Journal of Advances and Scholarly Researches in Allied Education

Vol. XI, Issue No. XXI, Apr-2016, ISSN 2230-7540

BIG DATA AND BIG DATA ANALYTICS: BASIC CONCEPT AND PERSPECTIVES

AN INTERNATIONALLY INDEXED PEER REVIEWED & REFEREED JOURNAL

www.ignited.in

Journal of Advances and Scholarly Researches in Allied Education
Vol. XI, Issue No. XXI, April-2016, ISSN 2230-7540

Big Data and Big Data Analytics: Basic Concept and Perspectives

Vakala Ramakrishna Sumant[1] Dr. Bechoo Lal[2]

[1]Pebble Hills University, USA

[2]Assistant Professor, Western College University of Mumbai

Abstract – This study focuses on the various states of studies towards Big Data analytic techniques and gives a better comparative analysis of various applications. Inference has been done for evaluating the performance efficiency, limitations and the advantages of the different types of existing Big Data Analytic techniques. In the recent times the amount of data are generated and stored by various industries are rapidly increasing on the internet thus data scientists are facing a lot of challenges for maintaining a huge amount of data as the fast growing industries require the significant information for enhancing the business and for predictive analysis of the information. A larger amount of data gives a better output but also working with it can become a challenge due to processing limitations. This paper intends to define the concept of Big Data and stress the importance of Big Data Analytics.

Keywords: Big Data, Big Data Analytics, Database, Internet, Hadoop Project, limitations, Perspectives etc.

----------X----------

INTRODUCTION

We can associate the importance of Big Data and Big Data Analysis with the society that we live in. Today we are living in an Informational Society and we are moving towards a Knowledge Based Society. In order to extract better knowledge we need a bigger amount of data. The Society of Information is a society where information plays a major role in the economic, cultural and political stage (Elgendy, 2013). In the Knowledge society the competitive advantage is gained through understanding the information and predicting the evolution of facts based on data. The same happens with Big Data. Every organization needs to collect a large set of data in order to support its decision and extract correlations through data analysis as a basis for decisions. In this study we will define the concept of Big Data, its importance and different perspectives on its use (Herodotou, *et. al.,* 2009). In addition we will stress the importance of Big Data Analysis and show how the analysis of Big Data will improve decisions in the future.

REVIEW OF LITERATURE:

Big data is one of the "hottest" phrases being used today. Everyone is talking about big data, and it is believed that science, business, industry, government, society, etc. will undergo a thorough change with the influence of big data. Technically speaking, the process of handling big data encompasses collection, storage, transportation and exploitation (Dean, Ghemawat, 2010). It is no doubt that the collection, storage and transportation stages are necessary precursors for the ultimate goal of exploitation through data analytics, which is the core of big data processing. Turning to a data analytics perspective, we note that "Big Data" has come to be defined by the four V's — Volume, Velocity, Veracity, and Variety. It is assumed that either all or any one of them needs to be met for the classification of a problem as a Big Data problem. Volume indicates the size of the data, which might be too big to be handled by the current state of algorithms and/or systems. Velocity implies data are streaming at rates faster than that can be handled by traditional algorithms and systems. Sensors are rapidly reading and communicating streams of data. We are approaching the world of quantified self, which is presenting data that was not available hitherto. Veracity suggests that despite the data being available, the quality of data is still a major concern (Kubick, 2012). That is, we cannot assume that with big data comes higher quality.

1- Big Data Concept: The term "Big Data" was first introduced to the computing world by Roger Magoulas from O'Reilly media in 2005 in order to define a great amount of data that traditional data management techniques cannot manage and process due to the complexity and size of this data. Nowadays the Big Data concept is treated from different points of view covering its implications in many fields. According to MiKE 2.0, the open source

standard for Information Management, Big Data is defined by its size, comprising a large, complex and independent collection of data sets, each with the potential to interact. In addition, an important aspect of Big Data is the fact that it cannot be handled with standard data management techniques due to the inconsistency and unpredictability of the possible combinations (Dean, Ghemawat, 2010).

2- Big Data Analytics: The world today is built on the foundations of data. Lives today are impacted by the ability of the companies to dispose, interrogate and manage data. The development of technology infrastructure is adapted to help generate data, so that all the offered services can be improved as they are used. As an example, internet today became a huge information-gathering platform due to social media and online services (Ibrahim, *el. al.,* 2008). At any minute they are added data. The explosion of data cannot be any more measured in gigabytes; since data is bigger there are used etabytes, exabytes, zettabytes and yottabytes.

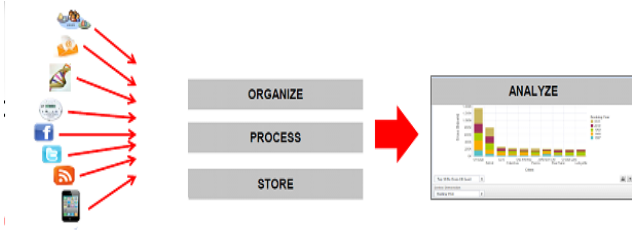

Fig 1- Big Data Management

The big data analytics initiative should be a joint project involving both IT and business. IT should be responsible for deploying the right big data analysis tools and implementing sound data management practices. Both groups should understand that success will be measured by the value added by business improvements that are brought about by the initiative.

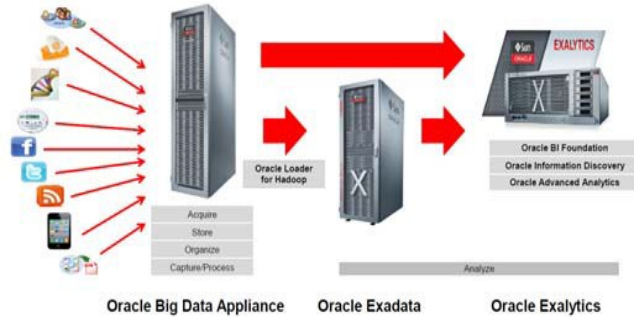

Fig 2- Oracle Big Data Solution

3- Big Data Analytics Software: Currently, the trend is for enterprises to re-evaluate their approach on data storage, management and analytics, as the volume and complexity of data is growing so rapidly and unstructured data accounting is for 90% of the data today. Every day, 2.5 quintillion bytes of data are created so much that 90% of the data in the world today has been created in the last two years alone (Russom, 2011). This data comes from various sources such as: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals, web and software logs, cameras, information-sensing mobile devices, aerial sensory technologies and genomics. This data is referred to as big data.

CONCLUSION:

The year 2012 is the year when companies are starting to orient themselves towards the use of Big Data. That is why this paper presents the Big Data concept and the technologies associated in order to understand better the multiple benefices of this new concept ant technology. In the future we propose for our research to further investigate the practical advantages that can be gain through Hadoop. The aim of this study is to evoke discussion rather than to provide a comprehensive survey of big data research. That is why this study presents the Big Data concept and the technologies associated in order to understand better the multiple benefices of this new concept and technology.

REFERENCES:

Elgendy, N. (2013): Big Data Analytics in Support of the Decision Making Process. MSc Thesis, German University in Cairo, pp. 164

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, pp. 261–272.

Jeffrey Dean and Sanjay Ghemawat (2010). "MapReduce: Simplified Data Processing on Large Clusters" OSDI.

Kubick, W.R. (2012): Big Data, Information and Meaning. In: Clinical Trial Insights, pp. 26–28.

Shadi Ibrahim, Hai Jin, Lu Lu (2008). "Handling Partitioning Skew in MapReduce using LEEN" ACM 51, pp. 107–113

Russom, P. (2011): Big Data Analytics. In: TDWI Best Practices Report, pp. 1–40.