



BIG DATA IN BUSINESS

Liviu Constantin STOICA

PhD, Academy of Economic Studies, Bucharest, email: stoica.liviu.constantin@gmail.com

Abstract *The present article titled "Big Data in Business" was done in two parts, that is, in the first part we conducted a study on Big Data, and in the second we studied Big Data in Business. In the first part we studied: we defined and realized the features of Bog Data, we described the benefits of using Big Data solutions, we made the Big Data architecture, we listed some advantages and disadvantages of Big Data. In the second part we made an overview and studied the implementation of Big Data solutions in the real world.*

Key words:

Business, databases, Big Data, solution, techniques

JEL Codes:

M1

1. INTRODUCTION TO BIG DATA

As long as the internet is constantly expanding and is a necessity nowadays to be connected with each other, with the latest news, latest information and much more, it attracts a huge number of users.

The term "database" first appeared in the 1960s, and is defined as an organized collection of data.

The emergence of computers brought a first revolution of databases, using magnetic tapes through which the data sequence could be followed, followed by magnetic discs with a much higher data access speed. [1]

At first, it was very hard to create a database because each application had to define its own handling code. This has led to the conclusion that there must be a separate system for handling application data. Therefore, in 1970, the data

management system appeared, and with the appearance of the SGBD1, two models, the hierarchical model and the network model appeared.

The second database revolution was the creation of the relational database model by Codd, which was to be the universal model used extensively.

The third database revolution is begun by the largest site, the well-known Google, which in 2005, even if the relational model was well established, did not deal with the large amount of data that Google owned. These conditions led to the realization of the Hadoop project based on concepts such as MapReduce and Big Table. After Google, Facebook followed using a partitioning system by dividing data into multiple databases, and the Dynamo was created by Amazon.

There are noSQL terms, which are unrelated databases and help store large data bases. NewSql, which is a more modern data management system. And finally, the term Big Data emerged in 2012 and referred to the means by which data is highlighted, and by this term we refer to the technologies and solutions it offers to handle large data sets.

a) Benefits of using Big Data solutions

When the traditional data, processing, and storage analytics did no longer handle large amounts of information, Big Data solutions came up with the ability to combine unrelated datasets and process large amounts of data unstructured in a short time.

Generally, data for Big Data environments is accumulated within an enterprise through apps, sensors, or different external sources. This data is either stored in a data warehouse to enrich existing data, or used directly through an enterprise application. The results obtained through Big Data's processing lead to a wide range of benefits and perspectives such as:

- operational optimization;
- intelligent actionable;
- identifying new markets;
- making accurate predictions;
- detection of frauds and losses;
- very detailed records;
- improving the decision-making process;
- scientific discoveries.

Clearly, Big Data solutions offer many potential benefits, but besides these benefits there are also

many disadvantages when Big Data is adopted. These disadvantages must be compared to the benefits offered and therefore the best decision to be chosen.

b) Definition and characteristics of Big Data

Big Data is not a simple technology, it is a combination of old and new technologies where companies can easily manage large amounts of data at high speed and in real time. [2]

As a data set can be considered Big Data, you have to meet some features. Some of these features were initially identified by Laney Doug in 2001. He published an article on the impact of the volume, speed and variety of electronic store data gathered in a data warehouse.

A set of data to be considered Big Data has to fulfill 5 features, and these are: [3]

- Data volume must be anticipated because a large amount of data requires another data storage and processing system. Everyday mankind generates about 2.5 exabytes of data. For example, some data sources that are responsible for generating this large amount of data are: online transactions, scientific experiments, sensors, and social media.
- Speed. In Big Data environments, data can run at very high speeds, and data sets can accumulate over an extremely short period of time in an enormous amount. From the enterprise's point of view, the data speed translates into the time period from the moment the data enters the enterprise, until it is processed. For this, the enterprise must have

highly elastic processing solutions and adequate storage capabilities.

- Variety of data refers to the different formats and types of data Big Data has to support. Variety of data is a real challenge for businesses because integration, transformation, processing and storage solutions must be found.
- Verity of data is their quality. The data being processed must be evaluated for their quality in order to eliminate erroneous data in order to have a correct and accurate analysis. For example, worthless data should be removed.
- Value is the usefulness of business data. The value of the data is closely related to the veracity of the data, as the values are more qualitative they will be retained. The value and time are inversely proportional, as the processing time of the data is greater to be transformed into meaningful information, the less value for the enterprise.

c) Data types processed

Data that Big Data processes can process can be generated by both man and computer. Man-generated data is generally about the interaction of a user with the system used, and data generated by the computer is a record of the user's activity, for example: the date when it was authenticated. Human and computer generated data can be in different formats.

The main types of data are:

- Structured data is usually stored in the form of tables, they are retained in this form because

they contain relationships with other tables and are usually represented in relational databases. [4]

- Unstructured data is data that does not have a pattern or schema, it can be textual, binary or links to other files. A text file may contain news or posts, and binary files can include media files such as images, audio, and video. These unstructured data can not be interrogated using SQL, but if we want to attach an audio file to a relational table, it can be put into the table as BLOB2 or using NoSQL, which is used to store structured data with unstructured data. [5]
- Semi-structured data is not a relational type, but it is usually hierarchically structured and retained in text files, such as JSON or XML. Because they are retained as text, they are much easier to process than unstructured data. [6]
- Metadata provides information such as the characteristics and structure of a data set. These data are generally generated by the computer. Their use in Big Data solutions is very important during processing because it provides information about the origin of the data. [7]

d) Management and Big Data Architecture

Until a few years ago, for many companies, it was very expensive to store all the data they needed to process them. But since Big Data has emerged, storing and processing large amounts of data is no longer difficult. Also now data storage can

be done in the Cloud and with a very good internet speed it is very efficient and much cheaper.

If businesses can analyze large volumes of data, with such quantity they can make patterns and anomalies, businesses can take advantage of these solutions. But the use of Big Data solutions does not just stop at business support but is used in many areas, science, research, government activities, astronomy, and beyond.

The figure below shows that data must first be captured after which they are organized and integrated. If this phase has been successfully completed, it may be possible to analyze it after the proposed problem.

To achieve the management of a large data base, a high-performance architecture is needed. It requires a lot of power and computing speed, because besides the fact that the data needs to be analyzed, large amounts of data are generated and must be stored, and this can lead to system slowdown or downtime.

Big components of the Big Data architecture are the interface and data streams. It is to be understood that data and data flows come from different external sources and for this it is necessary to use the APIs underlying the architecture without integrated services there could not be Big Data.

A first level of architecture is physical infrastructure. Physical infrastructure is the fundamental component for the operation and scalability of a Big Data architecture. To retain an unpredictable and unpredictable volume, physical architecture must be different from traditional

architecture. This architecture is based on a distributed computing model and means that data can be stored in different locations and can be linked to each other via the Internet.

A second important element is information security, the more important is the data analysis for a company, the importance of data security. An example is keeping data about private users. So security should be implemented from the beginning and not a later idea.

Another important component is the operational data source. Traditionally, the operational data source consists of highly structured relational databases. But now operational data sources also include unstructured data.

Services and data organization tools are another component of architecture. Data in very large quantities comes from a lot of sources, this data was expensive in the past, or even if a company had the ability to store huge amounts of information, they did not have the tools to use that information. The instruments at that time were very difficult to use and costly in terms of time.

e) Tools used in Big Data

With the evolution of computing technology, it is now possible to manage huge amounts of data, which in the past was only possible with supercomputers and these were very expensive. The Big Data step was made when companies such as Yahoo!, Google and Facebook realized that they need to value their data volume. So they had to create new technologies that managed to store, access and analyze large amounts of real-time data

in order to generate money on the basis of the information gathered, giving users benefits. The first tools that managed to do this were MapReduce, Hadoop and Big Table. [8]

- MapReduce was created by Google as an effective way to perform a set of functions on a large amount of data. The "Map" component is used to distribute problems and tasks and has the role of managing the placement of tasks so as to balance the system and avoid overloading. After the distribution step has been completed, the "Reduce" function brings together all the elements to provide the result.
- Big Table was developed by Google and is a distributed storage system designed to manage scalable, structured data. The data is organized in rows and columns, unlike a traditional relational database, Big Table is a distributed and persistent multidimensional map used to hold large volumes of data.
- Hadoop is an Apache framework and is developed on the MapReduce and Big Table concepts. This project is the foundation behind Yahoo's computing architecture. Hadoop was designed to paralyze data processing on computational nodes to increase computing speed and hide latency.

f) Traditional and advanced analysis techniques

For these data to make sense to a company, it requires different analysis approaches, depending on the problem. Some analyzes use traditional data warehouses, while other analyzes use advanced predictive analysis. Managing large

data requires different approaches to help the company have a successful plan in the future.

- Data warehouse analysis. Once a company sorts through large amounts of data, it is often necessary to take a subset of data that reveals a model and can be put in a form available to the firm. These data warehouses provide compression, multi-level partitioning, and parallel data processing arrays.
- Analysis of Big Data data. The ability to manage and analyze data pentabytes allows companies to work with data that could have a significant impact on the company. These require analytical engines that can manage distributed data and deliver optimized results to solve a business problem. For example, some organizations use predictive models that unite structured and unstructured data to predict and avoid fraud
- Reports and views. Organizations rely on the ability to create reports to understand the data gathered from the creation of tables with monthly sales to growth forecasts. Big Data changes how data is managed and used. If a company can collect, manage, and process data, they can use a new generation of tools that help the organization not only store data, but also understand how this data provides a context based on the issue. With Big Data, reports and data visualization become tools for analysis in the context in which data has an impact on the future.
- Big Data applications. Generally, companies expect data to be used for analysis and

accurately predict the next period. The data has been integrated into the industry to meet a general business goal, but with the release of Big Data, that has changed. Now, applications are developed that are especially designed to take advantage of the unique features of large data.

Some of the emerging applications are for areas such as health care, traffic management, manufacturing management, and more. For example, in the field of healthcare, Big Data applications can be used to monitor preterm infants to determine when intervention is needed. In production, Big Data applications can be used to prevent machines from closing during production. In traffic, these applications could reduce the number of bottlenecks, reduce casualties, save fuel, and reduce pollution.

g) Advantages and disadvantages in using Big Data

- Analyze. It is known that the computer was adopted to replace manual efforts. For example, "Google Analytics" scans any progression or regression from the user experience. Over time, keeping activities on a particular site may be an empowerment tool to get new strategies to attract traffic.

Advantages: Big Data is an exclusive data warehouse that helps build millions of creative and innovative ideas. Data analysis can be the basis of any business.

Disadvantages: Sometimes analyzing this data can be misleading.

- Access to data using surveys. Poll-survey research seeks to access more and more information. These polls provide authentic databases from end-users.

Advantages: Big Data is an amazing tool for getting the behavior and customer trends. By applying a survey by a firm, it can discover what should be changed or improved within the firm.

Disadvantages: It is quite costly to designate analysts, data mining experts.

- Update speed

Advantages: Updating data is essential for obtaining accurate information.

Disadvantages: Sometimes updating is too fast and the app does not match real data.

- BigData is enormous. While a word is written in the Google search field, a large amount of data is received as a result. It is sufficiently anticipated how large the data is accessed.

Advantages: A platform can have unlimited information.

Disadvantages: It is difficult to gather relevant information.

2. BIG BUSINESS DATA

a) Overview

By expanding common data sets that now include Big Data, structured data and unstructured data such as social media, navigation logs, sensor data, and application of text analytics and other tools, companies are able to understand customer behavior and preferences .

In telecommunications, using Big Data to create predictive models, companies have been able to gain a huge advantage using proprietary data to help them predict and better manage their customers. In this industry, keeping customers was a serious problem because customers like to switch suppliers looking for the best phones and the best deals and the telecoms companies wanted to understand these customers to identify and stop them .

They had data that they were analyzing but never analyzed how customers are calling. For example, customers who generally appeal to customers who generally receive calls, businesses did not know how long they were talking about, and which was the time when they spoke more. This information is very important for businesses, they have managed to identify a model to identify people who are more likely to leave, and companies are trying to offer these people special offers to prevent them from leaving.

Financial services organizations use Big Data from customer interactions, and they can rank them in multiple segments. Following customer rankings, these financial services organizations can create more relevant and sophisticated offers.

Insurance companies also use data analysis to understand the customer and provide him with a more tailored insurance solution based on his current behavior rather than placing it in a broader category. By trying to understand the market and people looking for insurance, it has led to the collection of a new database and product innovation.

For example, young drivers can opt for low premium insurance by wearing a black box placed in the car or a mobile app through which they can monitor how they drive their individual capabilities rather than fit them into an age segment . Or insurance companies also use Big Data to know when they issue a home insurance if they need an agent's visit.

Internet-based businesses develop information products that combine customer data to provide more appealing recommendations and discounts. Advertisers are pursuing social media to understand the reaction to some campaigns, promotions, and other forms of advertising.

Even hotels use data analytics to understand customers and improve their offer. For example, a chain of hotels has replaced the traditional method of providing a questionnaire to customers when they released the room, with data coming from social media where they analyzed people's opinion about their hotel.

In sports, for example in football, clubs such as Chelsea and Manchester United, employ data processing companies and data researchers who record each player's move to look for other ways to win. Players being very expensive, any way to find young talents is very useful to avoid over-wages. With the help of data analysis, in a football match every sliding, every step, every move can be recorded, reaching a number of 2000 events per match. The cameras can record the movements, the distance traveled, the speed and the acceleration of each player, and this data helps the

big clubs to discover good players in the lower leagues.

Here's how the Big Data business environment is attributed:

- Data sets are used perfectly, that is, some revenue comes from combining data sets;
- lower costs, ie less complexity, with sufficient flexibility for future requirements;
- stability. Because data volumes are massive, users need to access and interact with data easily. [9]

b) Implementing Big Data in the real world

Companies can get benefits when they use the data correctly and applying analytics tools to turn the data into critical business information. And I will continue to have some examples that have happened in the real world.

- A bank has managed to halve its personnel costs in a certain area. In the past, this bank relied on the fact that the best performances came from individuals graduating from an elite university in the area. The analysis of the data clearly showed that they were wrong. It turned out that candidates from universities that were not prestigious overcame candidates in that top university, which allowed the company to recruit the most talented individuals for less money.
- A company of goods and services uses social media analysis combined with other analysis tools to find the right candidates they need. By analyzing the social media profile, they can accurately specify the level of intelligence and

the emotional stability of the potential candidate.

- Another company has used different data analyzes to find the type of individual they want to recruit, and the type of individual they want to avoid. This company has discovered that the type of search engine used by potential candidates is very important. Thus, candidates who did not use the pre-installed search engine and used search engines such as Firefox or Chrome would be much more appropriate.
- Another asset company is able to predict how some key elements of staff involvement influence operating performance, customer satisfaction, and financial performance. In addition, the company can predict the extent to which a certain increase in staff satisfaction leads to revenue growth.
- A call center found that some salesmen who had a criminal record had far more satisfactory performances than those who did not have a criminal record. This company also noticed that people with fewer Facebook connections have better performances than people with many Facebook connections.
- Another organization has used analytics tools to scan and analyze the content of emails sent by employees, as well as analyze their posts in social media such as Facebook or Twitter, which allowed the organization to know precisely the level of employee involvement.
- LinkedIn is a business-oriented social networking, founded in 2002, is primarily used to create professional networks. LinkedIn uses

Big Data to develop services such as: finds people you might know, jobs you might be interested in, who viewed your profile, and more. A UK Formula 1 team, McLaren Racing, uses racing sensors that record real-time data to identify problems using predictive analysis, which leads them to correct some problems before they happen and too late to solve.

- Verizon is a telecommunications company in the United States. It uses Big Data to boost ads on your mobile phone. It uses a unique identifier that when the user logs on to the site, this identifier switches to action and allows the advertiser to use the information on the computer, so the advertisements will be sent using this information.
- Mint.com is a free website designed for the United States and Canada that offers personal financial management services. This site uses Big Data to provide users with information on spending by category and provides a perspective of money spent per week, month or year.
- Tesla is a car manufacturing company. This company implements sensors on cars, collected data is sent and collected at headquarters. The data collected is used by the research and development department to improve performance, machine maintenance and customer satisfaction. One example is that when the machine does not work properly, the company is notified of this and it will also notify the customer of going to a service.

3. CONCLUSIONS

As a result, Big Data is a highly needed data analysis solution. Several years ago, we did not have the necessary technology or it was very expensive to keep all the data captured. Even if we were able to keep all the data, the traditional instruments of that time were not able to make an analysis. Big companies like Google that collect a huge amount of daily data have been forced to develop new tools to process Big Data.

Today, Big Data is used in almost all areas. In the field of health, patients can be monitored to prevent various diseases and to be treated on time. In sports, different sensors are used to capture a competitor's skills. Other companies use a user's Internet history to create the right ads. Even machines use different sensors to prevent problems and let the owner know when to visit a service. These are just some areas where Big Data solutions are used.

So the use of Big Data solutions nowadays is very important for the development of companies as well as for people who have different daily activities.

4. BIBLIOGRAPHY:

[1] YichuanWang, LeeAnnKung, Terry Anthony Byrd, Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations, Technological Forecasting and Social Change, Volume 126, January 2018, Pages 3-13

[2] Senthilkumar SA, Bharatendara K Rai, Amruta A Meshram, Angappa Gunasekaran,

Chandrakumarmangalam S, " Big Data in Healthcare Management: A Review of Literature", American Journal of Theoretical and Applied Business 2018; 4(2): 57-69, doi: 10.11648/j.ajtab.20180402.14 ISSN: 2469-7834 (Print); ISSN: 2469-7842 (Online)

[3] Peter Géczy, BIG DATA CHARACTERISTICS, The Macrotheme Review 3(6), Summer 2014, A multidisciplinary journal of global macro trends, National Institute of Advanced Industrial Science and Technology (AIST), Japan, pp 94-104

[4] Raphael Wimmer, " Files as directories: some thoughts on accessing structured data within files", Proceeding Programming'18 Companion Conference Companion of the 2nd International Conference on Art, Science, and Engineering of Programming, pp 166-170, , Nice, France — April 09 - 12, 2018

[5] Bitty Balducci, Detelina Marinova, " Unstructured data in marketing", Journal of the Academy of Marketing Science, July 2018, Volume 46, Issue 4, pp 557–590

[6] Sapthami R, Anisha P Rodrigues, "Study on Handling Semi-structured Data using NoSQL Database", International Journal of Engineering and Innovative Technology (IJEIT) Volume 5, Issue 10, April 2016, ISSN: 2277-3754, ISO 9001:2008 Certified

[7] Phokion G. Kolaitis, " Reflections on Schema Mappings, Data Exchange, and Metadata Management", Proceedings of the 37th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, pp 107-109, Houston, TX, USA — June 10 - 15, 2018

[8] YichuanWang, LeeAnnKung, Terry AnthonyByrd, "Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations", Technological Forecasting and Social Change, Elsevier, Volume 126, January 2018, Pages 3-13

[9] Hafedh Ibrahim Alfouzan, " Big Data In Business", International Journal of Scientific & Engineering Research, Volume 6, Issue 5, May-2015, ISSN 2229-5518, pp 1351-1352