Chapter 6
Research Challenges in Big Data Analytics

Sivamathi Chokkalingam
Bharathiar University, India

Vijayarani S.
Bharathiar University, India

ABSTRACT

The term Big Data refers to large-scale information management and analysis technologies that exceed the capability of traditional data processing technologies. Big Data is differentiated from traditional technologies in three ways: volume, velocity and variety of data. Big data analytics is the process of analyzing large data sets which contains a variety of data types to uncover hidden patterns, unknown correlations, market trends, customer preferences and other useful business information. Since Big Data is new emerging field, there is a need for development of new technologies and algorithms for handling big data. The main objective of this paper is to provide knowledge about various research challenges of Big Data analytics. A brief overview of various types of Big Data analytics is discussed in this paper. For each analytics, the paper describes process steps and tools. A banking application is given for each analytics. Some of research challenges and possible solutions for those challenges of big data analytics are also discussed.

INTRODUCTION

The term Big Data refers to large-scale information management and analysis technologies that exceed the capability of traditional data processing technologies (Cloud Security Alliance 2013; Duggal & Paul, 2013). Big data sizes are increasing, ranging from a few dozen terabytes in 2012 to today many petabytes of data in a single data set. Big data applications are a great benefit to organizations, business, companies and many large scale and small scale industries. Some of the real time examples of Big Data are Credit card transactions made all over the world with respect to a particular Bank, Walmart customer transactions, Facebook users generating social interaction data, The New York stock exchange, Ancestry.com (the genealogy site), and The Internet Archive. Consider the following review. Today in every minute about (Skinner, 2015)

- More than 204 million email messages are passed.
- Over 2 million Google search queries are requested.
- 48 hours of new YouTube videos are downloaded.
- 684,000 bits of content shared on Facebook.
- More than 100,000 tweets are created.
- $272,000 spent on e-commerce.
- 3,600 new photos shared on Instagram.
- Nearly 350 new WordPress blog posts are created.

Advantages of Big Data

There are several advantages of big data. Some of the significant advantages of big data are:

- It can handle huge volume of data with high velocity and more variety.
- Easy integration of structured and unstructured data
- Ability to process semi-structured and unstructured data
- Easy to perform data analysis.
- Decreased cost of storage.
- Improved processing speed.
- Able to map the entire data landscape across a company with Big Data tools, thus allow analyzing the threats.
- Allows ever-narrower segmentation of customers and therefore possible to attain much more precisely tailored products or services.
- Sophisticated analytics can substantially improve decision-making, reduces risks and discovers valuable insights.

Need for New Technology

Some of the traditional data mining technologies fail to provide the tools to support big data. The reasons are

- Storing a very large quantity of data was not economically feasible.
- Cannot handle a very huge dynamic data.
- Unable to handle variety of data simultaneously.
- Performing analytics and complex queries on large, structured data sets was inefficient.
- Unable to analyze and manage unstructured data.
- Unable to integrate structured and unstructured data.
- Unable to discover information from unstructured data.

LITERATURE REVIEW

For literature review, see Table 1.

Related Content

Strategy as Action: From Porter to Anti-Porter
Milan Zeleny (2010). *International Journal of Strategic Decision Sciences (pp. 1-22).*
www.irma-international.org/article/strategy-action-porter-anti-porter/40996

Predictive Analytics in Digital Signal Processing: A Convolutive Model for Polyphonic Instrument Identification and Pitch Detection Using Combined Classification
Josh Weese (2014). *Emerging Methods in Predictive Analytics: Risk Management and Decision-Making (pp. 223-253).*
www.irma-international.org/chapter/predictive-analytics-in-digital-signal-processing/107908

A Cognitive Analytics Management Framework (CAM-Part 1): SAMAS Components, Leadership, Frontier Performance Growth, and Sustainable Shared Value
Ibrahim H. Osmanand Abdel Latef Anouze (2017). *Decision Management: Concepts, Methodologies, Tools, and Applications (pp. 151-227).*
www.irma-international.org/chapter/a-cognitive-analytics-management-framework-cam-part-1/176755

Decision-Making Approach for Catastrophic Scenario Selection in Disaster Recovery Planning
Jose Emmanuel Ramirez-Marquezand John V. Farr (2009). *International Journal of Decision Support System Technology (pp. 36-51).*
www.irma-international.org/article/decision-making-approach-catastrophic-scenario/3899

Planning Support System Project Management
 (2020). *Utilizing Decision Support Systems for Strategic Public Policy Planning (pp. 202-217).*
www.irma-international.org/chapter/planning-support-system-project-management/257630