

Novel Cloud Architecture to Decrease Problems Related to Big Data

Entesar Althagafy and M. Rizwan Jameel Qureshi

Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia
E-mail: entesar.althagafy@hotmail.com, anriz@hotmail.com

Abstract—IT companies are facing many difficulties and challenges when dealing with big data. These difficulties have surfaced due to the ever-expanding amount of data generated via personal computer, mobile devices, and social network sites. The significant increase in big data has created challenges for IT companies that need to extract necessary information and knowledge. Cloud computing, with its virtualized resources usage and dynamic scalability, is broadly used in organizations to address challenges related to big data and has an important influence on business in organizations. Furthermore, big data is changing the way organizations do business. This paper proposes novel cloud architecture to decrease problems related to big data. The proposed architecture is a combination of many big data infrastructures in the creation of a service. This architecture minimizes problems related to big data by improving performance and quality of service.

Index Terms—Cloud Computing, Big Data Problem, Big Data Infrastructure, AWS, Eucalyptus, Hadoop.

I. INTRODUCTION

In the past few years, a growing number of organizations are facing the challenge of explosive data growth, and the sizes of databases used have been rapidly increasing. The big data explosion is being generated by many sources such as online transactions, business processes, photos, videos through social network sites, web servers, and server logs. Processing or analyzing the huge amount of data to extract meaningful information is a challenging task [1].

Cloud computing has been driven mainly by the need to process this expanding mass of data in terms of Exabyte as we are near to the Zettabyte era [2]. One of the primary trends in cloud computing is big data. When an organization needs to store and access more data, cloud computing is one of the best solutions in terms of cost efficiency and rapid scalability, both of which are essential features when dealing with big data [2].

Cloud computing can be described as “on-demand network access to computing resources, provided by an outside entity” [2]. There are four types of service models for cloud computing: software as a service (SaaS), hardware as a service (HaaS), platform as a service (PaaS), and infrastructure as a service (IaaS). The SaaS model supplies businesses with applications that are

saved and operate in the cloud on virtual servers. The HaaS model is a cloud service that depends on the model of time sharing on mainframes and minicomputers. The PaaS model employs cloud computing to supply platforms for the evolution and use of specific applications. The IaaS model provides resources as services to customers (primarily businesses). The customer will pay on a per-use basis. The IaaS model will provide high flexibility because extra resources are always available [2-3].

The common types of implementation models for cloud computing are the private, public and hybrid clouds. A private cloud is run exclusively for a single organization and can be handled within the organization or by a third-party and hosted either externally or internally. The public cloud provides pay-as-you-go services, and services are rendered over a network. A hybrid cloud is a mixture of two or more private clouds and public clouds and provides the advantages of several implementation models [2]. The widespread usage of cloud computing to deal with the big-data technology is due to three significant reasons i.e., 1) the ability to examine the value of big-data, 2) less processing cost and 3) reduce hardware cost [2].

This paper proposes a novel architecture integrating the Amazon Web Service (AWS) Remote Cloud, Eucalyptus, and Hadoop. The proposed architecture is composed of hybrid infrastructure elements to minimize problems related to big data improving performance and quality of service (QoS).

The paper is organized as follows. Section 2 outlines the related work. Section 3 describes the problem statement. Section 4 depicts the details of the proposed solution. Section 5 contains the discussion and section 6 provides the validation of proposed solution.

II. RELATED WORK

Most companies are currently dealing with challenges related to big data. These companies are spending millions in order to extract useful information from big data for two reasons. First, having a strong information base makes it easier for management to make decisions and address challenges and issues associated with big data. Adnan et al. [4] proposed cloud computing as one of the solutions for the problems those companies are facing related to big data. These researchers encourage IT companies to use Hadoop architecture to establish

public/private clouds. These clouds will facilitate reliable distributed computing to process big data more effectively and efficiently. A setup is proposed by Adnan et al. [4] to comply efficiently with computational performance and the capacity requirements of incoming big data requests submitted by various clients.

A solution to big-data problems in small- and medium-sized companies through cloud computing is proposed by Purcell [5]. The major reasons for small- and medium-sized companies to use cloud computing are processing cost reduction, hardware cost reduction, and ability to test the value of big data.

Patel et al. [1] proposed using the Hadoop cluster as a solution to the big-data problem. Patel et al. [1] accomplished prototype implementation of the Hadoop Distributed File System (HDFS) for storage, the Hadoop cluster, and the MapReduce framework for analyzing large data sets by studying prototypes of big-data application scenarios. The initial results with these prototypes were promising with respect to addressing the big-data problem [1].

Padhy [6] proposed Hadoop and MapReduce for processing and analyzing big data in a cloud system. Hadoop HDFS is an open-source file system with flexible scalability and allows data storage in any form without the need to have data types [6]. MapReduce is a programming model for writing applications that immediately process massive amounts of data in parallel on big clusters of compute nodes [6].

Several big-data processing techniques are proposed by Ji et al. [7]. The techniques [7] are proposed from the aspect of big-data processing mechanisms and cloud data management. Ji et al. [7] proposed MapReduce strategies and applications as optimal techniques for big-data processing.

Big-data storage is an essential element of cloud computing. Hao and Ying [8] proposed the Hadoop platform and the MapReduce algorithm as a solution to big data storage, merging data-mining technology using the K-means. The objective of their work was to implement effectiveness analysis and application of their cloud-computing platform [8].

Ramamoorthy and Rajalakshmi [9] propose a system that provides service and data analysis based on a MapReduce algorithm with big-data analytics techniques. This system has three features: it automatically increases the number of users in the cloud, effectively analyzes the data stored on cloud storage, and decreases the complexity of services that are frequently used in cloud computing.

Ye et al. [10] propose a cloud-based big data mining and analyzing services platform by integrating R. The proposed services platform [10] provides rich data statistical and analytic functions. The architecture of the services platform consists of four layers: a services layer, a dataset processing layer, a virtualization layer, and an infrastructure layer.

Another common challenge confronted with big data is that it is collected from several databases, increasing the difficulty of processing via the usual database

administration tools or applications. In response to this challenge, Zhang and Huang [11] propose a framework for big-data visualization and analysis called the 5Ws model, which analyzes the 5Ws data dimensions and the density used for big data types. This process detects any correlations that conventional database administration tools are unable to reveal.

The NoSQL Database, one of the most-often used big-data analytics techniques, is proposed [2]. A NoSQL Database is composed of distributed non-relational data stores, horizontally scalable. The proposed technique has three characteristics: strong consistency, high availability, and partition tolerance [2].

Gu et al. [12] reviewed multiple open-source cloud storage platforms due to extensive usage of these platforms in cloud computing. Gu et al. [12] discussed newly open-source cloud computing systems like Hadoop, Cassandra, abiCloud, and MongoDB. The objective is to provide details of the platforms in terms of architecture, implementation technology, and functions. In addition, comparison and contrast of newly open-source cloud-computing systems are provided so that IT companies can select a suitable cloud storage platform as per their requirements [12].

Although big data is a relatively new concept, it affects IT companies and the way they do business [13]. Katal et al. [13] describe the new concept of big data while also covering its importance and several challenges and issues related to big data technology. Katal et al. [13] propose a solution to the big-data problem through the use of Hadoop architecture.

The current trends, characteristics, and analysis of big data are discussed by Ahuja and Moore [14]. The challenges in data management, collection, and storage in cloud computing are also covered. Hadoop, Cassandra, and Voldemort are platforms proposed by Ahuja and Moore [14] as being part of the solution to the big-data problem.

Shamsi et al. [15] analyze data-intensive cloud computing requirements and describe the challenges. Numerous solutions for requirements and challenges are provided. Shamsi et al. [15] provide a survey of the solutions and an analysis of their efficiencies in meeting the emerging needs of diverse applications.

Amiry et al. [16] propose a method to implement the Hadoop platform on a Eucalyptus infrastructure. The proposed method helps with and speeds up the implementation of a Hadoop platform and furthermore emphasizes the role of Eucalyptus open-source in providing services to the cloud customer [16].

A hybrid structure of cloud computing is used to build an interactive graphical user interface for Bangladeshi People Search (BDPS) [17]. The proposed framework is necessary and feasible for handling vast data. Amazon Web Service, Eucalyptus, and Hadoop are used in the proposed framework [17].

The concept of cloud computing is described by Jie et al. [18]. The architecture, implementation mechanism, characteristics, and various forms of cloud computing are also covered. Google techniques are used as an example

for cloud computing.

All the aforementioned papers have some limitation in

their solutions for problems related to big data. These limitations are presented in Table 1.

Table 1. The Summary of Limitations

Title	Limitations
Addressing Big Data Problem Using Hadoop and Map Reduce [1]	Does not cover performance evaluation and modeling of Hadoop applications on cloud platforms.
Big Data Analytics: Hadoop-Map Reduce & NoSQL Databases [2]	Does not provide an explicit comparison between Hadoop-MapReduce & NoSQL Databases.
Minimizing Big Data Problems using Cloud Computing Based on Hadoop Architecture [4]	The paper does not cover master cloud modeling and architecture. The paper does not solve the big-data storage problem after analysis of customer requests.
Big data using cloud computing [5]	The two major concerns of using cloud-computing models are data security and loss of control.
Big Data Processing with Hadoop-MapReduce in Cloud Systems [6]	Does not provide: A brief overview of cloud computing. A comparison between proposed platforms and other platforms.
Big data processing in cloud computing environments [7]	Does not provide a concrete realization of services or the platform.
Research of Cloud Computing based on the Hadoop platform [8]	Does not cover Hadoop Common and Hadoop YARN.
Optimized data analysis in cloud using BigData analytics techniques [9]	Does not provide a comparison between the proposed schemes with other works.
Cloud-Based Big Data Mining & Analyzing Services Platform Integrating R [10]	<ul style="list-style-type: none"> ▪ The proposed services [17] are incomplete. ▪ The services platform is preliminary.
5Ws model for Big Data analysis and visualization [11]	Proposed visualization methods cannot display big data patterns crossing numerous datasets.
An Overview of Newly Open-Source Cloud Storage Platforms [12]	Does not cover Google BigQuery, Rackspace Big Data Cloud, Voldemort, and Amazon Web Services.
Big data: Issues, challenges, tools and Good practices [13]	Does not mention the following challenges and issues: Flexibility and efficient user access. Data handling, locality, and placement. Elasticity. Power efficiency. Network problem.
State of Big Data Analysis in the Cloud [14]	Does not provide a brief overview of cloud-computing types and deployment models.
Cloud computing and its key techniques [18]	Does not describe cloud-computing types and deployment models.

III. RESEARCH PROBLEM

IT companies are dealing with a large mass of data and face many difficulties in big-data processing, analysis, and storage. The significant increase in big data has created challenges for IT companies that need to extract necessary information and knowledge. The research question is formulated as follows [4-5,13]:

How to improve affordability, computational performance, and capacity requirements of incoming big-data requests that require analysis and storage?

IV. THE PROPOSED SOLUTION

This research proposes an improved solution that is based on the recommendations of existing work [4,16-17]. Novel cloud architecture is proposed to decrease problems related to big data. The proposed novel architecture integrates the Amazon Web Service Remote Cloud, Eucalyptus, and Hadoop to improve performance and quality of service as shown in fig. 1. The proposed novel cloud architecture works as follows.

AWS Remote Cloud accepts all incoming big data requests and intelligently forwards these requests to the best proper Eucalyptus Local Clouds.

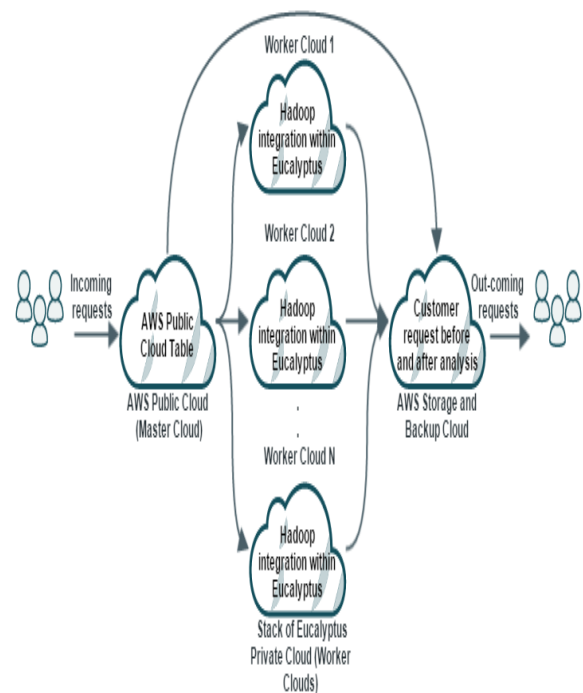


Fig.1. Mobility in SOA (Request-Response).

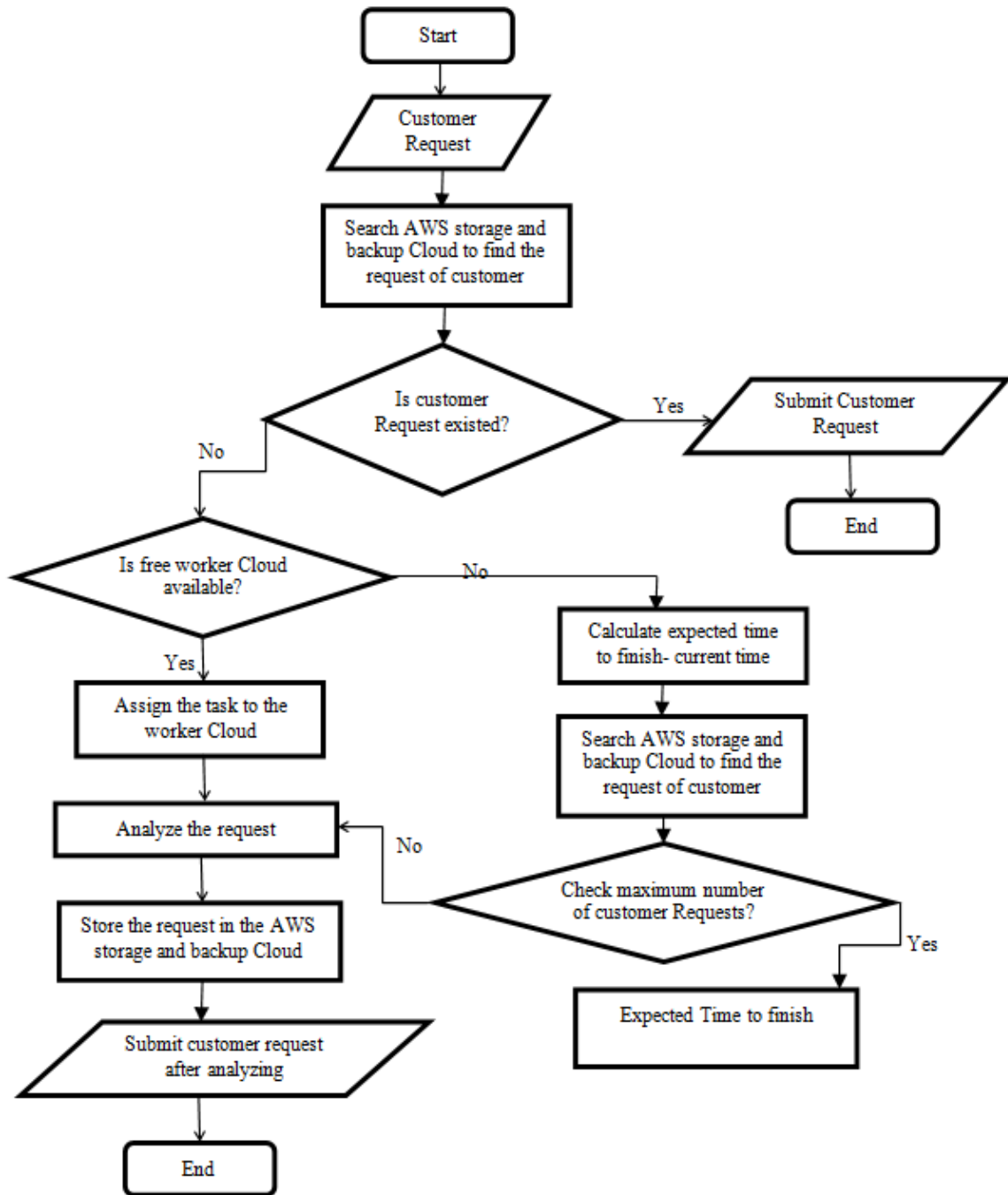


Fig.2. Flow Chart of the Proposed Solution.

Table 2. AWS Remote Cloud Table

Local Clouds	Expected TimeToFinish	Number Of Current Request	Number Of Max Request	Situation
Cloud1	Time1	Integer1	Integer1	Free / occupied
Cloud2	Time2	Integer2	Integer2	Free / occupied
....
CloudN	TimeN	IntegerN	IntegerN	Free / occupied

The calculation process between the CurrentTime and ExpectedTimeToFinish field in the AWS Remote Cloud table enables the fulfillment of these requests in the shortest possible time span. After a customer requests analysis, it will be automatically stored in the AWS Storage and Backup Cloud. The proposed novel cloud architecture is explained as follows.

A. AWS Remote Cloud (Master Cloud)

The infrastructure of the Remote Cloud will be implemented on the AWS cloud, which is a set of remote computing services that commonly form a cloud computing environment, offered by Amazon.com. The

AWS Remote Cloud is the master cloud. The master cloud will be used to receive incoming big data requests those are submitted by numerous customers. All incoming customer requests are transformed to the master cloud, which analyzes the big data request size and determines the best worker clouds depending on the calculation process between the CurrentTime and ExpectedTimeToFinish field in the AWS Remote Cloud table. The incoming request intelligently redirects to worker clouds by the master cloud.

B. Block of Eucalyptus Local Cloud (Worker Clouds)

The infrastructure of the Local Cloud will be implemented using open-source Eucalyptus. The Eucalyptus Local Cloud is the worker cloud. A block of worker clouds is registered along with the master cloud. One worker cloud contains a cluster of powerful machines. If the size of the cloud cluster grows, the performance and capacity increases. The block of the Eucalyptus Local Cloud (worker cloud) is based on Hadoop integration within Eucalyptus in order to obtain good performance results.

Eucalyptus is an open-source application for cloud-computing environments. Hybrid clouds consisting of remote and local clouds can be built by using Eucalyptus. The Eucalyptus system allows us to begin, manage, access, and stop entire virtual machines.

Hadoop is a MapReduce framework that works on the HDFS, accommodating high-throughput access to application data with the capability to store large amounts of data across thousands of servers.

C. AWS Storage and Backup Cloud

The infrastructure of the Storage and Backup Cloud will be implemented on the AWS cloud. The Remote AWS Storage and Backup Cloud will be used for backup and mirroring and will ensure backup through 7X24. All customers' requests before and after analysis are transformed to the Storage and Backup cloud and will be automatically stored. The least-frequently used requests will be automatically deleted from the cloud; the most-frequently used requests will remain in storage.

D. AWS Remote Cloud Table

This cloud architecture requires master cloud control using an AWS Remote Cloud table that consists of five fields.

- LocalClouds field represents a list of registered clouds within the Eucalyptus infrastructure. Any worker cloud available in the list can be added or removed.
- ExpectedTimeToFinish field represents the total expected time a particular worker cloud will take to finish the analysis of all customer requests that are being simultaneously analyzed; this is determined by the worker cloud itself and will be sent back to the AWS Remote Cloud.
- NumberOfCurrentRequest field represents the number of current customer requests that has been

accomplished by a particular worker cloud.

- NumberOfMaxRequest field represents the maximum number of customer requests that can be accomplished by a particular worker cloud in the same timeframe. This value depends on the number of machines in the cloud.
- Situation field represents the current situation of the worker cloud. It can either be free or occupied. If the status is free, the master cloud can assign the incoming request to that particular worker cloud. Conversely, if the status is occupied it means the cloud is analyzing one or more requests, and the master cloud needs to determine which worker cloud will finish first; that worker cloud then receives the next request. The worker cloud that will receive the next request is determined depending on the following equation.

$$R = E - C. \quad (1)$$

'R' denotes the upcoming request to worker cloud, 'E' shows the expected time to finish and 'C' expresses the current time. After applying the (1) to all worker clouds in the AWS Remote Cloud, the worker cloud, which has the least residual time to finish, will receive the next request. Then, the NumberOfCurrentRequest will be compared with NumberOfMaxRequest; if they are equal, this means the chosen cloud reaches the maximum number of requests that can be simultaneously analyzed. The master cloud then must choose another worker cloud to receive the request.

The five fields of the AWS Remote Cloud table are shown in Table 2. The Flow Chart for the complete sequences for analyzing and assigning requests to particular worker cloud is shown in fig. 2.

E. The Advantages of the Proposed Solution

The proposed novel cloud architecture improves computational performance against incoming big data requests from multiple customers.

- Can easily increase the storage capacity, performance, and processing power at any time by increasing the block of worker clouds.
- Can easily expand AWS Remote Cloud and AWS Storage and Backup Cloud storage capacity and processing power at low cost.
- The proposed novel cloud architecture is easy to set up at any company or university.
- Customers' requests after analysis are stored in the AWS Storage and Backup Cloud, protecting from any loss or damage.
- Eucalyptus is compatible with AWS and Hadoop.
- Hadoop and Eucalyptus are open-source and scalable platforms.

V. VALIDATION OF THE PROPOSED SOLUTION

The proposed architecture suggested earlier must be

validated. For this purpose, a closed-ended questionnaire was distributed among IT students and staff and it was comprised of twenty questions. Six goals were designed to collect data.

- Goal 1: Suitability of the proposed novel cloud architecture for decreasing big-data problems.
- Goal 2: The effect of the used clouds on cost saving.
- Goal 3: The efficiency of the scheduling algorithm used by the AWS Remote Cloud.
- Goal 4: Frequency with which AWS, Eucalyptus, and Hadoop are used in companies.
- Goal 5: The effect of saving requests before and after analysis with respect to saving time and storage space.
- Goal 6: The flexibility to add new clouds and increase storage space.

The questions were answered using a 5-point Likert scale. The responders were 35 IT students and staff members. Google tools and Microsoft Excel were used to collect and analyze data.

A. Goal 1. Suitability of the proposed novel cloud architecture to decrease big-data-related problems

The questions in this goal are designed to measure the suitability of the proposed novel cloud architecture to decrease the problems of big data. Table 3 shows that 41% of the respondents agree with the suitability of the proposed solution whereas 41% of the participants strongly agree. Furthermore, 12% of the software engineers neither agree nor disagree with the goal 1. However, 4% of the professionals disagree while 2% of the IT professionals strongly disagree. Fig. 3 shows the results of Table 3 graphically.

Table 3. Cumulative Analysis of Goal 1

Q. No	Very Low	Low	Nominal	High	Very High
1	0	0	5	15	15
2	1	2	0	15	17
3	1	3	7	15	9
4	1	1	5	12	16
Total	3	6	17	57	57
Avg.	2%	4%	12%	41%	41%

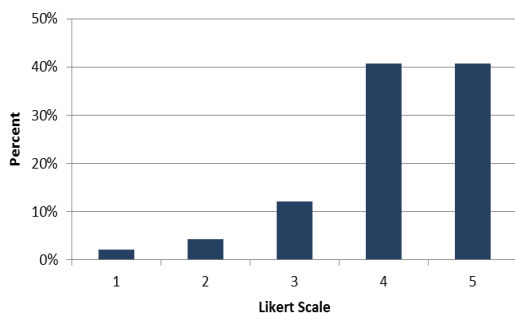


Fig.3. Cumulative Analysis of Goal 1.

B. Goal 2. The effect of the used clouds on cost saving

Table 4 gives information about goal 2. Table 4 shows that 34% of the professionals are strongly agreed with goal 2 and 44% the participants are agreed. Moreover, 16% of the software engineers neither agree nor disagree. 3% of the respondents disagree and 3% of the professionals strongly disagree with the goal 2. Fig. 4 shows the results of Table 4 graphically.

Table 4. Cumulative Analysis of Goal 2

Question No	Very Low	Low	Nominal	High	Very High
1	1	0	3	16	15
2	1	1	8	12	13
3	1	2	6	18	8
Total	3	3	17	46	36
Avg.	3%	3%	16%	44%	34%

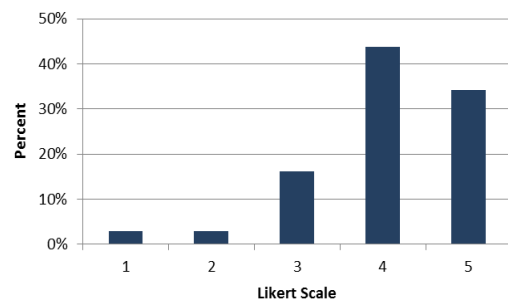


Fig.4. Cumulative Analysis of Goal 2.

C. Goal 3. The efficiency of the scheduling algorithm used by the AWS Remote Cloud

Table 5 shows 34% of the professionals strongly agree with goal 3 and 43% of the participants agree. Moreover, 16% of the software engineers neither agree nor disagree. 5% of the respondents disagree and 1% of the professionals strongly disagree with the goal 3. Fig. 5 shows this graphically as follows.

Table 5. Cumulative Analysis of Goal 3

Q. No.	Very Low	Low	Nominal	High	Very High
1	0	0	4	18	13
2	0	2	8	15	10
3	1	3	4	14	13
4	0	2	6	14	12
Total	1	7	23	61	48
Avg.	1%	5%	16%	43%	34%

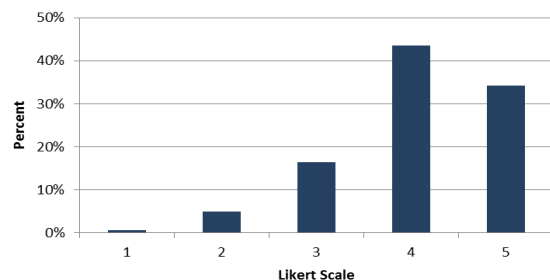


Fig.5. Cumulative Analysis of Goal 3.

D. Goal 4. Frequency with which AWS, Eucalyptus, and Hadoop are used in companies

Table 6 shows 42% of the professionals strongly agree with goal 4 and 49% of the participants agreed. Moreover, 8% of the software engineers neither agree nor disagree. 4% of the respondents disagree with the goal 4. Fig. 6 shows the results of Table 6 graphically.

Table 6. Cumulative Analysis of Goal 4

Question No	Very Low	Low	Nominal	High	Very High
1	0	0	2	15	18
2	0	1	2	20	12
3	0	1	4	16	14
Total	0	4	8	51	44
Avg.	0%	4%	8%	49%	42%

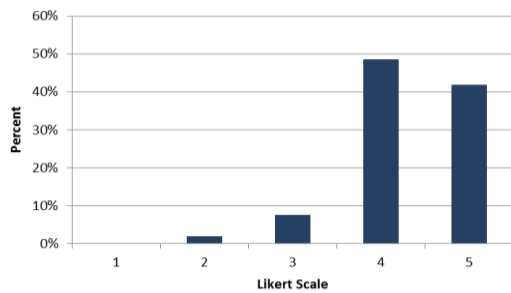


Fig.6. Cumulative Analysis of Goal 4.

E. Goal 5. The efficiency of the scheduling algorithm used by the AWS Remote Cloud

Table 7 shows 40% of the professionals strongly agree with goal 5 and 46% of the participants agree. Moreover, 11% of the software engineers neither agree nor disagree. 3% of the respondents disagree with the goal 5. 1% of the professionals strongly disagree with this goal. Fig. 7 shows the results of Table 7 graphically.

Table 7. Cumulative Analysis of Goal 5

Question No	Very Low	Low	Nominal	High	Very High
1	0	2	5	16	12
2	1	1	3	16	14
3	0	1	4	17	13
Total	1	3	12	48	42
Avg.	1%	3%	11%	46%	40%

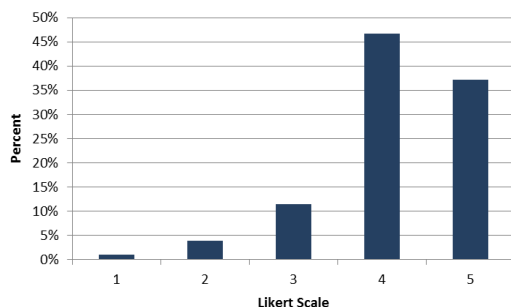


Fig.7. Cumulative Analysis of Goal 5.

F. Goal 6. The efficiency of the scheduling algorithm used by the AWS Remote Cloud

Table 8 shows 41% of the professionals strongly agree with goal 6 and 47% of the participants agree. Moreover, 10% of the software engineers neither agree nor disagree. 3% of the respondents disagree with the goal 6. Fig. 8 shows the results of Table 8 graphically.

Table 8. Cumulative Analysis of Goal 6

Question No	Very Low	Low	Nominal	High	Very High
1	0	1	6	15	13
2	0	2	2	16	15
3	0	0	4	17	14
Total	0	3	10	49	43
Avg.	0%	3%	10%	47%	41%

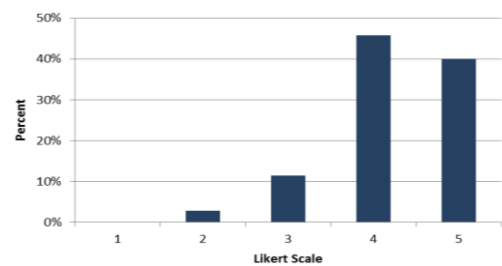


Fig.8. Cumulative Analysis of Goal 6.

G. Final cumulative analysis of 6 goals

Table 9 shows 38% of the professionals strongly agree with six goals and 45% of the participants agree. As such 83% support is available. Moreover, 13% of the software engineers neither agree nor disagree. 4% of the respondents disagree and 1% of the professionals strongly disagree with the six goals. Fig. 9 shows the results of Table 9 graphically.

Table 9. Final Cumulative Analysis of 6 Goals

Goal No.	Very Low	Low	Nominal	High	Very High
Goal 1	3	6	17	57	57
Goal 2	3	3	17	46	36
Goal 3	1	7	23	61	48
Goal 4	0	2	8	51	44
Goal 5	1	4	12	49	39
Goal 6	0	3	12	48	42
Total	8	25	89	312	266
Avg.	1%	4%	13%	45%	38%

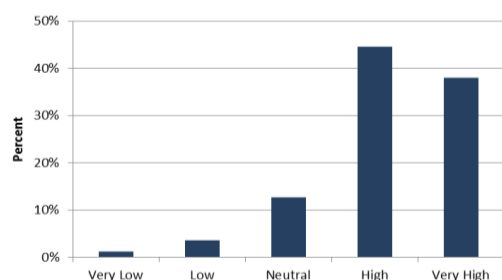


Fig.9. Final Cumulative Analysis of Six Goals

VI. CONCLUSION

Processing, analyzing, and storing big data are not easy tasks. It costs a lot of time and money when data is dealt with the wrong way. IT companies, whether small or large, are dealing with the problems related to big data. Many companies believe that cloud computing is one of the best solutions by considering that cloud computing has the advantages of cost efficiency and rapid scalability. The novel cloud architecture is proposed in this paper that integrates AWS, Eucalyptus, and Hadoop (which collectively) forming the best big-data infrastructure as one service. It is anticipated that this architecture will overcome the problems related to big data by reducing time and cost, improving performance and quality of service. The proposed architecture enhances computational performance against incoming big data requests from multiple customers and it is easy to set up at any company or university. A finding that was validated by this study, as 83% of the sample expressed approval of the proposed architecture.

REFERENCES

- [1] Patel A. and Birla M. and Nair U., "Addressing big data problem using Hadoop and Map Reduce," 2012 Nirma University International Conference on Engineering, Gujarat, India, pp. 1-5, 2012.
- [2] Pothuganti A., "Big Data Analytics: Hadoop-Map Reduce & NoSQL Databases."
- [3] Frantsvog D. and Seymour T. and John F. "Cloud Computing," International Journal of Management & Information Systems (IJMIS), vol. 16, no. 4, pp. 317-324, 2012.
- [4] Adnan M., et al., "Minimizing big data problems using cloud computing based on Hadoop architecture," 2014 11th Annual High-capacity Optical Networks and Emerging/Enabling Technologies (HONET), 2014.
- [5] Purcell B., "Big data using cloud computing," Holy Family University Journal of Technology Research, 2013.
- [6] Padhy R., "Big Data Processing with Hadoop-MapReduce in Cloud Systems," International Journal of Cloud Computing and Services Science (IJ-CLOSER), vol. 2, no. 1, pp. 16-27, 2012.
- [7] Ji C., et al., "Big data processing in cloud computing environments," 2012 International Symposium on Pervasive Systems, Algorithms and Networks (I-SPAN), San Marcos, Texas, 2012.
- [8] Hao, Chen, and Qiao Ying. "Research of Cloud Computing based on the Hadoop platform." Computational and Information Sciences (ICCIS), 2011 International Conference on. IEEE, 2011.
- [9] Ramamoorthy S. and Rajalakshmi S., "Optimized data analysis in cloud using BigData analytics techniques," 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), Tiruchengode, India, pp. 1-5, 2013.
- [10] Ye F, et al., "Cloud-Based Big Data Mining & Analyzing Services Platform Integrating R," 2013 International Conference on Advanced Cloud and Big Data (CBD), Nanjing, China, pp. 147-151, 2013.
- [11] Zhang J. and Huang M., "5Ws model for Big Data analysis and visualization," 2013 IEEE 16th International Conference on Computational Science and Engineering (CSE), Sydney, Australia, pp. 1021-1028, 2013.
- [12] Gu G., et al., "An overview of newly open-source cloud storage platforms," 2012 IEEE International Conference on Granular Computing (GrC), Hangzhou, China, pp. 142-147, 2012.
- [13] Katal A. and Wazid M. and Goudar R., "Big data: Issues, challenges, tools and Good practices," 2013 Sixth International Conference on Contemporary Computing (IC3), Noida, India, pp. 404-409, 2013.
- [14] Ahuja S. and Moore B., "State of Big Data Analysis in the Cloud," Network and Communication Technologies, vol. 2, no. 1, pp. p62, 2013.
- [15] Shamsi J., Khojaye M., and Qasmi M. "Data-intensive cloud computing: requirements, expectations, challenges, and solutions," Journal of grid computing, vol. 11, no. 2, pp. 281-310, 2013.
- [16] Amiry V., et al., "Implementing Hadoop Platform on Eucalyptus Cloud Infrastructure," 2012 Seventh International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), Victoria, Canada, pp. 74-78, 2012.
- [17] Tarannum N., and Ahmed N., "Efficient and Reliable Hybrid Cloud Architecture for big Database," International Journal on Cloud Computing: Services and Architecture (IJCCSA), vol. 3, no. 6, pp. 17-29, 2013.
- [18] Jie S. and Yao J. and Wu C., "Cloud computing and its key techniques," 2011 International Conference on Electronic and Mechanical Engineering and Information Technology (EMEIT), Heilongjiang, China, vol. 1, pp. 320-324, 2011.

Authors' Profiles



Dr. M. Rizwan Jameel Qureshi received his Ph.D. degree from National College of Business Administration & Economics, Pakistan 2009. He is currently working as an Associate Professor in the Department of IT, King Abdulaziz University, Jeddah, Saudi Arabia. This author is the best researcher awardees from the Department of Information Technology, King Abdulaziz University in 2013 and 2016. He is also honoured as the best researcher from the Department of Computer Science, COMSATS Institute of Information Technology, Pakistan in 2008.

Entesar Althagafy is a master degree student in the Department of Information Technology, King Abdulaziz University.

How to cite this paper: Entesar Althagafy, M. Rizwan Jameel Qureshi, "Novel Cloud Architecture to Decrease Problems Related to Big Data", International Journal of Computer Network and Information Security(IJCNIS), Vol.9, No.2, pp.53-60, 2017.DOI: 10.5815/ijcnis.2017.02.07