International Journal of Engineering Technology and Management Sciences[IJETMS]

Website: ijetms.in Issue:4, Volume No.4, July-2020 DOI: 10.46647/ijetms.2020.v04i04.006

Security Recommendations And Security Issues Associated With Big Data In Cloud Computing

Sambaiah.G[1], Hyma.Birudaraju[2]

[1]*Asst.Professor,CSE,Guru Nanak Institutions Technical Campus, Hyderabad,Telangana,India.*
[2]*Asst.Professor,CSE,Guru Nanak Institutions Technical Campus, Hyderabad,Telangana,India.*

[1]ganjisamba@gmail.com
[2]hymaomkaram@gmail.com

***Abstract*— In this paper, we discuss security surveillance for Big data, cloud computing, Map Reduce and Hadoop environment. The focal point is on security surveillance in cloud computing that are corresponding with big data. Big data applications are a admirable beneficent to organizations, business, companies and many large scale and small scale industries.We also talk about various attainable solutions for the issues in Hadoop and cloud computing security. Cloud computing security is being blossom at a rapid pace which incorporate with computer security, network security, information security, and data privacy.Cloud computing plays a very crucial role in keep safe data, applications and the related infrastructure with the help of technologies, policies,controls , and big data tools. Moreover, Cloud computing, big data and its applications, advantages are likely to illustrated the most hopeful new boundaries in science.**

***Keywords*—Big Data, Cloud Computing, Map Reduce, Hadoop, HDFS (Hadoop Distributed File System)**

I. INTRODUCTION

With the purpose of identify patterns and to analyze complex data it is very essential to securely store,manage and share large amounts of complex data. Cloud accompanies an explicit security challenge, i.e. the data administrator might not have any command of where the data is located. The reason beyond this administrative issue is that if one wants to get the advantages of cloud computing, he/she must also employs the allocation of resources and also the scheduling given by the controls. Hence it is required to protect the data in between the undependable processes. By considering that cloud involves wide spread complexity, we believe that rather than providing a exhaustive solution to securing the cloud, it would be ideal to make notable decoration in securing the cloud that will fundamentally provide us with a secure cloud.

MapReduce processes exceptionally huge amounts of data without being effected by traditional bottlenecks like network bandwidth by taking advantage of this data proximity. Hadoop, which is an open-source implementation of Google MapReduce, including a distributed file system, is responsible to the application programmer the abstraction of the map and the reduce. With Hadoop it is more straight forward for organizations to get a control on the large volumes of data being generated each and every day, but in spite of that can also create problems related to data access, monitoring, security, business continuity and high availability.

I.1 Cloud Computing

Cloud computing is a protector-ship, term used to refer to Internet based development and services. The cloud is a symbolism for the data on internet. A number of characteristics infrastructure, applications services and define cloud data:
Remotely hosted: Data or Services are hosted on someone else’s infrastructure. Ubiquitous: Services or data are accessible from anywhere.
Com modified: The result is a utility computing model similar to traditional that of traditional utilities, like gas and electricity.
Cloud computing is the provision of delivery of computing services over the Internet. Cloud services allow individuals and businesses to utilize software and hardware that are governed by third parties at remote locations. Examples of cloud services include web mail, online file storage, social networking sites, and online business applications. The cloud computing model provides access to information and computer resources from anywhere that a network connection is available. Cloud computing makes provision for a shared pool of resources, including data storage space, networks, computer processing power, and specialized corporate and user applications.

I.2. Architecture

A fundamental information about the architecture is given in this chapter, together with the explanations of relevant terms such as Frond/Back end or Middleware,virtualization. Virtualization is an approach of develop a software based and virtual representation of data model, such as virtual servers, applications, storage and networks , it is best described as primarily designating one computer to do the job of more than one computers by sharing the resources of that single computer across different environments. Virtual desktops and Virtual servers allows us to host multiple operating systems and multiple applications in restricted local systems and in remote systems located at different locations, freeing our business from physical and geographical limitations.

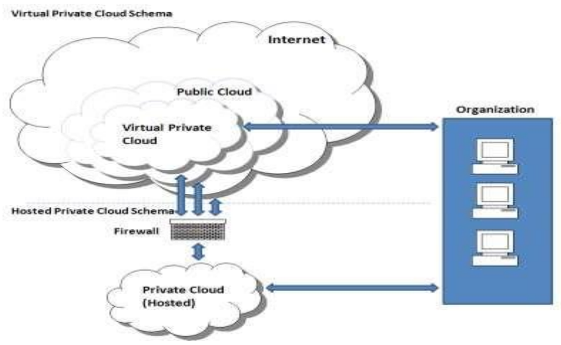
The Cloud Computing architecture can be split into two sections, the back end and the front end, connected together through a network, commonly Internet. The Front End consists of the client's computer and the application needed to access the cloud computing system. All cloud computing systems may not have the same user interface. Services like Web-based e-mail programs supports existing Web browsers like Internet Explorer or Firefox. Other systems have special applications that provide network access to clients.

The Back End of the system is depicted by various computers, servers and data storage systems that create the "cloud" of computing services. Fundamentally, Cloud Computing system could involves any program, from data processing to video games and each application may have its own server.

A central server controls the system, monitoring traffic and client demands to guarantee everything runs accordingly. It evolves a set of rules called protocols and uses a special kind of software called Middleware. The Middleware allows the networked computers to provide communication with each other.

Based on the range and area of cloud it is of two two types-Public cloud and Private cloud. Public Cloud is also known as external cloud and it is a model where services are available from a service provider over the Internet, such as applications and storage. These Public Cloud Services are available for free, as well as pay per usage or other monetized models. Private Cloud are also termed as Internal Cloud or Corporate Cloud and its computing architecture presuming that hosted services to a specified number of people behind a company's defensive firewall and it sometimes attracts criticism as firms still have to buy, develop, and manage some resources and consequently do not benefit from lower up-front capital costs, less hands-on management and the core concept of Cloud Computing.

FIGURE-1: PUBLIC AND PRIVATE CLOUDS



I.3What Is Big Data

Big data specifies the collection and successive analysis of any considerably huge collection of data that may contain (user data, machine data, and sensor data). When analyzed appropriately, big data can deliver new business perceptions, open new markets, and create competitive benefits. Compared with the structured data in business applications, the big data consists of following three major attributes:

•Variety—Makes available beyond structured data and provides semi-structure or unstructured data of all varieties, such as text, audio, video, log files, click streams and more.

•Volume—Appears in one size that is large. Organizations are chock full of data, easily acquiring hundreds of terabytes and petabytes of information.

•Velocity—Consistently must be analyzed in real time as it is streamed to an organization to fully take an advantage of the data's business value.

We can recognize two kinds of cost-saving efficiency enhancements:

✓ Operational efficiency: These progresses are measured by reduced costs to get the same or better results. With BD&A, this is due to more efficient methods of data integration, management, analysis, and delivery.

✓ Business processes: These gains are measured by the impact of new, better ways of conducting business, adding improvements to commercial transactions, acceptable management of communities, and suitable distribution of social, health care, and educational services.

I.4. Big Data Use Cases

There are many examples of big data use cases in virtually every industry conceivable. Some businesses have been more open minded of the technologies and more faster to integrate big data analytic into their everyday business than others. It is an evident that organizations incorporating this technology not only will see substantial first-mover advantages but will be considerably more graceful and advancement in the solutions and flexibility of their offerings.

Use case examples of big data solutions includes

•Financial services providers are employing big data analysis infrastructure to enhance their analysis of customers to help designate eligibility for equity capital, mortgage, insurance, or credit.

•Health care centers are managing and sharing patient electronic health registers and documentation from multiple sources—description, treatments, and demographics—and across many of practitioners. In addition, pharmaceutical companies and regulatory agencies are preparing big data solutions to track the specific drug efficacy and provide more efficient and shorter drug development processes.

•Telecommunications and utilities are using big data computations to analyze user behaviors and demand patterns for a preferable and more efficient power grid. They are also storing and analyzing environmental sensor data to provide an awareness of infrastructure weaknesses and provide refined risk management intelligence.

•Airlines and trucking companies also are using big data to track the fuel usage and traffic patterns across their fleets in real-time to improve efficiency and save costs.

TABLE-1: BIG DATA VERSUS TRADITIONAL DATA TYPES

Components	Traditional data	Bigdata
Architecture	Centralized	Distributed
Data volume	Terabytes	Petabytes to exabytes
Data type	Structured or transnational	Unstructured or semi-structured
Data relationships	Known relationship	Complex/unknown relationships
Data model	Fixed schema	Schema-less

1.5 Big Data Technologies (Hadoop)

The chief motive behind an implementation of big data is the software—both infrastructure and analytics. Foremost technology in the infrastructure is Hadoop. Hadoop is the big data management software infrastructure used to manage, distribute, catalog, and query data across multiple, horizontally mounted server nodes. Yahoo! Developed it based on an open source implementation of the data query infrastructure which is originated at Google, called MapReduce. It has a number of commercially sustained distributions from companies such as MapR Technologies and Cloudera. Hadoop is a framework for storing, processing, and analyzing vast amounts of distributed unstructured data. As a distributed file storage subsystem, Hadoop Distributed File System (HDFS) was designed to administrate petabytes and exabytes of data distributed over multiple nodes in concurrently.

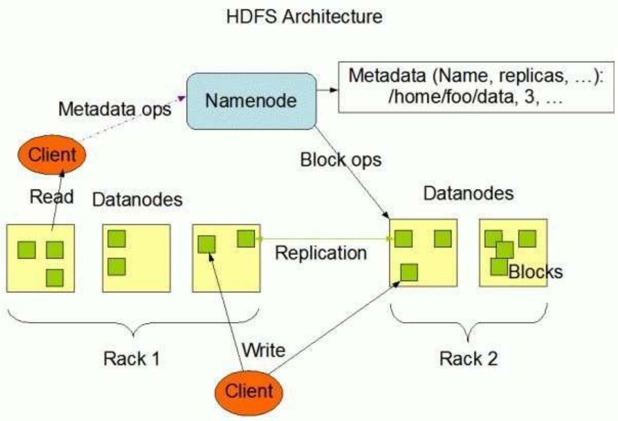
Hadoop Ecosystem

The Hadoop platform has two key services: a trustworthy, distributed file system called Hadoop Distributed File System (HDFS) and the hopped-up parallel data processing engine called Hadoop MapReduce. The compound mixture of HDFS and MapReduce gives a software framework for processing huge amounts of data in parallel on large clusters of commodity hardware in a reliable, fault tolerance manner. Hadoop is a conventional data processing framework which is designed to execute queries and also batch read operations against vast data sets those can scale from tens of terabytes to petabytes in size.

The Hadoop Cluster

Hadoop, which involves a distributed file system known as Hadoop Distributed File System (HDFS) and MapReduce, is a perspicuous big data technology that contains a extensible file system infrastructure and allows for the horizontal scale of data for very fast query, access, and data management.

FIGURE-2: HDFS ARCHITECTURE



1.6 MapReduce

MapReduce is a tremendously scalable, parallel processing framework that works along with HDFS. With MapReduce and Hadoop, compute is accomplished at the location of the data, rather than moving the data to the compute location; data storage and computation intersects on the same physical nodes in the cluster. MapReduce performs seies of action on extremely huge amounts of data without being forced by traditional obstructions like network bandwidth by taking advantage of this data proximity.

Key Features of MapReduce:

MapReduce is a dominant framework for processing distributed, large sets of structured or unstructured data on a Hadoop cluster. The key feature of MapReduce is its capable to perform processing across an whole cluster of nodes, with each node processing its local data.

Scale-out Architecture - Add servers to improve processing power

Security & Authentication - Works with HDFS and HBase security to ensure that only approved users can operate over the data in the system

Resource Manager - Utilizes data locality and server resources to determine perfect operations

Optimized Scheduling - Finishes the jobs according to prioritization

Flexibility – Procedures can be written in virtually using any programming language

Resiliency & High Availability - Multiple job and task trackers make sure that jobs fail independently and restart automatically

Big Data Analytics for Security:

This section expresses how Big Data is changing the analytics landscape. In particular, Big Data analytics can be taken advantage to improve information security and situational awareness. For example, Big Data analytics could be engaged to analyze network traffic, financial transactions, , and log files to identify abnormal situations and suspicious activities, and to connect multiple sources of information into a logical view. Data driven information security anticipates to handle fraud detection and anomaly-based intrusion detection systems. Fraud detection is one of the most visual evidence for Big Data analytics. Credit card companies have organized fraud detection for decades. In spite of, the custom-built infrastructure to mine Big Data for fraud detection was not efficient to adapt for other fraud detection uses. Off-the-shelf Big Data tools and techniques are now conveying attention to analytics for fraud detection in insurance, health care, and other fields. Based around the idea of data analytics for intrusion detection, the following expansion is anticipated:

- ✓ First generation: Intrusion detection systems – Security architects recognized the need for layered security (e.g., reactive security and breach response) because a system with 100% protective security is not possible.
- ✓ Second generation: Security information and event management (SIEM) – Monitoring alerts from different intrusion detection sensors and rules was a big burden in enterprise settings. SIEM systems composite and filter alarms from many sources and gives applicative information to security analysts.
- ✓ Third generation: Big Data analytics in security 2nd generation, SIEM Big Data tools have the potential to produce a significant advance in practicable security intelligence by minimizing the time for consolidating, correlating, and contextualizing various security event information, and also for correlating long-term historical data for forensic justifications.

II. MOTIVATION AND RELATED WORK

II.1.Motivation

In addition to the increasing reputation of the Could Computing environments, the security issues introduced through reconstruction of of this technology are also increasing. However, Cloud Computing offers many advantages and drawbacks, since it is unprotected from attacks. Attackers are frequently trying to discover loopholes to attack the cloud computing environment. The traditional security mechanisms which are used are reviewed because of these cloud computing spreads. Capability to visualize, inspect and control the network links and ports is required to make sure the security. Hence there is a necessity to consign in understanding the loop holes, challenges and components liable to attacks with respect to cloud computing, and progress with a platform and infrastructure which is less vulnerable to attacks.

II.2 Related Work

Hadoop is a cloud computing framework and it is Java based distributed system, is a recent innovative framework in the market. Since Hadoop is new and quiet being developed to add more features, there are many security issues which need to be addressed. Researchers have discovered some of the issues and started working on this. Some of the noteworthy outcomes, which is related to our domain and helped us to explore, are presented below. The World Wide Web consortium has proposed the importance of SPARQL which can be useful in diverse data sources. Later on, the idea of secured query was planned in order to improve privacy in privacy/utility tradeoff. In existence, Jelena, of the USC Information Science Institute, has stated that the queries can be processed as per the policy of the provider, rather than all query processing. Bertino et al published a paper that describes access control for XML Documents. In the paper, cryptography and digital signature technique are described, and techniques of access control to XML data document is highlighted for secured environment. Subsequently, he published another paper on authentic third party XML document distribution which extracted another trusted layer of security to the paradigm. Kevin Hamlen and et al explained that the data can be stored in a database with encrypted rather than plain text. The advantage of storing encrypted data is that even though intruder or hacker can get into the database, he or she can't understand the actual data. But, the disadvantage is that encryption involves a lot of overhead. Instead of processing the plain text in to cipher text, most of the operation will take place in cryptographic form. Hence the approach of processing cryptographic data added extra functionality to security layer.

IBM researchers also taught that the query processing should take place in a secured environment. Then, the use of Kerberos has been hugely effective. Kerberos is nothing but a model of authentication that has been developed at MIT. Kerberos uses an encryption technology accompanying with a trusted third party, an arbitrator, having capability to perform a secure authentication on an open network. To be more specific, it uses cryptographic tickets to keep clear of transmitting plain text passwords over the wire. Kerberos is worked based upon Needham-Schroeder protocol. Airavat has shown us some considerable advancement security in the Map Reduce environment. Roy and et al have used the access control mechanism along with differential privacy mechanisms. They have convinced the mathematical bound potential privacy violation which reduces information leak beyond data provider's policy. The above works have influenced me, and I'm analyzing various approaches to make the cloud environment more secure for data transfer and computation.

III. ISSUES AND CHALLENGES

Cloud computing comes with numerous security issues because it encompasses many technologies including databases, networks, virtualization, operating systems, transaction management, resource scheduling, concurrency control, load balancing and memory management. Hence, security issues of these systems and technologies are appropriate to cloud computing. For example,it is very important for the network which connects the systems in a cloud to be secure. Also, virtualization paradigm in cloud computing resultants in several security concerns

There are a several security threats associated with cloud computing that must be adequately addressed

Loss of governance. In a public cloud deployment, customers yield control to the cloud provider over a number of issues that may influence on security. Yet cloud service agreements may not offer a commitment to solve such issues on the part of the cloud provider service, thus there will leftover gaps in security defenses.

Responsibility ambiguity. Responsibility over facets of security may be split between the provider and the customer, with the potential for vital parts of the defenses to be left unprotected if there is a failure to allocate duty clearly. This split is most probably vary depending on the cloud computing model used (e.g., IaaS vs. SaaS).

Authentication and Authorization. The phenomenon that sensitive cloud resources are accessed from anywhere on the Internet raises the need to establish with assurance the identity of a user, especially if users now include customers, employees, contractors and partners . Strong authentication and authorization becomes a critical deal with.

Isolation failure. Multi-proprietorship and shared resources are describing characteristics of public cloud computing. This risk category includes the failure of methodologies separating the usage of storage, memory, routing and even reputation between tenants (e.g. so-called guest-hopping attacks).

Compliance and legal risks. The cloud customer's investment in obtaining certification may be lost, if the cloud provider cannot produce evidence of their own approval with the relevant requirements, or does not allowed audits by the cloud customer. The customer must investigate that the cloud provider has appropriate certifications in place.Conclusively, data mining techniques can be used in the security attacks detection in clouds. The challenges of security in cloud computing environments can be classified into data level, network level, user authentication level, and generic issues.

Network level: The challenges that can be categorized under a network level deals with network security and network protocols, such as distributed data, distributed nodes, Inter node communication.

Authentication level: The challenges that can be categorized under user authentication level deals with encryption/decryption techniques, authentication of applications, authentication methods such as administrative rights for nodes, and nodes, and logging.

Data level :The challenges under data level can be categorized are deals with data integrity and availability such as data safe guard and distributed data.

Generic types: The challenges which are categorized under general level are traditional security tools, and use of various technologies.

III.1 Distributed Nodes

Distributed nodes are an architectural issue. The computation is accomplished in any set of nodes. Basically, data is processed in those nodes which have the essential resources. Since it can turn up anywhere across the clusters, it is very difficult to identify the exact location of computation. Because of this it is very difficult to make sure the security of the place where computation is done.

III.2 Distributed Data

In order to moderate parallel computation,a large data set can be stored in many pieces over many machines. Also, redundant copies of data are made to provide guarantee for data reliability. In case a particular piece of data set is corrupted, the data can be recovered from its copies. In the cloud environment, it is extremely difficult to find exactly where pieces of data set file are stored. Also, these pieces of data are copied to another machines based on accessibility and administrative operations. In traditional centralized data security system, interpretative data is enclosed around various security tools. This cannot be applicable for cloud environments since all related data are not presented in one place and it changes.

III.3 Inter-node Communication

Many of Hadoop distributions uses RPC over TCP/IP for user data/operational data transfer between nodes. This occurs over a network, distributed around web world includes the wireless and wired networks. As a result, anyone can trap and they can change the inter node communication for breaking down the system.

III.4 Data Protection

Widespread cloud environments like Hadoop store the data as it is without encryption to enhance efficiency. If a hacker can access a set of nodes, there is no way to stop him to theft the critical data present in those machines.

III.5 Administrative Rights for Nodes

A node has administrative power and can access any data. This uncontrolled access to any data is very dangerous as a malicious node can theft or manipulate important user data.

III.6 Authentication of Applications and Nodes

Nodes can add clusters to improve the parallel operations. In case of no authentication, third part nodes can add clusters to theft user data or disturb the operations of the cluster.

III.7 Logging

In the unavailability of logging in a cloud environment, no activity can be recorded which modify or delete user data. No information is recorded like which nodes have added in to cluster, which Map Reduce jobs have run and also what changes are made because of these jobs. Without recording these logs, it is very difficult to find if someone has broken the cluster if any, authorized altering of data is done which needs to be recovered. Also, in the absence of logs, internal users can do nasty data manipulations without getting recognized.

III.8 Traditional Security Tools

Traditional security tools are designed for traditional systems where scalability is not much important as cloud environment. Because of this, traditional security tools which are evolved over years cannot be directly applied to this distributed form of cloud computing and these tools do not measure as well as the cloud scales.

III.9 Use of Different Technologies

Cloud consists of various technologies which has a lots of interacting complex components. These components include database, network, computing power, and many other stuff. Since the wide use of technologies, a small security weakness in one component can damage the whole system. Because of this diversity, providing security in the cloud is very challenging.

IV. THE PROPOSED APPROACHES

We have various security measures which would enhance the security of cloud computing environment. Since the cloud environment is a combination of many different technologies, We are proposing various solutions which collectively will make the environment well secure. The proposed solutions motivate the use of multiple technologies/ tools to reduce the security problems which are specified in previous sections. Security recommendations are designed such that they do not decrease the effectiveness and scaling of cloud systems.

The following security steps should be taken to ensure the security in a cloud environment.

IV.1 File Encryption

Since the data is present in the machines in a cluster, a hacker can theft all the critical information. Because of this, all the data stored should be encrypted. Encryption means converting the readable data into unreadable format. To provide more security for data, different encryption keys should be used on different machines and the key information should be stored centrally beyond robust firewalls. This way, even if a hacker is able to get the data, he cannot do understand meaningful information from it and misuse it. User data will be remain stored securely in an encrypted manner.

IV. 2 Network Encryption

All the network communication should be encrypted as per the specified industry principles. The RPC procedure calls which come around should happen over SSL so that even if a hacker can tap into network communication nodes of data sets, he cannot extract the actual useful information or manipulate packets.

IV.3 Logging

All the map reduce jobs which alters the data should be recorded in to log file. Also, the information of users, who is responsible for those jobs should be recorded. These logs should be regularly audited to find malicious activities are performed or any malicious user is altered the data in the nodes if any .

IV.4 Software Format and Node Maintenance

Nodes which run the software applications should be formatted frequently to avoid any virus present. All the application softwares and Hadoop software should be updated to keep the system more secure.

IV.5 Nodes Authentication

Whenever a node added into a cluster, it should be authenticated. In case of a suspicious node, it should not be permitted to join the cluster. Authentication methodologies like Kerberos can be used to recognize the authorized nodes from malicious ones.

IV.6 Rigorous System Testing of Map Reduce Jobs

After a developer invent a map reduce job, it is mandatory to comprehensively tested in a distributed environment instead of a single machine to make sure the robustness and stability of the job.

IV.7 Honeypot Nodes

Honey pot nodes should be made available in the cluster, which appear look like a regular node but is a trap. These honeypots can trap the hackers, malicious users and eliminate by taken necessary actions against them.

IV.8 Layered Framework for Assuring Cloud

A layered framework for securing cloud computing as shown in Figure (1) consists of the secure secure cloud storage layer, secure virtual network monitor layer secure cloud data layer, and the virtual machine layer. Cross cutting services are carried out by the policy layer, the reliability layer, the cloud monitoring layer and the risk analysis layer.

IV.9 Cloud Security Guidance

As customers transition their applications and data to the cloud, it is very difficult for them to maintain, or preferably overhead, the level of security they had in their traditional IT environment.

This section provides a conventional series of steps for cloud customers to assess and manage the security of their use of cloud services, with the aim of mitigating risk and delivering an signifcant level of support. The following steps will be discussed in detail below:

a) Make sure effective governance, risk and adherence processes exist

b) Business processes and Audit operational

c) Administrate people, roles and identities

d) Guarantees that proper protection of data and information

e) Execute privacy policies

f) Estimate the security provisions for cloud applications

g) Ensure cloud networks and connections are secure

h) Analyze security controls on physical infrastructure and facilities

i) Manage security terms in the cloud service agreement

j) Understand the security requirements of the exit process.

V. CONCLUSION

Cloud environment is extensively using in industry and research tendencies; thus security is an major aspect for organizations which are running on these cloud environments. Using our proposed approaches, cloud environments can be more secured for complex business operations. Cloud computing is obviously one of the vital enticing technology areas of the current times due, at least in part to its cost-efficiency and flexibility. In spite of, the improvement in activity and interest, there are considerable, persistent concerns about cloud computing that are blocking the momentum and will eventually compromise the vision of cloud computing as a new IT procurement model. Regardless of the advertised business and technical advantages of cloud computing, many potential cloud users have yet to join the cloud, and those major corporations that are cloud users are for the most part placing only their less sensitive data in the cloud. Insufficiency of control is unambiguousness in the cloud implementation – somewhat contradictory to the original promise of cloud computing in which cloud implementation is not applicable. Transparency is mandatory for regulatory reasons and to better concern over the potential for data beak-ups. Because of today’s detected lack of control, larger companies are testing the waters with smaller projects and less sensitive data. In other words, the prospective development of the cloud is not yet being realized, Log file system is one of the solution to keep data secure where we can record the every activity and recognize malicious activities in easy way.

References

[1] A, Katal, Wazid M, and Goudar R.H. "Big data: Issues, challenges, tools and Good practices.". Noida:2013, pp. 404 – 409, 8-10 Aug. 2013.

[2] N, Gonzalez, Miers C, Redigolo F, Carvalho T, Simplicio M, de Sousa G.T, and Pourzandi M. *"AQuantitative Analysis of Current Security Concerns and Solutions for Cloud Computing."*. Athens:2011., pp 231 – 238, Nov. 29 2011- Dec. 1 2011

[3] Hao, Chen, and Ying Qiao. *"Research of Cloud Computing based on the Hadoop platform."*. Chengdu, China: 2011, pp. 181 – 184, 21-23 Oct 2011.

[4] Y, Amanatullah, Ipung H.P., Juliandri A, and Lim C. *"Toward cloud computing reference architecture: Cloud service management perspective."*. Jakarta: 2013, pp. 1-4, 13-14 Jun. 2013.

[5] Ren, Yulong, and Wen Tang. *"A SERVICE INTEGRITY ASSURANCE FRAMEWORK FOR CLOUDCOMPUTING BASED ON MAPREDUCE."*Proceedings of IEEE CCIS2012. Hangzhou: 2012, pp 240 – 244, Oct. 30 2012-Nov. 1 2012

[6] Lu, Huang, Ting-tin Hu, and Hai-shan Chen. *"Research on Hadoop Cloud Computing Model and its Applications."*. Hangzhou, China: 2012, pp. 59 – 63, 21-24 Oct. 2012.

[7] Mahesh, Bhasutkar, Maninti Venkateswarlu, and M. Raghavendra. "End-to-end congestion control techniques for Router." 2011 International Conference on Communication Systems and Network Technologies. IEEE, 2011.

[8] Mahesh, B., and K. Shyam Sunder Reddy. "Router Aided Congestion Control Techniques." Second International Conference on Information Systems and Technology.

[9] Mahesh, B. "Dynamic Update and Public Auditing with Dispute Arbitration for Cloud Data." Journal of Advanced Database Management & Systems 4.3 (2017): 14-19.

[10] Mahesh, B., et al. "A Review on Data Deduplication Techniques in Cloud." Embedded Systems and Artificial Intelligence. Springer, Singapore, 2020. 825-833. [11] Xu-bin, LI , JIANG Wen-rui, JIANG Yi,

ZOU Quan *"Hadoop Applications in Bioinformatics."* Open Cirrus Summit (OCS), 2012 Seventh, Beijing, Jun 19-20, 2012, pp. 48 -52.

[12] Bertino, Elisa, Silvana Castano, Elena Ferrari, and Marco Mesiti. *"Specifying and enforcing access control policies for XML document sources."* pp 139-151.

[13] E, Bertino, Carminati B, Ferrari E, Gupta A , and Thuraisingham B. "Selective and Authentic Third- Party Distribution of XML Documents."2004, pp. 1263 - 1278.

[14] Kilzer, Ann, Emmett Witchel, Indrajit Roy, Vitaly Shmatikov, and Srinath T.V. Setty. *"Airavat: Security and Privacy for MapReduce."*

[15] *"Securing Big Data: Security Recommendations for Hadoop and NoSQL Environments."*Securosis blog, version 1.0 (2012)