Big Data Analytics in Supply Chain: A Literature Review

Mohamed Awwad, Pranav Kulkarni, Rachit Bapna and Aniket Marathe
Industrial and Systems Engineering Department
University at Buffalo, The State University of New York
Buffalo, NY 14260, USA
maawwad@buffalo.edu, pskulkar@buffalo.edu, rachitba@buffalo.edu, marathe2@buffalo.edu

Abstract

With the world moving towards the Industry 4.0 Standard, the number of machines, processes, and services generating and collecting large quantities of data will increase greatly in the future. This will give rise to Big Data, which is enormous amounts of data that cannot be processed with conventional computation techniques. To uncover patterns in the huge amounts of data and gain valuable insights on it, Big Data Analytics is devised. Supply Chain is a significant contributor to Big Data wherein the diversity of information is large. The data accumulated by Supply Chain contains information from the key entities such as manufacturing, logistics, and retail. The use of Big Data Analytics on a collection of such copious data sets can cultivate a proactive decision-making approach for predicting key opportunities and risks in Supply Chain. This paper discusses the various applications, benefits and the challenges of Big Data and Big Data Analytics in the Supply Chain.

Keywords
Big Data, Supply Chain, Big Data Analytics

1. Introduction

1.1 Big Data

According to Gartner (2012), Big Data is defined as:
"High-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation."
In addition to this, Big Data has been defined with the '3V's' concept by Laney (2001) :

- *Volume*: Until 2010, an estimated 13,000 Exabytes of data was generated wherein 1 Exabyte equals 1 Billion Gigabyte **(**Firican, 2017). The volume of the current Big Data datasets is a significant attribute as such datasets are considered out-of-scope in terms of prevailing traditional Database Management Techniques.
- *Velocity*: Velocity is the rate at which the data is being collected. With the increase in the accessibility of Internet over the globe and integration of machines and processes with Internet and centralized data collection techniques, the speed at which data is collected increases continuously.
- *Variety*: Along with the Volume and the rate of data generation, the type of data is also an important attribute in defining Big Data. Data collected can be structured or unstructured. Structured data is the systematic data collected from sources such as sales or financial transactions, reservation systems. Unstructured Data is the data generated from sources such as social media, emails, and communications **(**Technology**,** 2018**).**

In addition to the traditional Laney's definition, Big Data can be explained with two additional 'V's' with the 5V's concept. The additional two 'V's' being:

- *Veracity*: Veracity is often defined as the quality or trustworthiness of the collected data. Considering the accuracy of the collected data and analyzing it is important. Thus, when it comes to Big Data, quality is always preferred over quantity. To focus on quality, it is important to set metrics around the type of data that is collected and its sources.
- *Value*: Acquiring datasets of the Big Data scale involves substantial investment. Value of a dataset can be determined by estimating the insights that can be generated from the dataset post-analytics **(**XSNET, 2017**).**

1.2 Supply Chain

Supply Chain can be considered as a combination of four independent yet interlinked entities such as Marketing, Procurement, Warehouse Management and Transportation. Supply Chain Management is responsible for creating and maintaining the links of different entities in a business which are responsible for procurement of raw materials to ultimate end user delivery of the product (Halo, 2018). This paper focuses on the sources of generated data in Supply Chain, opportunities in Supply Chain from the analysis of collected data and the challenges in utilization of that data. In this review paper we will be discussing about the importance, potential opportunities and challenges of Big Data applications in Supply Chain and logistics.

1.3 Big Data Analytics

Big Data Analytics involves the use of advanced analytics techniques to extract valuable knowledge from vast amounts of data, facilitating data-driven decision-making. Big Data Analytics consists of three different levels of analytics. Each level of analytics has a different role and desired outcome. For this literature review, we consider the three levels of Big Data Analytics to be Prescriptive Analytics, Predictive Analytics and Descriptive Analytics. Currently, the level of consideration received by Prescriptive Analytics, followed by Predictive analytics with Descriptive Analytics receiving the least amount of consideration.

Prescriptive Analytics finds application data from processes such as Manufacturing, Logistics, Transportation and Warehousing along with newly introduced processes such as Cyber Physical Systems in the Industry 4.0 trend. Predictive analytics finds strong applications in Procurement, Risk assessment, Risk Management, Forecasting. Descriptive Analytics has the widest scope in terms of the number of processes in a system. Descriptive analytics finds application in development of effective and summarizing reports on raw data that is easy for human interpretation. The data used for descriptive analytics is mostly the Historical data (Nguyen, Zhou, Spiegler, Ieromonachou, & Lin, 2017).

2. Importance of Big Data in Supply Chain

With the technological advancements across the entities of Supply Chain, data generated is increasing at a fast rate. The information flow was documented in terms of physical documents until the use of Information Technology in Supply Chain. Now, majority of the information flow linked to the material flow is being documented in the form of digital structured data. As the scope of Supply Chain is currently worldwide, the volume of data collected from its numerous processes and the velocity at which it is being generated can be qualified as Big Data. In addition to this, entities such as marketing and sales are now relying on analysis of the unstructured data along with the structured data to gain better insights on customer needs and improve upon the cost aspects of Supply Chain processes.

The use of Big Data can offer a significant value in areas such as product development, market demand predictions, supplying decisions, distribution optimization and customer feedback. A summary of Big Data contribution in each business domain is exhibited in Table 1 (Ghosh, 2015).

Table 1. Percentage of Big Data contribution in business domains (Ghosh, 2015)

Name Of The Business Operations	Contribution To The Businesses (%)
Marketing	45
Operations	43
Sales	38
Risk Management	35
IT Analytics	33
Finance	32
Product Development	32
Customer Service	30
Logistics	22
HR	12
Other	12
Brand Management	8

3. Sources of Data in Supply Chain

A total of 52 mainstream sources of Big Data across the Supply Chain were considered for visualizing the classification of sources into Structured, Semi-Structured and Unstructured data in relation with the Velocity and Volume of data generation for each source, as shown in Figure 1.

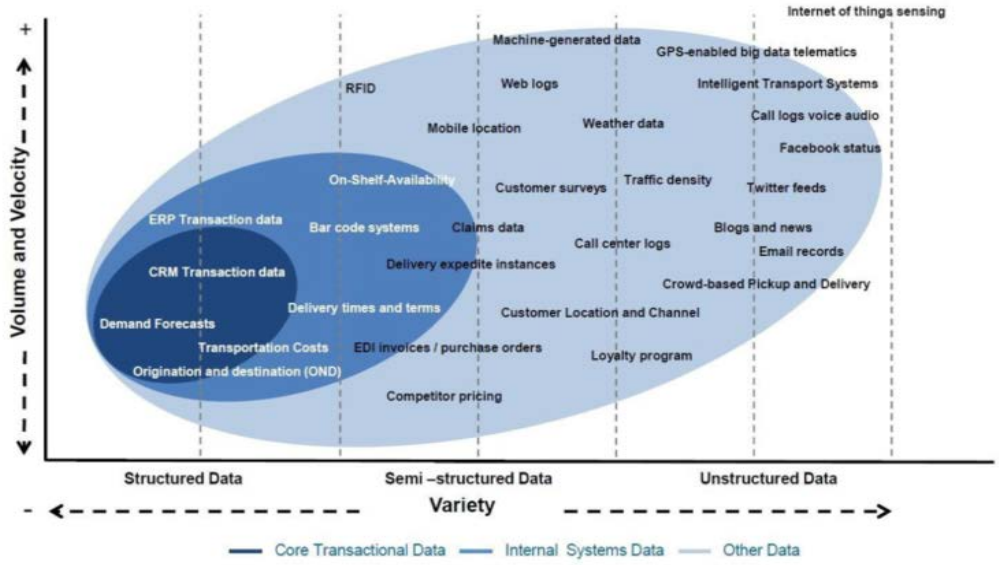

Figure 1. Classification of sources of data (Columbus, 2015)

According to a statistical observation about the growth of data from different sources, the data collected from Transactions, Social Media, Sensors and RFID scans or POS (Point of Sale) were recorded to be 88%, 43%,43%, 42% and 41% respectively, ranking them amongst the top 8 sources for data generation and analytics. The number of companies competing to utilize this data is increasing at a fast rate. The right utilization of this data is crucial to the transformation of the businesses across the globe, which has proved to be a paradigm for all the upcoming businesses. However, in some applications, such as maritime transportation and sea-based logistics, collection of data through RFID scans and sensors is very difficult due to the nature of the application or environment (Awwad & Pazour, 2018).

4. Potential of Big Data and Big Data Analytics

According to a study by Accenture (2014), companies with a disciplined strategy of utilizing Big Data Analytics have had bigger returns for their respective investments in Big Data Analytics as shown in Figure 2. As it is evident that a clear and systematic strategy towards Big Data Analytics can provide a good Return On Investment, the areas in Supply Chain such as Marketing, Procurement, Transportation, Warehousing can be tapped by Big Data and Big Data Analytics (Benabdellah, Benghabrit, Bouhaddou, & Zemmouri, 2016).

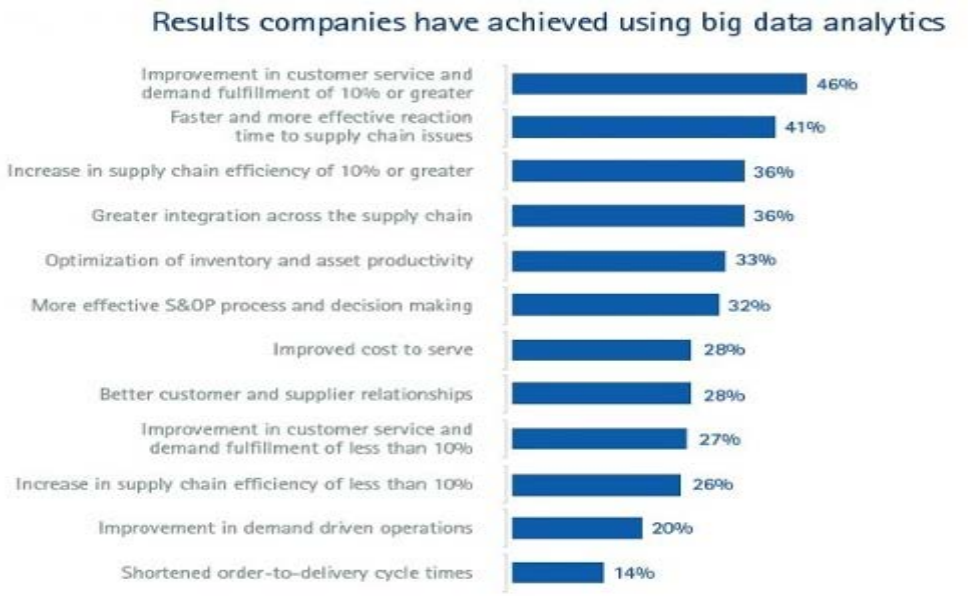

Figure 2. Results achieved using Big Data (Miguel & Gómez, 2016)

5. Opportunities for Big Data Analytics in Supply Chain

A significant motive of introducing Big Data Analytics in Supply Chain is to solve the prevailing problems that cannot be solved with traditional techniques. One of the significant challenges that Big Data and Big Data Analytics face in Supply Chain is the complexity of the process and the unstructured data evolved from it.
Supply Chain entities are interconnected by a significant physical flow that includes raw materials, work-in process inventories, finished products and returned items, information flows, and financial flow. Managing the increasing complexity in Supply Chains is necessary to companies to compete better in the global market. Complexity in Supply Chains is associated with material and information flows between the different Supply Chain entities. Traditionally, these flows are organized sequentially from supplier to customer. Today, information flows do not follow this linear form. Information flows rather now look like a simultaneous exchange, especially through electronic exchanges between all Supply Chain partners. A Supply Chain consists of many parts or elements of various types, which are linked to each other directly or indirectly. These various elements and their interrelationships are significant for complexity occurring in a system. The characteristics of complexity need to be considered for understanding its impact. The key characteristics of Supply Chain complexity are:

- *Number of Supply Chain entities*: All the different entities of a Supply Chain need to be considered. Computing data from numerous such entities can be difficult as the data collected increases with the increase in the number of considered entities.
- *Diversity*: A Supply Chain can be classified according to its homogeneity or heterogeneity.
- *Interdependency*: Interdependence between items, products and Supply Chain partners. Complexity increases in direct proportion to the increase of interdependence.
- *Variety*: Variety represents dynamical behavior of a system.
- *Uncertainty*: Uncertainty in a Supply Chain prevails due to the lack of knowledge about the whole system or any particular entity. Complexity of a Supply Chain increase with the increase in Uncertainty.

6. Applications of Big Data Analytics in Supply Chain

According to an article published on Computerworld, Prioritizing the development of a Big Data analytics strategy will help your organization overcome these Supply Chain challenges: By utilizing Big Data and Big Data Analytics, a Supply Chain should function with goals of Improving in areas such as prediction of customer needs, assessment of Supply Chain, efficiency of the overall Supply Chain, reaction time, Risk assessment (ComputerWorld, 2018).

- **Improving Predictions on Customer Needs**: Companies can lose their customers if it fails to fulfil the customer demands. Moreover, a company's reputation can be affected due to partial fulfillment or not fulfilling orders. In the age of the customer, offering the right product, to the right person at the right time and place is key to gaining (or retaining) customer satisfaction and loyalty. Smart organizations will leverage Big Data to get a full 360-degree view of your customer to better predict customer needs, understand personal preferences, and create a unique brand experience.
- **Improving Efficiency of Supply Chain**: Making use of Big Data analytics for appropriate, a Cost efficiency, cost reduction, and spend analytics will continue as top business priorities in Supply Chain management.
- **Improving Risk Assessment in Supply Chain:** Predictive Analytics is a major part of Big Data Analytics. With the help of Predictive Analytics can help assess the probabilities of occurrence of a problem, its potential impact. Predictive Analytics in Big Data can help in identification of Supply Chain risks by analyzing large volumes of historical data and risk mapping techniques. Appropriate predictions of the risks can help develop tools and techniques so as to minimize the effect of the potential risk.
- **Improving Traceability in Supply Chain:** Improved Traceability ensures better tracking of goods from production to retail. Improved traceability can help in integrating the different supply chain entities and maintain a better flow of goods. Better tracking capabilities give better control over the supply chain processes.
- **Improving the Reaction Time:** Ninety percent of companies say that agility and speed are important or very important to their business. The ability to quickly and flexibly meet customer fulfillment objectives is rated the second most important driver of competitive advantage across all industries. Embedding Big Data analytics in operations can have an impact on organizations' reaction time to Supply Chain issues (41 percent) and can lead to a 4.25x improvement in order-to-cycle delivery times, according to Accenture.

Along with the attribute specific opportunities and applications for Big Data Analytics, there are some process specific applications of Big Data Analytics which are shown in Table 2. In their paper, Benabdellah et al. (2016) discuss probable applications of Big Data Analytics in the processes such as Plantification, Supplying, Production, Distribution and Return.

Table 2. Applications Of Big Data Analytics

Serial Number	Process	Application
1.	Planification	i) Risk evaluation and resilience planning ii) Reduce the risk of infrastructure investments and contracted external capacities iii) Enabling the monitoring of performance, as well as improving planning and management functions.
2.	Supplying	i) Reduce storage capacity and distribution ii) Enabling more supplier networks that focus on knowledge collaboration as the value- add over just completing transactions. iii) Achieve granular levels on aggregated procurement patterns
3	Production	i) Market intelligence for small and medium-sized enterprises ii) The largest clusters of data are related to an automated sensing capability, connectivity and intelligence to material handling and packaging systems applications evolved iii) Getting back a real time capacity availability and providing a quicker response and vendor managed inventory
4	Distribution	i) Optimal routing ii) Real-time route optimization; address verification; crowd-based pick-up and delivery; environmental intelligence iii) Improve Supply Chain traceability iv) Real-time optimization of delivery routes v) Estimated lead times based on traffic conditions, weather vi) variables, real time marginal cost for different channels

		vi) Optimize logistics activities thanks to costs reduction, improved customer satisfaction and supply chain performance vii) Optimize manufacturing processes, shop-floor management and manufacturing logistics viii) Reduce lead times and minimize costs and delays, as well as process interruptions
5	Return	i) Reduction in driver turnover, driver assignment, using sentiment data analysis ii) Customer loyalty management; Continuous service improvement and product innovation iii) Benefits for the government (e.g., urban planning) and companies (e.g., localized advertising, optimized routing) iv) Creating an integrated view of customer interactions and operational performance, ensuring satisfaction of both sender and recipient v) Technology has made it more feasible than ever to access and understand customer data, as Big Data enables sensing of social behavior vi) Access and understand customer data, as Big Data enables sensing of social behavior vii) Know customers' perceptions of offered products and services and discover their unobservable characteristics

7. Challenges in Adopting Big Data Analytics for Supply Chain

The issues and challenges in adopting Big Data Analytics for Supply Chain can be broadly categorized into two categories such as Organizational Challenges and Technical Challenges **(**Arunachalam, Kumar, & Kawalek, 2017**).**

The issues and challenges under Organizational challenges are listed as:

- Time-consuming: Factors such as the volume of Big Data, Complexity of Supply Chain and interpretation goals for the datasets along with external factors such as lack of access to data contribute in making the analytics process time-consuming.
- Insufficient resources: For better results, the availability of real-time data is crucial. Supply Chain being a platform that generates complex cross-functional data for interlinked entities, collection and storage of cross-functional data should be streamlined.
- Privacy and security concerns: Data sharing across a Supply Chain Network is a major factor in collecting data from various sources, analyzing it and giving insights. Although, regional or global Supply Chain Networks might face difficulties in sharing data across its different sources due to various Privacy, Security laws concerned with sharing of data. Lack of shared data in such cases can affect the accuracy of the insights that Big Data Analytics might generate.
- Behavioral issues: Supply chain risk and cost of inventory can be an elevated risk if the decision makers react to insignificant changes in the physical world which would worsen the ''bullwhip effect". Due to the variety and volume of Big Data, there is an increased risk of decision makers identifying irrelevant correlations but statistically significant relations with insignificant causal linkage.
- Issues with Return on Investment (ROI): Volume and variety of the Big Data collected makes it difficult to estimate the value of the collected data. Performing analytics on Big Data requires a significant amount of investment for building the infrastructure. Due to the uncertainty on value of the data, there is an increased risk on the returns that the investment on infrastructure might produce.
- Lack of skills: The complexity of Big Data generated from Supply Chain source requires a combination of good domain knowledge analytics skills and the ability to interpret the usability of data. According to surveys, such a combination along with experience is difficult to find.

Technical challenges are listed as:

- Data scalability: Issue of Data Scalability is considered a major technical issue in the process of utilizing Big Data Analytics in any system. Inability of organizations to shift from traditional limited databases to

distributed databases or cloud storage databases affects the insights from Big Data Analytics as the amount of relative data is compromised.

- Data Quality: Quality of the stored and utilized data can affect the performance the results of the analytics techniques. Data being intangible and multidimensional based on its sources and applications. Dimensions of multidimensional dataset can be classified as intrinsic and contextual. For consistent and reliable results for decision making purposes, the quality of data should be consistent. The variety of data and type of sources for data in supply chain may affect the quality of the collected data.
- Lack of techniques: Incapability of a firm to utilize the data affects the robustness of the insights developed after analyzing the datasets. The techniques used to analyze, compute, forecast and visualize need to be altered or upgraded in accordance to the complexity or volume of data (Arunachalam et al., 2017).

8. Conclusions and Future Work

In this paper, the concept of Big Data and Big Data Analytics in Supply Chain is reviewed. The scale of Big Data is considered as the main reason for adopting it with Supply Chain. After studying the sources of Big Data generation in Supply Chain processes and activities, valuable insights regarding the potential of Big Data Analytics were uncovered. It was observed that combination of the complex data from supply chain activities and the scope of Big Data in terms of Volume, Variety, Velocity, Veracity and Value have practical applications that can solve some of the most prevailing challenges faced by supply chain even the recent years. Considering the adoption of Big Data Analytics, a relatively new phenomenon, it was found that the pace of creating infrastructure to sustain the increasing data needs to increase. It was found that the unavailability of professionals with appropriate skillsets can hinder the potential of Big Data Analytics in Supply Chain. As the complexity of the Supply Chain Networks around the globe increases, the Supply Chain industry along with the Data Analytics industry should work on developing new and effective models and techniques.

Given the high infrastructure costs for Big Data Analytics, a dedicated research on making Big Data Analytics more cost effective is possible by reducing the infrastructure costs for storing Big Data. To increase the volume and accuracy of the data generated from various processes such as manufacturing and logistics, improving the sensor accuracy in physical systems along with enhancements in the data integration technology amongst various business processes is necessary and can be a potential field of study for further research.

References

Arunachalam, D., Kumar, N., & Kawalek, J. P. (2017). *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*. doi:https://doi.org/10.1016/j.tre.2017.04.001

Awwad, M., & Pazour, J. (2018). Search Plan for a Single Item in an Inverted T k-Deep Storage System. *Military Operations Research, 23*(1), 5-22.

Benabdellah, A. C., Benghabrit, A., Bouhaddou, I., & Zemmouri, E. M. (2016, Nov. 29 2016-Dec. 2 2016). *Big data for supply chain management: Opportunities and challenges.* Paper presented at the 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA).

Columbus, L. (2015). *Ten Ways Big Data Is Revolutionizing Supply Chain Management*.

ComputerWorld. (2018). *Overcoming 5 Major Supply Chain Challenges with Big Data Analytics*.

Firican, G. (2017). *The 10 Vs of Big Data*.

Gartner. (2012). What Is Big Data?

Ghosh, D. (2015, 14-15 Sept. 2015). *Big Data in Logistics and Supply Chain management - a rethinking step*. Paper presented at the 2015 International Symposium on Advanced Computing and Communication (ISACC).

Halo. (2018). Descriptive, Predictive, and Prescriptive Analytics Explained.

Laney, D. (2001). *3 D Data Management: Controlling Data Volume, Velocity and Variety*.

Miguel, J., & Gómez, F. (2016). *Top Challenges for Big Data in the Supply Chain Management Process*.

Nguyen, T., Zhou, L., Spiegler, V., Ieromonachou, P., & Lin, Y. (2017). Big data analytics in supply chain management: A state-of-the-art literature review. *Computers & Operations Research*. doi:https://doi.org/10.1016/j.cor.2017.07.004

Technology, N. J. I. o. (2018). *Why Data is the Next Frontier of Innovation*.

XSNET. (2017). Updated for 2017: The V's of Big Data: Velocity, Volume, Value, Variety, and Veracity.

Biographies

Pranav Kulkarni is currently pursuing his Master of Science in Industrial Engineering degree from the University at Buffalo, The State University of New York, Buffalo, NY, USA. He has earned his Bachelor's of Engineering degree in Electronics and Telecommunication Engineering from the Savitribai Phule Pune University, India (formerly known as the University of Pune). His research interests include supply chain engineering, data analytics, data visualization, production planning, inventory control and lean manufacturing.

Rachit Bapna is currently pursuing his Master of Science degree in Industrial Engineering from the University at Buffalo, The State University of New York, Buffalo, NY, USA. Mr. Bapna holds a Bachelor's of Technology degree in Automobile Engineering from SRM University, India. He has spent several internship periods in big OEMs like Maruti Suzuki India Ltd., Honda Cars India and L&T Ltd. He has held the role of team lead at SAE Collegiate Club which is affiliated to SAE international. His keen research interests include supply chain engineering, advanced manufacturing, data analytics and lean manufacturing.

Mohamed Awwad is a Teaching Assistant Professor in the Department of Industrial and Systems Engineering at the University at Buffalo, The State University of New York, Buffalo, NY, USA. He received his Ph.D. and M.S. degrees in Industrial Engineering from the University of Central Florida, Orlando, FL, USA. Additionally, he holds M.S. and B.S. degrees in Mechanical Engineering from Cairo University, Egypt. Before his tenure with the University at Buffalo, Dr. Awwad held several teaching and research positions at the University of Missouri, Florida Polytechnic University, and the University of Central Florida. His research interests include applying operations research methods in the fields of logistics & supply chain, distribution center design, unconventional logistics systems design, healthcare and the military. He is a member of IISE, INFORMS and ASME.