Big Data in Cloud Computing

Hardik Singh, Hardik Jain, Bhavya Vora, Mayank Pujara, Shreeja Ravikumar[a]

[a]K.J Somaiya College Of Engineering Vidyanagar, Vidya Vihar East, Vidyavihar,, Mumbai, 400077, Maharashtra, India

Abstract

Big Data and Cloud Computing are two of the most important technologies of the day. Since data is being generated exponentially every day, Big Data has gained a lot of significance in any technology. The daily explosion of data means that it's better to have big data included in the applications. Whereas cloud computing is allowing users to use platforms according to their time, convenience and affordability. It is providing users ability to collaborate and work efficiently more than ever. Combining these two technologies can give a hands down advantage to the users in terms of knowledge and efficiency. Big Data, when used in cloud computing has applications in different fields such as Finance, Management, supply chain, planning, data storage, warehouses and many more. In this paper, we have discussed Big Data implementation and application in Cloud Computing. 4 V's in big data can be applied in Cloud computing to get better performance, higher input details, better insights, reliable and secure platforms at comparatively lower costs. Different analytics, technology involved in coupling of big data with cloud computing, the challenges involved in this process, trends applications of the domain and security factors involved are discussed in this paper.

Keywords: Big Data, Cloud Computing,

Contents

1. INTRODUCTION

Cloud is a server collection that is distributed across the internet to store, process and manage the data. Cloud computing provides the hardware and software services through internet. It enables big data to manage and allocate the stored data in proper manner[9]. Cloud computing provide the security to big data with hadoop, environment like file and network, encryption, logging, nods authentication, layered frame work for assuring cloud etc. Different types of models are using for mining the big data. Cloud computing approach is valuable as it has the adopted technologies to handle the large amount of data. It provides the interconnectivity between the devices and data that is further assist to exchange the data and connected to other devices. In 2014, the connected devices were 3.7 billions and it will reach at estimated 25 billions till 2020. Big data term is basically used to convert the data into information. The major purpose of big data is to store and manage, visualize and analyze huge amount of data per day.

2. CLOUD COMPUTING

Cloud computing delivers computing services such as servers, storage, databases, networking, software, analytics and intelligence over the internet for faster innovation, flexible resources, heavy computation, parallel data processing and economies of scale. It empowers the organizations to concentrate on core business by completely abstracting computation, storage and network resources to workloads as needed and tap into an abundance of prebuilt services.

The array of available cloud computing services is vast, but most fall into one of the following categories:

SaaS: Software as a Service Software as a service represents the largest cloud market and most commonly used business option in cloud services. SaaS delivers applications to the users over the internat. Applications that are delivered through SaaS are maintained by third-party vendors and interfaces are accessed by the client through the browser. Since most of SaaS applications run directly from a browser, it eliminates the need for the client to download or install any software. In SaaS vendor manages applications, runtime, data, middleware, OS, virtualization, servers, storage and networking which makes it easy for enterprises to streamline their maintenance and support.

PaaS: Platform as a Service Platform as a Service model provides hardware and software tools over the internet which are used by developers to build customized applications. PaaS makes the development, testing and deployment of applications quick, simple and cost-effective. This model allows business to design and create applications that are integrated into PaaS software components while the enterprise operations or thirty-party providers manage OS, virtualization, servers, storages, networking and the PaaS software itself. These applications are scalable and highly available since they have cloud characteristics.

IaaS: Infrastructure as a Service Infrastructure as a Service cloud computing model provides self-servicing platform for accessing, monitoring and managing remote data center infrastructures such as compute, storage and networking services to organizations through virtualization technology. IaaS users are responsible for managing applications, data, runtime, middleware, and OS while providers still manage virtualization, servers, hard drives, storage, and networking. IaaS provides same capabilities as data centers without having to maintain them physically.[1]

3. BIG DATA

Big data sports by the new technologies and architecture that is basically used for data capture, storage and analysis. Big data is implemented in various ways like email, mobile device output, sensor generated data and social media output. Big data has the need of large storage capacity. Hence, huge data could not be stored at local data base system. Another point is that data is in the form of different structure and use for different source. Big data is the concept of information and communication technology. It is put up the contribution to solve the problem of data warehouses and data mining. Here data mining is a big issue to handle the big database of data warehouses. According to Gartner, " Big data are highvolume, high velocity and high variety information assets that required new forms of processing to enable enhanced decision making, insight discovery and process optimization". Big data signify six terms- volume, variety, velocity, veracity, value and complexity.

a) **Volume**-it represents to the expanding of data beyond terabytes like transcation data, sensor data.

b) **Variety**-it refers to the collection of data from various different sources like machine, sensors etc. for example e-mails, audio-visuals, text document etc.

c) **Velocity**-How data is processed at fast speed. How fast data is accessed. Moreover IOT emerge the new type of data that is collection of sensor data and the control of equators. Big data is implemented in the form of different applications in the smart cities.

d) **Veracity**- it means collected data as different qualities with different accuracy, coverage and timeliness must be compliance.

e) **Value**-Big data is provided after the processing and analyzing data. Further store data can be used for further uses with the combination of other data sets.

f) **Complexity**- it manage the complexity of multiple sources data is linked, matches, cleansed and transformed before delivered [2]

4. RELATIONSHIP BETWEEN CLOUD COMPUTING AND BIG DATA

Storing and processing big volumes of data requires scalability, fault tolerance and availability. Cloud computing delivers all these through hardware virtualization. Thus, big data and cloud computing are two compatible concepts as cloud enables big data to be available, scalable and fault tolerant. Business regard big data as a valuable business opportunity. As such, several new companies such as Cloudera, Hortonworks, Teradata and many others, have started to focus on delivering Big Data as a Service (BDaaS) or DataBase as a Service (DBaaS). Companies such as Google, IBM, Amazon and Microsoft also provide ways for consumers to consume big data on demand.[3]

5. EXAMPLES SUCCESSFUL USE OF BIG DATA WITHIN CLOUD ENVIRONMENT

5.1. Nokia

Nokia was one of the first companies to understand the advantage of big data in cloud environments (Cloudera, 2012). Several years ago, the company used individual DBMSs to accommodate each application requirement. However, realizing the advantages of integrating data into one application, the company decided to migrate to Hadoop-based systems, integrating data within the same domain, leveraging the use of analytics algorithms to get proper insights over its clients. As Hadoop uses commodity hardware, the cost per terabyte of storage was cheaper than a traditional RDBMS (Cloudera, 2012). Since Cloudera Distributed Hadoop (CDH) bundles the most popular open source projects in the Apache Hadoop stack into a single, integrated package, with stable and reliable releases, it embodies a great opportunity for implementing Hadoop infrastructures and transferring IT and technical concerns onto the vendors' specialized teams. Nokia regarded Big Data as a Service (BDaaS) as an advantage and trusted Cloudera to deploy a Hadoop environment that copes with its requirements in a short time frame. Hadoop, and in particular CDH, strongly helped Nokia to fulfil their needs (Cloudera, 2012).[4]

5.2. RedBus

Redbus is the largest company in India specialized in online bus ticket and hotel booking. This company wanted to implement a powerful data analysis tool to gain insights over its bus booking service (Kumar, 2006). Its datasets could easily stretch up to 2 terabytes in size. The application would have to be able to analyse booking and inventory data across hundreds of bus operators serving more than 10.000 routes. Furthermore, the company needed to avoid setting up and maintaining a complex in-house infrastructure.

At first, RedBus considered implementing inhouse clusters of Hadoop servers to process data.

However they soon realized it would take too much time to set up such a solution and that it would require specialized IT teams to maintain such infrastructure. The company then regarded Google bigQuery as the perfect match for their needs, allowing them to:

- Know how many times consumers tried to find an available seat but were unable to do it due bus overload;
- Examine decreases in bookings;
- Quickly identify server problems by analysing data related to server activity;

Moving towards big data brought RedBus business advantages. Google bigQuery armed RedBus with real-time data analysis capabilities at 20

5.3. Tweet Mining

Noordhuis et al. used cloud computing to gather and analyse tweets. Amazon cloud infrastructure was used to perform all the computations. Tweets were crawled and later page ranking algorithm was applied. Page Ranking is used by Google to define the importance of a web page. On an overview, in Page Ranking Algorithm, if an author of a Web page A links to another Web page B, then author of Page A is likely to find Page B important. This way, the number of in-edges determine the importance of a certain page. Hence, more the number of inlinks to a page, the more its mportance. The same concept is applied to tweets, where instead of a web page, a tweet user following the other tweet is considered important. We need to crawl the Twitter social graph to compute PageRank. The data crawled had nearly 50 million nodes and 1.8 billion edges which is about two-thirds of the Twitter's estimated user base. [5]

6. MANAGING BIG DATA IN CLOUD COMPUTING ENVIRONMENT

Cloud Computing is an environment based on using and providing services. There are different categories in which the service-oriented systems can be clustered. One of the most used criteria to group these systems is the abstraction level that is offered to the system user. In this way, three different levels are often distinguished: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS)

Cloud Computing offers scalability with respect to the use of resources, low administration effort, flexibility in the pricing model and mobility for the software user. Under these assumptions, it is obvious that the Cloud Computing paradigm benefits large projects, such as the ones related with Big Data and BI . Focusing on the structure of the data management sector we may define, as the most suitable management organization architecture, one based on a four-layer architecture, which includes the following components :

1. A file system for the storage of Big Data, i.e., a wide amount of archives of large size. This layer is implemented within the IaaS level as it defines the basic architecture organization for the remaining tiers.

2. A DBMS for organizing the data and access them in an efficient way. It can be viewed in between the IaaS and PaaS as it shares common characteristics from both schemes. Developers used it to access the data, but its implementation lies on a hardware level. Indeed, a PaaS acts as an interface where, at the upper side offers its functionality, and at the bottom side, it has the implementation for a particular IaaS. This feature allows applications to be deployed on different IaaS without rewriting them.

3. An execution tool to distribute the computational load among the computers of the cloud. This layer is clearly related with PaaS, as it is kind of a 'software API' for the codification of the Big Data and BI applications.

4. A query system for the knowledge and information extraction required by the system's users, which is in between the PaaS and SaaS layers. [6]

7. DATA PROCESSING AND RESOURCE MANAGEMENT

7.1. Hadoop

Hadoop, is a free Java-based programming framework supports the processing of large sets of data in a distributed computing environment. It is a part of the Apache project sponsored by the Apache Software Foundation. Hadoop cluster uses a Master/Slave structure. Using Hadoop, large data sets can be processed across a cluster of servers and applications can be run on systems with thousands of nodes involving thousands of terabytes. Distributed file system in Hadoop helps in rapid data transfer rates and allows the system to continue its normal operation even in the case of some node failures. This approach lowers the risk of an entire system failure, even in the case of a significant number of node failures. Hadoop enables a computing solution that is scalable, cost effective, and flexible and fault tolerant. Hadoop Framework is used by popular companies like Google, Yahoo, Amazon and IBM etc., to support their applications involving huge amounts of data. Hadoop has two main sub projects – Map Reduce and Hadoop Distributed File System (HDFS) .

7.2. Map Reduce

Hadoop Map Reduce is a framework used to write applications that process large amounts of data in parallel on clusters of commodity hardware resources in a reliable, fault-tolerant manner. A Map Reduce job first divides the data into individual chunks which are processed by Map jobs in parallel. The outputs of the maps sorted by the framework are then input to the reduce tasks. Generally the input and the output of the job are both stored in a file-system. Scheduling, Monitoring and re-executing failed tasks are taken care by the framework .

7.3. Hadoop Distributed File System (HDFS)

HDFS is a file system that spans all the nodes in a Hadoop cluster for data storage. It links together file systems on local nodes to make it into one large file system. HDFS improves reliability by replicating data across multiple sources to overcome node failures.

7.4. Apache Spark

Apache Spark is designed to be a replacement for batchoriented Hadoop ecosystem to run-over static and real-time datasets. It is highly suitable for high throughput streaming applications where latency is not a major issue. Spark is memory intensive and all operations take place in memory. As a result, it may crash if enough memory is not available for further operations (before the release of Spark version 1.5, it was not capable of handling datasets larger than the size of RAM and the problem of handling larger dataset still persists in the newer releases with diferent performance overheads). Few research eforts, such as Project Tungsten, are aimed at addressing the efciency of memory and CPU for Spark applications. Spark also lacks its own storage system so its integration with HDFS through YARN or Cassandra using Mesos is an extra overhead for cluster confguration.

Other framework such as Apache Flink, Apache Storm, Apache Samza can also be used. [7]

8. APPLICATIONS

The big data application refers to the large-scale distributed applications which usually work with large data sets. In the span of big data, data exploration and analysis became a difficult problem in many industries. With large and complex data, conventional data processing applications fail to manage computation, prompting the development of big data applications.

8.1. Google Maps

Google's map reduce framework and apache Hadoop are the software systems for big data applications, in which these applications produce a large amount of intermediate data. Big data systems are primarily used in manufacturing and bio- informatics. Big data offers a transparent infrastructure for the manufacturing sector, allowing it to address uncertainties such as inconsistency, component performance and availability. A conceptual structure of predictive manufacturing starts with data acquisition in these big data applications, where various types of sensory data such as pressure, vibration, acoustics, voltage, current, and controller data can be acquired.

8.2. Integrating Historical Data

Big data in manufacturing is generated by combining sensory data with historical data. The input is the generated big data from the above combination. The parallel distributed computing system is combined with computer clusters and web interfaces in cloud computing. Software packages for Big Data include a rich range of tools and options that allow an individual to map the entire data field across the company, enabling the individual to evaluate the risks he or she faces internally. This is regarded as one of the most significant benefits, as big data ensures data security. This helps a person to identify potentially sensitive information that is not properly protected and ensures that it is processed in compliance with regulatory requirements.

8.3. Boosting company's performance

If big data are combined with predictive analytics, it produces a challenge for many industries. The combination results in the exploration of these four areas: Calculate the risks on large portfolios, Detect, prevent, and re-audit financial fraud, improve delinquent collections, execute high value marketing campaign.Companies may use big data to build new products and services, boost existing ones, and even invent completely new business models. Big data analytics can be used to obtain such benefits in a number of fields, including consumer intelligence, supply chain intelligence, performance quality, and risk management, and fraud detection. In the field of risk management, sectors such as investment or retail banking, as well as insurance, will benefit from big data analytics. Big data analytics can aid in the selection of investments by analysing the probability of gains versus the probability of losses, which is a crucial feature of the financial services industry. Internal and external big data may also be evaluated for the full and dynamic appraisal of risk exposures.

8.4. Prevention of fraud

Big data analytics can be used to detect and prevent fraud, especially in the government, banking, and insurance industries. While analytics are still widely used in automated fraud detection, organisations and sectors are increasingly looking to big data to improve their systems. They can use big data to match electronic data from various sources, both public and private, and perform faster analytics. Manufacturing, retail, central government, healthcare, telecom, and banking are few of the sectors that can benefit from big data analytics.[8]

9. ADVANTAGE OF THIS INTEGRATION

1. There are many advantages of integrating cloud computing to big date. Big data raises the need for multiple servers because of the massive data and size it works on, and it demands high velocity and variability. These multiple servers work in parallel to provide the big data high requirements.

2. Cloud computing already uses multiple servers and allow resource allocations. Because of that, it is a great fit to build the big data on these cloud multi-servers and make use of the resource allocation availability provided by the cloud environments which would result in a better efficiency for big data analysis.

3. Using cloud system as a storage for big data would improve the performance for both. As cloud systems are mainly based on remote multi-servers which makes it feasible to handle massive amounts of data simultaneously. This feature allows advanced analytics techniques to enable the big data to deal with massive amounts of data. The integration between cloud computing and big data would result in cost reduction.

4. While big data requires clusters of servers and volumes to support the massive amount of data. Cloud computing systems can work as the structure for all of these servers and volumes instead of creating new ones for big data which also provides more flexibility and scalability and eliminates the huge investments that would be invested on the big data computers and servers .[9]

10. ISSUES THAT ARE BEING FACE

Although big data solves many current problems regarding high volumes of data, it is a constantly changing area that is always in development and that still poses some issues. In this section we present some of the issues not yet addressed by big data and cloud computing. As the amount of data grows at a rapid rate, keeping all data is physically cost-ineffective. Therefore, corporations must be able to create policies to define the life cycle and the expiration date of data (data governance). Moreover, they should define who accesses and with what purpose clients' data is accessed. As data moves to the cloud, security and privacy become a concern that is the subject of broad research. Big data DBMSs typically deal with lots of data from several sources (variety), and as such heterogeneity is also a problem that is currently under study. Other issues currently being investigated are disaster recovery, how to easily upload data onto the cloud, and Exaflop computing. Within this section we provide an overview over these problems.

10.1. Security

Cloud computing and big data security is a current and critical research topic (Popovic Hocenski, 2015). This problem becomes an issue to corporations when considering uploading data onto the cloud. Questions such as who is the real owner of the data, where is the data, who has access to it and what kind of permissions they have are hard to describe. Corporations that are planning to do business with a cloud provider should be aware and ask the following questions:

a) Who is the real owner of the data and who has access to it?

The cloud provider's clients pay for a service and upload their data onto the cloud. However, to which one of the two stakeholders does data really belong? Moreover, can the provider use the client's data? What level of access has to it and with what purposes can use it? Can the cloud provider benefit from that data? In fact, IT teams responsible for maintaining the client's data must have access to data clusters. Therefore, it is in the client's best interest to grant restricted access to data to minimize data access and guarantee that only authorized personal access its data for a valid reason. These questions seem easy to respond to, although they should be well clarified before hiring a service. Most security issues usually come from inside of the organizations, so it is reasonable that companies analyse all data access policies before closing a contract with a cloud provider.

b) Where is the data?

Sensitive data that is considered legal in one country may be illegal in another country, therefore, for the sake of the client, there should be an agreement upon the location of data, as its data may be considered illegal in some countries and lead to prosecution. The problems to these questions are based upon agreements (Service Level Agreements – SLAs), however, these must be carefully checked in order to fully understand the roles of each stakeholder and what policies do the SLAs cover and not cover concerning the organization's data. This is typically something that must be well negotiated. Concerning limiting data accesses, (Tu et al., 2013) and (Popa et al., 2011) came up with an effective way to encrypt data and run analytical queries over encrypted data. This way, data access is no longer a problem since both data and queries are encrypted. Nevertheless, encryption comes with a cost, which often means higher query processing times.

10.2. Privacy

The harvesting of data and the use of analytical tools to mine information raises several privacy concerns. Ensuring data security and protecting privacy has become extremely difficult as information is spread and replicated around the globe. Analytics often mine users' sensitive information such as their medical records, energy consumption, online activity, supermarket records etc. This information is exposed to scrutiny, raising concerns about profiling, discrimination, exclusion and loss of control. Traditionally, organizations used various methods of deidentification (anonymization or encryption of data) to distance data from real identities. Although, in recent years it was proved that even when data is anonymized, it can still be re-identified and attributed to specific individuals (Tene Polonetsky, 2012). A way to solve this problem was to treat all data as personally identifiable and subject to a regulatory framework. Although, doing so might discourage organizations from using de-identification methods and, therefore, increase privacy and security risks of accessing data.

Privacy is undoubtedly an issue that needs further improvement as systems store huge quantities of personal information every day.

10.3. Heterogeneity

Big data concerns big volumes of data but also different velocities (i.e., data comes at different rates depending on its source output rate and network latency) and great variety. The latter comprehends very large and heterogeneous volumes of data coming from several autonomous sources. Data comes to big data DBMS at different velocities and formats from various sources. This is because different information collectors prefer their own schemata or protocols for data recording, and the nature of different applications also result in diverse data representations . Dealing with such a wide variety of data and different velocity rates is a hard task that Big Data systems must handle. This task is aggravated by the fact that new types of file are constantly being created without any kind of standardization. Though, providing a consistent and general way to represent and explore complex and evolving relationships from this data still poses a challenge.

10.4. Data Governance

Data Governance came to address this problem by creating policies that define for how long data is viable. The concept consists of practices and organizational polices that describe how data should be managed through its useful economic life cycle. These practices comprise three different categories:

1. Structural practices identify key IT and non-IT decision makers and their respective roles and responsibilities regarding data ownership, value analysis and cost management (Morgan Kaufmann, 2013).

2. Operational practices consist of the way data governance policies are applied. Typically, these policies span a variety of actions such as data migration, data retention, access rights, cost allocation and backup and recovery (Tallon, 2013).

3. Relational practices formally describe the links of the CIO, business managers and data users in terms of knowledge sharing, value analysis, education, training and strategic IT planning.

Data Governance is a general term that applies to organizations with huge datasets, which defines policies to retain valuable data as well as to manage data accesses throughout its life cycle. It is an issue to address carefully. If governance policies are not enforced, it is most likely that they are not followed. Although, there are limits to how much value data governance can bring, as beyond a certain point stricter data governance can have counterproductive effects.

10.5. Disaster Recovery

Data is a very valuable business and losing data will certainly result in losing value. In case of emergency or hazardous accidents such as earthquakes, floods and fires, data losses need to be minimal. To fulfil this requirement, in case of any incident, data must be quickly available with minimal downtime and loss.

For big corporations it is imperative to define a disaster recovery plan – as part of the data governance plan – that not only relies on backups to reset data but also in a set of procedures that allow quick replacement of the lost servers Successfully deploying big data DBMSs in the cloud and keeping it always available and fault-tolerant may strongly depend on disaster recovery mechanisms. [8]

11. BUSINESS CHALLENGES

11.1. Utilities: Power consumption prediction

Utility companies use smart meter to measure gas and electricity consumption. These devices generate huge volumes of data. A big data infrastructure needs to monitor and analyse power generation and consumption using smart meters.

11.2. Social Network: Sentiment analysis

Social networking companies such as Twitter needs to determine what users are saying and topics which are trending in order to perform sentiment analysis.

11.3. Telecommunication: Predictive analytics

Telecommunication provides need to build churn models which depends on the customer profiledata attributes. Predictive analytics can predict churn by analysing the subscribers callingpatterns.

11.4. Customer Service: Call monitor

Call center big data solutions use application logs to improve performance. The log files needs to be consolidated from different formats before they can be used for analysis.

11.5. Banking: Fraud Detection

Banking companies should be able to prevent fraud on a transaction or a user account. Big data solutions should analyse transactions in real time and provide recommendations for immediate action and stop fraud.

11.6. Retailers: Product recommendation

Retailers can monitor user browsing patterns and history of products purchased and provide a solution to recommend products based on it. Retailers need to make privacy disclosures to the users before implementing these applications. [10]

12. CONCLUSION

Big data and cloud computing play a huge role in the current digital world. The application of Big Data in Cloud Computing seems to have a huge potential in the coming years. While using Software as Service, typically, big data plays a pretty important role in giving insight, in cloud computing applications. Big Data when applied in cloud computing, has many applications in different fields. Some of these applications include improved analysis due to large data size, creation of an efficient infrastructure while reducing the cost in the long run and allowing better integrity and availability and security of the cloud platform, letting the businesses and platforms grow through the means of big data.

13. REFERENCES

[1]Manoj Muniswamaiah, Tilak Agerwala and Charles Tappert, "BIG DATA IN CLOUD COMPUTING REVIEW AND OPPORTUNITIES,"International Journal of Computer Science Information Technology (IJCSIT) Vol 11, No 4, August 2019.

[2]Dr. Harjinder Kaur, Dr. Major Singh Goraya, "Role of Big Data in Cloud Computing: A Review", International Journal of Engineering Research Technology (IJERT), Vol. 8 Issue 07, July-2019

[3]Pedro Caldeira Neves, Bradley Schmer, Jorge Bernardinoand Javier Cámara, "Big Data in Cloud Computing: features and issues",

[4] Parthil Jagani, Neelay Jagani, Suril Shah, "BIG DATA IN CLOUD COMPUTING: A LITERATURE REVIEW", International Journal of Engineering Applied Sciences and Technology, 2021 Vol. 5, Issue 11, ISSN No. 2455-2143, Pages 185-191

[5] Samir A. El-Seoud, Hosam F. El-Sofany, Mohamed Abdelfattah, Reham Mohamed, "Big Data and Cloud Computing: Trends and Challenges"

[6] Bernice M. Purcell, "Big data using cloud computing", Journal of Technology Research.

[7] Saeed Ullah , M. Daud Awan, and M. Sikander Hayat Khiyal,"Big Data in Cloud Computing: A Resource Management Perspective",Hindawi Scientific Programming Volume 2018, Article ID 5418679, 17 pages

[8] T.Sri Harsha,"Big Data Analytics inCloud Computing Environment ", International Journal of Scientific Engineering Research Volume 8, Issue 8, August-2017

[9] Amanpreet Kaur Sandhu, "Big Data with Cloud Computing: Discussions and Challenges", BIG DATA MINING AND ANALYTICS, ISSN 2096 -0654 03/06 pp 32 – 40,, Volume 5, Number 1, March 2022

[10] Venkatesh , Shrivatsa D, Nivedita Jalihal, "A Study on Use of Big Data in Cloud Computing Environment" , (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (3) , 2015, 2076-2078