Security issues in Big Data

[1]Tarunpreet Chawla, [2]Deepti Tak
[1]Assistant Professor, [2]Assistant Professor
[1]Computer Science & Engineering,
[1]Vivekananda Global University, Jaipur, India

Abstract : Privacy & Security concerns are growing as big data is becoming more and more accessible. The group and aggregation of huge quantities of various data are now possible. However, the tools and technologies that are being developed to manage these massive data sets are often not designed to incorporate sufficient security or privacy measures. So, a big question that arises is what security and privacy technology is adequate for the controlled and assured sharing for efficient direct access to big data. Making effective use of big data requires access from any domain to data in that domain, or any other domain it is allowed to access. This paper mainly focuses on the security issues in big data**.**

IndexTerms - **Big Data, Security, Privacy**

I. INTRODUCTION

The term Big Data refers to large-scale information management and analysis technologies that exceed the capability of traditional data processing technologies [1]. Big Data is differentiated from traditional technologies in three ways: the amount of data (volume), the rate of data generation and transmission (velocity), and the types of structured and unstructured data (variety). Big data implies performing computation and database operations for massive amounts of data, remotely from the data owner's enterprise. Recent technological advances and novel applications are making possible to capture, process, and share huge amounts of data – referred to as big data - and to extract useful knowledge, such as patterns, from this data and predict trends and events. Big data is making possible tasks that before were impossible.

II. CHALLENGES TO SECURITY IN BIG DATA

The biggest challenge for big data from a security point of view is the protection of user's privacy. Big data frequently contains huge amounts of personal identifiable information and therefore privacy of user's data and information is a major concern.Because of the big amount of data stored, breaches affecting big data have more devastating consequences.
An additional problem is that software commonly used to store big data, such as Hadoop, doesn't always come with user authentication by default. This makes the problem of access control worse, as a default installation would leave the information open to unauthenticated users. Big data solutions often rely on traditional firewalls or implementations at the application layer to restrict access to the information.
Securing big data comes with its own unique challenges beyond being a high-value target. The major difference between big data environments and traditional data environments include:

- The data collected, aggregated, and analyzed for big data analysis
- The infrastructure used to store and house big data
- The technologies applied to analyze structured and unstructured big data

III. MANAGING SECURITY IN BIG DATA

Big data is a relatively new concept and therefore there is not a list of best practices yet that are widely recognized by the security community. However there are a number of general security recommendations that can be applied to big data:

- Vet your cloud providers: If you are storing your big data in the cloud, you must ensure that your provider has adequate protection mechanisms in place. Make sure that the provider carries out periodic security audits and agree penalties in case those adequate security standards are not met.
- Create an adequate access control policy: Create policies that allow access to authorized users only.
- Protect the data: Both the raw data and the outcome from analytics should be adequately protected. Encryption should be used accordingly to ensure no sensitive data is leaked.
- Protect communications: Data in transit should be adequately protected to ensure its confidentiality and integrity.
- Use real-time security monitoring: Access to the data should be monitored. Threat intelligence should be used to prevent unauthorized access to the data.

IV. SECURITY METHODS FOR BIG DATA

Type Based Keyword Search for Security of Big Data: Big data provide many business opportunities to the information technology industry. Large scale applications of sensor networks, electronic health record systems, emails as well as social networks generate massive data each day. The volume of information collected and stored has exploded. Cloud computing system process extraordinary storage capacity and computation power and is promising to handle the big data processing system with its features. However, since the cloud data service provider system are distributed to share and process sensitive information assigned to them, a malicious data stealer might probably bring about serious privacy problems. Data encryption technology is used for boost information privacy protection. However, traditional encryption primitives (such as symmetric key encryption and public key encryption) are not capable to ensure the usability and hinder even authorized users from searching several keywords of encrypted files , it is difficult for the users to retrieve desired information from encrypted big data. So it is necessary to explore new cryptographic primitives to provide data encryption and search ability for big data era. Searchable encryption technology could fulfil the requirements to realize operability and data confidentiality, simultaneously. In this method, we provide a novel keyword search method to enable customers easily searching keywords from encryption-protection data. Moreover, the encrypted big data could be managed by different type that was assigned by data owner. Moreover, the access right can be given to others according to the user's willingness.

V. SECURING BIG DATA ENVIRONMENTS WITH VORMETRIC

Vormetric solutions for big data security enable organizations to maximize the benefits of big data analytics—while maximizing the security of their sensitive data and addressing the requirements of their compliance office. The Vormetric Data Security Platform offers the granular controls, robust encryption, and comprehensive coverage that organizations need to secure sensitive data across their big data environments—including big data sources, big data infrastructure, and big data analytic results. By delivering a single security solution that offers coverage of these areas, Vormetric enables security teams to leverage centralized controls that optimize efficiency and compliance adherence.

The Vormetric Data Security Platform offers capabilities for big data encryption, key management, and access control—featuring several product offerings that share a common, extensible infrastructure. Further, the solution generates security intelligence on data access by users, processes, and applications.

VI. CONCLUSION

Big Data is changing the way we perceive our world. The impact big data has created and it will continue to create ripple through all facets of our life.In this paper, I have highlighted the security and privacy problems in big data and how we can protect the information in big data. There is an immense scope in Big Data and a huge scope for research and Development.

REFERENCES

[1] http://gartner.com/it-glossary/big-data/
[2] http://www.cloudsecurityalliance.org, see Security Guide and Cloud Controls Matrix documents
[3] “Newly Emerging Best Practices for Big Data”, a Kimball Group Whitepaper by Ralph Kimball, Nov. 12, 2012.
[4] NIST Special Publication 800-145: “The NIST Definition of Cloud Computing”
[5] (IDG) Converged Storage: A Next Gen Storage Strategy for Big Data”, Hewlett-Packard Co., Nov. 12, 2012.
[6] “Newly Emerging Best Practices for Big Data”, a Kimball Group Whitepaper by Ralph Kimball, Nov. 12, 2012