

Big Data & Analytics: Security and Privacy challenges in Social Media

Naresh Kumar S¹, Shwetha S², Pramod kumar P³

Assistant Professor, Department of CSE, SR Engineering College, Warangal, India¹

Assistant Professor, Department of CSE, Sumati Reddy Institute of Technology for Women, Warangal, India²

Sr.Assistant Professor, Department of CSE, SR Engineering College, Warangal, India³

Abstract: Big data alludes to tremendous measure of advanced data gathered from various and distinctive sources. Security and protection will assume a critical part in big data research and innovation as we need to get information from numerous and diverse areas. Traditional security mechanisms are used to secure small scale static data, are insufficient. So the question arises which security and privacy technology is adequate for efficient access to big data.. In this paper, we concentrated on big data particular security and protection challenges. Fundamental desire from the engaged challenges is that it will expedite a novel concentration on big data infrastructure.

Keywords: Big data, security, privacy, social media.

Introduction

Big Data could not be portrayed just in terms of its size. However, to create a basic understanding, Big Data are datasets which can't be processed in conventional database ways to their size. This kind of data accumulation helps improve customer care service in many ways. However, such huge amounts of data can also bring forth many privacy issues, making Big Data Security a prime concern for any organization. Working in the field of data security and privacy, many organizations are acknowledging these threats and taking measures to prevent them

Big Data Example:

Social Media: 300M new photos/day uploaded on Face book; 300M Instagram users share 60M photos/day; >100 hrs/min video uploaded to YouTube



Fig. 1.1 Social media

I. CHARACTERISTICS OF BIG DATA

Big data can be depicted by the accompanying properties:

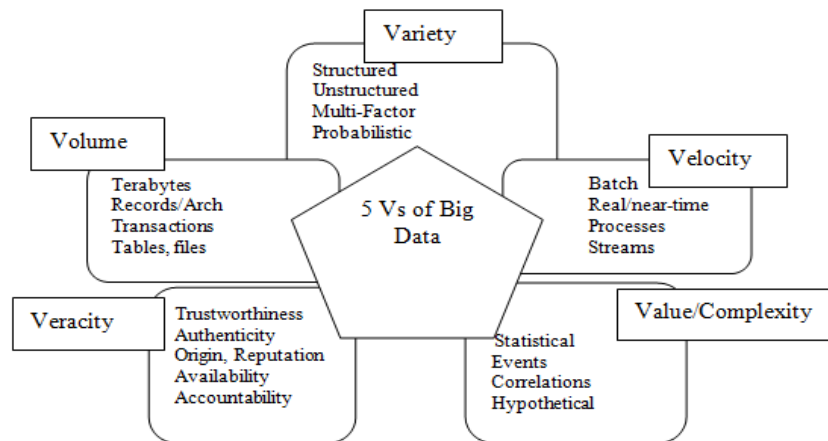


Fig. 2.1

1. Volume: The volume of big data is very important in the context. Many factors contribute to the increase in data volume: online transactions data, live streaming data from social media, customer feedback, data produced by employees, contractors, partners, and suppliers using social networking sites and data

collected from sensors etc.. In the past, excessive data volume was a storage issue. Other issues are how to determine relevant data within large data volumes and how to use big data analytics to generate value from relevant data.

A new style of IT emerging

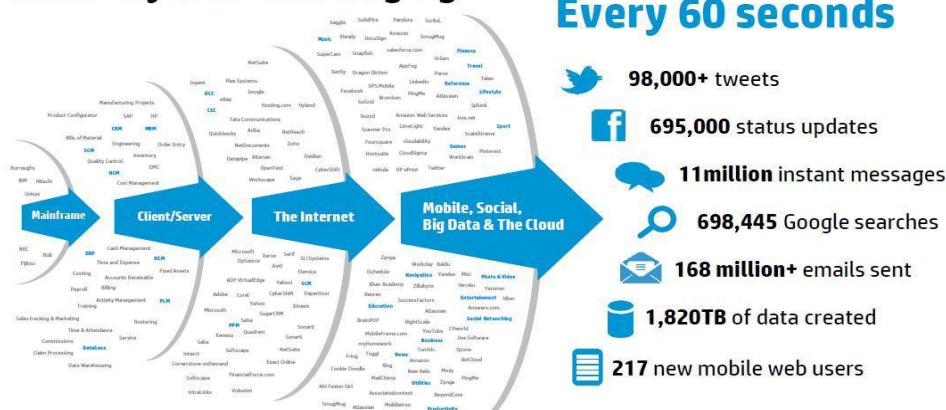


Fig. 2.2 Data gathered for every second

2. Variety: The next characteristic of big data is variety. Today data available in different types of formats (ex. Structural Data and Unstructured data). Structural data such as numeric data in traditional data bases and information created from business applications unstructured data such as text documents, email, video, audio, online transaction. Merging

managing and accessing different varieties of data is not easy task still.

3. Velocity: The term velocity means how fast the data is being produced and how fast the data needs to be processed to meet the demand and challenges. Data velocity is a challenge for most enterprises.

4. Variability: The term variability of big data refers to inconsistency of data. Along with the velocity and

varieties of data, data flows can be highly inconsistent with periodic peaks.

5. Complexity: Complexity of data needs to be considered, especially when large amount of data come from multiple sources. The data must be cleaned, merged, matched and transformed into required format before actual processing.

II. SECURITY AND PRIVACY CHALLENGES IN BIG DATA

Why Big Data Security Issues are Surfacing

Big data is nothing new to large organizations; however, it's also becoming popular among smaller and medium sized firms because of cost decrease and provided ease to manage data. Cloud-based storage has facilitated data mining and collection. We need to consider challenges to privacy and security threats while integrating cloud storage and big data.

The reason behind such ruptures may likewise be that security applications that are intended to store certain amounts of data can't the massive volumes of data that the previously mentioned datasets have. Additionally, these security advances are inefficient to oversee dynamic data and can control static data as it were. Therefore, just a regular security check cannot recognize security patches for continuous streaming data. For this reason, we need to consider privacy while data streaming and big data analysis.

Protecting Transaction Logs and Data

Data stored in storage medium, such as transaction logs and other sensitive data, may have anecdotal levels, but that's not enough. For example, the transfer of data between these levels gives the IT manager insight over the statistics which is being moved. Data size being continuously increased the scalability and availability makes auto-tiering essential for big data storage management. However, new challenges are being posed to big data storage as the auto-tiering method doesn't keep track of data storage location.

Validation and Filtration of End-Point Inputs

End-point devices are the principle factors for maintaining big data. Storage, processing and other required tasks are performed with the help of input data, which are given by end-points. Therefore, an organization should try to utilize an authentic and legitimate end-point device.

Securing Distributed Framework Calculations and Other Processes

Computational security and other digital assets in a distributed framework like Map Reduce capacity of Hadoop generally need security fortification. The two primary preventions for it are securing the mappers and protecting the data within the presence of an unauthorized mapper.

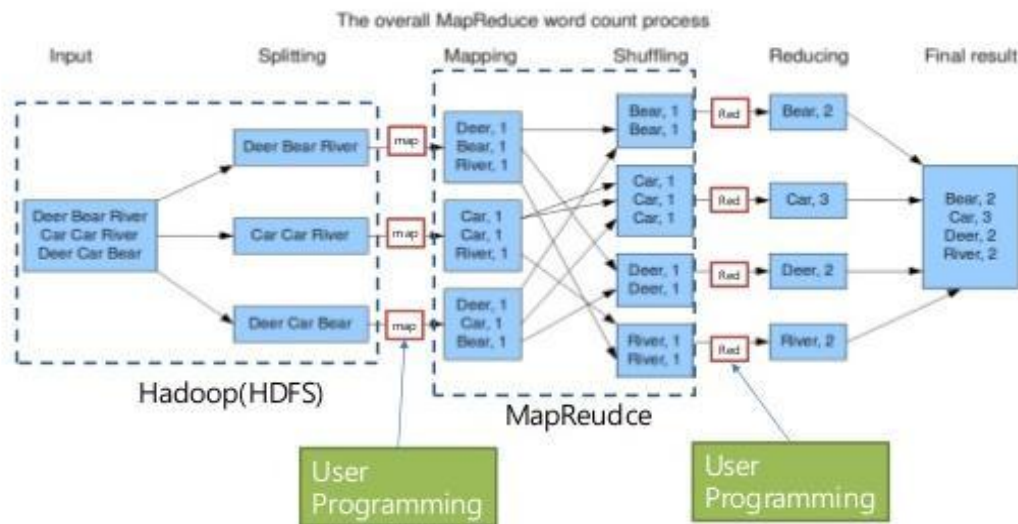


Fig 3.1 Map Reduce

Securing and Protecting Data in Real Time

Most organizations are unable to maintain regular checks, due to large amounts of data generation. However, it is most beneficial to carry out security checks and surveillance in real time.

Protecting Access Control Method Communication and Encryption

A secured data storage device is an intelligent tread in order to protect the data. However, most often data storage devices are vulnerable, it is necessary to encrypt the access control technique too.

Data Provenance

To catalog data, it is important to know about its origin, to determine the data origin accurately, validation, authentication, and access control could be gained.

Granular Auditing

Analyzing diverse kinds of logs could be advantageous and this information could be useful in perceiving any sort of digital assault or noxious action. Therefore, regular auditing can be beneficial.

Granular access control

Granular access control of big data stores by NoSQL databases or the Hadoop Distributed File System requires a obligatory access control and robust authentication process.

Privacy Protection for Non-Rational Data Stores

Data stores such as NoSQL have a lot of security vulnerabilities, which cause privacy threats. A prominent security flaw is that it is incapable to encrypt data during the tagging or logging of data or while disseminating it into different groups, when it is streamed or collected.

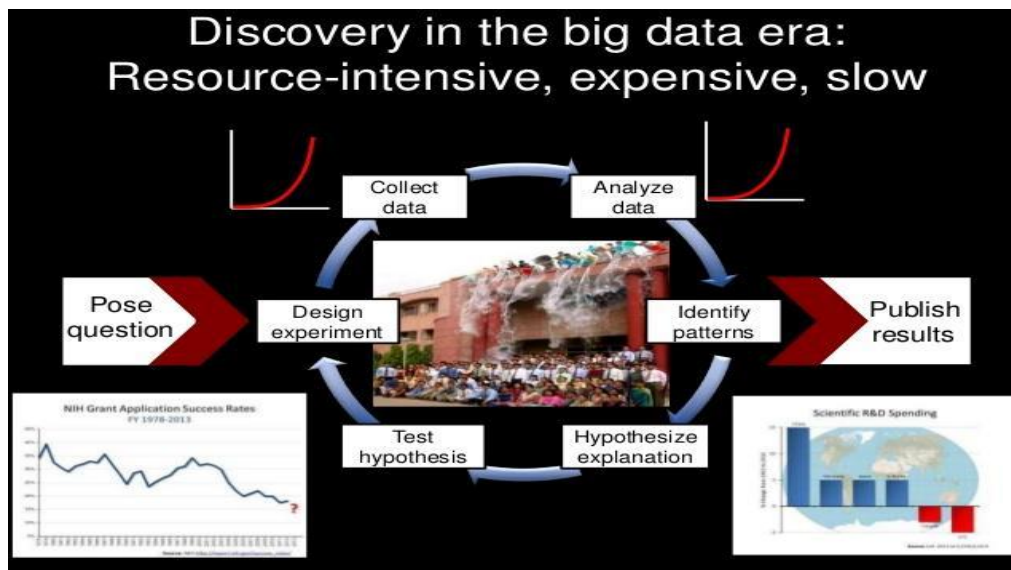


Fig. 3.2 Analysis or knowledge discovery in big data

III.SECURITY, TRUST AND PRIVACY FOR SOCIAL NETWORKING BIG DATA

Social media is the next battleground. Big social data can serve as: an early-warning system of trouble

ferment and a leading indicator of imminent action by a potential troublemaker.

Social Networking Big Data is a collection of very huge data sets with an incredible assorted variety of

kind from social networks. The emerging paradigm of social networking and big data provides gigantic novel way to deal with effectively adopting advanced networking communications and big data analytic schemas by utilizing the existing methods. The quick advancement of Social Networking Big Data conveys progressive changes to our day by day lives and worldwide business, which has been addressed by recent research. However, as attackers are taking advantages of social networks to accomplish their malicious goals, the security issue is also a vital worry when using Social Networking Big Data in practice.

Due to the intricacy and assortment of the Social Networking Big Data, there are two significant aspects of Social Networking Big Data. One is how to conduct social network analysis based on Big Data; the other is how to utilize Big Data analytic technique to make sure security of social networks using different security method. In Social Networking Big Data we have to conscious on following issue:

- Modeling on malicious information propagation with social influence analysis
- Secure social networking application with social influence analysis
- Privacy in management and analysis of social networking big data
- Prevention of malware propagation in social networks
- Modeling on the secure mechanisms of social networks
- Novel secure solutions for designing, supporting and operating social networks
- Trust evaluation in social networks with big data
- Threat and vulnerability analysis in social networks

- Secure social network architecture with big data
- Privacy protection in social networks with big data
- Secure social networking applications with big data
- Security design for social networks in big data
- Models, methods, and tools for testing the security of social networks
- Trust management in social networks with big data
- Spam problems in social networks with big data
- Detection for malicious information propagation in social networks

IV. CONCLUSION

We are living in a digital world of big data where massive amounts of heterogeneous, autonomous, complex and evolving data sets are constantly generated at unprecedented scale. In this paper, we have highlighted the security and privacy problems that need to be addressed for making big data processing and computing infrastructure more secure. To handle big data and to work with it and obtaining profit from it a branch of science has come up and is evolving, called Data Science. Data Science is the branch of science that deals with finding data from enormous sets of data, mostly unstructured and semi structured, by virtue of data inference and study. It's an insurgency that is changing the world and discovers application crosswise over different industries like finance, manufacturing, healthcare, sports, retail and communication. Social networking companies like Face book, Twitter, Search engine and digital marketing companies like Bing ,Yahoo and Google, and finance and e commerce companies like Amazon and EBay will require a plenty of data scientists. To the extent security is concerned

the existing technologies are promising to develop as newer vulnerabilities to big data arise and the need for securing them rises

REFERENCES

- [1] Stephen Kaisler, Frank Armour, Alberto Espinosa, "Big Data: Issues and Challenges Moving Forward", 2013 46th Hawaii International Conference on System Sciences.
- [2] Pandey, Shweta Tokekar, Vrinda, "Prominence of Map Reduce in Big Data Processing Communication Systems and Network Technologies "(CSNT), 2014 Fourth International Conference on April 2014 .
- [3] Dean, J., Ghemawat, "Map Reduce: Simplified Data Processing on Large Clusters." 6th Symposium on Operating System Design and Implementation (OSDI), pp. 137-150 , 2004.
- [4] Ghemawat, S. Gobioff, H., Leung, S.T, the Google File System. In: 19th ACM Symposium on Operating Systems Principles, pp. 29-43. Bolton Landing, New York ,2003
- [5] Xindong Wu , Gong-Quing Wu and Wei Ding " Data Mining with Big data ", IEEE Transactions on Knowledge and Data Engineering Vol 26 No1 Jan 2014
- [6] Chao Deng, Ling Qian, Meng Xu, Federated Cloud-based Big Data Platform in Telecommunications, Workshop on Cloud Services,Federation, and the 8th Open Cirrus Summit, , September 21 2012, San Jose, CA, USA
- [7] Stonebraker, M. and J. Hong. 2012. "Researchers' Big Data Crisis; Understanding Design and Functionality", Communications of the ACM, 55(2):10-11
- [8] The Economist. 2010. "Data, Data Everywhere", (online edition, February) <http://www.economist.com/node/15557443>

[9] <http://www.dataiversity.net/common-big-data-management-issues-solutions/> The Most Common Big Data Management Issues (And Their Solutions). By: A.RGuess. July 15 2014.

[10] Anuradha, G., "Suggested techniques for clustering and mining of data streams", Published in: Circuits, Systems, Communication and Information Technology Applications (CSCITA), 2014 International Conference on Date of Conference:4-5 April 2014

Authors



S. Naresh Kumar is working as a Assistant Professor in the Department of Computer Science & Engineering, SR Engineering College, Warangal, India. He received his M.Tech. Degree in Information Technology at JNTU Hyderabad, has 14 years of teaching experience. Mr.S Naresh Kumar research has been published in various international conferences and journals. His interested research area is Database, Network Security, Networking etc.



S Shwetha is working as a Assistant Professor in the Department of Computer Science & Engineering, Sumati Reddy Institute of Technology for Women, Warangal, India. She received her M.Tech. Degree in Software Engineering at JNTU Hyderabad, has 4 years of teaching experience. Mrs.S Shwetha research has been published in various international conferences and journals. Her interested research area is Networking etc.



Mr. Pramod Kumar P is working as a Senior Assistant Professor in the Department of Computer Science & Engineering, SR Engineering College, Warangal, India. He received his M.Tech. Degree in Software Engineering with distinction at JNTU Hyderabad. Mr. P Pramod Kumar has taught primarily introductory Programming courses, courses in Mobile communication, and courses that focus on professional communication for over 13 years. His research interests lie in the areas of Data Mining, Multimedia Analysis, Information Retrieval, Mobile Computing, Algorithms and Adhoc Networks. Mr.P Pramod Kumar's research has been published in various international conferences and journals.