

Big Data in Logistics

Matea Lekić¹, Kristijan Rogić¹, Adrienn Boldizsár², Máté Zöldy², Ádám Török^{2*}

¹ Faculty of Transport and Traffic Sciences, University of Zagreb, 10000 Zagreb, Trg Republike Hrvatske 14, Croatia

² Department of Transport Technology and Economics, Faculty of Transportation Engineering and Vehicle Engineering, Budapest University of Technology and Economics, H-1111 Budapest, Műegyetem rkp. 3., Hungary

* Corresponding author, e-mail: torok.adam@mail.bme.hu

Received: 25 June 2019, Accepted: 08 July 2019, Published online: 17 December 2019

Abstract

With certainty, we can say that we are in the process of a new big revolution that has its name, Big Data. Though the term was devised by scientists from the area such as astronomy and genomics, Big Data is everywhere. They are both a resource and a tool whose main task is to provide information. However, as far as it can help us better understand the world around us, depending on how they are managed and who controls them, they can take us in some other direction. Although the figures that bind to Big Data can seem enormous at this time, we must be aware that the amount of what we can collect and the process is always just a fraction of the information that really exists in the world (and around it). However, from something we have to start!

Keywords

Big Data, data collection, distribution network, supply chain

1 Introduction

The complexity of business operations on the global market today is characterized by extremely complex competitive relationships and the complexity of market sustainability risks. Technological procedures have come to acceptable fruition level, and variables of cost input are reduced to sometimes infinitely low levels in the structure of the final price of products and services. The quality of market communication has also been developed through various specialized and specialized models and forms. Competitiveness between supply chains is becoming one of the most significant modern business paradigms (Govindan et al., 2018).

An essential component of business efficiency, regardless of the security, political, natural, social and economic environment factors is the business data management and information system, the chapter on aspects that determine cost inputs in the supply chain of a particular business entity.

Stanford University Research pointed (Srića and Urbanija, 1999) out that the total human knowledge gained until 1900 doubled up to 1950, and since then, the amount of world knowledge doubled every five to eight years. Recent estimates say that large data warehouses tripled every two years, and according to research conducted by Gartner, 47 % of data surveyors qualified as one of the three most significant challenges of modern management. The findings above point to the fact that

modern business has reached the level at which information becomes the most significant focus of contemporary management (Luetić, 2013, p. 256).

By these mentioned facts, in this paper, it will be described by Big Data influence in logistic systems and how it can be implicated in anticipating user needs. It will be described by theoretically part and case study. In theoretically part, it will be described the meaning of Big Data, how it has been developed, which are possibilities of application Big Data in logistic and describing it by case studies, by practical side.

1.1 Understanding the meaning of Big Data

An exact definition of "Big Data" is challenging to nail down because projects, vendors, practitioners, and business professionals use it quite differently. With that in mind, generally speaking, Big Data is:

- large datasets
- the category of computing strategies and technologies that are used to handle large datasets (Jain et al., 2017).

The determination "large dataset" means a dataset too large to reasonably process or store with traditional tooling or on a single computer. It means that the standard scale of Big Data sets constantly shifting and varies significantly

from organization to organization. Today, there are new challenges at designing solutions by Big Data. By the term of new challenges, it means the massive scale, the speed of ingesting and processing, and the characteristics of the data that must be dealt with at each stage of the process. There exist "three Vs of Big Data" to describe some of the characteristics that make big data different from other data processing, first presented by Gartner's Doug Laney in 2001. There are:

- Volume
- Velocity
- Variety

However, if businesses want to leverage information as a production factor and strengthen their competitive position, there is far more to consider than this 3 V such as visualization, veracity, variability, and value. It is necessary a shift in mindset and application of the right drilling techniques.

The sustained success and market growth of powerhouses such as Amazon, Google, Facebook, and eBay provide evidence of a fourth production factor in today's hyper-connected world, the importance of sustainable information at the right time in the right place. That is the reason why, besides resources, labor, and capital, there is no doubt that information has become an essential element of competitive differentiation.

1.2 The possibility of application Big Data in logistic systems - Logistics and Big Data are a perfect match

The ideally placed for benefiting from the technological and methodological advancements of Big Data is the logistics sector. Today a lot of logistics providers manage a massive flow of goods and at the same time create large data sets (Fig. 1). For millions of shipments, every day is tracked across a global delivery network, their origin, and destination, size, weight, content, and location. However, does this

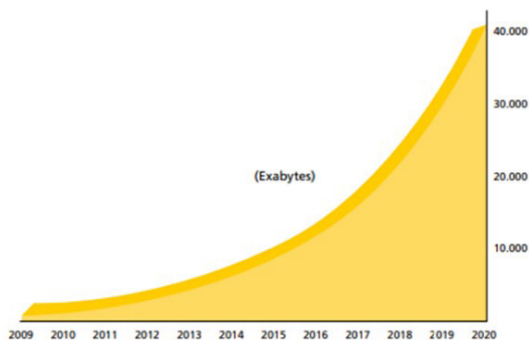


Fig. 1 Exponential data growth between 2010 and 2020 (Wegner and Kückelhaus, 2013)

data tracking fully exploit value? Probably not, most likely there is enormous untapped potential for improving operational efficiency and customer experience, and creating useful new business models. For example, the benefits of integrating supply chain data streams from multiple logistics providers could eliminate current market fragmentation, enabling powerful new collaboration and services. Big Data is a game-changing trend for the logistics industry, and many providers realize that (Anshari et al., 2018).

1.2.1 Logistics as a Data-driven business

A starting point for discussion of how to apply Big Data is to look at creating and consuming information. In the logistics industry, Big Data analytics can provide a competitive advantage because of five distinct properties. It is showing in the right way where Big Data can be most effectively applied in the logistics industry. They provide a roadmap to the well of unique information assets owned by every logistics provider. In this subtitled, we identify specific use cases that exploit the value of this information and contribute to operational efficiency, a better customer experience, or the development of new business models:

- Optimization to the core
- Tangible goods, tangible customers
- In sync with customer business
- A network of information
- Global coverage, local presence.

1.2.2 Last - mile optimization

A constraint in achieving high operational efficiency in a distribution network occurs at the "last mile" (Mikalef et al., 2019). This final hop in a supply chain is often the most expensive one. The optimization of last-mile delivery to drive down product cost is, therefore, a promising application for Big Data techniques. Two fundamental approaches make data analysis a powerful tool for increasing last-mile efficiency. In a first and somewhat evolutionary step, a massive stream of information is processed to maximize the performance of a conventional delivery fleet further. This is mainly achieved by real-time optimization of delivery routes. The second, more disruptive approach utilizes data processing to control an entirely new last-mile delivery model. With this, the raw capacity of a massive crowd of randomly moving people replaces the effectiveness of a highly optimized workforce (Wegner and Kückelhaus, 2013).

Route optimization on the last mile aims at saving time in the delivery process. Rapid processing of

real-time information supports this goal in multiple ways. Every delivery vehicle receives a continuous adaptation of the delivery sequence that takes into account geographical factors, environmental factors, and recipient status. Each driver receives from the navigation system, instant driving direction updates, guiding them to the next best point of delivery (Arunachalam et al., 2018).

1.2.3 Strategic network planning

From the point of a strategic level, the topology and capacity of the distribution network are adapted according to predict future demand. More accurate capacity demand forecasts increasing efficiency and reducing the risks of investing in storage and fleet capacity. Big Data techniques support network planning and optimization by analyzing large historical capacity and utilization data of orders by client, warehousing administration, and transportation routes. An influence of macroeconomic factors (such as industry-specific and regional growth forecasts) is included for more accurate prediction of specific transportation capacity demand.

The result of increasing predictive value and much higher volume and variety of information is a new quality of planning with expanded forecast periods – this effectively reduces the risk of long-term infrastructure investments and contracted external capacities. It can help with accelerating sales volume by exposing over-capacity, providing this as automated feedback (Boone et al., 2019).

1.2.4 Business intelligence in Croatian companies

From the sources of research, it can be concluded that companies are generally inclined to unify data and information at service levels or departments rather than processes and projects (Nedeliakova et al., 2016). Interestingly, the highest degree of agreement has been recorded in tourism and construction. The most significant tendency to organize data and information at the service level has been shown by large companies, while it is less prudent to work at the project level. By analyzing responses based on legal form, it is noticed that companies that have established a foreign founder prefer process unification, while limited liability companies are more inclined to unify at the level of their departments or services (Govindan et al., 2018).

2 Methodology - Influence of Big Data on changing customer experience

Achieving the view of what customer want and profile of the customer is the aspect of Big Data analytics that

currently attracts the most attention. For every business, the most important is to maximize profit and minimize risk (Čokorilo, 2008). One way could be to learn about customer demand and satisfaction. However, as organizations experience increased business success, sometimes happened that individual customer can blur into a large and anonymous customer. Companies often forgot on the individual customer and the analytics of Big Data is helping to win back individual customer insight and to create targeted customer value.

By applying techniques such as semantic text analytics, natural language processing, and pattern recognition, many logistics providers can avoid the problem of losing every single potential customer. Although business relationships in logistics are usually in benefit to the sender side, modern loyalty management must also target the recipient side. The poor service quality often has more influence on recipients, and their feedback influences sender selection for future shipments. An excellent example of that will be explained in Section 5 by a case study. Recurring customer complaints result in the vendor considering a switch of the logistics provider. To include recipients into loyalty management requires yet more data to be processed, especially in B2C markets.

3 Results - Case study of showing influence Big Data on the distribution network

In this article, business information was analyzed. Important data information was gathered in a warehouse, showing how many shipments were made, to which countries and cities. Full addresses (streets and zip codes for loading and unloading) were recorded for each client. Time stamps were recorded:

- the date and time when the shipment had arrived at the warehouse
- the date and time when it was on delivery.

Also volume and weight were recorded. Data is displayed in 34 columns and 56 802 rows, and it is related from 5. March 2018., to 7. June 2019. The data was in Excel format. Date and time were recorded as a string, so the transformation was required for further analysis. With this, the date and time could be analyzed as a dependent value of volume and time. The data includes national and national freight routes as well as freight picking lists. The paper will show the influence of time on weight, volume, and quantity, and also the interdependence of these three parameters.

3.1 Quantity

The mean value of a quantity that warehouse receiving and sending out from the warehouse is 12.71. It means that average quantity for this data's period we evaluated is 12.71 pieces. The standard error is 0.24, which represents the difference between the actual measured quantity and arithmetical mean. Median is 2, which shows the middle value in the dataset, At least 50 % of customers sent at least 2 packages. The mode is 1, which shows the most common value in the dataset was 1, which means mostly customers sent one package in the investigated period. Standard deviation is 59.43, which show the sample standard deviation measure for the data set. The simple variance of the pattern is 3532.225 and represents the measure of the variability of a data set that indicates how far different values are spread. Minimum quantity is 0 maximum 3 400. The minimum quantity is 0 that means that the package damaged or loss.

3.2 Volume

The mean value of the volume is 1.41 m³ the standard error of measuring is 0.01 m³. The middle value is 0.373 m³ that means 50 % of packages were smaller than 0.373 m³. Most common value is 0.048. The standard deviation measure for the data set is 4.718088863. The minimum value of the volume is 0 m³, and the maximum value is 101.839 m³.

3.3 Weight

The value of mean by the descriptive statistic of package weight is 171.2 kg, the standard error is 2.3 kg. The median is equal to 47.94 kg that means half of the packages were less than 50 kg. The mode value is 60 kg that means that is the most frequent mass of packages. Minimum weight is 0 kg, and the maximum is 16 420 kg.

4 Analysis of logistic performance

4.1 Quantity

From Fig. 2, it can be stated that the company needs less time to prepare for smaller quantities. Out of the 155 merchandise, the quantity becomes more and more so the warehouse needs more time to prepare the goods. On the vertical axis is the average of the time between arriving time in warehouse and time when cargo is ready for loading, and on horizontal axes is quantity.

Fig. 3 shows that the most common quantity is between 1 and 10. This means that the most significant quantity that came to the warehouse during the processing period ranged from 1 to 10. On the horizontal axis is quantity, and on the vertical axis is the frequency of repeating several quantities.

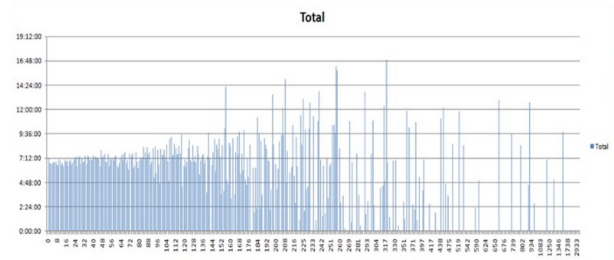


Fig. 2 Impact of the number of goods on time for packing goods
Source: own compilation

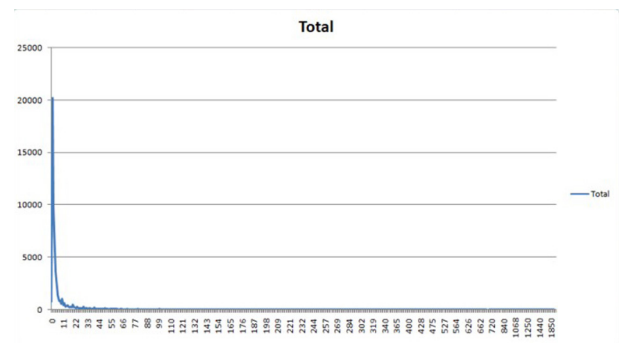


Fig. 3 Frequency of repeating quantity of goods
Source: own compilation

Fig. 4 shows that 0.042 package takes 1 m³ in the warehouse. In average, one packing using 23, 80 m³ of the warehouse. We improved it by linear regression. On the vertical axis is volume, and on the horizontal axis is quantity.

Fig. 5 shows that most of the quantities came to the warehouse in January and August, but also in March and May. Analytical data include the average incoming quantity of goods and the dates from March 5, 2018, to June 7, 2019. The minimum value of quantity has arrived at the warehouse was 0 at 22. June 2018. Moreover, the most significant value of quantity arrived at the warehouse was 400 at 08. October 2018. On the vertical axis is the quantity, and on the horizontal axis are the dates.

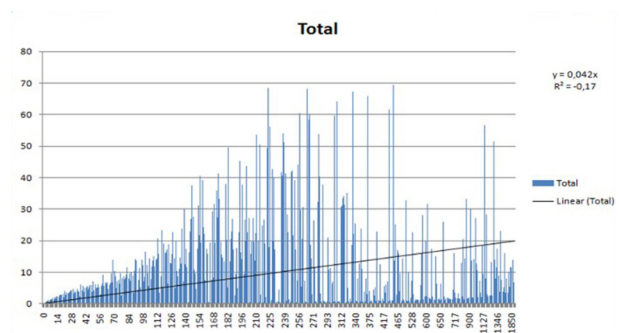


Fig. 4 Connection between volume and quantity in taking the place of warehouse
Source: own compilation

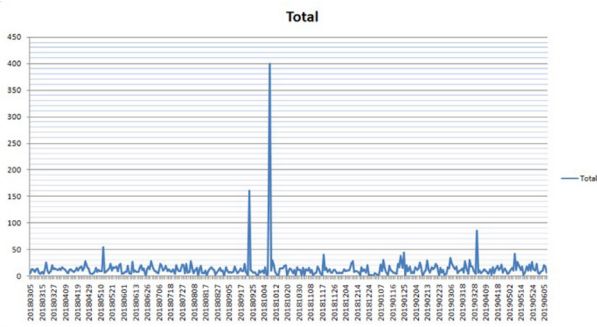


Fig. 5 Demand for quantity per months
 Source: own compilation

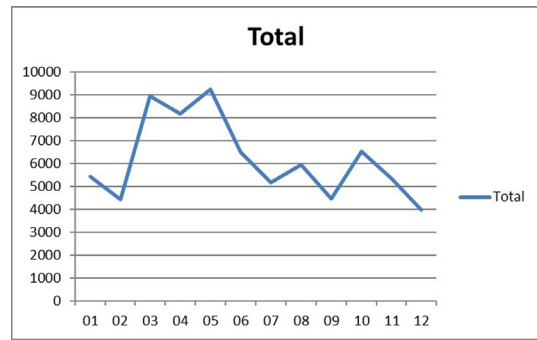


Fig. 8 Demand of volume per months
 Source: own compilation

4.2 Volume

Fig. 6 clearly shows that there is no correlation between volume and time for packing goods. On the vertical axis is time and on the horizontal axis is volume can be seen.

Fig. 7 shows that the most common value of the volume is 0.048 m³ and it has been repeated 2297 times. On the vertical axis in the frequency of repeating value of volume and on the horizontal axis is the value of volume.

Fig. 8 represents the periods of the year when the most significant value of volume comes to the warehouse. It is clear that it happens at springtime, in March, April, and May.

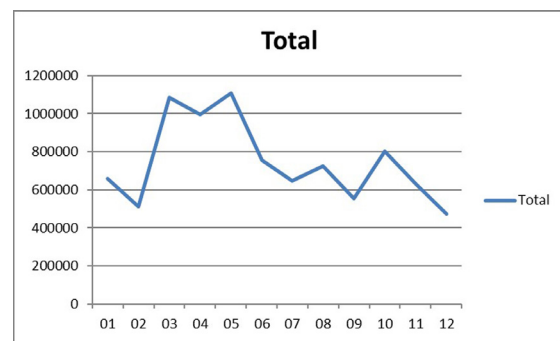


Fig. 9 Demand of weight per months
 Source: own compilation

4.3 Weight

Fig. 9 shows in which time of the year, there is the most significant value of weight. It is clear that during March, April, and May, the same pattern as volume.

The most common value of the weight was 60 kg, precisely 646 times. On the vertical axis is the frequency of repeating value of the weight and on the horizontal axis is the value of kilograms.

Fig. 10 it can be seen that the influence of weight on the period of packing goods is insignificant. There are no specific influences of weights on packing time as we can see in Fig. 11.

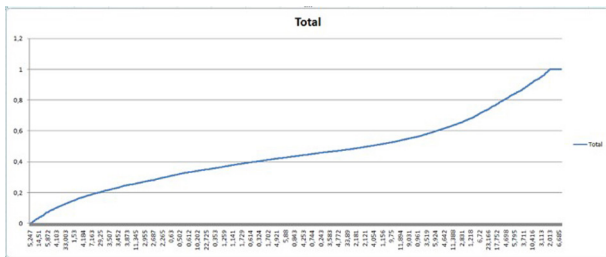


Fig. 6 Correlation between volume time for packing goods
 Source: own compilation

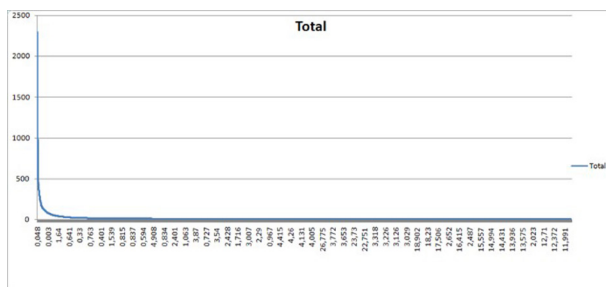


Fig. 7 Frequency of repeating value of the volume of goods
 Source: own compilation

5 Conclusion

There is no doubt that Big Data is a concept which is changing the world, at the first side business world, and it is clear

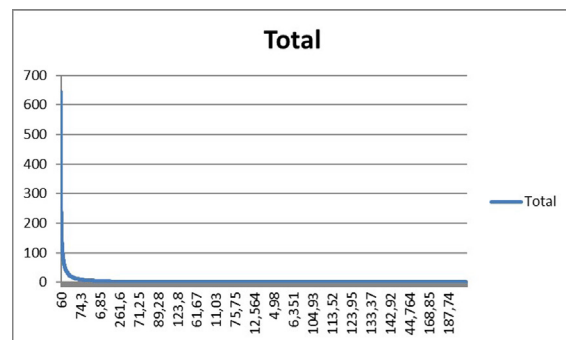


Fig. 10 Frequency of repeating value of weight
 Source: own compilation

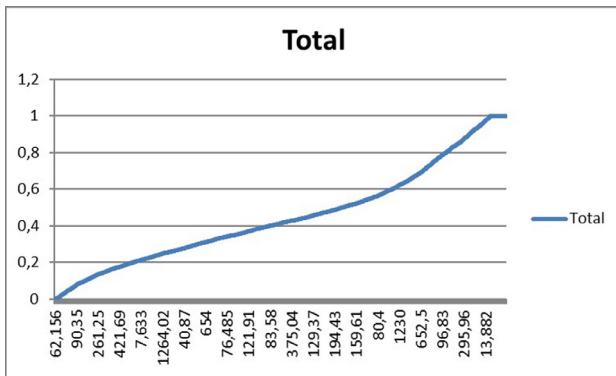


Fig. 11 Influence of weight on time for packing goods
Source: own compilation

that information has become the fourth production factor and essential to competitive differentiation. Of course, there are many obstacles to overcome, such as technical feasibility, data quality, speed of processing data, privacy, etc. There are tremendous opportunities to implement Big Data's concept into the logistics sector as we described in network planning, last-mile optimization, and changing the customer experience. The companies in today's market have to follow new trends if they want to survive. It is clear that it is not only important how much data one has, but what one should do with it.

Acknowledgment

Authors are much acknowledging the support of CEEPUS CIII-RS-0304-12-1920-M-133314 network.

References

- Anshari, M., Almunawar, M. N., Lim, S. A., Al-Mudimigh, A. (2018) "Customer relationship management and big data enabled: Personalization & customization of services", *Applied Computing and Informatics*, 15(2), pp. 94–101.
<https://doi.org/10.1016/j.aci.2018.05.004>
- Arunachalam, D., Kumar, N., Kawalek, J. P. (2018) "Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges and implications for practice", *Transportation Research Part E: Logistics and Transportation Review*, 114, pp. 416–436.
<https://doi.org/10.1016/j.tre.2017.04.001>
- Boone, T., Ganeshan, R., Jain, A., Sanders, N. R. (2019) "Forecasting sales in the supply chain: Consumer analytics in the big data era", *International Journal of Forecasting*, 35(1), pp. 170–180.
<https://doi.org/10.1016/j.ijforecast.2018.09.003>
- Čokorilo, O. (2008) "Risk management implementation in aircraft accident cost analysis", In: 12th Annual World Conference, Air Transport Research Society (ATRS) World Conference, Athens, Greece, Article ID 0087.
- Govindan, K., Cheng, T. C. E., Mishra, N., Shukla, N. (2018) "Big data analytics and application for logistics and supply chain management", *Transportation Research Part E: Logistics and Transportation Review*, 114, pp. 343–349.
<https://doi.org/10.1016/j.tre.2018.03.011>
- Jain, A. D. S., Mehta, I., Mitra, J., Agrawal, S. (2017) "Application of Big Data in Supply Chain Management", *Materials Today: Proceedings*, 4(2), pp. 1106–1115.
<https://doi.org/10.1016/j.matpr.2017.01.126>
- Luetić, A. (2013) "Poslovna Inteligencija i upravljanje opksbnim lancem" (Business intelligence and hazardous chain management), PhD Thesis, University of Split. (in Croatian)
- Mikalef, P., Boura, M., Lekakos, G., Krogstie, J. (2019) "Big data analytics and firm performance: Findings from a mixed-method approach", *Journal of Business Research*, 98, pp. 261–276.
<https://doi.org/10.1016/j.jbusres.2019.01.044>
- Nedeliakova, E., Sekulova, J., Nedeliak, I., Abramovic, B. (2016) "Application of Raymond Fisk Model in Research of Service Quality", *Communications-Scientific Letters of the University of Zilina*, 18(2), pp. 11–14.
- Srića, V., Urbanija, A. (1999) "Ustvarjalno mišljenje" (Creative thinking), *Gospodarski Vestnik*, Ljubljana, Slovenia. (in Slovenian)
- Jeske, M., Grüner, M., Weiß, F. (2013) "Big Data in Logistics", [pdf] DHL Customer Solutions & Innovation, Troisdorf, Germany. Available at: https://www.dhl.com/content/dam/downloads/g0/about_us/innovation/CSI_Studie_BIG_DATA.pdf [Accessed: 18 June 2019]