

Result Analysis of Big Data Analytical Architecture for Remote Sensing Application in Real time

Neha Sharma¹, Akhilesh Bansiya²

M.Tech Sch., Department of Computer Science, RKDF University, Bhopal, India¹

Asst. Prof., Department of Computer Science, RKDF University, Bhopal, India²

Abstract: In recent years, big data became a hot analysis topic. The increasing quantity of big data additionally will increase the possibility of breaching the privacy of individuals. Since big data need high procedure power and large storage, distributed systems are used. As multiple parties are involved in these systems, the risk of privacy violation is increased. There are varieties of privacy-preserving mechanisms developed for privacy protection at different stages (e.g., data generation, data storage, and data processing) of an enormous data life cycle. The goal of this work is to provide a comprehensive summary of the privacy preservation mechanisms in big data and present the challenges for existing mechanisms. Specifically, during this work illustrate the infrastructure of big data and also the state-of-the-art privacy-preserving mechanisms in every stage of the big data life cycle. Moreover, discuss the challenges and future analysis directions related to privacy preservation in big data.

Keywords: Big Data, DADU, DPU, land and sea area, Real-time, remote senses, RSDU.

I. INTRODUCTION

Due to recent technological development, the quantity of data generated by social networking sites, sensor networks, Internet, health care applications, and lots of different companies, is drastically increasing day by day. All the large quantity of information generated from totally different sources in multiple formats with very high speed is referred as huge data. Big data has become a very active analysis space for last number of years.

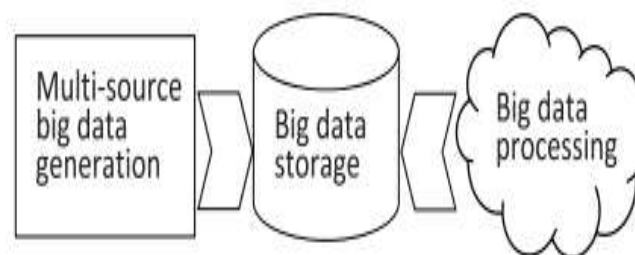


Fig.1 Big data Processing

The data generation rate is growing thus speedily that it's becoming very difficult to handle it using traditional ways or systems. Meanwhile, big information can be structured, semi-structured, or unstructured, that adds additional challenges once performing data storage and process tasks. Therefore, to the present end, it needs new ways that to store and analyse information in real time. Big data, if captured and analysed in a very timely manner, may be converted into actionable insights which may be of serious value. It will help businesses and organizations to enhance the internal decision making power and may produce new opportunities through information analysis. It may also help to promote the research and economy by transforming traditional business models and scientific values. Big data may be outlined in numerous ways that. For the scope of work use the definition given by International data Corporation (IDC) the term big data is outlined as "a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide kind of data, by enabling high-velocity capture, discovery, and/or analysis". Based on this definition, the properties of big knowledge are rejected by 3 V's, which are, volume, rate and variety, as shown in Fig. 1.1. Volume refers to the number of information generated. With the emergence of social networking sites, in this figure seen a dramatic increase within the size of the data. The speed at that new information are generated is usually characterized as velocity. A standard theme of big data is that the information are numerous, i.e., it will contain text, audio, image, or video etc. This diversity of data is denoted by selection. Despite big data can be effectively utilized for us to better understand the world and innovate in numerous aspects of human endeavors, the

exploding quantity of data has increased potential privacy breach. For instance, Amazon and Google will learn our shopping preferences and browsing habits. Social networking sites like Facebook store all the data about our personal life and social relationships. Common video sharing websites like YouTube recommends US videos based on our search history. With all the power driven by big data, gathering, storing and reusing our personal data for the aim of gaining commercial product, have place a threat to our privacy and security.

Big information has become the new frontier of information management given the amount of data today's systems are generating and consuming. It's driven the need for technological infrastructure and tools which will capture, store, analyse and visualize huge amounts of disparate structured and unstructured data. This info are being generated at increasing volumes from information intensive technologies yet as, but not limited to, the use of the internet for activities like accesses to information, social networking, mobile computing and commerce. Corporations and governments have begun to recognize that there are unexploited opportunities to enhance their enterprises which will be discovered from this data. Analytics once applied among the context of big data is that the strategy of examining large amounts of information, from a numerous vary of information sources and in various formats, to deliver insights which will alter decisions in real or near real time. Big data analytical approaches are used to recognize inherent patterns.

II. THEORY

A. Big Data Analytics

Big data has become the new frontier of data management given the number of information today's systems are generating and consuming. It's driven the necessity for technological infrastructure and tools which will capture, store, analyse and visualize large amounts of disparate structured and unstructured information. This information are being generated at increasing volumes from information intensive technologies as well as, however not limited to, the utilization of the net for activities like accesses to data, social networking, mobile computing and commerce. Corporations and governments have begun to recognize that there are fallow opportunities to enhance their enterprises which will be discovered from this information. Analytics once applied within the context of big data is that the method of examining large amounts of data, from a various variety of knowledge sources and in several formats, to deliver insights which will modify selections in real or close to real time. Massive information analytical approaches may be used to recognize inherent patterns, correlations and anomalies which might be discovered as results of integrating large amounts of information from different data sets. Analytics once applied within the context of big data is that the method of examining large amounts of information, from a spread of information sources and in several formats, to deliver insights which will modify decisions in real or close to real time. Various analytical ideas like data mining, natural language process, AI and predictive analytics will be used to analyse, contextualize and visualize the info. Big data analytical approaches will be used to recognize inherent patterns, correlations and anomalies which may be discovered as results of integration large amounts of information from different data sets. Big data analytics needs the utilization of recent frameworks, technologies and processes to manage it. Yet its arrival within the enterprise software house has created some confusion as business leaders try and understand the variations between it and traditional information warehousing (DW) and business intelligence (BI) tools.

There are vital distinctions and sufficient differentiating worth between BDA and DW/BI systems that build BDA unique. Gartner defines information warehouse as "a storage design designed to hold data extracted from transaction systems, operational information stores and external sources. The warehouse then combines that information in an aggregate, summary kind appropriate for enterprise-wide information analysis and reporting for predefined business desires."

III.METHOD

Personal information when combined with external datasets may lead to the inference of new facts about the users. Those facts may be secretive and not supposed to be revealed to others. Personal information is sometimes collected and used to add value to business. For example, individual's shopping habits may reveal a lot of personal information. The sensitive data are stored and processed in a location not secured properly and data leakage may occur during storage and processing phases. In order to ensure big data privacy, several mechanisms have been developed in recent years. These mechanisms can be grouped based on the stages of big data life cycle, i.e., data generation, storage, and processing. In data generation phase, for the protection of privacy, access restriction and falsifying data techniques are used. While access restriction techniques try to limit the access to individuals' private data, falsifying data techniques alter the original data before they are released to a non-trusted party. The approaches to privacy protection in data storage phase are mainly based on encryption techniques. Encryption based techniques can be further divided into attribute based encryption (ABE), Identity based encryption (IBE), and storage path encryption. In addition, to protect the sensitive information, hybrid clouds are used where sensitive data are stored in private cloud. The data processing phase includes privacy preserving data publishing (PPDP) and knowledge extraction from the data. In PPDP,

anonymization techniques such as generalization and suppression are used to protect the privacy of data. Ensuring the utility of the data while preserving the privacy is a great challenge in PPDP.

A. Attribute Based Encryption

The Encryption techniques which ensure send to end big data privacy in cloud storage system. In ABE access policies are defined by data owner and data are encrypted under those policies. The data can only be decrypted by the users whose attributes satisfy the access policies defined by the data owner. When dealing with big data one may often need to change data access policies as the data owner may have to share it with different organizations. The current attribute based access control schemes do not consider policy updating. The policy updating is a very challenging task in attribute based access control systems. The reason for that is once the data are outsourced to the cloud, the data owner would not keep the local copy in the system. If the data owner wants to update the policy, he has to transfer the data back to the local system, re-encrypt the data under new policy and store it back on the cloud server.

This process has got very high communication overhead and high computational cost. To solve the problem of policy updating, recently proposed a secure and variable policy updating outsourcing method. In data owner does not need to retrieve all the data and re-encrypt it. Instead the data owner can send the queries to cloud to update the policy, and the cloud server can update the policy directly without decrypting the data.

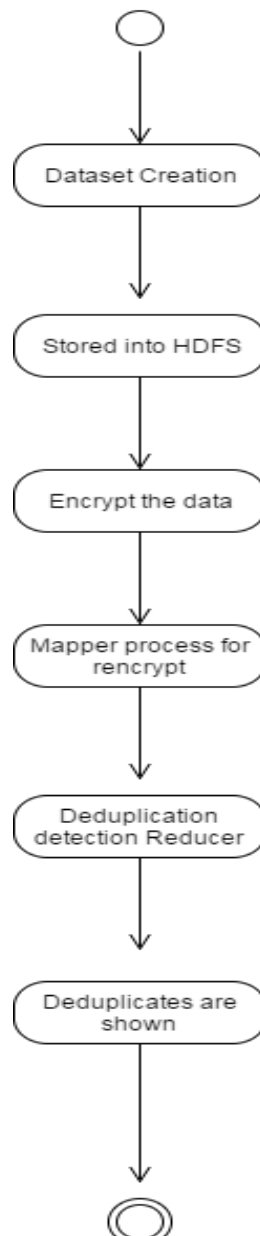


Fig.2 Flow of activity diagram

B. Data Processing through Map Reduce

Title Mapper Phase: Depending upon the Encrypted data has been send into the Mapper Phase as the input of the process for matching the data matching scenario been used for that.

Reducer Phase: The input of the mapper was sent to the shuffle process for the filtering of the whole data to obtain the result from the shuffle process. Then it been send to the sorting of the data are been deduplicated process. Finally the deduplicated process is achieved in a less time of processing.

C. Data Deduplication

Data Duplication by removing the personal details to preserve the privacy of users. It indicates that it would not be possible to identify an individual only from the anonymized data. However, due to the availability of huge volumes of data and powerful data analytic tools, the existing duplication techniques are becoming increasingly ineffective.

In big data scenarios, duplication needs to be more than just masking or generalizing certain fields. One needs to carefully analyse if the anonymized data are vulnerable to any attacks. For that, we need to study different attack models and information loss metric for big data anonymization. Moreover, most of the existing anonymization techniques are for static data, while much practical data is dynamic.

IV.RESULT

In this paper projected work implemented algorithms in simple java language using Beam-5.0 library also in Hadoop using MapReduce, at first in a single-node surroundings. Within the Hadoop implementation, Map function takes the image block offset as a key and also the image block (pixel values) as a value parameter.

Since Hadoop MapReduce cannot directly method image blocks, the entire product image data are converted into sequence file to be processed using MapReduce. In such some way, one line of the sequence file contains one image block. Map function performs parameters calculations on incoming block values and finally sends the block variety as a key and list of parameters results as a worth to the reduce function. Reduce function uses parameter results for performing decision-making on them.



Fig.3 terminal window

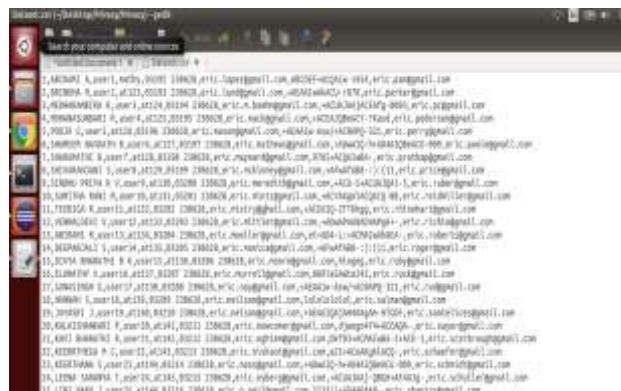


Fig.4 Dataset

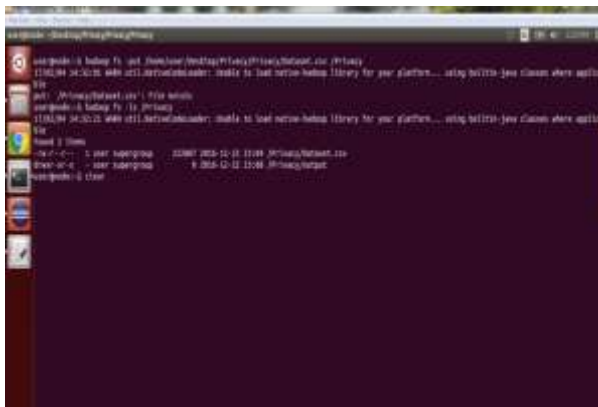


Fig.5 giving command on terminal window

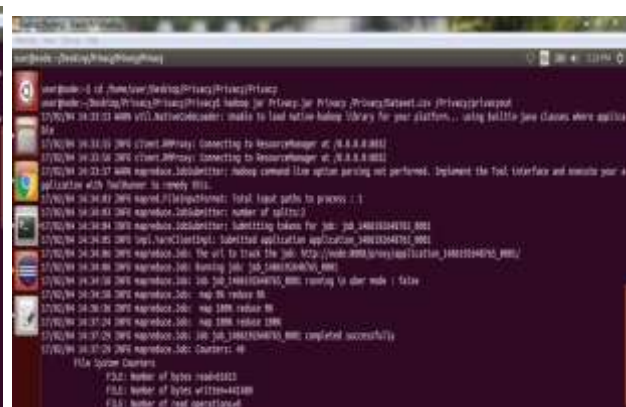


Fig.6 Pressure sensor value through MapReduce

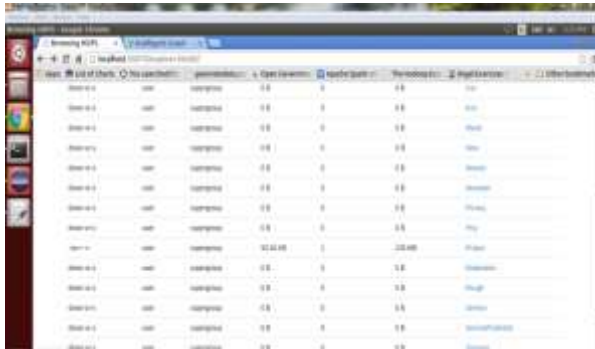


Fig.7 Data Analytics



Fig.8 3D bar Graph of Privacy in big data



Fig.9 3D Pie Graph of Privacy in big data

V. CONCLUSION

The amount information is growing everyday and it's impossible to imagine future generation applications while not producing and execution data driven algorithms. During this work conducted a comprehensive survey on the privacy problems once handling big data. In this proposed work investigated privacy challenges in every section of big data life cycle and mentioned some benefits and downsides of existing privacy preserving technologies within the context of big data applications. Lots of works are done to preserve the privacy of users from information generation to processing; however there still exist many open problems and challenges.

REFERENCES

- [1] Rathore, Muhammad Mazhar Ullah, Anand Paul, Awais Ahmad, Bo-Wei Chen, Bormin Huang, and Wen Ji. "Real-time big data analytical architecture for remote sensing application" IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8, no. 10 (2015): 4610-4621.
- [2] Katware, Ajay, Pankaj Patil, and Yogesh Lokare. "Efficient Analytical Architecture in Real-time Big Data for Remotely Sensing Application Using Hadoop Framework"
- [3] Shiva, P., P. Karthiga, and X. J. J. Anitha. "A System To Analysis Real Time Bigdata Using Topdown Specialization" International Education and Research Journal 2, no. 6 (2016).
- [4] Ancy, S. "Deployment of Architecture of Big Data in Real-Time for Implementation of Remote Sensing Application" Global Journal for Research Analysis 5, no. 6 (2016).
- [5] Kamble, Asmita, Pooja Chavan, Ashlesha Gaikwad, Shambala Honmode, and Sumezha Jagtap. "Processing of Real Time Big Data for Fremote Sensing Applications" International Journal of Engineering Science 4696 (2016).
- [6] Kolozali, Sefki, Maria Bermudez-Edo, Daniel Puschmann, Frieder Ganz, and Payam Barnaghi. "A knowledge-based approach for real-time IoT data stream annotation and processing" In Internet of Things Cyber, Physical and Social Computing (CPSCom), IEEE, pp. 215-222. IEEE, 2014.
- [7] D. Agrawal, S. Das, and A. E. Abbadi, "Big Data and cloud computing: Current state and future opportunities" in Proc. Int. Conf. Extending Database Technol. (EDBT), 2011, pp. 530-533.
- [8] J. Cohen, B. Dolan, M. Dunlap, J. M. Hellerstein, and C. Welton, "Mad skills: New analysis practices for Big Data" PVLDB, vol. 2, no. 2, pp. 1481-1492, 2009.
- [9] J. Dean and S. Ghemawat, "Mapreduce: Simplified data processing on large clusters" Commun. ACM, vol. 51, no. 1, pp. 107-113, 2008.
- [10] H. Herodotou et al., "Starfish: A self-tuning system for Big Data analytics," in Proc. 5th Int. Conf. Innovative Data Syst. Res. (CIDR), 2011, pp. 261-272.
- [11] K. Michael and K. W. Miller, "Big Data: New opportunities and new challenges [guest editors' introduction]" IEEE Comput., vol. 46, no. 6, pp. 22-24, Jun. 2013.
- [12] C. Eaton, D. Deroos, T. Deutsch, G. Lapis, and P. C. Zikopoulos, "Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data". New York, NY, USA: Mc Graw-Hill, 2012.
- [13] R. D. Schneider, Hadoop for Dummies Special Edition. Hoboken, NJ, USA: Wiley, 2012.
- [14] A. Cuzzocrea, D. Saccà, and J. D. Ullman, "Big Data: A research agenda" in Proc. Int. Database Eng. Appl. Symp. (IDEAS'13), Barcelona, Spain, Oct. 09-11, 2013.
- [15] R. A. Schowengerdt, "Remote Sensing: Models and Methods for Image Processing" , 2nd ed. New York, NY, USA: Academic Press, 1997.