

Chapter 1

The Basics of Big Data and Security Concerns

Sharvari C. Tamane

MGM's Jawaharlal Nehru Engineering College, India

Vijender K. Solanki

Institute of Technology and Science Ghaziabad, India

Madhuri S. Joshi

MGM's Jawaharlal Nehru Engineering College, India

ABSTRACT

The chapter is written on two important buildings, the basics of Big data and their security concern. The chapter is classifying in different sections. The chapter starts with the basic of big data and is concluded with security concern. The chapter is enriched with different category examples to make texts easy for author understanding. The chapter begins with the introduction of big data, their memory size followed by the examples. The chapter explains the category of big data in type of structured, semi-structured and unstructured data. The discussion on operational data service and big data application is also included to ensure the basic understanding to readers. The second portion of chapter which is based on security in big data. It's explaining the issues and challenges in big data. The section also focusing on the shift paradigm from cloud environment to big data environment changes and the problems encounter by organizations. The section discusses the framework issue and concluded with the necessity of understanding security in the big data, keeping in view of expansion of information technology infrastructure in the 21st century.

INTRODUCTION

Big data is a buzzword used to describe and transfer a massive volume of structured and/or unstructured data into knowledge. It is so huge that it is very difficult to process it using traditional database and software techniques. Big data helps companies to improve the performances of their applications and make faster intelligent decisions. Big data refers to Petabytes or Exabytes of data consisting of more than

DOI: 10.4018/978-1-5225-7501-6.ch001

Gigabytes to Terabytes of records of more than lac users, all from different sources. These sources may include: web, sales, customer contact center, social media, mobile data etc. One must be familiar with the other terms of data as shown in Table 1. Big data is always not referring to big volume of data but also refers to the technology that deals with the large amount of data and the infrastructure needed to store that data. When dealing with larger data sets organizations face difficulties in being able to capture, organize, manage and integrate big data as standard tools and procedures are not designed to search and analyze massive data sets. It may take so many minutes/days/years to transfer the data from one location to other. Businesses or enterprises expects more fast processing and transfer of data to perform different operations on it. Ninety percent of data currently available in the world is generated in last few years.

Characteristics of Big Data

Big data (Sharvari, 2015) is typically broken down by three main characteristics: Volume (i.e. quantity of big data), Velocity (i.e. processing speed of big data) and Variety (i.e. various categories of big data: structured or unstructured?). Even two more important characteristics are available, Veracity: (i.e. how accurate is that data in predicting business value?) and Value (conversion of data into actual business value). In a simple language, big data is defined as it is so large and comes in various categories that it becomes very difficult to process it using relational database management systems. It is the requirement of the industry that it should be provided with useful insight and gain correct content. Data should be processed with recent tools and produce important & meaningful information.

According to the study of IBM CMO (IBM, 2014), 2.5 exabytes of data is generated every day. Google generates approximately one Gigabytes of data, Twitter may generate greater than 500 Megabytes of users and 400 Megabytes of tweets and Facebook may generate more than 25 Terabytes of data daily. This leads to huge processing of data every day with new technologies like MapReduce.

Table 1. Bytes and bigger bytes

Data	Unit Size	Binary Size
1 bit	A binary digit	-
8 bits	1 byte or 10^0	2^3
1024 bytes	1 kilo bytes (1 KB) or 10^3	2^{10}
1024 KB	1 Mega Bytes (1MB) or 10^6	2^{20}
1024 MB	1 Giga Bytes (1 GB) or 10^9	2^{30}
1024 GB	1 Tera Bytes (1 TB) or 10^{12}	2^{40}
1024 TB	1 Peta Bytes (1 PB) or 10^{15}	2^{50}
1024 PB	1 Exa Bytes (1 EB) or 10^{18}	2^{60}
1024 EB	1 Zetta Bytes (1 ZB) or 10^{21}	2^{70}
1024 ZB	1 Yotta Bytes (1 YB) or 10^{24}	2^{80}
1024 YB	1 Bronto Bytes (1 BB) or 10^{27}	2^{90}
1024 BB	1 Geo Bytes or 10^{30}	2^{100}
...

10 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/the-basics-of-big-data-and-security-concerns/217819

Related Content

WSBen: A Web Services Discovery and Composition Benchmark Toolkit1

Seog-Chan Oh and Dongwon Lee (2009). *International Journal of Web Services Research* (pp. 1-19).

www.irma-international.org/article/wsben-web-services-discovery-composition/3131

Threat-Path Estimate-Based Watchword-Chunk Algorithm for Advanced Persistent Threat in the Cloud

Babu Pandipati and Praveen Sam R. (2022). *International Journal of Web Services Research* (pp. 1-32).

www.irma-international.org/article/threat-path-estimate-based-watchword-chunk-algorithm-for-advanced-persistent-threat-in-the-cloud/299021

Automatic Metadata Generation for Geospatial Resource Discovery

Miguel-Angel Manso-Callejo and Arturo Beltran Fonollosa (2012). *Discovery of Geospatial Resources: Methodologies, Technologies, and Emergent Applications* (pp. 78-110).

www.irma-international.org/chapter/automatic-metadata-generation-geospatial-resource/65110

Service Oriented Solution Modeling and Variation Propagation Analysis Based on Architectural Building Blocks

Liang-Jie Zhang and Jia Zhang (2013). *International Journal of Web Services Research* (pp. 39-61).

www.irma-international.org/article/service-oriented-solution-modeling-and-variation-propagation-analysis-based-on-architectural-building-blocks/108881

Mobile Web Services for P2P Computing

Fahad Aijaz (2010). *Developing Advanced Web Services through P2P Computing and Autonomous Agents: Trends and Innovations* (pp. 133-157).

www.irma-international.org/chapter/mobile-web-services-p2p-computing/43651