Distributed File System consists of many components, each of these components designed in different architecture and for different work. Hadoop can handle terabytes or more of data, and the application of their techniques are different from other technologies [22]. Hadoop store information on different computers, maybe hundreds of computers with an inefficient method which decrease the losing risk of information. The Hadoop is very beneficial when the data are distributed end executed on several servers.

II. CLOUD COMPUTING AND TECHNOLOGIES

The cloud computing term means a set of servers such as Platform as a service (PaaS) and infrastructure as a service (IaaS) that provide the resources and processors to the user in an efficient environment [3]. Cloud computing and their related technologies can split into two categories of technologies used by clouds. Also, cloud technologies refer to execute many tools of software based on cloud servers as Hadoop and Dryad, also framework of communication as HDFS (Hadoop Distributed File System), Amazon S3, etc. [15],[18]. There are many cloud computing as Nimbus and Eucalyptus that allow the companies to make the development of their individual and unique clouds to provide the best efficiency of resources. On the other side, cloud technologies provide parallel computing [1].

There are several applications used in cloud computing technologies, and the large part of these applications and data are applied and used in particle physics fields, information retrieval, etc. However, there are various techniques used to increase the performance of cloud computing. The cloud term widely used in some companies but without useful and understand the cloud completely [19]. The cloud tries to use technologies that make cloud computing more accessible and reached by anyone at any time. There are many technologies used to increase the performance of cloud computing, but the most important of them are Hadoop, Dryad, CGL Map Reduce, MPI, and CAP3 [20].

A. Hadoop Technique

Hadoop is one of the most common technologies used in cloud computing, and the Hadoop architecture is simple compared with other kinds of techniques. The main components of the HDFS are the master and slave architecture. Generally, the HDFS cluster has a single Name Node, Master-Slave, and many Data Nodes [15]. File system and the applications stored by HDFS separately. Each server connected by TCP protocols [21]. Figure 1 shows HDFS architecture.

The Dryad is another critical technology used by cloud computing. The Dryad is a system and consists of a set of extensions of languages that can create new developing programming on a colossal scale for distributed cloud computing [23]. On the other hand, it mainly considered the language domain-specific that depended on C++ libraries. The primary purpose of Dryad is to distribute the execution of parallel applications on the cloud. Also, it combines channels and vertices to create a new develop graph of dataflow then runs the application by using vertices of a graph on the computers to provide communication among them using TCP[3]. Dryad is created and designed to be very expressive, and it beats many of other cloud technologies like Google's map-reduce [6]. Also, it processes several complex problems of developing distributing applications on the extensive range as monitoring, creating and managing jobs, visualizing jobs, accounting, fault tolerance, and transporting information among vertices of the graph. The Dryad consists of the build, which allows programmers to configure the resources of the cloud or information center of different parallel applications. Furthermore, it provides many computers that may be more than thousands of machines for programmers [20],[24]. Figure 2 shows the monitoring in Dryad.

B. Dryad Technique

Dryad is another critical technology used by cloud computing. The Dryad is a system and consists of a set of extensions of languages that can create new developing programming on a colossal scale for distributed cloud computing [23]. On the other hand, it mainly considered the language domain-specific that depended on C++ libraries. The primary purpose of Dryad is to distribute the execution of parallel applications on the cloud. Also, it combines channels and vertices to create a new develop graph of dataflow then runs the application by using vertices of a graph on the computers to provide communication among them using TCP[3]. Dryad is created and designed to be very expressive, and it beats many of other cloud technologies like Google's map-reduce [6]. Also, it processes several complex problems of developing distributing applications on the extensive range as monitoring, creating and managing jobs, visualizing jobs, accounting, fault tolerance, and transporting information among vertices of the graph. The Dryad consists of the build, which allows programmers to configure the resources of the cloud or information center of different parallel applications. Furthermore, it provides many computers that may be more than thousands of machines for programmers [20],[24]. Figure 2 shows the monitoring in Dryad.

C. CGL Map Reduce Technique

It considered kind of stream of MapReduce that has ability executing in Java, this procedure give results in eliminating overhead linked related to file, communication when the file sent to end user directly by producer [15]. In addition, it adding to MapReduce several of improvements, also it is support long running map but it basically provides fast data from pub
network. In beginning of stage, the tasks of MapReduce are configured and the data will be ready for loading. All stages handle all task related to the data processing. Also, there is framework that used to transfer the results from the map task to reduce task directly and there is level which convert merge the results that derived from reduce task into new single data [25]. The users only can access to data of merged task. The MapReduce has other facility to distrust the data set which is variable to all tasks on the map; this procedure is very useful for analysis of data. The user divides and distribute the data and can read the information form any disk and file shared. The CGL computation can be individual step or iterative [19]. The iterative step the user has ability to call another MapReduce iteration by processing the results which achieved by merge state. In this time the data it not needs to be configured because the MapReduce task managed the data by framework already [6],[26]. Figure 3 shows CGL MapReduce.

D. CGL High Energy Physics (HEP) Technique

HEP technology used primarily in the study of the particle of interaction. This technology aims to find the most fundamental block and essential of matter [27]. The particle study used a large of data that need to stored and analyzed, and this procedure needs to vast experiment and evaluation results. This procedure cannot use Dryad technology because the input data is extensive and in binary format. This problem can overcome by dividing the information manually in each node of the cluster, and the data consist of the only name of file provide by each node that can be used [15]. The first step is analyzing the data by Appling function, which programmed for all input files in the ROOT script. They use the homomorphic filter because it can analyze several files at the same time. After that, can use apply operation to be applied for all data, and the output results will be multiple nature. In the next step, the application root method will be used to repeat over the data set, and the script is executing to give the results in a histogram. Finally, repeating all processes overall partially created by histogram to obtain the single histogram [23],[28]. Figure 4 shows HEP Application.

E. Cloud Burst Technique

Cloud Burst technology is one of the most important and latest high developing algorithms in parallel computing. This technology uses optimize mapping of end-user, and it has the ability to choose the filter of alignments of a report of the best read between the readings, which is given [1]. The cloud burst can be considered a tremendous scalable algorithm, and it has many advantages as providing easy work for the user when the user works on the enormous remote cloud computing [3]. It can determine a large number of reads in few times. Cloud bursting considered configuration put between the private and public cloud, which used in transfer information. This technology used when the companies used the private cloud, and the data reaches more than 100% of its capacity, the extra traffic directed to the public cloud, so the service is not interrupted [6]. The Figure 5 illustrates a cloud burst. The cloud bursting applied for a dynamic application that basically runs on the private cloud and converts to the public cloud. This technique making private clouds more efficient and reduce cost computation. The private clouds can be protected in computation and storage to provide the resources all time [21],[29].
Table 1. Comparison Among Cloud Computing Technologies

<table>
<thead>
<tr>
<th>Criteria / Technologies</th>
<th>Hadoop</th>
<th>Dryad</th>
<th>CGL Map Reduce</th>
<th>High Energy Physics (HEP)</th>
<th>Cloud Burst</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deployment</td>
<td>Use two nodes master and worker in each cluster to perform job</td>
<td>The analysis and store data are arranged at same node</td>
<td>The computation and storing done at same nodes like Dryad</td>
<td>Each task done in different nodes and all data analyzed to make single value</td>
<td>Connect the private and public cloud computing</td>
</tr>
<tr>
<td>Technology/Language support</td>
<td>Java or java services</td>
<td>Programmed by C#</td>
<td>Programmed by C++, C, Python</td>
<td>Programmed by C++, C, Python, MATLAB</td>
<td>Java or java services</td>
</tr>
<tr>
<td>Programming model</td>
<td>Only Map Reduce</td>
<td>Depending on (DAG)</td>
<td>Combine many stages with MapReduce</td>
<td>Cluster nodes and Map Reduce</td>
<td>Depending on (DAG), Only Map Reduce</td>
</tr>
<tr>
<td>Scheduling</td>
<td>Use FIFO Algorithm</td>
<td>Used Graph depending on network</td>
<td>Data local</td>
<td>Used Different algorithm as FIFO, Graph</td>
<td>Use several scheduling depending on the cloud</td>
</tr>
<tr>
<td>Handling errors and failure</td>
<td>Repeat the task to reduce error</td>
<td>Repeat executes vertices of graphs</td>
<td>By use histogram</td>
<td>Repeat the task to reduce error</td>
<td>Repeat the execution reduce task and map</td>
</tr>
<tr>
<td>Monitoring</td>
<td>The user is the monitor</td>
<td>----</td>
<td>Progress monitoring by programming</td>
<td>Monitoring by machines</td>
<td>No implement</td>
</tr>
<tr>
<td>Execution environment</td>
<td>Linux cluster, Future Grid</td>
<td>Windows HPCS cluster</td>
<td>----</td>
<td>Windows HPCS, Future Grid</td>
<td>Windows, Linux Unix</td>
</tr>
<tr>
<td>Communication</td>
<td>Http Technology</td>
<td>TCP Technology</td>
<td>All communication Technique</td>
<td>HTTP Technology</td>
<td>All communication Technique</td>
</tr>
<tr>
<td>Data storage</td>
<td>HDFS</td>
<td>NTFS and DFS</td>
<td>Google, Amazon S3</td>
<td>HDFS</td>
<td>Google, HDFS</td>
</tr>
<tr>
<td>Data handling</td>
<td>Using multi-machines on cloud</td>
<td>Use local disks or shared directories</td>
<td>Use shared storage on cloud</td>
<td>Use the different storage on cloud computing</td>
<td>Use shared storage on cloud</td>
</tr>
</tbody>
</table>

III. LITERATURE REVIEW

Kannadasan et al. [30] done some of the experiments and researching on the several technologies of cloud computing. Explore the Hadoop is very useful for the Dryad and CGL MapReduce in most cases. The Hadoop execution time is very fast in most applications compared with other techniques because the Hadoop mainly has a large number of partitions than other techniques. This feature significantly decreases the time of computation reduction phase. Also, when the data is enormous, the CGL MapReduce and the Hadoop will be more efficient, but Dryad will be working slowly compared with the Hadoop and CGL MapReduce. This reason is mainly returning to the storage of the mechanism used by Dryad, as illustrated in the HEP application. In another term, Hadoop is more user-friendly and is very easy to use when we compare it with the other techniques used for this purpose. The other feature that can benefit the users is that they increase the online storage mechanism, which provides the services of fast transfer time of computation analyzing of data procedure.

Ekanayake et al. [15] introduced a survey many of study on real and standard measurements related to clouds and cloud technologies. Study the different aspects of cloud computing of paralleling by using either traditional, or the new cloud technologies approaches. However, cloud computing technologies are very well working on most aspects of parallel processing. Additionally, when the data is enormous, the Hadoop is decidedly better than other technologies of the concept of computing of data, and quality of service is better for providing of service as reduce the fault and monitoring. All techniques try to simplify the implementation of the same problems. Evaluation of several techniques used to increase the performance of cloud computing such as Hadoop, Dryad LINQ, etc.

Kumar [25] Proofed that cloud computing's high-performance use in many aspects, especially allow scientists to
solve the considerable problems in sciences, business, and health problems by using cloud computing. Cloud computing provides high performance by using some techniques which increase the bandwidth, enhanced, networking, and capabilities of computing. Besides, cloud computing allows the users the high speed of searching with high performance and low cost by provides many resources as a cluster computer and set of CPUs. On the other hand, it provides unlimited resources, which mean the high-performance computing (HPC), also analyzing massive data by running multi-machines at the same time. Some companies can analyze the vast data more economically and fast scalable improvement solutions on cloud computing.

Shakeabubakor et al. [19] concluded that a few knowledge about cloud computing had enormous effects on the attitude towards trust and using cloud computing. Cloud computing impacts directly on performance and increases productivity by allows users to access their requirements from anywhere and anytime. Besides, cloud computing reduces the cost of the cloud application used by users. On the other side, it needs to increase the knowledge level of the users by put some roles and training. The researcher improves that cloud computing is the best way to provide the many resources and analyzing data at few minutes compared with the traditional approach of data processing. There are several techniques use to increase the performance of cloud computing as Hadoop and MapReduce.

Vecchiola et al. [27]Discussed the developing of technologies and the opportunities happen of high performance of computing science on the clouds. the cloud computing became the adoption technologies and new model today, the clouds has definitely one of the most popular technology which applied in many companies and service providers and also widely used by the end users. Although cloud is popular and widely used but it is still in growing stage and need some techniques to increase the performance and reduce the faults. The cloud computing Grids as Open Science Grid and EGEE produce the huge infrastructure and set of main way and tools. The aims of cloud are providing the resources and processing of data to the users. In other hand, increasing of high performance of cloud computing leads to use the cloud by many of companies and organization. Using some technologies to increase the performance of cloud computing as Hadoop and ASTAR.

Stantchev [21] Introduced the evaluation of how to increase the performance of cloud computing by using some technologies in advanced computing infrastructures, which represent cost and performance. Also, it provided experimental to the evaluation of the Amazon EC2 approach. Evaluation of Windows instances and its performance on cloud computing by using some technologies which allow Windows users to deal with cloud efficiently and processing enormous data in determining time. Also, merge the NFPs and service levels in the same approach and use various methods to make them work on the cloud more comfortable and powerful. The NFP of individual services provides information that is relevant and granular. This technique applied to the evaluation of Amazon Elastic Compute.

Attaran [22] showed that the development of technologies is providing the helping to companies and organizations to transform their working into cloud computing technology. This approach allows them to stay connecting with their customers, suppliers, and employees. The internet and servers’ techniques allow the user to create an interactive environment for the working. The internet provides the connecting to the servers from anywhere and anytime. On the other side, cloud computing is widely and extensively used to store and process massive data that provide services to companies in an efficient manner. While in the present, the number of uses clouds computing is not exhaustive but unquestionably provides the incentives of using cloud compared with traditional technologies. Cloud computing has advantages to increase information technology, infrastructure flexibility, powerful, easy access, and insurance keep the information. Also, cloud computing used some technologies that provide high performance in processing and analyzing massive data on cloud servers.

Rashid et al. [18] introduced a system design along with its analysis and requirements for a platform to apply the combination of both distributed parallel-processing and distributed cloud computing for solving (in minimum time) specific/huge problems of client that require special processing energy. The system consisted of several servers distributed geographically, thus for processing the tasks of a client some/all of the servers worked simultaneously. The servers were installed and distributed geographically between Iraq and Turky. A central server acts as a link between clients and all servers since the system is fully automatic this indicates that the registration of selves by the servers on the central server is done automatically. Meanwhile, the clients can make use of processing power provided by registered servers. A significant strength point is bing able to use the power of processing via cloud domain depending on the principles of distributed systems.

Asadi et al. [31] Proposed an analytical model and numerical evaluation for the performance and power of cloud infrastructure by utilizing the stochastic reward nets (SRNs). The purpose of making such a system is to analyze the influence of network traffics and temperature on the performance and strength of cloud computing systems, which was neglected in the majority of approaches that have been presented previously. In the purposed design, there are two categories for incoming tasks: hot and cold, manifesting various energy consumption patterns, which could fail because of the traffic and bandwidth constraints. By implementing the purposed system, when cold and hot tasks are assigned to virtual machines and taking into consideration the network traffic, there is the ability to compare and evaluate various resource allocation strategies in terms of performance and power.

Garg et al. [32] proposed an efficient model for data integrity auditing cloud computing. The purpose of proposing this protocol is to reduce the computational complexity of client during system setup phase of the auditing protocol, the suggested model is publicly verifiable based on the pairings characteristics of bilinear, also the dynamic operations on data are supported by this protocol. The protocol security relies upon the stability CDHP in a Random Oracle Model (ROM). Eventually, the evaluation of protocol performance was established by implementing a prototype of it and performing experiments. The experimentally obtained outcomes were compared to other existing protocols and have shown huge
adaptability and efficiency of the proposed protocol by clients which their resources were limited.

IV. DISCUSSION AND ASSESSMENT

With cloud computing technologies several applications are used, also the large part of these applications and data are applied and used in particle physics fields, information retrieval, etc. However, there are various techniques used to increase the performance of cloud computing. The cloud term widely used in some companies but without useful and understand the cloud completely. The cloud tries to use technologies that make cloud computing more accessible and reached by anyone at any time. There are many technologies used to increase the performance of cloud computing, but the most important of them are Hadoop, Dryad, CGL Map Reduce, MPI, and CAP3.

In this paper, we explained cloud computing in detail with several techniques used to increase the performance of cloud computing. Also, we explained Cloud Burst, Hadoop, Dryad, High Energy Physics (HEP), and CGL Map Reduce and compare among them according to various criteria. The comparison has shown that each technique has advantages and disadvantages. On the other side, we survey the study of seven authors; each author uses several techniques to evaluate the performance of the clouds. Explain the objectives of each one of them by discussing the goals for all authors and which methods they used. Table 2 explains the comparison of the performance of cloud computing.

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Research’s objectives</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ekanayake et al.</td>
<td>2015</td>
<td>Compute the measurement of high performance of cloud computing</td>
<td>Use some technologies and compared among them to explain which one is better for huge dat.</td>
</tr>
<tr>
<td>Shakeabubakor et al.</td>
<td>2015</td>
<td>Increase the person knowledge about the cloud computing and their technologies</td>
<td>Suppose the companies and organization to use cloud computing instead the traditional technologies.</td>
</tr>
<tr>
<td>Vecchiola et al.</td>
<td>2015</td>
<td>Rising a performance of cloud computing by using some technologies</td>
<td>Compared among some technologies and explain the best adoption technologies can used in companies.</td>
</tr>
<tr>
<td>Stantchev [21]</td>
<td>2015</td>
<td>Evaluation of cloud computing technologies by Windows applications.</td>
<td>Used some method and techniques to merge the NFPS and service levels in same approach to be suitable for windows application.</td>
</tr>
<tr>
<td>Kumar [25]</td>
<td>2016</td>
<td>Cloud computing used in many aspects as economy, Engineering and health</td>
<td>The cloud computing provides many of resources and cluster computer.</td>
</tr>
<tr>
<td>Attaran [22]</td>
<td>2017</td>
<td>Comparative among cloud technologies with the traditional approaches</td>
<td>Provide the best way to use the cloud by companies even the end user to keep their information in efficient manner.</td>
</tr>
<tr>
<td>Kannadasan et al.</td>
<td>2018</td>
<td>Use cloud computing technologies to explain the performance of clouds</td>
<td>Considers the Hadoop and MapReduce best technologies used to increase performance of cloud computing.</td>
</tr>
<tr>
<td>Rashid et al. [18]</td>
<td>2019</td>
<td>hing able to use the power of processing via cloud domain depending on the principles of distributed systems.</td>
<td>Using distributed parallel-processing and cloud computing for solving (in minimum time) specific/huge problems of client that require special processing energy.</td>
</tr>
<tr>
<td>Asadi et al. [31]</td>
<td>2019</td>
<td>analyze the influence of network traffics and temperature on the performance and strength of cloud computing systems</td>
<td>analytical model and numerical evaluation for the performance and power of cloud infrastructure by utilizing the stochastic reward nets (SRNs).</td>
</tr>
<tr>
<td>Garg et al. [32]</td>
<td>2020</td>
<td>reduce the computational complexity of client during system setup phase of the auditing protocol</td>
<td>an efficient model for data integrity auditing cloud computing which its security relies upon the stability CDHP in a Random Oracle Model (ROM).</td>
</tr>
</tbody>
</table>

V. CONCLUSION

Cloud computing is a future technology that will be used by everyone because it solves many of the problems facing the companies, organizations, and the end-user. In this paper, we survey the study of cloud computing technologies and methods that use to increase the performance of the cloud. Also, each technology has several features to increase the performance of
the cloud, and each of them uses different algorithms. However, each technology maintains the aspect of cloud computing, and each one used for a different purpose. We compare among several authors; each author uses different techniques for different purposes, and explain some technologies used to increase cloud performance and Comparative among them in several criteria, the most technologies used of maintaining performance are Hadoop, Dryad, CGL Map Reduce, High Energy Physics (HEP) and Cloud Burst.

In most cases, Hadoop is better and has the edge over all other technologies; the time of execution for Hadoop in the most application is faster because the number of the partition of data is significant. When the data is enormous, all technologies have proximately the same features, but the Dryad is very slow then other technologies. The Hadoop is also more accessible for users, and it works with the most operating systems. The High Energy Physics (HEP) also distinguishes for having an accuracy of information when the data is enormous. On the other side, the best one of them for achieving data from different cloud techniques is the Cloud Burst. Cloud computing continues growing, and there are many technologies used to increase the performance of the cloud, and each one of them has some advantages and disadvantages.

REFERENCES