Journal of Science and Technology
ISSN: 2456-5660 Volume 06,Special Issue 01, August 2021,

www.jst.org.in *DOI: https://doi.org/10.46243/jst.2021.v6.i04.pp42-46*

Privacy Violation Patterns in Non-Relational Databases

Shubham Dawkhar [1],Dr. Shwetambari Chiwhane[2]
[1,2]Dept. of Computer Engineering, NBN Sinhgad School of Engineering, Pune, Maharashtra, India
[1]Sdawkhar98@gmail.com
[2]shwetambari.chiwhane@sinhgad.edu

To Cite this Article
Shubham Dawkhar, Dr. Shwetambari Chiwhane, "A Privacy Violation Patterns in Non-Relational Databases", Journal of Science and Technology, Vol. 06, Special Issue 01, August 2021, pp42-46.

Article Info
Received: 15.07.2021 Revised: 24.07.2021 Accepted:10.08.2021 Published: 16.08.2021

Abstract: *In 21st century data became the new oil. Every decision taken by organizations, governments, individuals is influenced by data. With analysis of data, one can predict market behaviour, can forecast weather. With such high value, it is important to manage and access data. Data has been growing exponentially since the late 1990s. there are many databases system used to store this data. Before the 2000's mainly relational databases were used to store. But managing relational databases is expensive and complicated. To expand the capacity of relational databases hardware and skilled employees are needed. in the 2000s NoSQL databases were introduced. NoSQL database is fast, reliable, efficient, supports low-cost hardware up-gradation, and can store unstructured data. NoSQL databases are one the most preferred options as of today. With great speed and efficiency NoSQL rapidly became popular and widely used. But to achieve this function security is highly compromised. with a lot of security and privacy flaws data became more vulnerable to privacy breaching. many pieces of research have been done in the NoSQL domain to eliminate the flaws but all of them are scattered in this review paper we tried to synthesis the previous literature and compile all the repeating privacy violation patterns into one paper.*

Key Word*:* *NoSQL ,databases,privacy,security*

I. INTRODUCTION

Technological advancement has resulted into the creation of tremendous data. This data is responsible for growth of the organizations. today data is considered as the real wealth of the organizations. The Term NoSQL Database can be used as modern databases with high performance, Storage capabilities and scalability. These databases are reliable and efficient but they often lacks in privacy and security. In this review paper we synthesised the repeating patterns of privacy violations from previous literature available in the domain.

II. BACKGROUND

A. Database

A database is a separate application that is used to store data. Each database has one or more distinct APIs to create, access, manage, search and replace data. There are other data storing systems too, but they are not as efficient and fast as databases are.

There are many types of databases depending on the usage requirements. Few of them are as follows:

- Hierarchical databases
- Network databases
- Object-oriented databases
- Relational databases
- NOSQL databases

In this paper we are going to focus on NOSQL databases and privacy issues.

B. NOSQL Database:

A NOSQL database can be called as non-SQL or Non-Relational Database which offers mechanism to save and fetch data. This data is stored differently compared to relational database where data is stored in tabular form.

Simplicity, horizontal scaling of machines, better control over availability are the characteristics of NOSQL databases.

Advantage of Non relational Databases:

- Handles huge volume of data at highspeed with expansion focused architecture.
- Can hold Unstructured, Semi Structured and Structured Data.
- Simple Updates
- Development Friendly

C. Data Privacy

Data privacy can explain as how a data or information is handled based on its sensitivity and importance. For example, you may not hesitate to share your name while greeting other person, but you will think about any other personal information. While opening your bank account you will share information with bank without any second thought. In This modern world nowadays, data is the new oil. To the Organizations. All the decisions are takes places on basis of the data available. With such value connected with it data becomes the most important thing in this internet age.

Data privacy is the relation between handling of the data and the laws, rules and regulations. With public expectations and political issues. Generally, user's privacy may be compromised under the following situations:

- When personal information is jointly used with external datasets which can be shade light on latest personal information regarding to the individual person.
- While analysing personal information organization may know about the characteristics about the individual.
- When lack of governance and management occurs, sensitive data may get leak

D. Data Privacy in RDBMS

In 21st Century importance of data is growing tremendously as the decades passes. It is growing exponentially as the amount of data is growing need of infrastructure is also growing but maintaining infrastructure is expensive and energy hungry. Choosing right and most efficient is most important thing. NoSQL databases is easy to scaleup, efficient, fast and cost friendly option for data management.

NoSQL Databases provide high reading with minimal writing latency, efficiency in big data storage in distributed manner with High Scalability and availability and low operational cost. NoSQL was developed in late 2000's with flexible schema and horizontal scalability which is cheaper than relational database.

Thiughwhile providing performance, availability and great expansion, NoSQL databases lack in privacy features[2] .

There are various explorationdone in the field of NON RELATIONAL Databases privacy but they all are focused on single or few databases.

For ex.

For detection of irregular behaviour while conducting Map Reduce Queries with Hadoop various researches has been proposed [3,4] .

In this paper we review and sort prior research papers to identify various patterns of privacy breaching in NOSQL Database.

E. Types of NoSQL Databases

There are mainly four types of NoSQLdatabases. Each type is able to solve the problem which cant be solved with relational database. The types are document store(MongoDB,CouchDB),column-oriented database(H-Base, Cassandra) , graph database(Neo4j) and key-Value store(Redis).

III PRIVACY VIOLATIONS PATTERNS

In this section we present six privacy violation patterns narrowed down from various prior research done in NoSQL databases. The keywords we used to fetch prior research papers are: “Privacy”,” NOSQL”, “Mongo dB”, “Cassandra”, “on-relational Databases” “Security”. Recurring patterns were identified on distilled to Six patterns [1] .

These patterns are as follows:

- Hostile Query
  - NoSQL injection attack [5]
  - Insider attacks [6,7]
- Reidentification
  - Map Reduce [8]
  - Aggregation framework [8]
- Weak Validation
  - Poor password storage mechanism[9]
  - Limited to none Validation capabilities [10]
- Coarse grained Encapsulation
  - Partial authorization [11, 12]
- Vulnerable data in motion
  - Poor data security management due to distributed nature [11]
- Vulnerable data at rest
  - Weak encryption [6, 13]

1. Hostile Query

Hostile query introduction happens when aindividual breaks into system and modifies the original query to extract or change data with bad intention. Hostile query allows the person to Changing the back end of system with addition of data, deleting data or Modify data. Hostile Query modifications can be done using (1) NoSQL injection attacks [5] or (2) Insider attacks [6,7] .

2. Reidentification

Reidentification can be happen when a individual person or entity can be reidentified on the basis of output of the query result. Although many mechanisms are used to ensure safety of data this pattern can be implemented using sophisticated query where result output is small so reidentification can happen. Generallythis phenomenon prompts privacy troubles. MapReduce or aggregation framework [8]provide ways to attempt such sophisticated queries

3. Weak Validation

Validation is the process of Identification of user to grant the access to database where the user can access or modify the data.[12]. It is crucial method to ensure the privacy and security of the data to prevent any unauthorized access to the database. Unfortunately, NoSQL database generally provide the weak Validation to the user because (1) poor password mechanism [9] (2) Limited to nonexistence Validation capabilities.[10].

4. Coarse grained Encapsulation

Encapsulation or authorisation is the process of holding access to data and resources in the system. [11] normally done by associating different types of users with certain set of regulations based on their position and responsibilities.[12] . NoSQL databases support for Encapsulation. some types are roles and position based and few are data based. They can also be mixed.

5. Vulnerable data in motion.

In NOSQL data is stored at globally deployed shards in distributed manner. that's why computation take place at distributed environment. In such environment when data is in motion poses a risk to privacy because of low security mechanism used at the cluster. Which compromises cluster operations. While this is not the under control of the database the secure transfer of data to the various shards is. [11].

6. Vulnerable data at rest

Normally data always stayin motion in the NOSQL databases because of the distributed manner. But there are situations where data is not in the motion. NOSQL is designed to offer high availability and faster processing for big data compared to relational databases.[6]. Hence, to provide such functionalities privacy of data gets compromised. Balancing high performance and security is thus as an issue.[13] and weak encryption of data at rest or use of storage system prone to outsider attacks.

IV. SYSTEM EVALUATION

We picked eight widely used NOSQL databases. The databases used were: MONGODB, COUCHDB, REDIS, AEROSPIKE, CASSANDRA, HBASE, NEO4J, ORIENTDB. Table 1 summaries the analysis of all the databases across the six privacy breaching patterns. In the table (Y) indicate presence of pattern. (N) indicate none of the manifestation of patterns directly mitigated and (Y/N) indicate partial presence of the patterns. This table was created using analysis done by K. Goel, A. H. M. Ter Hofstede in the paper of Privacy-Breaching Pattein NoSQL Databases [1] .

Table no1 : System Evaluation Results

Patterns	MongoDB	CouchDB	Redis	Aerospike	Cassandra	Hbase	Neo4j	OrientDB
Hostile Query	Y	Y	Y	Y	Y	Y	Y	Y
Reidentification	Y	Y	Y	Y	Y	Y	Y	Y
Weak Validation	N	Y/N	Y/N	Y/N	Y/N	N	Y/N	N
Coarse grained Encapsulation	Y/N	Y	Y	Y/N	Y/N	N	N	N
Vulnerable data at rest	N	N	N	N	N	N	N	N
Vulnerable data in motion	N	Y	Y/N	N	N	N	N	N

V. CONCLUSION

Data Privacy and security is the real concern. Which has resulted in growing research in the field of security and privacy. NoSQL databases have many properties such as speed, efficiency, scalability but it lacks in the security and privacy. much research have been done but they all are scattered and unorganized. this paper used pattern bases approach to identify privacy issues in NoSQL by synthesising previous research done in the same field. This pattern opens up various paths for future research .by validating this patterns research can be done extensively in each pattern at deep level.

REFERENCES

[1]. K. Goel and A. H. M. T. Hofstede, "Privacy-Breaching Patterns in NoSQL Databases," in IEEE Access, vol. 9, pp. 35229-35239, 2021, doi: 10.1109/ACCESS.2021.3062034.

[2]. L. Okman, N. Gal-Oz, Y. Gonen, E. Gudes, and J. Abramov, ``Securityissues in NoSQL databases,'' in *Proc. IEEE 10th Int. Conf. Trust, Secur. Privacy Comput. Commun.*, Nov. 2011, pp. 541_547

[3]. C. Liao and A. Squicciarini, ``Towards provenance-based anomaly detection in MapReduce,'' in *Proc. 15th IEEE/ACM Int. Symp. Cluster, Cloud Grid Comput.*, May 2015, pp. 647_656.

[4]. S. N. Khezr and N. J. Navimipour, ``MapReduce and its applications, challenges, and architecture: A comprehensive review and directions for future research,'' *J. Grid Comput.*, vol. 15, no. 3, pp. 295_321, Sep. 2017.

[5]. B. Hou, K. Qian, L. Li, Y. Shi, L. Tao, and J. Liu, ``MongoDB NoSQL injection analysis and detection,'' in *Proc. IEEE 3rd Int. Conf. Cyber Secur. Cloud Comput. (CSCloud)*, Jun. 2016, pp. 75_78.

[6]. P. Raj and G. C. Deka, *A Deep Dive into NoSQL Databases: The Use Cases and Applications*, vol. 109. New York, NY, USA: Academic, 2018.

[7]. G. Kul, S. Upadhyaya, and A. Hughes, ``Complexity of insider attacks to databases,'' in *Proc. Int. Workshop Manag. Insider Secur. Threats*, Oct. 2017, pp. 25_32.

[8]. M. K. Srinivasan and P. Revathy, ``State-of-the-art big data security taxonomies,'' in *Proc. 11th Innov. Softw. Eng. Conf.*, 2018, p. 16.

[9]. K. Ahmad, M. S. Alam, and N. I. Udzir, ``Security of NoSQL database against intruders,'' *Recent Patents Eng.*, vol. 13, no. 1, pp. 5_12, 2019.

[10]. K. Grolinger, W. A. Higashino, A. Tiwari, and M. A. Capretz, ``Data management in cloud environments: NoSQL and NewSQL data stores,'' *J. Cloud Computing, Adv., Syst. Appl.*, vol. 2, no. 1, p. 22, 2013.

[11]. U. Saxena and S. Sachdeva, ``An insightful view on security and performance of NoSQL databases,'' in *Proc. Int. Conf. Recent Develop. Sci., Eng. Technol.* Singapore: Springer, 2017, pp. 643_653.

[12]. V. N. Gudivada, S. Jothilakshmi, and D. Rao, ``Data management issues in big data applications,'' *ALLDATA*, vol. 15, pp. 16_21, Apr. 2015.