A Review Paper on Big Data, Hadoop and Cloud Computing

Pankaj Saraswat

SOEIT, Sanskriti University, Mathura, Uttar Pradesh, India

Email Id- pankajsaraswat.cse@sanskriti.edu.in

ABSTRACT: *The concepts of Big Data and Hadoop are used to handle enormous amounts of data. Big Data is described as a vast volume of data, both organised and unstructured, that cannot be stored or processed using traditional data storage methods. Hadoop, on the other hand, is a technology for dealing with large amounts of data. The Apache Software Foundation has created an open-source framework. Cloud computing may be described as the network connections. BIG DATA has a nice plan as well as some drawbacks when it comes to dealing with big amounts of data. Nowadays, BIG DATA is more essential than ever before since everyone is keeping the data on the cloud and it is automatically saved. This is beneficial to organisations or any other major businesses that have many data. To save that, everyone wants the information to be safe and secure. As a result, this good technology has developed at a rapid pace. In addition, if you are looking for knowledge on a certain topic, you may search for it in that vast database. Because the data is both organised and unstructured, the findings include texts, videos, photos, captions, social media postings, and much more. Big Data, Hadoop, and cloud computing are the discussed in this paper.*

KEYWORDS: *Big Data, Cloud Computing, Data, Hadoop, Software.*

1. INTRODUCTION

There was no way to interact with anyone before the internet. The only method to go from one person to another is to write letters, which take a long time to arrive. However, now that everyone has access to the internet, we can read whatever we want and search for anything we want on our phones, computers, and other devices. Life has gotten easier because of the internet. The terms "big data," "Hadoop," and "cloud computing" spring to mind. Big data refers to a huge volume of data, or the majority of information, that is kept on the cloud. To both conserve and safeguard our information[1]. If anybody wishes to keep the information hidden or private, there is security. Traditional data, which may be classed as producing tables and databases, is referred to as structured and unstructured in the following paragraphs. Structured data is data that is in an executable format, whereas unstructured data is data that is not in an executable format and is found in places like social media, emails, and webpages[2].

Big Data and Hadoop are two technologies for dealing with enormous amounts of data. Data is made up of organised and unstructured data that cannot be stored or processed using traditional methods. Hadoop, on the other hand, is a large data management technology. This is now the most in-demand big data tool[3].

1.1 Characteristic of Big data are:

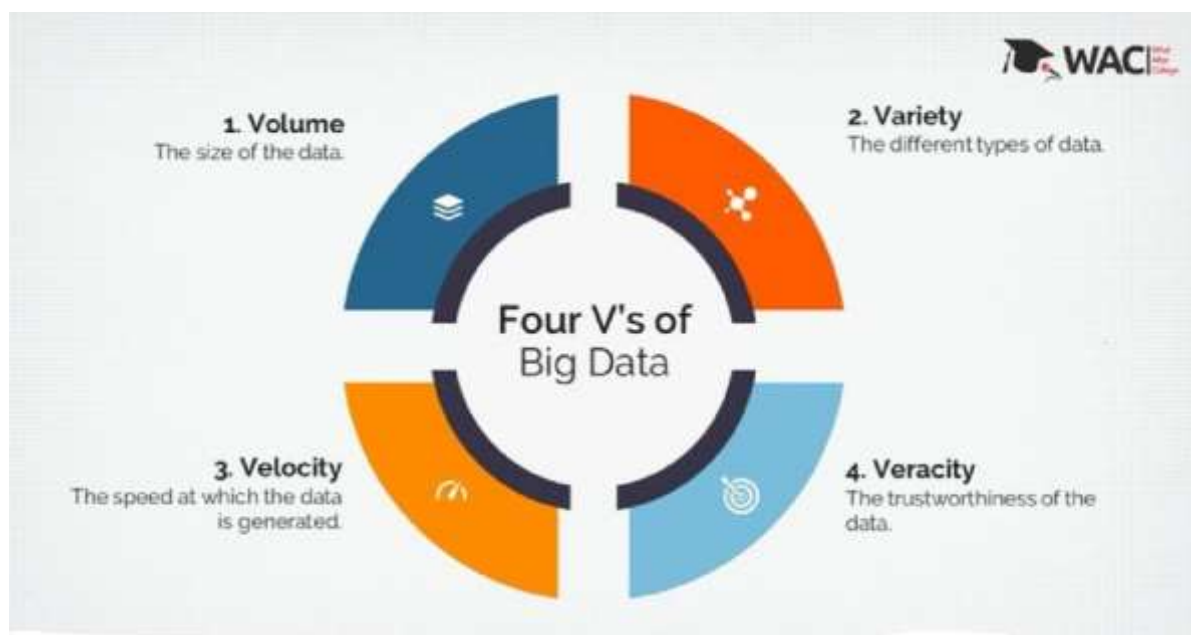Figure 1 shows the 4V characteristics of Big Data.

Figure 1: The above diagram shows the four V's characteristics of Big Data.

Big data may be defining the data sets. The size of big data captures different size and complexity. Which do the conventional technologies and tools study, as it collects the large amount of data in relational databases. Big data and Hadoop there are two main functionalities domain and slave[4]. Domain is also known as Masternode and slave is known as Namenode. In this slave are the storage space where the data is being stored and the masternode is that which have all the information about the slaves, as slave nodes are connected with the master node. Therefore, master node have all the information about the slave node that when is been excessed what is its use and what are the files stored in this so this is the work or master node and slave node[5] shown in Figure 2.
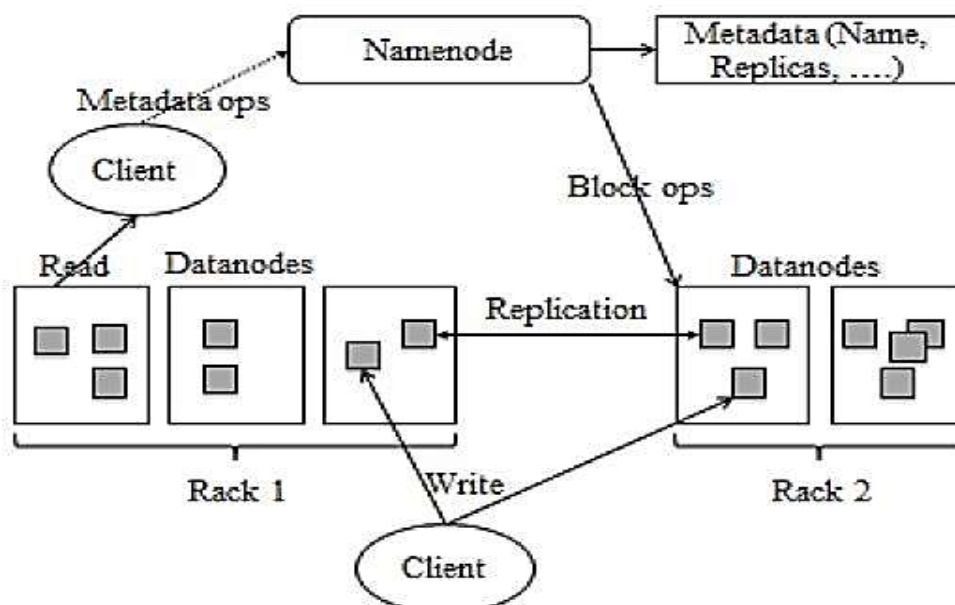

Figure 2: The above diagram shows the functionality of big data and hadoop is Datanode and Namenode.

Hadoop is a software, which is defined as an open source, which is used for framework to store the data. Hadoop is important as well because it has the ability to store the large amount of data. It is flexible, low cost, scalable etc.

1.2 Characteristics of Hadoop:

Figure 3 shows the characteristics of Hadoop which provides the reliable sharing storage and which is highly scalable and cost effective[6].
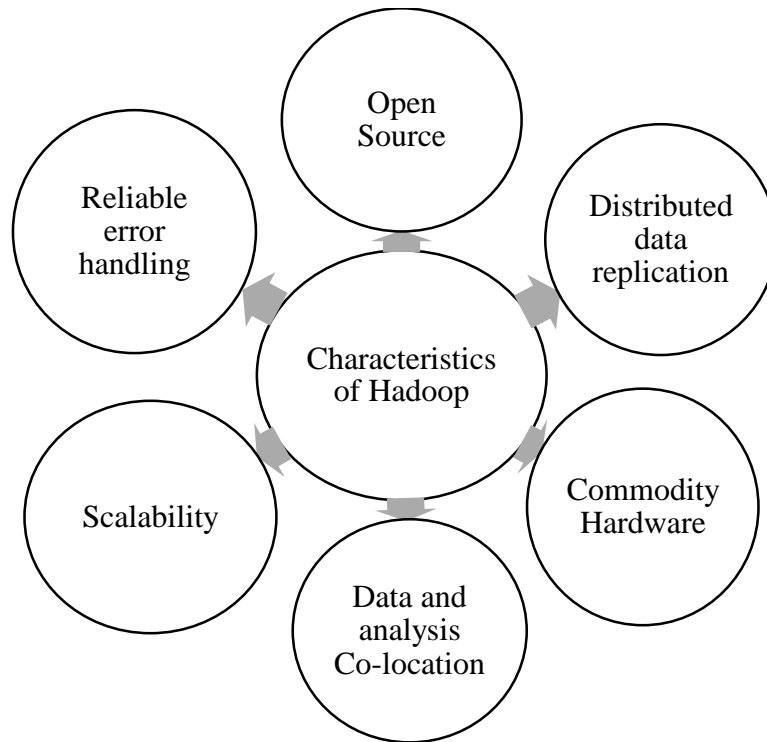

Figure 3: The above diagram shows the characteristic of the Hadoop which store the large amount of data.

Cloud refers to the network or internet. It is interconnected across the world. Cloud computing may be defined as to control the data, configure the data, and accessing the software and hardware resources. In this, data's are store online[7]. Mostly everyone nowadays are uploading the data's on interest so that it could be safe and can further be accessible in future in Figure 4.
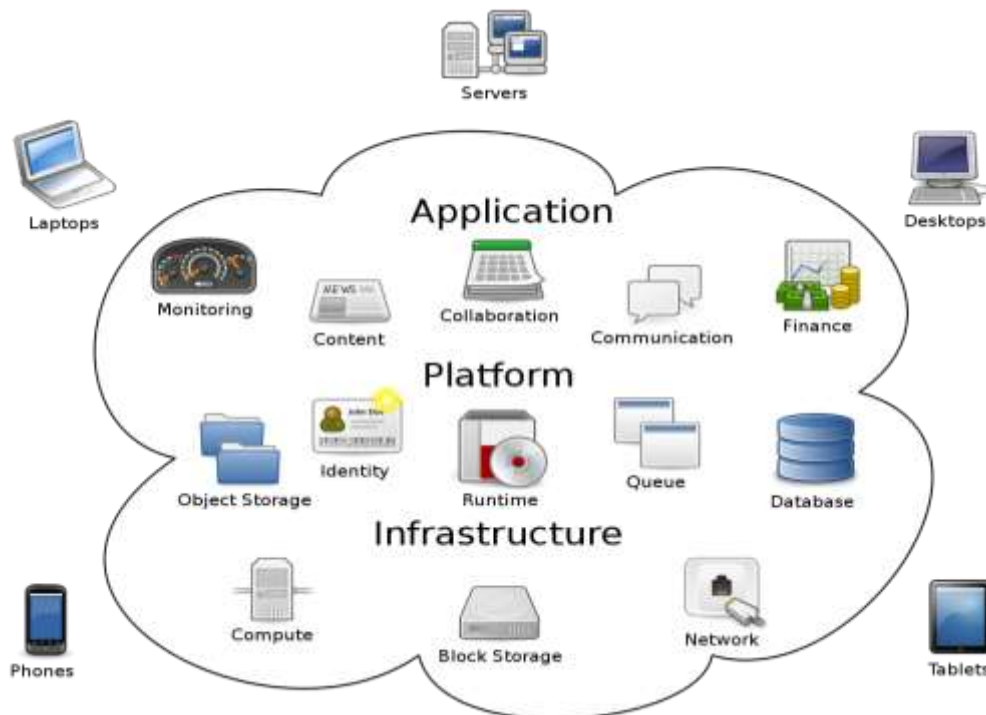

Figure 4: This diagram shows the cloud computing and the group of connected elements which are providing services and need to address one by one or managed by users.

Types of cloud shown in Figure 5. They are mainly of four types of cloud Private, Public, Hybrid and Community cloud[8].
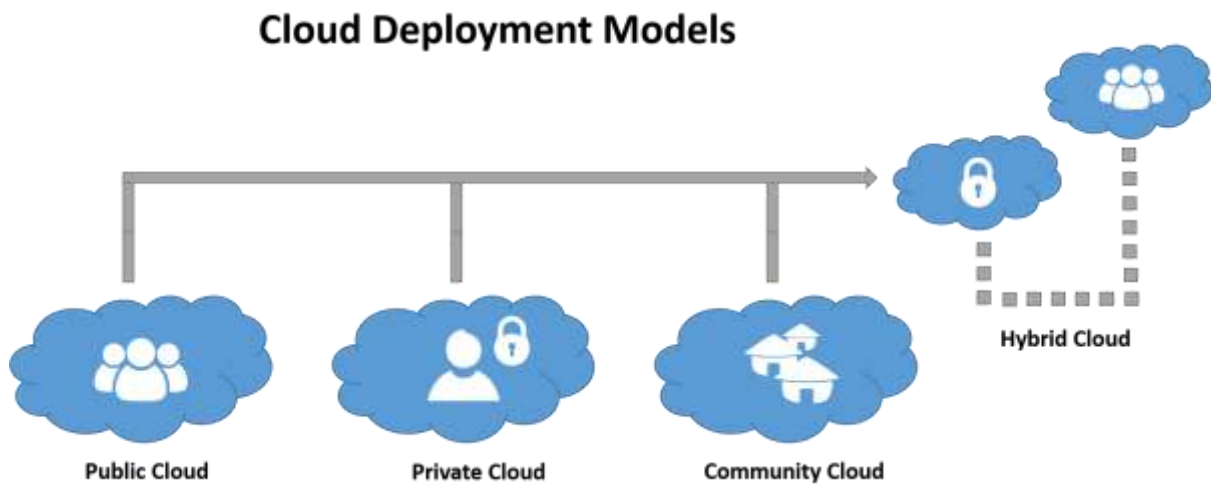
Figure 5: The above diagram shown the types of cloud.

Service model of cloud shown in Figure 6. They are mainly of three types, Infrastructure as a service(IAAS), Software as a service(SAAS), and Platform as a service(PAAS).

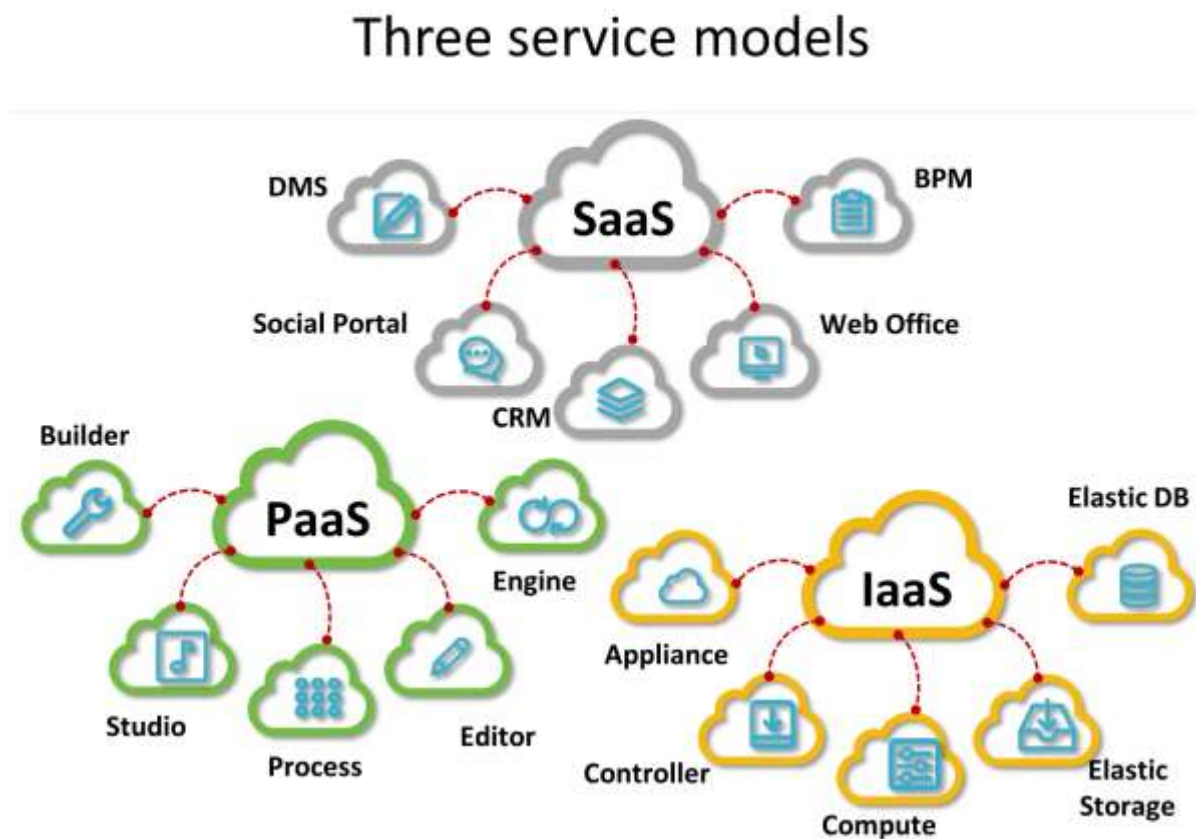**Figure 6: The above diagram shows the cloud services – PAAS, SAAS, IAAS.**

1.3 Characteristics of CC (Cloud Computing) shown in figure 7:

- Measured Service
- On demand self-services
- Resource polling
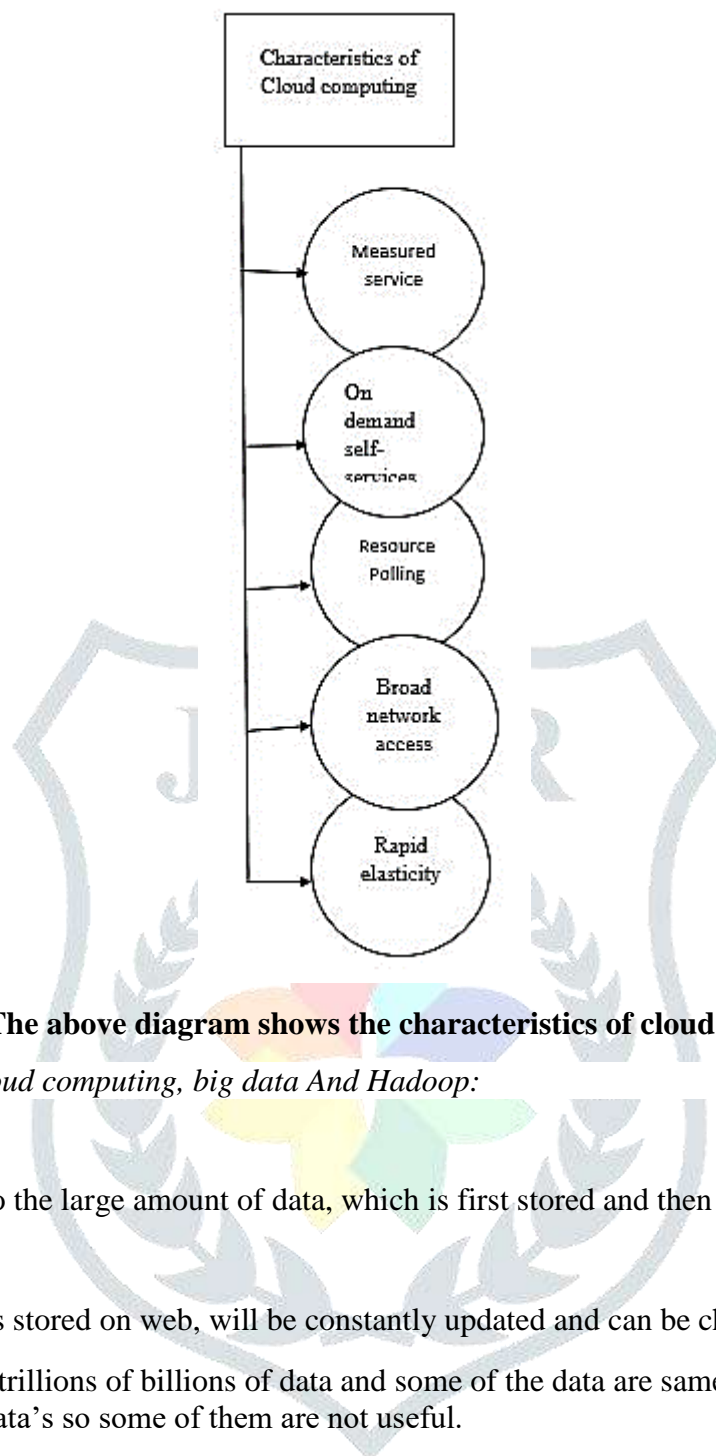- Board network access
- Rapid elasticity

Figure 7: The above diagram shows the characteristics of cloud computing.

1.4 Challenges of Cloud computing, big data And Hadoop:

1.4.1 Big Data:

- Here big relates to the large amount of data, which is first stored and then analysed.

- This is expensive.

- The data, which is stored on web, will be constantly updated and can be changed anytime.

- On web there are trillions of billions of data and some of the data are same as well so there is duplicate of the data's so some of them are not useful.

- There are not proper notes with proper information.

1.4.2 Hadoop:

- It is difficult to find the root of any problem.

- It is a low-level API.

- It can understand the multiple technologies, which is used for coding.

- It is real time operating system.

1.4.3 Cloud Computing:

- The main disadvantage of cloud computing is security.

- Many companies are facing lack of resources.

- To build the private cloud for business to hide the details of the organisation.

1.5 Advantages of Cloud Computing, Big Data And Hadoop and:

1.5.1   Big Data:

- Anyone can explore the data's present on the cloud.

- As everybody knows that every company are storing their data in online mode means on cloud or and other platform so they need to take care of their data's.

- It is cost optimised.

1.5.2   Hadoop:

- It is cost effective, flexible as well as scalable.

- The tools for processing the data can result much faster

1.5.3   Cloud Computing:

- The services mainly runs on the worldwide network.

- Makes data backup as well.

- This increases the work rate of or any tasks.

Big data, in general, is an extremely large file that may include trillions of billions of bytes of data or trillions of bytes of data in the form of a dataset. As a result, everyone asks, "What is that billions of data?" The answer is that everyone stores photographs or videos[9] from any site or social media account, so where did these photos come from and how were they uploaded. Every photo of yours or anybody else's, every post, is kept on the cloud, and since we have such a large population, everyone has something to store; therefore, all of that data is stored on the cloud. Large-capacity data can be found in a variety of places. Data volumes will rise in the future and will be transferred to the cloud, as data quantities will grow in the future[10].

2.   DISCUSSION

There was no way to interact with anyone before the internet. The only method to go from one person to another is to write letters, which take a long time to arrive. Everyone now has access to the internet, which allows them to read whatever they want and do searches on their phones, computers, and other devices. Life has gotten easier because of the internet. Big data and Hadoop, as well as cloud computing, are two current technologies that the author has covered. Nowadays, everyone is considerably more familiar with these languages and some of these sites. Because everyone is using them to store their data, they are in short supply. The author has addressed cloud computing, big data, and Hadoop in this article. The four V's of big data, big data problems, Hadoop, and cloud computing When you have a lot of data, everyone prefers this since these platforms are safe if you put them to private mode, and your papers are stored. You can also have an overabundance of them whenever you desire. Hadoop is an open-source framework, and cloud computing is the internet that links the entire globe. Big data refers to enormous amounts of data, and hadoop is an open-source framework. We have also spoken about the benefits and varieties of cloud, as well as their service model. All of the technologies' characteristics (big data, hadoop, cloud computing).

3.   CONCLUSION

The author has covered hadoop, big data, and cloud computing since they are extremely significant nowadays because it is quite difficult to upload data or research something about anything. All updates, whether organised or unstructured, are available there. Everything is connected to everything else. Nowadays, everyone has many data, and in order to keep it and avoid taking any risks, they all want to transfer it to the cloud. The author of this review article covered big data, which is defined as having a huge quantity of data saved on the cloud, hadoop, which is an open source framework, and cloud computing, which is defined as connecting to the rest of the world. What are the V's of big data, and what are the problems and benefits of cloud computing, big data, and Hadoop? Data volumes will continue to grow in the future, and more data will be transferred to the cloud. Data volumes will rise in the future and will be transferred to the cloud, as data quantities will grow in the future.

REFERENCES

[1]   R. T. Pvt, "Big Data and Hadoop Developer," pp. 219–226, 2015.

[2]   G. Vemuganti, "Metadata Management in Big Data," *Infosys Labs Briefings*, vol. 11, no. 1, p. 87, 2013.

[3]   S. D. Vidyasagar, "Engineering A Study on ' Role of Hadoop in Information Technology era ' ABSTRACT III Sem MCA , JNN Engineering College , Shimoga-577201 , Karnataka," no. 2277, pp. 100–101, 2013.

[4]   H. S. Bhosale and D. P. Gadekar, "A Review Paper on Big Data and Hadoop," *Int. J. Sci. Res. Publ.*, vol. 4, no. 10, pp. 1–7, 2014.

[5]   S. Shilpashree, R. R. Patil, and C. Parvathi, "'Cloud computing an overview,'" *Int. J. Eng. Technol.*, 2018, doi: 10.14419/ijet.v7i4.10904.

[6]   W. T. Wu, W. W. Lin, C. H. Hsu, and L. G. He, "Energy-efficient hadoop for big data analytics and computing: A systematic review and research insights," *Futur. Gener. Comput. Syst.*, 2018, doi: 10.1016/j.future.2017.11.010.

[7]   S. Abdul, "An overview on Big Data and Hadoop," *Int. J. Comput. Appl.*, 2016, doi: 10.5120/ijca2016912241.

[8]   S. Landset, T. M. Khoshgoftaar, A. N. Richter, and T. Hasanin, "A survey of open source tools for machine learning with big data in the Hadoop ecosystem," *J. Big Data*, 2015, doi: 10.1186/s40537-015-0032-1.

[9]   B. Saraladevi, N. Pazhaniraja, P. V. Paul, M. S. S. Basha, and P. Dhavachelvan, "Big data and Hadoop-A study in security perspective," 2015, doi: 10.1016/j.procs.2015.04.091.

[10]   S. Oliviandi, A. B. Osmond, and R. Latuconsina, "Implementasi Apache Spark Pada Big Data Berbasis Hadoop Distributed File System," *e-Proceeding Eng.*, 2018.