UDC 004.4'236

AN ANALYSIS OF MAIN SECURITY ISSUES OF BIG DATA

Dimiter G. Velev
Department of Information Technologies and Communications
University of National and World Economy, Sofia, Bulgaria
e-mail: dvelev@unwe.acad.bg

1. INTRODUCTION

Every day enormous amounts of data are generated daily smartphones, sensors, video cameras and other connected devices, adding them to the large store of information from traditional sources.

Big data is a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The challenges include capture, curation, storage, search, sharing, transfer, analysis, and visualization. The trend to larger data sets is due to the additional information derivable from analysis of a single large set of related data, as compared to separate smaller sets with the same total amount of data, allowing correlations to be found to determine business trends, quality of research, prevent diseases, link legal citations, combat crime and determine real-time roadway traffic conditions [1].

While the term may seem to reference the volume of data, that isn't always the case. The term big data -- especially when used by vendors - may refer to the technology (which includes tools and processes) that an organization requires to handle the large amounts of data and storage facilities [8]. The data is typically loosely structured data that is often incomplete and inaccessible.

As an emerging technology Big Data offers not only new possibilities, but it poses many real problems, the most important of which is Big Data security.

2. ARCHITECTURE OF BIG DATA

Analysis of literature, emerging trends, software tools and use-case scenarios shows that the architecture of Big Data is characterized with many dimensions [4, 5, 6]:

- Volume - Enterprises are awash with ever-growing data of all types, easily amassing terabytes of information. Many factors contribute to the increase in data volume – transaction-based data stored through the years, text data constantly streaming in from social media, increasing amounts of sensor data being collected, etc. In the past, excessive data volume created a storage issue. But with today's decreasing storage costs, other issues emerge, including how to determine relevance amidst the large volumes of data and how to create value from data that is relevant.
- Velocity - According to Gartner, velocity "means both how fast data is being produced and how fast the data must be processed to meet demand." Reacting quickly enough to deal with velocity is a challenge to most organizations. For time-sensitive processes such as catching fraud, big data must be used as it streams into the enterprise in order to maximize its value.
- Variety - Big data is any type of data - structured and unstructured, coming in all types of formats – from traditional databases to hierarchical data stores created by end users and OLAP systems, to text documents, email, meter-collected data, video, audio, stock ticker data and financial transactions. By some estimates, 80 percent of an organization's data is not numeric.
- Veracity - 1 in 3 business leaders do not trust the information they use to make decisions. Establishing trust in Big Data presents a huge challenge as the variety and number of sources grows.

- Variability - In addition to the increasing velocities and varieties of data, data flows can be highly inconsistent with periodic peaks. Seasonal and event-triggered peak data loads can be challenging to manage – especially with social media involved.
- Complexity - When dealing with huge volumes of data, it comes from multiple sources. It is quite an undertaking to link, match, cleanse and transform data across systems. However, it is necessary to connect and correlate relationships, hierarchies and multiple data linkages. Data governance can help determine how disparate data relates to common definitions and how to systematically integrate structured and unstructured data assets to produce high-quality information that is useful, appropriate and up-to-date.

Big Data technologies not only support the ability to collect large amounts of data, they provide the ability to understand it and take advantage of its value. The goal of all organizations with access to large data collections should be precise processing of the most relevant data and using it for optimized decision making. New technology advancements will enable companies to use up to the most of the possibilities of Big Data [2, 3, 7, 14]:

- Cheap, abundant storage and server processing capacity.
- Faster processors.
- Affordable large-memory capabilities.
- New storage and processing technologies designed specifically for large data volumes, including unstructured data.
- Parallel processing, clustering, virtualization, large grid environments, high connectivity and high throughputs.
- Cloud computing and other flexible resource allocation arrangements.
- NoSQL - a common term to describe data stores that house different types of structured and unstructured data in high quantities. The data stores are not accessed through the standard SQL language. The main advantages of this approach are scalability and availability of the data together with the flexibility of the data storage. With a technology where each data store is mirrored across different locations in order to guarantee constant up-time and no loss of data, these systems are commonly used to analyze trends.

3. MAIN SECURITY ISSUES OF BIG DATA

Since Big Data changes the rules for businesses, the security risks have become much greater. The analysis shows that the main security issues surrounding Big Data could be divided into the following groups [9, 10, 12, 13]:

- Data breaches – With more transactions, conversations, interactions and data now online, breaches involving Big Data could have far-reaching consequences and mean reputational damage, legal liability and even financial ruin. Cyber resilience and preparedness strategies are crucial for Big Data.
- Data in the cloud – The pressure for businesses to quickly adopt and implement new technologies such as cloud services, often to support big data's challenging storage and processing needs, comes with unpredictable risks and consequences. Importing data into a Big Data store in the cloud can result in the removal of permissions or confidentiality restrictions on the original data.
- Privacy – As huge amounts of data are generated, stored and analyzed, privacy concerns are becoming an even larger issue. Businesses need to start planning for new data protection requirements.
- Consumerization – Together with the growth of Big Data is the proliferation of new mobile devices used to gather, store, access and transfer data. The challenge for businesses is in managing and securing personal devices brought into the workplace by employees and balancing the need for security with productivity.

- Interconnected supply chains – Organizations are part of often complex, global and interdependent supply chains, which can be their weakest link. There is a key role for information security in coordinating the contracting and provisioning of business relationships, including outsourcers, offshorers and supply chain and cloud providers.
- There exist allegations that the different NoSQL systems have reduced security to a minimum from their systems. This lack of security is considered their feature and built with the idea that database administrators do not need to trouble themselves with security concerns. Hence, Big Data security should take into account questions, such as model maturity of Big Data, software maturity**,** staff maturity**,** client software, data redundancy and dispersion.

Companies are working on recommendations for providing Big Data Security. For example, Forrester has created a framework to help security and risk professionals control Big Data [11]. Forrester's Big Data Security and Control Framework breaks down the problem of securing and controlling big data into three steps:

- Definition of data based on its toxicity - Toxic data is any data that could be damaging to an organization if it leaves that organization's control. This allows security to properly protect data based on its classification once it knows where that data is located in the enterprise. Discovery and classification are critical - data discovery locates and indexes big data and data classification catalogs data to make it easier to control.
- Dissection and analyzing the data - Extracting information from these massive data sets will prove invaluable to the organization efforts and it should also anticipate using this data more efficiently to prioritize security initiatives and more effectively place the proper security controls.
- Protecting the data - Access control ensures the right user gets access to the right data at the right time; Inspecting data usage patterns can alert security teams to potential abuses; Disposing of data when the company no longer needs.

Data-protection products should be used that support both data-masking and encryption technologies [14, 15]. Data masking obscures sensitive data elements within data stores with false information, but keeps it meaningful for the application logic. The encryption and the access-control technology must allow users with different credentials to have the appropriate selective access to data.

The use of Big Data Analytics, which represents performing increasingly sophisticated analysis on massive amounts of data, predominantly unstructured, has the potential to generate significant industry productivity growth [12, 15]. At the same time, it promises benefits for information security while also presenting increased risks. Big Data analytics have the potential to reduce the growing number of cyber security risks and increase business agility. Businesses eager to adopt these new technologies for business benefit will be well advised to set out clear good practice guidelines for Big Data.

4. CONCLUSION

Big Data is not just a matter of size, but an opportunity to find insights in new and emerging types of data and content, to make business more agile and to answer questions that were previously difficult to resolve.

Big Data is very powerful, but dangerous at the same time. Data falling into the wrong hands can have devastating consequences. The planning of Big Data security must start immediately since such timely and early security building will reduce costs, risks and deployment efforts.

5. ACKNOWLEDGMENTS

The authors express their gratitude to the Science Fund of the University of National and World Economy, Sofia, Bulgaria for the financial support under the Grant NI 1-8/2011, titled

"Methodology for the Implementation of Web-based Integrated Information System for Risk Assessment Due to Natural Hazards".

REFERENCES

1. Big Data, http://en.wikipedia.org/wiki/Big_data.
2. Cisco, Big Data: Big Potential, Big Priority, http://newsroom.cisco.com/release/1158061.
3. James Manyika et al., Big data: The next frontier for innovation, competition, and productivity, http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.
4. IBM, What is big data?, http://www-01.ibm.com/software/data/bigdata/.
5. SAS, Big Data, http://www.sas.com/big-data/.
6. IDC, Big Data in 2020, http://www.emc.com/leadership/digital-universe/iview/big-data-2020.htm.
7. Oracle, Big Data for the Enterprise, http://www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf.
8. Big Data, http://www.webopedia.com/TERM/B/big_data.html.
9. Dave Beulke, 3 Big Data Issues: Security, Governance and Archiving, http://davebeulke.com/3-big-data-issues-security-governance-and-archiving/.
10. Robert Westervelt, 5 Security Requirements For Big Data Hadoop Implementations, http://www.crn.com/240152524/printablearticle.htm.
11. John Kindervag, Contributor, A framework for big data security, http://searchsecurity.techtarget.com/magazineContent/A-framework-for-big-data-security.
12. Big data issues: Big data analytics offers both rewards and risks, http://searchsecurity.techtarget.com/magazineContent/Big-data-issues-Big-data-analytics-offers-both-rewards-and-risks.
13. Steve Durbin, Does big data mean big risks for businesses?, http://gigaom.com/2012/06/17/does-big-data-mean-big-risks-for-businesses/.
14. Noa Bar-Yosef, Examining The Security Implications of Big Data, http://www.securityweek.com/examining-security-implications-big-data.
15. Randall Gamby, Contributor , Security big data: Preparing for a big data collection implementation, http://searchsecurity.techtarget.com/tip/Security-big-data-Preparing-for-a-big-data-collection-implementation.