OORJA
ISSN - 0974-7869 (Print)
ISSN - 2395-6771 (Online)

CLOUD-A FACILITATOR FOR BIG DATA

Ms. Vandana Nigam, Assistant Professor, S.S. Jain Subodh P. G. (Autonomous) College, Jaipur
Ms. Shalu J. Rajawat, Assistant Professor, S.S. Jain Subodh P. G. (Autonomous) College, Jaipur

Abstract

As the popularity of the internet increasing day by day, fiery growth of voluminous data also takes place very rapidly. Big data is a technique to process, handle and store voluminous structured, semi structured and unstructured data in an efficient and systematic manner. Cloud computing is an open space capable to hold such distributed big data that provides software, platform, service and infrastructure as per demand. It provides consistent, error liberal and scalable environment to port big data distributed management system. Traditional data management techniques are not sufficient to grip the large data as they suffer from poor scalability and compatibility, less fault acceptance and low performance under high pressure. Big data technique is capable to handle these data management problems but suffers from some trouble like security, safety and data recovery management. Big data and cloud computing together can handle such problems more intelligently and have significant advantages over such traditional system. Dealing with Big Data is a very tedious and time consuming mission that necessitates huge resources to handle it and cloud provide the necessary support for the same. Big Data processing is an intricate task that includes data collection, storage and analysis and cloud provides trustworthy, fault understanding, obtainable and ascendable environment for this Big data distributed management system. Although big data and cloud crack lots of our recent troubles but still they have some holes and problems. Big Data needs enhancement in Storage Issues, Data Transportation Issues, Data Management Issues, Processing Issues, and Data Security Issues which can be easily done by Cloud computing and cloud needs enhancement in huge computing, storage resources, encourage and accelerate the development of computing efficiently by Big Data. They are closely connected and can balance each other. In other words, Big data symbolizes the goods and the cloud symbolizes the container. This paper discusses how cloud computing facilitates Big data with its relationships and issues.

Keywords-Cloud Computing, Big Data, Big Data Issues

Introduction

Today's era is the era of data & information and Internet is the biggest source of data used by variety of users for diversified reason like search, research, study, knowledge gaining and other purposes. Organizations, institutions, firms and companies are producing voluminous data on regular basis for their clients, students, faculties and their members regularly and this data is added on the websites, so it can reach to majority of people beyond physical boundaries instantaneously. Internet is growing day by day very rapidly New web pages are added, older ones are modified and millions of web pages are accessed regularly from it. New and existing data not only contains customary text but it also contains video, voice, images and more complex contents. This data is very huge, diversified and different in many aspects and needs more sophisticated technology for handling.

Literature Review

Bernice M. Purcell (2014) in their paper described that big data is a very huge version of database and can't be handled by the traditional method of data keeping system hence some more advanced and any time available technique is required which can be availed by using cloud computing technology. Cloud Services Providers not only provide the necessary storage but also furnish the task of management and maintenance of the same at lower costs. They suggested that small to medium industries can keep their huge data on cloud without investing more on the services and infrastructure and without putting much cost on their financial resources.

Nabeel Zanoon, Abdullah Al-Haj, Sufian M Khwaldeh (2017) in their paper described that big data and cloud computing are complementary to each other and quoted that big data is a product and the cloud is the container. Big data deals with gigantic data and cloud is the online facility to keep data online and can be accessed with ease at low cost and along with that cloud provides flexible and distributed data management techniques.

Marcos D Assuncao, Rodrigo N. Calheiros, Silvia Bianchi, Marco A. S. Netto, Rajkumar Buyya (2014) in their research paper suggested that big data deals with complex massive data. Big Data analytics is a complex process that demands an expert brain that can deal with understanding the data, selecting the data and analyzing it with more advanced techniques to deal with the same. They explained that Cloud can play a key role for big data as it provides infrastructure to medium-small industries and suggested two improved models AaaS (analytics as a Service) for facilitating analysis for complex data and BDaaS (Big Data as a Service)

Marcos D. Assunc Rodrigo N. Calheirosb, Silvia Bianchic, Marco A. S. Nettoc, Rajkumar Buyya (2014), in their research paper explained that big data is dealing with structured, semi structured and unstructured data. Most of the data comes to the unstructured data category so it is expected to manage, analyze and maintain the data more intelligently without increasing the cost to a great extent especially for medium and small users, so cloud is a better choice for the same. Cloud technology offers better solution for data management, data integrity and data analysis as cloud provides more flexible services for handling big data shrewdly.

Babita Ahuja, Anuradha, Ashish Ahuja (2013) in their paper explained that hidden data is a high quality data that should be available to researchers for their research work. This data is a very huge in measure and generally remains unavailable to users.

Cloud Computing

According to Peter M. Mell, Timothy Grance (2011), described that Cloud Computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction. Cloud is probably the most promising technology in modern era which completely transforms the way with which the IT world works. Cloud service providers provide pool of computing resources in the form of servers, networks, storage, applications and services which are used by the geographically dispersed cloud users on demand against some charges. Cloud technology provides application, platform and software to users.

Big Data- Big data is extremely large size data sets which are very difficult to confine and manage with conventional data processing application software as they are not ample to deal with such

complex data. Big data is not a data but a huge quantity of data and handling of such huge data (approx. 1 TB or more data daily) requires faster methods to deal with. The giant size of big data requires much more speedy techniques to tackle it in minimum processing time and output of the processing will be return by the system within acceptable time. Big data is often characterized by three 3 Vs the tremendous quantity (volume) of data, the extensive range (variety) of data types and the swiftness (velocity) at which the data must be processed. This data can be divided in following categories-(Figure 1)

1. Structure Data- Structure data is organized data which is processed and arranged in some specific manner. It is helpful in making some decisions and conclusions or reaching on some actions. On internet only 2-4% is included in this category.
2. Semi Structured Data-An intermediate form of the data which is neither raw nor typed in traditional database method. Only some part of the data is organized in database system. About 16% data comes in this category.
3. Unstructured Data- Unstructured data is unorganized data which needs processing and operations to use it in an efficient way. More than 80% data on internet comes in this category.

Along with above mentioned categories, on the basis of accessibility (crawled by web crawler), data may be broadly divided into two types Visible data and hidden (unaccessed) data. (Figure 2)

Visible data-It is the data which is searched by crawlers and included in the list and index of the crawled pages, but the listing is refreshed periodically to get newly added data with modified contents.

Hidden Data-Hidden Data is the data which is not being crawled by the web crawlers. It is also known as deep data or deep web.

Benefits of Cloud Technology for Handling Big Data

Cloud Computing and Big Data both are ultimate mixture of benefits in today's world. This combination is successful because it can handle various troubles like handling voluminous data, accessing fast, availability of data anywhere any time and at any place which are essential feature that any organization would be looking for. Big data is a artifact and cloud is a container so combination of both serves at its best to the businesses. Few benefits are discussed here-

1. Agility: Traditional database system is old-fashioned, slower to access, store and retrieve data. If we want to store large amount of data that is rapidly increasing day by day on server then it takes plenty of time approx few weeks for installation and processing. A better solution for this problem is Cloud Computing that can store data in more managed way than conventional methods. It has faster speed for installing, retrieving and processing the big data. The voluminous data are stored on number of virtual servers within few minutes.
2. Affordability: To store Big Data, ample space or database is required and that must not be very expensive. This purpose is solved by cloud technology on which saving the data is cost effective and under expected budget. Before Cloud, companies paid a lot of money for setting Information Technology and hardware updating methods. Presence of Cloud Technology not only reduces the cost of handling huge data but also relinquish the organization from keeping, managing, updating, accessing, replicating and securing the data as all such factors are handled by cloud service providers and businesses have to pay

according to the use only.

3. Data Processing: Big data is a huge data be which is in structured, semi structured or unstructured form. This kind of data can't processed easily due to its giant size. An environment is necessitated which can complete the whole process in an easier manner and reachable to small, medium and big venture and that is cloud computing. Cloud provides infrastructure, services, platform, virtual servers and many more to handle big data efficiently and conveniently as per user's expectations.
4. Feasibility: For large data sets processing, Big data analytics required an environment which should be unlimited to handle continuously raising or falling data. So for satisfying this scale of measuring an ultimate solution is Cloud, which is perfect podium to fulfill giant task easily and more quickly than other available technologies. Cloud Technology offers more feasible environment to handle Big Data.
5. Elasticity: Cloud elasticity offered by Cloud Computing is ideal for big data. Cloud continuously can hold the huge amount of big data and diagnose its patterns and also look up related business strategies detecting by several cloud providers. Cloud distributed nature makes it suitable for big data handling as well as analysis as per current and future aspects.

Growth of Big Data in Cloud Technology

We are living in a new era of data age and surrounded everywhere from data. It is the essential part of our daily life and is increasing day by day, as the volume of data is increasing very rapidly it is necessary to handle it with caution and care. As data is biggest asset for any organization it is necessary to keep it secure and readily available at any time for current and future exploit. A useful technique or platform is strongly needed to cater the same and cloud technology is best available choice. In this section we discussed about how speedily data is increasing at fiery rate, so businesses and organizations are moving towards cloud to keep their giant data readily available at any moment. The statistics shows that organizations are uploading, downloading and using the data up to 2TB or more every day on servers which is big enough than expected. To handle such big data Cloud is undoubtly a suitable obtainable technology.

Figure 3 shows that, from 2010 to 2014 the growth of big data was increased at steady rate but from 2015 onwards the growth of big data is shooting very rapidly and expected to cross 60-61 zettabytes in year 2021. To handle such vast data Cloud would be the preferred choice and due to this reason the organizations believe to outsource the data rather than handling it in-house.

Figure 4 depicts that users preferably store their data on their portable devices and PCs. Use of secondary devices is lowering in comparison to them. Corporate, organization and businesses are moving towards cloud technology because the volume of their data is very gigantic than small users. PCs and small devices have their hardware limitation and volume of corporate data is many times bigger, hence more space is needed which can handle the big data with convenience and care. The graph indicates that contribution of cloud for data storage is increased sharply and crosses 42% or more in year 2021. Boosted rate of using cloud shows that cloud is again a suitable method to keep giant data on several servers than a single one.

Figure 5 shows the growth rate of mobile data on cloud was almost steady from 2010 to 2015 but it is

very clear from the graph that mobile data will occupy approx 40% of the total data stored (10 zettabytes) on the cloud in 2021 and there will be very strong growth of mobile data in coming time. The graph also depicts that cloud technology will probably be the most preferred and popular choice for keeping the mobile giant data for the same.

As is seen in Figure 6 In real time data, information is distributed immediately after assortment, so this data is to be stored and kept without delay as timeliness is the prime need of real time data. It is expected that real time data will grow approx 1.5 times the rate of overall data formation. From graph it is depicted that real time data will occupy almost half of the total data stored on the cloud servers.

Issues and Challenges Related with Big Data in Cloud Technology

In present scenario cloud technology and big data are complementary to each other and is a profitable combination to big, medium and small enterprises. Big data is capable to handle voluminous data more smartly than traditional methods on one side and on other side cloud endowed with platform, services and infrastructure which save capital investments of the organizations to a big extent.

Besides providing a lot of benefits, there are several issues and challenges related with big data and cloud technology. For Big data major hindrance factors are capturing the voluminous data, managing its cargo space, analysis of data, searching, transferring and sharing the data, making queries on vast data and maintain the privacy and security of the data are some big hurdles for big data. For cloud computing technology multi tenancy, hardware virtualization, security threats, load balancing, resource scheduling, concurrency control, compromised security and many more factors lowers its acceptance rate in the current scenario.

Big data puts meticulous demands on networks and storage servers for current and future needs. This huge data requires storage, accessing, handling and maintaining in a graceful and secure manner. Companies and businesses find it more convenient to outsource the data without investing more funds to manage and maintain the large volume. Cloud is the best available choice for the same as it's a vibrant technology to do complex computing on such substantial volume. Cloud supports big data analysis and accommodate large data, but is not free from issues and challenges. Various challenges are there that put very serious questions on the combination of both. Some of the combined issues and challenges are discussed here:-

1. Handling Vast Data with Cloud Technology-Organizations producing hefty data every day in Terra Bytes, which are very bulky and complex to handle with the existing cloud technology. Most of the cloud service providers and still managing the data with traditional methods, but analyzing, accessing, summarizing such big data with traditional methods is more cumbersome, and from this big volume of data analyzing only few important useful bytes is a more challenging task.

2. Heavy Load on Storage-Big data requires huge server space to put data on cloud. Storage and processing of voluminous data needs fault tolerance, scalability and availability. Cloud delivers all these factors through virtualization, but cloud storage is geographically dispersed so processing, querying and managing such large data puts heavy load on the storage and working capabilities. Successful data processing requires very fast and gigantic infrastructure at cloud service providers end. Adding new infrastructure and fast processing at cloud's end not only puts financial pressure but also desires innovation and technological advancement at regular basis to cope up the difference in

existing and potential requirements.

3. Issues Related with Capacity and Performance-Data is the most valuable asset for the organization. In today's era companies are facing data bang and need to analyze and summarize it in least time with proper output format expected by users, but physical hardware limitation and can handle it upto some extent only. Need of Big data analysis is increasing day by day at very rapid rate which highly affects the cargo space performance in virtualized distributed cloud and is probably one of the strong reasons that enterprises not taking interest in choosing cloud for big data processing.

4. Need of a Combined Design-Big data and cloud are not same and compatible to each other but they are different from each other in many respects. A combined architecture is expected which can easily club big data and cloud at one place, but building an integrated architecture is one of the most difficult tasks and it also needs to incorporate service based approach at cloud end and data tuning at big data end to solve architectural challenges faced by both.

5. Replication of Data at Cloud Data Centers- When data is stored on cloud data centers it is expected that data should be replicated without leaving nil errors during copying. Replication is required so that if some problems occur at one place then users can easily get their required data from replica without knowing and facing problems. During duplication, if errors exist then it can not only create new issues but break the trust of the users. So capturing such voluminous data, its storage, transferring data, sharing, visualization and analyzing of data as flat as possible are the major concerns which keeps its acceptance difficult to users.

6. Movement of Large Data Sets-Cloud is an on air technology where data moves from one place to another place as and when needed by the users, this moving data not only carries general information but also have some confidential data like passwords, debit or credit card numbers, personal details of person or organizations and uploading of voluminous data on cloud many times in a day etc. This moving data must be kept secure to protect it from unauthorized access and from unwanted access. Managing moving of voluminous data is also a big challenge for cloud and big data combo which again hinders its full growth.

7. Security Issues-Cloud and big data are seen as next generation paradigm in computation but security issues are major concerns for organizations, business and cloud technology. The biggest challenge of the cloud service providers is to keep the big data secure and protected. Everyday new denial of service attacks, insecure interfaces, virus attacks, phishing attacks, worms and digital pests attacks, cloud abuse and other means find new entry points in the system and keep attacking and damaging valuable data Ransomware are another problem which affect the company's status and resources. Statistics shows that about 40-42% of businesses experienced Ransomware events during last some years. To manage resources, data and security is probably the biggest challenge for cloud service providers. In near future accepting cloud at various levels may be the biggest task for service providers and security is most important concern for both service providers and users.

8. Lack of Standards-Standardization is a mean to apply rules globally but unfortunately cloud and big data missing such worldwide accepted standards. Lack of knowledge, different cloud vendors and their own implementation policies are some factors due to which time honored standards are not taking its final shape. Some standards are still available at cloud provider's level but are not accepted globally so there is a need to maintain worldwide acceptable standards to resolve this issue and to build such global standards high level of expertise is needed.

From above discussion it is clear that big data and cloud technology are in their infant stage and suffering from some challenges which not only act as hurdles for their growth but also prevent potential users to accept it with trust and reliability. Organizations and enterprises want steady Return on Investments but it is their responsibility also to protect data at own level. Combination of big data and cloud technology is a good example of need (handle vast data) and tool (cloud technology) but is still facing major issues discussed above.

Future Scope and Conclusion

It is clear that cloud technology for handling big data is more acceptable method than any other available technology. Cloud not only handles and takes care of such huge complex data more efficiently but also provides data as per demand of users. Besides providing data handling competently, cloud provides Services, Infrastructure and Application without investing more on them. This paper discussed various issues related to cloud and big data and found that security is the biggest issue among all of them. Handling such big multifaceted data, more superior techniques and sophisticated algorithms are to be developed in future so that trust of existing and potential users is increased. It is also clear from above discussion that cloud is more appropriate choice for storing, handling and accessing big data. Cloud technology is very useful especially for medium and small firms and organizations as various experiments can be performed on big data without making extensive promise of firms wherewithal. It is a launch pad of big data which is capable to handle it effectively and efficiently for present as well as future uses.

References

- Bernice M. Purcell (2014). Big data using cloud computing. Journal of Technology Research (page 01-08).
- Nabeel Zanoon, Abdullah Al-Haj, Sufian M Khwaldeh, (2017). Cloud Computing and Big Data is there a Relation between the Two: A Study. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 17.
- Marcos D Assuncao, Rodrigo N. Calheiros, Silvia Bianchi, Marco A. S. Netto, Rajkumar Buyya (2014), in their research paper. Big data computing and clouds: Trends and future direction. J. Parallel Distrib. Computing (79-80) (9-15) Journal of Parallel and Distributed Computing.
- Marcos D. Assunc¸ Rodrigo N. Calheirosb, Silvia Bianchic, Marco A. S. Nettoc, Rajkumar Buyya, (2014). Big Data Computing and Clouds: Trends and Future Directions" Journal of Parallel and Distributed Computing.
- Babita Ahuja, Anuradha, Ashish Ahuja, (2013). Hidden Web Data Extraction Tools. International Journal of Computer Applications (0975 – 8887) Volume 82 – No 15, November.
- Peter M. Mell, Timothy Grance (2011). The NIST Definition of Cloud Computing. https://www.nist.gov/news-events/news/ 2011/10/final-version-nist-cloud-computing-definition-published. Accessed on 2 June, 2017.
- Big Data and Cloud computing- challenges and opportunities. Available at http://bigdata madesimple.com/big-data-and-cloud-computing-challenges-and-opportunities/ on dated

2nd June 2017, accessed on 2 January, 2018.

- Jessica Scarpati, (2012). Big data analysis in the cloud: Storage, network and server challenges available at https://searchtelecom.techtarget.com/feature/Big-data-analysis-in-the-cloud-Storage-network-and-server-challenges.accessed on 10 January, 2018.
- David Reinsel, John Gantz and John Rydning, (2017). Total WW Data to Reach 163ZB by 2025 available at https://www.storagenewsletter.com/2017/04/05/total-ww-data-to-reach-163-zettabytes-by-2025-idc. accessed on 2 January, 2018.
- https://polestarlip.com/challenges with big data on cloud, accessed on 6 January, 2018.
- Yunchuan Sun,Junsheng Zhang, Yongping Xiong, and Guangyu Zhu, (2014). Data Security and Privacy in Cloud Computing; International Journal of Distributed Sensor Networks Volume 2014 (2014), Article ID 190903.
- http://bigdata-madesimple.com/big-data-and-cloud-computing-challenges-and-opportunities, accessed on 6 January, 2018.
- Jaydip Sen, Security and Privacy Issues in Cloud Computing. Innovation Labs, Tata Consultancy Services Ltd., Kolkata, India.
- Cloud computing challenges in 2017. available at http://www.opencirrus.org/cloud-computing-challenges-2017/ accessed on 25 April 2018.
- R. Velumadhava Rao, K. Selvamanib (2015). Data Security Challenges and Its Solutions in Cloud Computing Procedia Computer Science, Volume 48 www.sciencedirect.com, accessed on 20 April, 2018.

OORJA
ISSN - 0974-7869 (Print)
ISSN - 2395-6771 (Online)

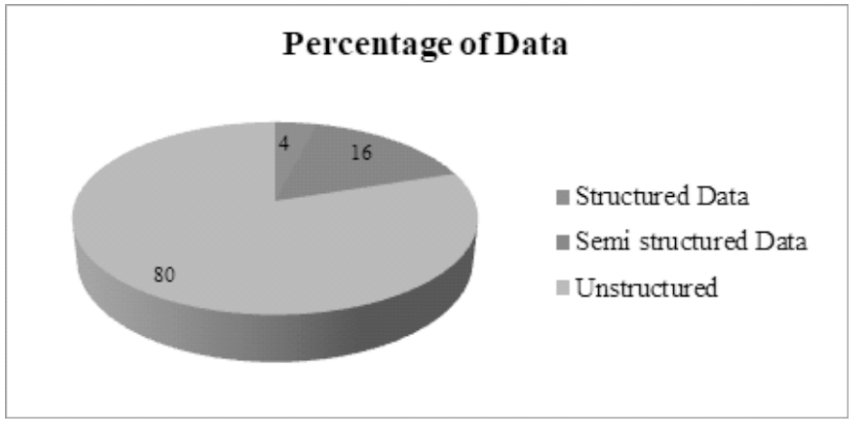

Figure 1 : Ratio of Structured, Semi Structured and Unstructured Data on Web
Source : Christine Taylor, (March, 2018). Structured vs. Unstructured Data. Retrieved from www.datamation.com

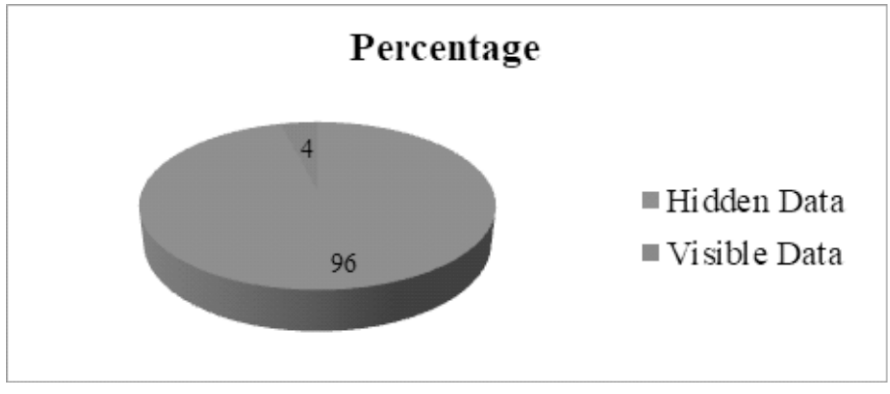

Figure 2 : Ratio of Hidden and Visible Data on Web
Source : Bergman, Michael K (2001).The Deep Web: Surfacing Hidden Value. The Journal of Electronic Publishing. Vol. 7 (1)

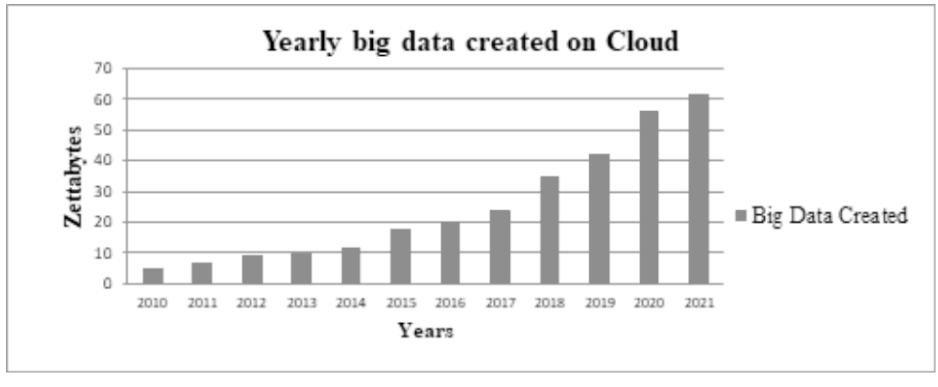

Figure 3 : Yearly Big Data created on Cloud
Source : Market Reports/Research (2018). Top 25 Cloud Backup Enablers Backup Review retrieved from https://www.storagenewsletter.com ›

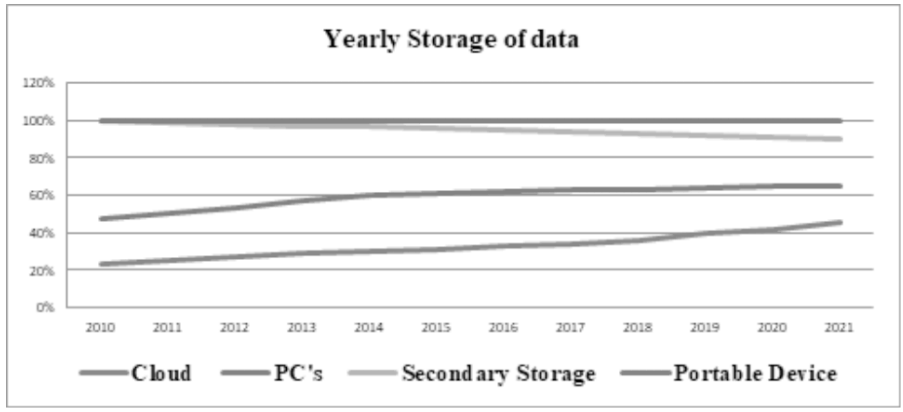

Figure 4 : Yearly Storage of Data on Different Devices and Media

Source : Market Reports/Research (2018). Top 25 Cloud Backup Enablers Backup Review retrieved from https://www.storagenewsletter.com ›

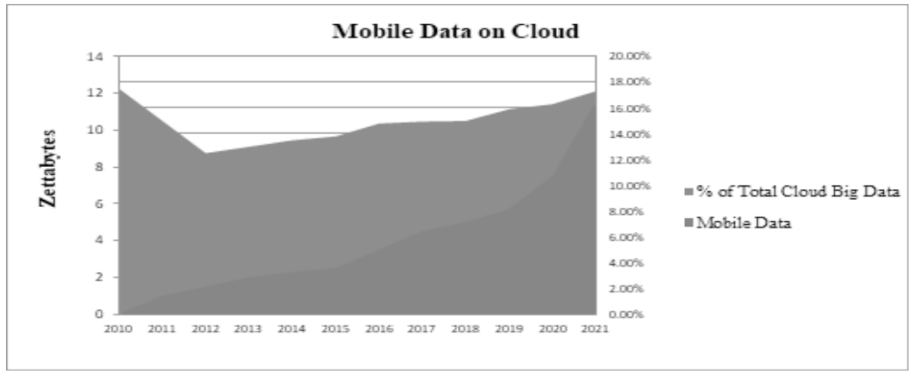

Figure 5 : Ratio of Mobile Data and Total Big Data on Cloud

Source : Market Reports/Research (2018). Top 25 Cloud Backup Enablers Backup Review retrieved from https://www.storagenewsletter.com ›

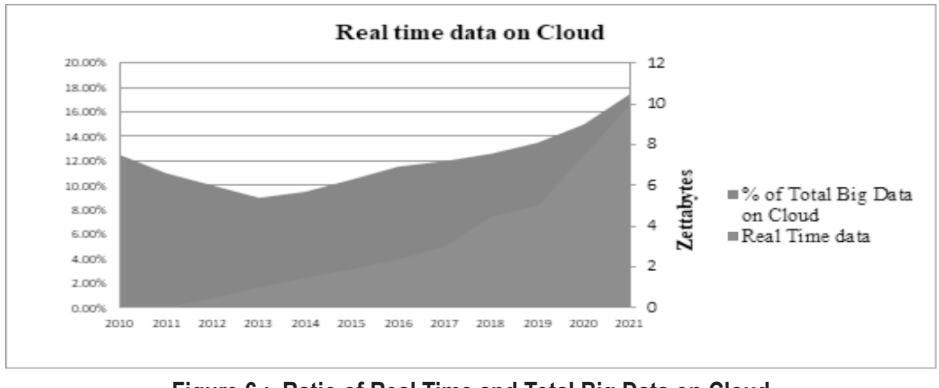

Figure 6 : Ratio of Real Time and Total Big Data on Cloud

Source : Market Reports/Research (2018). Top 25 Cloud Backup Enablers Backup Review retrieved from https://www.storagenewsletter.com ›