Dated: 19 April 2017
Author: Ayesha Asghar

Thesis Outline

BIG DATA MANAGEMENT, SECURITY & PRIVACY CHALLENGES

I. Introduction

A. Background

1. Definition: Big Data refers to humongous digital data which enables research and decision-support applications with extraordinary value.

2. Big Data attributes: The technological advancement and social indulgence of the information have changed to the fast generating data which has five attributes volume, velocity, variety, veracity, and value.

3. Sources of Big Data: Big Data is massive and created from several fields in the engineering, namely; Sciences, Internet of Things and Businesses.

B. Thesis statement: Big Data will face management, security and privacy challenges.

C. Expanded thesis: Businesses would be multiplying and become further reliant on Big Data, but at the same time, security, scalability, and privacy in Big Data would be a major concern.

II. Big Data Management, Security & Privacy Challenges

A. Big Data faces storage challenge.

1. An important concern is for the life science guys to create massive amounts of data and then to store it till needed. As more the data, better the analysis. Storing videos requires a lot of space, and high definition and 3D is non-linear data, which means more and more space is needed to store data and due to the humongous amount of data generated every day some industries are facing this issue (Robinson, 2012).

2. Big Data is big news, but many companies and organizations are struggling with the challenges of Big Data storage (RAID, 2015).

3. The way Big Data is stored and processed it becomes overwhelming to secure and protect it (Ryoo, 2016).

4. In the case of any unauthorized access to Big Data would be extremely damaging as the personal data of people would be accessible and released. In distributed environment, an application may need several datasets from various data centers and therefore confront the challenge of privacy protection (Jain et al., 2016).

B. Data transmission does not have added security in Big Data always.

1. Data and transaction logs are stored in multi-tiered storage media. Manually moving data between tiers gives the IT manager direct control over exactly what data is moved and when. However, as the size of data set has been, and continues to be, growing exponentially, scalability and availability have necessitated auto-tiering for Big Data storage management. Auto-tiering solutions do not keep track of where the data is stored, which poses new challenges to secure data storage. New mechanisms are imperative to thwart unauthorized access and maintain the 24/7 availability (CSA, 2012).

2. Extra security procedures and methods are required in automated data transfer, which often is unavailable (Panchenko, 2016).

C. Securing access control is an issue faced in managing Big Data.

1. When a system receives an enormous amount of data, it should validate it to ensure its integrity and accuracy, which is often not the case (Panchenko, 2016).

2. Some organizations cannot – or do not – Institute access controls to divide the level of confidentiality within the company (Panchenko, 2016).

3. The organizations when dealing with large amounts of data should make an intensive effort to obey the regulatory requirements. It is sensible to have added protection measures by ensuring the generation and storage of all user activities and system events. This data would be beneficial and is required for performing user and system audits (Pierson, 2016).

D. Data analysis is not performed consistently.

1. Owing to the massive size of Big Data the suggested thorough audits are not carried out on a routine basis (Panchenko, 2016).

E. The size of Big Data brings difficulty in its monitoring.

    1. Big Data is not routinely monitored and tracked due to its size (Panchenko, 2016).

    2. Real-time monitoring is designed to alert the company at the very first sign of an attack; however, the amount of feedback from SIEM (security information and event management) system, whose aim is to provide the big-picture feedback of the data, is enormous. Companies that have the resources to closely monitor this feedback and separate the real attacks from the false ones are rare (Vickery, 2016).

F. The architecture of Big Data is complex and faces security issues.

    1. One of the leading causes of Big Data security problems can be summed up in one word: variety. The more complex data sets are, the more difficult it is to protect. Big Data diversity can come from several different areas. Data forms could be structured or unstructured. Data sources may originate from email files, servers, cloud applications, mobile device data, and much more. Data consumers can be anyone from high-level executives to customers and business users. More diverse data simply means more work is needed to protect it (Buckley, 2015).

    2. In Big Data architecture, the data is usually stored on multiple tiers, depending on business needs for performance vs. cost. For instance, high-priority "hot" data will usually be stored on flash media. So, locking down storage will mean creating a tier-conscious strategy (Gross, 2016).

G. To provide complete security for Big Data is a huge challenge.

    1. Most distributed systems' computations have only a single level of protection, which is not recommended (Panchenko, 2016).

    2. Access control encryption and connections security can become dated and inaccessible to the IT specialists who rely on it (Panchenko, 2016).

    3. One of the best strategies for controlling access to information or physical space is having a single access point, which is much easier to secure than hundreds of them. The fact that Big Data is stored in such widely-spread places runs against this principle. Its vulnerability is far higher because of its size, distribution and a broad range of access (Ryoo, 2016).

    4. Several Big Data security issues center around the fact that user and service authentication protocols in native Hadoop are somewhat weak. This leaves Hadoop systems open to the risk of malicious data inputs and edits (Pierson, 2016).

5. Native Hadoop distributions offer data encryption capabilities for data-at-rest, but it's a bit trickier to protect data-in-motion (Pierson, 2016).

6. These are the heart of many Big Data environments; they find the patterns that suggest business strategies. For that very reason, it's particularly important to ensure they're secured against not just external threats, but insiders who abuse network privileges to obtain sensitive information – adding yet another layer of Big Data security issues (Gross, 2016).

7. One would think the increase in cyber-attacks would mean increased spending on IT security to protect Big Data, but that's not always the case. Most experts agree that around 10 percent of a company's IT budget should be spent on security, but the average is under 9 percent. Without the requisite resources, organizations will find it harder to protect their companies' data (Buckley, 2015).

H. In Big Data, a compromise on the customer's privacy is usual.

1. Unethical IT specialists practicing information mining can gather personal data without asking users for permission or notify them (Panchenko, 2016).

2. When law enforcement agencies collect information in the name of improved security, everyone is treated as a potential criminal or terrorist, whose information may eventually be used against them. The authorities already know a lot about us but could ask companies such as Apple, Google, and Amazon to provide more intelligence such as a decrypted version of our data, what search terms we are using and what we are buying online (Ryoo, 2016).

3. Companies are eager to deliver targeted advertising to you and tracking your every online move. Big Data makes this tracking easier to do, less expensive and more easily analyzed (Ryoo, 2016).

4. Many customers may feel uncomfortable with the idea that businesses can collect such detailed information about their identities, behaviors, motivations, and other sensitive facts. Some companies respond to these concerns with data masking policies and aggregating data sets, but those methods aren't always effective. Those with the right equipment could put data sets back together to re-identify individuals. Taking away privacy could lead to increased security risks, especially if the data involved contains sensitive corporate information (Buckley, 2015).

5. You would be surprised to know how the amount of data collected about each person can be processed and analyzed to provide a surprisingly complete picture. Attempts to make certain anonymous data are useless in protecting people's privacy, because there is so much data available, that

you can use some of it as a link for identification purposes. User information is in transit all the time, being accessed by the inside users, outside contractors, and business partners sharing it for research (Vickery, 2016).

6. When executing the security system of Big Data, it becomes a massive challenge to regard privacy concerns as the permission to control and analyze data remains (Vickery, 2016).

7. Big Data is a troubling manifestation of Big Brother by potentially enabling invasions of privacy, invasive marketing, decreased civil freedoms, and increase state and corporate control. A recent analysis of how companies are leveraging data analytics for marketing purposes identified an example of how a retailer could identify that a teenager was pregnant before her father knew. Similarly, anonymizing data for analytics is not enough to maintain user privacy (CSA, 2012).

I. Data skills gap are prevalent as Big Data is an emerging field.

1. There are not many people available who are experts in the field of Big Data, which leads to the shortage of data specialists and data scientists. The gravity of this situation rises as it 's hard to find the qualified people with necessary skills to handle the job (Buckley, 2015).

III. Case Study: Discussing a real-life scenario from industry to highlight the challenges and possible solutions.

IV. Conclusion

A. With all technological advancements, Big Data also has some gaps when it comes to privacy and security. The securing of all the Big Data components ensures the safety of Big Data itself.

B. The humongous size of Big Data calls for robust solutions, some tools & practices are readily available, but if not utilized in an integrated way they won't be useful in fully secure Big Data.

C. Privacy of customers can be achieved using real-time protection when the collection of data takes place.

D. Study and research encounter difficulty when it comes to creation and storage of Big Data. Challenges arise in data transfer and processing. Lastly, the usage and analysis of Big Data is an extensive trial. The need for high capability distributed architecture emerges for backing up Big Data's development in a secure setting protected from leaks.

E. To attain a fully secure and trustworthy environment for Big Data is a significant challenge for research purposes. Enabling protection between end users, applications, and Big Data is an emerging trend for the coming years.

References:

Buckley, J. (2015) *7 Big Data Security Concerns*. Available at: https://www.qubole.com/blog/big-data/big-data-security-concerns/ (Accessed: 18 April 2017)

CSA (2012) *Top Ten Big Data Security and Privacy Challenges.* Available at: https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Big_Data_Top_Ten_v1.pdf (Accessed: 17 April 2017)

Gross, G. (2016) 9 *Key Big Data Security Issues.* Available at: https://www.alienvault.com/blogs/security-essentials/9-key-big-data-security-issues (Accessed: 18 April 2017)

Jain, P., Gyanchandani, S. and Khare, N. (2016) *Big Data privacy: a technological perspective and review*. Available at: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-016-0059-y (Accessed: April 18 2017)

Panchenko, A. (2016) *Nine Main Challenges in Big Data Security*. Available at: http://www.datacenterknowledge.com/archives/2016/01/19/nine-main-challenges-big-data-security/ (Accessed: April 17 2017)

Pierson, L. (2016) *Big Data Security Issues in the Enterprise.* Available at: http://www.bmc.com/blogs/big-data-security-issues-enterprise/?blog-redirect-sw (Accessed: 18 April 2017)

RAID, INC. (2015) *Big Data Storage Challenges.* Available at: http://www.raidinc.com/blog/big-data-hpc/big-data-storage-challenges (Accessed: 17 April 2017)

Robinson, S. (2012) *The Storage and Transfer Challenges of Big Data.* Available at: http://sloanreview.mit.edu/article/the-storage-and-transfer-challenges-of-big-data/ (Accessed: 17 April 2017)

Ryoo, J. (2016) *Big Data security problems threaten consumers' privacy*. Available at: http://theconversation.com/big-data-security-problems-threaten-consumers-privacy-54798 (Accessed: 18 April 2017)

Sharma, A. (2011) *Big Data storage management challenges and how to deal with them.* Available at: http://www.computerweekly.com/tip/Big-data-storage-management-challenges-and-how-to-deal-with-them (Accessed: 17 April 2017)

Vickery, N. (2016) *Big Data Security Issues: Main Challenges in 2016*. Available at: http://www.datasciencecentral.com/profiles/blogs/big-data-security-issues-main-challenges-in-2016 (Accessed: 18 April 2017)