



IGNITED MINDS
Journals

*International Journal of
Information Technology
and Management*

*Vol. X, Issue No. XVI,
August-2016, ISSN 2249-
4510*

**BIG DATA AND BIG DATA ANALYTICS: BASIC
CONCEPT AND PERSPECTIVES**

AN
INTERNATIONALLY
INDEXED PEER
REVIEWED &
REFEREED JOURNAL

Big Data and Big Data Analytics: Basic Concept and Perspectives

Nitin Kumar Saran*

Research Scholar, SSSUTMS University, Madhya Pradesh

Abstract – Big data analytics is where advanced analytic techniques operate on big data sets. Hence, big data analytics is really about two things big data and analytics plus how the two have teamed up to create one of the most profound trends in business intelligence (BI) today. Let's start by defining advanced analytics, then move on to big data and the combination of the two. The size, variety, and rapid change of such data require a new type of big data analytics, as well as different storage and analysis methods. Such sheer amounts of big data need to be properly analyzed, and pertaining information should be extracted. The study was selected based on its novelty and discussion of important topics related to big data Analytics, in order to serve the purpose of our research.

Keywords: Big Data, Analytics, Techniques, Data Sets, Business Intelligence, Information, Important, etc.

-----X-----

INTRODUCTION

The Big – Data, Analytics, and Decisions (B-DAD) framework which incorporates the big data analytics tools and methods into the decision making process. The framework maps the different big data storage, management, and processing tools, analytics tools and methods, and visualization and evaluation tools to the different phases of the decision making process. Hence, the changes associated with big data analytics are reflected in three main areas: big data storage and architecture, data and analytics processing, and, finally, the big data analyses which can be applied for knowledge discovery and informed decision making. Each area will be further discussed in this study. However, since big data is still evolving as an important field of research, and new findings and tools are constantly developing, this study is not exhaustive of all the possibilities, and focuses on providing a general idea, rather than a list of all potential opportunities and technologies. Many organizations have realized the real-value benefits of Big Data and today they do have access to Big Data but they are facing significant challenges in processing and analyzing the wealth amount of data timely and effectively. More importantly, how to extract important information and knowledge from Big Data due to the sheer volume of the data in different forms (i.e. structured, semi-structured and unstructured) is an extremely challenging task. For many decades, the organizations have successfully applied relational database management systems (DBMS) for data storage and analysis. However, managing Big Data with its associative characteristics such as volume, velocity and variety is a challenging task for traditional

DBMS because DBMS are hard to scale with ever increasing data and only support structured data format. However, opportunity is available with the right technology platform, to store and analyze the Big Data timely and effectively.

REVIEW OF LITERATURE:

Bakshi, K. (2012) define, The data is uploaded to the storage from operational data stores using Extract, Transform, Load (ETL), or Extract, Load, Transform (ELT), tools which extract the data from outside sources, transform the data to fit operational needs, and finally load the data into the database or data warehouse. Thus, the data is cleaned, transformed, and catalogued before being made available for data mining and online analytical functions.

According to Manyika et al.'s (2011) research, big data can enable companies to create new products and services, enhance existing ones, as well as invent entirely new business models. Such benefits can be gained by applying big data analytics in different areas, such as customer intelligence, supply chain intelligence, performance, quality and risk management and fraud detection. Furthermore, Cebr's study highlighted the main industries that can benefit from big data analytics, such as the manufacturing, retail, central government, healthcare, telecom, and banking industries.

Cuzzocrea, A., Song et al. (2011) discusses, Map Reduce is a parallel programming model, inspired by the "Map" and "Reduce" of functional languages, which is suitable for big data processing. It is the core

of Hadoop, and performs the data processing and analytics functions. According to EMC, the MapReduce paradigm is based on adding more computers or resources, rather than increasing the power or storage capacity of a single computer; in other words, scaling out rather than scaling up. The fundamental idea of MapReduce is breaking a task down into stages and executing the stages in parallel in order to reduce the time needed to complete the task.

Russom, P. (2011) elaborate, Big data sizes are constantly increasing, currently ranging from a few dozen terabytes (TB) to many petabytes (PB) of data in a single data set. Consequently, some of the difficulties related to big data include capture, storage, search, sharing, analytics, and visualizing. Today, enterprises are exploring large volumes of highly detailed data so as to discover facts they didn't know before. Hence, big data analytics is where advanced analytic techniques are applied on big data sets. Analytics based on large data samples reveals and leverages business change. However, the larger the set of data, the more difficult it becomes to manage.

H. Herodotou et al. (2009) provides a technique to implement self-tuning in Big Data Analytic systems. Hadoop's performance out of the box leaves much to be desired, leading to suboptimal use of resource, time and money. This study introduces Starfish, a self-tuning system for big data analytics.

CONCLUSION:

The goal of Big Data analytics for security is to obtain actionable intelligence in real time. Although Big Data analytics have significant promise, there are a number of challenges that must be overcome to realize its true potential. The following are only some of the questions that need to be addressed:

1. Data provenance: authenticity and integrity of data used for analytics. As Big Data expands the sources of data it can use, the trustworthiness of each data source needs to be verified and the inclusion of ideas such as adversarial machine learning must be explored in order to identify maliciously inserted data.
2. Privacy: we need regulatory incentives and technical mechanisms to minimize the amount of inferences that Big Data users can make. CSA has a group dedicated to privacy in Big Data and has liaisons with NIST's Big Data working group on security and privacy. We plan to produce new guidelines and white papers exploring the technical means and the best principles for minimizing privacy invasions arising from Big Data analytics.
3. Securing Big Data stores: this document focused on using Big Data for security, but the other side of the coin is the security of Big

Data. CSA has produced documents on security in Cloud Computing and also has working groups focusing on identifying the best practices for securing Big Data.

REFERENCES:

- Bakshi, K. (2012). Considerations for Big Data: Architecture and Approaches. In: Proceedings of the IEEE Aerospace Conference, pp. 1–7.
- Cuzzocrea, A., Song, I., Davis, K. C. (2011). Analytics over Large-Scale Multidimensional Data: The Big Data Revolution! In: Proceedings of the ACM International Workshop on Data Warehousing and OLAP, pp. 101–104.
- H. Herodotou, H. Lim, G. Luo, N. Borisov, L. Dong, F. B. Cetin and S. Babu. Starfish (2009). A Selftuning System for Big Data Analytics. In CIDR, pages pp. 261–272.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H. (2011). Big Data: The Next Frontier for Innovation, Competition, and Productivity. In: McKinsey Global Institute Reports, pp. 1–156.
- Russom, P. (2011). Big Data Analytics. In: TDWI Best Practices Report, pp. 1–40.

Corresponding Author

Nitin Kumar Saran*

Research Scholar, SSSUTMS University, Madhya Pradesh

E-Mail – chintuman2004@gmail.com