Chapter 98
Impact of Big Data on Security:
Big Data Security Issues and Defense Schemes

Kasarapu Ramani
Sree Vidyanikethan Engineering College, India

ABSTRACT

Big data has great commercial importance to major businesses, but security and privacy challenges are also daunting this storage, processing, and communication. Big data encapsulate organizations' most important and sensitive data with multi-level complex implementation. The challenge for any organization is securing access to the data while allowing end user to extract valuable insights. Unregulated access privileges to the big data leads to loss or theft of valuable and sensitive. Privilege escalation leads to insider threats. Also, the computing architecture of big data is not focusing on session recording; therefore, it is becoming a challenge to identify potential security issues and to take remedial and mitigation mechanisms. Therefore, various big data security issues and their defense mechanisms are discussed in this chapter.

INTRODUCTION

To harness the power of big data, one requires an infrastructure that can manage and process huge volumes of structured and unstructured data in real-time and can preserve data privacy and security.

Security management and data access are primary concerns for both persistent and moving data. Persistent data security is usually managed in layers: Physical, Network, Application, and at the database. Data moving between applications and organizations need additional security to protect the data in transit from unauthorized access.

The focus of this chapter is to high light the impact of security and privacy aspects on capacity and performance implications of big data solutions, which must be taken into account for such solutions to be viable. This chapter gives an overview of root causes for security and privacy challenges in big data and their consequences. It also describes the defense mechanism aspects that must be considered to make secured Big Data environment and suggests good practices to be followed for Big Data solutions, both now and in future.

DOI: 10.4018/978-1-5225-8176-5.ch098

BACKGROUND

What Is Big Data?

The Smartphones, Science facilities, Readers/ Scanners, Programs/ Software, Social media, and cameras are working as data generation points in Healthcare, Security Systems, Traffic Control, Manufacturing Sector, Sales, Sensors, Telecommunication, On-line gaming, Location-based services and Trading are leading to Big Data. The data generated and collected from different sources is doubling in every two years. The Big data become increasingly important in enterprises, government, and sciences. The process of capturing, storing, filtering, sharing, analyzing and visualizing this voluminous data itself is a challenge in Big Data. The purpose of Big Data is to generate value from stored large volumes of information by processing it using analytical techniques. The Big data helps in generating revenue, better services, strategic decisions, executive efficiency, specify needs, determine new trends, and flourish new products.

Characteristics of Big Data

Big data is characterized by 5 Vs: volume, velocity, variety, veracity and value. Volume represents huge data; velocity represents rapidity of data; variety indicates data collected from variety of sources with different data types; veracity defines consistency and trustworthiness of data; and value capture greater insights into data and supports in decision making from huge data sets. Defining characteristics of Big data will be helpful in obtaining hidden patterns available in data.

Big Data Types

Big Data includes structured, semi-structured and unstructured data.

- **Structured Data:** Represents the formal structure of data associated with relational databases or any other form of data tables and which can be generated by humans or software or computers. Structured data are often managed with SQL. Structured data are easy to input, query, store, and analyze. Examples of structured data include numbers, words, and dates.
- **Semi-Structured Data:** Also called self-describing structure, contain marks such as tags to separate semantic elements. Also, the records and fields of data can be arranged in hierarchies. XML, JSON, EDI, and SWIFT are few examples of this kind of data.
- **Unstructured Data:** Has no pre-defined data model. Now a day 80% of data accounts for unstructured data in any organization and which includes data from e-mails, video, social media websites and text streams.

Similar to structured data, this unstructured data can be generated by human or machine. Human-generated data includes text messages, e-mails, and social media data. Machine generated data includes radar and sonar data, satellite images, security, surveillance, traffic videos and atmospheric data. Often data are generated by a combination of these three groups.