

# Chapter 16

## Big Data: An Emerging Field of Data Engineering

**Piyush Kumar Shukla**

*University Institute of Technology RGPV, India*

**Madhuvan Dixit**

*Millennium Institute of Technology, India*

### ABSTRACT

*In this chapter, Big Data provide large-volume, complex structure, heterogeneous and irregular growing data sets include multiple and autonomous different resources. In this chapter, With the growing improvement of networking sites, image information storing capacity become big issue too, Big Data concept are most growing expanding in all technical area and knowledge engineering domains, including physical, medical and paramedical sciences. Here a data-driven method consist demand-driven aggregation of information and knowledge mining and analysis, user interest prototyping, security and privacy aspects has been presented.*

### INTRODUCTION

Big data is concept more uses from various companies. Some examples are related with oil and gas refineries and mining industries, online social networks, multimedia data and business related transactions. More amount of data collected from different increasingly efficient data storing various devices as well as stored on fast-growing mass storage, people are keep to search to find solutions to collect and process the information more efficiently, and to find various values from the mass at the same time. When referring to big data research strategy problems, people often

support the 4 v's -- volume, velocity, variety, and value. These pose support more brand-new challenges to computer scientists nowadays (Ahmed & Karypis, 2012).

In 2004, Wal-Mart claimed to have the very large data warehouse with approx 500 terabytes storage. In 2009, eBay announce storage amounted to eight PB. Two years later, the Yahoo warehouse totaled 170 PB. Since the rise of digitization, various technical enterprises have amassed burgeoning amounts of digital information, including trillions of bytes of data about their customers, suppliers and operations (Aral & Walker, 2012). Information volume is also growing exponentially

DOI: 10.4018/978-1-4666-8387-7.ch016

## Big Data

due to the spread of machine-oriented data (data records, web-log files, sensors information) and from growing human involvement within the social networks.

Basically the growth of information will never stop. As per the 2011 IDC Digital Universe Study, 130 EB of data were created and stored in 2006. The amount grew to 1,327 EB in 2011 and is projected to grow at 43.6% to 7,913 EB in 2016. The growth of information constitutes the “Big Data” concept – a technological concept brought about by the rapid rate of data expansion and parallel generalization in technology that have given rise to a contribution of software and hardware products that are enabling users to analyze this information to produce new and more granular desire levels of information (Machanavajjhala & Reiter, 2012). A decade of Digital Universe Growth: Storage in Exabytes has been illustrated in Figure 1.

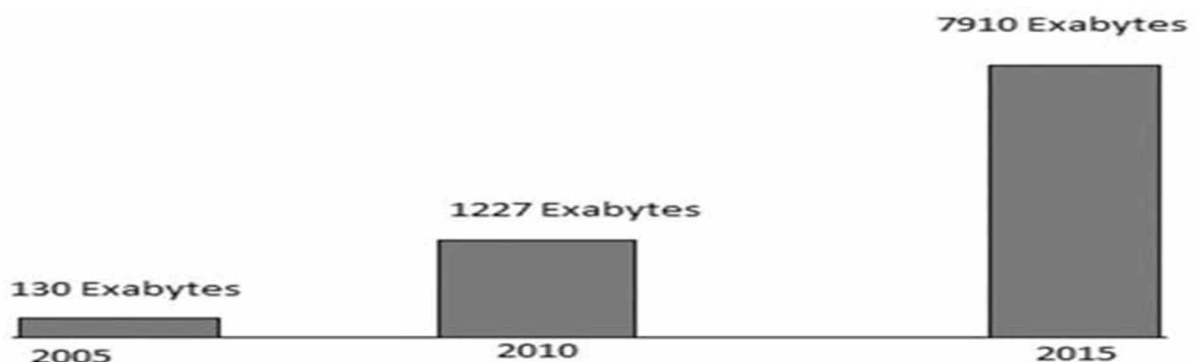
According to McKinsey, Big Data, perform to datasets whose sizes are beyond the ability of database software product to acquire, store, manage and study. There is no explicit definition of how big an image dataset should be in order to be considered large Big Data. New technology has to manage this Big Data concept. Basically, IDC defines Big Data technologies considered as a new generation of database technologies and architectures designed to extract data value from large volumes of a wide variety of data. Accord-

ing to Reilly, “Big data is information which exceeds the processing capacity of traditional database products. Basically, information is too big, moves too fast, or does not assemble the structures of desired database architectures. To gain value from these databases resource, there must be different way to process it” (Machanavajjhala & Reiter, 2012).

## CHARACTERISTICS OF BIG DATA

Big Data is not just about the volume of information but also includes information variety and velocity. Together, these three parameters form the three Vs of Big Data theory. Volume is synonymous form of the “big” in the term, “Big Data”. Volume is a relative term – some smaller-sized industries are likely to have mere gigabytes or terabytes of information storage as opposed to the PB or EB of data that big data global enterprises have. Data volume will continue to expand, regardless of the organization’s capacity. There is a natural definition for companies to store data of all categories: account data, biomedical data, research data and so on. Many of these organization datasets are within the terabytes range today but in early days they could reach PB or even EB (Aral & Walker, 2012). Data can come from a variety of resources and in a variety of category. With the expansion of

Figure 1. A decade of Digital Universe Growth: Storage in Exabytes (Machanavajjhala & Reiter, 2012)



17 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/big-data/134299](http://www.igi-global.com/chapter/big-data/134299)

## Related Content

---

### Recent Advances in Edge Computing Paradigms: Taxonomy Benchmarks and Standards for Unconventional Computing

Sana Sodanapalli, Hewan Shrestha, Chandramohan Dhasarathan, Puviyarasi T. and Sam Goundar (2021). *International Journal of Fog Computing* (pp. 37-51).

[www.irma-international.org/article/recent-advances-in-edge-computing-paradigms/284863](http://www.irma-international.org/article/recent-advances-in-edge-computing-paradigms/284863)

### Improved Distributed Energy Systems Based on the End-User Consumption Profile: A Review on How Consumers Can Drive the Energy Transition

Daniel Adrian Perez-Moscote and Mikhail Georgievich Tyagunov (2020). *Handbook of Research on Smart Technology Models for Business and Industry* (pp. 211-235).

[www.irma-international.org/chapter/improved-distributed-energy-systems-based-on-the-end-user-consumption-profile/259131](http://www.irma-international.org/chapter/improved-distributed-energy-systems-based-on-the-end-user-consumption-profile/259131)

### Resource Management in Sensor Cloud

Prashant Sangulagi and Ashok V. Sutagundar (2017). *Advancing Cloud Database Systems and Capacity Planning With Dynamic Applications* (pp. 158-175).

[www.irma-international.org/chapter/resource-management-in-sensor-cloud/174759](http://www.irma-international.org/chapter/resource-management-in-sensor-cloud/174759)

### FogLearn: Leveraging Fog-Based Machine Learning for Smart System Big Data Analytics

Rabindra K. Barik, Rojalina Priyadarshini, Harishchandra Dubey, Vinay Kumar and Kunal Mankodiya (2018). *International Journal of Fog Computing* (pp. 15-34).

[www.irma-international.org/article/foglearn/198410](http://www.irma-international.org/article/foglearn/198410)

### A Review of Quality of Service in Fog Computing for the Internet of Things

William Tichaona Vambe, Chii Chang and Khulumani Sibanda (2020). *International Journal of Fog Computing* (pp. 22-40).

[www.irma-international.org/article/a-review-of-quality-of-service-in-fog-computing-for-the-internet-of-things/245708](http://www.irma-international.org/article/a-review-of-quality-of-service-in-fog-computing-for-the-internet-of-things/245708)