INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 2 ISSUE 6 (2016) PAGES 1467 - 1471
RECEIVED : 24/11/2016. PUBLISHED: 08/12/2016
December 8, 2016

PERSPECTIVES ON BIG DATA AND BIG DATA ANALYTICS

[1]Gajalakshmi.D, [2]Daniel Sundarraj.P,
[1]M.Phil Scholar, Dept of computer science and Applications, KMG College of Arts & Science, Gudiyatham,
[2]HOD, PG and Research Dept of computer science and Applications, KMG College of Arts & Science, Gudiyatham.

Abstract:
Nowadays companies are starting to realize the importance of using more data in order to support decision for their strategies. It was said and proved through study cases that "More data usually beats better algorithms". With this statement companies started to realize that they can chose to invest more in processing larger sets of data rather than investing in expensive algorithms. The large quantity of data is better used as a whole because of the possible correlations on a larger amount, correlations that can never be found if the data is analyzed on separate sets or on a smaller set. A larger amount of data gives a better output but also working with it can become a challenge due to processing limitations. This article 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.

1. INTRODUCTION

Nowadays the Internet represents a big space where great amounts of information are added every day. The IBM Big Data Flood Infographic shows that 2.7 Zettabytes of data exist in the digital universe today. Also according to this study there are 100 Terabytes updated daily through Facebook, and a lot of activity on social networks this leading to an estimate of 35 Zettabytes of data generated annually by 2020. Just to have an idea of the amount of data being generated, one zettabyte (ZB) eguals 1021 bytes, meaning 1012 GB. [1] 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 economical, cultural and political stage. 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 article we will define the concept of Big Data, its importance and different perspectives on its use. 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.

2. 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. A study on the Evolution of Big Data as a Research and Scientific Topic shows that the term "Big Data" was present in research starting with 1970s but has been comprised in publications in 2008. [2] 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. [3] In IBM's view Big Data has four aspects:

Volume: refers to the quantity of data gathered by a company. This data must be used further to obtain important knowledge; Velocity: refers to the time in which Big Data can be processed. Some activities are very important and need immediate responses, that is why fast processing maximizes efficiency; Variety: Refers to the type of data that Big Data can comprise. This data can be structured as well as unstructured; Veracity: refers to the degree in which a leader trusts the used information in order to take decision. So getting the right correlations in Big Data is very important for the business future. [4] In addition, in Gartner's IT Glosarry Big Data is defined as high volume, velocity and variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making. There are a lot of definitions on Big Data circulating around the world, but we consider that the most important one is the one that each leader gives to its one company's data. The way that Big Data is defined has implication in the strategy of a business. Each leader has to define the concept in order to bring competitive advantage for the company.

3. BIG DATA ANALYSIS

The understanding of Big Data is mainly very important. In order to determine the best strategy for a company it is essential that the data that you are counting on must be properly analyzed. Also the time span of this analysis is important because some of them need to be performed very frequent in order to determine fast any change in the business environment. Another aspect is represented by the new technologies that are developed every day.

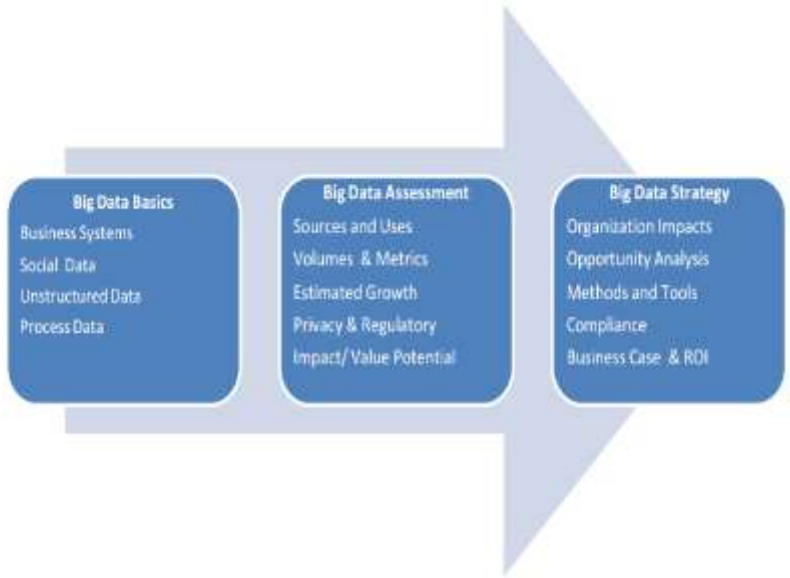

Fig.1. Developing a Big Data Strategy

Considering the fact that Big Data is new to the organizations nowadays, it is necessary for these organizations to learn how to use the new developed technologies as soon as they are on the market. This is an important aspect that is going to bring competitive advantage to a business. The need for IT specialists it is also a challenge for Big Data. According to McKinsey's study on Big Data called Big Data: The next frontier for innovation, there is a need for up to 190,000 more workers with analytical

expertise and 1.5 million more data- literate managers only in the United States. This statistics are a proof that in order for a company to take the Big Data initiative has to either hire experts or train existing employees on the new field. 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.

Fig.2. Big Data Management

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. 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. In order to manage the giant volume of unstructured data stored, it has been emerged the "Big Data" phenomena. It stands to reason that in the commercial sector Big-Data has been adopted more rapidly in data driven industries, such as financial services and telecommunications, which it can be argued, have been experiencing a more rapid growth in data volumes compared to other market sectors, in addition to tighter regulatory requirements and falling profitability. At first, Big Data was seen as a mean to manage to reduce the costs of data management. Now, the companies focus on the value creation potential. In order to benefit from additional insight gained there is the need to assess the analytical and execution capabilities of "Big Data". Big Data Management is based on capturing and organizing relevant data. Data analytics supposes to understand that happened, why and predict what will happen. A deeper analytics means new analytical methods for deeper insights.

4. ANALYSIS

Apache Hadoop is a fast-growing big-data processing platform defined as "an open source software project that enables the distributed processing of large data sets across clusters of commodity servers"[15]. It is designed to scale up from a single server to thousands of machines, with a very high degree of fault tolerance. Rather than relying on high-end hardware, the resiliency of these clusters comes from the software's ability to detect and handle failures at the application layer. Developed by Doug Cutting, Cloudera's Chief Architect and the Chairman of the Apache Software Foundation, Apache Hadoop was born out of necessity as data from the web exploded, and grew far beyond the ability of traditional systems to handle it. Hadoop was initially inspired by papers published by Google outlining its approach to handling an avalanche of data, and has since become the de facto standard for storing, processing and analyzing hundreds of terabytes, and even petabytes of data. Cost will certainly be a software selection factor as that's a big reason companies are adopting.

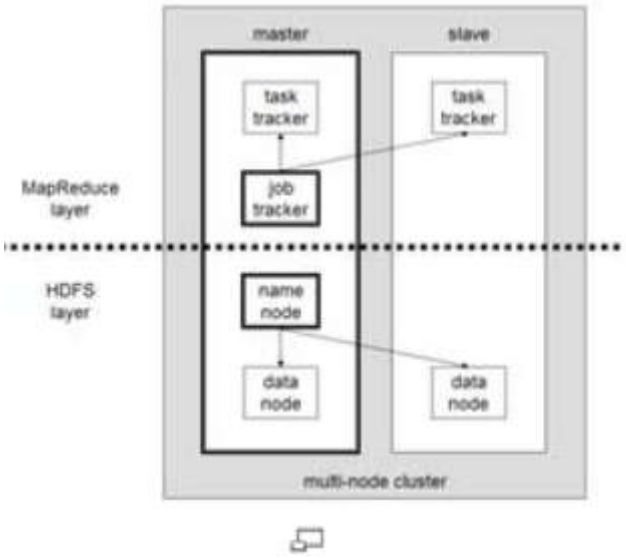

Fig.3. A multi-node Hadoop cluster

Hadoop; they're trying to retain and make use of all their data, and they're expecting cost savings over conventional relational databases when scaling out over hundreds of Terabytes or more. Sears, for example, has more than 2 petabytes of data on hand, and until it implemented Hadoop two years ago, Shelley says the company was constantly outgrowing databases and still couldn't store everything on one platform. Once the application can run on Hadoop it will presumably be able to handle projects with even bigger and more varied data sets, and users will be able to quickly analyze new data sets without the delays associated with transforming data to meet a rigid, predefined data model as required in relational environments.

CONCLUSION

The year 2012 is the year when companies are starting to orient themselves towards the use of Big Data. That is why this articol presents the Big Data concept and the technologies associated in order to understand better the multiple beneficies 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.

REFERENCES

[1] G. Noseworthy, Infographic: Managing the Big Flood of Big Data in Digital Marketing, 2012 http://analyzingmedia.com/2012/infograp
hic-big-flood-of-big-data-in-digital- marketing/
[2] H. Moed, The Evolution of Big Data as a Research and Scientific Topic: Overview of the Literature, 2012, ResearchTrends, http://www.researchtrends.com
[3] MIKE 2.0, Big Data Definition, http://mike2.openmethodology.org/wiki/Big _Data_Definition
[4] P. Zikipoulos, T. Deutsch, D. Deroos, Harness the Power of Big Data, 2012, http://www.ibmbigdatahub.com/blog/harnes s-power-big-data-book-excerpt
[5] Gartner, Big Data Definition, http://www.gartner.com/it-glossary/big-data/
[6]E. Dumhill, “What is big data?”, 2012 ,http://strata.oreilly.com/2012/01/what-is- big-data.html

[7] A Navint Partners White Paper, "Why is BIG Data Important?" May 2012, http://www.navint.com/images/Big.Data.pdf [8] Greenplum. A unified engine for RDBMS and Map Reduce, 2009. http://www.greenplum.com/resources/mapre duce/.

[9] For Big Data Analytics There's No Such Thing as Too Big The Compelling Economics and Technology of Big Data Computing, White Paper, March 2012, By: 4syth.com, Emerging big data thought leaders

[10] Big data: The next frontier for innovation, competition, and productivity. James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. McKinsey Global Institute. May 2011.