Chapter 3
Trends and Technologies in Big Data Processing:
An Overview

Amitava Choudhury

School of Computer Science, University of Petroleum and Energy Studies, Dehradun, India

Kalpana Rangra

School of Computer Science, University of Petroleum and Energy Studies, Dehradun, India

ABSTRACT

Data type and amount in human society is growing at an amazing speed, which is caused by emerging new services such as cloud computing, internet of things, and location-based services. The era of big data has arrived. As data has been a fundamental resource, how to manage and utilize big data better has attracted much attention. Especially with the development of the internet of things, how to process a large amount of real-time data has become a great challenge in research and applications. Recently, cloud computing technology has attracted much attention to high performance, but how to use cloud computing technology for large-scale real-time data processing has not been studied. In this chapter, various big data processing techniques are discussed.

INTRODUCTION

Technologies that run on big data have matured and gained impetus exploiting all the roles of big data. Most of the data sources are read-only if seen from the perspective of the analytical system. The history of data depicts that it changes over the series of time and is tampered knowingly or unknowingly (Padgavankar et al,2014). Thus, an efficient storage and processing mechanism is required for handling a tremendous amount of data that is being generated and collected each day. Data type and amount in human society is growing in amazing speed which is caused by emerging new services as cloud computing, internet of things and location-based services, the era of big data has arrived (Agarwal R. et al,2016;Yingyi Bu., et al, 2012). As data has been fundamental resource, how to manage and utilize big

data better has attracted much attention. Especially, with the development of internet of things, how to process large amount real-time data has become a great challenge in research and applications. Recently, cloud computing technology has attracted much attention with high-performance, but how to use cloud computing technology for large-scale real-time data processing has not been studied.

The processing techniques implied for big data work on terabyte or petabyte scales which is a collection of time-series data. The studies conclude that the data which changes over time is more useful for analysis and is preferred over the data belonging to a current state. The biggest challenges faced include an answer to two major questions. Firstly, how to store the vast amount of versatile data and Secondly how to deal with such a huge amount of heterogeneous data to get benefitted out of it. On the other hand, while dealing with big data few more issues need attention such as fraud detection, fault tolerance, and data security. Big data processing is done on large clusters of shared-nothing commodity architecture and needs the support of centers that can work with large-scale physical resources (Abouzeid A. et al, 2017; Ahmed M. Aly. Et al, 2012; Ivan et al, 2013). Before that, the applications that work on big data will need remote access to an increasingly diverse range of data resources. Not only this, there is the need for API's that should be able to hide the complexity of both data and hardware used.

Why big data?
Data is widely available. What is scarce is the ability to extract wisdom from it.

These lines from Hal Varian, Google's chief economist, 2010 correctly identify need to work on data. Big data is most commonly heard term these days. .Large data sets that are collected by various organizations and devices can be commonly referred to as big data. The term found its existence in 2000 and was given by Doug Laney((Katina M. et al, 2013).

The term big data stands for the data sets whose volume exceeds the capabilities of conventional tools to capture, manage and analyze data along with effective storage(Gracia G. et al, 2019). This data collected from various sources is of different size, types and different shapes. Thus, a proper mechanism is required to not only store but also process the data. The processing here can be preprocessing for storage or the mechanism to be followed for accessing and manipulating this large number of data. The data retrieved from various fields is complex such that traditional data processing mechanism are not inadequate to deal with the calculations that make the information out of data collected. Figure 1. describe the various sources of big data, mainly generated from various social networking websites and ecommerce Sources of big data:

- Social networking sites
- E-commerce site
- Weather Station
- Telecom company
- Share Market
- Airlines Industry
- Stock Exchange
- Search Engines

Related Content

Steganography Using Biometrics

Manashee Kalitaand Swanirbhar Majumder (2019). *Advanced Methodologies and Technologies in System Security, Information Privacy, and Forensics (pp. 326-347).*

www.irma-international.org/chapter/steganography-using-biometrics/213661

Hiding Information in the DNA Sequence Using DNA Steganographic Algorithms with Double-Layered Security

Vinodhini R. E.and Malathi P. (2022). *International Journal of Information Security and Privacy (pp. 1-20).*

www.irma-international.org/article/hiding-information-in-the-dna-sequence-using-dna-steganographic-algorithms-with-double-layered-security/300322

Video Surveillance Camera Identity Recognition Method Fused With Multi-Dimensional Static and Dynamic Identification Features

Zhijie Fan, Zhiwei Cao, Xin Li, Chunmei Wang, Bo Jinand Qianjin Tang (2023). *International Journal of Information Security and Privacy (pp. 1-18).*

www.irma-international.org/article/video-surveillance-camera-identity-recognition-method-fused-with-multi-dimensional-static-and-dynamic-identification-features/319304

Solutions for Securing End User Data over the Cloud Deployed Applications

Akashdeep Bhardwaj (2017). *Cybersecurity Breaches and Issues Surrounding Online Threat Protection (pp. 198-218).*

www.irma-international.org/chapter/solutions-for-securing-end-user-data-over-the-cloud-deployed-applications/173135

Secure Anonymous Systems and Requirements

 (2012). *Anonymous Security Systems and Applications: Requirements and Solutions (pp. 1-6).*

www.irma-international.org/chapter/secure-anonymous-systems-requirements/66332