IMPORTANCE OF BIG DATA IN MARITIME TRANSPORT

Stefania Loredana NITA[1]
Marius Iulian MIHAILESCU[2]
[1]Integrated Systems Department, Institute for Computers, Bucharest, Romania,
stefanialoredanani@gmail.com
[2]RLC, Miami, Florida, mihmariusiulian@gmail.com

Abstract: As the technology grows fast, big data have become a major topic nowadays. It is interesting due to its nature, namely different types of data combined into one entity, but also because of the opportunities that it brings. Knowledge means power, so a good analysis and mining over huge amounts of data could bring many benefits and decisions could be made based on real time predictions. Maritime transport represents an important part of our lives, and it could be improved integrating big data, because all data is collected and then analyzed, and would help to avoid power failures of different components, based on prediction and different environmental data could be obtained in real time, based on analysis of past data. In this paper we will present the major trends in big data and we will analyze how it could be integrated in maritime transport.
Key words: big data, maritime, analysis

I. INTRODUCTION

Nowadays, the size of digital data is 2.8 trillion gigabytes (GB), and 90% is collected just in the last two years. It is predicted that in 2020 the amount of data will be 40 trillion GB [1]. The sources of data could be anything: communication, transport, social media, wheatear and climate, search engines, GPS signals, online shopping, mobile devices, and many others. It is very important and useful that these data to provide powerful knowledges, which will help to make better decisions in different areas.

Maritime transport is a very important aspect of our lives, beginning with defense and ending with trips. All components and sensors of a ship provide powerful data about everything regarded of that ship, which could be analyzed in order to obtain patterns and to find out how that patterns influence the ship over the time. A great opportunity is when more ships are inter-connected and data that is allowed to be provided from every ship is available for others. The data obtained from all ships on the world is so diverse and has a huge amount, becoming big data. It could be analyzed for providing predictions regarding the behavior of ship's components, digital systems, locations of other ships, customer behavior and so on.

The paper is organized in three sections, as follows: the introduction section, big data section, in which we present definitions and tools, and big data in maritime transport, in which we present opportunities and challenged and provide some relevant examples in this area.

II. BIG DATA
A. Definitions

The term of big data is older than we expected. It was used for the first time in 90s, being spread by John Marshey, who considered big data as "*storage growing bigger faster DRAM*" (Dynamic Random Access Memory) [2]. Today, we refer to big data as data sets that are very large and complex, containing different types of data: structured, unstructured and semi-structured. The sources of big data could be anything, but nowadays database systems cannot handle such datasets. In 2001 in a report of META Group (now Gartner) big data have been defined as a data with three dimensions that grows very fast [3].
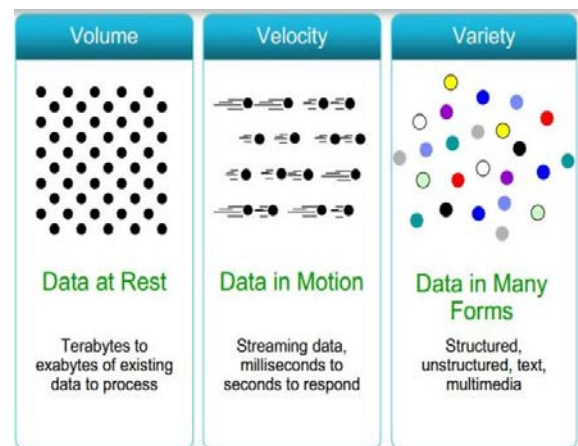

Figure 1. The three characteristics of Big Data [4]

The three dimensions of big data could be seen in Figure 1 and they are:
Volume. This characteristic refers to the large amount of data, obtained from very different and various sources, such as: machines, networks, social interactions.
Velocity. It is important to have many datasets, especially for prediction, but another important aspect is the speed at which the data is analyzed.
Variety. Data could have different forms, as structured, semi-structured or unstructured.
However, the above characteristics represent a first description of big data. Year by year, there

were added more Vs, such that, for the moment are seven Vs, coming from [5]:

Veracity. It refers to the accuracy of the data, and the ability to keep relevant data, preventing the system to accumulate useless data.

Variability. This characteristic is different from variety, in the sense that from the same data could be achieved different results applying the same analysis technique many times.

Visualization. It involves the ability that data to be processed and the results to be transformed into charts or graphics.

Value. Of course, this is the purpose of analysis of such amount of data.

The National Institute of Standards and Technology (NIST) definition for big data and big data paradigm are [6]:

"*Big Data consists of extensive datasets – primarily in the characteristics of volume, variety, velocity, and/or variability – that require a scalable architecture for efficient storage, manipulation, and analysis.*"

"*The Big Data paradigm consists of the distribution of data systems across horizontally coupled, dependent resources to achieve the scalability needed for the efficient processing of extensive datasets.*"

B. Tools

The main tools for big data are [7]:

- Storage and management. In order to store such amount of data, which is also variate it is need by specialized systems that need to allow among storage, the queries or running tools for analysis. Below, are the most used tools for storage and management:
  - Hadoop
  - Cloudera
  - MongoDB
  - Talend
- Cleaning. Because big data came from various sources, it is possible to exist useless data or redundant data, so, before analysis data should be cleaned. Dedicated tools are:
  - OpenRefine
  - DataCleaner
- Data mining. In this process there are found different patterns in datasets, whose purpose are predictions:
  - RapidMiner
  - IBM SPSS Modeler
  - Oracle data mining
  - Teradata
  - FramedData
- Data analysis. This is different to data mining. While the purpose of mining is to find patterns in data, analysis studies the effect of those patterns over dataset in time. Some tools are:

- Azure Machine Learning Studio
- Qubole
- BigML
- Statwing
- Data visualization. An important aspect of big data is to obtain values, so it is important to have the right tools for charts or graphs:
  - Tableau
  - Silk
  - CartoDB

C. Trends

According to TechRepublic, the main trends for big data in 2017 are [8]:

- Displacement to the cloud. Many small and middle companies have already moved to the cloud, but the large companies preferred own equipment and tools for data collecting and analysis. Together with growing amount and variety of data, it came many challenges in data processing. Thus, it is expected that many large companies to move to the cloud.
- Integration with Internet of Things (IoT). IoT represents a great data source. It comes from different devices, sensors, cameras and anything that could be connected to a network. For example, the popularity of drones have increased in the las few years, and now they are usually used for sites filming for different purposes.
- More analysis of unstructured data. Unstructured data consists mainly in photos, videos, documents, which represent a great source of historical facts about a company, showing their performance over the years.
- Security improvements. Data access permissions policies will be changed or improved such that every user from a company to have access just to the data they need to analyze.

To these, we could add [9]: more platforms of big data analysis, more investments in big data due to its variety, improved machine learning techniques and closeness to artificial intelligence.

III. BIG DATA IN MARITIME TRANSPORT
A. Challenges and solutions

Maritime Big Data provides many challenges, because big data itself is a relatively new technology. In Figure 2 there are summarized the main challenges [10].
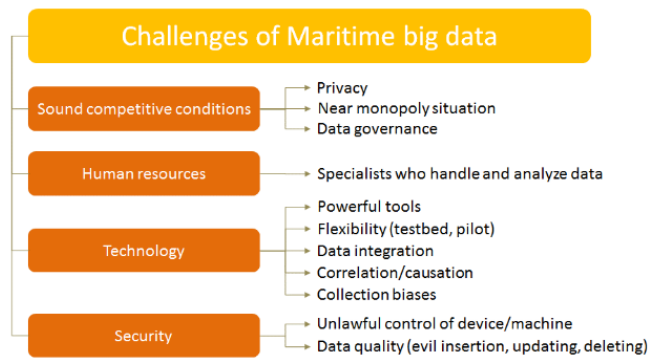
Figure 2. Challenges in maritime big data [10]

A source for these challenges is represented by the lack of well-defined standards for big data. If the data is collected and used by a community of a small numbers of companies, then it could be defined some rules such that to organize knowledge extraction in a secure way, but it is very possible to be needed a third party, which will establish through agreements or contracts who is the owner of the data and who could access the data. For the case in which the data is public and shared by everyone, it is obviously need of standards.

Further, it is needed data scientists, which analyze the collected data.

The next is represented by technology, because, as we have seen, the traditional database systems cannot handle big data. This represents a challenge only until a certain point, because we have seen in a previous section that there are many tools for big data handling.

Another important challenge is the security of the data, and this is really important. Because the data is shared by many entities, it should be established rigorous rules and policies for data accession, based on types of users with different rights. Moreover, data need to be protected against traditional attacks. A good solution could be mining and analysis over already encrypted data, but this is relatively new area of research and the current solutions need some improvements.

Now, let's see the advantages of adopting a big data system [11]:

- Power failures could be avoided based on predictions provided analyzing the data collected from sensors and components. Also, the maintenance schedule could be improved.
- Fuel quality is monitored in real time, and the performance could be improved if are made some analysis regarding fuel.
- The location of the ships could be found in real time, which could lead to a better management of supply chains. A good example of such platform is www.marinetraffic.com

- Both, physical and virtual attacks could be avoided, based on real-time predictive awareness.

In order maritime big data to be adopted, the things should be taken step by step. Computer Sciences Corporation (CSC) provides a data-driven approach, which contains five steps [11]:

1. Discover accessible data, its structure, and the possible values that could be obtained from it, and validation techniques.
2. Discover benefits form data mining and analysis and how could help to improve the company strategy.
3. Use many data sources for real time predictions which help to business decisions.
4. Adapt the business based on real-time prediction.
5. Make a proof of concept to find the importance of obtained values for the investments.

B. Examples of maritime big data

As we have seen, we can analyze datasets in order to obtain patterns or to discover the influence of these patterns in time. Nowadays almost 90% of the world commerce is realized through shipping transport [12].

Big data could have an important role in development of smart technologies for shipping industry, like communications between ships, a better planning of schedules, evaluation of performance for ships. Also, IoT will turn the way in which shipping companies work, if the accuracy of the data models will be very high.
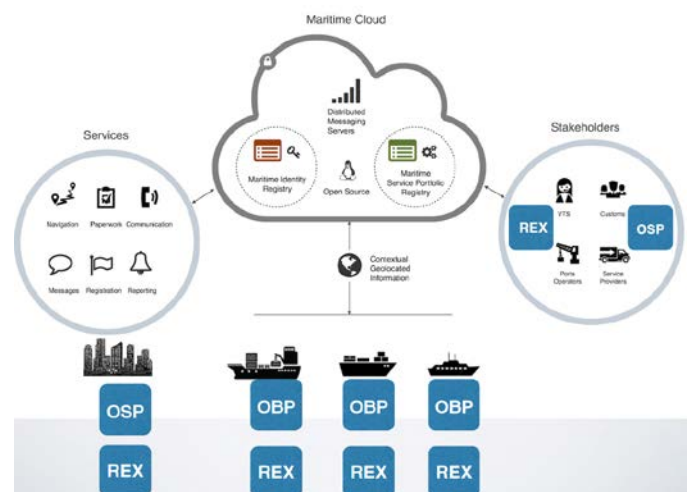

Figure 3. MARSEC system architecture [14]

In the last few years projects that involve big data have been initiated. In [13] and [14] it is described an IoT project-based developed for Oslo Fjord, in which data is collected from different sensors of the ships and is transmitted to the captain and the staff in real-time after a proper optimization. The results are promising: the routes have been

optimized, because the positions of ships are tracked in real time, the temperature of refrigerate containers (which need to have a constant temperature) is provided in real time, and, also the equipment is monitored in real time. In Figure 3 it is presented the architecture of the system (called MARSEC), where OBP means Open Bridge Platform, OSP means Open Shore Platform and REX means Route Exchange App.

A huge step was made by United States Pentagon, which released in 2016 a self-driving ship of 132-foot, that will be tested for two years off the San Diego coast by the Defense Advanced Research Projects Agency (DARPA) and Navy. The ship could travel 10.000 marine miles and its purpose is hunting hidden submarines or underwater mines [15].

Another example of collaboration between government and private sector is Maritime and Port Authority of Singapore (MPA), which have already initiated in 2013 a research and development program that helps shipping industry. In this project IBM was involved in 2015, whose task is to create a platform, in which data obtained from vessels or weather data is integrated in real-time. The purpose is to mine these data for obtaining different patterns for data point across MPA [16].

Another view of maritime transport is maritime tourism. In this area, Royal Caribbean International (RLC) intents to improve user experience through more technology. In 2014, the first smart ship of the world was released, namely Quantum of the Seas, which is equipped with Wi-Fi on sea, robotic bartenders, virtual balconies and Radio Frequency Identification (RFID) luggage tags [17]. All these are connected together. They could become an important source of information. After some mining and analysis techniques over collected data it could be obtained patterns in customers' behavior, which will reveal the performance over a period of time.

Conclusions

In this paper we have discussed about big data and its characteristics. We have seen that applying the right techniques of analysis and mining over huge datasets we can obtain useful information, such as patterns in the data or how that patterns influences the whole dataset (or the system from where data is collected). In marine, every day are collected various information about ships, their internal condition, or the interaction between ships. These data could be very useful and could improve the maritime if they are processed and knowledges are gathered in real time, helping for better decision making. We have seen that there are few projects in which big data or IoT are involved. We think this is an important step and big data will really help the marine.

Bibliography

[1] Esben Poulson, Big data in the digital age - the impact on ships and ports, The International Maritime Transport and Logistics Conference (MARLOG 5), 13-15 March 2016
[2] John R. Marshey, *Big data…and the next wave of IntraStress*, 25 April 1998.
[3] Doug Laney, *3D data management: Controlling data volume, velocity, and variety*, META Group Research Note, 2001, 6: 70.
[4] *Big data: What is and why it matters*, http://saphanatutorial.com/what-is-big-data/
[5] Eileen McNulty, *Understanding big data: The seven V's*, May 2014, http://dataconomy.com/2014/05/seven-vs-big-data/
[6] Nancy Grady, Wo Chang, *NIST Big data interoperability framework: volume 1, definitions*, 16 September 2015, http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1.pdf
[7] *All the best big data tools and how to use them*, https://www.import.io/post/all-the-best-big-data-tools-and-how-to-use-them/
[8] Mary Shacklett, *6 big data trend to watch in 2017*, 23 December 2016, http://www.techrepublic.com/article/6-big-data-trends-to-watch-in-2017/
[9] *Top 10 big data trends 2017*, https://www.tableau.com/resource/top-10-big-data-trends-2017
[10] Sadaharu Koga, *Major challenges and solutions for utilizing big data in the maritime industry*, 2015
[11] Anna Cebaseva, The possibilities big data could bring to maritime, 10 August 2006, https://blogs.csc.com/2016/08/10/the-possibilities-big-data-brings-to-maritime/
[12] *Big data boom in the shipping industry*, 22 December 2016, https://knect365.com/techandcomms/article/d8d85e1d-088a-4031-b235-2865f001a285/big-data-boom-in-the-shipping-industry
[13] Phillip Tracy, *Industrial Internet of Things maritime use cases*, 16 July 2016, http://www.rcrwireless.com/20160727/internet-of-things/maritime-industrial-internet-things-tag31-tag99

[14] Geir Fagerhus, Krystyna Wojnaro, Gisle Stava, *Buiding the Internet of Things @ sea*, The Maritime CIO Forum 2015, 2015

[15] Julie Watson, *Unmanned ships next on horizon?*, Associated Press, 5 May 2016, http://www.columbian.com/news/2016/may/05/unmanned-ships-next-on-horizon/

[16] *IBM and the Maritime and Port Authority of Singapore collaborate in research to improve maritime and port operations using analytics-based technologies*, http://www.mpa.gov.sg/web/portal/home/media-centre/news-releases/detail/6a054eff-e932-47bf-8558-ed73bc181ad9

[17] Hannah Sampson, How Royal Caribbean wants to make smart ships smarter, 15 July 2016, https://skift.com/2016/07/15/royal-caribbean-cio-on-making-smart-ships-smarter/