

Oct 28th, 9:00 AM - Oct 30th, 5:00 PM

The Importance of Big Data Analytics

Eljona Proko

Vlora University, elzavalani@gmail.com

Follow this and additional works at: <https://knowledgecenter.ubt-uni.net/conference>



Part of the [Databases and Information Systems Commons](#), and the [Information Security Commons](#)

Recommended Citation

Proko, Eljona, "The Importance of Big Data Analytics" (2016). *UBT International Conference*. 2.
<https://knowledgecenter.ubt-uni.net/conference/2016/all-events/2>

This Event is brought to you for free and open access by the Publication and Journals at UBT Knowledge Center. It has been accepted for inclusion in UBT International Conference by an authorized administrator of UBT Knowledge Center. For more information, please contact knowledge.center@ubt-uni.net.

The Importance of Big Data Analytics

Eljona Proko

Computer Science Department, Vlora University
Vlore, Albania
elzavalani@gmail.com

Abstract. Identified as the tendency of IT, Big Data gained global attention. Advances in data analytics are changing the way businesses compete, enabling them to make faster and better decisions based on real-time analysis. Big Data introduces a new set of challenges. Three characteristics define Big Data: volume, variety, and velocity. Big Data requires tools and methods that can be applied to analyze and extract patterns from large-scale data. Companies generate enormous volumes of poly-structured data from Web, social network posts, sensors, mobile devices, emails, and many other sources. Companies need a cost-effective, massively scalable solution for capturing, storing, and analyzing this data. They also need to be able to integrate their Big Data into their real-time analytics environment to maximize business value. Big Data was seen as a mean to manage and to reduce the costs of data management. In this paper we discuss definition, characteristics, challenges, and techniques for analyzing Big Data and the importance of Big Data. The main importance of Big Data consists in the potential to improve efficiency in the context of use a large volume of data, of different type. Big Data enables companies to create new products and services, and invent entirely new business models. A case study will be present to demonstrate the potential of Big Data Analytics: Oracle Application in Social Insurance. We conclude this paper that through better analysis of the large volumes of data that are becoming available, there is the potential for making faster advances in many scientific disciplines and improving the profitability and success of many enterprises.

Keywords: Big Data, Data Analytics, Challenges, Techniques, Oracle

1. Introduction

Big data is a new power that bring new data types and storage mechanisms. Big Data solutions are ideal for analyzing raw structured data, also semi structured and unstructured data from a wide variety of sources. Big Data platform gives the opportunity to extract insight from an immense volume, variety, and velocity of data, in context, beyond what was previously possible. Three characteristics define Big Data: volume, variety, and velocity. Big Data generally refers to data that exceeds the typical storage, processing, and computing capacity of conventional databases and data analysis techniques. As a resource, Big Data requires tools and methods that can be applied to analyze and extract patterns from large-scale data. The rise of Big Data has been caused by increased data storage capabilities, increased computational processing power, and availability of increased volumes of data, which give organization more data than they have computing resources and technologies to process. This paper begins by describing Big Data definition, characteristics and challenges. In section 3 we discusses

techniques for analyzing Big Data. Section 4 discusses the importance of Big Data. In section 5 we present a case study: Oracle Application in Social Insurance. Finally, the paper presents some conclusions.

2. Big Data Definition, Characteristics and Challenges

Big Data is currently defined using three data characteristics: volume, variety and velocity. It means that some point in time, when the volume, variety and velocity of the data are increased, the current techniques and technologies may not be able to handle storage and processing of the data. At that point the data is defined as Big Data. In the Big Data research, the term Big Data Analytics is defined as the process of analyzing and understanding the characteristics of massive size datasets by extracting useful geometric and statistical patterns. Big Data describes innovative techniques and technologies to capture, store, distribute, manage and analyze megabyte or larger sized datasets with high-velocity and diverse structures that conventional data management methods are incapable of handling.

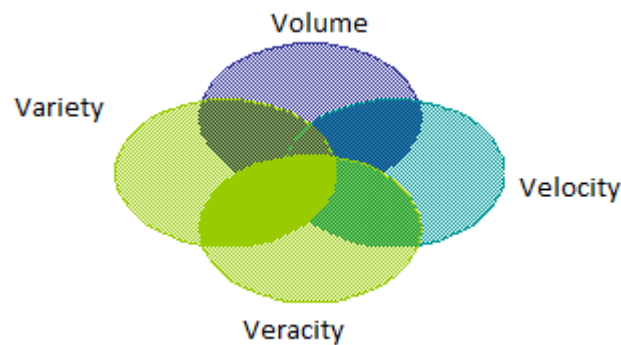


Fig. 1. Four characteristics of Big Data: volume, variety, velocity and veracity.

Characteristics

The first characteristic of Big Data, which is “Volume”, refers to the quantity of data that is being manipulated and analyzed in order to obtain the desired results. “Velocity” is all about the speed that data travels from point A, which can be an end user interface or a server, to point B, which can have the same characteristics as point A is described. “Variety” is the third characteristic of Big Data. It represents the type of data that is stored, analyzed and used. The type of data stored and analyzed varies and it can consist of location coordinates, video files, data sent from browsers, simulations etc. “Veracity” is the fourth characteristic of Big Data and came from the idea that the possible consistency of data is good enough for Big Data as is shown in figure 1. Current technologies software technologies try to overcome the challenges that “V’s” raises.

Challenges

This section discusses the most salient of the challenges.

Privacy is the most sensitive issue, with conceptual, legal, and technological implications. The costs of a data privacy breach can be enormous. Privacy is an overarching concern that has a wide range of implications for anyone wishing to explore the use of Big Data for development.

Access and Sharing. Sharing the information proves to be one of the most valuable characteristics of development. Every person and company has at their disposal large amount of information that can use it to serve their purposes. Everything is available only if everyone shares it. Regarding persons, there is a difference between what is personal and what can be made public.

Working with new data sources brings about a number of analytical challenges. The relevance and severity of those challenges will vary depending on the type of analysis being conducted, and on the type of decisions that the data might eventually inform.

A new class of business software applications has emerged whereby company data is managed and stored in data centers around the globe. While these solutions range from ERP, CRM, Document Management, Data Warehouses and Business Intelligence to many others, the common issue remains the safe keeping and management of confidential company data. By storing data on cloud systems is more convenient for companies in terms of cost.

Also, a cloud system is not only characterized by storage space, but as well for the speed of processing requested operations. The data security still remains a contentious issue in this case. These solutions often offer companies tremendous flexibility and cost savings opportunities compared to more traditional on premise solutions but it raises a new dimension related to data security and the overall management of an enterprise's Big Data paradigm.

Since big data will lose its value to current decision-making over time, and since it is voluminous and varied in content and structure, it is necessary to utilize new tools, technologies, and methods to archive and delete big data, without sacrificing the effectiveness of using your big data for current business needs.

3. Techniques for Analyzing Big Data

Big data analysis involves making “sense” out of large volumes of varied data that in its raw form lacks a data model to define what each element means in the context of the others.

It is a daunting task in most industries and companies that deal with big data just to understand the data that is available to be used, determining the best use of that data based on the companies' industry, strategy, and tactics. Also, these types of analyses need to be performed on an ongoing basis as the data landscape changes at an ever-increasing rate, and as executives develop more and more of an appetite for analytics based on all available information.

A common implementation that handles Big Data is MapReduce, presented in figure 2. MapReduce consists of two things: mapping and reducing. This is a technique that programmers use when they are confronted with large amount of data. By mapping a certain dataset is restructured into a different set of values. Reducing is a process that takes several “mapped” outputs and forms a smaller set of tuples.

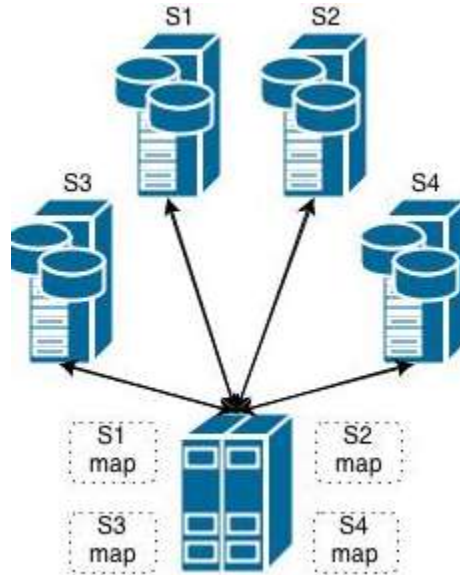


Fig. 2. MapReduce

4. Importance of Big Data

When big data is effectively and efficiently captured, processed, and analyzed, companies are able to gain a more complete understanding of their business, customers, products, competitors which can lead to efficiency improvements, increased sales, lower costs, better customer service, and/or improved products and services. The main importance of Big Data consists in the potential to improve efficiency in the context of use a large volume of data, of different type. Big data is widely seen as essential for the success of the design optimization of complex systems. Big Data can be used effectively in the following areas:

In information technology in order to improve security and troubleshooting by analyzing the patterns in the existing logs.

In customer service by using information from call centers in order to get the customer pattern and thus enhance customer satisfaction by customizing services.

In improving services and products through the use of social media content. By knowing the potential customers preferences the company can modify its product in order to address a larger area of people.

In the detection of fraud in the online transactions for any industry.

In risk assessment by analyzing information from the transactions on the financial market.

5. Case Study: Oracle Application in Social Insurance.

Oracle comes with a complete system solution for companies.

It starts from the basic ideas of Big Data sources, which can be traditional data generated by ERP systems, sensor data and social data, defined by the feedback that the company receives from customers.

Oracle consists of Hardware and Software from Oracle Corporation designed to integrate structured data and unstructured. This integration includes the registration of data from different sources that generate massive volumes of data. Also incorporated are different technologies to collect and analyze these data. A typical example is the use of this application in Social Security. The goal is to collect as much data regarding the files of pensioners. Registration of a folder starts from the registration of its request in ALSSH. It serves as the initial source of information from the national register of citizens who called upon the data of a person and his generalities, using its identification number. Another source of information are the years of work of a person who is taken from the archive of job years. Other sources of information are all the documents submitted by the person who made the referral, where the documents are scanned. Once registered applications and all documentation necessary for a person all along adopted through a user confirmation and had to pass benefits specialist who accepts them and made relevant calculations of the amount of pension in relation to claims made. Once the calculations are made conform the beneficiary satisfies the conditions that made mass confirmation of the pension calculation. Meanwhile after confirmation of the amount of pension becomes generation (transfer payment). The latter enables online over payment of social insurance in payment centers. Data collection consists of data which are provided by the National Register of Citizens, which is connected online to the civil state, and the data which are mainly provided by the beneficiaries of retirement. Figure 3 shows the view from user accounts that make registration of records in this application.

Fig. 3. Online user registration

The screenshot displays a web-based registration form with the following sections:

- Personal Data:** Fields for NID (highlighted in yellow), Email, Username, and Password.
- Work History:** Fields for 'Vencido Cables Pasados', 'No. VCP', 'Días VCP', 'No. Meses VCP', and 'Días Meses VCP'.
- Registration Details:** Fields for 'No. cartón', 'Fecha cartón', 'Agencia', 'Código', 'No. cartón', 'Formulario Preinscripción', 'Código Personal', 'No. de paso', 'Dí. Límite de ve. despes', 'Dí. Filio personal', 'Código pasaporte', and 'No. Seguro básico'.
- Insurance Status:** Fields for 'Seguro básico / Status Cartón', 'Regístrate / pa. preinscripción', 'Cual. Filio Básico', and 'Persono Actual'.
- Navigation:** Buttons for 'Documentos a Descargar', 'Móvil para más datos y personal', 'Tutor / Ay. Autorización Mapas', and 'Salir Cartón'.

Conclusions

Big Data concept and the technologies associated in order to understand better the multiple benefits of this new concept and technology. Big data has to do with volume as well as the speed of data storage. Since we are dealing with a transaction that efficiency leads to improvements in our services offering significant advantages. Big data allows organizations to create highly specific segmentations and to tailor products and services precisely to meet those needs. Big data enables companies to create new products and services, enhance existing ones, and invent entirely new business models.

References:

1. P. C. Zikopoulos, C. Eaton, D. deRoos, T. Deutsch, and G. Lapis, Understanding big data – Analytics for enterprise class Hadoop and streaming data, McGraw-Hill, 2012.
2. P. Zikopoulos, T. Deutsch, D. Deroos, Harness the Power of Big Data, 2012,
3. Oracle Information Architecture: An Architect’s Guide to Big Data, An Oracle White Paper in Enterprise Architecture August 2012
4. McKinsey Global Institute - Big data: The next frontier for innovation, competition, and productivity – June 2011
5. T. Chai, Y. Jin, and S. Bernhard, “Evolutionary complex engineering optimization: Opportunities and challenges,” IEEE Computational Intelligence Magazine, vol. 8, no. 3, pp. 12–15, 2013.
6. Y. Noguchi. The Search for Analysts to Make Sense of Big Data. National Public Radio (<http://www.npr.org/2011/11/30/142893065/the-search-foranalysts-to-make-sense-of-big-data>), Nov 2011.
7. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers. Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute, May 2011.