Cloud Computing Serving as a Solution to the IoT Generated Data

Fahmida Naveed, Muhammad Rizwan, Aysha Shabbir, Maryam Shabbir and Fahad Ahmad

***Abstract* – Nowadays all the devices seem to be connected to the Internet. There are billions and billions of Internet of Things (IoT) devices. These devices create an immense amount of data. IoT devices have limited processing capabilities and storage, and to handle all the data generated from billions and billions of these devices is a task itself. Handling the Big Data generated by IoT devices poses a problem and it is quite a significant one, it is one of the few problems presented in this paper. The paper presents the integration of the IoT with the Cloud, the challenges imposed by the IoT devices and how its integration with the Cloud Computing would help us, its advantages and problems are highlighted. A possible solution to handle all the Big Data is also proposed.**

Index Terms – IoT, Cloud Computing, Big Data, Hadoop

I. INTRODUCTION

Internet of Things (IoT) is considered the next uprising in the world of Internet. It is already being adopted in a number of dimensions of our daily lives. It is a way to communicate abundance of information among various devices to improve the quality of our everyday lives. A global networked framework is able to self-configure things in an extremely intelligent way. The world is headed for a time where all the devices would be connected to the Internet and would be able to communicate with each other without humans intervening in between. However, IoT also have a few limitations attached to it. It has a limited computing capacity with a limited storage [1].

On the other hand, how technologies are now accessed, managed and delivered has also been changed due to Cloud Computing. It offers services through which computing resources could be shared across the Internet. Experts think that Cloud Computing will aid in utility services in the coming time. The services offered by the Cloud are not free of cost, they come with a price and so it must be handled very wisely. The services are dependent upon the service providers [2].

There is a lot of diversity when it comes to IoT, whereas cloud based services are quite interoperable [1] [2]. Both technologies are being rapidly developed and taking over the world fast. They have unique characteristics. Since their characteristics are complementary, experts think that their integration is the best option for the future. It could be the next revolution for the future generations. The other, hence giving us a major break, could provide what lacks in one technology. Therefore, the integration of these technologies can be very beneficial.

Fahmida Naveed, Muhammad Rizwan, Aysha Shabbir, Maryam Shabbir and Fahad Ahmad are with Department of Computer Science Kinnaird College for Women, Lahore, Pakistan. Email: fahmidanaveed@ hotmail.com, mrizwan.hust@yahoo.com, arzoo_982@yahoo.com, maryamshabbir_564@yahoo.com, fahad.ahmad@kinnaird.edu.pk.
Manuscript received on Dec 29, 2018 revised on Feb 18, 2019, May 11, 2019 and accepted on Aug 25, 2019.

Though the integration of these technologies can result in a revolutionary technology, the integration is not that simple. Many hurdles are faced when integrating both of them. One of the few issues faced during their integration is of handling the big data. Big data management framework is required for its smooth handling. This paper would propose a solution for handling big data generated by IoT.

This paper gives a hybrid solution to the problem of big data management. It will actually comprise of both cloud and Hadoop technologies against (IoT generated data) big data management. Three noteworthy purposes behind organizations and communities concern to utilize cloud computing and Hadoop for big data analytics is cost limited processing, hardware charges downsizing, and capacity to examine big data value. IOT generated big data analytics demands an extensive level of secure storage space. Cloud computing and Hadoop based hybrid scheme against secure data storage is a suitable choice for communities. As, it provides on-demand computing with data confidentiality, integrity, availability ensuring algorithms which requires little administration exertion by the organizations. Due to Hadoop and cloud extensive security measures, the proposed hybrid solution would be greatly assistive against secure big data management. The rest of the paper is structured as: Section 2 defines the problem statement, Section 3 presents the elementary ideas, Section 4 discusses the advantages, applications and challenges of Cloud and IoT, Section 5 discusses the challenge of managing big data effectively and its proposed solution and the last section concludes the discussion.

II. PROBLEM STATEMENT

Big data is one of the challenges faced during the integration of these technologies. Big data management is a significant problem itself. How the data will be transferred from the IoT devices into the cloud, processed, analyzed with timeliness and storage poses great challenges. Extensive storage space is required for big data management with secure monitoring measures. The storage cost is keep on declining; the computing resources required for big data management can even now present budgetary troubles for low to high-level organizations.

III. BASIC CONCEPTS

A. Internet of Things(IoT)

Nowadays, everything and every object, ranging from lights, cars to home and home appliances are being turned into smart devices, as shown in figure1. All the physical devices are being converted to smart device, which in turn generate a lot of information. All the devices are connected to the Internet and have their own unique identity. Hence providing them with the ability to communicate with each

 ISSN – 1999-4974

other. The boundary between the digital and real world is being removed very rapidly.

Kevin Ashton in 1999 was the first person to state, "The Internet of Things has the potential to change the world, just as the Internet did. Maybe even more so" [2]. He realized the importance of IoT long before others did. Another definition explains the IoT as, "The IoT infrastructure allows combination of smart objects (i.e. wireless sensors, mobile robots, radio frequency identification systems, etc.), sensor network technologies, and human beings, using different but interoperable communication protocols and realizes a dynamic heterogeneous network that can be deployed in unreachable, or remote spaces (oil platforms, mines, forests, tunnels, pipes, etc.) or in case of emergencies, i.e. earthquake, fire, floods, radiations areas, etc. In this infrastructure, these different entities, or things, discover and explore each other and learn to take advantage of each other's data by pooling of resources and significantly enhancing the scope and reliability of the resulting services" [3].

IoT is usually described by real world and small things having limited storage and computing capacity, and with limitations such as reliability, performance, security, and privacy. It usually consists of sensors and small devices inserted in physical objects, which help connect them to the internet.

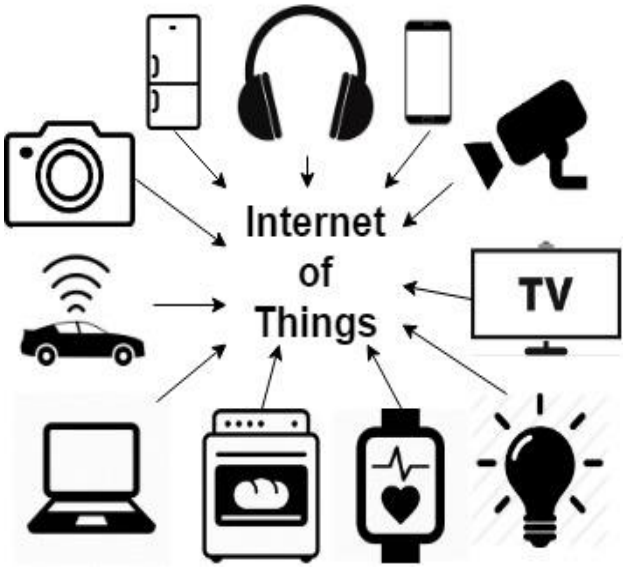

Fig. 1 Internet of Things

B. Cloud Computing

NIST described Cloud Computing as "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [2].

Cloud computing offers services which allow sharing of computing resources over the Internet. Cloud provides four types of services; in Software as a Service (SaaS) model the application is running on the Internet and operators can acquire the services on a pay-to-go basis; Platform as a Service (PaaS) model offers a platform in which one can create, test, and deploy specific software, with the resources available; the Infrastructure as a Service (IaaS) model offers structure for storage, hardware and servers. With the storage available in IaaS, the data is not only stored but it is also available globally through the Internet (e.g. Amazon Web Services); and finally, the Networks as a Service (NaaS) model, as many networks as required can be set-up [1]. It offers virtual networks. Figure 2. Represents the services and distinct type of clouds comprised by cloud computing.

Cloud Computing has three different deployment models; Public cloud, in which resources are accessible to users through the Internet. They are mostly held by a profitable organization; Private cloud is usually possessed by a single organization for specific purposes of its users. It provides a protected environment and more control; Hybrid Cloud is a combination of Private and Public clouds. It is offered to customers since it overcomes some of the limitations of each model.

Cloud computing can provide unlimited computing resources to users but they come with a cost [2]. The cost is affected by several factors such as, level of convenience of service, availability, scalability, elasticity, storage capacity and security, etc. Different cloud providers offer different pricing schemes built on coarse grained or fine-grained pricing schemes.

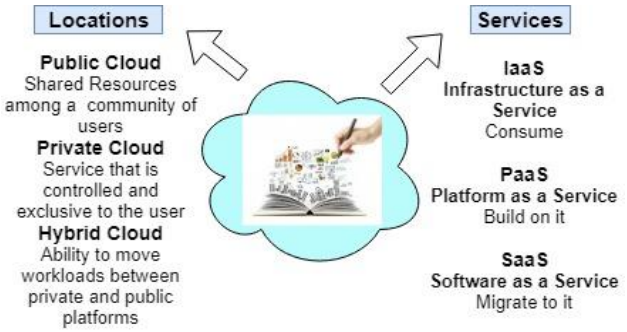

Fig. 2 Cloud Computing

C. Cloud-IoT

Cloud computing and Internet of Things (IoT) both are rapidly developing technologies. They both possess unique characteristics. However, they also have a few limitations. IoT is a chain of widely distributed devices connected through Internet, which have limited storage and processing abilities. Whereas, cloud computing offers unlimited storage and computing abilities. By integrating these two, we can overcome the limitations encountered by each of them and create a revolutionary technology. The Cloud-IoT paradigm offers smart usage of applications, data, and structure in an economical way. TABLE I provides the comparative characteristics of IoT and cloud computing.

Experts think that integrating them would result in a better platform, as their characteristics are complementary to each other [3].

 ISSN – 1999-4974

TABLE I CHARACTERISTICS OF CLOUD AND IOT.

Items	IoT	Cloud Computing
Characteristics	IoT is pervasive (things i.e real world objects are everywhere).	Cloud is ubiquitous (virtual resources are available everywhere).
Processing capabilities	Limited computational capabilities.	Virtually unlimited computational capabilities.
Storage capabilities	Limited storage or no storage capabilities.	Unlimited storage capabilities.
Connectivity	It uses internet as a point of convergence.	It uses internet for service delivery.
Big Data	It's a source of big data.	It's a means by which we can manage big data.

D. Big Data

The data produced nowadays by organizations are generating enormous amount of data. Simultaneously, vast amount of data is being gathered and attained from numerous resources and being stored.

Big Data can be defined as large complex datasets, which are hard to process using the existing database management tools [4][5]. Big Data comprises of three attributes; volume, variety, and velocity[6][7]. Experts believe that the Big Data will grow enormously with time, so we need to come up with solutions to handle the data issues that would be generated[8][9][10].

IV. ADVANTAGES, APPLICATIONS AND CHALLENGES OF CLOUD AND IOT INTEGRATION

A. Advantages:

The advantages of integrating IoT technology with Cloud Computing are discussed in the following TABLE II.

TABLE II ADVANTAGES OF CLOUD-IOT

Processing capabilities	Iot has limited processing capabilities whereas Cloud has unlimited. So processing the data generated by IoT devices is done with the help of Cloud. It also allows the analysis of the complex data.
Communication	IoT devices are IP-enabled and their communication with each other can hence be costly. So Cloud offers them cost-effective solution.
Storage capabilities	A huge amount of data is produced through the IoT devices and since IoT has limited storage, Cloud serves as a solution due to its attribute of unlimited storage.
Scope	The phenomena of IoT is leading to the Internet of Everything (IoE). The Cloud-IoT paradigm offers innovative applications and services, which are based on the growth of the Cloud via the IoT objects. It permits the Cloud to work with fresh real world scenarios.
New Abilities	IoT encompasses a variety of devices, technologies, and protocols which may be heterogeneous in nature. Hence, it can be tough to obtain security, availability, reliability, interoperability, scalability, and efficiency. Cloud not only resolves most of these problems but also provides further features like ease-of-access, ease-of-use, and reduced deployment costs.
New Models	The implementation of the Cloud-IoT paradigm allows new developments for smart services and applications: The implementation of the CloudIoT paradigm allows new developments for smart services and applications: SaaS, SAaaS, SEaaS, SenaaS etc.

B. Applications

The Applications of the Cloud-IoT are described through the Figure 3: [11]

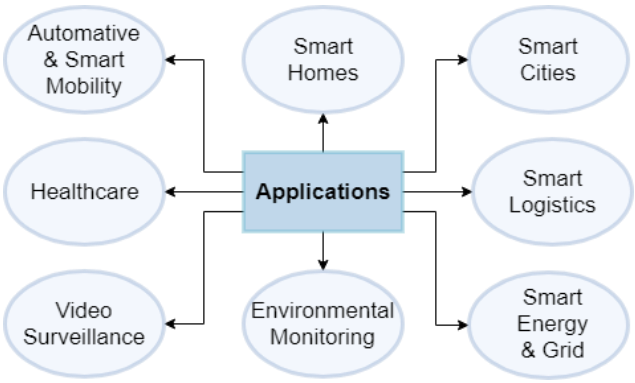

Fig. 3 Applications of Cloud IoT

C. Challenges

There are many challenges faced during the integration of these technologies. Some of the major problems are mentioned in the Figure 4 below:

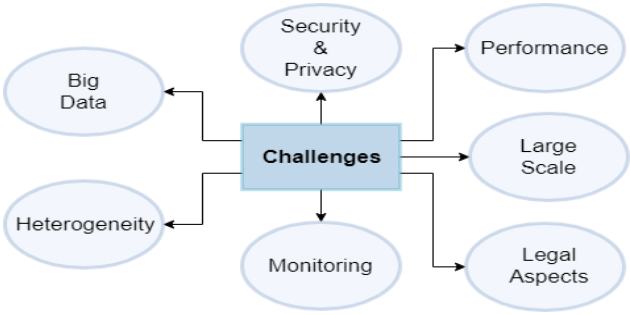

Fig. 4 Challenges of Cloud IoT

V. BIG DATA MANAGEMENT

Big Data as we know is characterized by its three main attributes; Volume, Velocity, and Variety. Volume refers to the size of data, velocity is the time between when the data was generated or gathered and the when it was available, and variety refers to the type of data i.e. structured or unstructured etc. [11] [12] [13]. Big data management is a method of dealing, managing and storing big data in such a form, which is comprehendible and provides easy access [14] [15]. Big data management is critical as it helps us make optimal and strategic decisions. The growth/progress of the underlying application or business depends on these decisions. The proposed solution to process this big data is explained as follows:

Two existing solutions for managing the big data are:

A. Hadoop

Hadoop is an open source solution [16]. It makes use of the distributed file system. It is executed on community hardware. The underlying architecture used by it is Hadoop Distributed File System (HDFS). This can be divided into

 ISSN – 1999-4974

three core parts; HDFS, Map Reduce, Hadoop common and Partition comparison module. HDFS connects all the files into a single huge file system. It tracks all the clusters.

Map Reduce [17] method can be broken down into two parts; map and reduce [18][19]. Map refers to chopping a task into smaller tasks i.e. sub problems. This way a complex problem is divided into smaller problems. These sub-problems are executed simultaneously on a separate server. The next process is reduce, which merges the outcomes of the sub-problems on different servers, this is also done in parallel. Map process is executed on the servers, on which the data resides. The outcomes of these processes is provided as input to the reduce process [20][21]. The overall architecture of Hadoop can be seen in figure 5.

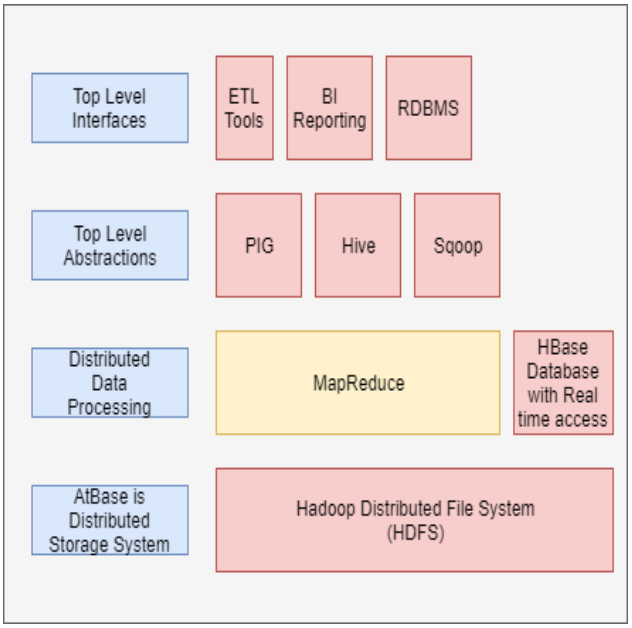

Fig. 5 Architecture of Hadoop

B. Cloud Services

The cloud offers cost-effective solutions for the management and retrieval of tremendous amounts of data. Cloud uses a number of virtual machines for the processing of huge amounts of data. The processing is done efficiently by dividing it into smaller parts and performing the function simultaneously on different machines quickly. Cloud is a potential solution for big data due to its flexible, powerful, and adaptive nature. Cloud makes use of the approach of sharing of computing resources. Some of the services offered by Cloud are:

a) **Iaas** (Infrastructure as a Service) in a public cloud: The formation of virtual machines with nearly unlimited storage and computing power could be delivered by IaaS.

b) **PaaS** (Platform as a Service) in a private cloud: In a public or private cloud environment it could allow to scheme, implement, and set up applications and services.

c) **SaaS** (Software as a Service) in a hybrid cloud: It offers a platform for analyzing the data.

d) **NaaS** (Network as a Service) is a plan of action for providing Wide Area Network(WAN), practically on a membership premise. Figure 6. shows distinct services provided by cloud computing.

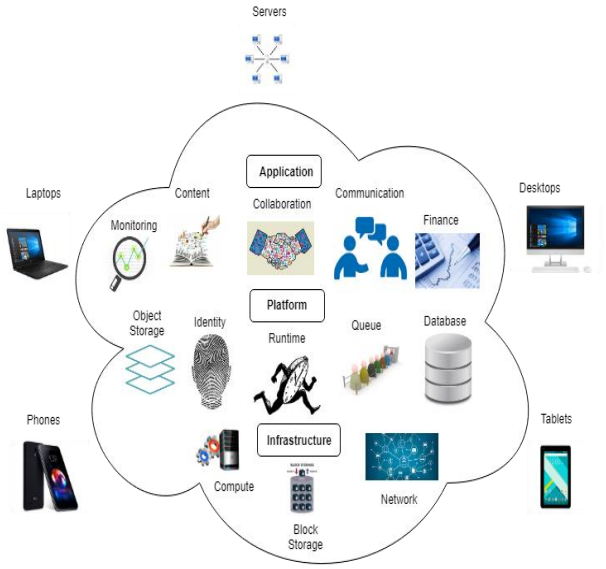

Fig. 6 Cloud Computing

C. Proposed Solution: A Hybrid Approach

The proposed solution to handle the big data produced by the IoT devices is a hybrid approach [22]. The Hybrid Solution uses both the Hadoop and the cloud together for the big data management [23] [24].

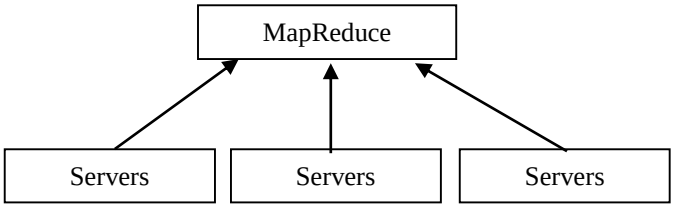

Fig. 7 Master-Slave Architecture

We can use the underlying architecture of the Hadoop. It makes use of Master-Slave approach, as shown in figure 7. To transfer and analyze the data using the MapReduce. The MapReduce requires an optimal environment where the Map process of dividing a task into smaller tasks and executing them to provide the reduce process with inputs, so that it can merge them up on the same servers. If they are done on the same servers then the process can be efficient, cost effective, and reliable and the retrieval of information can be done in real time.

ISSN – 1999-4974

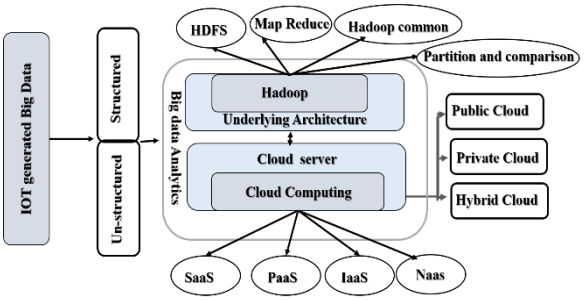

Fig. 8 Proposed solution

The services of Cloud can be used for this. The HDFS would perform its job of creating clusters and creating one big file system [21]. Then the Map Reduce process can be done on the virtual machines of the cloud. A complex task could be broken down into smaller simpler ones and can be executed in parallel on different machines and then the results can be integrated together.

The proposed solution as shown in figure 8. to the problem of Big Data Management could be of great value as Hadoop and Cloud services together can provide Flexibility, scalability, efficiency and reliability [25]. The core advantage of Hadoop is its fault tolerance. It does not pose any cost as it is open source, and cloud is a cost effective resource with unlimited storage and computing powers.

Cloud security comprises of a lot of strategies, controls, policies and technological means that cooperate to ensure cloud-based framework, it comprises preventive, deterrent, detective and corrective security controls. These security efforts are designed to secure data, raise regulative conformity and ensure clients' protection just as setting validation rules for each clients. From confirming access to traffic filtration, cloud security can monitor the demands of big data management. Hadoop configuration is based upon secure modes. Security highlights comprises of verification, Service Level Authentication, Data Confidentiality and Web Console Authentication. New Hadoop installments can be incredibly secure in terms of secure accession and data transmission scheme. The Hadoop system has assets to prop security measures. Ranger and Knox are two significant open source Apache projects. Knox gives a framework to overseeing security and supports security executions on Hadoop clusters. Ranger is centered on creating devices and methods to enable clients to convey and institutionalize security crosswise over Hadoop groups. It gives a centralized structure that can be utilized to monitor even at the resource level, for example, documents, databases, folders and even for particular columns or lines inside databases. It enables administrators to actualize group based or data type based access policies and so forth [26]. Hadoop's secure distributed file system and cloud's unlimited protective computation abilities could complement each other well. So the proposed hybrid scheme will be extensively helpful against secure big data management [27] [28] [29].

One limitation of this solution would be that Hadoop does not possess any kind of abstraction and it requires the MapReduce engineers to hand code for every single procedure.

VI. CONCLUSION

The IoT devices lack storage and processing capabilities, cloud can help in overcoming these issues. They can be of great application when used together. However, there are few challenges faced when integrating them such as security and privacy, heterogeneity, big data, monitoring, performance and legal aspects. A solution was proposed for the handling of the big data. Big data management is significant for organizations as it assists them to organize the data and use it for future purposes. Decisions and strategies are made after analyzing the big data, so it is critical to the people.

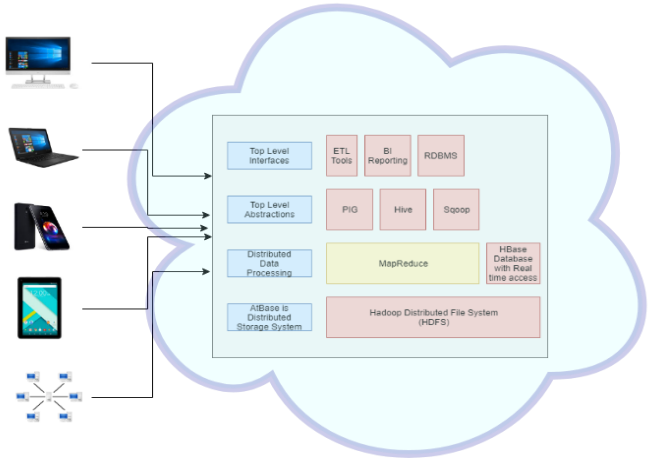

Fig. 9 Big data Management

The Solution proposed for handling the Big Data was as Figure 9. shows cloud with Hadoop Distributed File System (HDFS) and MapReduce. Hadoop is open source and hence does not pose any cost. It offers scalability. Cloud offers unlimited storage and processing capabilities. Hence, when both of these used altogether would be of great benefit.

REFERENCES

[1] S. M. Babu, A. Lakshmi and B. Rao, "A Study on Cloud based Internet of Things: CloudIoT," in Proceedings of 2015 Global Conference on Communication Technologies(GCCT 2015), 2015.

[2] L. L. Dhirani, D. T. Newe, p. E. Lewis and D. S. Nizamani, "Cloud Computing and Internet of Things Fusion: Cost Issues," in 2017 Eleventh International Conference on Sensing Technology (ICST), 2017.

[3] M. Aazam, I. Khan, . A. A. Alsaffar and . E.-N. Huh, "Cloud of Things: Integrating Internet of Things and Cloud Computing and the Issues Involved," in Proceedings of 2014 11th International Bhurban Conference on Applied Sciences & Technology (IBCAST) Islamabad, Pakistan, 14th – 18th January , 2014.

[4] I. L. a. K. Lee, "'The Internet of Things (IoT): Applications , investments , and challenges for enterprises," in Bus. Horiz., vol. 58, no. 4.

 ISSN – 1999-4974

[5] D. Gil, M. Johnsson, H. Mora, J. Szyma Nski " Review of the Complexity of Managing Big Data of the Internet of Things, Hindawi Complexity Volume 2019, Article ID 4592902, 12 pages https://doi.org/10.1155/2019/4592902, 2019.

[6] S. Kanchi, S. Sandilya, S. RamKrishna, S. Manjrekar and A. Vhadgar, "Challenges and Solutions in Big data management - An Overview," in 2015 3rd International Conference on Future Internet of Things and Cloud, Mumbai, 2015.

[7] H. F. Atlam, A. Alenezi, A. Alharthi, R. J. Walters and G. B. Wills, " Integration of Cloud Computing with Internet of Things: Challenges and Open Issues.," in IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom), 2017.

[8] M. C. N. H. W. Kyounghyun Park, "Web-based collaborative big data analytics on big data as a service platform," in 2015 17th International Conference on Advanced Communication Technology (ICACT), Seoul, 2015.

[9] K. K. S. K. J. P. HyeJung Moon, "Ecosystem Design of Big Data through Previous Study Analysis in the World: Policy Design for Big Data as Public Goods," in 2017 IEEE International Congress on Big Data (BigData Congress), Honolulu, 2017.

[10] Z. X. Z. J. X. J. YANG Mengke, "Challenges and Solutions of Information Security Issues in the Age of Big Data," in Security Schemes and Solutions, Beijing, 2016.

[11] S. G. A. G. Ranjit Kumar Behera, "Big-Data Empowered Cloud Centric Internet of Things," in 2015 International Conference on Man and Machine Interfacing (MAMI), Berhampur, 2015.

[12] F. A. J. Alberto Espinosa, "The Big Data Analytics Gold Rush: A Research Framework for Coordination and Governance," in 2016 49th Hawaii International Conference on System Sciences, Hawaii, 2016.

[13] Justin McHugh, P. E. (2017). Integrated access to big data polystores through a knowledge-driven framework. 2017 IEEE International Conference on Big Data (Big Data). Boston, MA, USA.

[14] Jun-Jie Zhang, L. Y. (2017). A research on enterprise crisis management innovation based on big data technology. 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA). Beijing, China.

[15] Mazumdar, S., & Dhar, S. (2015). Hadoop as Big Data Operating System -- The Emerging Approach for Managing Challenges of Enterprise Big Data Platform. 2015 IEEE First International Conference on Big Data Computing Service and Applications. Redwood City, CA, USA.

[16] Kim, J., & Park, N. (2015). Identification of the Optimal Hadoop Configuration Parameters Set for Mapreduce Computing. 2015 3rd International Conference on Applied Computing and Information Technology/2nd International Conference on Computational Science and Intelligence. Okayama, Japan.

[17] Mohammed, A. Q., & Bharati, R. (2017). An efficient technique to improve resources utilization for hadoop MapReduce in heterogeneous system. 2017 International Conference on Intelligent Communication and Computational Techniques (ICCT). Jaipur, India.

[18] Verma, C., & Pandey, R. (2016). Big Data representation for grade analysis through Hadoop framework. 2016 6th International Conference - Cloud System and Big Data Engineering (Confluence). Noida, India.

[19] Bante, P. M., & Rajeswari, K. (2017). Big Data Analytics Using Hadoop Map Reduce Framework and Data Migration Process. 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA). Pune, India.

[20] Bhardwaj, A., Singh, V. K., Vanraj, & Narayan, Y. (2015). Analyzing BigData with Hadoop cluster in HDInsight azure Cloud. 2015 Annual IEEE India Conference (INDICON). New Delhi, India.

[21] Singh, S., Raj, G., & Kaur, G. (2016). Analysis of HDFS RPC and Hadoop with RDMA by evaluating write performance. 2016 6th International Conference - Cloud System and Big Data Engineering (Confluence). Noida, India.

[22] Lee, Y.-J., Lee, M., Lee, M.-Y., Hur, S. J., & Min, O. (2015). Design of a scalable data stream channel for big data processing. 2015 17th International Conference on Advanced Communication Technology (ICACT). Seoul, South Korea.

[23] Mallika, C., & Selvamuthukumaran, S. (2017). Hadoop framework: Analyzes workload predicition of data from cloud computing. 2017 International Conference on IoT and Application (ICIOT). Nagapattinam, India.

[24] B. Jadhav, Archana B. Patankar, Opportunities and Challenges in integrating Cloud Computing and Big Data Analytics to E-governance. International Journal of Computer Applications (0975 – 8887 Volume 180 – No.15, January 2018.

[25] Taleb, I., & Serhani, M. A. (2017). Big Data Pre-Processing: Closing the Data Quality Enforcement Loop. 2017 IEEE International Congress on Big Data (BigData Congress). Honolulu, HI, USA.

[26] Introduction to Hadoop Security. (2019, June 26). Retrieved from bmc Blog: https://www.bmc.com/blogs/hadoop-security/.

[27] Tardío, R., Mate, A., & Trujillo, J. (2015). An iterative methodology for big data management, analysis and visualization. 2015 IEEE International Conference on Big Data (Big Data). Santa Clara, CA, USA.

[28] R. Dasoriya, "A Review of Big Data Analytics over Cloud " in Proceedings 2017 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia), 2017 IEEE.

[29] S. Banaeian Far, A. Imani Rad " Security Analysis of Big Data on Internet of Things" arXiv:1808.09491v1 [cs.CR], 28 Aug 2018.

 ISSN – 1999-4974