

Analysis of Security and Privacy Risk in Union of Big Data and Cloud Computing

¹Reetu Singh,

¹Assistant Professor (Guest Faculty)

¹Department of Computer Science, Institute of Engineering and Technology

¹Dr. Shakuntala Misra National Rehabilitation University, Lucknow, India

²Km Manisha Rai

²Student

²School of Computing Science and Engineering

²Galgotias University, Greater Noida, India

Abstract: Big data is a group of vast amount of unstructured datasets and data size but conventional management process can't handle the big data storehouse. With the increasing amount of knowledge, the request for massive data storage increases. By placing the information within the cloud that unstructured data is out there to anyone from anywhere. Cloud computing is an arising service-oriented framework for executing parallel and distributed computing over big data storehouse. Because of the increasing advantages of cloud computing in terms of storage, cost, and scalability and it's also concentrated by each data providers and organization for out sourcing their data from the native servers to remote cloud computing servers, which has become a standard drift. This elevate major worries about data security for cloud data storage and therefore the eagerness in provisions of extemporize the info privacy and consistency, which is give rise to the main barrier towards the acquiring of clouds services. In order to deal with this difficulty, this survey explores the problems and challenges towards big data storehouse, privacy issues, data protection and data accessing, managing the shared data within the cloud. Many organizations demand productive solutions for the store and analyze vast amounts of data. Cloud computing as an authorize agent gives flexible resources and significant economic benefits as lower operational expenses. This worldview elevates a broad scope of security and protection matter that must be taken into deliberation. Multi-tenure, loss of control, and trust are key problems in the cloud computing Environment. This paper reviews the present technologies and a vast arrangement of both prior and state-of-the-projects on cloud protection and security. We resolve the existing research as indicated by the cloud reference design organization, physical resource, resource control, and cloud administration management layers, in addition to judging the ongoing development for upgrading the Apache Hadoop privacy as one of the most important deployed big data framework. We additionally diagram the frontier research on security-preserving information-intensive applications in cloud computing such as security magnifies solutions and security threat modeling.

Index Terms: Big Data, Cloud Computing, Security and Privacy Issue, Virtualization.

Introduction

In present time data formation rate get increased very rapidly because most of the organization collects their data from various sources like from IOT devices, machines, and linked systems. This scenario leads the formation of huge amount of data and to use this data in better way organizations used big data technology to analyses these data and further generate new pattern from previous data. But when data amount get increased big data is all alone not sufficient to provide meliorate infrastructure to data. Here big data meet with cloud computing to provide improve privacy and security, better analysis, abstract infrastructure, cut the cost of data maintenance. Coalition of cloud computing and big data is profitable for organizations to accomplished their goals. The coalition of cloud computing and big data increase organization proficiency. Big data contain semi-structured, structured, and unstructured form of data which is label with implicit information and gathered from different sources. Here incessant increasing data generate issue of privacy and security, multiple occupancy, culpability that can give a result of data stealing, data disruption and damage of data. As a result to avoid these issues cloud computing with big data manages perceptive data, organize technical aspect and build data infrastructure.

To handle the impressible data cloud computing is work as a medium which operate huge amount of data which is gathered from various field of area and organization. The real world example of big data manage in cloud computing environment will be health sector. In health sector huge quantity of data is collect from different-different sources like from hospital document, hospital equipment's. This is attached to the internet of things, medical document of patients for cased based reasoning. Maintenance of this data from unauthorized access is important because this data is useful to find new pattern which can help to find conclusion. This conclusion is useful to provide better and fast services to patient and reduce cost of treatment in health sector. This paper gives an overview over current technology about use of cloud computing and big data and concern area related to their safety and privacy.

I. RELATED WORK

Big information is characterized by considering 3V's i.e. volume, selection and speed, growing to 7v's whereby illustration of data cannot be confined to plain systems. Within the Seventies, the term "Big Data" was coined, however rose in 2008. huge information defines a dataset wherever the info size is on the far side the standard database's capability to record, store, manage and analyze data [3]. huge information has no universally accepted definition of however massive it ought to be for classification as huge information. the info volumes square measure within the vary of petabytes (10¹⁵), exabytes (10¹⁸) and on the far side [4]. Information created, collected and organized in exabyte once a year. However, its creation and aggregation is fast and may approach to zeta-byte (10²¹) inside the approaching years. A review of huge information challenges and problems for data-centric applications expressed in sort and significance of data retrieved [5]. Also, this extended too several areas additional together with cloud computing, IOT (internet of things), social networks, tending applications etc. was verified helpful. huge information has advancements toward data management and handling challenges like huge information analysis, information diversity, information extraction and reducing, integration and cleansing and variety of different- different tools for analysis and mining [6]. As business areas square measure developing and there's a desire to recompile the financial framework, rethinking connections among producers, merchants, and customers of merchandise and enterprises [7]. In the year 2018, the cloud security become biggest concern for cloud researchers because of unauthorized activities square measure growing on in step with cloud users [8] projected new security design for cloud framework that give safer information transformation and shield information from information outflow. information house owners and cloud servers have completely different identities, this framework give information storage and have completely different security problems, associate degree freelance procedure needed to make certain that cloud information is hosted properly within the cloud server [9] mentioned completely different security techniques for secure information storage on cloud. Cloud computing uses "Utility Computing" and "Software-as-a-Service" to provide needed service by cloud user, cloud security could also be a main and vital truth, has various problems and downside connected it [10] delineated the list of parameters that square measure affected the safety and explore security problems and issues square measure faced by cloud service supplier and shoppers like information privacy, security problems and infected application.

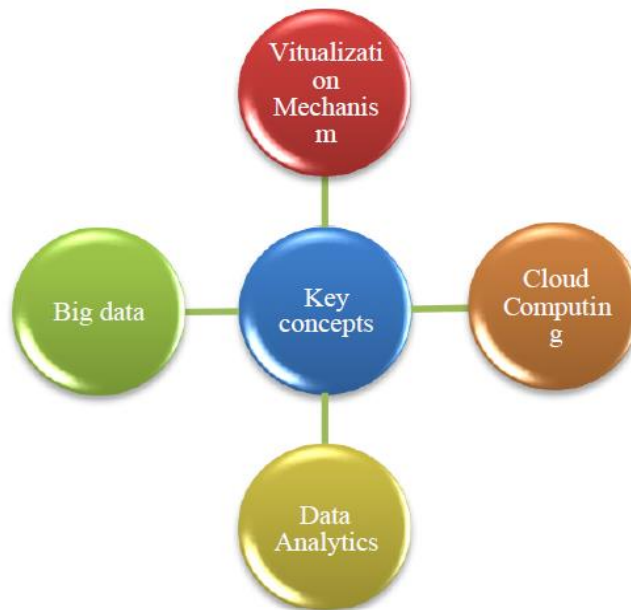
II. KEY CONCEPTS

While it's pragmatic and value effective to use cloud computing for data-intensive applications, there may be problems with privacy and security once mistreatment systems that don't seem to be provided in-house. to appear into these and realize applicable solutions, there square measure many key ideas that square measure wide employed in data-intensive clouds that for ought to be understood, like huge information infrastructures, virtualization mechanisms, kinds of cloud services, cloud computing and information Analytics.

A. Big Data

Computers turn out soaring rates of information that's primarily generated by the web of Things (IoT), Next-Generation Sequencing (NGS) machines, logical simulations and different origins of information that request economical building style for handling the new datasets. so as to address this terribly great amount of data, "Big Data" solutions like the Map/Reduce (MR), Google classification system (GFS), Hadoop Distributed classification system (HDFS) and also the Apache Hadoop are projected each as profit-oriented and ASCII text file. Key vendors within the IT trade like Oracle, IBM, Microsoft, Cisco, HP and SAP have preliminary customized these huge information solutions. there have been such a lot of definitions and packaging around "Big Data" at the rising points. throughout the previous couple of years, government agency fashioned the massive information operating group1 as a community with joint members from trade, academia, and government with the aim of developing a agreement definition, secure reference architectures, taxonomies, and technology roadmap. It identifies huge information characteristics as in depth datasets that square measure various, together with structured, semi-structured, and unstructured information from completely different domains (variety); massive orders of magnitude (volume); inbound with the quick rate (velocity); amendment in different characteristics (variability).

“Fig 1”:- Image related to key concept



B. Virtualization Mechanism

A hypervisor or virtual machine monitor (VMM) could be a pillar that resides between hardware and virtual machines to regulate the virtualized resource. It offers the means that to run variety of isolated virtual machines on an equivalent physical host. Hypervisors may be divided into 2 teams as follows.

a. Type 1

Here the hypervisor runs directly on the important system hardware, and there's no operating system (OS) under that. This approach is systematic because it eliminates any intermediary layers. Another advantage with this kind of hypervisor is that privacy levels may be improved by uninflected the guest Virtual Machines. That way, if a Virtual Machine is compromised, it will solely influence itself and cannot interfere with the hypervisor or different guest VMs.

b. Type 2

The second style of hypervisor runs on a hosted OS that offers virtualization services, like input/output (IO) device support and memory management. each VM interactions, like IO requests, network operations and interrupts, square measure handled by the hypervisor.

C. Cloud Computing

When considering cloud computing, we'd like to be understand the kinds of services that square measure offered, the means those services square measure delivered to those mistreatment the services, then many varieties of teams and other people that square measure involved cloud services. Cloud computing provide computing software system, platforms and infrastructures as services supported pay-as-you go models. Cloud service models may be deployed for on-request storage and computing power in numerous ways: Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS) and Infrastructure-as-a-Service (IaaS). Cloud computing service models are evolved throughout the previous couple of years inside a unique sort of domains mistreatment the “as-a-Service” thought of cloud computing like Business Integration-as-a-Service, Data-as-a-Service(DaaS), Cloud-Based Analytics-as-a-Service (CLAaaS) .

D. information Analytics

Big information analytics will advantages enterprises and organizations by determination several issues in producing, education, telecommunication, insurance, government, energy, retail, transportation, and health care. Over the previous couple of years, major IT vendors (such as Amazon, Google, and Microsoft) have provided virtual machines, via their clouds that customers might rent. These clouds utilize hardware resources and support live migration of Virtual Machines additionally to dynamic load-balancing and on-request provisioning. this suggests that, by dealing VMs via a cloud, the whole data center footprint of a contemporary enterprise will be reduced from the thousands of physical servers to a few hundred of a host.

III. ISSUES RELATED TO PRIVACY AND SECURITY

Cloud computing is a grid of several technologies at a same place because it's work as an on demand possibility of computer technique substances. These technologies include dataset, computer networks, operating systems, virtualization, memory control, scheduling, load balancing, and multi tenancy. These technologies enhanced the aspect of cloud computing like on demand service, combination of substances, wide networking, quick suppleness but also construct security and privacy risk which can cause the scenario of data stealing, damage of infrastructure, data disruption.

A. Issue Related to Security in Cloud Computing

Some issues related to cloud computing are given below:

a. Data Violation

Data violation is a concern issue in cloud computing because violation of confidential data harm the organization at financial level and also user of that organization lose their trust on them. To avoid data violation encryption can be a good option.

b. Grid Assurance

In cloud computing secure data is store and process on the software as a service level. Here this data is structured in the form of grid or network over the internet and to avoid data beaching at grid level this data must be secured by strong encryption technique.

c. Data Reigning

Data reigning is the important factors for those countries that use cloud services of that organization which is not belong to their country. In this scenario it's important for those countries that they demand for local level storage of their country data so they can monitor what kind of data is store by that organization and how they are going to use it.

d. Data Access

In cloud most of the data is accessible for everyone but to protect sensitive data from unauthorized access is a concern area and as a solution here we can use multi frame validation.

e. Denial-of-Service Attack

DOS attack is a type of attack which cannot be completely stop this kind of attack is stop the resources for user by temporary basis and it's restrain the cloud ability to execute. Regular observing and concern mitigation by cloud suppliers is best way to deny it.

f. Combined Cloud Computing Facilities

When cloud host combine their sources at that time they share their data, applications, assets, technical structure and all host not have a same kind of privacy and security between client and server so in a case if any cloud computing is effect by external attack it effect the other computing services also.

g. Hacking of Application Program Interface

Application program interface is a backbone of cloud computing system and it's establishing a communication between user and distributed cloud computing sys-tem. Hacking of API is harmful for the protection of data and traffic.

h. Immutable Data Damage and Deficient Data Alternate

Deficient data alternate can be destructive for organizations if they face immutable data damage because in this scenario they can loss their contact, stored data and it harm the business.

B. Issue Related to Security in Big Data

Safety of big data is very important for us but it has some security issues that are given below:

a. Liability of Feign Data Formation

In big data field formation of feign data is serious issue because when feign data is get mixed with real data it affect the quality of data and this kind of data is overall affect the performance and give incorrect result.

b. Missing Security Verification

Verification of big data at regular interval remove the security gap but most of the organization not perform this step at a serious

level and this issue affects the quality and transparency of data.

c. Verification of End-Point Devices

End-point input devices are important to execute the function of data storage, refinement so it's important for organization to use genuine and reasonable end point devices.

d. Inherent Existence of Indeterminate Mappers

When big data is gathered it's go through the mapper technique. Mapper as a function here processes the data and constitute modest slab of data. Input of mapper function is available in the form of key, value pairs. Here unauthorized access of mapper code can add unequal list of key pairs and get an access of sensitive information which can give the incorrect outcome.

e. Defensive Access Restrain

Data storage devices are assailable and it give easy access of the data which can harm the sensitivity of data. Use of encryption technique at access level can be a good option to prevent sensitivity of data.

V. SOLUTION OF SECURITY AND PRIVACY ISSUE

Some solutions for the security and privacy issue related to cloud computing and Big data are given below:

A. Security Solution for Cloud Computing

Some important security solutions for cloud computing are explained below:

a. Drafted Security Protocol Design

To secure the user data it's important to ensure that organization have proper security protection clause because it's develop sense of responsibility in organization and they made formalize security policy and if user data get stolen or data is get misused in any term so user have a right to take a legal action against them. Data security protocol is important in that case also when any country use cloud computing for data storage but cloud service provider company is belong to some other countries so in that scenario facilitator country also get an equal right to observe how service provider company using their data

b. Multifactor Substantiation

Multifactor substantiation is capable to provide illustrious security to data it works as a level of security wall. For example in an online banking system for online transaction banks provide the facility of password creation and one time password both for one time transaction because in any case if unauthorized person steal the account password of user account and try to do transaction so bank send one time password to real user number and they can verify the login to the data.

c. Restrain the Access of Data

Limitation of data access is helpful to prevent data from unauthorized access. Encryption and multi validation can be a good option to restrain the access of data.

d. Suitable Cloud Designs for Corporate

As per design cloud can be shared in two form private and public cloud. Public cloud gives access to multiple stakeholders while private cloud is design for single stakeholder. For large organization who keeps sensitive and confidential information private cloud can be a good option to prevent data.

e. Encryption of Data at Back-End

Still of most of the data in cloud get encrypted only on front-end. But for proper prevention we need to provide encryption at back-end also to enhance the security level of data and secure this data from unauthorized access.

f. Sheltered Data Destruction

Sometime depending upon the scenario data destruction is needed to avoid the situation of data leakage but for data destruction we need a proper authenticate system to prevent this data from wrong hands.

Security Solution for Big Data

Some important and good security solutions are given below:

a. Stop Feigns Data Formation

In big data when feign data is combining with real data its affect the quality of data and give false result. Here we can use feign detection technique to stop formation of feign data.

b. Testing of Mappers

Testing of mapping method is important to avoid addition of extra unequal list of key, value pair at the time of modest slab formation of data for testing we can use unit testing mapping as a method.

c. Select Authenticated Cloud Provider

When organization store data over cloud it's important for them to check security and privacy policy of cloud service provider for better prevention of their data.

d. Real Time Data Tracking

Real time data tracking is necessary to remove security loop hole in storage of huge data.

IV. CONCLUSION

In this paper we reviewed several safety and privacy issues related to big data and cloud computing. We try to describe four Key concepts related to Big data and Cloud computing like as data analytics and Virtualization Mechanism. We also discussed numerous safety and privacy issues related to cloud computing and big data as well as how we can solve or resolve these issues. In this paper we try to represent that how Big data and cloud computing will become the most important part of present day organization and without these two factors management of huge amount of data is very tough for us so big data and cloud computing play an very important role for managing the data generated from different-different. Resources.

REFERENCES

- [1] "Hypervisors, virtualization, and the cloud: Learn about hypervisors, system virtualization, and how it works in a cloud environment." Retrieved June 2015.
- [2] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for data storage security in cloud computing," in Proceedings of the 29th Conference on Information Communications, INFOCOM'10, (Piscataway, NJ, USA), pp. 525-533, IEEE Press, 2010.
- [3] Manyika, J., et al.: Big data: the next frontier for innovation, competition, and productivity,p. 1_137. McKinsey Global Institute, San Francisco (2011)
- [4] Kaiser, S., Armour, F., Espinosa, J.A., Money, W.: Big data: issues and challenges moving forward. In: 2013 46th Hawaii International Conference on Systems Science, pp. 995-1004(2013)
- [5] Agrawal, D., Bernstein, P., Bertino, E.: Challenges and Opportunities with Big Data 2011-1(2011)
- [6] Chen, M., Mao, S., Liu, Y.: Big Data: A Survey. Springer, New York (2014)
- [7] NIST Big Data Public Workinig Group: NIST Special Publication 1500- 4 NIST Big Data Interoperability Framework, Security and Privacy, vol. 4. September 2015.
- [8] A. Hussain, C. Xu, and M. Ali, "Security of Cloud Storage System using Various Cryptographic Techniques," International Journal of Mathematics Trends and Technology (IJMTT), vol. 60, no. 1, pp. 45-51, 2018.
- [9] A. Venkatesh and M. S. Eastaff, "A Study of Data Storage Security Issues in Cloud Computing," IJSRCSEIT, vol. 3, no. 1, pp. 1741-1745, 2018.
- [10]G. Jain and A. Jaiswal, "Security Issues and their Solution in Cloud Computing",Concepts journal of applied research(CJAR), vol. 02,no. 03, pp. 1-6, 2018.
- [11]o Carvalho, C. Miers, M. Naslund, and A. Ahmed, "A framework for authentication and authorization credentials in cloud computing," in Trust, Security and Privacy in Computing and Communications (TrustCom), 2013 12th IEEE International Conference on, pp. 509-516, July 2013.
- [12] R. Dua, A. Raja, and D. Kakadia, "Virtualization vs containerization to support paas," in Cloud Engineering (IC2E), 2014 IEEE International Conference on, pp. 610-614, March 2014.
- [13]Ali Gholami and Erwin Laure High Performance Computing and Visualization Department, KTH- Royal Institute of Technology, Stockholm, Sweden "BIG DATA SECURITY ANCloud Security Alliance (CSA), "Security Guidance for Critical Areas of Focus in Cloud Computing" version 3, 2011. Available at: <https://cloudsecurityalliance.org/guidance/csaguide.v3.0.pdf>

- [14] N. Mimura Gonzalez, M. Torrez Rojas, M. Maciel da Silva, F. Redigolo, T. Melo de Brito "PRIVACY ISSUES IN THE CLOUD" International Journal of Network Security & Its Applications (IJNSA) Vol.8, No.1, January 2016.
- [15] S. Subbalakshmi¹, Dr. K. Madhavi² ¹Research scholar, Department of CSE, JNTUCEA, Ananthapuramu, India. ²Associate Professor of CSE, JNTUCEA, Ananthapuramu, India "Security challenges of Big Data storage in Cloud environment: A Survey" International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 17 (2018) pp. 13237-13244 © Research India Publications. <http://www.ripublication.com>
- [16] R. Sumithra¹ Dr. R. Parameswari² ¹Research Scholar, Department of Computer Science, Vels Institute of Science Technology & Advanced Studies, Chennai 600 117, India, sumithraca@gmail.com ²Associate Professor, Department of Computer Science, Vels Institute of Science Technology & Advanced Studies, Chennai 600 117, India, dr.r.parameswari16@gmail.com "Security, Privacy Issues and Challenges in Big Data and Cloud Security: A Survey" Special Issue based on Proceedings of 4th International Conference on Cyber Security (ICCS) 2018.
- [17] Venkata Narasimhanukollu¹, Sailaja Arsi¹ and Srinivasa Rao Ravuri³ ¹Department of Computer Engineering, Texas Tech University, USA ³Department of Banking and Financial Services, Cognizant Technology Solutions, India "SECURITY ISSUES ASSOCIATED WITH BIG DATA IN CLOUD COMPUTING" International Journal of Network Security & Its Applications (IJNSA), Vol.6, No.3, May 2014.
- [18] N. Santos, K. P. Gummedi, and R. Rodrigues, "Towards trusted cloud computing," in Proceedings of the 2009 Conference on Hot Topics in Cloud Computing, HotCloud'09, (Berkeley, CA, USA), USENIX Association, 2009
- [19] Shazia Tabassam* Department of Computer Science, University of Agriculture, Faisalabad, Pakistan "Security and Privacy Issues in Cloud Computing Environment" Journal of Information Technology & Software Engineering Tabassam, J Inform Tech Softw Eng 2017.
- [20] N. Gonzalez, Miers C, Redigolo F, Carvalho T, Simplicio M, de Sousa G.T, and Pourzandi M. "A Quantitative Analysis of Current Security Concerns and Solutions for Cloud Computing.". Athens: 2011., pp 231 - 238, Nov. 29 2011-Dec. 1 2011
- [21] 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.
- [22] "Securing Big Data: Security Recommendations for Hadoop and NoSQL Environments." Securosis blog, version 1.0 (2012)
- [23] D. Perez-Botero, J. Szefer, and R. B. Lee, "Characterizing hypervisor vulnerabilities in cloud computing servers," in Proceedings of the 2013 International Workshop on Security in Cloud Computing, Cloud Computing '13, (New York, NY, USA), pp. 3-10, ACM, 2013.

