

THE ROLE AND IMPORTANCE OF DATA ANALYST IN USING LARGE VOLUMES OF DATA

PhD. Stefan IOVAN^{1,2}

⁷⁾ West University, Computer Science Department, Timisoara, ROMANIA

⁸⁾ Railway Informatics SA, Strategy Department, Bucharest, ROMANIA

Eng. Alina Anabela IOVAN, Railway Informatics SA, Development Projects Department, Bucharest, ROMANIA

ABSTRACT: Most of the organizations have on their hands mountains of data that they could use in making decisions. The possibilities are infinite when answering business questions by company's using data accumulated in Data Warehouse and, in the same manner, in data base unstructured or text format generated by Social Media. The answers are hard to find. Big Data concept is perceived by many as to be referring to data received from structured and unstructured data coming from the classic and new type sources, like social media, but all related to human activity [1]. Lately the data generated by sensors (M2M phenomenon) began to generate the largest volumes of information that need to be analyzed and read. Today's top analysts are more wiser than giving simple reports, because they also use creativity skills and insight informations for strategic analysis. Data wizers are using their ability in the activity including personal knowledge given by the experience with analytic process and associated applications. They not only understand the given tasks but also the needs of the companies management department, but also have the knowledge where to find those informations suitable for taking strategic decisions. With all this, most of the cases of data wizer positions are filled by people not qualified for extracting data for analysis or doing consolidation data structures at company level. In this domain the applications are helping improve the professional level. This paper wants to address this aspect of the necessary transformation of the Big Data analyst wizer into a specialist in the domain of large volumes informations analysis.

KEY WORDS: Big Data, data analyst, analysis and data modulation, contextual information, proactive research

1. INTRODUCTION

The discussions around the notion "Big Data" most of the times start from false premises'. Most of the discussions are regarding on technological platforms of processing Big Data concentrated on volume, variety, speed and leaving aside any reference of value.

In order to obtain value from Big Data is required to add contextual information, which can be obtained just through placing analytic capabilities in the hands of the ones that need to extract the value. In other words, Big Data is in need of 'humanization'.

To humanize means to transform something inaccessible in something usable, making difficult – easy, from complex – simple, from abstract – concretely. Meaning the process of putting the information into context to 'tell

the story' about whom and what generates that information.

Big Data must be extracted from the 'bits' and 'bytes' world in which are available just for experts and made accessible, useful and capable to 'tell the story'; just in this manner they must be converted into real insight for real business people. Big Data must be 'down to earth', right there the people that know the business can use that information to make decisions, revealing in this way their real value.

Big Data humanization depends on two critical elements:

1. **Data easy access:** the ability to access, integrate, analyze Big Data should be available for data and business analysts involved in making strategic decisions in the organization.

2. **Access to suitable instruments able to give context, that helps Big Data to ‘tell the story’:** Big Data could offer a convertible business image only if is linked by the full context offered by all available data and if can be run sophisticated analysis without the need of advance data science and statistic knowledge.

Working current flux of Big Data uses to many resources and humans to generate value. With most BI tools (**B**usiness **I**ntelligence) existing create reports and analysis only based on previous structured data. These are, usually, only internal data, without inside data from the market, competitive intelligence and geographic data. Because of this, the results tell only one side to the story.

Working current fluxes of Big Data have several components. The data must be acquired from countless sources and clean up. Then, the data must be sorted and linked in such a way that allows interrogations’. After words must be stored in the system as file types that accepts unstructured formats. The analysts and software developers have after to work together in statistics environments like **R**, **SAS** and **SPSS** to interrogate the data [2]. In the end, the result data can be visualize in several forms – as statistic report or, sometimes, as 2D or 3D visualizations. As result, working Big Data fluxes – as available today – are looking more as jumps than as fluent fluxes.

The problem is created by the situation that all this work is not lead by a business user working in team with data analysts. Executed mostly by a team of IT specialists from ‘behind *the scene*’ and each step of the process requires the involvement of someone else which has, usually, substantial setbacks with other projects. Big Data working fluxes involves a finite number of changes and reruns, resulting in delays due to huge requests from the advance analytical abilities persona.

The person closes to the business user, the data or business analyst, can’t perform by himself a large part of the work and in this way time that runs between interrogations and

insight information’s records significant delays. In fact, most of the times, decisions are taken based on limited information’s, before the results from working Big Data flux came back. Because of it, fast and independent access of data analysts and capabilities given and powerful tools, are the core of the humanization of data.

Following will discuss about the emancipation of the analyst role existing in company by upgrading them to wizard rank, capable to extract important insight, backing up the decisional factors with information’s that help taking fast and fully informed decisions.

1.1. Data analyst role and importance

Big Data humanization makes that ‘large volumes of data’ to become accesible for analysts in the companies, offering the abilities availables, most of the time, only for IT (**I**nformation **T**echnology) department. Making possible presenting the data as informations, easy accessible and related to the subjects. Is about an analysis based on Big Data effortless and in a natural manner.

In order not to focus strictly on software development and statistic abilities, data obtained can be humanize by adding right context and by offering simple and direct tools for building analitic applications. Big Data humanization means direct work with data, in such a way that they cand tell their story. The complete image leads, in this way, to insight business. Meaning, also, a new oportunity for data analysts to perfect their art and to extend the capacity of independent way of analysis. They become, from data wizers specialists in data analysis.

In the past, data analysts required advance knowlege of statistics and business. In this moment, being able to access to analytic and contextual data tools, which usually were accesible only to data experts and IT, the analysts can become masters of data. A wizerd is a person that takes raw materials (data in our case) and uses their skill, knowledge and vision to create something of unique value.

Data wizerds have a superior understanding not only over the data they observe, but also also the business and its problems. Because the data wizers understand data and business so well that they can provide the right tools to use, obtaining in this way the most suitable results and ways to aply , several times, for business problems.

Work of data wizerds is the essence of Big Data humanization. Data wizerds are ‘humanization masters’ when talking about Big Data. They create data and analytic working fluxs that revel the true story, regardless if they deliver an answer to a specific questions or a new application for business users at the end of the process.

Big Data humanization involves developing some basic principles when asked to create solutions that deliver real insight:

- **Gadering and integrate data in any source:** gadering systems of tranzactional data, Social media and ‘sensibile’ data are all taken into consideration along with the informations from Data Warehouse.
- **Finding pathers:** pathers hold te key for futture estimations. Is not requested to find perfect pathers down to the smallest details, but in general lines, when we are refering to unstructured data.
- **Insight presenting to decision factors:** insight is more valuable when is available for everybody. A store or storage manager knows his market because deals with it on a daily bases. Haled by powerful analytic tools and preview sly centralize data, he will be able to make decisions efficiently and informed.
- **Reuse of analytic IP:** a data wizerd can create a set of data which afterwords to make available for a large area of decision people, who may, on their turn, to shape, adapt andbuild based on it. Everytime the story extends and also, gains focus.

Purpose for Big Data humanization is to offer the oportunity to analysts in diffrent departments of business, permitting them to createanalytic working fluxs that can be reused. An analytic working flux is an important part of intelectual property of a company and is increaseing value as it

becomes more popular, reused and completed[3].

Such instruments are available for BI analysts by **Alteryx**, a software platform extensible from desktop to web wich offers the fastest and complete insight over the clients, businessand market, for big companies but also for medium ones, guvermental agences and academic environment. Solving business problems for nearly a quarter of million users worldwide, **Alteryx** leads the globel revolution from BI.

2. BIG DATA IN SENSORS ERA

Is considered by the data analysts that the given data by the automatic sensors will generate the next wave of growing in production and technological innovation. Statistic reports show, for example, that being engines witch are working on the planes in use worldwide have reached the point of generating 10 TB of data and information’s for every half of hour of functioning. A line plane with four engines can generate 640 TB of data for one flight across Atlantic. If the number is multiplied by over 25.000 flights every day, the obtained volume of data is overwhelming.

Until now, almost all this data was lost at the end of the flight, but things are going to change. Why? Thanks to the tools and technology of Big Data, now ca identify, store, redraw and analyze the data in an efficient manner regarding costs and time use. This will provide huge opportunities, for example in the area on maintenance and defects prevision for planes – with direct effects in reducing the flights delays and cancelation reasons due to technical defects on planes.

The example given above is not unique. The practice case of the data obtained by sensors can be now efficiently analysis, in real time, in several activity areas. For example, intelligent sensors used in electric networks can generate large volumes of data which can be used to increase productivity.

Thanks to some matching tools, the providers of electricity can red parameters from the network in different consumption points every

quarter of hour instead of once a month. This process not only solves the need of sending a person on the field, and also allows differential pricing of electric consumption regarding the pick hours.

Differential pricing can be used for settlement of the curve of consumption during the pick of loads, afterword eliminating the need of creating an extra electric load to cover the needs, in this way generating reduce in costs for the energy providers in terms of generating capacity and maintenance costs.

Above examines are not chose random. Clearly this examples can be found in different others areas. But this two are very common for Romania, a country with a national airline company and several energy providers. It remains to be seen if Romania will become the one where simple economic aspects will determine also the change in mentality regarding IT solutions investments in order to generate profit growth [4].

2.1. Using Big Data in agriculture

As agriculture and agricol industry is considered one of the strategic engines of operating the romanian economy in the future, one case study has showed that using Big Data in an agricol area in USA might raise the interest to use such solutions to increase efficiency of production also in Romania.

Concret, is about a project taking place in **Flint Valley River**, a land specific for agriculture from south-west of **Georgia**. At the finish of this project, the landlords from the are will receive extrem precice forecast, each one for his specific farm from three days ahead.

Around this time, a team of resurcers from Flint River Soil and Water Conservation District, American Agricultural Department, **Georgia University** and **IBM** are using Big Data sophisticated tools to analyze big volumes of forecast weather data, geografic, historical and of other nature. The purpose is to pathern the forcast phenomenons to a very precise manner then the ones we already have available today.

This kinds of forecasts will help the farmers to make smart decisions related to watering

solutions, sowing, harvest and soil fertilization, that will permit them to better use the water supply and increase the volume of the harvest.

Flint River Valley represents an important part of Agricola industry in the state of Georgia. The farms from this region of 27 districts represent an annual income of around 2 billions dollars at the revenues of United States. With approximately ten years back, the authority of water conservation from this region signed a partnership with American Agriculture Department, University of Georgia and other local, regional and state agencies to promote the practices of conservation the use water between farmers of the area.

The effort done in the meantime has already started to provide important results. One so called **VRI** technology (**V**ariable **R**ate **I**rrigation) developed by resources from Georgia's University has passed fast from the concept state to a commercial product. Based on GPS, this technology allows farmers to arrange the water sprinklers, in such a way that the water will be use efficient only in the needed areas by redirecting the jets.

The project linked to Big Data using for efficient forecast pattern is the second big project developed by the entities named above and has no release date yet. The outcome at the end of this project will generate improved forecasts and also mobility by using intelligent devices like pads and smart phones.

3. USING BIG DATA FOR REDUCEING THE FROUDE RISK AND BRIBE

Over 63% of seniors executives parts of top companies worldwide, state that is required to improve the anti-fraude and anti-mita procedures by using forensic data analysis – as shown in a global study [5].

For this study, 450 executives from 11 countries were questioned, including profesionals from the financial area responsible with internal and conformity audit and legal aspects, were asked regarding the

use of forensic data in the internal applications use by each.

Forensic data analysis in a company represents a detailed focus on specific aspects with the purpose to identify the elements that are not between the given parameters or under the internal rules, business ethics or active laws. The study has shown, also, that 87% of the ones questioned about the regulation tasks, including the ones regarding the laws of anti-corruption and the new procedure of enforcing them, represents a motic force for including and using forensic data analysis in conformity programs, almost half of them have indicated that evolutions in the regulation area are consider in the top five principal factors in this field.

Mita and corruption are reported of 65% of the questioned ones as the being the biggest risk, in alingment with what 74% have stated that they use forensic data analysis on a daily bases to fight against mita and corruption. Others area in wich is perceived a notable risk is defraud, like highjacking the actives and false financial reports, are also priority domains for using forensic data analysis.

Through the inccres of cross border working, between regulation authorities and law enforcement agency, which have as end result significant corporation bills and also convictions for the executives, the administration comity must encourage the management teams of the company to use the tools given by forensic data analysis included in on going conformity programs.

3.1. Forensic data analysis increase the evaluation of risks and detecting defrauded process

One of the most important benefits obtained through forensic data analysis, mentioned by 89% of the ones questioned is the ability to detect the potential deviations which were not able to detect before. This opinion is common for most of them, regardless of the rank they have in the company.

The results of this report show that, meanwhile companies are investing in implementing some forensic data analysis systems, many others might miss important

opportunity to improve the actions against defraud and bribe. Through increasing value of the mix between different sources of data and advanced analysis tools for forensic data analysis, companies have the possibility to obtain new and important information through analyzing their business data.

3.2. Missed opportunities of transforming data into valuable information

Regardless the general positive phenomenon over the efficiency of forensic data analysis, the study shows that the majority of companies are not working with big enough data in comparison with the corporative income.

Only 18% of intern audit specialists interviewed are working with large volumes over a million records. Between them, the ones in financial business, only 21% have declaimed to work with large volumes of data of approximately or over milion recordings, a low percent for a concentration segment of data.

Per assemble, 71% of the companies with revenues over 1 billion USD are working with sets of data under 1 million recordings. The usage of smaller volumes of data than expected, in comparison with the companies revenues, are an indication that many companies can miss important opportunity to prevent and identify defraud just because they are not working with large volumes of data.

Forensic data analysis advance tools, like statistic analysis and mining data technologies, are used only by 10% of the ones questioned. Is not, as result, a big surprise the fact that the bigger challenge mentioned regarding forensic data analysis is "obtaining the necessary expertise or the suitable tools".

From the study, as conclusion, that the support of the efforts of using forensic data analysis inside the company seems like a key task. So, 62% of the questions ones have stated that is necessary to improve awareness in management team over the benefits given by forensic data analysis. Interesting is the fact that the one responsible are not considering the costs as a major obstacle, only

10% stated that forensic data analysis might have prohibit costs.

Romanian companies have implemented IT technologies and IT integrated solutions for business. Anyway, the lack in awareness' thirds defraud risk are not leading to invest more in solutions for detecting and preventing defraud.

4. CONCLUSIONS

Collecting and using data challenge is a phenomenon that grows globally, as answer to the needs of companies and peoples. In this way, European Commission (CE) invites national governments to keep up with the revolution started by large volumes of data available (Big Data phenomenon and Cloud Computing) [8].

In this way, to help citizens and companies of UE to use at full potential the value given by data, CE will work together with the Parliament and Council to efficiently finish the law frame for data protection. At the end, a Directive will be adopted regarding security of networks and information to insure a higher level of trust, critical aspect of a successful economy based on applications and data.

The fear around Big Data have no real grownd. In every minute, globally, are generated 1,7 millions of bilions bytes of data, the equivalence of 360.000 DVD: meaning over 6 megabytes data per person, daily. As follows, data sector is growing every year with 40%, 7 times faster then world market of information and communications.

Large volumes of data are already in use for speeding the process of diagnostic for brain lesions, to spot the best position for Aeolian parks, to avoid traffic jams or to forecast agricultural harvest in countries in development.

The technological value and services for large volumes of data on world level will grow to 16,9 milliard USD in 2016 also due to data by creating new hundred of thousand on jobs in Europe. The companies that make decisions based on generated knowledge by data are registering a growth in productivity by 5-6%.

5. REFERENCES

- [1] **St. Iovan, Gh. I. Daian**, (2012) *New Challenges: “Big Data” and “Consumer Intelligence”*, Annals of the “Constantin Brancusi” University of Targu Jiu, Engineering Series, Issue 4/2012, (**CONFERENG 2012**), ISSN: 1842 – 4856, pag. 318 -3 29;
- [2] **St. Iovan, M. Litra**, (2013) *Decision Making in Big Data Era*, Proc. of 14th European Conference (**E_COMM_LINE 2013**), Bucuresti, Romania, ISBN-13: 978-973-1704-23-4;
- [3] **M. B. Dinu, St. Iovan**, (2014) *Harnessing Big Data Volumes*, Proc. of 7th Symposium “Durability and Reliability of Mechanical Systems”, (**SYMECH 2014**), Polovragi – Gorj, pag. 250 - 256;
- [4] **St. Iovan, Cr. Ivanus**, (2014) *Business Intelligence and the Transition to Business Analytics*, Targu Jiu: “Academica Brancusi” Publisher, *Annals of the “Constantin Brancusi” University, Engineering Series*, Issue 4/2014, (**CONFERENG 2014**), ISSN: 1842 – 4856, pag. 150-156;
- [5] **EY**, (2014), *Big risks require big data thinking*;
- [6] * * *, www.bigdatavalue.eu (accesat in mai 2015)
- [7] **Sefan Iovan**, (2015) *Big Data Security Problems*, Proc. of 16th European Conference (**E_COMM_LINE 2015**), Bucuresti, Romania, ISSN: 2392-7240, ISSN-L: 2066-6829;
- [8] **St. Iovan, Cr. Ivanus**, (2014) *“Open Source” Technologies used in Cloud Computing Implementations*, Proc. of 7th Symposium “Durability and Reliability of Mechanical Systems”, (**SYMECH 2014**), Polovragi – Gorj, pag. 32 - 38;
- [9] **Cr. Ivanus, St. Iovan**, (2013) *“Internet of Things” – A new Technological Evolution*, Targu Jiu: “Academica Brancusi” Publisher, *Annals of the “Constantin Brancusi” University, Engineering Series*, Issue 4/2013, (**CONFERENG 2013**), ISSN: 1842 – 4856, pag. 165 - 170;