e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:01/January-2023 Impact Factor- 6.752 www.irjmets.com

ANALYZING BIG DATA CHALLENGES AND SECURITY ISSUES IN DATA PRIVACY

S. Ramya*1, R. Sakthi Devi*2, Dr. P. Senthil Pandian*3, G. Suguna*4, R. Suganya*5, N. Manimozhi*6

*1,2,4,5,6Assistant Professor, Department Of Computer Science And Applications, Farouk Educational Trust, Perambai, Villupuram District, Tamilnadu – 605110, India.

*3HOD-Department Of Computer Science And Applications, Farouk Educational Trust, Perambai, Villupuram District, Tamilnadu – 605110, India.

ABSTRACT

The big data environment supports to resolve the issues of cyber security in terms of finding the attacker. There are security challenges of big data as well as security issues the analyst must understand. In this paper, the challenges faced by an analyst include the Data storage, fake data, Data access control, Data Management, Data Poisoning; Data privacy issues are well studied. It is more focused on Big data Security Practice to protect big data in terms of applying encryption capabilities.

Keywords: Data Poisoning, Data Privacy, Data Storage, Cyber Attack's, HDFC.

I. INTRODUCTION

The successful paradigm for the service oriented programming is the cloud computing. It has revolutionized the way of computing infrastructure's abstraction and usage. The elasticity, pay per use, low upfront investment, transfer of risks are few of the major enabling characteristics that makes the cloud computing the ubiquitous platform for deploying economically feasible enterprise infrastructure settings. Distributed databases had been the boon of vision for research for few decades. But changes in the data patterns and applications has made way for the new type of storage called key value storage which are now being widely used by various enterprises. In the domain of Map reduce and open source implementation of the same known as the Hadoop has been used by majority of the industry and academics. Hadoop increases the usability and performance .HDFS has become a Very helping tool to maintain and store the complex data. Big data has becoming more available and understandable to computers. What is big data? The question arrives. Big data is the representation of progress of the human cognitive processes, usually includes data sets with sizes that is beyond the current technology's capability. The data which is very fast, has various varieties and requires new type of the processing forms to enable decision making, insight discovery and optimization of process. In order for analyzing the data and for identification of patterns it is very important for us to store the data securely, manage and sharing of complex data on cloud. Since cloud involves extensive complexity, we feel it's ideal to make enhancements in securing cloud than showing holistic solutions. In this paper we provide a comprehensive background study of state of art systems. Identification of critical aspects in design of various systems and scope of the systems. We show up some approaches in security provision through a scalable system to handle large number of sites and also have the capability to process large and massive amounts of data. We also provide the status of big data studies and related works, aiming at providing a overview of managing big data and its applications.

II. BIG DATA ANALYTICS

Big Data in cloud refers to enormous size of the dataset perhaps in few dozens of terabytes and petabytes and thus working with them in a traditional local computer based Database Management System becomes enormously difficult. The ability to scale storage, visualize data, manage and capturing becomes very tedious and highly costly and thus use of cloud is the most apt solution. Many of the world's largest organization are storing all of their data on cloud. These enterprises are able to explore large volumes of highly detailed data so as to discover facts they didn't know with the help of inbuilt cloud features or deploying their own functionality on the cloud. Naturally businesses can benefit from large data with almost real time capability, and thus the cloud needs to have different data architecture, analytical methods, and tools.

A. Characteristics of Big Data– The feature characteristics of big data are divided into 4V's, namely Volume, Variety, Velocity, and Veracity. The first V, volume refers to the size of the data and how big the data is. This is

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:01/January-2023 Impact Factor- 6.752 www.irjmets.com

the primary and most looked attributing of big data. Velocity refers to the rate at which data is been collected or is changing. Some Big Data's like Stock Market prices are monitored and collected at a very high velocity with frequency as small as one second. The third feature Variety refers to from how many different sources the data is coming from, the data can be coming from logs, social media or even click streams. The last feature Veracity describes how good the data is. The quality of data is measured by observing patterns on how much data is inconsistent, missing, incomplete, approximated, decepted, ambiguous, or latent.

B. Storage Management and Cloud – There are several software packages available on cloud to facilitate cloud computing. Enterprise data warehouse can be used or if there is presence of unstructured data like large texts use of NoSQL can be used. Majorly Hadoop, Spark, Map Reduce, HBase are used. Hadoop is the most available programming framework, written in Java that supports processing of large amount of data. With Hadoop large amount of data can be analysed by use of clusters of servers on these servers we can have thousands of nodes running the application. Hadoop framework helps in risk of system failure even multiple nodes fail. The framework has a flexible and fault tolerant computing solution. The Hadoop Distributed File System (HDFS) defines a very efficient yet high tech file system where in the file system spans all nodes in Hadoop cluster for data storage and connects the file system on local nodes, this improves the reliability significantly. HBase is NoSQL software with Hadoop framework of HDFS. It is an open source database that was modeled after Google's Big table and like Hadoop is written in Java. It is widely used to store Big Data, more accurately when big data is of variety (unstructured). Spark is also an open source tool with unified analytics engine for large scale data processing. Spark provides an interface for programming entire clusters with implicit data parallelism and fault tolerance. As part of storing data on cloud with appropriate tools Azure HDInsight and Amazon EMR are most popular. Both these cloud services for storing data have proven to be enormously effective. They are cloud native; meaning they enable us ML services, create clusters of Hadoop, Spark, Map Reduce, and even HBase. When the velocity of data is high, cloud storage enables to scale workloads up and down accordingly.

C. Big Data Processing – There are four fundamental requirements for processing. 1. The primary requirement is the ability to load the data quickly. 2. Fast query processing. 3. Efficient utilization of storage space. 4. Strong adaptivity to highly dynamic workload. To satisfy all the four requirements efficiently, the cloud service providers help us by providing Map Reduce Software, both Azure HDInsight and Amazon EWS provide Map Reduce framework. The framework helps enormously in processing as it is a parallel programming model. The Map Reduce framework rather than increasing the storage capacity of a server or a computer, or increasing computational power, it adds more servers and computer. Therefore, the fundamental concept is that we do not scale up rather we scale out. In Map Reduce a task is broken down into stages and are executed parallely thus increasing the efficiency. The working is quite simple as the word suggests; the first word Map is used to "map", the smaller tasks and assign them appropriate key value pair. Like for example if we have unstructured data like text, the key could be any word and the value can be the number of occurrences of that word. Next is the reduce function. The reduce function performs collection and combination of the output generated by "map", by combing all values which share the same key value, to provide the final result of the computational task. This is very advantageous as cloud architecture is very fast and when clubbed with parallel processing, the performance is unmatched to a general local computer. When processing speeds are this high, we can analyze data in real time whilst getting the output in real time as well. Such a system when implemented on cloud is very advantageous, and Big Data with high velocity and high volume, companies, exchanges like NASDAQ, BSE, NSE can all benefit. The storage, analytics and processing all are carried out with more efficiency and lower cost when compared to traditional normal computers.

III. CHALLENGES

In spite of all the advantages of the integration between cloud computing and big data, there are some challenges and risks that ought to be thought while deploying big data on a cloud environment.

A. Data Storage – With the advancement of technology, we are able to wit- ness an exponential growth in data. However, most of the generated data is ignored or deleted because enough space is not available to store them. So, the primary challenge for Big Data analysis is storage mediums and better transmission rates. The available storage technologies do not possess the required ability to process Big Data. Storing data on traditional physical

storage systems is a complicated task as hard disk drives often fail, and traditional data protection mechanisms are not efficient. In addition to this, the velocity of Big Data must be such that the storage systems must be able to scale up quickly when required, which is actually difficult to achieve with these traditional storage systems. Due to this ever growing data, data mining tasks has increased considerably which has led to wide diversity of data. There's a need to pay more attention for designing storage systems and to make efficient data analysis tools that will provide guarantees on the output since the data is gathered from different sources. Moreover, machine learning algorithms can be designed for analyzing the data which will help in improving the efficiency and scalability. The unlimited storage along with high fault tolerance offered by Cloud storage services (such as: Amazon S3, Elastic Block Store) provides solutions to address Big Data storage challenges. But, it's very expensive to host and transfer Big Data on the cloud since the size of data is gigantic.

B. Data Transmission – Another challenge is how to move vast amounts of big data (let's take for example hundreds of terabytes of data) into a public cloud in a short period of time? How will we deal with the storage, reliability, privacy, and security issues? Transferring gigantic volumes of data in different stages of data life cycle poses challenges in each of these stages. Therefore, we need to devise smart pre-processing techniques and data compression algorithms to effectively reduce the data size before transferring the data. For transferring data from local data centers to cloud platforms, we need to develop efficient algorithms which will automatically recommend the appropriate cloud service (location) based on the geo temporal principles (since data can be at different locations) to maximize the data transfer speed while the minimizing cost.

C. Computational Complexities – For processing large volumes of data, we require dedicated computing resources, which we usually handle by the increasing speed of storage, network and CPU. However, the processing power and the computing resources provided by the traditional computing system is insufficient for processing the data. The virtually unlimited and on-demand processing power offered by cloud computing acts as a partial solution. However, shifting to the cloud results in some issues. First, the network bandwidth of cloud computing is very limited which affects the efficiency of computation over large volumes of data. Second, the data is dispersed at different locations which makes it difficult to gather it for pre-processing. The essential features of cloud computing such as virtualization, pooled resources of data and high computing power makes it a difficult task to track and ensure data locality, and hampers its ability to support data processing which involves intensive communication and exchange of data.

D. Data Security– Some security vulnerabilities arise due to the integration of Big Data and Cloud Computing. Also, the data security policies and schemes work with the structured data which is stored in conventional DBMS and aren't effective in handling highly heterogeneous and unstructured data. Therefore, we need to make effective policies for data access control and safety management so as to incorporate new data management systems and storage structures. Ensuring data confidentiality, integrity and availability becomes elemental in this cloud era since the data owners have limited control over the data and various resources. Heterogeneity is one of the most known Big Data's cloud security vulnerability. In many cases the deployment of Big Data requires it to deploy on a new cloud platform which will need new security tools to be developed as the existing security tools and practices won't work for such platforms. These security tools should include encryption, authentication, intrusion detection, access control, monitoring and event logging. Along with the security policies, while integrating Big Data to the cloud environment, consolidation plans should be taken into consideration.

E. Data Privacy – It's been noticed that the cultural challenge of cloud computing and big data lies in the aspects of privacy. According to many researchers, most of big data sources takes the styles of documents, messages, images, audio and video posts and also very sensitive information like an individual's location, behavior, transactions or companies tracking the employees' movement and productivity which are digitally recorded via social media implies that the most important resources for big data is relatively the social media and hence, accessing users' private information - which accounts for a major risk.

F. Different Conceptual Ideas of these Domains – The concepts of consolidation and resource pooling comprises the base concepts of Cloud Computing whereas, big data systems (such as Hadoop) are built on the shared nothing principle, where each node is self-sufficient and independent. Integrating big data with cloud computing technologies can lead businesses and educational institutes in a better direction for the future. Cloud computing has the capability to store enormous amount of data in various forms processing it at very large

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:01/January-2023 Impact Factor- 6.752 www.irjmets.com

speeds which will result in data that can guide the education and business institutes for fast development. Nonetheless, there is a huge concern regarding security and privacy issues while moving to the cloud environment which is the main reason why the educational and business institutes aren't willing to move to cloud.

IV. TECHNOLOGY

The cloud service types for Big Data analytics as a service includes infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS).

A. Infrastructure as a Service (IaaS) – To enable enterprises for allocating or buying time on shared server resources (which are often virtualized) for handling the computational and storage needs for Big Data analytics IaaS can be deployed on premise or via a cloud provider. Managing high-performance network, servers and storage resources is done by the cloud provider. Enterprises involved in Big Data do not need to maintain the hardware and software required for such performance. Hadoop, is an open-source solution, which employs distributed data storage and processing. The technologies that are used for IaaS purposes are: the Hadoop framework, or a NoSQL database, such as MongoDB, Apache Cassandra or Couch base technologies. IaaS solutions providers are Amazon Web Services, Windows Azure, Citrix Cloud Platform, Microsoft System Centre, OpenStack software, Rackspace etc.

B. Platform as a Service (PaaS) – For providing higher-level programming models and database systems PaaS is used. It provides tools and libraries for building, testing, deploying, and to run applications on cloud environment. Amazon ElasticMapReduce can be used which provides a basic Hadoop framework PaaS environment. Windows Azure's data service HDInsight brings Hadoop to the cloud environment coupled with Power Map, Power View and other Microsoft BI tools. Common PaaS offerings from Dynamo DB for NoSQL database services and AWS including Redshift for data warehousing. PaaS capabilities are also offered by Google such as: Big table, BigQuery.

C. Software as a Service (SaaS) – For delivering applications over the internet Software as a service (or SaaS) is used. SaaS offers jKool which provides business related cloud-based solutions and a real-time analysis of time-sensitive information. Concur is one of the fast growing TE SaaS company, which runs only a single instance of its software and contains preferences, history of millions of business travelers on a global scale for airlines, hotels, car rentals, taxi services, etc. Karmasphere also offers a pay-as you-go application which analyses data stored with Amazon S3 using Amazon Elastic Map Reduce.

V. SECURITY

A. Need for security in Big Data – Big data is used by too many of business but they may not have environment from perspective of the security. If any safety problem occurs to big data, it may come out with even more serious issue. Generally, companies use this technology to store data of zeta byte range regarding to the company. This potentially results in severe criticality for classification of information.to secures the data we either need to encrypt, log or use honeypot techniques. The challenge of detecting attacks and intruders, must be solved using big data style analysis. Analysis and computation of big data: Fastness is the main thing when we look up for database in the big data. However, the process may be hectic only because of the reason that it cannot traverse all related data in the whole database in a little time. While the big data is getting complex, the indices in the big data are aiming at the simple type of the data. The traditional series algorithm is inefficient for this big data.

B. Challenges of security in cloud computing – We live in the period of the big data where we can gather more information from daily life of human being. So far, researchers are unable to unify the features that are more essential to big data, many think that big data is something which we cannot process using existing technology, theory or any methods of such kind. However the world has become helpless since enormous amount of data is being generated by science, business and even society. Big data has posed many challenges to the IT industry.

C. Ways to tackle security problems

1. Encryption: Since the data in any computer will be present in a cluster, a person can easily steal the data from the system. This may become a serious problem for any company or organization to safeguard their very important data. To avoid this thing, we may go for encryption of the data. Different encryption mechanisms can

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:01/January-2023 Impact Factor- 6.752 www.irjmets.com

be used for different systems and the keys generated should be stored safely behind firewalls. By choosing this way the data of the customer is kept secure.

2. Node authentication: The node must be go from authentication whenever it joins the cluster. If the node turns out to be a malicious cluster then such nodes should not be authenticated.

3. Honeypot nodes: The honeypot nodes are disguised to be like a regular node but are a trap. It automatically traps the hackers and will not allow any harm to happen to the system or the data.

4. Access control: The various privacy and access control in the distributed environment will be a good measure of security. To prevent the information from leaking we use Linux operating system. The Linux is a feature that provides the mechanism for supporting access control security policy through the use of Linux Security modules in Linux kernels.

D. Ways to tackle security problems

Cloud computing helps in storage of data at a remote site so that we can maximize resource utilization. Therefore, it is very important for this data to protect and access should be given only to authorized people. Therefore, this amounts to secure third party publication of data that is required for data outsourcing, as well as for outside publications. In the cloud computing, the machine serves the role of a third party publisher, which stores the sensitive data in the cloud. The data needs to be protected, and the above techniques have to be used to ensure the timely maintenance of authenticity and completeness.

Big Data Security

The term big data security refers to practices and tools employed for the purpose of protecting data and analysis processes. The big data perimeter is typically divided into three categories:

- **Incoming data**—vulnerable while in transit
- **Storage data**—vulnerable while at rest
- **Output data**—processed for analysis, vulnerable in use. The goal for big data security is to prevent accidental and intentional breaches, leaks, losses, and exfiltration of huge amounts of data. Big data can be in the form of financial logs, health care data repositories, data lakes and archives, and in-progress business intelligence analyses.

VI. BIG DATA SECURITY ISSUES

Data Storage

Businesses are adopting Cloud Data Storage to move their data easily to expedite business operations. However, the risks involved are exponential with security issues. Even the slightest mistake in controlling the access of data can allow anyone to get a host of sensitive data. As a result, big tech companies embrace both on-premise and Cloud Data Storage to obtain security as well as flexibility. While mission-critical information can be stored in on-premise databases, less sensitive data is kept in the cloud for ease of use. However, to implement security policies in on-premise databases, companies require cyber security experts. Although it increases the cost of managing data in on-premise databases, companies must not take security risks for granted by storing every data in the cloud.

Fake Data

Fake Data generation poses a severe threat to businesses as it consumes time that otherwise could be spent to identify or solve other pressing issues. There is more scope for leveraging inaccurate information on a very large scale, as assessing individual data points can be a daunting task for companies. False flags for fake Data can also drive unnecessary actions that can potentially lower production or other critical processes required for running businesses. One way to avoid this is to ensure that companies should be critical of the data they are working on for enhancing business processes. An ideal approach is to validate the data sources by periodic assessments and evaluate Machine Learning models with diverse test datasets to find anomalies.

Data Privacy

Data Privacy is a big challenge in this digital world. It aims to safeguard personal or sensitive information from cyber- attacks, breaches, and intentional or unintentional data loss. Businesses must follow stricter Data Privacy principles with the help of access management services in the cloud, including very rigid privacy

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:01/January-2023 Impact Factor- 6.752 www.irjmets.com

compliance, to strengthen Data Protection. It is best to follow a few rules alongside implementing one or more Data Security technologies. The general rules know your data, having more grip over your data stores and backup, safeguarding your network against unauthorized access, conducting regular risk assessments, and training the users regularly about Data Privacy and Data Security.

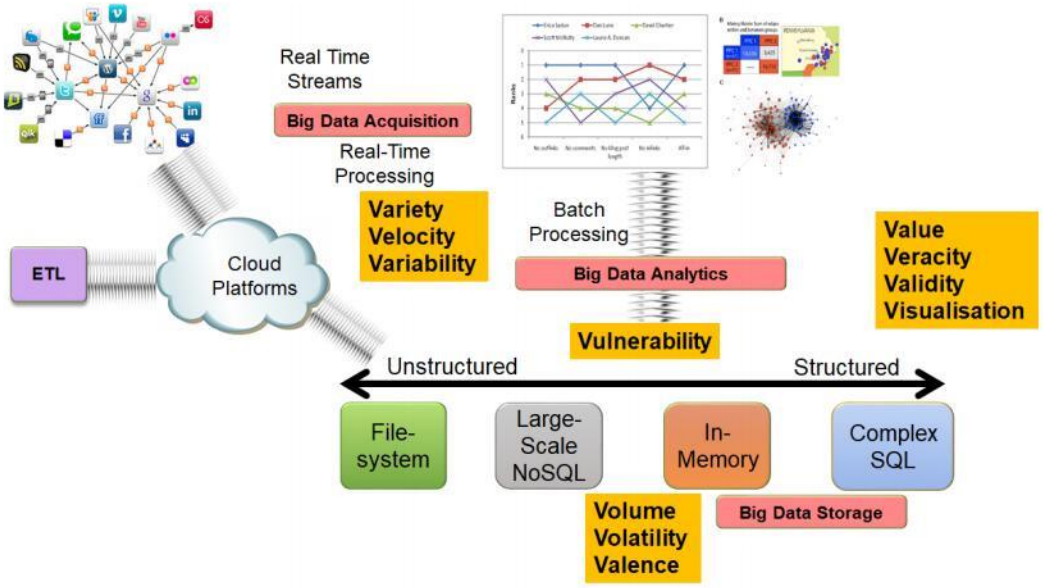

Figure 1

Data Management

A security breach can have crushing consequences on businesses, including the vulnerability of critical business information to a completely compromised database. Deploying highly secured databases is vital to ensure data security at all levels. A superior Database Management System comes with various access controls. While it is advisable to follow rigid and rigorous physical security practices, it is even more essential to follow extensive software-based security measures to safeguard data storage. A few methods to effectively achieve this goal are—practicing data encryption, data segmenting and partitioning, securing on-the-move, and implementing a trusted server. Besides, a few security tools can integrate with databases to automatically monitor data sharing and notify businesses when data has been compromised.

Data Access Control

Controlling which data users can view or edit enables companies to ensure not only data integrity but also preserves its privacy. But managing access control is not straightforward, especially in larger companies that have thousands of employees. However, a shift from on-premise solutions to cloud-based services has simplified the process of working with Identity Access Management (IAM). IAM does the job of controlling data flow via identification, authentication, and authorization. Following relevant ISO standards is a good starting place to ensure organizations meet the best IAM practices.

Data Poisoning

Today, there are several Machine Learning solutions like chatbots that are trained on a colossal amount of data. The advantages of such solutions are that they keep on improving as users interact. However, this leads to Data Poisoning, a technique to attack Machine Learning models' training data. It can be considered as an integrity attack as the tampered training data can affect the model's ability to provide correct predictions. The results can be catastrophic, ranging from logic corruption to Data Manipulation and Data Injection. The best way to beat the evasion is through outlier detection, wherein the injected elements in the training pool can get separated from the existing data distribution.

Employee Theft

Advance data culture has allowed every employee to hold a certain level of critical business information. While it boosts data democratization, the risk of employee leaking sensitive information, intentionally or

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:01/January-2023 Impact Factor- 6.752 www.irjmets.com

unintentionally, is high. Employee Theft is prevalent not only in big tech companies but also in startups. To avoid Employee Theft, companies have to implement legal policies along with securing the network with a virtual private network. In addition, companies can use a Desktop as a Service (DaaS) to eliminate the functionalities of data stored in local drives.

Hadoop

Hadoop is a free, Java-based programming frame work that aids in 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 reduces 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, fault tolerant and flexible. 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 namely Map Reduce and Hadoop Distributed File System (HDFS).

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 first divides the data into individual chunks which in turn are processed by Map jobs in parallel. The outputs of the maps sorted by the framework are then input to the reduce tasks. Usually 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.

Hadoop Distributed File System (HDFS)

HDFS is a file system that stretches over 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.

VII. ADVANTAGES

The big data allows an individual to analyze the threats he/she faces internally by noosing onto the entire data landscape over the company using the rich set of tools that the software supporting the big data provides. This is an important advantage of big data since it allows the user to make the data safe and secure. The speed, capacity and scalability of cloud storage provide a mere advantage for the company and organization. Big data even allows the end users to visualize the data and companies can find new business opportunities. Data analytics is one more notable advantage of the big data where in which the individual is allowed to personalize the content or to look and feel the real time websites.

VIII. SECURING BIG DATA ANALYTICS

Standard big data security best practices include:

- **Encryption**—the process of encoding information in a way that renders it useless for attackers. After the data is encrypted, the system generates keys. Only the right key can decrypt the data, and the system rotates keys. This security technique relies on the supposition that attackers won't be able to re-create the correct decryption key.
- **Tokenization**—the data is sent to a third-party mediator, which sends a token to the website. The tokenization system saves the information in a vault, and the website is not storing any financial information. This security technique relies on the supposition that attackers won't gain access to the tokenization system.
- **Next-Generation Firewall (NGFW)**—according to Garner, this is a "deep-packet inspection firewall". NGFW moves beyond stateful port/protocol inspection and blocking. NGFW is dynamic, and offers features such as application inspection, intrusion prevention, and cloud threat intelligence.
- For endpoint protection, organizations can make use of:
- Endpoint Protection Platforms (EPPs)—a passive layer of defense against known threats. Common EPP solutions make use of antivirus and Next-Generation Antivirus (NGAV), encryption, DLP, and NGFW. EPPs typically employ defense techniques such as signature matching, sandboxing, blacklisting and whitelisting.

- Endpoint Detection and Response (EDR)—an active layer of defense against endpoint threats. EDR solutions usually apply data collection, detection, and analysis techniques. The key goals of EDR are providing real-time threat intelligence, alerts, and forensics. Some EDR solutions provide automated responses and trace back mechanisms. If an EPP solution includes EDR tools, it gains active defense capabilities. The goal is to ensure that all points in the network are covered, so as to eliminate unauthorized access to your data.

IX. CONCLUSION

The majority of security attacks targeted data. In fact, by looking at security data and patterns, a safe deduction would be that the objective of attacks is almost always data. As attackers gain more advanced and sophisticated tools and techniques, we'll continue to see an increase in data breaches. In the digital sphere, data is a valuable commodity that can unlock financial information and credentials. Attackers can ransom data, sell it to the highest bidder, use it to launch another attack, delete it to damage the organization, and manipulate it for the purpose of spreading disinformation. Big data analysis repositories are especially vulnerable, and deserve a well-rounded security approach that covers all types of network points and users.

X. REFERENCES

[1] Hariharan, U. & Kotteswaran, Rajkumar & Pathak, Nilotpal. (2020). The Convergence of IoT with Big Data and Cloud Computing. 10.1201/9781003054115-1.

[2] Martian A.,Vulpe A.,Suciu G., Cranciunescu R.(2015) Big Data, Internet of Things and Cloud Convergence- An Architecture for Secure E-Health Applications. Article in Journal of Medical Systems. DOI:10.1007/s10916-015-0327-y

[3] Agrawal, Divyakant & Das, Sudipto & Abbadi, Amr. (2011). Big Data and Cloud Computing: Current State and Future Opportunities. ACM International Conference Proceeding Series. 530-533. 10.1145/1951365.1951432.

[4] Sengupta S., Kaulgud V., Sharma V. (2011) "Cloud Computing Security- Trends and Research Directions" in IEEE Computer Society,Pg 524-531. DOI:10.1109/SERVICES.2011.20

[5] Brevini, Benedetta. (2015). Book Review: To the Cloud: Big Data in a Turbulent World. Media, Culture & Society. 37. 1111-1113. 10.1177/0163443715596318

[6] Yadav S., Sohal A. (2017) "Review Paper on Big Data Analytics in Cloud Computing" in International Journal of Computer Trends and Technology (IJCTT) V49(3):156-160, July 2017. ISSN:2231-2803.