

Chapter 3

Introduction of Big Data With Analytics of Big Data

Preeti Bala

Institute of Management Studies, Noida, India

ABSTRACT

This chapter explained big data and how to do the data analytics in big data along with the basics of big data with its common traits. This chapter described technology that is used to handle the big data like, NoSQL, Hadoop, MangoDB, MapReduce, etc. This chapter explained the challenges that we are facing to handle big data and discussed some live projects in big data.

INTRODUCTION

The new word came into the existence “Big Data” is due to the fact that we are creating a huge amount of data every day. The Birth of Big Data can be traced back to 1998 in a silicon graphics slide deck by John Mashey with the title “Big Data and the next Wave of infra Stress.” Before analysing challenges and other aspects of big data, let us examine where big data comes from? The big Data comes from us .Its future is at the mercy and whims of our fingertips. Big Data is generated when formerly no digitized processes become digitized.

That is our life is getting digitized; we now use online banking, online shopping .e-learning, emails, and instead of the traditional methods. Big data comes relatively new types of data sources that previously were not mined for. Big data is also turned as devices become more integrated into our everyday lives. Everyone has a mobile phones, camera, laptop, takes pictures and makes videos and then uploads them to social networks, which keeps on adding to the huge bulk of data .we generate lots of big data as we willingly share information with the world through social media in the form of tweets, comments likes and many more. (Bifet, 2013)

The key enables for the growth of big data can be summarised as follows:

1. Increase in storage capacities
2. Increase in processing power

DOI: 10.4018/978-1-6684-3662-2.ch003

Introduction of Big Data With Analytics of Big Data

3. Availability of data

Two important trends in the era of Big Data are as follows:

1. The digitization of virtually “everything “.
2. Advanced analytics technologies and techniques that enable organization to extract insights from data with high levels of sophistication, speed, and accuracy.

Big Data generated usually share some or all the following features:

1. Digitally generated –IT will be a digital data
2. Passively produced –It may be a by product of our interaction with digital services in our day to day life
3. Automatically collected –There may be some system that collects and stores relevant data generated.
4. Geographically or temporally trackball- Mobile phone location data or call duration time.
5. Continuously analysed – Information can be analysed in real time for human well-being.
6. Doug Laney was the first person who highlighted the three dimensions of big data the three V”S

BIG DATA “THE THREE V’S”

1. **Volume:** Nowadays data is expanding rapidly to fill up the available space for storage .we are very familiar at the rate at which we copy videos and songs from one mobile to other mobile. Every minute, 72 h of videos are uploaded on you tube. So the data in increasing at a rate of pet byte datasets these days and Exabyte is not farway.
2. **Variety:** From excel tables and databases, data structure has changed to lose its structure and hundreds of formats such as pure text, photo audio, video, web, GPS data sensor data, relational database s, documents, SMS, pdf, and flash are added to the data set .No one no longer has control over the input data format. No structure can be imposed like in the past to keep control over analysis .As new applications are introduced, new data formats come up.
3. **Velocity:** In earlier days, companies used to batch process the data when the data were of the same kind .this is now becoming impossible because with the new sources of data such as social and mobile applications, batch processing breaks down .The data is now streamed into the server in real time in a continuous fashion, and the result is only useful if the delay is very short (Ward, 2013).

And while they cover the key attributes of big data, we believe organizations need to consider an important fourth dimension: veracity/variability .Inclusion of the veracity as the fourth Big Data attribute emphasizes the importance of addressing and managing the uncertainty inherent within some type of data.

One more V can also be included along with the 4-V definition of big data to make it complete.

Value:- The economics /business value of different data may vary ;the challenge is in identifying what data is valuable in extracting information out of that .Analyzing huge data sets can reveal stunning insights on consumer preferences and behaviour and can guide in business decision making and drive in revenue.

11 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/introduction-of-big-data-with-analytics-of-big-data/290976

Related Content

Clustering and Unsupervised Learning

Divyani Tirthyani, Sunil Kumar and Shally Vats (2024). *Recent Trends and Future Direction for Data Analytics* (pp. 119-139).

www.irma-international.org/chapter/clustering-and-unsupervised-learning/347268

Machine Learning Techniques for Healthcare Applications: Early Autism Detection Using Ensemble Approach and Breast Cancer Prediction Using SMO and IBK

Rajamohana S. P., Dharani A., Anushree P., Santhiya B. and Umamaheswari K. (2019). *Cognitive Social Mining Applications in Data Analytics and Forensics* (pp. 236-251).

www.irma-international.org/chapter/machine-learning-techniques-for-healthcare-applications/218400

Detecting Human Diseases Relatedness: A Spreading Activation Approach Over Ontologies

Said Fathalla (2020). *Data Analytics in Medicine: Concepts, Methodologies, Tools, and Applications* (pp. 471-486).

www.irma-international.org/chapter/detecting-human-diseases-relatedness/243128

SBASH Stack Based Allocation of Sheer Window Architecture for Real Time Stream Data Processing

Devesh Kumar Laland Ugrasen Suman (2020). *International Journal of Data Analytics* (pp. 1-21).

www.irma-international.org/article/sbash-stack-based-allocation-of-sheer-window-architecture-for-real-time-stream-data-processing/244166

Voluntary Reporting of Performance Data: Should it Measure the Magnitude of Events and Change?

Vahé A. Kazandjian (2018). *International Journal of Big Data and Analytics in Healthcare* (pp. 27-37).

www.irma-international.org/article/voluntary-reporting-of-performance-data/209739