IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, January 2026
Impact Factor: 7.67

# An Analysis of Big Data Privacy and Security in the Healthcare Industry

**Gopal Shankar[1] and Dr. Udai Shankar[2]**
[1]Research Scholar, Department of Computer Science and Engineering
[2]Associate Professor, Department of Computer Science and Engineering
Sunrise University, Alwar, Rajasthan

**Abstract:** *Healthcare organizations and all of its subsectors are facing an unprecedented flood of big data due to the ever-increasing integration of highly diverse enabled data generating technologies in the medical, biomedical, and healthcare fields, as well as the growing availability of data at a central location that can be used in any kind of organization, from hospitals to health insurance companies to pharmaceutical manufacturers. The healthcare sector is unable to fully use this data with its existing resources due to the overwhelming security and privacy concerns, despite the fact that it is being heralded as the key to improving health outcomes, gaining insightful knowledge, and cutting expenses. Nonetheless, big data management and use are essential to any healthcare organization's success. This study seeks to address some of the current data privacy, data security, user access procedures, and tactics while presenting the state-of-the-art security and privacy challenges in big data as applied to the healthcare business.*

**Keywords:** Security and Privacy, Big Data in Healthcare, Maintaining Privacy.

## I. INTRODUCTION

The healthcare business is experiencing a paradigm change as a result of the present trend toward digitizing processes and switching to electronic patient information. Big data is created when the amount of clinical data that is electronically accessible increases significantly in terms of complexity, variety, and timeliness. Big data holds the promise of supporting a wide range of unprecedented opportunities and use cases, including these key examples: clinical decision support, health insurance, disease surveillance, population health management, adverse events monitoring, and treatment optimization for diseases affecting multiple organ systems. These use cases are driven by mandatory requirements and have the potential to improve care, save lives, and lower costs.1, 2 Big data technology adoption in the healthcare industry has a number of opportunities and advantages, but it also presents some difficulties and obstacles. The security and privacy of sensitive data are indeed becoming more and more of a problem as a result of a number of expanding trends in healthcare, including cloud computing, wireless networking, clinician mobility, and health information interchange.

Furthermore, healthcare companies discovered that their patients' safety and the organization's security cannot be adequately safeguarded by a reactive, bottom-up, technology-centric approach to defining security and privacy needs. 3. Every healthcare institution has to adopt a proactive, preventative strategy and take steps with consideration for future security and privacy demands in order to avoid breaches of sensitive information and other sorts of security events. We will talk about a few noteworthy and well-received connected works in this essay. We will also go over some of the latest technological advancements, dangers to the security of health data, and how to mitigate these risks using innovative approaches. The privacy problem in healthcare will then be the main topic of discussion. Various rules and regulations set out by various regulatory agencies will be mentioned, along with some workable strategies and tactics utilized to protect patient privacy. Ultimately, we will outline some of the shortcomings of the strategies and tactics suggested to address security and privacy concerns before wrapping up the paper and outlining the further work we want to do.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, January 2026
Impact Factor: 7.67

## II. RELATED WORKS

In healthcare, a seamless integration of widely disparate big data technologies can lead to faster and safer patient throughput, deeper insights into organizational and clinical processes, increased efficiencies, and improvements in patient flow, safety, care quality, and overall patient experience all while keeping costs under control. That was the case with UNC Health Care (UNCHC), a non-profit integrated healthcare system in North Carolina that recently put in place a new technology that enables medical professionals to quickly access and use natural language processing to evaluate unstructured patient data. To extract insights and predictors of readmission risk for prompt intervention, better treatment for high-risk patients, and a reduction in re-admissions, UNCHC has actually accessed and analyzed enormous volumes of unstructured text found in patient medical records. Another example in the United States is the Indiana Health Information Exchange, a non-profit that links over 90 hospitals, community health clinics, rehabilitation facilities, and other healthcare providers in Indiana with a strong and safe technological network of health information. It makes it possible for medical data to follow a patient who is simply a part of a hospital system or one doctor's office.

Another example is the California-based Kaiser Permanente medical network, which has over 9 million members and is thought to handle enormous data quantities of between 26.5 and 44 petabytes. Six The Toronto Infant Hospital is one Canadian institution that uses big data analytics. By improving infant outcomes, this hospital has been able to reduce the risk of major hospital infections. This time around in Europe, specifically in Italy, the Italian drugs agency is a part of a national profitability program and gathers and evaluates a significant quantity of clinical data on costly new drugs. It could reevaluate the conditions of market access and drug costs in light of the findings. At the 2017 Annual Meeting of the American College of Medical Genetics and Genomics (ACMG), Sophia Genetics 8, a global leader in Data-Driven Medicine, announced that African hospitals have begun using its artificial intelligence to improve patient care, following the lead from Europe, Canada, Australia, Russia, and Latin America. Some of the most innovative medical facilities in Morocco, such as PharmaProcess in Casablanca, ImmCell, The Al Azhar Oncology Center, and The Riad Biology Center in Rabat, have begun integrating Sophia to expedite and analyze genomic data in order to identify disease-causing mutations in patients' genomic profiles and determine the best course of treatment.

As new SOPHiA users, they join a wider network of 260 institutions across 46 countries that exchange clinical insights about patient cases and demographics. This knowledge base is fed by biological discoveries to speed up diagnosis and treatment. 9. Automations have improved patient care efficiency and cut costs, but they also raise the risk of security and privacy breaches due to the growing amount of healthcare data. 2016 saw a 320% surge in hacking assaults on healthcare providers, and 81% of data compromised that year were directly the result of hacking attempts, according to CynergisTek's Redspin's 7th annual Breach Report: Protected Health Information (PHI).

Furthermore, the most common threat to hospitals has been determined to be ransomware, which is characterized as a kind of malware that encrypts data and keeps it captive until a ransom demand is paid. These results highlight the urgent need for providers to adopt a much more proactive and all-encompassing strategy to safeguard their information assets and counter the rising danger that cyberattacks pose to the healthcare industry.

**Big data security in healthcare**

Huge volumes of data are sent, stored, and maintained by healthcare institutions in order to facilitate the provision of effective treatment. Still, for many years, protecting sensitive data has been a difficult need. To make things worse, the healthcare sector is still among the most vulnerable to data breaches that are made public. Data breaches occur because attackers may really get sensitive information by using data mining techniques and processes and then making it public. The stakes are constantly increased as advanced methods are developed to circumvent security restrictions, even if putting security measures in place is still a complicated procedure. Therefore, it is essential that businesses put in place healthcare data security solutions that will both safeguard valuable assets and meet healthcare compliance requirements. technologies in use Healthcare data security and privacy are safeguarded by a variety of technologies. The most popular technologies are:

**Authentication:** The process of proving or verifying statements made by or about the topic are genuine and truthful is known as authentication. It is essential to any business since it protects user identities, grants access to corporate networks, and verifies that a user is who they say they are. In order to guard against man-in-the-middle (MITM) attacks,

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, January 2026 Impact Factor: 7.67

endpoint authentication is a feature of the majority of cryptographic protocols. For example, Secure Sockets Layer (SSL) and Transport Layer Security (TLS), which are cryptographic protocols, provide security for communications across networks like the Internet. Network connection segments are end-to-end encrypted at the Transport Layer using TLS and SSL. Numerous iterations of the protocols are extensively used in various applications such as voiceover-IP (VoIP), instant messaging, faxing over the Internet, email, and web surfing. Through the use of a mutually trusted certification authority, one may authenticate the server using TLS or SSL. Furthermore, any critical data may be 360° monitored using the Bull Eye algorithm. This method has been used to maintain relationships between original and copied data and to ensure data security. Critical data may only be accessed or written by those who are authorized. Paper 24 uses the one-time pad technique to present a fresh and straightforward authentication paradigm. It offers the removal of password transmission between servers. In a healthcare system, customer identities and the healthcare information provided by providers should be validated at the point of entry.

**Encryption:** Encrypting data effectively prevents unwanted access to private information. Its solutions safeguard and uphold ownership of data at every stage of its lifetime, from the data center to the endpoint (which includes administrators', doctors', and clinicians' mobile devices) to the cloud. To protect against security lapses like packet sniffing and storage device theft, encryption is helpful. It is vital for healthcare organizations or providers to guarantee that their encryption strategy is effective, user-friendly for both medical professionals and patients, and quickly adaptable to include new electronic health records. Additionally, there should be as few keys held by each side as possible. The right choice of appropriate encryption algorithms to ensure safe storage remains a challenging issue, despite the fact that several encryption algorithms (RSA, Rijndael, AES and RC6 20, 22, 23, DES, 3DES, RC4 21, IDEA, Blowfish, etc.) have been designed and implemented quite well.

**Data Masking:** Although masking substitutes an unidentifiable value for sensitive data components, it is not a true encryption technology, hence the masked value cannot provide the original value. It employs a method of suppressing or generalizing quasi-identifiers like date of birth and zip codes, as well as de-identifying the data sets or concealing personal identifiers like name and social security number. Therefore, one of the most often used methods for live data anonymization is data masking. The k-anonymity that was first put out by Swaney and Samrati 12,13 guards against attribute disclosure but not against identity revelation. P-sensitive anonymity, as proposed by Truta et al. (2014), offers protection against attribute and identity disclosure. Additional techniques for achieving anonymity include introducing noise into the data, rearranging cells within columns, and substituting k copies of a single sample for groupings of k records. One typical issue with these approaches is that they have trouble anonymizing large dimensional data sets (15, 16). This method lowers the cost of safeguarding a large data deployment, which is a great plus. Masking lessens the need to implement extra security measures on the data while it is stored in the platform during the safe data migration process from a secure source.

**Access Control:** Users may access an information system once they have been authenticated, but their access will still be controlled by an access control policy. These policies are usually based on the privileges and rights of each practitioner who has been given authorization by a patient or a reliable third party. Thus, it's a strong and adaptable system for giving users authorization. Sophisticated authorization controls are provided to guarantee that users can only carry out the tasks for which they are authorized, including cluster management, data access, and job submission. Numerous approaches have been put out to deal with the issues of security and access control. The two most often used EHR approaches are Role-Based Access Control (RBAC) 17 and Attribute-Based Access Control (ABAC) 18–19. When RBAC and ABAC are employed independently in the medical system, several limitations have been observed. Additionally, Paper 25 suggests an effective dynamic access control strategy for cloud-oriented storage that uses cipher text based on symmetric encryption algorithms (like AES) and CP-ABE. We advise using technologies in combination with other security strategies, such as encryption and access control methods, to meet the needs of fine-grained access control while maintaining security and privacy. 4. Healthcare and big data privacy Advanced persistent threats, which are targeted assaults on information systems with the primary goal of smuggling recoverable data by the attacker, have become more prevalent in recent years. Consequently, breaches of patient privacy are seen to be a major worry in the field of big data analytics, which presents a difficulty for businesses trying to handle these many, interrelated, and pressing issues. In actuality, data privacy establishes access to data based on privacy rules and regulations, which

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, January 2026
Impact Factor: 7.67

specify, for example, who may read personal data, financial, medical, or secret information. Data security, on the other hand, controls access to data throughout the data lifecycle. Concerns about patient privacy are raised by an occurrence that was covered by Forbes magazine 26. According to the article, a teenage girl received baby care vouchers from Target Corporation without her parents knowing about it. Because of this occurrence, big data must take privacy into account for analytics, and developers should be able to confirm that their apps adhere to privacy agreements and protect sensitive data even when the apps or privacy laws change. Therefore, the privacy of medical data is a crucial element that has to be given careful thought.

**Data protection laws**

In order to comply with the expanding body of relevant data privacy laws, healthcare organizations must now more than ever manage and secure personal information as well as handle their risks and legal obligations in regard to processing personal data. The laws and practices governing data privacy vary throughout nations. The Table below lists important aspects and data protection laws and policies in a few of the nations.

**Table 1: Data protection laws in some of the countries**

| Country | Law | Salient Features |
|---|---|---|
| U.S.A | HIPAA Act<br>Patient Safety and Quality Improvement Act (PSQIA)<br>HITECH Act | Requires the establishment of national standards for electronic health care transactions. Gives the right to privacy to individuals from age 12 through 18.<br>Signed disclosure from the affected before giving out any information on provided health care to anyone, including parents.<br>Patient Safety Work Product must not be disclosed [27].<br>Individual violating the confidentiality provisions is subject to a civil penalty.<br>Protect security and privacy of electronic health information. |
| EU | Data Protection Directive | Protect people's fundamental rights and freedoms and in particular their right to privacy with respect to the processing of personal data. [29] |
| Canada | Personal Information Protection and Electronic Documents Act ('PIPEDA') | Individual is given the right to know the reasons for collection or use of personal information, so that organizations are required to protect this information in a reasonable and secure way.[28] |
| UK | Data Protection Act (DPA) | Provides a way for individuals to control information about themselves. Personal data shall not be transferred to a country or territory outside the European Economic Area unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects. |
| Morocco | The 09-08 act, dated on 18 February 2009 | Protects the one's privacy through the establishment of the CNDP authority by limiting the use of personal and sensitive data using the data controllers in any data processing operation. [30] |
| Russia | Russian Federal Law on Personal Data | Requires data operators to take "all the necessary organizational and technical measures required for protecting personal data against unlawful or accidental access". |
| India | IT Act and<br>IT (Amendment) Act | Implement reasonable security practices for sensitive personal data or information. Provides for compensation to person affected by wrongful loss or wrongful gain. Provides for imprisonment and/or fine for a person who causes wrongful loss or wrongful gain by disclosing personal information of another person while providing services under the terms of lawful contract. |
| Brazil | Constitution | The intimacy, private life, honour and image of the people are inviolable, with assured right to indenization by material or moral damage resulting from its violation. |
| Angola | Data Protection Law (Law no. 22/11of 17 June) | With respect to sensitive data processing, collection and processing is only allowed where there is a legal provision allowing such processing and prior authorization from the APD is obtained |

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, January 2026 Impact Factor: 7.67

**Privacy preserving methods in big data**

Few traditional methods for privacy preserving in big data is described in brief here. Although These techniques are used traditionally to ensure the patient's privacy their demerits led to the advent of newer methods.

**De-identification:** a traditional method to prohibit the disclosure of confidential information by rejecting any information that can identify the patient, either by the first method that requires the removal of specific identifiers of the patient or by the second statistical method where the patient verifies himself that enough identifiers are deleted. Nonetheless, an attacker can possibly get more external information assistance for de-identification in the big data. As a result, de-identification is not sufficient for protecting big data privacy. It could be more feasible if develop efficient privacy-preserving algorithms to help mitigate the risk of re-identification. The concepts of k-anonymity 34, 36, 37, l-diversity 35, 36, 38 and t-closeness 34, 38 have been introduced to enhance this traditional technique.

**k-anonymity:** In this technique, higher the value of k, lower will be the probability of re-identification. However, it may lead to distortions of data and hence greater information loss due to k-anonymization. Furthermore, in kanonymization, if the quasi-identifiers containing data are used to link with other publicly available data to identify individuals, then the sensitive attribute (like Disease) as one of the identifier will be revealed. Various measures have been proposed to quantify information loss caused by anonymization, but they do not reflect the actual usefulness of data 39, 40.

**L-diversity:** It is a form of group based anonymization that is utilized to safeguard privacy in data sets by diminishing the granularity of data representation. This model (Distinct, Entropy, Recursive) 34,36, 41 is an extension of the k-anonymity which utilize methods including generalization and suppression to reduces the granularity of data representation in a way that any given record maps onto at least k different records in the data. The l-diversity model handles a few of the weaknesses in the k-anonymity model in which protected identities to the level of kindividuals is not equal to protecting the corresponding sensitive values that were generalized or suppressed. The problem with this method is that it depends upon the range of sensitive attribute. If want to make data L-diverse though sensitive attribute has not as much as different values, fictitious data to be inserted. This fictitious data will improve the security but may result in problems amid analysis. As a result, L-diversity method is also a subject to skewness and similarity attack 41 and thus can't prevent attribute disclosure.

**T-closeness:** is a further improvement of l-diversity group based anonymization. The t-closeness model(Equal/Hierarchical distance) 34, 38 extends the l-diversity model by treating the values of an attribute distinctly by taking into account the distribution of data values for that attribute. The main advantage of this technique is that it intercepts attribute disclosure, and its problem is that as size and variety of data increases, the odds of reidentification too increases. B. HybrEx Hybrid execution model 42 is a model for confidentiality and privacy in cloud computing. It utilizes public clouds only for an organization's non-sensitive data and computation classified as public, i.e., when the organization declares that there is no privacy and confidentiality risk in exporting the data and performing computation on it using public clouds, whereas for an organization's sensitive, private data and computation, the model executes their private cloud. Moreover, when an application requires access to both the private and public data, the application itself also gets partitioned and runs in both the private and public clouds. It considers data sensitivity before a job's execution and provides integration with safety. The problem with HybridEx is that it does not deal with the key that is generated at public and private clouds in the map phase and that it deals with only cloud as an adversary43.

**Identity based anonymization**

When anonymization, privacy protection, and big data methods 44 were effectively coupled to evaluate use data while safeguarding user identities, problems arose. In order to take advantage of the substantial advantages that come with cloud storage, Intel developed an open architecture for anonymization 44 that made it possible to use a range of tools for the de- and re-identification of web log data. Enterprise data has characteristics that vary from the typical examples in the anonymization literature 46 during the implementation architectural phase. Intel also discovered that the anonymized data was vulnerable to correlation attacks even when it concealed clear Personal Identification Information such as IP addresses and usernames. They discovered that there is a substantial correlation between User Agent information and specific users after examining the trade-offs associated with fixing these vulnerabilities. The needs, implementation, and experiences observed while applying anonymization to preserve privacy in corporate data

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, January 2026 Impact Factor: 7.67

examined using big data methods are described in this case study of anonymization implementation in a company. K-anonymity based criteria were employed in this analysis to assess the anonymization quality. Simultaneously, it was discovered that anonymization requires more than just hiding or generalizing certain variables; anonymized datasets must be thoroughly examined to see whether they are subject to security breaches.

## III. DISCUSSION

In this work, we have examined the security and privacy issues associated with big data by going over a few current methods and strategies that are likely to be very helpful to healthcare companies in securing security and privacy. By no means is this a comprehensive list. In this part, we emphasized the focus and limits of various methods and methodologies that were described in various articles. For example, considerable research efforts are required to solve particular privacy challenges in certain specific big data analytics using the technique proposed in 48, which developed an efficient and privacy-preserving cosine similarity computing protocol. An additional research study 44 explores the challenges and lessons learned from effectively fusing Big Data, anonymization, and privacy protection strategies to evaluate use data while safeguarding users' identities. Yet, it continues to use the correlation-attack-prone K-anonymity approach. In addition to these studies, 49 suggested a cloud-based, scalable, two-phase top-down specialization (TDS) method for anonymizing massive data sets utilizing the Map Reduce framework. However, it employs an anonymization method that is susceptible to a correlation attack. 50 Because consumer segmentation and profiling may easily lead to discrimination based on age, gender, ethnic origin, health condition, social background, and other factors, the range of privacy problems pertaining to big data applications that have been presented is also restricted. Furthermore, 51 suggested the use of a fast anonymization algorithm (FAST) to expedite the anonymization of large data streams. However, the design and implementation of this algorithm need further study, and great scalability and cloud compute power can only be attained by implementing it inside a distributed cloud-based framework. 52 presented a methodology as the last technique covered in this section. It offers data confidentiality, safe data sharing without the need for re-encryption, access control for insider threats, and forward and backward access control, which lowers the degree of trust in the cryptographic server (CS). In comparison to earlier models, the new models' increasing complexity and limitations make them harder to understand and harder to evaluate for dependability.

## IV. CONCLUSION

Big data has many chances to advance clinical treatment, knowledge acquisition, personal health management, and health research. However, a variety of barriers and difficulties, including as technological difficulties, privacy and security concerns, and a shortage of qualified personnel, prevent it from reaching its full potential in the healthcare industry. Researchers in this discipline regard big data security and privacy to be a major obstacle. We have covered a few instances of successful relevant global work in this study. Along with the benefits and drawbacks of the current privacy and security solutions in the context of big healthcare data, privacy and security challenges at each stage of the big data life cycle are also given. Apart from the strategies we have discussed, there are other ways that are being used to protect patient privacy in healthcare. These methods include the following: hiding a needle in a haystack 47; attribute-based encryption; access control; homomorphic encryption; storage path encryption; and so on. But the issue is constantly placed upon us. In light of this, our future views will be more focused on finding workable solutions to the big data privacy and security scalability issue in the healthcare age. In order to advance, we will attempt to resolve the conflict between security and privacy models by modeling various strategies and taking use of the MapReduce framework. to eventually aid in planning and decision-making processes.

## REFERENCES

[1]. Burghard C: Big Data and Analytics Key to Accountable Care Success. IDC Health Insights; 2012.
[2]. Fernandes L, O'Connor M, Weaver V: Big data, bigger outcomes. J AHIMA 2012:38–42.
[3]. David Houlding, MSc, CISSP: « Health Information at Risk: Successful Strategies for Healthcare Security andPrivacy » Healthcare IT Program Of ce Intel Corporation, white paper 2011.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, January 2026
Impact Factor: 7.67

**[4].** "UNC Health Care relies on analytics to better manage medical data and improve patient care." IBM press release. October 11, 2013.
**[5].** Indiana Health Information Exchange: http://www.ihie.org/ (Accessed Date: March 24, 2016).
**[6].** Transforming Healthcare through Big Data, Strategies for leveraging big data in the health care industry. Institute for health- 2013
**[7].** The Big Data revolution in healthcare, accelerating value and innovation – Peter Groves, Basel Kayyali, David Knott , Steve Van Kuiken –2013
**[8].** Sophia Genetics: « Product & Technology Overview » 2014
**[9].** Sophia Genetics: http://www.sophiagenetics.com/news/media-mix/details/news/african-hospitals-adopt-sophia-artificial-intelligence-to-triggercontinent-wide-healthcare-leapfrogging-movement.html (March 24, 2017)
**[10].** CynergisTek, Redspin : « BREACH REPORT 2016: Protected Health Information (PHI)» February 2017
**[11].** Rui Zhangand Ling Liu: '' Security Models and Requirements for Healthcare Application Clouds'' in IEEE 3rd International Conference on Cloud Computing, 2010
**[12].** L. Sweeney, "Achieving k-anonymity privacy protection using gener- alization and suppression," in International journal on uncertainity, fuzziness and knowledge based systems, vol. 10, 2002, pp. 571 – 588.
**[13].** P. Samrati, "Protecting respondents identities in microdata release," in IEEE transactions on knowledge and data engineering, vol. 13, 2001, pp. 1010 – 1027.
**[14].** T. M. Truta and B. Vinay, "Privacy protection: p-sensitive k-anonymity property," in Proceedings of 22nd International Conference on Data Engineering Workshops, 2006, p. 94.
**[15].** N. Spruill, "The confidentiality and analytic usefulness of masked business microdata," in Proceedings on survey research methods, 1983, pp. 602–607.
**[16].** S. Chawala, C. Dwork, F. M. Sheny, A. Smith, and H. Wee, "Towards privacy in public databases," in Proceedings on second theory of cryptography conference, 2005.
**[17].** Science Applications International Corporation (SAIC). Role-Based Access Control (RBAC) Role Engineering Process Version 3.0. 11 May 2004.
**[18].** Mohan, D. M. Blough, An Attribute-Based Authorization Policy Framework with Dynamic Conflict Resolution, Proceedings of the 9th Symposium on Identity and Trust on the Internet, 2010.
**[19].** M. Hagner. Security infrastructure and national patent summary. In Tromso Telemedicine and eHealth Conference, 2007.
**[20].** Federal Information Processing Standards Publication 197,"Specification for the Advanced Encryption Standards (AES)", 2001.
**[21].** S. Fluhrer, I. Mantin, and A. Shamir, "Weakness in the Key schedualing algorithm of RC4", 8th Annual International Workshop on Selected Areas in Cryptography, Springer-Verlag London, UK, 2001.
**[22].** J. Shafer, S. Rixner, and A. L. Cox. The Hadoop Distributed File system: Balancing Portability and Performance. Proc. of 2010 IEEE Int. Symposium on Performance Analysis of Systems & Software (ISPASS), March 2010, White Plain, NY, pp. 122-133.
**[23].** N. Somu, A. Gangaa, and V. S. Sriram, "Authentication Service in Hadoop Using one Time Pad," Indian Journal of Science and Technology, vol. 7, pp. 56-62, 2014.
**[24].** C. Yang, W. Lin, and M. Liu, "A Novel Triple Encryption Scheme for Hadoop-Based Cloud Data Security," in Emerging Intelligent Data andWeb Technologies (EIDWT), 2013 Fourth International Conference on, 2013, pp. 437-442.
**[25].** H. Zhou and Q. Wen, "Data Security Accessing for HDFS Based on Attribute-Group in Cloud Computing," in International Conference on Logistics Engineering, Management and Computer Science (LEMCS 2014), 2014.
**[26].** K. Hill, "How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did," Forbes, Inc., 2012.
**[27].** Data Protection Laws of the World. 2017 DLA Piper. [Online]. Available: http://www.dlapiperdataprotection.com

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, January 2026
Impact Factor: 7.67

**[28].** Challenges of privacy protection in Big Data Analytics –Meiko Jensen- 2013 IEEE International Congress on Big Data. 2013.
**[29].** Privacy and Big Data – Terence Craig & Mary E.Ludloff
**[30].** 30 Data protection overview (Morocco) – Florence Chafiol-Chaumont and Anne-Laure Falkman – 2013.
**[31].** Big Data security and privacy issues in healthcare – Harsh Kupwade Patil, Ravi Seshadri – 2014
**[32].** Sectorial healthcare strategy 2012-2016- Moroccan healthcare ministry.
**[33].** Big Data in Healthcare – Pranav Patil, Rohit Raul, Radhika Shroff, Mahesh Maurya – 2014
**[34].** Li N, et al. t-Closeness: privacy beyond k-anonymity and L-diversity. In: Data engineering (ICDE) IEEE 23rd international conference; 2007.
**[35].** Machanavajjhala A, Gehrke J, Kifer D, Venkitasubramaniam M. L-diversity: privacy beyond k-anonymity. In: Proc. 22nd international conference data engineering (ICDE); 2006. p. 24.
**[36].** Ton A, Saravanan M. Ericsson research. [Online]. http://www.ericsson.com/research-blog/data-knowledge/big-data-privacy-preservation/2015.
**[37].** Samarati P. Protecting respondent's privacy in microdata release. IEEE Trans Knowl Data Eng. 2001;13(6):1010–27
**[38].** Samarati P, Sweeney L. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical Report SRI-CSL-98-04, SRI Computer Science Laboratory; 1998.
**[39].** V. Iyenger, "Transforming data to satisfy privacy constraints," in Pro- ceedings of the ACM SIGKDD, 2002, pp. 279–288.
**[40].** K. LeFevre, R. Ramakrishnan, and D. J. DeWitt, "Modorian multidi- mensional k-anonymity," in Proceedings of the ICDE, 2006, p. 25.
**[41].** Sweeney L. K-anonymity: a model for protecting privacy. Int J Uncertain Fuzz. 2002;10(5):557–70.
**[42].** Ko SY, Jeon K, Morales R. The HybrEx model for confidentiality and privacy in cloud computing. In: 3rd USENIX workshop on hot topics in cloud computing, HotCloud'11, Portland; 2011. 43. Priyank J., Manasi G. and Nilay K. Big data privacy: a technological perspective and review. In Journal of Big Data2016.
**[43].** Sedayao J, Bhardwaj R. Making big data, privacy, and anonymization work together in the enterprise: experiences and issues. Big Data Congress; 2014.
**[44].** Yong Yu, et al. Cloud data integrity checking with an identity-based auditing mechanism from RSA. Future Gener Comp Syst. 2016;62:85–91.
**[45].** Oracle Big Data for the Enterprise, 2012. [online]. http://www.oracle.com/ca-en/technoloqies/biq-doto.
**[46].** Jung K, Park S, Park S. Hiding a needle in a haystack: privacy preserving Apriori algorithm in MapReduce framework PSBD'14, Shanghai; 2014. p. 11–17.
**[47].** Lu R, Zhu H, Liu X, Liu JK, Shao J. Toward efficient and privacy-preserving computing in big data era. IEEE Netw. 2014;28:46–50.
**[48].** Zhang X, Yang T, Liu C, Chen J. A scalable two-phase top-down specialization approach for data anonymization using systems, in MapReduce on cloud. IEEE Trans Parallel Distrib. 2014;25(2):363–73.
**[49].** Mehmood A, Natgunanathan I, Xiang Y, Hua G, Guo S. Protection of big data privacy. In: IEEE translations and content mining are permitted for academic research. 2016.
**[50].** Mohammadian E, Noferesti M, Jalili R. FAST: fast anonymization of big data streams. In: ACM proceedings of the 2014 international conference on big data science and computing, article 1. 2014.
**[51].** Wei L, Zhu H, Cao Z, Dong X, Jia W, Chen Y, Vasilakos AV. Security and privacy for storage and computation in cloud computing. Inf Sci. 2014;258:371–86