Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 4, Issue 12, December -2017

Achieving an Effective, Scalable and Privacy-Preserving Data Sharing Service In Big Data Environment

[1]Ch. Hrudaya Neeharika, [2]Y. Sunanda

[1]Assistant Prof, Dept. of CSE, Malla Reddy Institute of Technology & Science. HYD
[2]Assistant Prof, Dept. of CSE, Malla Reddy Institute of Technology & Science. HYD

Abstract *Security and privacy issues are amplified by the volume, assortment, and speed of Big Data. The decent variety of information sources, arrangements, and information streams, joined with the spilling idea of information obtaining and high volume make extraordinary security risks. This paper points of interest the security challenges when associations begin moving touchy information to a Big Data store like Hadoop. It recognizes the diverse danger models and the security control system to address and alleviate security hazards because of the distinguished risk conditions and utilization models. The structure sketched out in this paper is additionally intended to be distribution skeptic.*

Keywords: *Hadoop, Big Data, enterprise, defence, risk, Big Data Reference Framework, Security and Privacy, threat model*

1 Introduction

The expression "BigData" alludes to the gigantic measures of computerized data that organizations gather. Industry assesses on the development rate of information is generally twofold like clockwork, from 2500 Exabytes in 2012 to 40,000 Exabytes in 2020. Huge information isn't a Roleicular innovation. It is a gathering of traits and abilities. NIST characterizes Big Data as the accompanying: "BigData comprises of broad datasets, principally in the attributes of volume, speed, or potentially assortment that require an adaptable design for effective capacity, control, and examination" Securosis investigate includes extra qualities for a specific situation to qualify as 'Big Data'.
1. It handles a petabyte of data or more
2. It has distributed redundant data storage
3. Can leverage parallel task processing
4. Can provide data processing (MapReduce or equivalent) capabilities
5. Has extremely fast data insertion
6. Has central management and orchestration
7. Is hardware agnostic
8. Is extensible where its basic capabilities can be augmented and altered
Security and protection issues are amplified by the volume, assortment, and speed of Big Data. The decent variety of information sources, organizations, and information streams, joined with the spilling idea of information procurement and high volume make one of kind security dangers. It isn't just the presence of a lot of information that is making new security challenges for associations. BigData has been gathered and used by ventures for a very long while. Programming frameworks, for example, Hadoop empower designers and examiners to effectively use several processing hubs to perform information parallel figuring which was not there some time recently. Thus, new security challenges have emerged from the coupling of Big Data with heterogeneous arrangements of product equipment with item working frameworks, and ware programming foundations for putting away and registering on information. As Big Data grows at the diverse endeavours, customary security instruments customized to securing little scale, static information and information streams on firewalled and semi-disconnected systems are deficient. So also, it is hazy how to retrofit provenance in an endeavour's current framework. All through this record, unless unequivocally got out, Big Data will allude to the Hadoop system and its regular NoSQL variations (e.g. Cassandra, MongoDB, Couch, Riak, and so on.). This paper points of interest the security challenges when associations begin moving touchy information to a Big Data storehouse like Hadoop. It gives the distinctive danger models and the security control system to address and moderate the hazard because of the recognized security dangers.

2 Hadoop Security Weaknesses

Conventional Relational Database Management Systems (RDBMS) security has advanced throughout the years and with many 'eyeballs' surveying the security through different security assessments. Not at all like such arrangements, has Hadoop security not experienced a similar level of meticulousness or assessment so far as that is concerned and hence can guarantee little confirmation of the executed security. Another huge test is that today, there is no institutionalization or convenience of security controls between the diverse Open-Source Software (OSS) ventures and the distinctive Hadoop or Big Data sellers. Hadoop security is totally divided. This is genuine notwithstanding when the above gatherings execute a similar security highlight for the same Hadoop Role. Merchants and OSS gatherings' power fit security into the Apache Hadoop structure.

2.1 Top 10 Security & Privacy Challenges The Cloud Security Alliance Big Data Security Working Group has assembled the accompanying as the Top 10 security and protection difficulties to overcome in Big Data.
1. Secure calculations in conveyed programming structures
2. Security best practices for non-social information stores
3. Secure information stockpiling and exchanges logs
4. End-point input approval/sifting
5. Constant security observing
6. Versatile security protecting information mining and examination
7. Cryptographically implemented information driven security
8. Granular access control
9. Granular reviews
10. Information provenance

The above difficulties were assembled into four expansive Sections by the Cloud Security Alliance. They were: Infrastructure Security

- Secure computations in distributed programming frameworks
- Security best practices for non-relational data stores

Data Privacy

- Scalable privacy-preserving data mining and analytics
- Cryptographically enforced data centric security
- Granular access control

Data Management

- Secure data storage and transactions logs
- Granular audits
- Data provenance

Integrity & Reactive Security

- End-point input validation/filtering
- Real-time security monitoring

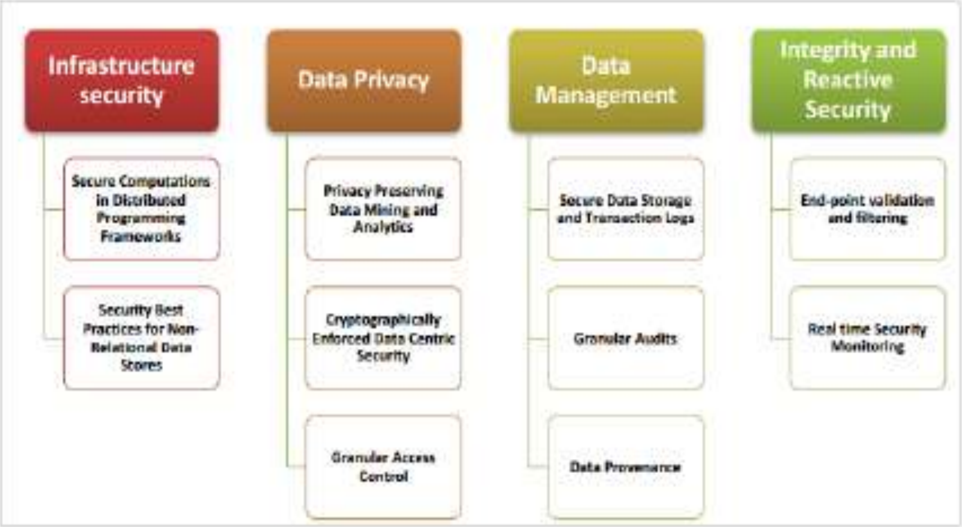

Infrastructure security
Secure Computations in Distributed Programming Frameworks
Security Best Practices for Non-Relational Data Stores
Data Privacy
Privacy Preserving Data Mining and Analytics
Cryptographically Enforced Data Centric Security
Granular Access Control
Data Management
Secure Data Storage and Transaction Logs
Granular Audits
Data Provenance
Integrity and Reactive Security
End-point validation and filtering
Real time Security Monitoring

Figure 1: CSA- classification of the Top 10 Challenges

2.2 Additional Security Weaknesses The prior Section with respect to Cloud Security Alliance list is a magnificent begin and this exploration and paper fundamentally adds to it. Where conceivable, exertion has been made to outline to the classes distinguished in the CSA work. This Section records some extra security shortcomings related with Open Source Software (OSS) like Apache Hadoop. It is intended to give the peruse a thought of the conceivable assault surface.

Infrastructure Security & Integrity

• The Common Vulnerabilities and Exposures (CVE) database just shows four announcing and settled Hadoop vulnerabilities in the course of recent years. Programming, even Hadoop, is a long way from idealize. This could either mirror that the security group isn't dynamic or that a big portion of powerlessness remediation happens inside the merchant situations themselves with no open detailing.

• Hadoop security configuration files are not self-contained with no validity checks prior to such policies being deployed, this usually results in the data integrity and availability issues.

Identity & Access Management

• Role Based Access Control (RBAC) arrangement records and Access Control Lists (ACLs) for Sections like MapReduce and HBase are typically designed through clear-text documents. These documents are editable by favoured records on the framework like root and other application accounts.

Data Privacy & Security

• All issues related with SQL infusion kind of assaults don't leave. They move with Hadoop Sections like Hive and Impala. SQL get ready capacities are right now not accessible which would have empowered Roleition of the inquiry and information

• Lack of local cryptographic controls for delicate information insurance. Oftentimes, such security is given outside the information or application stack.

• Clear-text information may be sent when conveying between DataNode to DataNode.

3 Big Data Security Framework

The accompanying Section gives the objective security engineering structure for Big Data stage security. The centre Sections of the proposed Big Data Security Framework are the accompanying:
1. Data Management
2. Identity & Access Management
3. Data Protection & Privacy
4. Network Security
5. Infrastructure Security & Integrity
The over '5 mainstays' of Big Data Security Framework are additionally decayed into 21 sub-Sections, each of which are basic to guaranteeing the security and moderating the security hazard and risk vectors to the Big Data stack. The general security system is demonstrated as follows.

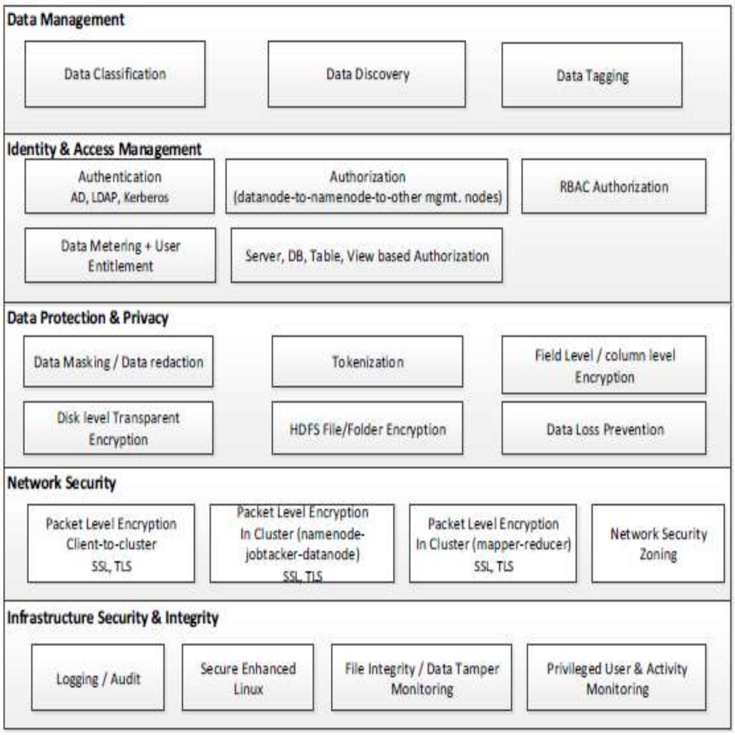
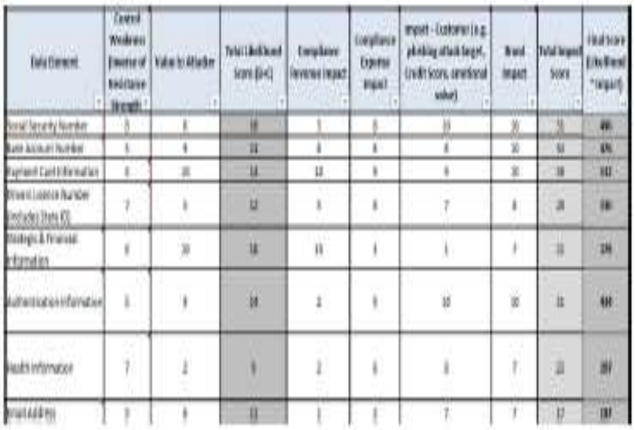

Data Management
Data Classification
Data Discovery
Data Tagging
Identity & Access Management
Authentication AD, LDAP, Kerberos
Authorization (datanode-to-namenode-to-other mgmt. nodes)
RBAC Authorization
Data Metering + User Entitlement
Server, DB, Table, View based Authorization
Data Protection & Privacy
Data Masking / Data redaction
Tokenization
Field Level / column level Encryption
Disk level Transparent Encryption
HDFS File/Folder Encryption
Data Loss Prevention
Network Security
Packet Level Encryption Client-to-cluster SSL, TLS
Packet Level Encryption In Cluster (namenode-jobtacker-datanode) SSL, TLS
Packet Level Encryption In Cluster (mapper-reducer) SSL, TLS
Network Security Zoning
Infrastructure Security & Integrity
Logging / Audit
Secure Enhanced Linux
File Integrity / Data Tamper Monitoring
Privileged User & Activity Monitoring

Figure 2: Big Data Security Framework

3.1 Data Management Data Management component is decomposed into three core sub-components. They are Data Classification, Data Discovery, and Data Tagging.

3.1.1 Data Classification Compelling information arrangement is likely a standout amongst the most essential exercises that can thus prompt successful security control usage in a Big Data stage. At the point when associations manage a to a great degree expansive measure of information, otherwise known as Big Data, by unmistakably having the capacity to distinguish what information matters, what needs cryptographic assurance among others, and what fields should be organized first for insurance, as a rule decide the accomplishment of a security activity on this stage.

1. Work with your lawful, protection office, Intellectual Property, Finance, and Information Security to decide every Roleicular datum fields. An open basin like wellbeing information isn't adequate. This activity urges the per user to go past the emblematic approach level exercise.
2. Play out a security control appraisal work out.
a. Decide area of information (e.g. presented to web, secure information zone)
b. Decide number of clients and frameworks with get to
c. Decide security controls (e.g. would it be able to be ensured cryptographically)
3. Decide estimation of the information to the assailant
a. Is the information simple to exchange on the underground market?
b. Do you have profitable Intellectual Property (e.g. a country state searching for atomic reactor outlines)
4. Decide Compliance and Revenue Impact
a. Decide break detailing prerequisites for all the Roleicular fields
b. Does loss of a specific information field keep you from working together (e.g. card holder information)
c. Gauge re-architecting cost for current frameworks (e.g. purchasing new security items)
d. Different costs like more incessant inspecting, fines and judgements and lawful costs identified with consistence
5. Decide effect to the proprietor of the PII information (e.g. a client)
a. Does the field cause phishing assaults (e.g. email) versus simply supplant it (e.g. loss of a Mastercard)
The following figure is a sample representation of certain Personally Identifiable Data fields



Data Element	Control Weakness (Inverse of Resistance Strength)	Value to Attacker	Total Likelihood Score (B+C)	Compliance Revenue Impact	Compliance Expense Impact	Impact - Customer (e.g. phishing attack target, Credit Score, emotional value)	Brand Impact	Total Impact Score	Final Score (Likelihood *Impact)
Social Security Number	[illegible]	8	16	[illegible]	[illegible]	10	10	[illegible]	[illegible]
Bank Account Number	5	9	14	8	8	8	10	34	[illegible]
Payment Card Information	8	10	18	12	6	6	10	34	[illegible]
Drivers License Number (Includes State ID)	7	5	12	5	8	7	8	28	[illegible]
Strategic & Financial Information	8	10	18	11	1	1	7	[illegible]	[illegible]
Authentication Information	5	9	14	2	9	10	10	31	[illegible]
Health Information	7	2	9	2	6	8	7	23	[illegible]
Email Address	3	8	11	1	1	7	7	17	[illegible]

Figure 3: Data Classification Matrix

3.1.2 Data Discovery The absence of situational mindfulness concerning delicate information could leave an association presented to noteworthy dangers. Distinguishing whether touchy information is available in Hadoop, where it is found and along these lines setting off the fitting information insurance measures, for example, information masking, information redaction, tokenization or encryption is vital.

• For organized information going into Hadoop, for example, social information from databases, or, for instance, comma-isolated esteems (CSV) or JavaScript Object Notation (JSON)- arranged documents, the area and grouping of touchy information may as of now be known.

• With unstructured information, the area, tally and order of delicate information turn out to be considerably more troublesome. Information revelation, where touchy information can be recognized and found, turns into a critical initial phase in information security.

1. Characterize and approve the information structure and mapping. This is all helpful prep work for information assurance exercises later
2. Gather measurements (e.g. volume checks, interesting tallies and so on.). For instance, if a document has 1M records yet it is copy of a solitary individual, it is a solitary record versus 1M records. This is exceptionally helpful for consistence however more vitally chance administration.
3. Offer this understanding with your Data Science groups for them to construct danger models, profiles which will be helpful in information exfiltration aversion situations.
4. On the off chance that you find grouping documents, work with your application groups to move far from this information structure. Rather use columnar capacity organizations, for example, Apache Parquet where conceivable paying little respect to the information handling structure, information mode, or programming dialect.
5. Construct contingent pursuit schedules (e.g. just write about date of birth if a man's name is found or Credit Card # + CVV or CC +zip)
6. Record for use cases where once delicate information has been cryptographically secured (e.g. scrambled or tokenized), what is the use case for the disclosure arrangement.

3.1.3 Data Tagging Comprehend the conclusion to-end information streams in your Big Data condition, Roleicularly the entrance and deRoleure techniques.

1. Distinguish every one of the information entrance techniques in your Big Data bunch. These would incorporate all manual (e.g. Hadoop administrators) or computerized techniques (e.g. ETL occupations) or those that experience some meta-layer (e.g. duplicate documents or make + compose).
2. Knowing whether information is coming in utilizing Command Line Interface or however some Java API or through Flume or Sqoop import of on the off chance that it is being SSH'd in is imperative.
3. Thus, take after the information out and recognize all the deRoleure Roles out of your Big Data condition.
4. This incorporates in the case of detailing employments are being go through Hive inquiries (e.g. through ODBC/JDBC), through Pig employments (e.g. perusing documents or Hive tables or HCatalog), or trading it out by means of Sqoop or duplicating by means of REST API, Hue and so forth will decide your control limits and put stock in zones.
5. The greater Role of the above will likewise help in information revelation movement and other information get to administration works out (e.g. to execute RBAC, ABAC, and so on.)

3.2 Identity & Access Management POSIX-style permissions in the secure HDFS are the basis for many access controls across the Hadoop stack.

3.2.1 User Entitlement + Data Metering Give clients access to information by halfway overseeing access approaches.
• It is critical to attach approach to information and not to the entrance technique
• Leverage Attribute based access control and secure information in light of labels that move with the information through heredity; consents choices can use the client, condition (e.g. area), and information characteristics.
• Perform information metering by limiting access to information once an ordinary edge (as controlled by get to models + machine learning calculations) is passed for a specific client/application.

3.2.2 RBAC Authorization Convey fine-grained approval through Role Based Access Control (RBAC).
• Manage information access by Role (and not client)
• Determine connections between clients and Roles through gatherings. Use AD/LDAP assemble Roleicipation and implement manages over all information get to ways

3.3 Data Protection & Privacy Most of the Hadoop appropriations and seller additional items bundle either information very still encryption at a square or (entire) document level. Application level cryptographic insurance (like field-level/section level encryption, information tokenization, and information redaction/veiling give the following level of security required.

3.3.1 Application Level Cryptography (Tokenization, field-level encryption) While encryption at the field/component level can offer security granularity and review following capacities, it comes to the detriment of requiring manual intercession to decide the fields that require encryption and where and how to empower approved decoding.

3.3.2 Transparent Encryption (disk / HDFS layer) Full Disk Encryption (FDE) anticipates get to by means of the capacity medium. Document encryption can likewise make preparations for (advantaged) access at the hub's working framework level.
• On the off chance that you have to store and process touchy or controlled information in Hadoop, information very still encryption ensures your association's touchy information and keeps at any rate the circles out of review scope.
• In bigger Hadoop groups, circles regularly should be expelled from the bunch and supplanted. Plate Level straightforward encryption guarantees that no intelligible leftover information remains when information is expelled or when circles are decommissioned.
• Full-plate encryption (FDE) can likewise be OS-local circle encryption, for example, dm-tomb

3.3.3 Data Masking/ Data Redaction Information masking or information redaction before stack in the normal ETL process de-recognizes personally identifiable information (PII) information before stack. Along these lines, no touchy information is put away in Hadoop, keeping the Hadoop Cluster conceivably out of (review) scope.
• This might be performed in cluster or constant and can be accomplished with an assortment of outlines, including the utilization of static and dynamic information veiling instruments, and in addition through information administrations.

3.4 Network Security The Network Security layer is disintegrated into four sub-Roles. They are information assurance in-travel and system zoning + approval Sections.

3.4.1 Data Protection In-Transit Secure interchanges are required for HDFS to ensure information in-travel. There are numerous danger situations that thusly order the need for https and anticipate data exposure or rise of benefit risk classes. Utilizing the TLS convention (which is currently accessible in all Hadoop circulations) to verify and guarantee protection of correspondences between hubs, name servers, and applications.
• An assailant can increase unapproved access to information by capturing correspondences to Hadoop comforts.

• This could incorporate correspondence amongst NameNodes and DataNodes that are free back to the Hadoop customers and thus can bring about qualifications/information to be sniffed.
• Tokens that are allowed to the client post-Kerberos confirmation can likewise be sniffed and can be utilized to mimic clients on the NameNode.
Following are the controls that when executed in a Big Data group can guarantee properties of information classification.

1. Bundle level encryption utilizing TLS from the customer to Hadoop group
2. Bundle level encryption utilizing TLS inside the group itself. This incorporates utilizing https between NameMode to Job Tracker to DataNode.
3. Bundle level encryption utilizing TLS in the group (e.g. mapper-reducer)
4. Utilize LDAP over SSL (LDAPS) instead of LDAP when speaking with the corporate undertaking indexes to counteract sniffing assaults.
5. Permit your administrators to design and empower encoded rearrange and TLS/https for HDFS, MapReduce, YARN, HBase UIs and so on.

3.4.2 Network Security Zoning The Hadoop bunches must be portioned into purposes of conveyance (PODs) with chokepoints, for example, Top of Rack (ToR) switches where organize Access Control Lists (ACLs) restrain the enabled activity to endorsed levels.
• End clients must not have the capacity to interface with the individual information hubs, yet to the name hubs as it were.
• The Apache Knox portal for instance, gives the ability to control movement all through Hadoop at the per-benefit level granularity.
• A fundamental firewall that ought to permit get to just to the Hadoop NameNode, or, where adequate, to an Apache Knox door. Customers will never need to discuss specifically with, for instance, a DataNode.

3.5 Infrastructure Security & Integrity The Infrastructure Security and Integrity layer is decayed into four center sub-Sections. They are Logging/Audit, Secure Enhanced Linux (SELinux), File Integrity + Data Tamper Monitoring, and Privileged User and Activity Monitoring.

3.5.1 Logging / Audit All framework/biological system changes interesting to Hadoop bunch should be examined with the review logs being ensured. Illustrations include:
• Addition/erasure of information and administration hubs
• Changes in administration hub states including work tracker hubs, name hubs
• Pre-shared mysteries or testaments that are taken off when the underlying bundle of the Hadoop dissemination or of the security arrangement is pushed to the hub keep the expansion of unapproved bunch hubs. At the point when information isn't restricted to one of the centre Hadoop Sections, Hadoop information security winds up having many moving Roles and high level of discontinuity. Thusly, there comes about a sprawl of metadata and review logs over all Roles. In an ordinary venture, the DBAs are normally utilized to put the security obligation at the table, line, Section, or cell level and keeping in mind that the design of document frameworks and framework heads, and the Security Access Control group is typically responsible for the more granular record level authorizations. However, in Hadoop, POSIX-style HDFS authorizations are every now and again imperative for information security or are on occasion the main intends to implement information security by any stretch of the imagination. This prompts questions concerning the reasonability of Hadoop security. Innovations suggestions to address information discontinuity:
• **Apache Falcon** is a hatching Apache OSS venture that spotlights on information administration. It gives graphical information heredity and effectively controls the information life cycle. Metadata is recovered and pounded up from wherever the Hadoop application stores it.
• **Cloud era Navigator** is an exclusive device and GUI that is a piece of Cloud era’s Distribution Including Apache Hadoop (CDH) circulation. CDH Navigator is an apparatus to address log sprawl, heredity and a few Roles of information revelation. Metadata is recovered and squashed up from wherever the Hadoop application stores it.
• **Zettaset Orchestrator** is an item to harness the general fracture of Hadoop security with a restrictive consolidated GUI and work process. Zettaset has its own Roleicular metadata archive where metadata from all Hadoop Sections is gathered and put away.
3.5.2 Secure Enhanced Linux (SELinux) SELinux was made by the United States National Security Agency (NSA) as an arrangement of patches to the Linux Kernel utilizing Linux Security Modules (LSM). It was in the long run discharged by the NSA under the GPL permit and has been received by the upstream Linux bit. SELinux is a case of a Mandatory Access Control (MAC) for Linux. Verifiably Hadoop and other Big Data stages based over Linux and UNIX frameworks have had optional access control. What this implies for instance is that a special client like root is transcendent.
• By implementing and designing SELinux on your Big Data condition, through MAC, there is approach which is authoritatively set and settled.
• Even if a client changes any settings on their home catalog, the arrangement keeps another client or process from getting to it.

• A specimen approach for instance that can be actualized is to make library records executable yet not writable or the other way around. Employments can write to/tmp area however not having the capacity to execute anything in there. This is an extraordinary approach to forestall summon infusion assaults among others.
• With strategies designed, regardless of whether somebody who is a sysadmin or a vindictive client can access root utilizing SSH or some other assault vector, they might have the capacity to peruse and compose a great deal of stuff. Be that as it may, they won't have the capacity to execute anything incl. possibly any information exfiltration strategies. The general suggestion is to run SELinux is lenient mode with customary workloads on your bunch, reflecting average use, including utilizing any devices. The notices produced would then be able to be utilized to characterize the SELinux strategy which subsequent to tuning can be sent in a 'focused on authorization' mode.

4 Final Recommendations

The accompanying are some key proposals in moderating the security dangers and dangers recognized in the Big Data biological community.

1. Select items and merchants that have demonstrated involvement in comparable scale arrangements. Demand merchant references for vast arrangements (that is, comparative in size to your association) that have been running the security controls under thought for your undertaking for no less than one year
2. Key columns are: Accountability, adjusting system driven, get to control driven, and information driven security is totally basic in accomplishing a decent general dependable security pose.
3. Information driven security, for example, name security or cell-level security for delicate information is favored. Mark security and cell-level security are coordinated into the information or into the application code as opposed to including information security afterward
4. Externalize information security when conceivable and utilize information redaction, information veiling or tokenization at the season of ingestion, or utilize information administrations with granular controls to get to Hadoop
5. Tackle the log and review sprawl with information administration apparatuses, for example, OSS Apache Falcon, Cloudera Navigator or the Zettaset Orchestrator. This accomplishes information provenance over the long haul

5. Conclusion

Hadoop and huge information are never again popular expressions in huge endeavours. Regardless of whether for the right reasons or not, endeavour information distribution centres are moving to Hadoop and alongside it come petabytes of information.

In this paper we have laid the foundation for directing future security evaluations on the Big Data biological community and securing it. This is to guarantee that Big Data in Hadoop does not turn into a major issue or a major target. Merchants pitch their advancements as the enchanted silver shot. Be that as it may, there are many difficulties with regards to conveying security controls in your Big Data condition. This paper likewise gives the Big Data danger display which the peruser can additionally grow and modify to their authoritative condition. It additionally gives objective reference engineering around Big Data security and spreads the whole control stack.

Hadoop and Big Data speak to a green field open door for security specialists. It gives an opportunity to advance beyond the bend, test and send your instruments, procedures, examples, and strategies before huge information turns into a major issue.

References

[01] T. Nishide, K. Yoneyama, and K. Ohta, "Attribute-based encryption with Roleially hidden encryptor-specified access structures," in Applied cryptography and network security. Springer, 2008, pp. 111–129.
[02] J. Li, K. Ren, B. Zhu, and Z. Wan, "Privacy-aware attribute-based encryption with user accountability," in Information Security. Springer, 2009, pp. 347–362.
[03] D. Boneh and B. Waters, "Conjunctive, subset, and range queries on encrypted data," in Theory of cryptography. Springer, 2007, pp. 535–554.
[04] J. Katz, A. Sahai, and B. Waters, "Predicate encryption supporting disjunctions, polynomial equations, and inner products," in Advances in Cryptology–EUROCRYPT'08. Springer, 2008, pp. 146–162.
[05] J. Lai, R. H. Deng, and Y. Li, "Fully secure cipertext-policy hiding cp-abe," in Information Security Practice and Experience. Springer, 2011, pp. 24–39.
[06] L. Lei, Z. Zhong, K. Zheng, J. Chen, and H. Meng, "Challenges on wireless heterogeneous networks for mobile cloud computing," IEEE Wireless Communications, vol. 20, no. 3, pp. 34–44, 2013.
[07] K. Zheng, Z. Yang, K. Zhang, P. Chatzimisios, K. Yang, and W. Xiang, "Big data-driven optimization for mobile networks toward 5g," IEEE Network, vol. 30, no. 1, pp. 44–51, 2016.
[08] S. Yu, K. Ren, and W. Lou, "Attribute-based text distribution with hidden policy," in Secure Network Protocols (NPSec'08 Workshop),IEEE, 2008, pp. 39–44.
[09] J. Lai, R. H. Deng, and Y. Li, "Expressive cp-abe with Roleially hidden access structures," in Proc. of ASIACCS'12. ACM, 2012, pp. 18–19.

[10] J. Hur, "Attribute-based secure data sharing with hidden policies in smart grid," IEEE Trans. Parallel Distrib. Syst., vol. 24, no. 11, pp. 2171–2180, 2013.

[11] A. Beimel, "Secure schemes for secret sharing and key distribution," Ph.D. dissertation, Israel Institute of Technology, Technion, Haifa, Israel, 1996.

[12] B. H. Bloom, "Space/time trade-offs in hash coding with allowable errors," Communications of the ACM, vol. 13, no. 7, pp. 422–426, 1970.

[13] EMC Big Data 2020 Projects http://www.emc.com/leadership/digital-universe/iview/big-data-2020.htm

[14] NIST Special Publication 1500-1 *NIST Big Data Interoperability Framework: Volume 1, Definitions* http://bigdatawg.nist.gov/_uploadfiles/M0392_v1_3022325181.pdf

[15] Securosis – Securing Big Data Security issues with Hadoop environments https://securosis.com/blog/securing-big-data-security-issues-with-hadoop-environments.

ABOUT AUTHORS:

Ch. Hrudaya Neeharika is currently working as assistant prof. in Computer Science Engineering Department, Malla Reddy Institute of Technology & Science
College, Hyderabad. She received her M.Tech in Computer Science & Engineering from CMR Institute of Technology, Hyderabad. She received her B.Tech in Computer Science & Engineering from Christu Jyoti Institute of technology & Science, Warangal.

Y. Sunanda is currently working as assistant prof. in Computer Science Engineering Department, Malla Reddy Institute of Technology & Science College, Hyderabad. She received her M.Tech in Computer Science & Engineering from Vathsalya Institute of Science & Technology, Bhuvanagiri . She received her B.Tech in Computer Science & Engineering from Narasaraopeta Engineering College, Narasaraopet.