HELLENIC REPUBLIC

UNIVERSITY OF MACEDONIA

Interdepartmental Program of Postgraduate Studied in Information Systems

Master Thesis

BIG DATA ANALYTICS IN SYPPLY CHAIN MANAGEMENT

Student name

THEODOROS TZIOUTZIAS

Submitted as required for the Master's Degree in Information Systems

January 2019

Abstract

In this research we study the role that Big Data Analytics play nowadays in the Supply Chain Management through the last decades. We are pointing out how the decision-making processes in logistics and the supply chain management is affected after the implementation of the Big Data analytics, the importance of the data quality, and the importance of the integrated information systems. In addition, we present how data generated, evolved and increased, coming from various sources in very big quantities, and also the importance of the Internet of Thing in todays' logistics.

We give a general view of Big Data, then we also expand to the supply chain procedures and then we combine that two very important terms. Moreover we present applications of Big Data Analytics in the Supply Chain Management from very important players in the market and finally we conclude.

Key words: Supply chain management, SCM, Big data, BD, big data analytics, BDA

Inscription

“To my beloved Georgia for her support and patience…”

Acknowledgements

"I would like to thanks all the people that help and support me to this MSc and especially my parents…"

Contexts

Figures Contexts

Tables Contexts

1. Introduction

Nowadays the term 'Big Data' (BD) is industry's hottest trend. BD refers to datasets that consists of large and complex data that traditional methods are unable to process. From many consultants and managers it is claimed that talking about BD evolution, we are talking about the 4th industrial revolution. BD is adding new roles in the industry and decision making processes. According to the Market Reports Center the investment on Big Data will account for over 65$ in 2018. Although there are many thought from practitioners that BD is an extension of forecasting or a traditional market research (Richey Jr et al., (2016)).

The science of Supply chain and Logistics is a dynamic, living and regularly changing science. In the early 1960's, logistics were defined as the physical distribution channel, while today logistics and Supply Chain Management (SCM) considered as the function that encompasses the entire organization (Ittmann, (2015)), focusing on the strategies that will led to a cost effective, reliable and predictable service.

Over the last years, firms realize the necessity to move towards in each sector of the economy with better accuracy. In order to achieve that data-driven insight considered to be the key to achieve the best possible decision making. Information Technology systems are developing and evolving rapidly becoming also more sophisticated. In addition, more and more powerful computers are being developing with great process power and with accordance to the abundance of the personal smart devices, the world is moving straight towards the 'Big Data' era. For example, according to (Zhong et al.,(2015, pp. 260-272)), during a flight with an Airbus A380, each engine generates 10 TB of data every 30 minutes, more than 12 TB of data are created form Twitter every day while through Facebook more than 25 TB of data are generating every day. Data are becoming more powerful helping companies to succeed, but in order to do that appropriate and proper analysis of the data through analytics should be done. The findings after that data process, play very important role and has already widespread in many different sectors, including SCM and logistics.

Companies have to adapt to that new business models and reconsider their position and their role in their value chain regarding the potential possibilities given by the utilization of BD to add value for their suppliers and their customers. There are also requirements concerning both the organization of the company and the information technology infrastructure.

It is expected that BD can improve in a high level the SC collaboration and SC processes like inventory management, transportation management and relationship management. Recent researched in BD projects in the field of SCM focus on tasks like faster tracking and classification of goods, gathering data for logistics and transportation scheduling and planning, and data analytics for "health checking" of creditors, keeping track of partners and suppliers compliance conditions using web mining.

Big Data can play very important role for the forecasting of delivery times and route optimization by considering different kind of data like traffic information, weather information or even driver characteristics. Other fields of application for the BD is the analysis of information received from sensors in order to achieve better warehouse management and more effective decision making. The always increased utilities of GPS systems in co-ordination with the further evolvement of sensor networks and the 'Internet of Things' (IoT) increase the capabilities of BD for the SCM. Data generated from so many personal and smart devices and logistics aiding equipment can led to the optimization of SCM processes, achieving better customer service and lower operational cost.

The tools for the data analysis that will provide companies with better mean to obtain value from the massive amount of data that produced and stored and get competitive advantage is the Big Data Business Analytics (BDBA). BDBA consists of two dimension, the dimensions of the BD and the dimension of the business analytics (BA). BD stands for the ability to process the data with the so called 5v's of BD, which are velocity, variety, volume, verification and value. We will analyze more these qualities in our research and we will expand them. Analytics refer to the technics like statistics, mathematics, simulation optimization, econometrics etc. and the abilities used in order to find the desirable findings, results and insights. By this way, analytics help companies to the decision making.

The aim of this research is to highlight how the area of SCM is impacted from the science of Big Data Analytics (BDA). We will also refer the various trends and concepts around BD and analytics and how big companies have already increase their competitiveness getting benefit from these new trends and opportunities of BDA.

2. Big Data Analytics

2.1. Data Definition

Before we get into the explanation of the BDA, we would like to answer to some simple question that will help us have a better understanding of the definition and our research in general. We begin with the explanation of the term data and what we call data.

According to Oxford English Dictionary, the term data is defined as *"facts and statistics collected together for reference or analysis"*. Another definition says that data is *"the quantities, characters, or symbols on which operations are performed by a computer, which may be stored and transmitted in the form of electrical signals and recorder on magnetic, optical, or mechanical recording media."* (Oxford Dictionary, 2018). Analyzing the first definition, according to (Arunachalam et al., 2017), data are facts and numbers collected for further analysis. The data process and analysis of these data can be made using traditional statistical and mathematical methods and tools without the need of a computer. As far as the second definition is concerned, it includes 'characters' and 'symbols' emphasizing the usage of the need of a computer in order to perform analysis. The second definition is also referring in the way that the data are transferred and stored in various forms.

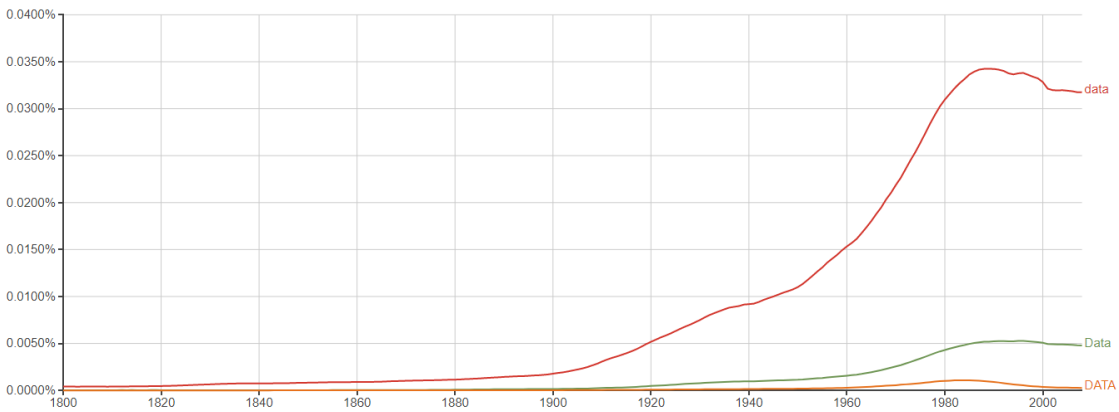

figure 1 Graphing the word 'data' in Google Ngram Viewer source: Google Books Ngram Viewer

The above diagram show us how the reputation of the terms "data", "Data" and "DATA" evolve through the years from 1800 till 2008.

2.2. Big Data Definition

The recent years, data come in many different formats from various sources. Most of the supply chain managers are accustomed to the structured data formats (operational number, accounting etc.). This qualitative data are translated into numeric formats easy to be processed from traditional databases. This operational data can be within the company or they can also be shared with integrated partners.

On the other hand, unstructured data, or data that is not pre-defined to be analyzed in a software, is the most challenging and less used data source. Examples of unstructured data is the e-mails, audio and video, social media posts, tweets and text documents.

In (Richey Jr et al., (2016)), it is stated that BD is data that you do not know what to do with. Nowadays, we get so big volume of data that it becomes so complex, and the key for success is to be able to know what to do with it. For example, you use complex data in many places in the SC, customer relationship management software or Procurement, everywhere. Each and every department within the business, they do not have pieces or sets of systems to make use of this data. The key for the success, is to know how much of this data you need really and there is just a “slew” of information. Sometimes, part of this data may just be useless and the key is to know what to do with it and how much of this information to use, and also how much capital to invest in IT so as to turn all of this into information.

According to Zhong et al.,(2016, pp. 572-591) the term Big Data is referring to the huge flood of data in the range of Exabyte and beyond. BD has extended the scope of technological capability to store, manage, process, interpret and visualize the amount of data.

It was in October of 1997 when the concept of BD was firstly exposed from the ACM digital library. It was used to define a visualization challenge for systems with large datasets in computer science. Since then the attention from the academia and the practitioners was great. It is estimated that BD will be one of the top 10 ambition markets for the coming century. According to the statistics, it is estimated that by 2022 the total global BD market will reach $118.52 growing at a compound annual growth rate of 26% from 2014 to 2020 (NewsOn6.com) (Zhong et al.,(2016, pp. 572-591)).

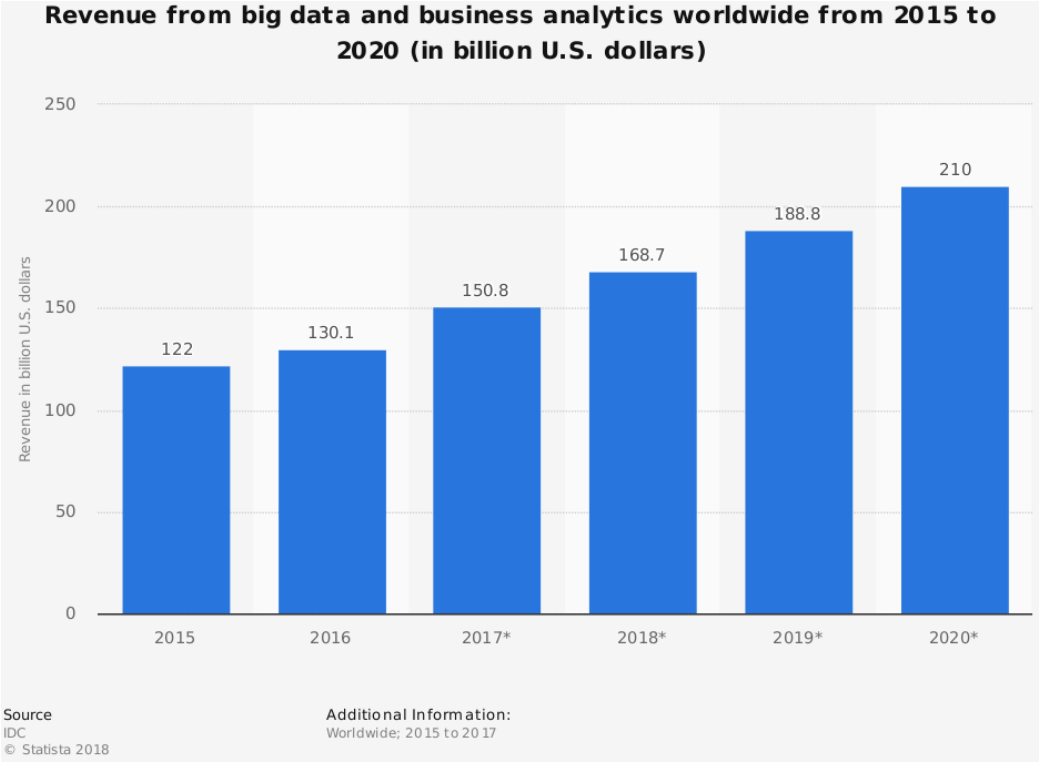

figure 2 source IDC Statista 2018

A sector that has been undertaking digitization for many decades is this of Service and Manufacturing Supply Chain Management (SM-SCM). This sector is evolving rapidly and the human activity is mainly involved in the needed procedures also to important fields like aeronautical to daily necessities. The necessity to improve the performance and the efficiency has led to the initialization of BD. The combination of SM-SCM and BD can led to a better decision making mechanisms when using natural resources. In order to achieve that, big investments have been done in order to encourage scientists, researchers and practitioners to study for Big Data's application, development and have a better understanding.

For many year, SM-SCM was focusing only on the collection and the storage of enormous size of data (Zhong et al.,(2016, pp. 572-591)). However, it was great challenge to make full use of such big size of data. As it is done in (Zhong et al.,(2016, pp. 572-591)), we will summarize the challenges of BD to the '5V's', three of them are main and the other two are additional, and we will give the typical characteristics of SM-SCM.

- ***Volume:*** The amount of data produces within the SC from all over the world in really enormous. A personal care manufacturer for example generates 500 data samples every 33ms, resulting in 152.000 samples per second, 9.000.000 per

minute, resulting in 4 trillion samples per year. The collection of so much data creates data floods to the data collectors and the storage facilities.

- ***Velocity:*** The velocity to process such big sizes of data produced in the SC is very important because data-driven decisions should be made as quick as possible. With the term velocity we are referring to the speed by which the data are collected, the reliability of the processes of data transferring, the efficient data storage, the time that is needed to discover and reach useful outcomes or data, as well as the fast decision making models and algorithms.
- ***Variety:*** Data in the SC are coming from various sources and in heterogeneous formats. New types of data are generating from sensors used in manufacturing, retail shops, highways and facilitated houses. For these data to be useful and able to be processed it is needed the use of a makeup language to change them into standard formats.
- ***Verification:*** All data collected for sure are not useful. There is always a part of a bad data produced in the SC procedures. From all data produced we are called to pick up only the good data. The verification and the selection of the good data has to be made under certain security points and authorities. Despite that the verification process is developed and designed as tool to automatically confirm the good quality of the data, some controls are complex enough to be addressed and they are constitute a challenge.
- ***Value:*** The value of the BD is difficult to be evaluated in the SC. Firstly, the extraction of value through the BD is a tough enough task because of the obstacles caused by volume, velocity, variety and verification. Secondly, it is hard enough to examine the impacts on the insights, the benefits and the business processes within the SC. Thirdly, the value of the reports extracted, the statistic results and the decisions obtained from BD is hard to be measured because of the big influence in micro and macro perspectives.

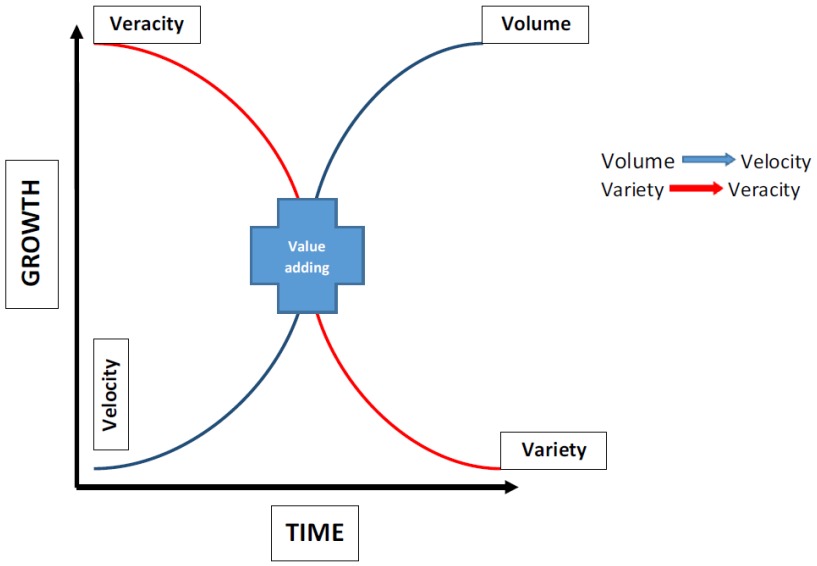

figure 3 Modulation of the five Vs of big data – an innovation S-curve model.
source: R. Addo-Tenkorang, P.T. Helo / Computers & Industrial Engineering 101 (2016) 528–543

2.3. Big Data Characteristics and Data Structures

In the previous section we analyze the V-model for the BD with its five characteristic V-features: volume, velocity, variety, verification/veracity and value have been considered. Based on these characteristics we will give the following definition for the BD, being considered as:

'Data whose scale, distribution, diversity, and timeliness require the use of new technical architecture and analytics to enable insights that unlock new sources of business value.' (Franczyk, 2014)

The data streams in logistics and SCM may be varying between structured, semi-structured and unstructured data. The structured data contains defined data formats, types and structures. An example of that kind of data could be the transaction data sourcing from Online Transaction Processing (OLTP) and Online Analytical Processing (OLAP).

The semi-structured data is characterized by textual data filed with discernable patterns, what enables data parsing. Examples of semi-structured data are common in self-describing XML data files defined by XML Schema.

The unstructured data are irregular data formats that is possible to be formatted, but only after the use of additional tools, effort and time. An example of such kind of data could be Wed click-stream data.

As unstructured considered the data that have no any inherent structure and can be stored in different types of files. Duo to the fact that part of data is unstructured, or in other

words, it lacks a structure appropriate for storage in conventional SQL databases, implying that other solutions are needed.

The conventional IT infrastructures in SCs include structured data in form of OLTP and OLAP systems. The common known traditional OLTP systems support the transactional systems with highly structured SQL databases, whereas the OLAP systems contain aggregated historical data in form of cubes. The role of OLTP systems is to deliver simple reports, while OLAP systems (Data warehouses) are suited for traditional Business Intelligence (BI) applications with reporting facilities on business statistics, performance, etc. on the basis of structured analytical historical data. On the opposite of partially low value BD the OLAP and OLTP provide only high quality data.

For the data scientists in the sector of logistics, the most important and critical activities are to provide services to other stakeholders, such as data engineers, data analysts, business analysts and the users in a line of business. Such activities are the reframing of business challenges as analytics challenges, design, implementation and deployment of statistical models and also the data mining techniques, especially for BD. A crucial aspect in data science is thereby included in creating insights leading to actionable recommendations to help business to gain a competitive advantage. (La Ponsie, 2011)

For SC design and management an advantageous skill set in the data science includes skills, where each may have different weight. Some examples of disciplines that should be covered are the following: forecasting, optimization and discrete event simulation, applied probability, statistics, analytical mathematical modeling, marketing, finance, economics, and accounting. For all the above mentioned disciplines, duo to BD utilization, the focus will be different than in a traditional approach. Skills with always increasing importance are these in conjunction with broad awareness of many different methods in comparison to the classical point of view. For instance, for the aims of forecasting the understanding, the application of qualitative and quantitative methods will become more important than the understanding of the underlying stochastic processes.

Predictive analysis is a subset of data science. It includes a variety of statistical techniques that enable to get benefit from the patterns found in transactional and historical data. In logistics it could held to optimize business operations, to identify business risks and security aspects, to predict new business opportunities and to fulfill the law of regulatory requirements. The business value of predictive data analysis-data science and data mining will be higher than gained from conventional Business Intelligence (BI) duo to

optimization, predictive modeling, forecasting and analysis of vast data resources like BD.

In logistics industry both qualitative and quantitative methods are being used by the predictive analysis in order to estimate the past and future behavior of the flow and storage of inventory, as well as the associated costs and service levels. On the other hand, the SCM predictive analysis also uses both quantitative and qualitative methods to improve SC design and competitiveness by estimating past and future levels of integration of business processes among functions of organization, as well as associated costs and service levels. The value of predictive analysis is not to be scoffed at. In collaboration with the appropriate analytics tools it may become a critical competitive asset.

2.4. Information Tsunami

In (Tan et. al., 2015), it is described that BD have an voluminous size of data, where the development of data is extremely fast and the scale of data is too varied, so that the conventional data management systems cannot process that data efficiently. During the year 2000, according to IBM, the total size of data around the world was estimated to be around 800.000 petabytes (PB). It is expected this amount to reach the 35 zettabytes (ZB) by 2020. It is very difficult for the conventional and traditional systems to store and analyze that explosion of data.

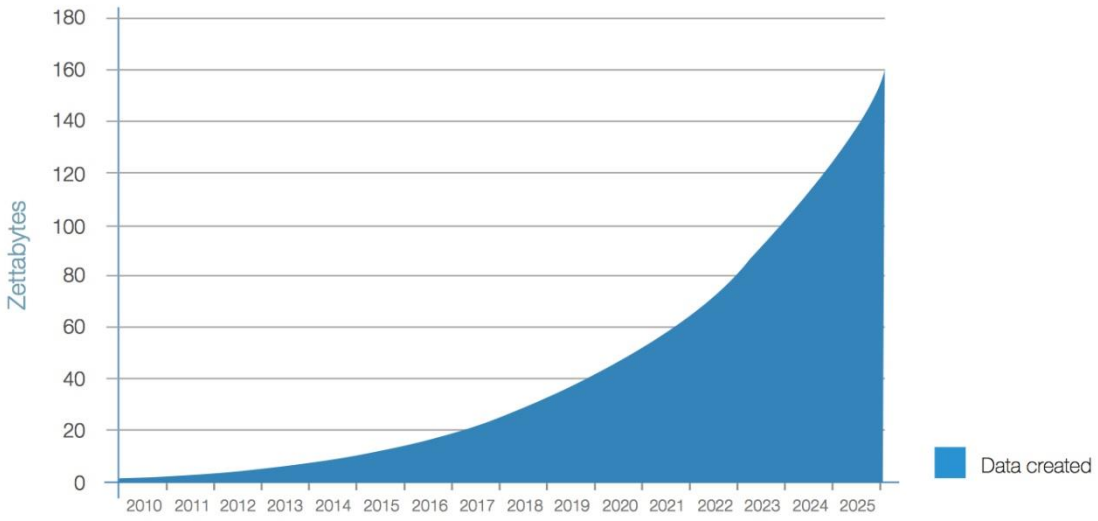

figure 4 Estimation of data volume till 2025. Source: https://www.statista.com

In addition, there are many new sources that provide us with different or even new types of data such as social media, texts, weblogs, global positioning systems and location information systems, data coming from sensors, graphs, videos and audio and other online

data. For the handling and the storage of this data new technologies and advanced equipment is needed. Moreover, the introduction of data from different sources made them more complex, introducing except from the well-known structured data, also semi-structured and unstructured data sourcing from the web mails, search indexes, social media, forums, sensor data etc. The big challenge is that the traditional analytic tools cannot deal with other than structured data (Tan et. al., 2015). It is estimated that the 80% of the available data is semi-structured or unstructured and it is very difficult to analyze it. However, in the last decade we live, the science of data is always developing. In order to succeed, the key is to be able to draw insights both from traditional and non-traditional data analysis methods. This ability to analyze all types of data is creating more value for the company and a general competitive advantage.

2.5. Data Collection

The process of data collection plays very important role for the future because without an effective and efficient approach for data collection, it is impossible to carry out decision making and data based analytics. Auto-ID technologies, smart sensors, RFID, internet based social networks and systems, digital and smart devices are some examples of data collection methods, but though there are many challenges to be overcome in the service and manufacturing industry field. Traditional methods like the use of paper and manual based data collection methods are still been used and they are very common in the service manufacturing sector and SCM (Zhong et al.,(2016, pp. 572-591)). These methods according to Zhong et al.,(2016, pp. 572-591) are still been used in developing countries such as Pearl River Delta in China. The data collected using these methods are used to incomplete, inaccurate and untimely as a result decision making based on this data to be usually unreasonable, unexecutable and unpresentable. On the other hand, diverse data collectors like digital devices and sensors, have specific data formats which are common heterogeneous, incompatible in other devices and unstructured. The technological development and the integration in the sector of data makes things more difficult in some situations. In Zhong et al.,(2016, pp. 572-591) it is presented as an example of two companies with activity in commercial service, when they try to merge their transactions and it was challengeable duo to the unstandarized data. What is more, in cases when huge amount of data needs to be collected simultaneously, the system may face problems duo to signal collisions and limited central processor’s capacity.

In order to overcome all the above problems, opportunities could be observed. Big steps have been done around data collection technologies such as Internet of Things (IoT)

enabled devices, smart Auto-ID, advanced data collectors that could help companies to collect and capture the data they need. These sensors-data collectors, may use biological recognition technologies to recognize the users, voice controlled system to facilitate the data collection operations, as well as adaptive mechanisms and machine learning that will allow the devices to operate and be used smarter and easier under various conditions. As far as the logistics sector is concerned, mobile smart data collectors are more suitable. Thus, IoT technologies can be applied using smart devices like cell-phoned or smart phones and other smart devices that could be used for collecting data. For example, if a box or a carton is equipped with Radio Frequency Identification (RFID) tag and a temperature sensor is placed into a smart container with active RFID reader, each box could be identified separately, and then if this container is being transported by a smart vehicle that has IoT enabled devices and Global Positioning System (GPS). A trend of the last years is also the wearable devices that use high intelligence and may be used in the manufacturing service and the SCM as data collectors to achieve better efficiency (Zhong et al.,(2016, pp. 572-591)).

The data standardization is very important and crucial, thus, those standards will be used from companies in various industries. That helps to open up an opportunity for standard parties, data warehouses, and operating system vendors to underline and highlight the importance and the chance to create various sets of standard data models, enabling the data transferring and information sharing. Industries around insurance market and banking, as well as pharmaceutical and automotive industries are more ideal for the application of standard models to customer information gathering. Models that would be able to collect data at the same time using the data standards and being able to deal with enormous datasets instantly, will be a very hot case study. In co-operation with advanced hardware design and software algorithms, parallel data gathering approach hopes to handle almost a terabyte per second. That give chances to the IT companies to study for new mechanisms or hardware devices that will be able to deal with data gathering-capturing very fast and reliable in the near future.

As it is stated in (Zhong et al.,(2016, pp. 572-591)) , smart devices will play critical role in the future for the data collection in the service and manufacturing sector and SCM sector. Some perspectives for the smart data collectors are the following:

- ***Multi-functional ubiquitous devices:*** That kind of devices are equipped with sensors or are high technological intelligent combining cloud computing, smart detection, auto-ID technology, temperature smart sensors, humidity, light sensors

for achieving multi-functions. They are designed for the delivery for example of light or temperature sensitive items and other special services (Zhong et al.,(2016, pp. 572-591)).

- ***Smart robotic collectors:*** These collectors are designed to operate under extreme conditions such as very low or very high temperatures, harmful environments like nuclear reactors etc. That kind of data collectors use very advanced technology that enable them to operate like a human that is able to capture data and sense everything like watching, smelling, listening, touching and tasting (Zhong et al.,(2016, pp. 572-591)).
- ***Intelligent adaptive devices:*** These devices are the most intelligent for data gathering using artificial intelligence and machine learning technologies. They have learnable skills to build up a knowledge base that will let them and guide them to operate autonomous for the collection of the required data. In the public transportation for example, that kind of technologies are used for easing the disabled people by providing them proper information like guiding the routine, carrying them upstairs or downstairs. Also these devices could gather information like who use a ticket card or smart ID cards like the location of the check point of the destination and the time (Zhong et al.,(2016, pp. 572-591))

2.6. Data Sources

Data sources are mainly divided into the *internal and the external data sources. Internal data* is mostly available in databases and business IT-systems, such as ERP system. The internal communication between the production and warehouse systems is also available as data streams, like radio-frequency identification (RFID) devices.

External data sources are also available as data streams, for example Datasets from data portals and Social Media like Instagram, twitter, Facebook etc. Social Media sources provide us with unstructured data and the data semantic is frequently changing and it's various. Most of the Social Media data is not directly accessible. Twitter for example, provides only a data stream that contains a specific number of tweets. Thus, organizations like DataSift can offer purchasable streams of data coming from social media (Ghemawat et. al. 2003). On the other hand, external sources like search engine trends are also less structured, but the provided APIs, like Google Search API, are mostly free to use. Some external data sources can be searched and captured using technologies like Web Sematic

(WS) as shown in (Berners-Lee, et al (2001)) ; (Rajapaksha, et al., 2008)); Pfisterer, et al., (2009)] or Linked Open Data (LOD) as described in (Smart Service Welt (2014)) .

Another example of Open Data sources, is Eurostat (epp.eurostat.ec.europa.eu). Eurostat provide us data free either for commercial or non-commercial use. For example we can find n political information about regions and countries, geological and geophysical data and other statistics. These sources are accessible through open data portals like the European Union Open Data Portal that represents a catalogue for further open data sources (Supply chain Council 2010). Next to the open data portals, there are also many other platforms that provide closed data. Closed data means that we need to purchase the datasets or to be licensed for access. Example platforms of data are:

Microsoft Azure Data Market is a cloud computing infrastructure. We can find both close and open data that are available through this data market platform. The datasets could be provided by many different organizations and companies.

Another open data platform that was founded in 2009.is Factual. Factual offers different data services like ad targeting or data mapping. All services are based on their global location database containing data for more than 65 million local businesses and places of interest spread all around the world.

A service platform focusing on Business Data management like contacts and company profiles is Data.com. The service offers access to millions of company profiles and allows current updates on the customer data.

2.7. Dimensions of data quality

Research suggests that data quality is comprised of several dimensions (Ballou and Pazer, 1985; Ballou et al., 1998; Pipino et al., 2002; Redman, 1996; Wand and Wang, 1996; Wang and Strong, 1996). Both Wang and Strong (1996) and Lee et al. (2002) organize data quality dimensions into two categories: the first one is intrinsic, that refers to attributes that are objective and native to the data and the second one is contextual, that refers to the attributes that are dependent on the context in which the data are used or observed. Contextual dimensions include quantity, value-added, relevancy (Wang and Strong, 1996), accessibility, believability, and reputation of the data (Lee et al., 2004, 2002). Measures of these dimensions have relied heavily on self-report surveys and user

questionnaires, as they rely on subjective and situational judgments of decision makers for quantification (Batini et al., 2009). Contextual dimensions of data quality lend themselves more towards information as opposed to data, because these dimensions are formed by placing data within a situation or problem specific context (Batini et al., 2009; Davenport and Prusak, 2000; Haug et al., 2009; Watts et al., 2009). Because we consider the quality of data, not information, as it moves through a production-like process, we limit our discussion of quality to consideration of the intrinsic measures of data quality.

The literature consistently describes intrinsic data quality along four dimensions: accuracy, timeliness, consistency and completeness (Ballou and Pazer, 1985; Batini e tal., 2009; Blake and Mangiameli, 2011; Haug and Arlbjørn,2011; Haug et al., 2009; Kahn et al., 2002; Lee et al., 2002; Parssian, 2006; Scannapieco and Catarci, 2002; Wang and Strong, 1996; Zeithaml et al., 1990). Below, we explore and use the aforementioned literature to define and describe these four dimensions.

Accuracy refers to the level to which data are equivalent to their corresponding 'real' values (Ballou and Pazer, 1985). This dimension can be assessed by comparing values with external values that are known to be or considered to be correct (Redman, 1996). A simple example would be a data record in a customer relationship management system, where the street address for a customer in the system matches the street address where the customer currently resides. In this case, accuracy of the street address value in the system could be assessed via validating the shipping address on the most recent customer order. No problem context or value-judgment of the data is needed: it is either accurate or not. Its accuracy is entirely self-dependent.

Timeliness refers to the degree to which data are up-to-date. Research suggests that timeliness can be further decomposed into two dimensions: first dimension is currency, or length of time since the record's last update, and the second dimension is volatility, which describes the frequency of updates (Blake and Mangiameli, 2011; Pipino et al., 2002; Wand and Wang, 1996). Data that are correct when assessed, but updated very infrequently, may still hamper efforts at effective managerial decision making ,for example, errors that occur in the data may be missed more often than not with infrequent record updating, preventing operational issues in the business from being detected early. A convenient example measure for calculating timeliness using values for currency and volatility can be found in Ballou et al. (1998), p.468, where currency is calculated using the time of data delivery, the time it was entered into the system, and the age of the data

at delivery which can differ from input time. Together, currency and volatility measures are used to calculate timeliness.

Consistency refers to the degree to which related data records match in terms of format and structure. Ballou and Pazer (1985) define consistency as when the "representation of the data value is the same in all cases" (p.153). Batini et al. (2009) develop the notion of both intra-relation and inter-relation constraints on the consistency of data. Intra-relation consistency assesses the adherence of the data to arrange of possible values (Coronel et al., 2011), whereas inter-relation assesses how well data are presented using the same structure. An example of this would be that a person, currently alive, would have for "year of birth" a possible value range of 1900 – 2013. This is an intra-relation constraint, while that person's record in two different datasets would, In both cases, have a field for birth year, and both fields would intentionally represent the person's year of birth in the same format. This is a inter-relation constraint.

Completeness refers to the degree to which data are complete and full in content, with no missing data. This dimension can describe a data record that captures the minimally required amount of information needed (Wand and Wang, 1996), or data that have had all values captured (Gomes et al., 2007). Every field in the data record is needed in order to paint the complete picture of what the record is attempting to represent in the 'real world. For example, if a specific customer's record includes a name and street address, but not any information for the state, city, and zip code are available, then that record is characterized as incomplete. The minimum data needed for a correct address record is not present. A simple percentage of complete versus incomplete records can then form a potential measure of completeness.

A summary of the data quality dimensions is presented in *Table1*. Once data quality measures are understood, these quality measures can be monitored for the improvement or adherence to standards. For instance, data can be characterized as either accurate or not. As soon as it characterized, there should be a way in place to monitor the long-term accuracy of the data.

In combination with the measuring and monitoring the other three data quality dimensions, help to ensure that the all the records in the dataset areas are accurate, timely, complete, and consistent as is practical.

Table 1 Dimensions of data quality

Data quality dimension	Description	Supply chain example
Accuracy	Are there any errors in the data?	Customer mobile phone number in a customer relationship management system responds to the easiest way to contact the customer any time
Timeliness **Consistency**	Are the data updated? Are the data presented in the same format?	Inventory management system reflects real-time inventory levels at each retail location. All requested delivery dates are entered in a specific format
Completeness	Are any necessary data missing?	Customer's VAT number in an information necessary to complete any shipment

By understanding the four inherent dimensions of data quality allow us to operationally define measures for these dimensions and use all the tools that are needed to actively capture and overcome all data quality problems. Such examples are, the total quality management approaches (Porter and Rayner, 1992; Redman, 1992), the statistical process control (SPC), process capability analyses (Veldman and Gaalman, 2013), and additional quality tools and theories that may help update the data quality management techniques, and investigation, using these techniques in the context of the data quality problem. To this end, tools from SPC have been considered as a physical way for the improvement and the monitoring of data quality over time (Jones-Farmer et al., 2013). Particularly, control charts could be used for the improvement of data quality, not batch-by-batch, but in the overall data production process. Although there are several quality methods that should be examined in a future data quality research, we suggest that SPC control chart methods might be most useful as an illustrative example of controlling and monitoring data quality in a SC DPB setting.

2.8. Big data harvest challenges

In (Tan et. al., 2015) BD is described as having an uncountable volume of data, where the scale of data is varying and the growth of the volume is extremely rapid, so that conventional database management systems cannot deal with this data efficiently. As it is was stated previously, according to IBM till 2020 it is expected the total volume of data to reach 35 zettabytes (ZB). That data explosion make it impossible for the traditional systems to manage such volume of data.

In addition, using different devices for capturing data has as result to collect data in various formats like texts, sensor data, graphs, video data and audio data and other online types of data. These varieties of data need different equipment to handle, process and store. The complexity is increased also because of the shifting from traditional structure formats to semi-structured and unstructured coming from e-mail, social media, log files, sensors etc. The challenge to that point is that traditional analytic tools cannot deal with this different structures of data (Zikopoulos and Eaton (2011)). It is believed (Tan et. al., 2015) that approximately the 80% of the data today is unstructured or semi-structured, setting them unable for further analysis. However, the success of a company depends on its ability to find out insights from various kinds of data that are available, either are structured or unstructured. The ability to be able to analyze all the types of data will give more value and create more opportunities (Tan et. al., 2015).

Every second huge amounts of data are produced, increasing the amount of data having short lifetime. This situation create the need for the companies to make almost real time analysis and making very fast decisions using this data. A literature review (Zikopoulos and Eaton (2011); Cohen et al., (2009); Huddar and Ramannavar (2013)) indicate Hadoop and MapReduce as suitable tools for managers in order to harvest BD. Apache Hadoop is an open-source application that allows users to easily use a distributed computing platform. This platform is capable to deal with large amounts of data in a scalable manner efficiently and reliable. The reliability of Hadoop is enhanced by maintaining multiple working copies of data and redistributing the failed node. In Hadoop multitasking is possible, processing parallel data, increasing the speed. Scalability is one of Hadoop characteristics because it can handle PBs of data. In addition, Hadoop in capable to run applications of data processing that need special infrastructure. According to (Vance, 2009), Hadoop offers high reliability and high fault tolerance to applications. About MapReduce, it is a programming model that can deal with large-scale data sets. It can run parallel computing and can be applied on Hadoop. It is used for the distribution of very big data sets along numerous servers.

However, it is very hard to analyze this massive volume of data in real time and produce useful information using the existing analytics (Bisson et al., 2010). Without doubt it can might help to produce a lot of information, but it is possible to be unreliable, unfocused or even inefficient. Big effort and a lot of time is needed to organize the information that comes up and to filter those that are relevant and viable. An analytic infrastructure is what is required to structure and relate various bits of information to the targeted information.

Therefore, except from generating vast amount of data using software, managers need tools and techniques to structure these data in order to create a better point of view for the problem and gain better insight after analysis. There are many sophisticated analytic techniques such as influence diagram, induction graphs, Burbidges' connectance concept and cognitive mapping that managers can use in order to make visual presentation of the problem (Tan et. al., 2015).

2.9. Big data storage

A major issue of BD storage is the limitation in the hardware that is used to carry large amount of physical data. Tools used currently are not capable to process the storage operations in seconds when facing huge volume of data. The way on how to store myriad of heterogonous or unstructured collected from various sources in logistics, manufacturing, service, and other fields, is really challengeable. Another big issue is the sets for holding super huge volume data which is difficult to insert, update, delete and query because such operations need a lot of time when you have to scan the whole data sets. How to efficiently keep the captured data in a set is very difficult if you have to use the common database or data warehouse technologies.

Cloud based storage and smart storage technologies may be two possible solutions to address the above mentioned challenges. Cloud based storage is able to provide virtually unlimited storage and flexibility to the stored data so as many applications and services to operate using the internet. Within this storage, cloud technology can use implicit security in online mode, an Iidentify-Based Encryption (IBE) to provide authentication, and a web service API to customize the usage. According to the smart storage from IBM, the mechanism is based on an infrastructure with instrumented, interconnected, and intelligent storage devices (Hill, 2012). Three basic tenets are included by the smart storage mechanism. They are efficient by design, self-optimizing, and cloud agile with the advantages of data centric, storage and data management confluence and storage information infrastructure-oriented (Gim, Hwang, Won, & Kant, 2015). Storage like this include self-configuration, self-protection, self-optimization, self-healing and self-management software intelligence so that huge data sets could be handled sufficiently.

According to (Zhong et al.,(2016, pp. 572-591)), future data storage will be carried out in two dimensions related to the media mechanism:

- ***Biological storage media:*** In the future, the magnetic and flash memory materials will be replaced by biological media that is able to keep the data by dynamic,

episodic, and semantic modes. This media deals with the incremental data storage with proliferation of smart cells which are the key components in an intelligent infrastructure, thus unlimited capacity and fast input and output could be achieved.

- ***Learnable storage mechanism:*** Based on the biological storage media, a learnable storage mechanism is able to simulate the ability of human to learn facts and relationships. Two main learning categories are included. The first one is the instrumental learning which is based on the storage behaviors from various smart cells. The other is motor learning that refines patterns according to the operations of smart cell practicing.

2.10. Big Data Analytics

The term Big Data analytics (BDA) is referring to the processes of examining voluminous amounts of variable types of data aiming to uncover hidden patterns, unknown correlations, hidden patterns and other useful knowledge or information (Rouse, 2012). BDA has been in the center of attention due to its ability to provide knowledge, information and patterns to increase business benefits, improve operational efficiency, and explore potential market (LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2013). Having as primary goal to assist companies in decision making processes, BDA allow users to analyze very big volumes of data coming from various sources like the Internet, databases, mobile-phone records and locations, smart-devices as well as sensor-captured information. In order to analyze such big amount of data, being in various formats, the technologies associated with BDA cover a wide range and form a core of open source software frameworks that are able to support the analysis of big data sets across clustered systems (Zakir, Seymour, & Berg, 2015).

Davenport and Harris (2007) underline three key attributes that can characterize analytics competitors. For organizations are usually used basic descriptive statistics and this is fairly straightforward, but between companies, the competition on analytics looks well beyond basic statistical analysis. According to Davenport and Harris (2007), the widespread use of predictive modelling, quantitative techniques and optimization is the most relevant key attributes that play the most important role towards greater profit potential, more profitable operations and better decision making. Many of these

enterprises optimize their SCs, and through that they become stronger and stronger. Many different alternatives are also simulated as well using 'what-if' analysis. Authors go further and introduce a more complicated definition about analytics that is presented below:

'By the term analytics we mean the extensive use of data, statistical and quantitative analysis, explanatory and predictive models and fact-based management that will lead us to drive decisions and actions'.

Analytics, that considered as a subset of business intelligence (BI) encompassing optimization, statistical analysis, predictive modelling and forecasting therefore provide inputs for decision making or even totally autonomous decisions.

The main questions addressed after the use of analytics, and are also presented by Davenport et al. (2010), are reflected in the table below.

Table 2 Analytical questions - supply chain examples, source (Ittmann, (2015))

Analytics	Past	Present	Future
Information	What happened?	What is happening now?	What will happen?
	(Reporting)	(Alerts)	(Extrapolation)
Insight	How and why did it happen?	What's the next best action?	What's the best/ worst that can happen?
	(Modelling, experimental design)	(Recommendation)	(Prediction, modelling, simulation)

Another one, different definition of analytics, is the one that was given by the Institute for Operations Research and Management Science in 2014. According to this definition analytics are:

'the scientific process of transforming data into insight for making better decisions'.

Dr. Gorman, a person with a strong quantitative background, from his side defines analytics into three categories:

- **Descriptive analytics** (what happened in the past):
    - Prepares and analyses historical data

- Identifies patterns from samples for reporting trends.

- **Predictive analytics** (what could happen):
  - Predicts future probabilities and trends
  - Finds relationships in data that may not be readily apparent with descriptive analysis.

- **Prescriptive analytics** (what is the best possible result on a set of circumstances):
  - Evaluates and determines new ways to operate
  - Targets business objectives
  - Balances all constraints.

Compering to the above mentioned categories, we notice that there are many similarities with the definition given by Davenport et al. (2010), Although, Gorman (2012) indicates clearly that the focus for those with a strong quantitative background is on prescriptive and predictive analytics. In addition, IBM also establish Social Media Analytics, referring to the analysis of data that originating from the external environment of an enterprise and which are not easily considered as transactional data. Examples of that kind of data would be data from the famous social media like Facebook, Instagram, Twitter, WhatsApp, etc. IBM also define Entity Analytics, which focuses on grouping and sorting data belonging to the same entity together (Dietrich et al. 2014).

One of the most recognizable and greatest software vendors internationally is SAS. That software developing company is being marketed and promoted by its self as the providers of analytic solutions. A range of components that are included in the SAS Analytics suite of software are presenting below (SAS 2013):

- **Predictive analytics and data mining:** The aim is to build descriptive and predictive models and make use of the results throughout the organization
- **Data visualization:** The target is to improve the effectiveness of analytics using dynamic data visualization
- **Forecasting, econometrics and time series:** Analyzes and helps to the prediction of future results based on historical models and results
- **Model management and monitoring:** Try to improve management processing, deploy, create and analytical models

- **Operations research:** Leverage optimization, simulation techniques and project scheduling so as to identify the actions that will offer the best possible outcomes
- **Quality improvement:** quality processes identification, measurement and monitoring over time
- **Statistics:** Use statistical data analysis to drive fact-based decision making
- **Text analytics:** Maximization and extraction of the value that is hidden in the unstructured datasets.

Computer power is getting bigger and bigger day after day, which allows the enterprises to solve and address of ever-increasing size. Through the use of that developed, though sophisticated computer technology more and more organizations are entering to the era of BD. This has prompted the realization that Davenport (2006) talk about, mentioning that there is value to be obtained from the analysis of the data which can be achieved through analytics. Davenport (2006) provides benefits that companies can achieve using this analysis of data. Those enterprises using BDA seize the lead in their fields, with examples given of companies that build their businesses on the ability to collect, analyze, and act on data. Without any doubt, adopting an analytics strategy will increase company's competitive advantage against its opponents and therefore also its profitability. Sanford (Dietrich et al. 2014) puts it succinctly: 'People respond to facts. Rational people will make rational decisions if you present them with the right data'.

In public sector or in complex private environments, decisions makers are not only dealing with voluminous data but also more complex issues. The ability to use business analytics (BA) in analyzing BD sets enables more informed and better decision making.

2.11. Tools for Big Data Analytics and applications

BDA is something new for the business life which needs some tools in place to handle huge volumes of data sets. Such tools are mainly used to identify trends, detect patterns and glean invaluable findings. To achieve marketing analytics, fraud detection and financial risk assessment, a mix of business intelligence (BI) tools has been used (Stedman, 2014). This tool integrates Oracle's Big Data appliance and Cloudera's Hadoop distribution for these purposes. Typical tools include GridGain for processing large amounts of real-time data, HPCC (High Performance Computing Cluster) for real-

time calculations, Storm for dealing with huge data sets with distributed real-time computation capabilities, and Talend for providing a number of BI service (Loshin, 2013). Harvey (2012) investigates the 50 top tools which are categorized into platforms and tools, databases and data warehouse, BI, data mining, and programming languages, which enable BDA to help enterprises to improve their processes and performance.

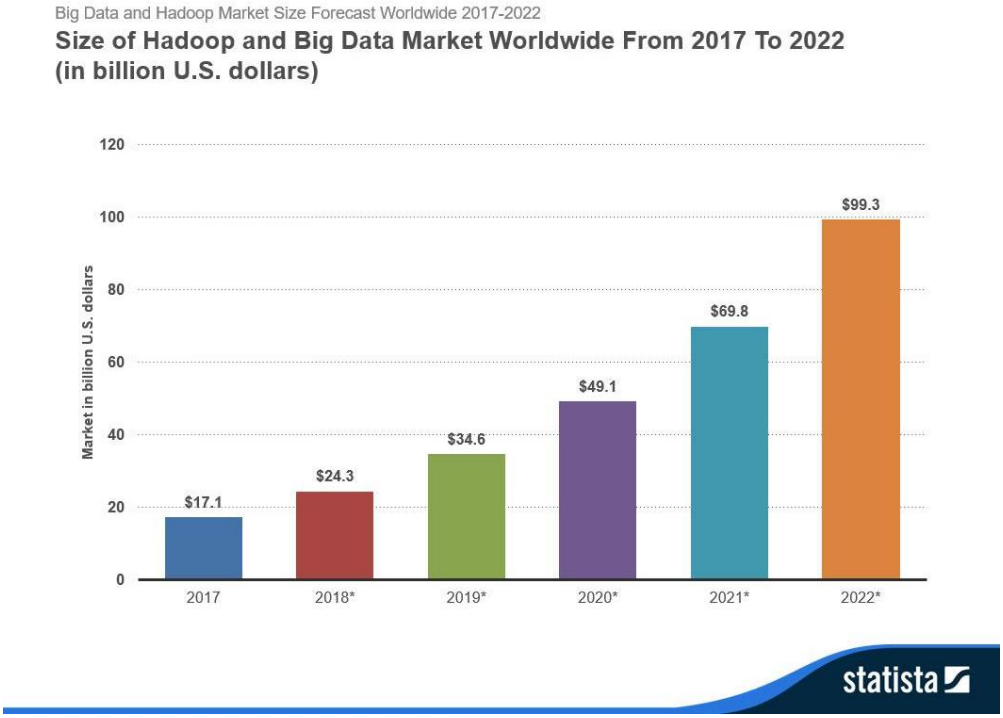

figure 5 Source: https://www.statista.com

BDA tools are enabled by suitable techniques which are attracted by plenty of large IT companies who are able to provide series of core technologies and solutions for different applications. Oracle Advanced Analytics (OAA), combining of powerful in-database and open source R algorithms, enables predictive analytics, advanced numerical computations and interactive graphics (Oracle, 2013). SAP High-Performance Analytic Appliance (HANA) uses parallel multicore processor technique to manage huge database so as to provide various predictive analytics solutions such as customer movements and market fluctuations (SAP, 2014). SAP HANA is comprised of a TREX (Text Retrieval and information EXtraction) search engine, an in-memory OLTP (Online Transaction Processing) database and a new architecture to enable users to advanced decision-making. Microsoft, to fully realize the BDA, provides a complete platform technique to integrate models, analyze and visualize tools to scale analysis companywide (Microsoft, 2014). Power BI for Office 365 and Microsoft Azure are two typical products which use such platform technique to build a visualization tool to uncover the patterns from huge data sets. IBM SPSS Modeler, another extensive predictive analytics platform technique, provides forecastable intelligence to assist systems, groups, individuals and enterprise to

the procedure of decision making (IBM, 2013). By using a set of decision management and optimization, entity analytics, advanced algorithms, IBM SPSS Modeler can play important role in taking the right decisions. Bhatti, LaSalle, Bird, Grance, and Bertino (2012) gave an investigation on BDA enabled techniques in terms of hardware and software perspectives.

Practices of BDA have been widely reported currently. One of the chief targets is to make full use of the data to achieve the ‘‘right data, for the right user, at the right time” (TDWI, 2014). Thus, different companies practice BDA given their specific situations and issues. McKinsey & Company, a global management consulting company, uses BDA to provide a rich set of services to plenty of firms in order to achieve sustainable improvements in performance. For example, in the financial area, with the use of BDA, this company helps small business banks to analyze consumer behaviors to implement digital innovation and upgrading services (Biesdorf, Court, & Willmott, 2013). Amazon, the Seattle-based e-commerce giant, currently leverages BDA to predict the customers’ behaviors so that the goods could be shipped to them before they decide to purchase them (Marr, 2013). We will analyze further that case in another chapter. Intel recently adopts BDA to accelerate deployment and development of wearable apps with data-driven intelligence by integrating a number of tools and algorithms from Inter alongside cloud-based data management system (Bell, 2014). Wide range of BDA practices has been used for finding potential markets, improving customer relationship management (CRM), increase profit margin, and carry out various predictions from the service and manufacturing sectors (Lardinois, 2014; Lohman, 2013; Salehan & Kim, 2016; Talia, 2013; Vera-Baquero, Colomo-Palacios, & Molloy, 2013; Zhong, Xu, Chen, & Huang, 2015).

2.12. Big Data enabled decision-making models

More and more decision making in SC, manufacturing and service require associated knowledge and information which could be gained from BD. Unfortunately, according to a survey conducted in 2015, the 55% of respondents believe that the BD decision models are not considered strategically at senior levels of their organizations (Jin et al., 2015). It is common for many data mining approaches to be packaged or upgraded and be termed as BD based decision-making models. In these cases, such approaches are not able to deal with the BD challenges in specific applications like service, manufacturing and SCM. In this case many challenges should be addressed. Firstly, decision making models may need

data for working our resolutions for different purposes such as optimization in strategical, operational and analytical levels. Although, facing huge volume of data, decision models need a long computational time. Secondly, the data driven decision models do not have evaluation criteria to examine effectiveness and efficiency. Current comparison with other solutions or models is possible not to be suitable for BD driven cases. Thirdly, a decision making model always focuses on a specific problem. Generic models are scarcely reported for solving multi-objective problems. In co-operation with BD, multifunctional models could be achieved.

New models and decision making mechanisms are possible opportunities to address the challenges. On the one hand, these new models can make full use of the BD sets through mining various information and knowledge which could be converted into parameters for optimal decision making. On the other hand, these new models can work quickly to get the optimal results. Such models could be used both in specific problem solving and multi-objective problems by using different BD driven mechanisms. With the historic data built into evaluation criteria, models can be then quantitatively verified.

Future BD driven decision making models would be implemented in two directions:

- ***Self-learning models:*** In the future, decision making models are capable to learn from massive BD for improving themselves. Deep Machine Learning (DML) will be embedded into decision-making models so that they can be equipped with continuous learning ability. Such ability could learn invaluable knowledge from different types of BD so that these models could be extended and fragmentized into new models.
- ***Smart decision-maker:*** Future decision making models are not working independently. A decision making model could invite associated models for collaboratively work through hierarchical or parallel fashion. With the intelligent ability to learn, a set of decision making models are able to form to be a smart decision maker to pick up precise data as parameters, figure our resolutions quickly and evaluate the results sufficiently.

2.13. Big Data trends and perspectives

The term BD firstly stated in the 1970's (Addo-Tenkorang & Helo (2016)), but since 2008 has seen an explosion of publications. Although the term BD is mainly associated with the computer science, research shows that is being applied to variable fields

including engineering, health, earth, arts and humanities and environmental science. In addition, mainframe computer systems are facing difficulties about storage and processing space because of the always increasing data volume. This motivates further the researchers to study more the trends and perspectives of BD, explaining the effective evolution in this area over the years. In the following figures is illustrating some significant findings by an Elsevier research about the trends and special issues on BD in 2012.

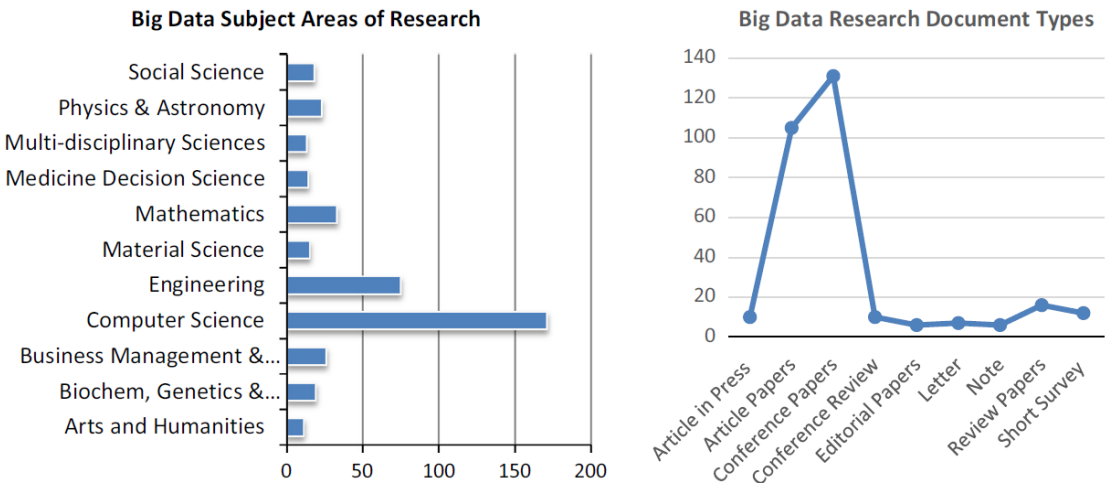

figure 6 Big Data Research Areas and Document Types.
source: Addo-Tenkorang, P.T. Helo / Computers & Industrial Engineering 101 (2016) 528–543

BD is a necessary attribute of the computation-intensive operation and analytics of the data size that can be stored by computer clusters and various cloud-computing system. The most important target for a cloud computing system is to be able to use as much as possible storage and computing resources under concentrated operational management, so as to provide BD applications with veracious computing capacity. Therefore, solutions for processing and storage for BD are provided by cloud computing analytics. Thus, the emergence of BD has also increased the advancement and operational expansion of Cloud computing. Cloud computing virtual storage technology has the ability to effectively maintain and analyze BD using parallel computing capacity to improve the processing and the efficiency of BD acquisition. Moreover, although there are also many similar analytical and operational technologies in cloud computing eminent in BD, there are some differences according to (Min et al., 2014) that include the following

- BD influences the operational decision making processes, while cloud computing systems transforms the IT architecture.
- BD depends on cloud computing as the foundation for smooth analytical operation.

- Cloud computing and BD have different target customers. Cloud computing is possible to be open source while BD is usually not as most of the data are private and classified.

According to Min et at. (2014), for the cloud computing and analytical products the main target customers are Chief Information Officers (CIO), who use the technology of cloud computing to offer advanced IT solutions in operation management and SCM (Gunasekaran & Ngai, 2004). From the other side, BD is targeting also CEOs (Chief Executive Officers), who use processed information that add value and help them to take informed strategic operational decisions. Taking strategic and well informed decisions positively impacts business operation in a more competitive way. With the cooperative advantage of BD and cloud computing technologies, efficient and effective data processing and analysis would certainly and increasingly complement each other. Cloud computing analytical and operating systems provide system-level resources, while BD provides functions similar to those of database management systems for efficient and effective data processing capacity. EMC President Kissinger stated that the application of BD should be based on cloud computing (Chen, Mao, Zhang, &Leung, 2014). Therefore, this implies that BD is expected to expand further into the arena of IoT when it will mature more. The evolution of DB has been massively propelled by the fast-moving growth of application demands and cloud computing advancement from virtualized technologies. Thus, cloud computing provides also computation and processing for BD but also itself is a service mode (Min et al., 2014).

2.14. The future of Big Data – Big Data II/Internet of Things (IoT)

The internet enabled integration of industrial physical products or objects and the digitization of operational activities into the networked society is a just a snapshot of what IoT tends to present (Rosemann, 2014). Therefore, enabling a powerful integrating network of industrial products or objects of all kinds using RFID sensors and readers allows the sensing of signals from such products or objects by analyzing incoming data streams and in return by controlling these products or objects remotely (Zeiler and DHL Solutions and Innovation, 2013; Zhong et al., 2013). Some examples include the sector of healthcare in specific developed countries, which patients are managed remotely, in the energy sector by the use of smart meters, and in manufacturing sector by the use for predictive maintenance (Kees et al., 2015). Moreover, Gartner (2013) mentioned that there will be 26 billion smart objects installed by 2020 that will create market opportunities that is estimated to be over $300 billion. The McKinsey Global Institute

(2013) also notice that the IoT is considered to be one of the most disruptive technologies that impacts the most industrial operations.

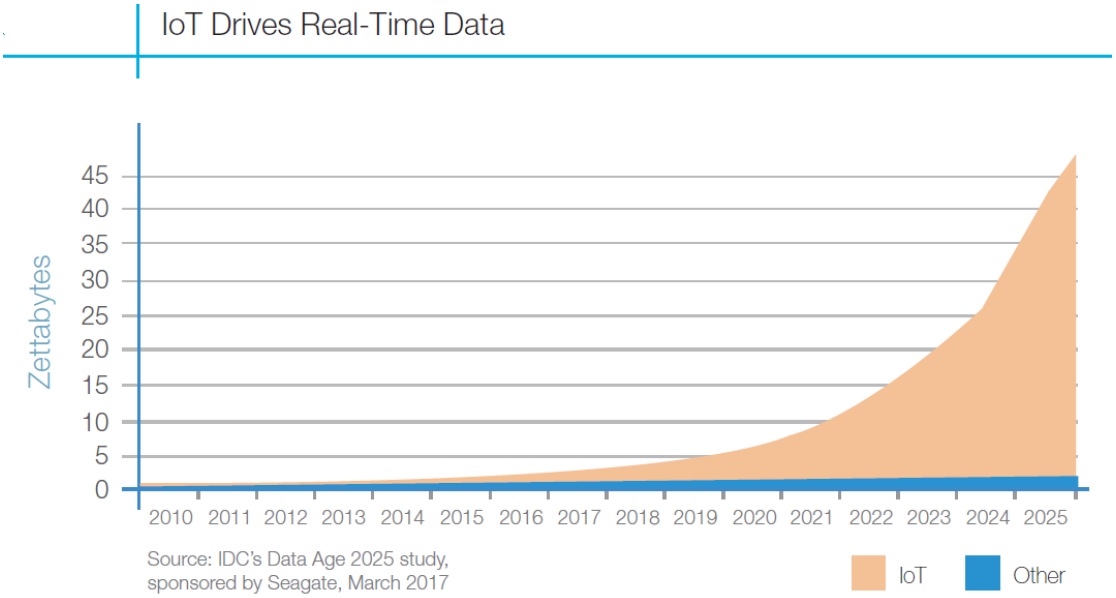

figure 7 Impact of IoT in real time Data.
Source: IDC's Data Age 2025 Study sponsored by Seagate, March 2017

The IoT paradigm shift of the information age create a voluminous amount of networking intelligent agents embedded into different devices and machines in the real world. Such intelligent agents distributed in different fields collect diverse amounts of massive data, such as manufacturing, production, operations, design, environmental, astronomical, logistical and geographical data. Mobile devices, public facilities, logistics facilities, home appliances and health care systems could all be data acquisition equipment in IoT. Quite recently, research has predicted that IoT data would be an internal part of BD by 2030 and the required quantity of intelligent agents will reach the one trillion, thus making data coming frot the Internet of Thing, the part with the highest important of BD, according to a forecast coming from HP. In addition, in a report from Intel it is pointed out that BD in IoT has three characteristics that conform to the normal or general operational BD paradigm:

- BD of IoT is useful only when it is processed and analyzed for value adding
- Various network terminals or points generating a variety of data
- BD generated by IoT is usually semi-structured or unstructured.

It is clear that industrial operators of IoT realize the importance of the potentials and essence of the capabilities for BD and it is understood that the success of IoT is hinged on the effective integration and synchronization of industrial BD and cloud computing systems. Therefore, recently there has been a compelling need to adopt BD in product

development principles and in industrial operational processes in order to enhance IoT applications, while the development of BD is already lagging behind in integration with cloud computing. It is widely recognized that these two technologies are interdependent and should be jointly developed, that means that the widespread deployment of IoT leads the high growth of data (Min et al., 2014). Kees, Oberlaender, Roeglinger and Rosemann (2015) define IoT as the connectivity of physical objects or industrial products, equipped with actuators and sensors, to the internet using data communication technology, enabling interaction with or among these products or objects after having reviewed a number of different definitions in their research, such as Boos, Guenter, Uckelmann, Harrison and Michahelles (2011).

According to Kees et al. (2015) IoT has been comprehensively discussed in terms of engineering related challenges (industrial operations and manufacturing SCM) (Atzori, Iera & Morabito, 2010; Kortuem, Kwasar, Fitton & Sundrammoorthy, 2010), and also from a business to business (B2B) perspective (product and services SCM) (Geerts and O'Leary, 2014). The Information Systems (IS) technology community has been rather passive with regard to researching the customer related business implication of IoT. Moreover, Wamba et al. (2015) researched how BD influence in a high level. Their findings from a longitudinal case study and systematic review show that, despite the strategical and the high operational impacts, there is a lack of empirical research to assess the business value of BD for the benefit of industrial operations or SCM. Therefore, in this research we are attempting to propose a framework for a sustainable, efficient and effective 'Big Data II' for application in IoT in industrial operations and SCM for industrial competitive advantage and innovation in the SC architecture (Gobble, 2013; Waller & Fawcett, 2013).

The IoT has come to the forefront in driving the rapid and voluminous growth of data both in terms of category and quality, giving also the opportunity for the development and the application of BD. On the other hand, the application of BD technology to IoT accelerates also the business models and the research advances of IoT (Min et al., 2014). Moreover, IoT can also enable improvement of customer experience, products and services, security, etc., if it is properly used. In addition, IoT has the potential to develop traditional business-to-customer interactions in a way that previously was not thought of (Kees et al., 2015) when networking sensors like RFID's are applied in a variety of electronic devices and machines to communicate and exchange information or data in real-time and in a real world activity (Zhong et al., 2015). Therefore, IoT represents a

creative disruption, something that begins to overthrow existing processes and technologies and bring forth a completely new way of working and managing electronic network activities (Kaushik, 2015). According to Kees et al. (2015) IoT is considered to be one of the most "hot" technologies as it integrates Internet-enabled physical objects into the networked society making these objects more and more autonomous partners in these digitized value chains. This is best implemented with the use of high level networking sensors like RFID's, thus indicating that the applications of RFID technology is indeed on the increase and are bound to offer new avenues for growth and new opportunities on the emerging frontier of efficient and effective SCM. RFID technology exists for many years, since World War II, but the last two decades it has emerged as the technology used in SCs (Ngai and Gunasekeran, 2007).

2.15. Industrial Awareness of Big Data Solutions

During 2011 there has been already a high interest or new proposed solutions for increasing the Supply Chain Visibility (SCV) based on BDA. But the existing security policies stopped the integration and the application of various data sources. The industry spotted high capabilities for solutions based on HDFS and MapReduce.

The situation about dealing with large amounts of data has not been changed. Existing data is often stored in large files, data silos and warehouses in web logs or in complex XML documents. All those files have to be transformed to existing structured data as they are demanded by conventional databases. The transformation is necessary, if a further usage is planned. The mentioned transformation is very time intensive and the stored data is limited for later usage. The distributed processing is needed to extract required information shortly from large datasets. This is not yet implemented due to the restricted accessibility of different internal data sources. Moreover, security polices mostly complicate the integration of external data in existing enterprise IT infrastructures.

Current solutions for increasing the SCV are optimizing the collaboration within the supply chain. Additionally, 44% of enterprises improved the internal visibility and 40% are optimizing their operations to improve monitoring, usability or efficiency as the (Heaney 2013) reports. This leads to reconsidering security policies to enable the access to different data sources for analytic methods. The understanding that data is valuable has highly increased throughout the industry. But the valuable information needs to be processed.

Following the Gartner survey (Kart et al., 2013), recently around 26-28% of manufacturing companies and retailers invest in BD solutions. In the transportation sector only 20% of the asked companies have already invested, but with 50% there is the highest value of planned invests within the next two years. The problem addressed with BD (summarized over all industries) is about 32% in improving risk management.

3. Logistics and Supply Chain

3.1. Defining the term of Supply Chain

According to Mentzer, Dewott, Keebler, Min, Nix, Smith and Zacharia, authors seems to agree better on the definition given for the 'supply chain' than the definition for 'supply chain management' (Cooper and Ellram 1993; La Londe and Masters 1994; Lambert, Stock, and Ellram 1998). La Londe and Masters suggest that a supply chain (SC) is a chain of organizations that pass materials forward. Generally, seperate independent organizations are involved in the manufacturing of a product or a service and all the procedures needed for this product to be placed it in the hands of the consumer or the end user in a SC. Raw materials and component producers, product assemblers, wholesalers, retailer merchants and forwarding companies are all parts of a SC (La Londe and Masters 1994). Similarly, Lambert, Stock, and Ellram describe the SC as the alignment of organizations that brings products or services on the market. Note that these concepts of SC include the end user as part of the SC.

Another term, define the SC as the network of organizations that is involved, through upstream and downstream linkages, in all the necessary activities that are needed to give value in the form of products and services delivered to the end user (Christopher 1992). To sum up, a SC consists of multiple organizations-chains, both upstream (i.e., supply) and downstream (i.e., distribution), and the final consumer.

For the purposes of this research, we give the following definition for the SC:

"As SC we defined the set of three or more entities, individuals, or organizations that are directly connected with the upstream and downstream flows of products, services, finances, and information from a source to a customer (Mentzer, Dewott, Keebler, Min, Nix, Smith and Zacharia, 2001)".

Encompassed with this definition, we can identify three classifications of SC according to its complexity:

- ***Direct supply chain:*** A direct SC is compromised by the enterprise, the supplier, and the end user that is involved in the upstream and maybe on downstream flows of products, services, finances, and information (Figure 9a).
- ***Extended supply chain:*** An extended SC is consisting for example from the suppliers of the final supplier and customers of the final customer, every part is involved in the upstream and probably in the downstream flows of products, services, finances, and information (Figure 9b).
- ***Ultimate supply chain:*** An ultimate SC includes all the entities involved in all the upstream and downstream flows of products, services, finances, and information from the ultimate supplier to the ultimate customer (Figure 9c).

In figure 9c is presented the worst case scenario of complexity that an ultimate SC can reach. In that paradigm, a third party financial provider is in charge of financing, assuming some of the risk, and offer financial advice. A third party logistics (3PL) provider is responsible to perform all the distribution activities needed between the two firms, and a market research organization is offering information and advice about the ultimate customer to a company as well as to back up the SC. This activities, very briefly present some of the multiple functions that complex SCs can and do perform.

It is critical to understand that implicit within these definitions is included the scenario for a SC to exist whether the flow of material is being managed or it is physical. If none of the organizations in Figure 9 actively implements any of the concepts discussed to manage the SC, then the SC as a phenomenon of business, still exists. Therefore, we draw a clearly distinction between SCs as phenomena that exist in organizations and the management of those SCs. The former is simply something that exists, often referred also as distribution channels, while the latter requires overt management efforts by the organizations within the SC.

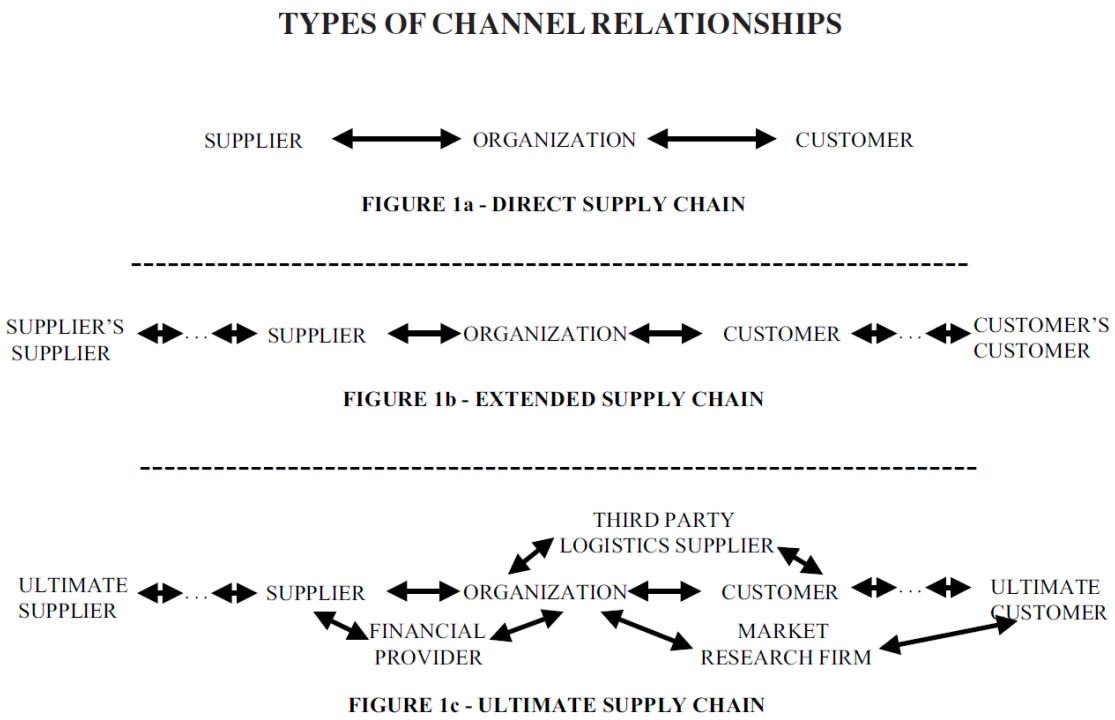

figure 8 source:"Defining Supply Chain Management" Mentzer, Dewott, Keebler, Min, Nix, Smith and Zacharia, 2001, Journal of Business Logistics, Vol.22,No.2,2001

Given the potential for various alternative SC configurations, it is vital to keep in mind that any organization can be part of many SCs. Wal-Mart, for example, can be part of the SC for clothing, for candy, for hardware, and for many other goods. This multiple SC phenomenon begins to explain the network nature that many SCs possess. For example, a telecommunication company might find NOKIA to be a customer in one SC, a partner in another, a supplier in a third, and a competitor in a fourth SC.

Note also that within the definition given of SC, the final consumer is considered a member of the SC. This point is important because it recognizes that retailers such as Wal-Mart can be part of the upstream and downstream flows that constitute a SC.

3.2. Logistics and Supply Chain Management definition

In our days, enterprises use the term SCM to signify the way in which SC processes are structured and managed. SCM is referring to the steps done so as the external material processes are managed. In that point we can distinct the incoming material flow and the outgoing materials flow. The outgoing material flow relates to the way in which finished products are distributed by the company to its customers. This is activity is commonly

denoted as physical distribution. The incoming material flow, obviously, covers all activities needed to optimize the goods flow up from suppliers to the point of consumption within the company itself. This activity used to be referred as materials management. In many cases, the scope of SCM goes one step beyond in that it also relates to materials flows optimization from the supplier's to the firm. As we see in that research, SCM has matured duo to the fact that advanced information systems have become available, which are able to track and trace complex materials flows in excellent detail.

SC thinking started several decades before, with logistics management. The term 'logistics' originally stems from military organizations and was already in use in the days of Louis XIV of France. Even then it was clear that the effectiveness of any military operation or organization did not depended solely on the power of weapons, the military power and the fighting spirit of soldiers, it was also affected by the possibilities of transportation and the efficient supply of food and ammunition. The rationalized consideration of the transportation and supply of food, aminations and materials was called logistics. The French military successes of that period of time were mainly because of the importance that was given to logistics.

Logistics and flexibility are going hand to hand. Flexibility is getting a lot of attention in many organizations today where functional thinking is dominating. In a functional organization individual departments such as production, product development, sales, purchasing, personnel and administration, essentially are managed as separate activities. In most cases managers responsible for each of these activity areas, report directly to the executive board. Each department has very specific tasks to manage that should be realized through limited resources, which are agreed in an annual departmental budget. Realization of these budgetary targets is an important factor in the assessment of departmental managers. They, therefore, often strive to realize their own budget targets, even at the expenses of other departments

This practice can easily lead to departments operating fairly autonomously so that cooperation of the whole is left to the interplay of forces between the departments themselves.

Logistics and SCM aim to counterbalance the shortcomings of the functional organization, by focusing on those processed through which customers can be better served. Logistics and SCM favor a process approach rather than a functional approach. All processes are aligned in order to meet specific customer needs and focused on creating

maximum customer satisfaction. Superior customer service, efficient customer complaint handling, and customer driven product development and innovation are important corner stones of logistics and SCM.

Logistics management is related to all materials flows, from the flows of purchased materials into facility, through the manufacturing process, and out to the customer. The starting point for any logistics process is the short-term sales plan and the related product plan. The total logistics function therefore includes short-term materials planning, the supply of raw materials and other purchased goods, internal transportation, storage and physical distribution. It may also include in some companies reverse logistics i.e. recycling packaging materials and surplus materials (Purchasing & Supply Chain Management: Analysis, Strategy, planning and Practice, A. J. Van Weele. Eindhoven University of Technology, 2000).

3.3. Logistics and supply chain operations

3.3.1. Demand planning.

A very important aspect in SCM is the processes management and operations management in order to meet demand and deal with variations in both process and demand. However, demand variability and process variation may become an obstacle in achieving a match between demand and process capacity. Effective capacity planning requires accurate demand forecasting, the ability to understand forecasts and translate them into capacity requirements and SC operations capable to meet anticipated demand. Therefore, demand planning is critical to SC operations planning (Chen and Blue, 2010; Wang et al., 2016).

The target of the demand planning is to analyze various customer segments in terms of brands, channels and product down to the Stock Keeping Unit (SKU) level, and develops models used for shaping demand and create revenue plans, which is the foundation for cooperative planning and forecasting with important SC partners (Jonsson and Gustavsson, 2008; Chen and Blue, 2010; Haberleitner et al., 2010). Demand planning is bi-directional allocation and aggregation, integration with product, brand and SKU level forecasting at different hierarchical levels through information sharing among partners and increasing SC visibility by allowing SC partners to access real-time inventory and sales information (Choudhury et al., 2008; Gallucci and McCarthy, 2009). Therefore, demand planning is not just forecasting, but involves operations and sales planning.

Demand forecasting on independent demand items predictive analytics using time-series approaches are necessary (Cheikhrouhou et al., 2011; Li et al., 2012). Exponential smoothing is also another method among the time-series methods, which is commonly used for both short-term and intermediate range forecasting since it can incorporate both seasonality and trend. Another method with high importance is the auto-regressive model, which achieves demand forecast in a period using a weighted sum of demand realizations in previous periods. In addition, intermediate range forecasting can also be realized from associative forecasting methods, particularly in service industries or in manufacturing the non-discrete items (Lu and Wang, 2010; Beutel and Minner, 2012). Using forecasts of sales, demand, and optimization techniques, sales and operations planning (S&OP) provides an integrative cross-functional management capability for production, marketing and inventory management to manage operational components and ensure customer promises (Feng et al., 2008; Chen et al., 2010; Sodhi and Tang, 2011; Lim et al.,2014).

3.3.2. Procurement.

In procurement a voluminous amount of data is generated from various sources and applications through spending, supplier performance assessments, and negotiation, whether internal or external. These data sources facilitate the use of advance analytics. In the case of DHL for instance, the combined use of external operational and macroeconomic data enhances its SC operations efficiency. SCA provides procurement decision makers with consistent, databased analysis for a wide variety of major decisions and business issues, e.g., material availability and quality problems (Souza, 2014). In the literature, the application of SCA in procurement is illustrated in the following aspects: (a) managing supply risks and (b) managing supplier’s performance.

SCA gives to the organizations the ability to distinguish between risks that have to be avoided, and risks that have to be taken by identifying trends and events through monitoring publicly available news or social media channels associated with suppliers or specific sourcing markets. Therefore, organizations can continuously gain updated information on suppliers and sourcing markets and quickly react to supply risks or changes even with contingency plans. Scholars focus on using methodologies or generic models to model supply risks or evaluate the impact of supply risk on SC performance (Kabak and Burmaoglu, 2013; Mishra et al., 2013; Zeotmulder, 2014). Other researches develop, inter alia, optimization based approaches and mathematical models to supplier relationship management with supply disruptions (Case, 2013; Khan, 2013).

SCA is also a powerful tool for helping firms to analyze, measure and manage their suppliers' performance for better sourcing (Oruezabala and Rico, 2012). Through comprehensively collecting and consolidating all forms of supplier data across global organizations, SCA can quickly analyze and evaluate suppliers' performance like quality for example, delivery guarantee and timeliness, and spend analysis, thus helping procurement organizations make informed decisions(Walker and Brammer, 2012; Yeniyurt et al., 2013).

3.3.3. Production.

SCA can give the chance to manufacturers to have better understanding of the various production costs involved and how the bottom line is influenced. The application of SCA can provide useful insights regarding the production capacity levels and inform decision makers if improvements are needed to maximize productivity (Jodlbauer, 2008; Heo et al., 2012; Noyes et al., 2014). SCA can also help manufacturers of multiple products adjust production to ensure that the correct mix of resources is allocated to the proper production lines. Furthermore, SCA is used by production analysts to identify material waste and manufacturing techniques and processes that can reduce or even eliminate material waste (Sharma and Agrawal, 2012). Therefore, SCA can be applied to production planning at both the tactical and operational level for aggregate planning and operations scheduling (Souza, 2014).

To enable aggregate planning SCA permits decision-making related to, inter alia, matching demand and supply, inventory management, and budget forecasting. After sales forecasts and resource requirements, the various alternate production plans are generated (Wang and Liang, 2005; Liu et al., 2011; Mirzapour et al., 2011; Li et al., 2013). Additionally, SCA provides useful insights to problems related to operations scheduling problems, which can be formulated as mixed integer linear programming problems (Wang and Lei, 2012, 2015; Wang et al., 2015). In routing problems for instance, SCA can help in e.g. modeling the sequence of operations and the work centers that perform the work and dispatching (Chen and Vairaktarakis, 2005; Sawik, 2009; Chen, 2010; Leung and Chen, 2013).

3.3.4. Inventory.

Business organizations are continuously collecting voluminous data sets coming from ERP systems because of Internet, software applications and electronic devices. Data generated in ERP systems includes historic demand and forecasting data, holding cost, replenishment lead times, the desired service level and fixed cost of placing a replenishment order. Challenges, like diverse organizational needs and supply and demand fluctuations, impact on inventory levels (Sage, 2013). SCA can help organizations well design modern inventory optimization systems needed to handle the most complex retail, whole-sale, and multi-channel challenges in inventory management (Hayya et al., 2006; Jonsson and Mattsson, 2008).

The use of SCA in Vendor Managed Inventory (VMI) systems enables collection, processing, and reporting on inventory data and can therefore inform decisions related to inventory performance improvement (Borade et al., 2013). SCA can also help to the accurately prediction of inventory needs and in responding to changing customer demands, utilizing statistical forecasting techniques (Downing et al., 2014; Wei et al., 2011), as well as the dramatically reduction of inventory costs (Babai et al., 2009). In addition, SCA is applied to address problems that occur within multi-echelon distribution networks (Wang and Lei, 2012, 2015; He and Zhao, 2012). It determines the appropriate inventory levels while taking under consideration factors such as demand variability at the network nodes as well as performance (e.g., lead time, delays, and service level) (Gumus et al.,2010; Guo and Li,2014). SCA helps obtain a holistic view at inventory levels across the SC, while taking into account the impact of inventories at any physical level or echelon on other echelons. SCA assists in decision making related to safety stock optimization (Fernandes et al., 2013; Guerrero et al., 2013).

3.3.5. Logistics

According to the Council of Supply Chain Management Professionals (Stroh, 2002), global logistics generates the massive amount of data as logistics service providers, shippers and carriers manage logistics operations. BD stemming from, for instance, mobile devices, RFID tags and EDI transactions (Swaminathan, 2012) can be harnessed for logistics planning purposes. This deals with the distribution of products from supply points (i.e., production facilities or warehouses) to demand points (i.e., retailers sites) through intermediate storage nodes (e.g., distribution centers). Logistics planning problems can be formulated as network flow problems where each arc represents a shipping mode with a given capacity and time period (Dong and Chen, 2005; Jharkharia

and Shankar, 2007; Grewal et al., 2008). Logistics details generated from different sources in distribution networks like shipping costs, forecasts on supply capacity at suppliers' plants, demand forecasts in demand points, and network capacity (Najafi et al., 2013). Because of supply disruptions and demand uncertainty, predictive analytics tools are essential to design SC flexibility into logistics operations.

In logistics planning it is fundamental to optimize both equipment and crew routing. The vehicle routing problem attempts to optimize the sequence of visited nodes in a route, for example as for a parcel delivery truck, for returns collection truck or for both (Drexl, 2012; Ozdamar and Demir., 2012). The optimal sequence considers the distances between each arc, left turns, expected traffic volume, and other constraints placed on the routes, like delivery and pick up time windows (Vidal et al., 2013). Although, vehicle capacities, multiple vehicles, tour-length restrictions, and delivery and pick up time windows among others complicate the planning of transportation and distribution operations in global logistics network (Li et al., 2010). Analytics methodologies and techniques are used to optimize the routing of shipments, vehicles, as well as crew (Novoa and Storer, 2009; Lei et al., 2011; Minis and Tatarakis, 2011) in order to balance between margin and transportation costs, and pay attention to safety and maintenance.

3.4. Analytic techniques in SCA

Based on our literature review and analysis, in this part we out-line popular techniques for SCA. As the central component of SCA, advanced analytics techniques are the basis for the success of SC strategies implementation, and daily operations for every business organizations. This taxonomy can be further developed in a future research.

3.4.1. Statistical analysis

Statistical techniques are consisting of two types of techniques: qualitative techniques and quantitative techniques. Qualitative methods are based on subjective judgment of experts or consumers. These methods are appropriate when past data are not available. Quantitative approaches are used to make predictions as a function of past data. Both methods are applied to short or intermediate-range decisions. Two widely used quantitative techniques in SCA are 'time series analysis and forecasting' and 'regression analyses. Time series analysis analyses data finding to extract meaningful patterns and statistics. Time series forecasting makes predictions of the future based on historically observed data. Regression analysis helps in understanding relationships and causality effects between variables.

BD is characterized by velocity, volume, variety, value and veracity which leads to the following challenges to BA (Fan et al., 2014): (a) volume accumulates data noise, and incidental homogeneity; (b) high volume creates high computational costs and algorithmic instability; (c) high variety requires various techniques and methodologies. These challenges result in heterogeneity, experimental variations and statistical biases. Therefore, more adaptive and robust procedures are required because traditional statistical methods were designed for moderate sample sizes and low-dimensional data, but not for massive data. Due to BD features, effective statistical procedures have received increasing attention for exploring BD.

3.4.2. Simulation

BD brings more challenges for modeling and simulation (Sanyal and New, 2013; Parashar, 2014). Firstly, depending on reductionism and causality, the basic simulation theory cannot meet concepts like boundary, target, constraints, and entity among others. Secondly, BD makes modeling methods difficult to perform well and requires new types of models because problems are more complex and need large amount of computation.

Although, modeling and simulation can benefit from BD (Be-laud et al., 2014; Pijanowski et al., 2014). SCA offers more in-depth analysis and processing, and new methods for the simulation problems with gigantic amounts of data. What is more, SCA makes it possible for modeling and simulating complex systems as it focuses on the inter relationship between SC operations, and emphasizes the analysis on integral data associated with SC integration. SCA can aggregate the disintegrated data from different SC operations and achieve global optimization (Ranjan, 2014).

3.4.3. Optimization

The use of optimization techniques as part of SCA helps for improving the accuracy of demand forecasting and SC planning, while creating challenges that relate, for example, on applying penalized quasi-likelihood estimators on high-dimensional data creates large-scale optimization problems (Slavakis et al., 2014). BD optimization is not only instable and expensive, but presents slow convergence rates, thus making traditional techniques difficult to succeed in SCA. To deal with the gigantic size of BD, therefore, it is needed to implement large-scale non-smooth optimization procedures, develop randomized and approximation algorithms and parallel computing based methods, and simplify implementations (Fan et al., 2014).

Conversely, optimization techniques are suitable for data analysis in LSCM. Optimization helps in analyzing highly complex dynamic systems with enormous data volumes, multiple constraints and factors and can gain insights that allow decision makers to make correct and appropriate decisions. Additionally, optimization helps analyze the measures of SC performance such as demand fulfillment and cost reduction, among others. Another one benefit associated with optimization is its flexibility because it can reveal new data connections and turn them into insights unlocking more business value from huge volume of data (Balaraj, 2013).

3.5. Demand driven supply networks

The principle of demand driven supply networks (DDSN) or demand driven value networks DDVN (Gartner) is not something entirely new. This has been an area of study with main aim to help us understand the impact of data latency and aggregation, as information propagates upstream the SC from the source of demand to the suppliers, namely the bullwhip effect.

Demand driven supply networks tend to be different from traditional SCs. They are highly adaptable, agile, responsive, and even predictive. According to Thalbauer (2014), as cited by Howells (2014), "the SCs in the future will not be chains at all, but will transform into demand networks". The term 'chains' has effectively been replaced with 'networks' which, according to Emmett and Crocker (2006), is an attempt to find out new terms that are more representative of the contemporary approach that best represents SCM practices today, focusing not just on the customers, but also on customers' customers.

Demand networks are customer-driven, and as they migrate their focus from the 'supply' to 'demand' side, they are characterized by the need and use of increasingly high levels of dynamic information, impacting almost every area of performance measurement, such as quality, cost, frequency, availability, , etc. In short, demand networks aim to deliver exactly what the customer wants, when they want it and as they want it, and also in the most competitive way. More accurately, when integrated with BD initiatives, demand networks have the potential to support organizations that may also offer what they think their customer may want, based on their profile data, actual purchasing history or other parameters, emanating from structured, semi-structured or unstructured data.

Such demand networks require increased levels of productiveness, pro-activeness, optimization and resilience. They are effectively integrated, managed and controlled at every level and every node of the demand network, ranging from sourcing, production, and product development through to shipment and reverse logistics.

These networks have moved as close as ever to the final customer of consumer on the one hand, and to intra-company operations, like marketing and customer service or new product development on the other. They are driven by data such as predictive data and BD analytics, enabled by smart technologies like the Internet of things (IoT), sensors and are delivered by new paradigms in logistics in many sectors, such as services and retail. In Figure 10 an example of complex and dynamic supply and demand network is presented (Wieland and Wallenburg, 2011).

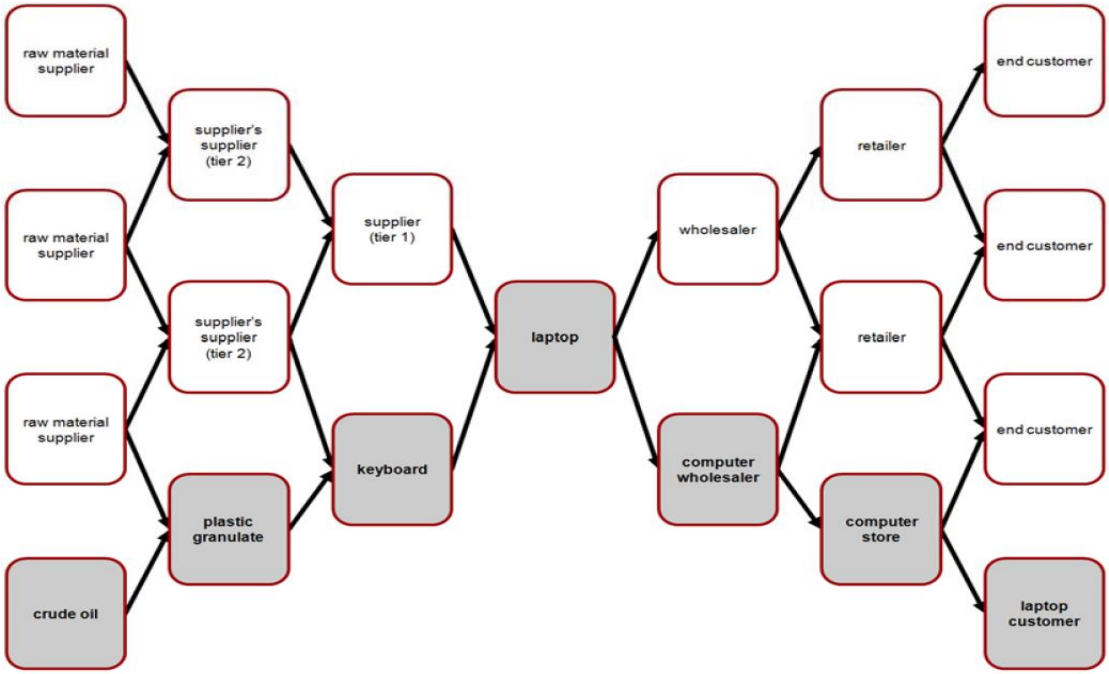

figure 9 Example of complex and dynamic supply and demand network.
source: Wieland and Wallenburg, 2011

A key enabler to optimize logistics for demand driven supply networks is a wide array of technologies which may be integrated directly with various enterprise management information systems, like enterprise resources planning (ERP), warehouse management systems (WMS), customer relationship management (CRM) systems, inventory and advanced planning and scheduling (APS) or other internet-enabled systems and applications.

Some systems incorporate technologies which incorporate advances in the area known as Internet of things (IoT), which use various tracking and sensing devices, such as radio frequency identification (RFID), location enabling technologies, global positioning systems (GPS) for example, geographic information systems (GIS) and other route optimization software.

The common theme of Logistics demand networks and its reliance on the above IS technologies is their dynamic on-demand nature. In a recent study reviewing the performance of IT systems, and specifically ERP, under dynamic market requirements (Tenhiala and Helkio, 2015), it is stated that the use of appropriate software is crucial in order to 'exploiting dynamic market conditions' (Sambamurthy et al, 2003). This ability is termed 'market capitalizing agility' in IS literature (Lu and Ramamurthy, 2011) and 'dynamic capabilities' in management literature (Teece et al, 1997).

Tenhiala and Helkio (2015) state that it is not clear what kind of software is appropriate for organizations that face dynamic market requirements. Currently, the bigger part of current organizations are subjected to dynamic market conditions, as a result of economic volatility of global markets, or other socio-political events. The dynamisms' of market requirements is further emphasized through the generation and accrual of voluminous data, emulating from several diverse sources, including sensor technology and Internet of things (IoT), web analytics and social media.

3.6. Supply Chain agility

Agility is being considered as a key factor through which a SC has the ability to adapt to the always changing market environment (Christopher and Towill, 2000). Plenty of different definitions of SC agility have been developed using conceptual models, interpretive structural modelling and normative indexes (Swafford et al., 2008; Gligor and Holcomb, 2012). It has been approached using several measures and dimensions like range, adaptability (Swafford et al., 2006), process integration, market sensitivity, collaborative planning (Agarwal et al., 2007), demand response (Braunscheidel and Suresh, 2009) data accessibility, swiftness, alertness and flexibility (Gligor et al., 2013). Among the large number of classifications of SC agility, two underlying dimensions are commonly expressed.

First, agility signifies the ability of a SC to react swiftly to unplanned or unexpected external situations. Responsiveness involves the need to perceive demand without latencies of distortions. The visibility of data information is therefore a fundamental characteristic of SC responsiveness as it increases the demand sensitivity. The inherent uncertainty of supply and demand and associated SC risks stipulate the need to be able to rapidly change SC operations. The rapid reaction and detection to SC risks and unexpected events is a characteristic of responsiveness that comes second. A third dimension of responsiveness relates to the speed with which organizations in a SC are able to deliver services or goods (Reichhart and Holweg, 2007).

A second characteristic of SC agility is based on the characteristic of the SC to show significant flexibility. This is the planned ability of cooperating organizations to adapt to expected demand uncertainty and deal with fluctuations and variation, by restructuring their operations, realigning their strategic objectives or reconfiguring their capabilities (Swafford et al., 2006). In the operations and SCM literatures flexibility is a well-established and complex construct. It is generally considered as the capability to change capacity to meet the expectations in case of changes in customer demand (volume flexibility), capability to change the variety of products and services that it produces at any time (variety flexibility), ability to introduce new or revised products (new product flexibility) and the capability to adjust the lead time of its products or services (delivery flexibility) (Reichhart and Holweg, 2007). Building on the above characteristics, the conceptual dimensions against which the capabilities of the proposed Multi-Agent Based Systems (MAS) considered in this research are shown in the following Figure 11.

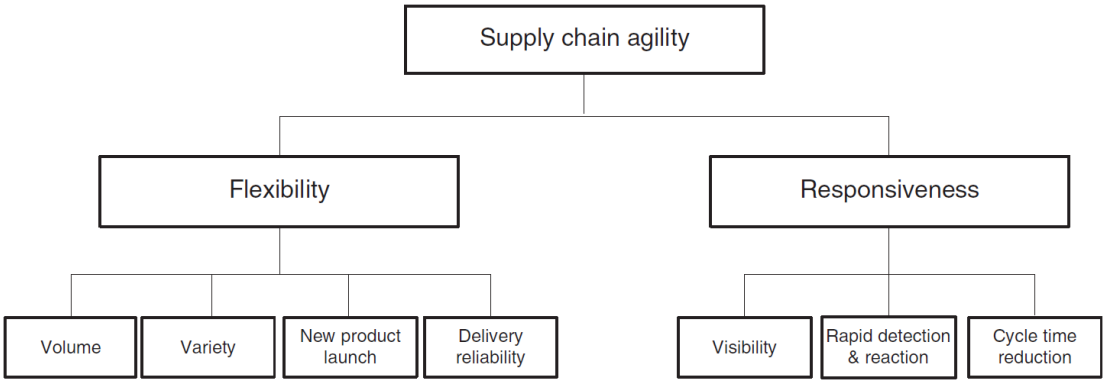

figure 10 Conceptualization of supply chain agility

source: Mihalis Giannakis, Michalis Louis, (2016) "A multi-agent based system with big data processing for enhanced supply chain agility", Journal of Enterprise Information Management, Vol. 29 Issue: 5,pp.706-727

3.6.1. Information systems as facilitators of supply chain agility

It is clearly documented in the literature that IT integration is a trigger for SC agility, flexibility, and finally higher business performance (Swafford et al., 2008). A big range of means have been forced so as to achieve SCM effectiveness: enterprise resource planning (ERP), e-commerce, and advance planning systems (APS) (Moyaux et al., 2006).

Successful e-commerce requires very advanced information systems, which can deal with the high complexity of SC processes and computational capabilities capable to analyze the 'big data' available to organizations today. The latest generation of traditional ERP and APS systems provide a high level of SC process integration through internet-based applications (Link and Back, 2015).

However, this integration are grouped either to the internal business processes, or to a dyadic context of collaboration (Botta-Genoulaz et al., 2005). Their computational realization of various types of business relationships is also limited (e.g. buyer/ vendor relation, CPFR). Therefore, they are bound by an inherent constraint to simultaneously facilitate different 'kinds' of co-operation and to provide efficient transition from one type to another. Little development has been done to find solutions for holistic cross-organizational co-operation. A considerable amount of funds and time is required in order to transform conventional e-business systems into cooperative SCM systems, noticing that only organizations that can afford the required high level of investment will be able to achieve responsiveness, and overall to reinforce their competitive advantage. The small parties of the SC would face considerable constraints in order to benefits from e-business, even if ERP systems manage to totally realize the concept of the extended enterprise. The computational capability of conventional information systems is also limited to analyze huge quantity of data, or data that is unstructured or too expensive and complex to process and exploit. They can incorporate some elements of data mining analysis, for example clusterization, and correlations of information items can be achieved (Berkovich and Liao, 2012), however, their ability to provide real-time analysis of data as well as generate knowledge from BD are nonexistent (Mayer-Schönberger and Cukier, 2013).

3.7. Supply chain Visibility (SCV)

3.7.1. A precise definition of SCV

The following definition describes the term Supply Chain Visibility (SCV) which captures and unifies the following important characteristics:

Supply chain visibility characterized as the identity, location and status of entities transiting the supply chain, captured in timely messages about events, along with the planned and actual dates/times for these events.

Each characteristic of the definition is important and needs further explanation. An entity is any object moving through the supply chain.

An entity can be an item (SKU), a form of packaging (e.g. pallets, cartons, packages, or boxes), an entire customer order, a form of encasement for the order (e.g. a pallet, container, tote, or returnable plastic container), a shipment (a collection of orders with a common destination and origin), a leading asset (e.g. a trailer, container, railcar or uniform load device (ULD) for an aircraft) or a vehicle (e.g. truck, aircraft, ship, or train).Entities form a natural hierarchy as illustrated in Figure 12.

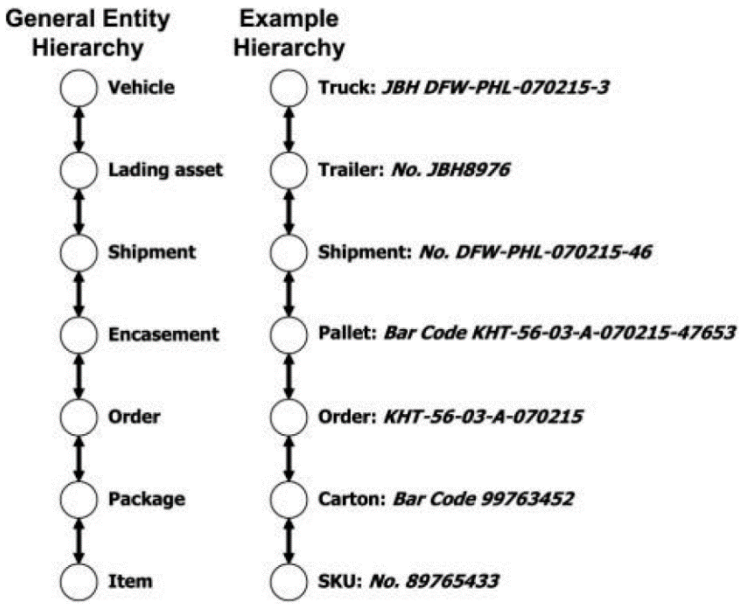

figure 11 entity hierarchy

According to this hierarchy it is implied that items are contained in packages. One or more packages constitute a customer's order. Each order is encased in pallets, totes, returnable plastic containers or any other form of encasement. Encasements use to be loaded into lading assets like container, trailer, ULD etc. And all lading assets are transported by vehicles (truck, aircrafts, trains or ocean vessels). The hierarchy is defined with association to the pairs of entity identifiers. This hierarchy and its constituent associations are not only natural, but also necessary. In case of an inquiry about an order in transit first knowledge is required about the vehicle that transports the lading asset, the encasements contained in the lading asset and the orders within the encasements. For example, in Figure 12 it is shown that the SKU 89765433 is associated with a carton with bar code label 99763452. That carton is part of a customers' order number KHT-56-03-A-070215. The carton is encased in a pallet with bar code label or RFID tag KHT- 56-03-A-070215-47653. This pallet is part of the shipment DFW-PHL-070215-46. The shipment is loaded in a trailer with license plate or reference number JBH8976 and the trailer mated to the vehicle truck, train etc. JBH DFW-PHL-070215-3. These associations are necessary and dynamic as the entity transits the SC.

The identity of an entity is simply a unique identifier or a unique reference for it. For example, a carton's serial number that is often encoded into a bar code or captured on an RFID tag, an order or shipment number, a trailer serial number or license number and an aircraft tail number.

Location has to do with the association of an entity to a position. Location is dependence with the entity and its position in the hierarchy. For example, when we have an inquiry about the location of a customers' order, it is possible to be necessary to know the location of the truck associated with the trailer that contains the order. Locations can be static or dynamic. An order that is waiting for loading into a container is static in warehouse. An order encased in a pallet within a trailer associated with a truck in-route exemplifies dynamic location because the truck is in motion.

Status describes the state of the entity. The possible state of an entity is dependent on the processes that affect the entity. For instance, status for an order might be 'waiting to be picked', 'picking in progress' or 'picked, packaged, awaiting loading'. Status for a trailer might be 'parked empty', 'being in loading', or 'parked loaded'. An event is a specific time when a task in a defined process is complete and, similarly, when the next task in a

process begins, it is the specific time when the location and the status of an entity changes. Therefore, processes define events. Examples of events are ‘order picking complete’, ‘trailer loaded’, ‘truck arrived/departed’ and ‘aircraft landed/parked/unloaded/empty’ etc.

A message is a way to establish a communication that contains all the required data about an entity. That data could be its global position, the identification number and its status. The way of communication is irrelevant. It can be an Electronic Data Interchange (EDI) transmission or an automatic data capture, automatic or manual updating of a web page even a telephone call or an e-mail.

Actual and planned times refer to the times when an event is scheduled to take place and when it actually occurs. Without this information in a message about an event, the inquiring party has no basis to judge if the occurrence of the event is problematic or normal. Therefore, this definition of SCV presupposes the existence of properly and strictly defined processes that specify events and steps in the SC, and a plan for the entity in question.

Summarizing, the definition proposed here is all about the sufficient and necessary information (identity, location and status) required for entities, even if they are stationary or moving, that are hierarchically organized, making their way through the SC. This requisite information is transmitted in messages about events, as defined within processes. Date and time of actual event that took place are compared to the corresponding planned date and time to provide transparently the implications for decision making.

3.8. Innovation in Logistics

For an organization the dimension of innovation and the dimension of time are two critical characteristics that give competitive advantages (Zalewski, 2010). In our case, time describes how frequent an enterprise for example presents a new edition of a product that is considerably upgraded than the previous edition of that product or develops and introduce in the market a completely new product. The development of that product has as a result to affect in a negative way its lifecycle contrary to conventional products. The different phases of product’s life cycle are getting shorter in the time dimension and the demand is getting rapid (Zalewski, 2015).

The phenomenon of the booming market, that has been developed to meet the needs of the increasing customer demand, affect the companies and impose them to change their organizational structure. The most critical change that is imposed is the one that shortens the product lifecycle. We meet this phenomenon especially in the automotive industry, and below we present the example of the Volkswagen Golf. The first generation of VW Golf, was in production since 1974 and stopped in 1993. In this period of time we consider also in the range of models the convertible version that was in production until 1993, and the Caddy van that was produced till 1992. In South Africa though, exactly the same car was produced continuously without any interruption till 2009. On the other hand, the 6th generation of VW Golf was in production line just for four years, from 2008 to 2012. Nevertheless, there were many extensions to the range of that model when produced, which means that the 6th edition of VW Golf were available in plenty combinations for engine, bodywork, equipment, comfort accessories etc.. This happened because the need to meet the needs of the constantly informed and demanding customers, was increasing day after day. So, car manufacturers in this case, in order to increase their portion in the market use to offer newer models in the market but with a poorer quality level than the previous models. This phenomenon is characteristic noticeable is the fast moving consuming goods. This led to the creation of simpler, cheaper and more attractive products to cover the needs and the demand of the less affluent and less prepared customers. From this point of view, we assume that the lifecycle of any product will be noticeable shorter than its previous edition-model. Examples of that products use to be technologically advanced and good value, however their shelf life is for a specific period of time or limited. That kind of market changes have as a result the appearance of the phenomenon of disruptive innovation (Zalewski, 2010).

Considering the sector of logistics, the term innovations is not, however, absolutely connected with the implementation of high-end IT solutions. A sign of modernity can also be a way of thinking. According to Doskonałość, (2010), innovative solutions in logistics can also be expressed by:

- continuous development, improvement and training of the team that carries out innovation and continuous verification of work and commitment
- being constantly careful and vigil over the quality of the innovative activities
- being always focused on the work done by the team which is responsible for the implemented practices and shared values

- activities that include the constant search to find out new and better ways to implement the logistics procedures
- Being satisfied with the work done and being also honest to customers, eliminating old bad habits, behaviors and barriers associated with changes in the area of logistics procedures.

The most critical drivers that led organizations and push them to invest in innovation and create new value in logistics are the organizational culture and the human capital.

At this point, it whore the while to mention the two greatest breakthrough innovations in logistics industry. The first one is the introduction of the container, which totally revolutionized the material flow and the second one is the development of RFID technology, which has contributed by bringing the transparency all along the SC. Pfohl also mention some additional factors that play important role for the success:

- the structure of regional networks, risk management, rotation means and flexibility
- increased customer claims such as lead time delivery services, product availability, product quality and reliability
- services prepared after taking into consideration consumer needs, and consequently rapid response to their requests
- segmentation of the SC focused on demand and specific needs of customers, which can help to the reduction of the stocks volumes, and thus the achievement of cost optimization
- safety requirements and potential hazards in the SC
- risk management in the SC
- strategies for sustainable development of enterprises with regard to environmental aspects

All these ideas, factors, proposals and trends that we mentioned above should be included and considered in the solutions that are characterized as innovative. For this reason economies of scale should be able to lead to solutions that allow logistics operators to meet the always increasing requirements of the consuming trends that we meet in the 21st century.

3.9. RFID Technology

The technology of Radio frequency identification (RFID) is an innovative tool that is used in a large extend to support logistics and supply chain procedures in manufacturing and distribution sector where production resources and goods equipped with RFID tags are evolved into smart manufacturing objects (SMOs) which are able to sense, interact, and play vital role in the creation of an ubiquitous environment. In these environments, huge amount of data could be collected and used to support decision making processes like logistics scheduling and planning.

Logistics within manufacturing sites like warehouse and production lines are rationalized by RFID technology so that materials' movements could be real-time tracked and visualized (Dai et al., 2012). The main application of RFID for item traceability and visibility is rudimentary. First of all, estimation of delivery time on manufacturing is crucial for the sales department when we get an order from a customer. That helps to ensure the shipping date, which has been estimated from past experiences and time studies. An estimation like this is not reasonable and practical given the difference of seasonal fluctuation and individual operators (e.g. peak seasons and off seasons). Secondly, RFID-enabled real-time manufacturing, scheduling and planning on production areas heavily relies on the arrival dates of materials, therefore, the decisions on logistics trajectory are very important. In a case for example, of a company that carries the decision using paper sheets manually it is common to be caused material delay. That has as a result many rescheduling and replanning, which greatly affect the production efficiency. Finally, the space on the manufacturing shop floor is restricted. As a result, the logistics trajectories of materials should be optimized. If, the procedures of logistics is not well-organized, high Work-In-Progress (WIP) inventory on manufacturing shop floors may be caused.

In order to address all the above issues, senior managers decide to explore a solution and proceed with the full use of RFID-enabled logistics BD. Unfortunately, they had to overcome several challenges. To begin with, manufacturing resources equipped with RFID tags are converted into smart manufacturing objects (SMOs). Every movement of these objects generate a very big amount of logistics data if you consider that SMOs have the ability to sense, interact, and affect each other to carry out logistics logics. The voluminous RFID-enabled logistics data closely relate to the complex operations on warehouses and production areas (Zhong et al., 2013). That leads to a great challenge for

further analysis and knowledge extraction. Secondly, the RFID-enabled logistics Big Data usually include some 'noise' such as incomplete, inaccurate, and redundant records, which could affect dramatically the quality and reliability of decisions. Thus, elimination of the redundancy is necessary (Zhong et al., 2013). However, current methods are not perfect for removing the above noises due to the high complexity and specific characteristics of RFID Big Data. Finally, mining frequent trajectory knowledge is significant for determining the logistics planning and layout of distribution facilities. However, the knowledge that can be extracted by RFID-enabled BD is sporadic. That means that a piece of information which indicated the detailed logic operations may be a result of hundreds of RFID records. It is very challenging for this creation to be achieved. In our research we propose a holistic BD approach to excavate the frequent trajectory from enormous RFID-enabled manufacturing data for supporting production logistics decision-makings. This approach comprises several key steps: warehousing able to deal with raw RFID data, cleansing mechanism for RFID BD, mining frequent patterns, as well as pattern visualization and interpretation.

3.9.1. Logistics operations within RFID-enabled ubiquitous manufacturing sites

In a RFID-enabled real-time ubiquitous manufacturing environment, logistics operations are rationalized and reengineered by SMOs. The upgraded operations could be briefly described as follows:

- In our case raw-materials are packaged with standard quantity for each batch, which is equipped with a RFID tag. An external logistics operator (ELO) uses a stationary reader to complete the binding process. After this process is completed, the RFID-labeled batches are delivered to the shop floor buffers, where the enter-in and enter-out movements are detected with the use of RFID devices.
- An internal logistics operator (ILO), on a shop floor, equipped with a mobile RFID reader to pick up the required materials and deliver them to a specific machine when he gets a logistics task. Using the mobile reader, machine operators and ILOs are able to execute the material hand over processing.
- After the materials are received, machine operators can carry on the processing. When the job is finished, an ELO is informed to move them to next processing stage using a mobile reader.

- At next processing stage, an ILO prepares a mobile reader to get the logistics jobs and moves the materials on the shop floor. The machine operators and ILOs proceed with the execution of the material handover over the mobile reader.
- The above mentioned steps are repeated until all the processing stages are fulfilled. The finished products will be delivered to the warehouse by an ELO, who uses a hand held RFID reader to fulfill the operations. In warehouse, a stationary reader deployed at finished products receiving area will be used for killing and recycling the tags.

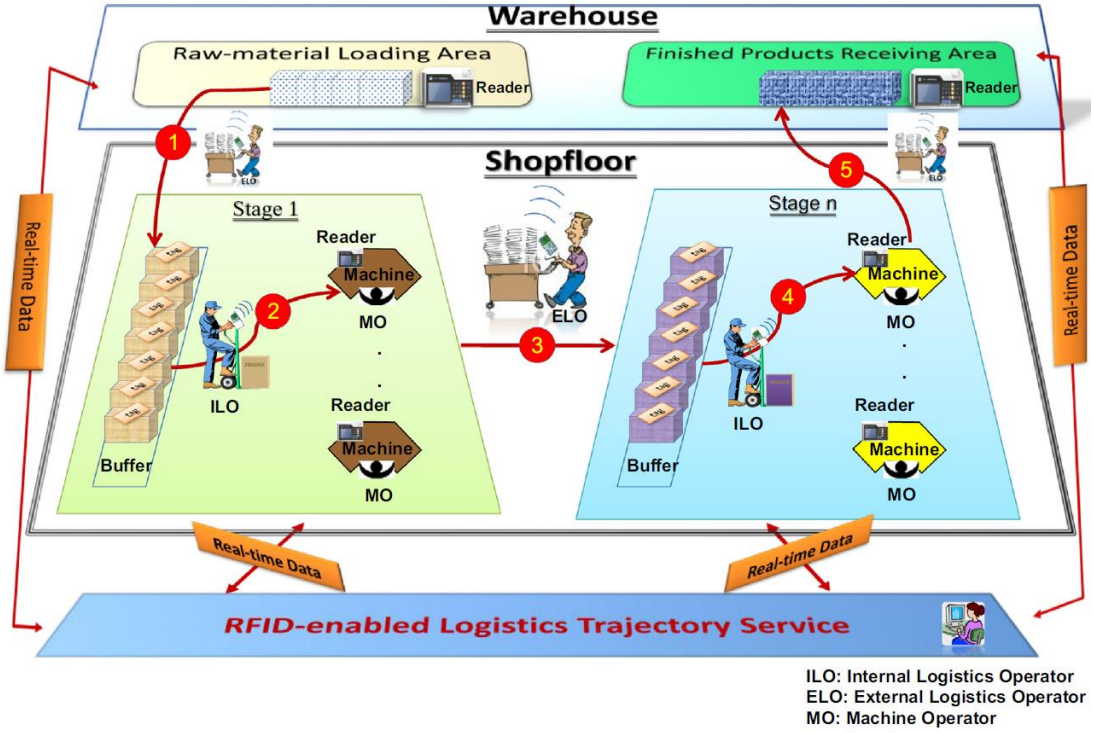

figure 12 RFID-enabled real-time logistics environment in manufacturing sites.

Source: (Zhong et al.,(2015, pp. 260-272))

4. Big Data and Supply Chain Management

4.1. Current research at the nexus of BDA and SCM

The gradual introduction of the internet in combination with the evolution in information technology (Bandyopadhyay et al., 2010), like storage, network, and telecommunication capabilities, is giving the chance to companies to have almost real time access to large amount of data, considering 'information as a strategic asset' (Mason-Jones and Towill, 1997, p. 140). It is normal to expect that recent technological advancements will continuously impact whole SCs, which become increasingly networked rather than the traditional linear setup, as a consequence adapting to the new information-centric production environment where 'information moves independently of product at internet

speeds' (Kuglin and Rosenbaum, 2001, p. 59). The emergence of the long proposed IoT (Ashton, 2009) indicates the beginning of this development as technologies and communication solutions are more and more integrated, from independent devices in a network to an intelligent object network in which the virtual and physical worlds interact. In this following technology evolution, also referred to as the fourth industrial revolution or 'industry 4.0','cyber-physical systems' are expected to propel the amount of data generated by and available to companies to unexpected new levels (Lee, 2008). This attempt make sense of this intangible growing mass of data, as we refer in a previous part, is known as BDA. Basically, analytics describes the application of advanced statistics to historical stored data, with main target to identify behavioral patterns which eventually enable the forecasting of future behavior to some extent (Shmueli and Koppius, 2011). The 'predictive' nature of BDA may probably represent a game changing advantage for SCs. As a result, companies may achieve cost advantages by applying BDA as unexpected equipment downtimes can be reduced significantly, allowing the companies to reduce buffer inventories. In this way partners are enabled to operate a leaner SC while eliminating supply risks. It is, however, imperative that information on potential unexpected malfunctions is shared with SC partners to allow timely mitigation actions across the chain.

As found in an extensive MIT research on BDAs comprising 3,000 business executives (La Valle et al., 2011), companies have to be able to analyze and manage the always growing amount of data in order to extract the relevant pieces and to feed that extracted information into their decision-making process. In turn, not being able to mine the data available and thus not having access to accurate, actual, and meaningful information represents a risk for companies and subsequently for the SC, as decisions need to be made on a reliable and evidence-driven basis (Ross et al., 2013). This holds especially true in regard to SCM, which in a very high level depends on the availability of up-to-date and accurate information for business execution (Gunasekaran and Ngai, 2004). Accordingly, the value and importance of information for effective and efficient SCM has been highlighted in detail in SCM research, most characteristically in regard to the information sharing (Fosso Wamba et al., 2015; Li and Lin, 2006).

However, given the steady growth of potentially relevant information, data from GPS sensors or customer data collected through mobile devices for instance, and the corresponding challenges to identify the most valuable and important items, it is

surprising that research on risks tied to and stemming from the use of information in SCs is scarce (Kache, 2015). Especially, the issue that the lack of up-to-date and 'right' information, although being an essential characteristic for the application of BDA, may pose a risk to the SC has gained little attention in the academic SC landscape, as identified in a literature review study by Kache and Seuring (2014).

Despite all the above, because of the novelty of the concept of BDA, the management research community is facing difficulties to grasp the value of this concept as other similar information management concepts exist, such as Business Intelligence (BI), Business Analytics (BA), or master data management (Chae and Olson, 2013; Otto et al., 2009). Nevertheless, from the authors' point of view, the concept of BDA, despite it has undisputed similarities to other concepts, as the analysis and assessment of data or its predictive touch for example, is different. Portraying a holistic approach, BDA rather supplements the existing concepts, offering an extension to the scope of these information management concepts. The supplementing, evolutionary character of BDA can be described along with two main dimensions, which Kache and Seuring (2017) labeled the 'origin of information' as the first dimension and the 'type of information' as the second dimension. The interaction of these two dimensions is displayed in Figure 14 below.

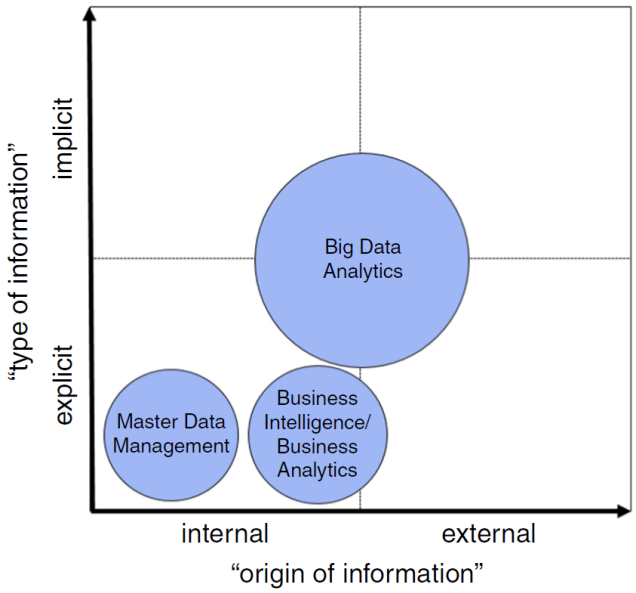

figure 13 Application of information management concepts based on the "origin of information" and "type of information" dimensions. Source (Kache and Seuring (2017))

Concerning the 'origin of information' dimensions describes if information stems from the internal or the external environment of the company. Many information management systems like master data management systems operate entirely on internal company-

related business information, but are usually susceptible to a lack of transparency (Otto et al., 2009). BA and BI systems could extend toward the utilization of external information like supplier inventory information, but are often facing problems coming from complexity issues and low degrees of cross-boundary standardization (Chae and Olson, 2013). The emergence of BDA represents an example shift, enabling better exploitation of the always increasing amount of data, as these systems are designed with a focus to include all information sources available, regardless if the information sourcing from inside or outside the central business environment.

The dimension of 'type of information' can be differentiated into implicit and explicit information. Manifest or explicit information is absolutely structured. Examples could be transaction data or inventory forecasts, where the information value is apparent from the beginning. On the other hand, implicit information, like social media content or equipment sensor readings, is rather intangible and latent. Therefore, it may not appear at first glance what the value and purpose of collecting the information is. The tangible value can be hidden and may only be operationalized through the application of statistical models, which enable the identification of patterns, as for instance machinery malfunction indicators. From Kache and Seuring (2017) point of view, the large-scale utilization of the structured, semi-structured and unstructured types of data, makes clear and implicit, without taking into consideration either the data were generated from the internal environment of the enterprise or the external environment like clients and customers, underlines the importance of the concept of BDA.

The increasing interest of BDA from the research community is clearly noticed by the always increasing number of SCM focused researches and conferences explicitly calling for research on the field of BDA. A Google Scholar literature search that uses as keywords the words 'supply chain'. 'big data' and 'analytics' indicate us that there are only few scholarly journal articles that takes into consideration BD and Analytics from a SCM perspective as it is stated by Rozados and Tjahjono (2014)(e.g. Davenport, 2006; Trkman et al., 2010; LaValle et al., 2011; de Oliveira et al., 2012; Waller and Fawcett, 2013; Hazen et al., 2014; Chae, 2015; Guo et al., 2015; Harris et al., 2015; Opresnik and Taisch, 2015; Tan et al., 2015; Fosso Wamba et al., 2015; Zhong et al., 2015). The relevance of these two topic is understood by the always increasing number of publication the recent years. As the body of literature is still limited, we summarize the exemplary researches mainly following the timeline of publication.

4.2. Main contributions so far to BDA and SCM

Davenport in 2006, write an article in which he built a case for leveraging analytics as a competitive advantage. In that research he concluded that the systematic collection, analysis, and action on data is not only beneficial on the corporate level but is also a key advantage to optimize from all the SC structure. This already points to BDA as a topic although without referencing the term.

Applying case research, the articles by Trkman et al. (2010) and de Oliveira et al. (2012) investigate both the advantages of analytics capabilities on SC's total performance. The work thereby adds to a better understanding of focused analytics application in a SC setup, laying the foundation for the applicability of BDA.

LaValle et al. (2011) are highlighting the importance and the need for enterprises to develop a data-driven business environment. By providing case examples, the authors underline the how import data analytics are for corporate and SC decision making.

From their side, Waller and Fawcett (2013), in a paper, they focus on the general applicability of data science, predictive analytics, and BD in combination with SCM. They present a range of practical applications, and identify potential further research areas understanding the need that more research is needed around the field of BD and SCM.

A research done by Hazen et al. (2014), investigates how important the data quality is for data analytics capabilities like BD and predictive analytics. As far as SCM is concerned, the authors notice that, as data are becoming more and more complex, a good control and monitoring of the data quality is the key for efficient, effective and successful decision making. By highlighting the relevance of the research they already have done, they promote a co-operative interdisciplinary approach that involves also IT experts and SCM professionals, in order to present and analyze all the challenges of data quality in a corporate setting with aim to achieve the best possible levels of data quality and also data integrity.

Chae in 2015 makes a proposal for a new analytical framework designed for the assessment of social media use in a SC setup. Relying on #Hashtags coming from Twitter and Instagram, the author studies how the social media posts can help enterprises to

handle the increasing demand in a better way, while simultaneously be able to extract valuable customer insights useful for the research and the development of new products. Although the presented approach is an interesting application of BD in a SC context that lacks a wider conceptual approach.

After Combining RFID technology and cloud computing technology, Guo et al. (2015) present an intelligent decision support system architecture for production scheduling and monitoring in labor-intensive distributed manufacturing environment. In this research Guo et al. address a key benefit of linking BDA and SCM, as the novel architecture approach showcases how the systematic integration of data-driven decision making in production and logistics operations increases information visibility and transparency in the whole SC.

In the article written by Harris et al. (2015), it is studied the positive impact of using communication and information technologies, such as BDA, to optimize for example the freight transport SC. About linking aspects like multimodal operations, barriers for technology adoption, and technological trends, authors provide a very interesting debate on how technological innovations can affect the SC integration. As far as data that can be used for managerial decision making may be well originating beyond the single enterprise's boundary, the SC integration perspective is the key factor to fully leverage the benefits of a data-driven SC ecosystem. Opresnik and Taisch (2015) present the advantages of BD to the business strategy of servitization. Designed to be capable to be applied in manufacturing industries, their proposed strategic framework underlines and sets in the center of discussion how organizations can increase the value of their portfolio by adding information as the third pillar together with the traditional aspects of 'product' and 'service'. Despite this, they do not explore further the association with the SC.

Tan et al. (2015) highlight the important role of BD as a source to achieve SC innovation. Being a very interesting contribution in research in the combination of BDA and SCM, the proposed analytic model approach gives the ability to enterprises to improve their SC innovative advantages against competitors with SC partners by making extensive and regular use of the insights from BDA. In an environment where the competition is getting stronger and stronger and it is also fast-paced, this research is promoting extension of the innovation capabilities into the SC seem like a promising strategy, considering that very

important knowledge and information may be available and unprocessed outside of the each organization.

Presenting findings from a systematic review and a longitudinal case study, the research by Fosso Wamba et al. (2015) provides a comprehensive overview of a range of BD concepts. On a more granular level, the authors assess how the generation of business value is supported by application of BD concepts where a key focus is on BD strategy as well as BD implementation. In this way, the paper prepares the ground for future research at the intersection of BDA and SCM. However, some crucial aspects such as the financial implications of BDA adoption in SCM or the potential of BDA in regard to enhancing the innovation and product design capabilities across the SC were not touched upon by the research.

Addressing the challenges connected with the voluminous amount of data, a major problem in modern manufacturing, Zhong et al. (2015) present an integrated approach for logistics management optimization on the shop floor level. Based on the smart manufacturing concept in the automotive industry, which uses production resources equipped with RFID facilities, they develop a holistic BD system combining a range of technical innovations so as to identify frequent patterns from the amount of RFID-enabled shop floor logistics data. Mining the previously invaluable RFID-enabled data streams, the paper thereby presents an interesting conceptual solution, which supports the identification and extraction of hidden information that may be relevant in order to make more informed decisions in logistics and production control.

4.3. Big data storage in SCM

As mentioned before, there are various forms and ways to store the vast amount of the big data generated from software and sensors in the industry of SC. The classically known rational database management system is a structured modelled of auxiliary systems consisting of servers and database management systems used for data storage and look up, analyze, manage and store the vast amount of data generated in a SC. In addition, BD storage accounts for the effective management and storage of voluminous data-clusters in a manner that is adding value, that is, in a more real-time accessible way and more reliable. The use of data-cluster storage systems or equipment enables the Direct Attached Storage (DAS). In DAS, different hard-disks and hard-drives are attached to the DBMAS and the Network Storage (NS) directly and come in two forms, the Network Attached

Storage (NAS) and Storage Area Network (SAN). NAS and SAN are both the network storage data-cluster storage systems that are both directly attached to the network enabling a unified network platform enabling real-time sharing and access to the data. SAN network storage is more independent network storage system that provides much better scalable and bandwidth data frequency accessibility. An example of SAN network storage system is the cloud-computing storage system (Addo-Tenkorang and Helo, 2016).

4.4. Decision Making

Manager's use to discuss the changing nature of how organizations make decisions and what is the role of BD in decision-making capabilities. Traditional decision making may have focused intensely on relationships. BD gives to the organizations the ability to move toward a more informed decision-making process where information supplants relationships.

With the use of BD, decision making still incorporates an important relational element that is called 'communication'. Being able to communicate organizational analysis results is of critical importance to sound decision making. In this point of view, human capital intersects with the decision making process as organizations look to specialized employees who can analyze and disseminate BD to key decision makers, allowing those specialists to combine key capabilities of analytics and decision making.

In (Richey Jr et al., (2016)) it is found that respondents view forecasting as a primary opportunity for the future. BD provides a way to evaluate an expanded set of projections and proceed with decisions that are strongly informed. A part of managers suggested that this improved forecasting may expand the organization's ability to address risks associated with decision making.

Human capital

Organizations working on the implementation of BD mentioned that management's mentality can limit their success. Quantitative analysis provides new information, but implementing ideas based on that information is limited when decision makers place greater emphasis on relationships than on quantitative data results. Employees willing to adopt BD and the BDA capabilities are in demand.

The lack of management support may have as a result delayed implementation of BD generated solutions. Despite this, at a lower level it is also necessary to find solutions to implement BD findings. Companies have to create actionable plans from BD and hire specially trained personnel who can implement them.

Hiring personnel throughout the organization that has the needed background about BD also benefits companies because they can monitor and improve the data collection process.

Risk and security governance

Numerous US SC managers indicate the importance of security governance, mentioning that 'security is huge' and 'since BD contain a voluminous amount of information, security is very important' .Moreover, managers find BD security to be a main concern, particularly as it is connected with the data ownership, data storage, accessibility, and data privileges, such as who can access data.

Despite the perceived risks, many managers feel sharing BD-related information provides substantial benefits to the members within the SC. More accurately, as it is stated in (Richey Jr et al., (2016)), German managers conveyed that even when organizations bears the greatest burden or risk of sharing information along the SC, the benefits outweigh the costs. Even though the challenge remains to find a balance between risk and potential outcome.

Although positive feelings toward BD sharing are commonly held, respondents also feel that some information should not be disclosed. For instance, several respondents propose that pricing information, even across SC partnerships, should remain confidential.

In any aspect, there are information that must remain undisclosed in order to protect an organization's solvency. While directors and managers indicate organizations keep information confidential for financial reasons or even as a matter of principle and privacy, others suggest that legal mandates and regulations prohibit the disclosure of specific information. However, some is required by law to be kept confidential, and that would be the first priority for any manager, but after that, it is also anybody's guess if you are really able to get that data or not. For instance, maintaining confidentiality is vital in specific industries, such as healthcare.

Other issues has to do with the problem about where to safely store data, and whether modern storage systems, such as cloud services, or private, internal systems offer sufficient security for storing information. Generally, US managers understand the inherent risks if they do not stay 'above board' with data privacy and security.

In other developed nations the opinion around sharing views similar to the US managers' opinions regarding data security. However, it appears that, more than anything else, stricter legal regulations influence managers' decision making regarding BD security in the SC in both developed countries and emerging markets.

As it is generally suggested, managers emphasize the legislation associated with data sharing more than the morality of disclosure. However, other problems arise over data ownership. Emerging economies, however, seem to be more concerned with privacy and protecting customer information. Among these countries, the general sentiment about sharing BD information across the SC was that, "it's strictly a no, no" (Richey Jr et al., (2016)). Specifically regarding information collected at the customer level, they were totally negative on sharing them. It is against personal privacy to use customers' data for organizational benefit without permission of the customers. Government regulations and social norms further reinforced and support this concern for privacy and security:

Big attention about data security is given by Chinese managers, ensuring that they limit data privileges to those who should be authorized to access to it, and only to portions of data relevant to given individuals. Furthermore, Chinese managers were largely reluctant to reveal broad types or examples of withheld or protected information, citing transparency and ethical reasons as obstacles.

Contrary to many other managers' views, some managers do not characterize BD use to be risky at all. For example, from a manager from Turkish it was revealed that his/her organization was extremely amenable to sharing any kind of information if the main achievement were to improve the organization's efficiency and make managers' and customers' life easier. In addition security or governance concerns are not mentioned. Generally it is indicated that potential opportunities and advantages derived from BD are of vital important, and it is believed that security requirements might actually make it even more challenging to achieve those benefits.

Furthermore, although some non-specified issued exists concerning what data can and cannot be shared, many managers believe that BD decreases risks associated with decision making. Not only is top leadership expected to make more knowledgeable decisions after mining BD, but respondents see the resulting information as internally empowering to the entire organization because it decreases risks associated with ill-informed decision making. Likewise, from other points of view, BD is considered as a potential tool for improving risk management efforts, which presently are difficult to manage thoroughly.

Storage

If we consider the aspect that nothing is invaluable, as it supported, then SC managers face the obstacle of storing enormous big amounts of data until it becomes useful, highlighting the need for adequate hardware to enable BD use. The value appears to be in recognizing that while the voluminous data undoubtedly provides greater opportunities for discovering more information about customers and SC partners than ever before. Data storage is one of the primary challenges resulting from BD accumulation. Several related logistical issues also arise because of the challenges associated with storing that huge volume of available data.

Intense is the controversy that raise the issue of long-term data storage against the short-term data storage. A part of managers' struggle with identifying and then filtering currently meaningful information without discarding information that may prove useful later. In addition, because the information's long-term value is unknown, managers indicate that they face another concern and challenge in determining what should be kept and what should be discarded.

SC managers in both emerging and developed economies express concerns over protection, the costs, and infrastructure needed for efficient data storage. Investment in disaster insurance is one example of additional financial burdens BD storage creates. Anxieties concerning capacity and the system retain-ability are other frequently cited concerns regarding BD storage. More accurately, organizations have to decide which data to retain at both macro and micro levels. The storage issue is expected to be more and more of a concern as firms lack physical space and means to store all information. It is

very important to know what to do with this data and how much of it to use, and also how much to invest in technology and to turn all of this into information.

In many industries that experience poor systems integration, in healthcare for example, managers acknowledge that more needs to be done with the collected and stored data. Nevertheless, few managers could identify strategies that might help make progress on this goal. A significant challenge is to know what to do with this great new tool called BD, because there is seemingly infinite information stored.

Another point of view indicates that innovation arises from little data as opposed to BD, and therefore when we start accepting information, we should be able to figure out what is little data and what is BD, what is useful and what is not useful. Therefore, what is the way in which we can discard the data, because some of it is not still being useful? It is possible not to be useful forever. Regardless of managers' beliefs pertaining to little versus BD, nearly every manager who discussed storage indicated that the ultimate goal involves converting stored data into meaningful information that can be used to improve decision making.

Operational efficiency

Decision makers see BD as useful and quite important in making efficiency in the organization. One advantage involves the ability to use information gathered to improve various SC partners' efficiency, accuracy, and collaboration. Managers in the manufacturing and transportation industries suggest that a best practice links customers and suppliers directly into a company's BD driven systems to increase efficiencies. In addition, they present other advantages resulting from integrating BD into firm systems.

In accurate, that kind of information improves organizations' efficiencies by improving decision making and inventory management. An example involves just-in-time inventory systems that, when properly implemented, span the SC and produce cost savings and optimum inventory levels. Decision makers indicate such a system reduces stock-out costs and associated opportunity costs, and allows for SC partners to track other partners' stock and capacity. A challenge for managers, however, is balancing the cost of BD management systems with the potential gains in efficiencies and performance.

Partner transparency

Managers also discussed transparency among SC partners as a desired communication tool. Increased transparency enhances firms to communicate with each other, and this facilitates the BD sharing.

Transparency is considered to be a way to gain SC partners' trust because the partners feel they can rely on the shared data because its provenance is clear. It is believed also that transparency is proactive. Companies that embrace transparency with SC partners often do so with specific purpose to realize particular value (Lamming et al., 2004).

They also support a move toward transparency as part of the rules of doing business within a modern SC (New, 2015). Researchers contend that companies have to be transparent in their SC, or else others will expose the lack of transparency to force transparency on the firm (New, 2010).

Other notable issues

Managers mention the following issues less frequently, but we consider them relevant given both the nature of BD and the detail and depth in which they discussed them.

One version of the truth

The fact that the accuracy and the quality of the data are frequently changing, and several different people manage input into various systems both present primary concerns to managers. More specific, these factors all contribute to the issue of determining the one side of the truth, a challenge faced by industrialized and emerging economies alike (see Dutta and Bose, 2015).

Many managers agree that they face a substantial challenge resulting from the lack of integrated systems across SC partners. In addition, information input manually as well as system disruptions and failures present additional room for error. A manager in the field of sporting goods industry argues strongly for the importance of a single truth.

For example managers from Germany, India and Turkey advocated perhaps most strongly for a unified system, specifically to preserve the integrity of information so that it retains its usefulness for decision making. They suggested that being able to access BD information seamlessly, from any location, and at any time, would improve existing systems. While no clear standards exist for determining the best system to implement,

several managers suggest the need for technology to simplify BD usability, both internally and externally. While some managers believe their organizations should be responsible for developing integrated systems, while others suggest that third-party providers must invest in developing superior data analysis tools which could be used more universally to create one truth.

What is more, while high variability and volume of data are nearly inevitable, a manager indicated that while it is important to gather as much data as possible, the quality of data is also important. Consequently, managers call for the cleansing of noisy databases which lack the sophistication to be truly functional, and for limiting the overall number of systems for quality assurance. The importance of such system integration could not be overstated, as decision makers use the information resulting from the pipeline of systems, so, it is needed to have clear data that can be accurately analyzed, summarized and distributed.(Richey Jr et al., (2016))

Climbing the learning curve

As firms begin to use BD, they face many challenges. Some managers have issues concerning poor guidance. From managers it is noted that no road map exists on how to use BD, and that all they know is that it is important. Other managers are still learning to deal with the complexity of sharing their findings.

Managers also point out that while BD might be a fashionable buzzword, not all firms are equipped to begin using it. In some cases, firms might lack technology. In other cases, the problem might be a lack of knowledgeable scientists available to analyze the data. Regardless of the reasoning, other issues may need to be addressed first, requiring influence of one capability over another, before BD can take place effectively.

The enormous volume of information that BD generates can create storage issues. One of the learning challenges is determining the appropriate tools needed to analyze varying amounts and types of data. Understanding which data are useful presents a formidable challenge for SC professionals wading into BD.

Customer orientation

Some managers suggest that BD applications improve the ability to assess customer perspectives, needs and, increase performance, and ultimately to improve customer

service. SC researchers have discussed that one of the future uses of BD could be for developing and customizing product offerings for customers, an aspect that the retail sector is always becoming more and more concerned about (Bhakoo et al., 2015, p. 395). Optimizing profits or increasing efficiencies is certainly a goal of a BD centered strategy. It is also suggested that by improving the customer relationship it is possible to increase the bottom line that involves.

Additionally, there is an opportunity to obtain data to help ‘capture the customers' perspective’ , predict customers’ future needs, and provide more targeted insights and services to them. For instance, using CRM packages and integrating them into daily operational systems increases the ability to predict and provide for customer needs. Nevertheless, it also allows organizations to be more proactive and innovative by concurrently addressing other market conditions such as times of economic downturn, tracked within the system. To that point of view, one director promotes the importance of an agile and self-correcting system because it constantly evolves and the system can recognize an issue by itself and solve the issue then for you, thereby decreasing service failures. As a manager in the agriculture industry suggests, it comes down to helping people manage their day-to-day better. Also, operations and performance are ideally improved for the firm day after day. Managers also propose that BD usage will more directly and positively affect customer-focused and customer-facing organizations.

4.5. SCM Big Data and the 3 main Vs

A full identification of data sources used in the business cases and methods for successful application obtained from the systematic review producing a list of 52 mainstream sources of BD across the SC. Each one of these sources is reported at least in one or more of the SCM four levers, with a level of incidence from 0, that indicate that does not appear in that lever, to 4, which means that it is core for processes at that level. In the same way, each data source was classified according to its reported volume and velocity in a scale from 0 to 4. Variety of data was described in a 3-level classification: Structured, Semi-Structured or Unstructured. However, these three subcategories are statistically dependent in the scores of a given data source, in order to facilitate analysis of some patterns of interest, they are reported separately.

In the following *Figure 15* it is shown the average volume and velocity versus the variety of the data sources in a model such as $D(Z \mid X) = f(X, s)$ with Z=0.5*(Volume+Velocity)* and *X=Variety*.

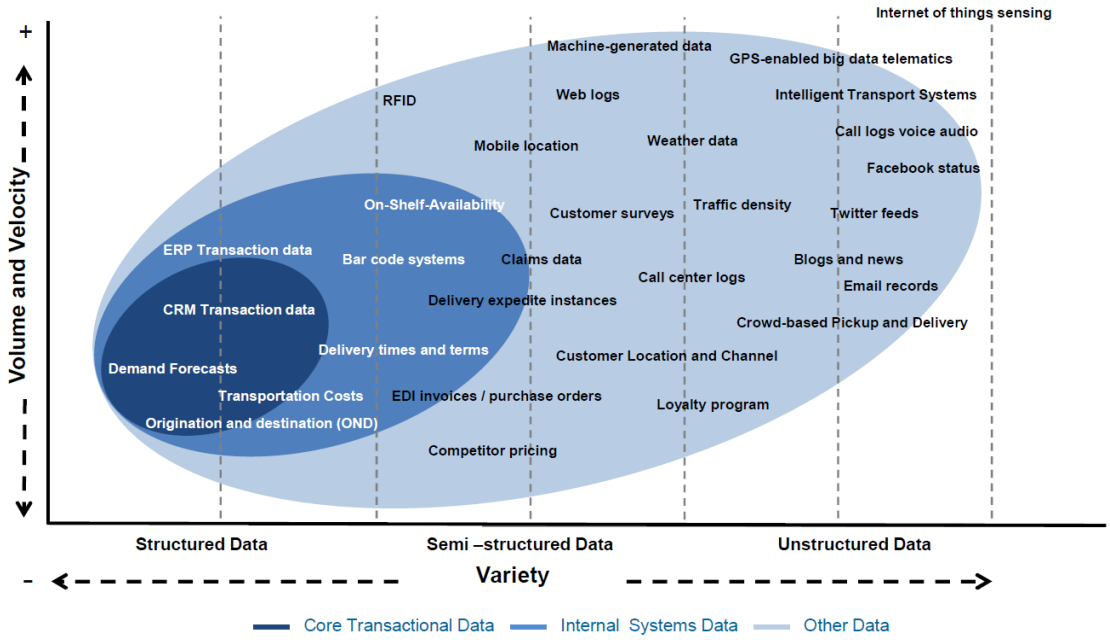

figure 14 SCM Data Volume and Velocity vs. Variety. Source (Rozados and Tjiahjono, 2014)

In the three shaded areas are included data sources that are considered as core transactional data, internal systems data or others, respectively. The frontier of these three areas has a much wider horizon when moving horizontally, along the variety of formats than when moving vertically on the other two dimensions. If the model *D* above is a linear regression, all parameters in vector *s* are strictly positive. Specifically means, that fact relates to a positive correlation between velocity of information and larger volumes in unstructured formats. This proposition is endorsed by many practitioners and researchers, and although there is no previous conclusive quantitative analysis, it is considered as rule of thumb that 80% of usable business information is unstructured (Roberts, 2010). The authentication of that direction in SCM has unambiguous implications in the approach to data management for BDA. Although transactional data in relational databases coming from various information systems, such as WMS, ERP, CRM or SRM remain the main source of internal information and have considerable high volumes, they are relatively a small part of the total data sources available for use. Consist just the 8 out of 52 in the taxonomy.

We can notice in the figure that a high concentration of points is located at the top right. That is because most of the customer interface data platforms are in this high volume of unstructured data region: social media, online surveys or mobile location devices. E-mail data is also another one example. E-mails are massively used nowadays as the primary communication and information tool, but, rarely used for analysis, when it certainly provides unstructured data as feedback about experiences with clients or suppliers (Ordenes et al., 2014).

SCM BD sources are mainly generated in unstructured formats as a result to be difficult to analyze with traditional IT tools. Although data management concentrated on expanding volume and velocity capabilities for transactional data, the number of core transactional data sources is relatively small. There is a contrast in SCM data sources between the relatively smaller variations of volume and speed against the larger ones in data variety, and a positive correlation between the unstructured formats and high volume and velocity.

4.6. Four levers in the Big Data Driven Supply Chain

BDA can be useful and operate across all SCM levers, conveying information from one area to another but the aggregation requires accuracy, completeness, consistency and timeliness (Hazen et al., 2014). For example, in the marketing area BDA captures and tracks demand through Point of Sale (PoS) data, transportation generates records from GPS devices, RFID data identifies goods in stock and electronic data interchange sends automatic buying orders to suppliers.

Marketing is transforming customer knowledge into an agile system that sends big volume of data flowing upstream in the chain (Jüttner et al., 2010). The close relationship with customers is achieved through the adoption of increasingly sophisticated methods that are used for analyzing customer data, and at this point of view, data sources that consist of social media, mobile applications, or loyalty programs can be found. All of these data sources are the enablers for the *sentiment analysis*. In a similar manner, recording omni-channel sales information can be facilitated by the cloud and electronic PoS, and by machine generated data that keep record of the transactions. According to Butner (2008) customer inputs should be better aligned to SCM systems, and that SC

managers tend to focus more on their suppliers than to their customers, but he also reflected that technology has made it more possible to be done than ever to access, analyze and understand customer data, as BD enables sensing of social behavior (Shmueli et al., 2014).

Procurement deals with the relationships at the upstream SC. Complexity issues with data, on this side might coming from globalized purchasing strategies with numerous transactions. On this level, a strong bond with the internal finance reporting led to the adoption of measures on spend visibility data, with main target the achievement of granular levels on aggregated procurement patterns. However, according to Ainsworth (2014), data on external expenditure are "often backward looking, often inconsistently categorized and not integrated with internal costs". In addition, could be more than 50% of a company's cost. A subgroup of data that is still to be fully integrated and appears in the taxonomy as semi-structured are the business documents that consists of purchase orders, shipping notices and invoices, could be sent through the EDI. Still et al. (2011) concluded that the procurement needs to activate the data sources not only for spending data management process, but also for the entire procurement function.

Warehouse management, specifically inventory management, has been dramatically developed after the successful introduction of RFID. In this dataset, the largest amount of data is related to an automated sensing capability, especially as the IoT and extended sensors, connectivity and intelligence to packaging systems applications and material handling evolved. Position sensors for providing information for on-shelf availability share space with traditionally SKU levels and BOMs.

Transportation analysis applying Operational Research (OR) models has been widely used for location, network design or vehicle routing using as input data origin and destination (OND), transportation costs or logistics network topology as 'static' data, as stated by Crainic and Laporte (1997). New alternatives to enable management and coordination in real time using operational data rely on direct and mobile sensing over shipments that are integrated into in-transit inventory, estimated lead times taking into consideration traffic conditions, weather variables, real time marginal cost for alterative channels, intelligent transportation systems or crowd-based delivery networks among sources of BD. A detailed analysis of the 3Vs in transportation data revealed to be the tool with proportionally higher speeds in data transition.

4.7. Issues and challenges of adopting and practicing BDA

4.7.1. Organizational challenges

Time-consuming: The initiation of Predictive analytics is time demanding and includes many stages of testing adapting and developing it to different cases (Blackburn et al., 2015). Bringing together experts from various fields with varied mindsets will be a challenging task. In complex systems such as the SC, BDA implementation requires consistent support from top management and key stakeholders, as it might need from 12 to 18 months to see catchable results. Getting access to data, that are collected and owned by different departments from inside the organization, validating, combining, and data cleansing will be a dull and monotonous process which requires exhaustive dedication from the project management team.

Insufficient resources: The data and analytics resource capabilities vary across organizations in a SC network. SC partners' do not have the needed IT infrastructure and capability to share data and information in real-time and discrepancies are caused. As stated in Dutta and Bose (2015), co-operation and cross-functional team formation between the stakeholders within an organization should be a priority for implementation of BD. Nevertheless, while forming an inter-organizational cross-functional team, the challenges that can be expected are competition within the SC network, incentives arrangements, principle-agent conflicts, data sharing policies, etc. Data-driven culture, that is fundamental BDA capabilities, and fact-based management, need to be encouraged across SC network as a strategy for effective and efficient utilization of BDA, and to create business value. Leadership also plays a critical role in successful application of BDA systems for SCM (Seah et al., 2010).

Privacy and security concerns: BD possess many concerns such as security, privacy, unethical use of BD and inefficient data processing (Hu et al., 2014), which would result in biased findings (Tien, 2012). SC professionals raised concern regarding privacy and data security, and support that out-dated regulations are one of the biggest obstacles that they should overcome in data sharing, especially consumer data (Richey et al., 2016). Privacy, security and data legislation could be of serious concern for multinational SCs, obligated to abide by the laws of different countries while sharing data across SCs (Alfaro

et al., 2015). Nevertheless, these challenges could be overcome by employing effective data governance initiative within the process of data integration and management.

Behavioral issues: From the perspective of the behavioral issues, the use of real-time information and data could be challenging because decision makers may excessively react to even small changes in the physical world which would make worst the 'bullwhip effect' and increase SC risk and cost of inventory (Tachizawa et al., 2015). Tachizawa et al. (2015) argued that concerning BD there is a risk of identifying many statistically significant but irrelevant correlations that do not have a causal linkage. BD focuses mainly on correlation but not on causation, which necessitates human inferences to solve important problems. However, Google has proposed a heuristic way to solve the correlation and causality problems (Radke and Tseng, 2015).

Issues with Return on Investment (ROI): Not clear advantages and ambiguity on ROI make stakeholders apprehensive about the implementation of BDA (Richey et al., 2016; Sanders, 2016). Achieving financial benefits from BDA is challenging too as it depends mainly on the 'downstream' employees who performs the task (Davenport et al., 2001). For example, analytics can help segment the market based on available data, but it is the sales and marketing team who has to believe in the data-driven knowledge and treat customers based on the segment types to make real change. In our scenario, Data-driven culture and the employees' absorptive capacity at individual level plays' a very important role in absorbing and assimilating the knowledge.

Lack of skills: Schoenherr and Speier-Pero (2015) through a survey identified potential barriers to using predictive analytics in SCM. The first and most important barrier is the inexperienced employees, then follows time constraints, lack of integration, lack of appropriate predictive analytics solution, and issues with change management. Moreover, it was found out that professionals who plan to use analytics in future and currently not using it have considered the lack of data and inability to identify suitable data as a prominent barrier, which relates to the situation of 'Data Poor and Information Poor'. Waller and Fawcett (2013) argued that data scientist requires a combination of both analytical skills and domain knowledge, which is difficult to find such combination as someone good in analytical skills may not be interested in learning domain knowledge. A recent study has also confirmed that lack of experts in BDA is a serious problem among SC professional (Richey et al., 2016).

4.7.2. Technical challenges

Data scalability: For the adoption of BDA Richey et al. (2016) identify data scalability as a major technical issue. Firms have to discard their data after a specific period of time so as to store newly generated data. Replacing relational databases which are limited regarding scalability with more advanced infrastructure such as Hadoop distributed databases, distributed file systems, cloud computing and parallel computing capability could be considered to tackle scalability issues. NoSQL database which has a high level of scalability is a better choice to deal with unstructured data generated from IoT data sources (Kang et al., 2016). Nevertheless, leveraging cloud-computing capability to store BD could incur more financial burden to companies as with increase generation of BD cloud storage utilization cost will also eventually increase. To avoid this, organizations could adopt strategies to optimize data collection process and reduce unwanted data generation right from the source (Rehman et al., 2016).

Data Quality: SC managers rely on data-driven insights for various reasons like for example to gain visibility, monitoring, process control, optimization, collaboration, etc. and, ultimately aim to obtain competitive advantage (Davenport, 2006; Hazen et al., 2014). Despite this, there are quality issues associated with the process of data production, which is often compared with the product manufacturing process (Hazen et al., 2014; Wang, 1998; Wang et al., 1995; Wang and Strong, 1996). Hazen et al. (2014) stated that poor data quality would hinder the data analytics activities and affect management decisions. Contrary to a physical product, data has the characteristics to be intangible in nature and measuring the data quality is a problem with many dimensions (Hazen et al., 2014). Hazen et al. (2014) discussed the research done by Wang and Strong (1996) and Lee et al. (2002), who classified the dimensions of data quality into intrinsic (accuracy, timeliness, consistency, and completeness) and contextual (relevancy, value-added, quantity, believe-ability, accessibility, and reputation of the data) dimensions. Concerns with the trustworthiness of the social media and web scraped data are also raised (Tan et al., 2015). The efficiency of the physical flow of material can be determined by the infrastructure quality such as transportation system, ports, technology, etc. (Bagchi et al., 2014). Thus, for effective flow of information, it is critical to have advanced data infrastructures and best practices of data management using techniques and tools like Hadoop, MapReduce and statistical process control.

Lack of techniques and procedures: Lack of quality data is not the only problem. There is an incapability of techniques to exploit the voluminous data properly. For example, in the case of demand forecasting techniques, significant attention is given specifically to endogenous time-serious variables for demand forecasting, and there is a lack of consideration of exogenous variables and information sources (Meixell and Wu, 2001). This evidently has inferences to develop better data management capabilities and methodology for forecasting demand to enhance SC operations. In Addition, there is a difficulty in considering expert judgement as a covariate in forecasting models. Furthermore, ordinal scales are used to measure opinions and experts' judgement. While, in practice, the available modelling approach of dependent and independent variables are intended for continuous or natural data as input. As discussed in Blackburn et al. (2015), very few studies like Fildes et al. (2009) have reflected on using expert judgement to increase the accuracy of SC forecasting.

4.8. Potential Benefits of BDA in SCM

According to a study done by (Capgemini Deutschland Holding GmbH, 2014), he identifies the communication between the machines, machine-to machine communication (M2M), as the issue with the most gained importance in the last year. M2M data is coming from sensors attached on products or machines, which are able to collect different types of data like positioning data, temperature, and usage data. As it is stated in a survey conducted in 2013, 11% of the requested companies state that they already use data sourcing from machine-to-machine communications. It is significant to notice that this percentage was increased in 2014 to 23%. In addition, 12% of the requested firms are in the status of implementing M2M technology and another 13% is planning to implement. The communication between the machines, enables automatic information exchange between different objects, such as production machines, cameras, containers, transport vehicles and their corresponding database. Possible use cases can contain activities like monitoring machines and areas, increasing the safety and the maintenance of facilities and even an automatic ordering system if demand is recognized. In the case of using the automatic ordering system, it can be achieved to be fully automated, up to a self-distribution of necessary goods. Regarding to the above mentioned classification, M2M communication will enable new business models and has the advantage to increase operational efficiency to very high levels.

The goal of the Smart Reusable Transport Items (smaRTI) research project that is mentioned in (Effizienzcluster Management GmbH, (2014)), was to increase the intelligence of the material flow. Hence, the identification and localization of handling units, like palettes, is achieved through the usage of Auto-ID technologies. A wide spreading of read points, like RFID or barcode reader, achieves increased transparency of freight deliveries between companies. When a handling unit is detected by the readers, the read points are proceeding with the generation of events, which are available in real-time and enabling better transparency and sped up processes. This is achieved because of planning optimization of deliveries and supported detection of bottlenecks and risks. Latencies can also be avoided or at least can be reduced. The project was finished in 2013, firstly implemented for a candy product SC with 18 read points shared between the manufacturer, a supermarket and handling unit pooling operator. The implementation focused in tracking 2500 handling units, which are transported on repeat between the three companies. Around 90.000 read events had been generated. The achievement of this pilot implementation was a complete transparent handling unit flow between the three participants. All companies notice many advantages and benefits by the roll out of the smaRTI in the whole SC, and not only to one product. Every handling unit of the shipment, and the transported goods could be directly located as soon they reached a read point location. The next step is to analyze the smaRTI data besides the business processes data to detect not only bottlenecks or risks along the SC, but also to find action within the process, which can be improved or even are not necessary and is possible to be excluded. An example of smaRTI would be some manual quality measurement for counting the received goods and comparing them with the expected numbers. Summarizing the smaRTI approach can decrease the loose of handlings units by 50% and can achieve reduction in stock up to 10% in the consumer goods domain.

Use cases concerning anticipatory shipping intention, as described in (Bubner, 2014), will achieve the speed up delivery times of goods and increase the use ratio of distribution capacities. One use case example is represented by DHLs volume forecast with predictive capacity utilization and planning. The parcel volume analytics give the chance to improve the prediction accuracy of the expected parcels and the freight within their network. This is realized by correlating data from various sources and with different degree of privacy protection. Some input data could be firm’s internal shipment history data and external

events, public holidays, Google search terms, weather condition and forecast, and the shopping habits of the online customers.

Another example is presented in (Spiegel et al., 2014), and it is about Amazons US Patent for Anticipatory shipping applied from December 2013. The aim of this patent was to ship goods prior to the customers' order so as to reduce delivery time. The prediction of an upcoming order was the key element of this patent. By patent several applications were enabled. First, a shipment is delivered to a destination area without knowing the final shipping address. During the shipment to a specific geographical location, this address will be completed depending on the placed orders in the meantime. This optimized distribution improves the lead time and the customer satisfaction and can help for increasing sales. Secondly, Amazon tried to match some of the goods while they are already in transit to a specific geographical location with current customer orders. During the transportation, the goods are added to orders, which have been placed, as expected, by a specific customer in this area. The motivation was to use the disadvantage of lower-cost transportation, non-expedited delivery, for buffering the speculatively selected items. If an item of these in transit fits to a customers demand, the item will be delivered within a short delivery time. This results in the reduction of transportation costs for both Amazon and the customer. The shopping process itself can play very important role in the forecast data generated for shipment and delivery. Some information influencing the forecasting could be specific web pages viewed by the customers and the duration of views, links hovered over with the mouse arrow and duration of hovering, the shopping cart, the wish lists and the relatedness to previous purchases, as well as in the case of a new product release, where no historical buying patterns from similar customers exist.

4.9. RFID Technology in SCM and BDA

4.9.1. RFID-enabled logistics data

Data coming from the RFID-enabled logistics control within logistics and production areas can be considered as a stream of tuples in the form of EPC. Location, operator, time and quantity, where Electronic Product Code (EPC) is the unique ID of a batch of materials, which could be read by an RFID reader. Location is referred on the exact position where the operations or events take place. An event means an actual RFID detection or an operation on RFID devices. Operator is the executor of the event. Time is

indicated when the event occurs. Quantity accounts for the standard amount of materials in a batch.

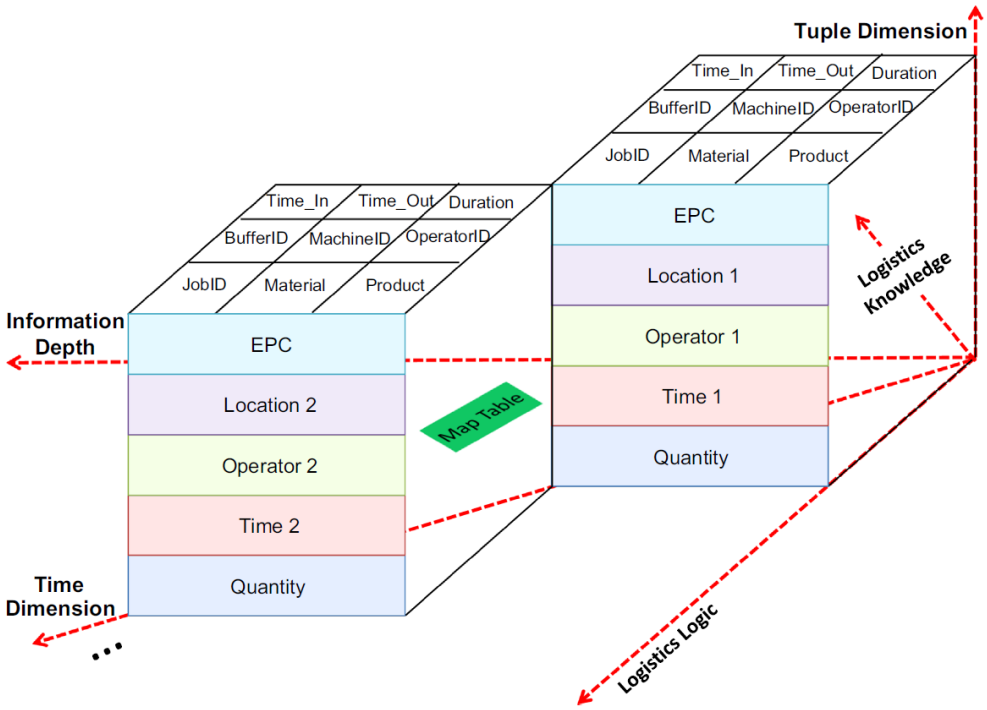

figure 15 RFID-cuboid in data warehouse. Source (Zhong et al.,(2015, pp. 260-272))

4.9.2. RFID logistics data warehouse

RFID logistics data warehouse is used for managing and storing the tuples according to a time sequence for addressing the complex logic relationship among enormous tuples since RFID generates enormous big amount of data at a glance of time on a continuous basis. The RFID-Cuboid is formed by various data records given the logical logistics operations. The main differences between the traditional database and RFID logistics data warehouse are the presence of data structure of the RFID-Cuboid and a Map Table which indicated the related records from various tables so as to preserve the meaningful data (Zhong et al., 2013). A Map Table is designed as a service in the warehouse to build up the RFID-Cuboid according to the predefined logics. For instance, when receiving an EPC, the Map Table is able to locate all the records in the data warehouse and then initiate a cuboid which is a cubic structure according to the logistics operations. In continue, the Map Table chains the cuboids given the time sequence so that all the logistics operations of the EPC identified material could be presented by the RFID-Cuboids. RFID-Cuboid plays a very important role in RFID logistics data warehouse. *Figure 16* demonstrates'

on the key principle of RFID-Cuboid, preserving the logistics paths at different abstraction levels. In tuple dimension, key attributes like EPC, Location, Operator, Time, and Quantity are presented. The tuple dimension is so abstract that it is very difficult to understand because these attributes are directly from the data warehouse with various data types structured, semi structured, unstructured etc. Thus, in information depth dimension, the attributes are converted into meaningful information which is shown on the top of each RFID-Cuboid. In the dimension of time, the RFID-Cuboids are chained according to the time stamp which records when the event occurred. What happened in an event is presented in logistics logic dimension that keeps the executed operations and procedures. With the chained RFID-Cuboids and detailed logistics logic, the entire information within the manufacturing sites are accumulated. In logistics knowledge dimension, valuables like production deviations, logistics trends and quantitative performance of machines and workers, could be exploited from the big number of RFID-Cuboids. Such valuables are very important for supporting advanced decisions like logistics optimization and planning.

5. Case studies and examples of implementation and use of Big Data Analytics in Supply Chains

In the sector of SCM and logistics we can draw many benefits from the application of technological advances but, also from the parallel methodological developments through BD analysis by the use of BA. Enterprises that their SC is totally controlled by them and those that outsource to third-party logistics (3PL) providers, have to deal with a very big inflow of freight, products, goods and services on an everyday basis while at the same time to manage also voluminous datasets that are generating from each step of a product into the SC. Numerous shipments are tracked daily from the place of origin to the place of the final destination, providing information about the content, the weight, the size, the route, the location etc., of each individual shipment, across a large numbers of networks (Watson, 2013). In this data, that provides the BD, it is contained potential value that needs to be taken so to get benefit from. This analysis and exploitation is possible still to a large extent unprocessed, but more and more enterprises are expressing their interest in these developments and experiencing the development of operational efficiency in combination with improved effectiveness and customer satisfaction.

Below in this section, we will present various examples and case studies, and we will illustrate what has been achieved and we will show what potential value can be extracted from the application of BDA. The aim target is not to cover all SC elements totally, but to provide a glimpse of what can be and has been achieved in all the fields of application. Although, a few actual initiatives implemented by existing firms are briefly discussed.

5.1. IBM – applying analytics to the Supply Chain

IBM is an international company that is established in more than 170 countries all over the world. IBM was one the organizations that recognize early the importance of data analytics and invest in the optimization of their SC (Dietrich *et al.* 2014). Many solutions have been developed. We will present below four examples are analyze briefly to in order to point out the variety of problems that were solved using a various BDA tools:

- **Quality early-warning system:** The target of this system is to allocate and prioritize quality problems much earlier than the traditional statistical process control. This is a BD analytic tool that is deployed upstream at suppliers, in IBM's operations and with products in the field. With the use of voluminous data coming from the SC and after analyzing them it was possible to avoid the phenomenon of rework, increase productivity, increase customers' satisfaction and guarantee better quality. This play very important role to the achievement of significant cost savings.

- **IBM buy analysis tool:** IBM by co-operating with many company partners understand the complexity and the difficulty of having the correct stock availability. This tool, provides supply and demand visibility, and also guarantee the best possible distribution channel management as well as making sure the right product is delivered to meet the customer needs, while at the same time minimizes the inventory levels.
- **Account receivable – next best action:** The aim of this IBM tool is to achieve optimization of the resources required to collect revenues with the assistance of BDA.

- **Supply chain social listening:** This is a very innovative tool that uses social media to monitor social channels and provide valuable and timely data on events

(e.g. disasters) which may lead to the disruption of the SC. This is also a way to obtain Information on products. There are already plenty potential uses and applications for this tool..

IBM presents the use of especially predictive and prescriptive analytics through real examples implemented over the last years.

5.2. DHL: Big data in logistics

Except from being a leading global parcel courier service companies that operates in a fast-changing and dynamic environment, DHL is also a firm that invest and looks for new trends and innovations in the global environment, like BDA. DHL was also a leader in the application of analytics tools. It strives to continuously innovate, in this way developing new strategies so as to remain competitive (Budned et al., 2013; BIGDIN LOGISTICS, 2013) and develop new services and business models. From a BD perspective they identified two major driving forces (BIGDIN LOGISTICS, 2013):

- How DHL can be transformed from a 'deep well of data' to a deep exploitation?
- How this can be used for operational efficiency improvement and customer service leading to new business models?

With the application of advanced IT technologies and analytics, DHL was able to analyze great amount of data much faster. Application cases where DHL implemented these new innovations and models include the following (BIGDIN LOGISTICS, 2013):

- **Last-mile optimization**: Last mile delivery was always the most sensitive and the less cost-effective procedure in the SC because goods should be deliver as fast as possible with the minimum possible cost. The efficiency of the last-mile delivery was improved by processing an enormous stream of data to attain:
    - *Real-time route planning optimization:* Rapid processing of real-time data, like weather and traffic, in the last mile achieves great time saving during the delivery process.
    - *Crowd-based pick-up and delivery:* An innovative solution that may arise by making commuters, students or taxi drivers the part of the SC that will make last-mile deliveries on the routes they are travelling anyway.
- **Predictive network and capacity planning:** Avoidance of excess or capacity shortages through:

- *Strategic network planning:* Distribution network adoption according to anticipated future demand.
- *Operational capacity planning:* With the use of real-time information about shipments. In this way, we can predict and allocate the resources that are required for the next 48 hours.

- **Customer value management:** After processing data deriving from the distribution networks we can limit customer attrition and understand customer demand through:
  - *Customer loyalty management:* To retain customers' BD analysis allows a comprehensive assessment of customer satisfaction by merging multiple extensive data sources.
  - *Continuous service improvement and product innovation:* Here the feedback received from the customers is used in a way to continuously improve the service quality and product innovation.

- **Supply chain risk management:** Real-time analysis of voluminous data streams can play critical role to forecast events and take measures to minimize the associated potential risk, This includes:
  - *Risk evaluation and resilience planning:* Continuous monitoring of SCs to detect performance deviations, etc.

- **B2B demand and supply chain forecast:** Using data sourcing from the flow of goods and the millions daily shipments, demand and SC requirements can be forecasted:
  - *Market intelligence for small and medium-sized companies:* By analyzing shipment records a market research service can be provided to small and medium-sized companies.
  - *Financial demand and SC analytics:* Data that are collected from global distribution networks provide insight into development perspectives, etc.

- **Real-time local intelligence:** A huge source of intelligence is sourcing from the pick-up and delivery shipment data:
  - *Address verification:* Verification of customer's address, a big requirement for online commerce.

- *Environmental intelligence:* Using various sensors on delivery vehicles we can produce environmental statistics like pollution and expansion.

All issues described above are very innovative, but also forward-thinking, presenting also all the great advantages that can be offered by the combination of BD and advanced analytics. In order to DHL remain competitive against competitors in this always developing and cut-throat environment, believes that the development of its future strategies utilizing BD with the appropriate methodologies and analytic tools is needed.

5.3. City of Stockholm Real-Time Intelligent Transportation Services

Another project that is based on IBM InfoSphere BDA intended to improve the quality of the transportation network in the city of Stockholm (Biem et al., 2010). Around 120,000 vehicles of trucks and taxis equipped with GPS devices were used to collect and send a big datasets every second combined with a map containing over 600,000 links (see Fig. 17).

The used BDA system in combination with the collected data with past traffic data and weather forecasts to generate more accurate predictions about future traffic conditions such as shortest-time routes in real-time. The results were used by the public, urban planners, police officers, firemen, etc.

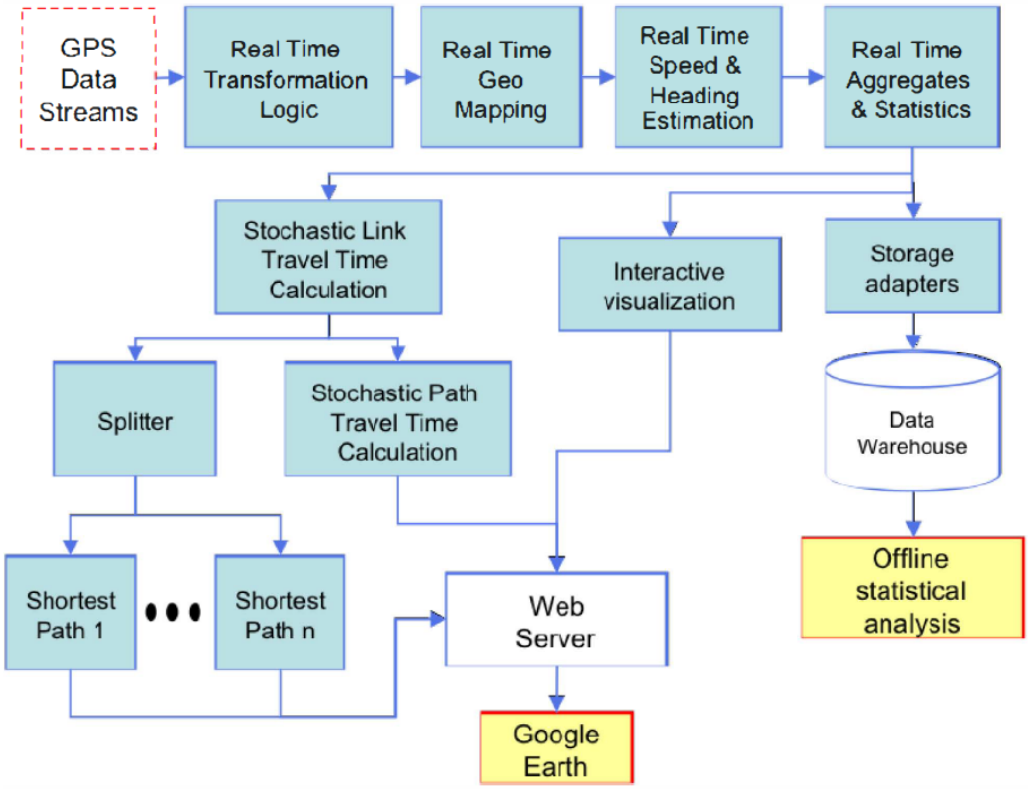

figure 16 Stockholm real-time monitoring of transportation network flow. Source (Leveling et al. (2014))

5.4. Cloud-Enhanced System Architecture for Logistics Tracking Services

The target of this project is to build a system that combines IoT technology, SaaS cloud architecture and BDA technologies for setting-up an effective and efficient real-time monitoring of customers shipments (Lin et al. 2013).

The data is harvested from smart devices: 2D QR codes, GPS locations and RFID electronic codes. Then, data is transferred using RSA encryption algorithm to protect customers' privacy using wireless networks. A BDA system using HBase as database is used to store all these unstructured data (see Fig. 18 and Fig. 19).

The use of traditional computing system to handle that kind of projects is impossible because of the large needs for storage capacity, processing power and bandwidth for such lot of unstructured logistics information.

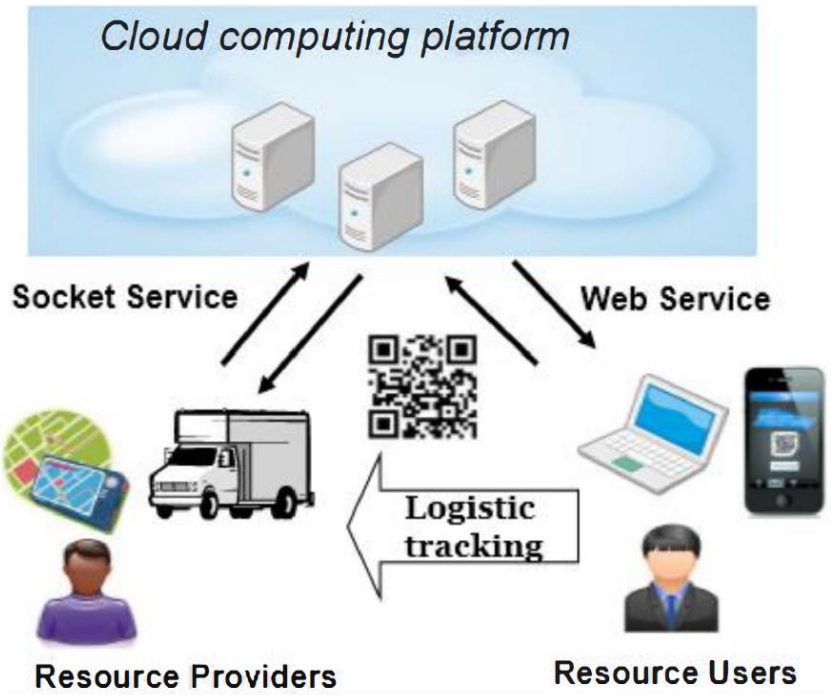

figure 17 Architecture overview of cloud-enhanced system architecture for logistics tracking services. Source [43], p. 546.

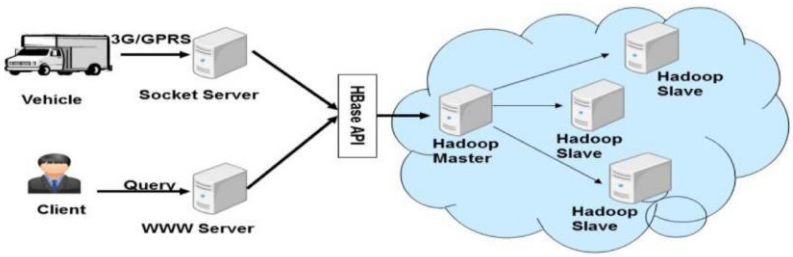

figure 18 The cloud computing structure of cloud-enhanced system architecture for logistics tracking services. Source [43], p. 547.

We can classify these transportation projects in two categories, projects for the improvement of the operational efficiency and projects that for the improvement of customer experience.

5.5. Proposed Big Data system for containers code recognitions

The evolution of container transportation industry is in our day's a very active industry with very large number of containers to be on the move every day. In order to be able to supervise the delivery of all these containers, unique identifiers codes are written on each

container, but the manual reading of these unique codes include lot of problems like slow speed and high error possibility.

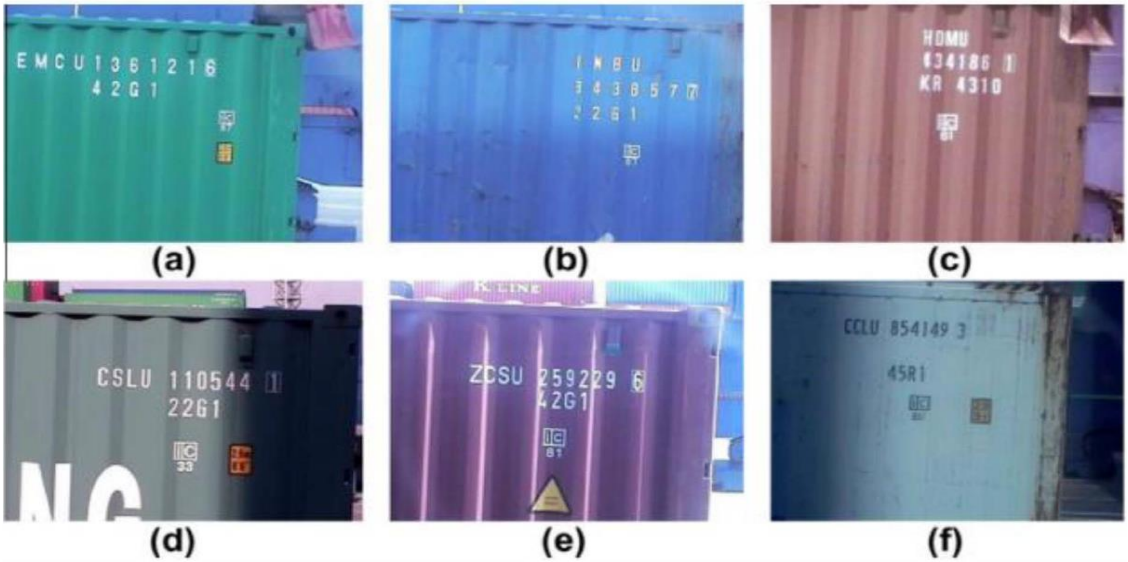

figure 19 examples of container codes. Source Google

As well, in contrast to vehicle license plate recognition systems, container-code recognition systems has more challenges to overcome because of the low code contrast and the large varieties of sizes, types, colors, positions, alignment and inter-spaces of these codes as we can check in f*igure 20*.

As it is stated in (Wu et al., 2012), automatic container code recognition has three requirements to operate: firstly text detection, secondly characters extraction and thirdly text recognition. Duo to text detection step is very important for the other steps, we chose a robust method insensitive for code contrast and other text variable characteristics like texture-based text detection method using Haar wavelet transform for text features extraction and Support Vector Machine (SVM) to classify these features into text and non-text regions (Sayahi et al., 2014).

In order to overcome the obstacle of high computation time of the employed text detection method (Sayahi et al., 2014) and text recognition method (Halima et al., 2010), which is a big difficulty for real time and industry applications including container transportation industry, authors took the decision to use Hadoop MapReduce to achieve a parallel execution model as a solution to reduce the computation time for the proposed system.

Concluding, the first step is to capture the container code using monitoring cameras or mobile devices and store them on Hadoop distributed system file (HDFS). Second step is a pre-processing and graying color image to be applied on the captured image. Then, the image should be decomposed in gray color in 20x20 pixel bloc. These blocs are analyzed and classified separately on different machines with the use of MapReduce programming model so as to extract the text regions. Another one step is used to separate code characters from the extracted text regions. Then, Optical Character Recognition (OCR) is applied on individual characters using MapReduce programming model and finally, the container code is recognized by merging these characters. The process is presented in *figure 21* step by step.

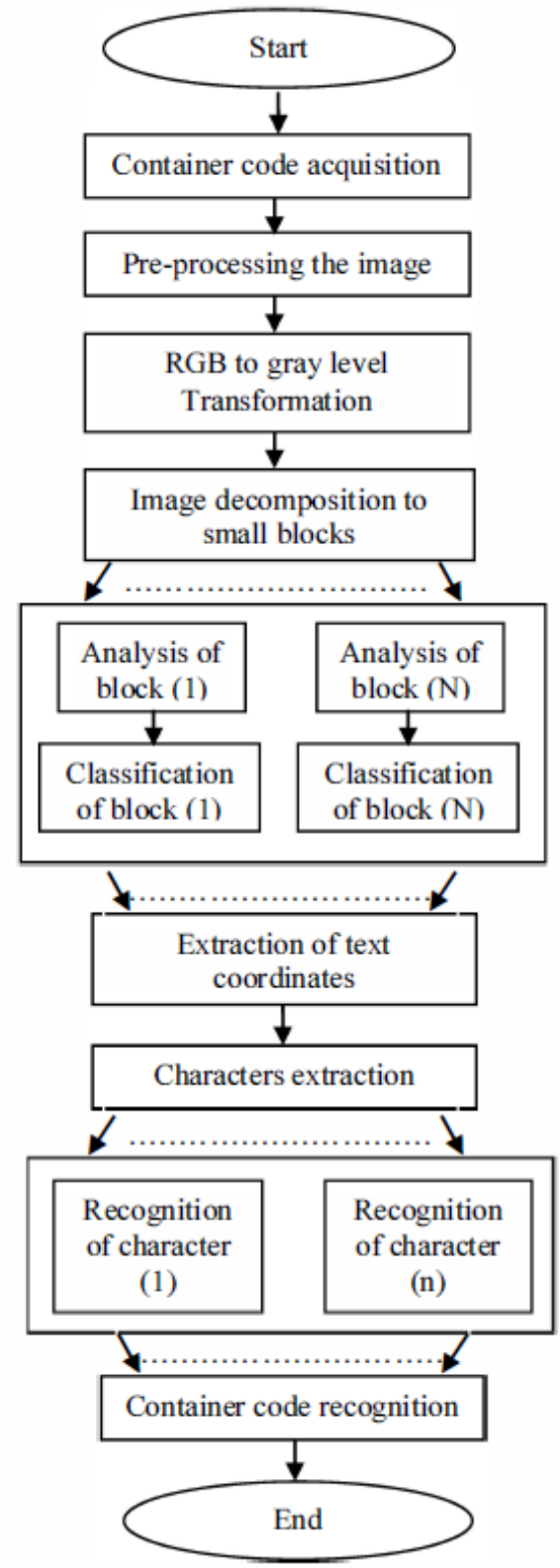

figure 20 Big data system for containers code recognition. Source (Ayed et al., 2015)

5.6. Big Data Distribution Concept

For the better presentation and understanding of the below presented customer concerns, we developed a framework that consists of three main components as we can see in the following *figure 22*. From right to the left: (1) personal smartphone or any other smart device of the customers , (2) a BDA computation center, residing in the data processing facilities of the delivery service provider, and (3) delivery vehicles, equipped with GPS technology.

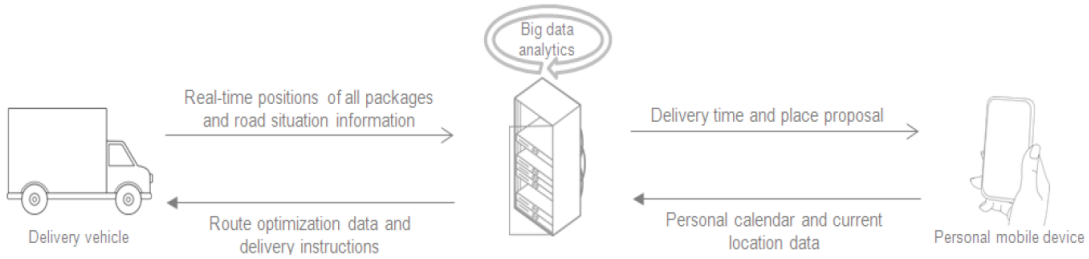

figure 21 Big data-aided concept for package delivery services (information flow). Source (Engel et al., 2014).

The information flow between the system components operates as described below:

- The customer shares' his personal daily calendar as source of their prospective location information and also global positioning data updates with the computation center through a dedicated mobile application, installed on their personal smart device.
- As far as the van is concerned, sensors are installed inside each delivery vehicle in order to register all the packages that are currently on board and send this information to the computational center. Additionally, delivery vehicles collect and send their location and road situation data constantly.
- Taking into consideration road situation information we understand real-time routing-relevant data like traffic information, locations of road construction sites and detours, weather conditions, locations of road construction sites and detours and so forth.

Concerning the dynamic customer data and road situation data, an analysis is conducted, providing more accurate routing information and computations needed for more efficient and accurate delivery schedule. As far the customer is concerned, positioning information is being used in order to compute and propose possible delivery places and estimated time of delivery, which are sent back to the customers' personal smartphone or smart devices. Therefore, customers receiving a notification in their smart devices with the suggested

locations for delivery and to choose one of them listed in their calendar suggesting also the time for delivery. After customers' selection of one of the possible options, the computation center recalculates the optimal route for the vehicle and sends a routing update to the vehicle driver regarding the booked delivery time slot information. For routing optimization and customer service, this process need to be conducted by all the customer that expect their shipment in a specific territory serviced by specific vehicle.

When the negotiation will all the customers of the territory are completed and the route is communicated, the driver of the delivery vehicle only receives following routing updates based on changing customers' locations and traffic and road updated. In case a customer decides to change its planned location in the calendar or data sourcing from the GPS sensor of customers' phone do not correspond to the planned place of delivery, this will be recognized by the mobile application. The user then will receive a new query if they are experiencing changes in their calendar, and if the user confirms, an update will be sent to the computational center where the route will be recalculated and the driver of the delivery vehicle will be notified accordingly about the next steps.

Below is described the technical side of the concept: Inside the vehicle, all packages are registered with its unique tracking number which is also used during loading and unloading with the use of RFID technology. All tracking numbers of the packages are collected and an inventory list of the truck responsible for each territory is created. This list is sent to the computation center using a dedicated communication interface via 3G wireless Internet connection. The computation center links tracking numbers to the customer data and recognizes the cases which users should be contacted for further information. In continue, users are notified with a short SMS or a notification to the application and are urged to launch the preinstalled application on their smart device or smartphone in order to start the communication with the computation center, which next sends specific information requests back to the application.

After receiving all information needed, the application installed in the smart device asks the user to grant temporary access to the calendar and GPS navigation sensor of its smart device. If the user authorizes the access, data is received by the mobile application and transferred using an encrypted channel to the computation center, where it is temporarily stored in a database. The secured communication channel is negotiated on top of a mobile broadband Internet connection (3G). It is important to underline that dynamic data, like

GPS positions of the delivery vehicles and road status readings are constantly being gathered and creating large amounts of data that need to real-time analysis for the achievement of route optimization. This is the reason why an in-memory database should be used in this concept. As soon as necessary information and location data from the customers is collected, real-time BD analysis methods are taking place, and a route optimization for the vehicle is calculated. Following, the users are notified over the already established secure communication channel about the delivery time and the possible places of delivery.

For the achievement of security and safety of user data, no personal information should be transferred over the communication network. For instance, the communication with delivery vehicles is limited to the package tracking number, while communication with the users' smart device is conducted using a separate secured communication channel for each connected user. No user location data should be accessible outside of the in-memory database at any time and must be deleted immediately after the delivery is fulfilled and the communication session with customer smart device is closed. Only the road situation information is constantly kept in the database in order to help the delivery route computation for all vehicles and improve route proposals based on historic data. We should note in this point that the customer remains the only and single owner of his location data and according to the concept, only temporarily gives access of this information to the computation center, specifically for the routing analytics and planning.

6. Conclusion and future research

In conclusion, BDA has the potential to transform and outperform traditional SCM practices. In this research we review plenty academic papers on BD and SCM and application models in non-academic sectors. In past researches, we see that BDA has mostly been explored from the technological perspective to rationalize its economic benefits, but in this research we emphasize the necessity to delineate BDA capabilities in SC to extract value from BD. The structured approach followed for reviewing the literature has revealed existing contributions of BDA and SCM research. Findings show a very important increase in the number of papers published in recent years around the field. Social media based academic research has emerged as an important discipline in SC field. Findings suggest that BDA could be beneficial if organizations can develop the right capabilities to effectively use the BD.

In this research the 2nd chapter has to do with understanding the term of data. We analyze the data harvesting process, data quality, data process, data safety and transferring. Then we introduce the conceptualization of BD and its capabilities. We review the BDA capabilities and characteristics and we expand to the BD enabled decision making models. Then, in the 3rd chapter, we make an introduction to SC. First we give a definition for the SC and then we analyze all the SC processes. We discuss about SC agility, the Bullwhip effect that is very common for most of the organizations, the SC visibility, and the application and introduction of the RFID technology. In the 4th part, we address the magnitude of understanding the provenance of BD in SC and the optimization of data generation processes. We underline the importance of integrating and standardizing data from heterogeneous sources to offer more coherent data sets to analytics systems. We proceed presenting different types of analytics and the importance of assimilating the findings into the business process. In comparison with descriptive and predictive analytics, prescriptive analytics require a negligible amount of human intervention (Puget, 2015), which will revolutionize the decision-making process. In addition, we address the role of BD driven SC in the sector of marketing, procurement, warehouse and transportation. What is more, from value creation and users' perspective, the significant role of data visualization and data-driven culture in increasing the flexibility and adaptability is discussed. Challenges of BD in the SCM are also presented. Finally, we have the presentation of real and hypothetical applications of BDA in the SC and the outcomes are presented.

Regarding the future research, qualitative approaches can provide detailed and rich views of phenomena, especially when those phenomena are new or imperfectly understood within the frameworks of extant theory (Corbin and Strauss, 2008). Qualitative methodology is well suited to describe constructs and relationships between constructs, but it does not describe the magnitude of the relationships. That has as a result, future research should employ quantitative research methods to estimate the degree to which different applications of BD influence firm performance.

A key finding of this research is that BD in SCML is often used as an operational instrument. Despite this, strategy literature has long recognized that influences, including capabilities, have differing effects based on levels of application: industry, enterprise, and small business unit (Chang and Singh, 2000; Short et al., 2007). While this sample

included representatives from a variety of organizational levels, in this study we do not investigate cross-level differences. Consequently, future research should seek multi-level samples explicitly to determine whether or not BD may provide varying resource value at different application levels.

Bibliography

Addo-Tenkorang, R., Helo, P. T., (2016). Big data applications/supply-chain management: A literature review. *Computer & Industrial Engineering, vol. 101,* pp 528-543

Agarwal, A., Shankar, R., & Tiwari, M. K. (2007). Modeling agility of supply chain. *Industrial marketing management*, *36*(4), 443-457.

Aho, A. M. (2015, August). Product data analytics service model for manufacturing company. In *International Conference on Knowledge Management in Organizations* (pp. 282-296). Springer, Cham.

Alfaro, L., Antràs, P., Chor, D., & Conconi, P. (2015). Make or buy decisions over upstream and downstream inputs: An investigation of firm boundaries along value chains. *VoxEu. org*, *14*.

Anderson, P. F. (1982). Marketing, strategic planning and the theory of the firm. *The Journal of Marketing*, 15-26.

Annonk, N. Y. (2013). Da Nang, Vietnam Turns to IBM to Transform City Systems.
Arnold, S. E. (1992). Information manufacturing: the road to database quality. *Database*, *15*(5), 32-39.

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*, 416-436.

Ashton, K. (2009). That 'internet of things' thing. *RFID journal*, *22(7)*, 97-114.
Atzori, L., Iera, A., & Morabito, G. (2010). The internet of things: A survey. *Computer networks*, *54*(15), 2787-2805.

Ayed, A. B., Halima, M. B., & Alimi, A. M. (2015, May). Big data analytics for logistics and transportation. In *Advanced Logistics and Transport (ICALT), 2015 4th International Conference on* (pp. 311-316). IEEE.

Babai, M. Z., Syntetos, A. A., Dallery, Y., & Nikolopoulos, K. (2009). Dynamic re-order point inventory control with lead-time uncertainty: analysis and empirical investigation. *International Journal of Production Research*, *47*(9), 2461-2483.

Bagchi, P., Lejeune, M. A., & Alam, A. (2014). How supply competency affects FDI decisions: some insights. *International Journal of Production Economics, 147*, 239-251.

Bakshi, K. (2012, March). Considerations for big data: Architecture and approach. In *Aerospace Conference, 2012 IEEE* (pp. 1-7). IEEE.

Balaraj, S. (2013). Optimization model for improving supply chain visibility. *Big Data: Countering Tomorrow's Challenges*, 9.

Ballou, D. P., & Pazer, H. L. (1985). Modeling data and process quality in multi-input, multi-output information systems. *Management science*, *31*(2), 150-162.

Ballou, D., Wang, R., Pazer, H., & Tayi, G. K. (1998). Modeling information manufacturing systems to determine information product quality. *Management Science*, *44*(4), 462-484.

Bandyopadhyay, S., Mehta, M., Kuo, D., Sung, M. K., Chuang, R., Jaehnig, E. J., ... & Fiedler, D. (2010). *Rewiring of genetic networks in response to DNA damage. Science, 330(6009)*, 1385-1389.

Barratt, M., & Oke, A. (2007). Antecedents of supply chain visibility in retail supply chains: a resource-based theory perspective. *Journal of operations management*, *25*(6), 1217-1233.

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM computing surveys (CSUR)*, *41*(3), 16.

Bell, L. (2014). IDF: Intel announces A-Wear to push big data apps via Internet of Things: The INQUIRER<http://www.theinquirer.net/inquirer/news/2364331/idfintel-announces-a-wear-to-push-big-data-apps-via-internet-of-things>.

Berkovich, S., & Liao, D. (2012, July). On clusterization of big data streams. In *Proceedings of the 3rd International Conference on Computing for Geospatial Research and Applications* (p. 26). ACM.

Berners-Lee, T. (2001). Business model for the semantic web. *W3C Semantic Web Activity background (2001)*.

Beutel, A. L., & Minner, S. (2012). Safety stock planning under causal demand forecasting. *International Journal of Production Economics*, *140*(2), 637-645.

Bhakoo, V., Singh, P. J., & Chia, A. (2015). Supply chain structures shaping portfolio of technologies: Exploring the impact of integration through the "dual arcs" framework. *International Journal of Physical Distribution & Logistics Management, 45(4),* 376-399.

Bhatti, R., LaSalle, R., Bird, R., Grance, T., & Bertino, E. (2012, June). Emerging trends around big data analytics and security: Panel. In *Proceedings of the 17th ACM symposium on Access Control Models and Technologies* (pp. 67-68). ACM.

Biem, A., Bouillet, E., Feng, H., Ranganathan, A., Riabov, A., Verscheure, O., ... & Moran, C. (2010, June). IBM infosphere streams for scalable, real-time, intelligent transportation services. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data* (pp. 1093-1104). ACM.

Biesdorf, S., Court, D., & Willmott, P. (2013). Big data: What's your plan?: Insights & Big Data Logistics A health-care Transport Capacity Sharing Model big_data_whats_your_plan>.

Bisson, P., Stephenson, E., & Viguerie, S. P. (2010). The productivity imperative. *McKinsey Quarterly*, 1-7.

Blackburn, R., Lurz, K., Priese, B., Göb, R., & Darkow, I. L. (2015). A predictive analytics approach for demand forecasting in the process industry. *International Transactions in Operational Research, 22(3),* 407-428.

Blake, R., & Mangiameli, P. (2011). The effects and interactions of data quality and problem complexity on classification. *Journal of Data and Information Quality (JDIQ), 2*(2), 8.

Boos, D., Guenter, H., Grote, G., & Kinder, K. (2013). Controllable accountabilities: the internet of things and its challenges for organisations. *Behaviour & Information Technology*, *32*(5), 449-467.

Borade, A. B., Kannan, G., & Bansod, S. V. (2013). Analytical hierarchy process-based framework for VMI adoption. *International Journal of Production Research*, *51*(4), 963-978.

Botta-Genoulaz, V., Millet, P. A., & Grabot, B. (2005). A survey on the recent research literature on ERP systems. *Computers in industry*, *56*(6), 510-522.

Bottani, E., & Rizzi, A. (2008). Economical assessment of the impact of RFID technology and EPC system on the fast-moving consumer goods supply chain. *International journal of production economics, 112(2),* 548-569.

Braunscheidel, M. J., & Suresh, N. C. (2009). The organizational antecedents of a firm's supply chain agility for risk mitigation and response. *Journal of operations Management*, *27*(2), 119-140.

Bruque Camara, S., Moyano Fuentes, J., & Maqueira Marin, J. M. (2015). Cloud computing, Web 2.0, and operational performance: the mediating role of supply chain integration. *The International Journal of Logistics Management, 26(3),* 426-458.

Bubner, N., Bubner, N., Helbig, R., & Jeske, M. (2014). Logistics trend radar, Delivering insight today. *Creating value tomorrow*.

Butner, K., Geuder, D., & Hittner, J. (2008). Mastering carbon management: balancing trade-offs to optimize supply chain efficiencies. *IBM Institute for Business Value*.

Capgemini Deutschland Holding GmbH, (2014). Studie IT-Trends 2014: IT-Kompetenz im Management steigt . Capgemini Deutschland Holding GmbH.

Case, L. (2013). Genpact Partners with Jaguar Land Rover to Optimize Their Procurement Processes. *Automotive Industries, 333*, 334.

Castiglione, A., Gribaudo, M., Iacono, M., & Palmieri, F. (2014). Exploiting mean field analysis to model performances of big data architectures. *Future Generation Computer Systems, 37,* 203–211.

Çetinkaya, S., & Lee, C. Y. (2000). Stock replenishment and shipment scheduling for vendor-managed inventory systems. *Management Science, 46(2),* 217-232.

Chae, B. K. (2015). Insights from hashtag# supplychain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research. *International Journal of Production Economics, 165*, 247-259.

Chae, B., & Olson, D. L. (2013). Business analytics for supply chain: A dynamic-capabilities framework. *International Journal of Information Technology & Decision Making, 12(01),* 9-26.

Chang, S. J., & Singh, H. (2000). Corporate and industry effects on business unit competitive position. *Strategic Management Journal, 21(7),* 739-752.

Chatfield, D. C., Kim, J. G., Harrison, T. P., & Hayya, J. C. (2004). The bullwhip effect—impact of stochastic lead time, information quality, and information sharing: a simulation study. *Production and Operations Management, 13(4),* 340-353.

Cheikhrouhou, N., Marmier, F., Ayadi, O., & Wieser, P. (2011). A collaborative demand forecasting process with event-based fuzzy judgements. *Computers & industrial engineering, 61*(2), 409-421.

Chen, A., & Blue, J. (2010). Performance analysis of demand planning approaches for aggregating, forecasting and disaggregating interrelated demands. *International Journal of Production Economics*, *128*(2), 586-602.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). *Business intelligence and analytics: from big data to big impact. MIS quarterly,* 1165-1188.

Chen, M., Mao, S., Zhang, Y., & Leung, V. C. (2014). Big data: related technologies, challenges and future prospects.

Chen, Z. L., & Vairaktarakis, G. L. (2005). Integrated scheduling of production and distribution operations. *Management Science*, *51*(4), 614-628.

Chen-Ritzo, C. H., Ervolina, T., Harrison, T. P., & Gupta, B. (2010). Sales and operations planning in systems with order configuration uncertainty. *European Journal of Operational Research*, *205*(3), 604-614.

Chernatony, L. D., Daniels, K., & Johnson, G. (1994). Competitive positioning strategies mirroring sellers' and buyers' perceptions?. *Journal of Strategic Marketing, 2(3),* 229-248.

Chongwatpol, J., & Sharda, R. (2013). RFID-enabled track and traceability in job-shop scheduling environment. *European Journal of Operational Research, 227(3),* 453-463.

Choudhury, B., Agarwal, Y. K., Singh, K. N., & Bandyopadhyay, D. K. (2008). Value of information in a capacitated supply chain. *INFOR: Information Systems and Operational Research*, *46*(2), 117-127.

Chow, H. K., Choy, K. L., & Lee, W. B. (2007). A dynamic logistics process knowledge-based system–An RFID multi-agent approach. *Knowledge-Based Systems, 20(4),* 357-372.

Christopher, M., & Towill, D. R. (2000). Supply chain migration from lean and functional to agile and customised. *Supply Chain Management: An International Journal*, *5*(4), 206-213.

Christopher, Martin L. (1992), *Logistics and Supply Chain Management,* London: Pitman Publishing.

Cohen, J., Dolan, B., Dunlap, M., Hellerstein, J. M., & Welton, C. (2009). MAD skills: new analysis practices for big data. *Proceedings of the VLDB Endowment*, *2*(2), 1481-1492.

Cooper, M. C., & Ellram, L. M. (1993). Characteristics of supply chain management and the implications for purchasing and logistics strategy. *The international journal of logistics management*, *4*(2), 13-24.

Coronel, C., Morris, S., & Rob, P. (2011). Database systems: design, implementation, and management. *Cengage Learning*, *9*.

Crainic, T. G., & Laporte, G. (1997). Planning models for freight transportation. *European journal of operational research, 97(3),* 409-438.

Croson, R., & Donohue, K. (2005). Upstream versus downstream information and its impact on the bullwhip effect. *System Dynamics Review: The Journal of the System Dynamics Society, 21(3),* 249-260.

Dai, Q., Zhong, R., Huang, G. Q., Qu, T., Zhang, T., & Luo, T. Y. (2012). Radio frequency identification-enabled real-time manufacturing execution system: a case study in an automotive part manufacturer. *International Journal of Computer Integrated Manufacturing, 25(1),* 51-65.

Daugherty, P. J. (2011). Review of logistics and supply chain relationship literature and suggested research agenda. *International Journal of Physical Distribution & Logistics Management, 41(1),* 16-31.

Davenport, T. H. (2006). Competing on analytics. *harvard business review*, *84*(1), 98.

Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Harvard Business Press.

Davenport, T. H., & Prusak, L. (2000). *Working knowledge: How organizations manage what they know*. Harvard Business Press.

Davenport, T. H., Harris, J. G., & Morison, R. (2010). *Analytics at work: Smarter decisions, better results*. Harvard Business Press.

de Oliveira, M. P. V., McCormack, K., & Trkman, P. (2012). Business analytics in supply chains–The contingent effect of business process maturity. *Expert Systems with Applications, 39(5),* 5488-5498.

De Souza, R., Zice, S., & Chaoyang, L. (2000). Supply chain dynamics and optimization. *Integrated Manufacturing Systems, 11(5),* 348-364.

Dejonckheere, J., Disney, S. M., Lambrecht, M. R., & Towill, D. R. (2003). Measuring and avoiding the bullwhip effect: A control theoretic approach. *European Journal of Operational Research, 147(3),* 567-590.

Dietrich, B. L., Plachy, E. C., & Norton, M. F. (2014). *Analytics across the enterprise: How IBM realizes business value from big data and analytics*. IBM Press.

Disney, S. M., & Towill, D. R. (2003). The effect of vendor managed inventory (VMI) dynamics on the Bullwhip Effect in supply chains. *International journal of production economics, 85(2),* 199-215.

Dong, M., & Chen, F. F. (2005). Performance modeling and analysis of integrated logistic chains: An analytic framework. *European Journal of Operational Research*, *162*(1), 83-98.

Doskonałość w logistyce (2010) [Excellence in logistics]. Interview with H.-Ch. Pfohlem. Eurologistics.

Drexl, M. (2013). Applications of the vehicle routing problem with trailers and transshipments. *European Journal of Operational Research*, *227*(2), 275-283.

Dutta, D., & Bose, I. (2015). Managing a big data project: the case of ramco cements limited. *International Journal of Production Economics*, *165*, 293-306.

Effizienzcluster Management GmbH, (2014). Smart Reusable Transport Items (smaRTI) - Intelligent Material Flow, (online) http://www.effizienzcluster.de/en/leitthemen_projekte/projekt.php?proPid=5

Ellinger, A. E., Natarajarathinam, M., Adams, F. G., Gray, J. B., Hofman, D., & O'Marah, K. (2011). Supply chain management competency and firm financial success. *Journal of Business Logistics, 32(3),* 214-226.

Emery, J. C. (1969). *Organizational planning and control systems: theory and technology*. Macmillan Pub Co.

Emmett, S., & Crocker, B. (2006). *The relationship-driven supply chain: creating a culture of collaboration throughout the chain*. Routledge.

Engel, T., Sadovskyi, O., Böhm, M., & Heininger, R. (2014). A Conceptual Approach for Optimizing Distribution Logistics using Big Data.

Fan, J., Han, F., & Liu, H. Challenges of Big Data Analysis. National science review. 2014; 1: 293–314.

Feldman, D., Schmidt, M., & Sohler, C. (2013). Turning big data into tiny data: Constant-size coresets for k-means, pca and projective clustering. In *Proceedings of the twenty-fourth annual ACM-SIAM symposium on discrete algorithms* (pp. 1434–1453). SIAM.

Feng, Y., D'Amours, S., & Beauregard, R. (2008). The value of sales and operations planning in oriented strand board industry with make-to-order manufacturing system: Cross functional integration under deterministic demand and spot market recourse. *International Journal of Production Economics, 115*(1), 189-209.

Fernandes, R., Gouveia, B., & Pinho, C. (2013). Integrated inventory valuation in multi-echelon production/distribution systems. *International Journal of Production Research, 51*(9), 2578-2592.

Ferrer, G., Heath, S. K., & Dew, N. (2011). An RFID application in large job shop remanufacturing operations. *International Journal of Production Economics, 133(2),* 612-621.

Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International journal of forecasting, 25(1),* 3-23.

Franczyk, B., (2014).Data science, predictive analytics and big data in logistics, SC design and management. *Universitaet Leipzig, Wirtschaftswissenschaftliche Fakultaet, Institut fur Wirtschaftsinformatic.* Presentation.

Fransoo, J. C., & Wouters, M. J. (2000). Measuring the bullwhip effect in the supply chain. *Supply Chain Management: An International Journal, 5(2),* 78-89.

Frehe, V., Kleinschmidt, T., & Teuteberg, F. (2014). Big Data in Logistics-Identifying Potentials through Literature, Case Study and Expert Interview Analyses. In *GI-Jahrestagung* (pp. 173-186).

Gallucci, J. A., & McCarthy, H. J. (2008). Enhancing the demand planning process with POS forecasting. *The Journal of Business Forecasting*, *27*(4), 11.

Gantz, J., & Reinsel, D. (2011). Extracting value from chaos. *IDC iview*, *1142*(2011), 1-12.

Gartner, J. (2013). Gartner says the Internet of Things installed base will grow to 26 billion units by 2020.

Garvin, D. (1987). Competing on the eight dimensions of quality. *Harv. Bus. Rev.*, 101-109.

Garvin, D. A., & Quality, W. D. P. (1984). Really Mean. *Sloan management review*, *25*.

Geary, S., Disney, S. M., & Towill, D. R. (2006). On bullwhip in supply chains—historical review, present practice and expected future impact. *International Journal of Production Economics, 101(1)*, 2-18.

Geerts, G. L., & O'Leary, D. E. (2014). A supply chain of things: The EAGLET ontology for highly visible supply chains. *Decision Support Systems*, *63*, 3-22.

Ghemawat, S., Gobioff, H., & Leung, S. T. (2003). The Google file system (Vol. 37, No. 5, pp. 29-43). ACM.

Ghosh, D. (2015, September). Big data in logistics and supply chain management-A rethinking step. In *Advanced Computing and Communication (ISACC), 2015 International Symposium on* (pp. 168-173). IEEE.

Giannakis, M., & Louis, M. (2016). A multi-agent based system with big data processing for enhanced supply chain agility. *Journal of Enterprise Information Management*, *29*(5), 706-727.

Gim, J., Hwang, T., Won, Y., & Kant, K. (2015). SmartCon: SmartCon: Smart Context Switching for Fast Storage Devices. *ACM Transactions on Storage (TOS)*, *11*(2), 5.

Gligor, D. M., & Holcomb, M. C. (2012). Understanding the role of logistics capabilities in achieving supply chain agility: a systematic literature review. *Supply Chain Management: An International Journal*, *17*(4), 438-453.

Gligor, D. M., Holcomb, M. C., & Stank, T. P. (2013). A multidisciplinary approach to supply chain agility: Conceptualization and scale development. *Journal of Business Logistics, 34 (2),* 94–108.

Gobble, M. M. (2013). The next big thing in innovation. *Research Technology Management*, 56(1), 64–66.

Goh, R. S. M., Wang, Z., Yin, X., Fu, X., Ponnambalam, L., Lu, S., & Li, X. (2013, August). RiskVis: Supply chain visualization with risk management and real-time

monitoring. In *Automation Science and Engineering (CASE), 2013 IEEE International Conference on* (pp. 207-212). IEEE.

Gomes, P., Farinha, J., Trigueiros, M.J., 2007. A data quality meta-model extension to CWM. *Fourth Asia-Pacific Conference on Conceptual Modeling. Australian Computer Society, Inc.*, 17–26.

Gorman, M.E., 2012, Analytics, viewed 11 September 2012, from http://www.informs.org/Participate-In-a-Community/Societies-and-Sections/Analytics/

Grewal, C. S., Sareen, K. K., & Gill, S. (2008). A multicriteria logistics-outsourcing decision making using the analytic hierarchy process. *International Journal of Services Technology and Management*, *9*(1), 1-13.

Griffiths, J. L., Phelan, A., Osman, K. A., & Furness, A. (2007). Using item-attendant information and communications technologies to improve supply chain visibility.

Guerrero, W. J., Yeung, T. G., & Guéret, C. (2013). Joint-optimization of inventory policies on a multi-product multi-echelon pharmaceutical system with batching and ordering constraints. *European Journal of Operational Research*, *231*(1), 98-108.

Gumus, A. T., Guneri, A. F., & Ulengin, F. (2010). A new methodology for multi-echelon inventory management in stochastic and neuro-fuzzy environments. *International Journal of Production Economics*, *128*(1), 248-260.

Gunasekaran, A., & Ngai, E. W. (2004). Information systems in supply chain integration and management. *European journal of operational research, 159(2),* 269-295.

Guo, C., & Li, X. (2014). A multi-echelon inventory system with supplier selection and order allocation under stochastic demand. *International Journal of Production Economics*, *151*, 37-47.

Guo, Z., Ngai, E., Yang, C., & Liang, X. (2015). An RFID-based intelligent decision support system architecture for production monitoring and scheduling in a distributed manufacturing environment. *International journal of production economics, 159,* 16-28.

Haase, P., Palma, R., & d'Aquin, M. (2007). Updated version of the networked ontology model. *Project Deliverable D*, *1*.

Haberleitner, H., Meyr, H., & Taudes, A. (2010). Implementation of a demand planning system using advance order information. *International journal of production economics*, *128*(2), 518-526.

Halima, M. B., Karray, H., & Alimi, A. M. (2010, September). A comprehensive method for Arabic video text detection, localization, extraction and recognition. In *Pacific-Rim Conference on Multimedia (pp. 648-659).* Springer, Berlin, Heidelberg.

Harris, I., Wang, Y., & Wang, H. (2015). ICT in multimodal transport and technological trends: Unleashing potential for the future. *International Journal of Production Economics*, *159*, 88-103.

Harvey, C. (50). Top open source tools for big data. *Datamation*.

Haug, A., & Stentoft Arlbjørn, J. (2011). Barriers to master data quality. *Journal of Enterprise Information Management*, *24*(3), 288-303.

Haug, A., Stentoft Arlbjørn, J., & Pedersen, A. (2009). A classification model of ERP system data quality. *Industrial Management & Data Systems*, *109*(8), 1053-1068.

Hayya, J. C., Kim, J. G., Disney, S. M., Harrison, T. P., & Chatfield, D. (2006). Estimation in supply chain inventory management. *International Journal of Production Research*, *44*(7), 1313-1330.

Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, *154*, 72-80.

Hazen, B. T., Skipper, J. B., Ezell, J. D., & Boone, C. A. (2016). Big Data and predictive analytics for supply chain sustainability: A theory-driven research agenda. *Computers & Industrial Engineering*, *101*, 592-598.

He, Y., & Zhao, X. (2012). Coordination in multi-echelon supply chain under supply and demand uncertainty. *International Journal of Production Economics*, *139*(1), 106-115.

Heaney, B. (2013). Supply chain visibility: A critical strategy to optimize cost and service. Aberdeen Group, 20.

Heo, E., Kim, J., & Cho, S. (2012). Selecting hydrogen production methods using fuzzy analytic hierarchy process with opportunities, costs, and risks. *International journal of hydrogen energy*, *37*(23), 17655-17662.

Hill, D. (2012). IBM smarter storage: What a smart idea. Mesabi Group Commentary, 1–6.

Hofman, W. (2011, August). Supply chain visibility with linked open data for supply chain risk analysis. In *Workshop on IT Innovations Enabling Seamless and Secure Supply Chains* (pp. 20-31).

Hofmann, E. (2017). Big data and supply chain decisions: the impact of volume, variety and velocity properties on the bullwhip effect. *International Journal of Production Research*, *55*(17), 5108-5126.

Hosoda, S., Ohira, T., & Nakamura, T. (2008). A monthly mean dataset of global oceanic temperature and salinity derived from Argo float observations. *JAMSTEC Report of Research and Development*, *8*, 47-59.

Houlihan, J. B. (1985). International supply chain management. *International journal of physical distribution & materials management*, *15(1)*, 22-38.

Howells, R. (2012), A 360 Degree Perspective On Customer Engagement, Forbes Business, online resource accessed on 20/08/2015 from: http://www.forbes.com/sites/sap/2012/11/13/a-360-degree-perspective-on-customer-engagement/

Howells, R. (2014), The emergence of the demand network, SAP Business Trends, on-line resource accessed 20/09/2015 from: http://scn.sap.com/community/business-trends/blog/2014/02/03/the-emergence-of-the-demand-network

Hu, C., Xu, Z., Liu, Y., Mei, L., Chen, L., & Luo, X. (2014). Semantic link network-based model for organizing multimedia big data. *IEEE Transactions on Emerging Topics in Computing, 2(3),* 376-387.

Huang, Y. Y., & Handfield, R. B. (2015). Measuring the benefits of ERP on supply management maturity model: a "big data" method. *International Journal of Operations & Production Management*, *35*(1), 2-25.

Huddar, M. G., & Ramannavar, M. M. (2013). A survey on big data analytical tools. *International Journal of Latest Trends in Engineering and Technology (IJLTET)*, 85-91.

Huh, Y.U., Keller, F. R. , Redman, T. C., Watkins, A.R., (1990). Data quality. *Inf. Softw. Technol.* 32(8),559–565.

Iacovou, C. L., Benbasat, I., & Dexter, A. S. (1995). Electronic data interchange and small organizations: Adoption and impact of technology. *MIS quarterly*, 465-485.

Institute for Operations Research and Management Science, 2014, *What is analytics?*, viewed 22 October 2014, from https://www.informs.org/About-INFORMS/What-is-Analytics

Ittmann, H. W. (2015). The impact of big data and business analytics on supply chain management. *Journal of Transport and Supply Chain Management*, *9*(1), 1-9.

Ivanov, D., & Sokolov, B. (2010). Conceptual Frameworks for Supply Chain Management. *Adaptive Supply Chain Management*, 19-33.

Jayathilake, D., Sooriaarachchi, C., Gunawardena, T., Kulasuriya, B., & Dayaratne, T. (2012, September). A study into the capabilities of NoSQL databases in handling a highly heterogeneous tree. In *Information and Automation for Sustainability (ICIAfS), 2012 IEEE 6th International Conference on* (pp. 106-111). IEEE.

Jeske, M., Grüner, M., & Weiß, F. (2013). BIG DATA IN LOGISTICS: A DHL perspective on how to move beyond the hype. *DHL Customer Solutions & Innovation*, *12*.

Jharkharia, S., & Shankar, R. (2007). Selection of logistics service provider: An analytic network process (ANP) approach. *Omega*, *35*(3), 274-289.

Ji, C. Q., Li, Y., Qiu, W. M., Jin, Y. W., Xu, Y. J., Awada, U., ... Qu, W. Y. (2012). Big data processing: Big challenges and opportunities. *Journal of Interconnection Networks, 13(03 and 04)*, 1–19.

Ji, C., Li, Y., Qiu, W., Awada, U., & Li, K. (2012, December). Big data processing in cloud computing environments. In *Pervasive Systems, Algorithms and Networks (ISPAN), 2012 12th International Symposium on* (pp. 17-23). IEEE.

Jin, X., Lee, X., Kong, N., & Yan, B. (2008, May). Efficient complex event processing over RFID data stream. In *Computer and Information Science, 2008. ICIS 08. Seventh IEEE/ACIS International Conference on* (pp. 75-81). IEEE.

Jin, X., Zong, S., Li, Y., Wu, S., Yin, W., & Ge, W. (2015). A domain knowledge based method on active and focused information service for decision support within big data environment. *Procedia Computer Science, 60*, 93–102.

Jodlbauer, H. (2008). A time-continuous analytic production model for service level, work in process, lead time and utilization. *International Journal of Production Research*, *46*(7), 1723-1744.

Jones-Farmer, L. A., Ezell, J. D., & Hazen, B. T. (2014). Applying control chart methods to enhance data quality. *Technometrics*, *56*(1), 29-41.

Jonsson, P., & Gustavsson, M. (2008). The impact of supply chain relationships and automatic data communication and registration on forecast information quality. *International journal of physical distribution & logistics management*, *38*(4), 280-295.

Jonsson, P., & Mattsson, S. A. (2008). Inventory management practices and their implications on perceived planning performance. *International journal of production research*, *46*(7), 1787-1812.

Jüttner, U., Christopher, M., & Godsell, J. (2010). A strategic framework for integrating marketing and supply chain strategies. *The International Journal of Logistics Management, 21(1),* 104-126.

Kabak, M., & Burmaoğlu, S. (2013). A holistic evaluation of the e-procurement website by using a hybrid MCDM methodology. *Electronic Government, an International Journal*, *10*(2), 125-150.

Kache, F. (2015). *Dealing with digital information richness in supply chain management: A review and a Big Data Analytics approach* (Vol. 8). kassel university press GmbH.

Kache, F., & Seuring, S. (2014). Linking collaboration and integration to risk and performance in supply chains via a review of literature reviews. *Supply Chain Management: An International Journal, 19(5/6),* 664-682.

Kache, F., & Seuring, S. (2017). Challenges and opportunities of digital information at the intersection of Big Data Analytics and supply chain management. *International Journal of Operations & Production Management, 37(1),* 10-36.

Kagermann, H., Riemensperger, F., Hoke, D., Helbig, J., Stocksmeier, D., Wahlster, W., & Schweer, D. (2014). Smart service welt: recommendations for the strategic initiative web-based services for businesses. *Berlin: Acatech-National Academy of Science and Engineering*.

Kahn, B. K., Strong, D. M., & Wang, R. Y. (2002). Information quality benchmarks: product and service performance. *Communications of the ACM*, *45*(4), 184-192.

Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013, January). Big data: Issues and challenges moving forward. In *System sciences (HICSS), 2013 46th Hawaii international conference on* (pp. 995-1004). IEEE.

Kang, Y. S., Park, I. H., Rhee, J., & Lee, Y. H. (2016). MongoDB-based repository design for IoT-generated RFID/sensor big data. *IEEE Sensors Journal, 16(2),* 485-497.

Kart, L., Heudecker, N., & Buytendijk, F. (2013). Survey analysis: big data adoption in 2013 shows substance behind the hype. *Gartner, 2013b, http://www. gartner. com/survey_analysis_big_data_ado_255160. pdf*.

Kart, L., Heudecker, N., & Buytendijk, F. (2013). Survey analysis: big data adoption in 2013 shows substance behind the hype. *Gartner, 2013b, http://www. gartner. com/survey_analysis_big_data_ado_255160. pdf*.

Katal, A., Wazid, M., & Goudar, R. H. (2013, August). Big data: issues, challenges, tools and good practices. In *Contemporary Computing (IC3), 2013 Sixth International Conference on* (pp. 404-409). IEEE.

Kaur, H., & Singh, S. P. (2018). Heuristic modeling for sustainable procurement and logistics in a supply chain using big data. *Computers & Operations Research*, *98*, 301-321.

Kees, A., Oberländer, A., Röglinger, M., & Rosemann, M. (2015). Understanding the internet of things: A conceptualisation of business-to-thing (B2T) interactions. In *The proceedings of twenty-third European Conference on Information Systems (ECIS)* (pp. 1–16). Germany: Münster.

Kepner, J., Arcand, W., Bergeron, W., Bliss, N., Bond, R., Byun, C., ... & McCabe, A. (2012, March). Dynamic distributed dimensional data model (D4M) database and computation system. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on* (pp. 5349-5352). IEEE.

Kessinger, C. & Pieper, H., 2013. End-to-End Analytics – Supply chain analytics company offers clients analysis, visualization, interaction and insight. *OR/MS Today Aug.*, 20–22

Khan, K. (2013). The transformative power of advanced analytics. *Supply Chain Management Review*, *17*(3), 48-49.

Kim, H. S., & Sohn, S. Y. (2009). Cost of ownership model for the RFID logistics system applicable to u-city. *European Journal of Operational Research, 194(2),* 406-417.

Koh, L. P., Koellner, T., & Ghazoul, J. (2013). Transformative optimisation of agricultural land use to meet future food demands. PeerJ, 1, e188.

Kortuem, G., Kawsar, F., Sundramoorthy, V., & Fitton, D. (2010). Smart objects as building blocks for the internet of things. *IEEE Internet Computing*, *14*(1), 44-51.

Krupnik, Y., (2013). 7 ways predictive analytics helps retailers manage suppliers. Predictive Analytics Times , 26 August 2013, viewed 07 November 2014, from http://www.predictiveanalyticsworld.com/patimes/7-ways-predictive-analytics- helps-retailers

Kuglin, F. A., & Rosenbaum, B. A. (2001). The supply chain network@ Internet speed: Preparing your company for the e-commerce revolution. Amacom Books.

La Londe, B. J., & Masters, J. M. (1994). Emerging logistics strategies: blueprints for the next century. *International journal of physical distribution & logistics management*, *24*(7), 35-47.

La Ponsie, Maryalene, "Data scientists: the hottest you haven't heard of" online https://www.aol.com/2011/08/10/data-scientist-the-hottest-job-you-havent-heard-of/?guccounter=1

Lambert, D. M., Stock, J. R., & Ellram, L. M. (1998). *Fundamentals of logistics management*. McGraw-Hill/Irwin.

Lamming, R. C., Caldwell, N., & Phillips, W. E. N. D. Y. (2004). Supply chain transparency. *New et al*, 191-208.

Lardinois, F. (2014). Google LAUNCHES BigQuery streaming for real-time, big-data analytics : Techcrunch<http://techcrunch.com/2014/03/25/google-launchesbigquery-streaming-for-real-time-big-data-analytics/>.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2013). Big data, analytics and the path from insights to value. *MIT Sloan Management Review, 52 (2),* 21–31.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT sloan management review, 52(2),* 21.

Lee, C.-H., Wang, Y.-H., & Trappey, A. J. (2015). Ontology-based reasoning for the intelligent handling of customer complaints. *Computers & Industrial Engineering, 84,* 144–155.

Lee, H. L., Padmanabhan, V., & Whang, S. (1997). Information distortion in a supply chain: The bullwhip effect. *Management science, 43(4),* 546-558.

Lee, Y. W., Pipino, L., Strong, D. M., & Wang, R. Y. (2004). Process-embedded data integrity. *Journal of Database Management (JDM), 15*(1), 87-103.

Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: a methodology for information quality assessment. *Information & management, 40*(2), 133-146.

Leung, J. Y. T., & Chen, Z. L. (2013). Integrated production and distribution with fixed delivery departure dates. *Operations Research Letters, 41*(3), 290-293.

Leveling, J., Edelbrock, M., & Otto, B. (2014, December). Big data analytics for supply chain management. In *Industrial Engineering and Engineering Management (IEEM), 2014 IEEE International Conference on* (pp. 918-922). IEEE.

Li, B., Wang, H., Yang, J., Guo, M., & Qi, C. (2013). A belief-rule-based inference method for aggregate production planning under uncertainty. *International Journal of Production Research, 51*(1), 83-105.

Li, D. C., & Lin, Y. S. (2006). Using virtual sample generation to build up management knowledge in the early manufacturing stages. *European Journal of Operational Research, 175(1),* 413-434.

Li, J., Tao, F., Cheng, Y., & Zhao, L. (2015). Big data in product lifecycle management. *The International Journal of Advanced Manufacturing Technology, 81(1–4),* 667–684.

Li, X., Tian, P., & Leung, S. C. (2010). Vehicle routing problems with time windows and stochastic travel and service times: Models and algorithm. *International Journal of Production Economics*, *125*(1), 137-145.

Lim, L. L., Alpan, G., & Penz, B. (2014). Reconciling sales and operations management with distant suppliers in the automotive industry: a simulation approach. *International Journal of Production Economics*, *151*, 20-36.

Lin, C. Y., & Ho, Y. H. (2009). RFID technology adoption and supply chain performance: an empirical study in China's logistics industry. *Supply Chain Management: An International Journal, 14(5),* 369-378.

Lin, X., & Zheng, X. (2013, May). A cloud-enhanced system architecture for logistics tracking services. In *Proceedings of the International Conference on Computer, Networks and Communication Engineering (ICCNCE'13)*.

Link, B., & Back, A. (2015). Classifying systemic differences between software as a service-and on-premise-enterprise resource planning. *Journal of Enterprise Information Management*, *28*(6), 808-837.

Liu, Z., Chua, D. K. H., & Yeoh, K. W. (2011). Aggregate production planning for shipbuilding with variation-inventory trade-offs. *International Journal of Production Research*, *49*(20), 6249-6272.

Logan, M. S. (2000). Using agency theory to design successful outsourcing relationships. *The International Journal of Logistics Management, 11(2),* 21-32.

LOGISTICS, B. D. I. (2013). A DHL perspective on how to move beyond the hype. *DHL December*.

Lohman, T. (2013). Big data and analytics trends for 2014 : ZDNet<http://www.zdnet.com/big-data-and-analytics-trends-for-2014-7000024260/>.

Loshin, D. (2013). *Big data analytics: from strategic planning to enterprise integration with tools, techniques, NoSQL, and graph*. Elsevier.

Lu, C. J., & Wang, Y. W. (2010). Combining independent component analysis and growing hierarchical self-organizing maps with support vector regression in product demand forecasting. *International Journal of Production Economics*, *128*(2), 603-613.

Lu, T., Guo, X., Xu, B., Zhao, L., Peng, Y., & Yang, H. (2013, September). Next big thing in big data: the security of the ICT supply chain. In *Social Computing (SocialCom), 2013 International Conference on* (pp. 1066-1073). IEEE.

Machuca, J. A., & Barajas, R. P. (2004). The impact of electronic data interchange on reducing bullwhip effect and supply chain inventory costs. *Transportation Research Part E: Logistics and Transportation Review, 40(3),* 209-228.

March, S. T., & Hevner, A. R. (2007). Integrated decision support systems: A data warehousing perspective. *Decision Support Systems*, *43*(3), 1031-1043.

Markham, S. K., Kowolenko, M., & Michaelis, T. L. (2015). Unstructured text analytics to support new product development decisions. *Research-Technology Management, 58(2),* 30-39.

Marr, B. (2013). The awesome ways big data is used today to change our world<https://www.linkedin.com/today/post/article/20131113065157-64875646-the-awesome-ways-big-data-is-used-today-to-change-our-world>.

Marr, B. (2013). The awesome ways big data is used today to change our world<https://www.linkedin.com/today/post/article/20131113065157-64875646-the-awesome-ways-big-data-is-used-today-to-change-our-world>.

Marz, N., & Warren, J. (2015). *Big Data: Principles and best practices of scalable real-time data systems*. New York; Manning Publications Co..

Mason, H. (2014). Inspiration day at the University of Waterloo Stratford, Campus<http://www.betakit.com/event/inspiration-day-at-university-ofwaterloo-stratford-campus/>.

Mason-Jones, R., & Towill, D. R. (1997). Information enrichment: designing the supply chain for competitive advantage. *Supply Chain Management: An International Journal, 2(4),* 137-148.

Mayer-Schnberger, V., & Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. *Houghton Mifflin Harcourt*.

McKinsey Global Institute (2013). Disruptive technologies: Advances that will transform life, business, and the global economy (Accessed on 24.10.2015) <https://www.sommetinter.coop/sites/default/files/etude/files/report_mckinsey_technology_0.pdf>.

Meixell, M. J., & Wu, S. D. (2001). Scenario analysis of demand in a technology market using leading indicators. *IEEE Transactions on Semiconductor Manufacturing, 14(1),* 65-75.

Mentzer, J. T., DeWitt, W., Keebler, J. S., Min, S., Nix, N. W., Smith, C. D., & Zacharia, Z. G. (2001). Defining supply chain management. *Journal of Business logistics*, *22*(2), 1-25.

Metters, R. (1997). Quantifying the bullwhip effect in supply chains. *Journal of operations management, 15(2),* 89-100.

Michaelides, Z. (2016). Big data for logistics and supply chain management. In *Production and Operations Management Society (POMS) Conference Proceedings in Orlando, Florida*.

Milan, P. (2015). Use big data to help procurement make a real difference <http://www.4cassociates.com>.

Min, C., Shiwen, M., & Yunhao, L. (2014). Big data: A survey. *Mobile Network Applications, 19,* 171–209.

Minis, I., & Tatarakis, A. (2011). Stochastic single vehicle routing problem with delivery and pick up and a predefined customer sequence. *European journal of operational research*, *213*(1), 37-51.

Mirzapour Al-E-Hashem, S. M. J., Malekly, H., & Aryanezhad, M. B. (2011). A multi-objective robust optimization model for multi-product multi-site aggregate production planning in a supply chain under uncertainty. *International Journal of Production Economics*, *134*(1), 28-42.

Mishra, A. N., Devaraj, S., & Vaidyanathan, G. (2013). Capability hierarchy in electronic procurement and procurement process performance: An empirical analysis. *Journal of Operations Management*, *31*(6), 376-390.

Moyaux, T., Chaib-Draa, B., & D'Amours, S. (2006). Supply chain management and multiagent systems: an overview. In *Multiagent based supply chain management* (pp. 1-27). Springer, Berlin, Heidelberg.

Najafi, M., Eshghi, K., & Dullaert, W. (2013). A multi-objective robust optimization model for logistics planning in the earthquake response phase. *Transportation Research Part E: Logistics and Transportation Review*, *49*(1), 217-249.

Nativi, J. J., & Lee, S. (2012). Impact of RFID information-sharing strategies on a decentralized supply chain with reverse logistics operations. *International Journal of Production Economics*, *136(2*), 366-377.

Neaga, I., Liu, S., Xu, L., Chen, H., & Hao, Y. (2015, May). Cloud enabled big data business platform for logistics services: a research and development agenda. In *International Conference on Decision Support System Technology* (pp. 22-33). Springer, Cham.

New, S. (2010). The transparent supply chain. *Harvard Business Review*, *88*, 1-5.

New, S. J. (2015). Modern slavery and the supply chain: the limits of corporate social responsibility?. *Supply Chain Management: An International Journal*, *20*(6), 697-707.

Ngai, E. W. T., Moon, K. K., Riggins, F. J., & Candace, Y. Y. (2008). RFID research: An academic literature review (1995–2005) and future research directions. *International Journal of Production Economics, 112(2),* 510-520.
Ngai, E. W., & Gunasekaran, A. (2007). A review for mobile commerce research and applications. *Decision support systems*, *43*(1), 3-15.

Nguyen, T., Li, Z. H. O. U., Spiegler, V., Ieromonachou, P., & Lin, Y. (2018). Big data analytics in supply chain management: A state-of-the-art literature review. *Computers & Operations Research*, *98*, 254-264.

Nouri Koupaei, M., Mohammadi, M., & Naderi, B. (2016). An Integrated Enterprise Resources Planning (ERP) Framework forFlexible Manufacturing SystemsUsing Business Intelligence (BI) Tools. *International Journal of Supply and Operations Management*, *3*(1), 1112-1125.

Novoa, C., Storer, R., (2009). An approximate dynamic programming approach for the vehicle routing problem with stochastic demands. *European Journal of Operational Research, 196(2),* 509- 515.

Noyes, A., Godavarti, R., Titchener-Hooker, N., Coffman, J., & Mukhopadhyay, T. (2014). Quantitative high throughput analytics to support polysaccharide production process development. *Vaccine*, *32*(24), 2819-2828.
Opera Solutions blog. June.

Opresnik, D., & Taisch, M. (2015). The value of big data in servitization. *International Journal of Production Economics, 165,* 174-184.

Ordenes, F. V., Theodoulidis, B., Burton, J., Gruber, T., & Zaki, M. (2014). Analyzing customer experience feedback using text mining: A linguistics-based approach. *Journal of Service Research, 17(3),* 278-295.

Oruezabala, G., & Rico, J. C. (2012). The impact of sustainable public procurement on supplier management—The case of French public hospitals. *Industrial Marketing Management*, *41*(4), 573-580.

Özdamar, L., & Demir, O. (2012). A hierarchical clustering and routing procedure for large scale disaster relief logistics planning. *Transportation Research Part E: Logistics and Transportation Review*, *48*(3), 591-602.

Papadopoulos, T., Gunasekaran, A., Dubey, R., Altay, N., Childe, S. J., & Fosso-Wamba, S. (2017). The role of Big Data in explaining disaster resilience in supply chains for sustainability. *Journal of Cleaner Production*, *142*, 1108-1118.

Parashar, M. (2014, June). Big data challenges in simulation-based science. In *DICT@ HPDC* (pp. 1-2).

Parssian, A. (2006). Managerial decision support with knowledge of accuracy and completeness of the relational aggregate functions. *Decision Support Systems*, *42*(3), 1494-1502.

Pfisterer, D., Romer, K., Bimschas, D., Kleine, O., Mietz, R., Truong, C., ... & Karnstedt, M. (2011). SPITFIRE: toward a semantic web of things. *IEEE Communications Magazine*, *49*(11), 40-48.

Pijanowski, B. C., Tayyebi, A., Doucette, J., Pekin, B. K., Braun, D., & Plourde, J. (2014). A big data urban growth simulation at a national scale: configuring the GIS and neural network based land transformation model to run in a high performance computing (HPC) environment. *Environmental Modelling & Software*, *51*, 250-268.

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, *45*(4), 211-218.

Poon, T. C., Choy, K. L., Chow, H. K., Lau, H. C., Chan, F. T., & Ho, K. C. (2009). A RFID case-based logistics resource management system for managing order-picking operations in warehouses. *Expert Systems with Applications, 36(4),* 8277-8301.

Publications<http://www.mckinsey.com/insights/business_technology/

Puget, R., & Baskiotis, N. (2015, October). Hierarchical label partitioning for large scale classification. In *Data Science and Advanced Analytics (DSAA), 2015. 36678 2015. IEEE International Conference on* (pp. 1-10). IEEE.

Quinn, J. B. (1981). Formulating strategy one step at a time. *The journal of business strategy, 1(3),* 42.

Radke, A. M., & Tseng, M. M. (2015). Design considerations for building distributed supply chain management systems based on cloud computing. *Journal of Manufacturing Science and Engineering, 137(4),* 040906.

Rajapaksha, S. K., & Kodagoda, N. (2008, December). Internal structure and semantic web link structure based ontology ranking. In *Information and Automation for Sustainability, 2008. ICIAFS 2008. 4th International Conference on* (pp. 86-90). IEEE.

Ramos, L. (2015). Semantic Web for manufacturing, trends and open issues: Toward a state of the art. *Computers & Industrial Engineering, 90,* 444–460.

Ranjan, R. (2014). Modeling and simulation in performance optimization of big data processing frameworks. *IEEE Cloud Computing*, *1*(4), 14-19.

Redman, T. C., & Blanton, A. (1997). *Data quality for the information age*. Artech House, Inc..

Reichhart, A., & Holweg, M. (2007). Creating the customer-responsive supply chain: a reconciliation of concepts. *International Journal of Operations & Production Management*, *27*(11), 1144-1172.

Richey Jr, R. G., Morgan, T. R., Lindsey-Hall, K., & Adams, F. G. (2016). A global exploration of big data in the supply chain. *International Journal of Physical Distribution & Logistics Management*, *46*(8), 710-739.

Richtárik, P., & Takácˇ, M. (2015). Parallel coordinate descent methods for big data optimization. *Mathematical Programming, 156(1),* 1–52.

Robak, S., Franczyk, B., & Robak, M. (2013, September). Applying big data and linked data concepts in supply chains management. In *Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on* (pp. 1215-1221). IEEE.

Robak, S., Franczyk, B., & Robak, M. (2014). Research Problems Associated with Big Data Utilization in Logistics and Supply Chains Design and Management. In *FedCSIS Position Papers* (pp. 245-249).

Ronen, B., Spiegler, I., (1991). Information as inventory: a new conceptual view. *Inf. Manage.* 21(4),239–247.

Rosemann, M. (2014). The internet of things – New digital capital in the hand of customers. *Business Transformation Journal, 9(1),* 6–14.
Ross, J. W., Beath, C. M., & Quaadgras, A. (2013). You may not need big data after all. *Harvard Business Review, 91(12),* 90-+.

Rouse, M. (2012). Big data analytics. Essential guide <http://searchbusinessanalytics.techtarget.com/definition/big-data-analytics>.

Rozados, I. V., & Tjahjono, B. (2014, December). Big data analytics in supply chain management: Trends and related research. In *6th International Conference on Operations and Supply Chain Management, Bali.*

Sage, (2013). Better inventory management: Big Challenges, Big Data, Emerging Solutions. http://na.sage.com//media/site/erp/responsive/resources/Sage-ERP-Better-Inventory-Management-wp.pdf. Accessed on April 20, 2015.

Sale, P. F., Agardy, T., Ainsworth, C. H., Feist, B. E., Bell, J. D., Christie, P., ... & Daw, T. M. (2014). Transforming management of tropical coastal seas to cope with challenges of the 21st century. *Marine Pollution Bulletin, 85(1),* 8-23.

Salehan, M., & Kim, D. J. (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems, 81,* 30–40.

Sambamurthy, V., Bharadwaj, A., & Grover, V. (2003). Shaping agility through digital options: Reconceptualizing the role of information technology in contemporary firms. *MIS quarterly*, 237-263.

Sanders, N. R. (2016). How to use big data to drive your supply chain. *California Management Review, 58(3),* 26-48.

Sanyal, J., & New, J. (2013, May). Simulation and big data challenges in tuning building energy models. In *Modeling and Simulation of Cyber-Physical Energy Systems (MSCPES), 2013 Workshop on* (pp. 1-6). IEEE.

Sarac, A., Absi, N., & Dauzère-Pérès, S. (2010). A literature review on the impact of RFID technologies on supply chain management. *International Journal of Production Economics, 128(1),* 77-95.
Sarac, A., Absi, N., & Dauzère-Pérès, S. (2010). A literature review on the impact of RFID technologies on supply chain management. *International Journal of Production Economics, 128(1),* 77-95.

Sari, K. (2010). Exploring the impacts of radio frequency identification (RFID) technology on supply chain performance. *European Journal of Operational Research, 207(1),* 174-183.

SAS, 2013, Advanced analytics, viewed 01 February 2013, from http://www.sas.com/ technologies/analytics/

Sawik, T. (2009). Coordinated supply chain scheduling. *International Journal of Production Economics, 120*(2), 437-451.

Sayahi, S., & Halima, M. B. (2014, August). An intelligent and robust multi-oriented image scene text detection. In *Soft Computing and Pattern Recognition (SoCPaR), 2014 6th International Conference of* (pp. 418-422). IEEE.

Scannapieco, M., & Catarci, T. (2002). Data quality under a computer science perspective. *Archivi & Computer, 2,* 1-15.

Schoenherr, T., & Speier-Pero, C. (2015). Data science, predictive analytics, and big data in supply chain management: Current state and future potential. *Journal of Business Logistics, 36(1),* 120-132.

Seah, M., Hsieh, M. H., & Weng, P. D. (2010). A case analysis of Savecom: The role of indigenous leadership in implementing a business intelligence system. *International journal of information management, 30(4),* 368-373.

Sharma, S., & Agrawal, N. (2012). Application of fuzzy techniques in a multistage manufacturing system. *The International Journal of Advanced Manufacturing Technology*, *60*(1-4), 397-407.

Shilling, H., (2013.). Big Data Takes the Travel Industry in New Direction.

Shmueli, E., Singh, V. K., Lepri, B., & Pentland, A. (2014). Sensing, understanding, and shaping social behavior. I*EEE Transactions on Computational Social Systems, 1(1),* 22-34.

Shmueli, G., & Koppius, O. R. (2011). Predictive analytics in information systems research. *Mis Quarterly,* 553-572.

Short, J. C., Ketchen Jr, D. J., Palmer, T. B., & Hult, G. T. M. (2007). Firm, strategic group, and industry influences on performance. *Strategic management journal, 28(2),* 147-167.

Singh, J., Powles, J., Pasquier, T., & Bacon, J. (2015). Data flow management and compliance in cloud computing. *IEEE Cloud Computing*, *2*(4), 24-32.

Slavakis, K., Giannakis, G. B., & Mateos, G. (2014). Modeling and optimization for big data analytics:(statistical) learning tools for our era of data deluge. *IEEE Signal Processing Magazine*, *31*(5), 18-31.

Sodhi, M. S., & Tang, C. S. (2011). Determining supply requirement in the sales-and-operations-planning (S&OP) process under demand uncertainty: a stochastic programming formulation and a spreadsheet implementation. *Journal of the Operational Research Society*, *62*(3), 526-536.

Souza, G. C. (2014). Supply chain analytics. *Business Horizons*, *57*(5), 595-605.

Spiegel, J., McKenna, M., Lakshman, G., & Nordstrom, P. (2014). Amazon us patent anticipatory shipping. *Amazon Technologies Inc*, *12*.

Stateczny, A., & Wlodarczyk-Sielicka, M. (2014, July). Self-organizing artificial neural networks into hydrographic big data reduction process. In *International Conference on Rough Sets and Intelligent Systems Paradigms* (pp. 335-342). Springer, Cham.

Stedman, C. (2014). Enterprises take a long view on big data programs and purchases. *TechTarget Search Business Analytics*.

Stevens, G. C., & Johnson, M. (2016). Integrating the supply chain… 25 years on. *International Journal of Physical Distribution & Logistics Management*, *46*(1), 19-42.

Swafford, P. M., Ghosh, S., & Murthy, N. (2006). The antecedents of supply chain agility of a firm: scale development and model testing. *Journal of Operations Management*, *24*(2), 170-188.

Swafford, P. M., Ghosh, S., & Murthy, N. (2008). Achieving supply chain agility through IT integration and flexibility. *International Journal of Production Economics*, *116*(2), 288-297.

Swaminathan, S. (2012). The effects of big data on the logistics industry. *Profit Oracle*.

Tachizawa, E. M., & Wong, C. Y. (2015). The performance of green supply chain management governance mechanisms: A supply network and complexity perspective. *Journal of Supply Chain Management, 51(3),* 18-32.

Talia, D. (2013). Toward cloud-based big-data analytics. IEEE Computer Science, 98–101.

Tan, K. H., Zhan, Y., Ji, G., Ye, F., & Chang, C. (2015). Harvesting big data to enhance supply chain innovation capabilities: An analytic infrastructure based on deduction graph. *International Journal of Production Economics*, *165*, 223-233.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic management journal*, *18*(7), 509-533.

Tenhiälä, A., Rungtusanatham, M. J., & Miller, J. W. (2018). ERP System versus Stand-Alone Enterprise Applications in the Mitigation of Operational Glitches. *Decision Sciences*, *49*(3), 407-444.

Tenhiala. A. and Helkio. P., (2015). Performance effects of using an ERP system for manufacturing planning and control under dynamic market requirements, *Journal of Operations Management36, pp.* 147–164.

Thalbauer, H. (2014), The Time to Move to a Demand Network Has Come, on-line resource accessed 20/9/2015 from: http://www.sap.com/bin/sapcom/downloadasset.the-time-to-move-to-a-demand-network-has-come-pdf.bypassReg.html

Tien, J. M. (2012). The next industrial revolution: Integrated services and goods. *Journal of Systems Science and Systems Engineering, 21(3),* 257-296.

Trkman, P., McCormack, K., De Oliveira, M. P. V., & Ladeira, M. B. (2010). The impact of business analytics on supply chain performance. *Decision Support Systems, 49(3),* 318-327.

ur Rehman, M. H., Chang, V., Batool, A., & Wah, T. Y. (2016). Big data reduction framework for value creation in sustainable enterprises. *International Journal of Information Management, 36(6),* 917-928.

Van Weele, A. J. (2000). *Purchasing & supply chain management: analysis, strategy, planning and practice*. Cengage Learning EMEA.

Vance, A. (2009). Hadoop, a free software program, finds uses beyond search. *New York Times. March, 16*.

Veldman, J., & Gaalman, G. (2014). A model of strategic product quality and process improvement incentives. *International journal of production economics, 149,* 202-210.

Vera-Baquero, A., Colomo-Palacios, R., & Molloy, O. (2013). Business process analytics using a big data approach. *IT Professional,* 1–9.

Vidal, T., Crainic, T. G., Gendreau, M., & Prins, C. (2013). Heuristics for multi-attribute vehicle routing problems: A survey and synthesis. *European Journal of Operational Research*, *231*(1), 1-21.

Viegas, J., (2013). Big data and transport. *International Transport Forum.* October.

Walker, H., & Brammer, S. (2012). The relationship between sustainable procurement and e-procurement in the public sector. *International Journal of Production Economics*, *140*(1), 256-268.

Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *Journal of Business Logistics*, *34*(2), 77-84.

Wamba, S. F., & Chatfield, A. T. (2009). A contingency model for creating value from RFID supply chain network projects in logistics and manufacturing environments. *European Journal of Information Systems*, *18(6)*, 615-636.

Wamba, S. F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How 'big data'can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, *165*, 234-246.

Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, *39*(11), 86-95.

Wang, G., & Lei, L. (2012). Polynomial-time solvable cases of the capacitated multi-echelon shipping network scheduling problem with delivery deadlines. *International Journal of Production Economics*, *137*(2), 263-271.

Wang, G., & Lei, L. (2015). Integrated operations scheduling with delivery deadlines. *Computers & Industrial Engineering*, *85*, 177-185.

Wang, G., Gunasekaran, A., Ngai, E. W., & Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics*, *176*, 98-110.

Wang, R. Y. (1993). *Information technology in action: trends and perspectives*. Prentice-Hall, Inc..

Wang, R. Y. (1998). A product perspective on total data quality management. *Communications of the ACM, 41(2),* 58-65.

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, *12*(4), 5-33.

Wang, R. Y., Storey, V. C., & Firth, C. P. (1995). A framework for analysis of data quality research. *IEEE Transactions on Knowledge & Data Engineering*, (4), 623-640.

Wang, R., Liang, T., (2005). Applying possibilistic linear programming to aggregate production planning. *International Journal of Production Economics, 98(3),* 328-341

Wang, S. J., Liu, S. F., & Wang, W. L. (2008). The simulated impact of RFID-enabled supply chain on pull-based inventory replenishment in TFT-LCD industry. *International Journal of Production Economics, 112(2),* 570-586.

Watson, M. (2013). *Supply chain network design: Applying optimization and analytics to the global supply chain*. Pearson Education.

Watts, S., Shankaranarayanan, G., & Even, A. (2009). Data quality assessment in context: A cognitive perspective. *Decision Support Systems*, *48*(1), 202-211.

Wieland, A., & Wallenburg, C. M. (2011). *Supply-chain-management in stürmischen Zeiten*. Univ.-Verlag der TU.

Windt, K., Böse, F., & Philipp, T. (2008). Autonomy in production logistics: Identification, characterisation and application. *Robotics and Computer-Integrated Manufacturing, 24(4),* 572-578.

Witkowski, K. (2017). Internet of things, big data, industry 4.0–Innovative solutions in logistics and supply chains management. *Procedia Engineering*, *182*, 763-769.

Wu, D. Y., & Katok, E. (2006). Learning, communication, and the bullwhip effect. *Journal of operations management, 24(6),* 839-850.

Wu, K. J., Liao, C. J., Tseng, M. L., Lim, M. K., Hu, J., & Tan, K. (2017). Toward sustainability: using big data to explore the decisive attributes of supply chain risks and uncertainties. *Journal of Cleaner Production, 142,* 663-676.

Wu, W., Liu, Z., Chen, M., Yang, X., & He, X. (2012). An automated vision system for container-code recognition. *Expert Systems with Applications, 39*(3), 2842-2855.

Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering, 26(1),* 97–107.

Xu, J., & Güting, R. H. (2013). A generic data model for moving objects. *Geoinformatica, 17(1),* 125–172.

Yao, Y., Evers, P. T., & Dresner, M. E. (2007). Supply chain integration in vendor-managed inventory. *Decision support systems, 43(2),* 663-674.

Yeniyurt, S., Henke Jr, J. W., & Cavusgil, E. (2013). Integrating global and local procurement for superior supplier working relations. *International Business Review, 22*(2), 351-362.

Zage, D., Glass, K., & Colbaugh, R. (2013, June). Improving supply chain security using big data. In *Intelligence and Security Informatics (ISI), 2013 IEEE International Conference on* (pp. 254-259). IEEE.

Zakir, J., Seymour, T., & Berg, K. (2015). Big data analytics. *Issues in Information Systems, 16(2),* 81–90.

Zalewski, R. (2010). Innowacje odwrotne. Towaroznawstwo 2.0 w działaniu na rzecz innowacji [Reverse innovation, Commodities 2.0 in action for innovation]. *Poznań: Wyd. Komisja Nauk Towaroznawczych PAN Oddział w Poznaniu*.

Zalewski, R. I. (2010). Czy nowe metody generowania innowacji ożywią współpracę nauka-przemysł. Marketing i Rynek, grudzień.

Zeithaml, V. A., Parasuraman, A., Berry, L. L., & Berry, L. L. (1990). *Delivering quality service: Balancing customer perceptions and expectations*. Simon and Schuster.

Zhao, R., Liu, Y., Zhang, N., & Huang, T. (2017). An optimization model for green supply chain management by using a big data analytic approach. *Journal of Cleaner Production*, *142*, 1085-1097.

Zhong, R. Y., Dai, Q. Y., Qu, T., Hu, G. J., & Huang, G. Q. (2013). RFID-enabled real-time manufacturing execution system for mass-customization production. *Robotics and Computer-Integrated Manufacturing, 29(2),* 283-292.

Zhong, R. Y., Huang, G. Q., & Dai, Q. (2014, May). A big data cleansing approach for n-dimensional RFID-Cuboids. In *Computer Supported Cooperative Work in Design (CSCWD), Proceedings of the 2014 IEEE 18th International Conference on* (pp. 289-294). IEEE.

Zhong, R. Y., Huang, G. Q., Dai, Q. Y., & Zhang, T. (2014). Mining SOTs and dispatching rules from RFID-enabled real-time shopfloor production data. *Journal of Intelligent Manufacturing, 25(4),* 825-843.

Zhong, R. Y., Huang, G. Q., Lan, S., Dai, Q. Y., Chen, X., & Zhang, T. (2015). A big data approach for logistics trajectory discovery from RFID-enabled production data. *International Journal of Production Economics*, *165*, 260-272.

Zhong, R. Y., Newman, S. T., Huang, G. Q., & Lan, S. (2016). Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives. *Computers & Industrial Engineering*, *101*, 572-591.

Zhong, R. Y., Xu, C., Chen, C., & Huang, G. Q. (2015). Big data analytics for physical internet-based intelligent manufacturing shop floors. *International Journal of Production Research*, 1–12.

Zikopoulos, P., & Eaton, C. (2011). *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media.