BIG DATA: CHALLENGES AND OPPORTUNITIES IN LOGISTICS SYSTEMS

Branka Mikavica[a*], Aleksandra Kostić-Ljubisavljević[a*], Vesna Radonjić Đogatović[a]

[a] University of Belgrade, Faculty of Transport and Traffic Engineering, Serbia

Abstract: *Advancements in telecommunications and computer technologies have led to exponential growth and availability of data, both in structured and unstructured forms. The term Big Data mainly refers to enormous datasets containing large amount of unstructured data that require more real-time analysis. Great potential and very useful values are hidden in this huge volume of data. The influence of Big Data is recognized in logistics services, turning large-scale data volumes into a unique asset capable of boosting efficiency in areas of the business. This paper analyses benefits and opportunities of Big Data in logistics systems. Challenges and risks that logistics systems, affected with this phenomenon deal with, are highlighted. The paper also proposes some efficient ways of exploiting the value of Big Data in logistics systems.*

Keywords*: Big Data, logistics systems, challenges, opportunities*

1. INTRODUCTION

With the permanent increase of data, scaling up to noteworthy amounts, generated by Internet-based systems, Big Data has emerged as a new research field. The core of Big Data paradigm is the extraction of knowledge from data as a basis for intelligent services and decision making systems. It encompasses many research topics, disciplines and it investigates a variety of techniques and theories from different fields, including data mining, machine learning, informational retrieval, analysis etc. Big Data has indubitably become important trend in logistics systems. Optimization of service properties, such as delivery time, resource utilization and geographical coverage, is a permanent challenge in logistics systems. Large-scale operations in logistics systems require data in order to work efficiently. Optimization results depend on information accuracy and availability. Integration of logistics providers and customer operations provides comprehensive knowledge related to supply chain risks. The transport and delivery network is important data source. Beside network optimization itself, network data may provide insight on global flow of goods. Hence, Big Data analytics emphasize a microeconomic viewpoint. Big Data concept provides multilateral analytics in order to enable the insight on customer experience and product quality.

The paper is organized as follows. After introductory remarks, a survey of some proposed Big Data definitions, tools and techniques is given.

* b.mikavica@sf.bg.ac.rs

The third section analyses potential benefits and opportunities of Big Data application in logistics systems. Steps for efficient exploitation of the values in Big Data are also presented in this section. The forth section observes challenges and risks that logistics systems need to cope with in order to ensure efficient business platforms. Concluding remarks are given in the last Section.

2. BIG DATA DEFINITIONS, TOOLS AND TECHNIQUES

Considering vast increase of global data and advancement of information technology which enable data generation, the term Big Data is mainly used to describe enormous datasets which typically include masses of unstructured data requiring more real-time analysis by Chen et al. (2014). It provides new opportunities for discovering new values, better understanding those values, but also raises new challenges including organization and management of these datasets. Beside masses of data, there are many other features determining the difference between Big Data and “massive data” or “very big data”. These differences and different concerns in scientific and technological aspect cause different definitions of Big Data. Big Data can be described as datasets which could not be captured, managed, and processed by general computers within an acceptable scope by Chen et al. (2014). In accordance with this definition, Big Data presents the next frontier for innovation, competition, and productivity by MGI (2011). It refers to such datasets which could not be acquired, stored, and managed by classic database software. This means that volumes of datasets conforming to the standard of Big Data are changing, increasing over time or with technological improvement. Also, these volumes in different applications differ from each other. Most common explanation of Big Data is 3Vs (Volume, Velocity, Variety) model by Laney (2001). Volume refers to vast data scale increasing with the generation and collection of masses of data. Timeliness of Big Data is described by Velocity. Hence, data collection and analysis must be rapidly and timely conducted in order to utilize the commercial value of Big Data. Variety indicates the various data types, including unstructured, semi-structured and traditional structured data as well. Another definition describes Big Data as a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling the high-velocity capture, discovery, and/or analysis by IDC (2012). Thus, characteristics of Big Data can be summarized as 4Vs (Volume, Velocity, Variety and Value). This definition is wide spread since it emphasizes the meaning and necessity of Big Data including exploration of huge hidden values from data with an enormous scale, various types, and rapid generation. Adding Veracity as a key characteristic of Big Data, 5Vs model is created by Hassanien et al (2015). More detailed definition by Beyer and Laney (2012) describes Big Data as high-volume, high-velocity, and/or high-variety information asset that require new forms of processing to enable enhanced decision making, insight discovery and process optimization. Hence, data set can be referred as Big Data if it is formidable to perform capture, storage, distribution, management, analysis and visualization on it at the current technologies.

There are many proposed techniques and technologies to capture value from Big Data, analyze and visualize these data by Philip Chen and Zhang (2014). However, they are not entirely capable to meet the requirements. These techniques and technologies combine a number of disciplines, including computer science, economics, mathematics, statistics and other expertise. Multidisciplinary methods are needed for determination of valuable information from Big Data. These techniques involve many different disciplines, such as statistics, data mining, machine learning, neural networks, social network analysis, signal processing, pattern recognition, optimization methods, visualization approaches and they overlap with each other. Statistics is used to collect, organize, interpret data and to provide numerical descriptions. These techniques are used to determine correlations and causal relationships between different objectives. However, standard statistical techniques are not entirely convenient for management of Big Data. Many extensions of classical statistical techniques or completely new methods have been proposed by Di Ciaccio et al (2012). Data mining is a set of techniques to extract valuable information from data, including clustering analysis, classification and regression. Big Data mining is more complex in comparison with traditional data mining algorithms. Machine learning is aimed to design algorithms that allow computers to evolve behaviors based on empirical data. Important characteristics of machine learning are discovery knowledge and making intelligent decisions automatically. Considering Big Data, both supervised and unsupervised learning algorithms need to be scaled up.

3. BENEFITS AND OPPORTUNITIES OF BIG DATA APPLICATION IN LOGISTICS SYSTEMS

The barriers to extracting value from Big Data can overcome through systematic plan by WEF (2014). The first step is to define responsibilities and roles for collection and analysis of data. The next step is determine how Big Data might be valuable, since the aim of Big Data is not data by itself, but discovering insights that can lead to valuable outcomes. Valuable business insight can be obtained from various sources, including social media, activity streams, machine instrumentation, operational technology feeds and data currently unused but have already been captured. These sources need to be explored in order to find new ways of capturing information, such as complex event processing, video search and text analytics. Big Data initiatives should be launched in business functions for which the potential payback is high. Functions such as marketing, customer service, supply chain management and finance are poised for the greatest growth. Another step is to match Big Data initiatives with compatible business functions. Issue of great importance is determination weather Big Data will yield valuable information unavailable through traditional business analytics. Also, complexities, priorities and technology architecture need to be assessed accordingly. Finally, multidisciplinary team of business and technology experts is the key for success.

The advancements in technological and methodological aspect of Big Data provide great benefits to the logistics sector. Logistics providers manage enormous flow of products thus creating vast data sets. Origin, destination, size, weight, content and location of shipments on every day basis are tracked across global delivery networks. There is great unutilized potential for improving operational efficiency, customer experience and creating new business models. Big Data analytics provides competitive advantage through properties which highlight where Big Data can be effectively applied in the logistics industry according to DHL (2014). Some of the most important fields of Big Data application in logistics systems are shown in Figure 1.

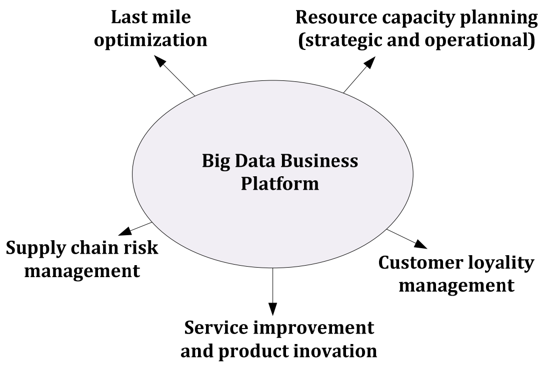

Figure 1. Big Data Application in Logistics Systems

Big Data analytics can accelerate business processes and increase the level of operational efficiency enabling last mile optimization. This goal can be achieved through real-time optimization of delivery routes or utilizing data processing in order to control an entirely new last-mile delivery model. For such purposes sensor-based detection of shipment items can be used or automatically changing of delivery routes according to current traffic conditions. The automated control of a large number of randomly moving delivery resources requires extensive data processing capabilities of Big Data techniques. This can decrease last-mile delivering costs, especially in rural and seldom populated areas. Optimal utilization of resources is one of the most important competitive advantages for logistics providers. Hence, logistics providers must apply resource planning on strategic and operational level. Planning on strategic level involves the long-term configuration of the distribution network, while operational level planning scales capacities on a daily or monthly basis. In both cases, Big Data techniques improve the reliability of the planning and enable logistics providers to optimally match demand and available resources. In order to significantly increase predictive value, a much higher volume and variety of data is exploited by advanced regression and scenario modeling techniques. This results in a new quality of planning with greater forecast periods. Hence, the risk of long-term infrastructure investments and contracted external capacities is reduced. In the case of planning on operational level, transit points and transportation routes must be efficiently coordinated on day-to-day basis. Operational planning tasks are often based on historical averages or even on personal experience. As a result, resource utilization is inefficient. Real-time information about shipments is aggregated to predict the allocation of resources. This data is automatically sourced from warehouse management systems and sensor data along the transportation chain. Also, detection of changes in demand is derived from externally available customer information. The prediction of resource requirements enables scaling capacity in each location. In addition, it reveals upcoming congestions on routes or at transit points which cannot be addressed by local scaling. The distribution network can become self-organizing infrastructure using Big Data analytics.

Acquisition of customer insight is of great importance in the aspect of Big Data analytics. Data from the distribution network carries meaningful value for the analysis and management of customer relations. Application of Big Data techniques enables understanding of customer demand. Big Data analytics allow a comprehensive assessment of customer satisfaction by merging multiple extensive data sources. In order to achieve the insight across the entire customer base, the logistics provider must merge multiple data sources. Big Data analytics are essential in creating an integrated view of customer interactions and operational performance, ensuring satisfaction of both sender and recipient. In order to get accurate results from customer feedback evaluation, it is necessary to aggregate information from as many touch points as possible. The open research problem in logistics systems and supply chain management can be analyzed from the perspective of managerial business components and the different category of stakeholder, where main business functions are forecasting, inventory management, transportation management, transport and human resources by Robak et al (2014). Issues such as prediction of time delivery, timely response to customer experience, real-time planning of capacity availability, inventory, customer and supplier relationship management can be addressed by Big Data. Considering data not only as information asset, but also as a strategic asset, organizations in supply chain management can realize economic value in the data using Big Data analytics through revenue generating activities by Rozados and Tjahjono (2014).

4. CHALLENGES AND RISKS OF BIG DATA IN LOGISTICS SYSTEMS

The effective use of Big Data techniques introduces great advantages in economy transformation, but also raises many challenges, including, among others, difficulties in data capture, storage, searching, shearing, analysis and visualization. These challenges need to overcome in order to exploit capabilities of Big Data. Computer architecture is one of the greatest challenges. Central Processing Unit (CPU) performance is doubling each 18 months, according to the Moore's law by Philip Chen and Zhang (2014). Performance of the disk drives is also doubling at the same rate, but rotational speed of the disk has slightly improved. In addition, the amount of information increases exponentially. This has a big impact on limitation of real-time values discovery from Big Data. Another important challenge related to the Big Data analysis includes data inconsistence and incompleteness scalability, timeliness and data security. Hence, data must be appropriately constructed and a number of preprocessing techniques, such as data cleaning, data integration, data transformation and date reduction need to be applied in order to alleviate noise and correct inconsistencies. Big Data has significantly changed data capturing and storing, including data storage device, data storage architecture, data access mechanism. The knowledge discovery process puts the highest priority on the accessibility of Big Data. In that sense, Big Data should be accessed efficiently and enabled to fully or partially break the constraint of computer architecture. Direct-attached storage (DAS), network-attached storage (NAS), and storage area network (SAN) are often used storage architecture. However, they have severe drawbacks and limitations in large-scale distributed systems. Optimizing data access is common way of improving the performances of data-intensive computing. This includes data replication, migration, distribution and access parallelism. When data volume is enormous, network bandwidth capacity is the bottleneck in cloud and distributed systems. Another issue related to cloud storage is data security. Data curation is aimed at periodically data discovery and retrieval, data quality assurance, value addition, reuse and preservation. This includes authentication, archiving, management, preservation, retrieval and representation.

In logistics systems, a comprehensive analytics framework requires integration of supply chain management, customer management, after-sales support and advertising by Kambatla et al (2014). Enormous amounts of multi-modal data including customer transactions, inventory management, store-based video feeds, advertising and customer relations, customer preferences and sentiments, sales management infrastructure, and financial data among others. Comprehensive deployment of RFIDs to track inventory, links to supplier's databases, integration with customer preferences and integrated financial systems provide improved efficiency. Big Data approach facilitates exploitation of RFID-enabled manufacturing data for supporting production logistics decision-makings by Zhong et al (2015). These applications mostly have relatively well structured and integrated datasets. Since infrastructure and data analysis is being performed in the same security domain, privacy and security issues are easier to handle. The major bottleneck in this domain is the development of analytics capable to scale vast amounts of multimodal data.

With increasing data volume, the probability that the data contain valuable and confidential information raises. Thus, information stored for the purpose of Big Data analytics is vulnerable for cyber criminal by Kshetri (2014). Besides, availability of personal data can be used to create value. Another important issue is determination of relevance within enormous data volume and usage of Big Data analytics for creating value from relevant data. Using such data, different products can be offered to different groups through quality discrimination and differential pricing. Big Data analytics enables determination of variables with a much higher correlation than in non-Big Data techniques. It also designs offerings and set prices based on such variables. Combination of structured and unstructured data for various sources, hidden connections between outwardly unrelated data can be revealed. Security problems also include intellectual property protection, personal privacy protection, commercial secrets and financial information protection by Philip Chen and Zhang (2014). Data protection laws are already established in most developing and developed countries. For Big Data related applications, data security problems are harder to deal with because of extremely large amount of Big Data and much more difficult workload of the security.

5. CONCLUSION

This paper presents overview of benefits, business opportunities of Big Data in logistics systems, but also emphasizes challenges and potential risks in this domain. Big Data analytics is still in the initial phase of development, since existing Big Data tools and techniques are not capable to entirely meet the requirements. More efforts from various fields of expertise need to be involved in order to more efficiently exploit all hidden values in vast datasets. There are numerous obstacles and challenges to deal with, such as data quality, privacy, technical feasibility, among others, before Big Data can achieve widespread influence in the logistics sector. In the long run, these challenges are likely to be solved, considering venturesome character of Big Data. Big Data have a potential to improve operational efficiency, customer experience and to create new business models. Therefore, it is reasonable to predict that Big Data will be important success driver in logistics sector.

ACKNOWLEDGMENT

This work is partially supported by Ministry of Education, Science and Technological Development of the Republic of Serbia under No. 32025.

REFERENCES

[1] Beyer, M., Laney, D. (2012).The Importance of Big Data: A Definition. Gartner.

[2] Chen, M., Mao, S., Liu, Y. (2014). Big Data: A Survey. Mobile Networks and Applications, 19 (2), 171-209.

[3] DHL. (2014). Big Data in Logistics

[4] Di Ciaccio, A., Coli, M., Ibanez, A., Miguel, J. (2012). Advanced Methods for the Analysis of Large Data-Sets, Springer.

[5] Hassanien, A.E., Snasel, V., Kacprzyk, J. (2015). Big Data in Complex Systems, Challenges and Opportunities, Springer.

[6] IDC. (2012). Market Analysis, Worldwide Big Data Technology 2012-2015 Forecast.

[7] Kambatla, K., Kollias, G., Kumar, V., Grama, A. (2014). Trends in Big Data Analytics. Journal of Parallel and Distributed Computing, 74(7), 2561-2573.

[8] Kshetri, N. (2014). Big Data's Impact on Privacy, Security and Consumer Welfare. Telecommunication Policy, 38(2014), 1134-1145.

[9] Laney. D., (2001). Meta Group. Application Delivery Strategies.

[10] McKinsey Global Institute. (2011). Big Data: The Next Frontier for Inovation, Competition, and Productivity.

[11] Philip Chen, C.L., Zhang, C.Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Sciences, 275(2014), 314-347.

[12] Robak, S., Franczyk, B., Robak, M. (2014). Research Problems Associated with Big Data Utilization in Logistics and Supply Chain Design and Management. Annals of Computer Science and Information Systems, 3(1), 245-249.

[13] Rozados, I.V., Tjahjono, B. (2014). Big Data Analytics in Supply Chain Management: Trends and Related Research. 6 International Conference on Operations and Supply Chain Management, Bali.

[14] World Economic Forum. (2014). The Global Information Technology Report: Rewards and Risks of Big Data.

[15] Zhong, R.Y., Huang, G.Q., Lan, S., Dai, Q.Y., Zhang, T. (2015). A Big Data Approach for Logistics Trajectory Discovery from RFID-enabled Production Data. International Journal of Production Economics, accepted manuscript.